research articles Journal of proteome research

Integrative Top-Down System Metabolic Modeling in Experimental Disease States via Data-Driven Bayesian Methods

Jung-Wook Bang,[†] Derek J. Crockford,[†] Elaine Holmes,[†] Florencio Pazos,[‡] Michael J. E. Sternberg,[‡] Stephen H. Muggleton,[§] and Jeremy K. Nicholson*,[†]

Department of Biomolecular Medicine, Division of Surgery, Oncology, Reproductive Biology & Anaesthetics, Sir Alexander Fleming Building, Imperial College, London SW7 2AZ, U.K., Structural Bioinformatics Group, Division of Molecular Bioscience, Imperial College, London SW7 2AY, U.K., and Computational Bioinformatics Group, Department of Computing, Imperial College, London SW7 2AZ, U.K.

Received June 7, 2007

Multivariate metabolic profiles from biofluids such as urine and plasma are highly indicative of the biological fitness of complex organisms and can be captured analytically in order to derive top-down systems biology models. The application of currently available modeling approaches to human and animal metabolic pathway modeling is problematic because of multicompartmental cellular and tissue exchange of metabolites operating on many time scales. Hence, novel approaches are needed to analyze metabolic data obtained using minimally invasive sampling methods in order to reconstruct the pathophysiological modulations of metabolic interactions that are representative of whole system dynamics. Here, we show that spectroscopically derived metabolic data in experimental liver injury studies (induced by hydrazine and α-napthylisothiocyanate treatment) can be used to derive insightful probabilistic graphical models of metabolite dependencies, which we refer to as metabolic interactome maps. Using these, system level mechanistic information on homeostasis can be inferred, and the degree of reversibility of induced lesions can be related to variations in the metabolic network patterns. This approach has wider application in assessment of system level dysfunction in animal or human studies from noninvasive measurements.

Keywords: Metabolic and regulatory networks • computational methods • molecular biology of disease

Introduction

The metabolic signature of intact biological systems, as expressed in their secreted fluids, is influenced strongly by genetic, environmental, and disease factors.[1–3] Most current systems biology approaches involve a bottom-up paradigm in which the behavior of a complex system is dissected from understanding interactions at the cellular level, e.g., in unicellular organisms.[3–5] Such approaches are not applicable to problems relating to screening for human disease or molecular epidemiological studies because it is only possible to sample real populations from the top, that is, with readily available biomaterials such as plasma and urine on a large scale. These fluids give a useful end-point metabolic signature that is representative of integrated system function or dysfunction related to disease; however, at present there are no statistical approaches capable of delivering network interaction information for a complex multicompartmental system. Probabilistic graphical models (Bayesian models) have been proposed as an efficient means of visualizing interactions at several levels of biomolecular organization influencing metabolic regulation and control.[6,7] Metabolic networks provide a potential framework for modeling dynamic biological events at an integrated or global systems biology level, where many different cell types interact to influence fluxes in extracellular metabolic pools. The metabolic information available in databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes) allows a starting point for the study of metabolism from a network perspective.[8] The modular architecture of such networks (grouping into cohesive subnetworks) and the relationship between these modules and classical pathways[9] have been reported. Substantial quantitative modeling work has been carried out on metabolic flux analysis,[10] but this is not readily applicable to data sets determined from biofluids, which represent integrated compartment interactions.

Most metabolic pathway analysis theories have been developed from *in vitro* studies on isolated cells where biological activities can be assumed, and fluxes through pathways can be followed using isotopic labeling. Various mathematical techniques have been applied to model such cellular biological systems including biochemical systems theory (BST), metabolic control analysis (MCA), and flux balance analysis (FBA).[11–14] In previous studies, gene expression data have been modeled using Bayesian techniques assuming complete knowledge of the structure of the network and measurement of cellular fluxes.[15,16] However, in order to understand system level

* To whom correspondence should be addressed. Tel: +44 (0)20 7594 3195. Fax: +44 (0) 20 7594 3226. E-mail: j.nicholson@imperial.ac.uk.
[†] Division of Surgery, Oncology, Reproductive Biology & Anaesthetics.
[‡] Division of Molecular Bioscience.
[§] Department of Computing.

10.1021/pr070350l CCC: $40.75

Published on Web 01/08/2008

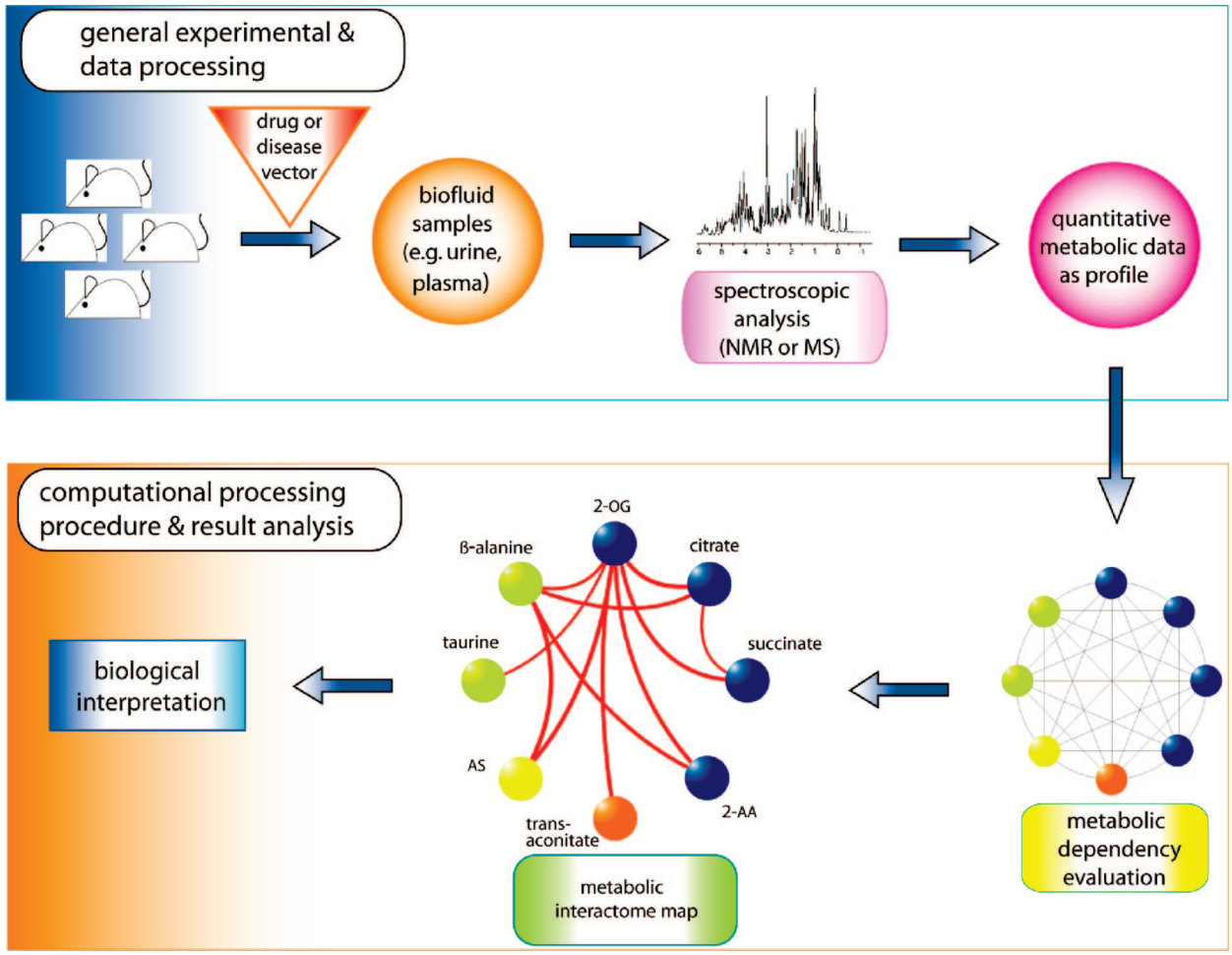

Figure 1. Procedure for generating a metabolic interactome map. For each toxin, a group of 30 Sprague–Dawley rats was divided equally into control, low-, and high-dose cohorts. Urine samples were collected and analyzed by ^{1}H NMR spectroscopy to determine relative levels of target metabolites. We constructed an interactome map using the dependency information calculated from these levels.

metabolic regulation in humans (or indeed animals used in toxicological, pharmacological, or functional genomic studies), a radically different measurement and modeling paradigm is required.[3] Therefore, in marked contrast to earlier attempts, we have provided a graphical representation of the perturbed metabolic network (an interactome map) directly from empirically derived metabolic data, reflecting integrated multicellular biochemical interactions at the whole system level.[2]

^{1}H NMR spectroscopy generates complex and highly reproducible metabolic signatures of biofluids that have been widely demonstrated to carry detailed information on a variety of pathophysiological conditions.[1,2,17] It has been shown that diverse metabolic signatures of homeostatic and pathological processes are carried in urine,[1,18] which is well suited for human studies as collection is noninvasive. A probabilistic graphical model permits the use of nonspatially registered metabolic information to describe the integrated system activities of many tissue compartments. Thus, we developed a novel interactome map approach that utilizes both probabilistic reasoning (Bayes' theorem) and graphical modeling (visualization) for describing the metabolic relationships.[19] These approaches are suitable for dealing with biological data since they provide a framework for incorporating prior knowledge, dealing with uncertainty, and combining competing models for explaining biological phenomena.[20,21] We use two model hepatotoxins with markedly different mechanisms of action, viz., hydrazine, a toxic metabolite of the antitubercular drugs isoniazid and iproniazid that produces steatotic liver pathology,[22–24] and α-napthylisothiocyanate (ANIT), which causes bile duct necrosis, cholangitis, and bile duct proliferation, resulting in secondary hepatic cholestasis.[25] The toxins were given at two dose levels to simulate hepatic disease states.

Metabolic Interactome Mapping. A metabolic interactome map is a novel graphical means of representing the homeostatic and pathological interactions of metabolites derived from multiple cellular sources at the organism level. This type of mapping is not constrained by or dependent on prior knowledge of pathway interactions at the cellular level or the conventional pathway formalism, but may nonetheless result in structural similarities owing to organism-wide connections in pathway control and regulation. The procedure for generating an interactome map is shown in Figure 1. Quantitative metabolic data are used to calculate dependencies between metabolites in basal and challenged biological systems. The interactome maps are cross-validated[26] to provide a measure of their robustness. Here, we use the word dependency to refer to a specific measure of the strength of the statistical association between variables (metabolite concentrations), but this has no *a priori* interpretation in terms of biological processes.

Experimental Procedures

Samples and Spectroscopy. Rat urine samples were analyzed by ^{1}H NMR spectroscopy as part of the COMET project,[27,28] and the resulting data were used here. Each COMET toxin study was performed by a specific collaborating center according to pre-agreed protocols, which included comparable experimental conditions to be applied (e.g., housing and diet). In each one, male Sprague–Dawley rats were randomly allocated to three dose groups (10 per group), giving 30 rats per study. The dose groups were control, low dose, and high dose. In one of the studies addressed here, hydrazine hydrochloride in 0.9% saline was administered orally at 30 mg/kg (low dose) or 90 mg/kg (high dose), and in the other, ANIT in corn oil was administered orally at dose levels of 12.5 mg/kg (low dose) or 125 mg/kg (high dose). Dose vehicles alone were administered to the control animals. Low dose levels were designed to invoke a threshold response in the animals, while high dose levels were designed to invoke a pathological response. All animal experiments were conducted according to specified U.K. national guidelines.

The time of dosing was defined as zero hours. Rats were housed in individual metabolic cages under controlled temperature, humidity, and light cycles. Urine samples were collected for time intervals spanning –24 to –16, –16 to 0, 0–8, 8–24, 24–48, 48–72, 72–96, 96–120, 120–144, and 144–168 h, with each sample labeled by the end time of the sampling interval. Half of each dose group was euthanized at 48 h and the rest at 168 h, giving 225 samples per toxin study (ignoring small discrepancies due to missing samples). The start of an experiment was timed so that the –24 to –16 and 0 to 8 h intervals occurred during daylight.

One dimensional ^{1}H NMR spectra of phosphate-buffered samples were measured at 600 MHz in a conventional manner using a standard 1D presaturation pulse sequence for water suppression.[29] The spectra were recorded into 32K data points and referenced to 3-(trimethylsilyl) propionic acid-d_4 (as sodium salt). Further details of experimental protocols and spectral acquisition parameters have been reported.[28,30]

Histopathology Assessment of the Liver. This was conducted by a team of COMET pathologists using preagreed criteria and vocabulary. The results obtained are consistent with published material.[22–25] The severity scale runs from 0 to 4. At low dose, there was a single hydrazine-treated animal that displayed a severity 1 focus of midzonal necrosis at 48 h postdose and two ANIT-treated animals that developed a severity 1 focal, single-cell necrosis at 48 h. The lesions at high dose, which was a dose level intended to induce overt tissue damage, are summarized below.

Hydrazine. Hepatocellular cytoplasmic vacuolation was found in varying severity from 2 to 4 in all animals euthanized at 48 h and from 1 to 3 in 3 of the 5 remaining animals at 168 h. The high dose of hydrazine induced midzonal necrosis at severity 1 to 2 at 48 h postdose in all animals, but no necrosis was observed at 168 h.

ANIT. At 48 h postdose, 4 of the 5 animals had bile duct necrosis at severity 2 to 3. The same animals also had severity 1 oval cell hyperplasia, a regenerative response, and periportal single-cell hepatocyte necrosis at severity 1 to 2. Periportal inflammation at severity 2 was observed in all animals euthanized at 48 h and in 4 of the 5 animals at 168 h. All rats continuing to the end of the 168 h study period demonstrated periportal fibrosis at severity 1 to 2, which was multifocal in 4 animals and focal in 1.

Data Analysis and Metabolic Interactome Map Construction. The detailed process of generating a metabolic interactome map consists of four steps.

Step 1: Metabolite Analysis. The frequency domain data obtained were integrated to find the intensities of the NMR signals due to 19 endogenous metabolites of interest. The exact positions of these signals in the spectra were identified and the signals integrated, using an in-house computer program employing mathematical transformations from the reference spectra of pure compounds.[31] The intensities were normalized by area to account for any variation in urine concentration. The 19 metabolites were chosen because their signals could be measured reliably using this method and because they are all significant in their own right.

There were 10 samples (time points) available for each animal euthanized at 168 h and 5 for each animal euthanized at 48 h. For each sample (characterized by subject, 1–30, and time point), the relative intensity for each metabolite was calculated as the ratio to the zero hour intensity (in that subject). Hence, the effects of irrelevant, interanimal response variation were minimized, and it was reasonable to treat every animal in a dose group equally. This step yielded a fold change matrix, with rows labeled according to sample (i.e., subject and time point) and columns according to metabolite. A logarithmic fold change matrix was obtained from this by taking the logarithm of each element. This is a standard technique when modeling variables that may have different dynamic ranges so that smaller ones may assume some importance in the model and proportional increase and decrease are treated equally. For each metabolite, the range of the logarithmic fold change was then divided into five equal intervals, which were labeled on an increasing scale from 1 to 5, and finally, the label of the interval enclosing each value was assigned as the corresponding quantized fold change value. This was done because the input to the interactome map consists of a matrix of discrete values.

Step 2: Model Construction. Interactome maps were constructed using a dependency measure derived from the Kullback−Leibler information divergence[32,33]

$$D(X, Y) = \sum_{X,Y} \left[P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \right] \geq 0 \qquad (1)$$

where x and y are the members of data sets X and Y, and P is the probability of the given members occurring (jointly or independently, according to the number of arguments). This is a specific form of dependency measure derived from information theory. Given that any form of functional relationship exists between X and Y, this form of dependency measure will be sensitive to it. In other words, it provides a measure of the statistical association between two variables, whether the association is linear or nonlinear, in contrast to the more conventional Pearson correlation coefficient, which is only sensitive to linear associations. It is important to realize that no prior biological knowledge is required in this calculation and that a strong dependency value does not necessarily imply a relationship between any associated biological processes: this is a matter of interpretation by biological domain experts upon consideration of the complete, validated interactome map.

Step 3: Metabolic Interactome Mapping. After obtaining the dependency measure for every metabolite pair in the data, the network was constructed by connecting pairs that have strong dependencies exceeding a predefined threshold value. For each

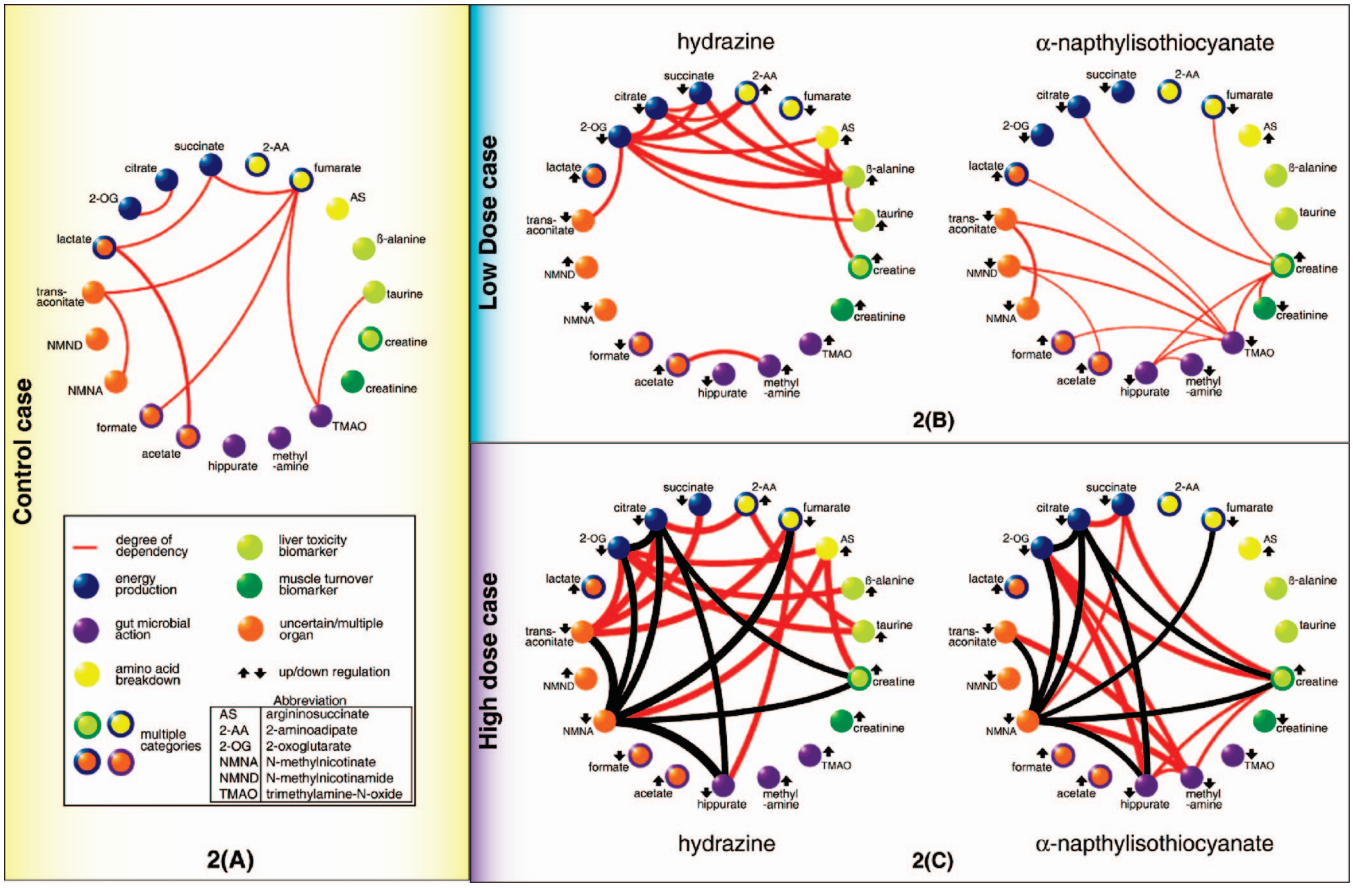

Figure 2. (**A**) Metabolic interactome map for the control case, based on spectra for 660 control animals, aggregated from several comparable COMET studies. Each node (O) and arc (−) represents a metabolite and the degree of dependency between the connected pair of metabolites, respectively. The width of a colored arc represents the degree of strength in dependency. The figure provides a reference reflecting normal metabolism, and highlights a number of expected relationships, e.g., tricarboxylic acid (TCA) cycle, and lactate and acetate. (**B** and **C**) Metabolic interactome maps for hydrazine and α-napthylisothiocyanate (ANIT) (2B: low dose, 2C: high dose). A black arc (−) indicates commonality of dependencies between the two toxins in the high-dose case (indicating a common liver damage end point). The meanings of the color-coded nodes are given in A. Arrow direction (↑ or ↓) indicates the increase or decrease of metabolite relative to the control.

toxin and dose regimen, this threshold was chosen so as to obtain a number of dependency relations of the same order of magnitude as the number of metabolites measured and hence to avoid overfitting the model to the data.

Step 4: Robustness Validation. Here, we employed 10-fold cross-validation,[26] which is a widely used test of model robustness. First, a data set is divided at random into 10 subsets. The first nine are used to build a model using the technique under test, which is used to predict the values in the tenth. The process can then be repeated nine times: each of the first to ninth subsets can take a turn as the one to be predicted, using a model built with the others. Therefore, 10 scores can be obtained, and the mean score is defined to be the result of the robustness test.

To make a fair test, the values corresponding to each toxin and to low or high dose samples were tested separately. Therefore, our low dose hydrazine data, consisting of 19 metabolite values (quantized fold changes ranging from 1 to 5) for each of 75 samples, were divided by sample at random into subgroups of approximately seven. An interactome map was built using the values for nine of these subsets and used to predict the values in the tenth subset, and the whole process repeated nine times as described above. The mean robustness score obtained for all metabolites was 67%; therefore, we can say that if predicting the value (1 to 5) of a metabolite using an interactome map, we would expect to get the right answer about 7 times out of 10 (i.e., 3.5 times better than guessing). The corresponding scores for low dose ANIT, and high dose hydrazine and ANIT, were 70%, 49%, and 52%, respectively. In a spirit of nonbias, data for all time points have been used in these calculations, including the predose samples; therefore, the scores are *conservative* compared to what might be obtained for a larger study in which comparable dose regimes and time points could both be tested separately.

Results and Discussion

Experimentally Induced Liver Injury. Prior to considering the system perturbed by toxins, interactome maps were calculated on the basis of spectra for 660 control animals, aggregated from several comparable COMET studies.[27] These provided a reference dependency network (Figure 2A) reflecting normal metabolism for Sprague–Dawley rats and highlighting a number of expected relationships such as those involving the tricarboxylic acid (TCA) cycle.

The networks for low and high dose hydrazine and α-napthylisothiocyanate treated animals are shown in Figure 2B and C, and represent the product of diverse metabolic interactions

Table 1. Selected Biochemical Interpretations of the Interactome Maps Obtained for Hydrazine and α-Naphthylisothiocyanate (ANIT) Perturbations

observed dependency	toxin, dosage, and suggested rationale for dependency
	hydrazine low dose case
2-aminoadipate to 2-oxoglutarate	disruption of lysine catabolism in liver and kidney, which relates to hippocampal neurotoxicity[24]
2-oxoglutarate to citrate and succinate	reduced 2-oxoglutarate availability due to transaminase inhibition[24]
β-alanine to taurine	• competition for the same renal transport system.[36] • Taurine is a primary biomarker of liver dysfunction, and β-alanine is a surrogate biomarker for liver damage; therefore, they should vary in concert in response to liver injury.[37]
methylamine to acetate	These both relate to the gut microbiotal effect on metabolism.[42] Methylamines derive from microbiotal conversion of choline and are associated with liver toxicity. Acetate is produced by colonic fermentation of dietary fiber
	hydrazine high dose case
hippurate to citrate and fumarate	The benzoylation of glycine to form hippurate is colocated with TCA cycle activity in the mitochondria; therefore, it is reasonable that hippurate has dependencies with TCA intermediates[24,38]
	ANIT low dose cases
creatine to citrate, fumarate, hippurate, creatinine and trimethylamine-*N*-oxide	Multiple cross-tissue interactions, e.g., citrate/fumarate/hippurate, is due mitochondrial colocation, and creatine to creatinine to muscle metabolism
trimethylamine-*N*-oxide to lactate, hippurate, formate, trans-aconitate, *N*-methylnicotinamide and creatine	Trimethylamime-*N*-oxide is a major amine excretory product in man, associated with gut microbiota;[43] the other metabolites have been associated with diet or energy metabolism[34] and therefore may link to trimethylamine-*N*-oxide through the gut microbiotal effect
	hydrazine and ANIT high dose cases
N-methylnicotinate hub; hippurate hub	These have many links in common for both toxins and therefore must say something about the high dose case in particular, i.e., are indicative of liver damage and/or severe disruption of the pathways between gut microbiota and host

in numerous tissues (Table 1). Results pertaining to the toxin-disrupted systems have been expressed as deviations from the basal network. Many dependencies center on major metabolites that form hubs in the network, for example, 2-oxoglutarate (hydrazine case, Figure 2B). The characteristic increase in the number of significant dependencies after toxin treatment is marked as is the variation with dose and partial convergence at high dose for both toxins (Figure 2C). It should be noted that the administration of a toxin may cause temporary changes in appetite in experimental animals,[34] but in our studies there was no evidence of fasting ketosis for either toxin even at high dose.

Pronounced differences in the metabolic responses at both doses of each compound are observed, as would be expected from their mechanistic toxicological differences. However, at high doses the interactome maps for each compound-disrupted system share a core pattern of perturbed metabolite dependencies (Figure 2C). This convergence of network interaction is consistent with the observation that during liver damage the mechanism-specific metabolic features of the network are partially replaced by features typical of a general metabolic failure pattern indicative of overt end-point cellular damage common to both toxins, that is, there are overriding convergent metabolic features that are a consequence of liver injury. At low doses, where the effects are readily reversible, the networks are highly toxin-specific (Figure 2B), that is, relate to pharmacological and mechanistic changes, and share no common pathway disruptions. Generally, the strengths of the metabolite dependencies vary considerably according to compound type and dose. Also, high dose hydrazine yields relationships with dependency values approaching unity, whereas ANIT at low dose gives weaker values (~0.25), and this is consistent with histopathology data suggesting that the level of observed damage is lower with ANIT. For both toxins, the specific metabolic networks observed at low dose collapse at high dose, and new dependencies are formed (Figure 2B and C).

***Short-Range* Dependency Relationships.** Many metabolite dependencies can be described as short-range, indicating linkage by a relatively small number of enzymatic steps in a specific metabolic pathway (these are particularly evident at low dose). These are amenable to biochemical interpretation using conventional knowledge of metabolism and generally occur within a single cellular system. For example, the robust dependency (hydrazine case, Figure 2B) between 2-aminoadipate (2-AA) and 2-oxoglutarate (2-OG) can be explained by disruption of the lysine catabolic pathway, where both metabolites are substrates of the enzyme 2-AA aminotransferase.[24] This is closely associated with the secondary neurotoxic effects of hydrazine due to pathway disruption in the hippocampus.[35] Likewise, the linkage of creatine to creatinine via the creatine phosphokinase catalyzed reaction to creatine phosphate is consistent with the strong dependency observed in the ANIT network (ANIT case, Figure 2B). The TCA cycle dependencies are also short-range in terms of the number of enzymatic steps between compounds and in the sense that the reactions are colocated within the mitochondria (hydrazine case, Figure 2B).

***Long-Range* Dependency Relationships.** We also observed what can be considered long-range dependencies between pairs of metabolites, that is, those that can only be connected via many intermediate enzymatic steps in a specific metabolic pathway, and hence are harder to explain in pathway terms. For example, β-alanine acts as a hub metabolite from which long-range dependencies to several TCA cycle intermediates derive (hydrazine case, Figure 2B); these can be connected via glutamate decarboxylase and L-aspartate in KEGG. Similarly, dependencies spanning multiple biochemical cycles can also be demonstrated; the dependency of argininosuccinate to 2-OG suggests a long-range interaction between the TCA and urea cycles.

Wormhole Effects. Most intriguing, however, are metabolite dependencies that involve multiple subsystems, physiological interactions, or transport mechanisms and that cannot be explained by conventional metabolic pathways at all, which can be thought of as wormhole connections. For example, the dependency between taurine and β-alanine (hydrazine case, Figure 2B) may occur because these compounds can compete for the same renal transport system.[36] This causes a physiological link at the renal tubule level so that taurine metabolism has a remote effect on β-alanine reabsorption from the urine. Individual modulations in the levels of urinary taurine and β-alanine have been previously reported as biomarkers of general hepatotoxicity, although the excretion of the latter has been attributed to concomitant renal failure.[37] The dependencies linking hippurate with citrate and with 2-OG (Figure 2C) also identify a shared cellular compartment (mitochondria); here, benzoic acid of gut microbiotal origin[17] is conjugated with glycine via a mitochondrial acetyl-Coenzyme A activation step to form hippuric acid,[38] in colocation with the topographical center for TCA cycle activity. This illustrates the ability of *in vivo* interactome maps to transcend species boundaries and allow virtual linkage of the mammalian primary metabolome (under host genomic control) with the cometabolome, which is partly under symbiotic gut microbial control.[17] The importance of transgenomic cometabolic interactions in the development of many disease states is now well documented,[1,2,17] and our new modeling approach appears ideally suited to probing such system complexity.

Conclusions

The use of metabolic interactome maps provides automated identification of metabolite dependencies for the rat undergoing toxic challenge from model drugs, enabling the visualization of short- and long-range metabolite dependencies, confirming known relationships between compounds, and identifying relationships that can result from interactions beyond currently established metabolic pathway maps. There is no need for prior knowledge of any underlying network of enzymes or transporters; therefore, the method could be used to generate hypotheses on undiscovered virtual pathways.

Data used here to construct these maps encompass all sampling times and hence do not directly capture the temporal progression of the toxic lesions, which we have previously modeled using other methods.[30,39–41,44] Nevertheless, the method could be extended to model the dynamics of dependency relationships.

Unlike mathematical models derived for single cell systems or simple organisms such as yeast, interactome maps of biofluids can capture the metabolic cross-talk between spatially disparate tissues and organs, providing a more holistic view of the metabolic status of complex organisms. Thus, with further development, a range of genetic and environmental factors including nutritional states, level of oxidative stress, age, hormone levels, and so forth might be visualized in terms of metabolic network activity.

Abbreviations: NMR, nuclear magnetic resonance; KEGG, Kyoto Encyclopedia of Genes and Genomes; BST, biochemical systems theory; MCA, metabolic control analysis; FBA, flux balance analysis; ANIT, α-napthylisothiocyanate; TCA, tricarboxylic acid; 2-AA, 2-aminoadipate; 2-OG, 2-oxoglutarate.

Acknowledgment. We thank C. Caulcott, N. Cooper, C. Rawlings, S.-I. Bang, T. Ebbels, and R. Chaleil for discussions and S. Islam for technical support. This work was supported by a Department of Trade and Industry (DTI) Beacon project and partially supported by METAGRAD. We are grateful to COMET researchers for histopathology results, in particular, Glenn Cantor, Eric Destexhe, Vibeke Aarup, and Zuhal Thorsteinsson, and to the COMET project as a whole for access to raw NMR data.

References

(1) Nicholson, J. K.; Wilson, I. D. Understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discovery* **2003**, *2*, 668–676.
(2) Nicholson, J. K.; Holmes, E.; Lindon, J. C.; Wilson, I. D. The challenges of modeling mammalian biocomplexity. *Nat. Biotechnol.* **2004**, *22*, 1268–1274.
(3) Nicholson, J. K. Global systems biology, personalized medicine and molecular epidemiology. *Mol. Syst. Biol.* **2006**, *2*, 52.
(4) Tomita, M. E-CELL: software environment for whole-cell simulation. *Bioinformatics* **1999**, *15*, 72–84.
(5) Hood, L. Systems biology: integrating technology, biology, and computation. *Mech. Ageing Dev.* **2003**, *124* (1), 9–16.
(6) Xia, Y. Analyzing cellular biochemistry in terms of molecular networks. *Annu. Rev. Biochem.* **2004**, *73*, 1051–1087.
(7) Green, M. L.; Karp, P. D. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinf.* **2004**, *5*, 76.
(8) Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; Barabási, A. L. The large scale organisation of metabolic networks. *Nature* **2000**, *407*, 651–653.
(9) Ravasz, E.; Somera, L.; Mongru, D. A.; Oltvai, Z. N.; Barabási, A. L. Hierarchical organization of modularity in metabolic networks. *Science.* **2002**, *297*, 1551–1555.
(10) Papin, J. A.; Price, N. D.; Wiback, S. J.; Fell, D. A.; Palsson, B. O. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **2003**, *28* (5), 250–258.
(11) Goodwin, B. C. *Oscillatory Organization in Cells, a Dynamic Theory of Cellular Control Processes*; Academic Press: New York, 1963.
(12) Hess, B.; Boiteux, A. Mechanism of glycolytic oscillation in yeast. I. Aerobic and anaerobic growth conditions for obtaining glycolytic oscillation. *Hoppe-Seyler's Z. Physiol. Chem.* **1968**, *349* (11), 1567–1574.
(13) Tyson, J. J.; Othmer, H. G. The dynamics of feedback control circuits in biochemical pathways. *Prog. Theor. Biol.* **1978**, *5*, 1–62.
(14) Edwards, J. S.; Ramakrishna, R.; Schilling, C. H.; Palsson, B. O. Metabolic Flux Balance Analysis. In *Metabolic Engineering*, Lee, S. Y., and Papoutsakis, E. T., Eds.; Marcel Dekker: New York, 1999.
(15) Friedman, N.; Linial, M.; Nachman, I.; Pe'er, D. Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **2000**, *7*, 601–620.
(16) Imoto, S.; Goto, T.; Miyano, S. Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression. *Pac. Symp. Biocomput.* **2002**, 175–186.
(17) Nicholson, J. K.; Holmes, E.; Wilson, I. D. Gut microorganisms, mammalian metabolism and personalized health care. *Nat. Rev. Microbiol.* **2005**, *3*, 431–438.
(18) Brindle, J. T.; Antti, H; Holmes, E. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. *Nature Med.* **2002**, *8* (12), 1439–1444.
(19) Pearl, J. Probabilistic reasoning in intelligent systems: networks of plausible inference, Morgan Kaufmann, San Mateo CA. 1988.

(20) Beaumont, M. A.; Rannala, B. The Bayesian revolution in genetics. *Nat. Rev. Genet.* **2004**, *5* (4), 251–261.
(21) Stoyanova, R.; Nicholson, J. K.; Lindon, J. C.; Brown, T. R. Sample classification based on Bayesian spectral decomposition of metabonomic NMR data sets. *Anal. Chem.* **2004**, *76* (13), 3666–3674.
(22) Bollard, M. E.; Keun, H; Beckonert, O.; Ebbels, T. M. D.; Antti, H.; Nicholls, A. W.; Shockcor, J. P.; Cantor, G. H.; Stevens, G.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. Comparative metabonomics of differential hydrazine toxicity in the rat and mouse. *Toxicol. Appl. Pharmacol.* **2005**, *204*, 135–151.
(23) Garrod, S.; Bollard, M. E.; Nicholls, A. W.; Connor, S. C.; Connelly, J.; Nicholson, J. K.; Holmes, E. Integrated metabonomic analysis of the multiorgan effects of hydrazine toxicity in the rat. *Chem. Res. Toxicol.* **2005**, *18*, 115–122.
(24) Nicholls, A. W.; Holmes, E.; Lindon, J. C.; Shockcor, J. P.; Duncan Farrant, R.; Haselden, J. N.; Damment, S. J. P.; Waterfield, C. J.; Nicholson, J. K. Metabonomic investigations into hydrazine toxicity in the rat. *Chem. Res. Toxicol.* **2001**, *14*, 975–978.
(25) Waters, N. J.; Holmes, E.; Williams, A.; Waterfield, C. J.; Duncan Farrant, R.; Nicholson, J. K. NMR and pattern recognition studies on the time-related metabolic effects of α-naphthylisothiocyanate on liver, urine and plasma in the rat: an integrative metabonomic approach. *Chem. Res. Toxicol.* **2001**, *14*, 1401–1412.
(26) Hastie, T.; Tibshirani, R.; Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*; Springer: New York, 2001.
(27) Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Antti, H.; Bollard, M. E.; Keun, H.; Beckonert, O.; Ebbels, T. M.; Reily, M. D.; Robertson, D.; Stevens, G. J.; Luke, P.; Breau, A. P.; Cantor, G. H.; Bible, R. H.; Niederhauser, U.; Senn, H.; Schlotterbeck, G.; Sidelmann, U. G.; Laursen, S. M.; Tymiak, A.; Car, B. D.; Lehman-McKeeman, L.; Colet, J.-M.; Loukaci, A.; Thomas, C. Contemporary issues in toxicology. The role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharmacol.* **2003**, *187*, 137−146.
(28) Keun, H. C.; Ebbels, T. M.; Antti, H.; Bollard, M. E.; Beckonert, O.; Schlotterbeck, G.; Senn, H.; Niederhauser, U.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Analytical reproducibility in ^{1}H NMR-based metabonomic urinalysis. *Chem. Res. Toxicol.* **2002**, *15*, 1380–1386.
(29) Claridge, T. D. W. *High-Resolution NMR Techniques in Organic Chemistry*; Elsevier: New York, **1999**.
(30) Keun, H. C.; Ebbels, T. M. D.; Antti, H.; Bollard, M. E.; Beckonert, O.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *Anal. Chim. Acta* **2003**, *490*, 265–276.
(31) Crockford, D. J.; Keun, H. C.; Smith, L. M.; Holmes, E.; Nicholson, J. K. Curve-fitting method for direct quantitation of compounds in complex biological mixtures using^{1}H NMR: application in metabonomic toxicology studies. *Anal. Chem.* **2005**, *77*, 4556–4562.
(32) Steuer, R.; Kurths, J.; Daub, C. O.; Weise, J.; Selbig, J. The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics* **2002**, *18* (Suppl. 2), S231–S240.
(33) Bang, J.-W.; Gillies, D. *Using Bayesian Networks with Hidden Nodes to Recognize Neural Cell Morphology*; Springer-Verlag: Tokyo, Japan, 2002.
(34) Connor, S. C.; Wu, W.; Sweatman, B. C.; Manini, J.; Haselden, J. N.; Crowther, D. J.; Waterfield, C. J. Effects of feeding and body weight loss on the ^{1}H-NMR-based urine metabolic profiles of male Wistar Han rats: implications for biomarker discovery. *Biomarkers* **2004**, *9*(2), 156–179.
(35) Wu, H. Q.; Ungerstedt, U.; Schwarcz, R. L-α-aminoadipic acid as a regulator of kynurenic acid production in the hippocampus: a microdialysis study in freely moving rats. *Eur. J. Pharmacol.* **1995**, *281* (1), 55–61.
(36) Jessen, H.; Sheikh, M. I. Renal transport of taurine in luminal membrane vesicles from rabbit proximal tubule. *Biochim. Biophys. Acta* **1991**, *1064* (2), 189–198.
(37) Waterfield, C. J.; Turton, J. A.; Scales, M. D.; Timbrell, J. A. Investigations into the effects of various hepatotoxic compounds on urinary and liver taurine levels in rats. *Arch. Toxicol.* **1993**, *67*, 244–254.
(38) Krähenbühl, L.; Reichen, J.; Talos, C.; Krähenbühl, S. Benzoic acid metabolism reflects hepatic mitochondrial function in rats with long-term extrahepatic cholestasis. *Hepatology* **1997**, *25* (2), 278–283.
(39) Holmes, E.; Bonner, F. W.; Sweatman, B. C.; Lindon, J. C; Beddell, C. R.; Rahr, E.; Nicholson, J. K. Nuclear magnetic resonance spectroscopy and pattern recognition analysis of the biochemical processes associated with the progression of and recovery from nephrotoxic lesions in the rat induced by mercury(II) chloride and 2-bromoethanamine. *Mol. Pharmacol.* **1992**, *42* (5), 922–930.
(40) Azmi, J.; Griffin, J.; Antti, H.; Shore, R. F.; Johansson, E.; Nicholson, J. K.; Holmes, E. Metabolic trajectory characterisation of xenobiotic-induced hepatotoxic lesions using statistical batch processing of NMR data. *The Analyst* **2002**, *127* (2), 271–276.
(41) Ebbels, T.; Keun, H.; Beckonert, O. Toxicity classification from metabonomic data using a density superposition approach: 'CLOUDS'. *Anal. Chim. Acta* **2003**, *490*, 109–122.
(42) Dumas, M. E.; Barton, R. H.; Toye, A.; Cloarec, O.; Blancher, C.; Rothwell, A.; Fearnside, J.; Tatout, R.; Blanc, V.; Lindon, J. C.; Mitchell, S. C.; Holmes, E.; McCarthy, M. I.; Scott, J.; Gauguier, D.; Nicholson, J. K. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (33), 12511–12516.
(43) Mitchell, S. C.; Bollard, M. E.; Zhang, A. Short-chain aliphatic amines in human urine: a mathematical examination of metabolic interrelationships. *Metab., Clin. Exp.* **2007**, *56*, 19–23.
(44) Ebbels, T. M. D.; Keun, H. C.; Beckonert, O. P.; Bollard, M. E.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. Prediction and classification of drug toxicity using probabilistic modeling of temporal metabolic data: the consortium on metabonomic toxicology screening approach. *J. Proteome Res.* **2007**, *6* (11), 4407–4422.

PR070350L