

山羊 β -乳球蛋白基因 5'侧区的 克隆、测序和序列分析

潘 玲* 陈建军 陈常庆

(中国科学院上海细胞生物学研究所 上海 200031)*

(中国科学院上海生物工程研究中心 上海 200233)

摘 要 采用 PCR 方法,首次从中国山羊肝组织扩增出 β -乳球蛋白 (β -Lactoglobulin, β LG) 基因的 5'侧区 867bp 片段,其中包括 770bp 的启动子和 97bp 的部分第一外显子,再克隆测序。然后,我们用计算机分析比较了山羊 β LG 基因的 5'侧区与绵羊的和乳牛的同源性,分别为 94.6% 和 88%。还发现上述三者该区域中的转录因子结合位点及其序列都很相似。而己知绵羊与乳牛的 β LG 基因启动子可指导外源基因在转基因动物中组织特异性地高效表达,这提示了山羊 β LG 基因启动子可能亦足以指导外源基因高效表达。本工作为以后用山羊 β LG 基因启动子,指导外源基因在转基因动物乳腺中高效表达打下了一定的基础。

关键词 山羊 β -乳球蛋白基因, 启动子, DNA 序列分析, PCR 扩增

用转基因动物的方法,生产真核目的基因蛋白,是近年来生物工程发展的重要方向,转基因乳腺表达更是倍受青睐。欲使外源基因在乳腺特异表达,则首先需要乳蛋白基因启动子作为指导。在转基因动物乳腺表达中使用的启动子主要有小鼠乳清酸蛋白 (WAP) 启动子^[1]、羊 β 乳球蛋白 (β -Lactoglobulin) 启动子^[2]、牛 α s1 酪蛋白 (α s1 casein) 启动子^[3]、 β 酪蛋白 (β casein) 启动子^[4]等,部分已证明可指导外源基因高效表达。而目前常用的转基因对象有山羊、绵羊、奶牛、兔等,考虑到转基因成本及生长周期与产量等因素,对我国来说,选用山羊为转基因对象可能更为合适。

鉴于 β LG 蛋白为羊、牛、猪、马等乳汁中的主要蛋白;同时,世界上许多实验室对羊 β LG 的组织特异性元件、激素调节等已作了许多研究^[5,6];而且,无论是转绵羊 β LG 片段^[2]还是绵羊 β LG 基因融合外源基因,都已获得高水平表达,如绵羊 β LG 融合人 α 1 抗胰蛋白酶 (α 1AT) 基因,在转基因羊中最高表达量超过 60mg/ml^[7],迄今仍为转基因动物领域中最高的表达量,用羊 β LG 指导人抗凝血因子 IX 也获得了表达^[8]。尽管采用绵羊 β LG 启动子指导外源基因已有高表达的例子,但使用山羊 β LG 启动子则可能更有利于指导外源基因在山羊乳汁中高效表达。

为此,我们首次从中国山羊肝组织,采用 PCR 扩增方法,获得了 β LG 基因的启动子及部分第一外显子序列(包括信号肽)共 867bp。并且与绵羊、乳牛 β LG 基因作了同源性比较,并分析了 β LG 5'侧区转录因子结合位点等,发现它们之间同源性很高,

“八、五”羊个体表达项目基金资助。

本文于 1996 年 5 月 24 日收到。

这提示了山羊 β LG 启动子可能亦足以指导外源基因, 在转基因动物中组织特异性地高水平表达。

1 材料和方法

1.1 材料

新鲜山羊肝组织购自上海郊区。绵羊新鲜血取自中科院上海动物中心。pUC19, M13mp19, JM103 等质粒、菌种由本实验室提供。Taq 酶购自英国 Amersham 公司。各种限制酶、T4 Ligase 购自德国 Boehringer 公司。蛋白酶 K, dNTP、RNA 酶、X-gal、IPTG 购自中国华美公司。DNA 序列试剂盒购自美国 USB 公司。³⁵S-dATP 购自英国 Amersham 公司。低熔点琼脂糖凝胶等主要生化试剂为美国 Sigma 公司或国产 AR 级试剂。DNA 自动测序仪为美国 ABI 381 型。

1.2 方法

1.2.1 组织 DNA 制备: 将山羊肝组织切成小块, 置入液氮。取出一部分, 研成细末, 移入 5ml 离心管, 加抽提缓冲液 (10mmol/L Tris-HCl, pH8.0; 0.1mol/L EDTA pH8.0; 0.5% SDS)、蛋白酶 K (终浓度为 100 μ g/ml), 混匀, 50 $^{\circ}$ C 水浴 5h, 时而旋动。再取出冷却到室温, 酚、氯仿抽提 2 次, 无水乙醇沉淀, 70% 乙醇洗 3 次, 干燥, TE 溶解, 测定 A₂₆₀、A₂₈₀ 方法见文献 [9]。

1.2.2 PCR 引物设计及 PCR: 由于绵羊 β LG 的 800 多 bp 的启动子序列和结构基因序列已有报道^[10], 考虑到结构基因序列应该变异很小, 因而选择在绵羊 β LG 信号肽 97bp 处设计了一个相同序列的 21bp 的反向引物 (引物 2), 交盖的信号肽部分也可作为相近启动子的确证。对于启动子区 DNA 序列, 已经知道牛乳球蛋白启动子^[11]和绵羊乳球蛋白启动子^[10]序列之间的同源性高达 89%, 估计山羊乳球蛋白启动子 DNA 序列应比之具有更高的同源性, 因而同样按照已报道的绵羊 β LG 启动子设计了一个 21bp 的上游正向引物 (引物 1)。考虑到可能有的 10% 的变异, 因而在 PCR 第一个循环中可采用低的退火温度和延伸温度, 或者将引物的位置再移动几个碱基, 即有可能保证引物 3' 末端与模板之间有足够的碱基配对, 使延伸反应得以顺利进行。

引物 1 ACC CGA ATT CCC GCT GCT CCT 21mer

引物 2 AGG CCT TTC ATG GTC TGG GTG 21mer

为了便于克隆, 分别在引物的两端加上 EcoRI、BamHI 酶切位点并加上 3 个保护碱基。

为了分析 PCR 结果, 以判定绵羊 β LG 序列的保守性, 设立了以绵羊白细胞 DNA 为模板的对照组。反应条件依 Saiki^[12], 如下:

DNA 0.5 μ g, 引物 1、2 为 50pmol, dNTP 2mmol/L 5 μ l, 10 \times Buffer 5 μ l, 加 Taq 酶 2.5u, 总体积 50 μ l。覆盖石蜡油。

98 $^{\circ}$ C 变性 7min 以后进入循环, 55 $^{\circ}$ C 1min, 72 $^{\circ}$ C 延伸 1.5min, 93 $^{\circ}$ C 变性 45s。30 次循环之后, 再 72 $^{\circ}$ C 延伸 7min。

1.2.3 PCR 扩增产物的克隆和测序: 将 PCR 扩增产物电泳分析, 低熔点胶回收片段, 再用 EcoRI、BamHI 酶切, 克隆到 pUC19, 筛选克隆、酶切鉴定, 双脱氧末端终止法

测定序列。

并且分别将 EcoRI/BamHI 片段克隆到 M13 mp18, M13 mp19, 送中国科学院上海植物生理研究所 DNA 自动测序仪测 DNA 序列, 以作为参照。

1.2.4 计算机分析山羊 β LG 基因序列: 用 PCGENE (Version 6.7) 对山羊 867bp 的 β LG 基因序列进行分析, 在 EMBL DATABASE (94) 中与绵羊、乳牛的 β LG 基因进行同源比较, 并搜寻和分析转录因子结合位点及其他特殊位点。

2 结果与讨论

用常规 PCR 方法成功地扩增出了绵羊和山羊 β LG 基因, 因而不必采用低的退火温度或移动引物位置。将绵羊和山羊的 PCR 产物走电泳, 发现它们的 PCR 扩增区带在同样的位置, 大约为 800~900bp (图 1), 可初步分析出绵羊和山羊的 β LG 基因在引物设计位置处保守性较强, 故对羊肝组织 DNA 扩增产物进行克隆、测序结果得全序列为 867bp, 包括 770bp 的启动子和 97bp 的部分第一外显子 (外显子中包括 54bp 的信号肽)。

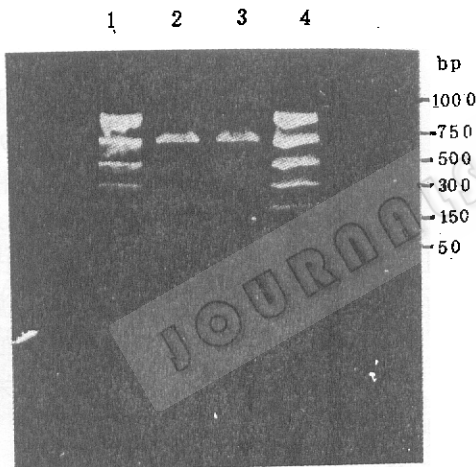


图 1 山羊、绵羊组织 DNA PCR 扩增电泳结果

Fig.1 Result of PCR amplification of goat and sheep tissue DNA

1, 4. PCR Marker, 2. Sheep sample, 3. Goat sample

不同种系的乳蛋白基因保守性很强。如 β -casein 基因, 小鼠^[13]和大鼠^[14]的同源性达到 69%, 与兔子^[15]的同源性为 72%, 而山羊^[4]和牛^[17]的同源性竟达到 95%, 包括 1.7kb 的 5' 侧区和内含子部分。人 α 乳球蛋白基因和大鼠的 5' 侧区非编码区的同源性高达 83%^[18]。牛^[11]和绵羊^[10] β LG 基因的同源性为 89%。所以, 考虑到乳蛋白基因的特点, 我们按绵羊 β LG 基因序列设计引物, 扩增出山羊 β LG 的 5' 侧区及部分第一外显子序列, 获得了成功。用 PCGENE 软件进行 DNA 序列分析、比较, 发现山羊和绵羊 β LG 的同源性达到 94% (表 1), 与乳牛 β LG^[11] 比较, 同源性为 89%。山羊、绵羊 β LG 的转录起始点以后的 97bp 同源性极高, 仅差一个碱基。

已经证实在转基因动物中, 绵羊 β LG 基因的 -406bp 区域足已获得高水平组织特异性表达, 并且以不依赖整合位点, 而与整合拷贝数相关的方式进行表达。用仅包含 -406bp 5' 侧区的绵羊 β LG 片段转基因小鼠, 乳腺表达量为哺乳期绵羊的 5%~300%。对 -5.4kb 的绵羊 β LG 5' 侧区进行序列分析, 发现两个弱乳腺组织特异性 DNAase I 高敏位点, 位于 -1.8kb (HS I) 和 -0.8kb (HS II), 一个强 DNAase I 高敏位点 (HS III), 位于 -300 到 +100bp 区域, 包含了启动子; -406 到 -40 区域内存在 5 个 NFI 结合位点, -133 到 -172bp 为 "milk box"; -406 区域内存在两个乳蛋白结合因子

(MPBF) 的结合位点; 另外还有一些激素结合位点^[5,6]。

表1 山羊和绵羊βLG基因同源序列比较

Table 1 Homology alignment of goat and sheep β LG gene

The two sequences to be aligned are:

GOAT β LG02:867 bp.

OALGB:7379 bp.

DE: OVIS ARIES BETA-LACTOGLOBULIN GENE; OS: OVIS ARIES (SHEEP).

Open gap cost : 5 Unit gap cost : 15

GOATBLG02	-				TTCCG-CTGCTCCTGAGGCTGCAGGC	-26	
OALGB	-	GTGCTCAGCAACACCCCAGCACCAGCA		CG		-55	
GOATBLG02	-	AGCTCGCTGTAGCCTGAGCCGGTGTGGAGGGGAAGTCTCTGGGAGACTTAAAATGT				-81	
OALGB					T	-110	
GOATBLG02	-	GGGAGGTGGGAGGTGGGAGGTTGGCCCTGTGGC-GTGCCCATCCCGTGTGCCTG				-135	
OALGB		A C		GCC	AC	-165	
GOATBLG02	-	CATGGAGCCCCAGTGTCTGCTCAGCCGTGCCCCGCCCGCAGGGT-CAGGTCACITT				-189	
OALGB		T-			GT	-219	
GOATBLG02	-	CCCGTCTGGGGTATTATGACCGTGTTCATTTCATTGCCAFTTTGCTACCC				-244	
OALGB		-	TC	GC		-273	
GOATBLG02	-	TAACTGGGCAGCAGGTGCTTGCAGAGCCCCGATACCGACCAGGTCCCCCCTGG				-299	
OALGB				T	T	-328	
GOATBLG02	-	AGCTCCACCTGAACCCCGTGTCCACCCCTGCCCCAGCCTGCAGAGGATGGGGTAC				-354	
OALGB		G	A		G - G	-388	
GOATBLG02	-	TGCAGAGATCCCTTCACCCAAGGCCACGGTCACATGGTTTGGAGGAGCTGGTGCC				-409	
OALGB						-437	
GOATBLG02	-	CAAGGCAGAGGCCACCCCTAGGACACACCTGTCCCCAGTGTGGCTCTGACCTG				-464	
OALGB			C			-492	
GOATBLG02	-	C-CTGTCTAAGAGGCTGACCCCGGAGGTCTCTGGA-CGGAAGCCAGCCCTGA				-517	
OALGB		TC		CA	C	G	-547
GOATBLG02	-	-CCCAGAGTCCAGACACCCACCTGTGCCCCNACTTCTGGGGTCTACCAGGAACCG				-571	
OALGB		A		CG		-602	
GOATBLG02	-	TCTAGGCCCCAGAGGGGGACTTCTGCTTGGCCCCGGATGGAAGAAGCCCTCCTAT				-626	
OALGB				TT		-657	
GOATBLG02	-	TGTCTCTGTAGAGGAAGCCACCCCGGGGCCCGGGATGAGCCAAGTAGGATTCCG				-681	
OALGB				T A	G	-712	
GOATBLG02	-	GGAACCTCGTGGCAGGGGGCCCGCCCGGGCTGGCTGGCTGGCAC--GCCTCTGT				-734	
OALGB		G	T	A	CT	TGC	-767
GOATBLG02	-	TATAAGGCCCCGAGCCCGCT-GTCTCAGCCCTCCACTCCCTGCAGAGCTCAGAAG				-788	
OALGB			A	T	CT		-822
GOATBLG02	-	CACGACCCAGCTGCAGCCATGAAGTGCCTCCTGCTTGCCCTGGGCCCTGGCCCTC				-843	
OALGB						-877	
GOATBLG02	-	GCCTGTGGCTTCCAGGCCATCAGG				-867	
OALGB		G		TCGTACCCAGACCATGAAAGGCCTGGACATCC		-932	

Identity : 820 (94.6%)

(Note: The bases of OALGB which are the same as that of GOATBLG02 are not written, only those different are noted; "-" delegates for "vacancy")

转录因子结合位点的点突变会影响转录因子对它的识别, 从而影响基因的活性。那么, 山羊βLG基因770bp5'侧区序列能否像绵羊βLG5'侧区序列那样, 足以有效地指导外源基因高水平、组织特异性地表达呢? 于是, 我们又在HSⅢ和MPBF结合位点等区域作了碱基同源比较, 并且参考了牛βLG的转录因子结合位点^[19], 对山羊βLG序列进行了分析(表2)。

从绵羊、山羊βLG基因的DNA序列比较和表2中, 可以看出两个基因的CAAT

box、TATA box、转录起始点和信号肽的序列完全一致，没有碱基差异；而且，几种转录因子结合位点的差异也很小；-172 到 -133bp 的 milk box 仅仅存在 3 个碱基的差异；HS III (-300 到 +100bp) 区域，存在散在的共 21 个碱基的差异，同源率为 94.8%。另外，山羊与绵羊的 TATAbox、转录起始点和翻译起始处序列也非常一致；而几种转录因子结合位点的差异也较小。这说明了 β LG 基因的结构基因与转录调控区等保守性都很强。因此从以上分析中，我们可以初步推断，山羊 β LG 770bp 的 5' 侧区序列，可能如同绵羊 β LG 基因一样，亦足以指导外源基因在转基因动物中组织特异性地高水平表达。当然，这仍需转基因动物表达和 EMSA (Electrophoretic mobility shift assays) 等实验进一步证实。

表 2 山羊与绵羊、牛 β LG 基因 5' 侧区的潜在转录因子结合位点序列比较表

Table 2 Potential transcription factor binding site sequence comparison of 5' flanking region of goat, sheep and cattle β LG gene

Factor	Consensus-sequence	Sequence in goat	Position (goat) /bp	Difference of the base number	
		β LG gene		G&S	G&C
MPBF	GGTTCNGGAACC	GGTTGGGCCCTGTGGC	-675	1	3
	GGGATTGGCCAACCGC	GGGTCTACCAGGAACC	-213	No	2
		GGATTCCGGAACC	-95	No	No
MGF	ANTTCTTGNA	ACTTCTGCTTGGCCT	-180	1	2
GR	GGTACANNNTGTT/CCT	GCCTCCTATTGTCCT	-151	No	No
		GCTCTTGACCTGCCCT	-319	2	2
Ap2	CCCCAGGC	CCCCAGCC	-446	No	No

Note: 1. MPBF: milk protein binding factor [6]

2. MGF: mammary gland factor [20]

3. GR: glucocorticoid receptor [21]

4. AP2: activator protein2 [22]

5. G: Goat; S: Sheep; C: Cattle

参 考 文 献

- [1] Burdon T, Wall R J, Shamay A *et al.* Mech. Dev. 1991, 36: 67~74.
- [2] Simons T P, Mcclenaghan M, Clark A J *et al.* Nature. 1987, 328: 530~532.
- [3] Meade H, Gates L, Kacy E *et al.* Biotechnology. 1990, 8: 443~446.
- [4] Roberts B, Ditullio P, Vitale J *et al.* Gene. 1992, 121: 255~262.
- [5] Whitelaw CBA, Harris S, Mcclenaghan M *et al.* Biochem. J. 1992, 286: 31~39.
- [6] Watson C J, Gordon K E, Robertson M *et al.* Nucleic Acids Research. 1991, 19: 6603~6610.
- [7] Wright G, Carver A, Cottom D *et al.* Biotechnology. 1991, 9: 830~834.
- [8] Clark A J, Bessos H, Bishop J O *et al.* Biotechnology. 1989, 7: 487~492.
- [9] Sambrook J, Fritsch E F, Maniatis. Molecular cloning, A Laboratory Manual 2nd ed. Cold spring harbor laboratory press. New York 1989.
- [10] Harris S, Ali S, Anderson A S *et al.* Nucleic Acid Research. 1988, 16: 10379~10380.
- [11] Alexander L J, Hayes G, Bawden W *et al.* Animal Biotechnology. 1993, 4: 1~10.
- [12] Saiki P K, Gelfand D H, Stoffel S. Science. 1988, 239: 487~491.

- [13] Yoshimura M, Banerjee M R, Oka T. *Nucleic Acid Research*. 1986, **14** (88): 8 224.
- [14] Blackburn D E, Hobbs A A, Rosen J M. *Nucleic Acid Research*. 1982, **10**: 2 295~2 307.
- [15] Thepot D, Deviney E, Fontaine M L *et al.* *Gene*. 1991, **97**: 301~306.
- [16] Bonsing J, Ring J M, Stewart A F *et al.* *Aust J Biol Sci*. 1982, **41**: 527~537.
- [17] Hall L, Emery D C, Devis M S *et al.* *Biochem J*. 1987, **242**: 735~742.
- [18] Wagner V A, Schild T A, Geldermann H. *Theor Appl Genet*. 1994, **89**: 121~126.
- [19] Schmitt-Ney M, Doppler W, Ball R K *et al.* *Mol Cell Biol*. 1991, **11**: 3 745~3 755.
- [20] Beato M. *Cell*. 1989, **56**: 335~344.
- [21] Mitchell J P, Tijian R. *Cell*. 1987, **50**: 847~861.

Cloning, Sequencing and Sequence Analysis of Goat β -lactoglobulin Gene 5' Flanking Region

Pan Ling* Cheng Jianjun Chen Changqing

(*Shanghai Institute of Cell Biology, Academia Sinica, Shanghai 200031*)*

(*Shanghai Research Center of Biotechnology, Academia Sinica, Shanghai 200233*)

Abstract We first reported the 867bp 5' flanking sequences of goat β -Lactoglobulin (β LG) Gene including the 770bp promoter and the 97bp part of the first exon. The fragment was isolated from Chinese goat tissue by PCR method, and cloned, sequenced. Comparing it with that of sheep and cattle in computer, we found their homology is 94.6% and 88% respectively, with high conservation of the transcription factor binding sites. It had been demonstrated, promoters of sheep β LG and cattle β LG could directed the strong tissue-specific and position-independent expression of exogenous gene in transgenic mice. Therefore, this goat β LG 5' flanking sequences, maybe, could direct heterologous protein expression in the transgenic animals. This work made the basis of using goat β LG gene promoter to direct heterologous protein expression in transgenic animal mammary gland.

Key words Goat β -Lactoglobulin, promoter, DNA sequence analysis, PCR amplification