Metallogenomics and biological X-ray absorption spectroscopy

Isabella Ascone,[a]* Roger Fourme,[b] Samar Hasnain[c] and Keith Hodgson[d]

[a]Department of Physics, University 'La Sapienza', Rome P. le A. Moro 5, 00185 Roma, Italy, [b]Synchrotron-SOLEIL, BP 48, Saint Aubin, 91192 Gif sur Yvette, France, [c]North West Structural Genomics Centre, Daresbury Laboratory, Warrington WA4 4AD, UK, and [d]SSRL, Stanford University, Stanford, CA 94305, USA. E-mail: isabella.ascone@uniroma1.it

An overview of the second special issue of the journal on biological applications of X-ray absorption spectroscopy (BioXAS) is presented. The emphasis is on the study of metalloproteins in the context of structural genomics programmes (metallogenomics).

Keywords: BioXAS; EXAFS; XANES; metalloproteins; metallogenomics.

1. Introduction

A number of structural genomics programmes focused on different organisms are currently in progress, in Asia, the USA and Europe. For instance, we can mention the programme concerning tuberculosis disease, which involves the *Mycobacterium tuberculosis* (MTB) structural genomics consortium (http://www.doe-mbi.ucla.edu/TB) in the USA; RIKEN's Protein 3000 project in Japan; SPINE – structural proteomics (http://www.spineurope.org); and X-TB – structural and functional genomics of *Mycobacterium tuberculosis* (http://www.pasteur.fr/SGM) among others in Europe.

These efforts involve several structural biology teams throughout the world, and their concerted work has already targeted about half of the MTB proteome for structural characterization. It is expected that this large-scale effort on the structure determination of proteins from MTB will ultimately provide a useful framework to understand mycobacterial biology and to speed up drug discovery.

Metalloproteins will constitute a significant fraction of these proteins and their study could be performed in the frame of a metallogenomics programme. Fig. 1 highlights the most important steps of metallogenomics: (*a*) identification of metalloproteins in genomic databases, (*b*) biochemical and molecular methods to produce proteins incorporating metal ions, (*c*) metalloprotein purification and (*d*) structural methods appropriate to the study of the metalloproteins and in particular metal site(s).

Currently there are no efficient methods to identify such proteins in genomic databases. A bioinformatic approach could be used exploiting metal-binding patterns of metalloproteins present in the Protein Data Bank (Andreini *et al.*, 2004) to search the list of essential genes for new putative metalloproteins.

Post-genomics programmes are based essentially on experimental methods such as protein crystallography (PX) and nuclear magnetic resonance (NMR). Detailed structural data of the metal sites in metalloproteins, at a resolution commensurate with their chemical properties, is essential to identify and modify the metal's functional role.

Experimental methods which can provide information on metal sites are mainly electron paramagnetic resonance (EPR) and X-ray absorption spectroscopy (XAS). EPR is applicable only to paramagnetic ions and accordingly some metals have no EPR signature. XAS provides electronic and structural information on each selected atom and is a technique mature for post-genomic studies (Ascone *et al.*, 2003).

The accuracy and precision of metal centers defined using XAS are comparable with those of small-molecule crystal structures, and are superior to those of most PX structures. The majority of submissions to the Protein Data Bank are for PX structures in the resolution range 1.5–3.0 Å. For example, only 2.9% of the Cu proteins are solved at a resolution better than 1.2 Å. The number of studies at very high resolution is limited, owing to a number of reasons, among which are: (i) the limited diffraction power of most crystals; (ii) photoreduction, or radiation damage, of the protein during data collection; and (iii) micro-heterogeneity of samples.

Metallogenomics could take advantage of new experimental approaches such as BioXAS as it overcomes these restrictions, providing subatomic resolution on the local metal environment. The ability to give accurate and/or unique information on sites which are generally of high relevance for the protein function are attractive features for metallogenomics. The synergism between PX and BioXAS has recently been exploited by several studies (*e.g.* Hasnain & Hodgson, 1999; Hasnain & Strange, 2003; Pohl *et al.*, 2003).

2. Metallogenomics and BioXAS

The concept of the BioXAS study weekend originated from the European BioXAS meeting held at Orsay (France) in July 2000, in which the discussion highlighted the need for a forum where cooperative efforts could be promoted at the European
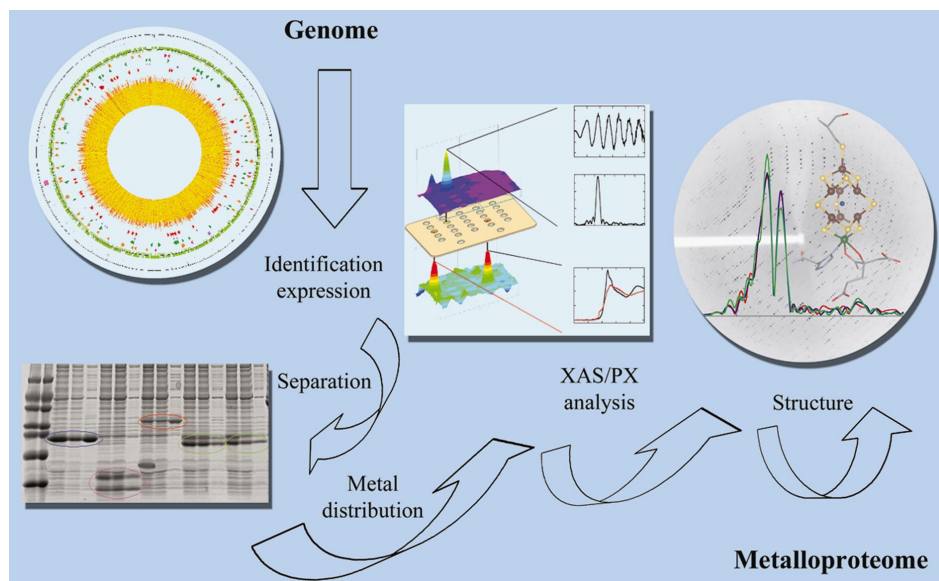
bioXAS and metallogenomics

Figure 1
The most important steps of the metallogenomics pipeline. This is a composite image including figures from this issue (Corbett *et al.*, 2005; Scott *et al.*, 2005).

level in order to develop BioXAS on all aspects covering data analysis, instrumentation, theory and applications.

The first study weekend was held at LURE in July 2001 on 'BioXAS and Structural Genomics' and was the basis of a special issue of the *Journal of Synchrotron Radiation* [(2003), Volume 10, Part 1]. The second study weekend was also held at LURE, in June 2003. New trends are discussed in the articles of this issue and some of these are highlighted here.

BioXAS developments have been identified in terms of a high-throughput (HT) XAS approach for the study of metallome (Scott *et al.*, 2005). Scott *et al.* have explored potential bottlenecks and benefits for creating a pipeline from genome to metalloproteome. The community has already progressed towards HT BioXAS studies focusing on some critical issues which are fundamental for such an approach.

Several articles have pointed out the difficulty of over-expressing metalloproteins in heterologous hosts (Hall *et al.*, 2005; Jenney *et al.*, 2005; Harris *et al.*, 2005). The presence of the correct metal and peculiar cofactors in a host organism is a critical factor in controlling protein insolubility and misfolding. This issue is a challenge for the metallogenomics community.

Considerations on data analysis and automation have progressed considerably in the last two years. Scientists have proposed new procedures to extract structural information from XANES spectra. Fitting of XANES data can be added to EXAFS fitting or can even, in part, replace them for much diluted metalloprotein samples which have signal statistics too low for EXAFS analysis. Following the presentation of the *MXAN* program at the first study weekend (Benfatto *et al.*, 2003), the present issue shows different examples of successful XANES fitting (Roscioni *et al.*, 2005; Wu *et al.*, 2005). In the *MXAN* program, the optimization is achieved by minimization of the square residual function in parameter space. J. Rehr

has more recently proposed the Bayes–Turchin approach instead of conventional least-squares fitting methods. This method, which takes advantage of *a priori* estimates of model parameters, thus improving the significance of fits, has the potential to provide an automated XAS analysis tool (Rehr *et al.*, 2005). Other methods for automation of BioXAS data treatment (Bunker *et al.*, 2005) as for automatic procedures in data quality control have been discussed. The last item is crucial for BioXAS experiments where signals from several (13 to 100) independent fluorescence detectors are averaged. The identification and elimination of contributions suffering from artificial deviations can be automatically removed from the data pool (Lippold *et al.*, 2005).

Finally, the successful combination of PX and BioXAS is illustrated by an interesting example (Corbett *et al.*, 2005). Local metal structure of the iron-molybdenum cofactor of nitrogenase in isolated MoFe protein has been determined by XAS. XAS results from crystalline and liquid samples have been correlated to PX investigations completing the structural determination with electronic information. This was possible by recording and analyzing *K*- and *L*-edge XANES spectra.

This complementarity also involves experimental set-ups as shown by recent and new devices integrating PX and BioXAS techniques in the same beamline. We can mention the ten-pole wiggler beamline (MPW10) at the CCLRC Daresbury Laboratory (recently in operation, www.nwsgc.ac.uk) (Cianci *et al.*, 2005), the wiggler beamline 9-3 at SSRL (Latimer *et al.*, 2005). At SOLEIL, a new facility under construction, a specific programme on BioXAS has been approved, with access on PROXIMA I, an undulator PX/BioXAS beamline.

3. Perspectives

Ideas discussed at the first meeting on the contribution of BioXAS to structural genomics have progressed significantly. The involvement of synchrotron radiation laboratories in BioXAS developments has increased. The Stanford Synchrotron Radiation Laboratory and CCLRC Daresbury Laboratory contributed to the organization of the second study weekend.

The community contributing to the second study weekend is wider; we can mention the participation of scientists from the USA (which had a strong delegation), Japan and China, thus setting the basis for a world-wide cooperation on metalloproteome. The coordination of efforts for BioXAS developments in the framework of metallogenomics has been useful

and it would be timely to give stronger support to cooperation through further initiatives.

References

Andreini, C., Bertini, I. & Rosato, A. (2004). *Bioinformatics*, **20**, 1373-1380.

Ascone, I., Fourme, R. & Hasnain, S. S. (2003). *J. Synchrotron Rad.* **10**, 1–3.

Benfatto, M., Della Longa, S. & Natoli, C. R. (2003). *J. Synchrotron Rad.* **10**, 51–57.

Bunker, G., Dimakis, N. & Khelashvili, G. (2005). *J. Synchrotron Rad.* **12**, 53–56.

Cianci, M., Antonyuk, S., Bliss, N., Bailey, M. W., Buffey, S. G., Cheung, K. C., Clarke, J. A., Derbyshire, G. E., Ellis, M. J., Enderby, M. J., Grant, A. F., Holbourn, M. P., Laundy, D., Nave, C., Ryder, R., Stephenson, P., Helliwell, J. R. & Hasnain, S. S. (2005). *J. Synchrotron Rad.* Submitted.

Corbett, M. C., Tezcan, F. A., Einsle, O., Walton, M. Y., Rees, D. C., Latimer, M. J., Hedman, B. & Hodgson, K. O. (2005). *J. Synchrotron Rad.* **12**, 28–34.

Hall, J. F., Ellis, M. J., Kigawa, T., Yabuki, T., Matsuda, T., Seki, E., Hasnain, S. S. & Yokoyama, S. (2005). *J. Synchrotron Rad.* **12**, 4–7.

Harris, R. L., Prudêncio, M., Hasnain, S. S., Eady, R. R. & Sawers, R. G. (2005). *J. Synchrotron Rad.* **12**, 13–18.

Hasnain, S. S. & Hodgson, K. (1999). *J. Synchrotron Rad.* **6**, 852–864.

Hasnain, S. S. & Strange, R. (2003). *J. Synchrotron Rad.* **10**, 9–15.

Jenney, F. E. Jr, Brereton, P. S., Izumi, M., Poole II, F. L., Shah, C., Sugar, F. J., Lee, H.-S. & Adams, M. W. W. (2005). *J. Synchrotron Rad.* **12**, 8–12.

Latimer, M. J., Ito, K., McPhillips, S. E. & Hedman, B. (2005). *J. Synchrotron Rad.* **12**, 23–27.

Lippold, B., Meyer-Klaucke, W., Meyer, T. & Henkel, G. (2005). *J. Synchrotron Rad.* **12**, 45–52.

Pohl, E., Haller, J. C., Mijovilovich, A., Meyer-Klaucke, W., Garman, E. & Vasil, M. L. (2003). *Mol. Microbiol.* pp. 903–915.

Rehr, J. J., Kozdon, J., Krappe, H. J. & Rossner, H. H. (2005). *J. Synchrotron Rad.* **12**, 70–74.

Roscioni, O. M., D'Angelo, P., Chillemi, G., Della Longa, S. & Benfatto, M. (2005). *J. Synchrotron Rad.* **12**, 75–79.

Scott, R. A., Shokes, J. E., Cosper, N. J., Jenney, F. E. & Adams, M. W. W. (2005). *J. Synchrotron Rad.* **12**, 19–22.

Wu, Z. Y., Tao, Y., Benfatto, M., Xian, D. C. & Jiang, J. Z. (2005). *J. Synchrotron Rad.* **12**, 98–101.