

# 植物生理学问题分析

刘沛雨 信息科学技术学院

信息论是一门研究信息传输和处理的数学理论。1948 年，克劳德·香农（C. E. Shannon）在《通信的数学原理》一文中首次提出了信息论的基本概念与理论框架，系统地介绍了信息的定义、信息的度量方法等基本问题，并提出了信息熵的概念。信息论允许我们在不完全了解细胞信号通路网络的情况下分析细胞信号的功能，从而对细胞信号转导有更为直观的理解。<sup>①</sup>本文将着重于细胞的信息处理功能，利用信息熵的概念和数学方法，从信息论的视角解读细胞行为。

细胞在本质上与计算机程序有类似之处。对细胞进行数学建模，我们可以将细胞视为一个信息处理程序：输入是细胞内外环境的刺激，输出是细胞对刺激做出的反应，而程序本身则是细胞在特定刺激下的信号通路网络。但细胞远比计算机程序复杂，二者最显著的区别在于计算机程序有着严格的逻辑性、精确性与可重复性，即对同样的输入执行程序得到的输出都是一致的；而细胞则并不如此，实验中细胞在不同状态下可能对同样的刺激做出不同的反应，这说明细胞这一信息处理程序本身具有随机性（如分子扰动和热力学波动等），而非预先设定好的精确的程序。细胞团中同样也有不确定性，其中单个细胞的状态不断发生变化，导致每个细胞对特定刺激的具体反应难以预测。对于具有不确定性的过程，需要通过引入随机性来进行分析，而信息熵可以很好地定量体现信息的不确定性，为研究细胞行为的随机性提供工具。

信息熵源于物理学中熵的概念，二者都可表征系统的不确定性。随机变量 $X$ 的信息熵 $H(X) = -\sum_x p(x) \log_2 p(x)$ ,  $x \in X$ ，由信息熵的计算公式可知当 $\forall x \in X, p(x) = 0$ 或 $p(x) = 1$ 时， $H(X) = 0$ ，此时信息的确定性最高（ $\forall x \in X$ ， $x$ 要么不可能发生，要么必然发生）；当 $\forall x \in X, p(x) = \frac{1}{n}, n = |X|$ 时， $H(X) = \log_2 n$ ，此时信息的不确定性最高（ $\forall x \in X$ ， $x$ 等概率发生）。从熵的角度看，生命的本质是通过能量输入维持自身低熵状态，细胞通过信号通路对刺激做出反应的目的也是维持自身的稳定，即维持自身内部信息的较高确定性。

---

<sup>①</sup> How Information Theory Handles Cell Signaling and Uncertainty. Science 338, 334-335 (2012). DOI: 10.1126/science.1227946

细胞作为一个信息处理系统，我们也可以根据信息熵的概念对其建模并利用数学方法分析。假设细胞内初始的信息熵为 $H_0$ ，某一刺激 $s$ 在该细胞中有 $n$ 种信号通路  $P = \{path_1, path_2, \dots, path_n\}$  对应细胞的  $n$  种不同反应  $A = \{action_1, action_2, \dots, action_n\}$ ，细胞对 $s$ 在 $path_i$ 下做出反应后的信息熵为 $H_i$ ，那么  $G_i = H_i - H_0$  表征 $path_i$ 对应的反应效果， $G_i$ 越小说明细胞在当前状态下根据 $path_i$ 做出的反应可以越好地维持自身低熵状态，因此从熵的观点来看，细胞会“选择”  $path_k, k = \operatorname{argmin}_i \{G_i | i = 1, 2, \dots, n\}$  来作为在当前状态下应对刺激 $s$ 的主要信号通路，对应的反应为 $action_k$ 。下面考虑含有 $m$ 个细胞的细胞集合  $C = \{cell_1, cell_2, \dots, cell_m\}$ ， $C$ 中每个细胞由于所处状态不完全相同（即初始的信息熵不完全相同），在同一刺激 $s$ 下也可能通过 $P$ 中不同的信号通路做出不同的反应来维持自身较低的信息熵水平，假设主要做出反应 $action_k$ 的细胞数目为 $m_k, \sum_{i=1}^n m_i = m$ ，此时 $C$ 整体的信息熵 $H_C$ 是 $m_k, k \in \{1, 2, \dots, n\}$ 的多元函数，可以通过数学方法分析当 $m_k, k \in \{1, 2, \dots, n\}$ 分别取何值时 $H_C$ 最小，从而定量预测 $C$ 中细胞对刺激 $s$ 反应的具体情况。<sup>①</sup>

以上的例子体现了信息论在预测细胞行为中的作用，通过数学建模我们可以量化细胞内的不确定性水平和各种随机性（如分子扰动和热运动等）对细胞信号转导的影响，可以帮助我们更好地理解信号转导的机制，还为预测细胞团中单个细胞的行为提供了数学方法。由此可见，信息论是可以用于解读细胞行为的，通过引入随机性可以更好地模拟信号转导的实际过程，从而更精确地预测细胞行为。除了预测细胞行为，信息论还可以用于细胞信号转导网络的拓扑结构分析和量化细胞信息流，为生物学研究提供了一个全新的视角。

但这一方法也有明显的局限性，生理状态下细胞所处环境十分复杂，细胞与环境的信息熵难以量化，并且数学的分析方法虽严格但也与实际情况相去甚远，理论的分析可能很难应用于实际。以前文的建模为例，推理的大前提是“细胞需要维持自身的低熵水平”，但是在多细胞生物中并非如此——胚胎发育过程常常伴随着细胞的程序性死亡，这一过程也受细胞信号的调控，显然这与推理的前提相悖的。因此通过信息论预测、解读细胞行为仍存在一些问题有待解决。

---

<sup>①</sup> 以上两段有关概念参考自 Shannon C., *Bell Syst. Tech. J.* 27, 379 (1948).