

鸡局部黏着斑激酶基因 5'UTR 及启动子的克隆鉴定

李淑锋^{1,2}, 王利祥¹, 张大鹏¹, 华子春^{1*}

(1. 南京大学医药生物技术国家重点实验室, 江苏 南京 210093; 2. 东南大学基础医学院, 江苏 南京 210009)

摘要: 为了研究鸡局部黏着斑激酶 (FAK) 基因表达调控的分子机制, 运用 cDNA 末端快速扩增技术确定了鸡胚成纤维细胞 FAK 基因的转录起始位点, 并发现 FAK 基因 mRNA 的 5'非翻译区 (5'UTR) 存在 4 种剪切形式。分析这 4 种剪切形式的 5'端序列, 发现它们都不影响蛋白的编码, 但可能影响 mRNA 的翻译效率。另外, 将启动子序列 935 bp 片段克隆构建到荧光素酶报告基因表达载体 pGL3-Basic 中, 转染鸡胚成纤维细胞, 测定启动子的转录活性。通过启动子序列系列缩短的克隆, 将核心启动子区定位在转录起始位点上游 -662 至 +7 的 669 bp 片段内, 该启动子的典型特征是: 没有 TATA 盒, 而富含“GC”碱基。以上结果为 FAK 基因的表达调控研究提供了分子基础。

关键词: 局部黏着斑激酶基因 (FAK); 5'非翻译区 (5'UTR); 可变剪切; cDNA 末端快速扩增技术

中图分类号: Q752 **文献标识码:** A **文章编号:** 1000-2030 (2007) 04-0102-06

Cloning and characterization of 5'UTR and promoter region of chick focal adhesion kinase gene

LI Shu-feng^{1,2}, WANG li-xiang¹, ZHANG Da-peng¹, HUA Zi-Chun^{1*}

(1. State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing 210093, China;

2. College of Basic Medical Science, Southeast University, Nanjing 210009, China)

Abstract: To understand the mechanism of chick focal adhesion kinase gene (FAK) expression and regulation, the transcription start sites and four kinds alternatively spliced 5' end (5'untranslated region, 5'UTR) of FAK mRNA in chick embryo fibroblast cells were identified by RACE. The analysis of the 5' end mRNA sequences showed that they all would not affect the encoded protein sequence, they may influence the mRNA translation efficiency. The 935 bp fragment with FAK promoter was also cloned and placed upstream of the luciferase reporter gene in pGL3-Basic vector and transfected into chick embryo fibroblast cells. Serial deletion constructs revealed that a 669 bp region (-662 to +7) was required for the maximal FAK promoter activity. The promoter sequence showed high "GC" content and lacked TATA box. These results will provide a molecular basis for understanding the mechanism of FAK gene expression.

Key words: focal adhesion kinase gene (FAK); 5'untranslated region (5'UTR); alternative splicing; RACE

FAK 是一种相对分子质量为 125 000 的非受体酪氨酸激酶, 是细胞外基质的下游信号激酶, 主要参与细胞和 ECM 的局部黏附, 是整合素介导的细胞外基质分子向细胞内传递的重要信号蛋白^[1-2], 所以命名为局部黏着斑激酶 (focal adhesion kinase, FAK)。FAK 既可以通过整合素结合和传播来自 ECM 的信号, 也可以通过特殊受体, 结合和传递来自可溶性生物活性因子的信号。FAK 与一些信号及细胞骨架分子结合后, 即被活化和磷酸化。这些分子包括整合素、Src 激酶羧基端、踝蛋白、 α 辅肌动蛋白等。从鸡、鼠、人获得的 FAK 的 cDNA 中推导出的蛋白质顺序发现, FAK 有高达 95% ~ 97% 的同源性, 表明 FAK 基因在进化中的高度保守^[3-5]。FAK 蛋白可分为 3 个功能区: N 端区、激酶区和 C 端区^[6]。

FAK 蛋白在肿瘤发生过程中也扮演了非常重要的角色^[7-9], 因此认识 FAK 的表达调控机制不仅有助于理解其进化上的演变, 也可以有助于了解癌症的发生机制, 并为肿瘤治疗提供潜在的靶点。

自 FAK 发现以来, 对它的生物活性、作用机制等已有较多研究, 但对 FAK 基因的表达调控特点知之甚少。除了已鉴定出人的 FAK 基因的启动子外^[10], 没有其他物种的任何调控研究。因此, 笔者试图

收稿日期: 2007-06-14

基金项目: 国家自然科学基金项目 (30330530, 30425009); 教育部基金 (TRAPOYT, SRFDP 20030284040)

作者简介: 李淑锋, 博士, 讲师, E-mail: shufengli3000@yahoo.com.cn.* 通讯作者: 华子春, 教授, 主要从事分子生物学研究,

E-mail: huazc@nju.edu.cn。

从基因转录及转录后加工水平来研究 *FAK* 基因在鸡中的表达调控机制。本试验采用 5'RACE (cDNA 末端快速扩增 rapid amplification of cDNA ends, RACE) 方法, 在鸡胚成纤维细胞中首先确定其转录起始位点, 进而克隆了 *FAK* 基因的 5'非翻译区 (5'untranslated region, 5'UTR), 并通过启动子序列的荧光素酶报告基因表达系统, 确定了核心启动子区, 为进一步探讨该基因表达调控的机制奠定基础。

1 材料与方 法

1.1 鸡胚成纤维细胞 CEF 的制备

根据文献 [4], 取 10 日龄鸡胚, 无菌取出鸡胚, 弃去眼睛和内脏团, 用灭菌剪刀将鸡胚剪成 1 mm^3 的碎块, 胰酶消化, 分散细胞, 再用灭菌纱布过滤细胞悬液, 加入含 10% 小牛血清的 DMEM, 置 $37\text{ }^\circ\text{C}$ 5% CO_2 的培养箱中培养。

1.2 RNA 的提取

将收获的细胞用 Trizol 试剂 (Invitrogen 公司) 提取细胞总 RNA。当裂解完全后, 将裂解液移至 1.5 mL 离心管中, 每管 1 mL。在上述离心管中加入 200 μL 氯仿, 振荡混匀后室温下静置 2 min; 离心取水相移至另一支 1.5 mL 离心管中, 加入 500 μL 异丙醇, 混匀后室温下静置 10 min; 离心弃上清液, 加入 75% 乙醇漂洗沉淀, 离心弃上清液, 沉淀物干燥后溶于适量焦碳酸二乙酯处理的超纯水中, 取适量经稀释后测定 A_{260} 与 A_{280} , 计算 A_{260} 与 A_{280} 的比值, 确定 RNA 的含量。用甲醛变性凝胶电泳鉴定 RNA 的质量。样品 $-70\text{ }^\circ\text{C}$ 保存备用。

1.3 5'RACE 克隆 5'上游序列与产物的克隆测序

采用 BD SMART RACE cDNA Amplification kit (Clontech, CA) 克隆 *FAK* 基因 5'端 cDNA。根据已发表的鸡的 *FAK* 基因 5'上游序列^[4], 设计 2 个下游引物 GSP₁ 和 GSP₂, GSP₁ 和 GSP₂ 都位于编码区。GSP₁: 5'-tagtgaaaaaccttaggactgct-3', GSP₂: 5'-gtccacaatctctg tattatgct-3'。以总 RNA 为模板, 以 GSP₂ 为引物进行 cDNA 合成, 再进行巢式 PCR 扩增反应 (引物为 GSP₁ 和试剂盒中通用引物)。将扩增出的片段与 pGEM-T Easy 载体连接, 转化 Top10 感受态细菌, 接种于含 X-Gal、IPTG、氨苄西林的固体培养皿, $37\text{ }^\circ\text{C}$ 过夜。挑取白色菌落, 用 PCR 鉴定。鉴定的上游引物: 5'-tgggctaagaatgatgacagatacc-3', 下游引物为 GSP₁。选阳性克隆送 Invitrogen 公司测序。利用 NCBI 公共数据库, 将所得序列进行 BLAST 分析比对, 确定 mRNA 5'端第 1 个核苷酸, 即转录起始位点。

1.4 启动子-Luciferase 报告质粒的构建

PCR 扩增 CEF 细胞 *FAK* 启动子区 $-928/+7$, 上游引物加 *Kpn* I 酶切位点, 下游引物加 *Sma* I 酶切位点。将酶切 PCR 产物, 克隆到荧光素酶表达载体 (Promega) pGL3-Basic 上。其他 *FAK* 启动子截短片段的报告质粒插入片段 ($-662/+7$, $-351/+7$, $-208/+7$, $-168/+7$) 以类似的方法产生, 使用不同的上游引物 (加 *Kpn* I 酶切位点) 和共同的下游引物 (加 *Sma* I 酶切位点) (引物序列见结果图)。构建的启动子荧光素酶报告质粒分别命名为 p-928、p-662、p-351、p-208 和 p-168。所有克隆经测序验证正确。

1.5 细胞培养和转染

细胞生长于含有 10% 小牛血清的 DMEM (Dulbecco's modified Eagle's medium) 培养液, 在含有 5% 二氧化碳细胞培养箱中 $37\text{ }^\circ\text{C}$ 进行培养。以 $2 \times 10^5\text{ mL}^{-1}$ 的密度接种细胞至 24 孔板。生长 24 h, 至细胞密度达到 80% 时, 利用 Lipofectamine 2000 转染试剂共转染启动子荧光素酶报告基因表达质粒 (表达 firefly luciferase) 和内参 pRL-TK 质粒 (表达 renilla luciferase)。每种样品设 3 个平行试验。24 h 后收集细胞, 用于荧光素酶活性分析。

1.6 双色荧光素酶活性分析

含有 *FAK* 启动子不同长度片段的报告质粒和内参 pRL-TK 质粒被瞬时转染到 CEF 细胞, 24 h 后收集细胞, 在冰上用细胞裂解液裂解细胞 40 min; 离心取上清液用于荧光素酶活性分析, 分别检测 firefly luciferase 活性和 renilla luciferase 活性。双色荧光素酶试剂盒购自 Promega 公司, 使用 Turner Biosystems 照度计测量荧光强度。启动子的相对活性用 firefly luciferase 活性与内参 renilla luciferase 活性标准化后, 再以相对于标准化后的 pGL3-Basic 的活性来表示。

2 结果与分析

2.1 细胞总 RNA 的提取

采用 Trizol 试剂提取 CEF 细胞总 RNA 的 A_{260}/A_{280} 为 1.95, 取 1 μg RNA 进行甲醛变性凝胶电泳, 显示清晰的 28S、18S 和 5.8S 3 条 RNA 带, 说明所获 RNA 样品完全能够满足制备 RACE cDNA 的需要。

2.2 5'RACE 法获取 FAK 基因转录起始位点和 5'端 cDNA 序列

用合成的特异引物和 5'RACE 试剂盒, 仅一轮扩增即从 CEF 细胞中获取了约 250 bp 左右的 5'RACE 产物。将此产物直接克隆到 pGEM-T 载体中, 随机挑取 14 个白色克隆进行电泳鉴定 (图略)。对所选克隆进行测序, 并将序列与 NCBI 数据库数据进行对比分析, 结果见图 1。将结果与 FAK 5'侧翼基因组 DNA 序列 (ref | NC_ 006089.1) 比较, 发现存在 4 种 5'端 mRNA 形式, GenBank 的登录号分别为 AY916079、AY916080、AY916081 和 AY916082。序列分析发现这 4 种 5'端前导序列都不影响蛋白的编码框。分别命名这 4 种 mRNA 形式的 5'端非翻译区为 α -UTR、 β -UTR、 γ -UTR 和 δ -UTR。为了描述的方便及与以前的命名一致, 把含有标准阅读框起始密码子 (ATG) 的外显子命名为 exon1, 上游非编码外显子按其在基因组的分布位置, 分别命名为 exon -1、exon -2、exon -3、exon -4 和 exon -5, 相应的各个外显子片段的大小和内含子大小显示在图 1-A 中。 α -UTR、 β -UTR 和 γ -UTR 这 3 种 UTR 是由我们新发现的外显子 exon -1、exon -3、exon -4 和 exon -5 可变剪切而组成, 而且转录的起始分别从 exon -5a、exon -5b、exon -5c 开始。含 δ -UTR 的 mRNA 比以前报道的鸡 FAK mRNA (NM_ 205435) 稍长, 从 exon -2 外显子转录起始。比对结果还发现: 外显子 exon -1、exon -2、exon -3、exon -4 和鸡 2 号染色体 FAK

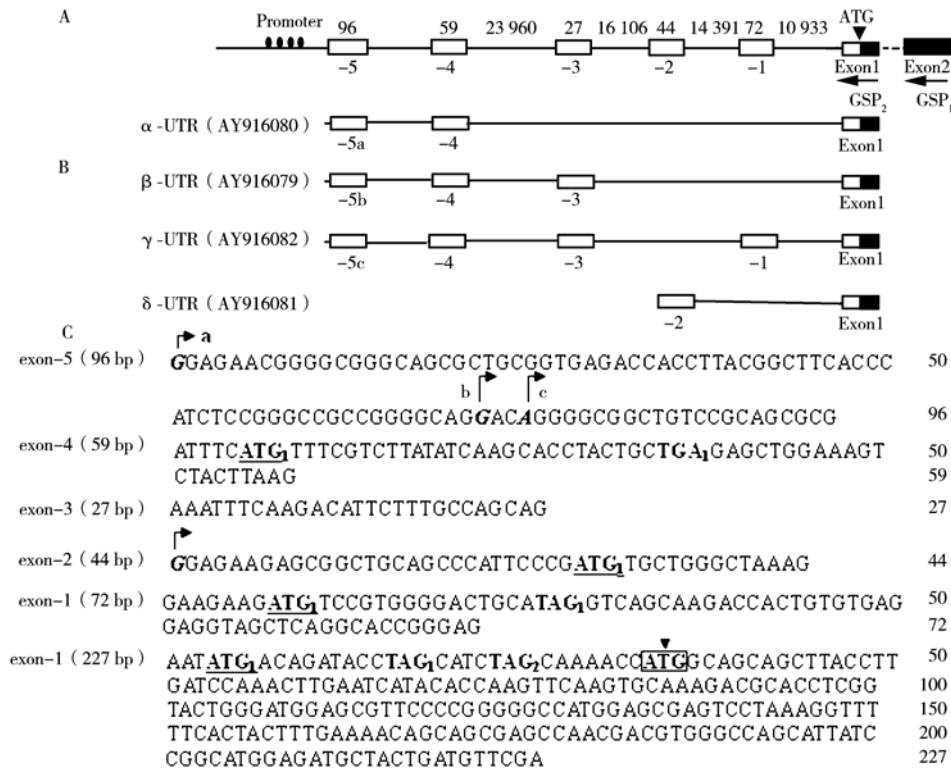


图 1 鸡 CEF 细胞中 FAK 基因 mRNA 的 5'端序列

Fig. 1 Sequences of 5' end of FAK mRNA in chick CEF cell

A. 鸡 FAK 基因 5'端基因组结构和外显子相对位置。内含子、外显子等片段大小表示在相应示意图上方。B. FAK 基因 mRNA 的 4 种 5'端形式。C. FAK 基因 5'端外显子序列, exon -5 和 exon -2 内转录起始位点用弯箭头显示。标准阅读框的起始密码 ATG 用方框加三角箭头显示。

A. Genomic organization covering the leader sequences of the FAK gene of chick is shown. The lengths of DNA fragments are shown on top of each genomic gene structure. B. Schematic representation of the 5' end of four kinds FAK mRNA in chick CEF. C. The sequences of exon-1, exon-2, exon-3, exon-4 and exon-5. The position of the transcription initiation sites are indicated by bent arrows in exon-5 and exon-2. The standard translation start codon ATG is indicated by box and triangle arrow.

基因组 DNA 序列一致。但 exon - 5 与鸡的基因组比对没有一致序列，这可能是由于鸡的基因组数据库不完善，部分区域还未测通的原因。推测 exon - 5a、exon - 5b、exon - 5c 可能位于同一个外显子中，但从不同位点起始转录。其他所有外显子剪切位点均符合 GT/AG 的剪接法则。因 exon - 5 在基因组上的确切位置还不能确定，因而 exon - 5 和 exon - 4 之间的内含子大小也未确定，故没标示在图 1 - A 中。

2.3 启动子的克隆分析

为了确定 exon - 5 在基因组上的确切位置，对鸡 *FAK* 基因 5'端基因组未测通的区域，在两侧设计引物，以鸡 CEF 提取的基因组 DNA 为模板，进行 PCR 扩增，将 PCR 产物测序。图 2 所示结果证实了我们原来的猜想，exon - 5 确实在这个未测完全的部分，exon - 5a、exon - 5b、exon - 5c 是由同一个外显子的不同转录起点形成的。把 exon - 5 第 1 个碱基命名为 +1，即最前端的转录起始位点，其上游区域即为启动子区。exon - 5 也符合 GT/AG 的剪切法则。通过与基因组比对，进一步确定了 exon - 4 与 exon - 5 之间的内含子大小为 14 356 bp。



图 2 鸡 CEF 细胞中 *FAK* 基因上游基因组部分序列的测序与转录起始区的确定

Fig. 2 Sequencing the chick genomic DNA sequence upstream the *FAK* gene and mapping the transcription initiation region of the *FAK* gene

双下划线部分为引物序列，阴影部分的序列为原来基因组中未测完全的部分，带方框部分是 exon - 5。箭头指示转录起始位点。Double underlined are primers used for PCR. Shaded are the gap that are not shown in the NCBI database. The exon-5 is shown in box. Bent arrows indicate the transcription initiation sites.

为了鉴定 *FAK* 基因 5'端启动子区的活性，构建了 exon - 5 上游系列截短的侧翼序列荧光素酶报告质粒（图 3）。p - 662 在 CEF 细胞中显示了最高的启动子活性，启动子逐渐缩短的插入片段的荧光素酶报告质粒 p - 351、p - 208、p - 168，其荧光素酶活性逐渐减小，而延长片段的报告质粒 p - 928，其荧光素酶活性不再增加。这些结果说明 exon - 5 上游侧翼序列 - 662 到 + 7 区域对启动子的活性很关键，是核心启动子区域。利用 TFSEARCH 和 TRANSFAC 软件预测这个区域的转录因子结合位点，结果有很多转录因子 sp1、AP1 等结合位点。这个启动子不含 TATA 框，而含有 GC 丰富区，这和人 *FAK* 基因启动子是一致的。这也说明了 *FAK* 基因不仅在编码区保守，在非编码的调控区也存在保守性。

2.4 *FAK* 非编码区二级结构的预测

为探究获得的 *FAK* 基因的 4 种 mRNA 形式是否编码不同的蛋白质，分析所获得的 4 种 mRNA 的 5'端序列，发现都存在非标准的短的上游阅读框，但都不影响 exon1 上起始的标准阅读框。说明这 4 种 mRNA 的表达蛋白是一致的。接着用 RNA structure 生物软件对 5'非编码区进行二级结构预测，可以看到 5'非编码区由茎环和发夹环互相连接形成复杂的二级结构。从自由能可以看出形成的二级结构比较稳定（表 1），其中 α-UTR 和 γ-UTR 的自由能最低，二级结构最为稳定，这些茎环结构和上游阅读框的存在都不同程度地影响翻译过程中核糖体的

表 1 鸡中 *FAK* 基因 mRNA 5'非翻译区的分析

Table 1 Analysis of 5'UTR of *FAK* mRNA in chick

| 5'非翻译区 5'UTR | 长度/bp Length | 上游阅读框 uORF | 自由能/kJ · mol ⁻¹ Free energy |
|-----------------|-----------------|---------------|---|
| α-UTR | 186 | 2 | -270.9 |
| β-UTR | 143 | 2 | -124.8 |
| γ-UTR | 212 | 3 | -233.2 |
| δ-UTR | 76 | 1 | -68.7 |

顺利滑动和翻译的进行,可能在减弱翻译的效率上起精细调控作用。这些结果也说明 *FAK* 基因在转录后的可变剪切加工水平和翻译水平上都存在调控作用。

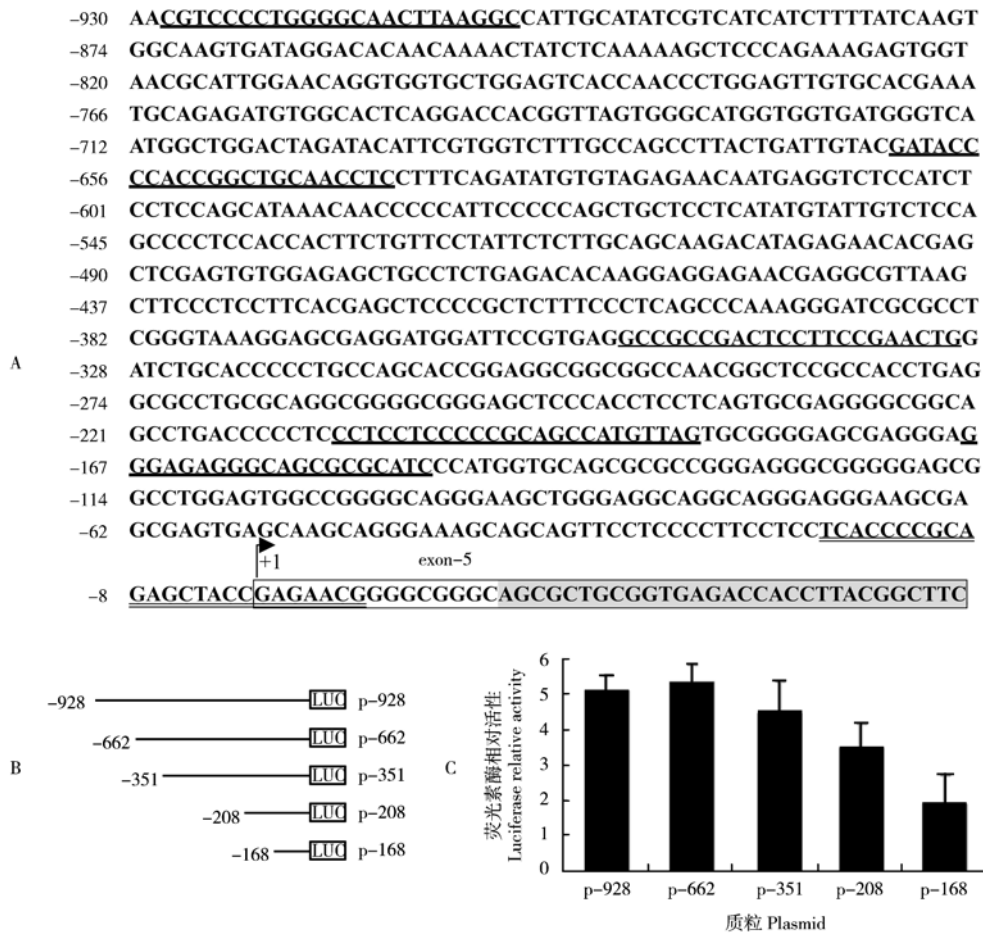


图 3 鸡 CEF 细胞中 *FAK* 基因核心启动子区的确定

Fig. 3 Identification of the core promoter region of *FAK* gene in CEF cell

A. 鸡 *FAK* 基因启动子区的序列。单下划线和双下划线分别表示构建启动子活性报告质粒所用的上游引物和下游引物,箭头指示第一个转录起始位点。B. 含有 *FAK* 启动子不同长度片段的报告质粒。C. *FAK* 启动子不同长度片段的报告质粒的相对活性。

A. The sequence of the chick *FAK* promoter region. Underline and double underline mark the primers used for promoter constructs. Bent arrow indicates the first transcription initiation site. B. Schematic maps of *FAK*-luc constructs containing various length of *FAK* promoter regions. C. The promoter activity of *FAK*-luc constructs containing various length of *FAK* promoter regions.

3 讨论

FAK 蛋白是生物体内少有的长序列蛋白,到目前为止,从人、小鼠到果蝇、酵母中均发现它的存在^[11-14],即使在进化关系较远的人和斑马鱼之间仍然具有很强的保守性^[15]。

为了研究 *FAK* 基因的表达调控机制,获取全长 cDNA 序列是功能研究的基础。RACE 技术是获得全长 cDNA 的高效、可靠的经典方法。本研究通过 RACE 技术首次克隆获得鸡的 *FAK* 基因 mRNA 的 5'端序列。本文报道的 4 种 5'端 mRNA 的序列都比原先报道的 mRNA (NM_ 205435) 的 5'端要长。

本研究确定了 *FAK* 基因 mRNA 最前端 exon -5 在基因组的位置,也随之确定了 exon -5 上游的启动子区的结构,通过启动子荧光素酶报告基因表达质粒的克隆和启动子活性的检测,证明 -662 到 +7 的区域对于启动子的活性是关键区域。鸡的启动子序列与人的启动子区域很相似,也不含 TATA 框,而富含 GC 盒,这反映了 *FAK* 转录调控的保守性。

FAK 基因 mRNA 的多种可变剪切形式和多个起始位点的存在,揭示了其复杂的转录及转录后加工机制。在本文报道的 4 种 *FAK* 基因 mRNA 形式中,3 种形式最前端都位于 exon -5 中,而第 4 种 mRNA 的 5'端则位于 exon -2 中。这 4 种 mRNA 是否都由 exon -5 上游的启动子转录,然后通过可变剪切加工

而形成成熟的 mRNA 形式呢, 还是在 exon - 2 上游还有另一个启动子, 通过双启动子来转录不同的 mRNA 呢? 这个问题仍需要进一步的研究。有趣的是, 在已报道的人 FAK 基因的 mRNA 的 5'端也存在多种可变剪切形式。说明 FAK 基因在转录后存在可变剪切的机制也是保守的。另外, 还可通过多种形式的 mRNA 上游非翻译区序列调控 FAK 基因的翻译效率^[16]。

总之, FAK 基因的表达调控体现在多个环节, 包括转录、前体的可变剪切、翻译及翻译后的磷酸化修饰。本研究对鸡胚成纤维细胞中 FAK 基因最前端核心启动子区的确定, 为以后进一步研究转录因子对鸡 FAK 基因的转录调节奠定了基础。对 4 种可变剪切形式的 5'UTR 的分析也为 FAK 翻译调控的深入研究提供了初步依据。

参考文献:

- [1] Hanks S K, Polte T R. Signaling through focal adhesion kinase [J]. *BioEssays*, 1997, 19: 137 - 145
- [2] McLean G W, Carragher N O, Avizienyte E, et al. The role of focal adhesion kinase in cancer: a new therapeutic opportunity [J]. *Nat Rev Cancer*, 2005, 5: 505 - 515
- [3] Schaller M D, Borgman C A, Cobb B S, et al. pp125FAK, a structurally distinctive protein tyrosine kinase associated with focal adhesions [J]. *Proc Natl Acad Sci USA*, 1992, 89: 5192 - 5196
- [4] Devor B B, Zhang X, Patel S K, et al. Chicken and mouse adhesion kinases are similar in structure at their amino termini [J]. *Biochem Biophys Res Commun*, 1993, 190: 1084 - 1089
- [5] Whitney G S, Chan P Y, Blake J, et al. Human T and B lymphocytes express a structurally conserved focal adhesion kinase, pp125FAK [J]. *DNA Cell Biol*, 1993, 12: 823 - 830
- [6] Xu L H, Yang X, Craven R J, et al. The COOH-terminal domain of the focal adhesion kinase induces loss of adhesion and cell death in human tumor cells [J]. *Cell Growth and Differentiation*, 1998, 9: 999 - 1005
- [7] Sood A K, Coffin J E, Schneider G B, et al. Biological significance of focal adhesion kinase in ovarian cancer: role in migration and invasion [J]. *Am J Pathol*, 2004, 165: 1087 - 1095
- [8] Hess A R, Postovit L M, Margaryan N V, et al. Focal adhesion kinase promotes the aggressive melanoma phenotype [J]. *Cancer Res*, 2005, 65: 9851 - 9860
- [9] McLean G W, Komiyama N H, Serrels B, et al. Specific deletion of focal adhesion kinase suppresses tumour formation and blocks malignant progression [J]. *Genes Dev*, 2004, 18: 2998 - 3003
- [10] Golubovskaya V, Kaur A, Cance W. Cloning and characterization of the promoter region of human focal adhesion kinase gene: nuclear factor kappa B and p53 binding sites [J]. *Biochim Biophys Acta*, 2004, 1678: 111 - 125
- [11] Burgaya F, Girault J A. Cloning of focal adhesion kinase, pp125FAK, from rat brain reveals multiple transcripts with different patterns of expression [J]. *Mol Brain Res*, 1996, 37: 63 - 73
- [12] Fox G L, Rebay I, Hynes R O. Expression of DFak56, a *Drosophila* homolog of vertebrate focal adhesion kinase, supports a role in cell migration *in vivo* [J]. *Proc Natl Acad Sci USA*, 1999, 96: 14978 - 14983
- [13] Palmer R H, Fessler L I, Edeen P T, et al. DFak56 is a novel *Drosophila melanogaster* focal adhesion kinase [J]. *J Biol Chem*, 1999, 274: 35621 - 35629
- [14] Schaller M D. Biological signals and biological responses elicited by the focal adhesion kinase [J]. *Biochim Biophys Acta*, 2001, 1540: 1 - 21
- [15] Henry C A, Crawford B D, Yan Y L, et al. Roles for zebrafish focal adhesion kinase in notochord and somite morphogenesis [J]. *Developmental Biology*, 2001, 240: 474 - 487
- [16] Singh S, Bevan S C, Patil K. Extensive variation in the 5' UTR of Dicer mRNAs influences translational efficiency [J]. *Biochem Biophys Res Commun*, 2005, 335: 643 - 650

责任编辑: 周广礼