The International Symposium on Optimization and Systems Biology August 8-10, 2007, Beijing, China

Welcome to

OSB 2007

Local Organizer



Organizers





Sponsors





Computational Systems Biology on Networks and Dynamics



Luonan Chen (陈洛南)

Institute of Systems Biology, Shanghai University Department of EEE, Osaka Sangyo University Institute of Industrial Science, University of Tokyo

At molecular level

From computational viewpoint

Systems Biology?

Instead of analyzing individual components or aspects of organism,

systems biology is to study an organism, viewed as a **dynamical** or **interacting network** of genes, proteins and biochemical reactions,

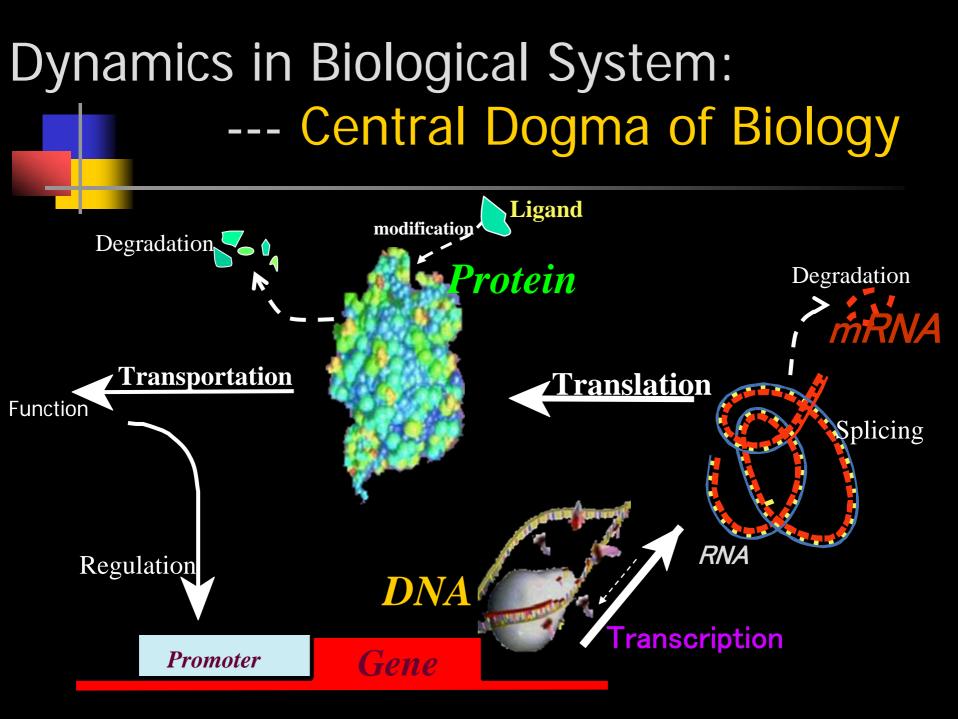
which eventually give rise to life.

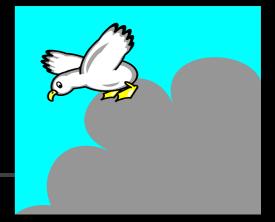


Keywords

- System (component->system)
- Integration
- Network
- Dynamics (static->dynamic)
- Interaction

from theoretical and engineering perspectives





Overview

- Motivation
- Areas
- Perspective

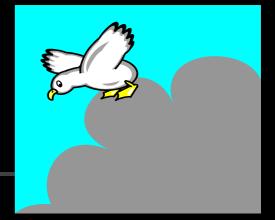
Pursue frontier works from theoretical and engineering perspectives



Motivation

Motive: Burgeoning high-throughput data are driving the integrative study from describing phenomena to understanding design principle, mechanism, from studying components to understanding functional network for biological systems, entire cell, organ, and even organisms.

- New Frontier: systems biology
- Area: An emerging multi-interdisciplinary field



Overview

Motivation

Areas

Perspective



Areas of Research

Three areas of research

- 1. Molecular systems biology (Dynamics, Network)
- 2. Molecular biology

(Synthetic Bio, Experiments)

 3. Computational systems biology (Static, Tool, Database)

interdisciplinary

1. Molecular Systems Biology

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.

Instead of analyzing individual components or aspects of organism, systems biology is to study an organism, viewed as a dynamical or interacting network of genes, proteins and biochemical reactions which give rise to life.



Keywords

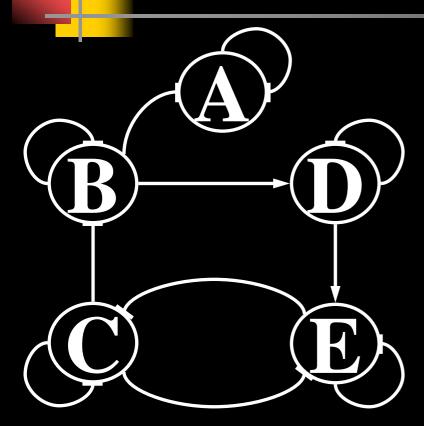
Dynamics
Network
Integration
Engineering
Optimization

The main topics on Dynamics and Networks

- Nonlinear dynamical models and nonlinear analysis at molecular level.
- Quantitative simulation of cellular dynamics.
- Molecular communication
- Inferring gene regulatory network, and signal pathways of biological systems
- Finding motifs and conserved substructures of protein interaction networks, GRN

Physical Biology '06, Physical Review E '04, IEEE Trans. on CAS '02

Modeling gene regulatory network



Gene Network Biochemical Reaction Network Mass Action Law

Difficulties

Nonlinearity, noises and delays

Models

Monotone Dynamical Systems Lur'e systems

- Nodes are mRNAs, proteins
- Arrows represent interactions
- Monotone dynamical system

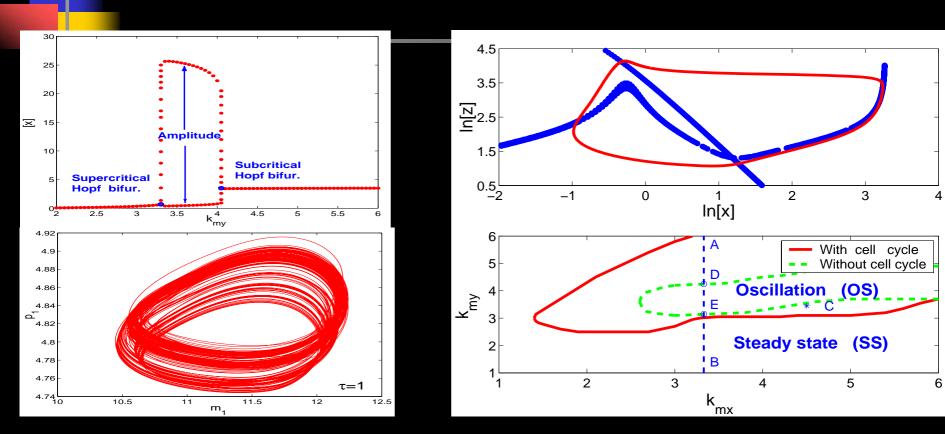
reduce complexity without significant approximation

Network Theory

PRE '05, Bioinformatics '06, Journal of Biology Rhythm '05

Nonlinear Theory

Nonlinear Analysis



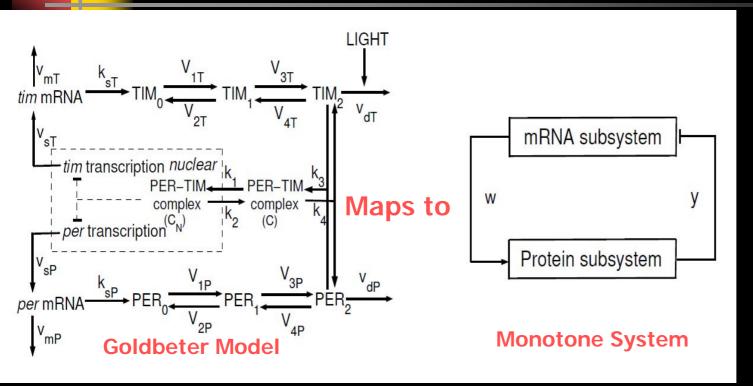
Effects of Cell Cycle on Cellular Dynamics (degradation factor)

Apply nonlinear theory, e.g. stability and bifurcation analysis to identify essential mechanism of cell cycle, chemotaxis, quorum sensing (gain insights)

Journal of Theoretical Biology '05, Journal of Biology Rhythm '05 Con

Control Theory

Reducing complexity by control theory



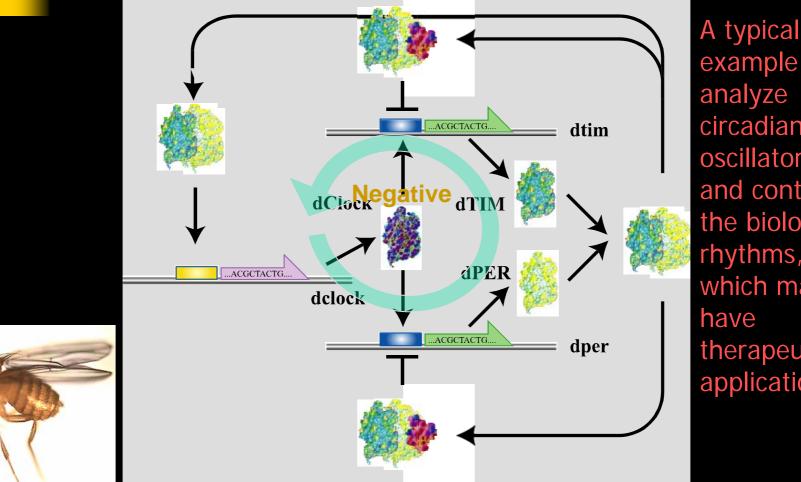
By applying a wealth of control theory and exploiting the special structures, many biological systems can be reduced to simple forms

(a) Scheme of the model for circadian oscillations in *Drosophila* involving negative regulation of gene expression by PER and TIM. (b) Its feedback closure form with inputs and outputs, by which the regulation mechanism can be understood and the system can be significantly reduced.

Jet lag

A schematic illustration

Analysis of Circadian Rhythm



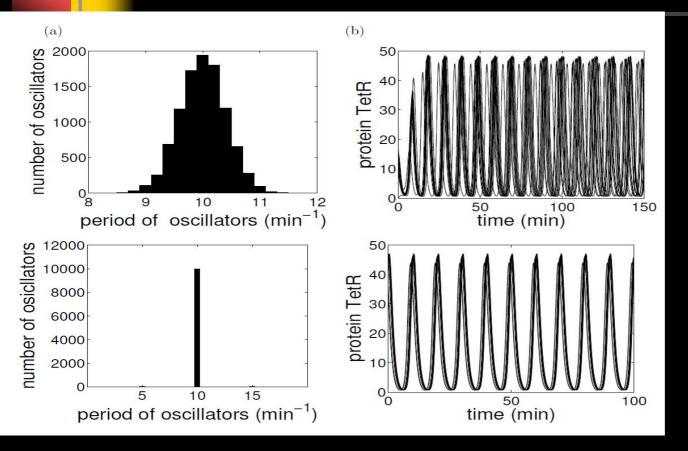
example is to analyze circadian oscillators, and control the biological rhythms, which may therapeutic application.

Drosophila

Natural oscillator

Bioinformatics '06

Control cellular dynamics by impulse inputs (Synchronization)



Control strategy is an effective way to change the cellular dynamics, for instance, synchronizing biooscillators by impulse control.

Entrainment of the 10,000 coupled repressilators by periodically injecting coupling AI into the common extracellular medium. (a) Asynchronous oscillations of protein TetR for 10 randomly chosen oscillators in the absence of injection. (b) Synchronization induced by the periodic injection

Cellular communication: essential for all living organisms at molecular level

For multicellular organisms

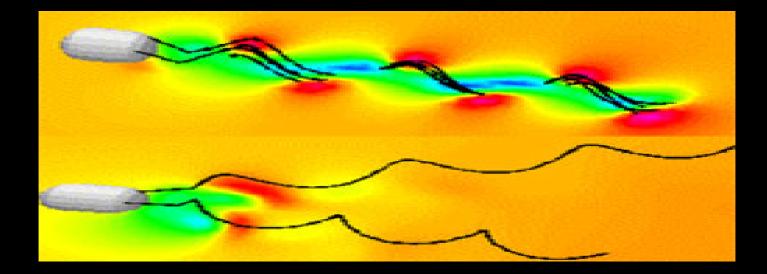
 complex pattern structures are created from identical and unreliable components via communication



Communication is accomplished by first transmitting individual cell information via signal molecules to neighboring cells, then exchanging information among these signal molecules and further generating a global cellular response at the level of tissues, organs and bodies

collective behavior, cooperative behavior, synchroniza tion

For bacteria - various social behaviors are displayed due to intercellular communication



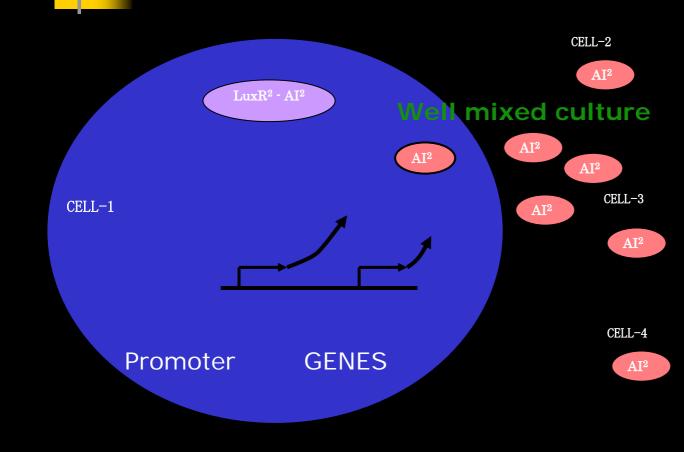
E.coli

Signal transduction in Chemotaxis: communication strategy

Physical Review Letter '05, Bioinformatics '05

Stochastic Process

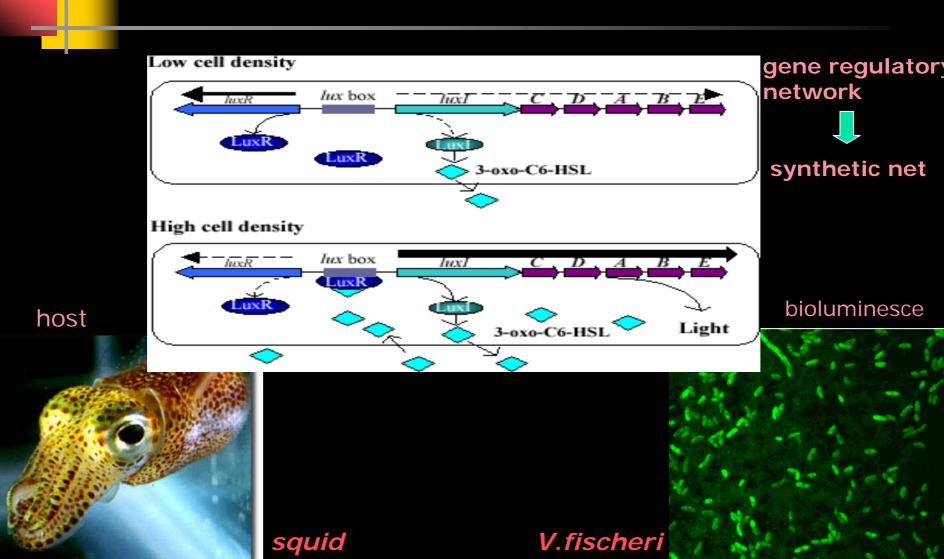
Cell communication is accomplished by diffusing Als to environment, which enter cells as signal molecules to regulate gene expression.



Such circuit is engineered on plasmids and assumed to grow in E.coli, based on quorum sensing.

The ability to 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 molecular communication is an essential topic with systems biology, which requires both mathematical and biological knowledge.

Quorum sensing : coupling between environment and cells A bacterium is effectively communicating, to detect and respond to signals (autoinducer: AI) produced by the surrounding bacteria



Commentary to PRL by NATURE

NATURE Vol 439 5 January 2006

NEWS & VIEWS

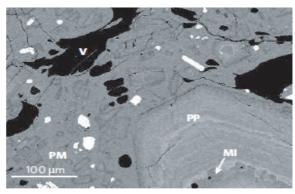


Figure 1 | Photomicrograph of a plagioclase phenocryst with a melt inclusion. This backscattered electron image shows the microtexture of a Mount St Helens lava-dome sample erupted in September 1981. PM, plagioclase microlite; PP, plagioclase phenocryst; MI, melt inclusion; V, vesicle (relict vapour bubble).

plagioclase phenocrysts form in a magma reservoir but then grow considerably during an eruption. In their new paper¹, they support and expand this idea with analyses of the melt inclusions found in phenocrysts (Fig. 1).

Melt inclusions are intriguing aberrations of 'normal' crystal growth. They preserve information about the history of the phenocryst, but are difficult to interpret because the conditions in which they form are unknown, the degree of chemical communication with the outside melt may vary, and post-entrapment processes can modify the original compositions. To address these problems, Blundy and Cashman¹ analysed more than 100 inclusions from six eruptions of Mount St Helens during 1980 that varied in intensity from quiescent lava effusion to sustained explosive activity.

They interpret a large range in observed

conditions, and the remainder as the result of the eruption itself in which partial degassing of the melt happened too quickly for crystallization to occur. There is an alternative interpretation, however, that is consistent with the traditional understanding of plagioclase phenocrysts. Variable H_2O may arise from fluctuations in the CO₂ and H_2O content of magma within the reservoir. Such fluctuations would corroborate the view of the reservoir as an open system subject to periodic influxes of new magma¹⁰.

Most importantly, Blundy and Cashman's interpretation of the melt-inclusion data reinforces the idea that crystallization during an eruption affects the style of intermediateintensity eruptions^{8,11}. The correlation between crystallinity and melt-inclusion H₂O content raises an intriguing chicken-and-egg issue, however. Does degassing-induced crystallization occur only when ascent rates and flow regimes in the magma conduit produce conditions favourable for rapid crystal growth? Or through its influence on magma viscosity, can the solidification process reduce the intensity of an eruption already in progress? Events at Mount St Helens in 1980 have inspired many studies, the latest being this report¹ detailing changes in melt chemistry at unprecedented spatial and temporal resolution. With the volcano again obliging investigators with new magma since early 2004, a renewed effusion of research is sure to follow. Julia E. Hammer is in the Department of Geology and Geophysics, University of Hawaii, 1680 East-West Road, Honolulu, Hawaii 96822, USA. e-mail: jhammer@soest.hawaii.edu

- Blundy, J. & Cashman, K. Geology 33, 793-796 (2005).
- Swanson, S. E., Naney, M. T., Westrich, H. R. & Eichelberger, J. C. Bull. Volcanol. 51, 161–176 (1989).
- Cashman, K. V. Contrib. Mineral. Petrol. 109, 431–449 (1992).
- Swanson, S. E. Am. Mineral. 62, 966–978 (1977).
- Hammer, J. E. & Rutherford, M. J. J. Geophys. Res. B107, 8-1–8-24 (2002).
- Couch, S. Am. Mineral. 88, 1471–1485 (2003).
- Geschwind, C. & Rutherford, M. J. Bull. Volcanol. 57, 356–370 (1995).
- Hammer, J. E., Cashman, K. V., Hoblitt, R. P. & Newman, S. Bull. Volcanol. 60, 355–380 (1999).
- Blundy, J. & Cashman, K. Contrib. Mineral. Petrol. 140, 631-650 (2001).
- Gardner, J. E., Carey, S., Rutherford, M. J. & Sigurdsson, H. Contrib. Mineral. Petrol. 119, 224–238 (1995).
- Gardner, C. A., Cashman, K. V. & Neal, C. A. Bull. Volcanol. 59, 537-555 (1998).

Harmonies from noise

Michael Springer and Johan Paulsson

Do random environments make for random responses to them? Mathematical models suggest that this is not always the case — adding noise could create synchronous oscillations in cell-cell signalling systems. PLoS Computational Biology '06

Transient Resetting: A Novel Mechanism for Synchrony

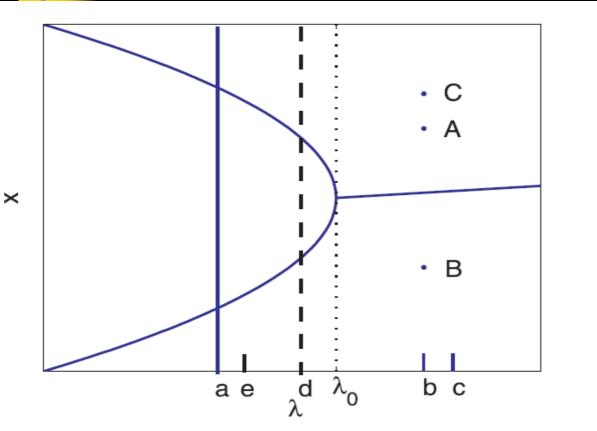


Figure 1. A Schematic Bifurcation Diagram with a Normal (Supercritical) Hopf Bifurcation for Illustrating the Transient Resetting Mechanism DOI: 10.1371/journal.pcbi.0020103.g001

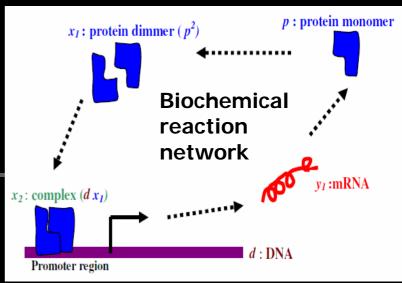
For uncoupled biological oscillators with Stimuli, this mechanism can unify and extend many existing results on (deterministic and stochastic) stimulus induced synchrony. Transient resetting is a possible mechanisn for synchronization i many biological organisms, which might also be further used in the medical therapy of rhythmic disorders.

Independent Oscillators

Bioinformatics'05

Quantitative Simulation in a Cell

Stochastic Simulation



Simulation for such network requires enormous computational power.

- Biochemical master equation
- Monte Carlo simulation for cellular dynamics
- Fokker-Planck equation
- Langevin equation, SDE
- Cumulant equation

ODE

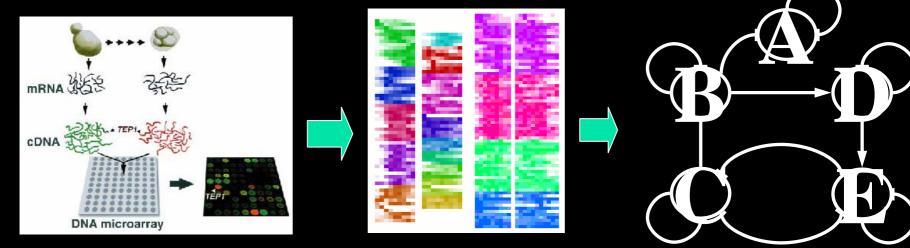
In contrast to the qualitative analysis, quantitative computation provides the detail information. Biological system can be viewed as a biochemical reaction network, which generally can be defined as master equation with both stochastic fluctuations and discrete changes.

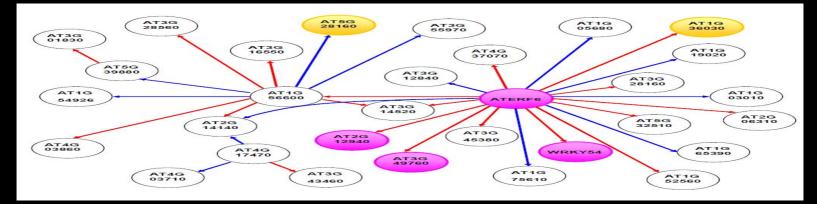


Reverse Engineering

Inferring Gene Networks

Microarray technologies have produced tremendous amounts of gene expression data

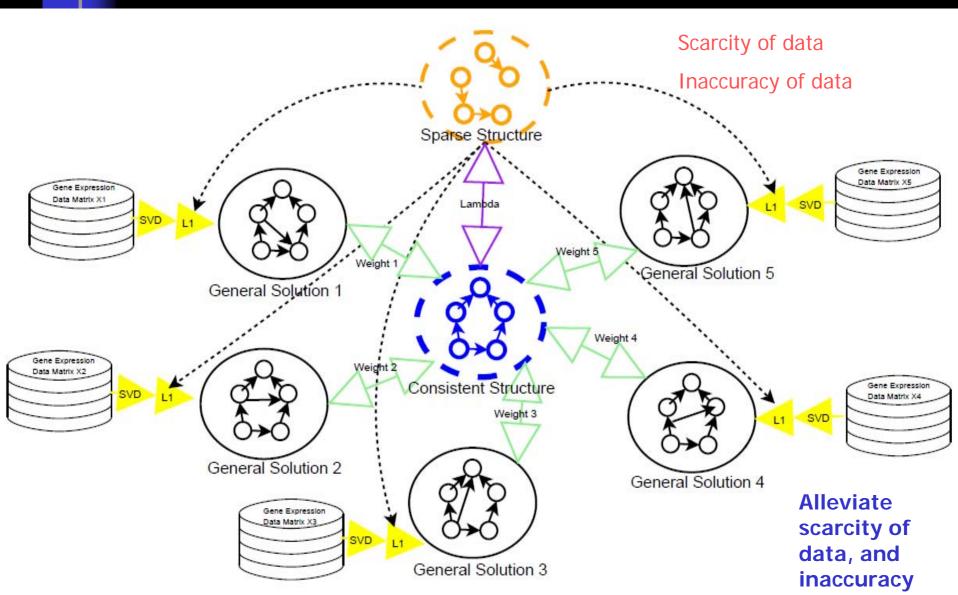




Inferred network in Arabidopsis thaliana, from stress response datasets in shoots.

Software, GRNInfer, is available on line !

Inferring Gene Network by Multiple Datasets



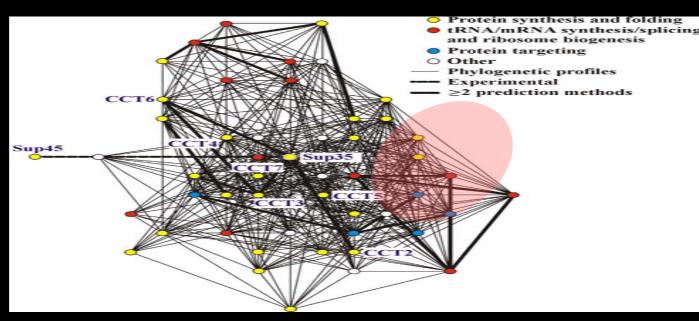
Proteins '05

Software is available APM !

Inferring Protein Network from experiment data

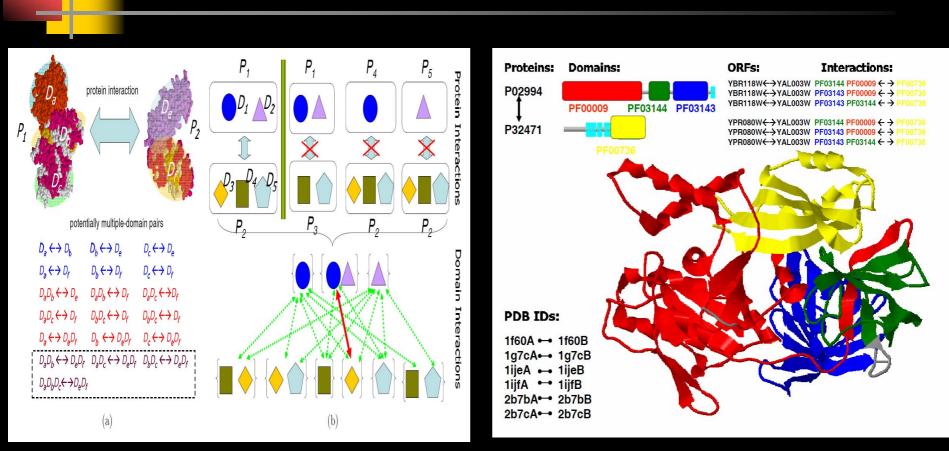
To elucidate protein interaction networks is one of major goals of functional genomics for whole organisms.

Due to recent advances of biotechnology, data of protein-protein interaction can be generated by experiment, such as two-hybrid assay, coimmunoprecipitation and chIP-chip approach.



Sup 35 Interaction Network

BMC Bioinformatics '07



Multi-domain interactions

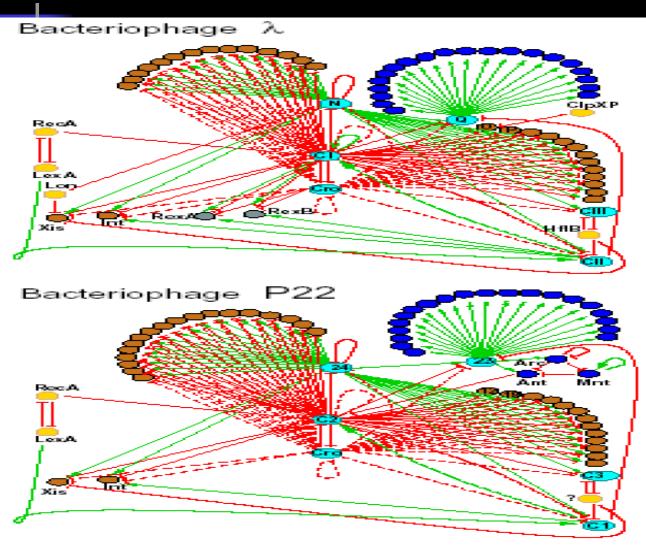
Super-domains, strongly cooperative domains

Domain cooperation plays an important role in facilitating the protein-protein interactions

Bioinformatics '07, Proteomics'07

Software, MNAligner, is available!

Network motif, signal pathway, evolution



Various organisms differ not only because of differences of constituting proteins, but also because of their networks. Hence, it is essential to address similarities and differences in networks by comparative network analysis

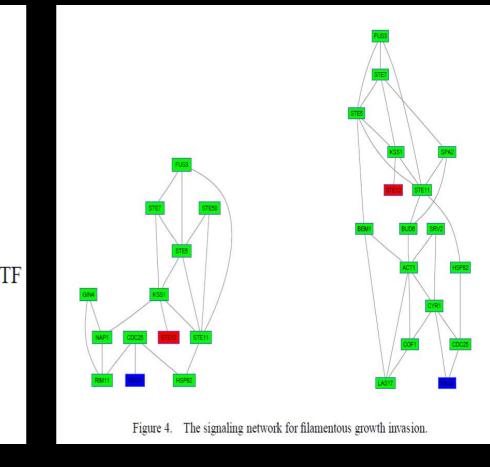
Alignment of gene and protein networks

Scale-free, small-world

APBC'07

Identify signaling network from PPI and Microarray data

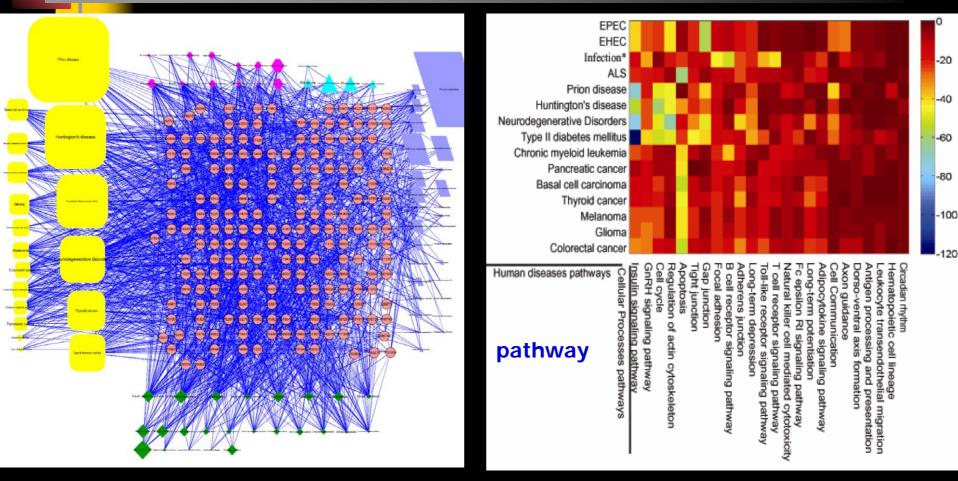
$$\begin{array}{ll} \text{Min} & \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} a_{ij} e_{ij} + \lambda \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} e_{ij} \\ \text{s.t.} & e_{ij} \leq x_i \\ & e_{ij} \leq x_j \\ & \sum_{j} e_{ij} \geq 1, \text{ if } i \text{ is a membrane protein or TF} \\ & \sum_{j} e_{ij} \geq 2x_i, \text{ if } i \text{ is not a membrane protein or TF} \\ & x_i = 1, \text{ if } i \text{ is a membrane protein or TF} \\ & x_i \in \{0, 1\}, \ i = 1, 2, \cdots, |V| \\ & e_{ij} \in \{0, 1\}, \ i, j = 1, 2, \cdots, |V| \end{array}$$



By mathematical programming, we can identify the signal pathway or network

Hubs and network motifs: diseases and their pathogenesis

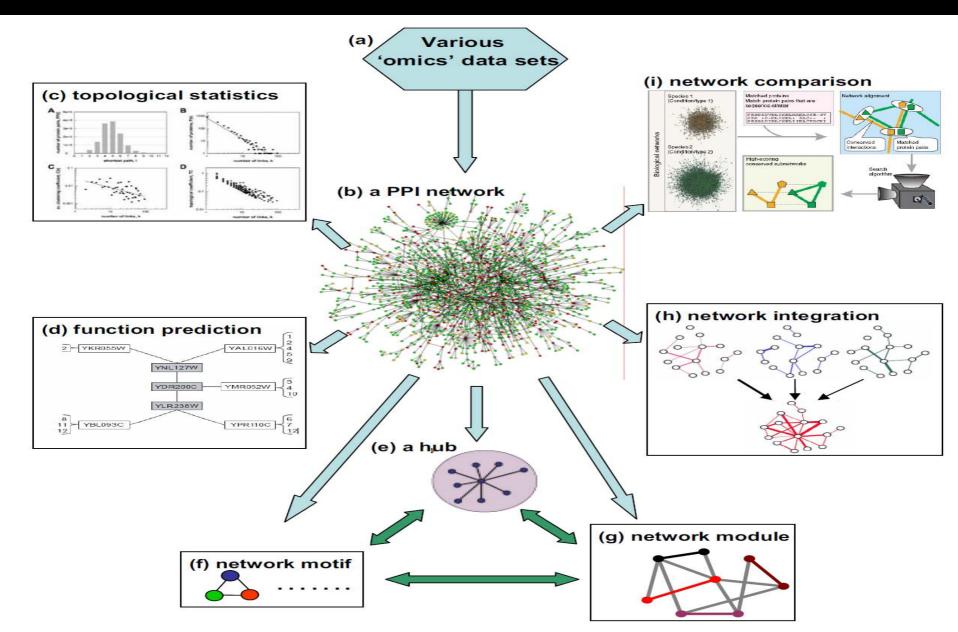
Connectors: derived from a network with motifs and hubs



One mechanism of diseases may be attributed to the connectors that activate or initiate interactions between disease pathways and causation pathways, thereby eventually leading to diseases.

Proteomics'07

Research for Molecular Networks



2. Molecular Biology (experiments, synthetic biology)

 Synthetic biology is a new area of research that combines science and engineering in order to design and build novel biological functions and systems.

Aim: Research is aimed to combine knowledge from various disciplines including molecular biology, engineering and mathematics to design and implement new cellular behaviors.
Goal: The goal is both to improve the quantitative understanding of natural phenomenon as well as to foster an engineering discipline for obtaining new complex cell behaviors in a predictable and reliable fashion.

Main Topics on Engineering

- Gene switching network and experiment
 Gene oscillating network and experiment
 Biosensor network and related experiment
- Study on differentiation of stem cells
- Drug target screen and identification

based on synthetic biology

Approach by Synthetic Biology Forward engineering

In contrast to the reverse engineering, the synthetic gene network is a forward engineering

Analyze and understand essential structure mechanism with simple environment

Designing Gene Switch Journal of Theoretical Biology '04 (toggle switch in *E.coli*)

- 1. A state
- 2. Switching

 Experiment data well agreed with theoretical results, which implies that mathematical model is a powerful tool for designing network

Ex.: **A gene toggle switch**

 \checkmark A gene oscillator

Applications

- Medicine (gene therapy)
- Biotech
- Bio-computer

Mathematical model can be used to design an artificial gene network which has desirable functions.

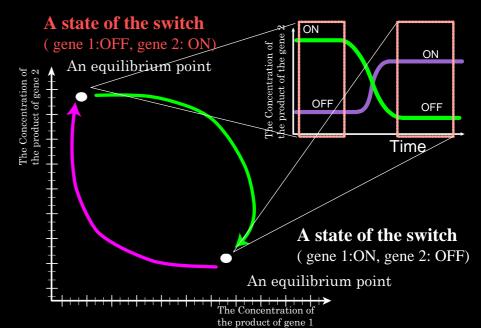


An artificial gene network which has multistable states !

A mathematical
 interpretation

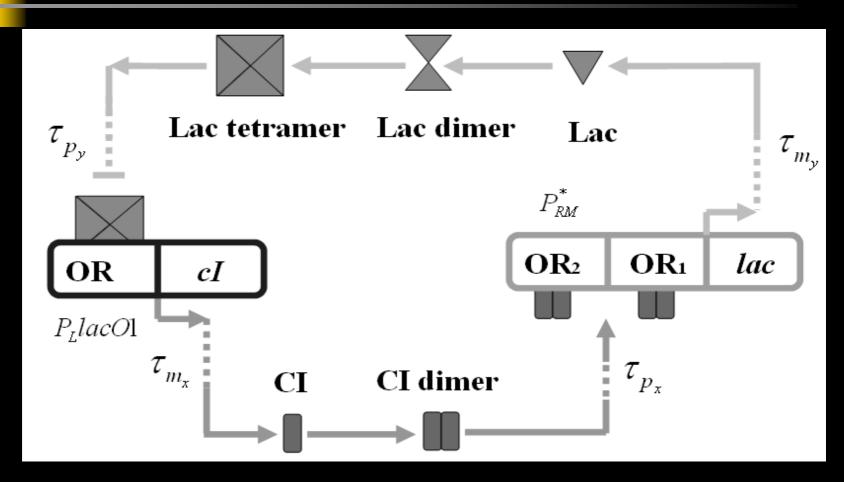
1. A stable equilibrium point.

2. Transition between two stable equilibrium points



Systems Biology '04

Designing Gene Network (Oscillating Network)

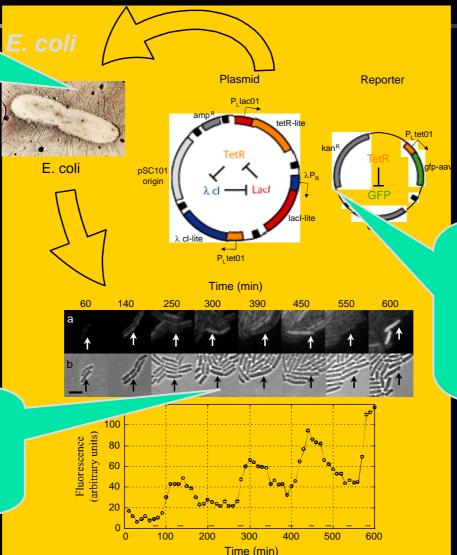


There exist various periodic oscillations with different time scales ranging from less than a second to more than a year, which may allow for living organisms to adapt their behaviors to a periodically varying environment

Repressilator

transfect cloned gene network into a host

By GFP, measure the dynamics of gene expression

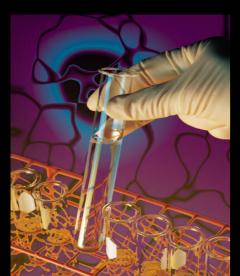


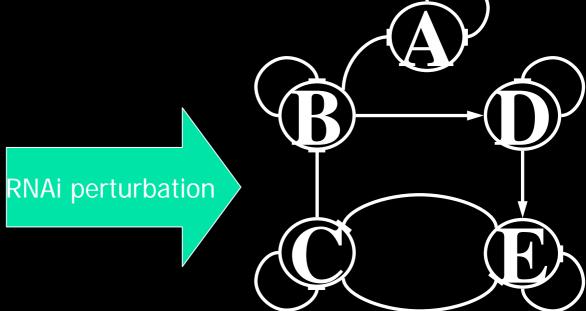
A typical oscillating network is repressilator, which is constructed on plasmids, and transformed into simple organisms, E.coli.

Construct artificial gene network on plasmids (vectors)

Refining network by RNAi

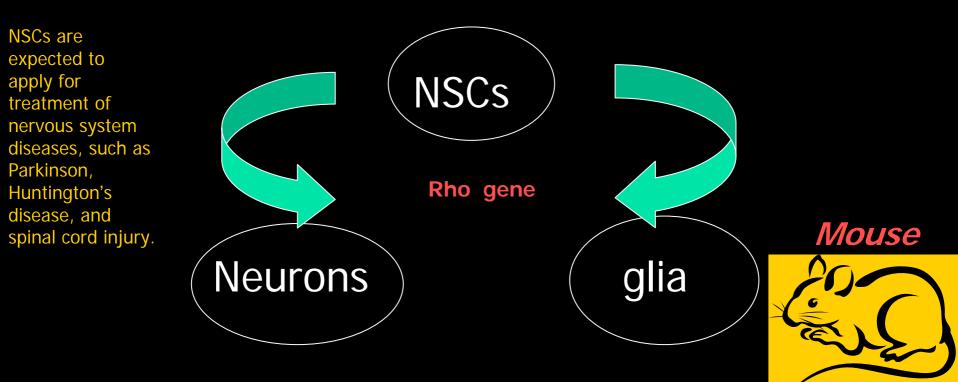
RNAi silence experiments are also conducted in several key or pivotal genes (with dense connections to others) to improve the confidence.





Neural Stem Cells (NSC)

Differentiation : subtype of progenitor cells in the nervous system



3. Computational Systems Biology (static problem, tools, database)

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

--Research in computational systems biology often overlaps with bioinformatics.



Main Topics on Optimization

- Protein structure analysis and function prediction
- Haplotype reconstruction
- Identifying protein by Mass Spectrometry
- Inferring protein-protein interaction network by experimental data
- Neuron image processing (Neuroinformatics)

Software PTG is available on line !

Bioinformatics'05, Current Bioinformatics'06

Disease genes mapping Drug design

SNP & Haplotype

NP-hard class

<u>SNP</u>: Single Nucleotide Polymorphism
 <u>Haplotype</u>: A set of closely linked genetic markers present on one chromosome which tend to be inherited together (not easily separable by recombination).



Set of SNP polymorphisms: a SNP haplotype

Tree-grow method

Parsimony principle, phylogenetic tree, frequency

Software, SAMO, is available on line !

BMC Structural Biology '06

Protein Structure Alignment

Optimally superimpose two structures



 2. Further find the regions of closest overlap in a three-dimension space.

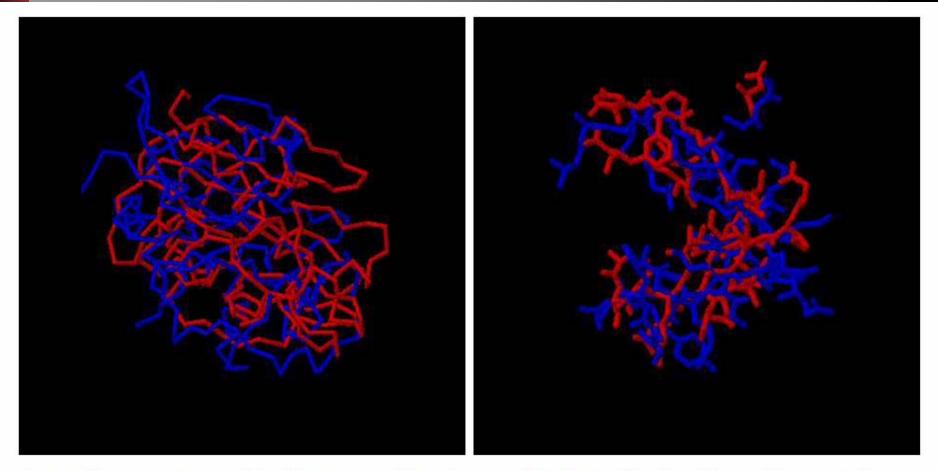


*Combinatorial optimization problem : NP-hard

*Pairwise sequence alignment: polynomial time algorithm (DP)

Structure alignment plays a key role in protein structure prediction, fold family classification, motif finding, phylogenetic tree reconstruction

Local Alignment to Identify Active Site



(a) Alignment result of 1tpo and 2act.

(b) Detail of active site match.

Structure- Function Relationships

 Can we predict the function of protein molecules from their sequence?

sequence > structure > function

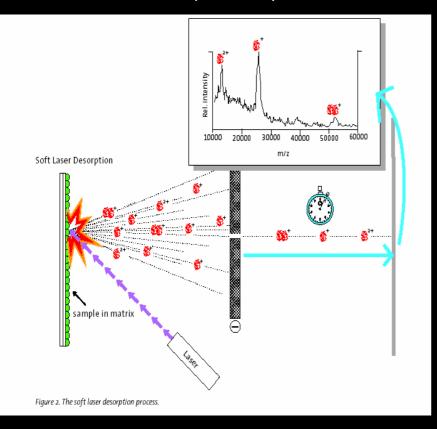
- Conserved functional domains = motifs
- Prediction of some simple 3-D structures (α-helix, β-sheet, membrane spanning, etc.)
 - Secondary structure: 80% accuracy

ELECTROPHORESIS '06

Software is available on line!

Mass Spectrometry to identify protein

Matrix-Assisted Laser Desorption/Ionization (MALDI)



itput						
Rank	Protein ID	Score	Hit Peptides	Coverage%	MW	Length
	P04064	56.782813	11	0.343750	68308.000000	640
	Q07262	33.069087	8	0.145455	55290.000000	491
	Q28653	25.133507	9	0.116883	68090.000000	586
	Q14738	23.449877	10	0.120482	69991.000000	602
	Q8MIS5	21.075731	5	0.217391	28867.000000	256
	094823	19.024397	11	0.073826	165390.000000	1461
	P56932 Q97EY6	17.160975 15.439052	5	0.090909	51982.000000 23505.000000	453 205
	Q97ET0	10.439002	4	0.125000	23303.000000	205
otein View						
â	15	30 44		69 70 80	96	101 106
G <mark>FENEK</mark> NSSILSK	LATNEELGENSPYFDGW	K AYDNDPFHPLK NPNG	WIQMGLAENQLCFDLIEEWI	K R NPNASICTTEGIK S	FR AIANFQDYHGLPEF	R SAIAK FMEK
	22		199 162 164	185 199	196 200 202	
r gor vtfdper '	A/MAGGATGANETIIFCLAD	TGDAFLVPSPYYPAFNR	DLR WR TGVQLIPIPCDSS	NNFQITTK AVR EAYENA	IQK SNIK VK GLILTN	PSNPLGTTLDR
				7 300 :		
220		208				
220			29 28 VHIVYSLSK DMGLPGFR V			EMLSDER
223 FLK NLLTFTNQHM 338 3	IIHLVCDEIYAATVFNTPQF\ 9 342 343 345	VSIAEILDDETSHCNK DL 357 360	VHIVYSLSK DMGLPGFR V 378 38	GIVYSFNDAWNCAR KI	MSSFGLVSTQTQYLLA 413	430
223 FLK NLLTFTNQHM 338 3	IIHLVCDEIYAATVFNTPQF\ 9 342 343 345	VSIAEILDDETSHCNK DL 357 360	VHIVYSLSK DMGLPGFR V	GIVYSFNDAWNCAR KI	MSSFGLVSTQTQYLLA 413	430
220 FLK NLLTFTNQHM 309 30 /SNFLTESSK R L 439 439	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 FLK NLLTFTNQHM 309 30 /SNFLTESSK R L 439 439	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	.VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 FLK NLLTFTNQHM 309 30 /SNFLTESSK R L 439 439	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 FLK NLLTFTNQHM VSNFLTESSK R L 438-439	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 FLK NLLTFTNQHI 308 37 /SNFLTESSK R L 438 439 PPERK / DODES	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 FLK NLLTFTNQHI VSNFLTESSK R L 408 49 Peak	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
221 FLK NLLTFTNQHI 200 30 30 25NFLTESSK R L 200 40 1Peak 000 4 Int	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 FLK NLLTFTNQHI VSNFLTESSK R L 408 49 Peak	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
221 FLK NLLTFTNQHI 200 30 30 25NFLTESSK R L 200 40 1Peak 000 4 Int	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 TLK NLLTFTNQH 398 3 SNFLTESSK R 498 49 Peak 000 Å Int 500 000	IIHLVCDEIYAATVFNTPQF\ ୨୦୪୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 FLK NLLTFTNQHM 308 32 (SNFLTESSK R L 304 59 19 60 19 60 10 6 10 7 10 7 10 10 7 10 7 1	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 TLK NLLTFTNQH 398 3 SNFLTESSK R 498 49 Peak 000 Å Int 500 000	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
22 TLK NLLTFTNOHH SNFLTESSK R L 90 40 90 40 90 90 90 90 90 90 90 90 90 9	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 TLK NLLTFTNQHI 308 32 VSNFLTESSK R L 304 42 1Peak 000 Å Int 500 - 500 - 500 -	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
22 LK NLTFTNOHH SNFLTESSK R L 90 40 90 40 90 90 40 90 40 90 90 90 90 90 90 90 90 90 9	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 TLX NLTFINGH '30 ST 2005 K R L '90 ST 2005 K R L '90 A 2005 K R L R '9	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 19	MSSFGLVSTQTQYLLA 413	430
220 TLK NLTFINGHH 303 37 SINFLTESSK R L 4000 Å Int 5000 Å 5000 Å 5000 Å 5000 Å	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 31	MSSFGLVSTQTQYLLA 413	430
220 TLX NLTFINGH '30 ST 2005 K R L '90 ST 2005 K R L '90 A 2005 K R L R '9	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 31	MSSFGLVSTQTQYLLA 413	430
220 TLX NLTFINHH 328 32 (SNFLTESSK R L 42 40 40 (Peak 1000 Å 101 1000 Å 101 101 101 101 101 101 101 101 101 10	IIHLVCDEIYAATVFNTPQF\ ୨୦୧୦୦୦୦୦୦ AKRHKHFTNGLEEVGI ଶାୟସ କରେସେ	VSIAEILDDETSHONK DL 357 300 IK CLR <mark>SNAGLFOWMDLF</mark> 461 468 467	VHIVYSLSK DMGLPGFR V 378 38 RPLLK ESTFDSEMSLWR VI 489 4	GIVYSFNDAWNCAR K 3 19 395 INDVK <mark>LNVSPGSSFDCQ</mark> 31	MSSFGLVSTQTQYLLA 413	430

ProteinDecision

Protein identification through mass spectrum data is an important domain in proteomics

Proteins'07

Function prediction by Machine Learning

Table 1: Summary of data descriptions										
Data sets	# Samples	Positive samples		Negative samples		Classification				
		Min	Max	Min	Max	tasks				
PCB00019	1357	10	168	592	670	55				
PCB00020	11944	10	771	566	587	246				
PCB00022	11944	10	1013	555	587	191				
PCB00028	11373	10	1301	503	573	199				

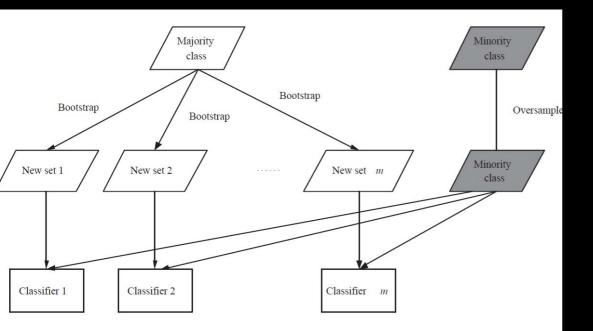


Figure 2: Schematic for EnClassifier

Imbalance data: a new algorithm to overcome the imbalanced problem with a new sampling technique and a committee of classifiers. Then, classifiers trained in different feature spaces are combined together to further improve the accuracy of protein function prediction

Website for Papers and Software

<u>http://www.isb.shu.edu.cn</u>
 <u>http://intelligent.eic.osaka-sandai.ac.jp</u>
 <u>http://zhangroup.aporc.org/</u>

List of Major Publications

[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] C.Li, L.Chen, K.Aihara, "Stability of Genetic Networks with SUM Regulatory Logic," *IEEE Trans. on Circuits and Systems - 1*, 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 - 1*, **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:10.1186/1752-0509-1-02007.
- [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*, 2, 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, 20
 [18] Zhou, T., Chen, L., Aihara, K. Molecular Communication through Stochastic Synchronization Induced by Extracellular Fluctuations, *Phy 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, *Bioinformatic* 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, WANG Y, ZHANG X-S, CHEN L. Alignment of molecular networks by integer quadratic programming[J]. *Bioinformatics*, in press, 2007.
- [Z3] 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, 2000
- [25] Li, Z., Zhang, X-S., Chen, L. Unique Optimal Foldings of Protein on a Triangular Lattice, Journal of 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. Bioinformatic 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*, March, 2007. In press.
- [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] P.Charusanti, X.Hu, L.Chen, D.Neuhauser, J.Distefano, 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] ZHANG S, JIN G, ZHANG X-S, CHEN L. Discovering functions and revealing mechanisms at molecular level from biological networks[J]. *Proteomics*, in press, 2007.

Acknowledgment

- Chinese Academy of Sciences: Prof. X-S.Zhang, Drs.L-Y.Wu, Y.Wang, Z-P.Li, etc.
- University of Tokyo: Prof. K.Aihara, Dr. X-M.Zhao, Dr. T.Kobayashi, etc.
- Shanghai University: Dr. R-Q. Wang, Prof. Z.Liu, etc.
- Osaka Sangyo University: Dr. R-S.Wang

Upcoming Books

Optimization !

 L.Chen, R.Wang, X.Zhang, *Biomolecular Network*: Computational Methods and Applications in Bioinformatics and Systems Biology, *John Wiley & Sons*, 2008

Dynamics !

 L.Chen, R.Wang, C.Li, K.Aihara, *Modelling Biomolecular Networks*: Structures and Dynamics.
 Springer-Verlag, London, 2008.

■ 陳、呉、王、章、系統生物学導論、*科学出版社*、2008

Thank you





Enjoy rich cultural heritage of Beijing, and further China during your stay !