Chemical Crystallography and Structural Chemistry

(VO 270287)

Lecture 7

14th May 2020

Dr. Tim Grüne
Centre for X-ray Structure Analysis
Faculty of Chemistry
University of Vienna

tim.gruene@univie.ac.at

Previous Lecture

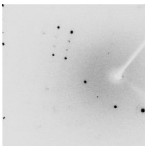
1. Phasing: Solving a crytal structre
2. direct methods
3. Patterson map

Today's Lecture

1. Model building
2. Refinement
3. Constraints & Restraints

Refinement: chemically sensible structure

Data collection ➡ **Data integration** ➡ **Data Scaling**



```
0 0 -1      2.7  0.9
0 0  1      4.0  1.0
0 0 -2 1'257.0 35.5
0 0 -2 1'600.0 42.7
```

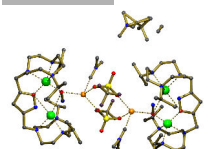
```
0 0 -1      2.8 0.55
0 0  1      3.8 0.63
0 0 -2 1'432.0 95.7
0 0 -2 1'282.0 85.9
```

several GB

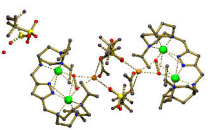
several files, 100's MB

1 "hkl"-file, 50MB

➡ **Phasing** ➡ **Refinement**



Starting model



Chemically sensible model

Model Building & Refinement

- Refinement optimises computationally the parameters of the structure with respect to the data
- Model building make modifications that are too large for computer optimisation, e.g.
  1. Addition or removal of atoms
  2. correction of atom types
  3. modelling of disorder and multiple conformations

Structure parameters

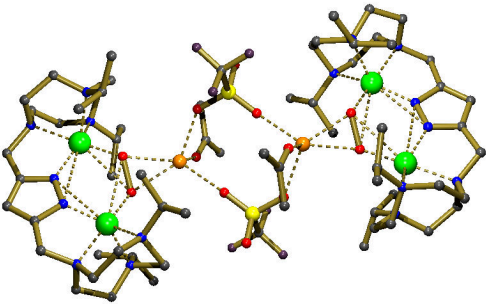
A "structure" consists of a set of parameters, *i.e.* numbers. Refinement improves these numbers for make the structure better correspond to the data.

```
          X        Y        Z    occ.
N2A   2 0.8142   0.9066   0.8201 11.00000 =
   0.0497   0.0413   0.0363 -0.0136  -0.0041 -0.0063
     U11      U22      U33      U23      U13      U12
```

1. 3 atom coordinates x, y, z
2. 6 atomic displacement parameters ADP $\begin{pmatrix} U_{11} & U_{12} & U_{13} \\ & U_{22} & U_{23} \\ & & U_{33} \end{pmatrix}$
3. possibly 1 occupancy parameter for disorder

9 Parameters per atom of the asymmetric unit are being refined, plus extra parameters in case of disorder or other special circumstances

Example for parameters



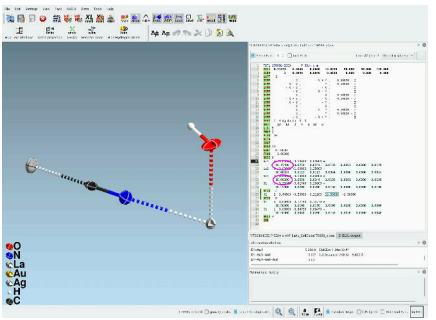
- $C_{34}H_{63}Cu_2F_6N_8NaO_{9.5}S_2$
- 62.5 non-hydrogen atoms
- 724 parameters
- hydrogen atoms are “special”

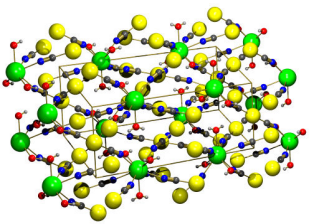
Atom occupancy

- Asymmetric unit: average of all asymmetric units of the crystal
- Molecules do not always strictly follow symmetry
- Some atoms sit on "special position", i.e. fix points of symmetry elements. Their occupancy is devided by the multiplicity of the symmetry element
- e.g. atom on three-fold axis: occupancy 33 %

Atom occupancy

Example: $La[Ag_{0.39}Au_{0.61}(CN)_2]_3 \cdot H_2O$ (Colis et al, Dalton Trans. (2005), 675-679) 39% of all unit cells contain Ag, 61% contain Au at the same position

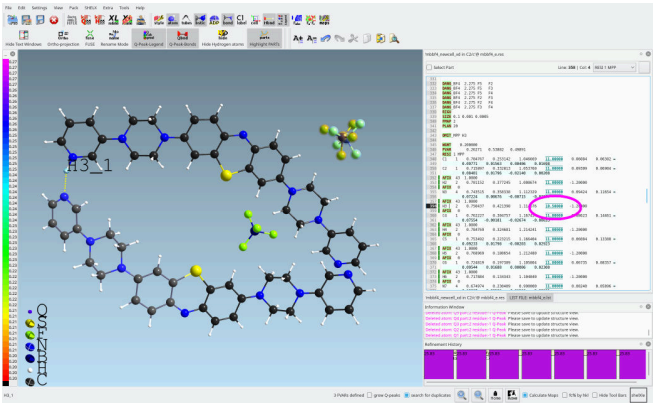




occ(Au1) + occ(Ag1) = 0.154 + 0.096 = 1/4 with 4-fold multiplicity

Atom occupancy

Example: Disordered BF_4^- and one H-atom on special position



Refinement = improvement of parameters

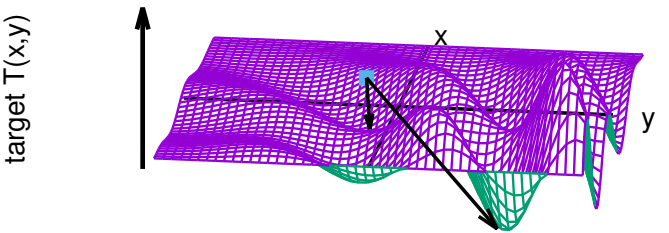
Computationally, refinement minimises the discrepancy between the observed data I_{obs} and the calculated data I_{calc}. I_{calc} is calculated from the model parameters, mainly atom coordinates x, y, z and atomic dispersion parameters ADPs

$$T(\vec{x}_i, U_i, (\text{occupancies}, \ldots)) = \sum_{(hkl)} w(hkl) |I_{obs}(hkl) - I_{calc}(hkl)|^2$$

$w(hkl)$ downweights untrusted reflections, typically $w(hkl) = 1/\sigma_I(hkl)$.
Note: different refinement programs use different target functions.

Least-square-minimisation

The shape of the target function $T = \sum_{(hkl)} w(hkl)|I_{\text{obs}}(hkl) - I_{\text{calc}}(hkl)|^2$ enables optimisation based on least-squares method (`L.S.` command in SHELXL).

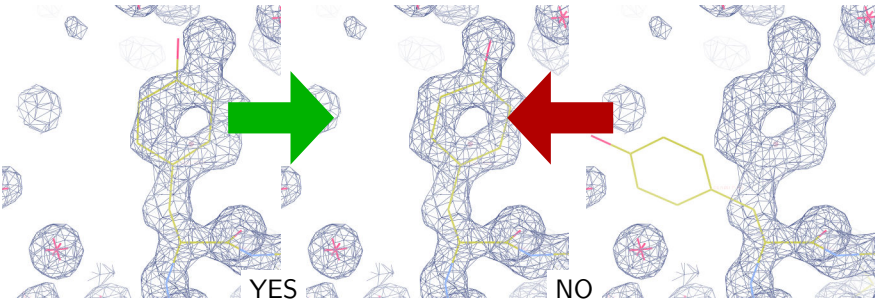


The algorithm finds the next minimum, but cannot jump across humps.

Advantages and limiations of refinement

- Refinement finds the "next" local minium
- only small changes in the structure
- does not add or remove atoms
- no change of element types
- one never knows whether the optimum is reached. However, for small molecules, the starting model usually converges to a good model.

The “next” local minium



Model building

- manual modifications a "help" refinement cross local humps
  - large movements of individual atoms (out of local traps)
  - delete wrong atoms
  - add missing atoms
  - correct atom type
- model building = add chemical understanding to the model
- graphic programs (Olex2, shelXle, ...)
- guided by the electron density map

Iterative process: improve model -> refine -> improve model -> refine -> ...

Electron density map and difference map

$$\rho(x,y,z) = FT(|F_{\text{obs}}(hkl)|, \phi_{\text{calc model}}(hkl)$$

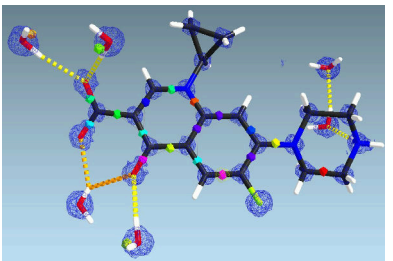
Fourier transformation from measured structure factor amplitudes $|F_{\text{obs}}(hkl)|$ and calculated phases $\phi_{\text{calc model}}(hkl)$
This model should follow this map.
The map

$$\Delta\rho(x,y,z) = FT(|F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|, \phi_{\text{calc}}(hkl))$$

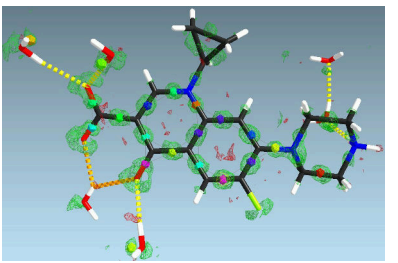
is called **difference map**. It reveals discrepancies between the model and the data.
Model building and refinement aim at reducing these discrepancies.

Example map: Ciprofloxacin

Structure of Ciprofloxacin, *J. Holstein, C. B. Hübschle, and B. Dittrich, Cryst. Eng. Comm. 2012 (14), pp. 2520–2531* (ultra high resolution 0.43 Å)



$\rho(x, y, z)$ (usally blue mesh)



$\Delta\rho(x, y, z)$ (usually green / red mesh)

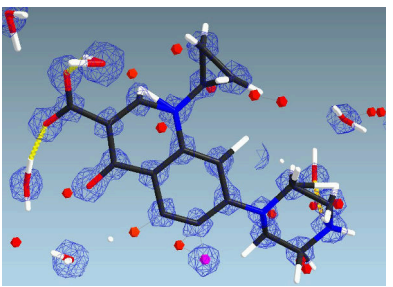
positive $\Delta\rho$: Model misses something. SHELXL places **Q-peaks**

negative $\Delta\rho$: model contains too much

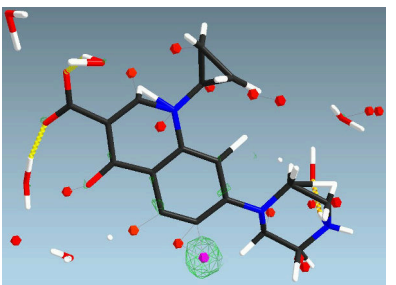
Example map: Ciprofloxacin

Structure of Ciprofloxacin, *J. Holstein, C. B. Hübschle, and B. Dittrich, Cryst. Eng. Comm. 2012 (14), pp. 2520–2531* (ultra high resolution 0.43 Å)

- data resolution truncated to 0.9 Å
- Fluorine atom F removed from model



$\rho(x, y, z)$ (blue mesh)

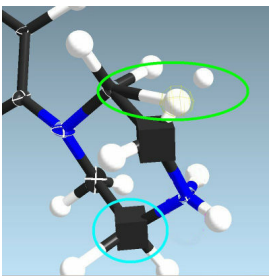


$\Delta\rho(x, y, z)$ (red / green mesh)

Refinement without restraints

$$T(\vec{x}_i, \mathbf{U}_i) = \sum_{(hkl)} w(hkl)(I_{\text{Daten}}(hkl) - I_{\text{Modell}}(hkl))^2$$

This formula carries out **unrestrained refinement**, purely taking experimental data into account. With poor data, this **can** cause



- unrealistic bond distances and bond angles
- negative ADPs (cubes) are physically meaningless
- refinement can produce non-sense results

Unrestrained refinement, example

Unrestrained refinement of protein structure with 1.4 Å resolution

Data to parameter ratio

Example Ciprofloxacin, ($a = 9.5$Å, $b = 9.9$Å, $c = 11.0$Å, $\alpha = 94.2°$, $\beta = 100.2°$, $\gamma = 91.3°$)

- $FC_{17}N_3O_9H_{30}$: $60 \times 9 = 540$ parameters

0.43 Å resolution 26'308 reflections. $26'308 : 540 = 48.7$ data points per parameter: very high data to parameter ratio, data sufficient to produce chemically sensible structure

0.8 Å resolution 2'926 reflections. $2'926 : 540 = 5.4$ data points per parameter: low data to parameter ratio, data insufficient to produce chemically sensible structure

Chemical sensible part need to be **restrained** -> restrained refinement

Restrained refinement

Except for at very high resolution, the refinement program has to be told some chemistry to make sure the model remains chemically meaningful. There are two different types how this can be accomplised:

Constraints Express an equality and permit no deviation from fixed value

Restraints Express similarity and provide some flexibility from target value.

Restraints are much more common than constraints

Constraints

- The structure of $La[Ag_{0.39}Au_{0.61}(CN)_2]_3 \cdot H_2O$ has either gold or silver at one location.
- In every unit cell there is always one atom at this location

$$occ(Au) + occ(Ag) = 1$$
$$occ(Au) = 1 - occ(Ag)$$

- Only the occupancy of silver has to be determined. The occupancy of gold can be calculated (or *vice versa*)
- remark: the program SHELXL uses the command FVAR ("free variables") to realise constraints.

Each constraint reduces the number of parameters by 1

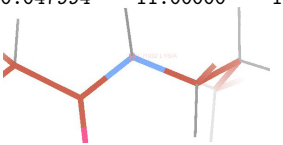
Important constraints

negative ADP value, mainly for hydrogen atoms: U(HA)= 1.2*U(CA)

```
CA    1    0.673087    0.878303    0.111632    11.00000    0.31129
HA    6    0.679625    0.855075    0.095775    11.00000   -1.20000
```

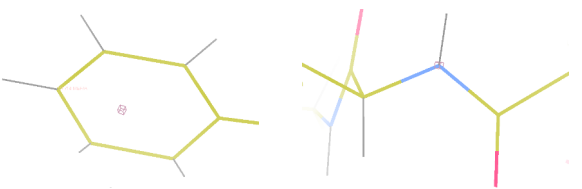
hydrogen positions: AFIX

```
N     3    0.611916    1.012005    0.052456    11.00000    0.18165
AFIX  43
H     6    0.628491    1.011598    0.033498    11.00000   -1.20000
AFIX   0
CB    1    0.622779    1.076653    0.067974    11.00000    0.18216
AFIX  23
HB1   6    0.608063    1.103479    0.072220    11.00000   -1.20000
HB2   6    0.641195    1.080130    0.047994    11.00000   -1.20000
AFIX   0
```



AFIX: riding atom model

- Except for at very high resolution ($d \ll 0.8$ Å), hydrogen atoms are invisible to X-rays
- the positions of most hydrogen atoms can be calculated: bond distances are known from spectroscopy, positions are determined by reducing steric clashes
- Advantages: hydrogen atoms do not add parameters, the contribute to VdW repulsion (BUMP command), they have a small, but non-zero contribution to the scattering.



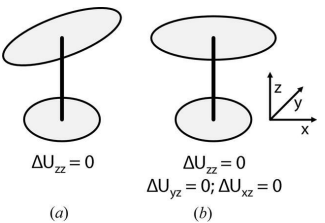
Restraints: Geometry

- restraints can be expressed as inequality "$\leq$"
- best known restraints: R. A. Engh, R. Huber, *Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement*, Acta Crystallogr. (1991), **A**47, pp. 392–400; *e.g.*

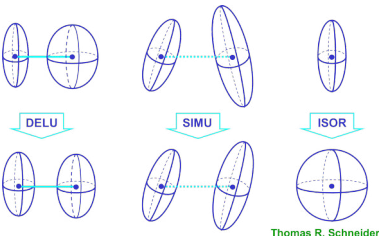
$$|d(N, C_\alpha) - 1.458\text{Å}| \leq 0.02 \qquad |d(C_\alpha, C_\beta) - 1.521\text{Å}| \leq 0.02$$

Restraints: ADP values

- restraints for ADPs: chemical bond affects thermal vibrations



RIGU command, Thorn *et al.*, Acta Cryst (2012), A68, 448–451



DELU, SIMU, ISOR (Thomas R. Schneider)

Restraints resemble data

Restraints are treated with additional terms to the target function:

$$T(\vec{x}_i, \mathbf{U}_i) = \sum_{hkl} w_{hkl}(I_{\text{Daten}}(hkl) - I_{\text{Modell}}(hkl))^2 + W \sum_{\text{N.B. } i} w_i(T_i^{\text{Daten}} - \langle T_i \rangle)^2$$

Restraints act like additional data points

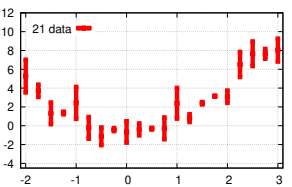
- W weights restraints and observed data
- the higher the resolution, the lower weight W
- the expected mean values $\langle T_i \rangle$ can be derived statistically from high resolution structures, or sometimes can be computed quantum chemically

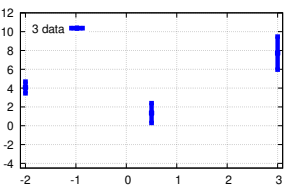
Example: Stabilisation through restraints

Two hypothetic measurements:

Experiment 1: high resolution, 21 pairs of measurements $(x_1, y_1), \ldots, (x_{21}, y_{21})$ and errors $\sigma_1, \ldots, \sigma_{21}$

Experiment 2: low resolution, 3 pairs of measurements $(x_1, y_1), \ldots, (x_3, y_3)$ and errors $\sigma_1, \ldots, \sigma_3$





Example: Stabilisation through restraints

Testing two models:

Model 1: $g(x) = g_2x^2 + g_1x + g_0$

Model 2: $h(x) = h_3x^3 + h_1x + h_0$

Either model has three parameters, g_0, g_1, g_2 and h_0, h_1, h_3 respectively. These parameters correspond to *e.g.* the model coordinates (x_i, y_i, z_i), or the ADPs U_i.
We will fit both models to the data to find out which model better represents the data.

Example: Stabilisation through restraints

Least-squares-minimisation:

$$\text{minimiere: } \sum_{i=1}^{N} \frac{1}{\sigma_i^2}(y_i - g(x_i))^2 \qquad \text{model 1}$$

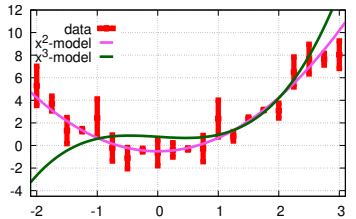
$$\text{minimiere: } \sum_{i=1}^{N} \frac{1}{\sigma_i^2}(y_i - h(x_i))^2 \qquad \text{model 2}$$

- Experiment 1: $N = 21$ data points
- Experiment 2: $N = 3$ data points

We will start with the high resolution experiment 1

Example: Stabilisation through restraints

experiment 1: high resolution; high data to parameter ratio = 21:3=7

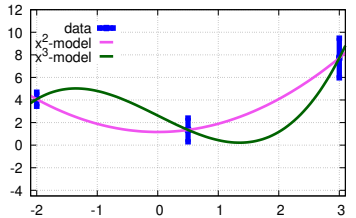


Model 1: $1.2x^2 + 0.0x - 0.5$
rmsd: 1.07
Model 2: $0.5x^3 - 0.3x - 0.8$
rmsd: 4.74

The *root mean square deviation* rmsd between model and data corresponds to the crystallographic $R1$ value.
The lower rmsd 1.07 clearly favours model 1. The pink curve also visually fits the data better than the green curve.

Example: Stabilisation through restraints

experiment 2: low resolution, low data to parameter ration = 3:3 = 1



model 1: $0.7x^2 + 0.0x + 1.2$

rmsd: 0

model 2: $0.5x^3 - 2.7x - 2.6$

rmsd: 0

When there are as many parameters as data points, any model can be fitted perfectly to the data. We cannot distinguish between the two models

Example: Stabilisation through restraints

experiment 2: low resolution with constraint

For some reason we know that the data must pass through the point $(0, 0)$. For the two models this means

$$\begin{aligned} 0 &= g(0) \\ &= g_2 * 0^2 + g_1 * 0 + g_0 \\ \Rightarrow g_0 &= 0 \end{aligned}$$

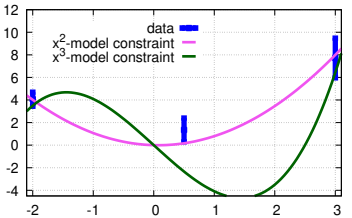
and analogously

$$h_0 = 0$$

The constraint reduced the number of parameters, only two parameters per model

Example: Stabilisation through restraints

experiment 2: low resolution with constraint



model 1: $0.9x^2 - 0.1x$
rmsd: 1.13
model 2: $0.8x^3 - 4.9x$
rmsd: 3.7

Due to the constraint, data to parameter ratio = 3:2 = 1.5. Now there is an *rmsd*, and it favours (again) the first model.

Summary refinement & model building

- The chemical model cannot be calculated directly from the data (*cf.* phase problem)
- phases are calculcate from the model
- model phases and observed data yield the electron density map, and electron difference map
- model building improves the model in large steps
- refinement optimises the model against the data
- medium resolution data or poor quality data require restraints and constraints in order to create a chemically sensible model

End of lecture