

物理学和生物学(下)*

郝柏林^{1,2}

(1 复旦大学理论生命科学研究中心 上海 200433)

(2 中国科学院理论物理研究所 北京 100080)

3 生物问题的诸多层次

对生命现象的研究必须在众多层次上进行. 最微观的层次包含分子和原子的相互作用, 从小分子(糖、脂肪、核苷酸、氨基酸)、钙镁钾钠等金属离子、水分子, 到生物大分子(DNA, RNA, 蛋白质以及它们的复合体)的结构和相互作用, 特别是中心法则所涉及到的复制、转录、剪接、翻译, 以及翻译之后的修饰、运输, 还有各个阶段上的查错和修补. 然后是各种基本构件所组成的途径和网络: 调控网络、代谢网络、免疫网络、信号传导网络等等. 再往上是细胞、组织、器官乃至个体, 从生物化学到生理学的研究. 更上层次则是种群动力学、生物多样性、生态系统, 以及最近颇为时髦的生物复杂性和系统生物学的研究. 各个层次上的问题, 其生物、物理和数学的“含量”有所不同. 一般说来, 底层上的物理更多, 而顶层涉及较多的数学模型. 瞄准一定层次进行研究, 就是物理学中的粗粒化方法. 适当的粗粒化可以揭示精确规律, 而过分注重细节会掩盖事物的主流. 正确的研究策略, 是从问题出发, 寻求和创造合适的工具. 下面从不同的层次选几个例子做简要的讨论.

3.1 分子手性破缺

构成蛋白质的 20 种氨基酸, 除了最小的甘氨酸以外, 都有左手和右手两种异构体, 因而都具有光学活性. 手性分子互为镜像, 它们不可能靠连续的位置变换重合到一起. 偏振光穿过含手性分子的溶液时, 偏振面会发生旋转. 实验室中用化学方法合成的氨基酸都是左右手性各含一半, 但生物体里的氨基酸绝大多数都是左旋的, 天然蛋白质都由左旋氨基酸组成. 然而, 天然蛋白质和 DNA 分子却多数具有右手螺旋结构. 这是从“大爆炸”后发生正反粒子对称破缺以来, 在生命物质层次上表现出来的对称破缺. 地球上所有生物具有一致的手性, 大家才可能处于同一条食物链的不同位置. 否则, 餐厅也必须分

“左”“右”, “左派”不能消化和吸收右旋食品.

如果在陨石中发现某种单一手性的有机分子, 就可能预示宇宙中还有生命存在. 至今还没有这样的报道. 然而, 生物分子的手性破缺是怎样发生的? 许多学者仍然在研究这个根本性的问题. 例如, 有人试图把它和弱相互作用中的宇称不守恒联系起来. 然而, 从非平衡相变和对称破缺的一般理论看, 发生破缺的具体原因可能是次要的, 而对破缺状态的放大机制则起着决定作用. 我们不妨用另外一个层次的生物现象说明这个观点.

蚂蚁是研究昆虫社会行为的良好对象. 有人设计过如下的实验. 在距离蚁巢同样远的两个完全对称的地点, 置放等量食物. 两条对称道路各自经过一座小桥, 因此可以明确分开. 实验开始不久, 两处食物都被发现, 两条路径上有数目大致相等的蚂蚁在搬运食物. 然而, 经过一段时间后, 出现了对称破缺: 大批蚂蚁集中到一条道路上. 要阐明最初的破缺是怎样发生的, 既不容易、也无大必要. 重要的是蚂蚁间交换两地信息的次数, 一旦发生破缺就会继续放大. 正如单轴磁性材料从顺磁到铁磁的相变, 铁磁相的具体磁化方向其实是由通常略而不计的细微因素如地磁场或杂质决定的.

3.2 分子马达

车轮并不是人类独有的发明, 大自然早就进化出精巧的蛋白质转动机器. 细胞中的生物化学反应过程都需要消耗能量. 这些能量储存在三磷酸腺苷(简称 ATP, 即腺三磷)分子中. 从外界摄入的营养, 如葡萄糖, 在降解过程中合成一定数目的 ATP 分子. 腺三磷提供能量后成为腺二磷(ADP). ADP 要重新“充电”成 ATP, 才能继续发挥作用. 这充电过

* 摘自《大学物理(当代物理学前沿专题部分)》(第 2 版, 高等教育出版社), 征得作者和出版者同意
 国家科技部“九七三”(批准号: G2000077308)、国家自然科学基金(批准号: 30170232)资助项目

程是在称为 ATP 合成酶的蛋白质分子马达(见图 2)里实现的. ATP 合成酶由 F_0 和 F_1 两个蛋白质复合体亚基组成. 下面的 F_0 亚基嵌在线粒体内膜里, 而 F_1 亚基像是一台直流电机的定子, 由两种共 6 块蛋白质组成. 在三度对称体的中心, 有一条从 F_0 亚基出来、贯穿 F_1 亚基的蛋白质轴. 正氢离子即质子穿膜移动造成的电位梯度, 导致中轴旋转. 中轴转到一定角度, F_1 亚基的一块蛋白张开, 纳入一个 ADP 分子, 把它重新变成 ATP, 并在另一个角度释放出去. 在这个分子马达中, 化学能和机械能的转化效率很高, 因此现在仍是重要的研究对象.

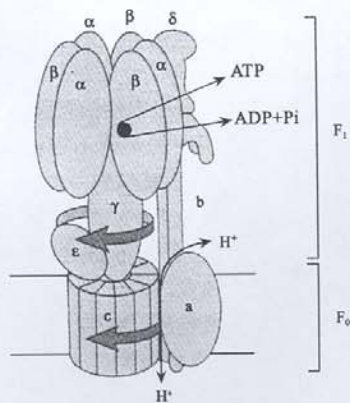


图 2 ATP 合成酶蛋白质分子马达示意图
 [图中希腊字母标示各个蛋白质块. F_0 亚基嵌在膜中. 本图改画自生物能和生物膜杂志(J. Bioenergetics and Biomembranes)2000 年第 32 卷 442 页 Y. Sambongi 等的文章]

分子马达不仅是转动类型的. 许多真核细胞里存在由蛋白质镶嵌成的细胞骨架. 这些骨架不仅支撑起细胞腔体, 协助实现细胞运动, 而且还充当细胞内输送物质的轨道. 原来有些特定的分子被包裹在小小的膜泡里, 膜泡借助由蛋白质形成的“双脚”或“多足”, 在细胞骨架上“行走”, 把物质输送到目的地.

无论是定向的旋转或行走, 都发生在随机变化的背景上. 在随机噪音背景上产生定向的、看起来像是有目的性的运动, 是发生在生物活动各种层次上的普遍现象. 非平衡统计物理学中经典的朗之万方程, 近些年发展的随机共振理论, 都会对解释这类现象有所贡献.

细胞在液体中的运动, 在更大尺度上提出“马达”问题. 许多细菌在液体环境中有趋向食物和躲避毒物的行为, 即“趋化性”(chemotaxis). 这通常是借助鞭毛和纤毛的运动实现的. 鞭毛是一条长在细

胞端部的尾巴, 其穿膜部分由若干蛋白质组成的, 具有一定的对称性, 像是一套装配精巧的“滚柱轴承”. 这是规模更大、结构更复杂的一类分子马达. 纤毛分布在细胞膜外, 数目较多, 它们的随机或协调摆动, 使细菌在原地停留或往某一方向移动. 需知单细胞的细菌, 并没有什么神经系统来控制各部分的动作, 看来有目的的运动是化学物质的浓度梯度所导致的一系列物理效应的后果.

3.3 蛋白质折叠

前面介绍“中心法则”时已经提到, 在核糖体中合成的蛋白质, 要折叠成特定的结构, 才能发挥生物功能. 特别是起催化作用的各种酶, 往往折叠成接近球形的特定形状. 形状不对, 就不能发挥作用, 甚至还会导致病变. 疯牛病就是一种蛋白质折叠病.

细长的高分子链, 由于侧支集团围绕化学键的旋转, 可以形成各种空间构象. 随着分子量增加, 可能的构象数目迅速上升. 构象数目通过熵 S 影响热力学性质, 出现在自由能的公式 $F = E - TS$ 中(这里 E 是内能, T 是温度). 要理解橡胶的弹性, 就必须考虑构象数目的贡献. 同一种高分子的集合, 其个体长度可以不一样, 而具有一定的分子量分布. 达到一定浓度的高分子溶液, 在温度变化时会发生溶胶凝胶相变. 就单个大分子而言, 它会处于延展和卷曲两种状态. 同处于卷曲状态的两个高分子, 其构象可能颇为不同. 高分子的构象统计学和橡胶的弹性力学早就是发展成熟的学科.

相对于高分子构象, 蛋白质折叠是完全不同的问题. 一个特定的蛋白质由固定的氨基酸序列组成, 它的长度和分子量都是确定的. 这是蛋白质的一级结构. 新合成的肽链的某些段落会以较快的速度形成局部的二级结构. 常见的二级结构有 α 螺旋和平行或反平行的 β 片, 它们都是由隔一定距离的氨基酸之间的氢键维系着. 这些含二级结构的肽链再以较慢的速度折叠成最终的三级结构. 来自多个蛋白质的三级结构会再形成多个亚基的复合体, 是为蛋白质的四级结构.

从蛋白质的一级结构即氨基酸的字母顺序, 预测二级结构已经有一些较好的算法和程序. 进一步预测三级结构, 则仍然是颇具挑战性的难题. 首先, 已经测定的近 2 万个蛋白质的三维结构, 只包含不超过一千种基本的折叠单元. 为何基本折叠单元的种类这么少? 这是一个物理学家们感兴趣的问题, 现在已经有了部分答案. 原来, 稳定的折叠单元不仅要满足某种能量最低的条件, 而且要具有惟一性, 即

同一条一级序列不允许折叠到两个以上能量同样低的结构。那些折叠结果不惟一的序列,已经在进化过程中被淘汰。

其次,生物学家更感兴趣的问题是特定的一级序列已知的蛋白质会折叠成什么样子,发挥何种生物功能。现在全世界每天可以从测序得到的 DNA 序列中翻译出上千条可能的蛋白质序列,而所有测定蛋白质结构的实验室目前每个月总共只能给出约 200 个结构。用实验方法确定所有重要蛋白质的结构,绝非短期内能做到的事情。用理论方法从蛋白质一级结构预测其二级和三级结构,是当前的紧迫课题。

决定蛋白质折叠的基本相互作用力,其物理本质是清楚的。然而从“第一原理”出发,直接计算折叠过程仍然超出现有的最大计算机的能力。美国 IBM 公司在 2000 年提出一个名为“蓝色基因”的计划,要用 5 年时间,研制一台有 1 百万个处理器的大规模并行系统,以便用 3 天 CPU 时间,算出一个实际蛋白质的折叠过程。这一计划即使顺利实现,也只能表明人类对蛋白质折叠过程的认识水平,而不可能在近期内用于解决日常问题。在相当长的时期内,人们还必须依靠已经积累的实验数据,用各种粗粒化方法做推测。这就进入后面将要介绍的生物信息学领域。

3.4 生物膜的形状

生物细胞和细胞里的各种“细胞器”都包裹在由磷脂形成的双分子膜里。这些生物膜可能具有奇形怪状,例如带棱带角的膜泡,细长的波纹软管,或是复连通的、高“亏格”的表面,而不同于简单的球形。“亏格”是一个拓扑学概念,球和椭球的亏格为 0,环的亏格为 1,等等。从理论上说明各种生物膜的形状,是远非平庸的课题。血红细胞的形状是一个实例。

人体中的血红细胞的形状,是一个上下两面都凹陷进去的旋转对称的圆饼。这种形状很不容易从理论上说明,因为任何借助表面张力和考虑膜内外压力差的理论都会导致向外膨胀的球形。历史上有过多种并不成功的尝试。然而,注意到磷脂膜本身既非液体,亦非固体,而更接近液晶(液晶),生物膜的液晶模型获得很大成功。它不仅给出了血红细胞形状的解析表达式,而且说明了许多其他类型的生物膜形状。

生物膜的形状问题,是说明描述尺度和粗粒化方法的极好的实例。一提到生物膜,人们就会想到嵌在膜上或穿过膜的各种各样的蛋白质、孔洞和离子通道。然而研究膜的形状时,其实没有必要考虑这些

细节,其原因在于它们是尺度相差甚远的两类对象。膜蛋白和蛋白质镶嵌而成的各种通道,是纳米(10^{-9} m)尺度的对象。研究离子通道时,可以在无穷大的平面上模拟,曲率其实没有影响。膜泡通常是微米(10^{-6} m)尺度的对象。研究膜泡时根本看不清嵌在膜上的纳米级对象,膜蛋白的种类或数量只是略为改变宏观描述中使用的一些物质参数。把几何尺度相差千倍的因素混淆在一起,不是提高描述精度,而是模糊研究重点,导致顾此失彼的后果。恰当的粗粒化描述可以得出精确结论。这是一条重要的方法论原则。

生物膜的液晶模型,是连续介质理论的成功应用。单个的生物大分子在适当的粗粒化处理中,可以当成连续对象。连续介质理论用于说明 DNA 和 RNA 分子的力学性质,近来也获得许多与实验符合的结果。二维生物膜和一维大分子的连续介质描述,都是活跃的研究领域。感兴趣的读者可以参阅本文参考文献中列出的一本小册子。

3.5 粘菌的自组织行为

盘基网柄菌(*dictyostelium discoideum*)俗称粘菌,是研究细胞发育、运动和信号传导的方便的模式生物。在生存环境良好时,它们作为单细胞生活,靠细胞分裂增殖。当一个或多个粘菌发现环境恶化、不利于生存时,就向周围发送信号。这信号就是环腺苷酸小分子 cAMP。收到信号的粘菌,也向其周围发送同样的信号。在一片互相收发的信号背景中,自发形成一个聚集中心,大家都向中心靠拢。大量粘菌聚集后,长出一支像多细胞植物的菌株,菌株顶部的小球体破裂后,成熟的孢子随风飘向远处,其中到达有利的生活环境者,继续生长分裂。这样,靠“牺牲”约 80% 的个体,给其余个体提供远走高飞的机会,延续着粘菌物种。

从大量粘菌以自己为中心向四面发送 cAMP 信号,最初形成均匀的随机背景,后来突然自发产生一个聚集中心。这是一种以自发对称破缺为特征的非平衡相变,是自组织行为的典型事例。人们对此进行过许多实验和理论研究。

3.6 生物系统的标度性质

生命现象所涉及的对象,从生物大分子、细胞器、细胞、组织、器官、个体、种群到生态系统,其尺度跨越近 20 个数量级。生物系统的一些特征量如何随尺度变化,是一个值得注意的问题。

考察一个独立生活的单细胞生物。为了维持细胞膜内的生物化学反应,它必需通过细胞膜同外界

交换物质. 不难估计细胞膜所包容的体积和质量, 也容易算出细胞膜的面积, 从而得到维持单位体积或质量所必需的表面积. 如果把这一比例关系简单地外推到像一个人或一棵树那么大的质量, 维持生命活动所需的表面积是颇为巨大的. 树木要生长出由粗到细的枝叶, 高等动物要发育出从主动脉、支动脉到微血管的循环系统, 从气管、支气管直到肺泡的呼吸系统等等, 其物理原因概在于此. 体积比例于线度的3次方, 面积比例于线度的平方. 因此, 如果某个特征量 Q 随线度 L 的变化规律是

$$Q \propto L^{n/3},$$

而其标度因子 $n/3$ 中的 n 的值接近整数, 这倒是不足为奇的.

然而, 近几年的观察和统计表明, 有许多特征量, 例如一个生物系统的代谢速率 R , 其标度因子却是 $1/4$ 的倍数, 即:

$$R \propto L^{1/4}.$$

虽然已经有一些解释这件事的理论模型, 但争论目前仍在继续.

4 生物信息学和计算生物学

生物有形, 人所共识. 地球上自然之美, 多数来自生物本身和它们的营造. 生物有数, 则经历了漫长的认识过程. 1885年恩格斯在比较各门自然科学使用数学的情况时, 曾经写道: “生物学中的数学 = 0”¹⁾. 如果不考虑当时在生物学界还未引起注意的孟德尔 (Mendel G J) 遗传定律中的排列组合分析, 这一观察大致是正确的.

4.1 生物数据

然而, 生物不仅“有数”, 而且是各种各样的大数. 其实, 仅仅生物多样性的宏观数据就是相当可观的. 地球上现存多少物种? 从几百万到几亿的估计都有. 细菌是地球上最成功的物种, 早在出现人类之前, 它们就已经生活了三十多亿年, 并且为其他物种的生存创造着环境; 人类如果把环境破坏到自身无法继续生存, 细菌们可能还会愉快地存活几十亿年. 细菌究竟有多少种类和个体? 我们只对其中与人类生活和疾病密切相关的种属略有所知. 2002年11月初, 在核酸序列数据库中有13.6万种生物至少有一条序列被收存, 其中只有5万种有某些分类学信息. 目前地球上大约有5千种哺乳动物, 它们是曾经生存过的20多万种哺乳动物中的胜出者. 如果把研究范围延伸到古生物物种, 那真是“不计其数”.

不过, 现代生物学数据的主体是分子水平的核酸序列、蛋白质序列和结构数据. 仅以中国科学院基因组学研究所, 即代表我国承担了国际人类基因组计划1%测序主攻任务最近又完成了水稻全基因组框架图和精细图的“华大基因”为例, 其测序能力已经达到每天3000万碱基对, 也就是每天产生 3×10^7 、每年 10^{10} 个字母. 全世界生物数据的产出量, 目前已经达到每年 10^{15} 字节, 而且还在继续增长. 国际上三大核酸序列数据库之一的 GenBank, 在2002年12月15日发布的第133版包含2231万条序列, 共计285亿个字母. 几年之内, 生物学就会成为人类科学活动中产生数据最多的领域.

核酸序列数据中最重要、近几年增长最快的部分, 是所谓完全基因组, 即一个生物体的全部遗传密码. 现在已经发表了近100种细菌和古细菌的完全基因组. 真核生物基因组已经完成的有酵母、线虫、果蝇、拟南芥等. 正在进行大规模测序或已完成了工作草图的物种有人、水稻、小鼠、按蚊、疟原虫等. 一个人的基因组有 3.2×10^9 碱基对, 编码3万多个基因. 水稻基因组有 4.3×10^8 碱基对, 可是其基因数目可能比人多一倍. 仅仅为了存放这些数据, 就需要大量设备. 英国的 Sanger 测序中心估计, 今后5年中每年的硬盘增长量为100TB, 即 10^{14} 字节.

4.2 大数和计算机

我们已经列举了不少大数, 需要对它们再多一点感觉. 首先是计数单位, 每3个数量级换一个名字和缩写: 10^3 称千 (Kilo, K)、 10^6 称兆 (Mega, M)、 10^9 称吉 (Giga, G)、 10^{12} 称太 (Tera, T)、 10^{15} 称拍 (Peta, P)、 10^{18} 称艾 (Exa, E). 其次, 这样的大数是怎样一个概念呢? 从大爆炸产生我们所在的宇宙, 大约已经过去 4×10^{17} 秒, 即0.4个“艾秒”. 有人估计地球上所有存活过和现存的智人所说的话语, 不超过1艾个字. 可见, 生物学所涉及的数字已经处在人类曾经设想过的大数的边缘. 为了处理这些数据、从中提取知识, 必须依靠人类自学会用火以来的又一项伟大发明——电子计算机.

把电子计算机与用火相提并论, 并不过分. 因为人类在一切活动领域里, 都没有可以同计算能力增长速度相比拟的成绩. 1944年研制的 ENIAC 计算机, 每秒可以完成330次乘法运算, 能用短于炮弹飞行的时间把弹道计算出来. 这在当时真是巨大的成绩. 我国前几年研制的神威计算机, 每秒可以执行

1) 恩格斯. 于光远等译. 自然辩证法. 北京: 人民出版社, 1983. 172

3840 亿次浮点运算,曙光 3000 计算机每秒可作 4 千亿次浮点运算. 每秒 10 万亿,即 1“太”次浮点运算的计算机也已问世. 50 多年间,人类的运算能力提高了 10 亿倍,而步行与乘坐喷气飞机,速度之差不过几百倍. 试问还有那个领域有如此的飞跃?

4.3 基因组信息学

生物的许多道理藏在数据中. 直接从生物数据出发的研究领域现在称为生物信息学. 由于我们已经出版过关于生物信息学的专书(见参考文献),这里不再详述,而只选列几件与基因组研究有关的事情.

大规模测定 DNA 序列,要先把总量大于基因组长度若干倍的原始材料,打碎成 500 个字母左右的短片分别测序,最终再拼接成数百万字母乃至更长的染色体序列. 基因组中的大量重复序列给拼接造成巨大困难. 人们虽已做过不少工作,仍然需要发展更有效的数学理论和算法.

成功拼接出的 DNA 序列,是由 4 种字母组成的不分段落、没有标点符号的“天书”. 要从这些天书中找出编码蛋白质的基因,不编码蛋白质但起着重要作用的各种 RNA,以及各种控制调节基因表达的信号. 马尔可夫和隐马尔可夫模型、动态规划、神经网络、各种判别和聚类方法、语言学分析等等,都已用于寻找基因. 目前一切较好的这类程序都严重地依赖于“训练数据”,即已经积累的生物知识. 随着大规模测序成为家常便饭,人们必须研制尽量少依赖训练数据的、能在寻找过程中自我完善的算法.

从基因注释得到的大批蛋白质序列,早已不可能只靠实验室工作来确定结构、发现功能. 用计算机来预测蛋白质的结构和功能,已经是十分活跃的研究领域. 这里包括蛋白质折叠问题. 由于前面已经讲述过,此处从略.

基因芯片的发展,使人们可以同时研究大量基因的表达过程,比较它们在正常生理条件和病理条件下的表达差异. 这里的许多算法还没有走出聚类分析,而人们期望从这些数据中能提取各种网络的结构和参数.

随着完全基因组数据的生长,比较基因组学的课题也越来越多. 由现存物种数据回溯进化过程的研究,也面临新的挑战.

在这一切研究的背后,是大量数据的采集、质量检验、数据库的建立和管理. 只有发展出对用户友好的、跨越各种计算机平台的数据库和应用软件系统,才能使生物工作者充分享用数据、进行创造. 这里计算机科学有着广阔的用武之地.

4.4 计算生物学

生物信息学和计算生物学这些提法往往被交替使用,不过,它们之间是有一些差别的. 一般说来,需要从已知方程或模型出发的研究,更接近计算生物学;建立在数据库搜索和比较基础上的工作,较多地属于生物信息学. 科学研究不应从定义出发,重要的是提出和解决问题. 对于一个细胞里全部生物化学反应的数值模拟,乃至对组织、器官和整个生物体的模拟,都已经提到研究日程上. 对人体组织和器官的生理与病理状态的计算机模拟,最终会解放大量的实验动物. 那时的生物学和医学就更离不开数理科学和计算机了.

虽然已经有过牛顿方程和麦克斯韦方程这样的伟大理论篇章,19 世纪的物理学仍被物理学家们自认为是实验科学. 20 世纪上半叶,随着量子力学和相对论的建立,物理学成为理论和实验密切结合的科学. 20 世纪下半叶,电子计算机的迅猛发展为自然科学提供了认识世界的强大新工具. 物理学既为计算技术的发展准备了材料、元件和理论,又是这一发展的最早受益者. 于是,物理学终于成为鼎立在实验、理论和计算三足之上的成熟的科学. 生物是物,生物学的发展也会从物理学的既往得到启发.

5 发展交叉科学、培养广谱人才

瞻望新世纪的自然科学发展,我们要特别强调培养“广谱”科学人才的必要性. 大学中的学科设置和研究机构的分工,总是落后于科学的发展. 时代需要培养一代新人,他们具有良好的数、理、化基础和丰富的生物学知识,熟悉计算机和网络技术. 他们不把自己局限在某一学科领域,而是热心关注和解决自然界中人类尚未明白的问题. 这里的目标越出了传统意义下的生物物理或理论生物物理,而指向广义的理论生命科学. 在目前的教育体制下,从哪一门学科开始培养并不重要. 关键是在思想上要兼容而不排他,还要有从事大科学的团队精神. 我国学术界要特别提倡大家都努力做好自己的事情,同时尊重别人的工作. 专业上相差较远的学者要努力吸收其他领域中可能有益的方法和概念,而不轻易地宣称对自己“没有用处”. 近距离的同行更要互相支持而不彼此拆台. 这样,交叉学科在我国才会更健康地发展.

5.1 思维方法取长补短

为了促进数理科学和生命科学的交叉研究,两大“阵营”的学者们还要注意彼此在思维方法上的差异,取长补短、理解对方. 数理科学已经具有较完

整的基本法则的体系,演绎推理在研究中起着较重要的作用;生命科学基于大量的事实,几乎任何规律都有例外,归纳和例证仍是主要方法.物理学者惯于进行简化,通过粗粒化来突显基本规律,生物学者注重具体事实和细节.物理学尊重因果论,经常从原因出发探索后果;生物学较多从事实和结果出发,回溯已经逝去的进化过程,甚至有时使用“目的论”(teleology)的语言.

其实,物理学在研究复杂现象时,也曾从“目的”出发来表述方程应满足的条件.描述布朗运动的郎之万方程是最早提出的随机微分方程.它有两个参数:摩擦系数和随机力的强度.这两个参数不能任意给定,而必须满足一个含有温度的比例关系,才能保证布朗粒子的速度最终与环境达到热平衡.这是著名的涨落耗散定理的早期实例,也是从“终值”而不是“初值”出发的发人深省的实例.

5.2 粗粒化和排列组合

我们在本章中多次提到粗粒化.其实我们对自然现象的研究,只能分层逐次地进行.“粗粒化”是不可避免的方法,而它往往导致符号和符号序列.符号和符号序列又自然地要求动用离散数学宝库中的法宝.于是语言学、图论和组合学等等接踵而来.这里要特别讲一下排列组合问题.

19世纪孟德尔豌豆杂交实验的重大贡献,就在于证明性状遗传是组合而不是混合.排列组合是利用有限种模块产生无穷多结果的手段,大自然在进化过程中不断利用了组合的威力.DNA中的遗传编码就是字母组合.内含子和外显子的交替剪切,导致多种蛋白质序列的组合.许多不同的蛋白质包含大同小异的若干个结构域的选择排列.图3给出一个教科书中的经典例子.现在这类数据越来越多.

在更高的层次上,类似模块的重新组合屡见不鲜.多个蛋白质亚基组成更大的复合体,其中有些亚基是相同的.前面提过的细菌鞭毛的“轴承”部分,是由大量蛋白质复合体排列成9度对称的形状.真核生物细胞核孔是蛋白质镶嵌的具有8度对称的“八宝琉璃井”.

水稻的籼稻和粳稻两个栽培变种,已经隔离多年.东汉许慎所著《说文解字》一书中就对它们的形状差别作过描述.然而,两个变种中负责光合作用的叶绿体基因组却相差甚少.这说明它们的许多性状差别来自比基因更高的层次.人和家鼠的基因组大小和基因数目都很接近,两者差异只能源于更高层次的排列组合,包括基因在染色体上的排列顺序和

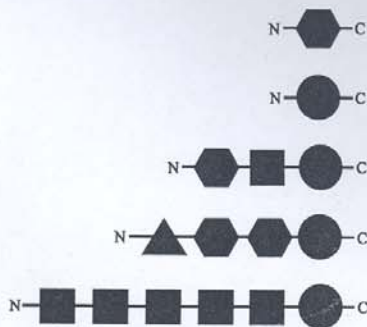


图3 几种丝氨酸蛋白酶的结构域组合
(各种几何形状代表不同的结构域.从上往下分别是表皮生长因子、胰凝乳蛋白酶、尿激酶、凝血因子 IX 和纤维蛋白溶酶原.字母 N 标明蛋白质链的氨基端, C 为羧基端.本图改画自 B. Alberts 等著《细胞的分子生物学》,1994 年第 3 版第 123 页)

它们表达过程中的时空顺序、基因的对交替剪接,以及蛋白质复合体等更高层次上的排列组合.

然而,我们不能在现实世界中期待完整、纯粹、干净的数学结构.相反,人们处处遇到模糊和随机.统计方法已经在生物学中发挥着重要作用,超越纯粹统计的尝试仍然离不开统计的辅佐.统计物理学在研究复杂“死物”的 100 多年里,积累了丰富经验,它对认识生物必定还大有用武之地.

5.3 前辈学者的良好榜样

在物理学和生命科学的交叉方面,我国老一代学者已经为我们树立过良好榜样.我国前辈植物生理学家汤佩松(1903—2001)和统计物理学家王竹溪(1911—1983)都是 1955 年的首届中国科学院院士(学部委员).他们早在抗日战争时期的昆明西南联大就合作建立了水分在植物中运动的热力学理论,论文发表在 1941 年的美国《物理化学杂志》上¹⁾.这一超越当时认识水平的重要成果,被忽略了约 40 年.20 世纪 80 年代才重新在国际上引起注意.我国目前的研究队伍和设备条件,比起 60 多年前的旧中国,真是天壤之别.我们应当而且可以做得更好.希望寄托在青年学子身上.

参 考 文 献

- [1] 郝柏林,张淑誉.漫谈物理学和计算机.北京:科学出版社,1988,1992[Hao B L, Zhang S Y. On Physics and Computers. Beijing: Science Press, 1988, 1992 (in Chinese)]
- [2] Schrödinger E. What is life? The Physical Aspect of the Living Cell. London: Cambridge University Press (many printings since 1944 in various languages)

1) Tang P S, Wang C S. J. Chem. Phys., 1941, 45, 443

- [3] 欧阳钟灿,刘寄星. 从肥皂泡到液晶生物膜. 长沙: 湖南教育出版社,1994 [Ouyang Z C, Liu J X. From Soap Bubbles to Biomembranes. Changsha: Hunan Education Press, 1994 (in Chinese)]
- [4] Murphy M P, O'Neill L A J eds. What is life? The Next Fifty Years. London: Cambridge University Press, 1995,1997
- [5] 郝柏林,刘寄星主编. 理论物理与生命科学. 上海:上海科学技术出版社,1997,1999 [Hao B L, Liu J X eds. Theoretical Physics and Biology. Shanghai: Science Technology Press, 1997,1999 (in Chinese)]
- [6] 郝柏林,张淑誉. 生物信息学手册. 上海:上海科学技术出版社,初版,2000;第2版,2002 [Hao B L, Zhang S Y. Handbook of Bioinformatics. Shanghai: Science Technology Press, 1st ed. 2000;2nd ed. 2002 (in Chinese)]
- [7] 郝柏林. 生物信息学浅谈. 上海:上海科技教育出版社,2002 [Hao B L. A Tale on Bioinformatics. Shanghai: Science Technology Education Press, 2002 (in Chinese)]