Apotheosis, not apocalypse: methods in protein crystallography

Victor S. Lamzin,[a]* Anastassis Perrakis,[b] Gerard Bricogne,[c] Jiansheng Jiang,[d] S. Swaminathan[d] and Joel L. Sussman[e]

[a]European Molecular Biology Laboratory, Hamburg Outstation, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany, [b]European Molecular Biology Laboratory, Grenoble Outstation, c/o ILL, BP 156, 6 Rue Jules Horowitz, 38043 Grenoble CEDEX 9, France, [c]MRC Laboratory for Molecular Biology, Hills Road, Cambridge, England, [d]Biology Department, Brookhaven National Laboratory, Upton NY 11973-5000, USA, and [e]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Correspondence e-mail:
victor@embl-hamburg.de

Automation of the steps in deriving a complete structural model from crystallographic data was discussed at a workshop *Toward Automation of Structure Determination in Macromolecular Crystallography* that was held on 20–21 June 1999 at the Biology Department of the Brookhaven National Laboratory under the joint auspices of the US Department of Energy, Brookhaven Science Association and the European Molecular Biology Laboratory (EMBL). The workshop gathered some 25 world-recognized experts, actively working in the field, for a frank exchange of ideas and a thorough analysis of the current status and the future requirements.

The motivation for such a meeting was to identify potential stages requiring intensive human effort in high-throughput crystal structure determinations within structural genomics projects. Several such projects are about to be scaled up from the pilot level to a more massive production mode (see, for example, the National Institute of General Medical Sciences' Structural Genomics Initiative, http://www.nigms.nih.gov/funding/psi.html). The workshop focused on one of the well known bottlenecks: the determination and refinement of three-dimensional structures subsequent to the wet-lab stage of obtaining suitable crystals of biological macromolecules. As in the human genome project, large-scale high-throughput structural genomics also requires an assembly line approach and automation of as many steps as possible.

A widely shared opinion is that all that is needed is the tying together of existing methods and programs with a user-friendly interface, assembling currently known protocols into a 'black box' set of modules; in this way, the task would essentially be performed. However, it was strongly pointed out at the workshop that the task of automation should not be viewed from too narrow a point of view. Instead, the discussions brought to light a need to re-examine the whole process of structure determination, including data collection, phasing, model building and refinement, from concepts to ergonomics. There was a consensus that while until the recent past this field of investigation was relegated to a few research groups with very limited resources, a renaissance of methods research is now beginning, largely driven by the need for fundamentally new approaches to keep up with the requirements of high-throughput projects.

The essential goal is high-throughput structure determinations and for this the performance of any automatic software will be crucially dependent on the quality of experimental data. Since the vast majority of the data will be recorded at synchrotron sources (Helliwell, 1998), the first step towards automation is the provision of on-site data-analysis facilities directly coupled with data collection. This will allow rapid decisions to be made on the feasibility of the data collection: whether it has to be completed or if there is a need for use of another crystal or even the next protein in the queue. Following the paradigm established by scientists at several beamlines, the end result of the crystallographic data collection should be at least an integrated data set and preferably an electron-density map or a molecular model. The experimental goal must be to obtain sufficient data to solve the structure to the level of detail required, not merely as much data as the available time will allow. This increase in local capabilities would benefit not only structural genomics projects but also the wider structural molecular biology community.

The synergy between molecular biology and the MAD technology has completely changed the concept of the crystallographic phase problem and has helped to clarify where the greatest effort should be placed to achieve maximum speed in structure determination together with an improvement of quality of crystal structures. On the other hand, the number of entries in the PDB, the database of macromolecular structures, is increasing exponentially. It is anticipated that this growth will accelerate further with the launch of high-throughput projects. Developments in methodology are foreseen in the very near future that will exploit the individual structural entries as well as the topological motifs for the use as search models in the 'generalized' molecular-replacement methods for solving new structures.

A major excitement that has emerged during the meeting was the realization that phasing, model building and refinement in crystal structure determination are much more tightly coupled than they are currently believed to be. Routine but elaborate use of computer graphics for manual construction of models has

now reached the stage where time-consuming steps urge for automation. A presentation of methods aiming at objectively automating electron-density map interpretation without a need for interactive graphics was one of the highlights of the workshop. The methods described now appear to be applicable to lower resolution X-ray data; these are often the most difficult structures to unequivocally determine and refine. The use of diverse databases ranging from primary sequences and topological folds even to shapes of electron densities to aid in interpreting available experimental data is emerging. The main focus in the use of computer graphics by the research scientist should be in the detailed analysis of the models, in particular those features which are specific to a structure of interest, *e.g.* post-translationally modified amino-acid residues, bound ligands and multimolecular assemblies, rather than primarily in the labour-intensive building of the model.

At the end of the meeting it was felt that the best way to foster development of the methods of structure determination would be to encourage *ad hoc* collaborations. This is specifically meant to include close interactions and feedback between end-users and software developers. This will provide a testing ground for the various new approaches and methodological developments. One of the major conclusions was the realization that improvements in the streamlining of structure determination will lead to a re-evaluation and reconsideration of the underlying methodology that will in the future have a strong impact on the ability to address more challenging biological problems. The impact of these newly developing methods will have a direct impact on the cutting edge of areas of modern molecular biology that involve multicomponent complexes.

Fears have been expressed that the emerging emphasis on high throughput might result in 'the end of macromolecular crystallography as we know it'. For instance, the choice of targets of unknown function might lead to inferior accuracy in their crystal structures. We recognize that the responsibility for accurate construction and thorough validation of macromolecular models will naturally shift from investigators to the methods developers. We believe that this challenge will be met and that conventional projects will benefit from these developments – hence the title of this report.

Details of the meeting are available from http://vivaldi.bio.bnl.gov.

References

Helliwell, J. R. (1998). *Nature Struct. Biol.* **5** (*Suppl.*), 614–617.