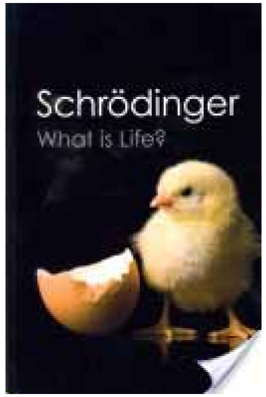
Genetic, Protein, Metabolic Networks

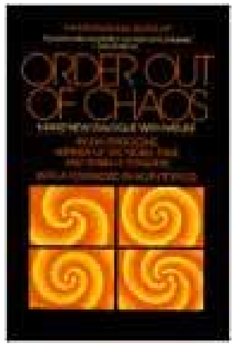
Which choice is amenable to modeling?

Masanori Arita (National Institute of Genetics)

Biology, the final frontier…

- Erwin Schroedinger "What is Life?" 1944
- Ilya Prigogine "Order out of Chaos" 1984
- ??? "Life out of Math" 2024





But, even single-protein phenomena have not been solved…

- Protein structure folding
- Temperature-compensation mechanism
- Molecular recognition

Success in Biology

- Great success in "DNA"

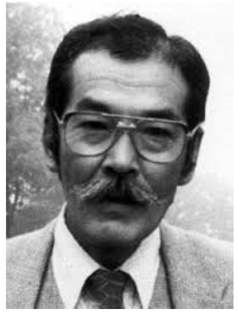
Genetics

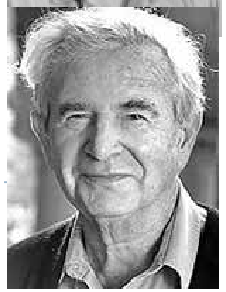
- Neutral theory of evolution (Motoo Kimura)
- Evolution by duplication (Susumu Ohno)

Informatics

- Dynamic programming
- P-value by random walk (Karlin & Altschul)

Other molecules need "killer" models.





Is Network the killer approach?

Most natural networks are:

Scale-free

- Node degree exhibits power-law distribution.
- Exponent λ of $y = x^{-\lambda}$ is between 2 ~ 3.

  (Barabasi *Science* 1999)

Small-world

- Local friends are clustered.
- Diameter of the network is extremely small.

  (Watts & Strogatz *Nature* 1998)

Clustering and path properties

Average path length

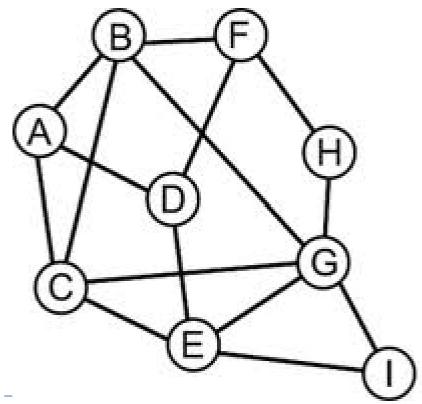
- Average number of steps between all pairs of nodes

- Random network has O (log n) path.

Clustering coefficient

- Density of edges among neighboring nodes

- Random network has a very low density.



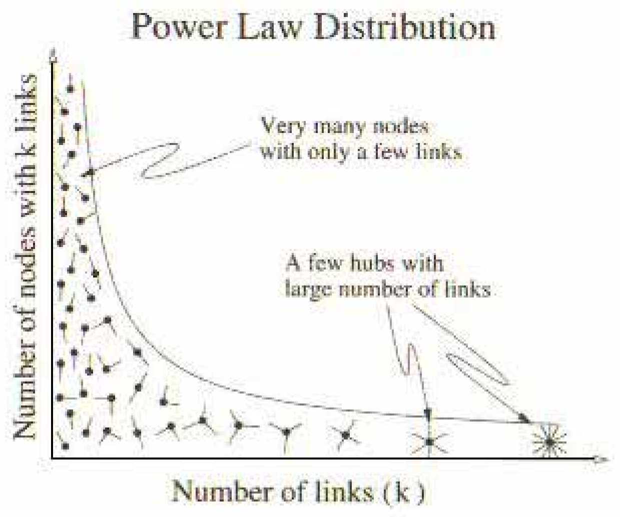
Albert, R, Barabasi, AL, *Reviews of Modern Physics* **74**, 47 (2002)

TABLE I. The general characteristics of several real networks. For each network we have indicated the number of nodes, the average degree $\langle k \rangle$, the average path length ℓ, and the clustering coefficient C. For a comparison we have included the average path length ℓ_{rand} and clustering coefficient C_{rand} of a random graph of the same size and average degree. The numbers in the last column are keyed to the symbols in Figs. 8 and 9.

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Why power law?

- Power law
  - lognormal distribution
  - multiplicative process



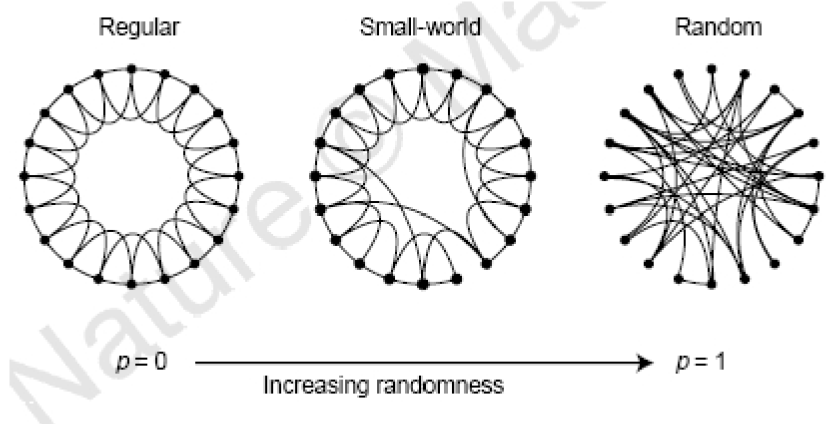
Real world is governed by multiplication, not by addition.

The central limit theorem:
"cumulative distribution approaches to lognormal (i.e. power law) distribution"

(E Fox-Keller *Bioessays* 2004, Arita *J Biochem* 2004)

Why small world?

- a little randomness + regular structure
  = small world



Real world is governed by
local interaction, not by global (i.e. regular).

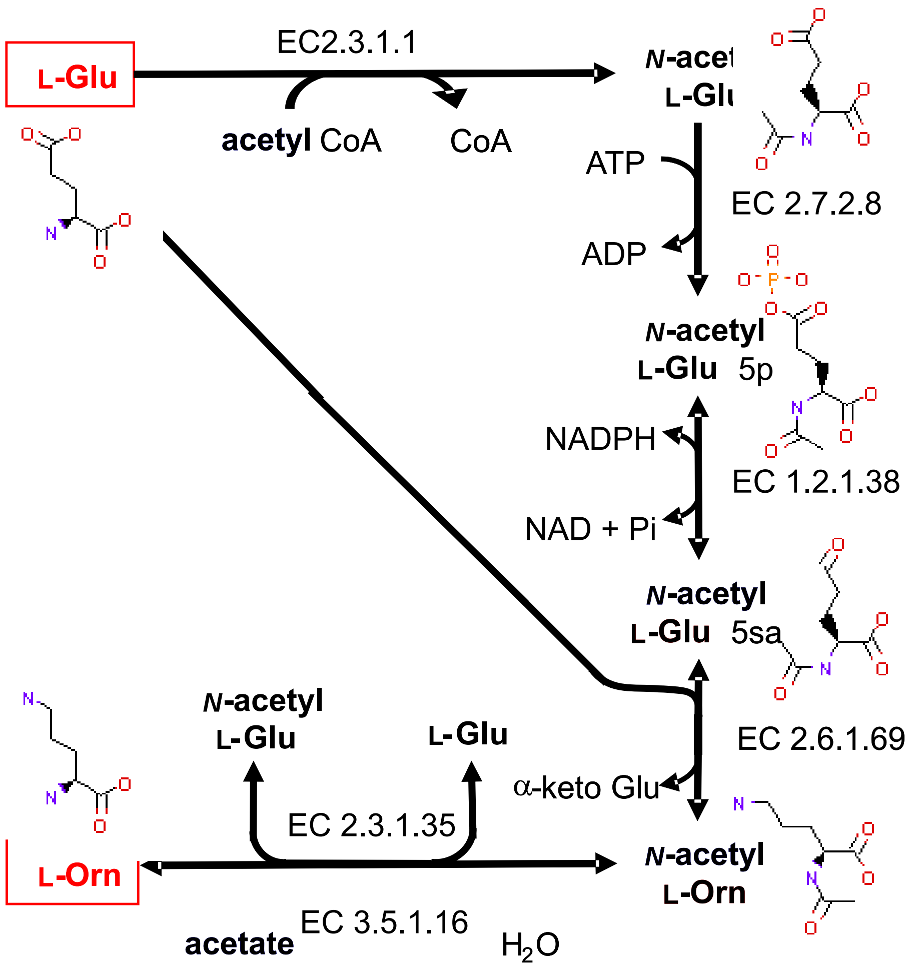
Numerical analysis:
"random noise make the world drastically smaller"

So the morals are

- In biology, scale freeness is a cumulative effect of complex interactions,
  and small worldness is due to noise…

- We need to look at more “simple” processes.
- Look at details of the target.

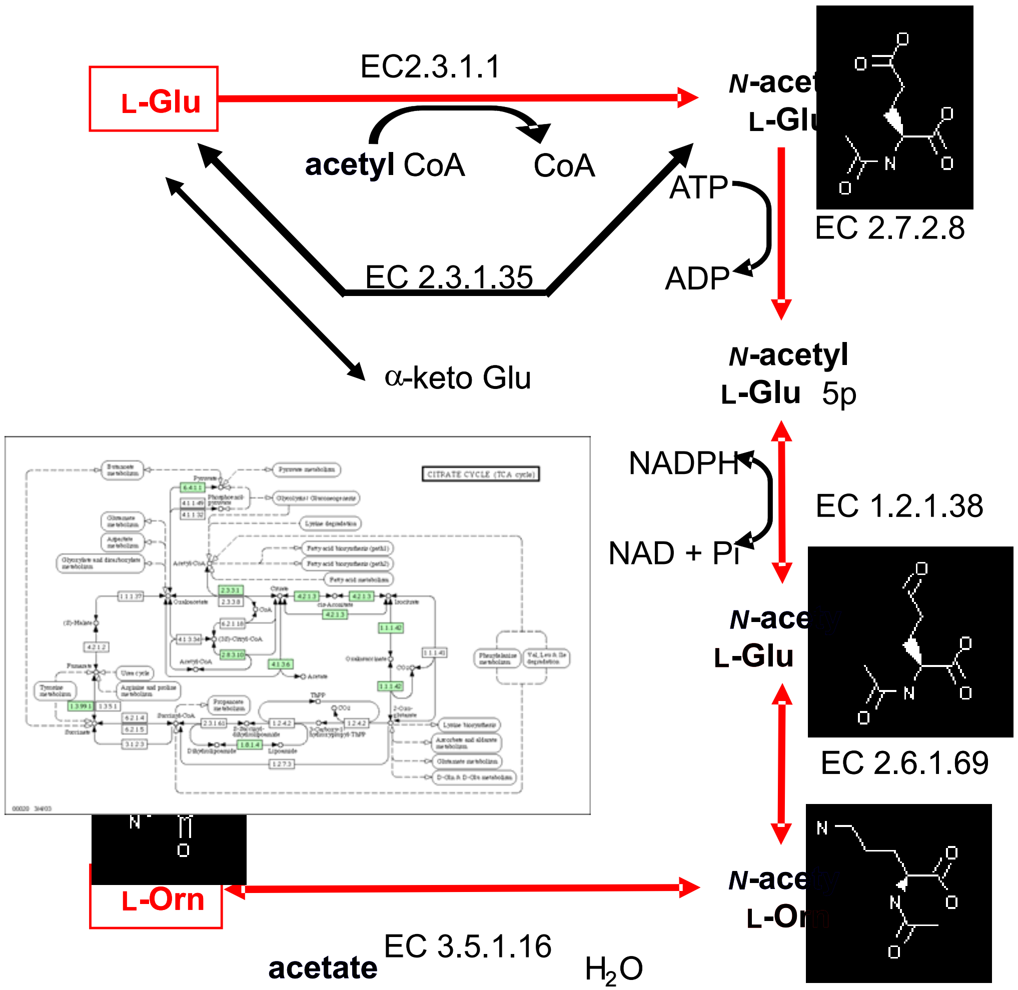
Complexity of the metabolism



Amino acid glutamine is converted to ornithine in 5 steps.

Glutamine is used twice to provide 2 nitrogens.

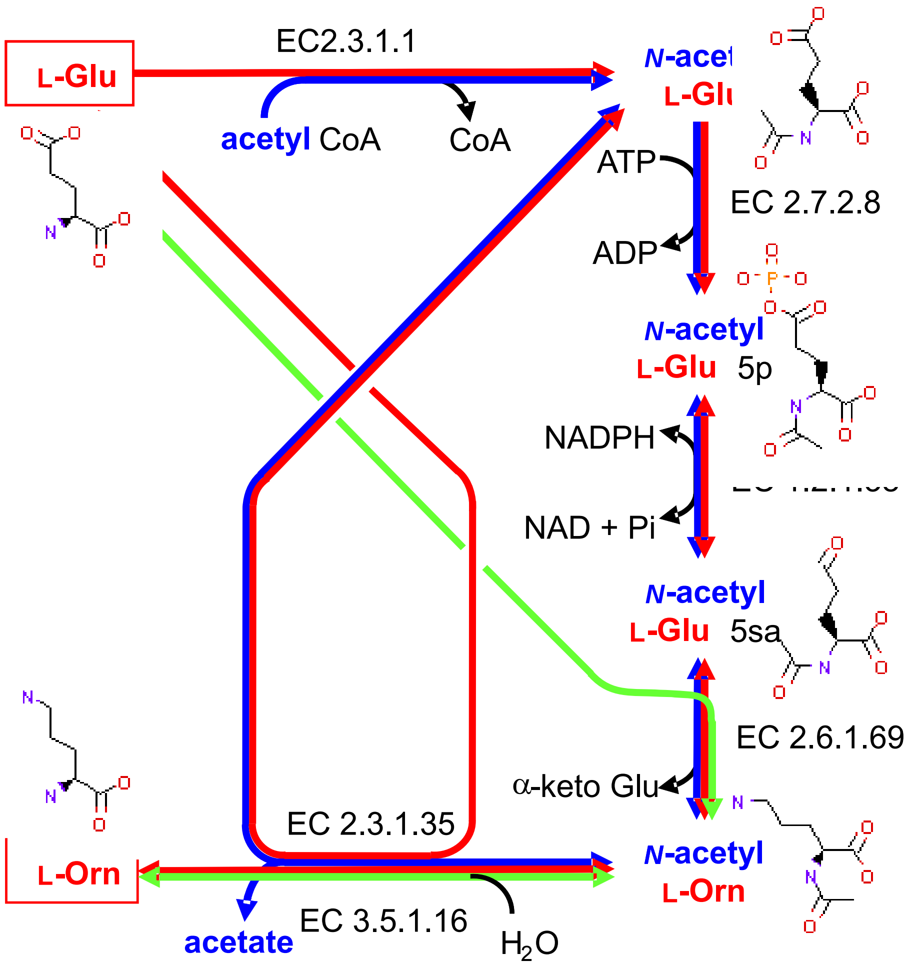
Textbook notation



Textbooks often show the red part only.

Databases too.

Ideal explanation (for this pathway)



Red: carbon

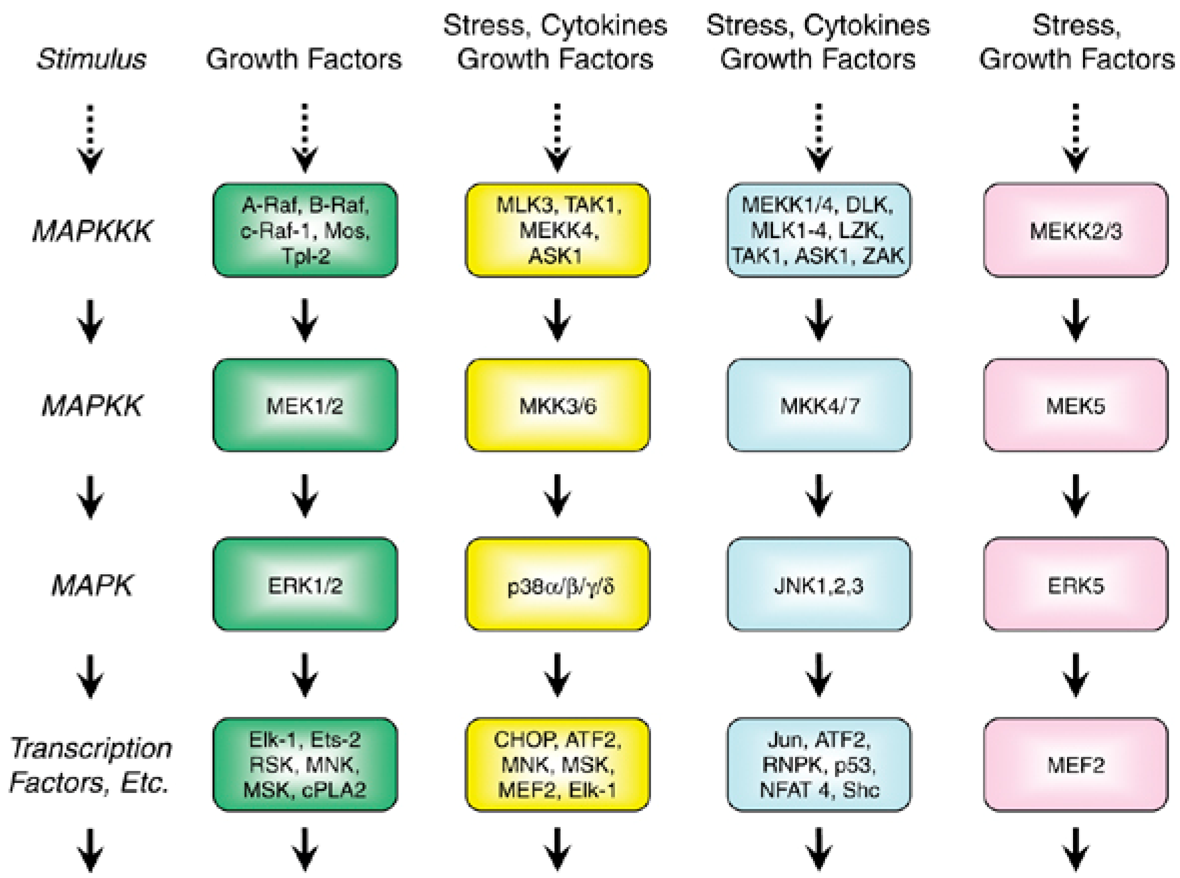
Blue: acetate

Green: nitrogen

Metabolism is neither small-world nor scale-free.

Complexity of protein network

- Famous MAPK cascade has 4 types in human.



Textbook information

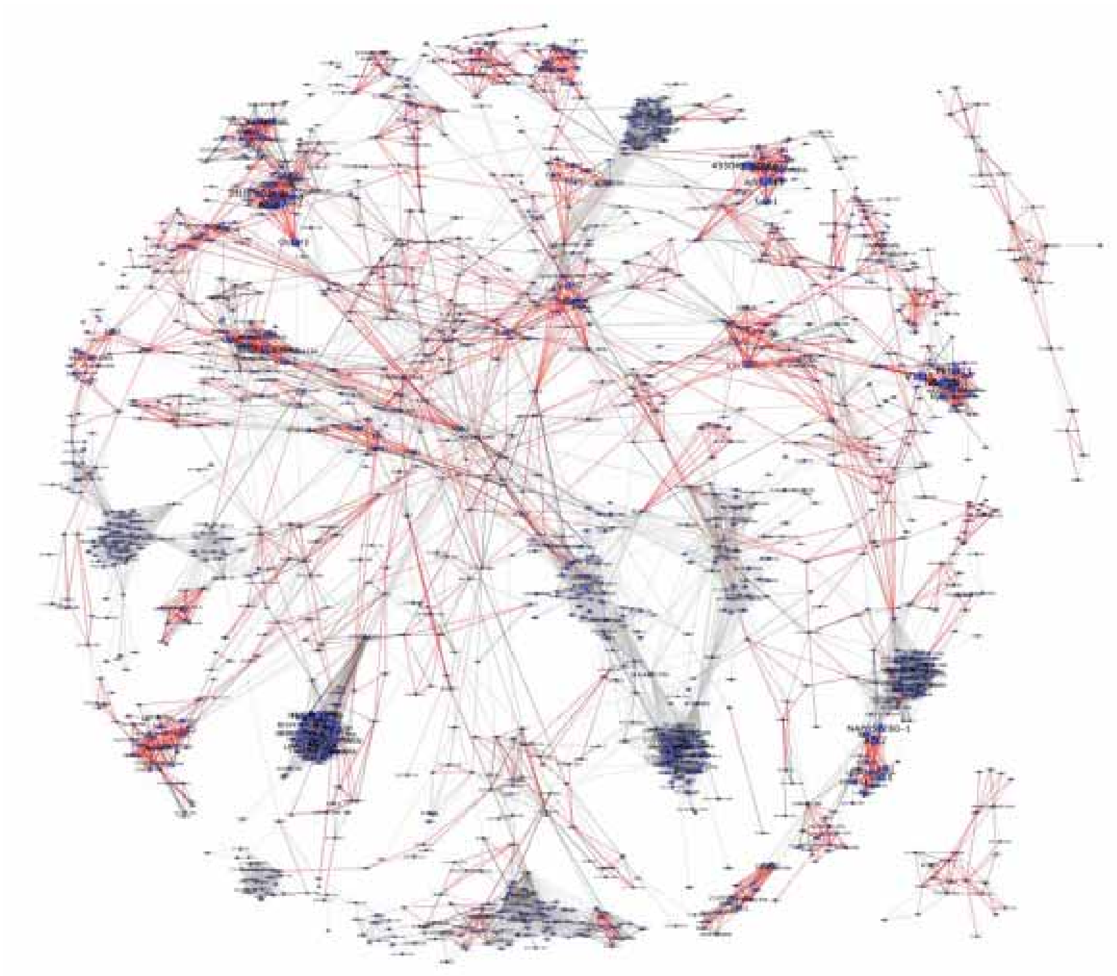
- MAPK (Mitogen-activated Protein Kinase) has multiple instances, under different names.
- Databases contain new/old names.

- Ontology is still a big issue.

  Mike Ashburner "Biologists would rather share a toothbrush than share a gene name"

Complexity of the gene network

- Dependence?



Finding a good target is difficult.

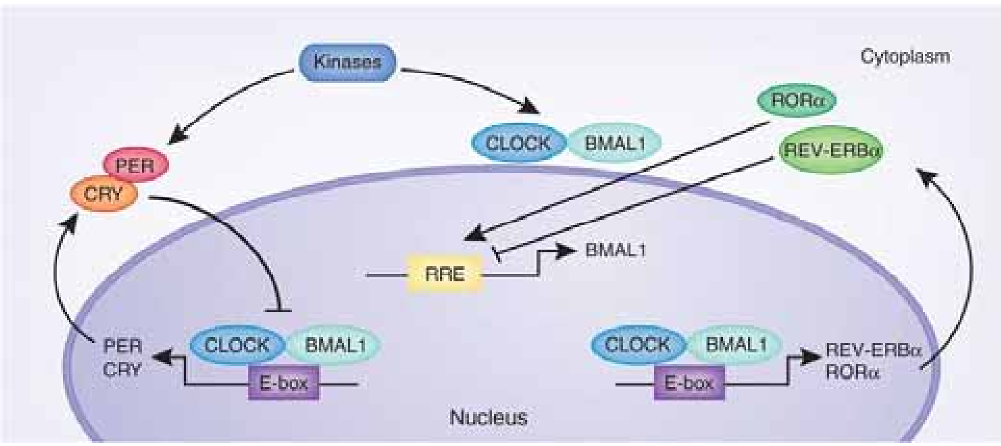
- Single protein/molecule topics
  - Protein folding
  - Fragmentation in mass spectrometry
  - Retention in chromatography
  - Catalytic site design …
- Regular and robust mechanism
  - Biosynthesis of metabolites (tracer study)
  - **Circadian oscillation**
  - :

Circadian mechanism

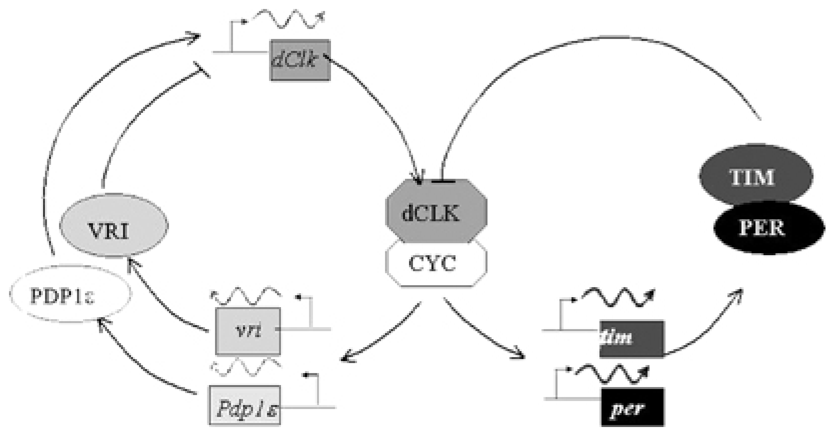
(with Dr. Yoshihiko Hasegawa)

- Gene expression of 24 h ± 1 cycle
- Entrained by (sun) light
- Mechanism is species-dependent

human

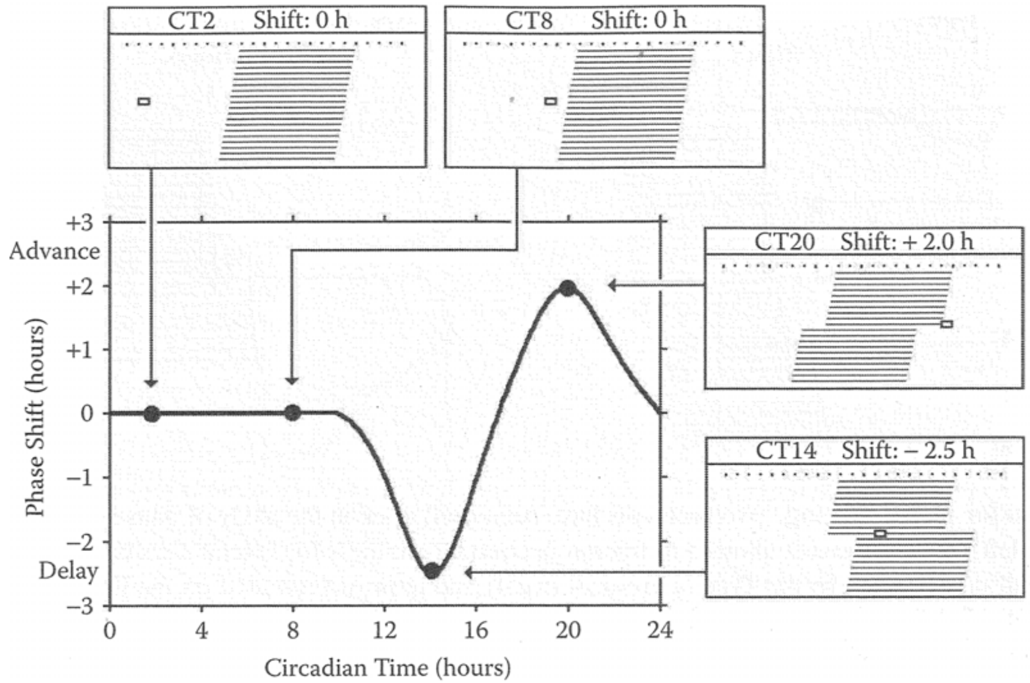


fly



Phase response curve

- Same perturbation results in different shift of the rhythm. PRC shows its relation.



[Refinetti, Circadian physiology, *Taylor & Francis*, 2005]

The model

- Deterministic rate equation of limit cycle

light sensitive parameter

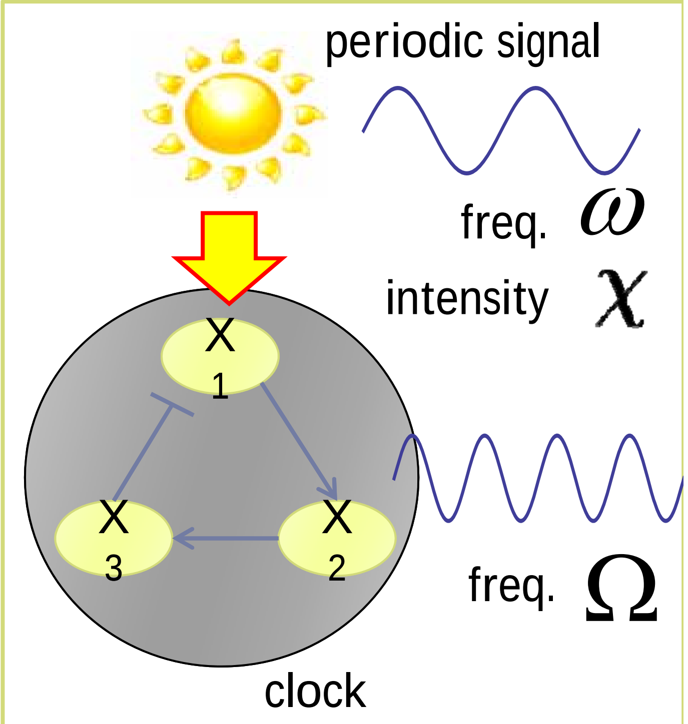
$$\frac{dx_i}{dt} = F_i(x;\rho) + Q_i(x)\xi_i(t)$$

Molecular species

multiplicative term

White noise

$\rho \to \rho + d\rho$



- The purpose is to analyze:
  - Regularity … period variance
  - Entrainability … range to be able to synchronize to external stimuli

Regularity and entrainability

- Regularity

$$\mathcal{V}_T \simeq \mathcal{V}_\phi \left(\frac{T}{2\pi}\right)^2 = \frac{T^3}{4\pi^3}\int_0^{2\pi} d\theta \sum_{i=1}^{N} U_i(\theta)^2 Q_i(\theta)^2$$

- Entrainability

$$\mathcal{E} = \Theta(\psi_M) - \Theta(\psi_m)$$

$$\chi\Theta(\psi_m) + \Omega < \omega < \chi\Theta(\psi_M) + \Omega$$

$$\psi_M = \mathrm{argmax}_\psi \Theta(\psi) \quad \psi_m = \mathrm{argmin}_\psi \Theta(\psi)$$

$$\Theta(\psi) = \frac{1}{2\pi}\int_0^{2\pi} d\theta\, Z(\psi+\theta)p(\theta)$$

$$Z(\phi) = \sum_{i=1}^{N} U_i(\phi)\frac{\partial F_i(\phi;\rho)}{\partial \rho}$$

(Arxiv 1305.0623)

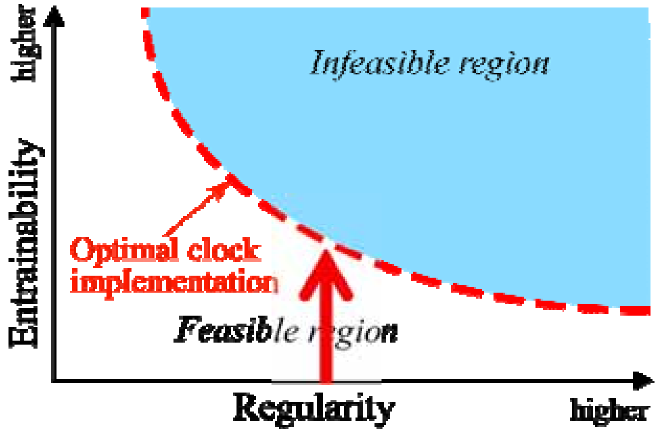
Variational method

- Maximize entrainability under constant regularity

Lagrange multiplier

$$\mathcal{L}[U] = \mathcal{E}[U] - \lambda \mathcal{V}_T[U]$$

Entrainability Regularity



Optimal phase-response curve (PRC)

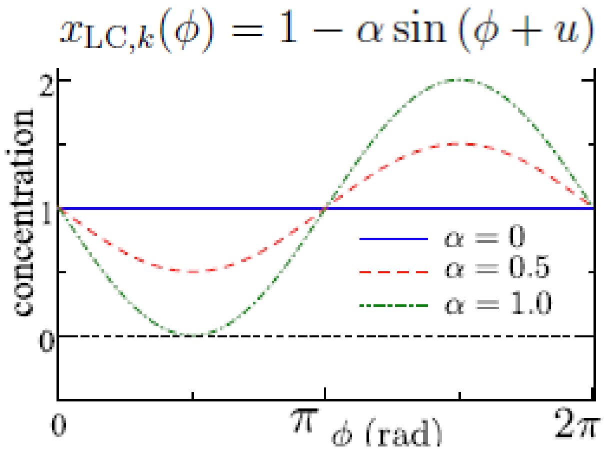
iPRC $$U_i(\phi) = \frac{\pi^2}{T^3\lambda} \frac{p(\phi - \psi_M) - p(\phi - \psi_m)}{Q_i(\phi)^2} \frac{\partial F_i(\phi;\rho)}{\partial \rho}$$

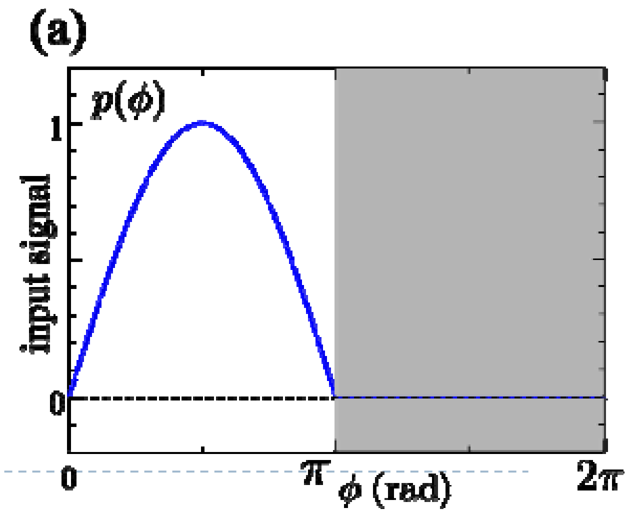
input signal

related to light entrainment mechanism

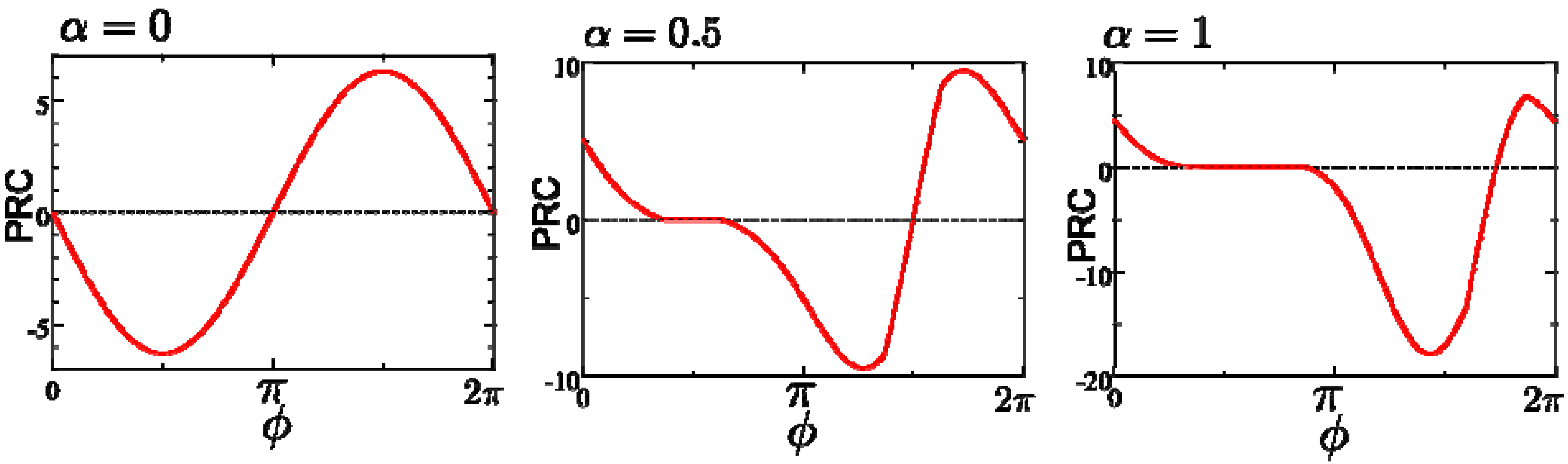
Model parameters

- T
  - Period of the oscillation
- σ_T
  - Variance of the period
- q
  - Noise intensity
- α
  - Amplitude of the key molecule affected by light





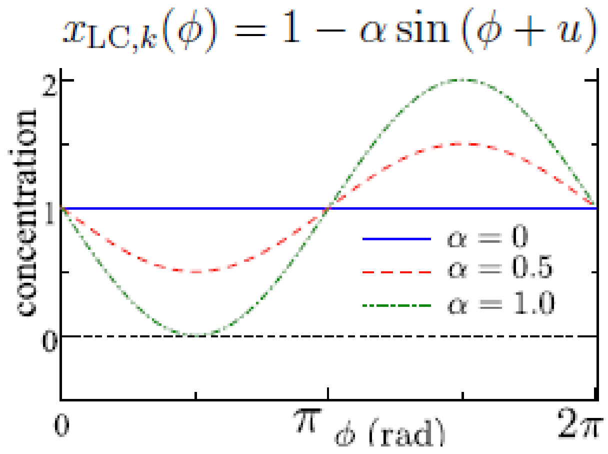
Optimal PRCs

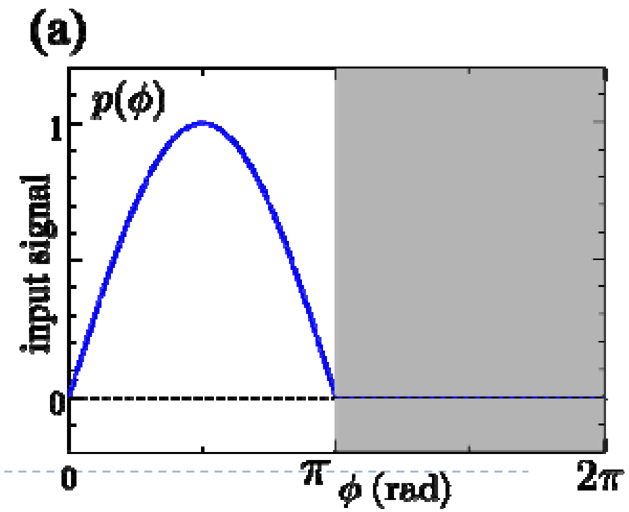


- Dead zone exists, a time during which light stimuli neither advance nor delay the clock
- Dead zone always exists for $\alpha > 0$

Intuitive explanation

- To optimize regularity, zero sensitivity is optimal (dead zone only).
- To achieve entrainability, some response is necessary (actual PRC).
- Light response is effective when the key molecule is abundant.





Observation

- Only under the solar input, we observe the dead zone.
- Dead zone appears for a wide parameter range.

→ **Actual circadian clocks are optimal for synchronization to daylight** !

(Hasegawa & Arita *Interface* accepted)

- No dead zone in peripheral clocks.
- When the pulse is long, we do not observe true PRCs.

Take home messages

- Do not work on noisy data.
- Choice of the modeling target is important.
- Keep the model simple.

For math details,

Dr.“Yoshihiko Hasegawa” (Google Sites)