32nd European
Crystallographic Meeting
ECM 32 ♦ Vienna Austria ♦ 18. – 23. 08. 2019

IUCr's Committee on Data Workshop
Data Science Skills in Publishing

"Overview of the new opportunities in and a harmonisation of peer review of 'data with validation report with article narrative' practices"

John R Helliwell
University of Manchester, UK
18th August 2019

Narrative Data Validation report Publish?

Narrative

Data

Validation report

*IUCr chemistry journals have followed the **exemplary practice of checking all three of the above** and are an inspiration to other fields of crystallography to follow their lead*

Dr Madeleine Helliwell
Acta Cryst C Coeditor
2000-2004

As EinC IUCr Journals 1996-2005 everyday I could see Madeleine's Acta C editorial work on her submissions by authors

At IUCr Congress Geneva in 2002 as EinC IUCr Journals I proposed to the Open Commission on Biological Macromolecules that Acta Cryst D should adopt Acta Cryst C's procedure of allowing editor and referees to see

article with data with validation report

This proposal was rejected by those present

Later I realised, in my refereeing, why not insist on having from the authors, via the editor, the to-be-released PDB coordinates and structure factors as well as the PDB Validation Report and article?[1]

Helliwell, J. R. (2018). Data science skills for referees: biological X-ray crystallography *Crystallogr. Rev.* **24**, 263-272.

So, a referee's rights for Open Data!

Let me describe my experiences these last two years or so

- How do I proceed in my refereeing?
- Do all publishers cooperate?
- Do my referee's reports get published in this new era of Openness?
- Is there a future for open data for referees? ie is it sustainable?

How do I proceed in my refereeing?

- *I follow what I describe in my Cryst Rev (2018) review:-*
  - Comment on the submitted article
  - Comment on the PDB Validation Report, which is good for the general checks of the data and authors' model
  - Calculate my own electron or nuclear density maps; I can then roam through the **whole unit cell (not just portions selected by the authors)** and at the **largest peaks of density that are uninterpreted** by the authors in their model
  - I can calculate the anomalous difference density map if I wish
  - Check by eye the processed diffraction data file eg general details, zones of reflections to see if there are major regions of reflection intensities data incompleteness
  - Check by eye the coordinates file for any strangeness
  - Maybe request the authors to reprocess their raw diffraction data eg if $\langle I/\sigma(I) \rangle$ looks to be too large ie data in outer edges are unused
  - Recommend "Deposit your raw diffraction images and quote the doi in your paper" as per the IUCr DDDWG Final Report

Do my referee's reports get published in this new era of Openness?

- Yes, so far just once, by Nature Comms 2018:-
  - https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-018-06957-w/MediaObjects/41467_2018_6957_MOESM1_ESM.pdf

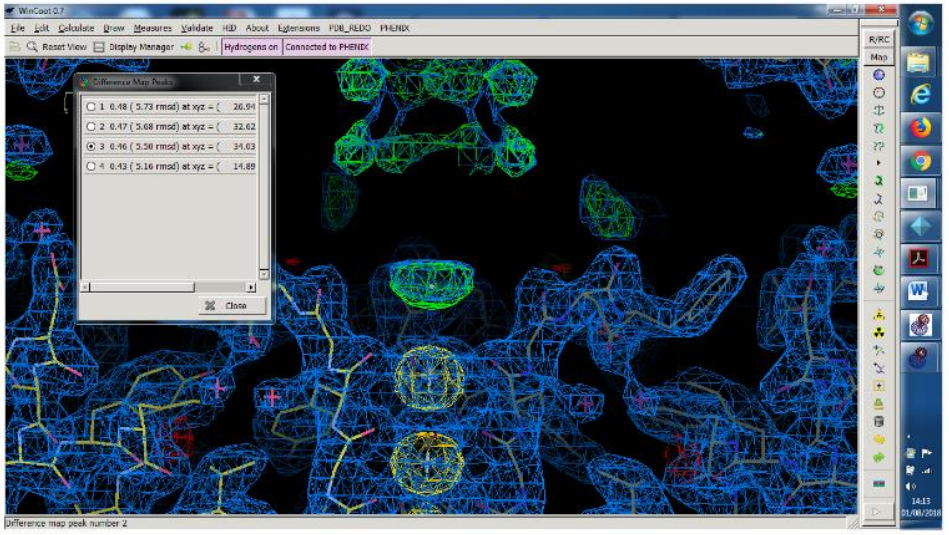
What is the peak 3 in the above screen shot? A bound water? [Ie since there is no anom peak on it.] It looks as if it may be functionally interesting. A comment is needed in the article. Water.

We have added a water molecule at this position into the structure and we added this sentence into the manuscript.

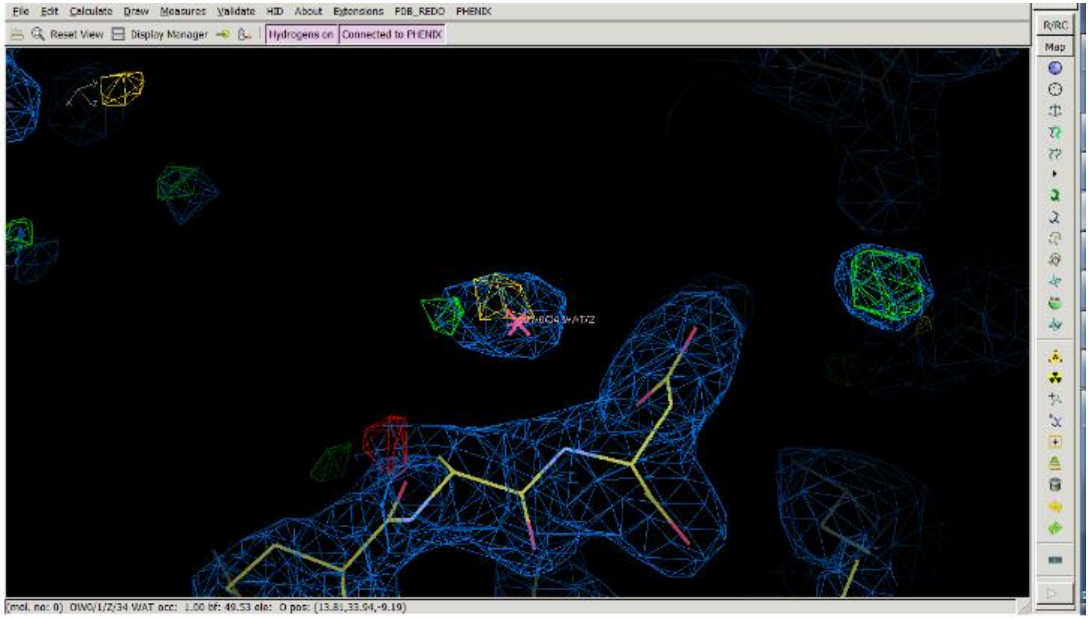
"On the top and bottom of the K^+ ions in the selectivity filter are bound water molecules, the structure and experimental data has deposited into the protein data bank with the accession code 6DZ1."

The anomalous difference Fourier map above was calculated in Phenix.

The Fo-Fc peaks list calculated now by Phenix is:-



An anomalous difference density map

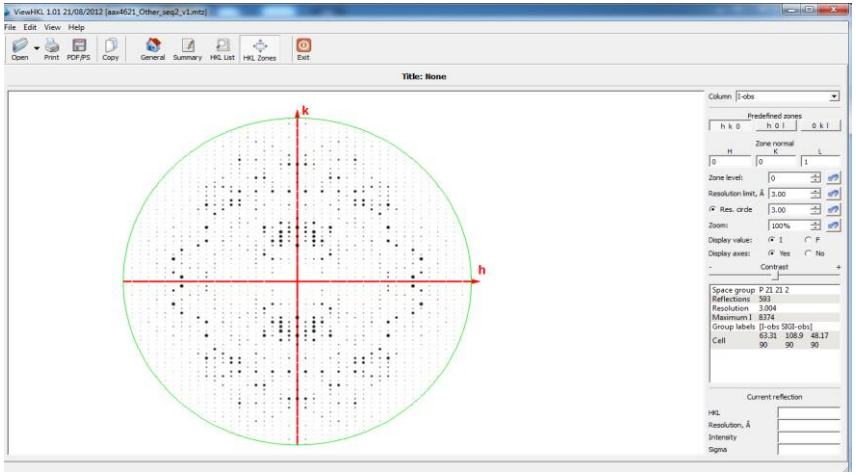


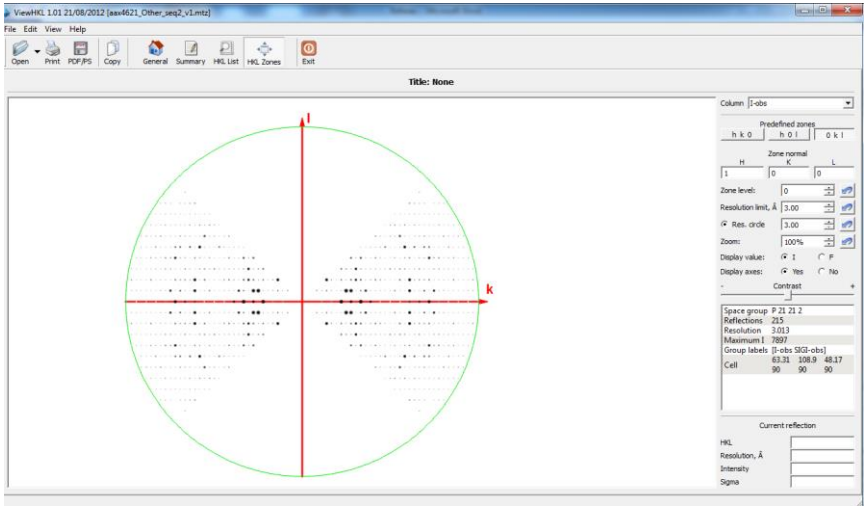
These three bound waters need reassignment to ions, their identity as a cation or anion accordin
their charged neighbour.

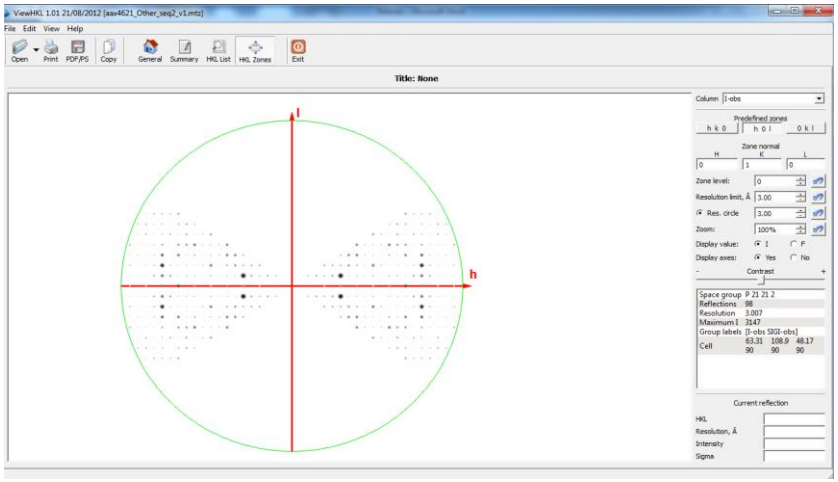
Response from authors:-

The three bound water molecules have been reassigned as K^+ ions in the structure.

Diffraction data completeness check; electron crystallography example







Admittedly an extreme example!

Do all publishers cooperate?

- Those that do:-
  - IUCr Journals, RSC, ACS, Nature Comms, Science Advances
    - Those that do not:-
      - »Elsevier (three times); MDPI Crystals

Is there a future for open data for referees? ie is it sustainable?

- Ok, I am (semi) retired and so I can spend a day on a refereeing task, and I find it very satisfying to do this thorough set of checks
  - But it is true I haven't been asked to referee a ribosome crystal structure ie something that is very large in the scope of the task
  - I have however refereed a submission for IUCrJ with five underpinning crystal structures, three in the PDB already and two new ones. All five needed revisions to the PDB data files
  - Fully employed referees, ie with a 'day job', could referee fewer submissions in favour of being more thorough with the tasks they do take on

In conclusion

- I think the Acta Cryst C way is much better than what we in MX have now from our journals
- So, I commend that we should check everything as referees before we let our journals and database put the versions of record into stone!
- Referees' reports should be published not least so that readers can see how thorough the reports are and if the editor has overseen a proper implementation of the referees' reports

The version of record:-

Thankyou