DARESBURY LABORATORY

INFORMATION QUARTERLY for PROTEIN CRYSTALLOGRAPHY

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

Number 13 JULY 1984

Contents

Editor:	Pella Machin	Science & Engineering Research Council, Daresbury Laboratory, Daresbury, Warrington WA4 4AD, England
Deputy Editor for Imperial College:	Dr. Alan Wonacott	Imperial College of Science & Technology, The Blackett Laboratory, Prince Consort Road, London SW7 2BZ
Deputy Editor for Birkbeck College:	Dr. David Moss	Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX

Editorial

The articles in this edition give an overview of the current research projects in various UK Protein Crystallography groups, with additional papers on specific computing projects. Our thanks once again go to all those who have contributed to this Newsletter.

Pella Machin

PROTEIN CRYSTALLOGRAPHY AT LEEDS

A.C.T.North, A.J.Geddes, D.Akrigg, F.Körber

E.Eliopoulos, M.Harris and N.Gammage

The binding of DPG analogues to deoxy haemoglobin

In the early 1970's P.Goodford, C.Beddell and their colleagues from the Wellcome Laboratories selected the effector binding site in deoxy haemoglobin as a model drug receptor to test a new rationale in drug design. They developed several compounds based on diphenylethane di-aldehyde (DIPEDAL) which proved to be about as potent as diphosphoglycerate but presumably reacted with the residues in the DPG binding in a different way, i.e. by covalently crosslinking the N-termini of the two β-chains and forming additional salt bridges with the positively charged residues in the effector site[1].

The binding of two of the drugs, a bisulfite (I) and an oxyacetic acid derivative (II) of DIPEDAL has now been investigated crystallographically in a 5.5Å difference study. Although these compounds have no apparent clinical value, related drugs may prove useful for preserving the oxygen binding properties of stored blood and thus enhancing the effectiveness of blood transfusions. The relatively short 'shelf life' of blood donations can probably also be prolonged by the latter compounds.

Crystals of the protein/drug complexes were obtained by co-crystallisation from PEG 6000 and zwitterionic buffer in the presence of a five-fold excess of compound I or II. The crystals grew in space group $P2_12_12$ with a = 97.3, b = 99.2 and c = 65.8Å and were isomorphous with those described by Ward et al.[2] but have slightly different cell dimensions from the crystals used in the high resolution study of T-state haemoglobin[3]

from which the phases were calculated. Diffraction data for the native proteins and the two derivatives were collected with a CAD4 diffractometer on one crystal each.

The two highest features in the difference map for compound I are situated near the β-N-termini at the entrance to the central cavity, indicating that the drug crosslinks the β-chains as predicted. The exact nature of this interaction (i.e. covalent or ionic) cannot be established at the present resolution of 5.5Å, but the N-terminal nitrogens and the side chains of His 2 and Lys 82 can be coordinated to the peaks whilst the participation of His 143 is doubtful. The absence of density in the 'bridge' region between the two features can be explained by the flexibility of the drug about the torsional angles of the ethane bridge, allowing different conformations of the molecule whilst its end groups remain fixed. A survey of all features in the difference map reveals no other probable binding site and thus emphasises the specificity of the drug.

The binding of compound II could be demonstrated less clearly although the DPG binding site is filled with positive electron density. Two modes of binding can be proposed, one as predicted crosslinking the β-N-termini, the second one covalently linking the N-terminus of one β-chain to the side chain of Lys 82 of the other subunit. Again no secondary binding site has been identified.

It is proposed to extend these studies to higher resolution and to examine the binding of other compounds in the series.

Structural Studies of L.casei Dihydrofolate Reductase (apo-enzyme)

Various structures of dihydrofolate reductase (DHFR) complexes with inhibitor molecules and the haloenzyme structure have been reported[4,5,6]. Comparison with the apoenzyme structure should be of interest because the protein seems to possess enough flexibility to undergo a major conformational change upon the binding of either substrate analogues or cofactor.

Small crystals ($V \sim 0.5 \times 10^{-3}$ mm^3) of the apoenzyme had been obtained some time ago and synchrotron data collections on the native and a $Pt(CN)_4$ derivative have been completed. The enzyme crystallises in space group $P6_1$ or $P6_5$ with one molecule per asymmetric unit and unit cell dimensions of $a \equiv b = 53.4$Å and $c = 110.4$Å and gives useful diffraction to about 2.8Å.

Films were scanned and processed with the standard suite of processing using a Scandig 3 scanner interfaced to a PDP11/45 in an offline mode. Difficulties were encountered during the processing of the film data due to high background, small spot size and the slow scanner response at 50μ and 100μ scan increments. An alternative algorithm to obtain the intensities without prior prediction of the spot position and thus enable 25μ scans to be made is currently being developed.

Crystallographic studies of E.coli glutamate dehydrogenase

Large crystals of E.coli glutamate dehydrogenase (GDH) have been obtained by pulsed microdialysis. Although the crystals do not diffract too well on conventional sources, the resolution limit with synchrotron radiation is about 2.1Å. Approximately 50° of rotation camera data to 3Å resolution have been collected so far. The crystals used for the synchrotron data collection were orthorhombic with cell dimensions

of a = 155Å, b = 207Å, c = 100Å. This is different from the parameters published earlier which were determined from precession photographs on a conventional source[7]. Changes in the axis ratios have been observed which might indicate motion in the (a,b) - plane. This phenomenon is currently under investigation.

β Lactoglobulin

Bovine β-lactoglobulin is a dimeric protein with a molecular weight of 2x18400. In solution it undergoes a pH dependent transition between two alternative structures. The structures of two different crystal forms crystallised on either side of the transition at pH 7.5 (lattice Y) and pH 6.5 (lattice X) respectively, are currently under investigation, lattice Y by Dr L Sawyer in Edinburgh, lattice X by ourselves.

Lattice X (space group P1) has unit cell dimensions 38.1, 49.7, 56.6Å and 122.2, 97.5, 104° with one dimer with pseudo two-fold symmetry per asymmetric unit.

Rotation camera data for the native and two mercurial derivatives (PCMBS and DCMNP) have been collected to a resolution of 2.5Å.

Electron density maps at a resolution of 2.8Å have been generated using phases (I) from isomorphous replacement and (II) from fitting the partially determined lattice Y structure to satisfy the pseudo two-fold symmetry, the common heavy atom positions and a 6Å isomorphous map. These maps are currently being interpreted. Meanwhile, we have collaborated with Dr Sawyer in building a model of the lattice Y structure at 2.8Å resolution, using our Evans and Sutherland PS300 system.

Insulin

Computer graphic modelling has been performed on synthetic insulin isomers with unnatural disulphide bridge pairing, and the observed properties have been rationalised in terms of observed conformational restrains.

Some proposed synthetic relaxin C-peptides have also been modelled, in order to select the most appropriate sequence to be used in the recombinant production of the prohormone.

Attempts have been made to purify prorelaxin of the natural sequence produced by recombinant techniques in E.coli.

A Hendrickson-Konnert refinement of cubic insulin to 1.7Å is currently being performed. In this form of insulin, the asymmetric unit is a single insulin molecule, in contrast to the well-known rhombohedral forms in which the asymmetric unit is a dimer in which there are conformational differences between the two monomers. Surprisingly, statistical disorder is present in the cubic crystal structure and is consistent with adjacent molecules related by 2-fold rotation axes being non-equivalent at their interface.

Development of Computer Graphics facilities

In order to facilitate the representation of 3-dimensional information on molecular structure, we have been developing a stereo viewer based on spectacles with 'lenses' consisting of liquid crystal shutters synchronized to the display of alternate images on the screen. Prototypes of these viewers are about to undergo evaluation in several laboratories.

References

1. Beddell, C.R. et al., Br.J.Pharmac. 57 (1976) 201.
2. Ward, K.B. et al., J.Mol.Biol. 98 (1975) 161.
3. Brzozowski, A., et al., Nature 307 (1984) 74.
4. Matthews, D.A., et al., Science 197 (1977) 452.
5. Volz, K.W., et al., J.Biol.Chem. 257 (1982) 2528.
6. Stammers, D.K. et al., in 'Chemistry and Biology of Pteridines, Pteridines and Folic Acid Derivatives, J.A.Blair (ed)., de Gruyter, Berlin (1983) 567.
7. North, A.C.T. et al., in 'Synchrotron Radiation, appendix to the Daresbury Annual Report 1982/83', reference 53.

Bristol - Work in Progress

Hilary Muirhead, Stephen Marshall and Deborah Clayden.

Pyruvate Kinase

The amino-acid sequence has been fitted to the 2.6Å electron density map of the muscle enzyme and the structure has been refined to give an R-factor of 0.28. The overall homology between the muscle and yeast sequences is about 42% and side-chains close to the active site are highly conserved. We are building a model of the active site and investigating sub-unit and domain interactions.

Glucose 6-Phosphate Isomerase (PGI)

We have collected 2.1Å data at Daresbury for native PGI, PGI + PCMB, PGI + Selenate and PGI cocrystallised with substrate (Glucose-6-Phosphate). We are in the middle of processing the data.

-----oOo-----

Herman Watson

The following structures are at the refinement stage:

Yeast Phosphoglycerate Kinase
Yeast Phosphoglycerate Mutase
Human Glyceraldehyde Phosphate Dehydrogenase

A two derivative map at 3.0Å has been calculated for human muscle Aldolase though it has not yet reached an interpretable quality.

COREL Refinement of apo and 1 NAD per tetramer

B. stearothermpohilus GAPDH

Dr Andrew Leslie (Imperial College, London)

The constrained-restrained parameter structure-factor least-squares refinement program CORELS (Sussman et al., 1977) has been used with considerable success in the refinement of two different crystal forms of GAPDH. The apo-enzyme crystallises in space group $P2_1$ with the entire tetramer (MW 145000) in the asymmetric unit. A 6Å resolution m.i.r. map suggested that the subunit structure of the apo-enzyme is related to that of the holo-enzyme by a five degree rotation of the coenzyme-binding domain (172 residues) with respect to the catalytic domain (162 residues). Both structures possess 222 molecular symmetry. Using the refined holo-enzyme coordinates as a starting model, the apo-enzyme structure was refined using CORELS. Initially each subunit was divided into three rigid groups corresponding to the two domains and the C-terminal helix, and refined against 6Å resolution X-ray data. Further refinement was then carried out at 4Å resolution, with rigid groups corresponding to alpha helices or strands of beta sheet. A "shift-averaging" procedure was developed to impose 222 molecular symmetry (Leslie, 1984) which improved both the rate of convergence and the stability of the refinement. The final coordinates were used to calculate structure factors to 3Å resolution (R=30.2%), and an $(F_o-F_c),\alpha_c$ map clearly indicated the deficiencies in the rigid-body model. These were suprisingly few in number and were limited to polypeptide loops linking elements of secondary structure. The quality of the difference map, which was averaged over the four subunits of the tetramer, was good enough to permit unambiguous identification of forty water molecules per subunit, even at 3Å resolution. Further refinemnt will be carried out using the Hendrickson Konnert program. The conformational change in going from the holo-enzyme to the apo-enzyme structure corresponds to an approximate rigid body rotation of the coenzyme-binding domain by 4.5^{o}, with an r.m.s. shift in alpha-carbon positions of 1.5Å.

The 1 NAD per tetramer enzyme crystallises in space group $P2_12_12$ and again the asymmetric unit contains the entire tetramer. The apo-enzyme coordinates were used as a starting model, and refined using CORELS at 6Å and then 4Å resolution. The final coordinates gave an R factor of 30.4% for data to 3Å resolution. The positional parameters were then refined for eleven cycles using the Hendrickson Konnert program. This refinement eliminated all the distortions in stereochemistry introduced by the CORELS refinement, and reduced the R factor to 23.8% at 3.0Å. Manual rebuilding is now required. The 1 NAD enzyme structure can be described as a tetramer in which three of the subunits adopt the apo or "open" conformation while the fourth subunit, containing the bound NAD, adopts the holo or "closed" conformation.

The structural results have demonstrated the essentially rigid nature of the secondary structural elements in

GGMADP

different forms of this enzyme, which undoubtedly contributed to the success of the rigid body description used in the structure refinement. The structural results will be published in detail elsewhere (Leslie & Wonacott, 1984).

References

Sussman, J.L., Holbrook, S.R., Church, G.M. & Kim, S.H. (1977) Acta Cryst. A33, 800-804.
Leslie, A.G.W. (1984) Acta Cryst in press
Leslie, A.G.W & Wonacott, A.J. (1984) J. Mol. Biol. in press

SYNCHROTRON RADIATION STUDIES OF VERY SMALL CRYSTALS

Janet Hails, Marjorie Harding, John Helliwell
Liverpool University/Daresbury

The project aims to make use of the high intensity of the SRS and the SRS protein crystallography equipment to record single crystal diffraction data for crystals of size much smaller than could be studied with conventional Xray sources. We planned to start with crystals of medium molecular weight compounds (di- and tri-saccharides, zeolites, which are simple to handle since they are stable in air and can be mounted on glass fibres), and work upwards to oligo-nucleotides, oligo-saccharides and related complexes.

One good run on the Wiggler Station has yielded 15 packs of oscillation camera data, representing a total of 75° of data, for a disaccharide (from Bill Mackie and Tony North, Leeds); the crystal dimensions were 25x5x150 μm, volume 19×10^3 μm^3. The first impression of the photographs is that they are blank, but closer inspection reveals minute spots. From this set of photographs the unit cell, previously unknown, has been determined. Modifications to most stages of the film scanning and protein data processing programs are in progress to cope with these minute spots - but in the meantime we have used quick visual estimates of 250 reflection intensities to attempt a structure solution. For this crystal we used exposure times of 400 sec/deg for λ = 1.10Å and a resolution of 1.18Å equivalent to 20 hours for a full data set of 180° (at 1.80 GeV, 170 mA). The diffracted intensity is proportional to the crystal volume and to its scattering power, which can be estimated from $\Sigma f^2/V_{cell}$, as well as to λ^2 and the incident beam intensity. Somewhat longer exposure times would have been desirable to record more of the weaker reflections, but this nevertheless affords a basis for estimating times for other compounds and crystals, and the practicability of recording these data on the SRS.

Another approach to the investigation of very small crystals which has just recently become possible on the new Daresbury wiggler protein crystallography station is through Laue photographs. The intensity of radiation in the white beam is enormous; an exposure of 15-30 sec was sufficient to obtain a good Laue photograph of a crystal of volume 25×10^3 μm^3 in single bunch mode of the machine at 20 mA current. A continous sweep of wavelengths of ~ 0.2 - 4Å are available, thus many orders of reflection are recordable for one orientation of the crystal. We are exploring the possibilities (and computational problems!) of using these for measuring intensities.

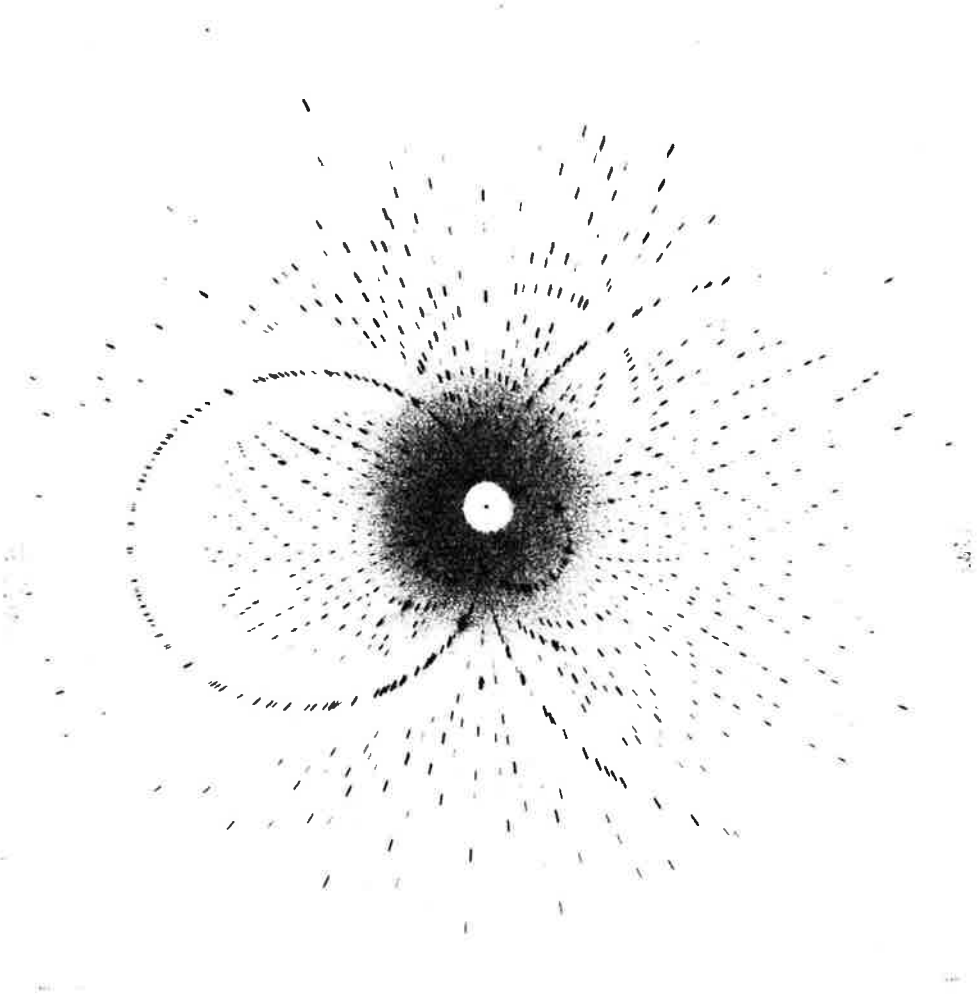


Figure: Laue diffraction pattern recorded on the SRS wiggler protein crystallography workstation from a single crystal of size $40 \times 40 \times 100\ \mu m^3$ of an iridium cluster complex, space group C2/c $a = 34.5$ Å, $b = 17.6$ Å, $c = 23.0$ Å, $\beta = 124^{\circ}$.
SRS 1.8 GeV, 20 mA (single bunch mode) wiggler 4.5 Tesla.
Incident λ range $\sim$ 0.2 - 4.0 Å continuous, exposure time $\approx$ 1 minute, crystal rotated 1°, CEA reflex 25.

Ferritin Around at Daresbury

G.C.Ford, P.M.Harrison, D.W.Rice, J.M.Smith, J.L.White

The iron storage molecule Ferritin has a hollow, spherical protein shell consisting of 24 subunits arranged with 432 symmetry, and an iron core of up to 4500 Fe atoms in an inorganic hydrous ferric oxide-phosphate complex. Ferritin forms octahedral crystals from 2% solutions of Cadmium and some other metals. The structure of apoferritin (the protein shell with the iron core removed) has been refined to 2.6Å resolution and this shows two different types of channel through the protein shell which may provide the entrance ports for the iron. These 4Å diameter channels are around the 3- and 4-fold symmetry axes and are hydrophilic and hydrophobic respectively. The 3-fold channel contains two bound cadmiums from the crystallisation medium coordinated to 3 Asps and 3 Glus. Along the dimer interface on the inside surface of the molecule is a groove containing many charged side chains, most of which are disordered in our map.

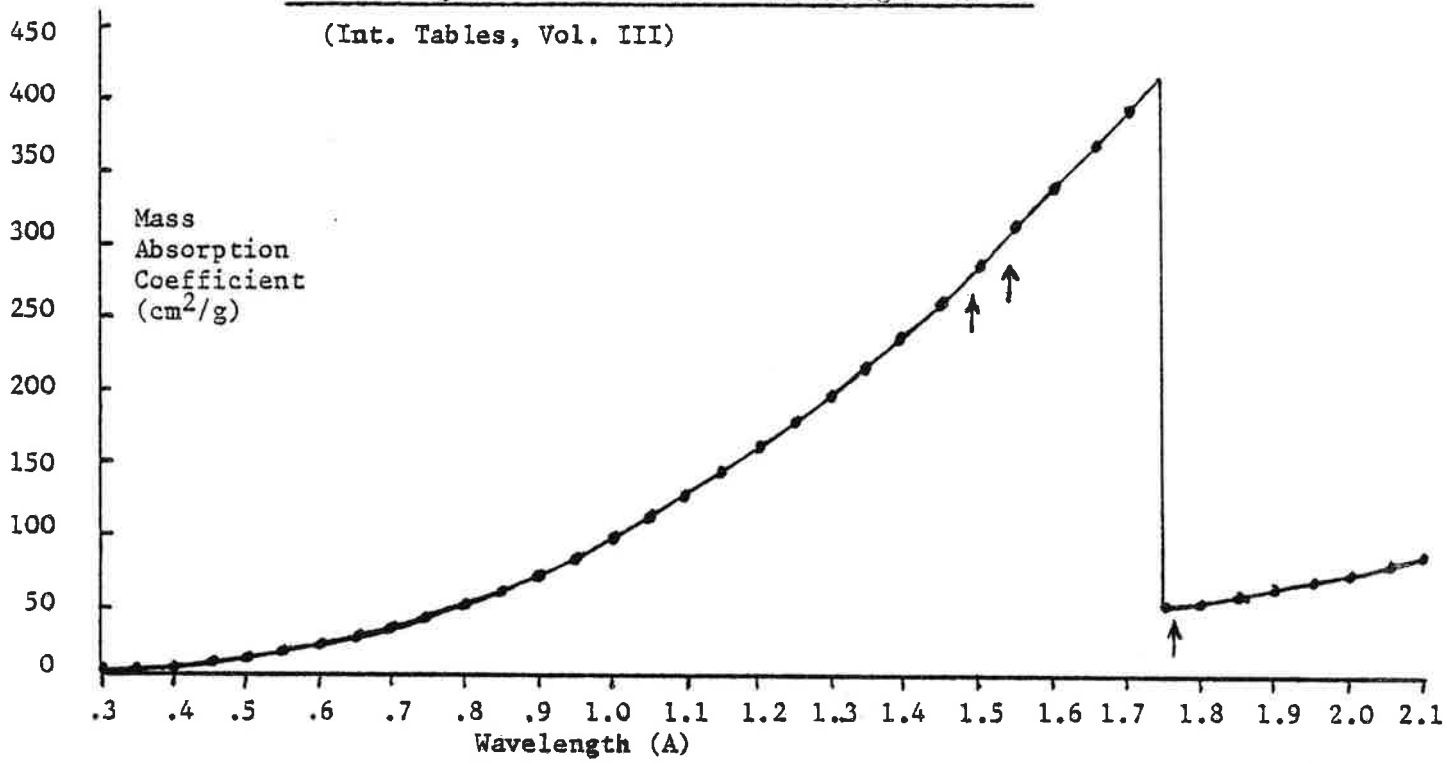
We wish to know whether the entrance for the iron atoms to the interior cavity is via the 3- or 4-fold axis channels, and where on the inside surface the iron core nucleates (the dimer interface being the most attractive possibility). To answer these questions we must examine the structures of holoferritin and complexes of apoferritin with metal ions such as Zinc and Terbium which are known to inhibit iron uptake. In the past, efforts to look at holoferritin have been frustrated by the high absorption and fluorescence of X-rays by the iron core. As can be seen in figure 1, the mass absorption coefficient (μ/ρ) for iron at 1.5Å is 288, and at 1.488Å wavelength on the SRS ferritin crystals died very quickly, either during setting or within the first 2 or 3 oscillation films, and the background scattering was enormous. This is in contrast to apoferritin crystals, which normally live for 24 films and have very low background.

During the recent period when the SRS was having difficulties (!) and running at low power (where there is only reasonable flux at longer wavelengths) we collected ferritin data at 1.76Å which was greatly improved. Unfortunately, the films cannot be directly compared as the crystal used in the 1.76Å experiment was both larger and drier than those previously tried at 1.488Å and thus should have been somewhat better anyway. However, even allowing for this, the improvement is striking. At the longer wavelength above the iron edge we got 12 good quality data films with a dramatically lower background.

At 1.75 (μ/ρ) for Fe is only 49, a factor of 6 lower than at 1.5Å, so the heating due to absorption of X-rays in the iron cores, especially at the high fluxes on the SRS, should be greatly reduced as well as the fluorescence. The experiment to determine the true difference in lifetime due to absorption effects by working with identical crystals at the two wavelengths will shortly be performed on the wiggler PX 9.6 station.

Mass Absorption Coefficient vs Wavelength for Fe

(Int. Tables, Vol. III)



DEVELOPMENT OF A VERSION OF THE HENDRICKSON-KONNERT PROGRAM FOR USE BY CCP4

John W Campbell (Daresbury Laboratory)

Introduction

Extensive use has been made of the Hendrickson-Konnert restrained least squares refinement program for protein structures by the UK Protein Crystallography groups. The original versions of the program obtained from Dr W Hendrickson have undergone many modifications and the versions in the UK have proliferated. The CCP4 Working Group II had, for some time, identified a need to prepare a version of the prograsm containing all the required features which would be compatible with the Fast Fourier Transform least squares refinement program SF. Though a certain amount of effort had already been spent in attempting to carry out this task it was apparent that more resources were required to complete it in a reasonable time and it was proposed that CCP4 should fund such a project if a suitable person could be found.

Background to the Existing Versions of the Hendrickson-Konnert Program

The Hendrickson-Konnert refinement procedure involves two programs, PROTIN which analyses the geometry of the protein structure and which prepares an input file for the program PROLSQ which carries out the refinement. The existing versions of the program at Daresbury were based on two versions of the program supplied by Dr W Hendrickson. The first of these, PRSQB, was used for larger structures and was space group specific. The second, PRSQN, was suitable for smaller structures, was space group general and allowed for the refinement of occupancies and a form of anisotropic temperature factors. Extensive modifications were made to PRSQB by Dr W Pulford (University of Oxford) to allow for the refinement of large structures by storing part of the least squares matrix on disk. Both versions were modified to run on the Cray computer and to handle the standard CCP4 file structures. A version of PRSQB, compatible with the Fast Fourier Transform refinement program was prepared by Eleanor Dodson (University of York) and it was this version that was taken as the starting point for the project. It was intended that the new version should be in ANSI standard FORTRAN 77 and should meet the standards for the CCP4 program suite. Such a version should be easily implemented on a wide variety of computers including the Cray-1.

Modifications made by Dr H Tsukada

We were fortunate that Dr H Tsukada from Imperial College, London was willing to take on the project in the six weeks he had available between a visit to Japan in May and his return to Japan at the beginning of July. During a busy six weeks he was able to incorporate the following features into PROLSQ.

(a) Re-written space group specific routines making best use of sorted data and catering for an extended range of space groups.

(b) A space group general (P1) routine making use of symmetry operation specified in the control data.

(c) Refinement of variable occupancy factors.

(d) Provision for specifying neutron form factors.

(e) Parameterisation of arrays within the program to allow for easy expansion or contraction of the program capacity.

(f) Completion of the conversion of the program to ANSI standard FORTRAN 77 and related CCP4 standards.

(g) Printing of new diagnostics for the conjugate-gradient solution.

(h) Provision for applying pre-selected dumping factors instead of those selected by the R-factor test.

Further Work at Daresbury

The six weeks for which Dr Tsukada was available was less than the eight weeks originally proposed for the project and some further work has been carried out including the following:

(a) Documentation of PROLSQ and PROTIN to CCP4 standards (in progress)

(b) Preparation of a version of PROTIN meeting CCP4 standards.

(c) Provision for overriding the default form factors in PROLSQ with those provided by the user for other atom types.

(d) Provision for overriding the default atom types for the Van-der-Waals contact distances used in PROTIN.

Future Development

Two items which have not yet been tackled and which may require fairly extensive effort are:

(a) Provision for coping with alternate conformations.

(b) Provision for dealing with inter-molecular contacts.

CURRENT PROJECTS IN THE LABORATORY OF MOLECULAR BIOPHYSICS, OXFORD.

Below is a brief list of projects currently under study in our laboratory. A few model building studies in progress are also listed for interest. The molecular weights or amino acid numbers refer to the amount of protein in the asymmetric unit. Film data reduction is performed on our PDP 11/70; crystallographic calculations are performed on the ICL 2988; a few programs e.g. CORELS are run on the University VAX 780 and all graphical molecular modelling is carried out on the PS2.

1. Phosphorylase b

Louise Johnson, Paul McLaughlin, Dave Stuart, Janos Hajdu, Ravi Acharya

P43212 128.5 116.4
Amino acids:841

i) Refinement of 2A native on the CRAY using Hendickson-Konnert.
ii) Post refinement and absorption corrections for oscillation data.
iii)Low temperature studies on native and substrate complexes.
iv) Activity of phosphorylase in the crystal.
v) Probes for catalytic mechanism.
vi) -100 C with glycerol refinement.
vii)Heptulose-2-phosphate binding; long and short soaks.
viii)Heptenitol binding.
ix) Mn binding.
x) Oligosaccharide binding (2.5 A)

All binding studies to 3.0 A unless otherwise stated.

xi) Model building of E.Coli sequence into rabbit structure.
xii)P21 crystal form from ammonium sulphate, potentially in the "R" state: crystallisation and data collection.

2. 6-phosphogluconate dehydrogenase (PGDH)

Margaret Adams, Sheila Gover, Richard Pickersgill, Katherine Pelly, Grant Ellis

C2221 72.72 148.15 102.91
Amino acids: 466

i) Native refinement to 2.8A. R=0.38.
ii) V-cassette data analysis to 2.0A.
iii)Ternary complex of PGDH with phosphogluconate and NADPH. 2.7A data collection.
iv) Ternary complex as above but with NADP bound. 3.0 A data collection.

3. Glucose-6-phosphate dehydrogenase

Margaret Adams, Mohan Bhadbhade

P3121 105.8 225.1
Amino acids: 450

i) Crystals grown, heavy atom searches under way.

4. Ribulose-1,5-bisphosphate carboxylase
Margaret Adams, Richard Pickersgill, David Phillips

C2221 158.5 158.6 203.4
One large subunit 55,000 and one small subunit, 15,000.

i) Diffraction to 1.6A, early stages of stabilising crystallisation.

5. Fc fragment of rabbit IgG

Brian Sutton, David Phillips, Peter Artymiuk

P21 68.85 72.50 60.40 105.1
Amino acids: 440

i) Refinement in progress on CRAY, including carbohydrate to 2.7A.

6. Fv fragment of mouse myeloma protein M315 (IgA)

Sue Collett, Brian Sutton, David Phillips

C2 59.6 56.6 137.9 99.6
amino acids: 2 x 222

i) Solution of structure by real space molecular replacement of Fab fragments into map using graphics. Subsequent refinement using CORELS. Phases to 3.0A.
ii) Haptan binding studies to Fv. 3.0A.

7. Model building studies on monoclonal antibodies.

Rissa de la Paz, Brian Sutton, David Phillips

i) Seven monoclonal antibodies have been raised by Mike Darseley and Rob Ryan (under Tony Rees in our laboratory), to the antigenic loop region of lysozyme: residues 60 to 80. Two have been fully sequenced and model building studies are under way using the graphics to model the antibody-antigen interaction.

ii) Crsytallisation trials of Fv + loop molecules for x-ray studies under way.

8. Human lysozyme

Peter Artymiuk, Colin Blake

P212121 57.13 60.99 32.89
Amino acids: 130

i) Refinement to 1.5A on CRAY, interpretation of solvent structure.

9. Hen egg white lysozyme

Helen Handoll, Peter Artymiuk, David Phillips

P43212 79.01 79.01 38.25
Amino acids: 129

i) Refinement to 1.6A and solvent structure.
ii) HEWL plus tri-NAG substrate refinement to 2.0A.
iii)High temperature HEWL: P212121 with cell dimensions 56.3 73.8 30.4. Refinement to 1.5A. In collaboration with Jon Berthou ,Paris.

10. Tortoise egg white lysozyme

Colin Blake, Simon Evans, Oliver Galley

i) Neutron diffraction studies at Grenoble in collaboration with Sax Mason. Data set to 2.8A collected, joint x-ray and neutron refinement.
ii)Binding study of NAG-NAM-NAG-NAM at 2.5A.

11. Baboon milk lysozyme

Helen Handoll, Peter Artymiuk, David Phillips

P212121 56.80 62.35 33.25
Amino acids: 130

i) Refinement at 3.0A on CRAY, with model building.

12. Triose phosphate isomerase

Peter Artymiuk, David Phillips

P212121 106.01 74.76 61.74
Amino acids: 2 x 247

i) Refinement to 2.5A.

13. Manganese superoxide dismutase

Colin Blake, Mike Parker

P21212 72.4 111.1 51.1
Amino acids: 2 x 203

i) Structure solution by isomorphous replacement and molecular replacement.

14. Prothrombin fragment 1

Karl Harlos, Colin Blake

P41212 77.7 84.9
Amino acids: 156

i) Finding stable crystals
ii) Native data to 3.5A collected

15. Phosphoglycerate kinase (PGK)

Colin Blake, Simon Evans

P21 50.8 106.9 36.3 98.6
Molecular weight: 45,000

i) Tertiary complex: crystallisation with ATP.

16. Beta-lactamases

David Phillips, B. Samraoui, Rosemary Todd, Brian Sutton

b-lactamase I: C2 143.9 35.8 52.7 97.0
Molecular weight: 28000

i) Improvement of initial map at 2.8A.
ii)Substrate analogue binding of 6-B-bromopenicillanic acid. Data collected.

b-lactamase II: P43(1)212 74.5 154
Molecular weight: 28000

i) Early stages.

17. Prealbumin

Colin Blake, Rissa de la Paz

P21212 43.5 85.7 66.0
Molecular weight: 54,000

i) Structures of 4 complexes with thyroid hormone analogues to 1.8A.

18. Seal myoglobin

Helen Scouloudi

i) Refinement to 2.5A.

19. Alpha lactalbumin

David Phillips, Dave Stuart, Steve Wilkins

P21212 35.5 69.1 46.1
Molecular weight: 14400

i) Refinement using CORELS first at 6.0A then extending to 2.7A.
ii) Hendrickson-Konnert refinement to 1.7A.
iii)Phase extension using maximum entropy techniques.

In addition to the above crstallographic studies, the following solution scattering studies are in progress under Andrew Miller who is now at Edinburgh. Most of his students will still be based in Oxford for the next few months before moving north.

Andrew Miller, Rob Alecio, Yvonne Jones, Jeremy Bradshaw, Milton Stubbs

20. Collagen

i) Improvement of the model against x-ray fibre data, in collaboration with Bruce Fraser, Australia.
ii) Computer modellling of fibril packing.

21. Virus Structures

i) Iridescent virus 29
ii) Tick-bone encephalitus virus
iii)Influenza virus
Small angle and neutron scattering experiments.

22. Synthetic lipids in multilayers

For completeness we briefly mention the cell biology studies in progress in th laboratory.

Tony Rees, Mary Gregoriou, Rob Moore, Sharon Smith, Neil Simister, Mike Darsley, Rob Ryan

23. Epidermal growth factor receptor.

i) Raising of monoclonal antibodies to EGF receptor.
ii) Purification of receptor.
iii)Physical studies: attempts to produce 2-D arrays for image reconstruction.
iv) Mobility measurements.
v) 2-D mapping of receptor.

24. Studies on growth factor receptors in differentiation.

25. Monoclonal antibodies raised against lysozyme (see 7 above).

26. Fc receptors and immunoglobulin transport.

Theoretical studies within the laboratory include:

27. Electrostatic interactions in proteins.

Neil Rogers, Mike Sternberg (now at Birkbeck), David Phillips

i) Analysis of different dielectric models.
ii) The role of the alpha helix dipole.
iii) Calculation of electrostatic effects in cytochrome C, in collaboration with Geoff Moore, Inorganic Chemistry, Oxford.

28. Maximum Entropy methods

Steven Wilkins, Australia (visitor until September 1984).

i) The application of maximum entropy techniques to phase extension, using the first constraint of a subset of known isomorphous phases.

29. Molecular dynamics

Peter Artymiuk, Janet Stockwell, David Phillips

i) Analysis of a 100 ps simulation carried out on the CRAY last year in collaboration with Carol Post and Mertin Karplus, Harvard, USA.
ii) Comparison with x-ray thermal factors and nmr data.

30. Graphics

Garry Taylor

i) FITZ extended for use in real space molecular replacement: Fv structure and beta-lactamase.
ii) Model building IgM pentameric structure with Brian Sutton.
iii)JIG a program for visualising molecular dynamics frames.
iv) Investigation of cytochrome C2 structure and comparison with nmr data in collaboration with Geoff Moore, Inorganic Chemistry Oxford.
v) Preliminary studies into site directed mutagenesis on lysozyme to produce cellulase activity, in collaboration with Gordon Lowe, Dyson Perrins Laboratory, Oxford.

We will be moving into a new laboratory in December where a purpose built computer suite has been planned. We will be installing a VAX 11/750 system with an Evans & Sutherland PS330 graphics system. The move will also allow us to become integrated into the University ring network, and into the JANET X.25 network. There we will hope to expand our current collaborations with other laboratories within Oxford. The VAX will become a focus for many biophysical techniques including image reconstructions from em photographs (with David Shotton and Helen Saibil, Zoology Department), site directed mutagenesis programs using gene and protein data base searches and the interactive

molecular graphics (Tony Rees, Gordon Lowe and Stephen Waley), research into the role of dynamics in cellular events (Chris Dobson, Inorganic Chemistry, Oxford) by analysis of nmr data with structural information.

The ICL 2988 will continue to perform our normal protein crystallographic computing. The VAX will be networked to the 2988 making file transfers much easier.

PROTEIN SEQUENCE ANALYSIS

Pella Machin (SERC, Daresbury Laboratory)

A copy of the Protein Sequence Database of the Protein Identification Resource (PIR, Georgetown University Medical Centre, Washington, USA) is available on a VAX computer at Daresbury. The software distributed with the system includes the following programs:

SEARCH — to compare a user specified segment with all segments of the same length of every sequence in the database.

ALIGN — to determine a best alignment of two protein (or two nucleic acid) sequences by computing the maximum match score using a version of the Needleman and Wunsch algorithm.

RELATE — is designed to detect unusual similarity between sequences by comparing all possible segments of a given length from one sequence with all segments of the same length from the second sequence.

PRPLOT — is a generalised version of the algorithm by Hopp et al PNAS 78 3824 (1981). Amino acid scoring values are averages over a specified sequence length are plotted versus sequence position.

DOTMATRIX — uses a RELATE type algorithm to produce a DOTMATRIX comparison of two sequences.

PSQ — is an interactive program for locating and examining an entry in the database.

This protein sequence database contains a total of 2676 sequences as shown overleaf.

Group	Number of Sequences	Number of Residues
Total	2676	526,466
Eukaryotes	1721	272,609
Mammals	947	165,416
Plants	169	22,143
Fungi	95	23,864
Prokaryotes	505	99,280
Animal Viruses	230	100,438
Plant Viruses	34	14,643
Bacteriophages	186	39,496

The PIR Nucleic Acid Sequence Database and associated retrieval system (including reformatted GenBank (TM) and EMBL databases) will also be available in the near future.

Anyone wishing to use this system should contact me for further information.

MORIA: A Space Filling, Molecular Modelling Program

by

S.Zurek, G.R.Mant and E.Pantos

SCIENCE AND ENGINEERING RESEARCH COUNCIL
DARESBURY LABORATORY

MORIA is a program that draws one-point perspective, space-filling, models of molecules or other structures that can be represented by a collection of spheres given basic co-ordinate information. It utilises the depth buffer method which enables it to use simple algorithms for drawing spheres.

There are two program versions. The first is written in "portable" Fortran and runs both on the AS/7000 and the VAX. The main algorithm uses techniques published in the literature. A single light source at a user defined point is used to calculate highlighting.

The depth buffer approach uses two arrays , one is a frame buffer which holds a copy of the picture that is displayed on a graphics device, and the depth buffer which holds the depth values in world space. It allows each sphere to be drawn one at a time and this leaves scope for optimisation since incremental methods can be used to calculate shading or depth across a sphere. Other algorithms, e.g., the scan-line algorithm, draw the picture in such a way that spheres cannot be drawn singly and some opportunities for optimization are lost.

The second version of our program makes certain assumptions about the nature of molecules it draws. The algorithm is strictly correct only for parallel projections, but distortions should not be noticeable unless views very close to the molecule are required.

It first determines the approximate radius of each projected sphere. This radius is then used as an index to tables which are either computed at the start of a session or loaded from disk. They consist of pairs, a frame buffer and a depth buffer for each sphere size. Tables for spheres of all sizes, from a radius of 1 pixel upwards to some limit, are created. With these tables the algorithm becomes very simple. No CPU intensive depth calculations are required and in fact the only depth related calculation is the addition of the z co-ordinate of the sphere's centre (distance from viewing point) to each depth value.

The table setup routines are written in C and the initialization and drawing ones in VAX assembler. CPU figures for different image sizes for the example in Fig.1 are given below.

Overheads for coordinate input are not included.

CPU time (seconds) for the example of Fig.1b (1443 spheres)

Version	128x128	256x256	512x512
AS/7000	1	3.5	15
VAX	20	46.5	169
VAX Fast	2.3	3.5	12

The size requirements for the storage of tables for various maximum values of radius are given below in Kbytes. The size includes all the pairs from 1 up to and including the maximum radius.

Max.Radius:	10	20	30	40	50	60
Size :	4.5	33.5	110.5	255.5	503	865

As the maximum radius increases the law of diminishing returns applies and the algorithm is limited in its capacity to draw large spheres. Fortunately for most structures (100 or more spheres) the maximum radius rarely exceeds 20 in normal views, and will only be larger if the structure is magnified.

Further details on both program versions will be given elsewhere. Work is now in progress to enhance the user interface to the program. At present, it can read coordinate values in free format (x, y, z orthogonal coordinates, radius and atom type index), or datasets from the Brookhaven Protein Structure Database. It will soon be possible to read datasets from the Cambridge Small Molecule Database, too. Viewing parameters may be altered interactively for subsequent iterations. Frames may be output either on a SIGMA ARGS 7000 display (VAX versions only) or in a file for subsequent display. The AS/7000 version may be run in the batch to produce large numbers of frames on mag tape, which may then be taken away for viewing at a different installation.

ARGS pictures may be printed on a video printer (Polaroid prints or 35 mm film). We soon hope to provide the equipment for video recording and editing of sequences of frames.

A number of support programs accompany MORIA.

MORIA on the VAX is closely linked to the ARGS, as the device forms an essential part of the interactive facility. It produces pictures in which each pixel is an integer in the range 0-255. Each atom type is assigned a smaller range , (32 levels per atom type for up to 8 different atoms in the molecule). These values can then be directly interpreted as 256 logical colours by the ARGS.

PALETTE, is a program that can be used to setup the ARGS Look-up Table so that ranges of intensity values describing a particular atom type are rendered in shades of a chosen colour.

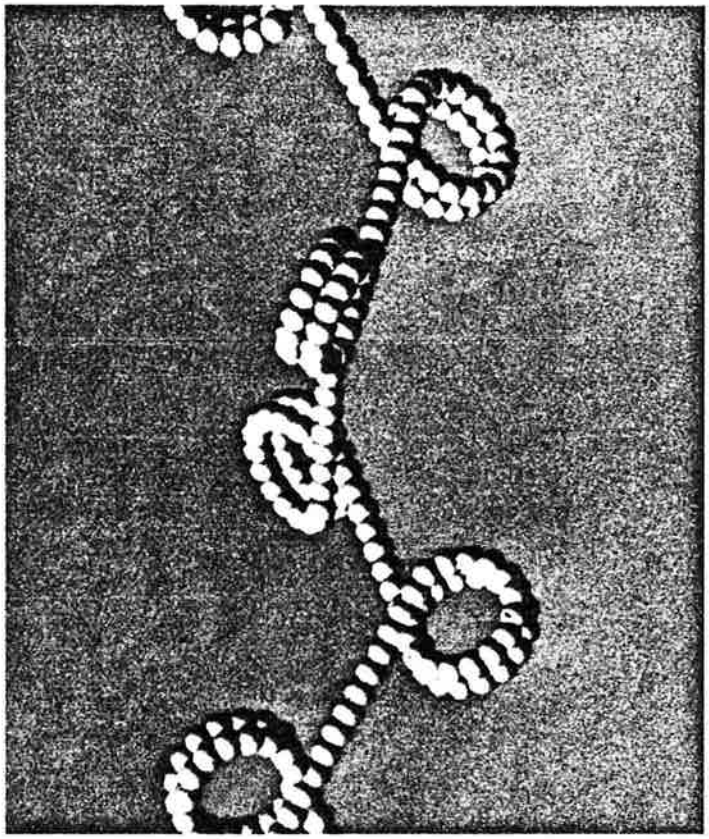
FLASHIT takes advantage of the hardware in the ARGS so that three pictures can be displayed in instant succession, thus giving the impression of motion and depth.

MORICOM enables the creation of files suitable as command input for MORIA. The user can set up a sequence of instructions to generate several views of a molecular system for automatic display and/or output to file.

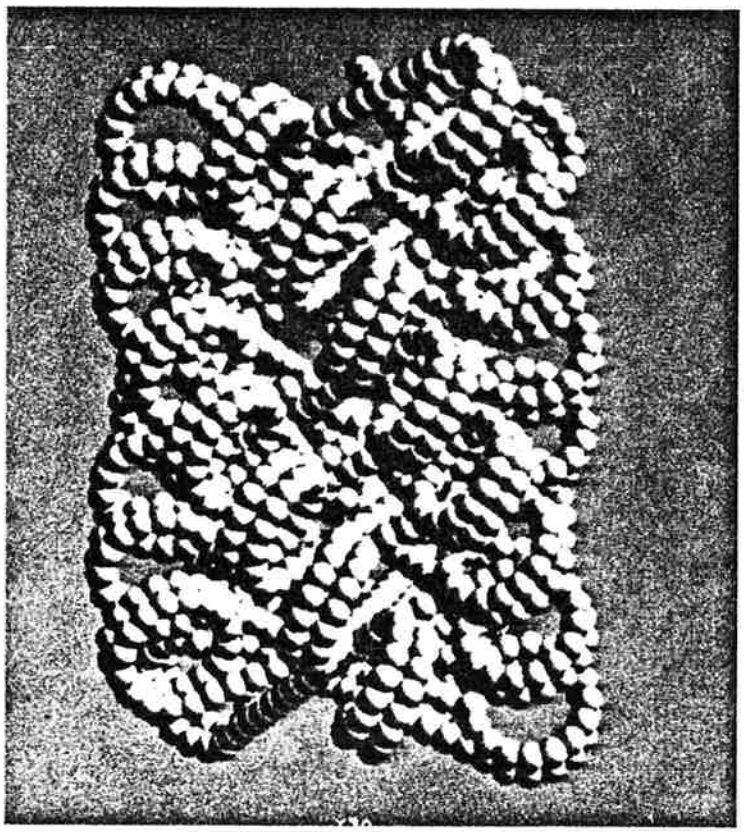
DISPIC may be used to display on the ARGS a precalculated sequence of frames that are stored in a single sequential file.

CREATE3D can be used to reformat individual image files into STARLINK BDF 3D files which may then be accessed either sequentially or at random using the STARLINK software utilities.

Fig.1 One of the many models for Chromatin, a) extended, b) compressed state. Each sphere represents a base pair in the DNA thread. The nucleosomes surounded by the double ring structures are not shown.



(a)



(b)

Fig.1