

基因及基因表达调控

Genes and Regulation of Gene Expression

敖世洲

(中国科学院上海生物化学研究所, 研究员)

生物体的性状是怎样决定的, 又是如何代代相传的? 这是过去一百多年来遗传学提出的主要问题, 现在对这个问题已经有了一个基本的答案。所有生物体均由细胞组成。一种生物的每一个细胞内含有同样组套的DNA, 构成了此种生物的基因组。基因组的DNA序列是从单细胞受精卵发育成为多细胞整体的蓝图。决定人类文化至关重要的学习、语言、记忆等智能因子亦编码在DNA序列之中。DNA序列还编码着突变和变异, 它是人类患病或对疾病易感染的原因

DNA是由四种脱氧核苷酸组成的很长的双链多聚体。用A、G、C和T表示四种核苷酸, 它们分别代表四种核苷酸的碱基: 腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶。在细胞内DNA与一些蛋白质复合构成染色体。染色体因其长度和核苷酸排列顺序不同而异。每个DNA分子含有很多作为功能单位的基因, 基因在DNA分子上以特定的次序排列。大多数基因编码着决定细胞性状的蛋白质分子——酶或结构组分。在细菌中, 一个基因的编码序列是连续的, 而在哺乳动物细胞中, 编码区(外显子)通常由非编码区(内含子)所间隔。每个外显子编码着一个大蛋白质分子中的不同结构域(domain)。单倍体的人类基因组约为30亿碱基对(bp), 编码蛋白质的序列不会超过基因组总顺序的5%。大量DNA序列位于基因之间或内含子之中, 部分非编码顺序对基因活动起调节控制作用, 有些可能在DNA组装成染色体时起重要作用。

细胞的表型、生物体的性状是通过基因的一系列活动决定的。这些活动包括: DNA复制, 它不仅关系到基因的增殖, 细胞分裂, 还涉及免疫反应、分化、组织再生和细胞癌变等诸领域。染色体

DNA复制的控制决定细胞周期性及生物发育过程等。基因重组是DNA序列的重新安排, 它是少数几个引起基因组改变的途径之一, 它对于新性状的产生, 脊椎动物中新抗体的形成及一些疾病的病理过程都是重要的。人们也可以利用基因重组人工地将基因片段引入细胞或生物体的基因组中。DNA的修复和变异, 对生物体保持和保护编码在基因中的信息正确性来说是重要的。然而DNA及它编码的信息一直受到环境的影响, 而且在细胞正常代谢过程中, 如DNA复制和重组中也会导入一些错误。生物体通过DNA修复机制来修正错误。DNA的修复和变异与人类的健康有直接关系。基因贮存着细胞和生物体的所有信息, 这些信息转为RNA和蛋白质的过程即为基因表达。基因表达的调节和控制是指导生物生长、发育和分化, 以及从简单受精卵发育成复杂完整生物体的关键。

二

在进行基因表达时, 基因首先被转录成mRNA, 然后再翻译成蛋白质。这是生物学中最基本的规律。正是在这个规律基础上, 把从微生物一直到人的各类生物统一成为一个大的生物系统。这种转录和翻译是忠实的, 从而决定了生物的遗传特性。但是基因表达的程序、时间和位置是受不同层次的调控元件所控制的。这种控制不仅决定了基因表达的数量, 而且决定了基因表达的时、空秩序性。生物的正常生长、发育和分化都是由于基因受控有序表达的结果。一旦这些调控机制出现差错, 就会导致各种病变, 甚至产生肿瘤。

原核生物基因表达的调控机制研究比较清楚。真核基因表达的调节控制比原核远为复杂, 是目前分子遗传学研究的热点之一。真核基因的表

达,在多层次,受多种因子协同作用,调控的主要部位在基因的转录水平。目前的研究正集中在基因转录的顺式作用元件(cis-acting elements)和反式作用因子(trans-acting factors),以及它们的相互作用等方面。

RNA聚合酶II转录的编码蛋白质的基因,其顺式作用元件包括转录起始位点及其上游30bp处的TATA序列,上游几百bp处的CCAAT序列或GGGCGG序列(GC box),或其他基因的特异调控序列。在基因上或下游远端,或内含子内有增强子(enhancer)序列。在酵母细胞内,类似于增强子的称上游激活序列。有些基因还含有负调控序列抑制子(silencer)等。上述各种特异的核苷酸序列都是反式蛋白质因子作用的靶位点。已被分离纯化或鉴定的蛋白质因子有几百种,新的因子还在不断被发现。结合在TATA附近核苷酸序列上的因子称为通用转录因子,有TFIIA、TFIIB、TFIID、TFIIE等。结合在上游特异核苷酸序列上的蛋白质因子称为转录调控因子,有SPI、CTF、Ap-1, Ap-2, Oct-1, Oct-2, CREB等。

三

反式作用的蛋白质因子一般都具有不同的功能区(domain)。这些功能区含有几十到几百个氨基酸残基。有的功能区是结合DNA特异序列必需的,有的功能区是激活基因转录必需的。不同功能区有自己的特征结构。研究转录调控因子结构和功能关系,对阐明真核基因表达调控的分子机制是基本的。

蛋白质因子DNA结合区的结构有多种不同的类型。早年在原核基因的激活蛋白和阻遏蛋白中发现了螺旋—转角—螺旋(helix—turn—helix)的结构。通过蛋白质与DNA复合物的X光晶体衍射分析,这些蛋白质DNA结合区含有两个 α 螺旋,其间为 β 转角所隔。在氨基酸序列上,它们有很大的可变性,但在高级结构模式上是很稳定的。两个 α 螺旋中,一个是识别螺旋,它的氨基酸残基直接与靶DNA大沟的碱基专一结合,另一螺旋穿过大沟,与DNA形成非特结合。在真核生物中,有一些调控因子属于同源域(homeodomain)蛋白,这一类蛋白质DNA结合区中含有一段长约60个氨基酸的稳态区,能形成螺旋—转角—螺旋的结构,很类似原核的激活蛋白和阻遏蛋白。

锌指(Zinc finger)结构也常见于一些调控因子的DNA结合功能区,其中大约每30个氨基酸就有一对半胱氨酸和一对组氨酸,前者在反平行 β 折叠中,后者在 α 螺旋中,它们能与 Zn^{++} 形成配位键。每一个重复单位形成一个指状结构。不同的因

子,重复单位不等。另外还有一类锌指含有两对半胱氨酸,不含有组氨酸。亮氨酸拉链(leucine zipper)为另一类蛋白质因子DNA结合区的结构类型。在这些蛋白质的DNA结合区中,含有4~5个亮氨酸,其间相距7个氨基酸残基,在 α 螺旋上每两圈就出现一个亮氨酸,这些亮氨酸排成一行。两个蛋白质分子的两个相应的 α 螺旋间,靠亮氨酸残基的疏水作用力形成一条拉链。这些蛋白质因子都是以双体形式与DNA靶位点相结合。新近又发现两种新的DNA结合区的结构模式。一种是免疫球蛋白基因增强子结合蛋白等的DNA结合区含有螺旋—环—螺旋(helix—loop—helix)结构。另一种是CTF家族的DNA结合区含有很多正电荷的 α 螺旋。蛋白质因子的DNA结合区多种结构模式的存在,反映了它们识别序列的多样性和作用方式的复杂性。

反式作用蛋白质因子激活基因转录的功能区一般由20~100个氨基酸残基组成。已发现三种不同的结构类型。一种是酸性激活区,一些蛋白质因子的这一功能区中含有很多酸性氨基酸,在形成 α 螺旋时,表面上带有很多负电荷。氨基酸替换实验表明,激活转录的水平与负电荷的变化有关,增加激活区的负电荷数,能提高激活转录的活性水平。另外,在Sp1的两个主要激活区中含有丰富谷氨酰胺的结构;在CTF家族转录激活的功能区中是脯氨酸丰富的肽段。三种不同的转录激活区的结构类型,可能代表了真核基因转录调控中的不同反应机制。

四

在基因转录起始过程中,涉及很多蛋白质与DNA、蛋白质与蛋白质之间相互作用的复杂反应。不同基因受不同组合的蛋白质因子的协同作用。研究这些反应规律,对阐明真核基因表达调控的分子机制也是很基本的。原核细胞中阻遏蛋白和激活蛋白与靶DNA的作用有高度亲和性的专一结合,而在真核生物中,调控因子与DNA的作用不像原核那样简单和专一,是比较复杂和灵活的。有些调控因子可以与序列相似性很小的几个DNA位点,以近似的亲和性相作用。或者一种DNA特异序列能作为多种蛋白质因子的靶位点。这两种作用方式的存在,增加了单独或协同调节基因表达的精确性和灵活性。特别是对那些诸如细胞生长、发育和分化过程起重要作用的调节和控制,精确性和灵活性是非常重要的。一些蛋白质因子结合DNA特异序列以二聚体形式作用,有的是不同蛋白质分子形成的异源二聚体。异源二聚体的形成,实际上产生了新的蛋白质,这就可能增

加新的识别专一性,或在两种生理条件下影响基因的表达。

另一些蛋白质因子需经物理或化学因素的诱导才具有活性。如热休克转录因子与DNA的结合或激活基因的转录,两种活性必须经热的诱导才能表现。甾体激素受体在细胞质内以非活性形式存在,经激素诱导,它才进入细胞核调节基因的表达。一些受代谢物阻遏或诱导的基因,代谢物的作用也是影响蛋白质因子DNA结合区或转录激活区的结构与功能作用来控制基因的。在其中的一些诱导过程中,常涉及蛋白质的磷酸化反应。不同的蛋白激酶对一些蛋白质因子的功能表现起很重要的作用。

蛋白质因子与特异DNA序列的作用还涉及蛋白质因子之间的反应。前已提到一些因子必须以二聚体或异源二聚体的形式才能结合在DNA上。几种通用转录因子TFⅡA、TFⅡB、TFⅡD、TFⅡE及RNA聚合酶Ⅱ与TATA附近的DNA序列相结合时,各因子之间的作用遵守一定的时空秩序性。首先TFⅡD结合在TATA序列上,TFⅡA结合在它的上游,接着TFⅡB结合在转录起始位点附近。只有上述三种因子分别结合在相应的位置上,RNA聚合酶Ⅱ才能与它们形成一个新的复合物。最后TFⅡE结合在RNA聚合酶Ⅱ的下游,形成一个基本的转录起始复合物。如果顺序颠倒,上述各种起始复合物的中间体都不能形成。在酵母交配型的控制中,不同种类的蛋白质因子的相互作用,能形成不同的作用表面。它们阻遏或激活不同类型的基因表达,以决定酵母的不同交配类型。

一个基因上下游不同调控序列间的距离,有的很远,达几千bp。结合在DNA不同部位的蛋白质因子通过什么方式发生相互作用的呢?一种得到广泛支持的看法认为,结合在DNA不同部位上的蛋白质因子可以通过两个结合位点之间的DNA形成一个环(loop)而发生相互作用。在调控中起作用的是两个结合在DNA不同位点上的蛋白质,而不是DNA环本身。早几年有实验表明,结合上游DNA序列上的调控因子的酸性激活区能与结合在TATA序列上的TFⅡD发生作用,形成稳定的复合物。新近一些报道指出,酸性激活区和谷氨酰胺丰富的激活区与TFⅡD之间的作用,需要通过一种蛋白质接头(adaptor)或辅激活因子(co-activator)才能相互作用。

五

“七五”期间,我国科学工作者也开展了真核基因转录调控研究。从不同的细胞系统,研究了不

同基因的上游调控序列,发现了一些新的元件。如人珠蛋白基因上游20kb片段中决定细胞特异和发育特异的高效表达元件,雄激素调节的高度组织专一性的调控元件,乙肝病毒基因组新的增强子序列和酸性磷酸酯酶基因的上游激活序列等。这些调控元件都具有自己的特征结构,能与细胞内特异的蛋白质因子相结合,并受一定的生理因素影响。通过突变改造,能提高调控序列的功能作用,促进基因的高效表达。这些研究已形成自己的特色,并受到国际同行的关注。

总的说来,真核基因表达调控研究目前仍处在探索阶段,国际上很多研究室正在为阐明基因表达的调控模式作出努力,发展是迅速的。要从分子水平上认识真核基因表达的调控机制,就必须从不同的细胞系统,寻找特定基因的调控元件和因子,分析它们结构和功能的关系,了解它们相互作用的规律,以及各种生理因素对它们功能作用的影响等。在未来十年中,这些问题将会有突破性进展。人们将全面认识基因表达的分子机制,阐明细胞发育和分化的控制机理,了解某些疾病包括肿瘤在内的发病机制,提出针对这些疾病诊断和治疗的新方法,并在细胞工程和基因工程等生物技术领域发展新的概念和新的技术,使它们的面貌发生全新的变革。

(责任编辑 蔡德诚)

(上接第62页)

精力为不履行合同的去经济法院诉诸法律。最后影响从事转化工作的积极性。电气绝缘研究室的一位副教授无可奈何地说,他们就遇到过这种不履行合同的情况,但毫无办法。看来,健全法制,改善科研成果转化的外部环境也是不容忽视的一个因素。

稳定科研队伍是当务之急

西安交通大学的教学科研人员对此感受很深,参加座谈的人几乎一致认为,高校科研队伍的后劲不足,尤其是年轻的科研人员相当不稳定。有一个系,200多科研人员,出国的就达43人,每年有20人左右去深圳,许多科研工作无人干。

一位55岁的副教授深有体会地说,现在搞科研,什么事都得我们这些老头老太太自己跑腿,到仓库领个料也得亲自干,因为没有年轻人,或者有,多数也不愿干。由于高校的科研人员工资收入相对偏低,长期得不到解决,有些系很难留住有水平的青年研究人员,造成科研队伍本身的结构很不合理,这也是影响科研工作本身和成果转化的一个内部因素。

看来,影响科研成果迅速转化成生产力的因素是多方面的,西安交通大学在成果转化中的成功经验和反映出的问题具有一定的代表性,我们希望,这些问题能引起有关部门的重视,并逐步加以解决。

(本刊记者 刘先曙)