

“From Isolated Programs to a Full Package”

Or

“Should/Can Software be developed by a team?”

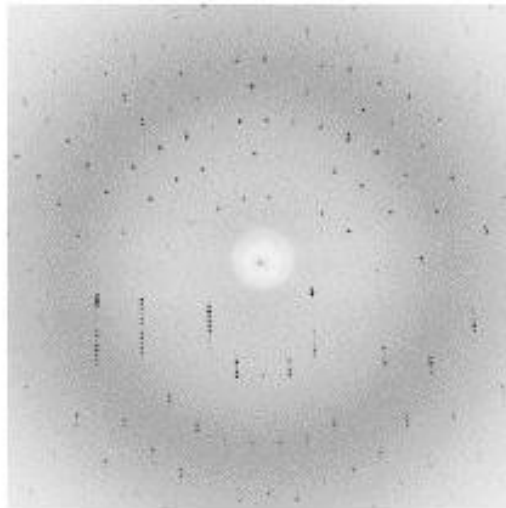
***Eleanor Dodson
August 2017***

The task for crystallographers: Crucial steps in X-ray structure solution

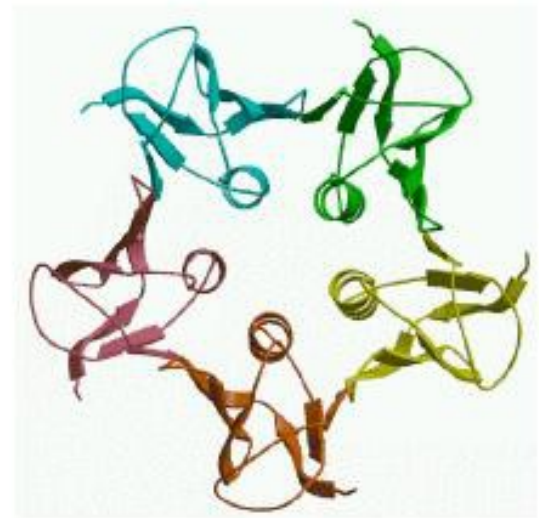
Crystal



Diffraction



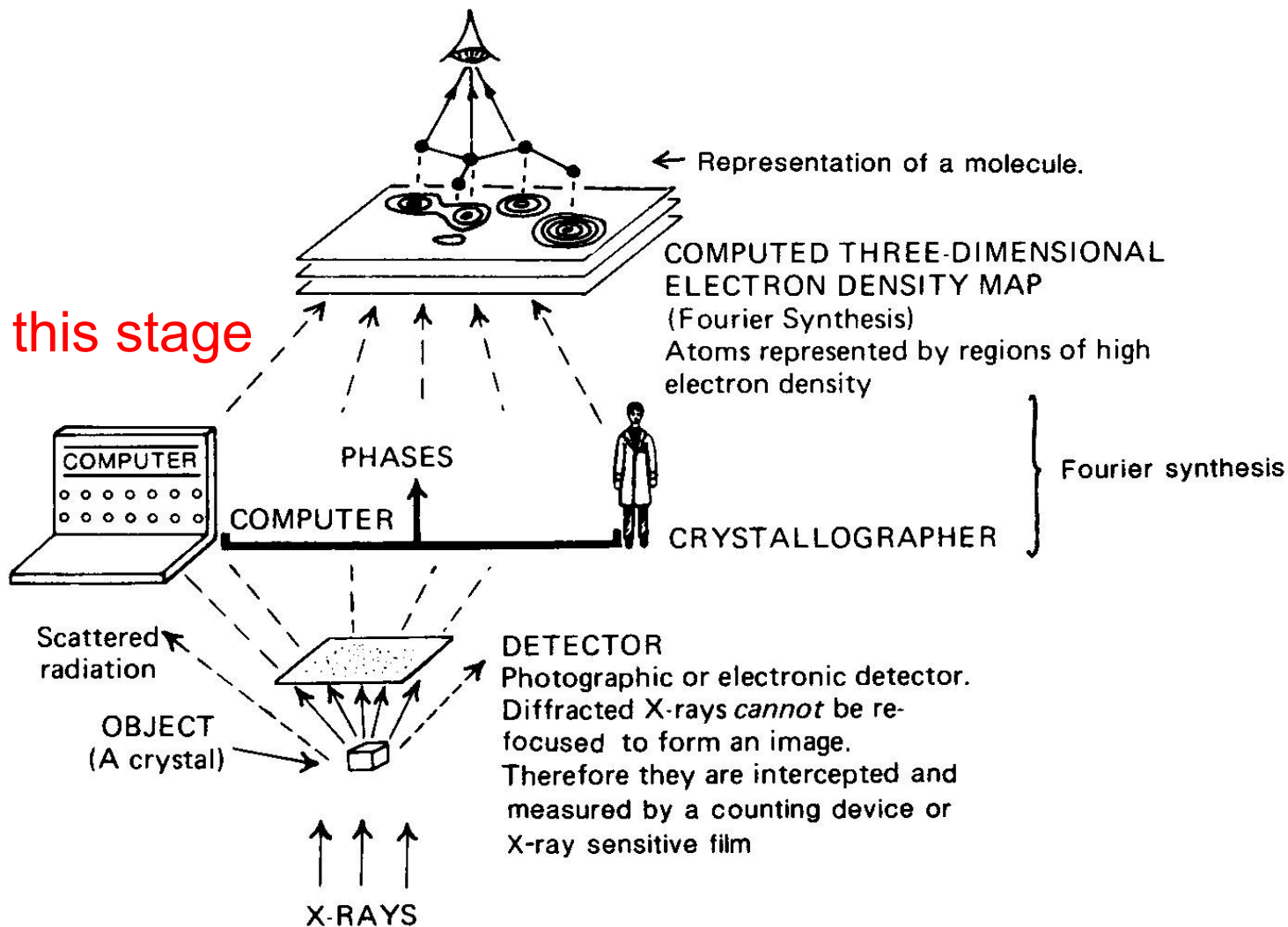
Interpretation



Crucial steps in X-ray structure solution

Figure by Jenny Glusker – DCH's "best" student

Maths at this stage



Summary slides from York crystallography course

Marek Brzozowski

Structure:

✓ known unit cell and its chemical content
(-) unknown positions of atoms

Diffraction pattern:

✓ known amplitudes $|F|$
(-) unknown phases

$$\rho(xyz) = 1/V \sum_h \sum_k \sum_l |F_{hkl}| \exp [-2\pi i (hx + ky + lz) + i \alpha_{hkl}]$$

1) Need to measure and process I_{hkl} and convert to F_{hkl}

Structure:

✓ known unit cell and its chemical content
(-) unknown positions of atoms

Diffraction pattern:

✓ known amplitudes $|F|$
(-) unknown phases

$$\rho(xyz) = 1/V \sum_h \sum_k \sum_l |F_{hkl}| \exp [-2\pi i (hx + ky + lz) + i \alpha_{hkl}]$$

2. Phase-problem solution Methods α_{hkl}

i) Direct methods (purely computational):

- do not require any extra information – use experimental data set $|F_{hkl}|$ only.
- trying to find some statistical relationship between strong reflections
- estimate phases for them and expand them for the whole data set
- these cycles can be repeated till electron density maps make sense.
- **organic crystals only**

Phase-problem solution Methods cont.

(ii). Heavy atom methods (purely computational or experimental):

our **(in)organic** compound already **has heavy atom** (e.g. Fe, Pt, Au, Hg, Ru etc.):

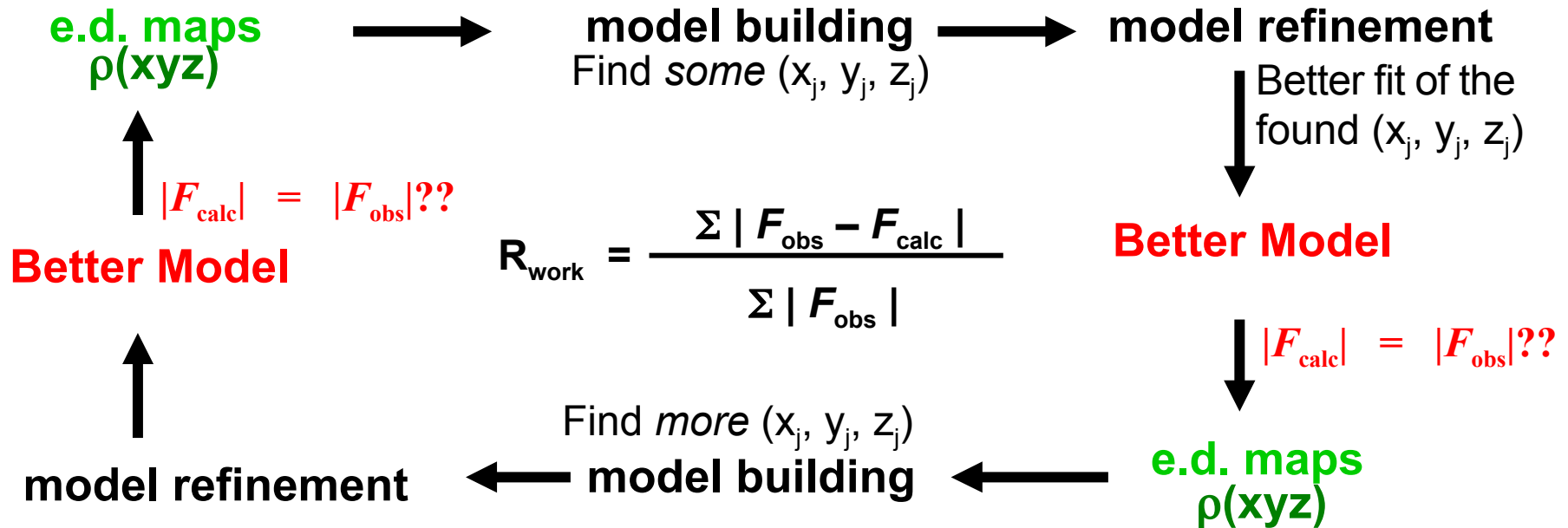
direct methods together with mathematical **Patterson method** can determine the position of the heavy atom in our crystal, which will greatly help in proper estimation of the phases.

- **Good for proteins crystals as well** (lots of solvent content there – 50%)

(iii) **Molecular Replacement method** – **for protein crystals only:**

if we already know structure of a similar protein.

3-4. Model building – refinement of the phases/structure



$$\rho(xyz) = 1/V \sum_h \sum_k \sum_l |F_{hkl}| \exp [-2\pi i (hx + ky + lz) + i \alpha_{hkl}]$$

$$\rho(xyz) = 1/V \sum_h \sum_k \sum_l |F_{hkl}| \exp [-2\pi i (hx_j + ky_j + lz_j)]$$

$$F_{calc} = \int_v \rho(x,y,z) \exp [2\pi i (hx + ky + lz)] dv$$

$$|F_{calc}| = |F_{obs}|?? \quad I_{calc} = I_{obs}??$$

Bayes Theorem (1763)

This is our aim..

To provide the **best model** with the highest probability, given a set of **observations** and some **prior knowledge**.

$$P(\text{Model}; \text{Obsns}) = \frac{P(\text{Model}) P(\text{Obsns}; \text{Model})}{P(\text{Obsns})}$$

What was and is needed? Lots of software..

Data recording, integration, processing, analysis

Fast Fourier routines

Experimental phasing

Density modification for Phase improvement

Molecular Replacement

Model building

Refinement

Validation

Utilities Etc Etc..

What was and is needed? Lots of software AND...

Communication, then more communication, then more..

From users to developers:

What are our problems? Poor diffraction? New ligand? ...

From developers to users.

What can you believe from the analysis?

Software contributions from 1960s (list incomplete!)

Data measurement, integration and processing

(Arndt, Wonnacott, Rossmann, Evans, etc etc etc)

Fast Fourier Transforms

(Lyn Ten Eyck, Bricogne, etc, etc etc)

Experimental Phasing – MIR, MIRAS, SIRAS, MAD, SAD

(Blow, Dodson, Matthews, Evans, Hendrickson, Otwinowski, etc etc etc)

Molecular Replacement

(Rossmann, Blow, Crowther, Navaza, Vaguine, Read, etc etc etc)

Phase Improvement and Model Building

(Wang, Kam, Cowtan & Main, Sheldrick, etc, etc)

Refinement

(Konnert & Hendrickson, Rollett, Isaacs & Agarwal, Baker & Dodson...)

And more ...

Graphics for model building

First Alwyn Jones, FRODO then O. Emsley – COOT

Bioinformatics

Graphics for display, analysis

Validation (and graphics) .

Now endless access to information



A THREE-DIMENSIONAL MODEL OF THE MYOGLOBIN MOLECULE OBTAINED BY X-RAY ANALYSIS

By Drs. J. C. KENDREW, G. BODO, H. M. DINTZIS, R. G. PARRISH and H. WYCKOFF

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge

AND

D. C. PHILLIPS

Davy Faraday Laboratory, The Royal Institution, London

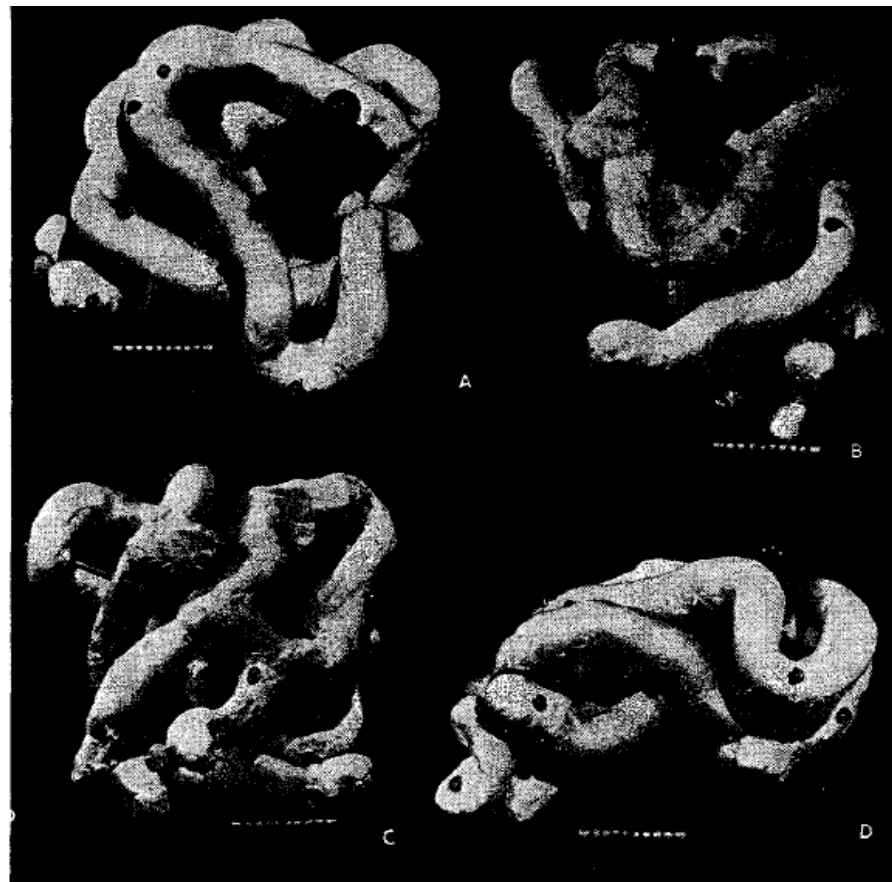


Fig. 2. Photographs of a model of the myoglobin molecule. Polypeptide chains are white; the grey disk is the haem group. The three spheres show positions at which heavy atoms were attached to the molecule (black: Hg of *p*-chloro-mercuri-benzene-sulphonate; dark grey: Hg of mercury diammine; light grey: Au of auri-chloride). The marks on the scale are 1 Å. apart

DCH & Siv Ramasechan – 1965 our Anomalous Dispersion Guru

S.Ramaseshan Advanced Methods of Crystallography ed by
G.N.Ramachandran (Academic, New York 1964)



Oxford Laboratory 1972?



Peter Main – Physics York 1964-2015

Direct Methods guru and superb programmer



Basic Outline of CCP4

- **Project established in 1979 and funded by UK Research Councils**
-
- **Project aims**
-
- **Provide software for macromolecular Crystallography. (MX)**
 - **Encourage collaborative development of software for MX.**
 - **Promote the teaching of MX.**
 -
 - **Provide organisation and guidance for projects.**

Why needed in 1979?

- **Self Interest - New MX labs needed mutual support**
-
- **With Maintenance and installation :**
-
- **With developing new algorithms. Accepted that no one group had/has a monopoly and that methods will be superseded.**
-
- **With having a forum for discussion, to distribute updates quickly, and to get feed back from users..**
-
- **All (small) funds were spent on meeting face to face. 3 monthly talking shops.**

1982 We had excellent computing facilities?



Zygmunt Derewenda refining haemoglobin

Finances – somewhat hazy!

1979 £3,500 – travel and 50% of Pella Machin

1980s on – funds for support staff – VERY important! Industrial contributions grow..

1998 Support plus funds for 5 years for Garib Murshudov

2003 Support plus funds for 5 years for Eugene Krissenel

2008 Support plus funds for 5 years for Paul Emsley

2013 Support plus funds for 5 years for Kevin Cowtan

Always some money to subsidise Study Weekends

Why has it survived?

- **Easy to update – modular approach means routines can be replaced without disrupting the working suite**
-
- **There is a great deal of developer support. Programs are distributed quickly; help with style, preparing documentation, etc etc**
-
- **There is a great deal of user support. The approach has been genuinely cooperative; users are in the driving seat, and program authors respond. Study weekend, workshops mutually beneficial**
-

Workshop 1998



•CCP4 Study Weekend - 2006



Bangalore 2008



What helps algorithm development?

- **Common Data structures (formats?)**
- **Library routines for data handling, crystallographic operations (Symmetry, FFTs, etc) :**
 - **Libraries must be accessible to developers and there needs to be a way to update them to add routines, and to debug.**
 - **They must be well documented and curated.**
 - **Common output style helps users**

- **Library routines and data objects -1979-2017**

CCP4 libraries originally in Fortran

Now in C++

much more comprehensive, professional and robust.

(Cctbx clipper mmdb etc..)

Accessible via scripting languages – Python,
Scheme.

*Raises a Q. Should legacy code be rewritten to
use new libraries. Would anyone do it!*

Example of algorithm development - 2015

Email from Kevin Cowtan: Hi folks,
Do we need some standardisation in input of
sequence types. At the moment:

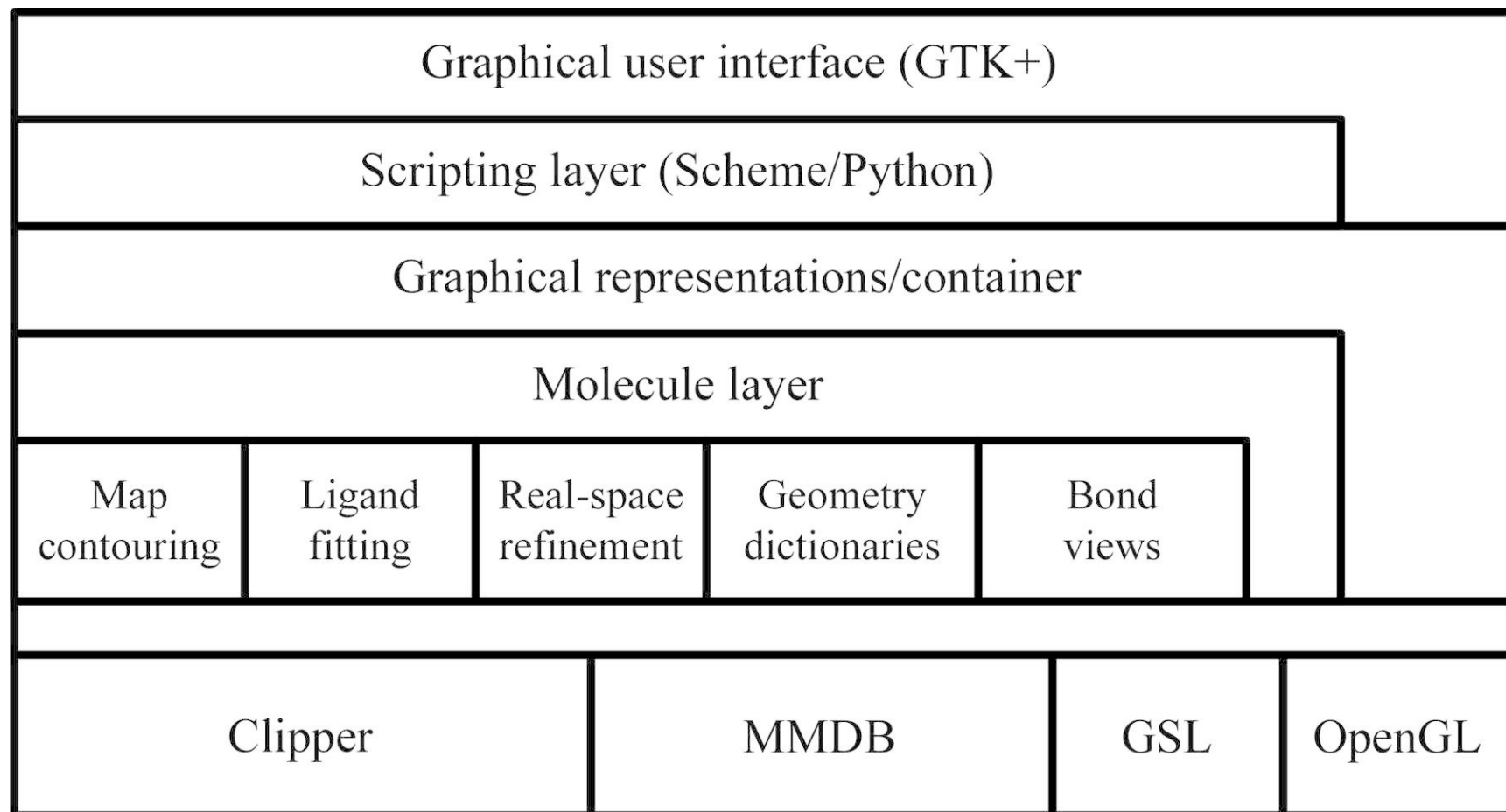
1. Buccaneer, Parrot use **CseqDataFileList**
 2. CTruncate, MolRep, MrBump use **CSeqDataFile**
 3. Phaser and Matthews use **CAsuComponentList**
- Unfortunately changing any of these are going to be painful (in terms of old jobs not working). On the other hand the incompatibility means that data objects are not transferable from job to job.

- **Library routines and data objects -2017**

The new model of linking robust, well documented crystallographic routines via a scripting language has made software development much easier and more robust.

The new model of linking robust, well documented crystallographic routines via a scripting language has made installation and maintenance MUCH more complex!!

Coot architecture, showing the layers of functionality from the user interface down to some of the low-level libraries. The libraries are described in more detail in the main text (§§3.4, 3.5, 4.3 and ...



Tristan Croll Update:

Looks like I have working **PyPI-acceptable Linux** builds for all current major **Python versions (2.7, 3.4, 3.5, 3.6 - 3.3 apparently doesn't support Numpy, so it's out)**, and a reasonably portable build system to rebuild them on command. It's a bit brute force right now - it rebuilds everything from scratch for each Python version. Will try to clean it up and document it properly over the next few days - it's a bit of a steep learning curve. The last problem I spent this afternoon banging my head against is that the typemapping from `std::string` to `clipper::String` that Stuart worked out only seems to work for **Python 2.7 - in Python 3.x** it crashes (all the way back to the shell) with the error:

```
terminate called after throwing an instance of 'std::logic_error'
```

```
  what(): basic_string::_S_construct null not valid
```

... the moment I try to run, for example, `CIFfile.open_read(filename)` - so a little more work to do there. For now I've commented those out and reverted to the standard built-in typemap, which seems to work in most cases (and doesn't crash as drastically where it doesn't - to be honest I can't remember where it was crashing)

Cheers Tristan

Developing User support 1

1997? GUI - scripting in TclTk – (*now obsolete..*)

Prepared scripts with on-screen guidance

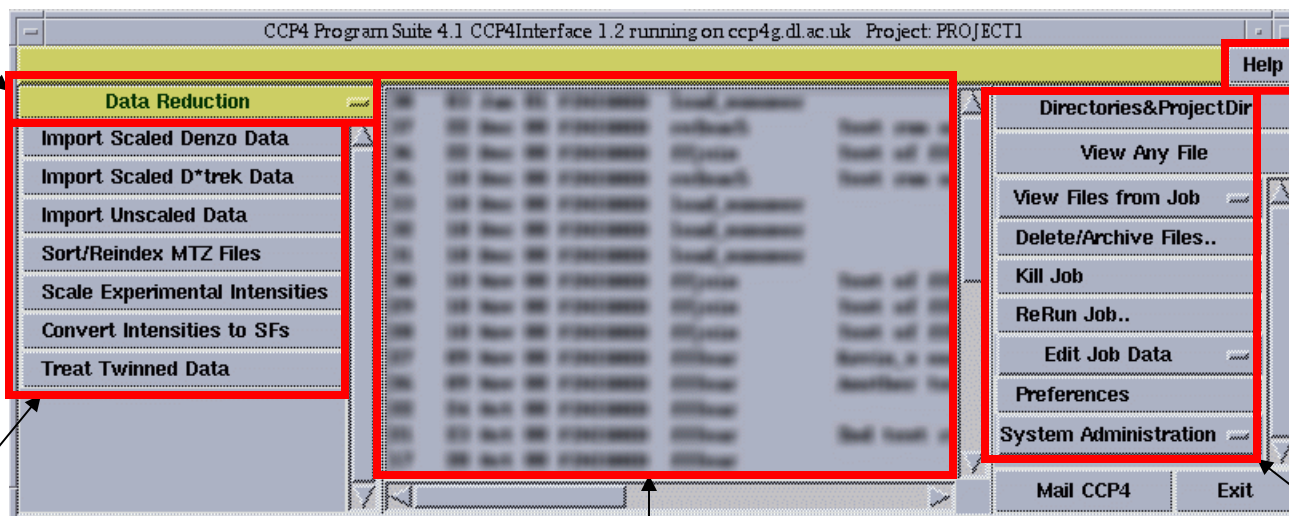
Reported results graphically..

Kept a record of what was done

Main CCP4i Window

Modules

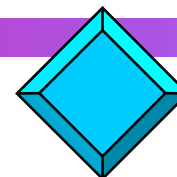
On-line help



Tasks

Job Database

Tools & Utilities



CCP4

User support 2

Visualisation now essential

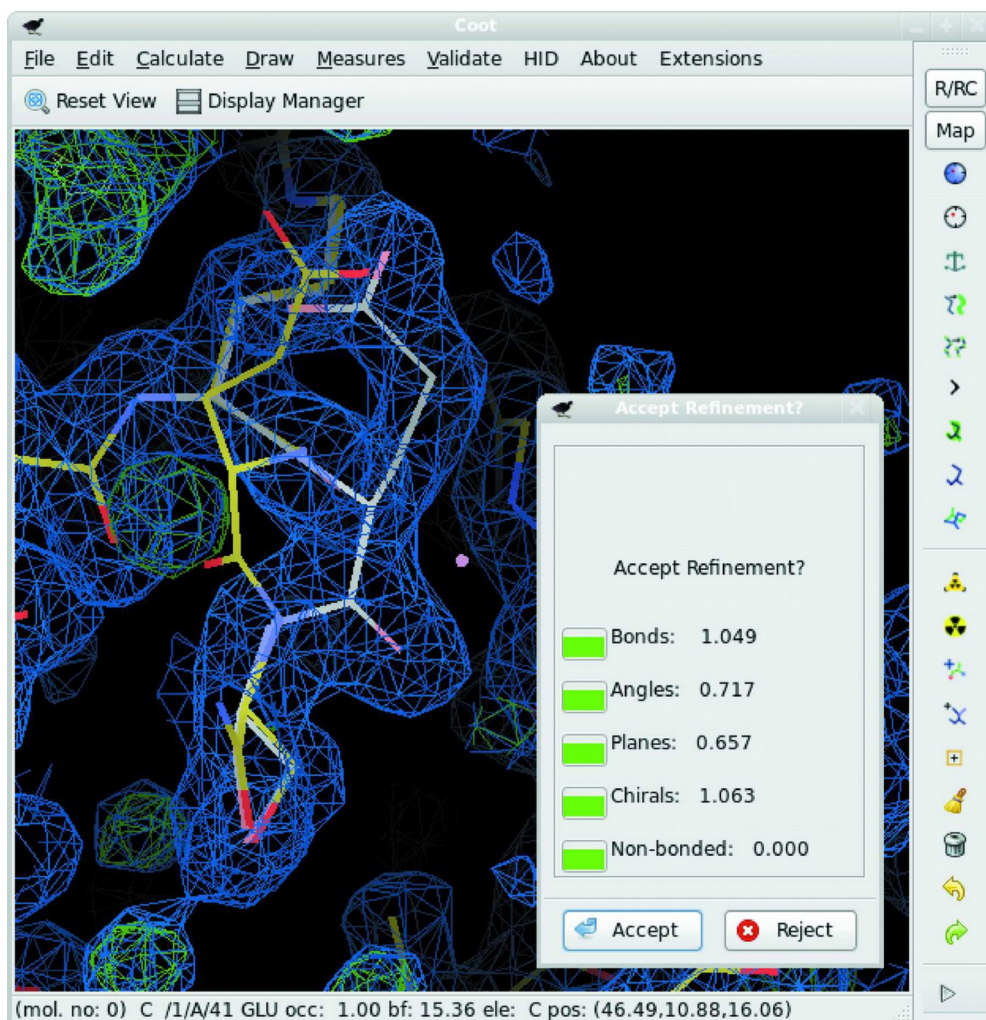
For Model building

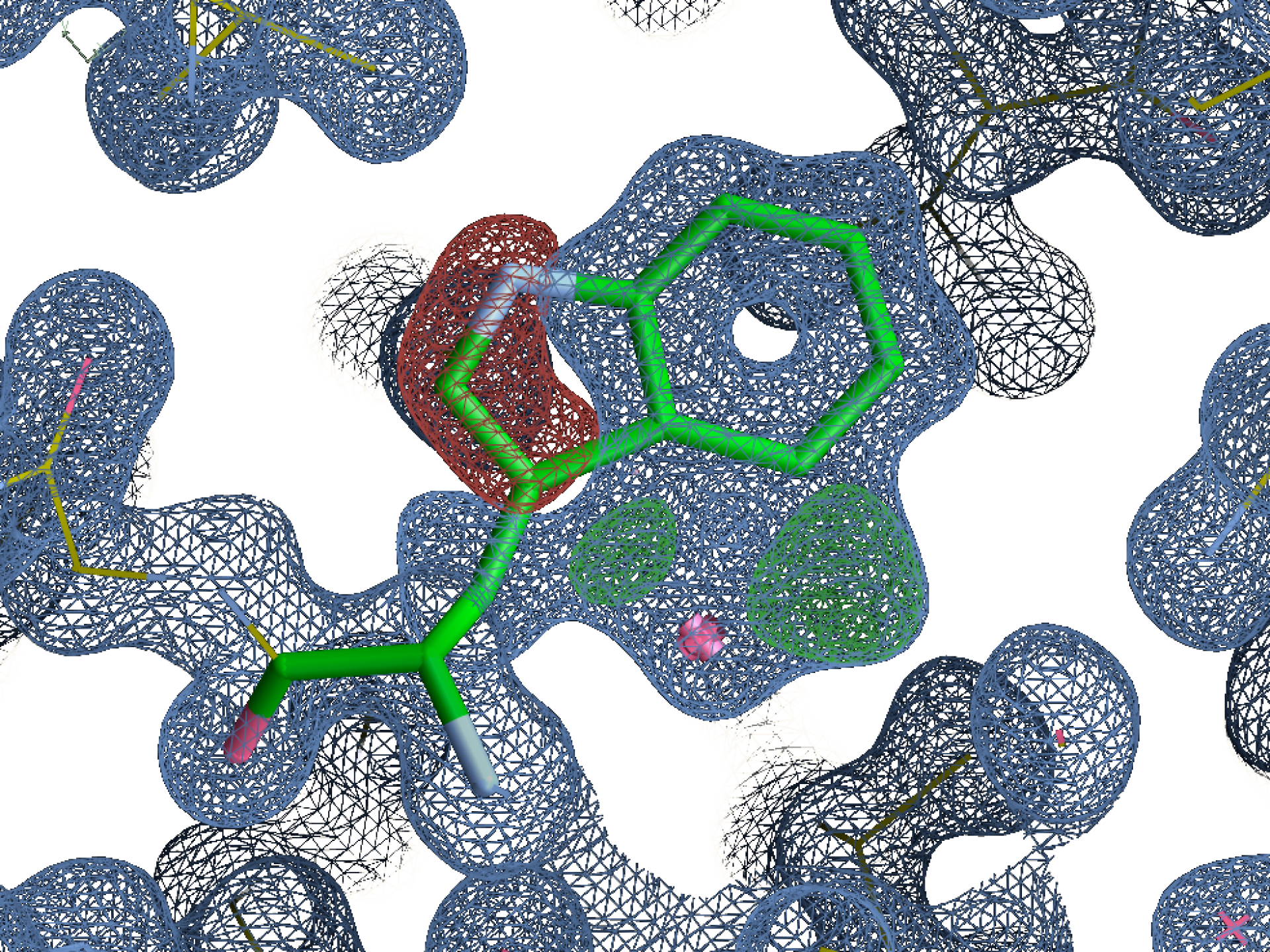
For Validation

For Structure analysis

Bioinformatics

COOT: Real-space refinement of a mispositioned residue. The coloured bonds show the original structure. The white bonds show the refined atoms after dragging and refinement. The coloured boxes in the pop-up ...





User support 3

Visualisation now essential

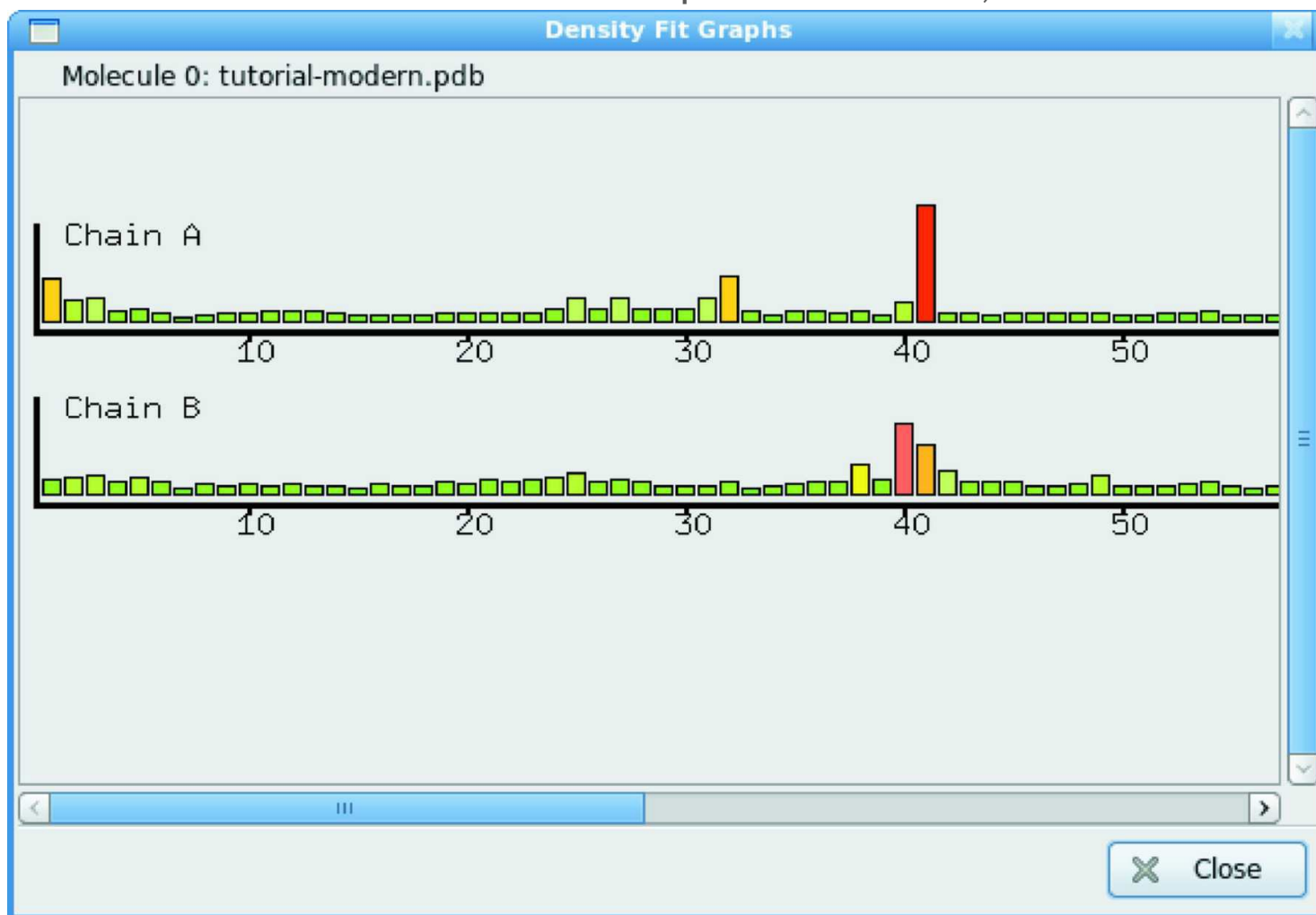
For Model building

For Validation

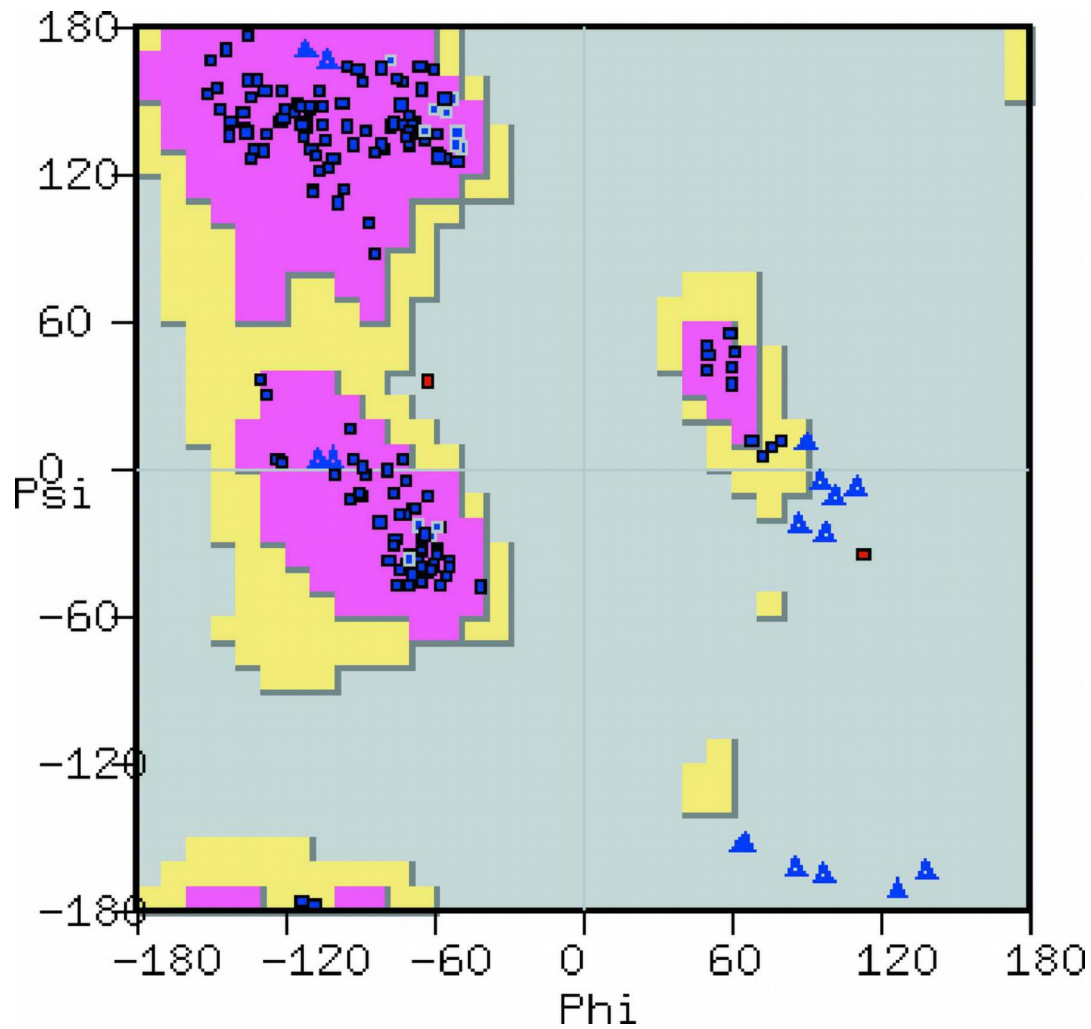
For Structure analysis

Bioinformatics

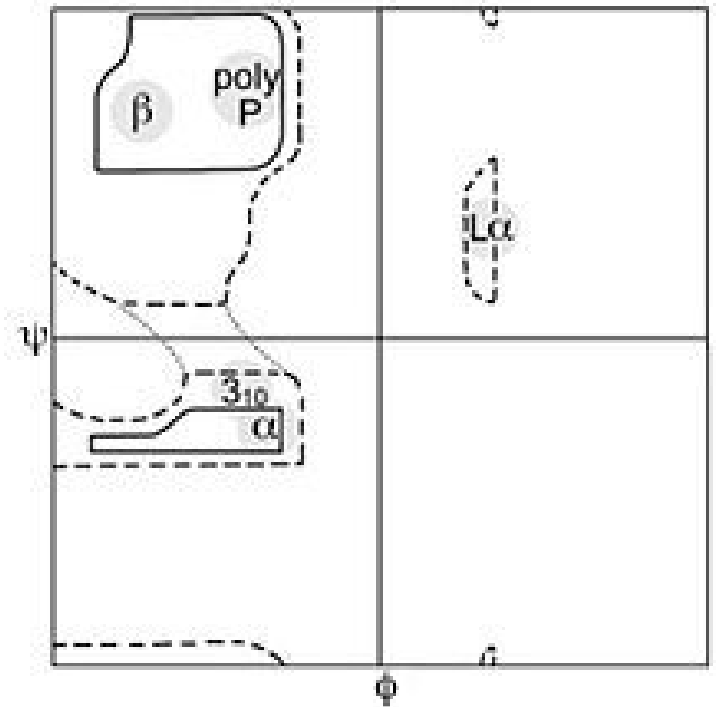
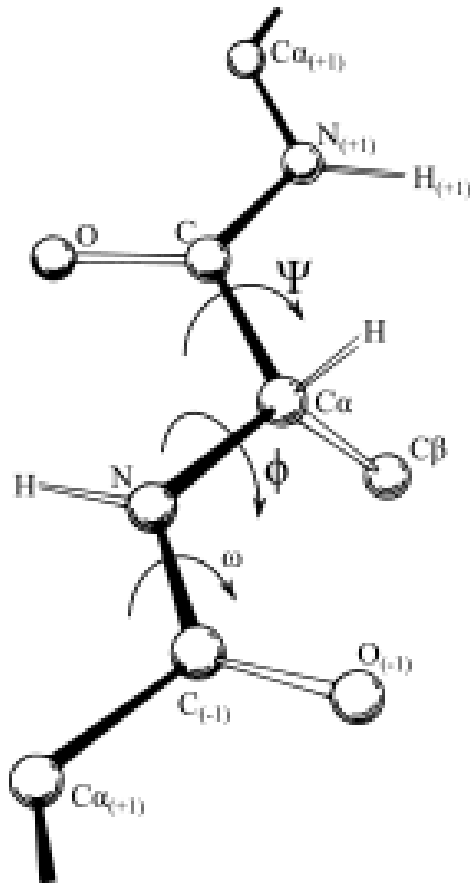
COOT: A typical validation graph. Bars represent individual residues in a chain, with an indication of quality for the residue being given by both the size and colour of the bar. The plot is interactive, ...



COOT: Screenshot of a classical Ramachandran plot showing all residues, with the axes defining the ϕ and ψ angles (angles in degrees). Preferred regions are coloured in pink, allowed regions in yellow and the ...

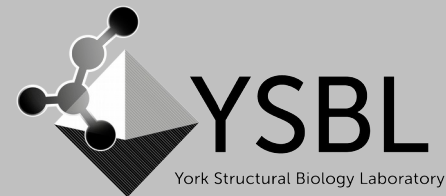


Ramachandran Figures 1963
But drawn by hand!





UNIVERSITY
of York



Privateer

More validation, re-
refinement and

analysis of
carbohydrate
structures

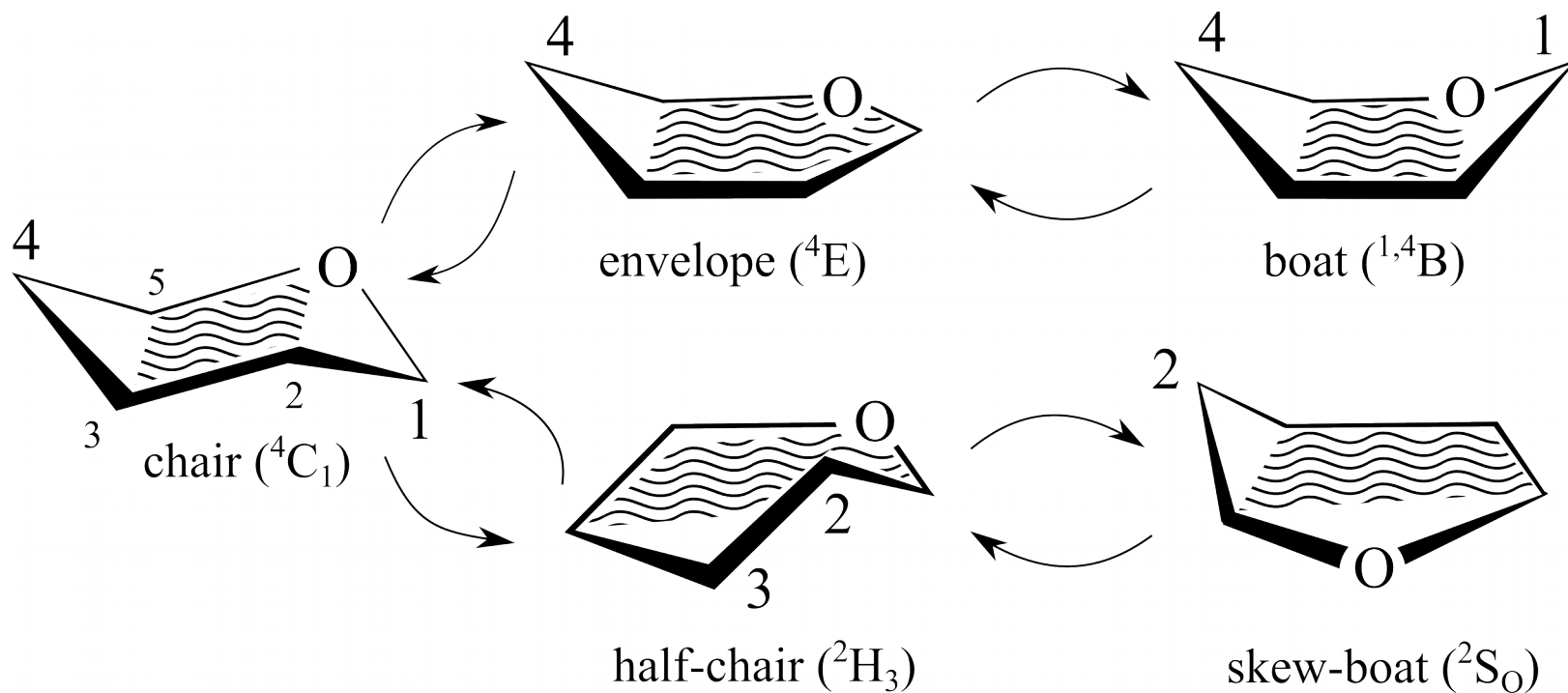
Jon Agirre

Carbohydrates

More challenging than building protein
Frequently ignored by structural biologists
Historically unsupported by software developers
"Cinderella molecules"

Enough is enough. Let us change that!

Conformations



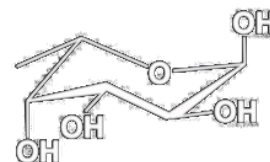
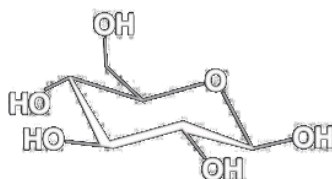
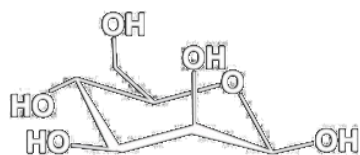
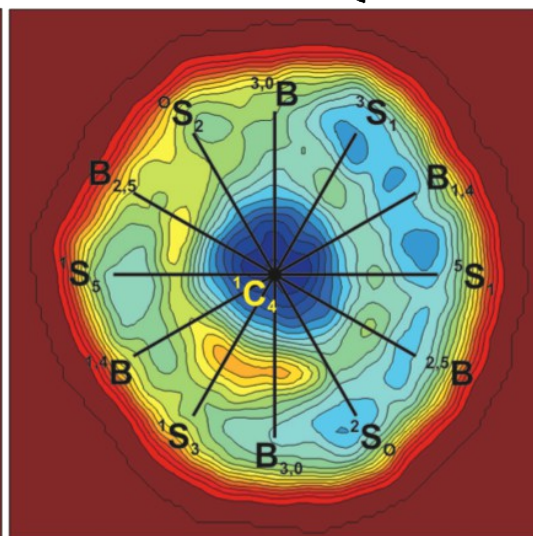
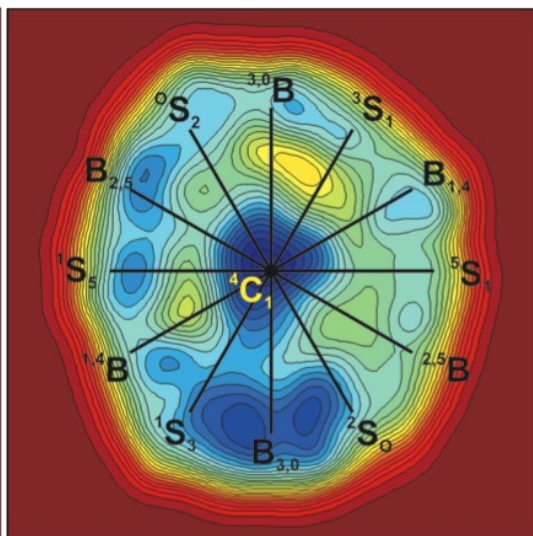
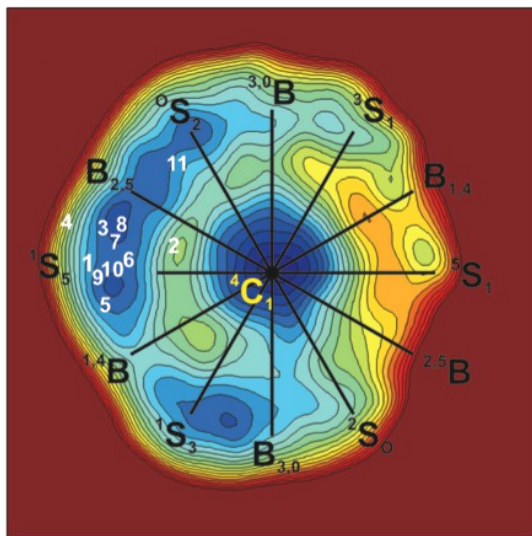
Wavy lines = atoms are roughly coplanar

Conformational preferences

0.0



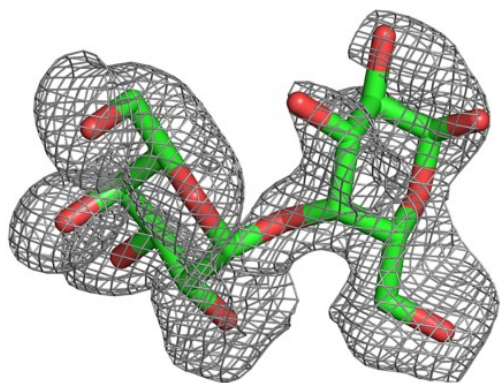
20.0 (kcal/mol)



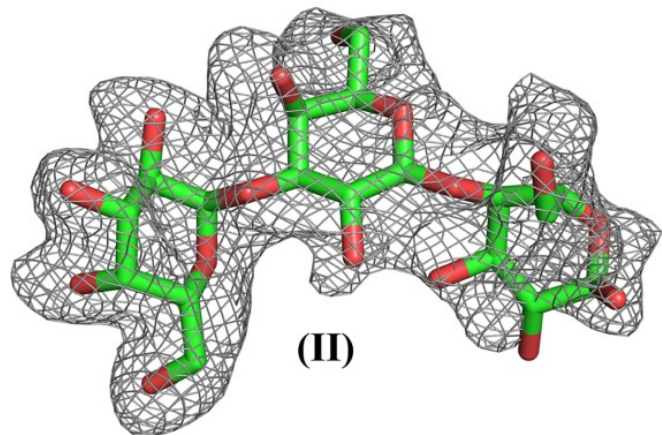
beta-D-mannopyranose beta-D-glucopyranose alpha-L-fucopyranose

Ardevol, Biarnés, Planas and Rovira. JACS 132(45), 2010

Always show
positive omit
density- sjs1



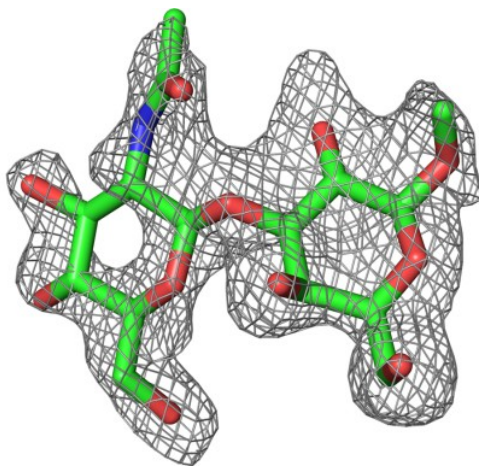
(I)



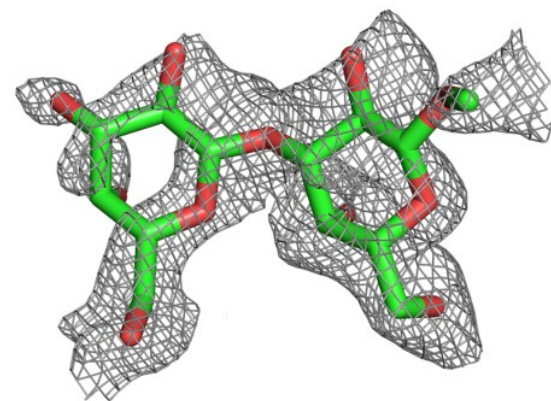
(II)

Abhinav, Sharma,
Surolia & Vijayan,
2016, IUBMB Life
68(12):971-979

Abhinav, Sharma,
Surolia & Vijayan,
2017, IUBMB Life
69(2):72-78



(I)



(II)

User support 4

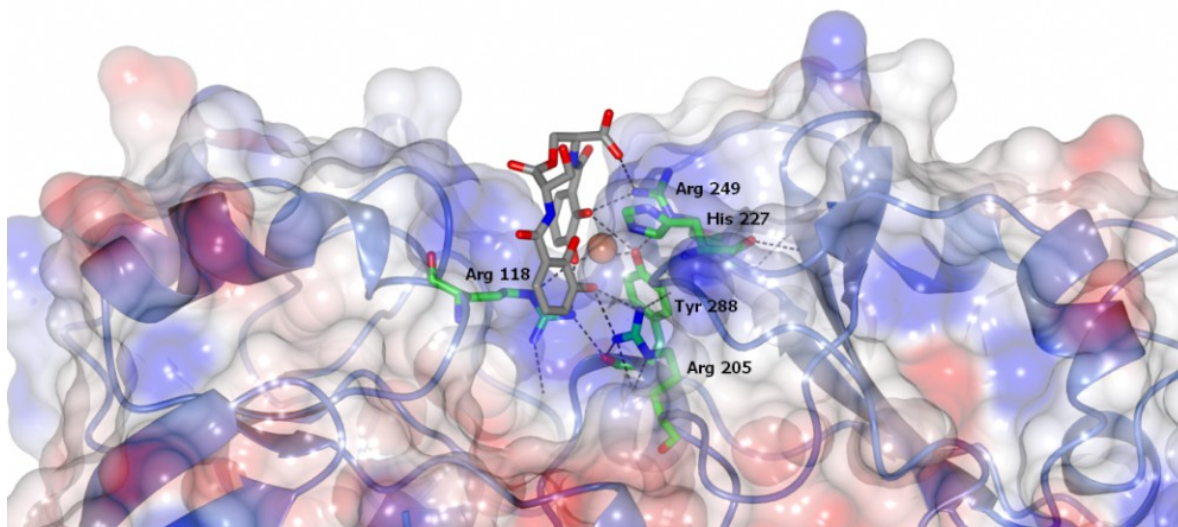
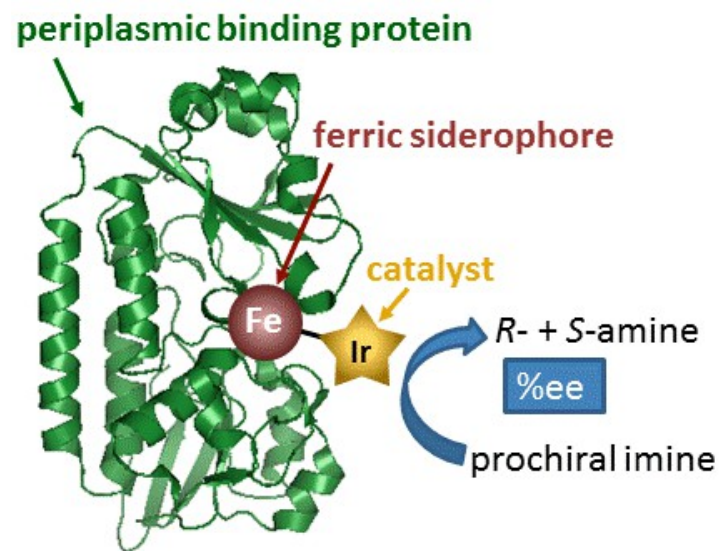
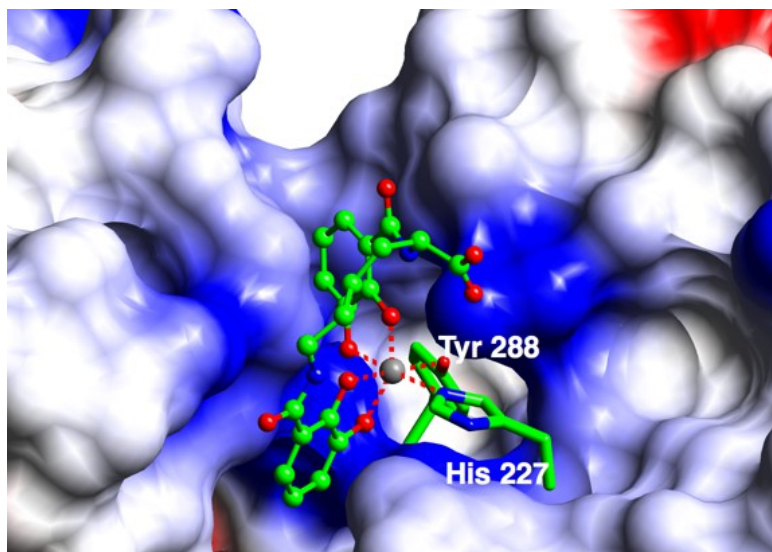
Visualisation now essential

For Model building

For Validation

For Structure analysis and Bioinformatics

CCP4MG Figures



CCP4MG Documentation

Introduction - loading and viewing PDB or mmCIF files. Essential reading.

Surfaces - selecting and colouring and electrostatic surfaces

Superpose proteins - instamagic protein superposition

Creating publication quality images

Importing Model Analysis Data

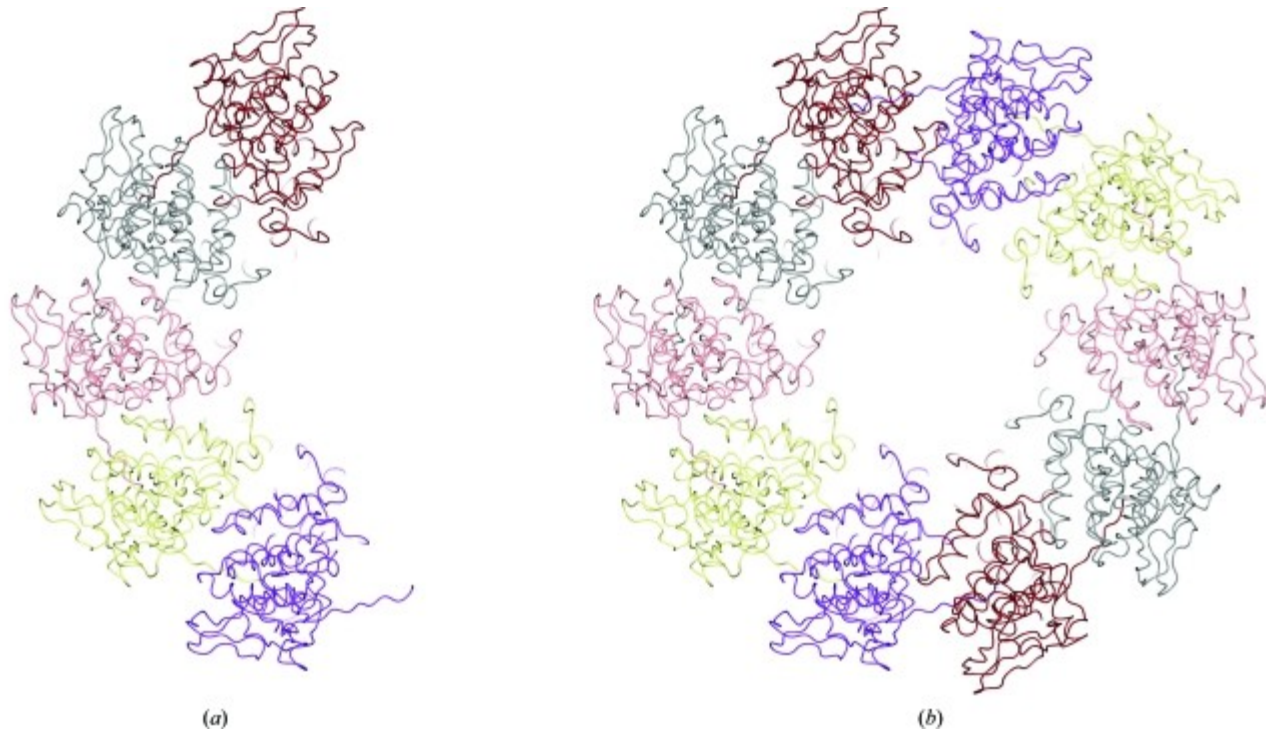
Electron density maps - better representation for images and packing diagrams

Adding Text and Extra Images

Ligand Binding Site Selection

Customised Atom Colouring

Investigating a structure - HBonds, Contacts, Solvent Accessibility, Geometry, Distances



Assembly prediction by **PISA**.

(a) The original model (PDB entry 3hhz; Green & Luo, 2009 ▶).

(b) The assembly predicted by PISA.

Why Automation?

- **Speed of data acquisition and computing.**
-
- **There is a move towards using MX as the new spectroscopy – a handy reliable(?) technique.**
-
- **Many user errors can be avoided by better self-correcting methods .**
-
- **GUIs attempts to do some of this**
-
- **But automation requires design by good crystallographers – how to train them?**
-
- **Project Management and book keeping become essential**

But more complexity – need mix of developers and comp.Sci. experts!

CCP4i2 - the new graphical user interface to the CCP4 program suite

Author(s) Lots!!: *Liz Potterton, Jon Agirre, Charles Ballard, Kevin Cowtan, Eleanor Dodson, Phil Evans, Huw Jenkins, Ronan Keegan, Eugene Krissinel, Kyle Stevenson, Andrey Lebedev, Stuart McNicholas, Rob Nicholls, Martin Noble*, Navraj Pannu, Christian Roth, George Sheldrick, Pavol Skubak, Johan Turkenburg, Villi Uski, Frank von Delft, David Waterman, Keith Wilson, Martyn Winn and Marcin Wojdyr*

Abstract: The CCP4 software suite for macromolecular structure determination by X-ray crystallography groups together many programs and libraries that, by means of well-established conventions, interoperate effectively without adhering to strict design guidelines. Because of this inherent flexibility, users are often presented with diverse, even divergent, choices for solving every type of problem. Recently, CCP4 introduced CCP4i2: a modern graphical interface designed to help structural biologists to navigate the process of structure determination, with emphasis on pipelining and streamlined presentation of results. In addition, CCP4i2 provides a framework for writing structure solution scripts that can be built up incrementally to create increasingly automatic procedures.

Need more people like Jon Agirre – PRIVATEER author:

CV:

Education

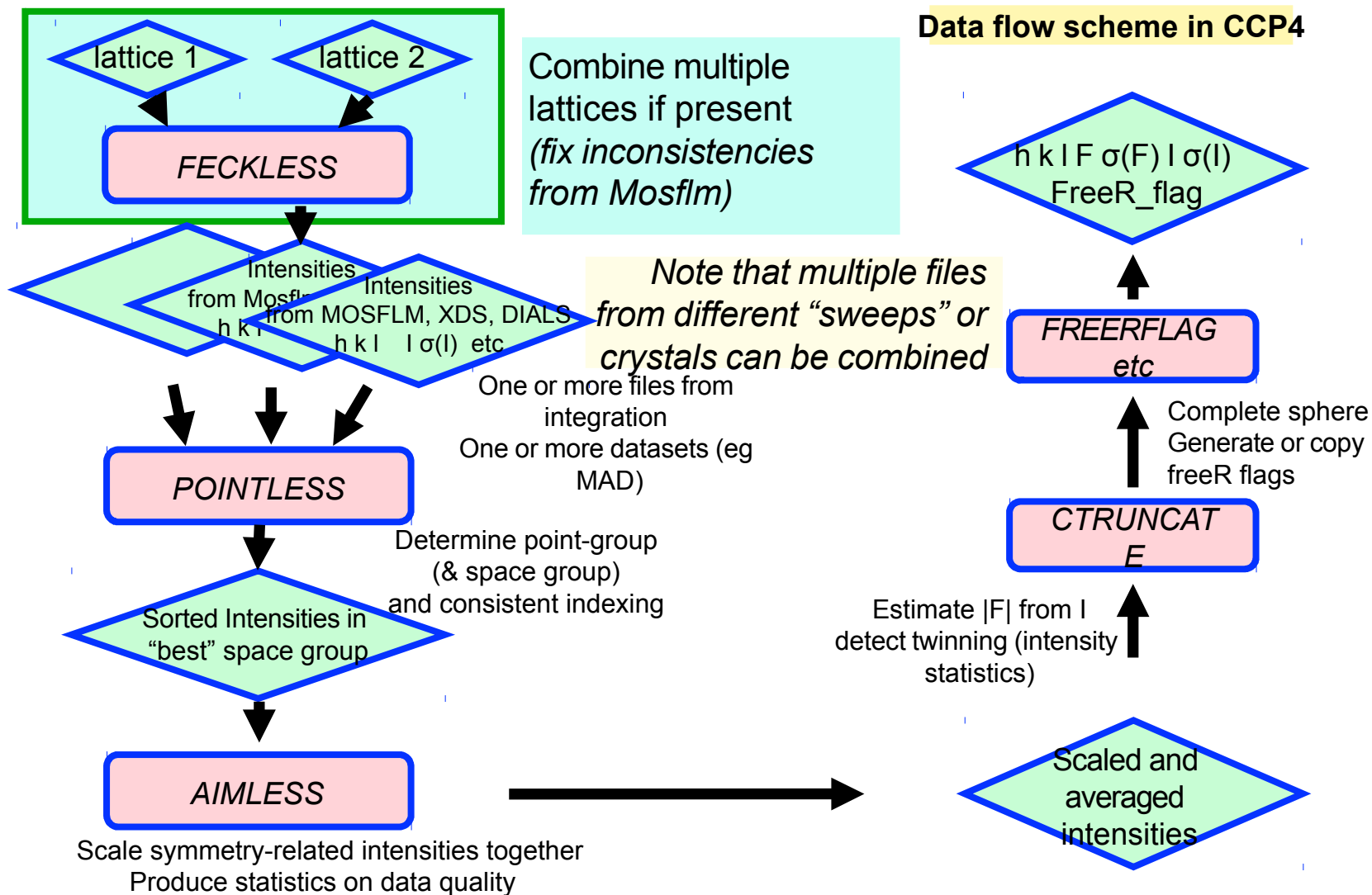
2004–2009 PhD in Biophysics, Unidad de Biofisica, Spanish Research Council (CSIC) and University of the Basque Country (UPV/EHU), Bilbao, Spain

Thesis: Structural basis of the stability, infectivity and disassembly process of Triatoma Virus.

1996–2002 Computer Engineer, University of the Basque Country, Spain

Specialised in parallel architectures. First-class honours.

Example of a good well documented informative pipeline for data processing provided in CCP4 GUI2.



Alternative indexing

If the true point group is lower symmetry than the lattice group, alternative valid but non-equivalent indexing schemes are possible, related by symmetry operators present in lattice group but not in point group (*note that these are also the cases where merohedral twinning is possible*)

eg if in space group $P3$ (or $P3_1$) there are 4 different schemes
(h,k,l) or (-h,-k,l) or (k,h,-l) or (-k,-h,-l)

For the first crystal, you can choose any scheme
For subsequent crystals, the autoindexing will randomly choose one setting, and we need to make it consistent: **POINTLESS** will do this for you by comparing the unmerged test data to a reference dataset (merged or unmerged, or coordinates)
Note that the space group from the reference will be assumed to be correct

The screenshot shows the POINTLESS software interface with several key settings highlighted in red boxes:

- File path:** /Users/pre/Projects/Xtal/l2demo/Data/amphtest.mtz
- Crystal name:** AmpNT
- Dataset name:** IP6_2
- Batches in file:** 1001 - 1238
- Resolution range (Å):** [] to [] Maximum resolution in files 3.50Å
- Options for symmetry determination:** Match index to reference data
- Reference data are:** Reflection list
- Reflections:** amph1_P3121_scala1: se1_peak imported by job 36
- Free R set:** ..is not used

Viewing the output statistics (job report from ccp4i2)

1. key summary

▼ Key summary

Selecting space group P 21 21 21
as there is a single space group with the highest score

Solution probability: 0.853, Confidence 0.835 (high resolution limit for symmetry testing 3.252A)

NOTE: the final selected symmetry and cell have alternative indexing schemes, but no reference data has been given
Possible alternative indexing operators (with cell differences in A): [h,k,l] (0.00), [-h,l,k] (0.71), [-k,h,l] (0.73), [l,h,k] (1.25), [k,l,h] (1.25), [l,k,-h] (1.44)
If you already have a matching dataset, you should choose it as a reference set to get consistent indexing

Key statistics for Dataset: I2demo/Brp/pk

Resolution of input data: 2.79A, resolution estimate 2.87A
Rmeas: overall 0.151, inner bin 0.069
In outer bin: Mean(I/sdI) 0.8 CC(1/2) 0.266
Anomalous CC(1/2) in inner bin 0.753
Significant anomalous signal extends to a resolution of 3.74A (above CCanom threshold 0.15)

Warning: Possible twinning, twin fraction estimates from Britton plot 0.20, from H-test 0.23

No evidence of possible translational non-crystallographic symmetry

Some anisotropy detected. This may have an effect on statistics.

Warning: Completeness test shows some issues.

No ice rings found.

Warnings:
red, bad;
orange, maybe OK;
green, OK

Viewing the output statistics (job report)

“Table 1”

2. main summary

Space group determination

scores for individual symmetry elements may detect pseudo-symmetry ...

... or suggest twinning

Space group determination
 WARNING: the L-test suggests that the data may be twinned, so the indicated Laue symmetry may be too high
 Rough estimated twin fraction: 0.096

Selecting space group P 21 21 21
 as there is a single space group with the highest score

Solution type: space group

| | |
|--------------------------------|------------|
| Group name | P 21 21 21 |
| Reindex | [h,k,l] |
| Space group confidence | 0.835 |
| Laue group confidence | 0.938 |
| Laue group probability | 0.948 |
| Systematic absence probability | 0.900 |

Scores for each symmetry element

Lattice group name P 4 3 2

Reindex operator from input to lattice: [h,k,l]

| Likelihood | CC | R | | Symmetry |
|------------|------|-------|-----|-----------------------------|
| 0.913 | 0.88 | 0.097 | | identity |
| 0.901 | 0.87 | 0.109 | *** | 2-fold l (0 0 1) {-h,-k,l} |
| 0.917 | 0.88 | 0.090 | *** | 2-fold k (0 1 0) {-h,k,-l} |
| 0.913 | 0.88 | 0.103 | *** | 2-fold h (1 0 0) {h,-k,-l} |
| 0.214 | 0.56 | 0.191 | | 2-fold (1-1 0) {-k,-h,-l} |
| 0.051 | 0.06 | 0.588 | | 2-fold (0 1-1) {-h,-l,-k} |
| 0.052 | 0.13 | 0.494 | | 2-fold (1 0-1) {-l,-k,-h} |
| 0.223 | 0.57 | 0.193 | | 2-fold (1 1 0) {k,h,-l} |
| 0.051 | 0.11 | 0.512 | | 2-fold (1 0 1) {l,-k,h} |
| 0.051 | 0.06 | 0.562 | | 2-fold (0 1 1) {-h,l,k} |
| 0.053 | 0.04 | 0.720 | | 3-fold (1-1-1) {-k,l,-h} |
| 0.054 | 0.04 | 0.712 | | 3-fold (1 1-1) {-l,h,-k} |
| 0.053 | 0.04 | 0.701 | | 3-fold (1-1 1) {l,-h,-k} |
| 0.053 | 0.04 | 0.704 | | 3-fold (1 1 1) {k,l,h} |
| 0.203 | 0.55 | 0.198 | | 4-fold l (0 0 1) {-k,h,l} |
| 0.050 | 0.10 | 0.509 | | 4-fold k (0 1 0) {l,k,-h} |
| 0.052 | 0.05 | 0.565 | | 4-fold h (1 0 0) {h,l,-k} |

Data internal consistency statistics

Summary of merging statistics for dataset I2demo/Brp/pk

| | Overall | Inner | Outer |
|-------------------------------|---------|-------|--------|
| Low resolution limit | 57.82 | 57.82 | 2.94 |
| High resolution limit | 2.79 | 8.81 | 2.79 |
| Rmerge(within I+ /I-) | 0.140 | 0.064 | 2.085 |
| Rmerge(all I+ and I-) | 0.154 | 0.076 | 2.218 |
| Rmeas (within I+ /I-) | 0.151 | 0.069 | 2.336 |
| Rmeas (all I+ & I-) | 0.161 | 0.080 | 2.352 |
| Rpim (within I+ /I-) | 0.057 | 0.026 | 1.013 |
| Rpim (all I+ & I-) | 0.045 | 0.023 | 0.754 |
| Rmerge in top intensity bin | 0.063 | | |
| Number of observations | 178940 | 5691 | 18047 |
| Number unique | 14045 | 501 | 1955 |
| Mean(I)/sd(I) | 9.7 | 33.5 | 0.8 |
| Half-set correlation CC(1/2) | 0.998 | 0.997 | 0.266 |
| Completeness % | 99.5 | 97.6 | 96.9 |
| Multiplicity | 12.7 | 11.4 | 9.2 |
| Anomalous completeness % | 98.5 | 98.1 | 90.8 |
| Anomalous multiplicity | 6.4 | 6.9 | 4.7 |
| DelAnom CC(1/2) | 0.664 | 0.753 | -0.026 |
| Mid-Slope of Anom Probability | 1.020 | | |

Download as CSV file



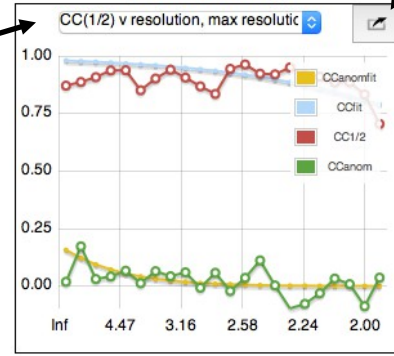
... to be a significant anomalous signal so anomalous flag was switched ON

Viewing the output statistics (job report)

3. The most important graphs

pull-down to change graph

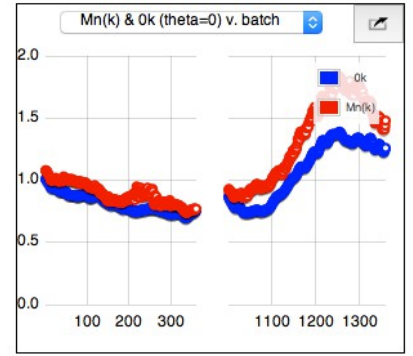
Analysis as a function of resolution
Plot of $CC(1/2)$ vs. resolution may indicate a suitable resolution cutoff, and indicate presence of an anomalous signal (but check anisotropy)



Analyses by resolution

pop out separate graph viewer

Analyses against Batch may show radiation damage, and which parts of the data should be removed (but consider completeness)

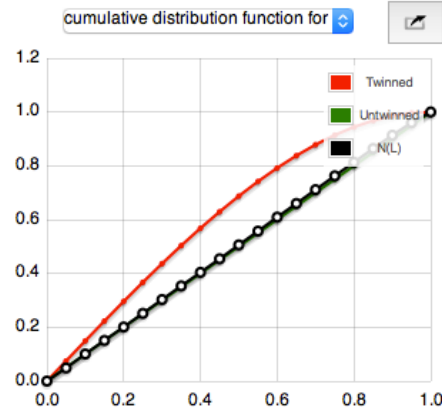


Analyses by batch

4. more details in folders, closed by default

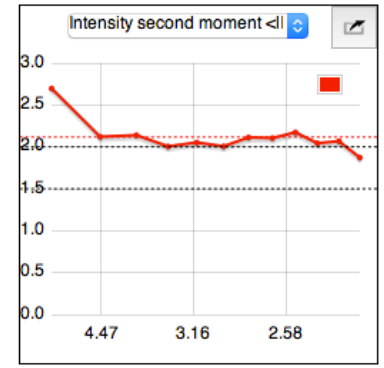
- ▶ Details of space group determination
- ▶ Other merging graphs
- ▶ Details of merging
- ▶ Intensity statistics: twinning tNCS etc

L-test for twinning



Analyses for twinning

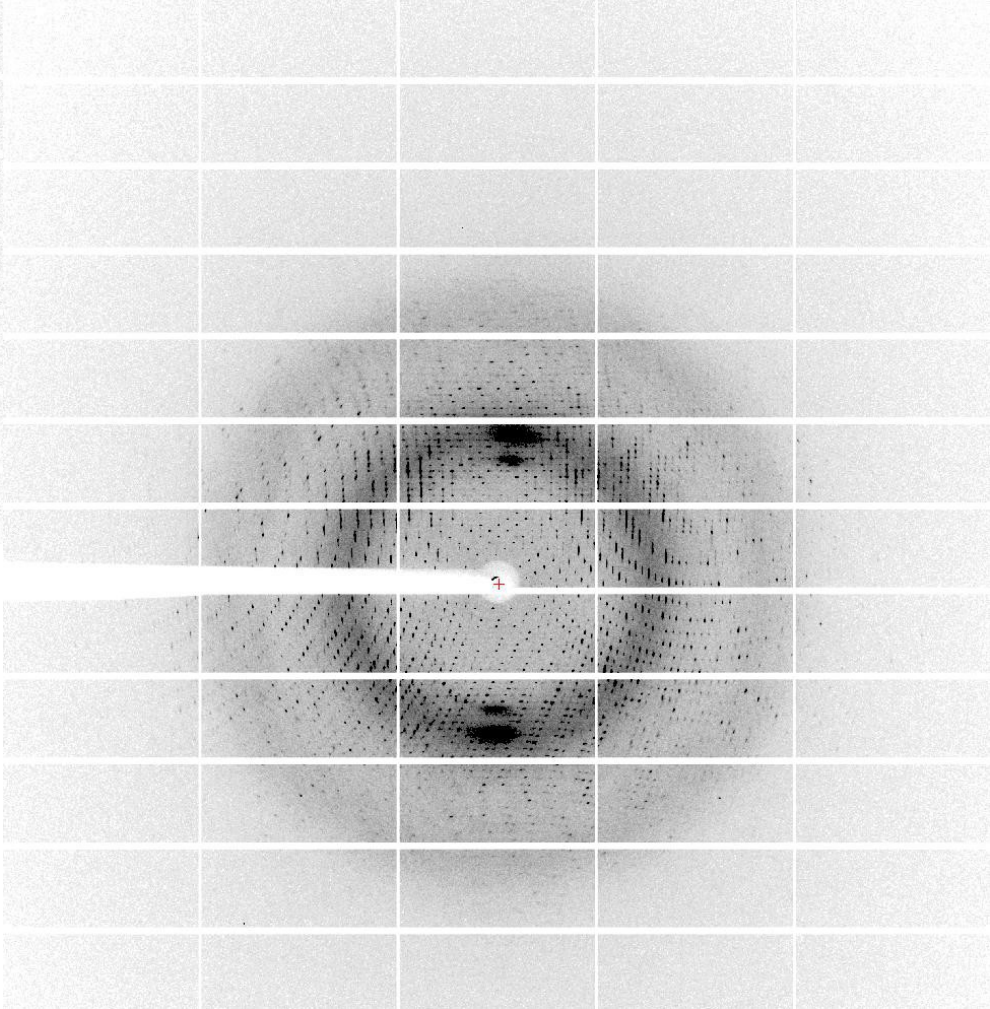
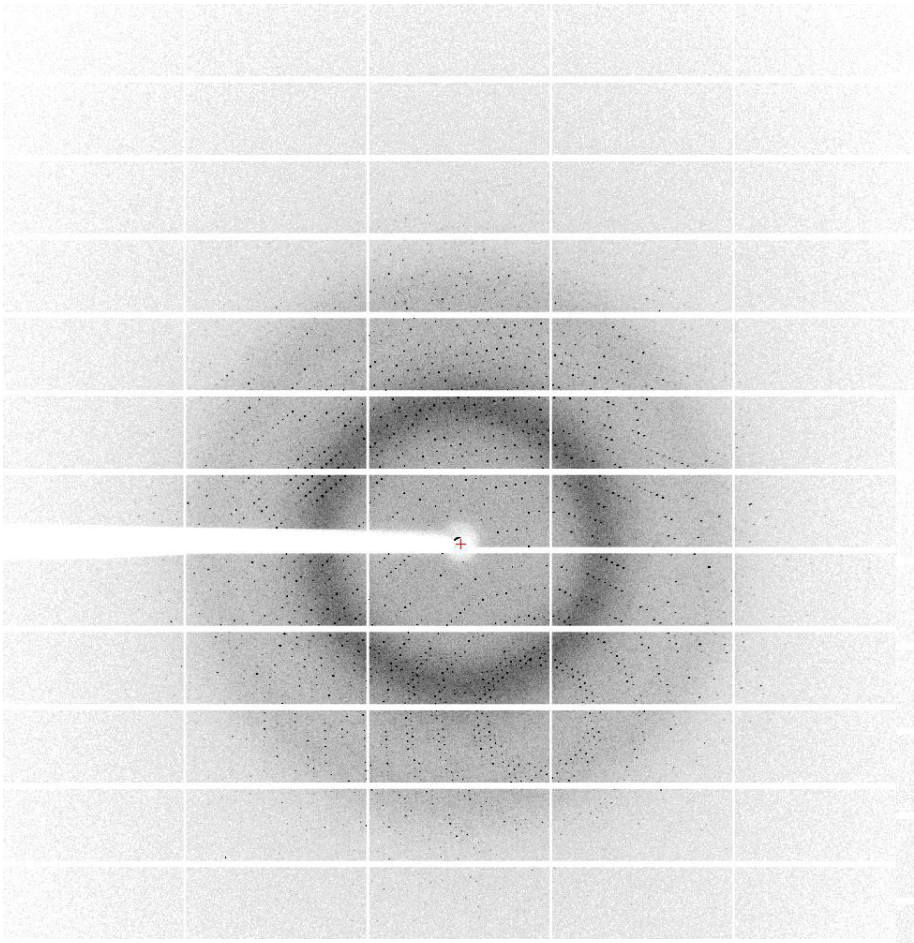
Acentric intensity moments



Values for these data, and for ideal data (untwinned or twinned)

| Operator | Value | Untwinned | Perfect twin |
|---|--------|-----------|--------------|
| $\langle I^2 \rangle / \langle I \rangle^2$ | 2.117 | 2.000 | 1.500 |
| $\langle I^3 \rangle / \langle I \rangle^3$ | 7.082 | 6.000 | 3.000 |
| $\langle I^4 \rangle / \langle I \rangle^4$ | 32.680 | 24.000 | 7.500 |

But a warning -
You still may need to look at
diffraction images!



Why Automation? Consider Speed

- **We now can test many many unlikely hypotheses quickly and with luck hit on a likely one.**
-
- **This requires very robust target functions, book keeping and reporting.**
-
- **Validation after automated structure solution becomes much more important..**
-
-

Cooperation – is it possible? desirable?

- **Advantages**
- **Can speed up developments if library routines are well documented and accessible :**
 -
- **Shared efforts for maintenance and distribution extend the code lifetime. Bugs identified & fixed more frequently**
 - **Common style helps users**
 - **Organising crystallographic data is not easy and requirements change**
 -

Cooperation – is it possible? desirable?

- **Disadvantages**
- **Time consuming – consultation essential.**
- **Computer nerds not always good at this!**
-
- **Needs commitment by developers of algorithms and libraries– often faster to make a quick cludgy fix than learn to use library; new routines may need to be added to libraries**
-
- **Harder to get credit for individual contributions**

- Friendly Discussion Amongst Developers??



Incomplete Acknowledgements

All the people I have worked with.

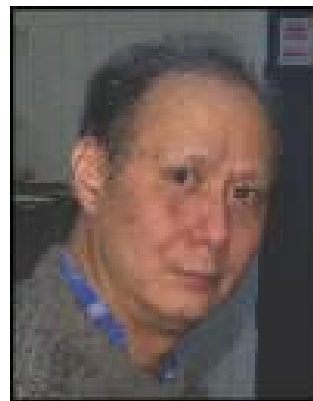
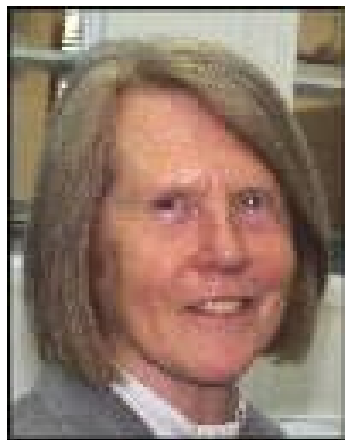
(Guy, Vijayan, Bakers, Phil Evans, Keith Wilson, Neil Isaacs,)

Pictures from York laboratory, family , friends, Hieronymus Bosch.

• A few algorithm contributors



Garib Murshudov. Alexei Vaguine. Randy Read, Paul Emsley, Jorge Navaza



Eugene Krissinel

me

Yao

Kevin Cowtan

Phil Evans ...

CCP4 Employees 2005



Happy Mapping



Plan

Macromolecular crystallography: from isolated programs
to full package

Algorithms – Fortran – simple.

CCP4 WG2 – Talking shop

GUI1 – talking shop still

Now Analysis –

Graphics complicated – Scheme? Python? QT5?

Reporting from Algorithms and providing information flow

Complicated – XML?? JSON??? PHIL files???

GUI2 – more sophisticated – farther from algorithms..

See example of AUTOMATION –

POINTLESS? AIMLESS? CTRUNCATE? FREERFLAG

Others – CRANK2 – exptl phasing

BUCCANEER? REFMAC? COOT

Dialogue between developers? Not so easy..

Remote working great but harder to nag..

Developers need MX + comp sci + bioinfor + chemistry

+ ..

OR a lot of good friends..