DARESBURY LABORATORY

INFORMATION QUARTERLY for PROTEIN CRYSTALLOGRAPHY

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

Number 1 OCTOBER, 1979

Contents

	Page
Editorial	1
Minutes of the Working Group 2 Meeting	3
Status of Programs	6
Definition and Description of 'NA2' Format	12
Least Squares Refinement Using FFT	14
Distribution	15

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

EDITORIAL

Readers are probably aware that various Collaborative Computational Projects (CCPs) have been set up within SRC. A CCP in Protein Crystallography has recently been formed and the purpose of this newsletter is to report on the progress and activities of this particular CCP.

The work of the project is supervised by two Working Groups. Working group 1 is concerned with the overall direction of the project whereas Working group 2 is more involved with the detailed work of program implementation. Two research associates have been appointed to work on the project, Dr John Campbell will be working at the Daresbury Laboratory (from Nov 1st 1979) and Dr Bhat is already working with Professor David Blow's group at Imperial College.

Both Working groups met during the Wye College Workshop (Aug 28-31, 1979) and the minutes of the Group 2 meeting are included in this newsletter for information. With the help of the various groups Alan Wonacott has compiled a detailed list of the programs used by protein crystallographers, with notes concerning their availability. The program list as published here will be input to the Daresbury computer so that it can be updated when necessary and can be inspected by computer users. The data format proposed by the Imperial group at the Working group meeting is also included here with suggested improvements incorporated.

We hope to issue a newsletter 4 times each year and expect future issues to consist of contributions from the various

university groups. I have taken this opportunity of adding a brief contribution of this type to the end of this newsletter, describing the current state of implementation of the Isaacs-Agarwal refinement program on the IBM 370/165 computer at Daresbury.

Please note that the main aim of the newsletter is to keep people informed of the work being carried out by the CCP and of which programs are available for general use. We will therefore rely upon your contributions for future editions!

Minutes of a meeting of Working Group 2 of the protein crystallography collaborative computational project which was held at Wye College, Kent, at 20.00 on 30th August, 1979.

Present:

Dr. Alan Wonacott	(Chairman)	Imperial College, London
Dr. Bhat		Imperial College, London
Dr. David Moss		Birkbeck College, London
Dr. Ian Tickle		Birkbeck Collge, London
Mrs. Eleanor Dodson		York University
Dr. Phil Bourne		Sheffield University
Dr. Bob Stansfield		Sheffield University
Dr. Keith Wilson		Oxford University
Dr. John Helliwell		Keele University/SRC Daresbury
Dr. Phil Evans		MRC Cambridge
Dr. Klaus Bartells		Munich
Dr. Wolfgang Steigemann		Munich
Dr. John Campbell		SRC Daresbury
Miss Pella Machin	(Secretary)	SRC Daresbury

1. The chairman explained that the collaborative computational project (CCP) in protein crystallography had been approved and that the meeting had been called to formulate a plan of work for the RAs and university groups associated with the project. He explained that 2 members of the Munich group of protein crystallographers who were at the Wye Workshop had been invited to the meeting as they had experience with setting up a large protein program system.

2. The chairman circulated a draft list of protein crystallography programs.

   (i) Additional programs were added to the list.
   (ii) It was noted which programs were already at the Daresbury Laboratory (DL).
   (iii) Availability of documentation was noted.

   The modified program list will be published in the CCP Newsletter.

3. The various program versions were discussed in some detail, and consideration was given to which programs should be implemented at Daresbury and with what priority.

   (i) It was agreed that in the short term the work of implementing programs should be apportioned between the RAs and the groups. A group with particular interest in one stage of data processing should be responsible for implementing that stage, and could choose the program version.
   (ii) In the long term a comprehensive suite of programs should be developed with standardised input and data formats. If a program existed in many versions then the best features of each should be combined into one standard version. However it was agreed that many parallel programs should remain available if the methods used differed (such as in refinement).

4. It was recommended that some programs be implemented as soon as possible (if they were not already available) and the following division of work was agreed.

Film Evaluation (excluding absorption)	- John Helliwell
Scaling and correction of film data (PROTPAIR, PASCAL)	- Imperial
Diffractometer data handling (DIFVAL, PDPLUS)	- Birkbeck
Scaling and merging of 3D data (ROTAVATA, AGROVATA)	- Imperial
Fast Fourier Transform implementation	- Birkbeck
Rationalisation of FFT versions	- SRC Daresbury
MIR phasing (PHARE, PHASE, REFINE)	- Birkbeck
Isaacs-Agarwal Refinement	- SRC Daresbury
CONTOUR	- Birkbeck

It was agreed that Ian Tickle should be responsible for checking the various versions of EREFN, and that Keith Wilson should write to Sygusch about his 'phase' program.

5. The meeting considered various aspects of general program systems.

(i) Comparison of large program systems linked at job control level with those linked at program level - the general feeling of the meeting was to favour the former method.

(ii) The relative merits of the IBM sort and a standard machine independent FORTRAN sort routine were discussed. Although a final decision was not reached it was noted that the IBM sort is very efficient and fast. The problems of transferring to another machine would not be great provided the sort remained an individual job step and was not required to be an integral part of a program.

6. Alan Wonacott and Bhat proposed a standard data format for reflection data. Details of this format will be published in the CCP Newsletter. Discussion followed on the problems of choosing either a rigid or flexible system.

(i) The Munich group used a very rigid system and concealed the complexities of the actual format from the user.

(ii) The Boss system allowed more flexibility, for example concerning the number of derivatives which could be stored.

(iii) The proposed new format (based on the original Cambridge format) allowed flexibility and was easy to use.

In general it was agreed that the proposed format was very satisfactory and that it should be adopted with a few modifications.

7. A format for coordinate data was requested. It was agreed that this could be fairly rigid and it was generally accepted that Eleanor Dodson's format be adopted.

The format was:

Atom no, X, Y,Z,B,Occupancy, Z, residue no, Atom name, residue name

in format (I5, 4F10.5, F5.2, I5, I10, A4, 1X, A3)

8. The meeting considered the way in which changes should be made to the Daresbury programs and suggested the following:

   (i) Trivial errors should be corrected immediately.

   (ii) Major modifications should be made by first taking a copy of the "standard" program, then modifying and testing the changes. The modified version should then be introduced either at the discretion of the Daresbury RA or after discussion at a working group meeting.

   (iii) A file should be set up on the Daresbury computer under TSO (identifier PC) for messages and information. The Daresbury RA should be responsible for keeping the file in order but access would be available to all groups, to read the file or append new information to it.

9. Future meetings were discussed and it was agreed that

   (i) There should be about 4 meetings each year.

   (ii) Meetings should be held where most convenient probably London or Daresbury.

   (iii) Meetings should consist of progress reports from RAs and university groups and discussion on some special topic under consideration for example refinement methods.

10. It was agreed that a newsletter containing the program list and data format be circulated as soon as possible.

11. It was suggested that the next meeting should be arranged for some time after the Daresbury RA had started work, such as the end of November.

STATUS OF PROGRAMS ALREADY/TO BE IMPLEMENTED AT DARESBURY

FUNCTION	PROGRAM NAME	BRIEF DESCRIPTION	IMPLEMENTATION STATUS	DOCUMENTATION STATUS	RESPONSIBILITY
Basic Data Processing-Film	PROTIN) PAIROT)	Scales packs of rotation film data Apply corrections, calculate statistics on scaling and Rsym.	C	B	AW
	PASCAL	Scales packs of film data (rotation or precession) which have been pre-corrected for transmission etc.	C	C	AW
	OSCAR) FILMPACK)	Oxford programs for off-line processing of rotation photographs.	C	B	JH
Diffractometer	DIFVAL	Validation of diffractometer data	A	B	JW/DM
Scaling & Merging of 3D data					
	ROTOVATA	Determines relative scale and temperature factors using modified Fox & Holmes.	B	B	TB
	AGROVATA	Merges dataset. Calculates statistics	B	B	TB
	HRS4	Scaling of datasets. Hamilton, Rollett & Sparks	A	B	IT/DM
	SHELLS	Scaling of datasets in shells	A	C	IT
	TRUNCATE	Applies Bayesian statistics	C	A	TB
	CAD	Collect assorted data	C	-	
	KSCALE	Kraut scaling of heavy atom derivatives	A	A	JW/DM
	ANSC	Calculates scale factors native/derivative as functions of several variables. Analyses scales	C	-	AW

FUNCTION	PROGRAM NAME	BRIEF DESCRIPTION	IMPLEMENTATION STATUS	DOCUMENTATION STATUS	RESPONSIBILITY
Fast Fourier Transforms	FFT	Ten Eyck FFT. Available for various spacegroups. Versions with modified I/O and D/A files.	A	A	IT/TB
Structure Factors	P1SF	Fast Fourier structure factors. Available for P1, $P2_1$, $P2_12_12_1$, $P2_12_12$, $P4_12_12$, $P3_121$, R3	A	B	IT/TB/ED
	FCALB) FCALC)	Structure factor calculator (overall B for FCAL, individual B's for FCALB)	A	B	TS/IT/DM
Heavy Atom Phasing & Refinement	PHARE	MIR phasing and refinement	A	A	TB
	PHASE	Calculation of isomorphous phases	A	B	IT
	REFINE	Refinement of heavy atom positions	A	B	IT
Fourier Refinement		Agarwals & Isaacs Jack			
	SFLSP	Konnert - restrained least squares refinement	A	B	AM/DM
	CORRELS	Constrained/restrained refinement	C	-	AW
Model Drawing	RIBBONS	McLachlan ribbon program	B	C	TS
	DRAW	Tony North's program	C	B	K.A.
	PLUTO	Molecule drawing	B	B	IT
	ORTEP	" "	-	-	
	CPK	Space filling diagrams (Doug. Richardson's algorithm)	C	C	AW

FUNCTION	PROGRAM NAME	BRIEF DESCRIPTION	IMPLEMENTATION STATUS	DOCUMENTATION STATUS	RESPONSIBILITY
Map handling					
	ISOLATEM	Isolates protein molecules	A	C	T.S./J.W.
Generation	ATOMAP	Calculate map from model. General	C	B	TB
	GENED	Map generation	A	B	TB
Shew planes	KEBAB	Bricogne double sort method	A	B	TB
	ROTMAP	Rotates maps	A	B	IT
Plotting	CONTOUR	Map contouring	A	C	TS
Bilder input	CONPROG)	Convert FFT output for "BILDER"	A	B	TB
	MAPMIX)	input	A	B	TB
Model Stereo-	RSRP	Diamond realspace refinement	C	A	AM
Chemistry	MODELFIT	Regularisation	A	B	TS
	EREFN	Levitt energy refinement	B	A	TS/LP
	DISTANG	Protein geometry	B	B	BG
	RAMA	Ramachandra plot			
	BALUPLOT	Balusubramanian plots of ϕ,ψ	A	C	TS
Small Molecule	CRYSTALS (Oxford)	Small molecule package	A	A	IT
	SFLS	Structure factor least squares	-	-	
	MULTAN)	Direct methods	A	A	IT
	YZARC)		C	B	IT
	SFFP	Classical structure factor/Fourier program	C	B	PL

FUNCTION	PROGRAM NAME	BRIEF DISCRIPTION	IMPLEMENTATION STATUS	DOCUMENTATION STATUS	RESPONSIBILITY
Molecule Fitting	EZIFIT)	McLachlan's programs	A	C	IT/LP
	ROTPROG)	Least squares fitting of 2 molecules			
	BITFIT)	Rotates coodinates	A	C	IT/LP
	BITSEC)	Least square fitting			
		Plotting	C	C	IT
Rotation Function	DLM	Crowther's Rotation Function	A	B	IT
	ALMN	Converts to spherical	A	B	IT
	FRFSUM	Rotation map	A	B	IT
Translation Function	TRANSFN	Crowther Translation function	B	C	IT
Phase Recombination	COMBIN	Recombines two sets of phases using A,B,C,D coeffts.	A	B	TB

FUNCTION	PROGRAM NAME	BRIEF DESCRIPTION	IMPLEMENTATION STATUS	DOCUMENTATION STATUS	RESPONSIBILITY
Ancilliary	PRIME 1	Converts SMOG O/P into line transmittable form	A	A	TS
	PRIME 195	Receives PRIME1 O/P at R.L. and sets up input for FR80	A	A	AB
	VIEW	Converts FR80 O/P into Tektronix format	A	B	TS/AB

NOTES:

Implementation status

A - No known imperfections in Daresbury implementation

B - Working with some deficiencies

C - Not yet working at Daresbury

Documentation status

A - Accurate machine readable documentation

B - Documentation not machine readable or has imperfections

C - Back-of-envelope documentation or nothing

Responsibility

The initials of people who know about the use or who have written the program.

Initials

A.B.	Tony Berraston+	J.H.	J. Helliwell	A.W.	A.J. Wonacott
T.S.	Trevor Sewell	J.W.	John Watson	A.M.	Andy Morffew
B.G.	Barry Gelatly	I.T.	Ian Tickle	T.B.	T. Bhat
L.P.	Lawrence Pearl	D.M.	David Moss	P.L.	P. Lindley
E.D.	Eleanor Dodson	K.A.	Kevin Adolphe		

+ Rutherford

DEFINITION AND DESCRIPTION OF 'NA2' FORMAT

- Alan Wonacott and T N Bhat

IT IS PROPOSED TO INTRODUCE A GENERAL FORMAT FOR REFLEXION DATA WHICH WILL EVENTUALLY BE USABLE WITH ALL PROGRAMS CONCERNED WITH:-

A) SCALING AND MERGING OF DATA
B) COMPUTATIONS OF R FACTORS, DIFFERENCE FOURIERS ETC.
C) ARCHIVING OF DATA

A SET OF ROUTINES WILL BE PROVIDED FOR READING AND WRITING OF THESE FILES WHICH WILL BE INCORPORATED IN THE STANDARD PROGRAMS AND MAY BE USED IN ANY FUTURE PROGRAMS THAT ARE WRITTEN.

FILES WRITTEN IN 'NA2' FORMAT WILL CONSIST OF:-

1) FIXED LENGTH RECORDS - RECFM=FB,LRECL=2N

WHERE N IS ANY POSITIVE INTEGER DETERMINED BY USER OR CALLING PROGRAM THE RECORD LENGTH WHEN WRITING A FILE IS DETERMINED BY THE CALLING PROGRAM. IN READING FILES, THE RECORD LENGTH IS READ FROM THE FILE.

2) ALL VALUES IN A RECORD ARE STORED AS 2 BYTE INTEGERS. A 'COLUMN NUMBER' M IS DEFINED SUCH THAT INFORMATION IN BYTES 2M-1 AND 2M IS REFERRED TO IN ANY RECORD.

3) IN NORMAL USAGE THE FIRST THREE COLUMNS WILL CONTAIN H,K,L ALTHOUGH THERE IS NO RESTRICTION ON THE CONTENTS OF ANY GIVEN COLUMN.

4) FILES WILL NORMALLY SORTED ON THE FIRST THREE COLUMNS. FOR THE PURPOSES OF SCALING AND MERGING OF DATA SETS, COLUMNS 4 AND 5 WHICH WILL CONTAIN MISYM (FLAG FOR FULL OR PARTIAL REFLEXION + CODE FOR SYMMETRY OPERATION TO TRANSFORM TO ASYMMETRIC UNIT OF RECIPROCAL SPACE) ,IXL (SERIAL NUMBER ASSOCIATED WITH DATA FROM A PARTICULAR FILM OR CRYSTAL) ARE ALSO SORTED.

THE RECORDS CONTAINED WITHIN A FILE CAN BE OF THREE TYPES.

TITLE RECORD

HEADER RECORD

DATA RECORD

TITLE RECORD

THIS ENABLES A USER TO ASSOCIATE A TITLE AND ANY RELEVANT COMMENTS WITH EACH DATASET. THE NUMBER OF BYTES OF TITLE INFORMATION THAT IS INCLUDED WITH A PARTICULAR DATASET IS CONTROLLED BY THE USER.IT IS WITHIN THE RANGE 2-800 BYTES.IF MORE THAN ONE TITLE EXISTS IN A PARTICULAR DATASET (FOR EXAMPLE IN A MERGED DATASET) THEN THE SUM OF THE LENGTHS OF ALL THE TITLES CANNOT EXCEED 800 BYTES.

IF MORE THAN 800 BYTES OF TITLE INFORMATION IS ENCOUNTERED IN 'READ MODE' THEN READING OPERATION IS NOT INHIBITED, BUT THE EXTRA TITLE INFORMATION IS IGNORED.

IN 'WRITE MODE', ONLY 800 BYTES OF TITLE ARE WRITTEN OUT.

COLUMNS 1,2 AND 3 OF EACH TITLE RECORD ARE ZEROS.
COLUMN 4 CONTAINS DATASET SERIAL NUMBER

COLUMN 5 CONTAINS AN INDIVIDUAL TITLE SEQUENCE NUMBER.
THE REMAINDER OF THE RECORD CONTAINS TITLE INFORMATION.

COLUMN	1	2	3	4	5	6 ---------- N
					SEQ. NO.	
	0	0	0	SER.NO.	1	TITLE OF
	0	0	0	SER.NO.	2	DATASET .....
	0	0	0	SER.NO.	3	.....
	0	0	0	SER.NO	4	.....

THE SERIAL NUMBER APPEARING IN COLUMN 4 IS USED TO IDENTIFY THE TITLE CORRESPONDING TO EACH SET OF DATA WHENEVER ONE DEALS WITH MERGED DATASETS. THE SEQUENCE NUMBER IS MADE USE OF IN THE REASSEMBLY OF TITLE INFORMATION.

THE POSITION OF THE SERIAL NUMBER CAN BE MOVED LEFT OR RIGHT IN A RECORD BY A USER IF HE WISHES TO DO SO. FOR EXAMPLE, ONE MAY WISH TO SORT IN COLUMN 5. THEN SERIAL NO. SHOULD BE MOVED TO COLUMN 6.
THE SUBROUTINE WRITING THE TITLE WILL AUTOMATICALLY ADJUST THE REMAINING COLUMNS IN CASES WHERE USER ALTERS THE POSITION OF THE SERIAL NUMBER.

HEADER RECORD

THE PURPOSE OF THIS RECORD IS TO ENABLE A USER TO ASSOCIATE EACH COLUMN OF THE DATASET WITH SOME SUITABLE ABBREVIATED NAME OR SYMBOL. (EG. OBSERVED AMPLITUDE BY 'FO' OR 'FOBS'). THEN SOME OTHER PROGRAM MAKING USE OF THE DATASET CAN DECODE THESE SYMBOLS OR NAMES AND DETERMINE THE COLUMN NUMBER FROM WHICH DATA SHOULD BE EXTRACTED.
THE STRUCTURE OF THESE RECORDS IS IDENTICAL TO THAT OF TITLE RECORDS EXCEPT THAT THESE RECORDS HAVE '1' IN COLUMN 3 FOR IDENTIFICATION. IN ADDITION BLANKS ARE USED AS SEPARATORS BETWEEN SYMBOLS.

DATA RECORD

THESE RECORDS CONTAIN THE NUMERICAL INFORMATION CODED AS TWO-BYTE INTEGERS. AS WITH TITLE RECORDS, THEY HAVE A SERIAL NUMBER ASSOCIATED WITH THEM IN THE SAME COLUMN AS THE TITLE RECORD. COLUMNS PRECEEDING THAT FOR THE SERIAL NUMBER MAY ONLY CONTAIN POSITIVE INTEGERS GREATER THAN 1.
THUS FOR REFLEXION DATASETS WITH H,K,L CODED IN THE FIRST THREE COLUMNS IT IS NECESSARY TO BIAS H,K AND L WITH A POSITIVE NUMBER (EG. 512) SO THAT CHARACTER MODE SORTING CAN BE USED WITH THE IBM SORT/MERGE PACKAGE.
INFORMATION FOR A PARTICULAR (HKL) VALUE CAN EXIST AS A SINGLE RECORD OR MAY EXTEND TO MULTIPLE RECORDS WHICH MUST THEN HAVE DIFFERENT SERIAL NUMBERS.

SUBROUTINES WILL BE PROVIDED FOR:-

A) WRITING TITLE AND HEADER RECORDS.
B) FOR THE READING AND REASSEMBLY OF TITLE AND HEADER RECORDS.
C) FOR READING AND IDENTIFYING THE COLUMN CORRESPONDING TO A PARTICULAR HEADER SYMBOL.

ADDITIONAL SPECIAL RECORDS

IT WILL BE POSSIBLE TO INCLUDE ADDITIONAL RECORDS FOR STORING INFORMATION SUCH UNIT CELL DIMENSIONS; THESE RECORDS WOULD BE TREATED IN MUCH THE SAME MANNER AS TITLE OR HEADER RECORDS.

SUGGESTIONS AS TO THE TYPES OF EXTRA INFORMATION REQUIRED WOULD BE USEFUL.

Least Squares Refinement using the FFT

Pella Machin Daresbury Laboratory

A Least Squares refinement program based on the Isaacs-Agarwal method using fast fourier transform techniques has been obtained from Eleanor Dodson at York University. Originally there were different program versions for different space groups, and for different size problems. The versions for different amounts of core have now been combined - the user requests a work array size with a data card and the program automatically extends to this size and allocates arrays. We are trying to simplify the versions needed for different space groups.

1. The version for space group 92(P4,2,2) has been implemented and tested on the IBM computer and is now available.

2. Similar changes will now be made to the codes for space groups 2(P-1), 4(P2,), 19(P2,2,2,), 152(P3,21), and tests will be run if test data are available from York. We do not forsee further problems with these additional space groups.

The program (for space group 92) can be accessed using the clist PC.CLIST(SFRF92). We hope to document this work both on paper and on the computer in the near future. Meanwhile please contact myself or Mike Elder at the Daresbury Laboratory (0925-65000 ext 528) if you have any queries regarding this refinement program.

Distribution

If you are reading someone else's copy of this Information Quarterly or if you know of someone who is not on the distribution list but should be, please fill in and return the form below.

To:

Pella Machin

Daresbury Laboratory

Science Research Council

Daresbury

Warrington

WA4 4AD

Please add me to the distribution list for subsequent issues of Information Quarterly for Protein Crystallography.

Name

Address