

后基因组时代的 思考

吴
家
睿
著

上海科学技术出版社

图书在版编目(C I P)数据

后基因组时代的思考 / 吴家睿著. —上海 : 上海科学技术出版社, 2007.1
ISBN 7-5323-8521-3

.后吴基因组 - 文集
.Q343. 1-53

中国版本图书馆 CIP 数据核字 (2006) 第 062730 号

上海世纪出版股份有限公司 出版、发行
上海科学技术出版社
(上海钦州南路 71 号 邮政编码 200235)
新华书店上海发行所经销
上海 × × × 印刷有限公司印刷
开本 889 × 1194 1/32 印张 7.875
字数 : 140 千字
2007 年 1 月第 1 版
2007 年 1 月第 1 次印刷
印数 : 1 - 2000
定价 : 28.00 元

如发生质量问题, 读者可向工厂联系调换

自序

经过英国格林尼治天文台的子午线在1884年被定为世界本初子午线，以它作为零度经线划出了东西半球。格林尼治天文台特意沿着子午线做了一个标志，让人们有机会在那里同时跨越东西半球。笔者就曾经在那里做过一次跨越动作。零度经线不仅仅是一种空间坐标，更重要的是被用作时间坐标。全世界就是以它为基准制定出了国际通用的格林尼治标准时间。因此，当21世纪降临之际，零度经线是新千年的“官方起点”。在笔者看来，当生命科学从20世纪进入到21世纪之时，也有一条“零度经线”，即“人类基因组计划”。在人类基因组计划之前，那是以分子生物学为核心的经典生物实验科学时代，而人类基因组计划的完成则迎来了以“组学”(omics)为主的“后基因组时代”。

笔者生而逢时，既得到了系统的经典生物实验科学的教育与训练，又参与了后基因组时代的科学研究工作。在这个“跨世纪”的过程中，笔者就像20世纪初的物理学家面对物理学危机那样，深切地感受到了生命科学中众多发现的震撼、各种观点的冲击、诸多问题的困扰，由此引发了笔者一次次的思考。在一次交谈中，上海科学技术出版社的编辑叶剑先生建议笔者把这些思考整理出来。感谢《科学》杂志编辑部的支持，为笔者在《科学》杂志设立了

自序

一个名为“后基因组时代的思考”的专栏。这个栏目为笔者的各种想法提供了一个表达的平台，也由此导致了这本书的诞生。

在整理这四年多发表的专栏文章中，笔者看到了自己认识过程的运动轨迹；这其中有许多不成熟的想法，有许多前后不一致之处，更有许多有待完善之处。笔者在重读这些文字的过程中，进一步体会到了那句名言：“以今日之我否定昨日之我，以明日之我否定今日之我。”为了给自己和读者提供一份真实的心路历程，笔者决定在编辑这本书时，除了文字的校对和体例的修订外，尽可能地保留文章的本来面目，即使文章中的某些论点看上去已经过时甚至不正确。

除了专栏文章外，本书中还选录了若干篇笔者的专栏之外的文章，包括发表在其他杂志上的相关文章。如果这本小册子能引发读者对“后基因组时代”的关注和思考，那么笔者的目的也就达到了。

吴家睿

2006年3月

自序

系统生物学面面观

- 2 世纪之交的生物学基础研究
- 7 生命的复杂性和复杂的生命科学
- 14 系统生物学面面观
- 24 生物系统的分子构建
- 31 后基因组时代的病毒研究
- 35 “活力论”的复活
- 45 合成生命之梦
- 53 走向新的综合：生命科学领域的小科学与大科学之关系
- 63 生物系统的复杂性与简单性
- 70 论 Systems Biology 的译名问题
- 78 横看成岭纵成峰 —— 简论系统生物学的学派

新时代 大科学

- 86 人类基因组学的新战略：人类全基因组的随机测序
- 93 并非咬文嚼字 —— 蛋白质组概念的分析
- 98 新时代 大科学
- 104 蛋白质组研究：无尽的挑战
- 113 两种大科学的抉择
- 120 “生命之书”的解读
- 130 走出“基因决定论”的误区

目录

137	最小基因组与生命起源
144	编撰人类基因组的“百科全书”
153	复杂的 RNA 与复杂的基因组
160	向生物小分子进军

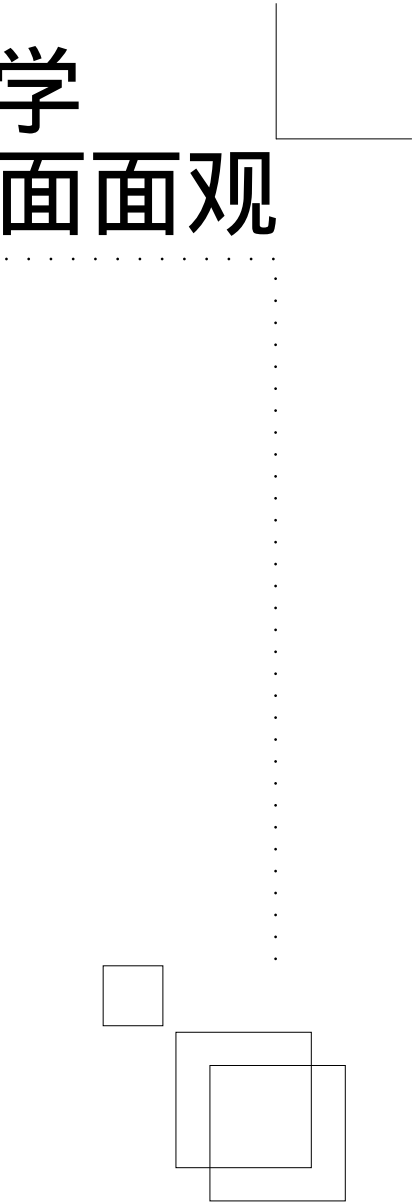
后基因组时代的交叉科学

168	后基因组时代的交叉科学：从 Bio-X 到 X biology
174	抽象的价值 —— 数学与当代生命科学
182	他山之石，可以攻“异” —— 化学遗传学漫谈
189	生命世界的“第三定律”

创造历史的人

198	生命科学研究人才培养战略刍议
208	科学研究的动机
211	细胞周期的调控
218	模型 · 问题 · 方法
225	科学规划三问
232	通向生命科学未来的路线图
242	创造历史的人

系统生物学 面面观

A decorative graphic consisting of a solid L-shaped line at the top right, a vertical dotted line extending downwards from the horizontal part of the L-shape, and three overlapping squares at the bottom right.

世纪之交的生物学基础研究

20 世纪中期，随着蛋白质空间结构的 X 射线解析和 DNA 双螺旋的发现，开始了一个分子生物学时代，对遗传信息载体核酸和生命功能的执行者蛋白质的研究成了生命科学研究的主要内容。经过科学工作者半个世纪的努力，生命科学已成为自然科学中最重要的学科。随着新世纪的降临，有必要对生物学基础研究的现状和发展趋势进行一番探讨。

生命科学的主导力量

分子生物学的诞生使传统生物学研究转变为现代实验科学。今天它仍然是生命科学的主导力量。随着分子生物学的飞速发展，带动了生物学各分支学科的研究向分子水平深入。一方面是在分子水平上对细胞活动、遗传、发育等各种生命现象进行解释，出现了分子发育生物学、分子神经生物学等新兴学科。另一方面是把分子生物学研究手段推广到生物学各分支学科，尤其是一些宏观生物学学科如进化论、分类学、生态学，形成了分子进化生物学、分子系统分类学和分子生态学等。这些学科用实验

的方法研究传统的生物学问题，微观研究和宏观研究得以紧密结合。

生命科学的研究模式

经典生物学研究以小型实验室和“单干”的研究模式为主。而现在，生物学界的研究模式则出现了重要的变化。首先是大规模的集约型研究的出现。随着人类基因组计划等“大科学”的实施，世界各国建立了许多大型研究中心，如英国的桑格（Sanger）中心、美国华盛顿大学基因组测序中心、中国的国家人类基因组南方和北方研究中心。这些研究中心通常像一个大企业，拥有大量的仪器设备，如数十台甚至数百台DNA测序仪，针对一个特定的目标进行研究。

其次，由于研究内容的深入和研究范围的加大，多个实验室间的合作研究方式已成为当前的主要潮流。在研究论文的署名中，共同第一作者、共同通讯作者等已越来越常见。对于基因组研究和蛋白质组研究等“大科学”而言，大规模的跨单位、跨地区、跨国家的联合研究已成为主要方式。甚至在细胞信号转导这类经典的“单干”领域，也出现了由美国、加拿大和英国的52个实验室组成的“信号转导联军”（Alliance for Cellular Signaling），准备对G蛋白介导的和与其相关的细胞信号转导系统进行系统和全面的研究。目前，这些新的研究模式已成为推动生物学快速发展的主动力。

生命科学的思维方式

分子生物学的成功曾经使还原论和线性思维成为生物学家、尤其是实验生物学家的主要认识工具：生命的活动遵循着物理、化学的基本规律；一个基因决定一种蛋白质，从而决定一种生物学功能。然而，现在的生物学家对生命的思考和认识却有了新的角度。由于基因组研究和蛋白质组研究等“整体性研究”方法的出现，研究的视野已经从一两个基因或蛋白质的行为扩展到了成千上万个基因或蛋白质的表现。此外，由于复杂系统理论和非线性科学的发展，关注的对象已不再停留于一条代谢途径或信号转导通路，而是提升到了细胞活动的网络和生物大分子之间复杂的相互作用关系这个层次。因此，生物学思想和方法论正在从局部观向整体观拓展，从线性思维走向复杂性思维，从注重分析转变为分析与综合相结合。

生命科学的研究技术

20世纪中叶，通过X射线衍射仪的应用，解析了DNA的双螺旋结构和蛋白质的空间结构，从而推动了分子生物学的产生。随后的生物学研究对技术和设备有了迫切的需求。每一次技术的进步都有可能带来一个生物学研究的突破。如果没有DNA序列分析仪、质谱仪、超级计算机，基因组研究和蛋白质组研究只能是纸上谈兵。这些技术/仪器可以分为两大类。一类是以生物学原理为主设

计的,例如聚合酶链反应(PCR)技术、基因芯片和蛋白质芯片等。另一类则是基于物理、化学原理的仪器,如电子显微镜、X射线衍射仪、质谱仪等。当前生物学的发展越来越依赖于高新技术。可以说,在“后基因组时代”,谁拥有先进技术,谁就是生物学研究的“领头羊”。

生命科学的交叉研究

现代生命科学的发展不仅离不开仪器设备,而且也不离不开其他非生物学学科的参与。在上个世纪初,正是由于一批物理学家、化学家进入生物学研究领域,才导致了分子生物学的诞生。现代生物实验科学经过了50多年的自我发展,研究越来越深入,问题越来越复杂,现有的生物学理论和方法暴露出了越来越大的局限,又到了一个需要多学科交叉的新阶段。例如,随着基因和蛋白质数据近年来爆炸式的增加,产生了一门新兴学科——生物信息学,来对这些数据进行处理和研究。又比如,对细胞和神经等复杂系统和网络的研究导致了数学生物学的诞生。当前数、理、化和生物学领域的科学家的交流和合作也许孕育着又一次学科的“大爆炸”。多学科交叉将推动生物学自身以及自然科学其他学科的发展。

生命科学的投入产出

传统的生物学基础研究,如孟德尔或摩尔根的遗传学研究,既不需要大的经费支持,也没有明显的经济效

益。可是现代生命科学研究却按照全然不同的方式进行。首先，对生物学研究的投入越来越大。例如在美国 1995 年自然科学的总研究经费中，65% 以上被用于生命科学研究。近年来美国国立卫生研究院(NIH)保持着每年 15% 的经费增长，到 2003 年时预计将达到 200 亿美金的年度拨款。其次，生命科学对社会的产出也在迅速增加。在分子生物学研究初期，大部分的研究工作都是在“象牙塔”里进行；到了遗传工程时代，基础研究开始了与应用研究和产业开发的结合。而在“后基因组时代”，许多在过去被视为基础研究的工作一开始就与应用紧密联系在一起；例如一些商业公司在对基因组的 DNA 序列进行测定或对蛋白质组中的蛋白质进行分析的同时，把所得到的研究成果直接变为具有很高利润的商业数据库。基因或蛋白质的信息已成为各大制药公司追求的重要目标。今后，生物学基础研究与应用研究的结合会更紧密，研究成果向产业化转化的速度会更快。

以上六个方面虽然不能说覆盖了当前生物学基础研究的所有特征和动向，但可以认为反映了它的基本现状和主要趋势。也许还可以把这些特点再概括为“一个潮流”和“一种态度”：顺应后基因组时代的新潮，以积极求变的心态从事研究。

(原载于《生命的化学》2001 年第 3 期)

生命的复杂性和复杂的生命科学

复杂性 (complexity) 是最近几年学术界比较流行的术语 , 但什么是复杂性却没有几个人能够说清楚。郝柏林先生曾统计过 , 对复杂性的定义已不下 30 种之多 , 但仍很不完备。换个角度来说 , 给一个概念定义得过于严格也常常会限制人们的思考 , 尤其像复杂性这样含义丰富的术语。正如美国著名物理学家费恩曼 (R. P. Feynman) 所指出的 , “ 要给一个概念下一个非常严格的定义往往是不值得的 ”。因此 , 笔者在文中不准备讨论复杂性的内涵 , 而试图对生命的复杂性进行描述 , 并分析这种复杂性对生命科学研究的影响。

生命的复杂性

2002 年 5 月 , 德国一个研究小组在英国《自然》周刊上宣布 , 他们在冰岛北部海下 120 米深处发现了一种迄今所知最小的生命。这种被称为“纳古菌”(*Nanoarchaeum equitans*) 的微生物属于古菌的一种 , 其基因组的 DNA 碱基对仅有 50 万左右。假设一个基因的平均长度为 1 000 个碱基 , 那么这个细菌所拥有的基因大约是 500 个。如果每

个基因编码一种蛋白质,那么这个细菌最多可以拥有500种蛋白质。这篇文章告诉我们,哪怕是最小、最简单的生物体,也是由许多执行不同功能的组分构成的。因此,生命复杂性的第一个特征是,生命是一种复合体,不可能由一个成分(一种基因或蛋白质)构成。

作为复合体的生命,其组成成分并非简单的堆积,而是彼此间有着广泛的相互作用。芽殖酵母的基因组拥有6 000种左右的蛋白质,但到目前为止人们已经知道的酵母蛋白质与蛋白质相互作用联系种类,就已接近1万。据保守的估计,在酵母细胞里至少存在着3万种蛋白质与蛋白质相互作用。随着各种物种基因组全序列分析的完成,人们看到,生命的复杂程度并不与基因数目成正相关。如水稻基因组可能编码蛋白质的“开放阅读框”(open reading frame, ORF)大约是4.6万到5.5万,而人的ORF数目则低于4万。显然,在生命从低级到高级的进化过程中,有可能是蛋白质之间的相互作用在起主导作用,越是高级的生命,其相互作用越广泛。由此可以概括出生命复杂性的第二个特征——组分之间有着广泛的相互作用。换言之,生命的本质是由组成元素之间的关系所决定,而非组成的物质本身。

生命组成成分的相互作用很早就被人们所认识,但是最初的观点,是把这些相互作用看成一些线性的联系,即甲因子作用于乙因子,乙因子再作用于丙因子,依次递推。比如过去的教科书里,生化代谢途径或者信号转导过

程常常被描述成直线式的作用图。但近年来,科学家已承认,生命内的这些相互作用不是直线式的,而是交错编织形成的网络。这种广泛存在的相互作用网络引出了生命复杂性的第三个特征:次序和层次。由于生命中各组成成分有着稳定的相互作用,从而形成了有序的结构,也就是人们常提到的“自组织”。生命自诞生那天起,就是一个与外部环境相对独立的系统,并且通过与外界交流物质和能量维持其有序性。随着生命的逐渐演化,次序发展出了层次:各种生物大分子相互作用并形成了不同的功能区域(细胞器等),这些功能区域组合成细胞,各种执行不同功能的细胞又汇聚成组织,组织与组织的结合又产生器官,最终形成了多细胞的生物体。

生物体的每一个层次都拥有特定的行为或性质。这类行为或性质不存在于构成它的组成成分里,而是由组成成分间的相互作用所形成。因此,生命复杂性的第四个特征是,整体比它的部分之和更大。研究复杂性系统的科学家把这种现象称为“涌现”(emergence)。生命系统的涌现是属于非线性的,即不能通过简单地叠加构成成分的行为推导出系统的行为。此外,系统组成部分的微小改变常常会被迅速放大并导致系统状态的改变,即所谓的“蝴蝶效应”:地球一端的一只蝴蝶扇一扇翅膀,可能改变地球另一端的气候。2001年9月,美国《科学》周刊发表的一个研究报告很能说明这一点。澳大利亚科学家通过测定1918年因“西班牙流感”而死的病人组织中流感病

毒的基因序列，发现其强烈的感染性可能是由于人流感病毒编码血凝素的基因中的一小段序列，与猪流感病毒的血凝素基因序列重组所导致的。也就是说，有史料记载的最大一次瘟疫（死亡人数超过第一次世界大战的死亡人数）——“西班牙流感”，其起因仅仅是流感病毒内一个基因的一段序列的改变。

这种“整体大于部分之和”和“涌现”性质，是生命诞生及其进化的主要动因，生命通过改变自身以适应变化着的外部环境。这种特性使得地球上形成如此繁多的生物物种，使得人类这样高级的生命形态能够从原始的细菌进化而来。在这个意义上，美国科学家霍兰（J. H. Holland）提出，“适应性造就复杂性”。因此，生命复杂性的第五个特征是，系统具有开放性，可以在过程中不断地演化。生命不是一种简单的“自稳态”系统——通过负反馈的调节控制来稳定自身的状态，从而适应外界的变化；而是一种远离平衡状态的开放系统——通过不断地形成新性质或新功能来适应外界的挑战或改变。

复杂的生命科学

20 世纪的生命科学研究主流是分子生物学。这是一种“简单的”生命科学，其研究基础建立在还原论和线性思维上。还原论的基本前提是，在由不同层次组成的系统内，高层次的行为是由低层次的行为所决定的。具有还原论观点的生物学家通常认为，只要认识了构成生命的分

子基础,如基因和蛋白质,就可以理解细胞或个体的活动规律。对于他们来说,生命是一种线性活动,即生命现象是由各个构成成分的孤立活动来决定的,组分之间的相互作用常常被忽略不计。

尽管分子生物学在近半个世纪取得了显著的成绩,但是并没有完全揭示出生命的奥秘。究其原因就在于以上所讨论的生命之复杂性,这是“简单的”生命科学研究方法难以处理的。今天,科学家正在试图发展出能够研究“网络”、“涌现”等复杂性行为的“复杂”的生命科学,如功能基因组学、蛋白质组学、数学生物学和系统生物学等。

复杂的生命科学的另一层含义是,生命科学理论的形成和发展并不遵循物理学或化学等“正统”科学理论的运行模式。

著名科学哲学家波普尔(K. R. Popper)曾提出,区别科学与非科学的标准是“可证伪性”:只有可以被反驳、被证伪的理论才是科学理论。科学的历史就是一系列的猜想、反驳、再猜想、再反驳。可证伪性的关键是,其命题必须是一个全称命题。例如,陈述“所有天鹅都是白的”具有可证伪性,因为发现一个黑天鹅就可以否定这个陈述。而“一些天鹅是白的”则无法证伪。

生命科学的发展历史表明,生命科学理论充满了例外。DNA双螺旋发现以后,人们一度认为,遗传信息是从DNA传递到RNA,再到蛋白质的。但随后的研究表明,遗传信息也可以从RNA传递到DNA。起初大家公认具有催