

## 自我复制的统计物理学

Jeremy L. England

引用本文: *J. Chem. Phys.* 139, 121923 (2013); doi: 10.1063/1.4818538

在线查看: <http://dx.doi.org/10.1063/1.4818538>

查看目录: <http://jcp.aip.org/resource/1/JCPSA6/v139/i12> Published by the AIP Publishing LLC.

---

关于J. Chem的更多信息。理论物理。期刊首页: <http://jcp.aip.org/>

期刊信息: [http://jcp.aip.org/about/about\\_the\\_journal](http://jcp.aip.org/about/about_the_journal) Top 下载: [http://jcp.aip.org/features/most\\_downloaded](http://jcp.aip.org/features/most_downloaded) Information 作者: <http://jcp.aip.org/authors>

## ADVERTISEMENT



Explore the **Most Cited**  
Collection in Applied Physics

AIP  
Publishing

# 自复制的统计物理

Jeremy L. England

美国马萨诸塞州剑桥市马萨诸塞州大道77号6C楼,麻省理工学院物理系02139

(收稿2013年4月28日;2013年8月1日录用;2013年8月21日在线发布)

自我复制是每一种生物共有的能力,简单的物理直觉告诉我们,这种过程必然是由熵的产生推动的。在这里,我们通过推导与热浴耦合的系统中自我复制过程中产生的热量的下界,来保证这种直觉的严格性和定量性。我们发现,物理上允许的产热速率的最小值是由复制子的生长速度、内部熵和耐久性决定的,我们讨论了这一发现对细菌细胞分裂的影响,以及对生物前自我复制核酸的出现的影响。©2013年作者(年代)。除另有说明外,所有文章内容均采用知识共享署名3.0未移植许可协议。[<http://dx.doi.org/10.1063/1.4818538>]

## INTRODUCTION

每一种生物都可以通过与周围环境交换能量和物质来复制自己。所有这些自发的“自我复制”例子的一个共同特征是它们在统计学上的不可逆性:很明显,一个细菌变成两个细菌的可能性要比两个细菌以某种方式自发地恢复为一个细菌的可能性大得多。从物理学的角度来看,这一观察结果包含了一个有趣的暗示,即自我复制体的性质如何必须受到热力学定律的约束,热力学定律规定了不可逆性总是伴随着熵的增加。然而,长期以来,用普遍的术语来谈论生命系统的统计物理学一直被认为是具有挑战性的,因为它们总是远离热力学平衡,因此不需要在微观安排上服从简单的玻尔兹曼概率分布。面对如此不受约束的生物组织多样性,我们完全有理由担心,每一个给定的生物自我复制过程的特殊机械复杂性,可能会压倒我们用一般理论来阐述的能力。

在开始前进的道路时,我们应该考虑一个粒子数 $N$ 和体积 $V$ 固定的系统,它与逆温度 $\beta$ 的热浴相接触。如果我们给这个系统的微态 $i, j$ 等加上标签,并分别与每个这样的微态关联能量 $E_i, E_j$ 等,那么潜在的哈密顿动力学的时间反转对称性告诉我们,以下详细的平衡关系在热平衡状态下保持:<sup>1</sup>

$$\frac{e^{-\beta E_i}}{Z(\beta)} \pi(i \rightarrow j; \tau) = \frac{e^{-\beta E_j}}{Z(\beta)} \pi(j^* \rightarrow i^*; \tau). \quad (1)$$

这里,  $i^*$  是  $i$  的动量反转伙伴,  $Z(\beta)$  是系统的正则配分函数。跃迁矩阵元素  $\pi(i \rightarrow j; \tau)$  是在时间  $t = \tau > 0$  时发现系统处于微态  $j$  的条件概率,假设它在更早的时间  $t = 0$  时开始于微

态  $i$ 。因此,上述关系表明,当系统松弛到热平衡并在微观状态上实现玻尔兹曼分布( $p(i) = e^{-\beta E_i} / Z(\beta)$ )时,连接  $i$  到  $j$  的正反方向电流的概率相等。

进步来自于认识到在从  $i$  到  $j$  的转变过程中排入热浴的热量由  $\Delta Q_{i \rightarrow j} = E_j - E_i$  给出,而且量  $\beta Q_{i \rightarrow j}$  是热浴的熵在这一转变过程中变化的量。因此,我们可以这样写

$$\frac{\pi(j^* \rightarrow i^*; \tau)}{\pi(i \rightarrow j; \tau)} = \exp[-\Delta S_{bath}^{i \rightarrow j}]. \quad (2)$$

方程(2)建立了过渡的不可逆性(即向前方向发生的可能性比相反方向发生的可能性大多少)与在向前轨迹过程中周围环境中增加的熵量之间的微观详细关系。此外,应该强调的是,虽然这一结果是从详细平衡(仅在平衡状态下成立)的陈述中得出的,但它本身对两个微观状态之间的任何过渡都是有效的,因此适用于任意远离平衡状态的非驱动系统的松弛动力学。

这项工作的主要目的是表明,上面说明的不可逆性和熵产生之间的微观详细的定量关系对远非平衡的宏观过程(如生物自我复制)具有重要的一般热力学后果。在过去非平衡统计力学结果的基础上,<sup>2</sup>我们将首先推导出宏观不可逆过程的热力学第二定律的概括。随后,我们将利用这一结果来证明自我复制子在其大小、增长率、内部熵和耐久性方面的热输出的下界。通过对经验数据的分析,我们将进一步证明,这一下限在与真实微生物和其他自我复制体的功能相关的尺度上运行。

## 宏观的不可逆性

我们的目标是确定两个任意复杂粗粒度状态之间宏观转变的热力学约束。要做到这一点，我们必须首先考虑概率、热量和熵在微观层面上是如何相互关联的。方程(2)在一种特殊情况下建立了这样的关系，但结果证明这种联系要普遍得多，因为已经证明<sup>2</sup>对于所有热浴耦合的、时间对称驱动的非平衡系统，其动力学由缺乏任何弹道惯性感的扩散运动主导，任何轨迹从 $x(0)$ 开始，并经过微观状态 $x(t)$ 随时间 $\tau$ 服从

$$\beta \Delta Q[x(t)] = \ln \left[ \frac{\pi[x(t)]}{\pi[x(\tau-t)]} \right]. \quad (3)$$

在这里， $\beta \equiv 1/T$ 以自然单位设置热浴的温度 $T$ ， $\pi[x(t)]$ 是给定 $x(0)$ 的轨迹 $x(t)$ 的概率， $Q$ 是在 $x(t)$ 过程中释放到热浴中的热量。这个公式抓住了热量和不可逆性之间的基本微观关系：一个向外运动的轨迹比它的时间反转的可能性越大，宇宙的熵就越有可能通过将热量耗尽到周围的浴缸中而增加。此外，应该强调的是，这一结果适用于受时间对称外部驱动场影响的系统。这意味着系统的稳态概率分布不必是玻尔兹曼分布，甚至不必是其稳态概率通量满足详细平衡的分布，才能使不可逆性和熵产生之间的这种基本关系成立。

通过固定轨迹的起点和终点( $x(0) = i, x(\tau) = j$ )，我们可以平均从 $i$ 到 $j$ 的所有路径上正向热量的指数权重，并得到 $-\ln \pi(j \rightarrow i; \tau) / \pi(i \rightarrow j; \tau) = \exp[-\beta Q_{\tau i \rightarrow j}]_{i \rightarrow j}$ 。这个微观规则(其中Eq.(2)显然是特殊的、非驱动的情况)必须具有宏观后果，为了研究这些后果，我们必须形式化用宏观术语谈论我们感兴趣的系统意味着什么。为了做到这一点，我们首先假设系统中存在一些粗粒度的、可观察到的条件 $\mathbf{I}$ ，例如标准“系统在一轮细胞分裂开始时精确地包含一个健康的、指数生长阶段的细菌”。如果我们在一组受控条件下准备系统，然后发现满足 $\mathbf{I}$ 的标准，我们可以立即与 $\mathbf{I}$ 关联一个概率分布 $p(\mathbf{I}|\mathbf{I})$ ，这是系统处于某些特定微观状态 $\mathbf{I}$ 的隐式概率，假设它是在受控条件下准备的，然后观察到它处于宏观状态 $\mathbf{I}$ 。

现在假设我们让一个时间间隔 $\tau$ 过去，同时保持我们的系统与温度相反的热浴接触 $\beta$ (在 $\pi(i \rightarrow j)$ 的后续表达式中隐式 $\tau$ 将被省略)。在这一点上，我们可以为系统的粗粒度观察引入第二个准则(例如，“在细胞分裂开始时，系统恰好包含两个健康的指数生长期细菌”)，我们可以称之为 $\mathbf{II}$ ，如果我们观察到这个准则得到满足，我们可以立即将概率分布 $p(j|\mathbf{II})$ 定义为系统

处于微观状态 $j$ 的可能性，假设最初在 $\mathbf{I}$ 中准备的系统随后在一定时间内 $\tau$ 繁殖，然后发现系统处于宏观状态 $\mathbf{II}$ 。

构造了我们感兴趣的宏观状态的概率定义之后，我们现在还可以定义相关的量，使我们能够对不可逆性给出宏观定义。具体地说，我们可以写

$$\pi(\mathbf{I} \rightarrow \mathbf{II}) = \int_{\mathbf{II}} dj \int_{\mathbf{I}} di p(i|\mathbf{I}) \pi(i \rightarrow j), \quad (4)$$

和

$$\pi(\mathbf{II} \rightarrow \mathbf{I}) = \int_{\mathbf{I}} di \int_{\mathbf{II}} dj p(j|\mathbf{II}) \pi(j \rightarrow i). \quad (5)$$

这些概率中的第一个 $\pi(\mathbf{I} \rightarrow \mathbf{II})$ 给出了根据 $\mathbf{I}$ 准备的系统在时间 $\tau$ 后被观察到满足 $\mathbf{II}$ 的可能性。第二个概率 $\pi(\mathbf{II} \rightarrow \mathbf{I})$ 给出了在另一个时间间隔 $\tau$ 之后再次观察到相同系统满足 $\mathbf{I}$ 的可能性，将这两个量放在一起并取它们的比值，从而为我们量化了从 $\mathbf{I}$ 自发传播到 $\mathbf{II}$ 的不可逆性：

$$\begin{aligned} \frac{\pi(\mathbf{II} \rightarrow \mathbf{I})}{\pi(\mathbf{I} \rightarrow \mathbf{II})} &= \frac{\int_{\mathbf{I}} di \int_{\mathbf{II}} dj \left( \frac{p(j|\mathbf{II})}{p(i|\mathbf{I})} \right) p(i|\mathbf{I}) \pi(j \rightarrow i)}{\int_{\mathbf{I}} di \int_{\mathbf{II}} dj p(i|\mathbf{I}) \pi(i \rightarrow j)} \\ &= \frac{\int_{\mathbf{I}} di \int_{\mathbf{II}} dj p(i|\mathbf{I}) \pi(i \rightarrow j) \frac{\langle e^{-\beta \Delta Q_{ij}} \rangle_{i \rightarrow j}}{e^{\ln \left[ \frac{p(j|\mathbf{II})}{p(i|\mathbf{I})} \right]}}}{\int_{\mathbf{I}} di \int_{\mathbf{II}} dj p(i|\mathbf{I}) \pi(i \rightarrow j)} \\ &= \left\langle \frac{\langle e^{-\beta \Delta Q_{ij}} \rangle_{i \rightarrow j}}{e^{\ln \left[ \frac{p(j|\mathbf{II})}{p(i|\mathbf{I})} \right]}} \right\rangle_{\mathbf{I} \rightarrow \mathbf{II}}, \end{aligned} \quad (6)$$

其中 $\langle \dots \rangle_{\mathbf{I} \rightarrow \mathbf{II}}$ 表示从初始集合 $\mathbf{I}$ 中的某个 $i$ 到最终集合 $\mathbf{II}$ 中的某个 $j$ 的所有路径的平均值，每个路径都由其似然加权。以通常的方式( $S \equiv -p_i \ln p_i$ )定义每个集成的香农熵 $S$ ，我们可以构造 $\Delta S_{int} \equiv S_{\mathbf{II}} - S_{\mathbf{I}}$ ，它测量正向反应的内部熵变化。由于 $e^x \geq 1 + x$ 对于所有 $x$ ，我们可以重新排列(6)来写

$$\left\langle e^{-\ln \left[ \frac{\pi(\mathbf{II} \rightarrow \mathbf{I})}{\pi(\mathbf{I} \rightarrow \mathbf{II})} \right] + \ln \left[ \frac{p(j|\mathbf{II})}{p(i|\mathbf{I})} \right]} \langle e^{-\beta \Delta Q_{ij}} \rangle_{i \rightarrow j} \right\rangle_{\mathbf{I} \rightarrow \mathbf{II}} = 1 \quad (7)$$

并立即得到

$$\beta \langle \Delta Q \rangle_{\mathbf{I} \rightarrow \mathbf{II}} + \ln \left[ \frac{\pi(\mathbf{II} \rightarrow \mathbf{I})}{\pi(\mathbf{I} \rightarrow \mathbf{II})} \right] + \Delta S_{int} \geq 0. \quad (8)$$

如果 $\mathbf{I}$ 和 $\mathbf{II}$ 对应于包含所有微观状态 $\mathbf{I}$ 的相同群，则有 $\pi(\mathbf{II} \rightarrow \mathbf{I}) = \pi(\mathbf{I} \rightarrow \mathbf{II}) = 1$ ，上述关系将简化为热力学第二定律的简单表述：平均而言，系统的熵变 $\Delta S_{int}$ 加上浴体的熵变 $\beta \langle \Delta Q \rangle_{\mathbf{I} \rightarrow \mathbf{II}}$ 必须大于或等于零，即宇宙的平均总熵变必须为正。然而，我们在这里使用克鲁克斯的微观关系所展示的是，从任意态系综 $p(i|i)$ 到未来系综 $p(j|\mathbf{II})$ 的转变的宏观不可逆性对相关熵的产生设定了更严格的界限：宏观过程越不可逆(即越负



$\ln \pi(I \rightarrow II)$

), 跃迁的必定是最小的总熵产生。此外, 由于该公式是在非常一般的假设下推导出来的, 因此它不仅适用于自我复制, 还适用于大范围的粗粒度起始和结束状态之间的转换。从这个角度来看, Eq.(8)中的结果与信息论中关于熵产生的过去界集terms,<sup>3-5</sup>以及由擦除一点信息所产生的热量的著名的Landauer界密切相关。<sup>6</sup>

## 自我复制的一般约束

上述第二定律的推广适用于从计算机操作到分子马达动作的各种非平衡、热浴耦合的情况。然而, 这一结果被证明是特别有价值的, 可以作为一个视角, 通过它来重新审视生物自我复制现象。很久以前, 对进化建模的兴趣催生了丰富的文献, 探索自我复制对种群动态和达尔文竞争的影响。在这类研究中, 达尔文“适应度”的概念经常被用于比较非相互作用种群中不同的自我复制体: 根据定义, 为了最快地复制自己而与环境相互作用的复制体更“适合”, 因为连续几轮的指数增长将确保它们在未来种群中占任意大的比例。<sup>7</sup>

这里首先要注意的是, 刚才描述的这种指数增长是一个高度不可逆的过程: 在一个选择性的横扫中, 最适合的复制子在一个种群中占据主导地位, 从定义上讲, 未来几乎看起来与过去非常不同。令人高兴的是, 我们现在能够对这种不可逆性的热力学后果进行定量分析。因此, 让我们假设有一个简单的自我复制体生活在逆温度 $\beta$ 下, 其种群 $n$ 服从如下形式的主方程

$$\dot{p}_n(t) = g n [p_{n-1}(t) - p_n(t)] - \delta n [p_n(t) - p_{n+1}(t)], \quad (9)$$

其中 $p_n(t)$ 表示在时刻 $t$ 有 $n$ 个总体的概率, 且 $g > \delta > 0$ 。为简单起见, 我们假设由速率参数 $\delta$ 介导的衰变事件相当于复制子回到其形成环境中的精确反应物集, 并且我们还假设从这样一组反应物中形成一个新的复制子会使系统的内部熵改变一定量 $\Delta s_{int}$ , 并且平均向周围环境释放一定量的热量 $q$ 。在本讨论中, 我们将进一步忽略考虑进化种群中自发突变的影响所带来的额外复杂性层。<sup>7</sup>

对于 $n(t=0)=1$ , 我们期望系统的行为遵循形式为 $n(t) = n(0)e^{(g-\delta)t}$ 的确定性指数增长路径。无论 $n$ 的确切值是多少, 一个特定的复制体在短时间 $dt$ 内产生 $\pi(I \rightarrow II)$ 的概率应该是 $g dt$ , 而一个新生的副本在相同长度的间隔时间 $(\pi(II \rightarrow I))$ 内衰减回原处的概率应该是 $\delta dt$ 。因此

, 代入式(8), 我们有

$$\Delta s_{tot} \equiv \beta \Delta q + \Delta s_{int} \geq \ln[g/\delta]. \quad (10)$$

作为题外话, 这里应该指出, 我们已经避免在上面包含一个来自系统中粒子数量的虚假乘法因子。例如, 人们可能会认为, 在将一个粒子转化为两个粒子的情况下, 逆转回一个粒子的概率率应该是 $2\delta$ , 由此产生的熵产的界为 $\ln(g/\delta) - \ln 2$ 。要了解为什么不应该包括 $\ln 2$ , 重要的是要认识到经典物理系统中的粒子具有彼此区分的位置。因此, 我们要么可以将所讨论的过程视为约束自我复制和粒子混合的熵产生(在这种情况下, 混合熵项抵消了2的因素), 要么我们可以在所讨论的跃迁中定义我们的粗粒度状态, 以便我们只考虑刚刚诞生的复制子的反转的概率电流。无论哪种方式, 相关的边界都是 $\ln(g/\delta)$ 。

上述结果当然与我们的预期一致, 即为了使这个自我复制因子能够实现净增长, 我们必须要求 $g > \delta$ , 这反过来又与自我复制相关的总熵产设定了一个正的下界。然而, 通过在固定的 $\delta$ 、 $\Delta s_{int}$ 和 $q$ 处重新排列表达式, 可以看到更多的结果

$$g_{\max} - \delta = \delta (\exp[\beta \Delta q + \Delta s_{int}] - 1). \quad (11)$$

换句话说, 一个自我复制体的最大净增长率是由三个因素固定的: 它的内部熵( $s_{int}$ ), 它的耐久性( $1/\delta$ ), 以及在复制过程中散逸到周围浴体的热量 $q$ 。

以下几点评论是有条理的。首先, 让我们考虑比较两种不同的复制子, 它们具有固定的 $\Delta s_{int}$ 和 $\delta$ , 具有不同的热值 $q$ 和 $q$ 。如果 $q > q$ , 那么显然 $g_{\max} > g_{\max}$ ; 耗散更多热量的复制子有相应更快生长的潜力。此外, 根据能量守恒, 我们知道这种热量必须以两种不同的方式之一产生: 要么来自最初储存在构建复制子的反应物中的能量(例如通过糖的水解), 要么来自某些时变的外部驱动场对系统所做的功(例如通过光合作用期间对光的吸收)。换句话说, 从统计物理学中精确考虑得出的基本热力学约束告诉我们, 自我复制子的最大潜在适应度取决于它如何有效地利用其环境中的能量来源来催化其自身的繁殖。因此, 生殖适应性与高效代谢密切相关这一经验性的生物学事实, 现在在物理学上有了明确而简单的基础。

这里需要提出的一个稍微微妙的观点是, 自我复制体不仅可以通过 $q$ 更有效地促进其生长, 还可以通过 $\delta$ 和 $\Delta s_{int}$ 降低其生长成本, 从而提高其最大潜在增长率。在所有条件相同的情况下, 自我复制体的持久性越差或组织性越差, 为了达到一定的生长速度, 它必须获得的最低代谢能量就越少。因此, 在自我复制体之间为主宰未来人口而展开的竞争中, “成功”的一个策略就是“更简单”

建设和更容易自发退化。当然，在两个自我复制体在内部熵或持久性(如病毒和兔子)方面存在巨大差异的极限下，以达尔文适应度进行比较的基础就变得更弱了。然而，在形式和结构相似的竞争者之间的竞争中，值得注意的是，降低复制子生长的最低“代谢”成本的一种策略是用可能更快“磨损”的新成分来替代。

## 自我复制的分布

持久性在复制适应性中的作用的一个简单证明来自多核苷酸的案例，它为研究熵产生在自我复制动力学中的作用提供了丰富的应用领域。<sup>8</sup>最近的一项研究使用体外进化来优化自我复制的RNA分子的生成速度，该分子的形成伴随着单个主链连接反应和单个焦磷酸基团的离开。<sup>9</sup>我们可以合理地假设，与此相反的反应，即带高度负电荷的焦磷酸必须攻击核酸的磷酸主链，其进行速度会比简单的水水解反应慢。通过对RNAase a底物UpA自发降解的早期(20天)指数动力学进行线性拟合，测量了这种水解的速率，测量的半衰期约为4年。<sup>10</sup>因此，随着自我复制因子1小时的倍增时间，我们可以估计该系统 $\ln[g/\delta] \geq \ln[(4\text{年})/(1\text{小时})]$ 。由于在相似的环境浓度下，结扎反应将底物的混合熵交换为焦磷酸盐的混合熵，因此我们也可以假设在这种情况下，反应的内部熵的变化是负可解的。因此，我们可以估计热界为

$$(\Delta Q) \geq RT \ln[(4\text{ years})/(1\text{ h})] = 7\text{ kcal mol}^{-1}. \quad (12)$$

由于实验数据表明，反应的焓在10千卡摩尔<sup>-1</sup>,<sup>11,12</sup>附近，看来这个分子的运作相当接近由它的组装方式所设定的热力学效率的极限。

为了强调这一点，我们可以考虑一下，如果同样的反应是用DNA来实现的，那么这种结合可能是什么，因为在水溶液中，DNA比RNA在抗水解方面更稳定。<sup>13</sup>在这种情况下，我们将有 $Q \geq RT \ln[(3 \times 10^{-7}\text{年})/(1\text{小时})] = 16\text{ kcal mol}^{-1}$ ，这超过了结扎反应的估计焓，因此在热力学上是禁止的。这个计算说明了DNA和RNA之间的显著差异，就每个分子参与由简单的三磷酸构建块推动的自我催化复制反应的能力而言：DNA的耐久性要高得多，因此在熵产方面需要付出更高的每碱基热力学成本<sup>14</sup>，以便在所有条件相等的比较中使其生长速度与RNA相匹配。

这里的关键点是，如果一个自我复制的核酸以速率常数 $g$ 催化其自身的生长，并以速率常数 $\delta$ 降解，那么该分子应该能

够通过与 $1/(g-\delta)$ 成比例的倍增时间的自我复制表现出指数增长。然而，热力学只对比值 $g/\delta$ 设定了一个界限，这意味着在复制的“代谢成本”保持固定的情况下， $g-\delta$ 可以任意变大。因此，令人惊讶的是，在适当的情况下，RNA的更大脆弱性可以被视为一种适应性优势，即使所讨论的系统是外部驱动的，即使复制子在其内部自由度上保持非玻尔兹曼分布，这也适用。此外，我们预计DNA和RNA之间的热结合差应该在反应期间连接的碱基数量中大致呈线性增加，这迫使DNA复制子的最大可能增长率在所有条件相等的比较中与其RNA竞争对手相比呈指数级缩小。这一观察结果无疑是令人感兴趣的，因为过去的论点是基于其他理由，即RNA，而不是DNA，一定是生物出现前自我复制核酸的材料。<sup>12,15</sup>

这里考虑的单个核酸碱基连接的具体例子是有指导意义的，因为在这种情况下，不可逆性和熵产生之间的关系进一步简化为一种可识别的形式。我们可以合理地定义我们的粗粒化状态，使该反应的起始和结束产物都处于局部热平衡状态，从而在给定其化学键约束的情况下，它们的所有自由度都是玻尔兹曼分布的。仅在这种情况下，细致平衡要求比正向和反向利率被锁定的吉布斯自由能变化的反应通过famil-iar公式 $\ln[k_f/k_r] = -\beta G = -\beta H + \beta \Delta S_{\text{int}} = \beta Q + \beta \Delta S_{\text{int}}$ 。<sup>8</sup>在这种情况下，耐用性和经济增长率之间的关系有一个优雅的描述的过渡态理论：一个激活障碍较低的正向会降低在相反的方向。然而，这里的关键点是，尽管自由能和反应速率之间的关系仅在低卡平衡假设下才能得到，但我们在这里推导出的不可逆性方面的熵产边界不等式即使适用于系统中许多自由度开始和结束热平衡之外的复制过程的情况。

另外，还可以指出，对上述核酸生长的叙述可能提出naïve反对意见，理由是它关于DNA最大可能生长速度的结论似乎与经验事实不符，即DNA显然能够以比每小时一个磷酸二酯连锁更快的速度进行复制；过程全酶DNA聚合酶III以其每秒约1000个碱基对的惊人催化速度而闻名。<sup>16</sup>这个谜题的解决在于认识到在这里提出的蛋白质辅助DNA复制场景中，聚合酶组装首先以atp依赖的方式加载到DNA上，使用由各种蛋白质亚基组成的甜甜圈形状的夹子将酶不可逆地附着在链上。因此，在聚合酶可以极快地催化DNA链的伸长的同时，也必须考虑到DNA水解的逆反应要发生很多



与溶液中分离的DNA相比,更容易与拴在链上的酶结合。因此,这个例子强调了在计算不可逆性时如何定义粗粒度状态**I**和**II**时必须注意的问题。

## 细菌细胞分裂

多核苷酸的结扎提供了一个相对简单的测试案例,其中跟踪新复制子的形成可以简化为监测单个分子事件的进展(即磷酸二酯连锁的形成);然而,不可逆性和熵产之间的宏观关系同样适用于复杂得多的情况。实际上,我们现在要论证的是,这些基本的热力学约束甚至与整个单细胞生物的生长和分裂有关。

我们首先考虑制备一个大型系统,该系统最初包含单个大肠杆菌,浸入富含营养的培养基样品中,并与细菌细胞的最佳生长温度( $1/\beta \equiv T \sim 4.3 \times 10^{-21} \text{ J}$ )。<sup>17,18</sup>的热浴接触,我们可以进一步假设细胞在其分裂周期开始时处于指数生长阶段,并且,虽然整个系统的体积和质量是固定的,因此没有颗粒与任何外部溶液交换,容器的壁也不移动,但细菌周围营养介质的组成和压力模拟了向地球大气开放的通风良好的样品的组成和压力。如果我们用标签**I**来总结上面描述的实验条件,我们可以立即说,有一定的概率 $p(I|I)$ ,该系统处于某种特定的微观状态**I**(能量为 $E_i$ ),因为它是在宏观条件**I**下通过某种标准程序制备的。虽然这个概率很可能不可能从头算出来,但原则上它可以通过与微生物学家进行漫长的实验磋商来测量(至少在思想实验中)。

由于这里定义的“系统”不仅是细菌,还包括周围的肉汤,其中包含它将吃的食物和它将呼吸的氧气等,因此有三点是初步的。首先,展望我们对熵产的表述,我们必须考虑——在生长和分裂过程中,与细菌的吸入和排泄相关的熵变是系统内部熵变 $\Delta S_{\text{int}}$ 的一部分。有些出乎意料的结果是,例如,如果细菌依靠代谢反应生存,而代谢反应伴随着熵的大幅增加(例如通过厌氧气体evolution<sup>19</sup>),那么形成新细菌的总内部熵变实际上可能是正值。

其次,必须指出,内部香农熵 $S_{\text{int}} \equiv -\ln p \equiv -p \ln p$ 与我们熟悉的平衡统计力学中熵的概念之间的联系并不明显,需要详细说明。在玻尔兹曼分布系统中,由于内能与概率之间的既定关系,内部香农熵与与周围环境交换的热量之间有着自然的联系,这一点早已很清楚。直到最近才证明 $S_{\text{int}}$  continues

服从任意非平衡跃迁的一般热力学关系: $\Delta S_{\text{int}} \leq Q_{\text{ex}}/T_0$ 。这里, $\Delta Q_{\text{ex}}$ 是所谓的余热,它测量的是在转变过程中伴随各种耗散稳态的热量之上的额外热量演化。<sup>20</sup>也许更重要的是,我们期望即使远离平衡,香农熵仍然是系统中“统计无序”的度量,它是平衡系统。看到这一点的最简单方法是考虑一个统一的开始和结束分布的情况,也就是说,对于在各自的集合中具有非零概率的所有状态*i*和*j*, $p(i|i) = p_I$ 和 $p(j|II) = p_{II}$ 。显然,在这种情况下 $\Delta S_{\text{int}} = \ln[p_I/p_{II}] = \ln[\pi_I/\pi_{II}]$ ,也就是说,它只是测量**II**中比**I**中有多少更多或更少的状态。此外,更普遍的情况是,香农熵影响着相变前后相空间中体积的对数尺度比较。因此,影响这个远离平衡的量的因素应该与接近平衡时相同,即当粒子被安排为属于所讨论的系综时,系统中粒子可用的不同可能位置和速度的数量。因此,例如,气体部分体积的变化仍然与这个内部熵是如何变化的问题相关。

最后,应该指出的是,对于像大肠杆菌这样的微生物(它不进行任何形式的外部驱动的光合作用过程)来说,系统 $\pi(i \rightarrow j; \tau)$ 在任何时间间隔内 $\tau$ 都可以被认为服从(2)。换句话说,虽然细菌本身似乎是由化学反应物和产物的外部电流驱动的,但整个系统(包括营养物质)根本不被驱动,而只是与周围的热浴交换热量。因此,完全相反地,我们期望这个系统的最终稳态(对于任何体积和粒子数固定的系统,在无限长的时间内与热浴接触)将是微观状态上的玻尔兹曼分布,在这种分布中保持详细的平衡。那么,使这里感兴趣的场景成为一个远离平衡过程的原因仅仅是它的初始条件 $p(i|i)$ 对应于微观状态上的高度非玻尔兹曼分布。从这个意义上说,在松弛到平衡的漫长过程的最初阶段发生的细菌生长仅仅是短暂的,其燃料来自系统开始时微观排列中储存的化学能。当详细的平衡开始时,我们预计系统中最初的所有细菌早已长得超过了它们的容器,饥饿、死亡,并分解成它们的组成部分。

现在考虑一下,如果我们从某种微观状态*i*开始,然后让物质在 $\tau_{\text{div}}$ 的时间间隔内繁殖,这是单轮生长和细胞分裂的典型持续时间,那么我们的系统会发生什么。从生物学角度来看,该系统的预期最终状态是明确的:两个细菌而不是一个细菌漂浮在介质中,周围的各种原子重新排列成新的分子组合(例如,一些氧气转化为二氧化碳)。我们可以将与这样一个宏观结果相对应的未来状态的集合标记为**II**。假设,系统最初是在**I**中准备的,随后确实在**II**中结束,任何

微态 $j$ 将具有有限似然, 我们可以称之为 $p(j|II)$ 。

对于上面介绍的细菌细胞分裂过程, 我们的集合 $\Pi$ 是一个富含营养的培养液, 其中包含两个在分裂周期开始时处于指数生长阶段的细菌细胞。为了利用式(8)中的关系, 我们需要估计 $\pi(II \rightarrow I)$ , 即在时间 $\tau_{div}$ 之后, 我们最终会在安排 $I$ 中结束的可能性, 其中系统中只有一个新形成的细菌, 而另一个细胞以某种方式转换回构建它的食物。当然, 我们从来没有观察到细胞进行无数的生化反应是向后进行的, 就像我们从来没有看到冰块在沸水锅中自发地形成一样。尽管如此, 在非常完善的统计力学理论的假设中隐含着这样的事件具有非零(尽管非常小)发生的概率,<sup>1</sup>并且这些可能性可以使用我们可以通过实验测量的事件的概率率从上面限定。这是可能的, 因为系统的粗略物理特征足以对感兴趣的热力学量做出合理的估计; 由于我们最终感兴趣的是限定这个过程产生的热量, 所以我们只关心我们在对数尺度上估计的概率的影响; 我们必须将我们的概率估计改变许多个数量级, 才能看到对相应热界的任何影响。

第一部分相对容易想象: 虽然我们可能无法计算出细菌波动到肽大小的片段并呼吸一定量的二氧化碳和水的确切概率, 但我们可以确信它比细菌中所有肽键自发水解的可能性要小。令人高兴的是, 后一种可能性可以根据这类键的数量 $n_{pep}$ , 分裂时间 $\tau_{div}$ 和肽键半衰期 $\tau_{hyd}$ 来估计。

一个额外的复杂性来自于这样一个事实, 即细菌通常以每单位时间内形成的 $r = n_{pep}/\tau_{div}$ 肽键的速率生长。<sup>18</sup>为了建立一个细菌在一段时间内分解的概率模型, 我们需要指定在这段时间内细菌开始分解的时间, 这里有两个不同的因素发生碰撞。我们等待分解开始的时间越短, 它在这段时间内发生的可能性就越小。然而, 我们等待的时间越长, 在正常生长过程中合成的新肽键就越多, 因此, 随后整个细胞在剩余时间内自发解体 $\tau_{div}$ 的可能性就会呈指数级缩小。在这两种相互抵消的效应之间, 一定存在一个使解体概率最大化的最佳时间; 因此, 假设每个肽键以概率率 $\sim \tau_{hyd}^{-1}$ 独立于其他肽键自发水解, 因此我们可以将时间 $t$ 的总细胞水解概率 $p_{hyd}$ 建模为

$$\ln p_{hyd} \simeq (n_{pep} + rt) \ln[t/\tau_{hyd}]. \quad (13)$$

当 $t_{max}$ 满足 $\tau_{div}/t_{max} + \ln[t_{max}/\tau_{hyd}] + 1 = 0$ 时, 该数量达到最大值, 我们可以对两个输入时间尺度的选定值进行数值计算。在此之后

, 我们只需进行评估

$$|\ln p_{hyd}| \simeq |n_{pep} \ln[t_{max}/\tau_{hyd}]| = n_{pep}(\tau_{div}/t_{max} + 1). \quad (14)$$

我们现在已经处理了其中一个细胞的死亡。处理一个存活的细胞更具挑战性, 因为我们已经假设这个细胞在不断地生长, 我们不应该错误地认为这样的反应可以被一个小小的扰动所阻止或暂停。在大肠杆菌中, 指数生长期的开始之前是一个可以持续几个小时的滞后期,<sup>17</sup>在此期间, 基因表达发生了实质性的改变, 以便重新装备细胞, 以便在可用的代谢底物的推动下进行快速分裂。<sup>21</sup>因此, 我们可以适当地认为, 所讨论的细胞是一种优化的成分混合物, 准备参与营养代谢和蛋白质合成等不可逆反应。

因此, 我们可以认为, 在这些反应的进展中, 自发的、持续的暂停(持续时间 $\tau_{div}$ )的可能性确实非常小: 如果细胞的每个酶蛋白成分都拒绝亚底物扩散到其活性位点的每次尝试(假设小分子在蛋白质之间的扩散时间为 $\tau_{diff} \sim 10^{-8} \text{s}$ <sup>22,23</sup>), 我们预计 $|\ln p_{pause}| \propto |n_{pep}(\tau_{div}/\tau_{diff})|$ 将超过 $|\ln p_{hyd}|$ 数量级。然而, 我们必须考虑最有可能的 $\Pi \rightarrow I$ 过渡的另一种机制: 可能是细胞生长和分化可能在一个时间略低于 $\tau_{div}$ 。<sup>18</sup>如果后续此类事件, 最近的子细胞内有回文分解为它的组成营养物质(与对数概率最多的 $\ln p_{hyd}$ ), 我们会完成 $\tau_{div}$ 的间隔只有一个, 最近分裂, processively增长细菌在我们的系统, 也就是说, 我们会返回给我。因此, 通过细菌生物学为我们提供的一个进入 $I$ 的后门, 我们可以这样说

$$\ln \pi(\Pi \rightarrow I) \leq 2 \ln p_{hyd} \simeq -2n_{pep}(\tau_{div}/t_{max} + 1). \quad (15)$$

得到上述结果后, 我们现在可以参考我们为这个自我复制过程产生的热量所设定的界限, 并写出

$$\beta(Q) \geq 2n_{pep}(\tau_{div}/t_{max} + 1) - \Delta S_{int}. \quad (16)$$

这个关系表明, 在细胞自我复制过程中进化的热量不仅取决于将周围介质的分子组分排列成一个新的有机体所需的熵的减少, 而且还取决于这一过程发生的速度(通过分裂时间 $\tau_{div}$ )和我们必须等待新组装的结构开始分解的时间(通过 $t_{max}$ )。此外, 我们现在可以量化每个因素对最终结果的贡献程度, 以 $n_{pep}$ 为单位, 假设细菌的干质量为0.3皮克, 我们估计其为 $1.6 \times 10^{-9}$ 。<sup>24</sup>

大肠杆菌在溶原菌肉汤(多肽和葡萄糖的混合物)上以最大速率生长的单个分裂周期产生的总热量为 $\beta Q = 220n_{pep}$ 。<sup>17</sup>我们预计细胞分裂的内部熵变的最大贡献来自于氧气到二氧化碳的等摩尔转化(因为二氧化碳的分压明显较低)

大气), 以及在肉汤中自由漂浮的氨基酸被限制在细菌蛋白质内部的特定位置。我们可以通过注意  $\ln(v_{\text{co}_2}/v_{\text{co}}) \sim 6$  来估计第一个因素(增加熵)的贡献。从各种代谢物中释放碳也通过围绕振动和旋转自由度进行洗选来增加熵, 但我们只期望这会对每个碳原子分解的熵进行一些顺序单位修改。同时, 肽合成代谢降低熵值: 假设在1%的色氨酸肉汤中, 氨基酸开始时的探索体积为  $v_i = 100 \text{ nm}^3$ , 并最终紧密折叠在蛋白质的某些  $v_f = 0.001 \text{ nm}^3$  亚体积中, 我们得到  $\ln(v_i/v_f) \sim 12$ 。鉴于细菌在分裂过程中消耗的氧分子数量大致等于它创造的新细胞中氨基酸的数量,<sup>17,25</sup> 我们可以任意设置一个慷慨的上界-  $S_{\text{int}} \leq 10n_{\text{pep}}$ 。

为了比较这对(8)中的不可逆性项的贡献, 我们假设细胞分裂时间为20分钟, 在生理pH值下, 肽键的自发水解寿命为0.02 ~ 600年, 这将产生  $t_{\text{max}} \sim 1$  分钟和  $2n_{\text{pep}}(\tau + 1) = 6.7 \times 10^{-\langle e \rangle}$ , 这个量至少要比大几倍。我们经常认为生物自组织必须克服的主要熵障碍是以适当的方式组装生物组件的成本。

然而, 在这里, 我们有证据表明, 有氧细菌呼吸的成本相对较小,<sup>19</sup> 并且由于自我复制反应的绝对不可逆性, 它在大量生产不易分解成其组成部分的副本时, 大大超过了这种成本。

更重要的是, 这些计算还表明, 大肠杆菌产生的热量小于其生长速度、内部熵产和持久性所决定的绝对物理下界( $220n_{\text{pep}}/42n_{\text{pep}}$ )的六倍。鉴于细菌是其环境的复杂传感器, 可以非常有效地适应各种不同环境的生长, 我们不应该对它没有针对其中任何一种环境进行完美优化而感到惊讶。相反, 值得注意的是, 在单一的环境中, 有机体可以如此有效地将化学能转化为自己的新副本, 即使它要产生四分之一的热量, 它也会突破热力学可能的极限! 这一点尤其重要, 因为我们在计算  $p_{\text{hyd}}$  时故意低估了逆反应速率, 这并没有考虑到不可能自发地将二氧化碳转化回氧气。因此, 未来对  $\beta Q$  下限的更准确估计可能会揭示大肠杆菌是一种比目前看起来更加特别适应的自我复制因子。

与此同时, 值得记住的是, 除了前面分析中探讨的热力学约束外, 任何自我复制因子都面临其他约束。任何复制自身的实体都是由一组特定的材料构成的, 这些材料具有一组偶然的属性和相互作用或相互组合的模式。纯粹从本工作导出的热力学第二定律的推广角度来看, 在消耗相同热量的情况下,

*E. coli* 的生长速度可能比现在快得多。一旦考虑到细菌如何复制的偶然性(以及, 例如, 它们如何使用核糖体来准确地合成蛋白质机制, 使它们能够在分子 level<sup>27</sup> 上发挥作用), 就有理由预计, 生物体实际上面临着更多限制性的生长障碍, 这些障碍来自整个过程所依赖的限速反应的速度。

## 闭幕词

细胞分裂的过程, 即使是在像细菌这样古老而流线型的生物中, 也是如此复杂得令人眼花缭乱, 以至于物理学可以对这一切发生的速度做出任何有约束力的声明, 这可能会让人感到惊讶。这一切成为可能的原因是, 恒温浴中的非平衡过程遵循将正向和反向转变概率与产热联系起来的一般定律。<sup>2</sup> 此前, 此类定律已成功应用于理解复制“信息”分子(如核酸)的热力学。<sup>8</sup> 然而, 在这些情况下, 考虑到DNA在RNA和蛋白质的产生中所扮演的明确角色, 系统分子结构的信息内容更容易被视为理所当然。

我们在这里瞥见的是, 熵产生和跃迁概率之间的潜在联系具有更普遍的适用性, 只要我们认识到, 只有当观察者决定如何对系统中的“自我”进行分类时, “自我复制”才可见: 只有当粗粒度方案确定了每个微状态下存在多少个对象的副本时, 我们才能用概率能力的术语来谈论该类型对象影响其自身复制的一般趋势, 并且使用任意数量的不同方案, 同一系统的微状态可以是粗粒度的。然而, 无论采用何种方案, 由此产生的随机种群动力学必须遵循相同的一般关系, 即纠缠着热量、组织和持久性。我们可能希望, 这一见解能激发未来的工作, 澄清非平衡系统中自然选择所遵循的一般物理约束。

## 致谢。

作者感谢 C. Cooney, G. Crooks, J. Gore, A. Grosberg, D. Sivak, G. Church, A. Szabo 和 E. Shakhnovich 的有益评论。

<sup>1</sup> C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. (Springer, 2003).

<sup>2</sup> G. E. Crooks, *Phys. Rev. E* **60**, 2721 (1999).

<sup>3</sup> R. A. Blythe, *Phys. Rev. Lett.* **100**, 010601 (2008).

<sup>4</sup> A. Gomez-Marín, J. M. Parrondo, and C. Van den Broeck, *Phys. Rev. E* **78**, 011107 (2008).

<sup>5</sup> G. Verley, R. Chétrite, and D. Lacoste, *Phys. Rev. Lett.* **108**, 120601 (2012).

<sup>6</sup> R. Landauer, *IBM J. Res. Dev.* **5**, 183 (1961).

<sup>7</sup> M. A. Nowak, *Trends Ecol. Evol.* **7**, 118 (1992).

<sup>8</sup> D. Andrieux and P. Gaspard, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9516 (2008).

<sup>9</sup> T. A. Lincoln and G. F. Joyce, *Science* **323**, 1229 (2009).



- <sup>10</sup> J. E. Thompson, T. G. Kutateladze, M. C. Schuster, F. D. Venegas, J. M. Messmore, and R. T. Raines, *Bioorg. Chem.* **23**, 471 (1995).
- <sup>11</sup> C. A. Minetti, D. P. Remeta, H. Miller, C. A. Gelfand, G. E. Plum, A. P. Grollman, and K. J. Breslauer, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14719 (2003).
- <sup>12</sup> H. J. Woo, R. Vijaya Satya, and J. Reifman, *PLoS Comput. Biol.* **8**, e1002534 (2012).
- <sup>13</sup> G. K. Schroeder, C. Lad, P. Wyman, N. H. Williams, and R. Wolfenden, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4052 (2006).
- <sup>14</sup> D. Y. Zhang, A. J. Turberfield, B. Yurke, and E. Winfree, *Science* **318**, 1121 (2007).
- <sup>15</sup> W. Gilbert, *Nature (London)* **319**, 618 (1986).
- <sup>16</sup> Z. Kelman and M. O' Donnell, *Annu. Rev. Biochem.* **64**, 171 (1995).
- <sup>17</sup> H. P. Rothbaum and H. M. Stone, *J. Bacteriol.* **81**, 172 (1961).
- <sup>18</sup> P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, *Curr. Biol.* **20**, 1099 (2010).
- <sup>19</sup> U. von Stockar, T. Maskow, J. Liu, I. W. Marison, and R. Patino, *J. Biotechnol.* **121**, 517 (2006).
- <sup>20</sup> T. Hatano and S. Sasa, *Phys. Rev. Lett.* **86**, 3463 (2001).
- <sup>21</sup> D. E. Chang, D. J. Smalley, and T. Conway, *Mol. Microbiol.* **45**, 289 (2002).
- <sup>22</sup> D. Brune and S. Kim, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3835 (1993).
- <sup>23</sup> S. C. Blacklow, R. T. Raines, W. A. Lim, P. D. Zamore, and J. R. Knowles, *Biochemistry* **27**, 1158 (1988).
- <sup>24</sup> F. C. Neidhardt, *E. coli and Salmonella: Cellular and Molecular Biology* (ASM Press, 1990), Vol. 1.
- <sup>25</sup> C. L. Cooney, D. I. Wang, and R. I. Mateles, *Biotechnol. Bioeng.* **11**, 269 (1969).
- <sup>26</sup> A. Radzicka and R. Wolfenden, *J. Am. Chem. Soc.* **118**, 6105 (1996).
- <sup>27</sup> C. G. Kurland and M. Ehrenberg, *Annu. Rev. Biophys. Biophys. Chem.* **16**, 291 (1987).