

I, III, IV 及 V 型胶原基因结构及其表达调控顺序

杜卫东 张月娥

(上海医科大学病理学教研室 200032)

胶原是细胞外基质的主要成份,与FN、LN、PGs等一起对实质细胞起支持、连接、营养及保护等功能。目前胶原家族中至少已发现了14种以上的胶原类型,其基因分别位于不同的染色体上^[1,2]。本文就 I, III, IV 及 V 型胶原基因的研究近况作一简要概述。

1 I, III, IV 及 V 型胶原的分子及超分子结构

胶原是 α 链组成的异构或同构三聚体, I, III, V 型胶原为纤维性胶原,主干是由肽多聚体(Gly-X-Y)_n组成,其中甘氨酸占1/3,脯氨酸及羟脯氨酸约占1/5~1/4, Gly-X-Y 重复单位内无间隔成份存在。每一前胶原分子包括三部分:羧基末端、胶原主体和氨基末端,两端常还各有一个球形结构的端肽(前体特异区)^[3,4],一经被分泌到细胞外,该端肽常由特异性肽酶加以水解^[2]。稳定及适当折叠三螺旋结构的形成取决于 α 链内甘氨酸及Y部位羟脯、羟赖氨酸含量的多少。胶原分子内或分子间一般是通过赖氨酸氧化反应及醇醛或醛胺反应以多种共价键而相互交联的,电镜下 I, III 及 V 型胶原为杆状,呈纵向错位排列,每分子长4.4D(300nm),由一段67nm或其倍数的D区相互重叠,未重叠部分的两端留有一0.6D(40nm)的缺口。IV型为基膜胶原,结构上与纤维性胶原大致相同,约75%的氨基酸序列为Gly-X-Y的重复单位,基本上仍为一杆状结构。氨基末端有一长约30nm的7S区,该区由半胱氨酸丰富区、三螺旋片段和短的非三螺旋片段组成。电镜下IV型胶原分子长

为400nm。通过7S区和NC1区与其它IV型胶原分子相互作用形成聚合物。氨基末端,4个IV型胶原分子通过其7S区,两两反向平行排列;羧基末端,2个IV型胶原分子通过半胱氨酸丰富区尾尾相接形成立体网架。中心三螺旋区氨基酸序列的一个最突出特点是Gly-X-Y序列的中断,这种中断序列 $\alpha 1(IV)$ 有21个, $\alpha 2(IV)$ 有23个。这种结构的存在使得IV型胶原分子有较大的易弯性,形成富有弹性的三维网架结构^[4,5]。

2 I, III, IV 及 V 型胶原的基因结构

I, III, V 型胶原基因结构从原始多细胞生物到哺乳类动物大致相同,其DNA序列是由50个以上的外显子构成。所有外显子的长度均为9bp的倍数,如45,54,99,108,162bp,分别编码5,6,11,12和18个Gly-X-Y三螺旋区。编码胶原主体三螺旋区的基因结构一般包括44个外显子,其中23个长54bp,8个108bp,1个162bp,5个45bp,5个99bp。三螺旋的始端和终端分别由一个连接外显子所编码,这个连接外显子也编码端肽和前肽序列。由于这两个连接外显子编码Gly-X-Y重复序列数量上的不同,使得肽链三螺旋区的长度略有差异(1014~1029个氨基酸)。I, III, V 型胶原基因外显子均准确地从Gly密码子起始,到Y部位的氨基酸终止,故三种胶原基因外显子的长度及排列高度一致。但也有例外:① $\alpha 1(I)$ 胶原基因中第33/34位点上的一个108bp外显子取代了其它胶原基因相应位点上2个55bp的外显子;②人和小鼠 $\alpha 1(III)$ cDNA序

列近三螺旋氨基末端多了一个Gly-X-Y重复序列^[6],Ⅳ型胶原外显子的大小变化较大,这可能是由于:①Gly-X-Y重复序列中有许多间隔;②三螺旋外显子上存在有割裂密码子等原因所造成。 $\alpha 1(\text{Ⅳ})$ 基因羟基末端的2/3序列中外显子几乎都含有割裂密码子,但在氨基末端7S区的三螺旋基因结构中,外显子无中断或割裂密码子,并严格遵循9bp法则^[7,8]。同样, $\alpha 2(\text{Ⅳ})$ 链羟基末端外显子的组成方面亦有较大差异。

不同种属不同类型胶原之间C前肽具有高度保守性。C前肽为球状结构,其中有C-蛋白酶切割位点和N-连接碳水化合物附着位点^[2]。C前肽内半胱氨酸残基的空间位置在不同前胶原链之间保持恒定^[7],从C前肽氨基末端开始,头4个半胱氨酸(C1~C4)参与链内二硫键的形成,C5~C8参与链间二硫键的形成,但 $\alpha 2(\text{Ⅰ})$ 缺乏C₂, $\alpha 2(\text{Ⅴ})$ 缺乏C3^[9]。C1~C4内半胱氨酸残基的量决定了 α 链形成同构三聚体和(或)异构三聚体^[7]。形成异构三聚体者,如 $\alpha 2(\text{Ⅰ})$ 、 $\alpha 2(\text{Ⅴ})$,链内二硫键有3个半胱氨酸残基;形成同构三聚体者,如 $\alpha 1(\text{Ⅰ})$ 、 $\alpha 1(\text{Ⅲ})$,则有4个半胱氨酸残基。C前肽由4个外显子编码(第49~52外显子),第51~52外显子长度相同,故C前肽的长度主要是由第49和第50外显子来决定的。连接外显子(第49外显子)编码三螺旋末端、端肽和C前肽的起始部,使球状结构与 α 链序列间的羟基末端相移行^[9]。三螺旋序列长度(45~63bp)上的差异认为主要是由于连接外显子的不同所造成的。

与C前肽相比,N前肽的长度和结构存在较大差异。N前肽一般由4个片段组成二信号肽Cys丰富球状区、短三螺旋区和短的球状区。 $\alpha 1(\text{Ⅰ})$ 、 $\alpha 1(\text{Ⅲ})$ 和 $\alpha 2(\text{Ⅴ})$ 链这4个片段呈纵向依次排列^[9]。 $\alpha 1(\text{Ⅰ})$ 和 $\alpha 1(\text{Ⅲ})$ 基因的第2外显子编码56~71个Cys残基的Cys丰富区。 $\alpha 2(\text{Ⅰ})$ 基因该

区缺乏,其信号肽直接与短三螺旋区相连,Cys丰富区功能不详。N前肽的三螺旋区长短不一, $\alpha 1(\text{Ⅲ})$ 有39个氨基酸,而 $\alpha 2(\text{Ⅴ})$ 有79个氨基酸。另外 $\alpha 2(\text{Ⅴ})$ 的N前肽三螺旋区还存在2个间断区。基因组编码N前肽的结构变化较大,不同胶原基因N前肽的外显子数量不一,一般可有5~8个^[9,10]。 $\alpha 1(\text{Ⅲ})$ 和 $\alpha 2(\text{Ⅰ})$ 基因N前肽内三螺旋区由2个外显子编码, $\alpha 1(\text{Ⅰ})$ 基因还包括另一个36bp的外显子。在上述外显子中唯一保持恒定的多为第6外显子,该外显子含有N前肽与三螺旋区相连接的序列^[9]。N前肽的功能目前不太清楚,有人认为N前肽在纤维发生和胶原生物合成的反馈机制方面起重要调节作用^[2],它能够决定胶原纤维的直径或在翻译水平上调节胶原的合成。 $\alpha 1(\text{Ⅳ})$ 和 $\alpha 2(\text{Ⅳ})$ 基因其5'端差异较大, $\alpha 1(\text{Ⅳ})$ 第1内含子较大,>30kb, $\alpha 2(\text{Ⅳ})$ 则较小,仅305bp,两者5'端非翻译序列均可形成茎环结构,但几乎无同源性; $\alpha 2(\text{Ⅳ})$ 非翻译片段较长(281bp),ATG起始密码子位于第2外显子, $\alpha 1(\text{Ⅳ})$ 非翻译片段较短(141bp),ATG位于第1外显子。非翻译区结构上的这种差异能影响RNA的稳定性以及翻译效率,提示该区具有重要的调控作用。人和小鼠 $\alpha 1(\text{Ⅳ})$ 和 $\alpha 2(\text{Ⅳ})$ 基因是以头对头的方式排列^[11],中间隔以130bp的基因间区,该间隔序列具有高度保守性,同源性达88%,当 $\alpha 1(\text{Ⅳ})$ 和 $\alpha 2(\text{Ⅳ})$ 基因线性化后,可呈明显的对称性。

3 I, Ⅲ, Ⅳ, Ⅴ型胶原基因表达的调控元件

鸡 $\alpha 1(\text{Ⅰ})$ 胶原基因分析发现在启动子和第1内含子5'端常有一些少见的DNA序列:一个位于从转录起始点-181到5'端,覆盖570bp的范围,为25个拷贝的23bp长的前后重复序列,其功能不详。另一个位于第1内

含子内,为2个S1核酸酶高敏感区,由1个“卫星样”重复序列构成,这两个S1核酸酶高敏感位点于3'端到第1外显子的250~350核苷酸范围内,它相当于鼠 $\alpha 1(I)$ [12]、 $\alpha 2(I)$ [13]和鸡 $\alpha 2(I)$ [13]基因DNase I高敏感区,该位点的缺失可使鼠 $\alpha 1(I)$ 基因转录减少20~100倍。人类 $\alpha 1(I)$ 基因则无这一位点。研究发现 $\alpha 1(I)$ 基因还有以下特点:① $\alpha 1(I)$ 基因第1内含子+820~+1093内有一个增强子,这一增强子紧靠能取消抑制效应的序列(+544~+854) [14,15];② $\alpha 1(I)$ 启动子-625~-161序列似乎为抑制效应所必需 [16];③小鼠 $\alpha 1(I)$ 启动子-222~3700序列及第1内含子内含有几个正性和负性调节元件:+550~+950为负性,-222~+1为阳性,+950~+1200和+1200~+1400为弱正性或负性 [15,16],这些序列发挥刺激还是抑制效应主要取决于不同细胞内DNA结合蛋白的含量多少。 $\alpha 2(I)$ 基因的调控机制与 $\alpha 1(I)$ 相似,小鼠 $\alpha 2(I)$ 基因在与转录起始位点有关的-240~+110内有一个DNase I高敏感区 [17];第1内含子+418~+1524有1个强的增强子结构 [18],该增强子具有细胞特异性,能限制其在NIH3T3细胞内的表达,其作用与方向和位置无关;把小鼠 $\alpha 2(I)$ 启动子-2000~+54的序列与一标记基因融合,含有该嵌合基因的小鼠即能表达这一标记基因。与 $\alpha 1(I)$ 和 $\alpha 2(I)$ 基因相比 $\alpha 1(II)$ 5'端含有散在的同源区,最主要的同源序列是靠近转录起始位点处1个50bp的反向重复片段,它可以形成茎环结构 [19],这一结构影响着 $\alpha 1(I)$ 、 $\alpha 2(I)$ 及 $\alpha 1(II)$ mRNA的转录效率。小鼠 $\alpha 1(II)$ 启动子缺失分析表明 $\alpha 1(II)$ 和I型胶原基因的调控机制有所不同, $\alpha 1(II)$ 启动子受位于-80~-350区域内的负性顺式调节元件的调控, $\alpha 1(II)$ 启动子头80bp足以使其高度表

达,但 $\alpha 2(I)$ 启动子被切除到约100bp时,可导致其表达的降低 [20]。比较 $\alpha 2(I)$ 和 $\alpha 1(II)$ 启动子头80bp和104bp顺序时发现在TATA Box周围有一高度保守区; $\alpha 2(I)$ -80有一CCAAT序列, $\alpha 1(II)$ 则缺如; $\alpha 1(II)$ 在-40有一NF1结合位点,在-52有一SV40增强子核心序列,而这些序列在 $\alpha 2(I)$ 头104bp内缺如 [20,21]。小鼠 $\alpha 1(II)$ 启动子内有2个正性顺式调节元件:①Jun/AP1结合位点,位于-122~-106;②DNA连接蛋白特异结合位点,位于-83~-61 [22]。由于 $\alpha 1(II)$ 基因缺乏与CBF结合的位点,这一DNA连接因子的存在,可类似CBF的功能,在 $\alpha 1(II)$ 基因转录的调控中起重要作用。小鼠 $\alpha 1(I)$ 和 $\alpha 2(I)$ 基因转录起始位点上游区存在着2个反式作用因子的特异结合点:①CCAAT结合因子(CBF),该因子为异构二聚体,A因子Mr.40 000~42 000,B因子Mr.34 000,二者均为DNA结合所必需。CBF可以与小鼠 $\alpha 1(I)$ -96~-100及 $\alpha 2(I)$ -80~-84内的CCAAT Box结合,可以保护约26bp [23];②核因子(NF1)与CBF不同,可以与小鼠 $\alpha 2(I)$ 启动子-315~-295 [24]及 $\alpha 1(II)$ -40结合,NF1及CBF可直接参与稳定态mRNA的形成。小鼠和人 $\alpha 1(IV)$ 和 $\alpha 2(IV)$ 基因的转录是受同一双向启动子启动的,这一双向启动子被-130bp的序列割裂,其内无TATA Box。IV型胶原基因的增强子位于第1内含子内,长约200bp,由于这一增强子内含有5'-CGCCATCA-3'序列,与cAMP反应元件(CRE)5'-TGACGTCA-3'有同源性,所以,IV型胶原基因增强子可能具有微弱的CRE结合位点效应。 $\alpha 2(IV)$ 基因转录起始点-106有CAATT Box,在小鼠和人均具有高度保守性。用该双向启动子构件转染未分化T9细胞时,常可表达高CAT活性,用5'氟胞苷处理后则CAT活性最大,提示甲

基化对Ⅳ型胶原基因活性的调控有重要作用。甲基化影响基因转录的机制认为可能与DNA构型变化^[26]和(或)甲基化作用引起反式作用因子的效应改变^[26]有关。甲基化的部位 $\alpha 2(\text{I})$ 位于第3内含子, $\alpha 1(\text{I})$ 则位于转录起始位点的上游区,但也有人认为在双向启动子内,因为该区域有较多的CpG,有利于甲基化的形成。小鼠和人双向启动子CpG的含量分别为23%和30%。

另外小鼠和Ⅳ型胶原基因还含有能与反式作用因子SP1结合的GC Box,该Box亦具有双向性。小鼠有2个GC Box,人则有1个,这可能是由于人5'-CGGCGG-3'序列位于鼠SP1结合位点的相同部位,它可以与SP1结合位点结合而发挥效应。NF1结合在小鼠双向启动子区域的2个GC Box上,推测这2个SP1结合位点亦参与Ⅳ型胶原基因转录的调节机制。

参 考 文 献

- 1 Miller EJ, Gay S. In: Cunningham LW, et al. *Methods in enzymes in enzymology*. Vol 144. Orlando, Florida: Academic Press, 1987, 3~74
- 2 Bornstein P, Sage H. Structurally distinct collagen types. *Annu Rev Biochem*, 1980, 49: 957
- 3 Lee ST, Smith BD, Greenspan DS. Construction of a full-length cDNA encoding human pro- $\alpha 2(\text{I})$ collagen and its expression in pro- $\alpha 2(\text{I})$ deficient W8 rat cells. *J Biol Chem*, 1988, 263: 13414
- 4 Pihlajaniemi T, Tryggvason K, Myers JC, et al. cDNA clones coding for the pro- $\alpha 1(\text{I})$ chain of human type Ⅳ procollagen reveal an unusual homology of amino acid sequences in two halves of the carboxyl-terminal domain. *J Biol Chem*, 1985, 260: 7681
- 5 Hofmann H, Voss T, Kuhn K. Localization of flexible sites in thread-like molecules from electron micrographs; comparison of interstitial, basement membrane and intima collagen. *J Mol Biol*, 1984, 172: 325
- 6 Ala-Kokko L, Kontusaari S, Baldwin CT, et al. Structure of cDNA clones coding for the entire pro- $\alpha 1(\text{I})$ chain of human type Ⅲ procollagen; differences in protein structure from type Ⅰ procollagen and conservation of codon preferences. *Biochem J*, 1989, 260: 509
- 7 Soyninen R, Huotari M, Ganguly A, et al. Structural organization of the gene for the $\alpha 1$ chain of human type Ⅳ collagen. *J Biol Chem*, 1989, 264: 13565
- 8 Killen PD, Burbelo P, Sakurai Y, et al. Structure of the amino-terminal portion of the murine $\alpha 1(\text{IV})$ collagen chain and the corresponding region of the gene. *J Biol Chem*, 1988, 263: 8706
- 9 Ramirez F. Fibrillar collagen genes; structure and expression in normal and disease states. *Ann NY Acad Sci*, 1990, 580: 74
- 10 Vuorio E, de Crombrughe B. The family of collagen gene. *Annu Rev Biochem*, 1990, 59: 839
- 11 Kaytes P, Linda W, Theriault N, et al. Head-head arrangement of murine type Ⅳ collagen genes. *J Biol Chem*, 1988, 263: 19274
- 12 Breindl M, Harbers K, Jaenisch R. Retrovirus-induced lethal mutation in collagen Ⅰ gene of mice is associated with an altered chromatin structure. *Cell*, 1984, 38: 9
- 13 Mckeon C, Schmidt A, de Crombrughe B. A sequence conserved in both chicken and mouse $\alpha 2(\text{I})$ collagen promoter contains sites sensitive to S1

- nuclease. *J Biol Chem*, 1984, 259 : 6636
- 14 Rossouw CMS, Vergeer WP, du Plooy SJ, et al. DNA sequences in the first intron of the human pro- $\alpha 1(I)$ collagen gene enhance transcription. *J Biol Chem*, 1987, 262 : 15151
- 15 Bornstein P, McKay J. The first intron of the $\alpha 1(I)$ collagen gene contains several transcriptional regulatory elements. *J Biol Chem*, 1988, 263 : 1603
- 16 Bornstein P. Interactions between the promoter and first intron are involved in transcriptional control of $\alpha 1(I)$ collagen gene expression. *Mol Cell Biol*, 1988, 8 : 4851
- 17 Liao G, Szapary D, Setoyama C. Restriction enzyme digestion identifies discrete domains in the chromatin around the promoter of the mouse $\alpha 2(I)$ collagen gene. *J Biol Chem*, 1986, 261 : 11362
- 18 Rossi P, de Crombrughe B. Identification of a cell-specific transcription enhancer in the first intron of the mouse $\alpha 2$ (type I) collagen gene. *Proc Natl Acad Sci USA*, 1987, 84 : 5590
- 19 Yamada Y, Mudryj M, de Crombrughe B. A uniquely conserved regulatory signal is formed around the translation initiation site in three different collagen genes. *J Biol Chem*, 1983, 258 : 14914
- 20 Schmidt A. Transcriptional control of the mouse $\alpha 2(I)$ collagen gene: functional deletion analysis of the promoter and evidence for cell-specific expression. *Mol Cell Biol*, 1986, 6 : 347
- 21 Mudryj M, de Crombrughe B. Deletion analysis of the mouse $\alpha 1(II)$ collagen promoter. *Nuclear Acid Res*, 1988, 16 : 7513
- 22 Ruteshouser EC, de Crombrughe B. Characterization of two distinct positive cis-acting elements in the mouse $\alpha 1(II)$ collagen promoter. *J Biol Chem*, 1989, 264 : 13740
- 23 Hatamochi A, Paterson B, de Crombrughe B. Differential binding of a CCAAT DNA binding factor to promoters of the mouse $\alpha 2(I)$ and $\alpha 1(II)$ collagen genes. *J Biol Chem*, 1986, 261 : 11310
- 24 Oikarinen J, Hatamochi A, de Crombrughe B. Separate binding sites for nuclear factor 1 and a CCAAT DNA binding factor in the mouse $\alpha 2(I)$ collagen promoter. *J Biol Chem*, 1987, 262 : 11064
- 25 Keshet I, Lieman-Hurwitz J, Cedar H. DNA methylation effects the formation of active chromatin. *Cell*, 1986, 44 : 535
- 26 Becker PB, Ruppert S, Schutz G. Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. *Cell*, 1987, 51 : 435

(收稿: 1993-09-23 修回: 1994-11-28)

(本文编辑: 程春开)

(上接第105页)

- atit. *Gastroenterol*, 1987, 92(2) : 304
- 18 Sekar B, Thyagarajan SP, Thirunalasundari T, et al. Non-organ specific and organ specific autoantibodies in HBV associated glomerulonephritis. *Indian J Med Res*, 1989, 89(7) : 221
- 19 Johnson RJ, Couser WG. Hepatitis B infection and renal disease, clinical immunopathogenetic and therapeutic considerations. *Kidney Int*, 1990, 37(2) : 663

(收稿: 1993-02-15)

(本文编辑: 程春开)