Daresbury Laboratory

JOINT CCP4 AND ESF-EACBM NEWSLETTER ON PROTEIN CRYSTALLOGRAPHY

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

Number 28 May 1993

Contents

CCP4 News ...2
D. Love

New German Post Codes ...3

Beamtime at HASYLAB in 1993 ...4

The structure determination of yeast peroxisomal thiolase ...5
M. Mathieu & R.K. Wierenga

Deconvoluting Laue Multiple Diffraction Spots by a Real-Space Density Modification Method ...8
Q. Hao & M.M. Harding

Density measurement of crystals of *Bacillus subtilis* lipase after staining with Coomassie Brilliant Blue using the Ficoll density gradient method ...9
M. Blaauw, S. Ransac & B.W. Dijkstra

Recent Activity at Pavia: Superoxide Dismutases and Myoglobins ..12
K. Djinovic-Carugo, M. Rizzi & M. Bolognesi

Protein Crystallographic measurements on BL9 (Troika) of the ESRF, February-March 1993 ..14
M.S. Lehmann, J. Als-Nielsen, G. Grübel & J.F. Legrand

Applications of Laue Diffraction and other small Molecule Studies with Synchrotron Radiation ..22
G.M.T. Cheetham, I.M. Dodd, M. Harding, B.M. Kariuki & H. Quan

Integrated Data and Knowledge Base for Protein Structure and Sequence: The European Efforts ..24
S. Wodak

The Effect of Cryogenic Treatment on the Cell Dimensions of Ribosomal Crystals ..26
J. Harms, F. Schlünzen, K. v. Böhlen, J. Thygesen, S. Meyer, I. Dunkel, B. Donzelmann, H.A.S. Hansen, A. Zaytzev-Bashan, A. Dribin, G. Kryger, G. Thoms, N. Volkmann, H. Bartels, W.S. Bennett & A. Yonatt

Editor: Keith S. Wilson

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

Location of the redox centers in hydrogenase as determined by X-ray crystallography at 5 Å resolution ..30
A. Volbeda, C. Piras, M.H. Charon, E.C. Hatchikian, M. Frey & J.C. Fontecilla-Camps

Progress on the crystal structure of the glycyl tRNA synthetase from *Thermus thermophilus* ..34
D. Logan, V. Cura, D. Kern & D. Moras

Compression of X-ray images ..39
J.P. Abrahams

Macromolecular Crystallography at the SRS ..42
C. Nave, M. Papiz, P. Rizkallah, S. McSweeney, J. Campbell, S. Kinder, P. Atkinson & P. Lindley

SRS Specialist User Group for Laue Diffraction - April, 1993 ..46
J. Campbell & B. Kariuki

Structural studies of some transthyretin variants ..49
A.M. Damas

Crystal Structure of liganded and unliganded Forms of Plasma Retinol-Binding Protein ..52
G. Zanotti, R. Berni, S. Ottonello & H.L. Monaco

Computing Difference-Patterson Maps with X-PLOR ..55
C. Vonrhein

Masks Made Easy ..56
G.J. Kleywegt & T.A. Jones

Simple example of the molecular replacement technique: The structure determination of 3-phosphglycerate kinase from *Bacillus stearothermophilus* ..60
G.J. Davies

Minutes of the Working Group 1 Meeting of CCP4 held at Daresbury on January 29, 1993 ..68

VIIIth European Symposium on Materials and Fluid Sciences in Microgravity, Brussels 12-16 April, 1993 ..72

Data Collection Strategy on the MAR Image Plate Scanner ..74
P.A. Tucker

LURE Synchrotron Radiation Facility (Orsay, France) Instruments for macromolecular crystallography ..77
R. Fourme

2.8 Å Crystal structure of Catalase HPII from *Escherichia Coli* ..79
J. Bravo, J. Tormo, N. Verdaguer, I. Fita, J. Switala & P.C. Loewen

Positions available ..82

ESRF USER'S MEETING, 24-25 September 1993 ..86

A Meeting on Complex Structures and their Functional Implications, 7-11 October 1993 ..88

CCP4 News

Dave Love
SERC Daresbury Laboratory
Warrington WA4 4AD, UK
d.love@dl.ac.uk

New CCP4 Secretary

Wojciech Wolf resigned as CCP4 secretary in November '92 to return to Poland, where we wish him well. CCP4 users will doubtless be grateful for his efforts on your behalf during his time at Daresbury. Dr. Susan Bailey (currently at Manchester University) has been appointed to replace him and is due to take up the position at the end of April '93. She may be familiar to some as she worked at Daresbury for CCP4 in the past.

Progress with the program suite

A new major release of the suite is now overdue but unfortunately it has not been possible to make another since the last newsletter. Several minor releases have been made and announced over the network to correct bugs and configuration problems; the latest includes a working configuration for HP-UX. The suite is reported to work fast on DEC Alpha machines, running both VMS and OSF/1 and we are currently trying to establish what changes, if any, need to be made for this.

A good deal of work has been done towards the next release, however. This includes separating the suite into 'supported' and 'unsupported' programs and dropping quite a number that were obsolete or duplicated the functionality of others. Please note that the preliminary lists of programs in these categories given the last newsletter have changed somewhat. Major new additions currently finished or being worked on include versions of Jorge Navaza's molecular replacement package `AMoRe`, Victor Lamzin's automatic refinement system and the UCL/Birkbeck `procheck` package for stereochemistry checking. A much-expanded manual is being worked on and there has been a lot of cleaning up done.

Following acquisition of a workstation for CCP4 use we will shortly be able to offer an improved service for distribution by Internet ftp and, eventually, we hope, an improved e-mail service for the CCP4BB list. The ftp service is currently awaiting final approval from the local network police and will be announced on CCP4BB when available; more will be available, including access to individual files in the source tree.

1992 Study Weekend Proceedings

A considerable number of attendees have not received their copy of the 1992 CCP4 Study Weekend on Molecular Replacement. It's not currently clear why not, but if you attended and haven't received a copy, please contact the CCP4 secretary at Daresbury unless you have already asked.

IMPORTANT

Please note our new post code valid from 1st July 1993

European Molecular Biology Laboratory
c/o DESY
Notkestrasse 85, Geb. 25 A
Letters to be addressed to: **22603** Hamburg
Shipments to be addressed to: **22607** Hamburg
Telephone: 040/89902-0
Telex. 215 124 (desy d)
Fax: 040/89902-149

BEAM-TIME AT HASYLAB 1993

In principle the DORIS ring has now re-started operation as a dedicated source of Synchrotron Radiation.

The EMBL beam-lines are scheduled to be available to users from July onward subject to satisfactory operation of DORIS. Please send requests for beam proposal forms for Protein Crystallography, Small Angle Scattering and EXAFS to

Dr. Keith S. Wilson
E M B L
c/o DESY
Notkestraße 85
22603 Hamburg

Tel.: 0049-40-899020
FAX:0049-40-89902149
E-mail:Keith@EMBL-Hamburg.DE

A priorities committee to assess proposals will be held later in the year.

The structure determination of yeast peroxisomal thiolase

M.Mathieu & R.K.Wierenga
EMBL, Meyerhofstrasse 1, D-6900 Heidelberg, Germany

Thiolases are ubiquitous enzymes involved in various different metabolic pathways. Several types of thiolases have been observed in eukaryotic and prokaryotic cells with differences in substrate specificity and intracellular localization.

We are interested in 3-ketoacyl-CoA thiolase from the peroxisome of yeast Saccharomyces cerevisae, a dimeric protein of 42 kd and 417 amino-acids per subunit. This protein functions only in the peroxisome and catalyses the thiolytic cleavage of 3-ketoacylCoA.

It is active as a dimer, and crystallizes as a dimer in the space group $P2_12_12_1$, using the microdialysis method (1).

A first dataset was collected to 2.7 Å, and we used it with the GLRF package (2) to find the local two-fold axis of that dimer.

We first used the self-rotation option, for a kappa of 180°. Different resolutions shells were tried, different protein radii and different cut-off values for the intensities of the reflections. One consistent peak was found at 1.8 sigma above the background.

Since we had collected at the time only one (poor) derivative ($KAu(CN)_2$), we used the SIRAS phases for the phased translation option of GLRF (3). That derivative gave a phased dataset till 4 Å, with figures of merit (centric and acentric respectively) of 0.57 and 0.37. That was not enough to find the position of the 2-fold axis.

Another derivative was eventually found, $KAuCl_4$, which did bind to the same sites as observed for $KAu(CN)_2$, but with higher occupancy. Combining the two derivatives (MIRAS) resulted in a figure of merit of 0.63 and 0.41 at 4 Å, and 0.52 and 0.34 at 3 Å. We used again option tfun2 of GLRF, at 5, 4, and 3Å, with datasets phased to the corresponding resolution, for each enantiomeric form of the heavy atom sites. We did not have any result at 5 Å, but we could see a consistent peak at 4 and 3 Å for one of the enantiomeric forms.

That position of the two-fold axis was then checked visually by looking at a map of a unit cell centered on the predicted molecular centre. Some features in the map indeed agreed with the presence of a local two-fold axis.

Having found the orientation and the position of the local two-fold axis, we optimised them using a program from the "O" package (4), a_rt_improve. The shifts were quite significant :

old orientation = 78° 63° 180° ; translation : 76.47 -77.95 118.46 (in Å)

new orientation = 76.6° 62.3° 180° ; translation : 73.95 -79.77 115.14

Subsequently, the new orientation/position of the local two-fold axis was used in averaging procedures, using A and CCP4 programs, to improve the map (table 1). The averaging procedure consisted of 10 cycles of averaging, map inversion and calculation of a new Fo, α_c map.

Table 1 : Correlation coefficient between subunit-1 and subunit-2 in the MIRAS map obtained with KAuCl4 and KAu(CN)2

before optimization of the 2-fold axis	0.274
after optimization	0.360
after averaging around the 2-fold axis	0.761

The improved phases obtained were then used in a Difference Fourier procedure to screen through all heavy atom datasets already collected, in order to try to find out if these datasets could be used as additional heavy atom derivatives. Three weak derivatives were found by that method : $KAuI_4$, $K_3UO_2F_5$ and $(NH_4)_2U_2O_7$. It is interesting to note that all gold derivatives are going to the same two sites, related by the local 2-fold axis. The uranyl sites are situated rather far away from the 2-fold axis of the dimer, but two of them, present in both derivatives, are related by the 2-fold axis. The final MIRAS-statistics (6 derivatives) are summarised in table 2.

Table 2 : Refinement statistics (to 3.1 Å)

Figure of merit (to 3.1 Å) : 0.4580 (acentric), 0.5994 (centric)

	Soaking time	Concentration	Rdifference	nb. of sites	Rcullis (centricdata)	Phasing power centr./acentr.
$KAuCl_4$ (1)	2 nights	1 mM	11.1 %	2	0.75	1.2 / 0.9
$KAuCl_4$ (2)	1 night	1 mM	11.4 %	2	0.81	0.8 / 0.5
$KAu(CN)_2$	3 nights	1 mM	9.8 %	2	0.80	0.8 / 0.6
$KAuI_4$	2 nights	1 mM	20 %	2	0.97	0.6 / 0.5
$K_3UO_2F_5$	1 night	2 mM	20.1 %	2	0.86 (to 4Å)	0.7 / 0.5 (to 4Å)
$(NH_4)_2U_2O_7$	1 week	1 mM	26.6 %	3	0.94	0.5 / 0.4

The map obtained from these 6 derivatives was used to make a better mask of the dimer, which was used in a new averaging procedure. The phaseshifts between the MIR-phases and the phases from the averaged map are shown in table 3. Note that the overall phase shift is almost random.

Table 3 :Phase shift between the MIRAS and the averaged phases (as a function of the figure of merit of the MIRAS phases)

Figure of merit	nb of reflections	Phase shift
0.0 - 0.1	1749	118.428
0.1 - 0.2	1731	114.033
0.2 - 0.3	1485	109.450
0.3 - 0.4	1301	101.637
0.4 - 0.5	1226	98.728
0.5 - 0.6	1252	87.657
0.6 - 0.7	1247	79.081
0.7 - 0.8	1420	75.526
0.8 - 0.9	1423	66.414
0.9 - 1.0	1077	54.045
Total	13911	92.635

We could then start building a polyalanine model into the electron density, at 3.1 Å resolution. This incomplete model was refined with TNT (5) at 3.1Å with the "real space" option, and then used in phase combination. The process of phase combination using map phases and model phases was repeated several times, starting always with the same map (the averaged MIRAS map), till it was possible to see a part of the sequence of thiolase.

At the moment, 85 % of the sequence has been placed in the density, and phase extension procedures are on the way.

(1) J.Ph.Zeelen, R.K.Wierenga,R.Erdmann and W.-H.Kunau
J.Mol.Biol. (1990) 215, 211-213

(2) L.Tong and M.G.Rossmann
Acta Cryst. (1990) A46, 783-792

(3) L.Tong, H.Choi, W.Minor and M.G.Rossmann
Acta Cryst. (1992), A48, 430-442

(4) T.A.Jones, J.-Y.Zou, S.W.Cowan and M.Kjeldgaard
Acta Cryst. (1991), A47, 110-119

(5) D.E.Tronrud
Acta Cryst. (1992), A48, 912-916

DECONVOLUTING LAUE MULTIPLE DIFFRACTION SPOTS BY A REAL-SPACE DENSITY MODIFICATION METHOD.

Q. Hao* and M. M. Harding
Chemistry Department, Liverpool University, Liverpool L69 3BX, U.K.
*also SERC Daresbury Laboratory, Warrington WA4 4AD, UK.

In a Laue diffraction pattern 10-20% of the spots result from the exact superposition of two or more reflections which are 'harmonics', e.g. hkl, $2h2k2l$ ..... For the solution of large or difficult structures the intensities of the remaining 80-90% of the reflections, measurable as singles, may not be sufficient, and evaluation of the intensities of the components of the multiple spots is therefore important.

A new procedure for this deconvolution using real-space density modifications on the Patterson map is given. This is a further development based on a procedure in reciprocal-space related to Direct Method (Hao, Campbell, Harding and Helliwell, *Acta Cryst.*, 1993, **A49**, in the press). A Patterson map is calculated using singles (origin peak height A and other highest peak height B). The map is then modified by removing negative densities and squaring with the origin peak height being held at AxB. A Fourier transform of the modified map evaluates intensity values for the multiples. This procedure can be repeated until convergence is reached.

It has been tested with Laue diffraction data from 4-Zn insulin and Cytochrome C Peroxidase (CCP). 304 and 1134 reflection intensities were evaluated from multiple spots of insulin and CCP respectively; the mean fractional differences (on F), showing the agreement with the high quality monochromatic data are, 0.24 and 0.21.

The program is now available at Daresbury. It can be used in conjunction with the standard CCP4 and the LAUE software at Daresbury.

Density measurement of crystals of *Bacillus subtilis* lipase after staining with Coomassie Brilliant Blue using the Ficoll density gradient method.

Mieke Blaauw, Stéphane Ransac and Bauke W. Dijkstra

Laboratory of Biophysical Chemistry, Department of Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

We are engaged in the 3D-structure determination of *Bacillus subtilis* lipase (E.C. 3.1.1.3). So far, relatively small crystals have been obtained with polyethylene glycol (PEG) 4000 as precipitant in the presence of Li_2SO_4 and octyl-β-D-glucoside at pH around 9. In one case, a 3.5 Å data-set was obtained which is complete up to 4.5 Å resolution. The X-ray diffraction analysis indicated the presence of either 4 or 8 protomers per unit cell.

The measurement of crystal density provides information to determine the volume fraction of the crystal which is occupied by protein, and hence the number of protomers within the unit cell [1]. The volume fraction of the crystal occupied by protein, ϕ_p , is given by

$$\phi_p = \frac{\rho_c - \rho_s}{1/v_p - \rho_s} = \frac{n\ v_p\ M}{N\ V} \qquad (1)$$

where ρ_c and ρ_s are the densities of the protein crystal and the solvent (free and bound) (g/cm^3); v_p, the partial specific volume of the unsolvated protein (cm^3/g); n, the number of protomers per unit cell ; M, the molar weight of one protomer (g/mol); N, Avogadro's number (6.02 x 10^{23}) and V the volume of the unit cell (cm^3).

Hence, n can be derived as

$$n = N\ V \frac{\rho_c - \rho_s}{M\left(1 - v_p\ \rho_s\right)} \qquad (2)$$

The unit cell volume V is determined from the X-ray diffraction measurements; ρ_c is usually measured in a calibrated density gradient and v_p is often approximated as 0.373 cm^3/g [1, 2].

A convenient technique to measure protein crystal density, first described by Westbrook [3, 4], makes use of a density gradient of aqueous solutions of Ficoll 400 (Pharmacia), a large polymer (M=400 kDa) of highly cross-linked sucrose. In such a gradient, the small amount of mother liquor in the crystal is rapidly exchanged against water [5]. If ρ_s is 1.0 g/cm^3, the number of protomers per unit cell is given by

$$n = N\ V \frac{\rho_c - 1}{M\left(1 - v_p\right)} \qquad (3)$$

Especially in the case of loosely packed crystals, a continuous increase in the apparent crystal density has been observed, probably due to the intrusion of Ficoll into the crystal [5]. The correct value can be determined by measuring the density as a function of time and extrapolation back to t=0 [5]. According to this improved Ficoll density gradient method, gradients covering a range from 15 to 60% (w/w) Ficoll in 1 mM NaN_3 were prepared in centrifuge tubes. The gradients

were calibrated with small droplets of mixtures of water-saturated carbon tetrachloride and toluene, whose densities had been determined pycnometrically. Crystals inside a small droplet of mother liquor were introduced at the top of the gradient, and centrifugation was started immediately up to an acceleration of 400 g in a swinging bucket rotor. This centrifugation procedure was repeated at least 7 times. After each run the positions of the calibration droplets and the protein crystal were determined.

Due to the small size of our *Bacillus subtilis* lipase crystals (plates with dimensions of about 0.3 x 0.1 x 0.05 mm^3) it was a problem to localize the crystals in the Ficoll gradient. In order to improve their visibility in the gradient, we stained the lipase crystals overnight at 22°C in a solution of 0.05 % (w/v) Coomassie Brilliant Blue (CBB) in mother liquor prior to the density measurement. We also stained crystals of hen egg-white lysozyme (E.C.3.2.1.17) (Boehringer Mannheim) and compared the densities of the stained crystals to those of non-stained crystals.

The densities determined for crystals of both proteins, without or after staining with CBB, are given in Table I together with the derived values of n and ϕ_p. The values of the densities indicated the presence of 8 protomers per unit cell for each protein tested. Although the values of the densities of individual crystals of the lipase varied considerably, it was well possible to distinguish between n = 4 and n = 8. The values of n determined for the lysozyme crystals agreed with the literature value of n=8. The values of ϕ_p lie in the normal range [4]. Staining of crystals with CBB strongly facilitates the localization of small crystals in the Ficoll gradient. The densities measured for unstained and CBB-stained lysozyme were essentially the same.
In conclusion staining of crystals with CBB greatly facilitates the crystal density determination of small crystals by the Ficoll density gradient method.

Table I. Crystallographic parameters amd protein content of protein crystals.

Protein	M_{proto} (kDa)	Space group	Unit cell volume ($Å^3$)	CBB staining	Crystal density [a] (g/cm^3)	Protein volume fraction (%)	n (protomers in unit cell)	
							derived from density	from X-ray analysis
Bacillus subtilis lipase	19.3	P_1	435,380	yes	1.15 - 1.17 [b]	42 - 48	7.7 - 8.8	4 or 8
Hen egg-white lysozyme [c]	14.6	$P4_12_12$	237,100	no	1.203	57	7.6	8
				no	1.203	57	7.6	
				no	1.223	62	8.3	
				yes	1.215	60	8.0	
				yes	1.219	61	8.1	

[a] extrapolated value to t = 0.

[b] densities found for about 10 crystals ranged between these values

[c] crystallographic data from Blake *et al.* [6]

References

1. Matthews B.W. (1985) Determination of protein molecular weight, hydrtion, and packing from crystal density. *Meth. Enzymol.* **114,** 176-187.

2. Mikol V. & Giegé R. (1992) The physical chemistry of protein crystallization. *In: Crystallization of nucleic acids and proteins. A practical approach* (Ducruix A. and Giegé R., Ed.). pp. 219-239. Oxford University Press, New York.

3. Westbrook E.M. (1976) Characterization of a hexagonal crystal form of an enzyme of steroid metabolism, Δ^5-3-ketosteroid isomerase: A new method of crystal density measurement. *J. Mol. Biol.* **103,** 659-664.

4. Westbrook E.M. (1985) Crystal density measurements using aqueous Ficoll solutions. *Meth. Enzymol.* **114,** 187-196.

5. Bode W. & Schirmer T. (1985) Determination of the protein content of crystals formed by *Mastigocladus laminosus* C-phycocyanin, *Chroomonas spec.* phycocyanin-645 and modified human fibrinogen using an improved Ficoll density gradient method. *Biol. Chem. Hoppe-Seyler* **366,** 287-295.

6. Blake C.C.F., Koenig D.F., Mair G.A., North A.C.T., Phillips D.C. & Sarma V.R. (1965) Structure of hen egg-white lysozyme. *Nature* **206,** 757-761.

RECENT ACTIVITY AT PAVIA: SUPEROXIDE DISMUTASES AND MYOGLOBINS.

Kristina Djinovic-Carugo, Menico Rizzi, Martino Bolognesi

Department of Genetics & Microbiology, University of Pavia.
Via Abbiategrasso, 207. I-27100 Pavia (Italy)

Following two of the different research lines active in our laboratory, we have recently crystallized and solved the structures of Xenopus laevis Cu,Zn superoxide dismutase, and of loggerhead sea turtle myoglobin.

a) X.laevis superoxide dismutase-b (X-SOD). The recombinant protein, expressed in pKK322-2 plasmid, under control of the *trc* promoter, was kindly provided by the Biology Department at the University of Rome "Tor Vergata" (Prof.Desideri and Rotilio). The enzyme shows a classical Cu,Zn SOD primary structure, with about 70% conserved residues as compared to bovine erythrocyte SOD. The dimeric protein was crystallized from PEG solutions, at pH .6.0 at 27° C. The crystals were shown to belong to the orthorhombic space group $P2_12_12_1$, with cell constants .73.33, 68.86 ,59.73 Å, one dimer (32,000 Mr) per asymmetric unit. A diffraction data set was collected at 3.0 Å resolution on one crystal, of approximately 0.1x0.3x0.1 mm^3, using a MarResearch image plate scanner at Farmitalia-Carlo Erba (Milano, Italy - we thank Dr.Charles Collyer for help during the data collection stage). The R-merge factor for the observed 51,860 reflections, which were reduced to .6238 , was 8.2%.

Structure solution was pursued by molecular replacement, using the program package AMORE (Navaza, 1987), and, as search molecule, the molecular model of the semisynthetic (bovine) cobalt-zinc SOD dimer, previously solved at 2.0 Å resolution in our laboratory (Djinovic, et al., 1992). Both the rotation and the translation searches produced prominent peaks which allowed to locate unambiguously the model in the X-SOD cell. The corresponding R-factor, calculated without any modification with respect to the search model was 0.42, for the data in the 15.0 - 3.0 Å resolution range.

Refinement of the molecular structure of X-SOD is in progress. A major improvement has been achieved by means of simulated annealing techniques (Brunger, 1992), which allowed to refine the starting structure up to R = 0.19, overnight, runnig X-PLOR on a DEC Alpha computer, for a total simulation time of 2.875 ps. Inspection of the electron density with manual correction of minor structural details is under way.

b) Sea-turtle Myoglobin (Mb). A fishing accident in Sicily has provided us with a generous supply of myoglobin from the logger-head sea-turtle Caretta caretta. Since no myoglobin structure from reptiles has been studied by X-ray crystallography, so far, we have undertaken the analysis of this structure, in line with our previuos studies on lower phyla globins (Bolognesi, et al., 1989).

The protein was crystallized, as many myoglobins, from 3.2 M ammonium sulfate, at pH 8.0, at 4° C. The crystals grow to a considerable size (1.0x0.7x0.4 mm^3) in about one week. Diffraction data to 2.0 Å resolution were collected by conventional diffractometry (ENRAF-NONIUS CAD4), using one crystal. The crystals belong to the monoclinic space group $P2_12_12_1$, with unit cell constants 37.50, 61.10, 75.20 Å, one molecule per asymmetric unit.

At the time of structure solution, only 50% of the amino acid sequence was available. By comparison with other sea turtle myoglobins (2 sequences known), it was obvious that the number of residues conserved, with respect to sperm whale Mb, was of the order of 70%. The molecular model of sperm whale Mb (Takano, 1977), was therefore used for the molecular replacement search, without modifications. The program suite AMORE was used throughout (Navaza, 1987), and virtually unique solutions were found both for the rotational as well as for the translational searches (using data in the 15.0 - 3.0 Å resolution shell). The R factor, calculated after proper positioning of the search model in the sea turtle Mb cell, is 38.7 %, for the data in the 15.0 - 3.0 Å resolution range. The corresponding electron density is quite clear, and may substantially help in completing the amino acid sequence of this protein.

Bolognesi, M., Onesti, S., Gatti, G., Coda, A., Ascenzi, P., & Brunori, M. (1989). J. Mol. Biol., 205, 529-544.

Brunger, A.T. (1992). XPLOR Manual, Yale Univ., USA.

Djinovic, K., Coda, A., Antolini, L., Pelosi, G., Desideri, A., Falconi, M., Rotilio, G. & Bolognesi, M. (1992). J. Mol. Biol. 226, 227-238.

Navaza, J. (1987). Acta Cryst., A46, 645-653.

Takano, T.(1977). J. Mol. Biol., 110, 537-568.

Protein crystallographic measurements on BL9 (Troika) of the ESRF, February-March 1993

M.S. Lehmann*, J. Als-Nielsen$, G. Grübel and J.F. Legrand#

ESRF, Av. des Martyrs, F38043, Grenoble, France

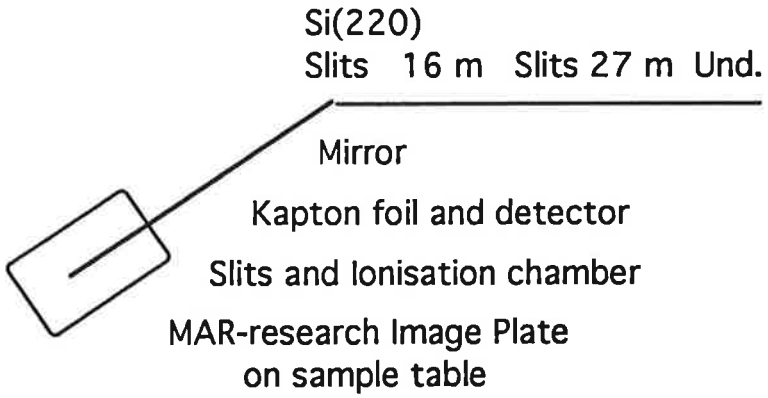
Introduction

The beam-line 9 (Troika) is located at the 1.6 m long undulator device ID10 (period: 46 mm) of the ESRF. It is planned for three consecutive monochromator housings, each with a single, X-ray transparent, monochromator crystal. In the first set-up only one housing was available, and the monochromator was either Diamond(111) in transmission or Si(220) in reflection.

At present the beam-line is undergoing commissioning tests, and the protein crystallographic measurements were part of these. Two periods were allocated, namely 5 days in February and 5 days in March.

The three stations of the Troika beam-line are foreseen to be multipurpose instruments operating independently. They will each be equipped with a sample support, a detector and (optionally) a crystal analyzer system, so they can either in themselves function as diffractometers or they can serve as the support for yet another diffractometer. In the last case the monochromator is used to select the beam, while the crystal orienter and (multi) detector is located at the sample position.

The set-up for protein crystallography was as following:



---

* On Leave from ILL, Grenoble, France. Person to contact for further information about these tests.

$ Permanent address: Riso National Laboratory, Roskilde, Denmark

On leave from ILL, Grenoble, France

The distance from the monochromator to the sample was 135 cm and the detector diameter was 18 cm.

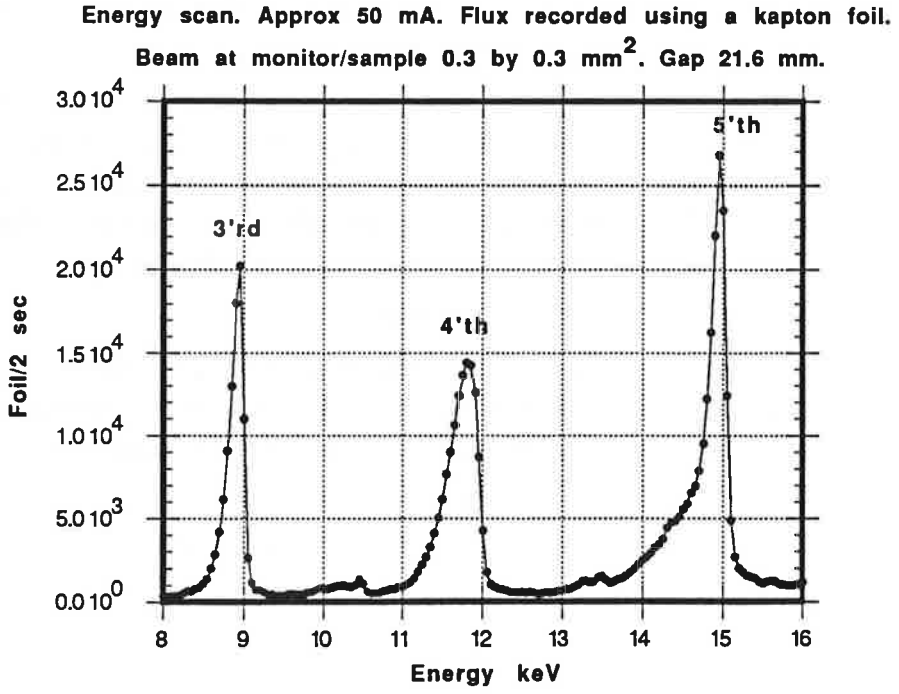
For the present experiments the thin, water-cooled Si(220) monochromator recently developed by A. Freund, ESRF was used rather than Diamond, mainly because it gave the best energy resolution for the MAD experiments.

The experiments were done in tight collaboration with the EMBL/Grenoble who supplied the MAR-research image plate as well as technical and crystallographic support (J. Allibon, C. Berthet, J.-M. Bois, S. Cusack, F. Dauvergne, R. Leberman, D. Pognant, B. Rasmussen, J. Sedita, A.W. Thompson and C. Wilkinson). Similarly, there was extended support from the technical staff of Troika (P. Feder and J. Linderholm), from many services of the ESRF, and ILL helped out with the loan of a VAX workstation and advice (A. Barthélemy).

Installation phases

The Mar-Research image plate was installed at BL9 and a mini-cluster was created using two VAX 3100 workstations. The cluster was equipped with exabyte and TK50 back-up units.

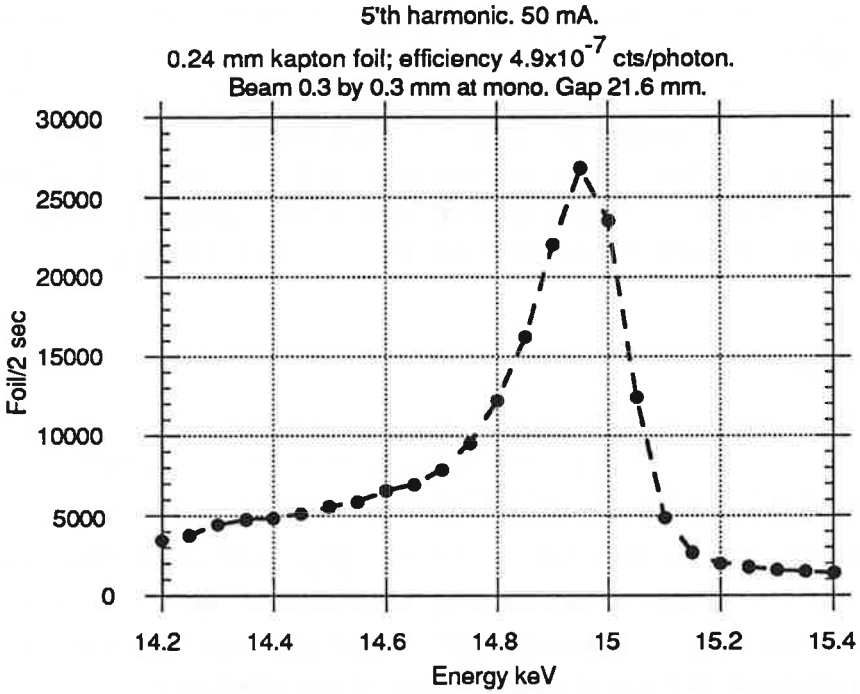
The undulator spectrum for a given gap between the magnets is recorded with a Θ-2Θ scan of the instrument where the flux can be estimated from the scattering off a kapton foil placed in the beam (see figure above). A typical spectrum is shown below[$]



[$] The discriminator of the detector was set to accept the whole range of energies. The lower harmonics are therefore contaminated with higher orders and thus overestimated.

The wavelength (or energy) selected for the experiment is chosen to be near one of the maxima, preferably the 3'rd or 5'th harmonic. The positions of these can be changed by changing the gap. An increase in gap will increase the energy.

It is worth noting that the harmonics are very narrow. Below is shown in detail the 5'th harmonic, which was used for protein crystallography in March.



From this graph the flux on the sample can be obtained, but it should be remembered that at this early stage the estimate might still be off, perhaps up to a factor of three.

The maximum of the 3'rd harmonic was used to optimize the location of the monochromator in the primary beam. For an undulator this is important because the beam characteristics such as flux, polarization and homogeneity varies strongly across the beam. Slits were set with values of 0.5 by 0.5 mm^2 at 26 m from source and 0.3 by 0.3 mm^2 before the monochromator.

The instrument, including the image plate was then aligned without mirror. This part was much facilitated due to a remote control unit for the image plate slits build at the EMBL for this purpose.

Finally a Si-C mirror was added in order to remove higher order contamination, and the beam was aligned to enter the image plate in the horizontal plane.

Protein crystallographic measurements

The crystals studied came from different projects under way at the EMBL/Grenoble, at the groups of O. Dideberg (LCM) and J. Fontecilla-Camps (LCCP) of the neighbouring Institut de Biologie Structurale (IBS) and at ILL. In some cases full data set were collected. In other cases only a few diffraction patterns were recorded for test or because the crystals did not survive. A simple gas-flow cooling device was used at several occasions to prolong the life-time of the crystals.

In February the undulator was set with a 27 mm gap giving a 3'rd harmonic with λ=0.787 Å. Two days were used to collect full or partial data sets on:

Seryl-tRNA synthetase *Thermus thermophilus* co-crystallized with a seryl-adenylate analog (EMBL, 2.6 Å)
Seryl-tRNA synthetase *Thermus thermophilus* with ATP analogue and serine hydroxamate (EMBL, 2.5 Å)
Hydrophobic protein tetramer complex (ILL, 2.7 Å, partial)

At the same time tests were made on:

Seryl-tRNA synthetase *E. Coli* grown in citrate (EMBL)
Seryl-tRNA synthetase *E. Coli* with tRNAser (EMBL)
EfTu-Ts complex (EMBL)
Orthinine carbamoyl transferase (OTCase) (IBS/LCM)
Chimeric form of OTCase (IBS/LCM)
Tcell receptor (IBS/LCCP)

In addition a few measurements were carried out on small crystals of catalase (150μ on the edge). The divergence of the beam is approximately 35 μrad and there should thus be no observable variation of the spot-size as a function of sample-detector distance. Indeed, there was not. The full width at half height of the spots was around 300 μ, i.e. 2 pixels. The typical full spot size was around twice this value, i.e. four pixels. No further tests were done as measurements on fine focus beam-lines will be done later.

In March with λ=0.829 Å (5'th harmonic for a gap of 21.6 mm) a little over two days were used for measurements and gave data on:

Seryl-tRNA synthetase *Thermus thermophilus* co-crystallized with a Ap4A, a secondary reaction product (EMBL, high (2.4 Å) and low order data)
Seryl-tRNA synthetase/tRNAser complex from *E. coli* (EMBL, 3.5 Å, partial)
Seryl-tRNA synthetase/tRNAser complex from *T. thermophilus* (EMBL, 2.7 Å, partial)

Complex of elongation factor EfTu-EfTs from *E. coli* (EMBL, 2.8 Å, partial)
Mutant (E103A) of ornithinine carbamoyltransferase of *Ps. aeruginosa* (IBS, 2.8Å, partial)
Mutant (E103G) of ornithinine carbamoyltransferase of *Ps. aeruginosa* (IBS, 2.8Å)

Test were made on:

Small crystals of penicillin binding 5 of *E. coli* (IBS/LCM)
Human trypsin (IBS/LCCP)
Porcine lipase (IBS/LCCP)

It should be noted that the partial data sets will be useful in analysis combined with previous data measured on the same systems.

During the MAD measurements a set of data was recorded to 3 Å with λ=1.385 Å on a crystal of β-Lactoglobulin (from E. Stura, Scripp's Clinic).

Data analysis is now under way. An example of the studies is Seryl-tRNA synthetase *Thermus thermophilus* co-crystallized with 5'-O-[N-(L-seryl)-sulfamoyl]adenosine (H. Belrhali, A. Yaremchuk and S. Cusack, EMBL). Space group: $P2_1$, unit cell 87 x 127 x 64 $Å^2$, β=109°. One needle crystal was used in three places to give 107 oscillation pictures of each 1.5°. Measurement time for one frame was around 3 min. Completeness to 2.6 Å resolution was 87 %, the R_{merge} was 6 % using MOSFLM for integration and varied from 5 % for the low order data (5 Å) to 12 % for the high order data (2.6 Å). In total 102000 reflections were recorded, of which 13000 were partials, and the averaged set held 34700 reflections. At present the substrate/inhibitor has been localized with densities of ten times the r.m.s. noise and refinements and analysis are well under way.

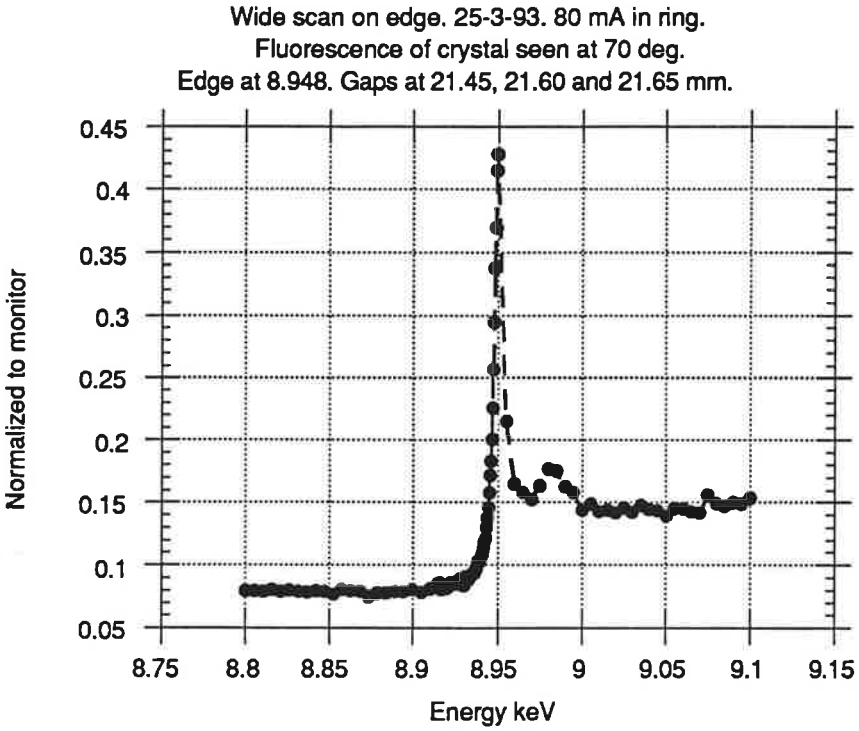
Multiple anomalous dispersion (MAD)

The last days in each run were devoted to MAD measurements proposed by W.H. Hendrickson and L. Shapiro, Columbia University, N.Y. and J. Als-Nielsen, with the participation of G. Grübel, M.S. Lehmann and A.W. Thompson.

The subject of study was a 102 residue calcium binding protein, where the calcium atom can been replaced by a lanthanide. In the present experiment ytterbium was chosen. The crystals studied are hexagonal, and grow as hexagonal needles.

For this a longer wavelength was chosen corresponding to the nominal value of the L_{III}-absorption edge for Yb^{3+}, ≈ 8.948 keV (the edge was defined as the value at half height). In February the undulator was set to have maximum for the second harmonic (27.0 mm gap) and in March the third harmonic was chosen (21.6 mm gap).

Edge scans were then carried out on an ytterbium containing foil placed at the monitor position and a sample crystal mounted on the diffractometer. A wide scan has the following features:



The total width in energy of the 3'rd harmonic is comparable to the range of this scan, and it was therefore carried out at three different gaps to get maximum flux from the undulator.

The wavelengths of the originally planned experiment were designed all to lie within one harmonic of the undulator and the first measurements, in February, were done in this mode. Better measurements could however be done if the gap was changed to allow for a larger range of wavelengths. In March, for the second run, the machine group of the ESRF therefore set up a procedure that made it possible to position the gap for the different measurement points.

In total there were thus two modes of operation. For fixed gap, four energy values were chosen for the measurements, namely about 100 eV below the edge, at the ascending edge below the maximum of the white line, at the maximum of the white line, and at the descending edge above the maximum.

When the gap was varied five values were chosen: near the ascending edge below the maximum of the white line, at the maximum of the white line, at the descending edge above the maximum, at the descending edge above the second smaller maximum and finally well above the edge (+600 eV) near the minimum between the L_{II} and the L_{III} edge.

For each fill of the synchrotron the edge position was checked doing a limited edge scan. This did vary only little, the fluctuations being of the order of ± 0.2 eV.

For each of the four or five positions in energy a ϕ scan was done after which ϕ was increased and the procedure repeated. The ϕ oscillations were typically 2.1° per frame with steps of 1.8° in the angle. The limiting resolution was set to 2.5 Å. The radiation damage was non-negligible, and one location on the needle crystal gave from two to four full sets of data, thus up to about 8° in angle. In the very best case 5 such sets of data were recorded on one crystal.

The wavelengths used for the four energy scans with fixed undulator gap were[$] :

Comment	Energy in keV	Wavelength in Å
Below edge	8.8480	1.4013
Ascending edge	8.9476	1.3857
Maximum of edge	8.9506	1.3852
Descending edge	8.9537	1.3847

For the variable gap method the five wavelengths used for measurements were[#] :

Comment	Energy in keV	Wavelength Å	Gap in mm
Asc. edge	8.9476	1.3857	21.6
Maximum	8.9506	1.3852	21.6
Desc. edge	8.9537	1.3847	21.6
2nd Desc. edge	8.9950	1.3784	21.7
Above edge	9.640	1.2862	22.5

This allowed to have both constant and maximum intensity, and worked well.

The data used in further analysis were all measured in March. In total 7 crystals were used to record 68° of data around the hexagonal axis, and 30° orthogonal to this. For 5 of these crystals the gap was varied during the measurement.

Intensities have now been integrated using MOSFLM, and R_{sym} for the best data is around 6 %. Data analysis is under way. The anomalous scatterer has been located, and the first phased maps are being studied.

[$] The last three values correspond to extrema in f'.

[#] The first four values correspond to extrema in f'.

Concluding remarks

The measurements showed that an undulator beam-line was well suited for the experiment. It was easy both to install and align the diffractometer on the two-axis base available, and adjustment of the mirror did not cause any major problem either.

For a few days real measurements on proteins were done under conditions that approach routine work, and the undulator set-up with a single monochromator is as expected satisfactory for this kind of data collection.

The background did not seem to be a problem, mainly due to the small size of the beam entering the experimental hutch. During the experiment only a temporary beam-stop for the direct beam was installed just behind the monochromator housing, and it was most likely the main source of background. This was handled by enclosing the sample environment of the image plate in a box with 1.5 mm lead, but when the final beam-stop for the main beam is available, no protection should be needed.

The MAD measurements proved to be feasible, and it was possible to set the sample table with the image plate so that the beam repeatedly illuminated very near the same part of the crystal. Successive tests of the location of the Yb edge showed reproducibility of the order of ± 0.2 eV.

The measurements were also very helpful for the standard tests of programs and mechanical parts that is a normal ingredient of all commissioning, and worked to satisfaction.

No focusing devices were used, so there is of course still an improvement in flux to expect from focusing multilayers and a curved monochromator. It is also easy to imagine this coupled with automatic change of the undulator gap, taking full advantage of the flexibility of this insertion device.

Acknowledgment. We thank J. Susini, ESRF for loan of the Si-C mirror.

APPLICATIONS OF LAUE DIFFRACTION AND OTHER SMALL MOLECULE STUDIES WITH SYNCHROTRON RADIATION

GMT Cheetham, IM Dodd, Marjorie Harding, BM Kariuki, Hao Quan
Chemistry Department, Liverpool University,
PO Box 147, Liverpool L69 3BX, UK

Some recent work at Liverpool is outlined below; although the compounds studied are 'small molecules', some aspects of the methods could be of interest for larger biological molecules.

Laue diffraction for complete structure determination.

The structure determination has been completed for a small crystal of a newly synthesised compound using only Laue diffraction photographs. This required unit cell determination, for which Laue diffraction has not normally been used, as well as the measurement of intensity data and structure solution; it also required refinement taking anomalous scattering contributions into account (in Fc). In the Laue method, different reflections are measured at different wavelengths; the anomalous scattering factors, f' and f", and therefore the structure factors, vary with wavelength, and this variation can be substantial for heavy atoms.

For the unit cell determination, film packs were recorded on SRS workstation 9.5 with a palladium foil attenuator (0.1 mm) in the incident beam. The axial ratios and the cell angles were derived from these photographs using gnomonic projections[1]. The minimum wavelength present was sharply defined by the Pd absorption edge (0.509 A), and this allowed the unit cell to be established on an absolute scale to an accuracy of *ca* 0.5% [2].

Intensity data were measured from five film packs using the Daresbury Laue software suite[3]. For these film packs recorded on workstation 9.7, the incident beam was attenuated by 0.2 mm Al and 0.114 mm Cu; these attenuators shifted the spectral distribution to shorter wavelengths, effectively 0.24-0.7 A, reducing absorption, radiation damage, and background due to air scattering. [The experiment to determine unit cell could also have been done on station 9.7, but was actually done when we were using station 9.5 for other work.] 12183 individual intensity measurements were merged to give 7163 unique reflections from which most of the structure was found and refined to R=0.12. However one phenyl group and a solvent molecule could not be located in the difference electron density maps.

At this stage the program SHELXL-92 became available to us (for which we are most grateful to Prof George Sheldrick); structure factors were recalculated taking proper account of

the values of f′ and f" at the appropriate wavelength for each reflection; the resulting difference map revealed the missing atoms. Finally an absorption correction was applied[4] and the unmerged data was refined using SHELXL-92 to give R1=0.075. The compound was shown to be $AuOs_3(CO)_8PPh_3dppm.PF_6.0.5C_6H_5Cl$

Monochromatic Data Collection for very small crystals

Using SRS workstation 9.6, a wavelength of 0.9 A, and the FAST diffractometer the smallest crystal for which a structure has been solved had dimensions 10x10x30 μm. This allowed the previously unknown chemical constitution of a new gold cluster compound to be established as $Au_{10}(PPh_3)_7(S_2C_2(CN)_2)_2$; refinement converged to R=0.064 for 3747 reflections[5].

Data collection and structure determination and/or refinement have been achieved for two aluminophosphate crystals of dimensions *ca* (20 μm)^3 and for a sucrose crystal 10x10x15 μm; for these the high intensity (and low beam divergence) of synchrotron radiation are probably essential, together with a diffractometer such as FAST. For somewhat larger crystals of aluminophosphates, e.g. (50 μm)^3, experiments have shown that a rotating anode source can give good data for structure solution[5].

1. P D Carr, D W J Cruickshank, M M Harding (1992) J Appl Cryst 25, 294-308.

2. I M Dodd, P D Carr,, M M Harding (1993) J Appl Cryst 26, *in the press*.

3. J R Helliwell *et al* (1989) J Appl Cryst B22, 483-497.

4. S J Maginn, M M Harding and J W Campbell (1993) Acta Cryst B49, *in the press*.

5. M Helliwell, V Kaucic, GMT Cheetham, MM Harding, BM Kariuki and PJ Rizkallah (1993) Acta Cryst B49, *in the press*.

knowledge/information, obtained by computing various quantities/properties from the raw atomic coordinates and sequence data. This includes f.e. a repertoire of residue contacts between secondary structures, domains and subunits, exhaustive collection of protein folding motifs obtained by automatic analysis and their detailed description (atomic interactions and sequence patterns); data on evolutionarily related proteins, etc.

3. Address the problem of links and interfaces between structural databases and other types of databases and modelling tools, which are presently virtually nonexistent. This includes prototype designs of links between the structural databases to databases of DNA and protein sequence. It also includes interfaces between relational systems and (a) conventional modelling programs for interactive use; (b) object–oriented front–ends; (c) Logic (Prolog) based front–end or (d) direct graphics–based (X–Windows/ Motif) user interfaces.

With the project now entering its third year, we can start to evaluate what it is or is not producing. The task of defining a standard conceptual schema for macromolecules has turned out to be more difficult than originally anticipated, as consensus could not be readily reached for defining certain types of molecular objects and concepts. At present the core schema, the one defining the most basic and commonly used concept and objects, has been agreed on. We hope to extend this agreement to higher level descriptions of protein data items (families of folding motifs, descriptions of topological features, etc.) by the end of 1993. It was furthermore encouraging to find out recently, that the established core schema and the data structure of the macromolecular CIF dictionary (the Crystallographic Information File format shortly due to replace the widely used Brookhaven PDB format), displayed only minor differences. Contacts established with the IUCr task force who is finalizing the CIF dictionary, have led to a very useful exchange of information with mutual benefits.

Progress in other areas has been smoother: object–oriented front–end tools have been mounted onto relational databases (e.g. SESAM) and shown to provide a flexible working environment. Prototype Object Oriented user interfaces implemented in C++, have been developed a the EMBL, which feature fast access to PDB files and simple 2D and 3D graphic display tools. Notable progress has also been made in generating and representing 3D structural information: a comprehensive classification of protein fold families has been carried out (EMBL, UCL) and a means of storing information on spatial proximity of atoms/residues, which combines cubing algorithms and residue neighbour lists have been devised.

A very crucial issue could however not be adequately addressed in this project. It is the issue of efficient procedures and criteria for checking and validating the data stored in the structural databases and from which valuable information and knowledge are extracted. Scientists grope with these problems as best they can, but there clearly is a lack of standard criteria and efficient automated procedures for data validation. With the aim of addressing this pressing issue more effectively a new European project was initiated in 1993 within the BIOTECH programme, with a very active role taken by members of the crystallographic and NMR communities, the chief producers of structural data.

THE EFFECT OF CRYOGENIC TREATMENT ON THE CELL DIMENSIONS OF RIBOSOMAL CRYSTALS

Jörg Harms°, Frank Schlünzen°, Klaus von Böhlen°, Jesper Thygesen°, Susanne Meyer°,Ilona Dunkel#, Beate Donzelmann°, Harly A.S. Hansen°, Anat Zaytzev-Bashan^, Alex Dribin^, Gitay Kryger^, Gesine Thoms°, Niels Volkmann°, Heike Bartels°, William S.Bennett° & Ada Yonath^°

° Max-Planck-Res.-Unit for Ribosomal Structure, Hamburg, Germany
Max-Planck-Institute for Molecular Genetics, Berlin, Germany
^ Dept. of Structural Biology, Weizmann Inst., Rehovot, Israel

SUMMARY

We report here our attempts at quantitative assessment of the effect of the shock freezing process on two properties of ribosomal crystals: shrinkage of the cell dimensions and isomorphism. We also refer to qualitative observations concerning the resolution and the mosaic spread of shock-frozen crystals.

Ever since we introduced cryo-temperature crystallographic data collection, we suspected that the inherent flexibility of the ribosome may not only allow significant shrinkage of the unit cell but also may lead to variability in its magnitude. However, independent determination of the cell dimensions of each part of a segmented crystal show that the shock cooling process was rather reproducible and did not harm the isomorphism.

These findings are most important for the continuation of our studies as they decrease our previous concern about the lack of isomorphism which may be caused by the mechanical and chemical stresses involved in the process of shock freezing.

I. Introduction

Ribosomes are the universal supramolecular assemblies responsible for the translation of genetic information, encoded in mRNA, into proteins. A typical bacterial ribosome contains more than 250.000 atoms, has a molecular weight of about 2.3×10^6 and a sedimentation coefficient of 70S. It is composed of two subunits of unequal size (small = 30S, m.w.:0.85×10^6, and large = 50S, m.w. = 1.45×10^6), which associate upon the initiation of the biosynthetic process. Bacterial ribosomes contain three chains of rRNA (about 5500 nucleotides), accounting for two thirds of its mass and 57-73 different proteins, depending on the species.

Systematic exploration of crystallization conditions, combined with sophisticated seeding, led to reproducible growth of ordered three-dimensional crystals of ribosomes, their intact subunits, their mutants and their functionally active complexes, diffracting best to 2.9 Å (von Böhlen et al., 1991). All crystals suitable for crystallographic studies are of ribosomal particles from halophilic or thermophilic bacteria, and in all cases, the crystalline ribosomal particles retain their integrity and biological activity for long periods despite their natural tendency to disintegrate rapidly. The large unit cell dimensions and the extremely weak diffracting power of the ribosomal crystals dictate the

use of intense synchrotron radiation for all steps in data collection, including the characterization of the crystals and the search for suitable derivatives (for review see Berkovitch-Yellin et al., 1992).

At ambient temperatures, the radiation damage of ribosomal crystals is so severe that most of the reflections beyond Bragg spacings of 10-15 Å decay almost instantaneously. Initially, before this property was noticed, this extreme sensitivity led to the incorrect conclusion that the measurable diffraction of all the ribosomal crystals was limited to that resolution. Only after observing the first indications for a higher resolution, a special procedure was designed, according to which precise alignment was skipped and each individual crystal was exposed only once. The combination of randomly oriented crystals, a high mosaic spread and single exposures resulted in wrong assignments of the space group, and, more seriously, in repeated failures to collect complete data sets, even when 263 were exposed (experiment performed at X11/EMBL/DESY, October, 1985).

The radiation damage of the ribosomal crystals was virtually eliminated by collecting the crystallographic data at cryogenic temperatures (about 90-100 K) from shock frozen crystals, immersed in very small amounts of solvents, which, upon cooling, solidify as amorphous materials (Yonath et al., 1988; Hope. et al., 1989; von Böhlen et al., 1991). An experimental procedure was designed to accommodate the unique features of the ribosomal crystals: anisotropic morphology (at least one very thin dimension), extreme softness and high fragility. Prior to cooling, the crystals are transferred to media similar to their original stabilizing solutions, but with somewhat higher viscosity. Often a cryosolvent has to be added to avoid the formation of ice crystals. A thorough empirical search to establish individual pre-cooling treatment for each crystal type was essential. The variables being the type and compositions of the added materials, the fashion of addition and the time course of the treatment (Hope et al., 1989).

All ribosomal crystals obtained so far are soft, flexible and easily deformed. When mounted on glass fibers, as is routinely done for cryo-temperature data collection of crystals of non-biological compounds or for average-size proteins, they bend around the glass fiber and loose their internal order. To suit this we constructed a variety of microspatula which allow mounting crystals in desired orientations. The more elaborate ones are made of double layers of thin glass plates ("double-layer" or "sandwich" spatula). These provide extra protection from drying and from bending stresses resulting from surface tension effects, created upon pulling crystals from the pre-cooling treated stabilizing solution.

Best results were obtained from crystals, immersed in their specific pre-cooling solutions and plunged into liquid propane near its solidifying temperature. A special apparatus was constructed, allowing transfer of the cooled crystals from the propane to the X-ray camera, where they are surrounded by a nitrogen gas stream at cryo-temperature throughout data collection. Under these conditions irradiated frozen ribosomal crystals show no signs of decay over periods longer than the time needed to collect a complete diffraction data set (days or weeks). To resume interrupted diffraction experiments, we constructed a device for preserving irradiated crystals in solid propane/liquid nitrogen chambers for extremely long periods, even a few years. To estimate the influence of the storage at cryo-temperature, the diffraction data of a fresh crystal was compared with those collected after 24 hours of irradiation and after 153 days of storage in solid propane. In both, no intensity changes were detectable.

II. The effect of shock freezing on the cell dimensions

For a large number of crystals of biological macromolecule it was shown that the shock cooling treatment effects the unit cell dimensions. In some cases, where these values could be determined at both ambient and cryo-temperature, significant shrinkage, 3-20% of the original length was reported (Shakked et al., 1990).

As most of our serious data collection experiments were carried out with synchrotron radiation at cryo-temperatures, accurate determination of the cell dimensions was not possible. Only a few

months ago were we able to determine the cell dimensions at ambient temperature, from limited data collected to around 10-12 Å, from one crystal of the large ribosomal subunits from Haloarcula marismortui, using the GX21 rotating anode coupled with an imaging plate. In this crystal form (space group $C222_1$), the a and the b axes (214 Å and 303 Å respectively) were found to be quite conserved in most of the so far measured crystal, whereas the c axis (around 570 Å), show some variability. Thus, variations of up to 1.5% were commonly observed for it between the frozen crystals.

Comparison of the cell dimensions determined at ambient-temperature with the average values obtained from the shock-frozen crystals by synchrotron radiation, showed shrinkage of up to 2.3% of the a and b axes. The ambient temperature c axis was found to be at the upper level of the average length found for it in frozen crystals. Therefore accurate analysis of the susceptibility of the c axis to the shock cooling procedure is still not possible.

A word of caution is due. These values should be considered as most preliminary, since the ambient temperature cell dimensions were determined from only one crystal and since a change of 1.5% of the c axis amounts to a change of 9 Å, which reaches the order of magnitude of the radius of globular proteins of a molecular weight comparable to that of average ribosomal proteins.

III. The reproducibility of the Freezing Procedures

Previous experience showed that even crystals grown under the same conditions from the same ribosomal preparation may show some variability in the cell dimensions. It was not clear whether this variability is an inherent property or induced by the cooling, since the rather lengthy pre-cooling treatment and the shock freezing, coupled with the inherent flexibility of the ribosome may result in gross structural variability or in apparent changes in cell dimensions.

To address this point, we halved a relatively large crystal of the 50S subunits from H. marismortui, and shocked cooled each of its halves separately. The two halves were positioned in the beam in a similar orientation, and data were collected around the cell axes of each part (12° around the a and b axes of each half). Differences of magnitude (0.25%) similar to the experimental errors in the determination of all cell axes were observed. Therefore it was concluded, that at least for this particular crystal, the shock cooling did not introduce changes. Thus, this preliminary experiment show that the process of shock cooling is rather reproducible, and if indeed it introduces shrinkage or other changes, these are of a similar magnitude in different experiments.

We plan to verify this result by repeating the experiment using crystals of different sizes, cut to even as well as non-even fragments. Unfortunately, the cutting procedure has still to be refined. Currently most of the segmented crystals are somewhat damaged presumably due to the development and propagation of strain caused by the segmentation, although we are experimenting under very fine and mild conditions.

No matter what is the source of the real, or apparent non-isomorphism of the ribosomal crystals, its existence may complicate phasing by methods which rely on difference maps (such as MIR and SIR). The cut crystals may be instrumental for overcoming these problems. Given we achieve smooth and non-damaging cutting, a procedure is currently being developed for collecting pairs (or larger number) of data sets from the same crystal. Thus, one segmented crystal should give rise to both ambient- and cryo-temperatures cell dimensions. Furthermore, for minimizing non-isomorphism effects, native data will be collected from one part of a cut crystals and the other part(s) will be measured after being soaked in solutions of potential heavy atom derivatives as well as of different densities which may facilitate phasing by contrast variation.

IV. The effect of shock freezing on resolution and mosaicity

The significant heterogeneity of the ribosomal crystals makes a quantitative assessment of the effect of shock-freezing on crystal properties impossible because the same crystal can not be exposed both at ambient and at cryo-temperatures, due to the irreversibility of the effects of radiation damage. However, for ribosomal crystals, it is clear that the mosaicity of properly shock-cooled crystals is conserved throughout the shock-freezing procedure.

The resolution limits of the diffraction patterns measured at cryogenic temperatures should not be higher than those obtained at ambient temperature, since these limits reflect the intrinsic conformational heterogeneity of the crystal. On the contrary, it is likely that the various steps in the cooling procedure, with their associated thermal, osmotic and mechanical stresses, may lead to a deterioration of crystal order. However, as the thermal motion within the frozen crystals is limited, and as at cryo temperature, the crystal can be irradiated for a long time, an apparent improvement in resolution was observed, resulting from the detection of a larger number of the higher resolution weak reflections.

Acknowledgement:

We thank Dr. B.C. Wang for stimulating discussion and our colleagues: Drs. Z. Berkovitch-Yellin, F. Frolow, H. Hope, C. Kratky, F. Franceschi, I. Agmon, M. Eisenstein, I. Sagi, I. Levin, Y. Halfon, T. Arad, M. Laschever, J. Piefke, J. Müssig, B. Romberg, C. Glotz, C. Paulke, N. Böddeker, R. Hasenbank and R. Albrecht, for their inspiration and participation in these studies.

Data were collected at EMBL/DESY, Hamburg; CHESS, Cornell U.; SSRL, Stanford U. and PF/KEK, Japan. Support was provided by NIH (GM 34360), BMFT (05 180 MP BO), DARA (50 QV 8606 1) and the Kimmelman Center for Macromolecular Assembly. AY holds the Martin S. Kimmel Professorial chair.

References:

Berkovitch-Yellin Z., Bennett W.S. and Yonath A., 1992, CRC Rev. Biochem. & Mol. Biol. 27:403

Berkovitch-Yellin Z., Hansen H.A.S., Weinstein S., Eisenstein M., von Böhlen K., Agmon I., Evers U., Thygesen J., Volkmann N., Bartels H., Schlünzen F., Zaytzev-Bashan A., Sharon, R., Levine I., Dribin A., Kryger G., Bennett W.S., Franceschi F. & Yonath A., 1993, in "Synchrotron Radiation & Mol. Biol.", (N. Sakabe, Ed.) Oxford U. Press,

von Böhlen, K., Makowski I., Hansen H.A.S, Bartels H., Zaytzev-Bashan A., Meyer S., Paulke C., Franceschi F. and Yonath A., 1991, J. Mol. Biol. 222:11

Hope, H., Frolow, F., von Böhlen, K., Makowski, I., Kratky, C., Halfon Y., Danz, H., Webster, P., Bartels, K., Wittmann, H.G. and Yonath, A., Acta Cryst. B345, 190 (1989)

Shakked Z., Guerstein-Guzikevich G., Zaytzev A., Eisenstein M., Frolow F. and Rabinovitch D., 1990, in "Structure and Methods, Vol. 3: DNA and RNA" (eds: R.H. Sarma & M.H. Sarma) Adenin Press.

Yonath, A., Frolow, F., Shoham, M., Mussig, J., Makowski, I., Glotz, C., Jahn, W., Weinstein S., and Wittmann, H.G., J. Cryst. Growth, 90, 231, (1988)

Location of the redox centers in hydrogenase as determined by X-ray crystallography at 5 Å resolution

A. Volbeda, C. Piras, M.H. Charon, E.C. Hatchikian[$], M. Frey & J.C. Fontecilla-Camps,

Laboratoire de Cristallographie et de Cristallogénèse des Protéines,
Institut de Biologie Structurale (CEA,CNRS), 41 avenue des Martyrs,
38027 Grenoble Cedex 1, France.

[$]Laboratoire de Chimie Bactérienne, CNRS,
BP71, 13277 Marseille, Cedex 9, France.

Hydrogenases are metalloenzymes which catalyze the reversible cleavage of molecular hydrogen. They are found in a large number of prokaryotes and eukaryotes and are involved in a great diversity of basic metabolic pathways. These include the coupling of H_2-oxidation to nitrogen fixation, methane formation or sulfate reduction, H_2-production as a means of disposing of excess electrons produced by other reactions, and biological energy conservation by the formation of a proton gradient across the cytoplasmic membrane. The hydrogenases may be classified in two main groups: "iron only" hydrogenases, and nickel-containing hydrogenases[1].

So far, no X-ray structure has been reported for either class of hydrogenase, although crystals suitable for high resolution X-ray diffraction were already obtained more than six years ago[2,3]. The last three years we have been working on the crystal structure determination of the periplasmic nickel containing hydrogenase from the sulfate reducing bacterium *Desulfovibrio gigas*. This extensively studied enzyme is a heterodimer of two subunits with molecular masses of 62 kDa and 28 kDa, respectively. The enzyme exchanges electrons with a tetrahaemic cytochrome c_3 [4]. Four redox centers have been characterized by means of EPR and Mössbauer techniques: two [4Fe-4S] clusters, one [3Fe-4S] cluster and one Ni-ion[5,6,7]. The Ni has been shown to be directly involved in the binding of H_2 [8,9].

Although *D. Gigas* hydrogenase, unlike many other hydrogenases, has the attractive feature of being relatively oxygen stable, the enzyme is not active in the presence of air. Two different deactivated forms exist, termed "unready" and "ready", the first being the most prevalent one when oxygen is present[10]. The second form can be rapidly activated, but the conversion of the "unready" to the "active" form is a long process requiring several hours of incubation under reducing conditions. The activation process involves the reduction of the [3Fe-4S] cluster and the Ni-ion, and is probably accompanied by a change in the coordination state of the latter.

Recently, we obtained a new triclinic crystal form of *D. gigas* hydrogenase (conditions first found by C. Abergel, unpublished results, and later modified by us), with cell dimensions a=64.3Å, b=94.3Å, c=69.3 Å, α=89.7°, β=103.3° and γ=90.5°. These cell dimensions are

much smaller than those found for the previous monoclinic form[3], and consequently the new form shows a significantly better diffractive power. There are two hydrogenase molecules per unit cell, the solvent content being ≈40%. We collected a large number of native and derivative data sets on a XENTRONICS/SIEMENS area detector, and processed these with the program XDS[11]. In addition, several data sets were collected at the LURE synchrotron and processed afterwards with the MOSFLM package[12]. The best crystals diffracted to ≈2.5 Å resolution.

A major difficulty for the structure determination is the instability of the hydrogenase crystals and their sensitivity to radiation damage. No less than 8 different native data sets had to be merged together (using the programs SCALKB and KBAPLY from the BIOMOL package developed in Groningen, The Netherlands) in order to obtain a first 3.3 Å resolution data set with >90% completeness (R_{mrg} = 8.9% on intensities). For the same reason only incomplete data sets could be obtained for heavy atom derivatives, as it is normally not a good idea to merge different derivative data sets together. In addition, native data sets collected afterwards using crystals obtained from independently purified batches of the enzyme showed a significant non-isomorphism with the first native data set.

In spite of such difficulties, the first isomorphous difference Patterson map could be solved at 5 Å resolution for a Hg derivative, showing 4 heavy atom binding sites. Using in addition cross derivative Pattersons and standard difference Fourier techniques, we found 1 site for a Pr, 2 sites for a Au and, so far, 10 sites for a Pt derivative. Patterson maps calculated from intensity differences between the Pr and the other three derivatives proved to be very helpful as they gave strong negative peaks for the cross vectors between the single Pr site and the main sites of the respective other derivatives. In this way it was quite easy to fix a common origin for all the heavy atom sites. The correct handedness of the heavy atom distribution could be determined by including the anomalous information.

Refinement of the heavy atom parameters was started with the old CCP4 version of the program PHARE, but it became clear soon that the new maximum likelyhood version[13,14] of the same program gave much better results, especially with respect to the refinement of occupancies. From then on the old version was used only in the final phase calculation step, in which the overall derivative scale factors were adjusted to remove correlations between heavy atom and native best phases. The resulting MIR phases having an average figure of merit of 0.6 at 5 Å resolution appeared to be of good quality, as shown by the results described below.

An MIR map was calculated at 5 Å resolution and subjected to 5 cycles of solvent flattening[15,16], using a solvent content of 30% and including phase combination with the MIR phases. A selected part of the resulting solvent flattened map is displayed in figure 1. The map shows good contrast between solvent and protein, and inside the protein region there appear to be long stretches of connected density, probably belonging to α-helices and/or β-strands. The two heterodimers are clearly distinguishable, although we have not yet identified the boundary between the small and the large subunit.

The 8 highest features in the map, showing densities of 6 to 9 times the rms value, are evenly distributed over the two molecules: 4 in one heterodimer (fig. 1) and 4 in the other.

Using these features together with the determined heavy atom distribution, we have found a non-crystallographic symmetry operation which superimposes them with very good correspondance. Our conclusion is that they probably represent the four redox centers. This is confirmed by an anomalous difference map calculated with 90° shifted MIR phases (fig. 1), in which the 6 highest peaks, having densities of 7 to 9 times the rms value, overlap almost exactly with 6 of the 8 highest peaks in the native map. These must therefore represent the 6 iron-sulfur clusters (3 in each heterodimer). The remaining 2 higher peaks in the native map may be tentatively assigned to the Ni-ions. Noteworthy, one of them corresponds to the 7th peak in the anomalous map (peak height 5 times the rms value)! It appears that the iron sulfur clusters and the nickel ion are located in an almost evenly spaced array. If the given assignment is correct, it is not difficult to imagine electrons jumping from one redox center to the next, as the average distance between subsequent centers is ≈12 Å. At the moment we are trying to extend the MIR phases to higher resolution by using density averaging techniques and hopefully we will be able to start model building in the near future.

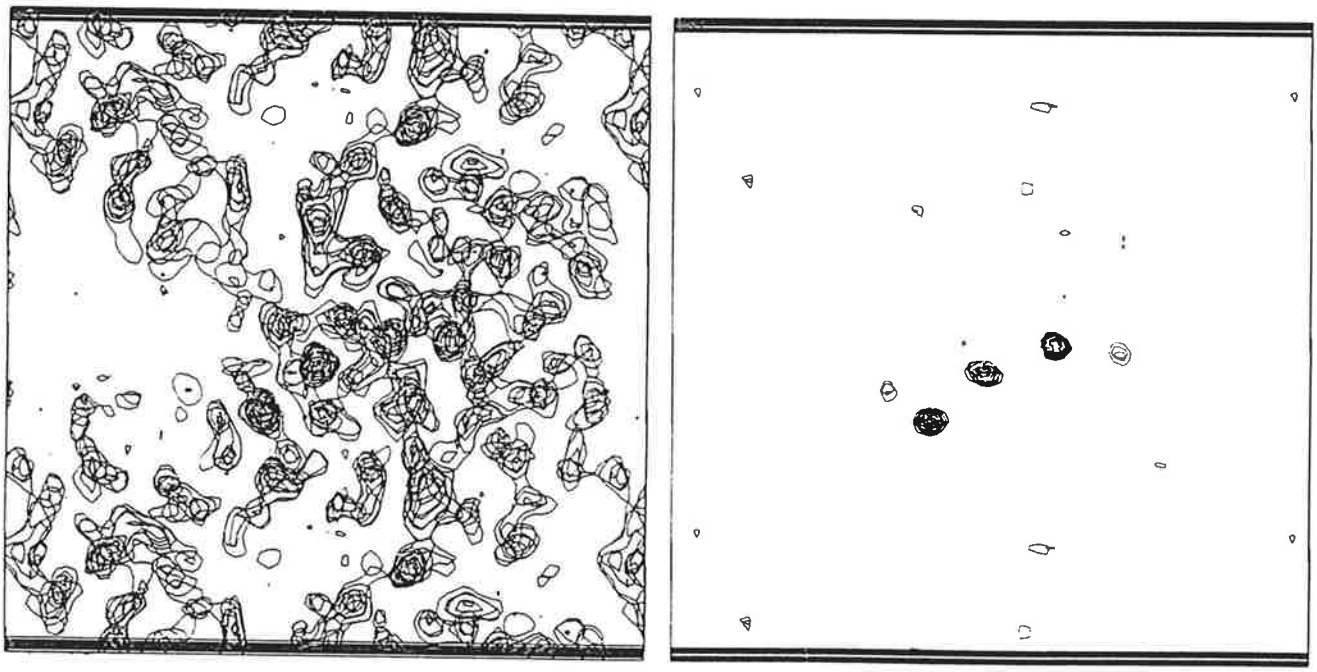


Figure 1. Left: Hydrogenase, part of 5 Å resolution native MIR solvent flattened map, contoured at multiples of 1 sigma, 8 sections stacked on top of each other. Right: Native anomalous difference map using 90° shifted MIR phases, same 8 sections, contoured at multiples of 1 sigma, with starting level of 3 sigma. The peaks are completely independent from the found heavy atom positions. Preliminary assignment from left to right (based on peak height and shape of density): [4Fe-4S], [3Fe-4S], [4Fe-4S] and Ni.

References.

1. G Fauque, HD Peck Jr, JJG Moura, B-H Huynh, Y Berlin, DV DerVartanian, M Teixeira, AE Przybyla, PA Lespinat, I Moura & J LeGall (1988), *FEMS Microbiol Rev* **54**, 299-344.
2. Y Higuchi, N Yasuoka,M Kakudo, Y Katsube, T Yagi & H Inokuchi (1987), *J Biol Chem* **262**, 2823-2825.
3. V Nivière, EC Hatchikian, C Cambillau & M Frey (1987), *J Mol Biol* **195**, 969-971.
4. V Nivière, EC Hatchikian, P Biano & J Haladjian (1988), *Biochim Biophys Acta* **935**, 34-40.
5. EC Hatchikian, M Bruschi & J LeGall (1978), *Biochem Biophys Res Commun* **82**, 451-461.
6. J LeGall, PD Ljungdahl, I Moura, HD Peck Jr, AV Xavier, JJG Moura, M Teixeira & DV Dervartanian (1982), *Biochem Biophys Res Commun* **106**, 610-615.
7. R Cammack, D Patil, R Aguirre & EC Hatchikian (1982), *FEBS Lett* **142**, 289-292.
8. JW vd Zwaan, SPJ Albracht, RD Fontijn & EC Slater (1985), *FEBS Lett* **179** , 271-277.
9. C Fau, M Teixeira, J Moura, I Moura, B-H Huynh, J LeGall, HD Peck Jr & BM Hoffman (1991), *J Am Chem Soc* **113**, 20-24.
10. VM Fernandez, EC Hatchikian & R Cammack (1985), *Biochim Biophys Acta* **832**, 69-79.
11. W Kabsch (1988), *J Appl Cryst* **21**, 916-924.
12. AGW Leslie (1992), *ESF/CCP4 Newsletter* **27**, 30.
13. G Bricogne (1991) in "Isomorpous Replacement and Anomalous Scattering, Proceedings of the CCP4 Study Weekend 25-26 January 1991 compiled by W Wolf, PR Evans & AGW Leslie, pp 60-68.
14. Z Otwinowski, ibid, pp 80-85.
15. BC Wang (1985), *Methods Enzymol* **115**, 90-112.
16. AGW Leslie (1987), *Acta Cryst* **A43**, 134-136.

Progress on the crystal structure of the glycyl tRNA synthetase from *Thermus thermophilus*

Derek Logan[1], Vincent Cura[1], Daniel Kern[2] & Dino Moras[1]

[1]UPR de Biologie Structurale & [2] UPR SMBMR, IBMC du CNRS,
15 rue René Descartes, 67084 Strasbourg Cedex, France

The aminoacyl tRNA synthetases (aaRS) constitute a family of enzymes which carry out the ATP-driven charging of amino-acids onto their cognate tRNA molecules, prior to participation of the adduct in the translation process. The aaRS were recently divided into two classes based on the detection of structural motifs in sequence alignments[1]; this classification was subsequently confirmed by determination of two crystal structures[2,3].

The functional form of the glycyl tRNA synthetase (GlyRS) from *Thermus thermophilus* is a dimer of molecular weight 110–130kDa. We have at present very little biochemical information on the enzyme in this organism, though sequence alignments[1] of the α subunit of its homologue in *E. coli* (an $\alpha_2\beta_2$ heterotetramer) have shown it to be a synthetase of class II, belonging to the third subgroup[4] along with the phenylalanine and alanine enzymes ($\alpha_2\beta_2$ and α_4 respectively). The sequence in *Thermus thermophilus* is under determination in this laboratory. In *E. coli* sequences, the 'dimerisation' motif characteristic of the other class II subgroups could not be identified in GlyRS. Thus, in addition to the elucidation of common elements in substrate recognition and specificity in this family of enzymes, a particular goal of this work is to ascertain whether there may exist alternative modes of dimerisation for class II aaRS.

Crystals measuring up to $0.7 \times 0.7 \times 0.4$mm have been obtained in hanging or sitting drops from a mixture of PEG and NaCl. When exposed to synchrotron radiation at the wiggler station W32 of LURE, they diffract to 3.0Å. Spontaneous nucleation results in two very closely related crystal forms, indistinguishable visually and possessing the same orthorhombic cell: a=125Å, b=254Å, c=104Å. The space groups are $C222_1$ and $P2_12_12_1$, the asymmetric unit of the latter being twice that of the former. We decided to concentrate our efforts on the centred space group, since it presents certain advantages for data collection, though this is negatively compensated by a lack of non-crystallographic symmetry. Also, most derivatives tried in $P2_12_12_1$ showed strong evidence of non-isomorphism. On the basis of initial calculations of the possible V_m, we had believed that there must be a dimer in the asymmetric unit of $C222_1$, but heavy atom sites (in both space groups, see below) correlate perfectly with exactly half of this. Fortunately the solvent content is thus very high, approximately 75%. Large crystals in $C222_1$ can be grown by successive macro-seeding experiments starting with crushed crystals already exposed to X-rays and known to be in the correct spacegroup. Available native data are summarised in Table 1.

The packing in the primitive spacegroup is strongly pseudo C-centred, as might be expected from the essentially identical crystal morphologies: the native Patterson (see Figure 1) contains a large peak at $(0.488,\frac{1}{2},0)$, at 19% of the height of the origin peak, and self-rotation functions calculated within a variety of resolution ranges fail to reveal any molecular 2-fold

space group	resolution (Å)	source	no. crystals	$R_{merge}(I)$ (%)	completeness (%)
$C222_1$	3.0	LURE, IP	5	9.9	91.9
$C222_1$	3.4	rot. anode, IP	1	10.2	97.9
$P2_12_12_1$	4.0	rot. anode, IP	2	9.1	99.9

Table 1: Summary of native data statistics.

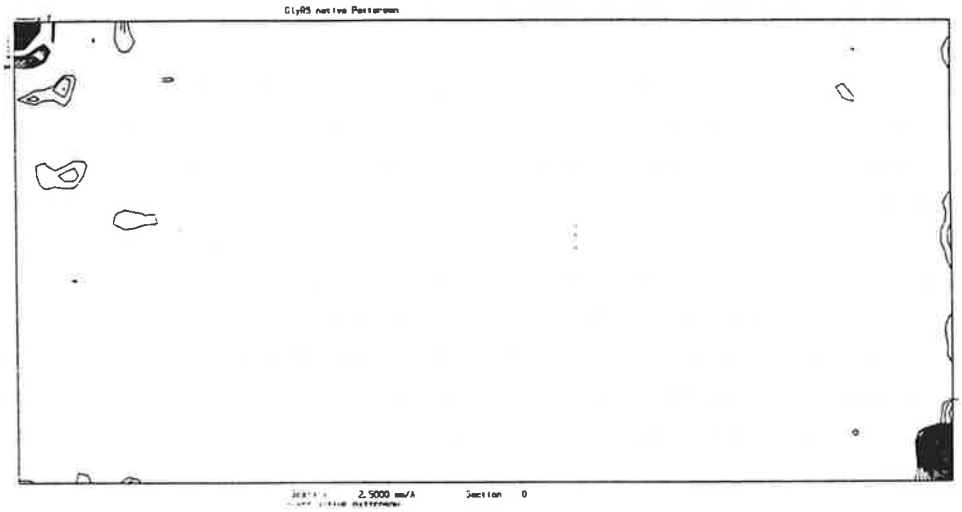


Figure 1: The native Patterson for GlyRS in $P2_12_12_1$.
Section $z = 0$. Resolution limits are 15–4Å, section limits from 0 to $\frac{1}{2}$ in x and y, contour levels are from σ in steps of σ

axes which are not masked by their proximity to the crystallographic dyads. Thus there was already some evidence to suggest that the dimer in $P2_12_12_1$ had its axis aligned almost parallel to y.

Table 2 presents a summary of heavy atom data collection and refinement. In every case, exactly half the number of expected unique, non-crystallographically related sites were found, suggesting that the asymmetric unit contained only one dimer in $P2_12_12_1$ and one monomer in $C222_1$. The unique positions for the former are (x, y, z) and $(0.469 - x, 0.004 + y, \frac{1}{2} - z)$, confirming the postulated alignment of the molecular 2-folds. The $P2_12_12_1$ derivatives have been used to phase to 5Å but as yet have not been exploited further. For tested derivatives whose difference Pattersons proved uninterpretable, 'derivative' Pattersons showed a slight movement of the pseudo-C peak towards $(\frac{1}{2}, \frac{1}{2}, 0)$, suggesting that packing in $P2_12_12_1$ is rather sensitive to heavy atom reagents.

Fewer isomorphism problems were encountered in $C222_1$. Samarium nitrate gave a beauti-

derivative	data coll.	resolution (Å)	completeness (%)	R_{merge} (%)	<FID> (%)
		$C222_1$			
7.5mM $Sm_2(NO_3)_3$	IP, lab	3.6	88.7	9.1	11.9
1mM PCMBS	IP, lab	4.0	79.4	11.9	23.6
0.75mM PCMBS	IP LURE W32	3.6	72.1	7.7 (8.7*)	26.9
		$P2_12_12_1$			
5mM $Yb_2(SO_4)_3$	Xentronix lab	4.5	75.3	11.2	37.7
1mM PCMBS	Xentronix lab	4.0	92.3	9.5	27.2

derivative	sites	occupancy	resolution	php(a)	php(c)
		$C222_1$			
$Sm_2(NO_3)_3$	0.095 0.185 0.261	2.50	12–4.0	0.4	0.1
PCMBS (lab)	0.326 0.150 0.299 0.397 0.364 0.069	4.86 4.47	13–4.6	1.4	0.5
PCMBS (LURE)	ditto	6.24 (2.39*) 5.80 (1.80*)	13–4.0	1.3	0.6
		$P2_12_12_1$			
PCMBS	0.0705 0.3465 0.1995 0.4192 0.3496 0.3006	2.55 2.38	12–5.0	1.0	0.7
$Yb_2(SO_4)$	0.1295 0.1905 0.2366 0.3188 0.1862 0.2620	4.83 4.34	12–5.0	1.1	0.7

Table 2: Summary of derivative data statistics

php(a), php(c) are phasing powers on acentric and centric reflections respectively. * denotes value for anomalous data. <FID> = mean fractional isomorphous difference. Statistics obtained from final phasing cycle of MLPHARE.

fully clear Patterson, but unfortunately the single site lies near a special position ($z=\frac{1}{4}$) and contributes only weakly to the phasing (see Table 2). The other derivatives were solved by a combination of vector search and cross-phasing techniques. Both laboratory and synchrotron PCMBS data sets were kept as separate derivatives, as both were incomplete. In this way, 97% of the observed reflections to 4Å are phased by at least one of the two. It is also possible that differences in occupancy between the two sets gave rise to some weak phase information. A small amount of anomalous scattering data were collected at the synchrotron (about 25% of the unique set), although the crystal morphology rendered alignment difficult in many cases. These contribute very little to the phasing, but have allowed the correct hand to be determined. The mean figure of merit on these three phase sets is 0.41 for acentric reflections and 0.59 for centrics.

Figure 2 shows a slab of the 4Å electron density map, perpendicular to the y axis and chosen

Figure 2: A slab through the electron density for GlyRS at 4Å.
Sectioning is perpendicular to y, and the slab is 13Å thick. The molecular 2-fold axis $(\frac{1}{2}, y, \frac{3}{4})$ is marked by a filled ellipse. The contour levels are from σ in steps of σ. Limits are from 0 to 1 in both x and z.

to coincide approximately with the centre of a dimer. Large solvent channels are observed in the other projections (not shown), and all reveal very few inter-dimer contacts. This would presumably explain the tendency of GlyRS to crystallise in two closely-related spacegroups (we can regard the packing in $P2_12_12_1$ as a 'defective' version of that in $C222_1$), and the sensitivity of the packing to binding of heavy atoms. The shape of the dimer is roughly that of an elongated ellipsoid, in keeping with the rather asymmetric form of all synthetases thus far observed[4]. The 5Å MIR map in $P2_12_12_1$ has the same overall form, allowing for the origin shift of $(\frac{1}{4}, 0, 0)$.

It is not certain at the moment whether further derivatives will be required. Solvent flattening alone is not normally considered a powerful tool for phase extension. However, 10 cycles of classical solvent flattening using the CCP4 package, using a loosely-defined envelope (Wang radius 15Å) and a conservative solvent content of 50% have refined the phases at 4Å to values with a mean figure of merit of 0.92 and an R-factor between $|F_{obs}|$ and $|F_{calc}|$ of 19%. In view of this improvement, averaging between the two crystal forms to the current data limit for the primitive form of 4Å seems unlikely to contribute substantially to the quality of the phases. Preliminary attempts at phase extension to 3.5Å using SQUASH have shown stable behaviour with good statistics, provided a large number of steps are used. We are also currently attempting to introduce partial structure information into the 4Å phases by use of a density correlation function to place the core of the class II synthetase structure (the central 5-stranded β-sheet and two of the other structural motifs characteristic of this class) into the observed map.

References

1) Eriani G., Delarue M., Poch, O., Gangloff J. & Moras D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature **347**, 203–206.

2) Cusack S., Berthet-Colominas C., Hartlein M., Nassar N. & Leberman R. (1990) A second class of synthetase styructure revealed by X-ray analysis of *Escherischia coli* seryl-tRNA synthetase at 2.5Å. Nature **347**, 249–255.

3) Ruff M., Krishnaswamy S., Boeglin M., Poterszman A., Mitschler A., Podjarny A., Rees B., Thierry J.C. & Moras D. (1992) Class II aminoacyl tRNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. Science **252**, 1682–1689.

4) Moras D. (1992) Structural and functional relationships between aminoacyl tRNA synthetases. TIBS **17**, 159–164.

Compression of X-ray images

Jan Pieter Abrahams, MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK, E-mail: jpa@uk.ac.cam.mrc-lmb

The advent of larger and faster area-detectors with larger dynamic ranges is in many instances accompanied by a swell of data in the form of diffraction images. Described is a compression algorithm which reduces the storage space taken up by these images. The data is exactly recoverable after compression and both compression and decompression are fast and do not require extra memory or disk space. Typical images collected in the lab on a MAR-Research imaging plate scanner can be compressed to 20% to 30% of their original size without losing any information. Compressed images can easily be transported from one computer to another without being affected by the byte order of either machine. The algorithm is compatible with any type of header.

Although the algorithm is written in ANSI C, it can effortlessly be combined with FORTRAN code. It has already been incorporated into the following programmes: IMSTILLS, MOSFLM, MADNES, IPDISP, a local version of XIPS and CLIPS, a locally developed alternative to the latter. The algorithm was tested on a variety of machines, running under UNIX and VMS.

On the faster machines like the Indigo R4000 or the Alliant FX/2800, reading a compressed image and decompressing it takes about the same time as reading the original image: the amount of time lost due to decompression is made up by the time gained in reading a shorter file. Compression is similarly fast.

The Algorithm

The algorithm is based on the assumption that the pixel values of a diffraction image are highly correlated: the value of a pixel with indices (i; j) will usually not deviate too much from the rounded mean of the pixels with indices (i - 1; j - 1), (i; j - 1), (i + 1; j - 1), and (i - 1; j). Therefore, the difference between the value of a pixel and the mean value of its neighbours will usually be smaller than the actual value of the pixel. Since it takes less bits to store a smaller value, the image can be compressed.

The actual compression takes place after calculating differences as described above. These differences are coded in stretches, each stretch starting with six defining bits, three of which encode the number of pixels present in the stretch (1, 2, 4, 8, 16, 32, 64, or 128 pixels), while the other three encode the number of bits each pixel takes (in the

present implementation either 0, 4, 5, 6, 7, 8, 16, or 32 bits). The total number of bits in a stretch is therefore the sum of the six defining bits and the product of the two values they encode. The best size in pixels of a stretch is determined by comparing the total number of bits of two consecutive stretches of equal length with the number of bits of the combined stretch. If the sum of the number of bits of the two small stretches is less than the number of bits in the combined stretch, the first of the two small stretches is written to disk and compression continues from the start of the second stretch. If, however, the long stretch contains less bits, it in its turn is treated as the first half of a stretch twice its size and the procedure described above is repeated. As already indicated, stretches never exceed 128 pixels, so when this length is reached, comparison stops, and the stretch is written to disk.

The Implementation

The location of the array containing the image, its dimensions and a filename are passed to the compression routine. The routine opens the specified file in "append" mode, therefore one has to write any header information to the file and close it prior to calling the compression routine. There are two separate compression routines: one compresses two-byte pixels and one compresses four-byte pixels. Similarly there are two decompression routines. Compressed images with four-byte pixels cannot be restored into images with two-byte pixels if any of the original pixels overflow two bytes. It is therefore a good idea to truncate any values larger than $2^{18} - 1$ and to divide values larger $2^{15} - 1$ by -8 prior to compression if one suspects that an image of four-byte pixels might be restored into one consisting of two-byte pixels.

A compressed image starts with the identifier "\nCCP4 packed image, X: ...., Y: ....", where X and Y preceed the number of fast and slow coordinates of the original image. The decompression routine scans a specified file until this identifier is found, skipping any potential header information. At the moment programs capable of reading compressed images recognize them by the filename extension ".pck".

Compressed images are stored as byte streams, and therefore one does not have to keep track of the byte order of the machine on which the data is stored: compressed images can simply be copied as binary files, provided that the header only contains ASCII data.

The routines can be found in two source files: pack_c.c and pack_f.for. The former contains the main code written in ANSI C, the latter contains a FORTRAN interface.

Compression Efficiency

One of the axioms of information theory is that random noise cannot be compressed without loss of information. This allows one to estimate the lower boundary of the compression rate. As a rule of thumb, one can determine this lower boundary by calculating the number of bits needed to represent twice the standard deviation of the

background of the image and dividing this by the number of bits actually used per pixel. Extensive tests with diffraction images containing spots in a fairly uniform background have shown that this lower boundary is about 80% of the actual attained compression rate. Although some compression techniques like those using wavelet transforms, are capable of compressing data even further, they do not allow the exact reconstruction of the original, which makes their use for the compression of diffraction images debatable.

Day to Day Use

The most noticeable effect of compressing one's images is that suddenly three to five times as many are needed to fill up all available disk space. One will also notice that images will have varying sizes, the size mainly being determined by the intensity of the X-ray background. On some of the older VAX machines, decompression tends to be rather slow (two to three minutes for an image of 1200 by 1200 pixels), so in that case one might consider storing original images. Because the result of compressing an image is independent of the byte-order of the machine used, transporting images from one computer to another is further facilitated.

Availability

A file containing the extensively documented code for both pack_c.c and pack_f.for can be retrieved through anonymous ftp from internet 131.111.84.16. Log in as ftp, password ftp, cd pub and copy over the file "pack.all".

Acknowledgements

I am greatly indebted to Andrew Leslie, Phil Evans and Peter Daly for their encouragement, help and enthusiastic criticisms.

MACROMOLECULAR CRYSTALLOGRAPHY AT THE SRS.

Colin Nave, Miroslav Papiz, Pierre Rizkallah, Sean McSweeney,
John Campbell, Steve Kinder, Steve Buffey, Phil Atkinson & Peter Lindley.

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

The SRS is a 2 GeV electron storage ring with λ_c = 3.9 Å and circulating currents of 200 - 300 mA. In addition a 5 Tesla superconducting wiggler magnet is used as a wavelength shifter to give λ_c = 0.9 Å. The SRS was upgraded in 1987 to give a reduced source size of 2.4 mm fwhm horizontally and 0.3 mm in the vertical direction. The machine runs at high currents for approximately 5000 hours per year. During the past 10 years four facilities have been developed for macromolecular crystallography. Each facility has been optimised for a particular type of data collection, but particularly in the earlier beamlines, some flexibilty was in-built to gain experience of different types of experiment. The facilities are currently used for both "state of the art" diffraction data collection from single crystals of macromolecules and the less routine experiments such as multiwavelength anomalous scattering phase determination and time-resolved Laue methods. In the former case demand far exceeds the time available and time is allocated on a peer-review procedure. All the protein crystallographic groups in the UK use the facility, as well as many from overseas. Agreements with the Medical Research Council, Sweden and the EEC provide access for users not supported by the SERC. These agreements not only assist in funding the exisiting facilities, but also enable developments to be pursued.

The SRS is currently in a 10 week shut-down so that the installation of a second wiggler magnet operating at 6 Tesla can be completed; unfortunately there are no macromolecular beam lines planned for this new magnet. It is anticipated that user beam will recommence in early June of this year.

Each of the four stations is briefly described below. The beam intensity is quoted for a circulating current of 200 mA taking into account losses in the beamline; measured values are in reasonable agreement with these figures.

Station 7.2 : Monochromatic.

- Fused quartz mirror giving 1:1 vertical focussing.
- Bent triangular Ge(111) monochromator with 4 mrad horizontal acceptance giving 9:1 horizontal focussing at 1.488 Å.
- $\Delta\lambda/\lambda$ = 0.0004.
- Focal spot size 0.5 mm x 0.3 mm [H x V].
- Intensity 6 x 10^{11} photons/sec/mm^2.

Detector.
- Arndt-Wonacott oscillation camera and film data collection.

Station 7.2 was constructed in 1980 to use radiation from a bending magnet and is available for approximately 75% of the time for macromolecular crystallography. Some wavelength tunability was incorporated into the station design, but it is now operated at a fixed wavelength of 1.488 Å for macromolecular data collection. There are no resources currently available to develope the station. However, it would be a relatively simple task to replace the film detector by an automatic on-line, or off-line image plate system, and development will be undertaken when resources permit.

Station 9.5 : Focussed Laue and Tunable Monochromatic.

All the 9.x stations utilise the radiation emergent from the three-pole supercooled wiggler magnet operating at field strengths of 5 Tesla.

- Platinum coated fused quartz toroidal mirror
- Wavelength range 0.45 - 2.60 Å.
- Water cooled channel cut Si(111) monochromator.
- $\Delta\lambda/\lambda = 0.00015$ Å.
- Aperture 1.4 mrad horizontal, 0.1 mrad vertical.
- Focal spot size 1.5 mm x 0.4 mm [H x V].
- Calculated intensity, 3.6 x10^{11} photons/sec/mm^2 at 1.0 Å within a band pass of 0.00015.

Detectors.

- Modified Arndt-Wonacott oscillation camera and film data collection for Laue work with ultra-fast shutter and opening times down to 50 μsecs.
- Mar-Research 18 cm image plate detector system installed and operating for macromolecular data collection.

Station 9.5 has been constructed via a collaboration with the Swedish Research Council. The station is commissioned for focussed Laue experiments and give the expected reduction in exposure time compared with station 9.7. Several MWAD data sets have also been collected and are undergoing evaluation. The station routinely operates for conventional data collection with the image plate system at wavelengths optimised for anomalous dispersion. Data has been collected from virus crystals (*i.e.* foot-and-mouth-disease virus) to 3.5 Å resolution.

Station 9.6 : Monochromatic.

- Platinum coated fused quartz mirror giving 1:1 vertical focussing.
- Bent triangular Si(111) monochromator with 3 mrad horizontal acceptance giving 8:1 horizontal focussing at 0.895 Å.
- $\Delta\lambda/\lambda = 0.0004$.
- Focal spot size 0.5 mm x 0.3 mm [H x V].
- Intensity, 1.3 x 10^{12} photons/sec/mm^2.

Detectors.

- Enraf -Nonius Fast TV detector with DEP image intensifier.
- R-Axis II Image Plate system.

A new cylindrical mirror has recently been installed which gives a factor of 2 improvement in intensity with respect to the old mirror; the latter had been damaged by use over a number of years. The station has been used to develop Laue and multiwavelength methods, but is principal use is as a fixed wavelength station at 0.895 Å. The reduced wavelength on this station (compared to station 7.2) gives reduced absorption errors, a reduction in radiation damage, and enables high quality, high resolution data to be collected.

The fast TV detector has been upgraded with the DEP image intensifier and is capable of resolving some 80 orders of diffraction. The device should enable automatic data collection from medium sized unit cells with a minimum dead time and 3-D profile fitting; it is complementary to the image plate system. The R-Axis II system can resolve some 200 orders of diffraction, but because of the necessity of using the fine scan (caused by the high collimation of the synchrotron beam) has a dead time unacceptably long for some users of station 9.6; the data quality is high. This device will probably be replaced by a 30 cm Mar-Research single plate system in the near future.

Station 9.7 : Laue.

- Wavelength range 0.2 - 2.6 Å.
- Intensity, 10^{10} photons/sec/mm^2 at 1.0 Å with a band pass of 0.00015.

Detector.

- Modified oscillation camera with high speed shutter and opening times down to 20 msec.

This station is available approximately 25% of the time for macromolecular crystallography. Its advantage with respect to station 9.5 is the availbility of wavelengths < 0.5 Å.

Support Facilities.

In addition to the 4 crystallographic stations, various support facilities are available to assist users as follows;

- Crystal cooling. Cooling of crystals between +10 and -20^{o}C and down to near liquid nitrogen temperatures can be achieved on all the beam lines.
- A microdensitometer is available for film scanning.
- The Biological Support Laboratory is available, by prior request, for sample preparation and mounting. The laboratory has a full range of microscopes. By arrangement, protein isolation, purification and crystallisation and/or other biochemical experiments, can also be undertaken by users in the BSL.

- stations 9.5 and 9.6 are equipped with specific workstations for data acquisition and processing, but a computing infra-structure is available for users, by arrangement, for all aspects of data processing.
- Daresbury Laboratory maintains the CCP4 programme suite and advice is available.

In-house Macromolecular Research Programme.

Macromolecular crystallographic facilities at the SRS come under the auspices of the joint Medical Research Council and Science and Engineering Research Council Biological Structure Programme. This programme is supplemented by a vigorous in-house research activity. The in-house programme not only aims to develope new techniques such as MWAD, Laue methods and cryogenic crystallography, but also undertakes structure-function studies in a number of areas including plasma proteins, proteins of the vertebrate eye lens, DNA binding proteins, light harvesting proteins and snake toxins. The in-house programme enjoys collaborations with a number of Higher Education Institutes including the Universities of London (Birkbeck College), Keele, Glasgow, Liverpool and Manchester. Scientists wishing to explore the possibility of undertaking biological research at Daresbury Laboratory should contact Professor Peter Lindley in the first instance.

SRS SPECIALIST USER GROUP FOR LAUE DIFFRACTION - APRIL, 1993

John Campbell (SERC Daresbury Laboratory), Benson Kariuki (University of Liverpool)

The second meeting of the SRS Specialist User Group for Laue Diffraction was held at the Daresbury Laboratory on Thursday April 15'th, 1993. It was chaired by Dr Marjorie Harding from the University of Liverpool. Scientific talks were presented in the morning and the experimental facilities and current state of the Laue data processing software were reviewed in the afternoon. We were particularly pleased to be able to invite Ilme Schlichting from the Max Planck Institute, Heidelberg to the meeting.

Scientific Section

a) Time Resolved Laue Diffraction Studies on Cytochrome p450 and Myoglobin

Ilme Schlichting, Max Planck Institute, Heidelberg

Ilme first discussed, in general terms, the problems associated with crystallographic kinetics and time resolved studies. She then described an attempt to investigate the oxygen complex (half life around 10 minutes at 5 degrees Kelvin) involved in the reaction mechanism of Cytochrome p450. The experimental details were described together with specroscopic studies used to follow the reaction. Laue data had been collected and processed but so far the oxygen is not visible in the resulting maps. In the second part of her talk, she described work done in preparation for Laue studies to investigate the Myoglobin complex with Carbon Monoxide. The problems of getting crystals which would give good Laue diffraction patterns when cooled to low temperatures was discussed.

b) Deconvolution of Multiples and Anomalous Dispersion Refinement

Hao Quan, University of Liverpool & Daresbury Laboratory

Quan first described two methods he had developed for deconvoluting Laue harmonics data using direct methods. The first was carried out in reciprocal space and the second was carried out in real space using modified Patterson functions. Results were presented for the 'red' crystal, Insulin and Cytochrome C Peroxidase.

In the second part of his talk, Quan described the use of wavelength dependent structure factors used for refining structures which contained heavy atoms with large anomalous dispersion. George Sheldrick had incorporated such a facility in his SHELXL92 version of SHELX. Results were reported for the refinement of an organometallic compound containing one gold and three Osmium atoms (AUOS).

c) Prediction of Laue Patterns with Streaky Spots

Don Nguti, University of Keele

In some cases where spots are streaked on the Laue diffraction patterns, the streak lengths are not uniform across the pattern or not simply a function of the distance from the centre of the pattern. Don described

how, by modelling the reciprocal lattice points using three different values for mosaicity along the reciprocal cell axes it was possible to generate the types of streaked patterns observed in such cases. The patterns from Met-tRNA Synthetase were used as a successful example of such an approach.

d) Choice of Wavelength Ranges for Laue Diffraction

Liz Duke, University of Oxford

Liz offered some thoughts on the choice of wavelength range for use in Laue diffraction; in particular the advantages of reducing the wavelength range from say 0.2 to 2.1 Angstroms to an 'octave' in the range of say 0.6 to 1.2 Angstroms were considered. The advantages of reducing the number of multiples present need to be offset against having to collect more images to get the same coverage of reciprocal space. It might be particulary advantageous to take this approach with image plates where the UNSCRAM method of deconvoluting harmonics would not normally be available.

Review of Experimental Facilities

a) Station 9.7

Simon Clark summarised facilities on the station which had been made available since the previous meeting. These were as follows:

1) A Molecular Dynamics Image Plate reader was available for use.

2) 5" x 5" image plate holders were being made; these would fit on the existing camera carousel. It would be possible to fit 6 of these simultaneously on the image plate reader. A total of six plates would be available. Three or four A4 size image plates were also available.

3) A new furnace and a new high pressure cell were available.

4) New 'pincer' software was available for station control.

b) Station 9.5

Sean McSweeney described work carried out to enable more control over the bandpass used for Laue work. The limitations of filters such as Aluminium were discussed. He then described work carried out in collaboration with Manchester University in the use of very thin semi-transparent mirrors to cut off the higher wavelengths whilst allowing transmission of the lower wavelengths. Experiments had been carried out using 1.5 micron thick Mylar films. Discussions were taking place about some possible alternative materials. To achieve a sharp cutoff, the films need to be very flat; they also need to be very thin to reduce the transmission path length as much as possible. Some test exposures were taken with Concanavalin A. It was hoped that a useable system could be made available within 12 months.

He also reported that John Helliwell had experimented with his 'toastrack' detector on the station.

There is a possiblilty of a large (300mm) MAR image plate system being ordered at some stage; Off-line image plates will also be available for use.

Software Developments

a) Daresbury Laboratory Laue Software Suite

Before giving some details on the current state of the Laue software, John Campbell reported on a test carried out to deconvolute Laue harmonics data, recorded on image plate, using the varying nature of the wavelength normalisation curve. The calculations were carried out using a modified version of the LAUENORM program with a Lysozyme Laue image as the test example.

The Laue programs had now been converted to use MTZ reflection data files instead of the former LCF files.

The Laue Demonstation test data sets and documentation had been updated. It now included an image plate example. The data were distributed with the software Suite.

The Laue documentation had been updated and some copies of the complete documentation were available.

The X-windows based LAUEGEN program (version 1.0) was now released and documented. Further developments were planned.

b) Automatic Data Processing

Marjorie Harding shared some thoughts with the meeting on how the process of Laue data processing might be made more automatic; this would include automatic refinement procedures and iterative determinations of improved soft limit values. The aims would be to increase throughput and make processing easier for the non-specialist.

Some discussion followed on how far this process could be taken and whether there were drawbacks if the approach became too 'black box'. It was planned to include at least some of the ideas in future developments of LAUEGEN. There was support for the idea of carrying out all the parameter refinements at the LAUEGEN stage and providing an 'integration module' which would do the integration without any additional refinement steps.

Structural studies of some transthyretin variants

Ana Margarida Damas, Instituto de Ciencias Biomédicas Abel Salazar, Universidade do Porto, Portugal.

The objective of this research is to improve our knowledge of the molecular basis of transthyretin related amyloidosis. Human transthyretin (TTR) is a plasma protein involved in the transport of thyroxin and vitamin A. In 1983 a variant TTR was described as the major component of the amyloidogenic material present in patients with Familial Amyloidotic Polineuropathy (FAP) (Saraiva *et al*, 1983). This disease, first described by Andrade in 1952 is characterised by a lower limb neuropathy with a predominant involvement of the peripheral nervous system (Andrade *et al*, 1970).

Over the past few years other pathogenic and non pathogenic mutations have been described in the literature. Most of the pathogenic variants are related to amyloidosis but the clinical symptoms are not the same for all the variants. They include polineuropathies, cardiac failures, vitreous problems or carpal syndrome (Saraiva *et al*, 1991). A point mutation, which occurs in different places of the polypeptide chain was described for most of the variants.

A molecular graphics study was performed in order to access the spatial positions of the mutations in the TTR molecule. This analysis revealed that the modifications occur mainly at the surface but they are not confined to the same region of the molecule (Damas *et al*, 1990).

Crystallographic studies of the variants should clarify the mechanisms that lead to the aggregation of the protein into fibrils. We have started crystallographic studies on TTR Met30 and TTR Ile122 variants because they are the most frequently expressed and although they both produce amyloid they lead to distinct clinical symptoms. The first variant is associated with FAP and Portugal is the largest focus of the disease in the world; the second mutant produces Familial Amyloidotic Cardiomiopathy (FAC), a disease characterised by cardiac failure and amyloid deposits in the heart and was found in the black population (Gorevic *et al*, 1989).

The three-dimensional structure of wild type TTR was determined by Blake (1978), who described it as a tetrameric structure composed of four identical subunits, each of them with 127 aminoacids. There is a channel running through the molecule where the two equivalent binding sites for the thyroxin hormone are located. The exact location of the binding site for retinol binding protein is not known.

The variants were expressed and purified (Furuya *et al*, 1991) and crystals with dimensions suitable for X-ray diffraction experiments were grown from ammonium sulphate solutions. The crystals are isomorphous to the wild type protein, they belong to the orthorhombic space group $P2_12_12$ and the deviation of the cell dimensions from the values reported by Blake is less then 10%. They are stable in the X-ray beam and diffracted to 2.5Å and 2.7Å on a conventional X-ray source. The data were collected on a Enraf-Nonius FAST area detector diffractometer using graphite monochromatic Cu Ka radiation from a Nonius FR571 rotating anode generator. In both cases coordinates of the native protein were used as a starting model for the calculation of Fourier maps and their analysis revealed close similarities between these mutants and the native protein.

The results of the X-ray crystallographic studies of recombinant TTR Met30 have been reported (Terry *et al*, 1993). In this study, a major alteration was described for the position of Cys10 which becomes more exposed to the solvent. It was suggested that these residues might form disulphide bridges between different molecules, giving rise to the fibrils.

Hamilton and co-workers have also reported their X-ray crystallographic studies on TTR Met30 isolated from a homozygous patient (Hamilton *et al*, 1992). Their results show an increase in the unit cell volume, consistent with larger hydrogen bond lengths related to the interaction between the dimers. They have indicated the existence of structural differences between the wild type and the mutant protein but, none of these variations was considered large.

Fourier maps of recombinant TTR Ile122 are currently being interpreted.

Future experiments will be focused toward collecting higher resolution data on these and other derivatives. We hope that these studies provide the information necessary to understand the importance of each substitution in the amyloidogenesis phenomena.

An on-going close collaboration exists between our group and Blake´s group, in Oxford.

References

Andrade, C., Arakis and Block, W.D. (1970) *Arthr Rheum*, **13**, 902-915.

Blake, C.C.F., Geisow, M., Oatley, S., Rerat, B. and Rerat, C. (1978) *J. Mol. Biol.* **121**, 339-356.

Damas, A., Terry, C. and Blake, C. (1990) *Arq. Med.*, **3**, 55-57.

Furuya, H., Saraiva, M.J.M., Gawinowicz, M., Alves, I., Costa, P. P., Sasaki, H., Goto, I. and Sakaki, Y. (1991) *Biochemistry*, **30**, 2415-2421.

Gorevic, P., Prelli, F., Wright, J., Pras, M. and Frangione, B. (1989) *J. Clin. Invest.* **83**, 836-843.

Hamilton, J., Steinrauf, L., Liepnieks, J., Benson, M., Holmgren, G., Sandgren, O. and Steen, L. (1992) *Biochim. Biophys. Acta*, **1139**, 9-16.

Saraiva, M. J. M., Goodman, D. S., Costa, P. P., Canfield, R. E. and Birken, S. (1983) *Clin. Res.* **31**, 533.

Saraiva, M. J. M. (1991) *Neurom. Disorders*, **1**, 3.

CRYSTAL STRUCTURE OF LIGANDED AND UNLIGANDED FORMS OF PLASMA RETINOL-BINDING PROTEIN

Giuseppe Zanotti, Rodolfo Berni*, Simone Ottonello* and Hugo L. Monaco&

Department of Organic Chemistry, University of Padova and Biopolymer Research Center, C.N.R., 35131 Padova, Italy; *Institute of Biochemical Sciences, University of Parma, 43100 Parma, Italy; &Department of Genetics and Microbiology and CISMI, University of Pavia, 27100 Pavia, Italy

Retinol-Binding Protein (RBP), the specific carrier of retinol in plasma, has been isolated from several vertebrates: mammals, chicken and fish. In every case, plasma RBP has been found to be a single polypetide chain containing one single binding site for retinol. Its function is to transport the vitamin from the liver to specific cell surface receptors and it circulates in mammalian plasma bound to another protein, Transthyretin (TTR), formerly called prealbumin, as a 1:1 molar complex[1]. The 3-D structure of the human *holo*RBP in the orthorombic crystal form is known[2]. We have investigated the crystal structures of human and bovine *holo*RBPs and of the unliganded forms.

Human Retinol Binding Protein (RBP) (trigonal crystal form)[3]. The three-dimensional structures of the liganded and unliganded forms of human plasma Retinol Binding Protein (RBP) in the trigonal crystal form have been solved at 2.5 Å resolution. The final model of RBP complexed with retinol (*holo*RBP, space group R3, **a=b**=104.0 Å, **c**=74.4 Å) has a crystallographic R factor of 0.176 for 9652 reflections. The unliganded form, obtained through a purification procedure which included steps based on hydrophobic interaction chromatography, crystallized isomorphously with *holo*RBP and its structure has been refined to an R factor of 0.190 for 9614 reflections. The structure of the trigonal *holo* protein is quite similar to that of the orthorhombic form: the r.m.s. deviation of all the equivalent α-carbons in the two chains is 0.53 Å.

Bovine Retinol Binding Protein (bRBP)[4]. The bovine retinol-RBP complex (*holo*bRBP) has molecular mass (21 kDa), amino acid composition, absorption and fluorescence spectra and binding affinity to TTR very similar to those of the human complex. However, the bRBP-TTR complex has the peculiar property of being significantly dissociated when plasma proteins are run through an ion-exchange DEAE-sephadex column during the classical procedure used for the purification of RBP from most species. The three-dimensional structures of bovine plasma Retinol-Binding Protein (bRBP) complexed with retinol (space group $P2_12_12_1$, **a**=46.08, **b**=49.12, **c**=76.10 Å) and of the unliganded protein prepared *in vitro* by extracting retinol with ethyl ether (space group $P2_12_12_1$, **a**=46.55, **b**=48.97, **c**=76.87 Å) have been solved at 1.9 and 1.7 Å resolution, respectively. The final crystallographic R factors are 0.190 for *holo*bRBP and 0.196 for the unliganded bRBP.

The model for the bovine *holo*protein is quite similar to that of the human protein, with which it exhibits 92 per cent sequence similarity. The r.m.s. deviations

between the α-carbons in the two proteins is 0.31 Å. The retinol binding site is almost completely preserved. The loops that surround the opening of the β-barrel are also particularly conserved, in contrast with the presence of several substitutions in parts of the RBP molecule opposite to the opening of the calyx that binds retinol.

Despite the fact that unliganded bovine RBP was prepared and crystallized using procedures completely different from those used to obtain the unliganded human RBP, the conformational differences between unliganded and liganded forms of bRBP are almost identical to those previously found between the same forms of human RBP. The model of unliganded bRBP is shown in Fig. 1, superimposed to that of *holo*bRBP. The r.m.s. deviation is 0.31 Å between α-carbons of bRBP and 0.56 Å between the two unliganded proteins. It is evident from the figure that, excluding some small differences in the region from amino acid 61 to 68, which is a highly mobile area and not well defined in both proteins, the only real difference between the *holo*protein and the unliganded model involves amino acids from 34 to 37. In particular, the amino acids that undergo a substantial conformational change are Leu 35 and Phe 36. This situation is very similar to that described for the liganded and unliganded forms of human RBP in the trigonal crystal form. Fig. 2 illustrates the movements of the two amino acids side chains: Phe 36, that in the *holo*protein points toward the interior of the cavity, is positioned in the unliganded protein in the place previously occupied by the hydroxyl group of retinol; Leu 35, owing to a rotation of the main chain, points now clearly towards the exterior of the protein. This last movement is made possible by the absence of retinol in its binding site.

Interaction of the unliganded bRBP with TTR. We have verified that the unliganded bRBP, obtained by retinol extraction with an organic solvent, exhibits distinct, although not remarkably different, elution profile as compared to liganded bRBP, when both forms are subjected to chromatography on a human TTR-Sepharose 4B affinity column. Both forms are retained by the affinity matrix at high ionic strength. However, unliganded bRBP is eluted at approximately 10 mM NaCl, whereas the elution of the bovine *holo*protein occurs at almost negligible ionic strength. As the chromatographic behavior is not remarkably different for both forms of RBP, presumably their affinities for TTR are not drastically different as well. Therefore, the quite limited conformational differences that we have found between liganded and unliganded bRBPs may be responsible for binding affinities to TTR not particularly different for the two RBP forms. Finally, the finding that the conformational differences between the liganded and unliganded forms of bRBP are confined to a limited region of the RBP molecule, coupled to the observation that such changes affect the interaction with TTR, represents the strongest evidence obtained so far that the loop comprising residues from 32 to 37 is part of the site that binds TTR.

REFERENCES

1. Goodman, D. S., 1984, in *The Retinoids"*, Sporn, M.B., Roberts, A.B. and Goodman, D. S., eds, Academic Press, Inc., Orlando, pp. 41-88

2. Cowan, S.W., Newcomer, M.E., and Jones, T.A., 1990, PROTEINS: Struct., Func., Genet. **8**, 44-61

3. Zanotti, G., Berni, R., Ottonello, S. and Monaco, H.L. (1993) *J. Mol. Biol.* **230,** in

press
4. Zanotti, G., Berni, R. and Monaco, H.L. (1993) *J. Biol. Chem.* in press.

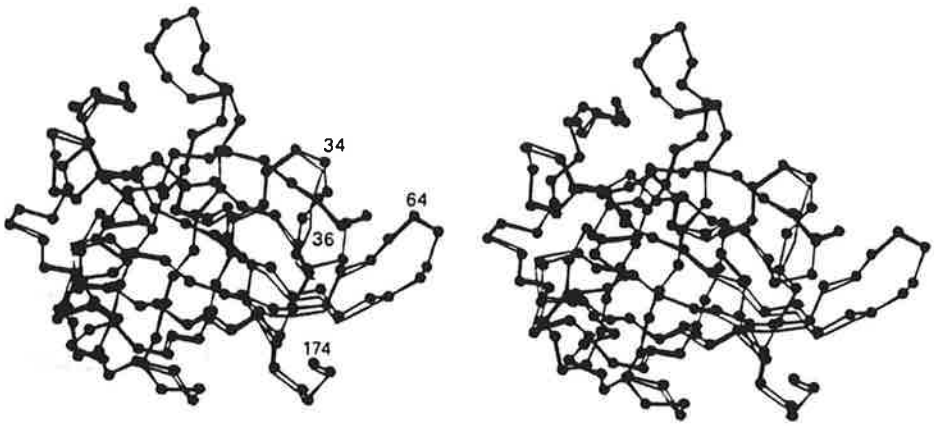


Fig. 1. alpha-Carbon chain trace of *holo*bRBP (black dots) superimposed to the unliganded protein (thin line).

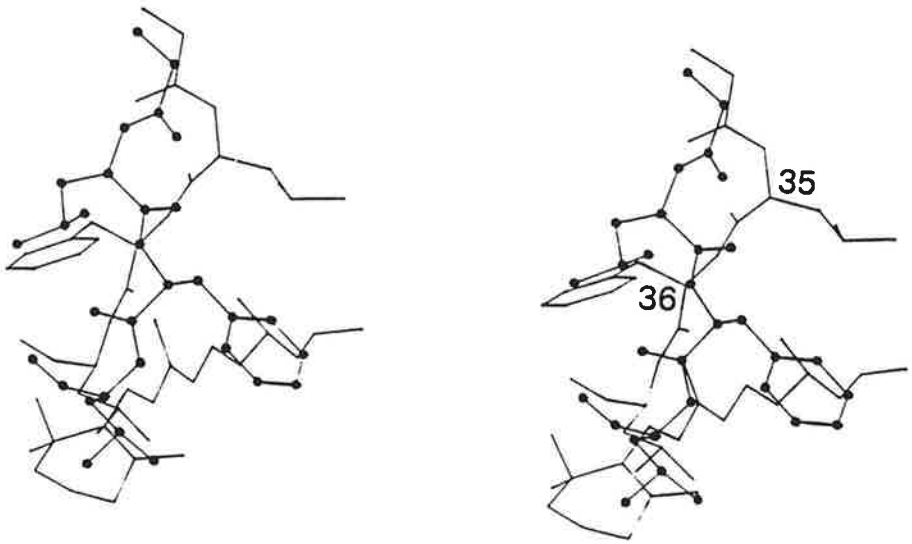


Fig. 2. Conformation of amino acids 35 and 36 in *holo*bRBP (thin line) and in the unliganded protein (black dots). Side chain of Phe35 in the latter is in the position occupied by retinol molecule in the *holo*protein.

Clemens Vonrhein

Institut für Organische Chemie und Biochemie
Albertstr.21
D-W-7800 Freiburg im Breisgau
E-mail: vonrhein@bio5.chemie.uni-freiburg.de

Computing Difference-Patterson Maps With X-PLOR

Fast and easy computation of difference-Patterson maps can be done with the widely distributed program suite X-PLOR [1]. To compute such a map it is necessary to store the native and derivative structure factors in 'FOBS' and 'FCALC', respectively. For a difference-Patterson map only (FPH-FP) differences for common reflections should be calculated. This can be achieved by multiplying each difference with: a), 1 (if FPH_{hkl} and FP_{hkl} are not zero), and b), 0 (if FPH_{hkl} or FP_{hkl} are zero), respectively. An example input-file is shown below. The scaling of derivative to native data can be carried out in X-PLOR itself or by other programs like 'CMBISO' [2] or 'FHSCAL' [3] which allow a fast transfer of the data into X-PLOR-format.
Scaling within X-PLOR is not very accurate. Clear solutions for weakly occupied heavy atom sites with unambiguous cross-peaks could be achieved only with 'CMBISO' and 'FHSCAL'. Nevertheless, highly occupied sites can be interpreted even in derivatives which are poor with respect to completeness and R_{sym}.

```
remarks COMPUTING DIFFERENCE-PATTERSON MAP

xrefine

  a=46.78 b=56.90 c=80.99 alpha=90.0 beta=103.76 gamma=90..0
  symmetry=(x,y,z)                                    { symmetry operators without  }
  symmetry=(-x,y,-z)                                  { translational part          }
                                                      { e.g. P21(unique b)          }
  nreflections=12000
  reflection @derivative.fob end                      { derivative dataset          }
  resolution 20.0 4.0

  do amplitude (fcalc=fobs)                           { stores derivative data in   }
  do phase (fcalc=0.0)                                { FCALC with phase=0          }

  reflection @native.fob end                          { native dataset              }
  resolution 20.0 4.0                                 { stored in FOBS              }

  do scale (fcalc=fobs)                               { scaling of derivative to    }
                                                      { native data                 }

  do amplitude (fcalc=(fcalc/(fcalc+1E-12))*(fobs/(fobs+1E-12))*((fcalc-fobs)^2))

                                                      { calculation of |FPH-FP|^2   }
                                                      { only for common reflections }

  method=fft
  fft
    grid=0.25
  end
  map                                                 { computing of formatted map  }
    automatic=false                                   { for the whole unitcell      }
    extend=unit
    output=xplor.map
  end
end
stop
```

[1] Axel T. Brünger, X-PLOR (Version 3.1) Manual (Yale University, New Haven, CO, 1992).

[2] William Furey, PHASES Manual (University of Pittsburgh, Pittsburgh, USA).

[3] CCP4, SERC UK Collaborative computing Project no.4 (Daresbury Laboratory, Warrington, UK, 1992).

MASKS MADE EASY

Gerard J. Kleywegt & T. Alwyn Jones
Department of Molecular Biology
Biomedical Centre, University of Uppsala
Uppsala - Sweden

One of the prerequisites for using density-modification techniques (such as solvent-flattening and NCS-averaging) is the availability of a high-quality molecular envelope. We call these envelopes "masks" and use them in our local set of density-averaging programs (**1,2**). A mask is a "logical print" of the molecule on a grid which has points set to "1" if they are inside the molecular envelope, and to "0" otherwise.
Up until recently, masks were typically generated from a PDB or BONES file and subsequently edited with "O" (**3**). This editing process tended to be tedious and time-consuming and many users were less than fond of it. We therefore wrote a new program, called MAMA (MAsk MAnipulation) with which masks can be generated within seconds and turned into high-quality envelopes within the hour.

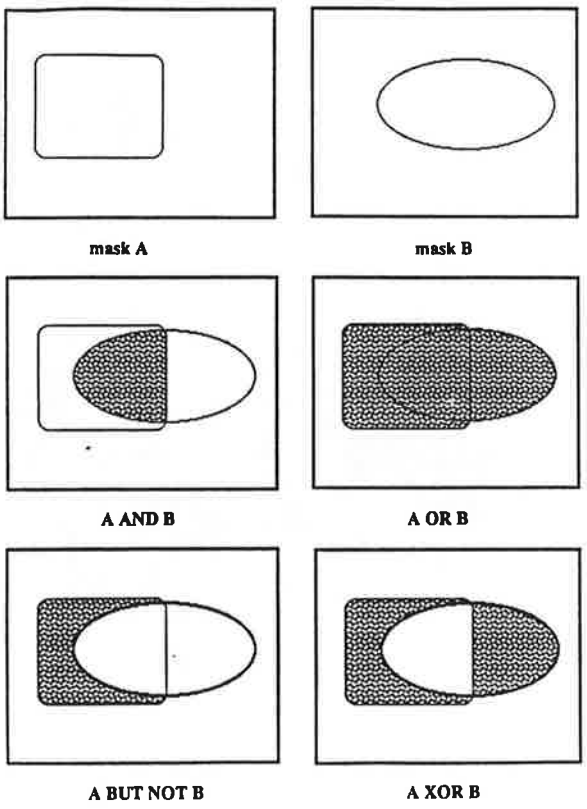


Figure 1) Illustration of some of the logical operations that can be performed on masks with MAMA.

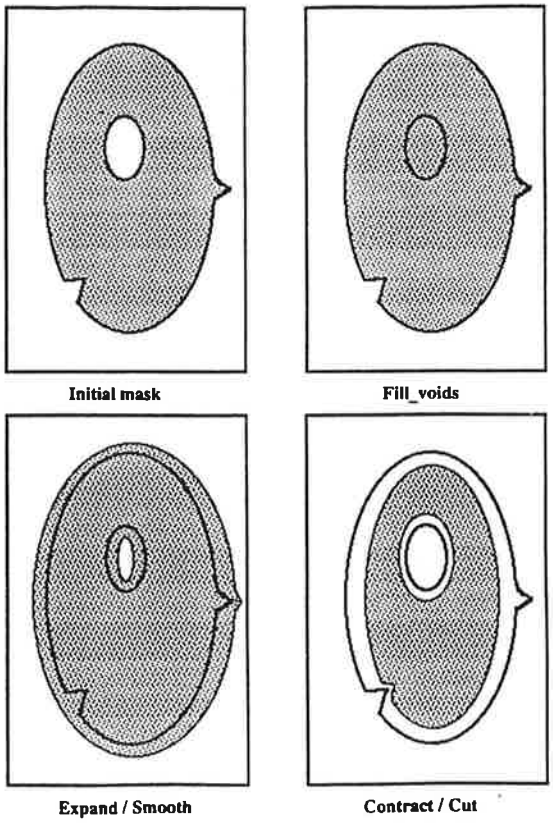


Figure 2) Illustration of some of the mask-manipulation tools available in MAMA.

GENERATING MASKS

Masks can be generated in three ways with MAMA:

1 - from a PDB file
2 - from a BONES file
3 - from one or more existing masks

In the first two cases, each (BONES) atom gets a radius associated to it, and all grid points which lie within this radius from the atom's position are included in the mask (*i.e.*, set to "1"). The third option enables users to "recycle" good masks and even to carry them over from a structure in one spacegroup to, for example, a mutant structure in a different spacegroup.

A unique feature of MAMA is that it is able to transform a mask into a different spacegroup, unit cell, and/or position while keeping the volume of the mask virtually constant (typically, within 0.3 %). This feature is a *conditio sine qua non* for multiple- spacegroup averaging where one wants to edit a mask in all two, three or more spacegroups and therefore requires a convenient mechanism to move the mask from one spacegroup to the next with as little distortion as possible. Alternatively, one may for instance combine two monomer masks into one large dimer mask, etc. Some of the operations which have been implemented to combine masks are: AND, OR, XOR, BUTNOT and UNITE (see Figure 1).

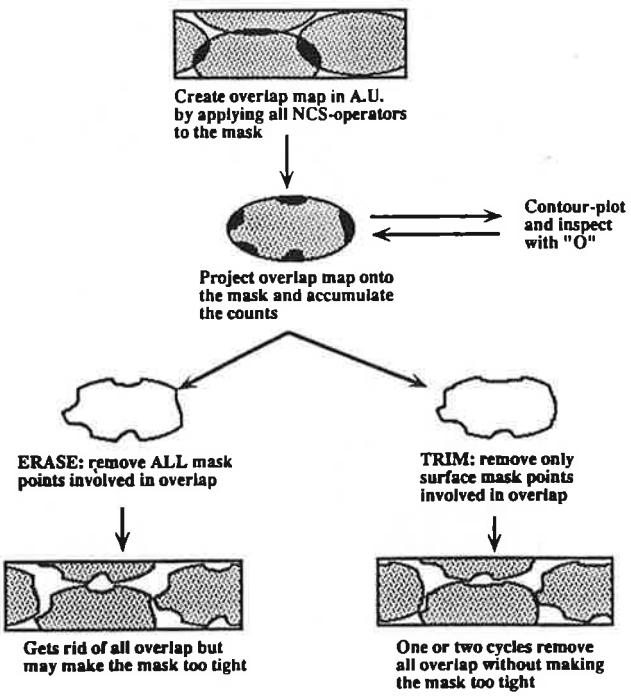


Figure 3) Methods for removing parts of a mask which overlap due to (non-) crystallographic symmetry.

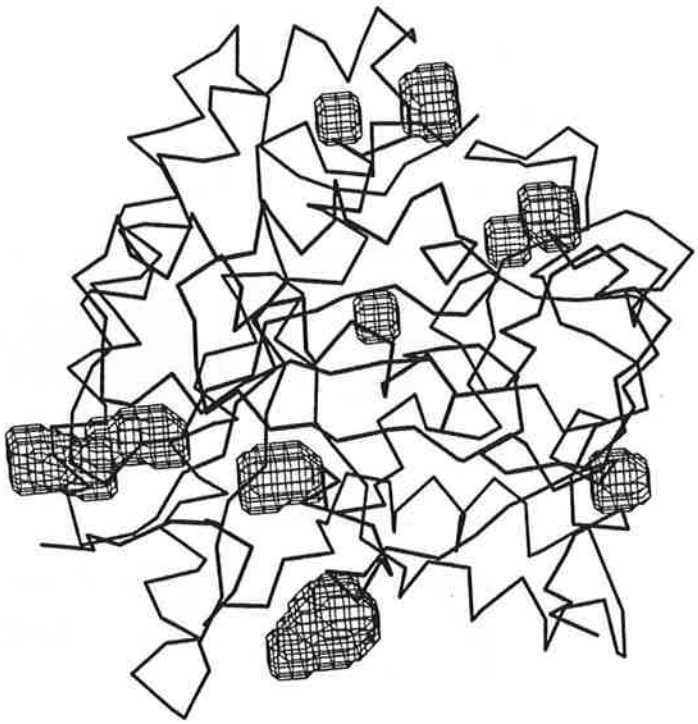


Figure 5) Example of the visualisation of cavities, in this case in CBH II (J. Rouvinen *et al.*, *Science* **249**,1990, 380).

IMPROVING MASKS

We define a "high-quality mask" as one which:

a - covers (almost) all atoms in the molecule(s)
b - has no internal voids
c - has no "droplets" which are not connected to the bulk of the mask
d - is fairly smooth
e - does not extend too close to the borders of its grid
f - has little overlap with (non-) crystallographically related copies of itself.

MAMA contains options which check criteria (a) and (e). Internal voids (b) and isolated "mask droplets" (c) can also be removed automatically by MAMA. Criterion (d) can be assessed by visual inspection with "O" or by calculating the surface-roughness of the mask with MAMA. In order to make a mask smoother (and larger) or rougher (and smaller), simple binary-image processing techniques have been implemented (in particular, EXPAND and CONTRACT operations; see Figure 2).

MAMA also contains an option to remove overlap areas, *i.e.* parts of the mask which overlap with some (usually) other part of the mask due to (non-) crystallographic symmetry (see Figure 3). Of course, the "final touches" to a mask will often still be applied within "O" (for example, making room for insertions and sidechains in a mask generated using a molecular-replacement search model). Nevertheless, the generation of a high-quality mask, using MAMA and "O", has become a simple exercise which can be carried out within one or two hours (see Figure 4).

SPIN-OFFS

After MAMA had been written, it turned out that the program has a number of unexpected additional capabilities:

1 - clever use of the image-processing and logical mask operations enables emulation of Delaney's cavity-detection program (**4**); see Figure 5.

2 - the overlap-removal option produces an "overlap map" which may be contoured in "O". This can be used to detect inter- molecular contacts due to (non-) crystallographic symmetry.

3 - by generating masks around two different, but aligned, molecules, one may assess the similarity of the two molecules' shapes: SI = Nab / SQRT (Na * Nb), where Nab is the number of points which both masks have in common, Na is the number of points in mask A and Nb in mask B; see Figure 6.

AVAILABILITY

MAMA is part of a forthcoming averaging package, RAVE (**2**), which will accompany "O". For more information, contact TAJ (E-mail: "alwyn@xray.bmc.uu.se").

REFERENCES

(1) T.A. Jones, "a, yaap, asap, @#*? A set of averaging programs", in *"Molecular Replacement"* (E.J. Dodson et al., Eds.), SERC Daresbury Laboratory (1992), pp. 91-105.

(2) G.J. Kleywegt & T.A. Jones, "Convenient single and multiple-crystal real-space averaging of macromolecular electron-density maps", to be published.

(3) T.A. Jones, J.Y. Zou, S.W. Cowan & M. Kjeldgaard, "Improved methods for building protein models in electron density maps and the location of errors in these models", *Acta Cryst.* **A47** (1991), 110-119.

(4) J.S. Delaney, "Finding and filling protein cavities using cellular logic operations", *J. Mol. Graph.* **10** (1992), 174-177.

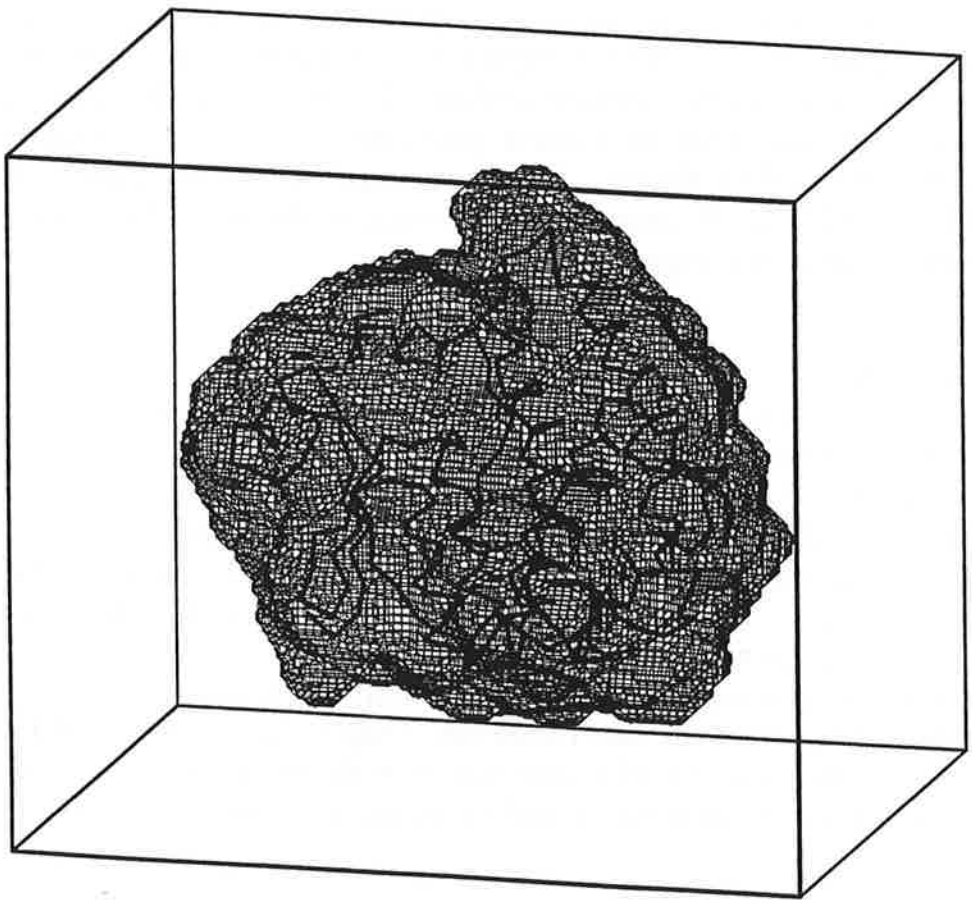


Figure 4) A high-quality mask around a monomer of human alpha-class glutathione transferase A1-1 (I. Sinning *et al.*, *J. Mol. Biol.*, in press).

Simple example of the molecular replacement technique: The structure determination of 3-phosphglycerate kinase from *Bacillus stearothermophilus*

by

Gideon J. Davies

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

Molecular replacement is now an extremely widely used method for the determination of protein structures from experimentally determined structure factor amplitudes. Before going on to examine some of the more complicated applications of the molecular replacement method, described later in this booklet, it is, perhaps, useful to study a simple "test case" and analyse the factors which can lead to success or failure of the method.

1. INTRODUCTION

The introduction of the molecular replacement method revolutionised protein structure determination. Given a suitable model structure, gone was the need for derivative screening, gone was the need for nasty heavy-metal reagents and gone were the problems of non-isomorphism. Whilst these facts are more-or-less true, what replaced these problems was a whole new set, which in the hands of the inexperienced or simply unlucky could lead to the failure to solve the protein structure, or worse still, to a completely incorrect structural determination. The structure determination of 3-phosphoglycerate kinase (PGK) from *Bacillus stearothermophilus* was not such a case, but analysis of the the molecular replacement method, as applied here, gives some insight into the use and applications of the method as well as suggesting reasons why the method can often fail to give the correct solution.

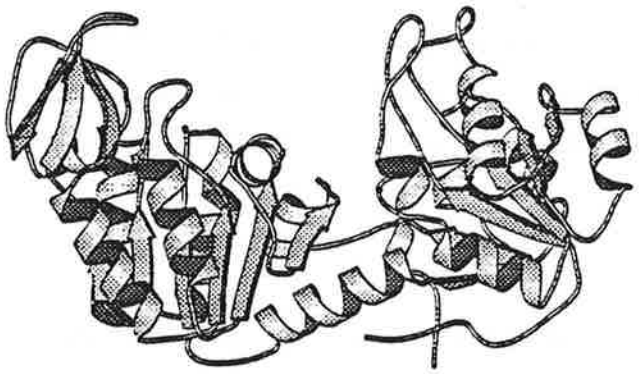
The molecular replacement method as used to solve the structure of similarly related proteins is a six-dimensional problem, normally split into two three dimensional calculations, the *rotation* and *translation* functions. As demonstrated by David Blow and Eleanor Dodson in the study weekend of 1985, these rotation and translation functions are most easily visualised by trying to superimpose two similar coffee cups ! In this paper we examine how two simple aspects of the rotation and translation function calculations, the accuracy of the trial structure and the completeness of the

observed data, can lead to failure to correctly determine the correct rotation and translation parameters when proteins rather than coffee-cups are involved.

2. THE SYSTEM

The work described in this paper was performed as part of a study into the factors involved in the thermal stability of proteins in Herman Watson's laboratory in Bristol. Phosphoglycerate kinase is a monomeric glycolytic enzyme which catalyses the first substrate level phosphorlylation in glycolysis. The enzyme consists of two domains (Figure1) which are believed to undergo a conformational change during catalysis. Two PGK structures had previously been solved at the time of this study, those from horse muscle (Banks *et al.*, 1979) and yeast (Watson *et al.*, 1982).

FIGURE 1. The structure of phosphoglycerate kinase



The structure of 3-phosphoglycerate kinase (PGK). The two domains of PGK are thought to change their orientations depending on the state of ligation, and the presence of ions such as $(SO_4)^{2-}$.

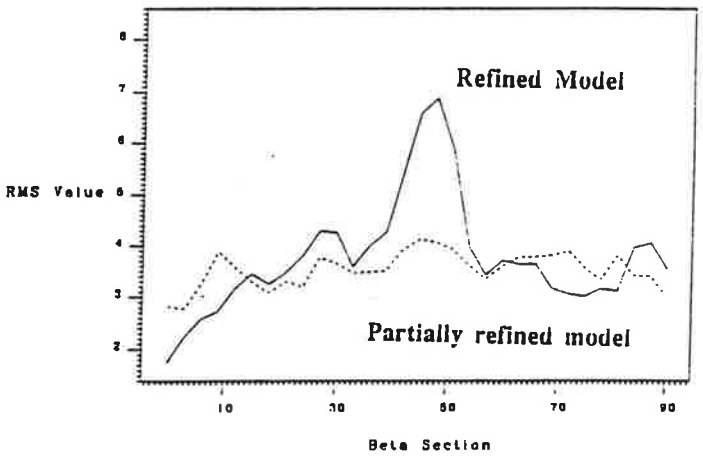
Crystals of *B. stearothermophilus* PGK were grown from polyethylene glycol in the presence of the nucleotide substrate MgATP. The space group was $P2_1$, with cell dimensions a=40.5Å, b=74.0Å, c=68.5Å and β=99.8°. There was one monomer of PGK in the asymmetric unit giving a solvent content of about 50%. The simplicity of this system made it an ideal candidate for structure solution by molecular replacement.

3 (a) THE ROTATION FUNCTION: *choice of model structure*

Initial attempts at molecular replacement with a completely unrefined model, the first build into the MIR map of yeast PGK, failed to give any significant features in the rotation function calculations. As beginners to molecular replacement this came as a surprise to us, for whilst the model obviously contained errors, these were generally small and randomly located across the whole molecule and we were only using data between 8 and 5Å. It is only when we performed trial calculations and found that the errors in an initial MIR built structure are often sufficiently large to prevent that model being used to solve *its own* rotation function that we began to appreciate the

need for a well refined model when trying to solve homologous structures ! Often, an initial MIR build will show a great discrepancy between the low resolution F_{calc} and F_{obs} which can be overcome with just a few cycles of refinement. An example of the effects of a partially refined model can be seen below in Figure 2. Two models were used, one is the refined Yeast PGK model (Watson *et al.*, unpublished) R=19% at 2.0Å and the other is a partially refined model with an R-factor of about 38% at 2.5Å. The two models are not vastly different, the differences are generally quite small but they are sufficient to make a large difference in the rotation fuction calculations.

FIGURE 2.



Rotation function output for *B. stearothermophilus* PGK. Calculations were performed with the MERLOT program (Fitzgerald, 1988). Data from 8-4Å were used together with an outer radius of Patterson integration of 25Å. This radius was chosen so as to maximise the number of self-vectors but limit the inclusion of Patterson origins from adjacent unit cells.

3(b) THE ROTATION FUNCTION: *data quality and completeness*

The rotation function, as calculated, measures the degree of overlap between two Patterson functions, one calculated from the model structure and the other calculated from the observed structure factor amplitudes. Patterson functions, calculated using F^2 terms, are obviously dominated by the stronger F's. The need for complete and accurate observations of the strong terms is clear, but is one that is often overlooked in data collection for molecular replacement calculations. It is all too easy to collect the highest resolution data possible from a crystal whilst neglecting the stronger and lower resolution F's, which often saturate the detector if good high resolution data are being collected. Even with the increased dynamic range of modern detectors, such as image-plates, measuring complete data from a well diffracting crystal may need as many as 3 data collection runs with varying degrees of exposure.

Data on the *B. stearothermophilus* PGK crystals, which diffract to beyond 1.6Å resolution, was collected in three stages. Initially a film dataset to 3.4Å was collected on station PX9.6 at Daresbury. Care was taken to ensure that there were no overloads, *i.e.*, that all the strong diffraction terms were measured. Two datasets were then collected at the EMBL Hamburg outstation using the Hendrix-Lentfer imaging-plate

scanner. A Table showing the completeness of the data, in the range used for molecular replacement calculations, is shown below:

Table 1 Completeness and quality of the *B. stearothermophilus* PGK data

RESOLUTION (Å)	Completeness	R_{merge}	
10.0	89%	0.030	
7.0	96%	0.028	
5.8	95%	0.028	
5.0	95%	0.029	
4.5	95%	0.031	
4.0	96%	0.032	(overall to 4Å, 0.030)

In order to assess the importance of the strong F's, and to simulate the effect of "overloads" (those reflections not measured due to detector saturation) different numbers of the highest intensities were removed from the rotation function calculations. The results are shown in Table 2 below:

Table 2. The effect of omitting strong observed amplitudes from the rotation function calculations

% of the strongest terms removed	Position of CORRECT answer in the peak list
0%	**1st (!)**
8%	not in top 100
6%	not in top 100
3%	20th
1%	6th
0.5% (only 15 reflections absent)	3rd

These results are quite shocking ! Even with a simple system, such as this, the absence of as little as 0.5% of the strongest observed intensities caused the correct answer to start to disappear into a sea of incorrect solutions. What is slightly more perplexing is

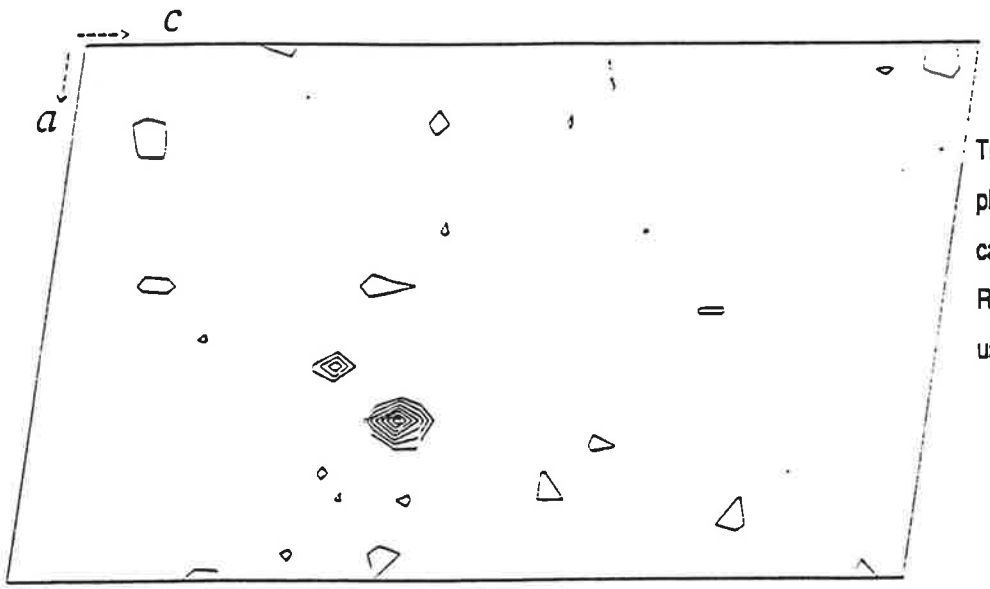
that the rotation function calculations carried out with the incomplete data are not featureless by any means. They contain a selection of *incorrect, but quite convincing* solutions, many of which could lead to a great deal of work running translation functions and the possibility of an incorrect structure.

The importance of these strong observed intensities cannot be over-stressed. Indeed using the, not insubstantial, benefits of viral point-group symmetry Rossmann and colleagues have demonstrated that it is possible to determine the orientation of a virus particle with less than 1% of the observed, *but strong*, diffraction data (Tong and Rossmann, 1990).

4. THE TRANSLATION FUNCTION

The *B. stearothermophilus* PGK crystals, described here, are monoclinic space group $P2_1$ with a single molecule in the asymmetric unit. The undefined origin along the crystallographic *b* axis means that a translation function search is limited to one over the *a-c* plane. With a single molecule in the asymmetric unit a simple R-factor serach was used using the CCP4 program RSEARCH (Figure 3). It gave one answer with an R-factor some 5% lower than at other positions on the search grid.

FIGURE 3. Translation function for *B. stearothermophilus* PGK



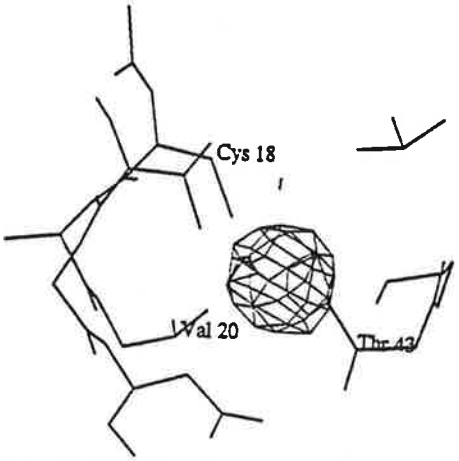
Translation function map in the *a-c* plane for *B. stearothermophilus* PGK calculated using the CCP4 program RSEARCH. Data from 10-5Å were used.

5. CONFIRMATION OF SOLUTION AND INITIAL REFINEMENT

It is useful to have independent means of assesing the validity of a molecular replacement solution. These are discussed in more detail in Eleanor Dodson's article later in these proceedings but those which were applicable to the *B. stearothermophilus* PGK structure determination are discussed here. Upon

obtaining the initial solution, the correctness of the packing of the molecules was assesed both on the graphics using FRODO and computationally using DISTANG (CCP4). There were no particularly bad steric clashes. We were fortunate during the *B. stearothermophilus* PGK structure determination to have a single heavy metal derivative dataset. This was a single-site mercury acetate derivative. A Hg difference Fourier, calculated with phases derived from the molecular replacement solution, gave a single large peak. Not only was this in the same position as the peak found in the difference Patterson synthesis, but superimposition of the model coordinates showed it to be only 4Å away from the sulphur atom of the single cysteine of *B. stearothermophilus* PGK suggesting that the solution was essentially correct (Figure 4). It should be noted, however, that peaks in the difference Fourier can be obtained when the molecular replacement solution is only partially correct, so they should be treated with some caution (Evans, 1985). For instance, in the example discusssed here an translation that was misplaced by 1/4 of a unit cell would give rise to a set of phases half of which would be correct.

FIGURE 4. Hg difference Fourier calculated with phases from the molecular replacement model.

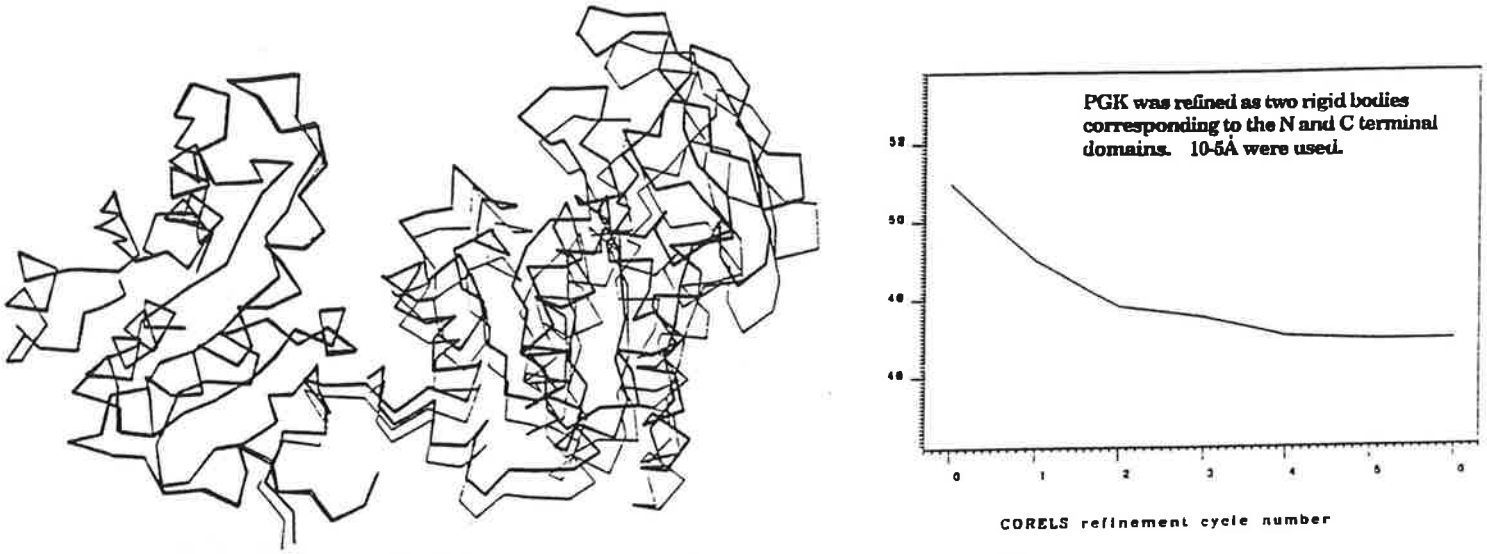


Section of the PGK Hg difference Fourier. The density peak shown is 14σ higher than any other feature in the synthesis. The cysteine sulphur to Hg distance is approximately 4.5Å.

Having arrived at a sensible solution to the rotation and translation function problems, what is the most appropriate refinement method to choose ? For a simple two domain protein such as PGK a simple application of rigid body refinement seemed sensible. In addition, there was independent evidence to suggest that the domains would have a changed orientation due to the ligation state of the enzyme. Rigid-body refinement can now be performed by a number of programs (CORELS, XPLOR, TNT, *etc*) but in this study constrained rigid body refinement was carried out using CORELS, written by Joel Sussman (Sussman *et al.*, 1977, and for review see Leslie, 1985). *B. stearothermophilus* PGK was treated as two independent domains linked by a single "peptide" restraint in the centre of the helix linking the two domains. Initial rigid body refinement of the rotation and translation parameters for

the two domains using data between 15 and 7Å proved to be unstable, so data from 10-5Å were used. Seven cycles of refinement gave an R-factor drop of some 4% and gave a movement of the two domains of approximately 4.5° (Figure 5).

FIGURE 5. Rigid body refinement of *B. stearothermophilus* PGK.



Rigid-body movement of the two domains of PGK after refinement with CORELS. The pre-CORELS structure is shown in faint lines and the structure after CORELS in bold.

R-factor fall during the 7 cycles of CORELS rigid-body refinement.

It is our experience in this laboratory that initial rigid-body refinement of the molecular replacement solution is a most sensible way to proceed. Not only does the refinement leads to a better solution but it helps to "filter-out" incorrect molecular replacement solutions (see Derewenda 1985, 1990). It is often the case with protein structure refinement that initial difficulties with conventional refinement turn out to be due to a global change in domain orientation that could, perhaps, have been avoided with some sensible initial rigid-body refinement (see Swift *et al.*, 1991).

5. DISCUSSION

The advantage of molecular replacement is that it should be simple ! What we have tried to show here that it can be simple, but that errors in the model and the data can lead to molecular replacement becoming rather more tricky. If a couple of lessons are to be learned then they would be (1) always start with the best model and (2) measure all the strong intensities. Film and image processing programs such as DENZO and MOSFLM allow for the inclusion of a best estimate of the intensities of "overloads" in the final output file. Whilst one may not wish to use these estimates for refinement

the
lue
ed"
hat

y in
are
iew

oss

vo

of

s

an

nd