

## Introductory overview: X-ray absorption spectroscopy and structural genomics

Isabella Ascone,<sup>a\*</sup> Roger Fourme<sup>b</sup> and S. Samar Hasnain<sup>c</sup>

<sup>a</sup>LURE, Bâtiment 209E, Université Paris-Sud, 91898 Orsay CEDEX, France, <sup>b</sup>Synchrotron SOLEIL, Bâtiment 209H, Université Paris-Sud, 91898 Orsay CEDEX, France, and <sup>c</sup>Daresbury Laboratory, Warrington WA4 4AD, UK.  
E-mail: isabella.ascone@lure.u-psud.fr

A special issue of the journal is presented, dedicated to biological applications of X-ray absorption spectroscopy (BioXAS) and examining the role of this technique in post-genomic biology. The issue confirms that BioXAS has come of age and it can be expected to make a significant contribution in the structural genomics effort on metalloproteins, which are estimated to make up about 30% of proteins coded by genomes.

**Keywords:** EXAFS; XANES; BioXAS; metalloproteins; structural genomics.

### 1. Current status of BioXAS

The biological applications of X-ray absorption spectroscopy (BioXAS) remain an active area of research at many synchrotrons and has been extensively reviewed (see, for example, Hasnain & Hodgson, 1999). This special issue on 'BioXAS and Structural Genomics' follows the study weekend meeting on the 'Contribution of BioXAS to structural genomics: developments in theory and refinement methods' that was held at LURE (<http://www.lure.u-psud.fr/Congres/BioXAS-SWE/>) on 30 June–1 July 2001. The concept of the study weekend originated from the European BioXAS community discussion at the second European BioXAS meeting held at Orsay in July 2000. The discussion highlighted the need for a forum where collaborative efforts could be promoted so that the technique can be advanced and accessed by non-experts from the wider biological community. The focus on developments in theory and refinement methods is one of the clear examples where collaborative efforts are essential in much the same way as has been achieved by the Crystallographic Computing Project Number 4 (CCP4) in protein crystallography (PX).

This issue is likely to prove a resource for beginners as well as experts as it provides for the first time a 'comprehensive' documentation of the current status of the BioXAS field. It is divided into three sections. The first section consists of contributions on the Paris-Sud initiative, a canonical structural genomics programme (Quevillon-Cheruel *et al.*, 2003), the complementarity of XAS and X-ray diffraction, experimental requirements of BioXAS (Ascone *et al.*, 2003), and CCP4 (Winn, 2003), an initiative which may be of interest for XAS developments. The second section is devoted to theoretical aspects of the XAS technique. The final section contains articles on applications of BioXAS to macromolecules and low-molecular-mass complexes for pharmaceutical applications; as such it presents a showcase for the wider biological community. For present purposes, and throughout this issue, a (protein-bound) metal site consists of one or more metal ions and all protein side chains and exogenous bridging and terminal ligands that define the metal ion environment. Such sites can be classified into five main basic types

with the following functions: structural, storage, electron transfer, anion transport binding and catalysis (Holm *et al.*, 1996). 'Biological' metals which occur in the five types of sites above include magnesium, calcium, all members of the first transition series (excluding scandium, titanium and chromium) and molybdenum, tungsten, cadmium and mercury. These metals and their ligands constitute prosthetic groups that are usually covalently bound by endogenous ligands provided by amino acid side chains. Protein structure and environment modulate properties such as electronic structure, redox potential, detailed stereochemistry and ligand chemistry in order to achieve a specific function. About 30% of proteins coded by genomes are metalloproteins. An estimated 30–50% of all enzymes carry a protein-bound metal centre located mostly at the catalytic site. Approximately one-third of all proteins and enzymes purified to apparent homogeneity require metal ions as co-factors for biological function.

XAS can provide information on the electronic structure of the metal atom (oxidation state, orbital occupancies) for which protein crystallography, even at atomic resolution, is unable to shed light directly. The protein sample can be a crystal, a powder, a slurry or a solution; insoluble proteins (*e.g.* membrane proteins) are not excluded. This flexibility is of particular significance in structure–function studies of biological systems as many of the functional states can be probed without being constrained by the requirement to obtain a diffraction-quality crystal of the protein in different states. As a local method focused on a small portion of the protein structure, a limitation of the technique in some sense, the strict chemical and structural homogeneity of the sample required by crystallography may be somewhat relaxed. Thus, in a structural genomics programme, a significant proportion of the proteins which may not yield to homogenous preparations can be investigated by BioXAS. Of course, the metalloproteins which, despite their homogeneity and solubility, may not yield to crystallization can again be probed.

The experimental requirements for BioXAS are among the most stringent at the synchrotron radiation facilities. This extreme demand arises from the fact that the ratio between the number of investigated metal atoms and other atoms (atoms of the protein and the solvent) is extremely low, in the range  $10^{-4}$ – $10^{-5}$ . This has catalysed the development of sophisticated instrumentation from highly stable high-resolution monochromators to compact energy-resolving fluorescence detectors. The BioXAS data are obtained in the fluorescence mode as a rule rather than in the absorption mode as for many materials science samples. The desirable signal-to-noise ratio in fluorescence spectra is of the order of  $10^2$  for XANES and  $10^3$ – $10^4$  for EXAFS. The high dilution of the absorbing species has deleterious effects on the signal-to-noise ratio, and radiation damage has to be carefully taken into account (Ascone *et al.*, 2003). Data analysis is complicated by both complexity and disorder of these structures as well as the high demand placed by the biological question at hand which often necessitates quantification of subtle structural changes upon substrate/ligand binding or redox change. Hence, the method is permanently pushed to its limits, catalysing advances in instrumentation as well as in analysis and refinement methods. Thus, for example, the application of XAS to proteins has driven the development of multi-element energy-dispersive detection systems based on Ge/Si detector (Cramer *et al.*, 1988). The importance of multiple-scattering pathways in biological systems was realised early on (Perutz *et al.*, 1982; Strange *et al.*, 1987). Following theoretical advances, modern programs include full multiple-scattering calculations. The current state is reported in theoretical papers of this issue (Natoli *et al.*, 2003; Rehr & Ankudinov, 2003; Di Cicco, 2003; Benfatto *et al.*, 2003; Joly, 2003). The theoretical developments have

been complemented by advances in analysis methods. For example, the method of constrained and restrained refinement implemented some ten years ago (Binsted *et al.*, 1992) has become standard practice. The concept has been further extended where three-dimensional structural information of a crystallographic structure is used as a starting model and the stereochemistry is retained throughout the refinement (see *e.g.* Cheung *et al.*, 2001).

Current XAS data analysis is based on the comparison of theoretical calculations with accurate experimental data. Several approximations are generally used in theoretical calculations:

- (i) The dipole approximation for the absorption matrix element.
- (ii) The one-electron approximation, for which the many-body excitation process is described by the transition of an optical electron from the core orbital into a continuum state. Many-body effects are accounted for by the imaginary part of the one-electron complex effective potential through the Hedin–Lundquist approximation. Other many-body effects associated with excitations of two or three electrons are presently introduced *via* empirical corrections.
- (iii) The muffin-tin approximation for potentials, where the photoelectron interacts with spherically symmetric non-overlapping potentials centred on each neighbouring atom, embedded in a constant potential region. This approximation is valid at about 20 eV beyond the absorption edge; corrections are required for quantitative calculations of the near-edge structure.

Despite the advances in data-analysis programs, further effort is required for new developments (Natoli *et al.*, 2003). This implementation will benefit XANES and also EXAFS analysis since, in metalloproteins, the EXAFS signal is concentrated at low- $k$  values of the spectrum.

The quantitative analysis of the full XANES spectrum, including the absorption-edge region, has encountered difficulties for quite some time. Only recently (Della Longa *et al.*, 2001) has the first fitting of the XANES region using full multiple-scattering analysis been presented (Benfatto *et al.*, 2003).

This approach is promising for BioXAS application since the XANES signal is almost two orders of magnitude more intense than the EXAFS signal, making it possible to consequently reduce the protein concentration. The problem of the time-consuming algorithms for calculating the absorption cross section in the framework of the full multiple-scattering approach can be solved through an automated XANES analysis (Rehr & Ankudinov, 2003). A rigorous validation criteria for XAS structure determination is still much needed. In PX, the community has worked hard to come up with a comprehensive list of criteria, *e.g.*  $R_{\text{free}}$  as well as overall quality of structure *via* the Ramachandran plot and  $G$ -factor.

## 2. Perspectives

BioXAS has emerged slowly and its potential is far from being fully realised. There are several reasons for this situation. Data collection cannot be performed with a laboratory source, and is totally dependent on access to an advanced synchrotron radiation facility. Indeed, BioXAS has developed in parallel with synchrotron radiation facilities, as have the first theoretical tools, almost two decades later than crystallography. The theory of XAS is extremely complex and is still in development, involving the theory of electronic states in solids. At the present stage, it is generally not possible to make a structural analysis from an XAS spectrum without substantial prior information. The situation differs radically from the elegant simplicity of crystallography, a textbook application of the Fourier transform theory of periodic objects, where the solution of the phase problem now has a rather systematic approach with the advent of synchrotron-

radiation-based MAD and SAD methods. BioXAS has been shadowed by crystallography, which has the fascinating capability to give a straightforward, global and reliable picture of the three-dimensional architecture of the macromolecule. This situation is not specific to XAS: for many years, structural biology using synchrotron radiation has sometimes almost reduced to macromolecular crystallography. The situation is now different. It becomes clear that the combination of several techniques with complementary possibilities is required to tackle more complicated macromolecules and complex assemblies, closer to physiological conditions and where functional, and not only structural, aspects have a growing importance (see, for example, Hasnain & Hodgson, 1999; Hasnain & Strange, 2003). The challenge of high-throughput studies in post-genome programs also contributes to the reconsideration of traditional approaches. Accordingly, PX and NMR tend to diversify partnerships with other structural techniques, including neutron and X-ray solution scattering, cryoelectron microscopy, near-field microscopy, circular dichroism and mass spectrometry. BioXAS has made such methodological and theoretical progress that it is potentially able to bring its unique contribution to the comprehensive tool box now available to structural biologists.

In structural genomics programs, the XAS study of metalloproteins will require systematic procedures. The first step is the identification of potential metal-coordinating sites in the genetic information through bioinformatic analysis, based on databases of amino acids and ligands associated with metallic sites and biomimetic complexes. These databases will have to be completed and purposely reorganized. Another issue is the expression of a potential metalloprotein with incorporation of the biologically relevant metal, which is an important post-transcriptional event, followed by biochemical tests of the incorporation of metal and ligands. It is now known that many of the metalloproteins engage specific chaperons for metal incorporation. Thus, among copper proteins, superoxide dismutase and cytochrome *c* oxidase each have their own metal chaperon protein. At this stage, the nature and extent of existing information and insights obtained by model chemistry formulation of structural models will often be sufficient to build one or a few models as starting points for refinement using XAS data.

The real question is how to express more fully this potential. As discussed above, BioXAS is not limited primarily by experimental considerations, provided that access to state-of-the-art beamlines with appropriate beam time is possible. Presently, these conditions are clearly not satisfied, especially in Europe, and a significant involvement of BioXAS in structural genomics programmes will require investments in new dedicated equipment at third-generation synchrotron radiation facilities. This message is shared by most contributors to this volume. The incorporation of BioXAS capability in the new NWSGC PX beamline at Daresbury Laboratory on a multipole wiggler and in the Paris-Sud structural genomics programme are steps in this direction. Another question is the size of the scientific community. BioXAS has been pushed by few groups, and the community remains under-critical in size at the level of a single country, except perhaps in the USA. In this context, the determination to reach a better coordination of efforts and strengthening of the support and demand from users has emerged in the last few years. The vigorous development of structural genomics studies is an incentive to have a deeper reflection on a collaborative BioXAS effort and we hope that this issue will serve its due role in this attempt.

## References

Ascone, I., Meyer, W. & Murphy, L. (2003). *J. Synchrotron Rad.* **10**, 16–22.

- Benfatto, M., Della Longa, S. & Natoli, C. R. (2003). *J. Synchrotron Rad.* **10**, 51–57.
- Binsted, N., Strange, R. W. & Hasnain, S. S. (1992). *Biochemistry*, **31**, 12117–12125.
- Cheung, K.-C., Strange, R. W. & Hasnain, S. S. (2001). *Acta Cryst. D* **56**, 697–704.
- Cramer, S. P., Tench, O., Yocum, M. & George, G. N. (1988). *Nucl. Instrum. Methods*, **A266**, 586–591.
- Della Longa, S., Arcovito, A., Girasole, M., Hazemann, J. L. & Benfatto, M. (2001). *Phys. Rev. Lett.* **87**, 155501.
- Di Cicco, A. (2003). *J. Synchrotron Rad.* **10**, 46–50.
- Hasnain, S. S. & Hodgson, K. O. (1999). *J. Synchrotron Rad.* **6**, 852–864.
- Hasnain, S. S. & Strange, R. W. (2003). *J. Synchrotron Rad.* **10**, 9–15.
- Holm, R. H., Kennepohl, P. & Solomon, I. (1996). *Chem. Rev.* **96**, 2239–2314.
- Joly, Y. (2003). *J. Synchrotron Rad.* **10**, 58–63.
- Natoli, C. R., Benfatto, M., Della Longa, S. & Hatada, K. (2003). *J. Synchrotron Rad.* **10**, 26–42.
- Perutz, M. F., Hasnain, S. S., Duke, P. J., Sessler, J. L. & Hahn, J. E. (1982). *Nature (London)*, **295**, 535–538.
- Quevillon-Cheruel, S., Collinet, B., Zhou, C.-Z., Minard, P., Blondeau, K., Henkes, G., Aufrère, R., Coutant, J., Guittet, E., Lewit-Bentley, A., Leulliot, N., Ascone, I., Sorel, I., Savarin, P., de La Sierra Gallay, I. L., de la Torre, F., Poupon, A., Fourme, R., Janin, J. & van Tilbeurgh, H. (2003). *J. Synchrotron Rad.* **10**, 4–8.
- Rehr, J. J. & Ankudinov, A. L. (2003). *J. Synchrotron Rad.* **10**, 43–45.
- Strange, R. W., Blackburn, N. J., Knowles, P. F. & Hasnain, S. S. (1987). *J. Am. Chem. Soc.* **109**, 7157–7162.
- Winn, M. D. (2003). *J. Synchrotron Rad.* **10**, 23–25.