

Techniques: Application of systems biology to absorption, distribution, metabolism, excretion and toxicity

Sean Ekins, Yuri Nikolsky and Tatiana Nikolskaya

GeneGo, 500 Renaissance Drive, Suite 106, St Joseph, MI 49085, USA

It is widely recognized that either predicting or determining the absorption, distribution, metabolism, excretion and toxicity (ADME/Tox) properties of molecules helps to prevent the failure of some compounds before they reach the clinic. Consequently, there has been considerable research into developing better *in silico*, *in vitro* and *in vivo* methods and models. Toxicogenomics, proteomics, metabonomics and pharmacogenomics represent the latest experimental approaches that can be combined with high-throughput molecular screening of targets to provide a view of the complete biological system that is modulated by a compound. The functional interpretation and relevance of these complex multidimensional data to the phenotype observed in humans is the focus of current research in toxicology. Multiple content databases, data mining and predictive modeling algorithms, visualization tools, and high-throughput data-analysis solutions are being integrated to form systems-ADME/Tox. In this review, we focus on the most recent advances and applications in this area.

An introduction to systems-ADME/Tox

In our quest to cure important diseases we encounter the complexity of the whole organism. At the molecular level, a coordinated system of transporters, channels, receptors and enzymes act as gatekeepers to foreign molecules; this system affects the absorption, distribution, metabolism, excretion and toxicity (ADME/Tox) of a molecule in humans (Figure 1). Understanding the interactions between small molecules and their molecular targets should improve our ability to predict the toxic consequences that are responsible for the withdrawal of many marketed drugs and late-stage failures of drugs in development [1].

The focus is now on preclinical ADME/Tox studies, but the complexities of different model systems mean that better predictive approaches are needed [2], preferably based purely on molecular structure [3,4]. Accurate predictions of toxicity mechanisms are complicated because the whole organism comprises thousands of endobiotic and xenobiotic molecules that interact in different cellular organelles and tissues. Species

differences in protein expression and ligand specificity should also be considered. Methods are needed that account for the complexity of the biological data and enable prediction of the complete system, including metabolic, regulatory, signaling and transport processes [5,6]. Systems biology uses the relationships between all elements rather than approaching them separately, and attempts to unite biological fields [7]. This approach can be taken from either the 'top down' (by using a conceptual framework to integrate data) or the 'bottom up' (by combining individually modeled biochemical processes) [8]. Interpreting ADME/Tox in the systems context might improve our understanding and, ultimately, predictions of toxicity. Potentially, the perturbing effect of a molecule on the complete biological system can be observed either experimentally, using high-throughput screening against many proteins, or theoretically, using computational models [9]. This should enable an understanding of the effects of binding to multiple proteins simultaneously. The iterative approach of data generation and modeling cycles can create dynamic hypotheses with broader applications than static models. Systems-ADME/Tox requires the collection of high-throughput data, including global gene-expression, protein content and metabolic profiles for the same samples plus genetic, clinical and phenotypic data. To date, empirical data have been used to build computational models and 'score' many virtual molecules for enzyme inhibition [10]. These computational predictions require multidimensional analysis to target affinity and improve the efficiency of lead selection [11]. Computational predictions might also be used as input parameters for systems-biology models to quantitatively predict clearance or molecule disposition.

In this review, we describe the integration and relevant applications of several types of biological data, databases, computational algorithms and predictive methods. We illustrate that the systems-ADME/Tox approach can use data from all experiments and computational methods to provide a deeper understanding of the effects of xenobiotic and endobiotic molecules on the recursive human ADME/Tox properties.

OMICS data and their integration

Presently, in addition to high-throughput screening assays for binding to receptors and other proteins of interest, much data generated in at least four 'OMICS'

Corresponding author: Ekins, S. (sean@gene.go.com).

Available online 3 March 2005

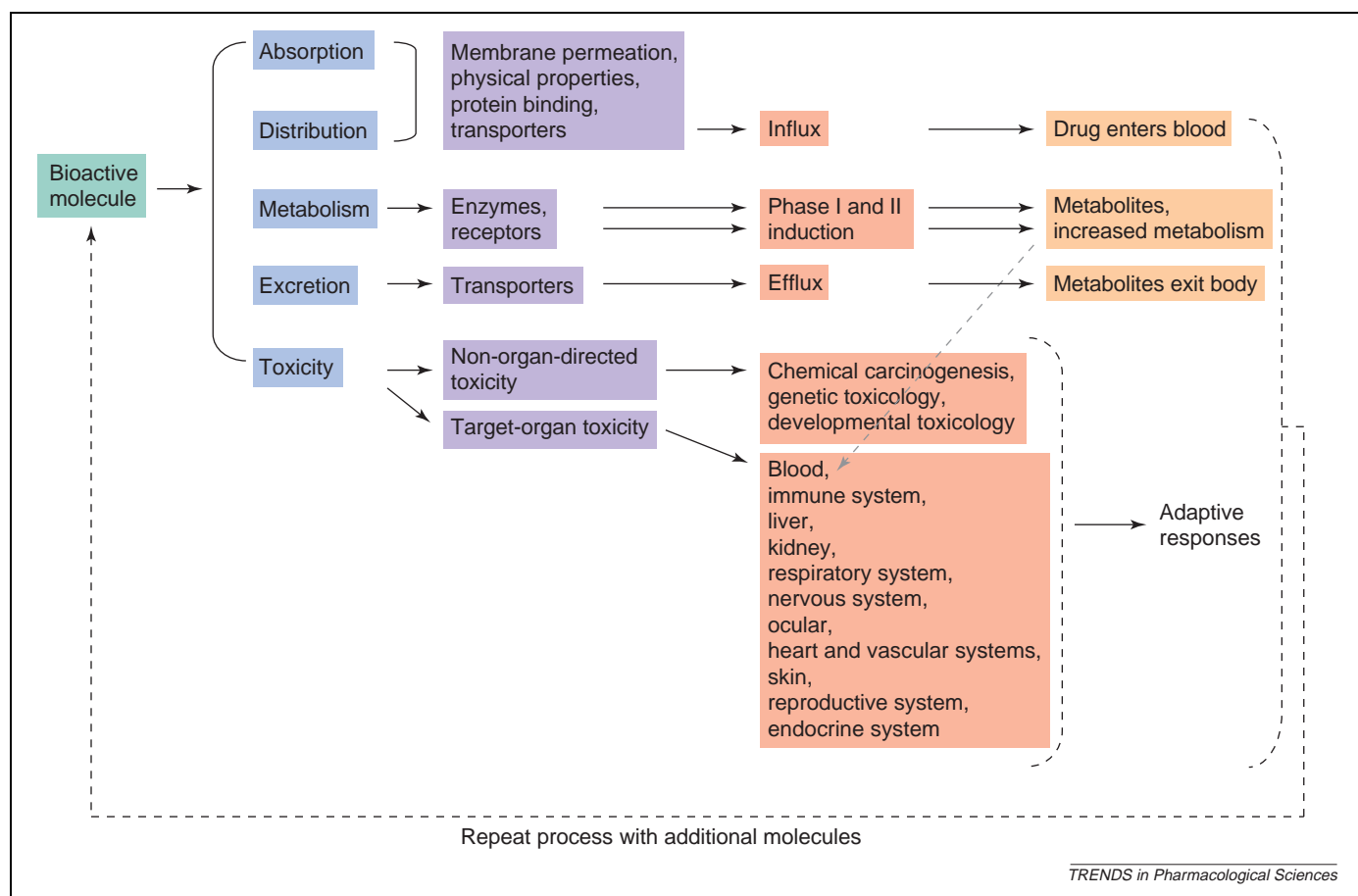


Figure 1. The iterative ADME/Tox optimization process. This figure demonstrates that a bioactive molecule is required to possess many favorable ADME/Tox properties before it can become a drug, and indicates the multidimensional nature of drug discovery. The proteins and endpoint associated with each ADME/Tox function are outlined. Adaptive responses represent the transcriptional and post-transcriptional effects following a toxic insult. Solid arrows represent the links between ADME/Tox properties, functions and endpoints. The grey dashed line represents reactive metabolites that can cause toxicity.

areas are related to the interaction of a drug with proteins involved in ADME/Tox. These areas are toxicogenomics, proteomics, metabonomics and pharmacogenomics.

Toxicogenomics

Toxicogenomics deals with global changes in gene expression in response to either a drug or a toxin, and is usually measured using microarrays (mRNA transcription). Drug discovery provides large numbers of molecules and data regarding the binding of these molecules to multiple proteins [12] and the quantity of toxicology information used in decision making during the drug discovery process has increased because of toxicogenomics. Generally, these studies can be divided into predictive toxicology and mechanism-based risk assessment [13]. Changes in the global patterns of gene expression in animals and in cells in response to drugs given at different doses and time-points identify 'signature' genes that can be used as predictors of toxicity [13] in humans.

In addition to predictive toxicology, the second toxicogenomics area to use microarrays is mechanistic toxicology. This is aimed at understanding the biochemical and biological responses in a particular type of toxicity and is important for compound risk assessment. Most biologically active compounds affect the function of cells at many

levels, including nuclear hormone receptors (NHRs), which regulate the transcription of genes that encode cytochrome P450 enzymes (CYPs), phase II enzymes, transporters and other genes [14]. Microarray data have been used to identify a detailed mechanism of benzene-induced hematotoxicity and leukemogenicity that involves inhibition of p53-mediated apoptosis and oxidative stress [15] and that can be illustrated using both static (gene ontology) (Figure 2) and dynamic models using the software that is described later. Normally, carcinogenicity is evaluated in mouse bioassays that take up to 2 years, but microarray methods have been used to identify gene markers that speed up the identification of carcinogens. In an *in vivo* toxicology study, cDNA microarray data have been generated following 5-day, repeat-dose treatment of rats with several prototype rodent genotoxic and non-genotoxic carcinogens, and two non-carcinogenic hepatotoxicants. Correlating the gene-expression data from this study with the known carcinogenic potential of these compounds has identified transforming growth factor β -stimulated clone 22 and NAD(P)H CYP450 oxidoreductase as molecular markers of non-genotoxic carcinogenicity in rodents [16].

Proteomics

Assessing the effects of a compound on the activity and concentration of proteins can provide results that

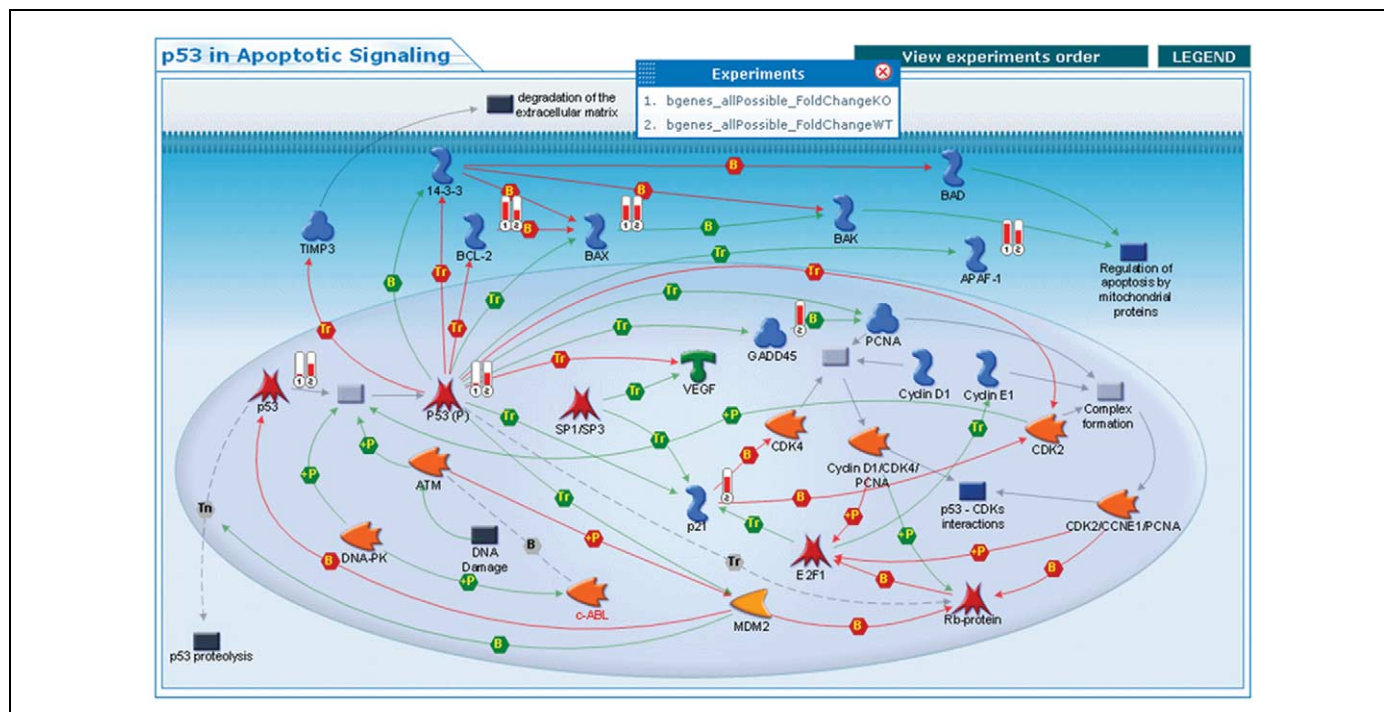


Figure 2. Visualizing OMICS data in networks. Visualization and network analysis of differential-expression data induced by a 2-week exposure to benzene indicates that hematotoxicity and leukemogenicity occurs via inhibition of p53-mediated apoptosis and oxidative stress [15]. Microarray-data mapping on the p53 apoptotic signaling map in MetaCore™ (<http://www.genego.com>) demonstrates the difference in expression of cell-cycle pathways between p53-knockout mice (1) and wild-type mice (2) (the height of the red bars indicates relative differences in expression). The colored, differently shaped nodes on the network represent enzymes (orange), transcription factors (red) and binding proteins (blue). The small colored hexagons encode the interactions between two connected nodes [e.g. binding (B), transcriptional regulation (Tr) and phosphorylation (+P)]. The green lines indicate a positive effect whereas the red lines indicate a negative effect.

complement gene-expression data and are more consistent with the overall mechanism of toxicity [17]. Proteomics deals with the quantitative and qualitative measurement of protein concentrations in whole-tissue samples [18]. This is important because post-translational modification, proteolysis and other dynamic processes that cause functional changes in proteins means that the presence of a mature mRNA transcript does not always correspond to the presence of the active protein [19]. Expression-proteomics studies use two-dimensional (2D) protein electrophoresis and mass spectrometry to identify a set of proteins (hundreds to several thousand per sample) that are expressed in a particular tissue under different conditions [20–22], and they are a main method in both predictive and mechanistic toxicology studies [23]. A recent proteomics study of a proprietary lead compound that caused steatosis in rat liver identified 22 proteins in liver, many of which are involved in pathways that lead to the accumulation of acetyl-CoA and triglyceride. This indicates that steatosis is the result of perturbing the β -oxidation pathway in this tissue [24].

Metabonomics and metabolomics

Metabolites represent the endpoint of the response of an organism to a stimulus (Figure 1), and metabolic profiling is the most direct measure of physiology [25]. Metabolomics relates to the understanding of metabolic regulation and flux in cells, and metabonomics is a larger scale, systemic determination of biochemical profiles and their regulation in biofluids and tissues [26]. Both methods analyze all the low-molecular-weight molecules in cells.

The quantification and identification of every metabolite in a cellular system is difficult because of the lack of analytical techniques that are reproducible, robust and automated, the chemical heterogeneity of metabolites, the lack of automated extraction techniques, and the low resolution of existing hardware [27].

Pharmacogenomics

Pharmacogenomics uses genome-wide approaches to identify variations in the networks of genes that determine how individuals respond to a drug, and its efficacy and toxicity. During the past few decades many genes have been linked directly to the mechanisms of response and we know that 20–95% of the variability in drug response is inherited [28,29]. Examples of key genes that govern drug efficacy in humans and examples of ADME/Tox proteins that exhibit polymorphic variability to drug response [28,30] are also growing and their identification will assist drug discovery.

Integration

A few groups have integrated more than one OMICS dataset in a study. Examples include the assessment of toxicogenomics of bromobenzene in rats using both mRNA and protein, which shows a modest overlap in results and the complementary nature of the data [31]. Evaluating acetaminophen overdose in mice using microarray expression and proteomics [32] shows that protein concentrations change rapidly but RNA expression lags and there is little correlation between the two, possibly because of the small array size. In another study in mice,

microarray data and high-resolution ^1H nuclear-magnetic-resonance spectra from intact liver, tissue extracts and plasma after dosing with acetaminophen were used to identify the biochemical changes associated with hepatotoxicity. Increased hepatic glycolysis was observed, which is consistent with gene-expression data for lipid and energy metabolism [33]. The effect of carbon tetrachloride in rats has been evaluated using microarray and 2D-gel proteomics at several time points and doses [34]. Low doses of carbon tetrachloride caused the rapid upregulation of known markers of stress, early transcriptional factors and DNA-damage-control genes. These gene changes correlated with 21 out of 22 proteins, with at least one gene being differentially expressed [34]. Although it is, perhaps, too early to see the benefits of this integration in all but the few published cases [35], a comprehensive analysis of biological systems requires the integration of all the data to identify molecular markers of different toxic endpoints. The integration of metabolomics with gene-expression and proteomics data to unify metabolic regulatory networks and infer condition-specific metabolic networks [36] indicates that traditional metabolic pathways can be visualized as networks of subsystems [37].

New algorithms for integrative data analysis

After generating data by most high-throughput methods, commercially available clustering techniques [GeneSpring™ (<http://www.silicongenetics.com/>), Guided Analytic™ application (<http://www.spotfire.com/>) and Rosetta Resolver™ (<http://www.rosettahbio.com/>)] are applied to extract the toxicity signatures as a phylogenetic tree in which the branch length corresponds to the similarity between datasets [38]. Recently, several approaches have been developed that enable more advanced, functional analysis of high-throughput molecular data. These algorithms can be used to combine protein-interaction information and expression data to find condition-specific modules in protein networks. This is achieved by clustering gene-expression data and mapping the resulting clusters onto interaction networks that are obtained from independent sources [39]. Other methods include superparamagnetic clustering (see Glossary), which identifies tightly connected sets of nodes [40], simulated annealing [41], probabilistic graphical models [42] and Monte Carlo optimization [43]. Most of the research in this area relies on interaction data from yeast two-hybrid assays. Thus, analysis is limited to direct physical interactions and also contains many false positives. Information on the type and direction of interaction is, generally, not represented in such datasets.

More recently, additional developments have been used to connect interacting, differentially expressed genes in condition-specific, functional, 'signature networks' [37]. Combining comprehensive databases with powerful analytical and network-building tools has resulted in the development of integrated, high-throughput, data-mining suites such as Pathway Assist™ (<http://www.ariadnegenomics.com/>), PathArt™ (<http://jubilantbiosys.com/>), MetaCore™ (<http://www.genego.com>) and Pathways Analysis™ (<http://www.ingenuity.com/>).

Glossary

Monte Carlo optimization: Random generation of values for uncertain variables to simulate a model. In the context of networks this method is used to find highly connected nodes in networks.

Nodes: An object (genes and molecules) connected in a network.

Probabilistic graphical models: A combination of graph theory and probability theory. They represent multivariate joint probability distributions via a product of terms, each of which involves only a few variables. The structure is represented by a graph that relates variables that appear in a common term.

Simulated annealing: A technique to find a good solution to an optimization problem by trying random variations of the current solution. A worse variation is accepted as the new solution with a probability that decreases as the computation proceeds. The slower the cooling schedule or rate of decrease, the more likely the algorithm is to find an optimal or near-optimal solution.

Superparamagnetic clustering: This method assigns a 'spin' to each node and spins in a highly connected cluster fluctuate in a correlated fashion, which is used to identify these nodes.

These tools enable visualization of the global cellular mechanisms that account for differences in expression. Most use manually curated content about the physical interactions between proteins in humans, which enables different levels of cellular functionality to be captured as either maps of the current biological knowledge or custom-built interaction networks (Figure 2).

Complex computational models have been generated iteratively to simulate networks that regulate transcription and metabolism in *Escherichia coli* [44]. Researchers have also constructed minimal-cell models as a first step to multicellular organisms, which are likely to provide insights into pharmacokinetics and pharmacodynamics [45]. Although the mathematical algorithms derived for simple organisms and pathways are unlikely to be reliable predictors of toxicity in mammals, we can learn from the gene responses of these organisms to xenobiotics. Yeast-cell-based assays are used increasingly for high-throughput screens to test for toxicity of xenobiotics because they mimic the well-conserved stress response [46], and produce chemical and genetic profiles of compounds that provide insights into their mechanisms of action [47,48]. The combination of algorithmic analysis of these types of higher throughput experimental approaches will enrich the mammalian data that are available.

Databases, modeling and predictive tools: ADME/Tox

Since the 1970s, industry and academia have organized databases on proteins, enzyme-encoding genes, and metabolic- and cell-signaling pathways (Figure 3) [49]. Although there have been limited efforts to organize ADME/Tox data, separate focused databases of ADME-associated proteins and pathways such as PharmaGKB [50], the nuclear receptor database [51], human membrane-transporter database [52] and the ADME-AP database [53] can be integrated. Commercial databases such as Metabolite™ (<http://www.mdl.com>), Metabolism™ (<http://www.accelrys.com>) and BioFrontier/P450™ (<http://www.fqspl.com.pl/>) represent a broad collection of metabolic data that are useful for calculating probabilities for a metabolic reaction [54]. Several companies market content databases of gene-expression profiles, histopathology, multi-parameter clinical chemistry tests and morphology in rat organs after treatment with several hundred marketed drugs and toxicants that use different array

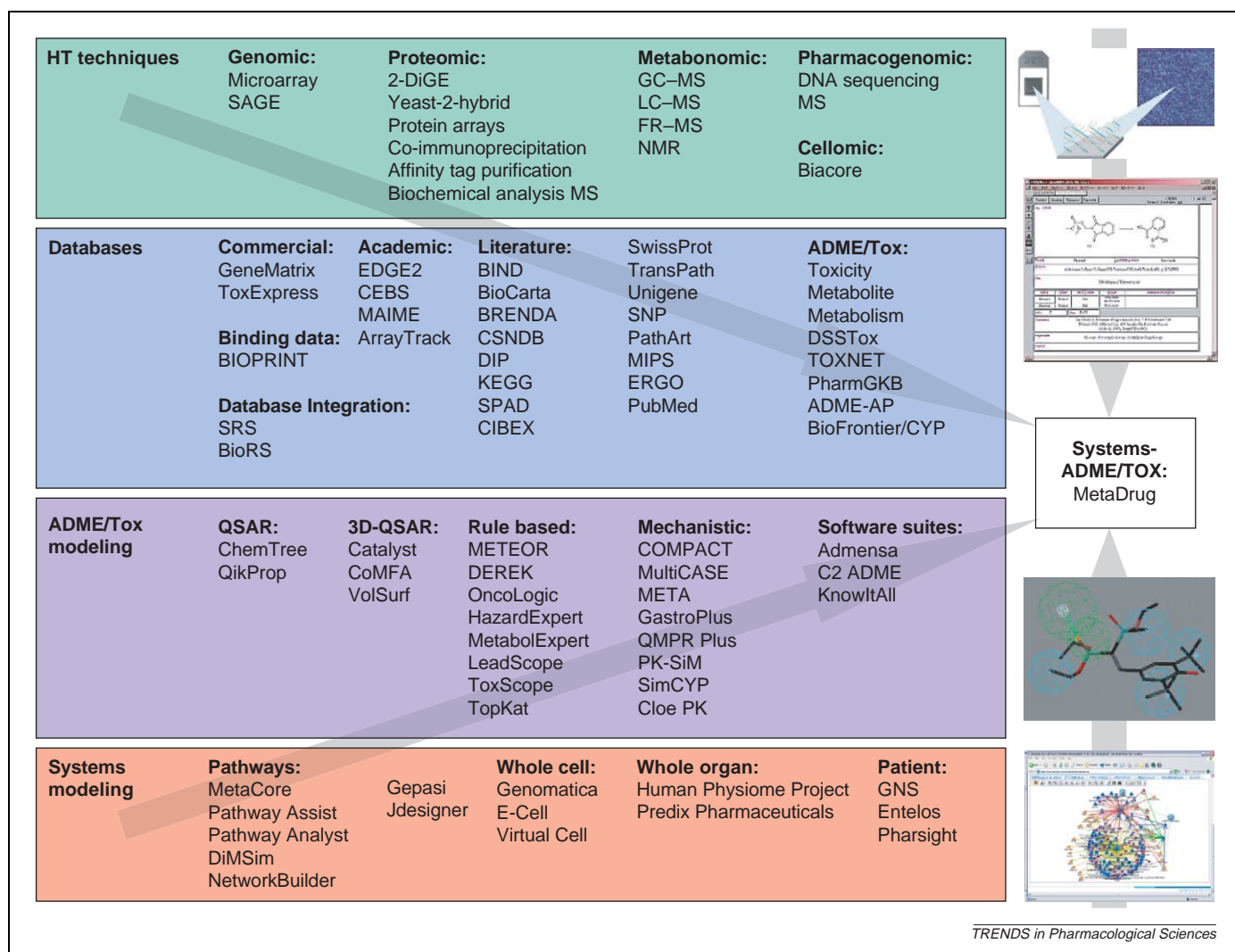


Figure 3. Developing systems-ADME/Tox. The convergence of high-throughput techniques, databases, small-molecule modeling technologies and systems modeling technologies forms the foundation of systems-ADME/Tox. Abbreviations: 2-DiGE, two-dimensional gel electrophoresis; FR-MS, Fourier transformation-mass spectrometry; GC-MS, gas chromatography-mass spectrometry; HT, high throughput; LC-MS, liquid chromatography-mass spectrometry; MS, mass spectrometry; NMR, nuclear magnetic resonance; SAGE, serial analysis of gene expression.

platforms and represent predictive toxicogenomic screening tools. These include Iconix (Drug Matrix™; <http://www.iconixpharm.com/>), GeneLogic (ToxExpress™; <http://www.genelogic.com/>) [55], Curagen (<http://www.curagen.com/>) and Icoria (formerly Paradigm Genetics; <http://www.icoria.com/>). Pattern-recognition analysis depends, ultimately, on the content and quality of these databases coupled with data on the phenotypic endpoint.

Many ongoing projects will contribute to the availability of public databases. The most promising are two databases that are being developed at the National Institutes of Health (NIH) and the Food and Drug Administration (FDA). The National Center for Toxicogenomics at the National Institute of Environmental Health Sciences (NIEHS) and The European Molecular Biology Laboratory-European Bioinformatics Institute are developing a comprehensive Chemical Effects in Biological Systems (CEBS) knowledgebase (<http://www.niehs.nih.gov/nct/cebs.htm>) [56,57]. This will accommodate gene-expression profiles, proteomics and metabolomics data in the extended open-source CEBS Systems Biology object model, and enable complex queries [57,58]. Similar goals

are being pursued by the National Center for Toxicological Research at the FDA in their development of the ArrayTrack database [59]. Finally, the EDGE² database (<http://edge.oncology.wisc.edu/>), a public effort at the University of Wisconsin contains gene-expression profiles following treatment of mice with different toxic molecules [60]. Microarray approaches and database development (Figure 3) provide valuable methods and data to help understand toxicity, which is seen as increasingly important for future submissions to regulatory authorities [61].

Computational models are widely available for predicting ADME/Tox properties using software for either custom-model building [4,49] or pre-built modeling suites [Cerius²™ ADME (<http://www.accelrys.com>) and KnowItAll™ (<http://www.bio-rad.com>)] (Figure 3). Generally, these systems are based around quantitative structure-activity relationships (QSARs) that generate descriptors based on molecular structure and use computational algorithms to relate the key descriptors to the biological activity [62]. The accumulation of drug-metabolism data from the literature has resulted in expert systems for predicting metabolism with

products such as MetabolExpert™ (<http://www.compu-drug.com/>), META™ (<http://www.multicase.com/>) and METEOR™ (<http://www.chem.leeds.ac.uk/luk/>), with the caveat that these contain data from many different mammalian species. Simulation methods have also been developed, including physiologically based pharmacokinetic modeling (PBPK) and methods such as Cloe PK™ (<http://www.cypotex.com/>), GastroPlus™ (<http://www.simulations-plus.com/>), Simcyp™ (<http://www.simcyp.com/>) and others that include toxicokinetic methods. PBPK approaches can be used with either empirical data, *in vitro* data or *in silico* predictions to derive human pharmacokinetic parameters such as area under the curve (AUC) [63]. By contrast, computational approaches for predicting toxicity are studied infrequently [3,64] but are complementary to research on ADME parameters [62]. These methods for individual toxicology properties tend to be rule-based systems such as DEREK™ (<http://www.chem.leeds.ac.uk/luk/>), Hazard Expert™ (<http://www.compudrug.com/>), LeadScope™ (<http://www.leadscope.com/>) or the mechanistic methods COMPACT [65] and MultiCASE™ (<http://www.multicase.com/>).

Applications of systems-ADME/Tox

Traditional ADME/Tox studies provide a detailed understanding of individual proteins. Now, we consider whether the molecule also binds to receptors that affect the regulation of other proteins, and if it interferes with endogenous metabolic, regulatory and transport proteins. Alternatively, the primary metabolic route might be mediated by a polymorphic enzyme and affect the likely therapeutic dose. Many of these questions are now answered earlier in the drug-discovery process *in vitro*. Because xenobiotics and their metabolites influence multiple genes and pathways simultaneously, predicting a response in a heterogeneous population is complex and depends on drug dose, genes, physiological state and other factors. The transcriptional regulation of CYPs, phase II enzymes and many transporters are regulated by numerous NHRs, which affect endogenous metabolism, cell growth, proliferation and oxidative stress [66,67]. Therefore, combining different experimental and predictive approaches will help explain the metabolism and toxicity of unknown compounds.

The development of a systems-ADME/Tox platform (MetaDrug™; <http://www.genego.com>) that links empirical, biological-pathway-centered data, OMICS-based models and ligand-based QSAR models, is the latest step in this research [6]. From an input molecular structure, the major metabolites in humans are predicted and scored with multiple ADME/Tox models (predicting K_m , V_{max} , IC_{50} and K_i values). This can be used to filter molecules before visualization as objects on regulatory networks. The visualization tool also uses the scores for many proteins to predict a network of proteins that are likely to be responsible for ADME/Tox interactions and provides a visual perspective of the most important pathways (Figure 4). This software has also been applied to illustrate the complexity of NHR interactions in humans [14]. Toxicogenomics data (microarrays, protein profiles and metabolic profiles) and pharmacogenomics data [single nucleotide polymorphisms

(SNPs) and haplotypes] can also be visualized and compared on the same networks. This represents a valuable starting point for the development of rational, mechanism-driven hypotheses and models that uses a combined approach rather than relying on either a single type of high-throughput data or QSAR models. Systems-ADME/Tox methods might also help to determine the toxicity of molecules and the genes that might be involved [68] by reconstructing functional networks.

The future of systems-ADME/Tox

Studies of drug safety increasingly use OMICS data, which provides a starting point for a systems-ADME/Tox approach. However, many compounds can be co-administered and thus more studies that describe the effects of dosing with multiple compounds simultaneously are needed to indicate positive and negative interactions [69]. However, extensive experimental studies are expensive and require collaborations between different groups [70]. Presently, therefore, we are witnessing the parallel development and maturation of several databases [57], visualization software and predictive algorithms. The integration of these tools is likely to be synergistic, unlike the complimentary combination of OMICS methods seen to date. We suggest that the ideal study would include all types of experiments for the same sample and genetic background, such that the data are analyzed in a computational system to better understand and explain the discrepancies between the data types [49]. A systems-based approach to ADME/Tox might improve the efficiency of drug discovery by integrating these disparate technologies.

Each of the technologies described above faces considerable challenges. These include controlling experimental variability followed by effective verification, storage, utilization and dissemination of the massive amount of data that is generated. The data derived from animals and *in vitro* models (even human models) must be extrapolated to humans *in vivo*, which is another complex process. Generating computational predictions also requires validation before such models are accepted generally, which takes time and financial investment. Without doubt, the development of new technologies will benefit from the incorporation of relevant OMICS content to identify the specific interaction networks for groups of new chemical entities and toxins: linking the databases that are being developed, such as CEBS and commercially available databases, might be a first step toward this. This could be followed by providing reliable network comparisons using algorithms, and the resultant networks should reflect the responses to specific chemotypes and toxicophores. In parallel, the compounds for which toxicogenomics data are already collected should be processed through the chemical rules and models with the expectation that the resulting data intersects at the level of networks. This represents a starting point for the iterative generation of predicted interaction networks based on molecular structure.

Concluding remarks

Humans are sophisticated organisms who require that multiple cellular processes operate in parallel to ensure survival. Although understanding gene-expression

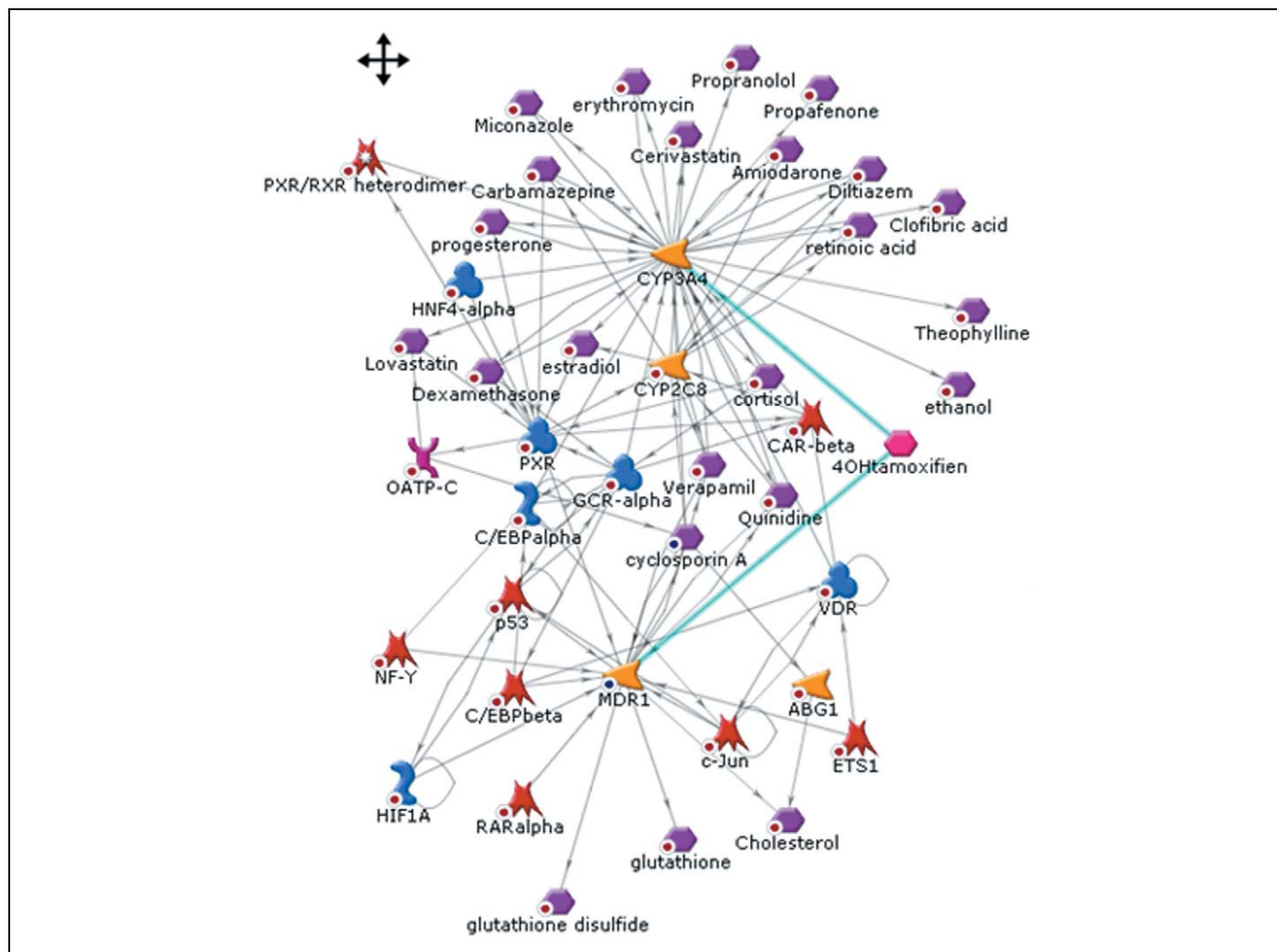


Figure 4. Simultaneous visualization of predicted and empirical protein–ligand interactions. A regulatory network generated with MetaDrug™ (<http://www.genego.com>) shows the predicted binding interactions for 4-hydroxytamoxifen (pink hexagon connected to thick blue lines) derived from QSAR models for P-glycoprotein (P-gp) and CYP3A4 built with literature data. The colored, differently shaped nodes represent enzymes (yellow/orange arrows), ligands (purple hexagons), transcription factors (red) and other proteins (blue). 4-Hydroxytamoxifen is known to be metabolized by (and inhibits) CYP3A4 [72] and to inhibit P-gp ($IC_{50}=7.4 \mu M$) [73].

networks *in silico* and *in vivo* in unicellular and multicellular systems is essential for drug discovery [71], these are not the only important factors. In conclusion, many separate tools and datasets are being integrated in software products that should improve our understanding of the effects of xenobiotics in humans because it is apparent that, used separately, top-down approaches and bottom-up methods are not sufficient.

Acknowledgements

S.E. gratefully acknowledges Maggie A.Z. Hupcey and Craig Giroux (Wayne State University) for editorial comments and suggestions, and our colleagues at GeneGo for software development. This work was supported by NIH grant 1-R43-GM069124–01.

References

- Rawlins, M.D. (2004) Cutting the cost of drug development? *Nat. Rev. Drug Discov.* 3, 360–364
- FDA (2004) Innovation stagnation: challenge and opportunity on the critical path to new medicinal products. <http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.pdf>
- Ekins, S. (2003) *In silico* approaches to predicting metabolism, toxicology and beyond. *Biochem. Soc. Trans.* 31, 611–614
- van de Waterbeemd, H. and Gifford, E. (2003) ADMET *in silico* modelling: towards prediction paradise? *Nat. Rev. Drug Discov.* 2, 192–204
- Ekins, S. *et al.* (2002) Towards a new age of virtual ADME/TOX and multidimensional drug discovery. *J. Comput. Aided Mol. Des.* 16, 381–401
- Bugrim, A. *et al.* (2004) Early prediction of drug metabolism and toxicity: systems biology approach and modeling. *Drug Discov. Today* 9, 127–135
- Hood, L. and Galas, D. (2003) The digital code of DNA. *Nature* 421, 444–448
- Ideker, T. and Lauffenburger, D. (2003) Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol.* 21, 255–262
- Ekins, S. *et al.* (2004) Applying computational and *in vitro* approaches to lead selection. In *Pharmaceutical Profiling in Drug Discovery for Lead Selection* (Borchardt, R.T. *et al.*, eds), pp. 361–389, AAPS Press
- Ekins, S. *et al.* (2003) Generation and validation of rapid computational filters for CYP2D6 and CYP3A4. *Drug Metab. Dispos.* 31, 1077–1080
- Shimada, J. *et al.* (2002) Integrating computer-based *de novo* drug design and multidimensional filtering for desirable drugs. *Targets* 1, 196–205
- Ulrich, R. and Friend, S.H. (2002) Toxicogenomics and drug discovery: will new technologies help us produce better drugs? *Nat. Rev. Drug Discov.* 1, 84–88
- Lord, P.G. (2004) Progress in applying genomics in drug development. *Toxicol. Lett.* 149, 371–375

- 14 Ekins, S. *et al.* A novel method for visualizing nuclear hormone receptor networks relevant to drug metabolism. *Drug Metab. Dispos.* (in press)
- 15 Yoon, B. *et al.* (2003) Mechanisms of benzene-induced hematotoxicity and leukemogenicity: cDNA microarray analyses using mouse bone marrow tissue. *Environ. Health Perspect.* 111, 1411–1420
- 16 Kramer, J.A. *et al.* (2004) Acute Molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. *Chem. Res. Toxicol.* 17, 463–470
- 17 Walgren, J.L. and Thompson, D.C. (2004) Application of proteomic technologies in the drug development process. *Toxicol. Lett.* 149, 377–385
- 18 Anderson, N.L. and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 19, 1853–1861
- 19 Klein, J.B. and Thongboonkerd, V. (2004) Overview of proteomics. *Contrib. Nephrol.* 141, 1–10
- 20 Fountoulakis, M. (2000) Two-dimensional electrophoresis. In *Encyclopedia of separation science, II/electrophoresis*, pp. 1356–1363, Academic Press
- 21 Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* 422, 198–207
- 22 Lahm, H.W. and Langen, H. (2000) Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis* 21, 2105–2114
- 23 Kennedy, S. (2002) The role of proteomics in toxicology: identification of biomarkers of toxicity by protein expression analysis. *Biomarkers* 7, 269–290
- 24 Meneses-Lorente, G. *et al.* (2004) A proteomic investigation of drug-induced steatosis in rat liver. *Chem. Res. Toxicol.* 17, 605–612
- 25 Glassbrook, N. *et al.* (2000) Metabolic profiling on the right path. *Nat. Biotechnol.* 18, 1142–1143
- 26 Nicholson, J.K. and Wilson, I.D. (2003) Understanding 'global' systems biology: metabolomics and the continuum of metabolism. *Nat. Rev. Drug Discov.* 2, 668–676
- 27 Goodacre, R. *et al.* (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* 22, 245–252
- 28 Evans, W.E. and McLeod, H.L. (2003) Pharmacogenomics—drug disposition, drug targets, and side effects. *New Engl. J. Med.* 348, 538–549
- 29 Weinshilboum, R. (2003) Inheritance and drug response. *New Engl. J. Med.* 348, 529–537
- 30 Trotta, R. *et al.* (2004) Trends in pharmacogenomics of drugs acting on hypertension. *Pharmacol. Res.* 49, 351–356
- 31 Heijne, W.H. *et al.* (2003) Toxicogenomics of bromobenzene hepatotoxicity: a combined transcriptomics and proteomics approach. *Biochem. Pharmacol.* 65, 857–875
- 32 Ruepp, S.U. *et al.* (2002) Genomics and proteomics analysis of acetaminophen toxicity in mouse liver. *Toxicol. Sci.* 65, 135–150
- 33 Coen, M. *et al.* (2004) Integrated application of transcriptomics and metabolomics yields new insight into the toxicity due to paracetamol in the mouse. *J. Pharm. Biomed. Anal.* 35, 93–105
- 34 Fountoulakis, M. *et al.* (2002) Modulation of gene and protein expression by carbon tetrachloride in the rat liver. *Toxicol. Appl. Pharmacol.* 183, 71–80
- 35 Klapa, M.I. and Quackenbush, J. (2003) The quest for the mechanisms of life. *Biotechnol. Bioeng.* 84, 739–742
- 36 Weckwerth, W. and Fiehn, O. (2002) Can we discover novel pathways using metabolomic analysis? *Curr. Opin. Biotechnol.* 13, 156–160
- 37 Barabasi, A.-L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113
- 38 Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
- 39 Tornow, S. and Mewes, H.W. (2003) Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Res.* 31, 6283–6289
- 40 Hanisch, D. *et al.* (2002) Co-clustering of biological networks and gene expression data. *Bioinformatics* 18, S145–S154
- 41 Ideker, T. *et al.* (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 18, S233–S240
- 42 Segal, E. *et al.* (2003) Discovering molecular pathways from protein interaction and gene expression. *Bioinformatics* 19, i264–i272
- 43 Spirin, V. and Mirny, L.A. (2003) Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12123–12128
- 44 Covert, M.W. *et al.* (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429, 92–96
- 45 Werner, E. (2003) In silico multicellular systems biology and minimal genomes. *Drug Discov. Today* 8, 1121–1127
- 46 Weiss, A. *et al.* (2004) High-throughput phenotypic profiling of gene-environment interactions by quantitative growth curve analysis in *Saccharomyces cerevisiae*. *Anal. Biochem.* 327, 23–34
- 47 Stockwell, B.R. (2004) The biological magic behind the bullets. *Nat. Biotechnol.* 22, 37–38
- 48 Parsons, A.B. *et al.* (2004) Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat. Biotechnol.* 22, 62–69
- 49 Ekins, S. *et al.* Systems biology: applications in drug discovery. In *Drug Discovery Handbook* (Gad, S., ed.), Wiley (in press)
- 50 Oliver, D.E. *et al.* (2002) Ontology development for a pharmacogenetics knowledge base. *Pac Symp Biocomput.* 2002, 88–99
- 51 Nakata, K. *et al.* (2002) A nuclear receptor database that maps pathways to diseases. *Genome Informatics* 13, 515–516
- 52 Yan, Q. and Sadee, W. (2000) Human membrane transporter database: a web-accessible relational database for drug transport studies and pharmacogenomics. *AAPS PharmSci* 2, E20
- 53 Sun, L.Z. *et al.* (2002) ADME-AP: a database of ADME associated proteins. *Bioinformatics* 18, 1699–1700
- 54 Boyer, S. and Zamora, I. (2002) New methods in predictive metabolism. *J. Comp. Aided Mol. Des.* 16, 403–413
- 55 Castle, A.L. *et al.* (2002) Toxicogenomics: a new revolution in drug safety. *Drug Discov. Today* 7, 728–736
- 56 Waters, M. *et al.* (2003) Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *EHP Toxicogenomics* 111, 15–28
- 57 Mattes, W.B. *et al.* (2004) Database development in toxicogenomics: issues and efforts. *Environ. Health Perspect.* 112, 495–505
- 58 Hood, E. (2003) Proteomics: characterizing the cogs in the machinery of life. *Environ. Health Perspect.* 111, A816–A825
- 59 Tong, W. *et al.* (2003) ArrayTrack—supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ. Health Perspect.* 111, 1819–1826
- 60 Thomas, R.S. *et al.* (2001) Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol. Pharmacol.* 60, 1189–1194
- 61 Hackett, J.L. and Lesko, L.J. (2003) Microarray data—the US FDA, industry and academia. *Nat. Biotechnol.* 21, 742–743
- 62 Ekins, S. and Swaan, P.W. (2004) Computational models for enzymes, transporters, channels and receptors relevant to ADME/TOX. In (Lipkowitz, K.B. *et al.*, eds), pp. 333–415, Wiley-VCH
- 63 Leahy, D.E. (2004) Drug discovery information integration: virtual humans for pharmacokinetics. *DDT: BIOSILICO* 2, 78–84
- 64 Greene, N. (2002) Computer systems for the prediction of toxicity: an update. *Adv. Drug Deliv. Rev.* 54, 417–431
- 65 Lewis, D.F.V. (1996) *Cytochromes P450*, Taylor & Francis
- 66 Ulrich, R.G. (2003) The toxicogenomics of nuclear receptor agonists. *Curr. Opin. Chem. Biol.* 7, 505–510
- 67 Ulrich, R.G. *et al.* (2004) Overview of an interlaboratory collaboration on evaluating the effects of model hepatotoxicants on hepatic gene expression. *Environ. Health Perspect.* 112, 423–427
- 68 Williams, D.P. and Park, B.K. (2003) Idiosyncratic toxicity: the role of toxicophores and bioactivation. *Drug Discov. Today* 8, 1044–1050
- 69 Borisy, A.A. *et al.* (2003) Systematic discovery of multicomponent therapeutics. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7977–7982
- 70 Clish, C.B. *et al.* (2004) Integrative biological analysis of the APOE*3-leiden transgenic mouse. *OMICS* 8, 3–13
- 71 Christopher, R. *et al.* (2004) Data-driven computer simulation of human cancer cell. *Ann. New York Acad. Sci.* 1020, 132–153
- 72 Desta, Z. *et al.* (2004) Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system *in vitro*: prominent roles for CYP3A and CYP2D6. *J. Pharmacol. Exp. Ther.* 310, 1062–1075
- 73 Bekaii-Saab, T.S. *et al.* (2004) Interactions of tamoxifen, N-desmethyl-tamoxifen and 4-hydroxytamoxifen with P-glycoprotein and CYP3A. *Biopharm. Drug Dispos.* 25, 283–289