

## Looking ahead with structural genomics

**Structural genomics initiatives aim to create a library of all existing protein folds. We take a look at the progress that has been made and what more needs to be done.**

The completion of multiple genome sequencing projects has driven high-throughput analysis on many levels. This includes structural genomics, genome-wide drives to catalog representative protein structures, or fold families. The proposed benefit of creating a fold library — that is, having at hand a representative structure from each of the major fold families — is that one could then model structures of homologous proteins that lack a three-dimensional structure, which might pave the way for more informed functional studies on proteins of interest.

Current structural genomics efforts can be roughly lumped into three groups: the Japan-based program, led by RIKEN, called Protein 3000; the Protein Structure Initiative (PSI), developed by the US NIH/NIGMS; and the efforts from the European research community, with programs such as SPINE and SGC, which are overseen in part by the European Commission. Over the past months, these three efforts have faced assessment by participating researchers and the funding agencies that help support them.

Protein 3000 began several years ago with the goal of hitting as much of the ‘fold space’ as possible, seeking to determine 3,000 structures, roughly one-third of the bioinformatically predicted 10,000 unique protein folds. This was facilitated by methodology improvements but the success of these endeavors has been slightly overshadowed with more recent predictions that the number of protein folds is actually triple that of previous estimates. The project is set to end in early 2007, and budget requests suggest a post-Protein 3000 shift of focus from high-throughput, factory-like structure determination to more traditional structural examination with complementary functional studies, as well as more attention to disease-related proteins of interest.

The PSI started seven years ago, with the long-term goal of making three-dimensional structures easily available after DNA sequence determination. The pilot phase aimed to streamline structure determination methods. Now, in its second year of phase 2, the emphasis has shifted to high-throughput structure determination, with a strong focus on improved bioinformatics-guided target selection. These efforts are already projected to outstrip the number of unique structures determined in the pilot phase.

The European consortia share similar goals to Protein 3000 and PSI—high-throughput methods to sample maximal fold-space—which are complicated by the intricacies that accompany international programs requiring funding from individual member countries. A recent workshop sponsored by the European Commission (see the Meeting Report on page 3 of this issue) made clear some of the scientific issues this funding agency would like structural genomics to address.

The most prominent criticism of structural genomics has been directed at the PSI, particularly regarding spending on structural genomics in the US, which has been hard to stomach when basic science programs

are suffering. Those in favor of structural genomics will argue that PSI represents only ~1%–3% of the total NIH operating budget. However, there are also more fundamental concerns shared by many prominent researchers in the structural biology community about the very concept of structural genomics. Stephen Harrison (Harvard Medical School and Howard Hughes Medical Institute) feels that the idea of structural genomics was flawed from the start. “The original justification for it was to enumerate all the different categories of protein folds. But this in itself was an inappropriate goal.” Homology modeling “can be done at some level, but that has already been shown to not be useful. Evolution plays with details. At the moment, you can’t even model the variable loops of one IgG molecule from another that shares a similar fold.” Andrzej Joachimiak, who heads the Midwest Structural Genomics Consortium, one of several research centers that are part of the PSI, argues that modeling efforts are being improved by new programs, including those coming out of the two new homology centers that are part of the PSI. “Traditional structural biology is just too slow,” Joachimiak says. “It took a long time to actually determine protein structures. Today, because of structural genomics efforts, structures can be determined much faster. What structural genomics brings is the application of high throughput methods, which can not only be applied to structural genomics efforts, but to regular structural biology projects.” Nevertheless, Harrison points to the fact that “most crystallographic and NMR efforts have required careful engineering of the protein or protein complex that yields a structure” and are therefore not suitable for high-throughput methods. He suggests that more energy be directed toward creating better expression ‘tricks’ for construct variation and improving ways to collect data from very small protein crystals.

Regardless of where one may stand on these issues, it is difficult to deny that structural genomics efforts have made contributions for the benefit of all, particularly in the area of methods development, with breakthroughs in cell-free expression systems and in expression of recombinant membrane proteins, and in crystallization robots, to name a few.

What is most striking about all of the structural genomics efforts is that they seem to go largely unnoticed by the majority of the research communities outside of structural biology. Although this is nothing new—structural biologists have long experienced a disconnect with many geneticists and molecular biologists, for example—it raises the question of exactly how useful structural genomics will be to the scientific community as a whole. Those familiar with protein structures are aware of the valuable contribution a three-dimensional structure can make to experimental design. A major goal must be to get the data into the hands of the larger research community, integrate them with functional and interaction analyses, and make non-structural biologists comfortable with looking at protein structures. Only then will we be able to truly judge the utility of structural genomics efforts. ■