**Low Resolution Crystallography in the Hurley lab**

James Hurley
LMB, NIDDK, NIH, DHHS

I recently had a visiting seminar speaker in my office, and was describing some of the new structures in the lab to him. I was excited at the time to have the structure of a biggish multiprotein complex and another biggish full-length mammalian signaling protein to show off. The resolution was low in both cases, but for the biological questions being asked about how domains and subunits interacted, the structures were still informative. My visitor commented that "Your lab seems to have a real problem with resolution." We have had a recent run of low resolution structures, whether due to bad luck, an increased focus on larger proteins and complexes, or both.

Axel Brünger wrote a review suggesting that this is a larger trend, a renaissance in low resolution crystallography (Brünger, 2005). From a crystallographer's point of view, it is never a good thing to be stuck with low resolution data. But from a structural biologist's viewpoint, things look different. The holy grail of the field is an integrated picture of the cell in which structural information is available at every size scale in the cell from atomic resolution crystal structures through light microscopy of whole cells. Our cryo EM colleagues are hard at work improving the resolution of their methods in order to bridge the gap. Low resolution x-ray crystallography also has a big part to play in the chain of imaging techniques as we try to bridge the gap from atomic to cellular structure.

Three examples from my lab follow:

## β2-Chimaerin

β2-Chimaerin is a GTPase activating protein for the small GTPase Rac. It is activated by the lipid diacylglycerol, which binds to the C1 domain. The protein has 466 amino acid residues. A total of 308 amino acids are in the SH2, C1, and RacGAP domains. None of the structures of the β2-Chimaerin domains had been solved, but structures were available for homologs of each domain with 35-45 % sequence identity. A MR search located the RacGAP domain (179 residues with about 35% identity) but it was not possible to locate the rest of the structure by MR. The protein was expressed in insect cells and the protocol we were using at the time for SeMet incorporation in insect cells produced poor yields that precluded using Se MAD. Bertram Canagarajah obtained phases at 3.8 Å using SAD from two native Zn ions and one atom of Xe (Canagarajah et al., 2004). The SH2 and C1 domains were located in real space searches. 91 residues of ordered connectors were manually built in, and the remaining 67 residues were never located. This density would have been very hard to interpret "from scratch", but given the positions of the known domains, it was feasible. The structure was refined against the best native data set, to 3.18 Å. Because the solvent content was 81 %, the ratio of observations to refined parameters was better than usual for a low resolution structure, and no unusual procedures were needed to refine the structure to a free R-factor of 0.290 with reasonable stereochemistry.

This case illustrates how the interpretation of a low-resolution experimental map is greatly facilitated if homologous structures are known for some portion of the structure. It also illustrates how a very high solvent content greatly facilitates refinement.

**ESCRT-II**

ESCRT-II is a trafficking complex that contains three subunits, Vps22 (233 residues), Vps25 (202 residues), and Vps36 (566 residues). At the start of the project, the stoichiometry of the subunits was unknown. None of the subunits had homology to known structures. SeMet protein yielded very poor crystals, and Aitor Hierro and I solved the structure by SIRAS using an Xe derivative at 3.6 Å (Hierro et al., 2004). Preliminary chain tracing coupled with Dali and VAST searches revealed a winged helix fold in one of the subunits. Comparison of this fold to the other subunits revealed that all three subunits consisted of small variations on a tandem repeat of a winged helix motif. This discovery was central to completing the chain trace, assigning the subunits, and determining the subunit stoichiometry (1 Vps22: 2 Vps25: 1 Vps36). We also discovered, and confirmed by protein sequencing, that the N-terminal 395 residues of Vps36 had been lost to proteolysis during the crystallization experiment. Secondary structure predictions using the PredictProtein server were critical for the sequence assignment, which was otherwise handicapped by the absence of SeMet markers. In hindsight, I think we should have tried harder to improve the SeMet crystals, if only for use as sequence markers. After the trace was completed, S anomalous data were collected and good agreement was found when the anomalous map was overlaid on the model. The C-terminal domain of Vps25 was the least well defined in the electron density, and not surprisingly gave the poorest S anomalous signal. Refinement at 3.6 Å was not completely satisfactory, with little map improvement was seen over the course of refinement. Once again a very high solvent fraction facilitated all-atom refinement, providing a higher number of observations than would have been present at this resolution for a protein of this size at a more typical solvent content.

This case illustrates the critical role of secondary structure predictions and of using partial structures to search the PDB for related folds.

**ESCRT-I**

ESCRT-I is another trafficking complex that contains three subunits, in this case, Vps23, Vps28, and Vps37. Intact ESCRT-I did not behave well enough to crystallize, so Mike Kostelansky used limited proteolysis to map a stable core consisting of the C-terminal 65 amino acids of Vps23, the N-terminal half of Vps28, and the C-terminal half of Vps37 (Kostelansky et al., 2006). Vps23 and Vps28 form a tight binary complex that can be purified and crystallized in the absence of Vps37. None of the core domains of these subunits had any homology to known structures. Mike crystallized and solved the structure of the Vps23:Vps28 subcomplex at 3.5 Å by SeMet MAD. The structure showed that the last seven amino acids of Vps28 were disordered. Rather than attempt to completely refine the structure at 3.5 Å, Mike engineered the seven-amino acid C-terminal deletion of Vps28 and obtained a new crystal form of the truncated binary complex that diffracted to 2.1 Å, and solved it by MR with the partially refined 3.5 Å structure. Information from these crystal structures was applied to engineer a form of the ternary complex that diffracted to 2.8 Å, which Mike solved by molecular replacement with his binary complex structure.

This story illustrates that a low resolution structure is not necessarily the end of the line. Low resolution analyses may be good enough to engineer variants that diffract well.

## Perspectives

Chain tracing, including sequence assignment, and refinement, both have difficulties at low resolution.

Chain tracing is assisted by applying all available knowledge: known structures of sequence homologs, structural homologies discovered after a partial trace, and secondary structure predictions are all invaluable. What will it take to automate low resolution chain tracing? Given the growing power of protein structure prediction for small structures (Schueler-Furman et al., 2005), it seems that it should be possible in principle to integrate structure prediction and automated density interpretation.

The availability of SeMet markers is invaluable as well. Given the increasing use of eukaryotic expression systems for large proteins and complexes that often diffract to low resolution, protocols for efficiently incorporating SeMet in these systems are important. We had problems reproducing the procedure of Bellizzi et al. (1999) in Sf9 cells, but others report success in *T. ni* (Hi-5) cells (Chen et al., 2005). My lab has switched from Sf9 to Hi-5 cells and we plan to revisit this procedure. The lack of side-chains at very low resolution impedes analysis, and B-factor sharpening has been advocated by Brünger and Harrison as a way to reveal side-chains (Brünger, 2005) Chen et al., 2005). I haven't tried this myself, but it sounds worth exploring, despite the obvious risk of amplifying noise.

In order to stabilize low-resolution refinements, Brünger advocates using a phased target function (the "mlhl" target function in CNS) (Brünger, 2005). I tried this after the fact for ESCRT-II using the flattened SIRAS phase probabilities (admittedly not great) and it helped to the tune of only ~ 1 % improvement in the free R factor. It might help more in a case of lower resolution, lower solvent content, or higher phase quality. While one should always strive to get the best phases, for low resolution structure determinations it is especially worth maximizing phase quality with an eye to using experimental phases in refinement.

## References


Bellizzi, J. J., Widom, J., Kemp, C. W., and Clardy, J. (1999). Producing selenomethionine-labeled proteins with a baculovirus expression vector system. Structure With Folding & Design *7*, R263-R267.

Brünger, A. T. (2005). Low-resolution crystallography is coming of age. Structure *13*, 171-172.

Canagarajah, B., Collucio Leskow, F., Ho, Y. S. J., Mischak, H., Saidi, L., Kazanietz, M. G., and Hurley, J. H. (2004). Structural Mechanism for Lipid Activation of the Rac-Specific GAP, β2-Chimaerin. Cell *119*, 407-418.

Chen, B., Vogan, E. M., Gong, H. Y., Skehel, J. J., Wiley, D. C., and Harrison, S. C. (2005). Determining the structure of an unliganded and fully glycosylated SIV gp120 envelope glycoprotein. Structure *13*, 197-211.

Hierro, A., Sun, J., Rusnak, A. S., Kim, J., Prag, G., Emr, S. D., and Hurley, J. H. (2004). Structure of the ESCRT-II endosomal trafficking complex. Nature *431*, 221-225.

Kostelansky, M. S., Sun, J., Lee, S., Kim, J., Ghirlando, R., Hierro, A., Emr, S. D., and Hurley, J. H. (2006). Structural and functional organization of the ESCRT-I trafficking complex. Cell *125*, 113-126.

Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., and Baker, D. (2005). Progress on modeling of protein structures and interactions. Science *310*, 638-642.