

随机数学理论与单分子和单细胞生物物理化学的交叉

葛颖

北京大学北京国际数学研究中心和生物医学前沿创新中心, 北京 100871

E-mail: haoge@pku.edu.cn

收稿日期: 2023-02-23; 接受日期: 2023-05-11; 网络出版日期: 2023-08-15

国家自然科学基金 (批准号: 11971037 和 T2225001) 资助项目

摘要 随机数学理论与生物物理化学的交叉是近二十年兴起的新方向. 本文首先简要回顾随机数学理论和生物物理化学—尤其是单分子和单细胞层面的生物物理化学—相交叉的历史, 然后分别从本文作者所熟悉的几个侧面介绍近二三十年以来该交叉学科的进展, 其中也包括本文作者课题组所做过的一些工作.

关键词 随机数学理论 生物物理化学 单分子 单细胞 学科交叉

MSC (2020) 主题分类 60G10, 60J28, 92C37

1 引言

随机过程是从 20 世纪初开始发展起来的一门数学分支, 从它诞生开始就一直与物理学、化学乃至生物学有着千丝万缕的联系. 例如, 在物理学方面, Einstein^[21] 首先从概率分布的角度研究了扩散和 Brown 运动, 这也被认为是随机过程理论的起源之一. 随后, 物理化学家 Langevin^[69] 引入了随机力 (白噪声), 提出了单个胶体粒子在溶液中运动的随机动力学方程, 称为 Langevin 方程, 这也是随机微分方程的起源之一. 时至今日, Langevin 方程依然是重要的分子动力学模型之一. 20 世纪 30 年代, 物理化学家 Onsager^[87] 利用随机过程的可逆性加上线性化思想推导出了著名的倒易关系, 是第一个在近平衡系统中广泛成立的规律, 并因此获得了 1968 年的诺贝尔化学奖. 之后, Onsager 和 Machlup 合作, 应用 Onsager 的理论进一步推导出了线性系统随机轨道的路径积分, 这个关系与后来随机过程中的大偏差理论有着千丝万缕的联系 (参见文献 [77, 88]). 20 世纪中叶, 化学家 Flory^[27, 28] 利用随机过程、统计物理和物理化学的知识, 系统性地定量分析了高分子聚合物的特性. Flory 是材料化学的鼻祖之一, 并因此获得了 1974 年的诺贝尔化学奖. 不仅如此, Flory 在研究单个聚合物分子物理化学性质的过程中, 还提出了自回避随机游动的数学模型, 而这个模型随后在随机过程领域引起了不小的反

英文引用格式: Ge H. The intersection of stochastic mathematical theory and single-molecule/single-cell biophysical chemistry (in Chinese). *Sci Sin Math*, 2023, 53: 1163–1180, doi: 10.1360/SSM-2023-0038

响, 至今该模型的尺度变换极限与 Schramm-Loewner 演化 (Schramm-Loewner evolution, SLE) 对应关系的严格证明仍然是随机过程领域尚未解决的重要问题之一. 一直以来, 聚合物模型都是离散概率论中的重要模型, 随机偏微分方程中也有连续化了的聚合物模型, 化学已经深刻地与严格随机数学理论相结合. 在 2022 年国际数学家大会上, 就有若干一小时和四十五分钟报告涉及聚合物模型.

在生物学方面, 最负盛名的例子是 20 世纪 40 年代, 实验生物学家 Luria 和物理学家 Delbrück 的工作, 他们利用随机过程中二项分布逼近 Poisson 分布的事实, 结合噬菌体的实验观察, 第一次在实验上证实了变异是随机自发发生的, 即在环境变化之前就产生了, 而且可以遗传, 并利用分支过程的随机模型计算出了细菌变异的速率 (参见文献 [75]). 因为这项工作, Luria 和 Delbrück 获得了 1969 年的诺贝尔生理学或医学奖. 要知道在那个年代, 遗传物质脱氧核糖核酸 (DNA) 的结构还未曾知晓. 直到近些年, 对于这种被称为 Luria-Delbrück 模型及其分布的研究依然还会发表在著名的交叉学科刊物和统计学刊物上 (参见文献 [52, 64]). 还有一个很有趣的例子是, 1972 年, 三位化学家 Magde 等^[79] 利用 Brown 运动的相关知识和自相关函数方法发明了荧光相干光谱 (fluorescence correlation spectroscopy, FCS) 的实验技术, 这已成为研究亚宏观化学反应和扩散过程的标准实验手段, 这也是随机过程知识对于实验技术手段的一次经典贡献.

早在 1940 年, 还有两项里程碑的工作对于日后随机过程和生物物理化学的交叉起到了极大的推动作用. 一项是物理学家 Kramers^[66] 利用 Langevin 方程对于描述单个大分子结构变化的异构化学反应建模, 并在 Reynolds 数比较小 (恰好对应于单个细胞和单个分子的情形) 时, 计算出了异构化学反应的反应速率, 可以说是第一次用随机模型验证了 20 世纪 30 年代提出的著名的化学反应过渡态理论. 但可惜的是, 那时人们对于大分子的化学反应知道的不多, 特别是对于完全由随机性驱动的化学反应更是几乎没有认识, 那个时候正是量子效应驱动的化学反应理论如日中天的时候, 其杰出代表是物理化学家 Marcus 和 Sutin^[81] 关于电子转移过程的精彩理论, 他也因此获得了 1992 年诺贝尔化学奖. 直到 20 世纪七八十年代, Kramers 的理论才慢慢被重新发现和推广, 那是因为人们可以观察到单个蛋白质分子的结构变化了, 而其结构变化之大已经脱离了量子驱动的范围, 必须用热运动和随机过程来描述和分析. 在这个过程中, 一些随机过程理论中所独有的方法和模型也被广泛应用, 如首达时方法以及与之相关的拟稳态理论, 还比如推广的 Langevin 方程, 这已经失去了 Markov 性, 且需要用到分数阶 Brown 运动. 关于 Kramers 理论的严格数学推导, 也是到了 2000 年左右才最终完成的, 当然数学家得到的结果要更一般, 不仅刻画了化学反应速率的渐近性质, 还刻画了化学反应等待时间的渐近指数分布性质 (参见文献 [30]). 2004 年, 哈佛大学寇星昌和谢晓亮课题组^[65] 合作, 利用推广的 Langevin 方程, 漂亮地分析了单个蛋白质分子内部各原子间距离随时间的随机变化规律, 这也是随机过程与生物化学实验相结合的典范之作.

1940 年的另一项里程碑式的工作是 Delbrück^[15] 提出了可以用化学中的主方程模型来刻画细胞的分裂和增殖以及变异. 以此为开端, 化学家们纷纷开始用化学主方程模型来刻画生物化学系统内各种化学物质的分子数目随时间的变化, 或者所有可以用类似化学反应来刻画的系统的随机动力学. 化学主方程是一类与化学反应相对应的、只取非负整数值的连续时间 Markov 跳过程, 其分布随时间的演化方程. 到了 20 世纪 60 年代, 这个模型引起了一些数学家的关注, 是物理学家也是数学家的 McQuarrie^[82] 在《应用概率期刊》上介绍了化学主方程模型, 而概率学家 Kurtz^[67] 利用他自己发展的一系列概率方法严格证明了化学主方程的大体积极限正好就是化学界著名的质量作用定律. 20 世纪 70 年代, Kurtz^[68] 还进一步证明了化学主方程在大体积下收敛到非时齐的扩散过程. 以 Kurtz 为代表的威斯康辛学派之后几十年一直没有停止对化学主方程数学性质的研究, 近些年来以应用数学家和概率学家 Anderson 为代表的一批数学家, 在化学反应系统的确定性和随机性模型的稳定性、平稳

分布及其数值算法方面都有不少进展 (参见文献 [1]). 在这些数学理论发展的同时, 从 20 世纪 70 年代末开始, 就有生物物理学家开始用化学主方程模型对于单细胞中的转录翻译等过程建立模型, 推导出了不少有趣的结果, 但是直到 20 世纪 90 年代末, 当人们开始可以测量到单个细胞中化学分子数目随时间的变化过程之后, 化学主方程模型才真正得到广泛应用, 而物理化学家 Gillespie^[48] 在 20 世纪 70 年代中叶提出的化学主方程随机模拟方法也一下子名声大噪, 其背后的原理是概率学家 Doob^[16] 在 20 世纪 50 年代证明了的数学定理. 到了 21 世纪, 随着测量技术及合成生物学的突飞猛进, 单分子和单细胞实验观测与随机数学模型相结合已经成为了诸多生物学论文的标准配置.

总之, 近十多年来, “随机数学理论与生物物理化学的交叉” 已经逐渐发展成为了一个独立的交叉学科方向, 这是因为随着单分子和单细胞实验技术的突飞猛进, 人们已经可以定量观测到从单个分子到单个细胞尺度的随机现象 (轨道), 使得随机过程非常自然地成为了描述这些亚宏观尺度现象的数学模型, 而且这些非平衡亚宏观生命活动中蕴含着丰富的未知随机规律. “随机数学理论和生物物理化学的交叉” 领域, 也是在过去这二十年间, 数学和生物学不断交叉融合的大背景下诞生的若干新兴学科之一. 2012 年 4 月, *Science* 杂志专门组织了一期计算生物学特刊, 其中有一篇综述文献讲的就是如何利用基因表达中的随机性来理解基因调控^[84]. 2015 年 11 月, 美国数学会的 *Notice* 上刊登了 3 篇生物数学特稿, 并在当期封面推荐阅读, 其中杜克大学数学教授 Reeds^[98] 的文章标题是 “Mathematical biology is good for mathematics”. 其实早在 2004 和 2005 年的美国数学会 *Notice* 上, 生物数学家 Reeds^[97] 及 Mackey 和 Santillán^[78] 就曾分别撰文叙述了数学和生物、物理、化学等的学科交叉. 2015 年 12 月, 本文作者也曾应邀在 *Cell* 杂志的 *Voices* 栏目中撰写短文向生物学家们介绍自己工作中所用到的数学工具、数学模型以及取得的科研成果, 表示生物学界也越来越重视数学在其中发挥的作用^[35]. 紧扣实际问题, 与具体实验数据和实验现象紧密结合, 近年来越来越成为交叉学科发展的趋势和方向, 而在这一过程中, 数学已经并将不断地为其他学科创造出真正有意义的应用.

本文作者将从自己的科研方向和研究内容出发, 从几个侧面综述对于随机数学理论与单分子和单细胞生物物理化学的这一交叉领域的理解. 还有很多重要的领域和方向, 因为本文作者并不十分熟悉, 没有做过或者刚开始涉及相关的科研, 这里不展开讨论, 如基于单细胞基因组的遗传学、基于单细胞转录组、三维基因组等各类组学数据的生物信息学等^[14], 都需要应用随机数学模型和统计学的知识. 下面介绍的交叉学科工作, 总体而言, 可以粗略地分成两类. 一类是理论味道较为浓厚的, 数学和物理的模式更加凸显, 追求的是理论上的创新, 发现新的生物物理化学规律, 或者给出已知的生物物理化学规律的严格证明及其成立的条件, 也比较重视数学的严格性. 而另一类则更多的是针对具体生物物理化学问题的建模、模拟和分析, 对于随机过程艰深理论的要求相对不高, 更在乎实用性而不是严格性, 不过背后都有严格数学理论的支撑. 在这类工作中, 难点在于理解实验和实验学家的需求, 以及如何将基于实验观察的科学直观转变为扎实的数学模型.

2 亚宏观尺度随机热力学理论及其应用

非平衡态热力学和统计物理理论从 20 世纪七八十年代开始进入了蓬勃发展的阶段, 特别是其中的随机模型可以用来描述亚宏观尺度的生物物理化学过程, 恰好契合了现代物理化学和生物化学高精度实验技术的发展潮流. 非平衡态随机数学理论, 能够很好地被应用于细胞层面重要的生物化学过程, 通过严格的数学论证探寻其中重要的规律, 并对生命现象获得更深刻的认识和理解.

生命体到底遵从什么样的物理学规律? 这是从 19 世纪末开始, 科学家们就很关心的问题. 20 世纪

40 年代, 物理学家 Schrödinger^[103] 提出了“负熵”的概念, 用于解释为什么生命体可以长期处于有序的状态. 将这一概念定量化和系统化的是物理化学家 Prigogine, 他从热力学角度提出了非平衡态定态和熵产生这些概念, 并给出了熵产生的一般表达式^[86,93]. 当然, 在 Prigogine 之前, Eckart、Bridgman 和 Onsager 等都作出了重要的贡献^[3,4,19,20]. 遗憾的是, Prigogine 的理论比较抽象和难于直接应用. 美国物理化学家 Hill^[54] 在 20 世纪 60 年代第一次将随机过程引入了亚宏观的单分子生物化学系统, 研究了细胞膜上的离子通道蛋白和肌肉中的肌动蛋白等, 并开创性地使用环动力学的观点, 真正阐释了熵产生本身的物理内涵和具体表达式. 后来, 物理学家 Schnakenberg^[102] 在比较一般的模型上推广了 Hill 的理论. 受 Hill 和 Schnakenberg 工作的启发, 并结合 20 世纪 30 年代 Kolmogorov 关于 Markov 过程可逆性的探讨, 从 20 世纪 70 年代末开始, 以钱敏教授为代表的一批中国数学家以平稳时齐 Markov 链和扩散过程为模型, 建立了非平衡定态的严格数学理论^[62,95].

这套数学理论的核心内容是平稳时齐 Markov 过程的时间可逆性, 也就是将时间倒过来, 得到的新的随机过程的统计性质是否与原过程相同, 相同就被称为可逆, 对应于热力学平衡态. 可逆性等价于物理量熵产生为 0, 也等价于每个状态环上的稳态流量都是 0 (参见文献 [62]). 更进一步地, 他们还第一次从轨道层面严格定义了熵产生, 即正过程和逆过程的 Radon-Nikodym 导数的对数, 并证明了其大数定律, 即沿轨道的熵产生为

$$Ep(t) = \frac{1}{t} \log \frac{dP_{[0,t]}}{dP_{[0,t]}^-} \rightarrow epr, \quad t \rightarrow \infty,$$

其中, $P_{[0,t]}$ 是原过程的测度, $P_{[0,t]}^-$ 是时间倒逆过程的测度, epr 是单位时间的熵产生率. 值得一提的是, 直到十几年后, 欧美才有统计物理学家开始在轨道层面定义热力学量, 并兴起了“随机热力学”这一新的领域 (参见文献 [13, 83, 104–106]).

但是, 平稳时齐随机过程只能描述单个的热力学定态, 无法刻画趋于定态的暂态过程以及两个定态之间的转移过程, 后者必须要利用非时齐或非平稳随机过程来描述, 其中就包括 20 世纪 90 年代左右发现的很多重要统计物理规律, 如刻画两个平衡态之间自由能改变的 Jarzynski 等式和 Crooks 等式等 (参见文献 [13, 53, 57]). 以这些等式为代表的一系列涨落定理, 被认为是过去 30 年统计物理学界最重要的发现之一, 因为它们将原本不等式形式的热力学第二定律改进成了等式, 揭示了远离平衡态系统的随机轨道必须满足的普遍统计规律. 2008 年, 以非时齐连续时间 Markov 链和多维扩散过程为模型, Ge 和 Jiang^[38] 给出了两个平衡态间转移过程所满足的 Jarzynski 等式、Crooks 等式和两个非平衡定态间转移过程所满足的 Hatano-Sasa 等式的数学描述及其严格证明. Ge 和 Jiang 的证明基于非时齐版本的 Feymann-Kac 公式. 我们所用的数学技巧后来还被其他一些物理学者用来证明更一般的涨落定理. 同样在 2008 年, 两位欧洲概率学家, 同时也是物理学家, Chetrite 和 Gawędzki^[10] 也给出了涨落定理的证明, 他们的方法在数学上更复杂, 结果也更一般, 涵盖了 Jarzynski 等式和 Hatano-Sasa 等式之外的涨落定理.

这些涨落定理作为热力学第二定律在非平衡态的等式表现形式, 对于过程的初分布和末状态等都有严格的条件限制, 但是热力学第二定律作为一个普适的物理规律, 理应有更一般的系综 (即期望的意义下) 不等式表现形式, 且对于每个瞬时都必须成立. 在 2007–2010 年, 包括本文作者和华盛顿大学钱纮教授在内的好几组国际研究人员, 几乎同时发现了熵产生的非负分解公式, 并由此得出了热力学第二定律在非平衡态情形下的一般系综表现形式, 即推广形式的 Clausius 不等式, 这要比平衡态热力学中的 Clausius 不等式 (即熵产生率非负) 更强 (参见文献 [24, 25, 32, 40]). 尽管所立足的物理现象并不相同, 有基于量子点 (quantum dot) 的, 有基于化学反应的, 但最终都是根据随机数学模型得到了同

样的分解式. 这个分解式也称为是热力学第二定律的三个侧面. 具体而言, 首先是在任意的时刻 t , 可以定义瞬时的熵产生率 $epr(t)$, 其次是它可以分解成两个非负项 Q_{hk} 和 $f_d = -\frac{dF}{dt}$, 其中, Q_{hk} 刻画的是外部的驱动力, F 是瞬时分布与平稳分布之间的相对熵, 因此 f_d 刻画的是相对熵变化的速率.

近十年来, 涨落定理又有了进一步的发展, 如发现了包含反馈机制或者对称性破缺的随机模型中各种热力学量所满足的等式关系等 (参见文献 [99, 100]). 与此同时, 还有一类叫作“热力学不确定关系”的不等式被提出来, 刻画的主要是随机过程中的随机流的方差与熵产生率之间的大小关系, 其严格证明用到了随机过程的 2.5 级大偏差原理, 其重要意义在于随机流是可以被测量的, 这类不等式至少可以让我们测量出熵产生率, 也即能量耗散率的某种下界 (参见文献 [49, 55]).

这几年, 围绕着随机化学反应系统非平衡态性质的研究也方兴未艾, 很多国际研究团队, 都在这方面有最新的进展 (参见文献 [94, 96, 119]). 单个细胞本质上就是一个亚宏观化学反应系统, 因此这些非平衡态理论上的发展, 都有可能帮助我们对于生命运动产生更深刻的认知. 我们从 2016 年就着手研究化学反应系统的热力学, 将随机非平衡态一般理论应用于亚宏观非线性化学反应系统的化学主方程模型上, 然后利用概率论和随机过程中的大偏差理论, 严格证明了在宏观极限下, 这套随机非平衡态理论中的 3 个最主要的热力学泛函的宏观极限分别是平稳分布的大偏差速率函数和宏观熵产生率, 以及一个之前不为人所知的化学热力学泛函^[43, 44], 即

$$\frac{d\phi^{ss}(x)}{dt} = cmf(x) - \sigma(x),$$

其中, $\phi^{ss}(x)$ 是平稳分布宏观极限下的大偏差速率函数, $cmf(x)$ 是 Q_{hk} 的宏观极限, $\sigma(x)$ 是 $epr(t)$ 的宏观极限.

这三者中, 宏观熵产生率 $\sigma(x)$ 是早就熟知的宏观热力学泛函, 而平稳分布的大偏差速率函数 $\phi^{ss}(x)$ 恰好是亚宏观自由能函数的宏观极限, 所以称之为推广的宏观自由能. 更进一步地, 推广的宏观自由能随时间的变化率等于第三个新的热力学泛函 $cmf(x)$ 减去宏观熵产生率, 由此, 可以称这个新的热力学泛函为推广的宏观自由能的输入速率, 且这一项是非负的, 它等于 0 等价于化学中判断平衡态的著名的 Lewis 环条件^[71]. 因此, 该新宏观热力学泛函也是对非平衡态的新的刻画, 且其表达式中含有平稳分布的大偏差速率函数, 因此仅从宏观层面研究确定性化学反应系统, 是不可能得到这样一个新的热力学泛函的.

过去这十年, 随机热力学领域还有两个有意思的发展. 一个是物理学家注意到有一些沿随机轨道定义的热力学量放到指数上可以构成鞅, 然后利用鞅的停时定理和 Doob 的鞅不等式可以得到这些热力学量更为细致的统计性质 (参见文献 [85]). 对于放到指数上并不构成鞅的热力学量, 2021 年有一项重要的工作指出, 可以增加一项, 使其变成鞅, 然后应用停时定理可以得到新的涨落定理和热力学第二定律的形式 (参见文献 [80]). 另一个是对于多尺度的随机过程, 研究其收敛到较慢尺度的随机动力学过程中, 热力学量的收敛是什么样的, 会不会出现反常的现象 (参见文献 [7]). 从 2014 年开始, 我们就着手研究多尺度二阶随机微分方程的非平衡态统计物理性质, 利用函数空间的算子投影和渐近展开的方法, 证明了当此二阶随机微分方程在过阻尼的情形下趋于所对应的一阶随机微分方程时, 熵产生率的极限的确会多出非负的一项 (参见文献 [34]). 这一项刻画的是空间温度不均匀情形下对熵产生率的贡献, 物理文献中被称为反常熵产生率. 最近, 我们又在此基础上, 以非时齐多尺度的扩散过程为模型, 发现和证明了其中蕴含的新热力学结构, 给出了非常一般的严格数学理论. 具体而言, 我们严格证明了被定义为原始过程的分布与另一个可比的扩散过程分布之间的对数 Radon-Nykodim 导数的一般热力学泛函, 在二阶平均化极限下的反常部分必然是鞅, 并借助于该鞅结构, 证明了反常部分所满足的停时积分涨落定理^[37]. 作为该数学理论的物理应用, 我们证明了沿轨道定义的熵产生 epr 和内部

热 Q_{hk} 作为重要的热力学泛函, 满足鞅结构, 并推广到了含有奇变量的情形, 涵盖了之前关于二阶随机微分方程 (也称 Kramers-Langevin 方程) 的所有相关物理结果.

从数学的角度来看, 这些热力学量的统计物理性质还有一些尚未得到的完整的严格数学证明. 例如, 在 1999 年, 统计物理学家 Lebowitz 和 Spohn^[70] 基于随机 Markov 模型提出了定态涨落定理, 其中的核心就是沿着随机轨道的熵产生满足大偏差原理. 后来, Jiang 等^[63] 证明了离散状态 Markov 过程的熵产生满足大偏差原理, 但是对于 Euclid 空间中的扩散过程, 其熵产生是否满足大偏差原理还未得到严格的证明, 主要是因为全空间并不是紧的. 我们在这方面做过一些努力, 例如, 2016 年, Chen 等^[8] 证明了一类二维扩散过程的熵产生满足大偏差原理. Wang 等^[112] 还证明了 Euclid 空间扩散过程熵产生的中偏差原理. 一般 Euclid 空间扩散过程沿轨道熵产生的大偏差原理仍然还是开放问题.

3 无反馈机制的单细胞中心法则随机模型及其应用

在 20 世纪八九十年代, 人们已经可以观测到单分子的运动, 而由于受到溶液分子热运动的影响, 单个分子的运动本身就是随机的, 对其动力学行为的定量刻画就必然需要建立随机模型. 到了 21 世纪, 随着荧光标记实验技术的突飞猛进, 可以高分辨率地记录下单个细胞内部单个化学物质或者若干种化学物质的分子数目随时间的变化. 而单个细胞内部的化学反应所处的时间尺度大约在毫秒到微秒之间, 而空间尺度大约在微米到纳米之间, 这正是随机规律起作用的时空尺度. 于是随机模型开始进入生物学领域, 围绕着单细胞中心法则给出了很多定量的、而又与实验十分吻合的精彩数学结果.

生物学中的中心法则, 就是脱氧核糖核酸 (DNA) 会自我复制, 可以转录成信使核糖核酸 (mRNA), 信使核糖核酸又可以翻译成蛋白质的过程. 因此最简单的中心法则随机模型 (不包括 DNA 复制), 就是将这个过程的一步都简化成一个一步的化学反应来处理 (图 1(a)). 对应到连续时间跳过程, 就是每个化学反应的等待时间是随机的, 服从指数分布, 该平均等待时间的倒数就称为反应速率. 这样建成了最简单的一个化学主方程模型, 以信使核糖核酸和蛋白质分子的数目来描述系统的状态. 当然, 这是一个极其简化的模型, 因为中心法则的每一步本质上都是很多很多步化学反应构成的, 即使在原核生物中也是如此. 如此简化的模型, 其在细节上的确无法完美地刻画中心法则的过程, 但是该模型整体上的一些定性结果, 甚至是定量结果, 却可以很好地符合实验, 这也是一件很神奇的事情.

单细胞转录翻译机制所对应的最简单化学主方程, 需要用二维向量 (m, n) 表示单细胞中有 m 个 mRNA 分子和 n 个蛋白质分子的状态. $p(m, n, t)$ 表示单细胞中有 m 个 mRNA 分子和 n 个蛋白质分

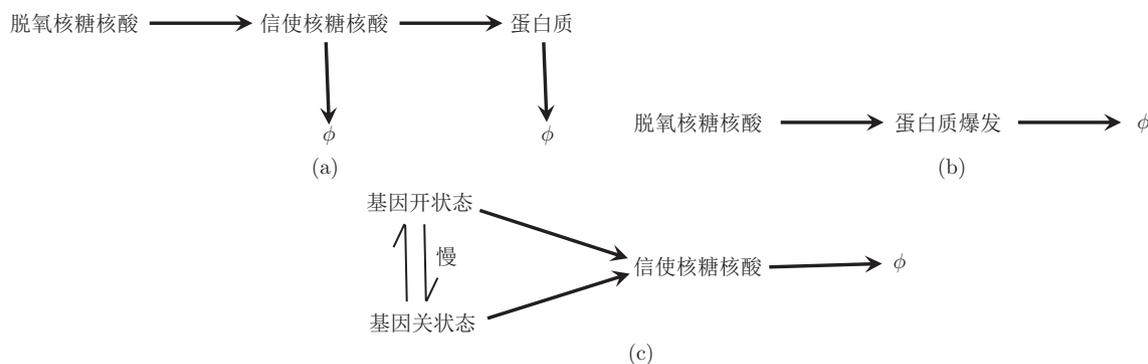


图 1 (a) 最简单的单细胞转录翻译模型; (b) 翻译爆发模型; (c) 转录爆发模型

子的概率, 满足方程

$$\begin{aligned} \frac{dp(m, n, t)}{dt} = & k_1 p(m-1, n, t) - k_1 p(m, n, t) + d_1(m+1)p(m+1, n, t) \\ & - d_1 m p(m, n, t) + k_2 m p(m, n-1, t) - k_2 m p(m, n, t) \\ & + d_2(n+1)p(m, n+1, t) - d_2 n p(m, n, t), \end{aligned}$$

其中, k_1 和 k_2 分别是转录和翻译的速率, d_1 和 d_2 是信使核糖核酸和蛋白质分子的降解速率.

2004 年, 瑞典生物学家 Paulsson (他本科是数学和生物的双学位) 求解了该化学主方程模型在平稳情形下单细胞内信使核糖核酸和蛋白质分子数目的均值和方差, 然后用变异系数的平方, 也就是分子数概率分布的方差除以均值的平方, 来作为度量随机性大小的量, 并以此厘清了当时互相矛盾的一些实验结果, 指导实验学家如何正确地处理数据^[89]. 除了均值和方差以外, 这个模型的化学主方程还可以计算出单细胞中信使核糖核酸和蛋白质分子数目之间的相关系数. 该相关系数在 2010 年由哈佛大学谢晓亮实验室第一次测量到, 他们发现大多数信使核糖核酸的数目和它们对应的蛋白质数目之间的相关系数都非常小^[108]. 根据以上这样一个最简单模型得出的相关系数的表达式, 可以推断, 该现象很可能是因为信使核糖核酸的降解速率很大, 使得同一时刻测到的蛋白质和信使核糖核酸数目之间并不存在对应关系.

单细胞内的单分子精度的实验测量到的都是真正的概率分布, 其信息远比若干个数字特征要多得多. 但是即使是如此简单的模型, 其概率分布的表达式却十分复杂, 含有特殊函数, 使得具体应用起来十分困难. 首先在这方面有所突破的是哈佛大学谢晓亮实验室^[5, 121], 他们在 2006 年首次观察到了单细胞中单个蛋白质分子合成过程中的爆发现象, 即在每等待一个随机的时间间隔后, 该单细胞内部会在很短的时间内密集地合成若干个该蛋白质分子, 然后再次进入等待状态 (图 1(b)). 经过精巧的实验测量, 实验学家发现每次蛋白质合成的爆发事件背后所对应的信使核糖核酸分子个数, 至少在细菌中可以近似为 1. 而且每个信使核糖核酸分子的生存时间相比较于相邻两个信使核糖核酸生成之间的时间间隔是很短的, 因此每一次的爆发强度, 其实就是每一个信使核糖核酸分子在其被降解之前所能产生的蛋白质分子数目, 而相邻两次爆发之间的时间间隔, 就是相邻两个信使核糖核酸分子合成的时间间隔.

于是可以建立一个只基于这些现象而且只关心蛋白质的生成与降解的唯象随机模型, 即延续最简单模型中的假设, 假设每隔一个指数分布的随机时间, DNA 会合成出一个信使核糖核酸分子, 但这个信使核糖核酸分子的生存时间极短, 可以忽略不计, 且在其生存时间内会以一定的概率分布产生出一定量的蛋白质分子. 现在的问题就是单个信使核糖核酸分子在降解之前能产生出的蛋白质分子数目服从怎样的概率分布呢? 这是一个典型的随机过程数学问题, 即假设每个信使核糖核酸分子是在等待一个指数分布随机时间后消亡, 而在这段生存时间内, 合成出蛋白质分子的过程服从随机过程中经典的 Poisson 过程. 于是可以计算出在一次蛋白质分子的爆发过程中产生蛋白质分子个数的概率服从几何分布. 其实, 这个分布早在 20 世纪 70 年代末, 就由单细胞随机模型的先驱们给出了文献 [2], 只是那个时候还完全是一个猜测. 终于, 基于翻译爆发现象给出了一个简化的只含有蛋白质数目的随机模型, 这是一个一维但非紧邻的连续时间跳过程. 有意思的是, 早在 2000 年, Paulsson 和 Ehrenberg^[90]就在并无任何实验启发的情形下构建了这个随机模型, 并计算出了其长时间后的平稳分布, 发现是负二项分布. 这是一个令人吃惊的结果, 因为一维非紧邻跳过程一般无法算出一个具有简单表达式的平稳分布, 不禁让人感叹大自然的神奇. 这个负二项分布中的两个参数也都具有生物学意义, 一个是每个细胞周期时间内平均的爆发次数, 而另一个则是每次爆发所生成的蛋白质分子数的平均值. 这两个数字都是可以在实验中直接测量的, 最后的平稳分布也是可以测量的. 在 2006 年, 谢晓亮实验室分别测

量了这 3 个量, 发现它们与该理论符合得很好^[5, 121]. 当然, 并不是说这就意味着该模型中的所有假设都是对的, 但这的确表明, 通过这样一个基本符合实验观测的简单模型, 可以抓住真正的复杂的生命过程中的一些重要定量性质. 2019 年, Jia 等^[61] 从数学上还证明了该模型正是最简单机制模型在一定条件下的极限, 进一步建立了这个唯象的一维模型与最简单的二维模型之间的联系.

在无反馈机制的情形下, 无论是最简单的二维模型, 还是唯象的一维模型, 单细胞内蛋白质分子数目的分布都已经定量刻画得很好. 但是, 有意思的是, 对于信使核糖核酸分子数目的分布, 最简单的化学主方程模型所给出的平稳分布是 Poisson 分布, 这是偏离实际的. 2010 年谢晓亮实验室发现, 实际中信使核糖核酸分子个数的平稳分布的方差明显大于均值, 这与 Poisson 分布很不一样^[108]. 其背后的机理被猜测和讨论了很多年. 2005 年, 普林斯顿大学的 Cox 实验室首次观测到了单个细胞中的单个信使核糖核酸分子的生成和降解现象, 并且发现了一种名为转录爆发的现象^[50]. 转录爆发, 顾名思义, 类似于蛋白质合成中的翻译爆发现象, 说的是, DNA 似乎存在着某种神奇的机制, 它会使得即使在完全不存在抑制蛋白结合的情形下, DNA 也会时不时地偷懒不工作, 即转录停止, 过一段时间后再重新开启. 而这种 DNA 自身具有多种不同的转录状态的现象, 并未包括在前面所述的最简单的机制模型中, 也许加入了这种 DNA 活跃和不活跃状态间的切换后, 信使核糖核酸分子数目的分布就会更接近事实. 的确如此, 早在 20 世纪 90 年代, 就有学者提出了具有两个不同的 DNA 转录状态的随机模型, 即俗称的两状态模型, 并通过简单的计算得到了信使核糖核酸的方差和均值, 在该模型下, 方差明显比均值要大. 但是两状态模型是一个一般模型, 并无法解释转录爆发这个现象背后的分子机制(参见文献 [91]).

要解释清楚该分子机制的发现和论证过程, 以及其中两状态随机数学模型又发挥了怎样不可或缺的作用, 需要从 2011 年讲起. 2011 年, 哈佛大学庄小威和谢晓亮实验室合作, 利用超高分辨率的显微成像技术, 发现即使是在细菌中, DNA 分子也不是自由的, 它们会被分成一段一段的并锚定在一些大的蛋白质分子上^[113]. 而早在 20 世纪 80 年代, 人们就发现在 DNA 转录的过程中, 处于核糖核酸 (RNA) 聚合酶前方的 DNA 链将聚集所谓的正超螺旋 (positive supercoiling), 通俗而言就是 DNA 双螺旋结构原本是每 10.5 个碱基对转一圈 (360 度), 而现在被转得越来越紧, 每转一圈的碱基对数目越少, DNA 形变就越厉害; 与此相对应的, 处于 RNA 聚合酶后方的 DNA 链将产生负超螺旋 (negative supercoiling), 即完成一圈的碱基对数目越来越多. 这两种 DNA 形变将由于 DNA 被锚定在一些大分子上而无法相互抵消. 因此细胞内需要有两种酶, 拓扑异构酶 (topoisomerase IA) 专门负责在 RNA 聚合酶的后方释放负超螺旋, 而旋转酶 (gyrase) 则专门负责在 RNA 聚合酶的前方释放正超螺旋. 关于 DNA 双螺旋结构的这种变化现在已经是经典分子生物学教材中的必备内容.

有研究表明, 拓扑异构酶的活性很高, 而旋转酶的活性不高, 而且旋转酶在活细胞内的分子数也并不十分多, 平均到每一段被锚定的 DNA 片段的话只有大约一个旋转酶分子. 因此可以合理猜测, 也许就是因为这一个旋转酶分子会时不时地从这段 DNA 片段上掉落下来, 从而导致在转录过程中积累起大量的正超螺旋, 最终从力学的角度阻止进一步的转录发生. 2014 年, 谢晓亮实验室发展了一套高通量的体外单分子荧光技术, 可以实时高分辨率地观测到单个分子上正在发生的转录过程. 通过这一技术发现, 转录过程在 RNA 聚合酶前段不断聚集起来的正超螺旋, 会渐渐地减慢 RNA 聚合酶的延伸速度, 并最终彻底阻止转录的起始. 而旋转酶和 DNA 分子的结合则可以使得转录得以继续. 因此, 之前的合理猜测, 至少在体外得到了充分的验证.

然而, 并没有这么简单. 生物是极其复杂的, 体外的单分子实验技术自然是高超的, 但是体外观察到的机制就一定是在体内该现象产生的主要原因吗? 而且棘手的是, 体外如此高分辨率的实验由于实验技术所限, 是无法在体内完成的, 那么又如何来验证该机制是否在体内发挥着重要作用呢? 这个时候

就需要结合两状态的随机数学模型. 本文采用了上面所述的两状态基因开关模型 (图 1(c)). 实验上虽然无法在体内完成十分高精度的单分子追踪实验, 但是可以精准地测出每一个单细胞中该信使核糖核酸分子的个数, 即可以测量出平稳分布来. 然而, 与最简单的机制模型一样, 两状态模型的平稳分布一般没有简单的表达式, 而没有具体表达式, 就无法与实验数据进行更为详细的对比和拟合. 不过实验上已经发现, 旋转酶分子和 DNA 分子的结合和解离, 相比于 DNA 处于转录活跃状态的转录速率而言, 是很慢的. 因此可以借助应用数学上常用的渐近分析的办法, 发现平稳分布可以一阶近似为一个在零点单独再加一个峰的 Poisson 分布, 称为零点膨胀的 Poisson 分布.

一旦有了平稳分布的具体表达式, 就可以直接利用实验数据来推断 DNA 转录处于活跃和不活跃状态的平均时间的比值 (体内的 FISH (fluorescent *in situ* hybridization) 实验精度很高, 而且采用了 Bootstrap 方法来计算该比值的误差), 而如果体外观测到的机制是正确的, 则这个比例就会随着旋转酶的不同浓度、旋转酶和 DNA 结合的不同强度而发生明显的变化. 经过分析实验数据发现, 该比值强烈依赖于细胞内的旋转酶浓度和活性, 且其变化的趋势与模型的预测完全相符. 在很多个基因上俱是如此. 综合以上的实验和数学模型分析, 我们间接但是严格地证明, 细菌里的转录爆发机制, 就是来源于旋转酶分子与 DNA 分子的不断随机结合与解离, 从而导致该 DNA 片段的超螺旋情形发生改变 (参见文献 [12]).

这里要特别说明的是, 在两状态模型的转录翻译模型中, 如果只观察蛋白质的生成, 那么只要 DNA 从转录不活跃状态转移到活跃状态的速率足够慢, 就依然可以看到翻译爆发的现象. 此时蛋白质的每次爆发都是来源于 DNA 处于活跃状态时所生成的若干个信使核糖核酸, 其在降解之前生成的蛋白质分子个数的概率分布和几何分布会略有不同. 在之前的翻译爆发实验以及模型中, 由于抑制蛋白的存在, DNA 处于活跃状态的时间非常短, 因此整体而言, 可以等效地认为每隔一个指数分布的等待时间, 就会产生一个信使核糖核酸的分子, 只是此时的参数 k_1 主要是由抑制蛋白的解离和重新结合速率来决定的.

我们从 2010 年初到 2013 年初还和谢晓亮教授实验室以及北京大学生物医学前沿创新中心多个课题组合作, 细致分析了新发现的 DNA 和蛋白质相互作用中的别构合作效应, 并提出了初步的理论分析, 不仅利用统计物理模型解释了该效应的“同时性”, 而且借助于原子层面的分子动力学模拟验证了该效应源于 DNA 的形变. 在这项工作中, 我们利用了 DNA 和两个转录因子 (transcription factor, TF) 相结合的四状态化学反应模型, 结合随机过程和统计物理的知识, 证明了平衡态下两个转录因子和 DNA 结合强度之间的相互影响只可能是互相促进或者互相抑制, 绝不可能出现一个促进另一个、而另一个反过来抑制这一个的现象. 这也与实验观测完全一致 (参见文献 [116]).

2022 年, 我们和北京大学张磊课题组合作, 进一步利用这样的四状态模型, 结合单细胞内类免疫缺陷病毒的随机动力学, 给出了 2014 年发现的噪声增强药物协同作用背后的一般机制. 噪声增强药物协同作用, 即具有噪声增强作用的小分子药物普遍都可以提高激活药物对于类免疫缺陷病毒的重激活效率. Guo 等^[51] 利用随机模型发现, 热力学平衡态无法产生该协同作用, 只有合适的能量输入才可以获得该协同作用, 而且这是一般机理, 并不需要假设两种药物分子层面的细节. 因此, 此项工作揭示了一种新的能量耗散与生物功能之间的联系.

4 正反馈机制下的单细胞地貌函数及表型迁移理论

前一节有关单细胞中心法则的随机模型, 都没有考虑调控和反馈, 而基因调控才是生命体实现各

种功能的根本. 具有调控机制的随机数学模型在过去这 20 年也得到了长足的发展. Jia 等^[59,60] 最近有一系列工作研究了带有反馈的化学主方程模型的性质, 主要基于平稳分布等的精确表达式. Elowitz 等^[23]、Dunlop 等^[18]、Eldar 和 Elowitz^[22] 作为单细胞随机性研究方面的先驱, 也多次结合精妙的单细胞实验和随机数学模型, 揭示随机性在生物功能上的作用.

我们课题组在这方面的研究工作主要集中在对于具有较强正反馈的生物化学反应系统, 建立了一套比较完整的多表型地貌函数理论以及单细胞表型迁移普适公式 (图 2(a)). 对于很多生物化学反应系统, 其分布经常会出现多个峰的情形, 每个峰对应于一种细胞的表型^[11] (图 2(b)). 多个表型的共存对于细胞应对不可预知的环境变化是非常重要的, 因此人们希望研究表型的稳定性及表型间的跃迁速率. 多表型 (稳态) 的产生机制大体上有两大类. 一类是类似前面提到的转录爆发机制, 即是由于 DNA 不同状态间的转移特别慢而导致的, 而这一类表型之间的跃迁速率也很简单, 由 DNA 不同状态间的转移速率所决定, 其本质不是由正反馈机制所决定的. 另一类产生多表型的机制则更为常见, 是由比较强的正反馈所形成的, 这样形成的表型一般都很稳定. 这一类表型之间的跃迁, 可以在随机性相对比较小时得到比较简单的表达式, 这与在 20 世纪后半叶才建立起来的随机过程中著名的大偏差理论, 特别是 Freidlin-Wentzell 型大偏差理论^[29], 有着紧密的联系, 其核心思想是如何刻画单条轨道的“概率密度”以及如何构造类似于能量函数的拟势函数.

过去十多年中单细胞实验技术的飞速发展, 使得人们发现, 至少在细菌中, 基因在活跃程度不同的状态间的切换既不像前人的某些模型中假设的那样, 要比转录和翻译的速率还要快得多; 也不像前人的另一些模型中假设的那样, 比细胞分裂的周期还要慢得多. 这种中间情形下的表型稳定性及表型间的跃迁速率将更难以研究, 其数学难度更大. 利用大偏差理论, Ge 等^[45] 研究了这类最符合实际情形的中间情形中单细胞的动力学行为. 首先将完整的化学主方程模型通过渐近展开的方法简化为随机切换的常微分方程组模型, 称作速率涨落模型, 其中仅保留基因在活跃程度不同的状态间切换的随机性. 该模型在数学上也称作分段确定的 Markov 过程 (piecewise-deterministic Markov process) 或者随机混合模型 (stochastic hybrid model). 此类随机模型曾被化学家用来刻画单个酶分子的动力学, 但是在之前还从未被用来刻画单细胞中的生化随机动力学过程. 然后在该简化模型下结合 Donsker-

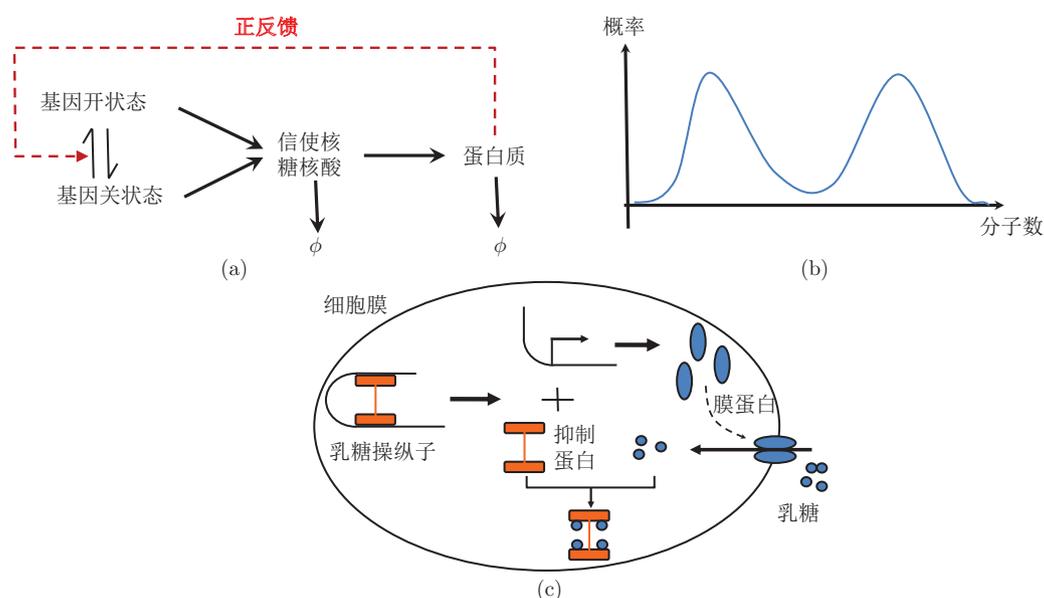


图 2 (网络版彩图) (a) 具有正反馈的单细胞转录翻译模型; (b) 分子数的双峰分布; (c) 乳糖操纵子

Varadhan 和 Freidlin-Wentzell 的大偏差理论, 我们推导出了这种情形下随机轨道的大偏差速率函数, 称为非平衡态的地貌 (landscape) 函数, 类似于平衡态统计物理中的能量函数. 该地貌函数不仅可以作为不同表型之间相对稳定性的定量刻画, 而且还预言了在随机性很小的情形下, 会出现非平衡态相变现象^[39,41,42]. 更进一步地, 该大偏差函数还给出了不同细胞表型间跃迁速率的新公式, 该公式类似于前面提到的单分子化学反应速率中著名的 Kramers 理论, 而且比以往的类似工作更为符合活细胞的实际且更加一般 (参见文献 [45]).

但是, 2015 年的这项工作中只是研究了基因具有两个不同状态的情形, 而真实的细胞中基因状态可能是多个, 这就需要对该理论进行推广, 而且多个基因状态的模型中的地貌函数是无法得到像两个基因状态模型中那样的解析表达式的 (参见文献 [107]). 2020 年, 我们和张云新教授课题组合作, 将涨落速率模型的非平衡态地貌函数理论, 推广到了多个基因状态的情形, 这种情形下的数学证明要复杂得多. 我们发现, 平稳分布的大偏差函数 (即地貌函数) 的导数恰好是某种推广的主特征值问题的唯一非平凡解. 根据这个性质, 证明了地貌函数具有之前在两个基因状态模型中发现的所有性质, 也能刻画不同的稳态间的相对稳定性以及表型迁移的速率; 还据此提出了地貌函数的数值解法, 且在一个三状态的例子中成功应用^[73]. 同时, 也尝试将这套理论真正应用于一个具体的生物系统, 发现并解决实际生物学机制问题. Ge 等^[47] 将速率涨落模型及其地貌函数理论应用于经典的生物学乳糖操纵子模型中 (图 2(c)), 发现虽然正反馈是单细胞呈现多个表型的主要原因, 但是基因状态间的随机切换会使得多个表型的状态更加稳定, 即使得出现多个表型共存现象的环境变量范围大幅度扩大, 更加有利于细胞对抗环境的变化. 更具体一点, 蛋白质表达量高的表型是被正反馈机制稳定住的, 而表达量低的表型则会被相对较慢的基因间切换速率所稳定住. 这两种稳定机制中所呈现的表型间迁移速率, 随着基因状态切换速率的加快, 其变化规律截然相反, 而这一点也正好可以被文献 [45] 提出的速率理论所解释 (参见文献 [47]).

这样一个地貌函数的概念, 其实早在 2004 年就由上海大学的敖平教授等在扩散过程模型里提出并应用^[9,126], 后来美国 Stony Brook 大学的汪劲教授课题组也在化学主方程模型和扩散过程模型下做过非常多这方面的工作^[26,72,115,117]. 2008 年, 纽约大学的 Vanden-Eijnden 和 Heymann^[110] 基于化学主方程模型提出了在体积趋于无穷时如何求解地貌函数的数值算法, 这就是著名的几何最小作用量方法 (geometric minimum action method, gMAM). 北京大学李铁军也在不少重要的合成生物学问题中用过这个概念并提出了具体的数值算法 (参见文献 [76,123,124]). 特别值得一提的是北京师范大学的 Hu^[56] 在 20 世纪 80 年代初就推导出了化学主方程的地貌函数所满足的方程, 是这个方向的前驱之一. 可惜的是, 速率涨落模型无法直接应用几何最小作用量方法来数值求解地貌函数, 因为其地貌函数的表达式含有好几重的优化, 这严重阻碍了这种模型及其地貌函数在具体问题中的应用, 本文作者正在积极地想办法解决.

地貌函数的概念也对与单细胞有关的其他生物学问题和方法有一定的启示. 例如, Zhang 等^[122] 在 2012 年的一项工作中, 就利用地貌函数的思想结合空间多细胞的动力学模型, 成功解释了斑马鱼发育过程中的边界锐化现象; Zhou 等^[125] 利用地貌函数和表型迁移的概念和方法, 发展出了一套在单细胞组学数据中刻画过渡态细胞的统计学方法.

5 学科交叉中涌现的数学问题

在随机数学和生物物理化学交叉的过程中, 涌现出了一些有意思的数学问题, 在前面所介绍的随

机热力学理论中, 已经涉及了一些, 下面介绍本文作者有着深切体会的一个例子, 抛砖引玉. 这是一个从生物物理单分子实验中观测到的有趣现象, 一步步发展成为比较系统的随机数学理论的工作.

首先是一篇 2005 年发表在 *Nature* 杂志上的实验论文, 测量的是单个分子马达蛋白沿着微管的随机运动, 发现该单分子往前和往后走一步都是 4 纳米的距离, 而且无论是往前还是往后, 其等待时间的分布居然完全一样 (参见文献 [6]). 之后有一些物理文献建立了离散时间的 Markov 链模型对其进行理论解释, 不过大多只是证明了向前向后的等待时间均值是相等的. 由于本文作者很熟悉哈佛大学谢晓亮实验室提出的非平衡定态单分子米氏酶动力学模型^[74,116] (图 3(a)), 是离散环上的连续时间紧邻随机徘徊, 同时本文作者也学习过单个分子马达蛋白的随机模型, 那也恰好是环上的紧邻随机徘徊, 往前或者往后在化学层面上就是完成了一个顺时针的环或者逆时针的环. 因此希望可以用环上的连续时间紧邻随机徘徊证明这篇实验论文中观测到的有趣现象, 而且似乎这也是原来随机过程理论中所不完全知道的. 根据后来我们能查到的资料, 概率学家 Samuels 和 Dubins 早在 20 世纪 70 年代和 90 年代, 在更特殊的紧邻随机徘徊模型中给出过同样的数学结论^[17,101], 只是 2008 年本文作者做这个工作时完全不知道这些工作. 本文作者是从轨道层面严格证明了正向环和逆向环的环形成时间的条件分布完全相同, 这在化学上也被称为广义 Haldane 等式^[31]. 后来, 本文作者又将这种环上的对称性推广到了具有某种特殊结构的多环结构^[33]. 2016 年, Jia 等^[58] 将同样的结论推广到了一般的连续时间 Markov 跳过程. 最近, 在北京大学吴秉杰的博士论文^[114] 中, 又将其推广到了较为一般的半 Markov 过程, 并且将 2015 年工作中的一些技术性条件加以削弱. 这一类等式是传统概率论和随机过程理论中所没有的, 但是其中实际上蕴含着该随机过程的一种深刻数学结构: 该过程各个环形成的动力学过程是一个 Markov 更新过程.

对于连续时间连续状态的随机过程, 概率学家 Pitman 和 Yor^[92] 在 1996 年的一篇论文中也得到了环上带有漂移的 Brown 运动满足的相同等式. 但其实该等式是一般成立的. 基于本文作者之前证明环上紧邻随机徘徊满足的环对称等式所使用的技巧, Ge 等^[36] 又证明了环上的一般扩散过程也满足这一等式. 其证明方法与文献 [31] 相同, 是定义了轨道空间的一种新的映射, 称为“拟倒逆映射” (图 3(b)), 其恰好是先完成正向环与先完成逆向环的轨道集合之间的一一映射, 且像映射和原像映射的“概率”之比为常数. 基于这种拟倒逆映射方法, 并利用 3 维 Bessel 过程的 Williams 轨道分解定理, 文献 [36] 先证明了 Brown 运动在拟倒逆映射下还是 Brown 运动, 由此可以推导出环对称等式. 借助该等式, 文献 [36,46] 还证明了沿状态空间中各个环的样本环流量和随机轨道能量耗散满足大偏差原

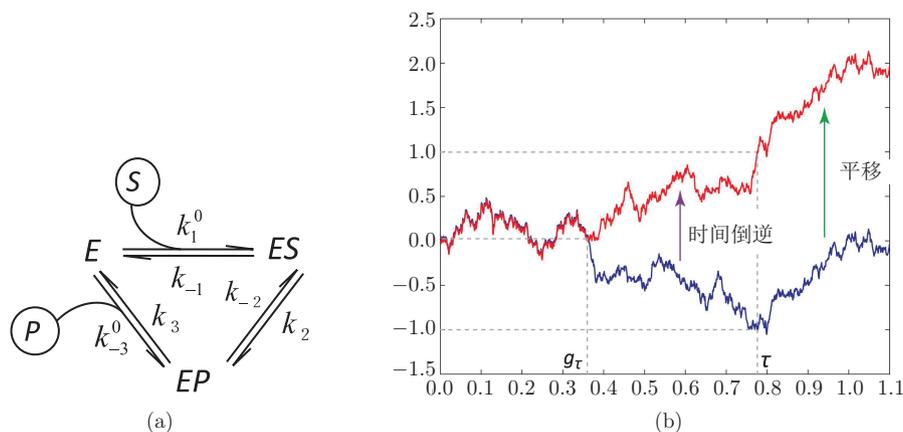


图 3 (网络版彩图) (a) 单分子酶动力学最简单的三状态单环模型; (b) 一维扩散过程的拟倒逆映射

理及诸多暂态和定态涨落定理, 其中 Gallavotti-Cohen 型涨落定理说的就是大偏差速率函数满足一种对称性. 这些结果是对时齐 Markov 链与扩散过程的环流理论的进一步深入发展. 对于高维的, 如二维环面上的扩散过程, 该等式不成立, 但是也许可以构造一种新的映射来得到另一种意义上的环对称性, 然而暂时还没有什么头绪.

在学科交叉的过程中, 不仅涌现出新的数学问题, 而且会涌现出新的需要数学来解释的算法理论问题. 例如, 由 Yang 等^[118]提出的温度积分增强抽样方法, 被成功地应用于多个具体生物化学问题, 取得了非常好的效果, 但是其到底为什么比已知的别的算法要好, 是一般情形下都要好还是只在特定条件下要好, 都是不清楚的. You 等^[120]从随机数学理论的角度详细研究了这个问题, 发现这种增强抽样方法, 其实就是熟知的模拟回火算法的无限切换极限. 更进一步地, 利用耦合扩散过程的 Donsker-Varadhan 大偏差理论, You 等^[120]严格证明了, 模拟回火算法的收敛速度是随着不同温度间切换速度递增的, 于是就给出了温度积分增强抽样方法之所以有效的严格数学基础.

在本文作者进行学科交叉的过程中, 所涌现出的数学理论问题并不太多, 而且大多呈散点状态, 还未连成一片. 希望在不久的将来可以有像混沌现象这样的兼具科学和数学双重意义的深刻发现在随机数学理论和生物物理化学的交叉中涌现出来, 从而开辟新的数学分支, 对随机数学自身的发展起到促进作用.

6 结论和展望

随机过程和生物物理化学交叉, 一方面可以借助随机过程在过去这一百年积累起来的丰富理论知识, 为亚宏观尺度的生物物理化学过程提供合理的抽象数学模型和数学分析方法, 揭示深层次的逻辑关系, 从而加深对实验现象的理解, 从中揭示出新的科学机制和规律, 同时也发展出数学新的应用; 另一方面, 在这些应用的过程中, 也会提出随机过程数学理论中原本没有的问题, 反过来促进随机过程自身的理论发展. 总而言之, 虽然迄今为止生物学对随机数学理论的反哺还不算多, 且大都零散, 但是随机数学理论的确已经在单细胞和单分子的生物物理化学领域中大展拳脚, 发展出了很多的崭新应用, 而且特别值得一提的是, 在这些应用过程中所使用的随机过程理论大都并不是浅显的, 需要研究者具有一定程度这方面的专业训练.

数学和生物物理化学的交叉学科, 往往不容易产生系统性的工作, 更多的是在每一个具体的生物物理化学问题的解决中, 发挥数学的优势. 但是坦白而言在 100 个我们接触到的生物物理化学领域的问题中, 可能只有不到 5 个是数学模型可以发挥作用并解决的科学问题, 剩下的大多数问题其症结都还在原学科内部, 数学工具鞭长莫及. 而就在这剩下的几个数学模型可以发挥作用的科学问题中, 可能有 3 个的背景知识是我们所缺乏的, 且不是一朝一夕可以弥补上的. 由于我们缺乏这方面的基本训练和知识, 从而就没有正确的直观, 因此所建立的数学模型有可能南辕北辙. 这就需要和合作者通力合作, 取长补短, 但即使如此, 可能最终也只有 2 到 3 个可以有所进展. 这就是做交叉学科困难所在, 找到合适的问题, 既在能力范围以内, 又具有真正的科学意义是很不容易的.

以上是本文作者从自身的知识背景和科研体会出发, 提到的一些对于随机数学理论和生物物理化学这一交叉学科的历史和现状的粗浅观点. 该方向方兴未艾, 还有很多没有解决的科学和数学问题亟待解决; 该领域也远未成熟, 它的发展依赖于数学和科学的深度交融, 需要熟悉和欣赏彼此的思维方式和解决问题的方法, 发挥各自的专长, 解决单一学科很难甚至是无法解决的复杂问题.

致谢 我从事交叉学科, 是我的博士导师钱敏先生和华盛顿大学钱纘教授带我入门的, 后来有幸在哈佛大学谢晓亮教

授实验室访问一年, 再后来返回北京大学后同时在北京国际数学研究中心和北京大学生物医学前沿创新中心工作, 使得我和实验学家进一步紧密接触, 也促成了很多的合作. 感谢复旦大学数学科学学院在本文作者刚博士毕业的时候对我从事交叉学科的支持, 感谢北京大学各位老师、同事多年来的支持. 本文作者曾分别申请到数学学部、化学学部和交叉学部的自然科学基金, 感谢国家自然科学基金委员会一直以来的大力支持.

我自己在随机数学和生物物理化学的交叉领域已经从事科研接近二十年, 前面所提到的理论和方法都被总结在两本书中, 都是我和华盛顿大学钱铨教授合作撰写的. 一本是中文教材《数学动力学模型: 在生物物理和生物化学中的应用》以及与之相对应的华文慕课网站上常年开放的在线课程 (<http://www.chinesemooc.org/mooc/4986>); 另一本是 2021 年 11 月出版的英文专著 *Stochastic Chemical Reaction Systems in Biology*, 里面涵盖了从基础知识到最新科研成果的一系列相关内容^[94]. 对于非数学专业的学生, van Kampen^[109] 的经典教材是随机过程很好的入门读物.

当然, 现在随机数学和生物物理化学的交叉, 还有一条新的途径, 那就是通过人工智能、神经网络、统计学习和深度学习这些工具. 北京大学鄂维南教授领导的“AI for Science”就是这样一个途径, 其中的优化算法也涉及很多随机数学理论, 如随机梯度下降法的理论基础等. 同时, 高维概率论也正在成为这些新兴领域中数学理论的语言和基石, 在各种应用中扮演着越来越重要的角色 (参见文献 [111]).

最后, 我想说, 随机数学理论和生物物理化学的交叉学科, 大有可为. 记得 2012 年我刚回到北京大学工作不久, 鄂维南教授有一次就跟我们说, 做应用数学, 要么创建非常深刻的理论, 要么非常接地气地解决实际问题, 尽量不要只是去做中间状态的科研. 这十年来, 我一直以此作为自己的追求. 当然我肯定距离此目标还很远, 但是这个目标就像一盏明灯, 一直在那里指引着我做交叉学科的努力方向. 我也将这句话和所有同道共勉.

参考文献

- 1 Anderson D F, Kurtz T G. *Stochastic Analysis of Biochemical Systems*. New York: Springer, 2015
- 2 Berg O G. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol*, 1978, 71: 587–603
- 3 Bridgman P W. Note on the principle of detailed balancing. *Phys Rev*, 1928, 31: 101–102
- 4 Bridgman P W. The second law of thermodynamics and irreversible processes. *Phys Rev*, 1940, 58: 845
- 5 Cai L, Friedman N, Xie X S. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 2006, 440: 358–362
- 6 Carter N J, Cross R A. Mechanics of the kinesin step. *Nature*, 2005, 435: 308–312
- 7 Celani A, Bo S, Eichhorn R, et al. Anomalous thermodynamics at the microscale. *Phys Rev Lett*, 2012, 109: 260603
- 8 Chen Y, Ge H, Xiong J, et al. The large deviation principle and steady-state fluctuation theorem for the entropy production rate of a stochastic process in magnetic fields. *J Math Phys*, 2016, 57: 073302
- 9 Chen Y C, Yuan R S, Ao P, et al. Towards stable kinetics of large metabolic networks: Nonequilibrium potential function approach. *Phys Rev E*, 2016, 93: 062409
- 10 Chetrite R, Gawędzki K. Fluctuation relations for diffusion processes. *Commun Math Phys*, 2008, 282: 469–518
- 11 Choi P J, Cai L, Frieda K, et al. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 2008, 322: 442–446
- 12 Chong S, Chen C, Ge H, et al. Mechanism of transcriptional bursting in bacteria. *Cell*, 2014, 158: 314–326
- 13 Crooks G E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E*, 1999, 60: 2721–2726
- 14 Cuomo A S E, Nathan A, Raychaudhuri S, et al. Single-cell genomics meets human genetics. *Nat Rev Genet*, 2023, 24: 535–549
- 15 Delbrück M. Statistical fluctuations in autocatalytic reactions. *J Chem Phys*, 1940, 8: 120–124
- 16 Doob J L. Markoff chains—denumerable case. *Trans Amer Math Soc*, 1945, 58: 455–473
- 17 Dubins L E. The gambler’s ruin problem for periodic walks. In: *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell*. Hayward: Institute of Mathematical Statistics, 1996, 7–12
- 18 Dunlop M J, Cox III R S, Levine J H, et al. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat Genet*, 2008, 40: 1493–1498
- 19 Eckart C. The thermodynamics of irreversible processes. I. The simple fluid. *Phys Rev*, 1940, 58: 267–269
- 20 Eckart C. The thermodynamics of irreversible processes. II. Fluid mixtures. *Phys Rev*, 1940, 58: 269–275
- 21 Einstein A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen (in German). *Ann Phys (Leipzig)*, 1905, 17: 549–560
- 22 Eldar A, Elowitz M B. Functional roles for noise in genetic circuits. *Nature*, 2010, 467: 167–173
- 23 Elowitz M B, Levine A J, Siggia E D, et al. Stochastic gene expression in a single cell. *Science*, 2002, 297: 1183–1186
- 24 Esposito M, Harbola U, Mukamel S. Entropy fluctuation theorems in driven open systems: Application to electron counting statistics. *Phys Rev E*, 2007, 76: 031132

- 25 Esposito M, van den Broeck C. Three detailed fluctuation theorems. *Phys Rev Lett*, 2010, 104: 090601
- 26 Fang X, Kruse K, Lu T, et al. Nonequilibrium physics in biology. *Rev Modern Phys*, 2019, 91: 045004
- 27 Flory P. *Principles of Polymer Chemistry*. Ithaca: Cornell University Press, 1953
- 28 Flory P. *Statistical Mechanics of Chain Molecules*. New York: Wiley-Interscience, 1969
- 29 Freidlin M I, Wentzell A D. *Random Perturbations of Dynamical Systems*. New York: Springer-Verlag, 1998
- 30 Gayraud V, Bovier A, Eckhoff M, et al. Metastability in reversible diffusion processes I: Sharp asymptotics for capacities and exit times. *J Eur Math Soc (JEMS)*, 2004, 6: 399–424
- 31 Ge H. Waiting cycle times and generalized Haldane equality in the steady-state cycle kinetics of single enzymes. *J Phys Chem B*, 2008, 112: 61–70
- 32 Ge H. Extended forms of the second law for general time-dependent stochastic processes. *Phys Rev E*, 2009, 80: 021137
- 33 Ge H. Multivariable fluctuation theorems in the steady-state cycle kinetics of single enzyme with competing substrates. *J Phys A-Math Theor*, 2012, 45: 215002
- 34 Ge H. Time reversibility and nonequilibrium thermodynamics of second-order stochastic processes. *Phys Rev E*, 2014, 89: 022127
- 35 Ge H. Mathematical laws of randomness: Expanding the biology toolkit. *Cell*, 2015, 163: 1306
- 36 Ge H, Jia C, Jiang D Q. Cycle symmetry, limit theorems, and fluctuation theorems for diffusion processes on the circle. *Stochastic Processes their Appl*, 2017, 127: 1897–1925
- 37 Ge H, Jia C, Jin X. Martingale structure for general thermodynamic functionals of diffusion processes under second-order averaging. *J Stat Phys*, 2021, 184: 17
- 38 Ge H, Jiang D Q. Generalized Jarzynski's equality of inhomogeneous multidimensional diffusion processes. *J Stat Phys*, 2008, 131: 675–689
- 39 Ge H, Qian H. Thermodynamic limit of a nonequilibrium steady state: Maxwell-type construction for a bistable biochemical system. *Phys Rev Lett*, 2009, 103: 148103
- 40 Ge H, Qian H. Physical origins of entropy production, free energy dissipation, and their mathematical representations. *Phys Rev E*, 2010, 81: 051133
- 41 Ge H, Qian H. Non-equilibrium phase transition in mesoscopic biochemical systems: From stochastic to nonlinear dynamics and beyond. *J R Soc Interface*, 2011, 8: 107–116
- 42 Ge H, Qian H. Landscapes of non-gradient dynamics without detailed balance: Stable limit cycles and multiple attractors. *Chaos*, 2012, 22: 023140
- 43 Ge H, Qian H. Mesoscopic kinetic basis of macroscopic chemical thermodynamics: A mathematical theory. *Phys Rev E*, 2016, 94: 052150
- 44 Ge H, Qian H. Mathematical formalism of nonequilibrium thermodynamics for nonlinear chemical reaction systems with general rate law. *J Stat Phys*, 2017, 166: 190–209
- 45 Ge H, Qian H, Xie X S. Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. *Phys Rev Lett*, 2015, 114: 078101
- 46 Ge H, Qian M, Qian H. Stochastic theory of nonequilibrium steady states. Part II: Applications in chemical biophysics. *Phys Rep*, 2012, 510: 87–118
- 47 Ge H, Wu P, Qian H, et al. Relatively slow stochastic gene-state switching in the presence of positive feedback significantly broadens the region of bimodality through stabilizing the uninduced phenotypic state. *PLoS Comput Biol*, 2018, 14: e1006051
- 48 Gillespie D T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Phys Chem*, 1976, 22: 403–434
- 49 Gingrich T R, Horowitz J M, Perunov N, et al. Dissipation bounds all steady-state current fluctuations. *Phys Rev Lett*, 2016, 116: 120601
- 50 Golding I, Paulsson J, Zawilski S M, et al. Real-time kinetics of gene activity in individual bacteria. *Cell*, 2005, 123: 1025–1036
- 51 Guo X L, Tang T, Duan M X, et al. The nonequilibrium mechanism of noise-enhanced drug synergy in HIV latency reactivation. *iScience*, 2022, 25: 104358
- 52 Hamon A, Ycart B. Statistics for the Luria-Delbrück distribution. *Electron J Statist*, 2012, 6: 1251–1272
- 53 Hatano T, Sasa S. Steady-state thermodynamics of Langevin systems. *Phys Rev Lett*, 2001, 86: 3463–3466
- 54 Hill T L. *Free Energy Transduction in Biology*. New York: Academic Press, 1977
- 55 Horowitz J M, Gingrich T R. Thermodynamic uncertainty relations constrain non-equilibrium fluctuations. *Nature Phys*, 2020, 16: 15–20
- 56 Hu G. Lyapunov function and stationary probability distributions. *Z Physik B-Condensed Matter*, 1986, 65: 103–106
- 57 Jarzynski C. Nonequilibrium equality for free energy differences. *Phys Rev Lett*, 1997, 78: 2690–2693

- 58 Jia C, Jiang D-Q, Qian M-P. Cycle symmetries and circulation fluctuations for discrete-time and continuous-time Markov chains. *Ann Appl Probab*, 2016, 26: 2454–2493
- 59 Jia C, Qian H, Chen M, et al. Relaxation rates of gene expression kinetics reveal the feedback signs of autoregulatory gene networks. *J Chem Phys*, 2018, 148: 095102
- 60 Jia C, Wang L Y, Yin G G, et al. Single-cell stochastic gene expression kinetics with coupled positive-plus-negative feedback. *Phys Rev E*, 2019, 100: 052406
- 61 Jia C, Zhang M Q, Qian H. Emergent Lévy behavior in single-cell stochastic gene expression. *Phys Rev E*, 2017, 96: 040402
- 62 Jiang D Q, Qian M, Qian M P. Mathematical theory of nonequilibrium steady states—On the frontier of probability and dynamical systems. *Lecture Notes in Mathematics*, vol. 1833. Berlin: Springer-Verlag, 2004
- 63 Jiang D Q, Qian M, Zhang F X. Entropy production fluctuations of finite Markov chains. *J Math Phys*, 2003, 44: 4176–4188
- 64 Kessler D A, Levine H. Large population solution of the stochastic Luria-Delbrück evolution model. *Proc Natl Acad Sci USA*, 2013, 110: 11682–11687
- 65 Kou S C, Xie X S. Generalized Langevin equation with fractional gaussian noise: Subdiffusion within a single protein molecule. *Phys Rev Lett*, 2004, 93: 180603
- 66 Kramers H A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 1940, 7: 284–304
- 67 Kurtz T G. The relationship between stochastic and deterministic models for chemical reactions. *J Chem Phys*, 1972, 57: 2976–2978
- 68 Kurtz T G. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes their Appl*, 1978, 6: 223–240
- 69 Langevin P. On the theory of Brownian motion. *C R Acad Sci (Paris)*, 1908, 146: 530–533
- 70 Lebowitz J L, Spohn H. A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics. *J Statistical Phys*, 1999, 95: 333–365
- 71 Lewis G N. A new principle of equilibrium. *Proc Natl Acad Sci USA*, 1925, 11: 179–183
- 72 Li C H, Wang J. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proc Natl Acad Sci USA*, 2014, 111: 14130–14135
- 73 Li J W, Ge H, Zhang Y X. Fluctuating-rate model with multiple gene states. *J Math Biol*, 2020, 81: 1099–1141
- 74 Lu H P, Xue L, Xie X S. Single-molecule enzymatic dynamics. *Science*, 1998, 282: 1877–1882
- 75 Luria S E, Delbrück M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 1943, 28: 491–511
- 76 Lv C, Li X, Li F, et al. Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process. *PLoS Comput Biol*, 2015, 11: e1004156
- 77 Machlup S, Onsager L. Fluctuations and irreversible process. II. Systems with kinetic energy. *Phys Rev*, 1953, 91: 1512–1515
- 78 Mackey M, Santillán M. Mathematics, biology, and physics: Interactions and interdependence. *Notices Amer Math Soc*, 2005, 52: 832–840
- 79 Magde D, Elson E, Webb W W. Thermodynamic fluctuations in a reacting system—Measurement by fluorescence correlation spectroscopy. *Phys Rev Lett*, 1972, 29: 705–708
- 80 Manzano G, Subero D, Maillet O, et al. Thermodynamics of gambling demons. *Phys Rev Lett*, 2021, 126: 080603
- 81 Marcus R A, Sutin N. Electron transfers in chemistry and biology. *Biochim Biophys Acta (BBA)-Rev Bioenergetics*, 1985, 811: 265–322
- 82 McQuarrie D A. Stochastic approach to chemical kinetics. *J Appl Probability*, 1967, 4: 413–478
- 83 Mou C Y, Luo J L, Nicolis G. Stochastic thermodynamics of nonequilibrium steady states in chemical reaction systems. *J Chem Phys*, 1986, 84: 7011–7017
- 84 Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. *Science*, 2012, 336: 183–187
- 85 Neri I, Roldán É, Jülicher F. Statistics of infima and stopping times of entropy production and applications to active molecular processes. *Phys Rev X*, 2017, 7: 011019
- 86 Nicolis G, Prigogine I. *Self-organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. New York: Wiley, 1977
- 87 Onsager L. Reciprocal relations in irreversible processes. II. *Phys Rev*, 1931, 38: 2265–2279
- 88 Onsager L, Machlup S. Fluctuations and irreversible processes. *Phys Rev*, 1953, 91: 1505–1512
- 89 Paulsson J. Summing up the noise in gene networks. *Nature*, 2004, 427: 415–418
- 90 Paulsson J, Ehrenberg M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Phys Rev Lett*, 2000, 84: 5447–5450

- 91 Peccoud J, Ycart B. Markovian modeling of gene-product synthesis. *Theor Population Biol*, 1995, 48: 222–234
- 92 Pitman J, Yor M. Bessel processes and infinitely divisible laws. In: *Stochastic Integrals*. Berlin-Heidelberg: Springer, 1981, 285–370
- 93 Prigogine I. *Introduction to Thermodynamics of Irreversible Processes*, 2nd ed. New York: Interscience, 1961
- 94 Qian H, Ge H. *Stochastic Chemical Reaction Systems in Biology*. Lecture Notes on Mathematical Modeling in the Life Sciences. Cham: Springer, 2021
- 95 Qian M, Hou Z T, et al. *Reversible Markov Process (in Chinese)*. Changsha: Hunan Science and Technology Press, 1979 [钱敏, 侯振挺, 等. 可逆马尔可夫过程. 长沙: 湖南科学技术出版社, 1979]
- 96 Rao R, Esposito M. Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics. *Phys Rev X*, 2016, 6: 041064
- 97 Reeds M. Why is mathematical biology so hard? *Notices Amer Math Soc*, 2004, 51: 338–342
- 98 Reeds M. Mathematical biology is good for mathematics. *Notices Amer Math Soc*, 2015, 62: 1172–1176
- 99 Roldán É, Martínez I A, Parrondo J M R, et al. Universal features in the energetics of symmetry breaking. *Nature Phys*, 2014, 10: 457–461
- 100 Sagawa T, Ueda M. Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Phys Rev Lett*, 2012, 109: 180602
- 101 Samuels S M. The classical ruin problem with equal initial fortunes. *Math Mag*, 1975, 48: 286–288
- 102 Schnakenberg J. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev Mod Phys*, 1976, 48: 571–585
- 103 Schrödinger E. *What is Life*. Cambridge: Cambridge University Press, 2012
- 104 Seifert U. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Phys Rev Lett*, 2005, 95: 040602
- 105 Seifert U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep Prog Phys*, 2012, 75: 126001
- 106 Sekimoto K. *Stochastic Energetics*. Berlin: Springer-Verlag, 2010
- 107 Shahrezaei V, Swain P S. Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci USA*, 2008, 105: 17256–17261
- 108 Taniguchi Y, Choi P J, Li G W, et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 2010, 329: 533–538
- 109 van Kampen N G. *Stochastic Processes in Physics and Chemistry*, 3rd ed. Amsterdam: North Holland, 2007
- 110 Vanden-Eijnden E, Heymann M. The geometric minimum action method for computing minimum energy paths. *J Chem Phys*, 2008, 128: 061103
- 111 Vershynin R. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge: Cambridge University Press, 2018
- 112 Wang F Y, Xiong J, Xu L H. Asymptotics of sample entropy production rate for stochastic differential equations. *J Stat Phys*, 2016, 163: 1211–1234
- 113 Wang W, Li G W, Chen C, et al. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*, 2011, 333: 1445–1449
- 114 Wu B J. Long-time behavior and cyclic kinetics of semi-Markov processes. PhD Thesis. Beijing: Peking University, 2023 [吴秉杰. 半马尔可夫过程的长时间行为与环动力学. 博士学位论文. 北京: 北京大学, 2023]
- 115 Wu W, Wang J. Landscape-flux framework for nonequilibrium dynamics and thermodynamics of open Hamiltonian systems coupled to multiple heat baths. *J Phys Chem B*, 2021, 125: 7809–7827
- 116 Xie X S. Enzyme kinetics, past and present. *Science*, 2013, 342: 1457–1459
- 117 Xu L, Patterson D, Staver A C, et al. Unifying deterministic and stochastic ecological dynamics via a landscape-flux approach. *Proc Natl Acad Sci USA*, 2021, 118: e2103779118
- 118 Yang L, Liu C W, Shao Q, et al. From thermodynamics to kinetics: Enhanced sampling of rare events. *Acc Chem Res*, 2015, 48: 947–955
- 119 Yoshimura K, Ito S. Thermodynamic uncertainty relation and thermodynamic speed limit in deterministic chemical reaction networks. *Phys Rev Lett*, 2021, 127: 160601
- 120 You Z Y, Li L Y, Lu J F, et al. Integrated tempering enhanced sampling method as the infinite switching limit of simulated tempering. *J Chem Phys*, 2018, 149: 084114
- 121 Yu J, Xiao J, Ren X, et al. Probing gene expression in live cells, one protein molecule at a time. *Science*, 2006, 311: 1600–1603
- 122 Zhang L, Radtke K, Zheng L K, et al. Noise drives sharpening of gene expression boundaries in the zebrafish hindbrain. *Mol Syst Biol*, 2012, 8: 613
- 123 Zhou P, Gao X, Li X, et al. Stochasticity triggers activation of the S-phase checkpoint pathway in budding yeast.

- Phys Rev X, 2021, 11: 011004
- 124 Zhou P, Li T. Construction of the landscape for multi-stable systems: Potential landscape, quasi-potential, A-type integral and beyond. *J Chem Phys*, 2016, 144: 094109
- 125 Zhou P, Wang S, Li T, et al. Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. *Nat Commun*, 2021, 12: 5609
- 126 Zhu X M, Yin L, Hood L, et al. Calculating biological behaviors of epigenetic states in the phage lambda life cycle. *Funct Integr Genomics*, 2004, 4: 188–195

The intersection of stochastic mathematical theory and single-molecule/single-cell biophysical chemistry

Hao Ge

Abstract The intersection of stochastic mathematical theory and biophysical chemistry is a new direction that has emerged in the recent two decades. In this paper, we first briefly review the history of the intersection between stochastic mathematical theory and biophysical chemistry, especially at the single-molecule and single-cell levels, and then present the progress of this discipline over the recent two to three decades in several directions with which the author is familiar, including some of the work done by the author's research group.

Keywords stochastic mathematical theory, biophysical chemistry, single-molecule, single-cell, intersection

MSC(2020) 60G10, 60J28, 92C37

doi: 10.1360/SSM-2023-0038