

增强子与真核基因转录起始

曲 善 乐

(中国科学院生物物理研究所, 北京)

提 要

许多真核基因的转录起始,除启动子外,还需要增强子。增强子是真核基因组中通过增强转录起始而参与基因表达调控的一类序列,近年不断发现于动物病毒和真核细胞中。对 SV40 增强子及 SV40 早区转录起始已经有比较深入的研究。本文综合近年研究进展,对增强子作用特点及结构特点等作概括介绍,并对其作用机理进行了讨论。

一、概 述

1. 真核基因转录起始研究及增强子的发现

真核基因表达及其调控是现代生物学研究中的一个中心课题,转录起始则是基因表达过程中最早发生的、关键性的步骤。在诸如细胞的分化发育、转化及肿瘤发生等复杂过程中,转录起始的速度和特异性具有十分重要的作用。但是,真核基因方面的研究过去一向落后于原核。近年来随着 DNA 重组与分子克隆技术和离体情况下在 DNA 序列中引起突变尤其是定点突变 (site-directed mutagenesis) 技术的发展,以及真核离体转录系统的建立,真核基因转录起始的研究才打开局面并取得重要进展。由 RNA 聚合酶 II 转录的大多数真核基因其启动子均含有一个 Goldberg-Hogness 序列 (Goldberg-Hogness box, 亦称 TATA box, 以下简称 GH 序列),这个序列位于距转录起始位点约 30 bp (—30) 处。近年研究表明,虽然许多启动子的 GH 序列是转录所必需的,但是有的启动子在其 GH 序列被删减后仍能继续转录。同时有证据证明 GH 序列与转录起始位点的准确定位有关。有关真核启动子研究还阐明,许多真核启动子中位于其 GH 序列上方的区域 (—30 至 —110 之间) 含有一个至几个为启动子功能

所必需的序列,这类序列通常称为上游序列。猿猴病毒 (SV40) T 抗原、疱疹病毒胸苷激酶及哺乳类 β -珠蛋白等基因的启动子均已有比较详细的研究。随着真核基因转录起始研究的逐步深入,人们又在某些基因更远的上游区发现了另一类重要序列——增强子。

增强子或称活化子,是存在于许多真核基因组中通过增进转录起始而参与基因表达调控的一类序列。八十年代初首先发现于 SV40 中^[1-4],随后在其它一些动物病毒及细胞基因组中相继发现有这类序列存在。植物细胞方面亦有报道。增强子在许多真核基因有效转录所不可缺少的,它在一些病毒的感染中起重要作用。在 B 淋巴细胞分化发育成浆细胞过程中, DNA 重排后,免疫球蛋白 (Ig) 重链基因及 κ 链基因其易变区编码序列转移到恒定区增强子附近,形成一种有活性的完整的基因而被选择性地表达。增强子还与肿瘤发生有关。禽类造血细胞组织增生病毒 (ALV) 可诱发鸡 B 淋巴细胞瘤,感染后 ALV 的 DNA 序列整合到细胞致癌基因 c-myc 附近致使该基因活化。起初认为这种活化是由于“启动子插入”,后来证明是插入的病毒增强子序列使上述基因自身的启动子活化所致。有关研究还表明细胞增强子也与致癌基因的活化有关。鼠类及人类许多淋巴

系统的瘤的产生涉及染色体移位,在这种移位中,DNA重排后致癌基因移至Ig基因恒定区附近,由于邻近的增强子的作用而使其转录活化。

2. 增强子研究中的短期表达测定 (transient expression assay)

在增强子研究中,其活性通常是通过克隆基因活体表达而进行测定的,一般采用短期表达测定方法。其步骤是:将增强子与基因一起组建质粒,用组建质粒(或病毒重组体)转染宿主细胞40—48小时后测定转录产物或表达的蛋白产物。用放射性DNA探针,转录的RNA可用S1核酸酶分析方法进行测定。用这种方法对同一凝胶进行不同曝光的放射自显影密度测定,可测到比野生型SV40转录水平低三个数量级的数值^[5]。测定蛋白产物的一种被广泛应用的方法是CAT短期表达测定。组建质粒时用大肠杆菌氯霉素乙酰基转移酶(CAT)编码序列与真核启动子区域相连;这样,转染细胞(在正常情况下真核细胞不含CAT)后便可用简便快速而灵敏的酶活测定方法定量地测定表达产生的CAT量。

此外,附带提一下,近年用离体转录系统研究增强子作用也有某些进展。离体转录研究表明,SV40增强子对转录的刺激作用涉及细胞转录因子。用Hela细胞核抽提液,在最适离子条件及DNA浓度,并在多胺化合物亚精胺存在下,SV40增强子可有效地增强SV40早区启动子的转录^[6]。

二、增强子作用的特点

1. 增强子增进转录起始的作用相对地不依赖于其所在位置与方向 增强子可以两种相反方向在基因的上游区、下游区或基因内(内含子中)起作用。SV40增强子的系列重组研究还说明,在包括两个或两个以上转录单位的重组体中,增强子优先活化邻近的同源或异源启动子。而且,增强子在与转录方向相背位置往往比在相向位置时的增强效率高^[7]。

2. 增强子只有与所作用的基因在同一

DNA分子序列中才能起增强作用,否则无作用 例如,用以下两种组建质粒共同转染宿主细胞,一种质粒含增强子(SV40增强子或Ig重链增强子),但不含基因,另一种质粒含包括启动子的基因(删减增强子序列的SV40早区基因或伴清蛋白基因),但不含增强子序列,共转染后观察不到增强子的增强作用^[4,8]。

3. 增强子可在距转录起始位点相当远的距离起增强作用 增强子至少可从几个kb的距离起作用,但其增强作用随距离增加而变弱。在SV40增强子与同源SV40早区启动子或异源伴清蛋白启动子之间插入不同长度DNA片段的研究^[9]表明,当插入片段在150bp范围内时,随着插入片段长度的增加,转录水平急剧降低直至降为10%以下。插入较长片段时其降低趋缓,插入275bp片段时降为4%。而插入更长片段650bp或3737bp时则均降为0.5%,但与不含增强子的重组体相比,仍能使转录至少增强10倍。据报道,Ig重链增强子可使相距17.5kb之远的V_H(重链易变区)启动子活化,并具有活化两个并列的相距15.8kb的启动子的能力^[10]。

4. 增强子对异源基因也有增强功能 前面的一些例子已经说明了这个问题。利用增强子可研究那些含有弱启动子基因的表达。例如,SV40增强子可使Igλ基因及β-珠蛋白基因等在Hela细胞中有效地表达。而在没有增强子存在情况下,上述基因在Hela细胞中的转录水平很难测出或非常低。

5. 增强子具有相对的种属特异性(宿主细胞特异性) 增强子在不同种属细胞中其增强活性有所不同。不同来源的增强子对不同种属细胞具有不同的相对特异性。例如,SV40增强子在猴肾CV-1细胞及HeLa细胞中增强转录的活性为鼠类肉瘤病毒(MSV)增强子的5倍多,而在小鼠L细胞中后者活性为前者的两倍半^[11]。

此外,Ig增强子具有组织特异性(细胞类型特异性),即:在淋巴细胞中有增进转录作用,在成纤维细胞或HeLa细胞中则无活性。有

的病毒增强子也有组织特异性,据报道,淋巴营性乳多空病毒 (LPV) 增强子在人体造血系统细胞中有增强活性,在成纤维细胞或上皮细胞中无活性^[12]。

上述种属及组织特异性表明增强子作用与细胞因子有关,这已为愈来愈多的证据所证明。

三、关于增强子的结构

SV40 增强子及 Ig 增强子等的研究结果表明,具有一定活性的增强子基本序列其长度可能为一百几十 bp (至少 100bp); 但具有充分活性的增强子则涉及更长的 DNA 序列 (200bp 以上)。已知的不同来源增强子之间虽然没有广泛的序列同源性,但存在某种短的共同性序列。在病毒增强子序列比较及 SV40 增强子随机诱变研究基础上,Weiher 等^[13] (1983) 提出 SV40 的 72bp 重复序列中及其他一些病毒增强子中的 5'-GTGG ^{AAA}_{TTT} G-3' 可能是对增强子活性起关键作用的“核心”序列。细胞 Ig 增强

	GTGG ^{AAA} _{TTT} G (或 GTGG ^{AAAA} _{TTTT} G)
SV40	GGTGTGGAAAGTCC
Py(多瘤病毒) i	CGTGTGGT ^{TTT} TTGCA
	ii GGCCTGGAATGTTT
MSV	TCTGTGGTAAGCGG
BKV	GTCATGGTTTGCT
BPV	GGAGTGGTGTGGAC
Ig 重链	GCTGTGGTTTGAAGAAGTGGTTTGA
Igκ 链 i	TTCTTGGTAAAGAA
	ii CTCTCGGAAAGTCC

图 1 增强子的“核心”序列

子中也含有这种序列(图 1)。此外,各种增强子还含有其他在功能上起重要作用的序列。

四、SV40 早区转录起始与 SV40 增强子

1. SV40 早区转录起始及 SV40 早区启动子区域有关序列的作用

真核基因表达及其调控方面知识有许多是来自动物病毒研究。由于病毒本身遗传容量有限,须依靠细胞的酶;故病毒基因表达调控,可以作为研究 RNA 聚合酶 II 转录的细胞基因调控基本机理的一种模式系统。SV40 是研究最多的一种动物病毒,其基因组为含 5243bp 的共价闭环分子。基因组编码序列分为两个区域,即早区和晚区,分别位于基因组的相对两个半边。它们以相反方向转录。早区编码大、小 T-抗原两种蛋白质,晚区则编码三种壳蛋白 VP1, 2 及 3。早区启动子区域位于早区编码序列与晚区编码序列之间。早区启动子包括 RNA 聚合酶 II 作用的两个重叠的启动子,即早-早区 (EE) 启动子和晚-早区 (LE) 启动子。早区在整个裂解感染周期中都进行转录。在感染早期,在转化细胞中是从早-早区起始位点 (EES) 起始转录,在晚期阶段,则从晚-早区起始位点 (LES) 起始转录。除转录起始位点外,早区启动子区域含有以下与转录起始有关的序列: GH 序列 (位于 SV40 基因组核苷酸位置 15—21), 21bp 重复序列 (位置 40—103, 是三个不完全重复的富含 GC 的序列,其中两个为 21bp, 一个为 22bp)。以及位于更远上游区的增强子 (图 2)。从 EES 起始转录时, GH 序列, 21bp 重复序列及增强子三者都是不可缺少的。而从 LES 转录的有效起始,也需要 21bp 重复序列和增强子,但不存在 GH 序列时亦可进行。21bp

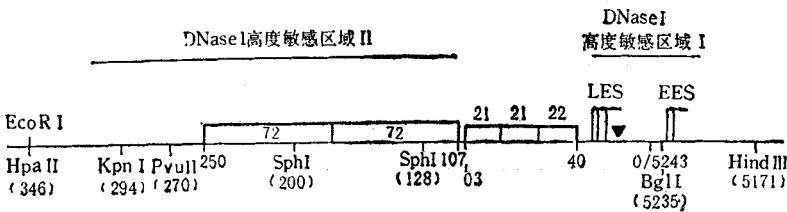


图 2 SV40 早区启动子区域(或称 ORI 区域): ▼: GH 序列

重复序列中含有 6 个与功能有关的 GC 短段 (CCGCCC), 这 6 个短段对从 EES 和 LES 转录所起作用不是等同的, 距 GH 序列较近的几个 GC 短段对从 EES 起始最为重要, 而从 LES 起始则主要依赖于上游的几个 GC 短段。增强子的存在可使早区转录水平增高三个数量级以上, 其所含的为活性所必需的序列短段对从 EES 和 LES 起始转录的作用则是相同的。

2. SV40 增强子的结构特点

最近, Zenke 等^[5](1986)对 SV40 增强子进行了系统定点突变及删减研究, 比较详细地阐明了其结构特点。他们用兔 β -珠蛋白编码序列取代大 T 抗原编码序列, 以消除表达产物大 T 抗原对早区启动子的影响。为便于定点突变, 将野生型 SV40 增强子的 72bp 重复序列 (位置 107—250) 中靠近 21bp 重复序列的一个 (位置 107—178) 准确地删减 (删减一个 72bp 后增强子活性降为原来的三分之一), 在此基础上进行系统的定点突变或删减。组建质粒并转染 HeLa 细胞后, 以一定长度的与由 EES 或 LES 开始的 mRNA 相对应的单链 DNA 作为保护性探针, 用核酸酶 S1 分析方法定量测定转录产物。实验结果表明, 具有基本活性的增强子序列至少应是包括单拷贝 72bp 序列及其上方邻接的近 30bp 的约 100bp 区域。这个约 100bp 序列又含有两个重要区域: A 和 B (图 3)。区域 A 或 B 单独只有很小的增强活性 (与增强子区域所有序列都不存在时相比, 单独的 A 或 B 仅能使转录水平增加 4—8 倍)。但 A 和 B 连在一起便产生飞跃式变化, 可使启动子活性增强 400 倍。而且其增强作用也相对地不依赖于 A 或 B 的相对方向 (A 或 B 可像整个增强子那样在两个相反方向起作用) 及两者之间在一定长度范围内

的距离 (A 与 B 之间插入 58bp 仍保持大部分活性)。因此, 实际上 A 和 B 是相对独立的两个结构域。两个 A 或两个 B 相连也比其单体的效率高, 均有一定程度的增强子活性。A 和 B 相连获得基本的增强子活性之后, 再增加 72bp 序列 (一个或少数几个) 则其活性将随 72bp 序列的增加而呈线性增加。增加一个 72bp 序列可使活性增加 3 倍多。若增加 94bp 序列 (位置 179—272 之间序列, 含结构域 A 和 B), 则比增加 72bp 更为有效。

SV40 增强子的显著特征之一是单个定点突变 (三个相邻核苷酸的非互补换异型碱突变) 没有能使其活性降低至原来的八分之一以下者。表明增强子的活性涉及多个序列。结构域 A 和 B 都含有在功能上起重要作用的序列, 而且是以重复形式存在。结构域 B 含有 5'-G^C_T TGTGGAA^A_T GT-3' 重复序列, 称为 GT 短段 I 和 II (图 3)。GT 短段 II 位于 72bp 序列上方的序列中。GT 短段 I 位于 72bp 序列中, 含有 Weiher 等提出的所谓“核心”序列。GT 短段 I 下方紧跟着有 5'-TCCCCAG-3' 重复, 其中上游的一个比较重要。5'-TGTGGAAA-GTCCCCA-3' 对 SV40 增强子的功能是非常重要的。其中的 5'-TGGAAAGTCCCC-3' 也存在于人的细胞巨病毒增强子中 4 个 18bp 直接重复序列中。GT 短段中的 5'-GGTGTGG-3' 也以重复形式出现在某些增强子中。结构域 A 最重要的特征是含有 5'-AAG^T_C ACGCA-3' 直接重复序列, 因序列中含限制性内切酶 SphI 位点, 故称为 Sph 短段 I 和 II。其它某些增强子也含有这种序列。上述这些序列短段其中任何一个

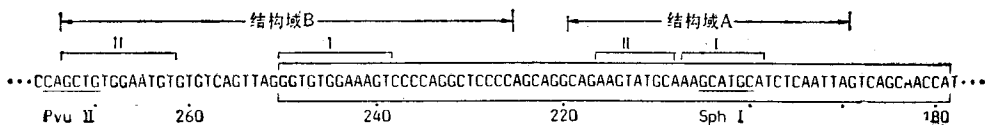


图 3 SV40 增强子的结构域 A 和结构域 B

方框内为 72bp 序列、结构域 A 含 Sph 短段 I 及 II, 结构域 B 含 GT 短段 I 及 II。

发生突变都会使增强子的功能受到严重损害。

除上述由结构域A和B组成的基本序列外,位置284—347之间序列可使增强活性加倍,其中298—347之间序列当结构域A和B处于完整状态时,对增强子的活性只起很小作用,但在增强子的某些区域被删减时可起一定的补偿作用。

3. 转录起始涉及的细胞因子

有关研究特别是离体转录研究表明,SV40早区的转录起始涉及某些不同的细胞转录因子特异地与控制区域的各有关序列结合。GH序列因子与GH序列结合可形成一种稳定的复合物,这是特异的转录起始的前提。转录因子Sp1与21bp重复区域特异结合,并促进从EES及LES转录。增强子因子则与SV40增强子结合并刺激由早区启动子或异源启动子的转录^[6,14,15]。增强子因子可与不同来源增强子结合,但其结合能力有差异。

五、免疫球蛋白基因增强子

(Ig基因增强子或Ig增强子)

已发现的Ig增强子有Ig重链增强子^[8,16]和Ig κ 轻链增强子^[17,18]。Ig基因分析表明,Ig多肽链是由分散在胚细胞一个染色体上的多个DNA区域编码。在B淋巴细胞分化发育成分泌抗体的浆细胞过程中,编码Ig κ 链易变(V)区及编码Ig重链V区的DNA序列分别通过VJ及VDJ重排,从远处转移到各自的恒定(C)区序列上游的增强子附近,从而形成完整的基因并导致它们的V区启动子活化。重排后,Ig重链增强子及 κ 链增强子位于各自基因主要内含子中。在胚细胞中Ig重链增强子是在J_H(重链连接区)与C μ (C区编码序列 μ)之间,在重排后成熟淋巴细胞中所发生的转换重组则使C μ 被C区其它编码序列(γ 、 ϵ 、 α 等)取代。在浆细胞中Ig基因转录产物可达全部mRNA的10%,而未重排的V区其它编码序列其转录水平则仍然是很难测出的。

Ig增强子的一个重要特点是具有组织特异性。不同增强子之间的竞争研究表明,在B淋

巴细胞中Ig增强子和SV40增强子可与共同的细胞因子结合;而成纤维细胞所含转录因子也能与SV40增强子结合,但不能与Ig增强子结合^[19]。另外有人报道Ig重链增强子的淋巴细胞特异性与其所含的一种序列有关^[20]。

六、关于增强子的作用机理

由于增强子具有前面介绍的一些特性,所以它的作用机理一直是人们感兴趣的一个问题。曾经有几种假说试图解释增强子的作用。一种假说认为,增强子为RNA聚合酶II提供了双向进入位点;另外一种意见则认为,增强子是通过改变染色质结构或DNA超螺旋结构而行使其功能的。此外还有其它说法。迄今,虽然有实验证明增强子所引起的稳定转录产物的增加确是由于转录起始增强所致^[21,22],而且也已证明,增强子是与细胞因子结合而起作用的,并与对核酸酶敏感的染色质结构的形成有关;但对其详细作用机理仍然不清楚。现将一些有关研究简略介绍并讨论如下:

在感染晚期感染细胞核中的SV40 DNA以微染色体形式存在。当用各种限制性酶消化时,BglI与HpaII位点之间的ORI区域(位置5235—346)先被切割。而且,一部分SV40微染色体含有一个无核小体区域(核小体袂)也位于ORI区域。用DNaseI所作研究说明ORI区域含有两个高度敏感区域,即区域I(包括转录起始位点及GH序列,大约几十个bp)和区域II(约在位置110—320)(图2)。为了研究DNaseI高度敏感区域及无核小体区域形成所需要的序列,Jongstra等^[23]将ORI区域或其部分序列插入已删减大T抗原内含子的SV40基因组中,组建一系列SV40突变体,转染CV-1细胞后分离DNA进行研究。结果表明,插入序列所导致的DNaseI高度敏感区域II的产生主要依靠位置113—270之间的增强子序列,但要产生足够长度的区域II还需要PvuII与HpaII位点之间序列(位置270—346)。将位置113—270之间片段删减一个72bp后仍可产生一个相应短的敏感区域,但其敏感程度较差。

由上述可见, 插入不同范围增强子序列所产生的敏感区域的敏感性与其增强活性呈一定的相关性(参看第四节)。区域 I 的产生则要求 21bp 重复序列和 HindIII-BglI 之间序列(位置 5171—5235)包括在相应的插入片段中。插入较长 ORI 区域序列可观察到无核小体区域。离体重建研究表明, 与 SV40 DNA 其它部分相比, ORI 区域不容易形成核小体。因此可以认为, 增强子及 ORI 区域中其它有关序列的存在, 使局部区域染色质结构成为对核酸酶高度敏感的某种开放性结构。这种开放性结构有利于与细胞因子相互作用。但是也有可能是增强子及 ORI 区域其它序列被细胞因子识别, 并特异结合而导致局部区域染色质结构改变。利用抗 Z-DNA 抗体所作研究证明, 在离体一定条件下, SV40 增强子可局部形成 Z-DNA^[24]。SV40 DNA 有三个主要抗 Z-DNA 抗体结合位点, 均在增强子序列上(分别在位置 130, 202 及 261), 若将 DNaseI 高度敏感位点放上去, 则可看到 DNaseI 高度敏感位点位于每个局部 Z-DNA 形成区两边约 20—30bp 处。但增强子作用过程中是否涉及 Z-DNA 形成, 尚待证实。离体转录研究表明, 在一定条件的离体情况下, SV40 增强子通过与细胞因子结合可刺激由异源启动子^[14]及同源启动子^[6]的转录。线型 DNA 和超螺旋 DNA 具有相同的作用^[14,15]。但在染色体重建情况下则失去增强作用^[15]。

最近报道^[25], 在 SV40 增强子与 21bp 重复区域或 21bp 重复区域与 GH 序列之间插入半个 DNA 螺旋周的奇数或偶数倍的片段, 观察其对转录的影响。短期表达测定结果表明, 上述两种不同插入对转录的影响有明显不同。插入奇数倍者往往影响较大, 说明上述序列之间需要一定的空间排列, 满足这种空间排列才能最有效地起始转录。由于增强子结构域 A 与 B

之间插入一定长度的片段(4—50bp)无显著影响, 故对 SV40 增强子与 21bp 重复序列而言, 增强子结构域 A (可能是其所含的 Sph 短段)与 21bp 重复区域的 GC 短段之间的空间排列是更为重要的。各种序列之间的空间关系, 实际上是它们所结合的各种细胞转录因子之间需要一定空间排列的反映, 这说明增强子所增进的转录起始, 除蛋白质与核酸之间相互作用外, 还涉及多元的蛋白质与蛋白质之间的相互作用, 包括不同转录因子之间以及它们与 RNA 聚合酶 II 之间的相互作用。

参 考 文 献

- [1] Benoist, C. et al.: *Nature*, **290**, 304, 1981.
- [2] Gruss, P. et al.: *Proc. Natl. Acad. Sci. USA*, **78**, 943, 1981.
- [3] Benerji, J. et al.: *Cell*, **27**, 299, 1981.
- [4] Moreau, P. et al.: *Nucleic Acids Res.*, **9**, 6043, 1981.
- [5] Zenke, M. et al.: *EMBO J.*, **5**, 387, 1986.
- [6] Wildeman, A. G. et al.: *EMBO J.*, **3**, 3129, 1984.
- [7] Waslylyk, B. et al.: *Cell*, **32**, 503, 1983.
- [8] Benerji, J. et al.: *Cell*, **33**, 729, 1983.
- [9] Waslylyk, B. et al.: *Nucleic Acids Res.*, **12**, 5589, 1984.
- [10] Wang, X. F. et al.: *Cell*, **43**, 659, 1985.
- [11] Laimins, L. A. et al.: *Proc. Natl. Acad. Sci. USA*, **79**, 6453, 1982.
- [12] Mosthaf, L. et al.: *Nature*, **315**, 597, 1985.
- [13] Weiher, H. et al.: *Science*, **219**, 626, 1983.
- [14] Sassone-Corsi, P. et al.: *Proc. Natl. Acad. Sci. USA*, **81**, 308, 1984.
- [15] Sergeant, A. et al.: *J. Mol. Biol.*, **180**, 577, 1984.
- [16] Gillies, S. D. et al.: *Cell*, **33**, 717, 1983.
- [17] Picard, D. et al.: *Nature*, **307**, 80, 1984.
- [18] Bergman, Y. et al.: *Proc. Natl. Acad. Sci. USA*, **81**, 7041, 1984.
- [19] Mercola, M. et al.: *Science*, **227**, 266, 1985.
- [20] Church, G. M. et al.: *Nature*, **313**, 798, 1985.
- [21] Treisman, R. et al.: *Nature*, **315**, 72, 1985.
- [22] Weber, F. et al.: *Nature*, **315**, 75, 1985.
- [23] Jongstra, J. et al.: *Nature*, **307**, 708, 1984.
- [24] Norheim, A. et al.: *Nature*, **303**, 674, 1983.
- [25] Takahashi, K. et al.: *Nature*, **319**, 121, 1986.

【本文于 1986 年 11 月 19 日收到】