

DARESBUURY LABORATORY

INFORMATION QUARTERLY

for

PROTEIN CRYSTALLOGRAPHY

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

Number 19

DECEMBER 1986

Contents

Editorial	1
Beginner's guide to using JANET and EARN (Rob Stansfield, Edinburgh)	3
Refinement of Mouse Dihydrofolate Reductase (Elias Eliopoulos, Leeds)	13
Using FRODO to build molecular models into MIR electron density maps (Peter Brick et al, Imperial College, London)	17
The Area Detector (Guy Dodson, York)	21
A brief description of the Image Plate as an area detector for protein crystallography (John Helliwell, York)	23
The Transputer and Molecular Simulations (David Fincham, York)	25
The EEC Co-operative Programming Workshop on Position Sensitive Detector Software (Ian Tickle, Birkbeck College, London)	27
Cytochrome b: a bacterial ferritin and cytochrome (John Smith, Sheffield)	29
LGEM and LCHK: Utility programs for investigating Laue photographs (Mike Elder, SERC Daresbury)	31
Curve fitting during Laue data normalisation (John Campbell, SERC Daresbury)	37
Intensity data collection from Laue photographs - a progress report (Marjorie Harding et al, Liverpool)	43
CCP4 Program Suite Changes (Sue Bailey, SERC Daresbury)	49
AT LAST! A structure from the FAST A collaboration between DL and the Universities of Edinburgh, York and Liverpool (Steve Andrews)	51

Editor: Pella Machin

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

Deputy Editor for Imperial
College: Dr. Alan Wonacott

Imperial College of Science & Technology,
The Blackett Laboratory, Prince Consort Road,
London SW7 2BZ

Deputy Editor for Birkbeck
College: Dr. David Moss

Department of Crystallography, Birkbeck College,
University of London, Malet Street, London WC1E 7HX

EDITORIAL

Rob Stansfield has been responsible for stimulating and collecting contributions for this newsletter and our thanks go to him and to the people who have written the articles.

I would like to take this opportunity to issue a final reminder about the CCP4/Daresbury meeting on Protein Crystal Data Analysis. This will be held at Daresbury on 23-24 January 1987 and the closing date for applications is 16 December 1986. The meeting will consider the computational aspects of processing (1) film data, measured via monochromatic oscillation and white beam Laue methods, and (2) data collected on electronic area detectors, including a report on progress on the EEC initiative on this topic. Copies of the (provisional) programme are available from Daresbury.

Pella Machin
1/12/86

Beginner's guide to using JANET and EARN

=====

being the experiences of a first-time node ... October, 1986

Rob Stansfield, Tel: 031-667-1011 x 2408
Departmental Computing Officer,
Department of Biochemistry, University of Edinburgh Medical School,
Hugh Robson Building, George Square, Edinburgh, EH8 9XD.

JANET: STANSF at UK.AC.ED.BIOVAX DTE: 000016000141
ARPA: STANSF at UK.AC.ED.BIOVAX via UCL-CS.ARPA
BITNET: STANSF at BIOVAX.ED.AC.UK via UKACRL.BITNET
UUCP: STANSF at UK.AC.ED.BIOVAX via UKC.UUCP

PREAMBLE

=====

These notes have been prompted by several requests from users, non-users and about-to-be users of JANET for a noddy guide to electronic mail, file transfer and the use of remote computers.

I am not an expert in communications, so these notes are only the result of three months' experience of being connected to JANET. They are necessarily prepared with the perspective of a VAX/VMS (V4.3) user, using VAX PSI (V3.1) and Coloured Books (V4.1), and (independent of the VAX) a JNT-PAD (V2.1). They are not explicit about accessing JANET from other networks, and may not be relevant to non-VAX, non-VMS installations. I hope that they will be complemented, eventually, by the experiences of users of different systems.

There is a wealth of information available, perhaps too much, and I have tried to extract the salient bits for practical application. The glossary given here is an amalgamation of extracts from the SOURCES listed below. The entries in it are a personal selection.

The networks with which these notes are concerned are JANET and EARN. If you do not know what these are then read the GLOSSARY first. Use the glossary for terms not explained in the text. Information on the gateway between JANET and EARN is available (by file transfer!) from a facility called NETSERV. Much of the following information was extracted from files from NETSERV (see SOURCES). In order to use these networks from a VAX running VMS, two software packages called "VAX PSI" and "Coloured Books" need to be installed. "JANET_MAILSHR" and "JTMP" are optional extras.

Note that although DEC's proprietary DECnet software need not be installed, PSI half-expects it to be there, and DECnet naming conventions (nodename::username) should be understood to be distinct from JANET and EARN naming conventions.

JANET, unlike PSS and IPSS, does not charge for calls, but excessive use at peak periods may worsen response times for you and for other users.

EARN node names

A full list of EARN, BITNET and NORTHNET names can be obtained from NETSERV. The names have no heirarchical structure. They have some

informal structure as in many cases the first few characters define the country, the next the site and the rest may indicate the operating system in use. The network protocols limit the name to eight or less characters.

JANET names

Names in the JANET network are registered with the Name Registration Scheme. Names are heirarchical. The top level is UK and the second level contains AC - meaning Academic Community. Site names are at the third level. Any structure below the third level is at the discretion of the site. The name components may have a 'standard form' or a 'short form'. The components of a name are separated by '.'. A typical standard form name would be:- UK.AC.RUTHERFORD.GEC-A The corresponding short form would be:- UK.AC.RL.GEC-A For use within JANET the UK.AC may be omitted.

SENDING MAIL

=====

On a single host

Examples:-

```
$ MAIL
MAIL> SEND
To:      FRED
```

```
$ MAIL file.ext FRED
```

Comments:-

Bog standard VMS mail

Fred's username

Non-interactive mail

Over DECnet

Examples:-

```
$ MAIL
MAIL> SEND
To:      VAXZ::FRED
```

```
$ MAIL file.ext VAXZ::FRED
```

Comments:-

Bog standard VMS mail

Node and username

Non-interactive mail

Between hosts at a single site on JANET

The choice is between (CBS) "POST /TO=recipient /AT=remotesite", (PSI) MAIL to "PSI%remotesite::recipient" or (JANET_MAILSHR) MAIL to "JANET%remotesite::recipient".

Examples: sending mail to Andrew Gracie whose address (using the short form of host name) is ERCC23@UK.AC.ED.ERCVAX. The sender and the recipient are both using hosts in Edinburgh.

Example:-

```
$ POST /TO=ERCC23 /AT=ERCVAX file.ext
$ POST /TO=ERCC23 /AT=UK.AC.ED.ERCVAX -
file.ext
```

```
$ MAIL
MAIL> SEND file.ext
To:      PSI%000015000010::ERCC23
```

```
$ MAIL
MAIL> SEND file.ext
```

Comments:-

Coloured Book Software
Both hosts are at the
site UK.AC.ED
UK.AC.ED is redundant.

PSI MAIL
A different method to
the DTE-address.

Logical name ERCVAX translates
to PSI\$MAIL_DTE_000015000010 in

To: PSI%ERCVAX::ERCC23 logical name table PSI%DTE_TABLE

\$ MAIL JANET_MAILSHR "fix".
MAIL> SEND file.ext A different method to
To: JANET%ERCVAX::ERCC23 the same address.

\$ MAIL The recipient is known
MAIL> SEND /EDIT to have a forwarding
To: JANET%ERCVAX::"A.Gracie" address for A.Gracie

Between hosts at different sites within JANET

Include at least the third level component of the NRS name for the remote site. Use any of CBS, PSI-MAIL or JANET_MAILSHR MAIL fix.

Examples: sending mail from Edinburgh to Pella Machin whose address is CDS@UK.AC.DARESURY.DLVB (standard form) or CDS@UK.AC.DL.DLVB (short).

Example:-	Comments:-
\$ POST /TO=CDS /AT=DL.DLVB file.ext	Coloured Book Software
\$ MAIL	PSI MAIL
MAIL> SEND /SUBJECT="Next CCP4 meeting"	A different method to
To: PSI%000001003000::CDS	the DTE-address.
\$ MAIL	JANET_MAILSHR "fix".
MAIL> SEND /SUBJECT="Next CCP4 meeting"	A different method to
To: JANET%DL.DLVB::CDS	the same address.

From JANET to EARN

Either use (CBS) "POST /TO=recipient%remotesite /AT=UK.AC.RL.EARN" or use (VMS) MAIL to "EARN%remotesite::recipient".

Examples: sending a mail message from Edinburgh both to Phil Bourne whose address is SYSTEM@CUCCFA and to Martha Teeter whose address is TEETER@BCCHEM. CUCCFA and BCHEM are EARN node names.

Example:-	Comments:-
\$ POST /TO=SYSTEM%CUCCFA /AT=RL.EARN - file.ext	Coloured Book Software
\$ POST /TO=TEETER%BCCHEM /AT=RL.EARN - file.ext	Using the JANET-EARN gateway UK.AC.RL.EARN
\$ MAIL	JANET_MAILSHR "fix".
MAIL> SEND /EDIT /SELF	Interactive mail
To: EARN%CUCCFA::SYSTEM,EARN%BCCHEM::TEETER	

Please note that there is no guarantee that you will be notified of non-delivery of mail once it has passed the JANET-EARN gateway!

Using a distribution list

You must either write your own command file to issue POST commands to a series of addresses read from a file, or use VMS mail. For JANET and EARN addresses in the list you need JANET_MAILSHR. Suppose you prepare a file called CCP4.DIS (using the default file extension) containing ...

```
! Don Akrigg
janet%uk.ac.LEEDS.BIOVAX::DON
! Sue Bailey
```

```

janet%uk.ac.DL.DLVD::CCP4
! Steve Briant
janet%uk.ac.BBK.CR::UBCG11
! William Cruse
janet%uk.ac.CAM.PHX::WBTC
! Eleanor Dodson
janet%uk.ac.YORK.CHEMVAX::XTAL
etc.

```

You also prepare a message in file REMINDER.TXT (again using the default file extension). Then you can send a reminder that contributions are invited for the next CCP4 Newsletter to everyone on the list using the following command.

```
$ MAIL REMINDER "@CCP4" /SUBJECT="CCP4 Newsletter"
```

You don't know the mail address of the person you wish to contact

Telephone him/her. Or write. Otherwise use one of these schemes.

If you suspect the person uses an EARN node, then with reference to the NETSERV file "FILE\$ LIST\$" (see SOURCES), transfer the appropriate file of node names for that country or state. For example "FEARN NODELIST" for nodes in France. From this list I can find, for example, the name FREMBL51 for the EMBL outstation in Grenoble and a contact for further information.

A list of JANET sitenames is available from NETSERV in "JANET SITELIST" but the NRS database from SUPPORT@UK.AC.RL.VE is usually more up to date. The most useful way of interrogating it is with

```
$ NRSHELP /SEARCH=ALL "string"
```

```
(NRSHELP = "$DISK:[DIRECTORY]NRS_HELP.EXE")
```

where "string" could be "OX" for all the computers in Oxford etc. (This will also give you the DTE-numbers for interactive work).

Finally, your system manager can put distribution lists, mail from Martha Teeter, lists of colleagues' addresses etc. all in one directory which can be read publicly using \$ SEARCH DISK:[DIRECTORY]*.* "string".

Networks other than JANET and EARN

Not much known. If you use a PSS or IPSS gateway from JANET somebody will have to pay British Telecom. According to the NETSERV file "JANET EARNGATE" there are a number of gateways from EARN to other networks and these are known to the JANET-EARN gateway, although testing of these gateways from within JANET has not taken place! They would be used like this.

```
$ POST /TO=recipient%domain1%domain2 /AT=RL.EARN file.ext
```

and the current list of gateways is

DEC	ARPA	UUCP	CCNET	CSNET	MFENET	MAILNET	GOV
EDU	COM	MIL	ORG	UWO	CDN	BELL	CERN
MLNET	SUNET	UNINETT.					

FILE TRANSFER

=====

This is accomplished using the Coloured Books' TRANSFER command for which you will have documentation. On JANET you can 'send' and 'fetch' files providing you know a password for a user on the remote machine.

Files may only be 'sent' to EARN nodes and may not be 'fetched'. The transfer of binary files is probably only meaningful if both the EARN

node and JANET node are IBM or IBM-compatible machines.

USING REMOTE COMPUTERS

=====

Interactively

Each computer connected directly to an X.25 network is given an address by the Network Authority, known as its "DTE-number" or "Network User Address". On JANET and on PSS these are always twelve digits in length, although leading zeroes may often be omitted.

Your terminal may be wired directly to a computer, or it may connect to a PAD such as the JNT-PAD. The JNT-PAD does not have enough memory to hold the names of all the nodes on JANET. Therefore the nodes you can access by name will be limited in number and probably chosen by someone else. Hence you will often need to know the DTE-number of the remote computer as well as a username and password. The JANET-EARN gateway does not support interactive processes.

Once you are logged on to a VAX running PSI and Coloured Books, there are two different host-based (software) PADs available. The PSI-PAD is usually invoked with one of the commands

```
$ SET HOST /X29 DTE-number
```

```
$ SET HOST /X29 name
```

where "name" is a logical assignment to a DTE-number in the logical name table `PSI$DTE_TABLE`. However, in order to use an explicit DTE-number you need `NETMBX` privilege.

The Coloured Books' PAD is usually invoked using "\$ PAD" or "\$ RUN PAD" to get the same prompt "PAD>" as appears on a JNT-PAD. A remote node is called using one of

```
PAD> CALL DTE-number
```

```
PAD> CALL name
```

This time, the names of hosts you may access are stored and controlled by a "network authorisation file". Normally a user has access to this file for interrogation but not modification. This may be suppressed in future versions. Again, in order to use a DTE-number you need `NETMBX` privilege. Read the manuals, then try:

```
$ RUN NET$DIR:NETAUTH
```

```
Enter User record command USER
```

```
Enter User record command LIST myusername
```

```
Enter User record command SITE
```

```
Enter Site record command LIST ME
```

```
Enter Site record command EXIT
```

The database for site names is not the same as, although it should correspond closely to, the one which can be interrogated by the `NRSHELP` facility from `SUPPORT@UK.AC.RL.VE` mentioned earlier.

Non-interactively

Submission of jobs to run on remote hosts is only possible using `JTMP`. We have no experience of this facility yet. `JTMP` must be installed at both the submitting and the remote hosts, and be known to the Network Authorisation File.

SOURCES

=====

Manuals

- o VAX PSI - Introduction - Introduction to Packet Switching (Chap.1)

A good introduction to PSDNs and their components. A glossary.

- o Coloured Books - User Guide - General Technical Description (Chap. 6)
Good for background. A glossary.

NETSERV Files

- o The file "JANET EARNGATE" from NETSERV.
\$ transfer rl.earn::"JANET EARNGATE" JANET.EARNGATE "NETSERV//193" ""
Essential reading for sending to EARN, BITNET etc.
- o The file "FILE\$ LIST\$" from NETSERV.
\$ transfer rl.earn::"FILE\$ LIST\$" F.LIST "NETSERV" ""
Essential reference list for further files that may be required in
order to send to EARN, BITNET nodes etc.

Software

- o PSI is purchased from DEC.
- o Coloured Books is purchased from DEC UK.
- o JANET_MAILSHR is free on request from Phil Taylor at RHBNC
(Royal Holloway Birkbeck New College) CHAA006@RHBNC.VAXA
- o NRS Database and updates are usually forthcoming from
SUPPORT@UK.AC.RL.VE

GLOSSARY

=====

ARPA and ARPANET

ARPANET is a large computer network whose development was sponsored by ARPA, the Advanced Research Projects Agency of the US Department of defense. ARPANET is now operated by the Defense Communications Agency on behalf of ARPA and other sponsoring US government agencies.

BITNET

Technically, EARN, BITNET (USA), NORTHNET (Canada) and GULFNET (Middle East?) are a single network. Access to them all from JANET is through the JANET-EARN gateway at the Rutherford Appleton Laboratory.

Blue book

The original UK definition of the FTP protocol for file transfer.

CAMTEC PAD

see JNT-PAD

CCITT (Comite Consultatif International Telegraphique et Telephonique)

An international consultative committee that sets international communications usage standards.

Character-Mode DTE

Same as non-packet-mode DTE.

Coloured Books Software

A set of protocol definitions developed in the UK Academic Community to allow for open systems communication in advance of the definition of international standards. Distributed for VAX/VMS by DEC under licence, and copyright to UWIST (University of Wales Institute of Science and

Technology) and St Andrews University. In the documentation it is often referred to as the FTP package, comprising FTP, PAD and POST facilities. Blue book (FTP), Grey book (POST), Green book (PAD) and Yellow book (Transport Service) are in here somewhere, but not Red book (JTMP).

DARPA (Defense Advanced Research Projects)
same as ARPA

DCE (Data Circuit-Terminating Equipment)
A CCITT X.25 term referring to the network equipment that provides functions to establish, maintain and terminate a connection and handle the signal conversion and coding between the data terminal equipment and the network. The switching exchange of the network to which DTEs are connected. (In non-X.25 usage, the term is synonymous with "modem").

DECnet
The collective name for the software and hardware products that allow various DIGITAL operating systems to be interconnected to form computer networks. DECnet is a LAN, typical use is "\$ SET HOST VAXA".

DTE (Data Terminal Equipment)
A CCITT term referring to the user's equipment (computer or terminal) connected to a DCE on a packet switching network for the purpose of sending and/or receiving data.

EARN (European Academic Research Network)
A computer network between academic institutions like universities or research centres for the purpose of world-wide communication. It consists of a set of independent host nodes connected by means of public leased lines in an open, unmeshed topology. A central computer in each country provides international connectivity and some central services. EARN provides world-wide connections to Western Europe, the Middle East, U.S.A. (BITNET), Canada (NORTHNET), Japan, South East Asia, and JANET. Its existence is due to the support of IBM, who have carried the costs for international and some national PTT-lines as well as for the central node computers of the EARN backbone network for the duration of four years. The UK has a single EARN node at the Rutherford Appleton Laboratory. This node is also connected to JANET. A gateway is provided between the two networks which allows the transfer of files and mail.

FTP (File Transfer Protocol)
The definition of the mechanism for transferring files across a network, used in Coloured books. Now used as a verb by the cognoscenti. For FTP package, see Coloured books.

Gateway
The connection between two individual packet switching networks. The connection provides a link through which a DTE can communicate with a DTE on a different network. A gateway is covered in CCITT Recommendation X.75. It might be a small dedicated computer, or it might be a single process among many running on a bigger computer.

Green book
The Green book defines preferred PAD parameters, and specifies the PAD service available using Transport Service (see TS29).

Grey book
The JNT Interim Mail Protocol used in the UK for mail transfers.

Host

A computer connected to a network which will provide you with services.

IPSS (International Packet Switchstream)

Operated by British Telecom. IPSS connects the UK public data network PSS to other national networks, in particular to Telenet and Tymnet in the US. There are real-money charges for using PSS and IPSS.

JANET (Joint Academic Network)

A computer network for the use of academic and research institutes in (mostly) the UK. It has connections to virtually all such sites although in many cases these connections are to local area networks to which the various nodes are connected. There are gateways to the public X.25 network PSS, to ARPA and to EARN. JANET is an X.25 network and provides facilities for file transfer, mail, interactive services and job transfer although the services on any system depend on the particular machines and on the local management. The protocols used are the so-called 'Coloured Books' ones. The network is expected to migrate to the use of ISO (International Standards Organisation) protocols over the next few years.

JANET_MAILSHR

This is a (free) "fix" to VMS mail that you can install on your VAX, provided that it runs Coloured Books, in order to use network mail from inside VMS Mail. The advantage is that you can retain the environment of VMS Mail (mail files, folders, auto-copy, distribution lists etc.) to organise your network mailing.

JNT

The UK Joint Network Team, sponsored by the Computer Board and SERC to manage academic network developments.

JNT-PAD

The JNT-PAD is based on a microcomputer which runs software for connecting terminals and computers on packet switched networks in accordance with the X.3, X.25, X.28 and X.29 protocols. CAMTEC Electronics Ltd. is the company contracted by JNT to build it.

JTMP (Job Transfer and Manipulation Protocol)

Otherwise known as Red Book, it is an optional extra to Coloured Books and requires a separate licence. It allows the user to generate and control tasks which run at remote network sites.

LAN (Local Area Network)

NETSERV

In common with other EARN backbone nodes, the Rutherford EARN node offers NETSERV which provides an EARN information service. The information files may be accessed with 'Blue Book' file transfer from JANET nodes. Those with accounts on the Rutherford IBM computer may access NETSERV as if they were on an EARN node.

Node

A terminal in a communications system. It may also be a switching point where information changes its format or method of transmission.

Non-Packet-Mode DTE

A DTE that is unable to handle data in packet form. This DTE must interface through a Packet Assembly/Disassembly (PAD) facility to connect to a PSDN. Also known as a Remote X.29 Terminal.

It is usually the terminal at which you are logged-on.
(Sometimes called character-mode DTE or start-stop-mode DTE).

NRS (Name, Registration Scheme)

This defines the organisation, structure and translations of names used in the UK Academic Community.

Packet

The unit of data switched through a PSDN; normally a user data field accompanied by a header carrying destination and other information.

Packet-Mode DTE

A DTE which can handle data in packet form. This implies a capability for assembling and disassembling packets. A computer is one type of packet-mode DTE.

Packet Switching

A way of transmitting messages through a communications network whereby long messages are split into shorter packets. Each packet is passed from the source machine via nodes to the destination machine. At each node the packet is received in its entirety, stored and then re-transmitted onwards.

PAD (Packet Assembly/Disassembly)

This is the hardware or software interface between a user's (asynchronous) terminal and a PSDN. That is, a PAD assembles the user's input characters into packets for network transmission, and disassembles packets of output characters into their component characters for output on the terminal.

Protocol

A set of rules that enables cooperating computer processes to communicate over a data network.

PSDN (Packet Switching Data Network)

A generic term used to refer to both public and private networks. For public networks, it takes on additional meaning as a set of equipment and interconnecting links that provides a public packet switching communications service to subscribers within a particular country. An alternative to a PSDN is a PSTN (Public Switched Telephone Network), the voice network generally reached by dialling a telephone.

PSI (Packetnet System Interface)

DEC's name for their X.25 software package.

PSS (Packet SwitchStream)

PSS is run by British Telecom and includes a packet switching network service as well as a set of PADs in various cities. There are real-money charges for using PSS and IPSS.

PTT (Postal, Telephone, and Telegraph Authority)

A country's nationalised communications system, for example British Telecom or the German Bundespost.

Red book - see JTMP

SERCNET

The private packet switched network operated by SERC that has been superseded by JANET.

Start-Stop-Mode DTE
Same as non-packet-mode DTE.

Transport Service
A protocol that allows addresses to be specified that span networks.
An alternative to using X.25 directly. (see TS29).

TS29
A variant of the Triple-X protocols that allow transparent terminal connections across different networks. In particular, Coloured books uses TS29 to make the Cambridge Ring usable alongside JANET.

Triple-X (XXX)
A contraction of X.3 / X.28 / X.29, the protocol standards documents defining the operation of a PAD, and hence terminal control, in an X.25 network. PSS terminals and hosts, and some JANET terminals and hosts, use Triple-X.

X.3
A CCITT recommendation that specifies the Packet Assembly/Disassembly (PAD) facility in a public data network.

X.25
A CCITT recommendation that specifies the interface between Data Terminal Equipment and Data Circuit-terminating Equipment for equipment operating in the packet mode on public data networks.

X.28
A CCITT recommendation that specifies the DTE/DCE interface for a non-packet-mode DTE accessing the Packet Assembly/Disassembly (PAD) facility in a public data network situated in the same country.

X.29
A CCITT recommendation that specifies procedures for the exchange of control information and user data between a packet-mode DTE and a Packet Assembly/Disassembly (PAD) facility.

X.75
The CCITT recommendation that specifies the procedures for communicating between PSDNs ie. gateways.

Yellow book
The definition of the Transport Service currently used in the UK Academic Community.

Elias Eliopoulos

The Astbury Department of Biophysics, Leeds

DHFR is one of the most widely studied enzymes; crystal structures are known for the enzyme from several different species and in complexes with cofactor NADPH and/or a variety of inhibitors. The main reason for this interest is that several clinically important drugs act by inhibition of this enzyme and it is hoped that structural studies may lead to more potent and less toxic inhibitors. Methotrexate (MTX), a widely-used anti-cancer drug, is an analogue of the natural substrate dihydrofolate. Trimethoprim (TMP), which contains the same diaminopyrimidine moiety as MTX but contains a trimethoxybenzene fragment instead of the benzoyl glutamate fragment of MTX, is an antibacterial agent whose efficacy arises from a remarkable selectivity between bacterial and mammalian DHFR. Pyrimethamine, a phenyl triazine, is an antimalarial which selectively inhibits the DHFR from the malaria parasite.

3-d structures are known for the following complexes of DHFR:

E.coli DHFR-MTX	1.7A	refined by Matthews, Kraut et al. at La Jolla.
L.casei DHFR-NADPH-MTX	1.7A	refined "
Chicken DHFR-NADPH	2.0A	refined "
Chicken DHFR-NADPH-TMP	2.2A	refined "
Chicken DHFR-NADPH-phenyltriazine	2.9A	"
E.coli DHFR-TMP	2.3A	refined by Beddell, Champness and Stammers at Wellcome Research Laboratories and in collaboration with La Jolla.
E.coli DHFR-NADPH-TMP	3.0A	Wellcome Research Laboratories.

Just recently the group at La Jolla have reported the structure of an E.coli plasmid DHFR which is totally different from all the above chromosomal DHFR's.

For the purposes of drug design it is, of course, important to know the structure of the mammalian enzyme, both for the design of tight binding inhibitors and to ensure that inhibitors of DHFR from microorganisms do not bind to the host enzyme. We have been collaborating with the group at Wellcome in studies of DHFR from the mouse L1210 tumour cell line and now report refined structures for the NADPH-TMP and NADPH-MTX ternary complexes.

Both complexes crystallise in space group P2₁ with unit cell dimensions a=43.7A, b=61.6A, c=41.9A, $\beta=116.3^\circ$ for the TMP and a=44.1A, b=61.7A, c=42.6A, $\beta=115.7^\circ$ for the MTX complex. Rotation camera data were collected to 2.0A for the TMP complex and 2.5A for the MTX complex. The structure of the TMP complex was initially determined by molecular replacement using the already known chicken liver DHFR. The unrefined mouse DHFR TMP model, together with the structure factors for both the TMP and MTX complexes were supplied to us by the Wellcome group.

Starting with the TMP complex, four rounds of Hendrickson-Konnert restrained least squares refinement were performed using PROLSQ with refitting and model building of the water structure carried out between rounds. The progress of the refinement is shown on figure 1. The final structure contains 1878 non-hydrogen atoms in total including 290 water molecules. The R-factor is 18% for the range of 10.0-2.0A. The mean uncertainty in atomic positions is 0.1A (displayed in figure 2 in the form of a Luzzati plot). Table 1 shows the

detailed statistics of the least squares refinement at cycle 52.

The structure of the MTX complex was determined using molecular replacement of the TMP complex. Two rounds of Hendrickson-Konnert restrained least squares refinement with interpolated model reconstruction led to an R-factor of 24% with a mean uncertainty in atomic position of 0.3Å for data extending between 10 and 2.5Å. The water structure is currently being fitted.

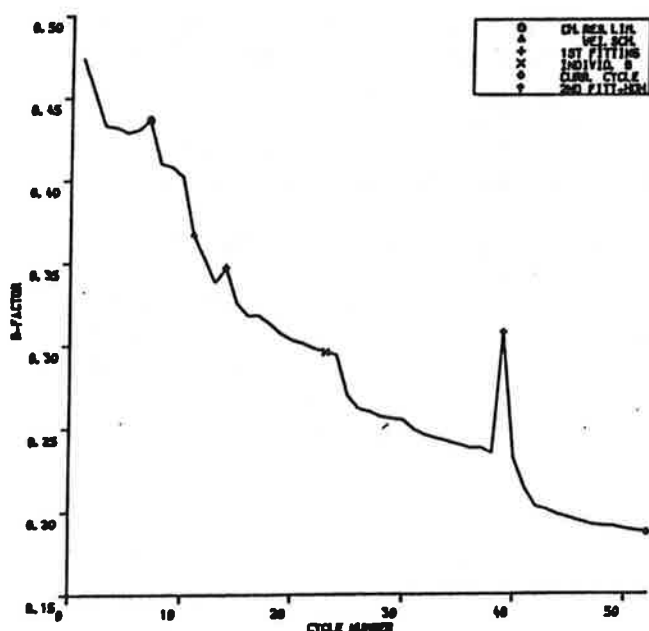
There are significant differences in the mode of binding of both TMP and MTX between the bacterial and mouse enzymes and between the binding of TMP to the chicken and mouse enzyme; we shall describe these in detail in a later report. It is of interest to note that the conformation of bound TMP is a local energy minimum conformation of the isolated molecule; conformational analysis of MTX is currently underway.

There are 21 sequence differences between mouse and human DHFR with no apparent insertions or deletions. 16 of these residues occur on the surface of the molecule and none occur in the substrate binding site. We have modelled the human sequence onto the mouse structure using FRODO and now believe we have a reasonable model for the human enzyme to use as a template for inhibitor design.

Elias Eliopoulos

Department of Biophysics
University of Leeds

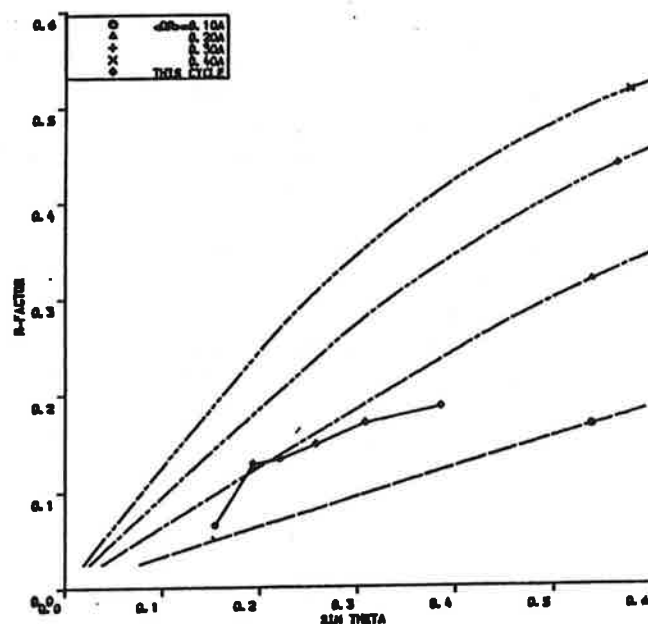
L1210-TMP R-FACTOR PLOT
CYCLES 1-52



R-FACTOR VS REF. CYCLE
FOR
RESOLUTION= 2-5 Å

Figure 1

L1210 LUZZATI PLOT
CYCLE 01210A14



R-FACTOR VS RESOLUTION
FOR
RESOLUTION= 2-5 Å

Figure 2

Table 1

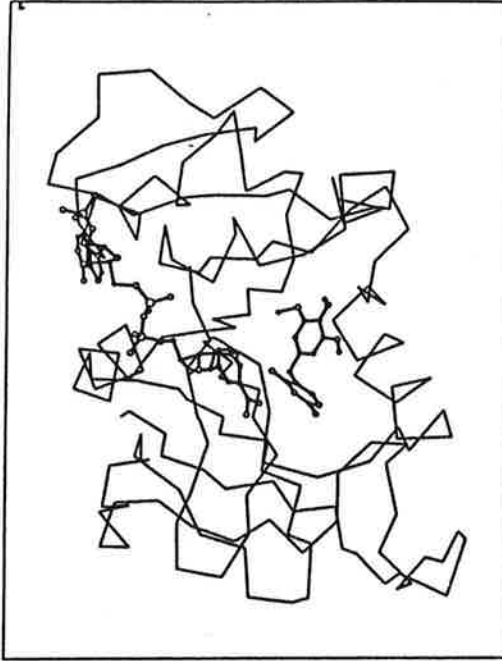
TABLE OF REFINEMENT PARAMETERS

Restraint group and restraint parameters	Standard deviation ⁺	r.m.s. Δ	Worst Δ	Number of parameters
Bonding distances (\AA)				
Bond length (1-2 neighbour)	0.02	0.030	0.09	1628
Angle-related distance (1-3 neighbour)	0.05	0.104	0.16	2202
Intra-planar distance (1-4 neighbour)	0.05	0.102	0.18	567
Planar groups				
Deviation from plane (\AA)	0.02	0.010	0.04	263
Chiral centres				
Chiral volume (\AA^3)	0.02	0.02	0.04	233
Non-bonded contacts (\AA) [*]				
Single torsion	0.2	0.246	0.742	509
Multiple torsion	0.2	0.299	0.705	928
Possible hydrogen bond (X...Y)	0.2	0.307	0.698	130
Torsion angles (deg.)				
Specified (e.g. helix ϕ and ψ)	-	-	-	-
Planar (e.g. peptide ω)	5.0	1.5	5.0	197
Staggered (e.g. aliphatic χ)	20.0	20.7		301
Transverse (e.g. aromatic χ_2)	20.0	24.2		20
Thermal factors (\AA^2)				
Main-chain bond (1-2 neighbour)	1.0	6.7		800
Main-chain angle (1-3 neighbour)	1.5	8.7		1023
Side-chain bond	1.0	7.1		828
Side-chain angle	1.5	9.3		1179
Restraints against excessive shifts				
Positional parameter (\AA^2)	0.3			
Thermal parameter (\AA^2)	0.6			
Diffraction data				
Structure factor modulus		26.7		8067

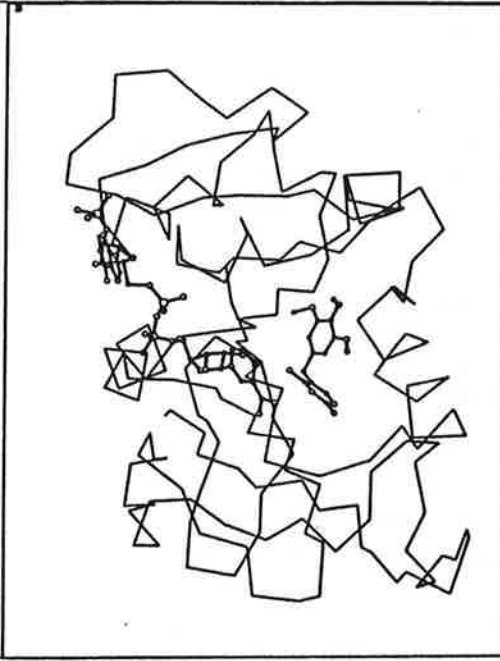
- ⁺ Each class of observation is assigned a standard deviation σ related to the desired breadth of the distribution. Weights w are expressed as variances ($w = 1/\sigma^2$). Δ = Deviation (modulus) from ideal residue. n.d. = not determined.
- ^{*} Ideal H-bond distances are O...O, 2.8 \AA : N...O, 2.85 \AA : N...N, 2.9 \AA . Non-bonded contact distances are ('...(' , 3.7 \AA : N...N, 3.1 \AA : O...O, 3.0 \AA : S...S, 3.6 \AA : ('(Carbonyl)...(' (carbonyl), 3.4 \AA . Distances for hetero-contacts are the averages of the relevant homo-contacts. These distances are modified by subtracting 0.3 \AA for contacts involving a single torsion angle.

Atoms	1828	Dmin \AA	Refs	Shell R	Sphere R
Variables	7313	5.0	3	0.064	0.064
R.Factor	0.188	4.5	301	0.129	0.129
		4.0	483	0.128	0.128
		3.5	799	0.139	0.133
		3.0	1395	0.173	0.148
		2.5	2423	0.223	0.171
		2.0	2663	0.258	0.188

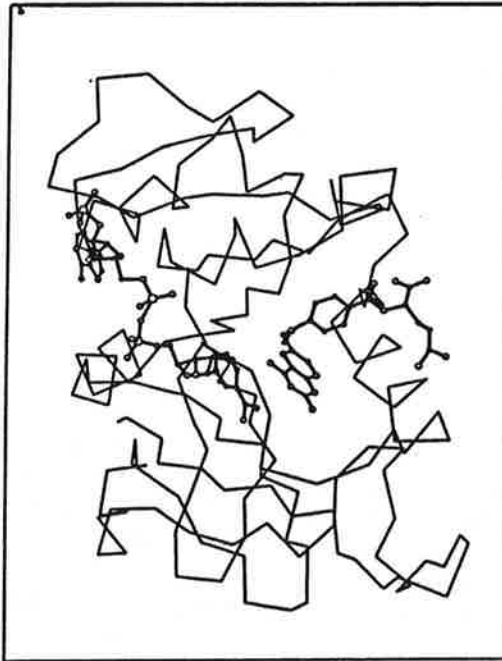
L1210 BONE WITH TOP AND HADPH



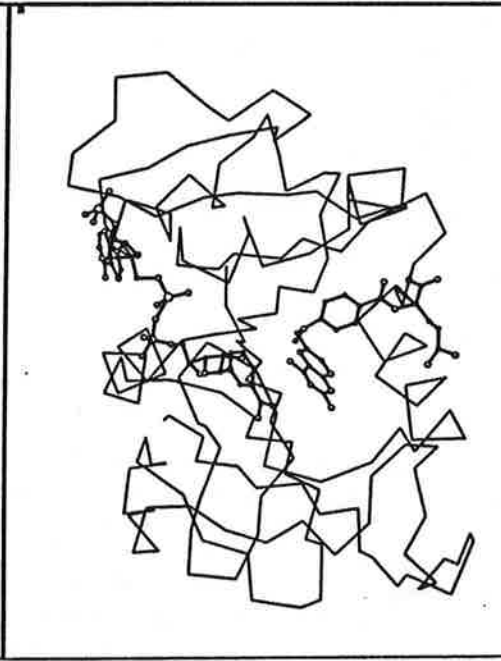
L1210 BONE WITH TOP AND HADPH



L1210 BONE WITH KIX AND HADPH



L1210 BONE WITH KIX AND HADPH



Using FRODO to build molecular models into mir electron density maps

Peter Brick, Kim Henrick, Andrew G.W. Leslie & Alan J. Wonacott.

Imperial College, London.

Recently Alwyn Jones has made available a new version of FRODO which contains two features not available in the standard version of the program (Ref. 1).

The first of these is an implementation of Jonathan Greer's algorithm for producing a skeletonised representation of an electron density map (Ref. 2). On the PS300 this skeleton is represented by bonded "bones" atoms, where the bones are individually labelled and are placed at grid points in the electron density map. Bones atoms are classified as belonging to the "main-chain" or "side-chain" (represented by different coloured bonds), and these atoms can be moved, bonds broken and remade or their classification changed using options on the graphics menu. The program stores both the coordinates and connectivity of the bones atoms.

When building a molecular model into an mir map the "bones" option is used in conjunction with a second new, "fragment-fitting", feature. This allows the user to match fragments from known structures to the skeleton structure. The fragments, which are typically 4 to 8 residues in length, are selected from a database containing 36 well refined protein structures. In this way it is possible to build a complete molecular model out of many fragments, each of which will automatically have good geometry and acceptable conformational angles.

The new FRODO has been used at Imperial to build molecular models of chloramphenicol acetyl transferase (CAT, MW 25,000) and xylose isomerase (MW 48,000). The CAT mir map was based on six derivatives with an overall figure of merit of 0.72 to 2.7 Å resolution. For xylose isomerase, four derivatives gave an overall figure of merit of 0.75 to 2.7 Å resolution prior to non-crystallographic two-fold molecular averaging. To make best use of the program we found the following strategy useful.

The first step is to display the electron density together with the skeleton and then edit the "bones" structure (skeleton) to eliminate connectivities which are clearly incorrect, and reclassify bone atoms as main-chain or side-chain where necessary. False connections between parallel strands of β -sheet or midway along helices can easily be identified. Once this has been done, the "main-chain" bones atoms alone are displayed, which should immediately reveal the location of α -helices and β -sheet in the structure, and allow the overall topology of the molecule to be determined.

In order to make use of the "fragment-fitting" procedure, further editing of the bones atoms is usually required. Individual bones atoms at tentative α -carbon positions are moved within the main-chain density so that the α -carbon bone to α -carbon bone distance lies within 0.2 Å of the correct spacing (3.8 Å). Omitting this step usually results in the program attempting to fit a shorter fragment than that required. Once this has been achieved, the user selects the α -carbon bone

atoms that are to be used in fitting the fragment.

The program lists the best twenty fits from the database (characterised by the rms deviation in α -carbon positions) and allows the user to inspect each of these individually on the graphics display, where the fragment from the database is represented by polyalanine. The user selects the fragment which best fits the map, both in terms of having an acceptably low rms deviation from the bones structure and also having the C^α to C^β bond lying within the side-chain density. If the map is good enough, the fit of the carbonyl groups to the density can also be assessed (this was the case for at least 50% of the residues in CAT). In the structures mentioned above, the rms deviation was typically in the range of 0.3 to 0.7 Å, depending on the length of the fragment.

When fitting α -helices, the fragment chosen from the database was always that of the highly regular helix in crambin (residues 7 to 16) even if this did not give the lowest rms deviation. This is justifiable because the accuracy in positioning the α -carbon bones atoms is, in our structures at least, certainly not adequate to determine whether or not the helix is slightly irregular. The fit of helical fragments to the density was generally so good that the only further adjustment required was to the side-chain torsion angles.

Fitting extended-chain segments was generally rather more difficult. Although rms deviations were generally lower than for fitting helices (because β -sheet conformations are more variable), the β -carbons were sometimes not in the side-chain density and this required correction using the "FBRT" menu option prior to adjusting side-chain torsion angles.

The fragment-fitting was most valuable in modelling turns, since the type of turn is not, of course, known in advance.

In spite of the need to edit the α -carbon bones prior to fragment fitting, this approach allows a model with good stereochemistry to be built significantly faster than the standard FRODO program. In addition, because the main-chain skeleton can be used to examine the connectivity and overall topology of the molecule, there is no need to prepare a "mini-map" which is normally used to follow the polypeptide backbone. Indeed the dominant TIM-barrel structure of xylose isomerase was recognised without ever preparing a stacked map, and although a mini-map was prepared for CAT it was essentially superfluous.

Inevitably there are some features of the procedure which could be improved. In particular, it is difficult to distinguish between the backbone amide and the β -carbon of the N-terminal residue of the displayed fragment, so occasionally this residue is inserted the "wrong-way round". Perhaps a more serious limitation is that it is not currently possible to select only a part of the fragment for insertion into the model. This is particularly frustrating when using the "DGLP" menu option, which allows fragment-fitting to specified α -carbon atoms in part of the model that has already been built. As an example, if residues 101-107 are part of a β -strand (101-102), turn (103-105), β -strand (106-107) structure, then it is sometimes convenient to build the β -strands first (ie including residues 101-102 and 106-107) and then try to fit the three residues in the turn itself (103-105). This is particularly true if the turn density itself is poor, in which case the turn

residues may be excluded until after the first round of refinement. By specifying the positions of the α -carbons of residues 101–102 and 106–107, the program will try to find a 7 residue fragment which best fits these four α -carbon positions. However, if the fragment is accepted, all seven residues will be replaced, which will frequently position the side-chains of the four residues already built in the wrong conformation.

In spite of these minor frustrations, the new FRODO is a very valuable aid in the initial stages of model-building and map interpretation. This makes it all the more regrettable that the many valuable features added by Phil Evans are not available in this version. Our pragmatic, though somewhat clumsy solution is to use the new version for the initial model-building and then to switch to Phil Evans' version during subsequent model-refinement.

REFERENCES

1. Jones, T.A & Thirup, S. "Using known substructures in protein model building and crystallography." *EMBO J.* 5, 819–822, (1986).
2. Greer, J. "Computer skeletonization and automatic electron density map analysis." *Methods in Enzymology*, 115, 206–224, (1985).

CCP4

THE AREA DETECTOR

Guy Dodson, Chemistry Department, University of York

The laboratory has purchased a Nicolet/Xentronics X400 HS area detector. The device arrived in the middle of October and is now being installed in the Physics Department under the close and attentive eye of John Helliwell. Commissioning will be carried out soon by Ron Burns, the original designer of the system.

The Xentronics detector is a high resolution multi-wire system using Xenon as the ionisable gas. It operates with a Xenon pressure of 4 atmospheres a radial and drift space the Beryllium window, which is curved; these features largely abolish parallax in the imaging of the crystal diffraction pattern (over an aperture of 10 cm). The characteristics of the chamber are optimal over the wavelength range 1.5 - 2.2Å, which includes the radiations most commonly used or considered for macromolecular studies in the home laboratory.

In deciding to buy the Xentronix for York we were conscious that there are advantages in complementing the FAST system already established at the SRS and on the Continent. The FAST is now operating at the SRS where its capacity to record short wavelength radiation efficiently makes it an especially ideal system. There is also a FAST system at Birkbeck and of course others at Jan Drenth's laboratory at Groningen and in Robert Huber's laboratory in Munich. We ourselves will be happy to make the Xentronics available to other University laboratories on an individual basis once we have established its general working characteristics in the York environment.

GGD18

CCP4ip.DO

A BRIEF DESCRIPTION OF THE IMAGE PLATE AS AN AREA DETECTOR FOR
PROTEIN CRYSTALLOGRAPHY

J R Helliwell, University of York and SERC, Daresbury Laboratory

In the past ten to twenty years a considerable effort has been expended in the development and implementation of electronic area detectors for X-ray crystallographic data collection from protein crystals. The requirements of these devices are exacting and cover the following aspects: number of resolution elements, full aperture and pixel size, absorption efficiency, noise, uniformity of response, linearity, dynamic range, spatial distortion and of course concerns as to cost, complexity and reliability. The technical challenge has widened following the increased interest in using different wavelengths at SR sources, and indeed all X-ray wavelengths for Laue diffraction. A lot has been achieved using the multi-wire proportional chamber and television devices and the fields of particle physics and observational astronomy have provided much experience. The drawbacks with existing devices commonly available involve a limited active area, insufficient dynamic range, noise, spatial distortions, non-uniformity of response or limited count rate capability. Good data is being collected but it has to be said that the ultimate detector system has not yet been realised. New possibilities in these areas exist of course; for example the use of the charge coupled device or new MWPC configurations (eg. the multi-stack idea of Jeavons).

Fairly recently a new technology in area detectors has become available. It originally developed for diagnostic radiography [Sonoda et al Radiology 148 (1983) p833]. There is considerable interest in this new detector known as the IMAGING PLATE. A variety of tests have been made on the plate by the X-ray diffraction group at the Japanese Photon Factory and reported by them recently [Miyahara et al NIM A246 (1986) 572-578]. Will this be the ultimate device?

The imaging plate is a flexible plastic plate coated with fine photostimulable crystals, BaFBr:Eu, of 150 micron thickness. It is manufactured by the Fuji Photo Film Co., Ltd.. The absorption of X-rays causes the formation of F-centres and which results in the storage of an image for a period of time (half life about 8 hours). The illumination of the centres by a laser beam results in luminescence well matched to the quantum efficiency of a photo-multiplier tube. The properties of the plate appear to overcome the limitations of existing area detectors. In particular it has a large active area (250 mm x 200 mm is standard), a pixel size of about 0.4 mm at the 1-2% peak light, a high absorption coefficient for X-rays

from Mo K α to Cu K α , a very low noise, a high dynamic range and good linearity. It is 10x more sensitive than DEF-5 film. It is of course re-useable!

There are computational simplicities for protein crystal image plate data acquisition and processing. In particular a uniform response and lack of spatial distortion avoids the need for look-up tables or use of polynomials for pixel by pixel corrections in routine data collection. The large aperture and dynamic range avoids the need for repeated settings of the detector and hence would reduce the total exposure or number of samples; this would represent a considerable saving in the volume of data that needs to be marshalled around. These simplifications are expected to accrue both conventional and SR source monochromatic data collection. For time resolved crystallography work with the monochromatic SR beam the data rate will be limited by the online image plate drum scanner. Hence film will probably still be used here unless a large number of plates is configured in an automatic cassette changer. For Laue SR data collection it is likely that an array of plates will be needed in a single exposure to collect the especially dense Laue diffraction patterns; the high absorption efficiency of the plate over the wavelength range used in broad band Laue time resolved experiments would reduce the exposure times by about an order of magnitude. On the new SRS protein crystallography workstation which is currently being developed (financed by the Swedish Research Council) millisecond data collection times are envisaged in Laue mode. If an image plate is used these would be reduced to 100 microseconds for a near complete protein crystal data set (high symmetry case).

The development of an image plate area detector system for routine data collection is being pursued by the Rigaku Corporation (under license from Fuji). Their system includes the use of two plates so that as one plate is exposed the other is read on a drum scanner. Precise interchange of the two plates takes place under computer control. Apparently sales of the system will start in March 1987 at a price of roughly £100,000 (excluding X-ray generator or data processing computer). Details should be obtained from Rigaku.

GGD19

THE TRANSPUTER AND MOLECULAR SIMULATIONS

David Fincham, Chemistry Department, University of York

"Molecular simulation" is a term covering a variety of techniques used to study a variety of problems. The standard methods are Metropolis Monte-Carlo and Newtonian molecular dynamics, but there are a number of variations such as stochastic dynamics and non-equilibrium molecular dynamics. Applications range over the whole of solid and liquid state physics, materials science and biophysics.

Simulation is a very expensive business in terms of the computer power required, and I believe it can only achieve its full potential as a powerful tool, particularly in many practical problems which are of interest to industry, when it can be made much cheaper. Currently most simulations are performed on supercomputers, even though the demand for memory and backing store is usually modest. The principal requirement is for raw CPU power, and it is clear that much more cost-effective computing can be achieved by the use of parallel or multi-processor designs. The ICL DAP has already played a very prominent role in molecular simulation, particularly for lattice models. I am now investigating whether another British invention, the Transputer, will be able to provide the next step forward in this field.

The Transputer is a very powerful 32 bit microprocessor, designed and produced by INMOS. It is a reduced instruction set computer (RISC) with a basic processing speed of 10 Mips (million instructions per second). A single transputer has been benchmarked at six times a VAX 11/780. For the technical, the device uses advanced CMOS 1.5 micron technology. There is an on-chip memory of 2 kbytes, and an interface to external memory chips (static or dynamic). The speed of this interface is 25 Mbytes/s. With its links (see below) the Transputer is a complete computer on a chip.

The most distinctive and important feature of the Transputer is the set of links or communication channels. Each Transputer has four links. A link can communicate with the outside world, by means of a link adapter chip which converts it to a standard serial or parallel interface. Alternatively, Transputers can be connected together by means of their links. The four links work simultaneously, each at 20 Mbits/s. By connecting Transputers together it is possible to create multi-processor computers of almost arbitrary speed, memory and peripheral bandwidth at modest cost. Although such parallel

computers can be built using conventional microprocessors (e.g. the INTEL hypercube) problems arise in achieving effective inter-processor communications.

What about programming: can we use multi-Transputer arrays effectively? The Transputer directly implements Occam, the parallel processing language originating with Professor Hoare's group in Oxford. The basic concepts in the language are the process and the channel. A program consists of a hierarchy of processes, from the individual assignment statements up to the complete program itself. At each level processes are explicitly executed sequentially or in parallel. Thus Occam is a much more flexible language than, say, DAP Fortran, which only allows parallelism at the statement level. Parallel processes communicate by means of channels, one process sending data over the channel and the other receiving it: the data are only sent when both processes are ready, so that synchronisation is automatic. In the case of processes running on different Transputers the channels are, of course, implemented over the links. However, parallel processes and channels execute in a logically identical manner on a single Transputer, and algorithms can thus be developed on a single Transputer before being transferred to an array.

At present, Transputers perform fixed point arithmetic in hardware, and floating point arithmetic in software. This means that their floating point performance is not quite as sparkling as my previous remarks might have lead you to expect. There are three solutions to this problem: a) use fixed point arithmetic for simulation, together with energy and force look-up tables: this is the method used in the Delft special purpose molecular dynamics computer; b) attach external additional floating-point chips to the Transputer; and c) wait for the promised floating-point Transputer.

To use arrays of Transputers effectively we need suitable parallel algorithms, and of course our DAP experience is a help here. My current research is examining a range of molecular simulation problems, from the Ising model to protein dynamics, to understand how they may best be implemented. I then plan to program a complete demonstrator system for polymer dynamics. This application is chosen because its heterogeneous interactions make it unsuitable for machines like the DAP; it should effectively test the ability of a Transputer array to do many different things at once.

GGD20

The "EEC Co-operative Programming Workshop on Position-Sensitive Detector Software (Phases I and II)" was initiated by Gerard Bricogne at L.U.R.E. and held there between May 26th and June 7th, 1986. The E.E.C. provided funds to cover both personal and computing equipment rental expenses. The aim of the workshop is ultimately to produce device-independent software for single-crystal X-ray and neutron diffraction intensity data collection by use of an area detector. Initially the scope is to be restricted to the fixed-wavelength method, although expansion to cover the Laue method may be considered later.

Phase I was planned to take 4 days and to cover all mathematical details of the procedures involved, from the characteristics of the incident beam to statistical analysis of the data collected. A full list of the topics considered is shown on the list overleaf. Consideration was also given to existing hardware architectures and associated software interfaces, to the user interface, and to the use of computer graphics. However, owing to the diversity of hardware it was immediately apparent that the software would have to be designed for a "virtual" goniometer and detector, the mapping to the real devices taking place via device-dependent handlers.

A draft specification arising out of Phase I, has recently been distributed by Gerard.

Phase II was planned to take the remaining 7 days, the aim being to produce a complete specification of all the software required, so that participants could be allocated the task, on return to their laboratories, of either writing a specific section of the code, or alternatively of providing it where feasible from routines already written.

Some 25 participants from France, W. Germany, Sweden, U.K. and U.S.A. took part in Phase I, the majority staying on for Phase II. Accommodation was provided in a hotel in Rue Gay Lussac (adjacent to the Jardin du Luxembourg, within walking distance of Paris city centre), and daily travel to Orsay was by Metro (45 mins each way). Evenings were spent in lively discussions over dinner, or beer at a street cafe, or simply sight-seeing in Paris, and I think I can speak for all participants if I say that the experience was most rewarding and enjoyable.

Ian Tickle
Birkbeck College
London

Session Topics.

- 1. Introduction.
- 0. Overall scope of the project.
- 1. Beam characteristics (wavelength, divergence, polarization).
- 2. Crystal characteristics (orientation, mosaicity, absorption).
- 3. Goniometer characteristics (geometry & control).
- 4. Detector characteristics (geometry & response).
- 5. Characteristics of the integrated system (calibration, missettings).
- 6. Coordinate mapping and associated derivatives.
- 7. Prediction of position and shape of reflexions.
- 8. Statistical methods for parameter estimation and tracking.
- 9. Processing of individual frames.
- 10. Dynamic storage on work file.
- 11. Profile estimation and its use in reflexion integration.
- 12. Data collection strategy and its interaction with post-processing.
- 13. Measures of system performance and data quality.
- 14. Connection with film processing.
- 15. Survey of hardware architectures.
- 16. Survey of software interfaces.
- 17. User interface.
- 18. Graphics.
- 19. Recommendations : methods and algorithms.
- 20. Recommendations : programming considerations.

Cytochrome b₁: a bacterial ferritin and cytochrome.

John M.A. Smith

Department of Biochemistry, University of Sheffield, Sheffield, S10 2TN.

X-ray studies on the eucaryotic iron-storage protein, ferritin, have been a feature of the protein crystallographic research carried out in Sheffield. Ferritins, isolated from a number of mammalian and amphibian species have been characterised crystallographically and the structure of horse spleen apoferritin has been refined against medium resolution data. Ferritin is found throughout the animal and plant kingdoms but has never been identified in the procaryotes. In 1979, however, Steifel and Watt (Nature, 279, (1979), 81) recognised several ferritin-like properties in cytochrome b_{557.5} of the bacterium, Azotobacter vinelandii. This, they declared, was the first of a new class of proteins, the bacterioferritins. In common with ferritin, bacterioferritin, when imaged by electron microscopy, has a transparent outer protein shell surrounding an electron dense core of mineral iron. The protein shell is roughly spherical and is formed by the aggregation of several identical protein subunits. Unlike ferritin, bacterioferritin, is a haem-containing protein.

Bacterioferritins have now been identified in 8 bacterial species from 6 genera and assigned tentatively in 4 more. The best characterised bacterioferritin is that isolated from Escherichia coli, for which a partial amino acid sequence is available, and it is with this protein that the crystallographic studies are furthest advanced. Originally identified spectroscopically as the haemochromogen component b₁ of the cytochrome system of E. coli by Keilin in 1934 (Nature, 133, (1934), 290), cytochrome b₁ was isolated and characterised independently by Deeb & Hager (J. Biol. Chem., 53, (1964), 1024) and Fujita et al. (J. Biochem. (Tokyo), 53, (1963), 282). The identity of this protein and the bacterioferritin isolated immunologically by Yariv et al. (Biochem. J., 197, (1981), 171) has only recently been proven with the crystallographic analysis of protein prepared by the two methods. Cytochrome b₁ is a membrane protein, and can be isolated in a soluble form from the particulate fraction of broken bacterial cells, albeit in poor yield, commonly less than 3mg protein per 1 kg bacterial paste.

Our studies on E. coli cytochrome b₁ have been in collaboration with Joe Yariv and Joe Gilboa of the Weizmann Institute of Science, Rehovot, Israel. Following Mossbauer spectroscopic studies on whole E. coli, they isolated and purified a bacterial iron-storage protein, a bacterioferritin, whose Mossbauer spectrum was identical with the dominant Mossbauer signal in the spectrum of the bacteria. The optical absorption spectrum of this protein was indistinguishable from that of cytochrome b₁. Crystallisation trials were undertaken with protein prepared free from non-haem iron by immuno-precipitation. A number of morphologically distinct crystal forms were obtained using different crystallisation conditions and techniques. Those characterised crystallographically include tetragonal, monoclinic, orthorhombic I and orthorhombic II forms of the ferri-haem protein and a tetragonal form of the ferro-haem protein.

Grown from ammonium sulphate solutions containing trace amounts of Mn²⁺, the tetragonal crystals have a regular distorted rhombic dodecahedral morphology, space group P4₂2₁2, cell dimensions a=b=210Å c=145Å, are stable towards X-radiation and diffract to high resolution. A high degree of pseudo-symmetry is apparent in the precession photographs of this form, and at low resolution the pattern can be indexed according

to the pseudo-space group $I432$ with $a=145\text{\AA}$. Although the highest molecular point symmetry required by this space group is 2, with molecules lying on crystallographic diads, the pseudo- 432 crystal symmetry may reflect 432 molecular symmetry. A space group transformation can be effected within these crystals by soaking in solutions of specific heavy metal compounds. The new space group is cubic, $I432$ with $a=146.9\text{\AA}$, and corresponds to the pseudo-space group of the tetragonal form. In this form the molecule has the full 432 point symmetry of the crystal lattice.

In all crystal forms the cell dimensions are consistent with a spherical molecule having a packing diameter in the range $117\text{--}128\text{\AA}$.

For practical reasons, it was decided to pursue the crystal structure of the tetragonal form in which it was hoped to utilize the non-crystallographic symmetry relating the 12 subunits of the asymmetric unit. Unfortunately, the search for heavy atom derivatives was frustrating, the soaks either having no discernible effect or causing the space group change. Finally, conditions were found for uranyl binding which gave small intensity differences only.

The very low resolution ($d>24\text{\AA}$) native tetragonal data were determined from small angle precession photographs. Given that the overall molecular architecture will be reflected in these very low resolution terms, these data were used to refine a hollow spherical model. The best fit to the data had an outer diameter of 120.6\AA , consistent with the range determined from unit cell packing considerations, and inner cavity diameter of 60.8\AA .

The low resolution ($d>6\text{\AA}$) data for native (collected with both crystal settings) and two different uranyl derivative crystals have been collected at the SRS. These data are currently being processed using the VAX package at Daresbury. The native self-rotation function shows a peak of $42/50$ ths the height of the origin peak corresponding to the expected 432 symmetry.

It is likely that the low resolution structure will be available in the near future and it remains to be seen whether the quality of the map will be sufficient to answer some of the puzzling questions posed by this protein. Of greatest interest will be the location of the haem groups in the protein shell. Haem determination has consistently indicated a content of $1/2$ haem per subunit (assuming a subunit molecular weight of 16kDa from SDS-PAGE). This stoichiometry is only reconcilable with the observed crystallographic 432 molecular symmetry if either haem groups lie on 2-fold subunit interfaces with two statistical haem orientations or if only one of any pair of subunits has its haem pocket occupied. In both cases the molecular symmetry is statistically generated. Further questions surround the bifunctional nature of the protein, how it functions as a ferritin when there is no significant homology between the partial sequence of cytochrome b , and known ferritin sequences, and how the molecule is associated with the membrane. The three-dimensional structure will hopefully provide the answers.

LGEN AND LCHK: UTILITY PROGRAMS FOR INVESTIGATING LAUE PHOTOGRAPHS

Mike Elder,
SERC, Daresbury Laboratory, Warrington, WA4 4AD

Two Fortran 77 programs for the ICL PERQ are summarised here:

- LGEN - a program which calculates and displays simulated Laue photographs and contains routines intended to help with their analysis, including an auto-indexing feature;
- LCHK - a program to analyse the coverage of reciprocal space that can be achieved by a series of Laue photographs in different orientations.

LGEN

This program displays a Laue simulation of the Perq screen. The user supplies Lattice parameters (reciprocal cell; lattice type and centering), crystal orientation (a standard orientation; rotation angles about beam, vertical and spindle axes following IDXREF conventions; spindle setting) and other parameters which affect the pattern (crystal-film distance; resolution limit; film radius; wavelength range recorded). The program displays the predicted pattern together with histograms showing the distribution of spots by wavelength and the number of multiplets due to the recording of harmonics.

For smallish crystals or low resolution data the reciprocal lattice coordinates of all possible reflections are stored and successive simulations at different spindle settings may be quickly displayed. If there is insufficient array space to hold all the reflections, or if the user changes a parameter which affects the number of reflections then the pattern is recalculated for each display requested.

There are four auxiliary commands which operate upon the current display.

- dmpscreen:** Plot the current Laue pattern on the attached Toshiba printer.
- ident:** Identify each spot to which the user points showing indices and wavelength for each of its harmonic components. See Figure 1 for an LGEN example illustrating this option.
- expand:** Re-display an enlargement of part of the pattern defined by the user.
- spotin:** Match spots on a Laue photograph taped to the pad and selected by the user with the puck against the displayed image and prepare a file of spot indices and coordinates for subsequent use.

AUTO-INDEXING

LGEN contains an `index` command which may be used to determine the orientation of a randomly orientated crystal from the observed Laue pattern. The user must know the lattice parameters and crystal-film distance accurately.

Indexing works by matching the angles between reciprocal lattice vectors for nodal spots (determined from the coordinates of these spots on the actual film) against a table of angles of potential nodal spots prepared in advance from the crystal lattice parameters.

There are three input methods available

- (1) ordinarily the user points at nodal spots on a film taped to the pad, having first oriented the film by pointing at the fiducials;
- (2) optionally, fiducial and spot coordinates may be read from a file;
- (3) as an aid to user confidence it is also possible to treat a displayed LGEN Laue simulation as the film and to point at fiducials and spots on the screen, in which case the program could be expected to determine the orientation actually used for the simulation.

The auto-indexing routine works well providing the crystal-film distance is known accurately (within about 1 mm) and provided that the user is able to identify nodals with small indices. Here small means less than about 5 since the tabulation of angles is limited to those arising from a total of no more than about 280 vectors. User patience will also be sorely tried if the angle table is large since the brute force matching algorithm is rather slow.

If the crystal is fairly well orientated so that the beam direction is close to a reciprocal axis then it may be hard to identify such nodal spots. In such cases the user will often know the approximate crystal orientation so the trick is to use LGEN to display a pattern that roughly resembles the film, use the `ident` command to identify spots with small indices, and then use `index` together with the option which invites the user to supply indices for the input nodals.

LCHK

This program is somewhat experimental in that it was put together rather quickly and has not been thoroughly tested. However, it is easy to modify and extend and it is my intention to respond to user requests or criticisms as quickly as possible.

LCHK calculates Laue patterns for a given crystal in a given orientation but does not display them. Instead, it determines what fraction of the symmetry unique set of reflections which could be recorded on film (depends on resolution limit, crystal-film distance, wavelength range) would actually be present for a particular orientation.

Input data is as for LGEN with some changes to aid clarity. The user must supply the Laue symmetry group where necessary. Another new parameter is a scale factor which multiplies the input reciprocal lattice coordinates in

order to reduce the time taken to analyse each crystal orientation. Apart from a minor truncation effect it does not change the fraction of symmetry equivalents recorded at a given orientation if the reciprocal lattice is scaled up.

There are a number of commands.

- rotate:** Determine the fraction of the symmetry unique reflections recorded at each of a number of steps about the spindle axis for an otherwise constant crystal orientation. The number of symmetry equivalents recorded at each step is analysed and there is a facility to determine the overlap between spindle steps. This should help to maximise the coverage of reciprocal space whilst minimising the number of photographs that are taken for a given experiment. Figure 2 illustrates this command in use.
- search:** Determine the fraction of symmetry unique reflections recorded at each point on a grid of crystal orientations defined by a number of steps about the two axes (spindle and vertical) perpendicular to the X-ray beam. This should help with the search for a crystal orientation which gives maximum coverage in one shot.
- wavvar:** Analyse the wavelength variation for those symmetry equivalent reflections which are recorded on a particular film in order to determine the contribution which this film would make to the wavelength normalisation process.
- varycf:** Determine the fraction of the symmetry unique reflections recorded for each of a series of crystal-film distances. The greater the crystal-film distance the better the spatial resolution of spots on a photograph: this command helps to determine the trade-off between spatial resolution and coverage.
- resoln:** Analyse the fraction of symmetry unique reflections recorded for different resolution ranges for a particular crystal orientation.

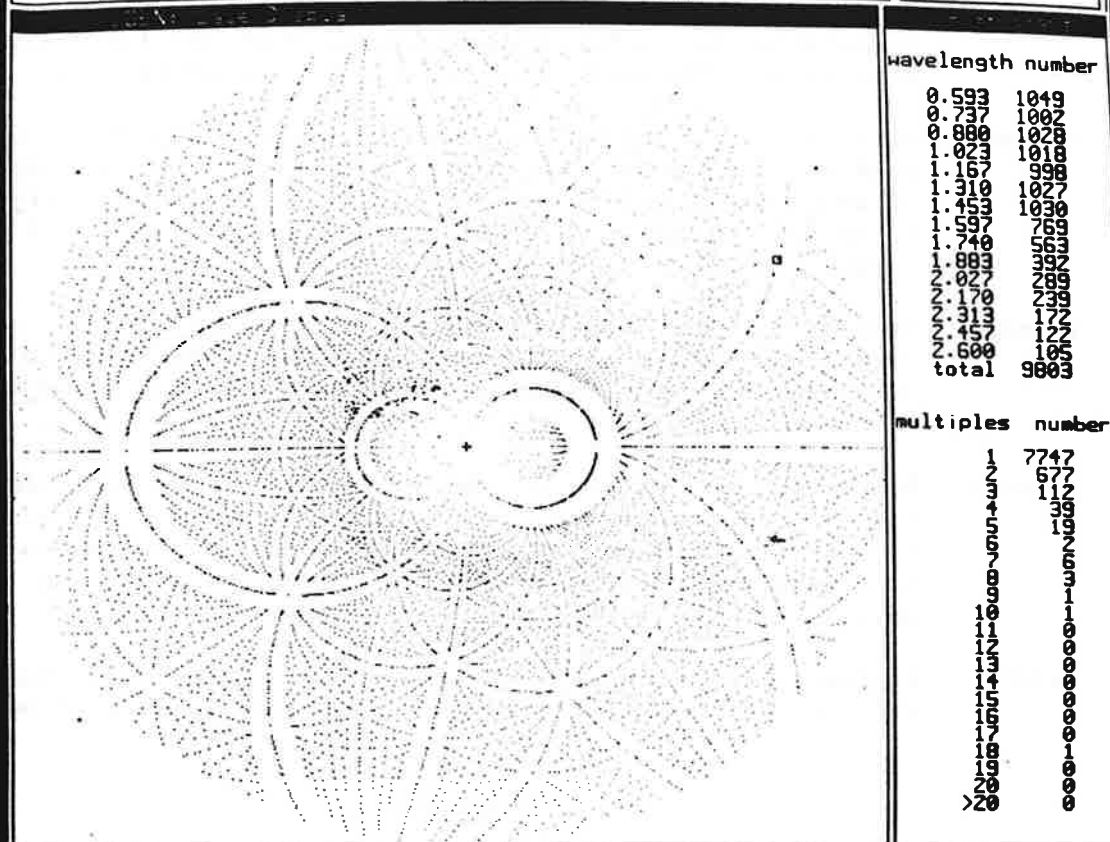
AVAILABILITY

Both programs are written in Fortran 77. They use a number of Pascal routines for window management, screen display and interaction with the screen via the puck. This Perq-specificity limits their portability but they are otherwise freely available. LCHK could be moved to a VAX system fairly easily and the CSE Applications Group at Daresbury could probably be persuaded to make this transfer if there were sufficient demand. LGEN could only be moved with loss of functionality since it is heavily dependent upon the bit-mapped screen.

ACKNOWLEDGEMENTS

Marjorie Harding had the original concept for LGEN. Pella Machin, John Helliwell and Janos Hajdu contributed clever ideas. The help and encouragement of all four is gratefully acknowledged.

datfil	pligen	a*	0.019852	wv min	0.45	spindle	30.0	start
		b*	0.016506	wv max	2.60	d-min	2.6	dmpscr
		c*	0.007380	cf(mm)	95.10	radius	59.3	ident.
		al*		wv-std	1.00	mosaic	0.200	expand
		be*		px(dg)	0.12			spot in
		ga*		py(dg)	0.13			index
savfil		mat	ort +as,+cs	pz(dg)	-0.63			stop



START REFLECTION CALCULATION
limit of 20000 reflections reached at h,k,l= -5 -15 1
do you wish to truncate calculation here (y/n)[n]?
OK. Will plot these, then return for more
Be advised that reciprocal space will not be saved
44.25 27.21: 5-10 5(2.523) 6-12 6(2.102) 7-14 7(1.802) 8-16 8(1.577)
9-18 9(1.401)

Figure 1: LGEN for the protein Pea lectin. The ident command is in use and the spot surrounded by the box cursor has been identified as harmonics of 1, -2, 1.

File	rec p. cell	geometr	d axis	rotate
datfil lchkt	a* 0.007280	wv min 0.45	spindle 0.0	rotate ← →
savfil	b* c* 0.008595	wv max 2.30	d min 2.4	search
sp dir +cs	al*	ct (mm) 90.00	radius 59.0	wavvar
bm dir +as	be*	scale 5.00		varycf
laue 4/m	ga*	px(dg) 15.00		resoln
cntrng P	latt tetrag.	py(dg) 18.00		stop
		pz(dg) 0.00		

Effect of rotating about spindle on the fraction of symmetry unique reflections recorded									
631									
Spindle	Fraction recorded	Distribution of symmetry equivalents							
		1	2	3	4+				
0.0	0.564 (356)	0.398	0.166	0.000	0.000	(461)			
22.5	0.571 (360)	0.414	0.157	0.000	0.000	(459)			
45.0	0.582 (367)	0.425	0.157	0.000	0.000	(466)			
67.5	0.580 (366)	0.425	0.155	0.000	0.000	(464)			
90.0	0.531 (335)	0.350	0.162	0.019	0.000	(461)			
Spndl	Add. fract. symm. uniq. captured	Fract. symm. uniques not recorded	Fractional overlap with previous spindles						
			1	2	3	4	5		
		1.000							
45.0	0.582	0.418							
22.5	0.257	0.162	0.307						
67.5	0.059	0.103	0.372	0.355					
90.0	0.029	0.074	0.288	0.437	0.341				
22.5	0.014	0.060	0.372	0.365	0.312	0.341			

```

Calculate and store the symmetry unique set of reflections
631 Reflections stored
This command analyses the effect of rotating about the spindle axis.
Enter number of steps (counting the present spindle setting as the first)
5
Enter step size in degrees
22.5
Analyse the overlap of the symmetry uniques?
Enter row in above table to use as base (0=end)
3
Analyse the overlap of the symmetry uniques?
Enter row in above table to use as base (0=end)
0

```

Figure 2: LCHK for the protein Pea lectin. The rotate command reveals that over 50% of the symmetry unique reflections are recorded at each spindle setting sampled for this crystal-film distance. Photographs taken at spindle setting of 0°, 45°, 67.5° would include 90% of the unique data.

SERC Daresbury Laboratory.

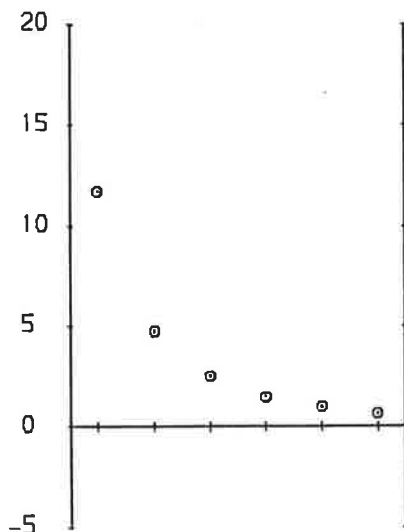
INTRODUCTION

The program LAUENORM (ref. 1) calculates a wavelength normalisation curve for a set of Laue data using symmetry equivalent reflections measured at different wavelengths. The reflection data are split into wavelength bins which are then scaled together using a Fox and Holmes scale factor refinement. The resulting scale factors are then fitted by a polynomial curve which is used to calculate scale factors for the individual reflections. The bin scaling and curve fitting may be iterated until convergence is achieved. This article describes some aspects of the curve fitting part of the procedure illustrated using part of a set of Phosphorylase data collected by Dr Janos Hadju.

THE CASE UNDER CONSIDERATION

The data for wavelength normalisation are normally divided into a number of wavelength ranges avoiding the Silver and Bromine absorption edges and, in this article, only the wavelength range of 0.49 to 0.92 Angstroms is considered. This range has a large change in scale factor and illustrates some of the problems of curve fitting which may arise.

The change of scale factor across the wavelength range is shown in the graph below.

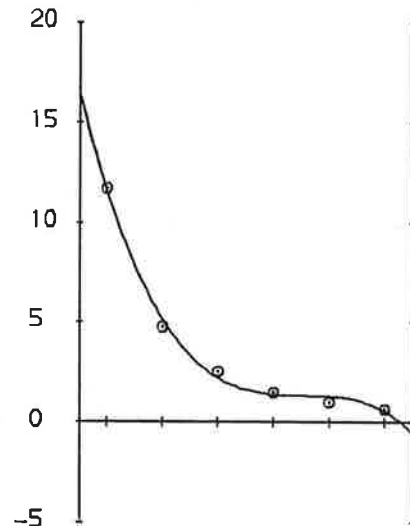


The graph shows a set of refined wavelength bin scale factors for the case where the wavelength range has been split into 6 wavelength bins (more will be said later on the effect of the number of wavelength bins chosen).

THE END OF THE RANGE PROBLEM AREA

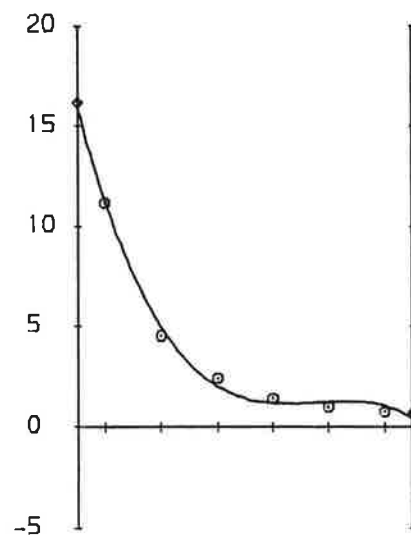
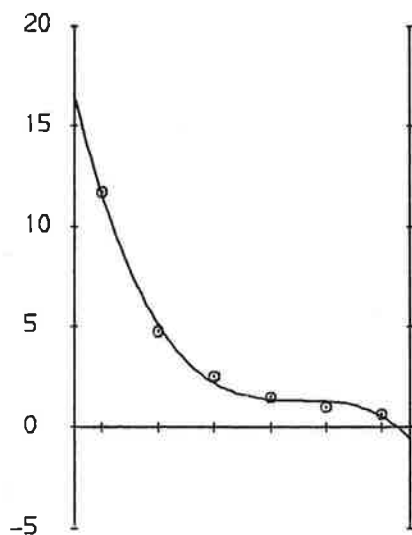
The scale factors between the wavelength bins are taken as being the scale factors for the wavelengths at the mid points of the bins. These scale factors are then fitted using a polynomial curve fitting routine in which the user may choose the order of the polynomial to be fitted. The program indicates the goodness of fit for a series of different orders to help the

user to make a suitable choice. There is however a potential problem in that the first half of the first bin and the second half of the final bin fall outside the range used to determine the polynomial with the result that the polynomial may behave badly in these regions. The problem is illustrated for the case where a third order polynomial is fitted to the set of scale factors shown above. The results of this are shown in the next graph.



At the end of the final bin, the curve trails off disastrously and even gives negative scale factors. This is in spite of the fact that in other respects the third order polynomial fits the points quite well. It is important therefore that the curve fitting at the ends of the range be examined closely even if the curve fitting in general looks satisfactory.

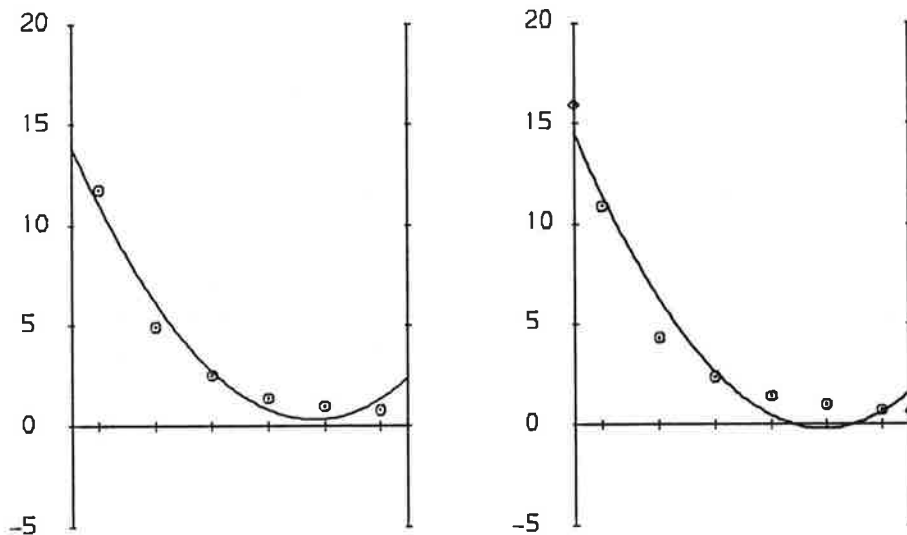
In order to try and improve the behaviour of the curve, a modification was made to the program to estimate two end points from examining the local trend of the scale factors near the ends of the range. The end point at each end of the range is estimated by making a quadratic extrapolation from the three points nearest to the end of the range. The two extrapolated points are added to the set of points before the curve fitting is performed. The results of fitting a third order polynomial without and with the added end points are shown in the graphs below (left and right respectively).



With the added end points, there was still some tendency for the curve to fall off at the high wavelength end of the range but it is generally much better behaved. This is shown in that the R-factor was reduced from 0.223 to 0.196 for the set of data throughout the range and from 0.755 to 0.218 for the data in the high wavelength bin.

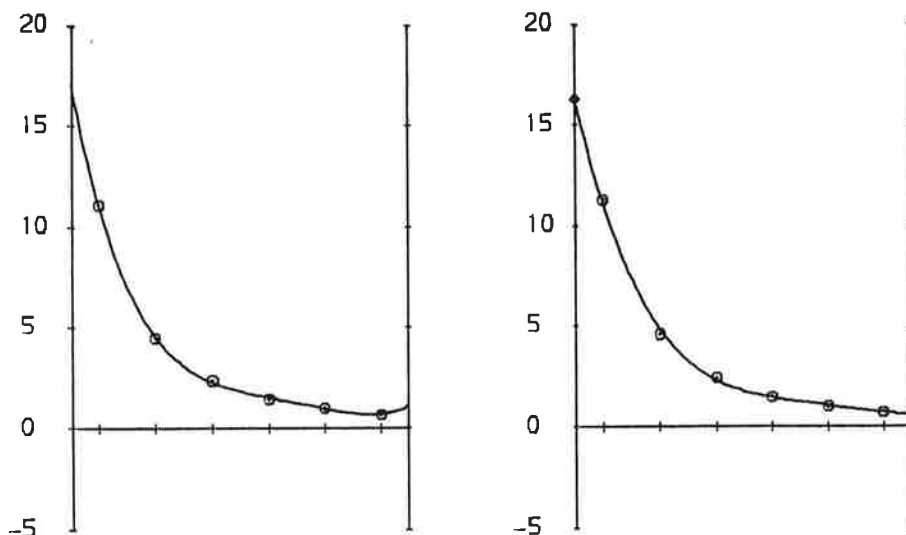
FITTING A SECOND ORDER POLYNOMIAL

During the first attempts at normalising the data in question, they were fitted with a second order polynomial. Even though, from the points to be fitted, it looked as if this might be appropriate, it does in fact give a rather poor fit and again the curve goes off in an undesirable direction at the high wavelength end of the range. In this case adding two extrapolated end points made the situation worse giving negative scale factors and then failing completely. The second order curve fitting without the added end points and an early iteration of the curve fitting with added end points are illustrated in the graphs below (left and right graphs respectively).



HIGHER ORDER POLYNOMIALS

The case of fitting a third order polynomial has already been considered in detail above. It was found that fitting a fourth order polynomial gave much better results and the curve was much better behaved even without any added end points. The addition of end points did however still make an improvement in the R-factors from 0.157 to 0.147 for the range and from 0.247 to 0.182 for the high wavelength bin. The curve fitting is shown in the graphs below.



Polynomials of order 5 and 6 were also tried. It may be noted here that one of the possible advantages of adding the end points is that it allows higher order curves to be fitted and this may be useful where only a small number of bins can be used.

The table below indicates the R-factors for different orders of curve fitting, without and with added end points.

Order =	2	3	4	5	6	
R-factors (range):	0.336	0.223	0.157	0.146	-	no end points
R-factors (range): disaster		0.196	0.147	0.154	0.152	added end points
R-factors (end bin):	0.351	0.755	0.247	0.222	-	no end points
R-factors (end bin): disaster		0.218	0.182	0.188	0.192	added end points

For the 5th order polynomial, the addition of the end points makes the R-factor for the range slightly worse though the R-factor for the end bin is slightly improved.

THE EFFECT OF CHANGING THE NUMBER OF WAVELENGTH BINS

One way to reduce the problem of end effects is obviously to increase the number of wavelength bins. It should be noted however that this may not always be possible as there need to be sufficient reflection overlaps between the bins to enable the inter-bin scale factors to be satisfactorily refined. In some cases it is likely that it may only be possible to use say four bins. For the case under consideration, there were plenty of data and it is hoped that the results, obtained for different numbers of bins and orders of fitted curves, may help in making sensible decisions for other cases. A table of results is given below:

No. bins =	4	6	8	12
Order of curve				
2	0.275 (0.42) 0.328 (0.62)	0.366 (0.35) disaster	0.476 (0.33) disaster	0.376 (0.22) disaster
3	0.192 (0.37) 0.177 (0.21)	0.223 (0.76) 0.196 (0.22)	0.260 (1.74) 0.229 (0.25)	0.288 (-0.9) 0.240 (0.39)
4	- - 0.177 (0.21)	0.157 (0.25) 0.147 (0.18)	0.163 (0.24) 0.159 (0.18)	0.164 (0.21) 0.166 (0.20)
5	- - 0.155 (0.16)	0.146 (0.22) 0.154 (0.19)	0.146 (0.19) 0.143 (0.16)	0.144 (0.19) 0.143 (0.17)
6	- - -	- - 0.152 (0.19)	0.150 (0.22) 0.147 (0.17)	0.142 (0.17) 0.144 (0.17)
	[430]	[198]	[101]	[40]

Key: R-factors R1 (R2)
 R3 (R4)

where R1 = R-factor for range not using added end points
 R2 = R-factor for end bin not using added end points

 R3 = R-factor for range using added end points
 R4 = R-factor for end bin using added end points

The numbers in square brackets give the number of reflections in the end (high wavelength) bin.

In some cases the R-factors at the end of the iteration were a bit higher than the lowest values at an earlier stage. This occurred particularly for the lower orders of polynomial where no end points were added.

When fitting a fairly low order polynomial (e.g. 3 or 4), an increase in the number of bins used makes the results worse. To obtain the benefit of increasing the number of bins it is necessary also to increase the order of the polynomial fitted. For the set of data considered, the fifth order polynomial would probably be the most appropriate to choose provided that sufficient bins were chosen.

CONCLUSIONS

- 1) The choice of curve fitting parameters is important in getting good scaling of the data as judged by R-factor analyses. It should therefore be done with some care.
- 2) Do not choose too low an order for the curve fitting and use at least a third order polynomial for the wavelength range considered. This means that at least 4 bins of data are likely to be required before reasonable results can be obtained.
- 3) It is important to examine closely the behaviour of a fitted curve at the ends of the range and to look at the R-factors for the data in the end of range bins. Bad fitting in these regions may not be obvious if the data are only considered as a whole.
- 4) In general it is best to use a reasonably large number of bins e.g. 8 provided that there are sufficient reflection overlaps between the bins to give good inter-bin scale factors.
- 5) If the number of bins is increased, it is important to increase the order of the polynomial fitted if a low order was used initially. Otherwise worse rather than better results may be obtained and problems of bad end of range behaviour may not be overcome.
- 6) If only a few bins can be used then the addition of end of range points as described in the article may be of considerable benefit. Thus it is possible to get an R-factor of 0.155 with 4 bins (using added end points and fourth order polynomial) compared with the best value of 0.142 with 12 bins and a sixth order polynomial.

- 7) If a large number of bins is used then it is probably not a good idea to add the end points, as random errors in the scale factors at the three points closest to the end may give misleading information about the local change of scale factor in that area.
- 8) The curve fitting process, showing the curve behaviour and goodness of fit, is much clearer in the accurately plotted graphs in this article than in the line printer output from the program. The program could probably, therefore, be improved in this area.

REFERENCES

- 1) "Determination of the Wavelength Normalisation Curve in the Laue Method" by J. Campbell, J. Habash, J.R. Helliwell and K. Moffat, Information Quarterly for Protein Crystallography, 18, July 1986.

Intensity Data Collection from Laue Photographs
A Progress Report

by Marjorie Harding and Steve Maginn, Department of Inorganic, Physical and Industrial Chemistry, Liverpool University
and Pella Machin, John Campbell and Ian Clifton, Applications Group, Daresbury Laboratory

Over the last two years programs and procedures have been developed for the measurement of reflection intensities from SR Laue diffraction patterns ^{1,2}. We report on progress with such measurements, towards the structure determination of an organometallic compound.

A complex of composition $\text{RhFe(CO)}_2\text{Cl(dppee)}$ where dppee is $\text{Ph}_2\text{P.C(=CH}_2\text{).PPh}_2$ was prepared in association with Dr A K Smith (IPI Chemistry Department, Liverpool); the largest crystal grown had dimensions $320 \times 60 \times 50 \mu\text{m}$, and since we could not easily get diffractometer intensity measurements from it we decided to record Laue photographs and attempt structure determination this way - as a pilot project for work with smaller crystals. The unit cell and symmetry were determined from oscillation and Weissenberg photographs, taken with CuK radiation. The axial ratios have been refined from the Laue photographs using the program GENLAUE.

monoclinic,

a = 18.237 Å formula weight 731

b = 19.806 Dm = 1.62 g cm⁻³

c = 8.256 Dcalc = 1.63 g cm⁻³

$\beta = 86.53^\circ$

e.s.d of ratio of cell edges, or angle $\sim 0.1\%$

e.s.d. of cell edges overall $\sim 1\%$

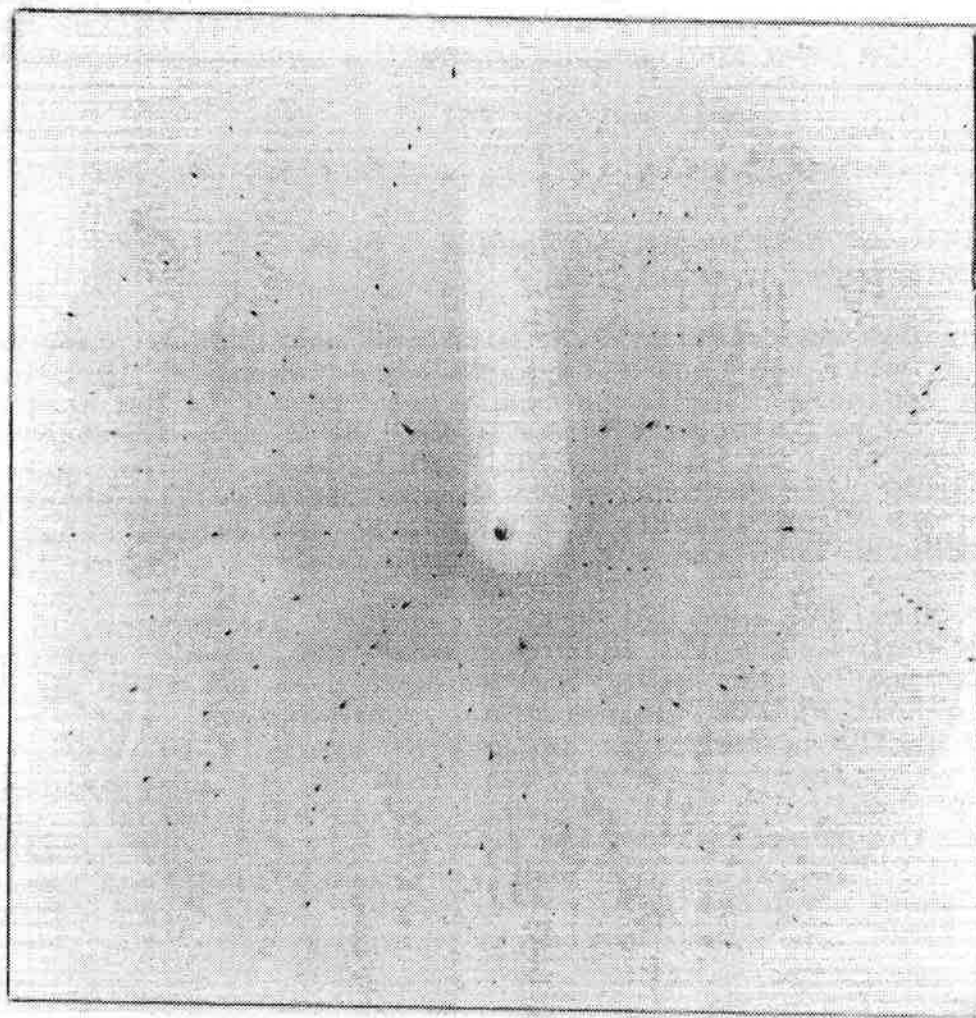
On the SRS workstation 9.7, with the oscillation camera set up by John Helliwell, ten Laue film packs were taken, at 18° intervals around the spindle (spindle parallel to c). The crystal film distance was 64 mm, exposure times 6 sec at 2 GeV, 158-146 mA; each film pack consisted of F,F,Al,F,Al,F,Al,F,Cu,F. The films were scanned ($50 \times 50 \mu\text{m}$ raster). A sample film is illustrated; the reflections are significantly elongated, an indication that mosaic spread is greater than for a typical good quality crystal.³

Pattern Prediction and Integration: Cell parameters, as well as orientation angles, crystal film distance etc, were refined with GENLAUE, typically to give an rms deviation of $< 0.10 \text{ mm}$ for 100 spots. The soft limits $d_{\text{min}} = 1.20 \text{ \AA}$, $\lambda_{\text{min}} = 0.30 \text{ \AA}$, and $\lambda_{\text{max}} = 2.1 \text{ \AA}$ were judged by the presence or absence of appropriate reflections in the pattern. Inspection of the final F_o F_c list suggests that these were probably not well judged. Integration of all predicted reflections to $2\theta_{\text{max}} = 43^\circ$ was done by MOSLAUE. The reflections are elongated radially; the frame, for background and spot, was square, 0.85mm across; it was the largest available in MOSLAUE at the time, but was not quite big enough for the strongest

reflections.

Film pack scaling: AFSCALE determined and applied the wavelength dependent scale factors (as well as L_p factors). At this stage too, the doublet and triplet spots were resolved by UNSCRAM.⁴ These yielded 300-350 singlet reflections and 30-50 intensities from doublet and triplet spots, with R factors (on intensity) between successive films in a pack in the range 10-15%

Wavelength Normalisation and Scaling/merging of all film packs: Each reflection intensity must be multiplied by a wavelength-dependent factor $1/f(\lambda)$ and a pack scaling factor $c(i)$, both of which may be determined empirically; the 3707 reflection intensities measured represent 830 unique reflections. It has been shown that for a protein film wavelength normalisation could be done satisfactorily within a single film pack, where the total number of reflections was 5000.⁵ Here, with fewer observations, the derivation of $f(\lambda)$ and $c(i)$ were done together, with a new program LAUENORM2. It is assumed that $f(\lambda)$ will be the same for all film



packs, and thus the number of observations to relate different wavelengths is adequate. Tables 1 and 2 give more details. Reflections measured at wavelengths close to the Ag and Br absorption edges were omitted; it was also found that measurements at $<0.55\text{\AA}$ seriously degraded the data so they, too, were omitted. The Rh absorption edge at 0.53\AA and the large variation of absorption below this may provide an explanation for the poor agreement of these reflections with the rest.

Some statistics:

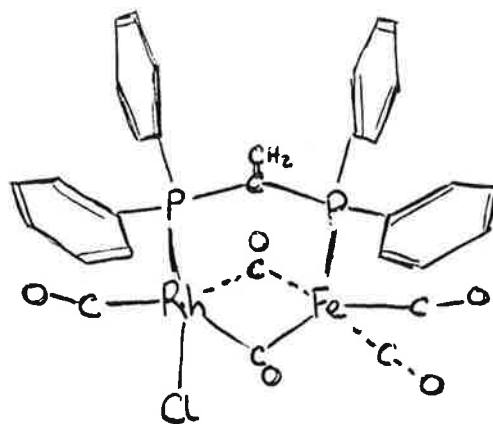
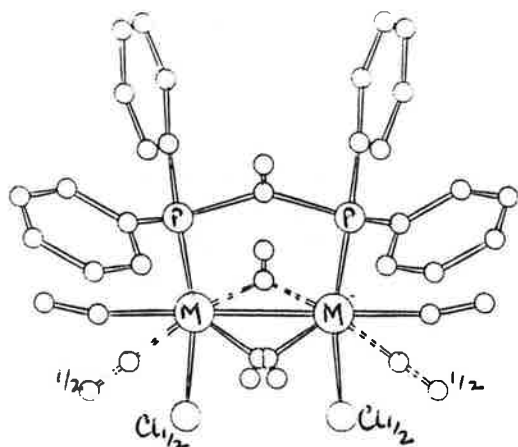
3707 reflection intensities measured,
838 intensities $< 1.5 \sigma(I)$ omitted,
431 measurements at $\lambda < 0.55\text{\AA}$ omitted,
334 measurements at $\lambda > 1.55\text{\AA}$ omitted, (they showed poor agreement in AFSCALE)

so 1318 measurements for $\lambda = 0.55$ to 0.91\AA were used, together with
786 measurements for $\lambda = 0.922$ to 1.55\AA .

These yielded 754 unique reflections, with R_{merge} (on I) = 0.14, of which 122 depended on the unscrambling of doublets and triplets. The data set lacks all axial reflections and a number of others with simple indices such as hh0, hhh, 2hhh, because these were only recorded as parts of quadruplets or higher multiplets. This has not caused difficulty in the interpretation of this simple Patterson series, and they are not essential for the least squares refinement.

Structure Solution: The apparent spacegroup was $B2_1$, or $B2_1/m$ (which could have been converted to $P2_1$ or $P2_1/m$, but we preferred the choice of axes with c parallel to the needle axis and camera spindle). $B2_1/m$ requires molecular symmetry m or $\bar{1}$; the former is unlikely, the latter impossible. A Patterson series yielded Fe and Rh positions, 2.8\AA apart, with the Fe-Rh vector parallel to the 2-fold screw axis. The first electron density map gave chemically sensible positions for two P atoms and Cl, three of the phenyl groups, and several carbon and oxygen atoms; refinement (in $B2_1$) reached $R=0.17$, but the remaining atoms were not evident and the structure did not deviate significantly from one with $B2_1/m$ symmetry. Therefore solution and refinement were repeated in space group $B2_1/m$. The program SHELX^a was used throughout: the metal atom was treated as Rh with a site occupancy of 0.8 (equivalent to the mean of Fe and Rh); in the refinement, benzene rings were given ideal geometry, P-C, C-C, and C-O bond lengths were restrained to take values close to those expected; anisotropic vibration parameters were used for Cl but no other atoms (the metal atoms do not appear to be particularly anisotropic). At present the best description of the compound is that shown over.

The F data appeared to be rather sharp and refinement lead to some negative B values. A damping factor, $\exp(-8\sin^2\theta/\lambda^2)$ was applied to all $F(hkl)$; then B values in the normal range for this type of compound were obtained ($B=4-10\text{\AA}^2$). No explanation has yet been found for this effect. The present R factor is 0.17 for 632 reflections all measured as singlets, or 0.20 for 751 reflections



The structure found (left) is assumed to represent the molecule on the right together with its mirror image.
(Dotted lines to poorly defined carbonyl groups.)

when the doublets and triplets are included. The Fe-Rh, metal-P and metal-Cl bond lengths are close to expected values, and the molecular packing is sensible. The chlorine appears to have a very large anisotropic vibration, and the location of two of the carbonyl groups is not entirely satisfactory because of the disorder.

Discussion: This is, we believe, the first attempt at intensity measurement for the determination of an unknown structure from SR Laue diffraction patterns, and it has been substantially successful. There are details of the structure which we have not yet established (two carbonyl positions) because of the disorder problem, and the high R factor can be attributed partly to this, but mainly to shortcomings in the intensity measurement procedures.

We can now see where improvements should be made. It is evident that reflections where the agreement between F_o and F_c is worst are ones with simple indices, e.g. $-1,0,3$ $-2,0,6$ which were ostensibly measured as singlets; this strongly suggests that our choice of d_{min} , λ_{min} and λ_{max} was not correct, and that some of these reflections are in fact multiplets. The films will be reprocessed with attention to this, and with more appropriate box sizes and profiles in the integration.

It is unfortunate that our trial crystal has turned out to be a disordered one, but it is very consistent with the accumulating evidence that small crystals are often poorly ordered in some way³.

A tendency to form a partially disordered structure would account for our difficulty in growing a bigger crystal.

References:

1. Information Quaterly of CCP4, No 15, Daresbury Laboratory, July 1985.
2. M.Elder, M.M.Harding, J.R.Helliwell, P.A.Machin et al. in preparation for Acta Cryst.
3. S.J.Andrews, J.E.Hails and M.M.Harding, Acta Cryst. (1987) A43, in the press.
4. S.Zurek et al, Information Quaterly of CCP4, No 16, 1985
5. J.Campbell, J.Habash, J.R.Helliwell and K.Moffatt, Information Quaterly of CCP4, No 18, July 1986.
6. G.M.Sheldrick, SHELX, A Program for Crystal Structure Determination, University Chemical Laboratory, Cambridge, 1976.

TABLE 1

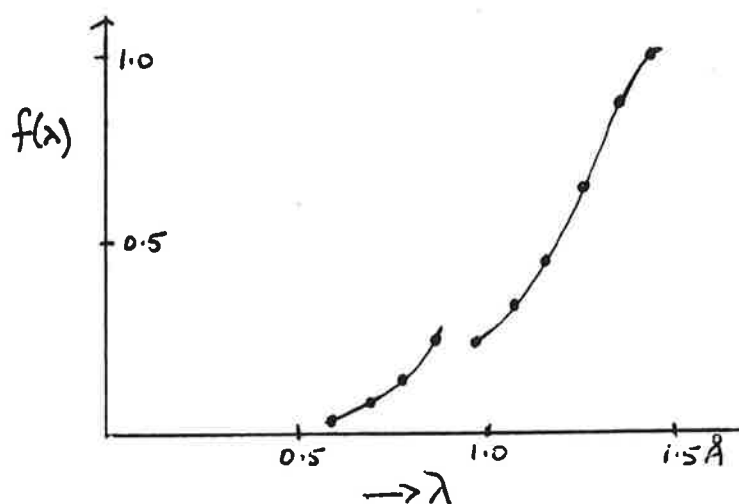
Overlap analysis - numbers of equivalent reflections measured at different wavelengths

λ bin λ mean	1 0.595	2	3	4	5	6	7	8	9	10 1.452
1	68	157	188	223	133	94	65	43	30	11
2		88	174	238	134	109	61	25	25	17
3			111	231	113	105	51	34	19	9
4				119	150	121	55	48	24	16
5					40	92	45	38	24	13
6						41	39	33	20	13
7							19	33	20	15
8								20	23	9
9									3	14
10										1

TABLE 2

Wavelength Normalisation curve found

mean	$1/f(\lambda)$	number
0.595 A	18.62	307
0.685	8.90	331
0.775	4.61	320
0.865	2.49	360
0.970	2.44	256
1.066	1.77	184
1.163	1.26	131
1.259	0.89	97
1.355	0.66	62
1.452	0.57	42



CCP4 Program Suite Changes

There is now a CCPHELP help library available which uses the standard Vax help facility. It gives information on the CCP4 programs.

The following is a list of programs which have been updated during the last 6 months. Would anyone requiring the updated versions of any of these programs please get in touch with me (Network address: CCP4@UK.AC.DL.DLVD).

PRTMAP

FFT81 Changes to make program work for space group P3121 (Number 152)

TSEARCH

REFINE

RFACTOR - old version will not run on Vax

MAKEDATA

GENSFC

HKPROLSQ.UTX updated to describe keyword control data

Sue Bailey.

AT LAST! A structure from the FAST

A collaboration between DL and the Universities of Edinburgh,
York and Liverpool

We have very recently analysed a data set collected in August on the SRS of a small molecule/small crystal.

A sample of Si3O7(2-) C4N2H6(2+) was received from Edinburgh with unit cell and space group. The CAD4 at Edinburgh gave only 25 measureable reflections on a Mo source.

A crystal of dimensions ca. 18x8x175 um was mounted on a glass fibre and placed on the FAST with the SRS running under standard multibunch conditions ($\lambda=0.9\text{\AA}$).

The orientation was obtained with our autoindexing routine which gave a triclinic cell equivalent to the conventional C-centred monoclinic lattice. (The standard CAD4 routines have now been incorporated which will do this). No particular axis was aligned with the goniostat axes.

The MADNES software was used to collect a 180 degree data set in 0.5 degree frames with a flat detector and no on-line analysis, the images being dumped to disk. Beam monitoring was carried out with the PX data acquisition software. The refined mosaic spread of the crystal was ca. 0.5 degree.

A second data set was obtained with a tilted detector over the same phi range with a higher detector gain, HVCA=4, HVII=4 (in the middle of the detector range), to collect the higher angle reflections (dmin ca. 1\AA).

Unit cell: a=13.57 b=4.90 c=22.46
beta=91.67

Space group: C2/c

Integrated intensities were extracted from the images using the off-line version of MADNES at which stage preliminary scaling was achieved using the beam monitor files. Reflection indices were transformed to the above cell and the systematic absences removed. Error flag rejection, application of symmetry operators and conversion to LCF files was performed with MADLCF written by John Campbell. This was run twice -

- 1) without weak or negative intensities for scale calculation with ROTAVATA:
(1919 refs. read, 908 rejected with error flag)
- 2) using all reflections for direct methods:

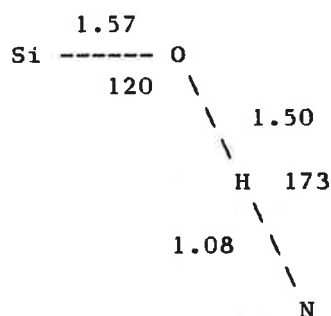
Each data set was sorted with CCP4:LCFSORT. 90 degree batches were used in ROTAVATA and these scale factors were applied to

the second sorted output via AGROVATA. This yielded 739 independent data but with a merging R-factor of only 22% on I, due to the great number of poorly measured weak reflections. These data will be profile fitted in the near future. With the weak and bad reflections omitted AGROVATA gave R = 8% with 446 independent data.

Data were reformatted using the LCFCARD utility in CCP4 and the structure solved using MULTAN80 running on defaults, implemented on the AS7000 at DL.

Structure refinement was carried out with SHELX76 also on the AS7000 to a final crystallographic R of 10.3%, weighted R of 9.4%, all non-H atoms having anisotropic thermal parameters (511 data with $F > 2\sigma(F)$, reflection data:refined variable ratio = 5:1). Hydrogen atom positions were calculated.

The structure (drawn with PLUTO on the SAS VAX at DL) is seen to consist of two-dimensional silicate nets parallel to the (ab)-plane, these layers being held together through the organic moiety by strong interaction between:



The nitrogen atoms were distinguishable in a difference Fourier from neighbouring carbons.

Although the data is obviously in need of some improvement it is encouraging to see that the combination of the FAST, the SRS, autoindexing and standard programs implemented at DL can be used to solve small molecule/small crystal structures.

Steve Andrews

