

· 专题五 ·

生命科学中的跨学科前沿

与会专家建议优先发展领域如下:

1 遗传语言研究在基因组结构和序列研究基础上,通过数学、信息科学、非线性物理和语言学的交叉渗透,揭示基因组 DNA 上控制发育和生命活动的遗传语言与程序。包括:基因信息与语言学;基因组复杂性分析;基因组功能的实验研究。

2 分子进化工程研究 在进化论思想指导下,应用计算分子生物学方法和基因工程手段,进行分子定向进化研究。包括:生物大分子和大分子复合体系结构和功能;分子识别的机制及相关技术的研究;定向分子进化研究。

3 发挥我国传统医药优势加强创新药物基础研究 结合中医药学基本理论和实践经验通过生命科学、化学和其他学科的交叉渗透和共同参与,采用现代的科学技术进行中药复方的化学成分、生理活性、组方原则和疗效机理的基础研究。

4 生物传感技术与仿生材料 包括:仿生功能薄膜的研究;基于微系统与分子系统的生物传感技术的研究;用于神经生物学、细胞生物学、发育生物学、医学与生物技术等领域的在体生物传感技术的研究。

5 计算神经科学 通过与数学、物理学和信息科学等学科的交叉渗透,开展脑功能的理论研究。包括:感知机制(视、听等)的理论研究;学习记忆及神经信息的存储、编码机理的研究;某些脑的高级功能(整合、注意等)的理论研究;感觉-运动系统的调控原理。

6 生物深层次信息的捕捉、提取与分析 依靠学科交叉的新概念,建立新的手段和方法,对生命体系在不同结构层次的大量与丰富的信息,进行系统、综合、关键和动态的捕捉、提取和分析。包括:建立空域、时域和频域相结合的活体的无创的和动态的分析手段和方法;基于知识集和信息集建立与生理相一致的同胚模型;建立正常与异常的重要生命器官的功能相关结构、信息载体和调制物质的活体研究方法。

基因组结构和遗传语言

王亚辉

(中国科学院细胞生物学研究所,上海 200031)

美国提出的“人类基因组作图和测序”的庞大计划,预计在本世纪末将完成其第一步目标,读出人基因组 DNA (3×10^9 碱基对)的全序列。下一步更为艰巨的任务是读懂基因组的工作语言——遗传语言的破译。

人和动物基因组结构是在长期进化过程中形成的,其功能则需要在发育过程中才能表现出来。基因组 DNA 荷载的遗传信息的特点是信息结构和物质结构的紧密联系。活细胞内, DNA 信息的表达是受位于染色质结构上的一系列核酸和蛋白质相互作用控制的。因此,对基

基因组结构和遗传语言的研究,必须结合发育和进化的研究进行。通过对人和其他动物(线虫、果蝇、小鼠等)基因组的总体结构(基因在染色体上的排序及各组间关系)和 DNA 序列的精细比较,将可能发现许多新的调节基因和在进化上保守的有重要功能的区域;再借助转基因动物,还可进一步检测这些序列的功能。此外,更为繁重的任务是需要与数学家、逻辑学家、非线性物理学家和计算机科学家合作,在对浩如烟海的序列资料进行比较分析和综合的基础上,去破译记载在基因组 DNA 上的“遗传语言”,从而阐明控制发育的遗传程序在染色体上的构建和操作规则,以及在进化过程中发生的变化。最终的目标是找到某种简明的数学形式来表述:记载在基因组 DNA 一维结构上的遗传信息如何控制生物体的三维形态发育;以及复杂性不断增加的动态发展过程。这无疑是对数学(系统理论、大组合复杂性和拓扑学等)和非线性物理学的一大挑战。

根据研究方法和侧重点的不同,遗传语言研究又可分为下述几个方面:

1 **基因信息学** 适应基因组计划和蛋白质工程的需要,建立对基因序列数据的储存和处理的计算机基因信息管理系统。如:(1)进行序列联配(Sequence alignment)(整体、局部、多重和柔性联配);(2)基因分析,包括功能位点(剪接、蛋白质结合序列、启动子、加强子等)、调控区的搜寻和预测、以及蛋白质折叠和三维结构的预测等;(3)基因组结构分析,包括大尺度联配、发育程序、进化“遗迹”等方面研究。

2 **基因组信息结构的复杂性** 利用非线性物理和数学等多学科方法,把基因组全序列作为一个整体系统,研究其信息结构的特性(数据信息与调控信息、序列频数、分布图式、长程关联、分形维数和自相似性等);与自复制、自修复、自调控等功能相关的操作方式(程序、中心指令集等),揭示基因组信息结构的复杂性及其构建和操作规则。

3 **语言学研究(Linguistic approach)** 把基因组 DNA 序列作为一种由“ATCG”四字母构成的书面语言,用语言学方法,对 DNA 序列进行形式的比较分析和语义学研究,以发现其“词法”和“句法”规则。目前以色列已编纂了第一部遗传语言词典(《Gnomic——A Dictionary of Genetic Code》),收集了 800 多个有生物学意义的“单词”。

4 **实验研究** 利用人工染色体(MAC)作载体,将重组的“词”和“句”输入到转基因动物受精卵内,检测其“语义学”功能。这是对上述理论研究结果进行验证的必要的办法。

人类基因组结构和遗传语言的研究将是下一世纪自然科学的前沿和竞争的焦点,对未来生物学、医学和生物技术的发展有根本的意义。目前国内随着“水稻基因组计划”的实施及参与人类基因组的研究,已经在计算机基因信息管理系统和基因信息学方面开始工作;理论研究方面也有人向国家自然科学基金会提出有关基因组信息结构的复杂性和动力学研究的建议;实验研究方面,中国科学院细胞所,早已开始动物染色体人工合成和个体表达系统研究,并设想将此系统用于遗传语言的研究。目前国际上遗传语言研究刚起步,应不失时机地及早组织多学科力量,迎头赶上,争占一席之地。

GENOME STRUCTURE AND GENETIC LANGUAGE

Wang Yahui

(Institute of Cell Biology, Academia Sinica, Shanghai 200031, China)