Refinement in cases of Twinning

Lothar Esser

National Cancer Institute

Despite working in crystallography for a number of years, it was only recently that I was asked to refine a structure against twinned data. The unfinished structure was handed to me after the post-doc who had been working on this project had left the lab. The crystal diffracted x-rays to 2.2 Å with intensities obeying the symmetry of space group R3. The structure was solved by molecular replacement before it was recognized that the crystal was twinned using a model of > 99% sequence identity. As the refinement did not progress as it should, it became clear that the data were twinned. Merohedral twinning in R3 is quite common and the extra 2 fold axis, that generates the twin domain, relates reflection indices of h, k, l to k, h, -l. Having no prior experience with the refinement of structures against twinned data, I tried all programs that were available to me. Knowing that SHELX has a long tradition in handling complex twinning problems and given its ability to handle macromolecules, shelxh was my first choice. In parallel with SHELX, I used phenix.refine of the PHENIX package. The development of phenix.refine is ongoing but it appears that it might develop into a major player in the field of macromolecular structure refinement. An important feature of phenix.refine is its ability to refine the twin fraction. The other two refinement programs that I used, refmac5 and CNS, require either an input of detwinned data or the prior knowledge of the twin fraction.

Although the starting model was quite complete, numerous loops and residue side chains needed adjusting. Therefore, it was important to me to be able to go back and forth between refinement and model building programs. Here the programs refmac5/CCP4i and phenix.refine stood out. No complicated map conversion is required and O or Coot work flawlessly with the generated maps or even mtz files (in the case of Coot). CNS with its refine.inp script works nearly as well but the giant sizes of the maps slow O down and for Coot, a map needs to be converted into ccp4-style using external scripts. Shelxpro's file handing and map generation is by far the most cumbersome. More than the existing differences in convenience however, what struck me as utterly surprising, if not shocking, were the differences in the quality of the maps produced by different programs. Of course, I'd be hard pressed to quantify the quality of a map by any objective means. In my subjective opinion however, a good map is characterized by its ability to point out the right from the wrong. To my surprise, SHELX maps are biased to the point of uselessness. No matter what I did, I obtained a gorgeous map back from refinement that confirmed virtually the whole structure as it was. The maps I am referring to were properly sigma-A weighted, unsharpened maps. Similarly disappointing was phenix.refine which produced what appeared to be FC maps. This problem was resolved by upgrading to the latest version of PHENIX. The current version 1.3b delivers maps that were

comparable in quality with those from CNS. However, no maps are more useful in revealing model errors or omissions than maps produced with refmac5 (and remember this comes from a dyed-in-the-wool CNS user).

Not strictly related to twin refinement, another issue for me was the handling of NCS restraints. There were two molecules in the asymmetric unit and maintaining reasonably tight restraints are important. SHELX uses torsion restraints to deal with NCS, which makes it exceptionally easy to setup and nearly worry free. However, being used to the fine-grained approach as implemented in CNS which allows the user to exclude single side chains (even single atoms if need be) I still feel more comfortable with CNS and phenix.refine. Phenix.refine uses syntax almost identical to that in CNS and even further facilitates the setup of inclusions or exclusions. Compared to this, the implementation of NCS in refmac5 with no control over side chains other than in broad terms like “tight, medium and loose” seems a bit crude.

So, what did I end up doing? Well, map quality trumps all. Using the best model available, I refined the twin fraction in phenix.refine and used this value to detwin the original data and convert them into the proper mtz format. Subsequently, refmac5 was used for model refinement. A re-refinement of the final model in phenix.refine showed that the twin fraction had not changed which eliminated any adjustment of the detwinned reflection file. Also, at no time, TLS or equivalent (grouped) anisotropic b-factor refinement was used. The crystallographic R-factors were about 22% (work) and 28% (free).

Summary:

Program	Twin fraction refinement	Map quality	NCS handling	Non protein ligand handling	Recommen-dation
shelxh release 97-2	yes	poor	very good	DFIX, DANG etc.	No[1]
phenix.refine 1.3b July07	yes	good	very good	CIF	N/A[2]
CNS 1.1	no	good	very good	Top/Param	Yes
Refmac5 5.2.0019	no	best	fair	CIF	Best[3]

[1]If rebuilding is required SHELX maps at low resolution (2.2 Å) are useless. If there is a way to obtain unbiased SHELX maps, please let me know.
[2]Unresolved issues with b-factor restraints prevent me from endorsing this program all the way. Future versions may get an unqualified recommendation.

Links:
Shelx twin http://shelx.uni-ac.gwdg.de/~rherbst/twin.html
Shelxh http://shelx.uni-ac.gwdg.de/SHELX/

Phenix.refine	http://www.phenix-online.org/phenix_wiki
CNS 1.1	http://cns.csb.yale.edu/v1.1/
Refmac5	http://www.ysbl.york.ac.uk/~garib/refmac/docs/index.html