

文章编号: 1000-5862(2016)01-0005-10

一个多 ON 基因系统的表达动力学分析

杨喜艳¹, 周天寿²

(1. 广东金融学院应用数学系, 广东 广州 510521; 2. 中山大学数学与计算科学学院, 广东 广州 510275)

摘要: 研究了一个转录水平上的基因表达模型, 它不仅考虑了启动子的复杂性, 而且考虑了相互作用转录因子的调控效果. 假定启动子区域有 2 个调控位点, 启动子有 1 个失活态和 2 个激活态(即 2 个不同的转录出口), 并考虑了绑定到不同调控位点的转录因子之间相互作用的 3 种代表性机制: 招募机制, 稳定机制和混合机制. 理论分析和数值模拟结果显示: 不同的转录出口能够导致 mRNA 分布峰的多样性; 3 种机制下的平均表达水平和平均爆发频率都相同, 但稳定机制诱导最大的表达噪声和平均爆发规模, 而招募机制诱导最小的表达噪声和平均爆发规模. 这些结果表明: 细胞表型的产生源是复杂的, 既与启动子结构有关也与调控机制有关.

关键词: 多 ON 基因系统; 总作用效能; mRNA 分布; 表达噪声; 爆发动力学

中图分类号: O 242; Q 141; Q 332 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2016.01.02

0 引言

具有相同遗传信息的群体细胞能够展示不同的表型, 这些多样化的表型对于细胞在复杂的环境中生存有着重要影响. 例如, 在突变的环境中, 差异化的表型能够增强群体细胞在波动环境中的适应性以及提高对外部刺激的鲁棒性^[1-3]; 在稳定的环境中, 单一化的表型能够最大限度地提高群体的适应性^[4-6]. 一般认为基因表达的随机性导致了群体细胞的多样性, 更确切地说, 基因表达噪声是导致具有相同遗传信息的细胞表型异质性的一个主要原因^[6-7]. 虽然近年来先进的实验技术允许对单细胞中的基因表达进行可视化观测^[8-9], 但是由于基因表达是个复杂的生化过程, 涉及到转录、翻译、降解、启动子状态之间的切换以及转录因子和聚合酶的作用^[10-12], 这些生化过程的具体细节还不清楚. 因此, 从理论上研究基因表达中的不同分子机制如何影响细胞表型的异质性是非常有意义的.

为了从理论上解释具有相同遗传信息的群体细胞展示出不同的表型, 2 种简单的基因表达模型已经被提出: 1) 描述构成式表达的单状态 ON 基因模型^[13-14], 即假设基因总是处于活性(ON)状态, 此时

基因产物(即 mRNA 或蛋白质)的数目服从泊松分布; 2) 描述爆发式表达的两状态 ON-OFF 基因模型^[15-17], 即假设基因启动子具有一个活性状态和一个非活性(OFF)状态, 此时基因产物的数目服从非泊松分布(由某个合流超几何函数描述). 为简单起见, 本文叫做合流超几何分布^[18-20]. 假如在这些基因模型的基础上再考虑自调控, 则基因产物的分布需要用更为复杂的若干合流超几何函数的组合来表示^[21-22]. 这些简单的基因模型已经很好地解释了某些实验现象, 但并不是全部. 另一方面, 一个基因可能有非常复杂的启动子结构. 例如, 实验上已经证实了真核细胞中的基因启动子状态能够达到 128 个^[23], 而且这些状态之间可以存在复杂的拓扑结构. 对于启动子具有复杂结构的基因模型, 很少有文献加以关注, 主要原因是研究具有挑战性. 本文将引入一个转录水平上的、启动子具有 2 个 ON 状态的基因模型, 并探讨细胞表型与基因表达之间的关系. 这种多 ON 基因模型可以在生物中找到原型, 如包括 3 个操作位点的溶源性噬菌体启动子结构^[24].

除了启动子结构之外, 转录因子的调控是影响基因表达的又一重要因素. 首先, 转录因子绑定到不同的 DNA 位点能够导致不同的启动子状态(这是启动子具有复杂几何结构的主要原因), 从而影响转录, 例如, 噬菌体 CI 蛋白和大肠杆菌 cAMP 受体

收稿日期: 2015-11-06

基金项目: 国家自然科学基金(91230204, 11005162, 11475273, 11526058, 2014CB964703), 2015 年广东省教育厅青年创新人才类(2015KQNCX103) 和 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)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.

蛋白与 RNA 聚合酶的作用能够增强转录^[25-26];其次,不同的调控方式能够导致不同的基因表达谱.例如,基因的非协作绑定能够诱导双峰表达^[27].很多的研究关注调控如何影响表达噪声的效应.例如,在均值不固定的情形下,增加负反馈强度会导致蛋白质噪声的增加^[28]或减少^[19],或出现一个 U 形结构^[29-30];在均值固定的情形下,增加反馈强度会导致蛋白质噪声的增加^[31]或减少^[32].总之,基因表达水平上的差异或是由转录因子绑定到不同的 DNA 位点引起,或是由其它不明的分子机制导致.尽管如此,转录因子的调控如何影响基因产物的分布(特别是在启动子具有多个 ON 的情形)是值得研究的.

近年来,大量的生物学实验表明:在细菌以及真核生物中,基因表达通常以爆发方式产生^[33-34],这里爆发方式指的是基因表达在一段时间内产生大量的基因产物,而随后基因会在一段时间内保持静默(即不表达).那么自然要问:基因为什么会选择这种方式表达,有何优势?一般认为爆发式基因表达是导致同源细胞表型多样性的主要原因^[2,34].例如,在对酵母菌的研究中发现,细胞群体的差异性源于启动子的慢切换以及爆发式表达^[35].尽管如此,转录因子的调控如何影响爆发动力学,进而影响群体细胞表型的变化(特别是在启动子具有多个活性状态的情形)仍然不清楚,需要进一步研究.本文将考虑 2 个调控位点的基因启动子以及启动子状态之间的切换具有快慢之分,并假设转录因子的调控效应直接影响启动子的切换率.

基于上面的文献分析,本文将引进一个新的基因模型,它既考虑了启动子结构的复杂性:启动子有 3 个状态,1 个 OFF 状态和 2 个 ON 状态;又考虑了转录因子的 3 种不同的调控机制:招募机制(re-recruitment mechanism, RM),即转录因子之间的相互作用增加了另一转录因子绑定到操作子的速率,但是并不影响转录因子从操作子解离的速率;稳定机制(stabilization mechanism, SM),即转录因子的相互作用减小了转录因子从操作子解离的速率,但是并不影响转录因子绑定到操作子的速率;混合机制(mixture mechanism, MM),即同时考虑了转录因子的招募和稳定功能.为了方便展示这 3 个机制对基因表达的效应,将引进用来刻画转录因子间相互作用的总作用效能^[36].虽然某些文献研究了 3 种机制下的表达噪声和双峰响应^[37-38],但是所涉及到的基因模型并没有考虑不同的转录速率(多个转录出口的主要特征),也没有考虑转录因子的相互作用对爆发动力学(表达差异性的主要源)的影响.

本文主要关注转录因子之间的相互作用对基因表达动力学的影响,并由此导致细胞多样性的产生.利用理论分析和数值模拟发现:转录因子的作用能够导致 mRNA 分布峰的多样化;3 种作用机制下的平均表达水平和平均爆发频率都相同,但稳定机制诱导最大的表达噪声和平均爆发规模,而招募机制诱导最小的表达噪声和平均爆发规模.总作用效能的增加或能够降低表达噪声或使之出现非单调性的结构;总作用效能的增加还能够导致平均爆发规模的增大但保持平均爆发频率不变.此外,启动子的慢切换能够导致更突出的爆发式表达和更加多样化的细胞表型.总之,转录因子之间的相互作用极大地丰富了基因表达动力学的内涵,可能是具有相同遗传信息的群体细胞产生不同表型的主要来源.

1 基因模型与分析结果

1.1 模型

考虑转录水平上具有 2 个绑定位点的基因调控系统,见图 1.这一系统假定基因有 3 个状态:1 个失活态 I 和 2 个激活态 A_1 和 A_2 .当转录因子绑定到调控位点时,在 RNA 聚合酶的作用下,能够导致基因产物(即 mRNA)的生成.同时,考虑激活型转录因

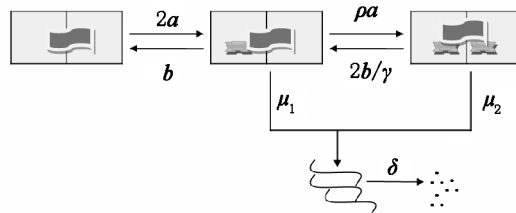
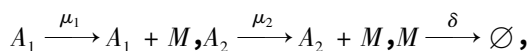
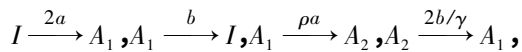


图 1 一个代表性的多 ON 基因模型示意图

子,并且其相互作用产生 2 种不同的功能:1) 招募功能,之前绑定到一个调控位点的第 1 个转录因子具有招募第 2 个转录因子绑定到第 2 个调控位点的能力;2) 稳定功能,即 2 个转录因子之间的相互作用使得其中一个转录因子离开绑定位点的解离速率更慢,因而更稳定.

为了引入要研究的模型,分别利用招募因子 ρ 和稳定因子 γ 来刻画这 2 种功能,见图 1.类似于文献^[1]的简化,合并转录复合物的形成和转录过程为单步,从而引入下列一套生化反应式



其中 a 和 b 分别表示转录因子的结合率和解离率,

μ_1 和 μ_2 是基因处于活性态的转录率, δ 是 mRNA 的降解率. 由于考虑的是激活型转录因子, 因此有 $\mu_2 > \mu_1$. 此外, 注意到由图 1 描述的基因系统包含了引言中提到的 3 种协作绑定机制. 特别地, 可用招募因子 ρ 和稳定因子 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, 为了方便, 记 $\sigma = \rho\gamma$ 表示总作用效能. 显然, 在 RM 中, $\sigma = \rho$, 在 SM 中, $\sigma = \gamma$ 以及在 MM 中, $\sigma = \rho\gamma$.

让 $P_0(m, t)$ 和 $P_i(m, t)$ ($i = 1, 2$) 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

$$\frac{dP(m, t)}{dt} = A_3 P(m, t) + D_3 (E - I) [nP(m, t)] + X_3 (E^{-1} - I) [P(m, t)]. \quad (1)$$

这里 I 是单位算子, E 和 E^{-1} 是平移算子 (或步长算子), 即对任意函数 g 和任意整数 n , 有 $E[f(n)] = f(n+1)$, $E^{-1}[f(n)] = f(n-1)$. 矩阵 A_3, X_3, D_3 分别称为启动子状态之间的转移矩阵、转录矩阵以及 mRNA 的降解矩阵, 其中

$$A_3 = \begin{pmatrix} -2a & b & 0 \\ 2a & -(\rho a + b) & 2b/\gamma \\ 0 & \rho a & -2b/\gamma \end{pmatrix},$$

$$X_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mu_1 & 0 \\ 0 & 0 & \mu_2 \end{pmatrix}, D_3 = \delta \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

1.2 概率分布

化学主方程很好地刻画了系统的随机行为, 因此, 求解生化主方程是非常有意义的, 这也是本文的主要理论工作之一. 近年来, 已有相关文献研究了主方程的进展及其在生化反应系统中的应用^[39-41]. 本文将应用概率-生成函数的方法求解主方程 (1). 首先, 计算与转移矩阵 A_3 及其余矩阵有关的几个特征值^[24]. 注意到 A_3 是一个 M 矩阵 (即每列元素之和为 0, 也称为零列矩阵), 故其特征值可表示为 0, $-\alpha_1, -\alpha_2$. 易知

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

另外, 对转移矩阵 A_3 的余矩阵 M_{ii} (即划去矩阵 A_3 的第 i 行和第 i 列后余下的矩阵), 设其特征值为 $-\beta_1^{(i)}, -\beta_2^{(i)}$ ($i = 1, 2, 3$), 通过简单计算, 可得

$$\beta_{1,2}^{(1)} = \left[(\rho a + b + 2b/\gamma) \pm \sqrt{(\rho a + b + 2b/\gamma)^2 - 8b^2/\gamma} \right] / 2, \beta_1^{(2)} = 2b/\gamma, \beta_{1,2}^{(3)} = \left[(\rho a + 2a + b) \pm \sqrt{(\rho a + 2a + b)^2 - 8\rho a^2} \right] / 2. \quad (3)$$

$$\sqrt{(\rho a + b + 2b/\gamma)^2 - 8b^2/\gamma} \Big] / 2, \beta_1^{(2)} = 2a,$$

$$\beta_2^{(2)} = 2b/\gamma, \beta_{1,2}^{(3)} = \left[(\rho a + 2a + b) \pm \sqrt{(\rho a + 2a + b)^2 - 8\rho a^2} \right] / 2. \quad (3)$$

需要指出的是, (2) ~ (3) 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, 为了讨论方便, 记 $\mu_2 = \mu_1 + \varepsilon$, 并对其进行规范化处理, 即 $\mu_1/\delta \rightarrow \mu_1, \varepsilon/\delta \rightarrow \varepsilon$. 本文后面内容所涉及的参数都是经过规范化的, 因此不再重复指出.

接下来, 利用概率-生成函数的方法求解主方程 (1), 并可获得 mRNA 的静态概率分布 (推导过程略)

$$P(m) = e^{-\mu_1} \left(\sum_{n=0}^m \frac{1}{n!} \frac{\mu_1^{m-n}}{(m-n)!} \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} (-1)^{k-n} b_k \right), \quad (4)$$

这里

$$b_0 = 1, b_k = -\frac{\mu_1}{k} a_{k-1}^{(0)} + \frac{\varepsilon}{k} a_{k-1}^{(2)}, k = 1, 2, \dots,$$

$a_k^{(0)}$ 和 $a_k^{(2)}$ 满足下面的迭代关系

$$\begin{pmatrix} a_k^{(0)} \\ a_k^{(2)} \end{pmatrix} = \frac{1}{k(k + \alpha_1)(k + \alpha_2)} \cdot \begin{pmatrix} -\mu_1(k + \beta_1^{(1)})(k + \beta_2^{(1)}) & 2b^2\varepsilon/\gamma \\ -2\rho a^2\mu_1 & \varepsilon(k + \beta_1^{(3)})(k + \beta_2^{(3)}) \end{pmatrix} \begin{pmatrix} a_{k-1}^{(0)} \\ a_{k-1}^{(2)} \end{pmatrix}, \quad (5)$$

其中初值满足

$$a_0^{(i-1)} = \beta_1^{(i)} \beta_2^{(i)} / \alpha_1 \alpha_2, i = 1, 2, 3. \quad (6)$$

1.3 平均水平和表达噪声

虽然解析分布 (4) 式为所研究的系统提供了全面的随机信息, 但是其形式复杂, 不方便理解系统的随机行为. 另一方面, 利用概率分布的随机指标如均值和方差来分析系统的随机行为比较方便, 并可获得一些直观的结果. 这里, 基于前面的概率-生成函数的方法, 给出计算 mRNA 平均和其噪声的一般公式

$$\langle m \rangle = G'(1), \eta_m^2 = \frac{1}{G'(1)} + \frac{G''(1)}{G'(1)^2} - 1.$$

特别地, 对所研究的多 ON 基因模型, 有

$$G'(1) = \mu_1 - \mu_1 a_0^{(0)} + \varepsilon a_0^{(2)},$$

$G''(1) = \mu_1^2 - 2\mu_1^2 a_0^{(0)} + 2\mu_1 \varepsilon a_0^{(2)} - \mu_1 a_1^{(0)} + \varepsilon a_1^{(2)}$, 其中 $a_0^{(0)}, a_0^{(2)}$ 由 (6) 式给出, 而 $a_1^{(0)}, a_1^{(2)}$ 可从 (5) 式获得.

这样, mRNA 的平均水平可显示地表达为

$$\langle m \rangle = \frac{\sigma a^2 (\mu_1 + \varepsilon) + 2ab\mu_1}{\sigma a^2 + 2ab + b^2}. \quad (7)$$

这一共识表明: 当总效能 $\sigma = \rho\gamma$ 固定时, 3 种协作机制下的平均水平相同. 此外, 同时变化 a 和 b 但保持比值 a/b 不变, 3 种协作机制下的平均水平也是相同的. 另一方面, 由于噪声强度的表达式比较复杂, 因此这里省略其具体表达式, 但在后面的内容中, 将采用数值模拟来展示出 3 种机制对表达噪声的影响.

1.4 爆发规模和爆发频率

单分子分辨率技术的实验表明, 在细菌和真核细胞中, 基因表达通常以爆发的形式发生^[33-34]. 爆发规模和爆发频率是刻画爆发动力学 2 个重要统计量. 在普通的 ON-OFF 模型中, 平均爆发规模和平均爆发频率的计算公式为^[34]

$$\langle BS \rangle = k_{\text{transcription}} \cdot \langle \tau_{\text{on}} \rangle, \langle BF \rangle = 1 / \langle \tau_{\text{off}} \rangle,$$

其中 $\langle \tau_{\text{on}} \rangle$ 和 $\langle \tau_{\text{off}} \rangle$ 分别表示基因在 ON 和 OFF 状态的平均驻留时间. 此外, 更复杂的多状态基因模型中 ON 和 OFF 状态的平均驻留时间公式也已经被导出^[34]

$$\langle \tau_{\text{on}} \rangle = u_K A_{10} (A_{11})^{-2} Q^{(1)}(0),$$

$$\langle \tau_{\text{off}} \rangle = u_L A_{01} (A_{00})^{-2} Q^{(0)}(0).$$

这里, 假设基因有 L 个活性态和 K 个失活态. $u_K = (1, 1, \dots, 1)$ 和 $u_L = (1, 1, \dots, 1)$ 分别表示 K 维和 L 维行向量. $A_{ij} (i, j = 0, 1)$ 是启动子转移矩阵 A 的分块矩阵. 特别地, $A_{11} (A_{00})$ 刻画启动子在活性态(失活态)之间的切换情况, 而矩阵 $A_{10} (A_{01})$ 刻画启动子从活性态(失活态)到失活态(活性态)之间的切换情况. 此外, $Q^{(1)}(0) = (Q_1^{(1)}(0), \dots, Q_L^{(1)}(0))^T$ 表示从 OFF 到 ON 或从 ON 到 ON 的初始生存概率; 类似地, $Q^{(0)}(0) = (Q_1^{(0)}(0), \dots, Q_K^{(0)}(0))^T$ 表示从 ON 到 OFF 或从 OFF 到 OFF 的初始生存概率. 注意到: 在 ON(或 OFF)状态的平均驻留时间等于所有 ON(或所有 OFF)状态的平均驻留时间之和

$$\langle \tau_{\text{on}} \rangle = \sum_{i=1}^L \langle \tau_{\text{on}}^{(i)} \rangle, \langle \tau_{\text{off}} \rangle = \sum_{j=1}^K \langle \tau_{\text{off}}^{(j)} \rangle,$$

其中

$$\langle \tau_{\text{on}}^{(i)} \rangle = u_K A_{10} (A_{11})^{-2} Q_i^{(1)}(0) e_i^{(1)}, i = 1, 2, \dots, L,$$

$$\langle \tau_{\text{off}}^{(j)} \rangle = u_L A_{01} (A_{00})^{-2} Q_j^{(0)}(0) e_j^{(0)}, j = 1, 2, \dots, K.$$

这里 $e_i^{(1)} (e_j^{(0)})$ 表示第 $i (j)$ 个元素是 1 的 $L (K)$ 维单位列向量. 因此, 对于启动子具有多个启动子状态的基因模型, 平均爆发规模和平均爆发频率的计算公式可表示为

$$\langle BS \rangle = \sum_{i=1}^L k_{\text{transcription}}^{(i)} \langle \tau_{\text{on}}^{(i)} \rangle, \langle BF \rangle = 1 / \left(\sum_{j=1}^K \langle \tau_{\text{off}}^{(j)} \rangle \right).$$

这里 $k_{\text{transcription}}^{(i)}$ 表示启动子驻留在第 i 个 ON 状态时的转录率.

根据上面的计算公式, 通过简单计算(具体推导过程略), 可获得基因系统中平均爆发规模和平均爆发频率的计算公式

$$\langle BS \rangle = \mu_1 \left(\frac{2}{b} + \frac{2\sigma a + \gamma b}{2b^2} \right) + \varepsilon \left(\frac{1}{b} + \frac{\sigma a + \gamma b}{2b^2} \right), \langle BF \rangle = 2a. \quad (8)$$

这些公式表明: 当总作用效能和其它参数固定时, 越大的稳定因子能够导致越大的平均爆发规模, 但是平均爆发频率不变. 本文将基于所获得的结果来研究 3 种协作机制中的爆发动力学.

2 转录因子间的相互作用对基因表达动力学的影响

在下面的分析中, 将基于上面获得的分析结果来考察转录因子间的相互作用对基因表达动力学的影响, 重点考察转录因子的招募和稳定机制分别对 mRNA 分布、表达噪声及爆发动力学的影响. 为了清楚地展示出结果, 首先定义 3 种基因切换模式: 慢切换, 即与降解率 δ 相比, 结合率 a 和解离率 b 是非常小的; 快切换, 即与降解率 δ 相比, 结合率 a 和解离率 b 是非常大的; 中等切换, 结合率 a 和解离率 b 与降解率 δ 具有相同的数量级. 此外, 还定义一种非对称切换, 此时结合率 a 和解离率 b 不是同一数量级的. 在混合机制中, 选取 $\rho = \gamma = \sqrt{\sigma}$. 需要指出的是这种选择并不影响本文的定性结果.

2.1 对概率分布的影响

通过数值模拟, 发现转录因子的作用能够导致多样化的 mRNA 分布. 图 2 展示了不同切换情形下的 mRNA 的概率分布, 图 2(a) ~ (c) 分别相应于慢切换、中等切换和快切换的情形. 这里固定总作用效能 $\sigma = 9$ 以及结合率 a 和解离率 b 的比率 $a/b = 2/3$. 在每个子图中, 转录因子的结合率和解离率分别为 (a) $a = 0.04, b = 0.06$; (b) $a = 0.4, b = 0.6$; (c) $a = 4, b = 6$, 其它参数为 $\mu_1 = 2, \varepsilon = 6, \delta = 0.4$. 根据公式(7)可知, 图 2 中每个子图的 mRNA 的平均水平是一样的. 由图 2 观察到: 对于慢切换, 概率分布在 SM 和 MM 情形都是 3 个峰的多峰, 而在 RM 情形是双峰, 并且都有一个靠近原点的峰. 对

于中等切换,概率分布在 RM 和 MM 情形是单峰,但是在 SM 情形变成了双峰且 2 个峰都是远离原点

的. 对于快切换,3 种机制下的概率分布均呈现出单峰形状.

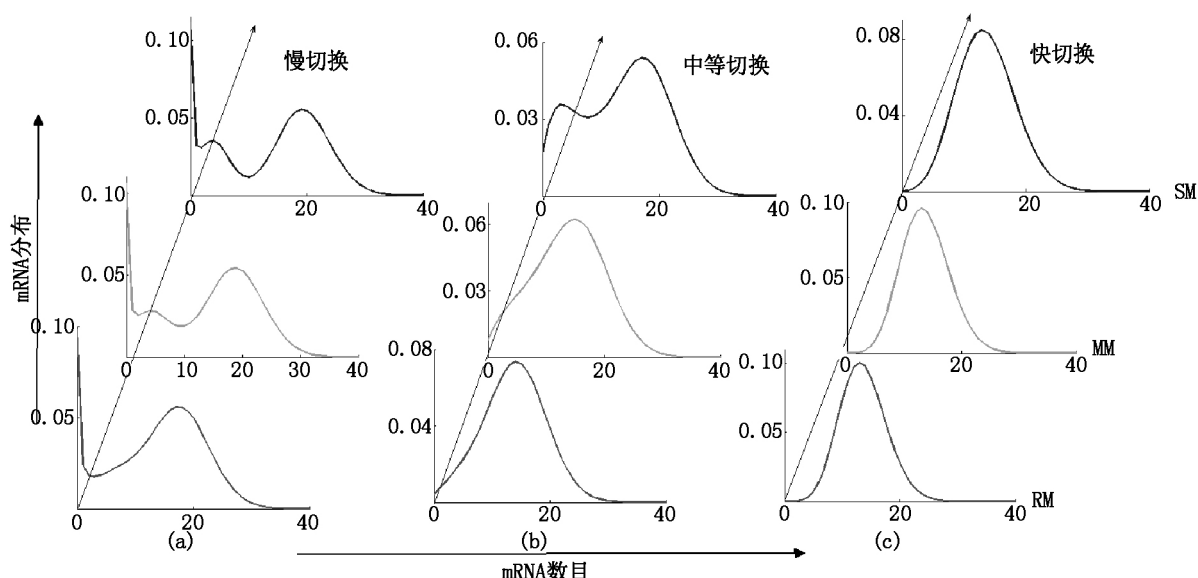


图 2 3 种基因切换下的 mRNA 概率分布

图 3 展示的是非对称切换下的概率分布,其中总效能固定在 $\sigma = 36$ 以及结合率 a 和解离率 b 的比率固定为 $a/b = 1/10$. 在每个子图中,转录因子的结合率和解离率分别为 (a) $a = 0.01, b = 0.1$; (b) $a = 0.1, b = 1$; (c) $a = 1, b = 10$, 其它参数与

图 2 中的相同. 由图 3 观察到: 随着结合率和解离率的同时增大,转录因子的作用能够导致概率分布从靠近原点的单峰变成靠近原点的单峰,以及远离原点的单峰.

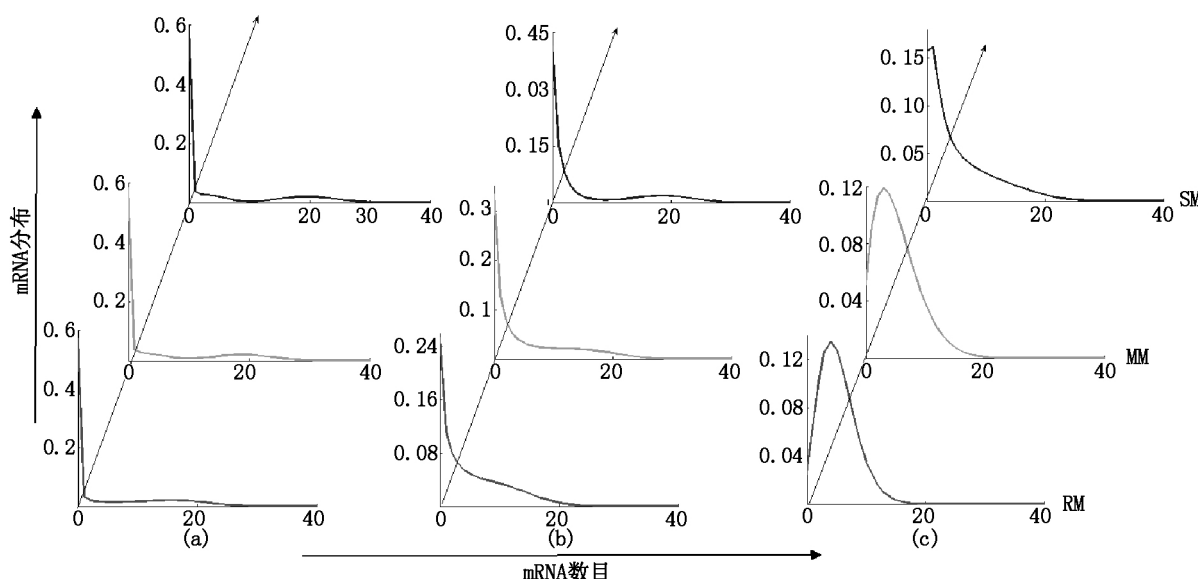


图 3 非对称切换下的 mRNA 概率分布

众所周知,基因产物的概率分布有着非常重要的生物学意义,即基因表达中的双峰或多峰分别代表了同源群体细胞中的不同表型^[4];表型之间的差异性影响着细胞生长的速率和对环境的适应性^[5,6]. 单峰刻画的是单一化的表型,这种表型有利于群体细胞在稳定的环境中生长;多峰刻画的是多样化的表型,这种表型能够增加细胞的适应性并有

利于群体细胞在波动的环境中生长. 因此,以上结果说明了转录因子间的相互作用为控制群体细胞表型提供了一种机制: 群体细胞不仅可以利用稳定机制诱导多峰来适应波动的生存环境,还可以利用招募机制诱导单峰来提高在稳定环境中的生存能力.

注意到在图 2 和图 3 中展示的是当 $a/b < 1$ 时的概率分布,事实上,当 $a/b > 1$ 时,转录因子间的

相互作用也能够导致 mRNA 分布的多样化. 特别是对多峰分布情形, 一个突出的特点是分子数目较大的峰占主导地位(这里省去了这种情形下的图形展示). 在下面的内容中, 也仅就 $a/b < 1$ 的情形进行讨论.

2.2 对表达噪声的影响

图 4 展示出不同切换情形下的表达噪声. 图 4 (a) 从左到右分别对应于慢切换、中等切换以及快切换的情形, 这里的参数与图 2 中的相同. 由图 4 (a) 观察到: 随着总作用效能的增大, 表达噪声总是

减小的. 图 4 (b) 是非对称切换下的 3 种调控机制情形, 这里的参数与图 3 中的相同. 观察到: 表达噪声还可以是总作用效能的非单调的函数(表达噪声有极大值). 由此可见, 转录因子间的相互作用和表达噪声之间的关系并不是唯一的, 既可以是单调递减的关系, 也可以是非单调的关系, 依赖于参数值的选择. 此外, 观察到 2 个子图的一个共同特点是 SM 有最大的表达噪声, MM 次之, RM 中有最小的表达噪声. 需要指出的是, 这是一个定性的结果, 并不依赖于参数值的选择.

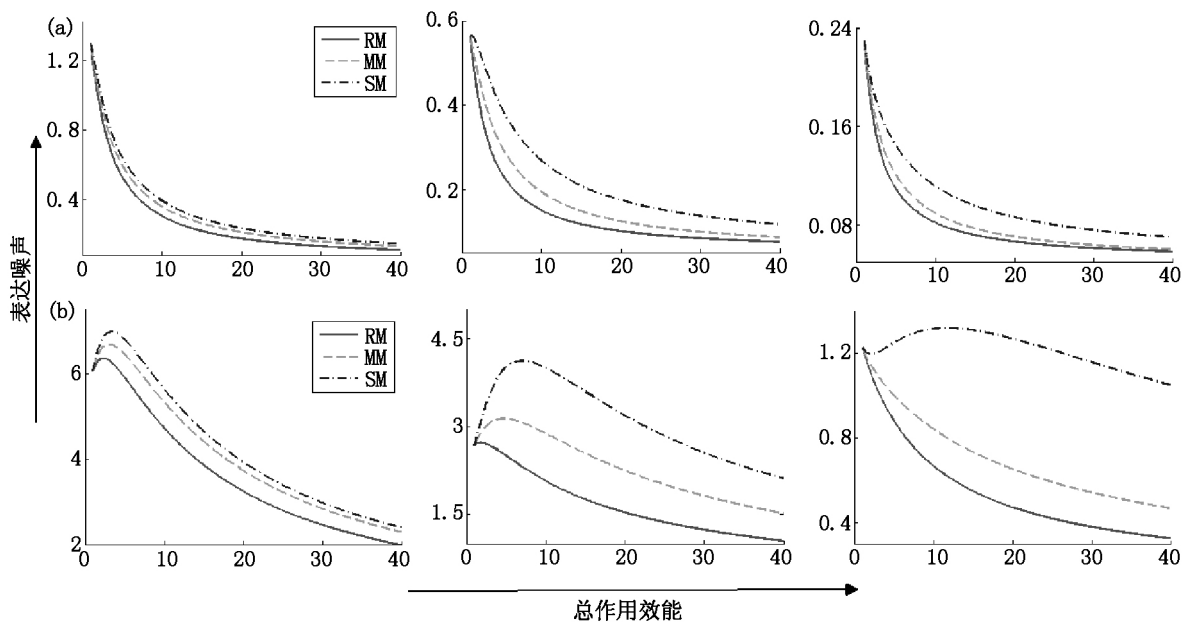


图 4 转录因子间的相互作用对表达噪声的影响

已有研究表明: 基因表达的噪声是促进群体细胞表型多样化的一种潜在机制^[42-43], 这里结合所获得的表达噪声和概率分布的结果进行进一步的分析. 由图 4 (a) 可看出, 当总作用效能从 0 到 40 变化时, 慢切换中噪声的变化幅度最大, 中等切换情形次之, 快切换中的噪声变化幅度最小. 相应地, 在概率分布图 2 中, 看到慢切换下有多达 3 个峰的概率分布出现, 中等切换中有 2 个峰出现, 快切换下都是单峰且概率分布的形状几乎没有变化. 另外, 对图 4 (b) 和图 3, 也可以进行类似的分析. 所有这些表明: 当总作用效能相同并且均值都相同时, 越大的表达噪声能够导致更加多样化的表型.

2.3 对爆发动力学的影响

大量的实验研究表明, 基因产物(mRNA 或蛋白质) 以爆发的方式生成^[32-34]. 虽然已有相关的文献研究了爆发规模和噪声之间的关系^[33,44], 但是转录因子间的相互作用如何影响爆发动力学还不清楚

下面的研究将表明转录因子间的相互作用对爆发动力学有不可忽略的影响, 它能够使爆发规模变大.

图 5 展示的是转录因子间的相互作用对平均爆发规模的影响, 图 5 (a) 从左到右分别对应于慢切换、中等切换和快切换的情形; 图 5 (b) 是非对称切换的 3 种调控机制情形. 可以看到平均爆发规模是总作用效能的一个线性单调递增函数, 并且 SM 和 RM 分别有最大和最小的平均爆发规模, MM 机制下的平均爆发规模介于 2 者之间. 此外, 发现每一行中子图形状一样, 并且纵轴呈现出以 1/10 的比率变化. 事实上, (8) 式中的平均爆发规模可改写为

$$\langle BS \rangle = \frac{1}{b} \left[2\mu_1 + \varepsilon + \frac{\sigma(2\mu_1 + \varepsilon)}{2} \cdot \frac{a}{b} + \frac{\gamma(\mu_1 + \varepsilon)}{2} \right],$$

可以看出, 当其它参数相同并固定比率 a/b 时, 3 种切换下的平均爆发规模的差异仅与因子 $1/b$ 有关. 显然, 越慢的解离速率能够导致越大的平均爆发规模. 此外, 从 (8) 式中可知, 越慢的结合率能够导致越大的平均爆发频率; 增加总作用效能并不改变平

均爆发频率. 这些结果表明, 转录因子间的相互作用对慢切换下的爆发动力学影响最大.

力学影响的动态分析结果(注: 这些结果是利用 Gillespie 随机模拟算法获得的^[43]), 这里仅展示 SM

图 6 展示的是转录因子间的相互作用对爆发动

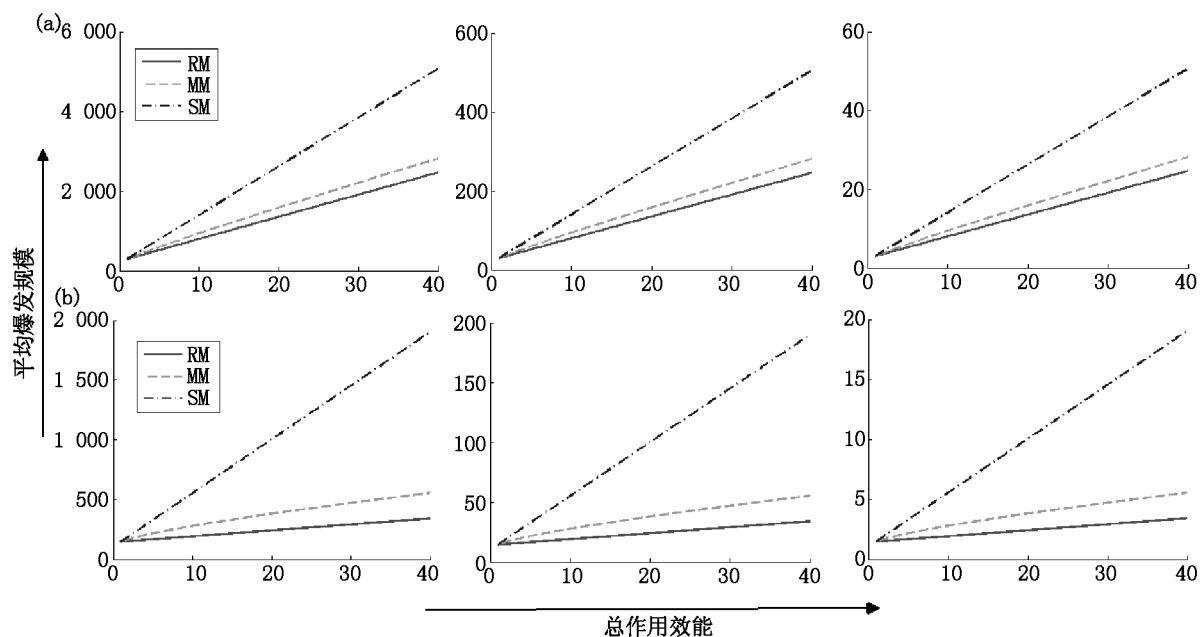


图 5 转录因子间的相互作用对平均爆发规模的影响

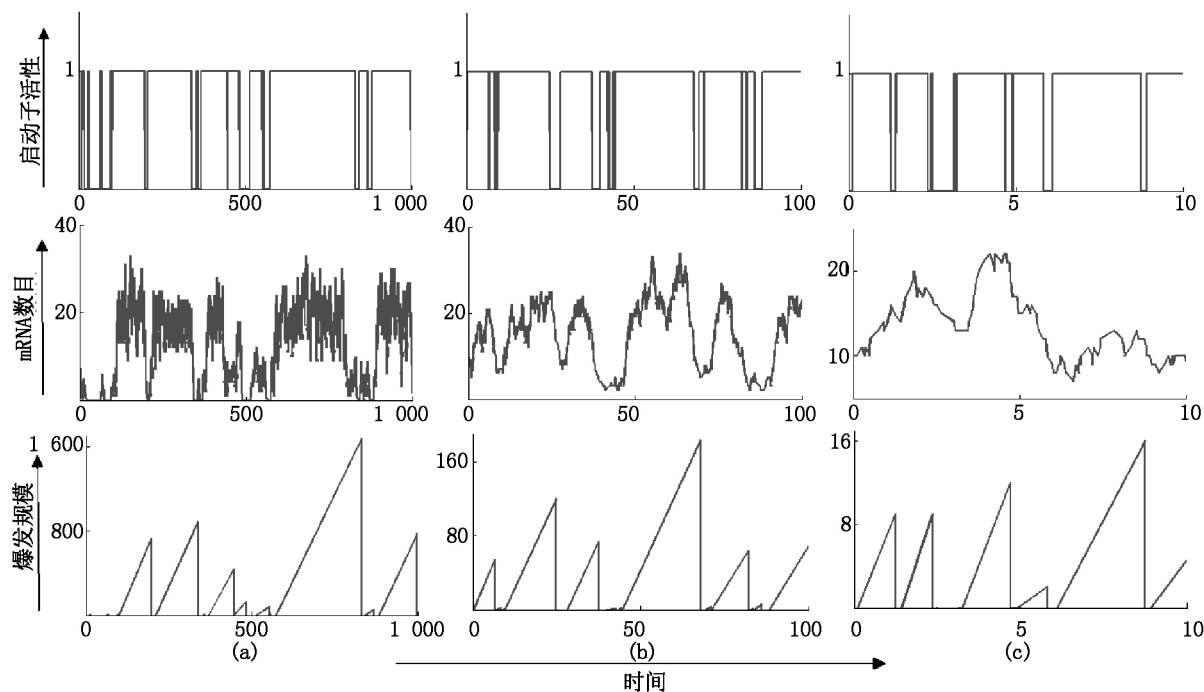


图 6 转录因子间的相互作用对爆发动力学影响的动态分析: 3 种切换下的情形

情形时的结果. 图 6(a) ~ (c) 分别相应于慢切换、中等切换和快切换的情形, 参数与图 2 中的相同, 因而 3 种切换下的平均水平是相同的. 最顶部一行展示的是启动子活性的时间变化图, 其中 0 表示 OFF 态, 1 表示 ON 态. 由此观察到: 在慢切换下, 基因在 ON 和 OFF 的驻留时间最长, 中等切换次之, 快切换

下的驻留时间最短. 第 2 行展示的是 mRNA 数目的时间序列图, 可以看到启动子切换快慢的不同表现出不同的特征: 从左到右, mRNA 分子数目分别在 3 个水平(其中有 1 个靠近 0), 2 个水平(都远离 0)和 1 个水平之间波动. 这与概率分布中峰的变化数目相一致: 分别对应于 3 个峰(其中 1 个靠近原点), 2

个峰(都远离原点)和单峰(看图2中顶部的3个子图).最底部一行展示的是爆发规模的时间序列图,得到与前面静态分析中相一致的结果,即慢切换下的爆发规模最大,中等切换次之,快切换下的爆发规模最小.

图7展示的是非对称切换下的3种调控机制情

形,这些动态结果再一次证实了转录因子间的相互作用对爆发动力学有非常重要的影响,特别是,在慢切换情形时影响最大.在慢切换情形下,基因在 ON 和 OFF 态的驻留时间最长;表型的多样性最明显;平均爆发规模和平均爆发频率最大.

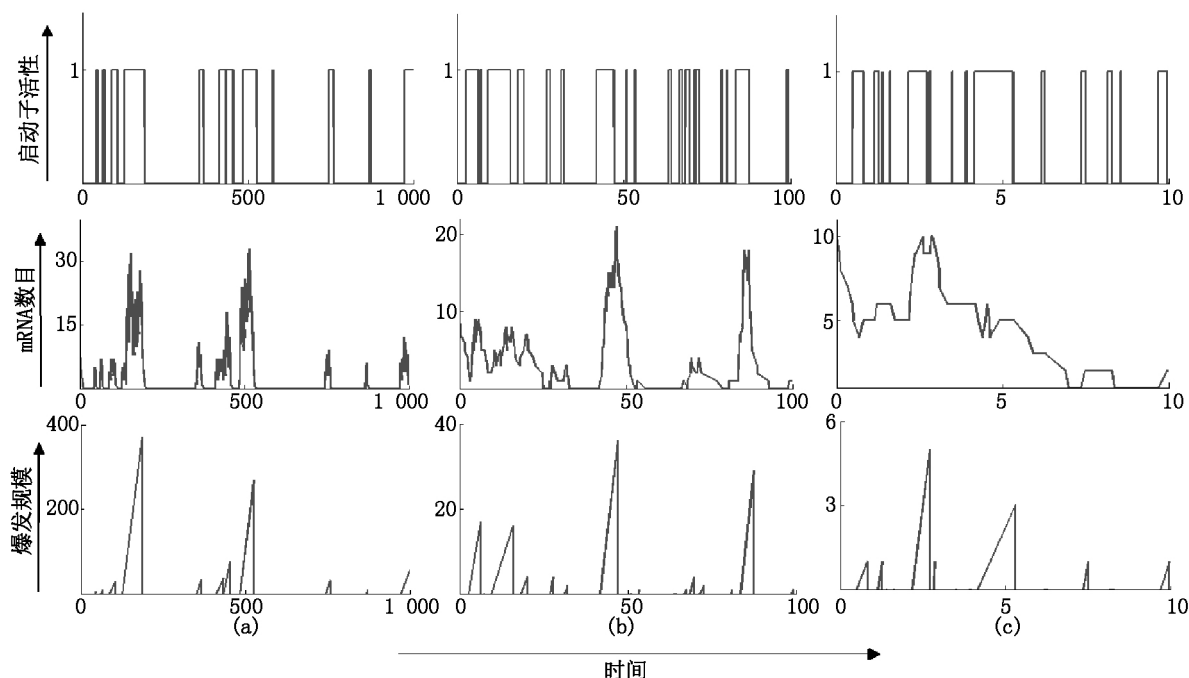


图7 转录因子间的相互作用对爆发动力学影响的动态分析:非对称切换下的情形

综合以上分析,发现转录因子间的相互作用对基因表达动力学具有重要影响:作用效能的增加能够导致爆发规模的增大,而越大的爆发规模导致差异性越大的细胞表型.此外,还发现爆发式表达在慢切换情形表现最为明显,这与先前关于酵母菌的研究结果是一致的^[3].本文的研究为群体细胞的生存提供了一种机制:群体细胞既可利用转录因子的稳定功能来提高其在波动环境中的生存能力,还可利用转录因子的招募功能最大限度地提高其在稳定环境中的生存能力.

3 结论与讨论

转录是一个复杂的过程,它涉及到转录因子与DNA聚合酶之间的一系列相互作用,以及涉及启动子在不同的活性态之间的切换.先前的研究表明启动子状态间的随机切换是导致同源群体细胞表型多样性的一个重要原因^[5],但是转录因子之间的相互作用如何影响转录动力和表达动力并未阐明.本文分析了一个多ON基因表达系统,考察了转录因子间的相互作用对基因表达动力学的影响,聚焦于转

录因子的3种作用机制对静态概率分布、表达噪声和爆发动力学的影响.通过细致分析发现:1) mRNA的平均水平在3种作用机制中是相同的,但是表达噪声在SM情形是最大的,在MM情形次之,在RM情形最小;2) 转录因子的作用能够导致多样化的概率分布,如靠近或远离原点的单峰,靠近或远离原点的双峰,以及靠近原点的多峰(实际是3个峰);3) 转录因子的总作用效能的增加能够导致爆发规模的增大,而爆发规模越大,群体细胞的表型异质性越强.特别地,转录因子间的相互作用在慢切换下爆发式表达的效果最为明显.总之,本文的结果表明转录因子间的相互作用在基因表达中是一个不可忽视的因素,其在控制基因表达谱以及在诱导细胞表型多样性方面起着重要的作用.

需要指出的是:本文的模型在几个方面进行了简化.1) 假设基因启动子仅有2个激活态和1个失活态,但是更复杂的启动子结构,比如多ON和多OFF是可能的^[2d];2) 没有考虑某些细化的生化过程,如转录和降解可以是多步的;3) 忽略了一些其它的生化过程,比如选择性剪切^[6d]、核驻留^[67]、反

馈调控^[2]等等. 这些过程或因素可能对于基因表达动力学行为也会产生重要影响,期待进一步探索和研究.

4 参考文献

- [1] Kussell E, Leibler S. Phenotypic diversity, population growth and information in fluctuating environments [J]. *Science*, 2005, 309(5743): 2075-2078.
- [2] Acar M, Mettetal J T, van Oudenaarden A. Stochastic switching as a survival strategy in fluctuating environments [J]. *Nat Genet*, 2008, 40(4): 471-475.
- [3] Fraser D, Kærn M. A chance at survival: gene expression noise and phenotypic diversification strategies [J]. *Mol Microbiol*, 2009, 71(6): 1333-1340.
- [4] Lehner B. Selection to minimise noise in living systems and its implications for the evolution of gene expression [J]. *Mol Sys Biol*, 2008, 4(1): 170.
- [5] Hung M, Chang E, Hussein R, et al. Modulating the frequency and bias of stochastic switching to control phenotypic variation [J]. *Nat Commun*, 2014, 5: 4574.
- [6] Eldar A, Elowitz M B. Functional roles for noise in genetic circuits [J]. *Nature*, 2010, 467(7312): 167-173.
- [7] Kærn M, Elston T C, Blake W J, et al. Stochasticity in gene expression: from theories to phenotypes [J]. *Nat Rev Gen*, 2005, 6(6): 451-464.
- [8] Raj A, Peskin C S, Tranchina D, et al. Stochastic mRNA synthesis in mammalian cells [J]. *PloS Biol*, 2006, 4(10): 1707-1719.
- [9] Spiller D G, Wood C D, Rand D A, et al. Measurement of single-cell dynamics [J]. *Nature*, 2010, 465(7299): 736-745.
- [10] Sanchez A, Choubey S, Kondev J. Noise and robustness in prokaryotic regulatory networks [J]. *Annu Rev Microbiol*, 2010, 64(64): 257-275.
- [11] Cai Long, Friedman N, Xie X Sunney. Stochastic protein expression in individual cells at the single molecule level [J]. *Nature*, 2006, 440(7082): 358-362.
- [12] Suter D M, Molina N, Gatfield D, et al. Mammalian genes are transcribed with widely different bursting kinetics [J]. *Science*, 2011, 332(6028): 472-474.
- [13] Larson D R, Zenklusen D, Wu Bin, et al. Real-time observation of transcription initiation and elongation on an endogenous yeast gene [J]. *Science*, 2011, 332(6028): 475-478.
- [14] Yunger S, Rosenfeld L, Garini Y, et al. Single-allele analysis of transcription kinetics in living mammalian cells [J]. *Nat Methods*, 2010, 7(8): 631-633.
- [15] Ko M S H. A stochastic model for gene induction [J]. *J Theor Biol*, 1991, 153(2): 181-194.
- [16] Kepler T B, Elston T C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations [J]. *Biophys J*, 2001, 81(6): 3116-3136.
- [17] Paulsson J. Models of stochastic gene expression [J]. *Phys Life Rev*, 2005, 2(2): 157-175.
- [18] Zhang Jiajun, Chen Luonan, Zhou Tianshou. Analytical distribution and tunability of noise in a model of promoter progress [J]. *Biophys J*, 2012, 102(6): 1247-1257.
- [19] Zhang Jiajun, Nie Qing, He Miao, et al. An effective method for computing the noise in biochemical networks [J]. *J Chem Phys*, 2013, 138(8): 1843-1852.
- [20] Zhang Jiajun, Zhou Tianshou. Promoter-mediated transcriptional dynamics [J]. *Biophys J*, 2014, 106(2): 479-488.
- [21] Huang Lifang, Yuan Zhanjiang, Liu Peijiang, et al. Feedback-induced counterintuitive correlations of gene expression noise with bursting kinetics [J]. *Phys Rev E*, 2014, 90(5): 052702.
- [22] Liu Peijiang, Yuan Zhanjiang, Huang Lifang, et al. Roles of factorial noise in inducing bimodal gene expression [J]. *Phys Rev E*, 2015, 91(6): 062706.
- [23] Saiz L, Vilar J M G. Stochastic dynamics of macromolecular-assembly networks [J]. *Mol Syst Biol*, 2006, 2(1): 2006. 0028.
- [24] Ptashne M. A genetic switch: gene control and phage lambda [M]. New York: Cold Spring Harbor Laboratory Press, 2004.
- [25] Joung J K, Koepp D M, Hochschild A. Synergistic activation of transcription by bacteriophage-lambda cI-protein and E. coli cAMP receptor protein [J]. *Science*, 1994, 265(5180): 1863-1866.
- [26] Bintu L, Buchler N E, Garcia H G, et al. Transcriptional regulation by the numbers: applications [J]. *Curr Opin Genet Dev*, 2005, 15(2): 125-135.
- [27] Ochab-Marcinek A, Tabaka M. Bimodal gene expression in noncooperative regulatory systems [J]. *Proc Natl Acad Sci USA*, 2010, 107(51): 22096-22101.
- [28] Becskei A, Serrano L. Engineering stability in gene networks by autoregulation [J]. *Nature*, 2000, 405(6786): 590-593.
- [29] Zhang Jiajun, Yuan Zhanjiang, Zhou Tianshou. Physical limits of feedback noise-suppression in biological networks [J]. *Phys Biol*, 2009, 6(4): 1100-1106.
- [30] Carey L B, van Dijk D, Sloot P M A, et al. Promoter sequence determines the relationship between expression level and noise [J]. *PloS Biol*, 2013, 11(4): e1001528.
- [31] Stekel D J, Jenkins D J. Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression [J]. *Bmc Syst Biol*, 2008, 2(2): 147-154.
- [32] Paulsson J. Summing up the noise in gene networks [J]. *Nature*, 2004, 427(6973): 415-418.

- [33] Friedman N, Cai Long, Xie X Sunny. Linking stochastic dynamics to population distribution: an analytical framework of gene expression [J]. Phys Rev Lett, 2006, 97 (16): 168302.
- [34] Dar R D, Razooky B S, Singh A, et al. Transcriptional burst frequency and burst size are equally modulated across the human genome [J]. Proc Natl Acad Sci USA, 2012, 109(43): 17454-17459.
- [35] Blake W J, Balázsi G, Kohanski M A, et al. Phenotypic consequences of promoter-mediated transcriptional noise [J]. Mol Cell, 2006, 24(6): 853-865.
- [36] Dellarole M, Sánchez I E, de Prat Gay G. Thermodynamics of cooperative DNA recognition at a replication origin and transcription regulatory site [J]. Biochemistry, 2010, 49 (48): 10277-10286.
- [37] Gutierrez P S, Monteoliva D, Diambra L. Role of cooperative binding on noise expression [J]. Phys Rev E, 2009, 80(1): 1274-1287.
- [38] Gutierrez P S, Monteoliva D, Diambra L. Cooperative binding of transcription factors promotes bimodal gene expression response [J]. PloS One, 2012, 7(9): e44812.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] McAdams H H, Arkin A. Stochastic mechanisms in gene expression [J]. Proc Natl Acad Sci USA, 1997, 94(3): 814-819.
- [43] Ozbudak E M, Thattai M, Kurtser I, et al. Regulation of noise in the expression of a single gene [J]. Nat Genet, 2002, 31(1): 69-73.
- [44] Hansen A S, O'Shea E K. Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression [J]. Mol Sys Biol, 2013, 9(1): 953.
- [45] Gillespie D T. Exact stochastic simulation of coupled chemical reactions [J]. J Chem Phys, 1977, 81(25): 2340-2361.
- [46] Wang Qianliang, Zhou Tianshou. Alternative-splicing-mediated gene expression [J]. Phys Rev E, 2014, 89(1): 390-400.
- [47] Kallehauge T B, Robert M C, Bertrand E, et al. Nuclear retention prevents premature cytoplasmic appearance of mRNA [J]. Mol Cell, 2012, 48(1): 145-152.

The Analysis of Expression Dynamics in a Multi-On Gene System

YANG Xiyan¹, ZHOU Tianshou²

(1. Department of Applied Mathematics, Guangdong University of Finance, Guangzhou Guangdong 510521, China;

2. School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou Guangdong 510275, China)

Abstract: The relationship between cellular phenotype across a genetically identical population and gene expression difference has being a very active topic. Here, a gene expression model at the transcription level is studied, considering not only complexity of promoter structure but also interactive effect between transcription factors. The promoter region is assumed to contain two regulatory sites and the promoter has one inactive state and two active states (i. e., two different transcription exits). In addition, based on interaction between transcription factors binding to the different regulatory sites, three representative mechanisms are considered, including recruitment mechanism, stabilization mechanism and mixture mechanism. Theoretical analysis and numerical simulation results show that different transcription exits can lead to mRNA distribution peaks of different modes; the mRNA mean level and mean burst frequency are identical under these three mechanisms, but stabilization mechanism induces the highest expression noise and the largest mean burst size, and recruitment mechanism induces the lowest expression noise and the smallest mean burst size. These results indicate that the source of cellular phenotype is complex, which is related to both promoter structure and regulation mechanisms.

Key words: multi-on model; total interaction efficiency; mRNA distribution; expression noise; bursting kinetics

(责任编辑: 王金莲)