Early prediction of drug metabolism and toxicity: systems biology approach and modeling

Andrej Bugrim, Tatiana Nikolskaya and Yuri Nikolsky

Many of the drug candidates that fail in clinical trials are withdrawn because of unforeseen effects of human metabolism, such as toxicity and unfavorable pharmacokinetic profiles. Early pre-clinical elimination of such compounds is important but not yet possible. An ideal system would enable researchers to make a confident elimination decision based purely on the structure of a new compound, and incorporate and use multiple pre-clinical experimental data to support such a decision. Currently available resources can be split into three categories: (i) structure–activity relationships (SAR) computational models based on compound structure; (ii) 'pattern' databases of tissue or organ response to drugs, compiled from high-throughput experiments; and (iii) 'systems biology' databases of metabolic pathways, genes and regulatory networks. In this review, we outline the advantages and drawbacks of each of these systems and suggest directions for their integration.

Andrej Bugrim*
Tatiana Nikolskaya
GeneGo
500 Renaissance Drive
Suite 106
St. Joseph
MI 49085, USA
*e-mail: andrej@genego.com
Yuri Nikolsky
Chemical Diversity Labs
11588 Sorrento Valley Road
San Diego
CA 92121, USA

▼ Small-molecule drug discovery is an increasingly costly and time-consuming process. There are several economic and technological reasons for this but the inefficiency is mainly due to the fact that almost half of drug candidates fail in clinical trials. Up to 25% of failed compounds are withdrawn because of metabolic or pharmacokinetics issues or toxicity. Such problems remain difficult to predict from current pre-clinical *in vitro* and *in vivo* testing in cell lines, such as human hepatocytes, and animal models.

Compounds at different stages of pre-clinical testing also undergo *in silico* ADME/TOX (absorption, distribution, metabolism, excretion and toxicity) evaluation, varying from straightforward solubility and reactivity filtering to advanced modeling for brain–blood barrier permeability and compound metabolism. The ultimate goal is to be able to predict human metabolism, toxicity and pharmacokinetics based solely on the structural formula of novel compounds. Such a system would allow the elimination of unfavorable compounds early in development, at the level of lead selection and, ultimately, initial discovery chemistry used in bioassays. Unfortunately, there is no computational system capable of accomplishing these tasks, and many practitioners are skeptical of such developments in the near-future. However, we believe that the creation of comprehensive and reliable *in silico* systems for the evaluation of human metabolism and relevant toxicity is technically feasible and that such systems will be developed in the near-future. Furthermore, the basic 'module' technologies for such systems already exist, and the critical mass of experimental data are compiled in the available databases.

Three broadly defined technologies can be considered as the components of such systems. First, there is the computational modeling of metabolism and toxicity based on chemical structure of lead compounds. These models use vast sets of experimental data (toxicity end points, metabolic products and intermediates) for elucidation of empirical rules describing the response to xenobiotics. The data are used for building structure–activity relationships (SAR) within groups of similar compounds and mathematical models (e.g. neural networks, non-linear regressions, and so on) are built for effective separation of compounds. Second, there are powerful high-throughput experimental technologies that catalog whole genomes, tissues and organ responses to external stimuli. These techniques enable the parallel sequencing of all relevant genes (polymorphism and haplotype analysis), the measurement of mRNA transcription of

1359-6446/04/$ – see front matter PII: S1359-6446(03)02971-4

an entire genome (transcriptomics), the evaluation of expressed proteins and peptides (proteomics) in a tissue, or the quantitation of thousands of metabolites as the final outcome of enzymatic transformations in a given tissue or body fluid (metabonomics). A great amount of such high-throughput data relevant to drug action has been accumulated by life science companies and is in the public domain. Finally, there are databases of cellular functional pathways and enzymatic reactions, assembled from experimental literature. These systems biology databases cover information such as metabolic pathways and networks, eukaryotic signal transduction, transcriptional regulation, membrane receptors and protein interactions. Over half of experimentally validated human and mammalian genes are incorporated in these databases.

Individually, none of these technologies is sufficient for the task of predicting ADME/TOX behavior based solely on chemical structure. For example, empirical structure-based models do not deal with the underlying mechanisms of toxicity and metabolism, and SAR rules are relevant only for a narrow chemical space around the original structure. Genomics-based systems are based on 'pattern recognition' – novel compounds are assumed to have similar toxicity as a reference compound with a similar profile; in reality, this is true only for structurally similar molecules and the techniques therefore have limited predictive value. Furthermore, the pathways databases do not reflect spatial or temporal patterns of pathway activity, and therefore reflect the functional potential of a cell or tissue rather than the real activity in response to stimuli.

Because the understanding of metabolism and toxicity is a systems-level problem, the solution will require integration of these technologies into one comprehensive environment, connecting the relevant chemistry with key cellular metabolic pathways. In this review, we describe the most significant developments in all three areas and outline a possible way for integration of these technologies into a single analytical system.

In-silico prediction of drug metabolism and related toxicity – the current state of affairs.

A major step in assessing the potential toxicity of a novel chemical entity is the prediction of entry and fate of the compound in human metabolism. There are two distinct sides to this problem. The first task is to predict nominal metabolic transformations of a molecule. The second is to understand which enzymes will actually be involved, predict relative importance of alternate reaction routes and finally, their interactions with endogenous metabolism.

Over the years several computational tools have been developed that handle various aspects of metabolism prediction. The primary group of tools can be described as 'rule-based' metabolism prediction systems [1–5]. These tools are mainly targeted at the prediction of metabolic routes and metabolite structures. The foundation of the rule-based metabolism prediction systems is formed by a collection of expert-derived patterns or model transformations. The input is usually supplied as a molecular structure that is analyzed for the presence of structural elements that will enable or prevent its metabolism along specific routes.

The major advantage of this approach is that mechanistic detail and the structures of intermediary metabolites of drug biotransformations are explicitly predicted. Moreover, collections of rules in these systems provide linkages between various structural elements related by metabolic transformations. It is possible that they could also be used to reverse-engineer compounds. For example, one could design compounds that contain substructures that are known to undergo desirable metabolic transformations, while avoiding undesirable ones.

A significant problem encountered in this approach is combinatorial explosion of the number of metabolites as one makes an attempt to predict metabolic trees beyond the first step. This problem is somewhat reduced when different features in the software are implemented to partially reduce the number of predicted metabolites, including higher-level rules to evaluate the likelihood of a particular biotransformation [6], stability analysis to filter out metabolites and routes that lead to unstable products [1], and the use of genetic algorithms to prioritize metabolites [7], among others.

Despite these weaknesses, *in silico* prediction of metabolic transformations is an important first step in predicting early in the discovery process how a drug will behave in the human body. However, many nominally predicted pathways never materialize because of numerous factors, including poor enzyme–substrate binding, enzyme inhibition, low reaction rates and competing reactions. A further concern in this area involves the growing appreciation that there can be significant differences in drug metabolism as a result of subtle differences in enzymatic phenotype. Thus, the metabolism of the same drug studied in the same species using the same method can still vary significantly from individual to individual based upon normal [8] or disease-related differences in phenotype [9].

From the systems biology perspective, two other areas of *in silico* modeling must be merged with metabolic fate prediction to successfully reconstruct drug behavior. First, an adequate reconstruction of the endogenous metabolic and signaling systems with which drug leads interact is necessary. Additionally, it is important to be able to reliably

model interactions between compounds (including the original compound or its predicted metabolites) and proteins participating in those pathways.

QSAR modeling

Much of the work in QSAR modeling has been focused on potential interactions of small molecules with major classes of proteins such as drug metabolizing enzymes [10–12], transporters [13], channels [14] and receptors [15,16]. These models usually correlate physicochemical descriptors with measured properties such as compound–enzyme binding and substrate potential. In a series of recent reviews and publications, QSAR models for substrates of several major drug-metabolizing cytochromes have been described; for example, Lewis and co-workers developed models for substrates of human cytochrome P450s [12,17]. Among the factors that have been found to be important for compound metabolism are frontier orbital energies and pi-pi stacking interactions.

Another effort to build QSAR models for major human drug metabolizing cytochromes has been recently undertaken jointly by our companies. We have collected and analyzed data on over 2200 compounds that are described in the literature as substrates and non-substrates of cytochrome P450 enzymes. [10]. Thirty-nine molecular descriptors have been identified as correlating with the ability of the compound to become a substrate of one of the major human cytochromes. Our neural network-based models show 76% accuracy in predicting substrates for major human cytochromes. We have built both QSAR models for individual isozymes, and an integral model for predicting metabolism by the whole complement of P450 enzymes in human liver microsomes.

Although cytochromes represent the main gateway for xenobiotics into human metabolism, there are many other enzymes that catalyze later-stage reactions. The enzymes involved at this stage are predominantly transferases with a broad spectrum of substrate specificity [18–20]. Importantly, many of these enzymes also participate in metabolic pathways of many endogenous compounds such as steroid hormones. Drug metabolism might affect these pathways by titrating away common enzymes and therefore changing the rate of normal metabolic fluxes (Figure 1). The later-stage metabolic fate of compounds is an area that systems biology, with its emphasis on reconstructing cellular pathways and networks, is uniquely suited to elucidate.

High-throughput 'new biology' and toxicology.

During the 1990s, a new set of methods were developed to greatly increase the amount of experimental data from one

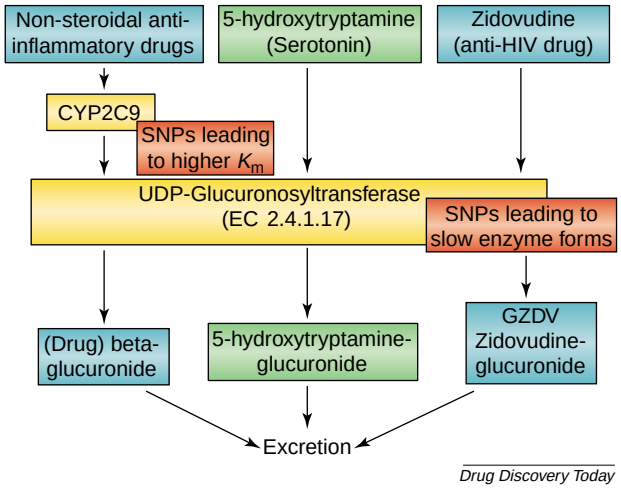

Figure 1. Interactions between drug metabolism and endogenous pathways. UDP- glucoranysyltransferase participates in both drug bioconjugation and metabolism of endogenous compounds (such as neuromediators and steroid hormones). This intersection between xenobiotic and endobiotic pathways might be the source of adverse health effects. The unfortunate combination of SNPs in CYP2C9 (upstream reaction), leading to high K_M, and in UGT1A6 (gene coding UDP- glucoranysyltransferase), which result in slow enzyme forms, could lead to dangerous overloading of this enzyme if drugs are taken simultaneously.

experiment. These methods include automated DNA sequencing, microarray analysis of gene expression and protein profiling. Although these techniques have a wide variety of applications, we will consider only those related to toxicology here.

Toxicogenomics

This involves investigating the response to xenobiotics at the genomics level (mainly gene expression). Whole-genome and specialty microarrays are the tools of choice in toxicogenomics studies. These studies classify toxicity based on gene transcriptional patterns, compiling large datasets of responses that are complemented with well-characterized toxicological endpoints. Patterns from new chemicals can be compared with those in the database to extrapolate the probability of toxicity.

DNA microarrays are a highly informative, robust, rapid and cost-efficient tool for the toxicological evaluation of compounds. They are applied widely to determine cellular response to toxicants and to elucidate toxicity mechanisms [21–23]. However, as a predictive tool, toxicogenomics profiling is useful only for a limited chemistry space; structures of novel compounds have to be similar to those in the databases. Because chemical diversity of active molecules is a major variable in drug discovery [24], toxicity of

new compounds often cannot be extrapolated from the profiles of existing drugs.

Proteomics

This is defined as the high-throughput separation, display and identification of proteins [25] and their interactions [26,27] and is increasingly applicable in predictive toxicology and drug discovery in general [28–31]. The global protein and peptide profiling is usually done by 2D-PAGE electrophoresis separation followed by mass spectrometry and NMR detection. Application of proteomics to toxicogenomics is based on the assumption that particular groups of xenobiotics would induce specific patterns of protein expression. The method is useful both for investigative studies of new targets for toxicants and predictive profiling. SAR analysis can be performed within these groups of compounds and new compounds can be compared with the established patterns. Proteomics is considered advantageous over mRNA expression because it allows the assaying of body fluids and therefore fast, non-invasive detection of biomarkers. Furthermore, proteins are closer to toxicology endpoints than mRNA because not all mRNA is translated into proteins and proteins can undergo post-translational modifications [32].

Metabonomics

Metabonomics is defined as the large-scale (organism level) study of metabolic profiles. The principle of metabonomics is powerful and straightforward. Human metabolism is the backbone of cellular organization and the principle effecter of complex cellular signaling in response to stimuli. As such, any stress or intervention, particularly one as profound as treatment with bioactive drugs molecules, affects the expression of metabolic enzymes and, over time, metabolite composition and relative concentration in body systems [33]. These changes are complementary and interrelated with the responses of the organism to gene mutation, disease condition or drug intervention [34,35].

Although technically challenging, large-scale metabolite profiling is accessible by several methods including gas spectroscopy–mass spectrometry [36], NMR analysis [37] and HPLC [38]. High-throughput methods are routine at larger drug companies and offered as a service by specialist companies [39]. The global metabolite content can be monitored by *in vivo* toxicological animal testing during lead selection and optimization. Similar human metabolite profiling has been used throughout clinical development for discovery of clinical safety biomarkers [40].

One important advantage of metabonomics technology is that it deals with the 'end points' of protein action – the changes in metabolite concentrations. The other 'omics' experiments measure the 'intermediate' steps in the response of the organism to the xenobiotics (mRNA and protein expression). Although all three kinds of high-throughput experimental data do not directly relate to classical toxicological end points, such as animal survival or mutation rates, the level of metabolic profiling is believed to be the most direct means of measuring toxicity *in vitro*.

It has been argued that multivariable temporal modeling of metabolism (for example, using neural networks [41]) is useful for the connection of gene- and protein-level events to pathological effects at the organism level [40,42]. However, the modeling and the interpretation of these data are complicated. One challenge is the complexity and spatial dispersion of affected cell types. Another challenge is that the feedbacks in many metabolic pathways are not reflected at the protein concentration or gene expression level. Furthermore, every high-throughput experimental method suffers from different kinds of technical hurdles that distort the response. Therefore, the outcome is not directly comparable to the gene or protein level.

It is believed that these drawbacks could be reduced if the three major kinds of high-throughput data (transcriptomic, proteomic and metabonomic) are related to some common denominator, for example metabolic maps, such as those provided by KEGG [43], MPW [44] and the more comprehensive functional pathways database MetaCore (http://www.genego.com). The drug industry and the academic community have embraced these efforts as the reasonable next step. However, there are several obstacles to this goal. Most important among these is that many endogenous and xenobiotic compounds can be inter-converted by non-enzymatic chemical reactions and extra-genomic elements [42]. In addition, foreign compounds interact with over 500 species of human commensal microflora [45], and mammalian metabolic control functions are dispersed among different cell types in different compartments and physiological conditions. Consequently, it is probably not possible to completely characterize the metabolic fate of a xenobiotic. However, the distribution of a drug can be described as a probabilistic function with the major fluxes being determined, which would already be a vast improvement over the current predictive value that can be generated by high-throughput data.

System-level reconstruction of endogenous pathways

Systems biology was defined by its founders as the integration of experimental data generated by high-throughput

platforms (genetic, transcriptomic, proteomic, metabonomic) to understand function through different levels of biomolecular organization [46]. In the narrowest sense, this term describes the computational systems and models that deal with the analysis of the vast experimental data accumulated during the past ten years of 'new biology'. These systems have multiple applications in drug discovery, including the validation of drug targets via cross-referencing of experimental data, pharmacogenomics, polymorphism haplotyping and the understanding of drug metabolism and toxicity.

In a more general sense, systems biology has come to mean a more integrated approach to drug discovery where the effects of a lead compound on whole pathways and networks of pathways, not just the effects on isolated proteins, are evaluated as early as possible in the drug discovery process. According to systems biology, any successful platform for predictive ADME/TOX needs to consider drug metabolism within the context of the endogenous cellular processes. Here we include a brief overview of the advances in the area of pathway reconstruction and modeling.

Pathway databases

A handful of public and commercial efforts have systematically emphasized various aspects of general biochemistry, metabolism and cell signaling. The first tier of these projects is essentially reference databases. These resources provide access to data on individual enzymatic reactions and their parameters (BRENDA [47], EMP [48]), multi-step metabolic pathways (MPW [49], EcoCyc and HumanCyc [50]) or human-curated maps for entire functional blocks [KEGG [43] and Biocarta (http://www.biocarta.com)].

Although these resources represent a good general reference, they also possess significant limitations. Most importantly, in many cases the amount of data contained in them is insufficient to be used effectively in building coherent reconstructions. According to our estimate, KEGG (probably the most comprehensive of pathways databases) captures products of ~2500 human genes, a mere 15% of human genes with known function. Also, the collection of pathways in these databases (with the exception of EcoCyc, HumanCyc and Biocarta) is not organism-specific because the data have been collected from many species. The human-specific maps in databases such as KEGG are actually generic pathways with information on human enzymes superimposed, leading to many errors and inconsistencies.

Pathway reconstruction platforms

This second class of resource provides an added functionality to pathway databases by reconstructing condition-specific pictures of relevant pathways. These reconstructions, dubbed metabolic reconstructions (MR), have been pioneered by Selkov and co-workers [50,51]. Metabolic blocks corresponding to the genetic component of the organism are connected into wire diagrams for each different bacterial species. For some better-studied bacterial species, such as *Escherichia coli* and *Bacillus subtilis*, over half of all identified genes have been connected into MR models.

Several investigations have confirmed the power of MR models in predicting gene functions and finding novel pathways [52,53]. Currently, MR is implemented in the commercial package ERGO, developed and distributed by Integrated Genomics (http://www.integratedgenomics.com) and the public system WIT [54]. These platforms contain a collection of metabolic pathways that can be used to reconstruct the metabolism of over 500 bacterial and eukaryotic species.

Although MR and ERGO are powerful tools, they are limited mainly to bacterial metabolism. Many of the available tools that allow manipulation of mammalian and human data are gene expression analysis systems with an added capability to map data onto pathways. However, the currently available commercial products are not designed for building system-level reconstructions and must rely on the pathway data that are available primarily in KEGG and Biocarta, with the previously mentioned limitations.

Recently GeneGo introduced a novel type of systems biology software dubbed MetaCore. MetaCore relies on the proprietary collection of metabolic and cell signaling pathways that are being constantly extracted from the literature and updated by a dedicated team of experts. This collection is significantly more comprehensive than those available in KEGG and Biocarta. Currently, ~45% of annotated human genes can be mapped onto MetaCore pathways.

MetaCore content is accessed on two levels. Expertly curated maps of major functional blocks in cell signaling and metabolism represent the top level. On the lower level, a user can explore the whole space of reactions and interactions by using network-generating tools. The major advantage of MetaCore lies in its ability to concurrently visualize and cross-validate multiple types and sets of molecular data within the framework of its maps and networks and observe functionally meaningful patterns in data (Figure 2). Functional analysis in MetaCore is still limited by the linking of data to pre-assembled maps. Although large in number (over 250 in the current version), these maps probably represent just a handful of 'windows' into the real workings of the living cell. The next step would be to create algorithms that are capable of inferring condition-specific functional blocks by analyzing the whole space of

Figure 2. Mapping high-throughput molecular data onto pathways helps to find functionally meaningful patterns and to reconstruct disease-specific pathways. In this example microarray data from glaucoma were mapped onto human pathways using MetaCore™ platform. Data are visualized as 'thermometers' next to protein icons. Red and blue fillings indicate that the corresponding genes are over- and underexpressed, respectively, in disease. Data on this map show a consistent functional pattern where inhibitors of NF-κB are overexpressed, whereas genes downstream of this transcriptional factor are underexpressed. Interactions on the map are color coded for inhibition (red) and activation (green). Symbols within hexagonal shapes in the middle of interactions denote mechanisms (e.g. P, phosphorylation; Tr, transcriptional regulation; B, binding)

molecular interaction networks alongside high-throughput data. Several such algorithms have been recently described, primarily for the analysis of interaction networks from yeast two-hybrid assays [55–57]. However, their applicability to the reconstruction of human and mammalian pathways has yet to be proven.

To demonstrate the utility of pathway reconstruction technology in ADME/Tox studies, we reverse-engineered the toxic effects of pyrazinamide, the drug used for treatment of multidrug-resistant latent tuberculosis infection. Pyrazinamide is known for severe side-effects related to the accumulation of uric acid [Ref. 58 and CDC (http://www.cdc.gov/od/oc/media/pressrel/r010831.htm)]. We reconstructed the network of metabolic reactions around uric acid using MetaCore tools. Because pyrazinamide elevates the cellular concentration of NAD^+ *in vivo* [59], we built a map of NAD^+ metabolism and networks around NAD^+ to reveal possible connections. The final map of possible metabolic effects of pyrazinamide is shown in Figure 3. It demonstrates that inhibition of enzymes 2.4.2.30, 2.4.2.31 and 4.1.1.45 (crosses, Figure 3) is responsible for substantial accumulation of NAD^+. This, in turn, shifts up the substrate balance in the purine metabolic pathway towards the production of urate (Figure 3, marked by an arrow on the right side). In addition, pyrazinamide can play a role as amino group donor, which also contributes to the effect (via 2.4.2.1). Purine metabolism is substantially different between mice and human. Therefore, uric acid, as the final product of this pathway, accumulates in humans but not in mice, which convert it into allantoin. All these data, although scattered in publications, were available at the time of pyrazinamide clinical trials. However, the synthesis of these was not obvious for the developers of this drug and the FDA, who approved the drug for market. We believe that similar pathway-centered analysis will be useful when any new compound is to be tested in humans.

Next steps – predictive ADME/Tox in the systems biology era

The theoretical and experimental approaches described in this review address different aspects of assessing metabolism and toxicity of novel drug candidates. What is needed is a platform that integrates different types of *in silico* modeling with high-throughput molecular assays and that enables the reconstruction of a system-level picture of the behavior of a drug. We envision the next generation platforms for predictive ADME/Tox as integrated systems that can process different types of inputs and present researchers with an overview of drug metabolism and drug-induced processes. Major components of such a system should include:

- A suite of QSAR models for making a quantitative prediction about the interaction of a test molecule with important classes of human enzymes and other proteins.
- A module for prediction of the metabolic fate of a compound, with metabolite structures and their relative abundance.
- A comprehensive database on the endogenous human and mammalian metabolic and regulatory pathways.
- A suite of software tools for integrating multiple types of high-throughput molecular data and for mapping them onto pathways. This platform should support concurrent mapping of gene expression, metabonomic and proteomic data and potentially include software for automated extraction of affected pathways.

A proposed workflow for ranking drug candidates using such a system is presented in Figure 4.

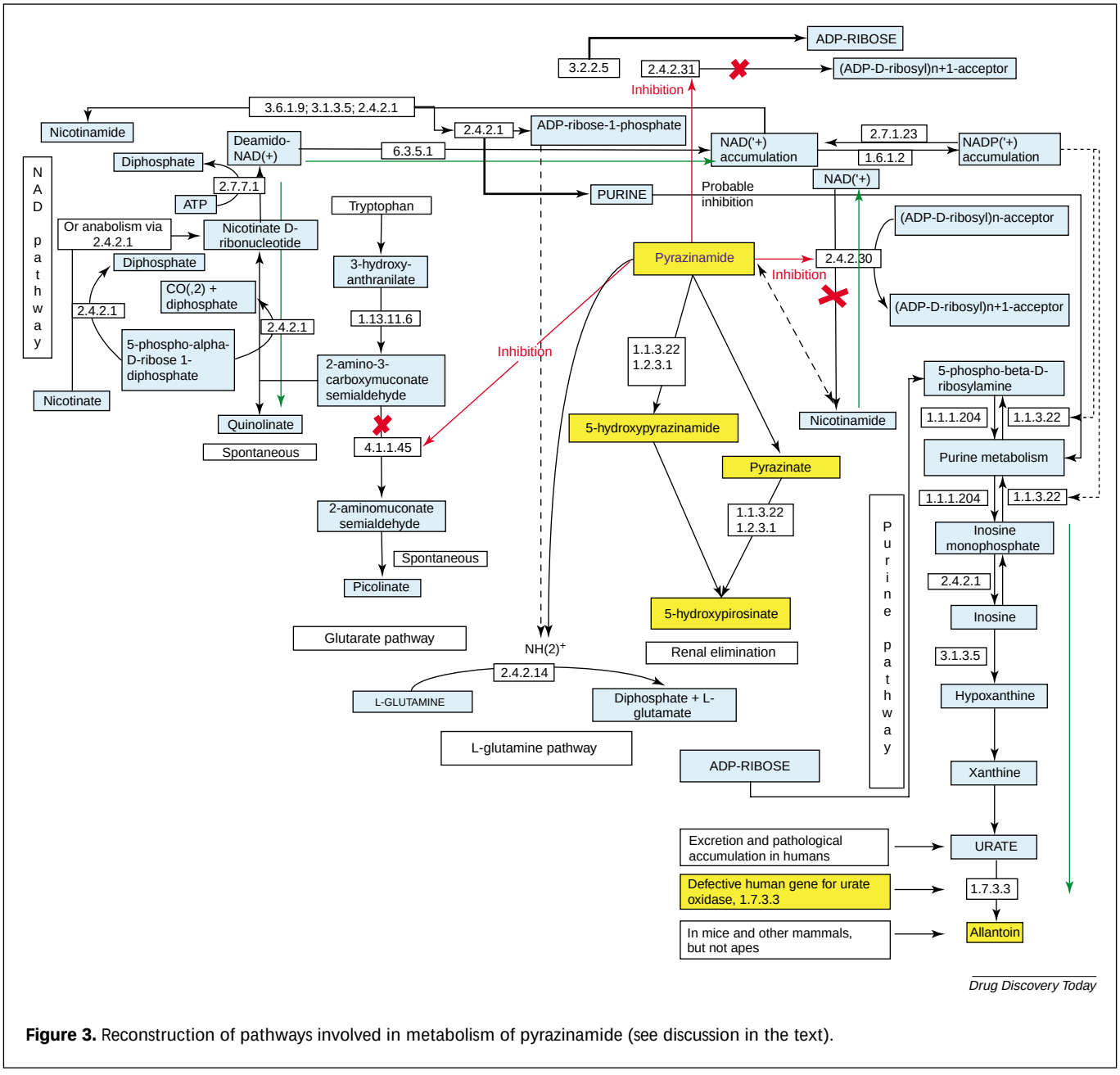

Figure 3. Reconstruction of pathways involved in metabolism of pyrazinamide (see discussion in the text).

A set of drug candidates is first presented to the QSAR and metabolic fate prediction modules. At this stage, the ability of a drug candidate to become a substrate, inhibitor or an inducer for various cytochrome P450s is tested. The system also predicts the structure of the first-pass metabolites and transition states that precede them. Predicted structures are fed back to the QSAR module, where the relative distribution of reaction rates along different reaction routes is computed based on models that relate transition state descriptors to relative reaction rates; metabolites are prioritized based on relative predicted reaction rates.

Priority metabolites could be fed back to the QSAR and fate-prediction modules to generate second-pass metabolites. Valuable information about metabolic fate and major metabolites that are generated by these steps is then placed into the context of reconstructed pathways. Information on predicted structures, their biological activity and the enzymes involved is merged with information on endogenous pathways. Simultaneously, metabonomic and toxicogenomic data are mapped onto the pathways, which serve as a skeleton on which other types of information are assembled.

Currently, this approach is being implemented in the development of a novel platform for predictive

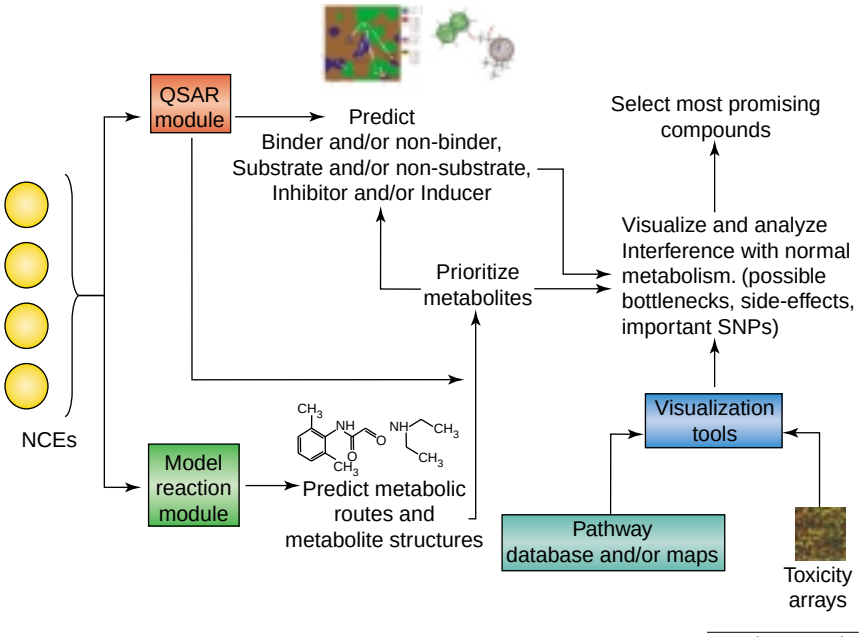

Figure 4. Proposed functionality of the next-generation ADME/Tox platform (see discussion in the text)

ADME/Tox called MetaDrug. This platform, which is being developed by GeneGo, gives the user an ability to assemble all kinds of predictions along with toxigenomics and metabonomics profiles around the framework of pathways. Consistently affected pathways can be identified after performing such mapping. Different types of data and prediction could be cross-validated with much better efficiency when considered in the context of pathways. For example, a QSAR-based prediction of enzyme inhibition could be compared with changes of the metabolites in pathways that involve this enzyme. Potential bottlenecks would be identified, raising 'red flags' about the compound in question.

Concluding remarks

In summary, we believe that the data and methodologies exist to greatly improve ADME/Tox prediction of novel chemical compounds early in the drug discovery process. The integration of modeling, high-throughput and systems biology approaches will allow true breakthroughs in *in silico* ADME/Tox assessment.

References

1 Talafous, J. *et al.* (1994) META. 2. A dictionary model of mammalian xenobiotic metabolism. *J. Chem. Inf. Comput. Sci.* 34, 1326–1333
2 Klopman, G. *et al.* (1994) META. 1. A program for the evaluation of metabolic transformation of chemicals. *J. Chem. Inf. Comput. Sci.* 34, 1320–1325
3 Darvas, F. (1988) Predicting metabolic pathways by logic programming. *J. Mol. Graph.* 6, 80–86
4 Darvas, F. (1987) MetabolExpert, an expert system for predicting metabolism of substances, in QSAR. In *Environmental Toxicology* (Keiser, K., ed.), pp. 71–81, Riedel, Dordrecht
5 Greene, N. *et al.* (1999) Knowledge-based expert systems for toxicity and metabolism prediction: DEREK,StAR and METEOR. *SAR QSAR Environ. Res.* 10, 299–314
6 Judson, P.N. *et al.* (1996) Using new reasoning technology in chemical information systems. *J. Chem. Inf. Comput. Sci.* 36, 621–624
7 Klopman, G. *et al.* (1997) META. 3. A genetic algorithm for metabolic transform priorities optimization. *J. Chem. Inf. Comput. Sci.* 37, 329–334
8 Hayashi, S. *et al.* (1991) Genetic polymorphisms in the 5′-flanking region change transcriptional regulation of the human cytochrome P450IIE1 gene. *J. Biochem. (Tokyo)* 110, 559–565
9 Kato, S. *et al.* (1995) Cytochrome P4502E1 (CYP2E1) genetic polymorphism in a case-control study of gastric cancer and liver disease. *Pharmacogenetics* 5, S141–S144
10 Korolev, D. *et al.* (2003) Modeling of human cytochrome p450-mediated drug metabolism using unsupervised machine learning approach. *J. Med. Chem.* 46, 3631–3643
11 de Groot, M.J. and Ekins, S. (2002) Pharmacophore modeling of cytochromes P450. *Adv. Drug Deliv. Rev.* 54, 367–383
12 Lewis, D.F. *et al.* (1998) An improved and updated version of the compact procedure for the evaluation of P450-mediated chemical activation. *Drug Metab. Rev.* 30, 709–737
13 Zhang, E.Y. *et al.* (2002) Modeling of active transport systems. *Adv. Drug Deliv. Rev.* 54, 329–354
14 Ekins, S. *et al.* (2002) Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channel. *J. Pharmacol. Exp. Ther.* 301, 427–434
15 Ekins, S. *et al.* (2002) A ligand-based approach to understanding selectivity of nuclear hormone receptors PXR, CAR, FXR, LXRalpha, and LXRbeta. *Pharm. Res.* 19, 1788–1800
16 Ekins, S. and Erickson, J.A. (2002) A pharmacophore for human pregnane X receptor ligands. *Drug Metab. Dispos.* 30, 96–99
17 Lewis, D.F. *et al.* (1998) Further evaluation of COMPACT, the molecular orbital approach for the prospective safety evaluation of chemicals. *Mutat. Res.* 412, 41–54
18 Beken, S. *et al.* (2001) Effects of L-proline on phase I and phase II xenobiotic biotransformation capacities of rat and human hepatocytes in long-term collagen gel cultures. *Altern. Lab. Anim.* 29, 35–53
19 Chen, T.L. *et al.* (2000) Effects of propofol on functional activities of hepatic and extrahepatic conjugation enzyme systems. *Br. J. Anaesth.* 84, 771–776
20 Jackson, H. (1996) Baculovirus-mediated high level expression of a human thiopurine methyl transferase. *Biochem. Mol. Biol. Int.* 38, 357–364
21 Thomas, R.S. *et al.* (2001) Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol. Pharmacol.* 60, 1189–1194
22 Amin, R.P. *et al.* (2002) Genomic interrogation of mechanisms underlying cellular responses to toxicants. *Toxicology* 27, 555–563
23 Waring, J.F. *et al.* (2002) Identifying toxic mechanisms using DNA microarrays. *Toxicology* 27, 537–550
24 Kubinyi, H. (2003) Drug research:myths, hype and reality. *Nat. Rev. Drug Discov.* 8, 665–668

25 Blackstock, W.P. and Weir, M.P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17, 121–127
26 Ho, Y. *et al.* (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* 415, 180–183
27 Gavin, A.C. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147
28 Anderson, N.L. *et al.* (1996) The effects of peroxisome proliferators on protein abundance in mouse liver. *Toxicol. Appl. Pharmacol.* 137, 75–89
29 Chevalier, S. *et al.* (2000) Proteomic analysis of differential protein expression in primary hepatocytes induced by EGF, tumor necrosis factor or the peroxisome proliferators nafenopin. *Eur. J. Biochem.* 267, 4624–4634
30 Arce, A. *et al.* (1998) Changes in the liver protein pattern of female Wistar rats treated with the hypoglycemic agent SDZ PGU 693. *Life Sci.* 63, 2243–2250
31 Fountoulakis, M. *et al.* (2000) Two-dimensional database of mouse liver proteins: changes in hepatic protein levels following treatment with acetaminophen or its non-toxic regioisomer. *Electrophoresis* 21, 2148–2161
32 Kennedy, S. (2002) The role of proteomics in toxicology: identification of biomarkers of toxicity by protein expression analysis. *Biomarkers* 7, 269–290
33 Griffin, J.L. *et al.* (2001) Metabolic profiling of genetic disorders: a multi-tissue NMR spectroscopic and pattern recognition study into dystrophic tissue. *Anal. Biochem.* 293, 16–21
34 Smith, L.L. (2001) Key challenges for toxicologists in the 21st century. *Trends Pharmacol. Sci.* 22, 281–285
35 Raamsdonk, L.M. *et al.* (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* 19, 45–51
36 Choi, B.K. *et al.* (2001) Comparison of Quadropole, Time-of-flight and Fourier Transform mass analyzers for LC-MS applications. *LC GC* 19, 514–520
37 Nicholson, J.K. *et al.* (1999) Metabonomics: understanding the metabolic responses of living systems to pathological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29, 1181–1189
38 Vigneau-Calahan, K.E. *et al.* (2001) Characterization of diet-dependent metabolic serotypes: analytical and biological variability in rats. *J. Nutr.* 131, 924–932
39 Beecher, C. (2002) Metabolomics: the newest of the 'omics' sciences. *Innov. Pharm. Technol.* 2, 57–64
40 Nicholson, J.K. *et al.* (2002) Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* 1, 153–161
41 Holmes, E. *et al.* (2001) Metabonomic classification of genetic variations in toxicological and metabolic responses using probabilistic neural networks. *Chem. Res. Toxicol.* 14, 182–191
42 Nicholson, J.K. and Wilson, I.D. (2003) Understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discov.* 2, 668–676
43 Kanehisha, M. *et al.* (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46
44 Selkov, E. *et al.* (1998) MPW: the metabolic pathways database. *Nucleic Acids Res.* 26, 43–45
45 Gilmore, M.S. and Ferretti, J.J. (2003) The thin line between gut commensal and pathogen. *Science* 299, 1999–2002
46 Hood, L. and Galas, D. (2003) The digital code of DNA. *Nature* 421, 444–448
47 Schomburg, I. *et al.* (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 30, 47–49
48 Selkov, E. *et al.* (1996) The metabolic pathway collection from EMP: the enzymes and metabolic pathways database. *Nucleic Acids Res.* 24, 26–28
49 Selkov, E., Jr *et al.* (1998) MPW: the metabolic pathways database. *Nucleic Acids Res.* 26, 43–45
50 Karp, P.D. *et al.* (2002) The metacyc database. *Nucleic Acids Res.* 30, 59–61
51 Gaasterland, T. and Selkov, E. (1995) Reconstruction of metabolic networks using incomplete information. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 3, 127–135
52 Covert, M.W. *et al.* (2001) Metabolic modeling of microbial strains *in silico. Trends Biochem. Sci.* 26, 179–186
53 Selkov, E. *et al.* (2000) Functional analysis of gapped microbial genomes: amino acid metabolism of Thiobacillus ferrooxidans. *Proc. Natl. Acad. Sci. U. S. A.* 97, 3509–3514
54 Overbeek, R. *et al.* (2000) WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.* 28, 123–125
55 Rives, A.W. and Galitski, T. (2003) Modular organization of cellular networks. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1128–1133
56 Snel, B. *et al.* (2002) The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5890–5895
57 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
58 Papastavros, T. *et al.* (2002) Adverse events associated with pyrazinamide and levofloxacin in the treatment of latent multidrug-resistant tuberculosis. *CMAJ* 167, 131–136
59 Shibata, K. *et al.* (2001) Effects of dietary pyrazinamide, an antituberculosis agent, on the metabolism of tryptophan to niacin and of tryptophan to serotonin in rats. *Biosci. Biotechnol. Biochem.* 65, 1339–1346