The First International Symposium on Optimization and Systems Biology (OSB'07)
Beijing, China, August 8–10, 2007
, pp. 5–12

Computational Systems Biology on Networks and Dynamics

Luonan Chen*

Institute of Systems Biology, Shanghai University, Shanghai, 200444, China
Department of Electrical Engineering, Osaka Sangyo University, Osaka 574-8530, Japan
ERATO Aihara Complexity Modeling Project, JST, Tokyo 153-8505, Japan
Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan.

1 Introduction

One of major challenges for post-genomic biology is to understand how genes, proteins and small molecules interact to form cellular systems. It has been recognized that a complicated living organism cannot be fully understood by merely analyzing individual components, and that interactions and dynamics of those components or networks are ultimately responsible for an organism's form and functions. Instead of analyzing individual components or aspects of the organism, systems biology is to study an organism, viewed as a dynamical and interacting network of genes, proteins and biochemical reactions which give rise to life. In recent years, with rapid progress of biological science, many high-throughput technologies have been developed for systematically studying interactions or networks of molecules, such as microarray, the two-hybrid assay, co-immunoprecipitation and the ChIP-chip approach, which can be used to screen for protein-protein interaction (PPI) or to infer gene regulatory network. With increasingly accumulated data from those high-throughput technologies, molecular networks and their dynamics have been studied extensively from various aspects of living organisms. Those research works help biologists not only to understand complicated biochemical phenomena but also to elucidate the essential principles or fundamental mechanisms of cellular systems at the system-wide level. This paper focuses on reviewing our recent works related to systems biology from the theoretical and engineering perspective, in particular emphasizing on aspects of interactions, dynamics, networks and integration for biological systems.

2 Dynamics

Dynamics exist in living organisms at all levels. From both theoretical and experiment viewpoints, it is a greatly challenging problem in biological science to model, analyze and further predict the dynamic behaviors of biosystems. One of the

*Email: chen@eic.osaka-sandai.ac.jp

best studied dynamic or rhythmic phenomena so far is circadian oscillations, which are assumed to be produced by limit cycle oscillators at the molecular level from the gene regulatory feedback loops. With the rapid advances in mathematics and experiments concerning the underlying regulatory mechanisms, more sophisticated theoretical models and general techniques are increasingly demanded to elucidate dynamical behaviors in a cell at a system-wide level. Closely related to systems biology, synthetic biology is also a new area of research that combines science and engineering in order to design and build novel biological functions and systems, and requires the techniques of systems biology. Research in synthetic biology is aimed to combine knowledge from various disciplines including molecular biology, engineering and mathematics to design and implement new cellular behaviors. Recent progress in genetic engineering has made the design and implementation of artificial or synthetic gene networks realistic from both theoretical and experimental viewpoints. Actually, from the theoretical predictions, several simple gene networks have been experimentally constructed, e.g. genetic toggle switch and repressilator. Such simple models clearly represent a first step towards logical cellular control by manipulating and monitoring biological processes at the DNA level, and not only can be used as building blocks to synthesize the artificial biological systems, but also have great potential for biotechnological and therapeutic applications [34].

For this area, we proposed a general framework for modeling gene regulatory networks and developed theoretical techniques to design bio-molecular networks, which can be summarized as follows.

2-1 Modeling gene regulatory networks [1-3] and analyzing their dynamics (stability and bifurcation) based on nonlinear dynamical theory and control theory (LMI, or Lur'e systems).
2-2 A general framework to model biomolecular networks with the consideration of cell cycle in a cell based on impulsive differential equations [4].
2-3 Designing gene switches and gene oscillators based on theory of monotone dynamical systems [5-8].
2-4 Designing general synthetic gene regulatory networks by exploiting the dynamical characteristics of positive and negative feedback loops [9-10].

From viewpoint of nonlinear dynamics, there are three major difficulties to analyze a large-scale biological system, i.e. nonlinearity, noises and delays. In the above works, we mainly tackled the problems for nonlinearity and time delays by exploiting the special properties of cellular systems, and will discuss how to handle the problem of noise in next section.

3 Collective Behavior

Cooperative behaviors are essential coordinated responses resulting from an integrated exchange of information by cell communication in both prokaryotes and eukaryotes. The ability to cooperate or communicate between cells is an absolute

requisite to ensure appropriate and robust coordination of cell activities at all levels of organisms under an uncertain environment. To understand the mechanism of cooperative behaviors (such as chemotaxis and quorum sensing) for molecules is an essential topic with systems biology, which requires both mathematical and biological knowledge and insight. Generally, cooperative behavior, such as intercellular communication is accomplished by transmitting individual cell reactions via intercellular signals to neighboring cells and further integrating to generate a global cellular response at the level of molecules, tissues, organs and a body. However, all cell components exhibit intracellular noises owing to random births and deaths of individual molecules, and extracellular noises owing to environment fluctuations. Gene regulation in particular, is an inherently noisy process which involves stochastic fluctuations owing to low copy numbers of many molecules per cell and uncertainty of an external environment. Such stochastic noises may not only affect the dynamics of the entire system but may also be exploited by living organisms to actively facilitate certain functions, such as synchronization and communication. For this problem, we developed both numerical algorithms and analytical theory to analyze the cooperative behaviors of a population of individual cells induced by the noises by considering effects of stochastic fluctuations and signal diffusion processes. The main results are summarized as follows.

3-1 Synchronizing genetic oscillators by signalling molecules [11-12] in multicellular systems with or without couplings.
3-2 Cooperative behaviors of coupled non-identical cells with or without noises [13-14], and transient Resetting mechanism for biological synchrony [15].
3-3 Synchronizing a multicellular system by external input: an artificial control strategy[16].
3-4 Molecular Communication through Stochastic Synchronization Induced by Extracellular Fluctuations [17-18] based on master equation, cumulant equations and stochastic simulation.

Based on analysis of deterministic and stochastic dynamics, all of those works indicated that noises may play an important role in living organisms for actively mediating cooperative dynamics [19].

4 Integration

Large-scale microarray gene expression data and proteomics data provide new ways to learn gene regulation or construct gene regulatory networks or signal pathways. Due to the scarcity of data, an important challenging problem of reverse engineering is how to integrate various information sources for deriving a reliable molecular network, e.g. gene regulatory network. In contrast to the conventional methods that are mainly based on a single time-course dataset, we recently proposed a novel method GNRinfer [20] to infer gene regulatory network which integrates multiple time-course microarray datasets even with different conditions in a universal framework. The method theoretically ensures the derivation of the most consistent network

structure with respect to all of the datasets, thereby not only significantly alleviating the problem of data scarcity but also remarkably improving the prediction reliability. Actually, GNRinfer can be further extended to identify the conserved network patterns or motifs from the datasets of either the same species or different species.

For annotating gene and protein functions, we also recently developed a novel technique based on the integration of various data sources [35], which demonstrates a superior performance over the existing methods.

5 Molecular Networks and Interactions

There are two major problems with experimental data generated by high-throughput experimental biotechnologies, i.e., insufficient amount and low quality for the available data. So far various 'omics' data have been adopted to infer molecular networks. Recently, an interesting class of algorithms based on statistic analysis was adopted to discover protein interactions at the domain level. With training data, those methods first calculate the probability of each domain pair, and then predict PPI based on the domain-domain interaction (DDI) information. Since DDI has a clear biological implication, it has been widely adopted to derive the PPI. For this problem, recently by exploiting special structures and composition of experimental data, we have developed a new Association Probabilistic Method (APM) [21] based on domain interaction to infer PPIs, which outperforms other existing methods in terms of prediction quality and computational efficiency. APM can also be applied to multiple organisms' data by a simple manipulation [21].

Molecular networks orchestrate the sophistic and complex functions of the living cells. Various organisms differ not only because of differences of constituting proteins, but also because of architectures of their molecular networks. Hence, it is essential to address the similarities and differences in the molecular networks by comparative network analysis, which can directly be applied for analyzing signal pathways, conserved regions, discovering new biological functions or understanding the evolution of protein interactions. The problem of network alignment is to detect subnets that are conserved across species or within species by comparing two networks. Due to the high complexity to compare such molecular networks, most of the conventional approaches either restrict comparative analysis to special structures, such as pathways without loops, or adopt heuristic algorithms due to computational burden. To overcome such difficulty for computational complexity, we developed an alignment tool MNAligner [22, 38] based on an integer quadratic programming model to align networks in an accurate and efficient manner. The method is rather general and can be applied not only to unweighted and undirected networks, but also to weighted and directed networks [22, 38].

6 Optimization

Computational systems biology or bioinformatics is the use of techniques from mathematics, informatics and computer science to solve biological problems. Many

mathematical methods have been adopted in computational systems biology. In particular, optimization and statistics play a key role in analyzing and understanding biological mechanisms from system-wide viewpoints. Research in computational biology often overlaps with computational systems biology. For this area, we developed various optimization techniques (e.g. linear programming, nonlinear programming, dynamical programming, machine learning) to solve the problems of biology, which are summarized as follows.

6-1 Haplotyping based on the dataset of SNPs [23-24] by mathematical programming.
6-2 Unique optimal foldings of a protein on a triangular lattice based on graphic theory [25-26].
6-3 Revealing divergent evolution, identifying circular permutations, and Detecting Active-Sites by Protein Structure Comparison [27-28].
6-4 Multiple structure alignment for proteins based on mathematical programming [29].
6-5 Protein identification using Peptide Mass Fingerprinting data [30] by statistical analysis.
6-6 A systems biology perspective on signal processing in genetic network motifs [31].
6-7 A new geometric-topological method to measure protein fold similarity [32], in particular for Outer Membrane Proteins [33].
6-8 Analysis of protein surface patterns by pocket similarity network in a systems biology framework [36].

7 Conclusion

One of the grand challenges in Systems Biology is to build a complete and high-resolution description of molecular topography and connect molecular interactions with physiological responses. By studying the relationships and interactions between various parts of a biological system [37], e.g. metabolic pathways, organelles, cells, physiological systems and organisms, we aim eventually to develop an understandable model of the whole system, which is a key both for understanding of life and for application of human medicine, in particular from the theoretical and engineering perspective. All software and related documents are available from http://www.isb.shu.edu.cn, http://intelligent.eic.osaka-sandai.ac.jp, or http://zhangroup.aporc.org/ResearchBioinformatics or upon request from authors. All of works listed in this paper are collaborations with groups of Prof. Xiang-Sun Zhang (Chinese Academy of Sciences) and Prof. Kazuyuki Aihara (The University of Tokyo).

References

[1] Chen L., Aihara K.: Stability of Genetic Regulatory Networks with Time Delay, *IEEE Trans. on Circuits and Systems – I*, **49**, 602-608, 2002.

[2] Chen L., Aihara K.: A Model of Periodic Oscillation for Genetic Regulatory Systems, *IEEE Trans. on Circuits and Systems – I*, **49**, 1429-1436, 2002.

[3] Li C., Chen L., Aihara K.: Stability of Genetic Networks with SUM Regulatory Logic, *IEEE Trans. on Circuits and Systems - I*, **53**, 2451-2458, 2006.

[4] Chen L., Wang R., Kobayashi T., Aihara K.: Dynamics of Gene Regulatory Networks with Cell Division Cycle, *Physical Review E*, **70**, 011909, 2004.

[5] Kobayashi T., Chen L., Aihara K.: Modeling Genetic Switches with Positive Feedback Loops, *Journal of Theoretical Biology*, **221**, 379-399, 2002.

[6] Wang R., Jing Z., Chen L.: Modelling Periodic Oscillation in Gene Regulatory Networks by Cyclic Feedback Systems, *Bulletin of Mathematical Biology*, **67**, 339-367, 2005.

[7] Wang R., Chen L., Aihara K.: Construction of Genetic Oscillators with Interlocked Feedback Networks, *Journal of Theoretical Biology*, **242**, 454-463, 2006.

[8] Wang R., Chen L., Aihara K.: Detection of Cellular Rhythms and Global Stability within Interlocked Feedback Systems, *Mathematical Biosciences*, in press, 2007.

[9] Wang R., Zhou T., Jing Z., Chen L.: Modelling Periodic Oscillation of Biological Systems with Multiple Time Scale Networks, *Systems Biology*, **1**, 71-84,2004.

[10] Chen L., Wang R.: Designing Gene Regulatory Networks with Specified Functions," *IEEE Trans. on Circuits and Systems - I*, **53**, 2444-2450, 2006.

[11] Wang R., Chen L.: Synchronizing Genetic Oscillators by Signalling Molecules, *Journal of Biological Rhythms*, **20**, 257-269, 2005.

[12] Zhou T., Chen L., Wang R.: A Mechanism of Synchronization in interacting Multi-Cell Genetic Systems, *Physica D*, **211**, 107-127, 2005.

[13] Li C., Chen L., Aihara K.: Synchronization of Coupled Nonidentical Genetic Oscillators, *Physical Biology*, **3**, 37-44, 2006.

[14] Li C., Chen L., Aihara K.: Stochastic Synchronization of Genetic Oscillator Networks, *BMC Systems Biology*, doi:10.1186/1752-0509-1-6, 2007.

[15] Li C., Chen L., Aihara K.: Transient Resetting: A Novel Mechanism for Biological Synchrony, *PLoS Computational Biology*, DOI: 10.1371/journal.pcbi.0020103.eor, 2006.

[16] Wang R., Chen L., Aihara K.: Synchronizing a Multicellular System by External Input: An Artificial Control Strategy, *Bioinformatics*, **22**, 1775 - 1781, 2006.

[17] Chen L., Wang R., Zhou T., Aihara, K.: Noise-induced Cooperative Behavior in a Multi-Cell System, *Bioinformatics*, **21**, 2722-2729, 2005.

[18] Zhou T., Chen L., Aihara K.: Molecular Communication through Stochastic Synchronization Induced by Extracellular Fluctuations, *Phys. Rev. Lett.*, **95**, 178103, 2005.

[19] Springer M., Paulsson J.: Harmonies from Noise, *Nature*, **439**, 27-28, 2006.

[20] Wang Y., Joshi T., Xu D., Zhang X.-S. Chen L.: Inferring Gene Regulatory Networks from Multiple Microarray Datasets, *Bioinformatics*, **22**, 2413 - 2420, 2006.

[21] Chen L., Wu L.-Y., Wang Y., Zhang X.-S.: Inferring Protein Interactions from Experimental Data by Association Probabilistic Method, *Proteins*, **62**, 833-837, 2006.

[22] Li Z., Zhang S., Zhang X.-S., Chen L.: Alignment of Protein Interaction Networks by Integer Quadratic Programming. *IEEE EMBC06, SaD03.4, 5527-5530, 2006.*

[23] Li Z., Zhou W., Zhang X.-S., Chen L.: A Parsimonious Tree-Grow Method for Haplotype Inference, *Bioinformatics*, **21**, 3475-3481, 2005.

[24] Zhang X.-S., Wang R.-S., Wu L.-Y., Chen L.: Models and Algorithms for Haplotyping Problem, *Current Bioinformatics*, **1**, 105-114, 2006.

[25] Li Z., Zhang X.-S. Chen L.: Unique Optimal Foldings of Protein on a Triangular Lattice, *Applied Bioinformatics*, **4**, 105-116, 2005.

[26] Zhang X.,Wang Y., Zhan Z., Wu L., Chen L.: Exploring Protein Optimal HP Configurations by Self-Organizing Mapping, *J. Bioinformatics and Comput. Biol.*, **3**, 385-400, 2005.

[27] Chen L., Zhou T., Tang Y.: Protein Structure Alignment by Deterministic Annealing, *Bioinformatics*, **21**, 51-62, 2005.

[28] Chen L., Wu L.-Y., Wang Y., Zhang S., Zhang X.-S.: Revealing Divergent Evolution, Identifying Circular Permutations and Detecting Active-Sites by Protein Structure Comparison, *BMC Structural Biology*, doi:10.1186/1472-6807-6-18, 2006.

[29] Zhou T., Chen L., Tang T., Zhang X.-S.: Aligning Multiple Protein Structures by Deterministic Annealing. *J. of Bioinformatics and Computational Biology*, **3**, 837-860, 2005.

[30] Song Z., Chen L., Ganapathy A., Wan X.-F., Brechenmacher L., Tao N., Emerich D., Stacey T., Xu D.: Development and Assessment of Scoring Functions for Protein Identification Using Peptide Mass Fingerprinting Data. *ELECTROPHORESIS*, **28**, 864-870, 2007.

[31] Li C., Aihara K., Chen L.: A Systems Biology Perspective on Signal Processing in Genetic Network Motifs. *IEEE Signal Processing Magazine*, 24, 136-147, March 2007.

[32] Wu Z., Wang Y., Feng E., Chen L.: A New Geometric-Topological Method to Measure Protein Fold Similarity. *Chemical Physics Letters*, **433**,432-438, 2007.

[33] Wu Z., Feng E., Wang Y., Chen L.: Discrimination of Outer Membrane Proteins by a New Measure of Information Discrepancy. *Protein & Peptide Letters*, **14**, 37-44, 2007.

[34] Charusanti P., Hu X., Chen L., Neuhauser D., Distefano J.: A Mathematical Model of BCR-ABL Autophosphorylation, Signaling through the CRKL Pathway, and Gleevec Dynamics in Chronic Myeloid Leukemia, *Discrete and Continuous Dynamical Systems - Series B*, 4, 99-114,2004.

[35] Zhao X., Wang Y., Chen L., Aihara K.: Protein Domain Annotation with Predicted Domain-Domain Interaction Networks, *Protein & peptide letters*, in press, 2007.

[36] Liu Z.-P., Wu L.-Y., Wang Y., Zhang X.-S., Chen L.: Analysis of Protein Surface Patterns by Pocket Similarity Network. *Protein & Peptide Letters*, in press, 2007.

[37] Zhang S., Jin G., Zhang X.-S., Chen L.: Discovering Functions and Revealing Mechanisms at Molecular Level from Biological Networks, *Proteomics*, in press, 2007.

[38] Li Z., Zhang S., Wang Y., Zhang X.-S., Chen L.: Alignment of Molecular Networks by Integer Quadratic Programming. *Bioinformatics*, doi:10.1093/bioinformatics/btm156, 2007.