Editorial

J. Synchrotron Rad. (2000). **7**, 355

Phase II of the genomic effort

S. Samar Hasnain,[a]* Hiromichi Kamitsubo[b] and Dennis M. Mills[c]

[a]CCLRC Daresbury Laboratory, Warrington WA4 4AD, UK, [b]SPring-8, JASRI, Sayo-gun, Hyogo-ken 679-5198, Japan, and [c]Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA

Earlier this year the Human Genome Project participants announced that they had completed a first draft of the human genome sequence. The goal of that project was to determine the sequence of the ~3 billion bases that make up the human genome. If we consider sequencing as a Phase I of the worldwide genomic effort, then the foundations are clearly being laid for the Phase II gemonics activity, namely determining the structures of the proteins coded for by genes. Knowledge of protein structures at or near atomic resolution, which can only be achieved through X-ray crystallography and best achieved with synchrotron radiation, provides insight into that protein's biological function. Thus, structural genomics is seen as a pre-requisite to functional genomics even though significant progress can be made through advanced structure prediction methods being currently referred to as proteomics. Detailed structural knowledge can lead to a better understanding of protein's role in various biological and biochemical processes, in particular its primary specificity. Needless to say, the incentives, both intellectual and financial, for understanding the structure of proteins are high.

This structural genomics effort is a bold undertaking and will certainly stress the current capabilities of macromolecular crystallography beamlines worldwide, considering that an estimated 5000 to 10000 unique families or fold topologies will have to be solved over the next several years from the more than 100000 human genes. (A fold is a recurring structural pattern or motif.) Examination of the 13000-plus structures deposited in the Protein Data Bank reveals that to date less than 1000, and perhaps only a few hundred, unique folds have been included. To achieve the goal of determining 10000 unique fold structures over the next several years will require new approaches and styles in crystallographic studies. Workshops have already been organized by the crystallography community to discuss the prospects of this daunting task, and new ideas have been brought to the table for new and existing crystallography beamlines. No longer will each tiny crystal be mounted and oriented in the X-ray beam by hand, but robotic techniques will necessarily be used to speed up the task of data collection. Similar robotic techniques may be required for crystal growth as well in order to keep the supply of crystals ready to be processed through this structural assembly line. In addition, high throughput is likely to be achieved only if large numbers of proteins can be expressed and purified in an automatic manner. Improved data-handling capabilities will also need to be implemented to manage the several gigabytes of data that are typically generated in the solution of the structure of a single protein. On-line data processing and structure determination will need to become common place. To accomplish the increased throughput will require close interaction between biologists, engineers, physicists and computer scientists. Several pilot programmes have been launched across the world aimed towards high-throughput structure determination. Germany, Japan and the USA have taken the lead, and recently a structure genomics programme has been proposed at Daresbury with a multipole-wiggler MAD beamline and off-line protein production and crystallization factories. The first major commercial inititiative occurred recently at the Advanced Photon Source where an agreement was reached between a start-up company, Structural GenomiX, and the APS for the construction of a beamline dedicated to structural determination of genomes.

One of the intriguing aspects of this undertaking is that all of the above-mentioned approaches, *i.e.* robotic manipulation of samples, automated sample-alignment procedures and high data streams, will be directly applicable to many other scientific disciplines that use X-ray synchrotron radiation. Smart Instrumentation on synchrotron radiation beamlines will become common place. One can imagine straightforward applications of these approaches to studies such as topography, spectroscopy, microprobe techniques and parallel analysis of combinatorial chemistry/materials science samples. It is fair to say that, to date, much of the synchrotron radiation instrumentation development has been initiated by researchers in the physical sciences and adapted by the life scientists to perform their work. The challenge of structural genomics may reverse that flow of instrumentation expertise as the structural biology beamlines require innovative approaches and new tools if serious contributions to the structural genomics game are to be made in a reasonable time frame.

One of the founding editors (Professor John Helliwell) has retired from the *Journal of Synchrotron Radiation* and has been replaced by Dennis Mills. John will remain involved with the journal through his position as the Editor-in-Chief of IUCr journals.