predictive chemoinformatics

applications to the pharmaceutical industry

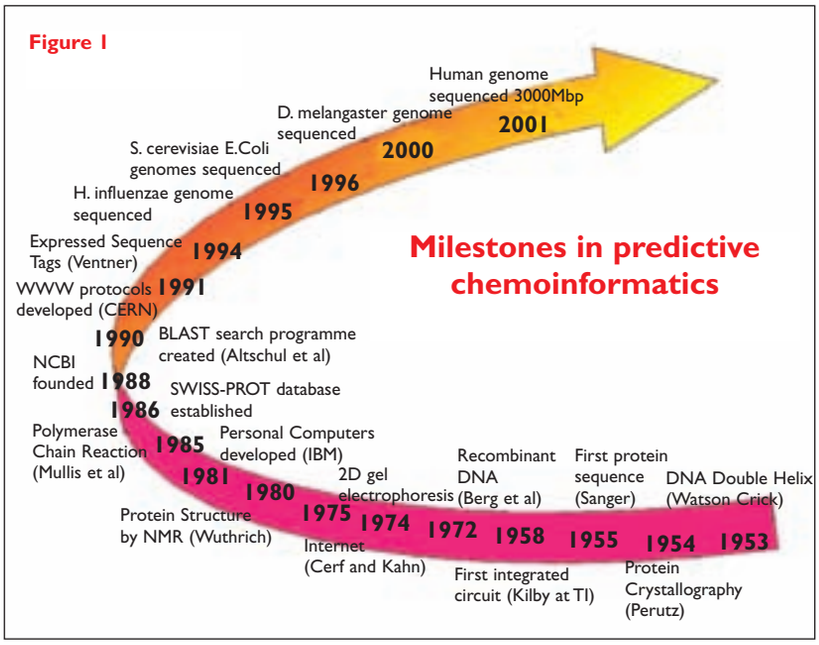
While significant advances in chemoinformatics present tremendous opportunities to improve human health, the future of chemoinformatics in the pharmaceutical industry is not without significant challenges.

By Dr Leslie J. Browne and Laurie L. Taylor

Chemoinformatics is the result of collective advances in chemistry, biology, computer sciences and statistics and refers to the electronic tools, methods and data used for analysis and predictive computation of drug effects on complex biological processes (**Figure 1**). Scientific milestones over the last 50 years which have contributed to the evolution of predictive chemoinformatics include the development of the DNA double helix model by Watson and Crick (1953); sequencing of the first protein-bovine insulin by Sanger (1955); protein crystallography by Perutz (1954); the first integrated circuit by Kilby at Texas Instruments (1958); recombinant DNA technology by Berg et al (1972); conception of the Internet by Cerf and Kahn (1974); development of 2-D gel electrophoresis (1975); identification of protein structure NMR by Wuthrich (1980); creation of the first personal computers by IBM (1981); polymerase chain reaction technology by Mullis et al (1985); creation of the SWISS-PROT database (1986); the founding of the NCBI (1988); creation of BLAST by Altschul (1990); development of WWW protocols by the CERN (1991); identification and significance of ESTs by Ventner (1991); sequencing of the entire genomes of H. influenzae, S.cerevisiae (12Mb), E. Coli (1995-1996); D melangaster (180Mb) in 2000; and the human genome 3000Mbp in 2001.

Advances in chemoinformatics related to genomics, proteomics and computer-assisted chemical modelling hold tremendous promise to improve human health. For pharmaceutical research and development, chemoinformatics provides the tools to compare expression of genes and proteins as well as complex signalling processes in disease and normal tissues and impacts concretely on selection of therapeutic targets (**Figure 2**). Differential gene and protein expression profiles related to a disease state (eg cancer) promise to help fine-tune diagnoses and improve the accuracy of prognostic indicators to best serve individual patient needs. Chemists can now design compounds with improved drug-like qualities through computerised structure/activity

Figure 1



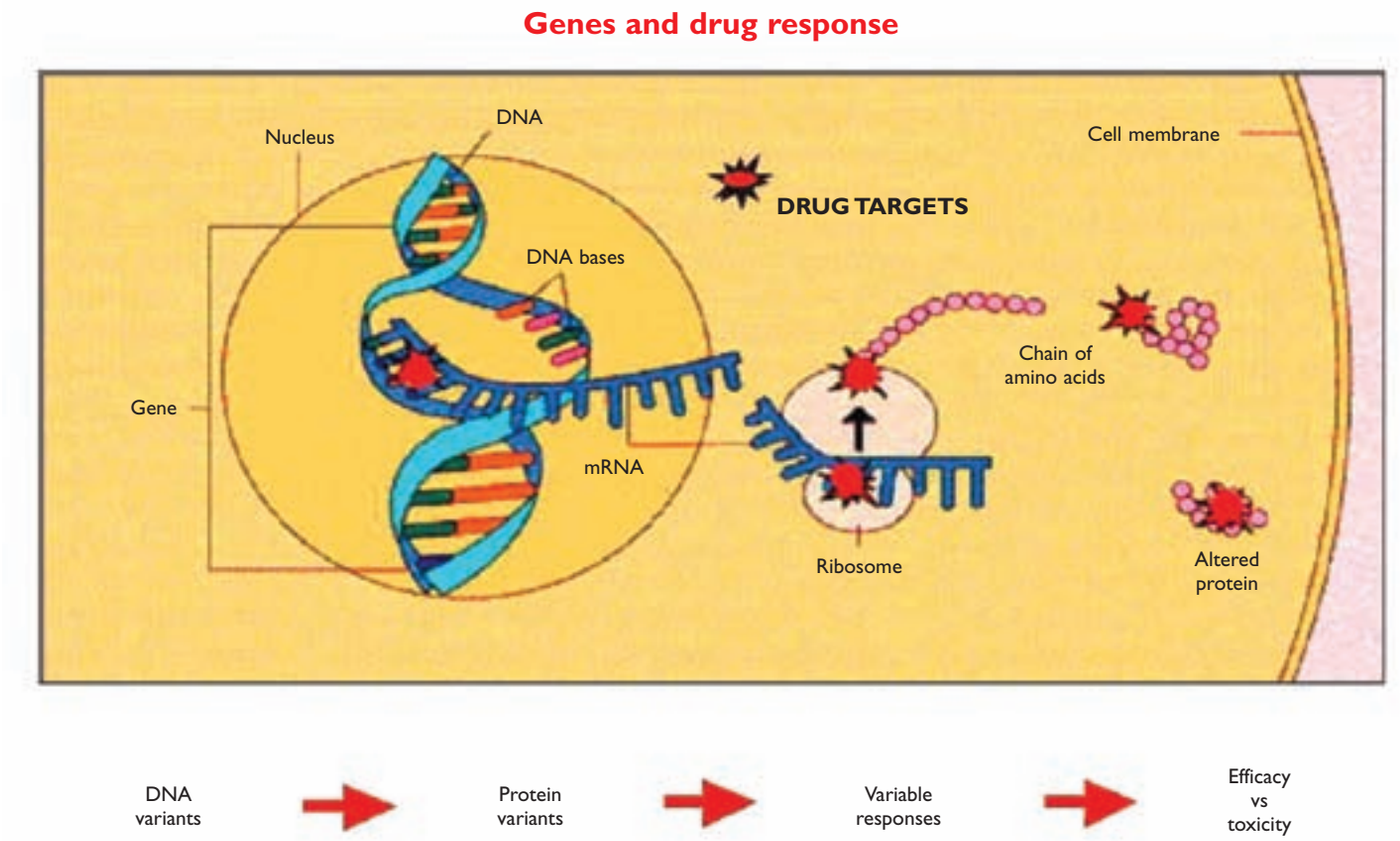


Figure 2
Cellular checkpoints for therapeutic intervention

modelling – often reducing the number of compounds tested, compared with conventional trial-and-error methods. Drugs themselves affect expression of a wide variety of genes and proteins, and individual patient responses to drugs differ in metabolism and toxicity.

Pharmaceutical companies are highly motivated to reduce the discovery-to-market time and cost. Increased R&D dollars dedicated to the business of discovering new therapeutics have not resulted in a correspondingly increased number of successful drugs on the market. The pre-market failure rate of drug candidates has been measured and remeasured from varying perspectives but always leads to the unavoidable conclusion that the process is inefficient. More than 50% of failures are due to lack of efficacy or unexpected animal toxicity. (**Figure 3**). It now costs an average of $800 million to bring a new product to market[1]. This includes, of course, the cost of the numerous failures and their consumption of R&D dollars – a cost that is passed on to the consumer. Since the failure rate is so high – about 1 in 10 drug candidates survives from initiation of clinical evaluation to market launch (**Figure 4**) – even a modest improvement to 1 in 5 halves the development cost. Many drug failures are the result of 'off target' activity, ie poor side effect profiles that offset the potential therapeutic effect. Structure-based design algorithms and structure-activity data of existing bioactive compounds facilitate the design of new compounds with the critical 'drug-like' qualities, in addition to potency and efficacy at the therapeutic target: a necessity for successful pre-clinical and clinical development. The ability to project the *in vitro* effects of a candidate drug into predictive models of broader *in vivo* systemic effects earlier in the discovery process, will benefit the industry by reducing failure rates, the developer by reducing costs and the consumer by helping get better drugs to the market.

Despite the expense and time committed to drug development, approved drugs have frequently been withdrawn from the market due to

severe adverse drug reactions (ADR). Between October 1997 and September 1998, a number of FDA-approved drugs were withdrawn, but not before being prescribed to 20 million patients in the US alone[2].

Importantly, the side effects that resulted in the ADR might have been measured and potentially designed out of the drug candidates had there been a means of identifying in advance the full spectrum of its potential side effects. While additional pre-market animal and human evaluation might decrease the number of drugs withdrawn from the market, the additional cost would be significant. In contrast, new chemoinformatics tools can be used to identify potential liabilities and benefits much earlier in the discovery process. Identifying and eliminating likely failures earlier permits efforts to be focused on higher quality compounds, resulting in more efficacious drugs produced at lower overall cost.

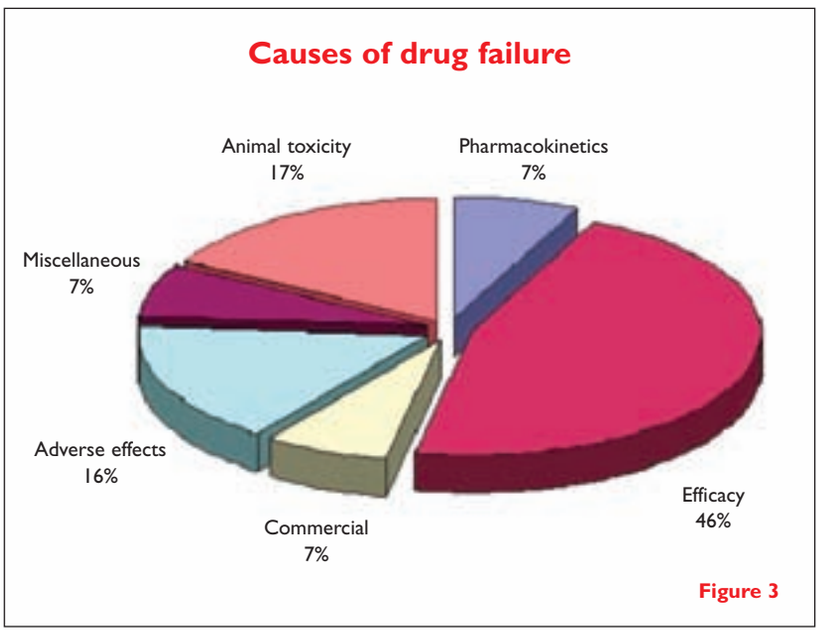


Figure 3

Chemogenomics applied to the discovery of new therapeutic agents

Overview: While the physiological response of animals to drug treatment is the mainstay of efficacy and safety evaluation for drug development, the nature of conventional pre-clinical evaluation methods means that only a few important physiological parameters can be assessed at a time. The new options provided by genomics and proteomics is to assess broadly the effect of a compound on the system as a whole by looking at the transcriptome and the proteome. As the tools are developed, it will be possible to look not only at mRNA in high throughput but also the resultant individual protein, its conformation and its phosphorylation state, etc to get the fullest possible picture of what is happening at the molecular level in response to compound treatment.

Chemogenomics – or 'pharmacology with genomics tools' – combines the strengths of traditional pharmacology and the mechanistic approach to drug discovery. Since an intact biological system is the focus of the evaluation, it is contextually information-rich. The effects of a compound are examined in the context of other biological processes it affects in addition to the target for which it was designed. For example, this approach allows for compensatory and regulatory mechanisms to influence the phenotypic outcome, as measured by the genomic response of the system. Furthermore, since the analysis views all, or at least a large proportion of, induced genomic changes within an organism, an

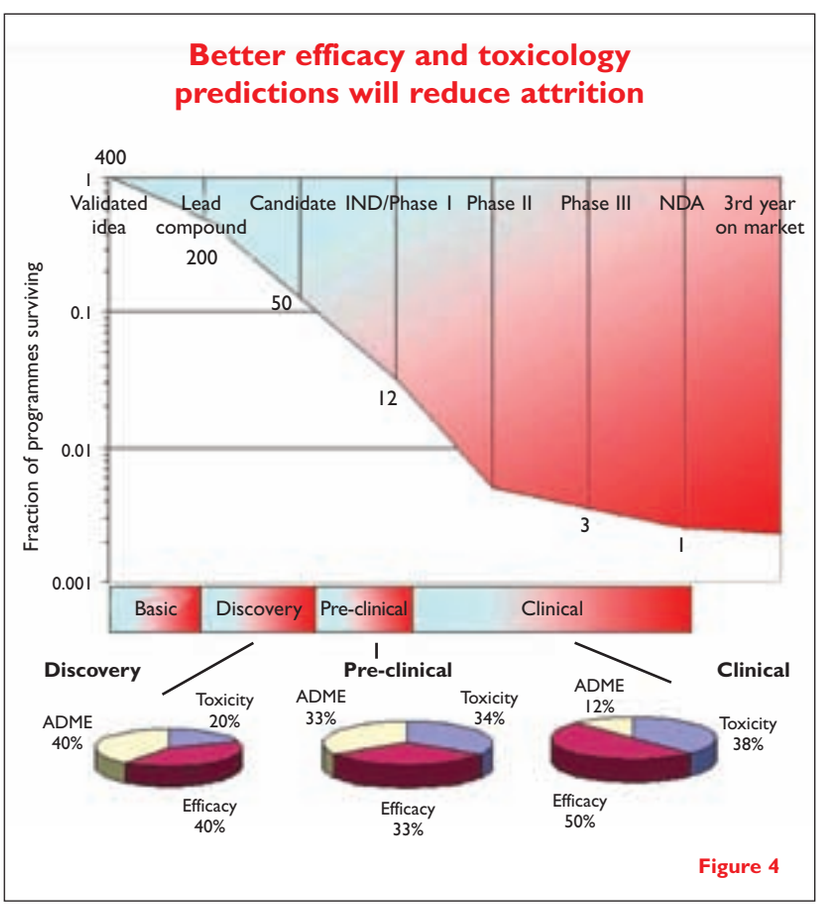
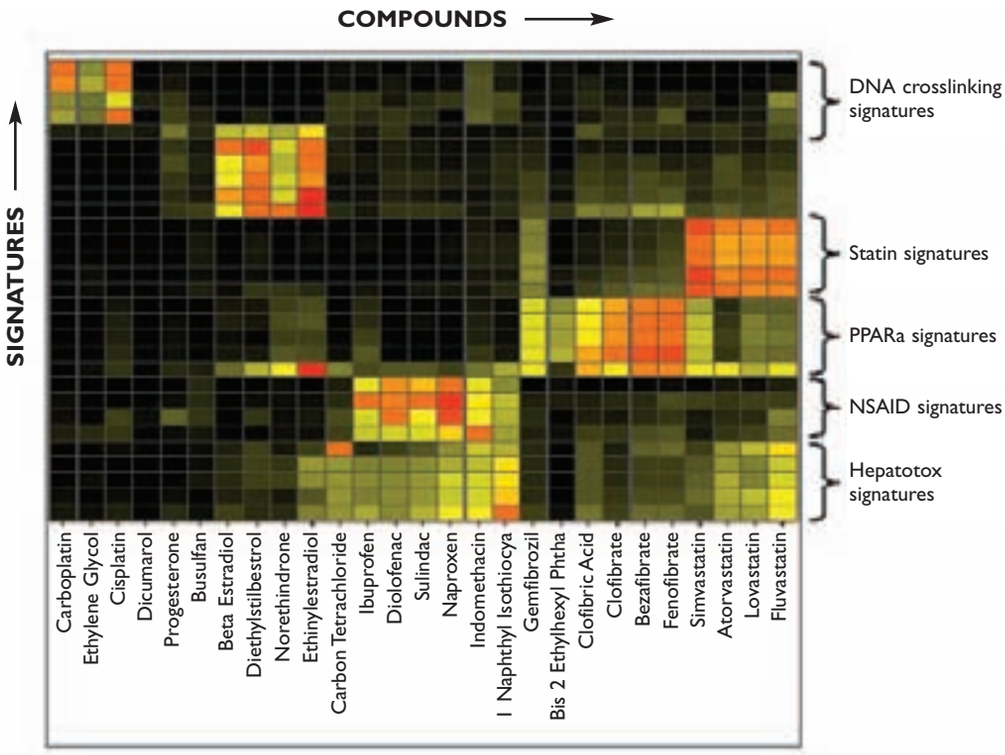


Figure 4

Figure 5
Gene expression class signatures



improved understanding of the breadth of compound action on target-related genes, as well as unrelated genes, is possible.

While the immediate promise of chemogenomics is to increase the efficiency of drug discovery and development by eliminating failures early, it offers the potential to improve drug quality by treating disease pathophysiology rather than symptoms. The use of gene expression profiles involving multiple genes, whose misregulation have been implicated in a disease state, represents a novel, although unproven, approach to drug discovery. Since many of the most important unmet medical needs are polygenic diseases, in which several genes contribute to the disease in a complex way, a drug discovery approach that identifies modulators of multiple genes in concert has the potential to uncover treatments for the underlying cause of the disease.

Chemogenomics, then, is the interaction of chemical compounds and living systems in terms of the induced genomic response. For example, instead of examining only a few changes in mRNA expression in a single experiment, an entire transcriptional state (10,000 or more changes in mRNA levels) of an organism may be analysed using microarray chips. The challenge is to interpret what these changes mean and how to use the information effectively to make key drug discovery decisions. One approach is to characterise the effects of existing, well-understood drugs in chemogenomic terms and translate the knowledge to prediction and interpretation of the effects of new drug candidates.

Applications: The key to the successful application of chemogenomics is interpreting the enormous amount of information obtained from each experiment. Although it is seductive to try to analyse the genomic profile of individual compounds, understanding the biology of classes of well-known compounds in genomic terms offers an improved platform on which to base understanding the profiles of new drug candidates.

Chemogenomic profiles: Diverse ways of extracting information from chemogenomic data are being developed using a variety of statistical and computational approaches. Certainly one approach to achieving this objective is to collect the genomic and pharmacological response of the target tissue or cell type to treatment with a chemical compound. Each compound profile is the compound's own signature of transcriptional and molecular pharmacological effects (**Figure 5**). While this has utility, extending the analysis to

look at compound families, eg related by a common therapeutic use, mechanism or by structural similarity, has even more value since it provides a means of extracting the biomarkers associated with a class effect, eg the therapeutic signature, as well as compound specific effects, eg side-effect signatures. The total activity profile of a compound comprises multiple signatures representing its structure, on- and off-target mechanistic effects, side effects and therapeutic effects. While such profiles clearly exist, the challenge is how to identify them and make use of them in making decisions that improve the quality of drug discovery and development. The principle challenge for chemo- and bioinformatics is to develop computational methods capable of deciphering information contained in chemogenomic profiles and effectively displaying the results for more effective 'next-step' decisions in drug candidate selection and development.

Toxicogenomics: A series of compounds known to cause a particular toxicity are employed in an *in vivo* or *in vitro* experiment to induce genomic changes, eg transcriptional, to derive the genomic profile. The resulting set of biomarkers, or signature, reflects genomic changes that represent the compound-induced phenotype. The signature contained in the chemogenomic profile of a drug candidate may indicate that the candidate possesses similar properties to compound classes from whose profiles a specific class signature was derived. When the signature is derived from known toxicants, it may be useful in predictive toxicology. This type of application, known as toxicogenomics – toxicology with genomics tools – is becoming a generally accepted approach in the pharmaceutical industry to identifying compounds with potential safety problems before they are evaluated in costly regulatory toxicology studies[3-5].

To date, investigations in toxicogenomics frequently involve *in vivo* studies in rats since this species is commonly employed as the primary model by the pharmaceutical industry. The presence of biomarkers of toxicity can alert investigators to potential overt toxicity, such as necrosis or organ pathology. By comparing an investigational compound's chemogenomic profile with known signatures of toxicity, it is possible to:

- Match drug candidate profiles against known toxicity profiles.
- Compare compounds by degree of toxicity.
- Anticipate compound-induced pathology.
- Elucidate the mechanism of toxicity in target organs.

***In vivo* and *in vitro* gene expression:** Gene expression profiles can be measured in both *in vivo* and *in vitro* experiments. The *in vivo* approach has the advantage for toxicogenomics that the pathological outcome can be measured in the intact animal and correlated to the genomic response. The principle disadvantages are the cost of whole animal experiments and the requirement for relatively large quantities of compound.

In vitro systems include the use of whole organs, tissue slices, primary cells, conditionally immortalised and immortalised cell lines and generally requires less test compound than *in vivo* test systems. The divergence of an *in vitro* surrogate system from an *in vivo* system has to be carefully considered depending on the application. While whole organ preparations and tissue slices may correlate better with *in vivo* models physiologically and metabolically, reproducibility and availability represent significant hurdles especially for high throughput gene expression analysis.

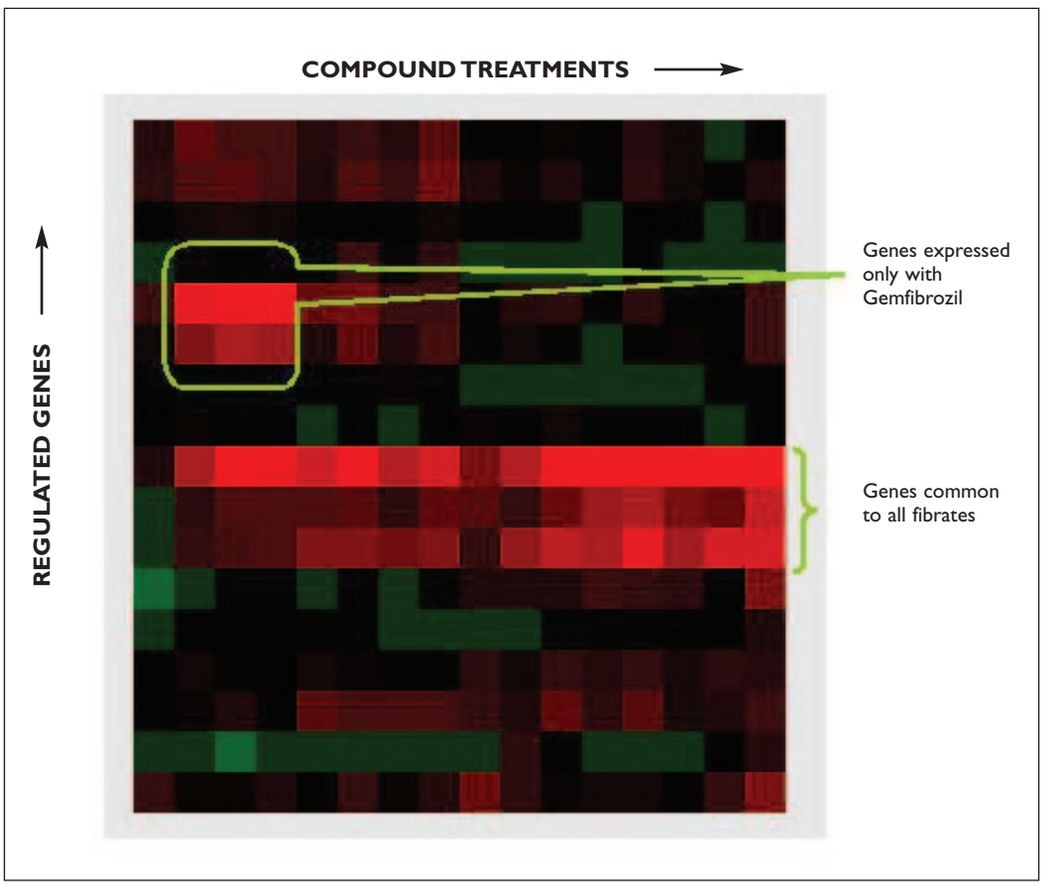
The principle disadvantages of *in vitro* cell systems are the lack of cellular heterogeneity and integrity of the whole organs from which they were derived and the fact they suffer almost universally from the lack of full metabolic capability. We have found *in vitro* experiments to be particularly useful for mechanistic studies when a phenotypic endpoint can often be measured (eg cell death as the functional endpoint of apoptotic gene expression).

For mechanism of action studies where metabolic competency is less important, *in vitro* cell lines offer considerable advantages:

- Reduced compound need – less than 100mg may be adequate for *in vitro* work.
- Ready access to human cellular systems – may be preferable to non-human mammals.
- Faster turnaround – 24 hours treatments may be sufficient.
- Higher throughput – cell culture amenable to miniaturisation.

Drug discovery: In the same way toxicity signatures are derived, biomarkers for other compound effects, eg mechanism of action, can be deduced. By assessing the effects of a broad range of chemical compounds, chemogenomics has yielded signatures for mechanistic, structural and therapeutic classes and even subtle off-target effects. Recent studies performed at Iconix and MDS Pharma Services demonstrate the predictive power of this

Figure 6
Genes differentially induced by Gemfibrozil



approach. By comparing the genomic profile for Gemfibrozil to those of the class of fibrates to which it belongs, it has been possible to identify the genomic responses uncommon to the class, but specific to Gemfibrozil. This approach identified a pathway regulated by Gemfibrozil but not by other members of the fibrate class, and may account for the effects of Gemfibrozil that differentiate it from other fibrates[6] (**Figure 6**).

In summary, the chemogenomics advantage applied to understanding the broad spectrum of effects of chemical compounds on a living system is to characterise, evaluate and prioritise compounds for further optimisation or, if necessary, elimination from further consideration. Future directions include extending our reach to develop surrogate genomic and proteomic markers for drug optimisation and design. Chemogenomic studies of drug effects on differential gene expression as well as post-translational protein expression will help address current challenges in the integration of our understanding of genomics and proteomics, eg correlation of gene to protein and the complexity of intracellular signalling, processing and regulatory pathways in health and disease.

Future considerations

The future of chemoinformatics in the pharmaceutical industry is not without significant challenges. Despite multiple public-access genomics, proteomics, sequence and functional pathway search engines and databases, large amounts of data reside in proprietary databases with restricted access and proprietary search algorithms. Collation, analysis and meaningful interpretation of disparate biological data present an arduous challenge to computer scientists and biologists alike. This fact is evidenced by the growing number of bioinformatics services and products offering database search, analysis and reporting tools as well as proprietary databases populated with gene, protein, pharmacological and molecular descriptors. In addition, many databases are populated with data generated under a variety of experimental conditions with varying degrees of accuracy and/or relevance to a given biological model or target. Analysis of complex biological systems from gene to the organism level will necessarily require development of mathematical and statistical algorithms to analyse very large data sets through close collaboration between biologist and mathematician.

Clearly, *in silico* predictive modelling, based on compound structure and target molecule motifs using chemoinformatics alone, will not obviate the need for validation through *in vitro* and *in vivo* experimentation. Similarly, *in vitro* approaches to studying key cellular disease pathways and drug effects on molecular targets afford critical but limited 'views' of complex interactions at any given point in time. Through *in silico* mathematical and computer analysis of thousands of these *in vitro* 'views', complex molecular interactions may be displayed simultaneously providing the ability to understand the effect of a single entity on a complex system.

In conclusion, key future applications of predictive chemoinformatics for the pharmaceutical industry are:

1 Identification, validation, testing and functional annotation of new protein and gene drug targets in disease and health.

2 Ensuring a healthy pipeline of new, improved drug leads for development.

3 Expanded understanding of complex biological systems.

4 Optimisation of structure-related activities for new as well as known compounds to reduce the number of compounds tested and therefore the cost and time to market.

5 Empowering clinicians and patients with informatics tools, such as differential gene and protein expression profiles as a function of patient age, health, family history and environment, to effectively tailor individual drug treatment regimes for improved patient care. DDW

Dr Leslie Browne *is currently Chief Operating Officer for Iconix, a company he joined in October 2001 from Gene Trace Systems where he held the same position. Before that Dr Browne spent more than a decade at Berlex/Schering AG, most recently as Corporate Vice-President, Berlex Laboratories, Inc and President of Schering Berlin Venture Corporation. Prior to this he was Vice-President, Head of Discovery Research, at Berlex Biosciences, having responsibility for drug discovery, including cell and molecular biology, protein chemistry, screening, medicinal chemistry and molecular and animal pharmacology. Before Berlex, Dr Browne was with Ciba-Geigy, where he invented Fadrazole, the first marketed non-steroidal aromatase inhibitor for the treatment of estrogen-dependent breast cancer. He also managed the cardiovascular research programme at Ciba-Geigy Ltd in Basle, where one of the group's achievements was the discovery of Diovan, the second angiotensin ll antagonist ever to be marketed. Dr Browne received his PhD from the University of Michigan, with a postdoctoral fellowship at Harvard University with the Nobel laureate Professor R.B. Woodward.*

Laurie Taylor *has more than 20 years of experience in pharmaceutical research, with special emphasis in lead discovery and pharmacology services. Ms Taylor joined MDS Panlabs (now MDS Pharma Services) in 1990 as an associate scientist in assay development. Today, she serves as Director, Lead Discovery, and is responsible for negotiating and implementing client contracts, identifying and developing new product areas and guiding the business unit's sales efforts. Ms Taylor's background includes management and scientific positions with Signal Pharmaceuticals and the departments of pharmacology and pathology at the University of Washington, Seattle. She has authored or co-authored numerous scientific articles and abstracts for peer-reviewed publications, including Tetrahedron, the Journal of Medicinal Chemistry and the Journal of the American Chemical Society. She has also presented papers and posters before Annual Conference on the Biotechnology of Microbial Products, the Annual Meeting of the American Society for Cell Biology, the Organic Chemistry Symposium and other associations. Ms Taylor earned her BA in zoology from the University of Washington. She is also a certified electron microscopist, San Joaquin Delta College.*

References

1 Outlook 2002. Tufts Center for the Study of Drug Development.

2 Lasser, KE et al. JAMA 287, 2215-2220 (2002).

3 Waring, JF et al. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. Toxicol. Appl. Pharmacol. 175, 28-42 (2001).

4 Waring, JF et al. Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity. Toxicol. Lett. 120, 359-368 (2001).

5 Furness, LM et al. Chemogenomics for Predictive Drug Assessment. In press. Toxicogenomics, Springer-Verlag (2002).

6 Browne, LJ, Furness, LM, Natsoulis, G, Pearson, C and Jarnagin, K. Chemogenomics: pharmacology with genomics tools, Targets, 1, 59-65 (2002).

Other references

Hansch, C, Hoekman, D, Leo, A, Weininger, D, Selassie, C. Chemo-Bioinformatics: Comparative QSAR at the Interface between Chemistry and Biology. Chem. Rev. 2002, vol 102, pp783-812, American Chemical Society.

Kitano, H. Systems Biology: A brief Overview. Science, March 1, 2002, Vol 295, pp1662-1664.

Noble, Denis. The rise of computational biology. Nature, Volume 3, June 2002, pp 460-463.

Richon, Allen B. A Short History of Bioinformatics. Network Science, www.netsciorg/Science/bioinform/feature06.html August 2002.