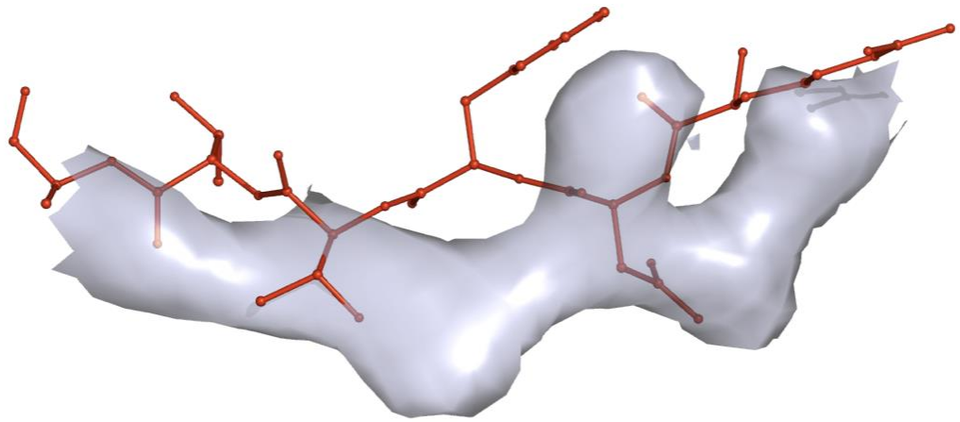
AsCA meeting,
6 December 2025

Refinement

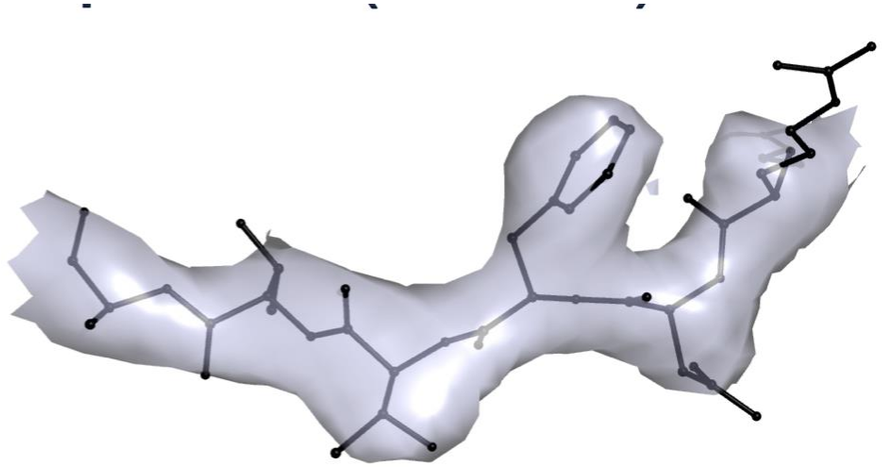
Dorothee Liebschner
Lawrence Berkeley Laboratory

Refinement

Initial model



Improved model



Fit atomic model to experimental data with the help of some known *a priori* information.

Refinement

Initial model

Experimental data

A priori knowledge

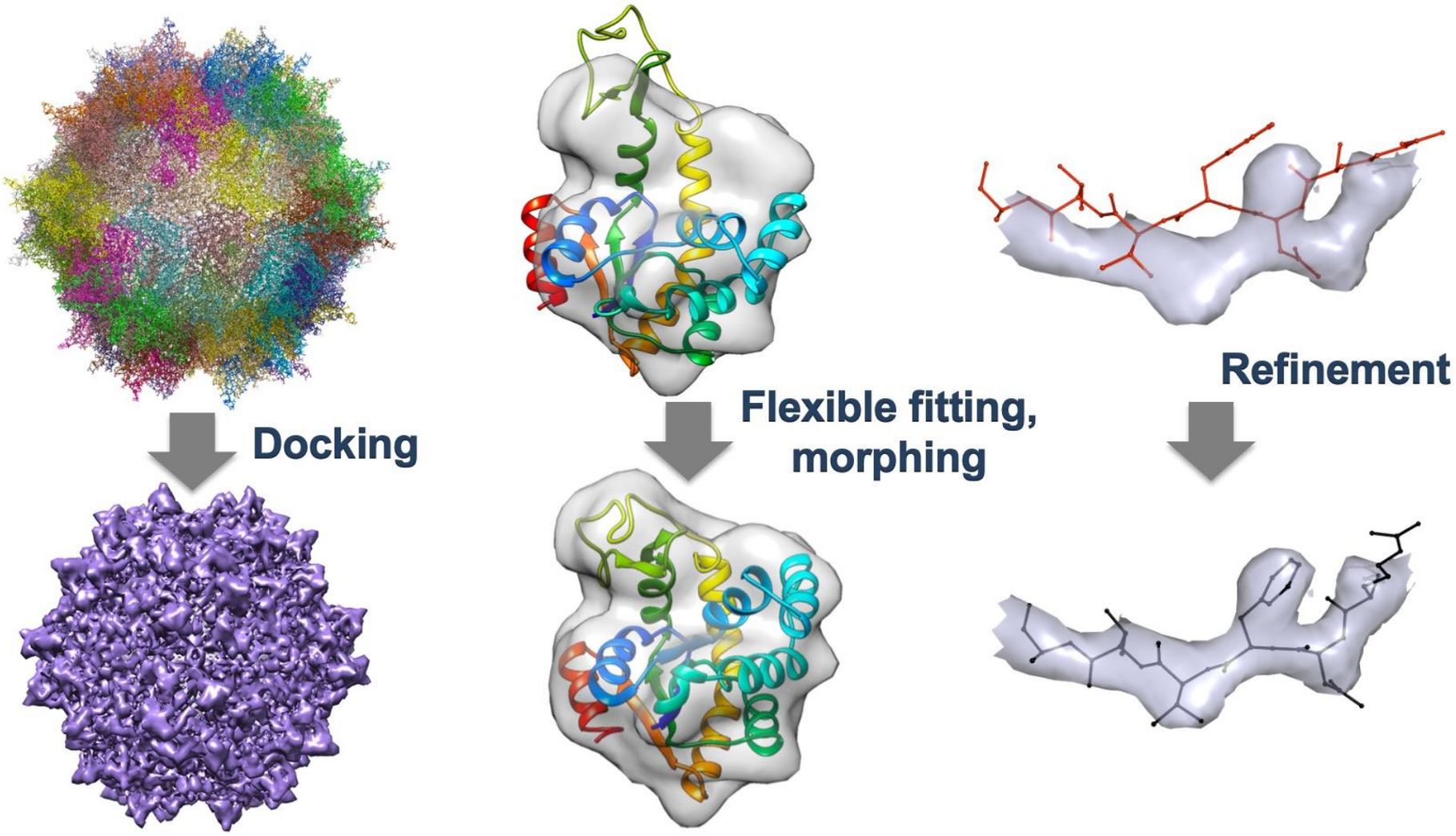
Calculate F_{model} from current model

Score F_{model} against F_{obs}

Optimize model so F_{model} are closer to F_{obs}

Improved model

Refinement



- Docking, morphing are **not** refinement
- Refinement is to fine-tune an already fine atomic model
- Refinement only applies small changes to the model (within the convergence radius of refinement, ~1Å)

Refinement: black box

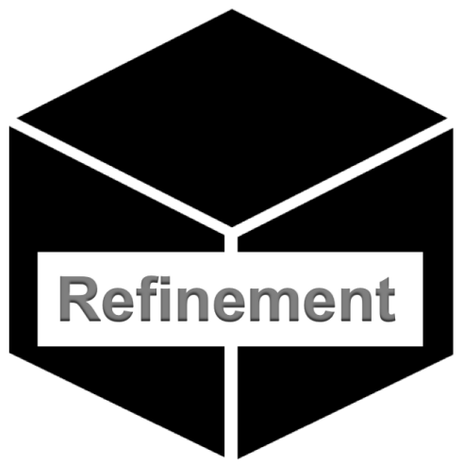
- Does it always work?
- Is it always as easy as poor model in, better model out?

Refinement: black box

- Does it always work?
- Is it always as easy as poor model in, better model out?

No. Because:

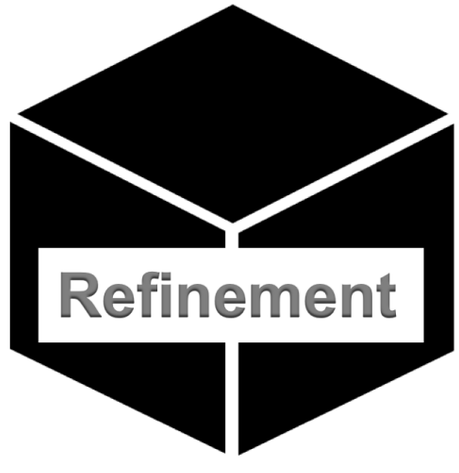
- Model parameterization is not easy.
- Default settings suit most common scenario (typical data resolution, model reasonably fits data)
- Less typical situations need customizations
  - Low/high resolution data
  - Incomplete models
  - Novel ligands



Refinement: black box

How do you know if...

- … refinement worked
- … you did it correctly?
- … the model is good enough to publish?



Do validation!

Standard validation protocols are designed to answer these questions.

Refinement

Use an *optimization* algorithm to minimize a *target function* of a set of *observations* by changing the *parameters* of a model.

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Introduction to macromolecular refinement

Dale. E. Tronrud

Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

The process of refinement is such a large problem in function minimization that even the computers of today cannot perform the calculations to properly fit X-ray diffraction data. Each of the refinement packages currently under development reduces the difficulty of this problem by utilizing a unique combination of targets, assumptions and optimization

Received 5 April 2004
Accepted 21 September 2004

Tronrud, D. E. Introduction to macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. 60, 2156–2168 (2004).

Refinement

Use an *optimization* algorithm to minimize a *target function* of a set of *observations* by changing the *parameters* of a model.

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Introduction to macromolecular refinement

Dale. E. Tronrud

Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

The process of refinement is such a large problem in function minimization that even the computers of today cannot perform the calculations to properly fit X-ray diffraction data. Each of the refinement packages currently under development reduces the difficulty of this problem by utilizing a unique combination of targets, assumptions and optimization

Received 5 April 2004
Accepted 21 September 2004

Tronrud, D. E. Introduction to macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. 60, 2156–2168 (2004).

Refinement

- *parameters* of a model
- *target function*
- *optimization* algorithm
- *observations*

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Introduction to macromolecular refinement

Dale. E. Tronrud

Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

Received 5 April 2004
Accepted 21 September 2004

The process of refinement is such a large problem in function minimization that even the computers of today cannot perform the calculations to properly fit X-ray diffraction data. Each of the refinement packages currently under development reduces the difficulty of this problem by utilizing a unique combination of targets, assumptions and optimization

Tronrud, D. E. Introduction to macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. 60, 2156–2168 (2004).

Refinement

- *parameters* of a model
- *target function*
- *optimization* algorithm
- *observations*

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Introduction to macromolecular refinement

Dale. E. Tronrud

Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

Received 5 April 2004
Accepted 21 September 2004

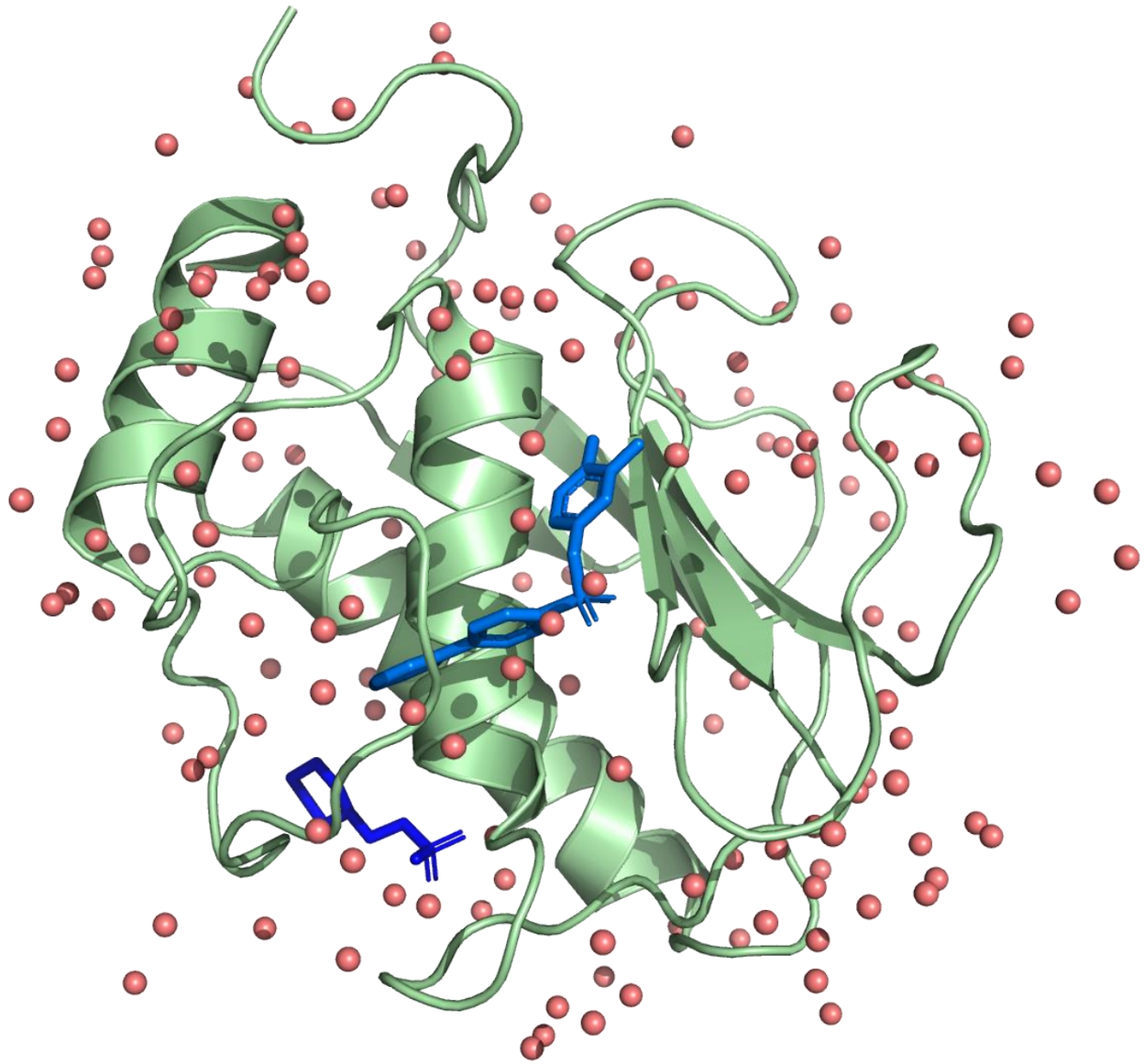
The process of refinement is such a large problem in function minimization that even the computers of today cannot perform the calculations to properly fit X-ray diffraction data. Each of the refinement packages currently under development reduces the difficulty of this problem by utilizing a unique combination of targets, assumptions and optimization

Tronrud, D. E. Introduction to macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. 60, 2156–2168 (2004).

Parameters of a model

Our model has parameters that describe the crystal and its content.

- **Atomic:**
  - coordinates
  - B-factors
  - occupancies
- **Non-atomic:**
  - bulk-solvent
  - anisotropy
  - twinning



Parameters of a model

Our model has parameters that describe the crystal and its content.

- **Atomic:**
  - coordinates
  - B-factors
  - occupancies

  } Saved in the model file (.pdb, .mmcif)

- **Non-atomic:**
  - bulk-solvent
  - anisotropy
  - twinning

Parameters of a model

Our model has parameters that describe the crystal and its content.

- **Atomic:**
  - coordinates
  - B-factors
  - occupancies
- **Non-atomic:**
  - bulk-solvent
  - anisotropy
  - twinning

Taken into account automatically by refinement program (e.g., bulk solvent)

Set by the user (e.g., twinning).

Atomic model parameters

To calculate a score (compare measured and model-based structure factor amplitudes), we need to compute structure factors from model parameters.

Ca atom in a Pro residue:

```
ATOM     25  CA  PRO A   4      31.309  29.489  26.044  1.00 57.79           C
```

$$\boldsymbol{F_{calc(atoms)}}(hkl) = \sum_{n}^{N_{\text{atoms}}} \left(q_n f_n(\boldsymbol{s}) e^{-B_n \frac{sin^2\theta}{\lambda^2}} e^{2\pi i \boldsymbol{s} \boldsymbol{r}_n} \right)$$

Atomic model parameters

To calculate a score (compare measured and model-based structure factor amplitudes), we need to compute structure factors from model parameters.

Ca atom in a Pro residue:

```
ATOM     25  CA  PRO A   4      31.309  29.489  26.044  1.00 57.79           C
```

31.309 29.489 26.044
Atomic coordinates (position)

$$\boldsymbol{F_{calc(atoms)}}(hkl) = \sum_{n}^{N_{\text{atoms}}} \left(q_n f_n(\boldsymbol{s}) e^{-B_n \frac{sin^2\theta}{\lambda^2}} e^{2\pi i \boldsymbol{s} \boldsymbol{r}_n} \right)$$

Atomic model parameters

To calculate a score (compare measured and model-based structure factor amplitudes), we need to compute structure factors from model parameters.

Ca atom in a Pro residue:

```
ATOM     25  CA  PRO A   4      31.309  29.489  26.044  1.00 57.79           C
```

$$\boldsymbol{F_{calc(atoms)}}(hkl) = \sum_{n}^{N_{\text{atoms}}} \left(q_n f_n(\boldsymbol{s}) e^{-B_n \frac{sin^2\theta}{\lambda^2}} e^{2\pi i \boldsymbol{s} \boldsymbol{r}_n} \right)$$

31.309 29.489 26.044
Atomic coordinates
(position)

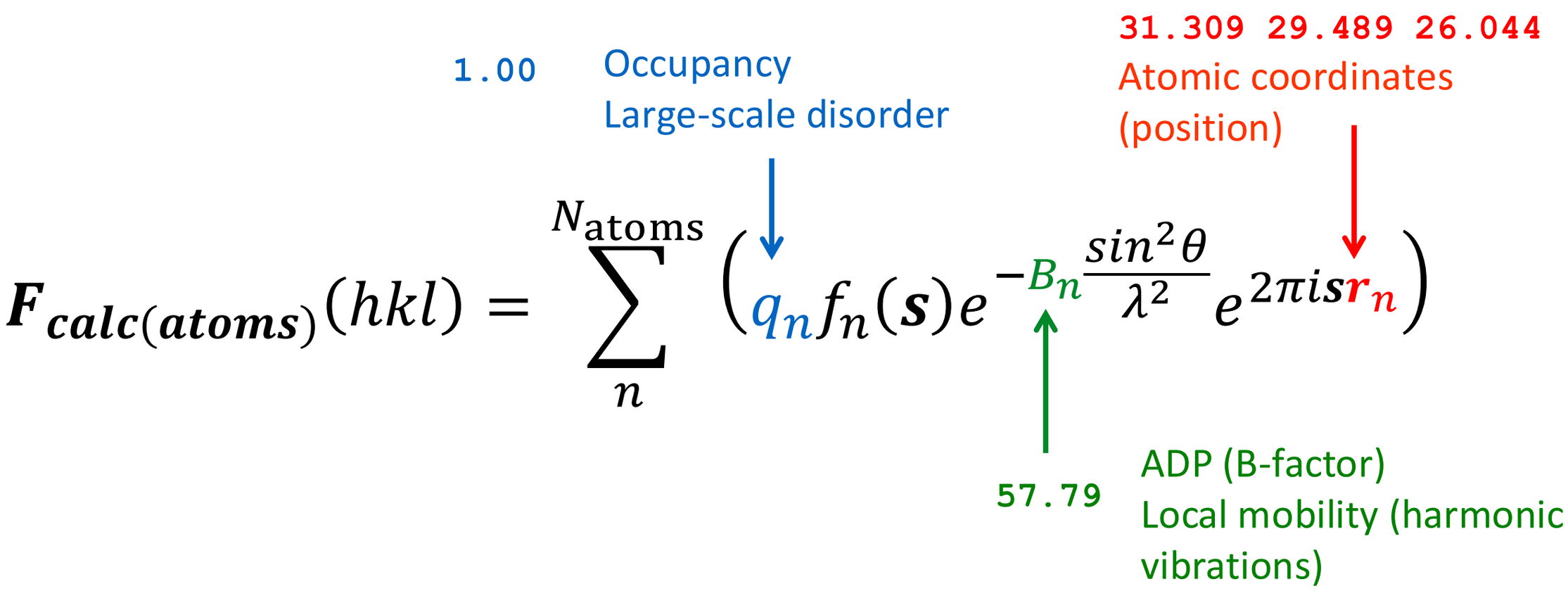
57.79
ADP (B-factor)
Local mobility (harmonic vibrations)

Atomic model parameters

To calculate a score (compare measured and model-based structure factor amplitudes), we need to compute structure factors from model parameters.

Ca atom in a Pro residue:

```
ATOM     25  CA  PRO A   4      31.309  29.489  26.044  1.00 57.79           C
```

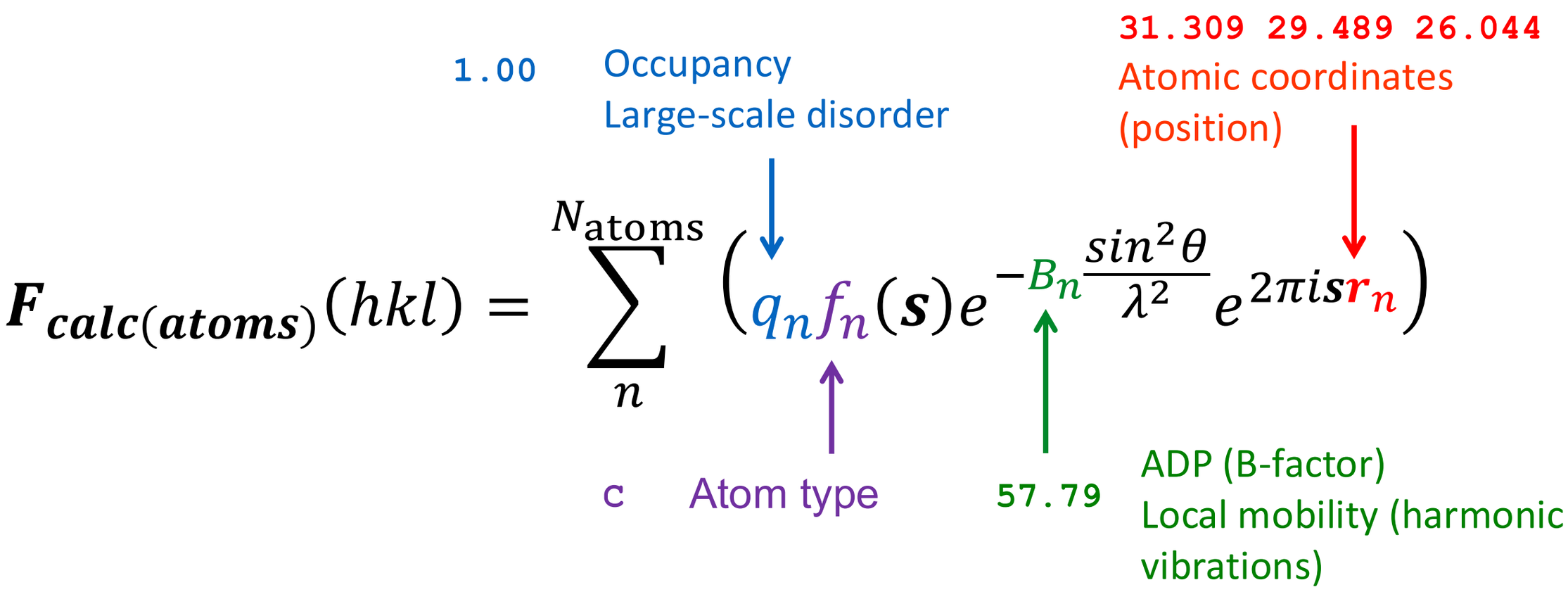


Atomic model parameters

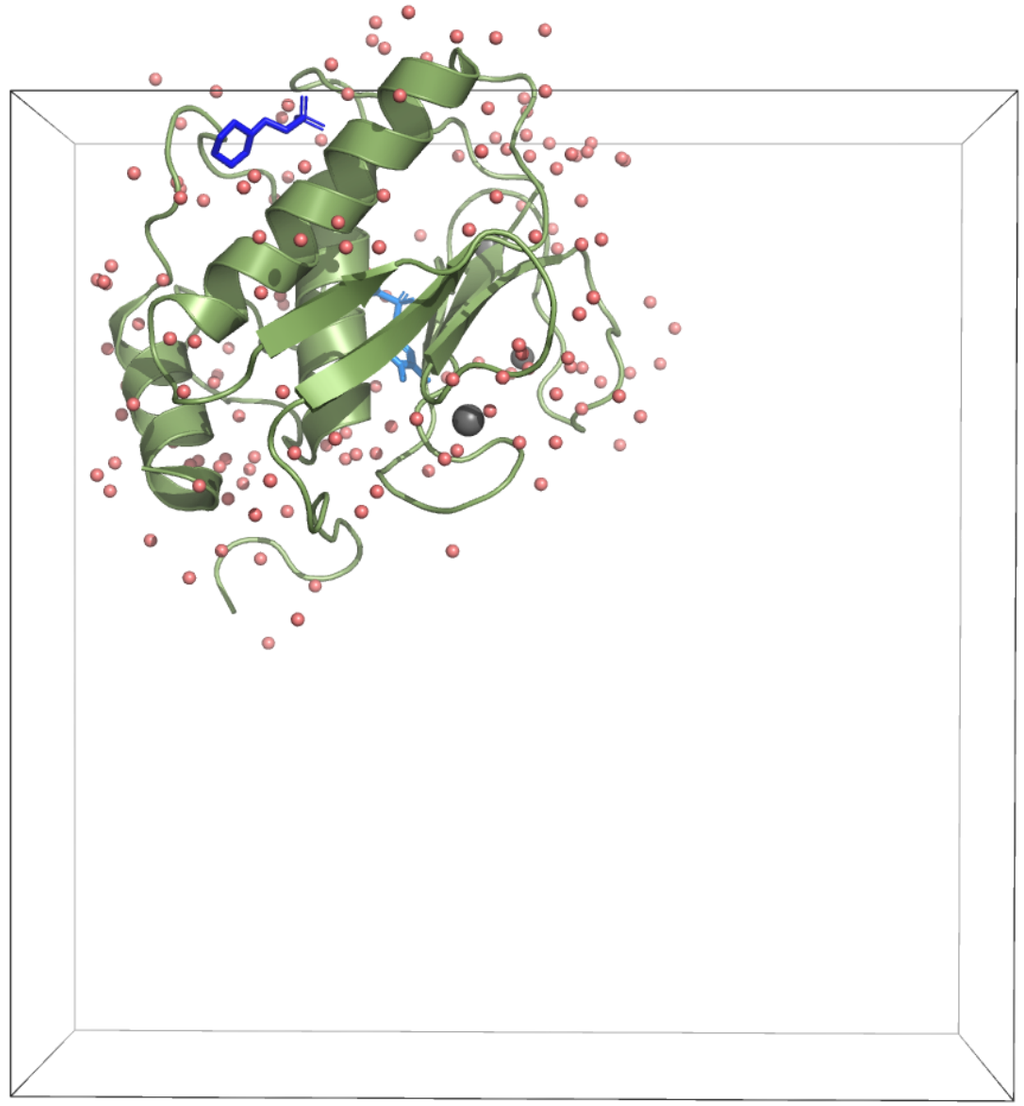
To calculate a score (compare measured and model-based structure factor amplitudes), we need to compute structure factors from model parameters.

Ca atom in a Pro residue:

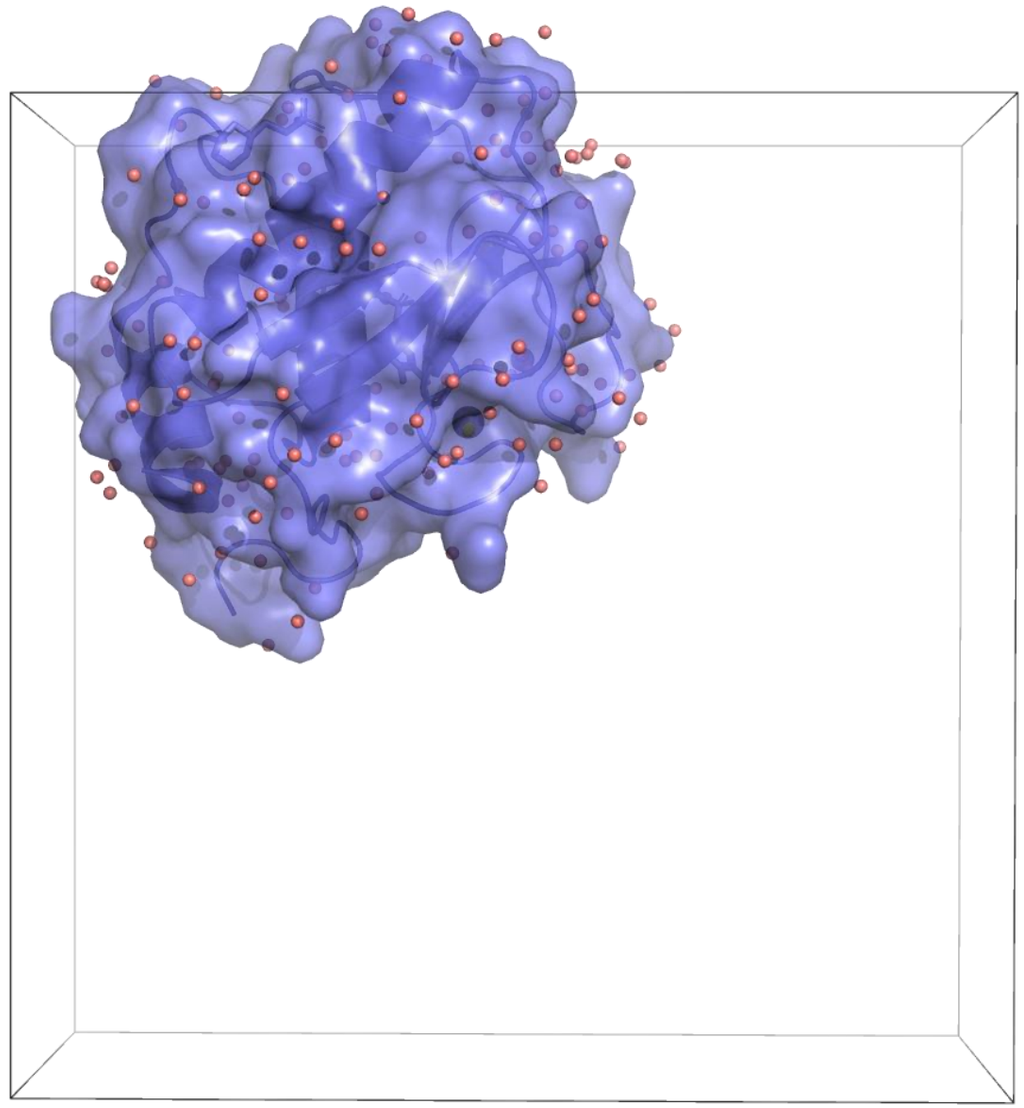
```
ATOM     25  CA  PRO A   4      31.309  29.489  26.044  1.00 57.79           C
```



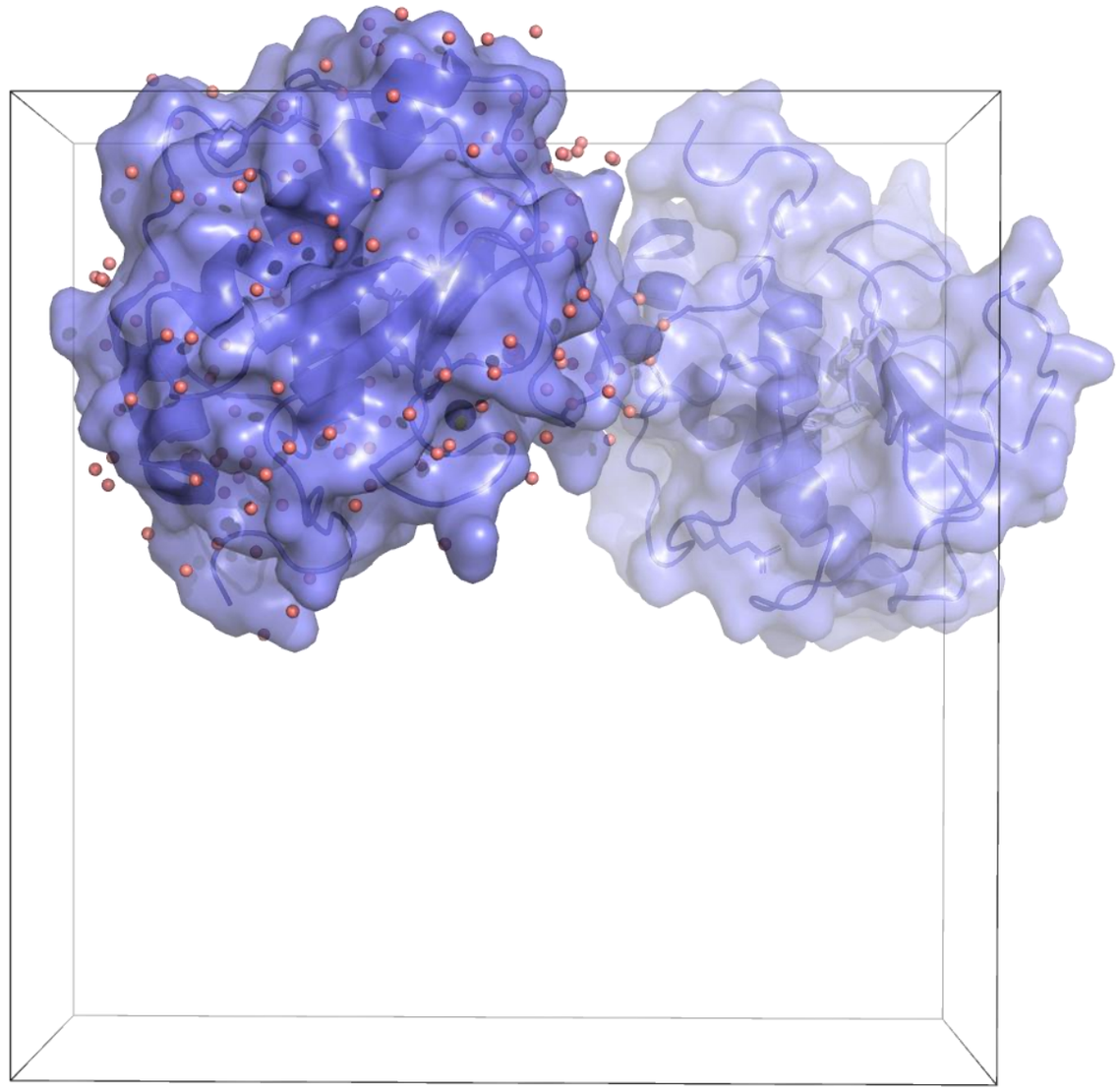
bulk solvent : non-atomic model parameter



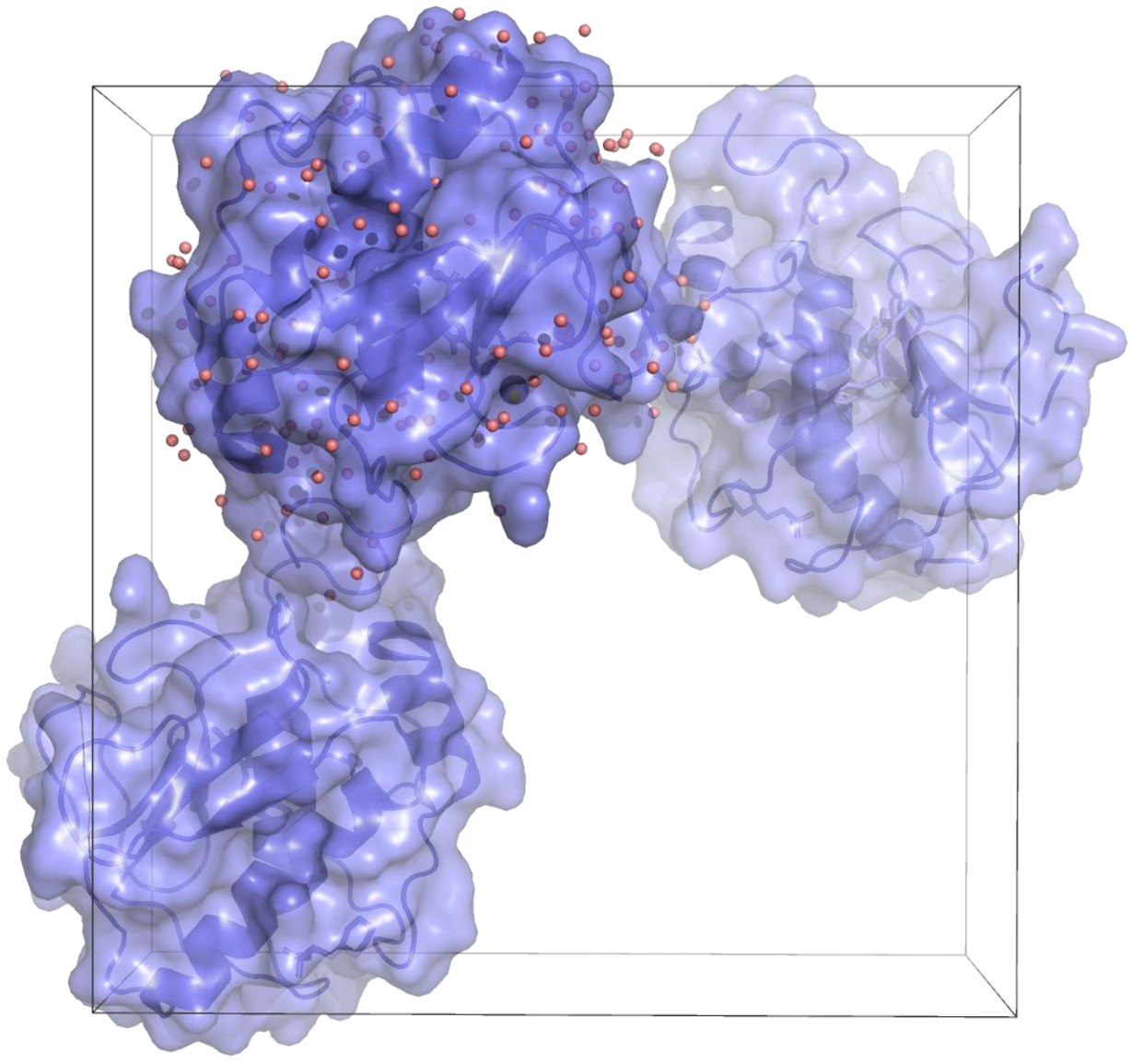
bulk solvent : non-atomic model parameter



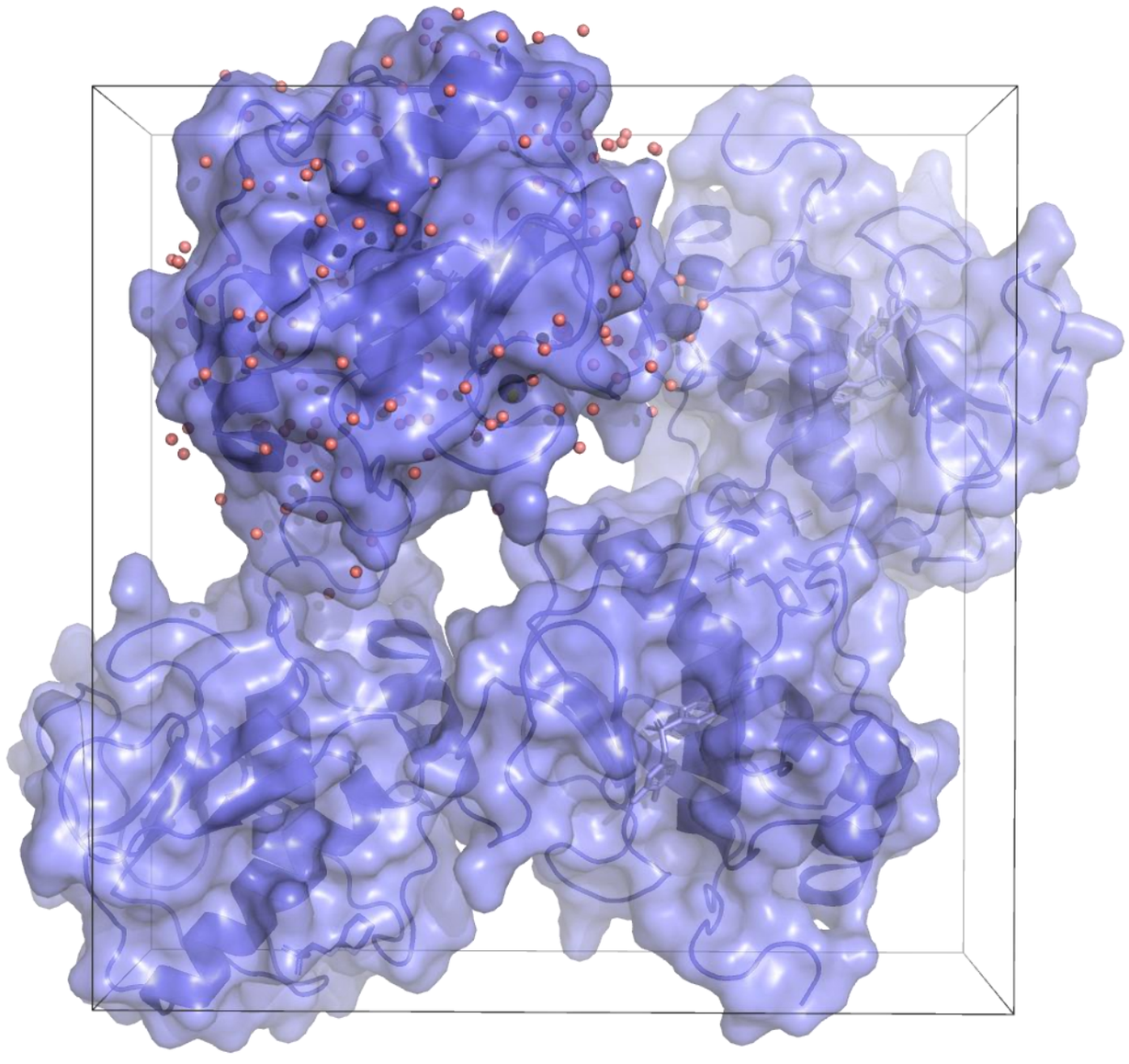
bulk solvent : non-atomic model parameter



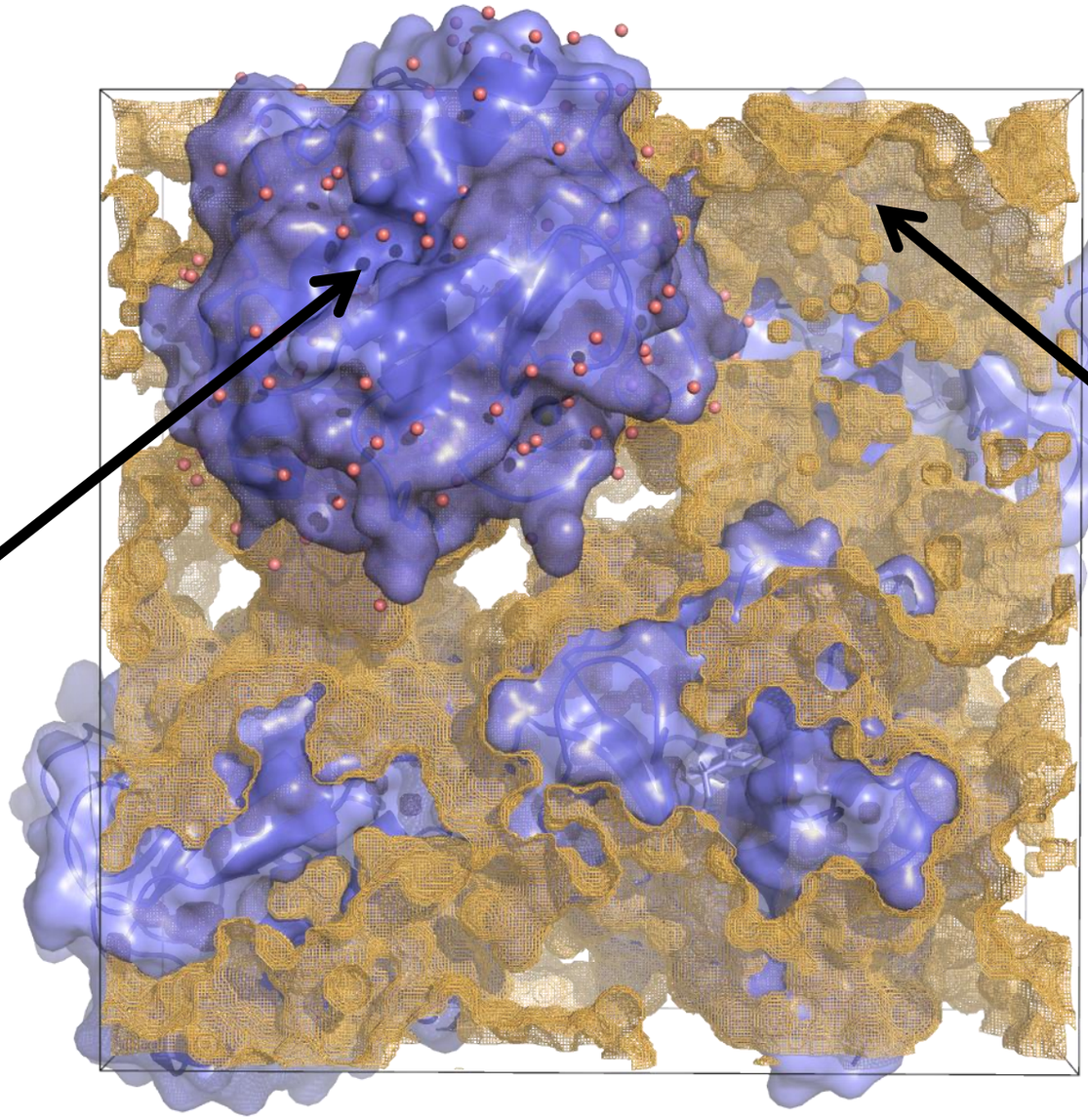
bulk solvent : non-atomic model parameter



bulk solvent : non-atomic model parameter



bulk solvent : non-atomic model parameter



Atomic model

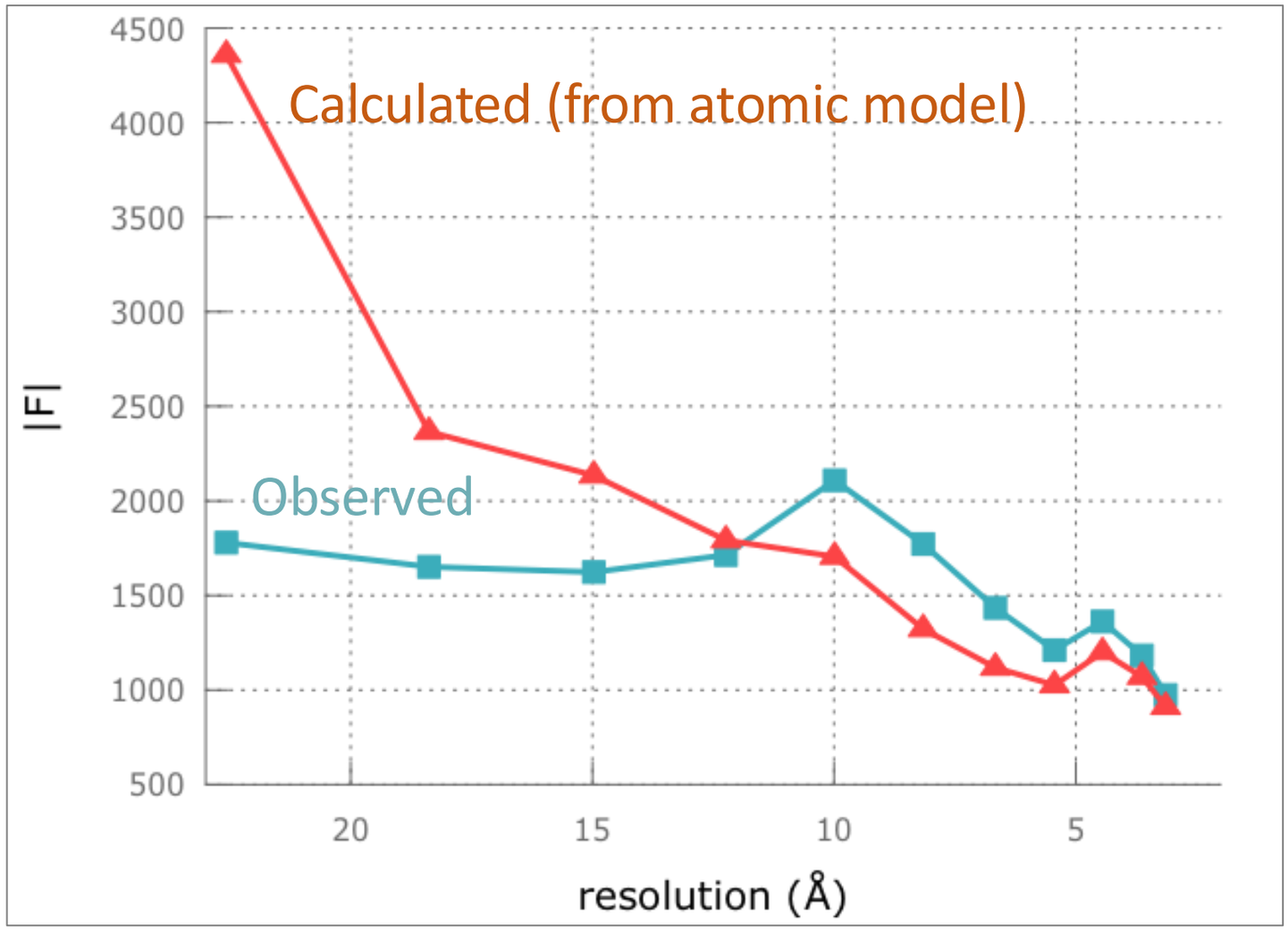
Bulk solvent

Constant density
~0.2-0.6 e/Å^3

Account for the bulk solvent in calculated structure factors:

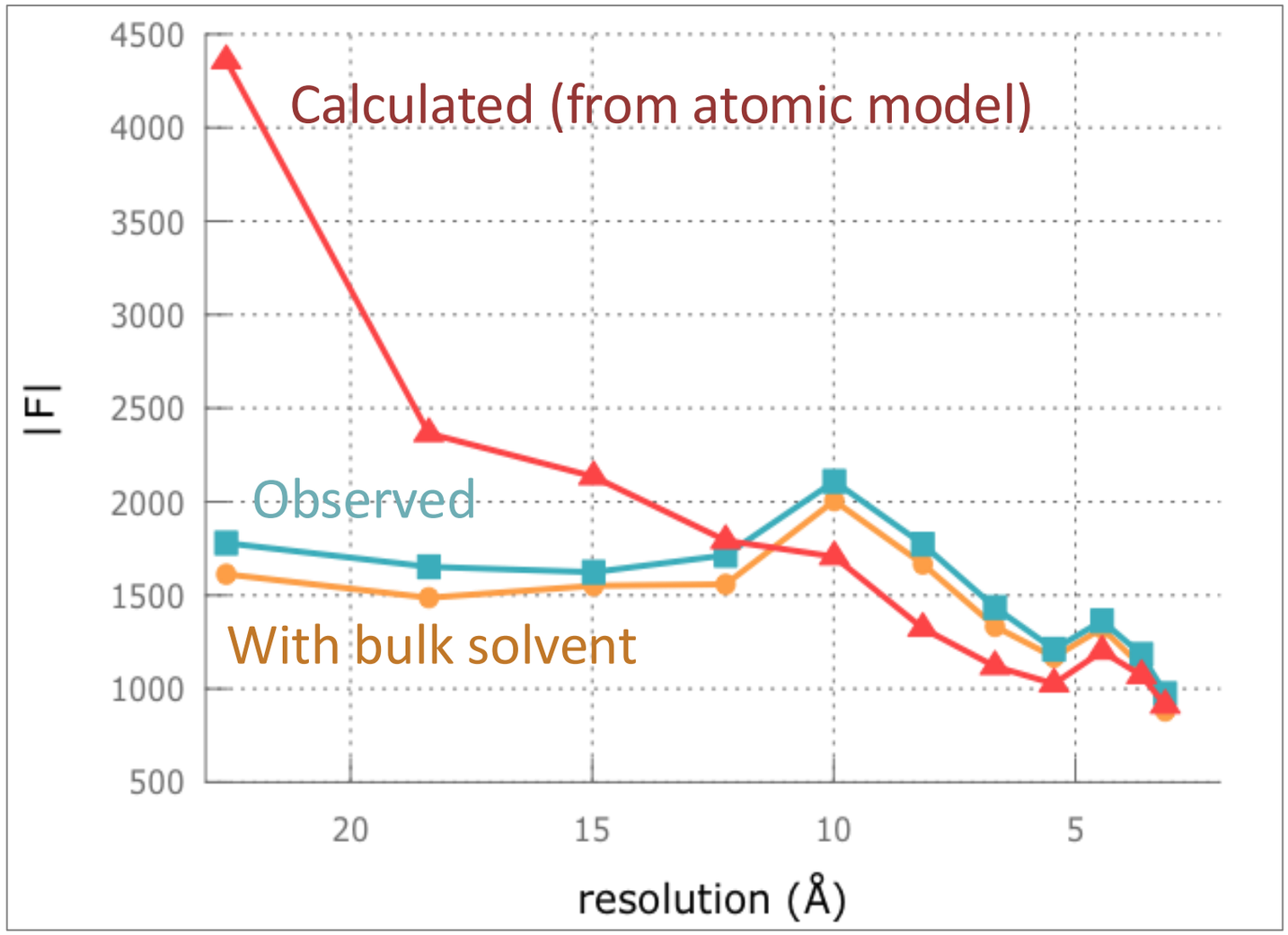
$$\mathbf{F}_{\text{model}} = k_{\text{overall}} \left(\mathbf{F}_{\text{calc(atoms)}} + \mathbf{F}_{\text{bulk solvent}} \right)$$

Why model the bulk solvent?



Bulk solvent significantly affects structure factor amplitudes at low resolution.

Why model the bulk solvent?

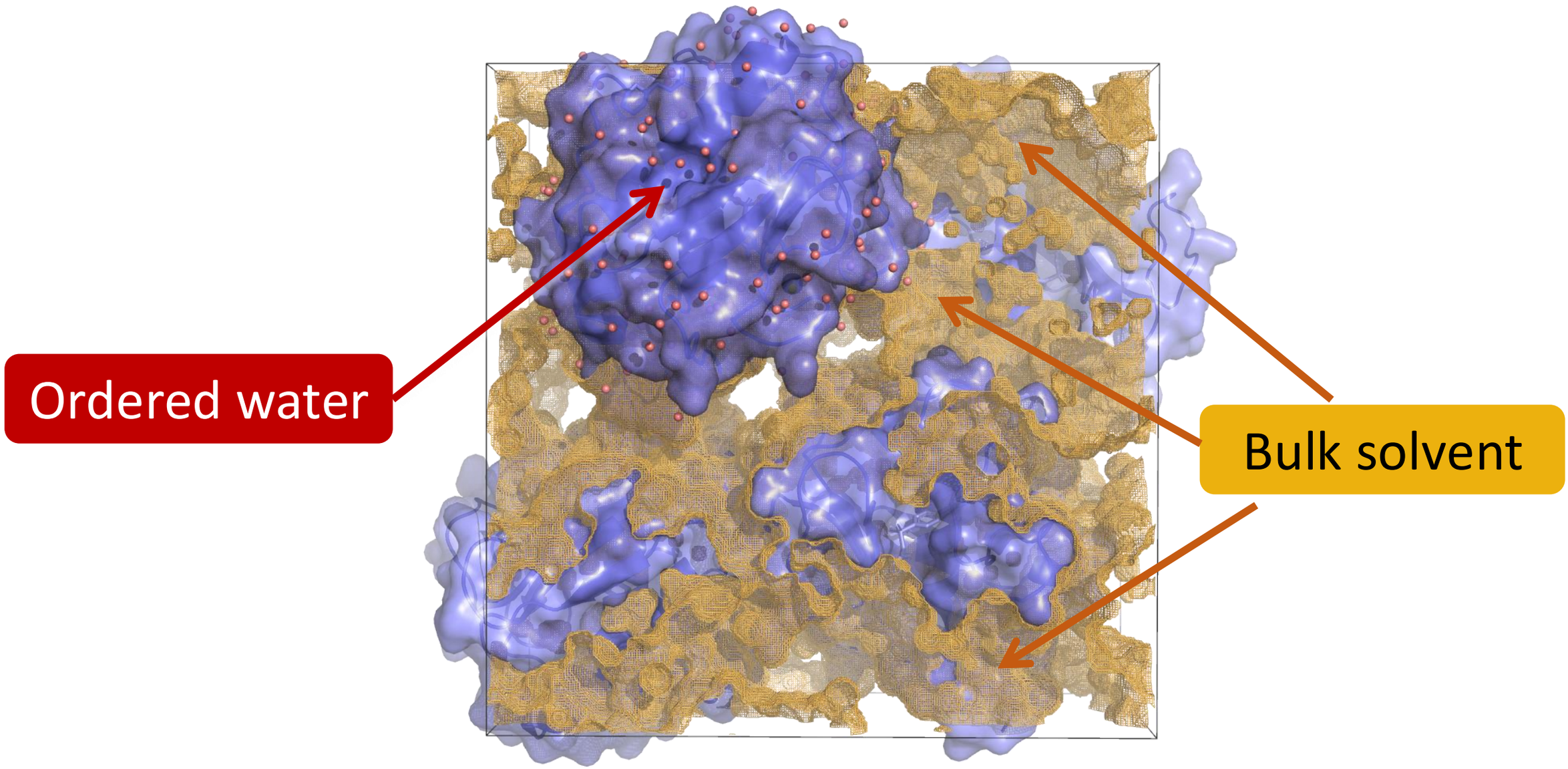


Without bulk solvent $R_{work} = 0.27$ $R_{free} = 0.31$

Using bulk solvent $R_{work} = 0.21$ $R_{free} = 0.24$

Bulk solvent vs ordered solvent

Bulk solvent is not the same as ordered solvent.



Refinement

- *parameters* of a model
- *target function*
- *optimization* algorithm
- *observations*

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Introduction to macromolecular refinement

Dale. E. Tronrud

Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

Received 5 April 2004
Accepted 21 September 2004

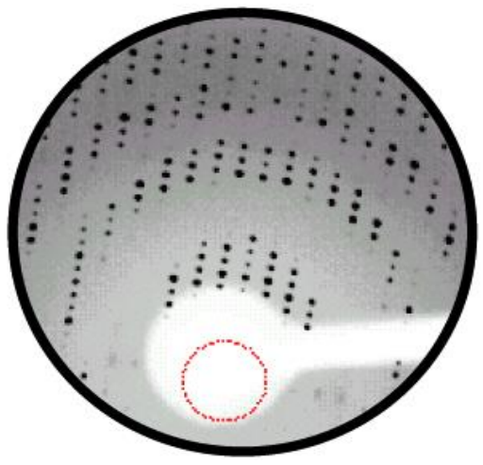
The process of refinement is such a large problem in function minimization that even the computers of today cannot perform the calculations to properly fit X-ray diffraction data. Each of the refinement packages currently under development reduces the difficulty of this problem by utilizing a unique combination of targets, assumptions and optimization

Tronrud, D. E. Introduction to macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. 60, 2156–2168 (2004).

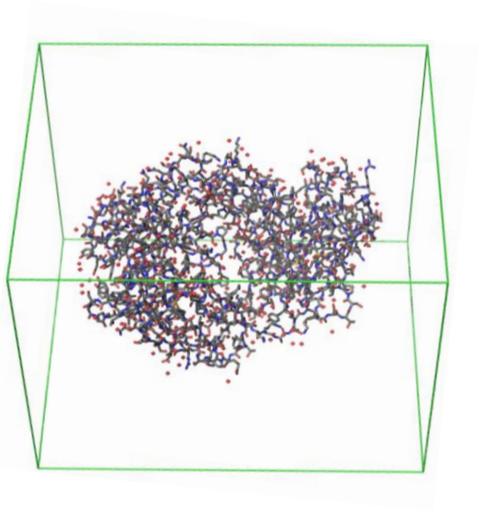
Target function

Score the model against the experimental data,. i.e. compare model-based and measured structure factor amplitudes.

Experimental data

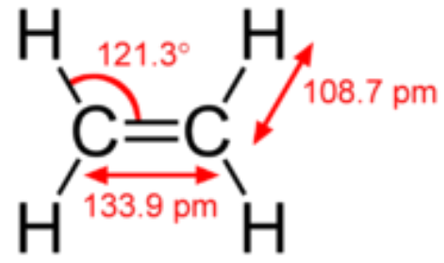


Model



A priori knowledge

Example:
Covalent geometry



$$T = T_{Data}(F_{obs}, F_{Model}) + wT_{Restraints}$$

Target function

Score the model against the experimental data,. i.e. compare model-based and measured structure factor amplitudes.

1. Least squares

$$T_{Data} = w_{xray} \sum_{hkl} \frac{1}{\sigma^2} \left(|F_{obs}| - |F_{\text{model}}|\right)^2$$

σ = Standard deviation

- Large difference between observed and calculated structure factor
  → The model is inaccurate.

- Small standard deviation σ
  → That observation will be important in the sum.

Target function

Score the model against the experimental data,. i.e. compare model-based and measured structure factor amplitudes.

1. Least squares

$$T_{Data} = w_{xray} \sum_{hkl} \frac{1}{\sigma^2} \left(|F_{obs}| - |F_{\mathrm{model}}|\right)^2$$

Assumptions of this approach:

- Errors obey a Gaussian distribution.

It is impossible for any set of parameters of an imperfect model (e.g. missing domain) to reproduce all the observed structure factors.

Target function

Score the model against the experimental data,. i.e. compare model-based and measured structure factor amplitudes.

2. Maximum Likelihood (ML)

“Maximize the probability (likelihood) that the observed data would be produced given the current model.”

Computing this exactly requires **fully modeling all sources of error**, but we cannot model all errors in complete generality.

→ Make assumptions about the nature of the uncertainties in the observations and the model parameters.

Method of choice for macromolecular crystallography.

Refinement

- *parameters* of a model
- *target function*
- *optimization* algorithm
- *observations*

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Introduction to macromolecular refinement

Dale. E. Tronrud

Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

Received 5 April 2004
Accepted 21 September 2004

The process of refinement is such a large problem in function minimization that even the computers of today cannot perform the calculations to properly fit X-ray diffraction data. Each of the refinement packages currently under development reduces the difficulty of this problem by utilizing a unique combination of targets, assumptions and optimization

Tronrud, D. E. Introduction to macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. 60, 2156–2168 (2004).

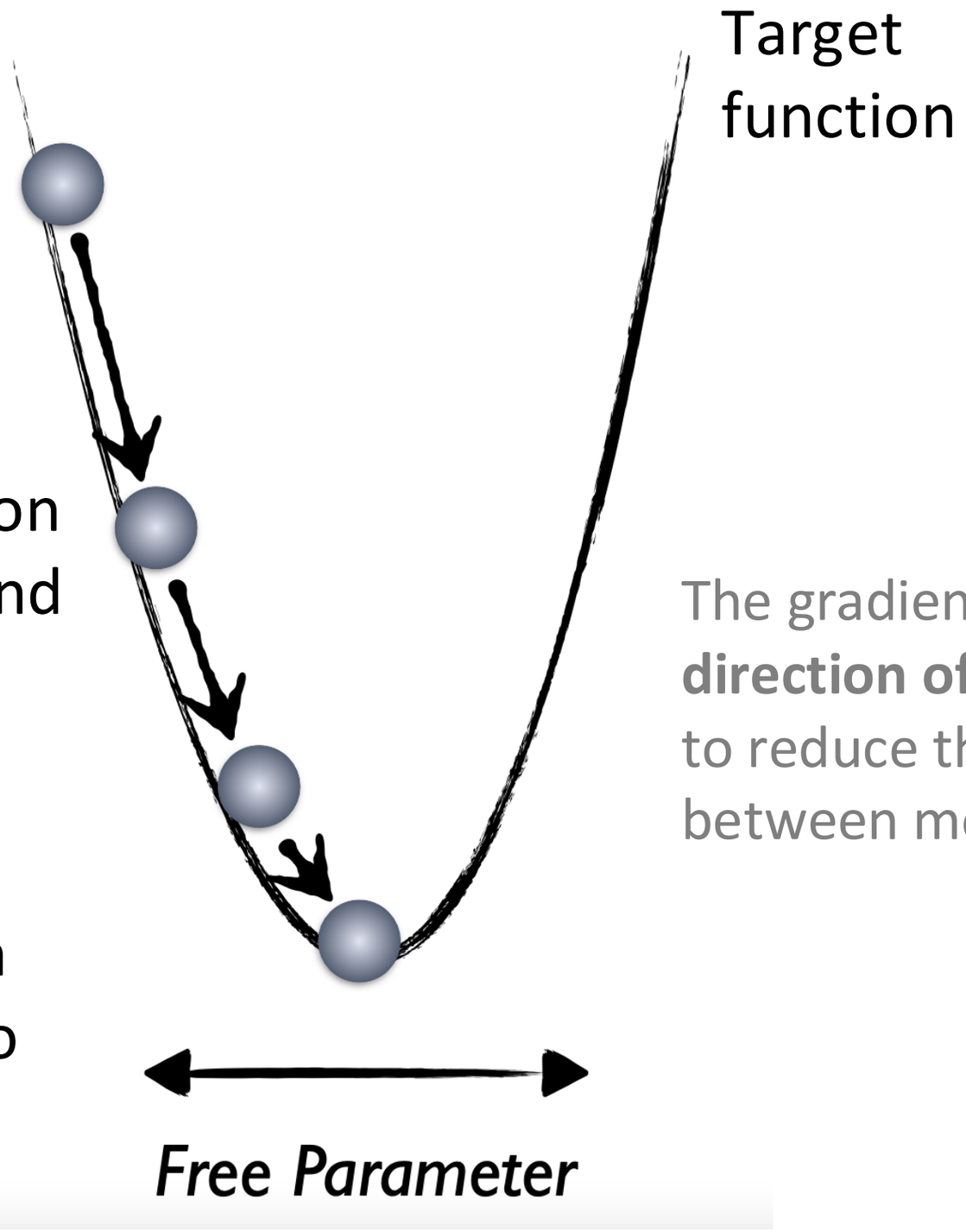
Optimization algorithm

The purpose of the optimization algorithm is to minimize the target function.

Starting model

- Evaluate target function
- Compute gradients (and second derivatives)

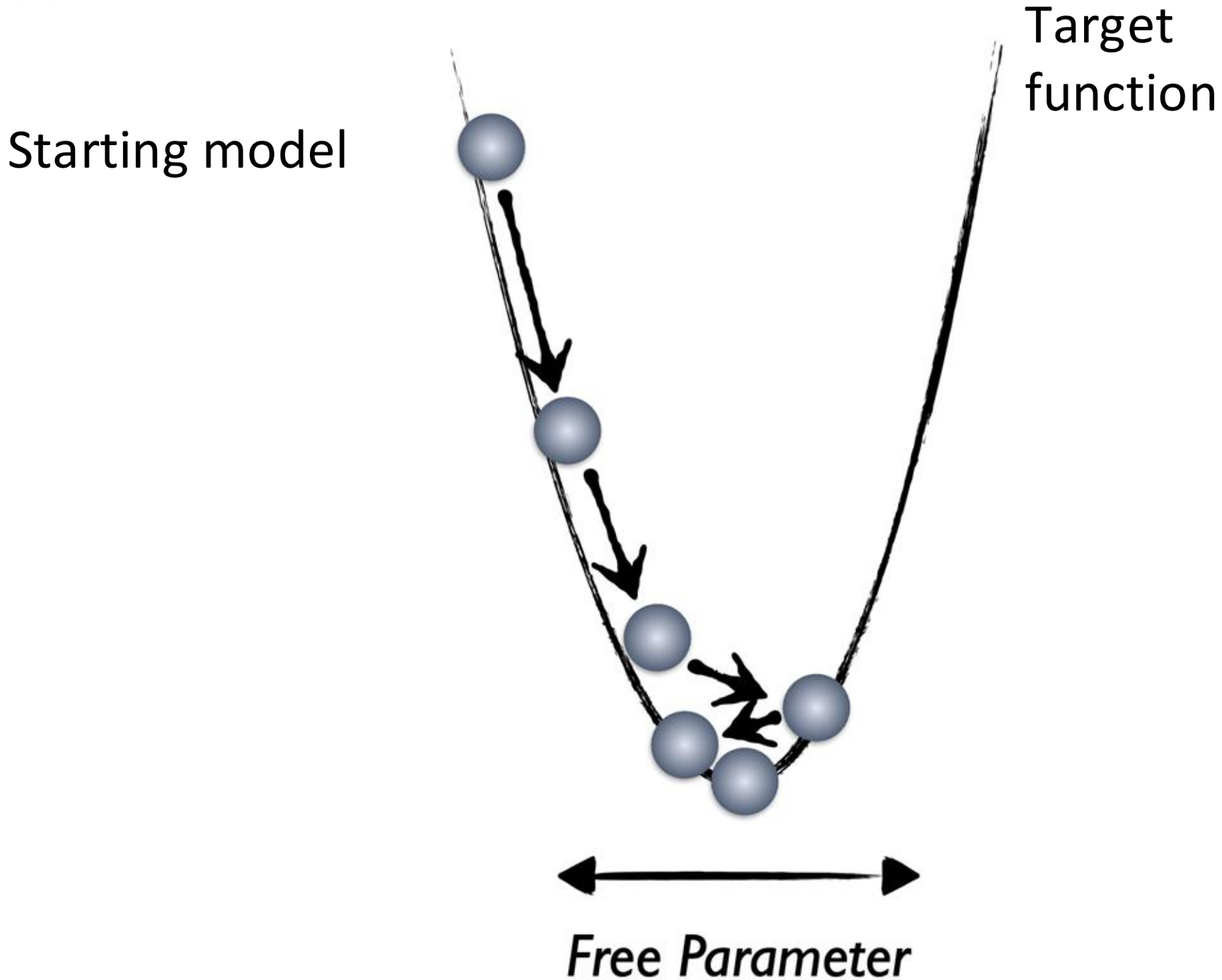
Optimization algorithm calculates how much to move each parameter.



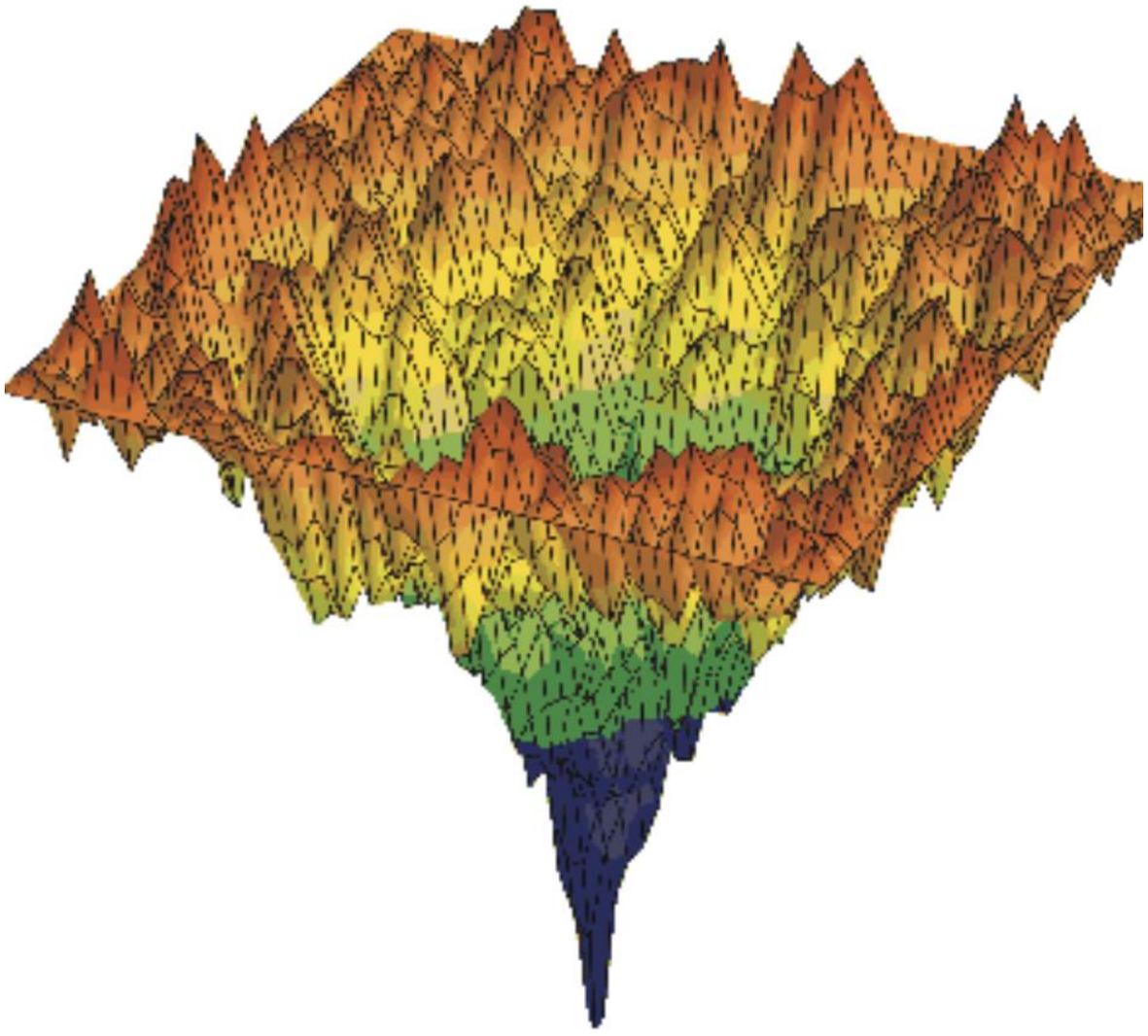
The gradients show the **direction of change** needed to reduce the disagreement between model and data.

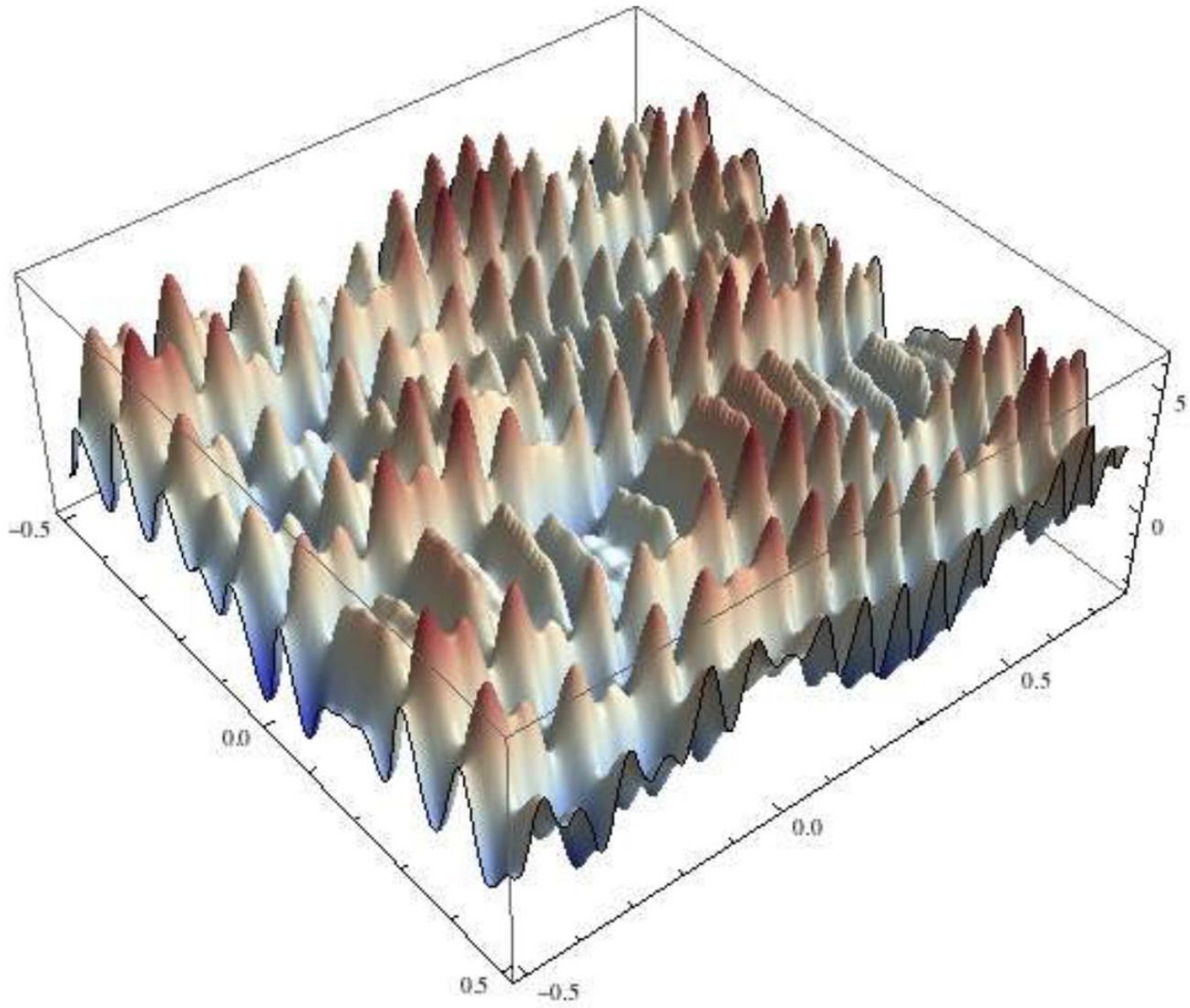
Optimization algorithm

The purpose of the optimization algorithm is to minimize the target function.



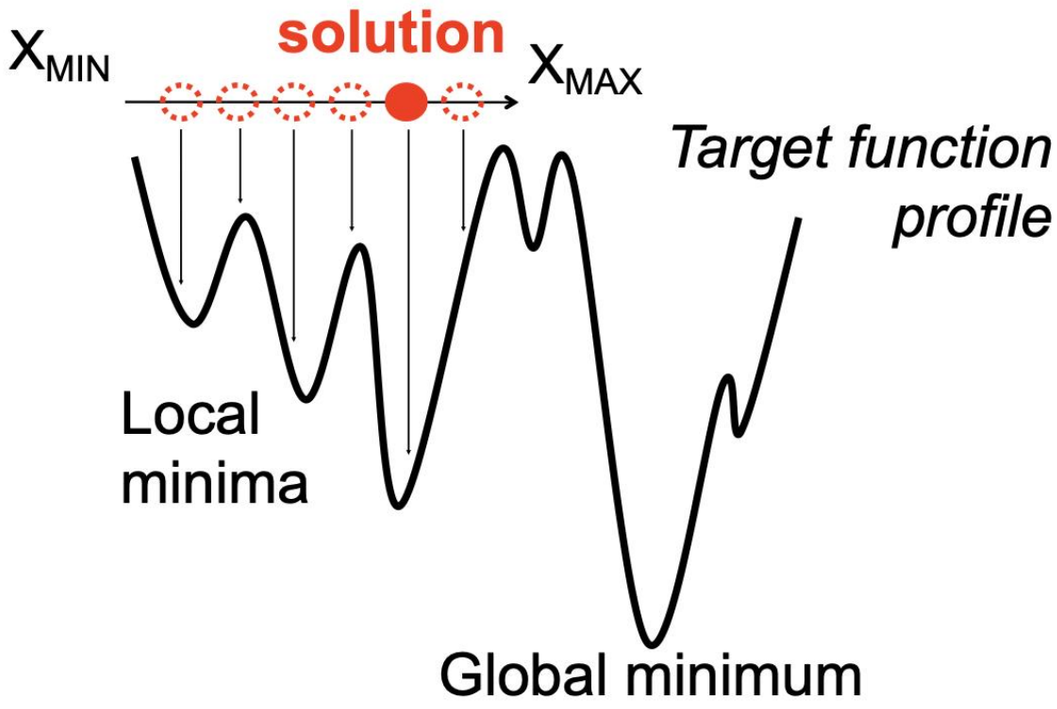
The target function is complex

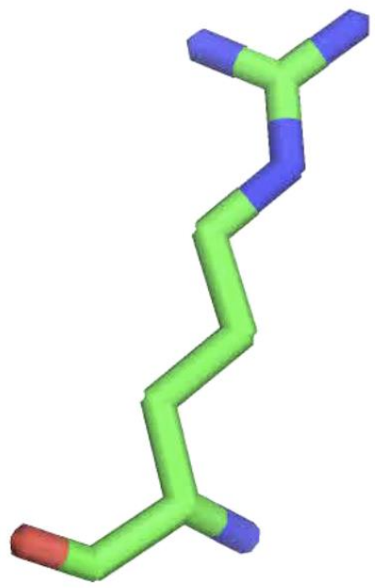


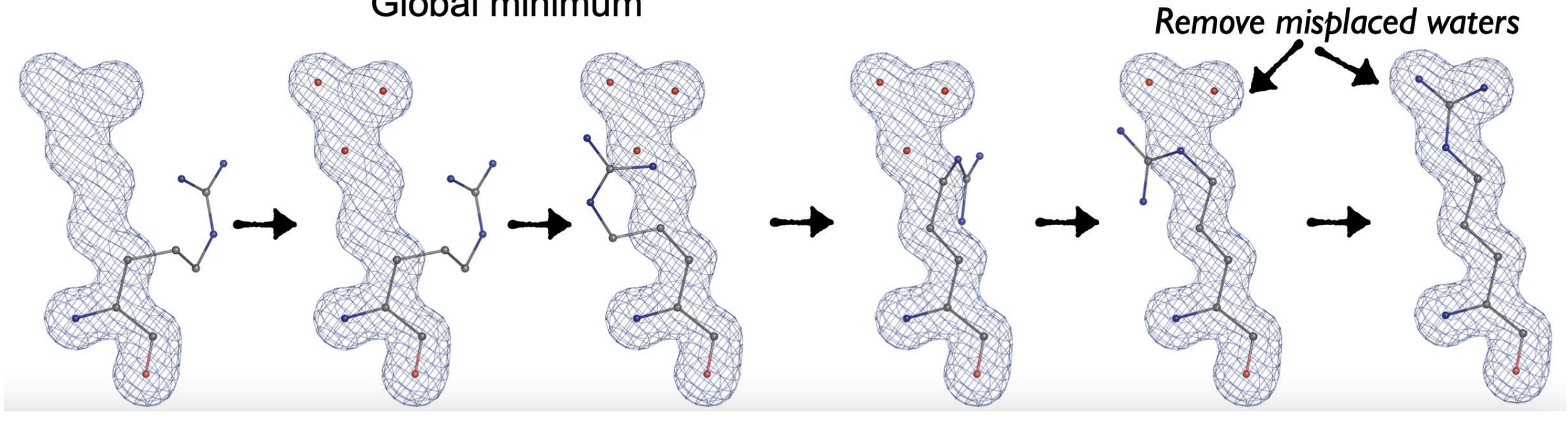


From Robb Seaton's "The Ultimate Guide to Simulated Annealing"

Systematic searching







Simulated Annealing

Physical process of **annealing** (metallurgy):
Heat a material and then slowly cool it so its atoms settle into a low-energy, stable structure.

In crystallographic refinement:
Introduce **controlled random changes** to model parameters (xyz, torsions). Introduce “heat” via a molecular dynamics simulation.

When to use:

- Poorly built structures early in refinement.
- Particular need to remove bias (changing Rfree).

Refinement

- *parameters* of a model
- *target function*
- *optimization* algorithm
- *observations*

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Introduction to macromolecular refinement

Dale. E. Tronrud

Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

Received 5 April 2004
Accepted 21 September 2004

The process of refinement is such a large problem in function minimization that even the computers of today cannot perform the calculations to properly fit X-ray diffraction data. Each of the refinement packages currently under development reduces the difficulty of this problem by utilizing a unique combination of targets, assumptions and optimization

Tronrud, D. E. Introduction to macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. 60, 2156–2168 (2004).

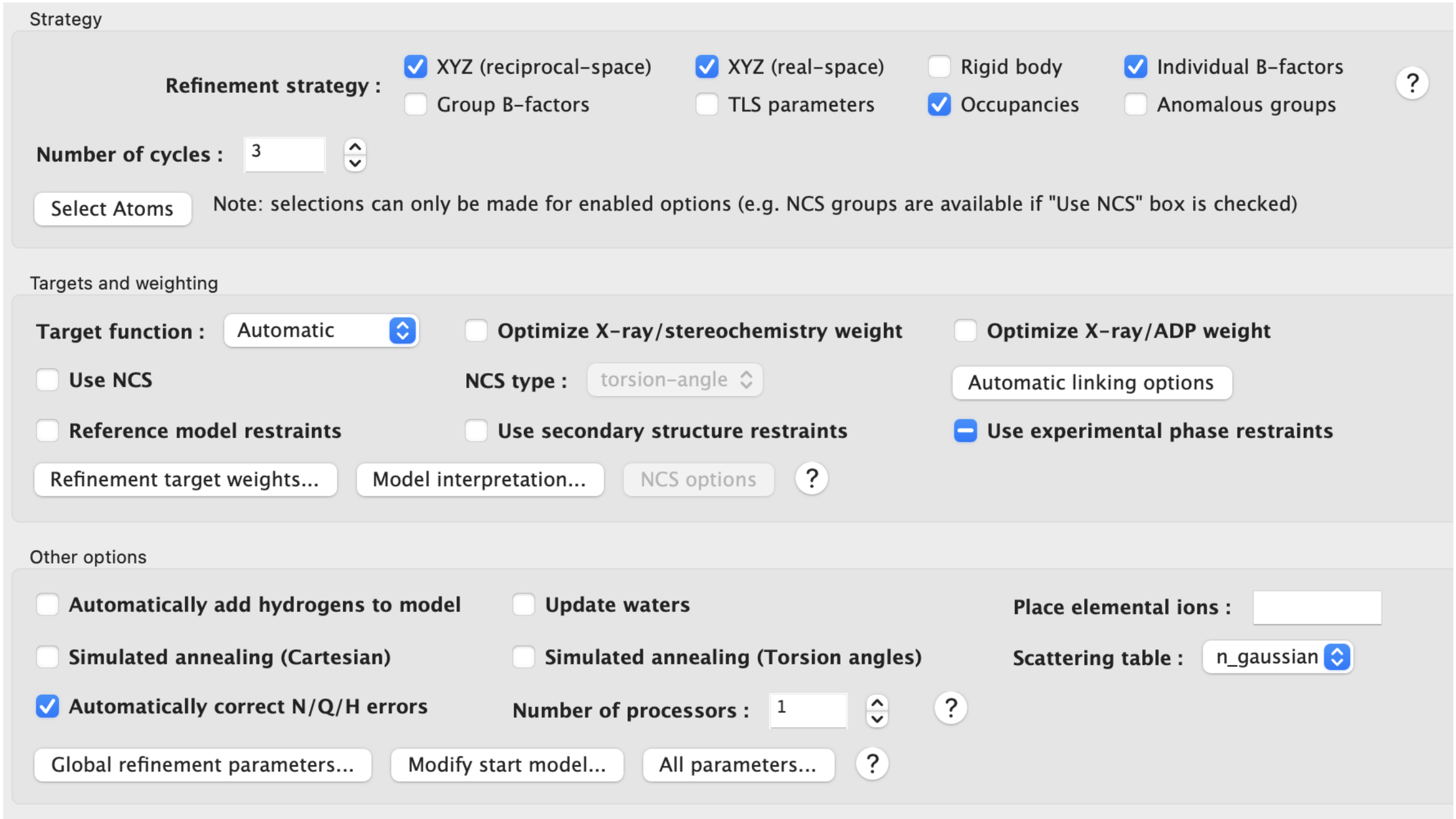
Observations

Observations = Everything known about the crystal

- structure-factor amplitudes
- unit-cell parameters
- standardized stereochemistry
- experimentally determined phase information

Restraints will add observations.

What to change in refinement?



Which options shall I use? Which parameters shall I change?

phenix.refine has >1k parameters.

What to change in refinement?

- *parameters* of a model
- *target function*
- *optimization* algorithm

} Set by refinement program

- *observations*

What to change in refinement?

- *parameters* of a model ←
- *target function*
- *optimization* algorithm
- *observations* ←

In practice, you'll change the model parameterization and modify the a priori information (observations) via restraints.

There is no recipe for what to change. Needs to be adapted to each case.

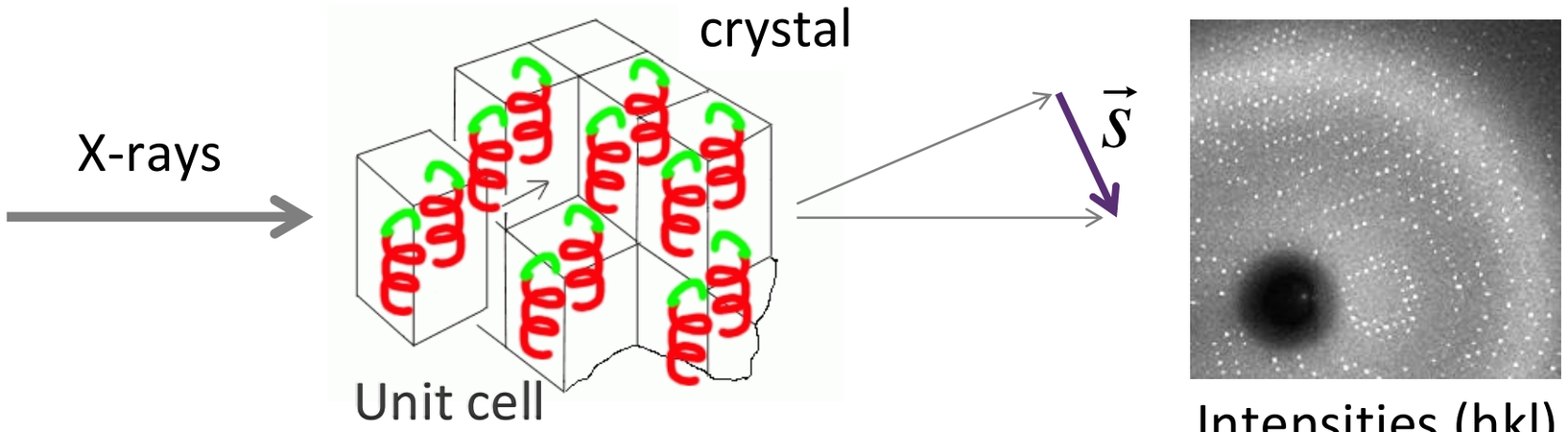
Consider refinement like an experiment. You try and analyze the result.

What to change in refinement?

- *parameters* of a model ←
- *target function*
- *optimization* algorithm
- *observations* ←

Some examples of model parameterization that can be changed.

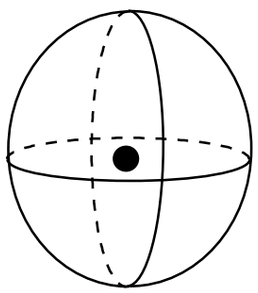
Atomic displacements



Average over time and space

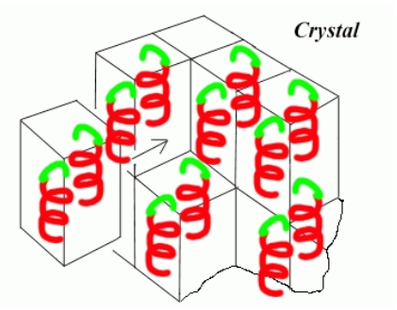
Time:

Atoms are in thermal motions around mean positions.



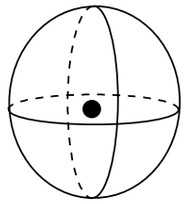
Space:

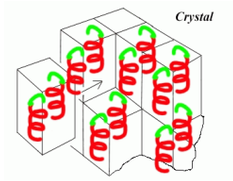
Small differences between unit cells.

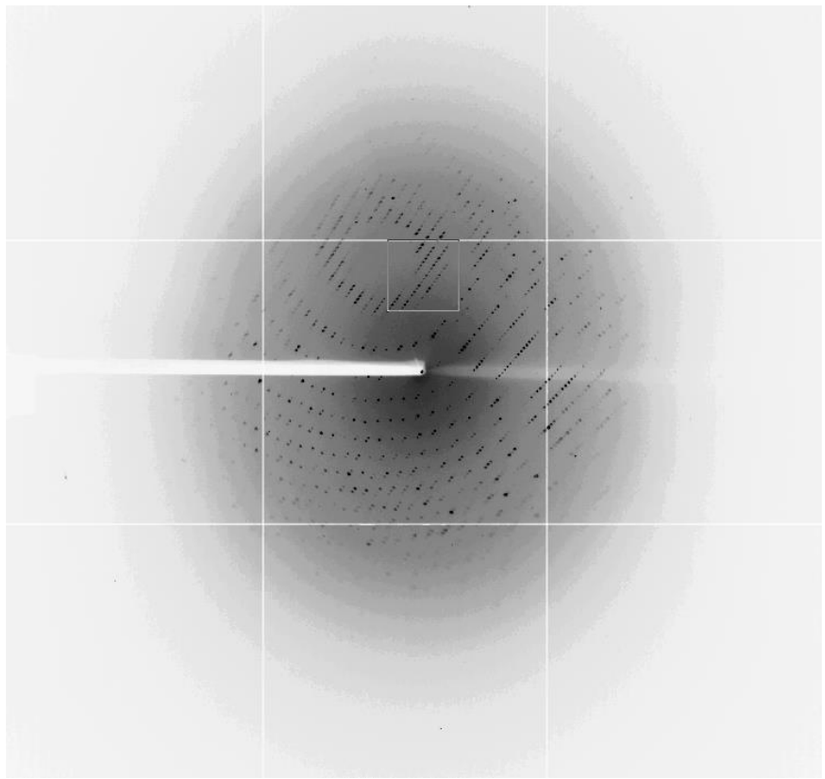


Atomic displacements

Displacements of atoms in the sample

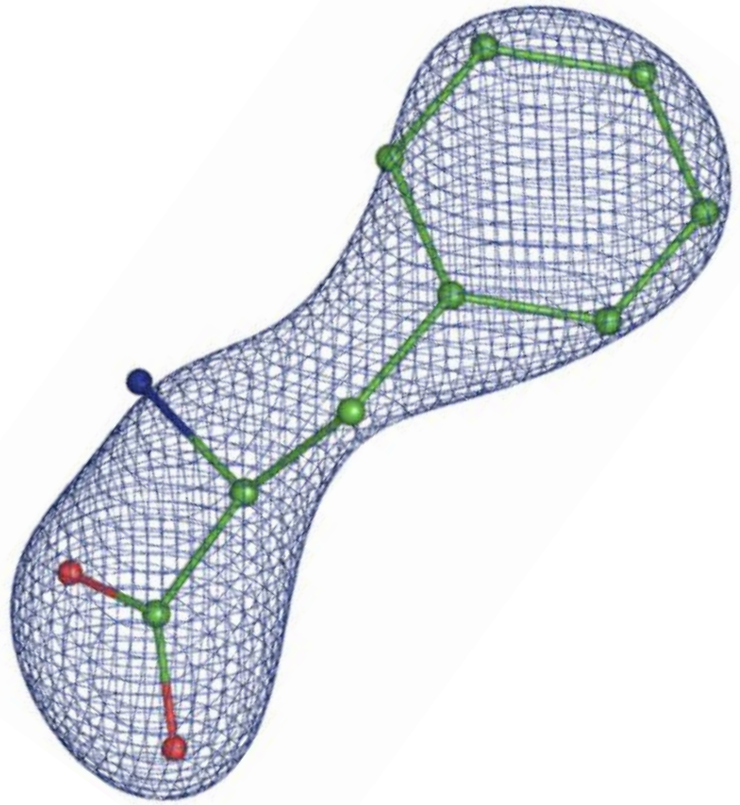






Reciprocal space:

High-resolution data vanish



Real space:

Density is blurred

Displacement need to be modelled

Atomic displacements

Superposition of several contributions:

$$U_{total} = U_{cryst} + U_{group} + U_{local}$$

Atomic displacements

Superposition of several contributions:

$$U_{total} = U_{cryst} + U_{group} + U_{local}$$

U_{cryst} = displacement of the crystal as a whole

Atomic displacements

Superposition of several contributions:

$$U_{total} = U_{cryst} + U_{group} + U_{local}$$

U_{cryst} = displacement of the crystal as a whole

U_{group} = concerted motions of multiple atoms (group motions)

Atomic displacements

Superposition of several contributions:

$$U_{total} = U_{cryst} + U_{group} + U_{local}$$

U_{cryst} = displacement of the crystal as a whole

U_{group} = concerted motions of multiple atoms (group motions)

U_{local} = small local atomic vibrations

Atomic displacements

Superposition of several contributions:

$$U_{total} = U_{cryst} + U_{group} + U_{local}$$

U_{cryst} = displacement of the crystal as a whole

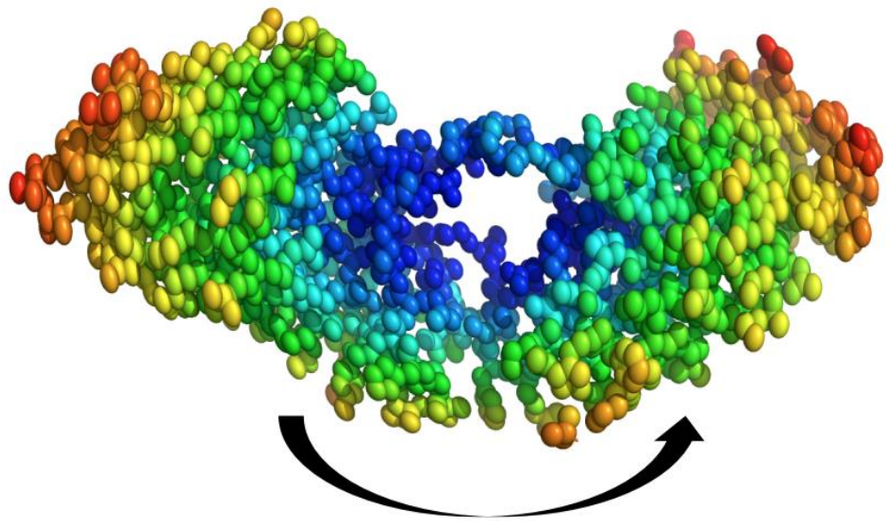
Done automatically

U_{group} = concerted motions of multiple atoms (group motions)

U_{local} = small local atomic vibrations

Atomic displacements: U_{group}

- TLS: rigid body displacements of molecules, domains, secondary structure elements



Displacement of a rigid body can be described by

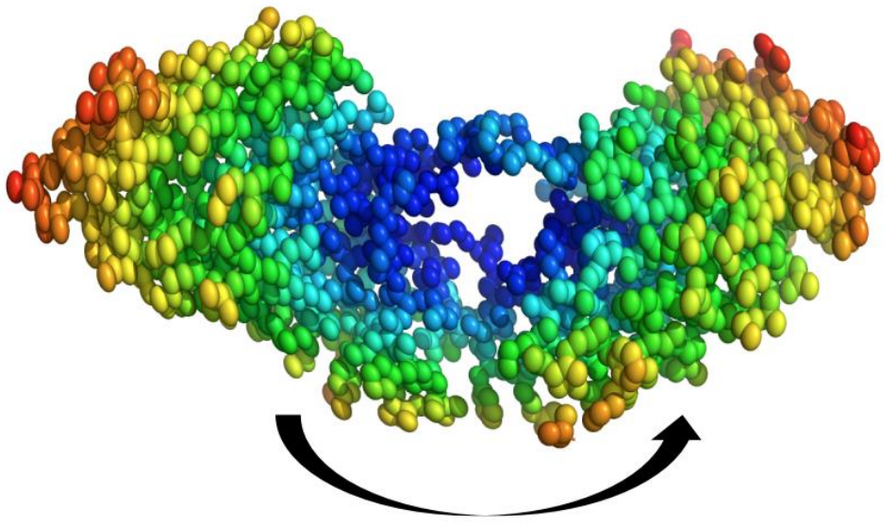
- Translation
- Rotation (Libration)
- Correlation between them (Screw)

➡ Partition the model into TLS groups

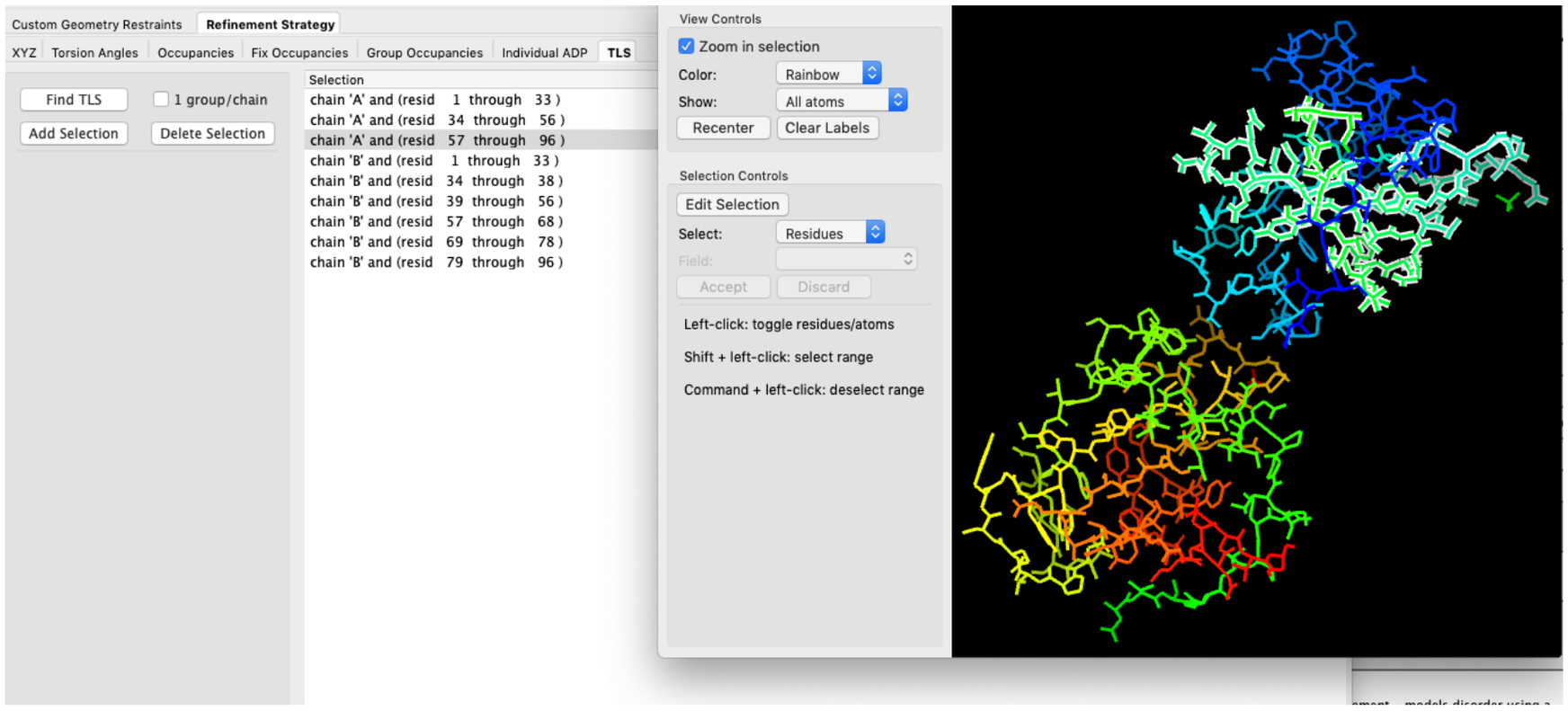
In phenix.refine, partinioning needs refined B-factors, so don’t turn on right after MR.

Atomic displacements: U_{group}

- TLS: rigid body displacements of molecules, domains, secondary structure elements

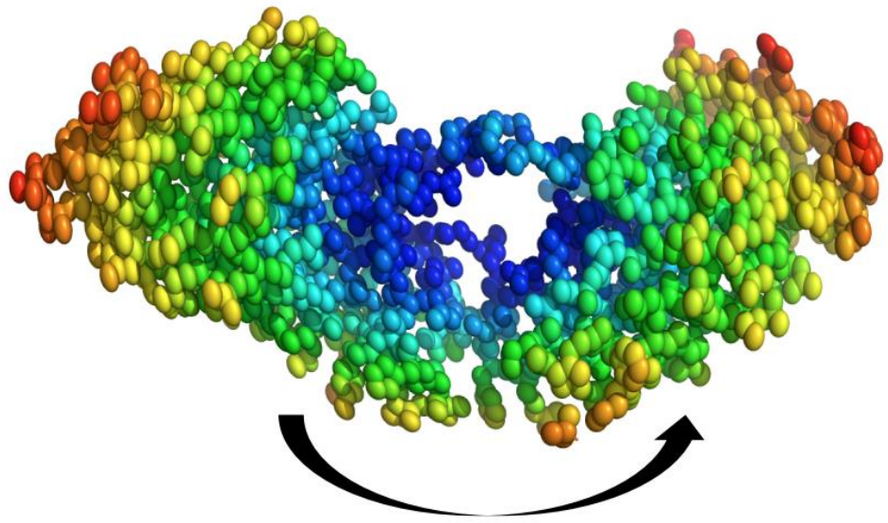


Can be done automatically in the Phenix GUI.

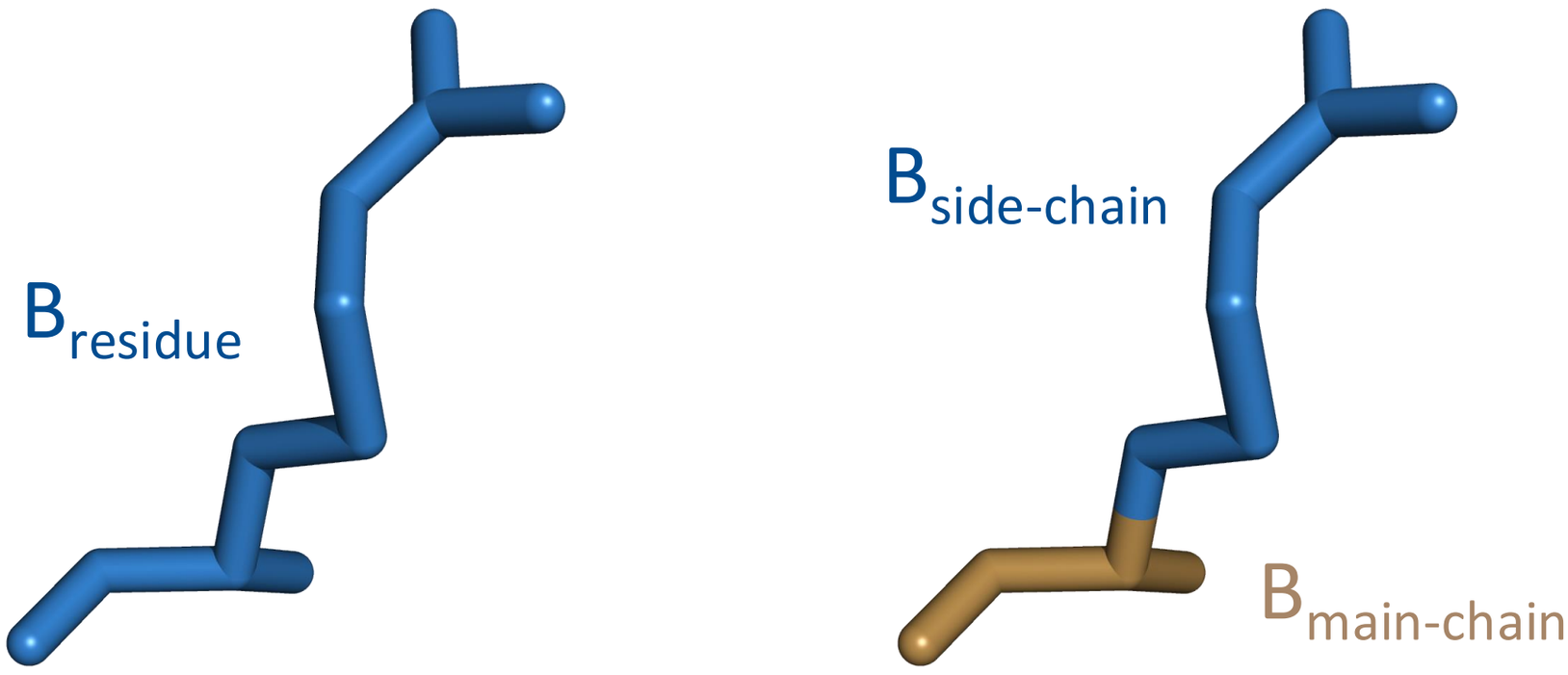


Atomic displacements: U_{group}

- TLS: rigid body displacements of molecules, domains, secondary structure elements

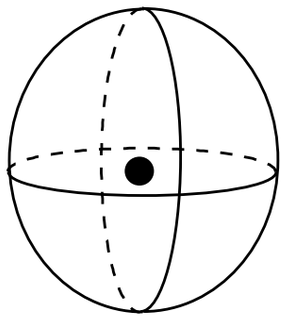


- Simple group isotropic model (one single B_{iso}), entire residue or main chain – side chain



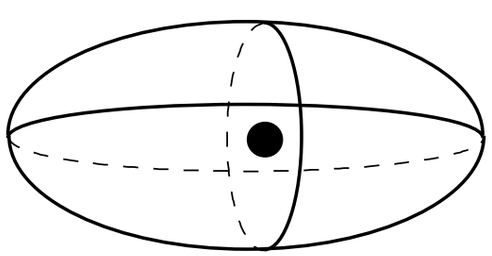
Atomic displacements: U_{local}

- Describe small local atomic vibrations
- ”per atom”
- Represent both thermal vibration, and variation in the atomic positions from one unit cell to the next



B_{iso}

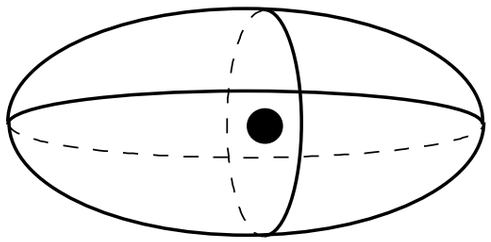
- related to the mean-square amplitude of vibration
- one isotropic displacement parameter per atom



B_{aniso}

- describe a probability distribution for the electron density with a 3d Gaussian
- 3x3 symmetric tensor (6 parameters)

Atomic displacements: U_{local}



- B_{aniso} is more “realistic”.
- But they **double** the number of parameters.

xyz+ occ + B_{iso}	xyz+ occ + B_{aniso}
3 + 1 + 1	3 + 1 + 6

- Requires more observations to be feasible (resolution).

Atomic displacements: which one to choose?

It depends…
(data resolution, data quality, data-to-parameter ratio,…)

Around 1.5Å: can try anisotropic B-factors
(all non-H atoms or only protein?)

TLS: can be applied at most resolution ranges.
(cannot do individual anistropic ADP and TLS)

Group B-factors: low resolution
(per residue or mainchain/sidechain?)

https://phenix-online.org/documentation/faqs/refine.html#b-factors-adps-tls

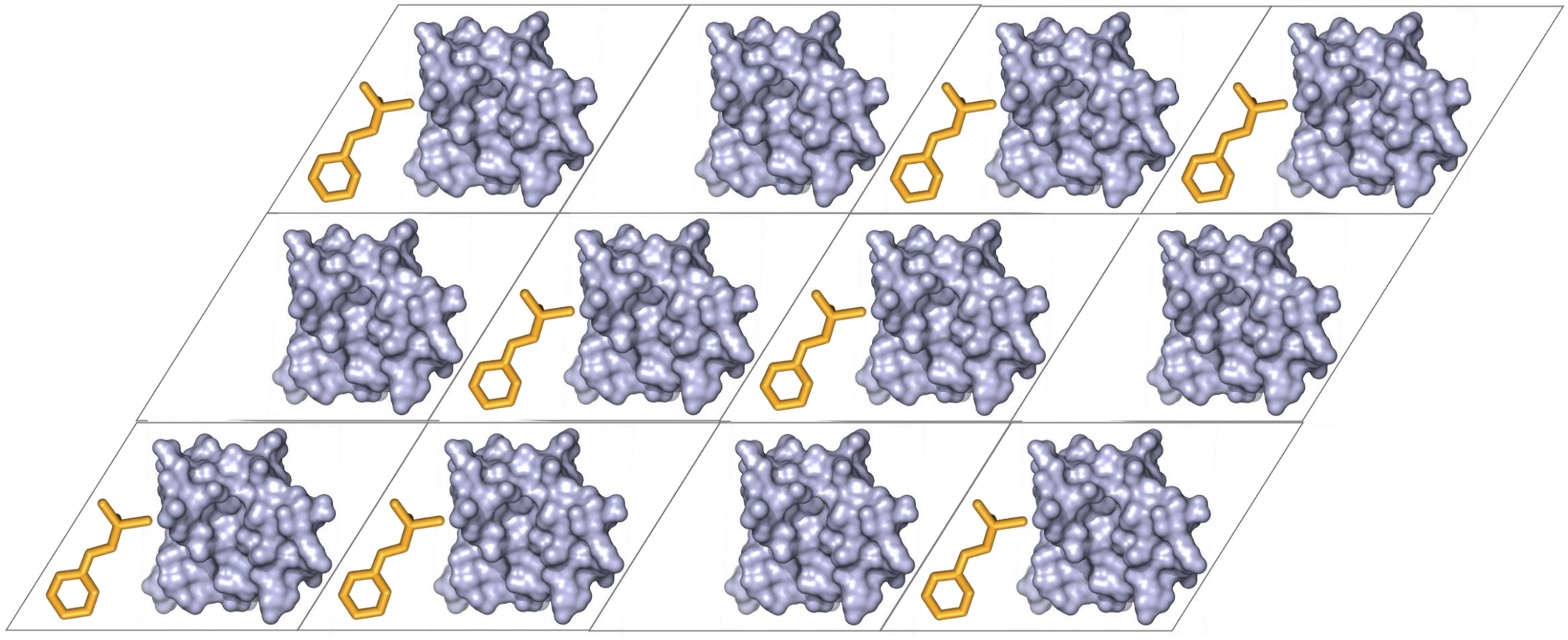
Occupancy refinement

Occupancy models disorder beyond the harmonic approximation:

Occupancy refinement

Occupancy models disorder beyond the harmonic approximation:

- Atoms are not present in every unit cell (e.g. ligand, ion)

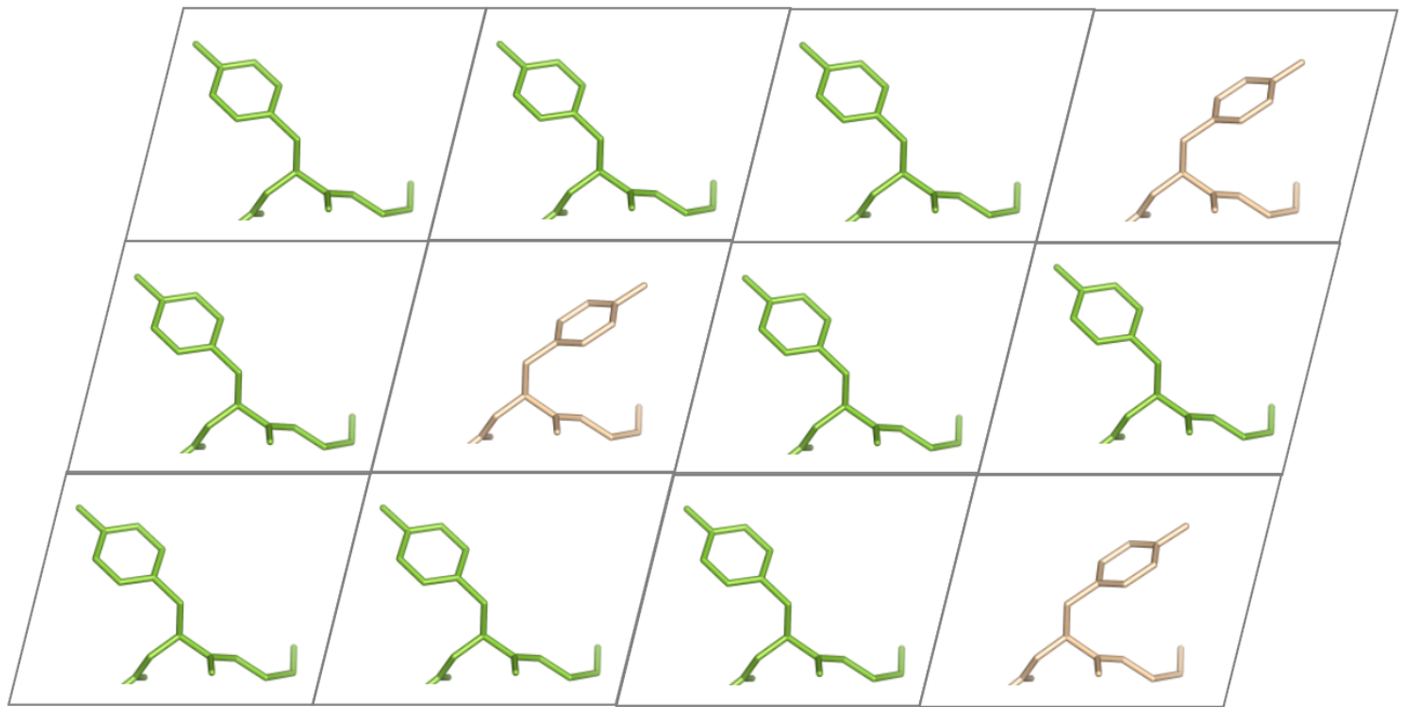


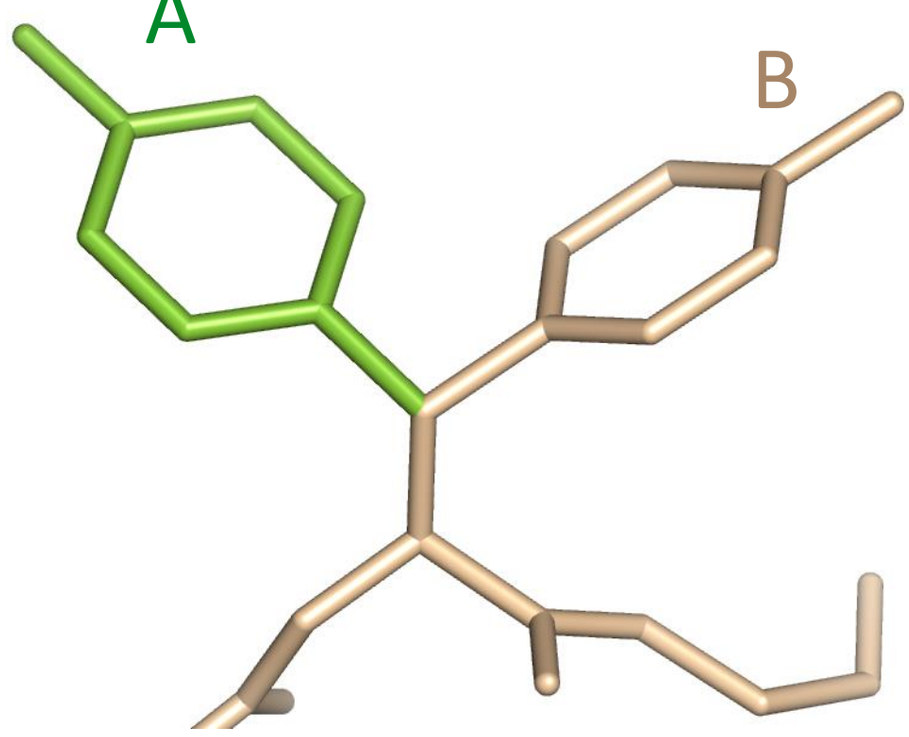
$occ_{ligand} < 1$

Occupancy refinement

Occupancy models disorder beyond the harmonic approximation:

- Atoms are not present in every unit cell (e.g. ligand, ion)
- Atoms are in alternative conformation





$occ_A + occ_B = 1$

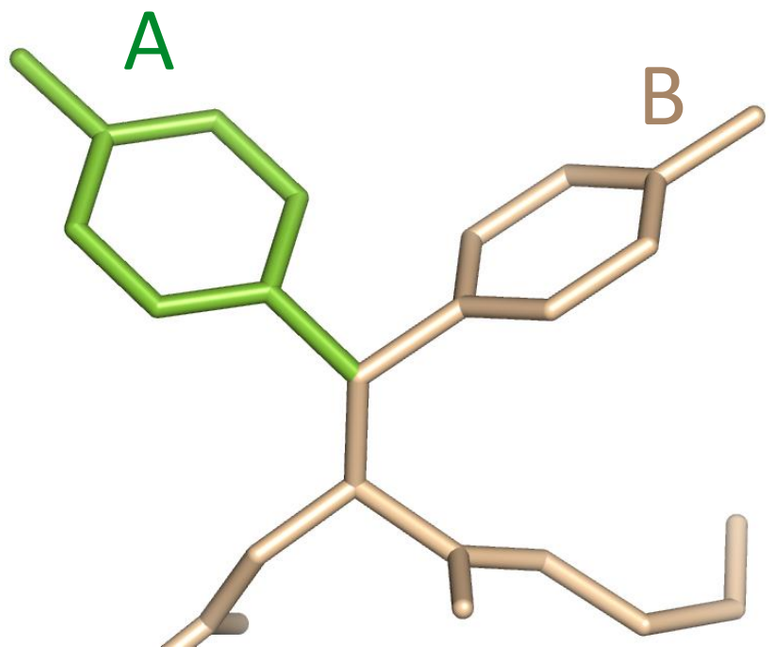
Occupancy refinement

How to refine the occupancy :

- Set the occupancy of atoms to a value < 1 (Coot, PDB tools, …)

How to refine alternative conformations:

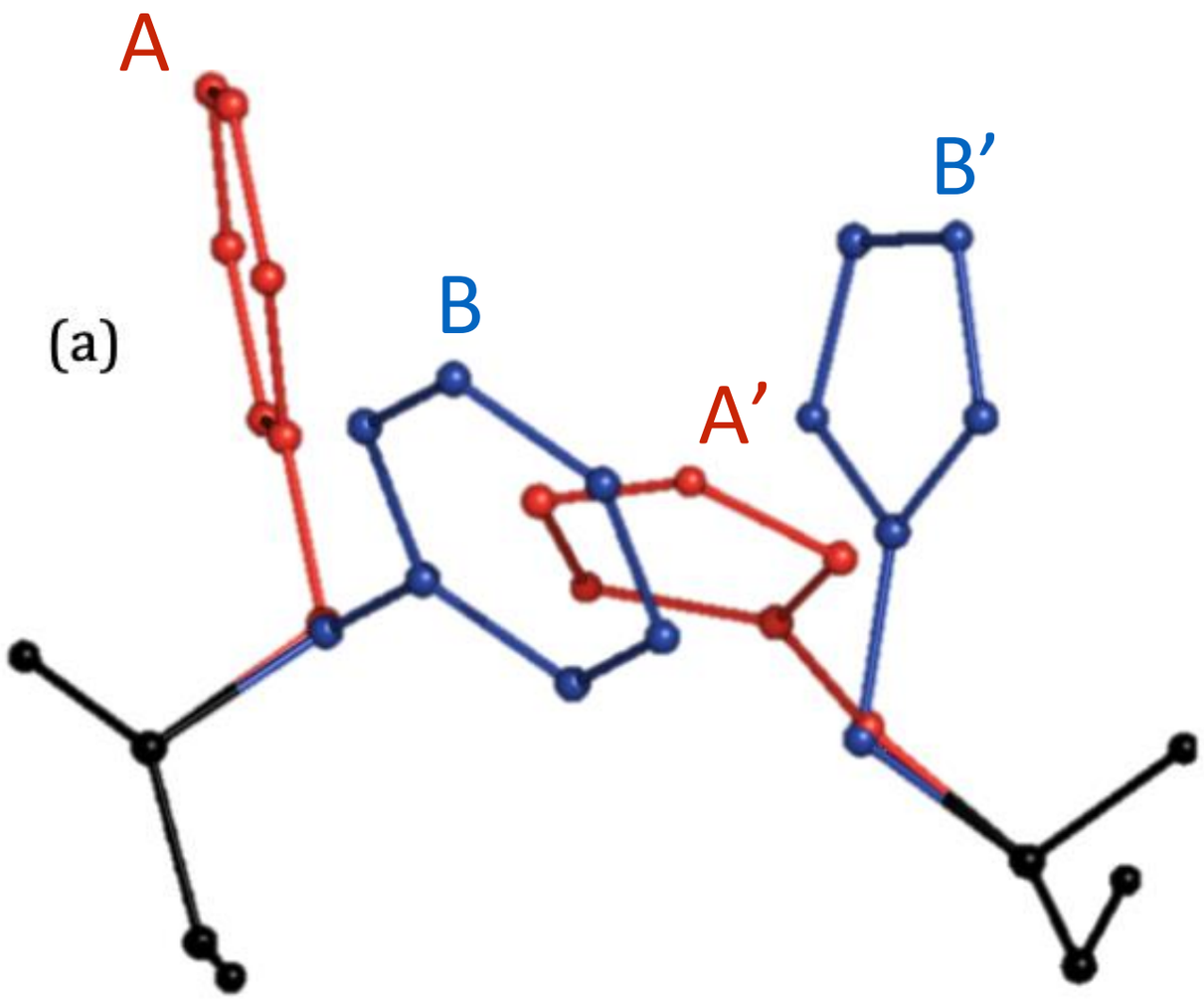
- Add alternative conformation in a molecular viewer (Coot)
- Activate occupancy refinement



phenix.refine:
Occupancy refinement is on by default.
Will only refine atoms with occ<1.

Occupancy refinement

Concerted (coupled) conformations



$occ_A + occ_B = 1$

$occ_{A'} + occ_{B'} = 1$

$occ_B + occ_{A'} = 1$

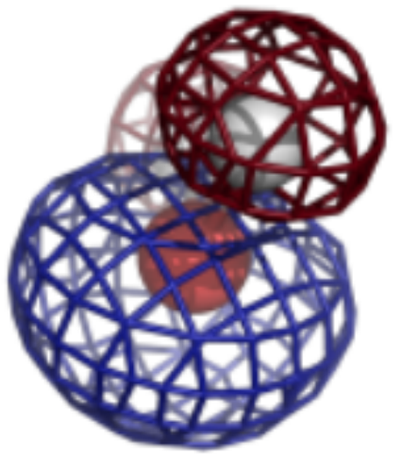
$occ_A = occ_{A'}$

You can couple the two alternative conformations.

```
refinement {
  refine {
    occupancies {
      constrained_group {
        selection = chain A and resseq 73 and altloc A or \
                    chain B and resseq 90 and altloc A or
        selection = chain A and resseq 73 and altloc B or \
                    chain B and resseq 90 and altloc B or
      }
    }
  }
}
```

Occupancy refinement

VOLUME SIX



COMPUTATIONAL CRYSTALLOGRAPHY NEWSLETTER

JULY MMXV

PHASER GUI, ALT. LOCS, BASE-PAIR STACKING

SHORT COMMUNICATIONS

13 typical occupancy refinement scenarios and available options in *phenix.refine*

Pavel V. Afonine

Lawrence Berkeley National Laboratory, Berkeley, CA 94720

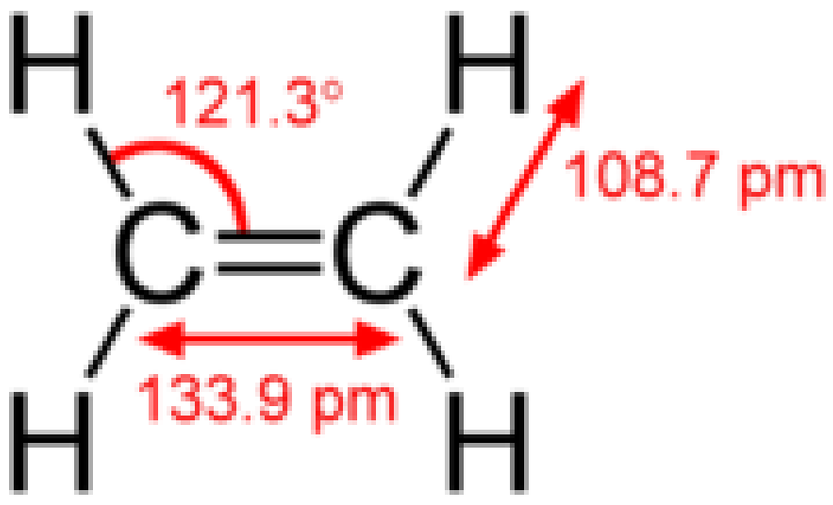
https://phenix-online.org/phenixwebsite_static/mainsite/files/newsletter/CCN_2015_07.pdf#page=12

Restraints: *a priori* knowledge

Restraints increase the number of observations.

Restraints modify the target function by creating relationships between independent parameters.

Example: restrained bond lengths



- the coordinates of the two atoms are independent
- restraint keeps their distance within a certain target value
- imposes a penalty if it deviates too much.

What kind of restraints can you use?

All resolution ranges

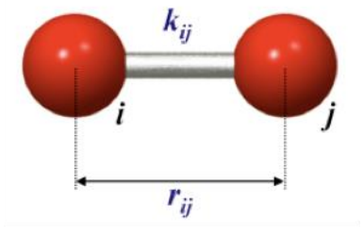
- Stereochemistry
- ADP

Low resolution

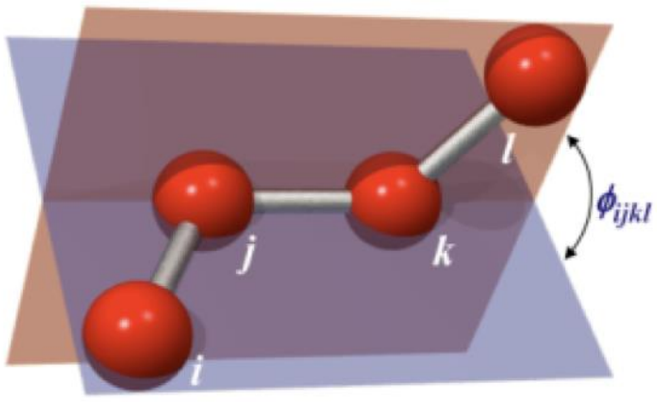
- Secondary structure restraints
- NCS
- Reference model
- Ramachandran

Restraints: *a priori* knowledge

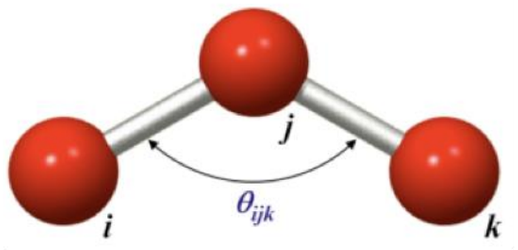
Chemistry



$\Sigma_{bonds}\ \omega(d_{model} - d_{ideal})^2$

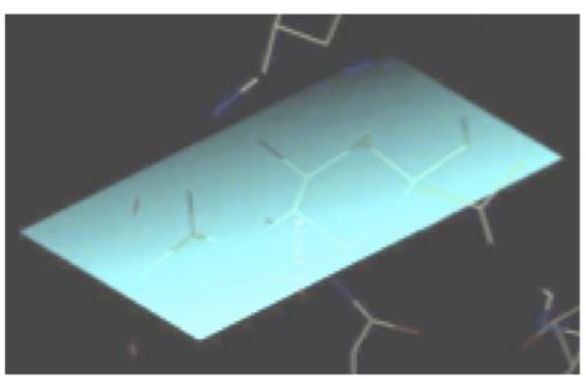


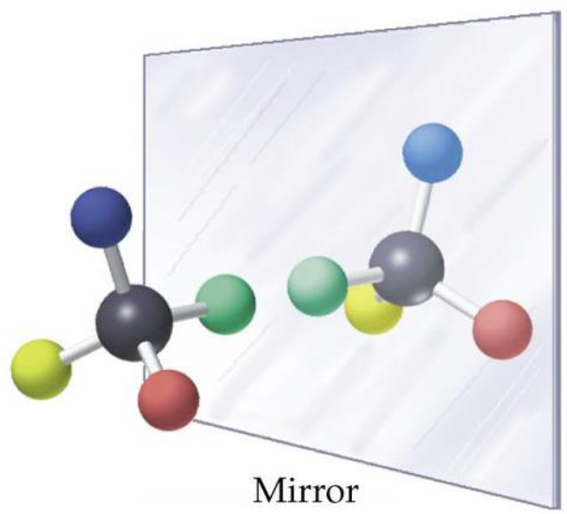
$\Sigma_{dihedrals}\ \omega(1 + \cos(n\chi_{model} + \chi_{shift}))$



$\Sigma_{angles}\ \omega(\theta_{model} - \theta_{ideal})^2$

Images from PumMa web site (http://www.pumma.nl)





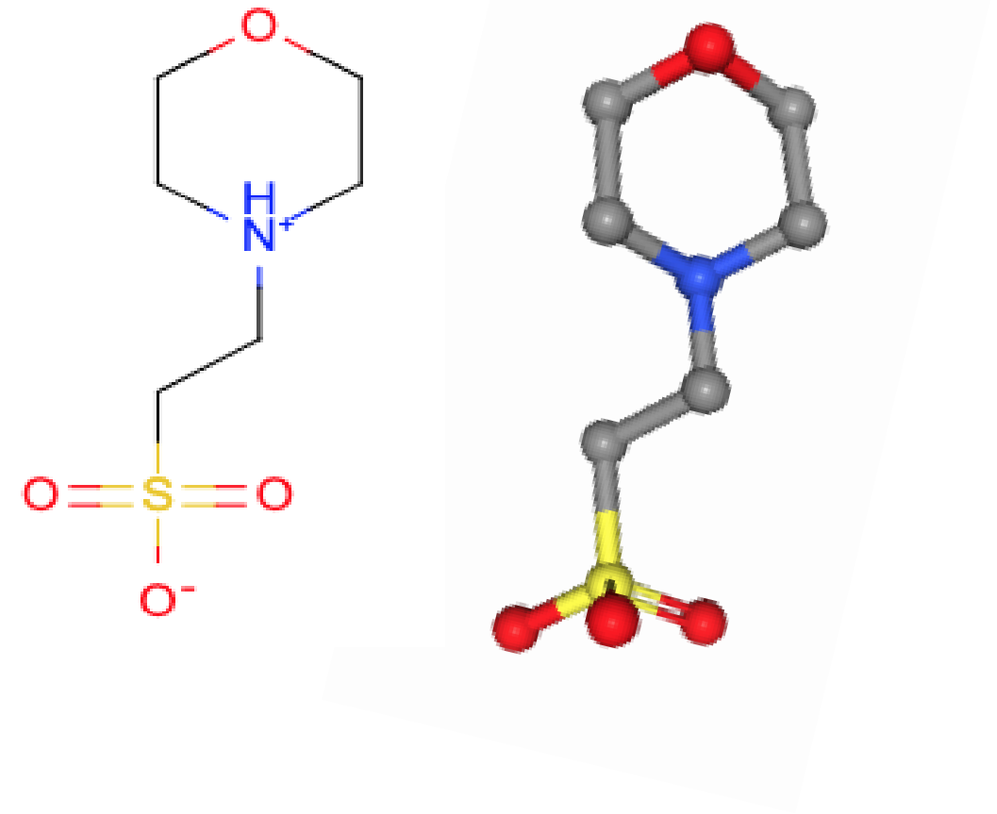
Volume (V) = $(r_N - r_{CA}) \bullet [(r_C - r_{CA}) \times (r_{CB} - r_{CA})]$

$E = \Sigma_{planes} \Sigma_{atoms}\ w\ (m \bullet r - d)^2$

Used automatically (no need to activate)

Restraints: Ligands

Restraints of common ligands are included in libraries.



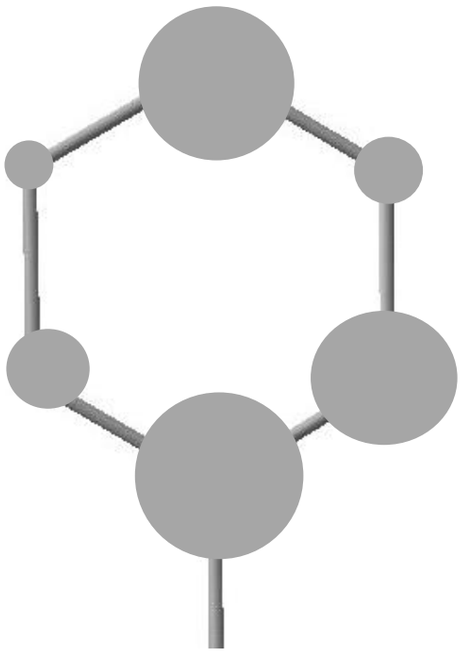
If novel ligand:

restraints need to be generated with a restraints generator

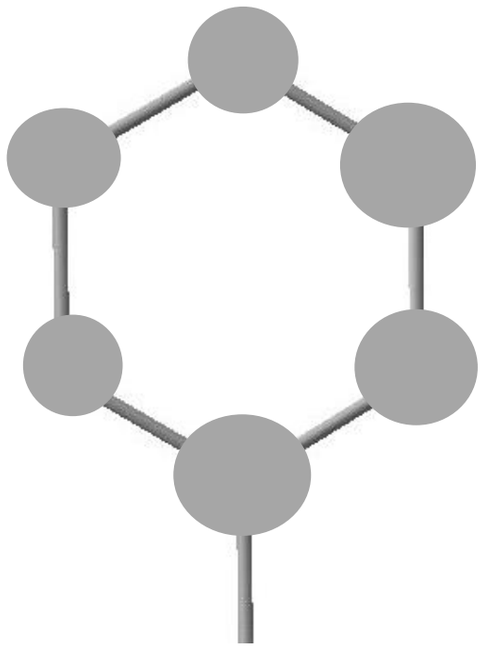
Restraints: ADP

Isotropic ADPs

Unlikely



Reasonable

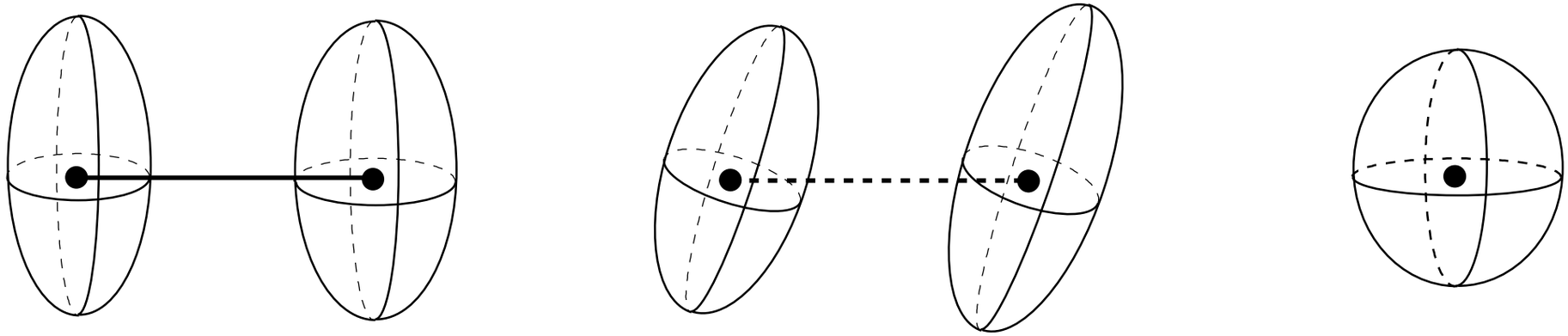


Used automatically (no need to activate).

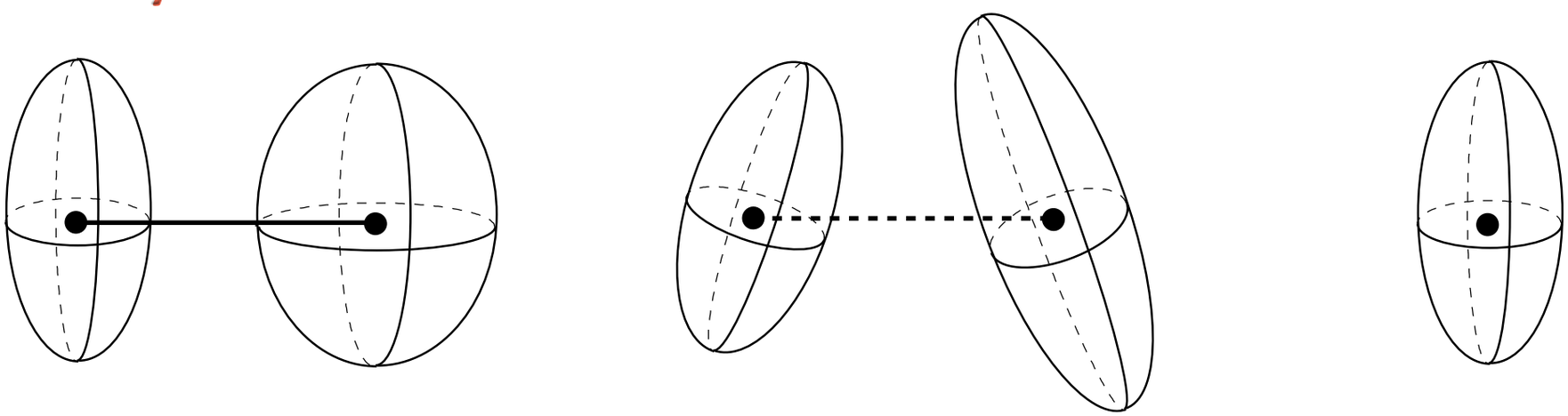
Restraints: ADP

Anisotropic ADPs

Reasonable



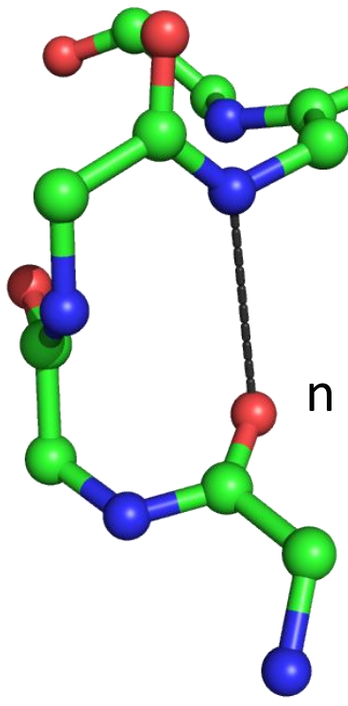
Unlikely



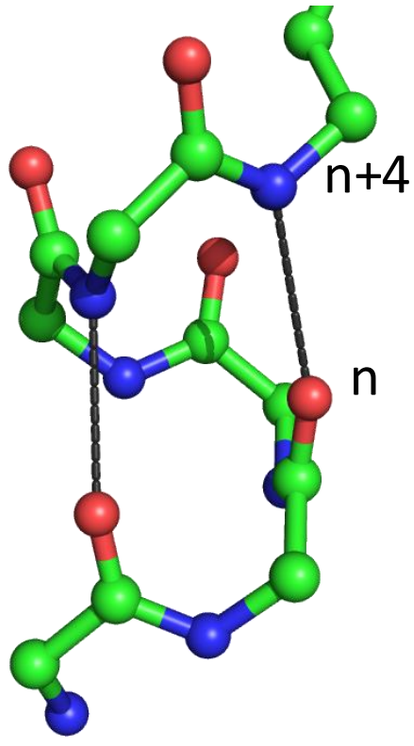
Pictures from Thomas Schneider

Secondary Structure Restraints

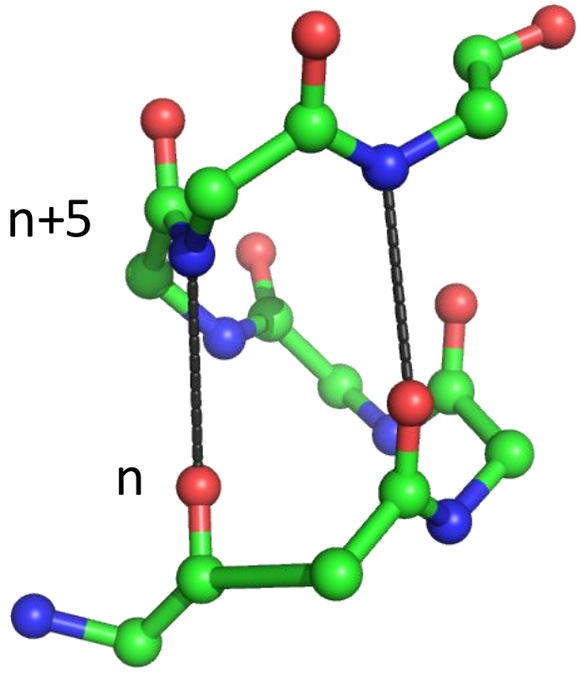
Helices



3_{10} helix



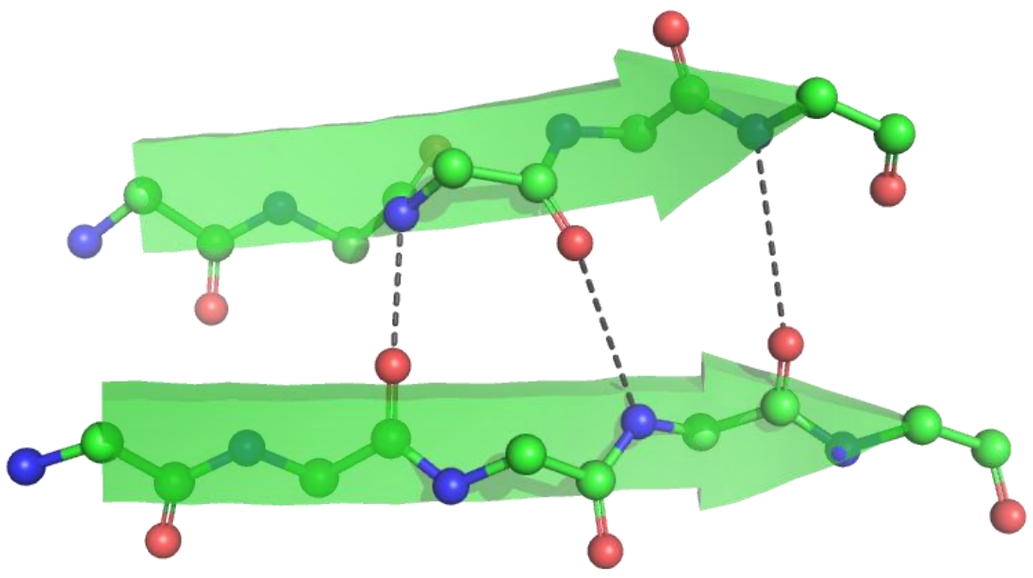
alpha helix



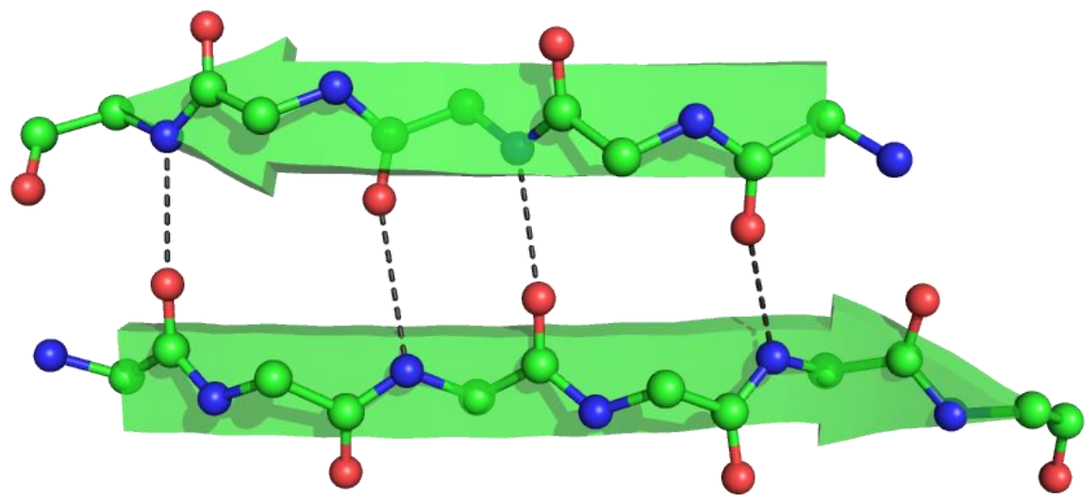
pi helix

Secondary Structure Restraints

Sheets



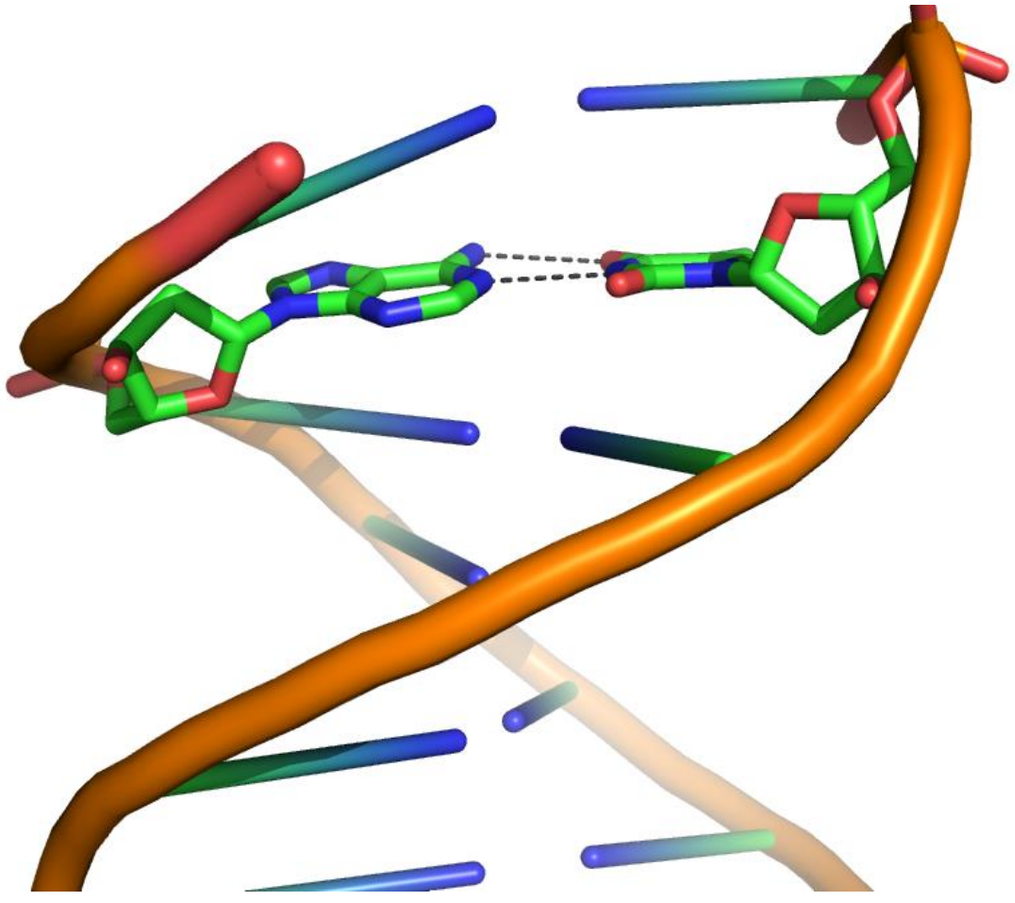
parallel



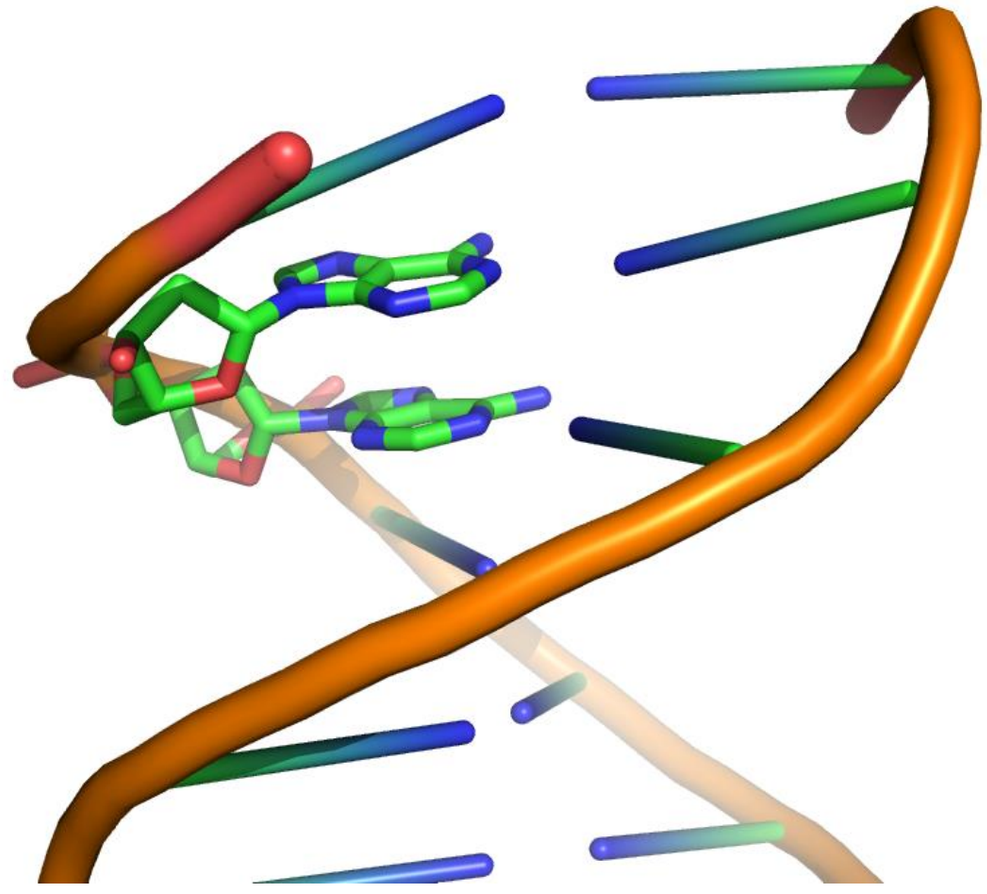
antiparallel

Secondary Structure Restraints

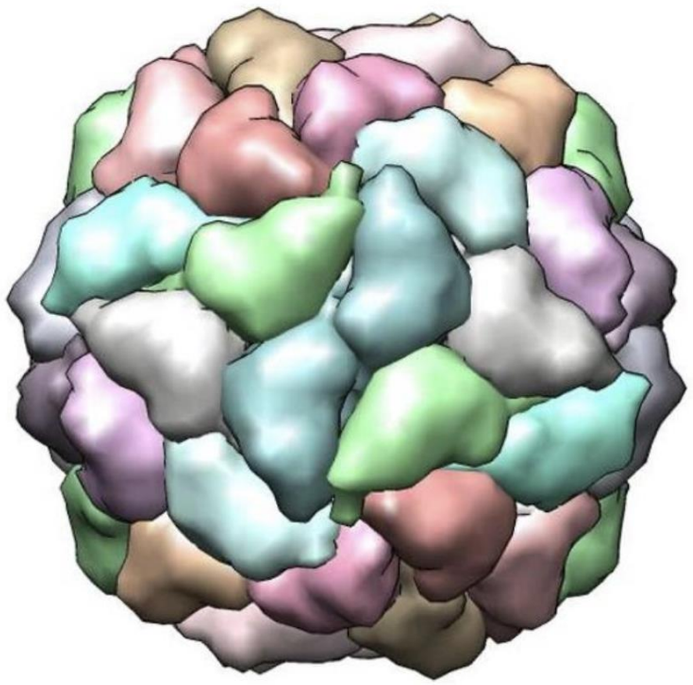
Base pairs

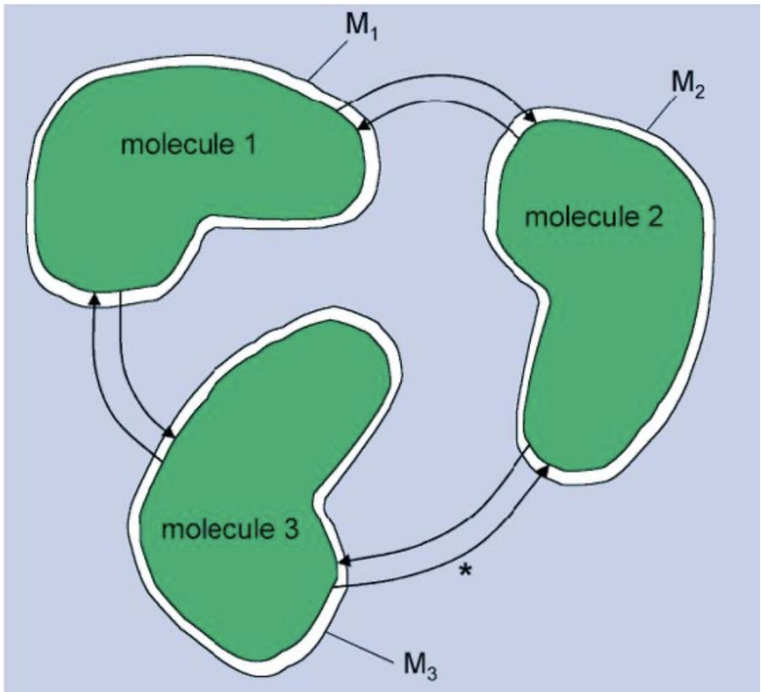


Stacking pairs



NCS – non crystallographic symmetry





Constrain mols 1, 2 and 3 to be identical.
Restrain mols 1, 2 and 3 to be similar.

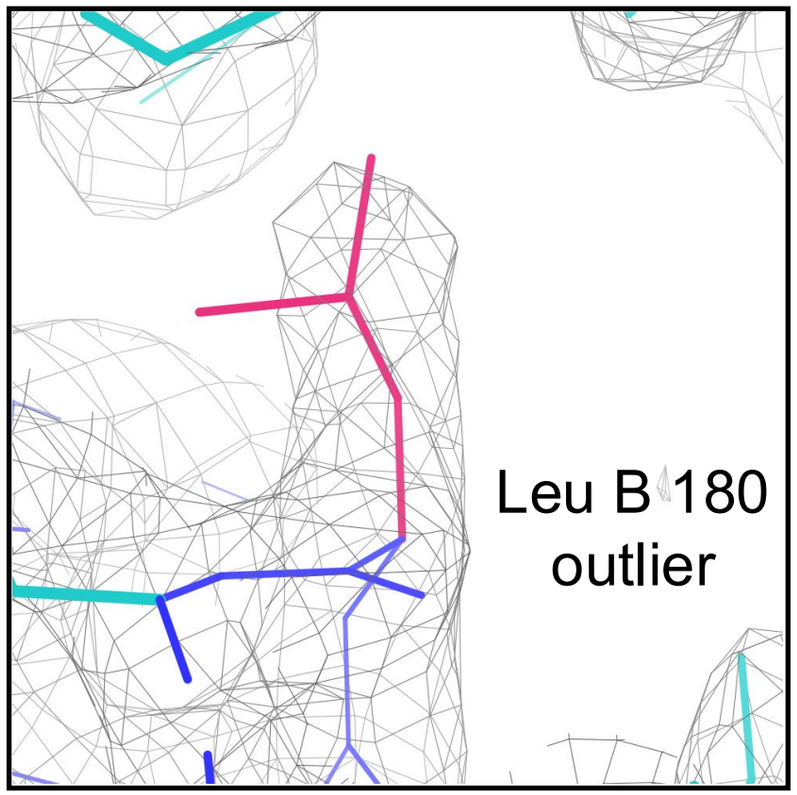
Cartesian restraints:
Atoms in NCS-related chains are restrained to the average xyz position.

NCS restraints

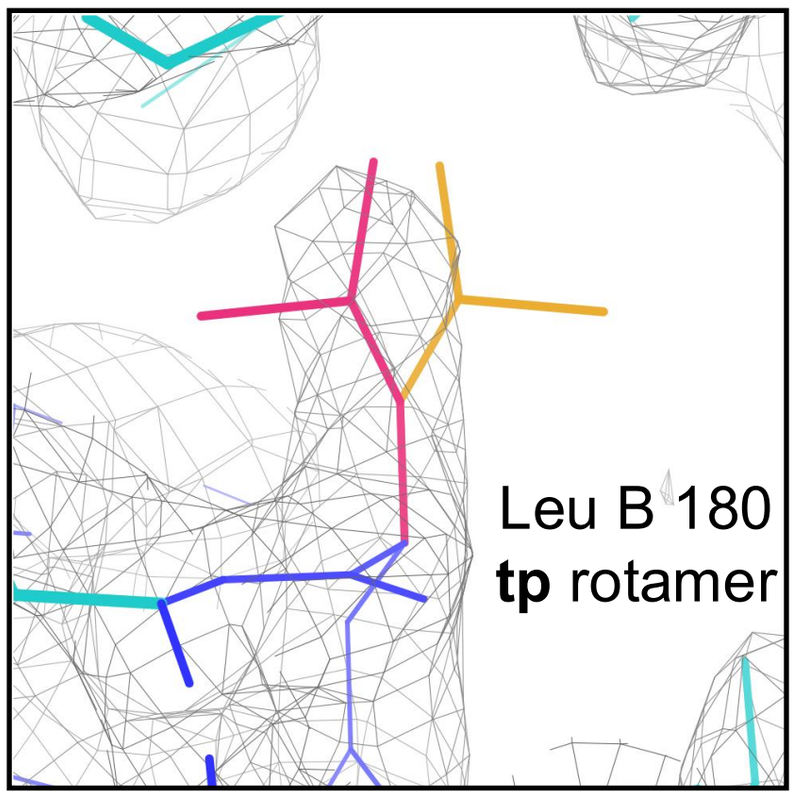
Torsion restraints:

Restrain dihedral angles;

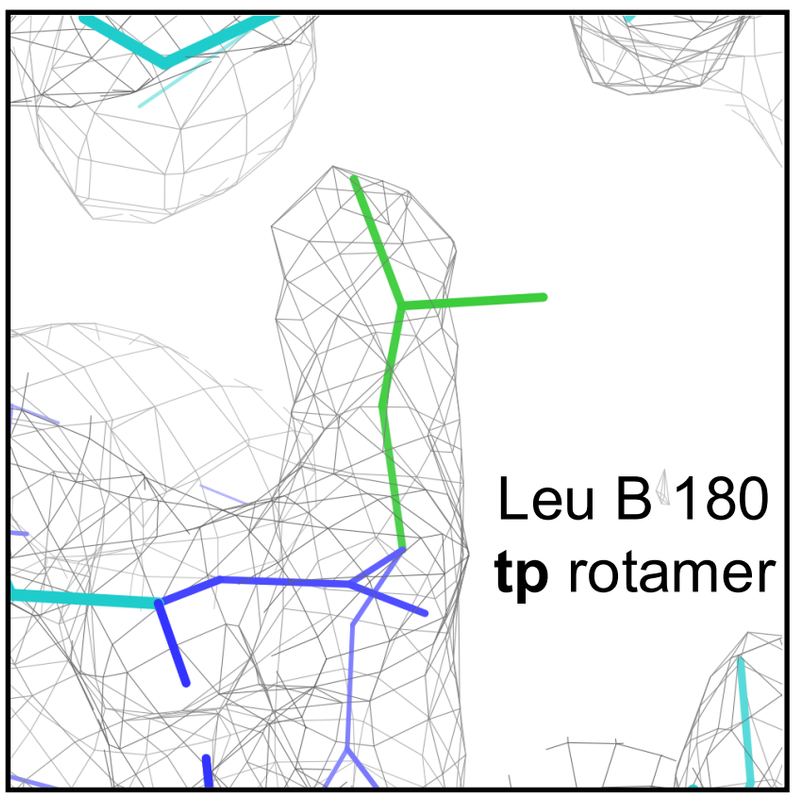
Allow them to be unrestrained if genuinely different.



1. Identify rotamer outlier



2. correct to corresponding rotamer in NCS-related chain by matching χ angles



3. verify rotamer is still correct match

1b04: 2.8 Å

DNA ligase

NCS – which one to use?

Consider:

- Does my model have NCS?

Constraints vs. restraints:

- cryo-EM: Was my map symmetrized?
- Is my data good enough to reasonably expect to see difference in NCS copies?

Torsion restraints are generally preferable of Cartesian restraints.

Reference model Restraints

When to use:

D

s

C

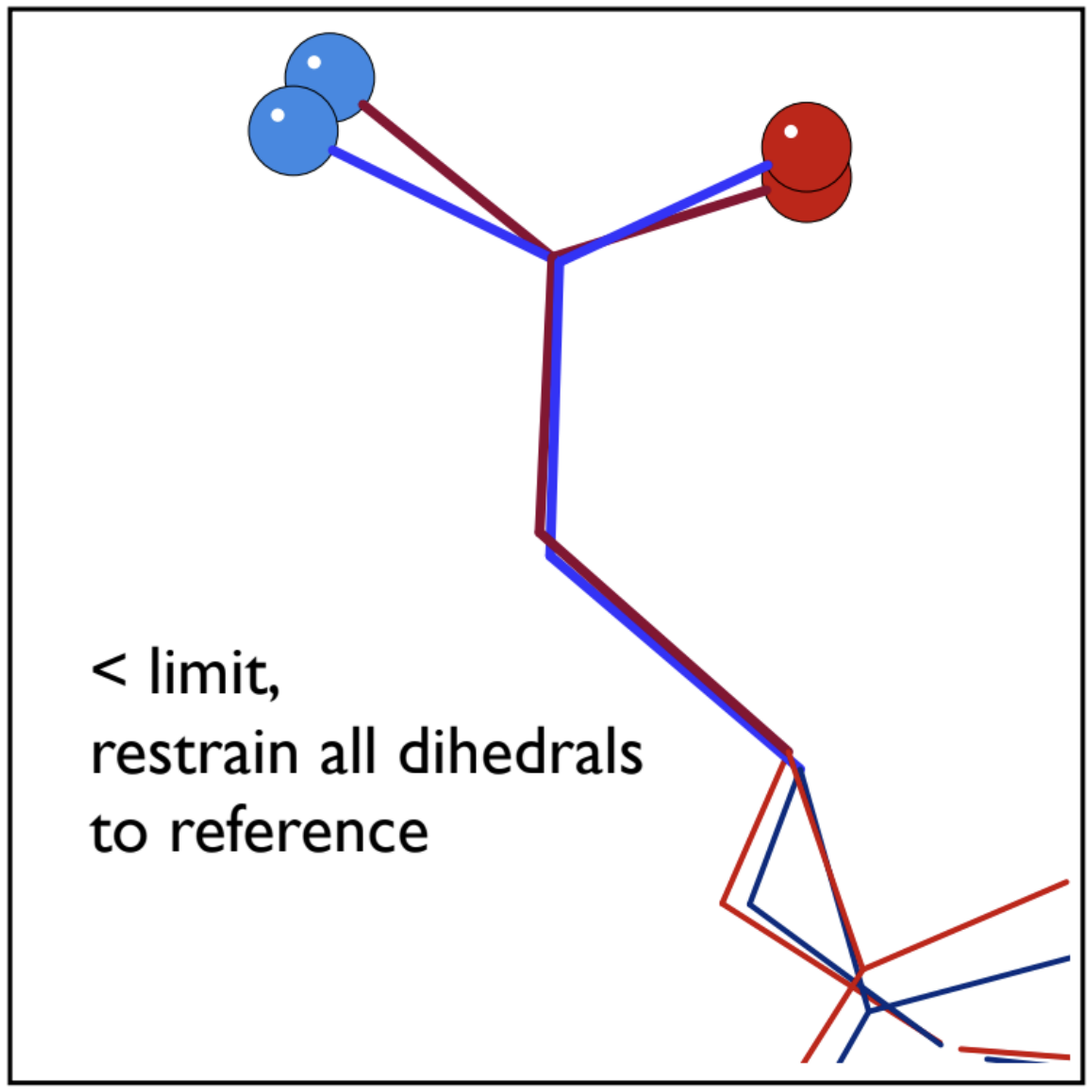
F

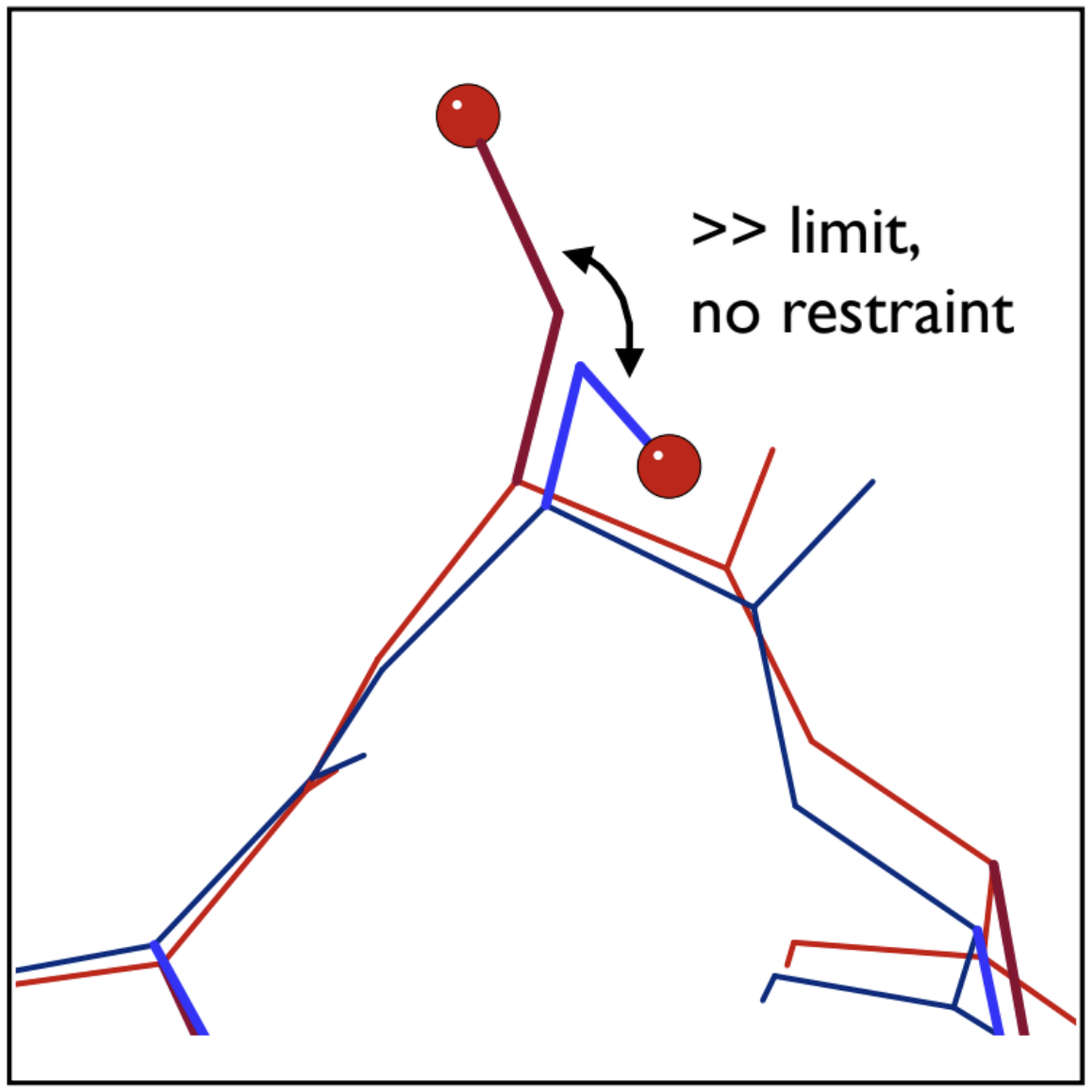
C

A

The 'limit' parameter

default: limit = 15.0°



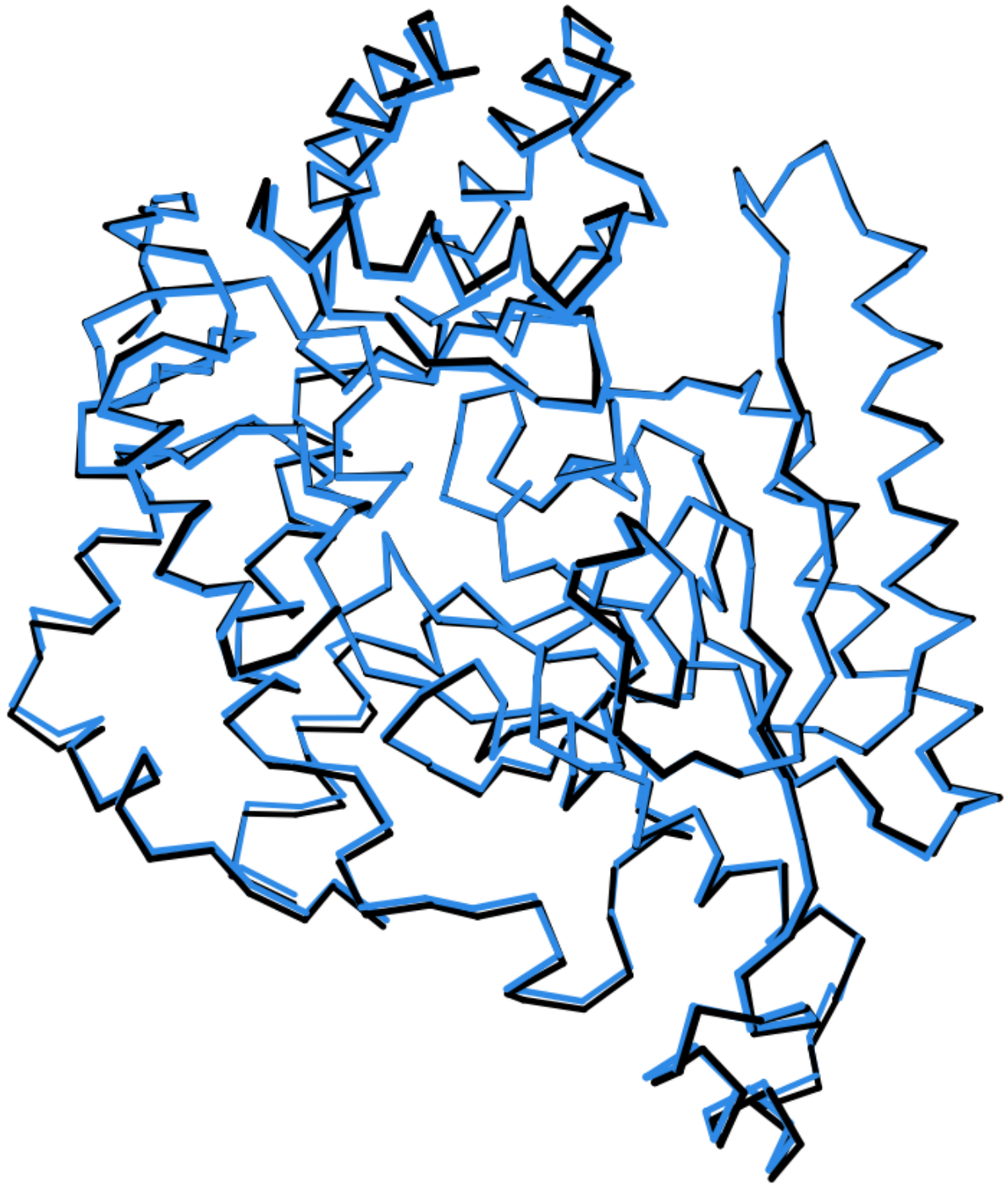


Headd JJ et al., 2012, *Acta Cryst.* **D68**:381-390

Reference model Restraints

1GTX: 3.0 Å

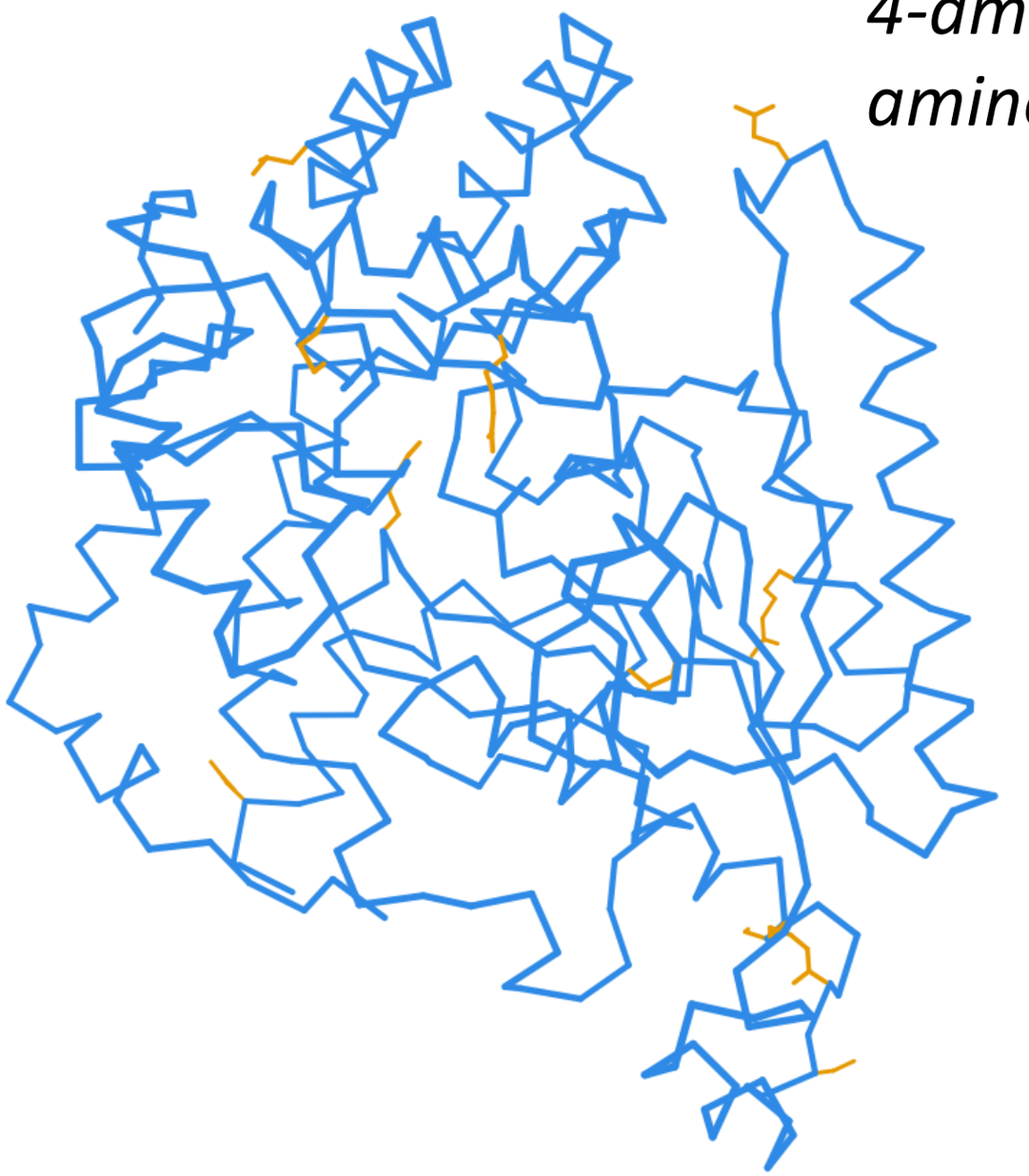
1OHV: 2.3 Å



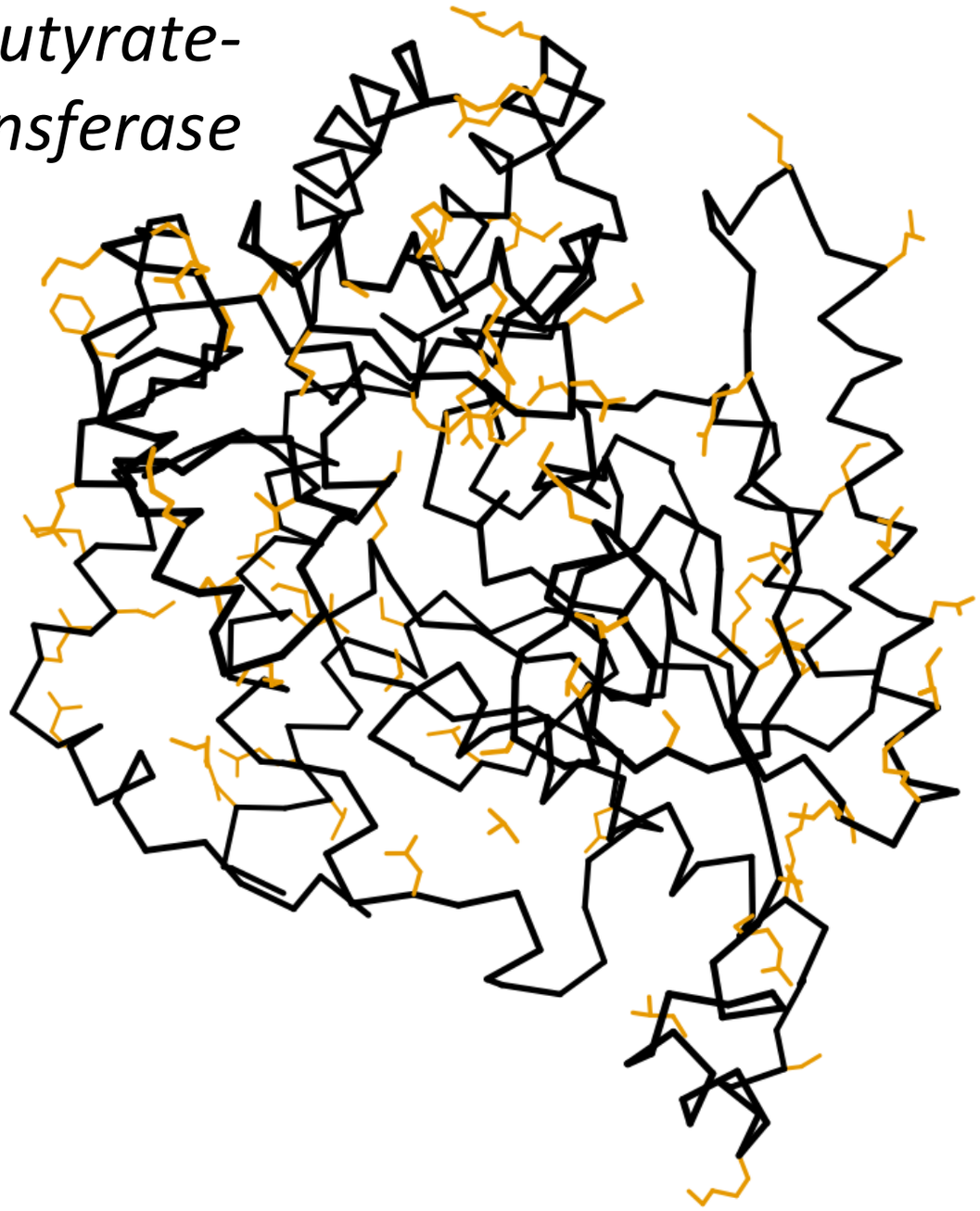
4-aminobutyrate-aminotransferase

Reference model Restraints

4-aminobutyrate-aminotransferase



1OHV: 2.3 Å



1GTX: 3.0 Å

Ramachandran plot restraints

- The backbone dihedral angles can be restrained to stay in the allowed regions of the Ramachandran plot
- Prevent the model from degrading at low resolution when the conformation is approximately correct.

→ Don’t use Ramachandran restraints to “fix” Ramachandran outliers.

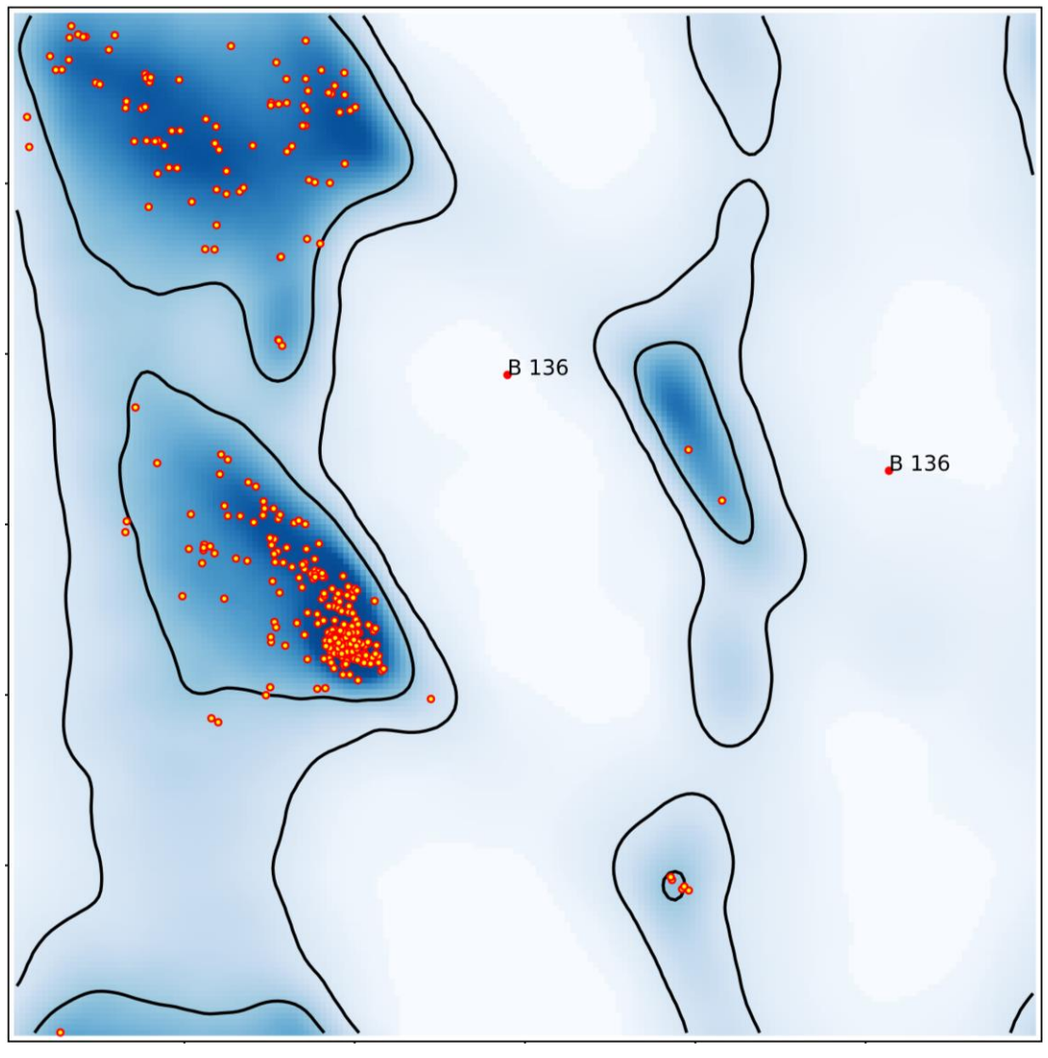
Keep in mind:

Don’t rely on Ramachandran plot for validation.

Two Ramachandran distributions

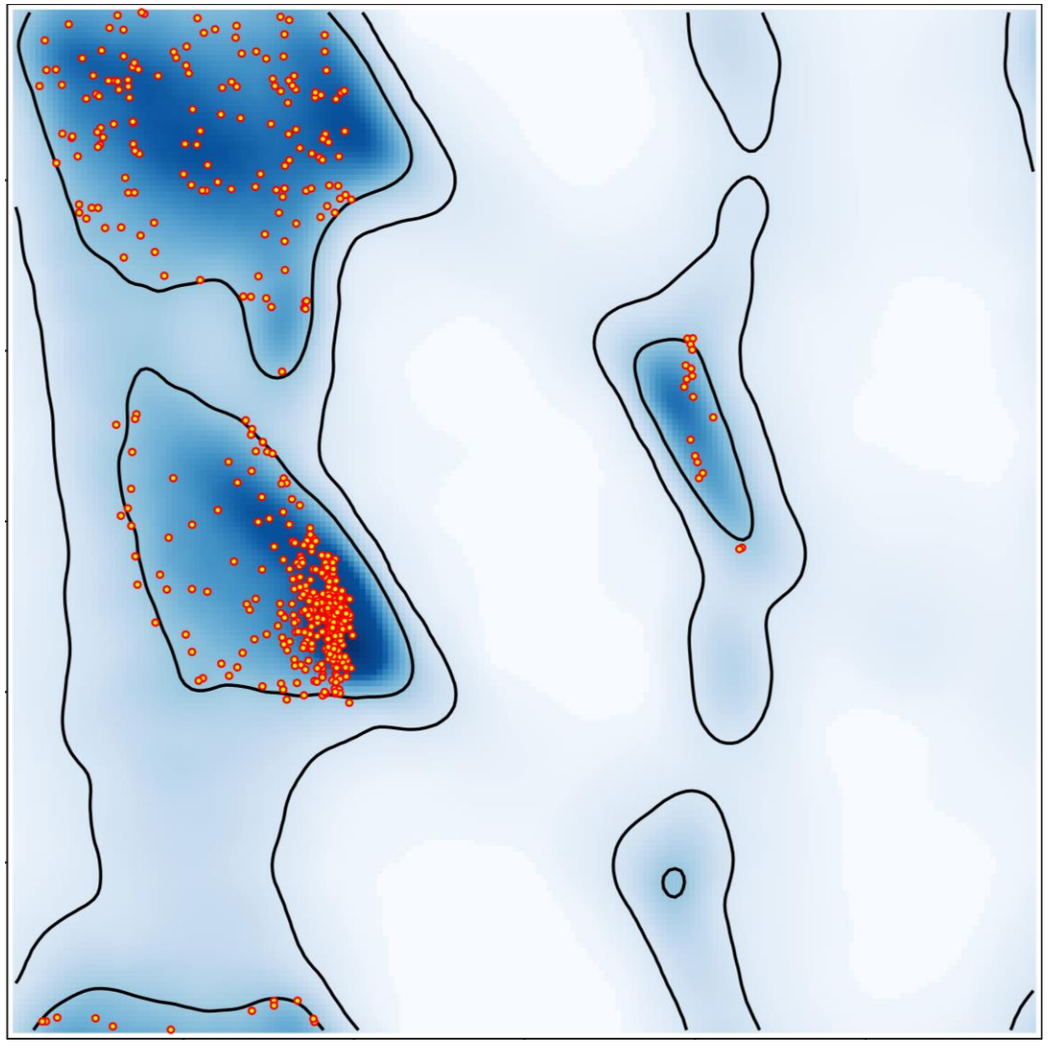
Model A

Favored	97.8 %
Allowed	1.95 %
Outliers	0.25 %



Model B

Favored	96.2 %
Allowed	3.8 %
Outliers	0.0 %



Global Ramachandran Score

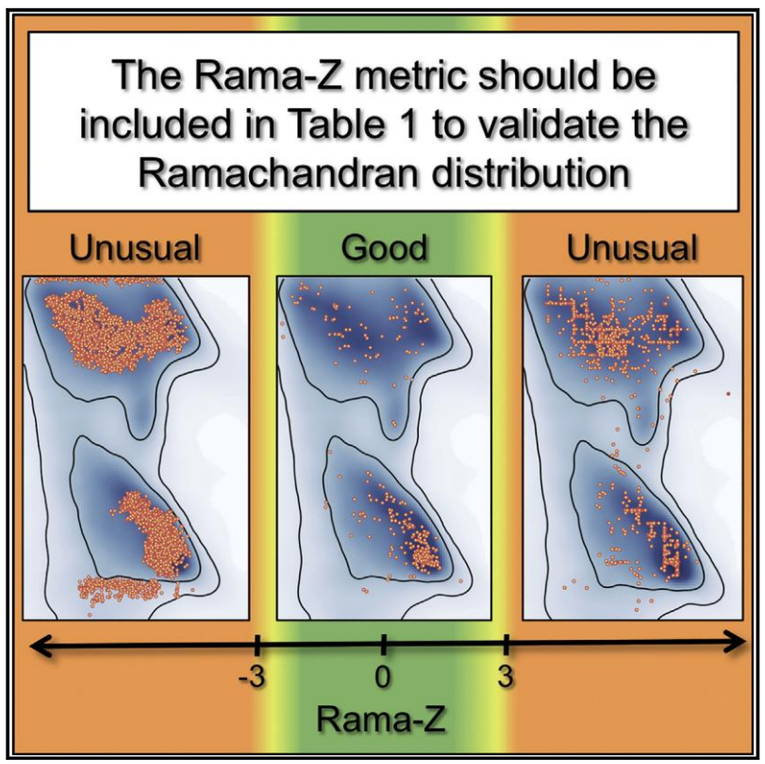
Resource

Structure

Structure, **28**. 1249-1258.

A Global Ramachandran Score Identifies Protein Structures with Unlikely Stereochemistry

Graphical Abstract



Authors

Oleg V. Sobolev, Pavel V. Afonine, Nigel W. Moriarty, Maarten L. Hekkelman, Robbie P. Joosten, Anastassis Perrakis, Paul D. Adams

Correspondence

osobolev@lbl.gov (O.V.S.), r.joosten@nki.nl (R.P.J.)

In Brief

Counting the number of Ramachandran outliers is not sufficient for protein backbone validation. Sobolev et al. revisited the underutilized Ramachandran *Z* score. The authors describe its reimplementation in Phenix and PDB-REDO and showcase its utility. They advocate including it in the validation reports provided by the Protein Data Bank.

Initially proposed in 1997!

CABIOS

Vol. 13 no. 4 1997
Pages 425–430

Objectively judging the quality of a protein structure from a Ramachandran plot

Rob W.W.Hooft, Chris Sander and Gerrit Vriend

Global Ramachandran Score

Model A

Favored	97.8
Allowed	1.95
Outliers	0.25

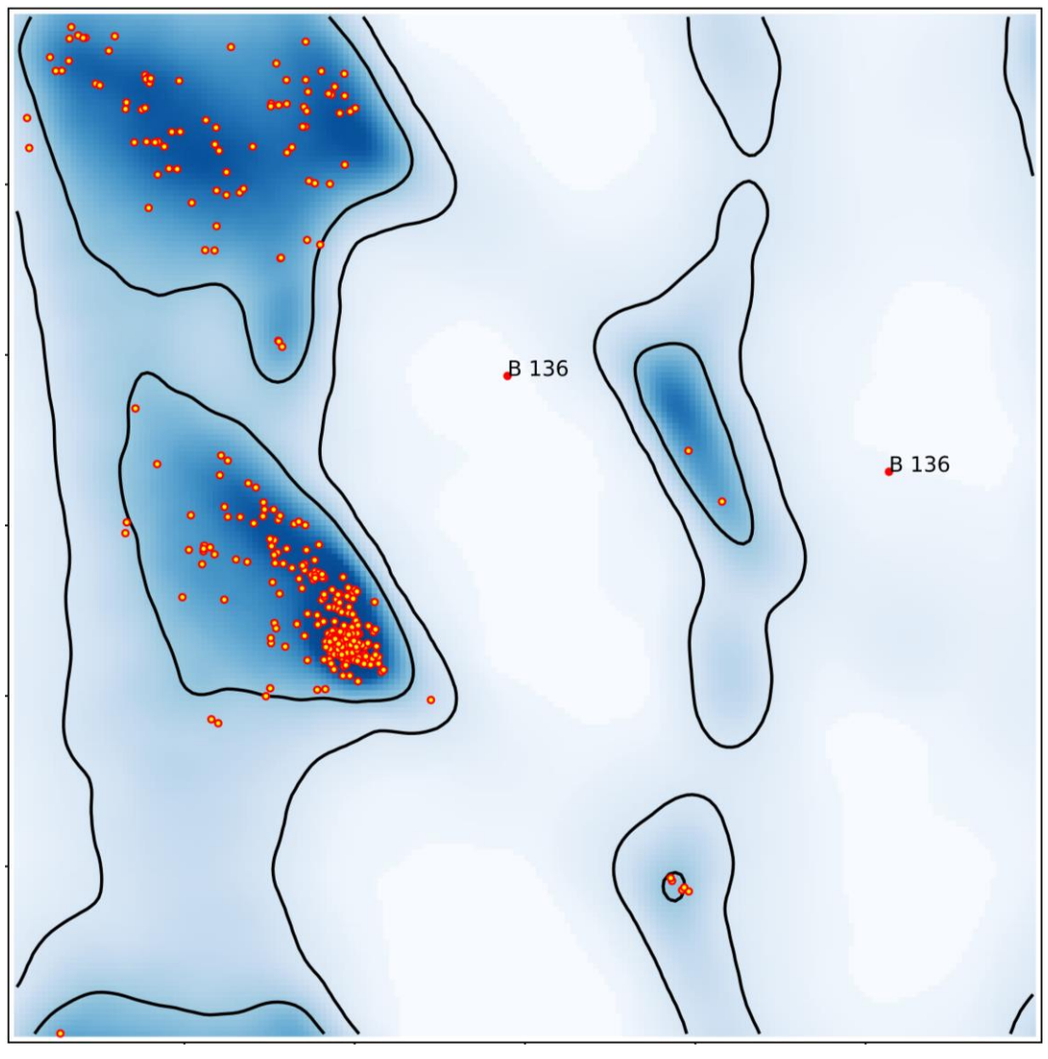
Rama z-score

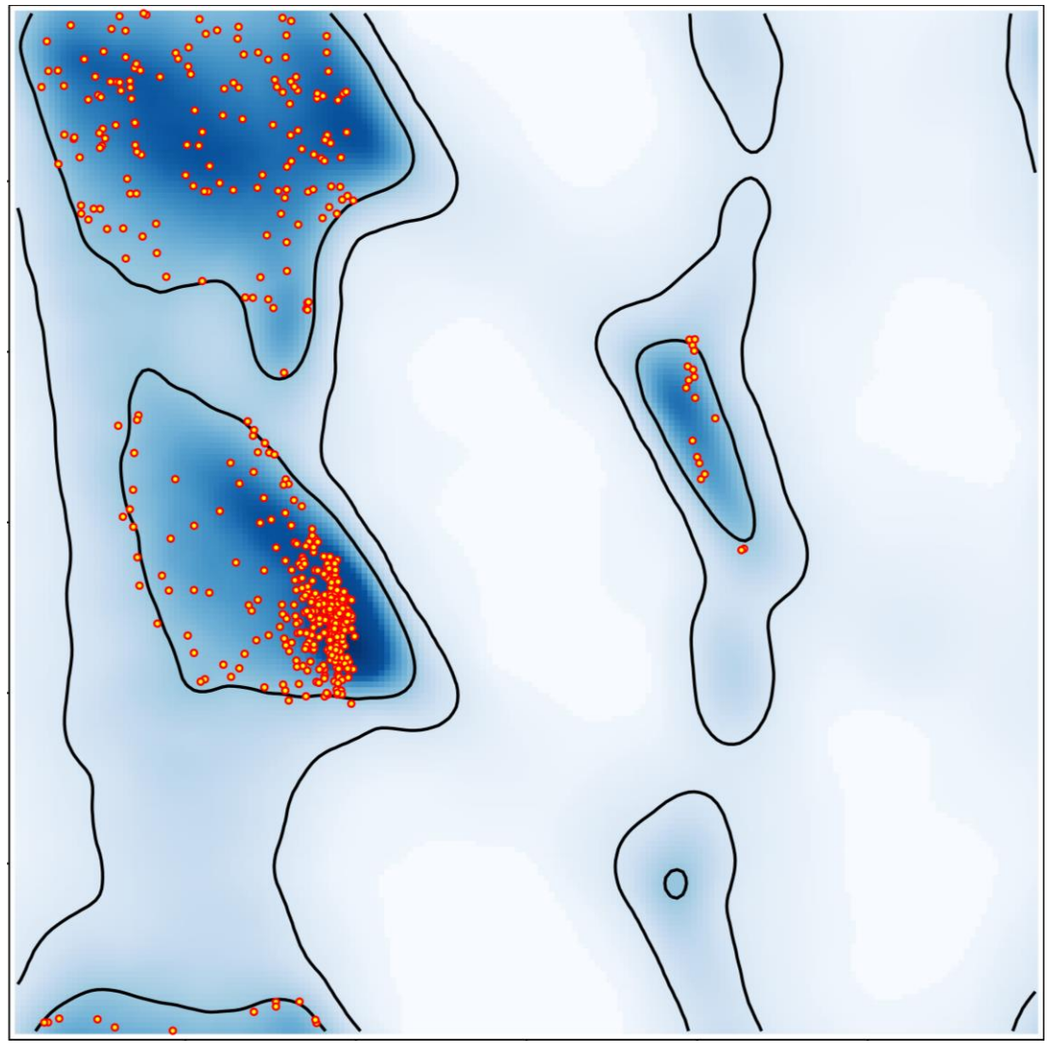
-0.19

-4.08

Model B

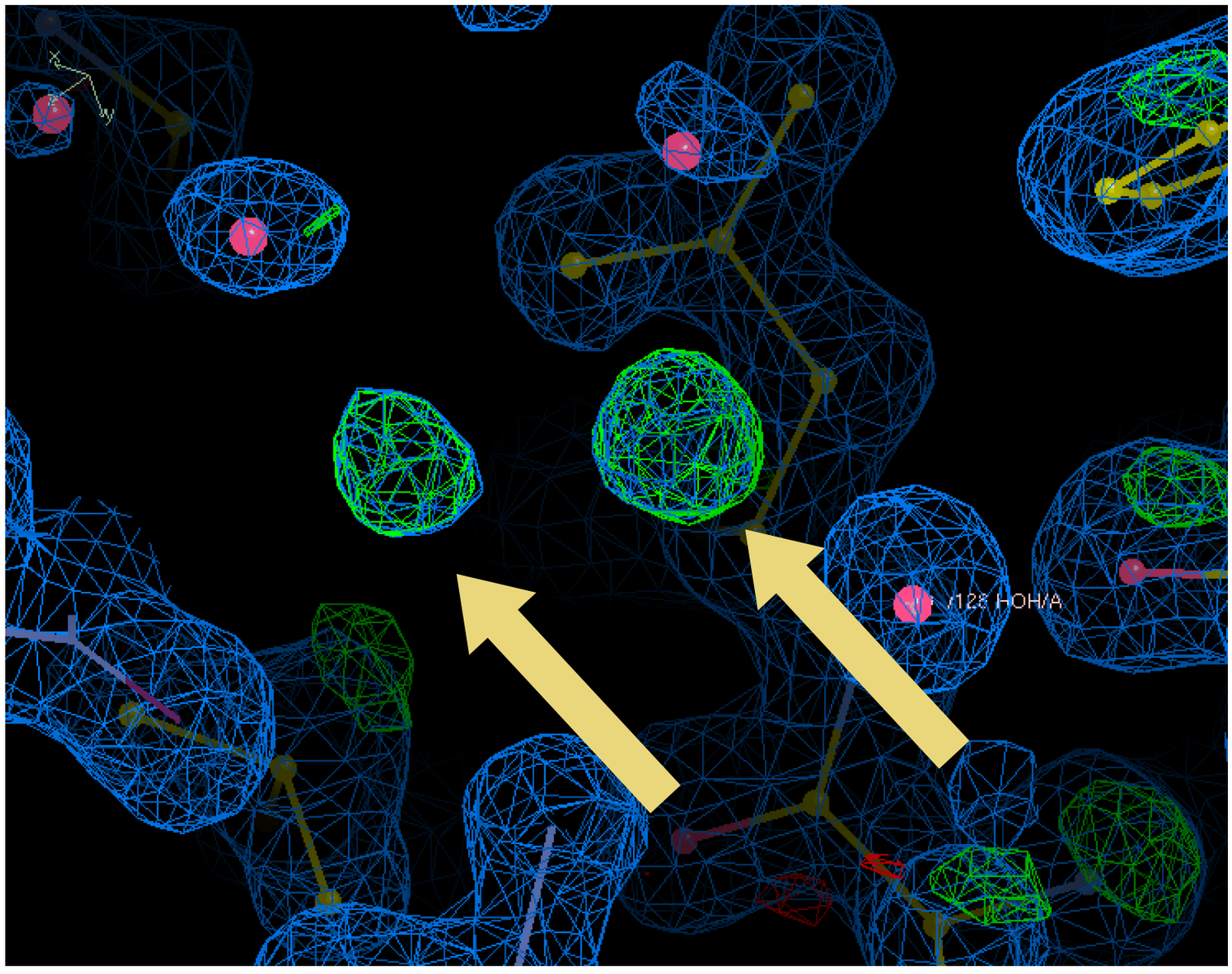
Favored	96.2
Allowed	3.8
Outliers	0.0





Automated water picking

Phenix.refine can build waters automatically.

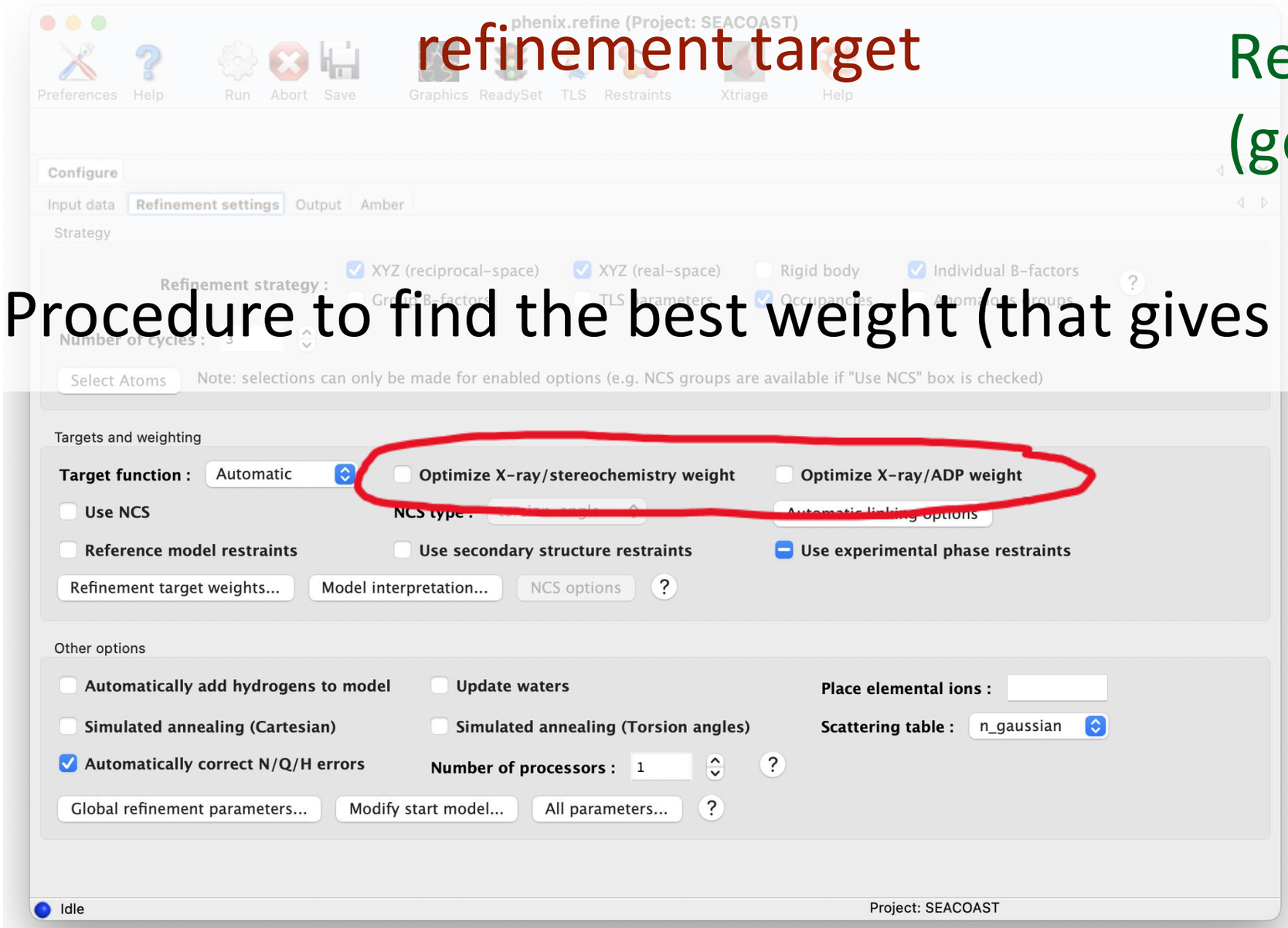


Weight optimization

$$T = w_1 T_{Data}(F_{obs}, F_{Model}) + w_2 T_{Restraints}$$

Crystallographic refinement target

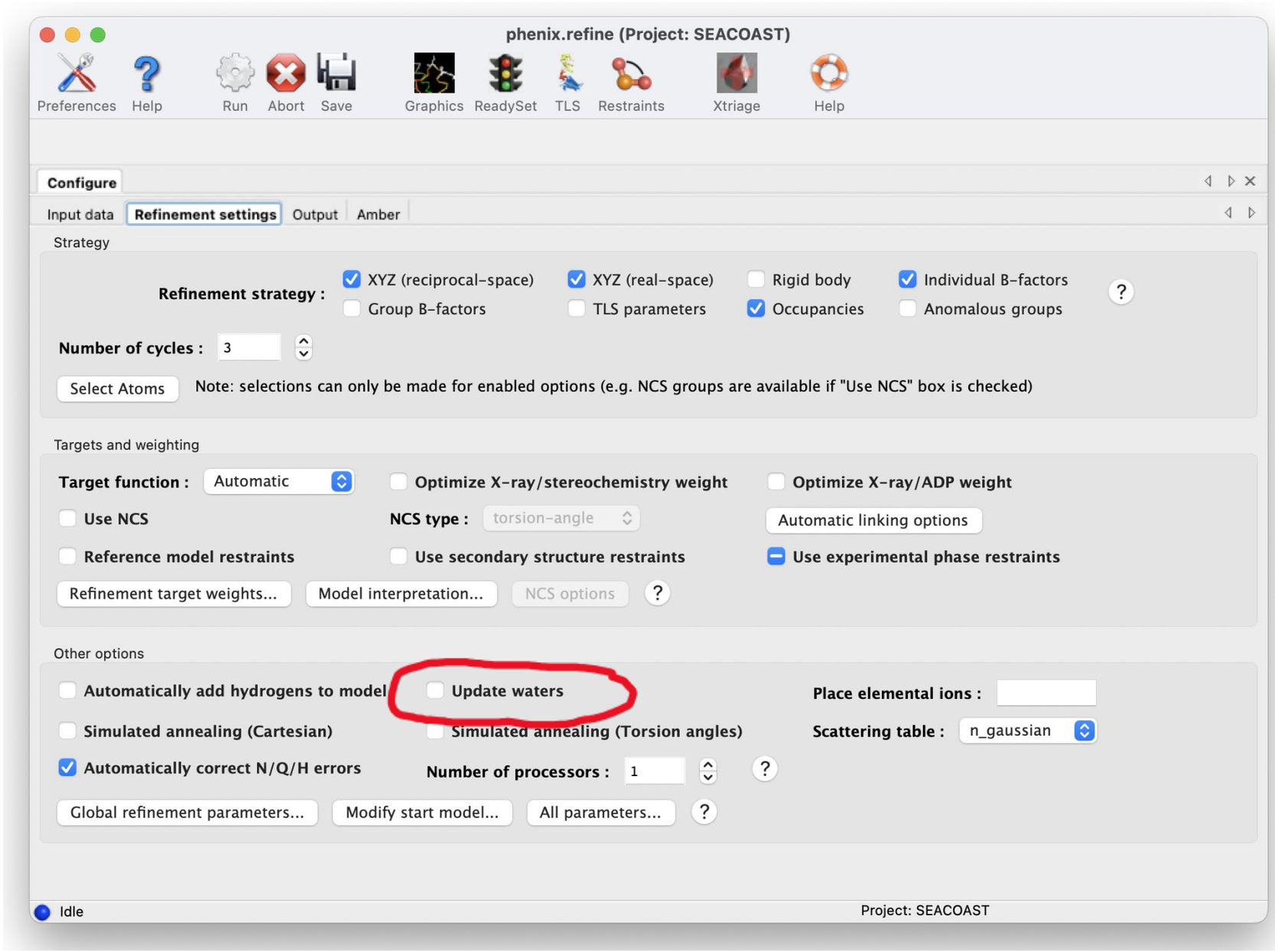
Restraints target (geometry, ADP)



Procedure to find the best weight (that gives lowest R_{free})

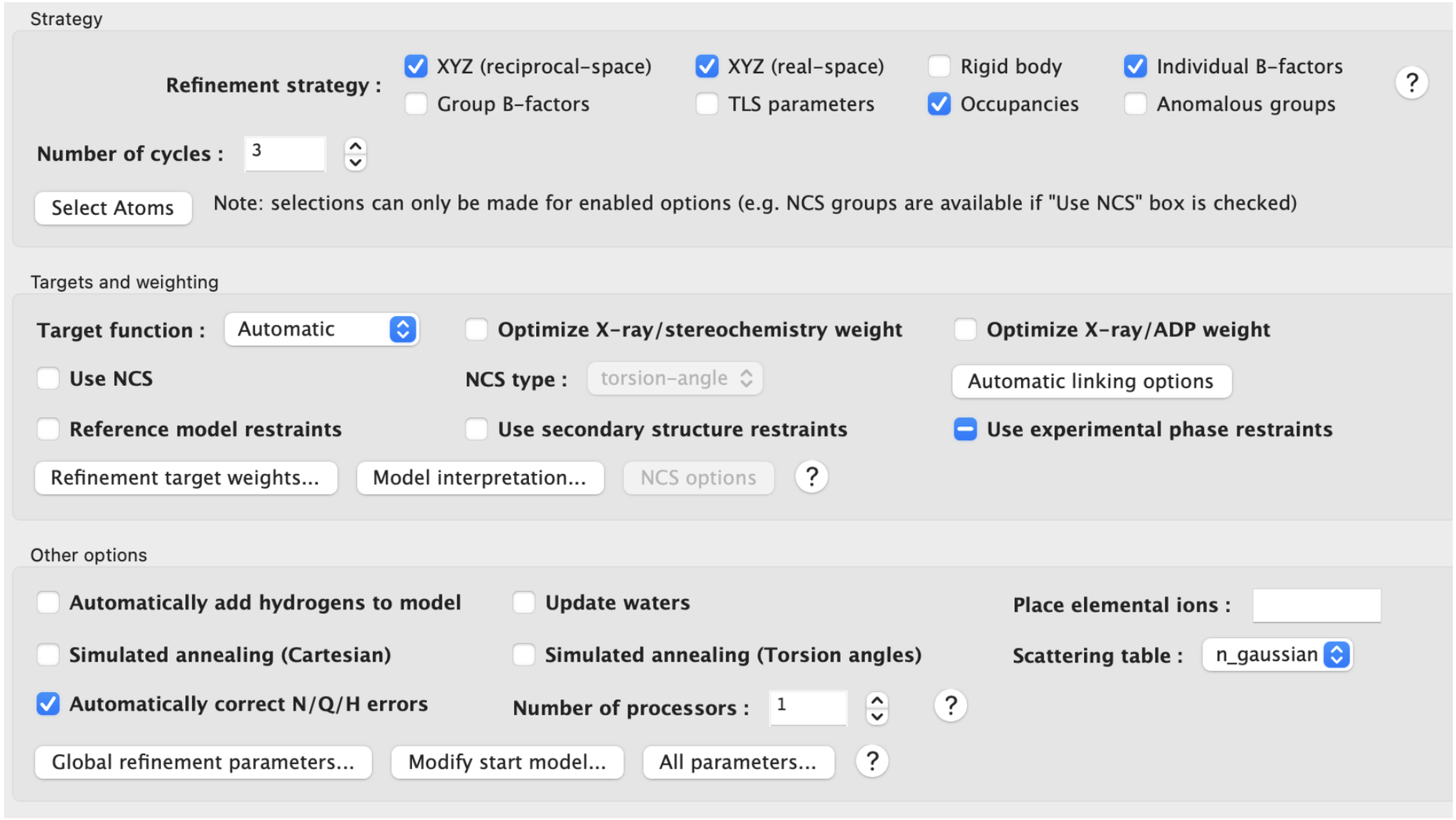
Automated water picking

Phenix.refine can build waters automatically.



You can then check them in a molecular viewer.

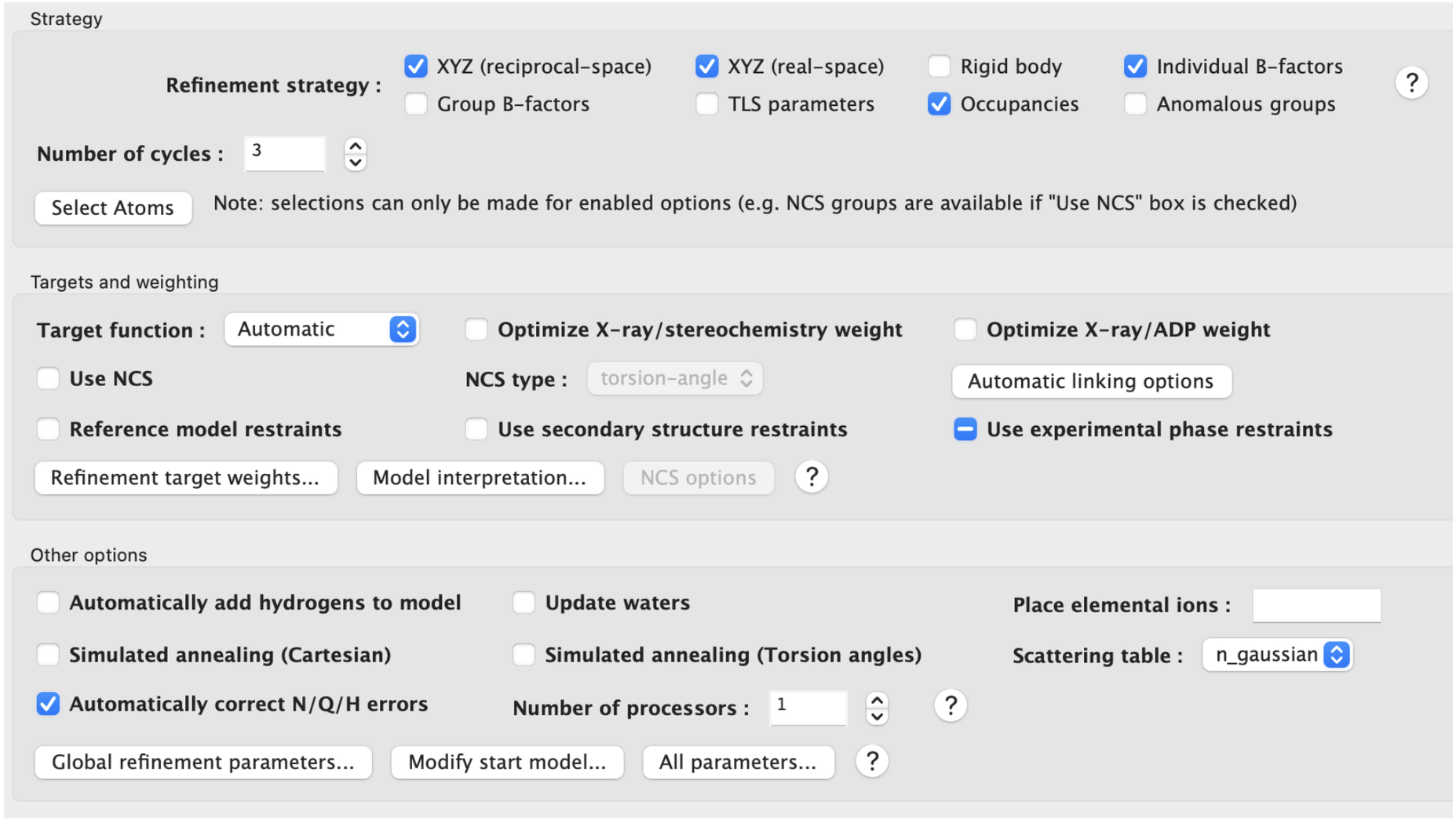
Deciding about a particular parameterization



There are no easy recipes:

- “Activate all boxes: the more options, the better.”
- “Always turn on TLS after 10 rounds of refinement/Coot model building.”

Deciding about a particular parameterization



Does it make sense to introduce this parameterization (now)?

Gradually increase the complexity of the model.

Using the same input model, try different parameterizations, then compare the results: don’t rely only on R-factors, but also check validation metrics.

The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Dorothee Liebschner, Nigel Moriarty, Billy Poon,, Oleg Sobolev

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

UTHealth

Matt Baker

University of Cambridge

Randy Read, Airlie McCoy

Duke University

Jane Richardson, Vincent Chen

An NIH/NIGMS funded Program Project

Liebschner D, *et al*., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Cryst. 2019 **D75**:861–877