

STRUCTURAL GENOMICS

Clear vision for a structure-seeking business

Tim Harris

I read with interest the editorial on structural genomics in the September issue of *Nature Biotechnology* and decided, as CEO of Structural GenomiX (San Diego, CA)—one such “smart, agile, and fast” startup in this area—that it was worthwhile setting out my views on the business of structure determination.

X-ray structure resolution is undergoing the same revolution that DNA sequencing underwent several years ago. The linearity of the previous approach to structure determination is being replaced by massively parallel systems for cloning, expression, protein purification, and protein crystallization. Dedicated X-ray diffraction beam lines suitable for very high-throughput multiwavelength anomalous diffraction (MAD) phasing experiments, such as the beam line Structural GenomiX is building at the Advanced Photon Source of the US Department of Energy’s Argonne National Laboratory (Chicago, IL), are being constructed, or existing beam lines are being adapted toward that end.

To date, little structural information has been available as an aid to drug discovery, but it is quite clear that the solution of many more structures will enfranchise the industry to use structural approaches in drug and compound discovery. It is worth recalling that obtaining the structure of influenza virus neuraminidase in Australia initiated the whole drug discovery effort at Glaxo (London) that culminated in Relenza, a potent inhibitor of that enzyme and successful drug for treating influenza. There are now many examples where structural information has not only helped to identify the function of a protein encoding an unknown open reading frame (ORF), but also validated a potential target. About 40% of the proteins predicted from genome sequences have no known function, and in bacterial genomes around 25% of the unknown ORFs are unique to any one genome¹. It is going to be very hard to identify the biochemical functions of these proteins by using either existing correlative and associative methods (such as expression arrays) or deletion studies. Determining three-dimensional structures will be an

increasingly useful way of implying biochemical function. There is no question that from both a cost and time point of view, structure determination will rapidly become competitive with other functional genomics and proteomics methods for implying function.

A successful structural genomics business can be developed by determining structures at very high throughput and adding considerable value to those structures by relevant chemical and biochemical annotation. Potential customers include not only the pharmaceutical and biotechnology companies, but also agricultural and industrial chemical companies. There is opportunity for strategic alliances both in different therapeutic areas and/or on certain protein families. A database of proprietary structures can also be created that will incorporate and add value to those structures in the public domain and include extensive biochemical and chemical annotation. Well-funded companies can also use structures to facilitate the design of small, but highly constrained, combinatorial libraries, which may also be accessible in database format.

The argument concerning commoditization of protein structures has been made before for expressed sequence tags (ESTs), cDNAs, genome sequences, and single nucleotide polymorphisms (SNPs). There is no sign that any of these have become commodities. Companies are continuing to sell genome sequence data to many customers, and it is as true today as it was six years ago that full-length cDNAs (which are hard to get for rare transcripts and cannot be obtained from genomic sequence) are valuable entities. I see no reason why commoditization of protein structures should happen any more quickly than it has for DNA sequences.

The public domain initiatives both in the US and in Europe will clearly add to the overall numbers of structures that are available. The whole process of structure determination is, however, inherently much more complex than DNA sequencing. Because such a large financial investment is required to produce structures at high

throughput, I believe that a concentrated, rather than a distributed, effort will succeed best.

There is also a move to form an industry consortium to increase the number of structures available for drug discovery. This appears to be modeled on the SNP consortium. However, structures are templates for drug discovery; SNPs are not. It seems unlikely that companies will want to share the structure of their validated targets in a consortium.

Intellectual property (IP) issues must also be worked through. In my view, the IP situation is quite clear: protein structures have been patented before and the claims therein substantiated. Thus, it will be possible to take out IP protection on those structures considered of high value for drug discovery, particularly when in complex with novel compounds. These structures can then be made available to paying customers.

Target selection is also an area of strategic importance. There are many protein families of considerable interest to the industry that have little or no protein structure information available. As stated in the editorial, membrane-bound proteins are particularly challenging, but there has been recent progress. For example, the structure of bovine rhodopsin, a member of the G-protein coupled receptor (GPCR) family, has recently been determined², and similar resolution structures may soon be available for other GPCRs. The genomics approach—that is, expressing many members of a family of proteins in multiple hosts—will lead to the determination of the structure of many more cytoplasmic and membrane-bound proteins.

In my opinion, the real key to structural genomics will be to use a focused and fast high-throughput technology platform to provide structures to the industry at an unprecedented rate and with considerable added value.

It seems unlikely that companies will want to share the structure of their validated targets in a consortium.

Tim Harris is president and CEO of Structural GenomiX, 10505 Roselle Street, San Diego, CA 92121 (tim@stromix.com).

1. Fraser, C.M. et al. *Nature* **406**, 799–803 (2000).
2. Palczewski K. et al. *Science* **289**, 739–745 (2000).