

文章编号:1000-5862(2020)02-0215-06

# 分子记忆对基因表达过程中的能量消耗影响

杨喜艳<sup>1</sup>, 张家军<sup>2\*</sup>

(1. 广东金融学院金融数学与统计学院, 广东 广州 510521; 2. 中山大学数学学院, 广东 广州 510275)

**摘要:**一方面,分子记忆广泛存在于基因表达过程中;另一方面,由热力学的观点,基因表达过程必然消耗能量. 这引起一个未探索的问题:分子记忆如何影响基因表达的能量消耗. 由此,通过分析一个代表性的、带记忆的基因表达模型,结果发现:分子记忆越强,基因表达消耗越多能量;极小势和能量消耗之间存在反比例关系,这表明基因表达越稳定,需要消耗的能量越多. 这些结果表明分子记忆是一个不可忽略的因素,它能有意义地影响基因表达过程中的能量消耗.

**关键词:**分子记忆;基因表达;能量消耗;极小势

**中图分类号:**O 242;Q 141;Q 332 **文献标志码:**A **DOI:**10. 16357/j. cnki. issn1000-5862. 2020. 02. 19

## 0 引言

生化反应网络的随机动力学是一个重要的科学课题. 在已有的研究中,生化反应网络的建模与模拟一般建立在马氏假说上,即生物分子的随机生成或降解仅与当前分子状态有关,而不受先前状态的影响. 这种无记忆的性质表明马氏反应动力学能够通过具有指数等待时间的常速率的泊松过程来描述<sup>[1]</sup>,并且马氏过程在数学处理上的可操作性也带来了许多重要的应用<sup>[2-3]</sup>. 然而,生化网络的随机反应过程通常是有记忆的非马氏过程. 如在基因表达过程中,反应物与环境相互作用能够产生由非指数等待时间所特征化的分子记忆<sup>[4-5]</sup>. 另外,反应事件间的延时也能够产生分子记忆<sup>[6]</sup>. 由于分子记忆和网络拓扑联合影响反应系统的行为,现有的有关马氏理论的工具和数学方法不能直接用来分析非马氏动力. 虽说当前已有学者关注非马氏过程在实验观测<sup>[7-8]</sup>、数学模型<sup>[9-10]</sup>和数值模拟<sup>[11]</sup>等方面的研究,但由于非马氏动力在处理上的困难,因而这些文献也只是涉及到特别的系统,很难应用到一般的生化反应系统中. 近年来,基于随机游走理论<sup>[12-13]</sup>,已有研究提出了针对分子记忆的应用于随机反应系统的

有用公式,并用于分析非马氏效应<sup>[14]</sup>.

另一方面,基因表达是非平衡的动态过程,需要消耗能量来保证转录翻译等细胞过程的特异性和有效性. 研究发现:能量消耗使得酵母菌的转录和翻译过程在一个合理可控的范围内<sup>[15]</sup>;基因表达调控中噪声信号的解码需要消耗能量<sup>[16]</sup>;控制基因转录表达的多个相互作用的信号通路也需要消耗能量<sup>[17]</sup>. 此外,有研究探讨了两状态基因调控模型中能量消耗和表达噪声之间关系<sup>[18]</sup>;并且有关复杂启动子结构中能量消耗也已被分析<sup>[19]</sup>. 这些研究表明:能量消耗在基因表达中起着非常重要的作用. 然而,这些有关能量消耗的研究是基于马氏过程的随机反应动力学提出的,即随机反应过程的等待时间是指数分布,因而可通过反应速率的变化来描述随机生化反应. 实际上,生化反应中的能量消耗是一个非马氏随机过程,而分子记忆的效应如何影响基因调控系统中的能量消耗至今还不清楚. 本文将就此问题展开研究,同时还考虑基因表达的稳定性与能量消耗的关系.

本文关注分子记忆效应下的基因自调控系统中的能量消耗,不同于马氏过程利用反应速率描述生化反应,这里应用反应过程中等待时间的概率密度函数来描述. 通过引进与马氏描述等价的有效转移

收稿日期:2019-08-05

基金项目:国家自然科学基金(11601094,11631005,11475273,11775314),广州市科技基金(201707010117),广东省自然科学基金(2018A0303130120)和广东省教育科学规划课题(2018GXJK119)资助项目.

通信作者:张家军(1978-),男,湖北麻城人,副教授,博士,博士生导师,主要从事计算系统生物学的研究. E-mail:zhjiajun@mail. sysu. edu. cn

速率,将具有分子记忆的非马氏过程转化为无记忆的  
马氏过程,并获得概率分布的静态解.最后在此基  
础上通过数值模拟来分析本文的相关结果,发现分  
子记忆效应极大地影响着自调控基因系统中的能量  
消耗,并且基因表达越稳定,能量消耗越大.本文的  
研究结果从生物学功能的角度揭示了具有分子记忆  
的基因调控系统中能量消耗的运行机制.

1 模型及分析结果

1.1 模型

为了定性地分析分子记忆效应与能量消耗的关系,考虑基因自调控模型

$$\text{ON} \xrightarrow{\psi_1(t;n)} \text{ON} + P, P \xrightarrow{\psi_2(t;n)} \emptyset,$$

其中  $P$  表示蛋白质,  $\psi_1(t;n)$  和  $\psi_2(t;n)$  分别表示蛋  
白质生成和降解反应中等待时间分布的概率密度函  
数,其中  $n$  表示蛋白质分子的数目,且假设非指数等  
待时间分布是 Gamma 分布,并考虑 2 种情形的分子  
记忆:1) 生成反应是非马氏过程,则  $\psi_1(t;n) =$   
 $((\lambda_1(n))^{k_1}/\Gamma(k_1))t^{k_1-1}e^{-\lambda_1(n)t}, \psi_2(n;t) = \lambda_2 n e^{-\lambda_2 n t};$   
2) 降解反应是非马氏过程,则  $\psi_1(t;n) =$   
 $\lambda_1(n)e^{-\lambda_1(n)t}, \psi_2(t;n) = ((\lambda_2 n)^{k_2}/\Gamma(k_2))t^{k_2-1}e^{-\lambda_2 n t}.$   
在 2 种情形中,  $\lambda_1(n)$  表示自调控.不失一般性,考  
虑如下形式的线性反馈:  $\lambda_1(n) = a + \lambda_1 n$  (正反  
馈),  $\lambda_1(n) = a - \lambda_1 n$  (负反馈),这里  $a$  表示基本生  
成率,参数  $\lambda_1$  量化调控效应或表示反馈强度.此外,  
用  $\Psi_i(t;n)$  表示  $\psi_i(t;n)$  的累积分布函数.注意到,  
当等待时间分布的密度函数  $\psi_1(t;n), \psi_2(t;n)$  均为  
指数分布的密度函数时,模型变为在无记忆的马氏  
情形下的基因自调控模型,已有很多的文献对该模  
型及其复杂模型做了研究<sup>[20-21]</sup>.虽然有相关文献确  
立了连续时间的随机游走理论<sup>[11-13]</sup>,但是有关非马  
氏动力分析的求解问题仍未解决.这里,考虑系统的  
稳态行为,即假设随机反应系统的静态概率分布存在.  
根据文献<sup>[14]</sup>,引进 2 个与马氏特征等价的有效  
转移率

$$K_1(n) = \frac{\int_0^{+\infty} \psi_1(t;n)(1 - \Psi_2(t;n))dt}{\int_0^{+\infty} (1 - \Psi_1(t;n))(1 - \Psi_2(t;n))dt},$$
  
$$K_2(n) = \frac{\int_0^{+\infty} \psi_2(t;n)(1 - \Psi_1(t;n))dt}{\int_0^{+\infty} (1 - \Psi_1(t;n))(1 - \Psi_2(t;n))dt}, \quad (1)$$

其中  $K_1(n)$  和  $K_2(n)$  分别表示有效生成速率和有效  
降解速率,从而可得到如下形式的静态概率主方程  
 $(E^{-1} - I)K_1(n)P(n) + (E - I)K_2(n)P(n) = 0,$  (2)  
其中  $E$  和  $E^{-1}$  是步长算子,  $I$  是单位算子,  $P(n)$  表示  
蛋白质分子的静态概率分布.方程(1)和方程(2)  
表明将具有分子记忆的随机反应过程转化为与之等  
价的无记忆马氏过程.

1.2 概率分布

为了求得静态概率分布,将方程(2)改写为  
 $(E - I)(K_2(n)P(n) - E^{-1}K_1(n)P(n)) = 0.$   
这隐含着  
 $K_2(n)P(n) - E^{-1}K_1(n)P(n) = -J,$   
其中  $-J$  表示从  $n$  到  $n - 1$  的概率净流.令  $n = 0$ ,得  
 $J = 0.$  因此,

$$K_2(n)P(n) = K_1(n - 1)P(n - 1).$$

从而可获得如下形式的概率分布

$$P(n) = P(0) \prod_{i=1}^n \frac{K_1(i - 1)}{K_2(i)}, \quad (3)$$

其中  $P(0)$  可通过保守条件  $\sum_{n=0}^{\infty} P(n) = 1$  计算得出.

经计算,对于生成反应是非马氏过程的自调控  
模型,有

$$K_1(n) = (\lambda_1(n))^{k_1} \lambda_2 n / ((\lambda_1(n) + \lambda_2 n)^{k_1} -$$
  
 $(\lambda_1(n))^{k_1}), K_2(n) = \lambda_2 n. \quad (4)$

对于降解反应是非马氏过程的自调控模型,有

$$K_1(n) = \lambda_1(n), K_2(n) = \lambda_1(n)(\lambda_2 n)^{k_2} / ((\lambda_1(n) +$$
  
 $\lambda_2 n)^{k_2} - (\lambda_2 n)^{k_2}). \quad (5)$

注意到在(4)式中当  $k_1 = 1, K_1(n) = \lambda_1(n)$  以  
及在(5)式中  $k_2 = 1, K_2(n) = \lambda_2 n$  时,模型变为马  
氏情形下的基因自调控模型.

1.3 量化能量消耗

从热力学观点来说,基因表达过程需要消耗能  
量;另一方面,在生化反应系统中,自由能的消耗速  
率能够通过生化反应中前后流的变化来计算<sup>[22]</sup>,即

$$\dot{W} = \sum_i (J_i^+ - J_i^-) \ln(J_i^+ / J_i^-), \quad (6)$$

其中  $J_i^+$  和  $J_i^-$  分别表示第  $i$  个反应中向前和向后的  
概率流.在马氏系统中,上述能量消耗速率可通过状  
态之间的转移概率以及在状态处静态概率分布来量  
化.对于非马氏系统中能量消耗的计算,类似于马氏  
系统,利用与非马氏等价的马氏主方程所获得的静

态概率分布及有效转移速率来测量分子记忆效应下的能量消耗速率,从而(6) 式可改写为

$$\dot{W} = \sum_m P(m) (K_1(m) \ln(K_1(m)/K_2(m)) + K_2(m) \ln(K_2(m)/K_1(m))).$$

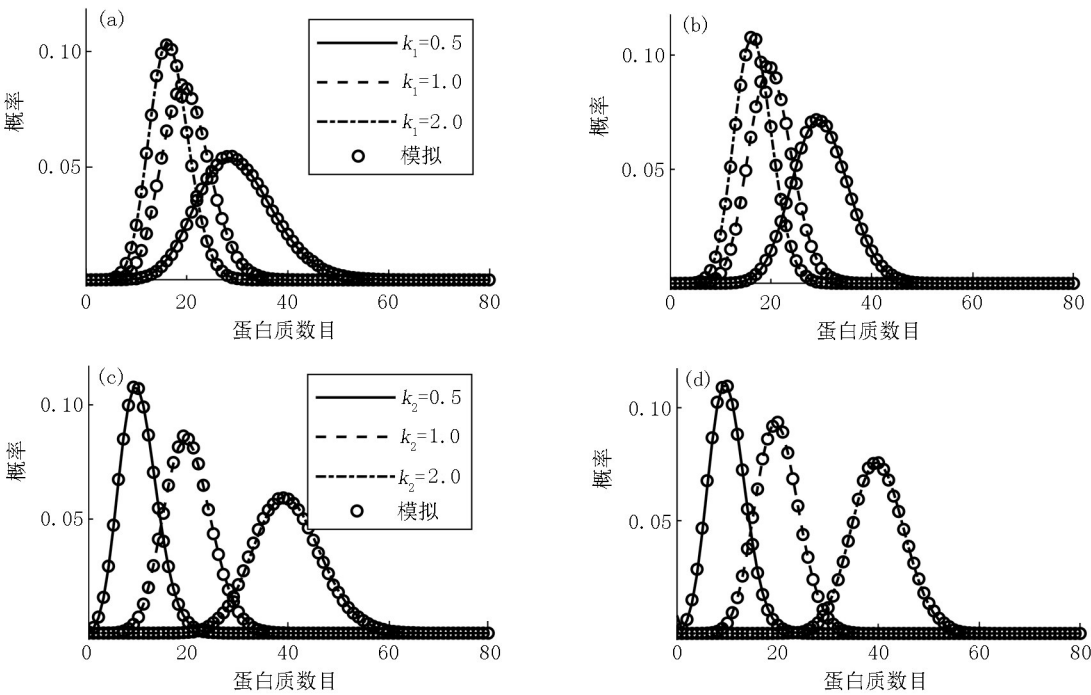
2 数值结果

下面将基于上面所得到的结果来考察分子记忆效应下能量消耗与基因表达稳定性的关系.

2.1 分子记忆对概率分布的影响

图1展示的是分子记忆效应对蛋白质概率分布

的影响, 其中曲线是理论预测(3), 圆圈是随机模拟<sup>[14]</sup>, 可以看到模拟结果很好地拟合了理论预测结果. 从图 1 中的第 1 行发现, 当生成反应是非马氏过程时, 随着 Gamma 分布中形状参数  $k_1$  的增加, 概率分布的峰值增加, 并且分布的图形左移, 这表明分子记忆能够导致蛋白质表达水平的降低. 然而, 当降解反应是非马氏过程时, 结果刚好相反: 随着形状参数  $k_2$  的增加, 概率分布的峰值降低, 并且分布的图形右移, 这表明分子记忆能够导致蛋白质表达水平的增加. 子图中的参数选取为  $\lambda_1 = 0.05, \lambda_2 = 0.50$ , 在所有的情形中, 变化参数  $a$  并固定非马氏反应中平均等待时间为 0.1 s.



注:这里固定非马氏反应中平均等待时间为 0.1 s,其中曲线表示理论结果,圆圈表示模拟结果.(a) 与(b) 中生成反应是非马氏过程,(c) 与(d) 中降解反应是非马氏过程;其中(a) 与(c) 是正反馈情形,(b) 与(d) 是负反馈情形.

图 1 分子记忆关于蛋白质概率分布的效应

为了分析基因表达稳定性与能量消耗之间关系,引进势函数(Potential function)  $U = -\ln P(m)$ . 从上面的分析可知,上述自调控模型的概率分布有一个局部最大值或者说有一个峰值. 相应地,势函数有一个局部极小值,称为极小势(Potential minimum). 极小势越小意味着表达越稳定. 此外,用来描述分子记忆效应的 Gamma 分布的概率密度函数中 2 个参数  $\lambda_1$  (或  $\lambda_2$ ) 和  $k_1$  (或  $k_2$ ) 分别为记忆系数(Memory coefficient) 和记忆指数(Memory exponent),并用这 2 个指标来研究分子记忆效应下

的能量消耗与基因表达稳定性之间的关系.

2.2 分子记忆效应对极小势和能量消耗的影响

图 2 展示的是当同时变化记忆系数和记忆指数时,极小势和能量消耗的变化情况,这里考察的是生成反应是非马氏的正反馈情形. 从图 2 中发现,随着记忆系数的增大,极小势增加而能量消耗减小;而随着记忆指数的增加,极小势降低而能量消耗增加. 这些结果表明越稳定的基因表达水平需要消耗越多的能量. 图 2 中参数选取为  $\lambda_2 = 0.5$ ,变化参数  $a$  并固定非马氏反应中平均等待时间为 0.1 s.

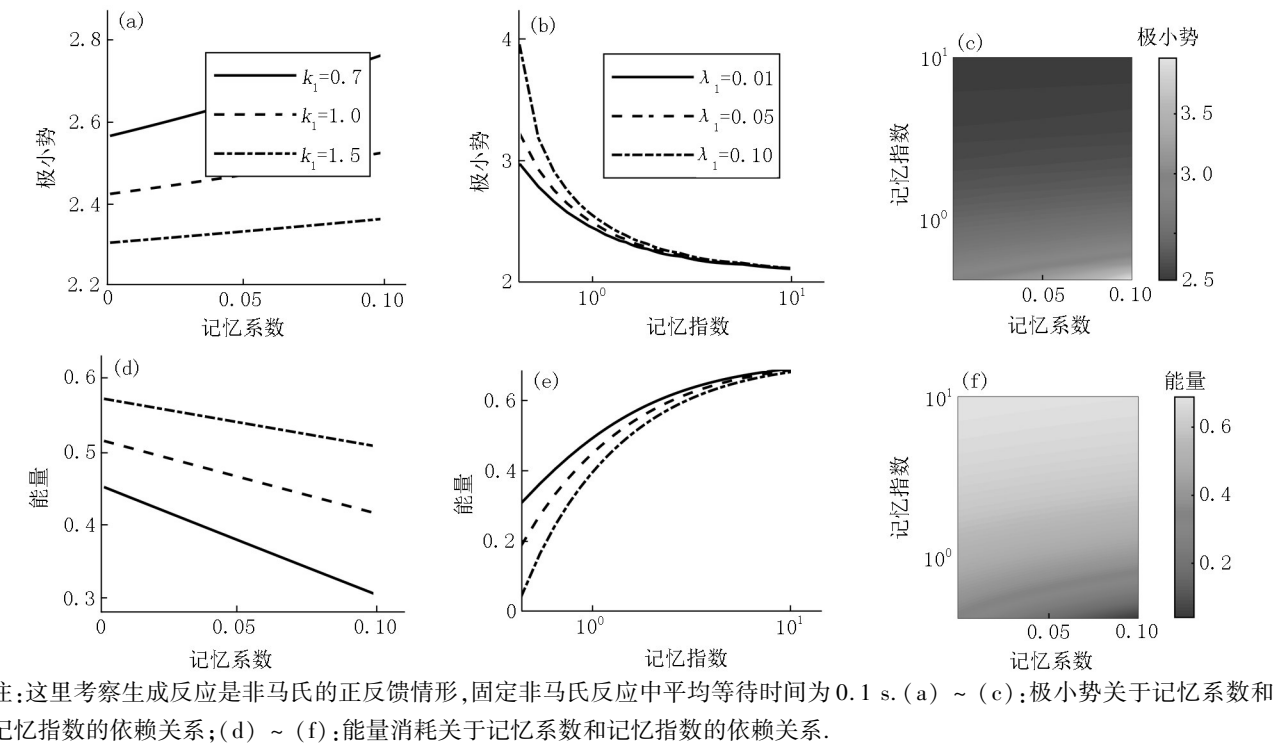


图2 分子记忆效应对极小势和能量消耗的影响,生成反应是非马氏的正反馈情形

图3展示的是在负反馈情形下,分子记忆效应对极小势和能量消耗的影响.与图2一样,假设生成反应是非马氏过程.通过与图2对比发现:相同之处是,记忆指数的增加能够降低极小势,但是增加能量消耗;不同之处的是,随着记忆系数的增加,极小势减小而能量消耗增加.从图3可看出,极小势和能量消耗之间呈反比例的关系,这再次说明越稳定的基因表达水平需要消耗越多的能量.图3中参数选取为 $\lambda_2 = 1.0$ ,变化参数 $a$ 并固定非马氏反应中平均等待时间为0.05 s.

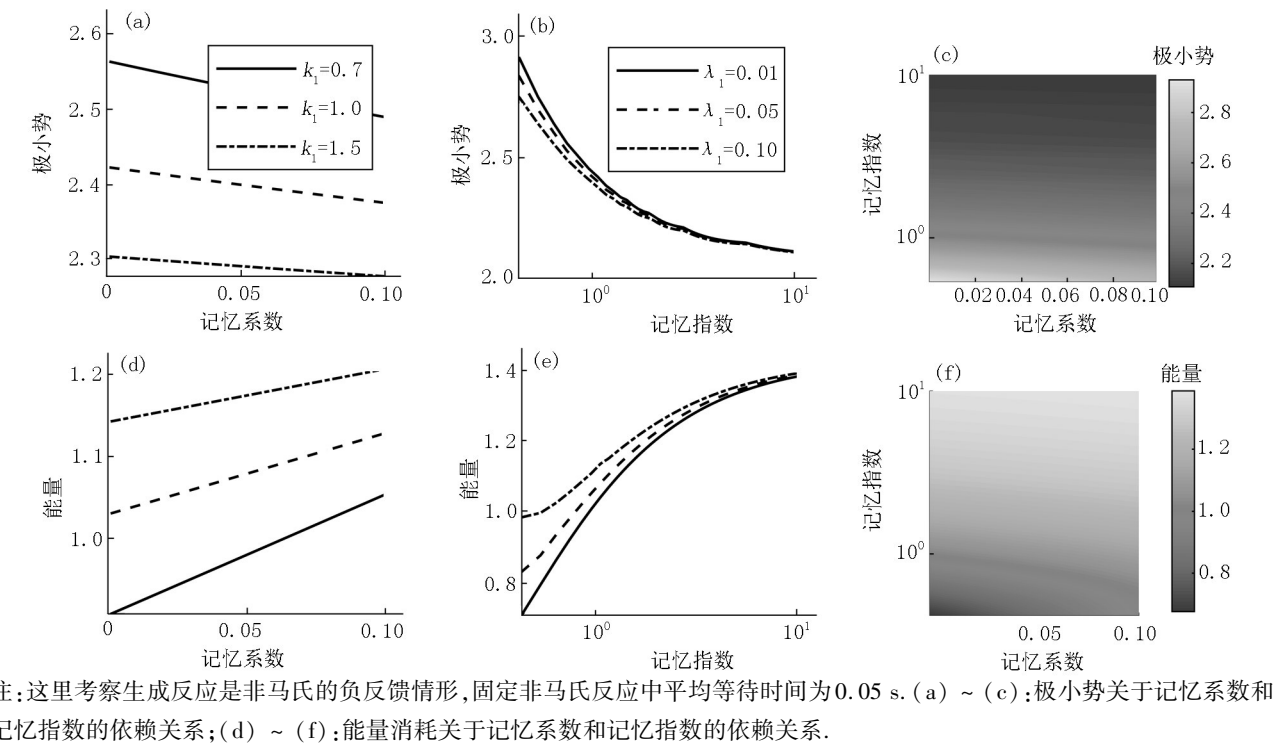


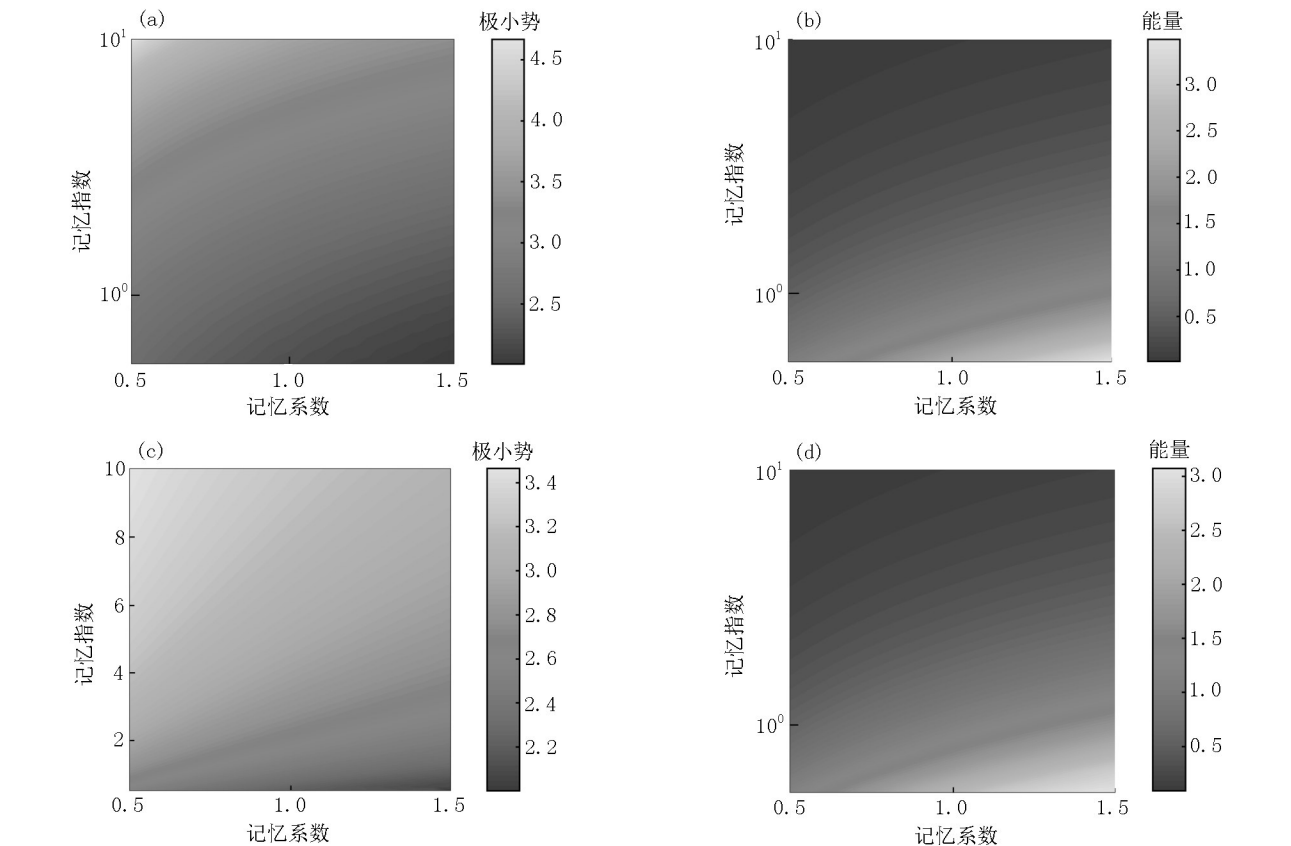
图3 分子记忆效应对极小势和能量消耗的影响,生成反应是非马氏的负反馈情形

接下来分析当降解反应是非马氏时,分子记忆效应对极小势和能量消耗的影响.图4展示了相关



结果,其中第 1 行表示正反馈情形,第 2 行表示负反馈情形. 从图 4 中可看出,无论是正反馈情形还是负反馈情形,随着记忆系数的增加,极小势降低,能量消耗增加;随着记忆指数的增加,极小势增加,能量

消耗减少,并且再次证实了越稳定的基因表达水平需要消耗越多的能量这一定性的结果. 图 4 中参数选取为  $\lambda_1 = 0.03$ ,变化参数  $a$  并固定非马氏反应中平均等待时间为 0.05 s.



注:这里考察降解反应是非马氏的情形,固定非马氏反应中平均等待时间为 0.05 s. (a)与(b):正反馈中极小势和能量消耗分别关于记忆系数和记忆指数的依赖关系;(c)与(d):负反馈中极小势和能量消耗分别关于记忆系数和记忆指数的依赖关系.

图 4 分子记忆效应对极小势和能量消耗的影响,降解反应是非马氏的情形

综合图 2 ~ 图 4 中的分析,发现分子记忆效应对自调控基因系统中能量消耗的影响是不同的. 当生成反应是非马氏时,越大的记忆指数要求越多的能量消耗;而当降解反应是非马氏时,记忆指数越大,能量消耗越少. 此外,除了非马氏生成反应中的正反馈情形,记忆系数的增大能够导致能量消耗的降低(如图 2(d)). 在其他情形中,有相反的结果:记忆系数的增大能够导致能量消耗的增加(如图 3(d),图 4(b)和图 4(d)). 另一方面,发现极小势和能量消耗之间有一个反比例关系,获得了基因表达的稳定性以能量消耗为代价这一定性结果.

3 结论与讨论

近年来,随着单分子测量技术的发展,基因表达动力学的研究已成为一个热门的研究课题. 已有研究表明基因表达的过程能够产生分子记忆<sup>[4-6]</sup>,即

具有非马氏特征. 现有的用来分析基因表达过程中马氏动力的理论和工具显然不能用来分析非马氏动力. 本文应用随机反应过程中等待时间分布的概率密度函数来描述随机反应过程,通过引进有效转移速率,将具有分子记忆的非马氏过程转化为与之等价的无记忆的马氏过程,并获得概率分布的静态解. 最后,在此基础上分析在分子记忆效应下的能量消耗与基因表达稳定性之间的关系. 本文所获得的这些定性结果表明了分子记忆效应对基因表达动力学有着非常重要的影响,相关研究结果为进一步理解复杂生物调控系统中的非马氏动力提供可借鉴的分析方法.

4 参考文献

[1] Van Kampen N G. Stochastic processes in physics and chemistry [M]. North-Holland: Amsterdam, 2007.

[2] Pardox E. Markov processes and applications; algorithms, networks, genome and finance [M]. New York:Wiley and Sons,2008.

[3] Andersson H, Britton T. Stochastic epidemic models and their statistical analysis [M]. New York:Springer,2000.

[4] Pedraza J M,Paulsson J. Effects of molecular memory and bursting on fluctuations in gene expression [J]. Science, 2008,319(5861):339-343.

[5] Jia Tao,Kulkarni R V. Intrinsic noise in stochastic models of gene expression with molecular memory and bursting [J]. Phys RevLett,2011,106(5):058102.

[6] Guérin T, Leveriner N, Bénichou O, et al. Mean first-passage times of non-Markovian random walkers in confinement [J]. Nature,2016,534(7607):356-359.

[7] Guérin T, Bénichou O, Voituriez R. Non-Markovian polymer reaction kinetics [J]. Nat Chem,2012,4(7):568-573.

[8] Stumpf P S,Smith R C G, Lenz M, et al. Stem cell differentiation as a non-Markov stochastic process [J]. Cell Syst,2017,5(3):268-282.

[9] Van Mieghem P, van de Bovenkamp R. Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks [J]. Phys Rev Lett,2013,110(10):108701.

[10] Kiss I Z,Röst G, Vizi Z. Generalization of pairwise models to non-Markovian epidemics on networks [J]. Phys Rev Lett,2015,115(7):078701.

[11] Masuda N,Rocha L E C. A Gillespie algorithm for non-Markovian stochastic processes [J]. SIAM Rev,2018,60(1):95-115.

[12] Aquino T, Dentz M. Chemical continuous time random walks [J]. Phys Rev Lett,2017,119(23):230601.

[13] Landman U, Montroll E W, Shlesinger M F. Random walks and generalized master equations with internal degrees of freedom [J]. ProcNatl AcadSci,1977,74(2):430-433.

[14] Zhang Jiajun,Zhou Tianshou. Stationary equations for non-Markovian biochemical systems [J]. Proc Natl Acad Sci, 2019,116(47):23542-23550.

[15] Wagner A. Energy constraints on the evolution of gene expression [J]. Mol Bio Evol,2015,22(6):1365-1374.

[16] Liu Peijiang, Wang Haohua, Huang Lifang, et al. The dynamic mechanism of noisy signal decoding in gene regulation [J]. Scientific Reports,2017,7:42128.

[17] Wang Haohua, Liu Peijiang, Li Qingqing, et al. Entangled signal pathways can both control expression stability and induce stochastic focusing [J]. FEBS Lett, 2018, 592(7):1135-1149.

[18] Huang Lifang, Yuan Zhanjiang, Yu Jianshe, et al. Fundamental principles of energy consumption for gene expression [J]. Chaos,2015,25(12):123101.

[19] Li Qingqing, Din A,Zhou Tianshou. Energy cost and mean dwell times for the activity of promoter with complex structure [J]. J Syst Sci Complex,2019,32(2):510-525.

[20] Peccoud J, Ycart B. Markovian modeling of gene-product synthesis [J]. Theor Popul Biol,1995,48(2):222-234.

[21] Shahrezaei V, Swain P S. Analytical distributions for stochastic gene expression [J]. Proc Natl Acad Sci,2008, 105(45):17256-17261.

[22] Qian Hong. Phosphorylation energy hypothesis; open chemical systems and their biological functions [J]. Annu Rev Phys Chem,2007,58(1):113-142.

The Influence of Molecular Memor on Energy Consumption in Gene Expression

YANG Xiyan<sup>1</sup>,ZHANG Jiajun<sup>2\*</sup>

(1. School of Financial Mathematics and Statistics,Guangdong University of Finance,Guangzhou Guangdong 510521,China;  
2. School of Mathematics,SunYat-Sen University,Guangzhou Guangdong 510275,China)

**Abstract:** On one hand, molecular memory exists extensively in gene expression. On the other hand, gene expression necessarily consumes energy from the viewpoint of thermodynamics. This raises an unexplored issue how molecular memory affects energy consumption in gene expression. By analyzing a representative stochastic model of gene expression with molecular memory, it is found that the stronger is the molecular memory, the more is the energy consumed. In addition, it is found that there is an inverse relationship between potential minimum and energy consumption, implying that more stable gene expression needs more energy consumption. These results indicate that molecular memory is an unneglectable factor and can significantly impact energy consumption in gene expression.

**Key words:** molecular memory; gene expression; energy consumption; potential minimum

(责任编辑:王金莲)