# Structural genomics in the spotlight

Structural genomics efforts have spurred the continuing development of new methods and technologies, benefiting a broad community.

One of the many areas of research that got a kick start from genome projects is the field known as structural genomics, the high-throughput endeavor of solving three-dimensional protein structures. Currently, about half of the novel structures being deposited to the Protein Data Bank were solved by structural genomics consortia. By most measures, the largest project is the Protein Structure Initiative (PSI) in the US, funded by the National Institutes of Health (see Commentary, p. 129 and Technology Feature, p. 203); there are also large consortia in Japan, Canada and Europe.

Not only does structural genomics differ in implementation and scale from traditional structural biology, but the overall goals differ as well. Structural genomics aims to develop a resource of representative protein fold structures to extrapolate (in theory) any protein structure by homology modeling. In contrast, hypothesis-driven structural biology seeks to understand biological function, often by solving protein structures. Challenging targets, such as large protein complexes, still require these concentrated, traditional efforts to solve. Thus the two fields are complementary.

Few would dispute the inherent value of a large repository of protein structures for basic biological research, not to mention for drug discovery. However, as with any large-scale effort, structural genomics has detractors who question whether the enormous monetary investment (the PSI budget alone is $60 million per year) in such projects will ever yield dividends in biological knowledge. Several prominent structural biologists doubt that homology modeling will be able to provide accurate structures and tell us something about the biology, which is often in the fine details.

Another frequent criticism is that structural genomics consortia have gone after the 'low-hanging fruit'—soluble, bacterial proteins that are relatively easy to express and purify—and that these targets are not biologically very interesting. The PSI in particular has responded to this concern by having its four large-scale production centers focus 15% of their effort on studying proteins of biomedical importance (of the investigator's choosing) and 15% of their effort on addressing community-nominated targets.

An undeniably beneficial result, however, is that structural genomics has triggered an abundance of technological developments, making the process of structure determination less of an art and more of a science. Through the development of new methods, automation, miniaturization and new software, the average cost, not to mention elapsed time, to obtain a structure has dropped by more than half. And quality has not fallen: a recent study showed that structures solved via structural genomics are on average of somewhat higher quality than those solved by traditional methods (*Acta Crystallogr.* **D63**, 941–950; 2007).

Considerable efforts from the PSI and other consortia are being focused on developing new methods to solve more challenging targets, such as membrane proteins and large protein complexes. These consortia are also actively addressing the methodological bottlenecks in the process to salvage the approximately 85% of targets that fall out of the pipeline. In addition, through the high-throughput generation of data, systematic evaluation of methodological efficacy is possible, allowing researchers to develop general solutions for protein expression and purification (see Review, p. 135), or crystallization (see Perspective, p. 147).

The new methods and technologies are not just for high-throughput structure-determination pipelines; they have also aided traditional biology. For example, the accelerated process from construct to three-dimensional structure allows researchers to rapidly investigate the biology of proteins of immediate interest for public health, such as those of the SARS coronavirus.

Consortia are making concerted efforts to give back to the broader community by rapidly depositing structures in the Protein Data Bank (often well before any paper is published), maintaining other databases, and making all methods and results freely available. Launching soon is the PSI Structural Genomics Knowledgebase, a centralized, user-friendly portal to all of the information collected by PSI efforts. The PSI is also developing a collection of 50,000 plasmid clones in the PSI Materials Repository, which will be available for a minimal fee.

In the future, we can expect to see the impact of structural genomics on other areas of research, such as in aiding the development of affinity reagents. It is perhaps too soon to tell what the biological impact of structural genomics will be—after all, obtaining results in biology takes some time—but the development of new methods and technologies will certainly be an enduring result.