

# Overview of CCP4 suite and current developments

Martyn Winn, STFC Daresbury Laboratory



1. CCP4 as a suite
2. Overview of CCP4 functionality
3. Future directions

# Philosophy



- ~200 independent programs from a wide pool of developers
- There may be several ways of doing a similar task  
e.g. Molrep/Phaser, DM/Solomon/Parrot
- Common routines placed in software libraries
- Infrastructure to guide user in applying programs

CCP4 exists to produce and support a world-leading, integrated suite of programs that allows researchers to determine macromolecular structures by X-ray crystallography, and other biophysical techniques. CCP4 aims to develop and support the development of cutting edge approaches to experimental determination and analysis of protein structure, and integrate these approaches into the suite. CCP4 is a community based resource that supports the widest possible researcher community, embracing academic, not for profit, and for profit research. CCP4 aims to play a key role in the education and training of scientists in experimental structural biology. It encourages the wide dissemination of new ideas, techniques and practice.

# In the beginning

- Programs only accessible via command line.
- Typically run via shell scripts.
- A crystallographic task requires chaining several programs together.
- Helped by the early adoption of standard file formats:

LCF → MTZ

PDB

CCP4 map

```
emacs@ccp4v.dl.ac.uk
File Edit Options Buffers Tools Help
$!
$!#####
$!          Set option COMBINE to combine phase information from
$! eg isomorphous replacement with partial structure
$!
$! Calculate structure factors.
$! File toxd_mir.mtz can be generated using mlphare.sh
$!
$pass/user CCP4_SCR:toxd_phase_mir.mtz hklin
$pass/user CCP4_SCR:toxd_sf_mir.mtz hklout
$pass/user ccp4_master:[ccp4.examples.toxd]toxd.pdb xyzin
$sfall
TITLE Structure factors calculated for toxd.
GRID 152 96 64 !div CELL by these should give ., 0.7 A
MODE SFCALC XYZIN HKLIN
RESOLUTION 37 2.1
BINS 60
RSCB 8.0 2.1
SFSG 19
LABIN FP=FTOXD3 SIGFP=SIGFTOXD3
LABOUT ALLIN FC=FCtoxd PHIC=PHICtoxd
END
$!
$!
$sigmaa HKLIN CCP4_SCR:toxd_sf_mir hklout CCP4_SCR:junk
TITLE Phase combination
RESOLUTION 40. 2. !# Resolution limits
RANGES 10 500 !# No of bins for analysing agst S,
ERROR
COMBINE PART 1

LABI FP=FTOXD3 SIGFP=SIGFTOXD3 PHIBP=PHI_mir WP=W_mir -
HLA=HLA HLB=HLB HLC=HLC HLD=HLD -
FC=FCtoxd PHIC=PHICtoxd

LABO PHCMB=PHCMB1 WCMB=FOMCMB1 FWT=FWT1 PHFWT=PHWT1
END
$!
$exit
$term:
$write sys$output "! run sf_calc first"
$exit
$term1:
$write sys$output "! run the mlphare.com procedure first"
@
-- sigmaa.com,v 45% L189 (Fundamental)-----
X menu-bar buffer
```



Liz Potterton

# Going graphical

CCP4 Program Suite 6.1.3 CCP4Interface 2.0.6 running on mdw45port Project: msc

**Data Reduction**

- Data Processing using Mosflm
- Import Integrated Data

Find or Match Laue Group

Scale and Merge Intensities

- Utilities
- Check Data Quality

5	14 Dec 08	FINISHED	refmac5	Restrained
4	14 Dec 08	FAILED	refmac5	Restrained
3	14 Dec 08	FINISHED	cad	[No title]
2	14 Dec 08	FINISHED	truncate	[No title]
1	14 Dec 08	FINISHED	import	[No title]

Directories&ProjectDir

View Any File

View Files from Job

Search/Sort Database..

Graphical View of Project

Delete/Archive Files..

Exit

Run Refmac5 Initial parameters from C:\xtal\msc\_project\CCP4\_DATABASE\5\_refmac5.def

Job title: Restrained refinement using isotropic B factors

Do: restrained refinement using no prior phase information

Input fixed TLS parameters: no twin refinement

Run CootFindwaters to automatically add/remove waters to refined structure

MTZ in: msc 1flm\_downloaded.mtz FP: F Sigma: SIGF

MTZ out: msc 1flm\_downloaded\_refmac2.mtz

PDB in: msc 1flm\_noligand.pdb

PDB out: msc 1flm\_noligand\_refmac2.pdb

LIB in: msc Merge LIBIN

Output lib: msc 1flm\_noligand.cif

Include keyword file: msc

Data Harvesting

Refinement Parameters

Setup Geometric Restraints

Setup Non-Crystallographic Symmetry (NCS) Restraints

No NCS restraints are currently defined

Monitoring and Output Options

Scaling

Geometric parameters

Run Save or Restore Close

Loggraph 2\_truncate.log

Cumulative intensity distribution (Acentric and centric)

Truncate style Wilson plot

Acentric moments of E for k=1,3,4

Centric moments of E for k=1,3,4

Cumulative intensity distribution

Anisotropy analysis (Yorgo Modis)

Tables in File

Graphs in Selected Table

Cumulative intensity distribution (Acentric and centric)

- ccp4i released 2000
- Program scripts prepared by task interfaces
- Convenient access to program output
- Management of user data into projects

Big step forward at the time !

# Automated structure solution ...

- Increasing sophistication of crystallographic programs
- Development of software pipelines
  - some intelligence, but also parameter scanning
- Faster computers to support pipelines

Crank

XIA2

MrBUMP

Balbes

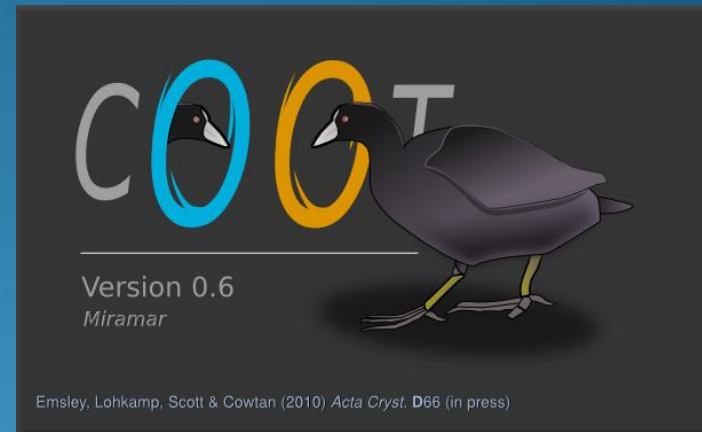
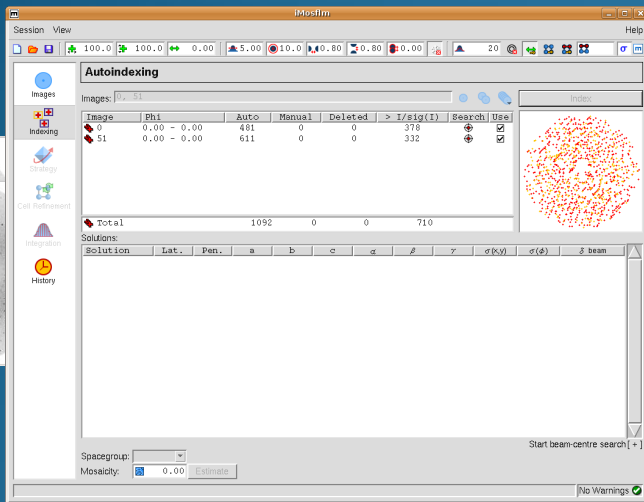
Potential to do most of the work for you. But ...

Always those difficult cases

You still need to get to know your structure!



# ...or interactive structure solution



Indexing / integration  
Laue group determination with  
Pointless ("QuickSymm")  
Scaling with Scala  
("QuickScale")

Model building  
Refinement with Refmac / Shelxl  
Validation with Molprobit tools  
(reduce/probe)

- Many ways to access the ~200 programs
- Programs hidden by graphical interfaces or become components of pipelines
- CCP4 programs used by many 3rd party pipelines

Now look at available programs in more detail ...



# MX structure solution

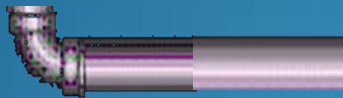
Data  
processing



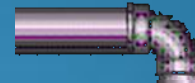
Experimental  
phasing



Molecular  
Replacement



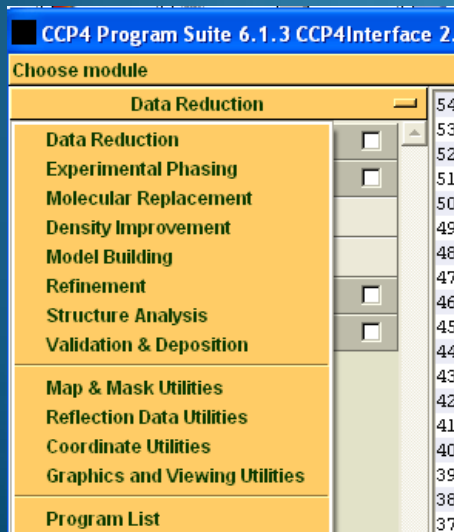
Phase improvement /  
Model (re)building



Model completion /  
refinement

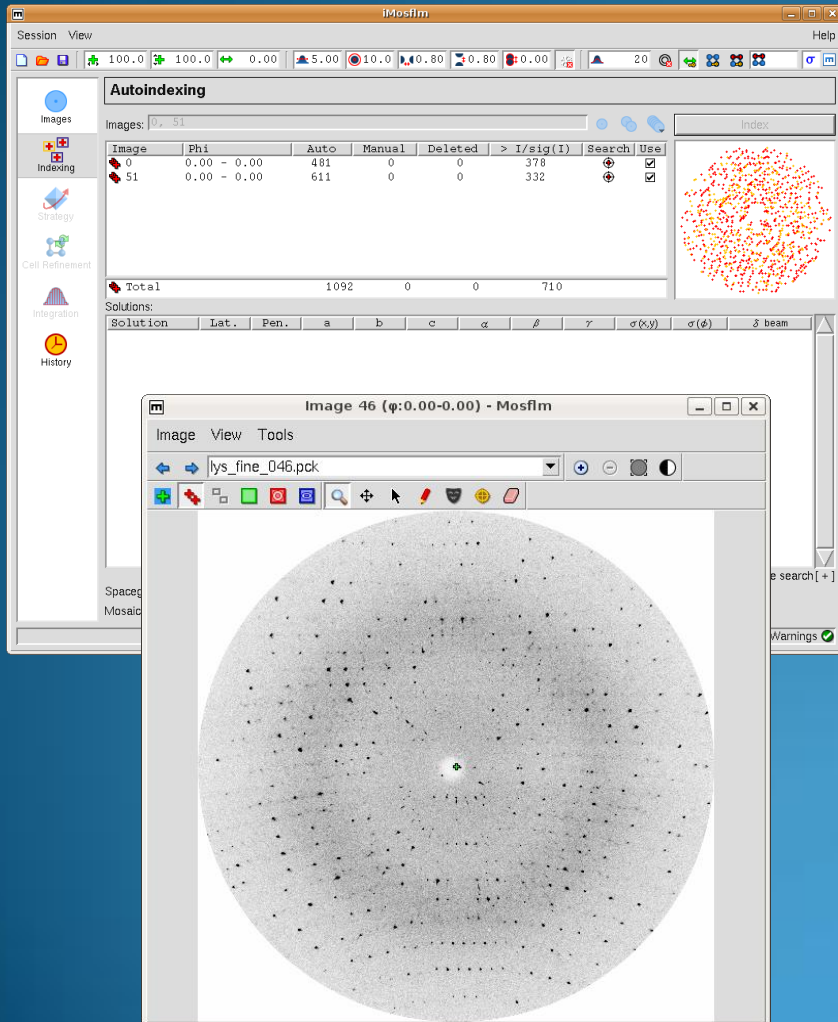


Validation



*In reality, not such a simple pipeline ...*

# Data processing (1)



Images → unmerged MTZ file

- Inspect images
- Identify spots and index
- Integrate
- Mosflm / iMosflm



Andrew Leslie



Harry Powell

Automation

XIA2

Multiple datasets →  
merged MTZ file

"semi-automatic" (describe  
experiment and go)



Graeme Winter

# Data processing (2)

## Identifying Laue group of data

pointless

## Scaling and merging reflections

scala

pointless/scala can be called from iMosflm

## Deriving structure factor amplitudes

ctruncate

$I(+)$  /  $I(-)$   $\rightarrow$   $F(+)$  /  $F(-)$

required for map coefficients, but intensities preferred for phasing/refinement

## How good is your data ??

check for twinning, incorrect spacegroup assignment, anisotropy, translational NCS etc.

scala, ctruncate, sfcheck



Phil  
Evans

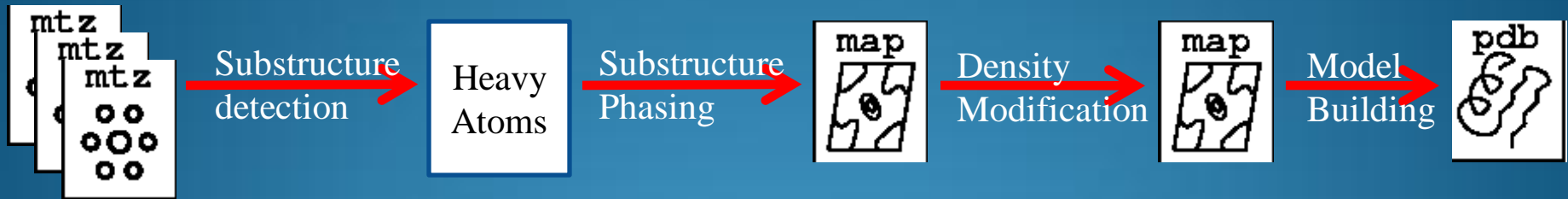
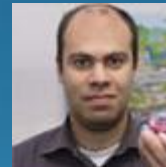


Norman  
Stein

# Experimental phasing (1)



Raj Pannu



Afro/Crunch2  
ShelxC/D

BP3

Solomon  
DM  
Pirate  
ShelxE

Buccaneer  
ARP/wARP

# Experimental phasing (2)

Other CCP4 phasing programs:

**Mrphare** -

Old but fast, still used e.g. in auto-Rickshaw

**Acorn** -

Ab initio structure determination using small starting structure  
(even randomly placed atom)

Requires atomic resolution, or artificially extended data

**Oasis** -

Dual-space direct methods for SAD/SIR

**Phaser** -

SAD phasing

Log-likelihood-gradient (LLG) maps to find additional anomalous sites

Alternate cycles of LLG completion and substructure refinement

Input anomalous atom sites and/or partial structure

⇒ combine experimental phasing with MR



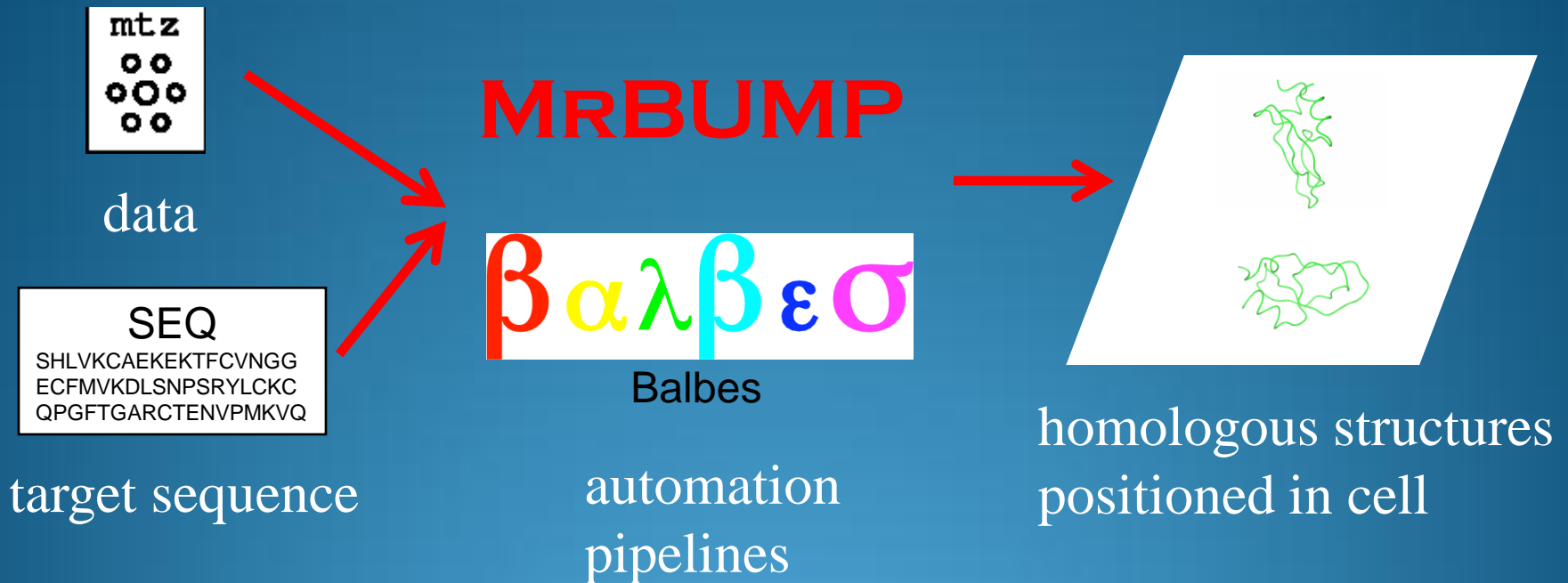
Eleanor Dodson



Randy Read



# Molecular Replacement (1)



Ronan Keegan



Fei Long



# Molecular Replacement (2)

Automation pipelines link existing programs.

## Data analysis:

No. copies in cell, rotational/translational NCS  
matthews\_coef, molrep, polarrfn, sfcheck, fft

## Model generation:

Find and edit homologous structures  
chainsaw, molrep

## MR programs:

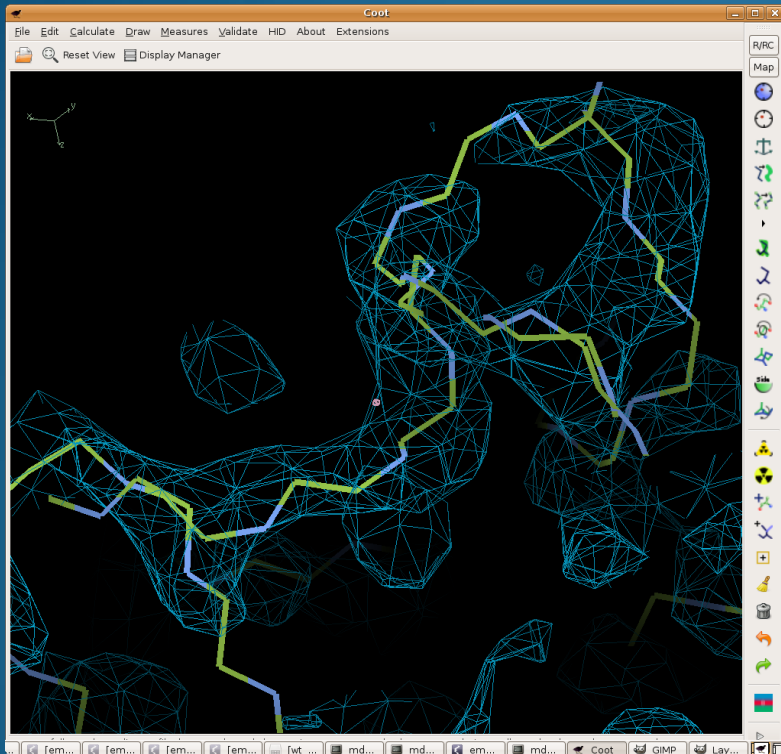
molrep - also multi-copy searches, map and NMR search models,  
phased RF and TF ...

phaser - also combined with SAD phasing



Alexei Vagin

# Phase improvement / model (re)building



## Phase Improvement or Density Modification

DM / Pirate / Parrot

Solomon

Dmmulti - multicrystal averaging

## Model building

Buccaneer

Rapper - loop modelling, model  
rebuilding

Sloop - loop building

Coot

Sequins - validates model side chains  
against electron density



Kevin Cowtan



# Model completion / refinement (1)

## Refinement with Refmac:

Optimise model parameters e.g.  $\{x,y,z\}$

You choose form of model (atomic vs group parameters, scaling function, etc.)

Refmac included in Arp/Warp, Buccaneer, Coot, etc.

## Latest developments:

- Twin refinement - automatic detection of significant twin operators and refinement of twin fractions
- Refinement against SAD data - anomalous scatterers included in XYZIN
- Automatic weighting Xray vs geometry
- External restraints, used in latest ARP/wARP

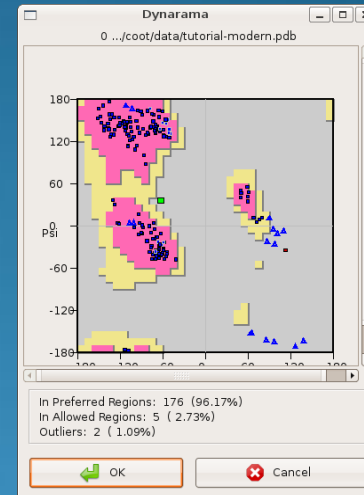


Garib Murshudov

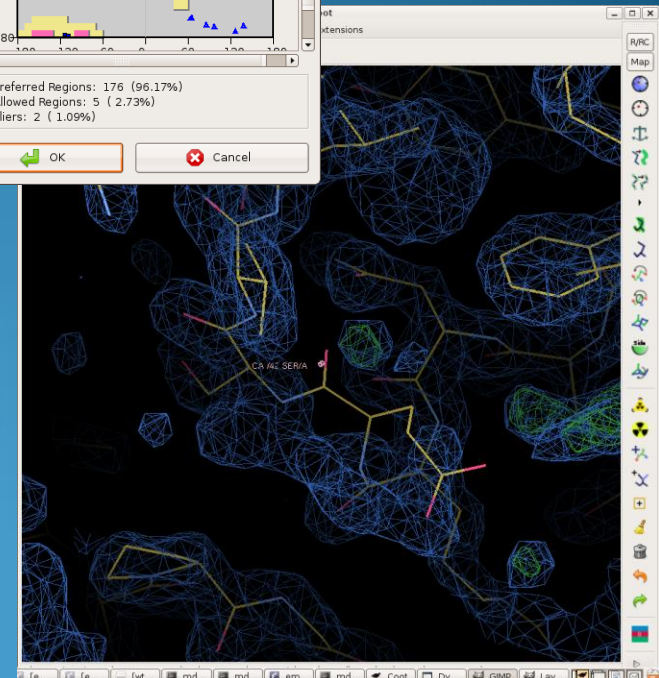
# Model completion / refinement (2)

## Coot

- Inspection of models and maps
- Model building
- Validation
- Real space refinement
- Refinement with Refmac
- NCS tools
- Sequence assignment
- Inclusion of Molprobit tools



Paul Emsley



# Validation

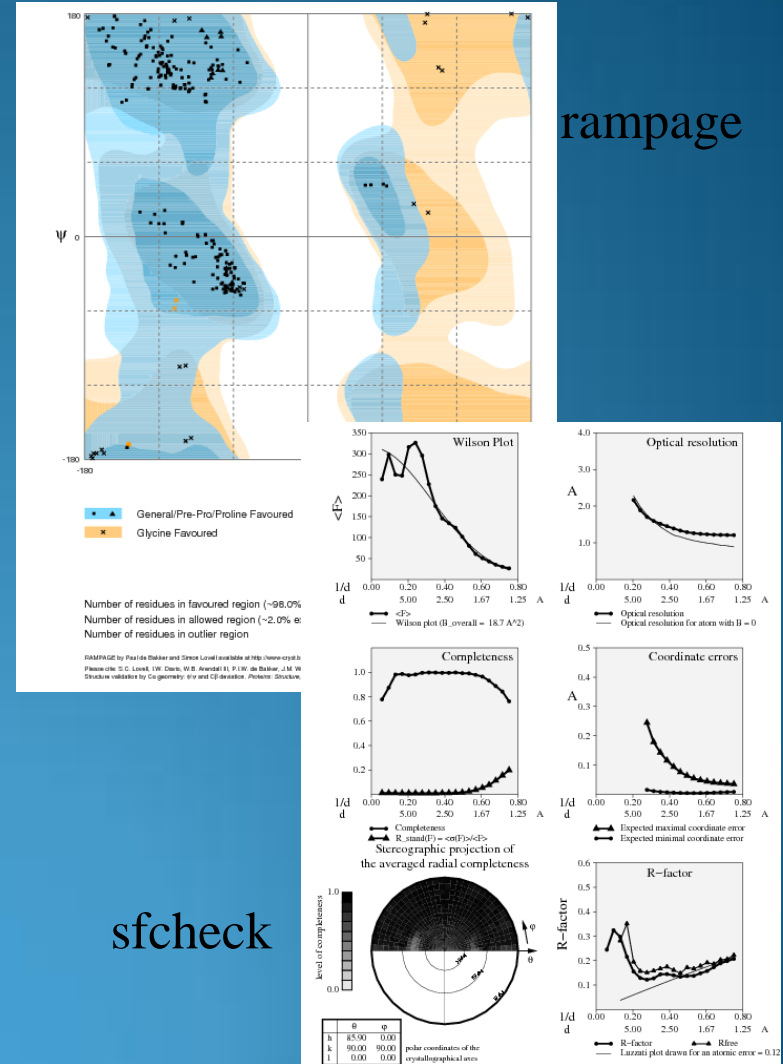
## Model:

appropriate level of detail?  
data/parameter ratio  
restraints  
geometry

## Fit to data:

R / Rfree  
Choice of free reflections (NCS?)  
Electron density support for ligands

rampage



sfcheck

# Deposition / publication

CCP4mg

New QT version

Publication quality images

Wide variety of representations

Surfaces and electron density supported

Easy to produce movies

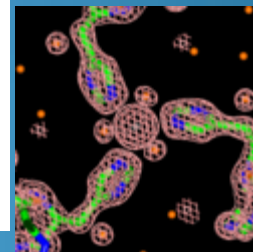
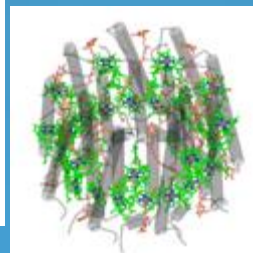
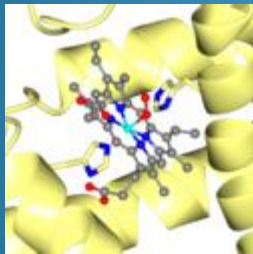
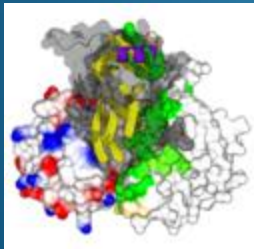
Transfer of structures/views to and from Coot



Liz Potterton



Stuart McNicholas



# Data organisation?

## A good idea!!

- More datasets, collected faster
- Can you remember what x2\_t7\_refmac1.mtz from Dec 08 is?
- Automation schemes produce files from many trials
- External requirements, e.g. BBSRC Data Sharing policy, journal requirements

## Tools available:

PiMS & xtalPiMS

synchrotron databases, e.g. ISPyB

ccp4i projects

lab databases

public databases e.g. TARDIS for images



Chris Morris



Rob Esnouf





# Getting help

The screenshot shows the CCP4 website interface. At the top, there is a navigation bar with links for Home, About CCP4, CCP4 Projects, Downloads, Documentation, Courses, Developers, and CCP4 people. Below this is a banner for the Collaborative Computational Project No. 4. The main content area is divided into several sections: Resources, CCP4 Documentation Wiki, CCP4 Software Suite, CCP4 Study Weekend 2010, Courses and Events, and PX Vacancy Postings. Red arrows point from external text to specific links and sections on the page.

wiki

You Are Here!

courses  
(CCP4 organised,  
CCP4 funded,  
CCP4 attended)

check here  
for current  
known issues

ccp4bb

# Future plans (1)

## New programs:

PRODRG\_ccp4, RAPPERtk, SLOOP, JLigand

## Distribution of CCP4:

- Monolithic suite & individual program updates
- Incremental updates e.g. via rsync
- Web services:
  - CPU-intensive applications, e.g. MrBUMP
  - Useful utilities
  - Updated programs

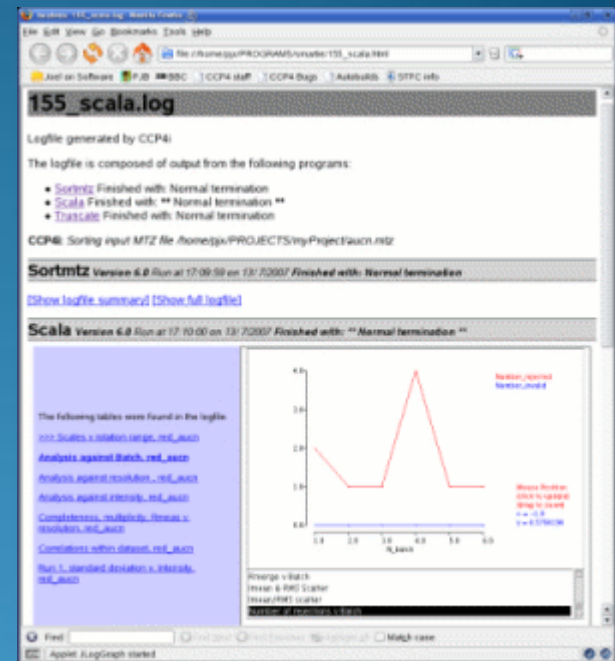
# Future plans (2)

## New GUI:

- Development just started ...
- Greater support for automation
- Better presentation of results

## Data management:

- Possible support for local back-end databases
- Possible support for remote databases / services - "web-based computing"



Log file reformatted by Baubles and displayed in web browser



# New core team



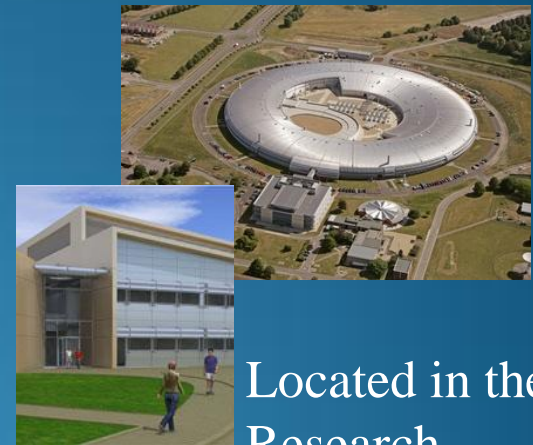
Eugene Krissinel



George Pelios



Natalie Zhao



Located in the  
Research  
Complex at  
Harwell

Aided by the old hands ...



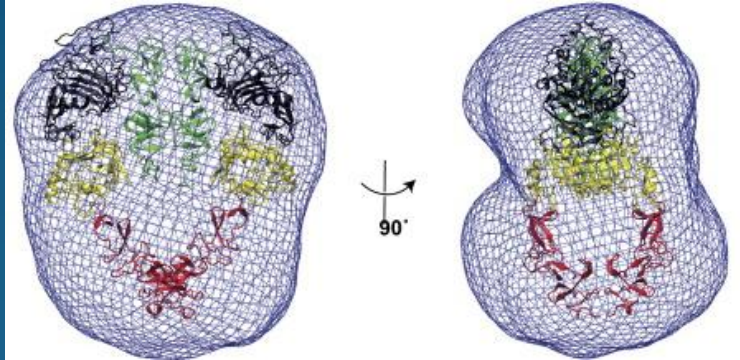
Charles Ballard



Ronan Keegan

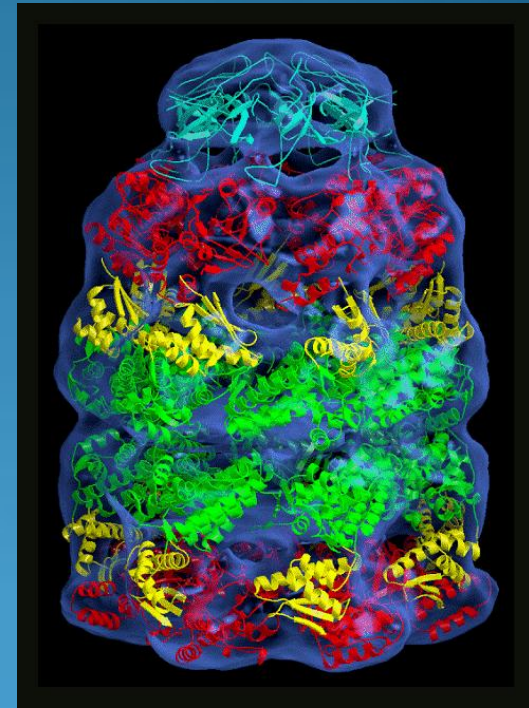
# Using Structures in other techniques

Small-angle X-ray scattering (**SAXS**)



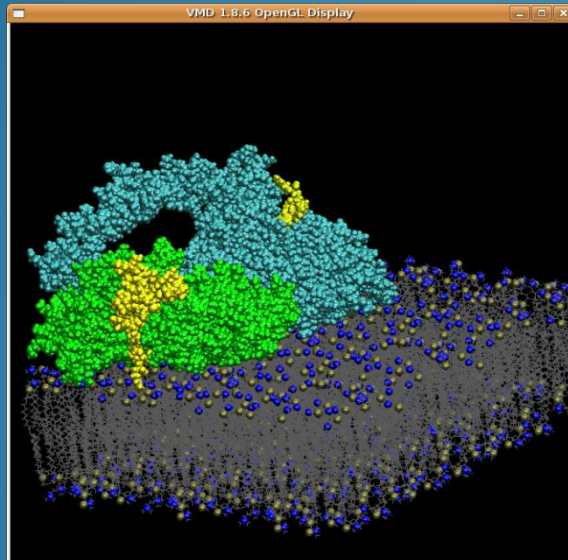
Dawson et al., *Structure* 15, 942–954 (2007)

Cryo-electron microscopy



Ranson et al. (2001) *Cell* 107, 869-879

Molecular  
Dynamics  
Simulations



Kaestner et al., *J. Struct. Biol.*, **167** 117–128 (2009)

# Structural Systems Biology?

Structures in a cellular context - combining structural genomics and systems biology.

System of 478 proteins:

120 expt.

336 automatic modelling

3 manual modelling

19 TM proteins

503 metabolites

645 reactions

⇒ non-random distribution of folds

## Three-Dimensional Structural View of the Central Metabolic Network of *Thermotoga maritima*

Ying Zhang,<sup>1\*</sup> Ines Thiele,<sup>2\*†</sup> Dana Weekes,<sup>3</sup> Zhanwen Li,<sup>1</sup> Lukasz Jaroszewski,<sup>3</sup> Krzysztof Ginalski,<sup>4</sup> Ashley M. Deacon,<sup>5</sup> John Wooley,<sup>6</sup> Scott A. Lesley,<sup>7</sup> Ian A. Wilson,<sup>8</sup> Bernhard Palsson,<sup>2</sup> Andrei Osterman,<sup>9</sup> Adam Godzik<sup>1,3,6‡</sup>

Metabolic pathways have traditionally been described in terms of biochemical reactions and metabolites. With the use of structural genomics and systems biology, we generated a three-dimensional reconstruction of the central metabolic network of the bacterium *Thermotoga maritima*. The network encompassed 478 proteins, of which 120 were determined by experiment and 358 were modeled. Structural analysis revealed that proteins forming the network are dominated by a small number (only 182) of basic shapes (folds) performing diverse but mostly related functions. Most of these folds are already present in the essential core (~30%) of the network, and its expansion by nonessential proteins is achieved with relatively few additional folds. Thus, integration of structural data with networks analysis generates insight into the function, mechanism, and evolution of biological networks.

Science, **325**, 1544 (2009)



# Partial acknowledgements

## **DL/RAL/DLS:**

Charles Ballard, Ronan Keegan, Norman Stein, Peter Briggs, Eugene Krissinel, George Pelios, Natalie Zhao, Graeme Winter, Chris Morris

## **MRC-LMB:**

Phil Evans, Andrew Leslie, Harry Powell

## **Cambridge:**

Randy Read, Airlie McCoy, Gabor Bunkoczi

## **York:**

Eleanor Dodson, Kevin Cowtan, Garib Murshudov, Alexei Vagin, Liz Potterton, Stuart McNicholas, Fei Long

## **Oxford:**

Paul Emsley, Martin Noble, Rob Esnouf

## **Leiden:**

Raj Pannu, Pavol Skubak



Science & Technology  
Facilities Council



MRC

Medical  
Research  
Council