The mathematical theory of large deviations provides a nonequilibrium thermodynamic description of complex biological systems that consist of heterogeneous individuals. In terms of the notions of stochastic elementary reactions and pure kinetic species, the continuous-time, integer-valued Markov process dictates a thermodynamic structure that generalizes (i) Gibbs' macroscopic chemical thermodynamics of equilibrium matters to nonequilibrium small systems such as living cells and tissues; and (ii) Gibbs' potential function to the landscapes for biological dynamics, such as that of C. H. Waddington's and S. Wright's.

Large Deviations Theory and Emergent Landscapes in Biological Dynamics

Hong Qian

Department of Applied Mathematics
University of Washington

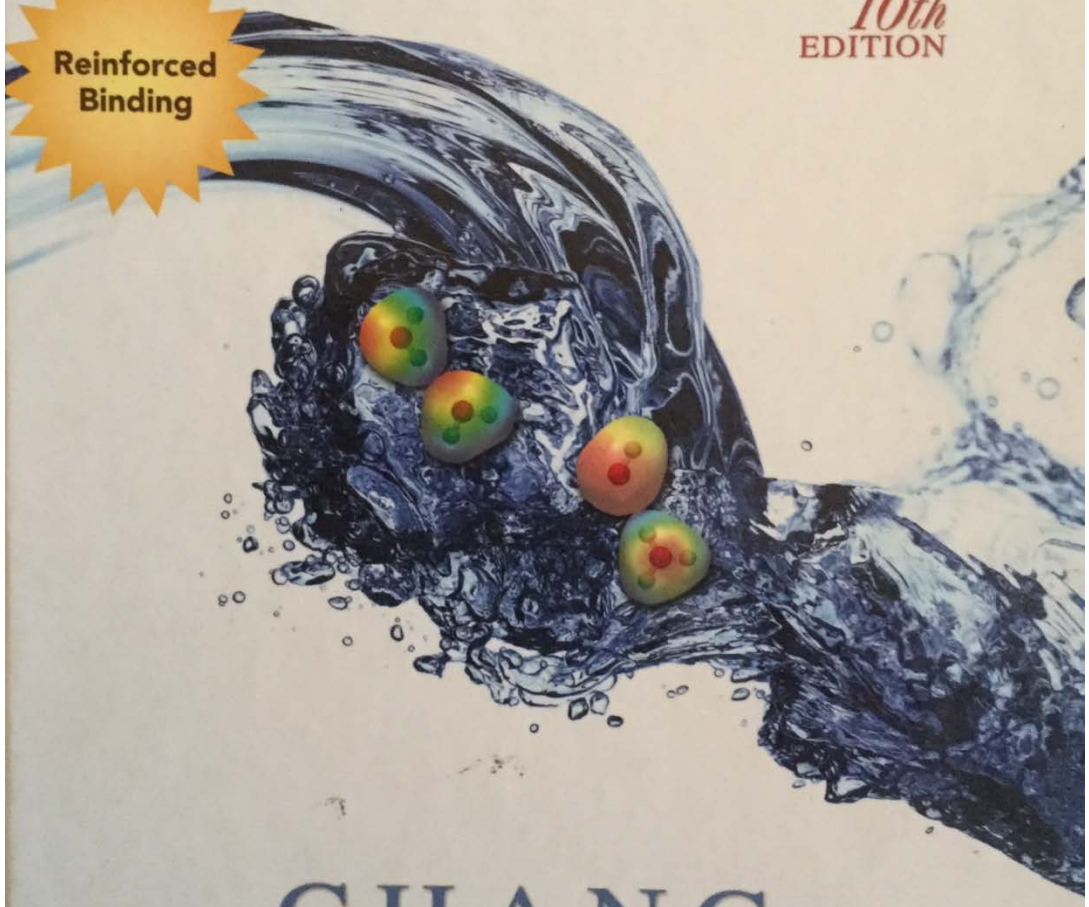
The biological world is stochastic

The maximum entropy principle (MEP) is a consequence of the *theory of large deviations* (Cover and van Campenhout, 1981). It is a twin of the Gibbs' statistical mechanics, which applies to only *thermodynamic equilibrium*.

CHEMISTRY

10th EDITION

Reinforced Binding



CHANG

Thermochemistry 228

Chemical Kinetics 556

How to represent (*e.g.*, describe) a biochemical or a biological system that consists of populations of individuals?

(i) classifying biochemical (or biological) individuals into populations of kinetic species;

(ii) counting the number of individuals in each and every "pure kinetic species";

(iii) representing changes in terms of "stochastic elementary processes".

A stochastic elementary process

$$A + 2B + C + E \xrightarrow{r_j} X + 3B + E$$

$$\boldsymbol{N} = (N_A, N_B, \dots, N_Z)$$

$$P\{\boldsymbol{N}(t + dt) = \boldsymbol{n} + \Delta\boldsymbol{n} | \, \boldsymbol{N}(t) = \boldsymbol{n}\}$$

$$= \begin{cases} r_j(\boldsymbol{n})dt + o(dt), if \ \Delta\boldsymbol{n} = \boldsymbol{\nu}_j \\ 1 - rdt + o(dt), if \ \Delta\boldsymbol{n} = \boldsymbol{0} \\ 0, \qquad otherwise. \end{cases}$$

A stochastic elementary process

$$A + 2B + C + E \xrightarrow{r_j} X + 3B + E$$

$$\boldsymbol{N} = (N_A, N_B, \ldots, N_Z)$$

instantaneous rate function

$$P\{\boldsymbol{N}(t + dt) = \boldsymbol{n} + \Delta\boldsymbol{n} \mid \boldsymbol{N}(t) = \boldsymbol{n}\}$$

$$= \begin{cases} r_j(\boldsymbol{n})dt + o(dt), if\ \Delta\boldsymbol{n} = \boldsymbol{\nu}_j \\ 1 - rdt + o(dt), if\ \Delta\boldsymbol{n} = \boldsymbol{0} \\ 0, \qquad otherwise. \end{cases}$$

stoichiometric coefficients

The mesoscopic stochastic description of a chemical reaction networks, elementary reactions in a continuously stirred reaction tank of volume *V*

$$v_{k1}^{+}X_1 + v_{k2}^{+}X_2 + \cdots + v_{kN}^{+}X_N$$

$$\underset{r_{-k}(\vec{n},V)}{\overset{r_{+k}(\vec{n},V)}{\rightleftharpoons}}$$

$$v_{k1}^{-}X_1 + v_{k2}^{-}X_2 + \cdots + v_{kN}^{-}X_N$$

On the border of this stochastic world, three major landmarks:

LLN: *law of large numbers,*
CLT: *central limit theorem,*
LDP: *large deviations principle.*

Law of Large Numbers

$$\lim_{M\to\infty} \bar{X}_M = \lim_{M\to\infty} \frac{X_1 + X_2 + \cdots + X_M}{M} = \mathbb{E}[X]$$

$$\lim_{M\to\infty} \frac{m_k}{M} = \mathbb{P}_k \, (k = 1,2, \ldots, K)$$

empirical mean value

stationary samples

expected value

$$\lim_{M\to\infty} \bar{X}_M = \lim_{M\to\infty} \frac{X_1 + X_2 + \cdots + X_M}{M} = \mathbb{E}[X]$$

empirical sample frequency

$$\lim_{M\to\infty} \frac{m_k}{M} = \mathbb{P}_k \, (k = 1,2, \ldots, K)$$

probability

- *statistical observables*
- *probabilistic concepts*

Central Limit Theorem

$$\lim_{M\to\infty} \frac{X_1 + X_2 + \cdots + X_M}{\sqrt{M}} = \infty$$

$$\lim_{M\to\infty} \left\{ \frac{X_1 + X_2 + \cdots + X_M}{\sqrt{M}} - \sqrt{M}\,\mathbb{E}[X] \right\}$$

$$= \mathcal{N}(0, \sigma^2)$$

Central Limit Theorem

$$\lim_{M\to\infty} \frac{X_1 + X_2 + \cdots + X_M}{\sqrt{M}} = \infty$$

$$\lim_{M\to\infty} \left\{ \frac{X_1 + X_2 + \cdots + X_M}{\sqrt{M}} - \sqrt{M}\,\mathbb{E}[X] \right\}$$

$$= \mathcal{N}(0, \sigma^2)$$

normal random variable

variance of X

Large Deviations Principle

$$f_{\bar{X}_M}(x;M) \to \delta(x-x^*), x^* = \mathbb{E}[X]$$

the premise

$$f_{\bar{X}_M}(x;M) \sim e^{-M\varphi(x)},$$

the existence

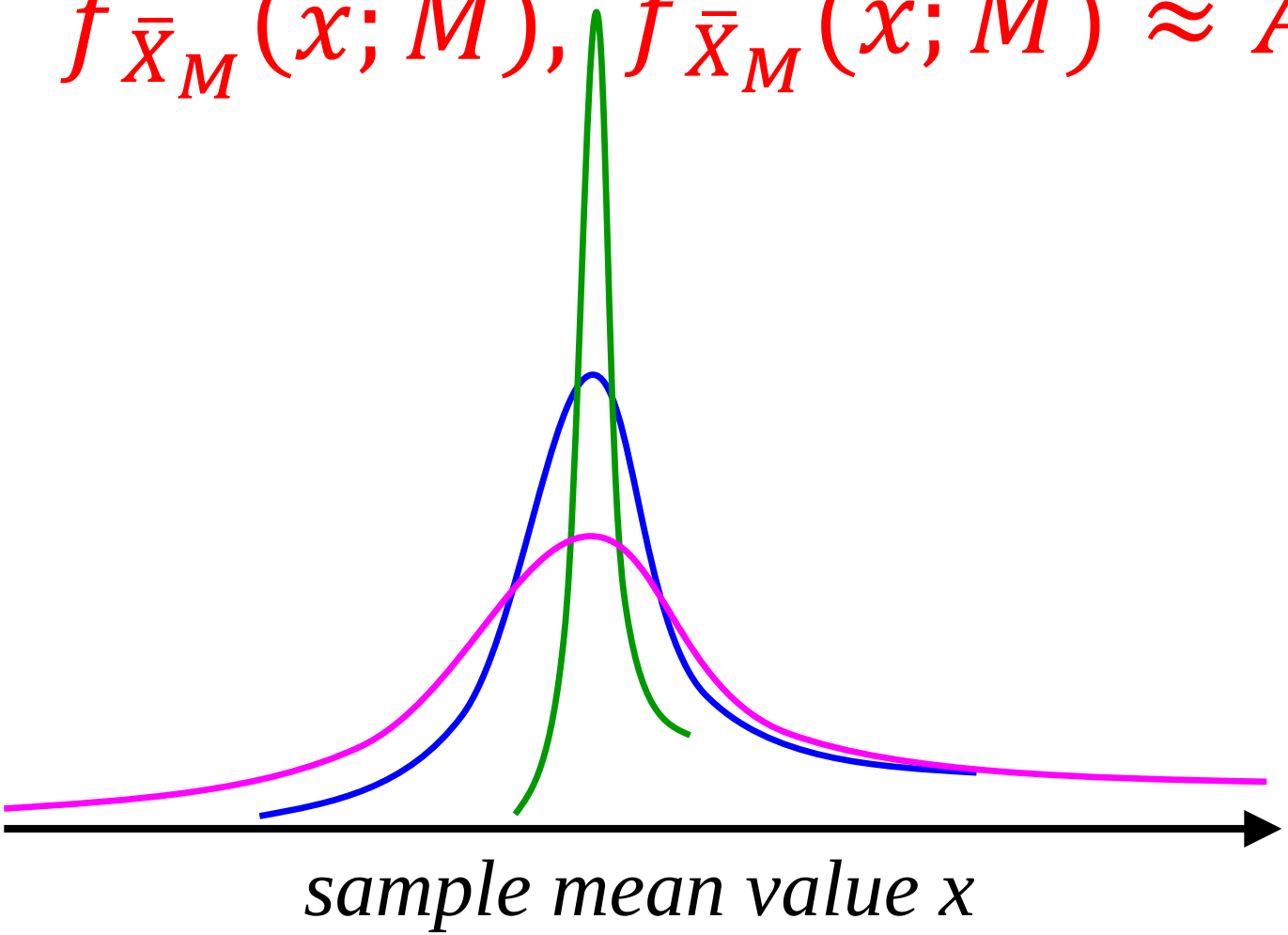
$$\min_{x\in\mathfrak{S}} \varphi(x) = \varphi(x^*) = 0,$$

$$\lim_{M\to\infty} -\frac{1}{M}\ln \mathbb{P}\{X_M \in \mathcal{A} \subset \mathfrak{S}\} = \min_{x\in\mathcal{A}} \varphi(x).$$

its property

When the M tends to infinite …

$$f_{\bar{X}_M}(x;M), \quad f_{\bar{X}_M}(x;M) \approx A\, e^{-M\varphi(x)}$$



The mesoscopic stochastic description of a chemical reaction networks, elementary reactions in a continuously stirred reaction tank of volume *V*

$$v_{k1}^{+}X_1 + v_{k2}^{+}X_2 + \cdots + v_{kN}^{+}X_N$$

$$\underset{r_{-k}(\vec{n},V)}{\overset{r_{+k}(\vec{n},V)}{\rightleftharpoons}}$$

$$v_{k1}^{-}X_1 + v_{k2}^{-}X_2 + \cdots + v_{kN}^{-}X_N$$

The *Law of Large Numbers* yields deterministic kinetics with ODEs

$$\frac{p_{Vz}(t)}{V} \to \delta(z - x(t))$$

$$\frac{dx_k(t)}{dt} = \sum_{j=1}^{M} \upsilon_{jk} \frac{r_{+k}(xV) - r_{-k}(xV)}{V}$$

$$= \sum_{j=1}^{M} \upsilon_{jk} (R_{+k}(x) - R_{-k}(x))$$

The *Large Deviations Theory*, or WKB ansatz:

$$\frac{p_{zV}(t)}{V} \to e^{-V\varphi(z,t)},\ \varphi(z,t) \geq \varphi(x(t),t) = 0,$$

$$\frac{\partial\varphi(x,t)}{\partial t} = \sum_{k=1}^{M} [R_{+k}(x) - R_{-k}(x)e^{-\boldsymbol{v}_k\cdot\nabla\varphi}] \times [1 - e^{\boldsymbol{v}_k\cdot\nabla\varphi}]$$

The *Large Deviations Theory*, or WKB ansatz:

$$\frac{p_{zV}(t)}{V} \to e^{-V\varphi(z,t)}, \quad \varphi(z,t) \geq \varphi(x(t),t) = 0,$$

first order in $\frac{\partial}{\partial t}$ *and* $\nabla\varphi$

$$\frac{\partial\varphi(x,t)}{\partial t} = \sum_{k=1}^{M} \left[R_{+k}(x) - R_{-k}(x)e^{-\boldsymbol{v}_k\cdot\nabla\varphi}\right] \times \left[1 - e^{\boldsymbol{v}_k\cdot\nabla\varphi}\right]$$

Macroscopic chemical equilibrium is between each and every local rate law and the global emergent potential

$$\frac{\partial \varphi(x,t)}{\partial t} = \sum_{k=1}^{M} \left[R_{+k}(x) - R_{-k(x)} e^{-\boldsymbol{v}_k \cdot \nabla \varphi}\right] \times \left[1 - e^{\boldsymbol{v}_k \cdot \nabla \varphi}\right]$$

$$\left[R_{+k}(x) - R_{-k(x)} e^{-\boldsymbol{v}_k \cdot \nabla \varphi^{ss}}\right] = 0!$$

$$\varphi^{ss}(x) = \sum_{j=1}^{N} x_j\left(\mu_j^o + \ln x_j\right)$$

It is the *equilibrium* Gibbs potential for ideal solution!

D.F. Anderson et. al., *Bull. Math. Biol.* 77, 1744-1767 (2015);
H. Ge and H. Qian, *Phys. Rev. E.* 94, 052150 (2016);
H. Ge and H. Qian, *J. Stat. Phys.* 166, 190-209 (2017).

Remember the term on the exponent?

$$e^{-\boldsymbol{v}_k \cdot \nabla \varphi^{ss}}$$

$$\mu_i(x) = \frac{\partial \varphi^{ss}(x)}{\partial x_i}$$

the chemical potential of species "*i*"

$$\boldsymbol{\upsilon}_k \cdot \nabla \varphi^{ss} = \sum_{i=1}^{N} \upsilon_{ki} \mu_i = \Delta \mu_k$$

the chemical potential difference of reaction "*k*"

$$\mu_i(x) = \frac{\partial \varphi^{ss}(x)}{\partial x_i}$$

entropy function

the chemical potential of species "*i*"

$$\boldsymbol{\upsilon}_k \cdot \nabla \varphi^{ss} = \sum_{i=1}^{N} \upsilon_{ki} \mu_i = \Delta \mu_k$$

the chemical potential difference of reaction "*k*"

$$\mu_i(x) = \frac{\partial \varphi^{ss}(x)}{\partial x_i}$$

entropy function

the chemical potential of spec

entropic force

$$\boldsymbol{\upsilon}_k \cdot \nabla \varphi^{ss} = \sum_{i=1}^{N} \upsilon_{ki} \mu_i = \Delta \mu_k$$

the chemical potential difference of reaction “*k*”

$$\varphi^{ss}(x) = \sum_{j=1}^{N} x_j\left(\mu_j^o + \ln x_j\right)$$

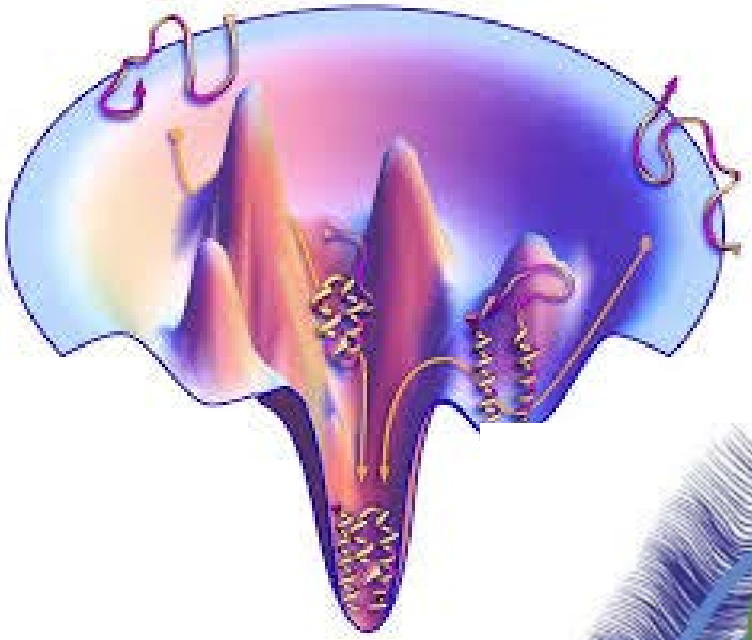
$$\left[R_{+j}(x) - R_{-j}(x)e^{-\boldsymbol{v}_j\cdot\nabla\varphi^{ss}}\right] = 0$$

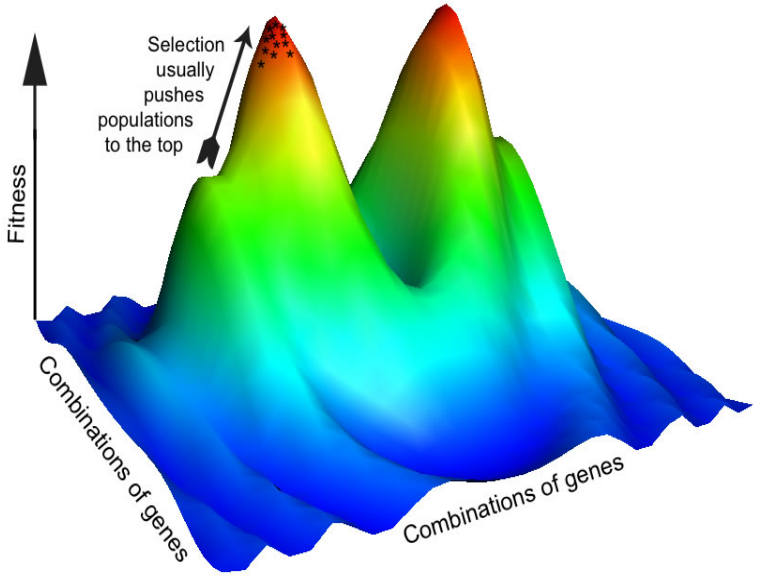
is equivalent to

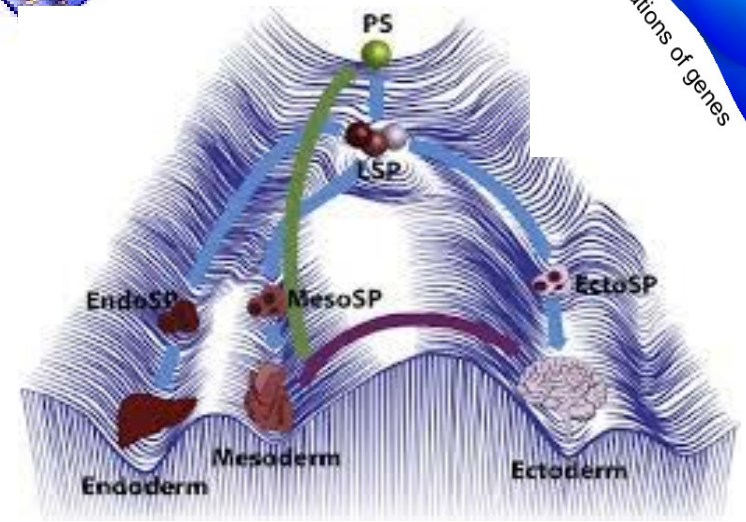
$$\ln\frac{R_{+j}(x)}{R_{-j}(x)} = -\Delta\mu_j(x), \text{ and}$$

$$\boxed{\ln\frac{k_{+j}}{k_{-j}} = -\Delta\mu_j^o}$$

In biology, all kinds of landscapes: protein folding (Frauenfelder), cell differentiation (Waddington), and evolution (Wright)







Summaries

- For complex systems which can be represented in terms of stochastic processes, entropy is a fundamental concept; and Gibbs energy comes the second;
- Gibbs' equilibrium chemical thermodynamics now has a stochastic kinetic foundation, and it is generalized to nonequilibrium settings;
- Chemical kinetics (how things change) and Gibbsian chemical thermodynamics (why things change) are general theory for any population dynamical systems, biological cells, tissues made of single cells, ecology, and infectious epidemics, etc.;
- It is a theory of "statistical counting" rather than "mechanics of the atoms within". The mathematics can be equally applied to other population dynamics.

Stochastic kinematics, *e.g.*, chemical kinetics, dictates chemical thermodynamics, of equilibrium and nonequilibrium states!

Thank you!