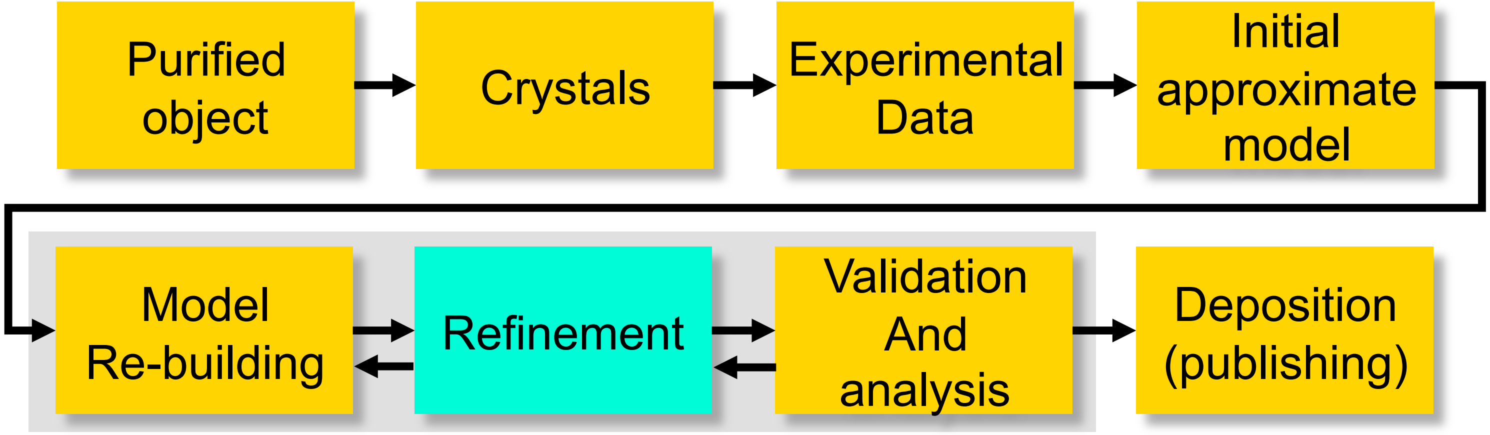
Crystallographic Structure Refinement

Pavel Afonine

Computation Crystallography Initiative
Physical Biosciences Division
Lawrence Berkeley National Laboratory, Berkeley CA, USA

Structure refinement

- **Crystallographic structure determination workflow**



Purified object

Crystals

Experimental Data

Initial approximate model

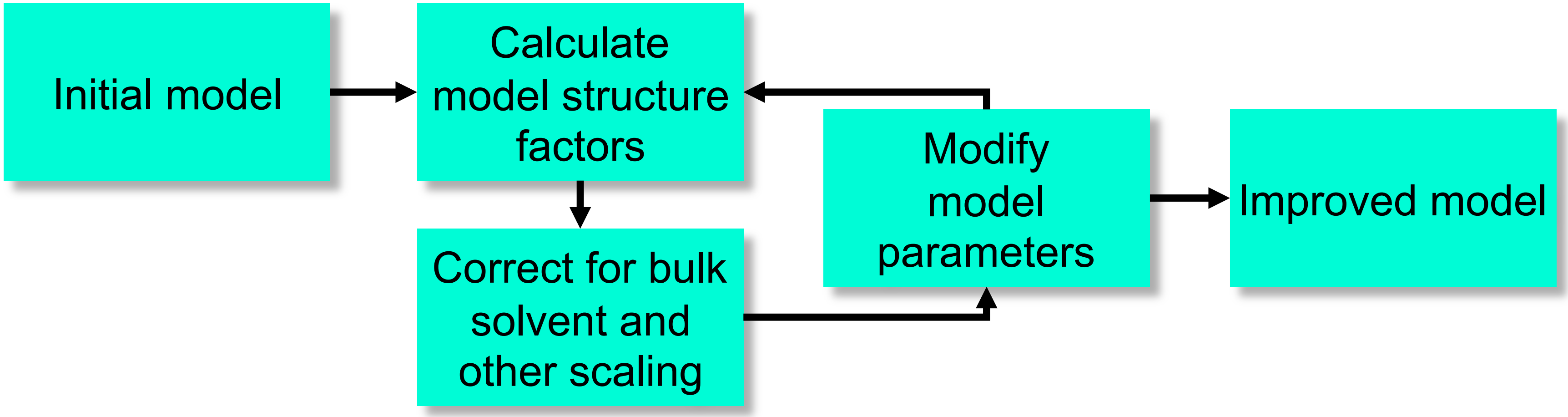
Model Re-building

Refinement

Validation And analysis

Deposition (publishing)

- **Structure refinement:** modify model parameters to describe the experimental data as good as possible



Initial model

Calculate model structure factors

Correct for bulk solvent and other scaling

Modify model parameters

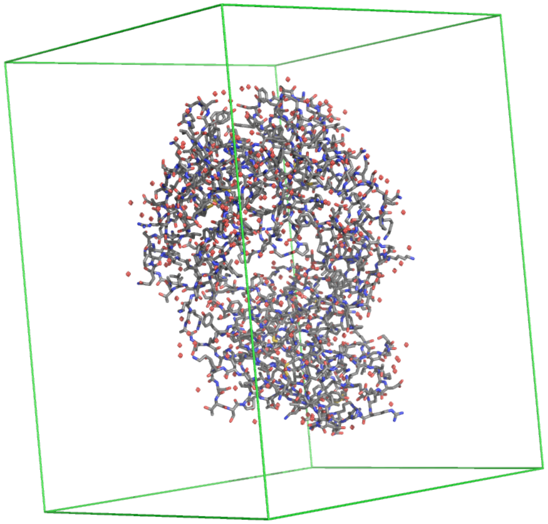
Improved model

Structure refinement

1. Model parameters
2. Optimization goal
3. Optimization method

Model parameters

Crystal (unit cell)



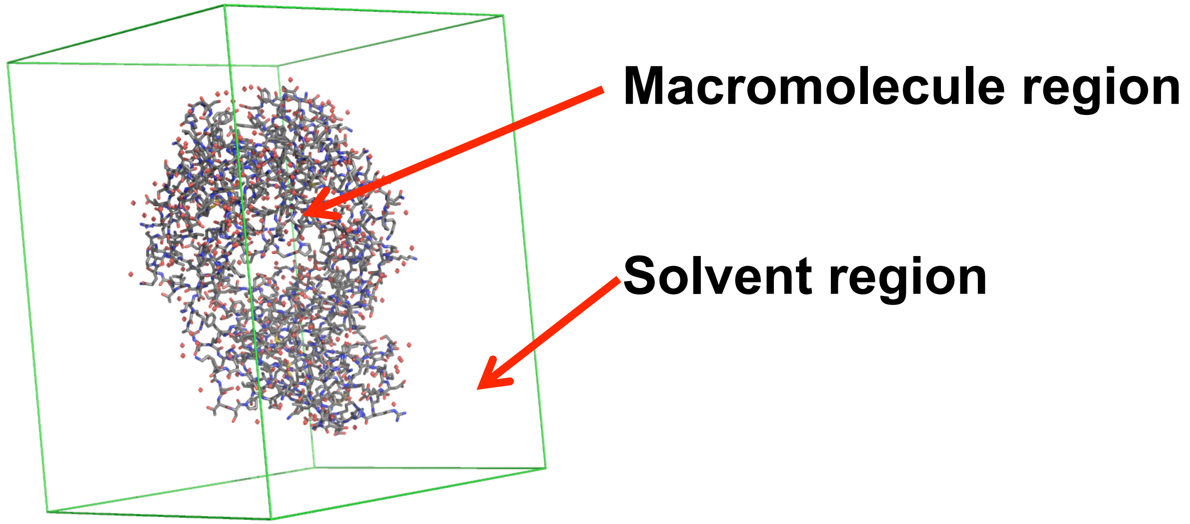
Non-atomic model parameters (Bulk solvent, anisotropy, twinning)

- Macromolecular crystals contain ~20-80% of solvent (mostly disordered)
- Crystal-specific: description of anisotropy or twinning

Atomic model parameters

- Position (coordinates)
- Mobility (ADP; Atomic Displacement Parameters)
- Disorder (occupancies)

Model parameters – Bulk solvent and anisotropy



Flat Bulk Solvent model (currently best available and most popular model):

- Electron density in solvent region is flat with average value k_{SOL} (e/Å^3)
- Solvent mask: a binary function: 0 in Macromolecular and 1 in Solvent region
- $\mathbf{F}_{MASK}$ are structure factors calculated from Bulk solvent mask
- Contribution to the model structure factor:

$$\mathbf{F}_{BULK} = k_{SOL} e^{-\frac{B_{SOL}\, s^2}{4}} \mathbf{F}_{MASK}$$

  - B_{SOL} is another bulk solvent parameter defining how deeply bulk solvent penetrates into a Macromolecular region

Non-atomic model (Bulk solvent and anisotropy)

- **Total model structure factor used in refinement, *R*-factor and map calculation:**

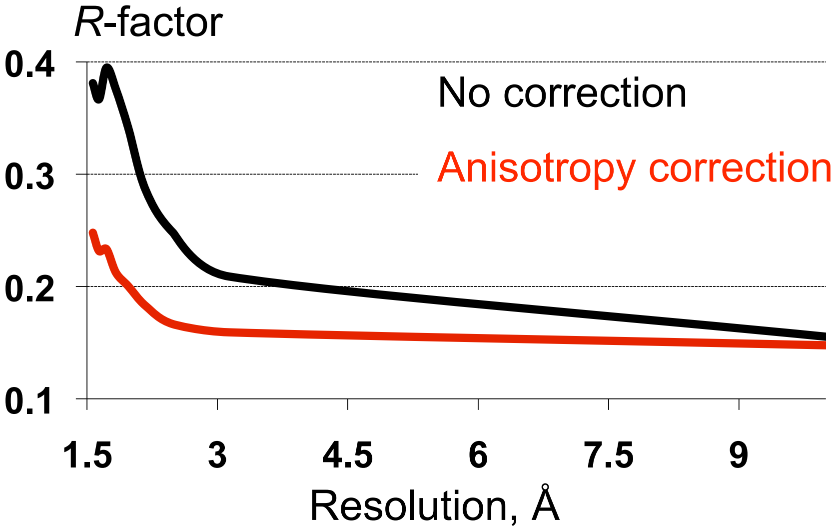
$$\mathbf{F}_{\mathrm{MODEL}} = k_{\mathrm{OVERALL}}\, e^{-\mathbf{s}\mathbf{U}_{\mathrm{CRYSTAL}}\,\mathbf{s}^{t}} \left(\mathbf{F}_{\mathrm{CALC\_ATOMS}} + k_{\mathrm{SOL}}\, e^{-\frac{B_{\mathrm{SOL}}\, s^{2}}{4}}\, \mathbf{F}_{\mathrm{MASK}} \right)$$

Anisotropy

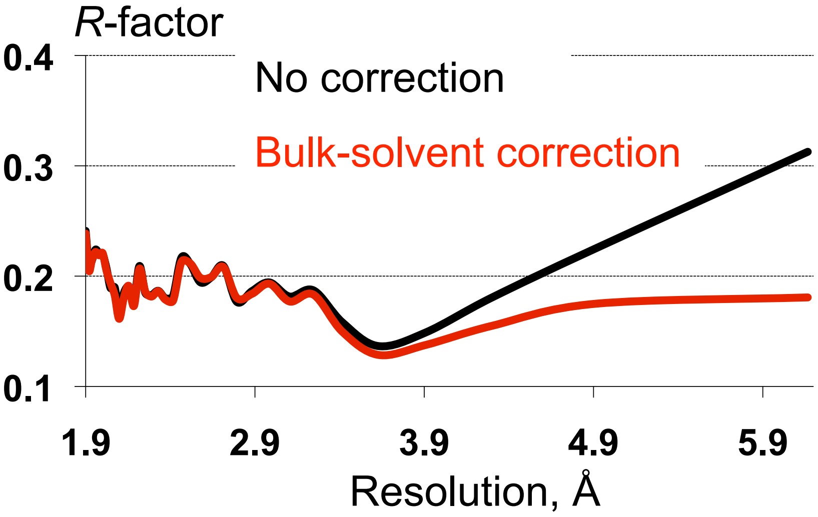
Bulk-solvent contribution

- **Contribution to *R*-factor :**

Effect of Anisotropic scaling (PDB: 2mhr)

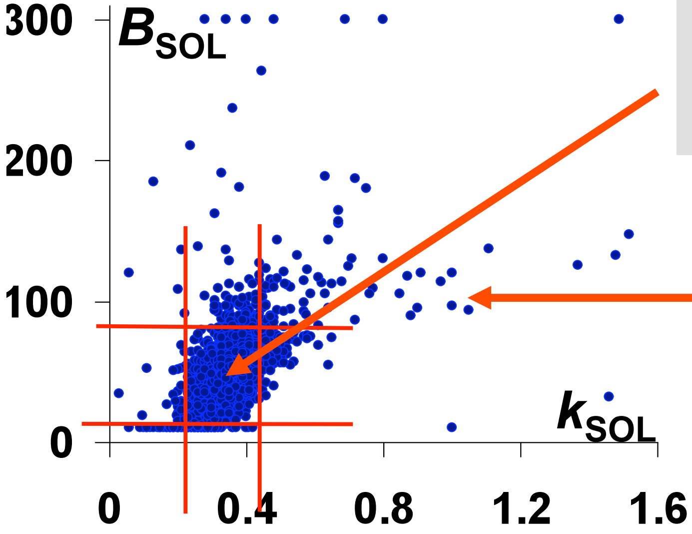


Effect of Bulk Solvent



Bulk-solvent parameters: k_{SOL} and B_{SOL}

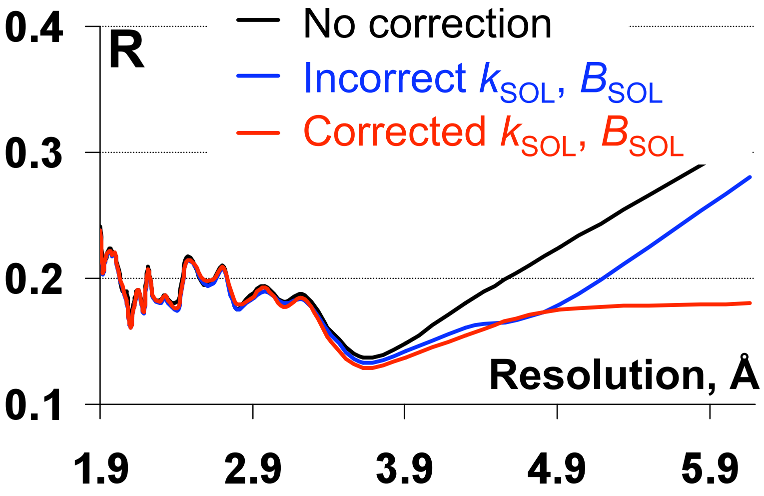
PDB survey



Average values:
k_{SOL} = 0.35 (e/Å^3)
B_{SOL} = 46.0 (Å^2)

Likely incorrect values

Bulk-solvent contributes to low resolution



Non-atomic model parameters: Twinning

- "Twins are regular aggregates consisting of crystals of the same species joined together in some definite mutual orientation" (Giacovazzo,1992).
- A twinned crystal contains two or more identical single crystals in different orientations.
- Parameterization:
  - *Twin law*: a description of the orientation of the different species relative to each other. This is an operator (matrix) that transforms the hkl indices of one species into the other.
  - *Twin fraction*: the fractional contribution of each component.
- The observed intensity is a weighted sum of the intensities of two reflections, **h** and **Th** (its twin mate): $I_{\mathrm{OBS}}(\mathbf{h}) = (1-\alpha)I(\mathbf{h}) + \alpha I(\mathbf{Th})$

$$\mathbf{F}_{\mathrm{M}}(\mathbf{h}) = e^{-\mathbf{s}\mathbf{U}_{\mathrm{CRYSTAL}}\,\mathbf{s}^{t}}\left(\mathbf{F}_{\mathrm{CALC\_ATOMS}} + k_{\mathrm{SOL}}\,e^{-\frac{B_{\mathrm{SOL}}\,s^{2}}{4}}\mathbf{F}_{\mathrm{MASK}}\right)$$

$$F_{\mathrm{MODEL}} = \left|\mathbf{F}_{\mathrm{MODEL}}\right| = k_{\mathrm{OVERALL}}\sqrt{\alpha\left|\mathbf{F}_{\mathrm{M}}(\mathbf{h})\right|^{2} + (1-\alpha)^{2}\left|\mathbf{F}_{\mathrm{M}}(\mathbf{Th})\right|^{2}}$$

Atomic model parameters

Extract from PDB file: example of an atom representation

```
ATOM      25  CA  PRO A   4      31.309  29.489  26.044  1.00 57.79           C
ANISOU    25  CA  PRO A   4    8443   7405   6110   2093    -24    -80       C
```

Atomic model parameters

- Position (coordinates)
- Local mobility (ADP; Atomic Displacement Parameters or *B*-factors):

  Diffraction data represents time- and space-averaged images of the crystal structure: time-averaged because atoms are in continuous thermal motions around mean positions, and space-averaged because there are often small differences between symmetry copies of the asymmetric unit in a crystal. ADP is to model the *small* dynamic displacements as isotropic or anisotropic *harmonic* displacements.

- Larger-scale disorder (occupancies)

  Larger displacements (beyond harmonic approximation) can be modeled using occupancies (“alternative conformations/locations”).

Atomic model parameters

Atomic model parameterization is defined by:

- quality of experimental data (resolution, completeness, …)
- quality of current model (initial with large errors, almost final, …)
- data-to-parameters ratio (restraints have to be accounted for)

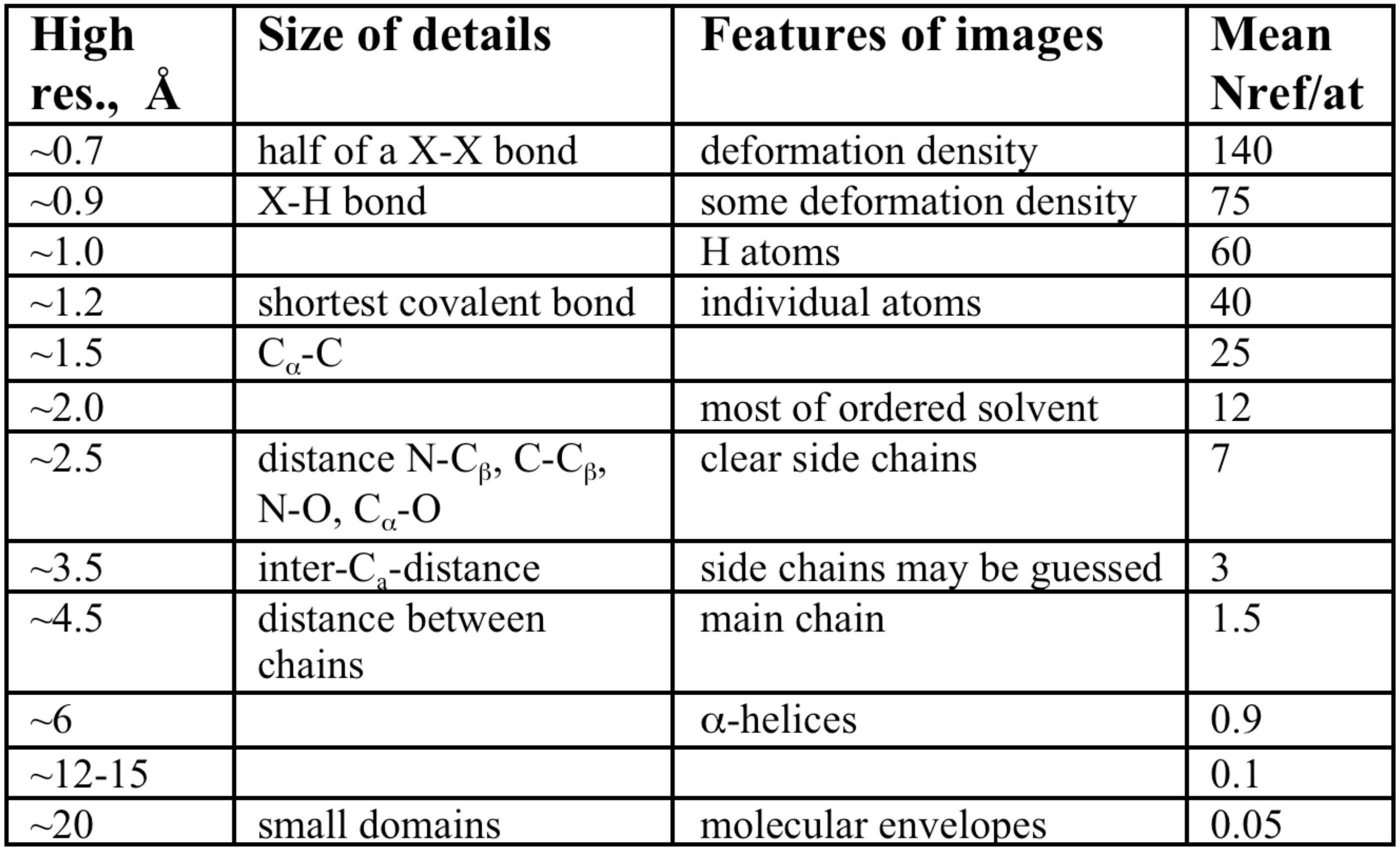
Can you see it in the map?

- ~0.01Å deviations from ideal bond lengths at resolutions ~ 2Å?
- anisotropy of individual atoms at resolutions 2Å and lower?

Model parameters

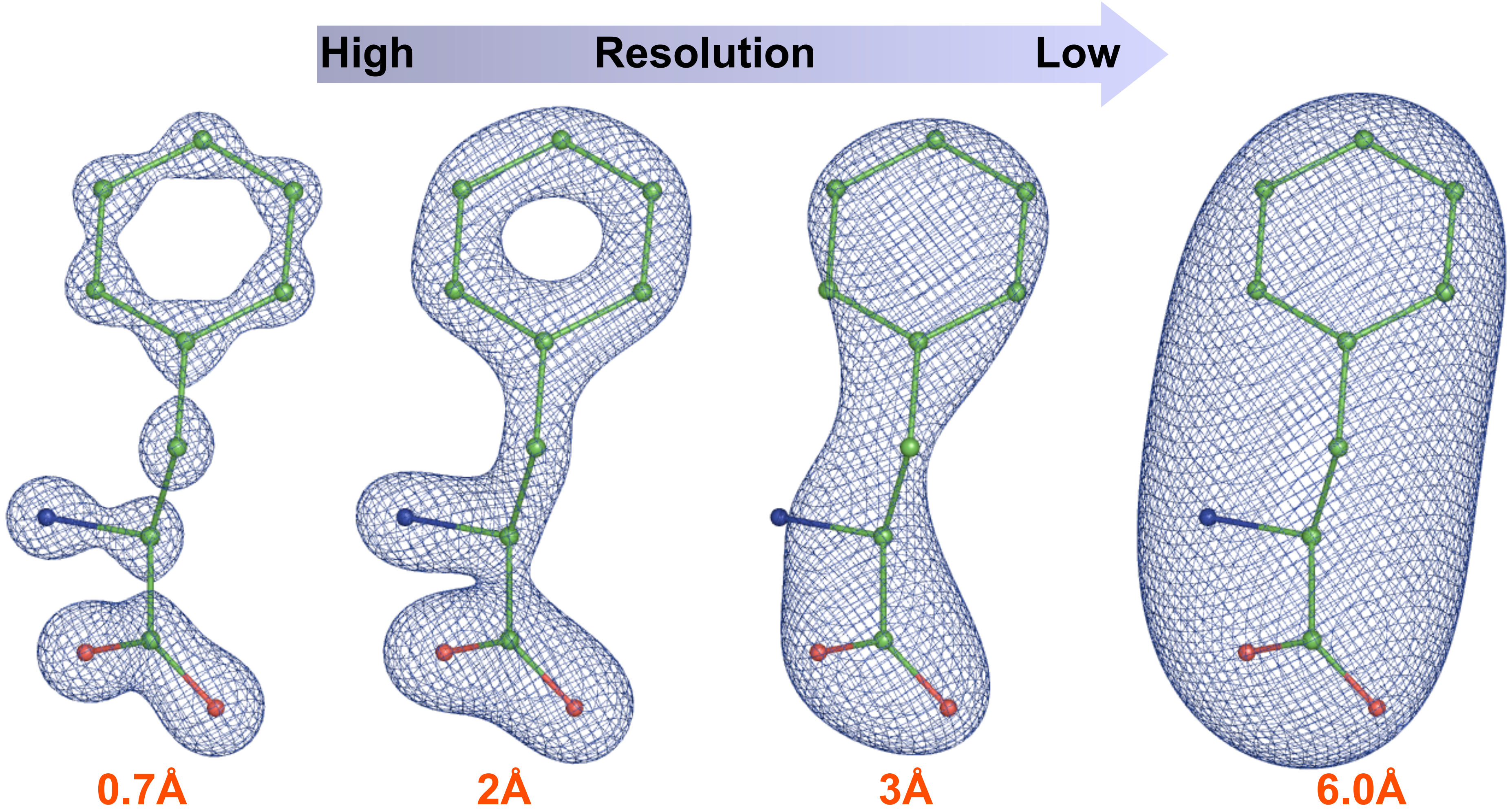
- Choice for model parameterization depends on amount of available data and its resolution

Key resolution limits and corresponding features



High res., Å	Size of details	Features of images	Mean Nref/at
~0.7	half of a X-X bond	deformation density	140
~0.9	X-H bond	some deformation density	75
~1.0		H atoms	60
~1.2	shortest covalent bond	individual atoms	40
~1.5	C_α-C		25
~2.0		most of ordered solvent	12
~2.5	distance N-C_β, C-C_β, N-O, C_α-O	clear side chains	7
~3.5	inter-C_a-distance	side chains may be guessed	3
~4.5	distance between chains	main chain	1.5
~6		α-helices	0.9
~12-15			0.1
~20	small domains	molecular envelopes	0.05

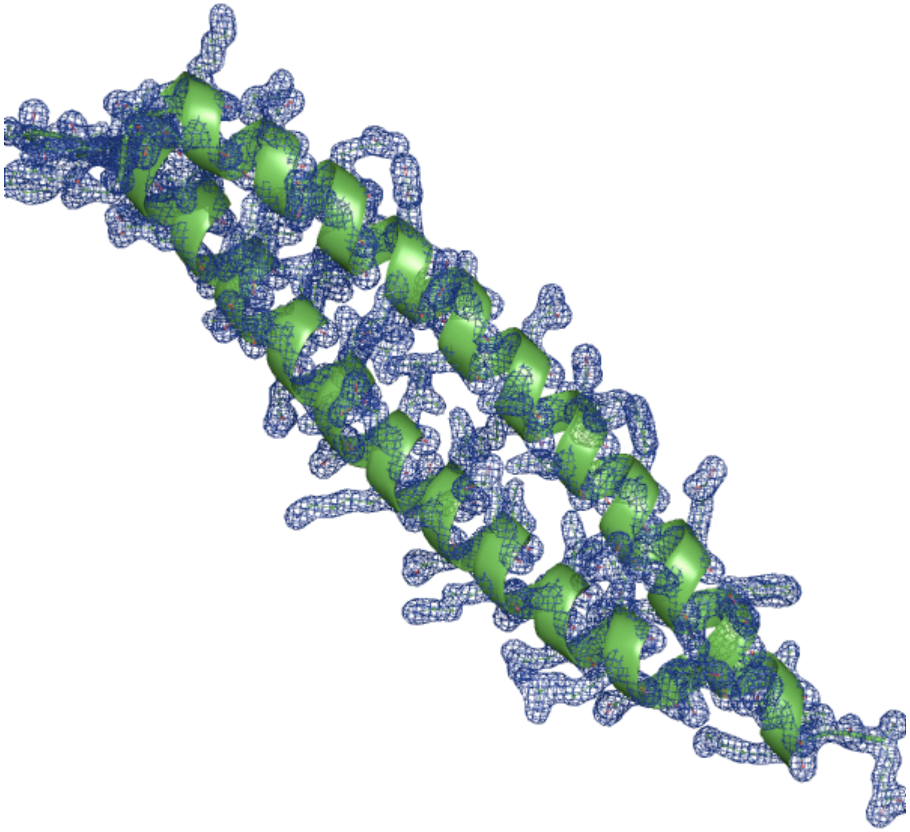
Data quality (resolution, completeness) defines how detailed the model is



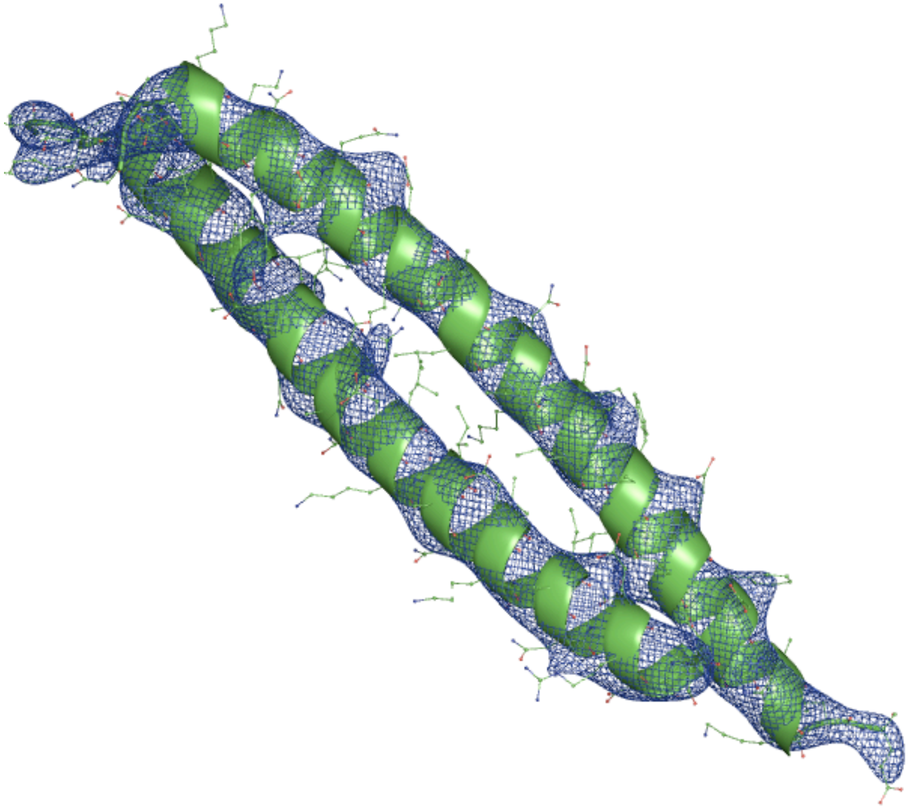
- **High resolution**: atomic models with or without restraints
- **Medium resolution**: atomic models with restraints or constraints
- **Low resolution**: atomic models with constrains or non-atomic models (cylinders for secondary-structure elements such as helices)

Data quality (resolution, completeness) defines how detailed the model is

High **Resolution** **Low**



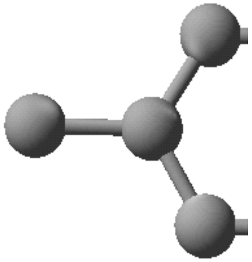
2Å



6.0Å

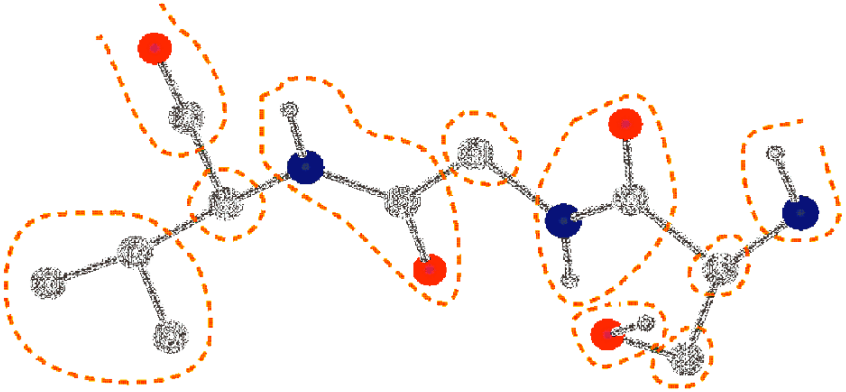
Model parameterization: coordinates

Individual atoms



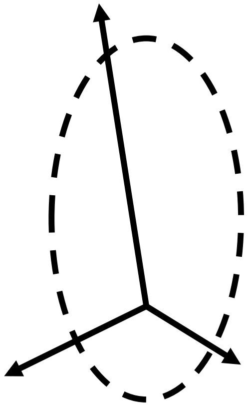
3 * Natoms

Constrained rigid bodies (torsion angle parameterization)



3 * Natoms / (7 …10)

Rigid body



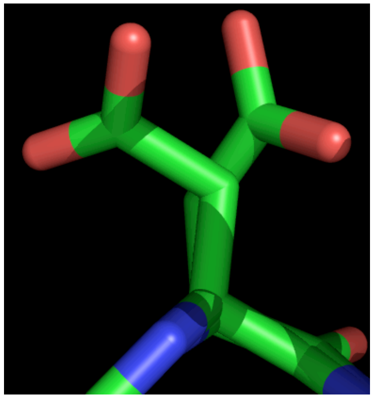
6 * Ngroups

High **Resolution** **Low**

Some *a priori* information may be needed:

- Stereochemistry restraints
- NCS restraints or constraints

Occupancy: large-scale disorder that cannot be modeled with harmonic model (ADP)



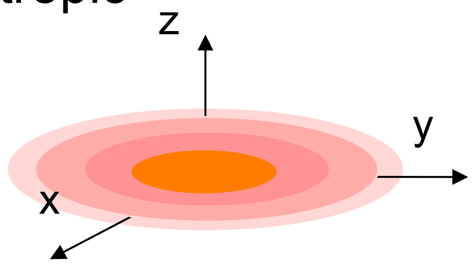
- Occupancy is the fraction of molecules in the crystal in which a given atom occupies the position specified in the model.
- If all molecules in the crystal are identical, then occupancies for all atoms are 1.00.

- We may refine occupancy because sometimes a region of the molecules may have several distinct conformations.
- Refining occupancies provides estimates of the frequency of alternative conformations.

```
ATOM      1  N  AARG A 192      -5.782  17.932  11.414  0.72  8.38           N
ATOM      2  CA AARG A 192      -6.979  17.425  10.929  0.72 10.12           C
ATOM      3  C  AARG A 192      -6.762  16.088  10.271  0.72  7.90           C
ATOM      7  N  BARG A 192     -11.719  17.007   9.061  0.28  9.89           N
ATOM      8  CA BARG A 192     -10.495  17.679   9.569  0.28 11.66           C
ATOM      9  C  BARG A 192      -9.259  17.590   8.718  0.28 12.76           C
```

Atomic Displacement Parameters (ADP or “B-factors”)

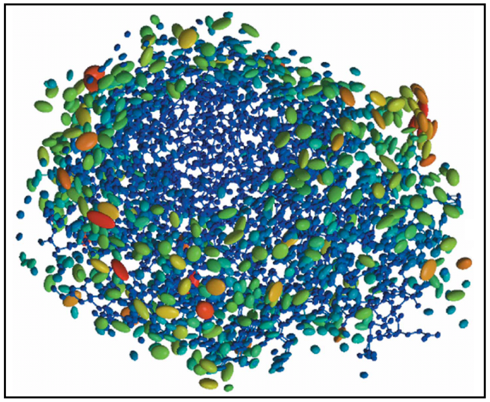
- Atomic displacements are anisotropic



$$\rho(\Delta \mathbf{r}) \sim \exp\{-\Delta \mathbf{r} \, , \, U^{-1} \Delta \mathbf{r}\}$$

$$U = \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{pmatrix}$$

- Hierarchy of atomic displacements



atom
residue
domain
molecule
crystal

- Typically three components are considered:

$$\Delta \mathbf{r}_{\text{total}} = \Delta \mathbf{r}_{\text{crystal}} + \Delta \mathbf{r}_{\text{domain}} + \Delta \mathbf{r}_{\text{atom}}$$

Model parameterization: Atomic Displacement Parameters ("B-factors")

- Hierarchy of atomic displacements

atom
residue
domain
molecule
crystal

Total atomic displacement

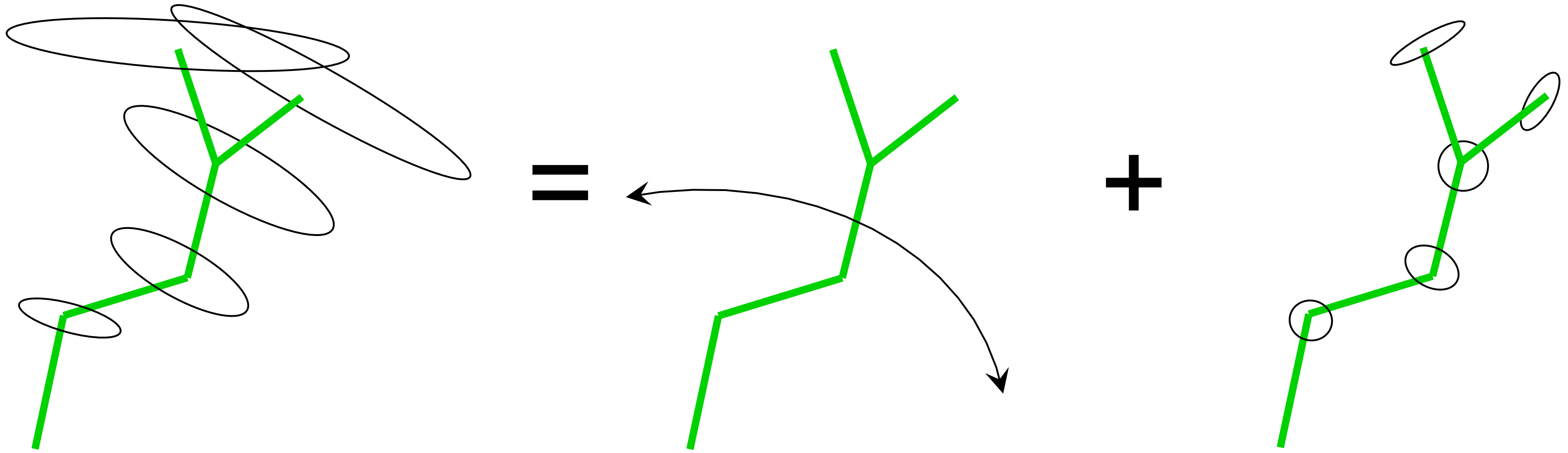
- Currently one can split and distinguish only a limited number of movements

$$\Delta \mathbf{r}_{total} = \Delta \mathbf{r}_{crystal} + \Delta \mathbf{r}_{domain} + \Delta \mathbf{r}_{atom}$$

Total movement ($\mathbf{U}_{\mathbf{TOTAL}}$)

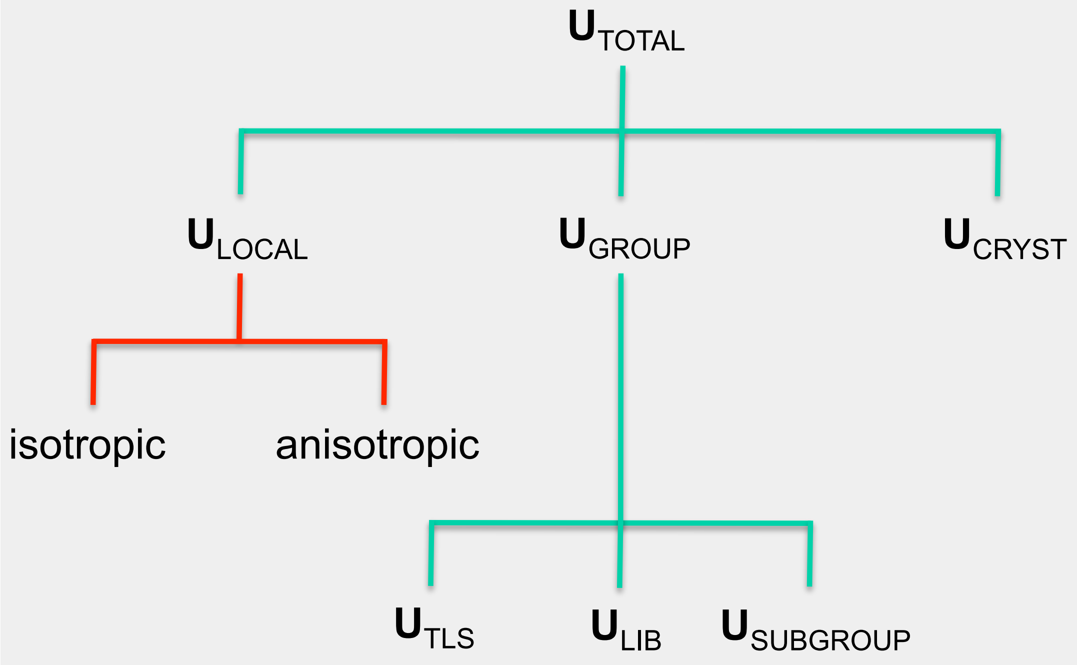
Collective movement ($\mathbf{U}_{\mathbf{TLS}}$)

Local movement of independent atoms ($\mathbf{U}_{\mathbf{ATOM}}$)



Atomic Displacement Parameters (ADP or “B-factors”)

- **Total ADP** $U_{\mathrm{TOTAL}} = U_{\mathrm{CRYST}} + U_{\mathrm{GROUP}} + U_{\mathrm{LOCAL}}$



- $\mathbf{U_{CRYST}}$ – overall anisotropic scale (6 parameters).
- $\mathbf{U_{TLS}}$ – rigid body displacements of molecules, domains, secondary structure elements. $U_{\mathrm{TLS}} = T + ALA^t + AS + S^tA^t$ (20 TLS parameters per group).
- $\mathbf{U_{LOCAL}}$ – local vibration of individual atoms.
- U_{LIB} – librational motion of side chain around bond vector.

Structure refinement

1. **Model parameters**
2. **Optimization goal**
3. **Optimization method**

Refinement target function

- **Structure refinement** is a process of changing a model parameters in order to optimize a goal (target) function:

$$T = F(\text{Experimental data}, \text{Model parameters}, \textit{A priori}\ \text{knowledge})$$

  - Experimental data – a set of diffraction amplitudes Fobs (and phases, if available).
  - Model parameters: coordinates, ADP, occupancies, bulk-solvent, …
  - *A priori* knowledge (restraints or constraints) – additional information that may be introduced to compensate for the insufficiency of experimental data (finite resolution, poor data-to-parameters ratio)

- Typically: $T = T_{\text{DATA}} + w{*}T_{\text{RESTRAINTS}}$
  - E_{DATA} relates model to experimental data
  - $E_{\text{RESTRAINTS}}$ represents *a priori* knowledge
  - w is a weight to balance the relative contribution of E_{DATA} and $E_{\text{RESTRAINTS}}$

- A priori knowledge can be imposed in the form of constraints so

$$T = E_{\text{DATA}}$$

Target function

$$T = T_{\mathrm{DATA}} + wT_{\mathrm{RESTRAINTS}}$$

- **Least-Squares (reciprocal space)**

$$T_{\mathrm{DATA}} = \sum_{\mathbf{s}} \mathbf{w}_s \left(F_{\mathbf{s}}^{\mathrm{OBS}} - kF_{\mathbf{s}}^{\mathrm{MODEL}} \right)^{\mathbf{2}}$$

  - Widely used in small molecule crystallography

  - Used in macromolecular crystallography in the past

- **Better option (for macromolecules): Maximum-Likelihood (reciprocal space)**

$$T_{\mathrm{DATA}} = \sum_{\mathrm{s}} (1 - K_{\mathrm{s}}^{cs}) \left(-\frac{\alpha_{\mathrm{s}}^2 \left(F_{\mathrm{s}}^{MODEL} \right)^2}{\varepsilon_{\mathrm{s}} \beta_{\mathrm{s}}} + ln\left(I_0 \left(\frac{2\alpha_{\mathrm{s}} F_{\mathrm{s}}^{MODEL} F_{\mathrm{s}}^{OBS}}{\varepsilon_{\mathrm{s}} \beta_{\mathrm{s}}} \right) \right) \right) + \\ + K_{\mathrm{s}}^{cs} \left(-\frac{\alpha_{\mathrm{s}}^2 \left(F_{\mathrm{s}}^{MODEL} \right)^2}{2\varepsilon_{\mathrm{s}} \beta_{\mathrm{s}}} + ln\left(cosh\left(\frac{\alpha_{\mathrm{s}} F_{\mathrm{s}}^{MODEL} F_{\mathrm{s}}^{OBS}}{\varepsilon_{\mathrm{s}} \beta_{\mathrm{s}}} \right) \right) \right)$$

- **Real space target**

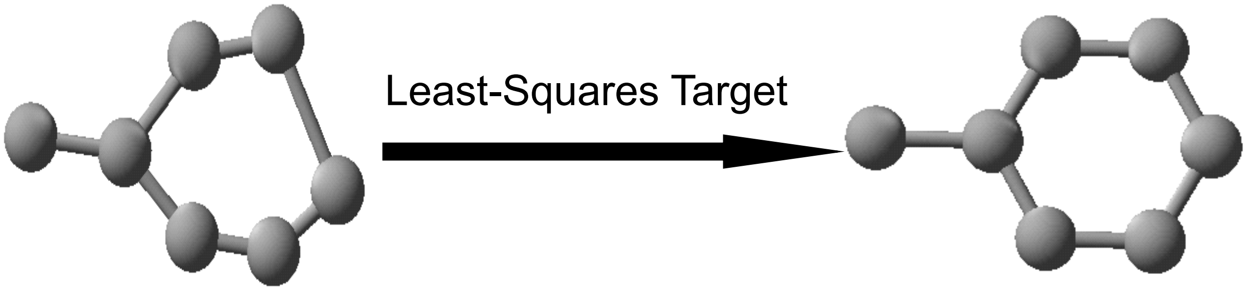
$$T_{\mathrm{DATA}} = \sum_{\text{grid points}} \left(\rho_{\mathrm{best}} - k\rho_{\mathrm{calc}} \right)^2$$

ρ_{best} - best available map: experimental, 2mFo-DFc

ρ_{calc} - calculated map from current atomic model

T_{DATA}: Least-Squares vs Maximum-Likelihood

- **Removable Errors** (never the case for macromolecular model, common for small molecules)

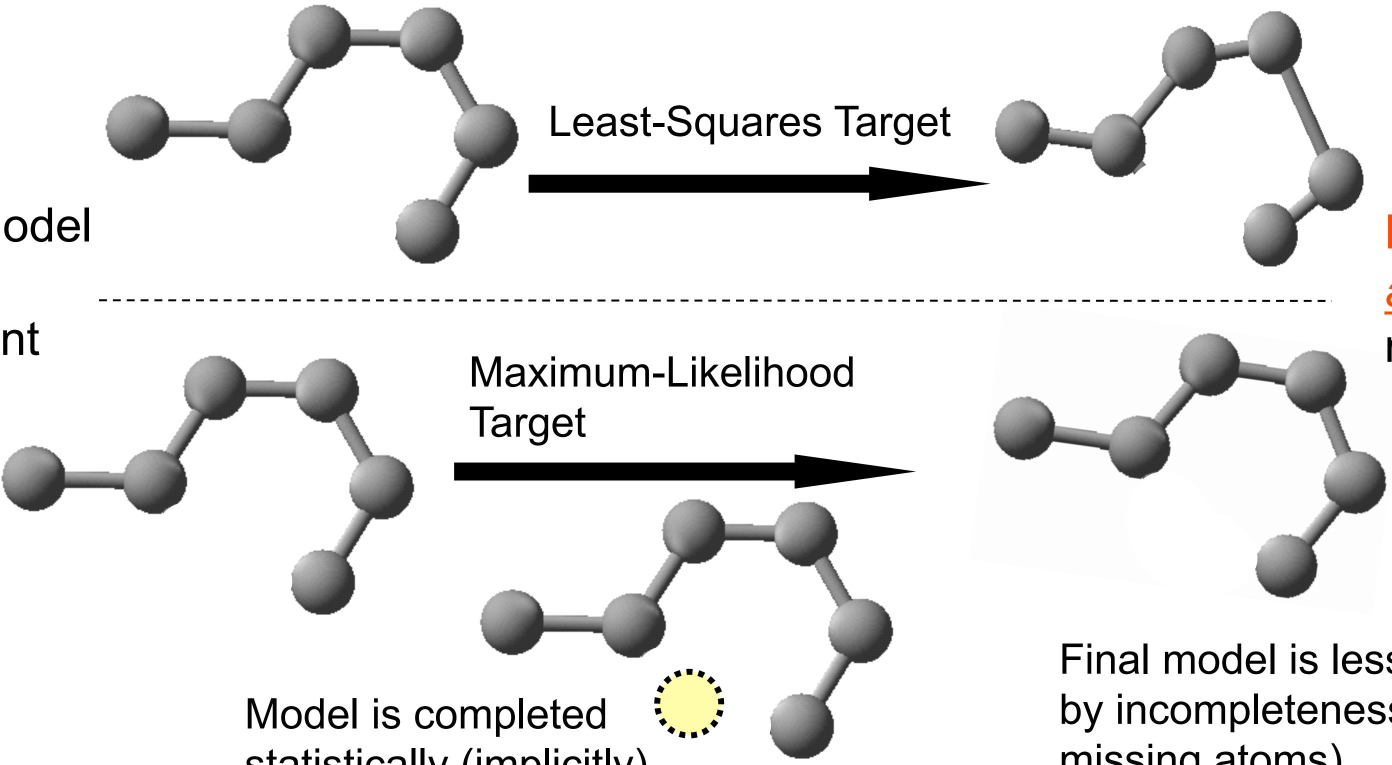


Complete model *before* refinement

Least-Squares Target

Complete model *after* refinement

- **Irremovable Errors** (always the case for macromolecular models)



Partial model *before* refinement

Least-Squares Target

Maximum-Likelihood Target

Partial model *after* refinement

Model is completed statistically (implicitly)

Final model is less affected by incompleteness (by missing atoms)

T_{DATA}: Least-Squares vs Maximum-Likelihood

- **Why Maximum-Likelihood target is better than Least-Squares** (in a nutshell):

  - ML accounts for model incompleteness (missing, unmodeled atoms) while LS doesn't;

  - ML automatically downweights the terms corresponding to reflections with the poor fit (poorly measured inaccurate F_{OBS}, high resolution reflections at the beginning of refinement, etc.)

- ***R*-factors in LS and ML refinement**:

  - *R*-factor is expected to decrease during LS based refinement, since the LS target and *R*-factor formula are very similar:

$$R = \frac{\sum \left| F_{\mathrm{OBS}} - F_{\mathrm{MODEL}} \right|}{\sum F_{\mathrm{MODEL}}} \qquad LS = \sum_{\mathbf{s}} \left(F_{\mathrm{OBS}} - F_{\mathrm{MODEL}} \right)^{2}$$

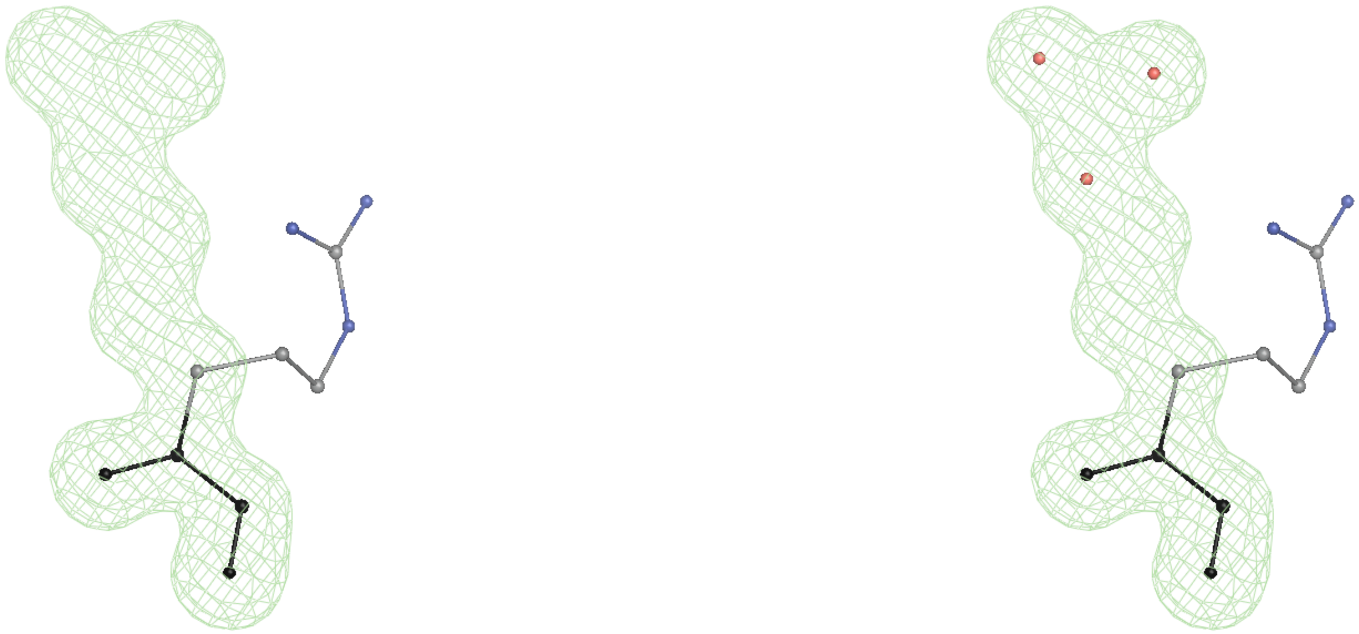
  - In ML based refinement the *R*-factor may eventually decrease (and this is what typically happens in practice) but this is not implied by the ML target function

Real-space refinement

Why real-space refinement ?

$$T = w \sum_{\text{grid points}} \left(\rho_{\text{best}} - k\rho_{\text{calc}}\right)^2 + T_{\text{RESTRAINTS}}$$

- Can be done locally (for example, for a residue or ligand), therefore can be very fast and target only problem places.
- Grid search can be used, therefore convergence radius can be increased compared to gradient-driven refinement or SA (discussed later)
- Ordered solvent (water) update can be enabled at earlier stage



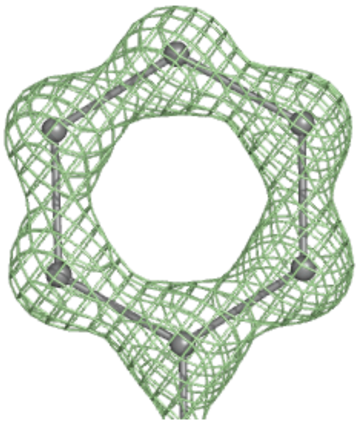
Real-space refinement

Real space refinement target $\mathrm{T} = w \sum_{\text{grid points}} \left(\rho_{\text{best}} - k\rho_{\text{calc}}\right)^2 + \mathrm{T}_{\text{RESTRAINTS}}$

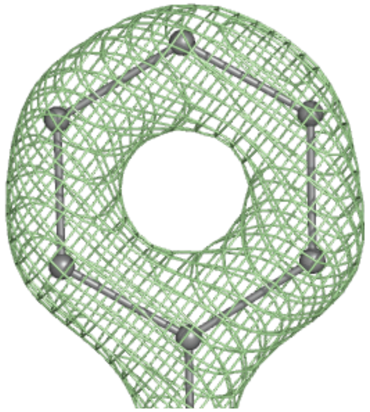
ρ_{best} is best available map: experimental, 2mFo-DFc, …

ρ_{calc} is calculated map from current atomic model

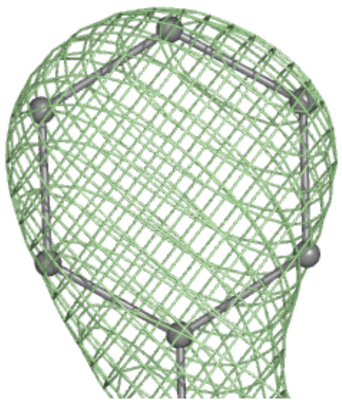
Fitting ρ_{calc} to ρ_{best} may be problematic because the exact ρ_{calc} computed from atomic model and its Fourier image (ρ_{best}) may look very different depending on resolution and data completeness:



exact



2 Å



3 Å

… so using the exact ρ_{calc} directly computed from atomic model may not be a good idea.

Real-space refinement

Solutions:

- Resolution and completeness dependent analytical functions for ρ_{calc} (M.S. Chapman; used in RSRef – a real-space refinement extension of CNS)
- Compute $\rho_{\text{calc}} = \text{FT}(\mathbf{F}_{\text{MODEL}})$ that naturally accounts for resolution and completeness.

Alternative – more simplistic target that moves atoms to the closest density peak:

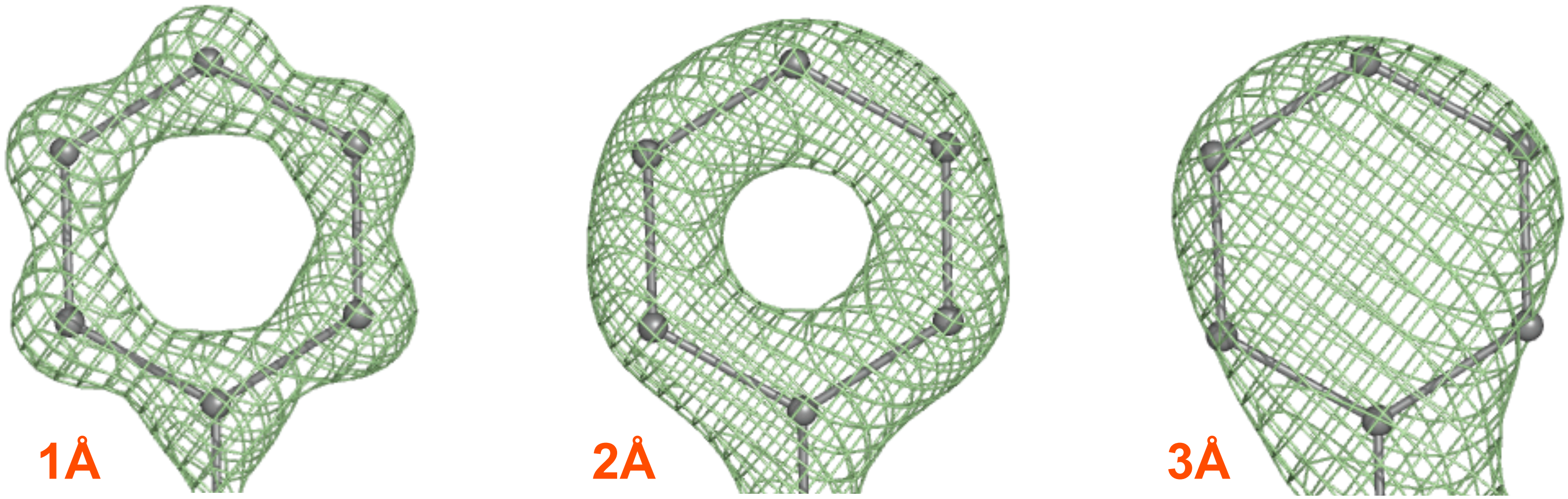
$$\mathrm{T} = -w \sum_{\text{atoms}} \rho_{\text{best}}\Big|_{\text{computed at atom center}} + \mathrm{T}_{\text{RESTRAINTS}}$$

- Fast (no need to re-compute ρ_{calc} each time an atom moved)
- Uses only one map (no issues related to dissimilarity of ρ_{calc} and ρ_{best} due to resolution)
- Less accurate since doesn't use shape of electron density (moves atoms to the closest density peak without considering how similar that peak is to expected density. May not be applicable at low resolution where the atomicity of the map is lost (no distinct peaks corresponding to atoms, but rather sphere and tube-like shapes).

Restraints in refinement of individual coordinates

$$E_{TOTAL} = w * E_{DATA} + E_{RESTRAINTS}$$

Fourier images at different data resolution



- *A priori* chemical knowledge is introduced (restraints) to keep the model chemically correct while fitting it to the experimental data at lower resolution (less resolution, stronger the weight *W*):

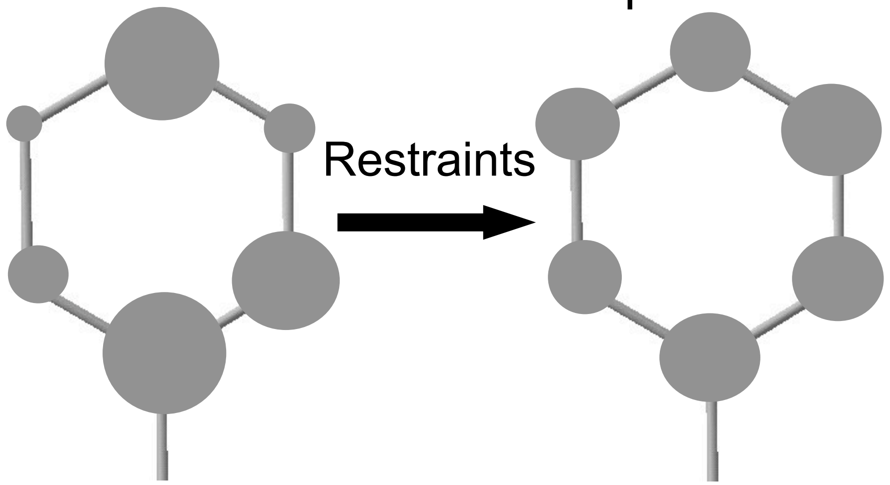
  $$E_{RESTRAINTS} = E_{BOND} + E_{ANGLE} + E_{DIHEDRAL} + E_{PLANARITY} + E_{NONBONDED} + \ldots$$

- Higher resolution – less restraints contribution (can be completely unrestrained for well ordered parts at subatomic resolution)

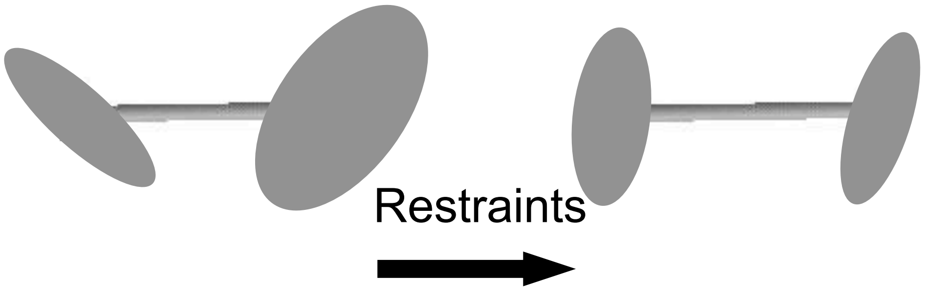
Restraints in refinement of individual ADP (Atomic Displacement Parameters)

$$E_{TOTAL} = w * E_{DATA} + E_{RESTRAINTS}$$

Refinement of isotropic ADP



Refinement of anisotropic ADP



✓ A bond is almost rigid, therefore the ADPs of bonded atoms are similar (Hirshfeld, 1976);

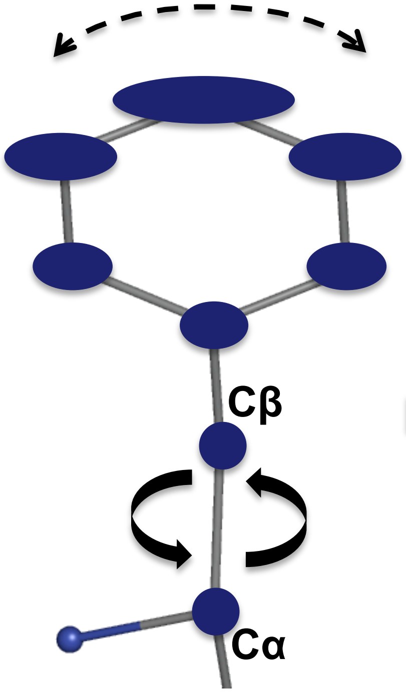
✓ ADPs of spatially close (non-bonded) atoms are similar (Schneider, 1996);

✓ The bond rigidity, and therefore the difference between the ADPs of bonded atoms, is related to the absolute values of ADPs. Atoms with higher ADPs can have larger differences (Ian Tickle, CCP4 BB, March 14, 2003).

$$E_{\mathrm{RESTRAINTS}} = \sum_{i=1}^{N_{atoms}} \left[\sum_{j=1}^{M_{atoms}} \frac{1}{r_{ij}^{\mathrm{distance\_power}}} \frac{\left(U_i - U_j\right)^2}{\left(\frac{U_i + U_j}{2}\right)^{\mathrm{average\_power}}} \Bigg|_{\mathrm{sphereR}} \right]$$

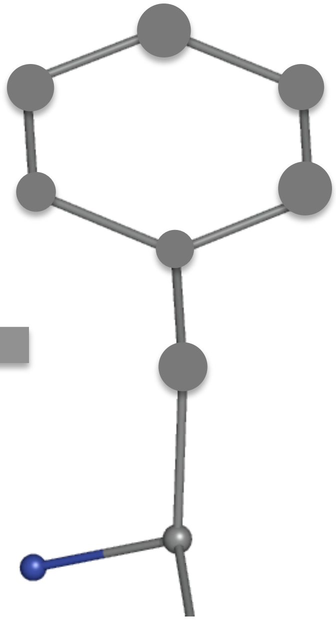
Restraints in refinement of individual ADP

- **A nuance about using similarity restraints**
  - Total ADP is: $\mathbf{U}_{\mathrm{TOTAL}} = \mathbf{U}_{\mathrm{CRYST}} + \mathbf{U}_{\mathrm{GROUP}} + \mathbf{U}_{\mathrm{LOCAL}}$
  - Similarity restraints should be applied to $\mathbf{U}_{\mathrm{LOCAL}}$
  - Applying it to $\mathbf{U}_{\mathrm{TOTAL}}$ is much less justified



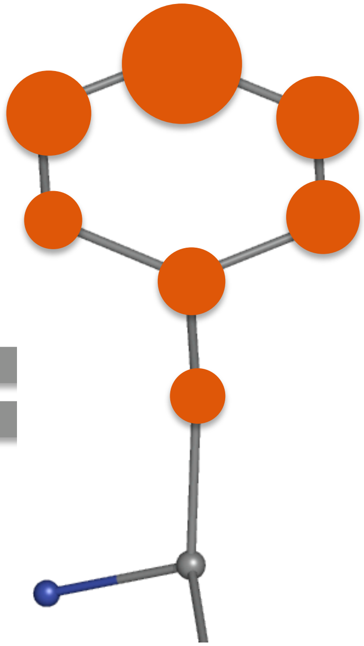
Rigid-body libration around Cα-Cβ bond vector, $\mathbf{U}_{\mathrm{GROUP}}$

+



Small local atomic vibrations, $\mathbf{U}_{\mathrm{LOCAL}}$

=



Resulting isotropic equivalent, $\mathbf{U}_{\mathrm{TOTAL}}$

Examples of constraints

- Rigid body refinement: mutual positions of atoms within a rigid groups are forced to remain the same, while the rigid group can move as a whole. 6 refinable parameters per rigid group (3 translations + 3 rotations).
- Constrained rigid groups: torsion angle parameterization. Reduction of refinable parameters by a factor between 7 and 10.
- Occupancies of atoms in alternative conformations: occupancies of alternate conformers must add up to 1.
- Group ADP refinement: mutual distribution of all B-factors within the group must remain the same. One refinable B-factor per group.
- Constrained NCS refinement: a number of *N* NCS related molecules or domains are assumed to be identical. Reduction of refinable parameters by a factor N.

- Do not confuse restraints and constraints

  Constraints: model property = ideal value

  Restraints: model property ~ ideal value

Refinement target weight

- Refinement target $E_{TOTAL} = w * E_{DATA} + E_{RESTRAINTS}$
  - the weight ***w*** is determined automatically
  - in most of cases the automatic choice is good

- If automatic choice is not optimal there are two possible refinement outcomes:
  - structure is over-refined: *Rfree-Rwork* is too large. This means the weight ***w*** is too small making the contribution of E_{DATA} too large.
  - weight ***w*** is too large making the contribution of restraints too strong. This results increase of *Rfree* and/or *Rwork*.
  - A possible approach to address this problem is to perform a grid search over an array of ***w*** values and choose the one ***w*** that gives the best *Rfree* and *Rfree-Rwork*.
- A random component is involved in ***w*** calculation. Therefore an ensemble of identical refinement runs each done using different random seed will result in slightly different structures. The *R*-factor spread depends on resolution and may be as large as 1…2%.

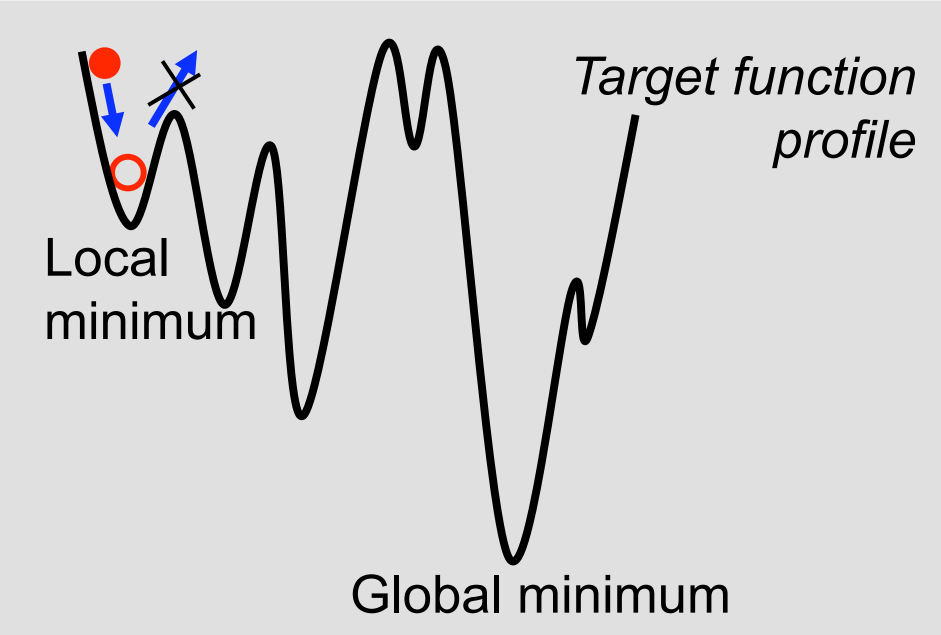
Structure refinement

1. **Model parameters**
2. **Optimization goal**
3. **Optimization method**

Refinement target optimization algorithms

- **Gradient-driven minimization**

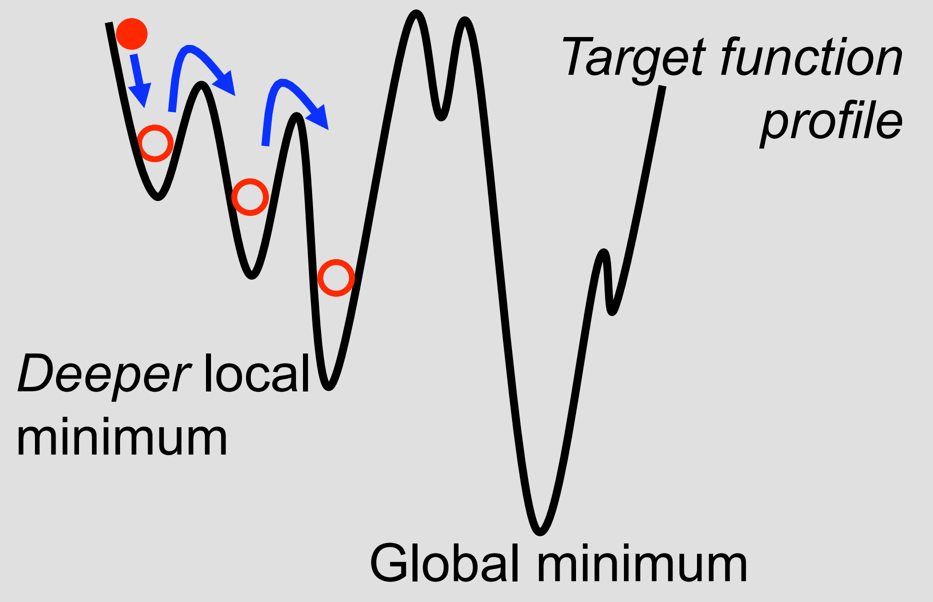
- Follows the local gradient.
- The target function depends on many parameters – many local minima.



Refinement target optimization algorithms

- **Simulated annealing (SA)**

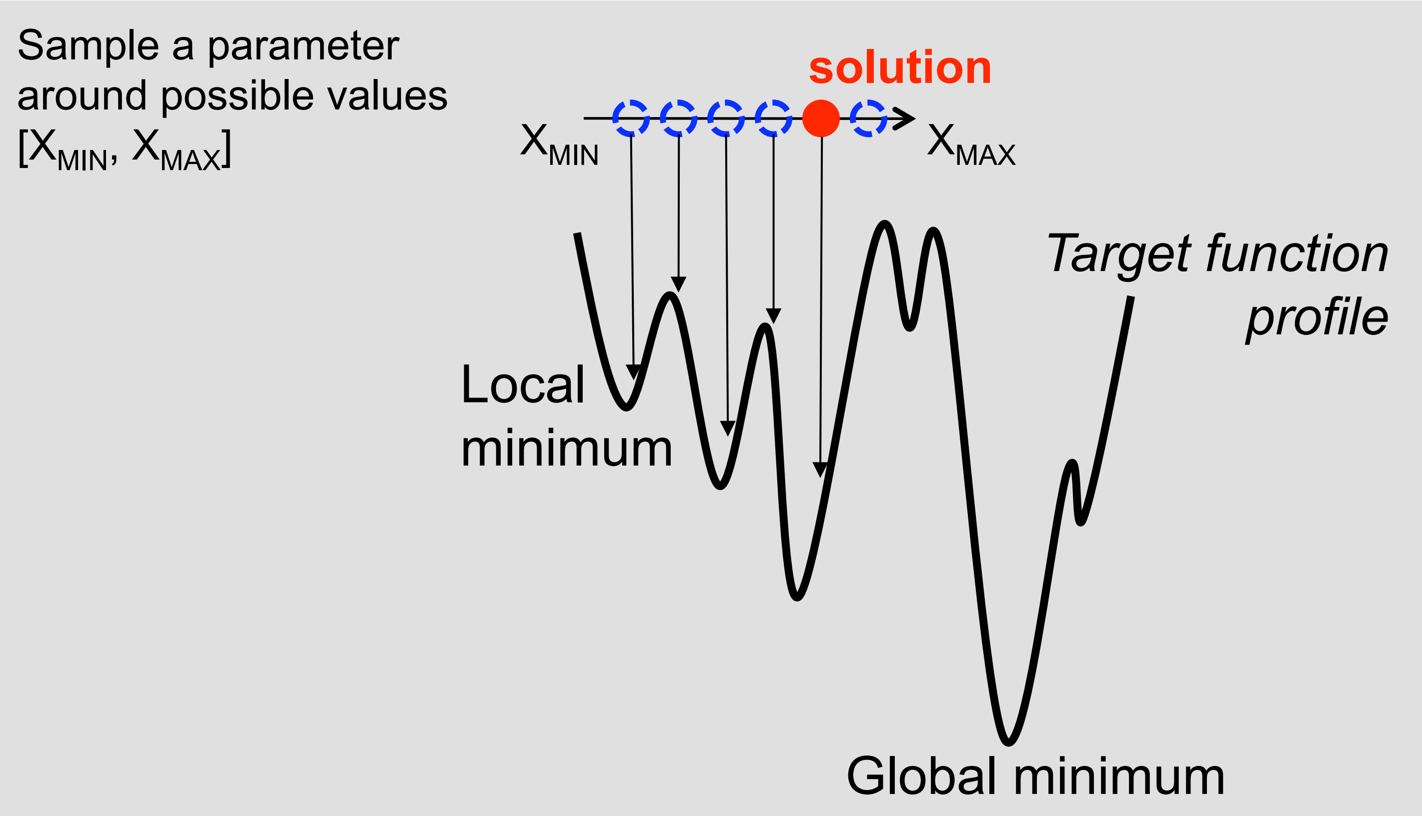
- SA is an optimization method which is good at escaping local minima.
- Annealing is a physical process where a solid is heated until all particles are in a liquid phase, followed by cooling which allows the particles to move to the lowest energy state.
- Simulated annealing is the simulation of the annealing process.
  - Increased probability of finding a better solution because motion against the gradient is allowed.
  - Probability of uphill motion is determined by the temperature.



Refinement target optimization algorithms

- **Grid search**

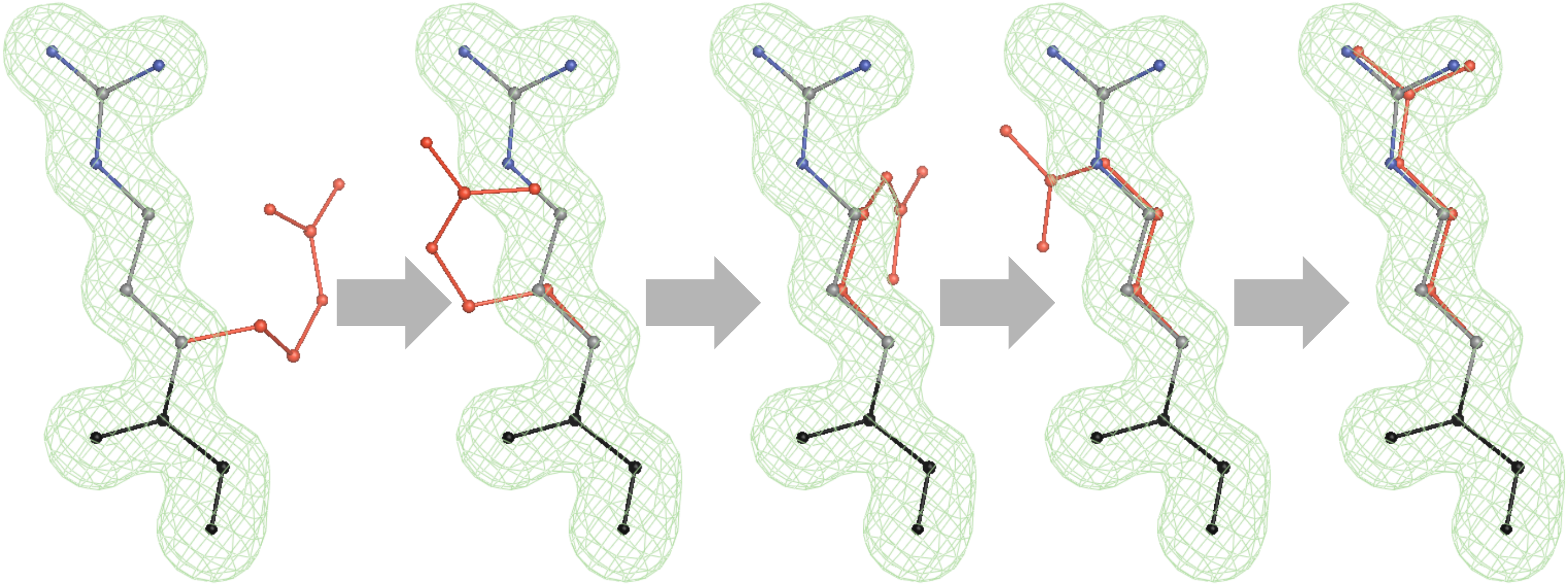
Robust but may be time inefficient for many parameter systems, and not as accurate as gradient-driven. Good for small number of parameters (2-3 or so), and impractical for larger number of parameters.



Sample a parameter around possible values $[X_{MIN}, X_{MAX}]$

Grid search examples

- **Real-space sampling to fit density**

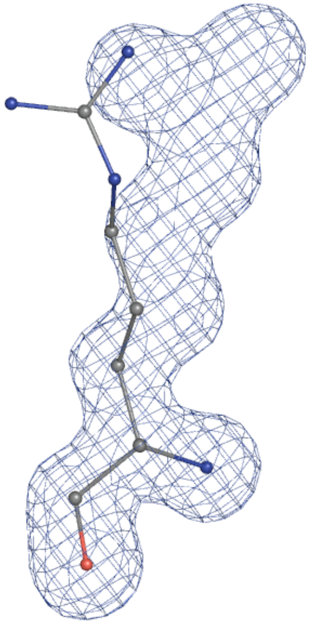


- **Finding bulk-solvent k_{SOL} and B_{SOL}**
  - k_{SOL} : [0.2, 0.6]
  - B_{SOL} : [10, 100]

$$\mathbf{F}_{\mathrm{MODEL}} = k_{\mathrm{OVERALL}} e^{-\mathbf{s}\mathbf{U}_{\mathrm{CRYSTAL}}\mathbf{s}^{t}} \left(\mathbf{F}_{\mathrm{CALC\_ATOMS}} + k_{\mathrm{SOL}} e^{-\frac{B_{\mathrm{SOL}} s^{2}}{4}} \mathbf{F}_{\mathrm{MASK}} \right)$$

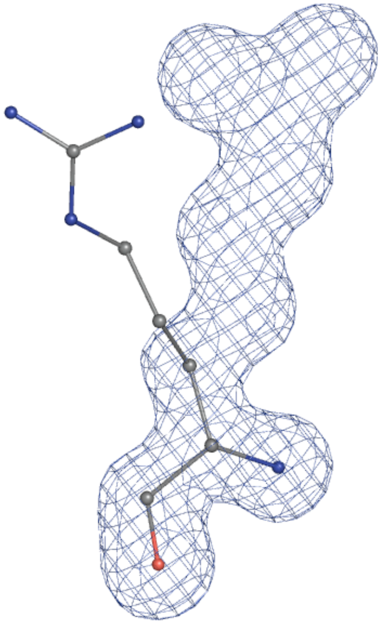
Refinement convergence

Minimization



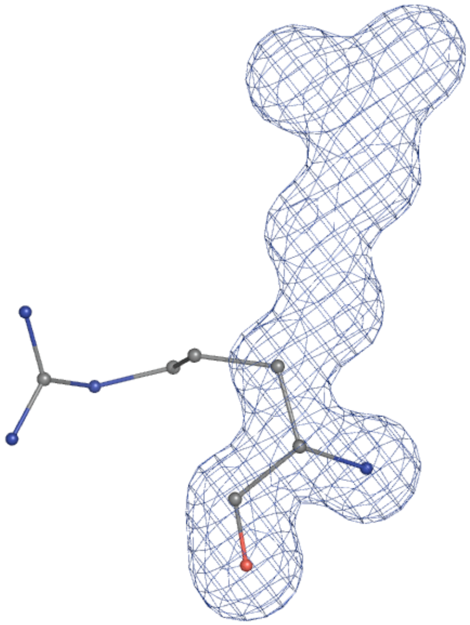
Both minimization and SA can fix it

Simulated Annealing



This is beyond the convergence radius for minimization

Real-space grid search



This is beyond the convergence radius for minimization and SA

Summary on refinement decisions

- **Parameterization:**
  - Coordinates: restraints *vs* constraints (Rigid body or its special case - Torsion angles)
  - ADP: aniso/isotropic, groups, individual, TLS
  - NCS: constrained, restrained, ignored
- **Optimization algorithm:**
  - Simulated annealing
  - Minimization (first or second derivatives methods)
- **Target function:**
  - Chemical information (chemical restraints, NCS similarity)
  - Maximum likelihood
  - Experimental phases
  - X-ray, neutron, joint X+N
  - Real-space

Refinement - summary

- **Refinement is:**
  - Process of changing model parameters to optimize a target function
  - Various tricks are used (restraints, different model parameterizations) to compensate for imperfect experimental data

- **Refinement is NOT :**
  - Getting a ‘low enough’ R-value (for supervisors or referees)
  - Getting ‘low enough’ B-values (for supervisors or referees)
  - Completing the sequence in the absence of density

Typical refinement steps

- **Input data and model processing:**
  - Read in and process PDB file
  - Read in and process library files (for non-standard molecules, ligands)
  - Read in and process reflection data file
  - Check correctness of input parameters
  - Create objects that will be reused in refinement later on (geometry restraints,…)
- **Main refinement loop (macro-cycle; repeated several times):**
  - Bulk solvent correction, anisotropic scaling, twinning parameters estimation
  - Update ordered solvent (water) (add or remove)
  - Target weights calculation
  - Refinement of coordinates (rigid body, individual) (minimization or Simulated Annealing)
  - ADP refinement (TLS, group, individual isotropic or anisotropic)
  - Occupancy refinement (individual, group, constrained)
- **Output results:**
  - PDB file with refined model
  - Various maps (2mFo-DFc, mFo-DFc) in various formats (CNS, MTZ)
  - Complete statistics
  - Structure factors

Literature

Maximum-likelihood in structure refinement

Lunin, V. Y. & Urzhumtsev, A. (1984). *Acta Cryst.* A**40**, 269-277.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140-149.
Read, R. J. (1990). *Acta Cryst.* A**46**, 900-912.
V.Yu., Lunin & T.P., Skovoroda. Acta Cryst. (1995). A51, 880-887.
Pannu, N.S., Murshudov, G.N., Dodson, E.J. & Read, R.J. (1998). Acta Cryst. D54, 1285-1294.
V.Y., Lunin, P.V. Afonine & A.G., Urzhumtsev. Acta Cryst. (2002). A58, 270-282
P. Afonine, V.Y. Lunin & A. Urzhumtsev. J. Appl. Cryst. (2003). 36, 158-159.

ADP

V. Schomaker & K.N. Trueblood. Acta Cryst. (1968). B24, 63-76.
F.L. Hirshfeld. Acta Cryst. (1976). A32, 239-244.
T.R. Schneider. Proceedings of the CCP4 Study Weekend, SERC Daresbury Laboratory, Daresbury, U.K., pp. 133-144 (1996).
M.D. Winn, M.N. Isupov & G.N. Murshudov. Acta Cryst. (2001). D57, 122-133.
R.W. Grosse-Kunstleve & P.D. Adams. J. Appl. Cryst. (2002). 35, 477-480.
P. Afonine & A. Urzhumtsev. (2007). CCP4 Newsletter on Protein Crystallography. 45. Contribution 6.
I. Haneef, D. S. Moss, M. J. Stanford and N. Borkakoti. Acta Cryst. (1985). A41, 426-433.
Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* 22, 510-516.
J Painter & E A Merritt (2006) Acta Cryst. D62, 439-450.
J Painter & E A Merritt (2006) J. Appl. Cryst. 39, 109-111.

Rigid body refinement

P.V. Afonine, R.W. Grosse-Kunstleve, A. Urzhumtsev and P.D. Adams. J. Appl. Cryst. (2009). 42, 607-615

Bulk-solvent modeling and anisotropic scaling

S. Sheriff & W.A. Hendrickson. Acta Cryst. (1987). A43, 118-121.
Jiang, J.-S. & Brünger, A. T. (1994). J. Mol. Biol. 243, 100-115.
A. Fokine & A. Urzhumtsev. Acta Cryst. (2002). D58, 1387-1392.
P.V. Afonine, R.W. Grosse-Kunstleve & P.D. Adams. Acta Cryst. (2005). D61, 850-855.

Refinement at subatomic resolution

Afonine, P.V., Pichon-Pesme, V., Muzet, N., Jelsch, C., Lecomte, C. & Urzhumtsev, A. (2002). CCP4 Newsletter on Protein Crystallography. 41.
Afonine P.V., Lunin, V., Muzet, N. & Urzhumtsev, A. (2004). Acta Cryst., D60, 260-274.
P.V. Afonine, R.W. Grosse-Kunstleve, P.D. Adams, V.Y. Lunin, A. Urzhumtsev. Acta Cryst. (2007). D63, 1194-1197.

LBFGS minimization

Liu, D.C. & Nocedal, J. (1989). Mathematical Programming, 45, 503-528. "On the limited memory BFGS method for large scale optimization"

Dynamics, simulated annealing refinement

Brünger, A.T., Kuriyan, J., Karplus, M. (1987). Science. 235, 458-460.
Adams, P.D., Pannu, N.S., Read, R.J. & Brünger, A.T. (1997). Proc. Natl. Acad. Sci. 94, 5018-5023.
L.M. Rice, Y. Shamoo & A.T. Brünger. J. Appl. Cryst. (1998). 31, 798-805.
Brünger, A.T & Adams, P.D. (2002). Acc. Chem. Res. 35, 404-412.

Target weights calculation

Brünger, A.T., Karplus, M. & Petsko, G.A. (1989). Acta Cryst. A45, 50-61.
Brünger, A.T. (1992). Nature (London), 355, 472-474.
Adams, P.D., Pannu, N.S., Read, R.J. & Brünger, A.T. (1997). Proc. Natl. Acad. Sci. 94, 5018-5023.

Electron density maps (Fourier syntheses) calculation

A.G. Urzhumtsev, T.P. Skovoroda & V.Y. Lunin. J. Appl. Cryst. (1996). 29, 741-744.

Monomer Library

Vagin, A.A., Steiner, R.A., Lebedev, A.A, Potterton, L., McNicholas, S., Long, F. & Murshudov, G.N. (2004). Acta Cryst. D60, 2184-2195.

Refinement

Tronrud. Acta Cryst. (2004). D60, 2156-2168.
Watkin. J. Appl. Cryst. (2008). 41, 491-522.
Afonine, P.V., Grosse-Kunstleve, R.W. & Adams, P.D. (2005). CCP4 Newsl. 42, contribution 8.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Acta Cryst. D53, 240-255.

Literature

Real-space refinement

Booth, A. D. (1946). Proc. Roy. Soc. London Ser. A, 188, 77-92.
Booth, A. D. (1947). Proc. Roy. Soc. London Ser A, 190, 482-489.
Cochran, W. (1948). Acta Cryst. 1, 138-142.
Cochran, W. *Acta Cryst.* (1951). **4**, 408-411
Cruickshank, D. W. (1952). Acta Cryst. 5, 511-518.
Cruickshank, D. W. (1956). Acta Cryst. 9, 747-753.
R. Diamond Acta Cryst. (1971). A27, 436
R. Diamond. J. Mol. Biol. (1974) 82, 371-391
R.J. Fletterick and H.W. Wyckof Acta Cryst. (1975). A31, 698
J.C. Hanson and B.P. Schoenborn J. Mol. Biol. (1981) 153, 117-146
S. Fitzwater and H. A. Scheraga Proc. NatL Acad. Sci. USA Vol. 79, pp. 2133-2137, March 1982
R.Diamond. (1985). Methods Enzymol. 115, 237-252.
S.Freer. (1985). Methods Enzymol. 115, 235-237.
Deisenhofer, J., Remington, S. J. & Staigemann, W. (1985). Methods in Enzymology, Vol. 115B, pp. 303-323.
A.G. Urzhumtsev, V. YU. Lunin and E. A. Vernoslova *J. Appi. Cryst. (1989). 22, 500-506*
T.A. Jones, J.-Y. Zou and S.W. Cowan, M. Kjeldgaard Acta Cryst. (1991). A47, 110-119
R.J. Read and J. Moult Acta Cryst. (1992). A48, 104-113
V.S. Lamzin and K.S. Wilson Acta Cryst. (1993). D49, 129-147
D.G. Levit, L.J. Banasza J. Appl. Cryst. (1993). 26, 736-745
V.Y. Lunin, M.M. Woolfson Acta Cryst. (1993). D49, 530-533
J.-Y. Zou, S.L. Mowbray Acta Cryst. (1994). D50, 237-249
M.S. Chapman Acta Cryst. (1995). A51, 69-80
M.S. Chapman and M.G. Rossmann Acta Cryst. (1996). D52, 129-142
V.Yu. Lunin and N.L. Lunina Acta Cryst. (1996). A52, 365-368
J.-Y. Zou and T. A. Jones Acta Cryst. (1996). D52, 833-841
E. Blanc and M.S. Chapman J. AppL CrysL (1997). 30, 566-567
M.S. Chapman and E. Blanc Acta Cryst. (1997). D53, 203-206
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Acta Cryst. D53, 240-255
E. Blanc, Z. Chen and M.S. Chapman, Direct Methods for solving macromolecular structures, 513-519, 1998
Z. Chen, E. Blanc and M.S. Chapman. Acta Cryst. (1999). D55, 219-224
Z. Chen, E. Blanc and M.S. Chapman. Acta Cryst. (1999). D55, 464-468
G. Zhou, T. Somasundaram, E. Blanc, Z. Chena and M.S. Chapman, Acta Cryst. (1999). D55, 835-845
T. Holton, T.R. Ioerger, J.A. Christopher and J.C. Sacchettini Acta Cryst. (2000). D56, 722-734
G.J. Kleywegt Acta Cryst. (2000). D56, 249-265
C.E. Wang Acta Cryst. (2000). D56, 1591-1611
Z. Chen and M.S. Chapman Biophysical Journal Volume 80 March 2001 1466-1472
L.F. Chen, E. Blanc, M.S. Chapman and K.A. Taylor Journal of Structural Biology 133, 221–232 (2001)
D.G. Levitt Acta Cryst. (2001). D57, 1013-1019
T.J. Oldfield Acta Cryst. (2001). D57, 82-94
A. Korostelev, R. Bertramc, and M.S. Chapmana, Acta Cryst. (2002). D58, 761-767
J.Z. Chen, J. Furst, M.S. Chapman and N. Grigorieff Journal of Structural Biology 144 (2003) 144-151
J.F. Hunta and J. Deisenhofer Acta Cryst. (2003). D59, 214-224
Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S. M., Van Roey, P., Agrawal, R. K., Harvey, S. C., Sali, A., Chapman, M. S. & Frank, J. (2003). *Cell*, **113**, 789-801.
F. Pavelcik Acta Cryst. (2003). A59, 487-494
S.X. Cohen, R.J. Morris, F.J. Fernandez, M. Ben Jelloul, M. Kakaris, V. Parthasarathy, V.S. Lamzin, G.J. Kleywegt and A. Perrakis Acta Cryst. (2004). D60, 2222-2229
L. Potterton, S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, G.N. Murshudov, S. Cohen, A. Perrakis and M. Noble Acta Cryst. (2004). D60, 2288-2294
J. Aishima, D.S. Russel, L.J. Guibas, P.D. Adams and A.T. Brungera Acta Cryst. (2005). D, 1354–1363
H. van den Bedem, I. Lotan, J.-C. Latombe and A.M. Deacon Acta Cryst. (2005). D61, 2-13
M.S. Chapman , F. Fabiola, A. Korostelev, M. Fenley Acta Cryst. (2005). A61, c114-c115
B. DeLaBarre and A.T. Brunger Acta Cryst. (2006). D62, 923–932
N. Furnham, A.S. Dore, D.Y. Chirgadze, P.I.W. de Bakker, M.A. DePristo, and T. L. Blundell Structure 14, 1313–1320, August 2006
S. Wlodek, A. G. Skillman and A. Nicholls Acta Cryst. (2006). D62, 741–749
K. Joosten, S.X. Cohen, P. Emsley, W. Mooij, V.S. Lamzin and A. Perrakis Acta Cryst. (2008). D64, 416–424
J. L. Knight, Z. Zhou, E. Gallicchio, D.M. Himmel, R.A. Friesner, E. Arnold and R. M. Levy Acta Cryst. (2008). D64, 383–396
D. Turk Acta Cryst. (2008). A64, C23
H. van den Bedem, A. Dhanik, J.-C. Latombe and A.M. Deacon Acta Cryst. (2009). D65, 1107–1117
J.J. Headd, R. M. Immormino, D.A. Keedy, P. Emsley, D.C. Richardson, J.S. Richardson J Struct Funct Genomics (2009) 10:83–93