DARESBURY LABORATORY

NEWSLETTER on PROTEIN CRYSTALLOGRAPHY

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

Number 23 June 1989

Contents

Editor: Kim Henrick

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

USE OF SIMULATED ANNEALING IN PROTEIN REFINEMENT

I Haneef
Astbury Department of Biophysics
University of Leeds
Leeds, LS2 9JT.

1. INTRODUCTION

Optimization problems are ubiquitous in many branches of science. In a nutshell: given a function $F(x)$ which depends on one or more independent variables x, one needs to find a set of variables x which minimize (or maximize) the value of F. Clearly, one is interested in a method that delivers the results with least amount of computational cost. Often, the computational effort is dominated by the cost of evaluating F (and its partial derivatives with respect to all variables, if such derivatives exist and/or are required by the optimization algorithm). Thus, from the computational point of view, optimization methods are required to deliver the result in least number of evaluations of F.

For the most general case, there exist a number of other problems. In particular, if F has very many minima, one is often interested in the global minimum. There is little or no theory about how to obtain the global minimum for a general function, or indeed, how to distinguish the global minimum from the many other 'local' minima. Classical solutions to this problem are variants of the multi-start method. In this method, very many local minima are obtained by starting from a large number of disparate starting points, and the minimum with the least value is chosen. The success of this method clearly depends on how finely one samples the configuration space. For systems with only a few degrees of freedom, multi-start methods can be very efficient and often effectively deliver the global minimum.

For large systems, the technique of simulated annealing appears to have considerable potential and has in recent years attracted significant attention. The method differs significantly from the classical methods. In classical methods a sequence of configurations $\{F\}$ is generated such that the nth configuration has value of F strictly lower than that of the previous configuration. In the annealing method, the latter requirement is relaxed; there exists a finite probability p that the nth configuration will have a value of F that is larger than that of the previous configuration. This effectively gives the system a chance to move out of a local minimum in favour of finding a better minimum. The probabilty p can be changed and is determined by the 'temperature' of the system; at high temperatures, the system has greater 'mobility' and can span larger regions of the configurational space in search of the global minimum.

2. APPLICATION OF SIMULATED ANNEALING TO PROTEIN REFINEMENT

Structure factor least-squares refinement of large molecules is an iterative process involving several rounds of programmed optimization step (typically several cycles of structure factor least-squares

refinement using programs such a PROLSQ, (Konnert and Hendrickson, 1980)) followed by manual rebuilding of the model using the interactive computer graphics program FRODO (Jones, 1978). Manual intervention is required because the least-squares procedures become trapped in local minima after only a few cycles. Manual rebuilding is an onerous task, often requiring the modeller to manipulate fragments of the molecule with very many degrees of freedom to obtain good fit to electron density, a task the human brain simply was not designed to do! Simulated annealing method, in principle, offers an opportunity to automate the whole process.

We have used the simulated annealing method to refine the structure of the complex of the monoclonal D1.3 Fab with hen lysozyme (resolution 2.8A, Amit *et al*, 1986). We continued the refinement from the best current model from the 2.8A data to see if poorly defined regions of this model could be improved. The complex structure is also being refined at Pasteur Institute (Paris) using data to 2.5A resolution; the 2.5A refinement, using conventional restrained structure factor least-squares refinement methods, has been used to rebuild an incorrectly modelled region of the structure. This region comprises 20 residues, residues 153-173 of the light chain, and is part of the protein beta-sheet. It was of interest to us to see if simulated annealing could correct this region of the structure using diffraction data only up to 2.8A.

The program employed was XPLOR (Brunger *et al* 1987). The objective function minimized by the program comprises the conventional residual between the observed and calculated structure factor amplitudes, and a potential describing the interatomic interactions. The protocol used was essentially that given by Brunger in the XPLOR user manual:

a. 160 steps of energy minimization
b. 1 pico-second molecular dynamics simulation at 2000K, using time step of 1 femto-seconds; the gradient vector component from the objective function was updated only if any atom moved by more than 0.2A.
c. 2 pico-second molecular dynamics simulation, with gradual decrease of the temperature from 2000K to 300K, with temperature rescaling at every 25 femto-seconds.
d. 40 steps of energy minimization.
e. Restrained B-factor refinement.

The R-factor dropped from 28% (initial model, from PROLSQ) to 20% after XPLOR refinement. The rms differences between initial and final structures is 0.8A (main chain atoms) and 1.6A (all atoms). Although some side chains have moved by several Angstroms (largest shift *ca.* 7A), the large drop in the R-factor is rather surprising.

The comparison of the electron density using phases from the new model and that using phases from the previous model shows some regions where there is substantial improvement in the structure. An example of improved regions of the structure is shown in Fig. 1. In the Fig. are shown old and new conformations for the loop region 130-140 in the heavy chain with the negative region of the Fo-Fc map superposed. In the pre-XPLOR map, Thr 135 was located in a negative region of the map; however,

there was no indication from the Fo-Fc map as to how one should rebuild this region. In the post-XPLOR map no negative regions are visible at this contour level.

One region of the Fab/HEL complex of particular interest was region 153-176 in the light chain. This region has been poorly defined from the very outset (when the structure was first solved). The best model has this region as two short loops at either end of a long beta-strand. Comparisons with other Fab structures suggests that there are two residues too many in the N-terminal loop and two residues too few in the C-terminal loop of this region. Thus a shift of two residues is suggested throughout the beta-strand - however, due to the general poor quality of the electron density maps in this region, it has not been possible to model build this region using the 2.8A resolution data. In the post-XPLOR maps, this region is still poorly defined. The main chain trace essentially follows that of the pre-XPLOR model (rms difference of ca. 1.3A), with no indication of shift of residues along the sequence. In the post-XPLOR model, many of the side chains show considerable shifts from the pre-XPLOR model. Examination of the electron density of this region shows that many of the side chains have moved closer to the main chain (some side chains moving by more than 6A); thus it would appear that these side chains have moved into regions of electron density occupied by the main chain atoms. We have observed similar movements of long side chains in other regions of the molecule.

3. SUMMARY

Simulated annealing refinement of D.13 Fab/HEL complex reduced the R-factor from 28% to 20% using all the diffraction data from 20A to 2.8A resolution. Examination of the electron density maps show that some regions (typically regions comprising some two to four residues) have improved markedly. There are, however, regions of the structure (mainly long flexible side chains) that have been moved out of density leading to some deterioration of the model structure. One region of the structure, of *ca.* 20 residues, which is known to be incorrect does not improve during the refinement. For improvement to occur, a large region of the structure would have to be moved in a correlated fashion to obtain correct fit to electron density. For this region at least, the simulated annealing method seems unable to produce correlated shifts to correct the structure; but rather, many of the side chains are pulled close to the main chain with the consequence that these side chains share the electron density with main chain atoms, thus perhaps contributing to the rather dramatic drop in the R-factor.

As a method of optimization, simulated annealing has attracted a great deal of attention and has been used to effectively solve a number of difficult global minima problems (e.g. Wille, 1984). However, considerable research effort is still required to ascertain useful and generally applicable protocols for problems where the objective function comprises several competing terms. In its use in refining protein structures there exists a practical problem of how the two terms (the diffraction data and the energy data) should be weighted with respect to each other; this clearly depends both on the resolution and quality of the diffraction data and the accuracy of the potential functions used.

For a system with large number of degrees of freedom, one may require much longer simulations since such a system needs to sample a much larger volume of configurational space to find the global minimum. Another problem that needs some attention is the temperature one needs to use during the heating stage of the simulation; clearly, if the local minimum from which the system needs to escape is very deep, one needs to use higher temperatures.

REFERENCES

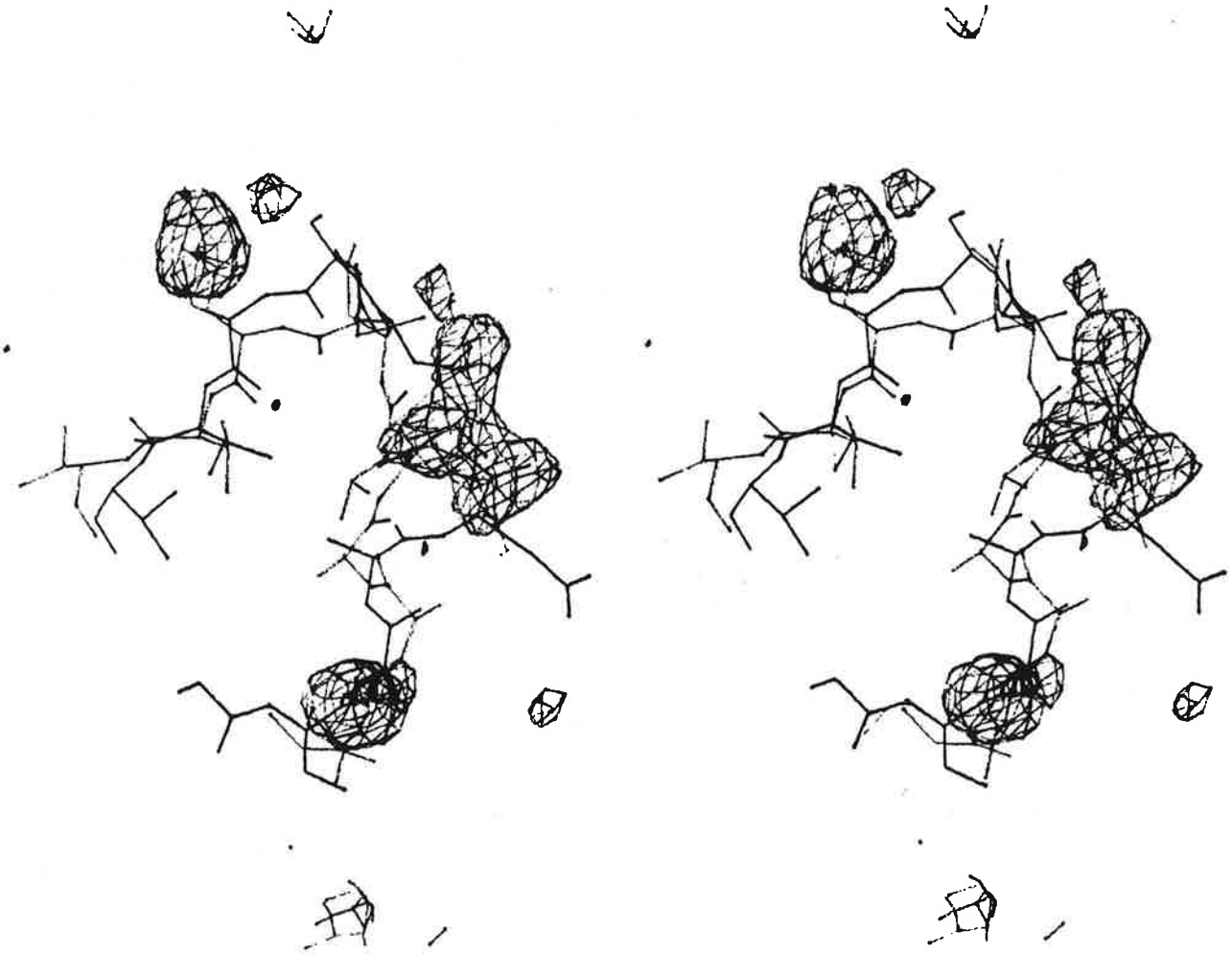
Amit A G, Mariuzza R A, Phillips S E V and Poljak R J (1986), Science ***233***, 747

Brunger A T, Kuriyan J and Karplus M (1987), Science ***235***, 458

Jones T A (1978), J Appl Cryst ***11***, 268

Konnert J H and Hendrickson W A (1980), Acta Cryst ***A36***, 344

Wille L T (1986), Nature ***324***, 46



An Experiment to Determine an Improved Strategy for Laue Photographic Data Collection

by Stephen J. Maginn,
Dept. of Chemistry,
Donnan Labs,
University of Liverpool

Following the successful solutions and refinements of the structures of two organometallic compounds by the Laue photographic method[1,2], it was decided to perform an experiment to optimise the data collection strategy so as to improve the chances of successful solution and refinement for further, more difficult samples. A number of different data sets would be taken for one crystal, of normal size and known structure, using different strategies.

The two parameters to be varied for each data set were to be firstly the arrangement of films and foils in the packs used, and secondly the presence or absence of a 200μm aluminium attenuator in the incident beam. The film packs used for data collection with the previous two compounds had consisted of six films interleaved with 200μm aluminium and 36μm copper foils, to increase the dynamic range available and provide some wavelength discrimination to aid in unscrambling multiplet reflections. However, since unscrambling is not performed for small molecule solutions and the presence of foils seemed to cause problems in data processing, particularly in INTLAUE at the twist, tilt and bulge refinement stage, it was decided to try a new pack arrangement consisting of six films interleaved with other, already developed films known as "spacers". It was also required to test the effect of inserting a 200μm aluminium attenuator in the incident beam; it was hoped that this would flatten out the wavelength normalisation curve and thus enable it to be more accurately determined.

The compound chosen for use in this experiment was proflavine hemisulphate hydrate. This organic compound of known structure (determination performed by diffractometer - Jones & Neidle[3]) has a large unit cell, giving a comparable spot density on the Laue exposures to the organometallic compounds normally studied. Also, it has a low linear absorption coefficient and forms large, easily obtainable crystals.

Details of the various different data sets taken, strategies used and processing results obtained are given in Table 1. Two crystals were used on different dates using station 9.7 at Daresbury. All the photographs produced were of excellent quality, showing small, round spots, in contrast to the streaked reflections found on Laue photographs of small crystals. Crystal B proved to be of inferior quality to crystal A, so data sets 1,2 and 3 cannot be directly compared with sets 4 and 5. In each crystal's case, a translation along the length of the crystal was performed between data sets, to avoid radiation damage effects.

Comparing data sets 1 and 3 shows the effect of the new pack arrangement. Processing results in GENLAUE and INTLAUE are seen to be dramatically better, particularly the twist, tilt and bulge (ttb) refinements in INTLAUE, clearly showing the deleterious effect of the foils. This effect is thought to be due to kinking and creasing in the foils - even brand new foils are often quite badly bowed. The price paid for replacing foils with spacer films is an increase in the number of overloads on F films, due to the reduction in dynamic range. However, the strongest spots tend to be multiplets anyway, and would therefore be discarded later in processing, so the advantage is clear. The LAUENORM merging R obtained for set 3, 0.070, is the best yet achieved for a small molecule data set.

A comparison of sets 1 and 2 shows the effect of the aluminium attenuator. Processing again proceeds more smoothly for set 2, and the LAUENORM normalisation curve is clearly less steep, although the effect is not as pronounced as was hoped. It is thought that the attenuator would help more at longer wavelengths where its absorption is greater (see Fig.1) and will cut down on radiation damage produced by such λ values.

The results of refinement show the same trends. No attempt was made at structure solution - so many light atoms in the assymetric unit would have made such a task with Laue data rather difficult! - initial positional parameters being taken from Jones and Neidle[3]. The same refinement strategy was followed for each set. On letting thermal parameters refine anisotropically, sets 1 and 2 gave non-positive definite atoms (i.e. the thermal parameters refined to physically meaningless values), but set 3 provided a satisfactory, reasonable refinement to convergence at $R = 0.076$. This compares very well with the value from diffractometer data of $R = 0.074$, although it must be stressed that Laue data has provided fewer reflections, so refinement is less reliable. Fig.2 shows one proflavine moiety complete with bond lengths from both set 3 and diffractometer data - it can be seen that trends are identical, and in fact all lengths are within 2σ of the previously determined values.

The beneficial effects of both the new pack arrangement and the attenuator individually seem clear. Sets 4 and 5 were taken using a different crystal of proflavine hemisulphate to investigate the effect of both together. Set 4 used the identical strategy to set 3, and set 5 had both the all-films pack and the attenuator inserted.

Set 5 indeed provided a more satisfactory processing and refinement than set 4, but unfortunately neither were up to the high standard of set 3, indicative of the inferior quality of crystal B. One pack in each set had to be discarded due to poor processing, leaving just three contributing to the final data set. Fewer atoms were allowed to refine anisotropically due to the smaller number of reflections in sets 4 and 5, so comparison with the results for sets 1,2 and 3 is not valid. However, the beneficial effect of the aluminium attenuator is again apparent on comparing sets 4 and 5.

So, for the purpose of small molecule structure solution and refinement, an optimum strategy for photographic data collection is the use of a film pack consisting exclusively of ten films, six of which contribute to the data set, and the use of an aluminium attenuator in the incident beam. Clearly, when all things are optimised as in set 3, the accuracy of Laue data can approach that of diffractometer data, showing the potential of the method.

Crystal Data

$C_{13}H_{12}N_3^+$ $0.5(SO_4^{2-}).1.75H_2O$, f.w. = 289.7, monoclinic, $P2_1/c$, $a = 12.703(1)$Å, $b = 19.940(2)$Å, $c = 21.487(2)$Å, $\beta = 92.24(2)°$, $Z = 16$, $\rho_c = 1.415 g/cm^3$, $\mu = 18.96 cm^{-1}$ (at $\lambda = 1.542$Å)

Cell lengths given and used in refinement here are from Jones and Neidle[3]. Those refined in GENLAUE for sets 1,2 and 3 were all within 2σ of these values; those for sets 4 and 5 were within 5σ.

References

1. M.M. Harding, S.J. Maginn, J.W. Campbell, I.J. Clifton, P. Machin; Acta Cryst. B44 (1988), 142

2. J.A. Clucas, M.M. Harding, S.J. Maginn; J. Chem. Soc. Chem. Comm. (1988), 185

3. A. Jones, S. Neidle; Acta Cryst. B31 (1975), 1324

	Crystal A, 0.3 × 0.1 × 0.6mm, 12/87			Crystal B, 0.5 × 0.4 × 0.3mm, 9/88	
Data Set	1	2	3	4	5
pack	films & foils	films & foils	10 films only	10 films only	10 films only
attenuator	none	0.2mm Al	none	none	0.2mm Al
SRS	180mA, 2GeV	147mA, 2GeV	140mA, 2GeV	175mA, 2GeV	174mA, 2GeV
exposure	10s	16s	10s	1.5s	2.0s
crystal-film distance	58.2mm	61.7mm	61.7mm	60.3mm	60.3mm
GENLAUE rms	0.065-0.095mm	0.041-0.072mm	0.041-0.055mm	0.027-0.08mm	0.035-0.08mm
INTLAUE ttb	better than 8.0 for A-E, worse than 12.0 for F	better than 5.0 for A-E, worse than 7.0 for F	better than 4.0 for all films A-F	all < 5.0	all < 5.0
nos. of overloads on F films	< 10	20-40	50-90	~ 10	50-70
LAUENORM curves:					
λ = 0.316	0.0025	0.0043	0.0030	0.0045	0.0068
0.348	0.0051	0.0088	0.0059	0.0082	0.0131
0.380	0.0099	0.0169	0.0119	0.0153	0.0243
0.412	0.0157	0.0264	0.0203	0.0261	0.0390
0.444	0.0286	0.0507	0.0329	0.0422	0.0656
0.526	0.0424	0.0633	0.0426	0.0540	0.0792
0.577	0.0770	0.1030	0.0757	0.0874	0.1312
0.628	0.1359	0.1716	0.1289	0.1434	0.2195
0.679	0.2241	0.2755	0.2087	0.2351	0.3410
0.731	0.3434	0.4029	0.3274	0.3714	0.4721
0.782	0.5090	0.5478	0.5032	0.5374	0.6377
0.833	0.7505	0.7587	0.7262	0.7078	0.9306
0.884	1.0000	1.0000	1.0000	1.0000	1.0000
no. of packs	4	4	4	3	3
no. of reflections in final data set	2029	2146	2125	1779	1555
merging R	0.108	0.087	0.070	0.076	0.071
SHELX final R	**0.097**	**0.087**	**0.076**	**0.1193**	**0.1126**

Table 1.

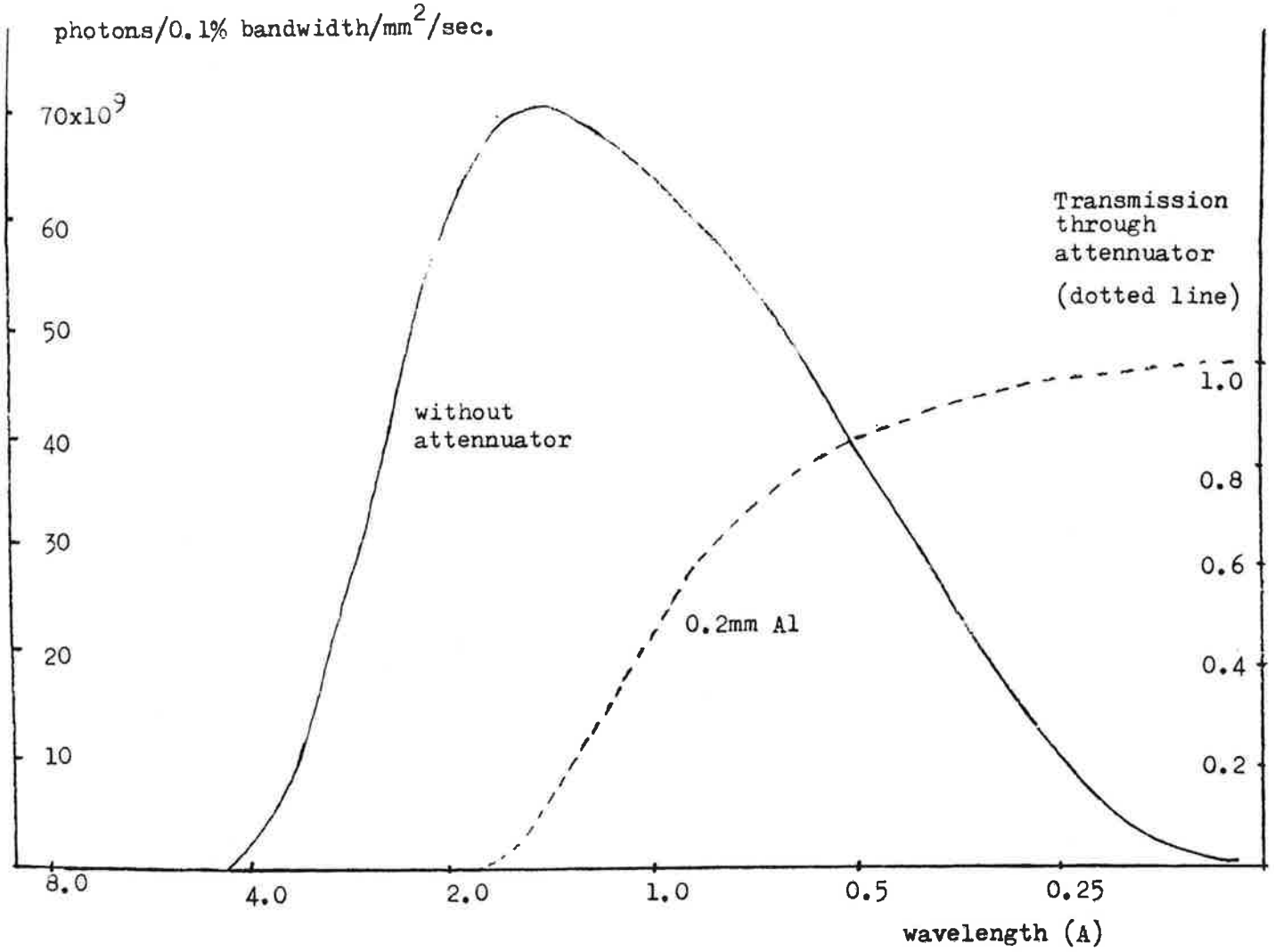
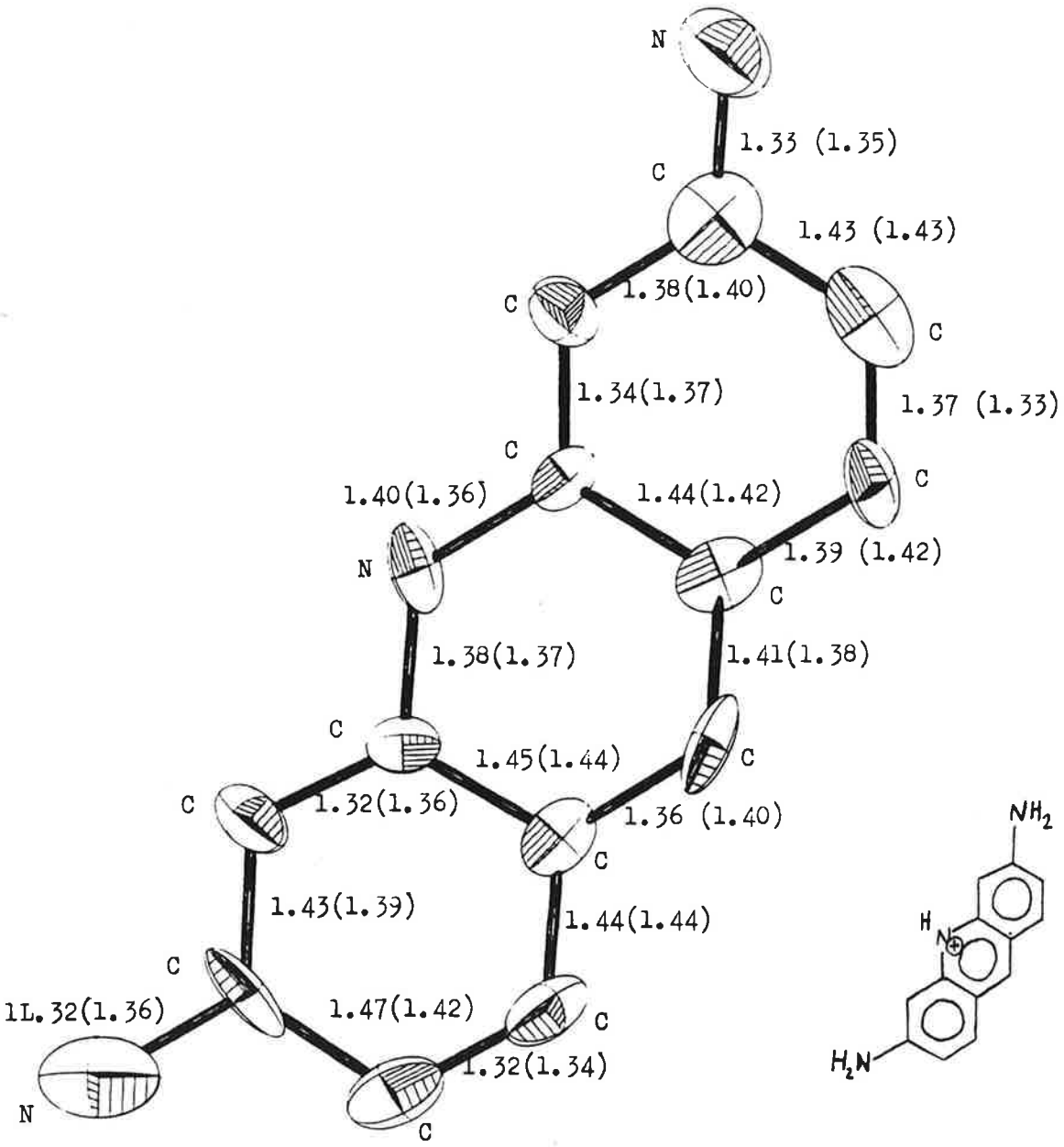


Fig.1 : The effect of the aluminium attenuator (dotted line) on the wavelength profile of the beam (station 9.7, 200mA)

Fig.2 : Bond lengths from Laue set 3 and diffractometer (in brackets), for one proflavine unit.



PROFLAVINE MOLECULE 1

LATTICE: A visual aid for finding symmetry in reciprocal space

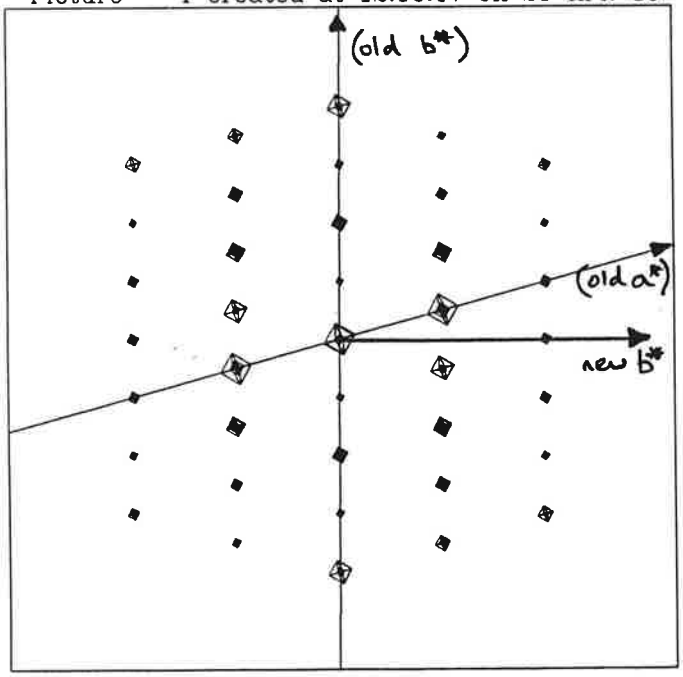
Garry Taylor

Laboratory of Molecular Biophysics, The Rex Richards Building,
South Parks Road, Oxford OX1 3QU. (Garry@UK.AC.OX.BIOP)

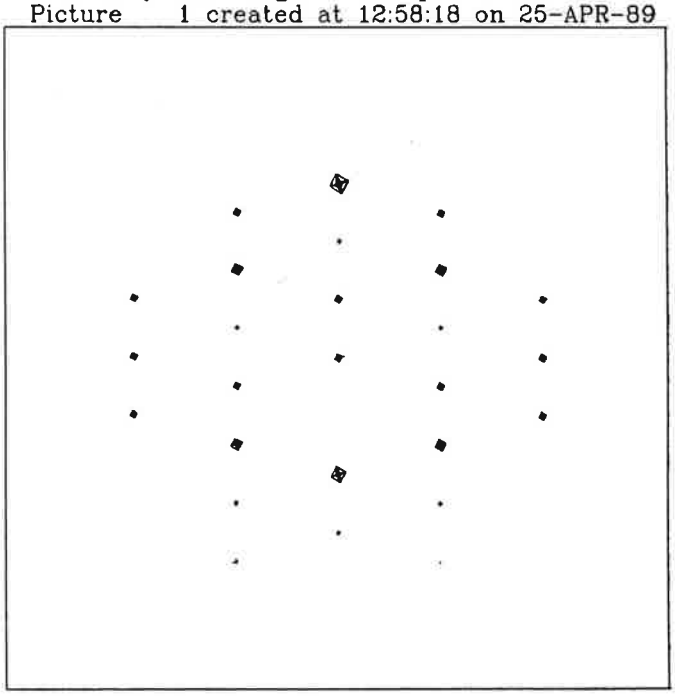
In the course of collecting crystal data for antigen binding fragments of antibodies (Phil Jeffrey & Bob Griest), a problem arose when deriving the unit cell during autoindexing. Data was collected 'blindly' on the Xentronix area detector, as the Fabs tend to be unpredictable as to their crystal form: we have two crystal forms of one Fab, and as crystals are precious, no photography is carried out beforehand. In one particular case, a P1 cell was predicted by the autoindexing routine of the XENGEN software, quite consistent with a minimum basis set of vectors. A self rotation function on this P1 set, revealed a strong 2-fold axis which lead us to question wether the crystal was of higher symmetry. To aid this process, the program LATTICE was written, which simply reads an LCF file of F's or I's and creates a weighted reciprocal lattice in a form suitable for reading into the VECTOR option of FRODO. Hence the lattice, together with vectors describing the reciprocal cell axes, can be rotated to search for symmetry. Fig. 1 shows the weighted lattice of a 'P1' form of Gloop4 (Gloop denotes one of 5 monoclonal antibodies raised against the loop peptide of hen egg white lysozyme-residues 60 to 84). The P1 cell had dimensions of a=39.4, b=72.0, c=83.7, α=91.3, β=76.4, γ=105.9. A mirror is clearly evident, shown more clearly in fig.2 where I have slabbed down and moved away from the origin. The spot distribution is consistent with a centred lattice. Rotating 90^{o} reveals a view down the 2-fold axis (fig.3), with a clear 2-fold symmetry between the lattice points shown in fig.4 where one plane is shown away from the origin. A careful exploration of reciprocal space indicated a reindexing of (h,k,l) to (2k+h,h,k+l). The new cell was estimated at approx. a=104, b=40, c=140Å with a β=130^{o}, symmetry C2. A new refinement of the cell gave: a=104.8, b=39.3, c=138.2, β=129^{o}.

This may be of use to others who do not know their crystal form before embarking on data collection, which is becoming a more common feature given the speed and convenience of area detector data collection. The program is available for free distribution.

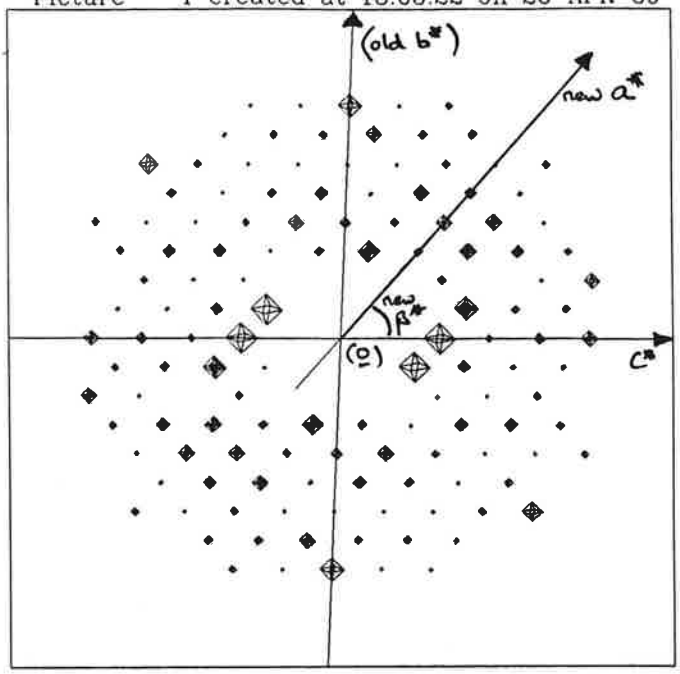
View of mirror for Gloop4 P1 data
Picture 1 created at 12:56:17 on 25-APR-89



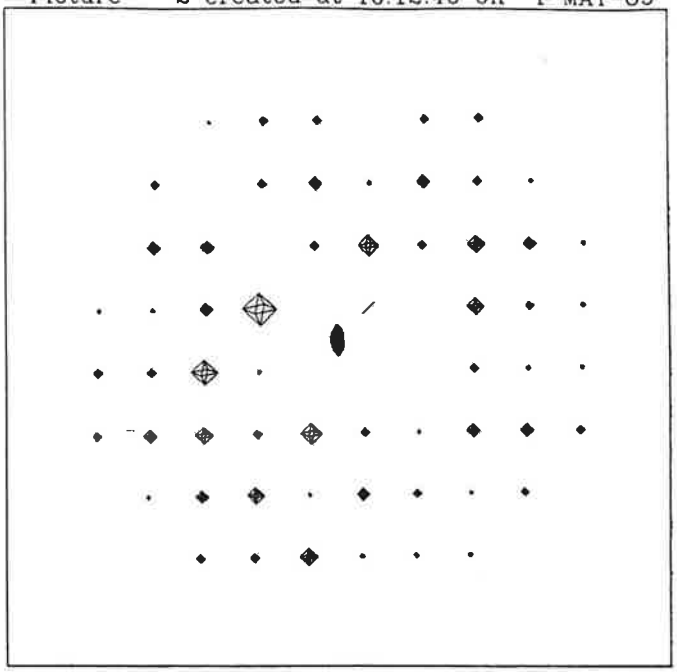
Slab away from origin of Gloop4 P1 data
Picture 1 created at 12:58:18 on 25-APR-89



Gloop4 P1 data: new C2 a*,c* plane
Picture 1 created at 13:03:22 on 25-APR-89



Gloop4 P1 data, new C2 a*c* plane
Picture 2 created at 10:12:46 on 1-MAY-89



AUTOMATED SEARCHING FOR SECONDARY STRUCTURE MOTIFS IN THE PROTEIN DATA BANK USING TECHNIQUES DERIVED FROM GRAPH THEORY.

Peter Artymiuk, Eleanor Mitchell*, David Rice & Peter Willett*

The Krebs Institute for Biomolecular Research, Departments of Biochemistry & of Information Studies(*), Sheffield University, Sheffield S10 2TN, U.K.

Introduction

As part of an interdisciplinary collaboration between members of the Krebs Institute in the Departments of Biochemistry and of Information Studies, Sheffield University, methods are being developed for the automatic identification of secondary structure motifs in 3-dimensional macromolecular structures in the Protein Data Bank. There is a pressing need to evolve methods for the rapid comparison of such structures, for the efficient detection of points of similarity and of common patterns of polypeptide chain folding ("motifs"), and ultimately to discover the principles that govern the folding of proteins.

Our work approaches the problem of identifying common folding motifs in a novel way: algorithms derived from the branch of mathematics known as Graph Theory have proved invaluable in the rapid searching of small molecule databases; we have applied related methods to the problem of protein structural analysis by representing the essential features of the protein folds of the entire Protein Data bank as labelled graphs, with the nodes of a graph corresponding to linear representations of alpha helices and beta strands, and the edges of a graph to the distance and angular relationships between them (Mitchell, Artymiuk, Rice and Willett, *J. Molec. Biol.*, accepted for publication). Query motifs are then defined in an analogous manner. The graph representations are searched using a modification of Ullman's subgraph isomorphism algorithm (*J. Association for Computing Machinery* 23, 31-42 (1976).) The program is named

POSSUM

= Protein On-line Substructure Search, Ullman Method

To achieve consistency in our creation of the database, helix and strand assignments were made using the Kabsch and Sander algorithm (*Biopolymers*, 22, 2577-2637 (1983)), except for Cα-only files where the depositors' assignments had to be used where available, or our own judgement where not. We have made such vectoral representations of the entire Protein Data Bank,

and these have been combined into a database. Search substructures ("motifs") can be created *de novo*, or extracted from existing structures.

Searches are fast: searching the entire PDB for a 9 element motif requires around 200 seconds on a microVAX II. The program writes out coordinates of arrow representations of the motifs, when they are found, together with control (.REC) files to enable their easy display using the CCP4 E&S PS300 version of Alwyn Jones' FRODO program.

It is a strength of the POSSUM program that the actual search is carried out totally independent of sequence order of secondary structure elements, although there is an option in the program which, if desired, allows the user to select for monitoring only those 'hits' which are in the correct sequence order.

We have also made provision for the inclusion of 'wild-card' values for interelement angles and distances in our searches. this permits the definition of much more schematic representations of motifs than is possible in fully-defined search patterns.

Testing the program.

In order to test the POSSUM program a number of previously identified motifs have been used to search the database. We only give one example here: the "beta (or TIM) barrel" which consists of 8 parallel beta-strands. This motif is easily spotted visually in a structure and is known to occur in 13 different protein structures (see Chothia, *Nature*, 333, 598-599 (1988)), of which only 5 are in the April 1988 version of the PDB (this figure excludes KDPG aldolase (1KGA) where the deposited coordinates are wrong, as pointed out by Richardson (*Biochem Biophys Res Comms* 90, 285-290 (1979)). A pattern corresponding to an 8-stranded beta barrel was created and, using typical tolerances of 40 degrees on inter-element angle, and 40 % on closest approach distance between elements, the program correctly found the motif in all 5 proteins: in triose phosphate isomerase (1TIM) itself, in pyruvate kinase (1PYK) in taka-amylase (2TAA), in enolase (1ENL) and in xylose isomerase (1XIA). The motif was not found in any other proteins in the PDB. This and a number of other tests indicate that the program works rapidly and correctly to identify known motifs in the PDB.

The program took 100 seconds of microVAX III cpu time to perform the TIM barrel search. The on-screen output of the program is fairly self-explanatory, and is shown on the next page (user responses in *italics*):

```
$ POSSUM
 IDCODE OF INPUT PAT FILE ?
BBAR
 Closest Approach or Midpoint Distance? [M/C*]
C
HEADER 8 stranded parallel beta barrel (a la TIM)
What angular tolerance [in degrees] do you want?
40
Distance discrepancy .. state A or % (eg 10A or 30%)
40%
PERCENTAGE distance discrepancy=    40.00000     %
 NAME OF OUTPUT *LP FILE ?

Is the sequence order of the elements in the pattern important? [y*/n]
Y
 Do you want to create the [.NEW] files on FRODO [y/n*]
N
****************** POSSUM ******************

Search code = BBAR
Aepsilon=     40.00     Depsilon=     40.00 %

 CLOSEST APPROACH DISTANCES USED

1ENLca ENOLASE (E.C.4.2.1.11) (2-PHOSPHO-*D-GLYCERATE HYDROLASE)   :     1 hits

1PYKca PYRUVATE KINASE (E.C.2.7.1.40)                              :     1 hits

1TIM   TRIOSE PHOSPHATE ISOMERASE (E.C.5.3.1.1)                    :     1 hits

1XIAca D-*XYLOSE ISOMERASE (E.C.5.3.1.5)                           :     1 hits

2TAA   TAKA-*AMYLASE A (E.C.3.2.1.1)                               :     1 hits

The  Total number of matches is        5
The number of files with hits is       5

Search code = BBAR
Aepsilon=     40.00     Depsilon=     40.00 %      CLOSEST APPROACH
OK
CPUTIME used = 99.49 seconds
$
```

More detailed information is available in lineprinter files, and by examination of selected motifs on the PS300.

Future developments

The POSSUM program works effectively in recognising known structural motifs in the PDB. A major aim for the future is to identify novel motifs in protein structures. At present the

motifs that have been identified in proteins are for the most part simple ones. They have been identified either because of their elegance and simplicity (eg: the beta barrel), because of clear sequence homology (calcium binding fold, immunoglobulin fold). or because of a relation to function (eg: NAD binding fold). We intend to use the program to search the database for new, less obvious motifs. These may prove to be related to function, or to represent especially stable arrangements of secondary structural elements.

THE PROTEIN CRYSTALLOGRAPHY LABORATORY AT THE UNIVERSITY OF GLASGOW

Neil Isaacs

The Department of Chemistry at Glasgow has a long and distinguished history in crystallography. Among the former staff and students of the Department are found such illustrious names as J.M. Robertson, M. Rossmann, D.W.J. Cruickshank and G.A. Sim. In 1975 the laboratory branched into protein crystallographic work with the appointment of Ian Swan to a lectureship. Following his untimely death in 1980, the protein laboratory lapsed, but the development of molecular sciences in other departments of the University led to a decision in 1988 to re-establish protein crystallography in the Department of Chemistry. Neil Isaacs has been appointed to a new Chair of Protein Crystallography to establish the protein crystallography laboratory and to head the crystallography section as a whole. Assisting him with the protein work are Andy Freer, Paul Mallinson and Deborah Harris. Paul will be supervising the computing facilities and will be the point of contact for CCP4. The laboratory will be housed in a wing of the Chemistry Building which is presently being refurbished and will have sufficient new equipment (as well as the use of the two existing CAD4 diffractometers) to enable it to function efficiently. Initially the unit will work on the structure of the glycoprotein hormone, human chorionic gonadotropin (HCG) and is aiming to develop new projects in areas as diverse as DNA binding proteins, shikinate pathway enzymes, and bacterial photoreaction centre complexes in collaboration with other departments in the University.

INFORMATION FROM ELECTRON DENSITY MAPS

Guy Dodson

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

and

Judith Everett

Structural Biophysics Unit, Physics Division, Liverpool Polytechnic, Liverpool, L3 3AF, England.

The arrival of graphics devices such as the Evans & Sutherland PS system has made the analysis of electron density maps much more convenient than by the inspection of contoured map sections. However the chicken wire system, which gives an impressive 3 dimensionality, does not so easily allow contouring. This presents few problems when the structure is well ordered and the electron density is consequently well defined. When the structure under study is poorly ordered and the electron density is correspondingly lower and spreading, then the chicken wire representation has disadvantages and it can be difficult to identify maxima which correspond to possible atomic positions. The difficulties with diffuse density has, generally, led crystallographers to use contour levels which cut it out. This simplification facilitates analysis of the structure but it can mean that details of more mobile residues and particularly water structure are lost. It is true that poorly ordered side–chains cannot usually be accurately defined (only the approximate structures are identified) and are an important element in a protein molecule's description.

The tendency to examine maps at high contour levels is supported by the use of the standard deviation given in the program FRODO. Most workers regard a value of 1σ (where σ is the rms density) as the noise level, but in reality structurally significant density occurs at lower levels than this. One convenient test for the error level in an electron density map is the appearance of the density in the protein's non–polar interior. Here the non–polar side–chains are generally packed together at van der Waals contact and exclude water molecules. Levels of contour which do not reveal peaks in these regions are usefully above noise.

Some recent experiences with several proteins illustrates the value not only in reducing the electron density contour levels to low levels but also in working with conventionally contoured map sections. If the contour levels are chosen appropriately the detailed structure of the more populated water positions can often be identified with reasonable ease. There are however great disadvantages in working with paper maps, not least in recording the coordinates of the introduced atoms. This and other difficulties have been overcome by working with contoured map sections displayed on a terminal.

A program has been developed by Judith Everett at Liverpool Polytechnic (Everett, Groves & Prax, 1989) for displaying map sections on Tectronix compatible equipment such as the Pericom. A selected map section is displayed and an option is chosen from the menu using a mouse. The option offered by this program, IMAP, are as follows;

MOVE	Interactive manipulation of atomic positions allows an atom to be moved to a new position
INSW	Interactive Insertion of water molecules
INSA	Interactive insertion of atomic positions
DIST	Interactive determination of interatomic distances displays a distance between two selected atoms
ZOOM	Any region of the displayed map section can be selected using the mouse and magnified
SECU	Map sections are displayed sequentially in the upwards direction
SECD	Map sections are displayed sequentially in the downwards direction
NAYB	An atom is selected from the current map section and all atoms within 3.5$\AA$ are listed together with their contact distances and the section they lie on. The position of these atoms may be displayed as a projection onto the current section.
PLOT	Outputs a hard copy of the section displayed on the section.
SAVE	Saves the current coordinates to a Brookhaven file.
END	Exits without saving

The program IMAP operates under the VAX/VMS operating system and is written in Fortran 77. It requires GINOF and NAGF libraries. The current software allows the user to select Tectronix drivers T4014, T4010 or T4111. Hard copy is obtained on Calcomp plotters using drivers CC81 or CC1051. The program also requires the use of a mouse.

The contouring levels can be selected to reveal the particular features under study most favourably but generally it is as easy to use the range from, say, zero to the maximum values of electron density. With this regime the proper appearance of the electron density for the whole molecule and the solvent is available for inspection. The system permits the positioning of an atom, say a water molecule, and then checks its contacts with those atoms already positioned, using the **NAYB** option. This allows ready examination of possible hydrogen bonding and van der Waals contacts and the identification of impossible contacts which imply either mistakes in the arrangement of the local structure or the introduced atom or the presence of partial occupancy.

Some examples of how the analysis of low electron density levels have revealed disorder and water structure are given.

Monoclinic insulin

The structure of a new hexameric form of human insulin, monoclinic insulin, containing six independent molecules, has been solved by molecular replacement and refined using 2.25Å resolution data to a residual of 0.22. It is currently being refined with 1.8Å resolution data and at present the residual is 0.19. Some disorder is evident in the A14–Tyr side–chain of subunit–A where two orientations for this side–chain are present (*Figure 1.* The position of the B30–Thr residue in tne subunit–B is also difficult to place because of very weak density. Over five hundred solvent water molecules have been determined and this has been greatly facilitated by the use of the program IMAP.

Despentapeptide insulin

The crystal structure of beef despentapeptide insulin has been determined, originally by molecular replacement, and refined with 1.5Å resolution data.

More recently at 1.2Å spacing a data set has been collected at the SRS on the FAST and the crystal structure further refined. This high resolution analysis has revealed clear disorder in some of the residues.

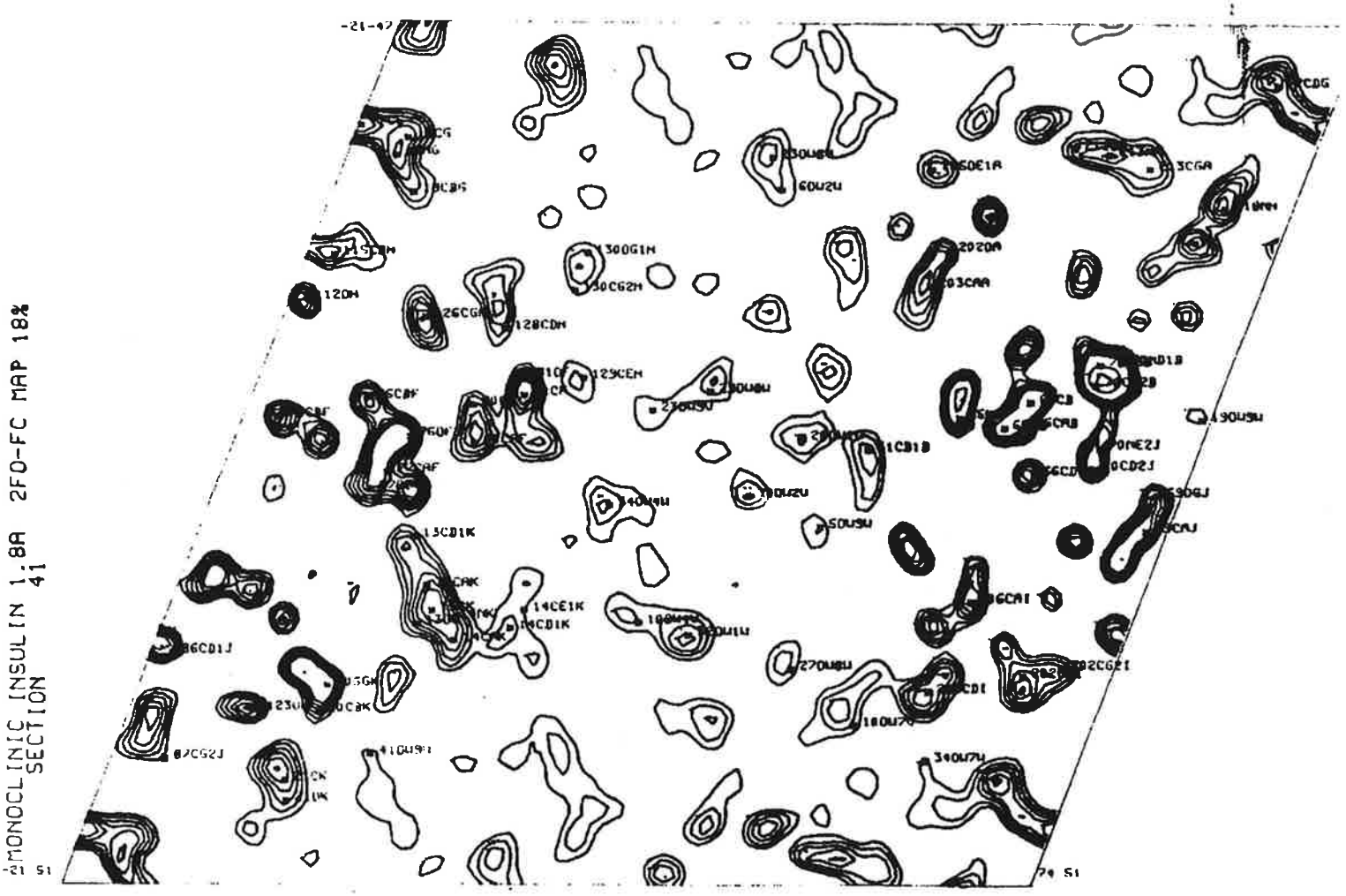


Figure 1. *Section of the electron density map at y = 41/48 for monoclinic insulin. Note the weak density for the A14–Tyr side-chain (lower left). Contours start at 0.5σ and increase at intervals of 0.5σ.*

The bacterial ribonuclease Sa

This enzyme has been carefully refined at 1.8Å resolution as the free enzyme and as the complex with the inhibitor 3′GMP. the crystal contains two molecules in the asymmetric unit but owing to lattice contacts only one of the molecules can bind the 3′GMP. The 3′GMP is very clearly seen in the one molecule and its conformation is apparently unambiguously defined. This has presented something of a puzzle since the

2′OH group which is attacked by the enzyme, is directed away from the catalytic group (glutamic acid, Glu73). It has been postulated by Chris Hill that an alternative bonding conformation exists in which the sugar moiety rotates by about 180°; this would bring the 2′OH group adjacent to the catalytic glutamic acid side-chain. Inspection of the electron density at low levels shows evidence for such a conformation.

Rhombohedral 2ZN insulin

The crystal structure of 2Zn insulin has been exhaustively refined (Baker *et al*, 1988) and many features of the molecule's conformation have been analysed. In particular seven clear examples of disordered side-chains have been identified in which the two alternative conformations are roughly equally populated. Other side-chains show evidence of alternative conformations which are much less occupied. A good example of this is the B25 phenylalanine side-chain of one of the two insulin molecules. The electron density in the region of this residue, contoured from $0.3e/\mathring{A}^3$, reveals one major conformation and a second much less populated conformation at χ_1 about 120° away *Figure 2.* This second conformation is sterically reasonable and indeed predicted in several calculations (Shoshana Wodak, and Axel Wollmer). Part of the water structure around neighbouring hexamers is illustrated in *Figure 3.*

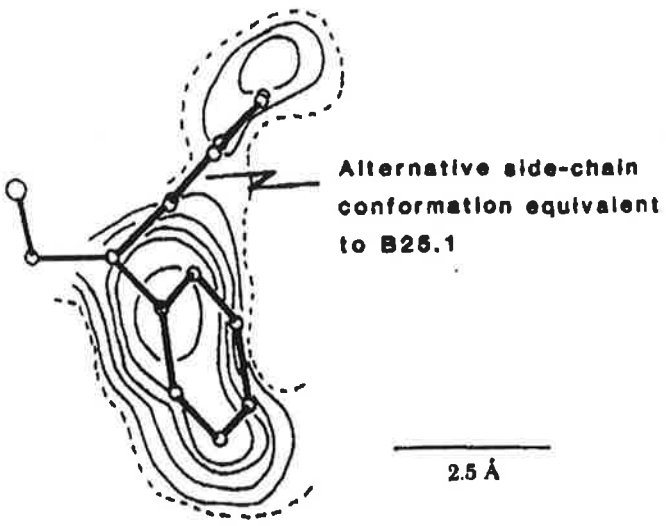


Figure 2. *The electron density at B25.2 Phe side-chain from the F_{obs} Fourier synthesis*

on section $z = 0$. The broken contours represent 0.3e/Å^3, the continuous contours are at 0.5, 0.6, 0.7, 1.0 and 1.3 eÅ^3 all on the absolute scale. An alternative conformation, roughly equivalent to that in molecule **1**, *can be seen to lie in low, but persistent, electron density.*

These examples should help to remind us that the electron density can contain evidence for the existence of structures which are not a major population in the crystal. Since these structures can sometimes carry clues about the protein's mechanism or structural behaviour, it is wise for protein crystallographers to extend their analysis of the electron density to the lowest significant levels.

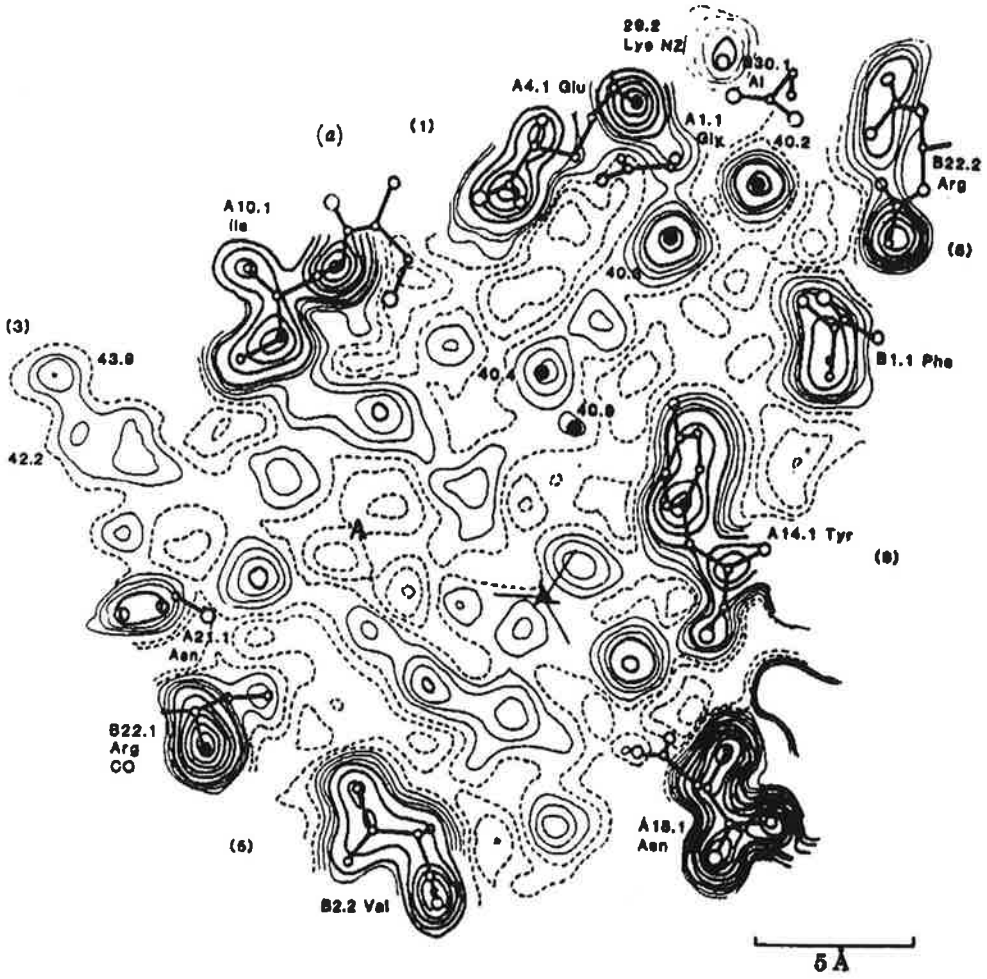


Figure 3. *Example of the solvent region between neighbouring hexamers in the structure of 2Zn insulin at 1.5Å resolution. The dotted contours are at 0, 0.1, 0.2 and 0.3 e/Å^3. Other contours at 0.4, 0.6, and 0.9 e/Å^3 and then at intervals of 0.3 e/Å^3.*

The FAST Bears Fruit

By P.J. Rizkallah

(Daresbury Laboratory and Liverpool University)

A recent three week run on Station 9.6 at Daresbury using the FAST area detector, saw the production of several crops of juicy data sets, harvested from many protein crystals. Microcrystal work has reached a new landmark as well, with the solution of a few structures that are otherwise inaccessible, and with the reduction of the required sample size to record lows, currently somewhere between 5 and 20 microns. Crystallites in the 30 micron range are dealt with routinely here.

I shall first enumerate the protein projects that got a move on:

- A data set from E. Coli matrix Porin/Iridate complex:

This project is a collaboration between P. Rizkallah (D.L.) and R. Pauptit (Basel), aimed at collecting anomalous data for Porin, a membrane protein. Some 5000 anomalous pairs have been measured, which are about a third of the total data set to 5 A resolution. More data collection is needed, and to a higher resolution, in order to start phasing. The merging statistics show an overall R-factor of 9.8% , and 8.1% with the anomalous pairs kept separate. This kind of quality, although not outstanding, has been much more difficult to attain from film data. Porin native and Iridate derivavtive data collected with the Xentronics area detector at Heidelberg compare very well with these data: Native to FAST derivative merge with an R-factor of around 18% , Native to Xentronics derivative 20% , Xentronics derivative to FAST derivative 12% . This indicates that the anomalous data are reproducibly consistent.

- A few data sets from Oligonucleotides/monopeptide complexes:

This project is a collaboration between Subirana's group in Barcelona (I. Fita and N. Verdaguer) and D.L. (P. Rizkallah and C. Nave). The nucleotides are six base pairs long, and the monopeptides are variable. The majority of these complexes are isostructural, but only one data set has been processed to date. The merging statistics are encouraging, and a solution may soon be available. Crystals of these complexes were

small in size, up to 100 microns, and other methods of data collection have not been very successful.

- *A data set from Pappaya Protease A:*

This project is run by R. Pickersgill of the AFRC Institute of Food Research at Reading. High resolution was obtained from this sample and a full data set was recorded. Preliminary processing readily yielded an orientation matrix, and reflection intensity estimation is being pursued actively. The use of the FAST for this project was aimed at gaining from its efficiency.

- *A data set from Monoclinic Beta-Lactoglobulin:*

This project is run by L. Sawyer, Edinburgh University. The crystals diffracted to 2 A resolution, and many crystals were used in the data recording. No processing has been performed as yet, and heavy atom derivatives are also planned.

- *A data set from Streptavidin:*

This data collection was made by P. MacLaughlin, MRC Unit Cambridge. The above structure is known, and the purpose was to provide a direct comparison between FAST data and data obtained otherwise. Although the processing was not fully optimised (ion-chamber corrections not incorporated), the merging statistics were very impressive, in the region of 4% .

- *Data collection from the Light Harvesting Protein complex of Rps. Acidophila:*

This project is run by M. Papiz here at Daresbury, in collaboration with a group from Glasgow. Plenty of data have been collected from crystals of this complex, and there is a publication in the press discussing results.

- *Data collection from Insulin derivatives:*

(Paul Holden, York University and DL). Several data sets have been collected, one of them to 1.2 A resolution. Structural refinement showed clear heterogeneous side chain disorder and extensive solvent networks. The crystallographic R-factor is in the region of 17% .

- *A data set from a Pt derivative of GMT:*

This project is run by P. Moody, & J. S. Cutfield, of York University. Processing is in progress.

- *A partial data set from a heavy atom derivative of the protein F1':*

Project run by R. Todd of the MRC unit in Cambridge. This data set was collected to low resolution, about 5 A, and processing is under way.

Another area where the FAST has been very useful, is Fibre diffraction. The Keele University group (W. Fuller and coworkers), are the most active in this application. Their main emphasis is on DNA fibres. C. Nave (DL) has also used this facility for recording diffraction patterns from rod virus fibres.

The small molecule microcrystal work has been very successful recently, after the not so encouraging experiences of a year earlier (See report in IQPC No. 22). Three structures are already featuring in publications that are in the making, and a few others are still being worked on:

- *Chako's Compound, a Rh - Iodine complex with a large organic component:*

This structure is a collaboration between M. Harding's group (P. Rizkallah and S. Maginn) and B. Heaton's group at Liverpool University. Crystals of this compound had been turned down by the QMC crystallographic service, and failed to give a reasonable diffraction pattern with the Laue method. A full data set to 1 A resolution was easily obtained with the FAST from a sample of dimensions 100x50x15 microns. The space group was determined to be P2/a. The structure was solved with automatic Patterson methods using SHELXS-86. Refinement with SHELX-76 reduced the crystallographic R-factor to 10.4% . A major feature of the structure is the non-symmetric substitution on the Rh atoms (Fig. 1a). The solvent molecule (methanol) bonded to a Rh atom, is probably involved in hydrogen bonding with the Cl on the neighbouring Rh, causing the cluster to hinge across the central plane of I atoms, a feature not observed in any of the related structures reported. A full packing diagram is shown in Fig. 1b.

- *Cheno Deoxy Cholic Acid (CDC), a component of human bile:*

This project is a collaboration between D.L. (P. Rizkallah), Liverpool University (M.M. Harding) and Birkbeck College (P.F. Lindley). This compound crystallises as thin fibres about 30 to 60 microns in diameter. Space group P6(5). Sealed tube and four-circle diffractometer yielded data to 1.5 A resolution only. This was easily extended to 0.9 A resolution using the FAST. Data were collected from two different crystallites. They had reasonable internal merging statistics (4.4 and 7.7% respectively), but were not in agreement with each other: Overall merging R-factor was 9.6% . The structure was solved with a novel rotation / translation search procedure in Patterson space, PATSEE, using the P2(1) structure as a model. Refinement and difference fouriers were done with CRYSTALS, and the final R-factor was 11.5% , with unit weights. The structure shows a vacant channel of dimensions larger than 12 A, where the solvent has total freedom of movement (Fig. 2). The orientation of the solvent molecule was different in each data set, which come from different crystallites. This explains the disagreement between them. The low solvent occupancy, 10% , and the variability of positions and orientation may explain some of the disorder in the plane normal to the crystallographic six fold axis. The different barrels are held together sideways by hydrogen bonding which involves all the Oxygen atoms in the structure. The weak nature of such interaction may add to the plane disorder, which is observed as streaking in a Laue photograph, thus making it impossible to process. Further data collection is needed to complete the unique set (at the moment only 25% is available), but the overall features have been well established.

- *C10BP, an organic semiconductor complex with TCNQ:*

This compound originated from the group of M.R. Willis, Nottingham University. The sample used was large and well ordered, space group P2(1)/n. It was used in order to debug the system at Daresbury of any potential problems. A data set was easily obtainable, with a very low merging R-factor of 2.7% . The structure was problematic because there is an inherent disorder in the aliphatic chain quaternising the N of the C10BP cation. The 10 atoms of the chain have half occupancy and a full freedom of motion. The consequence is a high thermal motion which renders all atoms beyond the fifth invisible. The final R-factor is in the region of 19% , but disorder modelling

and anisotropic refinement should bring it down to a reasonable level. More work is needed on this data set before it can be used for its original purpose. However, the features of the structure are well established (Fig. 3). The usual stacking habit of TCNQ molecules, in this type of complexes, is interrupted in this particular compound, generating isolated pairs of TCNQ's sharing a single electron donated from the cation. This has many implications on the electronic and magnetic properties of the complex, which makes it particularly interesting for the solid-state scientist.

- SAPO4, structure type 5, a zeolyte:

This is a collaboration between W. Kaucic of Ljubljana University, Yugoslavia, and D.L. (P. Rizkallah). This zeolyte is not different from the rest when it comes to crystal size. The sample used was 180x30x15 microns. The space group was found to be P6cc, isostructural with a related zeolyte reported by an American group. A large fraction of the data set was weak, but the unique data set contained around 38% of the total accessible. Merging R-factor is 9.8% . Structure solving is in hand.

- SAPO-44, a zeolyte:

This sample too was supplied by W. Kaucic. It is of a different structure type, Rhombohedral R32. Sample size was 20x30x30 microns, the smallest volume dealt with todate. Profile fitting has been applied to this data set, where many of the weak terms were retrieved to yield a complete data set to 0.8 A resolution. The merging R-factor of the profile fitted data is around 4% . The high symmetry of the unit-cell renders about two thirds of the data absent. The present data are all observed and above three s.d.'s. Structure solving is now in progress.

Many data sets from various other samples have been collected. Processing is now the bottleneck, and data acquisition has ceased to be the scourge of crystallographic studies.

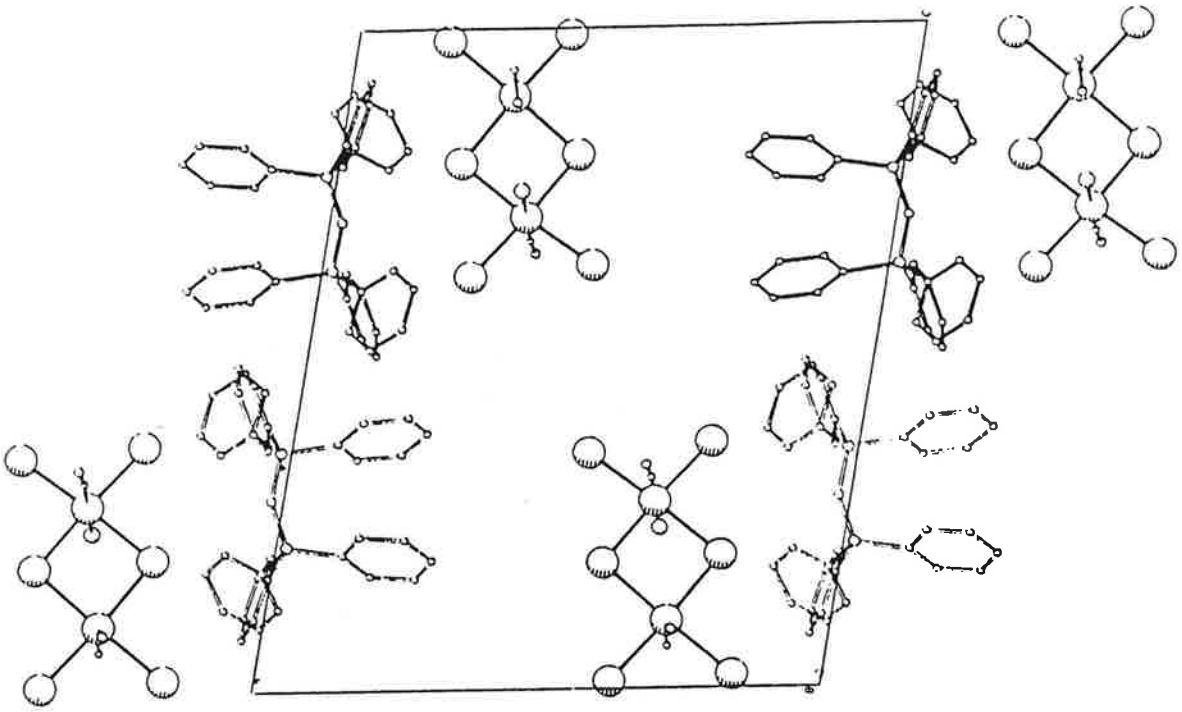
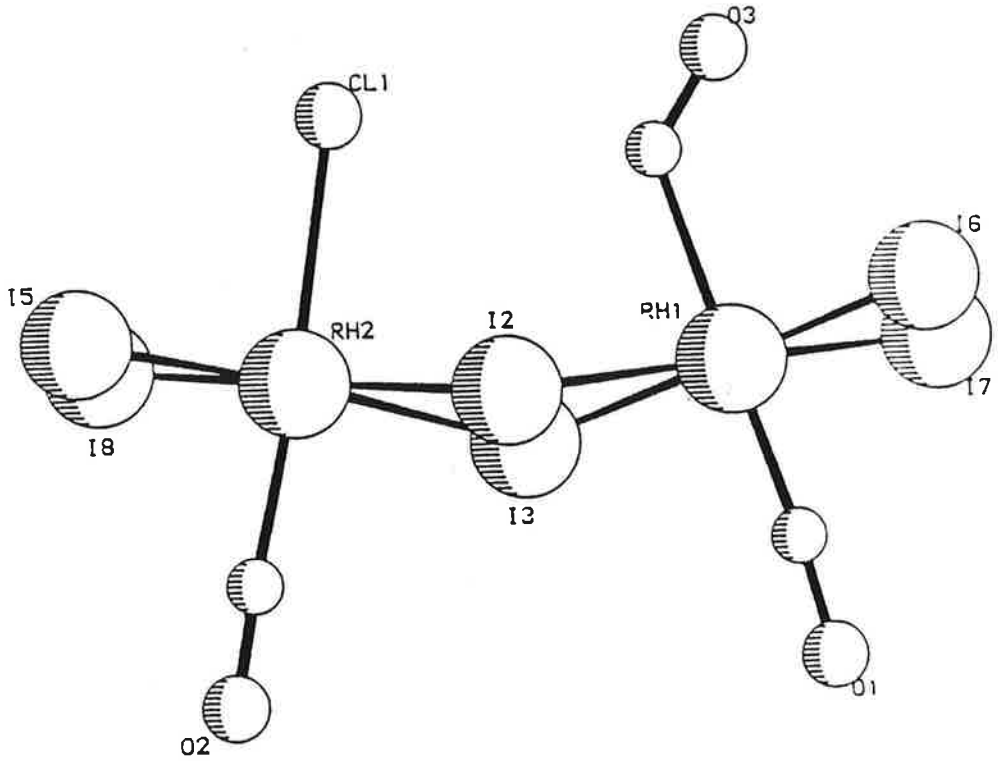


Fig. 1a The structure of the rhodium complex. Crystals are monoclinic, space group P2/a, with a = 17.53, b = 13.34, c = 21.41Å, β = 100.23°.

Fig. 1b The Sideways Interaction between the Out-of-Plane Ligands Causes Bending of the Cluster across the I-I bridge



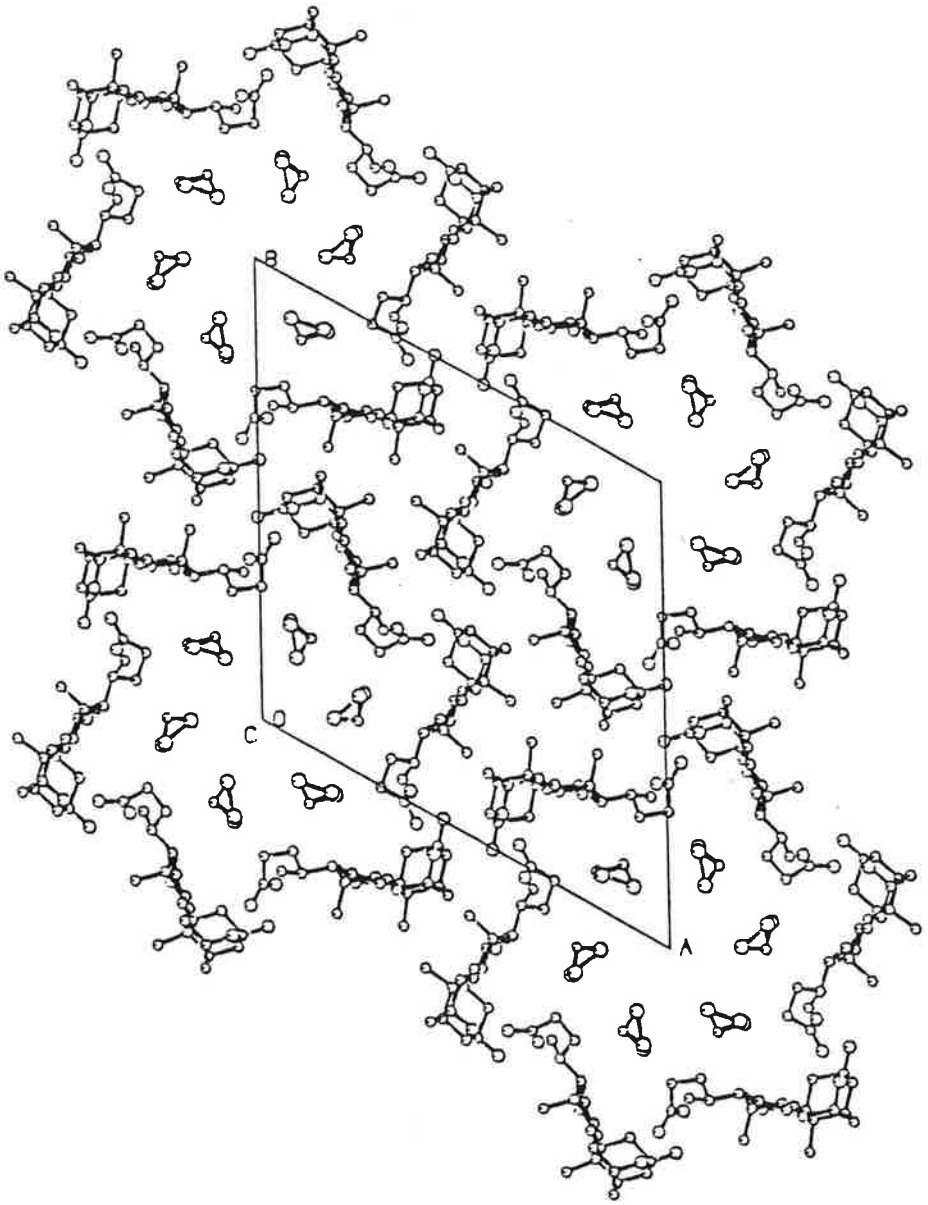


Fig. 2 Packing Diagram of CDC

Hexagonal P6(5), a=22.25, c=10.255

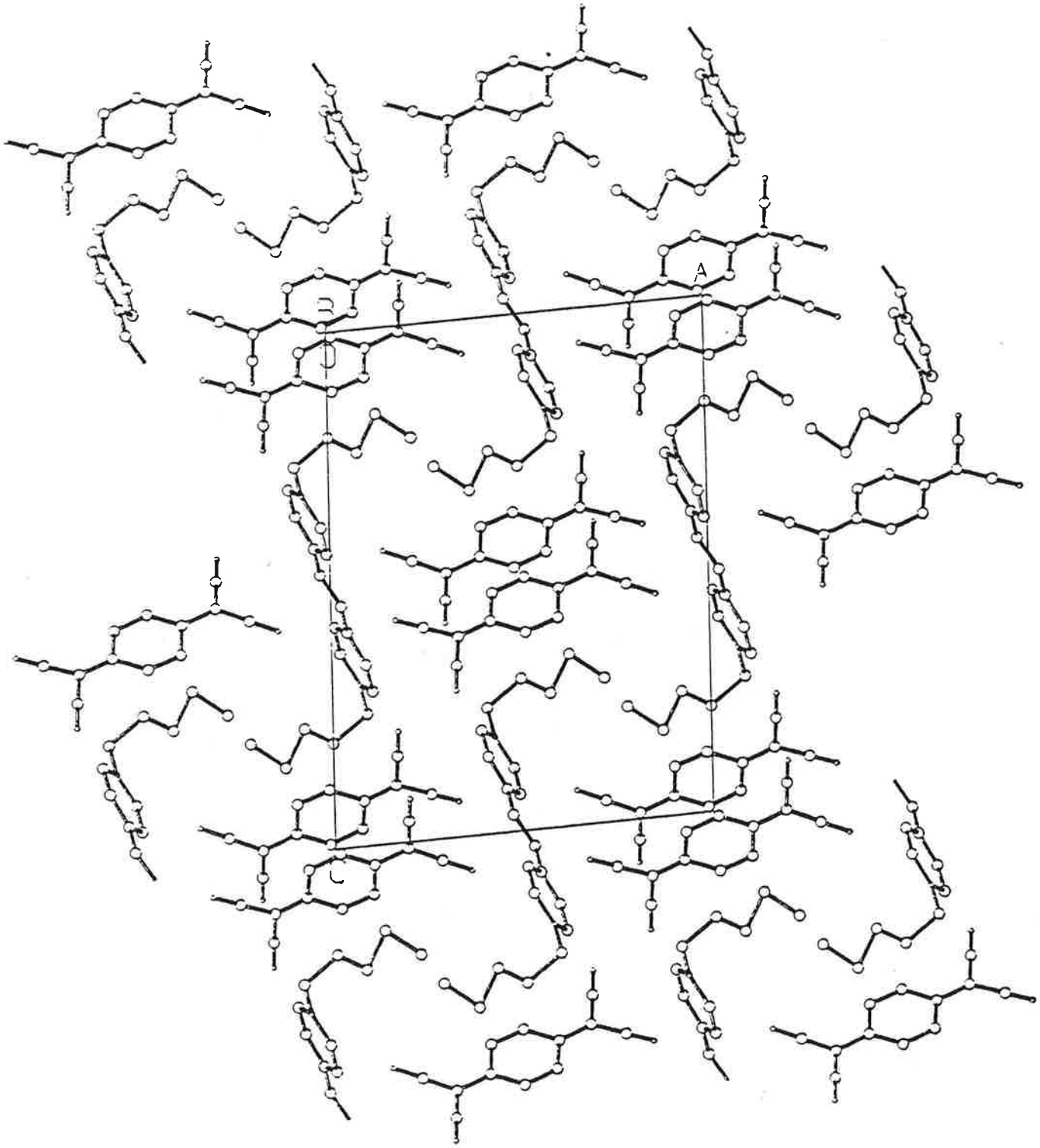


Fig. 3 Packing Diagram of C10BP.TCNQ

Monoclinic, P2(1)/n, a=13.857, b=8.485, c=18.745, beta= 94.83

CRYSTALLISATION OF OVALBUMIN: A NATIVE SERPIN

Penelope E Stein, William G Turnell, John T Finch, Andrew GW Leslie and Robin W Carrell

University Department of Haematology and MRC Laboratory of Molecular Biology, Cambridge, UK

Introduction

Ovalbumin, the major protein constituent of egg white, is a member of the serine protease inhibitor (serpin) superfamily, on the basis of primary sequence homology with the archetype of the family, α_1-antitrypsin. Ovalbumin has no known protease inhibitory activity and its biological role is unknown.

The serpins are susceptible to proteolytic cleavage in a loop of exposed peptide sequence aligned with the reactive site of α_1-antitrypsin. This cleavage is typically associated with a marked increase in stability that is believed to accompany a major structural transition from the native "stressed" conformation to a cleaved "relaxed" form. The crystallographic structure of cleaved α_1-antitrypsin has been determined to 3Å resolution (Loebermann et al., 1984), but repeated attempts to grow X-ray diffractable crystals of serpins in their native (uncleaved) form have proved unsuccessful, with the recent exception of antithrombin (Samama et al., 1987).

Recently we have shown that ovalbumin does not undergo a change in heat stability on cleavage (Biochem J, in press). Our findings are in agreement with those of Gettins (1989), who demonstrated using NMR spectroscopy that cleavage of ovalbumin is not accompanied by a large conformational change analagous to that shown by other serpins. Crystallographic studies of cleaved ovalbumin (plakalbumin) are in progress (T.Wright, unpublished). Native ovalbumin readily forms microcrystals and Miller et al. (1983) reported results of preliminary studies with crystals of native ovalbumin, although the small size of these crystals and poor reproducibility of growth made them unsuitable for further analysis.

If ovalbumin does not undergo a major conformational change on cleavage, then the crystallographic structures of either native or cleaved ovalbumin molecules would be expected to provide a model from which most of the folding common to all uncleaved serpins could be predicted. However, only in the structure of the native ovalbumin molecule might the conformation of the uncleaved "reactive site loop" be observed.

Protein Preparation and Crystallisation

We have crystallised native ovalbumin which we prepared in batches, each from a single newly-laid hens egg. Each batch of ovalbumin was further purified by ion exchange chromatography to reduce microheterogeneity arising from variations in the extent of phosphorylation of serines 68 and 344. The protein has been crystallised with its single carbohydrate side chain (at asparagine 292) intact.

Crystals of native ovalbumin have been grown from ammonium sulphate over the pH range 4.6 - 6.4, by various methods. Our largest crystals (1.5 x 0.4 x 0.3mm) were grown by the batch method at 37°C from solutions of ovalbumin 25mg/ml in 51% saturated ammonium sulphate, 40mM cacodylate buffer pH 6.4. Crystals appeared after 4-6 weeks.

X-ray diffraction

X-ray diffraction photographs showed that the crystals diffracted to at least 2.5Å resolution and precession photographs established that they were triclinic, of space group P1 (a=62.9A, b=84.7A, c=71.5A, α=87.7° β=104.0°, γ=108.5°). This crystal form was also obtained by Miller et al. (1983) using similar conditions to our own, above pH 5.8.

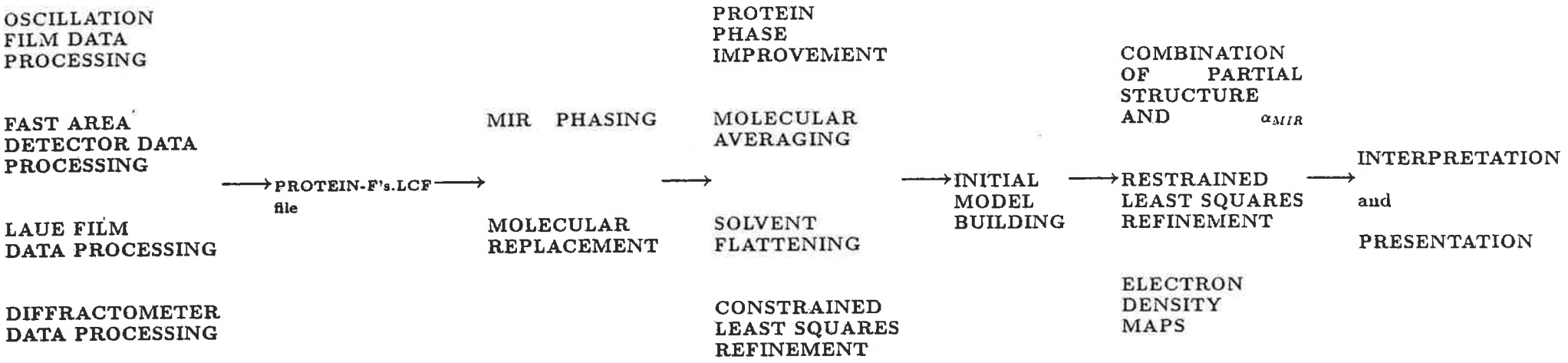
Crystal density was measured as 1.29g/cm^3 using a bromobenzene:xylene gradient column calibrated with solutions of sodium bromide. This indicates the presence of 4 ovalbumin molecules/unit cell (assuming a molecular weight for ovalbumin of 44KDa).

A native data set to 2.5Å resolution has been collected using an Enraf Nonius FAST diffractometer with rotating anode generator and the data collection program MADNES. A search for heavy atom derivatives is in progress.

References

1. Gettins P (1989) J Biol Chem 264:3781-3785
2. Loebermann H, Tokuoka R, Deisenhofer J, Huber R (1984) J Mol Biol 177:531-556
3. Miller M, Weinstein JN, Wlodawer A (1983) J Biol Chem 258:5864-5866
4. Samama JP, Delarue M, Moras D, Petitou M, Lormeau JC, Choay J (1987) Abstract 966. Thromb Haemost 58:264

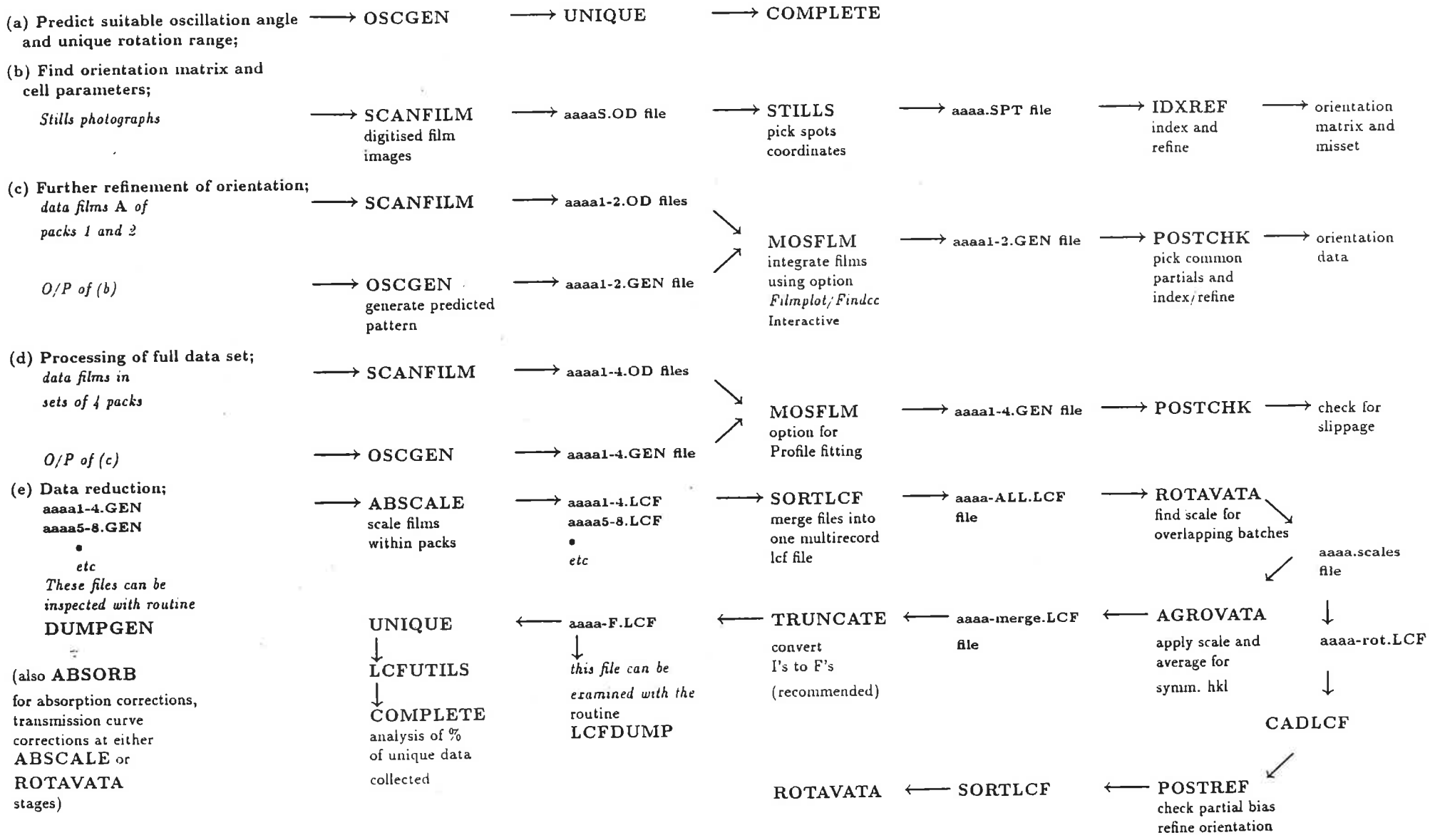
CCP4 OVERVIEW
in relation to programs for protein crystallography



FILM PROCESSING

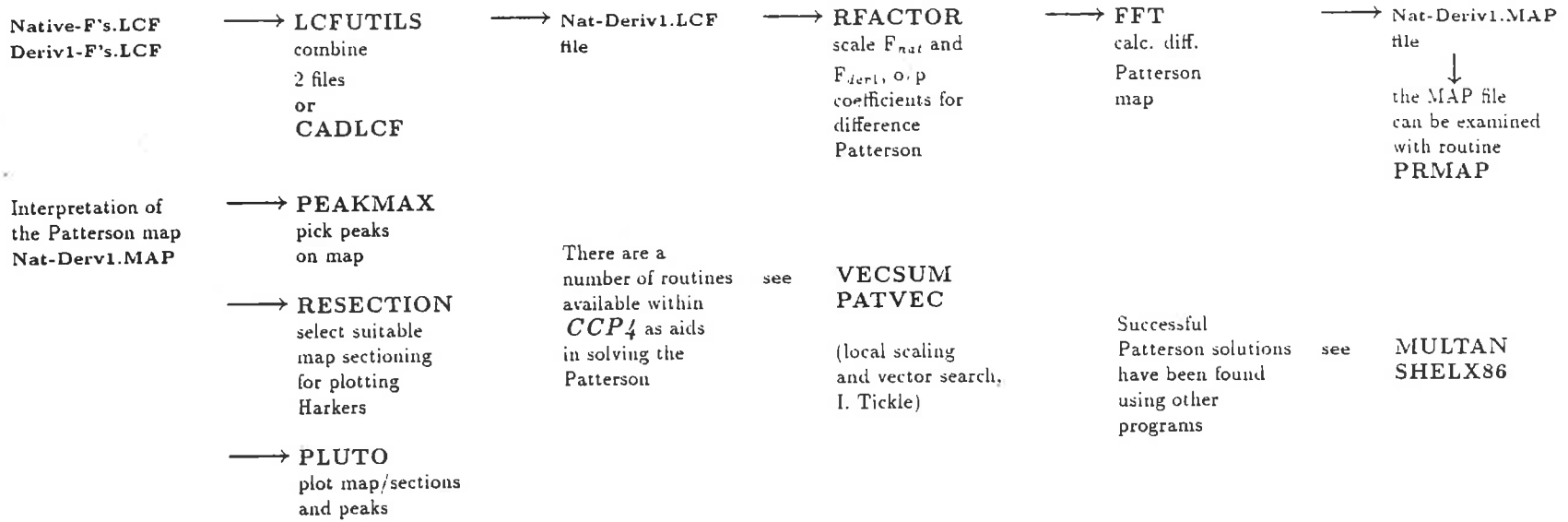
Imperial College

(see "*Information Quarterly for Protein Crystallography*", 18, July, 1986)



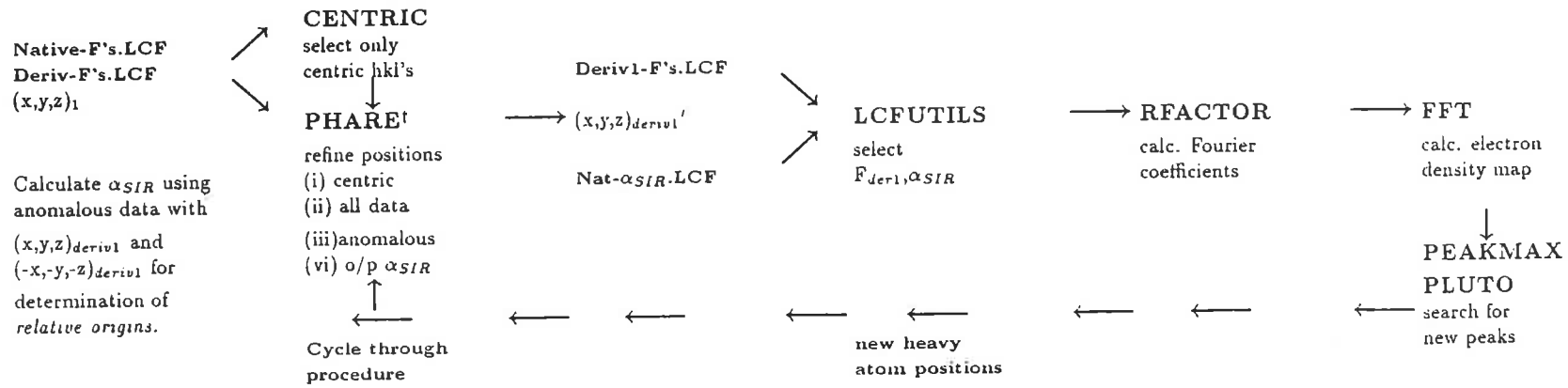
MIR PHASING

(a) Solving Patterson for first Derivative to give major site(s);



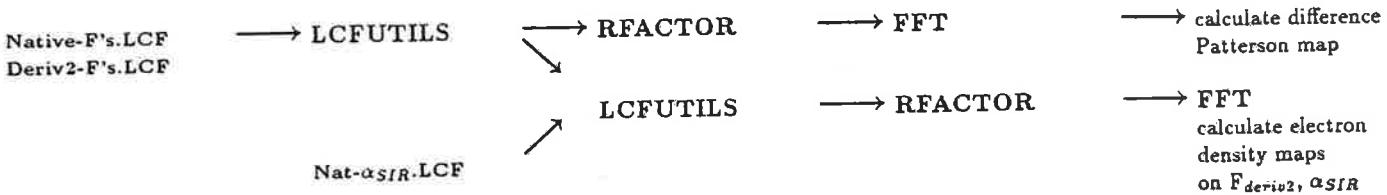
(b) SIR Phasing;

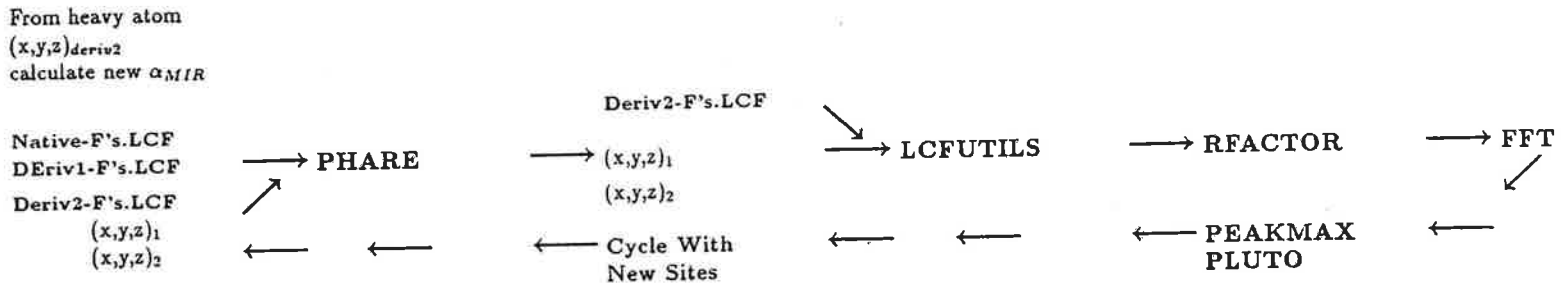
(Scaling of heavy atom derivatives from either **RFACTOR** or **DSCALE** see also **ANISOSC, ANSC, LOCAL, RSTATS** ans **SCALENEW**)



†**PHARE** program calculates phases and refines heavy atom positions; alternatives available in 2 steps, **PHASE/REFINE** this is preferred method.(see over)

(c) Use phases from 1st derivative
to solve for heavy atom positions
in subsequent data sets;



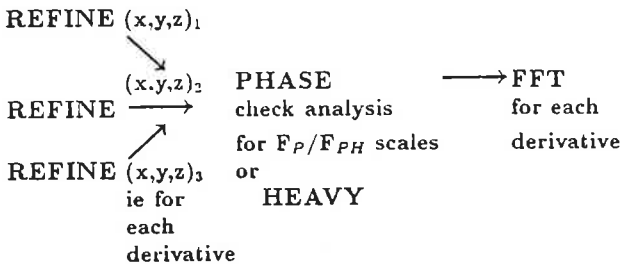


(d) Cycle through for all derivatives; (treating all derivatives together,
checking for bias of common sites.

(e) Final check for low occupancy sites,
use Residuals option in PHARE and
calculate maps with these coefficients.

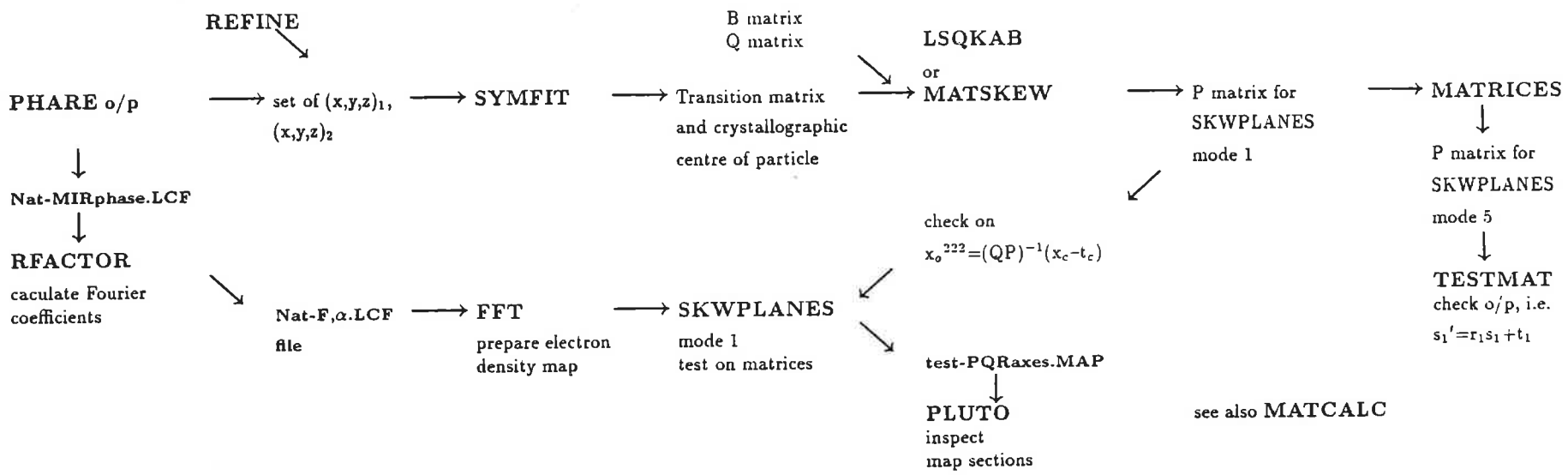
$$(|F_{PH}^{OBS}| - |F_{PH}^{CALC}|)\exp(i\alpha_{PH})$$

(f) Using programs REFINE and PHASE
(ie no cross refining)
see P.R. Evans Information Quarterly for Protein Crystallography Number 11, June 1983

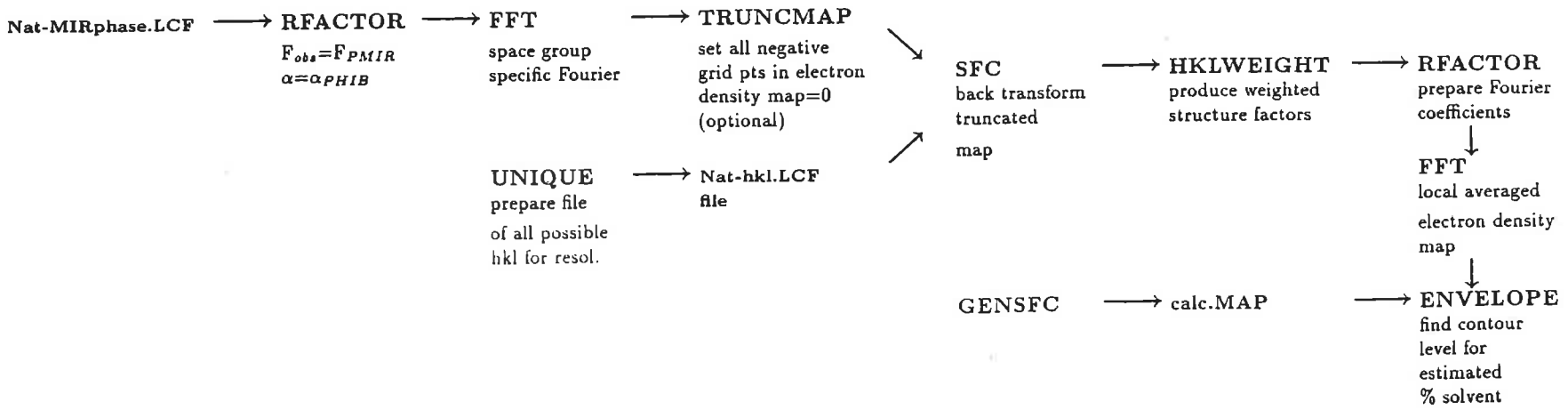


MOLECULAR AVERAGING -non crystallographic symmetry
(example of non-crystallographic 2–fold giving overall *222* symmetry

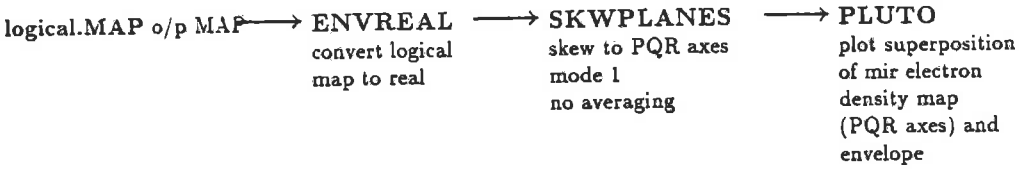
(a) Find required matrices from heavy atom coordinates;



(b) Find envelope for averaging with SKWPLANES mode 5;

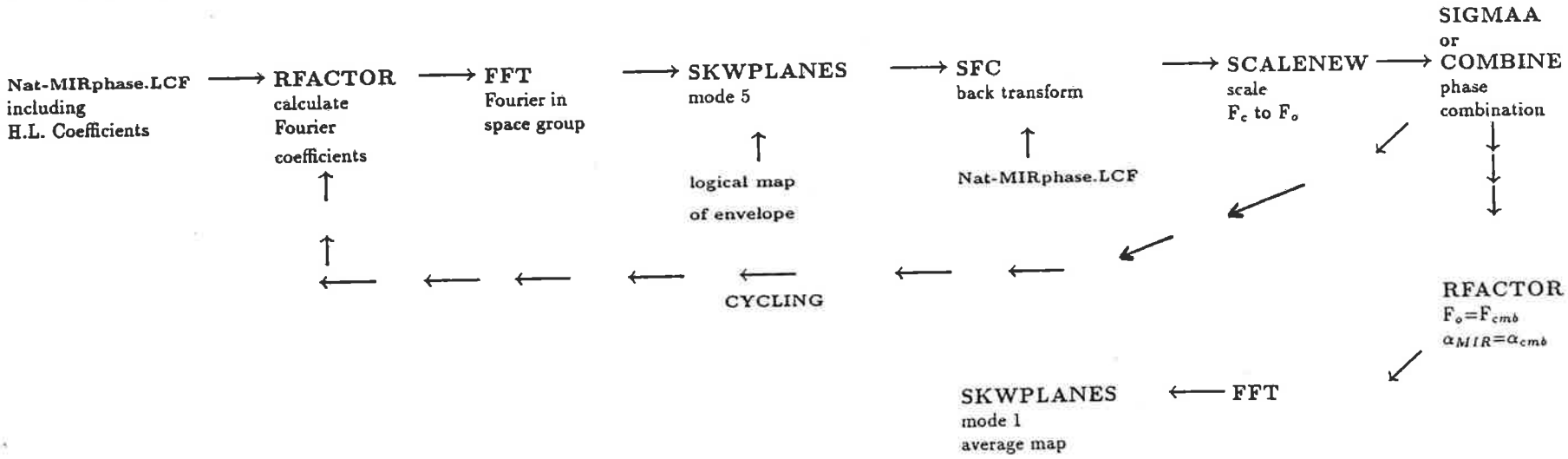


(c) Test on resulting envelope;



Alternatively, just use the local averaged density map produced above and in PLUTO use the contour level given by program ENVELOPE.

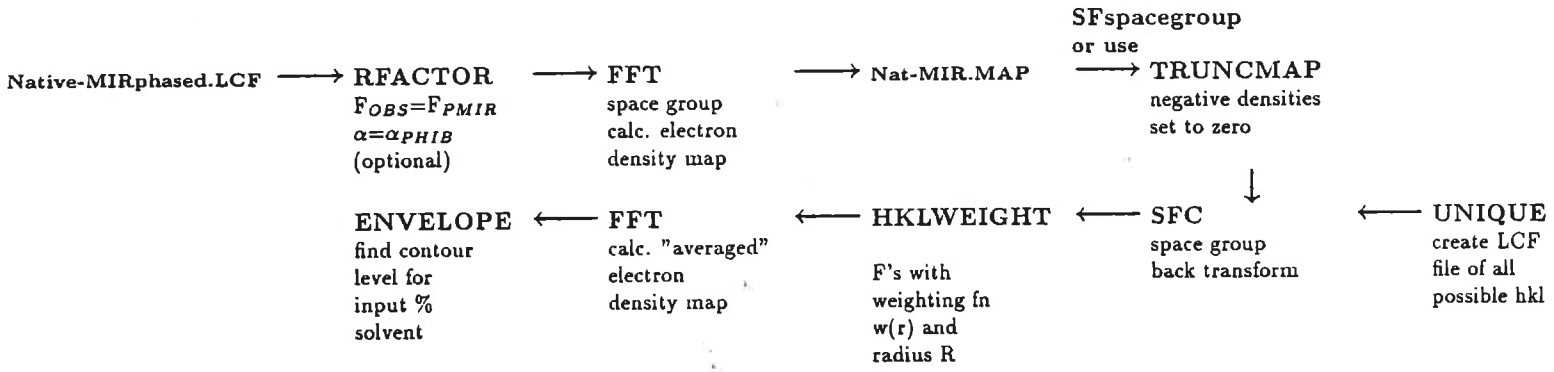
(d) Phase recombination from averaging non–crystallographic symmetry;



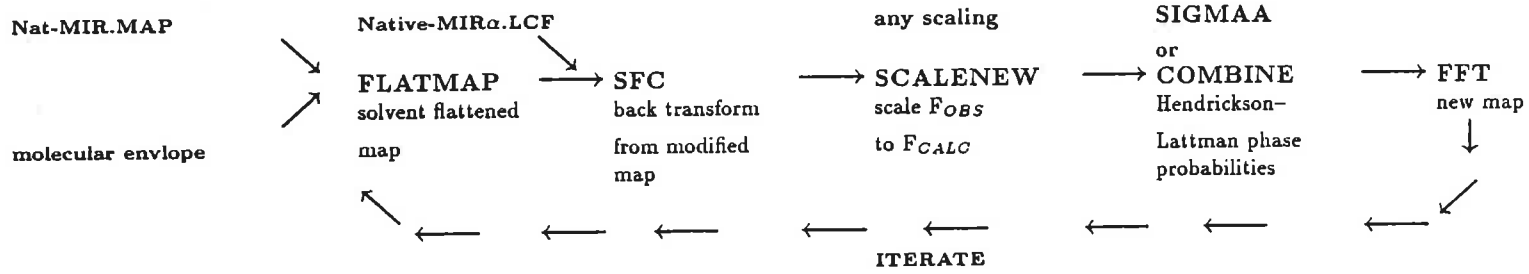
SOLVENT FLATTENING TO IMPROVE ISOMORPHOUS REPLACEMENT PHASES

Reciprocal Space version of Wang's Method,
(see B.C. Wang (1985) Diffraction Methods for Biological Macromolecules, (Wyckoff, Hirs, & Timasheff ed.) Methods in Enzymology, vol 115, Academic Press, and A.G.W. Leslie (1987) Acta Cryst., A43, 134–136)

(i) Determine Envelope;



(ii) Solvent flattening Iterations;



MOLECULAR REPLACEMENT
(see also MERLOT)

(a) Detection of non-crystallographic symmetry using Crowther's FFT Self Rotation function;

Native-F's.LCF → **ALMN** or **ROTATE** or **POLARRFN** → Run trials with different ranges in integration sphere and resolution limits → (ω,ϕ) angles for location of molecular symmetry axes → **RFPLOT** to plot 180° rotation sections as projections *no longer required*

(b) Crowther's Cross rotation function;

Model.BRK coordinates of model structure → **LIMITS or CDSLIM** find dimensions of model molecule → estimate P1 cell dimensions of model max. dims and radii of rotn. sphere → **GENSFC** calc. structure factors in P1 cell for model → Model-P1.LCF (Selected radii and resol. ranges) → **ALMN or ROTATE** (Native-F's.LCF) → (α,β,γ) rotation angles

(c) Translation function;† †Alternative is to use TFSGEN — general translation function (I. Tickle)

Model.BRK coordinates of model structure; rotation matrix from (b) → **LSQKAB** or **TRNSFMABG** Apply Euler angles to orient model → Model-Rot.BRK → **GENSFC** calc. structure factors of model in P1 cell → **PREPARE** calc relative phases for all symm hkl → **COLLATE** sort O/P LCF file from PREPARE (Native-F's.LCF) → **LCFUTILS** combine two files → **SELECT** scales F_{OBS}, F_{MODEL} → **TSEARCH** translation R-Factor → apply translation vector to model → **CORELS** constrained least squares reciprocal space refinement (Naive-F's.LCF) → **MAPSIG**

SFspacegroup → **CADLCF** → **TFSGEN**

INITIAL MODEL BUILDING

(a) It is useful to bring together all available information on the protein under study, this could include:–

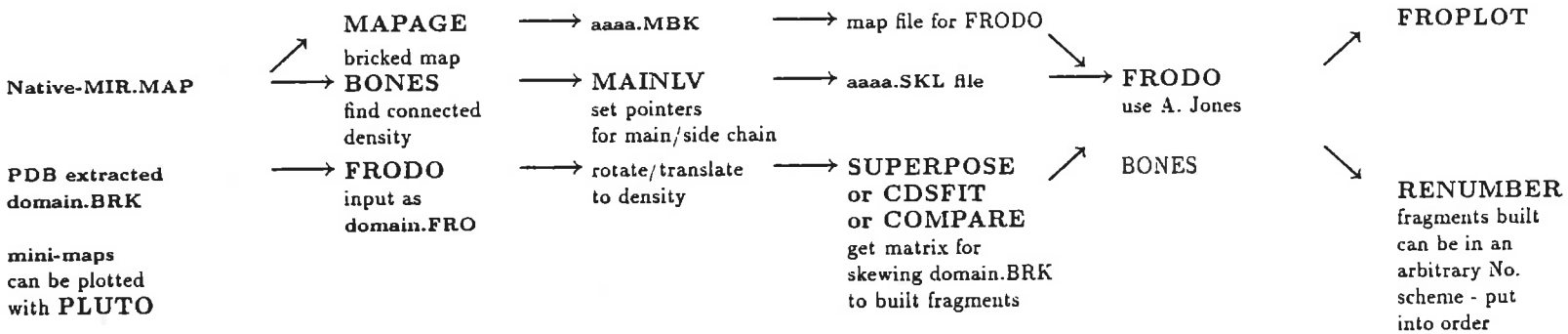
(i) All sequence information — programs are available within the Protein Identification Resource (PIR) for Searching for related sequences and Aligning of related available sequences.

(ii) Make use of available sequence information for predicting a consensus secondary structure pattern as a possible guide, see Leeds Structure Prediction library of programs.

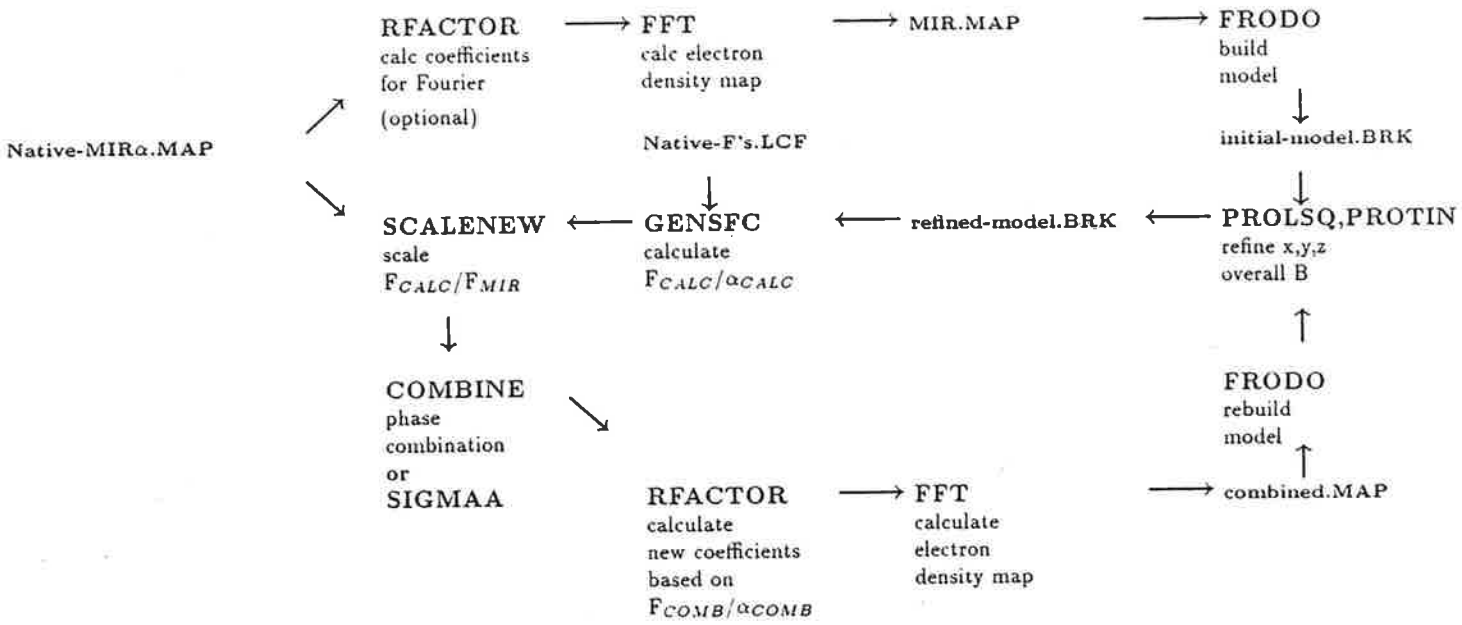
(iii) Use Protein Data Base files to extract fragments of expected secondary structure to act as mol object templates for model building, e.g. EF-hand calcium binding domains or the nucleotide binding domain

(iv) Input heavy atom positions into FRODO as a metal.FRO file as an aid to locate possible Cys or Met residues, also use known or predicted S—S bridge positions.

(b) Using FRODO; (programs RESECTION or SKEWPLANES may be required

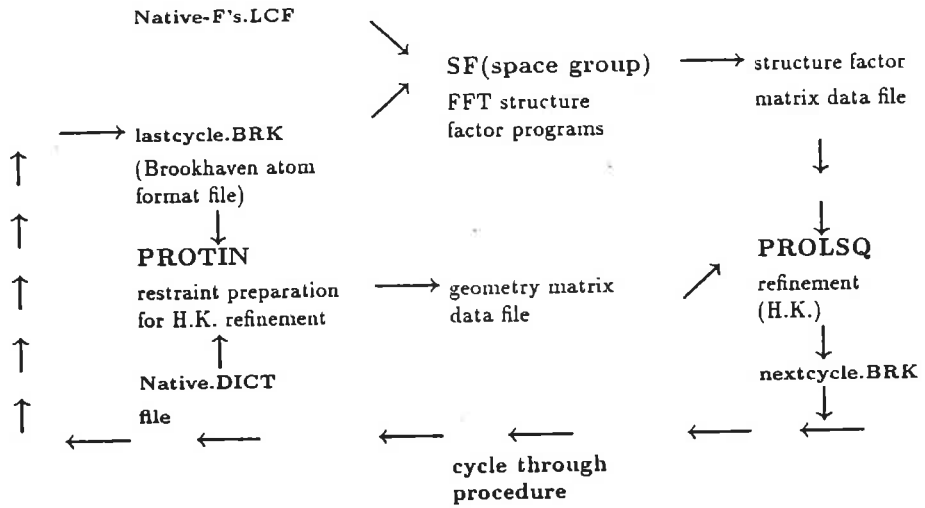


COMBINATION OF PARTIAL STRUCTURE AND ISOMORPHOUS PHASE INFORMATION

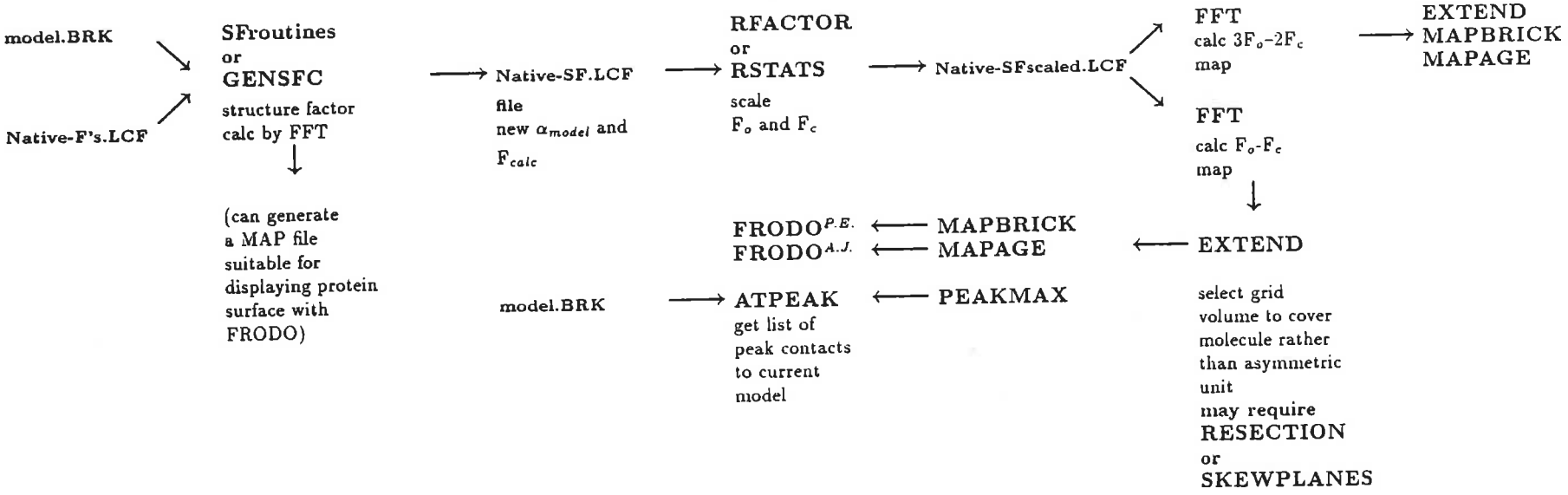


REFINEMENT TO DIFFERENCE ELECTRON DENSITY MAPS

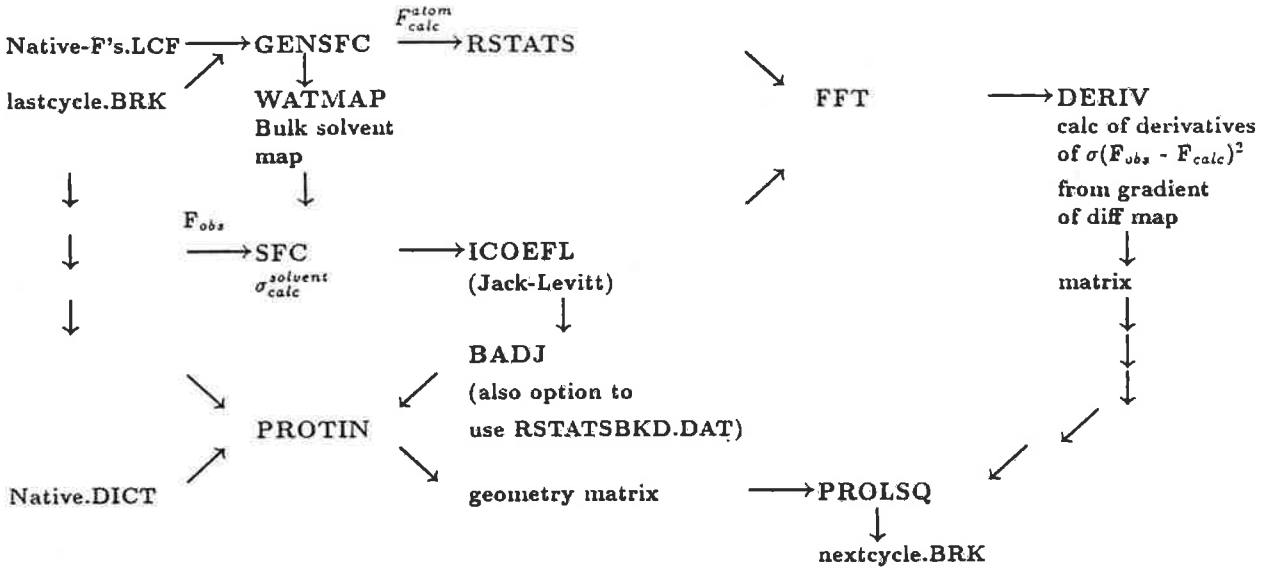
(a) Cycles of restrained least squares structure factor refinement (Hendrickson–Konnert)



(b) Difference map generation.



(c) **Alternative Refinement Procedure:**



INTERPRETATION AND MANPULATION OF COORDINATES

(a) BROOKHAVEN Files;

(i) BRKSORT	⟶	sort a brookhaven file into standard order for atoms within a residue required for LESK diagrams as this program requires DIAMOND I/P format atoms
(ii) BRK-TO-DIA	⟶	convert Brookhaven to Diamond format atom file
(iii) BRK-TO-TNT	⟶	convert Brookhaven file to 2 files, SEQUENCE and COORDINATES for TNT suite
(iv) BADDBRK	⟶	change temperature factors on BRK file
(v) CHAINBRK	⟶	change CHAIN IDENTIFIER on BRK file
(vi) LIMITS	⟶	find max and min in coordinates of a BRK file
(vii) SUPERPOSE	⟶	give matrix for best superposition of two sets of C_α BRK files
(ix) MPLANE	⟶	fit least squares planes to sets of atoms
(x) WATERSORT	⟶	renumber waters to match different Subunit Chains
(xi) CONTACT	⟶	list selected interatomic distances either intermolecular or intramolecular, analysis of solvent hydrogen bonding
(xii) DIFRES	⟶	compare either coordinates and/or temperature factors between different Subunits or different cycles of refinement or SUPERPOSE/SKEWED different sets of coordinates
(xiii) DSSP	⟶	determine H-Bonds and on this basis give regions of Secondary structure in a refined structure
(xiv) SURFACE	⟶	calculate degree of solvent accessibility
(xv) TORSION	⟶	calculate dihedral angles and plot a Ramachandran
(xvi) DISTANG	⟶	calculate sets of distances and angles

PICTORIAL PRESENTATION OF RESULTS

FRODO (A. Jones and P. Evans) ⟶ photographs of screen, ribbon diagrams

FROPLOT screen plotting to hardcopy device

LESK (A. Lesk) ⟶ "Arrow and Cylinder" secondary sructure diagrams

RIBBON (J. Priestle) ⟶ "Jane Richardson" secondary structure diagrams

SCHAKAL (E. Keller) ⟶ Space Filling Molecular Modelling Program

PLUTO ⟶ General plotting procedure for maps, stereoviews atoms etc

ORTEP2 (C.K. Johnson) ⟶ General atom plotting program superior to PLUTO for bond overlap