Structural genomics: shaping the future of drug design?

Stephen Anderson* and **John Chiplin**, GeneFormatics, 5830 Oberlin Drive, Suite 200, San Diego, CA 92121, USA, tel: +1 858 882 5900, fax: +1 858 450 1138, *steveanderson@geneformatics.com

The fourth annual *Structural Genomics in Pharmaceutical Design* meeting (24–25 October 2001, Princeton, NJ, USA) hosted by the Center for Advanced Biotechnology and Medicine (CABM; Piscataway, NJ, USA) and sponsored by the New Jersey Commission on Science and Technology (Trenton, NJ, USA), focused on the emerging field of structural genomics (or, as some would characterize it, structural proteomics). The first in this series (4–7 October 1998, Avalon, NJ, USA) was, in essence, the coming-out party for structural genomics, being the first major international meeting of this new discipline. Hence, it was fascinating to gather recently in Princeton to hear how the field is faring three years later.

Structural genomics initiatives

As part of a pilot project to explore technology development in structural genomics and to nucleate the public sector effort in this area, the National Institute of General Medical Sciences (Bethesda, MD, USA) founded several key national Structural Genomics Centers in the USA in 2000 and 2001. Four Principal Investigators of these Centers: Stephen Burley (Rockefeller University, New York, NY, USA), Wim Hol (University of Washington, Seattle, WA, USA), Andrzej Joachimiak (Argonne National Laboratory, Argonne, IL, USA) and Gaetano Montelione (CABM and Rutgers University, Piscataway, NJ, USA) reported on the progress made by their respective centers in the past year. The scientific promise of the NIH Structural Genomics Initiative is already beginning to be fulfilled; all of the Structural Genomics Center speakers spoke of discovering novel folds and of uncovering functional information both from structural homologies to existing proteins and from the structures of adventitiously co-crystallized ligands such as cofactors and substrates.

Joachimiak summarized the impressive capabilities, in terms of throughput and resolution, of the structural biology beamlines at the Advanced Photon Source (APS) synchrotron at Argonne where the Midwest Center for Structural Genomics (of which he is Principal Investigator) is based. As many as 15–20 X-ray diffraction datasets per day can be collected on one of the APS beamlines and, under favorable circumstances, the resolution can enter the sub-atomic realm (~0.6 Å), which enables the electron density from hydrogen atoms to be determined, carbons to be distinguished from nitrogen, and so on. Even at lower resolutions, side-chain structures can be determined with such high precision that the sequence of the protein can be checked.

Similarly, Montelione, with Clemens Anklin from Bruker Biospin (Billerica, MA, USA) described the technological strides being made in NMR spectroscopy, both in software for automating data analysis and in new hardware such as cryoprobes. Such innovations have enabled Montelione's *Northeast Structural Genomics Consortium*, which is focusing on eukaryotic targets of unknown function that represent large protein families, to achieve outstanding productivity in terms of its output of high-quality NMR structures.

Burley, speaking on behalf of the *New York Structural Genomics Research Consortium*, emphasized his team's efforts to concentrate on protein domains prioritized by their biological or medical interest. In particular, he recounted an enzyme-family based approach towards analyzing the three-dimensional (3D) structures of enzymes in the mammalian cholesterol-biosynthesis pathway that are 'downstream' of 3-hydroxy-3-methylglutaryl (HMG)-CoA reductase.

Montelione, Burley and Joachimiak also reported impressive throughputs in terms of protein expression, with ~1000 proteins cloned per year per consortium. Although attrition rates were not reported, approximately 50% of the expressed proteins appear to yield soluble product (in *Escherichia coli*). Therefore, it appears that the once-feared sample preparation bottleneck in structural genomics has been avoided – momentarily at least – because the current target-rich environment enables even large attrition rates to be absorbed by simply increasing the numbers of proteins attempted.

Protein folding

In silico *studies*

In the field of structure-based *in silico* studies, Mark Gerstein (Yale University, New Haven, CT, USA) and Jeffrey Skolnick (Danforth Plant Science Center, St Louis, MO, USA) reported some intriguing results. Those involved in new target discovery and validation will be interested to learn that gene-expression levels in yeast are correlated with gene-product subcellular localization, according to Gerstein. Using Bayesian statistical methods, his laboratory predicted protein localization from expression data with

1359-6446/02/$ – see front matter PII: S1359-6446(01)02125-0

>75% accuracy. If this observation can be generalized to human cells, it promises great potential for extracting increasingly useful drug-target information from DNA microarray data.

Another novel observation discussed by Gerstein was that of 'pseudofolds', that is, the folds of virtual proteins encoded by unexpressed pseudogenes in the genome. Pseudofolds are distributed differently from the folds of expressed proteins; for example, there are folds that are only represented in the pseudogenes, and not in any active protein-coding region.

Genome-wide studies

Continuing on the theme of protein folds on a genome-wide scale, Skolnick and coworkers have found that, for most genomes, 40–50% of the encoded polypeptides can be assigned to a known fold using sufficiently powerful threading algorithms. For the remaining proteins, Skolnick and his collaborators are developing *ab initio* fold-prediction methods that are sufficiently fast to be applied to the whole genome. At present these are limited to relatively small proteins or protein fragments of ≤150 amino acids. Purely computational approaches to structure and function determination such as those outlined by Skolnick are extremely fast and cheap compared with direct X-ray and NMR experimental methods (albeit, in some cases, only approximate), and could, therefore, complement and accelerate worldwide structural genomics efforts.

Structural genomics for drug discovery

Some fascinating insights into structural genomics in drug discovery were provided by Robert Powers (Wyeth-Ayerst Research and Genetics Institute, Cambridge, MA, USA), Sean Buchanan (Structural GenomiX, San Diego, CA, USA) and Raymond Salemme (3D Pharmaceuticals, Exton, PA, USA). These speakers represented groups working at the fertile intersection of genomics, structural biology and drug discovery, and their talks provided a glimpse of what the future holds as the field of structural genomics continues to evolve and impact research strategies in the pharmaceutical industry.

NMR- and mass spectrometry (MS)-based technologies

Powers highlighted the basic challenge for structural biology in the modern drug discovery paradigm: how can X-ray and NMR add value in a timely manner when done in parallel to, and in concert with, HTS, lead identification and lead optimization? Focusing primarily on the contribution that NMR can make, he identified two areas: rapid structure determination by NMR and NMR-aided compound screening.

X-ray crystallography offers an ultra-rapid method of obtaining a 3D structure in cases where suitable crystals form (as described by Joachimiak); however, if crystals do not easily form and the protein is sufficiently small and soluble, NMR can resolve this problem. Using oncostatin M as an example, Powers outlined how NMR resonance assignments and minimal constraints from sparse nuclear Overhauser enhancements (NOEs) or residual dipolar coupling data could be used to refine homology models (yielding a structure good enough to guide drug design) and/or to tackle structure determinations of proteins that are too large to solve using NMR alone.

Powers also described a novel hybrid MS–NMR-based compound screening method that requires considerably less protein than other methods. In this scheme, mass-tagged compound mixtures are mixed with the target protein and subjected to rapid gel filtration followed by MS on the excluded volume fraction; compounds eluting in this fraction are presumed to be bound to the target and MS identifies them. Once this has been done, chemical shift perturbation or other NMR experiments can be performed on the binders to map where on the target they interact, and then this information is fed back to the medicinal chemists for further refinement of the library.

X-ray crystallography-based approaches

Buchanan summarized his company's progress towards developing a high-throughput X-ray crystallography capability as a platform for drug discovery. He made the justifiable claim that structure-based (or at least structure-informed) approaches to drug discovery reduce the number of compounds that need to be screened and thus reduce the cost and time required to find valuable new chemical entities (NCEs).

Although they have focused primarily on bacterial proteins, to date, Buchanan and his colleagues have found, as have others, that 3D protein structures often provide valuable functional clues: 3D homologies, surface properties and co-crystallized natural small-molecules. Furthermore, once ligand-binding sites have been identified, the structures of the targets can also be used to assess 'drugability'.

A functional approach: differential scanning calorimetry

Salemme's talk was a counterpoint to this structure-based approach; he described how his group screens potential targets of unknown structure and function against compound libraries using differential scanning calorimetry. Bound ligands are detected based on their enhancement of the thermodynamic stability of the protein, and Salemme stated that their experience has shown that these tend to bind to functionally important clefts on the protein (e.g. active sites, allosteric regulatory sites, and so on). By screening unknown proteins against a library of 5000 known biological ligands and identifying which ligands bind, function can sometimes be inferred.

Although this approach might not obviate the need for getting structural data for targets, Salemme commented that 'good ligand-binding data is amazingly empowering'.

Final thoughts

In summary, the sense of this meeting was that the discipline of structural genomics, especially as it applies to functional characterization and drug discovery applications, has clearly taken root but is still in its infancy. Structural biology, despite many recent technological improvements that have produced quantum leaps of increased speed and decreased cost, is still a complicated and knowledge-intensive science. Thus, the acceleration of the rate of new structure determination is expected to be slower than the explosive growth rate of sequence-based genomic data. Nevertheless, the information-rich nature of protein structures and their high relevance to drug design presages a coming golden age of structure-based drug discovery in the post-genomic world.

Antisense Drug Technology: Principles, Strategies, and Applications

Edited by Stanley T. Crooke, Marcel Dekker, 2001, Price US$225.00, 929 pages, ISBN 0-8247-0566-1

Making drugs from antisense oligonucleotides (ASOs) is a compellingly simple concept which has been difficult to put into practice. Unlike most other drugs (small molecules or biologics) that target a specific protein, ASOs prevent protein synthesis by eliminating the template mRNA. The appeal of antisense lies in its simple structure–activity relationship; knowing only the nucleotide sequence of a drug target, one can, in theory, design a selective and potent inhibitor using simple base-pairing rules. In the early 1990s (the dawn of antisense drug R&D), it was anticipated that ASOs would find a place in the pharmaceutical armamentarium, and might ultimately transform traditional small-molecule drug discovery.

Many large pharmaceutical companies cultivated alliances to develop antisense therapeutics, including: Roche (Basel, Switzerland) and Searle (Skokie, IL, USA) with Hybridon (Cambridge, MA, USA); Ciba-Geigy (Basel, Switzerland) and Boehringer Ingelheim (Ingelheim, Germany) with Isis Pharmaceuticals (Carlsbad, CA, USA); Glaxo Wellcome (now GlaxoSmithKline, Greenford, UK) with Gilead (Foster City, CA, USA); and Baxter (Deerfield, IL, USA) and Johnson & Johnson (New Brunswick, NJ, USA) with Genta (San Diego, CA, USA).

However, the initial promise of those early days has not yet been fulfilled. Only one oligonucleotide-based drug is currently marketed (Isis' Vitravene™ for CMV retinitis), and controversy remains regarding its true mechanism of action in humans (see *Antisense Wars* by Karl A. Thiel available at http://biotech.about.com/cs/antisense/). In addition, Vitravene is given by intravitreal injection, which circumvents problems with delivery, pharmacokinetics, and the safety of ASO drugs administered by more typical routes.

Although antisense has recently become a popular tool to validate targets for small-molecule drug development, several companies continue to champion the use of ASOs as drugs themselves. The clear leader in this group is Isis Pharmaceuticals, whose CEO, Stanley Crooke, has edited the latest of several recent publications reviewing the antisense field. His effort, *Antisense Drug Technology*, comes close to becoming the definitive reference for the development of ASOs as therapeutics.

What is covered?

Antisense Drug Technology is successful at capturing the key topics involved in developing oligonucleotides into drugs. Part I is an excellent review of ASO basics – mechanism of action, antisense design and synthesis, medicinal chemistry and analytical methods. In Part II, the issues of pharmacokinetics and drug safety in animals and humans, and the steps required for preclinical and clinical drug development, are covered in great detail. Reviews of antisense clinical trials are relatively scarce in the primary literature, and so the extensive data and references are a helpful resource.

Part III, with 22 chapters, is a catch-all for many different aspects of antisense R&D. Topics range from alternative mechanisms (triplex DNA and ribozymes), through novel chemistry (second-generation 2′-modified oligonucleotides, locked nucleic acids, peptide nucleic acids and morpholino oligonucleotides), to delivery in different organs, and finally to ASO potential in various disease areas (respiratory, antiviral, anti-inflammatory, and cardiovascular). This section also includes a comprehensive review by Arthur Krieg (University of Iowa, Iowa City, IA, USA) of immunostimulation by oligonucleotides, a common drug effect that has confounded the interpretation of many antisense studies in animal models.

  PII: S1359-6446(01)02123-7