

DOI: 10.3724/SP.J.1005.2011.00337

# SV40PolyA 顺式活化基因元件中不完整茎环结构的发现和序列研究

王红钢<sup>1,2</sup>, 马欢<sup>1</sup>, 李珠<sup>3</sup>, 张彬<sup>3</sup>, 景向阳<sup>1</sup>, 张媛<sup>1</sup>, 吕占军<sup>1</sup>

1. 河北医科大学实验动物学部遗传研究室, 河北省实验动物重点实验室, 石家庄 050017;

2. 河南大学医学院生物化学与分子生物学教研室, 开封 475004;

3. 河北医科大学第一医院呼吸内科, 石家庄 050031

**摘要:** 研究室的前期工作发现, *Alu* 串联序列插入 pEGFP-C1 质粒的 *GFP* 基因下游, 瞬时转染 HeLa 细胞抑制 *GFP* 基因表达, 2F2R(来自 SV40PolyA 反序 5'端的第 2 个 60 bp)插入 *GFP* 和 *Alu* 串联序列之间可以解除 *Alu* 序列对 *GFP* 基因的抑制作用。文章通过删减 2F2R 发现, 45R(2F2R 5'端的 45 bp)、30R 和 22R 可以活化基因, 且二串联体活化基因作用高于单体。Secloop(2F2R 近中部的 22 bp)和 Poly4(2F2R 3'端的 30 bp)不能活化基因。30R 与 Poly4 用 9 碱基连接形成 30R-Poly4, 其活化基因作用低于 2F2R, 两个 22R 之间连接碱基数对活化 *GFP* 基因作用没有明显的影响。22R(5'-GTGAAAAAATGCTTTATTTGT-3')含有不完整的回文序列, 可以形成不完整的茎环结构, 包括一个 3 碱基 loop、3 bp 第一茎、2 碱基泡和 3 bp 第二茎。改变 22R 茎环结构的碱基突变明显影响其活化 *GFP* 基因的作用, 过多互补和过少互补的茎环结构均不利于活化基因, 提示适当的不完整茎环结构与活化基因有关。

**关键词:** 非编码序列; *GFP*; 茎环结构; *Alu*; SV40PolyA

## Detection and sequence analysis of an imperfect stem-loop structure in cis activating gene element from SV40PolyA

WANG Hong-Gang<sup>1,2</sup>, MA Huan<sup>1</sup>, LI Zhu<sup>3</sup>, ZHANG Bin<sup>3</sup>, JING Xiang-Yang<sup>1</sup>, ZHANG Yuan<sup>1</sup>, LU Zhan-Jun<sup>1</sup>

1. Department of Genetics, Hebei Key Lab of Laboratory Animal, Hebei Medical University, Shijiazhuang 050017, China;

2. Department of Biochemistry and Molecular Biology, Medical College, Henan University, Kaifeng 475004, China;

3. Department of Respiratory Medicine, The First Hospital of Hebei Medical University, Shijiazhuang 050031, China

**Abstract:** Our previous studies showed that tandem *Alu* repeats inhibited *GFP* gene expression when they were inserted into the downstream of *GFP* gene in pEGFP-C1 vector and HeLa cells were then transfected transiently. The sequence named 2F2R (second 60 bp from the 5' end of SV40PolyA antisense strand) eliminated the repression of *GFP* gene expression induced by *Alu* repeats when 2F2R was inserted between *GFP* and *Alu* repeats. In this study the deletion of 2F2R DNA showed that 45R (45 bp in 2F2R 5'end), 30R (30 bp in 2F2R 5' end) and 22R (22 bp in 2F2R 5' end) activated *GFP* gene expression, and the activating actions of the double tandem sequences were stronger than those of their corresponding single sequences. Secloop (22 bp near the center in 2F2R) and Poly4 (30 bp in 2F2R 3' end) sequences did not activate *GFP* gene

收稿日期: 2010-09-06; 修回日期: 2010-10-29

基金项目: 河北省自然科学基金(编号: C2008001065)和河北省支撑计划项目(编号: 08276101D-90)资助

作者简介: 王红钢, 博士研究生, 讲师, 专业方向: 基因表达调节。E-mail: whg1975316@sohu.com

通讯作者: 吕占军, 博士生导师, 研究方向: 衰老机理。E-mail: LSLAB@hebm.edu.cn

expression. The activating action of 30R-Poly4 sequence formed by ligating 30R with Poly4 by 9 bp was lower than that of 2F2R. The linking base number between two 22R sequences did not influence the *GFP* gene expression obviously. Sequence 22R (5'-GTGAAAAAATGCTTTATTTGT-3') contains an imperfect palindrome sequence and may form an imperfect stem-loop structure including a 3nt loop, 3 bp first stem, 2nt bulge, and 3bp second stem. The mutations changing stem-loop structure of 22R influenced the *GFP* gene activation significantly and neither the excessively stable nor excessively unstable stem-loop structures were in favour of *GFP* gene activation, which suggested that the suitably imperfect stem-loop structures had something with gene activation.

**Keywords:** non-coding sequence; *GFP*; stem-loop; *Alu*; *SV40PolyA*

真核和原核生物基因组中均存在非编码序列, 真核生物基因组中非编码序列较多, 原核生物基因组中非编码序列较少, 如人基因组中大于 95% 的序列属于非编码序列, 而大肠杆菌基因组中非编码序列约占 11%。非编码序列中蕴藏着重要的有关遗传和发育的信息, 但是仍然不清楚多数非编码序列的生物学作用<sup>[1]</sup>。*Alu*元件属于短串连重复序列(Short Interspersed Nuclear Elements, SINE), 由 7SL RNA 基因经融合、进化形成, 并通过转座大量扩增, 如今 *Alu*元件的数量约占人类基因组的 10%<sup>[2]</sup>。*Alu*元件有基因抑制作用, 人类烟碱乙酰胆碱受体 $\alpha 6$  亚单位基因的构成和启动子分析显示, 将*Alu*串连序列融合到  $\alpha 6$  基因启动子区域, 不管方向和拷贝数目如何, 均为转录抑制成分<sup>[3]</sup>。

基因侧翼的非编码序列存在高度结构特征, 含有重要的调控元件, 包括 5'UTR、3' UTR、转录终止信号、增强子等, 对基因表达产生重要影响<sup>[4,5]</sup>。位于 N4 噬菌体 RNA 聚合酶启动子部位的茎环结构有活化基因的作用<sup>[6]</sup>。所有的黄病毒(Flavivirus)基因组含有 3'末端的茎环结构, 这些茎环结构对于病毒复制是必需的, 已经证明这些区域可以结合病毒编码的和细胞中固有的蛋白质<sup>[7]</sup>。日本库蚊病毒(Nodamura) RNA2 的 3'末端存在茎环结构, 删除该茎环结构, 严重阻碍 RNA2 复制<sup>[8]</sup>。

将人基因组的 DNA 分成 7nt 小片段, 统计每一种小片段的个数, 发现互补的小片段有几乎相同的个数<sup>[9]</sup>, 提示在单链水平 DNA 也具有互补性, 可以形成茎环结构。用生物软件分析人基因组 DNA 单链, 只要具备一定的长度(如一百碱基), 几乎都可以折叠形成茎环结构。这些结果说明了 DNA 茎环结构在人基因组中普遍存在, 但是人基因组中茎环结构的

功能还不清楚。

本研究的前期工作发现, *Alu14*(14 个 *Alu* 同向串连)插入 pEGFP-C1 质粒的 *GFP* 基因下游可以抑制基因表达<sup>[10]</sup>, 在 *GFP* 基因和 *Alu14* 之间插入 *SV40PolyA* 反序(*PolyAas*, 240 bp)解除 *Alu14* 对 *GFP* 基因的抑制作用, 将 *PolyAas* 分为 4 段, 每段 60 bp, 第 2 个 60 bp (2F2R) 依然可以活化 *GFP* 基因<sup>[11]</sup>。本研究通过删减 2F2R 发现一个 22 bp 序列活化基因, 又通过对这个 22 bp 序列进行突变和删减, 研究 DNA 茎环结构对 *GFP* 基因活化的作用。选择 pEGFP-C1 作为本项目的真核细胞表达载体是因为 pEGFP-C1 带有荧光报告基因便于检测, 并且 pEGFP-C1 的多克隆位点位于 *GFP* 基因下游, 插入序列不会影响 *GFP* 读码。本研究的目的是探讨下游序列对 *GFP* 基因表达的影响, 使用 pEGFP-C1 质粒符合这一要求。

## 1 材料和方法

### 1.1 材料

p*Alu14* 质粒为本研究室前期工作构建, 系将 14 个 *Alu* 元件同向、正向串连插入 pEGFP-C1 质粒的 *GFP* 基因下游, pEGFP-C1 质粒为空载体, 不含有插入序列<sup>[12]</sup>。pEGFP-C1 质粒、DH5 $\alpha$  菌株和 HeLa 细胞均为本实验室保存。

### 1.2 方法

#### 1.2.1 合成引物和作为模板的 DNA 片段

委托 DNA 合成公司(赛百盛, 北京)合成带适当酶切位点的引物, 上游引物带 *EcoR* I / *Xba* I (或 *Hind* III / *Xba* I) 酶切位点, 下游引物带 *Kpn* I / *Nhe* I 酶切位点(表 1)。通过 PCR 扩增将酶切位点引入到扩增

产物中, 便于下一步将 PCR 扩增产物连入质粒。合成带有不同突变位点或突变部位的 DNA 片段作为模板(表 2)。

### 1.2.2 表达载体构建

用带有酶切位点的引物, 进行 PCR 扩增目的片段, 目的片段有两类: 一类是 pEGFP-C1 质粒中的 SV40PolyA 片段, 使用的引物见表 1, 扩增片段的位置见图 1; 另一类是合成的 DNA 片段(表 2), 使用的引物见表 1。构建表达载体涉及两个过程, 其一是将 PCR 扩增片段插入 pEGFP-C1 质粒, 构建单体和 2 串连体质粒; 其二是将构建的 pEGFP-C1 质粒中的插入片段(单体或 2 串连体)用限制酶切下, 经电泳、胶分离后插入 pAlu14 质粒的 GFP 基因和 Alu 串连序列之间, 即插入 GFP 下游, Alu 串连序列的上游。

具体方法是, 用 *EcoR* I /*Kpn* I (或 *Hind*III/*Xba* I)双酶切 pEGFP-C1 质粒, 同时用同样的酶切 PCR 扩增产物, T4 DNA 连接酶连接后得到插入一个片段的表达载体。*Nhe* I 酶切产生一个 5'CTAG 突出端, 可以与 *Xba* I 酶切产生的质粒大片段有效连接。但是连接以后的位点不能被其中的任何一个酶切断。我们利用此特性制备同向二串连体, 本文中涉及的所有 2 串连体(如 2F2R\*2, 45R\*2 等)均用这种方法制备。再将 pEGFP-C1 质粒中的插入片段(单体或 2 串连体)用 *Hind*III/ *Nhe* I 切下, 连入经 *Hind*III/ *Xba* I 酶切的 pAlu14 质粒中。经 PCR、酶切和测序(诺赛基因, 北京)鉴定正确以后用于实验。详见参考文献[12,13]。

文章中使用的表达载体名称含义举例; “p2F2R\*1-Alu14”代表在 pEGFP-C1 质粒的 GFP 基

表 1 构建表达载体的引物

引物名称	引物序列	注释
Poly60-2F(上游引物)	<i>EcoR</i> I <i>Xba</i> I 5'-ATCGGAATTCCTTAATCTAGATAAGTGAAAAAATGCTTTATT	扩增 2F2R,45R,30R
Poly60-2R(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGCACACTACGATAACGAAA	扩增 2F2R, Poly4
45R(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGCAAAGCAATAGCATCA	扩增 45R
30R(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGCCAAATTCACAAATA	扩增 30R
Poly4F(上游引物)	<i>EcoR</i> I <i>Xba</i> I 5'-ATCGGAATTCCTTAATCTAGATAATGATGCTATTG	扩增 Poly4
FirLoopF(上游引物)	<i>EcoR</i> I <i>Xba</i> I 5'-ATCGGAATTCCTTAATCTAGATAATGTGAAAAAA	扩增 22R, 19R, 16R
Poly22R(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGCACAAATAAAGCA	扩增 22R
SecloopF(上游引物)	<i>EcoR</i> I <i>Xba</i> I 5'-ATCGGAATTCCTTAATCTAGATAATGCTTTATTTGT	扩增 Secloop
SecloopR(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGCCACAAATTCAC	扩增 Secloop
1619MR(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGCAC	扩增 19R,16R
EcoXba (上游引物)	<i>EcoR</i> I <i>Xba</i> I 5'-ATCGGAATTCCTTAATCTAGA	扩增 17ntAT
KpnNhe(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGC	扩增 17ntAT
22R2F(上游引物)	<i>Hind</i> III <i>Xba</i> I 5'-ATCGAAGCTTAATCTAGAGTGAAAAAATGC	扩增 22R20, 22R21, 22R22, 22R24, 22R25, 22R28, 22R211
22R2R(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGCACAAATAAAGC	扩增 22R20, 22R21, 22R22, 22R24, 22R25, 22R28, 22R211
22RF(上游引物)	<i>EcoR</i> I <i>Xba</i> I 5'-ATCGGAATTCCTTAATCTAGATAATGTG	扩增 S2, S4, S5, S6, S7, S8, SecS2, SecS4, SecS5,SecS6,SecS7,SecS8,
22RR(下游引物)	<i>Kpn</i> I <i>Nhe</i> I 5'-ATCGGGTACCATGCTAGCAC	扩增 S2, S4, S5, S6, S7, S8, SecS2, SecS4, SecS5,SecS6,SecS7,SecS8

表 2 构建表达载体的合成模板

模板名称	合成模板的序列
19R	5'-AATGTGAAAAAATGCTTTATTGCTAGC-3'
16R	5'-AATGTGAAAAAATGCTTTGCTAGC-3'
17ntAT	5'-CTAGATAATAAAAAAATGCTTTATTGCTAGCAT-3'
22R20	5'-GAAAAAAATGCTTTATTTAAAAAATGCTTTATTGT-3'
22R21	5'-GAAAAAAATGCTTTATTTGAAAAAATGCTTTATTGT-3'
22R22	5'-GAAAAAAATGCTTTATTTTGAAAAAATGCTTTATTGT-3'
22R23	5'-GAAAAAAATGCTTTATTTGTGAAAAAATGCTTTATTGT-3'
22R24	5'-GAAAAAAATGCTTTATTTTGAAAAAATGCTTTATTGT-3'
22R25	5'-GAAAAAAATGCTTTATTTGTGTGAAAAAATGCTTTATTGT-3'
22R28	5'-GAAAAAAATGCTTTATTTGTGATGTGAAAAAATGCTTTATTGT-3'
22R211	5'-GAAAAAAATGCTTTATTTGTGCTAATGTGAAAAAATGCTTTATTGT-3'
S2	5'-ATAATGTGAAAAAATGCTTTATTGTGCTAGC-3'
S4	5'-ATAATGTGAAAAAATGCTTTTATTGTCTAGC-3'
S5	5'-ATAATGTGAAAAAATGCTTTTATTGTCTAGC-3'
S6	5'-ATAATGTGAAAAAATGCTTTTATTGTCTAGC-3'
S7	5'-ATAATGTGAAAAAATGCTTTTATTGTCTAGC-3'
S8	5'-ATAATGTGAAAAAATGCTTTTATTGTCTAGC-3'
SecS2	5'-ATAATGTGAAAAAATGCTTTATTGTCTAGC-3'
SecS4	5'-ATAATGTGAAAAAATGCTTTATTGTCTAGC-3'
SecS5	5'-ATAATGTGAAAAAATGCTTTATTGTCTAGC-3'
SecS6	5'-ATAATGTGAAAAAATGCTTTATTGTCTAGC-3'
SecS7	5'-ATAATGTGAAAAAATGCTTTATTGTCTAGC-3'
SecS8	5'-ATAATGTGAAAAAATGCTTTATTGTCTAGC-3'

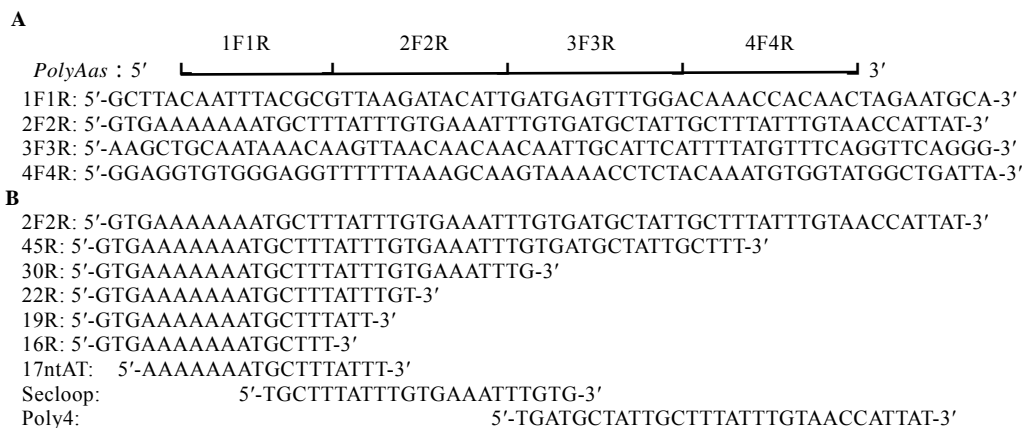


图 1 2F2R 及其删减序列的序列和位置

A: PolyAas 等分为 4 段, 各段的名称、位置和序列; B: 对 2F2R 进行删减突变, 各删减片段的名称、位置和序列。

因下游依次插入一个 2F2R 片段和 14 个 *Alu* 元件, 即在 pAlu14 质粒的 *GFP* 基因下游插入一个 2F2R 片段; “p2F2R\*2-Alu14”代表在 pAlu14 质粒的 *GFP* 基因下游插入 2 个 2F2R 片段。其余质粒命名遵从相同的原则。

### 1.2.3 细胞转染

用质粒(包括重组表达载体和 pEGFP-C1)0.4 μg

和 Lipofectamine<sup>TM</sup>2000(Invitrogen Carlsbad, USA) 2 μL 瞬时转染培养于 24 孔板的 HeLa 细胞, 于 37 ℃、5%CO<sub>2</sub> 环境中继续培养 36 h, 用 Trizol 试剂提取总 RNA 用于 Northern 印记检测。

### 1.2.4 Northern 印记检测

用 9N 随机引物和大肠杆菌 DNA 聚合酶大片段,

将 $\alpha$ - $^{32}$ P-dCTP掺入DNA(590 bp)中制备GFP探针。590 bp DNA片段扩增自GFP基因并经过琼脂糖凝胶电泳纯化,上、下游引物分别为;5'-GGGCGAGG GCGATG-3'和5'-CTTGTACAGCTCGTCCATGC-3'。提取转染细胞总RNA,甲醛变性琼脂糖凝胶电泳,将RNA转移到尼龙膜。 $^{32}$ P标记的GFP探针杂交、冲洗后通过放射自显影记录实验结果。然后,这种杂交过的尼龙膜用剥离液处理,去除 $^{32}$ P-GFP探针,用检测neo RNA的探针再次杂交,放射自显影作为转染效率的内参对照。Neo探针用随机引物法标记,标记模板为Neo基因PCR扩增产物(671 bp),扩增 671 bp Neo DNA的上、下游引物分别为;5'-CACACAACAGACAATCGGCTGCT-3'和5'-AGCGGCGATACCGTAAAGCAC-3'。

## 2 结果与分析

### 2.1 删减 2F2R 对活化基因的影响

本文中所要研究的序列来自PolyAas(240 bp),将PolyAas分成 60 bp一段,分别称为 1F1R、2F2R、3F3R和 4F4R,其位置和序列见图 1A。2F2R插入 pAlu14 的GFP和Alu序列之间可以解除Alu14 对GFP基因表达的抑制作用<sup>[4]</sup>。为了获得 2F2R中的何种序列可以活化基因的证据,本文中对 2F2R进行删减分析,删减碱基序列和位置见图 1B。

图 2 为 2F2R 3' 端删减碱基后,单体或二串连体插入 pAlu14 质粒 GFP 基因下游,转染 HeLa 细胞后 Northern 印记检测结果。可见单体和二串连体随着删减碱基的增加,活化基因作用下降,19R 和 16R

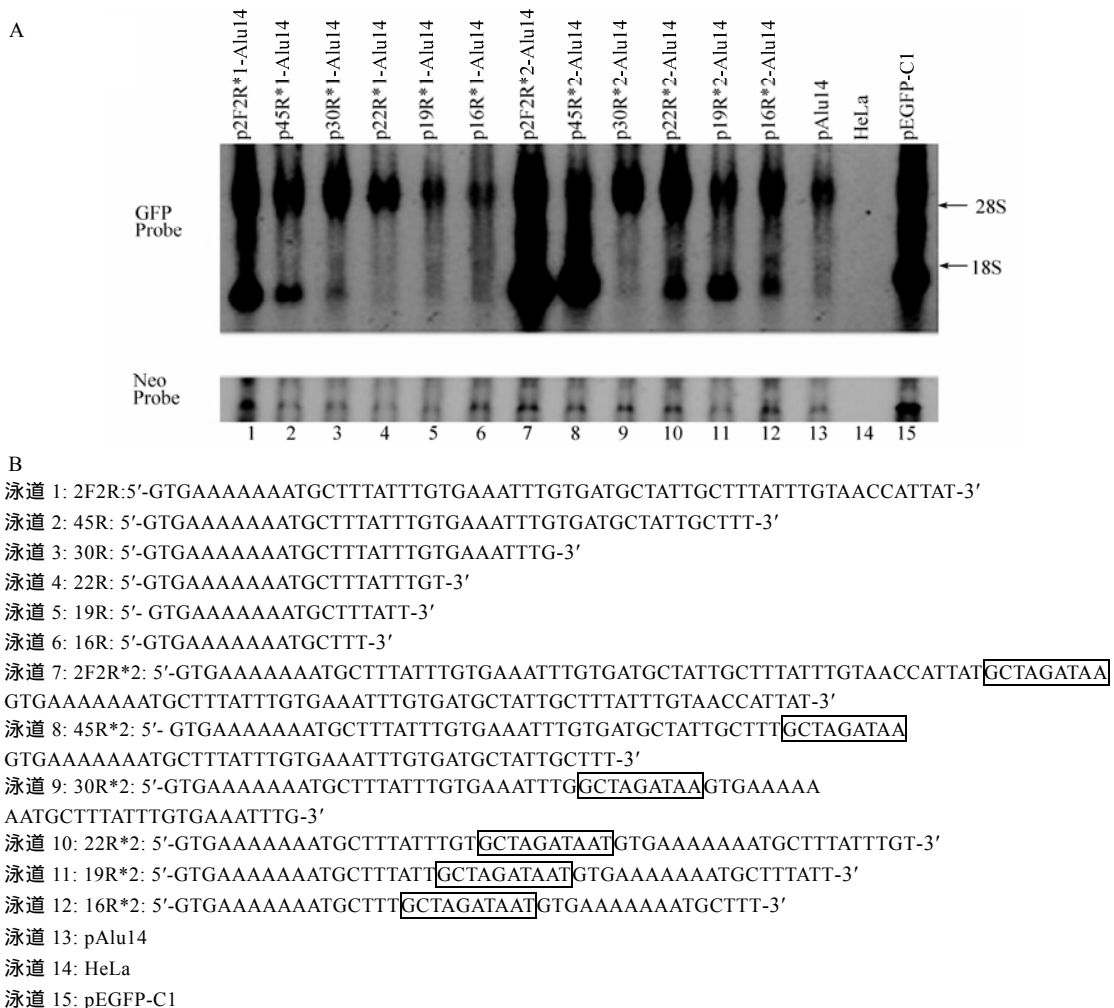


图 2 2F2R 及其删减序列插入 pAlu14 活化 GFP 基因作用的 Northern 印记检测

A: 2F2R 或者其突变序列以单体或二串连体形式插入 pAlu14 质粒的 GFP 基因和 Alu 串连序列之间,转染 HeLa 细胞, Northern 印记检测 GFP 报告基因的表达。B: 泳道对应的插入序列,方框中为连接 2 个单体的酶切位点连接后的序列。泳道 13、14 和 15 分别为插入 14 个 Alu 的 pEGFP-C1,未转染和空载体转染对照。

活化基因作用很弱。2F2R 及其删减序列二串连体的活化基因作用均高于其相应的单体(图 2A 泳道 7-12 vs. 泳道 1-6)。碱基删减和串连还影响转录终止位置, 单个 45R 的低分子转录物和高分子转录物各占约 1/2, 但是二串连体 45R 以低分子量转录物为主(图 2A 泳道 2 vs. 泳道 8); 二串连体 19R 明显出现低分子量的转录物, 与其单体表现不同(图 2A 泳道 5 vs. 泳道 11), 这可能是因为碱基删减和串连使转录终止位点的作用发生改变。30R 与 22R 单体和二串连体均以高分子量转录物为主(图 2A 泳道 3、4 和泳道 9、10)。图 2B 是插入 pAlu14 质粒中的片段。

图 3 是 2F2R 的其他两个位置的序列 Poly4 和 Secloop 的单体(在 2F2R 中的位置见图 1B, 插入序列见图 3B)及其二串连体插入 pAlu14 质粒的 Northern 印记检测结果。该两个片段插入 pAlu14 质粒与 pAlu14 比较均没有明显的活化 GFP 基因作用(泳道 1-4 vs. 泳道 5)。

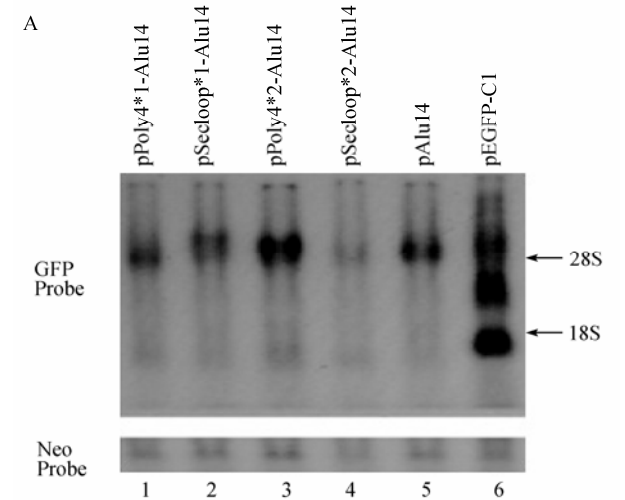


图 3 2F2R 其他位置片段插入 pAlu14 的活化 GFP 基因作用(Northern 印记检测)  
A: Poly4、Secloop 单体或二串连体插入 pAlu14 质粒的 GFP 基因和 Alu 串连序列之间, 转染 HeLa 细胞, Northern 印记检测 GFP 报告基因的表达。B: 泳道对应的插入序列, 方框中为连接 2 个单体的酶切位点连接后的序列。

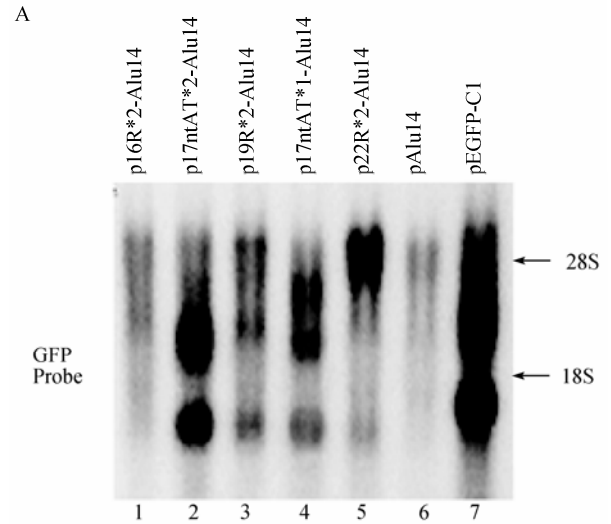


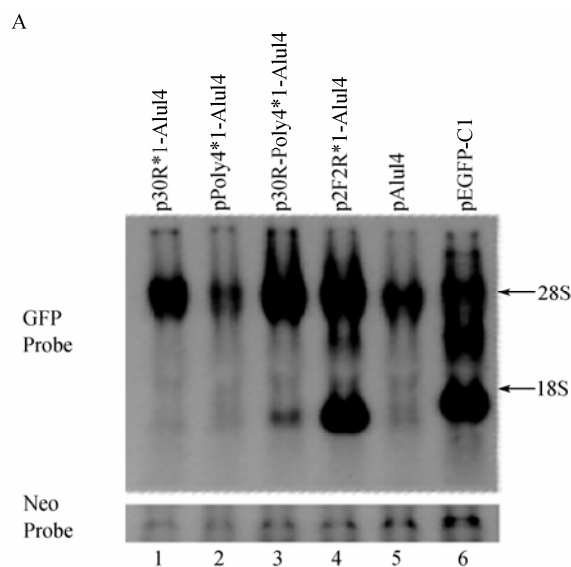
图 4 22R 及其突变序列插入 pAlu14 的活化 GFP 基因作用(Northern 印记检测)  
A: 22R 及其删减序列的二串连体插入 pAlu14 质粒的 GFP 基因和 Alu 串连序列之间, 转染 HeLa 细胞, Northern 印记检测 GFP 报告基因的表达。B: 泳道对应的插入序列, 方框中为连接 2 个单体的酶切位点连接后的序列。

图 4 是 17ntAT 与 22R 的比较, 区别在于删除了 22R 上游的“5'-GTG”和下游的“GT-3'”, 17ntAT 缺失了 22R 上游的 3 个碱基和下游的 2 个碱基, 仍然具有活化基因的作用, 且转录物的量不低于 22R(图 4A 泳道 2 vs. 泳道 5), 说明删除的这 5 个碱基对于活化基因并不重要。19R 等于是删除了 22R 下游的 3 个碱基(“TGT-3'”), 转录量明显下降(图 4A 泳道 3 vs. 泳道 5), 说明从 22R 下游向上游数第 3 个碱基(T)对于转录至关重要。在 19R 的基础上再删除下游的 3 个碱基(ATT-3')成为 16R, 转录量又进一步下降(图 4A 泳道 1 vs. 泳道 5)。与其他序列相同, 17ntAT 二串连体的转录量高于单体(图 4A 泳道 2 vs. 泳道 4)。22R

上游和下游的 5 个碱基尽管对转录量不产生影响,但是明显促进转录延伸,与 17ntAT 比较,22R 产生高分子量的转录物(图 4A 泳道 2 vs.泳道 5)。图 4B 是插入质粒的序列。

## 2.2 30R 与 Poly4 串连增加活化基因作用

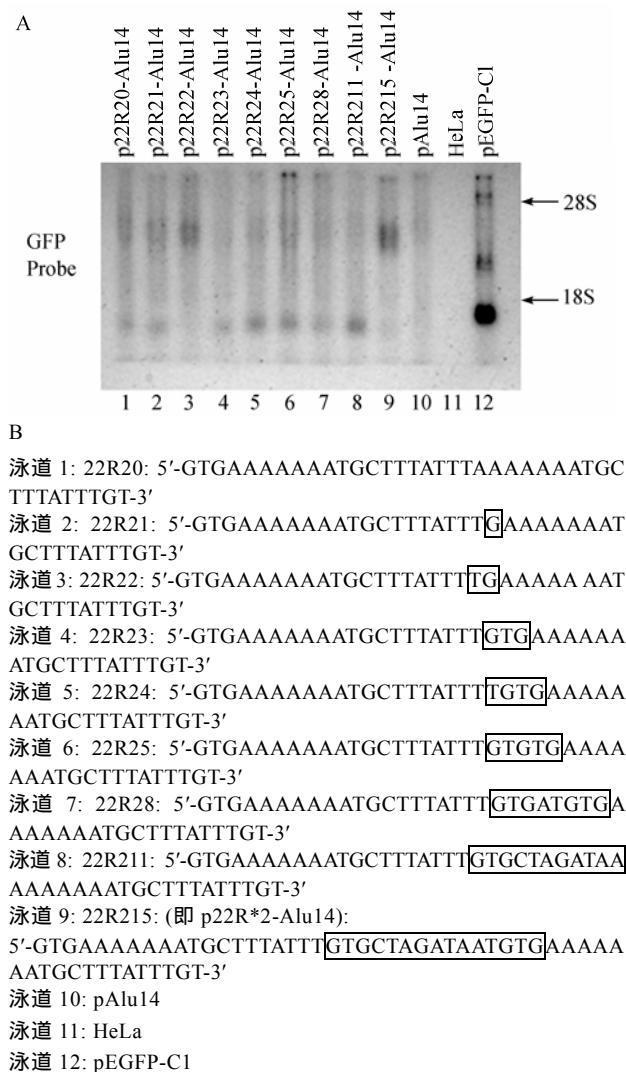
30R 可以活化基因(图 5A 泳道 1 vs.泳道 5), Poly4 不能活化基因(图 5A 泳道 2 vs.泳道 5)。尽管插入 Poly4 不能活化 *GFP* 基因,但是将 Poly4 接在 30R 下游可以增加 30R 的基因活化作用(图 5A 泳道 3 vs.泳道 1、2、5),说明 30R 与 Poly4 串连(30R-Poly4)有协同活化基因的作用。30R-Poly4 与 2F2R 相比,在中间加入酶切位点(9 碱基,图 5B),由于这 9 个碱基的插入使 30R-Poly4 高分子量转录物的量上升,转录物的总量有所下降(图 5A 泳道 3 vs.泳道 4)。图 5B 显示插入质粒的序列。



**图 5 30R 与 Poly4 串连增加活化 *GFP* 基因作用(Northern 印记检测)**  
A: 30R 与 Poly4 串连后插入 pAlu14 质粒的 *GFP* 基因和 *Alu* 串连序列之间,转染 HeLa 细胞, Northern 印记检测 *GFP* 报告基因的表达。B: 泳道对应的插入序列,方框中为连接 2 个单体的酶切位点连接后的序列。

## 2.3 两个 22R 之间连接碱基数对活化 *GFP* 基因的作用

改变两个 22R 之间的连接碱基数(将 22R 3'端 TG 和 5'端 GTG 也视为连接碱基),观察对 *GFP* 基因的活化作用,碱基序列如图 6B,两个 22R 之间的碱基数对活化基因作用没有明显的影响(图 6A),该特性不同于 2F2R 和 30R-Poly4 串连体的比较(0 碱基连接活化基因作用高于 9 碱基连接)。中间有 2 个碱基(图 6A,泳道 3)和 15 个碱基(图 6A,泳道 9)导致较多的高分子量转录物出现。



**图 6 两个 22R 之间连接碱基数对活化 *GFP* 基因的作用(Northern 印记检测)**

A: 合成 22R20-22R211 模板,用带有酶切位点的上、下游引物扩增,酶切后插入 pAlu14 质粒,测序正确后用于实验,转染 HeLa 细胞, Northern 印记检测 *GFP* 报告基因的表达。B: 泳道对应的插入序列,方框中为连接



接 2 个 22R 的序列 (将 22R 的 TG-3' 和 5'-GTG 也计入连接序列)。

## 2.4 茎长的作用

改变 22R 第一茎和第二茎的碱基数, 从 2 个碱基到 8 个碱基, 插入到 pAlu14 的 *GFP* 下游, 转染细胞后做 Northern 印记检测, 结果见图 7A, 插入序列见图 7B, 只有 22R (图 7A 泳道 13 vs. 泳道 1-12) 转录效率最高, 且高分子量转录物最多。

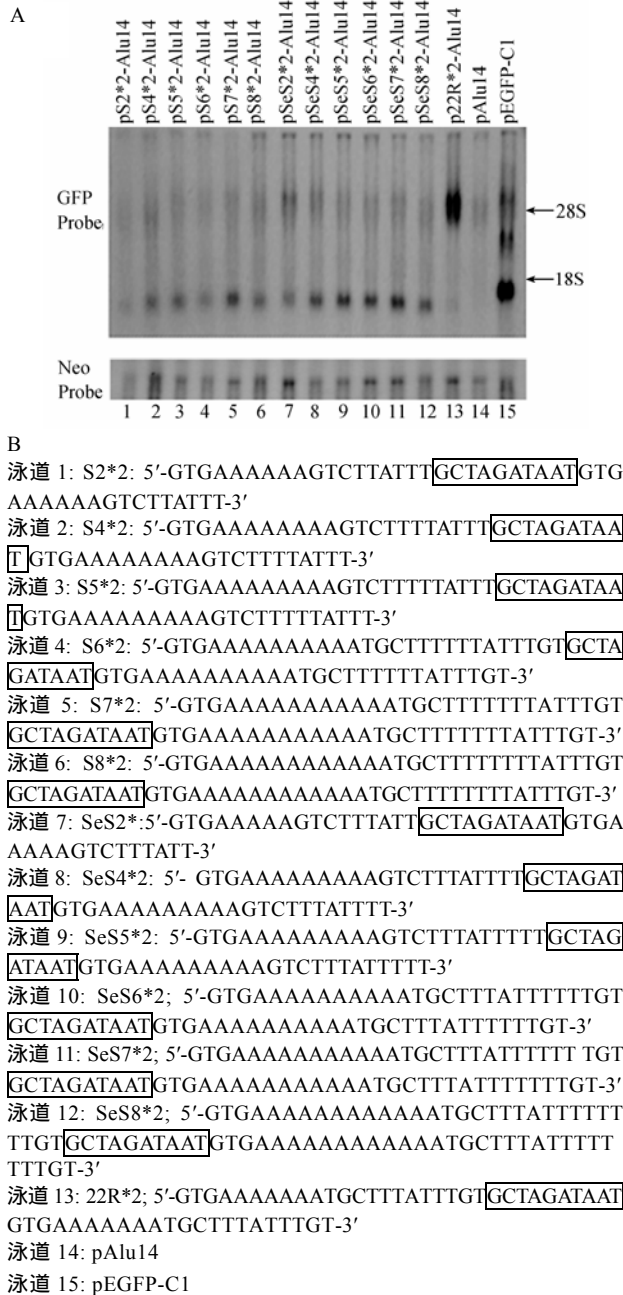


图 7 改变 22R 茎环结构中茎的碱基数对活化 *GFP* 基因的作用(Northern 印记检测)

A: 将不同茎长的 22R 序列二串连体插入 pAlu14 质粒, 转染 HeLa 细胞, Northern 印记检测 *GFP* 报告基因的表达。B: 泳道对应的插入

序列, 方框中为连接 2 个 22R 茎长突变体的序列。

## 3 讨论

已知质粒中的 *SV40PolyA* 在 HeLa 细胞中上调荧光素酶(Luciferase)报告基因表达<sup>[14]</sup>。我们以往的工作证明 *SV40PolyA* 及其片段(2F2R)可以解除 *Alu* 串连序列对 *GFP* 基因表达的抑制作用<sup>[11]</sup>。为了研究 2F2R 中何种序列在活化基因中起关键作用, 从 2F2R 的 3'端渐次删除碱基, 结果表明 45R、30R 和 22R 都可以活化基因, 且其二串连体的活化基因作用强于单体, 但是 19R 和 16R 没有明显的活化基因作用。2F2R 的近中部和下游的两段序列, Secloop 和 Poly4 没有明显的活化基因作用。

已知茎环结构与基因转录有关, N4 噬菌体 RNA 聚合酶(vRNAP)识别的启动子部位含有反向重复序列, 包括 5~7 碱基的茎和 3 碱基的环, 这样的结构对于 vRNAP 的识别是必需的<sup>[6]</sup>。本文中发现的 22R 包含不完整的回文序列, 可以形成不稳定的茎环结构, 可以分为 3 碱基的 loop、3bp 的第一茎、两碱基的泡和 3 bp 第二茎共 4 部分, 如图 8 所示。

从 22R 序列的 3'端删除 3 个碱基形成 19R, 活化 *GFP* 基因的作用显著减弱, 说明 22R 下游的这 3 个碱基(5'-TGT-3')对于活化基因有重要作用。从 22R 的 5'端删除 GTG 同时从 3'端删除 GT 形成 17ntAT、17ntAT 单体和二串连体插入 pAlu14 质粒均可以活化 *GFP* 基因, 说明删除的这 5 个碱基对于活化 *GFP* 基因不重要。17ntAT 是 22R 中形成不完整茎环结构的最小序列, 而 19R 和 16R 均可以破坏该种茎环结构并伴随 *GFP* 基因活化作用明显下降, 提示该种不完整茎环结构在 *GFP* 基因活化中起作用。本文中还观察到 22R 以高分子量转录物为主, 而 17ntAT 以低分子量转录物为主, 反映它们的转录终止位置发生改变。

5'和 3'非翻译区对于控制病毒的基因组功能有重要作用, 生物软件预测这些非翻译区含有不同的茎环结构, 多数在其茎中显示泡结构(出现非互补碱基), 属于不完整的茎环结构<sup>[7,8,15]</sup>。鼠肝炎病毒基因组第一茎环(MHV-SL-1)的第 35 碱基是 A, 该碱基在茎中是不互补的, 形成泡, 如果去掉这个碱基, 使茎完全互补, 病毒则不能复制<sup>[16]</sup>, 说明过于稳定的茎环不利于病毒复制。



所有生物的 DNA 都有负超螺旋(右手螺旋向相反

的方向旋转形成的结构), 负超螺旋有多余的能量,

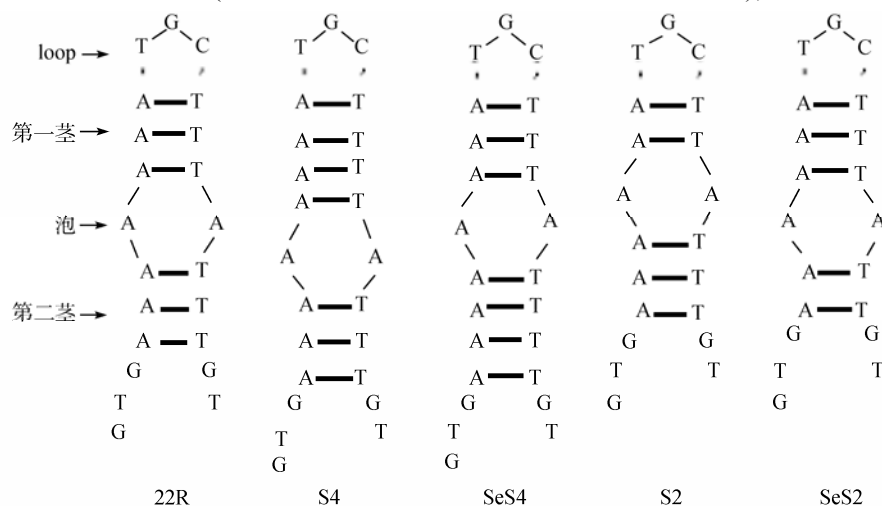


图8 22R 及其突变序列次级结构示意图

会转变为单链泡状结构, 如果泡状区域有回文序列则可以形成茎环结构, 通过这样的分析可以认为, 活体细胞中双链 DNA 反向重复序列可能形成小范围的茎环结构。

我们的初步实验表明, 茎环结构要有适当的大小才可以活化基因。如图8所示22R有适当的茎环结构, 可以活化基因; 而S4和SeS4过度互补, 活化基因作用减弱, S2和SeS2互补作用弱, 其活化基因作用也下降。关于为什么会有这种现象, 在此提出可能的假设; 我们的前期工作证明, *Alu*串连序列抑制GFP基因表达与促成染色质紧密包装有关<sup>[10]</sup>, 那么破坏染色质的紧密包装就可以活化基因, 图8中22R能够形成不完整茎环结构(回文序列不完全互补), 这种结构属于不稳定的茎环结构, 使DNA在不同形式的双链和茎环结构之间不断转化, 从而形成动态DNA, 不容易被紧密包装。核小体形成和染色质包装涉及蛋白质与DNA相互作用和蛋白质与蛋白质相互作用, 是一种多级反应过程, 例如DNA与组蛋白结合成核小体有人认为涉及20个以上的非共价键反应。形成茎环结构不仅改变了DNA的相对长度, 而且改变了DNA的相位, 会影响染色质包装。

适当的不稳定茎环结构活化基因, 有重要的生物学意义。在人的基因组中有许多不完整的茎环结构, 有可能是活化基因的顺式元件, 顺式元件通常要通过反式调节因子发挥作用, 寻找触发茎环结构

的反式调节因子是下一步的工作。基因表达调节在衰老过程中起重要作用, 细胞一代比一代老, 但是DNA的一级结构没有明显的改变(表观遗传)。虽然关于表观遗传有若干解释, 如DNA甲基化、组蛋白修饰、染色质构象等, 但是仍然不能全面解释和很好的控制表观遗传和衰老, 可能还存在其他未知的调控机制和调控成分。本文初步证明DNA的不完整茎环结构可以活化基因, 这可能是基因调控的一种形式, 将为衰老、表观遗传研究提供新的线索。

## 参考文献(References):

- [1] Costa FF. Non-coding RNAs, epigenetics and complexity. *Gene*, 2008, 410(1): 9–17. [\[DOI\]](#)
- [2] Polak P, Domany E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics*, 2006, 7: 133. DOI: 10.1186/1471-2164-7-133. [\[DOI\]](#)
- [3] Ebihara M, Ohba H, Ohno SI, Yoshikawa T. Genomic organization and promoter analysis of the human nicotinic acetylcholine receptor  $\alpha 6$  subunit (*CHNRA6*) gene: *Alu* and other elements direct transcriptional repression. *Gene*, 2002, 298(1): 101–108. [\[DOI\]](#)
- [4] Andreassi C, Riccio A. To localize or not to localize: mRNA fate is in 3' UTR ends. *Trends Cell Biol*, 2009, 19(9): 465–474. [\[DOI\]](#)
- [5] Mao JY, Li CY, Zhang YY, Li YH, Zhao YY. Human with-no-lysine kinase-4 3' -UTR acting as the enhancer

- and being targeted by miR-296. *Int J Biochem Cell Biol*, 2010, 42(9): 1536–1543. [\[DOI\]](#)
- [6] Dai X, Greizerstein MB, Nadas-Chinni K and Rothman-Denes LB. Supercoil-induced extrusion of a regulatory DNA hairpin. *Proc Natl Acad Sci USA*, 1997, 94(6): 2174–2179. [\[DOI\]](#)
- [7] Yu L, Markoff L. The topology of bulges in the long stem of the flavivirus 3' stem-loop is a major determinant of RNA replication competence. *J Virol*, 2005, 79(4): 2309–2324. [\[DOI\]](#)
- [8] Roskopf JJ, Upton JH 3rd, Rodarte L, Romero TA, Leung MY, Taufer M, Johnson KL. A 3' terminal stem-loop structure in *Nodamura virus* RNA 2 forms an essential *cis*-acting signal for RNA replication. *Virus Res*, 2010, 150(1–2): 12–21. [\[DOI\]](#)
- [9] Lu ZJ, Zhai Y, Wang XF, Song SX. DNA sequence composition on human X chromosome differing from that on chromosomes 6, 7, 8, 10, 11 and 12. *Acta Genet Sin*, 2003, 30(11): 1051–1060. [\[DOI\]](#)
- [10] Wang XF, Wang XY, Liu J, Feng JJ, Mu WL, Shi XJ, Yang QQ, Duan XC, Xie Y, Lu ZJ. *Alu* tandem sequences inhibit GFP gene expression by triggering chromatin wrapping. *Gene Genomics*, 2009, 31(3): 209–215. [\[DOI\]](#)
- [11] Yin K, Wang XF, Ma H, Xie Y, Feng JJ, Yang QQ, Lü ZJ. Impact of copy number of distinct SV40PolyA segments on expression of a GFP reporter gene. *Sci China Life Sci*, 2010, 53(5): 606–612. [\[DOI\]](#)
- [12] Wang XF, Jin X, Wang XY, Liu J, Feng JJ, Yang QQ, Mu WL, Shi XJ, Lu ZJ. Effects of L1-ORF2 fragments on green fluorescent protein gene expression. *Genet Mol Biol*, 2009, 32(4): 688–696. [\[DOI\]](#)
- [13] 段肖翠, 靳霞, 谢英, 焦宁, 刘静, 王晓燕, 吕占军. L1-ORF2 不同片段对报告基因表达产生不同影响. *遗传*, 2009, 31(1): 50–56. [\[DOI\]](#)
- [14] Xu ZL, Mizuguchi H, Ishii-Watabe A, Uchida E, Mayumi T, Hayakawa T. Optimization of transcriptional regulatory elements for constructing plasmid vectors. *Gene*, 2001, 272(1–2): 149–156. [\[DOI\]](#)
- [15] Nickens DG, Hardy RW. Structural and functional analyses of stem-loop 1 of the Sindbis virus genome. *Virology*, 2008, 370(1): 158–172. [\[DOI\]](#)
- [16] Li LC, Kang H, Liu PH, Makkinje N, Williamson ST, Leibowitz JL, Giedroc DP. Structural lability in stem-loop 1 drives a 5' UTR-3' UTR interaction in coronavirus replication. *J Mol Biol*, 2008, 377(3): 790–803. [\[DOI\]](#)

## •科学新闻•

### 我国首次查清小熊猫身份之谜

2011年3月28日,中央电视台综合频道(CCTV-1)《新闻30分》和新闻频道(CCTV-13)的《新闻30分》及《新闻直播间》栏目,以“我国首次查清小熊猫身份之谜”对云南大学云南省生物资源保护与利用重点实验室于黎研究员课题组研究成果进行了报道。

于黎研究员课题组一直致力于我国珍稀濒危物种小熊猫的分类及进化历史研究。近日,该课题组结合基因组信息,通过筛选大量基因标记,提出小熊猫是单独的一个科,而且与浣熊科和鼬科的关系最接近。这一研究结果不支持小熊猫属于浣熊科、熊科或臭鼬科的传统观点,为目前世界上备受争议的小熊猫进化地位提供了重要信息。而且此研究提供了一个使用大规模内含子数据进行哺乳动物分子系统学研究的成功范例,为脊椎动物分子系统学研究提供了新的核基因标记。文章发表在分子系统学领域刊物《*Systematic Biology*》(2011, 60: 175–187; 2009年影响因子8.48)。

此外,“云南电视台”、“北京科技报”、“中新社”以及“中国广播网”等媒体也陆续对此进行了报道和关注。