

• 研究报告 •

牛 α S1 酪蛋白基因启动区的克隆和序列分析

李 宁 吴常信 陈永福

(中国农业大学农业生物技术国家重点实验室, 北京 100094)

摘要 牛 α S1 酪蛋白是牛奶蛋白的最主要成份。本研究利用 λ EMBL3 载体构建了牛基因组文库, 并且从文库中分离克隆了牛 α S1 酪蛋白基因的启动区。利用自动测序仪, 对牛 α S1 酪蛋白基因 5' 侧翼 +298~-1082 的核苷酸顺序作了测定。经过与牛和其他物种的奶蛋白基因序列比较, 推断了牛 α S1 酪蛋白基因的乳腺组织特异性转录因子和一般性核转录因子的结合位点。此外, 本文还讨论了牛 α S1 酪蛋白基因启动区的利用前景。

关键词 牛基因组文库, α S1 酪蛋白基因, 启动区, 序列分析

Molecular Cloning and Sequencing of the 5' Flanking Region of Bovine α S1 Casein Gene

Li Ning Wu Changxin Chen Yongfu

(National Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100094)

Abstract Bovine α S1 casein is the most abundant bovine milk protein. A recombinant bacteriophage containing the entire 5' flanking region of bovine α S1 casein gene was isolated from a bovine genomic library constructed with bacteriophage λ EMBL3. The nucleotide sequence ranging from +298 to -1082 of bovine α S1 casein gene was determined with a DNA sequencer. The putative binding sites of mammary gland specific transcriptional factors and general nuclear transcriptional factors in bovine α S1 casein gene were found out by the consensus sequence comparison with other milk genes from bovine and other animal species. Furthermore, the potential utilization of bovine α S1 casein gene promoter in many aspects has been also discussed in the paper.

Key words Genomic library, Bovine α S1 casein gene, Promoter, Sequence analysis

酪蛋白是哺乳动物奶蛋白的主要成份。牛奶中存在着 4 种酪蛋白, 分别为 α S1 酪蛋白、 α S2 酪蛋白、 β 酪蛋白和 κ 酪蛋白, 其中 α S1 酪蛋白在牛乳蛋白中含量最高, 约为 13 克/升⁽¹⁰⁾。牛的 4 种酪蛋白基因成簇地位于牛第 6 号染色体, 物理长度在 200~300kb 之间, 基因排列顺序为: α S1 酪蛋白基因— β 酪蛋白基因— α S2 酪蛋白基因— κ 酪蛋白基因^(4,14)。目前, 对牛 β 酪蛋白基因和 α S2 酪蛋白基因的表达调控研究均有报道, 并在这些基因的启动区发现了乳腺特异性转录因子的结合位点和其他一些重要的调控元件^(1,5,6), 但对牛 α S1 酪蛋白基因的表达调控研究仍是空白。据估计, 牛酪蛋白基因座位存在座位控制区域(Locus Control Region), 因而使得酪蛋白基因可以成为一个完全独立的表达单位⁽⁹⁾。由于座位控制区域通常位于基因座位的 5' 侧翼, 显然对 α S1 酪蛋白基因启动区进行分析或以此为基点进行巡查将更易证实座位控制区域的存在。另一方面, 利用转基因技术在牛

乳腺中生产人类药用蛋白是生物反应器技术中最活跃和最有商业化前景的研究领域之一^{〔3,12〕}。要在乳腺中表达人类药用蛋白基因,转基因结合中必需有一个高效的、乳腺特异性表达的启动区,显然,牛 α S1 酪蛋白基因的启动区是最好的候选启动区。基于这些设想,在本研究中我们首先从牛基因组文库中克隆了 α S1 酪蛋白基因的启动区,并对启动区的部分核苷酸序列作了分析,为进一步研究奠定了基础。

1 材 料 与 方 法

1.1 实验材料和生化试剂

牛肝组织购自屠宰场。 λ EMBL3 和 pGEM-7Zf 克隆载体及相应的宿主菌购自 Promega 公司。限制性内切酶、DNA 修饰酶和其他生化试剂均购自商业公司。 $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ 和 $[\gamma\text{-}^{32}\text{P}]\text{TAP}$ 购自福瑞公司。

1.2 探针制备

根据有关参考文献^{〔18〕},由本室化学合成寡聚核苷酸作为探针,序列如下:

5' -TAGCTTGGTGGTATAATTAATAAAT-3', 位于牛 α S1 酪蛋白基因 5' 侧翼-547 至-569 的位置。合成的寡聚核苷酸通过末端标记³²P 作为筛选文库的探针。

1.3 牛基因组文库的构建和筛选

牛基因组 DNA 从肝组织细胞中提取,构建文库的方法按照参考文献〔13〕进行。在筛选文库时,每个平板(直径 9 厘米)含有 2 000~3 000 个噬菌斑,利用经末端标记³²P 的寡核苷酸探针进行原位杂交筛选。

1.4 Southern 杂交分析

利用牛 α S1 酪蛋白基因阳性克隆的 DNA 片段作为探针时,标记³²P 的方法为随机引物法,杂交温度为 62℃。Southern 杂交的操作程序同文献〔13〕。

1.5 序列分析

作核苷酸序列测定的亚克隆通过 Erase-a-Base 系统(Promega 公司)产生,测序反应按照 Taq Dye 试剂盒(PE 公司)说明书进行,核苷酸顺序确定在 370A 全自动序列仪(ABI 公司)上实现,序列比较分析利用 DNASIS(Hitachi 公司)软件系统进行。

2 结 果

2.1 牛 α S1 酪蛋白基因启动区阳性克隆的分离

牛基因组文库的重组噬菌体总数为 6.5×10^5 , 大约从其中 2.1×10^5 的重组噬菌体中分离到 2 个阳性克隆,分别命名为 bCAS1 和 bCAS2。阳性克隆经 *Sal*I 酶切分析发现, bCAS1 的 DNA 插入片段约为 9kb, 而 bCAS2 的 DNA 插入片段约为 14kb。通过用 *Hind*III、*Eco*RI、*Bam*HI、*Bgl*II 限制性酶进行单酶或双酶剪切分析 bCAS1 和 bCAS2, 构建了 bCAS1 和 bCAS2 的限制性酶切图谱(图 1)。

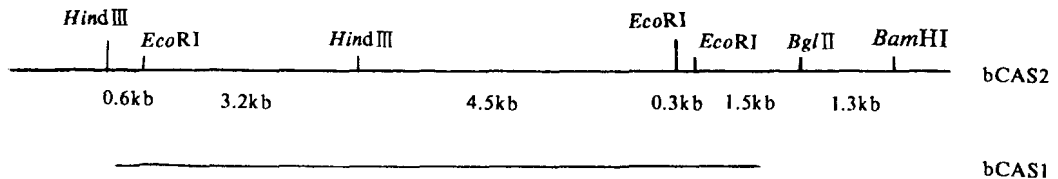


图 1 牛 α S1 酪蛋白基因启动区的限制性酶切图谱

2.2 牛 α S1 酪蛋白基因启动区阳性克隆的序列分析

将阳性克隆 bCAS2 的限制性酶切图谱与参考文献〔7,18〕报道的牛 α S1 酪蛋白基因的限制性酶切图谱相

比, 表明牛 α S1 酪蛋白基因启动区应该位于 bCAS2 克隆的 *Bgl*II 酶切位点的 5' 端。将 bCAS2 4.5kb *Hind*III-*Eco*RI DNA 片段、0.3kb *Eco*RI-*Eco*RI DNA 片段和 1.5kb *Eco*RI-*Bgl*II DNA 片段插入于 pGEM-7Zf 质粒的多克隆位点上, 相应的亚克隆分别命名为 pCASP1、pCASP2 和 pCASP3。我们对 pCASP1 和 pCASP3 插入 DNA 片段的部分核苷酸序列作了测定, 同时也测定了 pCASP2 插入 DNA 片段的全部核苷酸序列。经过与已发表的牛 α S1 酪蛋白基因的核苷酸序列^[7, 18] 进行比较, 发现本研究测定的牛 α S1 酪蛋白基因 5' 侧翼核苷酸序列是在 +298~-1082 的位置, 共计为 1380 个核苷酸(图 2)。测定的序列与文献〔7〕报道的序列相比, 变异率是 0.2%; 而与文献〔18〕报道的序列相比, 变异率是 0.7%。

```

-1082                                -1050
      at  tcctttctta taaacaatga gttgcaatca acaagttttt aaagctctca cttgtataga
-1000
tttattttta gcacataata tttttctaca atgtacaatg ccagttaatt ctaggagtc aattaagaat
-950
tggagagata ggaatttttt tcttttactt gtttacttta aaagatggaa aatcagagtt atggttttatt
-850
tttcgcaata tttaaaaatt ataattcttg aataactatt aattttaatt aaataatctg taatgagaat
-800
cctcctacca atgtaggaga cgtgagtttg actcccgggt agggaagata ccctgcagaa ggaaatggca
-700
accactcca atattattac ttgggaaatc ccatggacag aggagactgg caggctgcag tccatggggg
-650
tcacaaagaa ctggacacga cttagaaact aaacaacaac aatttatacc agaatgaatg aactagttac
-600
cacaactagt acacccaaaa tgaacaaaaa atagcttggt ggtataatta aaatgccacc aaaatttata
-500
caataattat attttcttt tgcaggaaaa agattagacc acatataatg taacttattt cacaaggtaa
-450
ataattataa taantaatat ggattaactg agttttaaaa ggtgaaataa ataatgaatt cttctcatgg
-350
tcttgtatgt taataaaaat tgaaaaattt tgaagacccc atttgtccc aagaatttca tttacaggta
MPBF
-300
ttgaattttt caaaggttac aaaggaaatt ttattgatat aataaatgca tgtttctata ataaccataa
-250
atctagggtt ttgttgggtt tttttttgt ttgttaattt agaacaatgc cattccattt cctgtataat
-150
gagtcacttc tttgtttaa actctcctta gaatttcttg ggagaggaaac tgaacagaac attgatttcc
milk box
-100                                -50
aatgtgagag aattccttaga atttaataa acctgttggt taaactgaaa ccacaaaatt agcattttac
MGF                                oct-1
taatcatgag gtttaaatag cttggaagca aaagtctgcc ATCACCTTGA TCATCAACCC AGCTTGCTGC
MPBF
+50                                +100
TTCTTCCCAG TCTTGGGTTT AAGgtattat gtatacatat aacaaaattt ctatgatttt cctctgtctc
+150
atctttcatt cttcactaat acgcagttgt aacttttcta tgtgattgca agtattggta ctttctatg
+200
ataactggt agcttaaaaa tatatttgca aatggtgata ctatctatct cagagctata ggtgaaaaat
+250                                +298
taaatacttt tataagacc aaattgatca ttttttaaac gaaaatttct tatatact

```

图 2 牛 α S1 酪蛋白基因 5' 侧翼核苷酸序列

外显子用大写字母表示; 转录因子结合位点在底部划线

2.3 牛 α S1 酪蛋白基因启动区的 Southern 杂交

随机抽样, 取 4 头牛提取基因组 DNA, 经用 *Eco*RI 和 *Hind*III 双酶完全消化, 然后进行 Southern 杂交, 探针为经过随机引物法标记³²P 的 pCASPI DNA。杂交结果(图 3)表明: 4 头牛的基因组 DNA 均有 4.5kb 的杂交带, 不存在非特异性杂交带。

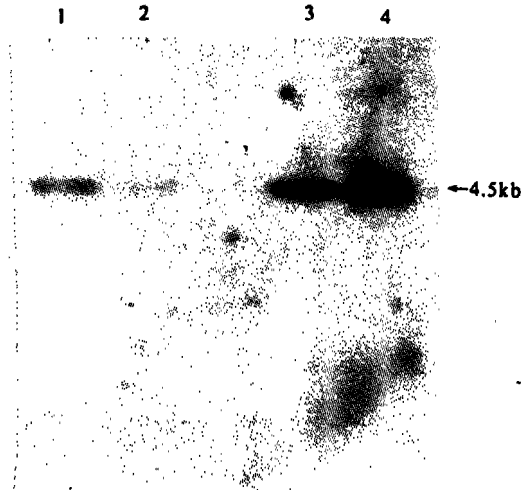


图 3 牛基因组 DNA 的 Southern 杂交结果

探针是 pCASPI 的插入 DNA 片段, 1~4. 4 头不同的牛基因组 DNA / *Hind*III + *Eco*RI

3 讨 论

本研究测定的核苷酸序列中, 包括了牛 α S1 酪蛋白基因的第二个外显子, 为 53 个核苷酸, 没有编码功能, 是牛 α S1 酪蛋白 mRNA 的 5' 非翻译区。两个具有潜在功能的 TATA 盒分别位于 -23~-29 和 -92~-98 的位置, 契合序列是 TTAAAT。通过与牛及其他物种的奶蛋白基因序列进行比较^[6, 11, 18], 推断出牛 α S1 酪蛋白基因的乳腺组织特异性转录因子(MGF)的结合位点可能位于 -87~-100 的位置, 核苷酸序列为 AATTC TTAGAATT。这段序列几乎在所有的奶蛋白基因中都是保守的(表 1)。另一种乳腺特异性转录因子叫做奶蛋白结合因子(MPBF), 首先从乳球蛋白基因启动区中发现^[16]。这种因子在牛 α S1 酪蛋白基因的结合位点可能有两个, 一个位于 -333~-346 的位置上, 其序列是 TGTCCCAAGAATT; 另一个位于 -11~-24 的位置, 序列是 ATAGCTTGGAAGC。在牛 α S1 酪蛋白基因的 -157~-148 的位置, 序列为 TCTCCTTAGAATTTCTTGGG, 叫做“奶盒”(milk box), 是一种类似于 CTF / NF-1 活性的一般性转录因子的结合位点。而另一种一般性转录因子 oct-1 在牛 α S1 酪蛋白基因序列上的核心结合位点可能位于 -47~-54 的位置, 序列则是 AATTAGCA。从以上序列分析的结果来看, 牛 α S1 酪蛋白基因的乳腺组织特异性转录因子结合位点及其他一些重要调控元件均位于 -400~-10 之间。

利用奶蛋白基因启动区进行乳腺特异性表达的转基因实验表明, 只要奶蛋白基因启动区包括了 -500~+500 的区域, 一般就能够进行高水平的乳腺特异性表达^[2, 8, 17]。本研究分离的阳性克隆 bCAS2 包含了牛 α S1 酪蛋白基因 5' 侧翼 10kb 以上的序列, 并且通过与参考文献发表的 α S1 酪蛋白基因酶切图谱相比, 可以推断 bCAS2 的 3' 端则位于牛 α S1 酪蛋白基因的第三个内含子中。因此, 本研究克隆的牛 α S1 酪蛋白基因启动区完全可以在乳腺组织中特异性地高效转录任何结构基因。

表 1 奶蛋白基因的乳腺组织特异性转录因子结合位点的比较

奶蛋白基因	位置 ¹⁾	MGF 结合区域
牛 α 1 酪蛋白	-100	AATTCCTAGAATT
牛 α 2 酪蛋白	-100	ACTTCTTAGAATT
牛 β 酪蛋白	-100	ATTTCTaGGAATT
牛 κ 酪蛋白	-110	tTTTCTTAGAAgT
牛乳球蛋白	-93	GATTCggGGAACC
牛乳清蛋白	-299	GCTTCCTAGAACC
羊 β 酪蛋白	-98	ATTTCTaGGAATT
羊乳球蛋白	-93	GATTCc _g GGAACC
鼠 β 酪蛋白	-100	ACTTCTTGGAATT
鼠 α 酪蛋白	-102	AATTCCTAGAATT
人乳清蛋白	-287	GCTTCCcAGAACC
契合序列 ²⁾		RNTTCYTRGAYY

1) 结合区域的第一个碱基位置; 2) 表示任何核苷酸, R 表示嘌呤, Y 表示嘧啶。

参 考 文 献

- 1 Altiock S, Groner B. β -casein mRNA sequesters a single-stranded nucleic acid-binding protein which negatively regulates the β -casein gene promoter. *Mol. Cell Biol.*, 1994, 14: 6004~6012
- 2 Doppler W, *et al.* Lactogenic hormone and cell type-specific control of the whey acidic protein gene promoter in transfected mouse cells. *Mol. Endoc.*, 1991, 5: 1642~1632
- 3 Ebert K M, *et al.* Induction of human tissue plasminogen activator in the mammary gland of transgenic goats. *Bio/Technology*, 1994, 12: 699~702
- 4 Ferretti L, *et al.* Long range restriction analysis of the bovine casein genes. *Nucleic Acids Res.*, 1990, 18: 6829~6833
- 5 Groenen M A M, *et al.* Multiple octamerbinding sites in the promoter region of the bovine α 2-casein gene. *Nucleic Acids Res.*, 1992, 20: 4311~4318
- 6 Hall L, *et al.* Organization and sequence of the human α -lactalbumin gene. *Biochem. J.*, 1987, 242: 735~742
- 7 Koczan D, *et al.* Genomic organization of the bovine alpha-S1 casein gene. *Nucleic Acids Res.*, 1991, 19: 5591~5596
- 8 Lee K F, *et al.* Tissue-specific expression of the rat β -casein gene in transgenic mice. *Nucleic Acids Res.*, 1988, 16: 5797~5801
- 9 Martin P, Grosclaude F. Improvement of milk protein quality by gene technology. *Livestock Production Sci.*, 1993, 35: 85~115
- 10 Mercier J, Vilotte J. Structure and function of milk protein genes. *J. Dairy Sci.*, 1993, 76: 3079~3089
- 11 Provot C, *et al.* Complete sequence of the bovine β -casein-encoding gene and interspecies comparison. *Gene*, 1995, 153: 259~263
- 12 Rexroad C E. Transgenic technology in animal agriculture. *Animal Biotech.*, 1992, 3: 1~13
- 13 Sambrook J, *et al.* Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989
- 14 Threadgill D W, Womack J E. Genomic analysis of the major bovine milk protein genes. *Nucleic Acids Res.*, 1990, 18: 6935~6942
- 15 Wagner V A, *et al.* DNA variants within the 5' -flanking region of milk-protein-encoding genes. *Theor. Appl. Genet.*, 1994, 89: 121~126
- 16 Watson C J, *et al.* Interaction of DNA-binding proteins with a milk protein gene promoter *in vitro*: identification of a mammary gland-specific factor. *Nucleic Acids Res.*, 1991, 19: 6603
- 17 Wright G, *et al.* High level expression of active human alpha-antitrypsin in the milk of transgenic sheep. *Bio/Technology*, 1991, 9: 830~834
- 18 Yu-Lee L, *et al.* Evolution of the multigene family: conserved sequences in the 5' flanking and exon regions. *Nucleic Acids Res.*, 1986, 14: 1883~1903

1995-10-31 收稿, 1996-04-11 修回。