MAD Data Collection for Structural Genomics: Studies of Proteins from *Haemophilus influenzae* at the IMCA-CAT Beamlines

K. J. Kim, G. J. Sahli, A. J. Howard
Illinois Institute of Technology, Chicago, IL, U.S.A.

Many of the gene products of completely sequenced organisms are hypothetical and so are of completely unknown function. Structural studies provide one means of obtaining functional information in these cases. A structural genomics project has been initiated aimed at determining the structures of about 65 hypothetical proteins from *Haemophilus influenzae* (HI) to gain an understanding of their function.[1] The use of third-generation synchrotron x-ray sources is playing a pivotal role in the HI genomics project.

The genes encoding candidates for structural determination from *Haemophilus influenzae* were cloned, expressed, and purified by scientists at the Center for the Advanced Research in Biotechnology (CARB) and National Institute of Standards and Technology (NIST). The native and heavy-atom derivative crystals grown by CARB scientists were transported to the IMCA-CAT beamline at the APS and mounted on the 17-ID and 17-BM beamlines. Multiple-wavelength anomalous diffraction (MAD) data sets were collected and processed with HKL2000 and X-GEN by IIT and CARB scientists. Raw and processed data were transported back to CARB, and structure determination was performed by CARB scientists with close discussion with IIT scientists at the APS.

For each multiwavelength experiment, a fluorescence scan was performed at the appropriate energy range for the heavy atom used. Three or four wavelengths were chosen from the f′ and f′′ plots based on the Kromers-Liberman transformations, and MAD data were collected at the chosen wavelengths. In general, data were collected at the inflection point of the f′′ spectrum, the peak of the f′′ spectrum, a wavelength about 60 eV below the inflection point, and sometimes a wavelength about 200 eV above it. Generally we attempted to achieve at least fivefold redundancy of the data and at least 94% coverage of the unique data. Typically the usefulness of the multiwavelength data was monitored during data collection by confirming that the anomalous R value ($=\Sigma \mid F^{+} - F - \mid / \Sigma <F>$) as larger for the peak data than for the low-remote data. CARB scientists perform the actual MAD phasing as the first phase of their structure-determination efforts.

Acknowledgments

Data were collected at beamline 17-ID in the facilities of the Industrial Macromolecular Crystallography Association Collaborative Access Team (IMCA-CAT) at the Advanced Photon Source. These facilities are supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Illinois Institute of Technology (IIT), executed through IIT's Center for Synchrotron Radiation Research and Instrumentation. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-ENG-38. The structural genomics project is supported by the National Institutes of Health through a Program Project Grant to investigators at CARB, IIT, and the Institute for Genomic Research (R. Poljak, overall PI). We thank these participants from CARB for expressing, purifying, and crystallizing the proteins and for determining the structures afterward: N. Bonander, S. Chu, G.L. Gilliland, O. Herzberg, K. Huang, J.Ladner, C. Lehmann, K. Lim, A. Teplyakov, M. Willis, and H. Zhang.

References

[1] E. Eisenstein, G.L. Gilliland, O. Herzberg, J. Moult, J. Orban, R.J. Poljak, L. Banerjei, D. Richardson, and A.J. Howard, Curr. Opin. Biotechnol. **11**, 25-30 (2000).