Analysis of Heterogeneous Regulation in Biological Networks

Irit Gat-Viks [1] [2], Amos Tanay [1] [2], Ron Shamir [1]

Keywords: genetic and metabolic regulation, high-throughput data analysis, model learning algorithms.

1 Motivation.

Biological systems employ heterogeneous regulatory mechanisms that are frequently interrelated. For example, the rates of metabolic reactions are strongly coupled to the concentrations of their catalyzing enzymes, which are themselves subject to complex genetic regulation. Such regulation is in turn frequently affected by metabolite concentrations. Metabolite-mRNA-enzyme-metabolite feedback loops have a central role in many biological systems and call for modeling and learning of heterogeneous regulatory mechanisms.

2 Method.

In this work we study steady state behavior of biological systems that are stimulated by changes in the environment (e.g., lack of nutrients) or by internal perturbations (e.g., gene knockouts). Our model of the system contains variables of several types, representing diverse biological factors such as mRNAs, proteins and metabolites. Interactions between biological factors are formalized as regulation functions which may involve all variable types and complex combinatorial logic. Our model combines metabolic pathways (cascades of metabolite variables), genetic regulatory circuits (networks of mRNAs and transcription factors protein variables), protein networks (cascades of post-translational interactions among protein variables) and the relations among them (metabolites may regulate transcription, enzymes may regulate metabolic reactions). We show how such models can be built from the literature and develop computational techniques for their analysis and refinement given a collection of heterogeneous high-throughput experiments. We develop algorithms to learn novel regulation functions in lieu of ones that manifest inconsistency with the experiments.

Our approach is innovative in several aspects:

• We model a variety of variables types, extending beyond gene network studies, that focus on mRNA, and metabolic pathways methods, that focus on metabolites. Consequently, our model can express effects of translation regulation and post translational modifications.

• Our approach allows handling feedback loops as part of the inference and learning process. This is crucial for adequate joint modeling of metabolic reactions and genetic regulation.

• We build an initial model based on prior knowledge, and then aim to improve (expand) this model based on experimental data. We show that formal modeling of the prior knowledge allows the interpretation of high throughput experiments in new level of detail.

• Our algorithms learn new transcription regulation functions by analyzing together gene expression, protein expression and growth phenotypes data.

[1]School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Email: {iritg,amos,rshamir}@post.tau.ac.il.

[2]These authors contributed equally to this work.

3 Experimental Results

Our methodologies and ideas were implemented in a new software tool called MetaReg. It facilitates evaluation of a model versus diverse experimental data, detection of variables that manifest inconsistencies between the model and the data, and learning optimized regulation functions for such variables. Graphical representations provide overview of results and highlight model inconsistencies. We used MetaReg to study the pathway of lysine biosynthesis in yeast. We performed an extensive literature survey and organized the knowledge on the pathway into a model consisting of about 150 variables. As part of model construction, we reviewed the results of many low throughput experiments and included in the model the most plausible regulation function of each variable. We assessed the model versus a heterogeneous collection of experimental results, consisting of gene expression, protein expression and phenotype growth sensitivity profiles. In general, model inference agreed well with the observations, confirming the effectiveness of our strategy. In several important cases, however, significant modeling discrepancies indicated gaps in the current biological understanding of the system. Using our learning algorithm we generated novel regulation hypotheses that bridge some of these gaps. We also showed that our method attains improved accuracy in comparison to extant network learning methods.