

Applying new analytical technologies

A recent meeting heard about the various applications of genomics, proteomics and metabonomics in drug development and discovery.

Geoffrey Phillips, honorary secretary of the Joint Pharmaceutical Analysis Group, reports

A comprehensive assessment of DNA microarray technology and its applications in drug discovery was presented by Colin Smith, professor of functional genomics at the University of Surrey. He said that conventional gene expression (transcription) tediously involves isolation of one gene string at a time whereas, in the “post-genomic era” of functional genomics, it is feasible to design specific DNA microarrays that represent individual genes and their cell complements. He used as an example the addition of labelled RNA in order to recognise particular spots in the array matrices. The importance of “gene expression profiling” arises from simultaneous quantification of the transcriptional activity of all genes in a cell population — which he likened to knocking out one gene causing a “house of cards” collapse.

Professor Smith distinguished a variety of microarray platforms with dual or single sample labelling and different fluorogen colour development. He presented an extensive series of applications of DNA microarrays; these included identification of unknown genes in biochemical pathways, novel cross-linked activity of different genes, and potential drug targets, as well as various monitoring operations and use as diagnostic tools in various human cancers.

He reviewed current developments in microarray gene-specific analysis and dynamic models of spliced elements, including several applications of microarray systems in the context of drug discovery and development. Gene expression profiling is important in cancer research and diagnosis — such as paediatric acute lymphoblastic leukaemia and in

identifying a “poor prognosis” gene signature for breast cancer. He asserted that microarray techniques have “redefined the drug discovery process and provided better therapies.”

Professor Smith defined chemogenomics as a study of genomic or proteomic response to chemical compounds by an intact (whole cell) biosystem, ie, “study of the ability of isolated molecular targets to interact with such compounds”. He concluded that “the future is bright, but complicated” for microarray techniques; there is a particular need for better data integration, especially between complex bioinformatic and chemoinformatic data. Asked about comparability of different systems, he accepted that operational procedures for different spotted arrays could lead to different results, perhaps prompting different clinical decisions.

Proteomics in drug discovery

Robert Massé, of MDS Pharma Services, Montreal, Canada, reviewed the role of proteomics in discovery and identification of safety biomarkers, primarily involving mass spectrometry (MS) and related techniques. Hitherto, in 20 years’ use as drug usage indicators, there had been little idea of identifying a set of markers that would allow clinicians to determine whether a trial subject would respond to a particular drug. “Today, that scene has changed dramatically,” said Dr Massé. Companies could save up to \$100m in their development costs if potential failure of the candidate product could be predicted early enough. He listed five clear “decision gates” — the initial combinatorial or biological synthesis, *in silico* pre-toxicology studies, formal toxicology assessment, clinical safety evaluation and, ultimately, the commercial launch — and even then there could be unexpected failure.

Research on biomarkers is now booming as scientists seek to discover, identify and use various biomarkers as diagnostic or prognostic tools, although he warned that the field is still evolving; the actual use of biomarkers in all phases of drug discovery and development, including clinical trials, is “by no means commonplace, even today,” he said. His overview of the biomarker discovery technology platforms that had been developed and implemented by MDS Pharma highlighted how these could play an enabling role in all phases of drug discovery and development processes.

Dr Massé exemplified the feasibility by a proof-of-concept case study, using the well-established nephrotoxicity of puromycin aminonucleoside in rats to identify potential protein safety biomarkers, noting that hepatic and renal safety are two of the most important components in a drug approval process.

Dr Massé then described miniaturisation, whereby protein chemistry “on a chip” ultimately defined an amino acid sequence for each protein, with the MS profile identifying biomarkers to use in high throughput analysis of new candidate drugs. He concluded that the results demonstrated the distinctive capabilities of his company’s integrated proteomic technology platform, which encompassed novel methodologies and tools for protein isolation, processing, identification and measurement that could be applied to a wide variety of biomarker discovery and determination issues across all phases of the drug discovery and development process.

Biomarkers in drug safety

Andrew Nicholls, of GlaxoSmithKline, Ware, Hertfordshire, described his concept of biomarkers in “metabolic profiling in drug safety”. He distinguished the roles of diagnosis, extent and prognosis of disease, prediction of clinical response and surrogate end-point prediction of therapeutic benefit. The optimum assessment of drug safety requires the identification of adverse events at the earliest possible stage of development and, here, analytical improvements for the study of “the gene-protein-metabolite triptych” critically provide novel opportunities for enhancing perceived understanding of drug action. As mass spectrometry (MS) methods have become more routinely applied, so proteomics and metabolomics are increasingly perceived as the two extremes of the metabolic profile. For better biological understanding, laser-induced-ionisation time-of-flight MS had been applied as a screening tool, rather than just an analytical probe. From a safety perspective, such technology held much promise. Nevertheless, he concluded, the “characterisation of the protein markers remained the major bottleneck to the application of this method”.

Dr Nicholls noted that most metabolic profiling studies have focused on the effects of classical toxins and large genetic variations. He emphasised that studies of subtle, recoverable effects and small genetic differences require exacting study design, careful sample handling, high analytical sensitivity and data interpretation to ensure the accurate assessment of the metabolic constituents. As work in metabolic profiling has progressed, various components have been identified that provide information as to the biological region of effect, or to the mechanism underpinning the observations. Improved knowledge has also shown that some “toxicity” markers are indicative of covariant biological effects or cellular adaptation to the induced effect. He claimed that such “house-keeping” biochemicals represent a powerful set of biological information when interpreted correctly and are not simply general markers of cellular dysfunction.

He outlined applications of multivariate statistical methods, including principal component analysis (PCA), that had been used to focus on key characteristics arising from drug-induced peroxisome proliferation in the rat. Speaking of the future, he envisaged clearer identification of metabolic constituents and a more comprehensive understanding of the cellular metabolic pathways and neighbourhoods that “form the underlying architecture of toxicity and disease”.

Using NMR and MS for metabolic profiling — a vital new area of research

John Shockcor, of Bruker BioSpin/Bruker Daltonics, Billerica, Massachusetts, reviewing tools for metabolic profiling, said he was a great advocate of using nuclear magnetic resonance (NMR) spectrometry and mass spectrometry (MS). Metabolic profiling has emerged as a vital new area of research, he said.

He commented that biochemical pathway charts ideally should be three-dimensional to reveal the full depths of interactions of enzymes with big molecules. Metabolic profiles of biological fluids and tissues contain a vast array of endogenous low-molecular weight metabolites. Their composition depends upon the sample type (plasma urine, bile etc) and factors such as the species, age, sex and diet of the organism from which the sample is derived — and even the time of day at which the sample is taken. Disease, drugs (and other biologically active molecules) perturb concentrations and fluxes in intermediary meta-

bolic pathways. He said the response to this perturbation involves adjustment of intracellular and extracellular environments in order to maintain homeostasis. Both the perturbations and the adjustments are expressed as changes in the normal composition of the biofluids or tissues that are characteristic of the nature or site of the disease process, toxic insult, pharmacological response or genetic modification.

Analytical techniques, particularly MS and NMR, provide spectral patterns that can be evaluated directly or with statistical methods such as PCA, to highlight both subtle and gross systematic differences between samples. Dr Shockcor maintained that understanding and evaluation of these observed biochemical changes over time could provide critical information on the mechanism of the perturbation. These data are also used to develop diagnosis and treatment for disease.

How to make sense of the “-omes”

Royston Goodacre, of Manchester University, helped the audience to “make sense of the -omes” through a description of explanatory machine learning for the rapid characterisation of biological systems. He traced them from study of the gene (genome) through messenger-RNA (four transcripts) to proteins (proteomes) and ultimately their metabolites, and illustrated this relationship with a picture of an iceberg floating with perhaps a 10th part visible above the ocean surface. He quoted Peter Drucker on information overload: “The fewer data needed, the better the information. An overload of information, that is, anything much beyond what is truly needed, leads to information blackout. It does not enrich, but impoverishes.”

Dr Goodacre confirmed that postgenomic science is “producing bounteous data floods” and the extraction of the most meaningful parts of these data is central to the generation of useful new knowledge. A typical transcriptomics, proteomics or metabolomics experiment could generate thousands of data points (samples multiplied by variables) of which only a handful might be needed to describe the problem adequately. He emphasised that current informatics approaches need to adapt and grow in order to make the most of the large amounts of data generated in postgenomic strategies. Especially necessary are good robust databases, good data, excellent visualisation methods and even better algorithms, with which to turn data into knowledge.

Dr Goodacre suggested that evolutionary algorithms are “ideal strategies for mining such data” to generate useful relationships, rules and predictions. These algorithms constitute explanatory supervised learning techniques from which are derived answers of biological interest, such as: “What metabolites have I measured in my metabolome that enables bacteria to be resistant to a specific antimicrobial?” He regarded these algorithms as particularly popular inductive reasoning and optimisation methods based on the concepts of Darwinian selection to generate and to optimise a desired computational function or mathematical expression to produce so called explanatory “rules”. Because the models are in English and penalise complex expressions, they could be made comparatively simple. Thus, they could be used to elucidate which inputs are important, thereby allowing selection of the most discriminatory and useful transcripts, proteins or metabolites.

Dr Goodacre illustrated these methods within the metabolomics area, referring to a form of metabolic fingerprinting with genetic algorithms. He showed that genetic programming could be used to detect a spore-specific chemical biomarker in bacterial spores and, in the area of food-technology, for the quantitative detection of metabolic markers for spoilage.

Impact of proteomics on drug discovery

The impact of proteomics on drug discovery was assessed by Hans Voshol, of Novartis Institute for Biomedical Research, Basel, Switzerland. He asserted that “separation methods are key in protein resolution”. In the 10 years since the term “proteomics” was first coined, the focus has been on high-throughput protein identification, with the ultimate goal to identify all proteins in the human proteome. While there has been huge progress with analytical, mainly hyphenated MS, methods, the real bottleneck — the reduction of sample complexity — “still awaited a quantum leap,” he said.

He discussed the possibilities and limitations of two different approaches for expression profiling of proteins. One approach performs separations at the protein level, usually by two-dimensional electrophoresis, which has the advantage of providing more inherent characterisation of proteins, because there is more sequence coverage and information on protein isoforms is retained. The alternative “shotgun”-type proteomics procedure begins by converting protein mixtures into peptides, followed by peptide fractionation, often online with tandem-MS. In both approaches, the identification is always based on fragments of the protein, usually peptides generated by tryptic digestion, because MS

can only yield the necessary accuracy and resolution in a limited mass range, far below that of intact proteins. The actual sequence of the protein is only partially “covered”, depending on how many peptides are recovered and assigned.

In tandem-MS sequencing, only small pieces of sequence information, perhaps just two amino acids in a single bipeptide, might be available for identification. Notwithstanding these restrictions, Dr Voshol claimed, the spectacular progress with MS has ensured that only rarely the identification of a protein of interest would be a limiting factor in proteomics.

Nevertheless, he believed, more research is necessary to turn high-throughput protein identification and other proteomic technologies into high-impact tools for pharmaceutical research in order to provide novel insights in disease processes and hence new drug targets. Multivariate data analysis would cope with say 20 to 100 samples and several hundred variables. Alternatively, they may seek particular patterns, replacing with targeted analysis and antibody arrays; but this was, he said, “still a long way from full pathway analysis”.

Another approach evolved from a focused proteomics technology platform to an integral part of a functional genomics environment. Dr Voshol illustrated this concept by outlining a recent case-study with bengamide-E. This had encompassed a wide spectrum of methods and tools that were integrated with “-omics” and were, he concluded, “pivotal for translating proteomic or genomic findings into novel biological insights”.

This meeting on genomics, proteomics and metabolomics in drug discovery and development was presented by the **Joint Pharmaceutical Analysis Group**. It took place on 2 December 2004 at the London headquarters of the Royal Pharmaceutical Society