

An Introduction to the CCP4 Software Suite: CCP4i, Files and Utilities

Peter Briggs
CCP4, CCLRC Daresbury Laboratory
p.j.briggs@ccp4.ac.uk

ACA 2004 Chicago
July 17th 2004

An introduction to the CCP4 software suite

Aims of this presentation:

- Provide an overview of the non-crystallographic aspects of the software
- Give inexperienced users an overview to get you started with CCP4
- Surprise more experienced users with some functions they didn't know about

Outline of this presentation

Overview of the CCP4 software suite

- What's new in CCP4 version 5.0.1
- Installing and using

Introduction to CCP4i the CCP4 graphical user interface

- Overview
- Project management tools
- Customisation

Overview of CCP4 file formats

- MTZ files
 - *Projects crystals and datasets*
 - *Data harvesting*
- File utilities
 - *Viewing*
 - *Manipulations*

CCP4 Resources

Overview of the CCP4 software suite

CCP4 suite consists of ~175 programs covering all aspects of macromolecular structure determination including:



- Data processing and reduction (**MOSFLM** & **SCALA**)
- *Experimental phasing*
- *Molecular replacement*
- *Density modification*
- Refinement (**REFMAC5**)
- Building (**CCP4mg**)
- Validation and analysis (into deposition with the PDB)

Much of the software is contributed by developers and scientists not funded by CCP4 and it is through their continued generosity and goodwill that the project survives!

Philosophy of the CCP4 software suite

- **Modular:**

- Each program covers a small range of functionality
- Data passed between programs via data files in standard formats
- Keywords control program function and provide additional data
- User decides on the sequence of programs to use for a particular task, e.g.

Data reduction:
Mosflm  Scala  Truncate

- **Inclusive & “redundant”:**

- Includes a number of different programs to do the same job
- Allows user to choose from different approaches

Downloading and installing the CCP4 software

Download from `http://www.ccp4.ac.uk/download.php`

- Installation instructions at `http://www.ccp4.ac.uk/dist/INSTALL.html`

Can build from source code:

- useful for customised installation

Binary installations are easiest:

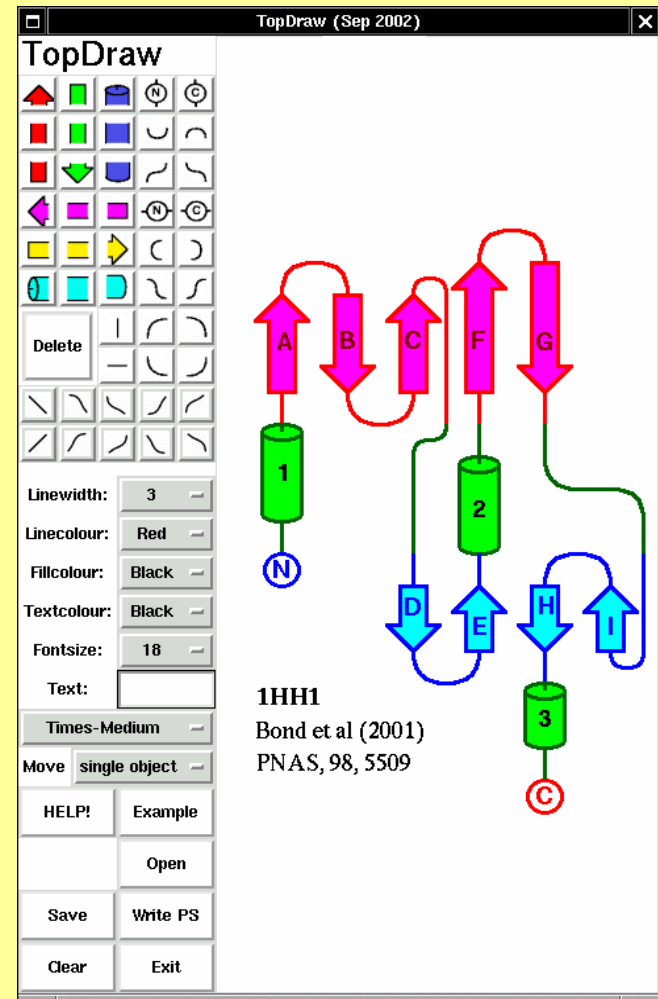
- For Macintosh and Windows: use the self-extracting packages
- On Windows:
 - recommend removing any previous installation first
 - admin privileges are required to install
- For Linux, Irix, OSF1/TruUnix64, SunOS
 - use `download-5.0.1.sh` script to download and install automatically

Note about licensing

- all users of 5.0+ need to agree to and return a new licence
- even if you already hold a licence for an earlier version

What's new in CCP4 5.0.1

- **topdraw** - sketchpad for drawing protein topology cartoons (see right)
- **dtrek2scala** - convert unmerged D*TREK data to input into Scala
- **bulk** - bulk-solvent correction for translation search in AMoRe
- **ncont** - search for protein contacts
- **pdbcur** - manipulate PDB files
- **tlsextract** - TLS parameters from PDB REMARKS
- **pdb_extract** - extract deposition information from logfiles (from RCSB-PDB)
- plus new major new core libraries



Running programs via scripts – an example

Program name *Input & Output files specified as
logical name-file name pairs*

Command line { `fft HKLIN toxd.mtz MAPOUT toxd_aupatt.map <<eof`
 Keyworded script { `TITLE Native patterson for Au derivative`
 `PATTERSON`
 `AXIS Y Z X`
 `RESOLUTION 100 2.5`
 `LABIN F1=FAU20 SIG1=SIGFAU20 F2=FTOXD3 SIG2=SIGFTOXD3`
 `.....`
 `END`
 `eof`

- Chapter 3 of the CCP4 manual covers this in detail
- Also lots of example scripts in the **\$CEXAM/unix/runnable/** directory
- Unix variants only – Windows uses graphical interface exclusively

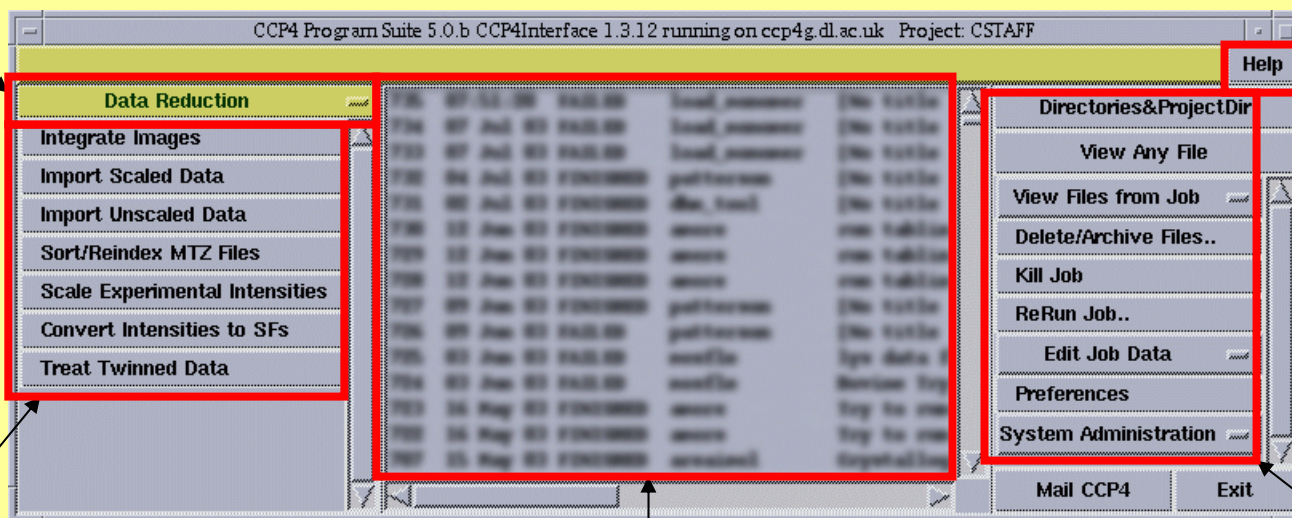
Introduction to CCP4i – graphical user interface

- Graphical user interface hides details of running programs
- Sits on top of the programs (user not locked-in)
- “Task-driven” rather than “program-driven”
- Key features:
 - Easy-to-use interfaces to major programs and utilities
 - Tools for file viewing and basic project management
 - Customisable
 - Integrated help system
- To start up
 - Unix: type `ccp4i` at the command prompt
 - Windows: launch using the CCP4 icon in the Start Menu
- Requires that Tcl/Tk and BLT are installed

CCP4i main window – quick tour

Modules

On-line help



Tasks

Job Database

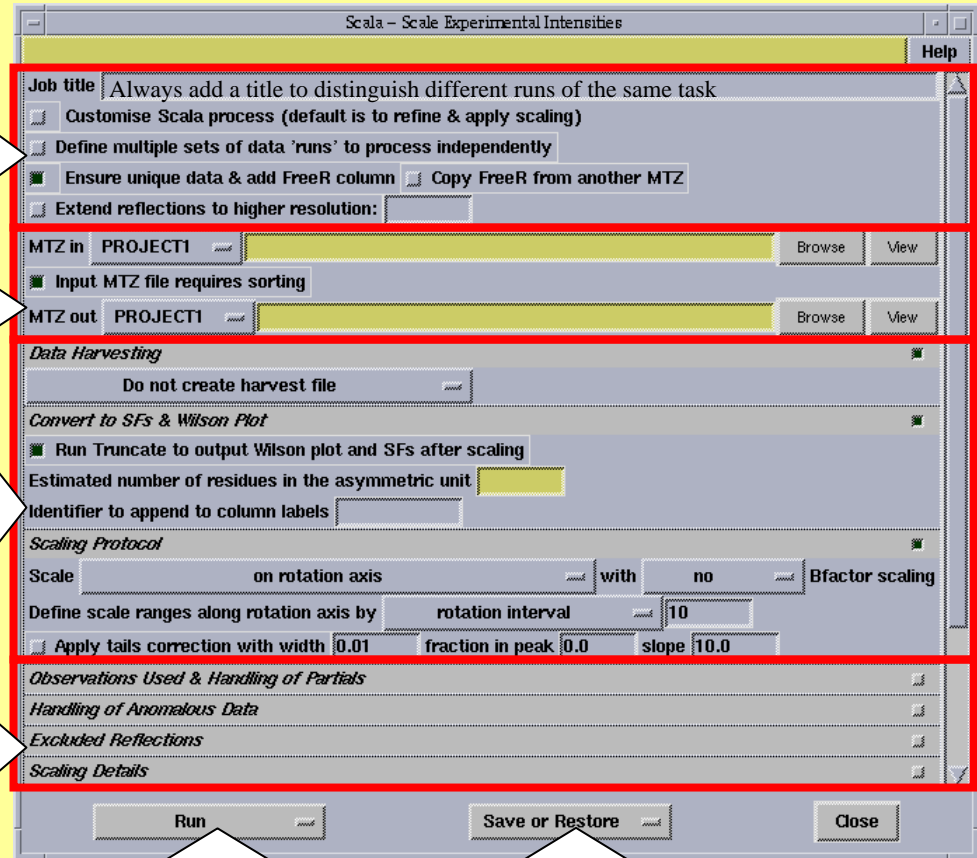
Tools & Utilities

Example of a CCP4i task interface

WORK FROM THE TOP DOWN



- Protocol folder**
Make the key decisions
- File folder**
Set input and output file names
- Open folders**
Parameters that should be checked by the user
Highlights indicate compulsory input
- Closed folders**
Advanced/infrequently used



- Run task
- Save/restore parameters

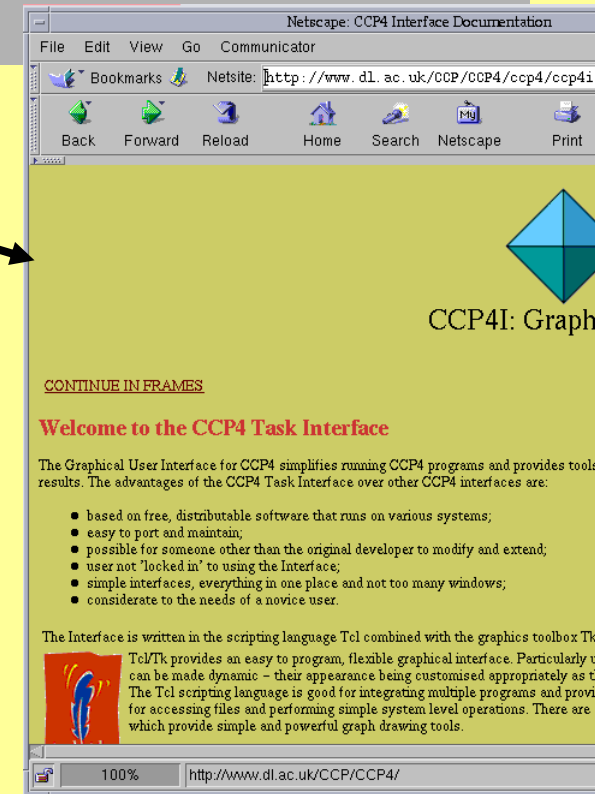
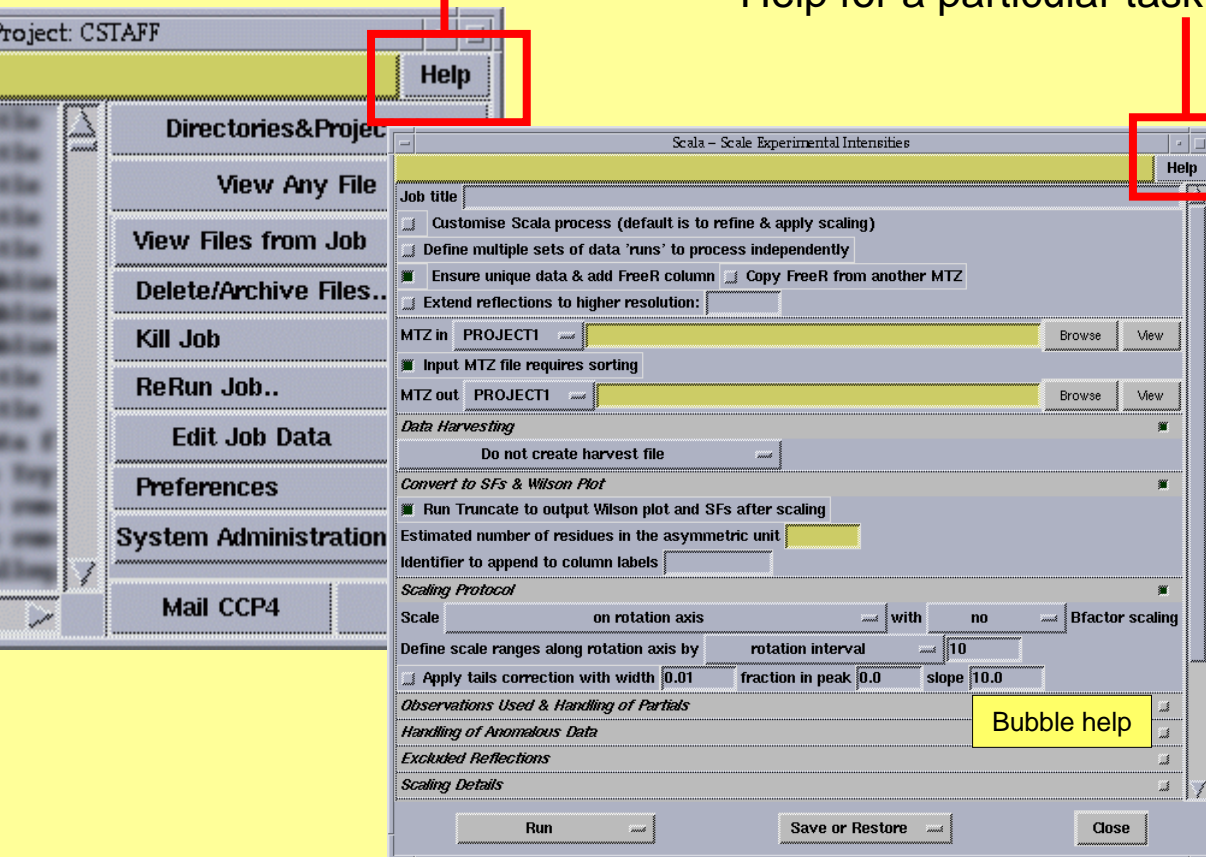
Defaults - "If it's not visible then it's not important"

Online help within CCP4i

General help from main window

Brings up relevant documentation in browser

Help for a particular task



Help with a particular option:
Right hand mouse button click over that option

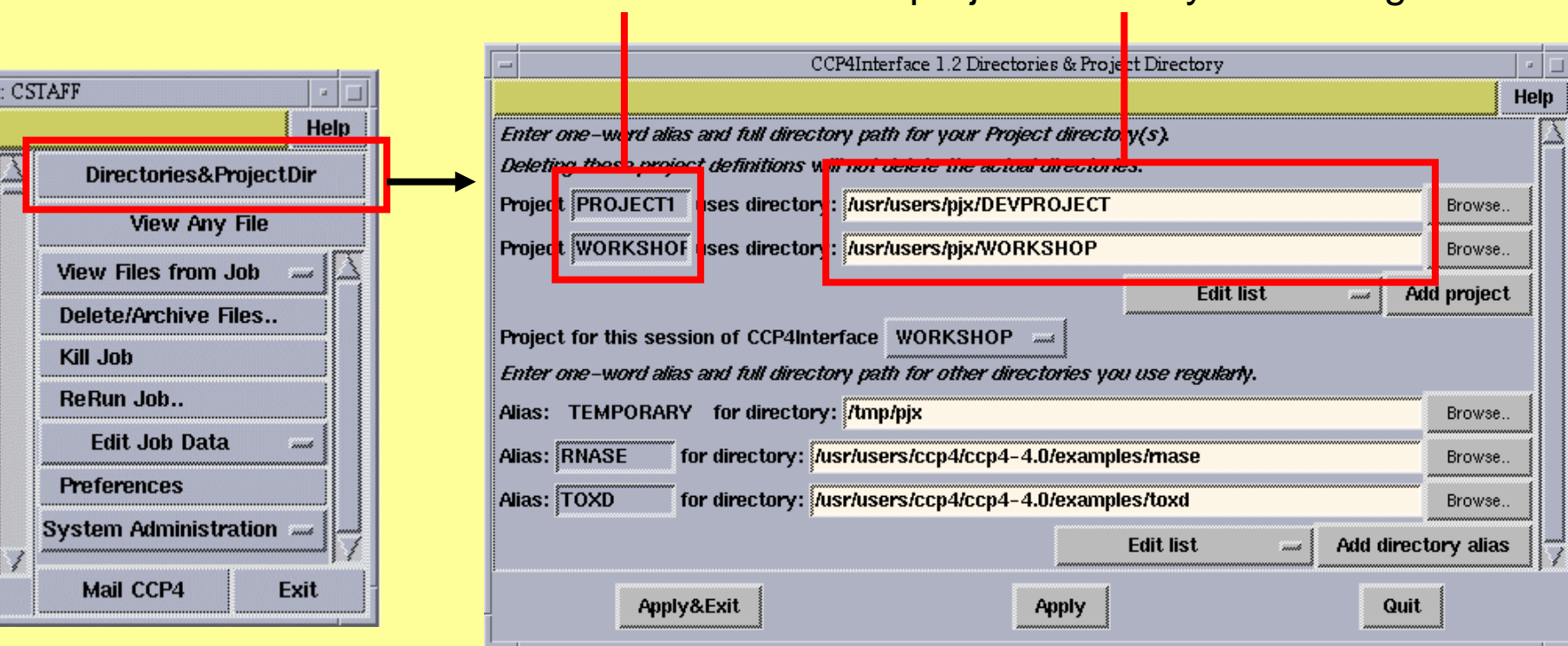
Project Management Tools in CCP4i

Why Project Management?

- Reminds you what you did six months ago
- Helps keep track of multiple projects and associated data
- Facilitates back-tracking (especially if things go wrong)
- Helps when depositing results & writing your paper

Setting up projects in CCP4i

One word alias for project directory containing data files



- All data files relating to one crystallographic project should be in a single project directory

Job database & Project History

CCP4 Program Suite 4.2.1 CCP4Interface 1.3.8 running on ccp4g.dl.ac.uk Project: CSTAFF

Task	Job ID	Date	Status	Input	Output
Molecular Replacement	512	23 Jul 02	KILLED	acorn	[No title
Cell Content Analysis	511	19 Jul 02	FINISHED	mlphare	Test run o
Analyse Data for MR	510	19 Jul 02	FINISHED	matthews	Solvent co
Beast - likelihood-based MR	509	17 Jul 02	FINISHED	mlphare	try anothe
Molrep - auto MR	508	17 Jul 02	FINISHED	mlphare	Restored j
Create Input SFs from Model	506	17 Jul 02	FINISHED	mr_analyse	Test of MR
AMoRe Model Database	505	17 Jul 02	FINISHED	matthews	Test run o
AMoRe	504	09 Jul 02	FINISHED	prephadata	[No title
Edit AMoRe Solution File	503	09 Jul 02	FINISHED	prephadata	test of el
Build AMoRe Output Model	502	25 Jun 02	FINISHED	scala	[No title
Edit Protein Structure	500	25 Jun 02	FINISHED	scala	Treat all
	499	24 Jun 02	FINISHED	refmac5	review wit
	498	24 Jun 02	FINISHED	refmac5	idealisati
	496	18 Jun 02	FINISHED	refmac5	Roberto St
	491	18 Jun 02	FINISHED	dm	Sacrificia

Help

Directories&ProjectDir

View Any File

View Files from Job

Delete/Archive Files..

Kill Job

ReRun Job..

Edit Job Data

Preferences

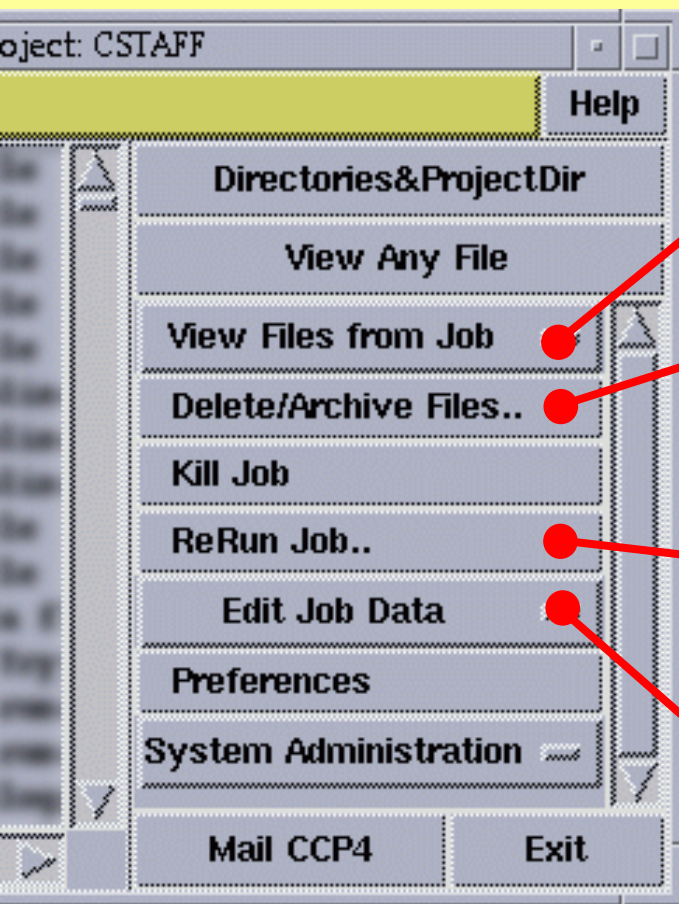
System Administration

Mail CCP4

Exit

- One job database per project
- Stores parameters used to run each task
 - Records date, status & input, output and logfiles for each job (project history)

Job database utilities



View files from *any* job in the database

Remove failed/unwanted jobs from the database and archive important data

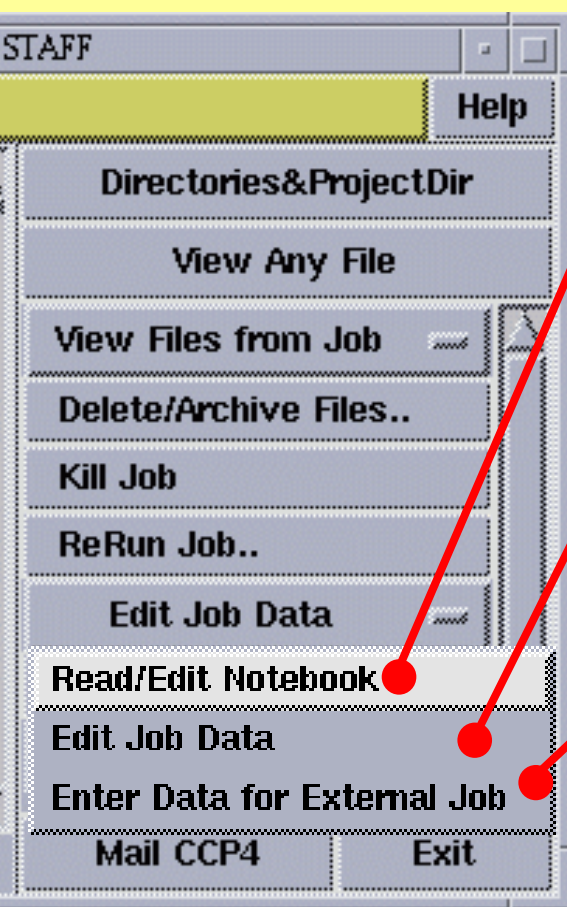
Rerun *any* job in the database (with the option of changing the parameters first)

- Use this to review parameters used in an earlier run

Keep the database up-to-date

- Add runs of “external” programs

Edit Job Data utilities



• Electronic Notebook

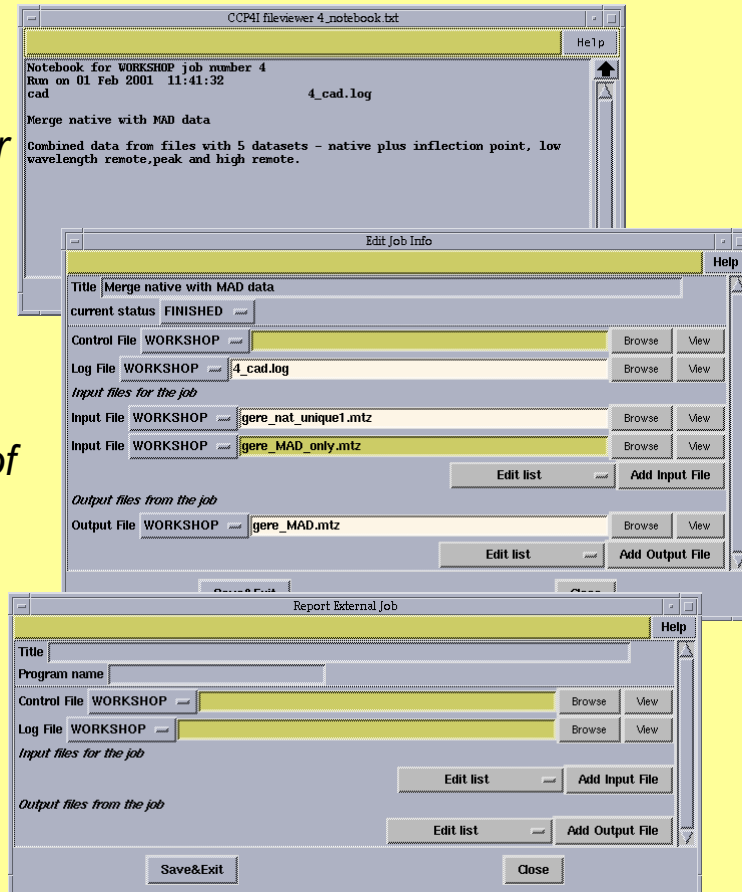
- Record information about a particular job for future reference

• Edit Job Data

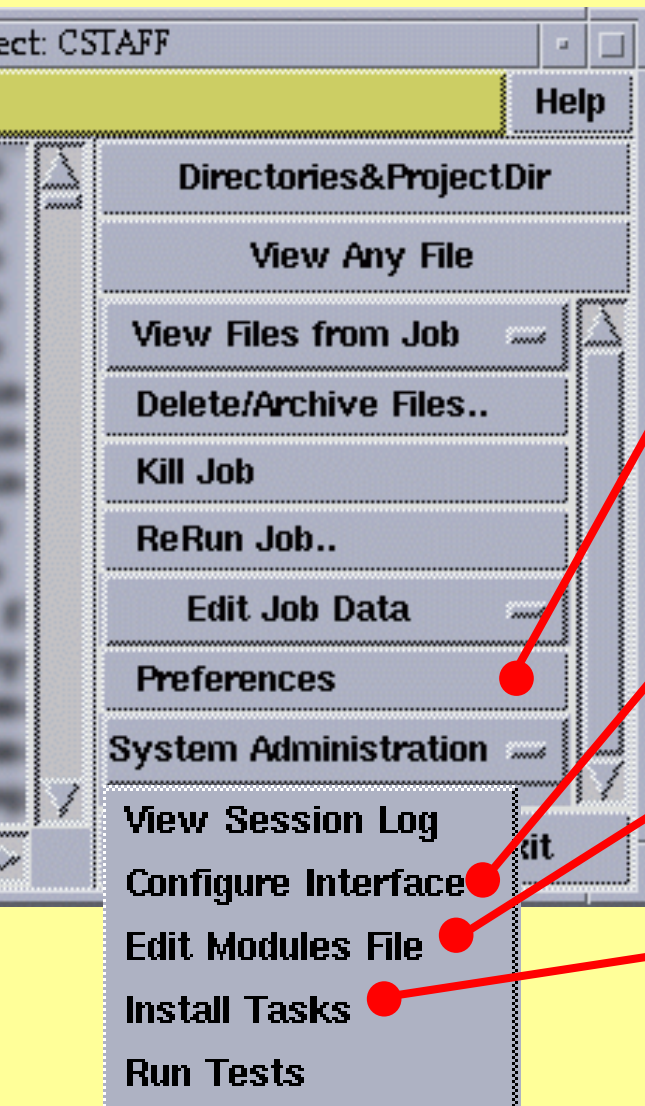
- Keep Job Database up-to-date
- Record changes e.g. of file locations

• Report External Tasks

- Record runs of non-CCP4(i) programs plus associated files
- Keep project history complete



Configuring and customising CCP4i



1. Preferences

- *Default viewers for PDB files and map files*
- *Data harvesting defaults*

2. Configure Interface

- *Maximum column lengths for menus*
- *Switch bubble help on or off*
- *Set name of web browser*
- *Explicitly define paths for programs*

3. Edit Modules File

- *Create new modules*
- *Add new references to existing tasks*
- *Requires some understanding of how tasks are referenced in CCP4i*

4. Install Tasks

- *Tracks tasks that are installed & lets you review/uninstall*

Preferences and Configure interface

1. Preferences

- Default options for deleting and archiving jobs
- Default file selection listing (alphabetic or by date)
- Map defaults including:
 - Format (O, CCP4, Quanta)
 - Location
- *Default viewers for PDB and map files*
- *Data harvesting defaults*

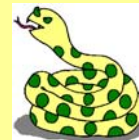
2. Configure Interface

- *Maximum column lengths for menus*
- *Switch bubble help on or off*
- *Set name of web browser (useful if it's not netscape!)*
- *Explicitly define paths for programs*
 - *useful for overcoming name clashes e.g. `dm` is a CCP4 program and a game under Linux!*
- Also configuring machines for remote running, printers, fonts etc

Edit Modules File and Install New Task

3. Edit modules file

- Customise lists of modules and tasks
- Requires some understanding of how tasks are referenced in CCP4i
 - Create new modules
 - Add new references to existing tasks



4. Install New Tasks

- Used e.g. by ARP/wARP, Phaser
- Tracks tasks that are installed and lets you review/uninstall

Overview of CCP4 file formats

Working Formats

- **MTZ:** *reflection data*
 - See following slides
- **PDB:** *coordinate data - based on PDB version 2.1 draft*
 - Officially for atomic position data
 - Also used semi-unofficially for storing other coordinate-based data
- **CCP4 map:** *electron density, pattersons, difference maps, masks*
 - Binary format so use `mapdump` to view header information
 - Map files can be large but are easily (re)generated from the original data

Other Formats

- **CCIF:** *coordinate data, harvest information, Refmac monomer dictionary*
- subset of the IUCr mmCIF dictionary
- **XML:** *(currently developmental) markup logfile information*

See **FILE FORMATS** section in documentation

<http://www.ccp4.ac.uk/dist/html/INDEX.html>

CCP4 Data File Formats: MTZ files

- **Store reflection data, e.g:**
 - Intensities
 - Structure factor amplitudes (observed/calculated)
 - Anomalous differences/Friedel pairs
 - Free-R flags (for cross-validation)
 - Phases, Figures-of-Merit etc
- **Binary format**
 - files are more compact & faster to read/write
 - need to use utilities to view and manipulate
- **Batch MTZ files are produced after integration e.g. from Mosfilm**
 - also referred to as multi-record files
 - contain multiple observations of the same reflection (“record”)
 - (*simplistically*) each batch corresponds to a diffraction image
 - perform data reduction steps to get standard MTZ file

MTZ file: tabular view

MTZ file can be thought of as a “table” of data

- columns = intensities, structure factors etc
- rows = values of each column associated with a reflection

Multiple **Crystals** within same file

Multiple **Datasets** within each dataset

Rows=reflections
(Miller indices)

			Crystal 1: name = "Native"				Crystal 2: name = "HgDeriv"	
			Dataset 1: Project="RNase" Name="D1"		Dataset 2: Project="RNase" Name="D2"		Dataset 1 ...	
H	K	L	F	Sig(F)	F	Sig(F)
0	0	0	49.2	0.5
0	0	2
0	0	6

Columns=quantities associated with reflections
e.g. intensities, structure factors, phases, FOM etc
Reference columns via their names (“labels”)

CCP4 Data File Formats: MTZ file header

- Use the `mtzdmp/mtzdump` program to view MTZ information
- Sample output from MTZ header:

```

* Title:
Dendrotoxin from green mamba (1dtx) - Tadeusz Skarzynski 1992...
} User-supplied descriptive title

* Number of Datasets = 4
* Dataset ID, project/crystal name, dataset name, cell dimensions, wavelength:
  1 TOXD /
  NATIVE
  73.5820 38.7330 23.1890 90.0000 90.0000 90.0000
} Dataset information
(names, associated cell &
wavelength)

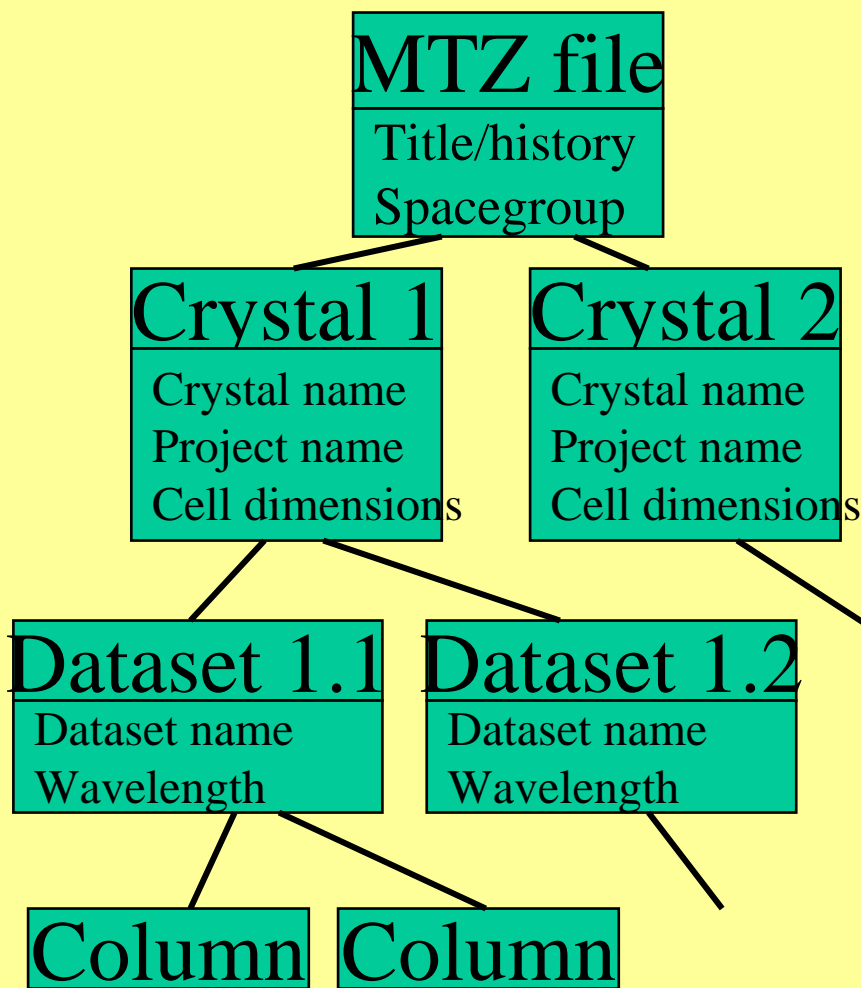
* Number of Columns = 14
* Column Labels :
H K L FTOXD3 SIGFTOXD3 ANAU20 SIGANAU20 FAU20 SIGFAU20 ... FreeR_flag
* Column Types : H H H F Q D Q F Q F Q F Q I
* Associated datasets : 1 1 1 1 1 2 2 2 2 3 3 4 4 1
} Column information
(labels, data types,
which dataset they
belong to)

* Cell Dimensions :
73.5820 38.7330 23.1890 90.0000 90.0000 90.0000
* Resolution Range :
0.00074 0.18900 ( 36.761 - 2.300 A )
* Space group = P212121 (number 19)
} Additional information

```

- **Other information not shown here includes: number of reflections, history etc**

MTZ data hierarchy: crystals, datasets and columns



Crystal: a physical crystal which was used to obtain data in one or more diffraction experiments

- *e.g. native, heavy atom derivative etc*

Dataset: data derived from a single experiment on a particular crystal

- *e.g. different MAD wavelengths*

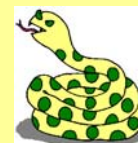
Column: a particular type of data associated with a dataset

- *e.g. experimental quantities (measured intensities) and data derived at various levels (observed structure factors, phases)*

Crystals Projects and Datasets in practice

Each crystal has an associated set of cell parameters

- In 5.0.1: the crystal cell is used by most programs



Each dataset has an associated wavelength

- many datasets can be associated with one crystal
- can be used automatically by some programs

Each dataset has an associated *project name*

- used by data harvesting

Set up crystals, projects, datasets when importing the data into MTZ format

- Add or edit later on using appropriate utilities

Data Harvesting in CCP4

Data Harvesting is the automatic capture of information by key programs in the structure determination process

- **mosflm, scala, truncate, mlphare, refmac5**
- data is recorded in *harvest files*
- at deposition time these files form an accurate record of how the final structure was obtained

Harvesting operates automatically - all you need to do is:

1. **Add project and dataset information to your MTZ file**

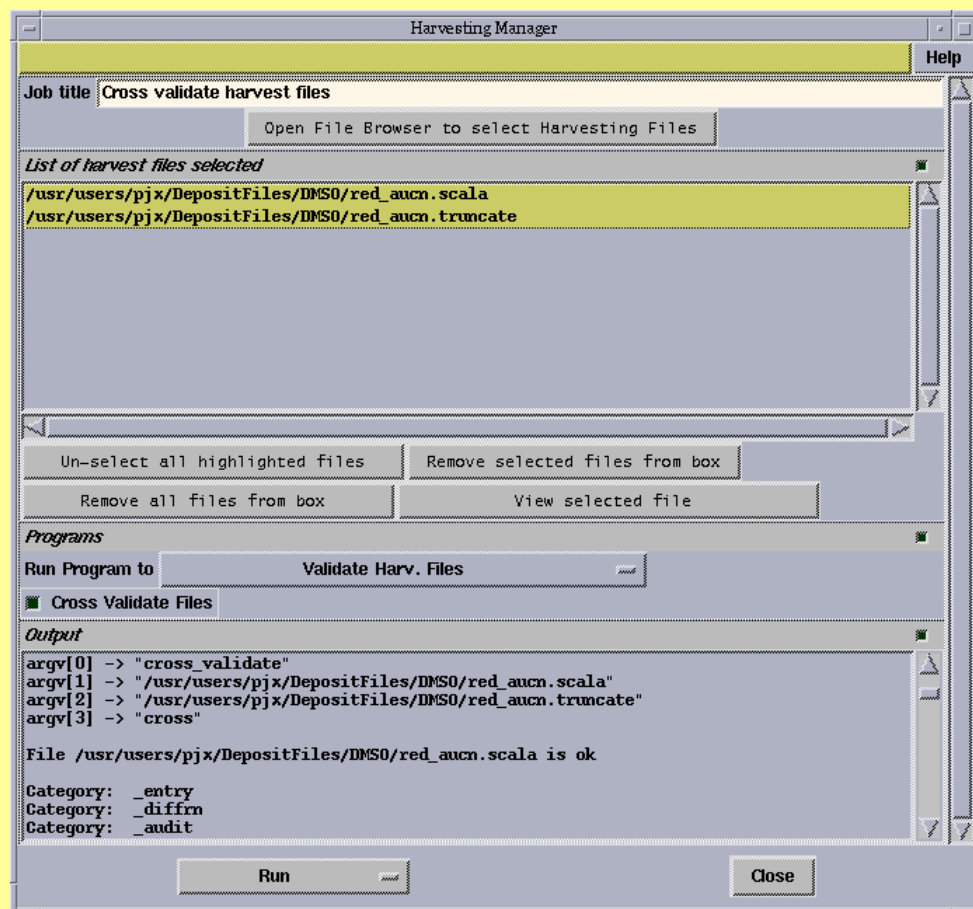
- when data is imported into CCP4 (or use utility programs)

2. **Switch on harvesting**

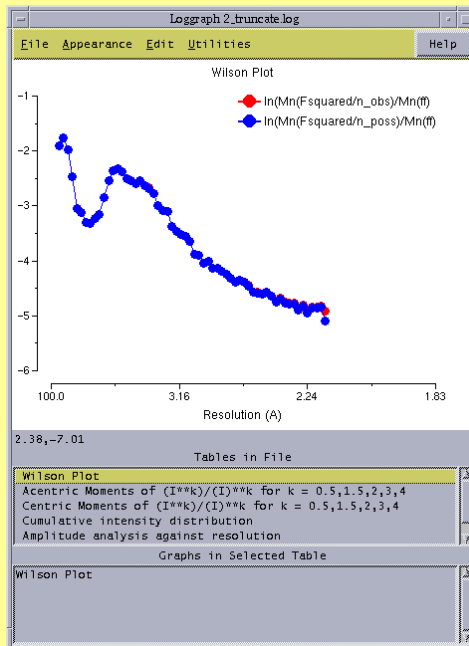
- use harvesting keywords in the programs, or
- in CCP4i – in individual tasks, or (better) in Preferences (default)

Data Harvesting Management Tool

- In the Validation&Deposition module of CCP4i
- Checking consistency and validity of harvest files prior to deposition
- Acts as an interface to `pdb_extract` to derive additional information for deposition from MTZ files, log files etc.

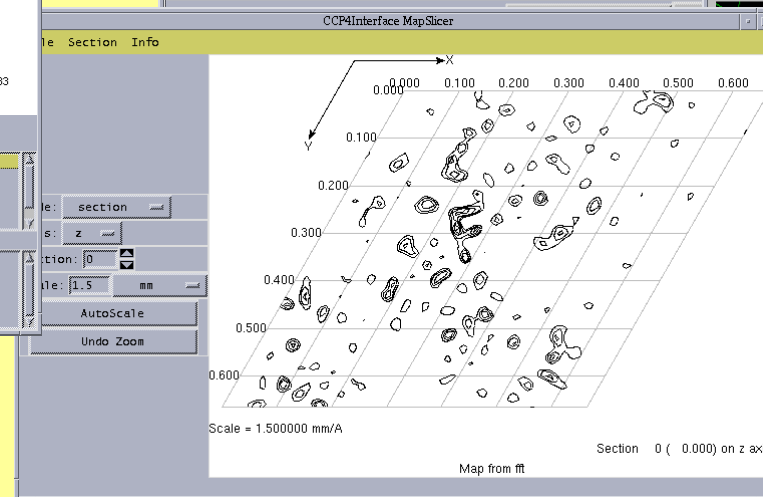
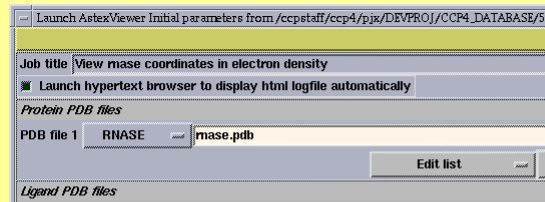


Utilities: graphical viewers

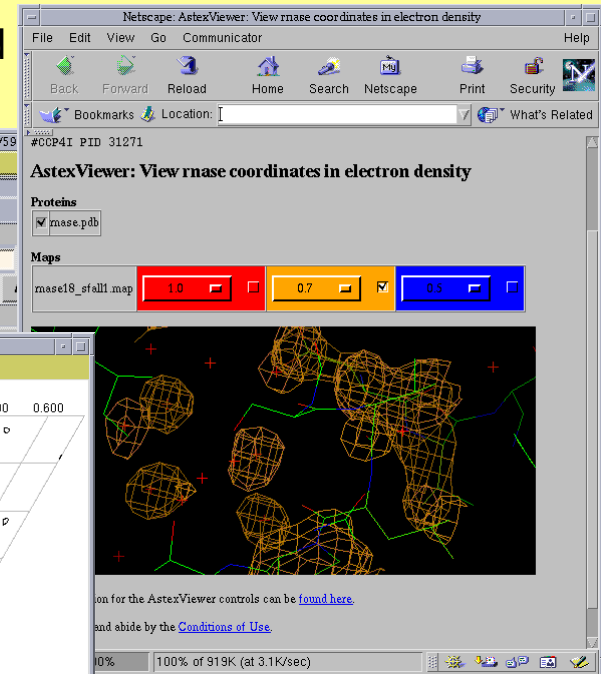


Loggraph: For graphs in CCP4 formatted logfiles

- **AstexViewer:** Java-based map-and-coordinate viewer



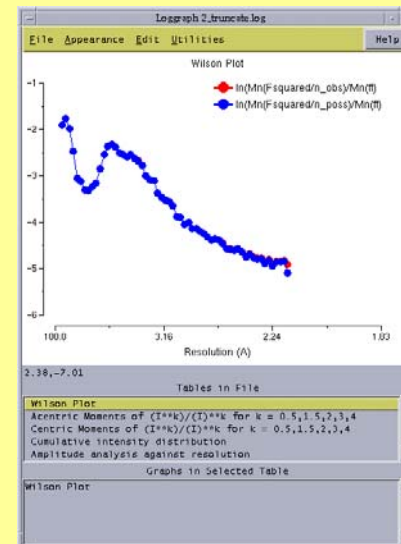
- **MapSlicer:** 2-d contoured sections through CCP4 maps



- **XtalView/Xfit launcher:** available for those who prefer to use XtalView - in CCP4i "Model Building" module

File viewing from within CCP4i

- From within the interface
 - **View Files from Job:** always uses default file viewer
 - **View Any File:** allows you to select from available viewers
- From Unix command line:
 - Use **ccp4i -v <filename>** to view a file in the default viewer
 - Useful for MTZ files (automatically runs mtzdump program to display header)
- HTML logfiles
 - Can be viewed as plain text or in HTML browser
- Loggraph
 - View tables and graphs in CCP4-formatted logfiles
 - Can also use **loggraph <filename>** at the command line



Utilities: file manipulations

MTZ files		
<i>Operation</i>	<i>CCP4i module and task</i>	<i>Program(s)</i>
Convert reflection data file to MTZ	Reflection utilities->Convert to MTZ and standardise (import)	f2mtz, cif2mtz, scalepack2mtz, dtrek2mtz
Convert from MTZ to other format	Reflection utilities->Convert from MTZ (export)	mtz2various
Add & edit crystals and datasets	Reflection utilities->Edit MTZ datasets	cad
Merge files	Reflection utilities->Merge MTZ files	cad
View contents	View any file (main window)	Mtzd(u)mp
General data manipulations	Reflection utilities->Edit MTZ files	sftools, mtzutils

Utilities: file manipulations

Batch MTZ files		
<i>Operation</i>	<i>CCP4i module and task</i>	<i>Program(s)</i>
Convert reflection data file to batch MTZ	Data reduction->Import unscaled data	combat , dtrek2scala
View Contents	View any file (main window)	mtzd(u)mp
General data manipulations	Data reduction->Modify/merge MTZ files	rebatch

Utilities: file manipulations

PDB files		
<i>Operation</i>	<i>CCP4i module and task</i>	<i>Program(s)</i>
Edit/manipulate	Coordinate utilities->Edit PDB file	pdbset
Convert from PDB to other formats	Coordinate utilities->Convert coordinate formats	coordconv
Convert from PDB to mmCIF	Not currently interfaced	coord_format
Repair broken files		
View contents	View any file (main window)	more (unix command) Rasmol astexviewer
Superpose coordinates	Coordinate utilities->Superpose molecules	lsqkab, topp

Utilities: file manipulations

Map and mask files		
<i>Operation</i>	<i>CCP4i module and task</i>	<i>Program(s)</i>
Generate maps	Experimental Phasing->Generate Patterson map Map & mask utilities->Run FFT – Create Map	fft
Generate mask	Map & mask utilities->Create/Edit Masks	ncsmask
View contents	View any file (main window)	Mapdump, mapslicer astexviewer
Manipulations	Map & mask utilities->various	Maprot, mapmask

Resources and support

- **Documentation**

- Installation: <http://www.ccp4.ac.uk/dist/INSTALL.html>
- Program documentation: <http://www.ccp4.ac.uk/dist/INDEX.html>
 - *Function index also available*
- Links to useful background material:
 - *Information on harvesting, scaling, twinning ...*
 - *“Roadmaps” for navigating the software*
- Extensive documentation accessible via CCP4i

- **Tutorials**

- Based around CCP4i
- Cover data processing & scaling, experimental phasing, MR, refinement

- **Bugs and problems**

- problems pages at <http://www.ccp4.ac.uk/problems.html>
- E-mail [**ccp4@ccp4.ac.uk**](mailto:ccp4@ccp4.ac.uk)

Summary: remember this!

- Binary installations for fast start up
- Use CCP4i project management tools
- Add project, crystal and dataset information in MTZ
- Switch on data harvesting
- CCP4 has many useful programs for file viewing and manipulations