

基因组工程研究

黄芳 郭礼和

(中国科学院生物化学与细胞生物学研究所 上海 200031)

一、基因组工程的诞生

1. 基因研究与基因工程的发展

随着生命科学和生物技术的发展,对基因的释义也在不断地发展和深化。最初将基因与遗传学性状相挂钩,提出所谓一个基因一个性状的概念,以后发展到一个基因一个酶、一个基因一个多肽、一个基因一个转录本等,今天看来这些概念都不能涵盖基因所有的内涵,因为基因还应该包括与转录相关的所有遗传结构信息。在结构上可以连续,也可不连续;在顺式功能上有的转录,有的不转录;在转录本上有的能翻译,有的不需要翻译;在翻译单元结构上根据细胞生理状况或其他原因,可以产生一种或多种蛋白或多肽,也可以通过翻译后加工产生多种蛋白或多肽;在翻译产物的功能上,它与细胞种类和生理状态紧密相关,同一种蛋白在不同情况下具有不同的生理功能,也就是说基因释义应是多义的。由于基因在形态结构上以及生理功能上内涵相当丰富,用简单的语言很难作出准确的定义,从信息学的角度,把基因概括成与转录相关的顺式遗传指令可能比较恰当。

1972年,斯坦福大学 Berg 实验室成功地完成世界上第一个体外基因重组实验,启动了重组 DNA 技术研究发展,直接导致了一种全新的生物技术学科和产业—基因工程。基因工程的诞生使得外源基因在细菌、酵母和动植物细胞中能进行表达,从而打破了物种的界限,在实验室内可以用工程原理和技术对生物直接进行遗传改造,以达到为生产实践直接所用之目的。1982年重组人胰岛素开始产业化,随后重组人生长激素、重组人凝血因子 VIII、重组疫苗等相继上市,基因工程对人类社会和经济发展起到了巨大的推动作用。

尽管基因研究进入基因组研究阶段,但目前还是停留在寻找新基因及其结构与功能的研究方面,也就是解析式的细化研究。细胞的生命活动不是一

个一个基因表达的简单组合,而是生理过程所需要整合的一群基因(少则几十个,多则成千上万)的协同活动,它们相互协调、制约和促进。同一基因在不同生理活动中所用的顺式遗传指令和功能会有很大差别。以一类生理活动为基础的一群相关基因的整合式研究目前还很少涉及。基因工程研究还处在一个或几个基因改造、利用的简单遗传操作阶段,生命活动复杂体系需要基因群体共同参与,有条不紊地完成生命有机体复杂的生化反应和生理活动,目前的基因工程还远远不能满足这种需求。

2. 基因组计划和生物信息学的发展

基因组是指细胞的染色体总和,1990年启动的人类基因组计划,旨在揭示人类所有的遗传结构,包括所有的基因(尤其是疾病相关基因)和基因外序列的结构。人类单倍体基因组序列含约 3×10^9 碱基对,分布在 23 条染色体上,2001年2月公布的序列研究数据已覆盖了 95% 以上的基因组结构,准确率高达 99.96%。现在公布的数据认为基因个数在 2.65 万至 4 万之间,约由 1% 的基因组结构来编码。人类基因组计划任务有望在 2003 年完成。尽管人类不同个体之间有不到 0.1% 遗传结构的差异(一百万个碱基中有 800 个碱基不同),但不同物种基因组间的差别是比较明显的,这种差别可作为进化的标志。人类基因组研究的意义决不仅仅如此,它还将极大地促进包括生物信息学(Bioinformatics)、功能基因组学(Functional Genomics)、蛋白质组学(Proteomics)、药物基因组学(Pharmacogenomics)和其他许多相关学科的发展^[1]。

生物信息(Biological Information)是以基因作为单元,研究个体中所有基因以及个体的全部信息^[2]。面对巨大的、日益增长的生物信息,计算机的加入成为一种必然。生物信息与计算机科学相结合从而诞生了生物信息学(Bioinformatics)。其内容集中在算法、模拟、数据库构建与分析、大范围基因序列的获取、SNP 分析、蛋白质结构预测、药物分子研究等,

因而除对生物信息进行管理之外,还涉及信息的整合、比较、推理、发现等。毫不夸张地说,这种建立在硅片上的研究新手段重塑了现代生物科学的新面貌。

随着人类及其他模式生物基因组计划的实施以及相关生物技术的发展,以基因组为基础对物种进行大范围修饰和改造即将成为一种可能。在这种形势下基因组工程(Genomic Engineering)呼之欲出,以便可同时对大量的基因群体进行操作,模拟某一生理过程,产生新的生命活力,改造物种,实现细胞和生物体能做什么,人类在实验室或工厂也能做什么的梦想。

二、基因工程与基因组工程的对比

1. 操作的基因和载体

基因工程和基因组工程,都是对基因进行工程性的遗传操作,但两者有着质的不同。基因工程常用的载体是质粒和病毒载体,克隆基因的容量有限,通常包含的基因只有一个或几个,而且长度较短,称为 Kb 级(指克隆 DNA 片段长度)工程性遗传操作。在基因组工程研究中,采用的载体是人工染色体,容纳的基因可达几十到几百个,甚至几千个,是基因群体克隆和表达,而且这种表达是按生理活动过程予以遗传控制。克隆 DNA 片段长度可达几千 Kb,也称为 Mb 级工程性遗传操作。

2. 克隆和扩增的宿主

重组后的基因或基因群体需要在适当的宿主内寄生,一般以下列中的一种方式存在:1)于宿主染色体之外,独立存在;2)整合到宿主的染色体内。用于基因工程研究的克隆和扩增的宿主有细菌(如大肠杆菌、枯草杆菌等)、真菌(如酵母等)、动植物细胞(如昆虫细胞、哺乳类细胞等),但大多是以大肠杆菌为主,而用于基因组工程研究的克隆和扩增的宿主也可以是细菌、真菌、动植物细胞,但主要以酵母和哺乳类培养细胞为主。

3. 工程性遗传操作手段

重组 DNA 技术是基因工程研究的基础,它依赖于质粒或病毒载体、限制性内切酶、DNA 连接酶和大肠杆菌宿主等。基因导入宿主的手段通常有转化、转导、转染和显微注射等。基因组工程在重组 DNA 技术应用基础上,发展人工染色体作为载体,建立遗传同源重组技术作为 Mb 级的 DNA 大片段切割和整合(包括改造和修饰)的基础,完善酵母和培养细胞的转化和融合技术,开拓干细胞和体细胞

克隆个体的策略。

4. 产物检测

基因或基因群体导入宿主后,对宿主需作基因和/或表达产物检测和分析。在基因工程研究中,利用限制性内切酶图谱、Southern 印迹、Northern 印迹、PCR、序列分析等手段分析基因表达,对其表达产物可以用蛋白分析、酶学分析、抗体或底物结合等手段来研究。在基因组工程研究中,操作的基因数量大,产物非常复杂,除掉用常规手段(如 PCR、序列分析等)来检测基因及表达之外,还需用高通量的研究手段如生物芯片(包括核酸芯片、蛋白/酶芯片、抗体芯片、底物芯片、配体芯片等)来检测基因群体、表达图谱及产物群。生物芯片的研制开始于 90 年代初,它是基于 Southern 印迹、Northern 印迹、Western 印迹、底物结合、配体/受体结合等杂交技术基础而发展起来的集多种学科于一身的新技术。生物样品群结合在片基上,样品群以微阵列(Microarray)形式分布,片基主要有玻璃片、半导体硅片、硝酸纤维膜、尼龙膜等。生物分子以阵列式固定在片基表面,它们可以是 DNA、RNA、蛋白质或其他分子(包括小分子)。目前有 95% 以上的微阵列研究工作是采用 DNA 分子,主要为 cDNA 分子和呈上升趋势的寡核苷酸片段。高通量性主要体现在 1 厘米的硅片或玻片上可以固定成千上万个探针,一次实验可以获得大量的信息,给产物群检测带来了极大的方便,成为基因组工程研究的重要手段。

基因工程和基因组工程研究在内容、形式、技术、方法学、信息分析等方面有显著的差别。这两种技术和方法的差异列在表 1 中。

表 1 基因工程与基因组工程

	基因工程 Gene Engineering	基因组工程 Genomic Engineering
目标基因	单个或几个	可达几十到几百、几千个
DNA 长度	Kb 级	Mb 级
操作手段	限制性内切酶、连接酶等	同源重组
片段检测	物理图谱、序列分析等	DNA Contigs、DNA 芯片等
克隆载体	质粒、噬菌体、病毒载体等	人工染色体
导入方式	转导、转染、注射	细胞融合、注射
克隆宿主	大肠杆菌	大肠杆菌、酵母、哺乳类细胞
表达宿主	各种细胞或动植物个体	各种细胞或动植物个体
产物形式	蛋白质	次生代谢产物、全新个体
检测方式	蛋白质检测或酶作用产物分析	代谢分析、生物芯片
信息水平	单个或几个信息	系统和网络信息

从上面的论述可以看出,基因组工程的原理、技术和方法是在基因工程、基因组计划以及生物信息学的研究基础上发展起来的,基因工程与基因组工程之间的研究既有差别又有相互重叠和交叉。应该说重组 DNA 技术和基因工程是基因组工程的基础和出发点;基因组工程是基因工程的延伸及归宿。

三、基因组工程技术

1. 基因组工程载体

由于基因组工程操作的是几十万到几百万对碱基的 DNA 大片段,因此只有大容量的载体才能满足这一需要。目前用于基因组工程的载体无一例外地都是人工染色体,包括噬菌体源人工染色体(P1 Phage-Derived Artificial Chromosome, PAC)、细菌人工染色体(Bacterial Artificial Chromosome, BAC)、酵母人工染色体(Yeast Artificial Chromosome, YAC)、哺乳类人工染色体(Mammalian Artificial Chromosome, MAC)、人源人工染色体(Human Artificial Chromosome, HAC)等。这些人工染色体作为载体各有利弊。

PAC 是基于 P1 噬菌体的一种克隆载体,由于 P1 噬菌体头部较大,可以包装约 110Kb 的 DNA,因此载体的容量在 100Kb 左右。PAC 的复制起始点通常只能复制这一大小的 DNA 分子,这样,每个细胞只能保持一份拷贝,所以重组后的 PAC 比较稳定^[3]。BAC 系统基于大肠杆菌的 F 因子。由于 F 因子的复制受到严格控制,每个细胞只能有一到两个拷贝,这明显地降低了 DNA 插入片段的重组,而且它的容量可达 1 千 Kb^[4]。重组后的 BAC 和 PAC 由于分子过大,不能利用通常的转化或转导方式,需要依赖电穿孔导入宿主细菌,这会导致高达 5% 的 DNA 插入片段发生改变,给使用这两种人工染色体进行克隆和扩增带来诸多不便。

YAC 的克隆片段容量在 100Kb 到 1-2Mb,而且酵母易与培养的哺乳类细胞发生融合。来源于酵母的染色体可以整合到哺乳类细胞的染色体上,并能稳定存在,这就使得 YAC 在哺乳类基因、基因簇和染色体结构元件的研究上大有作为。此类研究可在培养细胞或转基因小鼠体内进行。但 YAC 克隆也存在一些缺陷,如容易断裂、重组、丢失等,因而不稳定。另外酵母与哺乳类细胞的融合对细胞仍有一定的选择性^[5,6]。

MAC 可以容纳大染色体片段,并且在细胞或体内能稳定存在。目前还未见报道插入片段大于 1 千 Kb

的 MAC 研究工作,可能是这项研究刚开始的缘故。

HAC 是在细胞内构建的带有人染色体的着丝粒、端粒的微型染色体,其克隆片段容量可达 10Mb,是其他人工染色体所不可比拟的^[7]。其缺点是操作 HAC 非常繁琐。

2. 基于遗传同源重组的人工染色体拼接技术

上述人工染色体的克隆片段容量比起基因工程研究常用的克隆载体(如质粒、病毒载体等),已经相当大了,但有些真核基因的长度要超过几百 Kb,甚至几千 Kb,如此巨大的基因通常一个人工染色体不能包容的。建立基因文库时,需要对基因组 DNA 采用部分酶切,这会造成巨大的基因或基因簇克隆到若干个人工染色体上。体外操作 DNA 大片段容易断裂,用常规的酶切、连接的方法来获得完整的巨大基因或基因簇不现实。为了获取完整的巨大基因或基因簇,只能采用遗传同源重组技术将相互重叠的人工染色体在细胞内拼接成完整的基因或基因簇。噬菌体或大肠杆菌只在基因组特异位点上发生遗传重组,不能进行基因组随意的同源重组。遗传同源重组主要发生在细胞减数分裂期,故而只能发生在酵母和高等真核细胞内。在酵母中,YAC 克隆之间在减数分裂时进行遗传同源重组需要满足以下一些条件:

- 1) 它们位于配型相反的一种酵母体内(a, α 型);
- 2) 彼此之间存在一定程度的同源序列的重叠;
- 3) 有相同的极性,即载体一插入片段方向一致。

3. 两种常用的重组序列与重组酶系统

Cre/loxP 和 FLP/FRT 是两种常用的重组系统^[8,9,10,11,12]。前者来自 P1 噬菌体,后者来自于酵母,同属于重组酶系列的整合酶超家族。loxP 和 FRT 有相似的 34bp 序列结构,两端是 13bp 的反向重复序列,之间有一个 8bp 的不对称核心序列,决定 loxP 或 FRT 方向(见图 1)。Cre(Causes Recombination)重组酶基因和 loxP(locus of crossing over(x) in P1)序列位于 P1 噬菌体线性基因组的两个末端。FLP 重组酶由酵母 2 μ 质粒上 Flp 基因编码,FRT (Flp Recognition Target)序列是 Flp 识别位点。重组过程中两者都产生 Holliday 中间体^[8,13]。

除上述两种重组酶外,酵母的重组酶 R^[14]也已经在酵母和植物中被利用。有理由相信在比酵母更高形式的生物体中应有其他重组系统存在,若能发现,可以扩大同源重组技术研究和基因操作。

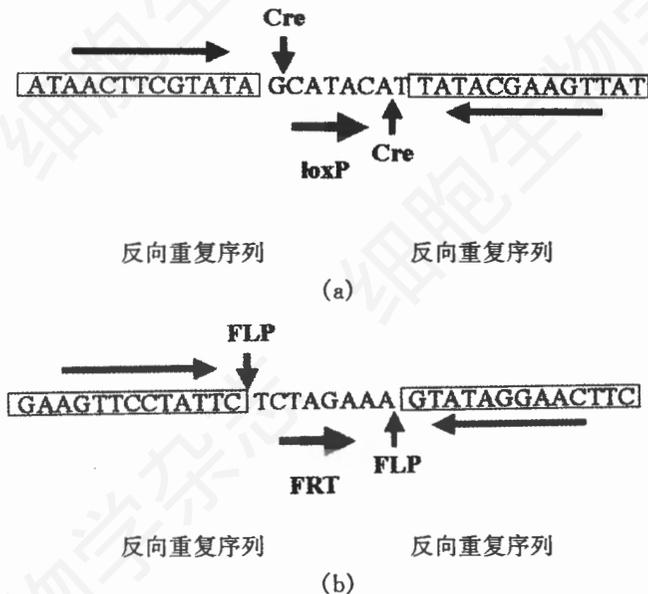


图1 (a) loxP 序列及 Cre 酶识别、切割 loxP 序列示意图;
(b) FRT 序列及 FLP 酶识别、切割 FRT 序列示意图。

对于存在两个 loxP (FRT) 位点的染色体或 DNA 分子, Cre (FLP) 重组酶有如下功能: 1) 删除/整合, 一个线形 DNA 分子或染色体上有两个同向的 loxP (FRT) 位点时, 经 Cre (FLP) 酶作用, 位点之间的片段被环化, 游离于线形 DNA 分子或染色体之外, 此时线形分子或染色体和环状分子上各留有一个 loxP (FRT) 位点, 其逆向过程就是一种整合的过程; 2) 倒位, 一个线形 DNA 分子或染色体上有两个反向的 loxP (FRT) 位点时, Cre (FLP) 酶作用后, 发生位点之间片段的倒位; 3) 转位, 两个线形 DNA 分子或染色体上各有一个或两个同向的 loxP (FRT) 位点时, Cre (FLP) 酶可以导致分子间片段的转位。我们实验室正在探索利用同源重组和 Cre/loxP 系统的删除功能来原位套取染色体大片段。目前同源重组实验都是利用胚胎干细胞或成纤维细胞来实现, 其他细胞是否可行, 也可值得一试。

4. 带有外源染色体大片段的新个体的构建

带有染色体大片段的人工染色体通过两条途径构建新的个体: 1) 人工染色体先导入受体细胞, 再经细胞注射或细胞核移植的胚胎工程操作, 可以构建新的个体; 2) 成熟的卵细胞注射: 直接注射受精卵, 或者以死精子为载体注射卵细胞, 然后通过胚胎发育产生新的个体^[15]。用胚胎干细胞作受体, 然后注射囊胚, 获得的新生个体其体内只有一部分细胞带有外源染色体大片段, 这种个体称为嵌合体。只有

转基因的干细胞进入新生个体的生殖腺系统才能传代遗传。经体细胞核移植或直接注射成熟的卵而产生的新生个体, 外源染色体大片段可以在新生个体所有细胞内存在, 因而可传代遗传。

四、基因组工程学的研究展望

1. 遗传语文——遗传、发育、分化的操作指令和程序

人类的基因组序列无疑像一本有 30 亿个字符的天书, 它是一堆毫无章法的核苷酸残基堆集, 还是一篇有序的遗传语文? 答案当然是后者。人类基因包括已知的和推测的约有四万个左右, 估计许多基因有 2-3 种剪接形式, 所以基因的总数在十万上下 (不包括基因重排、组合等产生的基因数量放大)。编码基因的序列仅占基因组序列的 1%, 99% 非编码序列的功能还不太清楚。细胞及个体有着极其复杂的生理过程 (包括发生、成长、衰老、死亡等), 还有与外界的物质、能量和信息的交换, 所有这些都依赖于各个基因的协同激活与关闭, 即依赖各个基因时空表达的程序。基因通常不会单独作业, 而是以基因群的形式彼此协调、制约, 相互作用共同完成细胞的生理活动。这一过程是如何实现的? 这正是遗传语文所要探究的内容。基因可以看作是一个个的单词, 单词需要按一定的语法结构和规则组合成句子、段落、章节和文章。这篇文章可能是一生命个体演化的剧本。不同物种的生命演化有着不同的剧本, 但它们之间所用的单词基本相同。例如人、鼠、鸡, 这三者在生命活动表现形式上是风、马、牛不相及, 但它们所用的基因 (单词) 在数量和结构上基本相同, 基因组大小也基本相同。用形象语言来说, 人、鼠、鸡三种物种用的是同一本词典, 但写出的剧本、剧情却完全不同。从遗传学角度来看, 这三种生命形式尽管结构基因差别不大, 但占 95% 以上的基因组非编码区却千差万别, 同源性不到 60%。另外在基因群体组织 (Organization) 上也有很大差异, 在染色体数量和结构方面也显著不同。从上所述可以看出仅仅研究基因组结构和各个基因的功能是远远不够的, 要想揭开生命本质的奥秘, 必须从基因群体表达的时空调控程序来研究, 也就是从遗传语文角度来探索基因群体时空表达的语文程序。这需要高通量的基因群体和信息网络实验处理系统和技术平台, 基因组工程学研究正是上述方法学的基础。

2. 细胞生命活动的网络 (整合生物学)

生命科学的发展从宏观观察走向实验, 从整体研究走向微观分析。目前发展趋势从分子水平走向

细胞水平,以致走向整体、综合水平,整合生物学或称整合生理学(Integrative Biology or Integrative Physiology)已经呼之欲出。这种整合应是生命活动的网络整合:基因调控网络、代谢网络、信号传递网络、物质和能量转运网络、神经网络、免疫系统网络...等等。这些网络的整合反映出“网中有网,生命之窗”的概念,透过这个窗口可以窥测到生命的本质和奥秘。这种整合研究与掌握大量的高通量技术平台是密不可分的。基因组工程研究就是整合这些高通量技术平台为整合生物学研究服务的。

3. 作为人类社会 21 世纪主导产业的希望之星——基因组工程产业

生物技术作为主导产业,将在 21 世纪取代信息技术,这一观点已被国际上许多国家政府和不同阶层人士所认识。就目前生物技术的发展水平来说,很难担当此重任。

生物技术作为产业对人类最大诱惑力,在于生物体不仅能在常温常压下高效地生产出几十万种甚至上百万种的产品,而且能量转化效率是目前的社会生产能力与之无法比拟的,更为奇观的是能生产思维,认识世界,改造世界。生物体(包括人)之所以有如此法力不在于上帝而是遗传信息。

遗传语文的研究需要同时操作大量基因。而大量基因同时操作,目前生物技术所用的基因工程无法适用,必须发展高通量的技术平台。这些高通量技术平台的组合,我们称它为基因组工程。生命科学进入高通量遗传信息研究需要基因组工程,而且生物技术对生物体进行高通量遗传信息改造也需要基因组工程。

如果基因组工程技术能够成熟,点石成金(生物采矿),化废为宝(将垃圾变成生产原料或方便地转化为能源),植物生产牛奶、动物生产人的组织器官,电脑变成人脑,到那个时代人类社会将会发生什么样的变化,这是难以想象的。基因组工程技术成熟

的标志就是细胞能做什么,人类在工厂里也能做什么或者基本能做。细胞的神奇魅力在于能按遗传语法操作指令同时指挥成千上万个基因有条不紊地工作,基因组工程的魅力不仅在能模仿细胞的功能,而且能改造它,按人类意愿去控制它、指挥它。

如果能够理解基因组工程这些内在的魅力,基因组工程作为 21 世纪生物技术的领头羊也就顺理成章。

参 考 文 献

- [1] Genetic Engineering News. 2001, Vol 21.
- [2] Brown T A. Genetics: A Molecular Approach. 3rd ed. Chapman and Hall Press, 1998.
- [3] Shepherd N S, Smoller D. In Setlow J K, ed. Genetic Engineering. Vol 16. Plenum Press, 1994, pp213-228.
- [4] Shizuya H, Bierren B, Kim U J, et al., 1992, *Proc. Natl. Acad. Sci.*, **89**:8794-8797.
- [5] Burke D T, Carle G F, Olson M V. 1987, *Science*, **236**: 806-812.
- [6] Huxley, C. In Setlow J K, ed. Genetic Engineering. Vol 16. Plenum Press, 1994, pp65-91.
- [7] Kuroiwa Y, Tomizuka K, Shinohara T, et al., 2000, *Nature Biotechnol.*, **18**: 1086-1090.
- [8] Torres R M, Kuhn R. Laboratory Protocols for Conditional Gene Targeting. Oxford University Press, 1997.
- [9] Sternberg N, Hamilton D. J. 1981, *Mol. Biol.*, **150**: 467-486.
- [10] Sauer B, Henderson N. 1988, *Proc. Natl. Acad. Sci.*, **85**: 5166-5170.
- [11] McLeod M, Craft S, Broach J R. 1986, *Mol. Cell Biol.*, **6**: 3357-3367.
- [12] Umlauf S W, Cox M M. 1988, *EMBO. J.*, **7**: 1845-1852.
- [13] Dixon J E, Shaikh A C, Sadowski P D. 1995, *Mol. Microbiol.*, **18**:449-458.
- [14] Araki H, Nakanishi N, Evans B R, et al., 1992, *J. Mol. Biol.*, **225**:25-37.
- [15] Perry A C F, Wakayama T, Kishikawa H, et al., 1999, *Science*, **284**:1180-1183.

IKK 激酶 (IKK) 结构与功能研究进展

董 少 忠

(中国医学科学院 中国协和医科大学医学生物学研究所 昆明 650118)

核转录因子 NF- κ B,是由 Rel 蛋白家族成员组成的二聚体。在哺乳动物中,这些蛋白包括 P50 (NF- κ B1)、P52 (NF- κ B2)、P65 (RelA)、RelB、c-Rel、

P105、和 P100 等^[1]。它们都含有一个由 N 端 300

李琦涵先生对本文进行了审校,谨此致谢。