UK National Crystallography Service

University of Southampton

The increasing diversity of small molecule data: can one size fit all?

Simon Coles

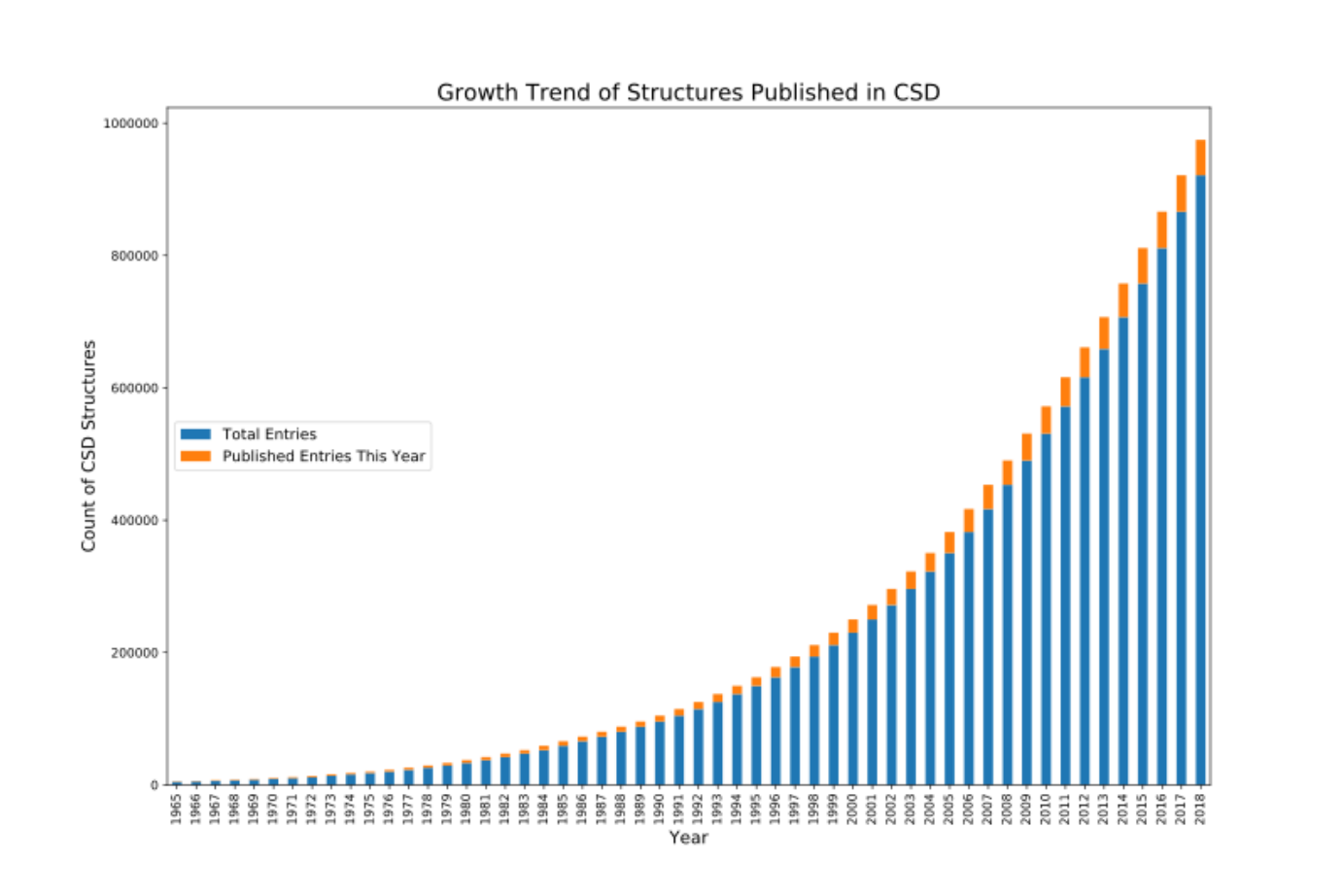
www.ncs.ac.uk
info@ncs.ac.uk

Small molecule crystallography ‘Quality Framework’

- The basis
  - >100 years of instrumentation development
  - >50 years of building highly valuable and curated results databases
  - >40 years of trusted common refinement processes
  - >30 years of agreed and maintained standards
  - >20 years of validation tools
- Well understood statistics for the suitability of a dataset and the fit of a model
- Automation e.g. CheckCIF / PLATON and Mogul

Small molecule crystallography ‘Quality Framework’

- Current validation and quality assessment processes for publication of crystal structures is largely based on:
  - (service) crystallography of the 1990’s
  - the final derived result i.e. submitted CIF



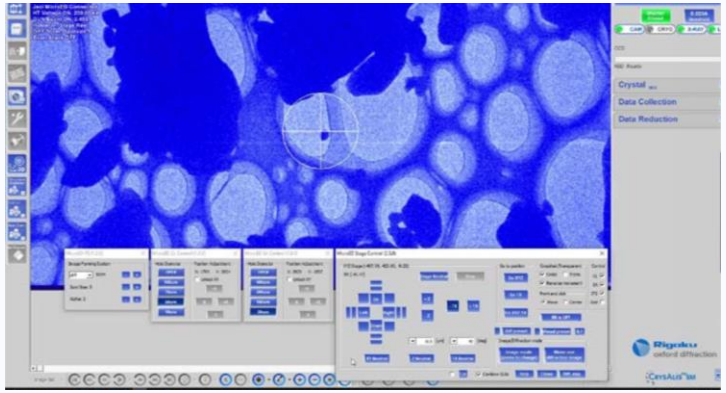
More recent challenges

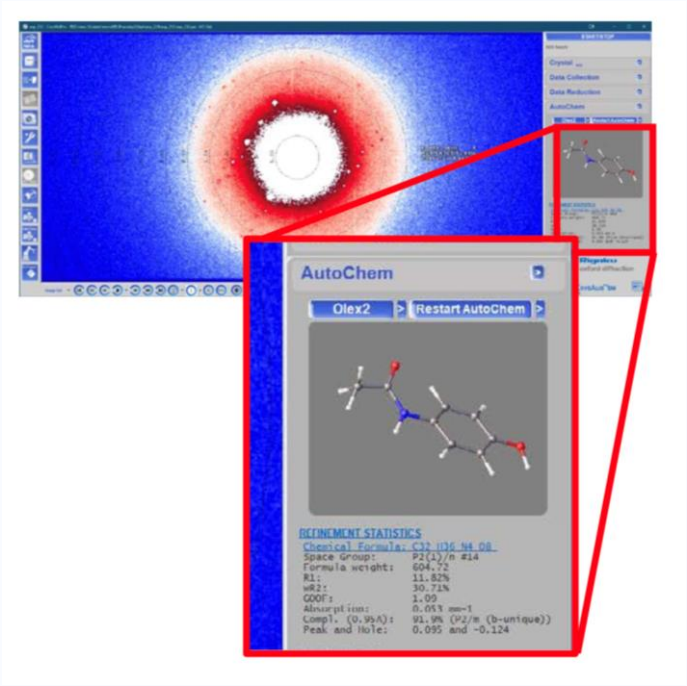
- More challenging samples – smaller, not ideally/usually crystalline
- Need to quickly answer questions to further chemistry research and development
- New methods for structure determination
  - Powder, NMR Crystallography, Electron Diffraction, XFEL, Crystal Sponge
- Dynamic and in-situ crystallography
  - High pressure, porous materials, photo-excitation, electric stimulus

- Established quality framework pushed beyond its limits…

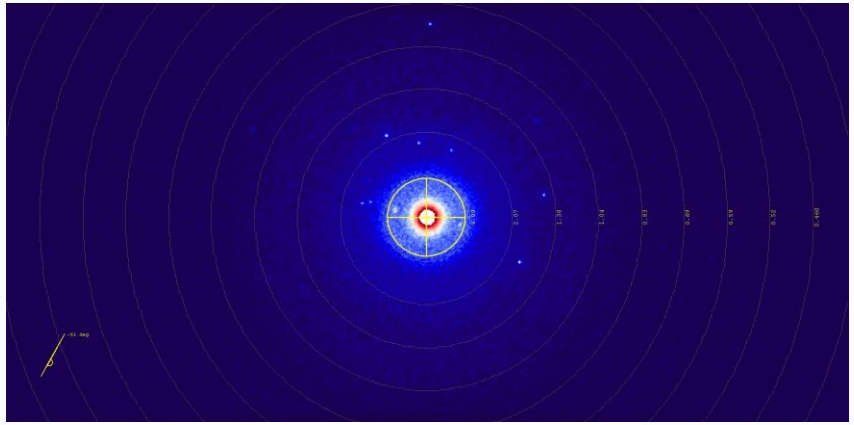
Electron diffraction – about to go viral

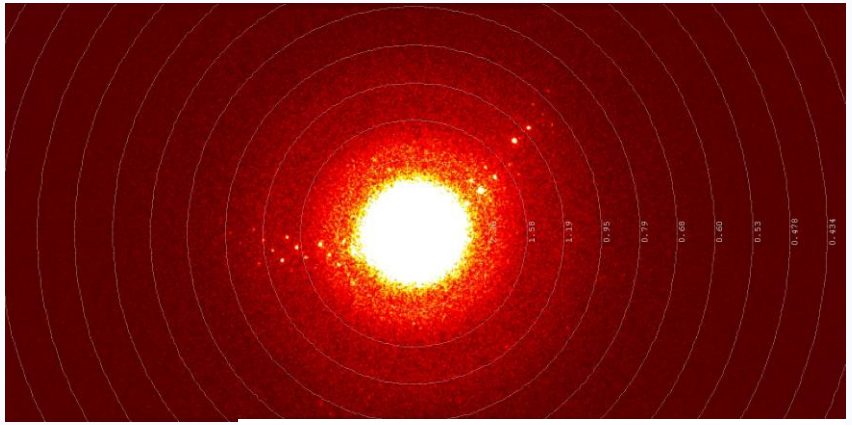
- A dedicated ‘electron diffractometer’

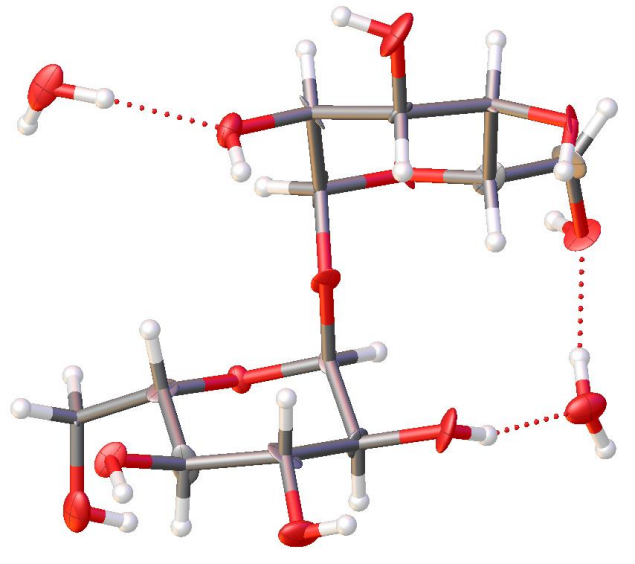


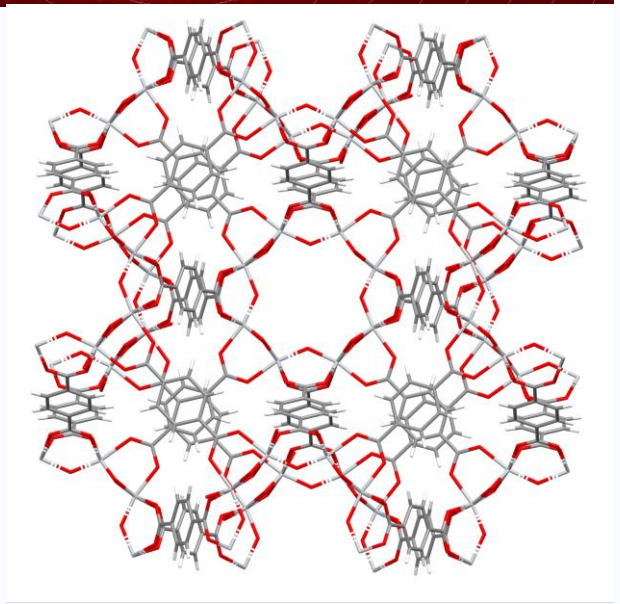


‘Routinely’ generating Chemical Crystallography results

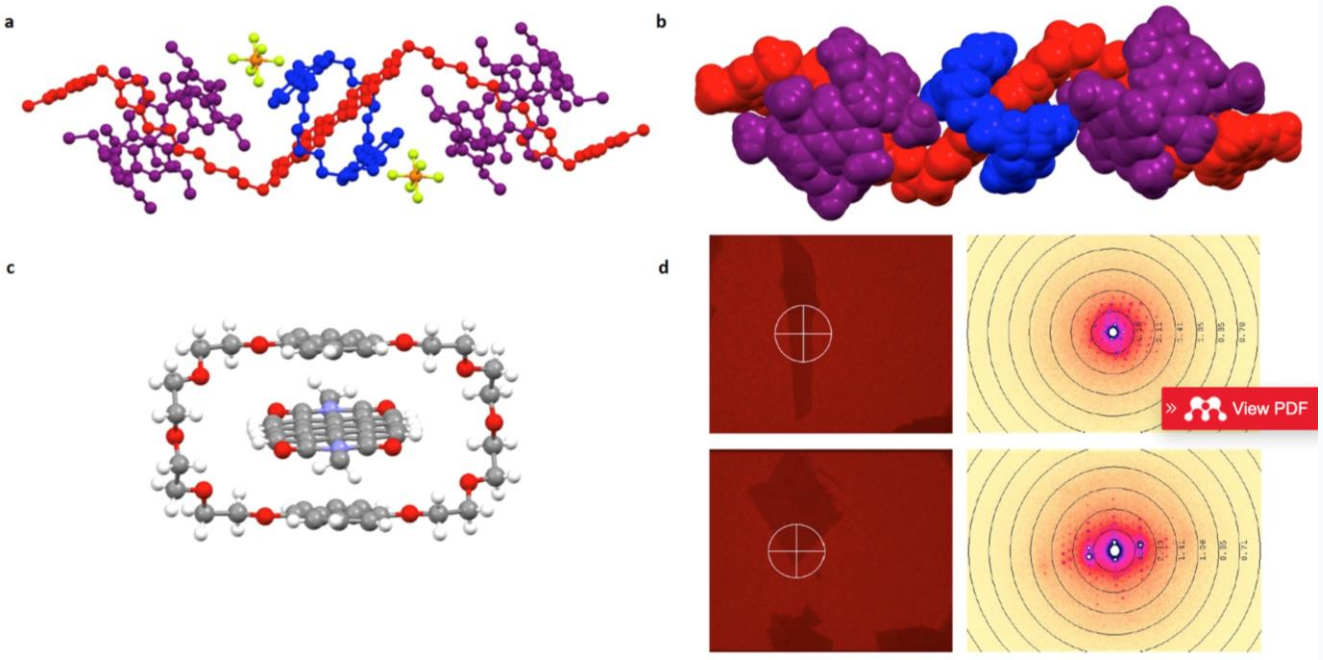








Ca 300 ED structures in the CSD (1.25Mth an ED structure)



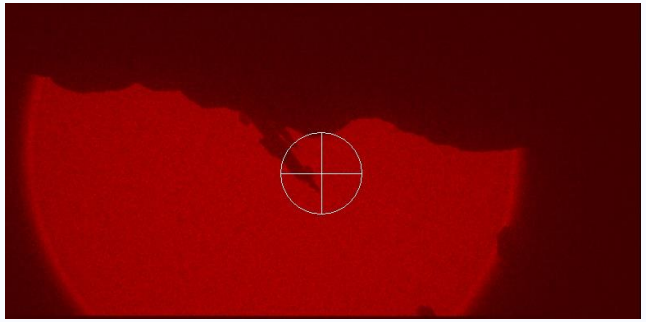
97091 reflections measured ($0.202^{\circ} \leq 2\Theta \leq 1.438^{\circ}$), 9201 unique ($R_{int} = 0.3529$, $R_{sigma} = 0.1495$)

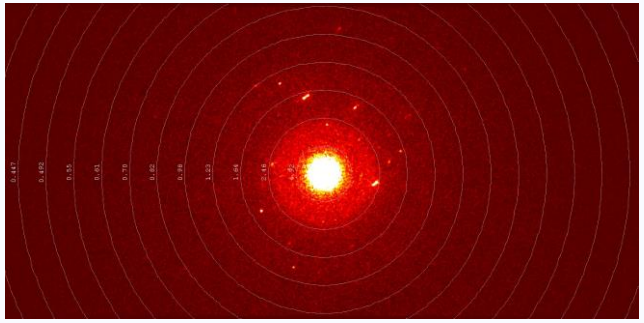
The final R_1 was 0.2607 ($I > 2\sigma(I)$) and wR_2 was 0.6034 (all data).

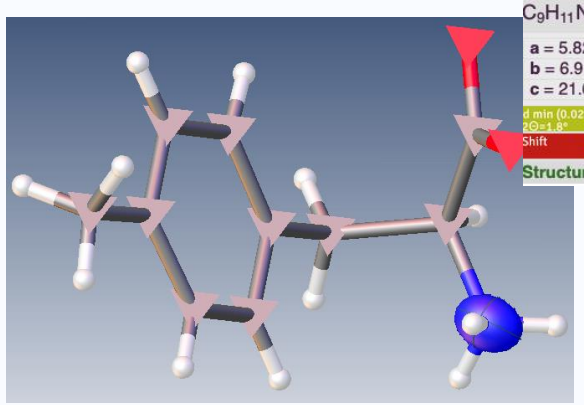
Pearce, N., Reynolds, K.E.A., Kayal, S. *et al.* Selective photoinduced charge separation in perylenediimide-pillar[5]arene rotaxanes. *Nat Commun* **13**, 415 (2022). https://doi.org/10.1038/s41467-022-28022-3

A typical approach to structure determination

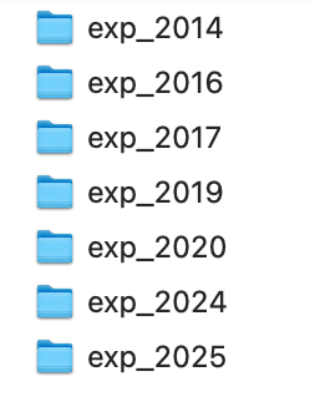
First try

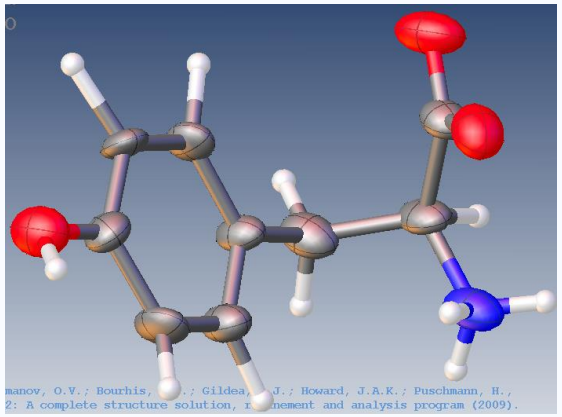






n[th] try (tries)





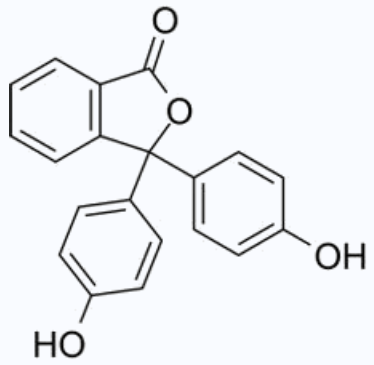
- Future reprocessing of raw ED data – multiple (dynamic) scattering, radiation damage to improve current models

Crystal Sponge

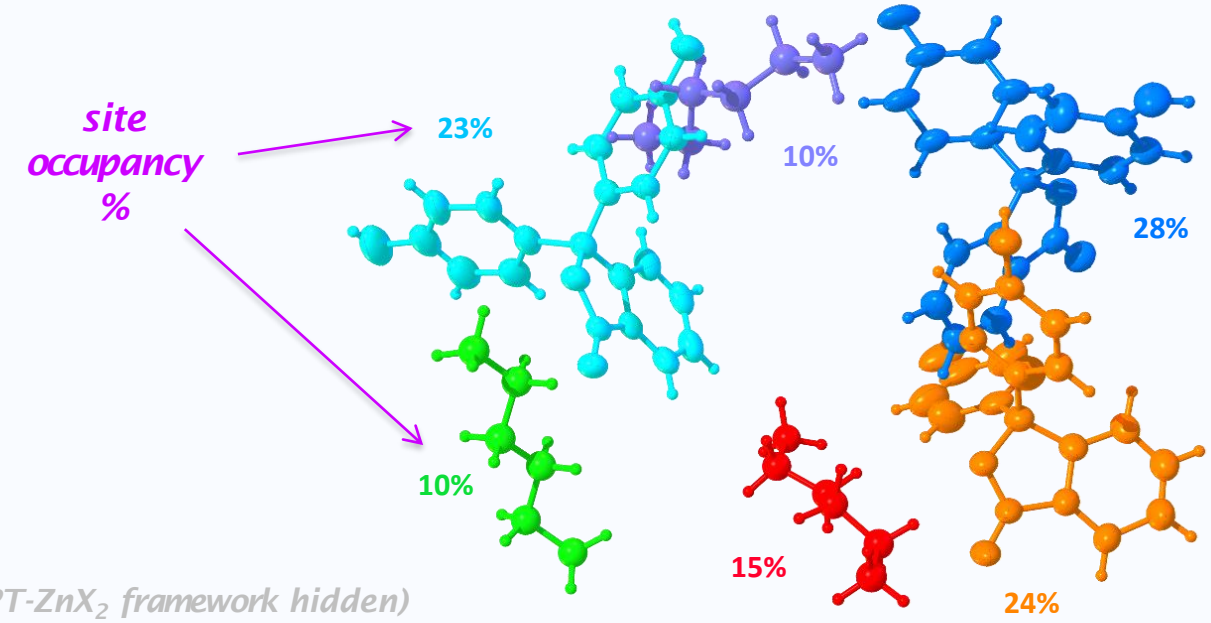
- A new and accessible approach to chemical / structural characterisation

- Determining ***molecular*** structures by encapsulating in porous materials
- Miniscule amounts of material (far less than required to grow a single X-ray size crystal)
- Determine structure of oils, gases, etc to 1000ths of an angstrom precision

Well-resolved result



phenolphthalein

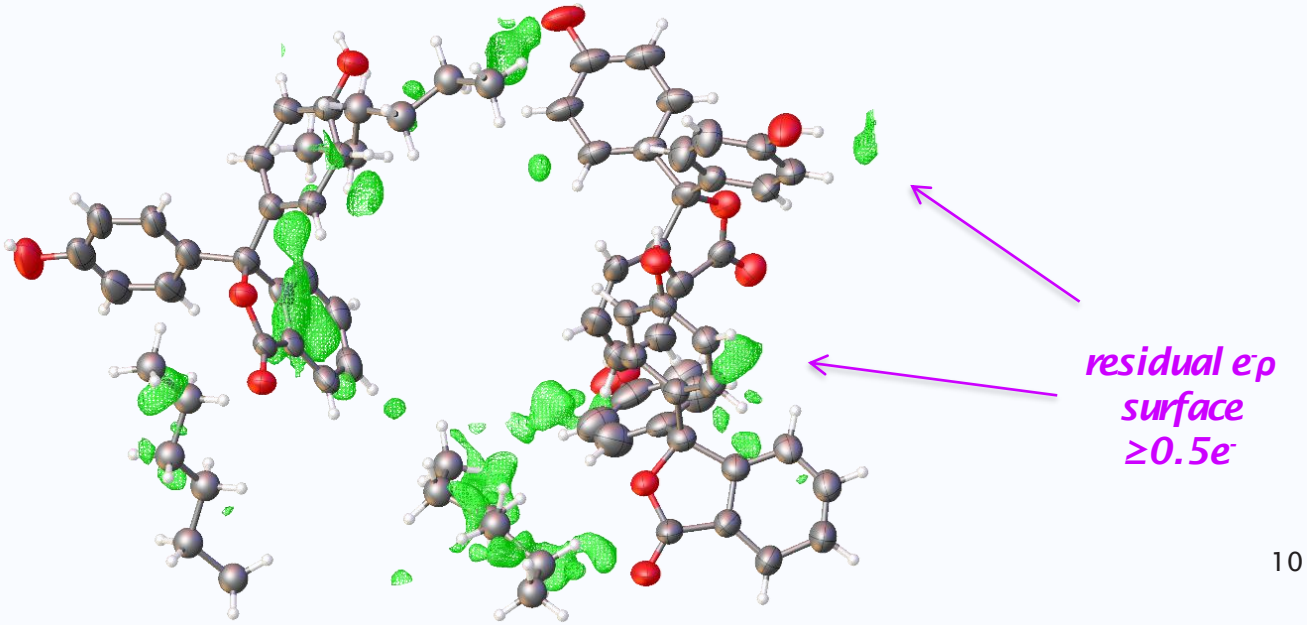


2021ncs7137c_jbo1a *C2/c*

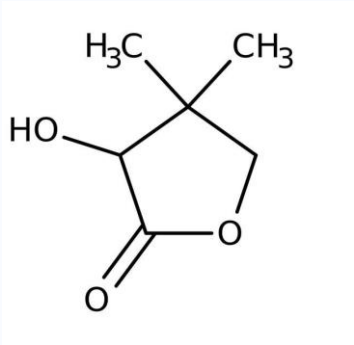
$C_{53.1}H_{39.4}Cl_6N_{12}O_3Zn_3$

a = 34.2990(4)	α = 90°	Z = 8	R_1	7.33	%
b = 14.4926(1)	β = 102.795(1)°	Z' = 1	wR_2	25.06	%
c = 30.8022(4)	γ = 90°	V = 14931.0(3)	11.5		

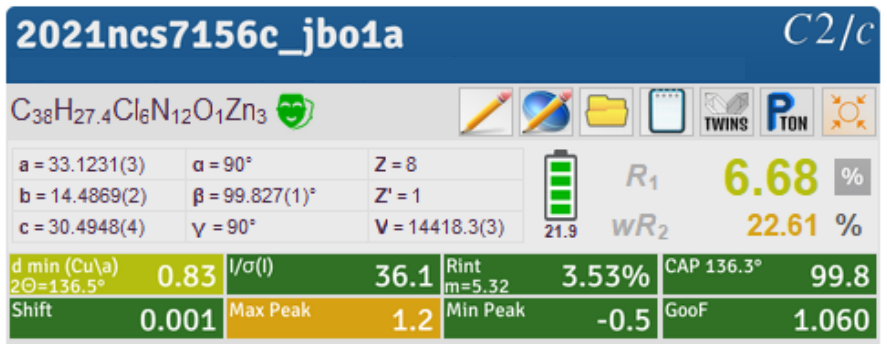
d min (Cu\a) 2Θ=136.5°	0.83	I/σ(I)	55.9	Rint m=5.38	2.46%	CAP 136.3°	99.8
Shift	0.001	Max Peak	1.6	Min Peak	-0.5	GooF	1.060

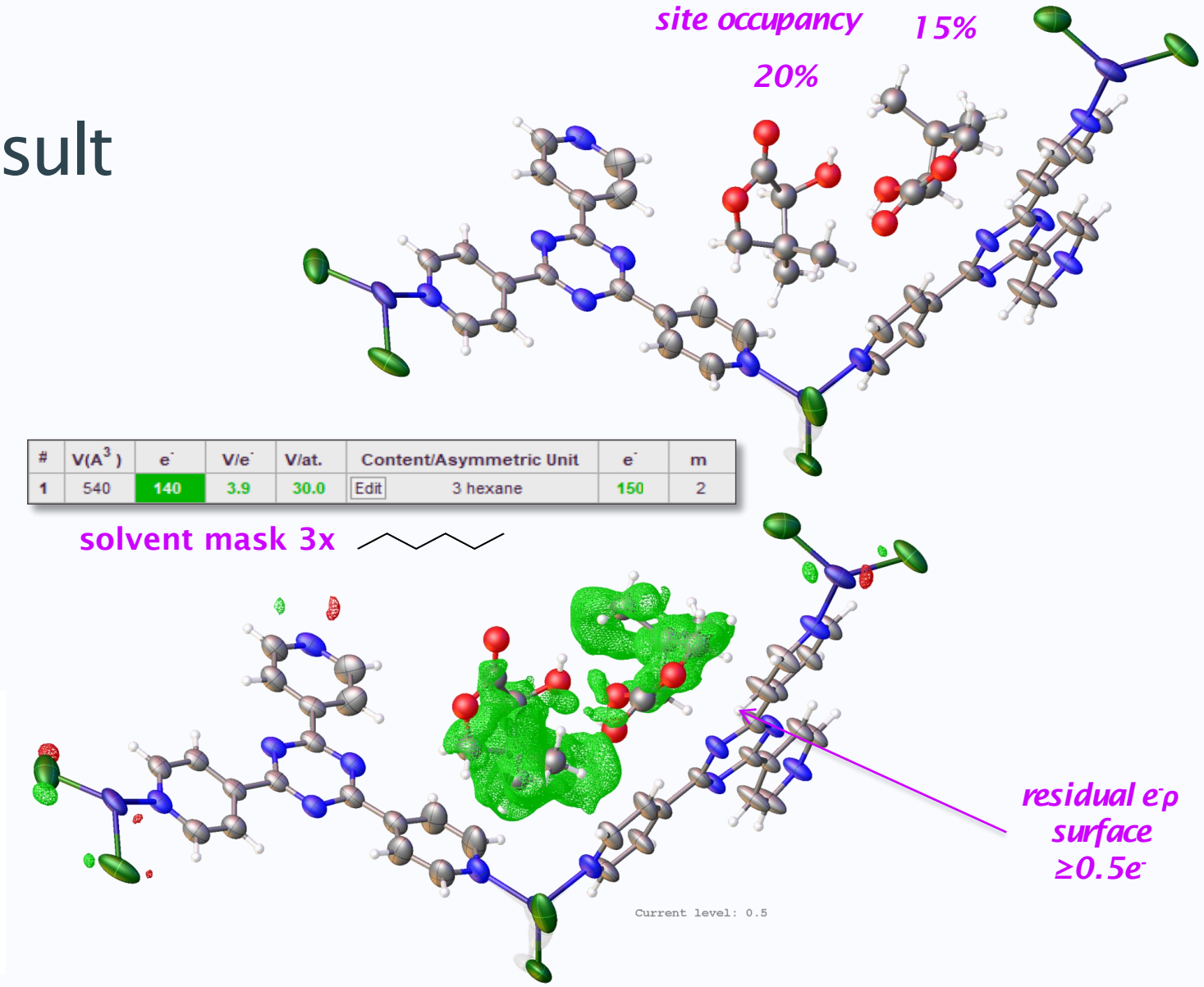


Less-well resolved result

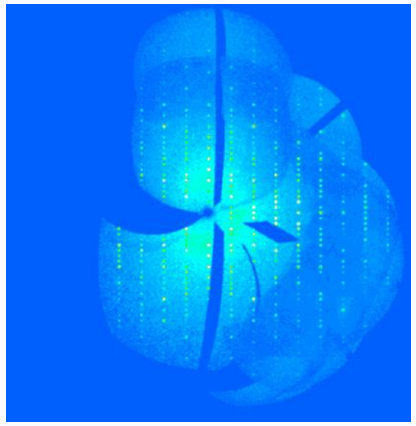
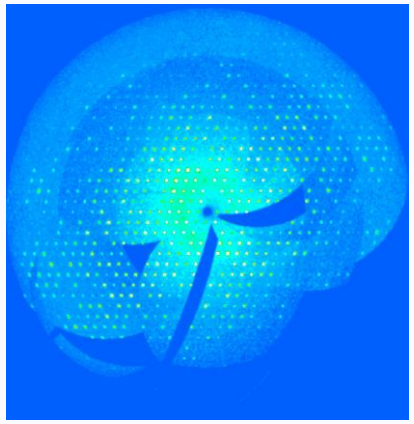
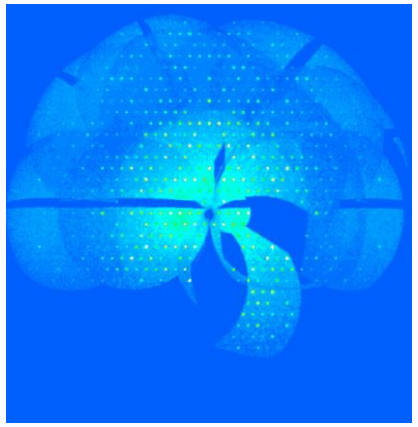
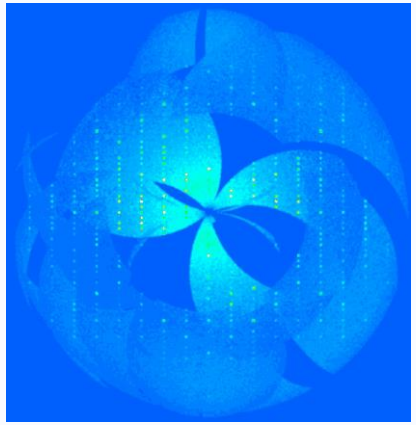
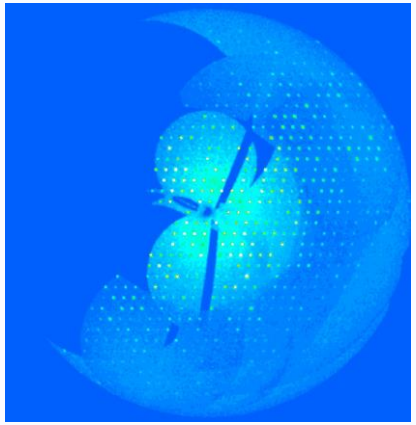
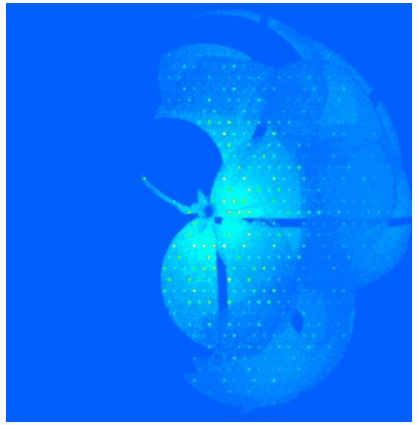


pantolactone





Further quality insight from raw data?

0kl	h0l	hk0	
			‘Good’ structure
			‘Poor’ structure

What reliable information can be taken from this?

Analyte	Sponge Type	Sample	Exchange Site 1 Dihedral Angle / °	Exchange Site 2 Dihedral Angle / °
	ZnI_2 cHEX	A	-61.1 (17)	-67.2 (31)
		B	-59.0 (18)	-67.6 (18)
		C	58.9 (19)	-64.5 (18)
		D	-61.0 (12)	-63.5 (14)
		E	62.4 (14)	62.7 (16)
		F	61.7 (13)	65.0 (15)
	ZnI_2 nHEX	A	60.4 (16)	-66.0 (19)
		B	61.3 (15)	63.0 (18)

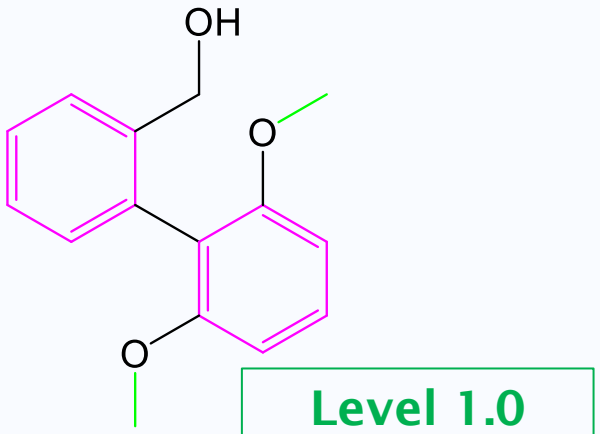
Data should be fit for purpose…

- Original purpose of structure determination
  - What have I made?
  - What is the reaction by-product?
  - How has my structure changed?
  - How does this material manifest these properties?
  - What are the driving forces behind structure formation, how do I control them?
- How others will reuse the results
  - Do chemicals like this exist (connectivity)
  - Starting point for follow on calculations (conformation)
  - Highly accurate structural features (precise structure)

- Crystal Sponge structures give a comparable accuracy to gold standard crystal structures
- Kept, e.g. in CSD, alongside gold standard crystal structures
- Are these structures used for follow-on research in the same way as / with gold standard crystal structures? What discernment going on?

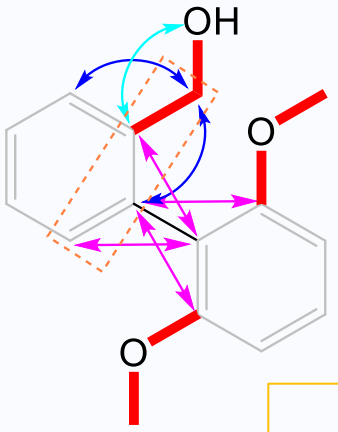
- Need to extend the current quality framework
- A structure grading system
  - For validation/publication, particularly to enable reuse – non-experts & data science
- Quantitative analysis of restraints / constraints applied
- Include properties of the primary / raw data?

Weighted sum averaged – 'good' example



Level 1.0

Exchange
Site A
74.1%
SIMU
RIGU



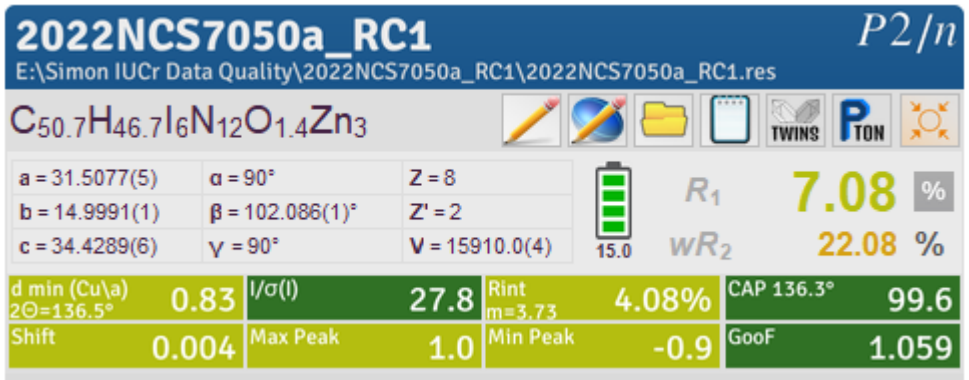
Level 2.7

Exchange
Site B
20.9%
SIMU
RIGU

AFIX 66 FLAT DFIX SADI DANG

SADI and DANG restraints are colour-coded to indicate the relevant relationships.

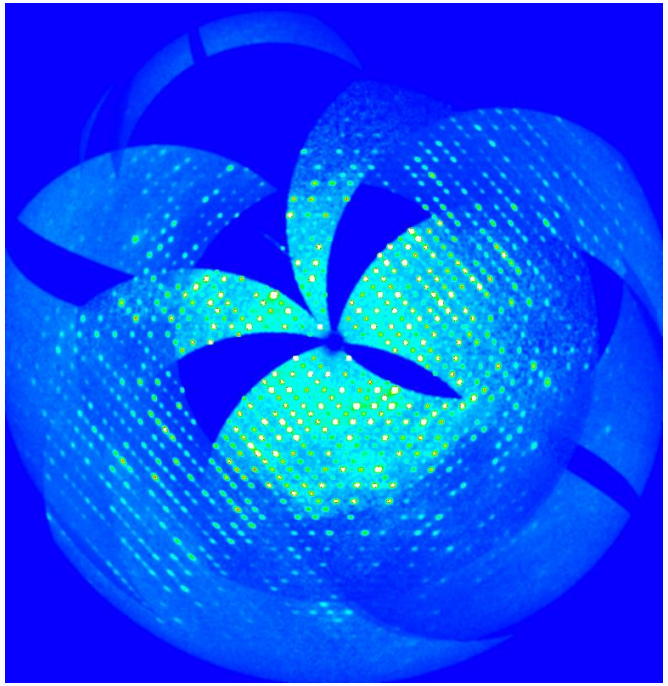
- Well defined pore (2 guests, 11 solvents, low level of unmodelled electron density



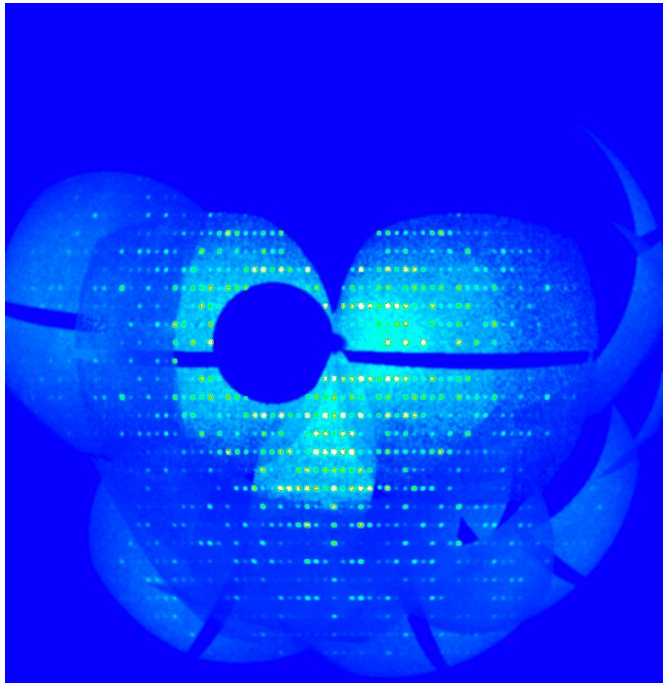
- How to grade a structure which has two guests of different quality?

How to include a factor from the raw data?

0kl



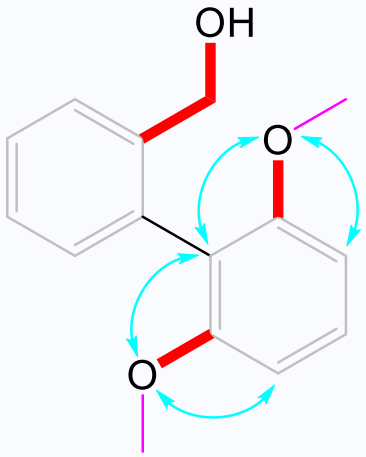
h0l



hk0

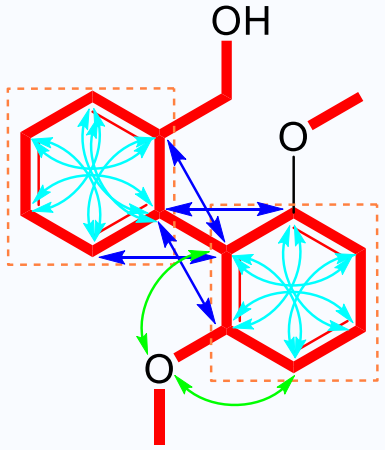
Weighted sum averaged – poorer example

Level 2.8



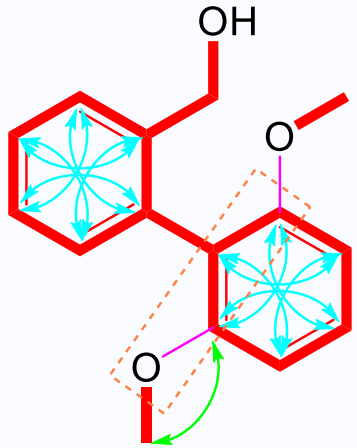
Exchange
Site A
49.3%
SIMU
RIGU

Level 3.2



Exchange
Site B
22.3%
SIMU
RIGU

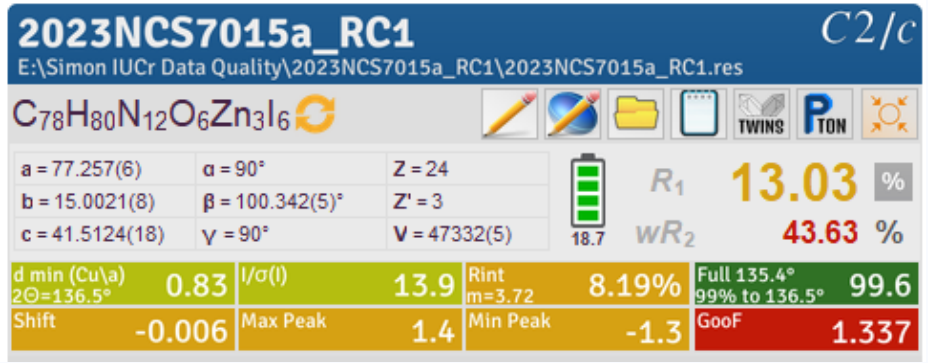
Level 2.7



Exchange
Site C
21.5%
SIMU
RIGU

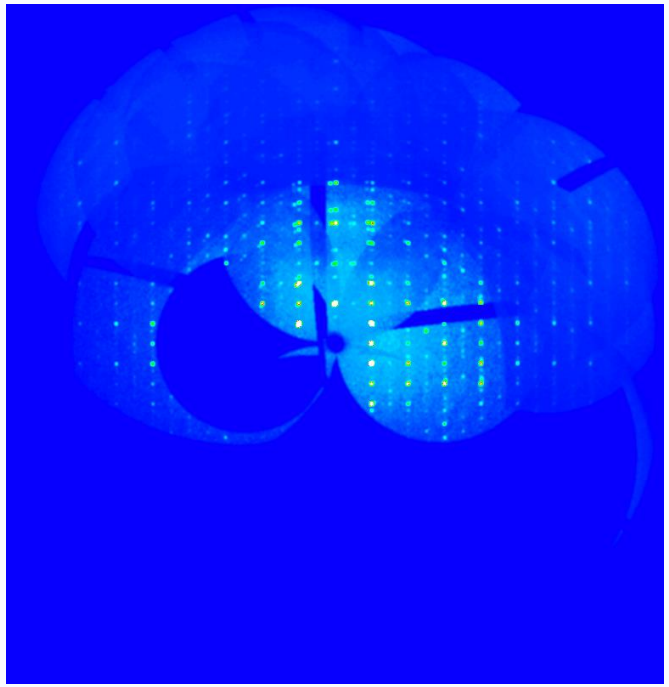
AFIX 66 FLAT DFIX SADI DANG

- ZnI_2 poorly defined with high residual density

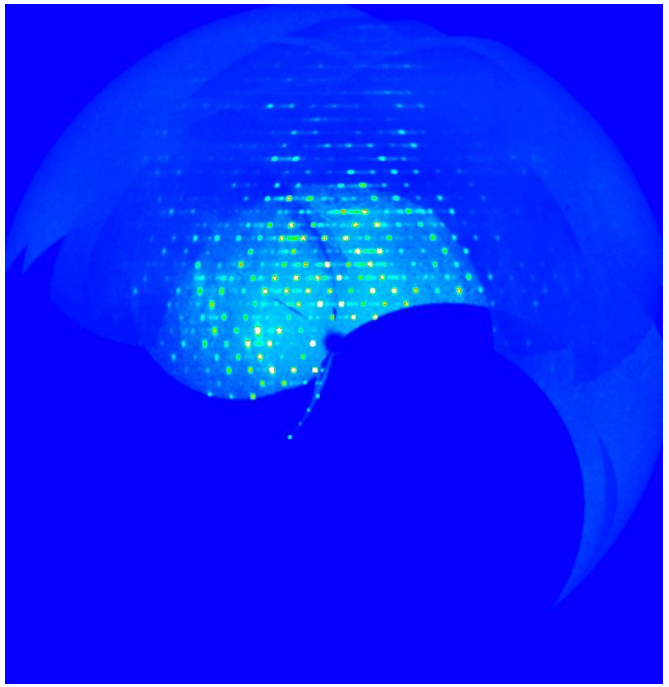


- Multiple exchanges
- Provides more insight into different conformations
- However, all require quite considerable restraints

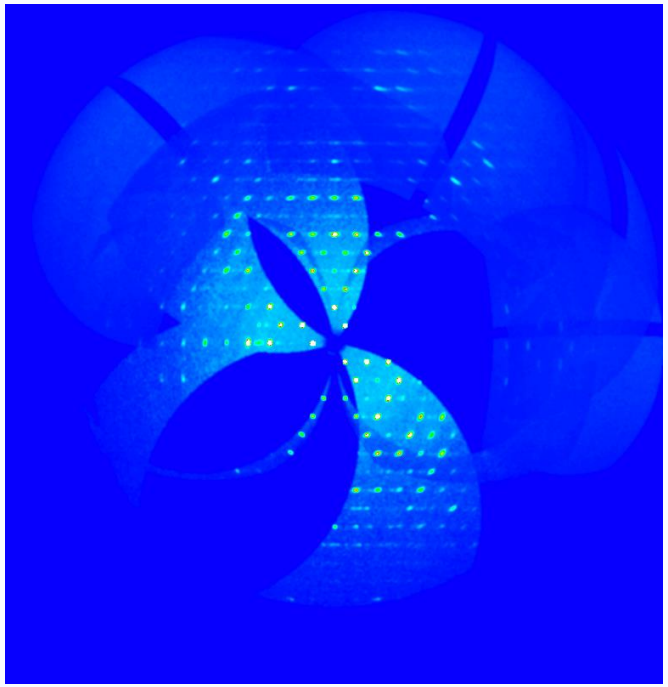
Quantitatively poorer raw data…



0kl

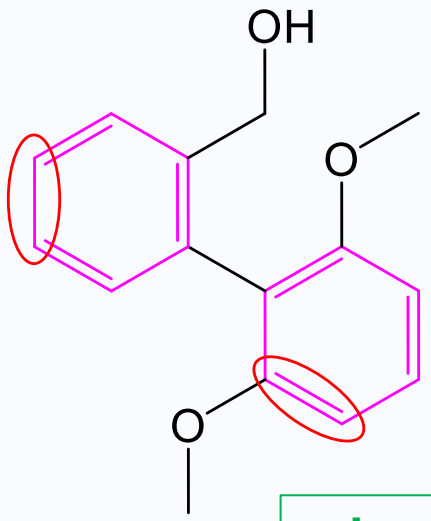


h0l



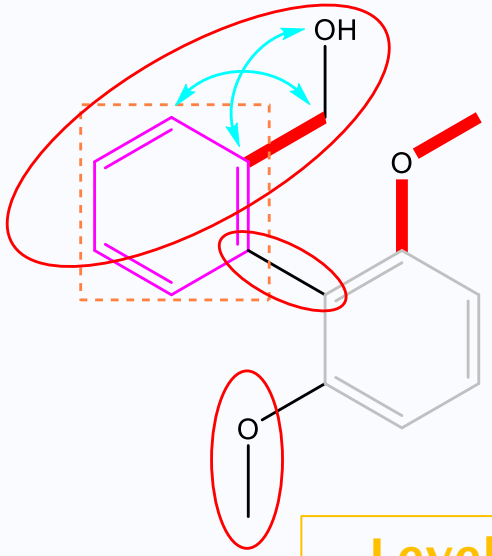
hk0

Minimising number of restraints / constraints



Level 1.0

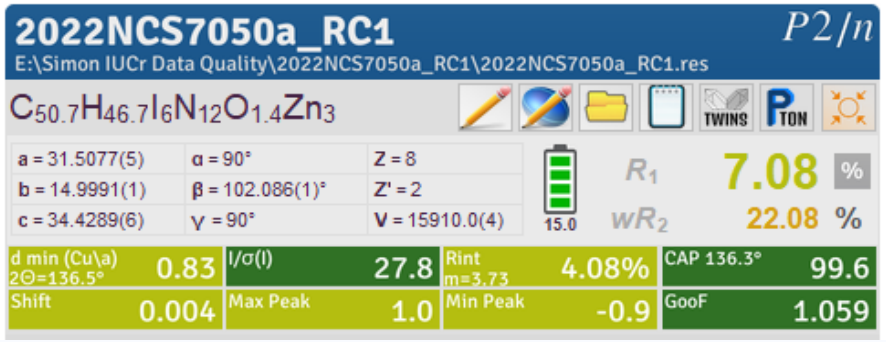
Exchange Site A
74.1%



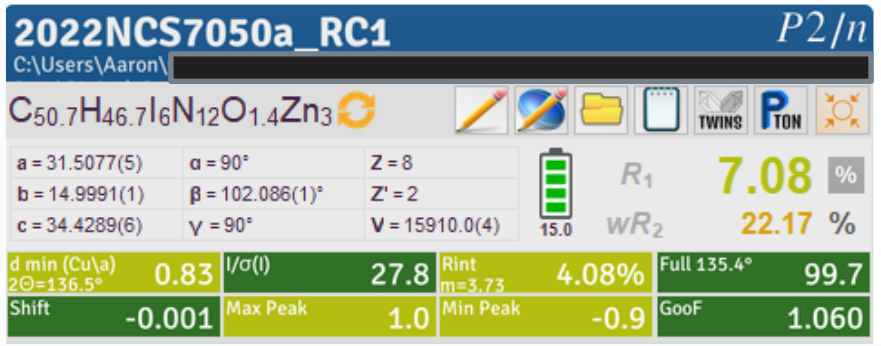
Level 2.4

Exchange Site B
20.9%

Prior



Minimal

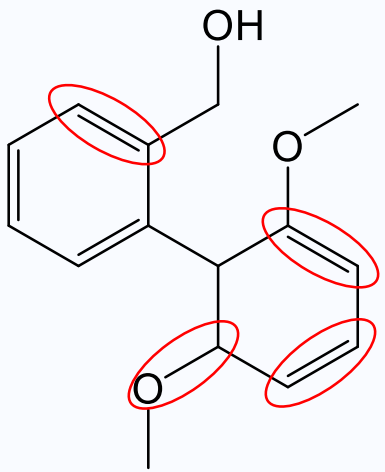


- Statistically significant changes in the model not reflected in conventional metrics.

AFIX 66 | FLAT | DFIX | SADI | DANG | >3σ shift

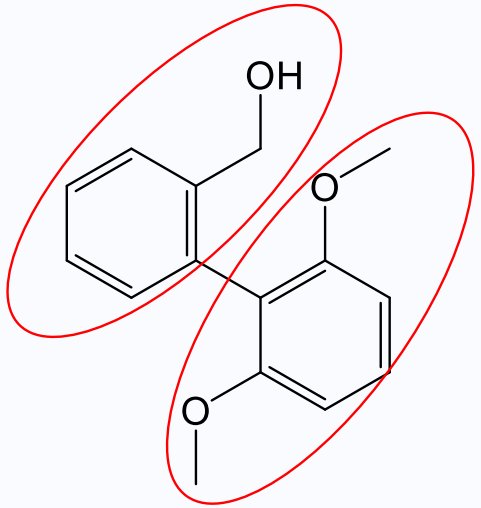
Minimising number of restraints / constraints

Level 1.0



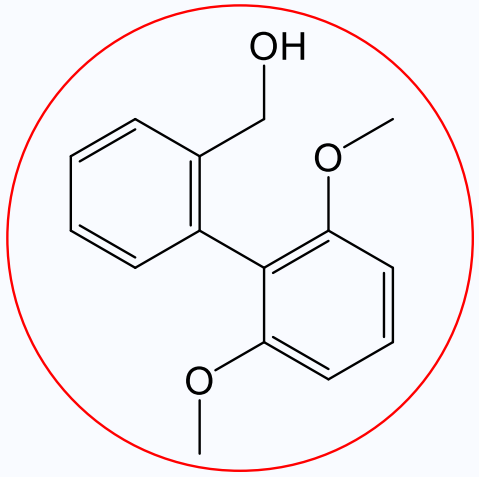
Exchange
Site A
49.3%

Level 1.0

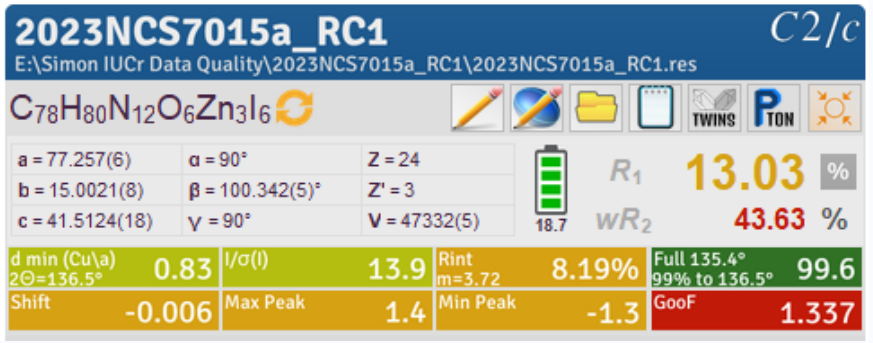


Exchange
Site B
22.3%
SIMU

Level 1.0



Exchange
Site C
21.5%
SIMU
RIGU



- Sites B and C do not have dramatic visual shifts but geometric features undergo statistically significant change
- Lower occupancy tends to indicate greater reliance on restraints/constraints - the impact of removing them results in greater statistical shifts (~5/6σ)

AFIX 66 FLAT DFIX SADI DANG >3σ shift

Conclusions and Further Work

- Increasing number of situations where conventional crystallographic metrics don't truly portray the quality of a model
- An extension to the established quality framework is necessary and viable – particularly to enable ***appropriate*** reuse of results
- Refinement of grading system necessary – particularly review of contributing factors
- How to combine contributing factors?
- How generally applicable is this approach?
- How to quantitatively include factors from raw data?

- Thanks to Rob Carroll and Aaron Horner for application of initial grading scheme