10th International Symposium and Summer School on Bioanalysis

Workshop in X-ray structure analysis

Practical parts II & III

Ivica Đilović & Dalibor Milić

Zagreb, 7 July 2010

Aims and objectives

You will:

- solve the crystal (and molecular) structure of a metal complex from the corresponding single-crystal X-ray diffraction data
- learn how to visualize and interpret crystal structures
- learn how to retrieve structural information from the databases of crystal structures

You will not:

- learn how to solve structures of biological macromolecules (proteins, nucleic acids, macromolecular complexes) – it is *more complicated!*
- learn how to solve structures from powder (polycrystalline) diffraction data – *it is even more complicated!!!*
- become an expert in X-ray structural analysis (just by attending this workshop)

Workshop materials

http://bioanalysis.chem.pmf.hr/workshop/

- Create a new folder in D:\
  - it is going to be your working folder
- Download files X.hkl and X.ins in your working folder

10th International Symposium and Summer School on Bioanalysis

Workshop in X-ray structure analysis

Practical part II − solving the structure

Dalibor Milić

Zagreb, 7 July 2010

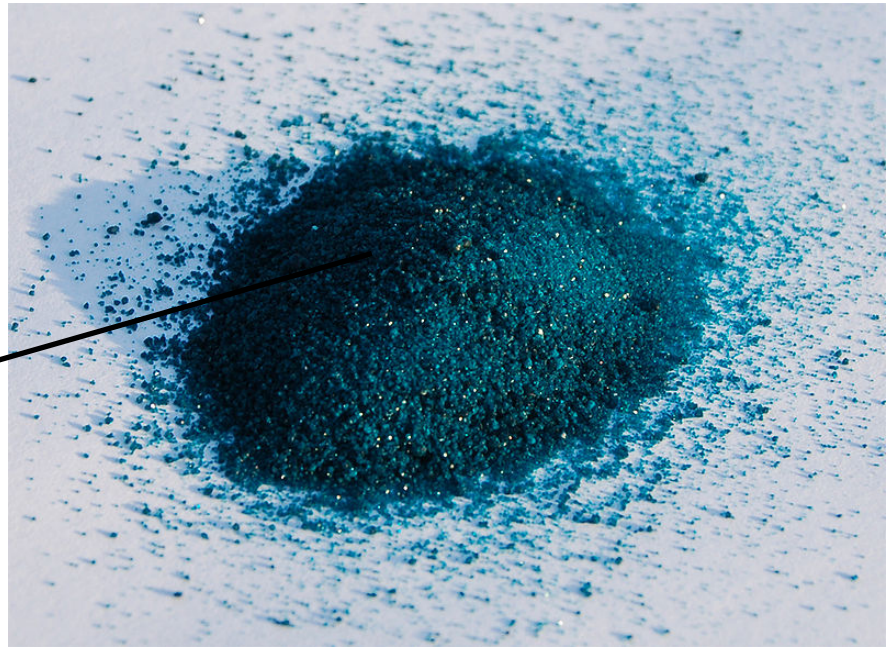
Problem

What is the crystal structure of chemical compound X?

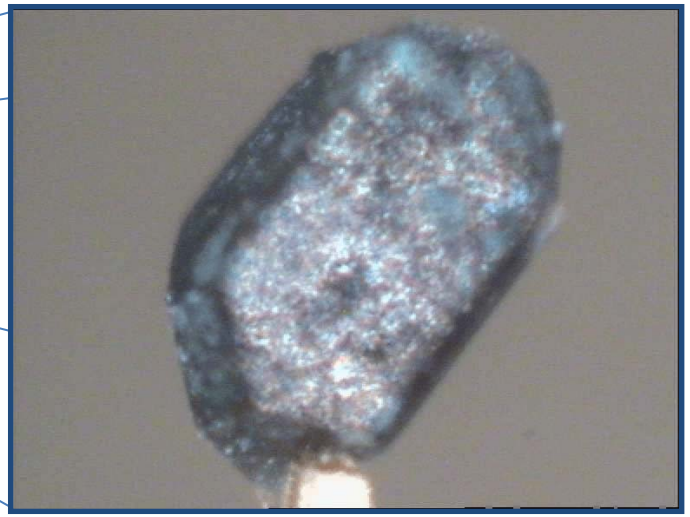
Chemical synthesis

1) Copper(II) hydroxide was dissolved in warm 10 % water solution of acetic acid.
2) The obtained clear solution was cooled off until bluish-green crystalline product **X** appeared.
3) Crystals **X** were filtered off and dried in air.

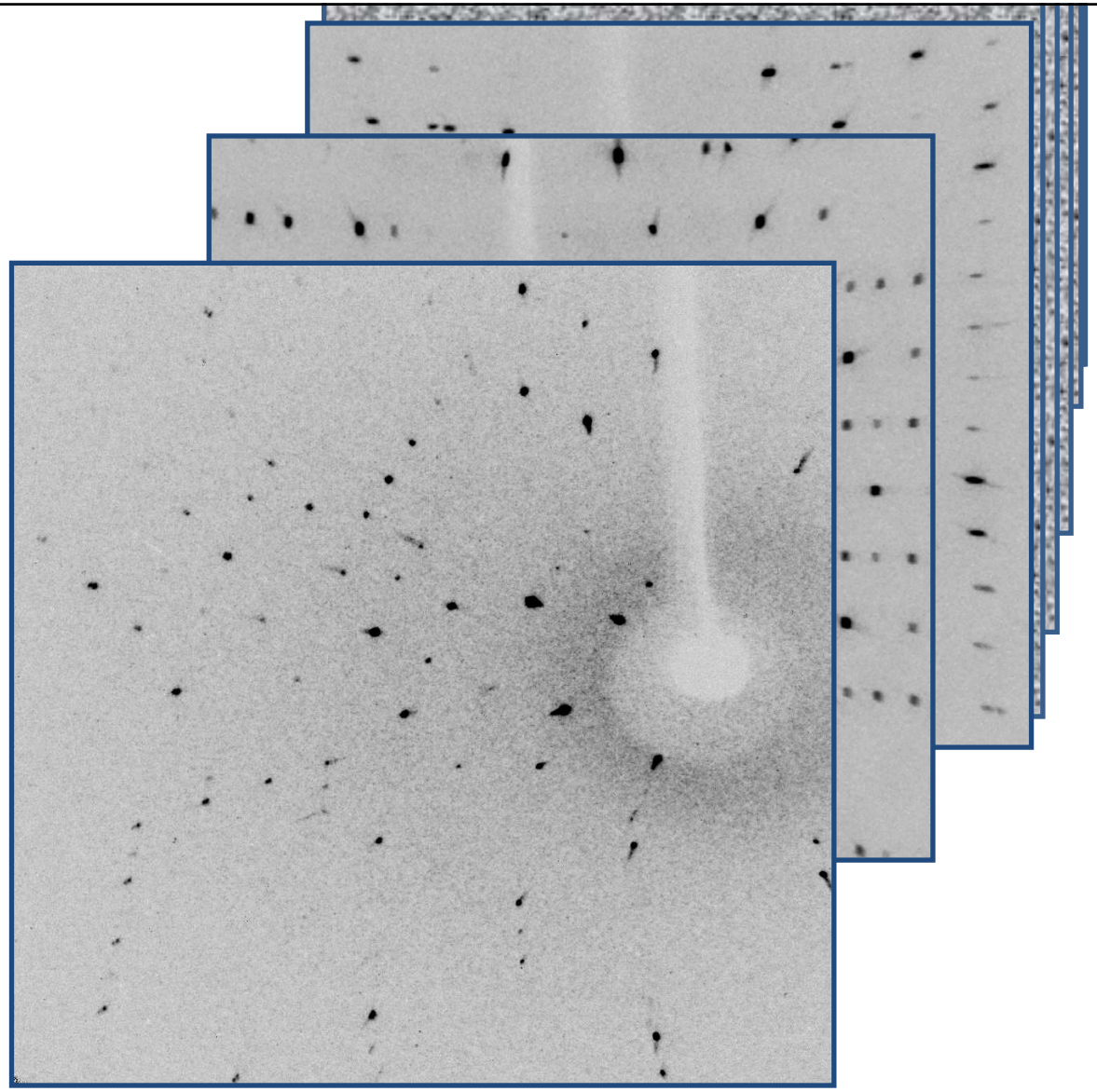
$$Cu(OH)_2(s) + CH_3COOH(aq) \xrightarrow{-\Delta T} \mathbf{X}(s)$$



Single-crystal X-ray data collection

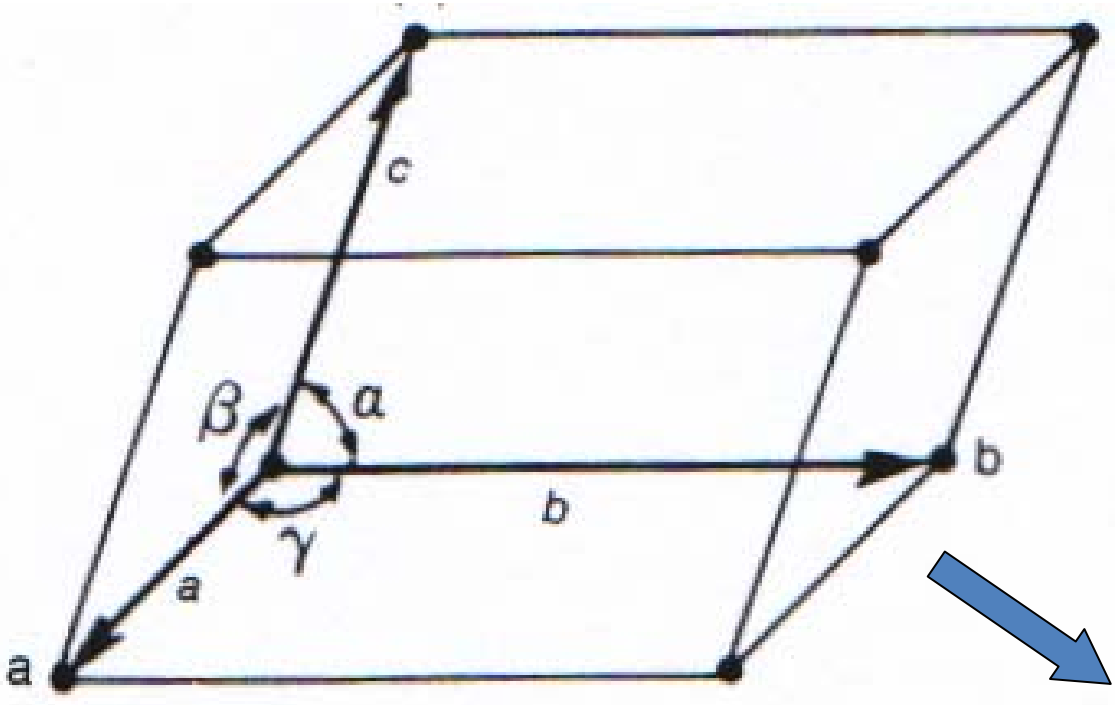


- Crystal size:
  $0.75 \times 0.85 \times 0.90\ mm^3$
- *Oxford Diffraction Xcalibur*
- λ(Mo-Kα) = 0.7107 Å
- *Sapphire 3* CCD-detector
- T = 295 K
- ω- and φ-scans by 1° steps
- exposure time: 1.5 s per step

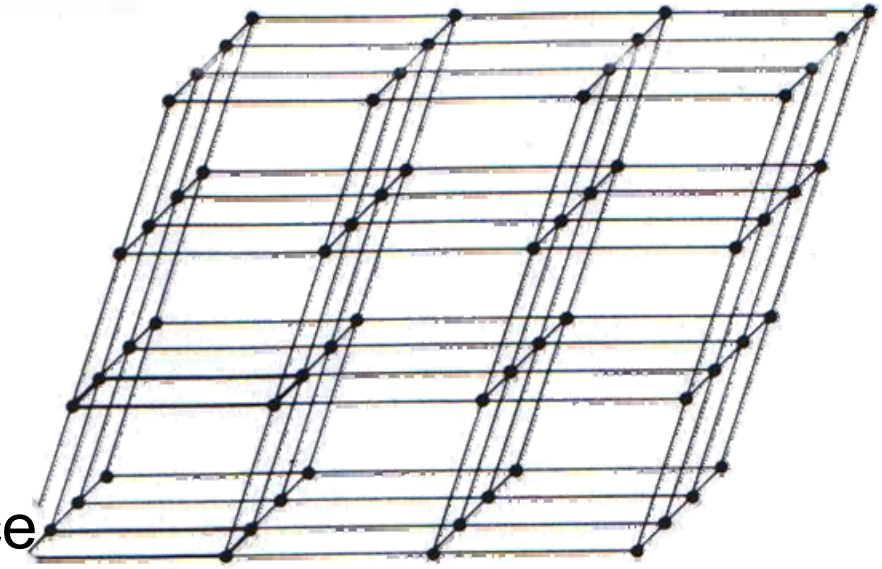


443 diffraction images (frames)

Crystallographic unit cell



Crystallographic unit cell

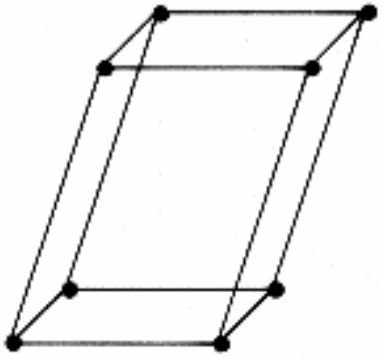


Crystal lattice

Basic crystallographic data

For crystal **X**:

Crystal system: **monoclinic**



$a \neq b \neq c$

$\alpha = \gamma = 90^{o}$

$\beta \neq 90^{o}$

Unit-cell parameters:

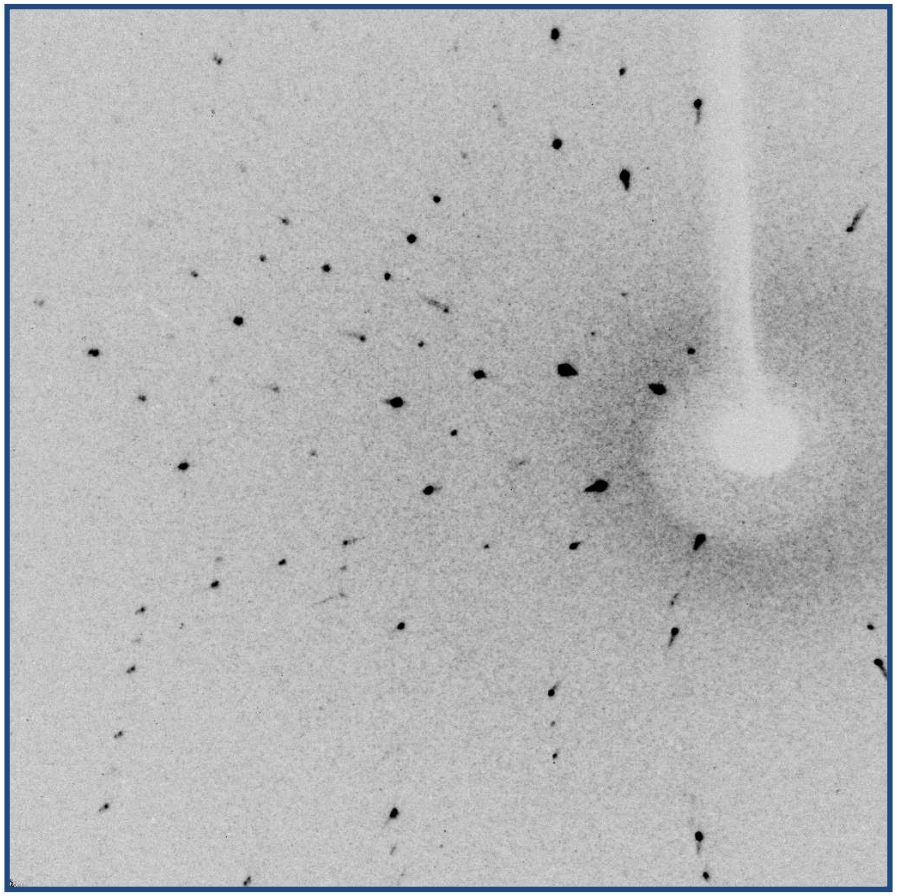
a = 13.1528(6) Å

b = 8.5440(3) Å

c = 13.8381(7) Å

$\alpha = \gamma = 90^{o}$

β = 117.036(6)^{o}



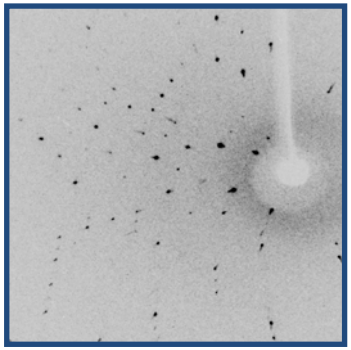
Known unit-cell parameters

Miller indices (*hkl*) can be assigned to each diffraction maximum

Data reduction

Integration of diffraction maxima intensities

$$I_{\text{int}}(hkl) = k\frac{\lambda^3 \Omega}{V^2} I_o LPTE|F(hkl)|^2$$



Applying of various corrections → Corrected intensities I_{corr}

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

Miller indices

Standard uncertainty of I_{corr}(hkl)

h	k	l	$I_{\text{corr}}(hkl)$	$\sigma[I_{\text{corr}}(hkl)]$
1	2	-1	19241.8	535.907
-1	2	1	18898.0	469.019
-1	2	0	3035.70	278.012
1	2	0	3167.05	359.579
1	-2	0	3317.19	281.786
-1	2	-1	2915.53	329.314
1	2	1	2987.65	340.650
...				

Data reduction

For crystal **X** (have a look at **X.hkl**):

4382 measured diffraction maxima (“reflections”)

1624 unique diffraction maxima (because of the crystal symmetry)

resolution(A)	# measured	# kept	# unique	average redundancy	mean F2	mean F2/sig(F2)	Rint	Rsigma
inf-1.70	666	592	162	3.7	5597159.76	114.01	0.019	0.020
1.69-1.35	651	590	162	3.6	3646108.60	74.97	0.023	0.021
1.35-1.17	555	505	162	3.1	2015393.58	44.66	0.027	0.026
1.17-1.07	512	479	162	3.0	1432040.76	32.14	0.031	0.029
1.06-0.99	479	447	162	2.8	1148818.32	24.59	0.032	0.033
0.99-0.93	414	393	162	2.4	1083019.98	23.24	0.048	0.046
0.93-0.88	423	400	162	2.5	868757.14	19.04	0.035	0.035
0.88-0.84	377	362	162	2.2	616733.00	13.51	0.043	0.044
0.84-0.81	368	354	162	2.2	629508.28	13.11	0.043	0.043
0.81-0.75	263	260	166	1.6	438961.95	9.16	0.054	0.055
inf-0.75	4708	4382	1624	2.7	2102554.44	44.13	0.026	0.027
inf-0.80	4539	4214	1500	2.8	2166848.92	45.48	0.026	0.026

Crystal space group

Systematic absences for **X**
(not obseved diffraction maxima):

hkl : $h + k = 2n + 1$
*h*0*l* : $h, l = 2n + 1$
0*kl* : $k = 2n + 1$
*hk*0 : $h + k = 2n + 1$
0*k*0 : $k = 2n + 1$
*h*00 : $h = 2n + 1$
00*l* : $l = 2n + 1$

} $2n + 1$ = any odd number

→ 2 space groups with such systematic absences:

Cc (no. 9)

non-centrosymmetric

C2/c (no. 15)

centrosymmetric

Crystal space group

E-statistics

Normalized structure factor: $|E_{hkl}| = \dfrac{|F_{hkl}|}{\sqrt{\varepsilon \sum_j f_j^2 T^2}}$

Centrosymmetric random structure: $\langle |E^2 - 1| \rangle = 0.968$

Non-centrosymmetric random structure: $\langle |E^2 - 1| \rangle = 0.736$

For **X**: $\langle |E^2 - 1| \rangle = 0.965$

Crystal structure of **X** is most probably centrosymmetric!

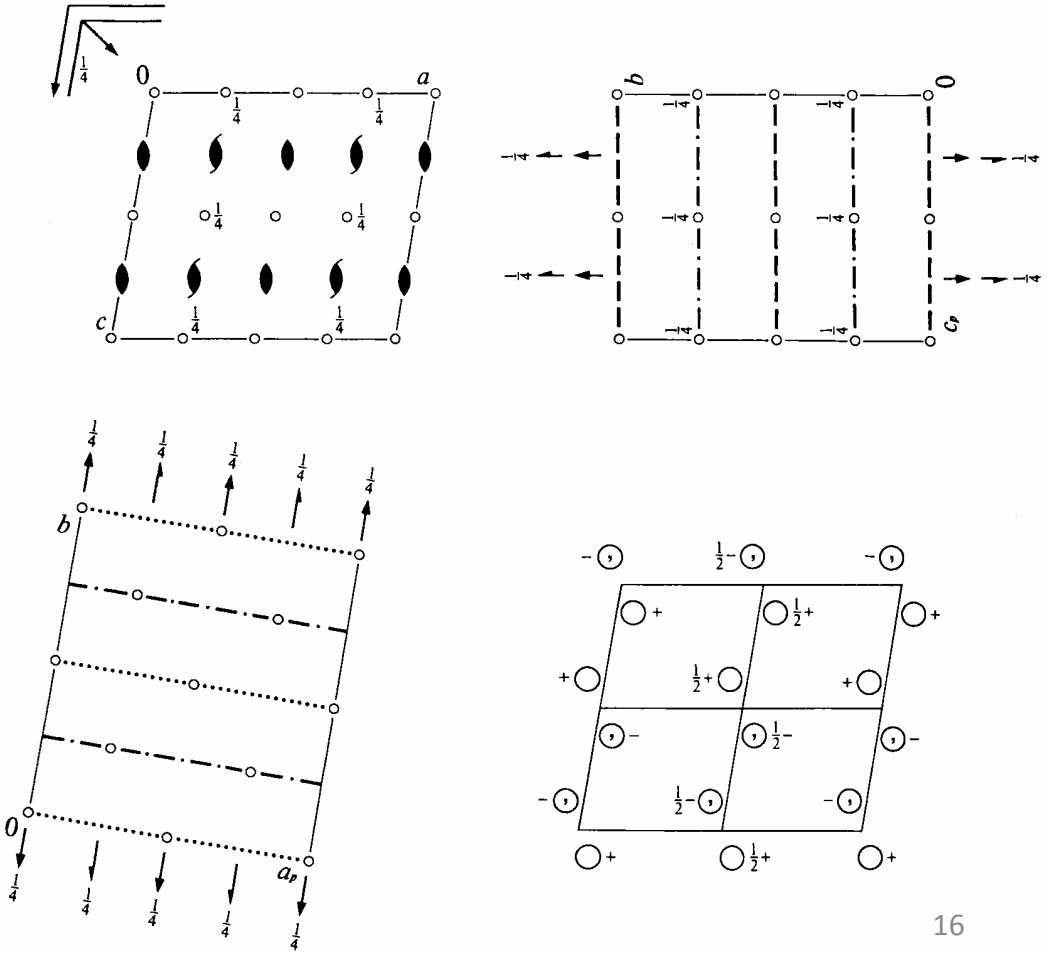
Crystal space group

~~***Cc*** (no. 9)~~

~~non-centrosymmetric~~

C2/c (no. 15)

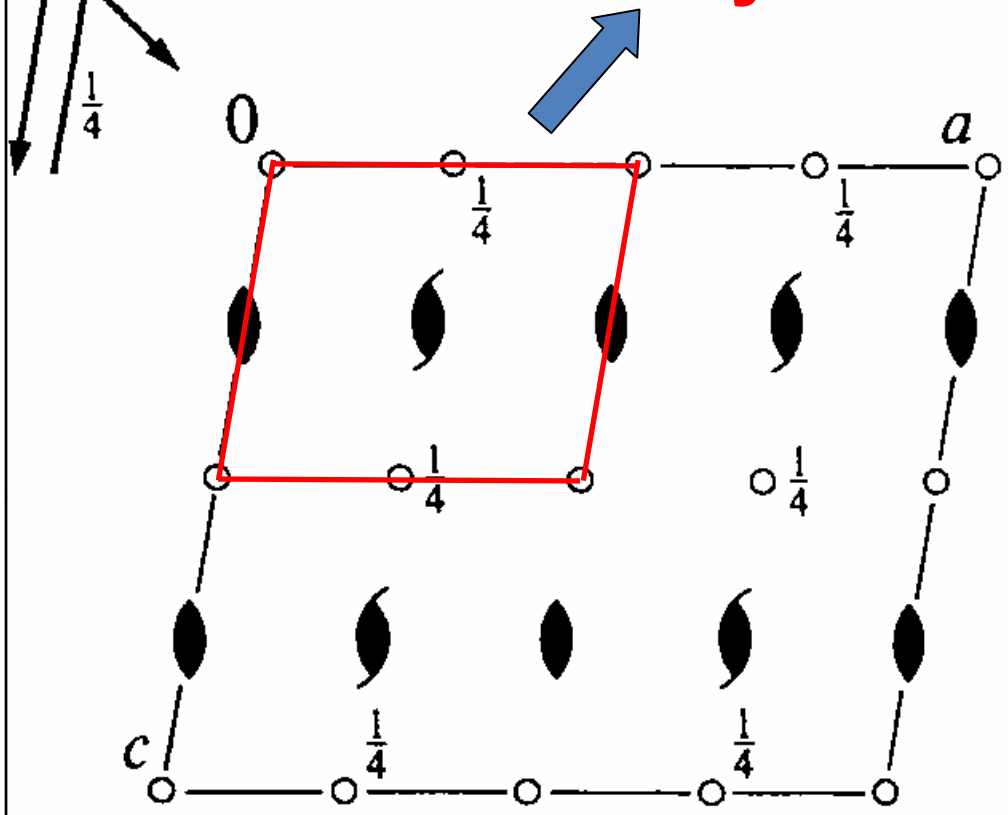
centrosymmetric



Crystal space group

C2/c

Asymmetric unit



- the smallest part of the unit cell from which, by application of all symmetry operations of the space group, the whole space is filled

For crystallographic analysis it is enough to use just an asymmetric unit

- an asymmetic unit can be made of:
  - a) just one whole chemical entity (*e.g.* a molecule)
  - b) more (different or same) chemical entities
  - c) the part of a chemical entity (*e.g.* one half of a molecule)

WinGX

- a software package for X-ray structure analysis
- free download from:
  http://www.chem.gla.ac.uk/~louis/software/wingx/

1. Start a new project in WinGX (**X.ins**)!
2. Check the space group of **X** in WinGX!
3. Open the INS file and have a look at it!

```
TITL X in C2/c
CELL  0.71073  13.1528   8.5440  13.8381   90.000  117.036   90.000
ZERR     4.00   0.0006   0.0003   0.0007    0.000    0.006    0.000
LATT   7
SYMM  - X,   Y, 1/2 - Z
SFAC  C    H    O    CU
UNIT  32   64   40   8
HKLF    4
END
```

Phase problem

Reciprocal space

$$\boldsymbol{F}(hkl) = \sum_{j=1}^{N} f_j(hkl)\exp[2\pi\mathrm{i}(hx_j + ky_j + lz_j)]$$

FT^{-1} ⇅ FT

Real space

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l \boldsymbol{F}(hkl)\exp[-2\pi\mathrm{i}(hx + ky + lz)]$$

Do not forget: $\boldsymbol{F}(hkl) = |F(hkl)|\exp[\mathrm{i}\varphi(hkl)]$!

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l |F(hkl)|\exp[-2\pi\mathrm{i}(hx + ky + lz) + \mathrm{i}\varphi(hkl)]$$

$I \propto |F(hkl)|^2$

PHASE PROBLEM ?

Solving the phase problem

Patterson method

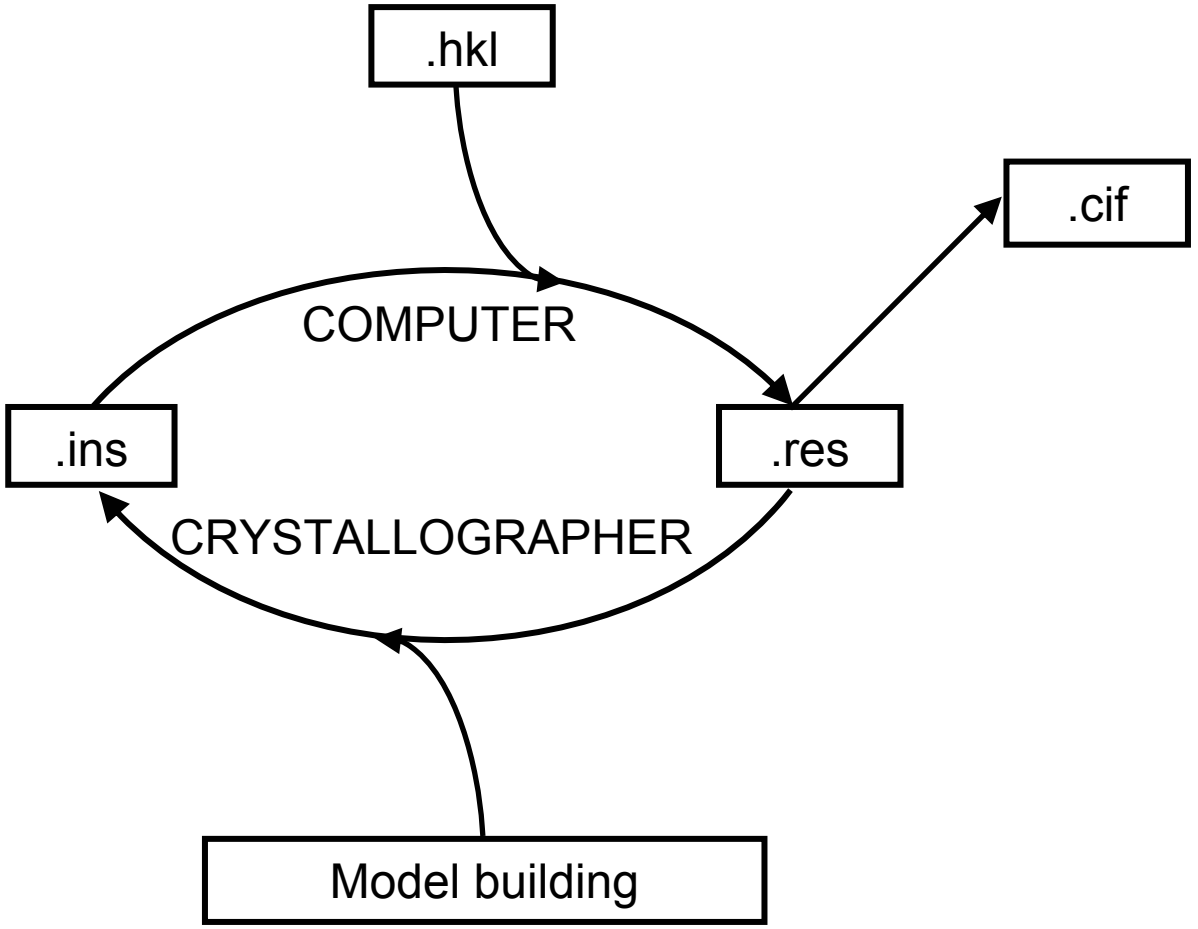
Direct methods

Charge flipping

...

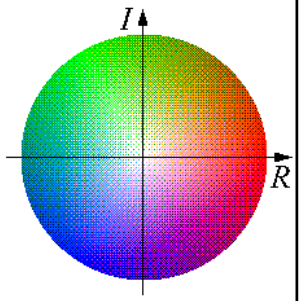
SHELX

- a group of programs for solving and refinement of crystal structures
- freely available at http://shelx.uni-ac.gwdg.de/SHELX/

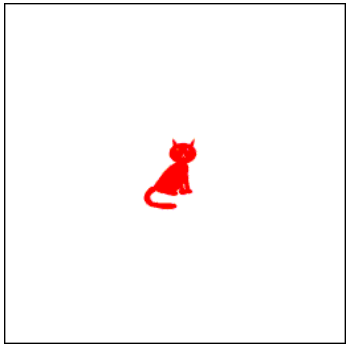


Fourier's recycling

Kevin Cowtan's Book of Fourier

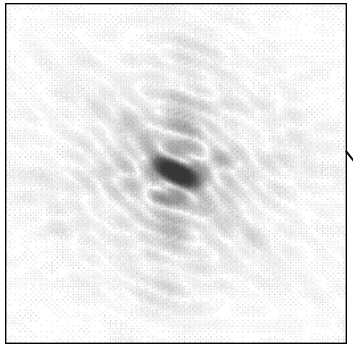


REAL STRUCTURE

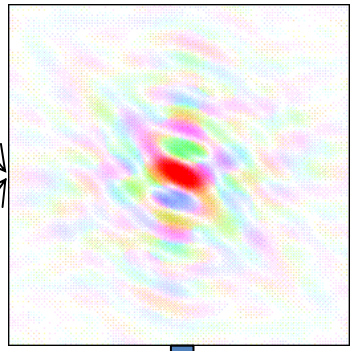


diffraction experiment

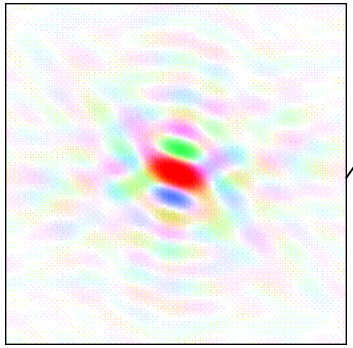
STRUCTURAL AMPLITUDES



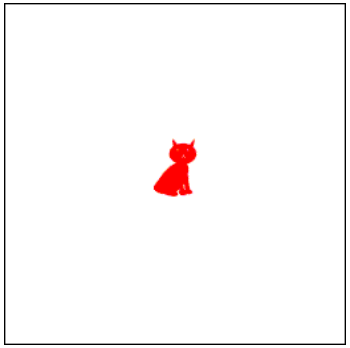
amplitudes



phases



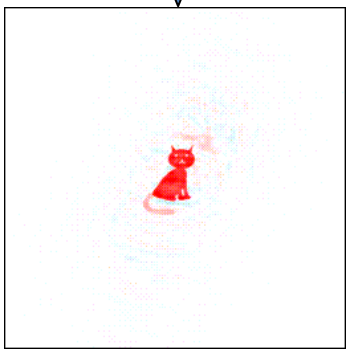
FT calculation



MODEL

STRUCTURAL AMPLITUDES & PHASES

FT $^{-1}$ calculation



Structure solving

Solve the structure of X by direct methods!

Vizualize the structure solution by SXGRAPH!

Build the atomic model from Q-peaks (maxima in difference electron density maps)!

```
MOLE    1
CU1     4   0.0477   0.0832   0.4521 11.000000   0.05
Q1      1  -0.0557   0.2531   0.5436 11.000000   0.05   259.07
Q2      1   0.1487  -0.0785   0.3646 11.000000   0.05   194.41
Q3      1   0.1861   0.0854   0.6013 11.000000   0.05   189.15
Q4      1  -0.0977   0.0717   0.3189 11.000000   0.05   145.61
Q5      1   0.1794   0.0187   0.6757 11.000000   0.05   113.28
Q6      1  -0.0043   0.2710   0.5026 11.000000   0.05   113.16
Q7      1   0.0950  -0.1235   0.4194 11.000000   0.05   107.67
Q8      1   0.1283   0.2112   0.3681 11.000000   0.05   105.39
Q9      1   0.0000  -0.0487   0.2500 10.500000   0.05    70.42
Q10     1   0.0993   0.2766   0.4192 11.000000   0.05    64.41
Q11     1   0.2898   0.0326   0.7928 11.000000   0.05    62.98
Q12     1   0.0000   0.1988   0.2500 10.500000   0.05    59.34
Q13     1  -0.1054   0.4080   0.5831 11.000000   0.05    57.75
Q14     1  -0.1648   0.2582   0.3986 11.000000   0.05    50.07
Q15     1   0.0040   0.4324   0.4953 11.000000   0.05    47.65
Q16     1  -0.1338   0.3497   0.6242 11.000000   0.05    43.86
Q17     1   0.1775   0.1514   0.6730 11.000000   0.05    42.69
MOLE    2
HKLF    4
END
```

Refinement

Minimization of function: $\sum w(Y_o - Y_c)^2$

- Y is usually $|F|^2$
- Y_o is observed value and Y_c is calculated value based on the actual model
- w is weighting parameter which is different for different *hkl*

• **least-squares method** - the most common refinement method in "small" molecule crystallography

• **atomic coordinates** (3 per atom), **atomic displacement parameters** (1 per atom, if it is isotropic; 6 per atom, if it is anisotropic), global **scale factor** between measured and calculated intensities

• sometimes additional **restraints** are added in the refinement

$$\sum w(Y_o - Y_c)^2 + \sum w_r(r_t - r_c)^2$$

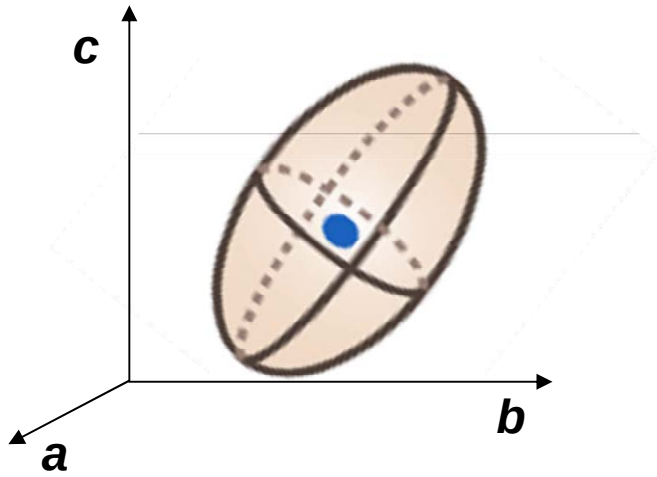
Refine the structural model of **X** by SHELXL!

Atomic displacement parameter

isotropic

$$T(\text{iso};hkl) = \exp\left[\frac{-8\pi^2 U_j \sin^2\theta}{\lambda^2}\right]$$

anisotropic



$$T(\text{aniso};\ hkl) = \exp[-2\pi^2(U_{11}h^2a^{*2} + U_{22}k^2b^{*2} + U_{33}l^2c^{*2} + 2U_{12}hka^*b^* + 2U_{13}hla^*c^* + 2U_{23}klb^*c^*)]$$

INS file – atomic coordinates section

atom name type

fractional coordinates: *x* *y* *z*

site occupancy factor

anisotropic atomic displacement parameters

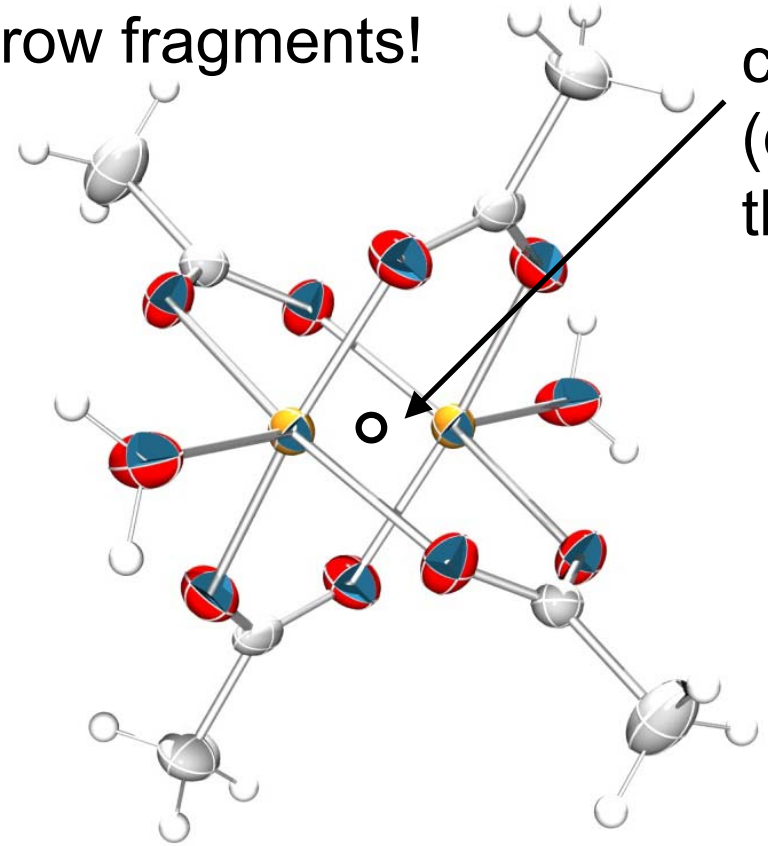
```
C1    1    0.182341    0.020876    0.680357    11.00000    0.02356    0.02529 =
       0.02135   -0.00270    0.00861    0.00185
C2    1    0.284465    0.035900    0.788864    11.00000    0.03667    0.06448 =
       0.02791    0.00137    0.00597   -0.01319
C3    1    0.064761   -0.249293    0.449754    11.00000    0.02053    0.02431 =
       0.01991   -0.00563    0.00027    0.00338
C4    1    0.104026   -0.394541    0.416166    11.00000    0.03921    0.03287 =
       0.04444   -0.01066    0.01566    0.00635
O1    3    0.099784   -0.060915    0.675734    11.00000    0.02571    0.03845 =
       0.02045    0.00333    0.00748   -0.00300
O2    3    0.183973    0.088391    0.600757    11.00000    0.02257    0.03450 =
       0.02042    0.00168    0.00657   -0.00355
O3    3    0.007686   -0.263932    0.500265    11.00000    0.03828    0.02655 =
       0.03951   -0.00146    0.02187    0.00271
O4    3    0.094874   -0.121007    0.425163    11.00000    0.03553    0.03043 =
       0.03857   -0.00359    0.02150    0.00254
O5    3    0.124263    0.207365    0.367223    11.00000    0.05091    0.04138 =
       0.03522   -0.01220    0.03005   -0.02112
CU1   4    0.049890    0.084004    0.455034    11.00000    0.02011    0.02218 =
       0.01933   -0.00003    0.00967   -0.00093
```

ORTEP

- Software for visualization of displacement ellipsoids
- Free download for Windows from: http://www.chem.gla.ac.uk/~louis/software/ortep3/

Visualize ADPs for the structural model of **X** by ORTEP!

Grow fragments!



crystallographic inversion center (center of symmetry) in the middle of the dinuclear complex

Structure evaluation

- **structure has to be chemically sound!!!**

- some parameters for evaluation – they measure fitting of structural model to experimental data

$$R = \frac{\sum \left| |F_o| - |F_c| \right|}{\sum |F_o|}$$

F_o – observed structure factor
F_c – calculated structure factor

***R*-factor** is usually lower than 5% for correctly determined structures

Weighted *R*-factor:

$$wR = \left[\frac{\sum w(Y_o - Y_c)^2}{\sum w Y_o^{\,2}} \right]^{1/2}$$

Goodnes of fit:

$$S = \left[\frac{\sum w(Y_o - Y_c)^2}{N - P} \right]^{1/2}$$

It is expected to be $S \approx 1$ for correct structural models

No. of data

No. of refined parameters

Hydrogen atoms

- X-ray diffraction gives information about electron density
- Positions of hydrogen atom nuclei do not correspond to maxima in electron density
- Hydrogen atoms (just 1 e^-) are poor X-ray scatterers

H atoms are frequently built in ideal calculated positions or treated with restraints during refinement

For **X** use:

HFIX 137 – generation of idealised methyl group, torsion angle is found by fitting to the electron density

```
DFIX   0.89 0.01 O5 H51 O5 H52
DANG 1.42 0.02 H51 H52
```

} Restraints for a water molecule

CIF file

- Crystallographic Information File – standard text format for exchange of crystallographic data
- for generation of CIF file after refinement just write in INS:

  OMIT −1.00 180.00 (instead of: OMIT 4.00 180.00)
  ACTA

That's it!

Reference

- the first structural report for X:

  J. N. van Niekerk & F. R. L. Schoening, *Acta Crystallogr.* **6** (1953) 227−232.