

DARESBUY LABORATORY

INFORMATION QUARTERLY

for

PROTEIN CRYSTALLOGRAPHY

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

Number 10

February 1983

Contents

Editorial	1
Oscillation Camera Data Processing	3
PERQ News	9
The Distributed Array Processor and the Future of Molecular Computing	11
Nucleic Acid Sequence Data	17
Current State of the CCP Program Suite	19

Editor: Pella Machin

Science 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

Editor's Note

We take this opportunity to circulate an up to date list of programs available through the CCP and to include some information on PERQ and DAP developments. Trevor Greenhough's report on synchrotron geometry corrections in data processing is also very timely, since most groups are currently working with oscillation films from the synchrotron.

University contributions are rather thin this time, but I hope this is only because everyone's effort has gone into the write up of the Bristol meeting on "Location and Refinement of Heavy Atom Sites" (held September 1982)! This write up should appear as the next newsletter, probably in a few weeks time.

OSCILLATION CAMERA DATA PROCESSING

Trevor Greenhough (Daresbury)

The implementation of the MOSCO suite of on-line film processing programs on the VAX at Daresbury Laboratory has been completed and several datasets have been successfully processed. Enhancements to the SCANDIG-3 and the Data General Nova 3/12 have been carried out in order to provide a full on-line scanning/processing service in the near future. SRS users are encouraged to make full use of these services and all enquiries will be welcomed.

The inclusion of synchrotron geometry (Greenhough and Helliwell, J. Appl. Cryst. (1982) 15, 493-508) into the prediction and refinement routines in MOSCO has led to major improvements in the processing of SRS film data. In some cases post-refinement via POSTCHK proved to be impossible or meaningless without the correct SR geometry, while in all cases the predictions are greatly improved. Furthermore, in the majority of cases a refinement of mosaic spread has given a narrowly defined value leading to much better prediction and post-refinement.

The synchrotron geometry is included in data processing by replacing the conventional reciprocal lattice volume radius:

$$\text{eps}_c = \frac{d^* \cos \theta}{2} \left[\eta + \gamma + \left(\frac{\delta \lambda}{\lambda} \right) \tan \theta \right]$$

where η is mosaic spread and λ beam crossfire (Greenhough and Helliwell, J. Appl. Cryst. (1982) 15, 338-351) by:

$$\text{eps} = \frac{1}{2} \left[(\delta d^*^2 + z \gamma_H)^2 + \gamma_V^2 y^2 \right]^{\frac{1}{2}} + \frac{d^* \cos \theta}{2} \left[\eta + \left(\frac{\delta \lambda}{\lambda} \right) \tan \theta \right]$$

where δ is the correlated component of spectral dispersion, γ_H and γ_V are the horizontal and vertical cross fire angles, $\delta \lambda / \lambda$ is the conventional type spectral dispersion and y and z are the crystal axis coordinates perpendicular to the beam with z the rotation axis.

Stills refinement where all partialities are assumed to be $\frac{1}{2}$ is of course unaffected by the size or shape of the rocking curve. Post refinement however depends on both functions; using only reflections nearly $\frac{1}{2}$ recorded reduces the dependence on the shape of the rocking curve but increases the uncertainty about the width. Since the majority of SRS data is collected near the Guinier position for 1.488 Å the most important factor to be considered is the highly asymmetric beam cross fire. A small component of the correlated spectral dispersion term is however present; in the case of optimised anomalous dispersion experiments a large contribution often needs to be accounted for.

The usual SRS set up yields:

$$\begin{aligned}\gamma_H &= 0.201^\circ \\ \gamma_V &= 0.013^\circ \\ (\delta\lambda/\lambda)_{\text{conv}} &= 0.00151 \\ \delta &= 0.00043\end{aligned}$$

but users should check for changes which occur from time to time.

The Manganese superoxide dismutase data (Colin Blake, Mike Parker; Oxford) provides a good example of the improvements gained with the use of synchrotron geometry during data processing. Collected at the Mn edge and very near to the Guinier position, this data is currently being processed on the VAX at Daresbury. The crystals seem to suffer quickly from radiation damage in the early exposures, and this is reflected in a steep rise in the value of the mosaic spread parameter required for good prediction.

Using a conventional geometry description of mosaic spread and beam cross fire and cycling through prediction, integration and post refinement for the first two packs leads to a 'refined' value of 0.27° for $(\eta + \gamma)$. The refinement is carried out by varying the value of $(\eta + \gamma)$ in repeated runs of IDXREF and plotting this value against the minimized quantity

$$\sum_i ((R_o - R_c)/d^*)^2$$

Convergence is considered to have been achieved when two successive runs of prediction, integration and post refinement yield the same misorientation angles and the same minimum residual in IDXREF for the same value of $(\eta + \gamma)$.

The process was repeated for the same two packs using the correct SR geometry for this case (asymmetric cross fire $\gamma_H = 0.15^\circ$, $\gamma_V = 0.013^\circ$) and a refined value of the mosaic spread η of 0.185° was obtained, as shown in the figure.

Table

	<u>Conventional Geometry</u>	<u>SR Geometry</u>
$\eta + \gamma$ (refined)	0.27°	-
η (refined)	-	0.185°
γ_H	-	0.15°
γ_V	-	0.013°
$(\delta\lambda/\lambda)_{\text{conv}}$	2.5×10^{-3}	0.77×10^{-3}
Prediction: Pack 1 Full	492	671
Partial	989	632
Pack 2 Full	492	672
Partial	984	613
Post refinement: NREF	72	56
In range	46	41
Out of range	26	15
$\sum_i ((R_o - R_c)/d^*)^2$	5.2×10^{-7}	4.0×10^{-7}
r.m.s. residual($^\circ$)	0.008	0.007
ϕ_x	0.286°	0.286°
ϕ_y	-0.672°	-0.672°
ϕ_z	0.503°	0.514°

MnSOD data processing. Vee Cassette data to $\sim 2.2 \text{ \AA}$ resolution at $\lambda = 1.89 \text{ \AA}$

A comparison of the two results is given in the table which shows that there are large differences between the two methods in crucial areas. While the final values of the misorientation angle ϕ_z differs by 0.01° , the prediction of fulls and partials differs dramatically. It was clearly seen that the conventional prediction gave far too many partials towards the vertical axis and too few as the cusp was approached. The $(\eta + \gamma)$ value of 0.27° is a compromise between the $(\eta + 0.013)$ needed in the vertical and the $(\eta + 0.15)$ needed towards the horizontal. This is a simplified view since the situation

is somewhat complicated by spectral dispersion, but the value of η obtained from the synchrotron geometry refinement gives a good prediction over the whole film. The somewhat sudden appearance of split spots on the data films in later packs is reflected by a sharp rise in the refined value of η to around 0.35° .

Other datasets which particularly benefit from SR geometry are bacterial 6-PGDH (Carr/Helliwell; Keele) which does not respond to post refinement in a sensible way with the conventional geometry, and native pea lectin (Papiz/Suddath; Keele/UAB) which gives a better post refinement with SR geometry. Obviously all datasets will benefit from much improved prediction.

Future improvements to the data processing facilities at Daresbury include the calculation of individual spot shapes and sizes for integration as outlined in Greenhough, Helliwell and Rule (J. Appl. Cryst., in press). This is particularly important for high resolution data collected on flat films, due to the spot inclinations introduced by the asymmetric SR cross fire.

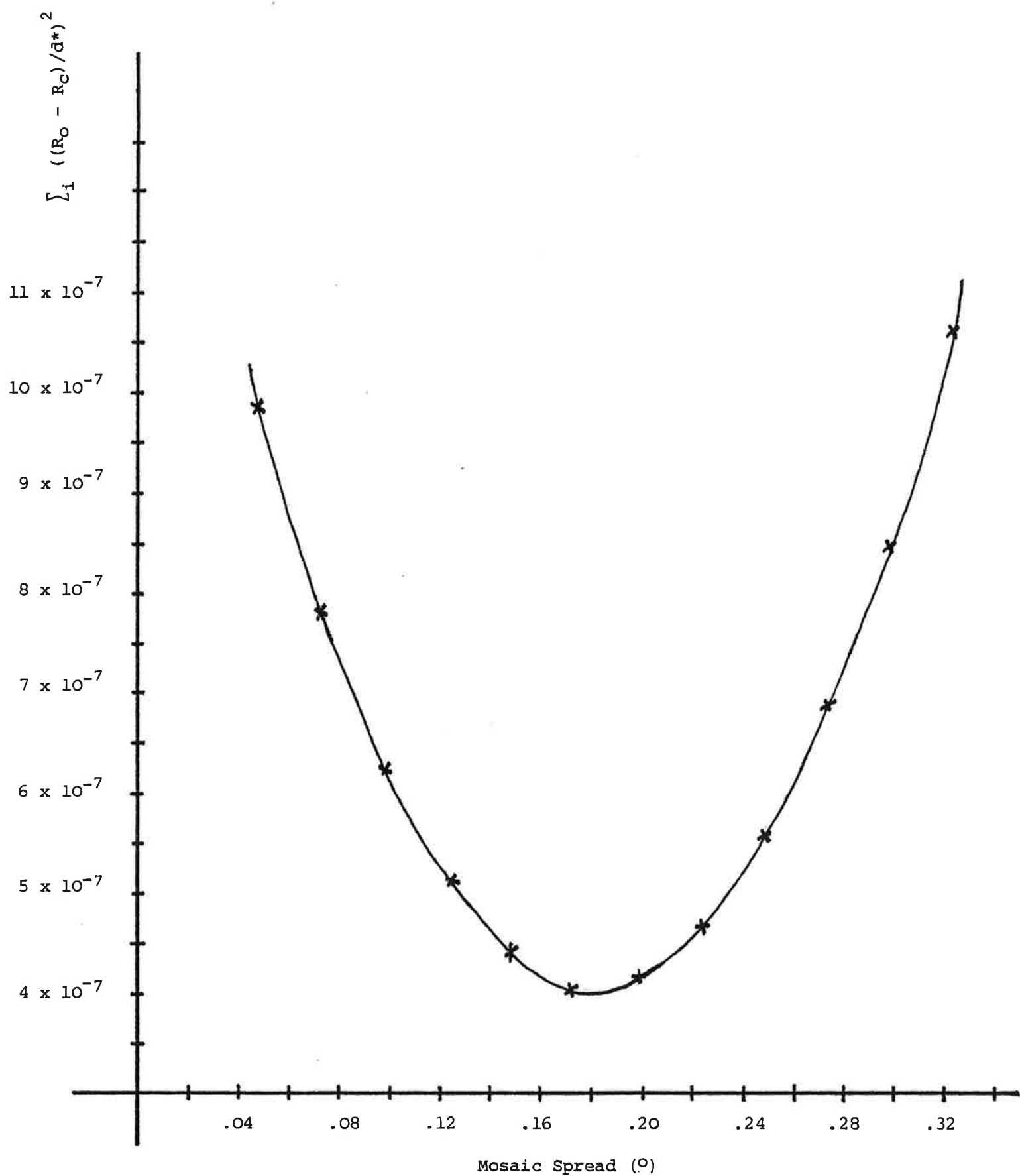


Figure: Mosaic Spread refinement for MnSOD using SR geometry with $\gamma_H = 0.15^\circ$, $\gamma_V = 0.013^\circ$, $(\delta\lambda/\lambda)_{\text{conv}} = 0.77 \times 10^{-3}$ in POSTCHK and IDXREF

PERQ NEWS

Pella Machin (Daresbury Laboratory)

PERQ computers are appearing all round the UK! Several protein crystallography groups have SERC funded PERQs (York, Birkbeck, Daresbury) and Computer Board funded PERQs are available in most university computer centres.

Guy Dodson convened a meeting of protein PERQ users on 17.1.83 at York University to stimulate an exchange of PERQ news and to plan and coordinate future work in order to avoid any duplication of effort. Representatives from York, Birkbeck College and Daresbury attended. The groups are already exchanging programs via floppy disks and the programs which are already available on the PERQ (or are under development) are listed below. (* = working now.)

*DISTAN	Rod Hubbard (York)
*MODELFIT	" " "
Protein PLUTO	David Vallely (York)
(Note: DV's version will be extended to include map contouring)	
*IDXREF/STILLS	Pella Machin (Daresbury)
GENVEE	Garry Taylor (Oxford)
*Small molecule PLUTO	Pella Machin (Daresbury)
*SHELX	Mike Elder (Daresbury)
FFT	Pella Machin, John Campbell (Daresbury)
LCF formats	" " "
ORTEP	Lawrence Pearl (Birkbeck)
Diffraction data processing	" " "

All groups are interested in developing programs utilising the graphics capabilities of the PERQ and it was agreed that a regular exchange of news on any progress on these activities was important.

PERQ computers were discussed further at the CCP Working Group meeting in February and David Fincham detailed his plans for the PERQ mini DAP (see his article for general information on DAP's).

THE DISTRIBUTED ARRAY PROCESSOR AND THE FUTURE OF
MOLECULAR COMPUTING

David Fincham, Queen Mary College

Introduction

The development of powerful computers has taken two distinct directions in recent years. One route has been to speed up the central processor unit of an otherwise fairly conventional computer architecture by the use of advanced technology, the provision of vector instructions and the pipelining of arithmetic operations. This is the approach of Cray and CDC and is familiar to members of CCP4.

An alternative is to move away entirely from the concept of a single central processor and instead to have a large number of simpler processors working in parallel (i.e. simultaneously). There are obvious advantages in this type of architecture, principally that the use of mass-produced chips makes it a much more cost-effective way of producing computational power by at least an order of magnitude. It is also possible to integrate the processors into the memory of the computer and so overcome the communications bottleneck between separate memory and processor modules. Furthermore it is easy to upgrade such a machine by adding more processors or replacing them with faster ones as integrated circuit technology advances, whereas vector processors are already employing state-of-the-art technology and it is becoming increasingly difficult to make them go any faster.

For these reasons it is widely believed that the future of computing lies with multi-processor architectures, and ICL with its Distributed Array Processor (DAP) is many years ahead of the world in this field.

The DAP is a true parallel processor, but machines such as the Cray, because of their pipelined arithmetic, have characteristics which are in some respects similar. Both types of machine require a rethinking of problems and algorithms in order for them to be used most effectively. However the Cray, in addition to its vector capabilities, includes a powerful scalar processor and this has provided an evolutionary route for program development which has been popular with users and helped in its commercial success.

The DAP being a more radical departure from traditional computer architectures does not allow such an evolutionary development. It has therefore appealed to the more adventurous users who are willing to put effort into rethinking their problem and learning to program in an enhanced version of Fortran. For university users there is a national DAP service based at Queen Mary College and a DAP Support Unit to provide assistance and basic software. A very wide range of applications are running on this machine; users find it gives very good performance and furthermore the features available in DAP Fortran make it easy and indeed fun to program.

For largely historical reasons the DAP has not been widely used by the Collaborative Computational Projects, though this is changing. However there are exciting developments in progress which will, I believe make the DAP type of computer of immense value to everyone involved in CCP4. My reason for writing this article is to alert you to these possibilities and I hope to interest you in beginning to explore what the DAP can do for you.

The DAP and DAP Fortran

The DAP at QMC has 4096 simple processing elements (PEs). Under the control of a Master Control Unit they perform the same operations simultaneously on their own data; each PE has 16K bits of memory (making 8 Mbytes in total). An important facility is the ability by the use of logical masks in the program to in effect switch off some of the processors for a given operation. For some applications the processors may be considered as being connected as a 64 x 64 square array, usually known as the PE plane.

The programming language used is an enhanced version of Fortran. Conventional Fortran only has scalar quantities and operations on vectors and matrices are performed by DO loops. Though this is possible on the DAP its performance arises from the possibility of defining objects which are said to have vector or matrix mode. Vectors have 64 elements, matrices are 64 x 64 (They can also be regarded as long vectors of 4096 elements). For example, in the following code

```
REAL A(,), B(,), C(,)
.
.
.
A = B+C
```

A, B and C are matrices, and the element-by-element addition of these is performed by a single statement.

The logical masks mentioned above are used in place of subscripts to index a part of a vector or matrix. For example the statement

$$A(B.LT.2.0) = B$$

will set the matrix A equal to the matrix B except that the assignment will be made only for those elements which are less than 2. Other elements will retain a previously defined value.

The language has efficient built-in functions operating on scalar, vector or matrix arguments, for interconverting between them, and, particularly important, for performing summations. For example

$$COUNT = SUM(B.LT.2.0)$$

simply adds up all the true values of the logical matrix argument of the SUM function.

Because the individual PEs only perform single bit arithmetic all other arithmetic is provided by software. Thus it is possible in DAP Fortran for scalar, vector and matrix variables to have any precision required from one bit upwards. Applications which can use low precision arithmetic, e.g. image processing, perform very rapidly (an order of magnitude faster than the Cray). For general floating point work the performance of the DAP lies somewhere between that of the CDC7600 and the Cray. However, in evaluating performance there are some surprises in store: for example, square root is faster than multiply!

There are other implications of the use of single bit processors. Take for example the formation of a histogram, a notorious example of a process which does not 'vectorise'. On the DAP it is possible to find which values fall within a certain range by examining them all in parallel. Usually only the most significant few bits need to be examined so this is a very rapid operation. In this type of approach the DAP is essentially being used as a content-addressable memory.

The array processing enhancements discussed above are extremely useful in simplifying programming and for this reason similar facilities will appear in the next Fortran standard. However, the new standard will of course allow arbitrary sized objects to be handled, whereas DAP matrices are of fixed size. For the time being the programmer has to cope with this by

mapping his problem in some way onto the PE plane. This is not difficult as masking can be used for objects of smaller size, and larger objects can be handled in blocks by using DO loops. Present users have been willing to do this because the DAP is the only machine which has any higher level language features at all, or because it is the only machine that can solve their problem at acceptable cost. However there is no doubt that many more people would consider this effort to be worthwhile if DAPs were more widely available.

The future

The present DAP is integrated very closely into a host computer which must be an ICL 2900 series mainframe. In fact the DAP is a specialised store module of such a computer, and the host computer sees it simply as memory. At the language level a program contains a host section in Fortran and a DAP section in DAP Fortran; they communicate by means of COMMON areas which are physically located in the DAP memory. The DAP has provided a very cheap method of upgrading an existing mainframe to supercomputer power and a number of sites have taken advantage of this. However its potential market has been limited to such sites and its cheapness has not therefore resulted in a mass market success.

In recent years an important phenomenon in the computing world has been the development of the micro-computer into the powerful professional workstation such as the Perq. More and more users are beginning to find that the most effective way of computing is to employ the high resolution graphics and fully interactive working which can be provided in a single-user environment.

Bearing these points in mind the expressed interest by ICL in attaching a DAP to the Perq is very exciting. Not only would this produce an enormous increase in computational power at modest cost. It should be clear from the previous description that the DAP, with its ability to process arrays of short fixed-point numbers at high speed in parallel, is potentially a very powerful device for raster-graphics processing. This must be the direction in which computing is going.

A machine integrating high performance graphics with number crunching would clearly be of great value to the work of CCP4. It is not necessary to wait until the hardware is available before thought is given to applications software, nor would it be desirable to do so. Because of the present

availability of Perqs and the mainframe DAP development work could start now. I am beginning to investigate possible applications in the general area of molecular science including the type of work covered by CCP4.

If you are at all interested in participating in this process, or would like more information about the DAP, I would like to hear from you. My address is DAP Support Unit, Queen Mary College, Mile End Road, London, E1 4NS, telephone 01-980-4811 Ext. 219. I hope I have convinced you that any effort expended at this stage will be a valuable investment for the future.

NUCLEIC ACID SEQUENCE DATA

Pella Machin (Daresbury Laboratory)

There is much interest in nucleic acid sequence data and in the development of a data base of such sequences. SERC asked me to investigate what data are currently available and how university scientists can access them. I report my initial findings here.

A nucleotide sequence data library has been established at the EMBL (G Hamm, Heidelberg) to collect, organise and make freely available the published body of nucleic acid sequence data. The initial data base contains nearly 600,000 nucleotides comprising some 600 separate entries. Each entry includes source, reported sequence features, sequence length and base composition, taxonomic classification of the organism and literature references. The library is designed to be read by both people and computer programs! The data base is available now on magnetic tape and I receive the associated newsletter for information.

There is a similar data base in America in the Los Alamos Laboratory and it has recently received funding from NIH. EMBL are working in close collaboration with this group and have agreed to exchange data collections. This means that the two collections should ultimately contain the same data, albeit in slightly different formats.

In the UK, Dr Olga Kennard (Cambridge) will be involved in the distribution of this data base. Dr Kennard writes: "the MRC has supplemented my program grant support to establish a centralised UK Facility which is to maintain a national network for the dissemination of the information accumulated in the EMBL Nucleic Acid Sequence Library and the free availability of programs for search and analysis. It is proposed in the first instance to mount the Library on the Cambridge University's IBM 3081 computer and to use the search program developed by Dr R Staden at the MRC Laboratory for Molecular Biology."

Finally I should report that we have at Daresbury a copy of Richard Feldmann's program NASSD - Nucleic Acid Sequence Synthesis and Display. This suite of programs is DEC system 10 specific (written in SAIL) and could be made available (on the ERCC DEC 10) via the SERC computer network, if sufficient interest is indicated, so let me know if you want to use it.

CURRENT STATE OF THE CCP PROGRAM SUITE

John W Campbell (Daresbury Laboratory)

Since the first release of the CCP program suite at the end of 1981, another fourteen programs have been added to the suite with the relevant data formats modifications made and documentation prepared. The 'User' and 'Program' documentation for each of the programs has now been transferred into machine readable form available on the private disk DLO251 on the NAS AS/7000 and available for distribution on magnetic tape.

There would now seem to be a fairly comprehensive set of programs in the suite and future development will be guided largely by the use that is made of the suite, with priority being given to assist those who are making active use of the suite or are attempting to transfer the suite to another machine. Programs will be added to the suite as the need arises. It is hoped that real use of the suite will lead to the ironing out of problems and that sufficiently prompt assistance can be given to those who find problems with programs not to discourage them from using such programs.

A current version of the CCP program list follows.

Note: The programs listed below are normally only available on the NAS AS/7000 at Daresbury unless otherwise stated.

The list of programs has been divided into 4 sections as follows:

- Section 1 Programs fully incorporated into the CCP program suite.
- Section 2 Programs under active development for the CCP program suite.
- Section 3 Special category programs.
- Section 4 Programs with no current plans for inclusion in the CCP program suite.

Within sections 1,2 and 4, the programs have been divided into the following categories:

1. Data Processing
 2. Scaling and Merging of 3D Data
 3. MIR Phasing and Heavy Atom Refinement
 4. Fourier Map Calculation and Structure Factor Calculation
 5. Structure Refinement
 6. Model Stereochemistry
 7. Map and Model Plotting
 8. Rotation and Translation Functions
 9. Phase Recombination
 10. Structure Prediction
 11. Map Manipulation
 12. Weighting Schemes
 13. Graphics Input
- UTILITIES

```

*****
*                               S E C T I O N  1                               *
*                               *                                               *
* PROGRAMS FULLY INCORPORATED INTO THE CCP PROGRAM SUITE                       *
*                               *                                               *
*****

```

NAME	DESCRIPTION	PROG	DIS	DOC	CON
---	-----	---	---	---	---
		STAT			

1. Data Processing

2. Scaling and Merging of 3D Data

AGROVATA	Merge and scale reflection datasets and calculate statistics	F	Y	A	JWC
ANISOSC	Scale derivative data using anisotropic temperature factor scaling function.	F	Y	A	JWC
ANSC	Analyse and scale derivative data	F	Y	A	JWC
CADLCF	Collect assorted data, apply symmetry transformation and format changes.	F	Y	A	JWC
RFACTOR	Scale datasets and calculate statistics. Option to prepare Fourier coefficients	F	Y	A	JWC
ROTAVATA	Determines scale and temperature factors for overlapping batches of data. (method of Fox and Holmes)	F	Y	A	JWC
TRUNCATE	Truncation of data output from AGROVATA	F	Y	A	JWC

3. MIR Phasing and Heavy Atom Refinement

PHASE	Calculation of isomorphous phases , (produces Hendrickson-Lattman coefficients). Birkbeck version.	F	Y	A	JWC
PHASE2	Calculation of isomorphous phases (Phil Evans' version)	F	Y	A	JWC
REFINE	Refinement of heavy atom positions (based on the Oxford Boss programs using FHLE). Birkbeck version.	F	Y	A	JWC
REFINE2	Refinement of heavy atom positions (Phil Evans' version)	F	Y	A	JWC

4. Fourier Map Calculation and Structure Factor Calculation

FFT	Fast fourier transforms, method of Ten Eyck. Available for subgroups of all space groups (Acta Cryst (1973) A29, 183). Birkbeck version.	F	Y	A	JWC
FFT81	Fast Fourier Transforms, method of Ten Eyck. Available for subgroups of all space groups. (Phil Evans' version)	F	Y	A	JWC
GENSFC	Generate electron density map from a coordinate set with optional structure factor calculation using the FFT routines. (Phil Evans' version)	F	Y	A	JWC
SF	Structure factor calculation using FFT. (Part of refinement program (see section 5 below))	F	Y	A	JWC
SFC	Reverse FFT to calculate structure factors from an electron density map	F	Y	A	JWC

5. Structure Refinement

SF	Isaacs-Agarwal fast Fourier refinement (Acta Cryst (1978) A34 791). Available for space groups : P1, P21, P212121, P21212, P41212, P3121, R3	F	Y	A	JWC
----	--	---	---	---	-----

6. Model Stereochemistry

DISTANG	Protein geometry (distances and angles)	F	Y	A	JWC
MODELFIT	Regularisation of peptide geometry (Acta Cryst (1976) A32 311)	F	Y	A	JWC

7. Map and Model Plotting

PLUTO	Fourier map and/or protein model plotting (highly modified version of Sam Motherwell's original program, mostly by Phil Evans)	F	Y	A	PAM
POLY	Multiple stereo views of alpha carbon backbones with automatic scaling for direct viewing.	F	Y	A	JWC

8. Rotation and Translation Functions

ALMN	York (Eleanor Dodson) version of Rotation function, using method of Crowther. All three stages are combined in one program. All symmetry related peaks are output. basic version allows up to 30 Bessel functions. Version ALMN60 available for 60 Bessel functions.	F	Y	A	PAM
SEARCH	Translation search and R-factor calculation by Eleanor Dodson	F	Y	A	PAM

9. Phase Recombination

COMBIN	Bricogne's method of combining Hendrickson-Lattman coefficients.	F	Y	A	JWC
--------	--	---	---	---	-----

10. Structure Prediction

PREDICT	Secondary structure prediction. Set of 6 programs from the University of Leeds (Collected by Elias Eliopoulos)	F	Y	A	JWC
---------	--	---	---	---	-----

11. Map Manipulation

EXTEND	Extend Fourier map (e.g. for PLUTO)	F	Y	A	JWC
--------	-------------------------------------	---	---	---	-----

12. Weighting Schemes

-

13. Graphics Input

-

UTILITIES

BIRKLCF	Convert Birkbeck binary reflection data to LCF format	F	Y	A	JWC
CARDLCF	Convert free format card image reflection data to LCF format	F	Y	A	JWC
DODLCF	Convert Dodson binary reflection data to LCF format	F	Y	A	JWC
EDITLCF	Edit/Inspect the reflection data in an LCF reflection data file	F	Y	A	JWC
KEELCF	Convert Keele merged binary SSM output reflection data to LCF format	F	Y	A	JWC
LCFDUMP	Dump/summarise the contents of an LCF file	F	Y	A	JWC
LCFSORT	Sort the reflection data in an LCF reflection data file (up to 5 sort keys)	F	Y	A	JWC
LCFUTILS	LCF utility program to create a new or rearranged or edited LCF file from 1 or 2 existing LCF files.	F	Y	A	JWC
MESSAGE	Message and program logging program (PCLOG)	F	Y	A	JWC
PASCLCF	Convert reflection data in 'PASCAL' output format to LCF format suitable for input to ROTAVATA and AGROVATA.	F	Y	A	JWC
PRTMAP	Print map file contents on the line printer	F	Y	A	JWC

```

*****
*              S E C T I O N  2              *
*                                              *
*  PROGRAMS UNDER ACTIVE DEVELOPMENT FOR THE CCP PROGRAM SUITE  *
*                                              *
*****

```

<u>NAME</u>	<u>DESCRIPTION</u>	<u>PROG</u>	<u>DIS</u>	<u>DOC</u>	<u>CON</u>

1. Data Processing

-

2. Scaling and Merging of 3D Data

DSCALE	Scaling of heavy atom derivatives. (6 methods are available e.g. weighted Kraut)	F		B	DSM
--------	--	---	--	---	-----

3. MIR Phasing and Heavy Atom Refinement

PHARE	MIR phasing and heavy atom refinement (Bricogne's method)	F		B	AJW
-------	---	---	--	---	-----

4. Fourier Map Calculation and Structure Factor Calculation

-

5. Structure Refinement

-

6. Model Stereochemistry

-

7. Map and Model Plotting

-

8. Rotation and Translation Functions

-

9. Phase Recombination

-

10. Structure Prediction

SEPLOT Plots of secondary structure prediction results.
 (from E.Eliopoulos, University of Leeds)

C JWC

11. Map Manipulation

-

12. Weighting Schemes

-

13. Graphics Input

-

UTILITIES

INTOLCF Convert Imperial College film data to LCF format

F

C AJW

```

*****
*                               S E C T I O N   3                               *
*                               *                                               *
* SPECIAL CATEGORY PROGRAMS *                                               *
*                               *                                               *
*****

```

NAME	DESCRIPTION	PROG	DIS	DOC	CON
----	-----	----	----	----	----
		STAT			

a) Restraigned Refinement of Protein Structures

PROTIN	Calculate protein geometry in preparation for PROLSQ, Hendrickson-Konnert. Only available on the Cray-1	F	N	B	JWC
PROLSQ	Hendrickson-Konnert restrained least squares refinement program. Available on the Cray-1 only. CCP has 2 versions using LCF formats. These are PRSQN for smaller proteins (up to 2000 atoms) is space group independent and allows for anisotropic temperature factors. PRSQB for larger proteins (up to 7000 atoms) is available for space groups 4,5,18,19,23 and 96. Other versions of the program are available at the University of Oxford, the University of Sheffield and at Imperial College.	F	N	B	JWC

b) Restraigned Refinement of Protein Structures

RESTRAIN	Restraigned refinement of protein structures. David Moss' program. Latest version is only available on the Cray-1 computer	F	N	B	DSM
----------	--	---	---	---	-----

c) Off-line film processing (Keele/Daresbury programs)

OSCAR	Off-line film data processing			B	TJG
FILMPACK	Combining data from two film packs			B	TJG
SSM	Scale, Sort and Merge program			B	TJG

d) Density modification programs

	Set of programs written by T.N.Bhat for his density modification procedure.			A	AJW
--	---	--	--	---	-----

e) Diffractometer data processing programs at Birkbeck College

VALDIF	Validate raw diffractometer data	F	A	IJT
DIFCOR	Calculate standard deviations and apply corrections	F	A	IJT
HRSLCF	Scale different batches or crystals and calculate statistics	F	A	IJT

f) Daresbury general utilities

COMPARE	Compare the contents of two program source files		A	JWC
DISKLIST	Produce a classified listing of the contents of a private disk pack at Daresbury.	N	C	JWC

g) Ancillary programs available from Brookhaven

TAPDIR	Print directory of Brookhaven Data base contents F. Bernstein, Brookhaven is the main contact at Brookhaven for all these programs	N	B	PAM
BENDER	Parameters for best wire models	N	B	PAM
CONNECT	Generate peptide connectivity	N	C	PAM
DGPLOT	Diagonal plots on the printer	N	C	PAM
DSTANCE	Distance calculations	N	C	PAM
FISIPL	Phi/Psi plots on the printer	N	B	PAM
NAMOD	Ball and stick model display	N	C	PAM
PHIPSI	Main chain torsion angles	N	B	PAM
STEREO	Extract X, Y, Z from stereo diagrams	N	C	PAM
TORSRU	Complete torsion angles	N	C	PAM
TOTALS	Validation of master record	N	C	PAM

h) Small molecule programs

SHELX	Structure solving package written by George Sheldrick, Gottingen.	N	B	ME
MULTAN78	Direct methods for structure determination written by Peter Main, York University.	N	B	ME
PLUTO	Molecular plotting program. Original version for small structure drawings, written by Sam Mother- well, Cambridge.	N	B	ME
DIRDIF	Direct methods using difference structure factors Acta Cryst (1979) A35, 765. Written by P. Beurskens at Nijmegen.	N	B	?

i) Crystallographic Program Package

ROCKS Crystallographic system from George Reeke at the
Rockefeller University, New York, encompassing many
aspects of protein crystallographic computing. Code
is available at Daresbury but the suite has not been
implemented.

N B JWC

```

*****
*               S E C T I O N   4               *
*               *                               *
* PROGRAMS WITH NO CURRENT PLANS FOR INCLUSION IN THE CCP PROGRAM SUITE *
*               *                               *
*****

```

NAME	DESCRIPTION	PROG	DIS	DOC	CON
----	-----	----	----	----	----
		STAT			

1. Data Processing

-

2. Scaling and Merging of 3D Data

NORMAL	Calculation of overall scale and temperature factors from MULTAN-78 (original author is Peter Main, York University)	N	B	?
--------	--	---	---	---

3. MIR Phasing and Heavy Atom Refinement

-

4. Fourier Map Calculation and Structure Factor Calculation

FCALC	Structure factor calculation using overall temperature factor. Subgroups available for all space groups.	B	DSM
FCALB	As FCALC only with individual structure factors	B	DSM
GENED	Map generation from model co-ordinates	B	AJW

5. Structure Refinement

-

6. Model Stereochemistry

BALUPLOT	Balusubramanian plots of ϕ, ψ	C	DSM
EREFN	Levitt energy refinement (Acta Cryst (1978), A34, 931)	B	DSM
MAIN	Generation of a polyaniline structure (Starting from C alpha positions ?)	B	DWR

7. Map and Model Plotting

RIBBONS	Ribbon program, method of McLachlan (o/p to FR80)	C	DSM
---------	---	---	-----

8. Rotation and Translation Functions

-

9. Phase Recombination

HLSIMW	Hendrickson Lattman phase combination with Sim weighting. Input is from the PHASE output, plus Fc's calculated by RESTRAIN. Its function appears to be very similar to COMBIN	B	IJT
--------	---	---	-----

10. Structure Prediction

-

11. Map Manipulation

ISOLATEM	Isolates protein molecules.	C	DSM
KEBAB	Skew planes using Bricogne double sort method Beware there are some space group specific parts	B	AJW
ROTMAP	Rotates map (applies the angles obtained from the rotation function and applies them to the electron density map)	B	IJT

12. Weighting Schemes

SIMW	Application of Sim's weighting scheme to partial structures.	B	IJT
------	--	---	-----

13. Graphics Input

CONPROG	Contours FFT output into a form suitable for Builder, Diamond's interactive graphics display programs	B	AJW
MAPMIX	Combines several outputs from CONPROG	B	AJW
AJCON	contours FFT output for FRODO on the graphics (ie for A.Jones program)	B	IJT

UTILITIES

-

NOTES ON THE ABOVE TABLE

Introduction

Most of the programs listed above are being developed in association with the Collaborative Computing Project in Protein Crystallography (CCP4). Many programs are third or fourth generation i.e. after initial coding they have been modified by several different laboratories.

The table is divided horizontally into sections on particular areas of computing. The meaning of the abbreviations and the content of the various columns of the table are given below:

1. Program name

This is an 8 letter code given as the program name for shorthand reference.

2. Program description

A few words to outline the purpose of the program, perhaps the program method, historical development and current limitations.

3. Program status (PROG STAT)

Two columns are used to describe the implementation status.

- F denotes a program version modified to use the standard data formats for the CCP program suite for reflection data (LCF format), coordinate data (Brookhaven format) and map data.
- * This code indicates that the program will very soon be ready for inclusion in the CCP program suite. The program has been implemented using standard data formats and the documentation has been prepared in standard format though not yet necessarily word processed.

4. Program distribution (DIS)

One column is available for one of three symbols..

- Y Program is available for distribution, i.e. a version with new formats is available and documentation is on the word processor.
- N The program is not available for distribution and will not be in future. This may apply to program packages obtained from elsewhere which are always distributed from the original site.
- blank A program which is not yet ready for distribution but which may be sometime in the future.

5. Documentation status (DOC)

One of three symbols may be used to state this.

- A Documentation in standard form is available on the word processor.
- B Written documentation
- C 'Back of envelope' documentation

6. Contact (CON)

The initials of the person responsible for the implementation at Daresbury are given if this person is still working in the UK or alternatively a member of his group is indicated. Usually initial questions should be directed to this person, who will know about program versions, methods of running the program, etc.

A key to the initials is given below :

JWC	John Campbell	DARESBUY	0925 - 65000 ext 528
ME	Mike Elder	DARESBUY	0925 - 65000 ext 350
TJG	Trevor Greenhough	KEELE/DARESBUY	0925 - 65000 ext 531
DSM	David Moss	BIRKBECK	01 - 580 - 6622
PAM	Pella Machin	DARESBUY	0925 - 65000 ext 528
DWR	David Rice	SHEFFIELD	0742 - 78555 ext 4242
IJT	Ian Tickle	BIRKBECK	01 - 580 - 6622
AJW	Alan Wonacott	IMPERIAL	01 - 589 - 5111 ext 1853

