DMD Fast Forward. Published on December 28, 2005 as doi:10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.

A COMBINED APPROACH TO DRUG METABOLISM AND TOXICITY ASSESSMENT

Sean Ekins, Sergey Andreyev, Andy Ryabov, Eugene Kirillov, Eugene A. Rakhmatulin, Svetlana Sorokina, Andrej Bugrim, Tatiana Nikolskaya.

GeneGo Inc., 500 Renaissance Drive, Suite 106, St. Joseph, MI 49085.

DMD Fast Forward. Published on December 28, 2005 as DOI: 10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.

Running title: **DRUG METABOLISM AND TOXICITY ASSESSMENT**

Corresponding Author:

Sean Ekins M.Sc., Ph.D, D.Sc.

Vice President, Computational Biology, GeneGo, Inc.

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

Phone 269-930-0974, Fax 269-983-7654,

Email sean@genego.com, ekinssean@yahoo.com

Text pages:	20
Supplemental Tables:	2
Supplemental Figures:	4
Figures:	4
References:	46
Words in Abstract:	249
Words in Introduction:	780
Words in Discussion:	2332

Abstract

The challenge of predicting the metabolism or toxicity of a drug in humans has been approached using in vivo animal models, in vitro systems, high throughput genomics and proteomics methods and more recently computational approaches. Understanding the complexity of biological systems requires a broader perspective rather than focusing on just one method in isolation for prediction. Multiple methods may therefore be necessary and combined for a more accurate prediction. In the field of drug metabolism and toxicology we have seen the growth in recent years of computational quantitative structure activity relationships (QSAR) as well as empirical data from microarrays. In the current study we have further developed a novel computational approach MetaDrug™ that: 1) predicts metabolites for molecules based on their chemical structure, 2) predicts the activity of the original compound and its metabolites with various ADME/Tox models, 3) incorporates the predictions with human cell signaling and metabolic pathways and networks and 4) integrates networks and metabolites, with relevant toxicogenomic or other high throughput data. We have demonstrated the utility of such an approach using recently published data from in vitro metabolism and microarray studies for Aprepitant, L-742694, Trovofloxacin, 4-hydroxytamoxifen and artemisinin and other artemisinin analogs to show the predicted interactions with CYPs, PXR and P-gp, the metabolites and the networks of genes that are affected. As a comparison we used a second computational approach MetaCore™, to generate statistically significant gene networks with the available expression data. These case studies demonstrate the combination of QSAR and systems biology methods.

Predicting the metabolism and toxicity of a drug in humans can use resources that include in vivo animal models, in vitro systems, high throughput genomics and proteomics methods (Gerhold et al., 2001; Thomas et al., 2001) to generate empirical data for analysis and decision making. The amount and complexity of the data being generated is increasing, requiring not only judicious use of which experimental methods to use but also novel tools for visualization and analysis. More recently within drug disposition and toxicology, in vitro approaches for generating data with drug metabolizing enzymes, transporters, ion channels and receptors has been used for computational approaches including quantitative structure activity relationships (QSAR)(Ekins and Swaan, 2004). These methods have been used widely and applied for predicting absorption, distribution, metabolism, excretion and toxicity (ADME/TOX) (Ekins et al., 2005d) properties either at the level of the individual protein (e.g. CYPs (Balakin et al., 2004a; Balakin et al., 2004b)) or specific properties (e.g. absorption (Zhao et al., 2001; Niwa, 2003)). Many of these ADME/Tox proteins are also known to be regulated by nuclear hormone receptors or other transcription factors.

We are also presently witnessing the beginning of a new approach which aims to understand organisms from computationally generated networks of protein and ligand interactions (Barabasi and Oltvai, 2004). To this point, high throughput data such as that derived from microarrays has mainly been visualized by clustering approaches (Eisen et al., 1998) which limits associations to the actual genes on the microarray and provides little if any information on the relationship of the genes to each other. In contrast, network building tools such as MetaCore (Ekins et al., 2005b) enable the analysis of such data in the context of all known interactions when using a database as the source. Various software resources have been applied to modeling the networks of nuclear hormone receptors and their connections with additional genes and small molecules using a manually curated database (Ekins et al., 2005c). A second study has indicated how a natural language processing method, CCNet was used to show the genes

regulated by the nuclear hormone receptor FXR (Apic et al., 2005). These automated methods enable a more complete understanding of the complexity of these transcriptional factors (Ekins et al., 2002; Ulrich, 2003; Plant, 2004). Ultimately the pathways generated rely on the quality of the content of the underlying database of literature interactions. These networks can also be used to overlay and explain experimental data from genomic and proteomic studies to further aid in analysis of this complex data. We are hence seeing a convergence of the different methods described above to create a field we have termed systems-ADME/TOX (Ekins et al., 2005d).

We have built on the previously described efforts to generate networks of nuclear hormone interactions (Ekins et al., 2005c) as well as interpret microarray data for MCF-7 cells treated with 4-hydroxytamoxifen (OHT) and estrogen (Ekins et al., 2005c; Nikolsky et al., 2005) to develop and apply a novel method for systems-ADME/TOX (Ekins et al., 2005d). This uses a subset of the MetaCore™ database, which is considerably enhanced with the previously described key drug metabolizing enzymes, their substrates, nuclear hormone receptors and other ADME/TOX related proteins to represent the backbone of the system termed MetaDrug™. In addition we have used integrated human drug metabolism reactions (Korolev et al., 2003) and QSAR methods (Ekins et al., 2003; Balakin et al., 2004a; Balakin et al., 2004b) to enable the inference of potential interactions from an input molecular structure. These predicted interactions can also be visualized on networks alongside the empirical data and high throughput data (such as microarray) when available. As there are only a very limited number of molecules for which there is a complete published dataset for drug metabolism including characterization of the enzymes involved and microarray or other high throughput data, we were restricted to datasets with a combination of human or animal data which we recognize is far from ideal. However using mechanisms within both MetaDrug™ and MetaCore™ to map gene orthologs for different species we are able to visualize this data. This mixture of data types, sources and species also presents some difficulty for

interpretation due to the differences in metabolism and toxicity between species, but there are presently few available alternatives available unless one has the resources available in a pharmaceutical company to generate such complete datasets. We have therefore analyzed recently published data from in vitro and microarray studies as test cases with this MetaDrug™ system. This preliminary study provides examples of how the integration of a database of ADME/Tox information, metabolism rules and QSAR methods may be used to generate predictions and analyze experimental microarray data relevant to drug disposition and toxicity.

Materials and Methods

Data annotation and software programming. The development of the underlying MetaCore™ and MetaDrug™ databases (GeneGo, St Joseph, MI) have been described in detail previously (Ekins et al., 2005b; Ekins et al., 2005c; Nikolsky et al., 2005). For example, MetaDrug™ is an interactive, manually annotated database derived from literature publications on proteins and small molecules of relevance to drug disposition and toxicology in humans. MetaDrug™ was developed with an Oracle version 9.2.0.4 Standard Edition (Oracle, Redwood Shores, CA) based architecture for the representation of biological functionality and integration of functional, molecular, or clinical information (Bugrim et al., 2004). Both MetaDrug™ and MetaCore™ software runs on an Intel-based 32 bit server running RedHat Linux Enterprise 3 AS (RedHat, Raleigh, NC) and the web server ran Apache 1.3.x/mod_perl (http://perl.apache.org/start/index.html). Software on the server side was written in Perl while the client side required HTML/JavaScript and the Macromedia Flash Player Plug-in (Macromedia Inc, San Francisco, CA).

In addition to the database of ADME/Tox related proteins and small molecules within MetaDrug™ we have integrated cheminformatics tools by incorporating the Accord (Accelrys, San

Diego) Oracle plugin for searching and querying the molecular structure database. A ChemDraw ActiveX version 8 or higher plugin (CambridgeSoft, Cambridge, MA) for structure sketching is also integrated in the MetaDrug™ interface. Approximately 70 human metabolic reaction rules (Supplemental Table 1, available online) are included in MetaDrug™ and represents an expanded version of the subset of rules previously published (Korolev et al., 2003). These metabolic reactions now include many other phase I and II reactions that have been described elsewhere (Ekins et al., 2005a). The prioritization of metabolites was achieved using a modified version of the method described previously (Boyer and Zamora, 2002) in which we have used the MetaDrug™ database to calculate the occurrence frequency of metabolites relating to the reaction rules. This occurrence frequency is then assigned as a negative log value to predicted molecules, the larger the score then the higher the frequency of similar metabolites observed in our database of literature metabolic information for humans. These rules were coded in a Perl script file used by the proprietary MetaDrug™ software. The panel of QSAR models (Ekins et al., 2005a) were generated using published data for various CYPs (Korolev et al., 2003), transporters, ion channels and nuclear hormone receptors (Ekins and Swaan, 2004) gathered from many sources. This data was then used with a recursive partitioning tool ChemTree™ (GoldenHelix, Boseman, MT) (Young et al., 2002; Ekins et al., 2003) to generate the proprietary models stored in MetaDrug. The QSAR models were also validated by leaving groups out or using other external test sets. The correlation or Spearman's Rho value was then used as assessment criteria for model utilization. These QSAR models were integrated in MetaDrug™, such that after sketching a molecule or selecting a file of structures, this could then be processed to generate metabolites and QSAR predictions as defined by the user. The similarity of the input molecules to those in the individual QSAR model training sets was calculated using the Tanimoto coefficient (Willet, 2003) using Accord software. The Tanimoto coefficient is: a /(a+b+c) were a= the number of bits common to both the query and target structures, b= the number of bits exclusively in the

query structure, c= the number of bits exclusively in the target structure. In this case a value of 1 indicates the molecule is identical to one in the training set. As this value decreases, the less similar the molecule is to molecules in the training set. The proteins in the MetaDrug database that relate to the specific QSAR models e.g. CYP3A4 were linked such that predictions could then be visualized as a network of interactions radiating from that protein. ChemTree™ was also integrated into MetaDrug™ to allow the user to generate QSAR models for integration within the software from their own data. MetaDrug™ and MetaCore™ can be freely evaluated by contacting GeneGo (www.genego.com).

Generation of metabolite and QSAR model predictions Molecules were either sketched in the ChemDraw pluggin window or loaded from a mol or sdf file (Figure 1). The molecules (Figure 2) were then processed through the user defined metabolite rules and QSAR models developed with literature data (CYPs, P-gp, PXR etc) (Balakin et al., 2004a; Balakin et al., 2004b). The user can specify which metabolic reaction rules and QSAR models are used as well as specify upper and lower prediction thresholds as a means to filter the molecules prior to visualizing on networks. The previously described network building algorithms (Ekins et al., 2005c) are used for visualizing the predicted interactions of metabolites or input molecules with the related proteins in MetaDrug™. We were able to use the known molecules with metabolic pathways to test the software and predict interactions with these proteins as a network (Ekins et al., 2005a).

Visualization of microarray data on gene networks in MetaDrug. We have previously described how data from microarray studies can be imported into MetaDrug™ on the client side as a tab-delimited file (Ekins et al., 2005c). In this study we have used microarray data from livers of rats after treatment with L-742694 (structurally similar to Aprepitant) (Hartley et al., 2004), human hepatocytes treated with Trovofloxacin (Liguori et al., 2005) and G_0-arrested MCF-7 breast cancer cells treated with OHT (Hodges et al., 2003). In addition we have visualized the TaqMan real time PCR data for human hepatocytes and

LS174T cells treated with Artemisinin (Burk et al., 2005) or other artemisinin analogs (Efferth and Oesch, 2004). In most cases it is possible to extract the significant up / down regulated genes with the accession numbers from the pdf files of the publication (Hartley et al., 2004), in other cases the data could be downloaded (Hodges et al., 2003) from the NIEHS website http://dir.niehs.nih.gov/microarray/datasets/home-pub.htm) or extracted from supplemental files (Liguori et al., 2005) before it was imported into MetaDrug™. In some cases it was necessary to convert the gene identifiers from accession numbers to Locuslink identifiers using DAVID (Dennis et al., 2003). The expression intensity values could then be uploaded in MetaDrug™ to generate a gene network using the auto expand network algorithm.

The auto expand algorithm starts with a number of root nodes as specified by the user and builds sub-networks around every object from the uploaded set consisting of nearest neighbors. The expansion halts when the sub-networks intersect. The objects that do not contribute to connecting sub-networks are automatically truncated and there is no user control over the size of the network. Each connection represents a direct, experimentally confirmed, physical interaction between the objects. If a user is building a network around one node only the network generally consists of the nearest neighbors and their connections within 1-2 steps. The auto expand algorithm provides a means to look up one or more genes of interest and identify regulatory cascades that lead to or from the gene(s) of interest. These networks may become quite complex so it is likely that some filtering may be necessary to simplify the visualization.

Microarray data from rats are mapped onto the human networks using the gene ortholog information within MetaDrug™. In some cases it was also possible to visualize predicted metabolite interactions with proteins and overlay the experimental expression data simultaneously.

Visualization of microarray data on gene networks in MetaCore and statistics. MetaDrug™ contains a subset of the MetaCore™ database of manually annotated interactions as well as only 2 of the 7 currently available algorithms. Therefore we have used all of the available microarray data gene lists previously analyzed with the MetaDrug™ 'auto expand' algorithm and additionally analyzed them in MetaCore™ using the 'analyze networks' algorithm. This algorithm builds on the Dijkstra's 'shortest path' algorithm and takes a list of root nodes and for each node creates shortest paths networks to the other root nodes in the list and stops the network at a size defined by the user in the advanced options. This process is repeated iteratively until every node from the list is included in at least one network. The end-result of this is that it essentially fragments the "super-network" using the chosen nodes, down into sub-networks. Each subnetwork is associated with a Z-score, G-score and p-value which rank the sub-networks according to saturation with the objects from the initial gene list. The Z-Score ranks the Analyze Network algorithm's sub-networks in regards to their saturation with genes from the experiment. A high Z-Score means the network is highly saturated with genes from the experiment. The G-Score combines the Z-Score and the sum of the squares of the interactions to and from each of the nodes not related to the initial list. The value for the K coefficient can be specified in the advanced options section for the Analyze Network algorithm. The G-Score downgrades the Z-Score if there are high degree nodes that are not from the experiment in the sub-network. So, in general, a highly positive G-Score means the network is highly saturated with genes from the experiment and the network contains few to no high degree nodes not in the experiments; and a highly negative G-Score means there are many high degree nodes in the network that are not from the experiment. The p-Value is used to initially rank the sub-networks. The P-values throughout MetaCore - for maps, networks and processes are all calculated using the same basic formula: a hypergeometric distribution where the p-value essentially represents the probability of particular mapping arising by chance, given the numbers of genes in the set of all genes on

DMD Fast Forward. Published on December 28, 2005 as DOI: 10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.

maps/networks/processes, genes on a particular map/network/process and genes in the experiment. This function uses the same variables as the Z-Score. The equation for the Z-score, G-score and p value calculations is described below.

$$p\text{-}Value = \frac{R!\,n!\,(N-R)!\,(N-n)!}{N!}\sum_{i=\max(r,R+n-N)}^{\min(n,R)}\frac{1}{i!\,(R-i)!\,(n-i)!\,(N-R-n+i)!}$$

$$Z\text{-}score = \frac{r-n\frac{R}{N}}{\sqrt{n\left(\frac{R}{N}\right)\left(1-\frac{R}{N}\right)\left(1-\frac{n-1}{N-1}\right)}} \qquad G\text{-}score = Z\text{-}score - \frac{4}{K\cdot\sqrt{n-r}}\sum_{i\in\{n/r\}}V_i^2$$

Where:

N	- total number of nodes in MetaCore database
R	- number of the network's objects corresponding to the genes and proteins in your list
n	- total number of nodes in each small network generated from your list
r	- number of nodes with data in each small network generated from your list
V_i	- number of links to/from i-th node
{n/r}	- denotes the set of nodes in a small network that are not related to user's list
K	- user-specified coefficient - used to "demote" networks with high-degree nodes that do not correspond to genes/proteins in user's list

Gene-ontology processes are also mapped to the gene list and individual networks (see below). The analyze networks algorithm is also used with raw data to present multiple pathways that may be statistically feasible for connecting the nodes from the input list with other nodes in the database via shortest pathways. The advantage of this network is that it may find a well connected cluster of root nodes without any predefined restrictions from the user and therefore presents more flexibility in the connections possible.

Results

Generation of metabolite and QSAR model predictions. In the current study we have used the MetaDrug™ platform (Figure 1) to generate metabolism and QSAR predictions for Aprepitant, L-742694, Trovofloxacin, OHT and Artemisinin. With Aprepitant, MetaDrug™ produced 37 possible metabolites including 1 of the 4 initial metabolites recently identified, namely the major initial metabolite Aprepitant M1 (predicted log OC -1). Aprepitant was predicted to have a relatively high affinity for CYP3A4 K_m (predicted, 15 μM; similarity score = 0.78) CYP3A4 Ki (predicted, 13.5 μM; similarity score = 0.78) and interact with PXR (predicted to bind, probability = 0.89 similarity score = 0.77). In the latter case using this classification model values closer to 1 represent PXR ligands and values closer to 0 are unlikely to be PXR ligands. In contrast 36 possible metabolites were predicted for L-742694 which was indicated to have an affinity for CYP3A4 K_m (predicted, 14.8 μM; similarity score = 0.75) CYP3A4 Ki (predicted, 8.1 μM; similarity score = 0.76) and interact with PXR (predicted to bind, probability = 0.58 similarity score = 0.77). MetaDrug produced 34 metabolites for Trovafloxacin including the major circulating metabolite Trovafloxacin glucuronide M1 (predicted log OC - 0.9), acetylated Trovafloxacin M3 (-1.3) and Trovafloxacin sulphate M4 (predicted log OC – 2.46). MetaDrug produced 28 metabolites for OHT including Endoxifen (predicted log OC -1) and 3,4-dihydroxytamoxifen (predicted log OC -0.97). OHT had a relatively high affinity for CYP3A4 K_m (predicted, 55 μM; similarity score = 1), is also an inhibitor for P-gp (predicted, 15.1 μM; similarity score = 0.81), is further metabolized by phenol and estrogen sulfotransferases (SULT1A1K_m model predicted 17.4 μM, similarity 1) and is an inducer of PXR (predicted to bind, probability 0.90, similarity score = 1). MetaDrug produced 17 metabolites for Artemisinin including dihydroartemisinin (Log OC -5) and many hydroxylated metabolites (log OC -0.81). In addition predictions with various QSAR models indicated that CYP2B6 (13.4 μM, similarity

0.63), CYP3A4 (60.2 μM, similarity 0.7), PXR (predicted to bind, probability = 0.94 similarity 0.63), P-gp substrate (0.53, similarity 0.7), P-gp inhibitor (64.5 μM, similarity 0.7) were important.

Visualization of microarray data on gene networks in MetaDrug. Microarray data from rat livers after treatment with L-742694 was uploaded into MetaDrug and 17 out of 44 gene identifiers mapped in the database (Hartley et al., 2004). These genes had been identified from a 25K rat microarray were a p-value (< 0.01) and magnitude of regulation (log 10 ratio > 0.5) were considered significant (Hartley et al., 2004). After generating a network this data was visualized alongside the predicted interactions derived computationally from the QSAR models for the similar molecule Aprepitant (Figure 3). In this case CYP3A4 was upregulated in the experiment with L-742694 indicative of induction and this enzyme was also predicted to metabolize Aprepitant. Microarray data from human hepatocytes treated with Trovafloxacin (Liguori et al., 2005) was uploaded into MetaDrug and 87 out of 141 gene identifiers mapped in the database. These genes had been selected from an Affymetrix human genome U133A array were the p-value less than or equal to 0.01 was considered significant (Liguori et al., 2005). In this case the microarray data does not appear to directly impact drug metabolism (data not shown). Microarray data from human MCF-7 cells treated with OHT (Hodges et al., 2003) was uploaded in to MetaDrug and 1446 out of 1617 gene identifiers mapped in the database (Supplemental Figure 1A). This data was obtained raw from the microarray studies previously published (Hodges et al., 2003). Both CYP3A4 and P-gp (MDR1) were up regulated in this dataset, indicating that this molecule may regulate its own transport and metabolism. TaqMan real time PCR data for human hepatocytes and LS174T cells treated with Artemisinin demonstrated induction of CYP3A4, CYP2B6 and MDR1 (Burk et al., 2005). This small dataset of uploaded genes was uploaded after assigning LocusLink identifiers. 3 of 3 genes were accessible (Supplemental Figure 1B). A second dataset used the NCI mRNA expression data in 60 cell lines after treatment with Artemisinin and related compounds (Efferth and Oesch, 2004). These genes

DMD Fast Forward. Published on December 28, 2005 as DOI: 10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.

were selected from an initial list of 170 genes following cluster analysis into 3 groups using the complete linkage method, were mRNA expression correlated with cytotoxicity IC_{50} data (Efferth and Oesch, 2004). The genes in cluster 1 and 2 were uploaded in MetaDrug and 28 of 36 genes mapped in the database some of which were present on the network of predicted interactions (Supplemental Figure 1C).

Visualization of microarray data on gene networks in MetaCore. The same microarray datasets uploaded in MetaDrug™ previously, were evaluated in MetaCore™ with the analyze network algorithm. The most statistically significant network based on the p-value (in parenthesis) as described above was generated in all cases. For L-742694 17 of 44 genes were uploaded and this gene list mapped onto the following GO processes using an approach similar to the EASE tool (Hosack et al., 2003): xenobiotic metabolism ($9.12e^{-13}$), steroid metabolism ($8.84e^{-12}$), electron transport ($8.53e^{-11}$), lipid metabolism ($1.98e^{-07}$), icosanoid metabolism ($4.79e^{-07}$), metabolism ($3.00e^{-06}$), estrogen catabolism ($1.02e^{-05}$), retinal metabolism ($2.03e^{-05}$), retinoic acid metabolism ($3.38e^{-05}$) and aldehyde metabolism ($5.07e^{-05}$). A network was generated from this gene list (Figure 4A, $p = 6.18e^{-36}$, Z-score 20.33) which in turn mapped to the following GO processes ; electron transport ($1.058e^{-20}$), steroid metabolism ($7.31e^{-10}$), xenobiotic metabolism ($1.04e^{-08}$), icosanoid metabolism ($1.94e^{-08}$), lipid metabolism ($8.51e^{-07}$), eye morphogenesis ($1.28e^{-05}$), drug metabolism ($1.32e^{-05}$), regulation of heart contraction ($6.57e^{-05}$), arachidonic acid metabolism ($7.89e^{-05}$), retinal metabolism ($7.89e^{-05}$), retinoic acid metabolism ($1.31e^{-04}$) and vitamin biosynthesis ($1.31e^{-04}$). The xenobiotic metabolism GO process was also highlighted on this network (Figure 4B).

For the Trovafloxacin dataset, 87 of 141 genes were uploaded and this gene list mapped onto the following GO processes signal transduction ($4.37e^{-09}$),protein-nucleus import, docking ($7.56e^{-07}$), intracellular signaling cascade ($3.88e^{-05}$), cyclic nucleotide metabolism ($8.18e^{-05}$), thioredoxin pathway ($1.02e^{-04}$), cGMP-mediated signaling ($3.04e^{-04}$), cGMP catabolism ($3.05e^{-04}$), germ cell migration ($3.05e^{-}$

DMD Fast Forward. Published on December 28, 2005 as DOI: 10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.

^{04}), protein transport (3.05e^{-04}) and protein-nucleus import (3.28e^{-04}). A network was generated from this gene list (Supplemental Figure 2A, p = 4.50e^{-44}, Z-score 21.36) which in turn mapped to the following GO processes: intracellular signaling cascade (9.453e^{-09}), signal transduction (7.398e^{-07}), calcium ion transport (1.808e^{-06}), dephosphorylation (6.793e^{-06}), protein kinase C activation (9.509e^{-06}), T-cell differentiation (3.408e^{-05}), regulation of cell cycle (8.594e^{-05}), G1/S transition of mitotic cell cycle (3.18e^{-04}), learning and/or memory (4.11e^{-04}) and transcription, DNA-dependent (4.11e^{-04}). The intracellular signaling cascade GO process was mapped on the network (Supplemental Figure 2B).

For the OHT dataset, 1358 of 1617 genes were uploaded and this gene list was mapped onto the following GO processes: protein amino acid phosphorylation (8.54e^{-59}), regulation of cell cycle (1.76e^{-53}), signal transduction (6.87e^{-45}), regulation of transcription, DNA-dependent (2.33e^{-33}), cell proliferation (1.16e^{-23}), DNA repair (3.16e^{-18}), organogenesis (1.61e^{-15}), cell surface receptor linked signal transduction (1.67e^{-15}), protein amino acid dephosphorylation (3.09e^{-15}) and cell adhesion (3.35e^{-15}). A network was generated from this gene list (Supplemental Figure 3, p = 5.19e^{-42}, Z-score 14.98) which inturn mapped to the following GO processes: regulation of cell cycle (1.81e^{-06}), positive regulation of cell proliferation (6.72e^{-06}), ER-overload response (1.51e^{-04}), induction of positive chemotaxis (1.51e^{-04}), collagen catabolism (4.11e^{-04}), chemotaxis (4.57e^{-04}), ovulation (5.23e^{-04}), response to unfolded protein (6.61e^{-04}), leading edge cell differentiation (8.90e^{-04}), cell motility (1.33e^{-03}).

For the Artmesinin dataset, 28 of 36 genes were uploaded and this gene list was mapped onto the following GO processes: electron transport (3.71e^{-09}), icosanoid metabolism (2.48e^{-04}), regulation of signal transduction (3.09e^{-04}), generation of precursor metabolites and energy (3.74e^{-04}), protein metabolism (4.52e^{-04}), protein targeting (8.07e^{-04}), cholesterol biosynthesis (1.16e^{-03}), steroid biosynthesis (1.56e^{-03}), response to stress (2.64e^{-03}) and positive regulation of cytotoxic T-cell differentiation (2.69e^{-03}). A network was generated from this gene list (Supplemental Figure 4A, p = 2.81e^{-41}, Z-score 30.67) which

in turn mapped to the following GO processes: electron transport (1.80e-06), leading edge cell differentiation ($1.08e^{-05}$), regulation of cell cycle ($1.01e^{-04}$), pentose-phosphate shunt ($1.57e^{-04}$), DNA repair ($5.01e^{-04}$), icosanoid metabolism ($9.25e^{-04}$), nucleotide metabolism ($1.40e^{-03}$), cell proliferation ($1.76e^{-03}$), response to oxidative stress ($2.16e^{-03}$), base-excision repair ($2.65e^{-03}$), negative regulation of protein kinase activity ($2.65e^{-03}$) and iron ion homeostasis (3.01e-03). The oxidative stress GO process was mapped onto this network (Supplemental Figure 4B).

Discussion

The human physiology represents a complex system, therefore it is difficult when attempting to understand the changes in gene, protein or biological activity upon challenge with a drug (Nicholson et al., 2004). While we would not advocate that this is an impossible task to try to predict, there have been an increasing number of extensive toxicogenomics datasets that have been published from which we can learn. These studies have generally applied clustering methods to interpret the data. Example datasets for molecules evaluated in this way either in vivo in animal species, isolated cells or cell lines are accumulating, resulting in the need for accessible databases and tools for expedited analysis, data mining and hypothesis generation. The MetaDrug™ database uses a novel architecture described previously (Ekins et al., 2005c), allowing the organization and visualization of biological and chemical information. This approach generates networks which are more informative than linear pathways as commonly described in biology text books. We have applied this database platform previously to visualize the nuclear hormone receptor interactions, providing a qualitative understanding of the current state of knowledge (Ekins et al., 2005c). We have used a second platform MetaCore which includes more network building algorithms to visualize microarray data for MCF-7 cells treated with OHT and estrogen for 24h

(Hodges et al., 2003). This enabled us to generate, compare and subtract networks to produce signature gene-networks for each molecule (Nikolsky et al., 2005).

In the current study we have extended the MetaDrug platform beyond a database of ADME/TOX related content to further include a rule based method to generate predicted metabolites (Korolev et al., 2003) and QSAR methods for predicting interactions with ADME/TOX related proteins (Balakin et al., 2004a; Balakin et al., 2004b; Ekins and Swaan, 2004) as well as other physicochemical properties (Ekins et al., 2005a). Recent testing of MetaDrug with 66 molecules (Supplemental Table 2, available online) indicates we capture at least 78.8 % (on average) of correctly predicted first pass metabolites. Much earlier testing on 28 of these molecules indicated approximately 73 % of metabolites were correctly identified (Ekins et al., 2005a). However there is still room for future improvement to minimize the number of total metabolites predicted, as well as the addition of further rules for metabolic reaction not currently captured (e.g. conjugation at the selenium atom in Ebselen). The metabolism predictions for Aprepitant, L-742694, Trovofloxacin, 4-hydroxytamoxifen and artemisinin and artemisinin, outlined earlier included the prediction of phase II metabolites. MetaDrug therefore represents a systems-ADME/TOX platform for the prediction of metabolism and interactions from molecular structure as well as the visualization and simultaneous analysis of multiple high throughput data types (such as microarrays) (Ekins et al., 2005d). As such, the approach is highly novel, integrating not only different algorithms for predictions but also data parsers, algorithms for network generation, visualization tools and data filters. This latter component enables the selection of genes associated with a specific species, tissue or organelle, for example.

To demonstrate the utility of such a platform we have analyzed several recently published datasets from in vitro and/or microarray studies for Aprepitant (Sanchez et al., 2004), L-742694 (Hartley et al., 2004), Trovofloxacin (Dalvie et al., 1997; Liguori et al., 2005), OHT (Crewe et al., 1997; Chen et al.,

2002; Desai et al., 2002; Hodges et al., 2003; Bekaii-Saab et al., 2004; Desta et al., 2004) and Artemisinin (Svensson and Ashton, 1999; Svensson et al., 2003; Efferth and Oesch, 2004; Burk et al., 2005). In all cases the data was taken from the papers following clustering or other statistical pre-analysis. However we have previously described the analysis of 'raw' microarray data without the need for clustering or other similar approaches (Nikolsky et al., 2005). We have used the molecular structure of the test molecules described above with MetaDrug generated metabolites, predictions for affinity to multiple ADME/TOX related proteins and auto-expand gene-networks of these predicted protein interactions alongside available microarray data. These gene networks were compared with an analysis performed with MetaCore™ using the analyze network algorithm with all available microarray data for each compound after the clustering performed in the published papers.

MetaDrug™ produced 37 possible metabolites for Aprepitant including a major initial metabolite Aprepitant M1. Aprepitant was predicted to have a relatively high affinity for CYP3A4 K_m (15 μM) which is comparable to the actual value (~10 μM). Similarly the predicted CYP3A4 K_i (13.5 μM) is close to the actual value (10 μM) and the predicted interaction with PXR may indicate that this is responsible for the induction of CYP3A4 as described in the package insert for this drug (http://www.fda.gov/cder/foi/label/2003/21549_Emend_lbl.pdf). A structurally similar drug L-742694 was shown to activate rat PXR (Hartley et al., 2004) and transcriptional profiling induced a battery of genes involved in drug metabolism and transport. These genes regulated by L-742694 are mapped on the human ortholog network derived from the QSAR predictions for Aprepitant. The high Tanimoto similarity values derived for Aprepitant and L-742694 for the CYP3A4 K_m, CYP3A4 IC_{50} and PXR compared to the training sets is indicative of a high degree of structural similarity with molecules in these training sets. The microarray data from rats treated with L-742694 (Figure 3A) enabled the visualization of upregulated CYP3A4, CYP3A4 is known to be induced by Aprepitant which is in turn metabolized by

this enzyme. The QSAR models predicted the role of this enzyme and incidentally predicted that this molecule may also bind PXR with a higher probability than the structurally similar L-742694. However there is currently no published indication whether Aprepitant binds to PXR.

MetaDrug™ produced 34 metabolites for Trovafloxacin including 3 of the 4 metabolites described in the literature (Dalvie et al., 1997) Trovafloxacin glucuronide M1, acetylated Trovafloxacin M3 and Trovafloxacin sulphate M4. Microarray data from human hepatocytes treated with Trovafloxacin was uploaded in to MetaDrug and 87 out of 141 gene identifiers mapped in the database. In this case the microarray data does not appear to directly impact drug metabolism pathways. MetaDrug™ produced 28 metabolites for OHT including Endoxifen and 3,4-dihydroxytamoxifen. OHT had a relatively high affinity for CYP3A4 K_m while the similarity score indicated this molecule is in the training set of the model (Desta et al., 2004). OHT is also an inhibitor for P-gp with a predicted IC_{50} (15.1 μM) quite similar to the actual value (7.4 μM) (Bekaii-Saab et al., 2004). OHT is known to be further metabolized via phenol and estrogen sulfotransferases and in this case the SULT1A1K_m model predicted a value of 17.4 μM while the similarity calculation indicated that this molecule is in this training set (Chen et al., 2002). OHT was predicted to bind to PXR and once again was indicated to be present in the training set (Desai et al., 2002). Binding to PXR would be expected to increase levels of CYP3A4 that is in turn involved in OHT formation from Tamoxifen (Crewe et al., 1997). Microarray data from human MCF-7 cells treated with OHT was uploaded and was mapped in the MetaDrug database (Supplemental Figure 1A). Both CYP3A4 and P-gp (MDR1) were up-regulated in this dataset, once again indicating that this molecule may regulate its own transport and metabolism. MetaDrug™ produced 17 metabolites for Artemisinin including one of the known metabolites observed in human plasma, namely dihydroartemisinin (Svensson and Ashton, 1999). Little else is known regarding the human in vitro metabolism of this compound. It is possible that CYP2B6 could be responsible for forming this metabolite as this occurs in the same location

DMD Fast Forward. Published on December 28, 2005 as DOI: 10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.

as the O-deethylation of β-arteether (Grace et al., 1998) which is mediated by the same enzyme. The hydroxylation of artemisinin may also be mediated by CYP2B6 which is known to be involved in numerous metabolic reactions (Ekins and Wrighton, 1999). In addition, predictions with the various QSAR models indicated that artemisinin binds CYP2B6, CYP3A4, PXR, and is unlikely to be a P-gp substrate (but is a weak P-gp Inhibitor). These predictions are in very good agreement with the metabolism data (Svensson and Ashton, 1999; Li et al., 2003) while recent studies suggested artemisinin binds PXR with an EC_{50} of 34 μM and inhibits P-gp mediated digoxin transport with an IC_{50} of 33 μM (Burk et al., 2005). This same study generated PCR data with human hepatocytes treated with artemisinin to show the induction of CYP3A4, CYP2B6 and P-gp (Burk et al., 2005). The gene expression data can be visualized alongside the predicted interactions to show the other transcriptional regulators of these proteins (Supplemental Figure 1B). A second data set derived from the NCI cell lines treated with artemisinin analogs and clustered (Efferth and Oesch, 2004) was also overlapped on the same network (Supplemental Figure 1C).

In 3 of the 4 cases presented we were able to visualize gene expression data alongside the predicted interactions in MD using the autoexpand algorithm. Using a second platform MetaCore, we were able to use a different network building algorithm, namely ‘analyze network’ which provides multiple significant small scale networks with statistical significance and enables the mapping of Gene Ontology data. The network with the most significant p-value was then generated in all cases. These networks do not allow the user to generate predicted molecules on the networks as in MetaDrug™, but they do provide considerable insight into the significance of the gene expression data. The L-742694 gene expression data set from rat liver was mapped on the human orthologs in MetaCore™ and indicated a significant link with metabolism (Figure 4A, B) as the data mapped to the metabolism based GO processes. This corresponds with the observation that L-742694 impacts the PXR responsive gene battery

and the structurally similar Aprepitant is metabolized by CYP3A4 as well as other CYPs (Hartley et al., 2004; Sanchez et al., 2004). The Trovafloxacin dataset, was linked with many GO processes from signal transduction to protein transport likely to be involved as part of an oxidative stress response (Supplemental Figure 2A,B) which perhaps strengthens the observations made after clustering the human hepatocyte gene expression data (Liguori et al., 2005). The OHT dataset mapped onto the GO processes related to the cell cycle (Supplemental Figure 3) which corresponds well with the microarray data from MCF-7 cells (Hodges et al., 2003). The Artmesinin analog microarray dataset was linked to metabolism, cell cycle and oxidative stress GO processes (Supplemental Figure 4) correlating well with the observations of cytotoxicity observed from clustering the NCI cell line data (Efferth and Oesch, 2004), and metabolism in human cells (Burk et al., 2005). The ability to highlight the genes involved with each GO process on the networks is a valuable approach for quickly identifying their position and relationships with other genes on the network. It is important to note that although we are able to map a large number of the genes uploaded into the software either directly or using DAVID, there will be future improvements in the database or using different identifiers instead of LocusLink (e.g. HomoloGene), which will enable more genes to be visualized. Another important consideration is the potential for species differences in receptor binding, metabolism and toxicity. One example we have used previously as an example to illustrate the utility of building species specific gene networks is the drug pyrazinamide. This blocks NAD+ metabolism to result in the accumulation of the toxic uric acid metabolite in humans but not mice (Bugrim et al., 2004). We have also previously indicated how MetaDrug can be used to simulate the effect of knock out or inhibition of a target gene, by simply removing it from a network. This would then open the possibility to allow the user to consider species differences or different genotypes.

From these test cases it is apparent that although we can suggest the majority of the major metabolites for these compounds we either do not identify some others or predict metabolites that have

not been identified to date. This will need to be rectified in future by allowing the user to add their own reaction rules to the software to generate metabolites currently missing. Over prediction will require the development of a machine learning algorithm based on the available human metabolism data or an expansion of the rules for metabolic pathways using reasoning (Button et al., 2003) or alternative approaches. The availability of multiple QSAR models for particular ADME properties is currently limited to published literature available to us. The capability to generate QSAR models with the software included in MetaDrug™ (not discussed here), will allow the user to incorporate their own data for ADME/Tox properties or therapeutic targets whether based around a single or multiple structural series. There is considerable flexibility in the user being able to add their own biological data (e.g. in vitro screening data, or unpublished protein interactions or other biological knowledge) into the Oracle™ database structure to further customize MetaDrug™.

In summary, we have developed and applied additional utilities that have been added to the MetaDrug™ software platform. This software suite now incorporates reaction rules for metabolite prediction, QSAR models and visualization tools in addition to a database of manually curated human ADME/Tox data. We have used this software to generate predictions for several drug-like molecules, and additionally we have visualized the experimental gene expression data for L-742694, Trovofloxacin, OHT and artemisinin using this software. Networks were also generated with MetaCore™ and the 'analyze network' algorithm to further aid in the construction of statistically significant visualizations of the gene expression data correlated with GO processes. The 'analyze network' algorithm has also been recently added to MetaDrug™ due to this demonstrated utility. The MetaDrug™ system represents a prototype for integrative or systems-ADME/Tox that builds on the database and network building tools such as MetaCore™. In future we will likely test more compounds not included in the software training sets as this data becomes available in the literature. At present we are quite limited for testing the software

relying on molecules and microarray data in the public domain and envisage that future NIH efforts to collate such information (Waters et al., 2003) will improve future modeling and validation studies in this area. The present version of MetaDrug^{TM} is focused on predicting human drug metabolism and interactions with ADME/Tox proteins although it is likely that future versions will be required to enable predictions for other species.

References

Apic G, Ignjatovic T, Boyer S and Russell RB (2005) Illuminating drug discovery with biological pathways. *FEBS Lett* **579:**1872-1877.

Balakin KV, Ekins S, Bugrim A, Ivanenkov YA, Korolev D, Nikolsky Y, Ivashchenko AA, Savchuk NP and Nikolskaya T (2004a) Quantitative structure-metabolism relationship modeling of the metabolic N-dealkylation rates. *Drug Metab Dispos* **32:**1111-1120.

Balakin KV, Ekins S, Bugrim A, Ivanenkov YA, Korolev D, Nikolsky Y, Skorenko SA, Ivashchenko AA, Savchuk NP and Nikolskaya T (2004b) Kohonen maps for prediction of binding to human cytochrome P450 3A4. *Drug Metab Dispos* **32:**1183-1189.

Barabasi A-L and Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5:**101-113.

Bekaii-Saab TS, Perloff MD, Weemhoff JL, Greenblatt DJ and von Moltke LL (2004) Interactions of tamoxifen, N-desmethyltamoxifen and 4-hydroxytamoxifen with P-glycoprotein and CYP3A. *Biopharm Drug Dispos* **25:**283-289.

Boyer S and Zamora I (2002) New methods in predictive metabolism. *J Comp-Aided Mol Des* **16:**403-413.

Bugrim A, Nikolskaya T and Nikolsky Y (2004) Early prediction of drug metabolism and toxicity: systems biology approach and modeling. *Drug Discovery Today* **9:**127-135.

Burk O, Arnold KA, Nussler AK, Schaeffeler E, Efimova E, Avery BA, Avery MA, Fromm MF and Eichelbaum M (2005) Antimalarial Artemisinin Drugs Induce Cytochrome P450 and MDR1

Expression by Activation of Xenosensors Pregnane X Receptor and Constitutive Androstane Receptor. *Mol Pharmacol* **67:**1954-1965.

Button WG, Judson PN, Long A and Vessey JD (2003) Using absolute and relative reasoning in the prediction of the potential metabolism of xenobiotics. *J Chem Inf Compu Sci* **43:**1371-1377.

Chen G, Yin S, Maiti S and Shao X (2002) 4-Hydroxytamoxifen sulfation metabolism. *J Biochem Mol Toxicol* **16**.

Crewe HK, Ellis SW, Lennard MS and Tucker GT (1997) Variable contribution of cytochromes P450 2D6, 2C9 and 3A4 to the hydroxylation of tamoxifen by human liver microsomes. *Biochem Pharmacol* **53:**171-178.

Dalvie DK, Khosla N and Vincent J (1997) Excretion and metabolism of trovafloxacin in humans. *Drug Metab Dispos* **25:**423-427.

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC and Lempicki RA (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4:**P3.

Desai PB, Nallani SC, Sane RS, Moore LB, Goodwin BJ, Buckley DJ and Buckley AR (2002) Induction of cytochrome P450 3A4 in primary human hepatocytes and activation of human pregnane X receptor by tamoxifen and 4-hydroxytamoxifen. *Drug Metab Dispos* **30:**608-612.

Desta Z, Ward BA, Soukhova NV and Flockhart DA (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.

Efferth T and Oesch F (2004) Oxidative stress response of tumor cells: microarray-based comparison between artemisinins and anthracyclines. *Biochem Pharmacol* **68:**3-10.

Eisen MB, Spellman PT, Brown PO and Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95:**14863-14868.

Ekins S, Andreyev S, Ryabov A, Kirilov E, Rakhmatulin EA, Bugrim A and Nikolskaya T (2005a) Computational Prediction of Human Drug Metabolism. *Exp Opin Drug Metab Toxicol* **1:**303-324.

Ekins S, Berbaum J and Harrison RK (2003) Generation and validation of rapid computational filters for CYP2D6 and CYP3A4. *Drug Metab Dispos* **31:**1077-1080.

Ekins S, Bugrim A, Nikolsky Y and Nikolskaya T (2005b) Systems biology: applications in drug discovery, in: *Drug discovery handbook* (Gad SC ed), pp 123-183, Wiley, New York.

Ekins S, Kirillov E, Rakhmatulin E and Nikolskaya T (2005c) A novel method for visualizing nuclear hormone receptor networks relevant to drug metabolism. *Drug Metab Dispos* **33:**474-481.

Ekins S, Mirny L and Schuetz EG (2002) A ligand-based approach to understanding selectivity of nuclear hormone receptors PXR, CAR, FXR, LXRa and LXRb. *Pharm Res* **19:**1788-1800.

Ekins S, Nikolsky Y and Nikolskaya T (2005d) Techniques: Application of Systems Biology to Absorption, Distribution, Metabolism, Excretion, and Toxicity. *Trends Pharmacol Sci* **26:**202-209.

Ekins S and Swaan PW (2004) Computational models for enzymes, transporters, channels and receptors relevant to ADME/TOX. *Rev Comp Chem* **20:**333-415.

Ekins S and Wrighton SA (1999) The role of CYP2B6 in human xenobiotic metabolism. *Drug Metab Rev* **31:**719-754.

Gerhold D, Lu M, Xu J, Austin C, Caskey CT and Rushmore T (2001) Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiol Genomics* **5:**161-170.

Grace JM, Aguilar AJ, Trotman KM, Peggins JO and Brewer TG (1998) Metabolism of beta-arteether to dihydroqinghaosu by human liver microsomes and recombinant cytochrome P450. *Drug Metab Dispos* **26:**313-317.

Hartley DP, Dai X, He YD, Carlini EJ, Wang B, Huskey SE, Ulrich RG, Rushmore TH, Evers R and Evans DC (2004) Activators of the rat pregnane X receptor differentially modulate hepatic and intestinal gene expression. *Mol Pharmacol* **65:**1159-1171.

Hodges LC, Cook JD, Lobenhofer EK, Li L, Bennett L, Bushel PR, Aldaz CM, Afshari CA and Walker CL (2003) Tamoxifen functions as a molecular agonist inducing cell cycle-associated genes in breast cancer cells. *Mol Cancer Res* **1:**300-311.

Hosack DA, Dennis G, Jr., Sherman BT, Lane HC and Lempicki RA (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol* **4:**R70.

Korolev D, Balakin KV, Nikolsky Y, Kirillov E, Ivanenkov YA, Savchuk NP, Ivashchenko AA and Nikolskaya T (2003) Modeling of human cytochrome p450-mediated drug metabolism using unsupervised machine learning approach. *J Med Chem* **46:**3631-3643.

Li XQ, Bjorkman A, Andersson TB, Gustafsson LL and Masimirembwa CM (2003) Identification of human cytochrome P(450)s that metabolise anti-parasitic drugs and predictions of in vivo drug hepatic clearance from in vitro data. *Eur J Clin Pharmacol* **59:**429-442.

Liguori MJ, Anderson LM, Bukofzer S, McKim J, Pregenzer JF, Retief J, Spear BB and Waring JF (2005) Microarray analysis in human hepatocytes suggests a mechanism for hepatotoxicity induced by trovafloxacin. *Hepatology* **41:**177-186.

Nicholson JK, Holmes E, Lindon JC and Wilson ID (2004) The challenges of modeling mammalian biocomplexity. *Nat Biotechnol* **22:**1268-1274.

Nikolsky Y, Ekins S, Nikolskaya T and Bugrim A (2005) A novel method for generation of signature networks as biomarkers from complex high throughput data. *Tox Lett* **158:**20-29.

Niwa T (2003) Using general regression and probalistic neural networks to predict human intestinal absorption with topological descriptors derived from two-dimensional chemical structures. *J Chem Inf Comp Sci* **43:**113-119.

Plant N (2004) Interaction networks: coordinating responses to xenobiotic exposure. *Toxicology* **202:**21-32.

Sanchez RI, Wang RW, Newton DJ, Bakhtiar R, Lu P, Chiu SH, Evans DC and Huskey SE (2004) Cytochrome P450 3A4 is the major enzyme involved in the metabolism of the substance P receptor antagonist aprepitant. *Drug Metab Dispos* **32:**1287-1292.

Svensson US and Ashton M (1999) Identification of the human cytochrome P450 enzymes involved in the in vitro metabolism of artemisinin. *Br J Clin Pharmacol* **48:**528-535.

Svensson US, Maki-Jouppila M, Hoffmann KJ and Ashton M (2003) Characterisation of the human liver in vitro metabolic pattern of artemisinin and auto-induction in the rat by use of nonlinear mixed effects modelling. *Biopharm Drug Dispos* **24:**71-85.

Thomas RS, Rank DR, Penn SG, Zastrow GM, Hayes KR, Pande K, Glover E, Silander T, Craven MW, Reddy JK, Jovanovich SB and Bradfield CA (2001) Identification of toxicologically predictive gene sets using cDNA microarrays. *Mol Pharmacol* **60:**1189-1194.

Ulrich RG (2003) The toxicogenomics of nuclear receptor agonists. *Curr Opin Chem Biol* **7:**505-510.

Waters M, Boorman G, Bushel P, Cunningham M, Irwin R, Merrick A, Olden K, Paules R, Selkirk J, Stasiewicz S, Weis B, Van Houten B, Walker N and Tennant R (2003) Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. *EHP Toxicogenomics* **111:**15-28.

Willet P (2003) Similarity-based approaches to virtual screening. *Biochem Soc Trans* **31:**603-606.

Young SS, Gombar VK, Emptage MR, Cariello NF and Lambert C (2002) Mixture deconvolution and analysis of Ames mutagenicity data. *Chemo Intell Lab Sys* **60:**5-11.

Zhao YH, Le J, Abraham MH, Hersey A, Eddershaw PJ, Luscombe CN, Butina D, Beck G, Sherborne B, Cooper I, Platts JA and Boutina D (2001) Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J Pharm Sci* **90:**749-784.

Funding and Competing Financial Interest

The development of MetaDrug™ was supported by a National Institutes of Health Grant 1-R43-GM069124-01 and 2-R44-GM069124-02 “In silico Assessment of Drug Metabolism and Toxicity”. MetaDrug and MetaCore are proprietary tools developed and licensed by GeneGo, Inc. All authors are employees of GeneGo Inc.

Figure legends

Figure 1. The process flow for using MetaDrug™. HT data = Microarray type data but could also be used with in vitro screening data.

Figure 2. Molecules tested in this study with metabolites discussed in the text. Those highlighted in bold were predicted by MetaDrug™.

Figure 3. Network visualizations. Aprepitant (pink hexagon) predicted interactions and microarray data for L-742694. Microarray expression data are overlaid on the key proteins involved in drug metabolism following rat treatment with L-742694 (Hartley et al., 2004)(red circles = up regulation). Interactions types between nodes on the network are hidden for purposes of clarity. Highlighted lines show predicted interactions. Ligands (purple hexagons) linked to transfactors (red), enzymes (yellow arrows) and transporters (blue) from the MetaDrug™ database. Highlighted lines show predicted interactions. Small colored hexagons on network edges represent functional interactions (green – positive, red- negative, black –unspecified effects).

Figure 4. A. Network visualization generated with MetaCore™ and the Analyze Networks algorithm using microarray expression data overlaid on proteins following rat treatment with L-742694 (Hartley et al., 2004)(red circles = up regulation). B. GO process highlighted for xenobiotic metabolism on Figure 3A network. Small colored hexagons on network edges represent functional interactions (green – positive, red- negative, black –unspecified effects).

DMD Fast Forward. Published on December 28, 2005 as DOI: 10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.

Supplemental Figures (available online along with supplemental tables at journal website only)

Supplemental Figure 1. Network visualizations (A) Network for OHT (pink hexagon) with microarray expression data are overlaid on the key proteins involved in drug metabolism, from published studies with human MCF-7 cells (Hodges et al., 2003) Effects are hidden for clarity. (B) Network for Artmisinin (pink hexagon) using PCR data from human hepatocytes (red circles) (Burk et al., 2005). Small molecules are hidden from this network for clarity. (C) Network for Artemisinin (pink hexagon) using Microarray data from the NCI cell lines (Efferth and Oesch, 2004). Small molecules are hidden from this network for clarity Ligands (purple hexagons) linked to transfactors (red), enzymes (yellow arrows) and transporters (blue) from the MetaDrug™ database. Highlighted lines show predicted interactions. Small colored hexagons on network edges represent functional interactions (green – positive, red- negative, black – unspecified effects).

Supplemental Figure 2 A. Network visualization generated with MetaCore™ and the Analyze Networks algorithm using microarray expression data overlaid on proteins following human hepatocyte treatment with Trovafloxacin (Liguori et al., 2005). B. GO processes highlighted for intracellular signaling cascade on Figure 4A. Red circles = up regulation, blue circles = down regulation. Small colored hexagons on network edges represent functional interactions (green – positive, red- negative, black –unspecified effects).

Supplemental Figure 3 A. Network visualization generated with MetaCore™ and the Analyze Networks algorithm using microarray expression data overlaid on proteins following MCF-7 cell treatment with 4-OHT (Hodges et al., 2003). Small colored hexagons on network edges represent functional interactions (green – positive, red- negative, black –unspecified effects).

DMD Fast Forward. Published on December 28, 2005 as DOI: 10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.

Supplemental Figure 4 A. Network visualization generated with MetaCore™ and the Analyze Networks algorithm using microarray expression data overlaid on proteins following cell line treatment with Artemisin analogs (Efferth and Oesch, 2004). B GO processes highlighted for oxidative stress on Figure 7A. Small colored hexagons on network edges represent functional interactions (green – positive, red-negative, black –unspecified effects).

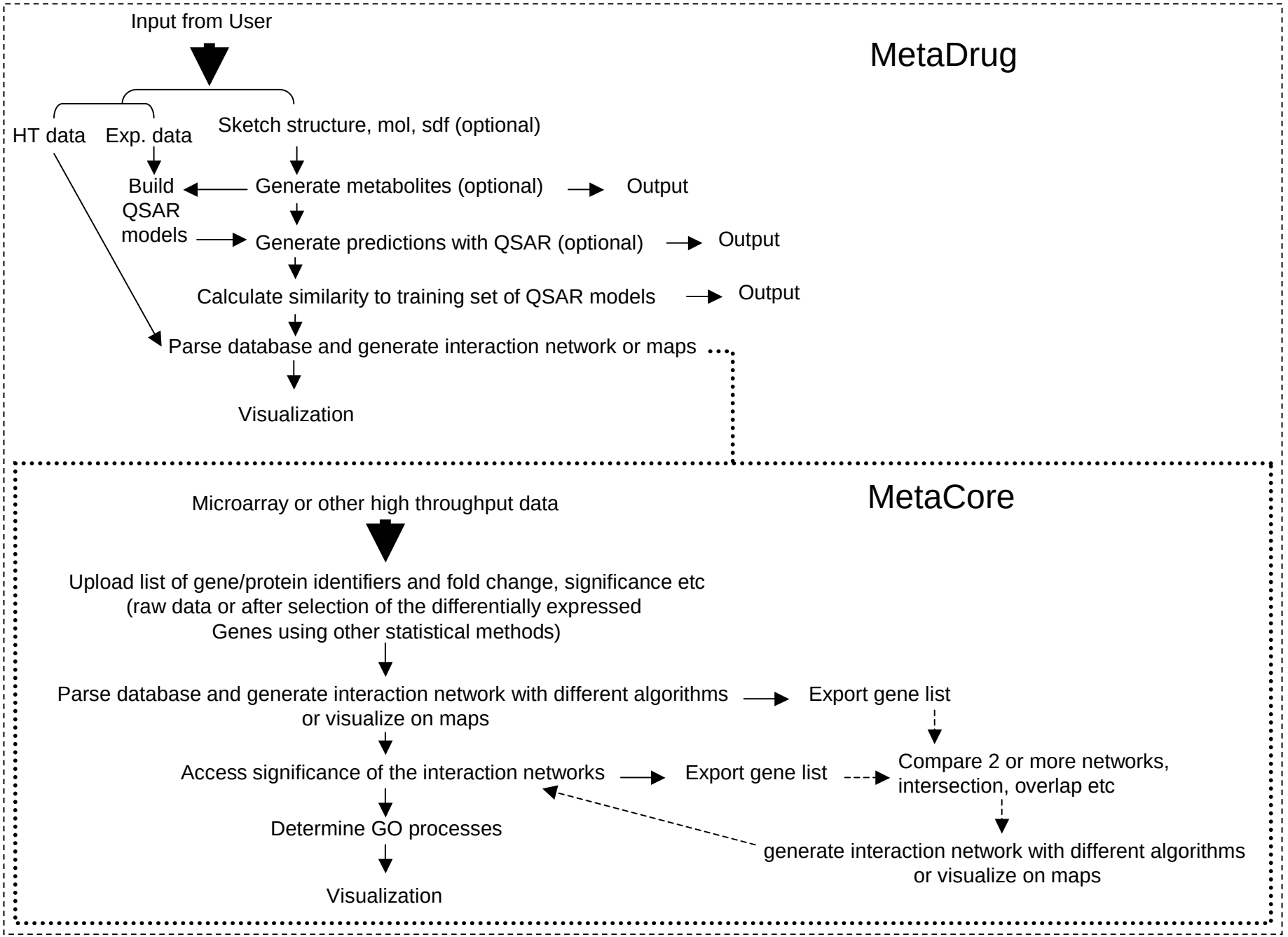
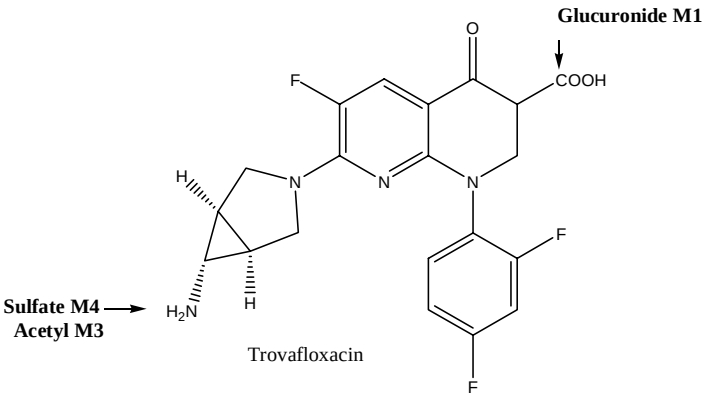
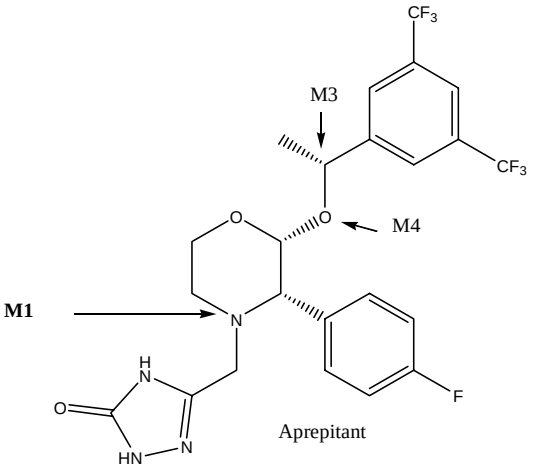
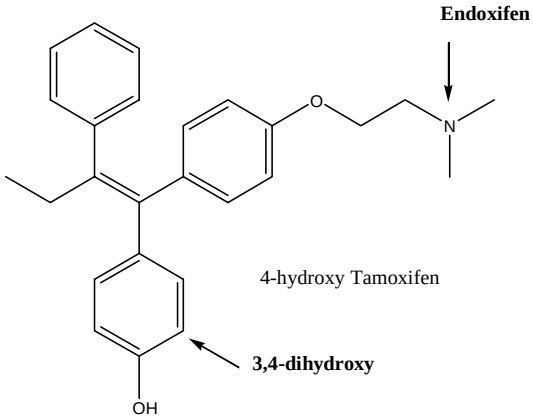


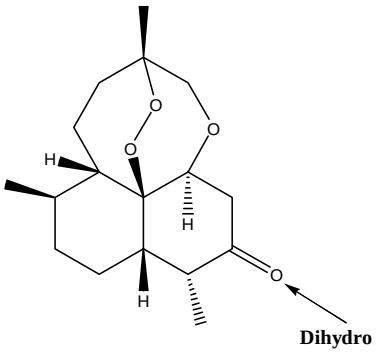
Fig 1

DMD Fast Forward. Published on December 28, 2005 as DOI: 10.1124/dmd.105.008458
This article has not been copyedited and formatted. The final version may differ from this version.









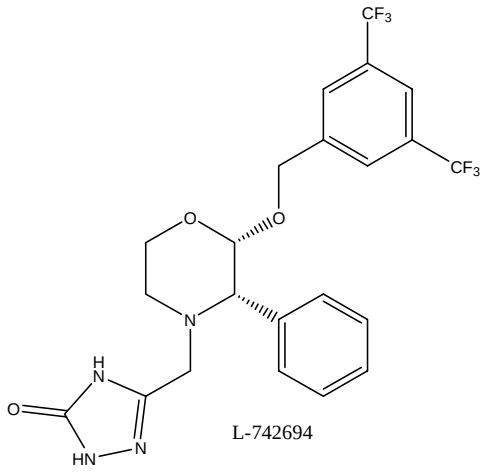


Fig 2

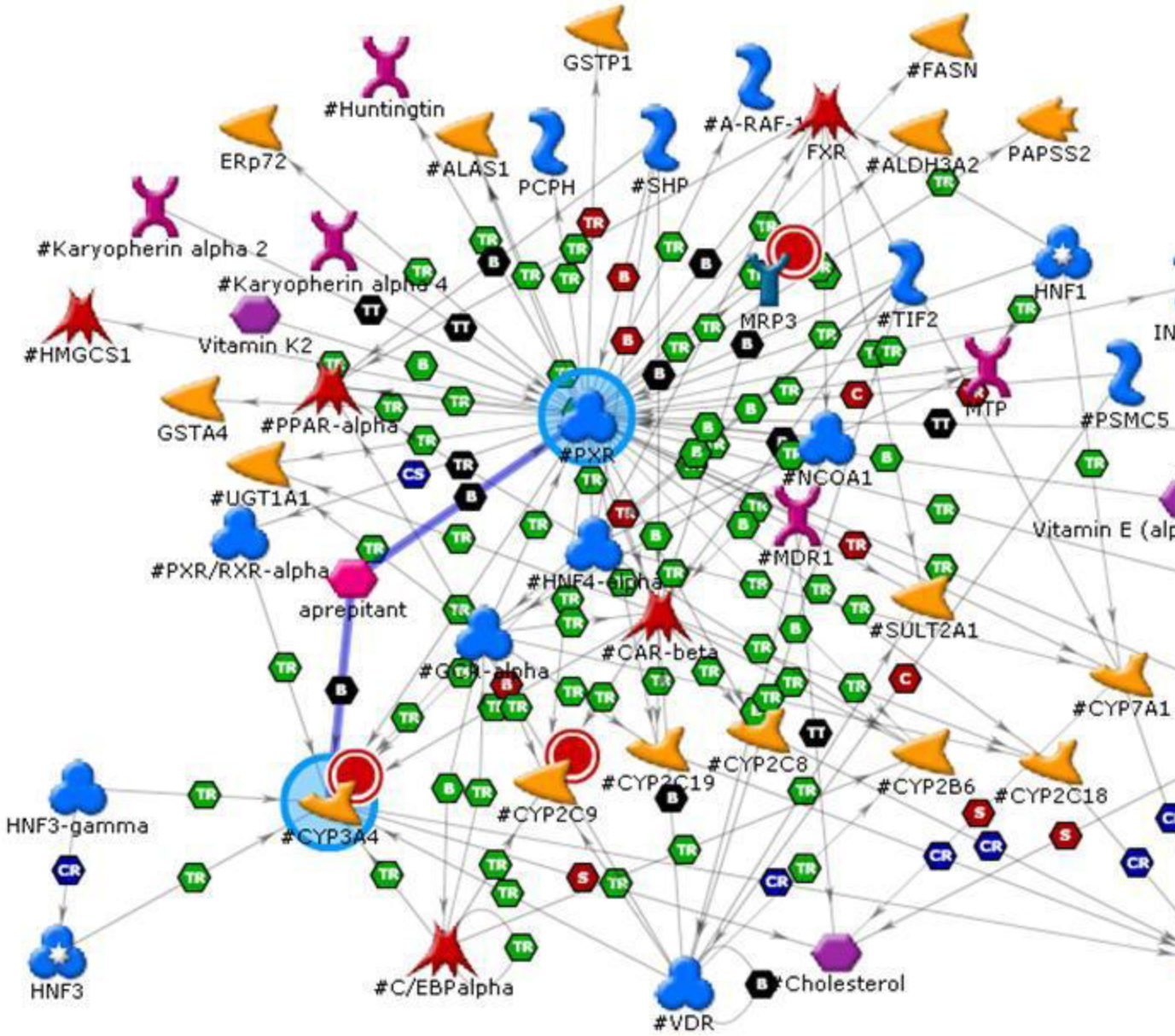


Fig 3

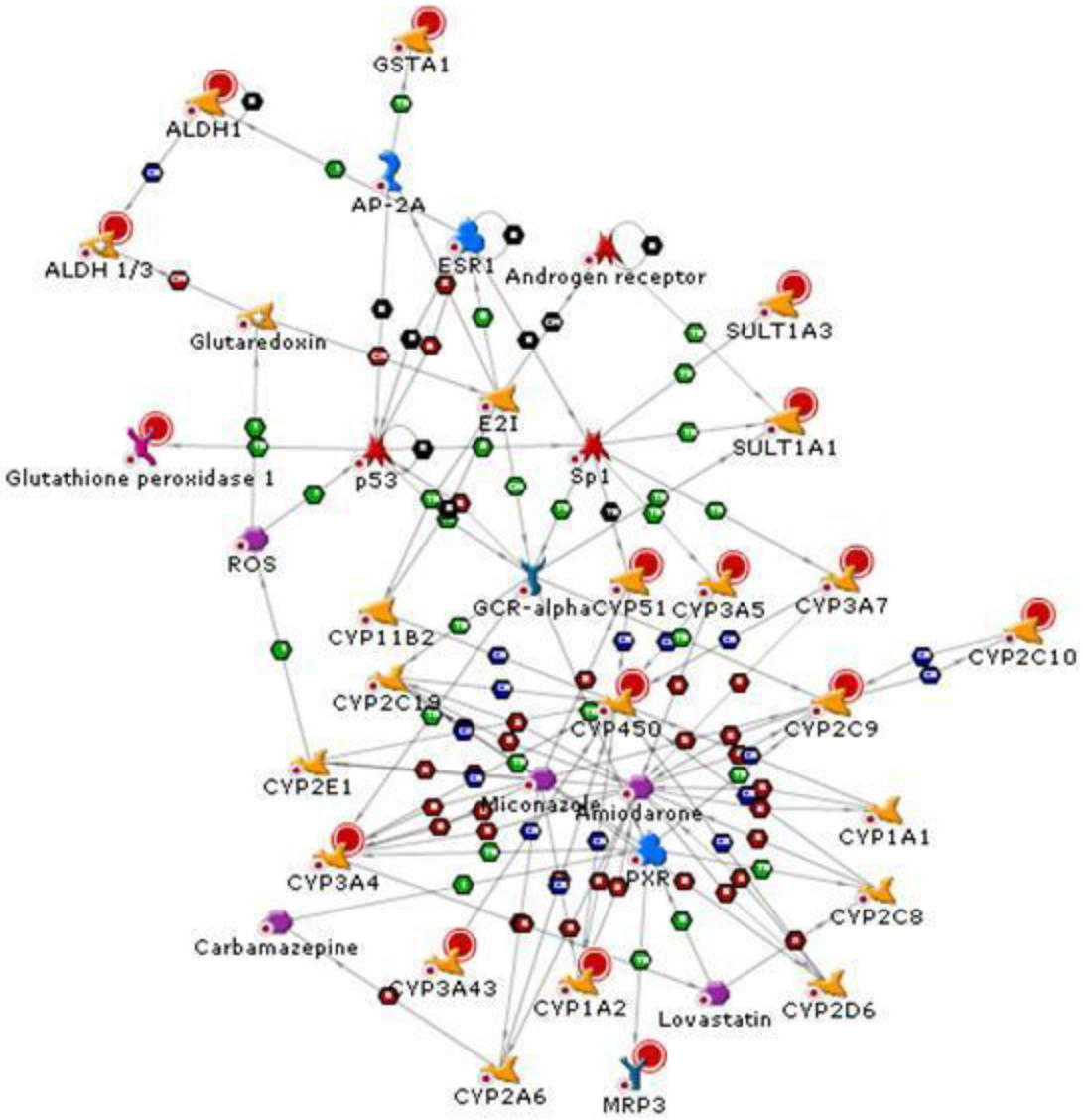


Fig 4a

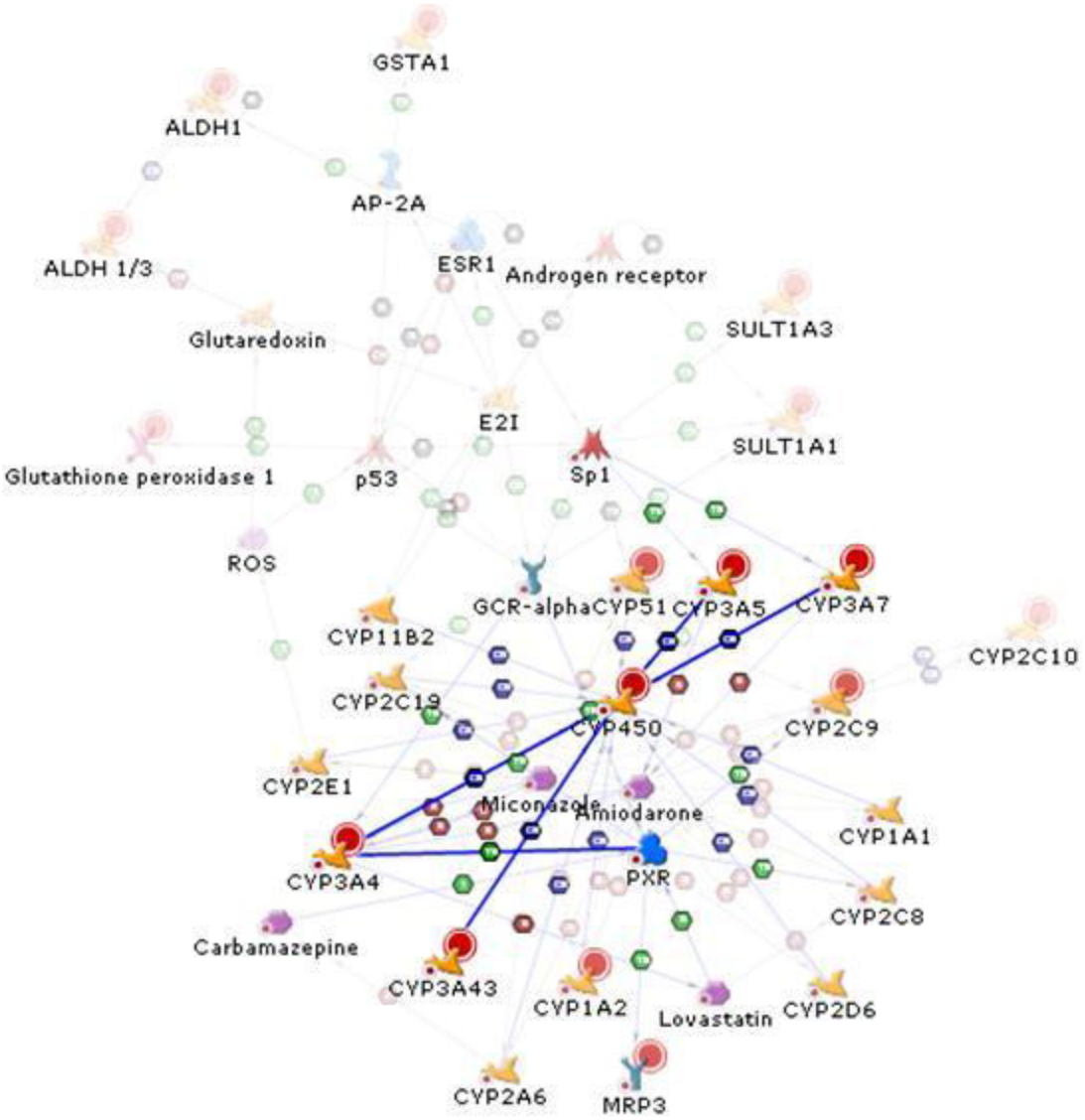


Fig 4b