

Kinematical structural refinement with 3D ED data

how to refine a crystal structure kinematically

Electron Crystallography Course Pisa

4th July 2023

<https://homepage.univie.ac.at/tim.gruene/research/seminars>

Dr. Tim Grüne
Centre for Chemical Structure Analysis
Faculty of Chemistry
University of Vienna
tim.gruene@univie.ac.at



Contents

1	Refinement against (electron) diffraction data	3
2	Kinematic Theory of Diffraction	19
3	SHELXL	32
4	SHELXL with electron diffraction data	47
5	Model quality and data quality: structure validation	55
6	Indicators for data quality	58
7	Indicators for model quality	65

1 Refinement against (electron) diffraction data

From data collection to structure

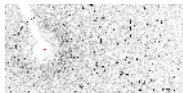
Data collection



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

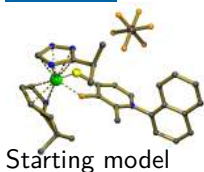
several GB

several files,
100's MB

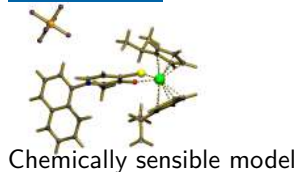
1 "hkl"-file, 50MB



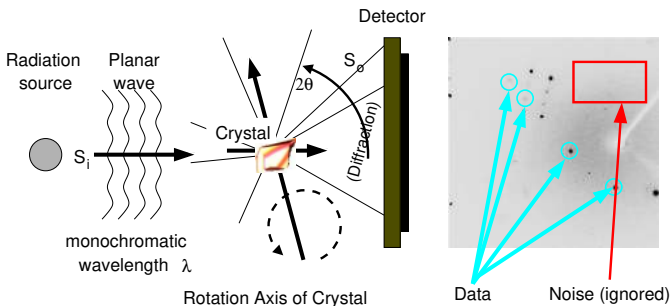
Phasing



Refinement



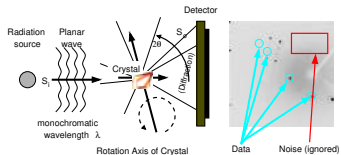
The (electron) diffraction experiment



Typical elements:

- Monochromatic, planar wave, wavelength λ
- Crystal rotates about a point, no translation
- Area detector
- Data frames show reflections ("spots")

The (electron) diffraction experiment



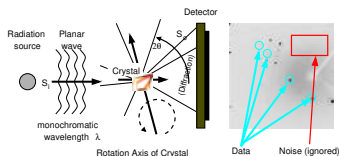
The ideal crystal produces a diffraction pattern, that consists of reflections (spots) Positions of the reflections can be described with the Laue equations:

$$(\vec{S}_o - \vec{S}_i) \cdot \vec{a} = h$$

$$(\vec{S}_o - \vec{S}_i) \cdot \vec{b} = k$$

$$(\vec{S}_o - \vec{S}_i) \cdot \vec{c} = l$$

The (electron) diffraction experiment



- \vec{S}_i : points from source to crystal, length $1/\lambda$
- \vec{S}_o : points from crystal to detector, length $1/\lambda$
- \vec{a} , \vec{b} , \vec{c} unit cell parameters
- h, k, l: Miller indices, integer numbers

Type of radiation (X-ray, neutrons, electrons) are not part of the Laue equations, nor is the atomic composition of the crystal.

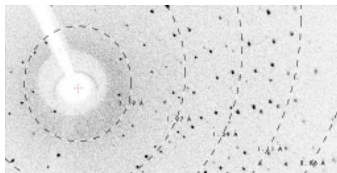
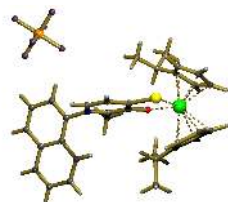
The (electron) diffraction DATA

- Data processing reads the frames
- Data processing determines for each reflection:
 1. Miller index h, k, l
 2. Intensity $I(hkl)$
 3. standard uncertainty $\sigma_I(hkl)$

0	0	1	8.185E+00	1.207E+00
0	0	2	5.758E+00	9.215E-01
0	0	3	2.747E+00	6.489E-01

- these number are our data.
- the molecular model is **not** data

Chemical Crystallography



Structural chemists, or structural biologists, are not interested in “spots”, but would like to learn about the molecule(s) inside the crystal. A link is needed between the molecule and the reflection intensities.

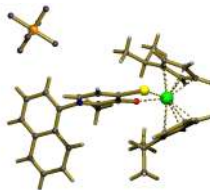
How to describe a molecule

- A molecule consists of atoms
- some of the atoms are bonded to one another
- there are different types of bonds (covalent, ionic, van-der-Waals interaction ...)
- The molecule is part of a crystal, and as crystal structure. The crystal has unit cell parameters a , b , c , α , β , γ , and belongs to one of 230 spacegroups (e.g. $C2/c$, $P2_12_12$, $Ia\bar{3}d$, ...)



How to describe a molecule

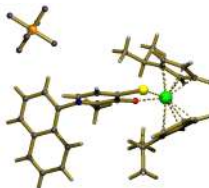
- A molecule consists of atoms
- An atom has an element type (C , O , Zr , Au ...)
- An atom has a position, i.e. it has coordinates (x, y, z)
- The atom vibrates, its thermal motion is described by the Atomic Displacement Parameter (ADP, Debye-Waller-factor)
- The atom position may not be fully occupied (especially solvent molecules), and share its position with another element type (especially metal ions, e.g. $Fe_{3-x}M_xO_4$ ($M=Cr, Mn, Co, Ni$))



How to describe a molecule

Note on graphical presentation of crystal structures

- Bonds between atoms are usually not described
- Graphics programs like Olex2, ShelXle, Coot, draw a bond when two atoms are closer than the sum of their van-der-Waals radii



The Structure factor

For every reflection (h, k, l) , a structure factor $F(hkl)$ is *defined* as the Fourier transform of content of the unit cell. “Content” refers to the type of interaction of the radiation:

X-rays interact with the electrons in the crystal, described by the electron density ρ : $F(hkl) = \int \rho(x, y, z)e^{2\pi i(hx+ky+lz)}$

- $\rho(x, y, z)$: local density of electrons, [$e/\text{\AA}^3$]
- High density at atoms, low density away from atoms
- contribution of nuclei irrelevant

electrons interact with the electrostatic potential ϕ in the crystal: $F(hkl) = \int \Phi(x, y, z)e^{2\pi i(hx+ky+lz)}$

- $\Phi(x, y, z)$: local electrostatic potential, [V]
- Combination of nucleic charge and electron cloud

Structure factor: Molecule and Intensities

How does the structure factor $F(hkl)$ bridge between the electron diffraction experiment and the molecules inside the crystal?

1. What is the connection between $F(hkl)$ and the chemical compound inside the crystal? — the Independent Atom Model
2. What is the connection between $F(hkl)$ and the diffraction pattern? — the Kinematic Theory of Diffraction

Computation of intensities $I_{\text{calc}}(hkl)$

- There are several ways to compute $I_{\text{calc}}(hkl)$ from the model
- some are more accurate and more complex
- some are less accurate and fast
- the IAM — independent atom model — assumes scattering from isolated atoms
- atomic scattering factors can be calculated
- the total $I_{\text{calc}}(hkl)$ results from the sum of the individual atoms

The Independent Atom Model (IAM)

The *Independent Atom Model* (IAM) is a powerful method to calculate the atomic structure factor $F(hkl)$. Each atom contributes independently from the others to $F(hkl)$.

$$F_{\text{calc}}(hkl) = \sum_{\substack{\text{atoms } j \\ \text{in u.c.}}} f_j(\theta) e^{-8\pi^2 U_j(\theta, \lambda)} e^{2\pi i(hx_j + ky_j + lz_j)} \quad (1)$$

f_j atomic *form* factor. Depends on atom element, on radiation type, and (sometimes) on wavelength. Decreases with decreasing scattering angle

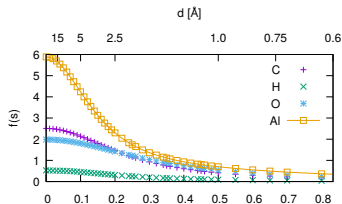
$U_j(\theta, \lambda)$ atomic displacement parameter (ADP, alias Debye-Waller factor): models thermal vibration of atoms

$e^{2\pi i(hx_j + ky_j + lz_j)}$ phase shift of the atom relative to the origin of the unit cell

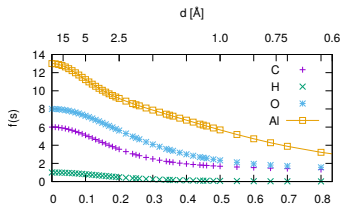
IAM: atomic scattering factors

Atomic scattering factors can be computed for isolated atoms (and isolated ions)

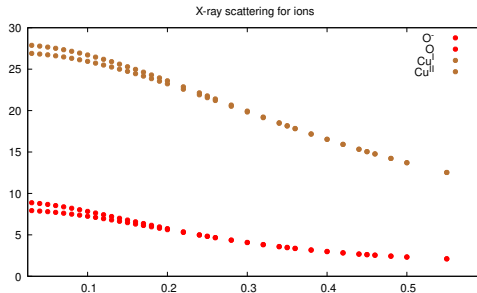
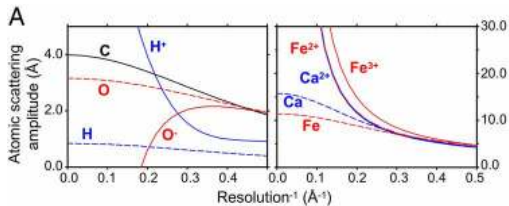
Electron diffraction



X-ray diffraction



Ionic scattering factors [1]



2 Kinematic Theory of Diffraction

Structure Factors

- Experimental **data** after processing: observed intensities $I_{\text{obs}}(hkl)$
- Structural **model** composed of atoms with element type, coordinates (x, y, z) , temperature factor B (Debye-Waller factor, ADP)
- Refinement programs (PHENIX, REFMAC, BUSTER, SHELXL, OLEX2, JANA) improve the model with respect to the data
- They must compute calculated intensities $I_{\text{calc}}(hkl)$

Link observed intensities to structure factor

The kinematic theory of diffraction is equivalent to the assumption that every electron (for electron diffraction) and every photon (for X-ray diffraction) undergoes one scattering event in the crystal.

This results in a simple relationship between the structure factor and the observed intensities:

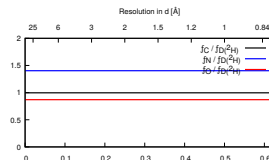
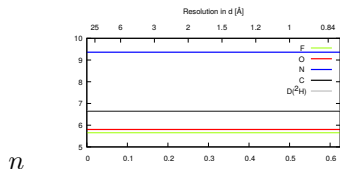
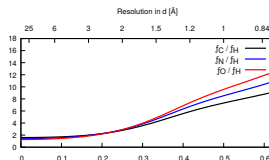
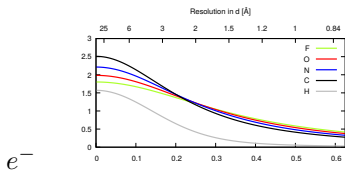
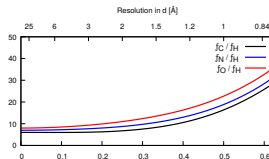
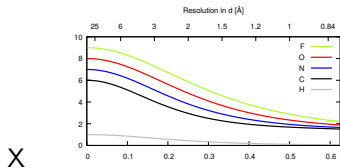
$$I_{obs}(hkl) \propto |F(hkl)|^2$$

In the kinematic theory of diffraction for X-rays [2],

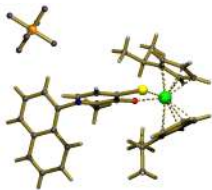
$$I_{obs}(hkl) = \frac{e^4}{m_e^2 c^4} \frac{\lambda^3 V_{crystal}}{V_{u.c.}^2} I_0 L P T E |F(hkl)|^2$$

In electron diffraction, the assumption $I_{obs}(hkl) \propto |F(hkl)|^2$ is less accurate, but still works for many purposes.

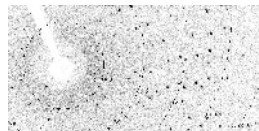
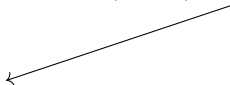
Comparison of scattering factors



“Kinematic” refinement



$$\xrightarrow{\text{IAM}} (x, y, z), U_{ij} \rightarrow F_{\text{calc}}(hkl)$$



$$|F_{\text{calc}}(hkl)|^2 \propto I_{\text{obs}}(hkl)$$

kinematic
 \longleftrightarrow

Refinement programs improve the model parameters $(x, y, z), U_{ij}$ with the aim to get a better match between $|F_{\text{calc}}(hkl)|^2$ and $I_{\text{obs}}(hkl)$

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

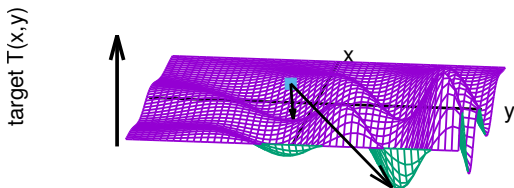
$$T(\vec{x}_i, U_i, (\text{occupancies}, \dots)) = \sum_{(hkl)} w(hkl) |I_{\text{obs}}(hkl) - I_{\text{calc}}(hkl)|^2$$

$w(hkl)$ 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).

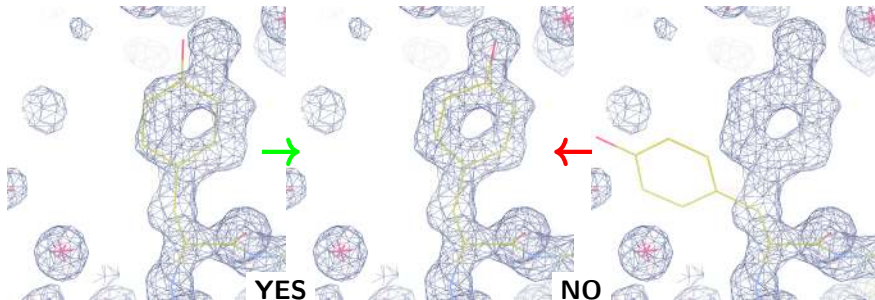


The algorithm finds the next minimum, but cannot jump across humps.

Advantages and limitations of refinement

- Refinement finds the “next” local minimum
- 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 minimum



Model building

- manual modifications “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, Crystals, JANA...[3, 4, 5, 6])
- guided by the map and difference map

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

Map and difference map

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

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.

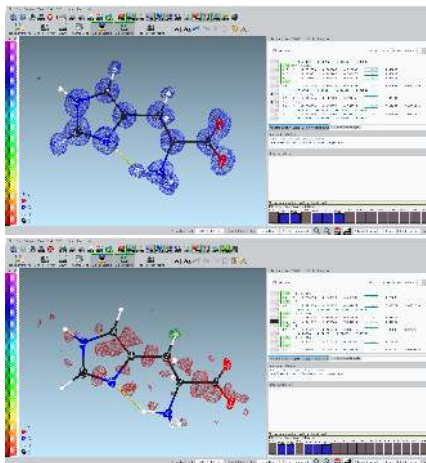
The map

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

is called **difference map**. It reveals discrepancies between the model and the data.

Model building and refinement aim at reducing these discrepancies.

Map and difference map



Traditionally, the map is shown as blue mesh; the difference map is shown as green (+, atom missing) and red (-, atom to many) mesh

Dynamic diffraction [7, 8, 9, 10]

- kinematic diffraction theory: $I_{obs}(hkl) \propto |F(hkl)|^2$ is the “normal” theory
- A better detailed description for $I_{obs}(hkl)$ is based on the *dynamic theory* of diffraction.
- Implemented in the refinement program JANA2020 [5]
- SHELXL, OLEX2 [3, 11]: based on the kinematic diffraction theory
- kinematic refinement is very fast, and particularly good to (metal-) organic compounds
- dynamic theory results in better fine details (positions of hydrogen atoms ...), better models for inorganic structures with heavy elements

3 SHELXL

Overview of SHELXL[11]

- Download and documentation shelx.uni-goettingen.de
- Author: George M. Sheldrick
- > 50 years development (SHELX-76 first official version)
- Code currently maintained by Isabel Uson (IBMB-CSIC, Spain)



The SHELX homepage

SHELX is a set of programs for the refinement of crystal (SM) and macromolecular (MM) crystal structures by direct crystal X-ray and neutron diffraction. These programs are available for both NO Windows, Unix-like and non-Unix systems. They are compatible with all modern versions of Linux, Windows and Mac OS X, and are free for academic use. For special cases are expected to be available for other operating systems and support for all CPUs.

The programs may be called from a GUI such as [shelxGUI](#), [shelxGUI2](#), [shelxGUI3](#), [shelxGUI4](#) and [shelxGUI5](#) or [shelxGUI6](#), [shelxGUI7](#), [shelxGUI8](#), [shelxGUI9](#), [shelxGUI10](#), [shelxGUI11](#), [shelxGUI12](#), [shelxGUI13](#), [shelxGUI14](#), [shelxGUI15](#), [shelxGUI16](#), [shelxGUI17](#), [shelxGUI18](#), [shelxGUI19](#), [shelxGUI20](#), [shelxGUI21](#), [shelxGUI22](#), [shelxGUI23](#), [shelxGUI24](#), [shelxGUI25](#), [shelxGUI26](#), [shelxGUI27](#), [shelxGUI28](#), [shelxGUI29](#), [shelxGUI30](#), [shelxGUI31](#), [shelxGUI32](#), [shelxGUI33](#), [shelxGUI34](#), [shelxGUI35](#), [shelxGUI36](#), [shelxGUI37](#), [shelxGUI38](#), [shelxGUI39](#), [shelxGUI40](#), [shelxGUI41](#), [shelxGUI42](#), [shelxGUI43](#), [shelxGUI44](#), [shelxGUI45](#), [shelxGUI46](#), [shelxGUI47](#), [shelxGUI48](#), [shelxGUI49](#), [shelxGUI50](#), [shelxGUI51](#), [shelxGUI52](#), [shelxGUI53](#), [shelxGUI54](#), [shelxGUI55](#), [shelxGUI56](#), [shelxGUI57](#), [shelxGUI58](#), [shelxGUI59](#), [shelxGUI60](#), [shelxGUI61](#), [shelxGUI62](#), [shelxGUI63](#), [shelxGUI64](#), [shelxGUI65](#), [shelxGUI66](#), [shelxGUI67](#), [shelxGUI68](#), [shelxGUI69](#), [shelxGUI70](#), [shelxGUI71](#), [shelxGUI72](#), [shelxGUI73](#), [shelxGUI74](#), [shelxGUI75](#), [shelxGUI76](#), [shelxGUI77](#), [shelxGUI78](#), [shelxGUI79](#), [shelxGUI80](#), [shelxGUI81](#), [shelxGUI82](#), [shelxGUI83](#), [shelxGUI84](#), [shelxGUI85](#), [shelxGUI86](#), [shelxGUI87](#), [shelxGUI88](#), [shelxGUI89](#), [shelxGUI90](#), [shelxGUI91](#), [shelxGUI92](#), [shelxGUI93](#), [shelxGUI94](#), [shelxGUI95](#), [shelxGUI96](#), [shelxGUI97](#), [shelxGUI98](#), [shelxGUI99](#), [shelxGUI100](#).

SHELXT - New and improved (SM) small molecule program

SHELXS - Classical method for (SM) small molecule

SHELXL - SM and MM refinement program compatible with SHELXS and SHELXD

SHELXD - Empirical direct method for (SM) small molecule refinement

Technical highlights of SHELXL

- fast, flexible, command line program
- available for Linux, Windows, Mac OS X
- very efficient language to describe chemical compounds in molecules, including complex situations, like disorder, partial substitution of metals (as e.g. in $\text{Fe}_{3-x}\text{M}_x\text{O}_4$ (M=Cr, Mn, Co, Ni))
- suitable for minerals, organic compounds, macromolecules,...
- suitable for X-ray, neutron, electron data

Using SHELXL

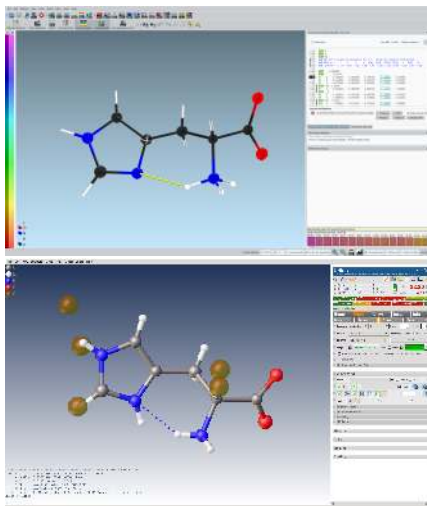
- command line program, the command

```
#> shelxl histidine
```

- requires the files `histidine.ins` and `histidine.hkl` in the same directory.
- SHELXL will create `histidine.res`, `histidine.fcf` and `histidine.lst` (and `histidine.cif`)

Using SHELXL

Graphical user interfaces: ShelXle [4], Olex2 [3]



Setting up SHELXL — hkl-file

- SHELXL requires the data, the *hkl*-file
- Each line correspond to one reflection with $h k l I \sigma_I$
- do not merge your reflections! SHELXL will merge internally

0 0 1 249.82 36.84

0 0 2 175.75 28.14

0 0 3 83.84 19.81

0 0 4 2487.24 309.80

...

-8 0 1 -1.43 5.37

-8 0 -1 2.75 3.33

-8 0 2 1.25 9.13

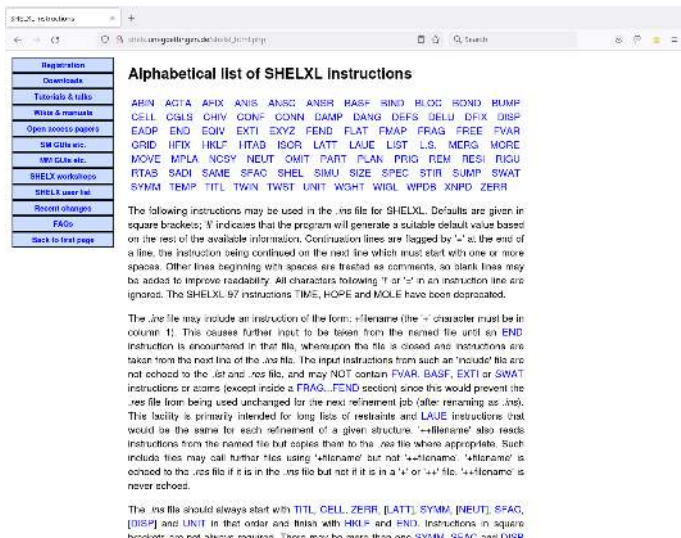
0 0 0 0.00 0.0

Setting up SHELXL — ins-file

- SHELXL requires an instruction file, the *ins*-file
- *ins* file contains the atom coordinates, experimental description and all commands for SHELXL

```
REM Formula found by SHELXT:  C9 N O
CELL  0.02508   5.2140   7.4410  19.1120   90.000   90.000   90.000
ZERR   4.000   0.0010   0.0015   0.0038   0.000   0.000   0.000
LATT  -1
SYMM  1/2-X, -Y, 1/2+Z
...
SFAC  C 0.1361 0.3731 0.5482 3.2814 1.2266 13.0456 0.5971 41.0202 0 =
...
O1   4   0.726057   0.495585   0.198134   11.00000   0.03368
O2   4   0.323826   0.492355   0.241326   11.00000   0.02855
```

SHELXL commands



The screenshot shows a web browser displaying the 'SHELXL instructions' page. On the left is a navigation menu with links such as 'Registration', 'Downloads', 'Tutorials & talks', 'Notes & manuals', 'Open access papers', 'SW GUIs etc.', 'MO GUIs etc.', 'SHELXL workshops', 'SHELXL user list', 'Recent changes', 'FAQs', and 'Back to first page'. The main content area is titled 'Alphabetical list of SHELXL instructions' and contains a list of instructions: ABIN, ACTA, APX, ANIS, ANSG, ANSR, BASF, BEND, BLOC, BOND, BUMP, CELL, CGLS, CHIV, CONF, CONN, DAMP, DANG, DEFS, DELU, DFIX, DISP, EADP, END, EQIV, EXTI, EXYZ, FEND, FLAT, FMAP, FRAG, FREE, FVAR, GRID, HFIX, HKLf, ITAD, ISOR, LATT, LAUE, LIST, L.S., MERG, MORE, MOVE, MPLA, NCSY, NEUT, OMIT, PART, PLAN, PRIG, REM, RESI, RIGU, RTAB, SADI, SAME, SFAC, SHEL, SIMU, SIZE, SPEC, STIR, SUMP, SWAT, SYMM, TEMP, TITL, TWIN, TWST, UNIT, WGHT, WIGL, WPDB, XMPD, ZERR.

The following instructions may be used in the .ins file for SHELXL. Defaults are given in square brackets; * indicates that the program will generate a suitable default value based on the rest of the available information. Continuation lines are flagged by '*' at the end of a line, the instruction being continued on the next line which must start with one or more spaces. Other lines beginning with spaces are treated as comments, so blank lines may be added to improve readability. All characters following '?' or '?' in an instruction line are ignored. The SHELXL 97 instructions TIME, HOPE and MOLE have been deprecated.

The .ins file may include an instruction of the form: *filename (the '*' character must be in column 1). This causes further input to be taken from the named file until an END instruction is encountered in that file, whereupon the file is closed and instructions are taken from the next line of the .ins file. The input instructions from such an 'include' file are not echoed to the .lst and .res file, and may NOT contain FVAR, BASF, EXTI or SWAT instructions or atoms (except inside a FRAG...FEND section) since this would prevent the .res file from being used unchanged for the next refinement job (after renaming as .ins). This facility is primarily intended for long lists of restraints and LAUE instructions that would be the same for each refinement of a given structure. '*filename' also reads instructions from the named file but copies them to the .res file where appropriate. Such include files may call further files using 'filename' but not '*filename'. 'filename' is echoed to the .res file if it is in the .ins file but not if it is in a '*' or '*+' file. '*filename' is never echoed.

The .ins file should always start with TITL, CELL, ZERR, [LATT], SYMM, [NEUT], SFAC, [DISP] and UNIT in that order and finish with HKLF and END. Instructions in square brackets are not always required. There may be more than one SYMM, SFAC and DISP

Important SHELXL commands: spacegroup

- there is no spacegroup command
- the symmetry is derived from
 - the lattice type LATT: 1=P, 2=I, 3=rhombohedral obverse on hexagonal axes, 4=F, 5=A, 6=B, 7=C.
 - a '-' sign for non-centrosymmetric space groups ('LATT -1')
 - a list of symmetry operators SYMM
- E.g. $C2/m$:

LATT 7

SYMM -X, Y, -Z

- E.g. $P\bar{1}$:

LATT 1

Important SHELXL commands: SFAC

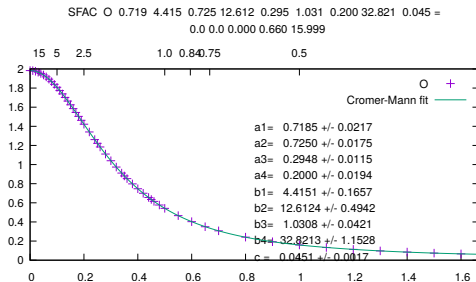
Reminder: $F_{\text{calc}}(hkl) = \sum_{\text{in u.c.}}^{\text{atoms } j} f_j(\theta) e^{-8\pi^2 U_j(\theta, \lambda)} e^{2\pi i(hx_j + ky_j + lz_j)}$

SHELXL describes the scattering factor $f_j(\theta)$ with the Cromer-Mann-parametrisation:

$$f(s) = \sum_{j=1}^4 a_j \exp(-b_j s^2) + c \quad (2)$$

($s = \sin \theta / \lambda$)

Important SHELXL commands: SFAC



- The nine Cromer-Mann parameters a_1 , b_1 , a_2 , b_2 , a_3 , b_3 , a_4 , b_4 , and c need to be fitted for each element.
- For X-ray diffraction, these are hardcoded into the SHELXL source code.
- Their values are not unique, only the curve $f(s)$ they describe, should be as close as possible to the scattering curve.

Important SHELXL commands: SFAC

- The order of SFAC commands is important
- Without values, they refer to X-ray diffraction and names should be existing element names

```
SFAC C H N O Fe
```

- With values, you can invent your own names

```
SFAC O 0.719 4.415 0.725 12.612 0.295 1.031 0.200 32.821 0.045 =  
0.0 0.0 0.000 0.660 15.999
```

```
SFAC O- -65.208 2152.211 -5.700 141.057 -80.892 -0.016 -19.125 609.408 =  
81.738 0.0 0.0 0.000 1.260 15.999
```

SHELXL: list of atoms

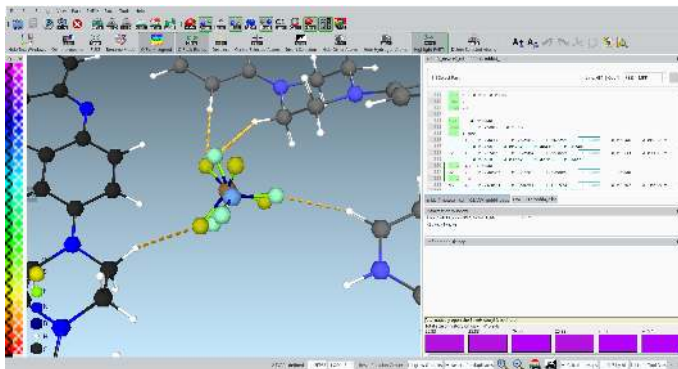
- Every line that does not start with one of the SHELXL commands, is considered an atom
- The first 4 character are the atom name. They are arbitrary and can even start with a number. The next number N refers to the N^{th} SFAC scattering command

O2	4	0.724492	0.495382	0.198080	11.00000	0.03256
N1	3	0.908998	0.679622	0.310716	11.00000	0.02241
C2	1	0.703321	0.423167	0.505136	11.00000	0.02232
Name	SFAC	X	Y	Z	10+occ.	Uiso

SHELXL: Free variables FVAR

- Any parameter with a value between -5.0 and 5.0 is refined by SHELXL
- Any parameter can be fixed and excluded from refinement by adding 10.0
 - Fully occupied atom: $occ = 11.0$
 - Fully occupied atom on two-fold atom: $occ = 10.5$
- Adding e.g. 20 (in general $n * 10, n \geq 20$) to a parameter multiplies the parameter with the second free variable (in general, n^{th} free variable)
- The most common use of free variables are distinct conformations

SHELXL: FVAR example



- BF_4^- anion can “wobble” about, but must be somewhere
- one molecule: occupancy grouped into 21.0, occupancy 54 %
- the other molecule grouped into -21.0, occupancy 100 % - 54 % = 46 %

4 SHELXL with electron diffraction data

overview

- Scattering factor: the SFAC command for ED
- bond distance of hydrogen atoms: AFIX command
- validation: what is a useful structure

ED: the SFAC command

- Cromer-Mann-parameters for electron scattering can be produced from X-ray and the Mott-Bethe formula [1]

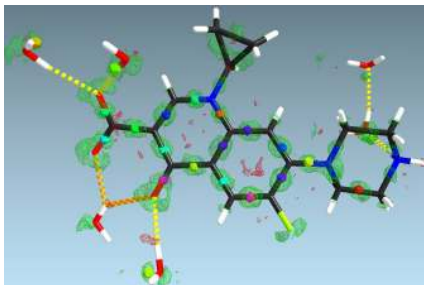
$$\begin{aligned} f_{\text{el}}(s) &= \frac{m_0 e^2}{8\pi\epsilon_0 h^2} \frac{Z_0 - f_X(s) + \Delta Z}{s^2} \\ &= 0.02393366 \frac{Z_0 - f_X(s) + \Delta Z}{s^2} \end{aligned}$$

ED: where to get the SFAC command

- Peng [12]
- my hand-made list <https://homepage.univie.ac.at/tim.gruene/research/ed/data/sfac-electron.dfx> (incomplete, additions are made upon request)
- web script at UCLA, <https://srv.mbi.ucla.edu/faes/>

Hydrogen bond distance

- Hydrogen atoms are usually placed in riding position, calculated from the geometry of the donor-atom
- In SHELXL: AFIX command: 'AFIX 33' = CH_3 , 'AFIX 13': $N - H$
- Bond lengths for hydrogens seem artificially shortened with X-ray diffraction.



model of ciprofloxacin from 0.43Å X-ray [13]

AFIX with electron diffraction

- The electron cloud is between the proton and the donor atom: H-bonds appear 10'ish % shorter with X-ray diffraction
- with neutron diffraction, internucleic distances have to be used, and there are some high-resolution data sets from neutron diffraction
- Add the command "NEUT" to the ins file after the SFAC commands produces a list of interneuclear distances

Default effective X-H and X-D distances for T = -173.0C for neutron diffraction

AFIX m =	1	2	3	4	4[N]	3[N]	15[B]	8[O]	9	9[N]	16
d(X-H) =	1.12	1.11	1.08	1.10	1.03	1.05	1.22	1.00	1.10	1.03	1.10

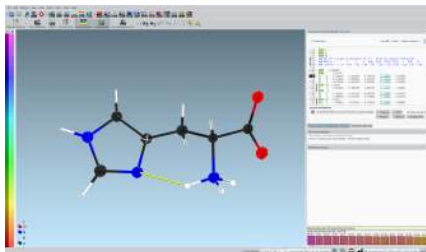
- append the distance to each AFIX command
- Note: distances are temperature dependent!

Meaning of the AFIX numbers explained at http://shelx.uni-goettingen.de/shelxl_html.php#AFIX

AFIX with electron diffraction

Extract from histidine data:

C7	1	0.668051	0.578240	0.320004	11.00000	0.01858
AFIX	13	1.1200				
H7	2	0.522826	0.668243	0.345236	11.00000	-1.20000
AFIX	0					
N2	3	0.871018	0.498076	0.553595	11.00000	0.02495
AFIX	43	1.0300				
H2	2	0.854260	0.490760	0.607218	11.00000	-1.20000
AFIX	0					



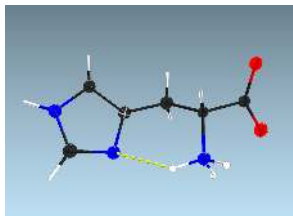
9 H-Atoms, R1 drops from 21.96 % to 21.13 %

AFIX with electron diffraction

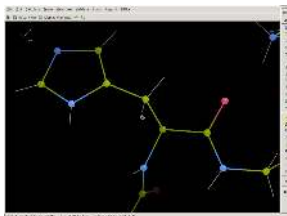
- “neutron” distances for H-atoms improve the model a lot
- “neutron” distances may yet not be the proper distances
- free refinement for H-atoms at high resolution may produces better estimates in the future

5 Model quality and data quality: structure validation

Atom coordinates \neq model accuracy



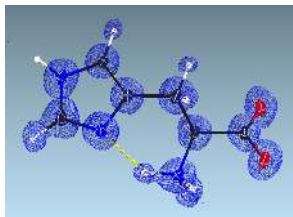
L-Histidine at 0.65 Å resolution



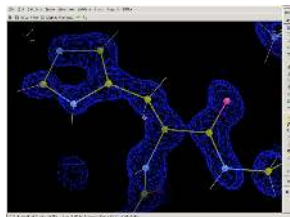
L-Histidine at 0.87 Å resolution
(PDB ID 7JULY)

The coordinates of the model do not reveal the data quality, nor the model quality.

Model coordinates = interpretation of data



L-Histidine at 0.65 Å resolution

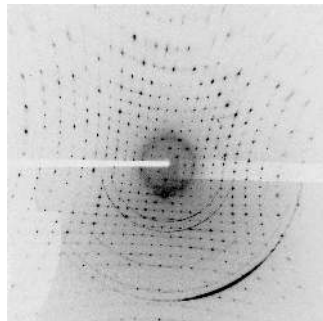
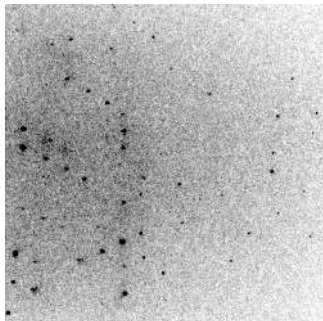


L-Histidine at 0.87 Å resolution
(PDB ID 7JULY)

Only in combination with the data can we judge the model quality

6 Indicators for data quality

Example data quality



Important quality indicators

Recommended reading: *Against Method: Table 1—Cui Bono?* Rupp [14]

R_{meas} relative difference between symmetry equivalent reflections and their mean value

data completeness : fraction of measured data w.r.t. theoretically possible data

multiplicity (*alias: redundancy*): how often every unique reflection was measured (on average)

signal strength $I(hkl)/\sigma_{I(hkl)} < 1$: noise

CC_{1/2} [15]

1. split data set into two random halves
2. calculated correlation coefficient between symmetry equivalent reflections

R-values for data

The classic data quality indicator is R_{int} , alias R_{merge} or R_{sym} :

$$R_{\text{int}} = \sum_h \sum_j \frac{|I_{hj} - \langle I_h \rangle|}{\langle I_h \rangle}$$

R_{int} mathematically increases with multiplicity, although data quality improves with multiplicity

R_{int} is typically shown in publications. It is, however, obsolete and should not be published. R_{meas} *alias* $R_{\text{r.i.m.}}$ should be published instead [16]:

$$R_{\text{meas}} = \sum_h \frac{n_h}{n_h - 1} \sum_j \frac{|I_{hj} - \langle I_h \rangle|}{\langle I_h \rangle}$$

Example data statistics (XPREP)

Resolution	#Data	#Theory	%Complete	Redundancy	Mean I	Mean I/s	Rmerge
Inf - 2.46	196	197	99.5	39.27	215.01	110.27	0.0300
2.46 - 1.13	1762	1825	96.5	14.86	75.32	42.01	0.0453
1.13 - 0.89	1972	2123	92.9	8.71	25.52	19.00	0.0895
0.89 - 0.77	2007	2258	88.9	6.81	10.84	10.39	0.1425
0.77 - 0.69	1864	2499	74.6	3.37	5.66	5.76	0.1885
0.69 - 0.62	2108	3360	62.7	2.24	2.88	3.29	0.2890
0.62 - 0.57	1929	3542	54.5	1.44	1.51	1.79	0.4191
0.57 - 0.54	1123	2367	47.4	1.10	0.90	1.14	0.5593

0.64 - 0.54	3720	7014	53.0	1.43	1.47	1.76	0.4170
Inf - 0.54	12961	18171	71.3	5.08	20.64	13.61	0.0514

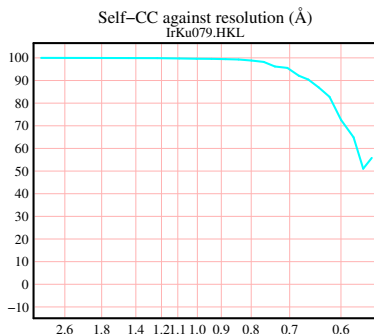
Merged [A], lowest resolution = 11.49 Angstroms

CC1/2, and resolution cut-off

A good quality crystal diffracts beyond the theoretical limit $d_{\min} = \lambda/2$. Resolution cut-off is not an issue, one can use all data. Large complexes, supramolecular structures, low quality crystals reach the diffraction limit before the theoretical limit. One has to decide where to cut the diffraction data.

- CC1/2 should be close to 100% throughout resolution range
- where CC1/2 drops below 70%, noise becomes significant, and data at higher resolution can be excluded from refinement
- $I/\sigma(I)$ should be about 2, where CC1/2 about 70%
- $I/\sigma(I)$ should be about 1, where CC1/2 about 30% (in cases very resolution cut-off is critical)

Example $CC_{1/2}$, and resolution cut-off



$CC_{1/2}$ vs. data resolution; plot generated with XPREP

7 Indicators for model quality

R-values for the model

$$R = R_1 = \sum_h \frac{||F_h(data)| - |F_h(model)||}{|F_h(data)|}$$

weighted intensity based R-value:

$$wR_2 = R_B = \sqrt{\sum_h \frac{|w_h(I_h(data) - I_h(model))|^2}{w|I_h(data)|^2}}$$

small molecules R_1 of the refined model 2-5 %.

supramolecules compounds, MOFs, ... R_1 of the refined model **can** be higher, 2-15 %

macromolecular compounds R_1 of the refined model 15-25 %

To a great extent, this discrepancy is due to the unmodelled solvent region in the latter two types of compounds

Goodness of Fit — GooF

$$GooF = \sqrt{\frac{\sum_h w_h (F_h^2(data) - F_h^2(model))^2}{n - p}}$$

- Takes number of parameters (p) and number of data (n) into account
- Ideally ≈ 1 , increases with worse model

model: residual density

SHELXL calculates the “highest peak” and “deepest hole” in the electron density map. Units are electrons, e.g. at the **beginning** of model building:

Electron density synthesis with coefficients Fo-Fc

```
Highest peak    4.95  at  0.5434  0.9981  0.3231  [  0.04 A from RU01  
]
```

```
Deepest hole   -3.34  at  0.0057  0.4976  0.3299  [  0.99 A from RU02  
]
```

~~~~~

```
Mean =      0.00,    Rms deviation from mean =      0.34 e/Å3
```

~~~~~

model: residual density

SHELXL calculates the “highest peak” and “deepest hole” in the electron density map. Units are electrons, e.g. for the **refined** model:

Electron density synthesis with coefficients Fo-Fc

Highest peak 0.50 at 0.6610 0.1969 0.4278 [0.69 Å from C006
]

Deepest hole -1.22 at 0.2635 0.6156 0.2132 [0.04 Å from P003
]

~~~~~

Mean = 0.00, Rms deviation from mean = 0.06 e/Å<sup>3</sup>

~~~~~

Acknowledgements

- Soheil Mahmoudi, Julian Maisriml, Maria Schmetterer, Nicole Bolehradski, (Chemistry, Uni Vienna)
- R. Miletich-Pawliczek, Chr. Lengauer (Mineralogie, Uni Vienna)
- Julian T. C. Wennmacher, Jeroen A. v. Bokhoven, et al. (ETH Zurich)
- Erik Frojdh, Bernd Schmitt, et al. (PSI Detector Group)
- Julian Holstein (TU Dortmund)
- Swiss Nanoscience Institute
- Sacha DeCarlo, Clemens Schulze-Briese (Dectris Ltd)



Julian Wennmacher, Soheil Mahmoudi

References

- [1] Koji Yonekura et al. 'Ionic scattering factors of atoms that compose biological molecules'. In: *IUCrJ* 5 (2018), pp. 348–353. DOI: 10.1107/S2052252518005237.
- [2] C. Giacovazzo, ed. *Fundamentals of Crystallography*. Oxford University Press, 1985.
- [3] Oleg V. Dolomanov et al. '*OLEX2*: a complete structure solution, refinement and analysis program'. In: *J Appl Crystallogr* 42 (2009), pp. 339–341. DOI: 10.1107/S0021889808042726.
- [4] Christian B. Hübschle, George M. Sheldrick and Birger Dittrich. '*ShelXle*: a Qt graphical user interface for *SHELXL*'. In: *J. Appl. Crystallogr.* 44 (2011), pp. 1281–1284. DOI: 10.1107/S0021889811043202.
- [5] Václav Petříček, Michal Dušek and Lukáš Palatinus. 'Crystallographic Computing System JANA2006: General features'. In: *Z. Kristallogr. - Cryst. Mater.* 229 (2014), pp. 345–352.
- [6] Paul W. Betteridge et al. 'CRYSTALS version 12: software for guided crystal structure analysis'. In: *J. Appl. Crystallogr.* 36 (2003), p. 1487. DOI: 10.1107/S0021889803021800.

- [7] J. Jansen et al. '*MSLS*, a Least-Squares Procedure for Accurate Crystal Structure Refinement from Dynamical Electron Diffraction Patterns'. In: *Acta Crystallogr. A*54 (1998), pp. 91–101.
- [8] A. Authier. *Dynamical Theory of X-Ray Diffraction*. IUCr Monographs on Crystallography. Oxford University Press Inc., New York, 2001.
- [9] L. Reimer and H. Kohl. *Transmission Electron Microscopy*. Physics of Image Formation. Springer-Verlag New York, 2008. ISBN: 978-0-387-34758-5.
- [10] Lukáš Palatinus, Václav Petříček and Cinthia Antunes Corrêa. 'Structure refinement using precession electron diffraction tomography and dynamical diffraction: theory and implementation'. In: *Acta Crystallogr A*71 (2015), pp. 235–244.
- [11] G. M. Sheldrick. 'Crystal structure refinement with *SHELXL*'. In: *Acta Crystallogr. C*71 (2015), pp. 3–8.
- [12] L.-M. Peng. 'Electron atomic scattering factors and scattering potentials of crystals'. In: *Micron* 30 (1999), pp. 625–648.
- [13] Julian Jacob Holstein, Christian Bertram Hübschle and Birger Dittrich. 'Electrostatic properties of nine fluoroquinolone antibiotics derived directly from their crystal structure refinements'. In: *CrystEngComm* 14 (2012), pp. 2520–2531. DOI: 10.1039/C1CE05966A.

- [14] B. Rupp. 'Against Method: Table 1—Cui Bono?' In: *Structure* 26 (2018), pp. 919–923.
- [15] P. Andrew Karplus and Kay Diederichs. 'Linking Crystallographic Model and Data Quality'. In: *Science* 336 (2012), pp. 1030–1033.
- [16] K. Diederichs and A. Karplus. 'Improved R -factors for diffraction data analysis in macromolecular crystallography'. In: *Nat. Struct. Biol.* 4 (1997), pp. 269–275.
- [17] Marco F. Taddio et al. 'Synthesis and Structure–Affinity Relationship of Small Molecules for Imaging Human CD80 by Positron Emission Tomography'. In: *J. Med. Chem.* 62 (2019), pp. 8090–8100. DOI: 10.1021/acs.jmedchem.9b00858.