Accelerated Communication

A NOVEL METHOD FOR VISUALIZING NUCLEAR HORMONE RECEPTOR NETWORKS RELEVANT TO DRUG METABOLISM

Sean Ekins, Eugene Kirillov, Eugene A. Rakhmatulin, and Tatiana Nikolskaya
GeneGo Inc., St. Joseph, Michigan

Received October 21, 2004; accepted December 16, 2004

ABSTRACT:
The increasing generation of biological data represents a challenge to understanding the complexity of systems, resulting in scientists increasingly focused on a relatively narrow area of study, thereby limiting insight that can be gained from a broader perspective. In the field of drug metabolism and toxicology we are witnessing the characterization of many proteins. Most of the key enzymes and transporters are recognized as transcriptionally regulated by the nuclear hormone receptors such as pregnane X receptor, constitutive androstane receptor, vitamin D receptor, glucocorticoid receptor, and others. There is apparent cross talk in regulation, since multiple receptors may modulate expression of a single enzyme or transporter, representing one of many areas of active research interest. We have used published data on nuclear hormone receptors, enzymes, ligands, and other biological information to manually annotate an Oracle database, forming the basis of a platform for querying (MetaDrug). Using algorithms, we have demonstrated how nuclear hormone receptors alone can form a network of direct interactions, and when expanded, this network increases in complexity to describe the interactions with target genes as well as small molecules known to bind a receptor, enzyme, or transporter. We have also described how the database can be used for visualizing high-throughput microarray data derived from a published study of MCF-7 cells treated with 4-hydroxytamoxifen, to highlight potential downstream effects of molecule treatment. The database represents a novel knowledge mining and analytical tool that, to be relevant, requires continual updating to evolve alongside other key storage systems and sources of biological knowledge.

The increasing generation of biological data using high-throughput methods in drug discovery necessitates the use of computational technologies including databases to store, analyze, interpret, and learn from this information (Navarro et al., 2003). Within drug disposition and toxicology, in vitro approaches for generating data with drug-metabolizing enzymes, transporters, ion channels and receptors can be used for predictive computer model generation (Ekins and Swaan, 2004). Many of these proteins are known to be regulated by nuclear hormone receptors (NHRs) or other transcription factors (Waxman, 1999; Moore and Kliewer, 2000; Goodwin et al., 2001; LeCluyse, 2001a,b; Staudinger et al., 2001a,b; Akinyama and Gonzalez, 2003; Mankowski and Ekins, 2003) affecting endogenous metabolism, cell growth, proliferation, and oxidative stress (Ulrich, 2003; Ulrich et al., 2004). The effect of these NHRs and other transcriptional factors on the toxic response and drug metabolism is complex and overlapping in a species-specific manner (Sonoda et al., 2003), with the same compounds working as agonists and antagonists on different receptors (Ulrich, 2003). Understanding the interactions of diverse ligands with these receptors (Sueyoshi et al., 1999; Spink et al., 2002; Mimura and Fuji-Kuriyama, 2003; Hartley et al., 2004; Tabb et al., 2004) and their impact on regulation of proteins has resulted in a simplistic schematic of the cross talk (Ekins et al., 2002). There have been considerable advances in the availability of software for visualizing complex gene networks. To date, several algorithms have been described in the literature for combining protein interaction information and expression data to find condition-specific modules in protein networks. These include different statistical methods to analyze data prior to mapping onto interaction networks (Tornow and Mewes, 2003), network clustering algorithms such as superparamagnetic clustering to identify tightly connected sets of nodes (objects connected to each other on a network, a component that can be a gene, small molecule, etc.) (Hanisch et al., 2002), simulated annealing (Ideker et al., 2002), probabilistic graphical models (Segal et al., 2003), and finding tightly connected clusters of nodes (cliques) using Monte Carlo optimization (Spirin and Mirny, 2003). These different approaches are useful when dealing with the idea of modular organization of large-scale networks of biological processes in which various types of cellular functionality are provided by relatively small, transient, but tightly connected networks of proteins (5–25 nodes) that are engaged in performing specific functions (Hartwell et al., 1999).

In the present study we describe the development of software with

ABBREVIATIONS: NHR, nuclear hormone receptor; ACC, acetyl-CoA; PPAR, peroxisome proliferator-activated receptor; AHR, aryl hydrocarbon receptor; FXR, farnesoid X receptor; LXR, liver X receptor; PXR, pregnane X receptor; RXR, retinoid X receptor; OHT, hydroxytamoxifen.
FIG. 1. A, MetaDrug database model. Borders enclose database entities related to core functional entities: red box represents entities related to component; blue box represents entities related to transformation; green box represents effects; gray box represents entities related to functional block. B, pathway representation and high-throughput data uploaded into the MetaDrug system. The first step of fatty acid biosynthesis, conversion of acetyl-CoA into malonyl-CoA, is shown on the right-hand side, and regulation of this metabolic reaction is on the left-hand side of the figure. Blue boxes 1 to 9 represent nine different levels of gene expression regulation. High-throughput data, microarray, proteomic, and metabolomic are represented as large rectangles, and arrows indicate corresponding objects of the database for which different types of the data could be uploaded and visualized. C, the legend for the different shapes presented in A and B.
a novel architecture (Fig. 1A) using an underlying manually curated database of key drug-metabolizing enzymes, their substrates, and nuclear hormone receptors. We have used this database to query for NHRs and build gene networks, as well as analyze high-throughput data such as those derived from microarrays. Since there has been considerable generation of data for NHRs in recent years (Tugwood and Montague, 2002; Ueda et al., 2002; Yamazaki et al., 2002; Rosenfeld et al., 2003; Johnson et al., 2004), we were keen to assess graphically the extent of the current knowledge and the complexity of the interactions. In addition, we provide an example of how the underlying database may be used to visualize microarray data relevant to drug disposition and toxicity. To our knowledge, this represents the first description of the development and applications of a novel interactive database for building gene networks relevant to drug disposition.

Materials and Methods

Data Annotation and Software Programming. An interactive, manually annotated database was derived from literature publications on proteins and small molecules of relevance to drug disposition and toxicology in humans (MetaDrug; GeneGo, St. Joseph, MI) and 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). Functional processes are the core objects in the database which can be of a different nature and have different relationships with molecular entities. We use three major types of functional processes: effects, transformations, and blocks. In addition, we introduce the notion of a component that describes molecular species or functional groups of molecules in their biological context (Fig. 1A), described as follows.

1. Component represents functional groups of molecules in biological systems. It is related to a molecular entity, localization, cell/tissue, and organism. Thus, the components represent biological molecules within their biological context. The molecular entity is treated in a broader sense than just being a specific chemical compound. In our representation, it could also be a group of molecules (e.g., a protein family or class of chemical compounds) or a molecular complex. This is particularly useful for representing the cellular processes, when the exact chemical composition or particular isoform of a protein participating in a pathway is unknown or ambiguous (e.g., Enzyme Commission numbers).

2. Transformation is an entity that is used to store information on biochemical reactions, transport, transcription, and translation or any biological process, with a primary function being to change the amount of a component (e.g., a reaction, in a broad sense) that is considered, in its particular environment, to be a subcellular compartment, tissue, and organism.

3. Effect is an entity that represents the influence that components exert on either transformations or other effects. Each effect has an agent (component), a target (transformation, another effect, or an entire block), a type, and a set of numerical values. The notion of effect is convenient for description of biological activity, whether or not its exact mechanism is known, since incomplete information can be stored, allowing reconstruction of cellular networks.
FIG. 2. A, a schematic description of the results of direct interaction networks and autoexpand networks. B, a direct interactions network generated with MetaDrug using all of the nuclear hormone receptors input. Each different-colored and -shaped node represents an individual gene connected by a vector. The type of interaction between nodes is encoded on the vector with a small hexagon of different colors. C, an autoexpand network generated with MetaDrug using all of the NHRs that act as transcriptional regulators of many genes involved in drug disposition and toxicity. Clusters of ligands (purple hexagons) linked to each NHR (large blue blobs), trans_factors (red), enzymes (yellow arrows), and transporters (blue) from the MetaDrug database are shown (the legend for the different shapes presented is described in Fig. 1C). Identifications of the location of clusters of ligands for the major NHRs are shown. An example of molecule connectivity is illustrated for guggelsterone (in rectangle) connected by red lines to FXR (top) and PXR (bottom).
4. **Block** is used to describe functional units, be it a particular category of metabolism, or any other functional process. Thus, blocks link together components, effects, and transformations that are functionally related. Blocks are hierarchical inasmuch as they may contain other blocks as elements. On the other hand, every element may be a part of more than one block. Blocks are linked to each other by shared elements. Assembling different entities within functional blocks enables rapid searches of functional links, and function-centered analysis of expression and other high-throughput molecular data (Fig. 1A).

To summarize, functional processes and components serve as the core information space-holders in our database, with many-to-many relationships between them. The corresponding molecular and mechanistic data are then linked to these space-holders as they become available. Functions serve as the “linking portals” for heterogeneous data. Once linked, the heterogeneous types of high-throughput data become a part of a larger system-level picture in which functional relations among them can be more easily established and elucidated (e.g., all proteins in a pathway and their genes with expression patterns). Every pathway and its elements (interactions, reactions, enzymatic functions) are linked to available molecular data (genes, proteins, compounds, expression data, single nucleotide proteins, etc.) and annotated with relevant information about their involvement and importance in a number of common human diseases. An illustration of how we can represent metabolic and signaling pathways in MetaDrug is shown in Fig. 1B. Here, we demonstrate the integration between these pathways and give an example of the space-holders that contain information on the gene expression regulation and protein activity. The first metabolic step of fatty acid biosynthesis, the conversion of acetyl-CoA into malonyl-CoA in mammals, is known to be differentially regulated in different tissues (Allred and Reilly, 1996; Abu-Elheiga et al., 1997). In addition, mammals have two major isoforms of acetyl-CoA carboxylase. ACCAlpha is the major isofrom in lipogenic tissues such as liver and adipose tissue, and ACCbeta is predominantly active in heart and skeletal muscle. The activity of both ACC isoforms is under long-term control at the transcriptional and translational levels and under short-term regulation by rapid covalent modification of the enzyme (inactivation and activation by phosphorylation and dephosphorylation, respectively) and by allosteric transformation by citrate or palmitoyl-CoA (feed-forward activation and feed-back inhibitors, respectively) (Allred and Reilly, 1996; Abu-Elheiga et al., 1997). These different levels of gene expression and protein activity regulation are represented by nine space-holders. In this case, space-holders serve as connectors between signaling pathways that regulate the activity of the muscle-specific acetyl-CoA carboxylase, which in turn becomes a part of the first metabolic reaction in fatty acid biosynthesis. Because we are not aware of any standard in the systems biology community for objects displayed on such metabolic and signaling pathways, we have created our own symbols to represent different types of objects such as enzymes, transcriptional regulators, etc. (Fig. 1C).

The MetaDrug software runs on an Intel-based 32-bit server running Red-Hat Linux Enterprise 3 AS (Red Hat, Raleigh, NC) and the web server ran Apache 1.3/mod_perl (http://perl.apache.org/start/index.html). Software on the server side was written in Perl, whereas the client side required HTML/JavaScript and the Macromedia Flash Player Plug-in (Macromedia Inc., San Francisco, CA).

**Generation of a Nuclear Hormone Receptor and Transcriptional Factor Gene Network.** The following transcriptional factors were queried in MetaDrug: PPAR, FXR/RXRA, estrogen receptor α, AHR, hepatocyte nuclear factor 4α, glucocorticoid receptor-β, MCR, constitutive androstane receptor-β, glucocorticoid receptor-α, LXR-α, constitutive androstane receptor/RXR heterodimer, hepatocyte nuclear factor 4, FXR, PXR/RXR heterodimer, PXR, AHR/aryl hydrocarbon receptor nuclear translocator heterodimer, PPARα/ LXRα, vitamin D receptor, and PPAR-α. These factors were initially visualized using a direct interactions algorithm (Fig. 2A) implemented in the software used to create subnetworks around every object from the uploaded set of nuclear receptors. The expansion of this network halts when the subnetworks intersect for each nuclear hormone receptor (Fig. 2A). The objects that do not contribute to connecting subnetworks are then automatically truncated (Fig. 2B). A second algorithm for autoexpansion of the interaction network around the transcriptional factors was also used, which finds the clusters of objects directly connected to each other (Fig. 2, A and C). Each connection represents a direct, experimentally confirmed, physical interaction between the objects.

**Visualization of Microarray Data on Gene Networks.** We have used data from a published study of G_{Q16}-arrested MCF-7 breast cancer cells treated with 4-hydroxytamoxifen (OHT) using the National Institute of Environmental Health Sciences ToxChip microarray, consisting of 1901genes (Hodges et al., 2003). The microarray data describing the up-/down-regulation of genes (Hodges et al., 2003) was imported into MetaDrug on the client side as a tab-delimited file. This file was obtained from the National Institute of Environmental Health Sciences website, http://dir.niehs.nih.gov/microarray/data-sets/home-pub.htm, and a network was generated around genes (cytochromes P450) of interest, relating to the metabolism of OHT. Genes from the microarray that were up- or down-regulated could then be visualized on this network (Fig. 3).
Results and Discussion

Computationally aided interpretation of biologically complex (genetic, proteomic, transcriptomic, and metabonomic) datasets generated with high-throughput technologies and the accumulated knowledge of human protein interactions represents the new paradigm of systems biology (Hood, 2003; Nicholson and Wilson, 2003). Datasets for molecules studied by researchers in drug disposition and toxicology are accumulating rapidly, and the resultant information can also contribute to the ongoing systems biology paradigm. For example, the Chemical Effects in Biological Systems initiative will ultimately provide microarray data for relevant compounds in a publicly accessible format (Waters et al., 2003; Mattes et al., 2004). Capturing and querying the various types of biological data requires an accessible database. The design of the MetaDrug database is unique compared with currently available genomic and proteomic databases based on genes or proteins as elementary objects (Bairoch and Apweiler, 2000; Benson et al., 2002), which typically do not provide functional links between these objects, making it difficult to query functionally related objects (e.g., all proteins in the pathway, all pathways involved in a disease, etc.). We have designed and implemented a novel database architecture (described earlier and in Fig. 1, A–C) allowing the organization and visualization of relevant biological and chemical information around the molecular entities, genes, proteins, transcripts, and compounds by connecting them through functional processes: reactions, pathways, and networks.

Our approach to understanding the effects of a molecule on the whole organism (and, in particular, drug disposition and toxicology) are derived from the work in the area of metabolic reconstruction (Selkov et al., 1997), in which the metabolic blocks (pathways and subsystems) corresponding to the genetic component of an organism are compiled and connected via intermediates into wire diagrams or metabolic reconstruction models. Our database and software therefore enable the user to 1) query the underlying database, 2) generate gene networks using algorithms based on datasets and the manually curated database of human protein–protein and protein–compound interactions, and 3) upload and simultaneously visualize on the gene networks microarray or other high-throughput data. Additionally, the user can filter the networks by removing an object, by type of interaction, or by tissue type, and the microarray data can be filtered based on the desired fold change threshold, etc.

We have applied this database platform to the NHRs, transcriptional factors, and their associated interactions with other proteins and small molecules relevant to drug disposition and toxicology. A direct interactions algorithm (Fig. 2B) was used to connect NHRs to each other and results in two clusters, a small AHR-centered network and a larger RXR-centered network with the remaining NHRs (Fig. 2B). These networks show the directionality of an interaction via the vector between two nodes. The type of interaction is also encoded on this vector (e.g., binding, transcriptional regulation, covalent modification, etc.). The heterodimerization of RXR with many of the transcriptional regulators appears to be a key node for connectivity. In contrast to this, a very complex network is generated with the NHRs using the autoexpand algorithm (Fig. 2C), which now encompasses proteins regulated by these NHRs, as well as connected small molecule ligands. This qualitative visualization represents the current state of knowledge around NHRs in terms of a network focused on the proteins of importance to drug disposition. For example, the ligand guggulsterone is shown highlighted and connected to FXR (as an antagonist) and LXR (as an activator) on the network (Fig. 2C), which corresponds to its known behavior (Owsley and Chiang, 2003).
would be far too lengthy to describe in detail the underlying connectivity of the transcriptional regulators in Fig. 2C compared with our previously published simplistic schematic of ligand-receptor interactions (Ekins et al., 2002). Such networks as Fig. 2C can, however, be filtered to remove objects as desired or queried simply by browsing over the object. These types of actions result in the retrieval of other annotated data for a particular node such as synonyms, biological functions, and connections to other public databases including PubMed, the human single nucleotide protein database, and others. By focusing on any one object, the interaction network can be expanded to highlight other nodes that may not be apparent on the current network. The collection of ligands presently stored in MetaDrug illustrates a molecule repository to be used for computational model building to predict binding to nuclear hormone receptors (Ekins and Erickson, 2002; Ekins et al., 2002; Xiao et al., 2002; Mankowski and Ekins, 2003; Tabb et al., 2004).

Microarray data can also be transposed on metabolic and signaling networks generated with this database. We have used data from published experiments in which MCF-7 breast cancer cells had been treated with OHT for 24 h (Hodges et al., 2003) to demonstrate this. OHT is known to be further metabolized via CYP3A4 (Destá et al., 2004) as well as by phenol and estrogen sulfotransferases in human liver tissues (Chen et al., 2002), and this molecule is also an inducer of PXR (Desai et al., 2002). The induction of PXR results in increased levels of CYP3A4, which is partly responsible for OHT formation itself (Crewe et al., 1997). A network was generated around the enzymes of interest relating to the metabolism of OHT, and the microarray data were visualized on this network. In the MCF-7 breast cancer cells, PXR and CYP3A4 were shown on a network to be slightly up-regulated following treatment with the low nanomolar doses of OHT to which the MCF-7 cells were exposed (Fig. 3). In this example, genes that were down-regulated are also shown as blue circles and genes up-regulated as red circles (Fig. 3). The illustrative map has directional information encoded in the vector between two objects, and for clarity, we have removed the additional information on the type of interaction between objects. It is possible for the user to highlight pathways of interest and to build more complex networks than shown here, which obviously only contain a fraction of our database content relevant to drug-metabolizing enzymes and transporters known to be involved in tamoxifen and 4-hydroxytamoxifen metabolism and transport. Ideally, an experiment using multiple doses of OHT to which the MCF-7 cells were exposed (Fig. 3). In this way, the database and software could be expanded to highlight other nodes that may not have been evaluated.

In summary, we have described how published data relating to transcriptional regulators of key proteins involved in drug disposition and toxicology in humans can be used to generate a database that can then be analyzed by a novel network visualization tool. A similar approach could be taken in future with data published for other species to compare the complexity and differences in the regulatory networks and aid in drug disposition and toxicity assessment.

Acknowledgments. Drs. Sergey Andreyev, Svetlana Sorokina, Tatjana Serebrijskaya, Andrej Bugrim, Roman Zuev, Andrej Ryabov, Yuri Nikolsky, and colleagues at GeneGo are acknowledged for contribution to this work. We gratefully acknowledge Michigan Life Science Institute for hosting our server.

References


Address correspondence to: Dr. Sean Ekins, Vice President, Computational Biology, GeneGo, Inc, 500 Renaissance Drive, Suite 106, St. Joseph, MI 49085. E-mail: sean@genego.com or ekinssean@yahoo.com