

## 牛 $\alpha$ -s1酪蛋白基因3'端的克隆鉴定及序列分析

陈瑞环 汪波 张玉芝\* 刘伟 张靖溥 劳为德

(中国科学院生物物理研究所, 北京)

本文用牛 $\alpha$ -s1酪蛋白基因 cDNA 为探针, 筛选了以EMBL3为载体构建的牛基因组文库  $2 \times 10^5$  pfu, 得到一个阳性克隆。分别以牛 $\alpha$ -s1酪蛋白基因cDNA全长、终止符前后序列为探针鉴定了该克隆, 制作了较为详细的限制酶谱, 亚克隆了相应片段, 并对含有终止符和 poly A 加成信号的有关亚克隆进行了序列分析, 结果表明: 该克隆含有牛 $\alpha$ -s1酪蛋白基因终止符下游3.5kb及上游10.2kb片段, 与现有的该基因的资料相比较, 限制酶谱存在部分差异, DNA序列有多处点突变, 并有少量缺失。点突变在内含子和外显子均有发生, 这是限制酶图谱差异的原因。

**关键词** 牛 $\alpha$ -s1酪蛋白基因; 3'下游区; 克隆与鉴定; 序列分析

乳汁蛋白基因启动子与外源基因融合产生转基因动物, 以期从动物乳汁中获取目的基因产物已成为真核基因表达的基因工程中引人注目的途径<sup>[1-6]</sup>。乳汁蛋白基因的5'端实现了外源基因在乳腺的特异性表达, 然而要获得高效表达, 尚有许多因素在起作用。虽然有研究探讨了顺、反式作用、内含子等对转基因动物中基因表达的影响<sup>[7-9]</sup>, 但总体来看, 有关依据和方法仍不清楚。牛 $\alpha$ -s1酪蛋白基因存在复杂的顺反式调控机制, 转录需经历复杂的剪接<sup>[10]</sup>, 从该基因启动区来看, 它不具有CAAT盒, TATA盒序列TTTAAAT为一弱启动区<sup>[10, 11]</sup>, 但牛 $\alpha$ -s1酪蛋白基因的mRNA和蛋白的表达量却很高, 因此推测除了5'端调控机制外, 内含子及3'端在表达过程中会起到一定的作用。我们已成功克隆了含有牛 $\alpha$ -s1酪蛋白基因5'端及其较长片段的上游区<sup>[12]</sup>, 并发现了5'端存在数个结合蛋白<sup>[13]</sup>, 本文报道包含最后一个外显子及其下游区的牛 $\alpha$ -s1酪蛋白基因3'端的克隆鉴定、限制酶谱、终止符及polyA加成信号附近的序列, 并与现

有的资料<sup>[5, 10]</sup>进行了比较。

### 材料与 方法

#### (一) 材料

牛基因组文库、牛 $\alpha$ -s1酪蛋白基因cDNA克隆(p184)由中国科学院动物研究所沈孝宙教授惠赠; 大肠杆菌菌株GM242 (dcm<sup>-</sup>)、DH5 $\alpha$ 、JM109为本室自存; T4 DNA连接酶, T7 polymerase序列分析试剂盒为promega公司产品; 限制酶为华美生物工程公司产品; 硝酸纤维素膜、随机引物标记试剂盒为Bio-Rad公司产品; <sup>32</sup>P-dCTP, <sup>35</sup>S-dATP为Du Pont公司产品; 其余试剂均为国产分析纯。

#### (二) 方法

1. 探针的标记、噬斑及Southern吸印、杂交质粒、噬菌体的提纯, DNA酶切与部分酶切, 亚克隆的制备: 均参照文献[14]进行。

本文于1991年7月16日收到。  
本工作由七·五攻关项目资助。  
\* 北京农学院畜牧系

2. DNA 序列分析: 参照试剂盒说明及文献[15]进行。

## 结 果

### (一) 牛基因组文库的筛选及阳性克隆的鉴定

用牛 $\alpha$ -s1酪蛋白基因cDNA p184(Pst I -1172bp)为探针筛选了以EMBL3 为载体, Mob I 部分酶切制备的牛基因组DNA 插入片段构建的基因组文库, 共筛选了 $2 \times 10^5$  pfu, 经过三轮筛选, 得到一个阳性克隆 $\lambda$ AZ1。 $\lambda$ AZ1 DNA经Sal I 和BamH I 双酶切得到两条插入片段, 分别长 5kb, 8.7kb, 与p184-1172bp杂交, 8.7kb片段有较强杂交信号, 5kb 片段有较弱杂交信号。如图 1。

### (二) 插入片段亚克隆及酶切图谱的制作

将 $\lambda$ AZ1 DNA Sal I 和BamH I 酶切所得插入片段(如图 1)分别亚克隆至pUC 19的BamHI/Sal I 位点, 形成pSB5和pBS 8.7。pBS8.7 DNA 经 Pst I 酶切产生的 1.7kb, 1.1kb, 0.9kb片段 分别克隆至 pUC19的Pst I 位点, 形成亚克隆pPP1.7, pPP1.1, pPP0.9; pBS8.7经 Pst I 酶切后与载体相连的部分自身环化, 形成pBP 5.0(如图 2)。pBP5.0经EcoR I 酶切所得之 1.4kb, 1.0kb片段(如图 3中 2)分别亚克隆至 pUC19之 EcoR I 位点, 形成亚克隆pEE1.4、pEE1.0。

经EcoR I, Pst I, EcoR I/Pst I 酶切(如图 4), pBS8.7含有 5 个 EcoR I 位点(包括pUC19本身 1 个)将 pBS8.7分为 2.6、1.4、1.0、0.25和6.15kb(3.45kb 插入片段加pUC19载体)等片段; pBS8.7 含 4 个Pst I 位点(包括pUC19本身 1 个)将pBS8.7分为0.9、1.1、1.7和7.7kb(5.0

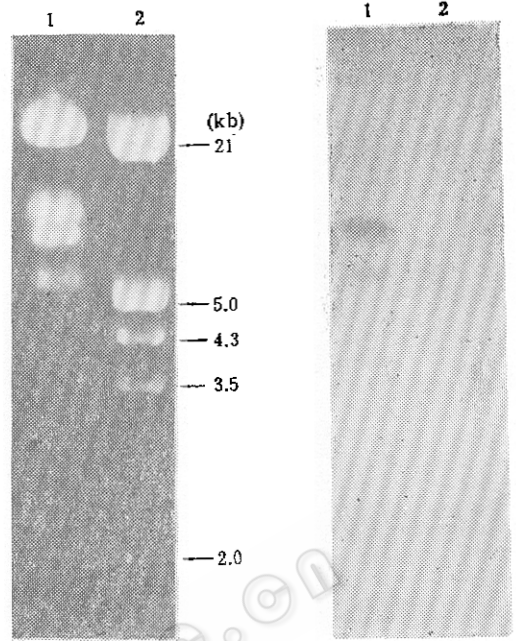


图 1 阳性克隆 $\lambda$ AZ1 DNA限制酶水解及Southern 杂交分析

Fig.1. Restriction endonucleases digestion and Southern hybridization of the positive clone  $\lambda$ AZ1  
Left: 0.5% agarose gel electrophoresis stained with EtBr  
Right: Southern blot hybridized with the probe of  $^{32}$ P-labeled bovine  $\alpha$ -s1 casein gene cDNA (1172bp Pst I fragment of p184)  
1.  $\lambda$ AZ1 DNA + Sal I + BamH I  
2.  $\lambda$ DNA + Hind III + EcoR I

kb插入片段加pUC19 载体)等片段; pBS 8.7的0.25kb EcoR I 片段的一端位于1.7 kb Pst I 片段内,另一端紧靠Pst I 位点。

经 BamH I /Bcl I、EcoR I 部分切/ Bcl I, Pst I 部分酶切(如图 5)结果表明 pBS8.7 插入片段含 1 个 Bcl I 位点, 该 Bcl I 位点不在EcoR I 片段内, 距0.25kb EcoR I 片段一端0.35kb; pBS8.7插入片段EcoR I 片段的顺序为 2.6kb-1.4kb-1.0kb-0.25kb-3.45kb; Pst I 片段的顺序为 5.0kb-1.7kb-1.1kb-0.9kb。

pBP5.0 经 Hind III, BamH I /Hind III 酶切, 如图5中4,5,结果表明距 BamH I

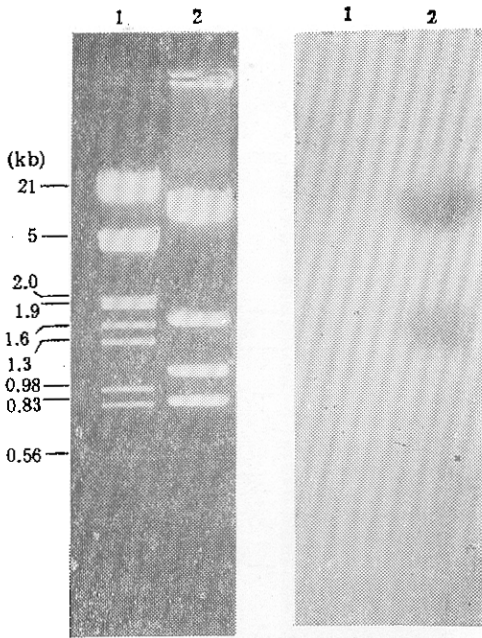


图2 亚克隆pBS8.5 DNA限制酶酶解及Southern杂交分析

Fig.2 Restriction endonucleases digestion and Southern hybridization of the subclone pBS8.5

Left, 0.8% agarose gel electrophoresis stained with EtBr

Right, Southern blot hybridized with the probe of  $^{32}\text{P}$ -labeled bovine  $\alpha$ -s1 casein gene cDNA (1172bp Pst I fragment of p184)

1.  $\lambda$ DNA + Hind III + EcoRI

2. pBS8.5 DNA + Pst I

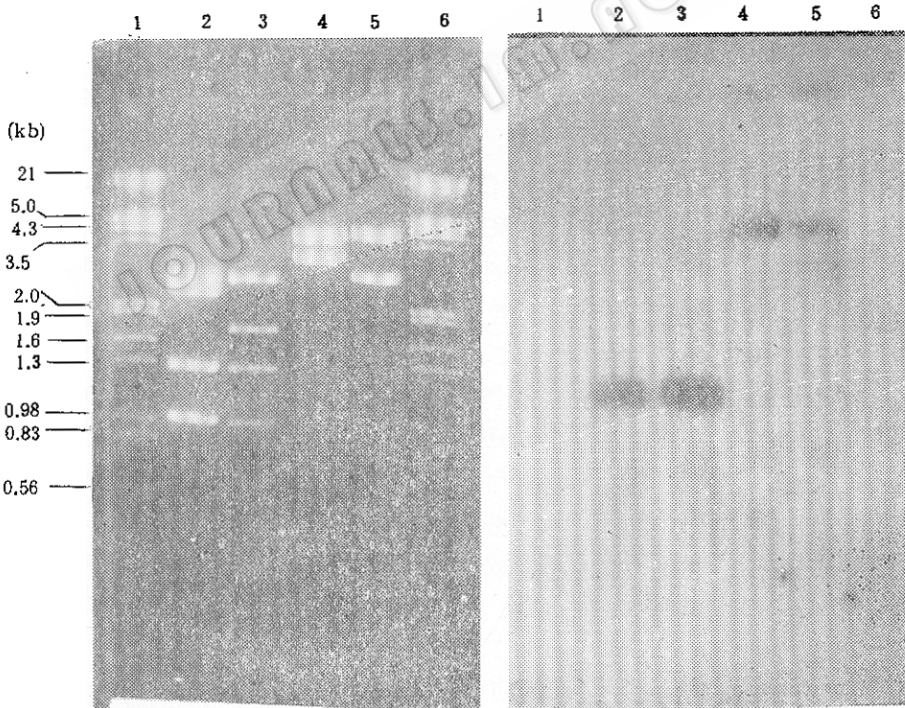


图3 亚克隆pBP5.0 DNA限制酶酶解及Southern杂交分析

Fig.3 Restriction endonucleases digestion and Southern hybridization of the subclone pBP5.0

Left, 1.0% agarose gel electrophoresis stained with EtBr

Right, Southern blot hybridized with the probe of  $^{32}\text{P}$ -labeled bovine  $\alpha$ -s1 casein gene cDNA (1172bp Pst I fragment of p184)

1,6.  $\lambda$ DNA + Hind III + EcoRI. 2. pBP5.0 DNA + EcoRI. 3. pBP5.0 DNA + EcoRI + Hind III,

4. pBP5.0 DNA + Hind III. 5. pBP5.0 DNA + Hind III + BamHI

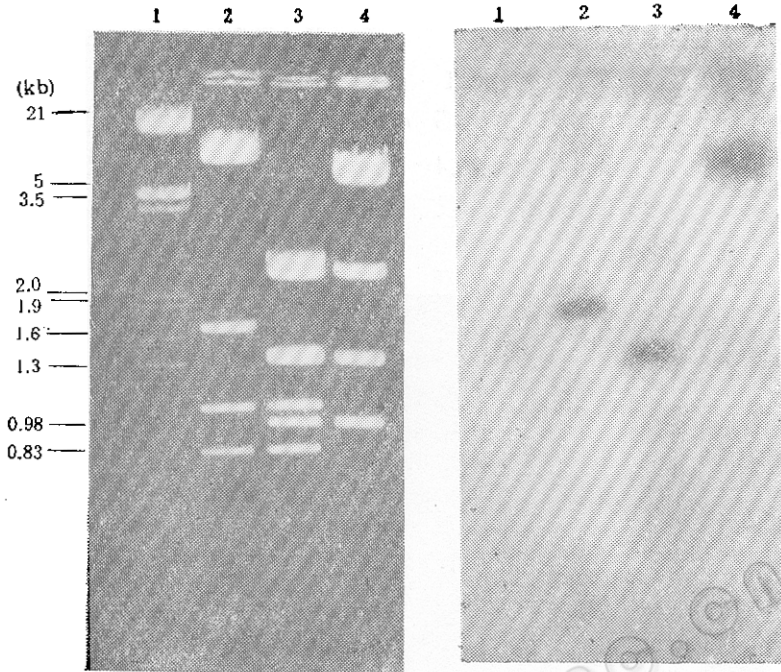
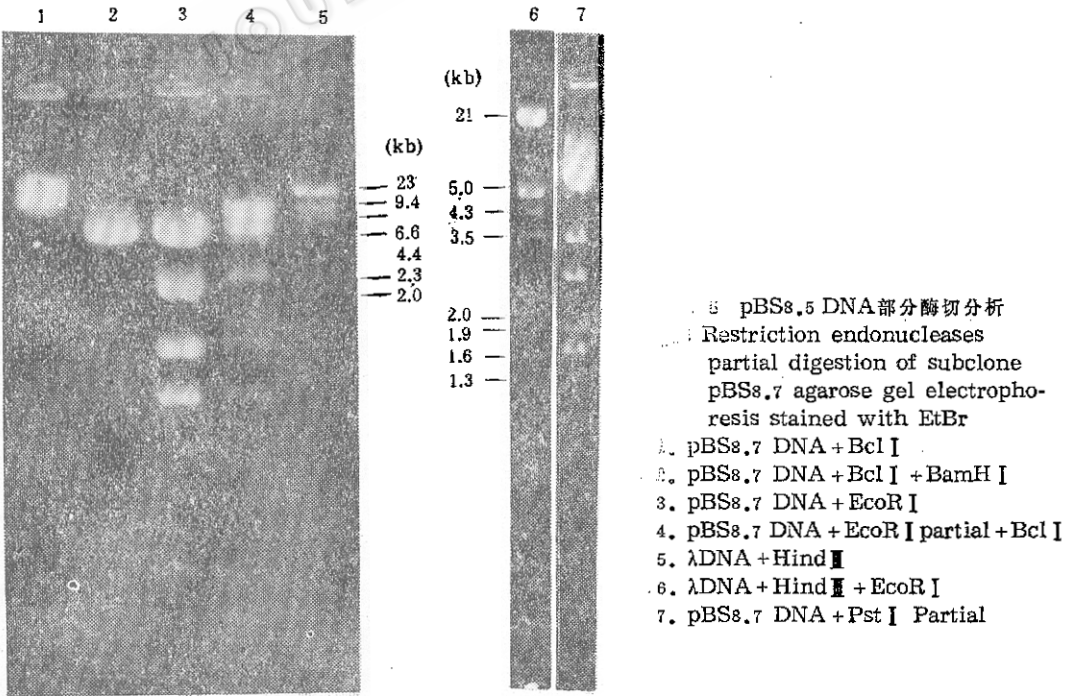


图 4 亚克隆pBS8.7 DNA限制酶降解及Southern杂交分析

Fig.4 Restriction endonucleases digestion and Southern hybridization of the subclone pBS8.5  
 Left: 0.8% agarose gel electrophoresis stained with EtBr  
 Right: Southern blot hybridized with the probe of  $^{32}$ P-labeled bovine  $\alpha$ -s1 casein gene cDNA which the fragment before terminator was removed(400bp Pst I/EcoR I fragment of p184)

1.  $\lambda$ DNA + Hind III + EcoR I , 2. pBS8.7 DNA + Pst I ,
3. pBS8.7 DNA + Pst I + EcoR I , 4. pBS8.7 DNA + EcoR I



5. pBS8.5 DNA部分酶切分析
6.  $\lambda$ DNA + Hind III + EcoR I
7. pBS8.7 DNA + Pst I Partial

位点0.6kb处有一Hind III位点。

pBS8.7, pEE1.4, pEE1.0经Kpn I酶切, 结果如图6中5, 图7所示, pBS8.7插入片段含2个Kpn I位点, 相距0.4kb, 分别位于pEE1.4, pEE1.0插入片段

内, EcoR I位点分隔Kpn I片段为0.1kb, 0.3kb。

综上所述,  $\lambda$ AZ1插入片段限制酶图谱如图8中1所示。与已发表的资料相比较, 有少数位点的变异, 如图8中1—3。

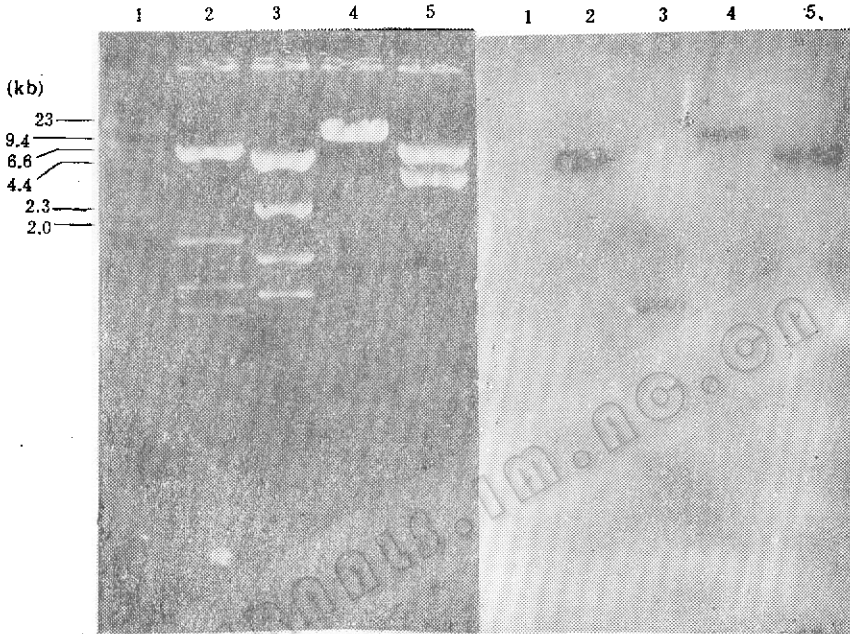


图6 亚克隆pBS8.7 DNA限制酶酶解及Southern杂交分析

Fig.6 Restriction endonucleases digestion and Southern hybridization of the subclone pBS8.7  
Left: 1.0% agarose gel electrophoresis stained with EtBr

Right: Southern blot hybridized with the probe of  $^{32}$ P-labeled bovine  $\alpha$ -s1 casein gene cDNA which the fragment after terminator was removed (700bp Pst I/EcoR I fragment of p184)

1.  $\lambda$ DNA + Hind III, 2. pBS8.7 DNA + Pst I, 3. pBS8.7 DNA + EcoR I,
4. pBS8.7 DNA + Bgl II, 5. pBS8.7 DNA + Kpn I

### (三) 克隆片段方向的确定

分别以牛 $\alpha$ -s1酪蛋白基因cDNA全长片段(p184 1172bp Pst I片段)、终止符前片段(p184 770bp Pst I/EcoR I片段)、终止符后片段(p184 400bp Pst I/EcoR I片段)为探针, 分别与 $\lambda$ AZ1, pBS8.7, pBP5.0各相应酶切片段杂交, 确定 $\lambda$ AZ1插入片段的方向, 即:  $\lambda$ AZ BamH I / Sal I 5kb片段靠近5'端, 8.7kb片段靠近3'端。

### (四) DNA序列分析

对可能含有牛 $\alpha$ s1酪蛋白基因终止信号的pEE1.0, pPP1.7插入片段分别再进行亚克隆, 并测定了相应的DNA序列, 确定了终止符前的一个外显子(外显子17), 含有终止符的外显子(外显子18), 和含有polyA加成信号的外显子(外显子19)部分的位置<sup>[10]</sup>。与现有的该基因的资料相比较<sup>[10]</sup>, DNA序列有多处的点突变和少量缺失, 某些酶切位点差异是由

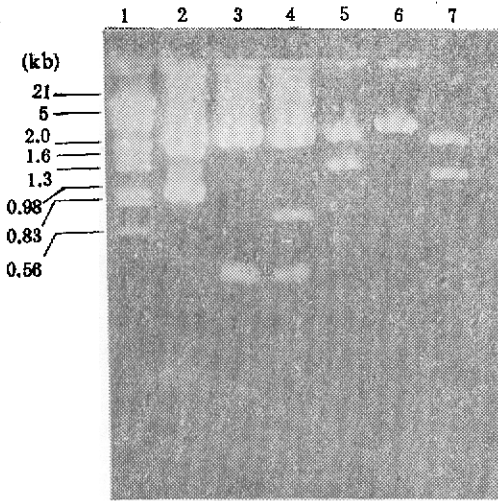


图 7 pEE1.4和pEE1.0限制酶酶切分析  
Fig.7 Restriction endonucleases digestion of subclone pEE1.4 and pEE1.0  
1.2% agarose gel electrophoresis stained with EtBr

1.  $\lambda$ DNA + Hind III + EcoR I
2. pEE1.0 + EcoR I
3. pEE1.0 + Kpn I
4. pEE1.0 + EcoR I + Kpn I
5. pEE1.4 + EcoR I
6. pEE1.4 + Kpn I
7. pEE1.4 + EcoR I + Kpn I

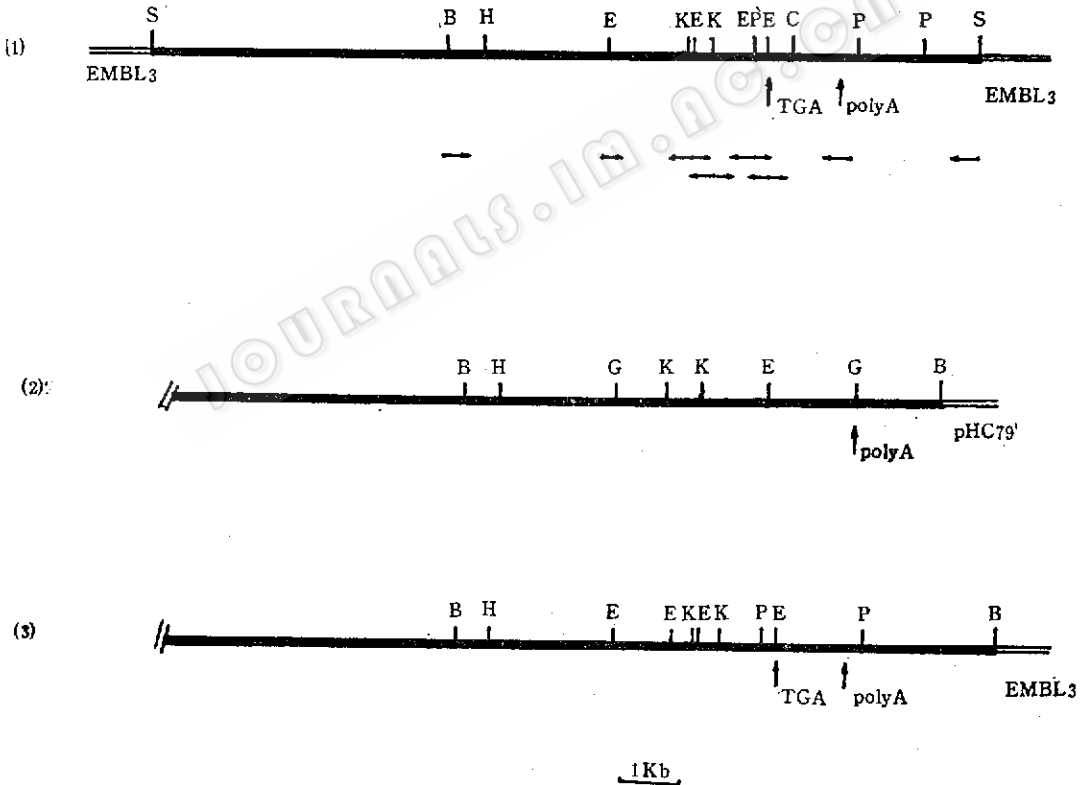


图 8 牛 $\alpha$ -s1酪蛋白基因克隆 $\lambda$ AZ1酶切图谱及其与已发表资料的比较

Fig.8 Restriction endonuclease map of the 3'-flanking region of bovine  $\alpha$ -s1 casein gene ( $\lambda$ AZ1) and comparison with those which had been published

B. BamH I; C. Bcl I; E. EcoR I; G. Bgl I; K. Kpn I; H. Hind III; S. Sal I;  
 ↑ PolyA, Polyadenylation site; ↑ TGA, Terminator; ←, Sequencing strategy;  
 □, Vector sequences; ■, Bovine  $\alpha$ -s1 casein gene and its down stream fragment  
 (1)  $\lambda$ AZ1, (2) by Meade(1990)<sup>[5]</sup>, (3) by Koczan(1991)<sup>[10]</sup>

1 AAAATTCAG ATATCATATT TATTATAATC ATTAGGATAG CCAACCACAG CAAAACITGA  
 2 ----- 14849

1 AATGCTTICT TTGGTTTTTC TAATCCATC AAAGAACATG TATTTTCTAC CTGAATATAT  
 2 ----- 14909

1 AAAGAATTAC ATTAATACAC AGAAACAAAT AAAAGGCCAT GATAGATTAA CAGAGGTACC  
 2 ----- 14989

1 TATTCAGAAA TTTCAGAAA GAAATAATGT TTTTGCACCA AACCTTATA GTTATGAAGG  
 2 ----- 15029

1 GGCAC TGGC TAGGAATTAG AGAATTCCTT TCTCTTTATT TTTGGCAITTT TTGAAATCCC  
 2 ----- C----- C----- 15089

1 TGTATTATTC TGGGACGAGG ACTCTCCCTT TCTTTGAGCT CATTGTCTTC ATGGAAAAC T  
 2 ----- 15149

1 AAAC TATAGA ACAGAAAAGC AAAATAACC TTCCTAGTC CAAAAACCA TGAGTTTATA  
 2 ----- G----- C----- 15208

1 CATTAGATAA TAATTA AAT CACACTACCA CTTTCCCTCG GACTTTGATA GCTGGACCTG  
 2 C----- I----- A----- 15289

1 TGTTAGTAAT GCAGGTICTT AGAGTTCAAT ACTGACTATT GATCTTAAAG AAGTGAITTA  
 2 ----- 15329

1 TTCATTACCC CTCAGGTACC CTAAGAAAAG CCGTTTCCAA AATTTGCTTA CTGTGACTTA  
 2 ----- C----- 15389

1 CCATAGGGAA GAGTTTGATT TCCTGTGTTT CCTCACAGAG TAAACATCTC TTGTGATGGC  
 2 ----- 15449

1 AATAGCCAATG TCTGAAATGA AGGCAATGAT TCATTTTCAG AGATTCAAAA CTGATTTCTC  
 2 ----- 15509

1 AFACACTGTT GCTTTTTCAA TGGTCTTTCT CTCTAGCTTT TCAGACAATT CTACCAGCTG  
 2 ----- 15589

1 GATGCCATC EATCTGGTGC CTGGTATTAC GTTCCACTAG GCACACAATA CACTGATGCC  
 2 ----- 15629

1 CCATCAITCT CTGACATCCC TAATCCCATT GGCTCTGAGA ACAGTGAAAA GACTACTATG  
 2 ----- 15689

1 CCACTGTGCTT CGTAAGTTCA TTTAAATGAC TGCATATTGT TGCTTATCA AAGGAAATAA  
 2 ----- 15740

1 AAGAAAACAT AATATAAAAA TAGAATTAGA ATAGCATGA CACATAAATG CTTAGTGTCC  
 2 ----- 15809

1 TATACTACAA TTTTCTGAAA TGGAAAATG ATGATAACTT TCTGGTATAT GGCTAATGTT  
 2 ----- 15869

1 AATCCATTAC TCAGGAACAT GTGGAGCAGT GCTATCTATT CGATAAGTGA TAATCAITCT  
 2 ----- 15929

```

1 GATGAAAATA CGAAAATTTT CTCTCCAAAG TAAAAATCA ACTTTATCCT CCTTCCACTT
2 ----- 15989

1 TTGCTAATCT TAAAATGCCT TTCTTCTGAT TATATCCATG ATATACATTA GAATTC AATG
2 ----- -T----- -G-T- 16049

1 TGGGGGATAA ACTGCAGATT TTGGCATTCC TAAAGTCCCA AITTGAAATC CTGATCTTTT
2 ----- -C----- 16109

1 TTTTTTTTTT TTT//GCTTA CTTGAAATAG TATA/TGATG CTTGTITTTA TAACCTTGAA
2 ----- -TT----- -A----- 16169

1 GGTGATTAAA TATAATAATC CATTAAAGCAT ACTGCTGGCA AAATTAGTGC TCATTTTTTC
2 ----- 16229

1 ATTTAGCAA ATTTTATTAC TGAATGCCIT ACTCACATTT TACCAATTTT TGCCTCTCA
2 -C----- -A----- 16289

1 CTAGAGTCAA GTGAATTCTG AGGGACTCCA CAATTATGGT CTTTGGTAAG TTGGAAACTG
2 ----- -G----- 16349

1 CTTGTCTAAT CATTGATCCT CTTTTCATAT GAGAGCTCAG TACAAAAGTA CAACGTGTAG
2 ----- 16409

1 ACTATAAAGT TGTITGCTG GTCTCTAGT CTAGCTATAT TTAGACACAT TACACTTAGA
2 ----- -A----- 16469

1 TAATATCAAT AATTAAATG GCTTCAACAT TTTTGTATT ATAGTAATAT CAAATTTAAG
2 ----- 16529

1 TAGCAATAAA ACTAGCTAAT TTTAATTAAT ATAATTTATT ATATTGAAAT TTTTATACAA
2 ----- 16589

1 GTATAGGAGT GTGGGTGTTA CTAATCTGG TGACCCCAAC AGTGGAAATCC TATTCITTAT
2 ----- 16649

1 GATCA ..... unsequenced..... CTTCATIAAA TTTCTGTATG GAAAATGTTT
2 --- ... intron-18,exon-19 --- ----- 17399

1 TAAAAGCCTT TGAATCACTT CTCTGTAAG TGCCATCATA TCAAATAATT GTGTGCATTA
2 ----- 17459

1 ACTGAGATT IGTCTTICTT CTTTCAATA AATTACATTT TAAGGCACTA TTCCTATTTT
2 ----- -G----- 17519

1 TTGCAITAT ICCAATGGAA GGAATTTACA CAACCTTGTG AGITTTGTAIG TATAAACAT
2 ----- -G----- 17579

1 TTTGTTTICA CTAAATTTTT ATGACATTTT CAACCACATT TTAATGAAAA AATICA AATG
2 ----- 17639

1 TTCCTTCTA GCTGATCCTG GTAGATTATA AACTGAGTCT AAGGTCTTTC ATTGAAGTC
2 ----- -A----- 17699

1 AACTGTTTAT AGAATATTTT CCATGTGGAC ATGGACATGG CTGCAG
2 ----- -A----- -G----- 17745

```

图 9 牛 $\alpha$ -s1酪蛋白基因3'下游区部分序列分析及比较  
 with those which had been published  
 Exons are underlined, terminator and polyadenylation signal boxed  
 1, A AZ1; 2, by Koczan(1991)1101



点突变所引起, 如图 9。

## 讨 论

牛  $\alpha$ -S1 酪蛋白基因结构基因部分长 17.5kb, 上下游可能起调控作用的部分也很长<sup>[5,10]</sup>。我们曾克隆了牛  $\alpha$ -s1 酪蛋白基因的 5' 端及其上游区<sup>[12]</sup>, 为构建乳腺特异性表达异源基因的框架, 我们还希望得到该基因的 3' 端部分。在酪蛋白基因中, 除了 5' 非翻译区、信号肽序列和酪蛋白激酶识别位点序列非常保守外, 其 3' 端最初数十个碱基也相当保守<sup>[18]</sup>。在几乎所有酪蛋白 cDNA 中, 终止符 (TGA) 恰好位于 EcoR I 识别序列 (GAATTC)

处<sup>[17-20]</sup>, 这为我们寻找牛  $\alpha$ -s1 酪蛋白基因终止信号提供了方便。经 Southern 杂交, 我们认为终止符及 polyA 加成信号可能位于 pEE1.0 和/或 pPP1.7kb 内。对 pEE1.0 和 pPP1.7 插入片段分别再进行了亚克隆并分析了序列, 确定了终止符 (TGA) 及 polyA 加成信号 (AATAAA) 的位置。

将  $\lambda$ AZ1 酶切图谱及近 3' 端序列与已发表的资料<sup>[6,10]</sup> 相比较, 发现存在一些限制酶位点的差异, DNA 序列有多处点突变, 点突变不仅在内含子中有, 外显子中也有, 这是引起限制性酶切位点差异的原因。但从总体看, 牛  $\alpha$ -s1 酪蛋白的基因 3' 端具有很高的保守性。

## 参 考 文 献

- [1] Simons, J.P. et al, *Nature*, 328:530—532, 1987.
- [2] Wright, G. et al, *Bio/Technology*, 9:830—834, 1991.
- [3] Ebert, K. M. et al, *Bio/Technology*, 9:835—838, 1991.
- [4] Buhler, Th. A. et al, *Bio/Technology*, 8:140—143, 1990.
- [5] Meade, H. et al, *Bio/Technology*, 8:443—446, 1990.
- [6] Andres, A. C. et al, *Proc Natl Acad Sci USA*, 84:1299—1303, 1987.
- [7] Palmiter, R. D. et al, *Proc Natl Acad Sci USA*, 88:478—482, 1991.
- [8] Choi, T. et al, *Mol. Cell Biol.*, 11:3070—3074, 1991.
- [9] Falcone, D. et al, *Mol. Cell Biol.*, 11:2656—2664, 1991.
- [10] Koczan, D. et al, *Nucl. Acids Res*, 19:5591—5596, 1991.
- [11] Yu-lee, L. Y. et al, *Nucl. Acids Res*, 14:1883—1902, 1986.
- [12] 陈瑞环等: 生物工程学报, 8(3):218—226, 1992.
- [13] 陈瑞环等: 生物化学及生物物理进展, 18:378, 1991.
- [14] Sambrook, J. et al, *Molecular Cloning*, Second edititon, New York, Cold Spring Habor Laboratory Press, 1989.
- [15] Wong, W. M. et al, *Nucl. Acids Res.*, 18:5573, 1990.
- [16] Bonsing, J. et al, *Aust. J. Biol. Sci.*, 41:527—537, 1988.
- [17] Stewart, A. F. et al, *Nucl. Acids. Res.*, 12:3895—3907, 1984.
- [18] Stewart, A. F. et al, *Mol. Biol Evol.*, 4:231—241, 1987.
- [19] McKnight, R. A. et al, *J. Dairy Sci.*, 72:2464—2473, 1989.
- [20] Boisnard, M. et al, *Biochimie*, 67:1043—1051, 1985.

## Cloning and Sequencing of 3'-flanking Region of Bovine $\alpha$ -s1 Casein Gene

Chen Ruihuan Wang Bo Zhang Yuzhi Liu Wei

Zhang Jingpu Lao Weide

(*Institute of Biophysics, Academia Sinica, Beijing*)

A positive clone was isolated from a bovine genomic library in EMBL3 with the probe of  $\alpha$ -S1 casein gene cDNA. The results from restriction endonucleases digestion, Southern hybridization with the probes of upstream and downstream from terminator of bovine  $\alpha$ -S1 casein gene cDNA suggested that the clone contained 3.5kb downstream and 10.2kb upstream fragments from the terminator of  $\alpha$ -s1 casein gene. Some divergence in restriction endonucleases sites, a few deletions and many point mutations were found in introns and exons when compared with those which had been published.

### Key words

Bovine  $\alpha$ -s1 casein, 3'-flanking region, sequence analysis