# Acta Crystallographica Section D

**Volume 70 (2014)**

**Supporting information for article:**

## Exploiting subtle structural differences in heavy-atom derivatives for experimental phasing

**Jimin Wang, Yue Li and Yorgo Modis**

### S1. Preparation of uranium derivatives and of pseudo native crystals

Soaking crystals in 5 mM $UO_2$ acetate changed the unit cell parameters from $a$ = 137.4 Å, $b$ = 56.2 Å, $c$ = 94.8 Å, β = 94.9˚ of the native data set to $a$ = 136.7 Å, $b$ = 54.5 Å, $c$ = 95.9 Å, β = 92.2˚ of the derivative with $\Delta b$ = 3.3% and Δβ = 2.7˚ (both in space group C2). Interestingly, this soaking also improved the resolution of the diffraction of many crystals from approximately 4 to 3.25 Å in the strong diffraction direction. We expected that soaking in lower U concentrations would induce the same change in unit cell dimensions but with lower U occupancy so that the resulting data sets could serve as pseudo-native sets in SIR phasing. Unfortunately, soaking in 0.2 mM U failed to induce any change in unit cell parameters, and also did not result in any detectable anomalous signals. Finally, we generated several isomorphous pseudo-native data sets with nearly identical unit cell parameters as the U derivatives by a back-soaking method: first soaking with 5 mM $UO_2$ acetate overnight and then back-soaking without $UO_2$ acetate for 30 s, 60 s or 30 min. Unfortunately, the resolution of the back-soaked native data sets at the three time points was reduced to 3.8, 4.2, and 6 Å, respectively. Nevertheless, these pseudo-native data sets were indeed useful to produce new independent phase information in our original study (Li *et al.*, 2013).

### S2. Effects of incomplete HA models on FOM during HA model refinement

Incomplete HA models can generate an overinflated average cosine of estimated phase errors, or figure of merit. Incomplete HA models are not uncommon during early stages of *de novo* HA structure determination (Silvian *et al.*, 1999). In order to obtain a clear solvent boundary for effective density modification and interpretable maps, it is important to work with a set of derivative data with unambiguous phase triangular relationship and phase probability curves in the lower resolution shells (up to 5 Å). Indeed, we found in a previous study that experimental maps calculated using only the 10% strongest reflections at 5.0 Å resolution show a clear solvent boundary, but maps calculated using the remaining 90% reflections did not (Silvian *et al*., 1999). This demonstrates that an average cosine of estimated global phase errors at low resolution may not reflect the interpretability of the solvent boundary in the experimental maps.

Another severe model bias problem associated with incomplete HA models is that the partial HA structure factors substantially differ at low resolution from complete HA structure factors so that phase triangular relationship are incorrect. Additionally, the unaccounted

isomorphous signals from the missing sites contribute to the individual and global lack of closure errors. Due to a combination of these factors, global lack of closure errors can vary by an order of magnitude in the lowest resolution bins during HA model refinement (Silvian *et al*., 1999). In contrast, with complete and correct HA models, the global lack of closure errors in general do not vary much as a function of resolution.

### S3. Multi-crystal multi-domain averaging

After obtaining the best phases using 3-domain 1-crystal averaging from merged U amplitudes, we added a non-isomorphous native data set in density modification to preform 2-crystal averaging. This native data set extended approximately to 3.2-Å resolution and had different unit cell parameters from the merged U derivative. We transferred the experimental maps and phases from the U derivative to the native data set using an identity matrix (which was further refined during averaging) to generate initial phases. Because of severe anisotropy in the native data and the lack of useful, strong isomorphous derivatives, the contribution of native amplitudes to the phase improvement of the U derivative in this procedure was limited. Nevertheless, this procedure provided sufficiently accurate phases for the native data set so that we could determine the other weak derivative substructures using isomorphous difference Fourier maps.

Three data sets from crystals of SeMet-substituted sE2 were isomorphous to the native set. One Se-Met data set was collected at the Se peak wavelength and extended to 3.5-Å resolution in the strong diffraction direction. The other two Se-Met data sets (Se-EMP and Se-TAMM) were collected at the Hg peak wavelength and extended to 4.0-Å resolution. Although the Se-EMP and Se-TAMM crystals were soaked in ethylmercurial phosphate and tetra-kis methyl mercury, respectively, neither compound appeared to bind to the protein. Even though domain II has one free cysteine residue, it is possible that binding of Hg may cause some disorder in the region near the free cysteine residue and thus bound Hg remained invisible. Nevertheless, by correlating isomorphous difference Fourier peaks between the three Se-Met data sets, most of the selenium sites could be unambiguously identified even though the estimated Se substitution level was below 50%. This identification greatly assisted sequence assignment during structure interpretation (Fig. S2). Attempts to include Se-Met SAD phases in averaging were unsuccessful.

### S4. Structure determination of the full length BVDV E2 ectodomain

Using the partially built sE2 model, we identified a molecular replacement solution for the structure of the full length E2 ectodomain (E2-ECD) (Li *et al.*, 2013). There was one disulfide cross-linked E2-ECD dimer in the asymmetric unit as in the sE2 structure. Using transformation matrices derived from the molecular replacement solution, we added the E2-ECD data set to multi-domain averaging calculations, now with three crystals (U123, Native, and E2-ECD), which also suffered from severe diffraction anisotropy (Fig. S1). Without prior knowledge of the location of domain I in the E2-ECD structure, we could not protect it from solvent flattening in the initial step. As a consequence, we could not identify its location or structure in the electron density. Nevertheless, this averaging provided the experimental phases that were useful for constrained refinement of the remainder of the E2-ECD structure.

After phase-constrained refinement, residual $F_o$-$F_c$ difference Fourier maps (where $F_o$ and $F_c$ denote observed and calculated amplitudes, respectively) revealed the missing domain I in both copies of the molecule. In one subunit, many structural features of domain I were recognizable. In the second copy, only the approximate location of domain I was identifiable. However, this allowed us to define manually the solvent boundary of the E2-ECD structure to protect the missing domain I in density averaging and solvent-flattening. This further improved the experimental phases for the E2-ECD structure, and provided better constraints for structure refinement. As a consequence, residual maps at 3.5-Å resolution now revealed additional features for the second copy of missing domain I, allowing a partial atomic model to be placed into the density. From this placement, the exact NCS matrix for domain I was then calculated for averaging. Notably, the E2-ECD data exhibited even more severe anisotropy than the sE2 data with a disparity in diffraction resolution of 3.5 and 6 Å in two orthogonal directions. Multi-domain three-crystal averaging attempts were carried out both at 3.5-Å and 4.0-Å resolution for the E2-ECD data set. The attempt at 3.5-Å resolution resulted in better densities for the missing domain I than the attempt at 4.0-Å resolution.

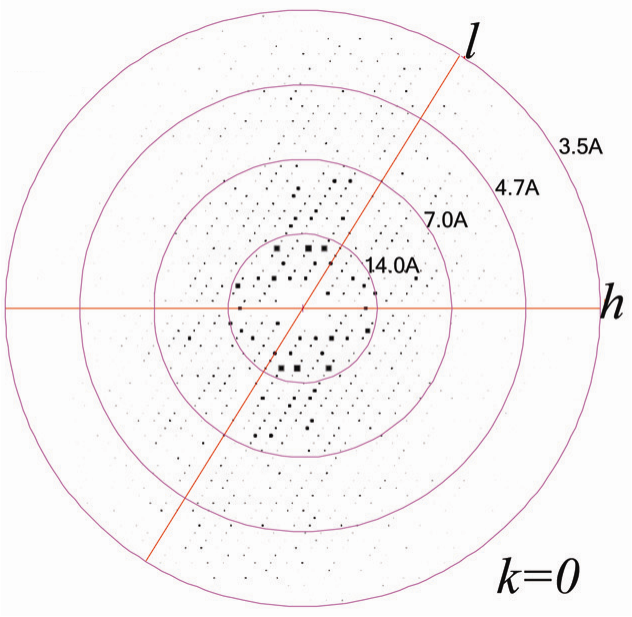


**Figure S1** . Severe anisotropy in the full length BVDV E2-ECD native data set.

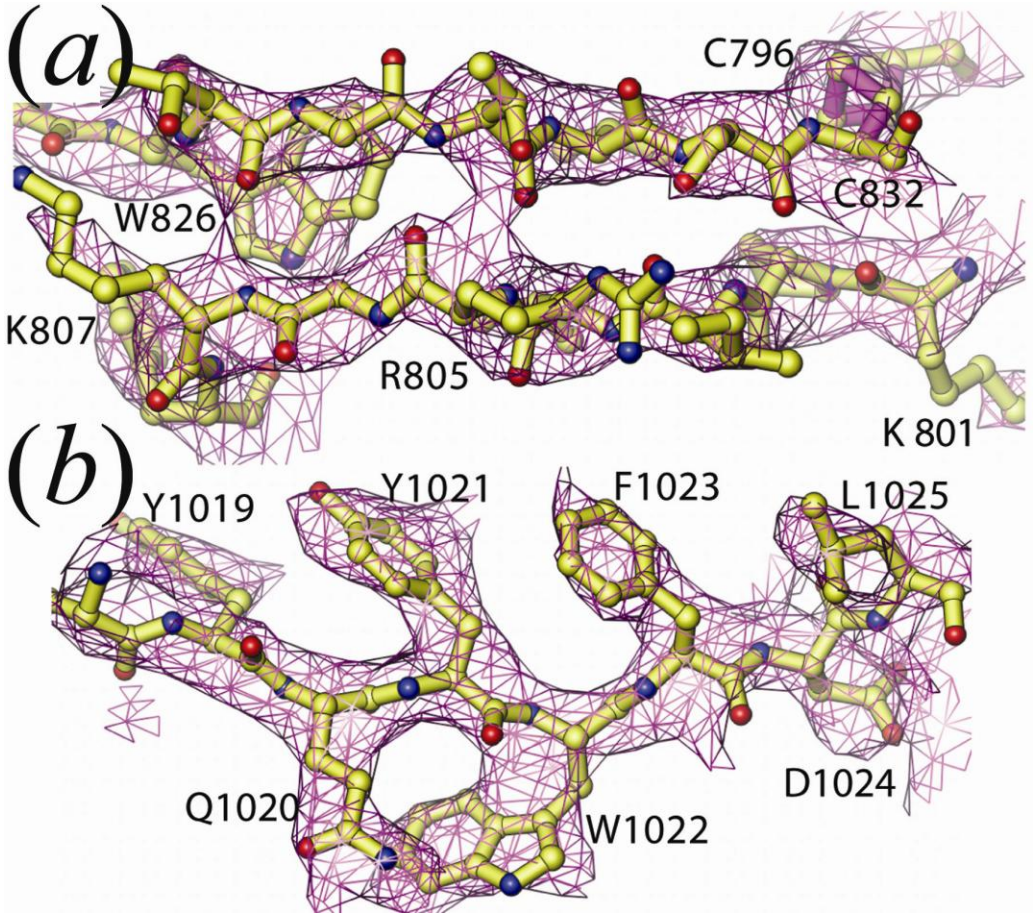


**Figure S2** Multi-crystal multi-domain density-averaged experimental maps. (a) Two β-strands from domain II. (b) A β-strand from domain IIIc. Maps are contoured at 1.75 σ.