Bioinformatics Center
– Mathematical Bioinformatics –

https://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)

Program-Specific Res
MÜNZNER,
Ulrike Tatjana Elisabeth
(Ph D)

Students

LIU, Pengyu (D3)
YU, Coleman (D3)
TAKAGI, Motoshige (D3)
LI, Ruiming (D2)
WANG, Feiqi (D2)
OHTOMO, Masahiro (D2)
NAKASHIMA, Shogo (D1)
KUMANO, Sou (M2)
YANG, Ziwei (RS)
MA, Yier (RS)

Guest Res Assoc

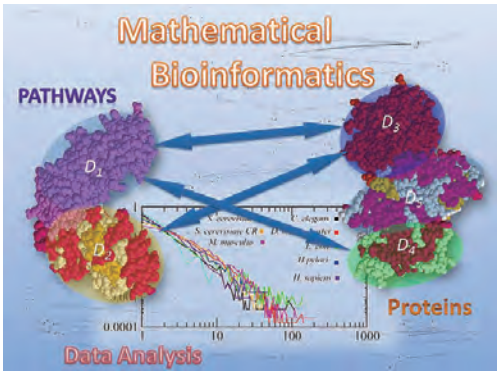
LEE, Jung-Yu (Ph D) National Chiao Tung University, China, P.R., 18 November–6 December
CHEN, Yun-Ti (Ph D) National Chiao Tung University, China, P.R., 18 November–6 December
LIN, Xiang-Yu (Ph D) National Chiao Tung University, China, P.R., 18 November–6 December

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.; Melkman, A. A., Identification of the Structure of a Probabilistic Boolean Network from Samples Including Frequencies of Outcomes, *IEEE Transactions on Neural Networks and Learning Systems*, **30,** 2383-2396 (2019).

Schwartz, J.-M.; Otokuni, H.; Akutsu, T.; Nacher, J. C., Probabilistic Controllability Approach to Metabolic Fluxes in Normal and Cancer Tissues, *Nat. Commun.*, **10,** [2725-1]-[2725-9] (2019).

Itami-Matsumoto, S.; Hayakawa, M.; Uchida-Kobayashi, S.; Enomoto, M.; Tamori, A.; Mizuno, K.; Toyoda, H.; Tamura, T.; Akutsu, T.; Ochiya, T.; Kawada, N.; Murakami, Y., Circulating Exosomal miRNA Profiles Predict the Occurrence and Recurrence of Hepatocellular Carcinoma in Patients with Direct-Acting Antiviral-Induced Sustained Viral Response, *Biomedicines*, **7,** [81-1]-[81-14] (2019).

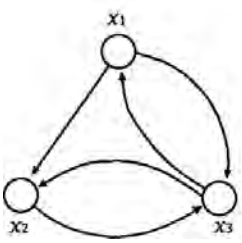
Marini, S.; Vitali, F.; Rampazzi, S.; Demartini, A.; Akutsu, T., Protease Target Prediction via Matrix Factorization, *Bioinformatics*, **35,** 923-929 (2019).

Hou, W.; Ruan, P.; Ching, W.-K.; Akutsu, T., On the Number of Driver Nodes for Controlling a Boolean Network When the Targets Are Restricted to Attractors, *J. Theor. Biol.*, **463,** 1-11 (2019).

Identification of the Structure of a Probabilistic Boolean Network from Probability Distribution of Samples

In order to understand dynamical behavior of biological systems, various kinds of mathematical models have been utilized. Among them, the Boolean network (BN) is one of the simplest non-linear models, in which 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. In a BN, each node corresponds to a gene or neuron, and 1 and 0 mean that genes/neurons are active and inactive, respectively. Although BN is an old model proposed more than 50 years ago, extensive studies are still being done. However, real biological systems contain noise and other stochastic factors. In order to cope with these stochastic factors, the probabilistic Boolean network (PBN) has been proposed and utilized, 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 (Figure 1). We have been studying 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 our previous work, we studied the identification of the structure (*i.e.*, set of Boolean functions assigned for each node) of a PBN using information only on the occurrences of samples. In this work, we make use of additional information obtainable from samples: the frequencies of occurrences of sub-tuples. We show that under this model, it is possible to identify a PBN for much broader classes of PBNs. In particular, we prove under a reasonable assumption that the structure of a PBN can be identified from among the class of PBNs that have at most three functions assigned to each node, but that identification may be impossible if four or more functions are assigned to each node. We also present an efficient algorithm for the identification of a PBN consisting of threshold functions from samples.



	Boolean Function	$p_j^{(i)}$
$f_1^{(1)}$	$x_1(t+1) = \overline{x_3(t)}$	0.7
$f_2^{(1)}$	$x_1(t+1) = x_3(t)$	0.3
$f_1^{(2)}$	$x_2(t+1) = x_1(t) \wedge \overline{x_3(t)}$	1.0
$f_1^{(3)}$	$x_3(t+1) = x_1(t) \wedge \overline{x_2(t)}$	0.8
$f_2^{(3)}$	$x_3(t+1) = x_2(t)$	0.2

Figure 1. Example of a probabilistic Boolean threshold network. Boolean functions are independently selected at each time step according to the prescribed probabilities shown in the table.

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.

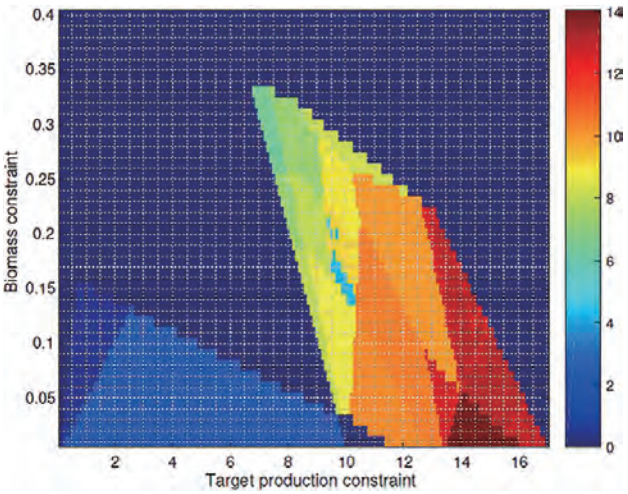


Figure 2. An example of the production rate heatmap for a target metabolite where each grid represents constraints for the ranges by the production rate and the growth rate.