

# Review of: "Native SAD Phasing at Room Temperature"

Sabine Botha<sup>1</sup>

<sup>1</sup> Arizona State University

**Potential competing interests:** The author(s) declared that no potential competing interests exist.

While the scope of the title "Native SAD Phasing at Room Temperature" promises some methodological, technical or analytical advancement in SAD phasing techniques, it is instead a very literal description of the contents of this article: three model protein datasets were collected at room temperature (RT) at the Advanced Photon Source beamline 24-ID-C and the structures solved using Single-wavelength anomalous (SAD) phasing. Additionally some analysis is presented that fits a model to the peak heights in the anomalous density maps with respect to the number of frames included in the phasing dataset, but the overall scientific impact is minimal.

While the manuscript is well written from a literary perspective it is unclear what its scientific purpose is intended to be. SAD phasing itself is not novel, and the authors state that "*native SAD phasing is commonly used as a proving ground for new technologies and data collection methods*". It can therefore be assumed that the SAD phasing itself is merely used as an assessment criteria of the data collection method, i.e. performing a helical scan at room temperature. However this data collection strategy in itself is also not new (for example, the P14 EMBL beamline at PETRAIII has been offering these "4d" scans for years).

The SAD phasing itself is performed using routine, well-established pipelines with default parameters, and the data is collected from large protein crystals which are not routinely obtainable for more complex and unknown systems. The data were collected using a helical scan at room temperature, but this is not well presented and it is the task of the reader to filter out this information. The authors state that radiation damage is not a problem with their data collection technique, but due to a lack of experimental detail and quantification, it is not possible to assess the validity of this statement, even for the large crystals used in this study.

If the data collection method is the intended focus of the manuscript, then it should be presented a lot more clearly and completely. It is not mentioned in the abstract, outlined in the introduction nor detailed in the Materials and Methods Section that this is a helical scan. It is briefly mentioned that the crystal gets translated to distribute the radiation dose, and that multiple sweeps are collected per crystal. However, the

lack of detail makes it impossible to judge the radiation impact of this collection strategy:

- Data was collected at 0.5% transmission. What was the actual photon flux at full (or at 0.5%) transmission for this X-ray energy?
- What was the beam size, was an aperture inserted, and how much was the crystal translated during an exposure? What is the overlap of the crystal volume between subsequent exposures? How many exposures impacted the same crystal volume?
- Were successive sweeps performed on the same part of the crystal or on a "fresh" sample volume?
- What is the actual estimated radiation dose per crystal? Both globally and locally. It is highly recommended to include some dose estimates from e.g. RADDOS (http://www.raddos.se/). How does this compare with conventional methods on similar sized crystals?
- Personally, I would also not consider a 0.2-0.5 mm protein crystal "fairly" large but very large. From the crystallization protocols, there also seems to have been a spread in crystal sizes up to 1mm! This being presented as a "single crystal" collection strategy, what were the **3D** sizes of the crystals that were actually used for data collection?

Further, the authors claim that there is no indication of radiation damage in their data, but Figure S2, the fact that *"For the PTP1B:TCS401 complex, we excluded the last 360 frames from the second pass due to weaker diffraction"* and that data from two crystals had to be merged for apo PTP1B to yield reasonable anomalous signal indicate the contrary. A detailed analysis of dose rates for the helical scan collection strategy (including the actual flux parameters) needs to be included before any conclusions can be drawn regarding the radiation damage effects on the crystal.

*"Data were collected at 6.5 keV to maximize the anomalous scattering signal within the accessible spectrum of the APS 24-ID-C beamline"*. However, from Table 1, the data was collected at a wavelength of 1.892 Å, which would correspond to an X-ray energy of 6.55 keV (rounded to one decimal this would be 6.6 keV). While this may seem like a minor detail, in this particular case it is the difference between measuring above as opposed to below the K-absorption edge for Manganese (theoretical estimate 6.54 keV)! Additionally, was a fluorescence scan performed to determine the experimental absorption edge of Manganese for DHFR at 24-ID-C?

It is further unclear why a high symmetry Spacegroup such as  $P4_32_12$  would yield such a low completeness overall (90.74%) for *"highly redundant data collection"*, or 85.52% for  $P2_12_12_1$ . A statement should be made whether there is a benefit in collecting redundant, yet incomplete data.

Is the low completeness related to the statement *"we excluded strong reflections lower than 6 Å in resolution during indexing due to a secondary diffraction pattern originating from the harmonic"*? Was this data excluded when finding the indexing solution, but the intensities nonetheless integrated beyond this, or were these reflections omitted completely? What was the criteria for being considered a "strong"

reflection?

A related minor point: “*we used local index assignment (index.assignment.method=local)*”; what does this mean for someone who does not routinely use *DIALS* for indexing? Is this akin to what is routinely done for serial crystallography, where individual patterns are indexed independently from one another? If yes, then it is clear why this yielded better indexing rates, however, would this change the definition of “redundancy” for this dataset, as reflections are longer integrated across images but considered as separate measurements?

The final point the authors make is that high-redundancy data collection improves anomalous signal. This is not a new observation, and a rather obvious statistical implication which has been demonstrated experimentally in numerous studies for both single crystal, multiple integrated wedge and serial crystallography (and the authors reference a number of these studies in their manuscript, such as the excellent and thorough study performed by Weinert and colleagues). While the slightly modified model the authors present for fitting the anomalous peak height dependance on the number of data frames included in the analysis is a good fit, the predictive power is very limited being based on already refined phases.

The experimental protocols for expressing, purifying and crystallizing the model system proteins is presented with exemplary detail and should be commended as the specifics of crystallization protocols are often not disclosed in sufficient detail in crystallography publications.

In conclusion, the manuscript presents an established data collection approach with some additional analysis by testing how well the data is compatible with SAD phasing, however the omission of crucial experimental parameters and details to support their claim of low dose data collection at room temperature render it unsubstantiated.