Metabolic profiling: pathways in drug discovery

George Harrigan, Pharmacia Corporation, 700 Chesterfield Parkway, Chesterfield, MO 63198, USA, tel: +1 636 737 6663, fax: + 1 636 737 7300, e-mail: george.g.harrigan@pharmacia.com

The Cambridge Healthtech Institute's *Metabolic Profiling: Pathways in Discovery* conference (3–4 December 2001, Chapel Hill, NC, USA) focused on developments in spectroscopic tools for the acquisition of metabolite data, pattern recognition and modeling studies based on such data, and specific examples of metabolic profiling applications.

Spectroscopic tools

The development of spectroscopic tools for high throughput analyses of selected biochemical pathways is crucial to the development of metabolic profiling. Whereas 'metabolic fingerprinting' or 'metabonomics' assumes it is not necessary to determine the levels of all individual metabolites for classification or response readouts, 'metabolite profiling' or 'metabolomics' [1] absolutely requires the identification of a class of metabolites that is as broad as possible.

Conference discussions on the technology available for analyzing metabolite concentration and fluxes focused primarily on mass spectrometry (MS) and NMR. Both, however, lack the resolving power to distinguish all cell or tissue metabolites in a single spectral measurement. By contrast, the potential of Fourier-transform ion cyclotron resonance MS, a technology that enables the discrimination of metabolites that differ by as little as 0.005 mass units, was demonstrated by Dayan Goodenowe of Phenomenome Discoveries (Saskatoon, Canada). During an investigation of phytochemical changes in strawberry fruit during ripening, quantitative changes in >6000 identified metabolites could be recorded simultaneously. This enabled the identification and monitoring of >20 known metabolic pathways associated with strawberry ripening. Phenomenome's metabolome data is exported to a proprietary bioinformatics and visualization system that organizes plant genotypes by metabolite expression and thereby allows direct correlation of changes in metabolism with changes in gene expression.

Integrating metabolomics, transcriptomics and proteomics

At present, capturing metabolic information by spectral monitoring of a wide range of metabolic pathways and correlating it with changes in the transcriptome and proteome still represents an intriguing challenge. This is particularly true if metabolic profiling is to be applied to functional analyses in humans. However, with current estimates of >500 human diseases with direct defects in metabolism, and with complex pathologies such as cancer and inflammation also known to involve pronounced metabolic changes, metabolic profiling clearly represents a worthwhile pursuit.

The metabolite analyses of mammalian biological fluids and tissues carried out at Beyond Genomics (Waltham, MA, USA) are based on the use of parallel NMR, LC–MS and GC–MS platforms targeting specific metabolite classes including, but not limited to, lipids, steroids, eicosanoids and bile acids. This approach was applied to a study on transgenic mice that overexpress the human ApoE3 gene; these mice serve as a model of atherosclerosis and coronary artery disease.

Pattern recognition algorithms applied to spectroscopic analyses of plasma from wildtype and transgenic mice enabled the rapid identification of significant metabolic differences. Specifically, certain triglycerides were elevated in the ApoE3 transgenic mice relative to wildtype strains, whereas lysophosphatidylcholine was decreased.

These findings imply a role for metabolic profiling in the discovery of disease biomarkers or biomarker fingerprints that could have value in clinical settings. Correlating such metabolome data with proteomic and genomic studies has enabled comprehensive annotation of the metabolic pathways implicated in apoE3-mediated atherosclerosis.

Paradigm Genetics (Research Triangle Park, NY, USA) is also exploiting metabolic profiling in functional analyses of the many novel plant and fungal genes they have identified from their FunctionFinder™ platform (http://www.paradigmgenetics.com). The company's metabolome analyses and pattern recognition strategy use parallel chromatographic-MS platforms that enable tissues and cell cultures to be characterized according to >600 identified metabolites. Integration with corresponding genomic and proteomic data is made possible by a comprehensive bioinformatics system currently being co-developed by Paradigm Genetics and LION Bioscience AG (Heidelberg, Germany). In an overview of the Paradigm Genetics approach, John Hamer illustrated a chemical genetic study on herbicide action that, in brief summary, demonstrated that the integration of both gene expression data and metabolic profile data were required to reproducibly cluster, by principal components analysis,

1359-6446/02/$ – see front matter PII: S1359-6446(02)02214-6

Arabidopsis phenotypes according to the mode of action of 18 different applied herbicides. Data presented by Hamer on the effect of chemical inhibition of the nuclear factor-κB (NF-κB) pathway on gene expression and the metabolic profiles of tumour necrosis factor–α (TNF-α)-stimulated human umbilical vein endothelial cells supported the need for integration of such data.

The requirement for integrating genomic and metabolome data to effectively establish phenotype clustering in the experiments mentioned here reflects, at least in part, the well-recognized challenges in using gene expression data to explain complex biochemical networks. A major disadvantage of gene expression clustering, for example, is that it can miss relationships between pathway genes if they are regulated to differing degrees. This was an issue addressed by Eberhard Voit (Medical University of South Carolina, Charleston, SC, USA) in an excellent overview of the contribution of the mathematically based biochemical systems theory (BST) [2], in understanding the regulation of metabolic pathways.

As an example, Voit pointed out that in a study on glycolysis in heat-shocked yeast the genes for glucose transport and phosphorylation are upregulated 5–20 fold, and yet phosphofructokinase expression is essentially unaltered. Clearly, the clustering of genes by relative expression would fail, in this instance, to reveal pathway correlations between phosphofructokinase and hexokinase. By contrast, BST was able to demonstrate that the observed, and somewhat non-intuitive, gene expression profile satisfied the primary metabolic goals of increased ATP, trehalose and NADPH production in heat-shocked yeast. It is anticipated that modeling metabolic profiling data by BST will add meaning and interpretative value to corresponding gene expression data.

Metabolic networks, of course, cannot be described mathematically in the same detail as physicochemical or engineering processes. As such, new approaches that can apply experimentally or theoretically derived constraints to such modeling need to be developed. The reader is referred to http://geneticcircuits.ucsd.edu for detailed discussions on this topic and also for a look at *in silico* representations of *Escherichia coli* and human red blood cells.

Databases

There was extensive conference discussion on integrated pathway–genome databases (PGDBs) that have been developed to describe the genome of organisms and their predicted metabolome [3]. Talks on these PGDBs focused primarily on the development of tools for interpreting large-scale gene, protein, and metabolite expression data to infer metabolic networks and tools for visualizing and computationally analyzing metabolic networks. The Pathway/Genome Navigator of SRI International (Menlo Park, CA, USA), for example, enables the diagrammatic description of pathways, chromosomes and operons. Its analysis operations allow pathway visualizations of gene expression data and global comparisons of metabolic networks. Such pathway visualizations enable the investigator to rapidly identify pathways that are switched on or off under the conditions of a gene array experiment.

At present PGDBs are also being promoted as tools to aid the discovery and development of antimicrobial (and by extension, other pharmaceutical) therapies. A total of ten PGDBs representing medicinally relevant microbial genomes are available through SRI at http://www.ecocyc.org.

The National Center for Genome Resources (Santa Fe, NM, USA) is building upon its PathDB project (www.ncgr.org/software/pathdb) to create a database and toolkit for interpreting gene expression and metabolite data through an integration system (ISYS) that supports synchronization with other publicly available PGDBs.

An intriguing complement to the databases previously discussed will be the Dictionary of the Human Metabolome project sponsored by Paradigm Genetics. This will serve as a central repository for all publicly available information (and proprietary Paradigm Genetic information) on small-molecule human metabolites. It is estimated that the number of entries will total >10,000 metabolites!

References

1 Fiehn, O. (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Compar. Funct. Genom.* 2, 155–168

2 Voit, E.O. and Radivoyetich, T. (2000) Biochemical systems analysis of genome-wide expression data. *Bioinformatics* 16, 1023–1037

3 Karp, P.D. *et al.* (1999) Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol.* 17, 275–281