

文章编号: 1000-5862(2016)04-0338-08

调控机制依赖的能量消耗

杨喜艳, 吴亚豪

(广东金融学院应用数学系, 广东 广州 510521)

摘要: 基因表达是一个动态非平衡过程, 必然要消耗能量. 故从能量消耗的观点研究了转录水平上的一个代表性的基因表达模型. 此模型不仅考虑了复杂的启动子结构(有2个调控位点), 还考虑了相互作用转录因子的功能(区分为招募、稳定和混合3种机制). 通过数学建模、理论分析和数值模拟, 研究发现: 招募机制最大化能量消耗(定义为熵的产生率), 稳定机制最小化能量消耗, 混合机制的作用介于二者之间. 此外显示出: 当启动子活性态的转录率变化时, 启动子活性与能量消耗之间呈现出反比关系, 即启动子活性越高(低), 能量消耗越少(多); 而当非活性态的转录率(或泄露率)变化时, 启动子活性与能量消耗之间呈现出正比关系, 即启动子活性越高(低), 能量消耗越多(少). 这些结果表明基因表达过程中的能量消耗是调控机制依赖的.

关键词: 基因表达; 能量消耗; 调控机制; 启动子活性

中图分类号: O 242; Q 141; Q 332 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2016.04.02

0 引言

生命活动必然要消耗能量. 从生物体内物理功能的实现(如细胞运动^[1])到特定生物学功能的实现(如控制细胞周期的生化振荡^[2]), 都需要消耗能量. 蕴含着能量消耗存在于生命活动的所有层面. 能量消耗在细胞系统中起着非常重要的作用, 如增加特异性^[3]、控制动力^[4]、减少变异^[5]以及响应信号^[6]. 近年来, 通过对感应-适应系统的研究发现, 适应性状态的稳定^[7]、短时间内强响应和长时间内精确适应性的获得^[8], 以及降低噪声的同时维持高度的敏感性^[9], 这些生化过程的实现都需要消耗能量. 在基因调控系统中, 需要通过生物计算来协调细胞内外的信号过程, 这意味着基因表达过程不可避免地会消耗能量^[10]. 已有研究探索了两状态基因模型中表达噪声和能量消耗之间的关系^[11], 但是对于启动子具有复杂结构的基因模型, 转录因子的作用如何影响基因表达过程中的能量消耗还不清楚. 因此, 从理论上探讨基因调控系统中的不同调控机制如何影响细胞能量消耗是非常有意义的.

著名的兰道尔原理认为不可逆的计算需要消耗能量^[12]. 基因表达中的基本生化过程如转录、翻译

等都是不可逆的, 因而需要消耗能量. 例如, 通过对酵母菌的研究发现, 能量消耗使得基因转录和翻译的速率在可控制的范围之内^[13]. 另有研究通过从高磷酸盐模块中回收能量的做法, 探测了基因表达过程中转录和翻译等基本生化反应所消耗的能量, 即蛋白质的合成大约消耗90%的能量, 而mRNA的转录过程需要不到10%的能量^[14]. 此外, 许多其它的生化过程如转录因子作用、染色体重组以及DNA甲基化都需要消耗能量用以保证转录的特异性、保真性和有效性^[15].

众所周知, 系统中非平衡态机制的出现(这意味着细致平衡的破坏)导致了能量消耗^[6, 16]. 已有研究表明精确适应性等重要生物功能的实现要求系统远离平衡态^[8]; 特别地, 基因表达过程中的每一步都依赖于一系列的非平衡态的过程, 来确保其特异性、保真性和有效性. 诚然, 非平衡机制在基因表达中起着非常重要的作用, 但是基因表达过程中动力学参数如何导致非平衡机制的出现(或打破细致平衡)还不清楚, 特别是在启动子具有复杂结构的模型中, 调控机制如何影响系统的能量消耗还需进一步研究. 基于此, 下面将研究复杂基因模型中转录因子的调控作用和能量消耗之间的关系.

本文考虑具有复杂启动子结构(2个活性态和

收稿日期: 2016-03-10

基金项目: 国家自然科学基金(11526058), 广东省教育厅青年创新人才类(2015KQNCX103), 2015年广东金融学院“创新强校工程”(高校重大科研成果培育计划类)和2016年广东金融学院创新强校规划资助项目.

作者简介: 杨喜艳(1983-), 女, 湖南邵阳人, 讲师, 博士, 主要从事系统生物学的研究.

1 个非活性态) 的基因表达模型,同时还考虑转录因子作用下的 3 种不同的分子调控机制: 招募机制、稳定机制和混合机制,故所提出的多状态启动子模型是具有生物学意义的: 1) 细胞中多状态基因启动子的存在已经在实验上得到证实^[17]; 2) 在生物中找到原型,如在 λ 溶源性噬菌体的启动子中包含了 3 个操作位点,其就属于这种多 ON 基因模型^[18]. 需要指出的是 P. S. Gutierrez 等^[19-20] 首先提出了转录因子的这 3 种调控机制,通过对概率分布矩的计算分析了 3 种机制下的表达噪声. 与该方法不同的是,本文利用生成函数的方法获得了启动子在各状态的精确概率分布,这也是用来计算系统能量消耗的关键. 虽然近年来的单分子实验能够对基因表达过程进行观测^[21-22],但是转录因子的作用、启动子活性与能量消耗之间的关系如何还不清楚,故将对此问题展开研究.

本文主要关注 2 类转录速率(即活性和非活性转录速率)下的能量消耗. 这 2 类转录速率也是系统打破细致平衡,消耗能量的关键动力学参数. 通过理论分析和数值模拟发现: 招募机制诱导最大的能量消耗,稳定机制诱导最小的能量消耗; 当活性态转录速率变化时,启动子活性和能量消耗之间呈反比关系,即启动子活性越高(低),能量消耗越少(多); 而当非活性态转录速率变化时,启动子活性和能量消耗之间呈正比关系,即启动子活性越高(低),能量消耗越多(少). 本文所获得的结果都是定性的,与参数的选取无关. 即利用定量化的手段研究基因表达过程中的能量消耗是非常有意义的,研究结果揭示了复杂基因调控系统中能量消耗的分子机制.

1 基因模型与分析结果

1.1 模型

为了清晰地展示转录因子的作用对基因表达和能量消耗的影响,下面考虑一个转录水平上的基因模型,如图 1 所示. 在这一模型中,启动子有 2 个调控位点和 3 个状态: 2 个活性态(或 ON 态) A_1 和 A_2 , 1 个非活性态(或 OFF 态) I . 假定在活性态时,转录因子与 RNA 聚合酶的作用能够导致基因产物(即 mRNA) 以较高的速率产生,而在非活性态时仍有少些基因产物的生成(即启动子有泄露). 同时,假定转录因子的作用直接影响启动子在 2 个活性态的切换率,从而导致 2 种不同的功能: 第 1 种称招募功能,即先前绑定到其中一个调控位点的转录因子能够招募其它转录因子绑定到剩余的调控位点; 第 2 种称为稳定功能,即 2 个转录因子间的相互作用导

致其中一个转录因子从绑定位点解离的速率更慢,从而更稳定.

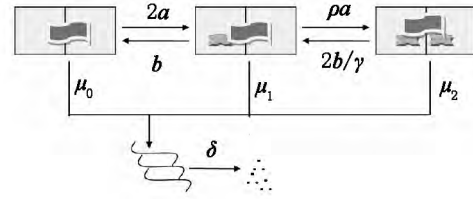
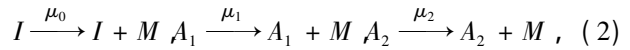
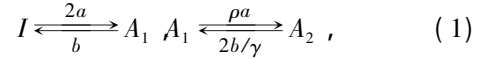


图1 2个调控位点的3状态基因模型示意图

为方便理解这个模型,利用下列生化反应来刻画:



反应(1)描述了启动子在3个状态间的切换,其中 a 和 b 分别表示转录因子的结合率和解离率, ρ 和 γ 分别刻画了转录因子的招募和稳定功能. 注意到在(1)中包括了转录因子的3种调控机制: 招募机制(recruitment mechanism RM), 稳定机制(stabilization mechanism SM)以及混合机制(mixture mechanism MM). 特别地,通过参数 ρ 和 γ 的不同取值来刻画这3种机制: 对于 RM $\rho > 1, \gamma = 1$; 对于 SM $\rho = 1, \gamma > 1$; 对于 MM $\rho > 1, \gamma > 1$. 此外,为方便,记 $\sigma = \rho\gamma$ 表示转录因子的总作用效能. 显然,在 RM 中 $\sigma = \rho$, 在 SM 中 $\sigma = \gamma$ 以及在 MM 中 $\sigma = \rho\gamma$. 反应(2)刻画了 mRNA 的转录,其中 μ_0 是基因处于非活性态的转录率(或泄露率), μ_1 和 μ_2 是活性态的转录率. 不失一般性,假设 $\mu_0 < \mu_1 < \mu_2$. 反应(3)描述了 mRNA 的降解.

让 $P_0(m, t)$ 和 $P_i(m, t)$ 分别表示基因在 I 和 $A_i (i = 1, 2)$ 状态于 t 时刻有 m 个 mRNA 分子的概率分布,则相应的主方程可表示为

$$\frac{dP(m, t)}{dt} = T_3(m) P(m, t) + \Delta_3(E - I) [mP(m, t)] + \Omega_3(E^{-1} - I) [P(m, t)], \quad (4)$$

这里 I 表示单位算子, E 和 E^{-1} 表示平移算子(或称为步长算子),即对任意函数 f 和任意整数 n ,有 $E[f(n)] = f(n+1)$, $E^{-1}[f(n)] = f(n-1)$. 矩阵 T_3, Δ_3, Ω_3 分别称为启动子状态之间的转移矩阵、mRNA 的转录矩阵以及 mRNA 的降解矩阵,其中

$$T_3 = \begin{pmatrix} -2a & b & 0 \\ 2a & -(\rho a + b) & 2b/\gamma \\ 0 & \rho a & -2b/\gamma \end{pmatrix},$$

$$\Delta_3 = \begin{pmatrix} \mu_0 & 0 & 0 \\ 0 & \mu_1 & 0 \\ 0 & 0 & \mu_2 \end{pmatrix}, \quad \Omega_3 = \delta \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

1.2 概率分布

化学主方程很好地刻画了系统的随机行为,已有相关文献对主方程在生化反应系统中的应用进行了分析和研究^[23-25],下面将应用概率-生产函数的方法求解主方程(4).考虑求静态解,通过引入下面的概率生成函数:

$$G_i(z) = \sum_{m=0}^{\infty} z^m P_i(m) \quad (i=0, 1, 2) \quad G(z) = \sum_{m=0}^{\infty} z^m P(m),$$

其中总生成函数 $G = G_0 + G_1 + G_2$.

把方程(4)转化为下面的偏微分方程

$$T_3 G - (z-1) \Delta_3 \frac{\partial}{\partial z} G + (z-1) \Omega_3 G = 0, \quad (5)$$

其中 $G = (G_0, G_1, G_2)^T$.

如果能够求得方程(5)的解析解,根据下面的关系

$$P_i(m) = \frac{1}{m!} \partial_z^m G_i(z) \Big|_{z=0} \quad P(m) = \frac{1}{m!} \partial_z^m G(z) \Big|_{z=0}, \quad (6)$$

则概率分布 $P_i(m)$ 、 $P(m)$ 的解析解也能够获得.为此首先计算跟转移矩阵 T_3 和它的余矩阵有关的几个特征值.由于 T_3 是一个零列矩阵(即每列元素之和为0),易知其特征值可表示为 $0, -\alpha_1, -\alpha_2$, 其中

$$\alpha_{1,2} = \left[(\rho a + 2a + b + 2b/\gamma) \pm \sqrt{(\rho a + 2a + b + 2b/\gamma)^2 - 8(\rho a^2 + 2ab/\gamma + b^2/\gamma)} \right] / 2, \quad (7)$$

另外,假设转移矩阵 T_3 的余矩阵 M_{ii} (即划去 T_3 第 i 行和第 i 列后剩下的矩阵)的特征值可表示为 $-\beta_1^{(i)}, -\beta_2^{(i)} \quad (i=1, 2, 3)$, 则易知

$$\begin{aligned} \beta_{1,2}^{(1)} &= \frac{(\rho a + b + \frac{2b}{\gamma}) \pm \sqrt{(\rho a + b + \frac{2b}{\gamma})^2 - \frac{8b^2}{\gamma}}}{2}, \\ \beta_1^{(2)} &= 2a, \quad \beta_2^{(2)} = 2b/\gamma, \quad \beta_1^{(3)} = \\ &= \frac{(\rho a + 2a + b) \pm \sqrt{(\rho a + 2a + b)^2 - 8\rho a^2}}{2}. \end{aligned} \quad (8)$$

注意到(7)~(8)式中的参数 a 和 b 都是规范化了的,即 $a/\delta \rightarrow a, b/\delta \rightarrow b$.此外,为方便讨论,记 $\mu_i = \mu_0 + \varepsilon_i \quad (i=1, 2)$ 并对其进行规范化处理,即 $\mu_i/\delta \rightarrow \mu_i \quad (i=0, 1, 2)$, $\varepsilon_i/\delta \rightarrow \varepsilon_i \quad (i=1, 2)$.然后应用概率-生成函数的方法求解生化主方程(4),从而获得 mRNA 的总概率分布(详细推导过程见附录)

$$P(m) = e^{-\mu_0} \left(\sum_{n=0}^m \frac{1}{n!} \frac{\mu_0^{m-n}}{(m-n)!} \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} (-1)^{k-n} b_k \right), \quad (9)$$

这里 $b_0 = 1, b_k = \frac{\varepsilon_1}{k} a_{k-1}^{(1)} + \frac{\varepsilon_2}{k} a_{k-1}^{(2)} \quad k=1, 2, \dots$, 并且 $a_k^{(1)}$ 和 $a_k^{(2)}$ 满足下面的迭代关系

$$\begin{pmatrix} a_k^{(1)} \\ a_k^{(2)} \end{pmatrix} = \frac{1}{k(k+\alpha_1)(k+\alpha_2)} \cdot \begin{pmatrix} \varepsilon_1(k+\beta_1^{(2)})(k+\beta_2^{(2)}) & \rho a \varepsilon_2(k+2a) \\ 2b \varepsilon_1(k+2a)/\gamma & \varepsilon_2(k+\beta_1^{(3)})(k+\beta_2^{(3)}) \end{pmatrix} \begin{pmatrix} a_{k-1}^{(1)} \\ a_{k-1}^{(2)} \end{pmatrix} \quad (10)$$

其中初值满足

$$a_0^{(i-1)} = \beta_1^{(i)} \beta_2^{(i)} / \alpha_1 \alpha_2 \quad i=1, 2, 3. \quad (11)$$

此外,还可获得3个状态的部分概率分布

$$P_i(m) = e^{-\mu_0} \left(\sum_{n=0}^m \frac{1}{n!} \frac{\mu_0^{m-n}}{(m-n)!} \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} (-1)^{k-n} a_k^{(i)} \right), \quad (12)$$

$$\text{这里 } a_k^{(0)} = \frac{\varepsilon_1 b(k\gamma + 2b) a_{k-1}^{(1)} + 2\varepsilon_2 b^2 a_{k-1}^{(2)}}{\gamma k(k+\alpha_1)(k+\alpha_2)}.$$

虽然(9)式和(12)式形式复杂,但是却提供了完整的随机信息.此外,根据下面的关系

$$\langle a_p \rangle = G_1(1) + G_2(1), \quad \langle m \rangle = G'(1),$$

其中 $G_1(1) = a_0^{(1)}, G_2(1) = a_0^{(2)}, G'(1) = \mu_0 + \varepsilon_1 a_0^{(1)} + \varepsilon_2 a_0^{(2)}$, 容易获得启动子活性和 mRNA 平均水平的显示表达

$$\langle a_p \rangle = \frac{\sigma a^2 + 2ab}{\sigma a^2 + 2ab + b^2}, \quad \langle m \rangle = \mu_0 + \frac{2ab\varepsilon_1 + \sigma a^2 \varepsilon_2}{\sigma a^2 + 2ab + b^2}. \quad (13)$$

方程(13)表明当总作用效能 $\sigma = \rho\gamma$ 和其它参数相同时,3种机制下的启动子活性和 mRNA 的平均表达水平都相同.

1.3 能量消耗公式

这节给出如图1所示的环路中能量消耗的计算公式.为方便,记

$$\omega_{01} = 2a, \quad \omega_{12} = \rho a, \quad \omega_{10} = b, \quad \omega_{21} = 2b/\gamma. \quad (14)$$

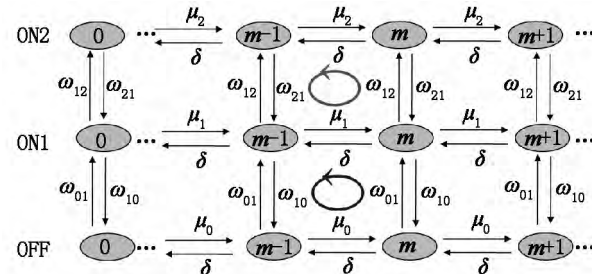


图2 环路的马氏过程示意图

图2所表示的是环路的马氏过程示意图,由图2看到在 OFF-ON1、ON1-ON2 的切换导致了2类不同的环流($\mu_1\delta$ 环和 $\mu_0\delta$ 环).接下来将指出:3个不同的转录速率是影响系统能量消耗的关键动力学参数.众所周知,在细致平衡系统中,反应环路的切换率遵循热力学关系.对本文的多 ON 基因系统,有如下的细致平衡条件

$$\mu_0 \omega_{01} \delta \omega_{10} = \mu_1 \omega_{10} \delta \omega_{01}, \mu_1 \omega_{12} \delta \omega_{21} = \mu_2 \omega_{21} \delta \omega_{12}. \quad (15)$$

把(14)式代入(15)式中,发现细致平衡条件变为 $\mu_0 = \mu_1 = \mu_2$. 根据先前的假设 $\mu_0 < \mu_1 < \mu_2$,表明环路的细致平衡被打破,系统远离平衡态,从而导致能量消耗. 图2中 $\mu_1 \rightarrow \delta$ 和 $\mu_0 \rightarrow \delta$ 的箭头表示系统存在2类能够导致熵产生和能量消耗的非零环流. 熟知,非平衡系统能够产生非零的熵生成率,用它可以刻画系统为了维持非平衡定态所消耗的能量. 因此,可利用熵生成率来计算系统中的能量消耗. 为此,引进下面的熵生成率公式^[10-26]

$$EP = \sum_{\theta, \theta'} P(\theta) W(\theta, \theta') \log \frac{W(\theta, \theta')}{W(\theta', \theta)},$$

这里 $P(\theta)$ 表示处于 θ 状态的概率, $W(\theta, \theta')$ 表示从状态 θ 到状态 θ' 的转移概率. 特别地,对本文的多ON基因系统,熵生成率可写成下面的形式

$$EP = \sum_m \left\{ P_0(m) \left[\mu_0 \log \frac{\mu_0}{\delta(m+1)} + \delta m \log \frac{\delta m}{\mu_0} + \omega_{01} \log \frac{\omega_{01}}{\omega_{10}} \right] + P_1(m) \left[\mu_1 \log \frac{\mu_1}{\delta(m+1)} + \delta m \log \frac{\delta m}{\mu_1} + \omega_{12} \log \frac{\omega_{12}}{\omega_{21}} + \omega_{10} \log \frac{\omega_{10}}{\omega_{01}} \right] + P_2(m) \left[\mu_2 \log \frac{\mu_2}{\delta(m+1)} + \delta m \log \frac{\delta m}{\mu_2} + \omega_{21} \log \frac{\omega_{21}}{\omega_{12}} \right] \right\}.$$

这里 $P_0(m)$ 、 $P_1(m)$ 以及 $P_2(m)$ 分别表示非

活性态I、活性态 A_1 和 A_2 处的静态概率分布,其具体表达式见(9)和(12)式.

上面的分析表明:泄露率 μ_0 、2个活性转录率 μ_1 和 μ_2 是影响系统能量消耗(或熵生成率)的关键动力学参数. 在接下来的内容中,将探讨当这些关键动力学参数变化时,系统能量如何消耗.

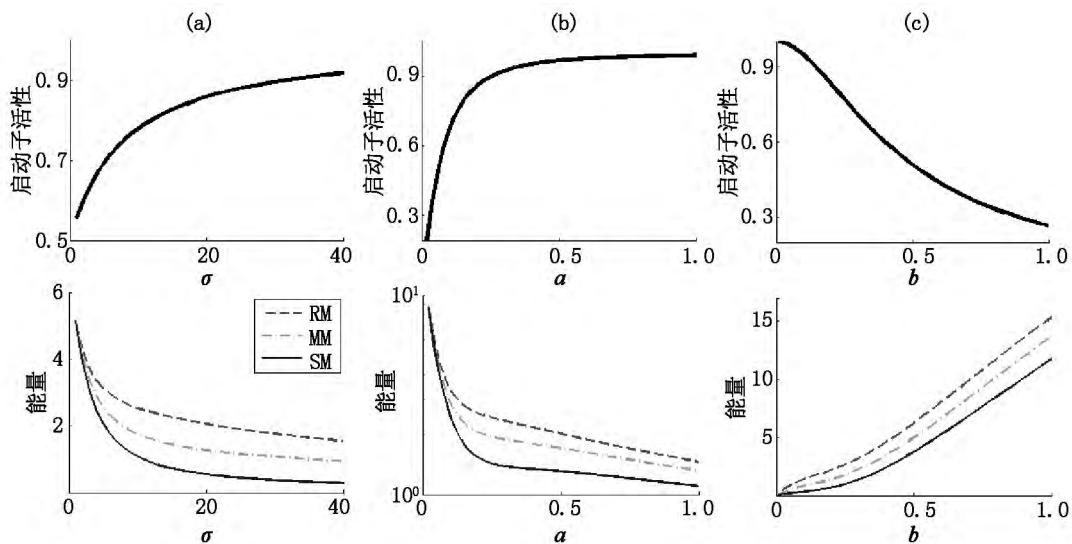
2 数值结果

下面将通过数值模拟来研究3种调控机制下的能量消耗,主要关注当上述关键动力学参数变化时,3种机制下的启动子活性与能量消耗之间的关系.

2.1 活性转录率的变化导致启动子活性和能量消耗呈反比关系

为了定性地说明活性转录率的变化对启动子活性和能量消耗的影响,保持mRNA的平均表达水平固定,同时变化活性转录率与影响启动子活性的3个参数:总用效能 σ 、结合率 a 以及解离率 b . 此外,定义一个新的参数 $\Delta\mu = \mu_2 - \mu_1$ 表示2个活性转录率的差,并考虑2种情形:(i)变化活性转录率 μ_1 同时固定 $\Delta\mu$; (ii)变化 $\Delta\mu$ 同时固定 μ_1 .

图3展示的是同时变化活性转录率和启动子的切换率(包括总作用效能 σ 、结合率 a 以及解离率 b)对启动子活性与能量消耗的影响. 这里固定mRNA



注: 这里固定 mRNA 的平均水平 $\langle m \rangle = 22$, 变化 2 个活性转录率并固定 $\Delta\mu = 10$. 图中第 1 行的粗线表示 3 种机制下的启动子活性相同. 图中第 2 行 RM 虚线、MM 虚点线以及 SM 实线分别表示招募机制、混合机制以及稳定机制下的能量消耗.

图3 活性转录率变化时,启动子切换对3种机制下的启动子活性和能量消耗的影响

的平均水平 $\langle m \rangle = 22$, 2 个活性转录率之差 $\Delta\mu = 10$. 子图中的参数分别为 (a) $a = 0.1$, $b = 0.2$; (b) $b = 0.2$, $\sigma = 4$; (c) $a = 0.1$, $\sigma = 16$; 其它参数为 $\mu_0 = 1$, $\delta = 0.4$. 从图3中可以看到:3种机制下的启

动子活性相同(图中第1行);但是能量消耗有不同的特征:招募机制下的能量消耗最多而稳定机制下的能量消耗最少(图中第2行). 此外,结合图3(a)~(c),可看到启动子活性与能量消耗之间有一个

反比例的关系: 启动子活性越高(低), 能量消耗越少(多). 另外, 注意到这些定性结果的获得并不依赖于所选取的参数.

为了更清晰地说明活性转录率的效应对启动子活性与能量消耗关系的影响, 下面考虑上述第 2 种情形, 即变化 $\Delta\mu$ 同时固定 μ_1 . 图 4 展示的是该情形下的数值结果. 这里固定 $\langle m \rangle = 20$ $\mu_0 = 1$ $\mu_1 =$

10 $\delta = 0.5$, 子图中参数为 (a) $b = 0.6$ $\sigma = 4$; (b) $a = 0.2$ $\sigma = 4$; (c) $a = 0.2$ $b = 0.6$. 从图 4 中可看到随着活性转录率的增加, 启动子活性降低, 能量消耗增加, 再次证实了当活性转录率变化时, 启动子活性与能量消耗之间的反比例关系. 此外, 图 4 还证实了上面的定性结论, 即招募机制和稳定机制分别有最大和最小的能量消耗.

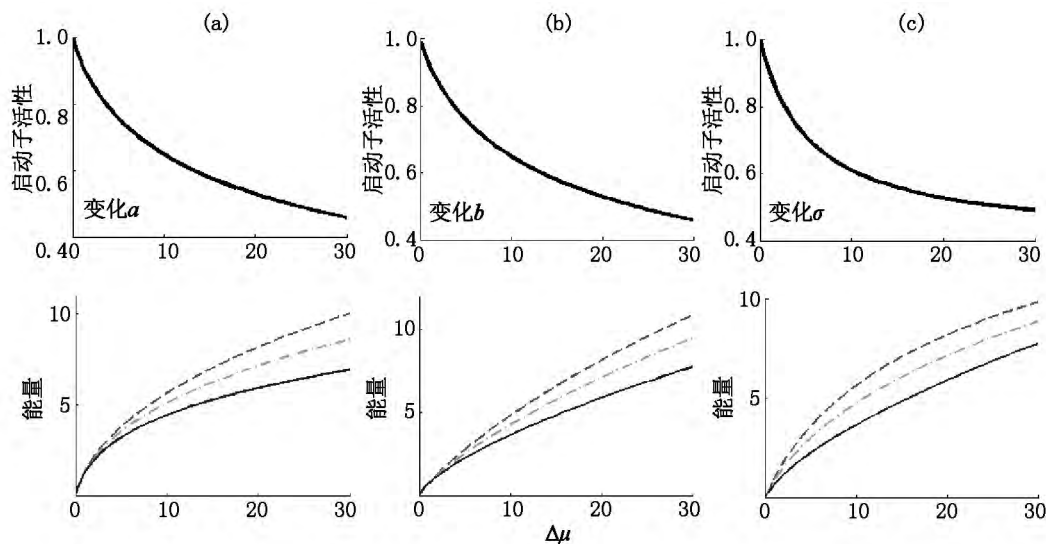


图 4 活性转录率之差 $\Delta\mu$ 对 3 种机制下的启动子活性和能量消耗的影响

2.2 泄露率的变化导致启动子活性和能量消耗呈正比关系

接下来, 将考察启动子泄露与能量消耗的关系. 图 5 展示了同时变化启动子泄露率和 3 个切换率对启动子活性和能量消耗的影响. 这里, 固定 $\langle m \rangle = 30$ $\mu_1 = 15$ $\mu_2 = 20$ $\delta = 0.5$, 子图中的参数为 (a) $a = 0.1$ $b = 0.3$; (b) $b = 3$ $\sigma = 16$; (c) $a = 0.1$,

$\sigma = 16$. 从图 5 中可看到启动子的活性与能量消耗之间是一个正比的关系: 启动子活性越高(低), 能量消耗越多(少). 这与前面活性转录率变化时的反比关系是不同的. 此外, 招募机制具有最大的能量消耗, 而稳定机制具有最小的能量消耗, 这一定性结论在泄露率变化时仍然成立.

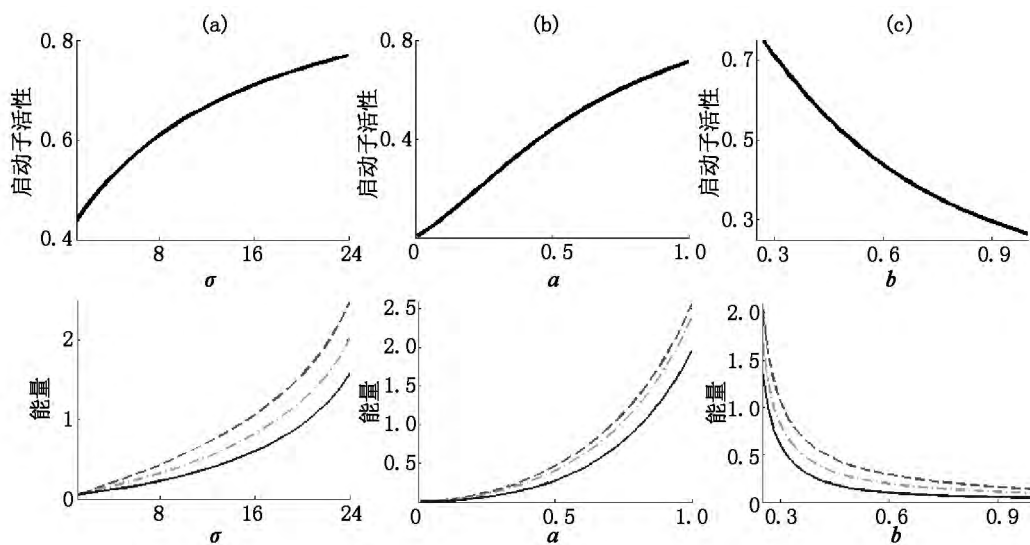


图 5 泄露率变化时, 启动子切换对 3 种机制下的启动子活性和能量消耗的影响

类似地, 为了说明启动子泄露的效应对启动子活性与能量消耗关系的影响, 在图 6 中画了启动子

活性和能量消耗关于泄露率的依赖曲线. 从图 6 中可看到: 随着泄露率的增加, 启动子活性降低, 能量

消耗减少.这也再次证实了图5中的结论,即泄露率变化时,启动子活性与能量消耗之间的正比关系.此外,3种机制下的能量消耗大小关系仍然成立.这

里,子图中的参数分别为(a) $b = 0.6$ $\sigma = 16$; (b) $a = 0.2$ $\sigma = 16$; (c) $a = 0.2$ $b = 0.6$.图6中其它参数与图5中的相同.

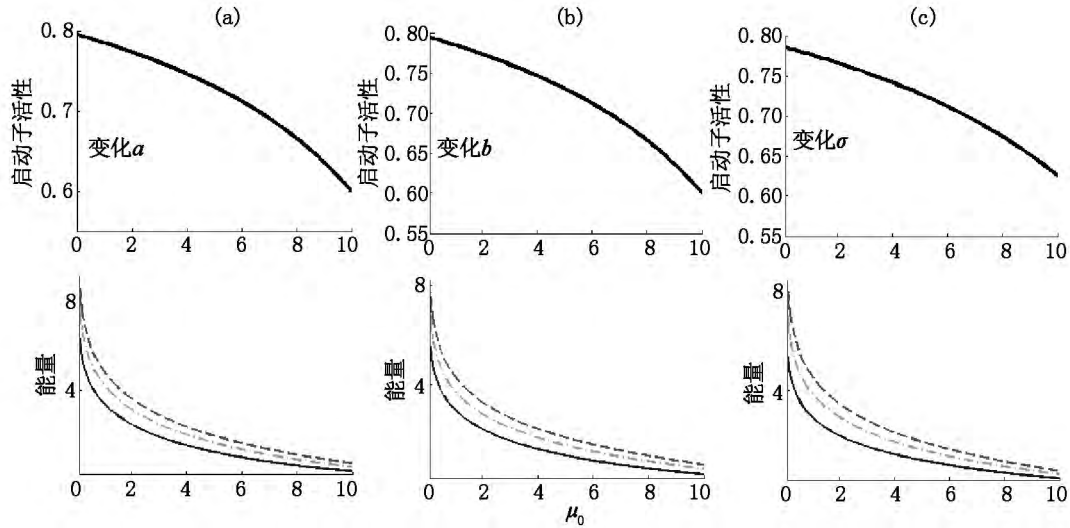


图6 启动子泄露对3种机制下的启动子活性和能量消耗的影响

综合2.1和2.2中的分析,发现转录因子的3种调控机制对基因表达过程中的能量消耗影响是不同的:招募机制下的能量消耗最多,而稳定机制下能量消耗最少.此外,启动子泄露也影响基因表达过程中的能量消耗.因此,基于分子生物学中能量消耗应遵循最小化准则^[14],本文的结果从理论上预测了一种较为理想的基因调控模型,应该是结合启动子泄露和转录因子的稳定功能的基因表达模型.

3 结论与讨论

近年来,随着单分子测量技术的发展,转录水平上的基因表达已成为一个热门的研究方向.转录过程是非常复杂的,它包括一系列的生化过程,如转录因子与聚合酶之间的作用、启动子在不同状态间的切换等.先前的研究主要集中在转录、启动子切换等重要事件对表达动力学(例如分子噪声)的影响.但是很少有研究关注基因表达过程中的能量消耗,特别是还没有文献关注转录因子的调控作用如何影响基因表达过程中的能量消耗.本文分析了一个在启动子区域有2个调控位点,并且启动子有一个非活性态和2个活性态的基因调控系统,研究了转录因子的作用机制对细胞能量消耗的影响,主要探讨了3种调控机制下的启动子活性与能量消耗的关系研究,获得了一些定性的结果:当mRNA平均水平相同时,转录因子招募机制下的基因表达消耗能量最多,稳定机制下的基因表达消耗能量最少;活性转录率的变化导致启动子活性与能量消耗之间呈一个反比的关系;而非活性转录率(或泄露率)的变化导致

启动子活性与能量消耗之间呈一个正比的关系.这些定性的结果从能量的观点表明了转录因子的作用在基因表达中起着非常重要的作用,控制着基因表达和细胞生物功能实现需要的能量消耗.

需要指出的是,本文的模型仅考虑了基因表达过程中的一些基本事件,忽略了某些细化和更为复杂的生化过程,如多部过程的转录和降解、细胞内外部信号的调控以及反馈调控等.这些过程或因素也会影响基因表达过程的效率和能量消耗,这有待于进一步的研究.

4 附录

静态概率分布的导出

为了从方程(5)中推导出静态概率分布,引进新变量 $s = z - 1$,并作变换 $F_i(s) = e^{-\mu_0 s} G_i(s)$ ($i = 0, 1, 2$), $F(s) = e^{-\mu_0 s} G(s)$,则方程(5)变为

$$\begin{aligned} -2aF_0 + bF_1 - sF_0'(s) &= 0, \\ 2aF_0 - (\rho a + b)F_1 + \frac{2b}{\gamma}F_2 + \varepsilon_1 sF_1 - sF_1'(s) &= 0, \\ \rho aF_1 - \frac{2b}{\gamma}F_2 + \varepsilon_2 sF_2 - sF_2'(s) &= 0. \end{aligned} \quad (A1)$$

展开函数

$$F_i(s) = \sum_{k=0}^{\infty} a_k^{(i)} s^k, \quad F(s) = \sum_{k=0}^{\infty} b_k s^k, \quad (A2)$$

把(A2)代入(A1)中,得

$$\begin{aligned} a_k &= \frac{(kI_3 - T_3)^*}{\det(kI_3 - T_3)} E_3 a_{k-1}, \quad b_k = \frac{\varepsilon_1}{k} a_{k-1}^{(1)} + \frac{\varepsilon_2}{k} a_{k-1}^{(2)}, \\ k &= 1, 2, \dots \end{aligned} \quad (A3)$$

其中列向量 $\mathbf{a}_k = (a_k^{(0)} \ a_k^{(1)} \ a_k^{(2)})^T$, 矩阵 $\mathbf{E}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \varepsilon_1 & 0 \\ 0 & 0 & \varepsilon_2 \end{pmatrix}$ 如果能够求得初值 $a_0^{(i)}$, 进而求出所

有的 $a_k^{(i)}$, 则 $b_k = \sum_{i=0}^2 a_k^{(i)}$ 和生成函数

$$G(z) = \left(\sum_{k=0}^{\infty} b_k (z-1)^k \right) e^{\mu_0(z-1)},$$

$$G_i(z) = \left(\sum_{k=0}^{\infty} a_k^{(i)} (z-1)^k \right) e^{\mu_0(z-1)}$$

也能够确定. 然后根据概率分布于生成函数的关系 (6) 获得平衡态的概率分布. 因此, 问题的关键是求解 \mathbf{a}_0 . 为此, 首先定义矩阵 \mathbf{T}_3 和余矩阵 \mathbf{M}_{ii} 的特征多项式分别为

$$\begin{aligned} f_{T_3}(\lambda) &= \det(\lambda \mathbf{I}_3 - \mathbf{T}_3) = \lambda(\lambda + \alpha_1)(\lambda + \alpha_2), \\ f_{M_{ii}}(\lambda) &= \det(\lambda \mathbf{I}_2 - \mathbf{M}_{ii}) = (\lambda + \beta_1^{(i)})(\lambda + \beta_2^{(i)}). \end{aligned} \quad (\text{A4})$$

易知 $\det(\mathbf{T}_3) = 0$, $\det(\mathbf{M}_{kk}) = \beta_1^{(k)}\beta_2^{(k)}$, $\text{rank}(\mathbf{T}_3) = 2$, $\mathbf{T}_3^* \mathbf{T}_3 = \det(\mathbf{T}_3) \mathbf{I}_3 = 0$, 并且矩阵 \mathbf{T}_3 的零空间是 1 维的. 利用矩阵 \mathbf{T}_3 的 Laplace 公式, 可知

$$\mathbf{T}_3 \begin{pmatrix} \det(\mathbf{M}_{11}) \\ \det(\mathbf{M}_{22}) \\ \det(\mathbf{M}_{33}) \end{pmatrix} = 0. \quad (\text{A5})$$

根据概率的保守性条件, 有

$$b_0 = \sum_{i=0}^2 a_0^{(i)} = 1, \mathbf{T}_3(0) \mathbf{a}_0 = 0. \quad (\text{A6})$$

结合 (A5) 和 (A6), 可得

$$\begin{aligned} a_0^{(i-1)} &= \det(\mathbf{M}_{ii}) \left(\sum_{i=1}^3 \det \mathbf{M}_{ii} \right)^{-1} = \\ &\beta_1^{(i)} \beta_2^{(i)} \left(\sum_{i=1}^3 \beta_1^{(i)} \beta_2^{(i)} \right)^{-1}. \end{aligned} \quad (\text{A7})$$

此外, 通过简单计算, 得

$$df_{T_3}(\lambda) / d\lambda \big|_{\lambda=0} = \alpha_1 \alpha_2. \quad (\text{A8})$$

另一方面, 根据 Jacobi 公式有

$$\begin{aligned} \frac{df_{T_3}(\lambda)}{d\lambda} \bigg|_{\lambda=0} &= \frac{d \det(\lambda \mathbf{I}_3 - \mathbf{T}_3)}{d\lambda} \bigg|_{\lambda=0} = \\ \text{tr} \left((\lambda \mathbf{I}_3 - \mathbf{T}_3)^* \frac{d \det(\lambda \mathbf{I}_3 - \mathbf{T}_3)}{d\lambda} \right) \bigg|_{\lambda=0} &= \\ \sum_{i=1}^3 \det(\mathbf{M}_{ii}) &= \sum_{i=1}^3 \beta_1^{(i)} \beta_2^{(i)}. \end{aligned} \quad (\text{A9})$$

联立 (A7) ~ (A9), 从而获得本文中的 (11) 式.

接下来推导 mRNA 的静态分布. 首先, 定义矩阵 $\mathbf{H} = (\pi_{ij})_{3 \times 3} = (k \mathbf{I}_3 - \mathbf{T}_3)^*$, 将其代入 (A3), 得

$$\begin{pmatrix} a_k^{(1)} \\ a_k^{(2)} \end{pmatrix} = \frac{\varepsilon_1 a_{k-1}^{(1)}}{\det(k \mathbf{I}_3 - \mathbf{T}_3)} \begin{pmatrix} \pi_{22} \\ \pi_{32} \end{pmatrix} + \frac{\varepsilon_2 a_{k-1}^{(2)}}{\det(k \mathbf{I}_3 - \mathbf{T}_3)} \begin{pmatrix} \pi_{23} \\ \pi_{33} \end{pmatrix}. \quad (\text{A10})$$

注意到

$$\pi_{ij} = (-1)^{i+j} \det(k \mathbf{I}_3 - \mathbf{M}_{ij}), \quad (\text{A11})$$

其中 \mathbf{M}_{ij} 是矩阵 \mathbf{T}_3 的元素 t_{ij} 的余矩阵, 通过简单计算, 可得

$$\begin{aligned} \pi_{12} &= b(k + 2b/\gamma), \quad \pi_{13} = 2b^2/\gamma, \\ \pi_{22} &= (k + \beta_1^{(2)})(k + \beta_2^{(2)}), \quad \pi_{23} = \rho a(k + 2a), \\ \pi_{32} &= \frac{2b}{\gamma}(k + 2a), \quad \pi_{33} = (k + \beta_1^{(3)})(k + \beta_2^{(3)}). \end{aligned} \quad (\text{A12})$$

又从方程 (A4), 可得

$$\det(k \mathbf{I}_3 - \mathbf{T}_3) = k(k + \alpha_1)(k + \alpha_2). \quad (\text{A13})$$

联立方程 (A10) ~ (A13), 可获得本文中 (10) 式. 最后, 利用本文中概率分布与生成函数的关系 (6), 可得本文中总的静态概率分布 (9) 和 3 个状态的部分概率分布 (12).

5 参考文献

- [1] Jülicher F, Ajdari A, Prost J. Modeling molecular motors [J]. Rev Mod Phys, 1997, 69: 1269.
- [2] Tsai T Y. Robust, tunable biological oscillations from inter-linked positive and negative feedback loops [J]. Science, 2008, 321: 126-129.
- [3] Murugan A, Huse D A, Leibler S. Speed, dissipation, and error in kinetic proofreading [J]. Proc Natl Acad Sci U S A, 2012, 109: 12304-12309.
- [4] Barton J P, Sontag E D. The energy costs of insulators in biochemical networks [J]. Biophys J, 2013, 104: 1380-1390.
- [5] Lang A H, Fisher C K, Mora T, et al. Thermodynamics of statistical inference by cells [J]. Phys Rev Lett, 2014, 113: 148103.
- [6] Qian Hong. Phosphorylation energy hypothesis: open chemical systems and their biological functions [J]. Annu Rev Phys Chem, 2007, 58: 113-142.
- [7] Lan Ganhui, Sartori P, Neumann S, et al. The energy-speed-accuracy trade-off in sensory adaption [J]. Nat Phys, 2012, 8: 422-428.
- [8] Lan Ganhui, Tu Yuhai. The cost of sensitive response and accurate adaptation in networks with an incoherent type-I feed-forward loop [J]. Interface, 2013, 10: 20130489.
- [9] Sartori P, Tu Yuhai. Free energy cost of reducing noise while maintaining a high sensitivity [J]. Phys Rev Lett, 2015, 115: 118102.
- [10] Mehta P, Schwab D J. Energetic costs of cellular computation [J]. Proc Natl Acad Sci U S A, 2012, 109: 17978-17982.
- [11] Huang Lifang, Yuan Zhanjiang, Yu Jianshe, et al. Fundamental principles of energy consumption for gene expression [J]. Chaos, 2015, 25: 123101.
- [12] Landauer R. Irreversibility and heat generation in the com-

- puting process [J]. IBM J Res Dev ,1961 ,5: 183-191.
- [13] Wagner A. Energy constraints on the evolution of gene expression [J]. Mol Bio Evol 2005 22: 1365-1374.
- [14] Schwanhäusser B. Global quantification of mammalian gene expression control [J]. Nature 2011 473: 337-342.
- [15] Pincus D. Keeping up with the omics: non-equilibrium models of gene regulation [J]. BMC Biol 2015 13: 9.
- [16] Tu Yuhai. The nonequilibrium mechanism for ultrasensitivity in a biological switch: sensing by Maxwell's demons [J]. Proc Natl Acad Sci U S A 2008 105: 11737-11741.
- [17] Saiz L ,Vilar J M. Stochastic dynamics of macromolecular-assembly networks [J]. Mol Syst Biol 2006 2: 0024.
- [18] Ptashne M A. Gene switch [M]. New York: Cold Spring Harbor Laboratory Press 2004.
- [19] Gutierrez P S ,Monteoliva D ,Diambra L. Role of cooperative binding on noise expression [J]. Phys Rev E 2009 , 80: 011914.
- [20] Gutierrez P S ,Monteoliva D ,Diambra L. Cooperative binding of transcription factors promotes bimodal gene expression response [J]. PloS one 2012 7: e44812.
- [21] Golding I ,Paulsson J ,Zawilski S M ,et al. Real-time kinetics of gene activity in individual bacteria [J]. Cell 2005 , 123: 1024-1036.
- [22] Spiller D G ,Wood C D ,Rand D A ,et al. Measurement of single-cell dynamics [J]. Nature 2010 465: 736-745.
- [23] 周天寿. 基因表达模型的研究进展: 概率分布 [J]. 江西师范大学学报: 自然科学版 2012 36(3): 221-229.
- [24] 周天寿. 概率主方程的研究综述 [J]. 江西师范大学学报: 自然科学版 2015 39(1): 1-6.
- [25] 雷锦誌. 生化反应系统的建模与分析 [J]. 江西师范大学学报: 自然科学版 2015 39(3): 230-239.
- [26] Thill P. Dynamic disorder and the energetic costs of information transduction [J]. J Chem Phys ,2014 , 141: 015102.

The Regulation-Mechanism-Dependent Energy Consumption

YANG Xiyan ,WU Yahao

(Department of Applied Mathematics ,Guangdong University of Finance ,Guangzhou Guangdong 510521 ,China)

Abstract: Gene expression is a dynamic non-equilibrium process ,inevitably consuming energy. Here ,a gene expression model at the transcription level is analyzed from the viewpoint of energy consumption. This model considers not only complexity of promoter structure (two regulatory sites) ,but also regulatory functions of interacting transcription factors (distinguished as three mechanisms: recruitment ,stabilization and mixture) . By mathematical modeling ,theoretical analysis and numerical simulation ,it is found that the recruitment mechanism maximizes energy consumption ,the stabilization mechanism minimizes energy consumption ,and the role of the mixture is between that of these mechanisms. In addition ,it is found that if the transcription rates at active states change ,then the relationship between promoter activity and energy consumption is inverse (i. e. ,the higher (lower) the promoter activity is ,the less (greater) is energy is consumed) ,but if the transcription rate (also called leakage rate) at the inactive state changes ,then this relationship is direct (i. e. ,the higher (lower) the promoter activity is ,the greater (less) is energy is consumed) . Altogether ,these results indicate that energy consumption in gene expression is regulation mechanism-dependent.

Key words: gene expression; energy consumption; regulation mechanism; promoter activity

(责任编辑: 王金莲)