The value of the availability of raw data for modulated structures

Or

How obviously good practice may be really helpful

Michal Dušek
Fyzikální ústav AV ČR, Na Slovance 2, 182 21 Praha

Storing of raw data is a good practice for any kind of structures.

For modulated structures, there are two specil points where raw data can help. With raw data, we can check

1) Satellite order

2) Twinning and number of modulation vectors

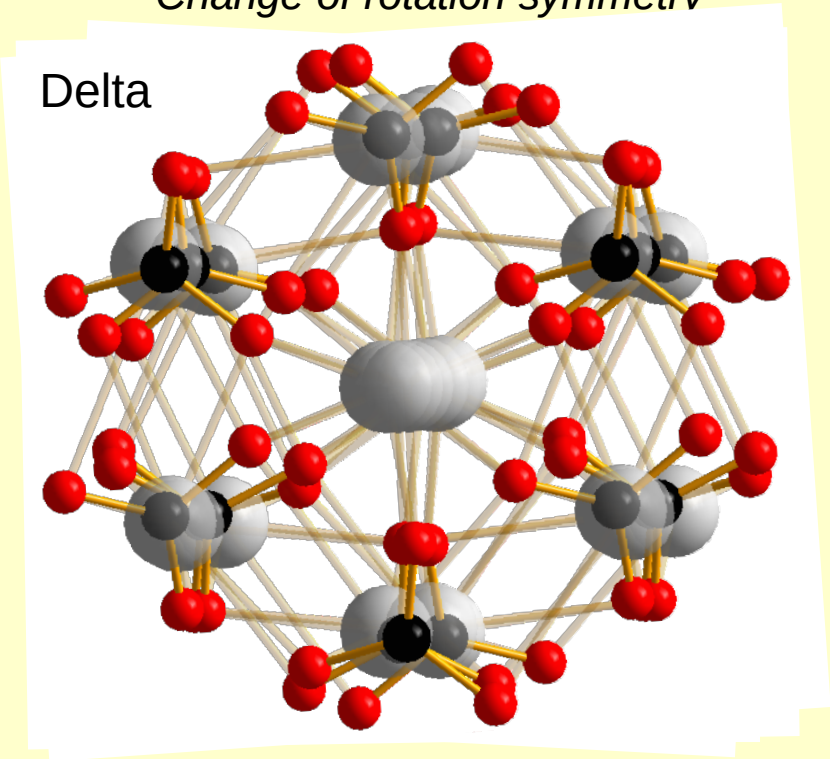
Examples that follow illustrate the importance of storing raw data.

Sodium carbonate – structure model improvement by 20 years of waiting

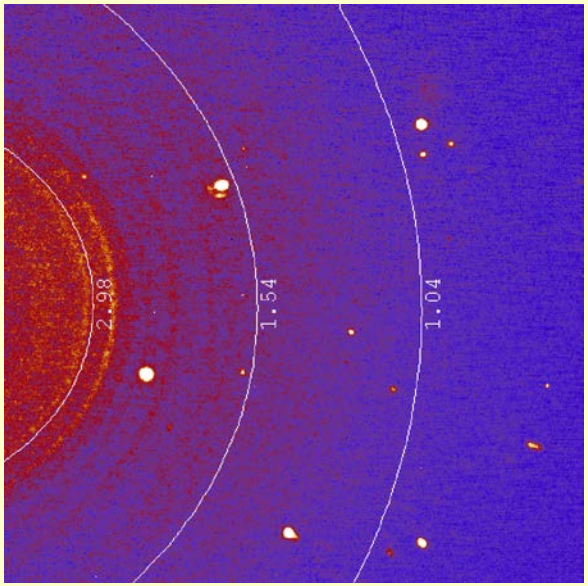
Alpha	757K	Beta	628K	Gamma	170K	Delta
9.02, 5.21, 6.50 90, 90, 90 $P6_3/mmm$		8.98, 5.25, 6.21 90, 90.33, 90 C2/m		8.92, 5.25, 6.05 90, 101.35, 90 C2/m(α0γ) q=(0.182,0,0.322)		8.90, 5.24, 6.00 90, 101.87, 90 C2/m(α0γ) q=(1/6,0,1/3)

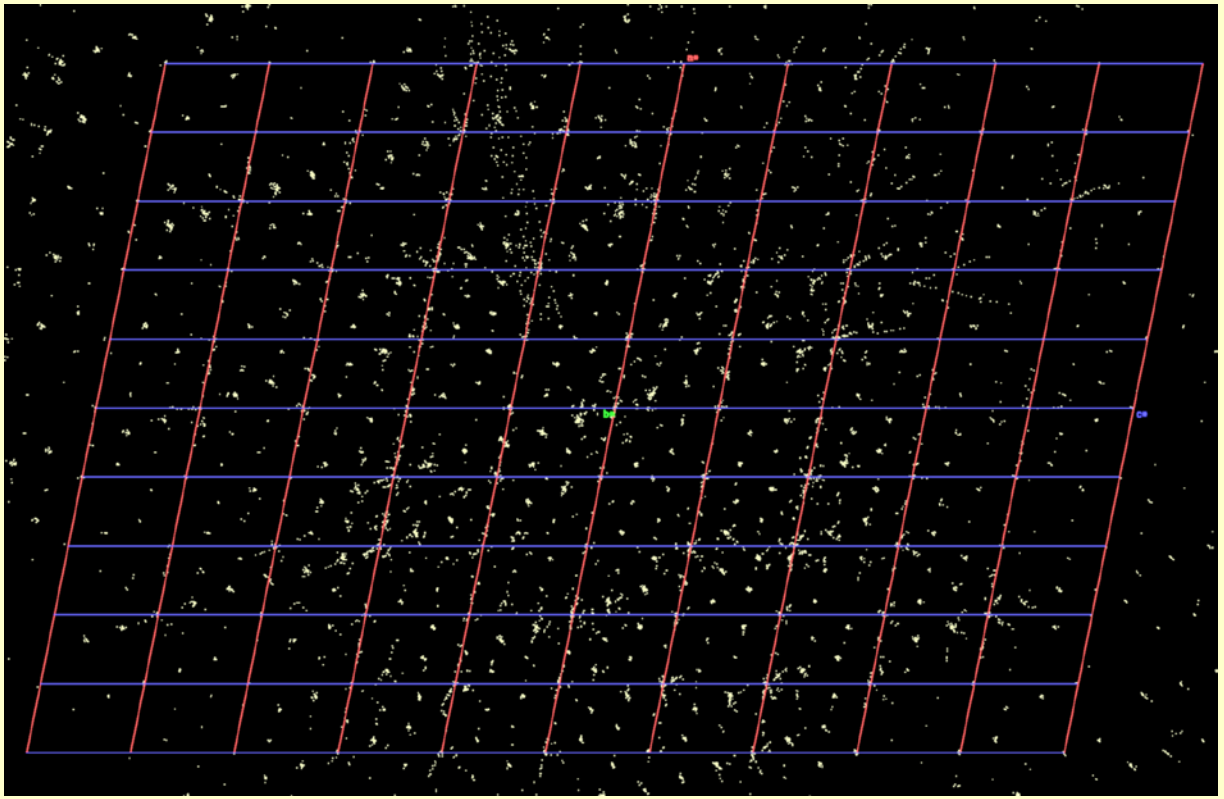
Change of rotation symmetrv

Change of translation symmetry



Data collection: 1999 Lausanne, Kuma CCD diffractometer
Scan width 0.5° in ω, full sphere, 111 hours

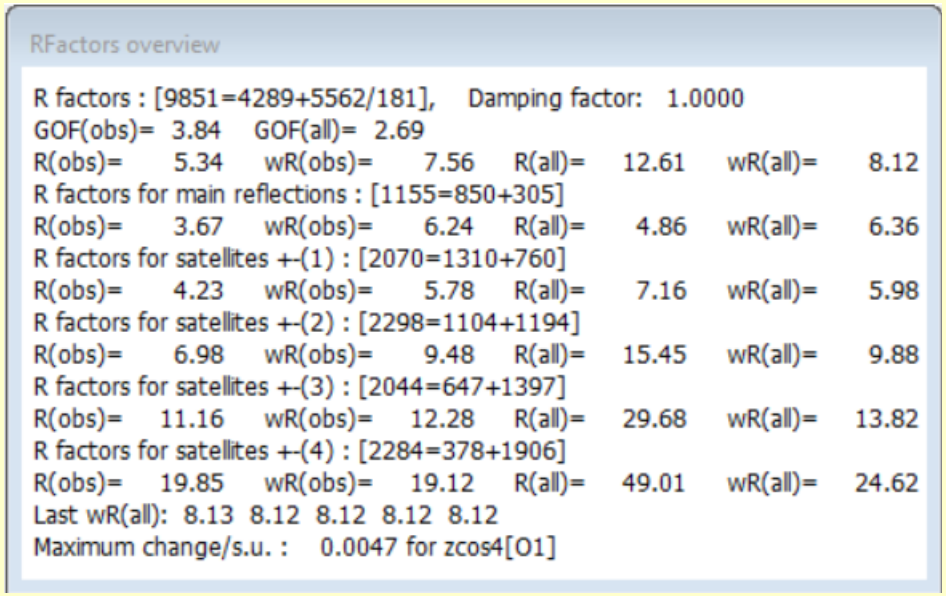




Large sample
Moderate data quality

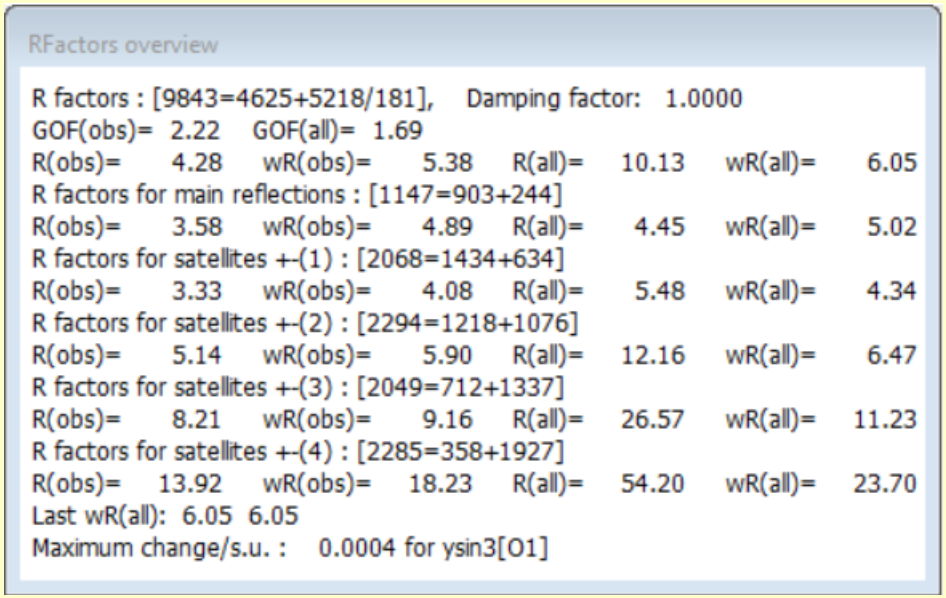
Projection of peak positions clearly reveals satellites

Data reduction by Crysalis 2001 / refinement with Jana2020



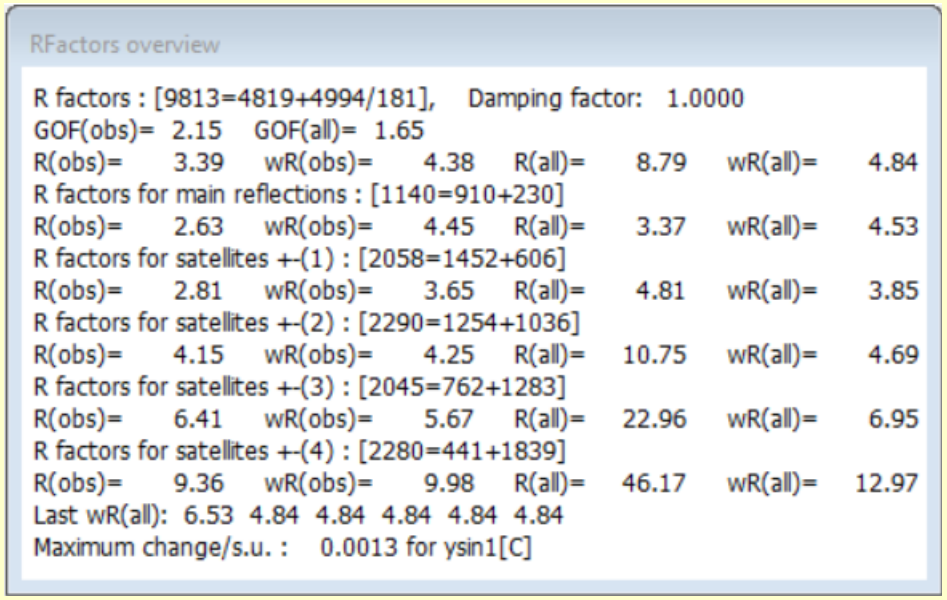
Good results. The poor fit for 4th order satellites corresponds with the fact they are weak

Data reduction by Crysalis 2010 / refinement with Jana2020



Nine years later:
GOF 3.84 → 2.22
R(all) 5.34 → 4.28
R(main) 3.67 → 3.58
R(sat4) 19.85 → 13.92

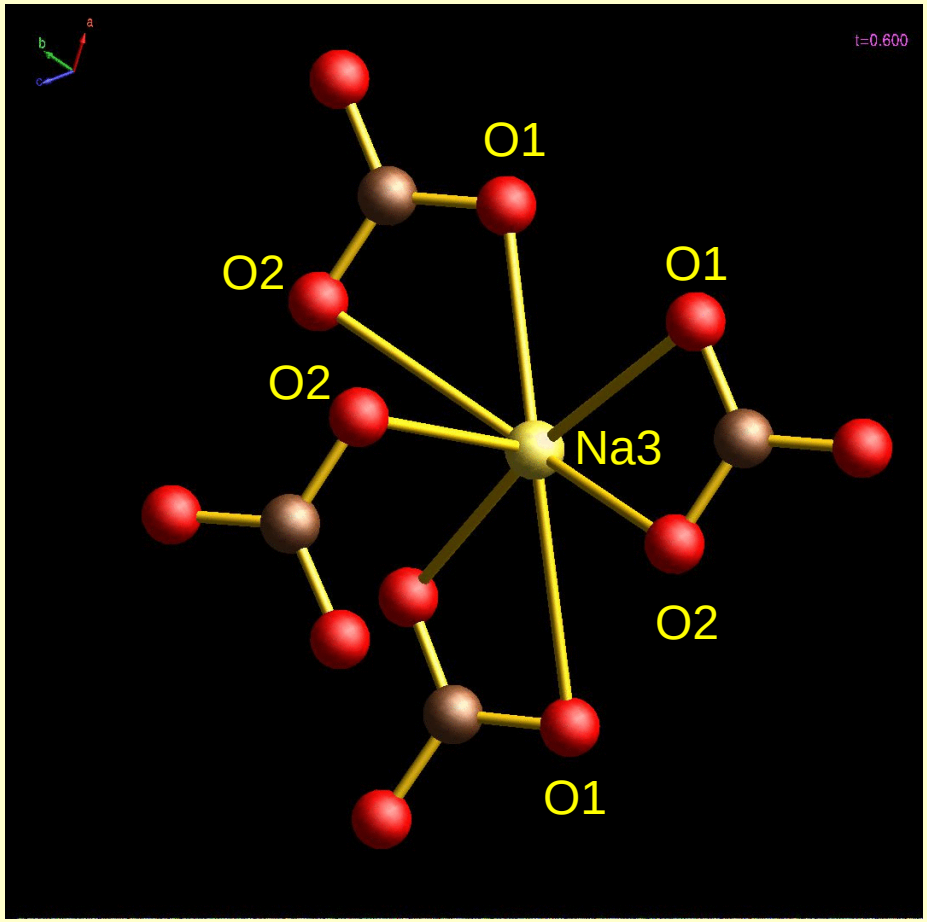
Data reduction by Crysalis 2021 / refinement with Jana2020



Twenty years later:
GOF 3.84 → 2.22 → 2.15
R(all) 5.34 → 4.28 → 3.39
R(main) 3.67 → 3.58 → 2.63
R(sat4) 19.85 → 13.92 → 9.36

The most modulated bond

Na3-O1 (Å)	2001	2010	2021
Average	2.632(2)	2.6323(18)	2.6313(15)
Min	2.403(3)	2.4052(19)	2.4029(16)
Max	2.909(3)	2.8979(19)	2.9063(16)

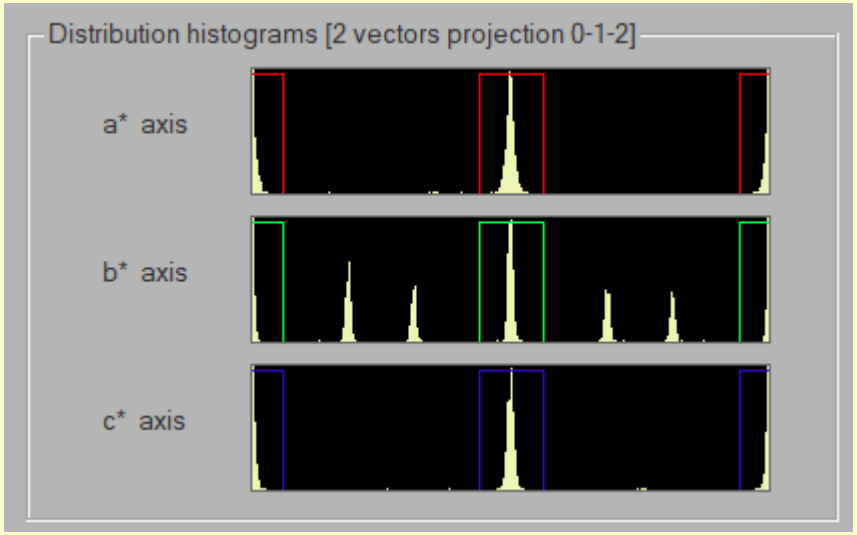


$Y(PO_4)_3$ reveals additional satellites

Jana2006 cookbook example 5.1.” Simple modulated structure with crenel”
Sample measure in 2007 with OD diffractometer, detector Sapphire2.

Symmetry C2/c(0β0)s0
Modulation vector (0 0.371 0)
Satellites up to the second order (very small intensity)

Results of automatic peak hunting

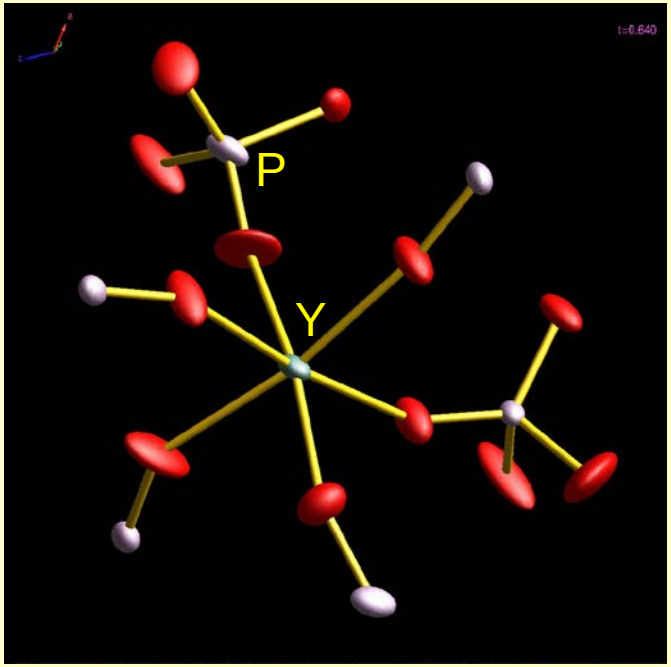


Good refinement fit

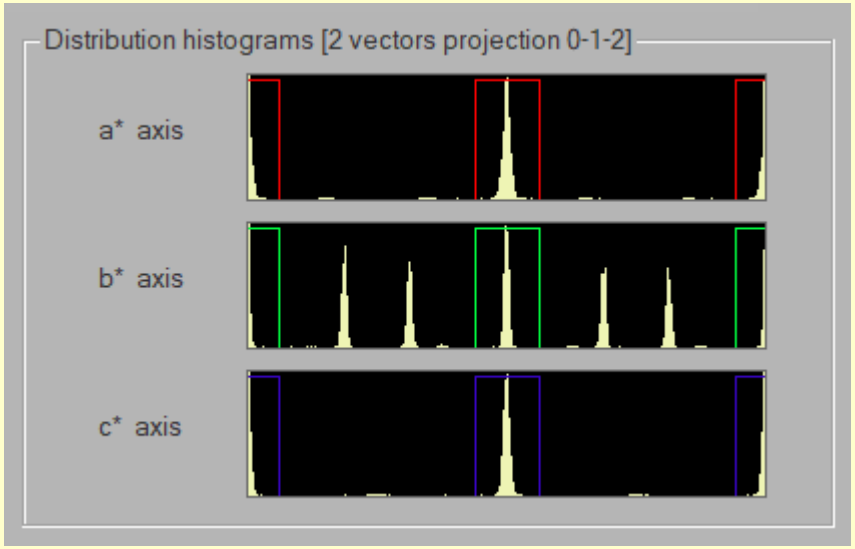
RFactors overview

R factors : [3952=2068+1884/297], Damping factor: 1.0000
GOF(obs)= 1.61 GOF(all)= 1.21
R(obs)= 3.54 wR2(obs)= 6.55 R(all)= 7.33 wR2(all)= 7.05
R factors for main reflections : [794=655+139]
R(obs)= 2.51 wR2(obs)= 5.73 R(all)= 3.17 wR2(all)= 5.80
R factors for satellites +-(1) : [1570=1043+527]
R(obs)= 3.66 wR2(obs)= 6.65 R(all)= 5.89 wR2(all)= 6.92
R factors for satellites +-(2) : [1588=370+1218]
R(obs)= 10.02 wR2(obs)= 22.11 R(all)= 23.61 wR2(all)= 28.11
Last wR2(all): 7.05
Maximum change/s.u. : 0.0193 for U22cos2[O5]

Troubles with ADP parameters

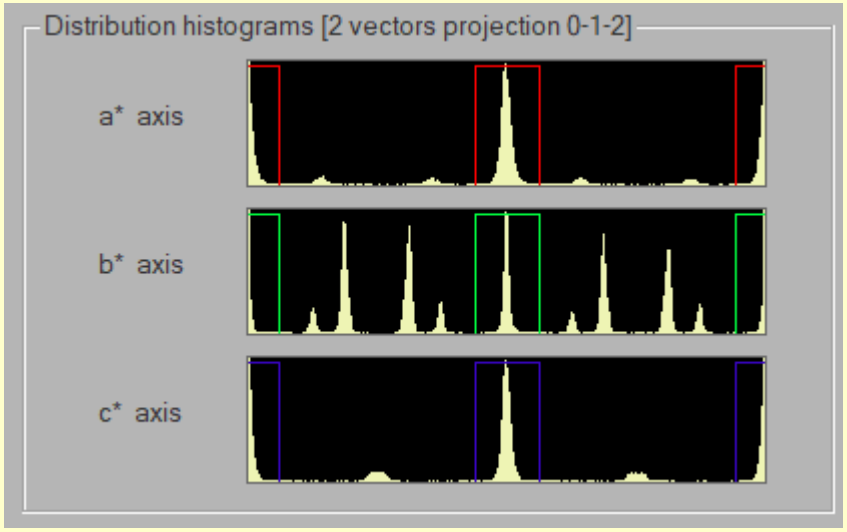


Crysalis peak hunting, type “Smart”, indicates additional spots

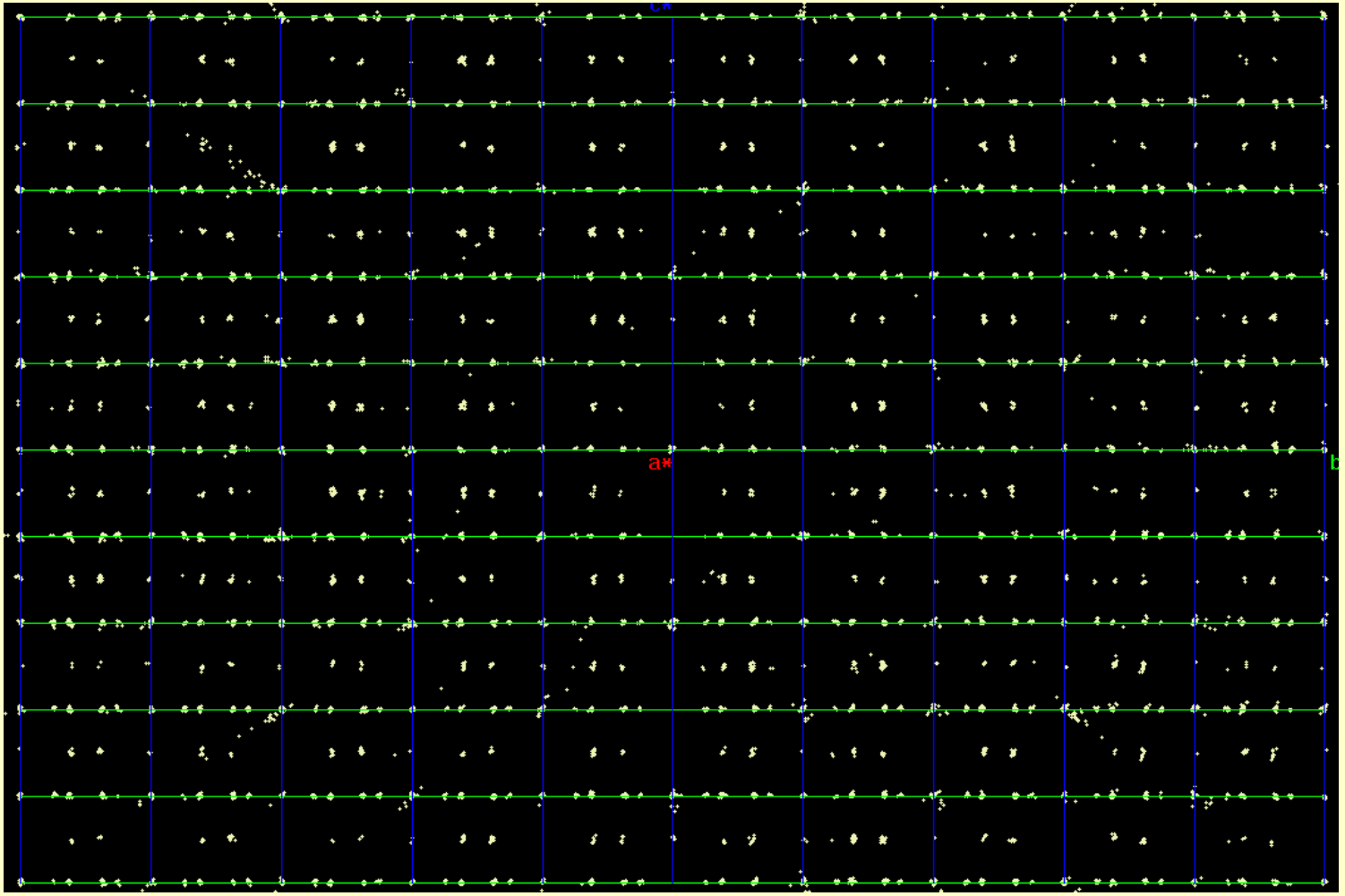


2016: new measurement with SuperNova/Atlas2

There are additional satellites with intensity below the one of the 2nd order satellites



There are two q-vectors
(0 0.371 0)
(0.287 0.389 0.52)
The structure is probably (3+2)d, not yet solved

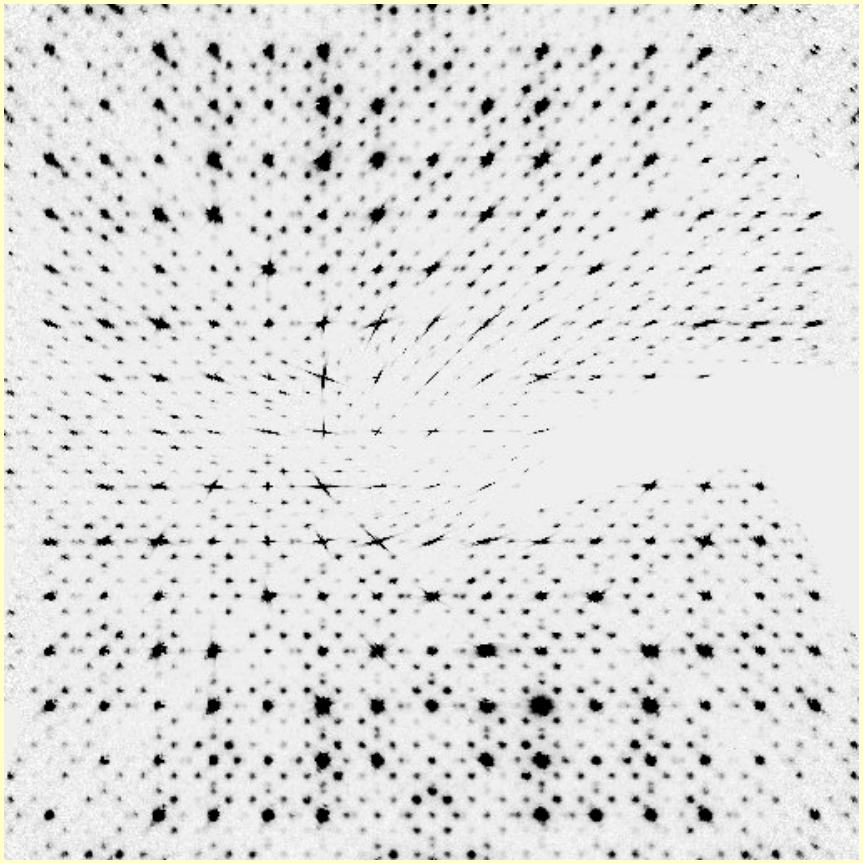


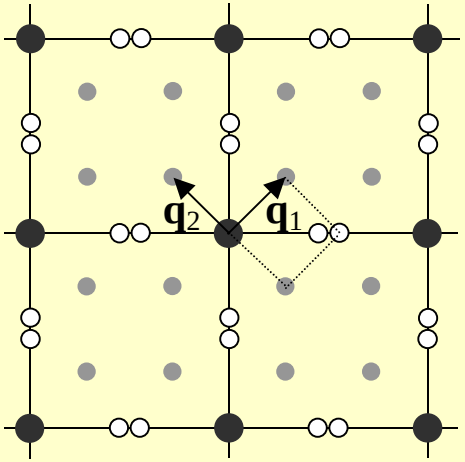
Cross satellites help to decide between a twin and (3+2)d structure

Natural melilite from San Venanzo, Umbria, Italy
Formula: $(Ca_{1.89}Sr_{0.01}Na_{0.08}K_{0.02})(Mg_{0.92}Al_{0.08})(Si_{1.98}Al_{0.02})O_7$

Superspace group: $P\text{-}42_1m(\alpha\alpha0, \alpha\text{-}\alpha0)$ + translation part

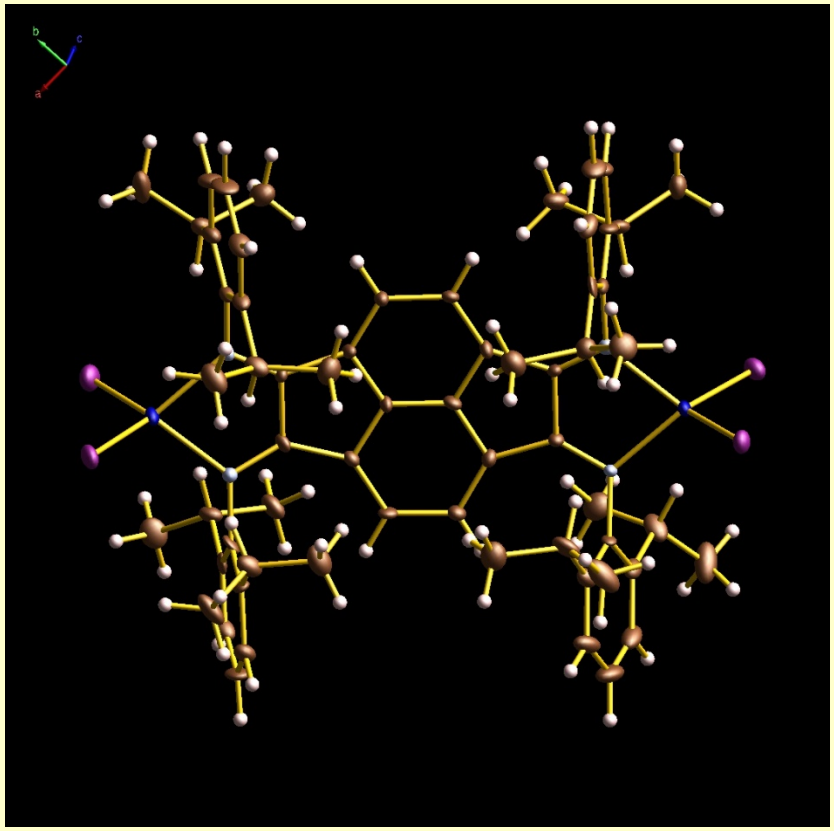
Cell parameters: a=7.860 (1), c=5.024 (1) Å
q vectors : 0.2815(3)(**a***+**b***) , 0.2815(3)(-**a***+**b***)





l=1

Data quality may not be clear from R and R_{int} factors

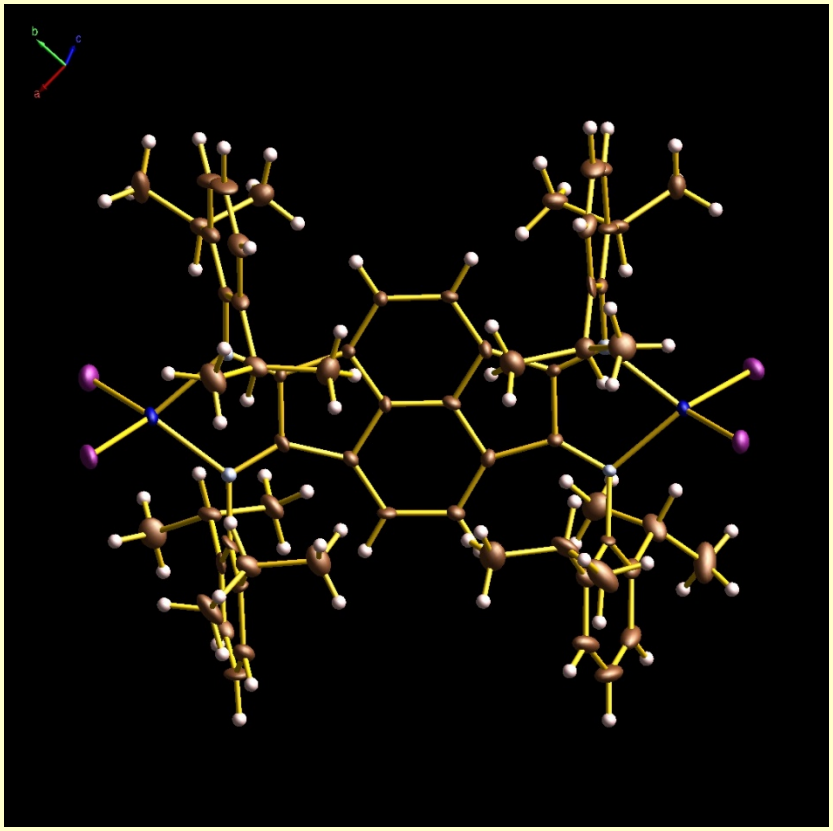


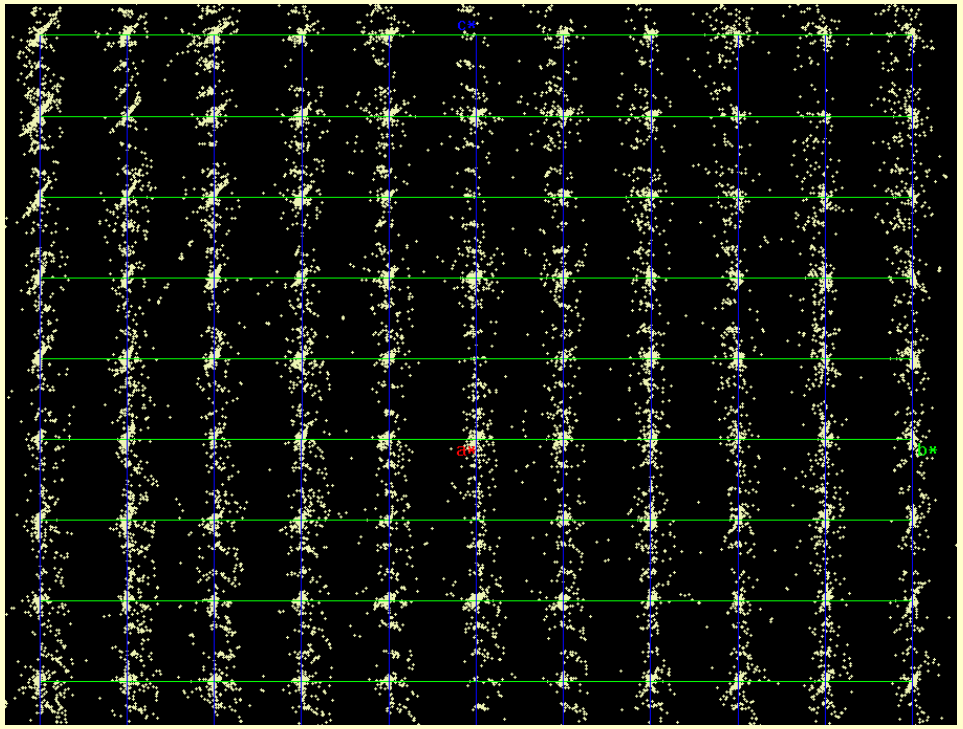
RFactors overview

R factors : [12378=10399+1979/650], Damping factor: 1.0000
GOF(obs)= 2.13 GOF(all)= 1.99
R(obs)= 6.59 wR2(obs)= 23.43 R(all)= 7.50 wR2(all)= 23.97
Last wR2(all): 23.97
Maximum change/s.u. : 0.0461 for GIso

R_{int}=8.6

Data quality may not be clear from R and R_{int} factors

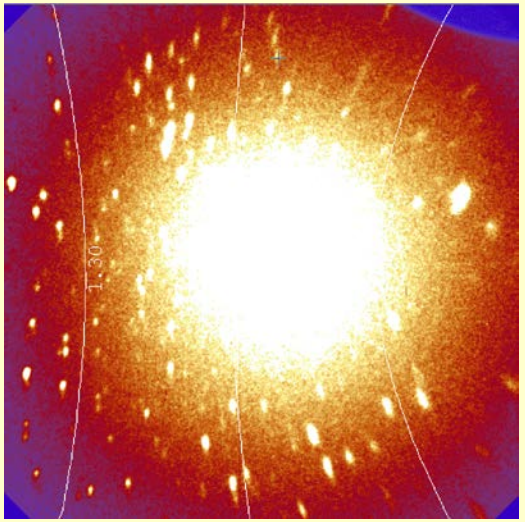




RFactors overview

R factors : [12378=10399+1979/650], Damping factor: 1.0000
GOF(obs)= 2.13 GOF(all)= 1.99
R(obs)= 6.59 wR2(obs)= 23.43 R(all)= 7.50 wR2(all)= 23.97
Last wR2(all): 23.97
Maximum change/s.u. : 0.0461 for GIso

R_{int}=8.6



How to store raw data?

In single laboratory level, storing of data needs just some disk space

Primitive solution in our lab in Prague:

Data from two Diffractometers

Backup

Hard disks replaced every 5 years for double sized

In the world-wide scale, it is a technical and diplomatic challenge

How to store raw data?

Experience with CIF project is discouraging: we have to embed the raw computer data to the platform-independent format

Any platform-independent CIF-like format for raw images will have the same problems. It can block possibility to use future features of data reduction programs for improvement of structure models.

The major diffractometer programs have very good back-compatibility and they are just a few

=> I suggest always storing raw data in the native format


Thank you for the attention