Model building, refinement and validation

Roberto A. Steiner[a]* and Bernhard Rupp[b]*

[a]Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London, SE1 1UL, UK, and [b]k.-k. Hofkristallamt, San Marcos, CA 92078, USA

Correspondence e-mail: roberto.steiner@kcl.ac.uk, br@hofkristallamt.org

The CCP4 Study Weekend 2011 was held at the University of Warwick on the 6–7 January. Following a long established tradition of discussing at the Study Weekend the most important crystallographic topics, the choice for 2011 was the 'Model building, refinement and validation' triple bill.

As a result of the extraordinary efforts of instrumentation and software developers, macromolecular crystallography has been greatly automated to a point where it can be treated as a black-box method with little input required by the user. This is not without danger. The complexity and size of problems tackled has kept pace with the developments, and competent work still requires a great deal of knowledge about the methods and, naturally, about the source of the remaining challenges. In the process of structure solution, model building is often a very satisfying step in which the scientist acquaints him/herself with the macromolecule(s) under study and biological hypotheses start taking shape. It is however not without hurdles and potential pitfalls, particularly when only data at limited resolution are available. The best possible electron-density maps which guide further model (re)building and afford additional biological insight are calculated as part of crystallographic refinement. In this step one maximizes the agreement between the X-ray data and our (atomistic) interpretation of the diffraction experiment. Although the stereochemical restraints employed during the crystallographic refinement try to maintain a chemically sound model, constant and concurrent validation is required. Robust and easily accessible tools for global and local quality analyses are critical to ensure that reliable structures are made available to the scientific community at large.

The Study Weekend 2011 was opened with a talk by Bernhard Rupp who provided an overview of the key challenges that are still present in the field of macromolecular crystallography with a particular focus on the three main topics of the meeting. The challenges originate from two major sources. On one hand there are difficulties resulting from the complex nature of the biomacromolecules self-assembling into an imperfect crystal (such as dynamic motion, disorder, limited diffraction, twinning and lattice modulation). On the other hand there are fundamental problems inherent in a highly multivariate parameter space with an often barely sufficient data-to-parameter ratio. Bernhard's talk was followed by one by George Sheldrick. George further developed the general themes outlined in the previous talk and described in more detail the foundations of crystallographic refinement emphasising the importance of proper restraints. After the introductory talks which set the scene for the topics to be discussed, the first session of the meeting focussed on model building. Kevin Cowtan discussed recent developments on the automated model building software *BUCCANEER* and Paul Emsley presented some of the new tools available in the very popular *Coot* package. Isabel Uson presented the innovative *ARCIMBOLDO* program for *de novo* phasing and building using many protein fragments. Alwyn Jones then went on to discuss pitfalls in low-resolution model building and some of the tools available to deal with them in his venerable model building program *O*. The session was closed by Willy Wriggers who described low-resolution model building and refinement tools (particularly relevant to EM and SAXS applications) in *SCULPTOR* and *SITUS* that take advantage of spatial coarse graining or tessellation methods. Day 1 of the meeting continued with a session on crystallographic refinement. Pavel Afonine gave an overview of the *PHENIX* package with particular emphasis on the extensive set of tools available in *phenix.refine* for macromolecular refinement. Oliver Smart then went on to discuss local structure similarity restraints (LSSRs) which are particularly useful for the automated setup of restraints between multiple NCS related copies in the asymmetric unit. Providing easier NCS restraint setup should convince crystallographers to use these somewhat underappreciated source of redundancies to the fullest. Also discussed were quantum-mechanics-based restraints

available in the new release of *autoBUSTER*. Another interesting approach was presented by Jeff Headd, who discussed knowledge-based restraints for low-resolution structure refinement in *phenix.refine*. Here, the generic distributions normally used are replaced with sequence-dependent restraints from a homologous high-resolution reference model. Secondary-structure-dependent restraints preserve structural features that often become distorted in low-resolution refinement. After the information-heavy evening session the participants continued scientific discourse during a fine conference dinner, which was followed by entertainment and dancing, when the spirits relaxed and sanity was re-established.

The morning session of Day 2 was dedicated to low-resolution refinement, twinning and complex cases. Axel Brunger presented the application of Deformable Elastic Network (DEN)-refinement and automated model building to the difficult case of the putative succinyl-diaminopimelate desuccinylase from *Corynebacterium glutamicum*. Garib Murshudov, the lead developer of *REFMAC*5, introduced several innovative approaches available in the latest version of the software to overcome the severe under-determination that hampers low-resolution refinement, including advanced secondary-structure-restraint approaches and map-reconstruction methods that enhance the signal without amplifying noise too much. One of the most insidious hindrances to refinement can be complex twinning. Pietro Roversi discussed an interesting case of tetartohedral twinning in triclinic crystals of the human complement factor I including structure solution by molecular replacement with *PHASER* and crystallographic refinement with *REFMAC*5. After a short tea break the focus shifted to the building and refinement of structures containing nucleic acids and ligands. This latter topic is of particular interest to academic laboratories and to industry sectors employing macromolecular crystallography for structure-based drug design. Judit Debreczni introduced new additions that have been made to *Coot* to enable better support for various aspects of ligand building, analysis and validation. *Coot* interfaces to external applications relevant for handling small molecules, such as the *CCP4* programs *JLigand*, *LIBCHECK* and *CPRODRG* for the description of novel monomers and links between residues and the *CSD* tool *Mogul* for validation of geometric parameters of ligands. The next talk in the ligand building session focussed on the *CCP4* program *Jligand*, presented by Andrey Lebedev. *JLigand* provides a graphical user interface that helps to create description of ligands and covalent links. It currently uses *Libcheck* to create descriptions of restraints and initial three-dimensional coordinates, and *REFMAC*5 for optimization of the coordinates. The complete description of a ligand can be created from the connectivity graph drawn in the *JLigand* GUI or imported from a smiles string, sdf, mol2 or mmCIF files. Model building and refinement of nucleic acid crystal structures differs to some degree from that of protein-only structures. Bill Scott explained that as a consequence of the great similarities between base pairs and canonical forms, all nucleic secondary structural elements tend to appear quite similar, making sequence assignment and backbone tracing more daunting. Accurate sequence data and biochemical constraints to augment and double-check a crystallographically derived structure are thus quite important. Victor Lamzin introduced a new automated NCS detection and extension tool for the *ARP/wARP* program together with other new features of this popular automated model building program.

The afternoon session was dedicated to validation, and lead into by Frank von Delft presenting approaches to NCS and cross-crystal structure (CCS) validation. Frank covered the use of NCS and CCS in both refinement and validation. Ian Tickle analysed and discussed means and statistics of electron-density map quality assessment. He pointed out that real-space R value and correlation coefficient suffer from the inability to distinguish accuracy and precision, and suggested an improved, likelihood based χ^2 measure termed the real-space difference density Z score (RS-DZ), a measure purely of the local model accuracy. Ethan Merrit then went on to remind everyone that in choosing and refining any crystallographic model, there is tension between the desire to extract the most detailed information possible and the necessity to describe no more than what can be justified on the basis of the observed data. It is therefore important to validate the choice of model parameterization analogous to validation of the stereochemistry. Programs relevant to the choice, construction, and validation of model parameterization include *PARVATI*, *TLSMD*, and *TLSANL*. Sameer Velankar then informed us about the future of validation at the wwPDB. He summarized the findings and recommendations of the X-ray Validation Task Force, and described the design and implementation plans for the wwPDB validation pipeline that will become the common deposition tool for structural data using all experimental techniques at all wwPDB deposition sites in 2012. In the last talk of the meeting, Robbie Joosten presented new developments within the *PDB_REDO* effort including new challenges in structure optimization and the possibilities for practicing crystallographers to proactively use this pipeline before submitting a structure model.

Concluding remarks of the meeting included an appeal by the organisers, pointing out that the many remaining challenges in refinement, model building, and validation require equally sophisticated software to handle them, which can only be produced if young talents step up to the plate and contribute to the truly collaborative effort that crystallographic software development always was and still is. As hard as the funding for crystallographic software development has become, the intellectual satisfaction from it and its far-reaching impact on the entire community are still well worth the effort.

The present issue collates original research articles based on the talks given at the CCP4 Study Weekend 2011. Not all authors felt that their oral contribution required an accompanying article. In particular, Bernhard and George felt that excellent introductory material on the meeting's topics is already available in the literature (for example, Sheldrick & Schneider, 1997; Tronrud, 2004; Rupp, 2009; Sheldrick, 2010).

The reader is therefore directed to the references provided to cover the basics. Finally, we would like to express our gratitude to the CCP4 staff, and in particular Shirley Miller, for their invaluable help with the practical aspects of the organisation of the 2011 Study Weekend. We also thank all our speakers for their excellent talks and all contributors to this issue of *Acta Crystallographica Section D*.

References

Rupp, B. (2009). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, 1st ed. New York: Garland Science.

Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Tronrud, D. E. (2004). *Acta Cryst.* D**60**, 2156–2168.