Low-Resolution MX Model Refinement

CCP4 Crystallography School and Workshop

14-24 November 2018

Sao Carlos, Brazil

Roberto A Steiner

roberto.steiner@kcl.ac.uk

KING'S College LONDON

University of London

with slides from Rob Nicholls

MRC Laboratory of Molecular Biology

Key aspects of refinement

- Objective function
- Method of optimization
- Model parametrization
- Prior knowledge

Objective (target) function

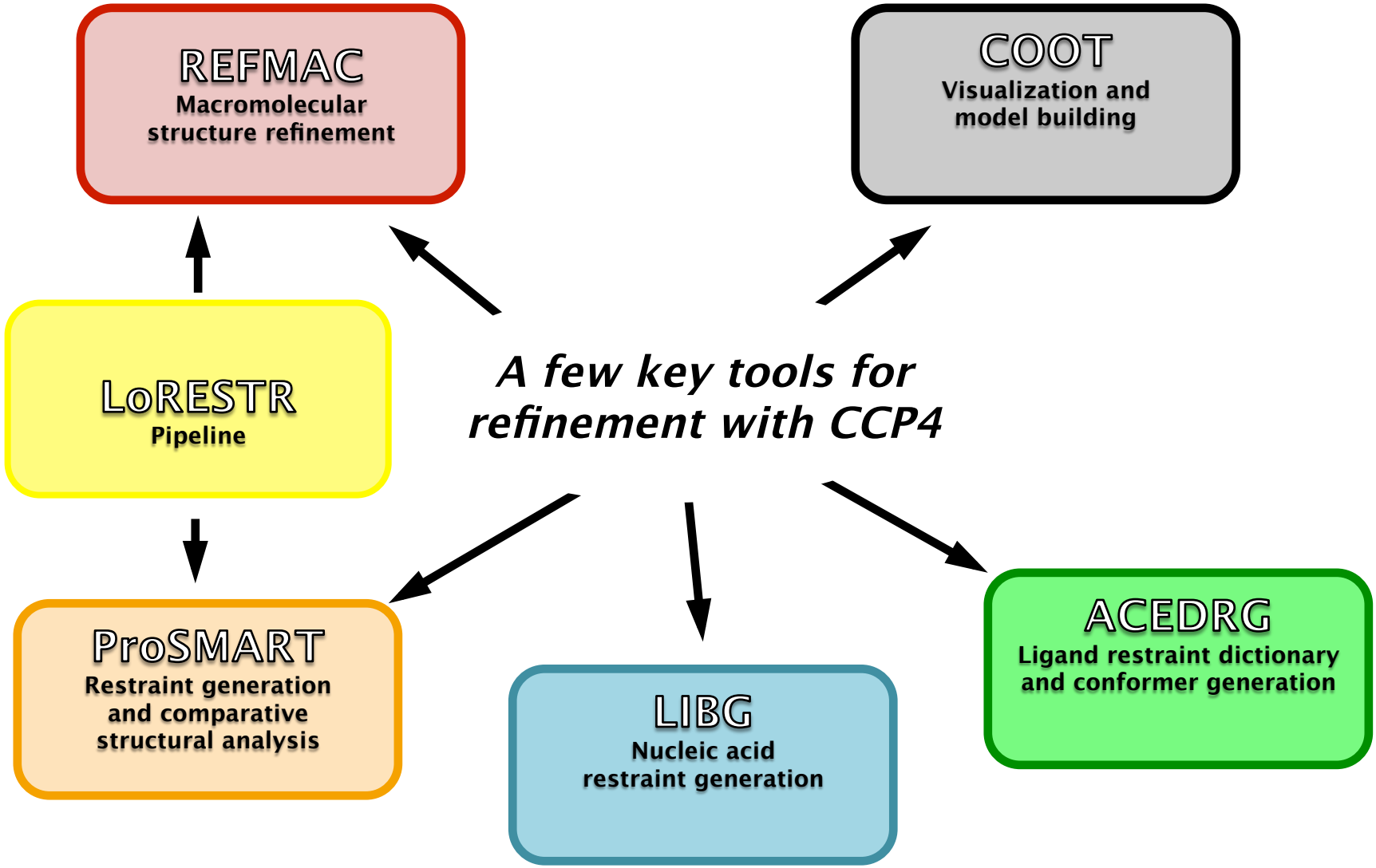
2. Target functions in *REFMAC*5

As in all other refinement programs, the target function minimized in *REFMAC*5 has two components: a component utilizing geometry (or prior knowledge) and a component utilizing experimental X-ray knowledge,

$$f_{\text{total}} = f_{\text{geom}} + w f_{\text{xray}}, \tag{1}$$

where f_{total} is the total target function to be minimized, consisting of functions controlling the geometry of the model and the fit of the model parameters to the experimental data, and w is a weight between the relative contributions of these two components. In macromolecular crystallography, the weight is traditionally selected by trial and error. *REFMAC*5 offers automatic weighting, which is based on the fact that both components are the natural logarithm of a probability distribution. However, this 'automatic' weight may lead to unrea-

$$f_{\text{total}} = -\log[P_{\text{posterior}}(\text{model}; \text{obs})]$$
$$f_{\text{geom}} = -\log[P_{\text{prior}}(\text{model})]$$
$$f_{\text{xray}} = -\log[P_{\text{likelihood}}(\text{obs}; \text{model})].$$


REFMAC
Macromolecular structure refinement
COOT
Visualization and model building
LoRESTR
Pipeline
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

Low Resolution Refinement

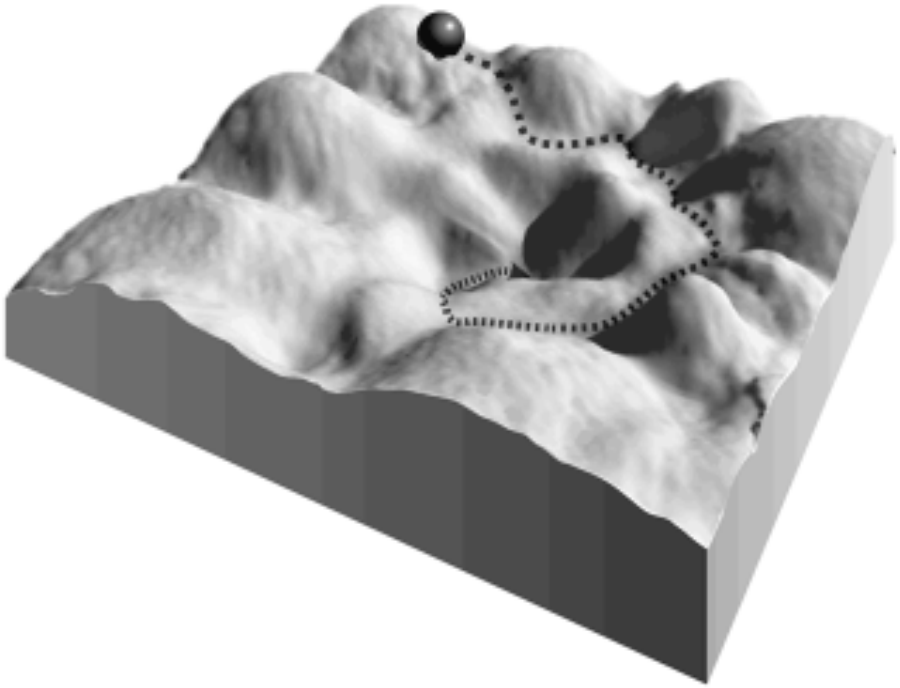
At low-resolution:

- Reflection intensities often noisy
- Limited data - poor observation:parameter ratio
- Refinement becomes unstable
- Overfitting - R-factors diverge

How to improve the observation:parameter ratio?

- Reduce number of parameters
- Increase number of restraints to regularise refinement

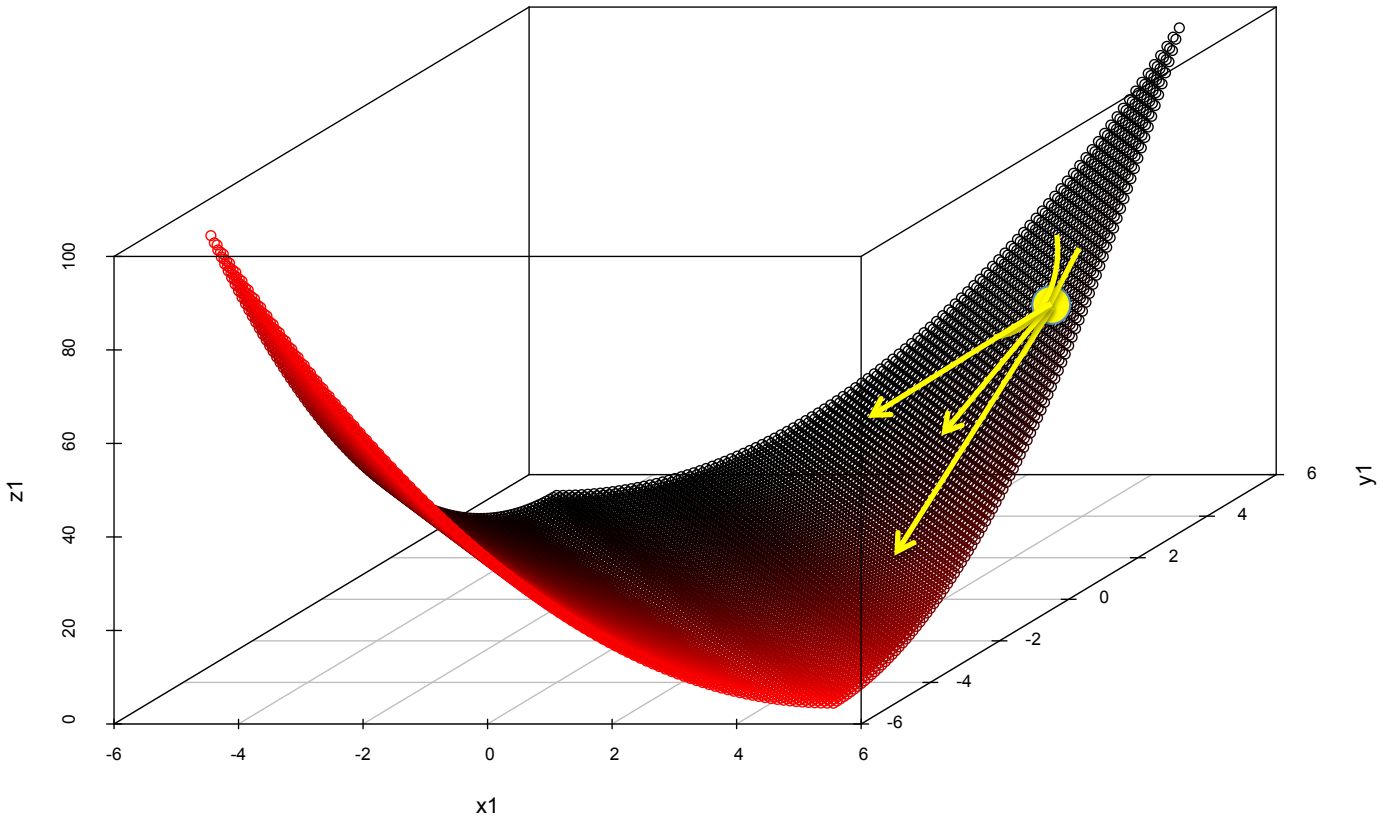
Regularisation



Regularisation

Example:

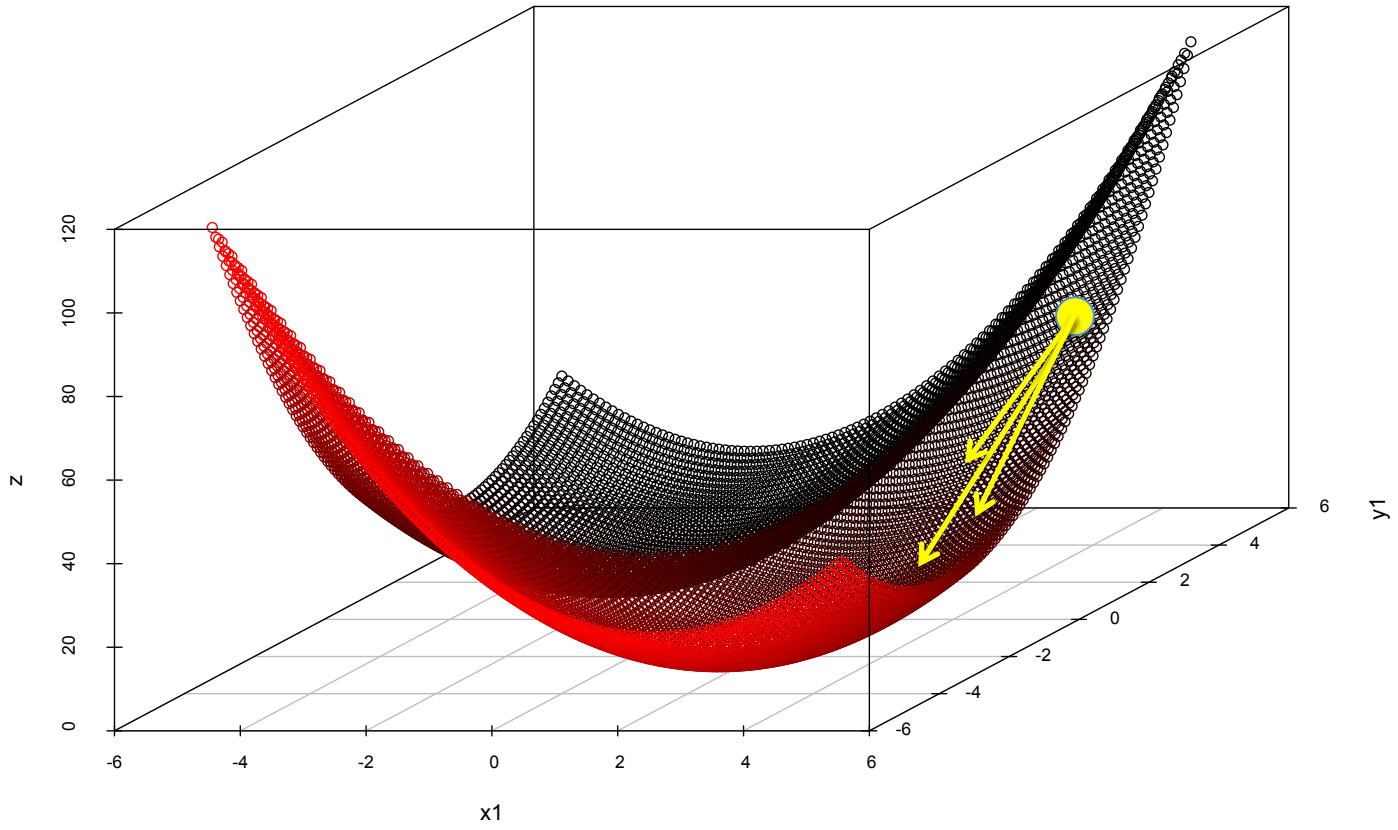
$$z = (x + y)^2$$



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**

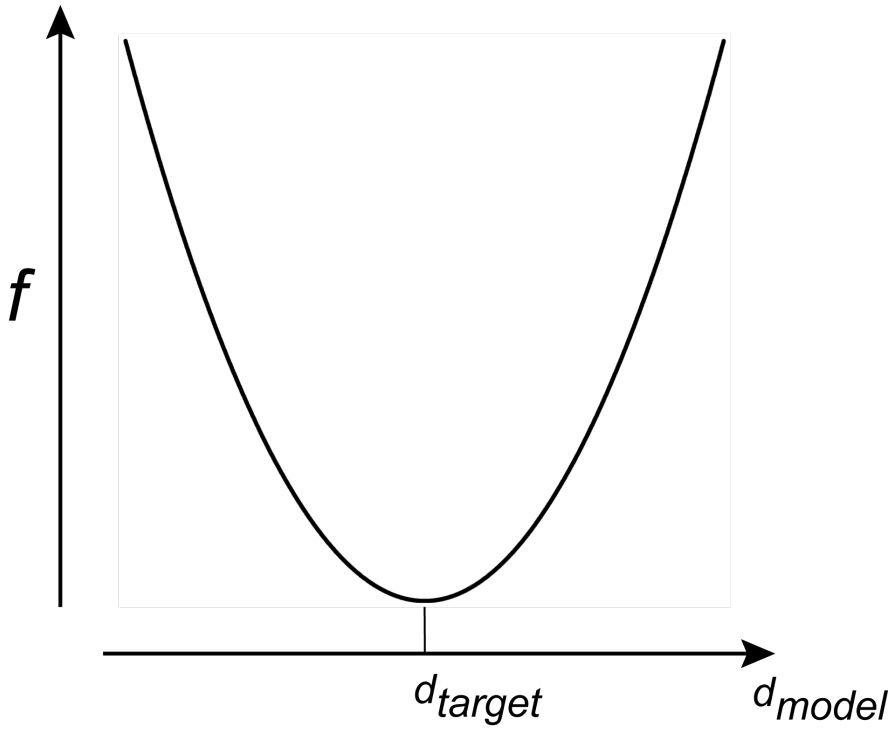
Regularisers with a target value

Some examples of restraints

Bonds/Angles

$$f_{bonds} = \sum_{bonds} \frac{1}{\sigma_{bond}^2} \left(d_{\text{model}} - d_{\text{target}} \right)^2$$

$$f_{angles} = \sum_{angles} \frac{1}{\sigma_{angle}^2} \left(\alpha_{\text{model}} - \alpha_{\text{target}} \right)^2$$



NCS

(Non-Crystallographic Symmetry Restraints)

Three ways of dealing with NCS:

1. NCS constraints
2. Global NCS restraints
3. Local NCS restraints

NCS

(Non-Crystallographic Symmetry Restraints)

Three ways of dealing with NCS:

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
3. Local NCS restraints

NCS

(Non-Crystallographic Symmetry Restraints)

Three ways of dealing with NCS:

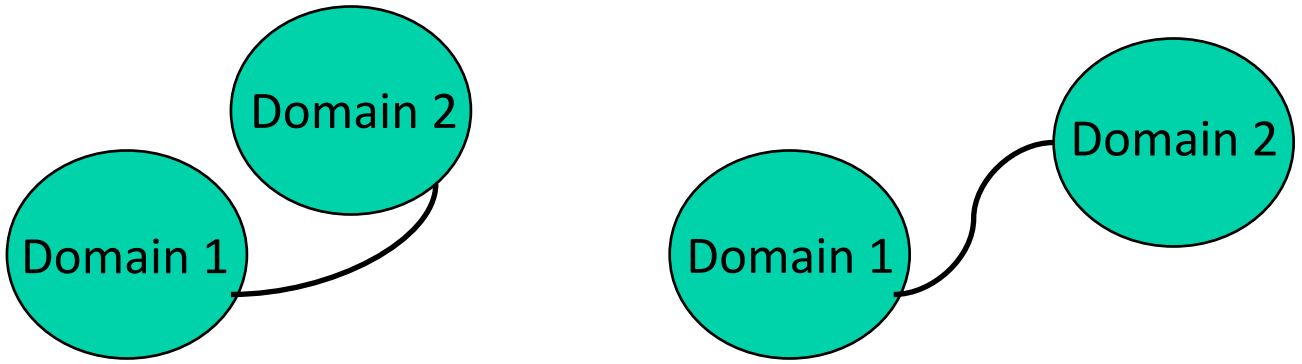
1. NCS constraints
2. Global NCS restraints
   - Molecules are superimposed
   - Difference between corresponding atoms are minimised
3. Local NCS restraints

NCS

(Non-Crystallographic Symmetry Restraints)

Three ways of dealing with NCS:

1. NCS constraints
2. Global NCS restraints
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



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

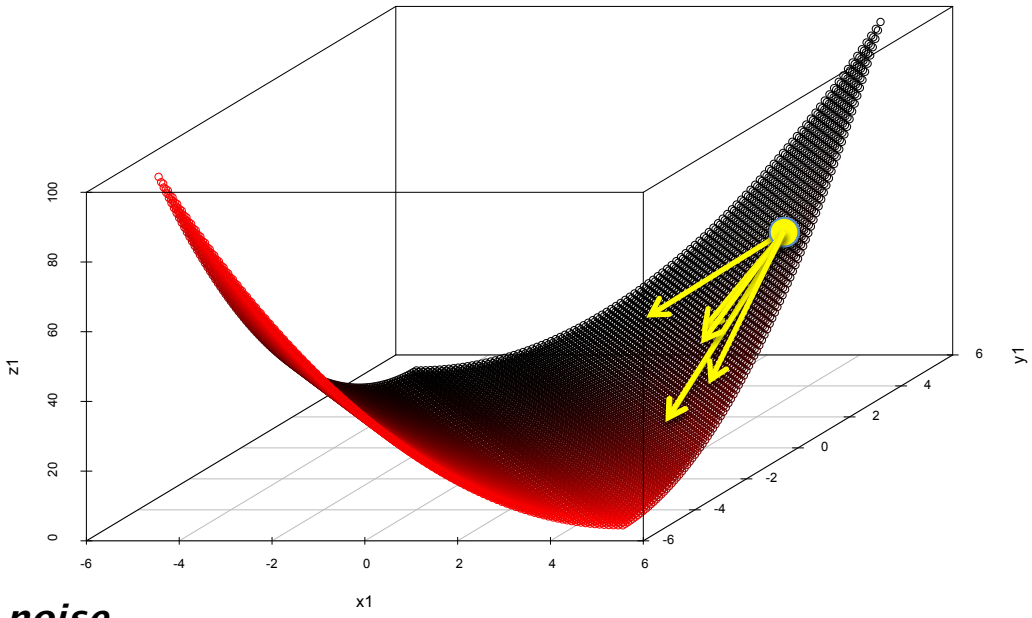
Regularisers without a target:

$$f = \sum_{\text{close atom pairs}} \frac{1}{\sigma^2}\left(d - d_{\text{current}}\right)^2$$

d : interatomic distance

d_{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

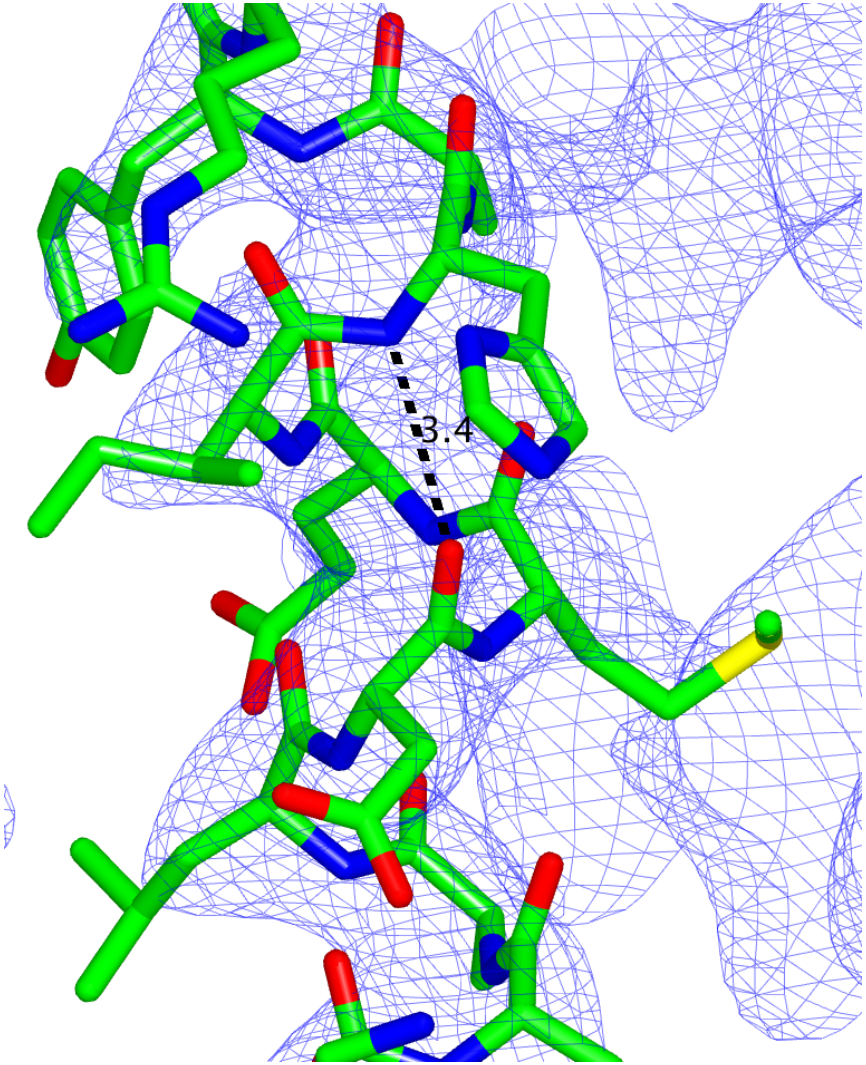
ProSMART

Injection of prior knowledge to aid new structure determination

- **External Restraints from homologous structures**
  - Protein or nucleic acid chains
- **Hydrogen bond restraints**
  - Protein backbone
- **Generic self-restraints**
  - Everything − protein, nucleic acid, ligand, water
- **Structure analysis**
  - Alignment & comparison - helps analyse differences between models

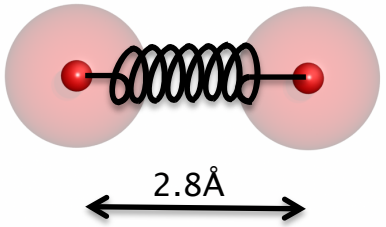
Independent of global conformation

ProSMART External Restraints



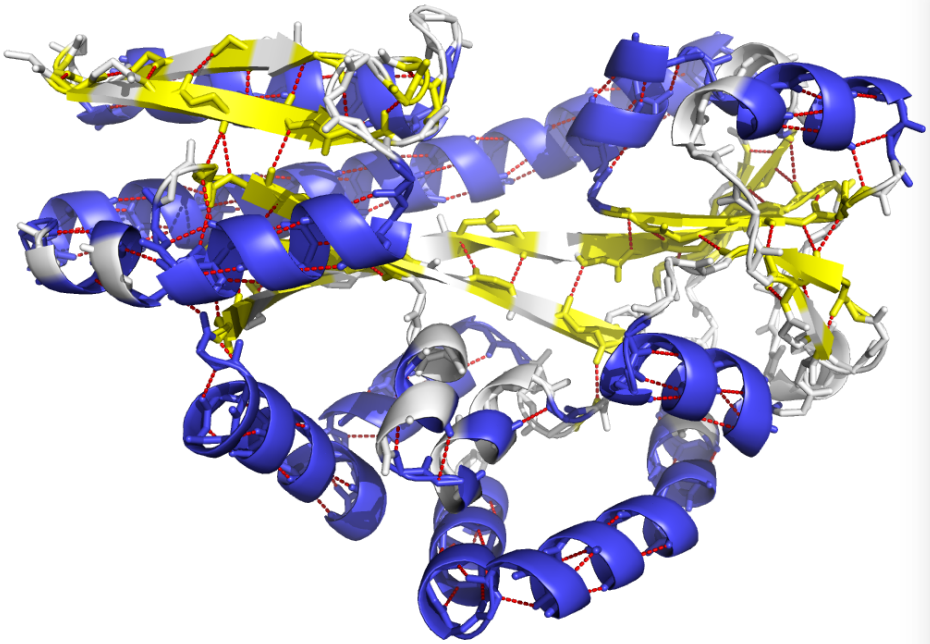
3g4w – 3.7Å

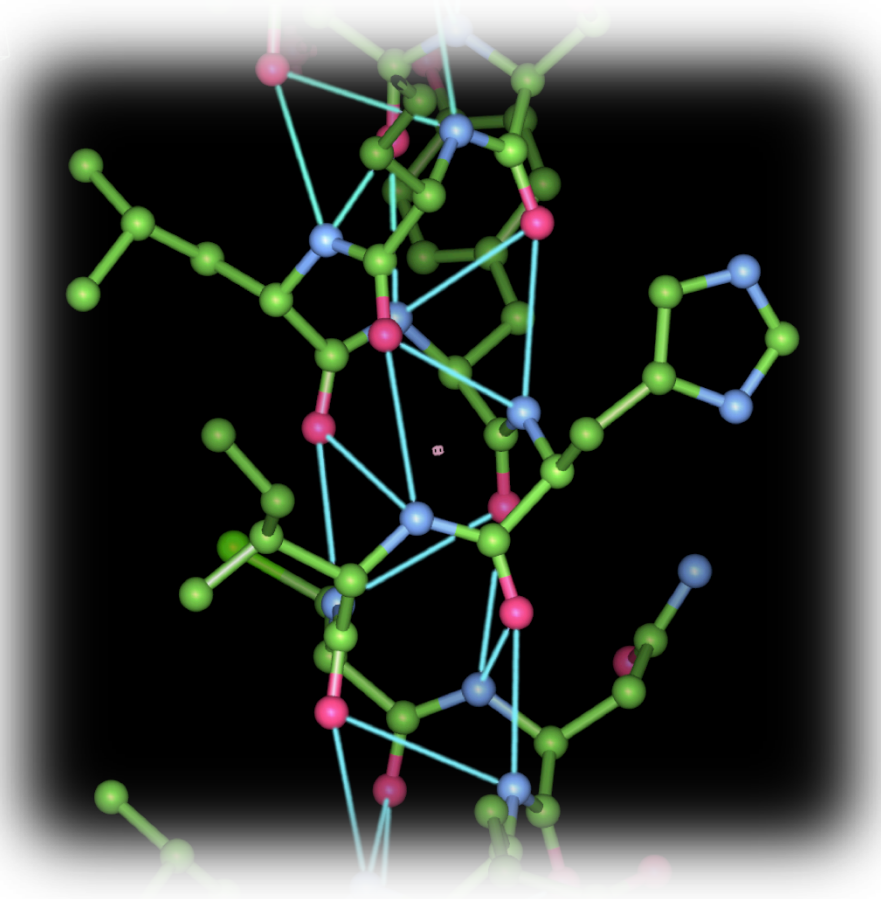
Prior information:



Stabilises structural features

ProSMART H-Bond Restraints



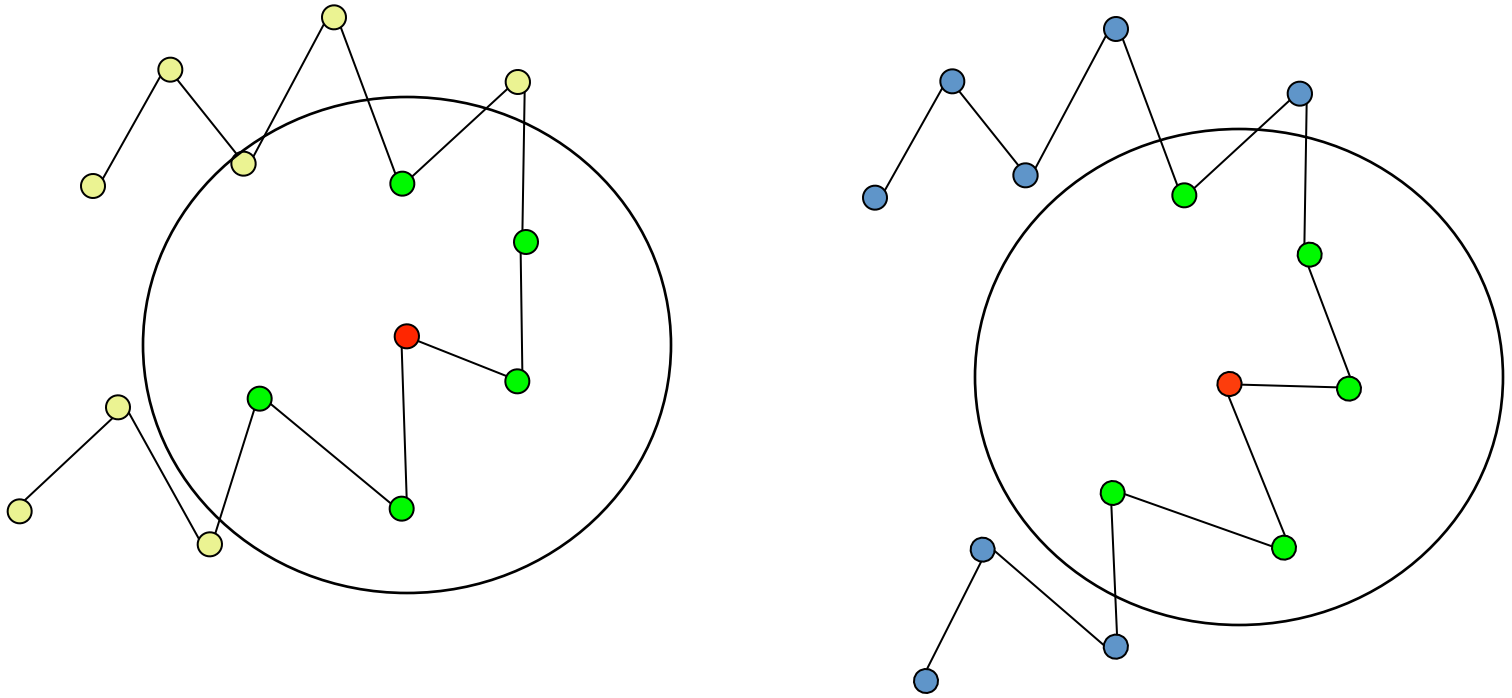


External Restraint Generation

When a reliable model of a close homologue is available

structure to be refined

known similar structure (prior)



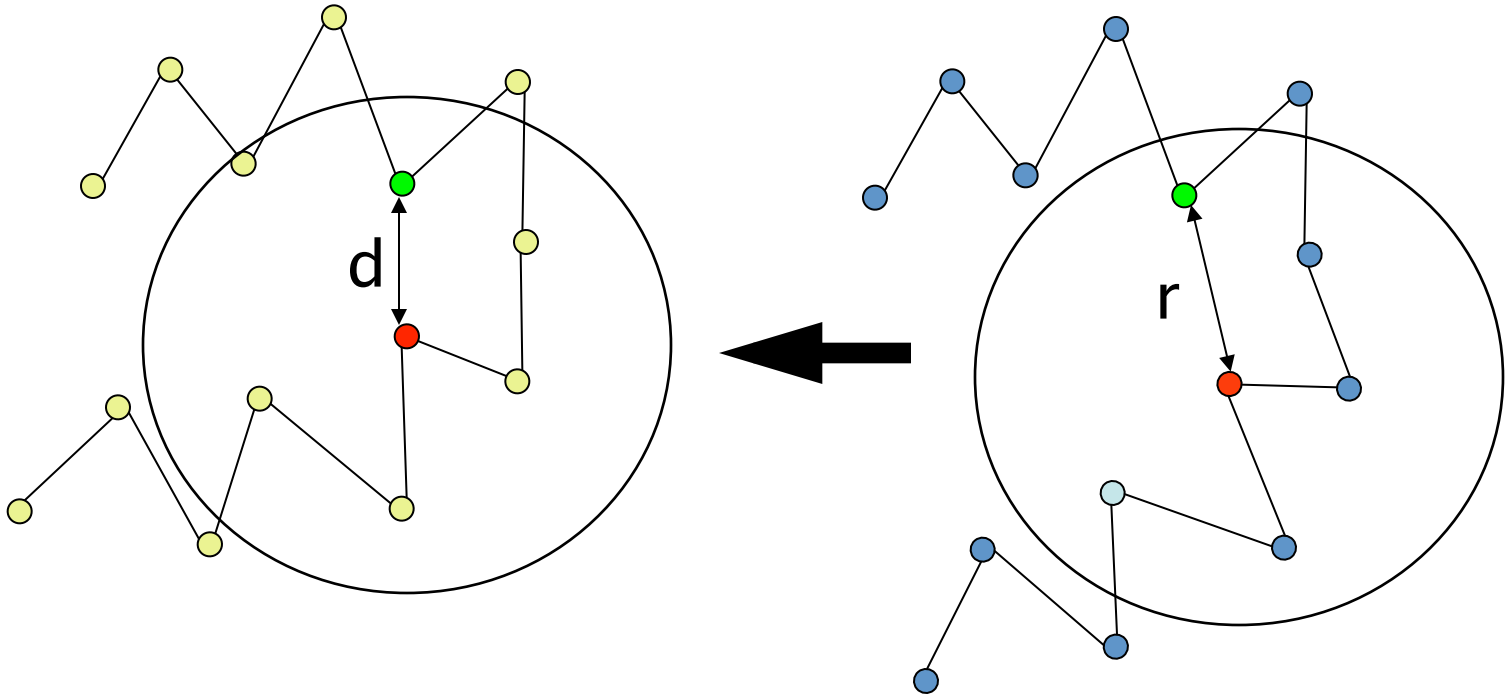
(abstract representation of an atomic model; circles = atoms)

External Restraint Generation

When a reliable model of a close homologue is available

structure to be refined

known similar structure (prior)

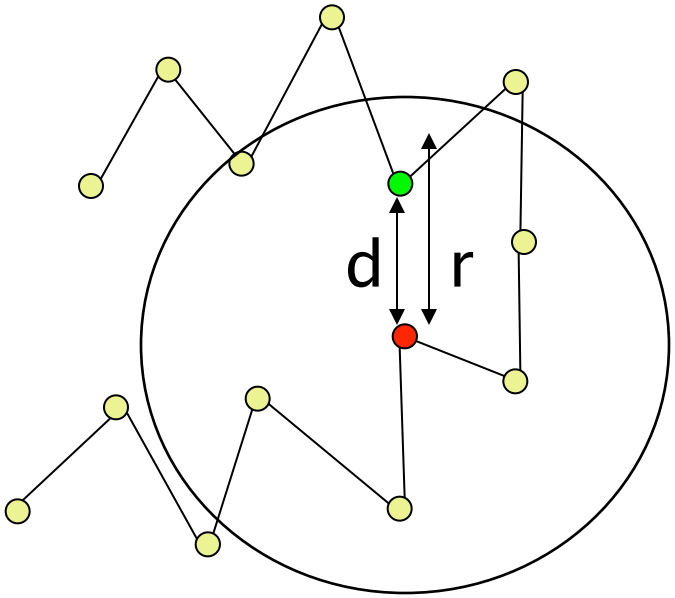


(abstract representation of an atomic model; circles = atoms)

External Restraint Generation

When a reliable model of a close homologue is available

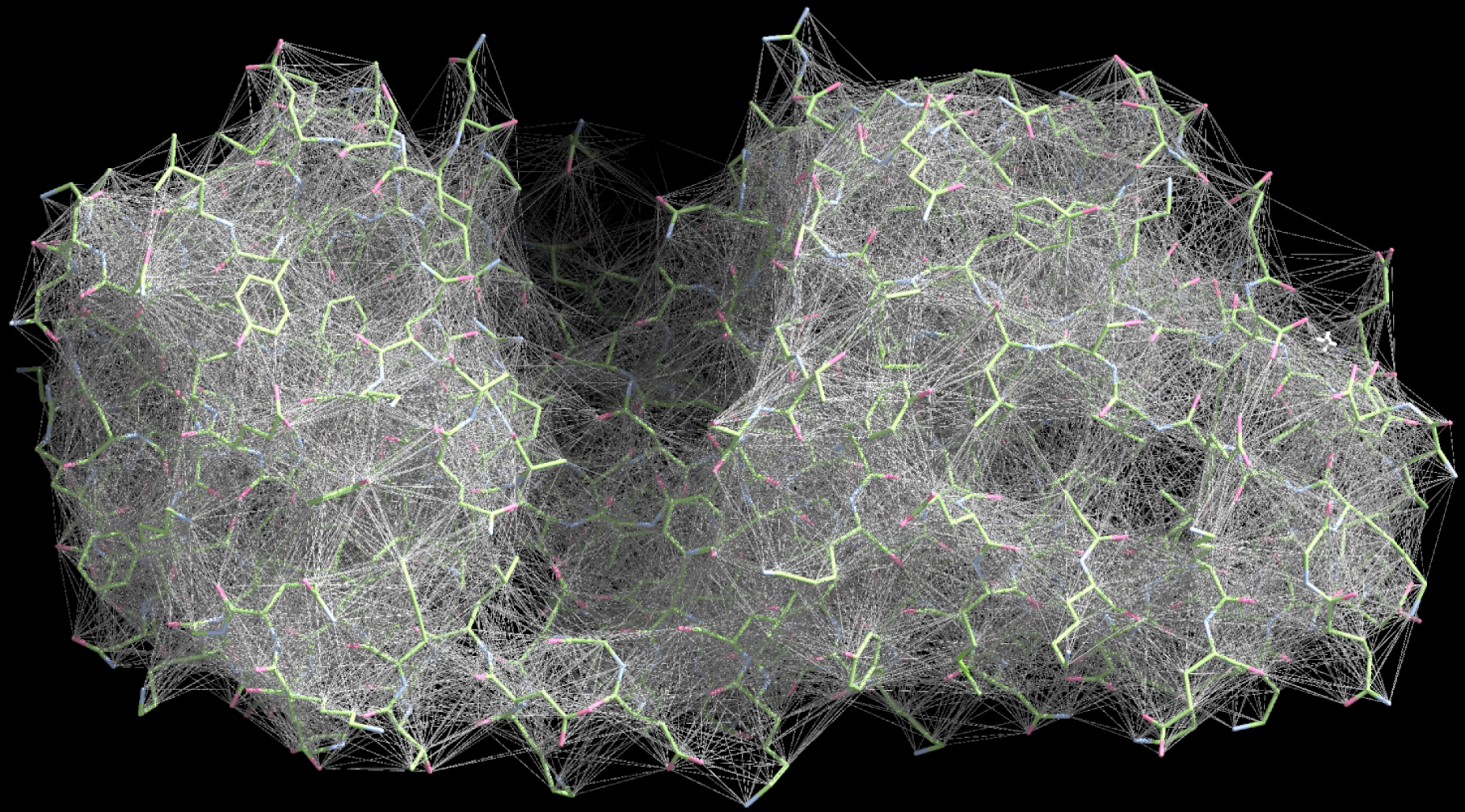
structure to be refined



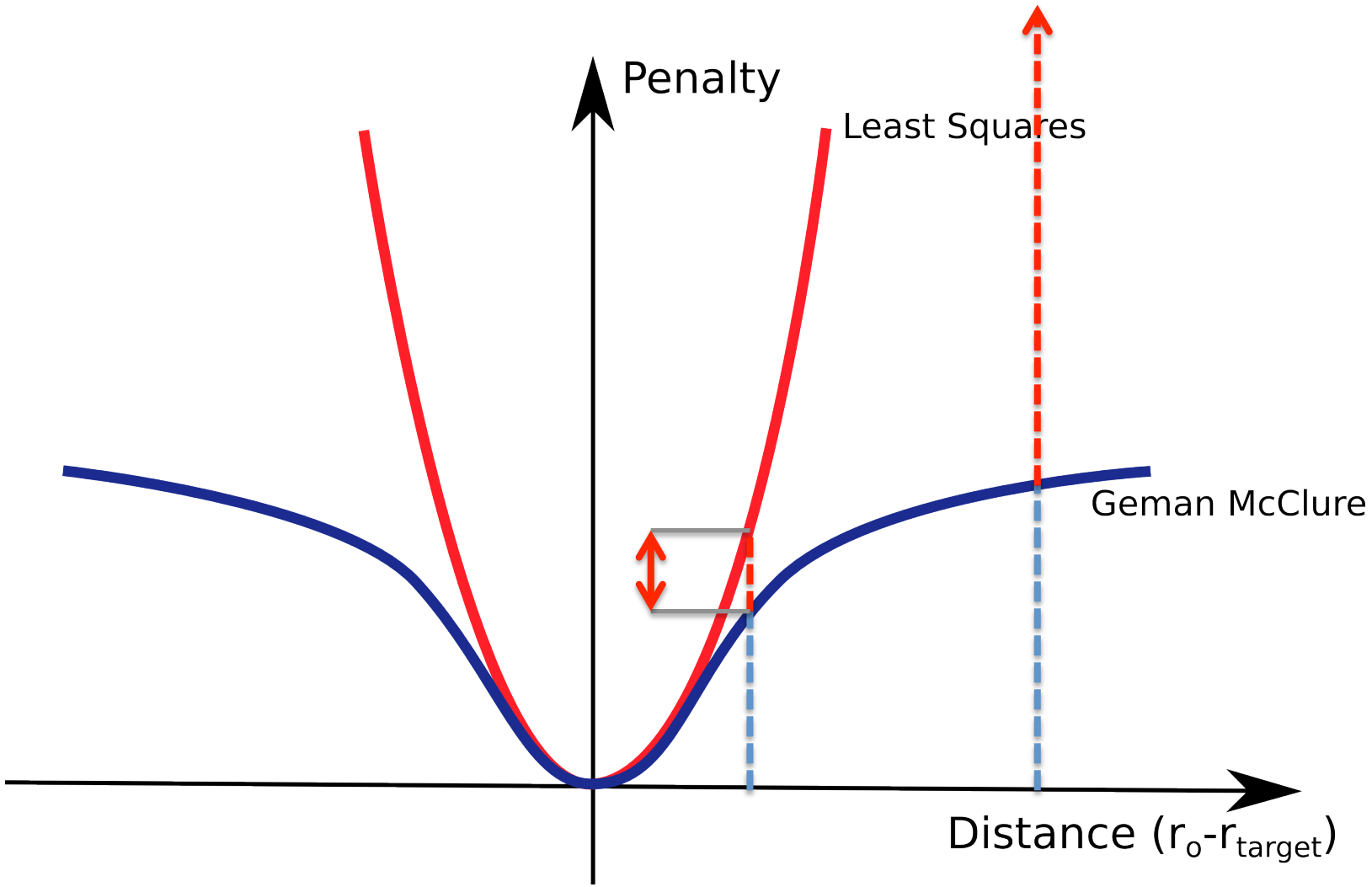
$d \sim N(r,\sigma^2)$

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

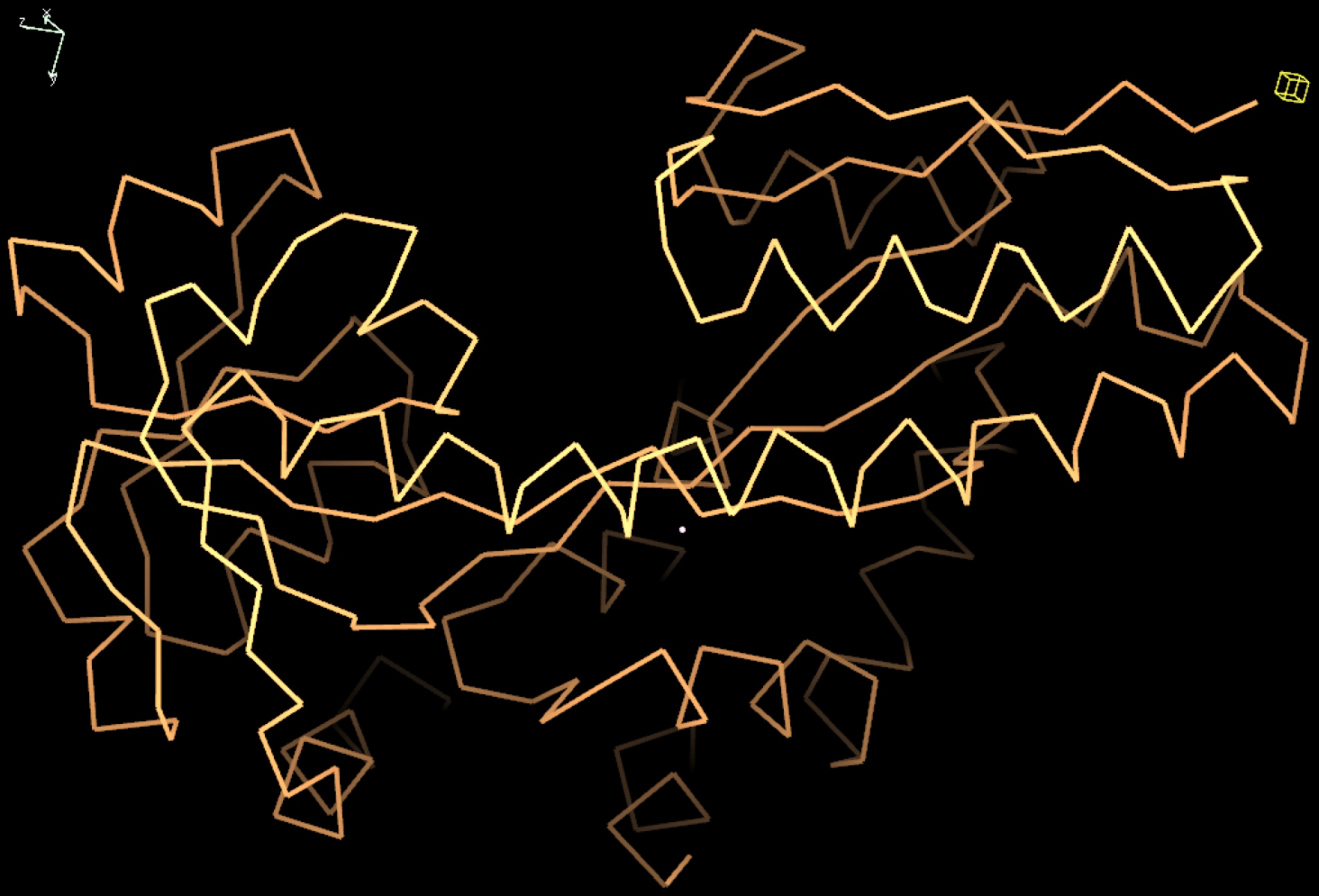
ProSMART Restraints



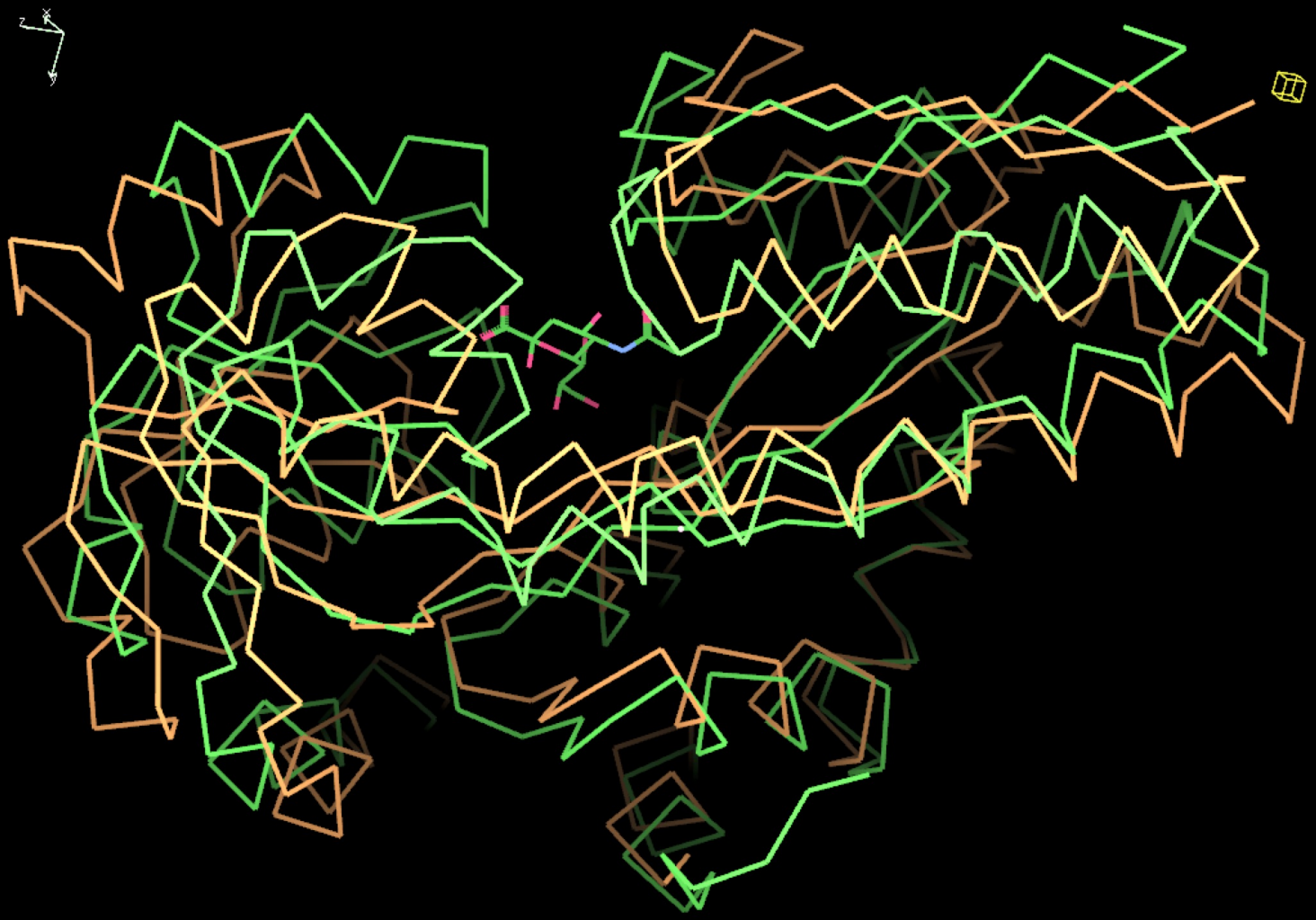
Robust Estimation



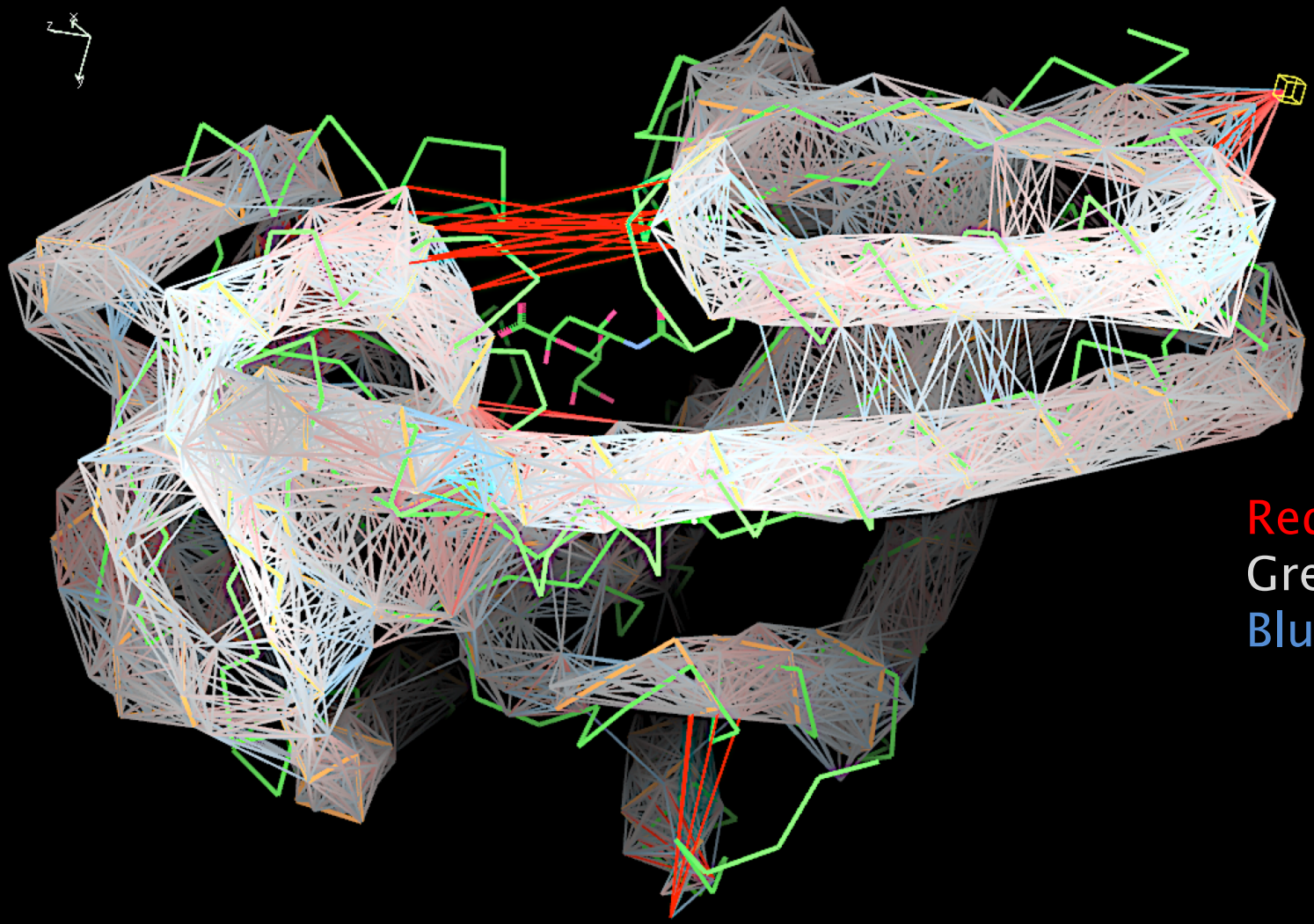
ProSMART Restraints



ProSMART Restraints

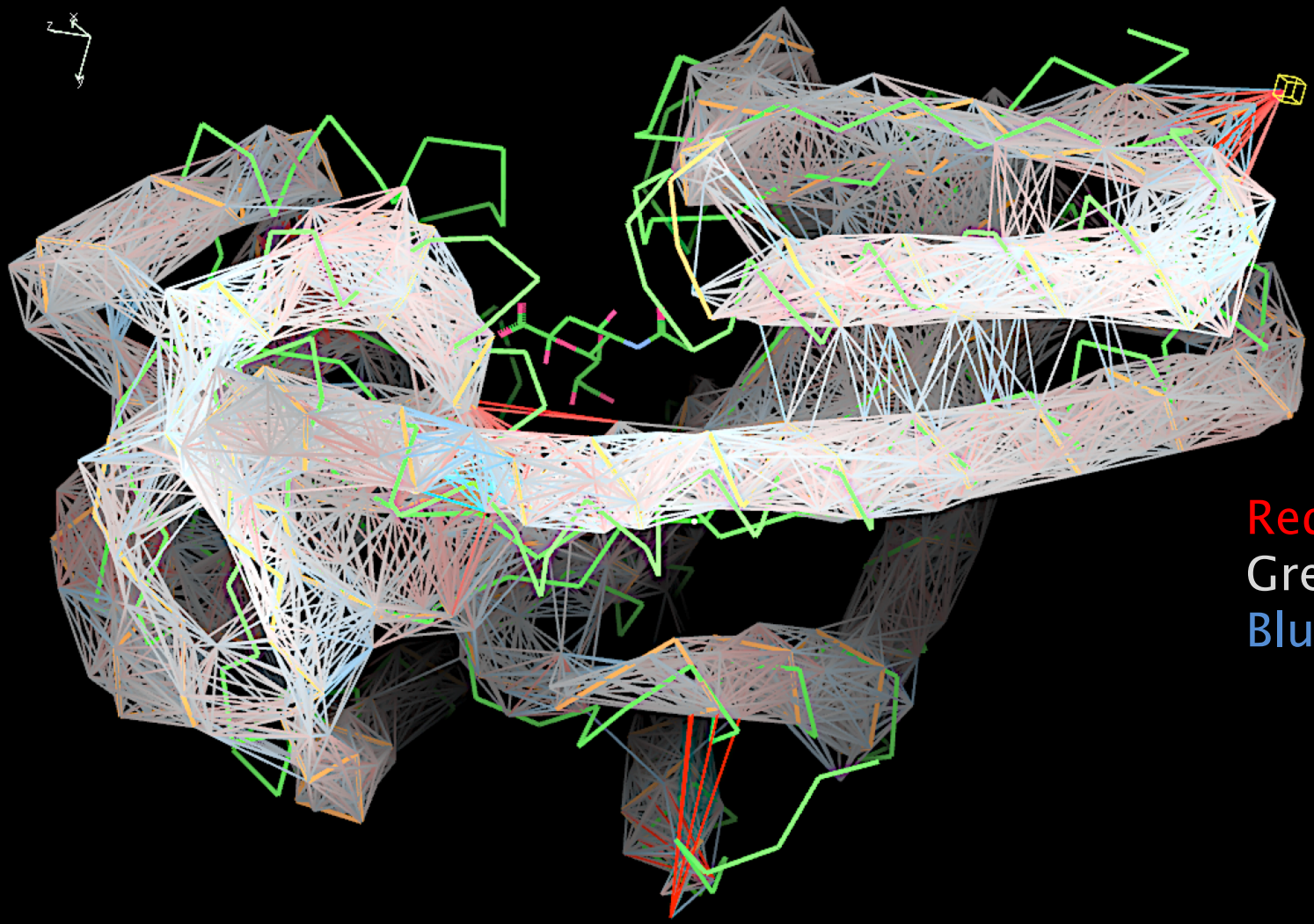


ProSMART Restraints



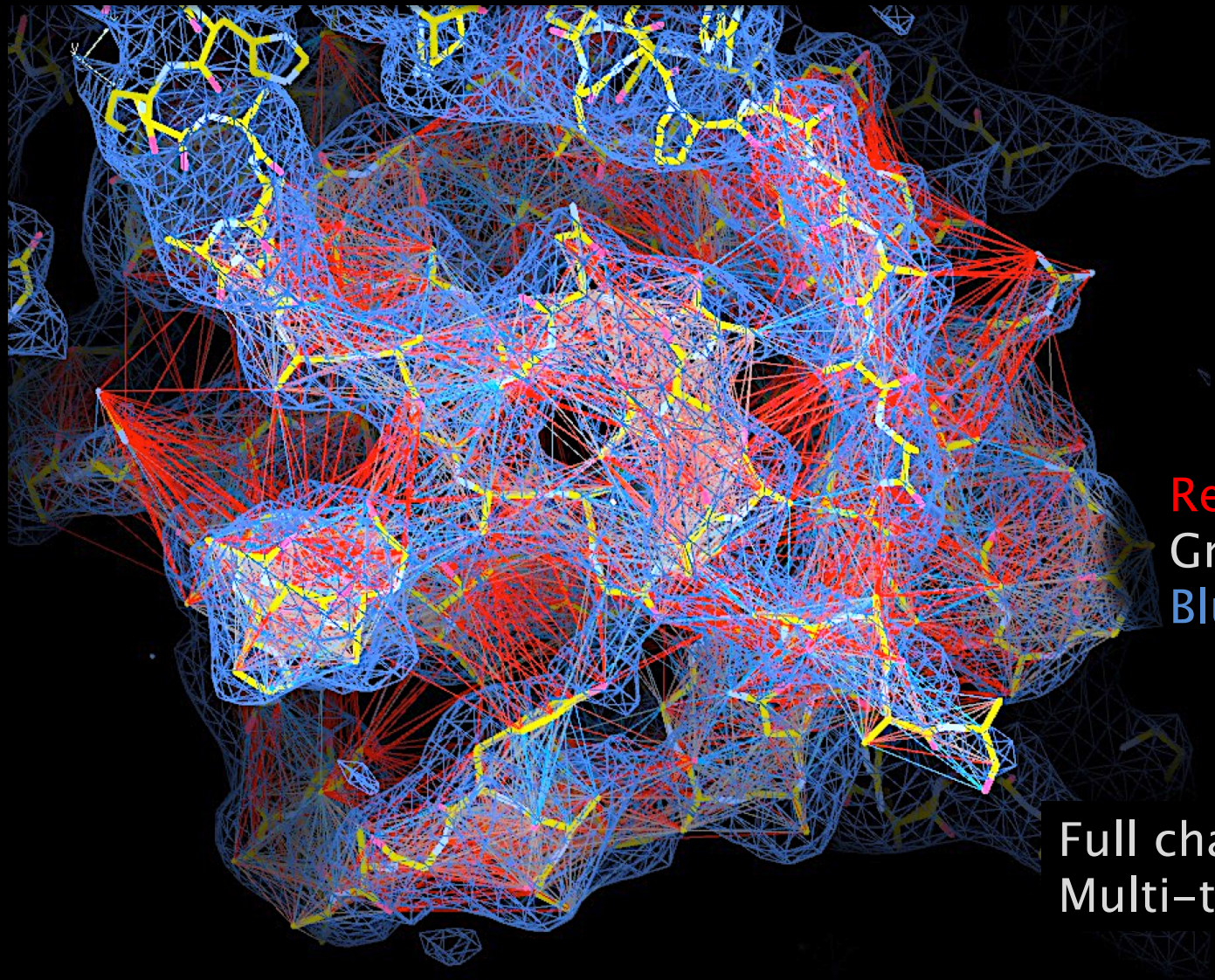
Red: long
Grey: similar
Blue: short

ProSMART Restraints



Red: long
Grey: similar
Blue: short

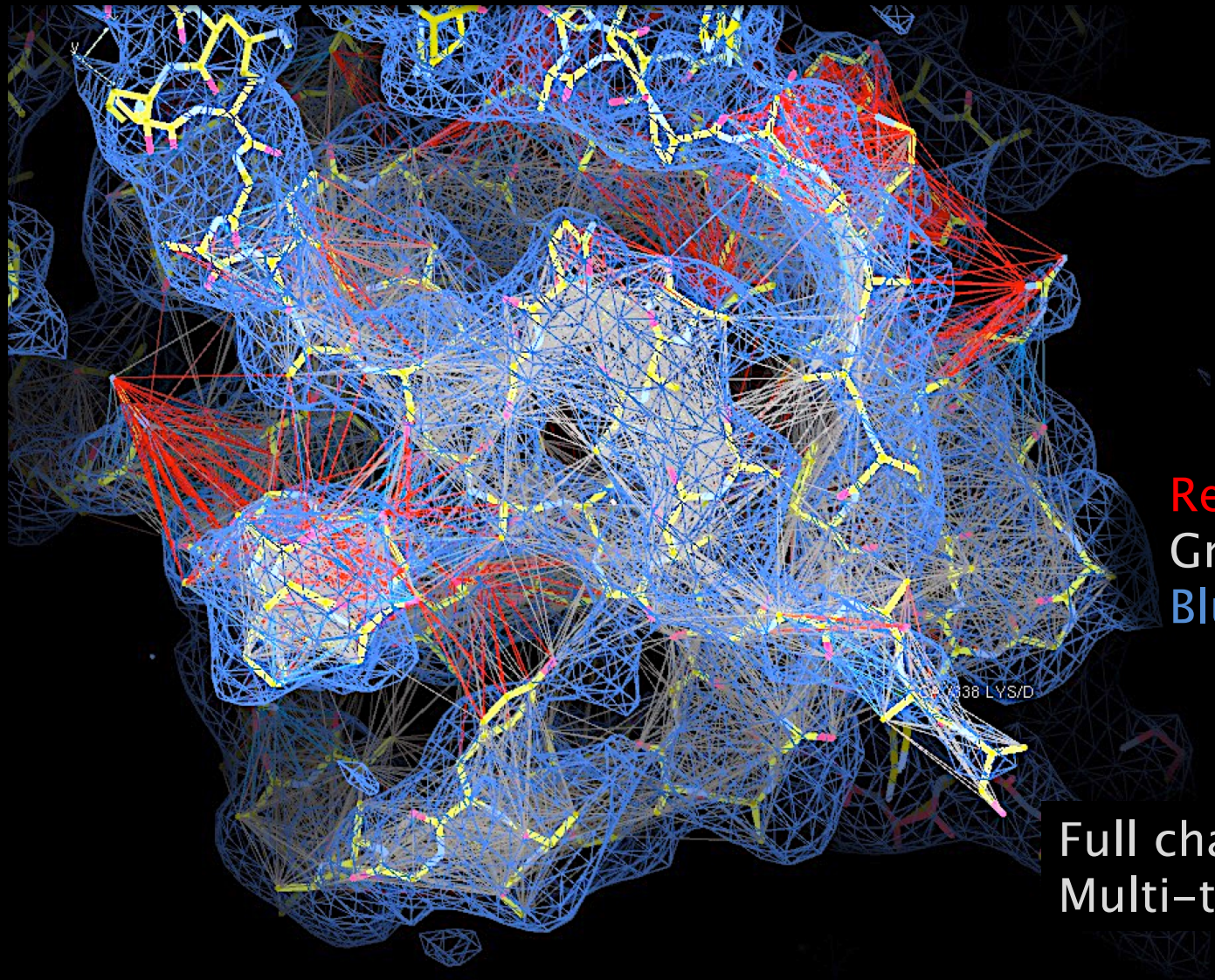
ProSMART Restraints in Coot



Red: long
Grey: similar
Blue: short

Full chain refine
Multi-threaded

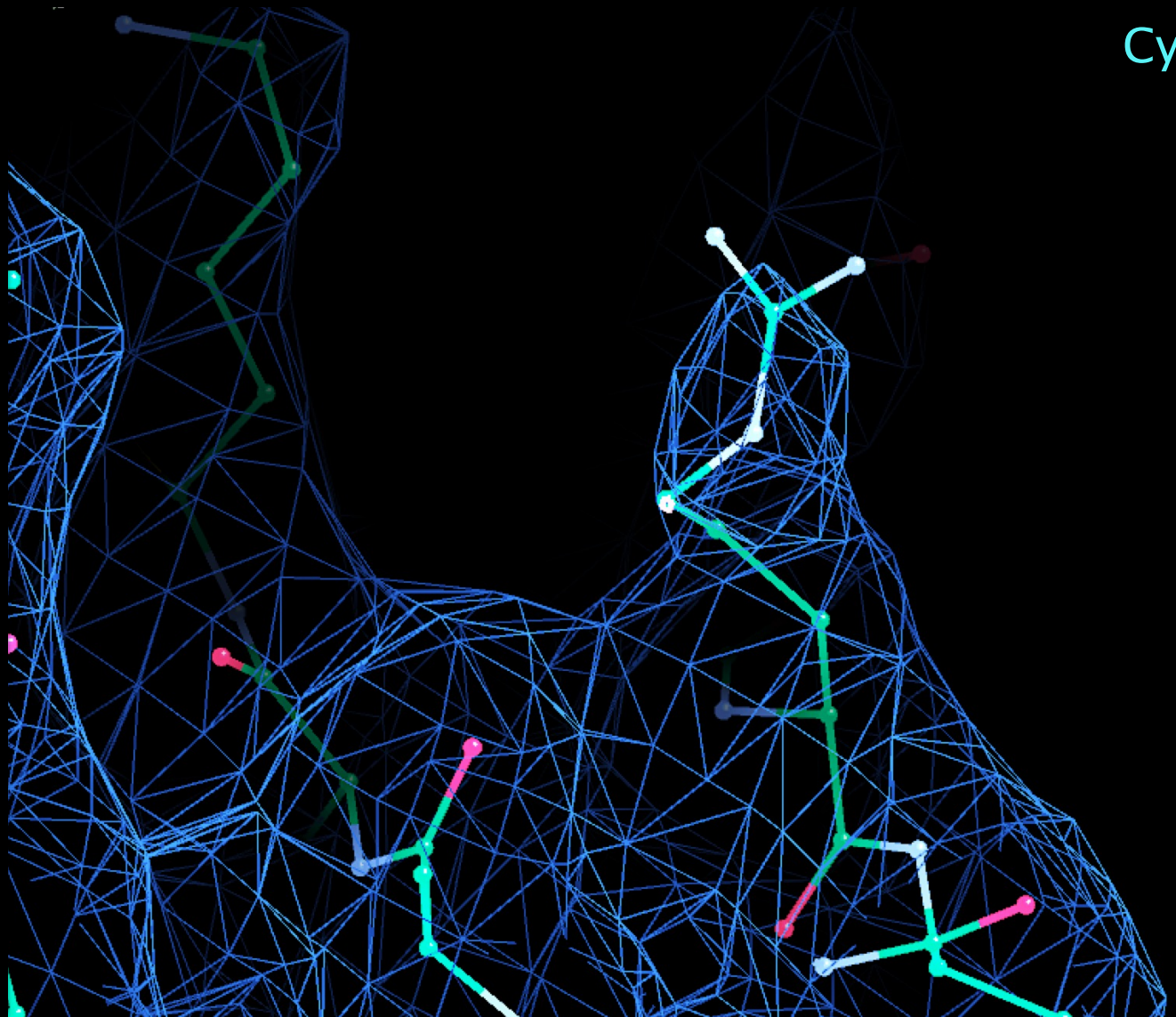
ProSMART Restraints in Coot



Red: long
Grey: similar
Blue: short

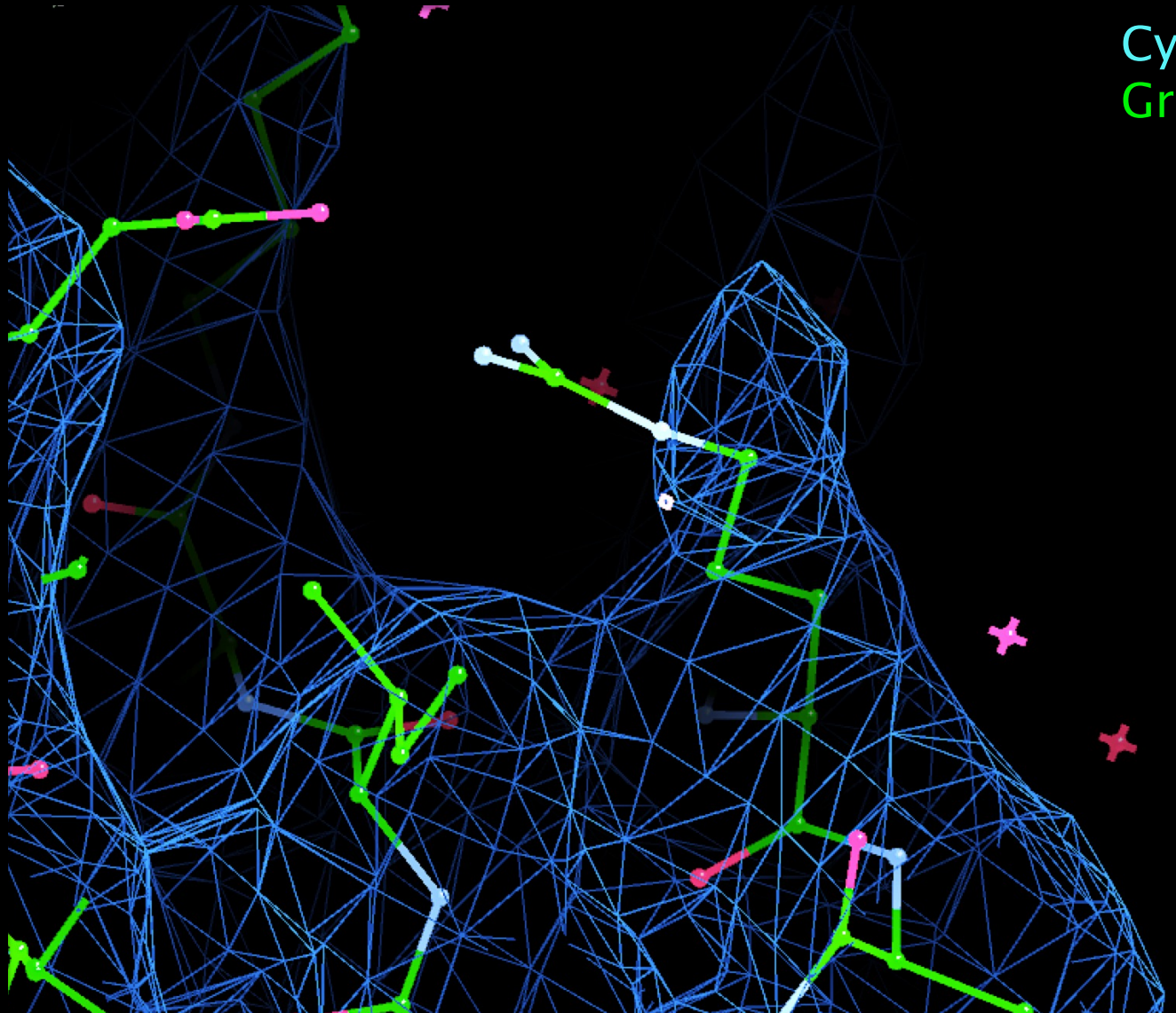
Full chain refine
Multi-threaded

Robust Estimation



Cyan: original

Robust Estimation



Cyan: original
Green: homolog

Robust Estimation

Cyan: original
Green: homolog

Robust Estimation

Cyan: original
Green: homolog
Yellow: refined

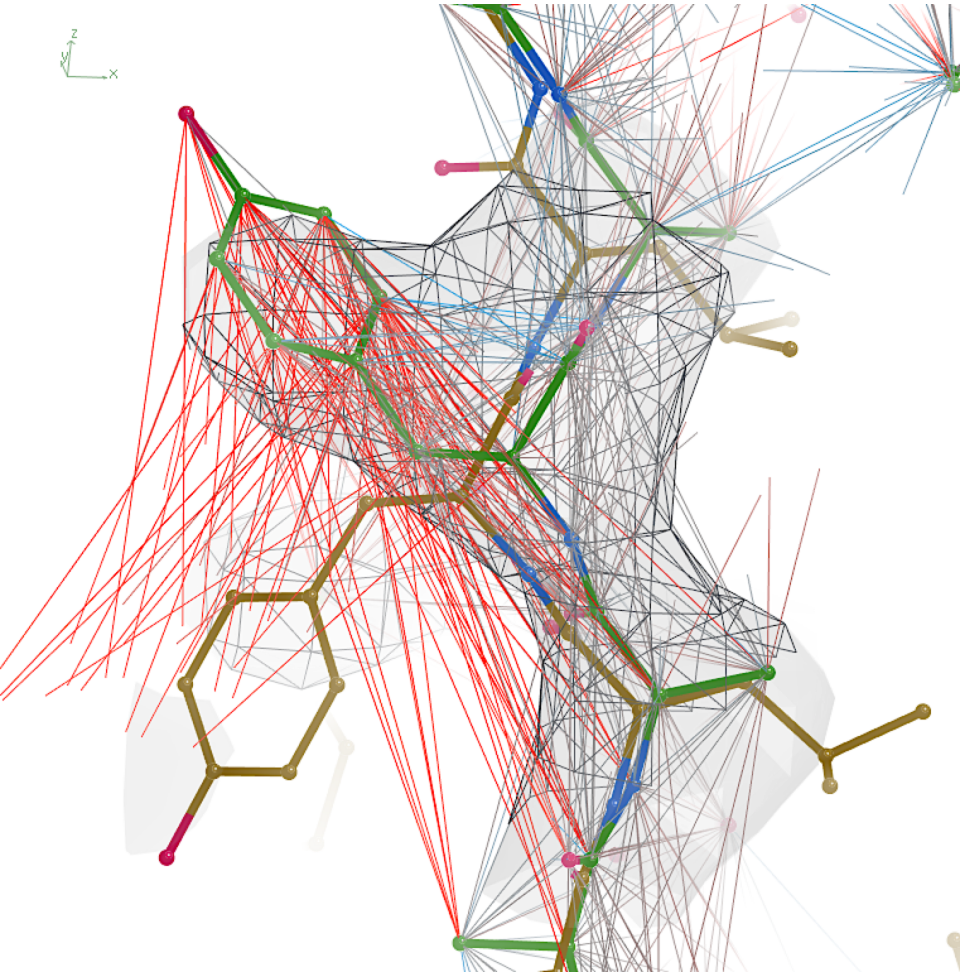
Red: long
Grey: similar
Blue: short

Robust Estimation

Cyan: original
Green: homolog
Yellow: refined

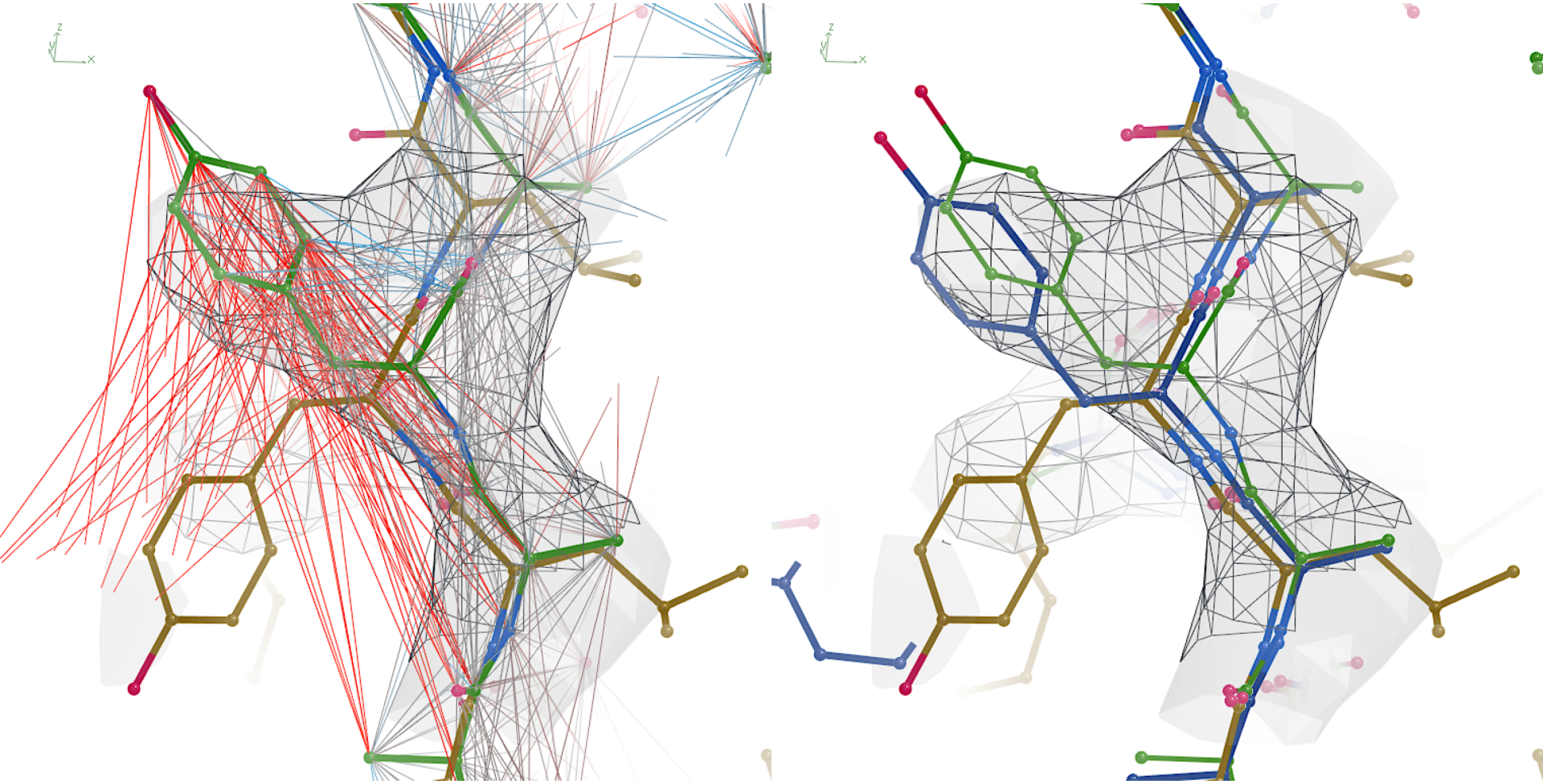
Red: long
Grey: similar
Blue: short

Robust Estimation



Green: original
Brown: homolog

Robust Estimation



Green: original
Brown: homolog
Blue: refined

Example: Ovotransferrin

1ryx – 3.5Å

Low-resolution refinement:

Weak signal
Noisy data

↓

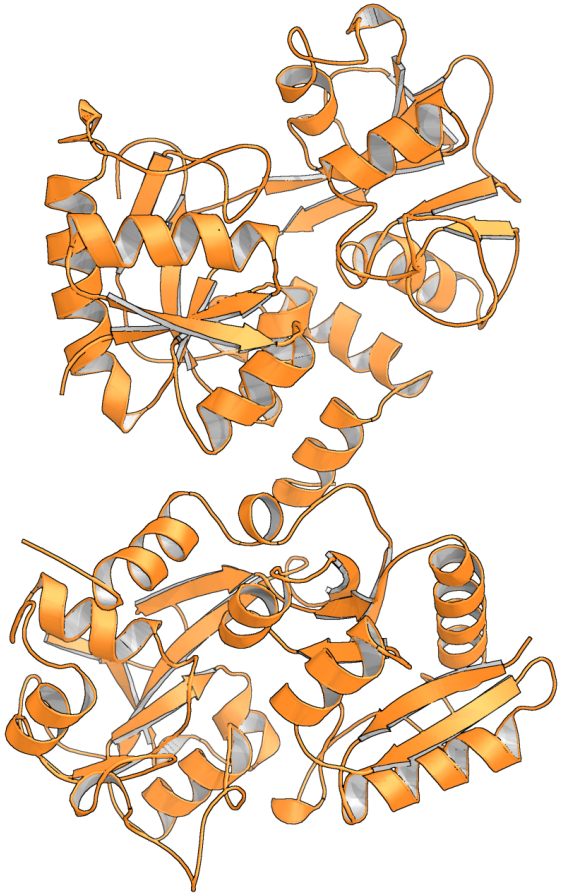
Unstable refinement

Result:
Poor quality model

Example: Ovotransferrin

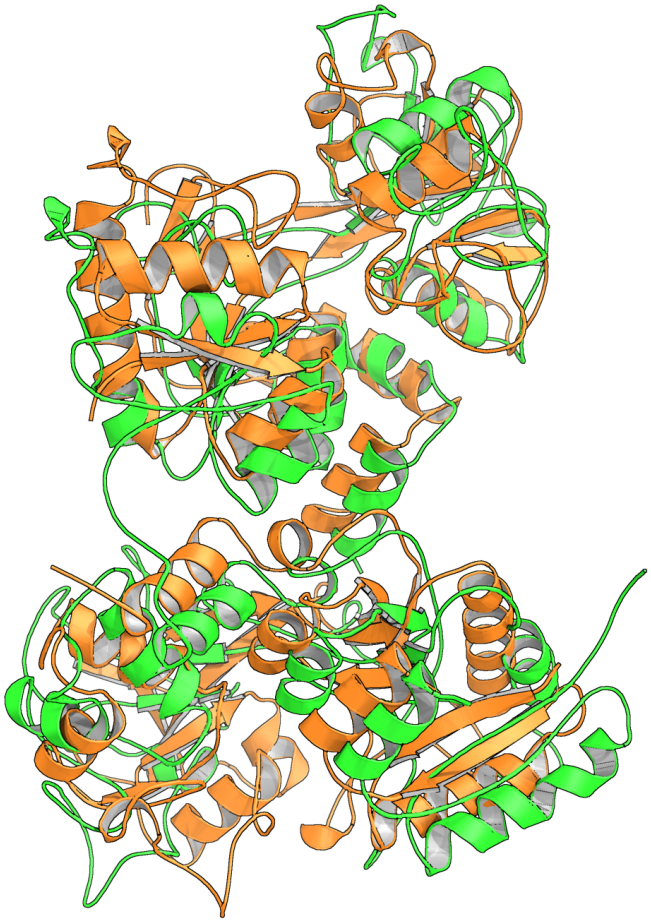
High-resolution homologue

1ryx – 3.5Å



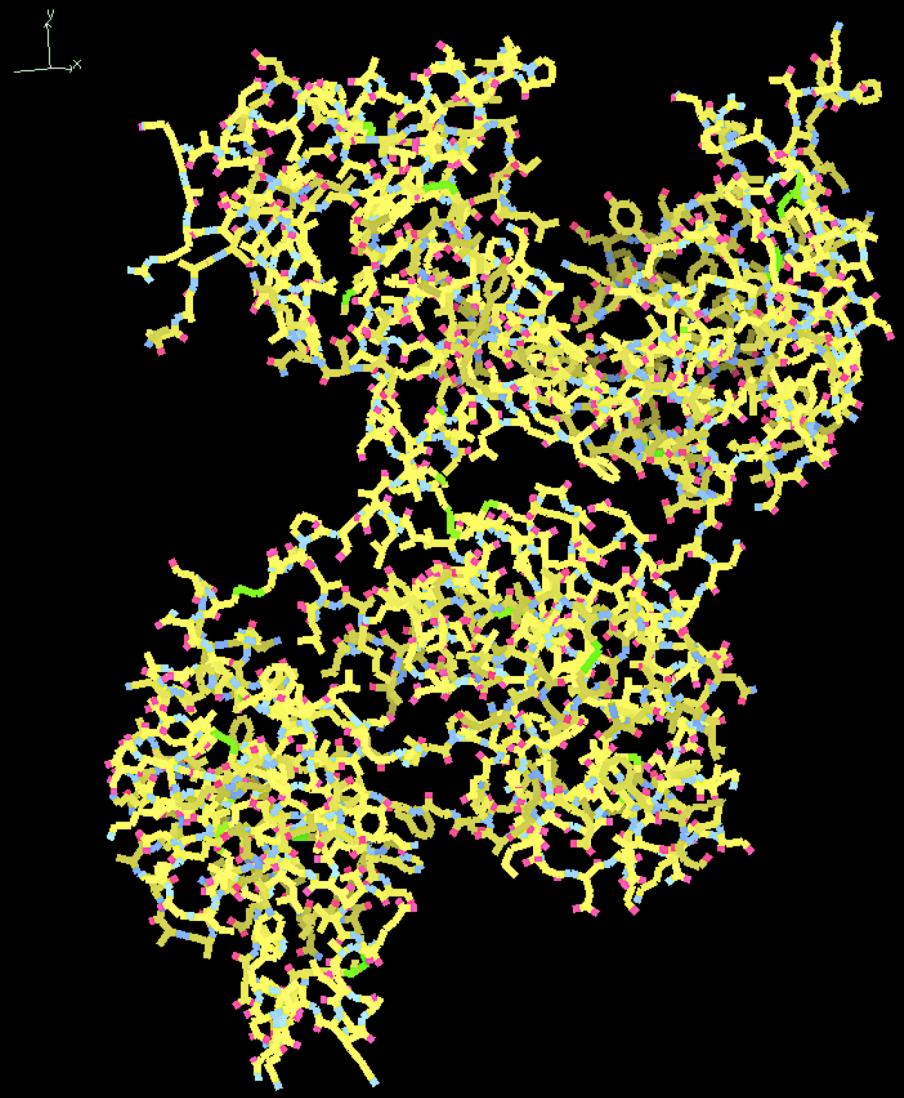
2d3i – 2.15Å

Example: Ovotransferrin



Models don’t superpose well

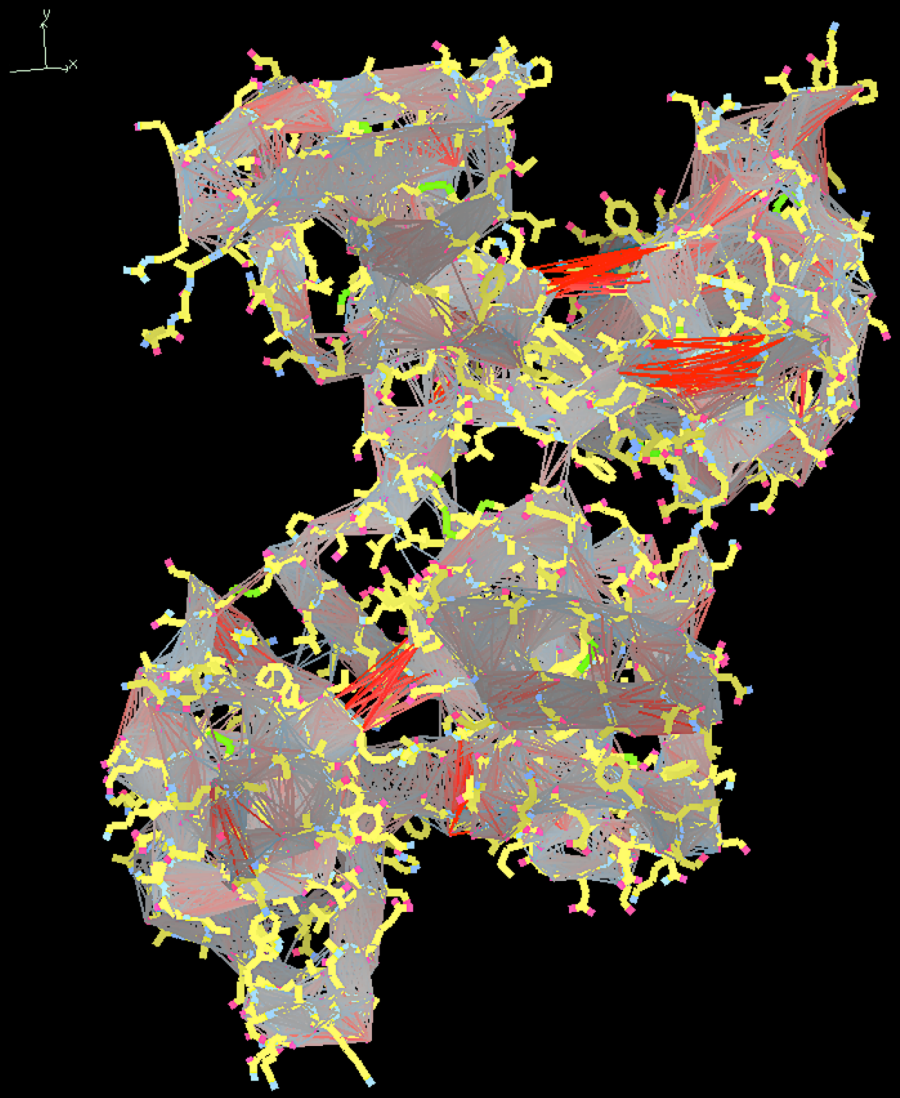
Example: Ovotransferrin



1ryx (3.5Å)

Example: Ovotransferrin

Restraints:
Backbone

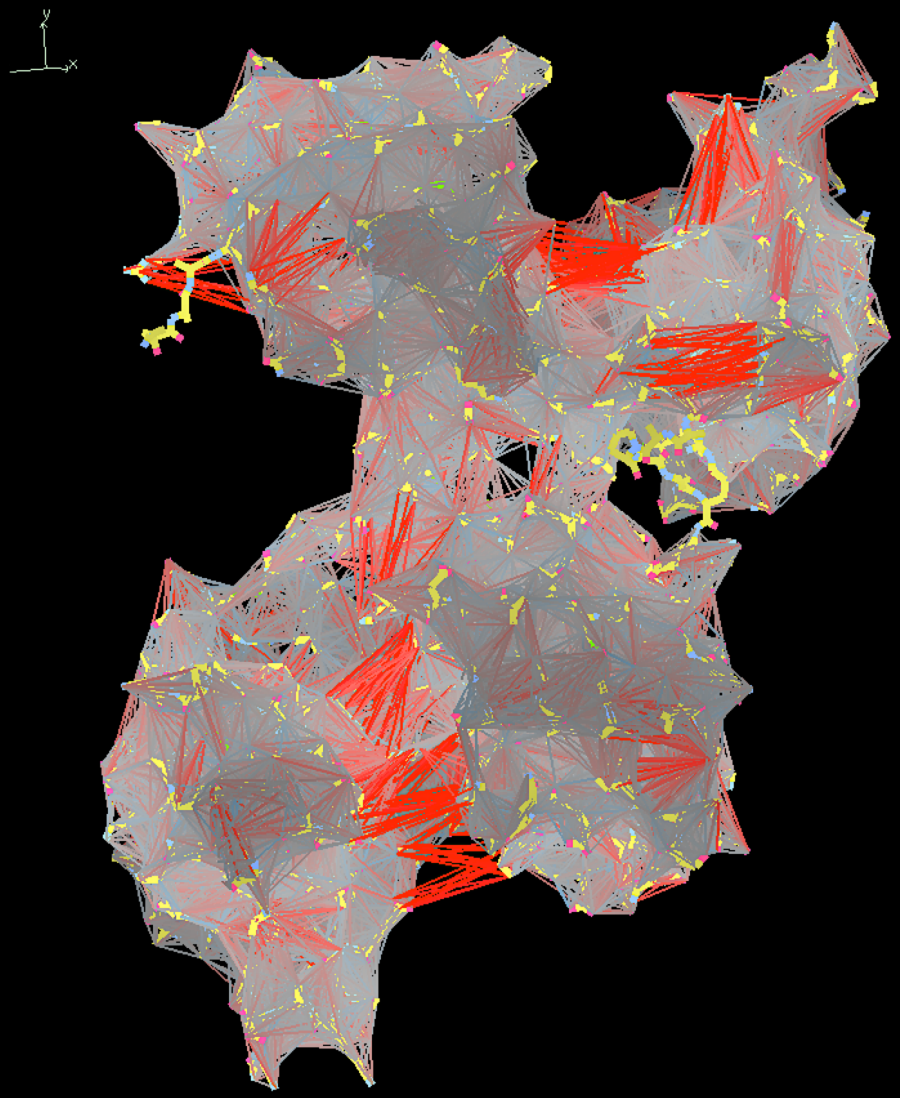


1ryx (3.5Å)
restrained to
2d3i (2.15Å)

Red: long
Grey: similar
Blue: short

Example: Ovotransferrin

Restraints:
- Backbone
- Side chains



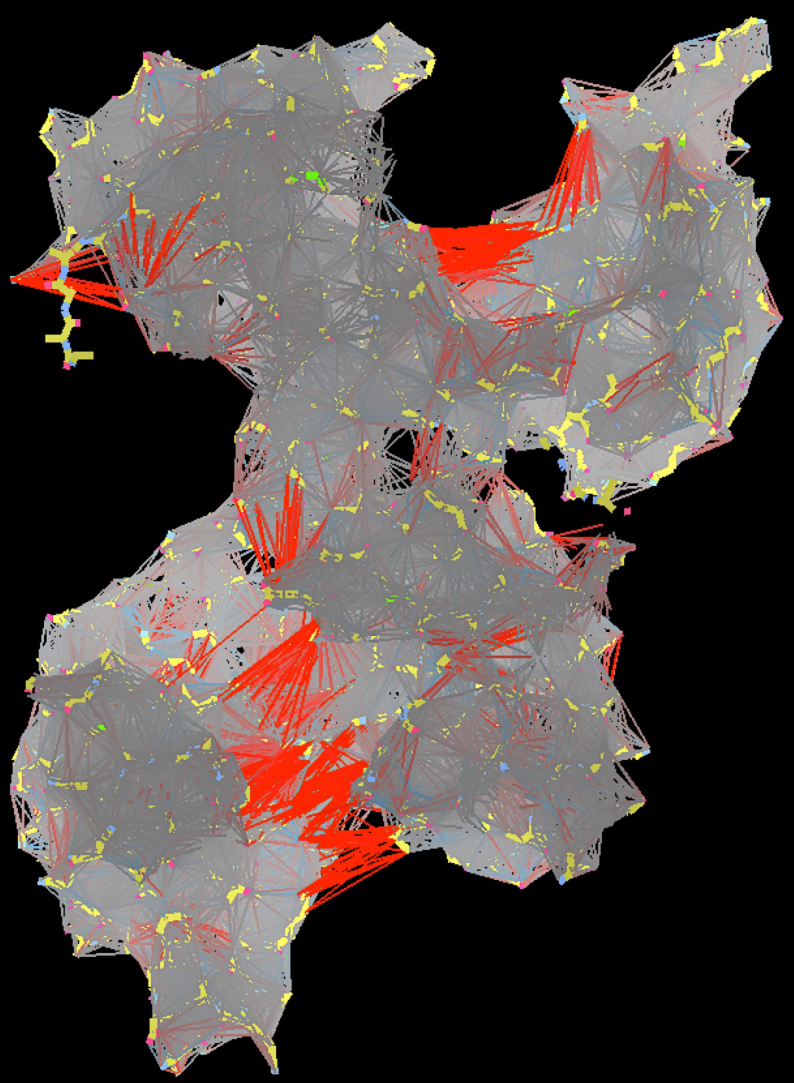
1ryx (3.5Å)
restrained to
2d3i (2.15Å)

Red: long
Grey: similar
Blue: short

Example: Ovotransferrin

Restraints:
- Backbone
- Side chains

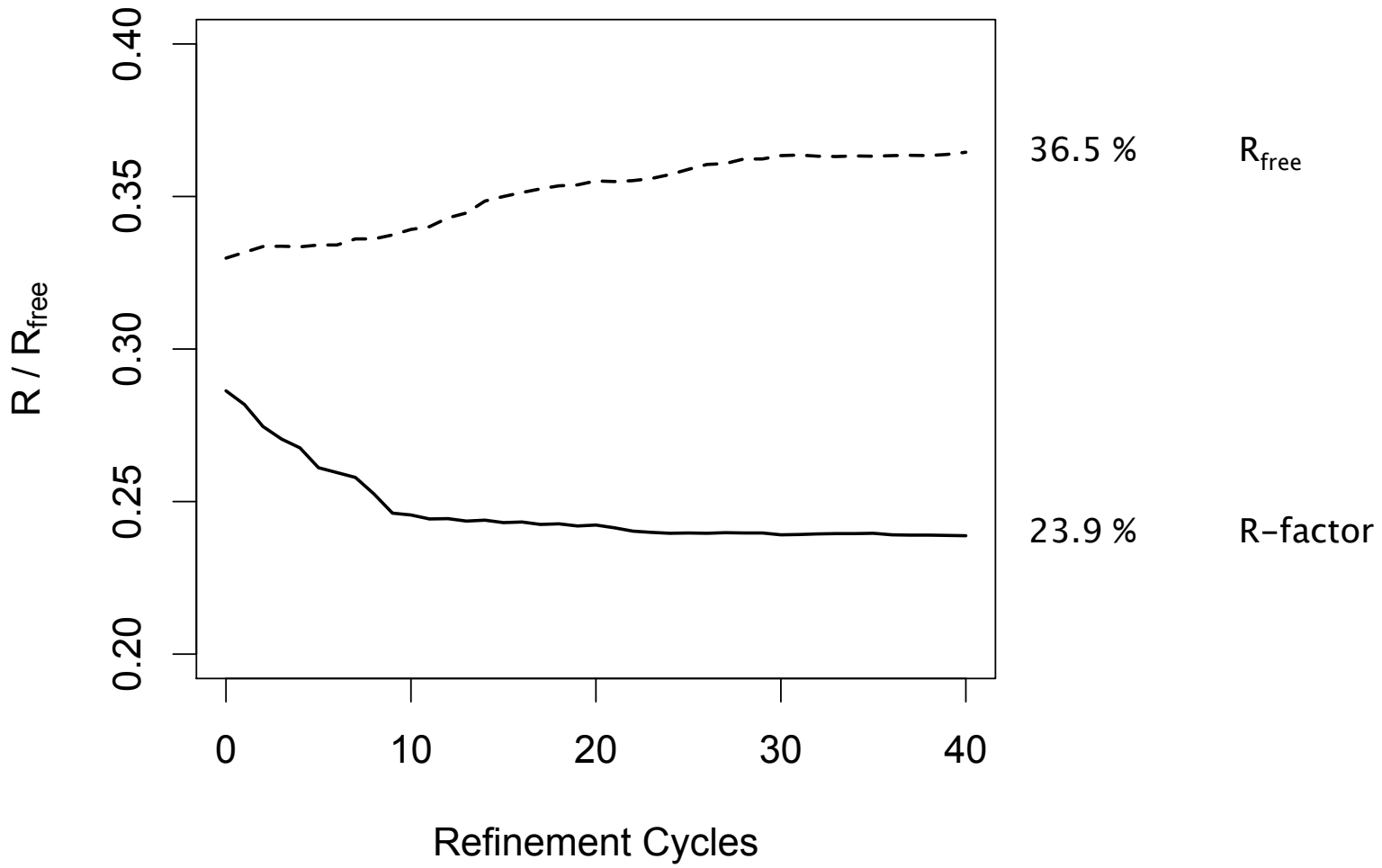
After re-refinement



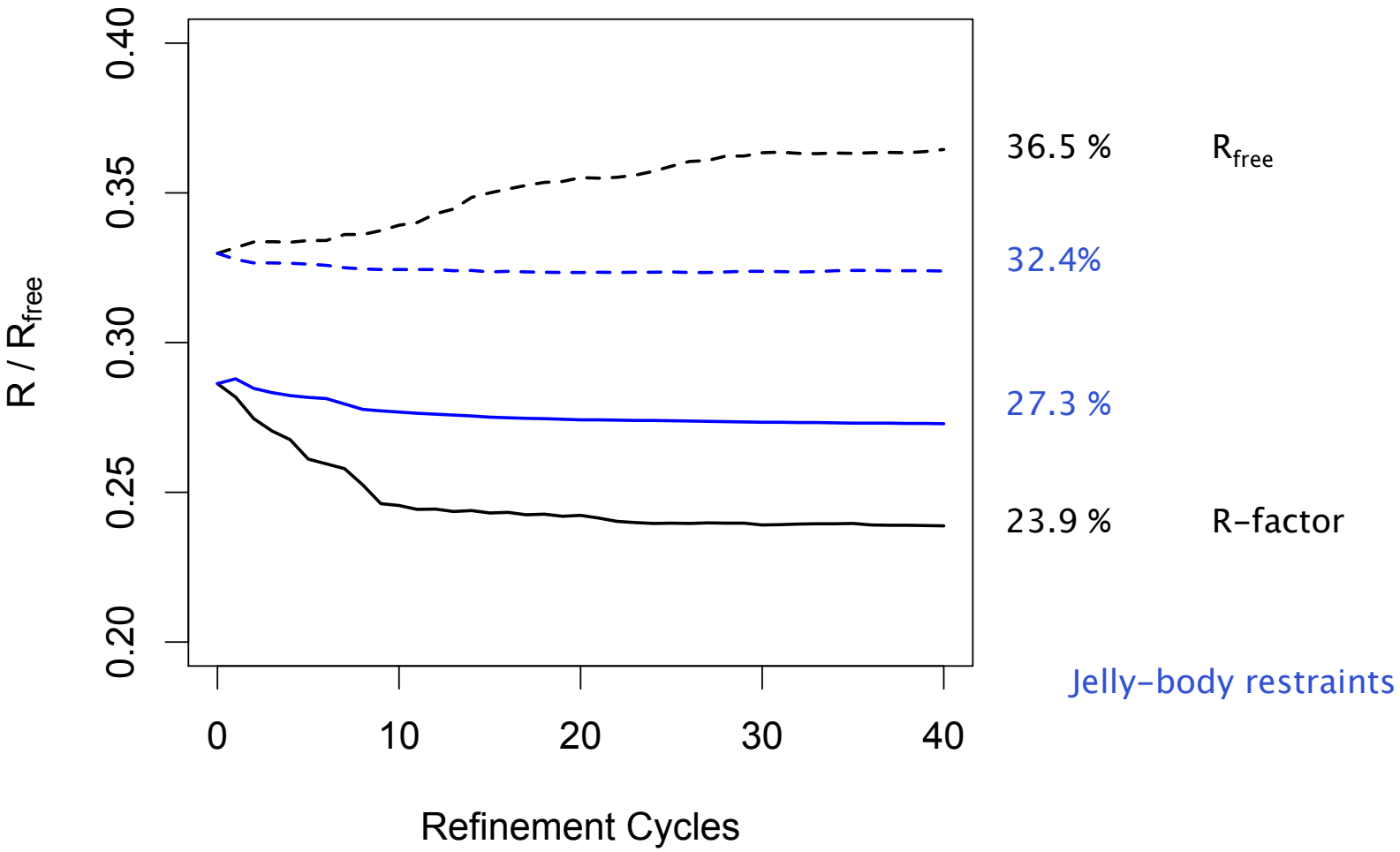
1ryx (3.5Å)
restrained to
2d3i (2.15Å)

Red: long
Grey: similar
Blue: short

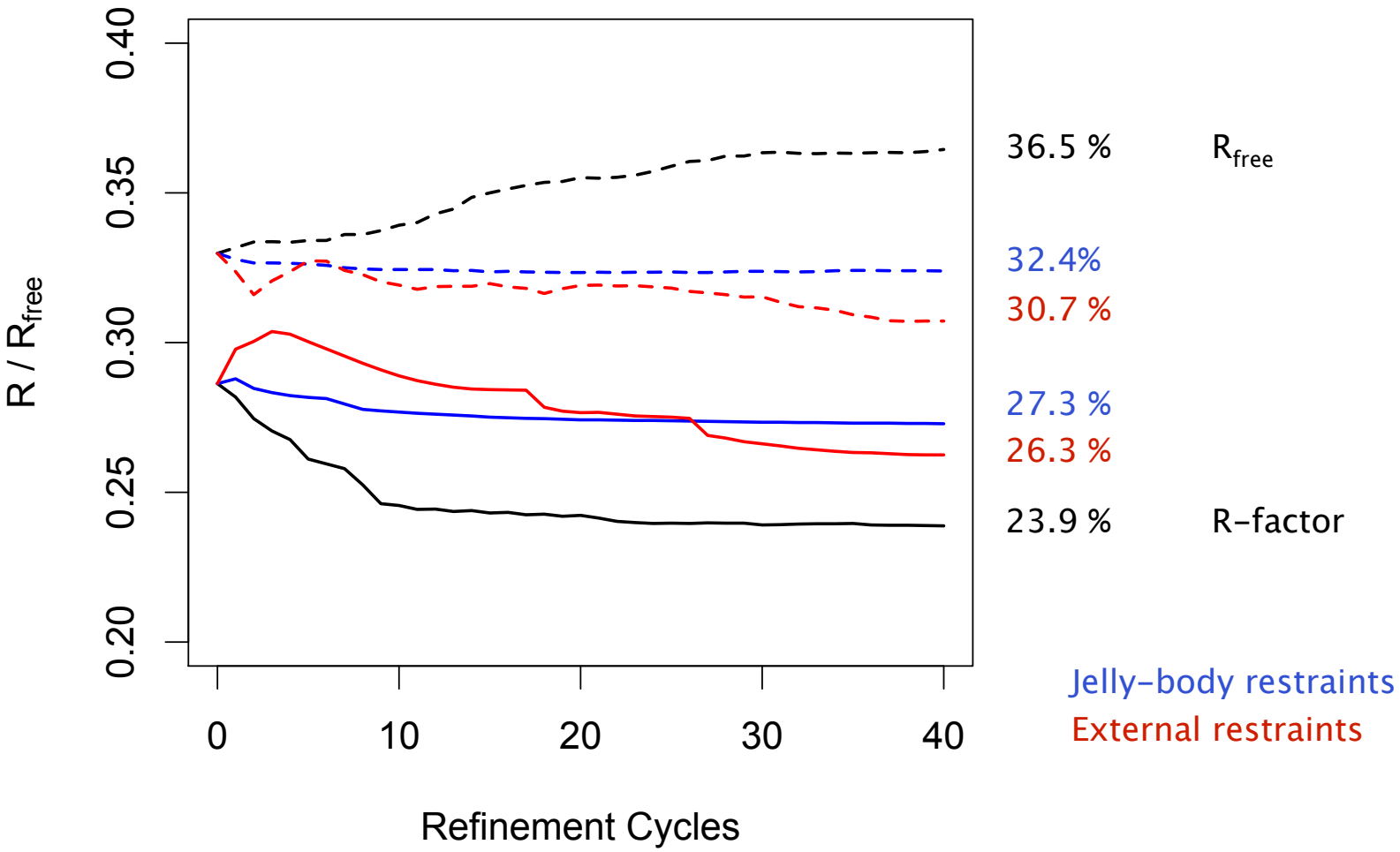
Example: Ovotransferrin



Example: Ovotransferrin



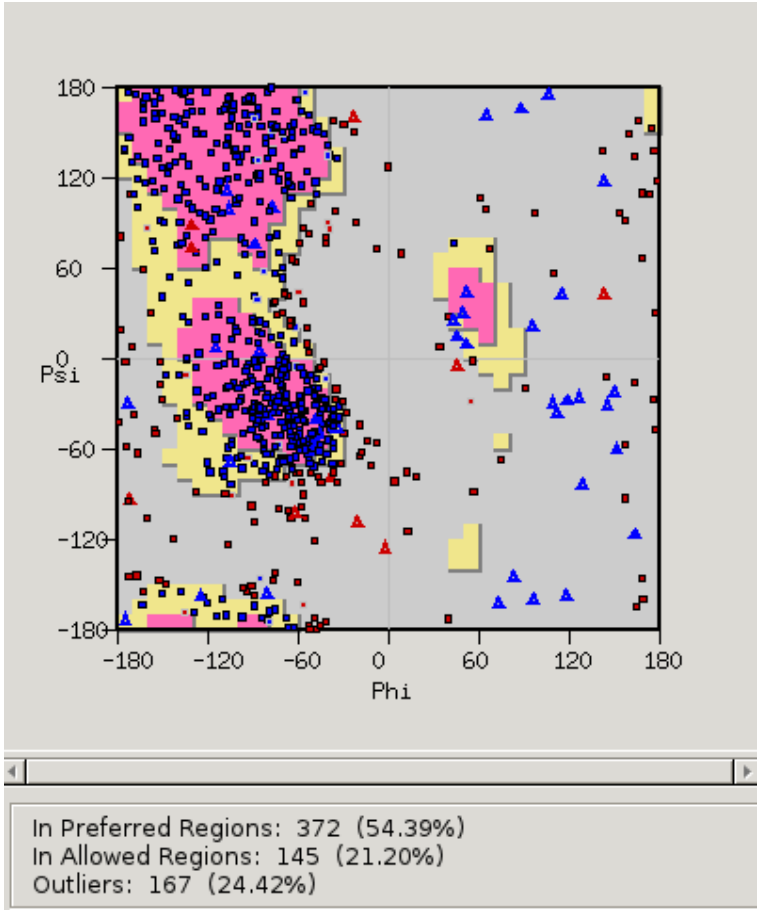
Example: Ovotransferrin



Example: Ovotransferrin

Original Structure

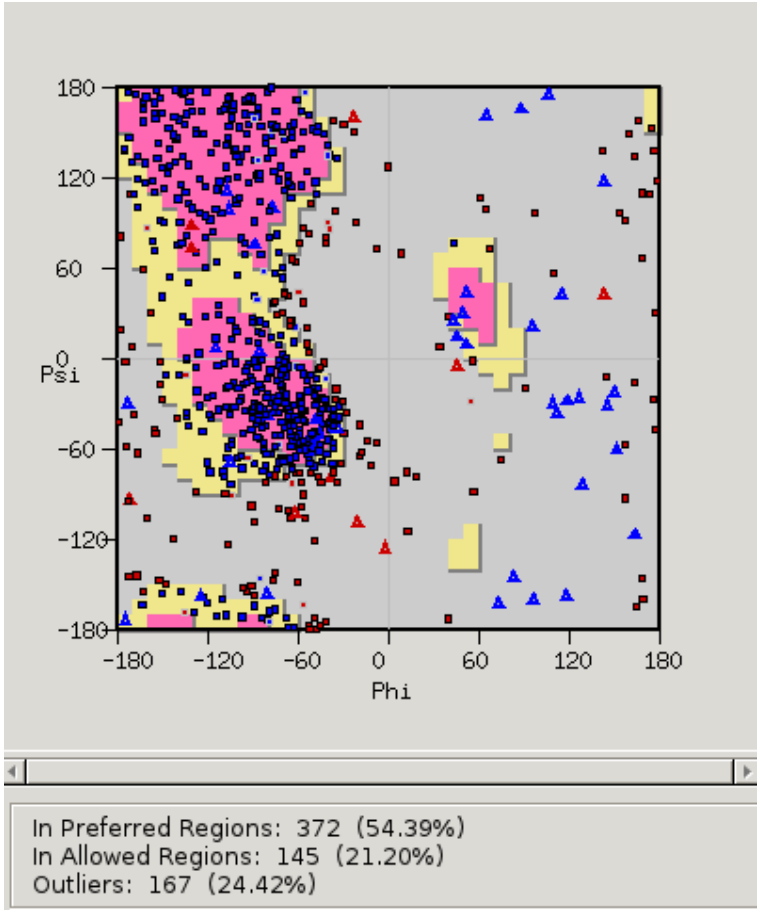
R/R_{free} : 0.286/0.330



Example: Ovotransferrin

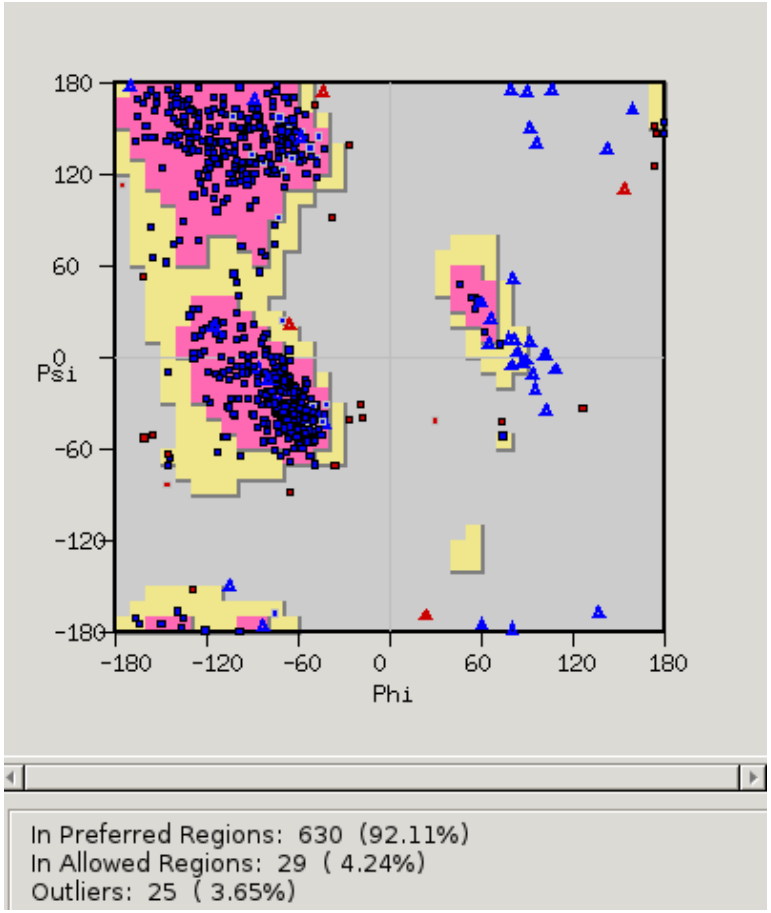
Original Structure

R/R_{free} : 0.286/0.330

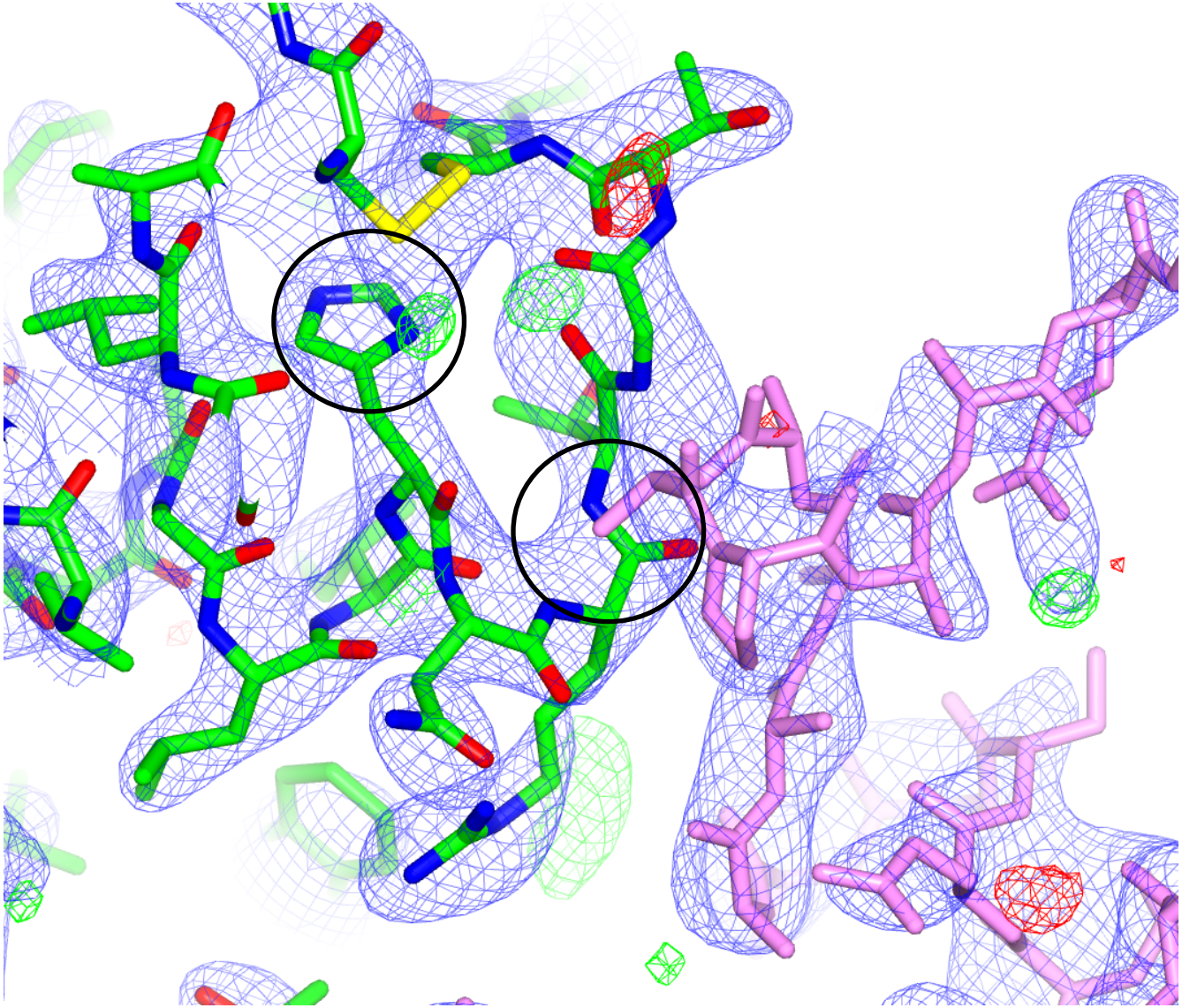


Re-refined with External Restraints

R/R_{free} : 0.263/0.307



Example: Ovotransferrin



1.3σ

Original Structure

R/R_{free} : 0.286/0.330

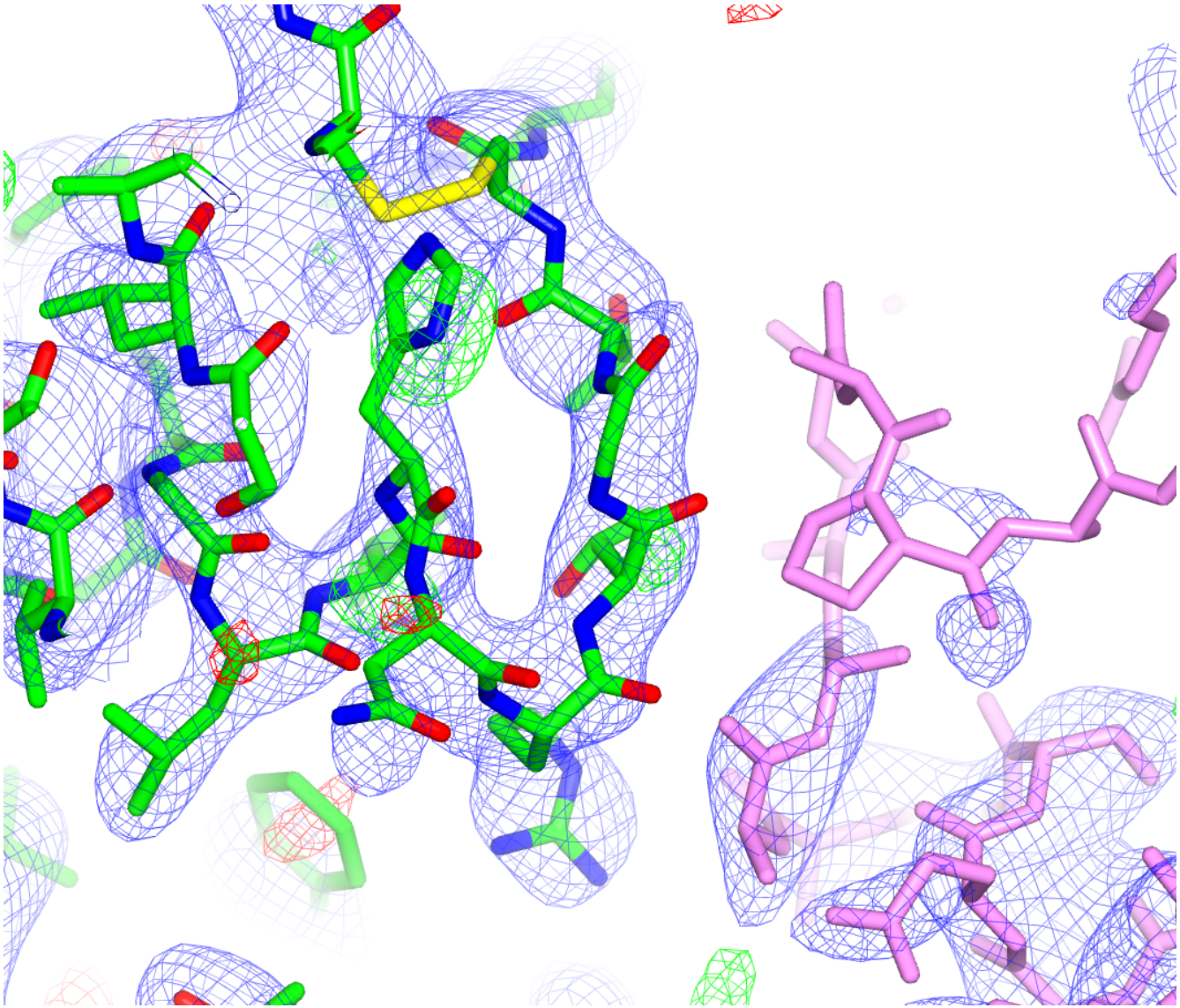
↓

External restraints

(40 cycles)

R/R_{free} : 0.263/0.307

Example: Ovotransferrin



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

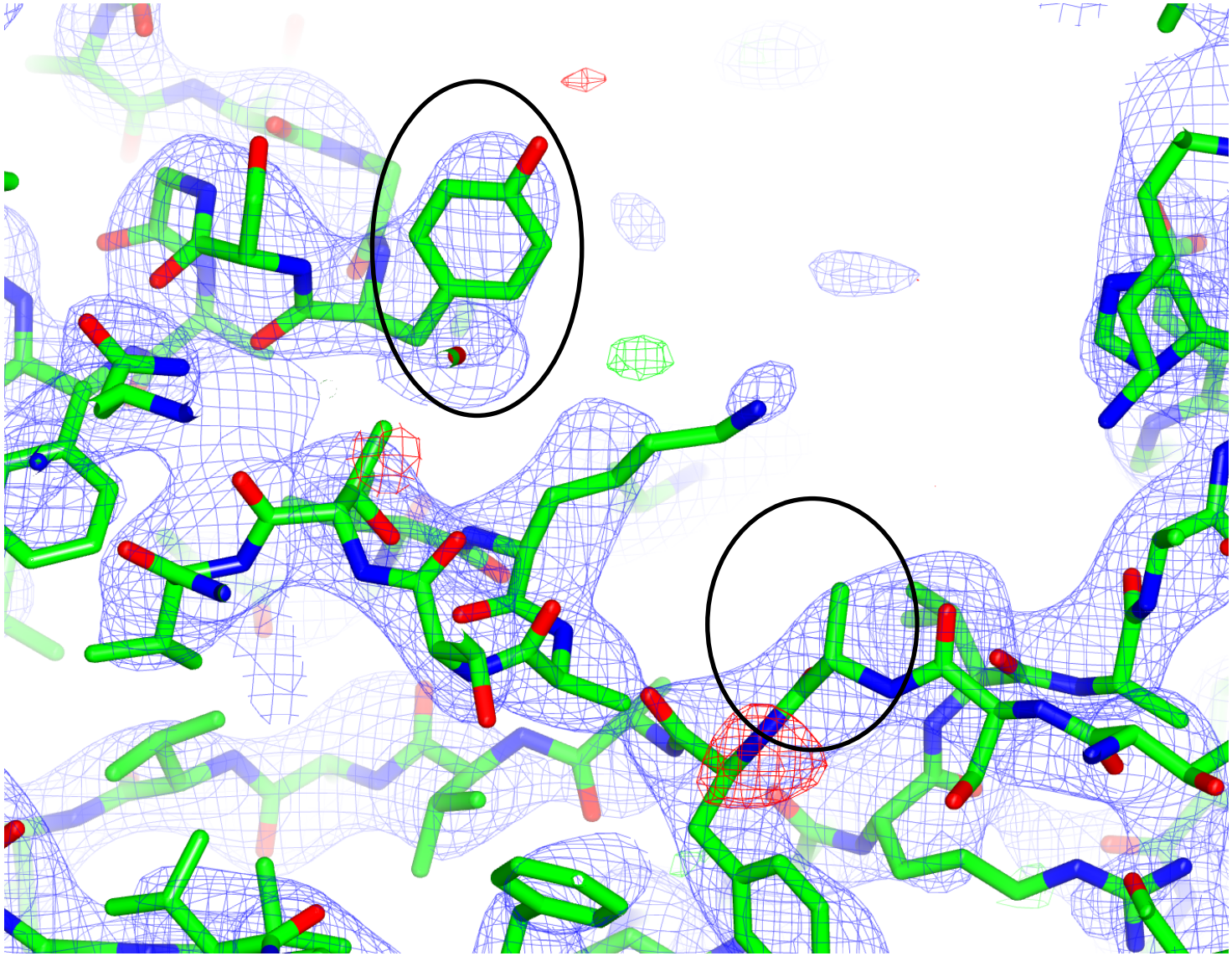
↓

Jelly body

(40 cycles)

R/R_{free} : 0.253/0.304

Example: Ovotransferrin



1.3σ

Original Structure

R/R_{free} : 0.286/0.330

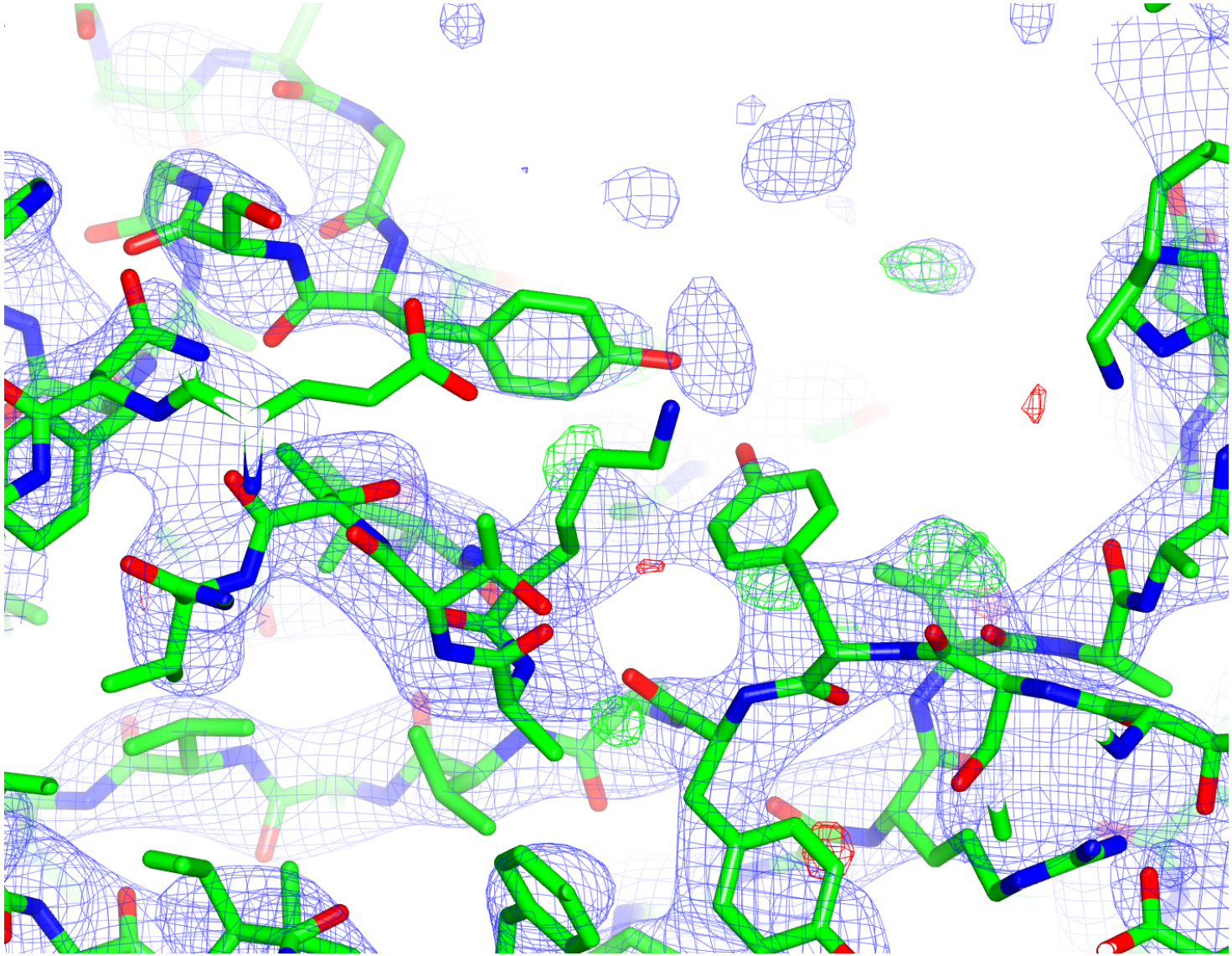
↓

External restraints

(40 cycles)

R/R_{free} : 0.263/0.307

Example: Ovotransferrin



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

Example: Ovotransferrin

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**

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

STRUCTURAL BIOLOGY

ISSN 2059-7983

Automated refinement of macromolecular structures at low resolution using prior information

Oleg Kovalevskiy, Robert A. Nicholls and Garib N. Murshudov*

MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, England. *Correspondence e-mail: garib@mrc-lmb.cam.ac.uk

Received 22 June 2016
Accepted 13 September 2016

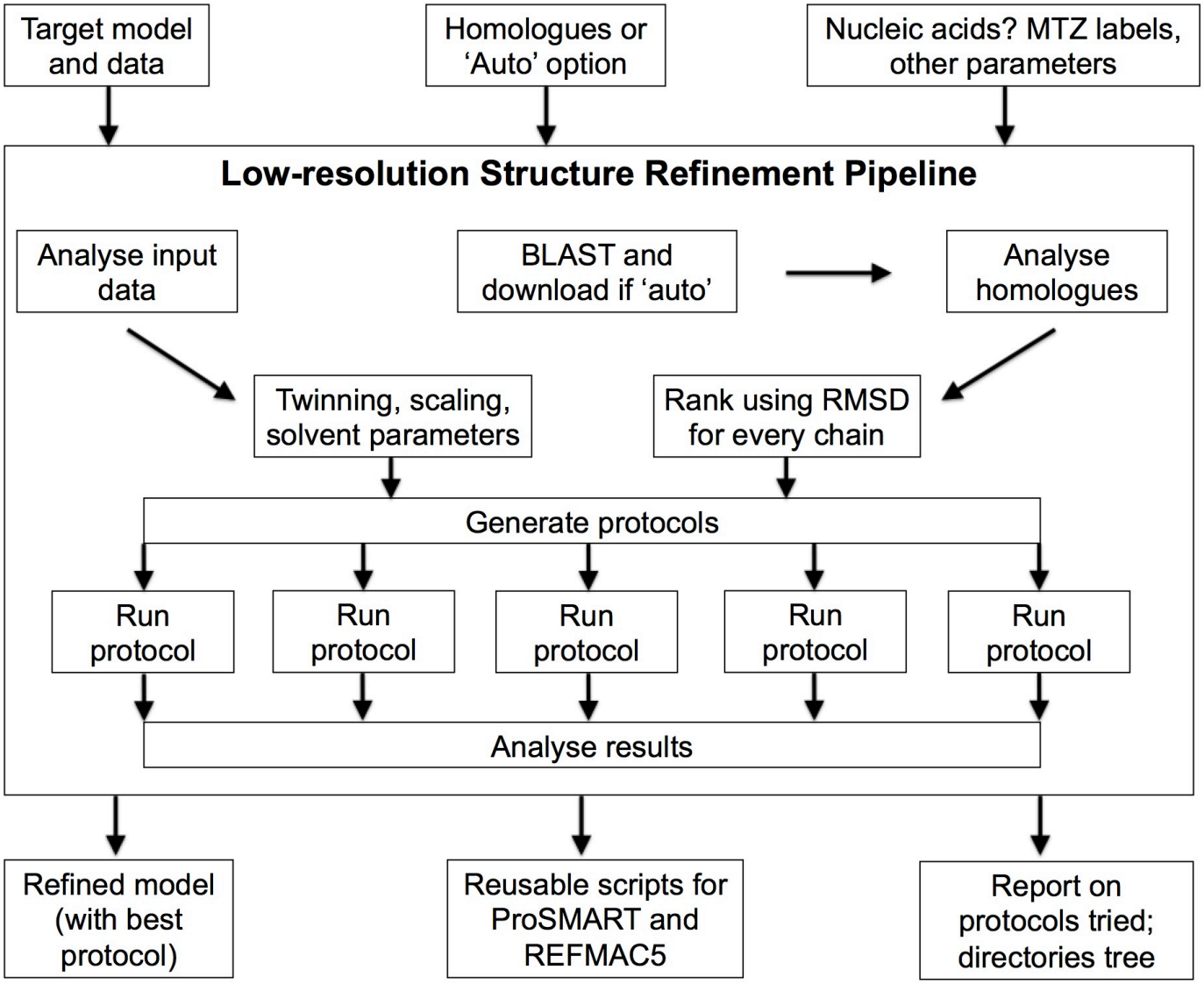
Edited by R. J. Read, University of Cambridge, England

Keywords: *REFMAC*5; *ProSMART*; low-resolution refinement; external restraints; *LORESTR*.

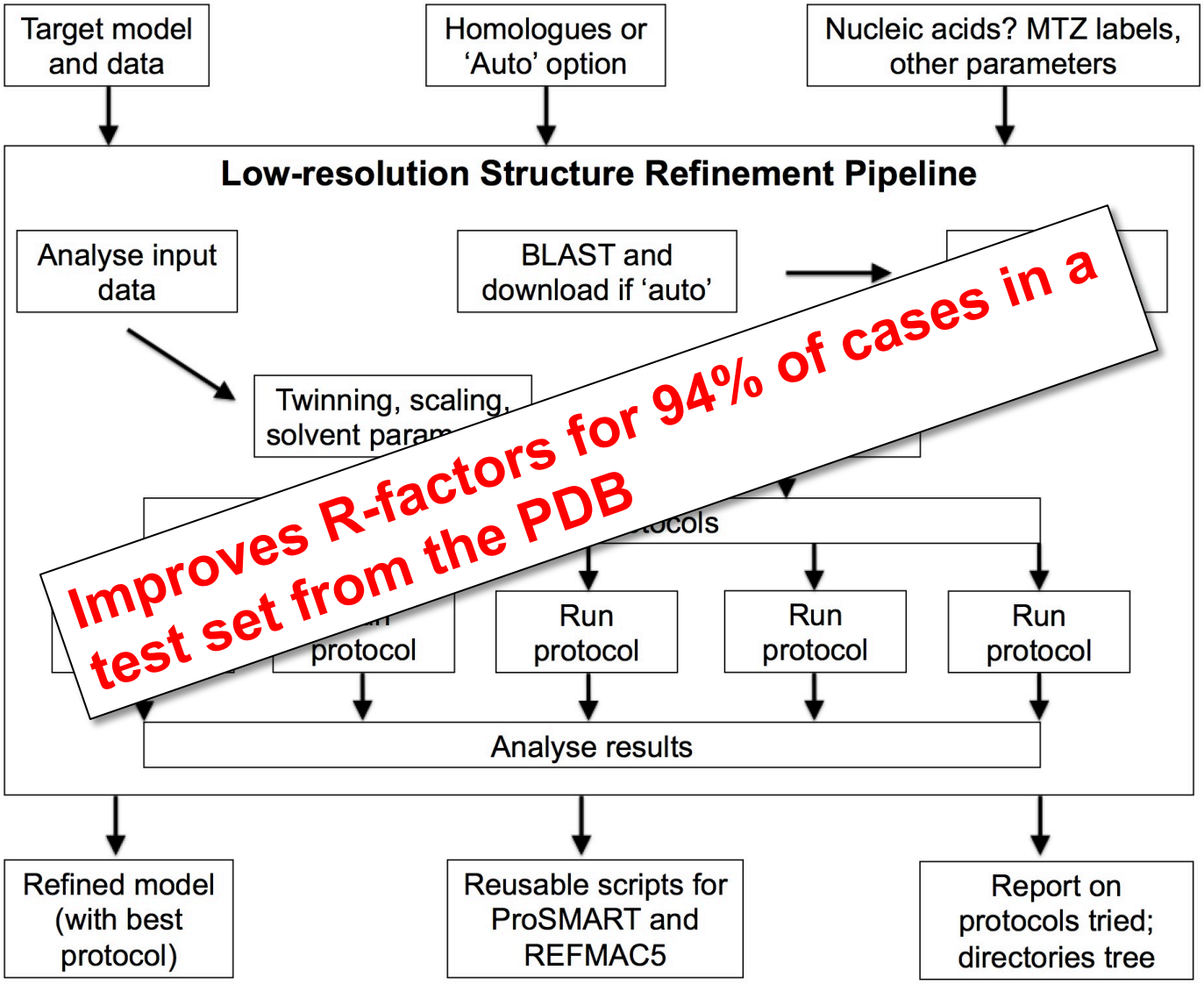
Supporting information: this article has supporting information at journals.iucr.org/d

Since the ratio of the number of observations to adjustable parameters is small at low resolution, it is necessary to use complementary information for the analysis of such data. *ProSMART* is a program that can generate restraints for macromolecules using homologous structures, as well as generic restraints for the stabilization of secondary structures. These restraints are used by *REFMAC*5 to stabilize the refinement of an atomic model. However, the optimal refinement protocol varies from case to case, and it is not always obvious how to select appropriate homologous structure(s), or other sources of prior information, for restraint generation. After running extensive tests on a large data set of low-resolution models, the best-performing refinement protocols and strategies for the selection of homologous structures have been identified. These strategies and protocols have been implemented in the *Low-Resolution Structure Refinement* (*LORESTR*) pipeline. The pipeline performs auto-detection of twinning and selects the optimal scaling method and solvent parameters. *LORESTR* can either use user-supplied homologous structures, or run an automated *BLAST* search and download homologues from the PDB. The pipeline executes multiple model-refinement instances using different parameters in order to find the best protocol. Tests show that the automated pipeline improves *R* factors, geometry and Ramachandran statistics for 94% of the low-resolution cases from the PDB included in the test set.

Automated pipeline - LORESTR

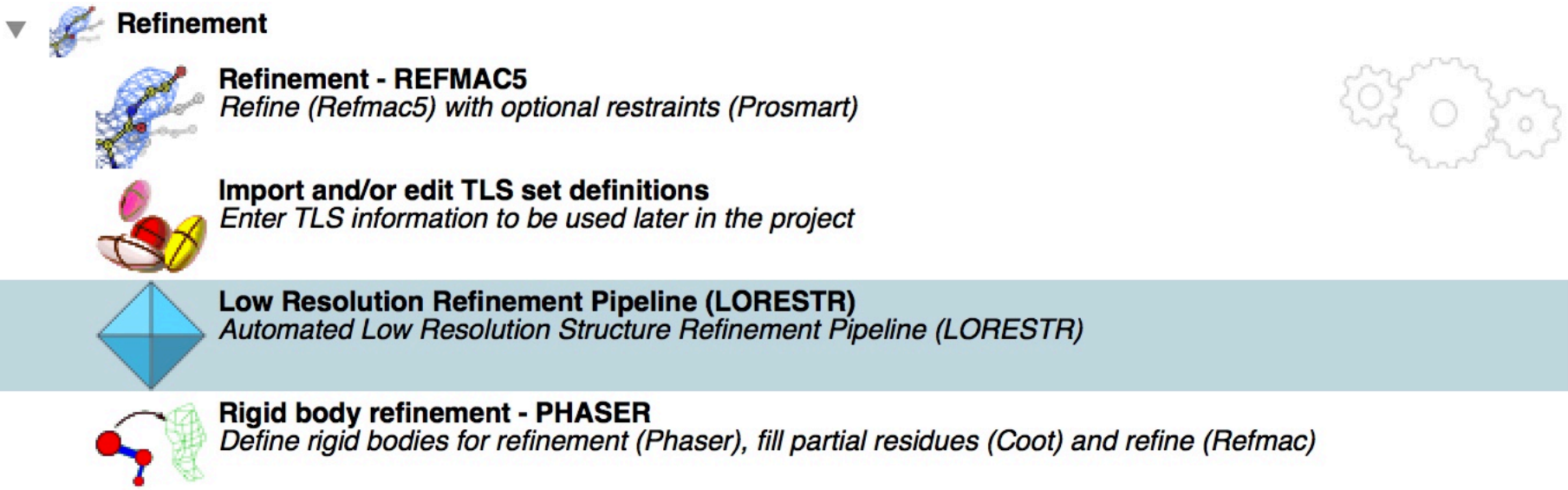


Automated pipeline - LORESTR



Automated pipeline - LORESTR

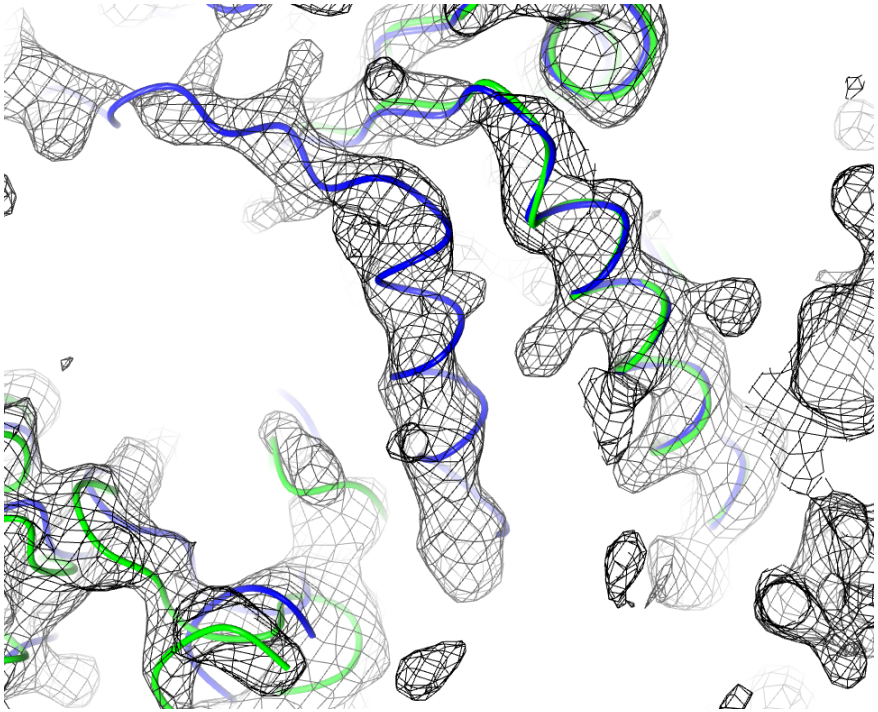
Available in CCP4i2:



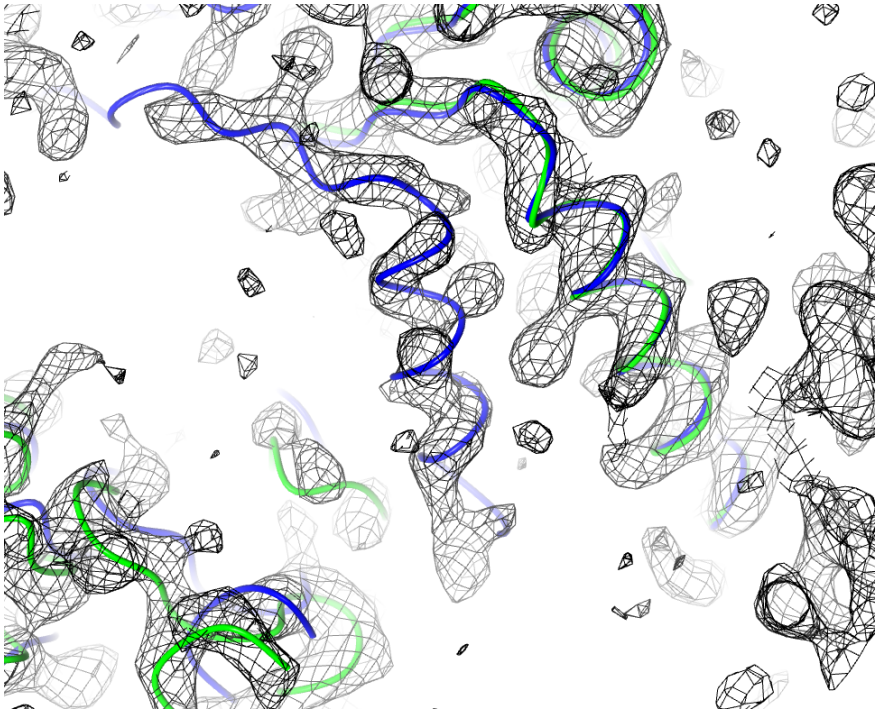
REFMAC Anisotropic Map Sharpening

Idea – remove an overall B value

Original Map



Sharpened map from REFMAC



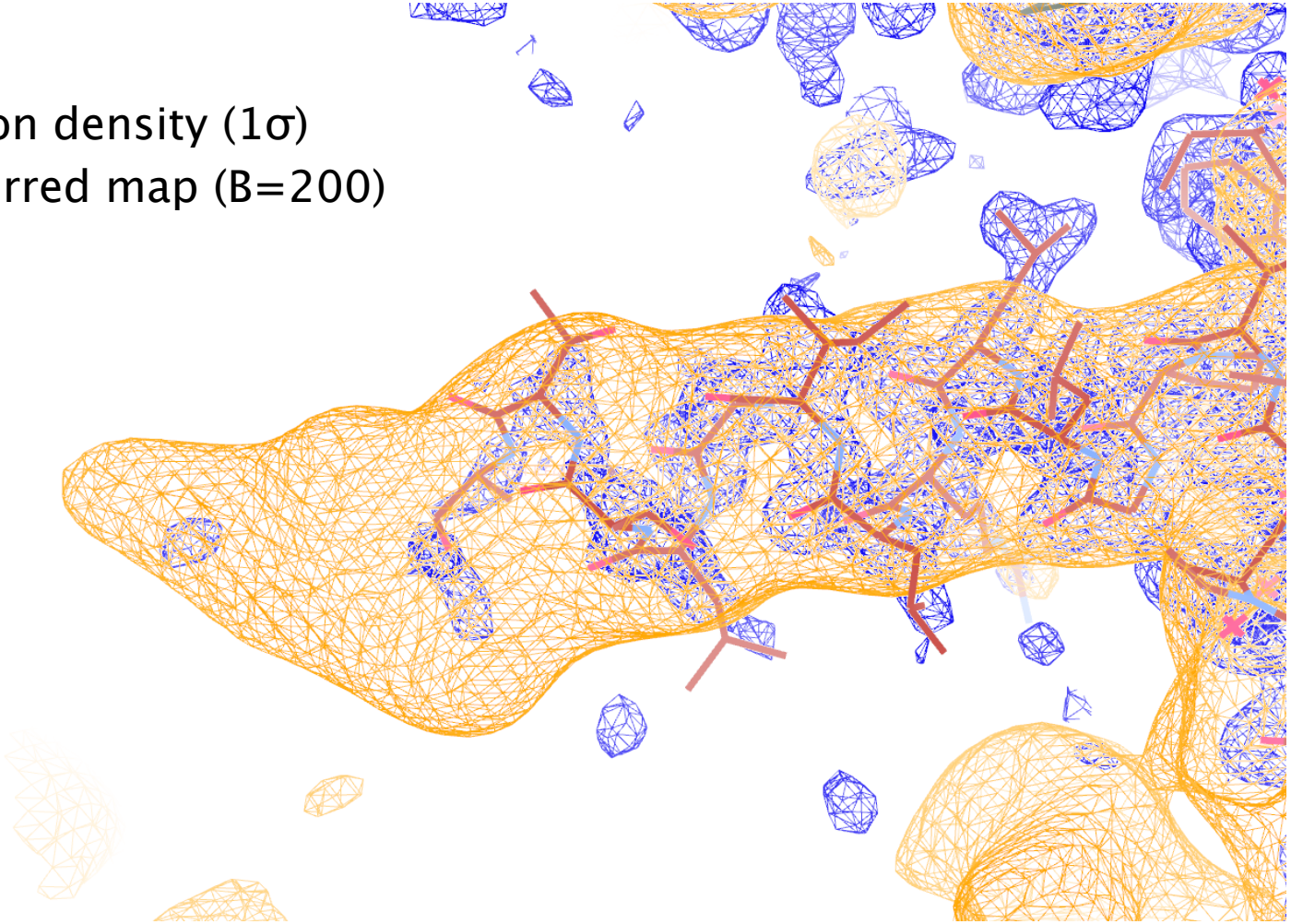
Green: original structure 2r6c (4.0Å) – helix unmodelled

Blue: homologous structure 2r6a (2.9Å)

Map Blurring

Idea – apply an overall B value

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



1s3e

Summary

External restraints to homologous structures can be useful

- Used by REFMAC5 for full-model refinement
- Visualised in Coot, and used for real space refinement
- If homologs are not available, use:
  - Generic h-bond restraints for protein
  - Generic base-pair/stacking restraints for DNA/RNA

Restrained refinement can be used for low-res & cryo-EM refinement

- Need lots of “extra” restraints to regularise refinement
- Jelly-body restraints are almost always needed

Other things to think about in cryo-EM:

- Multiple levels of blurring/sharpening helps, but care is needed
- Box size should be selected appropriately
- Divide & conquer pipeline for large models

Summary

Tools to help with model building and refinement:

REFMAC: Refinement, jelly body restraints, map sharpening/blurring

ProSMART: External restraints, comparative analysis

LIBG: Nucleic acid restraints

COOT: Visualisation & manipulation of restraints, map blurring
…also morphing, jiggle-fit, backrub rotamers…

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

What and When – Low-Res MX

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 and DNA/RNA 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

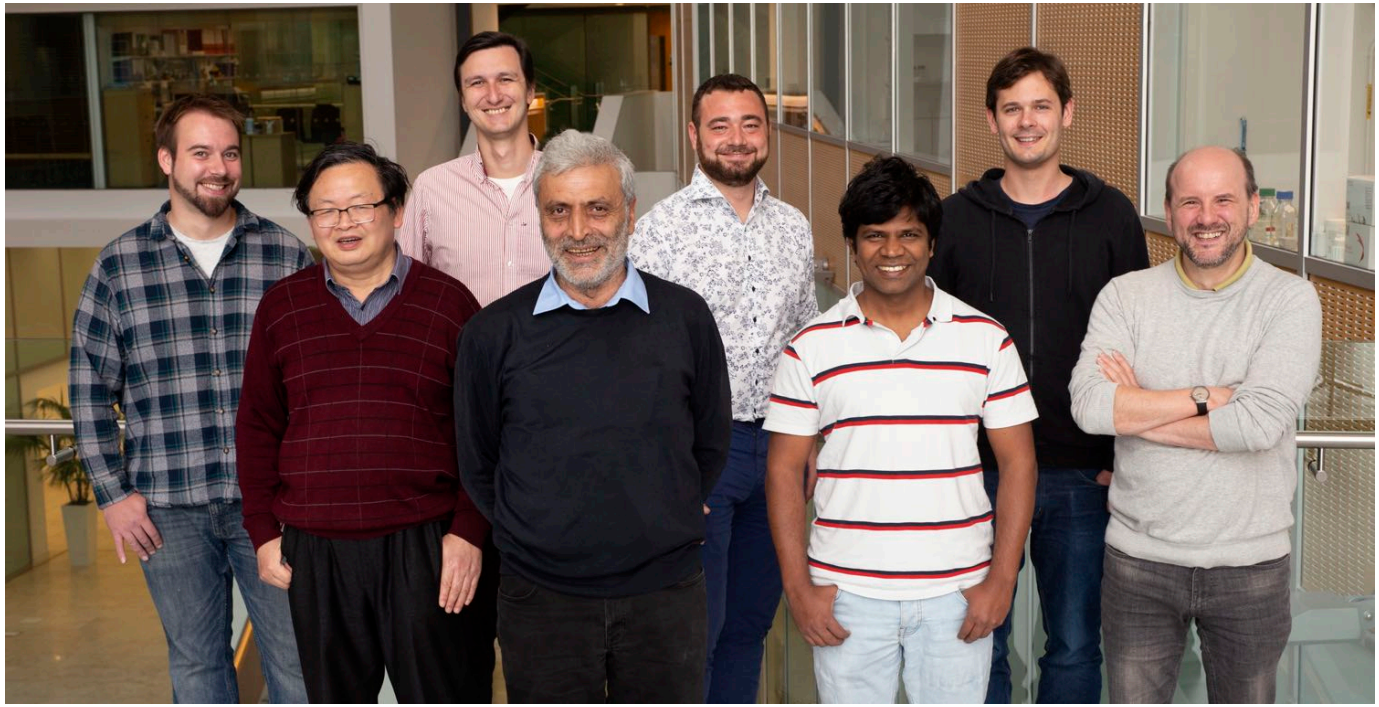
References

Low-resolution refinement with REFMAC, ProSMART & LORESTR:

- Kovalevskiy *et al.* (2018) Overview of refinement procedures within REFMAC5: Utilising Data from Different Sources. Acta Cryst D74.
- Nicholls *et al.* (2017) Low Resolution Refinement of Atomic Models Against Crystallographic Data. Protein Crystallography, 565-93.
- Kovalevskiy *et al.* (2016) Automated refinement of macromolecular structures at low resolution using prior information. *Acta Cryst.* D72, 1149-61.
- Nicholls *et al.* (2014) Conformation-Independent Structural Comparison of Macromolecules with ProSMART. *Acta Cryst.* D70, 2487-99.
- 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.
- Murshudov *et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.* D67, 355-67.

Acknowledgements

MRC-LMB Computational Structural Biology Group



Left to right:

Rob Nicholls
Fei Long
Oleg Kovalevskiy
Garib Murshudov
Michal Tykac
Rangana Warshamanage
James Parkhurst
Paul Emsley

MRC | Laboratory of Molecular Biology

CCP4