

医学分子遗传学

第二讲 人体基因的表达与调控

俞民澍 薛京伦

(复旦大学遗传学研究所,上海)

基因突变可以导致遗传性疾病,这是由于基因突变后,基因的表达发生了改变或失去控制,从而产生了一系列临床症状。因此,在研究遗传性疾病的分子病理机制之前,必须对正常人体基因的表达及其调控机制有很好的了解。但是,在相当长的一段时间里,人们对人体基因的结构特征、表达过程及其调控机制,只能根据从原核生物中获得的研究结果进行推测。1977年,真核生物基因结构的揭示和 mRNA 剪接加工现象的发现,使人们对人体基因结构与功能的认识向前迈进了一大步。近十年来,随着真核生物基因的分子遗传学研究不断深入,人们对真核生物基因的结构和转录、转录后加工、翻译及翻译后修饰等过程有了较为详细的了解,使得医学遗传学工作者有可能从基因表达与调控角度分析和研究遗传性疾病的分子病理机制。

真核基因的断裂结构

(一) 外显子与内含子

大多数真核生物基因都是由编码顺序和非编码顺序两部分组成。编码顺序称为外显子,非编码顺序称为内含子。在一个结构基因中,编码某一蛋白质不同区域的各个外显子不是连续排列在一起的,而是被长度不等的内含子分隔开来,形成镶嵌排列的断裂形式。因此又有人将真核生物基因称作断裂基因(interrupted gene)。

不同基因所含内含子的数目和大小相差非常悬殊。如胶原蛋白基因,长约 40kb,至少具有 40 个内含子,其中短的只有 50bp,长的可达 2000bp。有少数基因根本不含内含子,如 α 型和 β 型干扰素。内含子在结构基因中的具体作用尚不清楚,但已发现,在某些基因中,完全去除内含子后,照样可产生有活性的 mRNA 转录产物。然而,在另一些基因中,如 SV40T 抗原基因,一旦将基因中的内含子去除,便可阻止有功能的 mRNA 产物进入细胞质。采用几种不同的限制性内切酶,其中一部分限制性内切酶在内含子中没有识别位点,因而使得这一部分片段显得比较长,另一部分内切酶则能够将内含子裂解成两段,这样就多出了一些

片段,将这些结果综合起来进行分析,并与有功能的 mRNA 上限制性内切酶分析的结果进行比较,就可确定内含子的大小及其在基因中与外显子的排列方式。断裂基因的结构形式为一些蛋白质的编码区域提供了进行重组的机会,这对于真核生物基因的进化是极为有利的。胶原蛋白基因就是通过重复而产生的。

(二) 一个基因的内含子可以是另一个基因的外显子

在基因转录成为成熟的有功能的 mRNA 分子时,内含子通过剪接加工过程被去掉,保留在成熟 mRNA 分子中的外显子被拼接在一起,最终被翻译成蛋白质。因此,通过反转录酶的作用,由成熟 mRNA 产生的 cDNA 分子中,只含有外显子,不含内含子。外显子和内含子的关系并不完全是固定不变的。有时,在同一条 DNA 链上的某一段 DNA 顺序,当它作为编码某一条多肽链的基因时是外显子,但在作为编码另一条多肽链的基因时则成了内含子。这是由于 mRNA 剪接加工的方式不同所造成的。结果,同一个基因(严格地说是同一段 DNA 顺序)产生了两条或两条以上的 mRNA 链。在真核生物基因中,一个基因的内含子成为另一个基因的外显子,形成基因的差别表达,这是基因断裂结构上的一个重要特点。

(三) 外显子—内含子接头 (exon—intron junctions)

真核生物基因断裂结构的另一个重要特点是外显子—内含子接头高度保守的特异性碱基顺序。外显子—内含子接头就是指外显子和内含子的交界处(boundaries),或称边界顺序。它有两个重要特征:(1)一个内含子的两端顺序之间没有广泛的同源性,因此内含子的两端顺序不能互补。这说明,在剪接加工之前,内含子上游顺序和下游顺序不可能通过碱基配对形成发夹样二级结构。(2)外显子—内含子接头顺序相当短,但却是高度保守的一致顺序(consensus sequence)。这一顺序与剪接机制密切相关,它是 RNA 剪接的信号顺序。现已证明,每一个内含子 5' 端起始的两个碱基是 GT, 3' 端最后的两个碱基是 AG。这两个顺序是高度保守的,在各种真核基因内含子中都是一样的,

因此又把这种形式称为 GT-AG 法则。

位于内含子两端的接头顺序是不同的,因此可以定向标明内含子的两个末端。由于剪接加工过程是沿内含子自左向右进行的,所以将内含子的 5' 接头顺序称为左剪接位点,3' 接头顺序称为右剪接位点。有时也将前者称为供体位点 (donor site),后者称为受体位点 (acceptor site)。外显子-内含子接头的一致顺序几乎存在于所有的高等真核生物基因中。这表明,在高等真核生物基因中,可能存在着一个共同的剪接加工机制。但是,在线粒体基因和酵母 tRNA 基因中不存在这类一致顺序。这说明,在这些基因中,还可能存在着另一类剪接加工机制。一个真核基因结构除了外显子和内含子外,还包括一些共同的侧翼顺序 (flanking sequence),如启动子、增强子、多聚腺苷化顺序以及 mRNA 5' 端的帽等,这些结构对于基因的有效表达是必不可少的,它们起着调控基因表达的作用。

转 录

(一) 转录过程

从基因的 DNA 顺序合成 mRNA 分子的过程称为转录 (transcription)。转录是遗传信息从 DNA 流向蛋白质的关键步骤。真核生物基因开始进行转录时, RNA 多聚酶首先与启动子 (promotor) 结合,这是 mRNA 开始合成的信号。RNA 聚合酶使双螺旋 DNA 分子解旋,然后,沿着 DNA 模板,以 5'→3' 方向移动。到达转录起始点后,在 RNA 聚合酶的作用下,一个个核糖核苷三磷酸与 DNA 链上的互补碱基配对,合成 mRNA 链。这一合成过程一直持续到 RNA 聚合酶在 DNA 模板上遇到转录终止信号为止。但直到现在,真核基因转录终止信号尚未能鉴定出来。最近认为,真核基因转录的终止可能与转录出来的 mRNA 形成的特殊二级构象有关。当转录终止后, RNA 聚合酶释放出新合成的 mRNA 链。此时,合成的 mRNA 链未经任何修饰,相当于一个转录单位,称为初级转录产物。一旦初级转录产物合成后, mRNA 链的两端立即在细胞核内被修饰加工 (processing)。在 5' 端加上一个帽结构,在 3' 端加上一串腺嘌呤核苷酸,称为多聚 A (poly A) 尾。

(二) RNA 聚合酶 (RNA polymerase)

真核细胞中负责转录核基因的 RNA 聚合酶有三类: RNA 聚合酶 I、II 和 III。它们的分子量都在 500,000—700,000 左右,各由 10—15 个独立的亚单位所构成。RNA 聚合酶 I 转录核糖体 RNA 前体基因; RNA 聚合酶 II 转录编码蛋白质的基因,生成 mRNA; RNA 聚合酶 III 则转录 tRNA 前体基因和较少的核糖体 RNA。根据它们对抑制剂的敏感性,可以将这三类酶区分开来。细胞核中 RNA 聚合酶 II 的转录产物称为不均一性核 RNA (hn RNA)。大多数不均一

性核 RNA 都在细胞核内加工,然后才释放到细胞质中去。

(三) 启动子

启动子通常位于基因的 5' 端,是转录起始时 RNA 聚合酶结合的位置。启动子也可以决定转录两条 DNA 链中的哪一条链。目前,在编码蛋白质的真核基因 5' 端侧翼区中,已发现至少有三个启动子顺序。(1)TATA 顺序:位于转录起始位点上游 25—30bp 处。在这一顺序中,头两个核苷酸和最后第二个核苷酸是高度保守的, RNA 聚合酶 II 可能就精确地结合在这一位置上。有证据表明, TATA 顺序能保证转录起始位点的精确性,又对转录水平有定量效应。(2)CAAT 顺序:位于转录起始位点上游 70—80bp 处。RNA 聚合酶 II 可以识别这一比较保守的一致顺序。过去认为,这一顺序的改变对基因转录没有明显的影响,但近年来在对家兔 β 珠蛋白基因 CAAT 顺序的研究中发现,人工诱导 CAAT 顺序突变,可以使家兔 β 珠蛋白基因的转录水平降低。(3)远端因子 (distal element):位于转录起始点上游 800—1000bp 处。其基本顺序为 PuCPuCCC (Pu 指嘌呤)。这一顺序存在于 β 珠蛋白基因中,但在表达水平较差的 δ 珠蛋白基因中则没有发现。这一因子与疱疹病毒中胸腺嘧啶激酶基因和 SV40 病毒早基因区域中的启动子顺序非常相似。对家兔 β 珠蛋白基因中的远端因子进行人工定向诱变发现,诱变后 β 珠蛋白基因的转录水平至少可降低两倍,而且已经发现,在人类 β 珠蛋白基因中,这一顺序的突变也能严重影响基因的表达。

(四) 增强子 (enhancer)

增强子最早是从 SV40 病毒中分离得到的。它包括启动子上游与启动子重叠的一段 DNA 顺序。增强子的作用是使启动子指挥转录的能力大大增强,从而增强基因的转录效率。增强子的作用与所在的位置和方向基本无关,可在转录起始点前后 3kb 或 3kb 以上范围内发挥作用,作用方向可以是 5'→3',也可以是 3'→5'。各个基因中的增强子顺序差别较大,但都有一个基本的核心顺序。

增强子具有组织特异性,例如免疫球蛋白基因的增强子只有在 B 淋巴细胞内活性最高。在胰岛素基因和胰凝乳蛋白酶基因的增强子中都发现了很强的组织特异性。人体基因的转录与增强子的关系比较复杂。若将人体 β 珠蛋白基因引入异源性细胞如 HeLa 细胞中, β 珠蛋白基因的转录就必须有一个顺式增强子存在;而若将 α 珠蛋白基因引入 HeLa 细胞,它的转录就不需要增强子。进一步还发现,如果 HeLa 细胞中原先就存在病毒基因产物,而且是以反式形式存在,那么将人体 β 珠蛋白基因引入细胞后进行转录时,就不再需要增强子。因此,对人体基因转录与增强子的关系还有待进一步深入研究。

初级转录产物的加工

初级转录产物合成后就被包装在核糖核蛋白颗粒中,在其中进行一系列的 mRNA 加工步骤。初级转录产物的加工过程主要分为两步:(1) 5'端加帽和 3'端加 Poly(A) 尾;(2) 剪接。从初级转录产物中除去内含子,再将外显子拼接在一起。

(一) 初级转录产物 5' 端和 3' 端的加工

1. 5' 端加帽 真核生物 mRNA 与原核生物 mRNA 的主要区别之一在于,真核生物 mRNA 5' 端具有一个帽结构,即 7 甲基鸟嘌呤核苷酸与 mRNA 5' 末端连接。新合成的 mRNA 在 5' 端有一个 5'-三磷酸(通常是连结在腺嘌呤上),它可以被酶水解成为 5'-二磷酸,成为鸟嘌呤残基的受体,然后鸟嘌呤核苷酸在加帽酶——鸟嘌呤转移酶的作用下与这一受体相结合,接着,连接在 5' 端的鸟嘌呤残基经甲基转移酶作用,在第 7 位上进行甲基化,从而完成加帽过程。此过程进行得很快,在 mRNA 分子转录形成后,加帽过程紧接着就完成了。mRNA 的加帽有两个作用:(1) 促进 mRNA 和核糖体结合,这一作用是通过帽结合蛋白的介导完成的。帽结合蛋白能专一性地与帽结构相结合;(2) 增强 mRNA 的稳定性,使它免受 5' 核苷酸酶的作用。原核 mRNA 由于缺乏帽结构,被降解的速度远比真核 mRNA 为快。

2. 3' 端加 Poly(A) 尾 绝大多数真核 mRNA 在进行 5' 端加工时,同时还在 3' 端进行修饰,这种修饰方式就是在 3' 末端附加上一串腺嘌呤核苷酸尾,这一反应称为多聚腺苷酸化反应。整个过程是由 Poly(A) 聚合酶催化的。典型的真核生物 mRNA 的 3' 端尾由 100—200 个腺嘌呤核苷酸残基组成。多聚腺苷酸化反应发生在细胞核内,在初级转录产物刚形成后便进行。已经发现,编码组蛋白的 mRNA 通常不需要多聚腺苷酸化。大多数真核基因的 3' 端有一段顺序 AATAAA,多聚腺苷酸化开始于 mRNA 上这段顺序下游 10—15bp 处,在多聚腺苷酸化进行前,必须先先在 3' 端水解掉 10—15 个核苷酸,因此这段顺序称为 RNA 裂解信号,但它不是转录终止信号,转录过程可继续向下游进行。RNA 裂解信号的作用是指导核酸内切酶在信号下游 10—15bp 的特定位置点上裂解 mRNA,然后,在多聚酶的作用下,在 3' 端加上 100—200 个腺嘌呤核苷酸。RNA 裂解信号顺序的突变可以导致 mRNA 转录延长。人类 α 珠蛋白基因中的这一信号顺序发生突变后,产生了过长的 mRNA,由于这种 mRNA 很不稳定,结果引起 α 珠蛋白 mRNA 缺乏,导致 α 地中海贫血症。

多聚腺苷酸化的功能尚不清楚。可能与下列机制有关:(1) 在 mRNA 从细胞核转运入细胞质的过程中起作用;(2) 延长 mRNA 的功能寿命。已有实验

证明,多聚腺苷酸化 mRNA 的功能寿命比非多聚腺苷酸化的长,组蛋白 mRNA 经人工多聚腺苷酸化后寿命增长;(3) 多聚 A 尾能使内含子两端的剪接位点排列在一条线上,以便进行正确的剪接加工。

(二) RNA 剪接 (splicing)

hn RNA 的平均长度为 3000 核苷酸,一般蛋白质有 400 个氨基酸,因此,对一个蛋白质来说,实际上只需要有 1200 个核苷酸组成的编码顺序就够了,hnRNA 与实际编码顺序之间大小差别如此悬殊,是由于存在内含子的缘故。一般来说,两者大小可相差 100—10,000 个核苷酸。与外显子不同,即使内含子中的核苷酸顺序有许多改变,也不会因此而影响基因的功能。但是当内含子两端的外显子—内含子接头顺序中的核苷酸发生改变,虽然只涉及极少数几个核苷酸,却能严重破坏和影响基因功能。

在酶的作用下,初级转录产物在边界顺序上发生剪切 (excision),去除内含子,然后再将编码同一蛋白质的各个外显子拼接起来,这一过程就称剪接。剪接过程必须非常精确。在边界顺序上,只要有一个核苷酸发生错误,就会完全改变 mRNA 分子的阅读框,形成无意义的 mRNA 链,从一个 mRNA 初级转录产物中,可以同时去除几个内含子顺序。

(三) 差别剪接 (differential splicing)

初级转录产物剪切后的连接方式通常是一个外显子接头的 5' 端与另一个外显子接头的 3' 端连接。但是,某些 RNA 则能以不同方式加工,产生可以编码不同蛋白质的 mRNA。这样,内含子的存在就为实现这种加工方式提供了极大的灵活性,通过不同的剪接方式,能从同一初级转录产物中产生不同的蛋白质,这种剪接方式就称为差别剪接。免疫球蛋白基因转录过程中产生不同的 mRNA,就是差别剪接。

翻 译

蛋白质合成包括三种重要成份: mRNA、tRNA 和核糖体。mRNA 携带密码信息;tRNA 可以为 mRNA 上每个密码子选择相应的氨基酸;核糖体则可以将这些成份组合在一起,提供反应的场所,并使肽键形成。核糖体是一种具有多种酶的大颗粒,它由一个大亚单位 and 一个小亚单位组成,含有约 50% 的 RNA。核糖体内有两个沟,一个沟容纳新生的多肽链,另一个沟容纳 mRNA 分子。它们分别覆盖大约 30 个氨基酸和 35 个核苷酸,核糖体上有一个专门结合已经连结了新生多肽链的 tRNA 结合位点,还有一个 tRNA—氨基酸结合位点。氨基酸的连结受肽转移酶催化,肽转移酶也需要与核糖体结合后才能起催化作用。蛋白质合成包括三个步骤:(1) 起始。核糖体与 mRNA 上专一起始位点结合,tRNA 附着于起始密码子上,翻译水平的调控通常发生在这一阶段。(2) 延伸。氨基酸不

断地加到正在延长的新生多肽链羧基末端上,直到抵达终止密码子为止。(3) 终止。翻译终止,从核糖体上释放出已完全合成的新生多肽链。起始位点的不同,能够产生三种不同的阅读框,所以 mRNA 分子的翻译方式就有三种。但是,任何一种 mRNA 分子在翻译时只能采用一种方法,即只能择取一种阅读框。这样,翻译的关键就在于起始密码子上的起始步骤准确与否,原核和真核生物中的起始密码子都是 AUG。AUG 同时也是甲硫氨酸的密码子。携带甲硫氨酸的 tRNA 是一种特定的起始因子,它可以专一性地识别 mRNA 起始位置上的 AUG,所以,凡新合成多肽链的氨基末端都是甲硫氨酸。由于在 mRNA 的其他位置上也存在编码甲硫氨酸的 AUG 密码子,因此在翻译起始时,正确地选择 AUG 起始密码子就十分重要。

起始过程是相当复杂的,它是由许多蛋白质催化的多步反应,这些催化起始反应的蛋白质称为起始因子。翻译起始需要两种信号:(1) AUG 密码子;(2) mRNA 与核糖体结合的信号。在真核生物中,5' 帽结构可能是部分识别信号。在家兔的网织红细胞中,至少已分离出 7 种起始因子。这 7 种起始因子都是 mRNA 与核糖体结合、起始 tRNA 与 mRNA 和核糖体结合以及核糖体亚单位之间的连接等反应过程必不可少的成份。在肽链延伸到一定长度时,可遇到肽链终止信号——终止密码子。终止密码子有三个,分别是 UAA, UAG 和 UGA。随着最后一个 tRNA 肽键被水解,新生的多肽链释放出来,翻译结束。这时形成的肽链称为初级翻译产物。

翻译后修饰

初级翻译产物形成后,需要进一步加工,才能使它具有完整的功能,成为成熟的翻译产物。mRNA 仅仅是规定了多肽链中氨基酸的顺序,但是多肽链分子折叠成二级和三级结构的方式则是由翻译后修饰所决定

的。

(一) 前导顺序

在初级翻译产物中,有一部分蛋白质是由细胞内向细胞外输出的,称为分泌性蛋白质;还有一部分需要进入细胞膜内,称为细胞膜蛋白质。这两类蛋白质在初级翻译产物阶段,在它们的 N 末端都有一段大约由 15—25 个氨基酸组成的顺序,称为前导顺序 (leader sequence),也称信号顺序 (signal sequence) 或信号肽 (signal peptide)。在前导顺序中部是一串具有疏水性侧链的氨基酸,如亮氨酸、异亮氨酸、缬氨酸、苯丙氨酸和色氨酸等,它们对细胞膜具有高度亲和性。前导顺序的功能就是引导新生的多肽链穿过细胞膜,使一部分蛋白质从细胞中分泌出去,如胰岛素、血清清蛋白、抗体和消化道酶等;使另一部分蛋白质停泊在细胞膜的外表面,如细胞表面的组织相容性抗原。现在认为,前导顺序引导新生的多肽链穿膜的作用是通过一种叫做信号识别颗粒 (signal recognition particle, SRP) 来实现的。SRP 是一种小分子 RNA,它既能与新生多肽链的 N 末端区结合,又能与膜上的受体结合。通过这种方式,使新生肽链穿膜。一旦前导顺序穿膜,位于膜上的信号肽酶就会将前导顺序去除。也有人认为,可能有多种酶参与前导顺序的去除过程。

(二) 其他修饰

有些蛋白质是由几条多肽链组成的多聚体,而在初级翻译产物中只是一条多肽链,因此在翻译后,就需要去除多肽链中的一段或几段内部顺序,以便产生两条或几条多肽链,这些多肽链再以二硫键相连接,折叠成有功能的完整蛋白质,如胰岛素的合成。

对有些蛋白质,在翻译后需加入一些化学基团。如糖蛋白,必须在翻译后加入糖基,使它进行糖化。糖化作用是在前导顺序去除后就立刻进行的。还有如蛋白质的磷酸化、乙酰化等都是翻译后修饰的重要内容。