

文章编号: 1000-5862(2015)03-0230-10

# 生化反应系统的建模与分析

雷锦誌

(清华大学周培源应用数学研究中心, 北京 100084)

**摘要:** 很多细胞行为受细胞内基因表达和蛋白质相互作用等生物化学反应的调控. 这些细胞内的生化反应表现出明显的随机性, 它成为描述细胞行为的可计算建模中不可忽略的因素. 该文是关于计算系统生物学中随机模拟的基本理论和新进展的自洽综述. 本综述从生化反应系统的基本假设出发, 介绍关于生化反应内部噪声、外部随机扰动和包含时间滞后的反应过程的各种数学描述, 包括化学主方程、化学福克尔-普朗克方程、化学速率方程、化学郎之万方程等; 还介绍了相关的数值模拟方法, 包括随机模拟算法和 $\tau$ 跳跃算法.

**关键词:** 系统生物学; 随机模拟; 化学主方程; 生物化学反应; 随机微分方程

**中图分类号:** O 29 Q 61 O 175 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2015.03.02

## 0 引言

近几十年来, 生物技术的快速发展把人们带进了前所未有的对生命现象微观事实和动态过程的研究领域, 例如体内单分子测量技术<sup>[1-6]</sup>、基因编辑技术<sup>[7-8]</sup>和细胞重编程技术<sup>[9-12]</sup>等. 随着人们对这些微观过程了解的日益完善, 对其进行定量可计算建模不仅成为可能, 而且可以促进人们对生命现象背后的调控机制和动态规律的深刻理解. 在微观尺度, 生命现象主要表现为各种生化反应, 包括蛋白质的各种修饰、蛋白质与各种大分子之间的相互作用、各种大分子的结构变化、合成与分解、体内各种离子对蛋白质活性的调控等<sup>[13-21]</sup>, 与普通的化学反应(参加反应的分子个数通常在 $10^{23}$ 量级)不同, 这些生化反应所涉及的分子数量通常比较少(从几个到几千个), 表现出明显的随机性, 生命系统中的这些随机性对于细胞行为有重要意义<sup>[18, 22-23]</sup>. 因此, 在对微观生命系统的可计算建模过程中随机涨落是必须要考虑的因素<sup>[14, 21, 24]</sup>.

对于普通的参加反应分子的个数为 $10^{23}$ 量级的化学反应, 化学动力学可以很好地描述相关化学反应的动力学行为. 化学动力学的数学形式通常表现为常微分方程组的形式, 即反应速率方程, 或者是当考虑到反应物的空间扩散行为时表现为反应扩散方

程组. 这些确定性的方程主要用于描述反应物的浓度随时间变化的动力学. 然而, 对于细胞内的微观分子行为, 这些确定性方程却不足以准确描述系统的状态. 例如, DNA 的转录过程通常涉及相关单个基因启动子的激活或失活, mRNA 的翻译过程受制于核糖体与 mRNA 序列的结合. 这些过程的共同特点是参加反应的分子个数很少, 例如每个单拷贝基因的转录只涉及单个启动子的激活<sup>[24-26]</sup>, 细菌中每个基因所对应的 mRNA 数量只有少量的几十个<sup>[27]</sup>, 即使是数量较多的蛋白质的数量也通常在几千个的量级<sup>[28]</sup>. 因此, 考虑到每个分子每次反应的随机性, 这些生物化学反应系统的动力学行为表现出明显的不确定性. 对这些不确定过程的可计算建模需要不同于反应速率方程的数学工具和计算方法.

本综述将介绍用于模拟此类随机行为的随机动力学建模和计算模拟方法, 并讨论各种数学形式之间的关系<sup>[29]</sup>, 更详细内容可以参见文献[23, 30].

## 1 生化反应系统

在计算生物学的随机模拟中, 为简单起见, 可以把细胞内的分子行为近似为 1 个生化反应系统. 这里首先介绍 1 个生化反应系统的一般描述.

考虑 1 个具有固定体积  $\Omega$  和常温  $T$  的系统, 包含充分混合的  $N(\geq 1)$  种分子  $\{S_1, \dots, S_N\}$ . 系统的

收稿日期: 2015-01-15

基金项目: 国家自然科学基金面上(11272169)和重大研究计划(91430101)资助项目.

作者简介: 雷锦誌(1974-), 男, 广东台山人, 副研究员, 博士, 主要从事应用数学、动力系统与应用和计算生物学的研究.

状态用每种分子的个数表示:  $X(t) = (X_1(t), \dots, X_N(t))$ , 其中  $X_i(t)$  = 在时刻  $t$  分子  $S_i$  的个数 ( $t = 1, 2, \dots, N$ ).

这些分子通过  $M (\geq 1)$  个化学反应通道  $\{R_1, \dots, R_M\}$  改变系统的状态. 给定初始状态  $X(t_0) = x_0$ , 系统的状态由  $X(t)$  的动态过程给出. 显然, 因为每次反应的发生是随机事件, 系统状态  $X(t)$  是一个随机过程. 因此, 对系统的描述应该以系统状态的统计性质为主, 而不仅仅描述单个样本轨道.

首先, 假设每个反应一旦发生就会立刻完成. 这一假设是和包含时间滞后的化学反应区分开<sup>[31]</sup>. 关于包含时间滞后的反应系统, 目前的研究还不是很多, 将在后面讨论. 另外, 假设所有反应物充分混合以忽略空间不均匀性. 在这些假设下, 每个反应通道  $R_j$  与相应的趋向性函数  $a_j$  和状态改变向量  $v_j = (v_{j1}, \dots, v_{jN})$  关联. 趋向性函数和状态改变向量分别定义为

$a_j(x) dt$  = 给定  $X(t) = x$ , 反应  $R_j$  在  $\Omega$  内在下一个时间区间  $[t, t + dt)$  内发生的概率 ( $j = 1, 2, \dots, M$ ),

$v_{ji}$  = 当发生 1 次反应  $R_j$  时分子  $S_i$  个数的改变量 ( $j = 1, 2, \dots, M; i = 1, 2, \dots, N$ ).

根据定义, 趋向性函数和状态改变向量一起给出了反应  $R_j$  的数学描述. 因此, 根据所有反应通道的趋向性函数和状态改变向量可以得到反映系统状态变化的动力学方程.

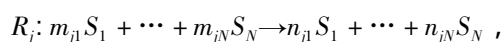
给定反应通道的状态改变向量可以很容易根据化学反应方程式给出. 只需要数一下每种分子在 1 次反应中被消耗或者产生的数量即可.

对反应通道的趋向性函数的定量描述与反应中所涉及的化学势的改变和化学动力学的随机行为密切相关<sup>[32]</sup>. 通常只能给出近似的描述. 一般地, 可以把函数  $a_j$  表示为如下的数学形式<sup>[33]</sup>

$$a_j(x) = c_j h_j(x), \quad (2)$$

这里  $c_j$  是反应通道  $R_j$  的化学反应常数. 定义  $c_j dt$  为参加反应  $R_j$  的反应物按完成反应所需分子个数的任意合适的组合在下一个无穷小时间区间  $dt$  内发生 1 次反应的概率. 这个概率  $c_j dt$  依赖于相应反应的化学势, 同时受反应体系的温度  $T$  和体积  $\Omega$  的影响. 通常具有形式  $c_j \propto e^{-\Delta\mu_j/k_B T}$ , 其中  $\Delta\mu_j$  为对应的化学势垒,  $k_B$  为玻尔兹曼常数. 在真实系统中, 大部分的化学反应都可以分解为单分子或者双分子的基元反应. 对单分子反应, 反应常数  $c_j$  与系统体积  $\Omega$  无关. 而对于双分子反应, 反应常数  $c_j$  与体积  $\Omega$  成反比. 对于 3 分子反应, 常数  $c_j$  通常正比于  $\Omega^{-2}$ <sup>[33]</sup>.

方程 (2) 中的函数  $h_j(x)$  表示给定系统的状态  $x$  下反应  $R_j$  所需的反应物的可能组合数. 这个组合数可以通过反应化学方程式和简单的组合计算得到<sup>[30]</sup>. 一般地, 对于化学反应



有

$$v_{ji} = n_{ji} - m_{ji}, \quad a_j(x) \propto \prod_{k=1}^N \frac{x_k!}{m_{jk}!(x_k - m_{jk})!}.$$

如果  $\forall k, j$ , 有  $x_k \geq m_{jk}$ , 则近似有

$$a_j(x) = c_j \prod_{k=1}^N x_k^{m_{jk}}. \quad (3)$$

在计算中为方便起见通常采用近似表达式 (3).

## 2 动力学方程内部噪声的数学描述

首先假设所有的反应趋向性函数都是不依赖于时间的, 即方程 (2) 中的  $c_j$  都是常数. 在这种情况下, 系统的随机性来源于系统内蕴的热力学涨落, 也称为内部噪声. 而与之对应, 如果系统的随机涨落是因为环境因素的随机变化引起的, 则称之为外部噪声. 这里先介绍关于内部噪声的数学描述, 而对外部噪声的数学描述在下一节讨论.

### 2.1 化学主方程

生化系统的状态向量  $X(t)$  的动态变化是 1 个  $N$  维整数格点空间上的带跳马氏过程. 其状态转移概率是由所有反应通道所对应的反应趋向性函数所定义的. 下面给出这个随机过程的数学描述. 为此首先定义概率密度函数

$$P(x, t | x_0, t_0) = \text{Prob}\{X(t) = x \text{ 如果 } X(t_0) = x_0\}.$$

这里总是采用大写字母表示随机变量, 而相应的小写字母为该随机变量的一个确定值. 在物理上, 假设有很多独立的生化反应系统 (也称系综), 每个反应系统都包含相同的分子, 并且体积、温度和初始状态均相同. 则在时刻  $t$ , 这些系统的状态可以是任意值. 概率密度函数  $P(x, t | x_0, t_0)$  表示在时刻  $t$  状态为  $x$  的系统的个数所占的比例.

通过概率密度函数可以得到关于系统动力学的精细描述. 例如任何一个关于系统状态的函数  $f(X)$  的期望可以表示为

$$\langle f(X) \rangle = \sum_{x \in \mathcal{X}} f(x) P(x, t | x_0, t_0),$$

其中  $\mathcal{X}$  表示系统的所有可能状态的集合. 注意这里的期望是表示在条件  $X(t_0) = x_0$  下的条件期望.

在任一时刻概率  $P(x, t | x_0, t_0)$  随时间的演化由 2 部分组成, 分别是状态为  $x - v_j$  的系统发生了

反应  $R_j$  使概率增加和状态为  $x$  的系统发生了任意 1 个反应使概率减少. 由此可推导出化学主方程<sup>[33-34]</sup>:

$$\frac{\partial}{\partial t} P(x, t | x_0, t_0) = \sum_{j=1}^M [P(x - v_j, t | x_0, t_0) a_j(x - v_j) - P(x, t | x_0, t_0) a_j(x)]. \quad (4)$$

方程(4) 是根据生化反应系统的反应趋向性函数和状态改变向量所定义的随机动力学的精确描述, 下面所介绍的方程都可以由此推导出. 如果可以由方程(4) 求解出  $P$ , 就可以得到关于系统统计性质的动态描述. 然而, 由于维数问题, 方程(4) 的精确解是很难求解的(关于求解方程(4) 的精确方法, 可以参见文献[35-36]).

## 2.2 福克尔-普朗克方程

如果在所考虑的生化系统中所有分子的个数  $X_i(t)$  都远大于 1, 则可以把状态变量  $X(t)$  近似为实数. 进一步假设函数  $a_j(x) P(x, t | x_0, t_0)$  关于变量  $x$  是光滑的, 则由泰勒展开到 2 阶近似可以给出化学福克尔-普朗克方程

$$\frac{\partial}{\partial t} P(x, t) = - \sum_{i=1}^N \frac{\partial}{\partial x_i} A_i(x) P(x, t) + \frac{1}{2} \sum_{1 \leq i, k \leq N} \frac{\partial^2}{\partial x_i \partial x_k} B_{ik}(x) P(x, t), \quad (5)$$

其中  $A_i(x) = \sum_{j=1}^M v_{ji} a_j(x)$ ,  $B_{ik}(x) = \sum_{j=1}^M v_{ji} v_{jk} a_j(x)$ .

下面还将由化学郎之万方程推导出这个化学福克尔-普朗克方程.

## 2.3 化学速率方程

由化学主方程(4), 方程两边乘以  $x_i$ , 并对所有  $x$  相加可以得到下面的化学系综平均方程

$$\frac{d \langle X_i \rangle}{dt} = \sum_{j=1}^M v_{ji} \langle a_j(X) \rangle \quad (i = 1, 2, \dots, N). \quad (6)$$

方程(6) 是关于(4) 式的系综平均的精确动力学方程. 然而, 当反应通道中存在高阶反应时方程(6) 通常是不封闭的.

若所有反应通道都是单分子反应, 则所有反应趋向性函数  $a_j(X)$  都是线性函数, 此时有

$$\langle a_j(X) \rangle = a_j(\langle X \rangle) \quad (j = 1, 2, \dots, N). \quad (7)$$

另一方面, 如果系统的随机性可以忽略(例如系统中每种分子的个数足够大), 方程(7) 也是近似成立的. 此时, 方程(6) 可以近似为确定性常微分方程组

$$\frac{dx_i}{dt} = \sum_{j=1}^M v_{ji} a_j(x) \quad (i = 1, 2, \dots, N), \quad (8)$$

其中  $x(t)$  的分量是实数. 令  $z_i(t) \equiv x_i(t) / \Omega$  ( $i = 1,$

$2, \dots, N$ ) 表示分子的浓度, 则有动力学方程

$$\frac{dz_i}{dt} = \sum_{j=1}^M v_{ji} \tilde{a}_j(z) \quad (i = 1, 2, \dots, N), \quad (9)$$

其中  $\tilde{a}_j(z) = a_j(\Omega z) / \Omega$ .

方程(8) 或(9) 就是在化学动力学中常见反应速率方程或称为化学速率方程. 这类方程的优点是容易进行数值求解. 然而, 在上面的推导过程中可以看到方程(8) 成立的条件是关系(7) 近似成立. 而对很多生物系统中的生化反应, 关系(7) 并不成立, 此时确定性的化学速率方程不能很好地描述系统的动力学行为. 有些情况下对同一个系统采用化学速率方程和采用化学主方程所得到的结果甚至是矛盾的<sup>[37-38]</sup>, 这就是著名的凯泽悖论(Keizer's paradox). 下面介绍的化学郎之万方程既可描述系统的随机性又便于进行数值求解.

## 2.4 化学郎之万方程

化学郎之万方程是由吉莱斯皮根据化学主方程和 2 个基本假设推导出来的随机微分方程组, 用于描述生化系统的随机动力学<sup>[33]</sup>.

假设系统的当前时刻  $t$  的状态为  $X(t) = x$ . 令  $K_j(x, \tau)$  ( $\tau > 0$ ) 表示反应通道  $R_j$  在时间区间  $[t, t + \tau]$  内发生的次数. 因为每次反应都会使分子  $S_i$  的个数改变  $v_{ji}$ , 分子  $S_i$  的个数在时刻  $t + \tau$  变为

$$X_i(t + \tau) = x_i + \sum_{j=1}^M K_j(x, \tau) v_{ji} \quad (i = 1, 2, \dots, N), \quad (10)$$

注意到这里  $K_j(x, \tau)$  是随机变量, 因此  $X_i(t + \tau)$  是随机的.

下面引入 2 个条件以得到随机数  $K_j(x, \tau)$  的近似分布.

条件 1 假设  $\tau$  充分小使系统状态在时间区间  $[t, t + \tau]$  内的变化不明显改变趋向性函数的值;

条件 2 假设  $\tau$  充分大使每个反应通道  $R_j$  在时间区间  $[t, t + \tau]$  内发生次数的平均值远远大于 1.

如果上述 2 个条件都满足,  $K_j(x, \tau)$  可以近似为正态分布随机数

$$K_j(x, \tau) \approx N_j(a_j(x) \tau, \sqrt{a_j(x) \tau} N_j(0, 1)),$$

其中  $N_j(m, \sigma^2)$  表示均值为  $m$  和方差为  $\sigma^2$  的正态分布的随机变量, 这里  $M$  个随机变量是相互独立的. 由此可以把方程(10) 改写为

$$X_i(t + \tau) = x_i + \sum_{j=1}^M v_{ji} a_j(x) \tau + \sum_{j=1}^M v_{ji} \sqrt{a_j(x) \tau} N_j(0, 1) \quad (i = 1, 2, \dots, N). \quad (11)$$

把时间间隔  $\tau$  记为  $dt$ , 并记  $dX_i = X_i(t + dt) -$

$X_i(t)$ . 然后, 引进  $M$  个独立的维纳过程(或称布朗运动) 满足  $dW_j(t) = W_j(t+dt) - W_j(t) = N_j(0, 1) \sqrt{dt}$  ( $j = 1, 2, \dots, N$ ), 则方程(11) 可改写为郎之万方程(或随机微分方程) 的形式

$$dX_i = \sum_{j=1}^M v_{ji} a_j(X) dt + \sum_{j=1}^M v_{ji} \sqrt{a_j(X)} dW_j \quad (i = 1, 2, \dots, N). \quad (12)$$

这里的随机微分方程(12) 就是由吉莱斯皮提出的化学郎之万方程, 方程(12) 满足初始条件  $X(0) = X_0$  的解是满足关系

$$X(t) = X_0 + \sum_{j=1}^M \int_0^t v_{ji} a_j(X(s)) ds + \sum_{j=1}^M \int_0^t v_{ji} \sqrt{a_j(X(s))} dW_j(s) \quad (13)$$

的随机过程  $X(t)$ . 这里的随机积分采用伊藤积分解释, 用于表示内部噪声<sup>[34]</sup>.

注意到在上面的推导过程中, 条件1 和条件2 是相互矛盾的, 这2 个条件对郎之万方程适用的系统提出了一些限制. 注意到条件2 蕴含当  $\tau$  很小时趋向性函数  $a_j(x)$  要足够大. 因此当系统中分子的个数很大时2 个条件是有可能同时满足的, 因为每个函数  $a_j(x)$  通常是状态变量  $x$  的分量的多项式. 然而, 在生物系统中的很多时候2 个条件是不能同时满足的, 此时关于化学郎之万方程的推导过程并不成立. 另一方面, 即使条件1 和条件2 不同时满足, 化学郎之万方程(12) 仍然是数学上具有明确定义的随机微分方程, 对该方程所描述的随机过程进行研究还是有意义的. 下面对这个问题进行讨论.

## 2.5 关于化学郎之万方程的讨论

在细胞内的很多生物化学反应中, 参加反应的分子个数很少, 而且反应都很慢, 例如基因表达或者基因调控网络的动力学过程等. 在这些系统中推导化学郎之万方程的2 个条件不能同时成立. 然而方程(12) 仍然可为描述系统的随机动力学行为提供有用的信息.

首先对方程(12) 取期望并注意到伊藤积分的性质  $\langle \sqrt{a_j(X)} dW_j \rangle = 0$ , 因此

$$d \langle X_i \rangle / dt = \sum_{j=1}^M v_{ji} \langle a_j(X) \rangle \quad (i = 1, 2, \dots, N).$$

又重新得到了前面由化学主方程推导出的方程(6). 这说明方程(13) 中的随机积分采用伊藤积分的合理性.

不同分子个数变化的相关函数定义为

$$\sigma_{ij}(t) = \langle X_i(t) X_j(t) \rangle - \langle X_i(t) \rangle \langle X_j(t) \rangle, \quad 1 \leq i, j \leq N, \text{ 则由化学主方程(4) 可以推导出 } \sigma_{ij}(t) \text{ 满}$$

足的化学系综相关方程<sup>[29]</sup>

$$d\sigma_{ij}/dt = \langle A_i(X) (X_j - \langle X_j \rangle) \rangle + \langle (X_i - \langle X_i \rangle) A_j(X) \rangle + \langle B_{ij}(X) \rangle. \quad (14)$$

另一方面, 由化学郎之万方程(12) 并使用伊藤公式<sup>[39]</sup>, 也可以得到相同的化学系综相关方程(14). 方程(14) 是由化学主方程推导出来的关于  $\sigma_{ij}(t)$  动力学方程的严格形式.

在统计平衡态附近, 如果随机涨落很小, 则近似地有

$$A_i(X) \approx A_i(\langle X \rangle) + \sum_{l=1}^N \frac{\partial A_i(\langle X \rangle)}{\partial X_l} (X_l - \langle X_l \rangle), \\ \langle B_{ij}(X) \rangle \approx B_{ij}(\langle X \rangle).$$

把上述近似表达式代入(14) 式, 并定义矩阵

$$\sigma = (\sigma_{ij}), \quad A = (\partial A_i(\langle X \rangle) / \partial X_l), \quad B = (B_{ij}(\langle X \rangle)),$$

则可以得到方程

$$d\sigma/dt = (A\sigma + \sigma A^T) + B. \quad (15)$$

方程(15) 给出了近平衡态附近相关函数所满足的线性近似方程. 这个方程也是涨落耗散定理的1 个例子<sup>[34]</sup>.

最后由化学郎之万方程(12), 也可以推导出(5) 式所给出的化学福克尔-普朗克方程<sup>[29, 34]</sup>. 因此, 化学郎之万方程和化学主方程都可以得到相同形式的化学郎之万方程.

比较化学郎之万方程和化学主方程, 有以下几点值得注意:

1) 化学郎之万方程和化学主方程都可以推导出相同形式的化学系综平均方程(6)、化学系综相关方程(14) 和化学福克尔-普朗克方程(5);

2) 虽然关于系综平均和相关函数的方程(6) 和(14) 的形式是相同的, 但是由于这里的系综平均都是关于概率密度函数  $P(x, t)$  的, 而该概率密度函数的求解需要分别由化学主方程或者化学郎之万方程得到, 因此由这2 类方程并不一定能得到相同的关于系综平均和相关性的动力学过程. 如果所有的反应通道都是单分子反应, 则对应的方程(6) 和(14) 本身是封闭的, 因此也就给出了相同的动力学过程;

3) 化学郎之万方程和化学主方程都对应于相同的化学福克尔-普朗克方程. 故可认为到2 阶近似, 化学郎之万方程和化学主方程给出了相同的关于概率密度函数  $P(x, t)$  的动力学演化过程.

## 3 动力学方程-外部噪声的数学描述

在上面关于内部噪声数学描述的动力学方程

中,假设方程(2)中的反应速率 $c_j$ 为常数.这一假设仅仅是对实际情况的简单近似.在实际问题的研究中经常需要考虑环境的随机涨落,因此这些反应速率也是随机和时变的.在这种情况下目前还没有很好的数学工具能对系统进行数学描述,在这里给出一些粗浅的看法,虽不成熟但也希望能引起研究人员对相关问题的关注.

由前面的讨论,反应趋向性函数的一般形式为

$$a_j(\mathbf{x}, t) = c_j h_j(\mathbf{x}), \quad (16)$$

其中 $h_j(\mathbf{x})$ 作为内蕴函数是由化学反应本身确定的,一般是固定的,而 $c_j$ 依赖于体系的温度和体积等环境因素,通常受到环境涨落因素的影响而表现为随机的.然而,由于环境因素的复杂性,要对这个反应速率进行精确描述一般来说是不可能的,下面给出这个随机过程的一个近似描述.

通常有形式 $c_j \propto e^{-\Delta\mu_j/k_B T}$ ,如果对化学势垒有随机扰动使 $\Delta\mu_j$ 变为 $\Delta\mu_j - k_B T \eta_j(t)$ ,其中 $\eta_j(t)$ 是随机扰动,则可以把反应速率 $c_j(t)$ 用以下函数

$$c_j(t) = \bar{c}_j e^{\eta_j(t)} / \langle e^{\eta_j(t)} \rangle \quad (17)$$

代替,其中 $\bar{c}_j$ 是常数,表述反应速率的平均值.

在很多情况下,随机扰动 $\eta(t)$ (这里忽略下标 $j$ )可以用奥恩斯坦-乌伦贝克过程表示为

$$d\eta = -(\eta/\tau) dt + (\sigma/\sqrt{\tau}) dW, \quad (18)$$

这里 $W$ 是维纳过程, $\tau$ 和 $\sigma$ 为正常数,分别表示自相关时间和方差.容易验证 $\eta(t)$ 是在任何时刻满足正态分布并且具有指数衰减的自相关函数<sup>[34, 40-42]</sup>

$$\langle \eta(t) \eta(t') \rangle = \frac{\sigma^2}{2\tau} e^{-|t-t'|/\tau}.$$

奥恩斯坦-乌伦贝克过程是有色噪声的1个例子.当 $\tau \rightarrow 0$ 则 $\eta(t)$ 趋向于高斯白噪声<sup>[42]</sup>.对奥恩斯坦-乌伦贝克过程 $\eta(t)$ 有<sup>[43]</sup>  $\langle e^{\eta(t)} \rangle = e^{\sigma^2/(4\tau)}$ .因此,反应速率(17)式可表示为 $c_j(t) = \bar{c}_j e^{\eta_j(t) - \sigma_j^2/(4\tau_j)}$ ,则反应趋向性函数(16)变为

$$a_j(\mathbf{x}, t) = \bar{c}_j e^{\eta_j(t) - \sigma_j^2/(4\tau_j)} h_j(\mathbf{x}), \quad (19)$$

其中 $\eta_j(t)$ 满足形如(18)的方程.

若 $\eta(t)$ 是1个奥恩斯坦-乌伦贝克过程,则反应速率 $c_j(t)$ 的静态分布为对数正态分布.对数正态分布已经被应用于很多问题中<sup>[44]</sup>,例如基因表达的速率满足对数正态分布<sup>[45-46]</sup>.

把反应趋向性函数(19)和方程(18)代入反应速率方程(8),可以得到下面的随机微分方程

$$dX_i = \left( \sum_{j=1}^M v_{ji} \bar{c}_j e^{\eta_j(t) - \sigma_j^2/(4\tau_j)} h_j(\mathbf{X}) \right) dt \quad (i = 1, 2, \dots, N), \quad (20)$$

$$d\eta_j = -(\eta_j/\tau_j) dt + (\sigma_j/\tau_j) dW_j \quad (j = 1, 2, \dots, M) \quad (21)$$

这里 $\bar{c}_j, \tau_j, \sigma_j$ 都是正常数.方程(20)~(21)推广了原来的反应速率方程,用于描述包含对系统参数的随机扰动的反应速率方程.

把反应趋向性函数(19)和方程(18)代入化学郎之万方程(12),得到下面的随机微分方程

$$dX_i = \left( \sum_{j=1}^M v_{ji} \bar{c}_j e^{\eta_j(t) - \sigma_j^2/(4\tau_j)} h_j(\mathbf{X}) \right) dt +$$

$$\sum_{j=1}^M v_{ji} \left( \bar{c}_j e^{\eta_j(t) - \sigma_j^2/(4\tau_j)} h_j(\mathbf{X}) \right)^{1/2} dW_j \quad (i = 1, 2, \dots, N) \quad (22)$$

$$d\eta_j = -(\eta_j/\tau_j) dt + (\sigma_j/\tau_j) dW_j \quad (j = 1, 2, \dots, M) \quad (23)$$

这里 $W_j$ 是独立维纳过程,而 $\bar{c}_j, \tau_j, \sigma_j$ 如前面一样都是正常数.这一组广义化学郎之万方程描述了同时包含内部噪声和对反应速率有随机扰动的外部噪声的随机动力学过程.

上面所给出的郎之万方程(22)和(23)同时包含系统状态和环境的涨落.对于如何由此方程得到只描述系统状态涨落的概率密度函数的演化方程(福克尔-普朗克方程)仍是数学上的困难问题.

## 4 具有时间滞后的生化反应系统

在前面所介绍的生化反应系统中假设所有反应通道都是瞬间完成的.但是在很多生化反应中这样的假设并不完全成立,例如基因的转录、蛋白质的翻译和降解等过程都不是瞬间完成的,而是需要一个明显的过程.为了描述此类包含过程的生化反应系统,需要把化学主方程(4)推广到包含时间滞后的情况.包含时间滞后的生化反应系统的研究还比较少,这里的内容主要参见文献[31, 47-48].

类似前面的讨论,考虑1个包含 $N$ 种分子 $\{S_1, \dots, S_N\}$ 的充分混合的生物化学系统,每种分子 $S_i$ 在时刻 $t$ 的分子数为 $X_i(t)$ ,共有 $M$ 个反应通道 $\{R_1, \dots, R_M\}$ ,每个反应通道的反应趋向性函数和状态改变向量分别为 $a_j(\mathbf{x})$ 和 $\mathbf{v}_j$ ( $j = 1, 2, \dots, M$ ).假设在这些反应中,前 $M_1$ 个反应 $\{R_1, \dots, R_{M_1}\}$ 分别有时间延滞 $\{\tau_1, \dots, \tau_{M_1}\}$ ,即这些反应如果发生,则需要一定的时间才能最终完成,其余的 $M - M_1$ 个反应则没有时间延滞.令 $P(\mathbf{x}, t)$ 表示给定初始状态 $\mathbf{X}(t) = \Phi(t)$ ( $t \leq t_0$ )下的概率密度函数为

$$P(\mathbf{x}, t) = \text{Prob}\{\mathbf{X}(t) = \mathbf{x} | \mathbf{X}(t) = \Phi(t), t \leq t_0\}.$$

关于概率密度函数 $P(\mathbf{x}, t)$ 推广的化学主方程<sup>[47]</sup>为

$$\frac{\partial}{\partial t} P(\mathbf{x}, t) = \sum_{j=1}^{M_1} \sum_{\mathbf{y} \in \mathcal{X}} a_j(\mathbf{y}) P(\mathbf{x} - \mathbf{v}_j, t; \mathbf{y}, t - \tau_j) -$$

$$\sum_{j=1}^{M_1} \sum_{y \in \mathcal{X}} a_j(y) P(x, t; y, t - \tau_j) + \sum_{j=M_1+1}^M \sum_{y \in \mathcal{X}} a_j(x - v_j) \cdot P(x - v_j, t) = \sum_{j=M_1+1}^M a_j(x - v_j) P(x, t), \quad (24)$$

其中  $P(x, t; y, t - \tau_j)$  表示系统在时刻  $t$  时的状态为  $x$  和在时刻  $t - \tau_j$  的状态为  $y$  的联合概率密度,  $\mathcal{X}$  表示系统的所有可能状态.

注意到关系

$$P(x - v_j, t; y, t - \tau_j) = P(x - v_j, t | y, t - \tau_j) P(y, t - \tau_j),$$

其中  $P(x - v_j, t | y, t - \tau_j)$  表示在  $t - \tau_j$  时刻系统的状态为  $y$  的条件下在时刻  $t$  系统的状态为  $x - v_j$  的条件概率, 则由方程 (24) 可以推导出下面的关于生化反应系统系综平均的动力学方程<sup>[48]</sup>

$$\frac{d\bar{X}}{dt} = \sum_{j=1}^{M_1} v_j \overline{a_j(X(t - \tau_j))} + \sum_{j=M_1+1}^M v_j \overline{a_j(X(t))}.$$

如前面讨论, 如果所有反应都是 1 阶反应, 或者随机涨落可以忽略, 则有以下近似等式

$$\overline{a_j(X(t))} = a_j(\bar{X}(t)).$$

这种情况下可以得到以下的时滞微分方程

$$\frac{dx}{dt} = \sum_{j=1}^{M_1} v_j a_j(x(t - \tau_j)) + \sum_{j=M_1+1}^M v_j a_j(x(t)).$$

这样得到了关于带有滞后反应的生化反应系统的化学主方程和对应的反应速率方程, 其它形式的动力学方程的数学形式目前还不清楚.

## 5 生化反应系统的随机模拟

前面给出了关于生化反应系统的数学描述, 这里介绍数值模拟方法. 这些模拟方法通过随机模拟产生随机过程  $X(t)$  的样本轨道, 如果可以得到足够的独立样本轨道, 则可以通过统计方法得到密度函数  $P(x, t)$  和其他统计量.

### 5.1 随机模拟算法

设系统在时刻  $t$  的状态为  $x$ . 定义发生下一个反应的概率密度函数  $p(\tau, j | x, t)$  为: 给定  $X(t) = x$ , 下一个反应在时间  $[t + \tau, t + \tau + d\tau)$  内在系统中某处发生, 并且所发生的反应为  $R_j$  的概率.

根据反应趋向性函数的定义 (1) 并由简单计算可以得到<sup>[49-50]</sup>

$$p(\tau, j | x, t) = a_j(x) \exp(-a_0(x)\tau) (0 \leq \tau < \infty, j = 1, 2, \dots, M), \quad (25)$$

其中  $a_0(x) = \sum_{k=1}^M a_k(x)$ .

这一概率密度函数的表达式给出了随机模拟算

法(也常被称为吉莱斯皮算法)的基础.

为产生满足分布 (25) 的随机数  $(\tau, j)$ , 可以按照下面的计算方法进行<sup>[50]</sup>: 首先产生 2 个单位区间内均匀分布的随机数  $r_1$  和  $r_2$ , 然后取

$$\tau = (1/a_0(x)) \ln(1/r_1),$$

$$j = \text{满足关系 } \sum_{j=1}^j a_j(x) > r_2 a_0(x) \text{ 的最大整数.}$$

下面是随机模拟算法的主要步骤(详细过程可参见文献 [30, 50]):

Step 1 初始化: 令  $X = x_0$  和  $t = t_0$ ;

Step 2 蒙特卡洛模拟: 产生一对满足分布 (25) 的随机数  $(\tau, j)$  (这里  $x = X$ );

Step 3 状态更新: 时间增加  $\tau$ , 并把系统状态  $X$  更新为  $X + v_j$ ;

Step 4 迭代: 转第 2 步或停止模拟.

因为方程 (4) 和 (25) 都是根据反应趋向性函数 (1) 严格推导出来的, 上面的随机模拟算法和化学主方程的描述等价. 这个算法模拟了生化系统的一个真实可能的时间序列.

### 5.2 $\tau$ 跳跃算法

在推导化学郎之万方程的过程中, 总是假设  $\tau$  足够小使下面的跳跃条件成立.

跳跃条件: 系统的反应趋向性函数在时间  $[t, t + \tau)$  内没有明显改变.

在这一条件下, 可以以时间步长  $\tau$  的跳跃来模拟系统的动力学过程

$$X(t + \tau) \approx x + \sum_{j=1}^M \mathcal{P}_j(a_j(x), \pi) v_j, \quad (26)$$

这里  $x = X(t)$ .

方程 (26) 是下面要介绍的  $\tau$  跳跃算法的基础<sup>[51]</sup>: 从当前状态  $x$  开始, 首先选取满足跳跃条件的  $\tau$  值, 然后对每个反应  $j$  产生满足均值为  $a_j(x)\tau$  的泊松分布的随机数  $k_j$ , 最后把系统状态由  $x$  更新为  $x + \sum_j k_j v_j$  并把时间增加  $\tau$ .

在应用  $\tau$  跳跃算法时需要解决 2 个问题: 首先是如何估计合适的  $\tau$  值, 其次是如何避免产生负的分分子数.

对  $\tau$  的选取首先要满足跳跃条件, 同时取值尽可能大以加速计算. 吉莱斯皮最早给出的选取  $\tau$  值的方法, 取  $\tau$  满足<sup>[51]</sup>

$$\tau = \min_{j \in [1, M]} \left\{ \varepsilon a_0(x) / \left| \sum_{i=1}^N \xi_i(x) b_{ji}(x) \right| \right\},$$

其中  $\xi(x) = \sum_{j=1}^M a_j(x) v_j$ ,  $a_0(x) = \sum_{j=1}^M a_j(x)$ ,  $b_{ji}(x) = \partial a_j(x) / \partial x_i$ .

吉莱斯皮在后来的工作中又对上述结果进行了改进,详细内容可参见文献[52-53].

在公式(26)中,如果状态改变向量 $v_j$ 的某个分量是负数,而泊松分布的随机数 $k_j$ 足够大,则有可能在模拟过程中产生负的分子数,这显然是不合理的.有很多关于 $\tau$ 跳跃算法的改进算法考虑了如何避免负分子数的问题,其中最简单的方法是用有界的2项分布随机数代替泊松分布的随机数<sup>[54-55]</sup>.文献[56]提出了另一个算法用于解决这一问题.在这一算法中把所有反应通道分为2类,分别是消耗反应和非消耗反应.消耗反应会消耗反应物,对应的状态改变向量有负分量.非消耗反应不会消耗反应物,因此状态改变向量都是非负的.在每一步模拟中对于非消耗反应可以采用正常的 $\tau$ 跳跃算法以估计跳跃时间 $\tau'$ .而对于消耗反应则采用随机模拟算法以得到对应的时间步长 $\tau''$ 和下一步发生反应的通道 $j_c$ .实际的跳跃时间 $\tau$ 则取为 $\tau'$ 和 $\tau''$ 的最小值.如果是前者则没有消耗反应发生,如果是后者,则只发生1次消耗反应.

下面给出这一改进的 $\tau$ 跳跃算法的主要过程(详细过程参见文献[56]):

Step 1 初始化:  $X = x_0$  和  $t = t_0$ ;

Step 2 确定消耗反应: 确定满足  $a_j(x) > 0$  并且  $v_j$  包含负分量的消耗反应  $R_j$ ;

Step 3 计算跳跃时间: 对非消耗反应计算相应的跳跃时间  $\tau'$ ;

Step 4 蒙特卡洛模拟: 对消耗反应根据相应的概率密度函数(25)产生随机数( $\tau'' j_c$ );

Step 5 确定下一步反应:

(a) 若  $\tau' < \tau''$ , 取  $\tau = \tau'$ . 对所有消耗反应  $R_j$ , 令  $k_j = 0$ . 对所有非消耗反应  $R_j$ , 产生均值为  $a_j(x)\tau$  的泊松分布随机数  $k_j$ ;

(b) 若  $\tau' \geq \tau''$ , 取  $\tau = \tau''$ . 取  $k_{j_c} = 1$ , 并对其余的所有消耗反应取  $k_j = 0$ . 对所有非消耗反应  $R_j$ , 产生均值为  $a_j(x)\tau$  的泊松分布随机数  $k_j$ ;

Step 6 状态更新: 时间增加  $\tau$ , 并把系统状态  $X$  更新为  $X + \sum_{j=1}^M k_j v_j$ ;

Step 7 迭代: 转第2步或停止模拟.

### 5.3 其它模拟方法

因为生物系统的复杂性和跨尺度性质,关于复杂生化反应的计算模拟是计算系统生物学领域值得研究的重要问题.除了上面介绍的随机模拟算法和 $\tau$ 跳跃算法以外,还有很多方法也被用于关于模拟生化系统的随机性.这些方法包括对包含不同时间

尺度问题的快慢尺度随机模拟<sup>[57-62]</sup>,确定性问题和随机过程的混合算法<sup>[15, 63-66]</sup>,稀有时间的数值模拟方法<sup>[67]</sup>和其它求解化学主方程的计算方法等<sup>[68-76]</sup>.此外,求解化学郎之万方程的数值计算方法可以参见文献[77].

## 6 总结

生化系统中系统的状态随反应发生的变化是1个随机过程,化学主方程可以对这一过程给出较好的统计描述.然而,对大的系统,由于维数问题,准确求解化学主方程很困难.关于主方程研究的最新进展及其在基因表达模型中的应用可参考文献[78-79].在不同的条件下可以推导出关于化学主方程的各种近似,包括化学福克尔-普朗克方程、化学速率方程、化学郎之万方程等,其中当随机涨落可以忽略时,化学速率方程可以很好地描述系统的动力学行为.当随机涨落较大时,在一定条件下化学郎之万方程可以对系统的随机行为给出合理的刻画.但是对很多生化反应系统而言推导化学郎之万方程的条件并不成立.因此在对给定的生化反应系统进行建模时如何选择合适的方程描述系统的动力学这一问题还没有明确的答案,还需要进一步研究.

如果考虑到环境的随机涨落,则在生化反应系统的描述中的反应速率都是时变的随机过程,在这种情况下由于外部随机因素的复杂性,没有简单的方法对系统的动力学行为进行描述.在有外部干扰的特殊的情况下可以采用对数正态分布的随机变量描述系统的反应速率常数.然而此时还有很多数学问题需要解决.

在生化反应系统中有很多反应并不是瞬间完成的,而是具有明显的过程.这类反应系统的随机行为并不是马尔科夫过程,对其进行描述需要具有时间滞后的化学主方程.然而相关的数学理论还很不完善,有很多问题值得进一步探讨.

对生化系统的随机模拟的研究已经有比较成熟的结果.随机模拟算法(SSA)作为对生化反应系统的“准确”模拟方法已经广泛应用于各种研究中.然而,这一方法在模拟大规模的生化反应体系时效率很低.目前有很多关于随机模拟算法的加速算法,关于随机模拟方法的综述可参见文献[14].

## 7 参考文献

- [1] Cai Long, Friedman N, Xie X Sunney. Stochastic protein

- expression in individual cells at the single molecule level [J]. *Nature Cell Biology* 2006 440( 7082 ): 358-362.
- [2] Choi P J ,Cai Long ,Frieda K ,et al. A stochastic single-molecule event triggers phenotype switching of a bacterial cell [J]. *Science* 2008 322( 5900 ): 442-446.
- [3] Elf J ,Li GeneWei ,Xie X Sunney. Probing transcription factor dynamics at the single-molecule level in a living cell [J]. *Science* 2007 316( 5828 ): 1191-1194.
- [4] Xie X Sunney ,Choi P J ,Li GeneWei ,et al. Single-molecule approach to molecular biology in living bacterial cells [J]. *Annual Review of Biophysics* 2008 37: 417-444.
- [5] Yu Ji ,Xiao Jie ,Ren Xiaojia ,et al. Probing gene expression in live cells one protein molecule at a time [J]. *Science* , 2006 311( 5767 ): 1600-1603.
- [6] Taniguchi Y ,Choi P J ,Li GeneWei ,et al. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells [J]. *Science* 2010 329( 5 ): 533-538.
- [7] Bassett A R ,Liu Jilong. Crispr/cas9 and genome editing in drosophila [J]. *Journal of Genetics and Genomics* 2014 , 41( 1 ): 7-19.
- [8] Mali P ,Esvelt K M ,Church G M. Cas9 as a versatile tool for engineering biology [J]. *Nature Methods* ,2013 ,10 ( 10 ): 957-963.
- [9] Amabile G ,Meissner A. Induced pluripotent stem cells: current progress and potential for regenerative medicine [J]. *Trends in Molecular Medicine* 2009 15( 2 ): 59-68.
- [10] Efe J A ,Hilcove S ,Kim J ,et al. Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy [J]. *Nature Cell Biology* 2011 13( 3 ): 215-222.
- [11] Takahashi K ,Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors [J]. *Cell* 2006 126( 4 ): 663-676.
- [12] Takahashi K ,Tanabe K ,Ohnuki M ,et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors [J]. *Cell* 2007 131( 5 ): 861-872.
- [13] Fiering S ,Whitelaw E ,Martin D. To be or not to be active: the stochastic nature of enhancer action [J]. *Bioessays* , 2000 22( 4 ): 381-387.
- [14] Gillespie D T. Stochastic simulation of chemical kinetics [J]. *Annu Rev Phys Chem* 2007 58: 35-55.
- [15] Li Hong ,Cao Yang ,Petzold L R ,et al. Algorithms and software for stochastic simulation of biochemical reacting systems [J]. *Biotech Prog* 2008 24( 1 ): 56-61.
- [16] McAdams H H ,Arkin A P. Stochastic mechanisms in gene expression [J]. *Proc Natl Acad Sci* ,1997 ,94 ( 3 ): 814-819.
- [17] McAdams H H ,Arkin A P. It's a noisy business! Genetic regulation at the nanomolar scale [J]. *Trends Genet* , 1999 15( 2 ): 65-69.
- [18] Raj A ,van Oudenaarden A. Nature ,nurture ,or chance: stochastic gene expression and its consequences [J]. *Cell* 2008 135( 2 ): 216-226.
- [19] Rao C V ,Wolf D M ,Arkin A P. Control ,exploitation and tolerance of intracellular noise [J]. *Nature* ,2002 420 ( 6912 ): 231-237.
- [20] Samoilov M S ,Price G ,Arkin A P. From fluctuations to phenotypes: the physiology of noise [J]. *Sci Ske* ,2006 , 2006( 366 ): 17.
- [21] Shahrezaei V ,Swain P S. The stochastic nature of biochemical networks [J]. *Curr Opin Biotechnol* ,2008 ,19 ( 4 ): 369-374.
- [22] Pedraza J M ,van Oudenaarden A. Noise propagation in gene networks [J]. *Science* ,2005 ,307 ( 5717 ): 1965-1969.
- [23] 周天寿. 生物系统的随机动力学 [M]. 北京: 科学出版社 2009.
- [24] Kaern M ,Elston T C ,Blake W J ,et al. Stochasticity in gene expression: from theories to phenotypes [J]. *Nat Rev Genet* 2005 6( 6 ): 451-464.
- [25] Elowitz M B ,Levine A J ,Siggia E D ,et al. Stochastic gene expression in a single cell [J]. *Science* ,2002 ,297 ( 5584 ): 1183-1186.
- [26] Paulsson J. Summing up the noise in gene networks [J]. *Nature* 2004 427( 6973 ): 415-418.
- [27] Golding I ,Paulsson J ,Zawilski S M ,et al. Real-time kinetics of gene activity in individual bacteria [J]. *Cell* 2005 , 123( 6 ): 1025-1036.
- [28] Moran U ,Phillips R ,Milo R. Snapshot: key numbers in biology [J]. *Cell* 2010 141( 7 ): 1262.
- [29] Lei Jinzhi. Stochastic modeling in systems biology [J]. *J Adv Math Appl* 2012 1: 76-88.
- [30] 雷锦誌. 系统生物学: 建模 ,分析 ,模拟 [M]. 上海: 上海科学技术出版社 2010.
- [31] Schlicht R ,Winkler G. A delay stochastic process with applications in molecular biology [J]. *J Math Biol* 2008 57 ( 5 ): 613-648.
- [32] Oppenheim I ,Shuler K E ,Weiss G H. Stochastic and deterministic formulation of chemical rate equations [J]. *J Chem Phys* ,1969 50( 1 ): 460-466.
- [33] Gillespie D T. The chemical Langevin equation [J]. *J Chem Phys* 2000 113( 1 ): 297-306.
- [34] van Kampen N. Stochastic processes in physics and chemistry [M]. 3rd ed. Amsterdam: North-Holland 2007.
- [35] Jahnke T ,Huisinga W. Solving the chemical master equation for monomolecular reaction systems analytically [J]. *J Math Biol* 2007 54( 1 ): 1-26.
- [36] Shahrezaei V ,Swain P S. Analytical distributions for stochastic gene expression [J]. *Proc Natl Acad Sci USA* , 2008 105( 45 ): 17256-17261.



- [37] Lipshtat A ,Loinger A ,Balaban N Q ,et al. Genetic toggle switch without cooperative binding [J]. *Phys Rev Lett* , 2006 ,96( 18) : 188101.
- [38] Vellela M ,Qian Hong. A quasistationary analysis of a stochastic chemical reaction: Keizer's paradox [J]. *Bull Math Biol* 2007 ,69( 5) : 1727-1746.
- [39] Øksendal B. Stochastic differential equations [M]. 6th ed. Berlin: Springer 2005.
- [40] Gillespie D T. Markov processes: an introduction for physical scientists [M]. San Diego: Academic Press ,1992.
- [41] Uhlenbeck G E ,Ornstein L S. On the theory of the brownian motion [J]. *Physical Review* ,1930 ,36: 823-841.
- [42] van Kampen N. Langevin-like equation with colored noise [J]. *J Statis Phys* ,1989 ,54( 5/6) : 1289-1308.
- [43] Crow E L ,Shimizu K. Lognormal distributions: theory and applications [M]. New York: Marcel Dekker Inc ,1988.
- [44] Limpert E ,Stahel W A ,Abbt M. Log-normal distributions across the sciences: keys and clues [J]. *Bio Science* , 2001 ,51( 5) : 341-352.
- [45] Rosenfeld N ,Young J W ,Alon U ,et al. Gene regulation at the single-cell level [J]. *Science* ,2005 ,307( 5717) : 1962-1965.
- [46] Sommer S S ,Rin N A. The lognormal distribution fits the decay profile [J]. *Biochem Biophys Res Commu* ,1979 ,90( 1) : 135-141.
- [47] Tian Tianhai ,Burrange K ,Burrange P M ,et al. Stochastic delay differentiation equations for genetic regulatory networks [J]. *J Comput Apply Math* ,2007 ,205( 2) : 696-707.
- [48] Zhang Xuan ,Jin Huiqin ,Yang Zhuoqin ,et al. Effects of elongation delay in transcription dynamics [J]. *Mathematical Biosciences and Engineering* ,2014 ,11( 6) : 1431-1448.
- [49] Gillespie D T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions [J]. *J Comput Phys* ,1976 ,22( 4) : 403-434.
- [50] Gillespie D T. Exact stochastic simulation of coupled chemical reactions [J]. *J Phys Chem* ,1977 ,81( 25) : 2340-2361.
- [51] Gillespie D T. Approximate accelerated stochastic simulation of chemically reacting systems [J]. *J Chem Phys* , 2001 ,115( 4) : 1716-1733.
- [52] Gillespie D T ,Petzold L R. Improved leap-size selection for accelerated stochastic simulation [J]. *J Chem Phys* , 2003 ,119( 16) : 8229-8234.
- [53] Cao Yang ,Gillespie D T ,Petzold L R. Efficient step size selection for the tau-leaping simulation method [J]. *J Chem Phys* 2006 ,124( 4) : 044109.
- [54] Chatterjee A ,Vlachos D ,Katsoulakis M. Binomial distribution based  $\tau$ -leap accelerated stochastic simulation [J]. *J Chem Phys* 2005 ,122( 2) : 024112.
- [55] Tian Tianhai ,Burrange K. Binomial leap methods for simulating stochastic chemical kinetics [J]. *J Chem Phys* , 2004 ,121( 21) : 10356-10364.
- [56] Cao Yang ,Gillespie D T ,Petzold L R. Avoiding negative populations in explicit poisson tau-leaping [J]. *J Chem Phys* 2005 ,123( 5) : 054104.
- [57] Cao Yang ,Gillespie D T ,Petzold L R. Accelerated stochastic simulation of the stiff enzyme-substrate reaction [J]. *J Chem Phys* 2005 ,123( 14) : 144917.
- [58] Cao Yang ,Gillespie D T ,Petzold L R. The slow-scale stochastic simulation algorithm [J]. *J Chem Phys* 2005 ,122( 1) : 014116.
- [59] Goutsias J. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems [J]. *J Chem Phys* 2005 ,122( 18) : 184102.
- [60] Hu Yucheng ,Abdulle A ,Li Tiejun. Bossted hybrid method for solving chemical reaction systems with multiple scales in time and population size [J]. *Commun Comput Phys* , 2012 ,12( 4) : 981-1005.
- [61] Rao C V ,Arkin A P. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm [J]. *J Chem Phys* 2003 ,118( 11) : 4999-5010.
- [62] Samant A ,Vlachos D G. Overcoming stiffness in stochastic simulation stemming from partial equilibrium: a multiscale monte carlo algorithm [J]. *J Chem Phys* 2005 ,123( 14) : 144114.
- [63] Alfonsi A ,Cancés E ,Turinici G ,et al. Exact simulation of hybrid stochastic and deterministic models for biochemical systems [R]. Paris: Research Report RR-5435 INRIA-00070572 2004.
- [64] Erban R ,Kevrekidis I G ,Adalsteinsson D ,et al. Gene regulatory networks: a coarse-grained equation-free approach to multiscale computation [J]. *J Chem Phys* ,2006 ,124( 8) : 084106.
- [65] Haseltine E L ,Rawlings J B. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics [J]. *J Chem Phys* 2002 ,117( 15) : 6959-6969.
- [66] Salis H ,Kaznessis Y. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions [J]. *J Chem Phys* 2005 ,122( 5) : 054103.
- [67] Lv Cheng ,Li Xiaoguang ,Li Fangting ,et al. Constructing the energy landscape for genetic switching system driven by intrinsic noise [J]. *PLoS One* 2014 ,9( 2) : e88167.
- [68] Deudlhard P ,Huisinga W ,Jahnke T ,et al. Adaptive discrete galerkin methods applied to the chemical master equation [J]. *SIAM J Sci Comput* ,2008 ,30( 6) : 2990-3011.
- [69] Engblom S. Galerkin spectral method applied to the chemical master equation [J]. *Commun Comput Phys* 2009 ,5

- (5): 871-896.
- [70] Engblom S. Spectral approximation of solutions to the chemical master equation [J]. J Comput Appl Math, 2009 229(1): 208-221.
- [71] Hegland M, Hellander A, Lötstedt P. Sparse grids and hybrid methods for the chemical master equation [J]. 2008, 48(2): 265-283.
- [72] Jahnke T. An adaptive wavelet method for the chemical master equation [J]. SIAM J Sci Comput 2010 31(6): 4373-4394.
- [73] Jahnke T, Huisinga W. A dynamical low-rank approach to the chemical master equation [J]. Bull Math Biol 2008, 70(8): 2283-2302.
- [74] Khoo C F, Hegland M. The total quasi-steady state assumption: its justification by singular perturbation and its application to the chemical master equation [J]. ANZIAM J 2008 50: 429-443.
- [75] Munsky B, Khammash M. A multiple time interval finite state projection algorithm for the solution to the chemical master equation [J]. J Comput Phys 2007 226(1): 818-835.
- [76] Zhou Tianshou, Zhang Jiajun. Analytical results for a multistate gene model [J]. SIAM J Appl Math 2012 72(3): 789-818.
- [77] Kloeden P E, Platen E. Numerical solutions of stochastic differential equation [M]. New York: Springer-Verlag, 1992.
- [78] 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.
- [79] 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.

## The Modeling and Analysis of Biochemical Reaction Systems

LEI Jinzhi

(Zhou Peiyuan Centre for Applied Mathematics, Tsinghua University, Beijing 100084, China)

**Abstract:** Many cellular behaviors are regulated by intracellular biochemical reactions such as gene expression and protein-protein interactions. The kinetics of these biochemical reactions are mostly random, and therefore the stochasticity is essential for the modeling of cellular behaviour in the field of computational biology. This paper is a self-contained review trying to provide an overview of stochastic modeling and recent advance in computation biology. This paper starts from basic assumptions in biochemical reaction systems and introduce mathematical formulations for modeling intrinsic noise, external noise and reactions systems with delay. Multiple dynamical equations are discussed, including chemical master equation, chemical Fokker-Planck equation, chemical reaction rate equation, and chemical Langevin equation, etc. Several numerical methods are introduced, including the stochastic simulation algorithm (SSA) and the Tau-leaping algorithm.

**Key words:** systems biology; stochastic simulation; chemical master equation; biochemical reaction; stochastic differential equation

(责任编辑: 王金莲)