Refinement with REFMAC5

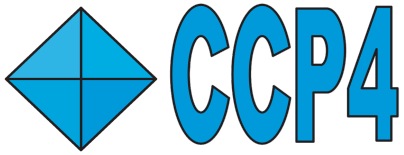
DLS-CCP4 Data Collection & Structure Solution Workshop

4th December 2022

Lucrezia Catapano
lucrezia@mrc-lmb.cam.ac.uk

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

MRC Laboratory of Molecular Biology



REFMAC5
Macromolecular structure refinement

COOT
Visualization and model building

A few key tools for refinement with CCP4

ProSMART
Restraint generation and comparative structural analysis

AceDRG
Ligand restraint dictionary and conformer generation

LibG
Nucleic acid restraint generation

LORESTR
Automated pipeline for low-resolution refinement

MRC-LMB, Cambridge:

Purpose of Refinement

Crystallographic refinement has two purposes:

- **Fit atomic model into observed X-ray crystallographic data**

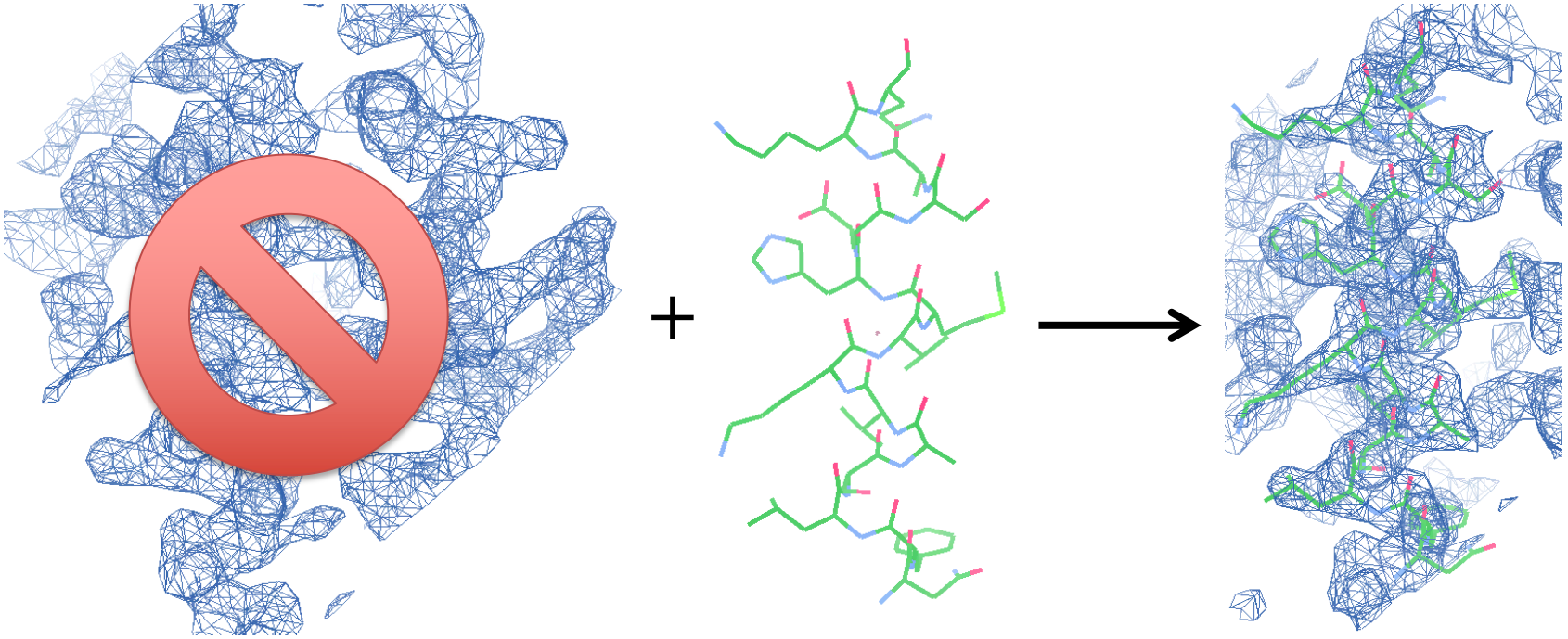
  Model should agree with the observed data

  Model must be chemically and structurally sensible

- **Calculate best possible electron density map**

  Allowing the atom model to be visualised, criticised and analysed

Purpose of Refinement



Data

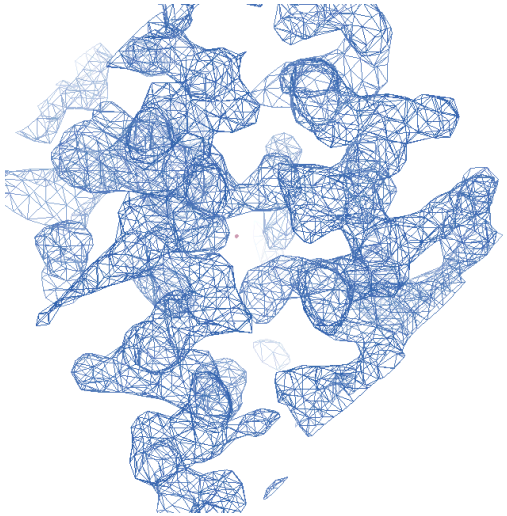
Atomic model

Fit and refine

Crystallographic Data

Different types of data:

- Amplitudes of structure factors from single crystals:
  Observed amplitudes: $|F_{obs}|$
  Estimated uncertainties: σ_{obs}



- Intensities/amplitudes from “twinned” crystals
- SAD – amplitudes available for $|F_+|$ and $|F_-|$
- Amplitudes available from multiple crystal forms

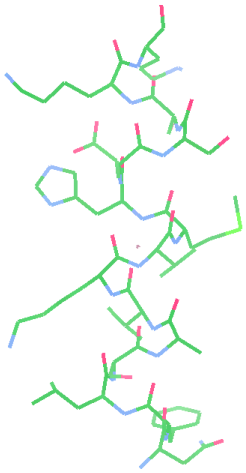
Model Refinement

We have observed amplitudes: $|F_{obs}|$

But we don't have phases: φ ←

- molecular replacement (MR)
- multiple or single anomalous diffraction methods (MAD or SAD)
- multiple or single isomorphous replacement methods (MIR and SIR)

Suppose we have a starting model:



FFT →

F_{calc} $|F_{calc}|, \varphi_{calc}$

calculated model structure factor

Idea:

Iteratively improve the model, optimising the agreement between $|F_{obs}|$ and $|F_{calc}|$

Purpose: improve phase estimates: φ_{calc}

Model Refinement

Idea:

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

Note – we are not actually refining against a density map

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

How to assess success, model quality?

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

Model Refinement

Refinement essentially tries to minimise the R-factor

How do we know that the model is reliable?

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

Such overfitting can happen if there are too many parameters

How to validate?

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

Map Calculation

REFMAC outputs coefficients for two types of maps:

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

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

Warning:

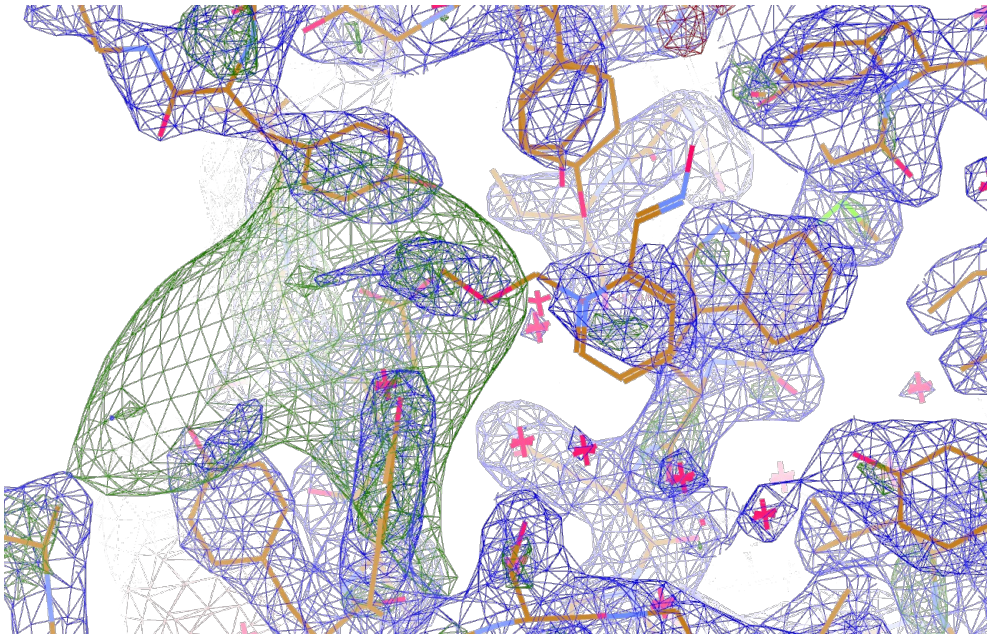
Model updated

↓

φ_{calc} updated

↓

Density changes



Note – contrast with Coot real space refinement, and REFMAC5 cryo-EM refinement

Model Refinement

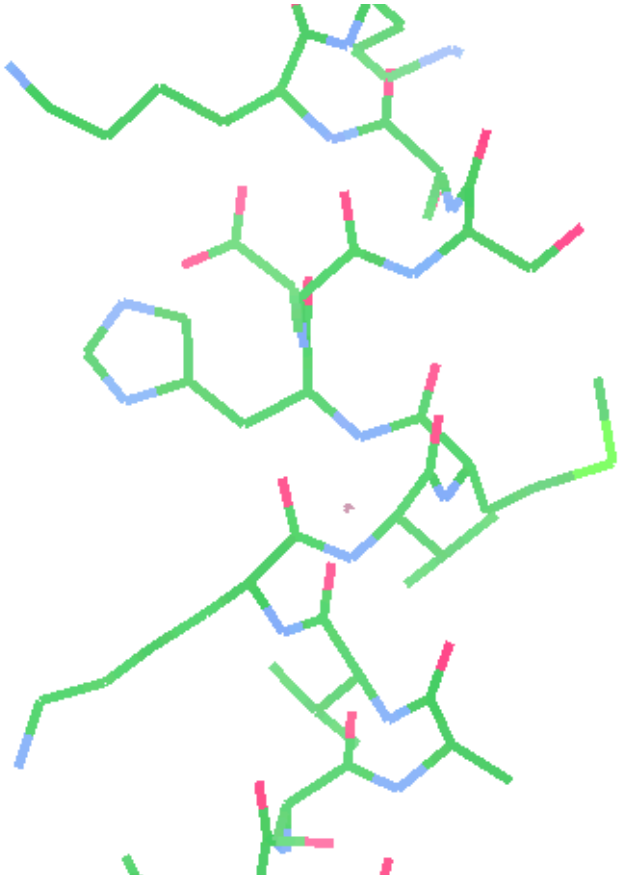
We now know:

- What sort of data we have
- How to assess model quality
- How to get phase estimates from the current model
- How to calculate electron density maps

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

Model Parameterisation

Standard refinable parameters



Atomic model:

- Position – (x,y,z) coordinates
- ADP (B or U factors)
- (Occupancies)

Overall parameters (scaling)

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

Note – different to data anisotropy (which is dealt with during data processing)

Model Parameterisation

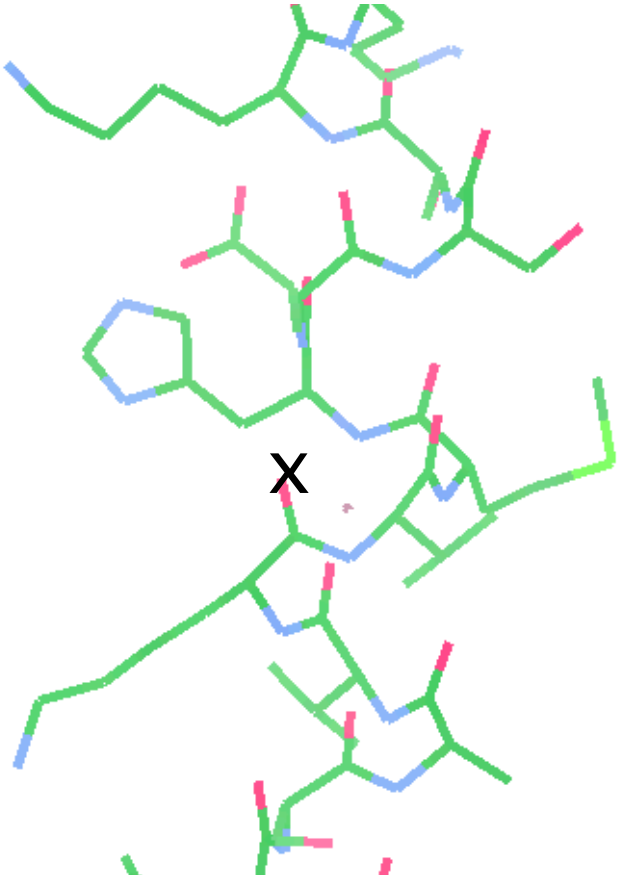
Standard refinable parameters

Atomic model:

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

Overall parameters (scaling)

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



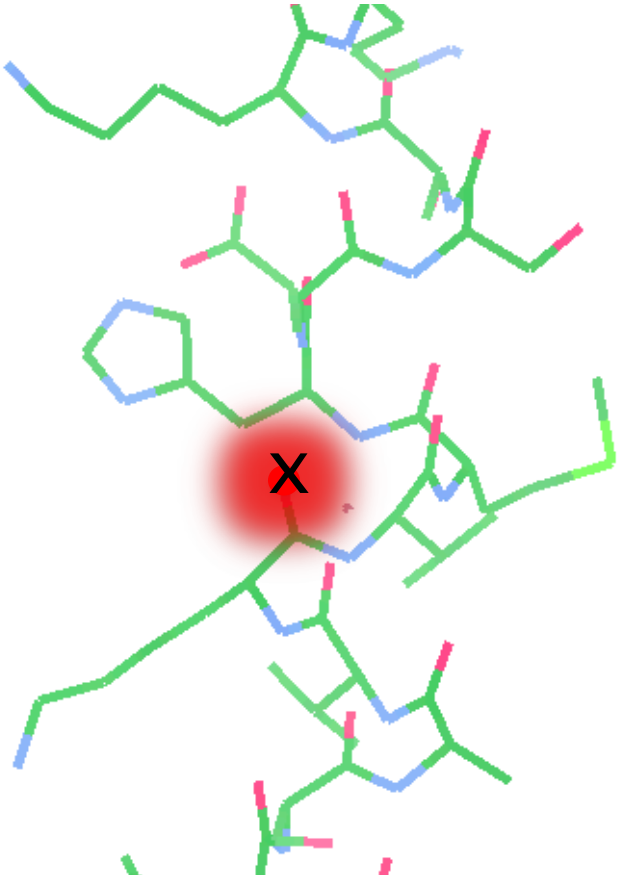
Note – different to data anisotropy (which is dealt with during data processing)

Model Parameterisation

Standard refinable parameters

Atomic model:

- Position – (x,y,z) coordinates
- **ADP (B or U factors)**
- (Occupancies)



Overall parameters (scaling)

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

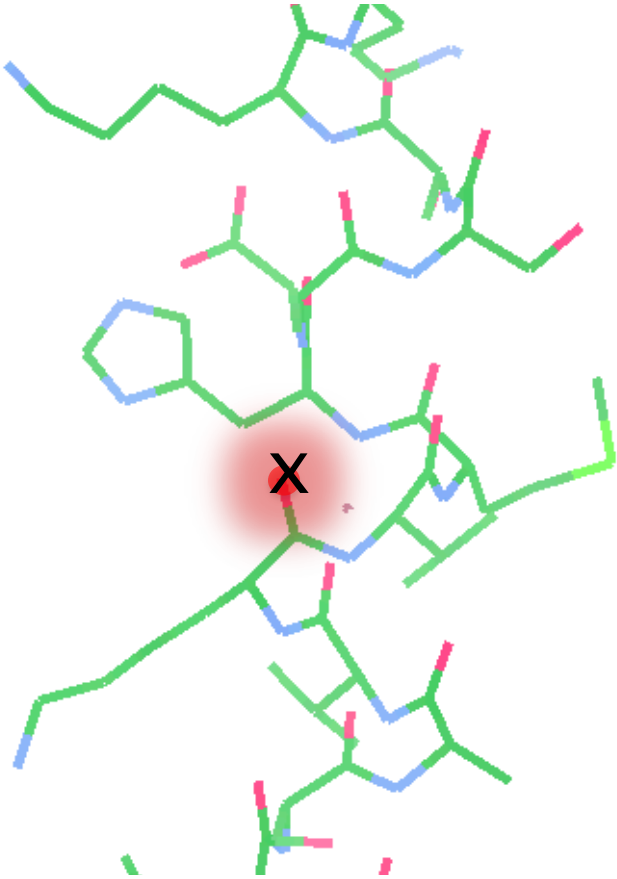
Note – different to data anisotropy (which is dealt with during data processing)

Model Parameterisation

Standard refinable parameters

Atomic model:

- Position – (x,y,z) coordinates
- ADP (B or U factors)
- **Occupancies**



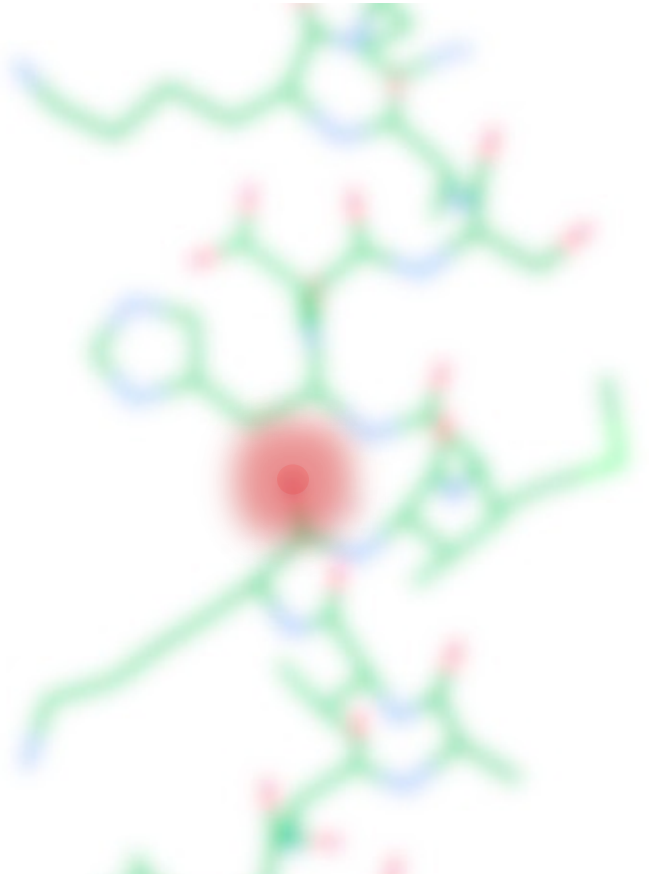
Overall parameters (scaling)

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

Note – different to data anisotropy (which is dealt with during data processing)

Model Parameterisation

Standard refinable parameters



Atomic model:

- Position – (x,y,z) coordinates
- ADP (B or U factors)
- (Occupancies)

<u>Overall parameters (scaling)</u>

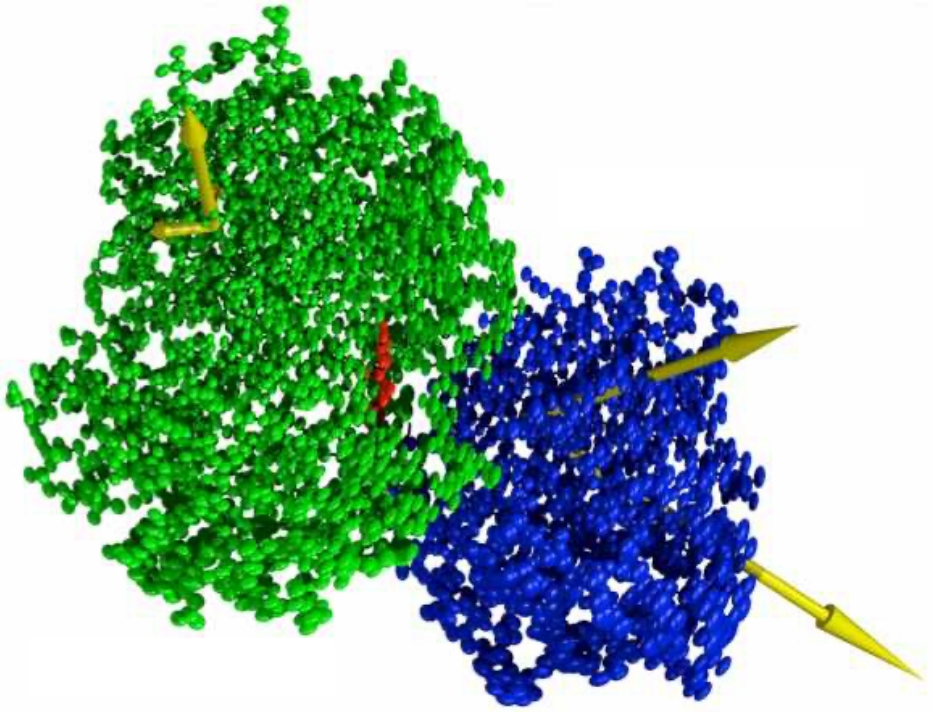
- Overall B-factor (and anisotropic U)
- Solvent treatment

Note – different to data anisotropy (which is dealt with during data processing)

TLS Groups

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

Suitable for medium/low resolutions, when full anisotropy is impossible



Per group (20 parameters):

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

Refined as a separate step

- Auto: one group per chain
- Define groups manually
- TLSMD webserver: http://skuld.bmsc.washington.edu/~tlsmd/

Model Refinement

We now have:

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

How do we refine the model?

Model Refinement

REFMAC5 uses a Maximum Likelihood approach

Crystallographic target functions have two components:

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

likelihood of the data

probability of the model

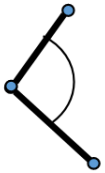
We also need prior knowledge (restraints)

These help ensure chemical and structural integrity

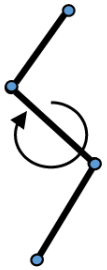
Restraints

Standard restraints (used by default) include:

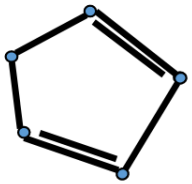
Bond lengths



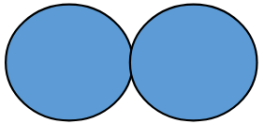
Angles



Torsion-angles



Planes



VdW repulsions

B-values

These help to ensure that the model is chemically sensible

Note – we generally deal with restraints, not constraints

Restraints

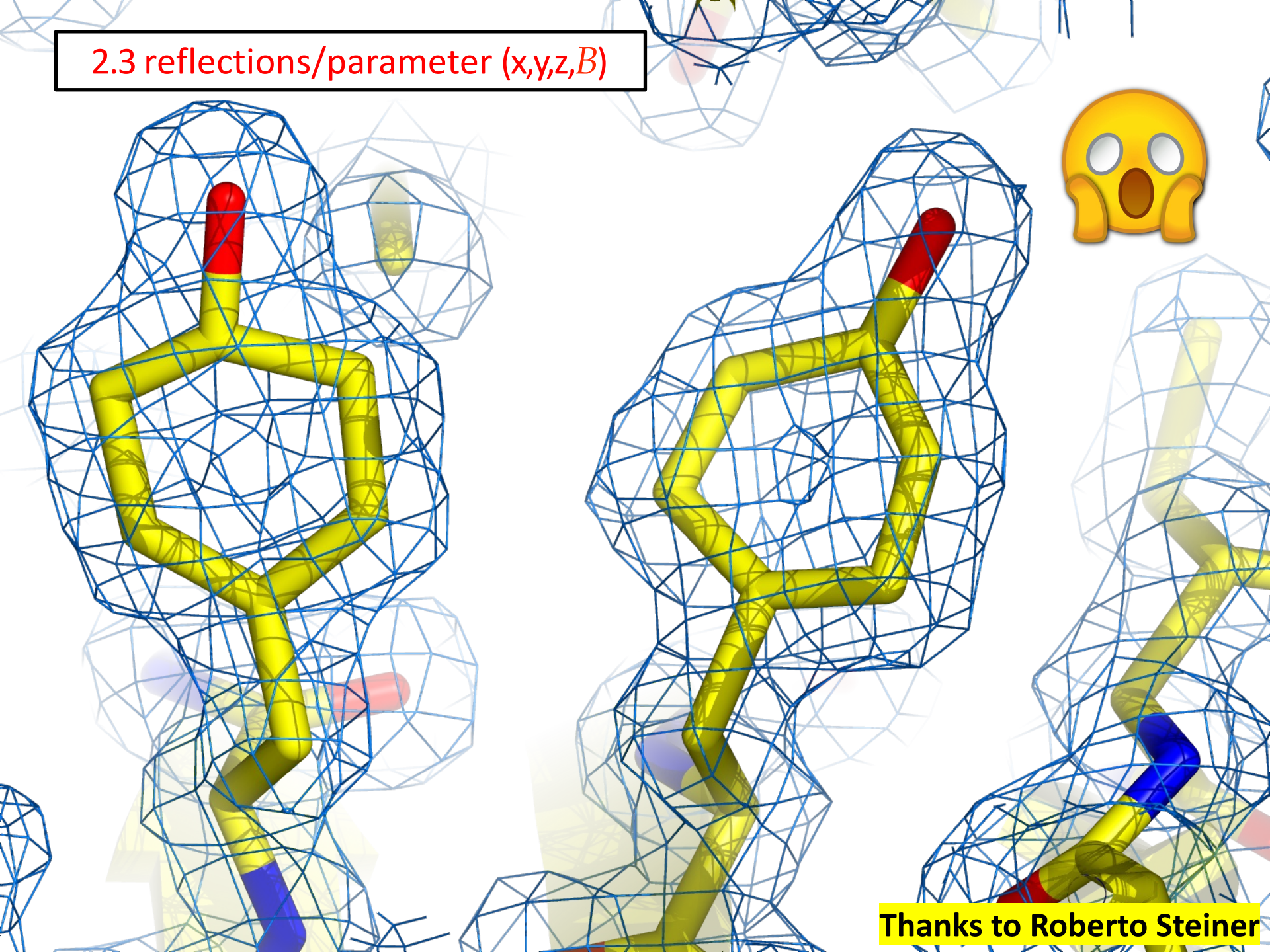
Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.

1.8 Å / 164 aa / 1540 non-H atoms / 14217 reflections

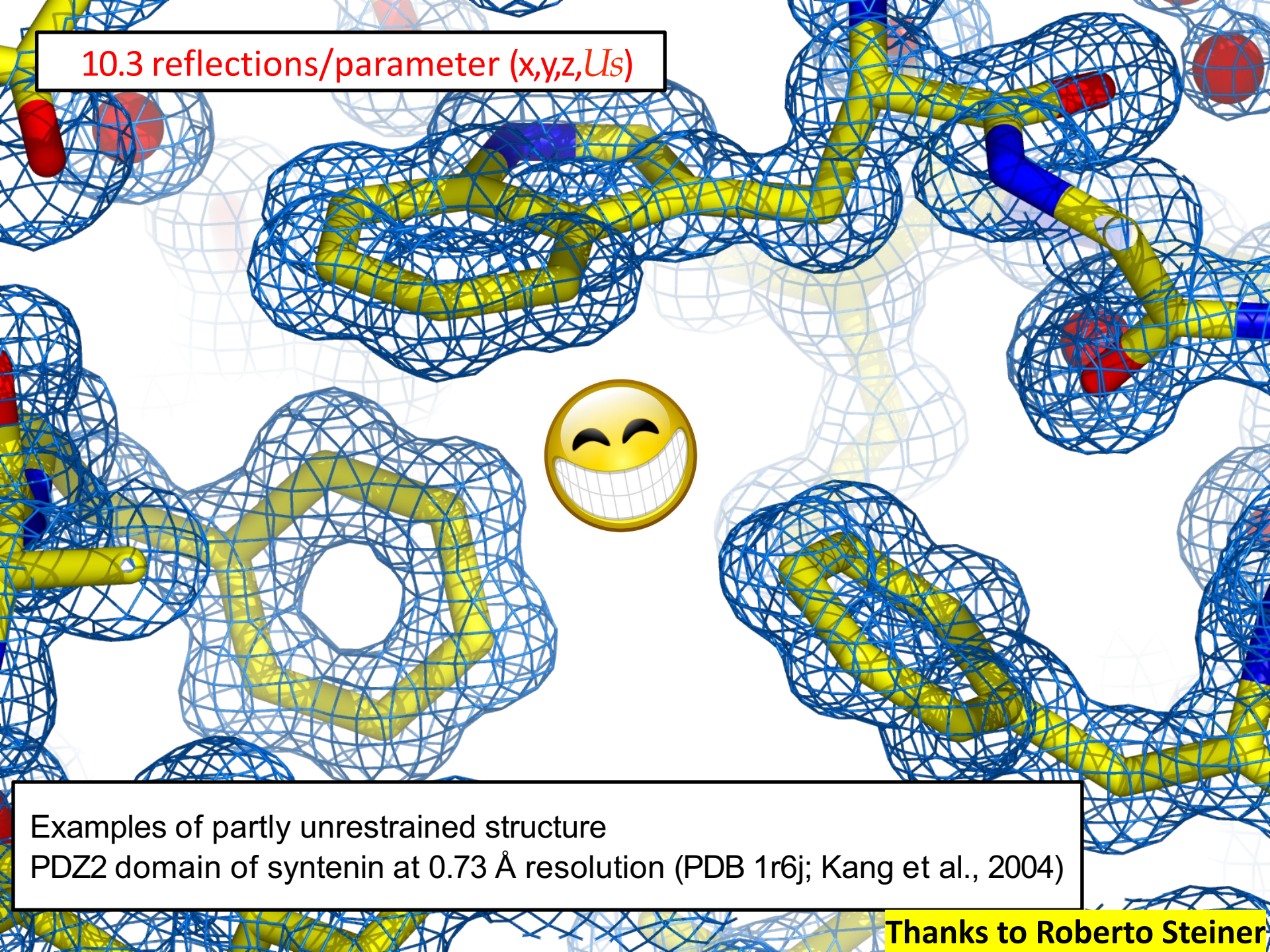
≈2.3 reflections/parameter (x,y,z,B)

2.3 reflections/parameter (x,y,z,B)



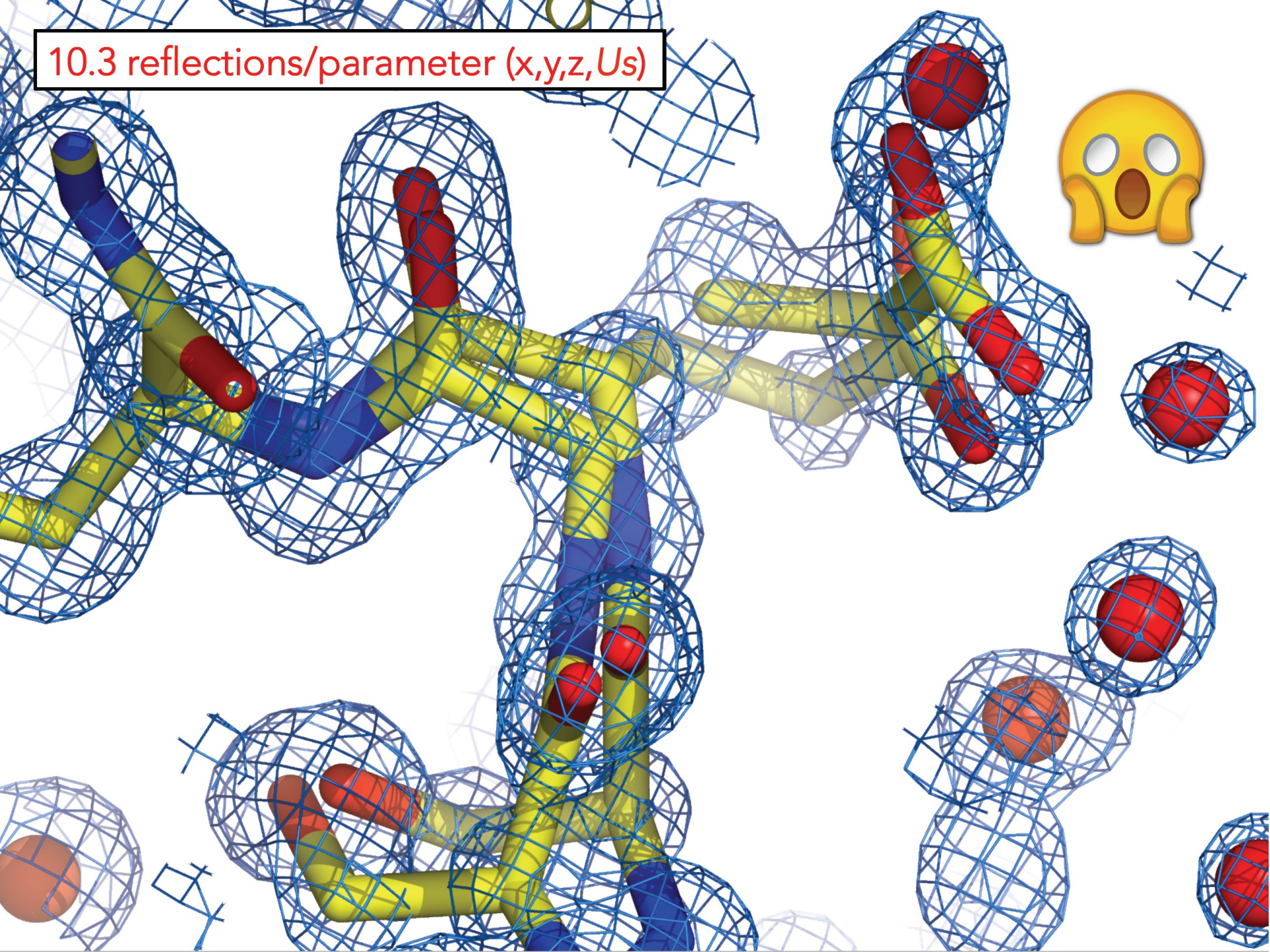
Thanks to Roberto Steiner

10.3 reflections/parameter (x,y,z,*Us*)



Examples of partly unrestrained structure
PDZ2 domain of syntenin at 0.73 Å resolution (PDB 1r6j; Kang et al., 2004)

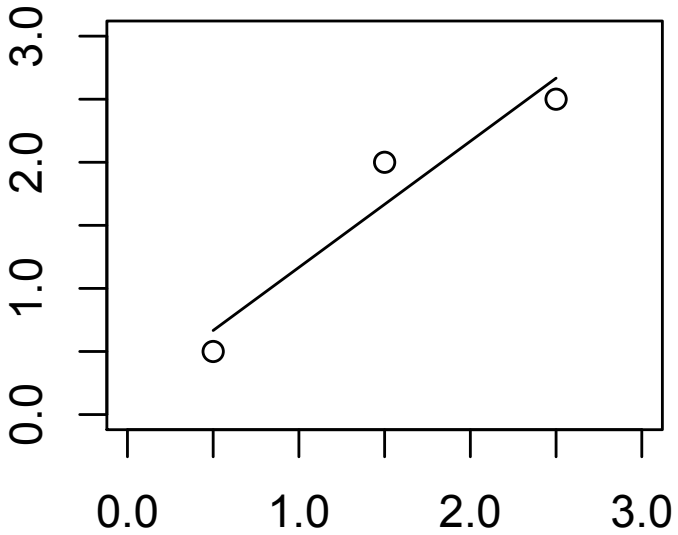
Thanks to Roberto Steiner


10.3 reflections/parameter (x,y,z,*Us*)

Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.

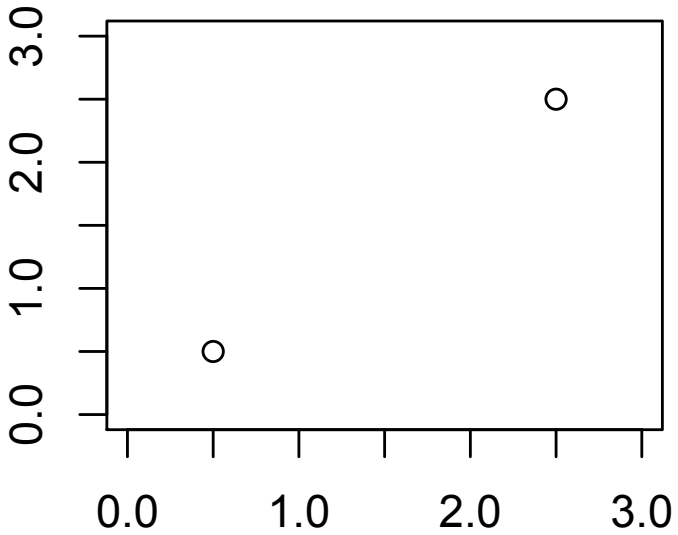


Example: Fitting a line $y = a + bx$

Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.



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

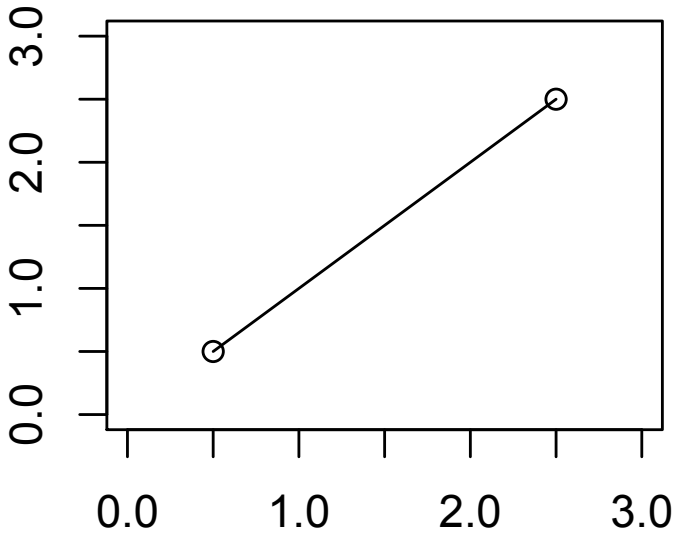
Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.

Can fit a line

Line is unreliable



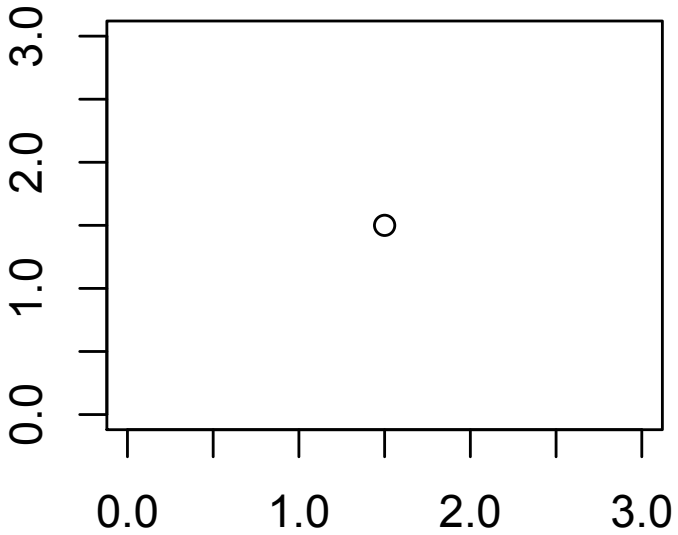
Overfitting
Model Bias

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

Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.



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

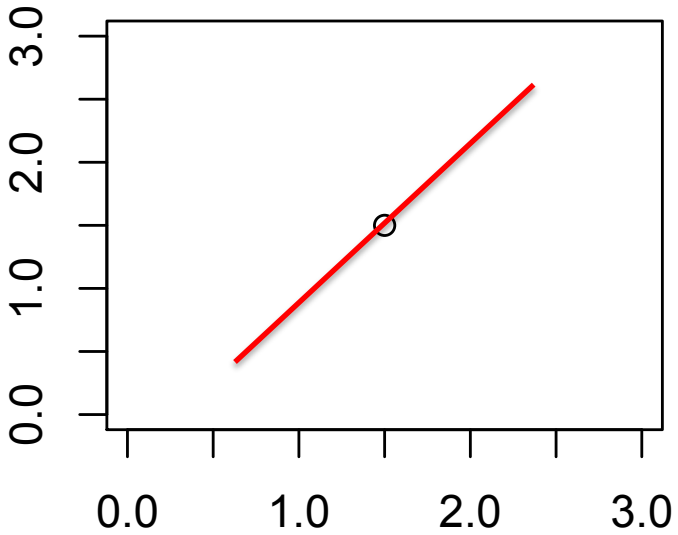
Restraints

Why introduce so many restraints?

Answer: to improve the observation:parameter ratio.

Insufficient observations!

Unstable refinement



Ill-posed problem

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

Restraints

How to improve the observation:parameter ratio.

1. Reduce number of parameters

Med-low resolution	:	Isotropic ADP	– 4 params per atom
High resolution	:	Anisotropic ADP	– 9 params per atom

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

Restraints

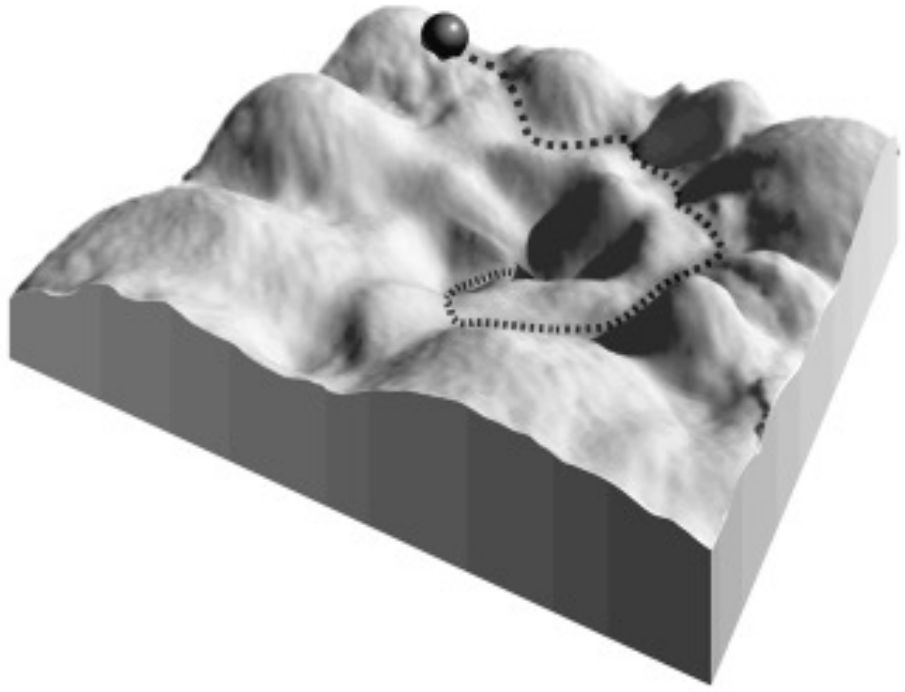
How to improve the observation:parameter ratio.

2. Increase number of restraints

Particularly useful at low-resolution:

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

‘Refinement problem’

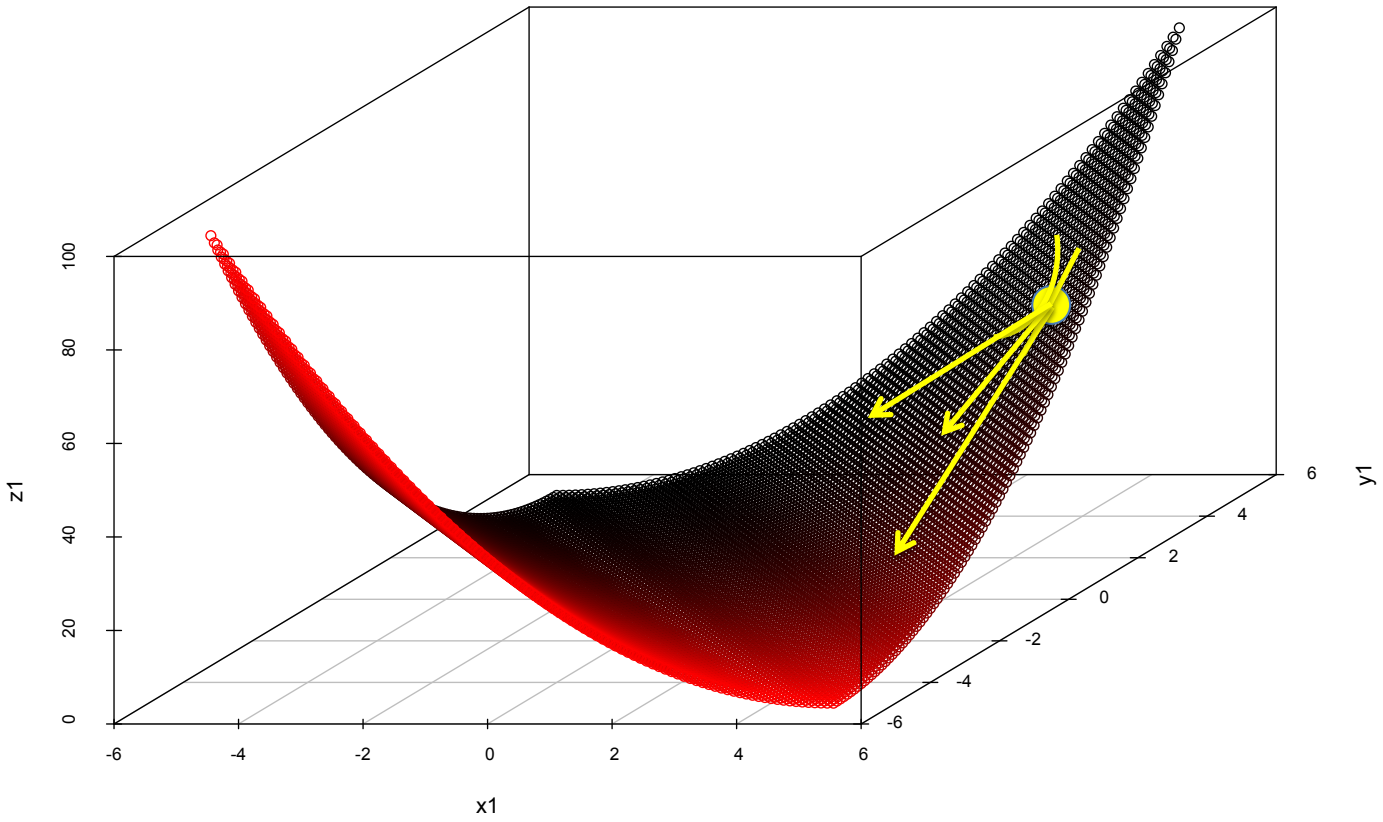


- 2D representation of refinement
- The ball represent the current set of parameters
- We want to minimize a function i.e. find the best set of parameters
- The ball could get stuck or go in the wrong way

Regularisation

Example:

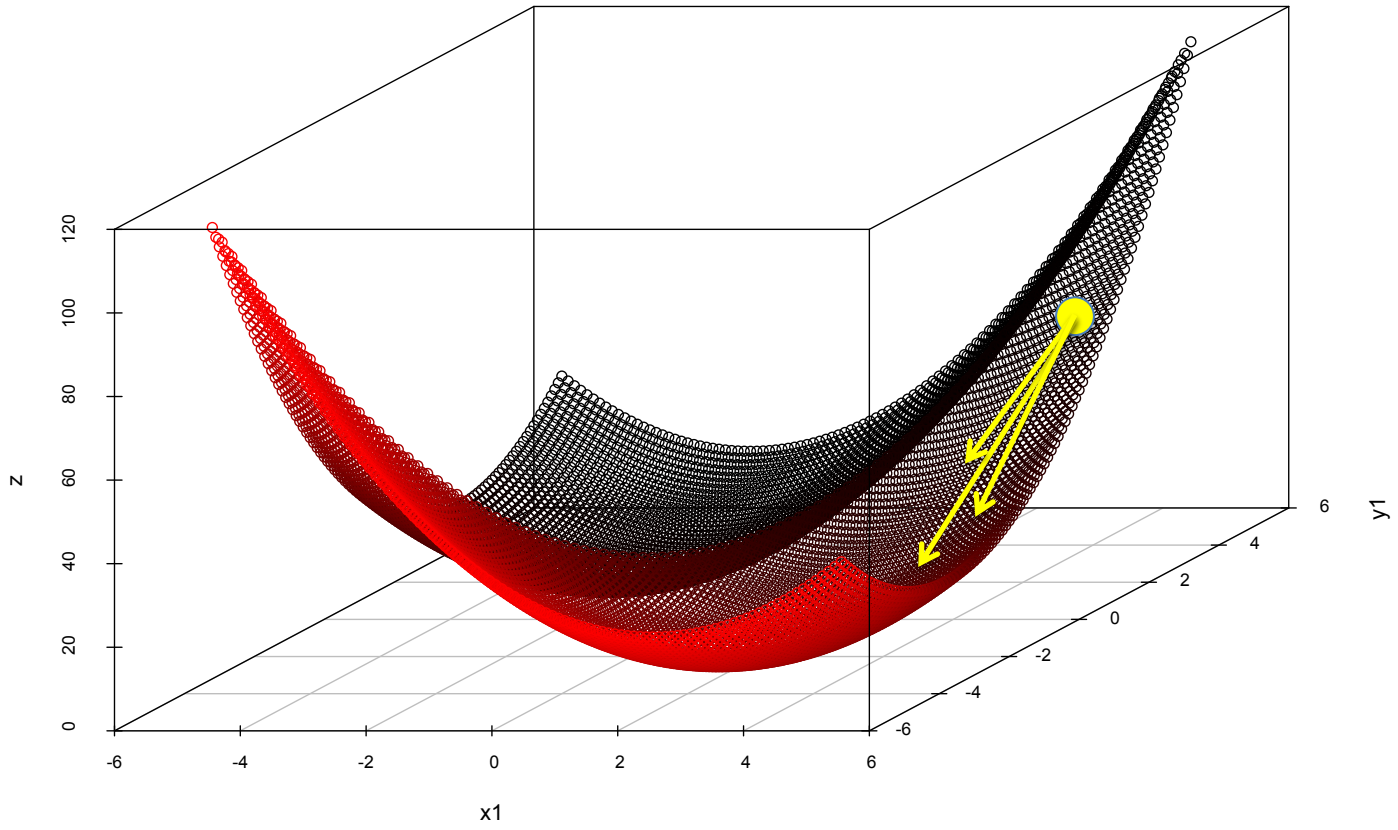
$$z = (x + y)^2$$



Regularisation

Example:

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



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

CCP4 Monomer Library

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

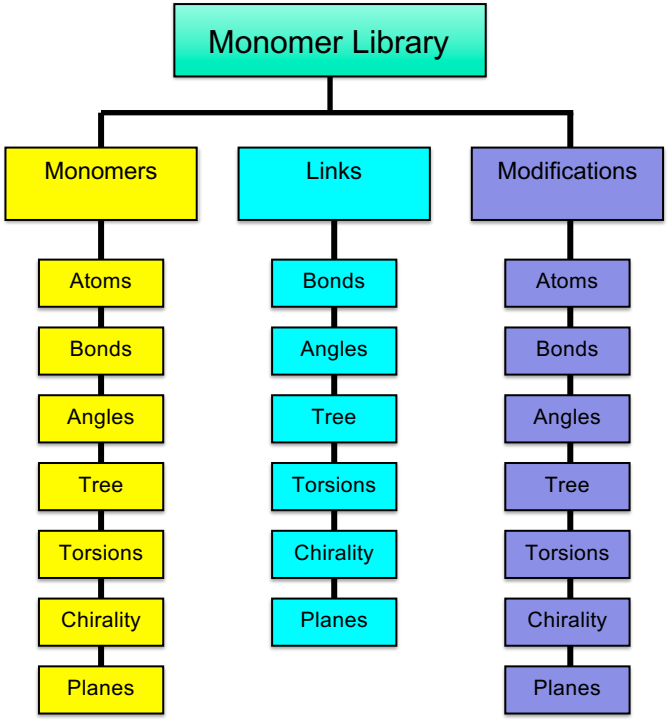
*REFMAC*5 dictionary: organization of prior chemical knowledge and guidelines for its use

Alexei A. Vagin, Roberto A. Steiner,‡ Andrey A. Lebedev, Liz Potterton, Stuart McNicholas, Fei Long and Garib N. Murshudov*

Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, England

One of the most important aspects of macromolecular structure refinement is the use of prior chemical knowledge. Bond lengths, bond angles and other chemical properties are used in restrained refinement as subsidiary conditions. This contribution describes the organization and some aspects of the use of the flexible and human/machine-readable dictionary of prior chemical knowledge used by the maximum-likelihood macromolecular-refinement program *REFMAC*5. The dictionary stores information about monomers which represent the constitutive building blocks of biological macromolecules (amino acids, nucleic acids and saccharides) and

Received 19 April 2004
Accepted 22 September 2004



- > **30,000 monomers** in the dictionary
- > **100** entries of **modifications and links**
- **CCP4-ML entries** are generated by **AceDRG**

Vagin, A. A. et al., (2004). *Acta Cryst.* D**60**, 2184-2195

CCP4 Monomer description

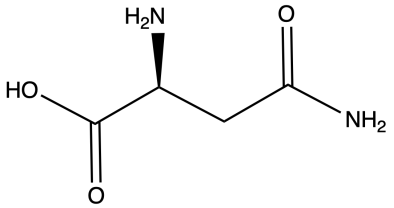
```
data_comp_list
loop_
_chem_comp.id
_chem_comp.three_letter_code
_chem_comp.name
_chem_comp.group
_chem_comp.number_atoms_all
_chem_comp.number_atoms_nh
_chem_comp.desc_level
ASN     ASN     ASPARAGINE     L-peptide     17     9     .
#
data_comp_ASN
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.charge
_chem_comp_atom.x
_chem_comp_atom.y
_chem_comp_atom.z
ASN     N       N       NT3     1       -1.744      -1.312      -0.254
ASN     H       H       H       0       -1.616      -1.398      -1.252
ASN     H2      H       H       0       -2.656      -1.069      -0.067
ASN     H3      H       H       0       -1.554      -2.155      0.164
ASN     CA      C       CH1     0       -0.804      -0.279      0.280
ASN     HA      H       H       0       -0.851      -0.300      1.265
ASN     CB      C       CH2     0       0.643       -0.579      -0.126
ASN     HB3     H       H       0       0.748       -0.392      -1.081
ASN     HB2     H       H       0       0.813       -1.533      0.010
ASN     CG      C       C       0       1.664       0.215       0.667
ASN     OD1     O       O       0       1.852       -0.042      1.855
ASN     ND2     N       NH2     0       2.326       1.179       0.041
ASN     HD21    H       H       0       2.974       1.613       0.460
ASN     HD22    H       H       0       2.122       1.406       -0.789
ASN     C       C       C       0       -1.247      1.109       -0.201
ASN     O       O       O       0       -1.284      1.385       -1.400
ASN     OXT     O       OC      -1      -1.576      1.981       0.603
loop_
_chem_comp_tree.comp_id
_chem_comp_tree.atom_id
_chem_comp_tree.atom_back
_chem_comp_tree.atom_forward
_chem_comp_tree.connect_type
 ASN      N      n/a    CA     START
 ASN      H      N      .      .
 ASN      H2     N      .      .
 ASN      H3     N      .      .
 ASN      CA     N      C      .
 ASN      HA     CA     .      .
 ASN      CB     CA     CG     .
 ASN      HB3    CB     .      .
 ASN      HB2    CB     .      .
 ASN      CG     CB     ND2    .
 ASN      OD1    CG     .      .
 ASN      ND2    CG     HD22   .
 ASN      HD21   ND2    .      .
 ASN      HD22   ND2    .      .
 ASN      C      CA     .      END
 ASN      O      C      .      .
 ASN      OXT    C      .      .
```

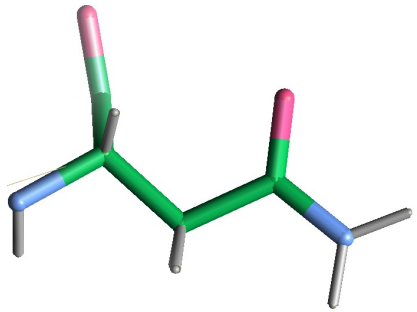
General category

Atom category

Tree category

ASN





CCP4 Monomer description

General category

```
data_comp_list
loop_
_chem_comp.id
_chem_comp.three_letter_code
_chem_comp.name
_chem_comp.group
_chem_comp.number_atoms_all
_chem_comp.number_atoms_nh
_chem_comp.desc_level
ASN     ASN     ASPARAGINE     L-peptide     17     9     .
#
data_comp_ASN
#
```

Atom category

```
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.charge
_chem_comp_atom.x
_chem_comp_atom.y
_chem_comp_atom.z
ASN     N       N       NT3     1       -1.744      -1.312      -0.254
ASN     H       H       H       0       -1.616      -1.398      -1.252
ASN     H2      H       H       0       -2.656      -1.069      -0.067
ASN     H3      H       H       0       -1.554      -2.155       0.164
ASN     CA      C       CH1     0       -0.804      -0.279       0.280
ASN     HA      H       H       0       -0.851      -0.300       1.265
ASN     CB      C       CH2     0        0.643      -0.579      -0.126
ASN     HB3     H       H       0        0.748      -0.392      -1.081
ASN     HB2     H       H       0        0.813      -1.533       0.010
ASN     CG      C       C       0        1.664       0.215       0.667
ASN     OD1     O       O       0        1.852      -0.042       1.855
ASN     ND2     N       NH2     0        2.326       1.179       0.041
ASN     HD21    H       H       0        2.974       1.613       0.460
ASN     HD22    H       H       0        2.122       1.406      -0.789
ASN     C       C       C       0       -1.247       1.109      -0.201
ASN     O       O       O       0       -1.284       1.385      -1.400
ASN     OXT     O       OC      -1      -1.576       1.981       0.603
```

Tree category

```
loop_
_chem_comp_tree.comp_id
_chem_comp_tree.atom_id
_chem_comp_tree.atom_back
_chem_comp_tree.atom_forward
_chem_comp_tree.connect_type
 ASN      N      n/a     CA     START
 ASN      H      N       .      .
 ASN      H2     N       .      .
 ASN      H3     N       .      .
 ASN      CA     N       C      .
 ASN      HA     CA      .      .
 ASN      CB     CA      CG     .
 ASN      HB3    CB      .      .
 ASN      HB2    CB      .      .
 ASN      CG     CB      ND2    .
 ASN      OD1    CG      .      .
 ASN      ND2    CG      HD22   .
 ASN      HD21   ND2     .      .
 ASN      HD22   ND2     .      .
 ASN      C      CA      .      END
 ASN      O      C       .      .
 ASN      OXT    C       .      .
```

Bond category

```
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.aromatic
_chem_comp_bond.value_dist_nucleus
_chem_comp_bond.value_dist_nucleus_esd
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
ASN       CB        CA        SINGLE    n         1.531     0.011     1.531     0.011
ASN       CA        C         SINGLE    n         1.533     0.011     1.533     0.011
ASN       C         O         DOUBLE    n         1.247     0.019     1.247     0.019
ASN       C         OXT       SINGLE    n         1.247     0.019     1.247     0.019
ASN       CA        N         SINGLE    n         1.488     0.010     1.488     0.010
ASN       CB        CG        SINGLE    n         1.514     0.010     1.514     0.010
ASN       CG        OD1       DOUBLE    n         1.229     0.012     1.229     0.012
ASN       CG        ND2       SINGLE    n         1.323     0.011     1.323     0.011
ASN       CB        HB3       SINGLE    n         1.089     0.010     0.979     0.016
ASN       CB        HB2       SINGLE    n         1.089     0.010     0.979     0.016
ASN       CA        HA        SINGLE    n         1.089     0.010     0.986     0.020
ASN       N         H         SINGLE    n         1.036     0.016     0.911     0.020
```

Angle category

```
_chem_comp_angle.atom_id_2
_chem_comp_angle.atom_id_3
_chem_comp_angle.value_angle
_chem_comp_angle.value_angle_esd
ASN          CA           CB           CG      112.981    1.50
ASN          CA           CB          HB3      108.904    1.50
ASN          CA           CB          HB2      108.904    1.50
ASN          CG           CB          HB3      109.076    1.50
ASN          CG           CB          HB2      109.076    1.50
ASN         HB3           CB          HB2      108.069    1.50
ASN          CB           CA            C      111.540    2.60
ASN          CB           CA            N      111.766    1.50
ASN          CB           CA           HA      107.983    1.50
ASN           C           CA            N      109.241    1.50
```

Torsion-angle category

```
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 ASN      chi1     N       CA      CB      CG       180.000   15.000   3
 ASN      chi2     CA      CB      CG      ND2        0.000   30.000   2
 ASN      hh1      CB      CG      ND2     HD22       0.000   30.000   2
ASN             sp2_sp3_1          O           C           CA          CB       0.000   10.00     6
ASN           sp3_sp3_10          CB          CA           N           H      180.000   10.00     3
```

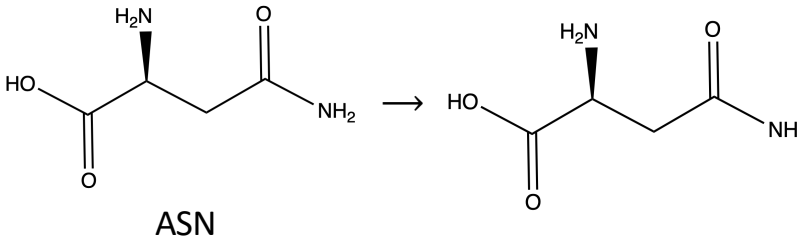
Chirality category

```
loop_
_chem_comp_chir.comp_id
_chem_comp_chir.id
_chem_comp_chir.atom_id_centre
_chem_comp_chir.atom_id_1
_chem_comp_chir.atom_id_2
_chem_comp_chir.atom_id_3
_chem_comp_chir.volume_sign
ASN    chir_1    CA    N    C    CB    positive
```

Plane category

```
loop_
_chem_comp_plane_atom.comp_id
_chem_comp_plane_atom.plane_id
_chem_comp_plane_atom.atom_id
_chem_comp_plane_atom.dist_esd
ASN    plan-1           C   0.020
ASN    plan-1          CA   0.020
ASN    plan-1           O   0.020
ASN    plan-1         OXT   0.020
```

CCP4 Modification description



ASN

```
data_mod_ASNm1

loop_
_chem_mod_atom.mod_id
_chem_mod_atom.function
_chem_mod_atom.atom_id
_chem_mod_atom.new_atom_id
_chem_mod_atom.new_type_symbol
_chem_mod_atom.new_type_energy
_chem_mod_atom.new_charge
ASNm1          delete         HD22       .          H          H          0
ASNm1          change         ND2        .          N          NH1        0

loop_
_chem_mod_bond.mod_id
_chem_mod_bond.function
_chem_mod_bond.atom_id_1
_chem_mod_bond.atom_id_2
_chem_mod_bond.new_type
_chem_mod_bond.new_value_dist
_chem_mod_bond.new_value_dist_esd
ASNm1          delete         ND2        HD22       single          .              .
ASNm1          change         CG         ND2        single          1.340          0.0100
ASNm1          change         ND2        HD21       single          0.895          0.0200

loop_
_chem_mod_angle.mod_id
_chem_mod_angle.function
_chem_mod_angle.atom_id_1
_chem_mod_angle.atom_id_2
_chem_mod_angle.atom_id_3
_chem_mod_angle.new_value_angle
_chem_mod_angle.new_value_angle_esd
ASNm1          delete         CG         ND2        HD22       .              .
ASNm1          delete         HD21       ND2        HD22       .              .
ASNm1          change         CG         ND2        HD21       118.696        2.17

loop_
_chem_mod_plane.mod_id
_chem_mod_plane.function
_chem_mod_plane.plane_id
_chem_mod_plane.atom_id
_chem_mod_plane.dist_esd
ASNm1          delete         plan-3         CG             0.020
ASNm1          delete         plan-3         HD21           0.020
ASNm1          delete         plan-3         HD22           0.020
ASNm1          delete         plan-3         ND2            0.020
```

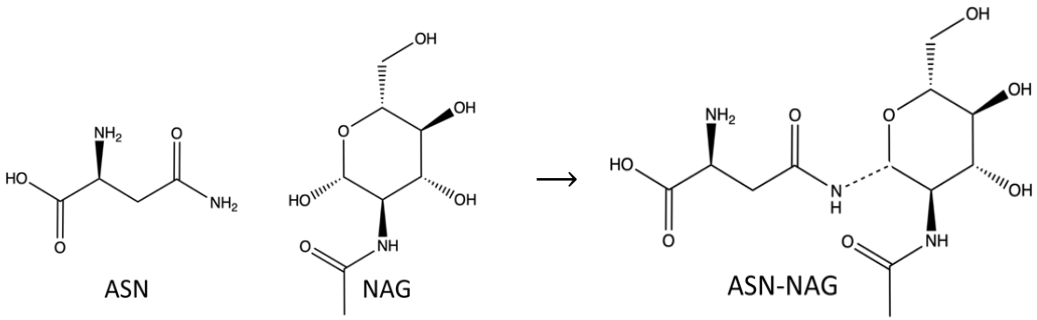
Atom category

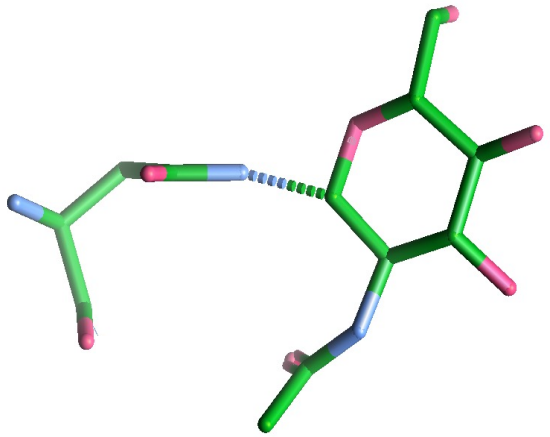
Bond category

Angle category

Plane category

CCP4 Link description





General category

```
loop_
_chem_link.id
_chem_link.comp_id_1
_chem_link.mod_id_1
_chem_link.group_comp_1
_chem_link.comp_id_2
_chem_link.mod_id_2
_chem_link.group_comp_2
_chem_link.name
NAG-ASN . DEL-O1 pyranose ASN DEL-HD22 . bond_NAG-C1_=_ASN-ND2
```

Bond category

```
data_link_NAG-ASN
loop_
_chem_link_bond.link_id
_chem_link_bond.atom_1_comp_id
_chem_link_bond.atom_id_1
_chem_link_bond.atom_2_comp_id
_chem_link_bond.atom_id_2
_chem_link_bond.type
_chem_link_bond.value_dist
_chem_link_bond.value_dist_esd
NAG-ASN 1 C1 2 ND2 single 1.439 .020
```

Angle category

```
loop_
_chem_link_angle.link_id
_chem_link_angle.atom_1_comp_id
_chem_link_angle.atom_id_1
_chem_link_angle.atom_2_comp_id
_chem_link_angle.atom_id_2
_chem_link_angle.atom_3_comp_id
_chem_link_angle.atom_id_3
_chem_link_angle.value_angle
_chem_link_angle.value_angle_esd
NAG-ASN 1 C1 2 ND2 2 CG 121.000 3.000
NAG-ASN 1 C1 2 ND2 2 HD2 119.000 3.000
NAG-ASN 1 O5 1 C1 2 ND2 112.300 3.000
NAG-ASN 1 C2 1 C1 2 ND2 112.700 3.000
```

Plane category

```
loop_
_chem_link_plane.link_id
_chem_link_plane.plane_id
_chem_link_plane.atom_comp_id
_chem_link_plane.atom_id
_chem_link_plane.dist_esd
NAG-ASN plane1 1 C1 .020
NAG-ASN plane1 2 ND2 .020
NAG-ASN plane1 2 CG .020
NAG-ASN plane1 2 OD1 .020
NAG-ASN plane1 2 CB .020
NAG-ASN plane1 2 HD21 .020
```

Nicholls, R. A. et al. (2021). *Acta Cryst.* D**77**, 712-726

Restraints for Ligands

Geometric restraints for protein / nucleic acids are pre-tabulated

- **Common/known structures are dealt with automatically**
  - CCP4/REFMAC monomer library has pre-computed descriptions

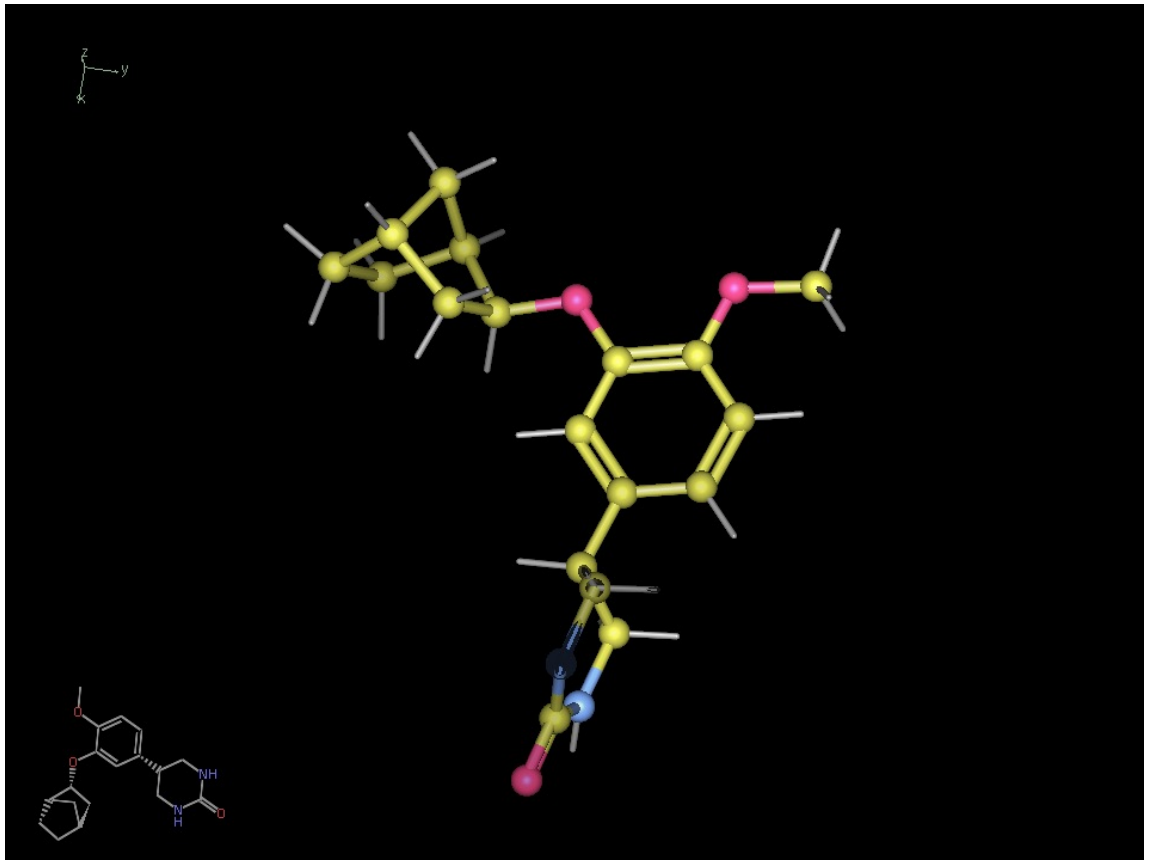
- **New ligands require description (CIF file)**
  - AceDRG

AceDRG

New atom types

Full 2nd order neighbour-based atom description

(3rd order in some cases)



AceDRG derives atom types from small molecular databases
These are tabulated, distributed as part of CCP4 for quick lookup

AceDRG

Functionalities:

(1) Restraint Dictionary Generator

- Uses restraint tables to generate restraints for given molecule
- Inputs – mmCIF, SMILES, MDL/SDF, SYBIL/MOL2
- Output – CIF – bond lengths, angles, torsions, planes, chiralities

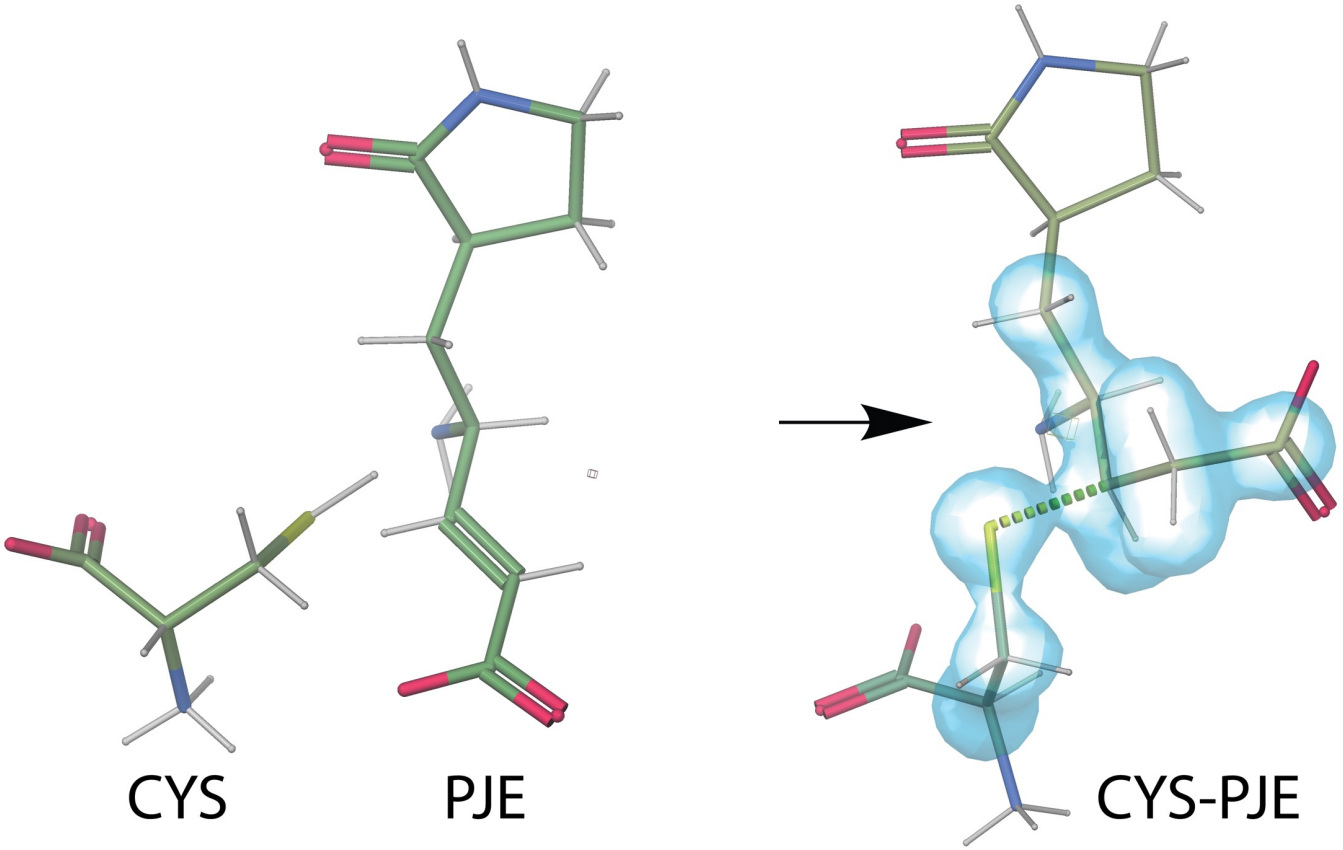
(2) Conformer Generator

- Generates coordinates from graph-based molecule description
- Generates one of the lower-energy conformations
- Refines conformation using Refmac
- Output – PDB

(3) **Link Creation**

- Creates link between two components
- May be from the CCP4 monomer library, or custom CIFs
- Separate operational mode – requires separate execution

Modelling Covalent Linkages



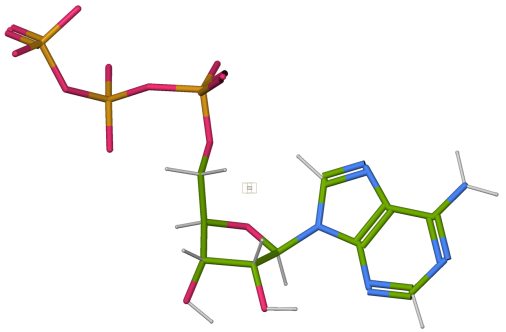
- Link between Cys[SG] and PJE[C20]
- PJE[C20]–[C21] bond order changed from double to single
- Various restraints in the local environment are changed – modifications

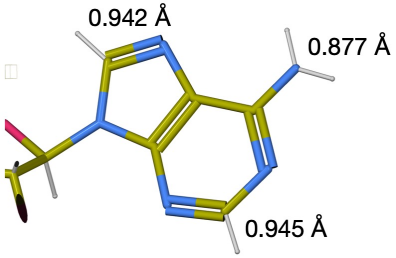
Inclusion of bond lengths to the H atom nuclei

```
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.aromatic
_chem_comp_bond.value_dist_nucleus
_chem_comp_bond.value_dist_nucleus_esd
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
```

new

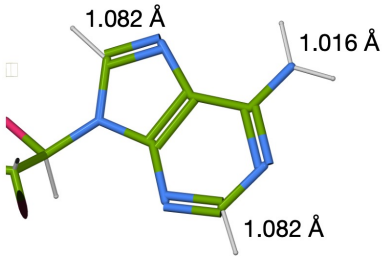
ATP	PG	O1G	DOUBLE	n	1.509	0.0200	1.509	0.0200
ATP	PG	O2G	SINGLE	n	1.509	0.0200	1.509	0.0200
ATP	PG	O3G	SINGLE	n	1.509	0.0200	1.509	0.0200
ATP	PG	O3B	SINGLE	n	1.614	0.0178	1.614	0.0178
ATP	PB	O1B	DOUBLE	n	1.493	0.0157	1.493	0.0157
ATP	PB	O2B	SINGLE	n	1.493	0.0157	1.493	0.0157
ATP	PB	O3B	SINGLE	n	1.601	0.0114	1.601	0.0114
ATP	PB	O3A	SINGLE	n	1.601	0.0114	1.601	0.0114
ATP	PA	O1A	DOUBLE	n	1.493	0.0122	1.493	0.0122
ATP	PA	O2A	SINGLE	n	1.493	0.0122	1.493	0.0122
ATP	PA	O3A	SINGLE	n	1.604	0.0133	1.604	0.0133
ATP	PA	"O5'"	SINGLE	n	1.604	0.0133	1.604	0.0133
ATP	"O5'"	"C5'"	SINGLE	n	1.450	0.0166	1.450	0.0166
ATP	"C5'"	"C4'"	SINGLE	n	1.509	0.0100	1.509	0.0100
ATP	"C4'"	"O4'"	SINGLE	n	1.451	0.0100	1.451	0.0100
ATP	"C4'"	"C3'"	SINGLE	n	1.535	0.0100	1.535	0.0100
ATP	"O4'"	"C1'"	SINGLE	n	1.409	0.0100	1.409	0.0100
ATP	"C3'"	"O3'"	SINGLE	n	1.422	0.0100	1.422	0.0100
ATP	"C3'"	"C2'"	SINGLE	n	1.531	0.0100	1.531	0.0100
ATP	"C2'"	"O2'"	SINGLE	n	1.411	0.0100	1.411	0.0100
ATP	"C2'"	"C1'"	SINGLE	n	1.525	0.0100	1.525	0.0100
ATP	"C1'"	N9	SINGLE	n	1.458	0.0200	1.458	0.0200
ATP	N9	C8	SINGLE	y	1.372	0.0200	1.372	0.0200
ATP	N9	C4	SINGLE	y	1.372	0.0100	1.372	0.0100
ATP	C8	N7	DOUBLE	y	1.310	0.0100	1.310	0.0100
ATP	N7	C5	SINGLE	y	1.388	0.0100	1.388	0.0100
ATP	C5	C6	SINGLE	y	1.408	0.0100	1.408	0.0100
ATP	C5	C4	DOUBLE	y	1.381	0.0100	1.381	0.0100
ATP	C6	N6	SINGLE	n	1.330	0.0100	1.330	0.0100
ATP	C6	N1	DOUBLE	y	1.354	0.0100	1.354	0.0100
ATP	N1	C2	SINGLE	y	1.339	0.0100	1.339	0.0100
ATP	C2	N3	DOUBLE	y	1.330	0.0100	1.330	0.0100
ATP	N3	C4	SINGLE	y	1.343	0.0100	1.343	0.0100
ATP	"C5'"	"H5'1"	SINGLE	n	1.089	0.0100	0.989	0.0200
ATP	"C5'"	"H5'2"	SINGLE	n	1.089	0.0100	0.989	0.0200
ATP	"C4'"	"H4'"	SINGLE	n	1.089	0.0100	0.981	0.0200
ATP	"C3'"	"H3'"	SINGLE	n	1.089	0.0100	0.992	0.0200
ATP	"O3'"	"HO3'"	SINGLE	n	0.970	0.0120	0.849	0.0200
ATP	"C2'"	"H2'"	SINGLE	n	1.089	0.0100	0.994	0.0200
ATP	"O2'"	"HO2'"	SINGLE	n	0.970	0.0120	0.849	0.0200
ATP	"C1'"	"H1'"	SINGLE	n	1.089	0.0100	0.984	0.0200
ATP	C8	H8	SINGLE	n	1.082	0.0130	0.942	0.0170
ATP	N6	HN61	SINGLE	n	1.016	0.0100	0.877	0.0200
ATP	N6	HN62	SINGLE	n	1.016	0.0100	0.877	0.0200
ATP	C2	H2	SINGLE	n	1.082	0.0130	0.945	0.0200





H electron positions

(H—X bond lengths from MX)



H nucleus positions

(H—X bond lengths from NMX/QMcalc)

New linkage entries

Category	Linkage ID	Monomer 1	Atom 1	Monomer 2	Atom 2	# Annotated	# Unannotated	Example
Lysine linkages	LYS-CYS	LYS	NZ	CYS	SG	0	354	6h27 : B73-70
	LYS-ASN	LYS	NZ	ASN	CG	139	82	4oq1 : A264-354
	LYS-RET	LYS	NZ	RET	C15	383	17	5u6g : G108-201
	LYS-PLP	LYS	NZ	PLP	C4A	1,264	17	3b8w : A34-1001
	pept-LYS	peptide	C	LYS	NZ	101	119	5emz : A76-B48
Histidine linkages	HIS_TYR1	HIS	ND1	TYR	CB	61	60	4enu : C392-415
	HIS_TYR2	HIS	NE2	TYR	CE2	68	64	3abk : A240-244
	HIS-FAD1	HIS	ND1	FAD	C8M	121	15	1w1s : A105-1535
	HIS-FAD2	HIS	NE2	FAD	C8M	171	19	6b58 : A44-601
Tyrosine linkages	MET-TYR	MET	SD	TYR	CE1	20	29	5whs : A264-238
	TRP-TYR	TRP	CH2	TYR	CE2	10	17	5whq : A90-238
Cysteine linkages	CYS-CYC	CYS	SG	CYC	CAC	326	48	4xxi : A196-301
	CYS-PEB	CYS	SG	PEB	CAA	215	57	3v58 : B158-204
	CYS-FAD	CYS	SG	FAD	C8M	150	23	5zao : A343-501
Disulphide bridges	Ddisul	DCY	SG	DCY	SG	134	0	5e5t : B29-62
	CYS-BME	CYS	SG	BME	S2	437	201	3cav : A148-329
Peptide to peptide-linking	pept-GYC	peptide	C	GYC	N	75	14	3ls3 : B61-62
	pept-CR8	peptide	C	CR8	N	115	2	3s05 : C61-64
	pept-MDO	peptide	C	MDO	N	122	0	3kdy : B151-152
	pept-CRO	peptide	C	CRO	N1	314	3	5mak : D64-66
	pept-NRQ	peptide	C	NRQ	N1	131	0	4h3l : A65-66
	pep-NH2	peptide	C	NH2	N	1,097	31	6q5p : F30-31
Peptide-linking to peptide	GYC-pept	GYC	C	peptide	N	89	0	6nqj : C63-65
	CR8-pept	CR8	C	peptide	N	118	0	4ljd : C64-65
	MDO-pept	MDO	C	peptide	N	119	1	6hqf : A205-203
	CRO-pept	CRO	C3	peptide	N	315	5	5mak : B66-68
	NRQ-pept	NRQ	C3	peptide	N	149	2	3nt3 : C63-66
	IAS-pept	IAS	CG	peptide	N	126	0	1dy5 : A67-68
Glycosidic and DNA/RNA linkages	ALPHA2-6	ketopyranose	C2	pyranose	O6	132	3	4fqc : H310-308
	GTP-p	GTP	O3'	dna/rna	P	237	1	4gv9 : E8-9
N3 inhibitor linkages	02J-ALA	02J	C41	ALA	N	23	0	
	PJE-010	PJE	C22	010	O	23	0	
	PJE-CYS	PJE	C20	CYS	SG	8	12	
	PJE-LEU	PJE	N5	LEU	C	23	0	

- Identified by Gemmi algo
- 34 links and 47 modifications added by using **AceDRG**

Nicholls, R. A. et al., (2021). *Acta Cryst.* D**77**, 712-726
Nicholls, R. A. et al., (2021). *Acta Cryst.* D**77**, 727-745

Regularisation

Use of available knowledge (prior information):

High-low resolution:

- Geometry restraints (chemical information)

Medium-low resolution:

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

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

- External restraints

Regularisation

Use of available knowledge (prior information):

High-low resolution:

- **Geometry restraints (chemical information)**

Medium-low resolution:

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

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

- **External restraints**

Regularisers with a target value

Regularisation

Use of available knowledge (prior information):

High-low resolution:

- Geometry restraints (chemical information)

Medium-low resolution:

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

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

- External restraints

Regularisers without an external target value

Jelly Body Restraints

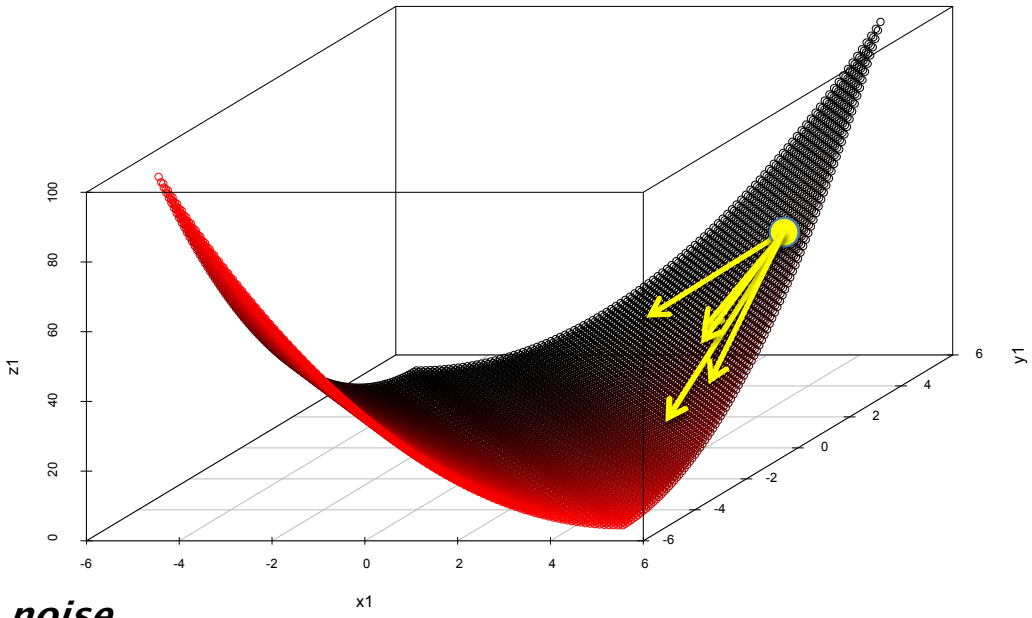
Regularisers without a target:

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

d : interatomic distance

d_{current} : current interatomic distance

σ : restraint standard deviation



Model should be less prone to fitting into noise

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

Typical: σ = 0.01–0.02

Distance threshold: 4.2Å