DDT Vol. 6, No. 15 August 2001

Proteomics: delivering new routes to drug discovery – Part 1

Kewal K. Jain, Bläsiring 7, CH-4057 Basel, Switzerland, tel/fax: +4161 692 4461, e-mail: jain@pharmabiotech.ch, URL: http://pharmabiotech.ch

The recent IBC conference on *Proteomics: Delivering New Routes to Drug Discovery* (14–17 May 2001, Philadelphia, PA, USA) provided a broad coverage of the application of innovative proteomic technologies for drug discovery. It would be difficult to review the large number of presentations during the four days and details of technologies are described elsewhere[1]. The first part of this conference review deals with the role of proteomics in drug development and structural genomics.

Functional genomics and molecular profiling

Lothar Germeroth (Cytos Proteome Therapeutics, Konstanz, Germany) described the company's protein-discovery technology as an alternative to conventional approaches to decipher gene function on a genomic scale. By applying highly sophisticated *in vitro* and *in vivo* assays, a virally reproduced library is rapidly screened for drug candidates. This functional genomics approach enables the reproduction of the entire human proteome as a large library of therapeutic human proteins, which enables target identification, target validation and small-molecule profiling.

The pitfalls of proteomic analysis caused by polymorphism mis-annotation and chimeric mRNAs were described by Christopher Southan (Gemini Genomics, Cambridge, UK). These anomalies can be uncovered by a detailed manual examination of all the sequences and literature links.

Proteomic technologies allow the identification of proteins associated with phagosomes – organelles that enable macrophages to participate in tissue remodelling, clear apoptotic cells and restrict the spread of intracellular pathogens[2]. The advantages of an organelle-proteomic approach is that it provides functional insights into signalling pathways and protein orientation. Similarly, John Bergeron (Caprion Pharmaceuticals, Montreal, Canada) described the use of subcellular proteomics for identification of novel post-transcriptional drug targets, which can be used for target validation, determination of lead mechanism of action and toxicity profiling.

Target validation in the post-genomic era

Strategic requirements for drug-target validation in the post-genomic era require an increase in speed as well as quality. High-quality cell models can facilitate this process but, frequently, cell lines are difficult to obtain in large numbers, difficult to transfect and undergo replicative quiescence.

Mark Lindsay (AstraZeneca, Macclesfield, UK) described developmental cell models for target identification and validation, in which conditional immortalization of genes can stimulate proliferation without differentiation, but when the proliferation stimulus is turned off the cells can differentiate again. Such immortalized cells are obtainable in large numbers and are easier to transfect or transform than primary cells, making them useful for target identification and prioritization. Specific targeting of a gene, protein or mRNA using antisense technology can modulate gene expression. However, these approaches are still 'hit-and-miss', and there are problems with toxicity and cell delivery. Alternative approaches include the use of DNA, peptide nucleic acid (PNA) or locked nucleic acid (LNA), which can be applied to gene targeting (by using triple-helix-forming oligonucleotides and zinc-finger proteins) and *in vivo* target validation (by acute application of oligonucleotides or genetic manipulation of ribozymes). Delivery to tissues is still a problem, but fusion of a protein to the protein transduction domain (PTD) of certain organisms results in direct delivery of active fusion protein to all tissues; *in vivo* delivery of a β-galactosidase–PTD fusion protein from the HIV protein, TAT, has been demonstrated[3].

High-throughput protein validation was discussed by Stewart Chipman (Immunex Corporation, Seattle, WA, USA). The protein-function screening process requires a panel of robust bioassays that are capable of quick, accurate and reproducible prediction of the immune function of novel proteins. An essential element is an inventory of purified, novel proteins as a consistent supply of biologically active material for screening. To this end, an information management system has been built to track project workflow, facilitate data analysis and disseminate scientific information.

High-throughput protocols for identifying highly active antisense molecules that inhibit gene expression were presented by Nicholas Dean (GeneTrove, Carlsbad, CA, USA), who highlighted potential applications of this approach

 PII: S1359-6446(01)01913-4

such as the *in vitro* and *in vivo* inhibition of interleukin-5 (IL-5)-mediated eosinopoiesis by murine IL-5 receptor-α (IL-5R-α) antisense oligonucleotide[4]. The use of antisense affords speed, selectivity and flexibility to gene functionalization and target validation. GeneTrove's drug discovery platform can be applied to any gene, provides predictable pharmacokinetics and enables a rapid transition from gene targeting to human therapy.

Post-genomic drug development

Integrated technologies – chemical, biological and robotic – and their role in post-genomic drug development were discussed by James Lorens (Rigel, South San Francisco, CA, USA). Vast retroviral libraries and functional screening assays are often used in drug discovery programmes. Rigel uses its rapid target identification technology to identify and characterize the potential drug targets that arise from such assays. Proteomics and two-hybrid screening technologies are then used to study the interactions of these targets, which provides immediate validation of the target in the context of a disease-specific cellular response. Rigel's post-genomics combinatorial-biology technology addresses many of the limitations of traditional genomics-based drug discovery by bypassing the need to know the identity or sequence of the genes to discover new drug targets and, ultimately, the small-molecule drugs that can modulate them[5].

The nematode, *Caenorhabditis elegans*, can be used as a live test-tube to identify the targets of compounds with unknown modes of action, find new chemical drug leads and discover the function of human genes, Thierry Bogaert (deVGen, Ghent, Belgium) explained. By feeding with bacteria expressing an appropriate gene, genes in disease-relevant pathways in *C. elegans* can essentially be silenced. This technology is used to validate candidate targets discovered by proteomics, gene-expression profiling and protein-linkage mapping. Validation is based on the restoration of selected disease pathways to normal function. This Function Factory™ technology is used to analyze reduced gene function or the effect of compounds, as well as to profile phenotypes of novel genes or compounds. This technology is applicable to many target classes including ion channels and scaffold kinases.

The Human Genome Project and microbial sequencing have provided several new targets for drug discovery, which include many proteins of unknown function. Approximately one-third of the genes in *Streptococcus pyogenes* have no identifiable function[6]. To meet these challenges, David Nelson (Anadys Pharmaceuticals, San Diego, CA) presented an approach to drug discovery that uses integrated biology and medicinal chemistry to advance from proteomics to drugs. The proprietary technology of the company is GATE (Genetics Assisted Target Evaluation), in addition to technologies for HTS of challenging targets for drug development that include: ATLAS (Any Target Ligand Affinity Screen) for protein targets; SCAN (Screen for Compounds with Affinity for Nucleic Acids) for RNA targets; and riboproteomics (systematic annotation of RNA–protein interactions that affect RNA metabolism: transport, splicing, translation and decay). Anadys is using these technologies to translate human and microbial genomic information into small-molecule therapeutics.

Structural genomics

The goal of structural genomics is to obtain useful, three-dimensional (3D) models of all proteins using a combination of experimental structure-determination and comparative model building. The shapes of most protein sequences are modelled on their similarity to experimentally determined protein structures[7]. Dennis Vitkup (MIT Genome Center/Harvard University, Cambridge, MA, USA) described target-selection strategies in structural genomics. Structural-genomic approaches include the organization of known protein sequences into domain families and the selection of family representatives as targets for structure determination by X-ray crystallography and NMR. The strategy that maximizes structural coverage requires about seven times fewer structure determinations compared with the strategy in which random target selection. With a choice of reasonable model-quality and the goal of 90% structural coverage, it would take ~16,000 carefully selected structure-determinations to construct atomic models for the majority of proteins[8]. This task could be accomplished within a decade provided that selection of targets is highly co-ordinated and significant funding is available. One of the examples given was the Pfam protein family database, a collection of protein family domains, the latest version of which contains 2000 families, 260,000 domains and 180,000 sequences. In *Escherichia coli*, 52% of the sequences have at least one Pfam.

Mark Gerstein (Yale University, New Haven, CT, USA) focused on analyzing functional genomic data as a finite list of protein 'parts' that can be either a protein fold or a protein family. A new web resource called PartsList (http://www.partslist.org) enables comparative protein-fold surveys[9]. It will rank the ~420 protein folds based on more than 180 structural attributes. The integration of whole-genome expression and protein–protein interaction data with structural information is a novel feature of this system. The 'parts' can then be used to interpret genomes and mine expression data.

The simultaneous high-throughput crystallographic and NMR spectroscopic analysis of pharmaceutically and medically important proteins, as used at the Berlin Protein Structure Factory (http://www.fu-berlin.de/psf), was described by Konrad Büssow (Max Planck Institute for Molecular Genetics, Berlin, Germany). This approach will close

the gap between genomic and structural information. Cheryl Arrowsmith (Integrative Proteomics, Toronto, Canada) also emphasized the role of NMR in structural proteomics and regards it as a link between proteins and chemistry that has the potential to be high-throughput.

Many structural genomic initiatives have focused initially on solving the structures of readily expressed proteins, but high-throughput protein production remains a challenge. Shane Teremi (Schering-Plough Research Institute, Kenilworth, NJ, USA) described how to build the protein-production infrastructure that addresses this problem. Such 'directed structural genomics' involves strategies in which various high-throughput technologies, including multiple expression-vectors, engineered host-cells and solubility sparse matrixes, have been incorporated at both the expression and purification stages. Schering-Plough's proprietary GATEWAY™ technology uses converted, destination-vectors (incorporating cleavable N-termini and green fluorescent protein C-termini) for protein expression and engineering. A protocol was presented for the entire process from target gene cloning and protein expression to growing crystals; it is possible to automate this process. The benefits of such an approach are that it can assess more targets simultaneously and rapidly determine the suitability of targets for structural studies.

The New York Structural Genomics Research Consortium has yielded structural information on five enzymes in the cholesterol biosynthesis pathway. Stephen Burley (Rockefeller University, New York, NY, USA) discussed the rapid expression, purification, crystallization and structural characterization of two of these enzymes. Comparative protein-structure modelling with these structures gave reliable homology models for >210 different proteins, giving an insight into biochemical function and protein evolution. The problem of target selection from >500,000 protein sequences is reduced to selection from ~30,000 families at 30% sequence identity, therefore, the families can be prioritized.

Structure-based drug design

Raymond Salemme (3-Dimensional Pharmaceuticals, Exton, PA, USA) described the DirectedDiversity™ technology, a combinatorial chemistry approach with structure-based drug-design technology that uses X-ray crystallography to directly visualize how drugs bind to a target receptor. This approach enables the simultaneous investigation and optimization of multiple drug properties throughout the process of drug design.

David Bailey (De Novo Pharmaceuticals, Cambridge, UK) described the use of structural genomics to expand the medicinal chemistry vision and predicted that human therapeutic targets defined by genomics could yield small-molecule therapeutics by 2005. The next wave of drug discovery will be driven by chemical genomics linked to exploratory ligand-design.

Assigning function from structure

Jeffrey Skolnick (Danforth Plant Science Center, St Louis, MO, USA) described a novel method for protein function prediction based on a sequence-to-structure-to-function paradigm. The structure-based approaches to predict protein function from amino-acid sequence can be applied to entire genomes and can be extended to ligand docking and quaternary structure prediction.

New technologies for merging structure-based drug design and structural genomics were discussed by Peter Rose (Pfizer La Jolla Laboratories, San Diego, CA, USA). Comparative modelling enables the generation of 3D models of target proteins from closely related homologues of known structure. These techniques can be used to evaluate putative targets for drugability, to facilitate the structure solution of new targets and to design selective inhibitors for human targets or broad-spectrum inhibitors for anti-infective targets.

Concluding remarks

This was a comprehensive proteomics conference, the size of which reflects the expansion in proteomics technologies and applications. In addition to the presentations highlighted here, there were parallel sessions on sample preparation and strategic corporate alliances. The second part of this conference report will be published in the 15th August 2001 issue of *Drug Discovery Today* and will deal with structure-based drug design and analysis of proteomic information.

References

1 Jain, K.K. (2001) *Proteomics: technologies and commercial opportunities* (7th edn), Jain PharmaBiotech Publications

2 Garin, J. *et al.* (2001) The phagosome proteome: insight into phagosome functions. *J. Cell Biol.* 152, 165–180

3 Schwarze, S.R. *et al.* (1999) *In vivo* protein transduction: delivery of a biologically active protein into the mouse. *Science* 285, 1569–1572

4 Lach-Trifilieff, E. *et al.* (2001) *In vitro* and *in vivo* inhibition of interleukin (IL)-5-mediated eosinopoiesis by murine IL-5R alpha antisense oligonucleotide. *Am. J. Respir. Cell Mol. Biol.* 24, 116–122

5 Peelle, B. *et al.* (2001) Intracellular protein scaffold-mediated display of random peptide libraries for phenotypic screens in mammalian cells. *Chem. Biol.* 8, 521–534

6 Ferretti, J.J. *et al.* (2001) Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4658–4663

7 Sanchez, R. *et al.* (2000) Protein structure modeling for structural genomics. *Nat. Struct. Biol.* 7 (Suppl.), S986–990

8 Vitkup, D. *et al.* (2001) Completeness in structural genomics. *Nat. Struct. Biol.* 8, 559–566

9 Qian, J. *et al.* (2001) PartsList: a web-based system for dynamically ranking protein folds based on disparate attributes, including whole-genome expression and interaction information. *Nucleic Acids Res.* 29, 1750–1764