# Bioinformatics Center
## – Mathematical Bioinformatics –

http://www.bic.kyoto-u.ac.jp/takutsu/index.html

Prof
AKUTSU, Tatsuya
(D Eng)

Assoc Prof
TAMURA, Takeyuki
(D Inf)

Assist Prof
MORI, Tomoya
(D Inf)

SPD (JSPS)
LIN, Chun-Yu
(Ph D)

SPD (JSPS)
MÜNZNER,
Ulrike Tatjana Elisabeth
(Ph D)

### Students

LIU, Pengyu (D3)
YU, Coleman (D2)
TAKAGI, Motoshige (D2)
LI, Ruiming (D1)
WANG, Feiqi (D1)
OHTOMO, Masahiro (D1)
CAO, Yu (M2)
MARUTA, Kunpei (M2)
PI, Wenya (M2)
KUMANO, Sou (M1)

### Guest Res Assoc

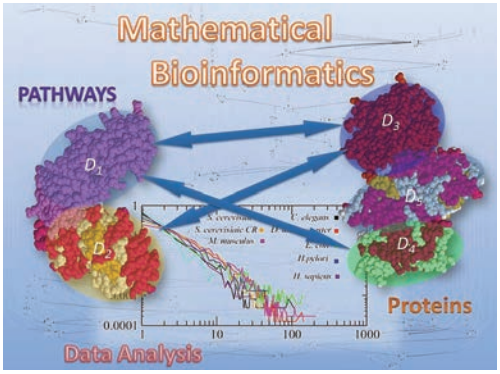
GUO, Wei-Feng (Ph D) Northwestern Polytechnical University, China, P.R., 22 Novemver 2018–1 February 2019

## Scope of Research

Due to the rapid progress of genome sequencing technology, whole genome sequences of organisms ranging from bacteria to human have become available. In order to understand the meaning behind the genetic code, we have been developing algorithms and software tools for analyzing biological data based on advanced information technologies such as theory of algorithms, artificial intelligence, and machine learning. We are currently studying the following topics: systems biology, scale-free networks, protein structure prediction, the inference of biological networks, chemo-informatics, and discrete and stochastic methods for bioinformatics.

### KEYWORDS

Complex Networks
Boolean Networks
Neural Networks
Chemical Graphs
Protein Informatics




### Selected Publications

Akutsu, T., Algorithms for Analysis, Inference, and Control of Boolean Networks, *World Scientific* (2018).

Melkman, A. A.; Cheng, X.; Ching, W. K.; Akutsu, T., Identifying a Probabilistic Boolean Threshold Network from Samples, *IEEE Transactions on Neural Networks and Learning Systems*, **29,** 869-881 (2018).

Akutsu, T.; Jansson, J.; Li, R.; Takasu, A.; Tamura, T., New and Improved Algorithms for Unordered Tree Inclusion, *Proc. 29th Int. Symp. Algorithms and Computation*, [27-1]-[27-12] (2018).

Li, J.; Nagamochi, H.; Akutsu, T., Enumerating Substituted Benzene Isomers of Tree-like Chemical Graphs, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **15,** 633-646 (2018).

Bao, Y.; Hayashida, M.; Liu, P.; Ishitsuka, M.; Nacher, J. C.; Akutsu, T., Analysis of Critical and Redundant Vertices in Controlling Directed Complex Networks Using Feedback Vertex Sets, *J. Comput. Biol.*, **25,** 1071-1090 (2018).

## Exact Identification of the Structure of a Probabilistic Boolean Threshold Network from Samples

Various kinds of mathematical models have been utilized for understanding dynamical behavior of biological systems. Among them, the Boolean network (BN) is a simple but well-studied discrete model, especially for modeling genetic regulatory networks and for neural networks. In a BN, each node takes a Boolean value, 0 or 1, at each time step, and the states of all nodes are updated synchronously according to Boolean functions assigned to nodes, where each node corresponds to a gene or neuron, and 1 and 0 mean that genes/neurons are active and inactive, respectively. The probabilistic Boolean network (PBN) is a probabilistic extension of BN in which multiple Boolean functions can be assigned to each node and one function is randomly selected at each time step according to the prescribed probability distribution. The probabilistic Boolean threshold network (PTBN) is a restriction of PBN in which every Boolean function is limited to threshold functions: $w_1x_1 + \ldots + w_nx_n \geq \theta$ (Figure 1). Note that Boolean threshold functions have been extensively studied for analyzing the ability of neurons and neural networks.

In our previous work, we studied the exact identification of the structure of a PBN from samples, which is potentially important for identifying genetic network structures in cells and/or neural network structures in brains. In this work, we extend our previous approach to the analysis of PBTN. We show that wide classes of PBTN with unit coefficients can be exactly identified from samples under reasonable constraints, which include: 1) PBTNs consisting of pairs of threshold functions with different numbers of input variables and 2) PBTNs in which any number of threshold functions can be assigned provided that all functions have the same number of input variables. We also show that the problem of deciding the equivalence of two Boolean threshold functions is solvable in pseudo-polynomial time but remains computationally hard (co-NP complete).

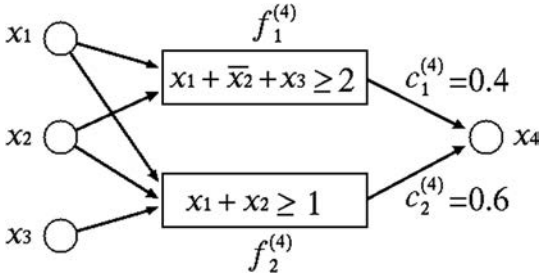



**Figure 1.** Part of a probabilistic Boolean threshold network. Two threshold functions are assigned to node $x_4$: one is selected with probability 0.4 and the other with probability 0.6.

## Grid-based Computational Methods for the Design of Constraint-based Parsimonious Chemical Reaction Networks to Simulate Metabolite Production: GridProd

Constraint-based metabolic flux analysis of knockout strategies is an efficient method to simulate the production of useful metabolites in microbes. Owing to the recent development of technologies for artificial DNA synthesis, it may become important in the near future to mathematically design minimum metabolic networks to simulate metabolite production.

We developed an efficient method for computing the design of minimum metabolic networks by using constraint-based flux balance analysis to simulate the production of useful metabolites. When the growth rate of this obtained parsimonious metabolic network is maximized, higher production rates are observed for many target metabolites when compared to existing methods. The set of reactions used in this parsimonious flux distribution consists of reactions included in the original genome scale model iAF1260. Under the conditions that the growth rate is maximized and the minimum cases of flux variability analysis are considered, the developed method produced more than 90% of metabolites, while the existing methods produced less than 50%.

The source code is freely available and is implemented in MATLAB and COBRA toolbox.

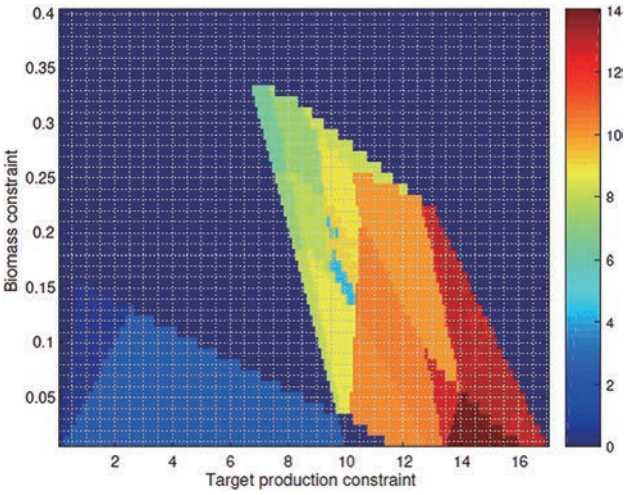



**Figure 2.** An example of the production rate heatmap for a target metabolite where each grid represents constraints for the ranges by the production rate and the growth rate.