系统生物学

(Systems Biology)

马彬广

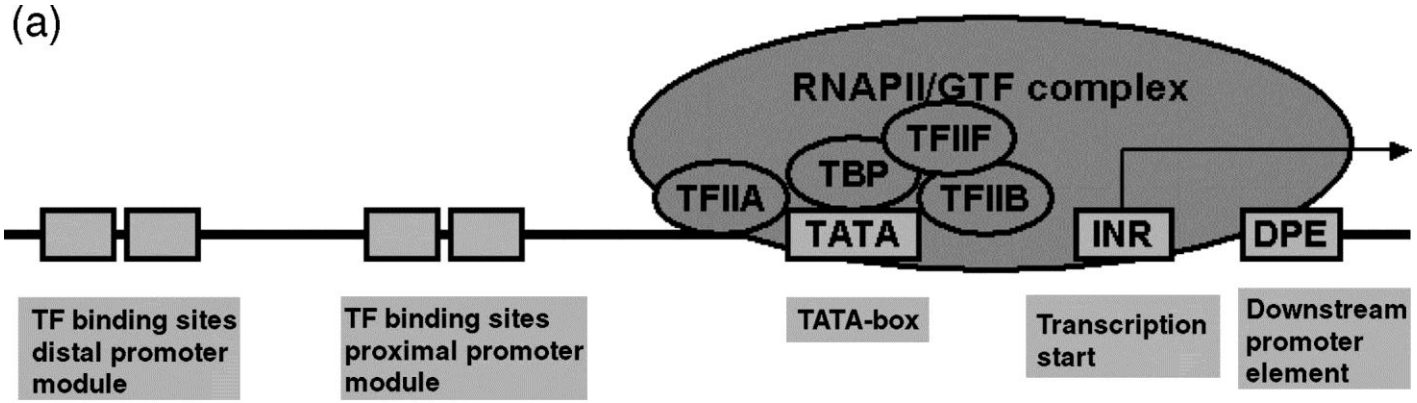
基因表达模型

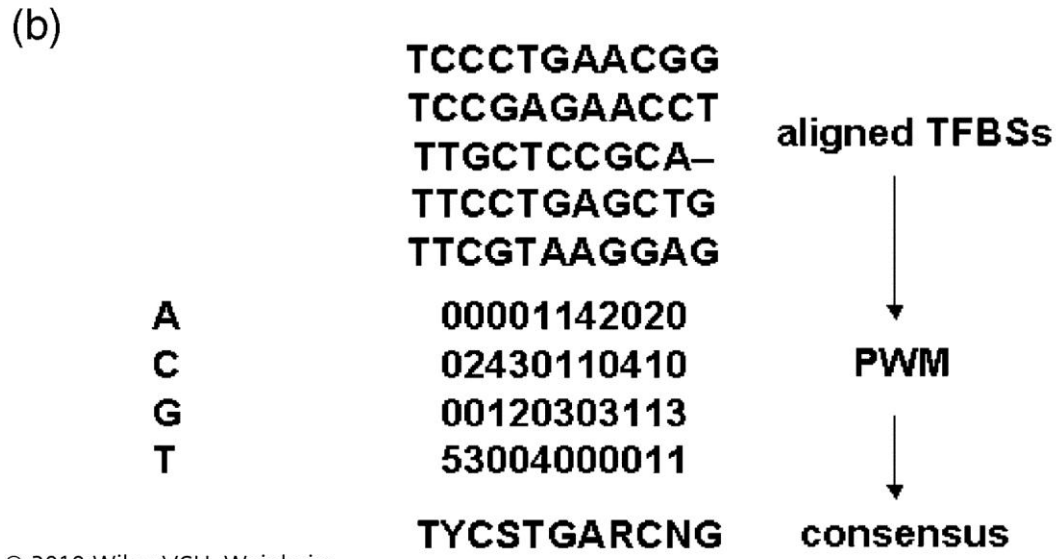
（第十二讲）

基因表达过程中的控制

- Transcriptional control: when and how often is a gene transcribed.
- RNA processing control: how is the RNA transcript spliced.
- RNA transport and localization control: which mRNAs in the nucleus are exported to cytosol and where in the cytosol are they localized.
- Translational control: which mRNAs in the cytosol are translated by ribosomes.
- mRNA degradation control: which mRNAs in the cytosol are destroyed.
- Protein activity control: decide upon activation, inactivation, compartmentalization, degradation of the translated proteins.

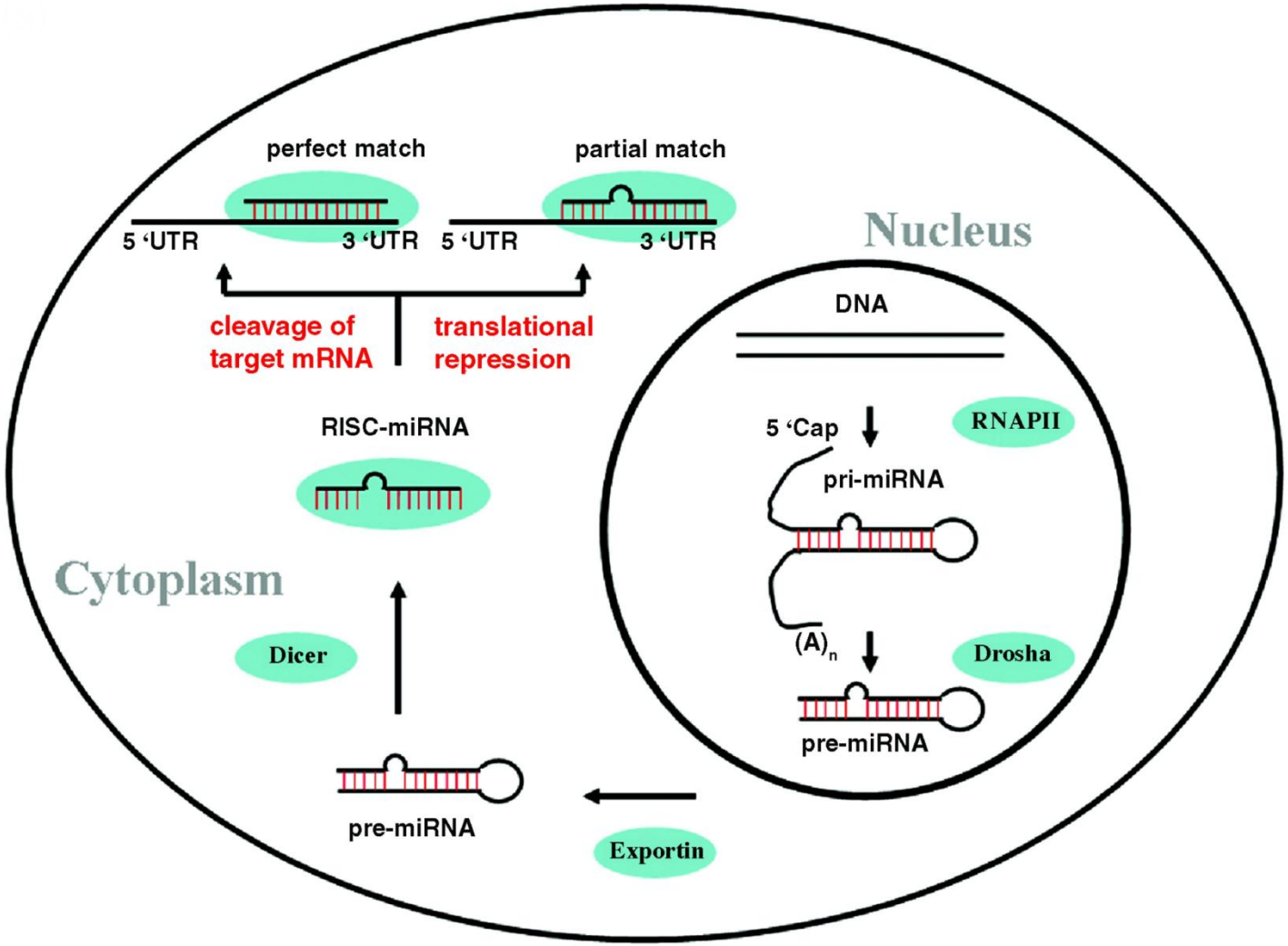
真核基因启动子的一般结构



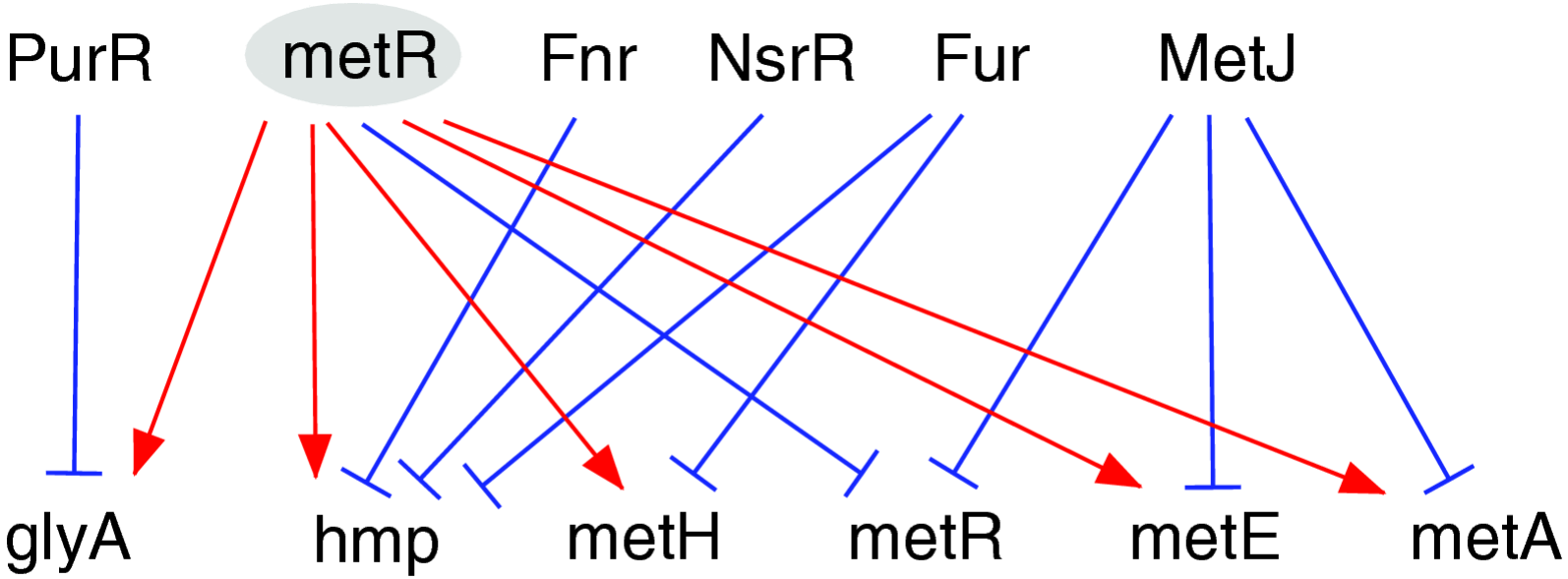


Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-06-01

真核细胞microRNA的转录后调控



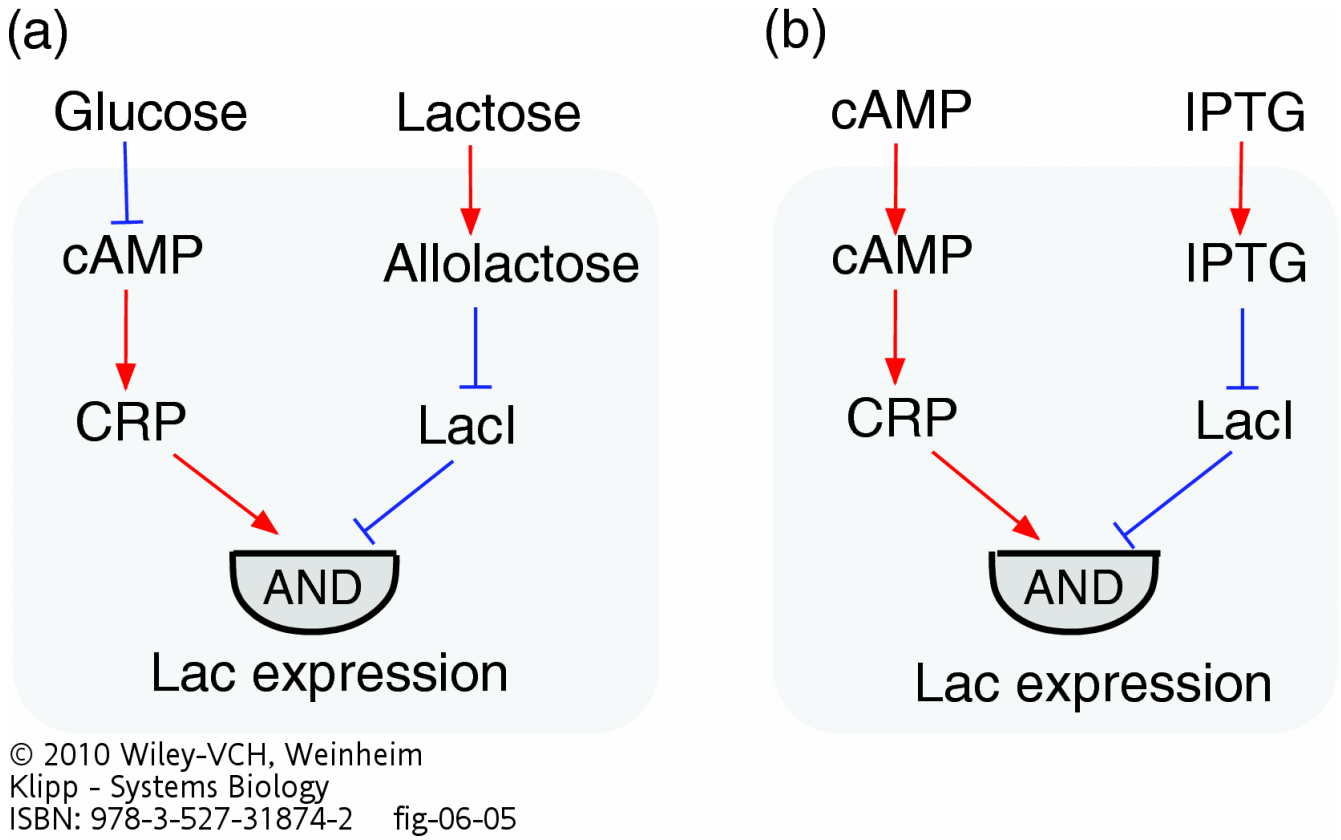
Genetic Network and Regulon



Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-06-04

Genetic network in *E. coli* comprising the transcriptional regulator MetR and its known target genes. MetR (gray ellipse) regulates a number of target genes (bottom), which therefore form the MetR regulon. Other regulators controlling these genes are shown on top. Arrows denote transcriptional regulation (blue: repression, red: activation). Data taken from the EcoCyc database.

The Lac Operon in Escherichia coli

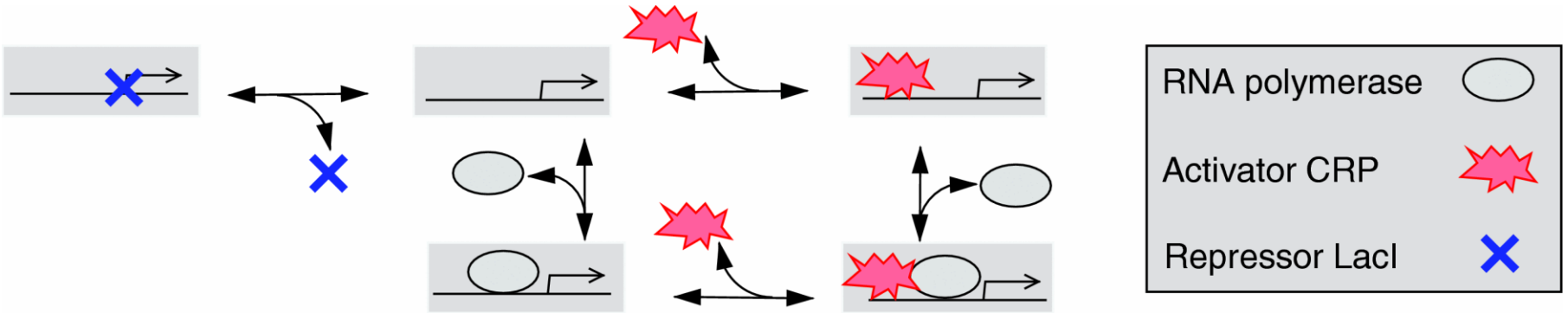


Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-06-05

The Lac operon is controlled by the transcriptional regulators CRP and LacI, which respond to extracellular levels of lactose and glucose. High expression of the Lac operon requires that lactose is present and glucose is absent.

In an experiment , the activities of CRP and LacI are regulated by extracellular levels of the ligands cAMP and IPTG. Effectively, both substances activate Lac expression.

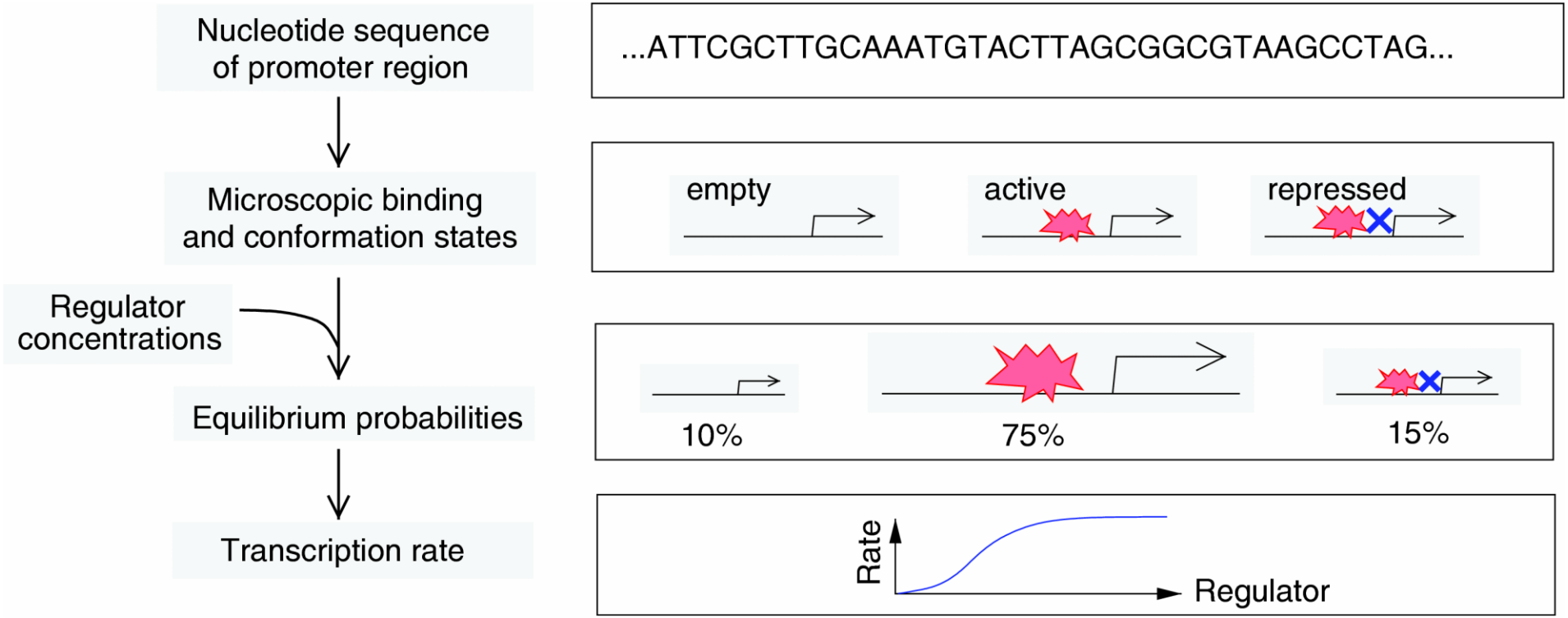
乳糖操纵子的微观状态描述



Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-06-06

Microscopic states of the Lac promoter (schematic model). The promoter can be bound by RNA polymerase, the activator CRP, and the repressor LacI. Bound activator increases the probability of polymerase binding (right). Transcription only occurs in states with bound polymerase (bottom). Bound repressor LacI inhibits binding of other molecules (left). In reality, the promoter sequence is much more complex: LacI can bind to several binding sites and cause DNA looping.

Binding Equilibrium and Transcription Rate



Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-06-07

Schematic relation between nucleotide sequence and transcription rates.

Gene Regulation Function

一般形式： $y(t) = f\left(\mathbf{x}(t),\ \mathbf{p}\right)$

假设：

1. 在所考虑的时间尺度上，存在平衡态，且平衡态下，各微观状态出现的概率依赖于各调控因子的浓度和它们与DNA序列的结合能。
2. 在每种微观状态下，转录过程以一定的速率随机启动。

经推导得出： $y = \dfrac{\sum_i w_i v_i}{\sum_i w_i},\ w_i = e^{-\frac{F_i}{K_B T}},\ F_i = E_i - TS_i$

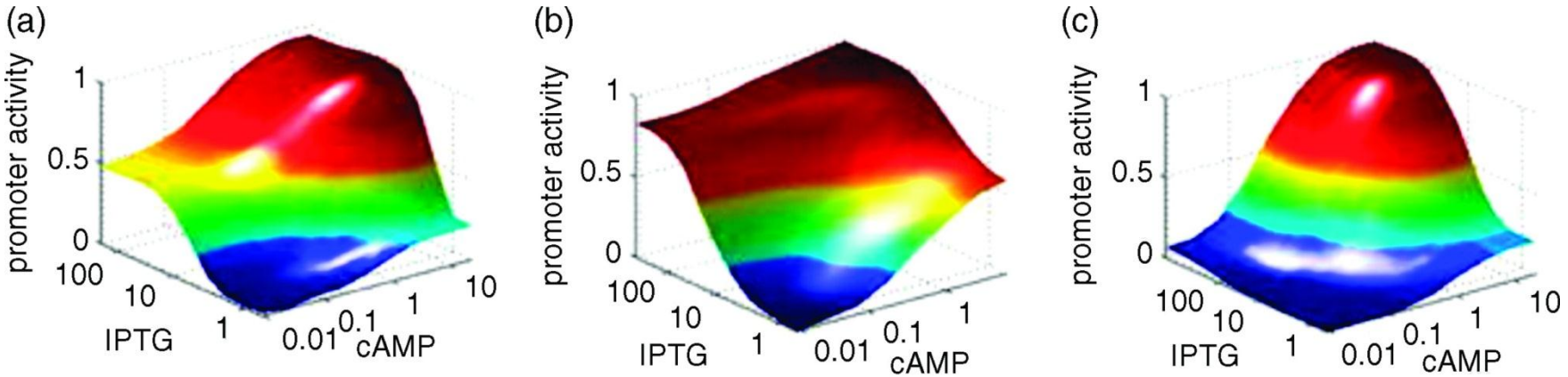
在实际应用中，有时用线性关系来近似表达基因调控函数:

$$y_i(t) = \sum_l a_{il} x_l(t)$$

i 表示当前的target gene, 而*l*表示调控因子。

Gene Regulation Function

$$f(x_1, x_2) = \frac{1 + a_1 x_1 + a_2 x_2}{1 + b_1 x_1 + b_2 x_2} \gamma$$



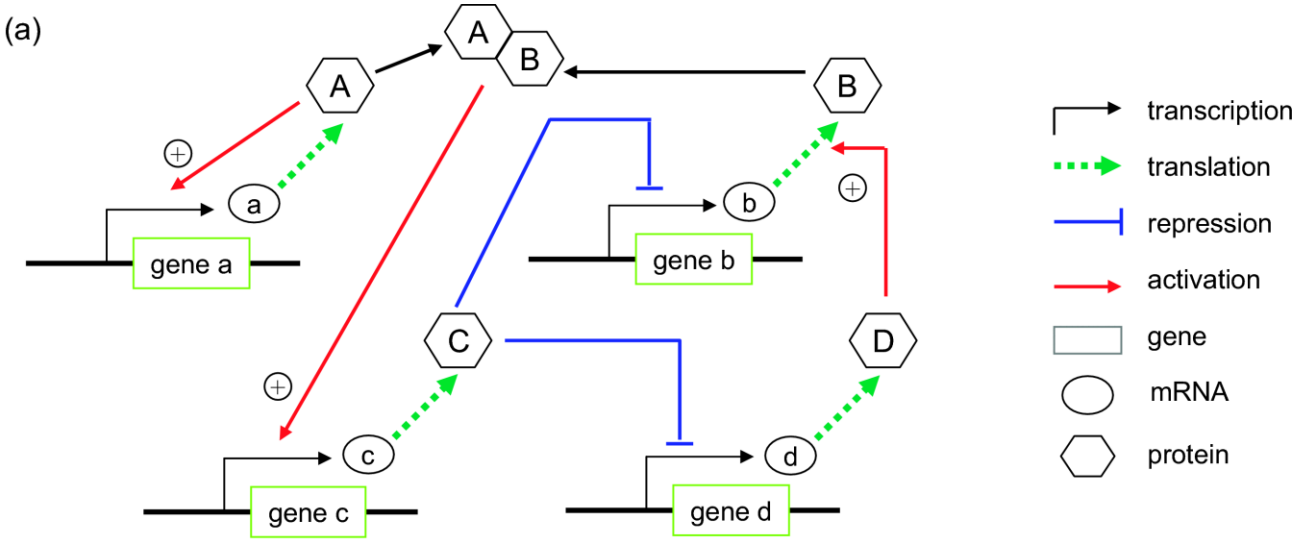
Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-06-08

Gene regulation functions of the wild type Lac operon and two variants obtained by altered promoter sequences. Measured gene regulation function in an *E. coli* wild-type strain. The strain U340 (obtained from a screen of *E. coli* strains with point mutations in the Lac promoter) shows an OR-like regulation function. Another strain, U339, shows an AND-like regulation function. From PLoS Biol, 4: e45.

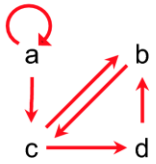
基因调控网络的多种建模手段

- Directed Graph G = (V, E).
- Bayesian Network.
- Boolean Network.
- ODE.
- Stochastic Process.

基因调控网络建模



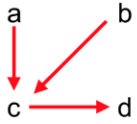
(b) Directed graphs



$V = \{a,b,c,d\}$

$E = \{(a,a,+),(a,c,+),(b,c,+),(c,b,-),(c,d,-),(d,b,+)\}$

(c) Bayesian network



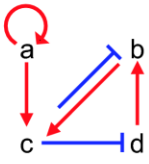
$p(x_a)$

$p(x_b)$

$p(x_c|x_a,x_b)$,

$p(x_d|x_c)$,

(d) Boolean network



$a(t+1) = a(t)$

$b(t+1) = (\text{not } c(t)) \text{ and } d(t)$

$c(t+1) = a(t) \text{ and } b(t)$

$d(t+1) = \text{not } c(t)$

Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-06-09

基因调控网络建模（Boolean Rules ）

Boolean Rules for a node with *K*=1 input

Input	Output			
A	0	A	not A	1
0	0	0	1	1
1	0	1	0	1
Rule	0	1	2	3

Boolean Rules for a node with *K*=2 inputs

Input		Output															
A	B	0	AND		A		B	XOR	OR	NOR		notB		notA		nAND	1
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
Rule		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

基因调控网络建模（Boolean Rules ）

The network

$$A \rightarrow B \rightarrow C \rightarrow D$$

has the maximal connectivity $K = 1$.

Let $A = \text{const.}$, $B = f_B(A) = \text{not } A$, $C = f_C(B) = \text{not } B$, $D = f_D(C) = C$ with the initial state $(A, B, C, D)(t_0) = (1, 0, 0, 0)$.

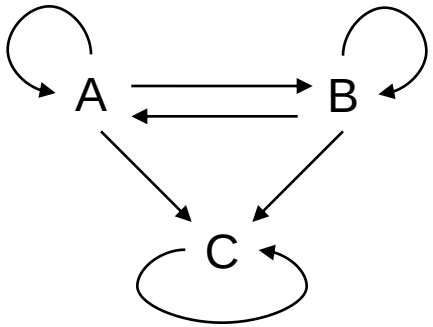
The following states are

$$(A, B, C, D)(t_1) = (1, 0, 1, 0)$$
$$(A, B, C, D)(t_2) = (1, 0, 1, 1)$$
$$(A, B, C, D)(t_3) = (1, 0, 1, 1)$$
$$\ldots \ldots \qquad \ldots$$
$$(A, B, C, D)(t_i) = (1, 0, 1, 1) \text{ for } i = 2, \ldots, \infty$$

After two steps, the system has attained a fixed point.

基因调控网络建模（Boolean Rules ）

The network



has 3 elements and $2^3 = 8$ states. Let the rules be

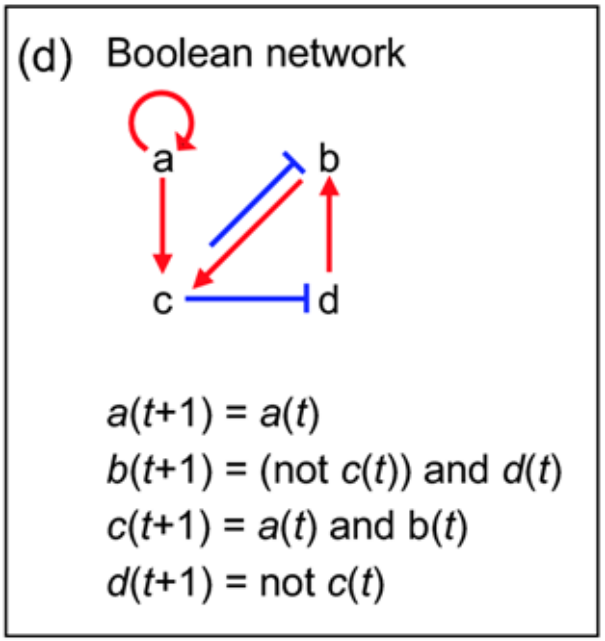
$$A(t+1) = A(t) \text{ and } B(t)$$
$$B(t+1) = A(t) \text{ or } B(t)$$
$$C(t+1) = A(t) \text{ or } (\text{not } B(t) \text{ and } C(t))$$

The Table below lists the successive states.

Successive states for the above Boolean network								
Current state	000	001	010	011	100	101	110	111
Next state	000	001	010	010	011	011	111	111

Fixed points: (000), (001), (010), (111);
Unreachable states from other state: (000), (001), (100), (101), (110).

基因调控网络建模（Boolean Rules ）



Apply the flowing Boolean rules

$$a(t+1) = f_a(a(t)) = a(t)$$
$$b(t+1) = f_b(c(t), d(t)) = \text{not } c(t) \text{ and } d(t)$$
$$c(t+1) = f_c(a(t), b(t)) = a(t) \text{ and } b(t)$$
$$d(t+1) = f_d(c(t)) = \text{not } c(t)$$

The temporal behavior is determined by the successive states as follows.

(0000) →(0001), (0001) →(0101), (0010) →(0000), (0011) →(0000),

(0100) →(0001), (0101) →(0101), (0110) →(0000), (0111) →(0000),

(1000) →(1001), (1001) →(1101), (1010) →(1000), (1011) →(1000),

(1100) →(1011), (1101) →(1111), (1110) →(1010), (1111) →(1010).

基因调控网络建模（Boolean Rules ）

- Start from 0: the system evolves to a steady state 0101;

- Start from 1: the system evolves toward a cyclic behavior of the following state sequence: 1000 → 1001 → 1101 → 1111 → 1010 → 1000.

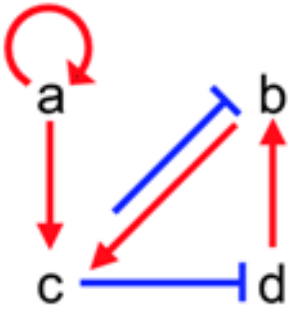
- The above steady states are called attractors, and other states are transient states.

- Kauffman revealed that the median number of attractors and the cycle length of the attractors are proportional to sqrt(N), N is the number of system components.

- Kauffman suggested to interpret the number of possible attractors as the number of possible cell types arising from the same genome.

- Ref: Journal of Theoretical Biology, 149: 467-505. and JTB 173: 427-440.

基因调控网络建模（ODE）



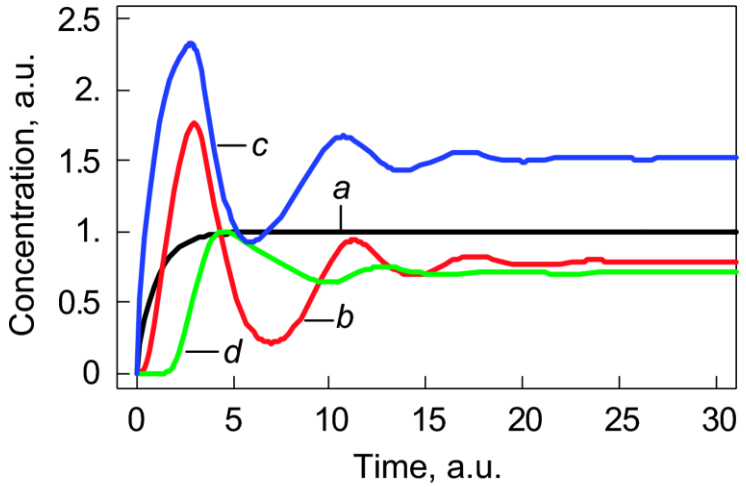
$$\frac{da}{dt} = f_a(a) = v_a - k_a \cdot a$$

$$\frac{db}{dt} = f_b(b,c,d) = \frac{V_b \cdot (d/K_b)^{n_d}}{(1+(d/K_b)^{n_d})(1+(c/K_{Ic})^{n_c})} - k_b \cdot b$$

$$\frac{dc}{dt} = f_c(a,b,c) = \frac{V_c \cdot (a \cdot b/K_c)^{n_{ab}}}{1+(a \cdot b/K_c)^{n_{ab}}} - k_c \cdot c$$

$$\frac{dd}{dt} = f_d(c,d) = \frac{V_d}{1+(c/K_{Ic})^{n_c}} - k_d \cdot d$$

Dynamics of the mRNA concentrations of the system. Parameters: *v*a = 1, *k*a = 1, *v*b = 1, *K*b = 5, *k*Ic = 0.4, *K*b = 0.1, *V*c = 1, *K*c = 5, *K*c = 0.1, *V*d = 1, *k*d = 1, *n*ab = 4, *n*c = 4, *n*d = 4. Initial conditions: *a*(0) = *b*(0) = *c*(0) = *d*(0) = 0.



ODE建模的局限性

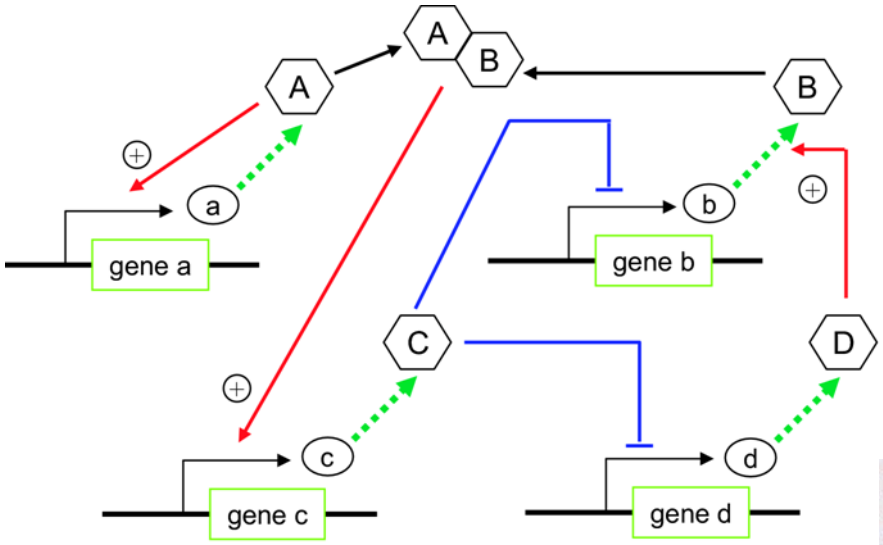
- ODE是变量连续变化的模型。但在实际的系统中，有些量是不连续的，连续变化只是一个近似。

- ODE是确定性的模型，没有考虑随机性，而在真实的系统中，随机性大量存在。

- 分子系统中不连续的例子：分子数目，特变是转录因子的拷贝数可能很低，事件过程也可以步进的方式变化。

- 随机性的例子：热运动、细胞内部状态的随机涨落、来自外部环境的随机因素等。

基因调控网络建模（随机过程）



Expression of protein A from gene a,	$a \rightarrow a + A$,	$k_A = 1.0$
Expression of protein B from gene b,	$b \rightarrow b + B$,	$k_B = 1.0$
Expression of protein C from gene c,	$c \rightarrow c + C$,	$k_C = 0.2$
Expression of protein D from gene d,	$d \rightarrow d + D$,	$k_D = 1.0$
Formation of protein complex AB,	$A + B \rightarrow AB$,	$k_{AB} = 1.0$
Inhibition of gene b	$b + C \rightarrow b_I + C$,	$k_{bI} = 1.0$
Activation of gene b,	$b_I \rightarrow b$,	$k_{bA} = 0.1$
Inhibition of gene c,	$c \rightarrow c_I$,	$k_{cI} = 0.1$
Activation of gene c,	$c_I + AB \rightarrow c + AB$,	$k_{cA} = 1.0$
Inhibition of gene d,	$d + C \rightarrow d_I + C$,	$k_{dI} = 1.0$
Activation of gene d,	$d_I \rightarrow d$,	$k_{dA} = 0.1$
Degradation of protein A,	$A \rightarrow$,	$k_{Ad} = 1.0$
Degradation of protein B,	$B \rightarrow$,	$k_{Bd} = 0.1$
Degradation of protein C,	$C \rightarrow$,	$k_{Cd} = 1.0$
Degradation of protein D,	$D \rightarrow$,	$k_{Dd} = 1.0$
Degradation of protein complex AB,	$AB \rightarrow$,	$k_{ABd} = 1.0$

基因调控网络建模（随机过程）

Stochastic simulations of the reaction network shown in Figure using the system .
(a) Individual simulation run,
(b) average over 100 simulation runs, both simulated with Gillespie's direct method, and (c) deterministic simulation.
Parameters: see text, initial abundances: gene a = 10; protein A = 0; gene b = 10; gene bi = 0; protein B = 0; gene c = 0; gene ci = 10; protein C = 0; gene d = 10; gene di = 0; protein D = 0; complex AB = 0.