the VS world in the past decade. With distributed computing comes the potential to increase available processor power by several orders of magnitude. The simple manner in which VS databases can be divided up onto separate processors makes VS a prime candidate application for such technology. The resultant increase in power available is already being used to screen massive libraries (http://www.chem.ox.ac.uk/cancer/thinksoftware.html).

Using current technology, however, larger libraries bring with them more noise. Further post-screen analysis becomes a particularly daunting prospect given the huge increase in the hit numbers produced. Another, perhaps more promising, long-term alternative is to perform better, rather than more, calculations. By increasing sampling and using more accurate descriptors of the forces that govern binding, the number of false positives should be dramatically decreased, thus increasing the hit rates derived from VS calculations.

One final comment on a practical element of VS application that is generally neglected in technical reviews on the subject. A primary objective of such screens is to dramatically refine the list of compounds that is put forward for screening. Such focussing is often used to buy-in compounds that complement those screened in HTS. In addition, this number is often small enough that screening samples can be freshly made up from solid store. As a result, the issues of liquid sample stability and concentration inaccuracies are dramatically reduced, enabling a second 'bite' at compounds already screened but possibly missed in HTS campaigns.

References

1 Good, A.C. *et al.* (2000) High-throughput and virtual screening: core lead discovery technologies move towards integration. *Drug Discov. Today (Suppl.)* 5, S61–S69

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

3 Böhm, H-J. *et al.* (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization: a promising alternative to random screening. *J. Med. Chem.* 43, 2664–2674

Andrew Good
Bristol-Myers Squibb
5 Research Parkway
Wallingford 06492, CT, USA

Conquering the proteome

Leodevico L. Ilag, Xerion Pharmaceuticals, Fraunhoferstr. 9, 82152 Martinsried, Germany, tel: +49 89 86 307 201, fax: +49 89 86 307 222, email: l.ilag@xerion-pharma.com

The fifth annual IBC *Proteomics and the Proteome* conference (18–20 February 2002, Geneva, Switzerland) began with a keynote presentation by Hanno Langen (F. Hoffmann-La Roche, Basel, Switzerland) who outlined state-of-the-art proteomics technologies. To maximize the identification of different proteins from complex biological material, such as biological fluids, he recommended that several fractionation steps (subcellular fractionation and classical chromatographic methods) should be performed before separation via two-dimensional (2D) gel electrophoresis. For example, albumin comprises 50% of the whole protein content of plasma and the other 40% includes immunoglobulins, transferrin and haptoglobulin. By removing these abundant proteins before separation, the sensitivity of subsequent 2D electrophoresis is increased 10-fold.

In his keynote address, Dennis Hochstrasse (Geneva University Hospital, Geneva, Switzerland) demonstrated interesting parallels between the way computers and cells generate and process information. Based on this, he shared the concept of connecting the virtual and the real worlds, that is, combining bioinformatics with experimental biology. He reiterated the problem of uneven distribution of protein concentrations inside the cell when resolving the identity of different proteins, and described a different approach to performing experiments or extracting proteins from large-scale batches of samples to isolate the least expressed proteins.

The use of free-flow electrophoresis (FFE) as a preparative method for fractionation and enrichment of proteins was described by Christoph Ekcerskorn (Tecan Group, Munich, Germany) as a way to reduce the complexity of samples. Enrichment facilitated the visualization of less abundant proteins for subsequent separation by 2D electrophoresis or LC–MS analysis. FFE has been used to process protein samples from intact cells, organelles, membranes and protein mixtures and is reported to be high throughput, with high quantitative recovery and high resolution.

Differential protein analysis without 2D electrophoresis was the topic of Pierre Thibault's presentation (Caprion Pharmaceuticals, Quebec, Canada). He

1359-6446/02/$ – see front matter PII:S1359-6446(02)02263-8

explained how MS and bioinformatics can be combined to measure relative protein abundance. Proteomic analysis is performed on human tissue where macrodissections of normal and viable diseased membranes are isolated from the same patient. Using immunoisolation methodology, plasma membranes are isolated to homogeneity and quality controlled by western blotting and microscopy. The proteins from the plasma membrane fractions are separated by SDS-PAGE and analyzed by tandem MS combined with proprietary bioinformatics software to generate comprehensive protein expression profiles. Results were presented that validate the approach in model systems, such as epidermal-growth factor (EGF) stimulation of rat plasma membranes and the transformation of U937 cells from monocytes to macrophages.

Ian Currie (Amersham Biosciences, Little Chalfont, UK) addressed a major issue in obtaining reliable data from differential protein analysis using 2D electrophoresis. He described how the 2D Difference Gel Electrophoresis (2D DIGE) system can run up to three samples per gel, prelabelled with a different molecular weight and p*I*-matched with corresponding dyes called CyDye. Co-mapping three samples in one gel removes variation between gels and reduces the number of gels that needs to be run. He demonstrated the benefits of using an in-gel standard to facilitate the distinction between gel-to-gel variation from true biological variation.

Thomas Neumann (Xzillion, Frankfurt, Germany) described the PST (Protein Sequence Tag) technique, which is based on specific isolation of either the N- or C-terminal peptides of a protein after modification and cleavage. The terminal tags can be used for differential protein analysis between diseased and normal tissues, and for subsequent identification of the protein without the need for additional peptide information.

Detecting the proteome

For future diagnostics applications, it is envisioned that protein chips will be miniaturized wafers that contain proteins or protein ligands that can be used to probe for a specific biomarker. Three major challenges to the development of protein chips were highlighted at the conference – surface chemistry, protein (or ligand) stability and detection methods.

Leigh Anderson (Large Scale Biology, Vacaville, CA, USA) described the early efforts of proteomics via the Human Proteome Index and how proteomics got supplanted by the genomics revolution. He commented that, despite the existence of a draft sequence of the genome, there is a severe lack of full-length cDNA clones for expressing proteins and generating antibodies. The cost per array should be less than $10 for it to be economically feasible. To address this issue, he described his company's efforts to use plants as expression systems for proteins at a rate of 50–100 different proteins per month.

Multianalyte protein expression analysis was discussed by Larry Cohen (Zymoyx, Hayward, CA, USA). The criteria for this approach is the parallel analysis of tens to hundreds of analytes, their direct measurement in complex mixtures such as serum or cell lysates, the use of small sample volumes (<20 μl), and sensitivity comparable with low throughput methods such as enzyme-linked immunosorbent assays (ELISAs), while maintaining high throughput and reliability.

Two issues regarding surface chemistry and proper orientation of the probe were addressed. A biotin-derivatized poly(L-lysine)-g-poly(ethylene glycol) (PLL) anchoring backbone tethered to a TiO_2 chip gave more uniform results compared with conventional microarray spots on nitrocellulose. Furthermore, proper orientation of the antibodies on the surface of the chip gave better signal intensities compared with randomly oriented antibodies.

Larry Gold (SomaLogic, Boulder, CO, USA) described the use of nucleic acids as probes for proteins. Photoaptamers are nucleic acid mimics containing halogenated bases that bind specifically and covalently to virtually any protein. Selection pressure during *in vitro* evolution can be applied for both binding specificity and specific photo-cross-linkability. Because nucleic acids are more stable than proteins, photoaptamers are readily arrayed on a chip surface. Analyte detection takes advantage of the chemical differences between proteins and aptamers, enabling universal protein staining without detection of aptamers within the array. Because of the covalent link formed between the photoaptamer and the analyte protein, extensive washing virtually removes background providing better signal:noise ratio. The use of photoaptamers was demonstrated by detecting femtomole amounts of HIV gp120 in a sample containing 5% serum.

Protein–protein interactions

A genome-wide approach to identifying protein complexes in *Saccharomyces cerevisiae* was presented by Bernhard Kuster (Cellzome, Heidelberg, Germany). Protein complexes are isolated by using Tandem Affinity Purification (TAP) tagging, in which an affinity tag is fused to a target gene by homologous recombination in yeast. Two steps of high affinity purification are used to finally capture the protein complex. The complexes are resolved by 1D gel electrophoresis and the proteins analyzed by MS using matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF). This approach covered 56% of known complex components while analyzing only 25% of the yeast genome. By contrast, yeast-two-hybrid screens covered only 10% of known complex components while analyzing 95% of the genome. This difference can be attributed to the fact that protein complexes are more than a sum of binary interactions and involve cooperative binding.

Donny Strosberg (Hybrigenics, Paris, France) described the use of automated, large-scale, yeast-two-hybrid technology to identify interacting proteins as potential drug targets. He presented data on the protein–protein interaction map of *Helicobacter pylori* where they were able to determine interactions for 46.6% of the proteome. From these results, it is possible to identify the Selected Interacting Domain (SID) that can be used as a tool for target validation. A SID against a particular target in *H. pylori* was used to inhibit the proper formation of the bacterium's flagellum.

A variation of the yeast-two-hybrid system that addresses membrane proteins was described by Igor Stagljar (University of Zürich, Switzerland). The split ubiquitin approach exploits the ability of the C-terminal (Cub) and N-terminal (Nub) domains of ubiquitin to associate non-covalently and restore ubiquitin function. This was adapted so that the membrane protein of interest (X) was fused to Cub and a reporter gene, while the potential interacting protein (Y) is fused to Nub (Nub-Y). Interaction between X and Y results in ubiquitin reconstitution and leads to the proteolytic cleavage and subsequent activation of an easily measurable reporter system. This method was successfully applied to erbB3 and a G-coupled-protein receptor.

Integrating technologies for target validation

Sam Hanash [University of Michigan, Ann Arbor, MI, USA, and President of the Human Proteome Organization (HUPO)] gave an overview of the objectives of HUPO. Its main focus is to manage and develop resources such as distributed reagents and standards, antibodies, protein microarrays and protein databases. The first major proteome project will be the human plasma or serum proteome because of its relevance to disease, amenability to multiple technologies and ease of standardization. Another key initiative is the construction of a Proteome Knowledge Database with universal user access to facilitate the cataloguing and annotation of all proteins.

The use of chromophore-assisted laser inactivation (CALI) in the identification and validation of drug targets was presented by Vic Ilag (Xerion Pharmaceuticals, Martinsried, Germany). This technology can validate correlative data from RNA and protein differential expression and *in silico* data. The application of CALI to the direct identification and validation of protein targets that are involved in the invasion of metastatic cells was described.

Josh LaBaer (Harvard Medical School, Boston, MA, USA) described the creation of FLEXGene, a repository of full-length cDNAs of all human genes. The aim is to automate different methods involved in cloning and protein expression, such as high-throughput protein expression and purification using fusions with affinity tags. The expression clones have been validated by demonstrating the expected phenotype in reporter gene assays. In addition, LaBaer described his group's efforts to provide an annotation system for different genes by integrating the FLEXGene repository with other information, such as medical literature and disease association. Although not yet comprehensive, the preliminary results demonstrated convincing clustering of relevant data associated with the set of genes presented.

An integrated approach to combining genetic and proteomic technologies with patient data to develop better tools for evaluating therapeutic strategies was discussed by Marc Reymond (Europroteome, Henningsdorg, Germany). His talk emphasized the need to account for inter-individual variability, as well as the need for careful selection and processing of clinical samples, and he highlighted the power of combining several technologies, standardizing data formats and linking all of this to clinical outcome.

Informatics

Hans-Werner Mewes [Institute of Bioinformatics, National Research Centre for Environment and Health (GSF), Germany] provided a critical assessment of how bioinformatics should be combined with experimental biology. He contends that bioinformatics must provide methods for its interpretation and introduced the term 'combinatorial bioinformatics', whereby it is necessary to combine different attributes of proteins to gain additional information or a better understanding of proteins. These attributes can be derived from different experimental data, such as differential expression analysis, structural information, protein–protein interactions and functional analysis. Such an approach can be used in the refinement of future experimental design and the formulation of new hypotheses that can be verified empirically.

Rolf Apweiler, from the European Bioinformatics Institute (EBI; Cambridge, UK – an outstation of the European Molecular Biology Laboratory), gave an informative overview of the different public databases and the need to integrate both public and private databases, giving the example of the Human Proteomics Initiative (SWISS-PROT–TrEMBL) between the Swiss Institute of Bioinformatics (SIB; Geneva, Switzerland) and the EBI. The goal of this initiative is to annotate all known human polymorphisms at the protein level, post-translational modifications, links to structural information, clustering, and the relationship of this information to biological pathways.

Industrial proteomics

High throughput approaches are necessary to rapidly sort through the complexity of the proteome. Several companies described their different industrial scale approaches, all of which employed scores of MS instrumentation, liquid chromatography systems and quantification methods, such as isotope-encoded affinity tags (ICAT) and DIGE. Although

some companies still use 2D gels, others claimed that they exclusively focus towards using liquid chromatography. It is noteworthy that at this scale, information technology becomes crucial, with most companies developing sophisticated hardware and data storage systems. The power of this approach was evident in the talk of Keith Rose (GeneProt, North Brunswick, NJ, USA) who demonstrated that in a period of six months his company was able to identify and validate at least two potential protein therapeutics.

Conclusions

A year after the publication of the draft human genome sequence, work on the human proteome is rapidly under way. Major challenges lie ahead, however, but it is rewarding to note that major technological advances are keeping pace. It is noteworthy that different aspects of the complex proteome are already being addressed at this early stage. I look forward to a progress update at the sixth annual IBC *Proteomics and the Proteome* conference next year.

1359-6446/02/$ – see front matter PII: S1359-6446(02)02228-6