Refinement

DLS-CCP4 Data Collection and Structure Solution Workshop

December 1-9 2015

Diamond Light Source, Oxfordshire, UK

Oleg Kovalevskiy
okovalev@mrc-lmb.cam.ac.uk

Rob Nicholls
nicholls@mrc-lmb.cam.ac.uk

MRC Laboratory of Molecular Biology

CCP4

Contents

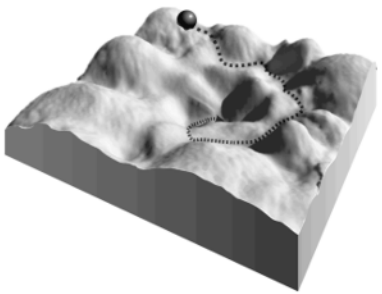
Purpose of Refinement

Crystallographic refinement has one major purpose:

<u>to fit atomic model into observed X-ray crystallographic data</u>

Model should agree with the observed data

Model must be chemically and structurally sensible



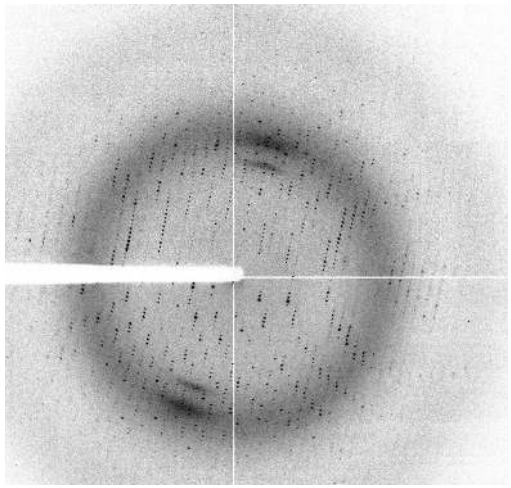
and one immediate most valuable consequence:

<u>to calculate best possible electron density map</u>

Allowing the atomic model to be visualised, criticised and analysed

Direct relation between model quality and phase quality (corresponds to the quality of the electron density maps)

Fourier Transform

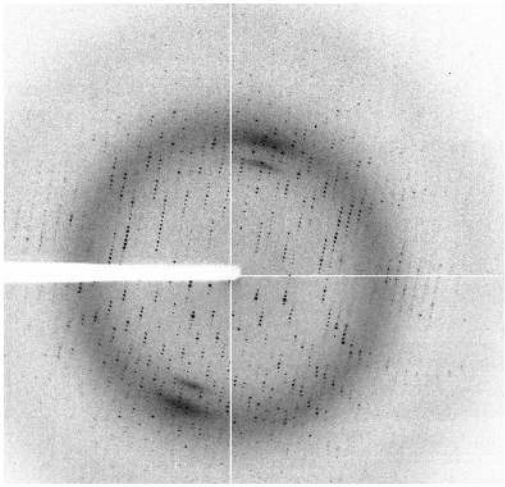


H, K, L	$\lvert F_{obs} \rvert$	ϕ
…		
5, 5, 5	348	-
5, 5, 6	392	-
5, 5, 7	157	-
5, 5, 8	312	-
…		

We have observed amplitudes: $\lvert F_{obs} \rvert$

But we don't have phases: φ

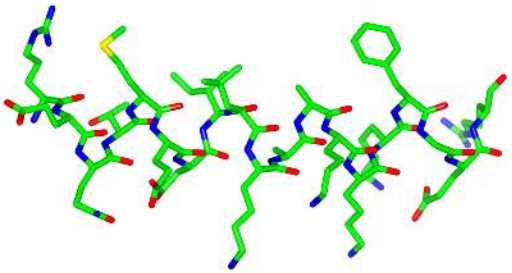
Fourier Transform



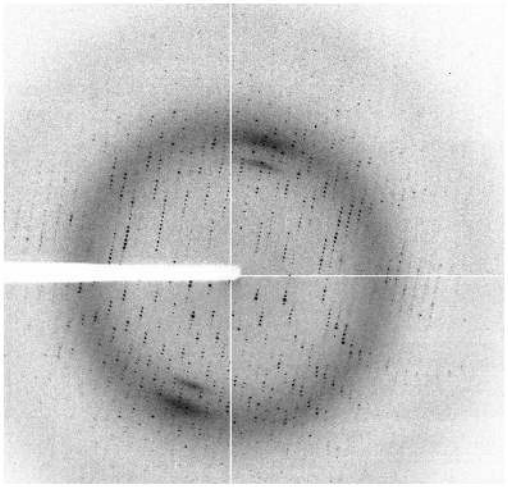
H, K, L	$\lvert F_{obs} \rvert$	ϕ
...		
5, 5, 5	348	-
5, 5, 6	392	-
5, 5, 7	157	-
5, 5, 8	312	-
...		

Suppose we have a starting model:

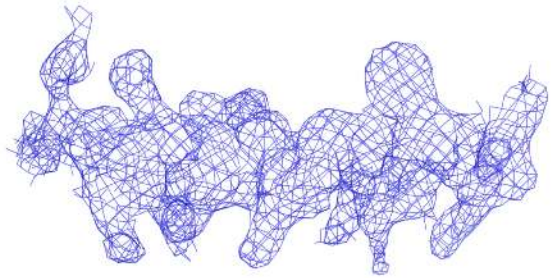
H, K, L	$\lvert F_{calc} \rvert$	ϕ_{calc}
...		
5, 5, 5	355	27°
5, 5, 6	387	8°
5, 5, 7	146	75°
5, 5, 8	340	31°
...		

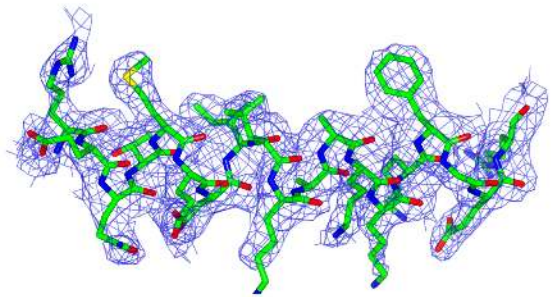


Model Refinement



H, K, L	$\lvert F_{obs} \rvert$	ϕ
...		
5, 5, 5	348	-
5, 5, 6	392	-
5, 5, 7	157	-
5, 5, 8	312	-
...		



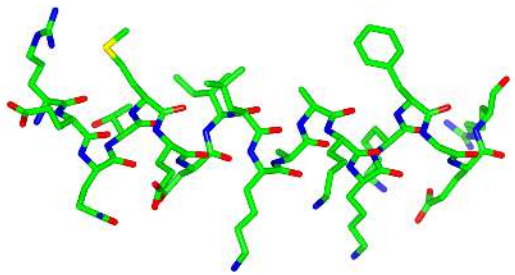


Idea:

Iteratively improve the model, optimising the agreement between $\lvert F_{obs} \rvert$ and $\lvert F_{calc} \rvert$

Purpose: improve phase estimates: φ_{calc}

H, K, L	$\lvert F_{calc} \rvert$	ϕ_{calc}
...		
5, 5, 5	355	27°
5, 5, 6	387	8°
5, 5, 7	146	75°
5, 5, 8	340	31°
...		



Model Refinement

Idea:

Iteratively improve the model to optimise the agreement between $|F_{obs}|$ and $|F_{calc}|$

*Note – we are **not** refining against a density map*

We are optimising the agreement between $|F_{obs}|$ and $|F_{calc}|$

How to assess correspondence between the model and experimental observations?

R-factor: $$R = \frac{\sum \left| |F_{obs}| - |F_{calc}| \right|}{\sum |F_{obs}|}$$

Model Refinement

Refinement essentially tries to minimise the R-factor

How do we know that the model is reliable?

What if we improve the amplitudes $|F_{calc}|$ but worsen the phases φ_{calc}?

Such overfitting can happen if there are too many parameters

How to validate?

- **R_{free}** – reserve a portion of data for cross-validation (usually 5%)
- **Chemical & structural validation** – ensure that the model is physically sensible
- **Inspect electron density map** – manual intervention

Map Calculation

Two types of maps:

- $2F_{obs}-F_{calc}$: *"standard" electron density* – represents crystal contents
- $F_{obs}-F_{calc}$: *difference density* – represents differences

Maps are calculated using phase estimates from the current model: φ_{calc}

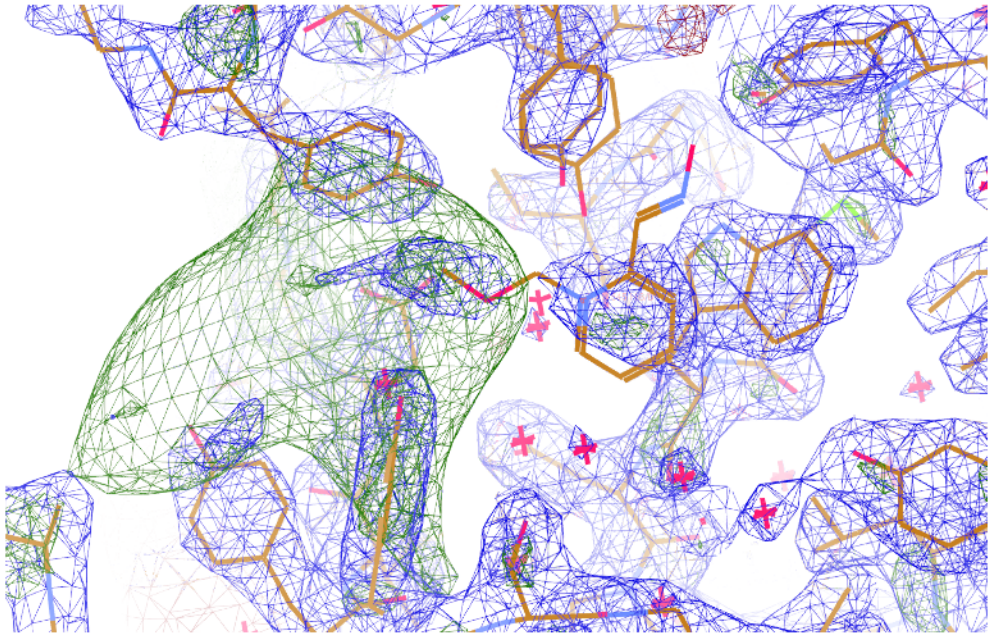
Model updated

↓

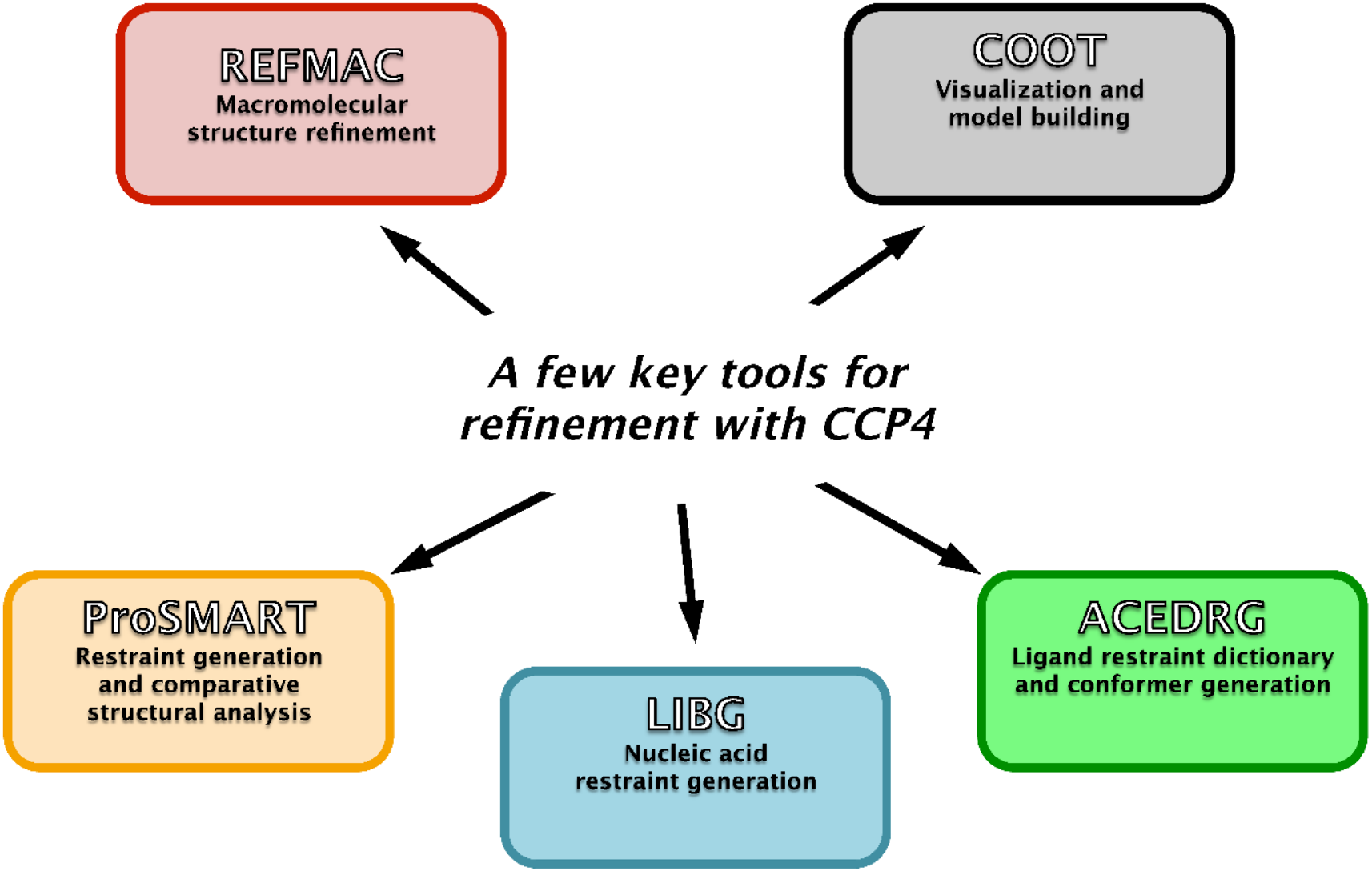
φ_{calc} updated

↓

Density changes

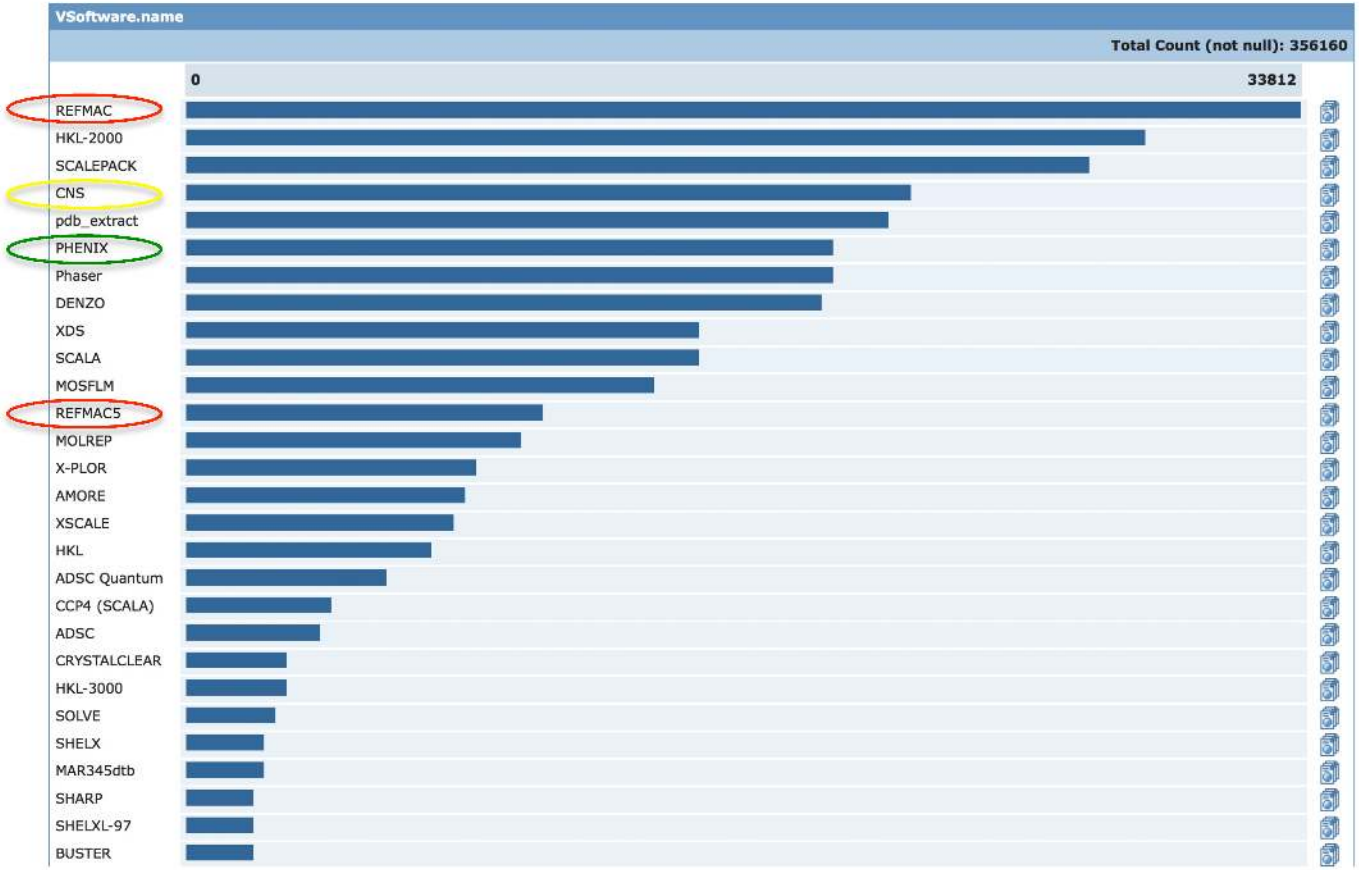


Note – contrast with real space refinement


REFMAC
Macromolecular structure refinement
COOT
Visualization and model building
A few key tools for refinement with CCP4
ProSMART
Restraint generation and comparative structural analysis
LIBG
Nucleic acid restraint generation
ACEDRG
Ligand restraint dictionary and conformer generation

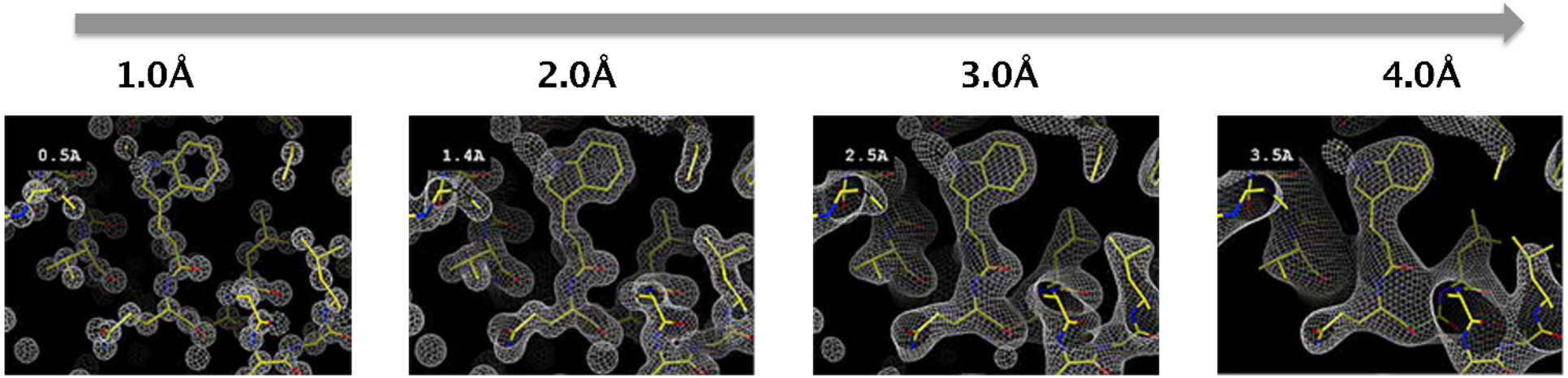
Available Refinement Programs

- SHELXL
- CNS
- REFMAC5
- TNT
- BUSTER/TNT
- Phenix.refine
- RESTRAINT
- MAIN
- MOPRO
- XD



Your Crystal Peculiarities

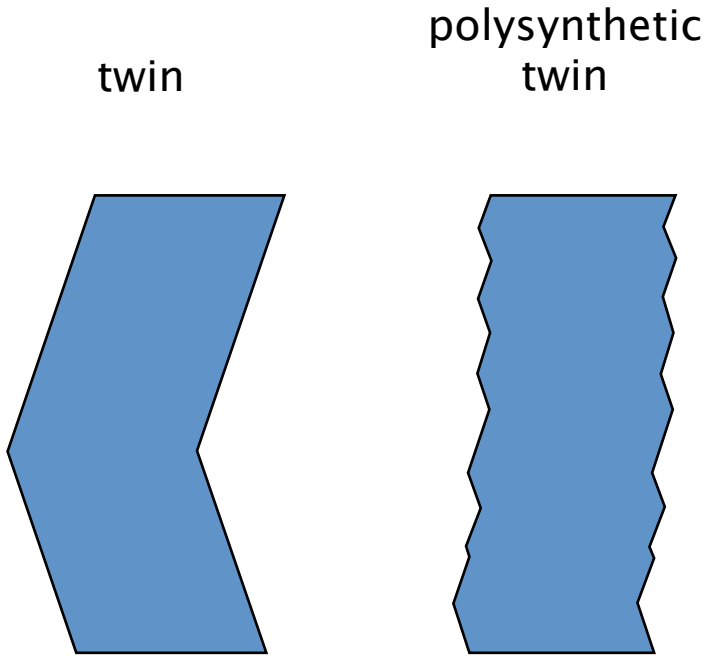
Refinement strategy will differ for different quality of original data and it can also exploit particular features of your crystal:



- Twinning
- Several similar subunits – NCS (at moderate and low resolution)

(There might be other peculiarities of your crystal like pseudo-translation, but they are not important at the refinement stage)

Twin Refinement



A single crystal can be cut out of the twin:	yes	no

Need to deal with polysynthetic OD-twin during refinement

Twin Refinement

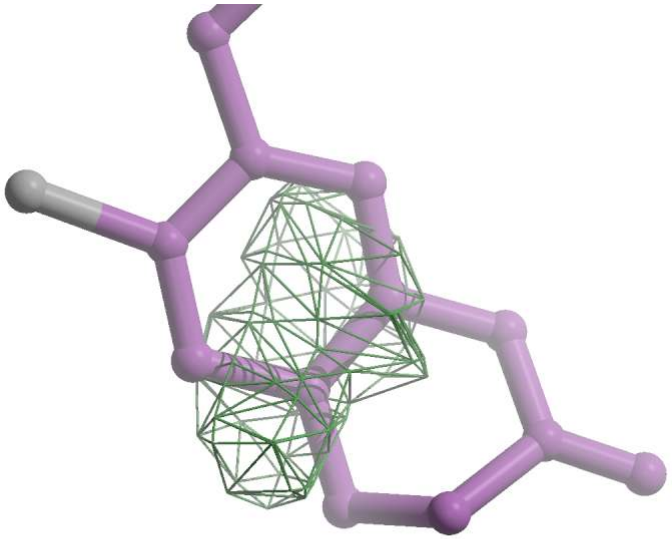
Twin refinement in REFMAC5 is automatic

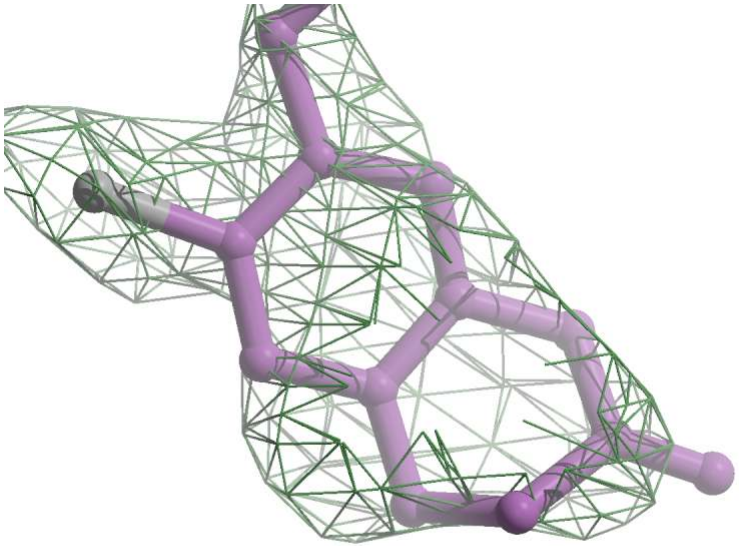
1. Identify potential twin operators
2. For each operator, calculate R_{merge} (R-factor comparing twin-related intensities)
3. If $R_{merge}>0.44$ remove this operator
4. Refine twin fractions
5. Keep only sufficiently large domains (default 7%)

Note - intensities can be used

Twin Refinement

Example: Where's the density for my ligand (2.15A)?





F_o-F_c (3σ)

R/R_{free} = 25.5/26.9%

After initial rigid body and restrained refinement

R/R_{free} = 15.9/16.3%

Re-refine with twin on (twin fractions: 0.6/0.4)

Borrowed from Ben Bax, GSK, Stevenage, UK

Model Refinement

We now know:

- What sort of data we have
- Particular features of your crystal (twinning, several subunits in the A.U., etc)
- How to get phase estimates from the current model
- How to calculate electron density maps

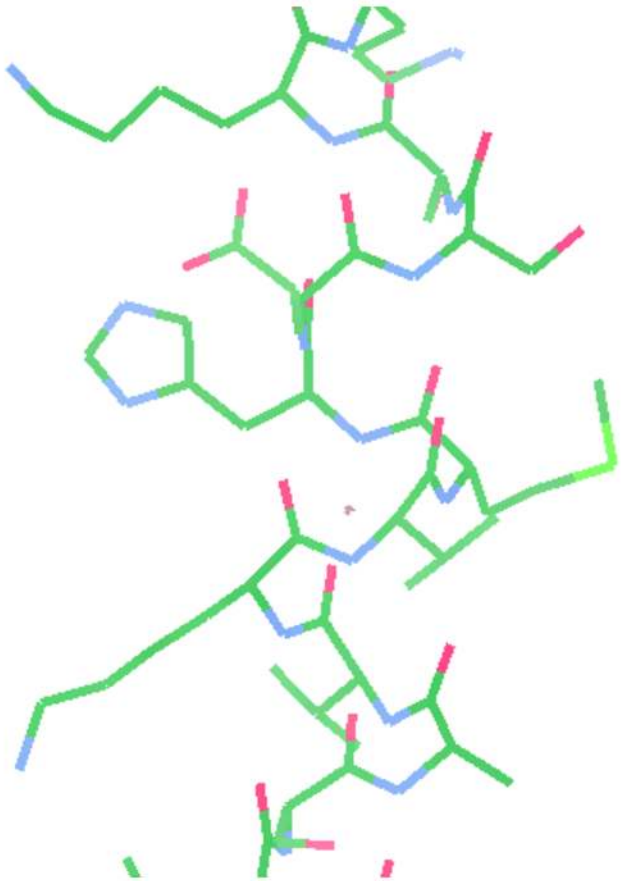
So what is the model, and how do we refine it?

```
ATOM      1  N   ASP A   8     -28.870  10.698  19.213  1.00 21.81          N
ATOM      2  CA  ASP A   8     -29.963  10.919  18.239  1.00 20.35          C
ATOM      3  C   ASP A   8     -30.625  12.268  18.547  1.00 18.04          C
ATOM      4  O   ASP A   8     -30.345  12.830  19.616  1.00 15.38          O
ATOM      5  CB  ASP A   8     -30.909   9.723  18.264  1.00 33.70          C
ATOM      6  CG  ASP A   8     -31.252   9.345  16.825  1.00 41.96          C
ATOM      7  OD1 ASP A   8     -31.072  10.248  15.981  1.00 46.18          O
```

…the model is not only this

Model Parameterisation

Standard refinable parameters



Atomic model:

- Position – (x,y,z) coordinates
- Uncertainty – B-factors
- (Occupancies)

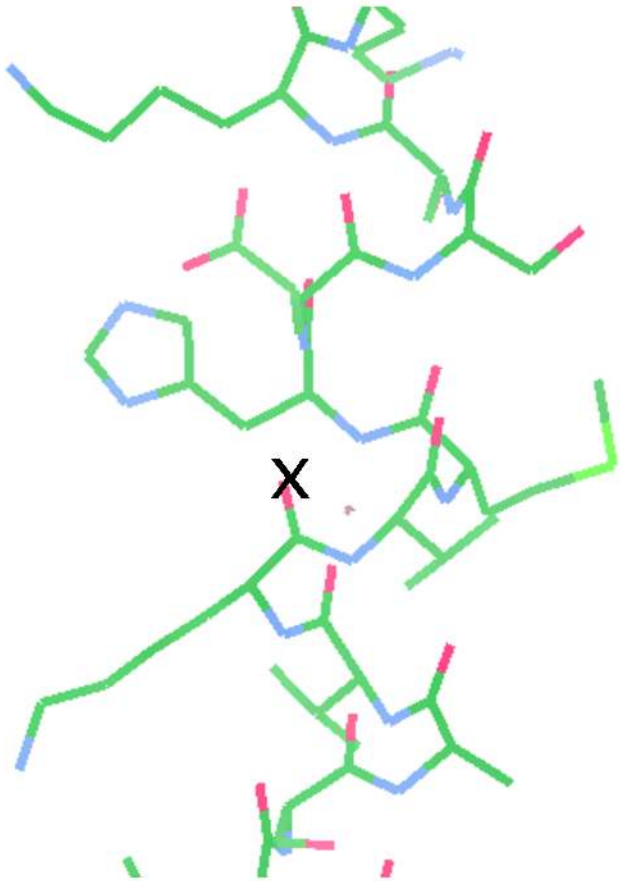
Overall parameters (scaling)

- Overall B-factor (and anisotropic U)
- Solvent treatment

```
ATOM      5  CB  ASP A   8     -30.909   9.723  18.264  1.00 33.70          C
ATOM      6  CG  ASP A   8     -31.252   9.345  16.825  1.00 41.96          C
ATOM      7  OD1 ASP A   8     -31.072  10.248  15.981  1.00 46.18          O
```

Model Parameterisation

Standard refinable parameters



Atomic model:

- **<u>Position – (x,y,z) coordinates</u>**
- Uncertainty – B-factors
- (Occupancies)

Overall parameters (scaling)

- Overall B-factor (and anisotropic U)
- Solvent treatment

```
ATOM      5  CB   ASP A   8      -30.909   9.723  18.264  1.00 33.70          C
ATOM      6  CG   ASP A   8      -31.252   9.345  16.825  1.00 41.96          C
ATOM      7  OD1  ASP A   8      -31.072  10.248  15.981  1.00 46.18          O
```

Model Parameterisation

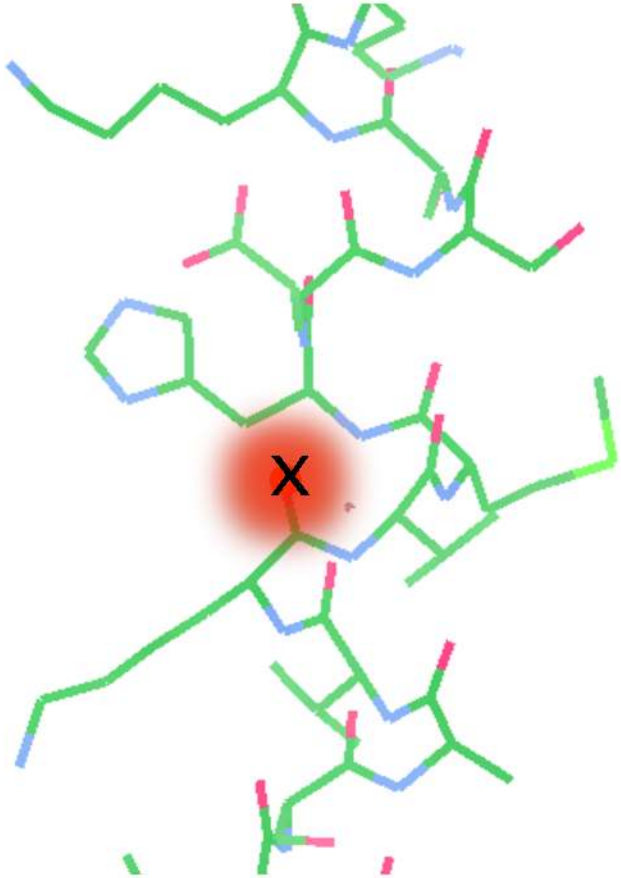
Standard refinable parameters

Atomic model:

- Position – (x,y,z) coordinates
- **<u>Uncertainty – B-factors</u>**
- (Occupancies)

Overall parameters (scaling)

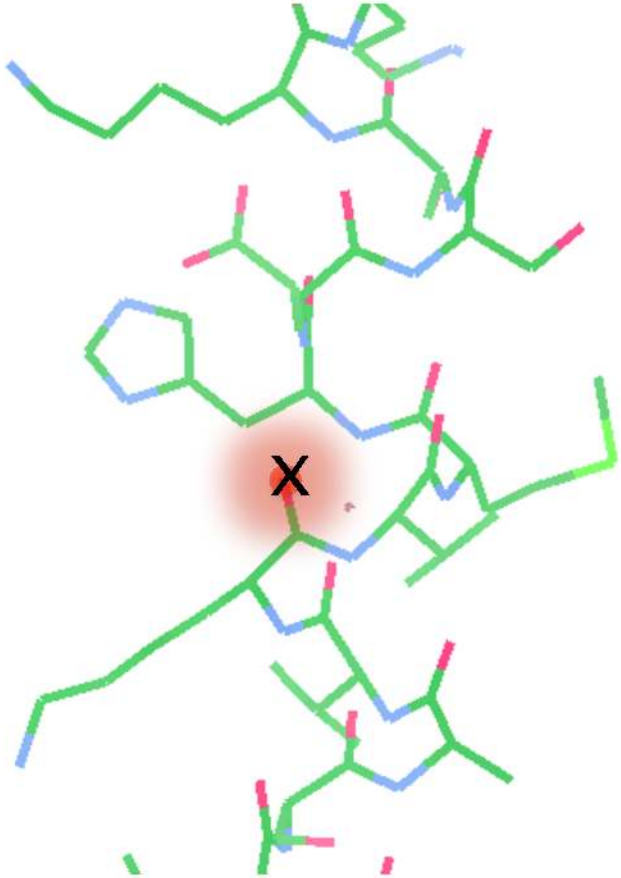
- Overall B-factor (and anisotropic U)
- Solvent treatment




```
ATOM      5  CB   ASP A   8      -30.909   9.723  18.264  1.00 33.70          C
ATOM      6  CG   ASP A   8      -31.252   9.345  16.825  1.00 41.96          C
ATOM      7  OD1  ASP A   8      -31.072  10.248  15.981  1.00 46.18          O
```

Model Parameterisation

Standard refinable parameters



Atomic model:

- Position – (x,y,z) coordinates
- Uncertainty – B-factors
- **<u>(Occupancies)</u>**

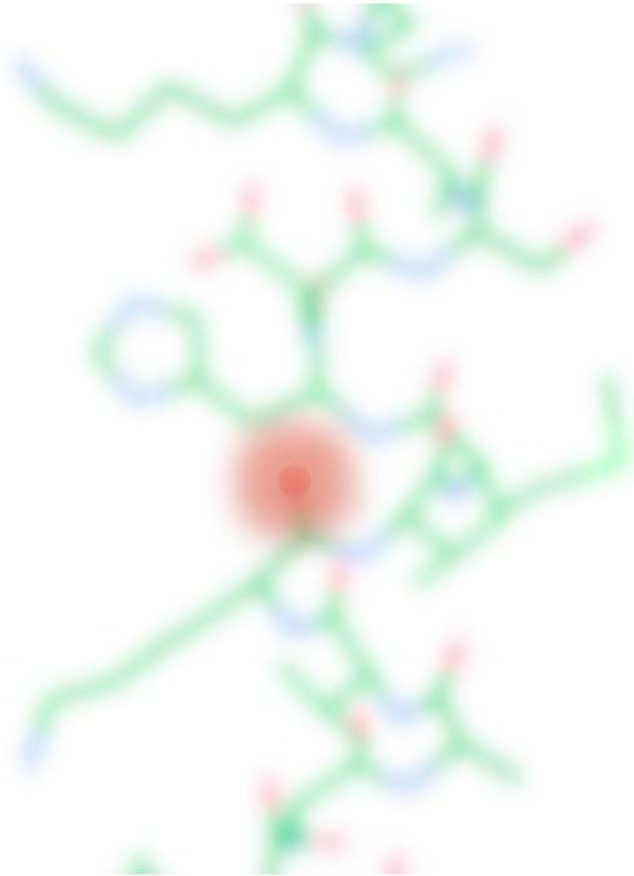
Overall parameters (scaling)

- Overall B-factor (and anisotropic U)
- Solvent treatment

```
ATOM      5  CB   ASP A   8      -30.909   9.723  18.264  1.00 38.70          C
ATOM      6  CG   ASP A   8      -31.252   9.345  16.825  1.00 41.96          C
ATOM      7  OD1  ASP A   8      -31.072  10.248  15.981  1.00 46.18          O
```

Model Parameterisation

Standard refinable parameters



Atomic model:

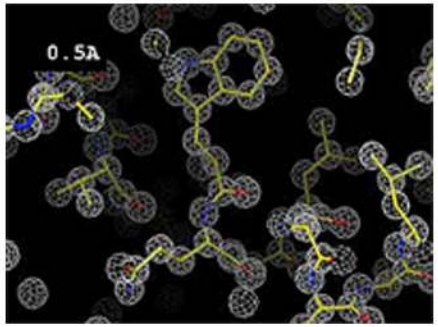
- Position – (x,y,z) coordinates
- Uncertainty – B-factors
- (Occupancies)

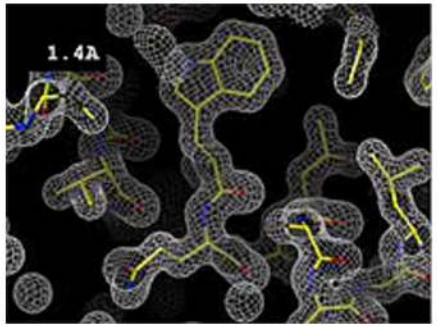
Overall parameters (scaling)

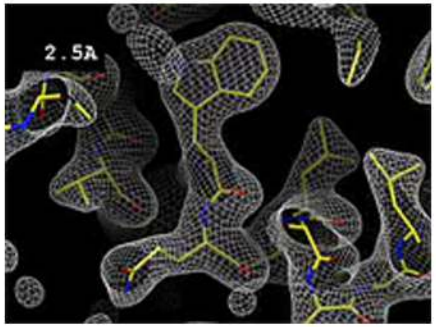
- **<u>Overall B-factor (and anisotropic U)</u>**
- Solvent treatment

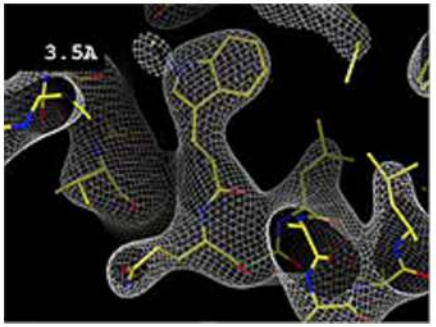
```
ATOM      5  CB  ASP A   8     -30.909   9.723  18.264  1.00 33.70          C
ATOM      6  CG  ASP A   8     -31.252   9.345  16.825  1.00 41.96          C
ATOM      7  OD1 ASP A   8     -31.072  10.248  15.981  1.00 46.18          O
```

Model Parameterisation



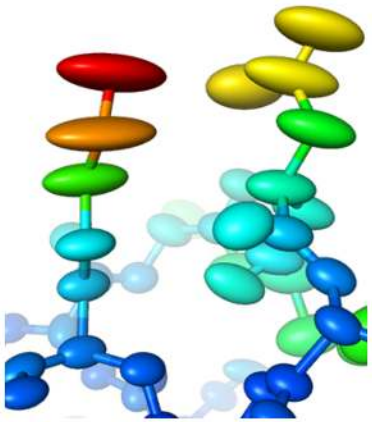


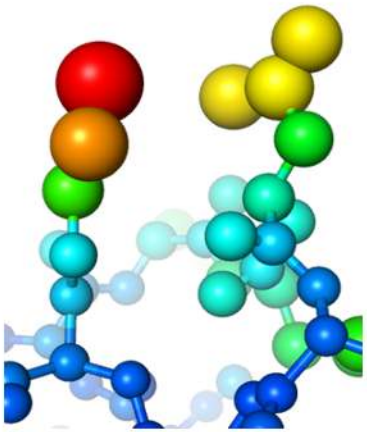




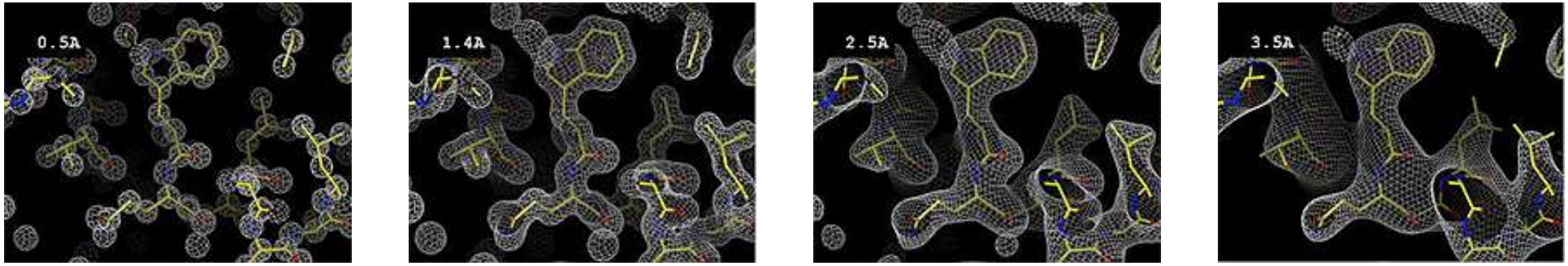
Notes on B-factors

- Overall B-factor and atomic B-factor are different!
- Atomic B-factors can be modelled in a different way according to the quality of the data: **anisotropic** (6 parameters per atom), **isotropic** (1 parameter per atom), **TLS** (20 parameters per group of atoms)





Model Parameterisation



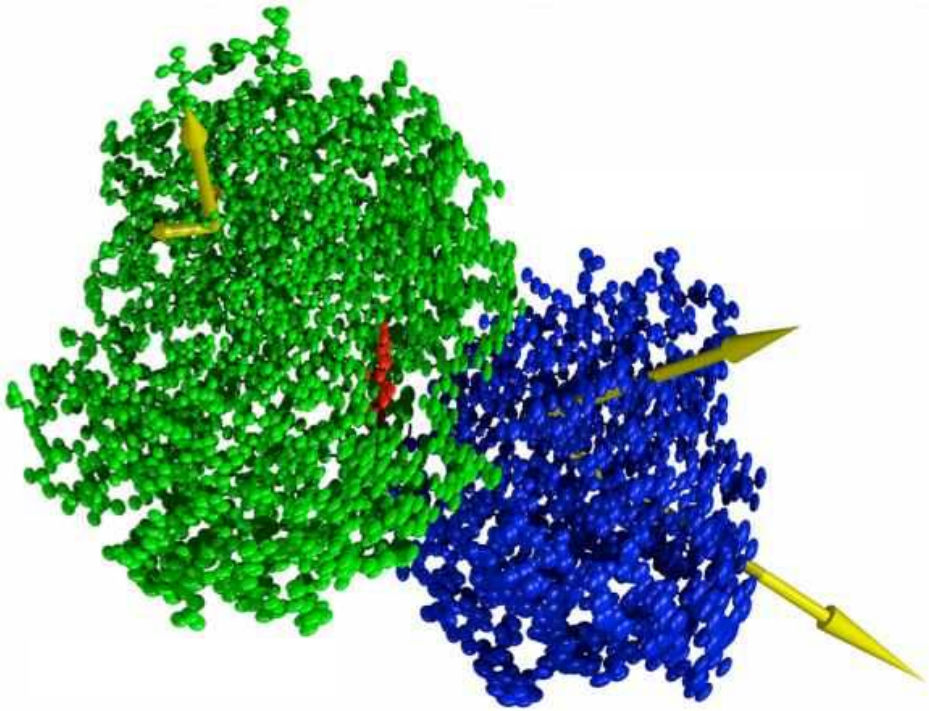
Notes on B-factors

- Overall B-factor and atomic B-factor are different!
- Atomic B-factors can be modelled in a different way according to the quality of the data: **isotropic** (1 parameter per atom), **anisotropic** (6 parameters per atom), **TLS** (20 parameters per group of atoms)
- B-factors are sometimes also referred to as atomic displacement parameters (ADPs) or thermal/temperature factors
- B-factors describe relative positional uncertainty
- Should not compare atomic B-factors between different models

TLS Groups

Describe rigid body motion – e.g. for chains/domains/subunits

Suitable for medium resolution, when full anisotropy is impossible



Per group (20 parameters):

- **T**ranslation – 6 parameters
- **L**ibration – 6 parameters
- **S**crew rotation – 8 parameters

Define groups using CCP4i

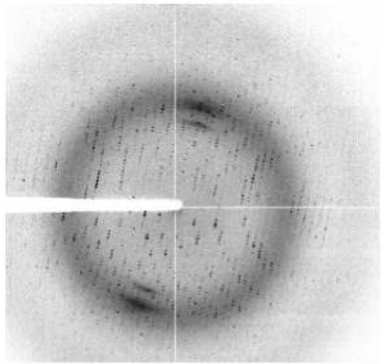
or TLSMD webserver:

http://skuld.bmsc.washington.edu/~tlsmd/

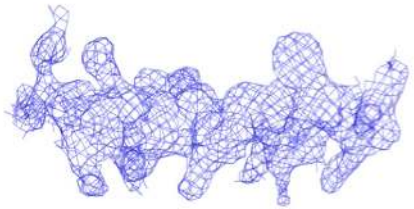
Overall Parameters: Scaling

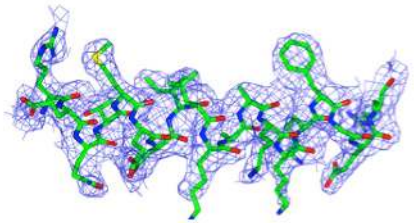
Problem:

- Observed and calculated amplitudes need to be brought to the same scale so that they can be compared



H, K, L	$\lvert F_{obs} \rvert$	ϕ
...		
5, 5, 5	348	-
5, 5, 6	392	-
5, 5, 7	157	-
5, 5, 8	312	-
...		



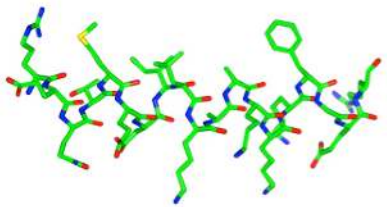


Idea:

Iteratively improve the model, optimising the agreement between $\lvert F_{obs} \rvert$ and $\lvert F_{calc} \rvert$

Purpose: improve phase estimates: φ_{calc}

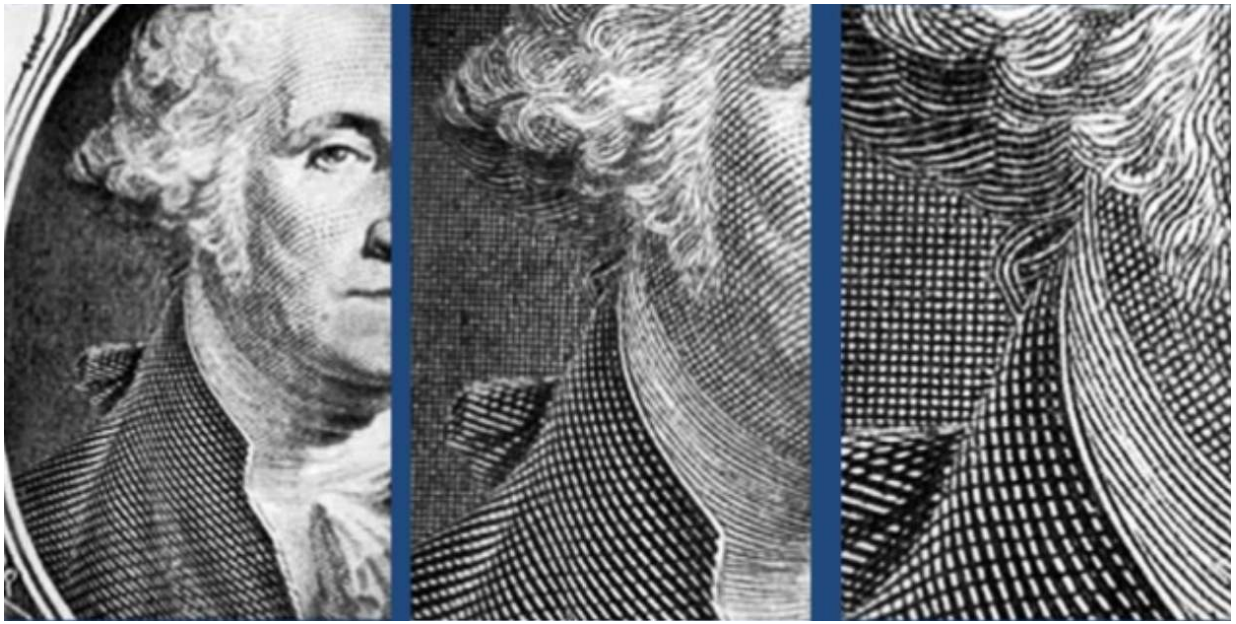
H, K, L	$\lvert F_{calc} \rvert$	ϕ_{calc}
...		
5, 5, 5	355	27°
5, 5, 6	387	8°
5, 5, 7	146	75°
5, 5, 8	340	31°
...		



Overall Parameters: Scaling

Problem:

- Observed and calculated amplitudes need to be brought to the same scale so that they can be compared



Overall Parameters: Scaling

Problem:

- Observed and calculated amplitudes need to be brought to the same scale so that they can be compared

Need to:

- Modify/scale F_{calc}
- Find a scaling function, with some parameters – "overall parameters"

Scale parameters are optimised in ML refinement, along with all other parameters

Overall Parameters: Scaling

Want to find some parameters that allow scaling of F_{calc} in order to better agree with F_{obs}

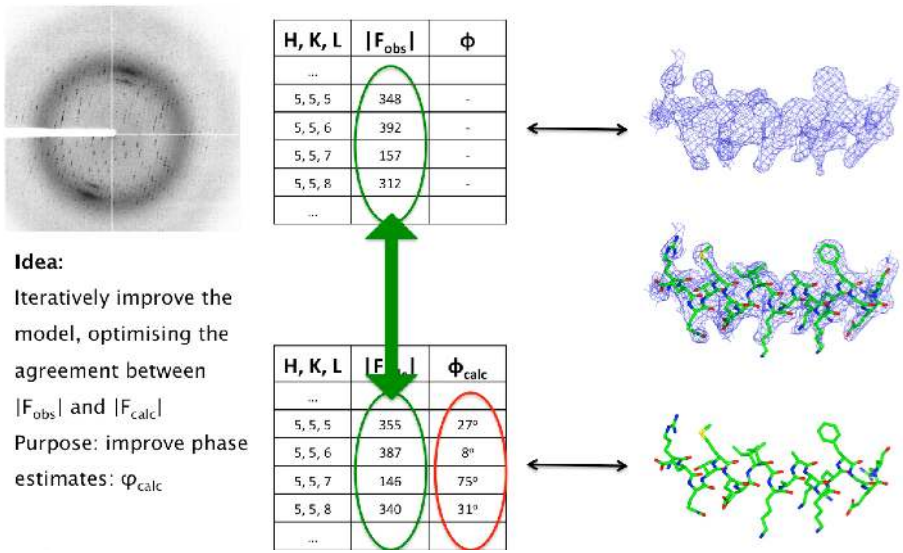
$$k\ e^{-Bs^2}\ e^{-s^T U s}\ F_{calc}$$

k : overall scale factor

s : reciprocal space vector

B : overall B-factor

U : anisotropic B-factor



Overall Parameters: Scaling

Want to find some parameters that allow scaling of F_{calc}
in order to better agree with F_{obs}

$$k \; e^{-Bs^2} \; e^{-s^T U s} \; F_{calc}$$

k : overall scale factor

s : reciprocal space vector

B : overall B-factor

U : anisotropic B-factor

But what about the unmodelled solvent regions?

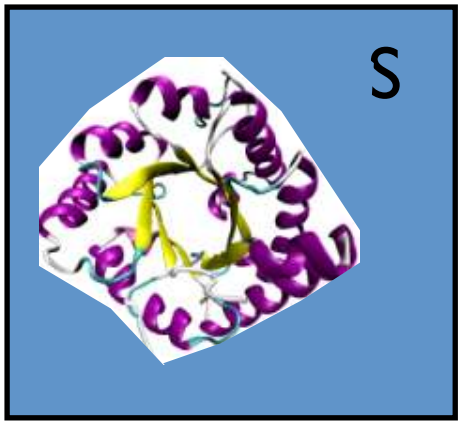
Problem: Our F_{calc} only corresponds to the ordered regions (protein)

Overall Parameters: Scaling

Mask-based bulk solvent correction

Idea:

- Protein region is masked out
- Solvent region is flattened (set to constant)
- Structure factors for solvent are calculated: $F_{solvent}$



$$F_{total} = F_{protein} + \alpha F_{solvent}$$

α : solvent scale factor

$$k\ e^{-Bs^2}\ e^{-s^T U s}\ (F_{protein} + \alpha F_{solvent})$$

k : overall scale factor

B : overall B-factor

U : overall anisotropic B-factor

Model Refinement

We now have:

- Data – to refine our model against
- Parameters to refine – describing the model

How do we refine the model?

Model Refinement

REFMAC5 uses a Maximum Likelihood approach

Maximum Likelihood (ML) target:

likelihood: $P(|F_{obs}|;F_{calc})$

prior

$$P(model;obs) \rightarrow max$$

$$P(model;obs) \propto P(obs;model)\ P(model) \rightarrow max$$

$$\log[P(obs;model)] + \log[P(model)] \rightarrow max$$

$$-\log[P(obs;model)] - \log[P(model)] \rightarrow min$$

Objective – minimise the negative log-likelihood

Model Refinement

REFMAC uses a Maximum Likelihood approach

Crystallographic target functions have two components:

$$f_{tot} = wf_{xray} + f_{geom}$$

$f_{xray} = -\log[P(obs;model)]$

$f_{geom} = -\log[P(model)]$

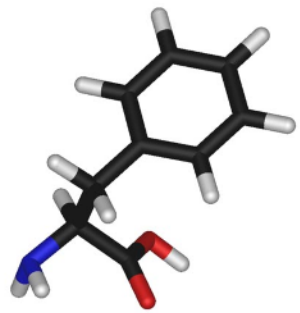
w : relative weighting

likelihood of the data

probability of the model

We have:

- Data – to refine our model against
- Parameters to refine – describing the model



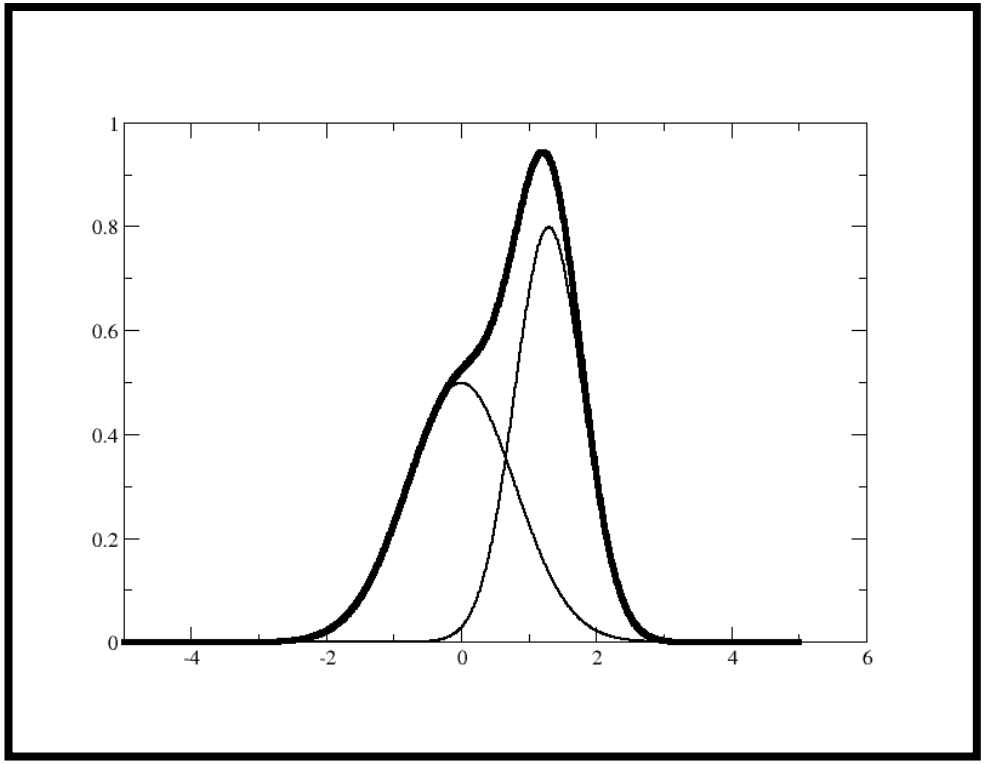
We also need prior knowledge (restraints)

These help ensure chemical and structural integrity

Why Restraints?

Example: two-atom ideal case

Distance between atoms 1.3Å. B-factors 20 and 50



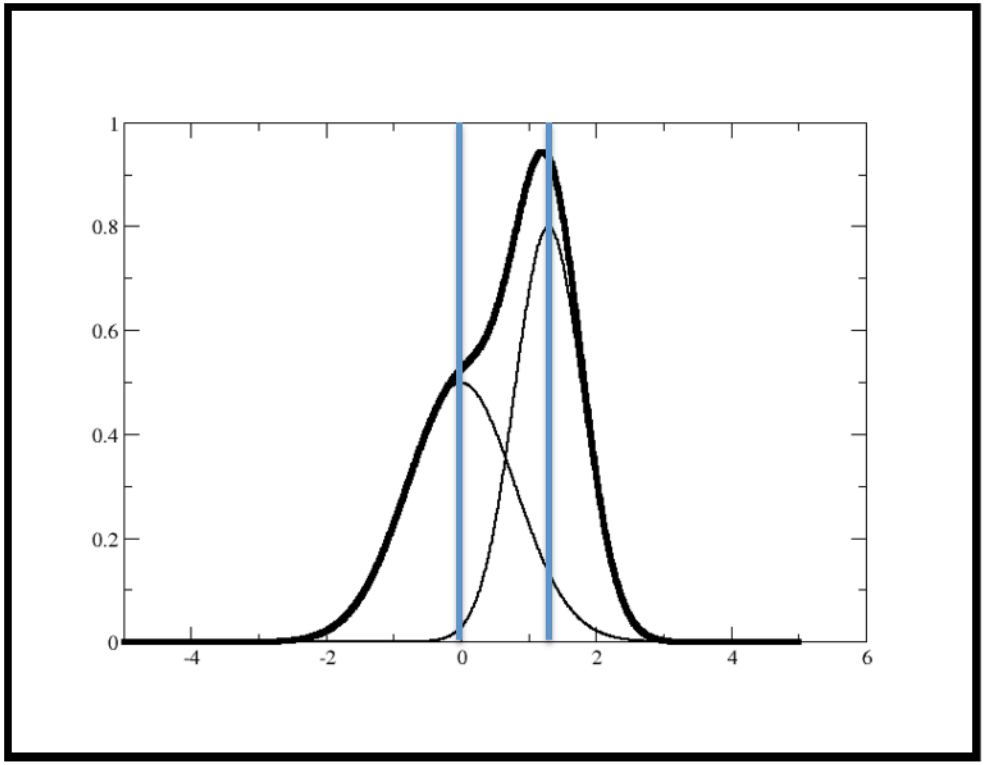
Thin lines - single atoms

Bold line - sum of the two atoms

Why Restraints?

Example: two-atom ideal case

Distance between atoms 1.3Å. B-factors 20 and 50



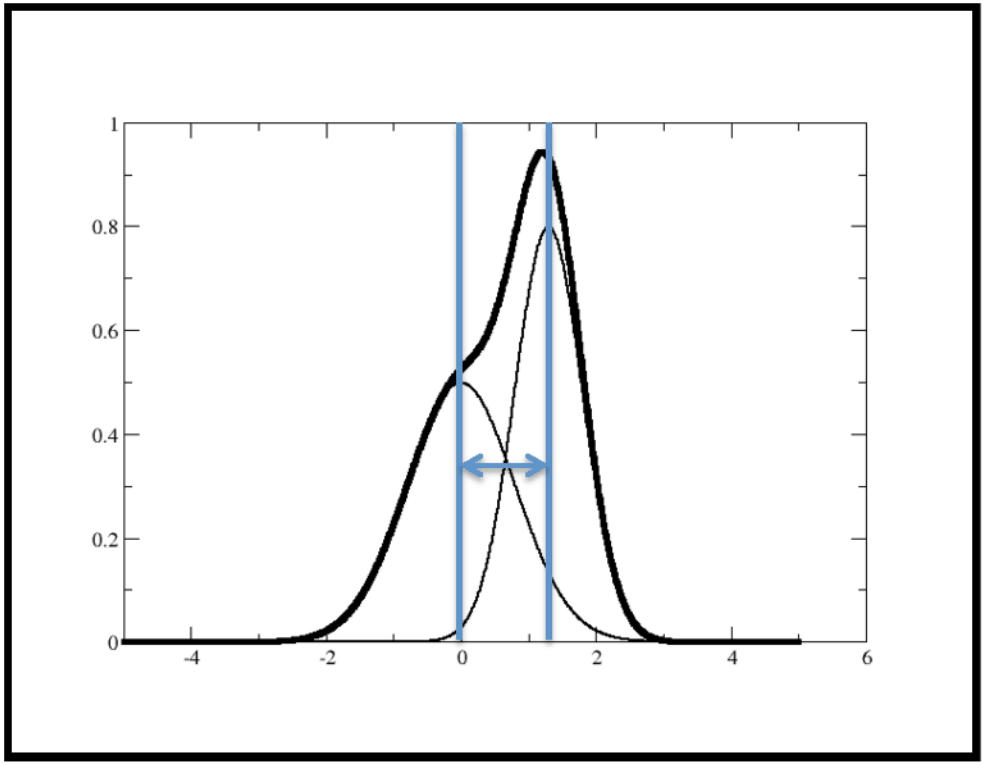
Thin lines – single atoms

Bold line – sum of the two atoms

Why Restraints?

Example: two-atom ideal case

Distance between atoms 1.3Å. B-factors 20 and 50



Thin lines – single atoms

Bold line – sum of the two atoms

Restraints

Standard restraints (used by default) include:

- Bond lengths
- Angles
- Chirals
- Planes
- Some torsion angles
- B-values
- VDW repulsions

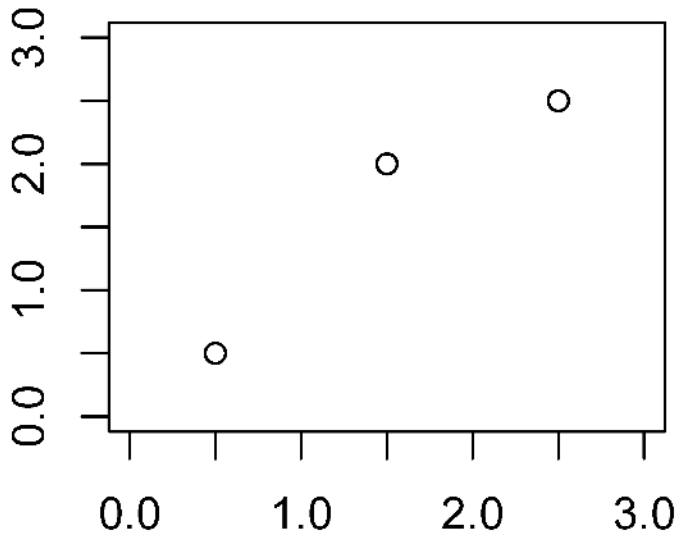
These help to ensure that the model is chemically sensible

Note – we generally deal with restraints, not constraints

Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.

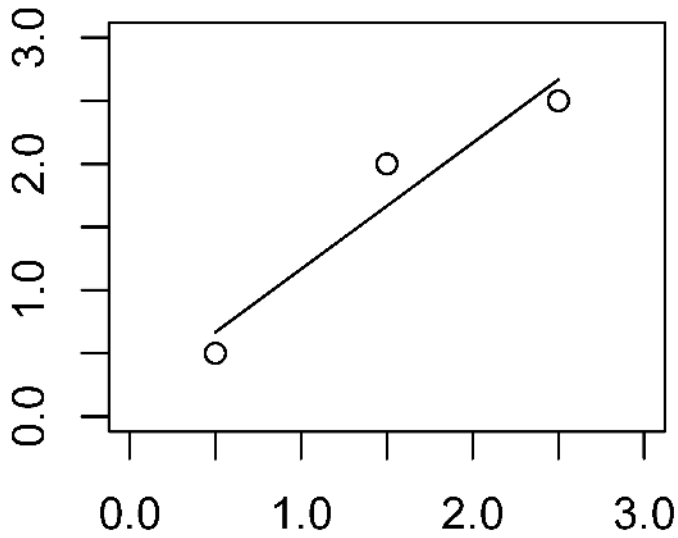


Example: Fitting a line $y = a + bx$

Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.

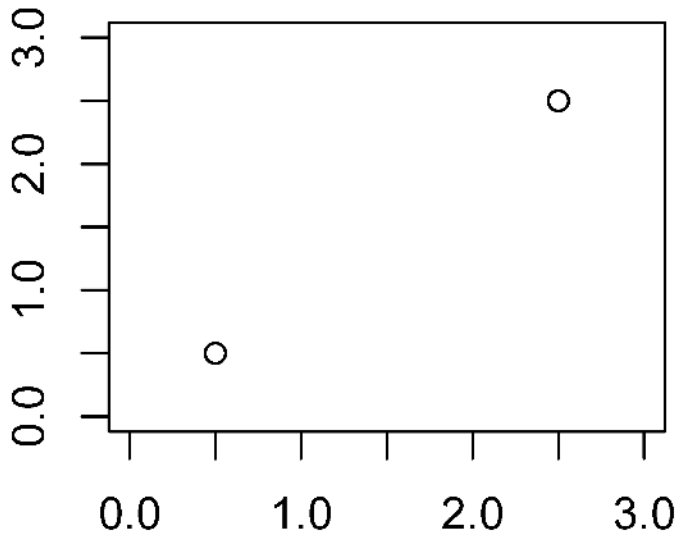


Example: Fitting a line $y = a + bx$

Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.



Example: Fitting a line $y = a + bx$

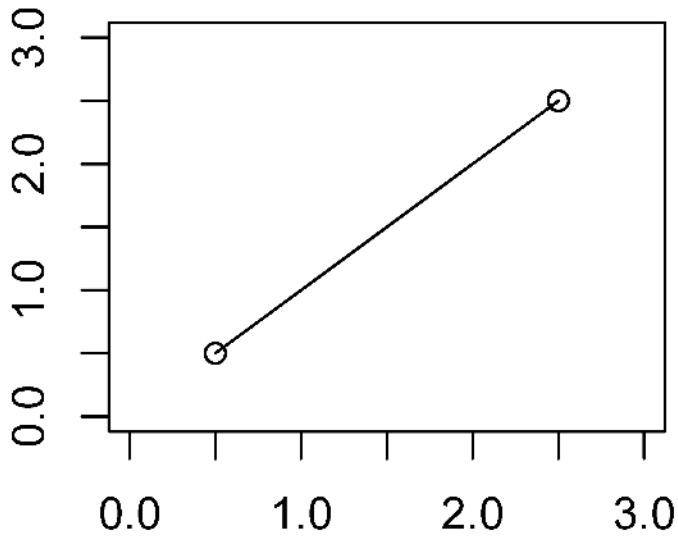
Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.

Can fit a line

Line is unreliable



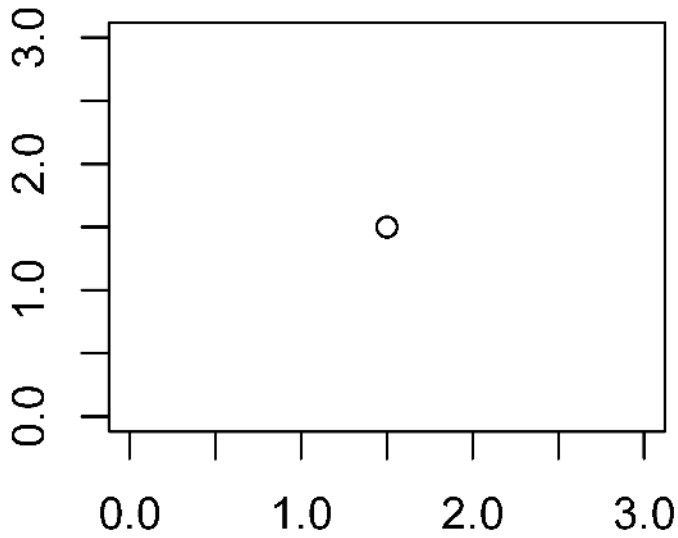
Overfitting
Model Bias

Example: Fitting a line $y = a + bx$

Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.



Example: Fitting a line $y = a + bx$

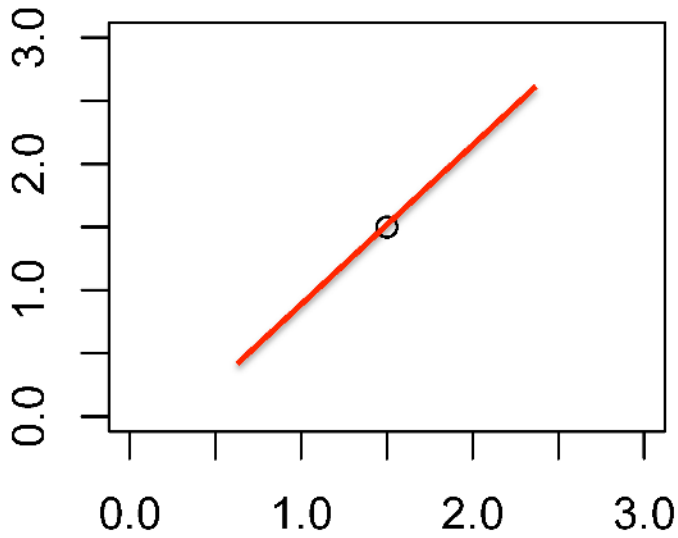
Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.

Insufficient observations!

Unstable refinement



Ill-posed problem

Example: Fitting a line $y = a + bx$

Restraints

How to improve the observation:parameter ratio.

1. Reduce number of parameters

N_{obs} is resolution dependent…

High resolution	:	Anisotropic B-factors	– 9 params per atom
Low resolution	:	Isotropic B-factors	– 4 params per atom

- TLS – 20 additional parameters per group
- Rigid body refinement – 9 parameters per body

Restraints

How to improve the observation:parameter ratio.

1. Reduce number of parameters

2. Increase number of restraints

   - B-value restraints
   - NCS restraints
   - H-bond and secondary-structure restraints
   - Restraints to homologous known structures
   - Nucleic acid base-pair and base-stacking restraints
   - Jelly-body restraints

Ligand Refinement

Geometric restraints for protein / nucleic acids are pre-tabulated

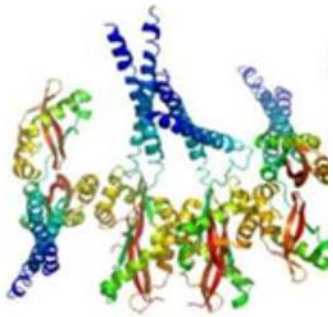
Ligands are more complicated

Need a source of prior information

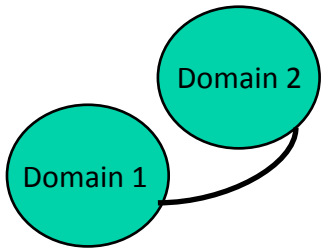
- Common/known structures are dealt with automatically
  - CCP4/REFMAC monomer library has pre-computed descriptions
- New ligands require description (CIF file)
  - New tool – ACEDRG

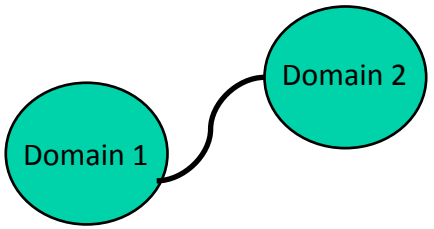
NCS
(Non-Crystallographic Symmetry Restraints)

1. NCS constraints
   - NCS-related copies are considered to be exactly the same
   - Only one set of atomic parameters per molecule is refined
2. Global NCS restraints
   - Molecules are superimposed
   - Difference between atoms are minimised



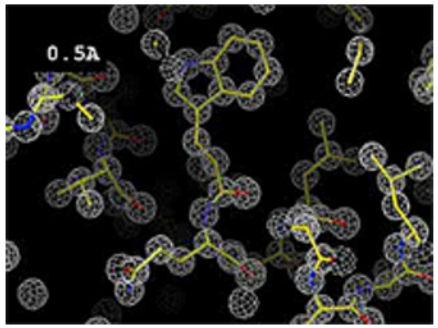
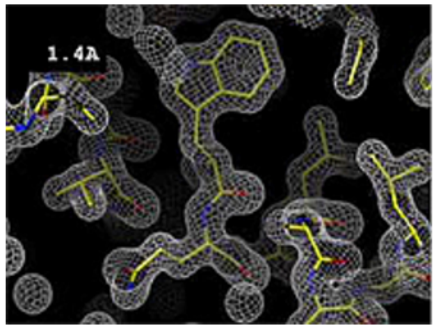
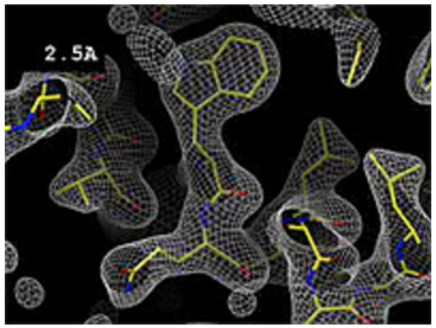
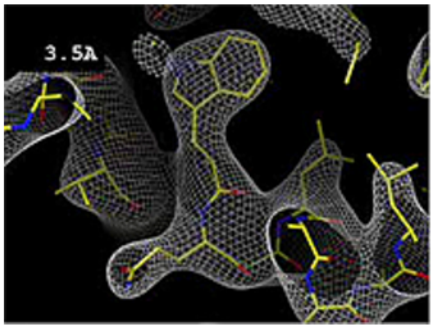
3. Local NCS restraints
   - Molecules are assumed to be locally similar
   - However, they may adopt (slightly) different global conformations
   - Restrain differences between local interatomic distances





Model Refinement

Refinement strategy will differ for different quality of original data and it can also exploit particular features of your crystal:

1.0Å	2.0Å	3.0Å	4.0Å
			
Unrestrained Anisotropic B Hydrogens	Chemical restraints Isotropic B TLS	Chemical restraints Isotropic B TLS	Chemical restraints Jelly-body restraints External restraints Isotropic B

- Twinning
- Several similar subunits – NCS (at moderate and low resolution)
- Available homologues for external restraint generation (low resolution)

Low Resolution Refinement

Low Resolution Refinement

- ~3Å or lower (tools are sometimes useful at slightly higher resolutions)
- Reflection intensities often noisy
- Limited data – poor observation:parameter ratio

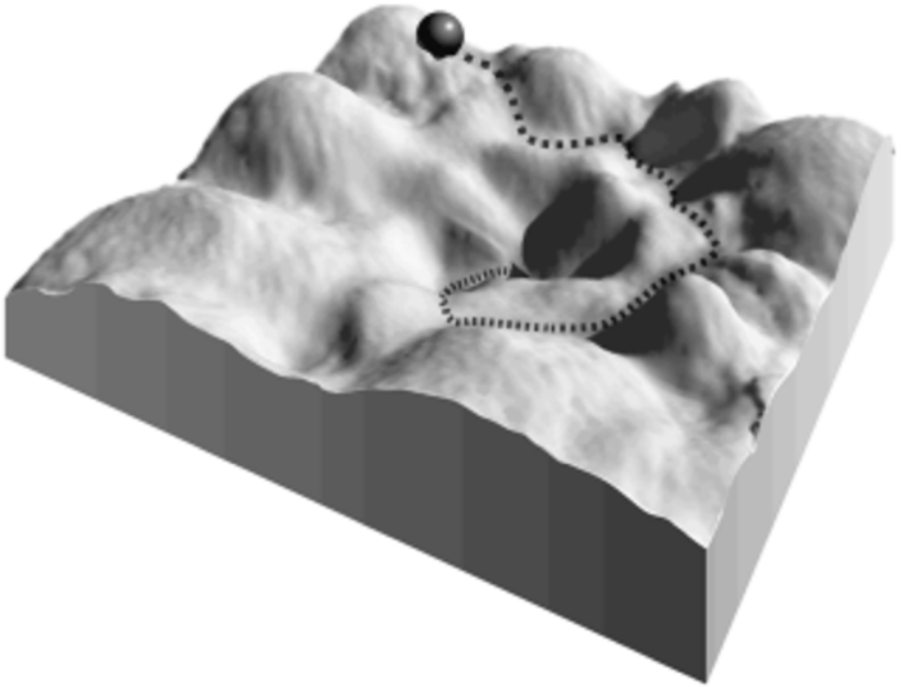
At low resolution, refinement often becomes unstable

Overfitting – R-factors diverge

To improve models, we need to:

- **Ease** manual **model building**
- Use of **prior knowledge** to **stabilize refinement** – ***Regularisers***

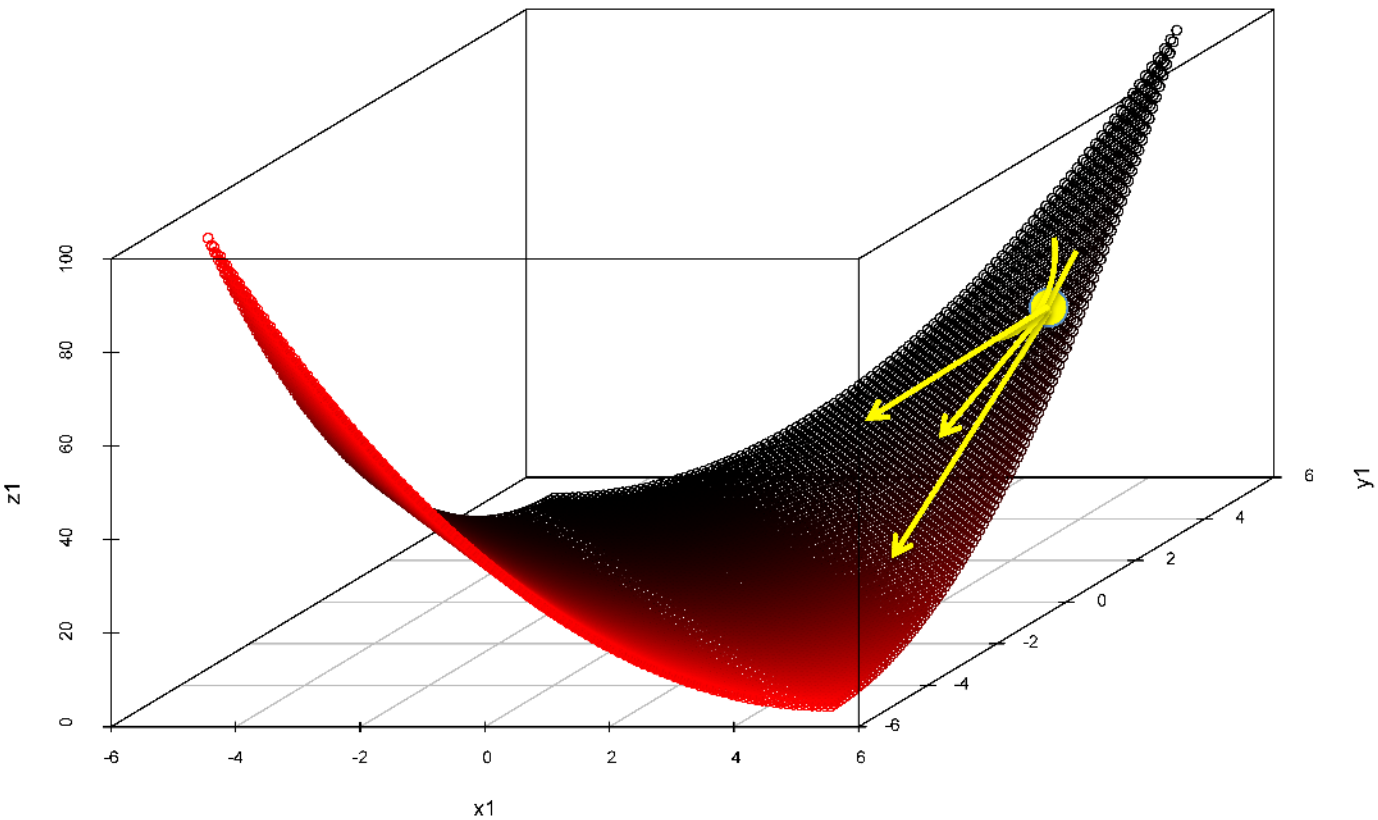
Regularisation



Regularisation

Example:

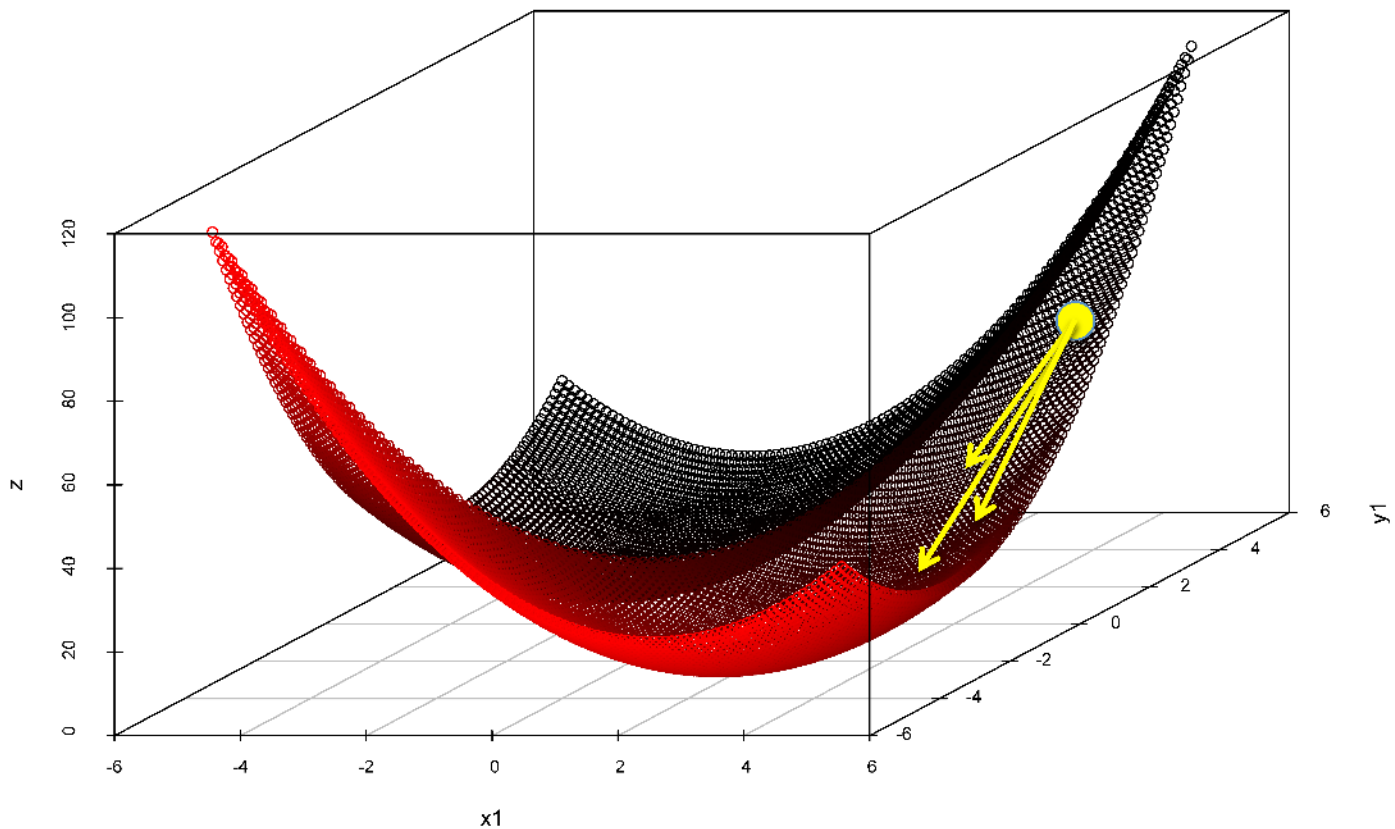
$$z = (x + y)^2$$



Regularisation

Example:

$$z = (x+y)^2 + (|x-y| - 4)^2$$



Regularise using prior information: $|x-y| = 4$

Regularisation

Use of available knowledge (prior information):

High-low resolution:

- Geometry restraints (chemical information)

Medium-low resolution:

- Local NCS restraints
- B-value restraints
- Jelly body restraints

Low resolution (and medium-low resolution model building):

- External restraints

Regularisation

Use of available knowledge (prior information):

High-low resolution:

- **Geometry restraints (chemical information)**

Medium-low resolution:

- **Local NCS restraints**
- **B-value restraints**
- Jelly body restraints

Low resolution (and medium-low resolution model building):

- **External restraints**

Regularisers with a target value

Regularisation

Use of available knowledge (prior information):

High-low resolution:

- **Geometry restraints (chemical information)**

Medium-low resolution:

- **Local NCS restraints**
- **B-value restraints**
- **Jelly body restraints**

Low resolution (and medium-low resolution model building):

- **External restraints**

Regularisers without an external target value

Jelly Body Restraints

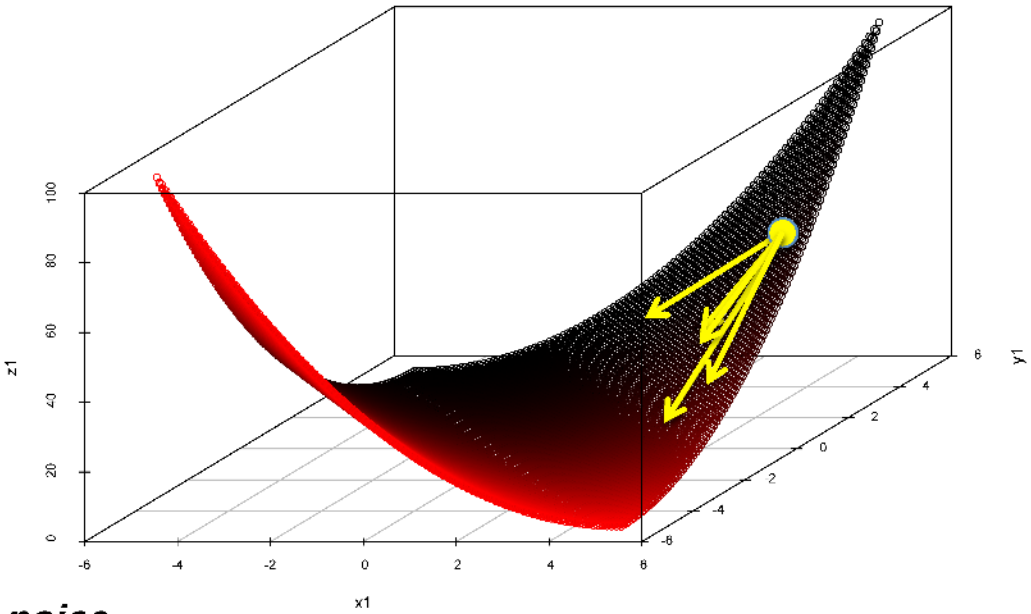
Regularisers without a target:

$$f = \sum_{\text{i : close atom pairs}} \frac{1}{\sigma^2}(|d_i| - |d_{i,\text{current}}|)^2$$

$|d_i|$: interatomic distance

$|d_{i,current}|$: current interatomic distance

σ : restraint standard deviation



Does not change likelihood function.

Does not change derivative.

Does change 2nd derivative – curvature.

Model should be less prone to fitting into noise

Should only work if parameters are near the minima (model is good)

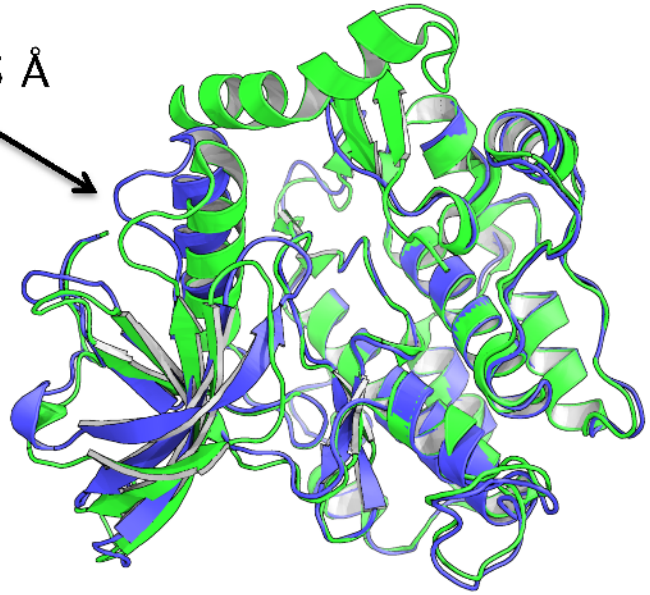
Typical: σ = 0.01–0.02

Distance threshold: 4.2Å

Jelly Body Restraints

Tyrosine Kinase

Positional shift up to 4.5 Å



Molecular replacement
(Molrep, 10s)

2z8c (2007) – 3.25Å

2auh (2005) – 3.2Å

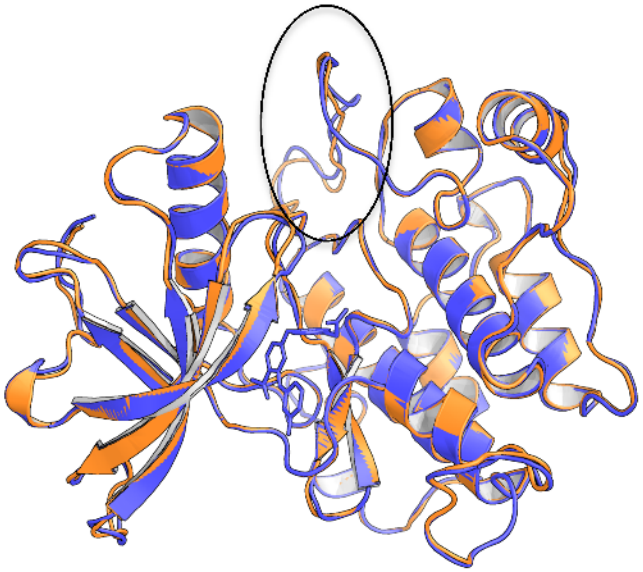
R_{work}/R_{free} : 21.7/29.2 %

R_{work}/R_{free} : 22.3/25.4 %

Jelly-body: R_{work}/R_{free} : 19.7/27.3 %

Jelly Body Restraints

Tyrosine Kinase



Molecular replacement
(Molrep, 10s)

2z8c (2007) - 3.25Å

R_{work}/R_{free} : 21.7/29.2 %

Jelly-body: R_{work}/R_{free} : 19.7/27.3 %

Molprobity clashscore : 28.3

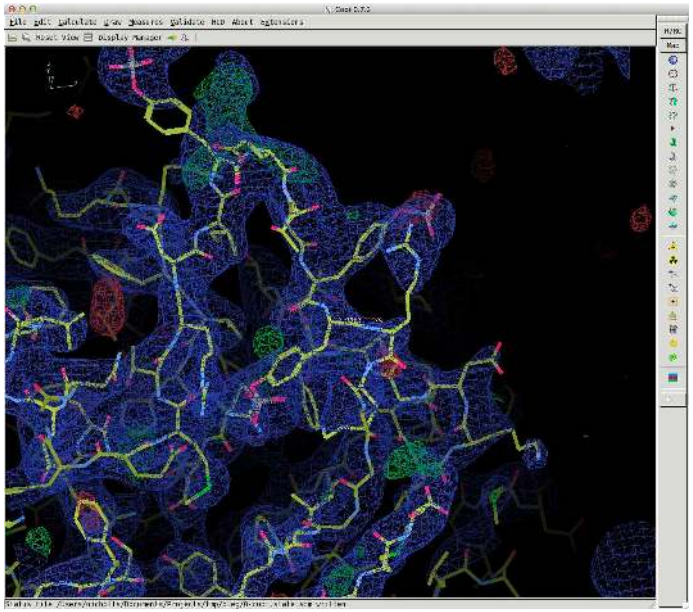
2z8c solved using 2auh,
re-refined using jelly-body

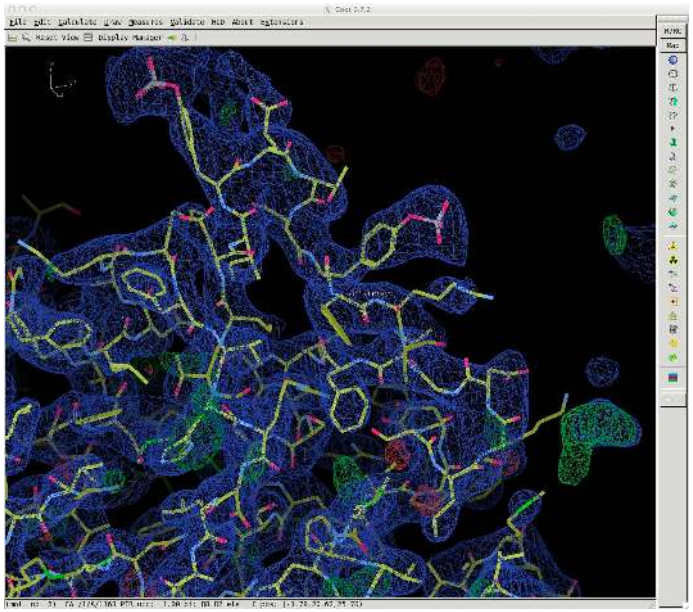
R_{work}/R_{free} : 20.2/25.8 %

Molprobity clashscore : 16.9

Jelly Body Restraints

Tyrosine Kinase





2z8c – 3.25Å

R_{work}/R_{free} : 21.7/29.2 %

Jelly-body: R_{work}/R_{free} : 19.7/27.3 %

Molprobity clashscore : 28.3

2z8c solved using 2auh, re-refined using jelly-body

R_{work}/R_{free} : 20.2/25.8 %

Molprobity clashscore : 16.9

Jelly Body Restraints

Conclusions:

(1) Jelly-body restraints are great at stabilising refinement,
and can have quite a wide radius of convergence…
but can't alone improve a model dramatically

(2) We can't rely on the reliability of models deposited in the PDB
(applies for more moderate-resolution models also)

(3) Structural information in the PDB can be useful in the
determination of new structures

ProSMART in MX Refinement

External restraint generation:

Injection of prior knowledge to aid new structure determination:

- External Restraints from homologous structures
- Hydrogen bond restraints

Restraints utilised by: REFMAC5, Coot

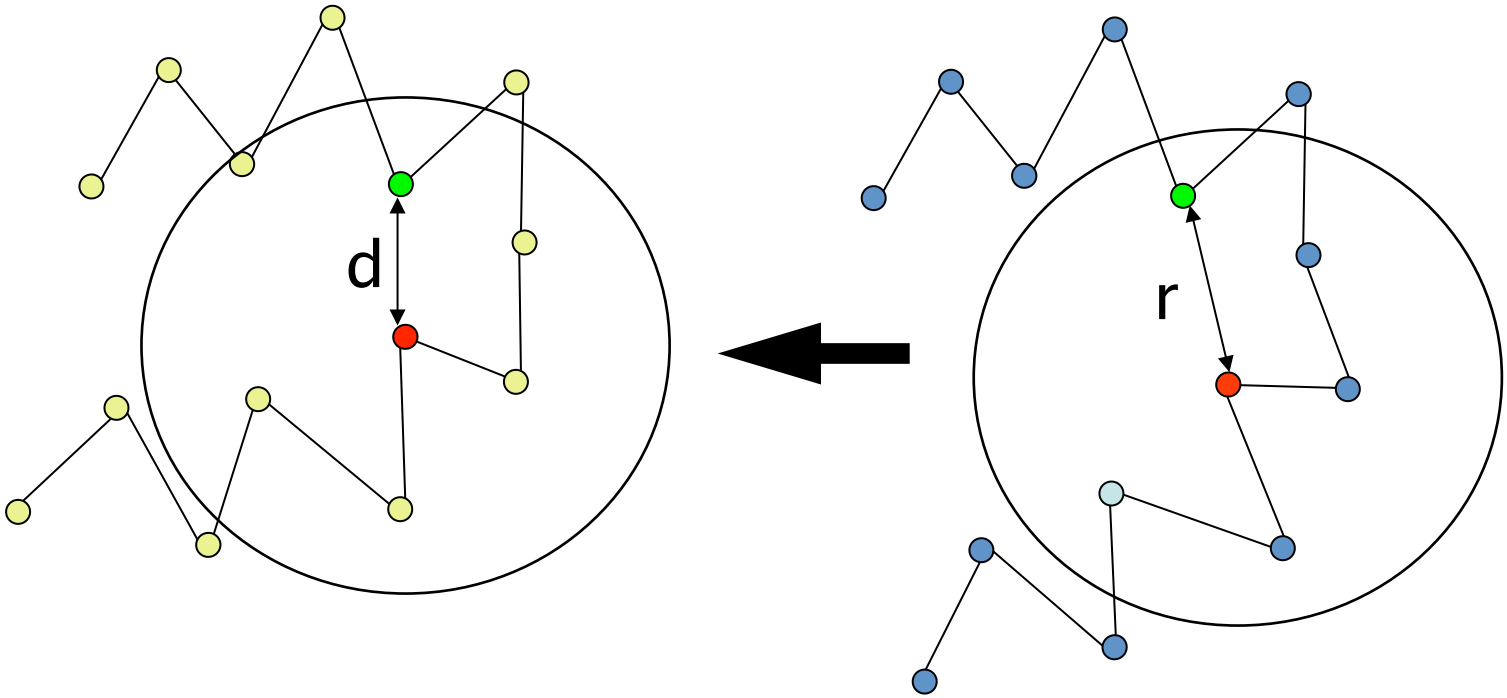
Executed from: CCP4i, CCP4mg, Coot, command line

Independent of global conformation

External Restraint Generation

structure to be refined

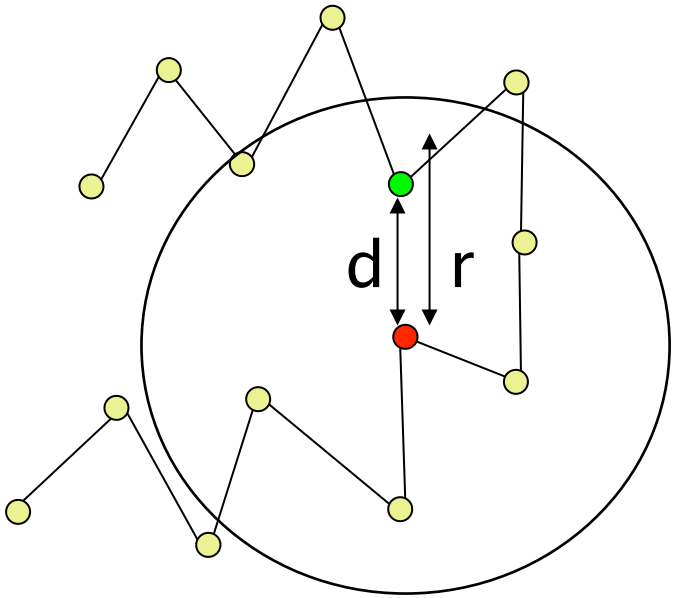
known similar structure (prior)



(abstract representation of an atomic model; circles = atoms)

External Restraint Generation

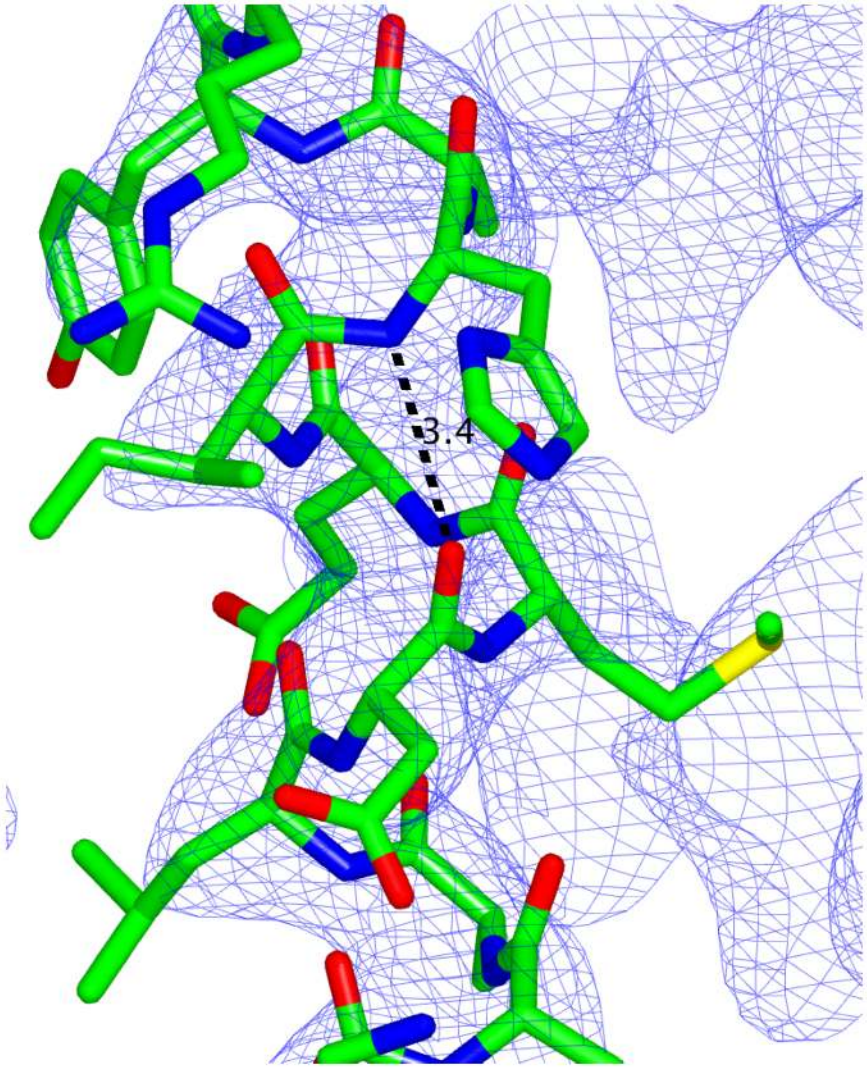
structure to be refined



$$d \sim N(r,\sigma^2)$$

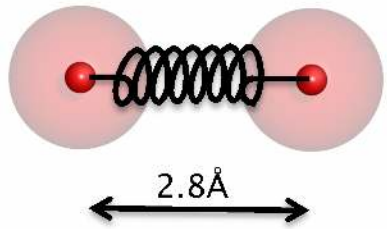
(abstract representation of an atomic model; circles = atoms)

External Restraints



3g4w - 3.7Å

Prior information:



Stabilises structural features

Motivational Example

Ovotransferrin

1ryx - 3.5Å

Low-resolution refinement:

Weak signal
Noisy data

↓

Unstable refinement

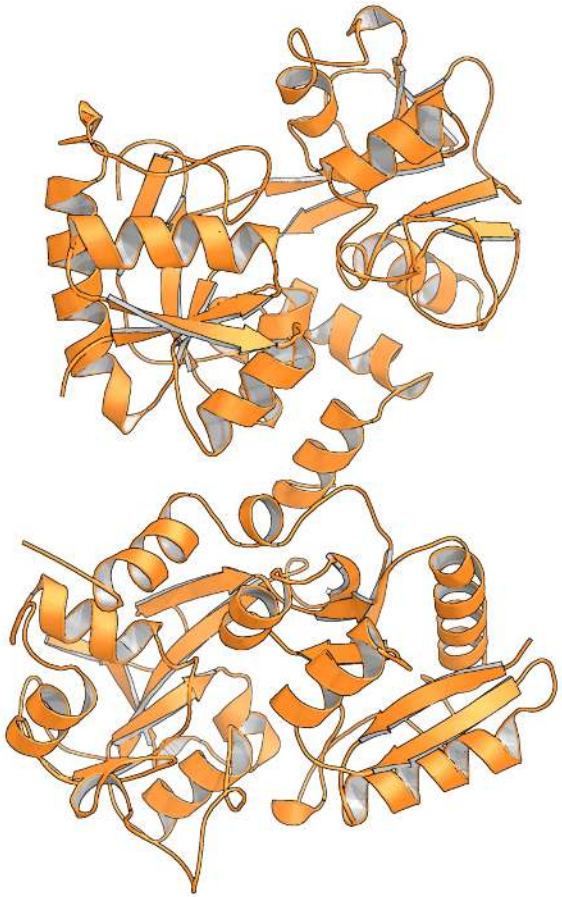
Result:
Poor quality model

Motivational Example

Ovotransferrin

1ryx – 3.5Å

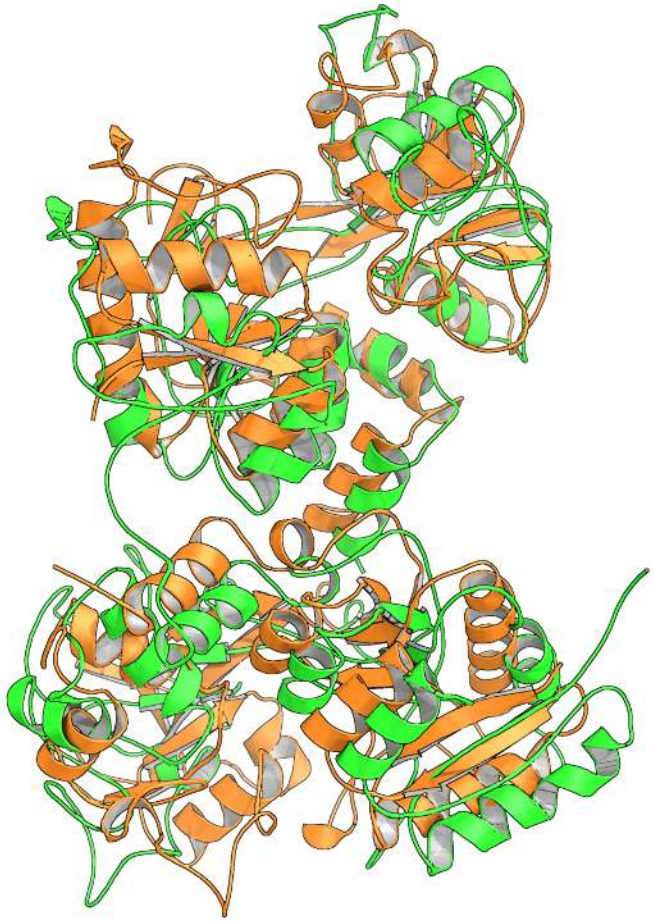
High-resolution homologue



2d3i – 2.15Å

Motivational Example

Ovotransferrin



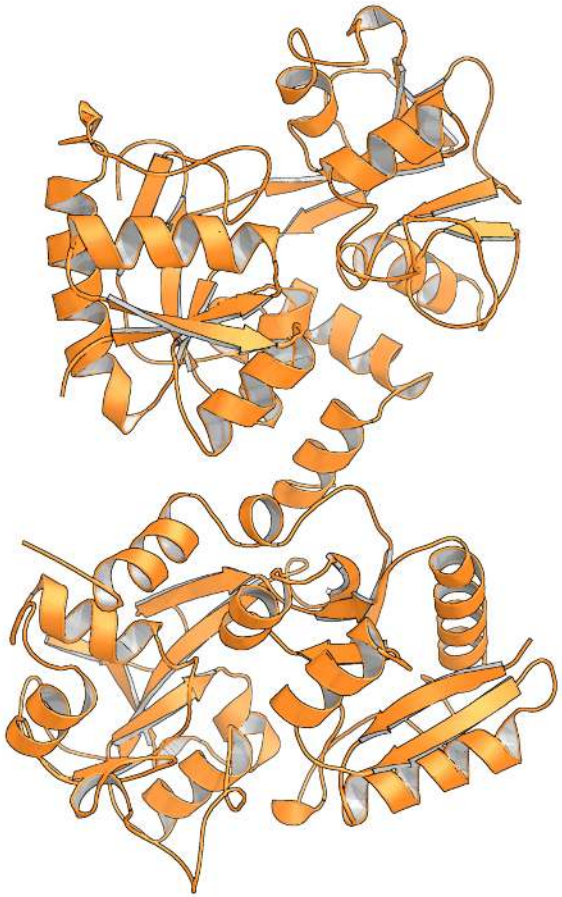
Models don’t superpose well

External Restraints

Ovotransferrin

1ryx – 3.5Å

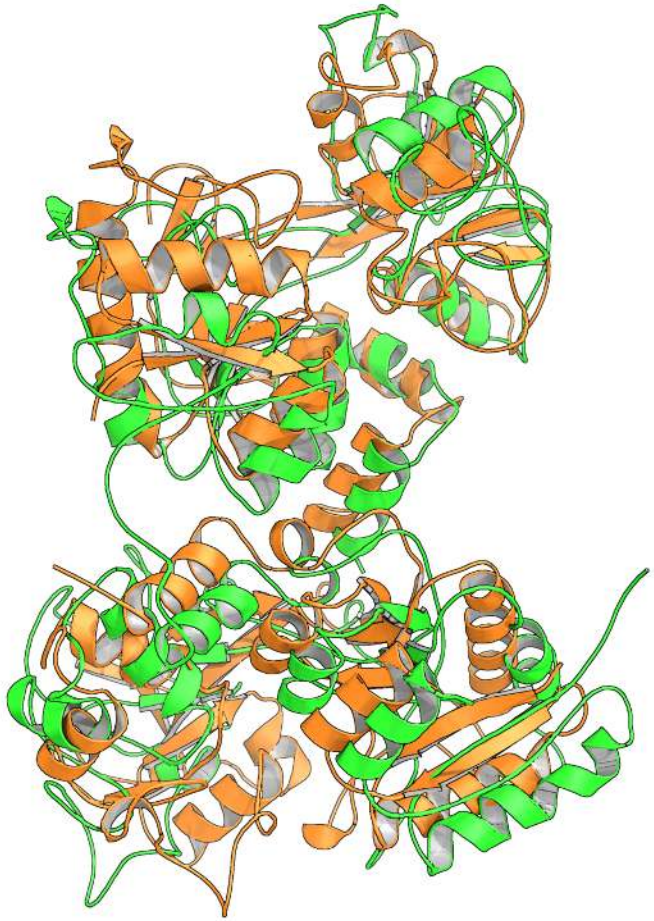
High-resolution homologue



2d3i – 2.15Å

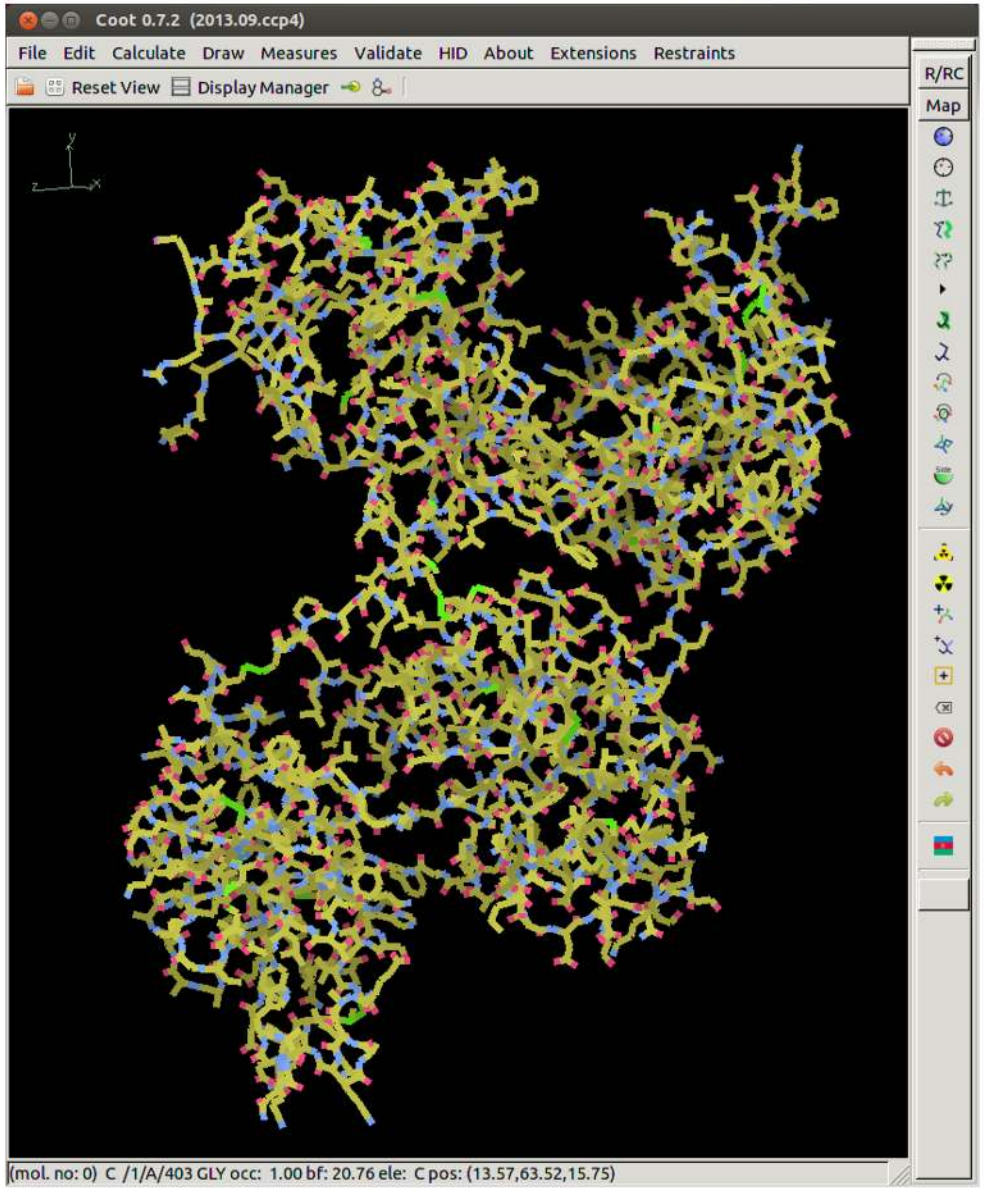
External Restraints

Ovotransferrin



Models don’t superpose well

ProSMART Restraint Visualisation in Coot



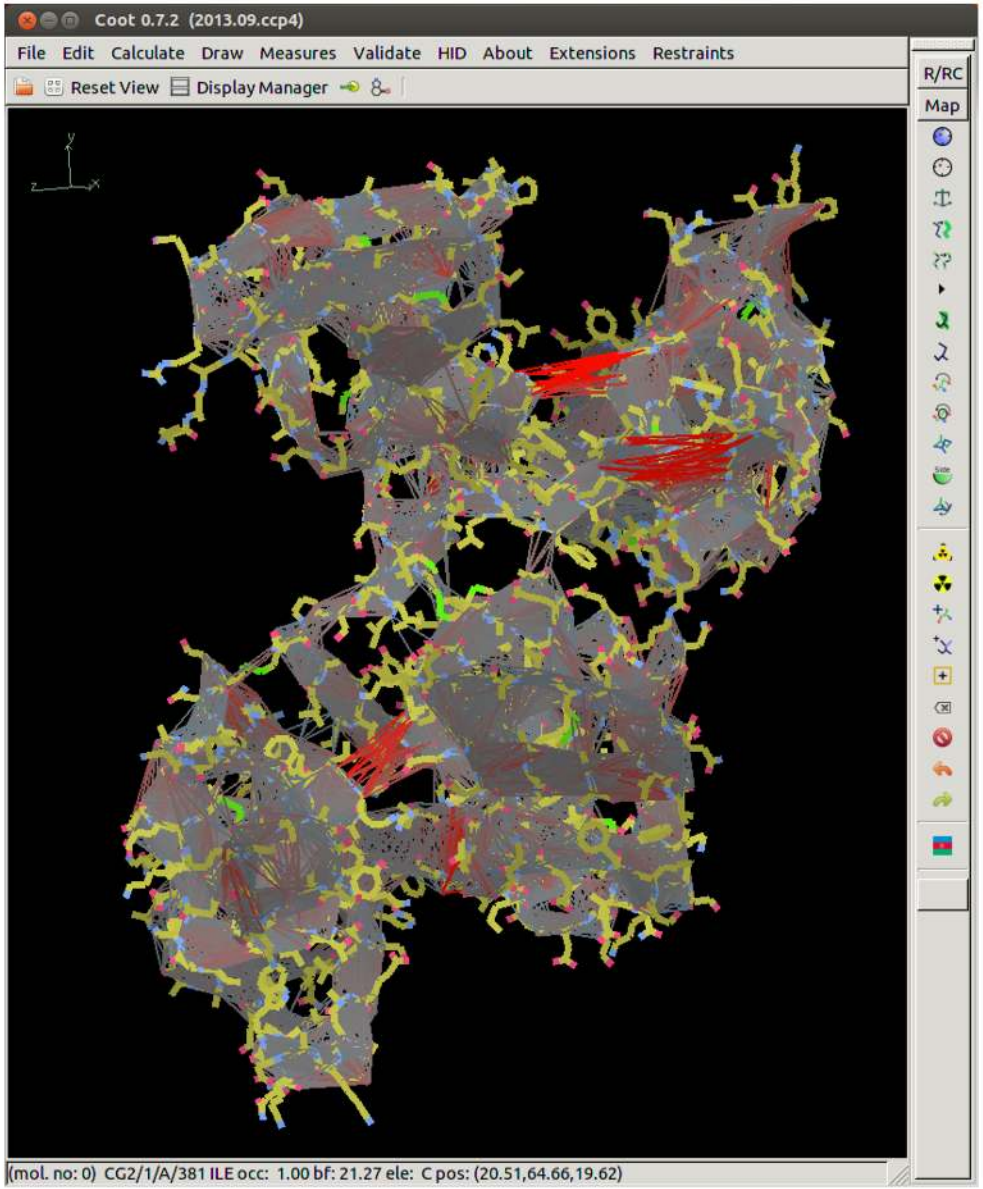
Ovotransferrin

1ryx (3.5Å)

Thanks to Paul Emsley

ProSMART Restraint Visualisation in Coot

Backbone
Restraints



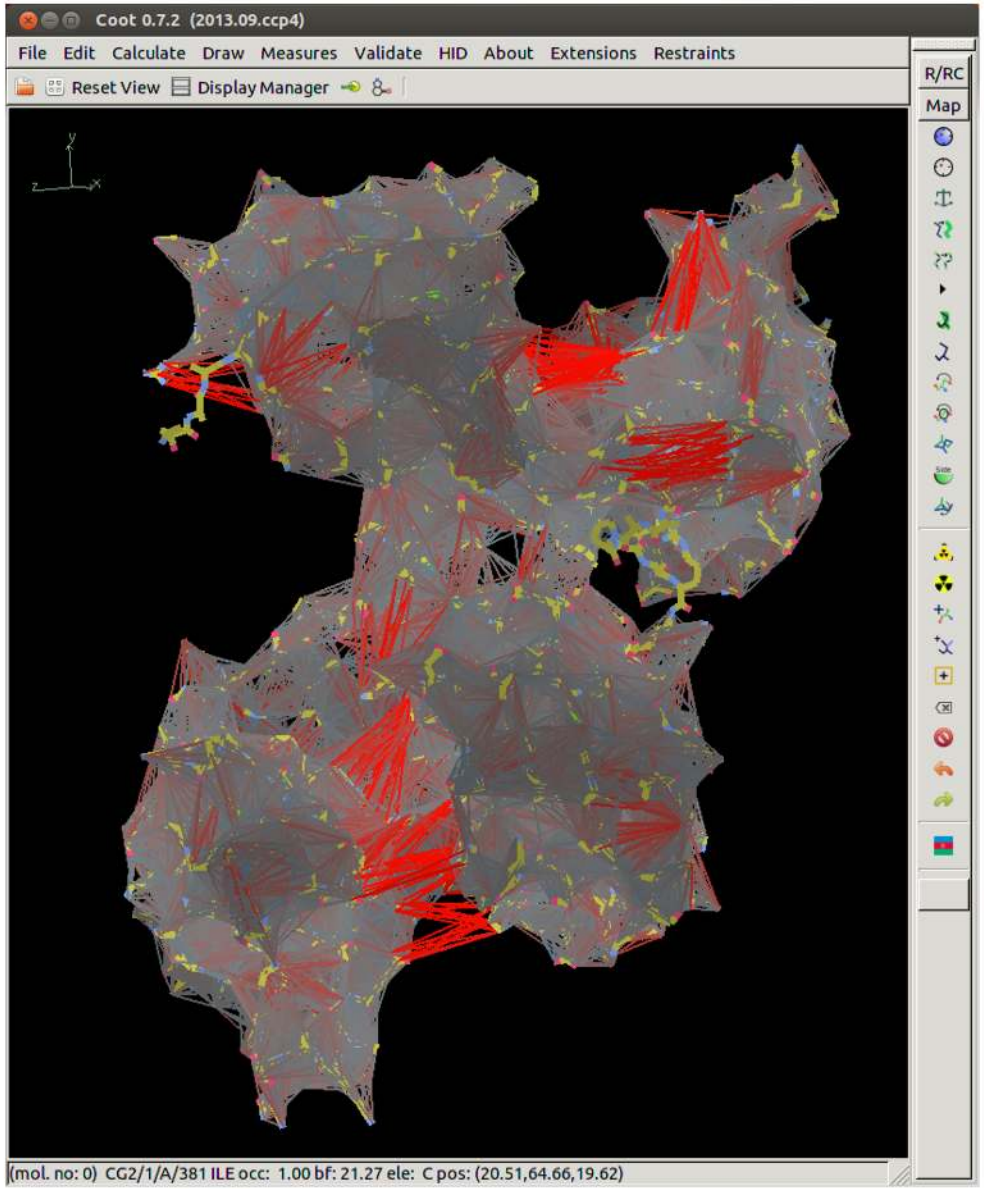
Ovotransferrin

1ryx (3.5Å)
restrained to
2d3i (2.15Å)

Thanks to Paul Emsley

ProSMART Restraint Visualisation in Coot

Backbone
& Side Chain
Restraints



Ovotransferrin

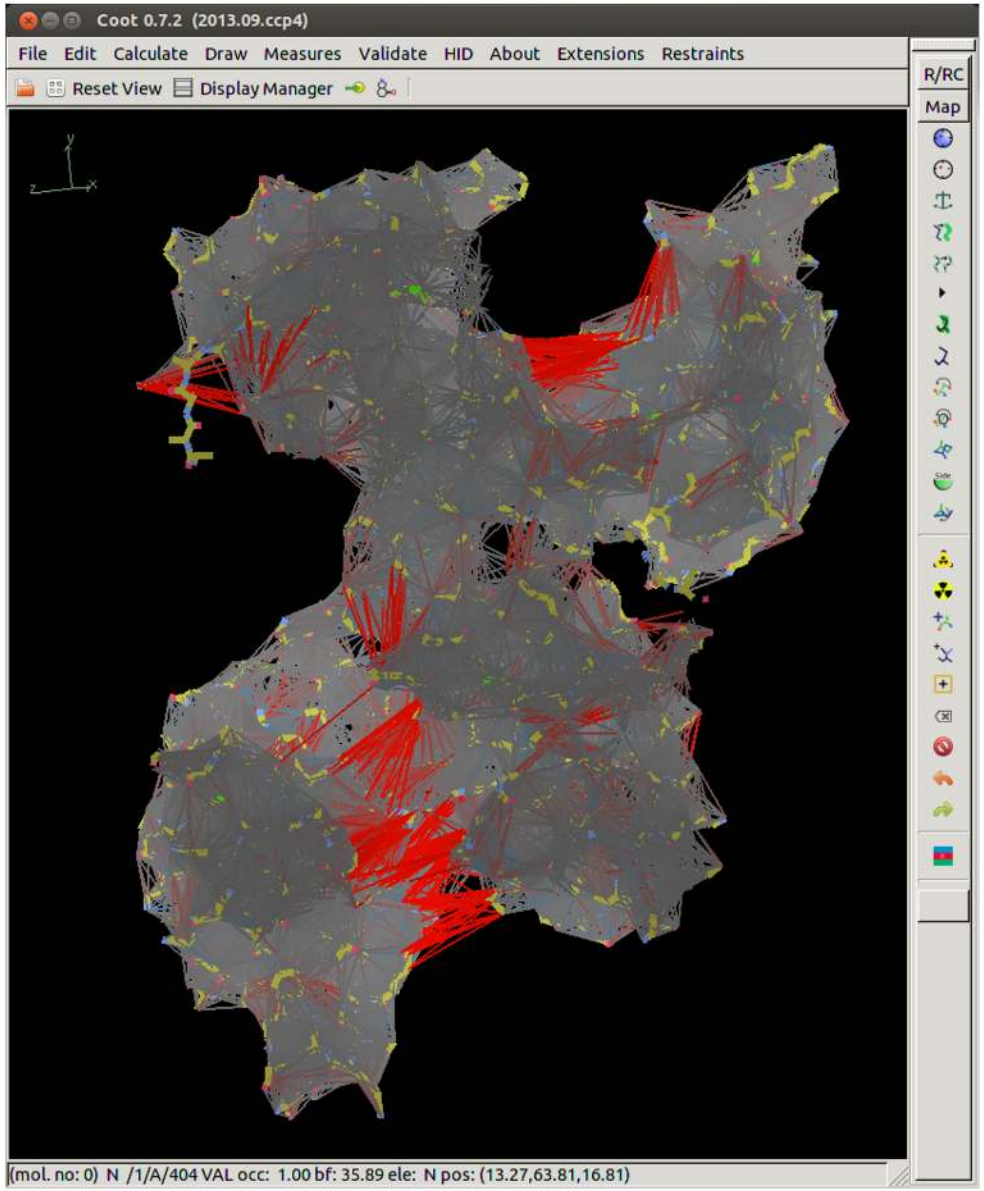
1ryx (3.5Å)
restrained to
2d3i (2.15Å)

Thanks to Paul Emsley

ProSMART Restraint Visualisation in Coot

Backbone
& Side Chain
Restraints

Regenerated after
re-refinement



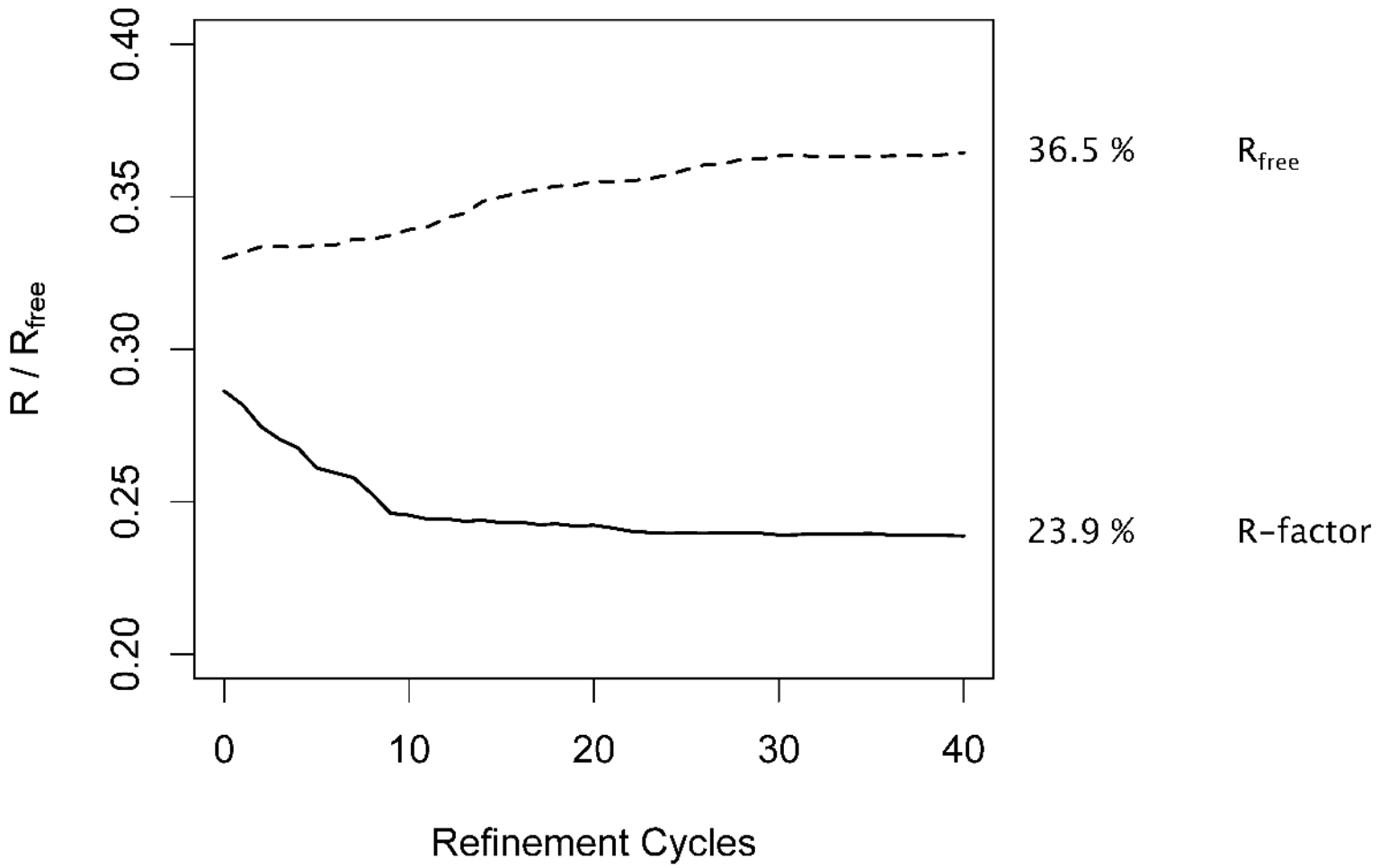
Ovotransferrin

1ryx re-refined
restrained to
2d3i (2.15Å)

Thanks to Paul Emsley

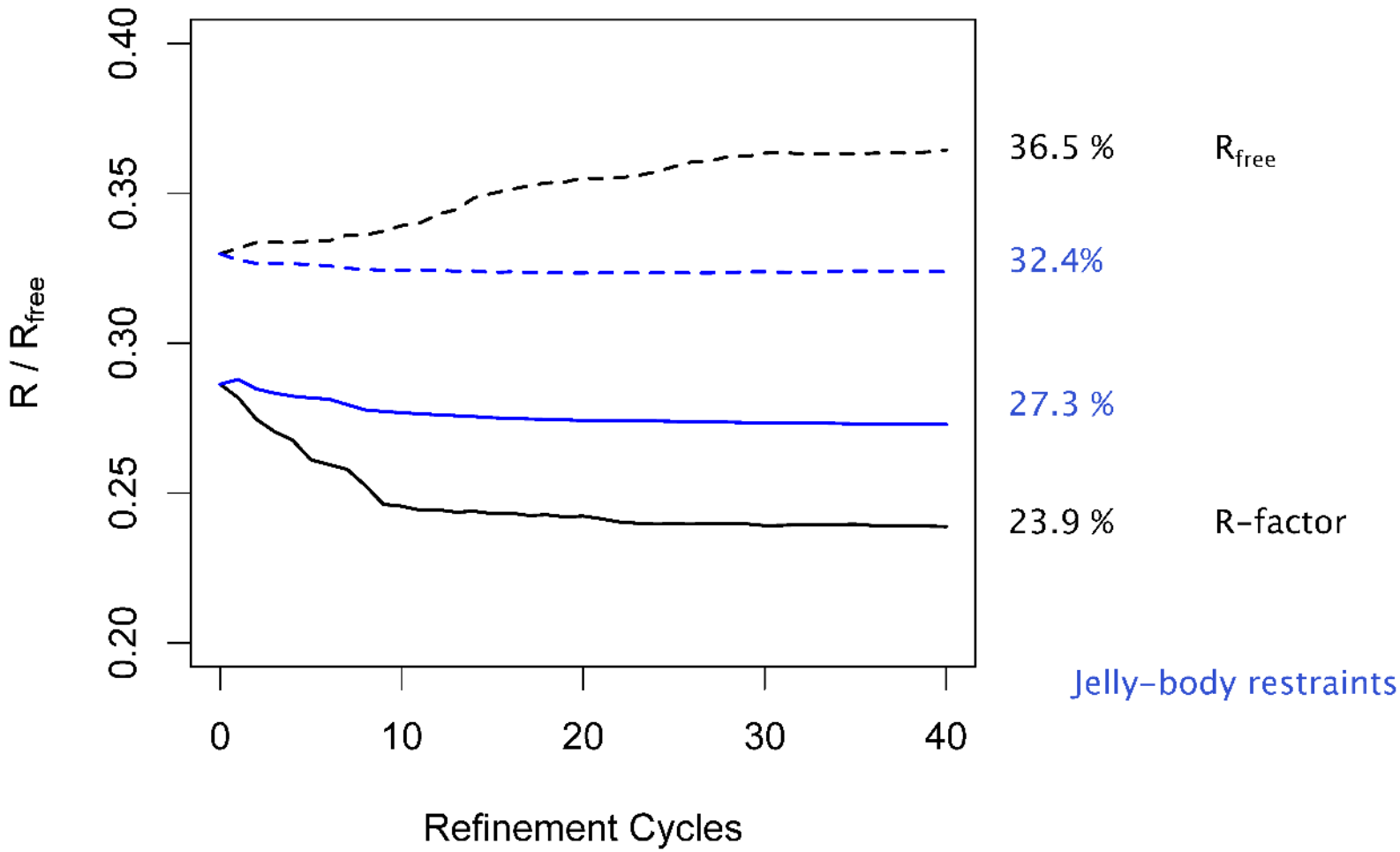
External Restraints

Ovotransferrin



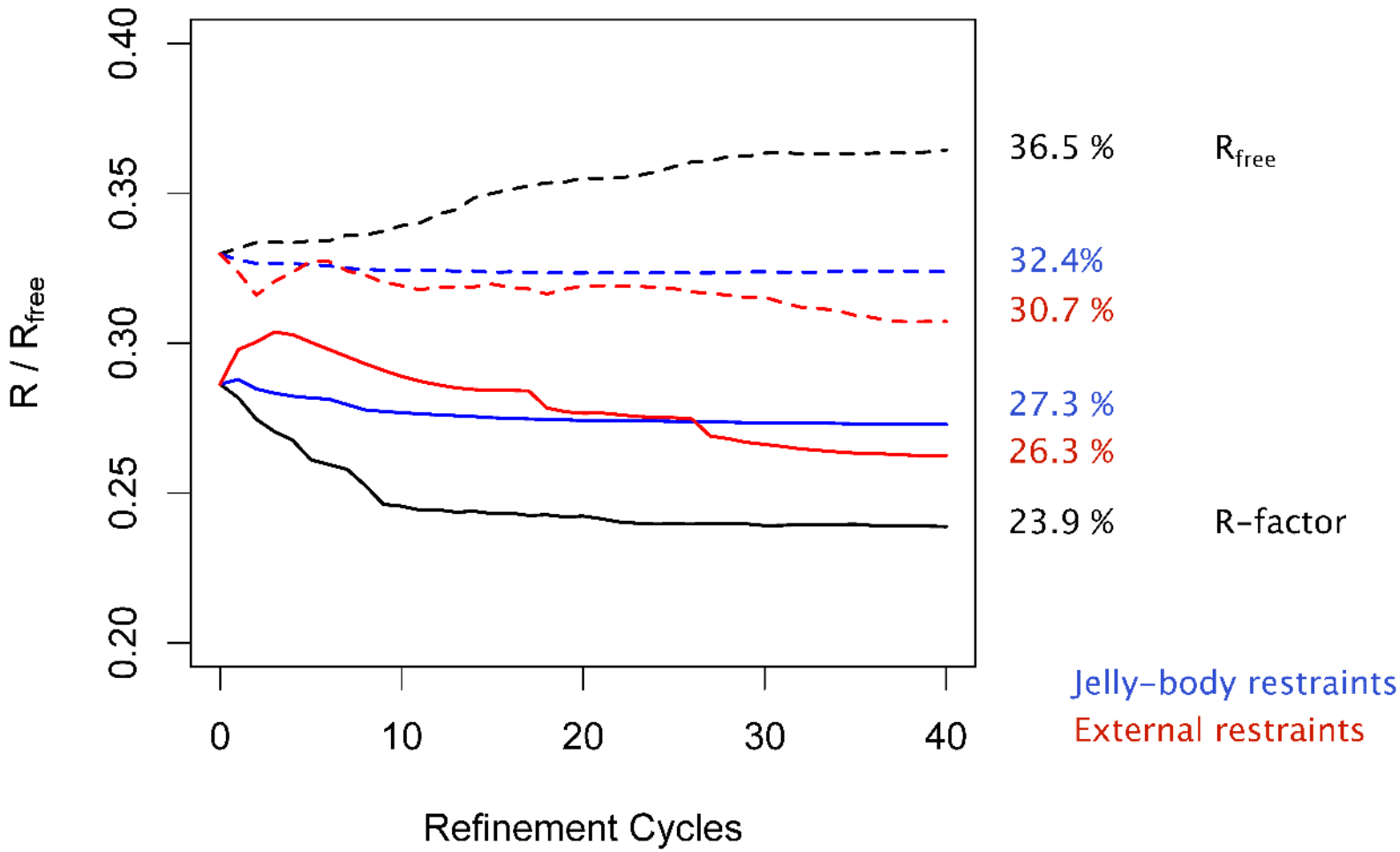
External Restraints

Ovotransferrin



External Restraints

Ovotransferrin

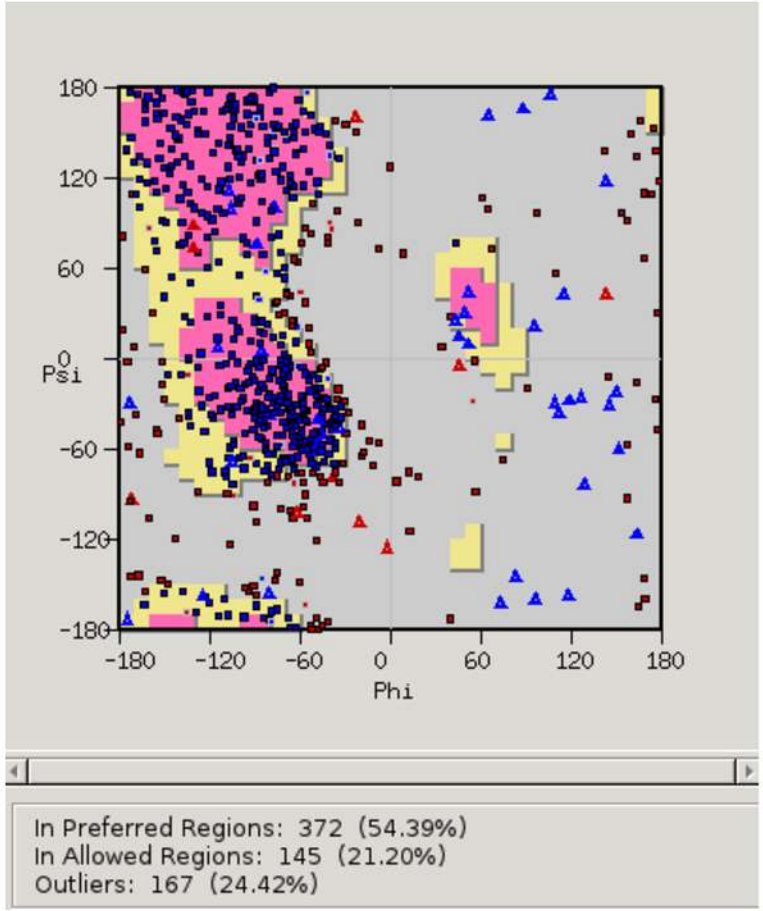


External Restraints

Ovotransferrin

Original Structure

R/R_{free} : 0.286/0.330

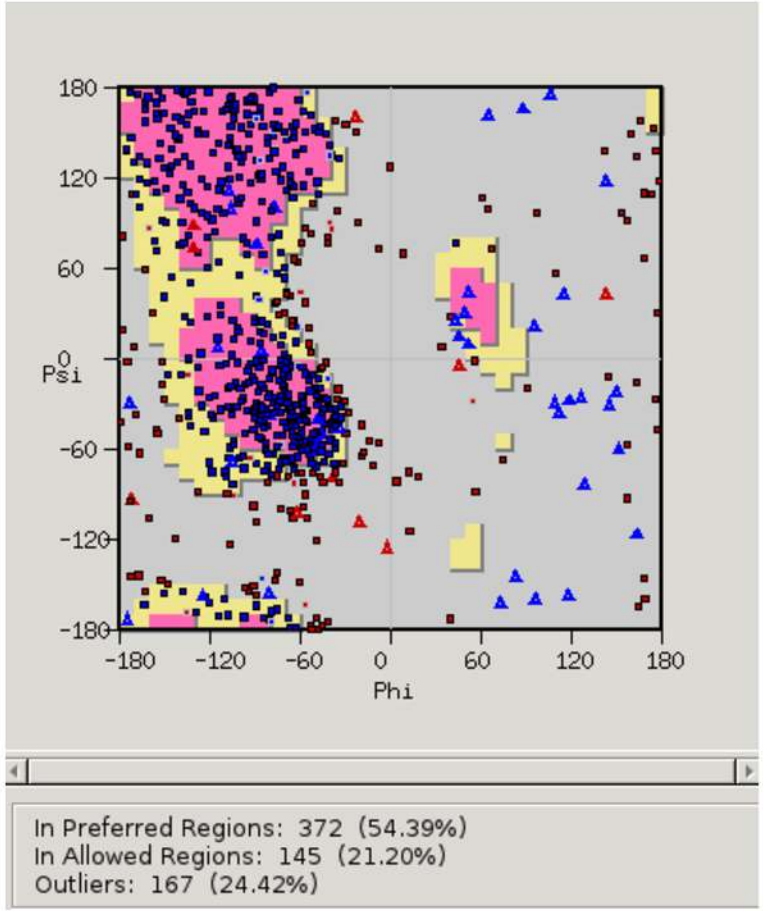


External Restraints

Ovotransferrin

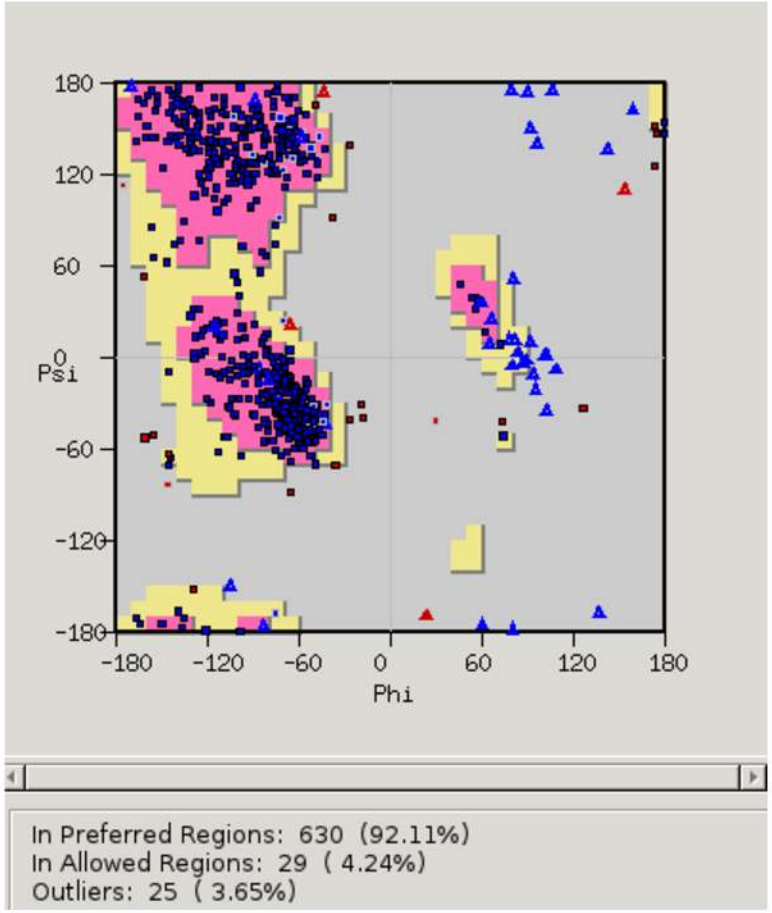
Original Structure

R/R_{free} : 0.286/0.330

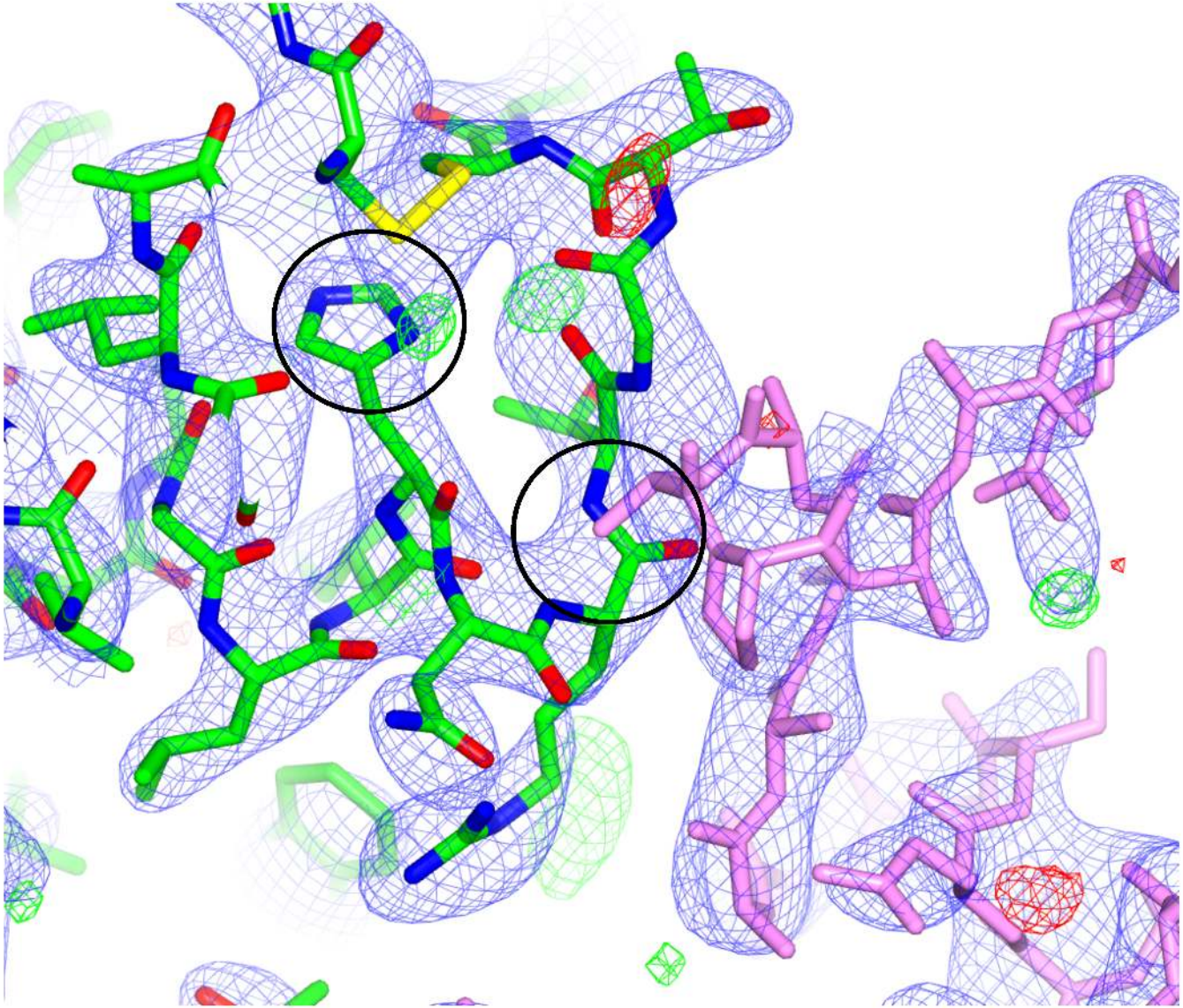


Re-refined with External Restraints

R/R_{free} : 0.263/0.307



External Restraints



1.3σ

Original Structure

R/R_{free} : 0.286/0.330

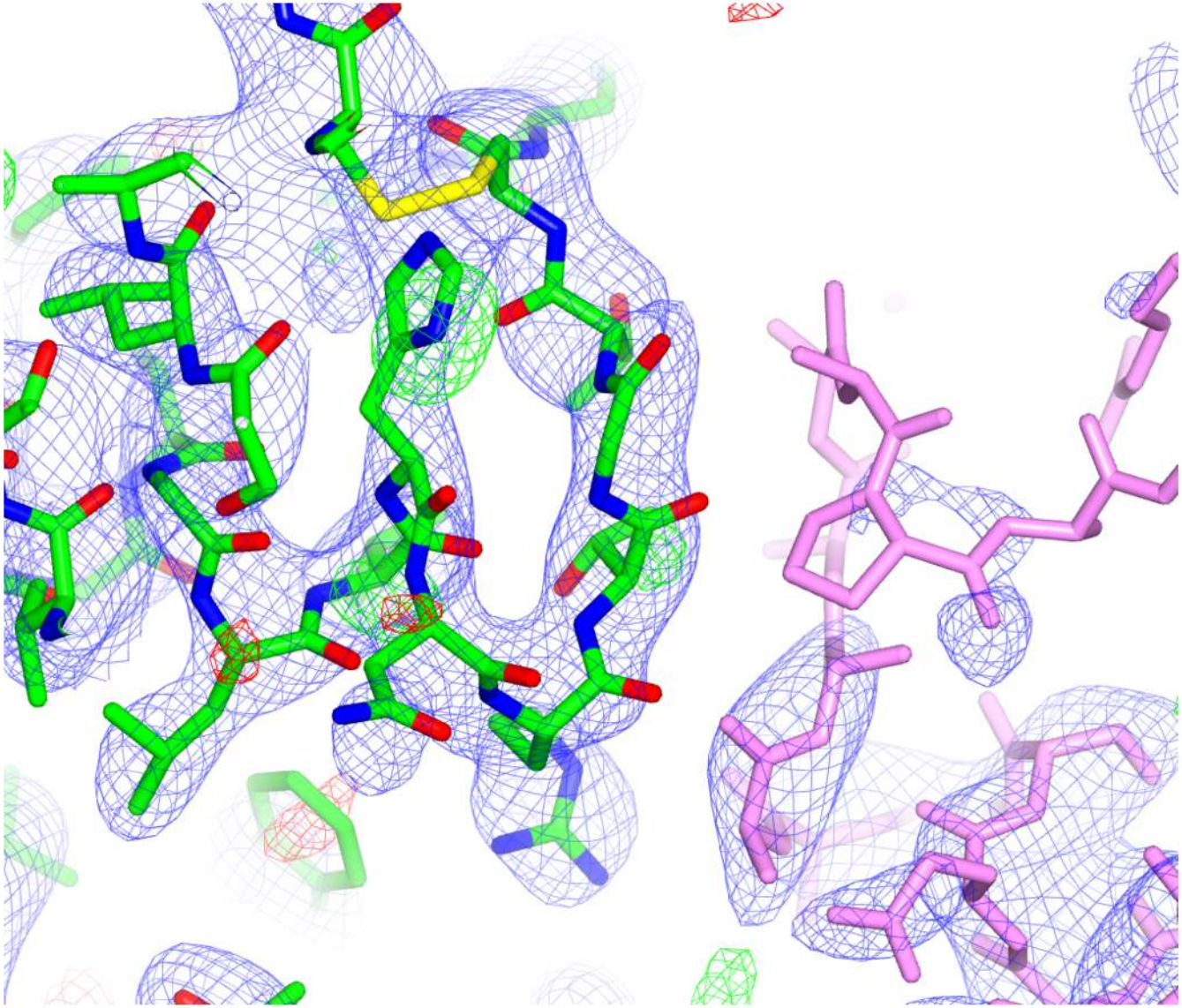
↓

External restraints

(40 cycles)

R/R_{free} : 0.263/0.307

External Restraints



1.3σ

Original Structure

R/R_{free} : 0.286/0.330

↓

External restraints

(40 cycles)

R/R_{free} : 0.263/0.307

↓

Modify

Real Space Refine

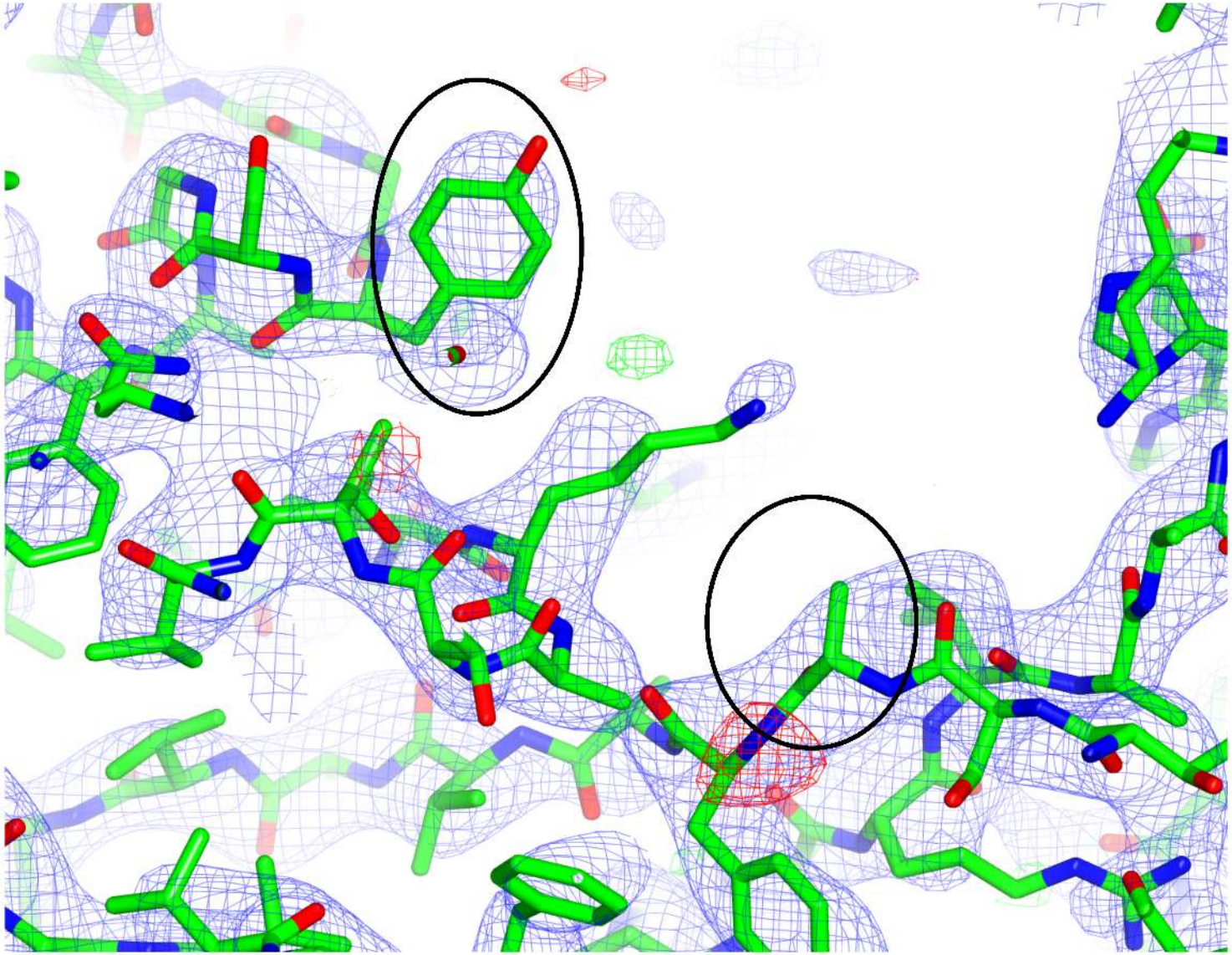
↓

Jelly body

(40 cycles)

R/R_{free} : 0.253/0.304

External Restraints



1.3σ

Original Structure

R/R_{free} : 0.286/0.330

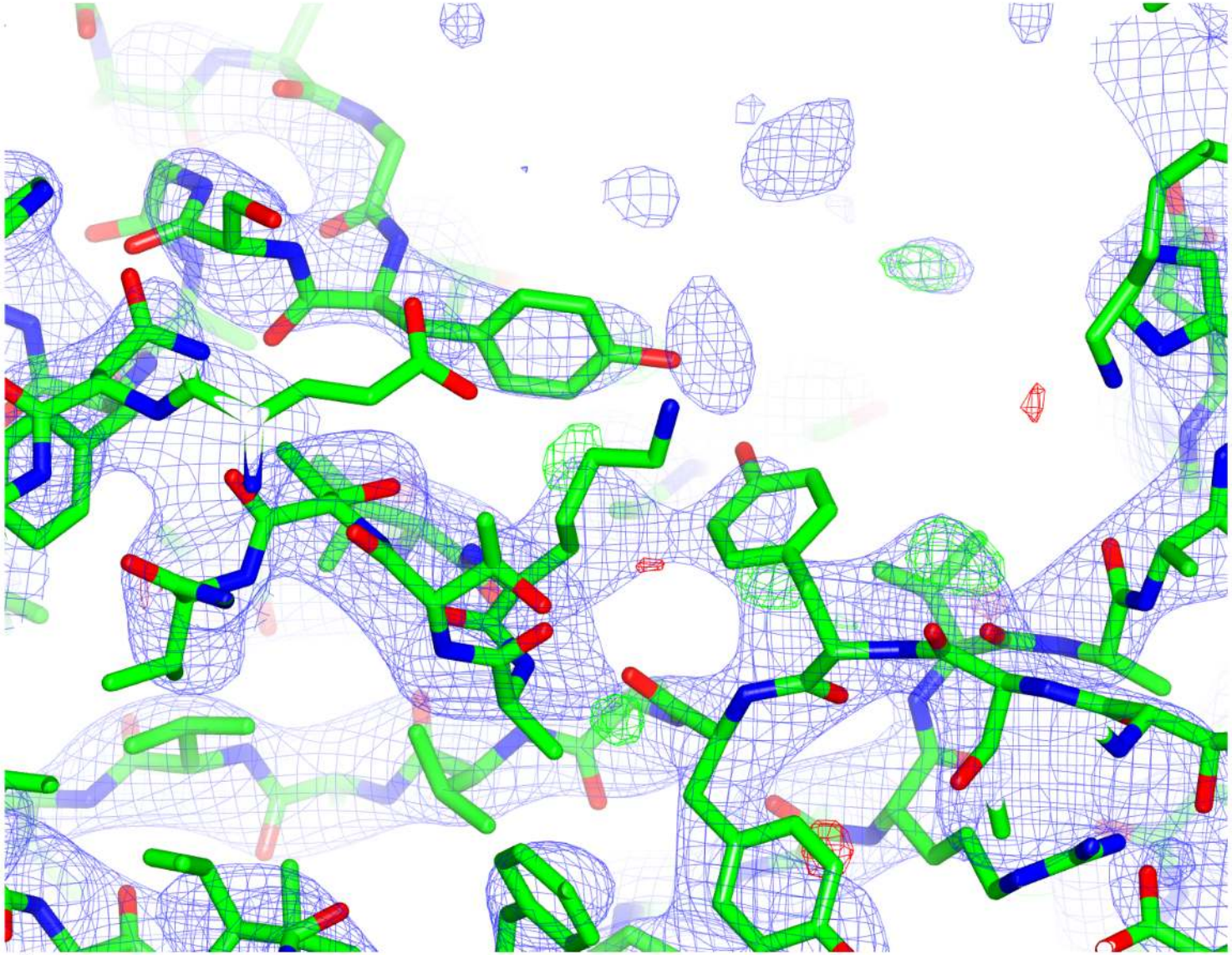
↓

External restraints

(40 cycles)

R/R_{free} : 0.263/0.307

External Restraints



1.3σ

Original Structure

R/R_{free} : 0.286/0.330

↓

External restraints

(40 cycles)

R/R_{free} : 0.263/0.307

↓

Build TYR92

Modify LYS209

↓

Jelly body

(40 cycles)

R/R_{free} : 0.252/0.307

External Restraints

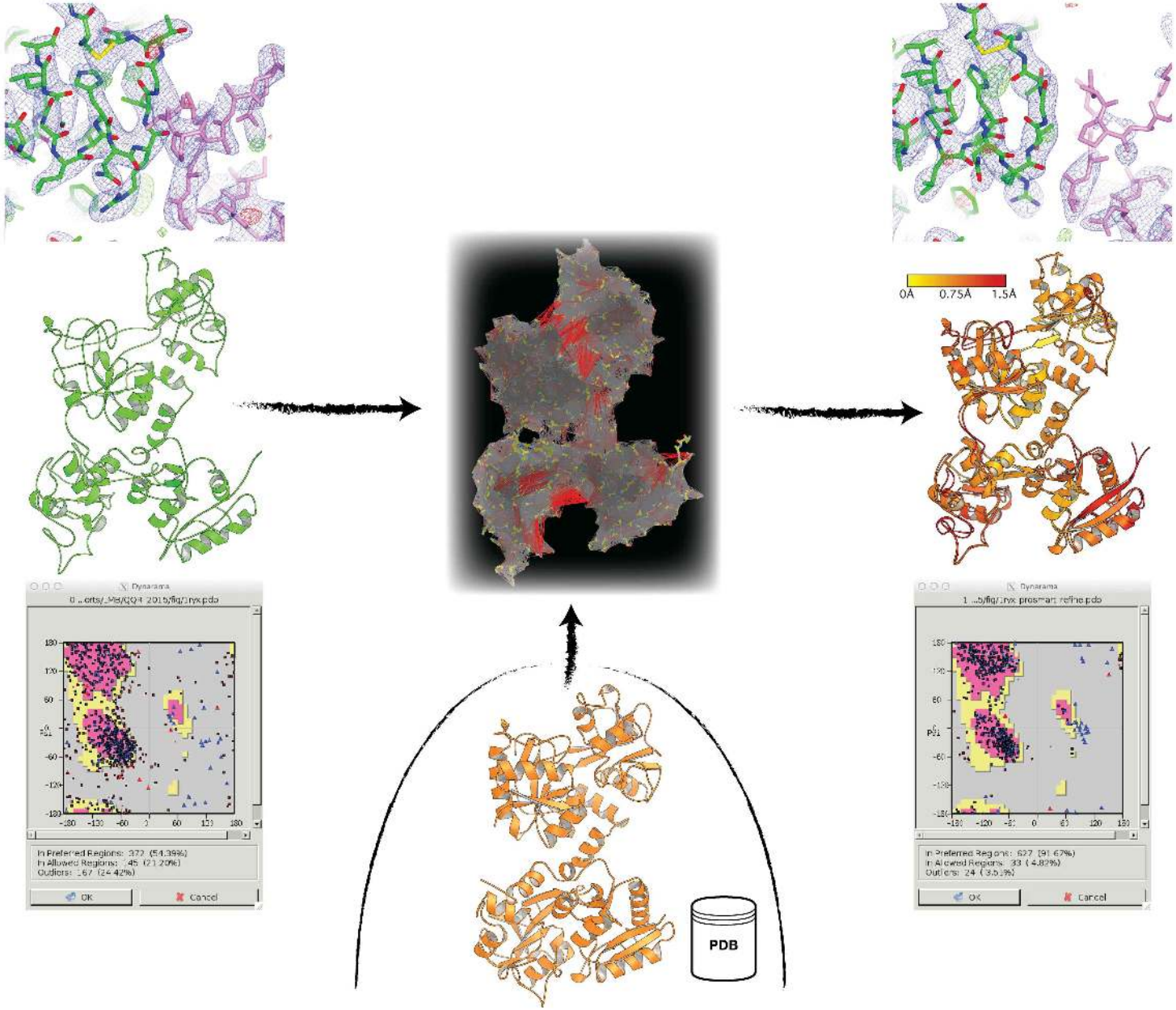
When refining at low resolution, check:

- Refinement statistics – ***Not always conclusive***
- Geometry – ***Not always conclusive***
- Electron density – ***Not always reliable***

Conclusion: **At low resolution, everything has to add up!**
Take care; reflect

Quality of prior information is important – **consider manual re-refinement**
– **PDB_REDO is useful**

Low Resolution Refinement – Recap



What if there are no high-resolution homologues?

We still need to stabilise refinement…

- Jelly-body restraints

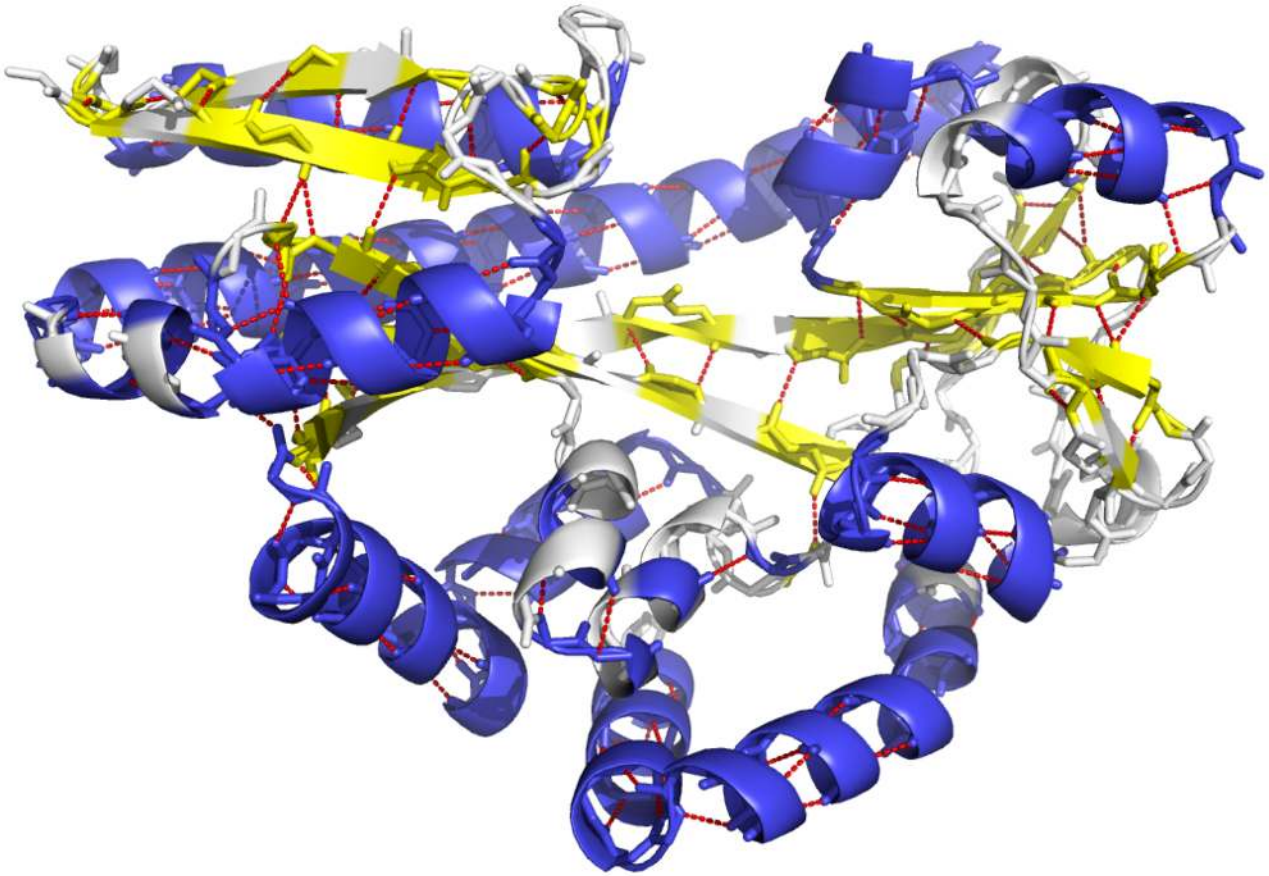
- Generic external restraints:

  - ProSMART – protein (secondary-structure)

  - LIBG – DNA/RNA (base-pair, base-stacking)

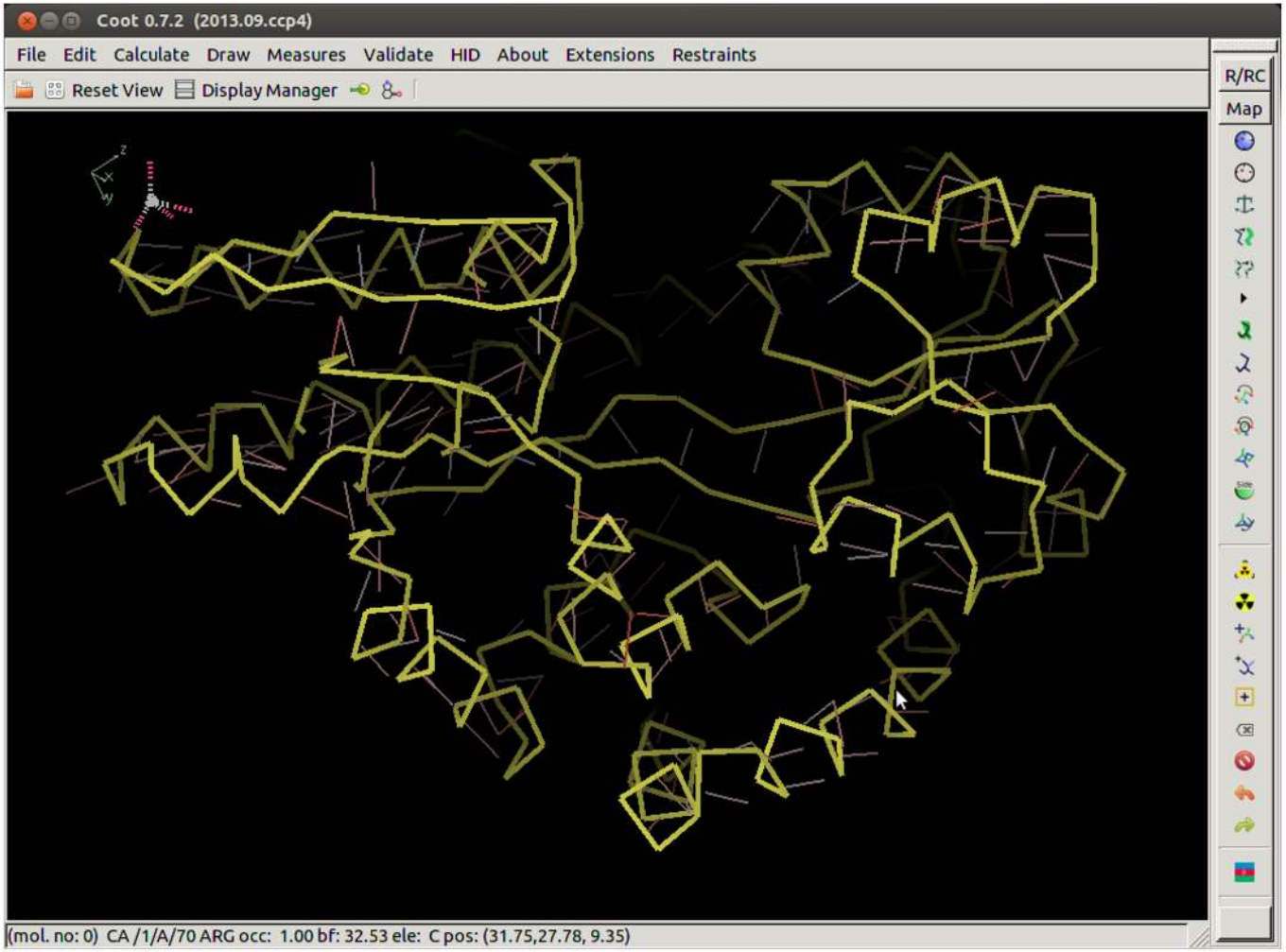
Hydrogen Bond Restraints

Visualisation of ProSMART hydrogen bond restraints using PyMOL:



Hydrogen Bond Restraints

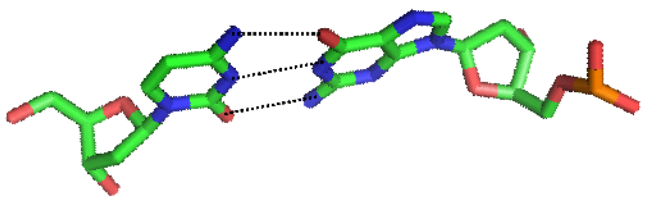
Visualisation of ProSMART hydrogen bond restraints using Coot:



LIBG Restraints for DNA/RNA

LIBG – for the generation of nucleic acid restraints

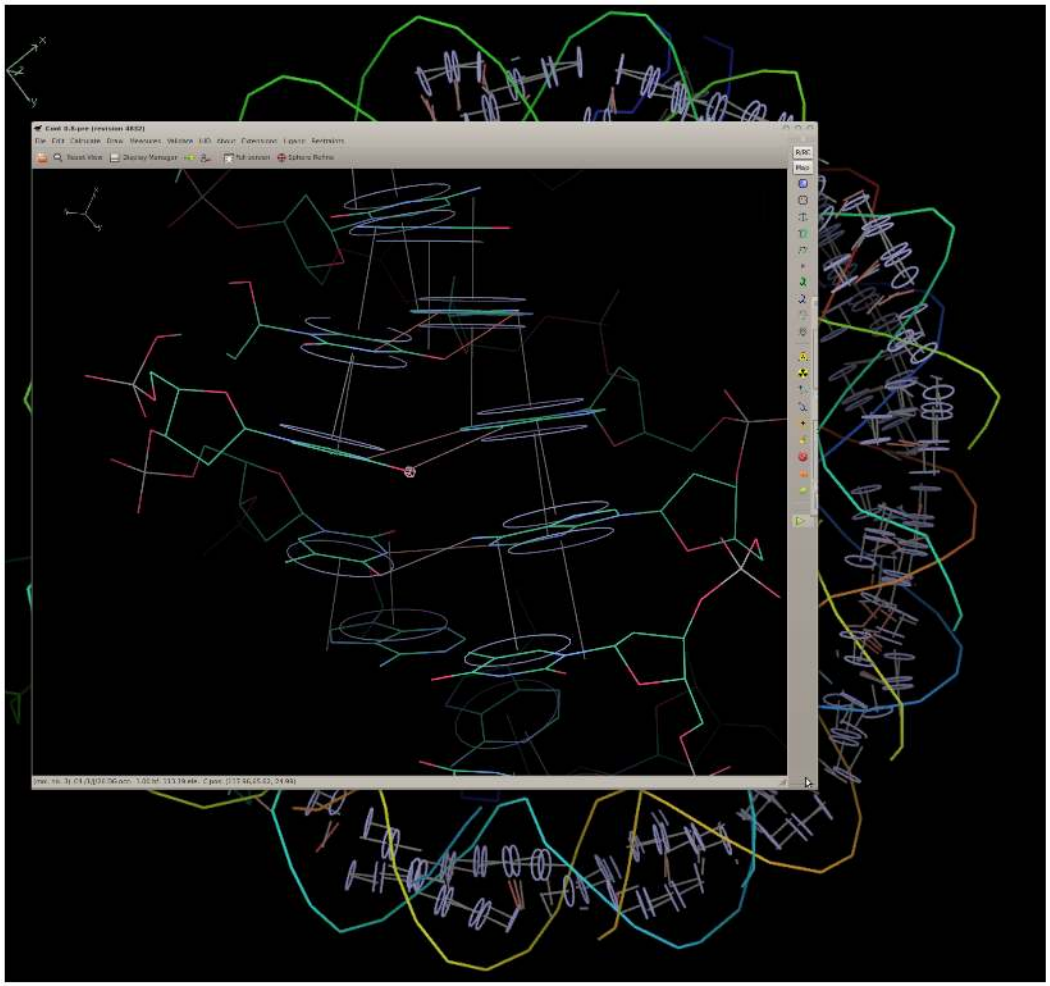
Base-pair restraints:



LIBG Restraints for DNA/RNA

LIBG – for the generation of nucleic acid restraints

Base-stacking restraints:
(parallel plane restraints)



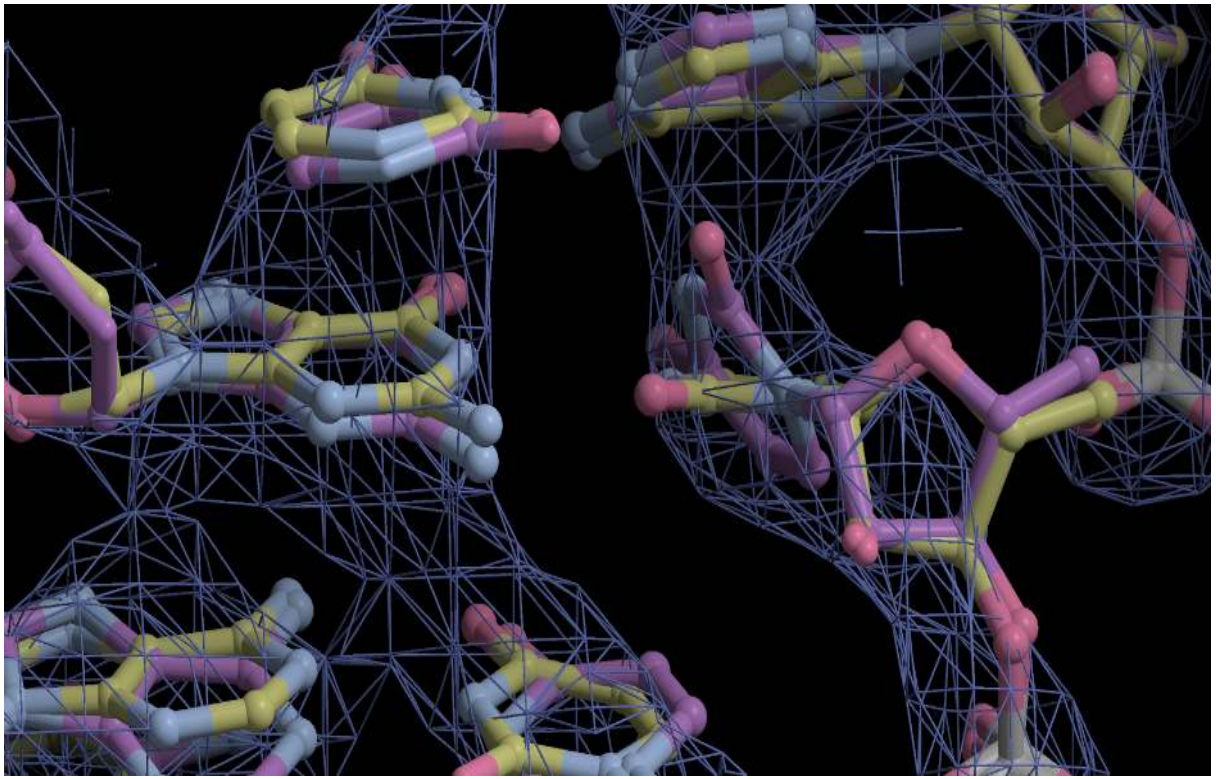
LIBG Restraints for DNA/RNA

LIBG – for the generation of nucleic acid restraints

Base-stacking restraints:
(parallel plane restraints)

Purple – refined without LIBG
Yellow – refined with LIBG

Example:
3.2Å cryo-EM



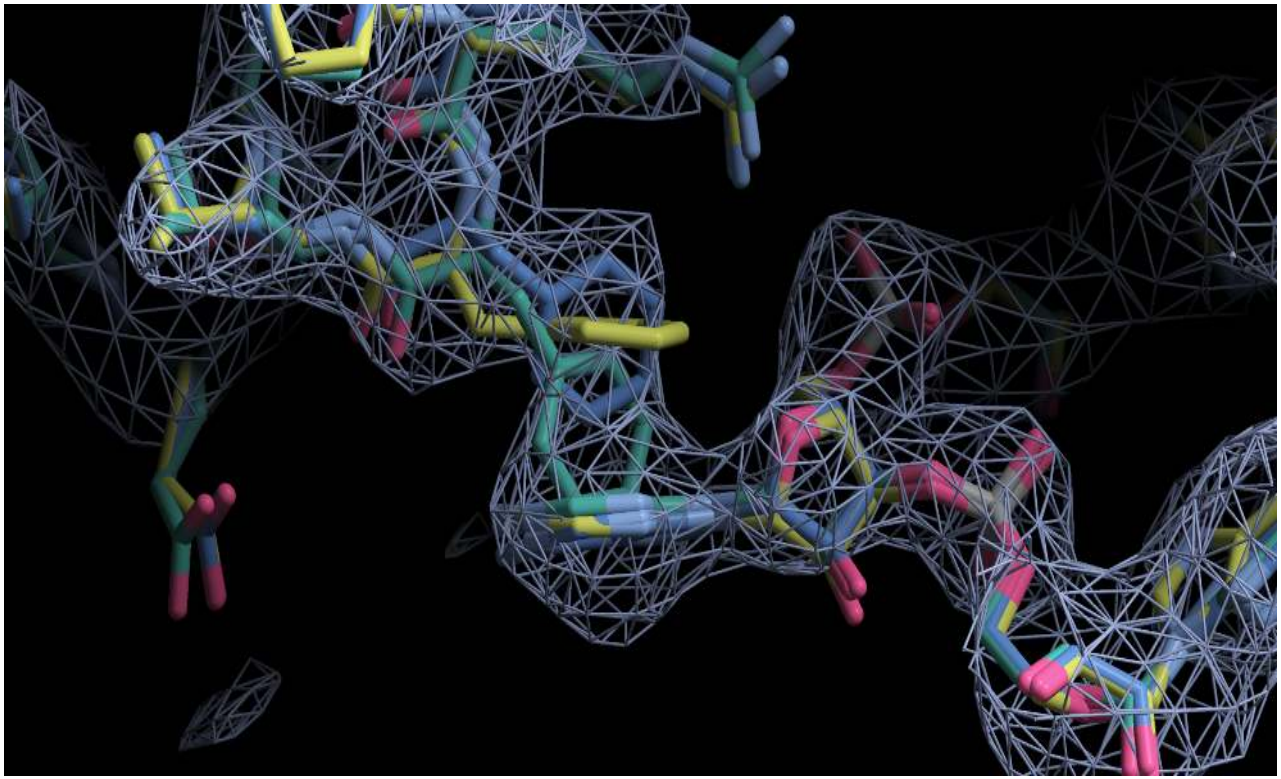
LIBG Restraints for DNA/RNA

LIBG – for the generation of nucleic acid restraints

Base-stacking restraints:
(parallel plane restraints)

Green – before refinement
Blue – refined without LIBG
Yellow – refined with LIBG

Example:
3.2Å cryo-EM



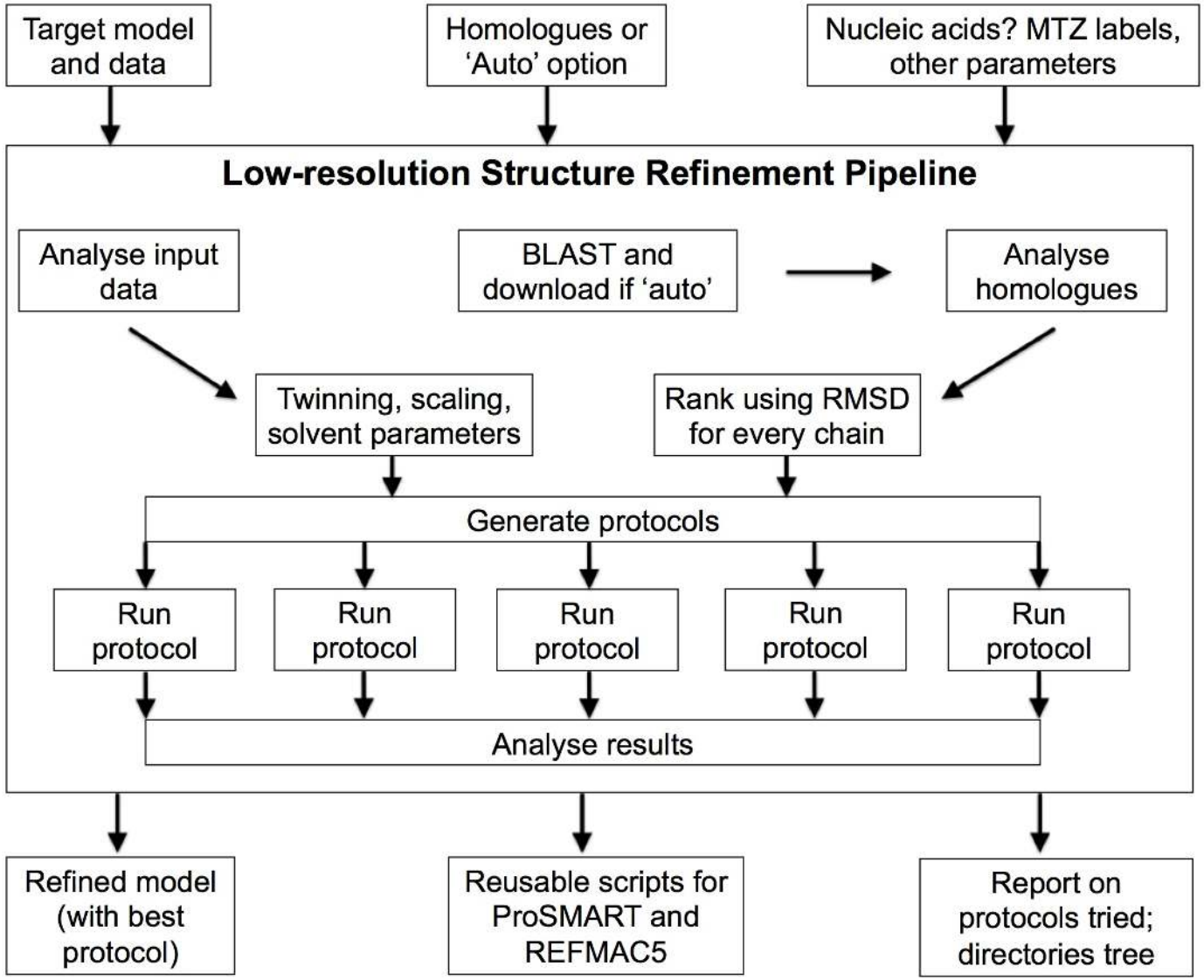
Automated pipeline - LORESTR

- Efficiency of ProSMART-generated restraints greatly depends on the homologues used
- If several homologues are available, substantial manual effort is required to find their optimal combination
- Other refinement parameters (scaling, solvent, etc) also affect efficiency of the process

Solution:

LOw-**RE**solution **ST**ructure **R**efinement

Automated pipeline - LORESTR



Automated pipeline - LORESTR

Target model and data

Homologues or 'Auto' option

Nucleic acids? MTZ labels, other parameters

```
Minimal run:
lrsr.py -p1 1.pdb -f 1.mtz

Manual supply of homologues:
lrsr.py -p1 1.pdb -f 1.mtz -p2 homologue1.pdb homologue2.pdb

Recommended - automated run with BLAST search:
lrsr.py -p1 1.pdb -f 1.mtz –auto

Automated run with BLAST search, saving disk space:
lrsr.py -p1 1.pdb -f 1.mtz –auto –save_space

Automated run with BLAST search, 4 CPUs:
lrsr.py -p1 1.pdb -f 1.mtz –auto –cpu 4
```

Refined model (with best protocol)

Reusable scripts for ProSMART and REFMAC5

Report on protocols tried; directories tree

Automated pipeline - LORESTR

Target model and data

Homologues or 'Auto' option

Nucleic acids? MTZ labels, other parameters

```
Minimal run:
lrsr.py -p1 1.pdb -f 1.mtz

Manual supply of homologues:
lrsr.py -p1 1.pdb -f 1.mtz -p2 h                                  ologue2.pdb

Recommended - automa                                  search:
lrsr.py -p1 1

Automated                 LAST search, saving disk space:
lrsr.py -p                pdb -f 1.mtz —auto —save_space

Automated run with BLAST search, 4 CPUs:
lrsr.py -p1 1.pdb -f 1.mtz —auto —cpu 4
```

Improves R-factors for 94% of the test structures from the PDB

Refined model (with best protocol)

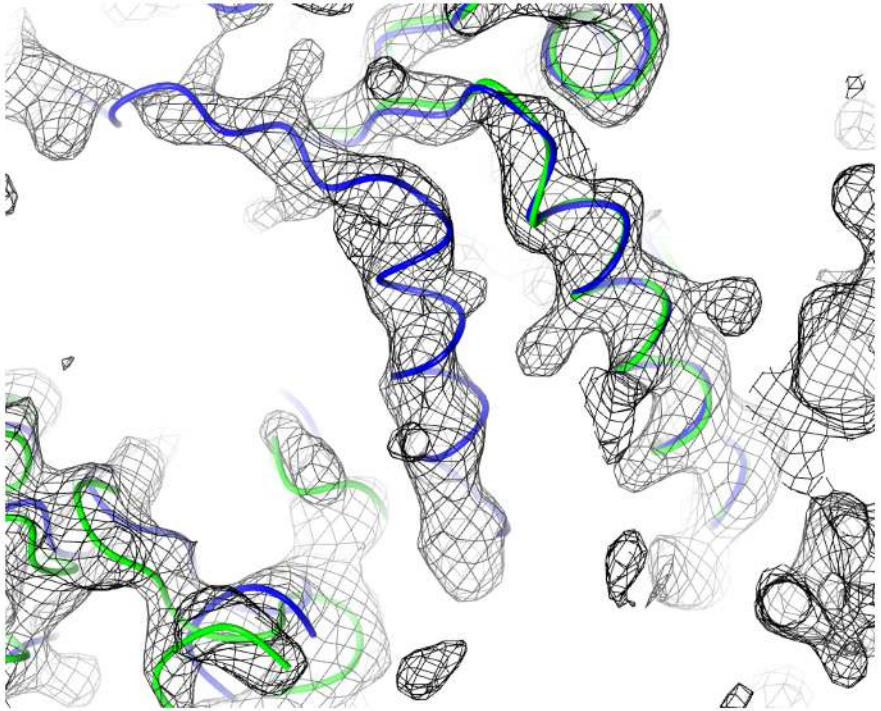
Reusable scripts for ProSMART and REFMAC5

Report on protocols tried; directories tree

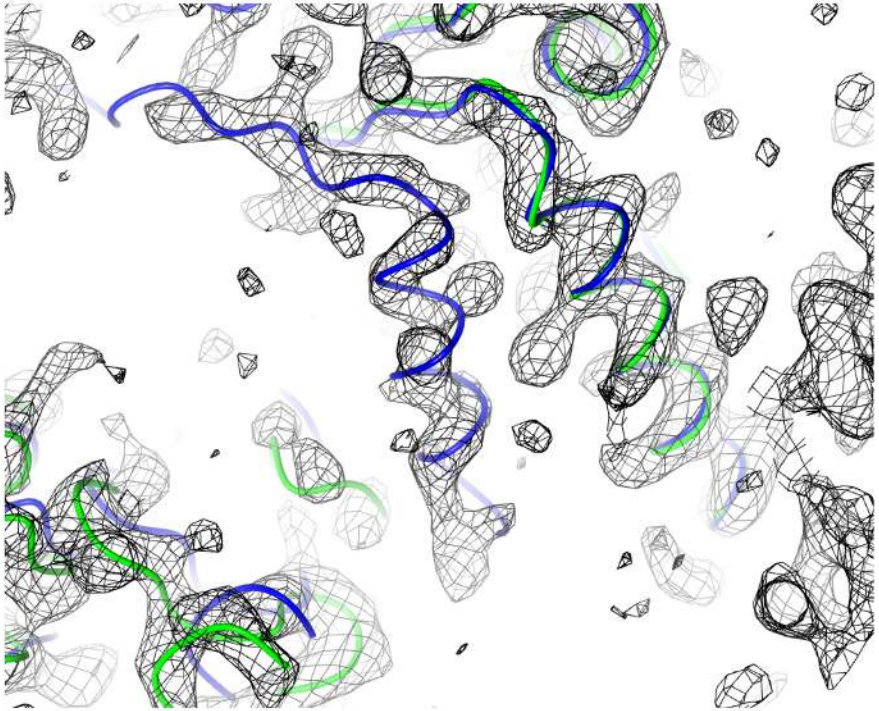
REFMAC Anisotropic Map Sharpening

Idea – remove an overall B value

Original Map



Sharpened map from REFMAC



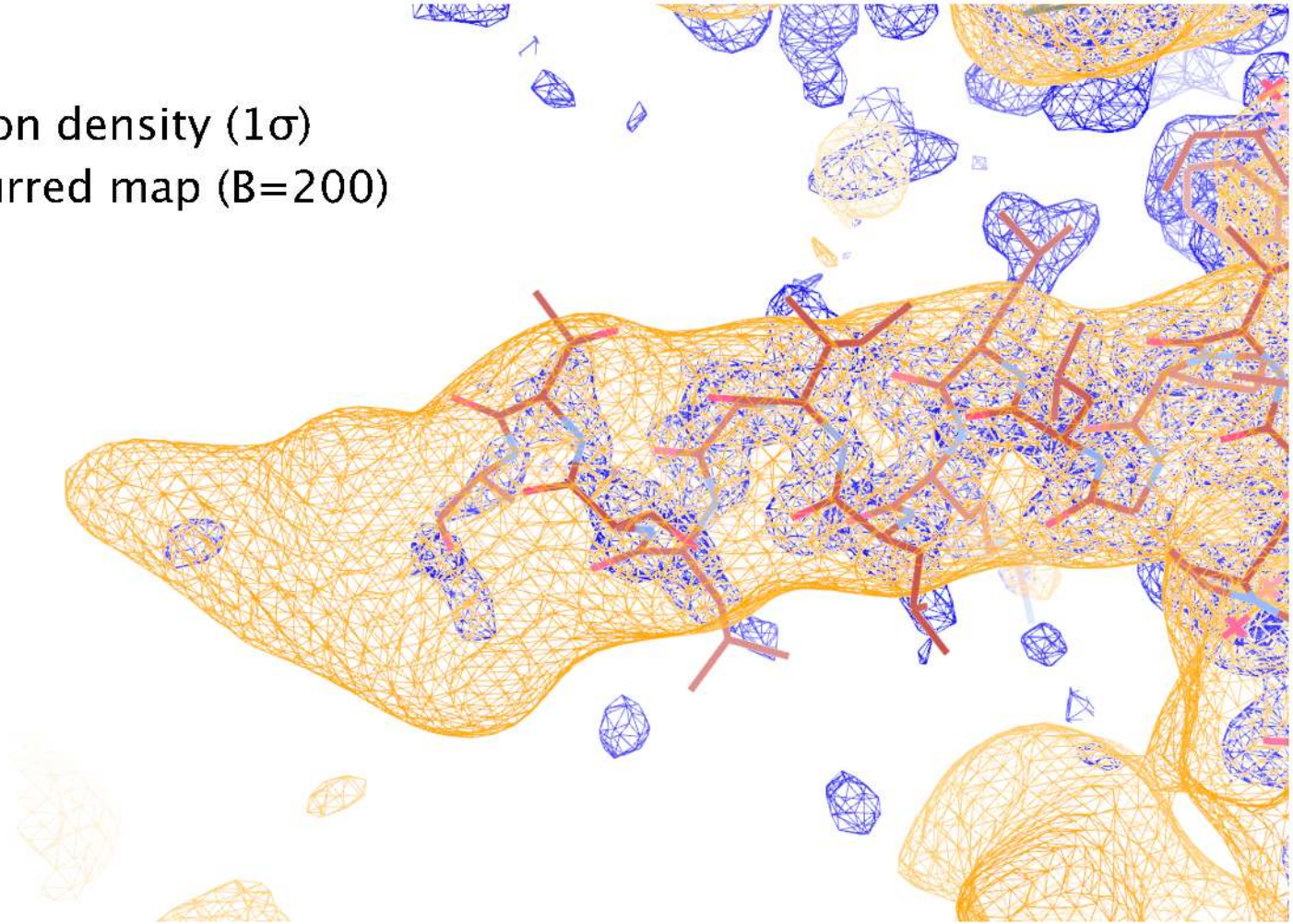
Green: original structure 2r6c (4.0Å) – helix unmodelled

Blue: homologous structure 2r6a (2.9Å)

Map Blurring

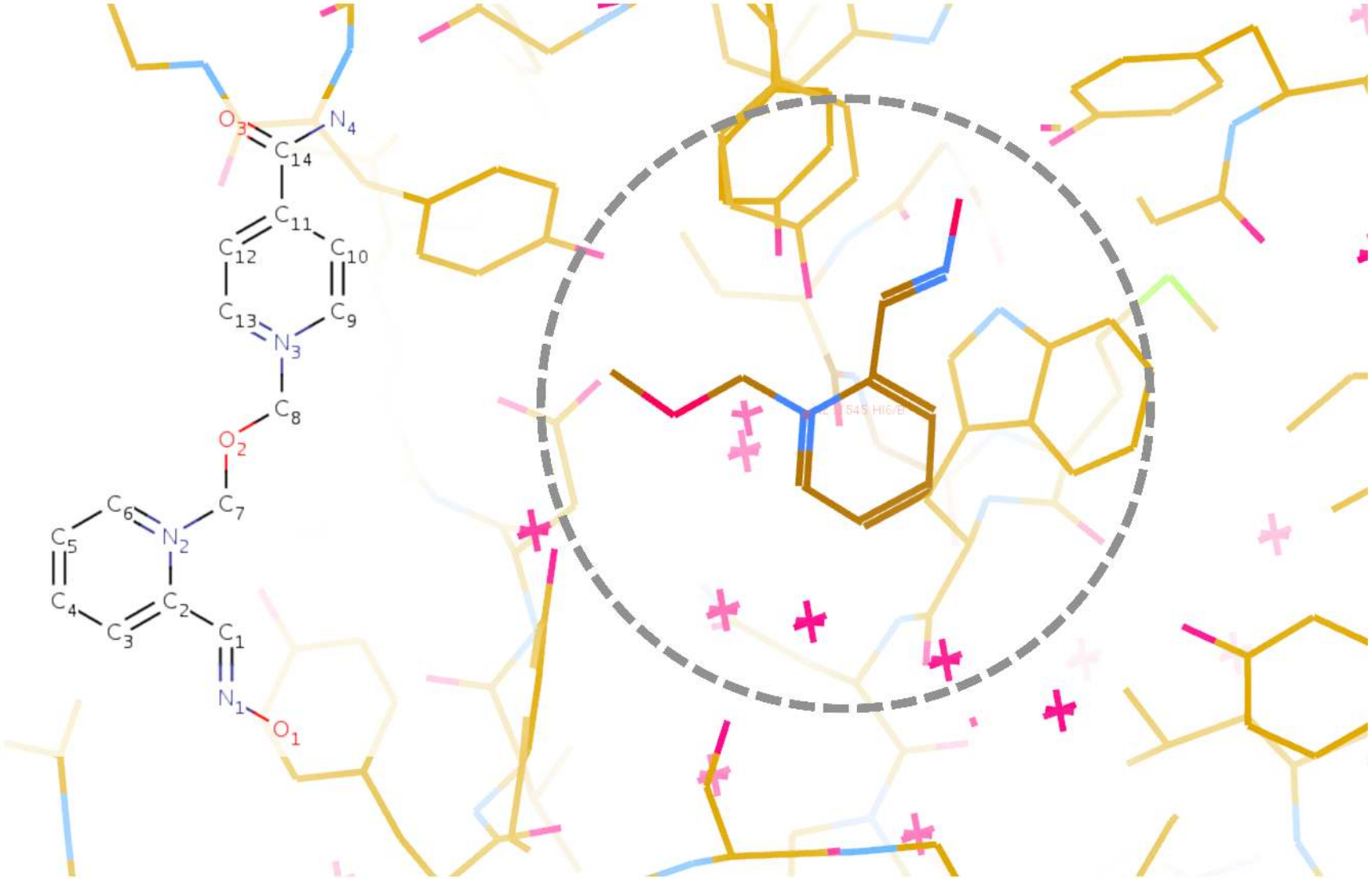
Idea – apply an overall B value

Blue – electron density (1σ)
Orange – blurred map (B=200)



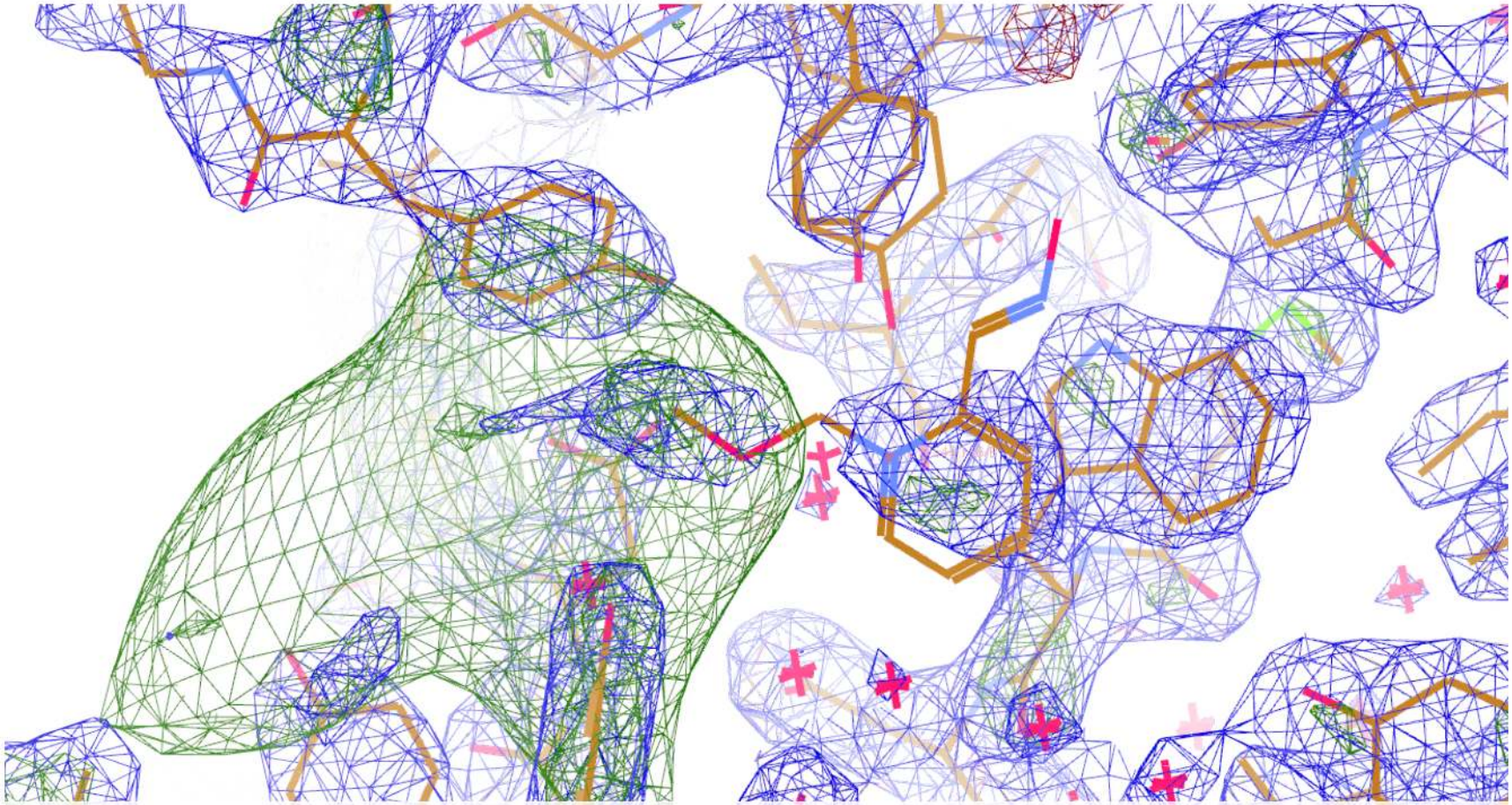
1s3e

Map Blurring



2whq

Map Blurring



Electron density at 1.7σ; Residual density at 3.5σ
Blurred residual map, B = 200

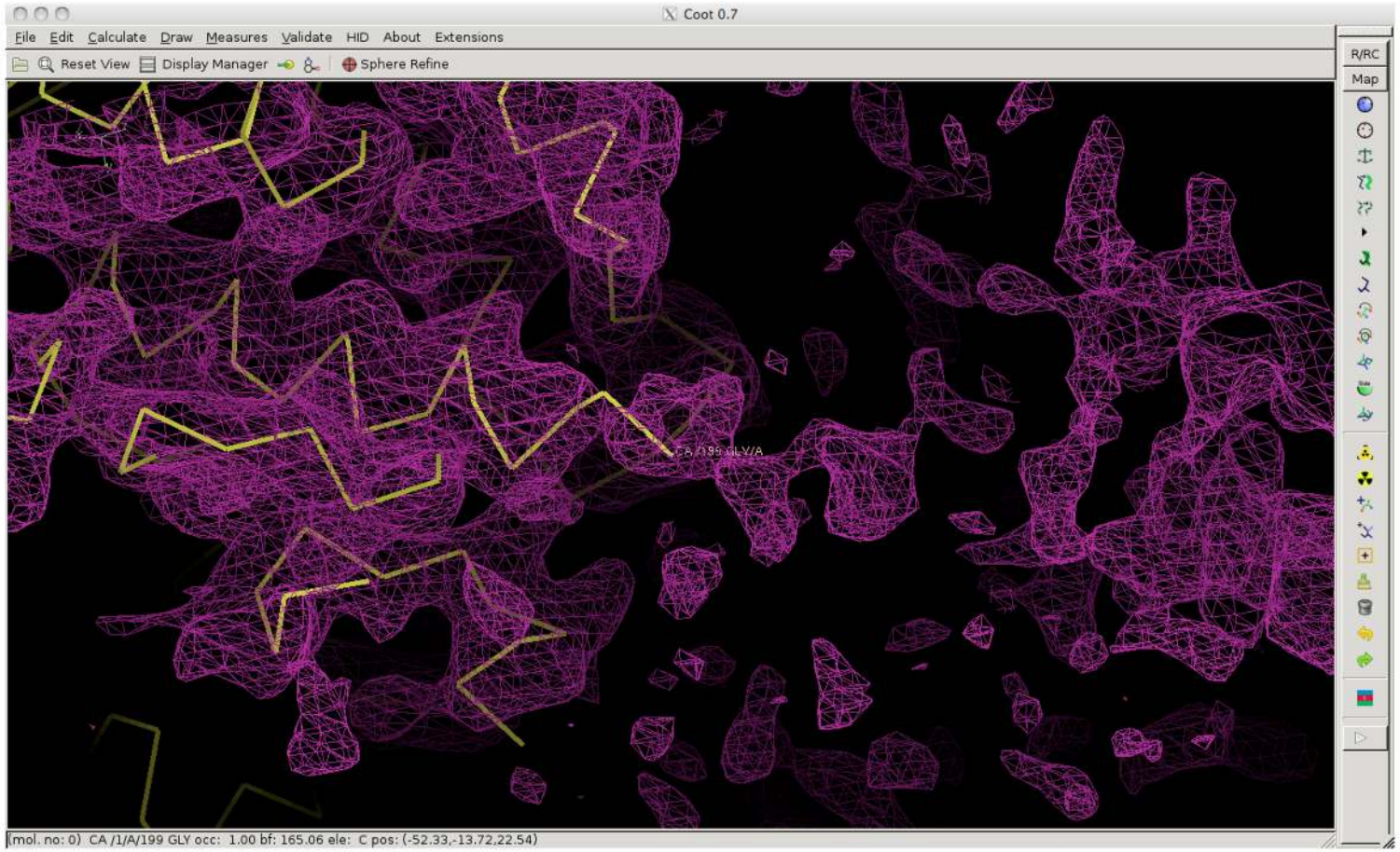
2whq

Map Sharpening/Blurring

Output Array of Sharpened Maps

Standard:

4iwo (2.6Å)

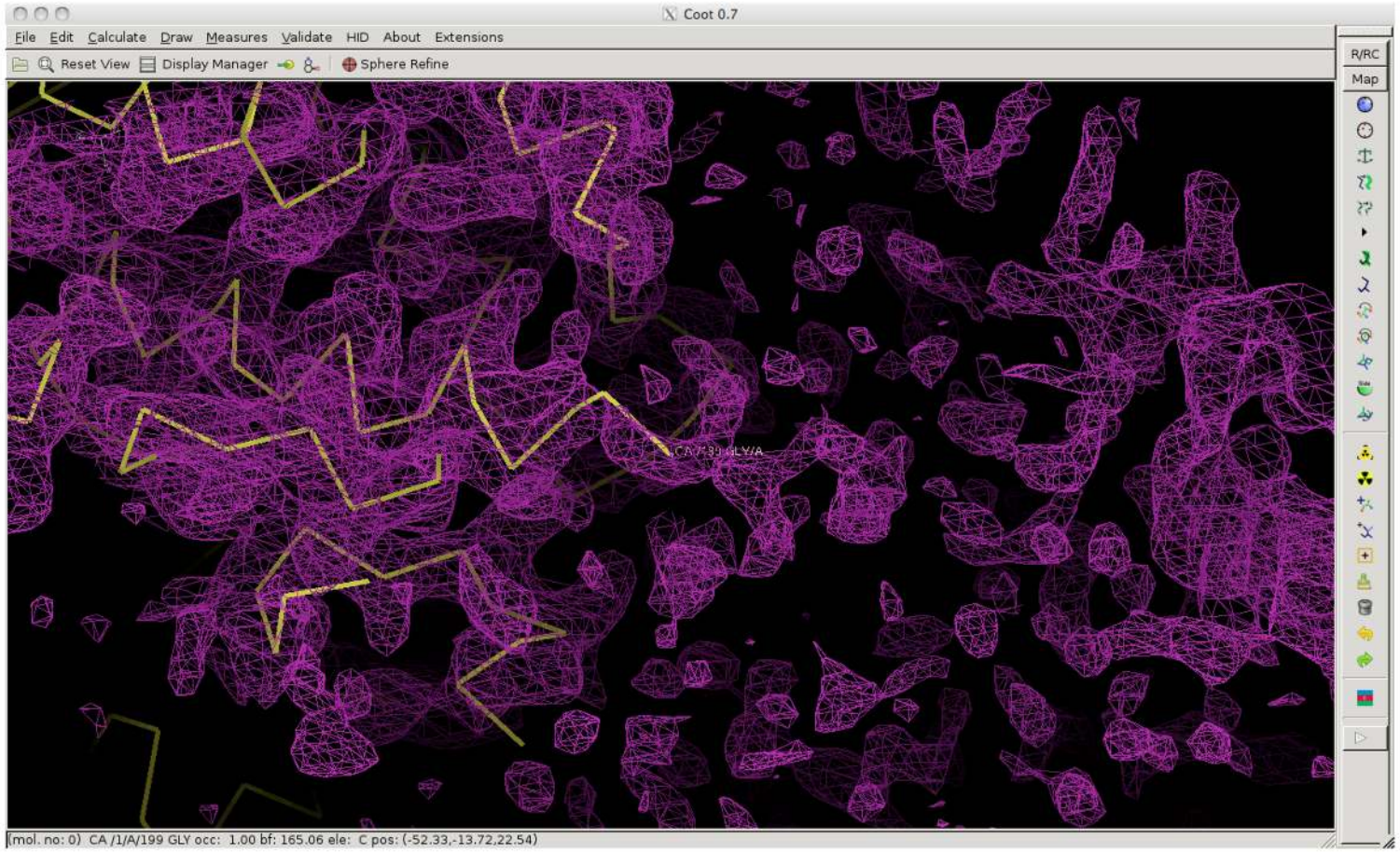


Map Sharpening/Blurring

Output Array of Sharpened Maps

Sharpen 50:

4iwo (2.6Å)

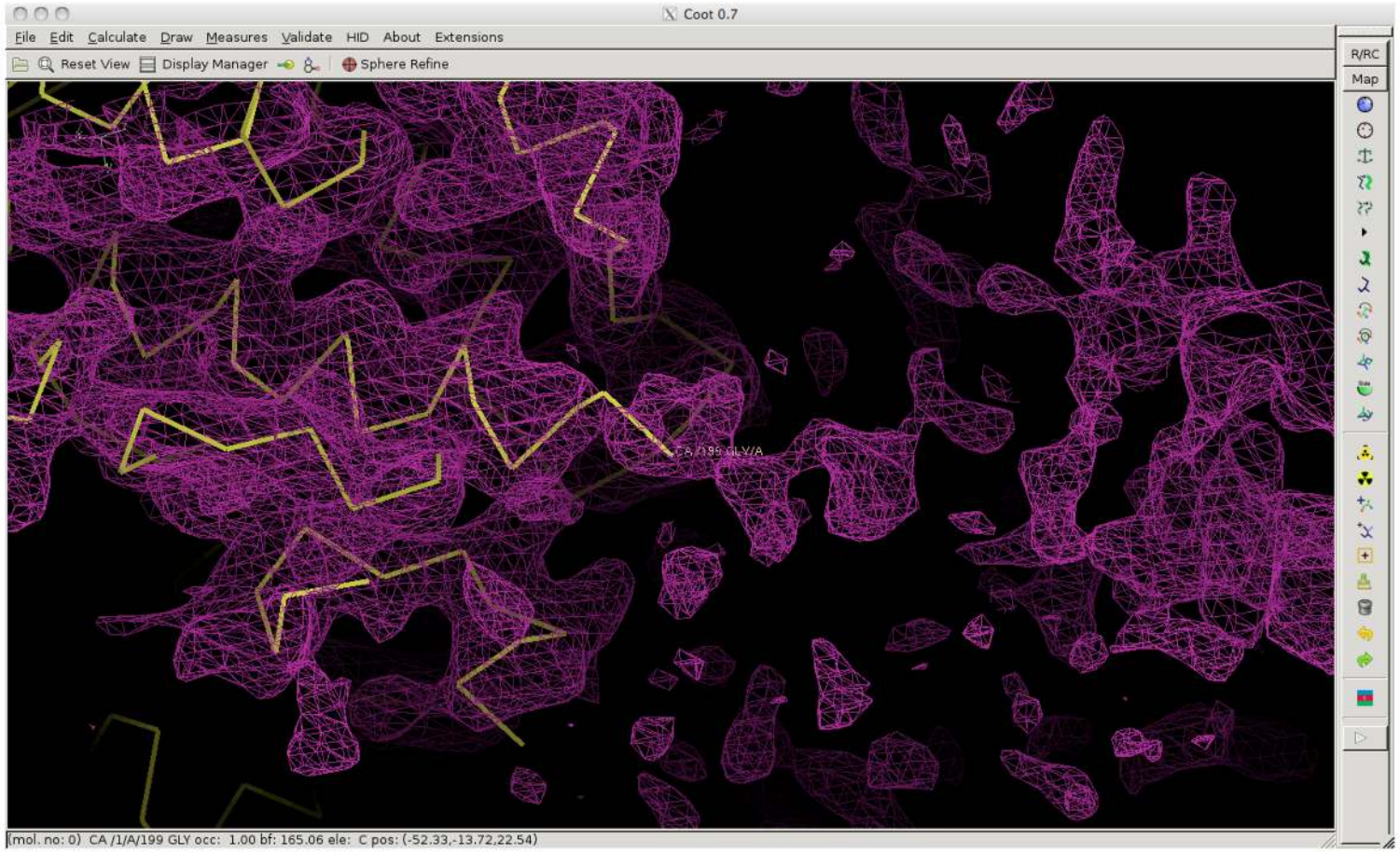


Map Sharpening/Blurring

Output Array of Sharpened Maps

Standard:

4iwo (2.6Å)

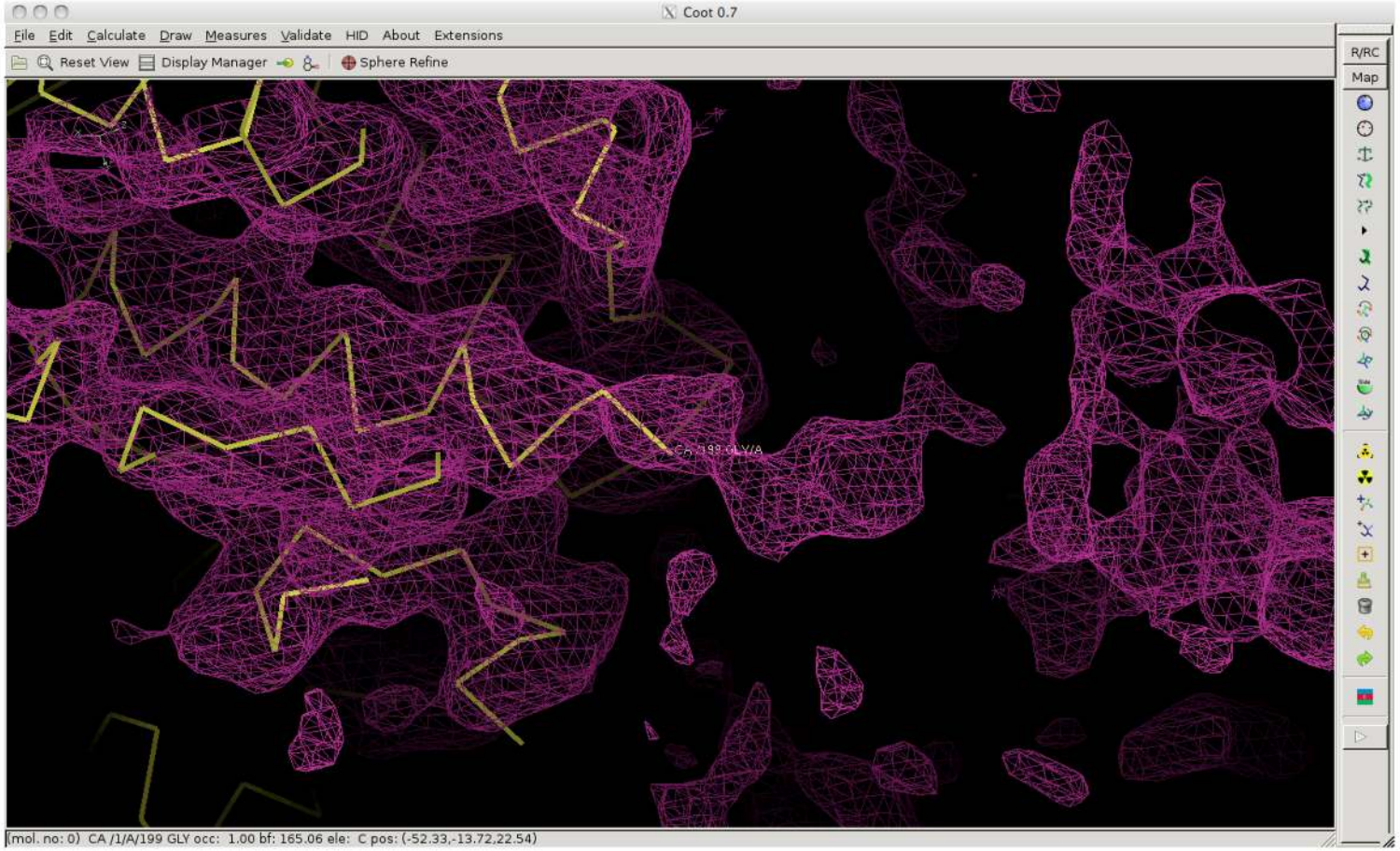


Map Sharpening/Blurring

Output Array of Sharpened Maps

Blur 50:

4iwo (2.6Å)

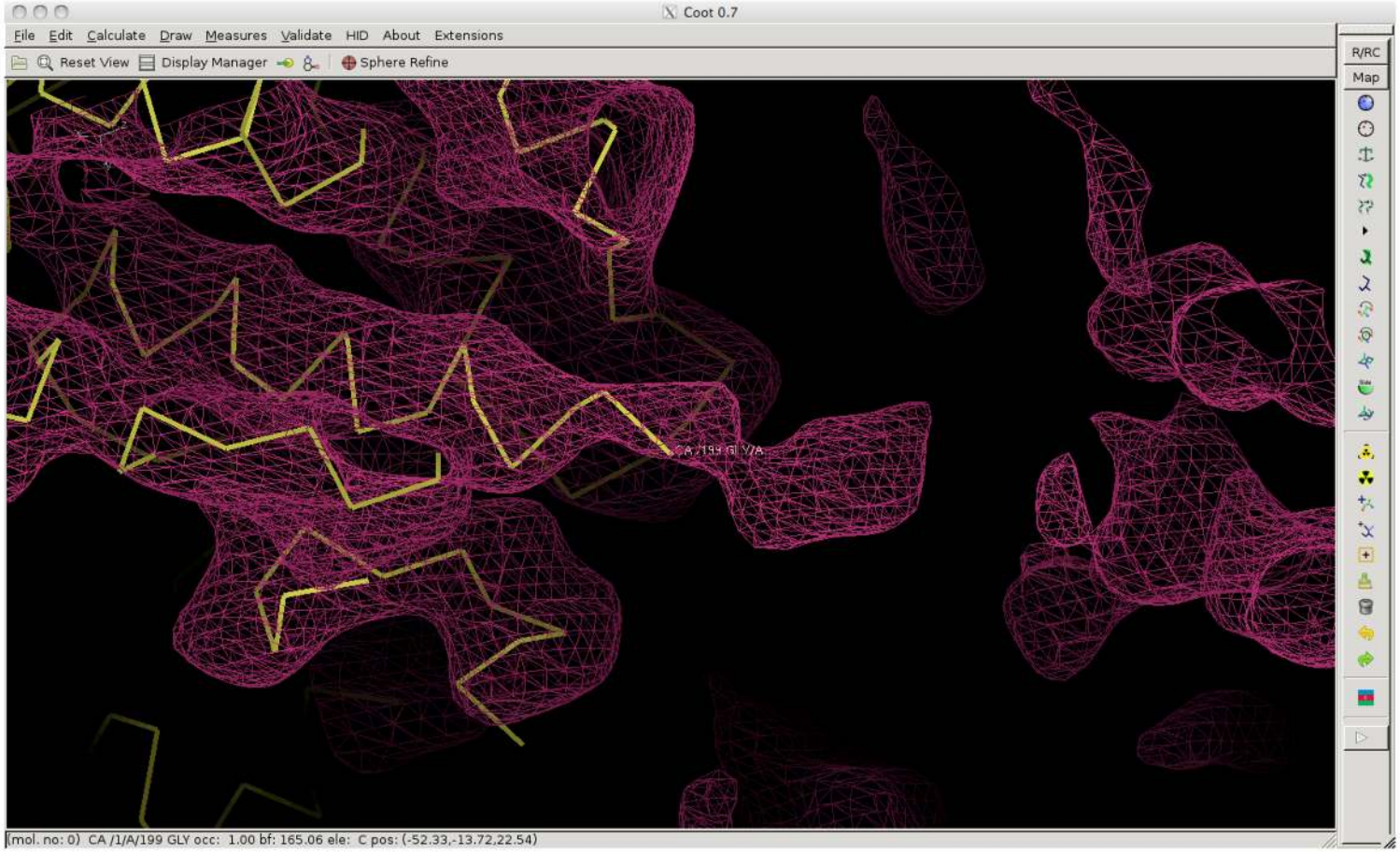


Map Sharpening/Blurring

Output Array of Sharpened Maps

Blur 150:

4iwo (2.6Å)



What and When

Early stages (e.g. straight after MR)

- Rigid body refinement
- Jelly body – sometimes up to 200 cycles

Medium stages – during model building

- Auto local NCS – wherever possible
- External restraints (40 cycles) – homologue available
- Otherwise, jelly body… but not together
- H-bond restraints – no homologue available
- Secondary structure conformation restraints – model building tool
- Add hydrogens (?)

Medium-final stages

- TLS – at medium resolutions
- Anisotropic B-factors – only at high resolution
- Twin refinement – only if you are sure

Final stages of refinement

Jelly body – around 20 cycles

Summary

CCP4 tools for refinement:

REFMAC: Refinement, map calculation

ProSMART: External restraints, comparative analysis

LIBG: Nucleic acid restraints

LORESTR: Automated low-resolution refinement pipeline *(released soon!)*

ACEDRG: Generating ligand description dictionaries and conformers

COOT: Automated & manual real-space refinement
Visualisation & manipulation of restraints, map blurring
…also morphing, jiggle-fit, backrub rotamers…

Many tools are applicable to cryo-EM as well as MX

Acknowledgements

Contact: **okovalev@mrc-lmb.cam.ac.uk**
nicholls@mrc-lmb.cam.ac.uk

Computational Crystallography Group:

Garib Murshudov
Paul Emsley
Oleg Kovalevskiy
Fei Long
Rob Nicholls
Andrea Thorn
Michal Tykac

Many Thanks:

CCP4 core team
Colleagues from MRC-LMB and CCP4
Users for feedback

Software available for Mac/Linux/Windows via CCP4

Computational Crystallography Group website:
www2.mrc-lmb.cam.ac.uk/groups/murshudov/

REFMAC5 tutorials: **www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/tutorials/refmac_tutorial/**

ProSMART tutorials: **www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/tutorials/prosmart_tutorial/**

REFMAC5:

Murshudov *et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.* D67, 355-67.

Low-resolution refinement with REFMAC5 and ProSMART:

Nicholls *et al.* (2013) Recent Advances in Low Resolution Refinement Tools in REFMAC5. Adv. Methods for Bio. Xtallography, 231–58.
Nicholls *et al.* (2012) Low Resolution Refinement Tools in REFMAC5. *Acta Cryst.* D68, 404-17.

ProSMART:

Nicholls *et al.* (2014) Conformation-Independent Structural Comparison of Macromolecules with ProSMART. *Acta Cryst.* D70, 2487-99.
Nicholls *et al.* (2012) An Overview of ProSMART. *CCP4 Newsletter on Protein Crystallography*.

Tools for cryo-EM fitting:

Brown *et al.* (2015) Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Cryst.* D71, 136-53.

MRC Laboratory of Molecular Biology

CCP4

wellcome trust

BBSRC bioscience for the future