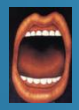
Overview of CCP4 suite and current developments

Martyn Winn, STFC Daresbury Laboratory



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

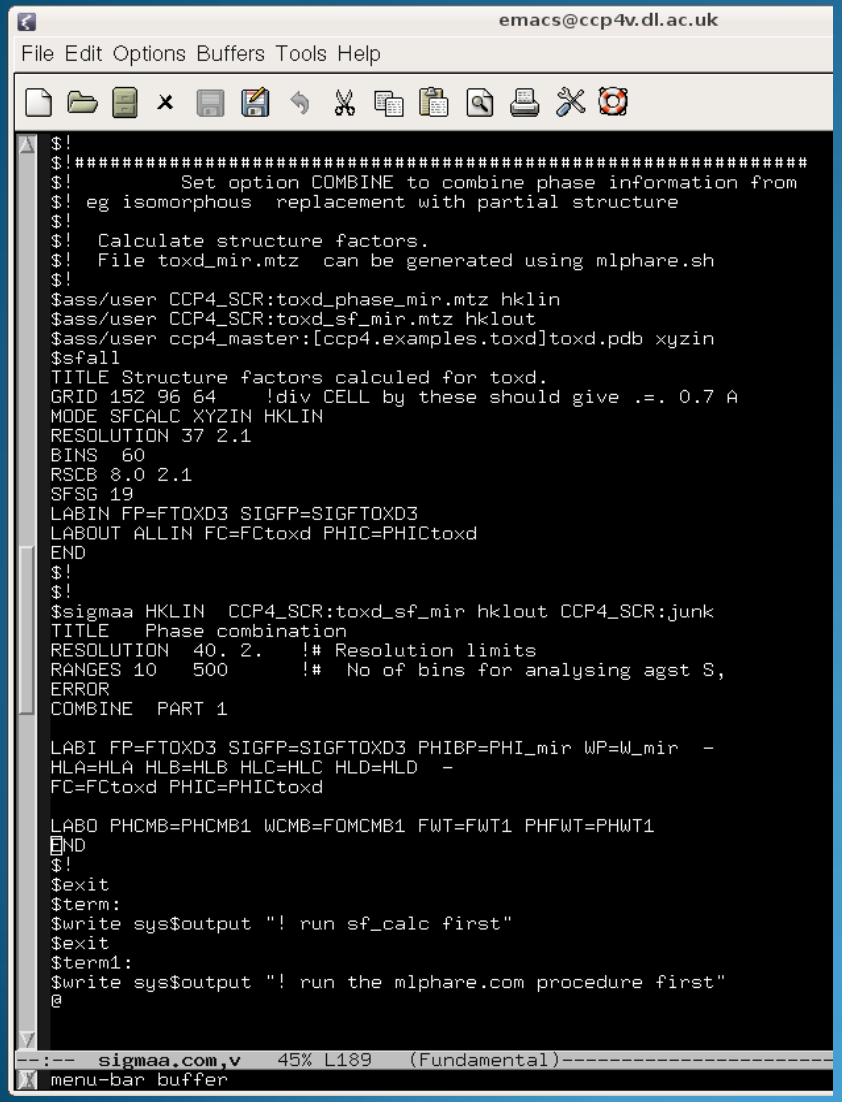
Philosophy

- ~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.

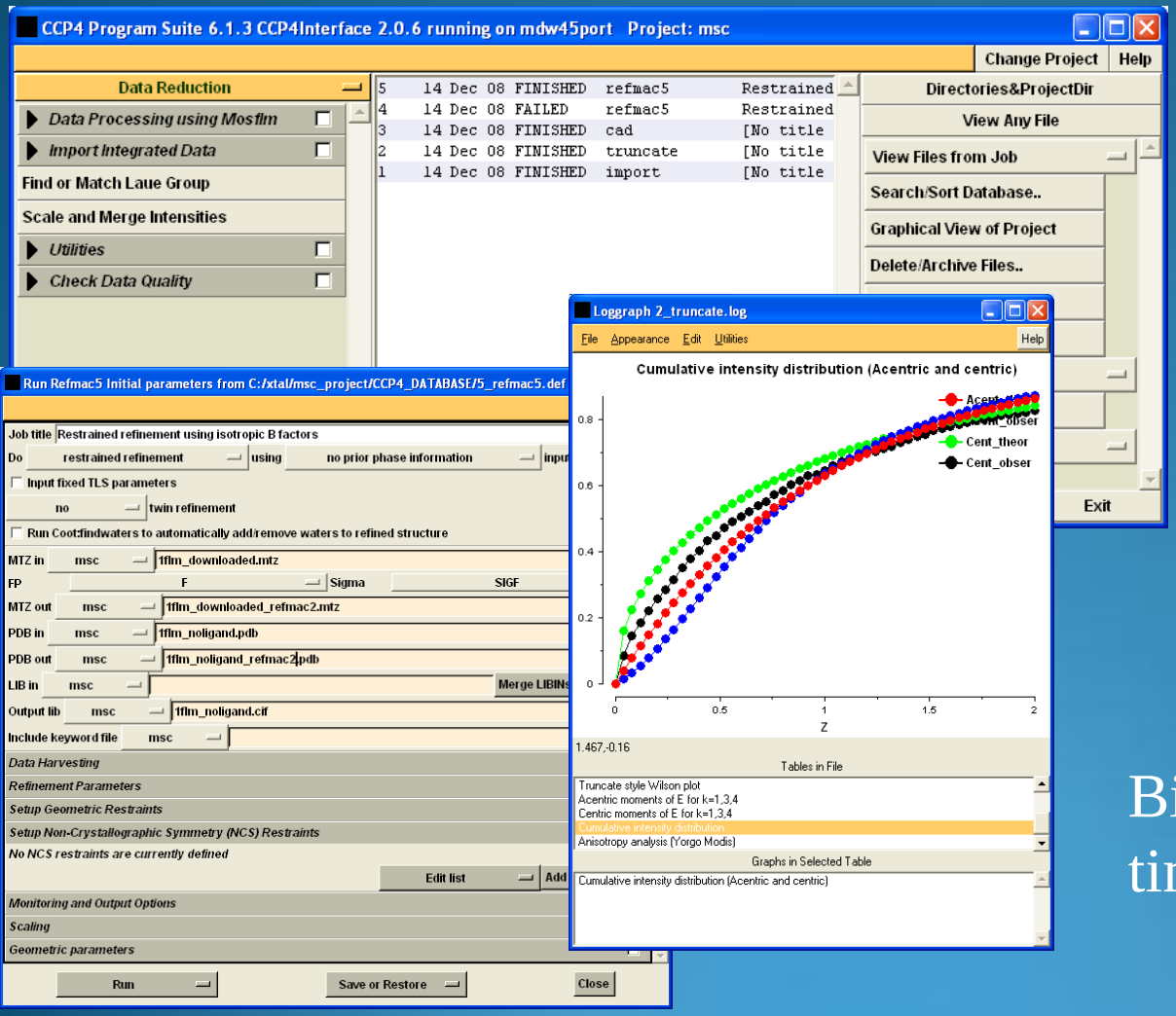
In the beginning



- 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

Liz Potterton

Going graphical



- 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

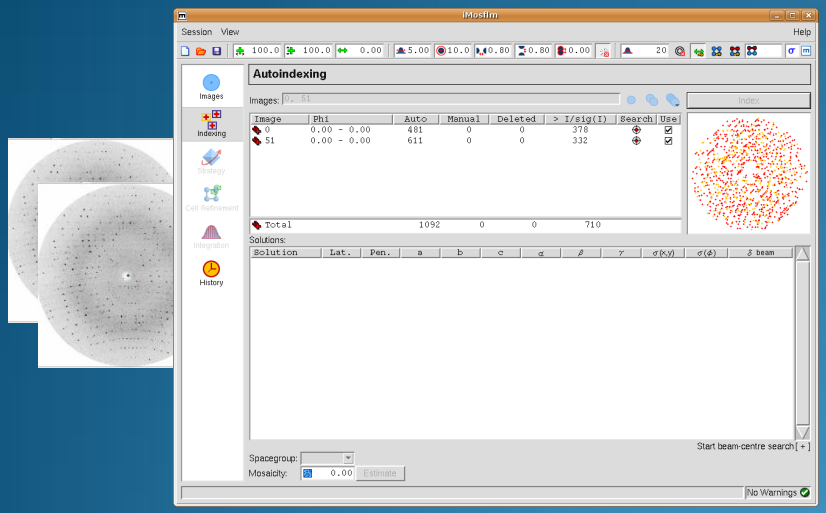
Crank
XIA2
MrBUMP
Balbes

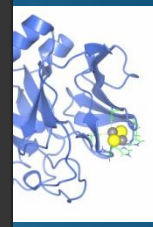
Potential to do most of the work for you. But ...

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 Molprobity 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

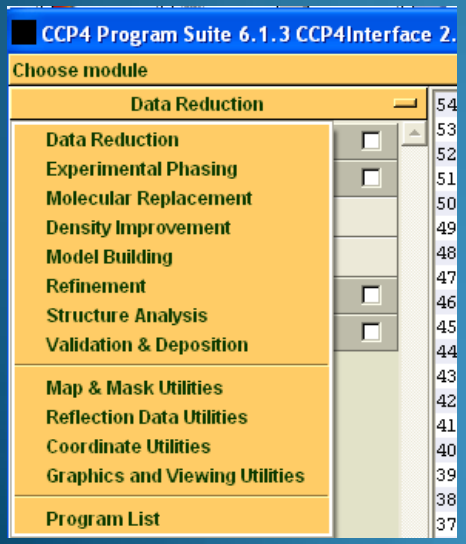
Experimental phasing

Molecular Replacement

Phase improvement / Model (re)building

Model completion / refinement

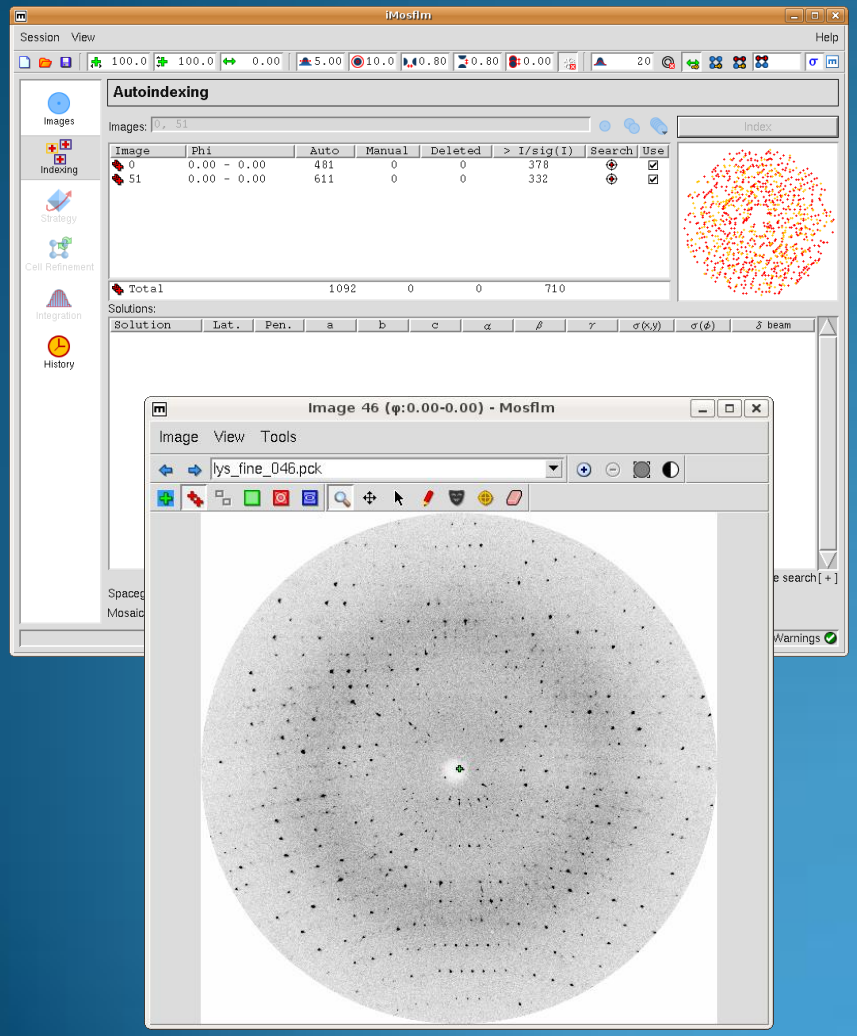
Validation



In reality, not such a simple pipeline ...

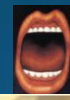
WORLDWIDE PDB
PROTEIN DATA BANK

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

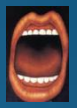
pointless

Scaling and merging reflections

scala

pointless/scala can be called from iMosflm

Phil Evans



Deriving structure factor amplitudes

ctruncate

I(+) / I(-) → F(+) / F(-)

Norman Stein

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

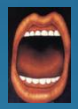
How good is your data ??

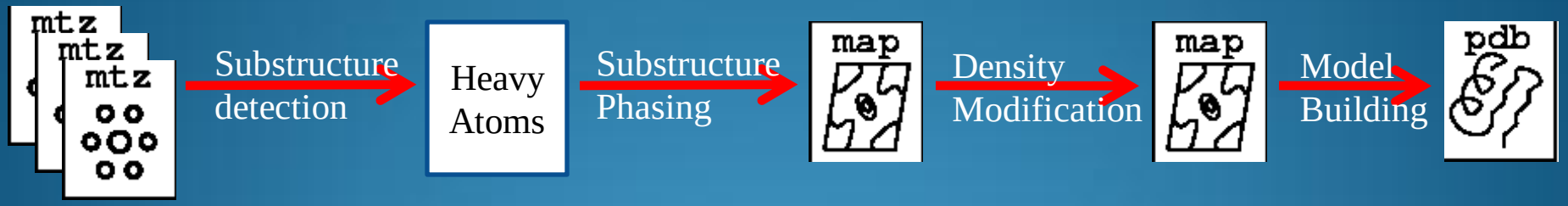
check for twinning, incorrect spacegroup assignment, anisotropy, translational NCS etc.

scala, ctruncate, sfcheck

Experimental phasing (1)

Raj Pannu





Afro/Crunch2
ShelxC/D

BP3

Solomon
DM
Pirate
ShelxE

Buccaneer
ARP/wARP

Experimental phasing (2)

Other CCP4 phasing programs:

Mlphare -

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

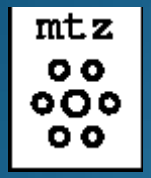
Input anomalous atom sites and/or partial structure

$\Rightarrow$ combine experimental phasing with MR

Eleanor Dodson

Randy Read

Molecular Replacement (1)



data

SEQ
SHLVKCAEKEKTFCVNGG
ECFMVKDLSNPSRYLCKC
QPGFTGARCTENVPMKVQ

target sequence

MrBUMP

βαλβεσ

Balbes

automation
pipelines

homologous structures
positioned in cell

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

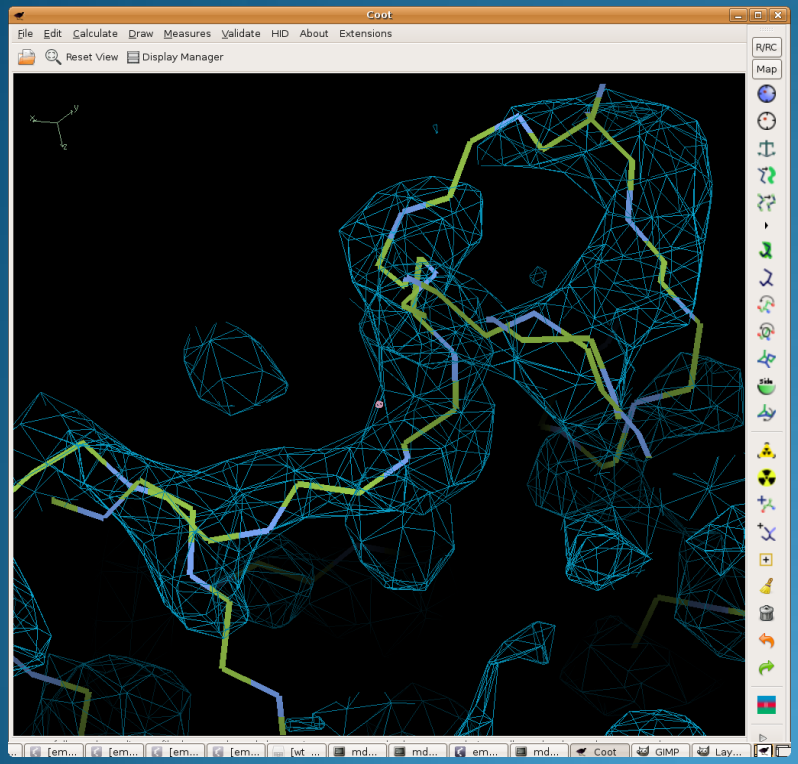
Alexei Vagin

MR programs:

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

phaser - also combined with SAD phasing

Phase improvement / model (re)building



Kevin Cowtan

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

Model completion / refinement (1)


Garib Murshudov

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.

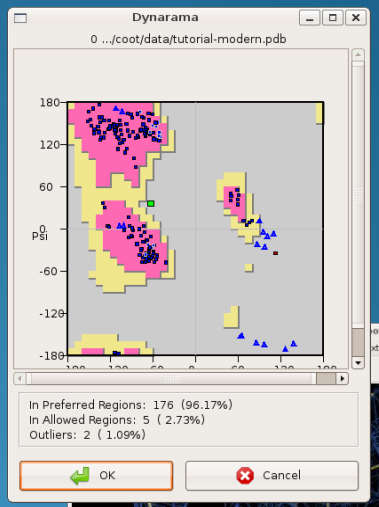
Latest developments:

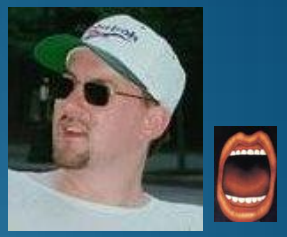
- Twin refinement - automatic detection of significant twin operators and refinement of twin fractions
- Refinement against SAD data - anomalous scatterers included in XYZIN
- Automatic weighting Xray vs geometry
- External restraints, used in latest ARP/wARP

Model completion / refinement (2)

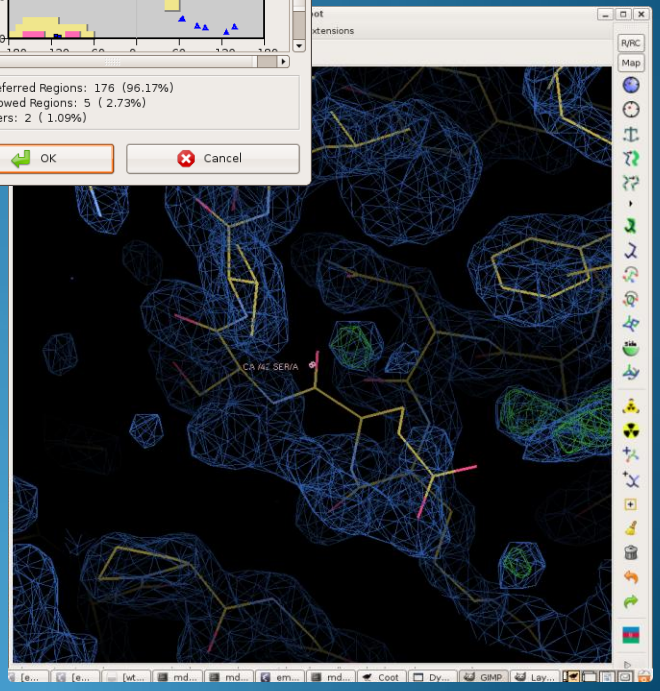
Coot

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





Paul Emsley



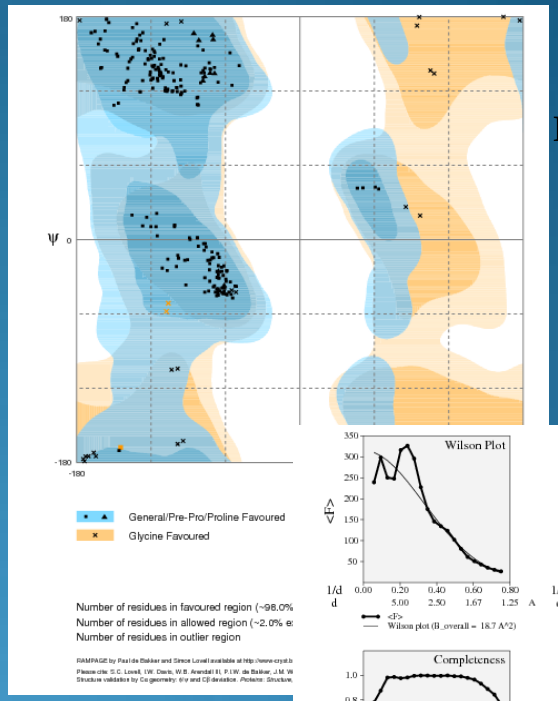
Validation

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

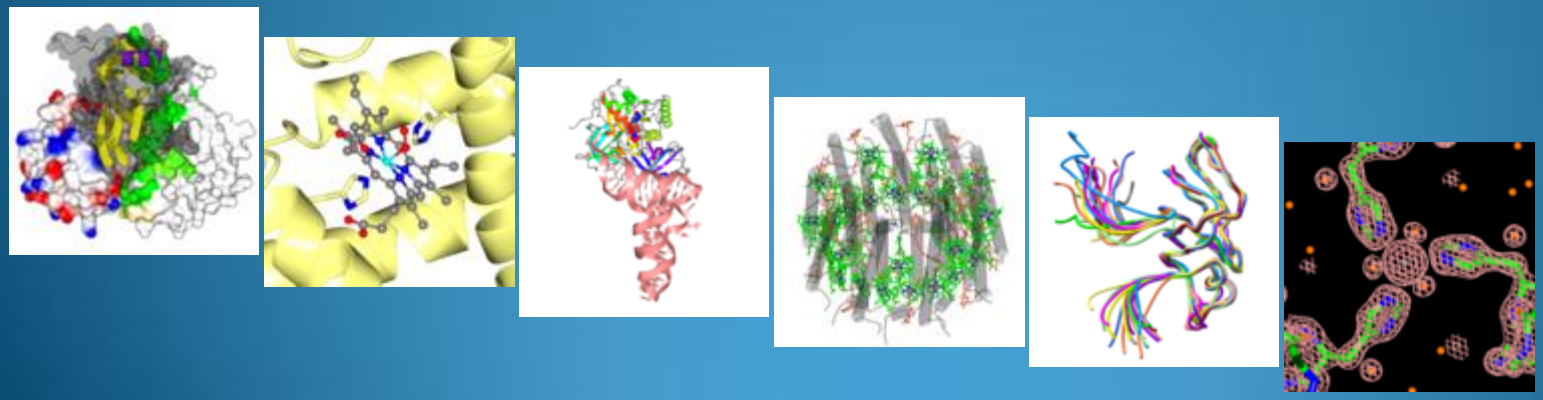
Deposition / publication

Liz Potterton

Stuart McNicholas

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



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

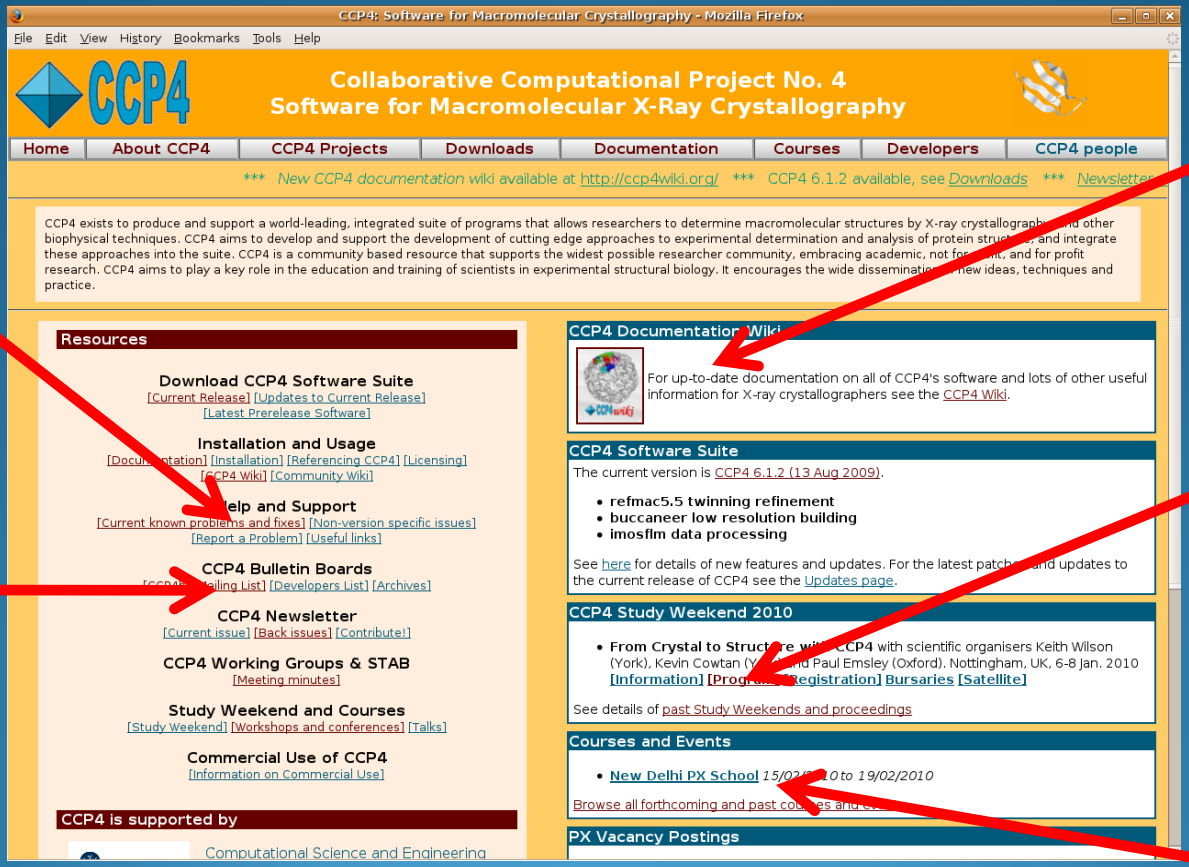
Chris Morris

Rob Esnouf

Tools available:

PiMS & xtalPiMS
synchrotron databases, e.g. ISPyB
ccp4i projects
lab databases
public databases e.g. TARDIS for images

Getting help



check here for current known issues

ccp4bb

wiki

You Are Here!

courses (CCP4 organised, CCP4 funded, CCP4 attended)

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

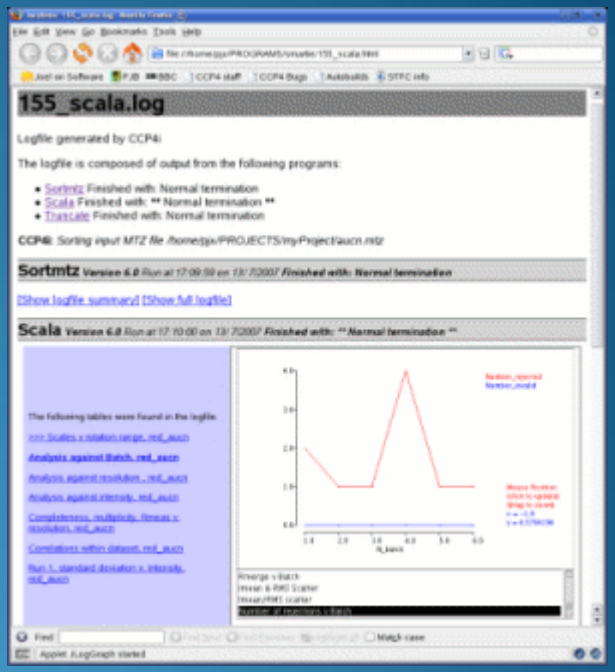
Future plans (2)

New GUI:

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

Data management:

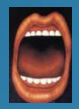
- Possible support for local back-end databases
- Possible support for remote databases / services - "web-based computing"



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

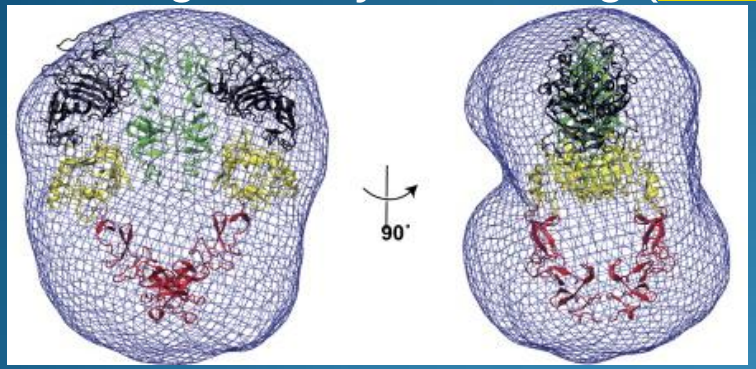
Aided by the old hands ...

Charles Ballard

Ronan Keegan

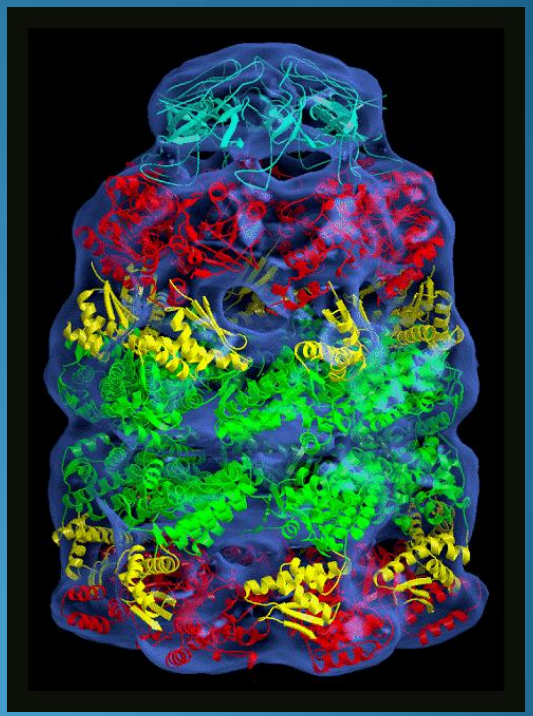
Using Structures in other techniques

Small-angle X-ray scattering (SAXS)



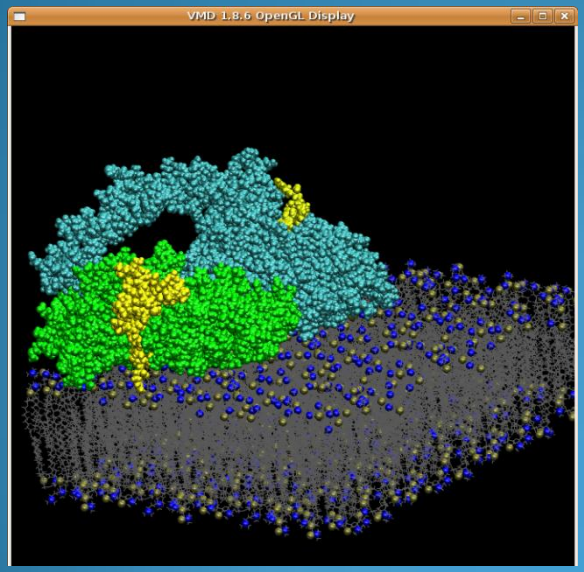
Dawson et al., Structure 15, 942–954 (2007)

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

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

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

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)

⇒ non-random distribution of folds

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