Already established tools, such as similarity searching, clustering, nonlinear mapping (including Kohonen maps) and other virtual screening tools, aid in the selection of the most promising sublibraries. If the 3D structure of the biological target is known, docking methods are applied in later stages, for example, GOLD, FlexX [3,4] and, more recently, FlexE [5], which simultaneously considers the flexibility of the ligand and the binding site.

However, filtering and virtual screening of huge libraries is a waste of time if drug-like libraries can be generated from scratch. The program Comb^{i}Gen, developed at the University of Innsbruck (Austria), combines and modifies several hundreds of structure-optimized, drug-specific ('privileged') fragments to produce, *de novo*, structurally diverse, virtual combinatorial libraries with a high percentage of drug-like molecules [6]; ~1,000,000 structures can be generated per CPU day at a standard personal computer.

Another strategy of the future will be the combinatorial design of ligands within their binding site, following a concept that was first formulated by Böhm *et al.* [7] and which inherently applies the algorithms of the program FlexX. Instead of docking thousands of ligands separately, a recent extension of the Flex software suite, called FlexX-C [8], constructs a whole library within the binding site of the protein, step-by-step. Although such approaches are highly appealing, all current docking methods suffer from the insufficient precision of the scoring functions. Further research in this direction is required to proceed with combinatorial ligand design.

References

1 Lahana, R. (1999) How many leads from HTS? *Drug Discov. Today* 4, 447–448

2 Rose, S. (2002) Statistical design and application to combinatorial chemistry. *Drug Discov. Today* 7, 133–138

3 Böhm, H.-J. and Schneider, G. (eds) (2000) *Virtual Screening for Bioactive Molecules. Methods and Principles in Medicinal Chemistry.* (Vol. 10), Wiley–VCH

4 Schneider, G. and Böhm, H-J. (2002) Virtual screening and fast automated docking methods. *Drug Discov. Today* 7, 64–70

5 Claußen, H. *et al.* (2001) Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 308, 377–395 (http://www.biosolveit.de)

6 Wolber, G. and Langer, T. (2001) Comb^{i}Gen, a novel software package for the rapid generation of virtual combinatorial libraries. In *Rational Approaches to Drug Design, Proceedings of the 13th European Symposium on Quantitative Structure–Activity Relationships,* (Höltje, H-D. and Sippl, W., eds), pp. 390–399, Prous Science (http://pharmazie.uibk.ac.at/combigen.html)

7 Böhm, H-J. *et al.* (1999) Combinatorial docking and combinatorial chemistry: design of non-peptide thrombin inhibitors. *J. Comput.-Aided Mol. Design* 13, 51–56

8 Rarey, M. and Lengauer, T. (2000) A recursive algorithm for efficient combinatorial library docking. *Persp. Drug Disc. Des.* 20, 63–81

Hugo Kubinyi
University of Heidelberg, Germany
and BASF AG
Ludwigshafen, Germany

Functional proteomics: separating the substance from the hype ▼

The completion of the draft sequence of the human genome by the public consortium and Celera Genomics (Norwalk, CT, USA) has stirred controversy regarding the lower than expected number of genes in humans. Although the debate will continue for the next several years regarding the 'true' number of genes, the consensus is that the next bottleneck is the identification of the genes' function(s). Proteins, the main executors of biological function, have been estimated to be magnitudes greater in both number and complexity than genes. From the pharmaceutical industry's point of view, proteins represent the majority of molecules targeted by the drugs in the market today. An immediate requirement in the pharmaceutical industry is to prioritize which protein targets are suitable bases for developing novel and effective therapies.

The new opportunity created by the plethora of gene sequences with unknown function and the urgent need to understand protein function has led to the frequent misuse of the term 'functional proteomics'. Increasing interest from the financial markets has compounded this misuse. As the buzz moves from genomics to proteomics, it is important to separate the substance from the hype. Originally, proteomics focussed on the identification of the protein complement of the genome [1] but has since been extended to include everything associated with proteins on a global scale [2]. It is thus necessary to subdivide proteomics to fully grasp this burgeoning field, and the effort by Blackstock and Weir [3] to do this represented the first attempt to clarify the jargon. This communication extends the subdivision of proteomics and concurs that dividing proteomics and understanding each segment will facilitate conquering the complexity of proteins, which will lead to a better understanding of their biological function(s). Extending from the Blackstock and Weir nomenclature, proteomics can be subdivided into the following categories: (1) Expression proteomics; (2) Cell-map or interaction proteomics; (3) Functional proteomics; and (4) Structural proteomics.

Expression proteomics

Expression proteomics involves the creation of quantitative maps of expressed proteins from cell or tissue extracts (by separation methods such as 2D gels and MS), similar to cDNA arrays for transcription profiling. An alternative terminology used is 'differential proteomics', which refers to the comparison of differences between biological samples. Expression of functional proteins and the advent of protein chips represent a burgeoning area of proteomics research and fall under this category. This comparative profiling of proteins, noting their

 1359-6446/02/$ – see front matter PII: S1359-6446(02)02275-4

presence or absence under different cellular or physiological states (e.g. diseased versus normal tissue), does not provide direct functional information on a particular protein that is differentially expressed. The problem with deducing functional relevance from differences in expression is difficult because it is not possible to conclude which differences provide true causal relationships to biological function.

Cell-map proteomics

Cell-map or interaction proteomics involves determining the subcellular location of proteins and protein–protein interactions by purifying organelles or protein complexes, followed by MS identification. This category includes other protein interaction technologies, such as the popular yeast two-hybrid method and variations based on cellular systems that accomplish the same objective. Although the identification of protein–protein interactions might provide an insight into protein function and could elucidate functional pathways that are relevant to disease, the information is not sufficient to delineate a protein's physiological role within a proper functional context. Binding of a protein with another protein to form a multimeric complex to execute a particular function represents a type of functional context for proteins, but to collectively term the identification of protein–protein interactions as 'functional proteomics' is misleading. Apart from their interaction with other proteins to form functional protein complexes, proteins can have other diverse functions, including their roles in enzymatic activities, molecular transport, and so on.

Functional proteomics

Although both aforementioned approaches frequently claim to elucidate the function(s) of a protein, they have been casually labelled as functional proteomics. By contrast, functional proteomics is the systematic perturbation or functional inactivation of proteins within a given physiological environment to address the potential role of the target protein in a cellular process, akin to gene disruption to elucidate the function of a gene [4]. Furthermore, it is becoming clear that proteins can perform multiple functions depending on the cellular location of the protein, the cell type in which the protein is expressed, multimeric state of the protein and substrate bound to the protein [5]. Approaches in functional proteomics include the use of specific blocking molecules, such as neutralizing antibodies and aptamers, pharmacological inhibitors, transdominant mutations and chromophore-assisted laser inactivation. Chemical genetics can be classified under this category because small synthetic organic molecules or natural products are used to elucidate the function of proteins [6].

Structural proteomics

The recent large-scale initiative to determine the 3D structures of proteins has been inappropriately labeled structural genomics. A more appropriate term is structural proteomics because it is the structure of the proteins that is being determined and not the nucleic acids corresponding to the gene sequence. Identification of all potential protein folds should provide the structural framework for understanding how a protein executes its function. However, identification of all protein folds only represents the first step because proteins can have multiple structural domains with corresponding functional properties; therefore, it will be necessary to link these two types of information. Because of the current lack of understanding of protein SARs, it is still not possible to deduce function from a protein's 3D structure with much confidence. This becomes more complicated in humans where most genes are fusions of different structural domains providing further functional diversity.

The goal of categorizing proteomics is to provide a systematic framework for studying the different and complex aspects of proteins. Solutions from the different segments should be recombined to achieve the higher goal of understanding one of the fundamental problems in biology – the relationship of protein structure-to-function, and the burgeoning field of 'systems biology'. This can only be achieved through the synergies from the different categories described previously. Expression proteomics provides an important step in cataloging differences between different physiological states, which can be the basis for determining which proteins are involved in a particular pathway via cell-map or interaction proteomics technologies. The association of the different proteins within a particular pathway can then be tested for functionality by different functional proteomics methods. Finally, information from the previous categories can be related to protein structure (structural proteomics) to form the basis for the rational design of safer and more effective therapies.

References

1 Wilkins, M.R. *et al*. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* 13, 19–50

2 Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* 405, 837–846

3 Blackstock, W.P. and Weir, M.P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17, 121–127

4 Ilag, L.L. *et al*. (2000) Chromophore-assisted laser inactivation (CALI) to validate drug targets and pharmacogenomic markers. *Drug Dev. Res.* 49, 65–73

5 Jeffery, C.J. (1999) Moonlighting proteins. *Trends Biochem. Sci.* 24, 8–11

6 Schreiber, S.L. (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* 287, 1964–1969

Jocelyn H. Ng and Leodevico L. Ilag*

Xerion Pharmaceuticals

Fraunhoferstrasse 9

82152 Martinsried

Germany

**tel: +49 89 86307 200*

fax: +49 89 86307 222