

Low resolution refinement in the program - REFMAC

**Garib N Murshudov
MRC-LMB
Cambridge**

Contents

- 1. Twin in REFMAC**
- 2. Problems of low-resolution refinement and regularisation**
- 3. Some tools for low resolution: ncs, external restraints,
“jelly” body**
- 4. ProSmart**
- 5. Map sharpening: General approach and some applications**

Available refinement programs

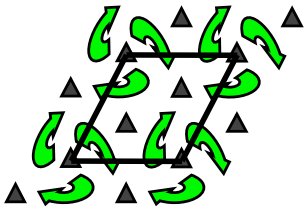
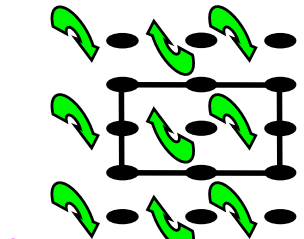
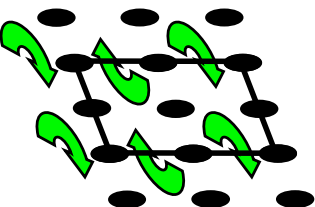
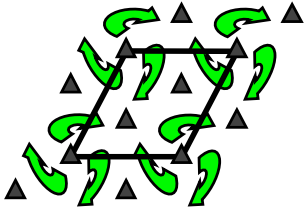
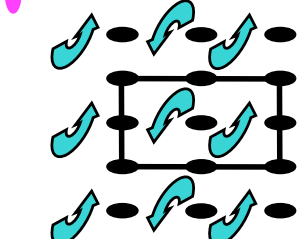
- SHELXL
- CNS
- REFMAC5
- TNT
- BUSTER/TNT
- Phenix.refine
- RESTRAINT
- MOPRO
- XD
- MAIN

What can REFMAC do?

- Simple maximum likelihood restrained refinement
- Twin refinement: Warning
- Phased refinement (with Hendrickson-Lattmann coefficients)
- SAD/SIRAS refinement
- Structure idealisation
- Library for more than 10000 ligands (from the next version)
- Covalent links between ligands and ligand-protein
- Rigid body refinement
- NCS local, restraints to external structures, jelly body
- TLS refinement
- Map sharpening: Inverse problem, Bvalues etc
- etc

TWIN

merohedral and pseudo-merohedral twinning

Crystal symmetry:	P3	P2	P2
Constrain:	-	$\beta = 90^\circ$	-
Lattice symmetry *: (rotations only)	P622	P222	P2
Possible twinning:	merohedral	pseudo-merohedral	-
Domain 1			
Twinning operator			-
Domain 2			-

Crystal lattice is invariant with respect to twinning operator.

The crystal is NOT invariant with respect to twinning operator.

Twin refinement in REFMAC

Twin refinement in refmac (5.5 or later) is automatic.

- Identify “twin” operators
- Calculate “Rmerge” ($\sum |I_h - \langle I \rangle_{\text{twin}}| / \sum I_h$) for each operator. If $R_{\text{merge}} > 0.44$ remove it: Make sure that twin plus crystal symmetry operators forms a group
- Refine twin fractions. Keep only “significant” domains (default threshold is 7%): Twin plus symmetry operators should form a group

Intensities can be used

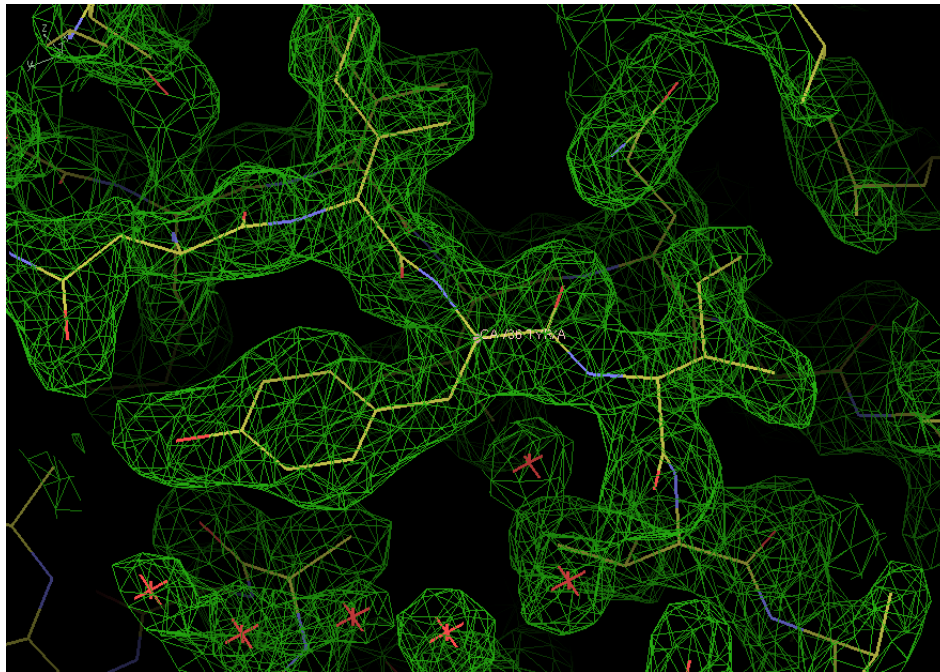
If phases are available they can be used

Maximum likelihood refinement is used

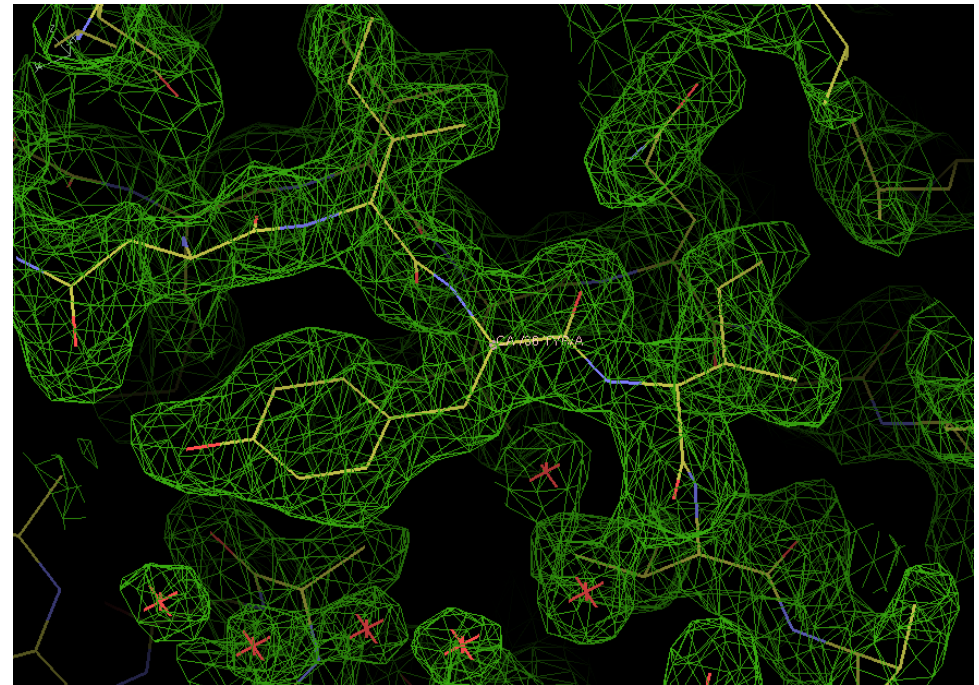
Electron density: 1jrg

Warning: Usually twin refinement reduces R factors
but electron density does not improve much

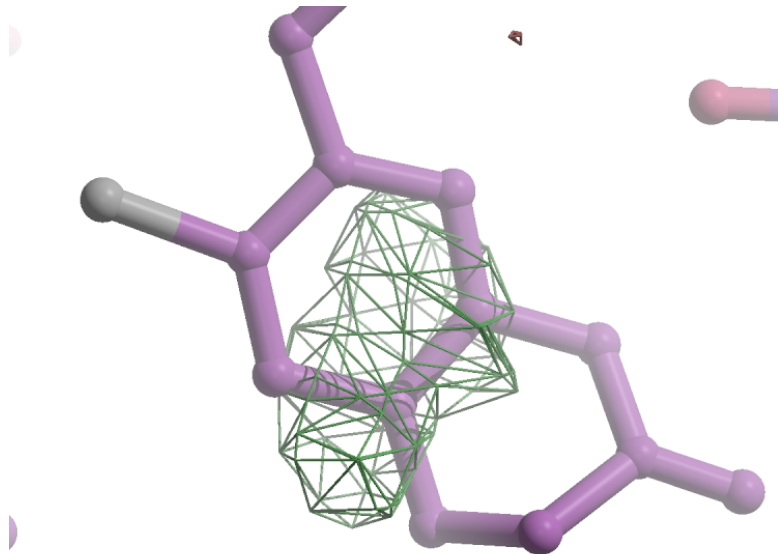
“non twin” map



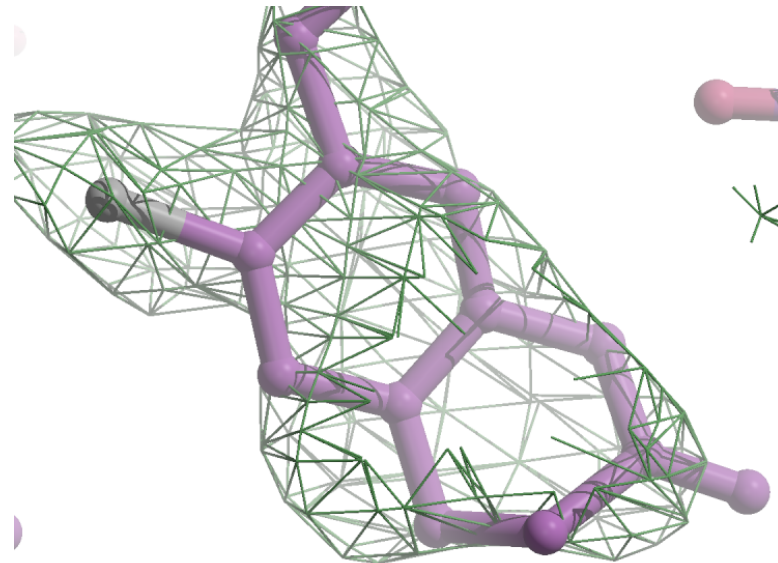
“twin” map



Where's the density for my ligand (2.15Å)?
(Slide is provided by B.Bax)



R-factor (R-free) 25.5%
(26.9%) – after initial rigid
body and restrained
refinement.
Fo-Fc – 3 sigma (does not look better at lower
sigma)



R-factor (R-free) 15.9% (16.3%)
– re-run restrained ref. with
twin on (refined twin fractions
0.6043/0.3957).
Fo-Fc – 3 sigma

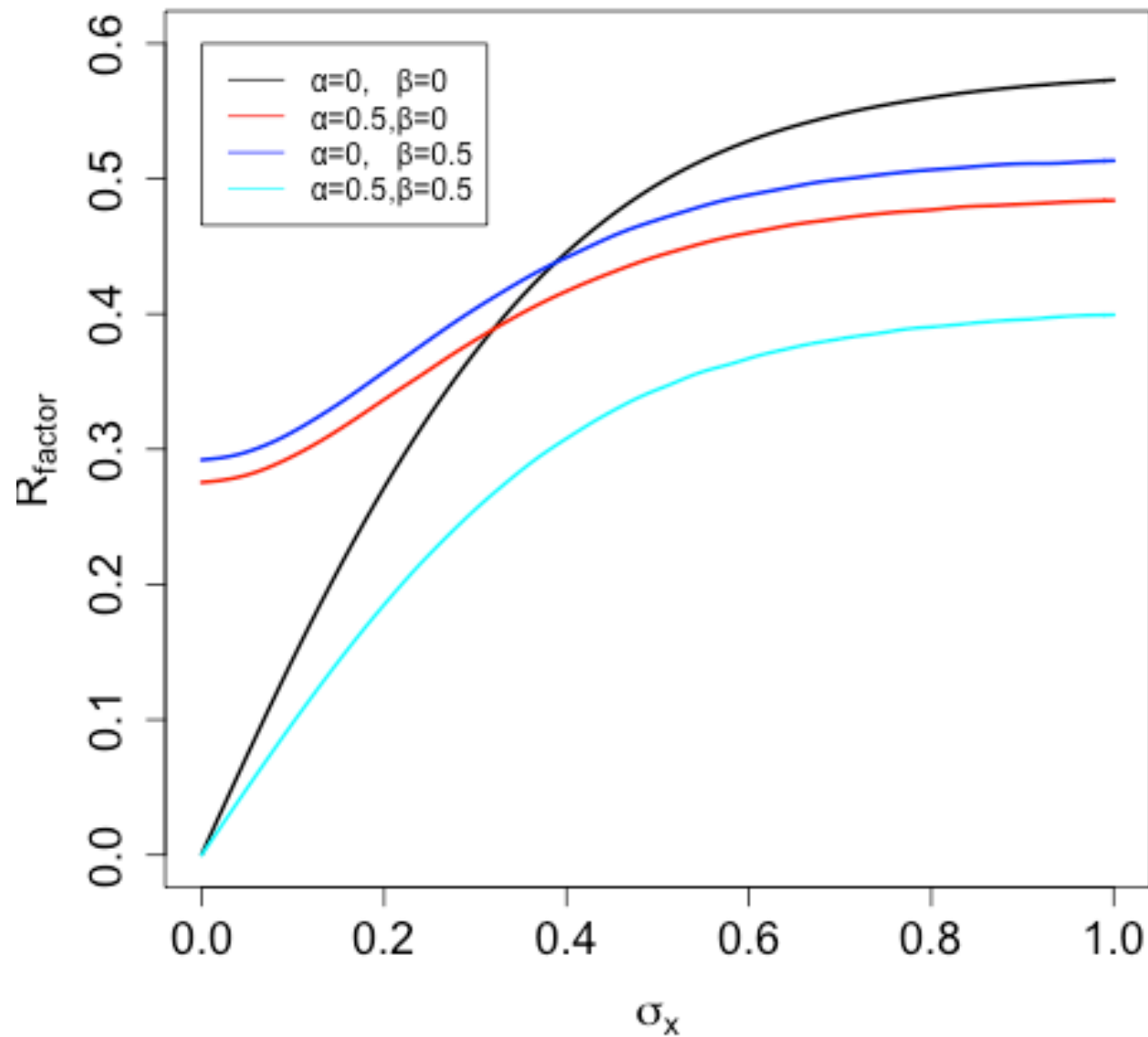
Twin: Few warnings about R values

Rvalues for random structures (no other peculiarities)

Twin	Modeled	Not modeled
Yes	0.41	0.49
No	0.52	0.58

Murshudov GN “Some properties of Crystallographic Reliability index – Rfactor:
Effect of Twinning” Applied and Computational Mathematics”, 2011:10;250-261

Rvalue for structures with different model errors:
Combination of real and modeled perfect twin fractions



Low resolutions refinement tools

Problems of low resolution refinement

- 1) Function to describe fit of the model into experiment: likelihood or similar
 - 1) Data may come from very peculiar “crystals”: Twin, OD, multiple cell
 - 2) Radiation damage
 - 3) Converting I-s to $|F|$ may not be valid: weak reflections, modulated crystals
- 2) Limited and noisy data: use of available knowledge
 - 1) High B value and spread of B values (!!!)
 - 2) Severe incompleteness of models
- 3) Smeared electron density with vanishing side chains, secondary structures, domains:
High B values and series termination

Use of available knowledge

- 1) NCS local
- 2) Restraints to known structure(s)
- 3) Restraints to current inter-atomic distances (implicit normal modes or “jelly” body)
- 4) Better restraints on B values

These are available from the version 5.6

Note

Buster/TNT has local NCS and restraints to known structures

CNS has restraints to known structures (they call it deformable elastic network)

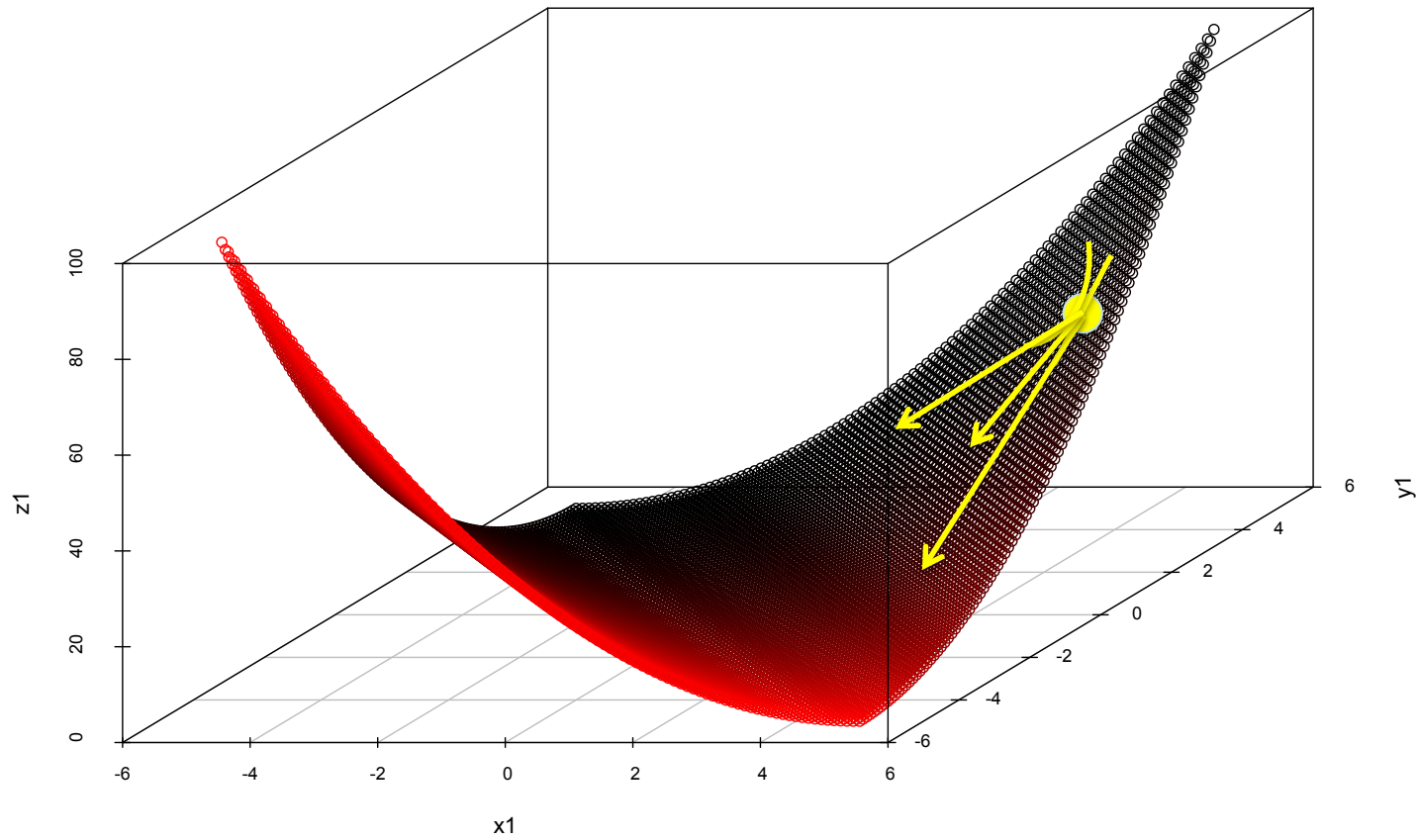
Phenix has B-value restraints on non-bonded atom pairs and automatic global NCS

Local NCS (only for torsion angle related atom pairs) was available in SHELXL since the beginning of time

Regularisation

Example:

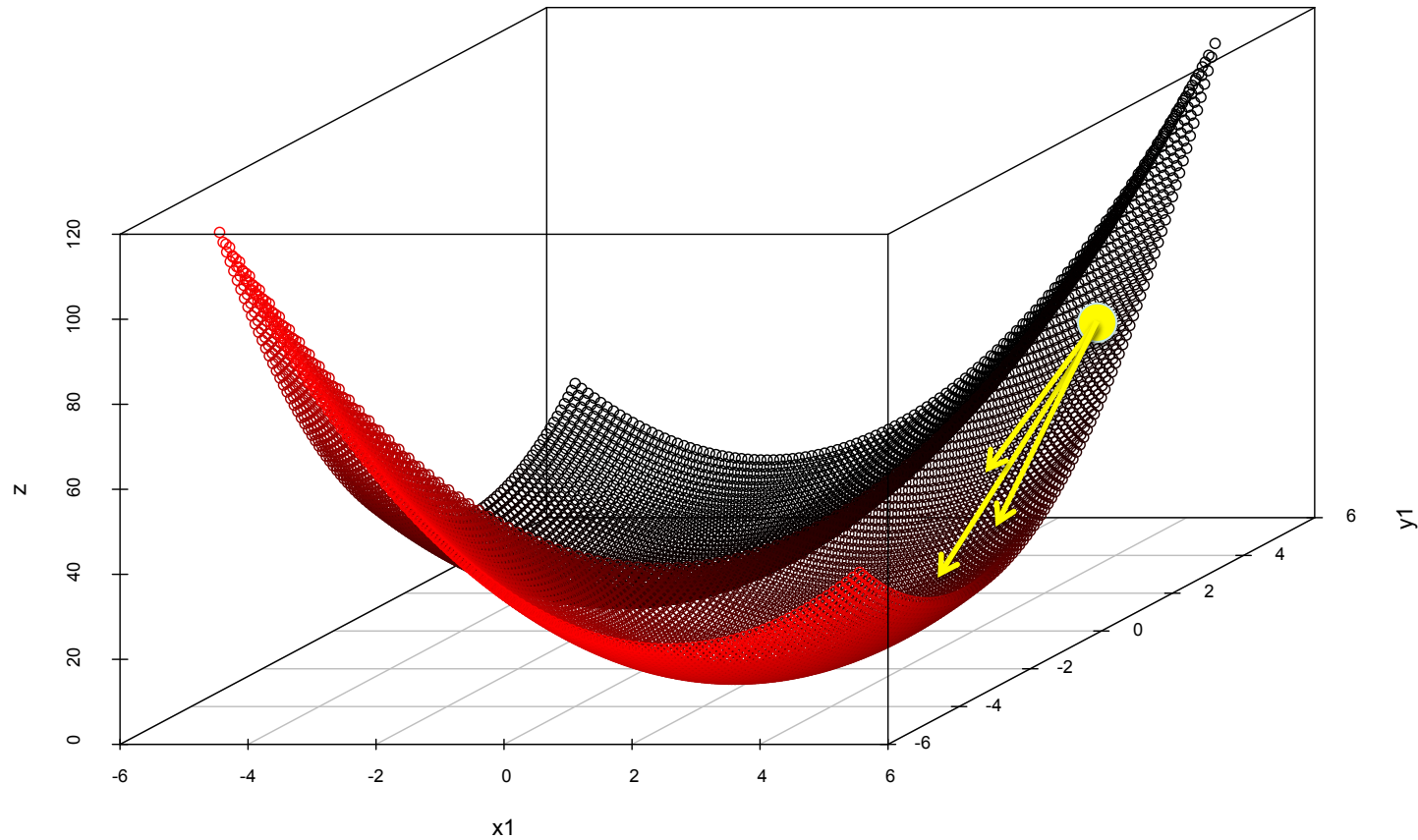
$$z = (x + y)^2$$



Regularisation

Example:

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



Regularise using prior information:

$$|x - y| = 4$$

Regularisation

Use of available knowledge (prior information):

Ultra high resolution:

- May not be needed (unrestrained refinement)

High-low resolution:

- Geometry restraints (chemical information)

Medium-low resolution:

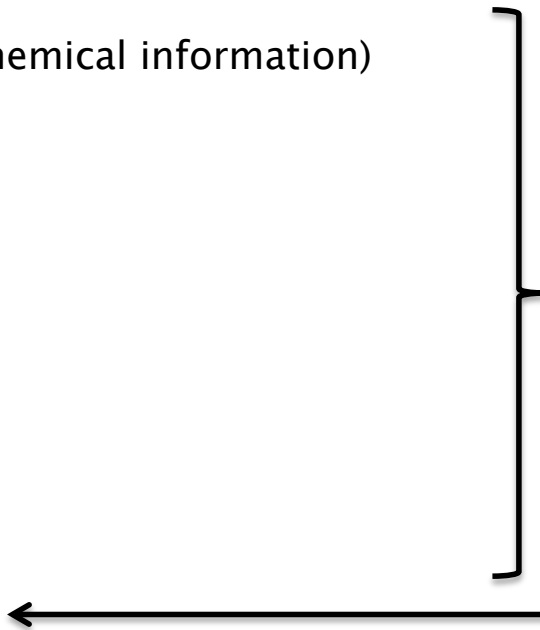
- Local NCS restraints
- B-value restraints

Low resolution:

- External restraints
- Jelly body restraints

Regularisers with a target

Regularisers without a target



Auto NCS: local and global

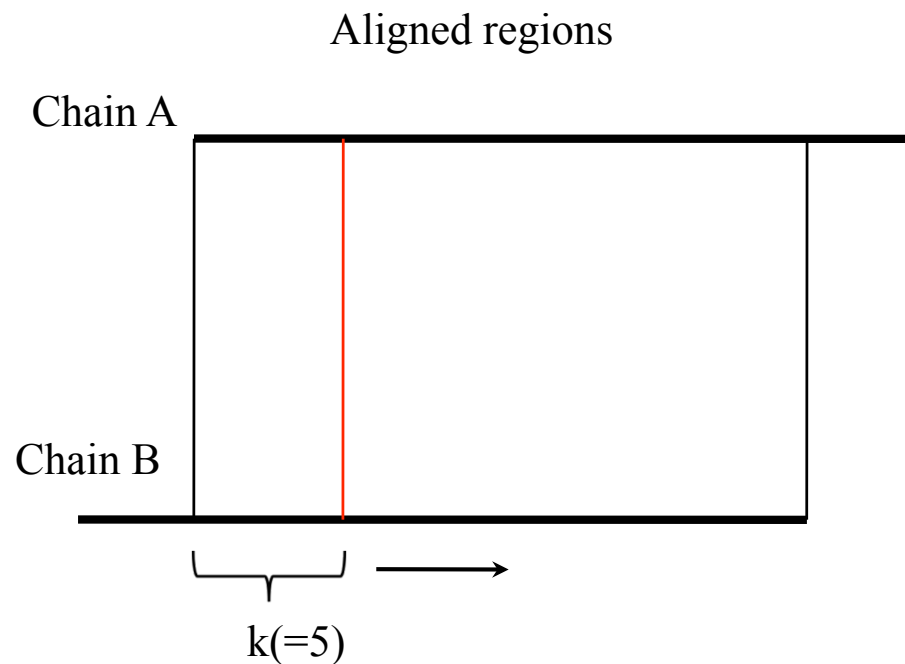
1. Align all chains with all chains using Needleman-Wunsh method
2. If alignment score is higher than predefined (e.g.80%) value then consider them as similar
- 3.Find local RMS and if average local RMS is less than predefined value then consider them aligned
4. Find correspondence between atoms
5. If global restraints (i.e. restraints based on RMS between atoms of aligned chains) then identify domains
- 6.For local NCS make the list of corresponding interatomic distances (remove bond and angle related atom pairs)
- 7.Design weights

The list of interatomic distance pairs is calculated at every cycle

Auto NCS

Global RMS is calculated using all aligned atoms.

Local RMS is calculated using k (default is 5) residue sliding windows and then averaging of the results



$$Ave(RmsLoc)_k = \frac{1}{N - k + 1} \sum_{i=1}^{N-k+1} RmsLoc_i$$
$$RMS = Ave(RmsLoc)_N$$

Auto NCS: Iterative alignment

Example of alignment: 2vtu.

There are two chains similar to each other. There appears to be gene duplication

RMS – all aligned atoms

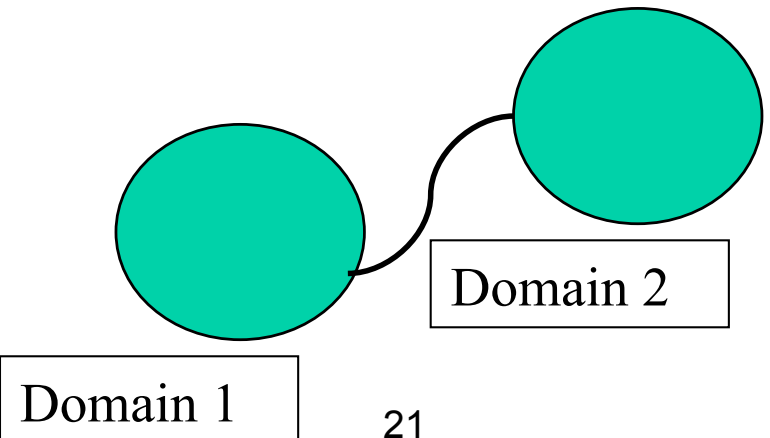
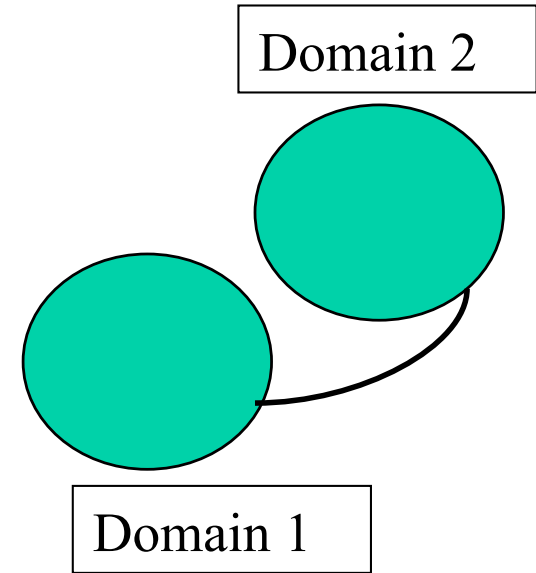
Ave(RmsLoc) – local RMS

```
***** Alignment results *****
```

: N:	Chain 1 :	Chain 2 :	No of aligned	:Score :	RMS	:Ave(RmsLoc):
: 1 :	J(131 - 256)	: J(3 - 128) :	126 :	1.0000 :	5.2409 :	1.6608 :
: 2 :	J(1 - 257)	: L(1 - 257) :	257 :	1.0000 :	4.8200 :	1.6694 :
: 3 :	J(131 - 256)	: L(3 - 128) :	126 :	1.0000 :	5.2092 :	1.6820 :
: 4 :	J(3 - 128)	: L(131 - 256) :	126 :	1.0000 :	3.0316 :	1.5414 :
: 5 :	L(131 - 256)	: L(3 - 128) :	126 :	1.0000 :	0.4515 :	0.0464 :

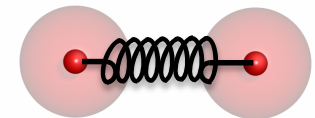
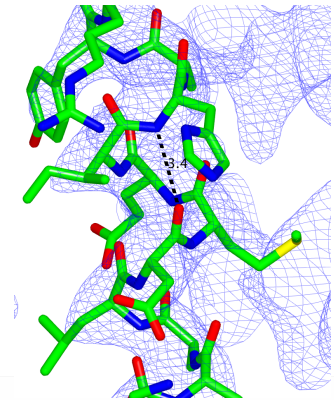
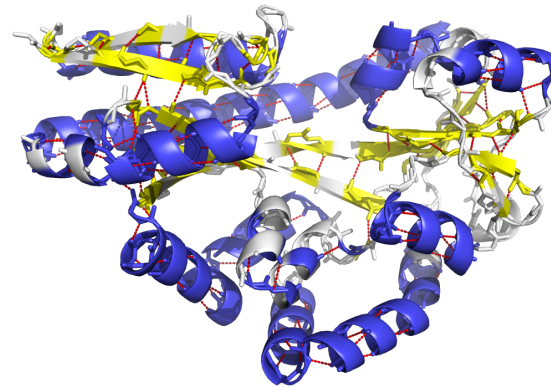
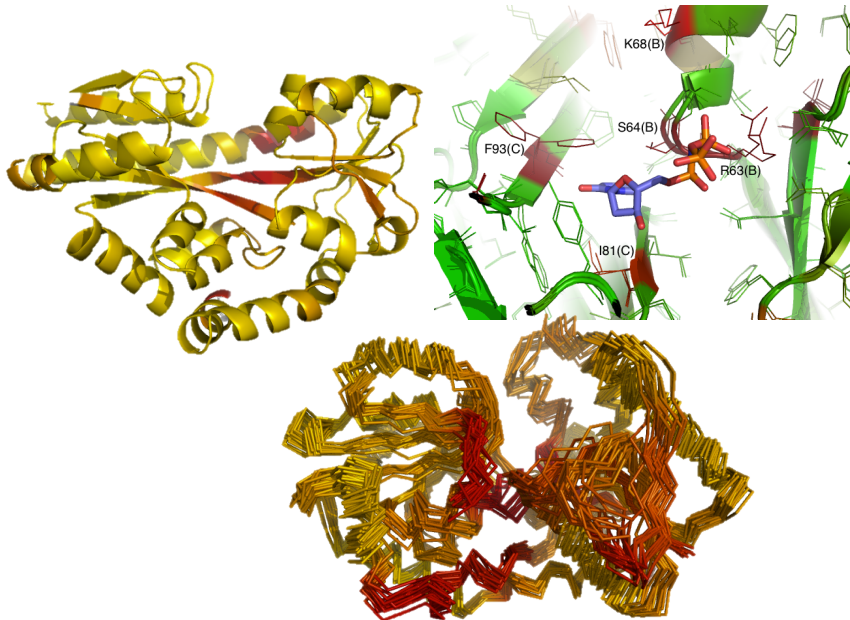
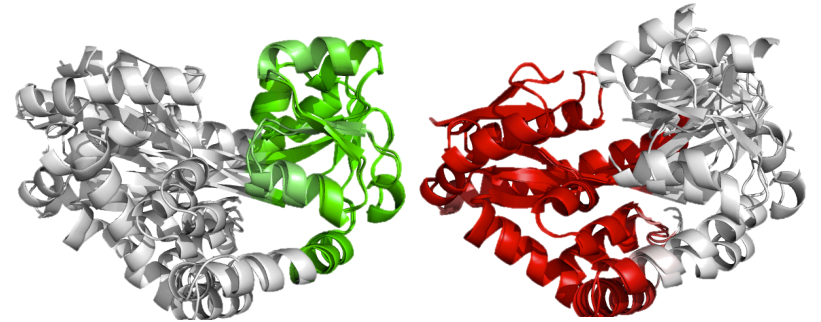
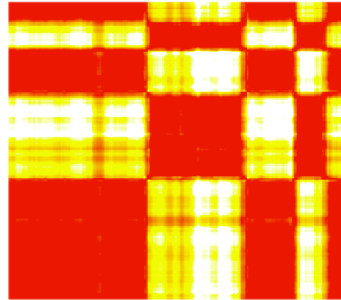
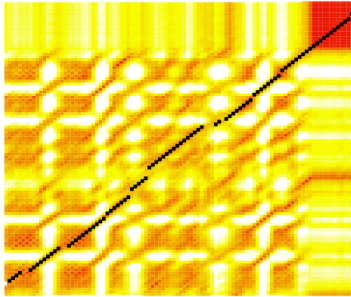
Auto NCS: Conformational changes

In many cases it could be expected that two or more copies of the same molecule will have (slightly) different conformation. For example if there is a domain movement then internal structures of domains will be same but between domains distances will be different in two copies of a molecule

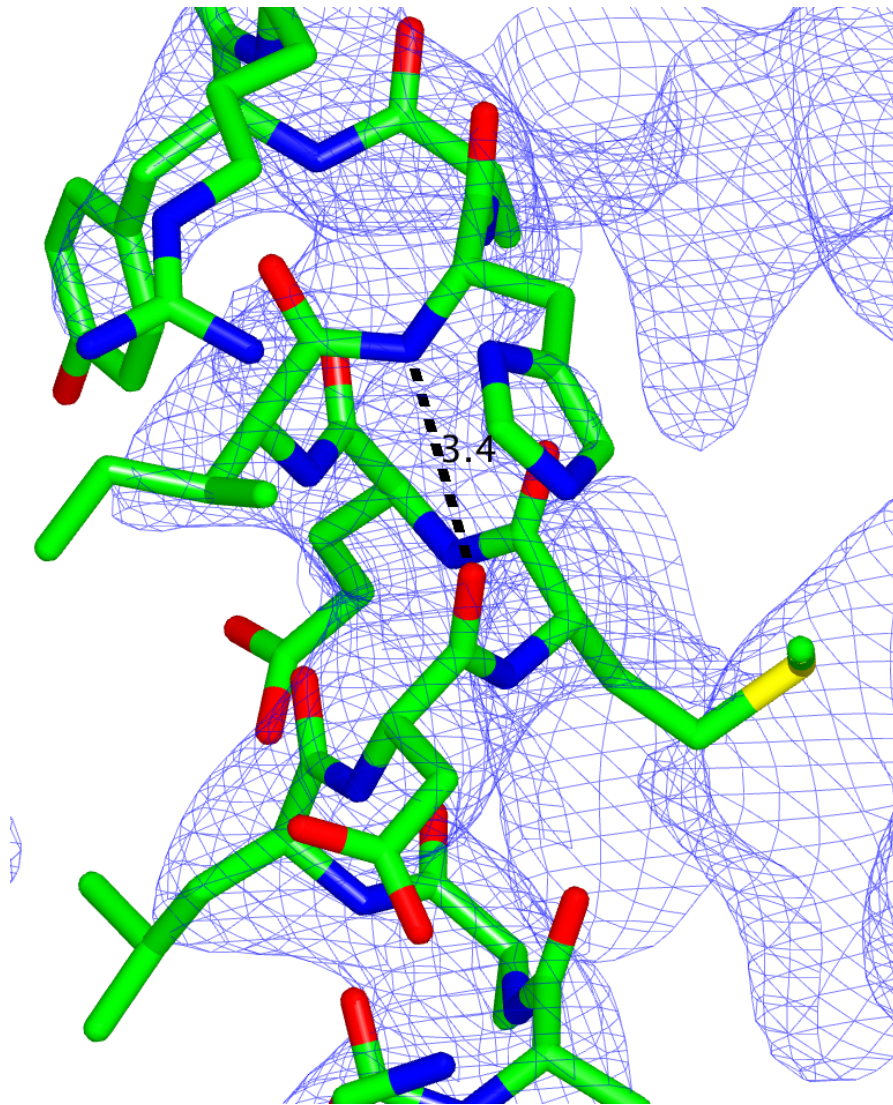


ProSMART (by R.Nicholls)

Conformation-independent comparative analysis of protein structures

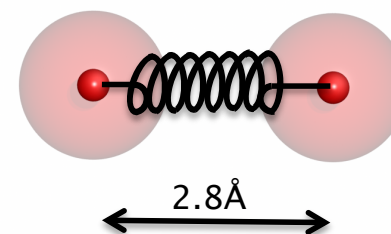


External Restraints



3g4w - 3.7Å

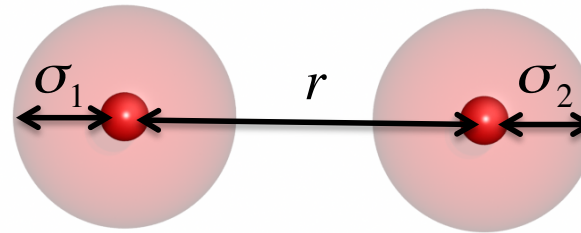
Prior information:



External Restraints

Restraint Parameter Estimation:

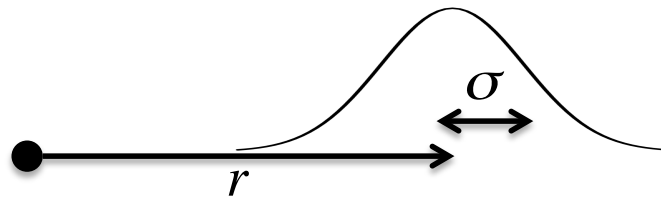
Atoms in homologous structure:



σ_i estimated as by Murshudov & Dodson (CCP4 Newsletter 1997)

Dependencies:

- B-value
- Resolution
- Model/data completeness
- Data quality

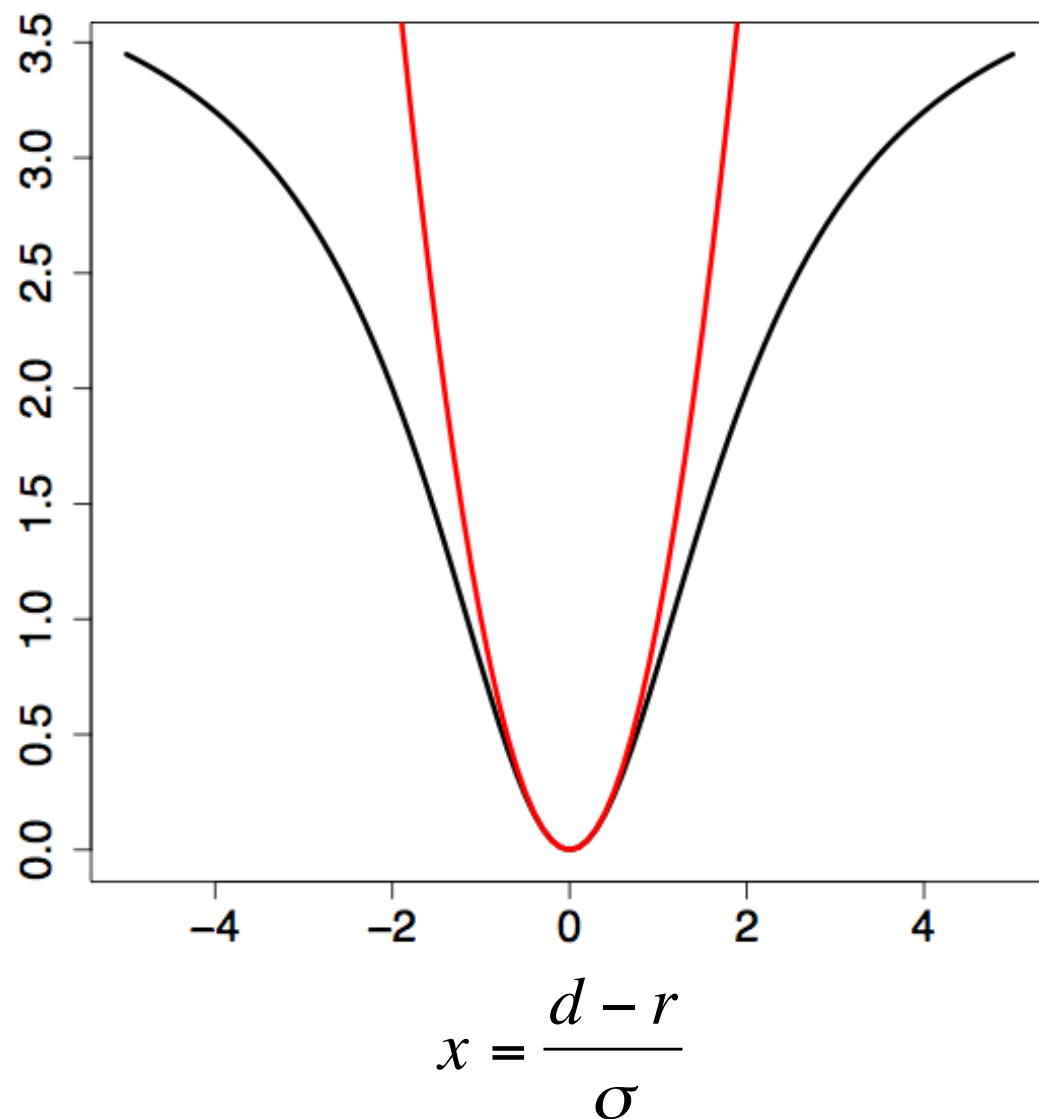


$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2c_{12}$$

0

Plus a modification that increases robustness to outliers (use of Geman-McClure function)

Use of Restraints in REFMAC5



Least squares: x^2

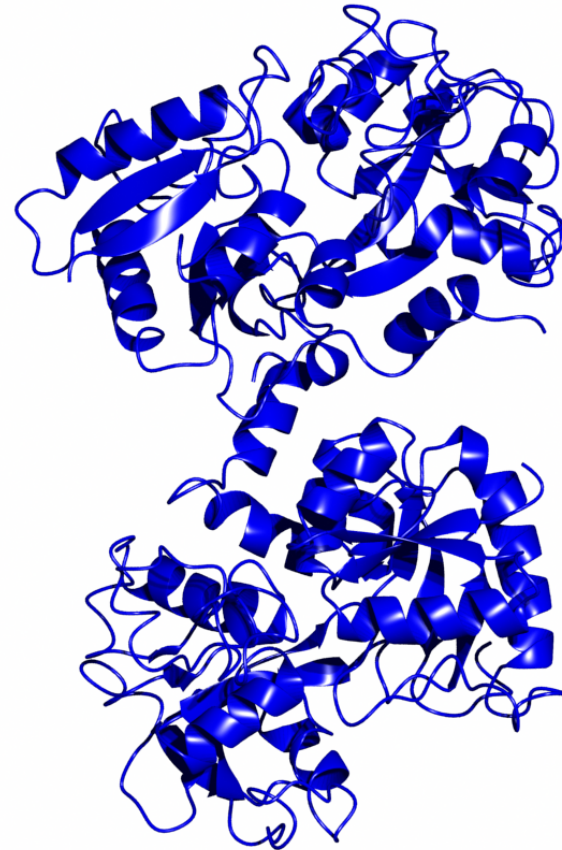
Geman-McClure: $\frac{x^2}{1 + wx^2}$

Example – Ovotransferrin 1ryx (3.5Å)

This example is given by Rob Nicholls



1ryx – 3.5Å

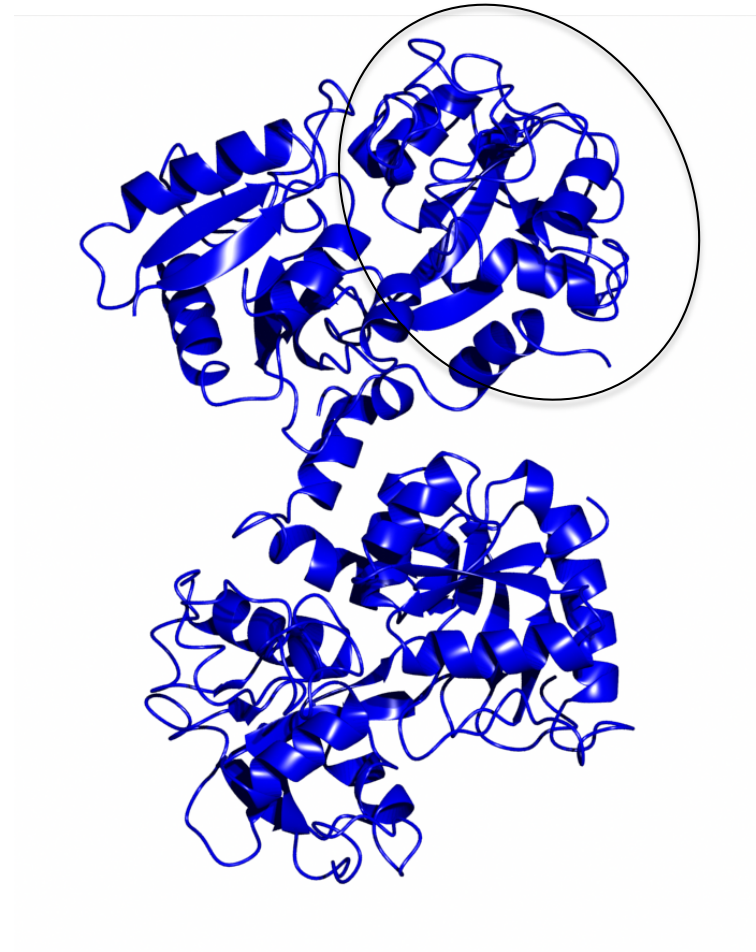


2d3i – 2.15Å

Example – Ovotransferrin 1ryx (3.5Å)



1ryx – 3.5Å



2d3i – 2.15Å

Example – Ovotransferrin 1ryx (3.5Å)

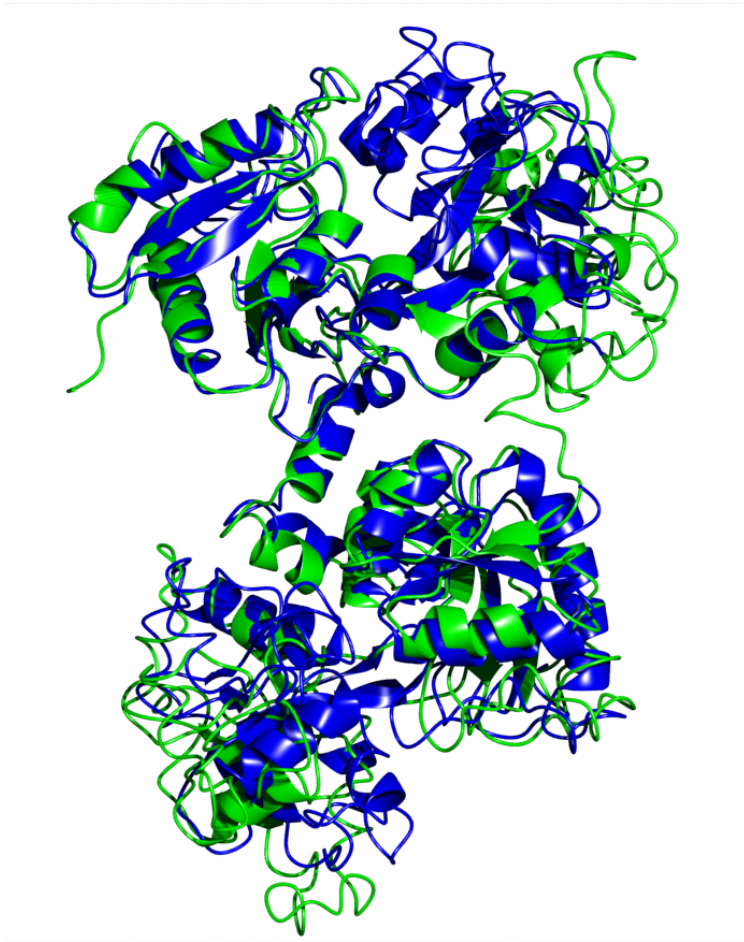


1ryx – 3.5Å

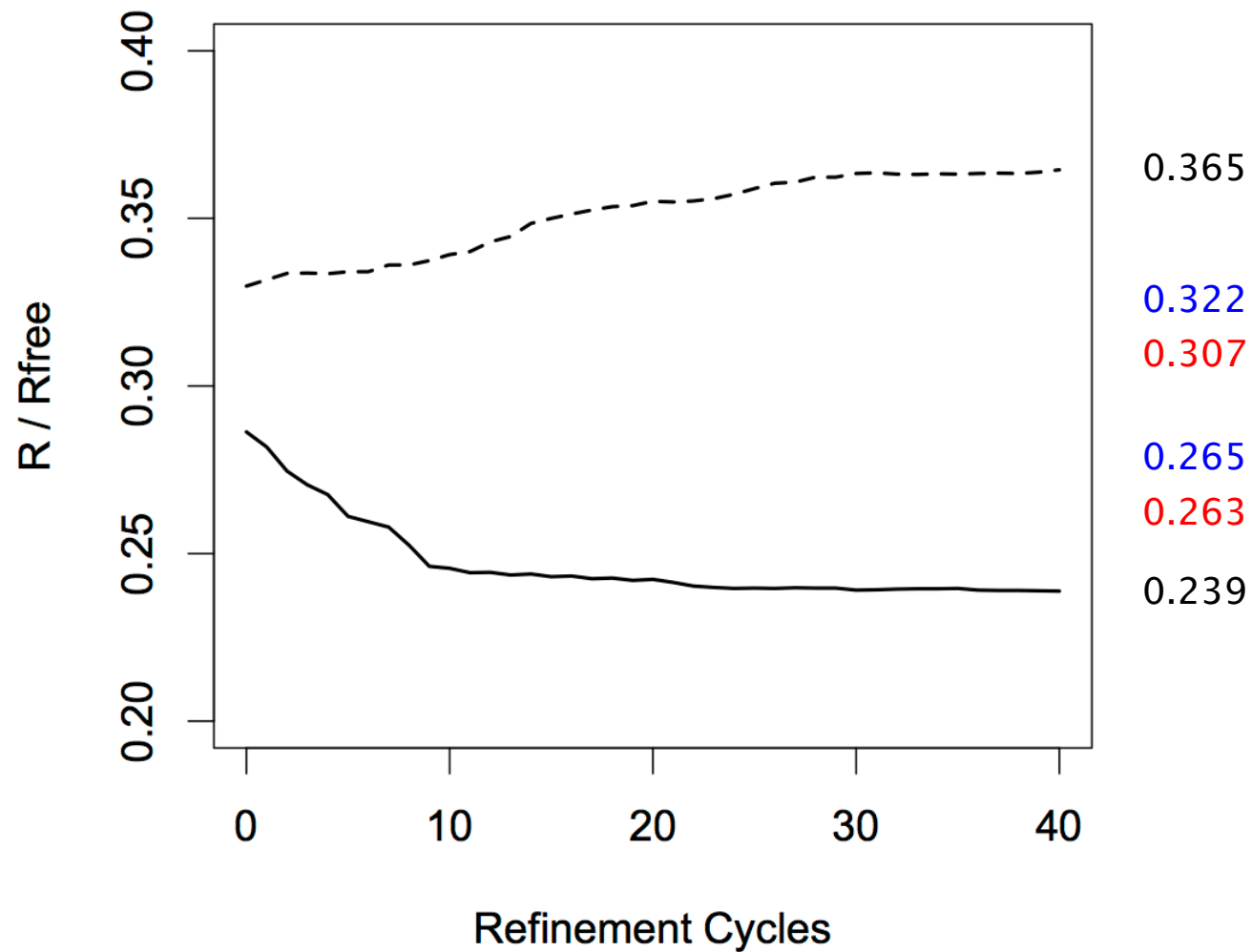


2d3i – 2.15Å

Example – Ovotransferrin 1ryx (3.5Å)



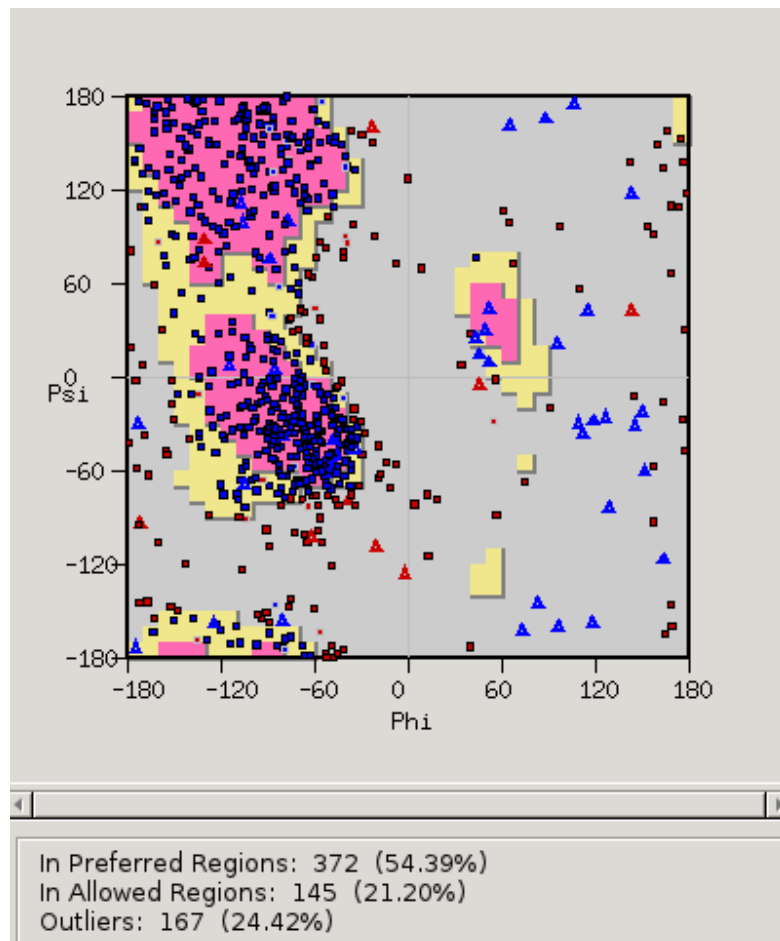
Example – Ovotransferrin 1ryx (3.5Å)



Example – Ovotransferrin 1ryx (3.5Å)

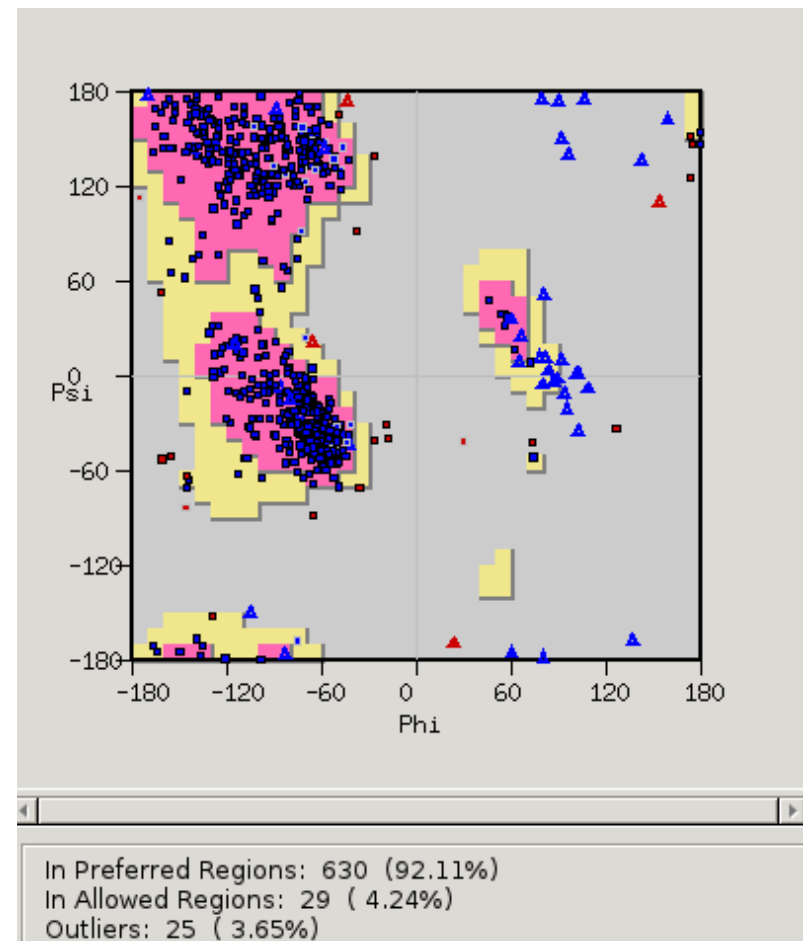
Original Structure

R/R_{free} : 0.286/0.330



Re-refined with External Restraints

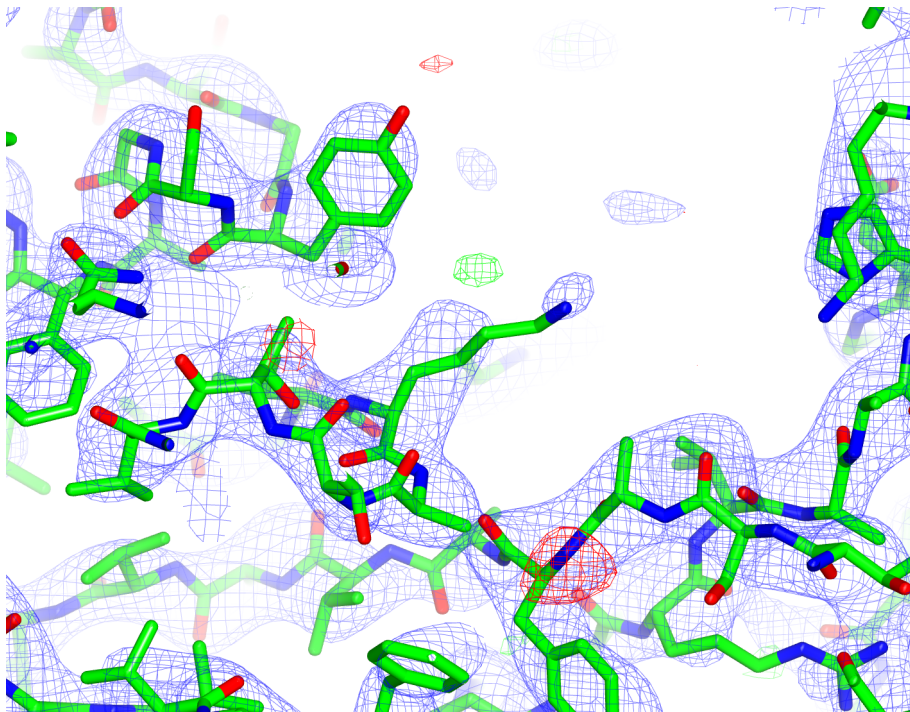
R/R_{free} : 0.263/0.307



Example – Ovotransferrin 1ryx (3.5Å)

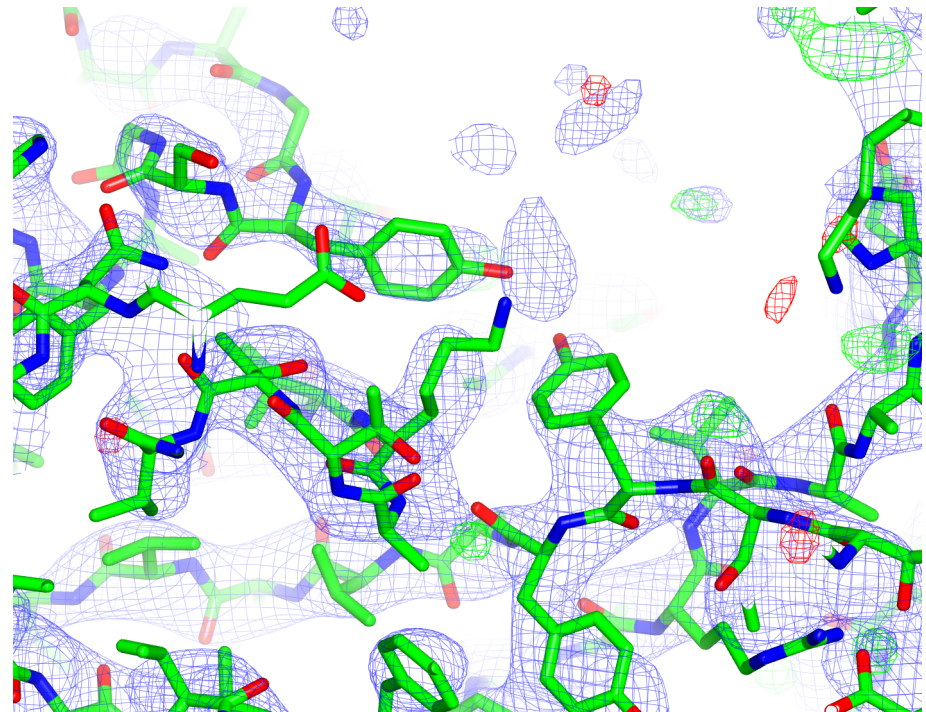
Original Structure

R/R_{free} : 0.286/0.330



Re-refined Structure

R/R_{free} : 0.252/0.307



Restraints to current distances

The term is added to the target function:

$$\sum_{pairs} w(|d| - |d_{current}|)^2$$

Summation is over all pairs in the same chain and within given distance (default 4.2Å). $d_{current}$ is recalculated at every cycle. This function does not contribute to gradients. It only contributes to the second derivative matrix.

It is equivalent to adding springs between atom pairs. During refinement inter-atomic distances are not changed very much. If all pairs would be used and weights would be very large then it would be equivalent to rigid body refinement.

It could be called “implicit normal modes”, “soft” body or “jelly” body refinement.

Example, after molecular replacement

3A resolution, data completeness 71%

Rfactors vs cycle

Black – simple refinement

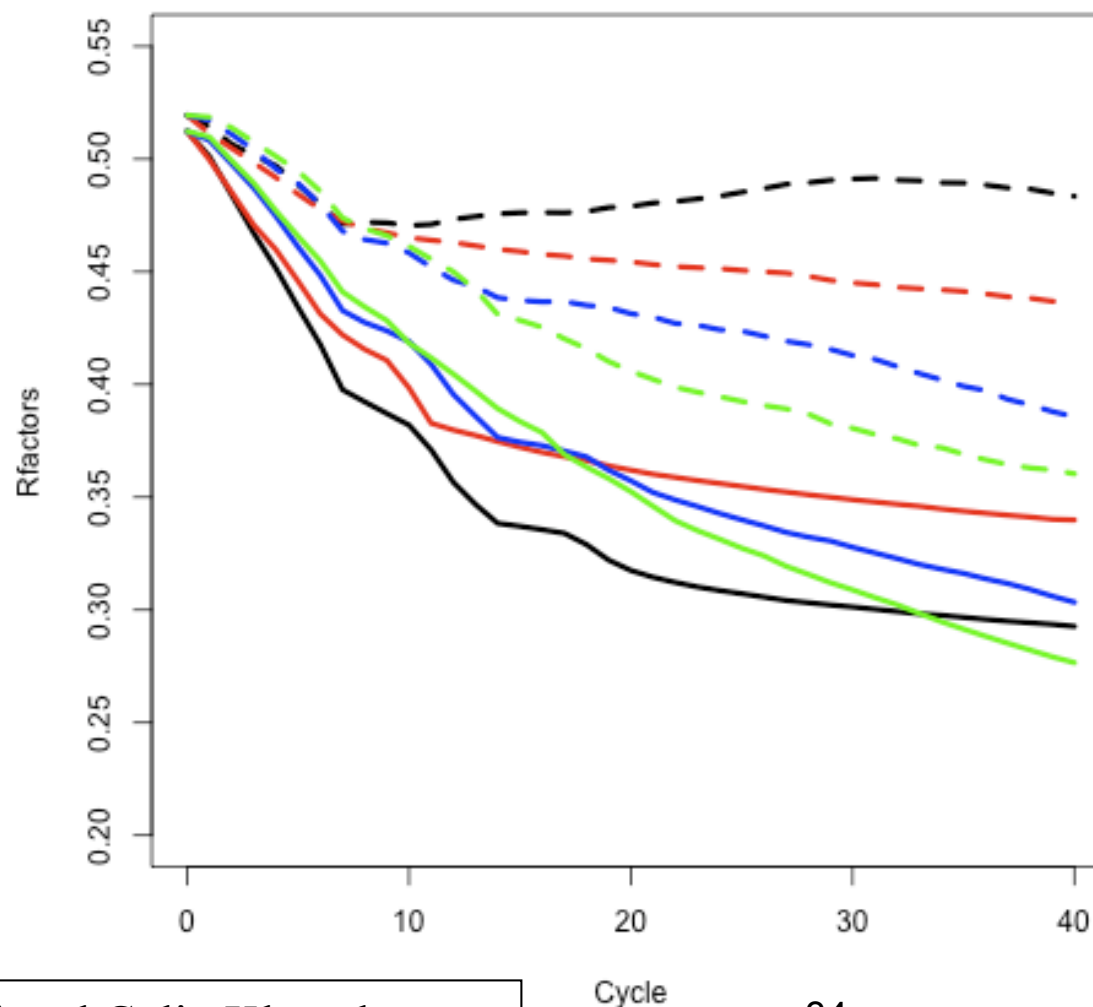
Red – Global NCS

Blue – Local NCS

Green – “Jelly” body

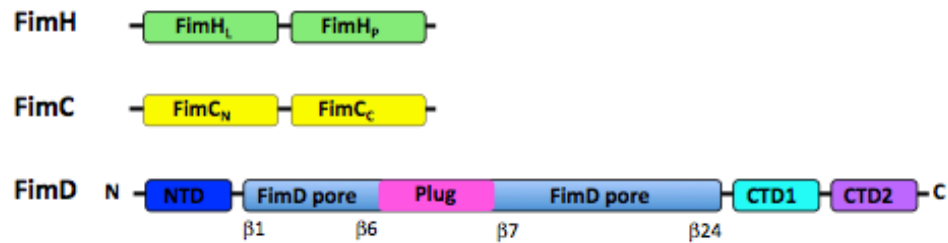
Solid lines – Rfactor

Dashed lines - Rfree

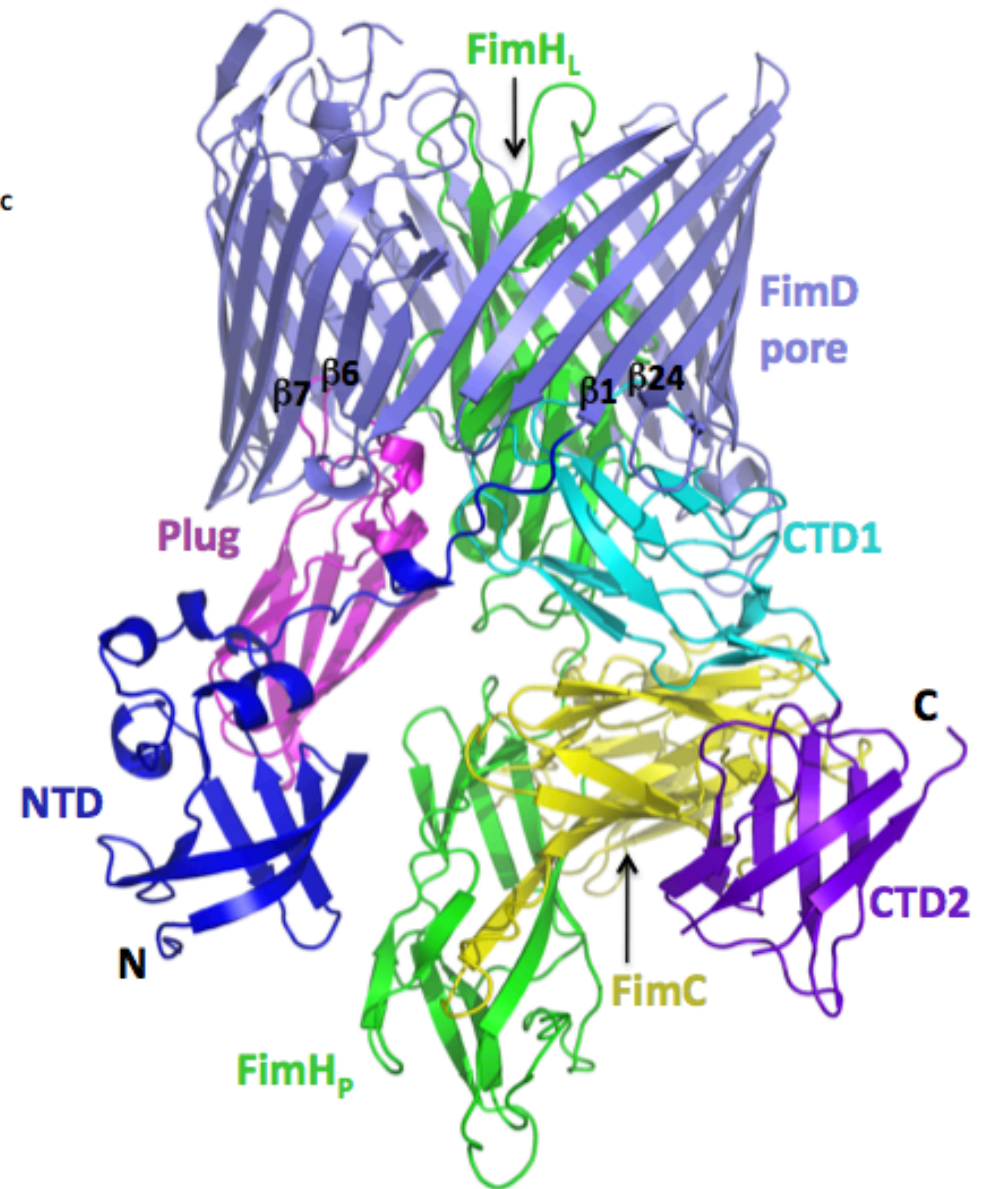


Data provided by: Marek Brzozovski and Colin Kleanthous

Effect of “jelly” body refinement: Example is provided by A. Lebedev



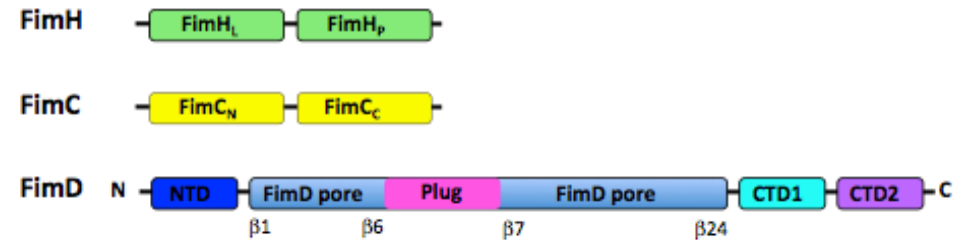
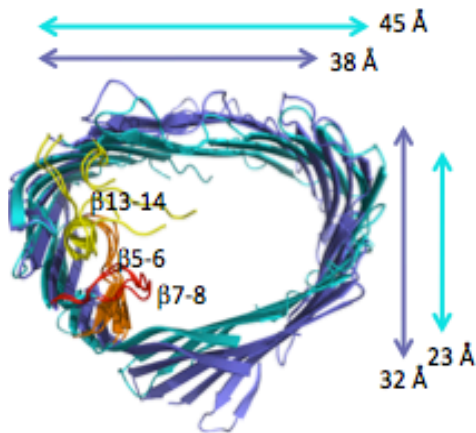
- Asymmetric unit two
- Resolution 2.8 Å



Phane et. al (2011) Nature, 474, 50-53

Usher complex structure solution

- 1. Conventional MR
 - FimC-N + FimC-C
 - FimH-L + FimH-P
 - FimD-Pore
- 2. Jelly body refinement (Refmac)
 - FimD-Pore



- 3. Fitting into the electron density
 - FimD-Plug
 - FimD-NTD
 - FimD-CTD-2
- 4. Manual building
 - FimD-CTD-1

Map Sharpening

Anisotropic Regularised Map Sharpening

Map sharpening attempts to remove overall B-value.

Can be considered as an inverse de-blurring problem:

$$\int k(x, y) \rho_0(y) dy = \rho(x) \quad \begin{array}{l} k(x, y) \text{ is unknown} \\ \text{Problem is ill-posed} \end{array}$$

Assume blurring is homogeneous: $k(x, y) = k(x - y, 0)$

Use regularisation: $\| \int k(x, y) \rho_0(y) dy - \rho(x) \|^2 + f(\rho_0(x)) \rightarrow \min$

Solution:

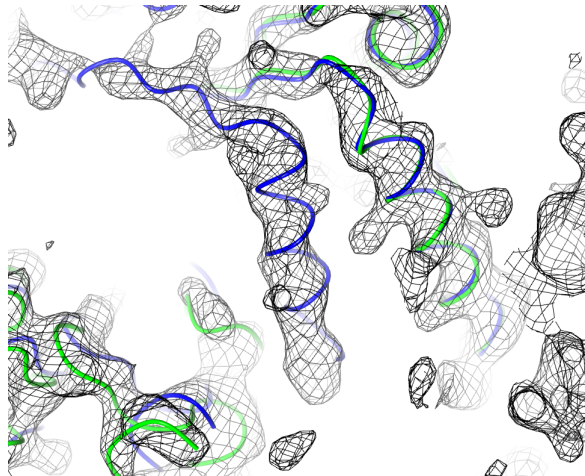
$$F_{0hkl} = \frac{K(s)}{K(s)^2 + \alpha(s)} F_{hkl} \quad K(s) = \exp\left(\frac{-s^T U_{deblur} s}{4}\right)$$

Anisotropic Regularised Map Sharpening

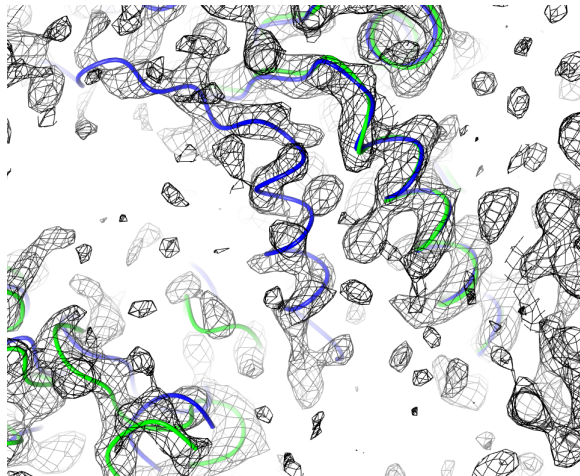
Example: 2r6c – 4Å
values

Sharpening parameter: median of atomic B

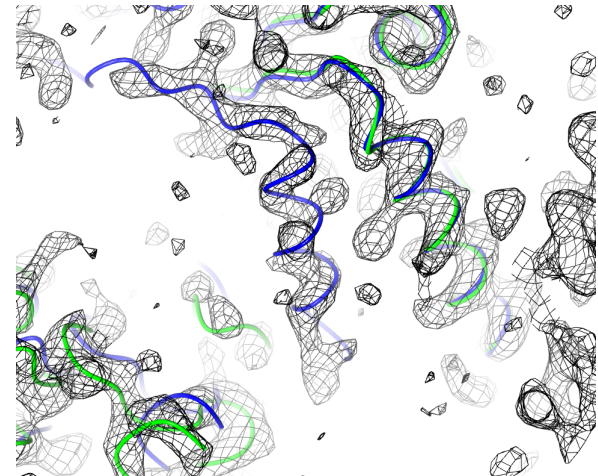
Original Map



Sharpened Map
No Regularisation



Sharpened Map With
Regularisation



Green: 2r6c – 4Å
Blue: 2r6a – 2.9Å

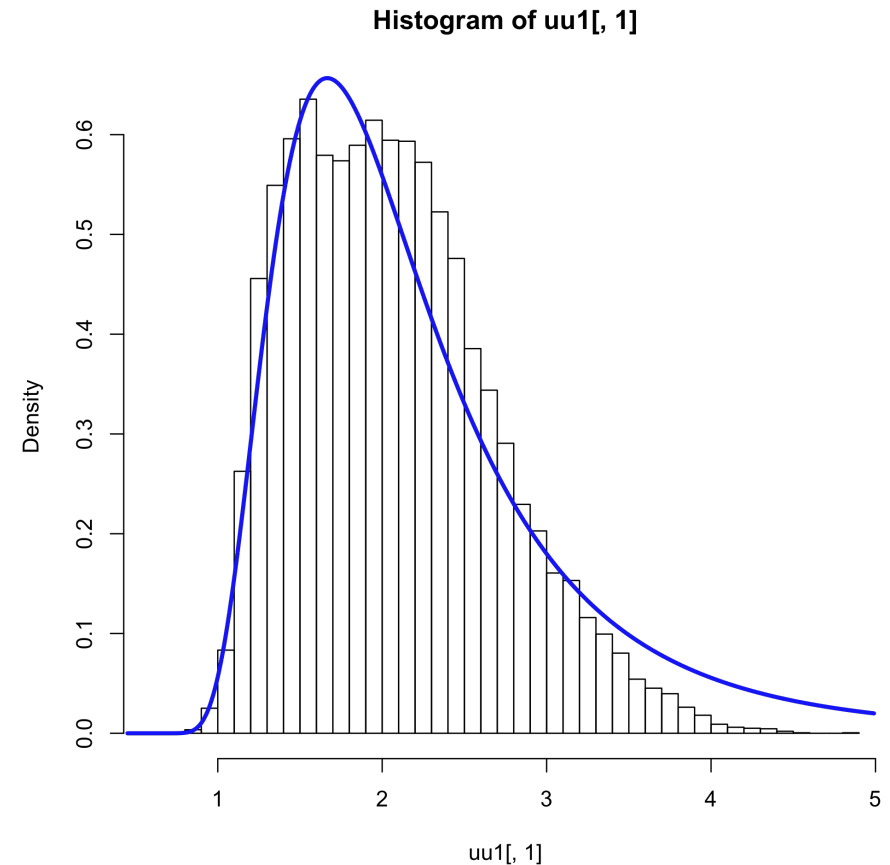
Effect of B value distribution

4Å resolution data.

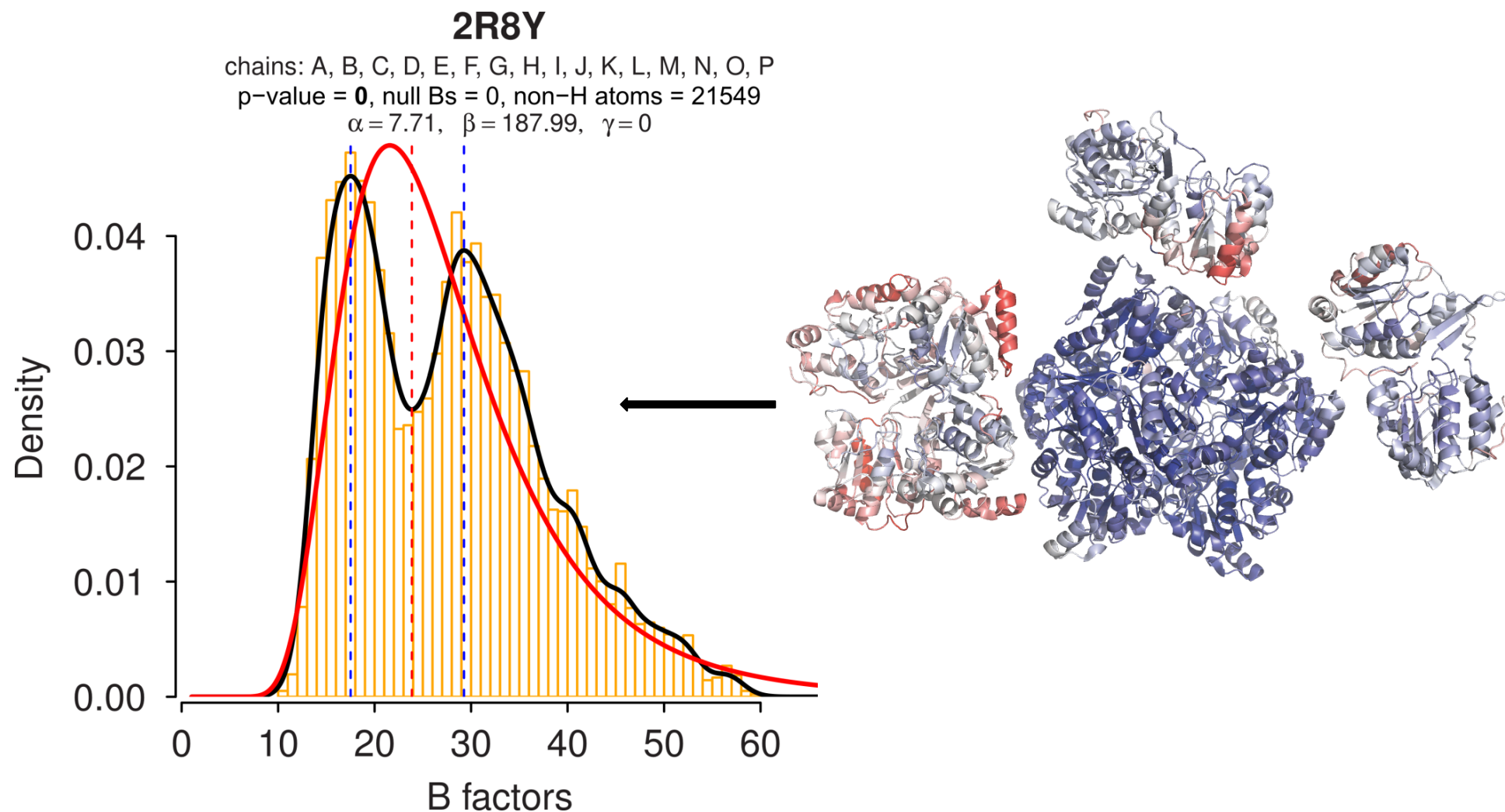
Histogram: empirical distribution
of B values

Blue line: Shifted inverse gamma
distribution

$$P(B) = \frac{\beta^\alpha (B - B_0)^{-\alpha-1}}{\Gamma(\alpha)} e^{-\beta / (B - B_0)}$$



Multimodality at chain level (1)



Conclusion

- Twin refinement improves statistics and occasionally electron density: Rfactors may be misleading
- Use of known structures improves reliability of the derived model: Especially at low resolution
- NCS restraints must be done automatically: but conformational flexibility must be accounted for
- “Jelly” body works better than I thought it should
- Regularised map sharpening looks promising.

Future work

- Reticular twinning, multiple cells, modulations
- Refinement in the presence of radiation damage
- Local TLS
- Resolution extension and reciprocal space regularisation
- Multicrystal refinement
- More restraints for RNA/DNA and carbohydrates
- NMR and X-ray refinement
- Etc

MURSHUDOV GROUP COMPUTATIONAL CRYSTALLOGRAPHY

STRUCTURAL STUDIES DIVISION, MEDICAL RESEARCH COUNCIL LABORATORY OF MOLECULAR BIOLOGY, HILLS ROAD, CAMBRIDGE, CB2 0QH, ENGLAND.

> HOME

HOME

PROJECTS

PEOPLE

SOFTWARE

LINKS

Proteins, nucleic acids and other biological macromolecules take part virtually in all processes within living organisms. Knowledge of their 3-dimensional structures is essential for understanding how they work. X-ray Crystallography is a powerful experimental technique that gives 3D structures with high accuracy. According to the Protein Data Bank more than 85% of all structures have been analysed using this technique.

Our research is centred on the development of efficient mathematical, statistical, computational algorithms for Macromolecular X-ray Crystal (MX) structure analysis. We implement the developed algorithms in the software tools and distribute them to the structural biology community. Most of our software is distributed to the community via UK based crystallographic software initiative - CCP4.

We use Bayesian method, which combines prior structural and chemical knowledge with the MX data and allows extraction of biologically relevant information from noisy diffraction data. Bayesian statistics has two components: likelihood through which new information from the experimental data to the model transferred and prior probability distribution that ensures models' consistency with the knowledge available about them.

Two examples of software developed in our group are: a maximum likelihood refinement program - REFMAC and an automatic structure solution pipeline - BALBES.

Acknowledgment

Alexei Vagin

York (now retired)

Andrey Lebedev

York (now works for ccp4)

Rob Nicholls

Cambridge

Fei Long

Cambridge

CCP4, YSBL, LMB people

REFMAC and other programs are available from CCP4 or from York's and Cambridge's websites:

www.ysbl.york.ac.uk/refmac/latest_refmac.html

<http://www2.mrc-lmb.cam.ac.uk/groups/murshudov/>

Latest ProSmart and external restraints

– Rob Nicholls (nicholls@mrc-lmb.cam.ac.uk)

DNA/RNA/sugar restraints

- Fei Long (flong@mrc-lmb.cam.ac.uk)

Distribution of F and truncate

Distribution of amplitudes of structure factors

$$P(F;atoms) = \begin{cases} \frac{1}{\sqrt{\pi\Sigma}} \exp(-\frac{|F|^2}{2\Sigma}) & \text{acentric reflections} \\ \frac{|F|}{\Sigma} \exp(-\frac{|F|^2}{\Sigma}) & \text{acentric reflections} \end{cases}$$

Truncate procedure finds expectation value of $F=J^{1/2}$ for the distribution

$$P(J;Io) = \frac{\exp(-\frac{(J - Io)^2}{2\sigma^2})P(J;atoms)}{\text{Normalisation}}$$

Where $J=|F|^2$ – ideal intensities, Io observed intensities σ observation uncertainty

When $\sigma \rightarrow \infty$ then we get expected value of $P(F;atoms)$

3zr5: refined to 2.1Å, P.Evans processed data to 1.8Å



Observed (truncated)

Beyond 2.3Å F $\rightarrow$ <F>

Beyond 2.3Å F $\rightarrow$ F_random