INFERRING THE HEPATIC NETWORK FROM EXPERIMENTAL MEASUREMENTS USING BAYESIAN NETWORK ANALYSIS

Christina Chan
Zheng Li

Michigan State University
Chemical Engineering and Material Science Department

INTRODUCTION

Numerous mathematical models, from abstract discrete Boolean networks to detailed biological mechanistic-based models, have been applied to represent biological networks. Thus far, these algorithms have been used predominantly to reverse engineer the genetic regulatory network from both experimental and simulated data [1]. The goal of these models is to aid our understanding of the underlying mechanism that governs the transition of system or biological states. However, current reverse engineering methods have their limitations. For example, Boolean network models capture well only the logical relationships in a system. Likewise, deterministic models require known, detailed biochemical and biological reactions and pathways, restricting its applicability to relatively small systems. What's more, all these reverse engineering methods are data driven, thus precluding statistical hypothesis testing of possible regulatory models [2]. So it is the purpose of this paper to combine data-driven reverse engineering and model-driven statistical testing methods to infer the biochemical network from several possible hypothetical models.

Using Bayesian network (BN) analysis we developed a framework to infer the regulatory network from experimental metabolic data and quantify the identified pathways with conditional probabilities. This method originally was applied to gene regulatory network by [1]. In our framework, data driven BN analysis was first applied to learn the raw metabolic regulatory network from the data itself. The inferred network was then compared with the known metabolic network to evaluate the ability of our framework to learn for example the TCA and urea cycles. Next, we identified direct and indirect pathways that causally influence an objective function, in particular, biological functions, such as triglyceride accumulation and the rate of urea synthesis. Methods, such as inductive causation (IC*), were applied to the sub-network composed of the identified pathways with the goal of finding possible latent variables in the network by discovering indirect pathways relevant to our objective function. Finally, based upon the causal relationships learned from the data and currently known biological associations, we postulated several alternative metabolic regulatory network models. The utility and accuracy of these alternatives models were evaluated by comparing their Bayesian metric scores, calculated based upon the experimental data, to assess 1) a model's goodness of fit to the data and 2) how well a model predicted the objective functions. Thereupon an optimal framework to define the metabolic regulatory network was selected based upon these criteria.

MATERIAL AND METHODS

All materials and biochemical assays, including oxygen uptake rate, employed to obtain metabolite measurements are described elsewhere [3]. Hepatocytes were isolated from adult female Lewis rats. The isolated hepatocytes were cultured in a collagen sandwich configuration and incubated in standard hepatocyte culture medium containing either 0.5 U/ml insulin (high insulin) or 50 µU/ml insulin (low insulin). The hepatocytes were cultured in this fashion for at least 6 days prior to plasma exposure. This interval is considered the pre-conditioning period. The six-day old-sandwiched hepatocyte cultures were subsequently exposed to unsupplemented, amino acid, hormone, or amino acid plus hormone supplemented plasma solution for an additional seven days (see Chan et al. [3] for details). Therefore, six combinations of pre-conditioning-plasma supplementation were evaluated. They were i) low insulin pre-conditioned and unsupplemented plasma, ii) low insulin pre-conditioned and amino acid supplemented plasma, iii) high insulin pre-conditioned and unsupplemented plasma, iv) high insulin pre-conditioned and amino acid supplemented plasma, v) high insulin pre-conditioned and hormone supplemented plasma, and vi) high insulin pre-conditioned and amino acid plus hormone supplemented plasma. A model for hepatocyte metabolism was created based on the known stoichiometry of the hepatic metabolic network. Metabolite measurements were applied to Metabolic Flux Analysis (MFA) to obtain intracellular fluxes.

The approaches developed to learn the BN structure from either experimental or simulated data generally fall into two categories: search and scoring based approach, or constraint-based approach

(dependency analysis). The BN is composed of a set of nodes, which represent the random variables, and a set of arcs connecting pairs of nodes, which denote a certain relation between the variables or events. By using search and scoring based methods, the algorithm starts with a graph without arcs, then new arcs are added and a new scoring function is calculated, if the score improves, the new arc is kept, otherwise it is eliminated, thus, the BN is learned through optimizing the scoring function. The constraint-based approach applies conditional independency test to uncover the dependencies within the data, which are used to reconstruct the BN. All the variables in the BN are taken as nodes and the links between the nodes represent the causal dependencies among the variables. The dependencies are quantified by conditional probabilities of each node given its parent nodes. Thus, the whole network is a representation of the joint distribution defined by a set of random variables. To develop a complete BN, we calculate the probability of every possible event, i.e. the joint distribution of the system, which are calculated as follows:

$$p(x1,..., x_n) = \prod_i p(x_i \mid pa_i) \quad (1)$$

where $p(x_i|pa_i)$ is the conditional probability of node i.

RESULTS

There were 36 samples with 6 groups/environmental conditions and 6 samples were collected for each condition to explore metabolic states through MFA and perform BN analysis on these states. For each sample, 34 metabolites were measured and 42 metabolic fluxes were estimated using MFA. All flux values including measurements and estimated fluxes are combined into one data matrix X (36×76). The data matrix X is then discretized into matrix D and BN analysis was applied to D to infer the regulatory metabolic network from the data. By applying the algorithm described in Materials and Methods, we were able to infer both the TCA and urea cycles from the flux data, as shown in Figure 1.

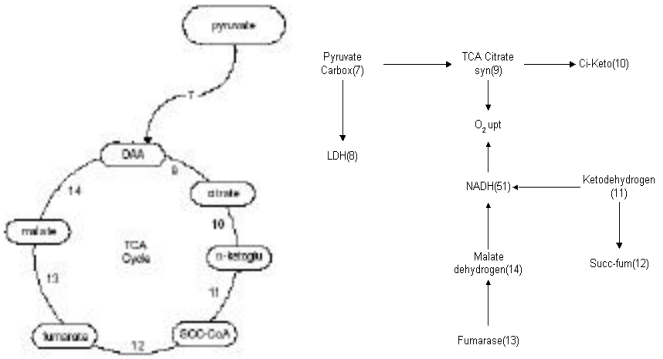
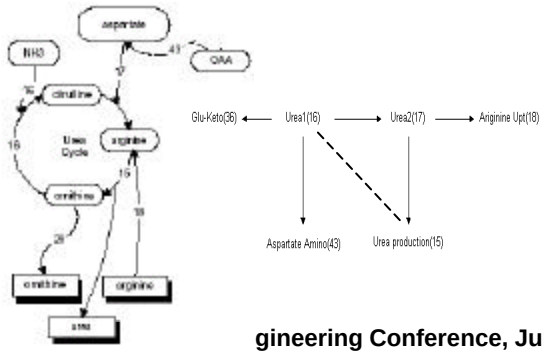


Figure 1. TCA cycle, top (right: actual; left: inferred) Urea cycle, bottom (right: actual; left: inferred)



The direct causal relationship between the TCA pathways 10-11 and 12-13 were not learned by BN analysis, but rather were identified as being coupled through oxidative phosphorylation (flux no. 51). Interestingly, in our model, pathway no. 10-11 combines two pathways (citrate - isocitrate - αketoglutarate) into one pathway (citrate - αketoglutarate), similarly with pathway no. 12-13.

The ability of BN analysis to infer the known metabolic network from flux data provides confidence in the causal relationships they identify. Oftentimes, the causal relations in biological systems are indirect or contain feedback loops that have yet to be discovered. BN analysis could identify some of these associations as latent variables. Latent variable detecting algorithm, such as IC*, was then applied to a sub-network related to a specific hepatocellular function, such as intracellular triglyceride accumulation, to obtain a more comprehensive sub-network than was used in MFA. Based on the latent variables identified, two alternative metabolic sub-networks were proposed and evaluated with Bayesian scoring function to determine the "closest" network based upon the data. To illustrate this concept we found from previous work [4] that the intracellular triglyceride accumulation was not very well described by the current MFA model. In this paper we used BN analysis to identify 1) the causal relationships relevant to intracellular triglyceride and 2) the latent variables that may exist and thus influence intracellular triglyceride accumulation. Combining the network inferred by BN analysis and the known pathways relevant to triglyceride metabolism, we postulated 2 alternative sub-networks:

The difference between the networks is the glutamine (GLN) pathway. The sub-networks were compared and network B was able to predict correctly intracellular triglyceride accumulation for 33 of the 36 cases vs. 30 of the 36 with network A. Thus GLN was identified as a latent variable that influenced intracellular triglyceride accumulation.

CONCLUSION

Reverse engineering networks with BN analysis could provide a framework for predictive modeling, hypothesis testing and systematic experimental verification to gain deeper insight into our understanding of biological systems.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation.

REFERENCE

1. Friedman, N., Linial, M., Nachman, I., and Pe' er, D., "Using Bayesian Networks to Analyze Expression Data", J. Comp. Bio., 7, 2000: 601-520.
2. Hartemink, A.J., Gifford, D.K., Jaakkola, T.S., and Young, R.A., "Using Graphical Models and Genomic Expression Data to Statistically Validate Models of Genetic Regulatory Networks", Pacific Symposium on Biocomputing 6, 2001: 422-433.
3. Chan, C., F., Berthiaume, F., Lee, K., and Yarmush, M. L., "Metabolic Flux Analysis of Cultured Hepatocytes Exposed to Plasma", Biotechnology and Bioengineering, 81, 2003: 33-49.
4. Chan, C., Hwang, D. H., Stephanopoulos, G. N., Yarmush, M. L., and Stephanopoulos, G., "Application of Multivariate Analysis to Optimize Function of Cultured Hepatocytes", (accepted, Biotechnology Progress).