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

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

IUCr Florence
August 23rd 2005

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 5.0.2
- What's coming in CCP4 version 6.0
- 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**)
- Graphics and building (**CCP4mg/Coot**)
- Validation and analysis (**PDBExtract**)

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.

E.g. data reduction starting in CCP4:

`Mosflm` ➡ `Scala` ➡ `Truncate`

Or alternatively starting outside CCP4:

HKL2000/Scalepack ➡ **`Combat`** ➡ **`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:
  - remove any previous installation first
  - admin privileges are required to install
- For Linux, Irix, OSF1/TruUnix64, SunOS
  - use **`download-5.0.2.sh`** script to download and install automatically

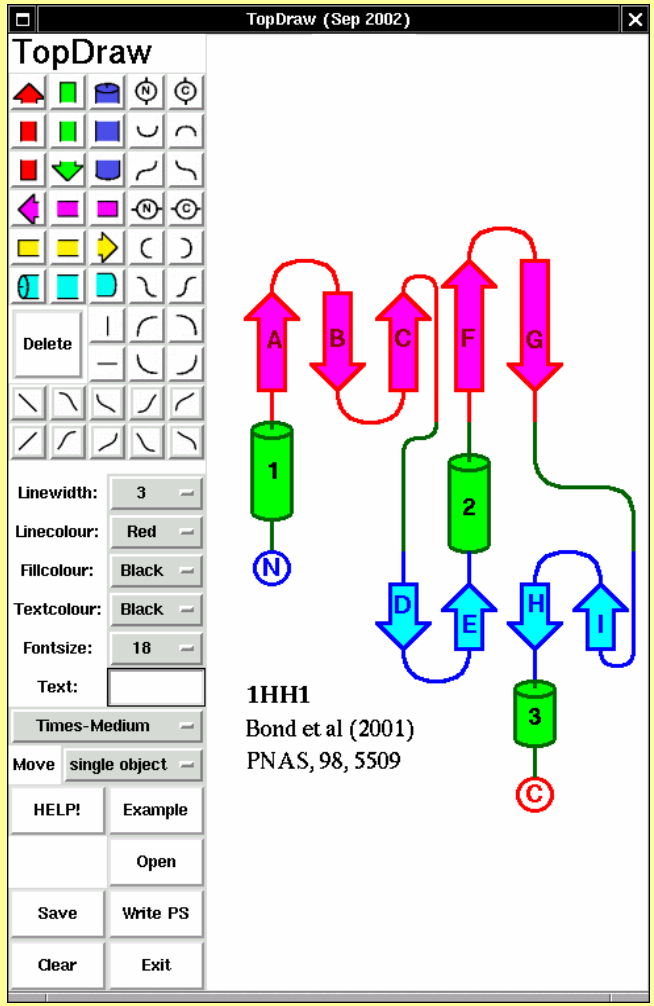
A Note about licensing

- *current academic licence has expired but no update available yet*
- *we will continue to honour the existing licence*
- *watch for announcements when update becomes available*

What's new in CCP4 5.0.2

- **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*



What's coming in CCP4 6.0

New packages:

- **CCP4MG:** CCP4 Molecular Graphics package
- **PHASER:** maximum-likelihood molecular replacement
- **Coot:** graphical model building tools
- **Pirate:** statistical phase improvement
- **Superpose:** secondary structure alignment
- **BP3**: heavy atom phasing & refinement
- **CHOOCH:** anomalous scattering factors from raw fluorescence spectra

Updates to **REFMAC5**, **MOLREP**, **SFCHECK**, **SCALA**, **PDBEXTRACT** and others

CCP4i:

- **CRANK:** automated structure solution via SAD, SIR, SIRAS
- **SHELXC/D/E** interface
- Database search and sort utility

Plus many bug fixes and minor improvements

Availability of CCP4 6.0

Test version 5.99.2 available:

- see **`http://www.ccp4.ac.uk/dev/releases.html`**

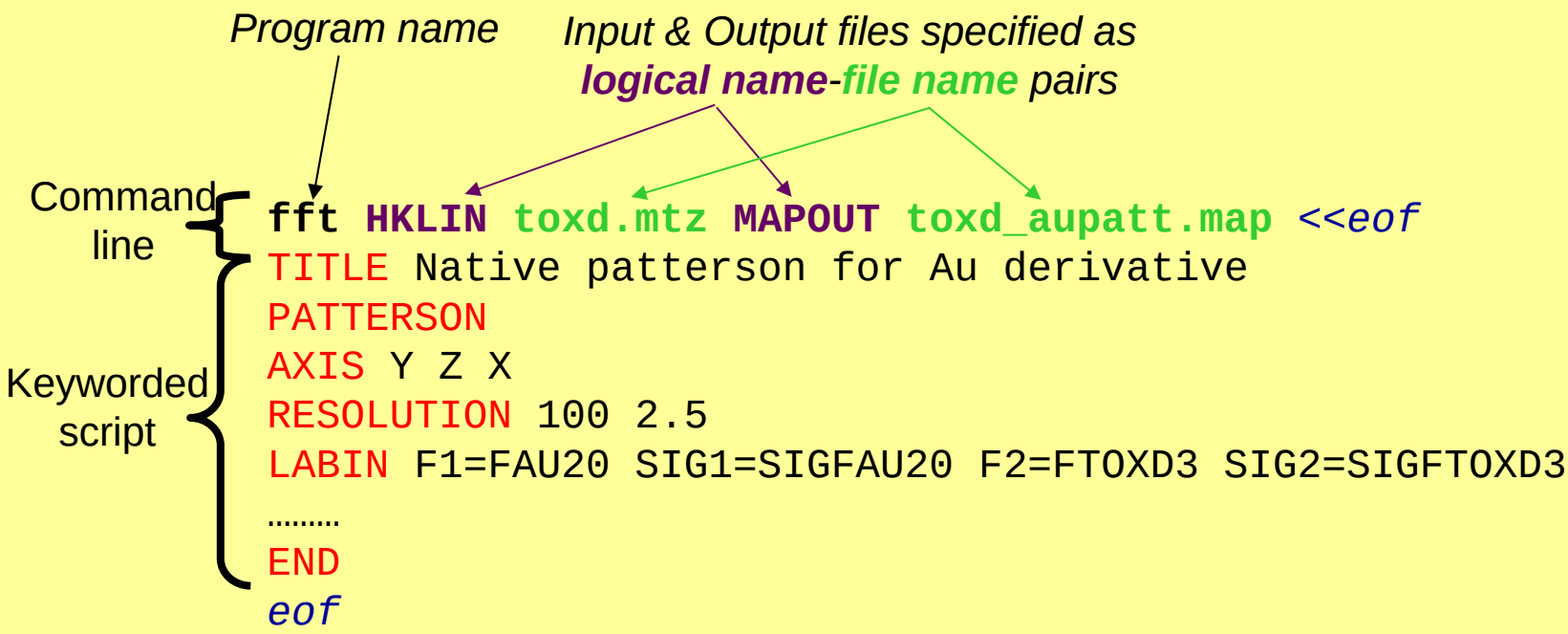
Downloads divided into a number of packages:

- Basic CCP4 (about the same as v5.0)
- **`Phaser`**
- ***`cctbx`*** *(libraries)*
- **`CCP4mg`**
- **`Coot`**
- **`CHOOCH`**
- *plus dependencies* (**`Tcl/Tk/BLT`**, **`Python`** …)

New download pages:

- allow user to select required packages and dependencies
- download a single file for installation
- source code and/or binaries

Running programs via scripts – an example



- *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
  - Allows mix-and-match approach (use both scripting & CCP4i)

- **Philosophy: *"Task-driven"* rather than *"program-driven"***

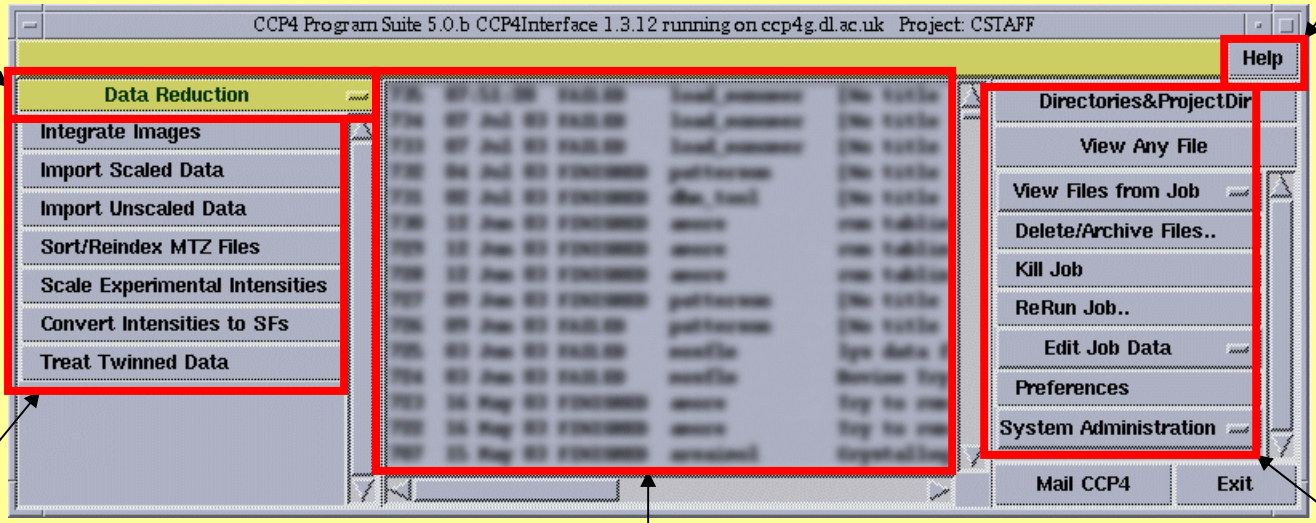
- **Key features:**
  - Easy-to-use interfaces to major programs and utilities
  - Tools for file viewing and basic project management
  - Customisable
  - Integrated help system

- **Requires that Tcl/Tk and BLT are installed**

CCP4i main window – quick tour

Modules

On-line help



Tasks

Job Database

Tools & Utilities

To start up CCP4i:

- Unix: type **`ccp4i`** at the command prompt
- Windows: launch using the CCP4 icon in the Start Menu

Example of a CCP4i task interface

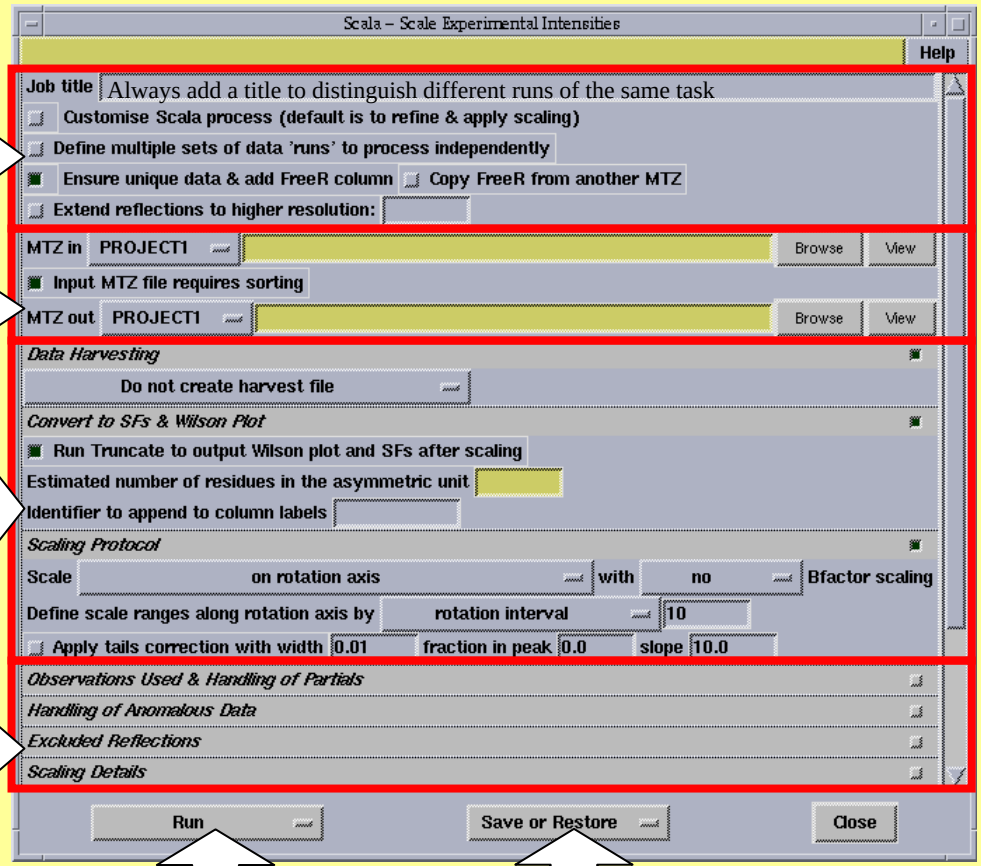
WORK FROM THE TOP DOWN

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

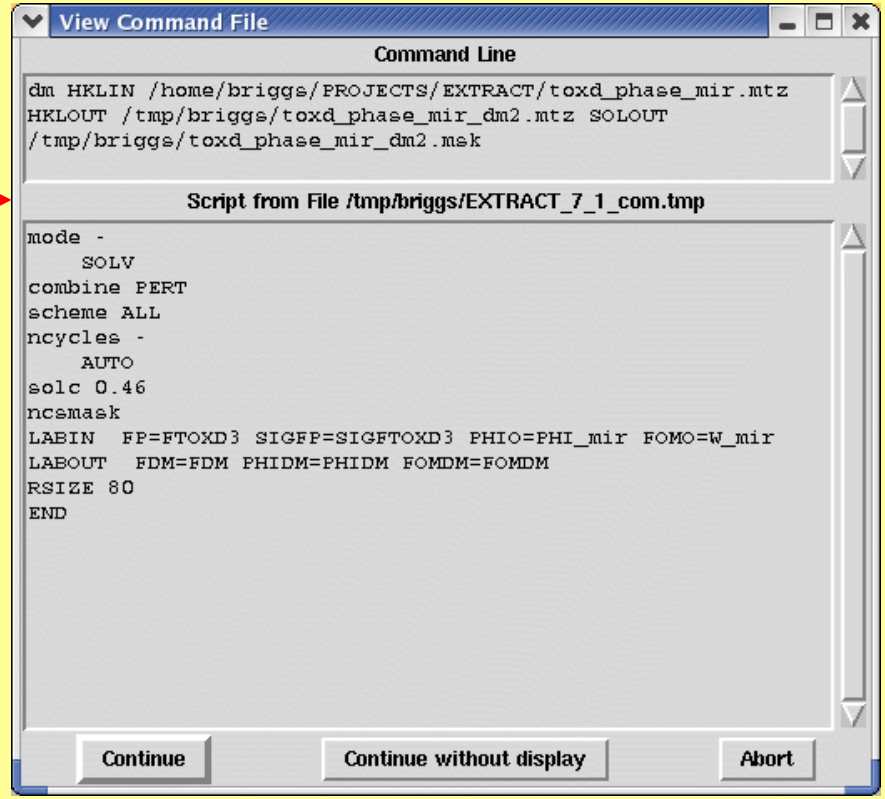


Run task

Save/restore parameters

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

Running tasks … back to scripts …



Run Now

- no further intervention required

Run&View Com File

- view (and edit) command line and scripts
- scripts also viewable from output files

Run Remote/Batch/Later

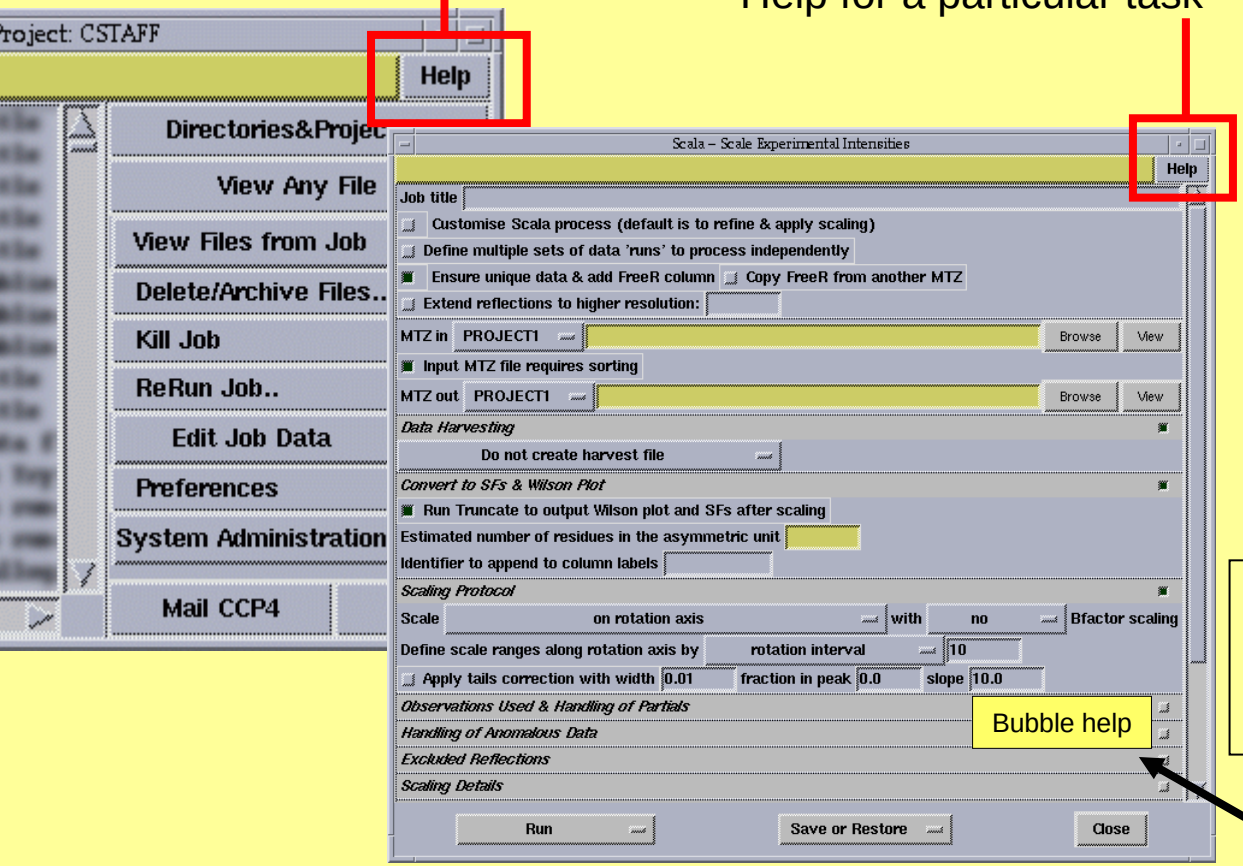
- use a remote machine or a batch queue or schedule task to run at a future date/time

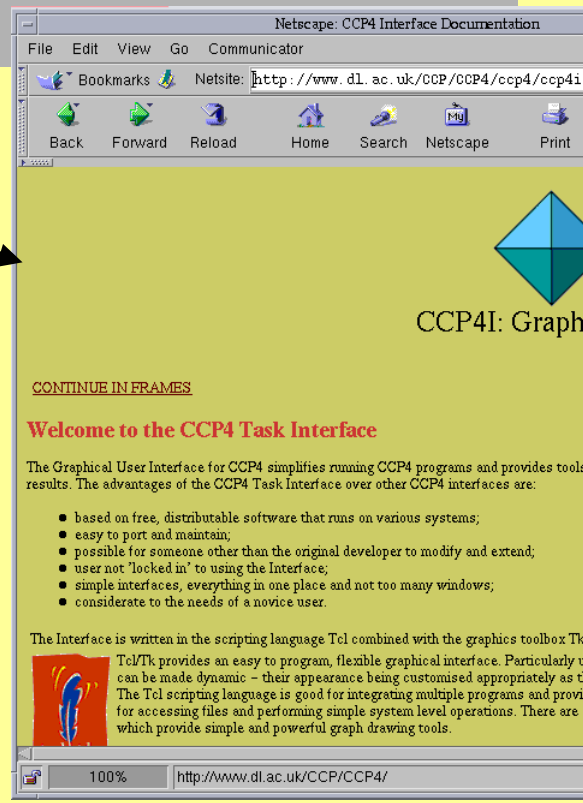
Online help within CCP4i

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

Can be switched off in
Configure Interface

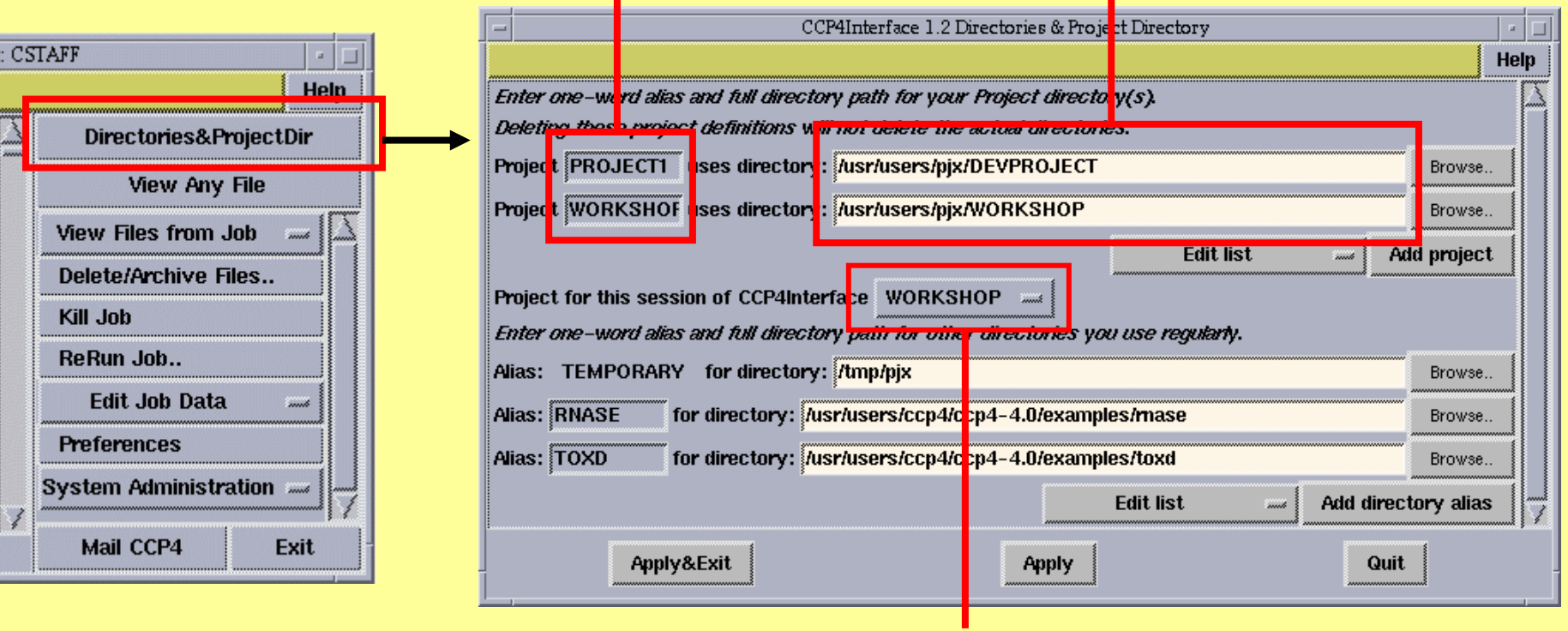
Project Management Tools in CCP4i

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

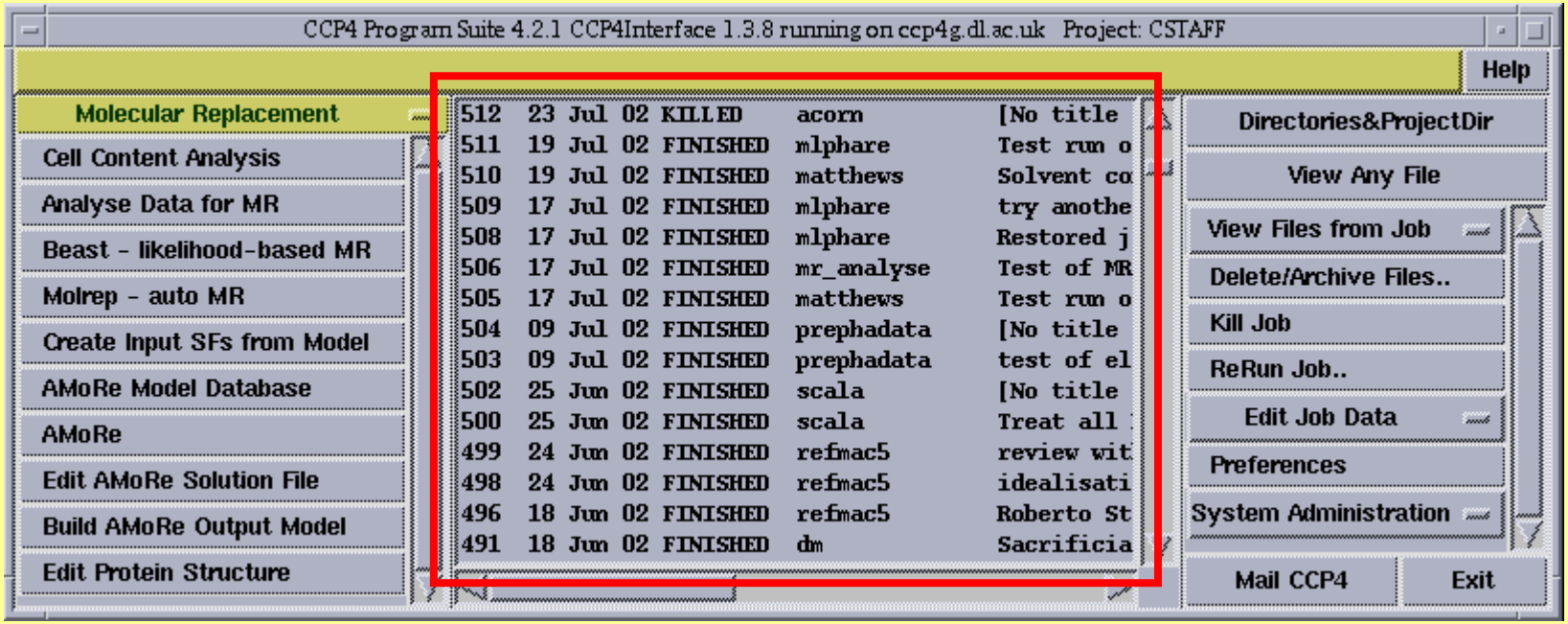


- *All data files relating to one crystallographic project should be in a single project directory*

Switch between projects

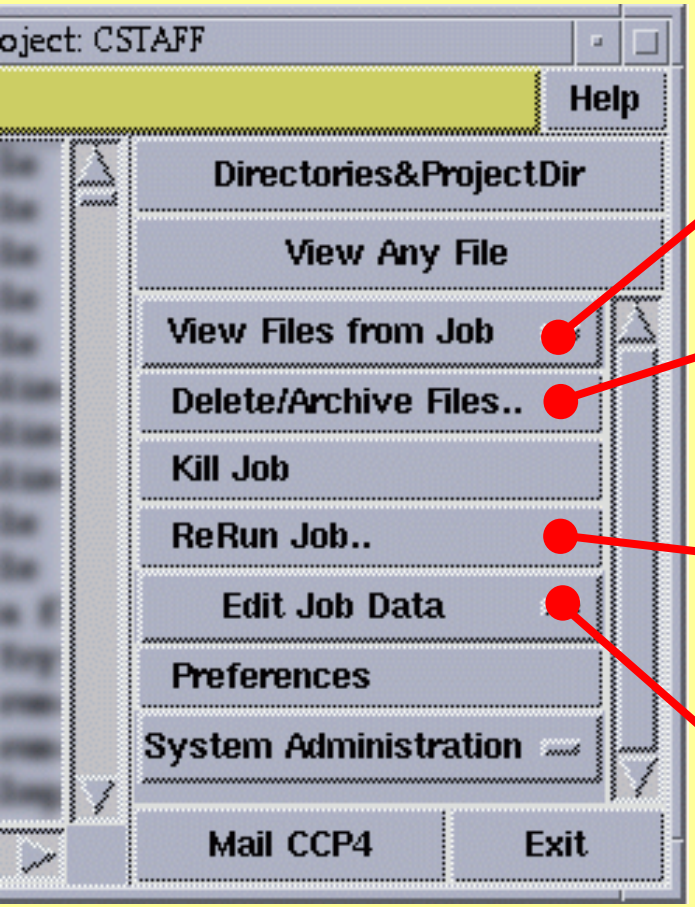
- in CCP4 6.0: also do this from the main window

Job database & Project History



- **One job database per project**
- **Stores parameters used to run each task**
  - Records date, status & input, output and logfiles for each job (project history)
  - In CCP4 6.0: new tool to search & sort database entries

Job database utilities



View files from *any* job in the database

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

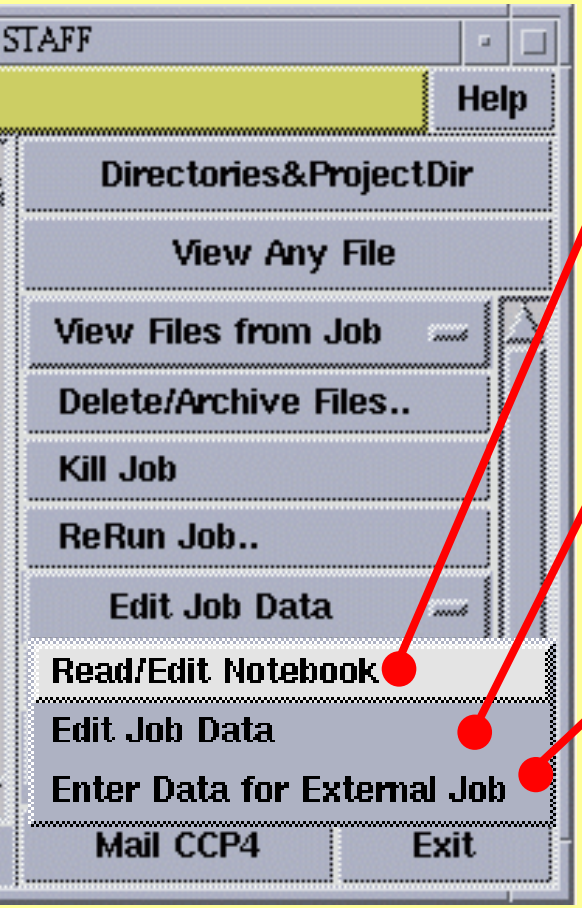
Rerun *any* job in the database (with the option of changing the parameters first)

- Use this to review parameters used in an earlier run

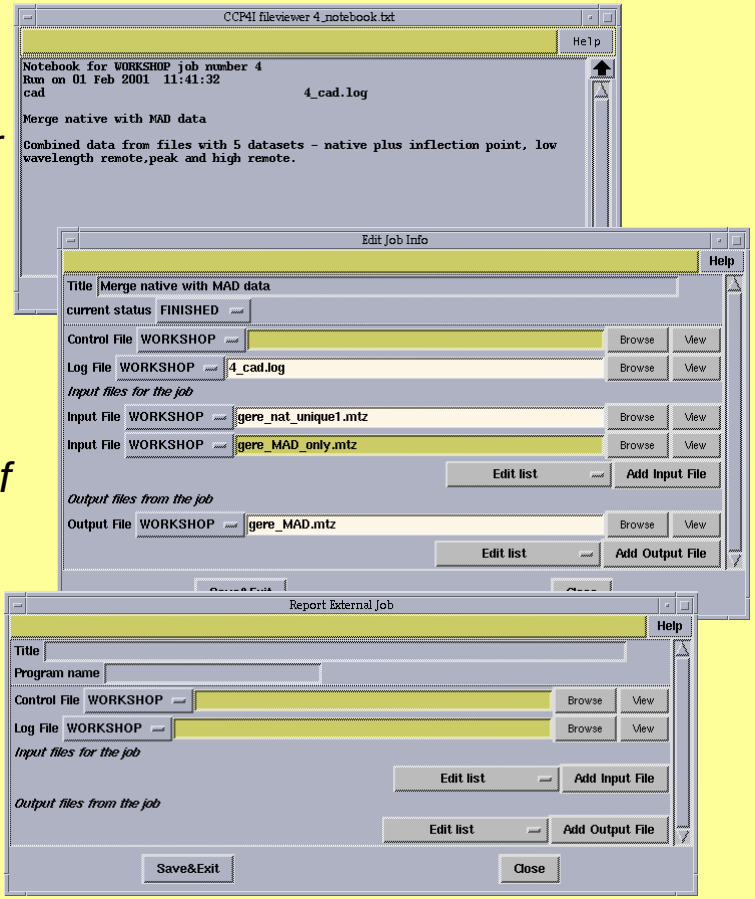
Keep the database up-to-date

- Add runs of “external” programs

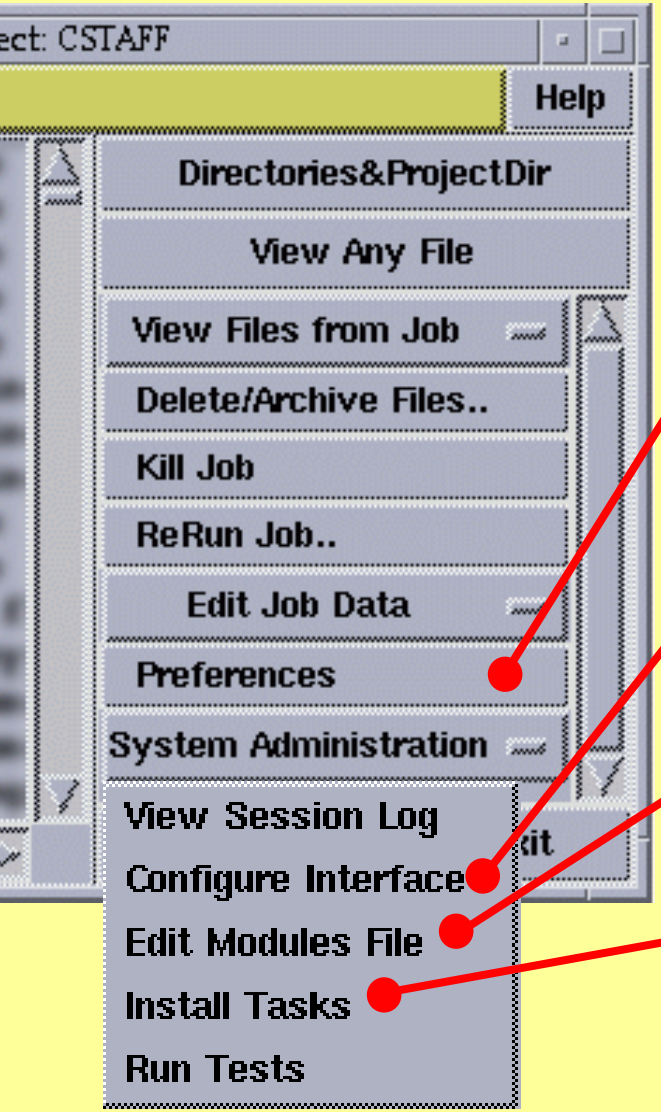
Edit Job Data utilities



- **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 and add new references to existing tasks*
- *! Requires some understanding of how tasks are referenced in CCP4i !*

4. Install Tasks

- *Used e.g. by ARP/wARP & Phaser*
- *Tracks tasks that are installed & lets you review/update/uninstall*

Preferences and Configure interface

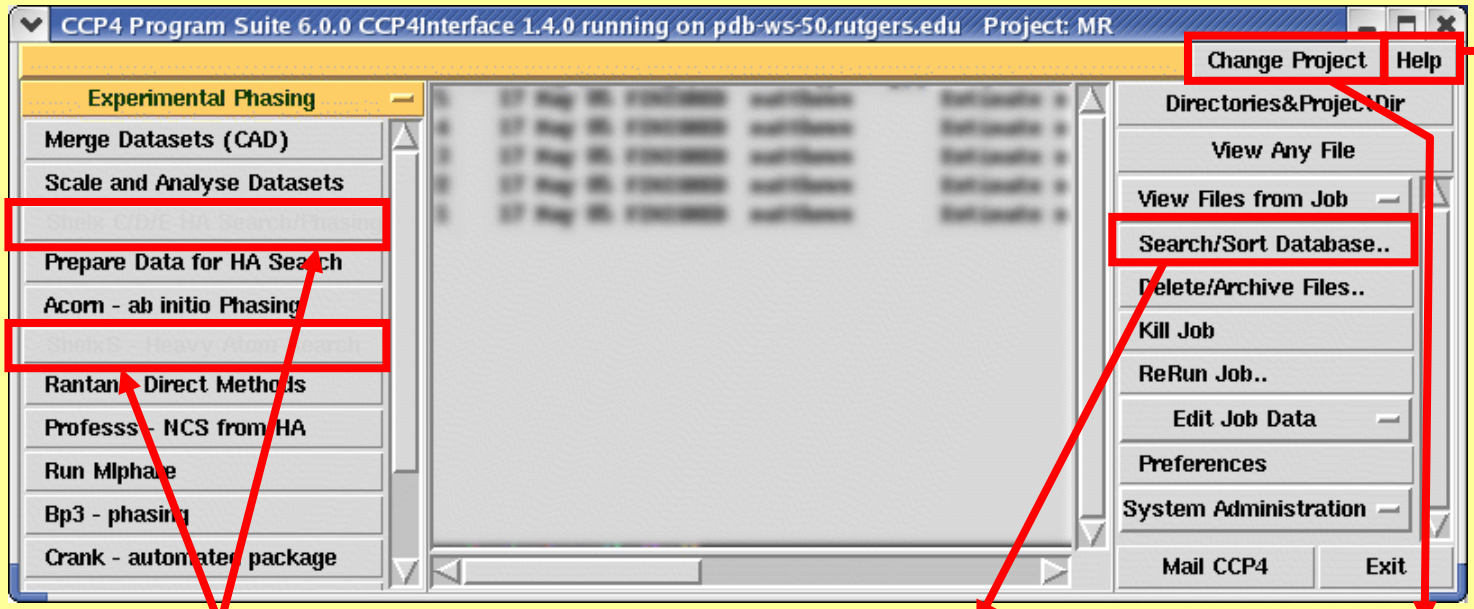
1. Preferences

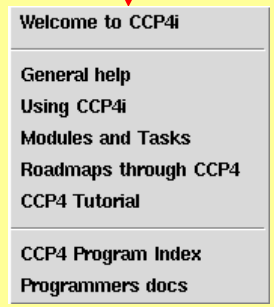
- Default options for deleting and archiving jobs
- Default file selection listing (alphabetic or by date)
- Map defaults including:
  - Format (O, CCP4, Quanta)
  - Location
- *Default viewers for PDB and map files*
- *Data harvesting defaults*

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!*
- Define batch queues & remote machines
- *Also configure printing, fonts etc*

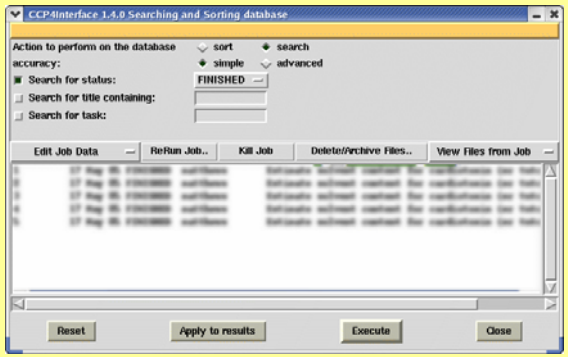
CCP4i – coming in CCP4 6.0





"Greyed out" tasks
- indicate that you need to install underlying software first e.g. SHELX

Database Search/Sort Tool



Quick switch between projects

Top level help
- split into topics

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
  - Can use **`mapslicer`** to view sections
  - 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 e.g.
`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
  - MTZ files are portable across different platforms

- **Batch MTZ files are produced after integration e.g. from `Mosflm`**
  - 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
- additional data groups together related columns

Multiple **Crystals** within same file

Multiple **Datasets** within each crystal

			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			…	…	…	…

Rows=reflections
(Miller indices)

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
```

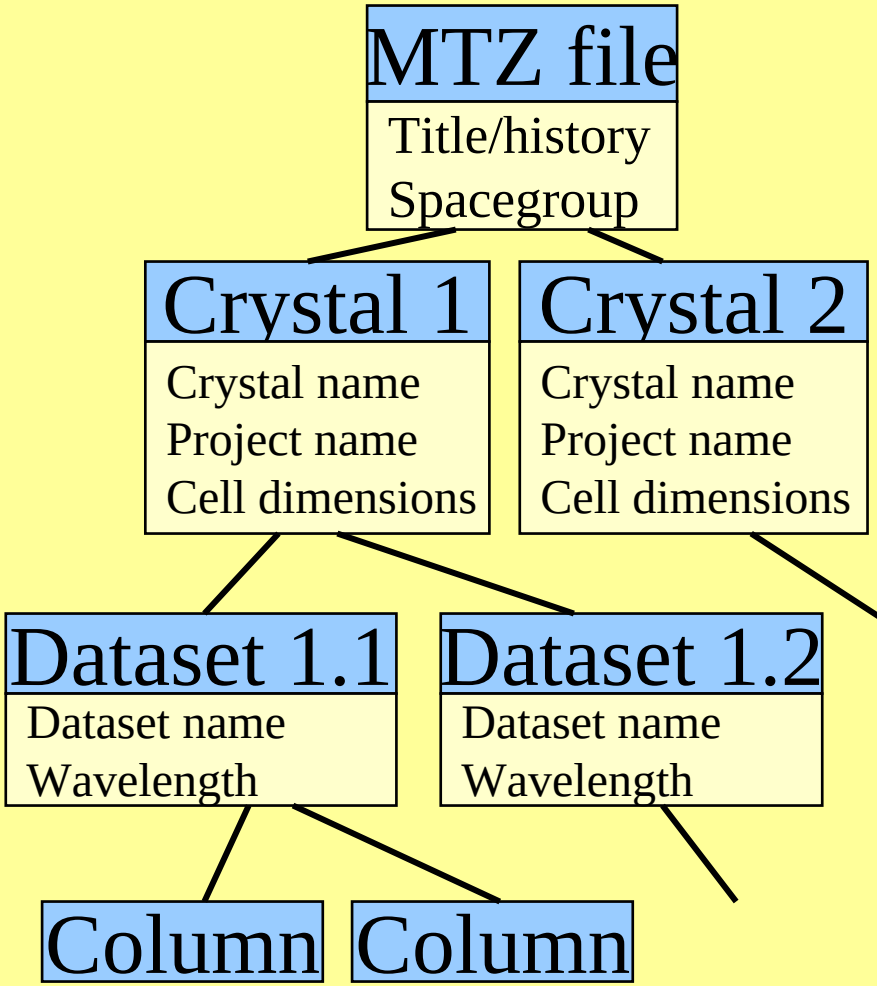
} Column information (labels, data types, which dataset they belong to)

```
* 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 (1)

Each crystal has an associated set of cell parameters

- **! In 5.0+ : the crystal cell is used by most programs !**
- *e.g. maps created by fft will have cell parameters taken from the parent crystal of the chosen MTZ column*

Each dataset has an associated wavelength

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

Each dataset also has an associated *project name*

- only used by data harvesting at present

All MTZ files also contain `HKL_base` dataset

- used to assign H K L columns
- other columns are assigned to `HKL_base` if not explicitly assigned to another dataset

Crystals Projects and Datasets in practice (2)

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

- using **`mosflm`**, **`scala`** etc or importing from **`scalepack`** etc

Or:

Add or edit later on using appropriate utilities

- Use the **`cad`** program or **`edit datasets`** task in CCP4i (**`Reflection Data utilities`** module)
- Allows you to set names and other attributes (cell, wavelength)

Crystal & dataset names

- should each be a single word
- only contain alphanumeric characters and underscores
- be no longer than 64 characters
- are case sensitive (i.e. **`rnase`** is not equivalent to **`Rnase`**)

See the DATA MODEL section in MTZ file format documentation
`http://www.ccp4.ac.uk/dist/html/mtzformat.html#datamodel`

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 mmCIF-format *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)

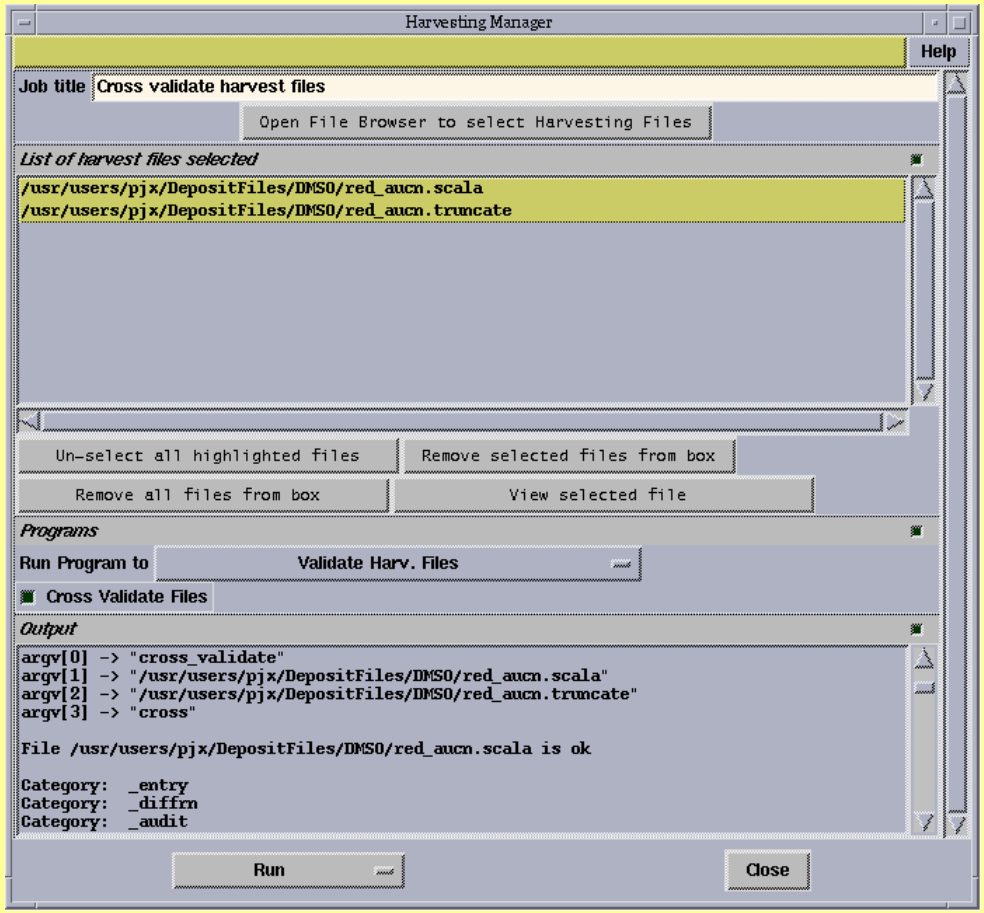
2. **Switch on harvesting**
   - use harvesting keywords in the programs, or
   - in CCP4i – in individual tasks, or (better) in Preferences (default)

Data Harvesting Management Tool

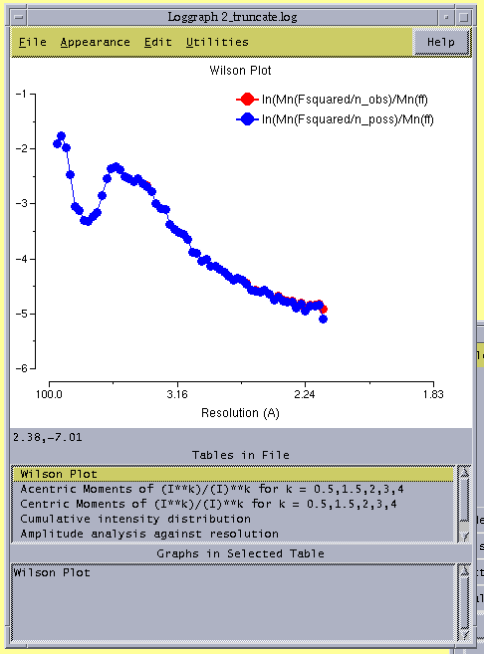
- In the **Validation&Deposition** module of CCP4i

- 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.

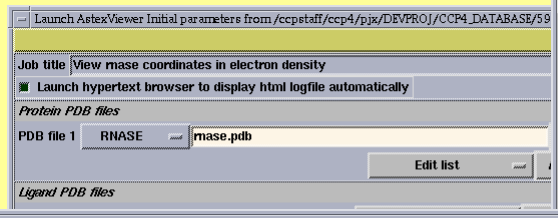


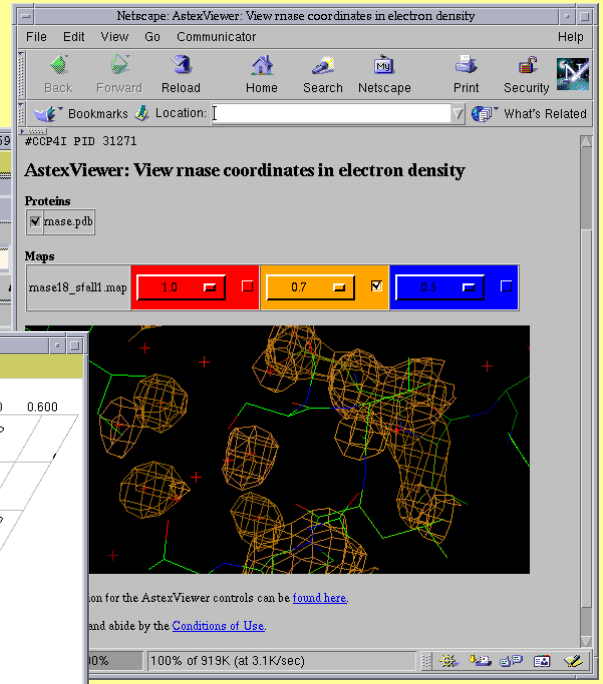
Utilities: graphical viewers

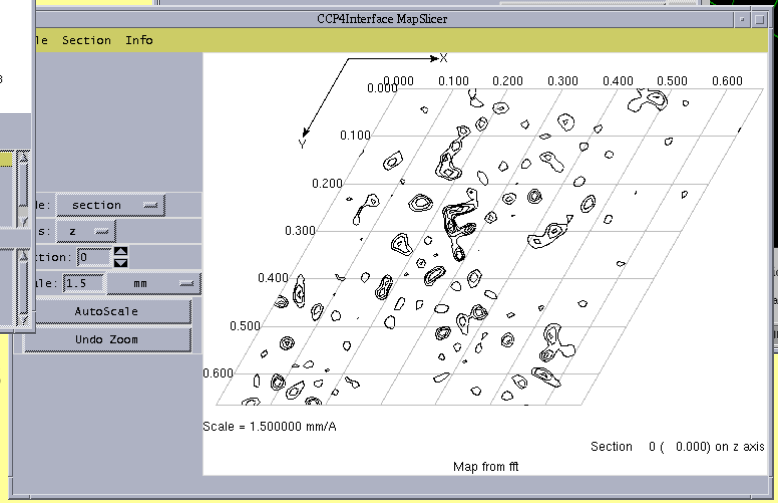


loggraph: For graphs in CCP4 formatted logfiles

- **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*

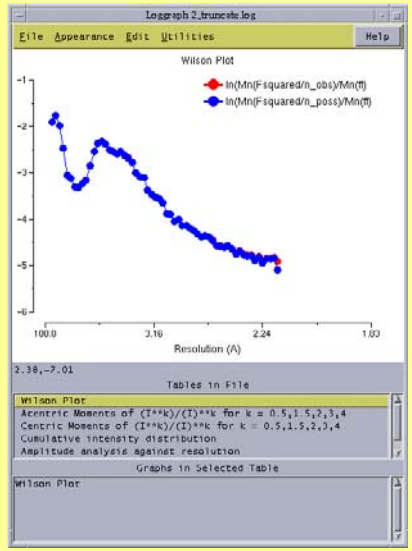
File viewing from within CCP4i

- *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*

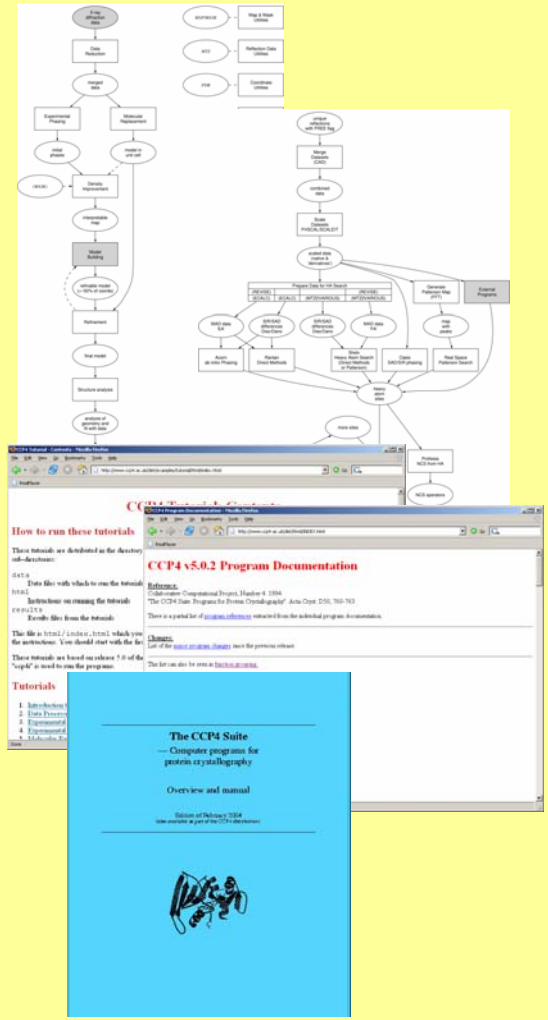
- Loggraph
  - *View tables and graphs in CCP4-formatted logfiles*
  - *Can also use* **loggraph <filename>** *at the command line*



Navigating the suite

Documentation (http://www.ccp4.ac.uk/docs.php):

- **Roadmaps**
- **Tutorials**
  - based around **`ccp4i`**
  - data processing/scaling, MAD, MR, refinement
- **Individual program documentation**
  - Function index
  - General background e.g. twinning, reindexing,
- **Postscript manual**
  - Slightly dated but still useful
  - Content distinct from program documentation



Runnable example scripts

- Part of the CCP4 distribution

Graphical user interface

- Also has extensive documentation

Utilities: file manipulations

MTZ files		
Operation	***CCP4i module and task***	***Program(s)***
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		
Operation	***CCP4i module and task***	***Program(s)***
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		
Operation	***CCP4i module and task***	***Program(s)***
Edit/manipulate	Coordinate utilities->Edit PDB file	**`pdbset, pdbcur`**
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		
Operation	***CCP4i module and task***	***Program(s)***
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`**

Other CCP4 Resources

Problems Pages

- known bugs/fixes with current release
- **`http://www.ccp4.ac.uk/problems.php`**

Bug Reports

- E-mail **`ccp4@ccp4.ac.uk`**

Other Problems

- General crystallography questions can go to ccp4bb
- **`http://www.ccp4.ac.uk/ccp4bb.php`**

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