Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Opportunities and challenges with the growth of neutron crystallography

Paul Adams[a,b] and Paul Langan[c]*

[a]Lawrence Berkeley National Laboratory, CA 94720, USA, [b]Department of Bioengineering, UC Berkeley, CA 94720, USA , and [c]Los Alamos National Laboratory, NM 87545, USA

Neutron crystallography is a newly blossoming field of structural biology. As described in the historical account given by Benno Schoenborn in this issue, its origins can be traced back to the Cambridge MRC tearoom in 1965 when it was discussed as the only plausible experimental method for directly determining the location of H atoms in proteins. Knowing the location of H atoms can provide information on the protonation states of amino-acid residues and ligands, the identity of solvent molecules, and the nature of bonds involving hydrogen. Neutron crystallography can also be used to identify H atoms that are exchanged with their isotope deuterium (deuteration) and the extent of this replacement, thus providing a tool for identifying isotopically labeled features, for studying solvent accessibility and macromolecular dynamics, and for identifying minimal protein folding domains. This unique information, which is often difficult or impossible to obtain using X-ray crystallography, is important for understanding protein function and enzyme mechanism. However, it wasn't until 1968 that neutron crystallography was first used to study a protein at the nuclear reactor run by Brookhaven National Laboratory. With limited equipment and a relatively low flux beam, it took several months to collect a 2 Å data set from the 17 kDa transport protein myoglobin using a crystal that was several tens of mm^3 in volume.

Over the years since that beginning, growth has been relatively slow mainly because there have been a limited number of available instruments and their fluxes have been relatively weak compared with X-ray beams, therefore requiring larger crystal sizes. However, the uniqueness of the type of information that can be provided by neutrons, and its complementarity to the information provided by X-rays, has given neutrons a small but important role in biology in the past. Recently, the prospects for this field have changed dramatically and there has been great increase in the application of neutrons in biology. This can be related to improvements in beamline instrumentation, neutron sources, data collection and sample preparation methods, and new approaches to and computational tools for structure determination. These advances are pushing practicable sample sizes down to fractions of a mm^3, data collection times down to a few days or even hours, and are allowing increasingly complex biological systems to be studied. Growth is bound to continue with the current construction of new and more powerful spallation neutron sources throughout the world and continued improvements to existing facilities. During this period of growth several new opportunities and also challenges have arisen. The recent International Conference on Neutrons in Biology reported in this issue was a rare opportunity for the neutron macromolecular crystallography community to come together to discuss these issues and to develop a common vision for the future development of the field.

One of the most immediate challenges discussed was an inconsistency in the way neutron structures are reported and deposited. Now is an ideal time to address this issue while the number of neutron structures is still relatively small and therefore easy to rehabilitate. Neutron structure refinement uses similar computational methods to those for X-ray structure refinement. However, the resulting models are typically more complex as they must explicitly model

hydrogen and/or deuterium atoms. When the crystal has only been partially deuterated atomic sites will typically contain a mixture of hydrogen and deuterium atoms, which must be accurately represented in the model. In addition, a joint X-ray and neutron (XN) structure refinement method has been developed for macromolecules. XN refinement was first applied to proteins by Wlodawer and Hendrickson (1982) and more recently it has been combined with advances in modern computational methodologies in a generalized implementation as described by Pavel Afonine *et al.* in this issue and Adams *et al.* (2009). This powerful approach generates a single model that is simultaneously related to two data sets (one X-ray, one neutron). These unique features of modern neutron structure refinement have the potential to create challenges for structure deposition at centralized databases.

The worldwide Protein Data Bank (wwPDB; Berman *et al.*, 2000) is an outstanding international resource that is the repository for macromolecular structures derived from multiple experimental techniques including X-ray and neutron crystallography, NMR, and electron microscopy. As such the wwPDB deposition sites must process very complex user files describing experiments ranging from the routine to the unique. In the past the lack of computational tools to appropriately refine neutron structures has lead to some structure depositions which poorly represent the atomic model fit to the data. Some of the problems that can be encountered have been reported by Afonine *et al.* (2010) elsewhere. Briefly, incorrect modeling of hydrogen/deuterium exchange, hydrogen/deuterium occupancies that do not sum to unity, mislabeled diffraction data (intensities are really amplitudes or *vice versa*) and negative occupancies can be found. We anticipate that these problems can be avoided by the use of new refinement tools, expressly developed for neutron structure refinement. The development of new tools will bring challenges in the form of joint XN structure refinement. Deposition of such structures in the wwPDB will need to capture a single model linked to two different diffraction data sets. Working with the RCSB it has been possible to create a deposition header that captures many facets of the joint refinement results. However, the information is currently stored as REMARK 3 records in PDB format files. In the future it will be desirable to provide the information about multiple data sets in a more content rich manner, for example the presence of multiple CRYST1 records to describe the two different data sets.

Ultimately, joint XN refinement methods will probably generate two structures (one for each experimental data set) that have been intimately linked by the refinement process. These structures, and their experimental data, will need to be deposited as a single entry and reported to users of the wwPDB as being connected. It is likely that this will become an increasing trend in macromolecular structure determination, where multiple sources of information will be used to generate structures. Therefore, we strongly support the wwPDB working with software developers and structural biologists to define the next generation of structure deposition tools and robust representations of structures to accommodate these future developments.

Another issue discussed was how to consistently report and use neutron data statistics. For IUCr journals, authors are encouraged to describe clearly the effective resolution of their data. At a minimum a table is usually required that reports the values of $R_{\rm merge}$, multiplicity, average I/σ and percentage completeness for the overall data set and as a function of resolution. From these statistics an effective resolution can be determined; typically the resolution at which the completeness of the observable data falls below 70% or the mean value of I/σ is below 3. We believe that the same reporting criteria should be applied to neutron data. However, we also believe that it is useful to report neutron data statistics, and to use neutron data, beyond the effective resolution because they can still have high information content. With both X-ray and neutron data an effective temperature factor causes diffracted intensities to decrease with $\sin\theta/\lambda$. For X-ray data there is the additional dependence of the atomic form factor on $\sin\theta/\lambda$. However, with neutron data the form factor remains constant with $\sin\theta/\lambda$. Therefore, the mean I/σ and completeness typically fall off more slowly for neutron data. Furthermore, since the relative contribution of the atomic form factor for a H atom falls off more rapidly with respect to $\sin\theta/\lambda$ than for other atoms contributing to X-ray data, but remains the same for neutron data, the information content available for the location of H atoms can still be high in neutron data even beyond the effective resolution when completeness is below 70%. For these reasons it can be appropriate to include X-ray and neutron data sets in joint XN refinement that have very different limits of resolution and completeness. Therefore, because of the higher information content of neutron data, we encourage the full reporting of neutron data statistics and the use of the neutron data in refinement beyond the limits typically defined using the criteria applied to X-ray data.

The development of new computational tools and approaches to neutron and XN refinement was widely recongnized at this meeting as an example of a collaborative effort which has greatly benefited the entire

neutron biomacromolecular community. However, there are other potential collaborative efforts that would also be widely beneficial. Typically, structural biologists find neutron crystallographic studies a challenge at many stages of the process. The common experience of scientists at this meeting who operate neutron macromolecular crystallography beamlines at facilities across the world is that by providing the structural biology community access not only to neutron beam time but also to facilities for deuteration, the expression of proteins, purification and crystallization, the synthesis of substrates with stable isotopes, and also support for data reduction and structure analysis, then the chances of a successful outcome are greatly increased. New and optimized methods for deuteration and crystallization were identified as key areas that would particularly benefit from more rapid development through stronger international collaborations between existing facilities.

The currently operational macromolecular crystallography beamlines, together with those being constructed or envisioned for the near future, have very different characteristics and capabilities. They are optimized for different resolutions, sample sizes, unit-cell sizes, levels of perdeuteration, and data-collection rates. Together they constitute a complementary international pool of instruments. While these beamlines remain heavily oversubscribed, it is important that the international community has open access to all of them, so that the most appropriate instrument can be best matched to any particular experiment. The beamline scientists present at this meeting agreed to work towards maintaining and enhacing this pool of complementary capabilities and to making it available to the whole structural biology community. We also believe that the field of neutron biomacromolecular crystallography is still in its infancy, and that there are further opportunities to explore, particularly with future long pulsed spallation neutron sources.

Finally, it was recognized that the true power of neutron crystallography is in its combination with X-ray crystallography and other highly complementary methods such as deuteration, NMR and quantum chemistry. X-ray and neutron crystallographic techniques provide complementary information on the structure and function of biological systems. However, in order to fully exploit these complementarites a mechanism has to be established to allow users to easily collect more than one type of experimental data set from their samples. Our long-term vision is for partnerships between central user facilities that would establish a highly collaborative and multidisciplinary capability that combines neutron and X-ray experimental methods in new ways with molecular biology, NMR, quantum chemistry and computational methods.

References

Adams, P. D., Mustyakimov, M., Afonine, P. V. & Langan, P. (2009). *Acta Cryst.* D**65**, 567–573.

Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Headd, J. J., Moriarty, N. W., Richardson, J. S., Richardson, D. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2010). *J. Appl. Cryst.* **43**, 669–676.

Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P. & Adams, P. D. (2010). *Acta Cryst.* D**66**, 1153–1163.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Research* **28**, 235–242.

Schoenborn, B. P. (2010). *Acta Cryst.* D**66**, 1262–1268.

Wlodawer, A. & Hendrickson, W. A. (1982). *Acta Cryst.* A**38**, 239–247.