

山羊 *MyoG* 基因启动子区序列分析与结构预测

刘铮铸¹, 巩元芳^{1*}, 张文香¹, 付志新¹, 张传生¹, 宋 瑜², 马艳华³

(1. 河北科技师范学院动物科技学院, 昌黎 066600; 2. 中国环境管理干部学院, 秦皇岛 066004;

3. 河北科技师范学院欧美学院, 秦皇岛 066004)

摘要: 为探明山羊 *MyoG* 基因的启动子序列与结构及转录调控机制, 本研究设计了 1 对引物, 采用 PCR 扩增、测序以及序列比对分析, 从波尔山羊基因组中扩增出一段长度为 969 bp 的山羊 *MyoG* 基因的 5'侧翼序列, 其 GC 含量为 50.6%。经过生物信息学分析, 确定了其转录起始位点及活性区域; 发现在转录起始位点上游-24 bp 处存在 1 个 TATA-box, 另外还发现了 1 个 GC-box、2 个 CAAT-box、2 个 E-box 和 1 个富含 A-T 碱基的模体结构; 该序列还包含 HSF、ADR1 和 cap 等转录因子结合位点。通过比对分析, 所得山羊 *MyoG* 基因 5'侧翼序列与牛、马鹿、人、小鼠和鸡同源序列的相似性在 37.22%~96.85% 之间, 说明不同物种间该序列具有一定的保守性。本研究为进一步探讨山羊 *MyoG* 基因的表达和调控机制奠定了理论基础。

关键词: 山羊; *MyoG* 基因; 启动子区; 序列分析; 结构预测

中图分类号: S827; Q523.8

文献标识码: A

文章编号: 0366-6964(2012)01-0029-08

Sequence Analysis and Structure Prediction of the Promoter Region of *MyoG* Gene in Goats

LIU Zheng-zhu¹, GONG Yuan-fang^{1*}, ZHANG Wen-xiang¹, FU Zhi-xin¹, ZHANG Chuan-sheng¹,
SONG Yu², MA Yan-hua³

(1. *College of Animal Science and Technology, Hebei Normal University of Science and Technology, Changli 066600, China*; 2. *Environmental Management College of China, Qinhuangdao 066004, China*; 3. *Europe & America College of Hebei Normal University of Science and Technology, Qinhuangdao 066004, China*)

Abstract: In order to explore the mechanism of goat *MyoG* gene transcription regulation by studying the promoter region of the gene, the 5' flanking sequence of *MyoG* from the Bore goat genome was attained by PCR amplification, sequencing and sequence alignment. The DNA sequence of 969 bp was amplified. The content of G and C in the sequence was about 50.6%. The sequence analysis showed that there were one TATA-box at -24 bp and one GC-box, two CAAT-box, two E-box and one A+T-rich motif. Meanwhile, some transcription factor binding sites including HSF, ADR1, cap were detected. The 5' flanking sequence of goat *MyoG* was compared with cattle, wapiti, human, mouse and chicken, the similarity was between 37.22% and 96.85%. It indicates that the 5' flanking sequence of *MyoG* is comparatively conservative among different species. This work would lay the theoretical foundation for further study for expressing and regulating mechanism of *MyoG* gene in goat.

Key words: goat; *MyoG* gene; promoter region; sequence analysis; structure prediction

收稿日期: 2011-02-14

基金项目: 河北省自然科学基金(C2007000743); 国家科技支撑计划(2008BADB2B04-8-4); 河北科技师范学院博士基金资助(2010YB006)

作者简介: 刘铮铸(1968-), 男, 内蒙古包头人, 教授, 博士, 主要从事动物分子遗传学研究, Tel: 0335-2039062, E-mail: liuzhengzhu@163.com

* 通讯作者: 巩元芳, 教授, 博士, E-mail: gongyuanfang1992@yahoo.com.cn

MyoG 基因是生肌调节因子 MRFs 家族的主要成员之一,该家族基因具有一个保守的碱性螺旋-环-螺旋结构 (bHLH),其包括肌细胞生成素 (*MyoG*)、生肌决定因子 D (*MyoD*)、生肌因子 5 (*Myf5*) 和生肌调节因子 4 (*MRF4*) 4 种转录调节因子,其中 *MyoG* 基因调控中胚层细胞分化为成肌细胞,再由成肌细胞融合为肌纤维这一过程,在肌细胞分化过程中起着中心调节作用^[1-4]。启动子是位于结构基因 5' 端上游的一段核苷酸序列,是 RNA 聚合酶识别并与模板 DNA 特异性结合的部位,保证转录起始的精准性^[5]。真核生物许多基因的启动子序列中常含有 TATA-box、GC-box 和 CAAT-box 等调控元件,这些调控元件的存在与分布对基因的正常转录有着不可分割的调控作用。因此,深入地研究启动子,对于了解包括 *MyoG* 基因在内的各种基因的结构及其表达调控等具有重要意义。

目前,在部分动物上人们对 *MyoG* 基因 5' 侧翼启动子区的序列结构、表达调控与动物的肌细胞生成机理等方面的研究已经有相关的报道, Malik 等^[6]对鸡 *MyoG* 基因 5' 端 833 bp 长的侧翼序列进行了分析,认为其中 131 bp 的 DNA 片段是包括 E-box 和肌细胞增强因子 2 (MEF-2) 的启动子,其启动子的结构、转录起始位点及转录调控特点与小鼠和人的非常相似。王珊^[7]和宋兴超^[8]等分别对牛和马鹿 *MyoG* 基因启动子区序列进行了克隆并进行了相应的生物信息学分析。但迄今为止,无论是绵羊还是山羊 *MyoG* 基因启动子区的研究还未见报道。为此,本研究通过 PCR 扩增及测序等技术,获得了山羊 *MyoG* 基因的启动子区序列,并对该序

列进行了生物信息学方面的分析,以期为山羊 *MyoG* 基因的结构、表达调控及其突变位点与山羊产肉性状等方面的研究提供理论依据。

1 材料与方法

1.1 材料

1.1.1 试验动物 试验用波尔山羊的肾脏采自秦皇岛市波尔山羊繁育改良研究中心。

1.1.2 试剂 LA *Taq* DNA Polymerase、dNTPs、TaKaRa Agarose Gel DNA Purification Kit Ver. 2.0、蛋白酶 K、DNA Marker DL 2000 均购自宝生物工程(大连)有限公司,其它均为国产分析纯。

1.2 方法

1.2.1 基因组 DNA 的提取 波尔山羊基因组 DNA 的提取见参考文献^[9],TE 溶解, -20 °C 冻存。

1.2.2 引物设计和 PCR 扩增 用在线软件 Primer 3 (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi), 根据 GenBank 数据库中登载的牛 *MyoG* 基因启动子区的同源序列 (HM452338) 设计上游引物: 5'-GTCTCTGCTTTCTCTCCTTC-3', 然后再根据本课题组提交到 GenBank 数据库中的山羊 *MyoG* 基因近 5' 端的编码区序列 (FJ607135) 设计下游引物: 5'-TCTTGCACACCTTACACGCC-3', 以波尔山羊基因组为模板获得山羊 *MyoG* 基因启动子区未知序列, 预期扩增产物约为 1 244 bp。引物委托宝生物工程(大连)有限公司合成。PCR 扩增的反应体系和反应条件见表 1。

表 1 山羊 *MyoG* 基因启动子区 PCR 扩增的反应体系和反应条件

Table 1 PCR amplification system and condition of promoter region of goat *MyoG* gene

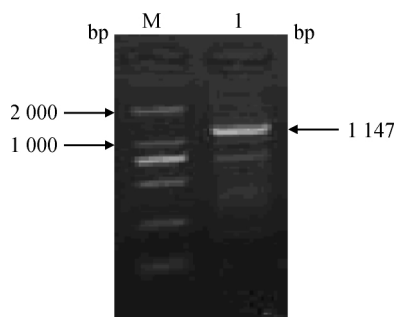
组成 Composition	用量/ μ L Volume	反应条件 Reaction condition		
山羊基因组 DNA	1	94 °C	5 min	1 cycle
10 $\times$ LA PCR Buffer II (Mg ²⁺ Plus)	5			
dNTP Mixture (2.5 mmol $\cdot$ L ⁻¹ each)	8	94 °C	30 s	
SP4 Primer (20 μ mol $\cdot$ L ⁻¹)	1	55 °C	15 s	35 cycle
SP2 Primer (20 μ mol $\cdot$ L ⁻¹)	1	72 °C	75 s	
TaKaRa LA <i>Taq</i> [®] (5 U $\cdot$ μ L ⁻¹)	0.5	72 °C	10 min	1 cycle
dH ₂ O	33.5	4 °C	∞	

1.2.3 *MyoG* 基因启动子区的结构预测及序列分析 利用分子生物软件 Chromas2.3 对已获得的 *MyoG* 基因启动子区序列的峰图进行校对,用 BioEdit(7.0.1.1)进行序列拼接,采用 DNAMAN、Neural Network Promoter Prediction (http://www.fruitfly.org/seq_tools/promoter.html)、TA-TA signal prediction (http://zeus2.itb.cnr.it/~webgene/wwwHC_tata.html)、转录结合位点 (<http://mbs.cbrc.jp/research/db/TFSEARCH.html>)以及 CpG island Searcher (<http://www.usc-norris.com/cpgislands2/cpg.aspx>)和 CpG Island Searcher (<http://zeus2.itb.cnr.it/cgi-bin/wwwcpwg.pl>)等软件对该序列潜在的顺式作用元件和反式作用元件及其 CpG 岛等进行预测分析,然后再利用 DNAMAN 将该序列与 GenBank 中牛 (HM452338)、马鹿 (FJ746497)、人 (AF050501)、小鼠 (M95800)和鸡 (X85744)的同源序列进行相似性分析。

2 结果

2.1 *MyoG* 基因启动子区未知序列的分析

利用所设计的引物,以 1 μ L 波尔山羊基因组 DNA 为模板,获得了较好的 PCR 扩增结果(图 1),并对扩增产物进行了测序,测序结果见图 2。



M. DNA 相对分子质量标准 DL 2000; 1. PCR 产物

M. DNA marker DL 2000; 1. PCR product

图 1 山羊 *MyoG* 基因启动子区扩增产物电泳结果

Fig. 1 The PCR amplification result of goat *MyoG* promoter

从图 1 可知,扩增产物约 1 200 bp,片段长度与预期大小基本一致。测序所得峰图经 Chromas2.3 校对,用 BioEdit(7.0.1.1)软件对测序结果进行了拼接,发现以上游引物正向测序获得序列长 886 bp,以下游引物反向测序获得序列长 798 bp,二者重叠 537 bp,经过认真比对确认此扩增产物为 1 147 bp。

再与本课题提交到 GenBank 数据库中的山羊 *MyoG* 基因序列 (FJ607135) 比对发现,该序列包括编码区序列 178 bp,由此可知,本研究得到山羊 *MyoG* 基因 5'侧翼序列 969 bp(图 2)。

2.2 山羊 *MyoG* 基因启动子区结构预测及生物信息学分析

通过 CpG Island Searcher 和 CpG Island Prediction 工具预测,发现山羊 *MyoG* 基因 5'侧翼调控区没有符合条件的 CpG 岛。通过 DNAMAN、TA-TA signal prediction 在线软件分析以及与已知鸡 *MyoG* 基因启动子区序列^[6]的进一步比对分析可知,山羊 *MyoG* 基因启动子区具有典型的真核生物启动子结构元件和序列特征,含有 TATA-box、GC-box、CAAT-box、E-box 和富含 A+T 碱基的模体结构(motif)等。图 3 中的黑体 G 是经 Neural Network Promoter Prediction 在线工具推测为山羊 *MyoG* 基因的转录起始位点。启动子的预测结果表明,山羊 *MyoG* 基因启动子为“GCGGG-GAGGGTTTAAATGGCACCCAGCAGTTGGCG-TGAGGGGCCGAGGA”,转录起始位点“G”位于起始密码子 ATG 上游 50 bp 处(图 3)。

山羊与牛、马鹿、小鼠、人、鸡 *MyoG* 基因启动子序列比较分析结果见图 4。

通过启动子在线分析软件 TFSEARCH 预测,发现山羊 *MyoG* 基因启动子区分数>90.0 的转录因子结合位点有 63 个,其中 HSF 11 个,ADR1 17 个,cap 8 个,STRE 4 个,CdxA 和 MZF1 均为 3 个,SRY、NIT2、Nkx-2 和 GATA-1 均为 2 个,GCN4、AP-1、AP-4、C/EBP、MATa1p、MyoD、GCR1、Dfd 和 c-Myb 均为 1 个。表 2 为主要转录因子及其序列和所在位置。

2.3 山羊 *MyoG* 基因启动子区与其它物种同源序列的相似性分析

用 DNAMAN 软件将所得的山羊 *MyoG* 基因启动子区与 GenBank 中牛 (HM452338)、马鹿 (FJ746497)、人 (AF050501)、小鼠 (M95800)和鸡 (X85744)等同源序列进行相似性分析,结果见表 3。山羊与各物种的相似性在 37.22%和 96.85%之间,表明该序列在不同物种间具有一定的保守性。对其序列进行进一步分析,发现山羊 *MyoG* 基因启动子区序列中各碱基组成存在差别,GC 含量为 50.6%,AT 含量为 49.4%,GC 含量略高于 AT 含量。

Promoter	CCTGGTGTCTAAATTGGGTAGGAAACAAGTTGAAAAGGTGACATGGGAAGACAGATAAAAT	60 bp
Upstream	CCTGGTGTCTAAATTGGGTAGGAAACAAGTTGAAAAGGTGACATGGGAAGACAGATAAAAT	
Downstream	-----	
FJ607135	-----	
Promoter	GATTACAGAGTCAGAAAAGAGAGGAGACAGAGGGCTTAATAGACCTTGACCTTGCCACTGTG	120 bp
Upstream	GATTACAGAGTCAGAAAAGAGAGGAGACAGAGGGCTTAATAGACCTTGACCTTGCCACTGTG	
Downstream	-----	
FJ607135	-----	
Promoter	GGGAAGATGGTGTATGTAGACTGTAAATTTCAATCTCTGCTTTAACCCCTGGTCAGCTGT	180 bp
Upstream	GGGAAGATGGTGTATGTAGACTGTAAATTTCAATCTCTGCTTTAACCCCTGGTCAGCTGT	
Downstream	-----	
FJ607135	-----	
Promoter	ATTAAGAGAAAACTTCCACTGACTCTTACCAAAAAAGAGAAAAAAGACCGAAAAATTC	240 bp
Upstream	ATTAAGAGAAAACTTCCACTGACTCTTACCAAAAAAGAGAAAAAAGACCGAAAAATTC	
Downstream	-----	
FJ607135	-----	
Promoter	CTTTTGCCCTCAGTTTATCCACAATCACATGAGATCTAAGAAATATATGATGTCAAAACTG	300 bp
Upstream	CTTTTGCCCTCAGTTTATCCACAATCACATGAGATCTAAGAAATATATGATGTCAAAACTG	
Downstream	-----	
FJ607135	-----	
Promoter	CTTTGAATACTTAAAGTAGAATCCATCTGCCAATCCTTTTACCCTGGACACTGAGTAT	360 bp
Upstream	CTTTGAATACTTAAAGTAGAATCCATCTGCCAATCCTTTTACCCTGGACACTGAGTAT	
Downstream	-----	
FJ607135	-----	
Promoter	CCATACATGCTGCGACACTGAAGTGCTGTCACCCCTCTGGAAGGGAATCTTTTGAAGAGT	420 bp
Upstream	CCATACATGCTGCGACACTGAAGTGCTGTCACCCCTCTGGAAGGGAATCTTTTGAAGAGT	
Downstream	-----	
FJ607135	-----	
Promoter	TTCACTCTACCAATACAGTCTGGGTAAAGTGCTATGAGGAAGCAGGGGGATGAGTAGACT	480 bp
Upstream	TTCACTCTACCAATACAGTCTGGGTAAAGTGCTATGAGGAAGCAGGGGGATGAGTAGACT	
Downstream	-----	
FJ607135	-----	
Promoter	GACTTCCTATGGCCCTTCCAGCCTAAGCCTGCTCCTTCCTTGCTCTCCCCACCCCCAC	540 bp
Upstream	GACTTCCTATGGCCCTTCCAGCCTAAGCCTGCTCCTTCCTTGCTCTCCCCACCCCCAC	
Downstream	-----	
FJ607135	-----	
Promoter	CCGTCTTCTTTGAGGCTAGCAGCAGGCAGGCCGCCAGCTAGGAGTAATTGAAAGGAGCA	600 bp
Upstream	CCGTCTTCTTTGAGGCTAGCAGCAGGCAGGCCGCCAGCTAGGAGTAATTGAAAGGAGCA	
Downstream	-----	
FJ607135	-----	
Promoter	GATGAGGGGAGAATGTGTGCTCCTCCCCAGCTCCCCGGCCCCACGGGGGCTGTCGAGAA	660 bp
Upstream	GATGAGGGGAGAATGTGTGCTCCTCCCCAGCTCCCCGGCCCCACGGGGGCTGTCGAGAA	
Downstream	-----	
FJ607135	-----	
Promoter	ATGAAAACTAGTCAGATTACACCCTGTGGCTCCTCCAGTGACAGGAGCCAGCTAGG	720 bp
Upstream	ATGAAAACTAGTCAGATTACACCCTGTGGCTCCTCCAGTGACAGGAGCCAGCTAGG	
Downstream	-----	
FJ607135	-----	
Promoter	GGTAGGCTGGCTGGGCGAGGGGCGGGGTGGTGTGTGACAGGGGAGAGGGAAGGGGAATC	780 bp
Upstream	GGTAGGCTGGCTGGGCGAGGGGCGGGGTGGTGTGTGACAGGGGAGAGGGAAGGGGAATC	
Downstream	-----	
FJ607135	-----	
Promoter	ACGTCTAATCCACTGTAAACGTCTTGATGTGTGACAGCAACAGCTTAGAGTGGGGCTCAAGTT	840 bp
Upstream	ACGTCTAATCCACTGTAAACGTCTTGATGTGTGACAGCAACAGCTTAGAGTGGGGCTCAAGTT	
Downstream	-----	
FJ607135	-----	
Promoter	TCTGTGGCGTTGGCTATATTTATCTCTGGTTCCATGCCAGCGGGGAGGGTTTAAATGGCA	900 bp
Upstream	TCTGTGGCGTTGGCTATATTTATCTCTGGTTCCATGCCAGCGGGGAGGGTTTAAATGGCA	
Downstream	-----	
FJ607135	-----	
Promoter	CCCAGCAGTTGGCGTGAGGGGCGCAGGAGCTTGGGGCTGGTGGCAGGAACAAGTCTCT	960 bp
Upstream	CCCAGCAGTTGGCGTGAGGGGCGCAGGAGCTTGGGGCTGGTGGCAGGAACAAGTCTCT	
Downstream	-----	
FJ607135	-----	
Promoter	GCTGACCCCATTGAGCTGTATGAGACCTCTCCCTACTTCTATCAGGAACCTCACTTCTAT	1 020 bp
Upstream	GCTGACCCCATTGAGCTGTATGAGACCTCTCCCTACTTCTATCAGGAACCTCACTTCTAT	
Downstream	-----	
FJ607135	-----	
Promoter	GACGGGGAGAACTACCTGCCTGTCCACCTCCAGGGCTTCGAGCCGCCAGGCTACGAGCGG	1 080 bp
Upstream	GACGGGGAGAACTACCTGCCTGTCCACCTCCAGGGCTTCGAGCCGCCAGGCTACGAGCGG	
Downstream	-----	
FJ607135	-----	
Promoter	GCTGAGCTCGGCCCTGAGCCCCGAGGCTCGCGTGCCCCCTTGAAGACAAGGGGCTGGGGCCC	1 140 bp
Upstream	GCTGAGCTCGGCCCTGAGCCCCGAGGCTCGCGTGCCCCCTTGAAGACAAGGGGCTGGGGCCC	
Downstream	-----	
FJ607135	-----	
Promoter	GCGGAGC	
Upstream	-----	
Downstream	GCGGAGC	
FJ607135	GCGGAGC	

阴影部分为重叠序列,ATG为起始密码子

The shadows are the overlapping sequence, ATG is the initiation codon

图2 正反向测序结果与已获得的 *MyoG* 基因启动子区序列、GenBank 数据库登载序列(FJ607135)的比对

Fig. 2 Alignment of sequences among upstream-, downstream-*MyoG*, promoter DNA obtained and *MyoG* DNA (FJ607135) in goats

```

-919 CCTGGTGTCTAAATTGGGTAGGAAACAAGTTGAAAAGGTGACATGGGAAGACAGATAAAAT
-859 GATTCAGAGTCAGAAAAGAGAGGAGACAGAGGGCTTAATAGACCTTGACCTTGCCACTGTG
-799 GGGAAAGATGGTGTGTGTAGACTGTAAATTTCAATCTCTGCTTTAACCTGGTCAGCTGT
-739 ATTAAGAGAGAAAACCTTCCACTGACTCTTACCAAAAAAGAGAAAACAGACCAGAAAAATTC
-679 CTTTTGCCCTCAGTTTATCCACAATCACATGAGATCTAAGAAATATATGATGTCAAAACTG
-619 CTTTGAATACTTAAAAGTAGAATCCATCTGC2CCAATCCTTTTACCCTGGACACTGAGTAT
-559 CCATACATGTCTGCAGACTGAAGTGCTGTCCACCTCTGGAAAGGGAATCTTTGAAGAGT
-499 TTCATCTCAC1CCAATACAGTCTGGGTAAAGGTGTATGAGGAAGCAGGGGGATGAGTAGACT
-439 GACTTCCTATGGCCCTTCCAGCCTAAGCCTGCTCCTTCTTGTCTTCCCCACCCAC
-379 CCGTCTTCTTTGAGGCTAGCAGCAGGCAGGBCCGCCAGCTAGGAGTAATTGAAAGGAGA1CA
-319 D2GATGAGGGGAGAATGTGTCTCTCCCCAGCTCCCCGGCCCCACGGGGGCTGTGAGAA
-259 ATGAAAAC TAGTCAGATTACACCTGTGGCTCCCTCCAGTGCACAGGAGCCAGCTAGG
-199 GGTAGGCTGGCTGGGGCAGGGGCCGGGTGGTGTGTGCAGGGGAGAGGGAAGGGGAATC
-139 ACGTCTAATCCACTGTAAACGTCTTGATGTGCAGCAACAGCTTAGAGTGGGGCTCAAGTT
-79 TCTGTGGCGTTGGCA2TATATTTATCTCTGGTTCATGCCAGCGGGGAGGGA1TTAAATGGCA
-19 CCCAGD1CAGTTGGCGTGAGGGGCCGAGCAA1GCTTGGGGGCTGGTGGCAGGAACAAGTCTCT
+42 GCTGACCCCATGG

```

A1 为 TATA-box; A2 为富含 A+T 碱基的模体结构; B 为 GC-box; C1 和 C2 均为 CAAT-box; D1 和 D2 均为 E-box; G 为转录起始位点; A 为山羊 *MyoG* 基因启动子区的 3' 端; ATG 为起始密码子

A1. TATA-box; A2. A+T-rich motifs; B. CAAT-box; C1, C2. GC-box; D1, D2. E-box; G. Transcription start site; A. The 3' end of the caprine *MyoG* promoter region; ATG. Initiation codon

图 3 山羊 *MyoG* 基因启动子区序列结构

Fig. 3 Nucleotide sequence and structure of the promoter region of the goat *MyoG* gene

3 讨 论

真核生物的启动子常含有 TATA-box、GC-box 和 CAAT-box 等调控元件, 它们的存在与转录调控有着内在的联系。本研究所得山羊 *MyoG* 基因 5' 侧翼调控区序列长 969 bp, 在此序列中发现 2 个富含 A+T 碱基的模体 (motif) 结构, 1 个 GC-box、2 个 CAAT-box 和 2 个 E-box, 同时进行了山羊与牛、马鹿、人、小鼠和鸡 *MyoG* 基因启动子区序列的比较分析, 得知山羊与牛、马鹿、人、小鼠和鸡 *MyoG* 基因启动子调控元件的结构和特点基本相似, 也与前人对以上物种的研究结果相一致^[6-8]。其中第一个 motif (5'-TTTAAAT-3') 位于 -30~-24 bp 处, 是该基因的 TATA-box, 此结果符合真核生物启动子 TATA 框一般位于 -34~-26 bp 处 (绝大多数在 (-31±2) bp 处) 的理论观点^[10-12], 也与牛、猪、马鹿、小鼠、人和鸡的启动子中 TATA 框的所处位置 -30~-24 bp 一致^[6-8]; 第二个 motif (5'-TATATTTAT-3') 位于 -65~-57 bp 处, 与牛、马鹿、小鼠、人和鸡的该 motif 所处位置 -65~-57 bp

一致, 在鸡上该 motif 是肌细胞增强因子 2 (MEF-2) 的结合位点^[6]。2 个 E-box 分别位于 -14~-9 bp 处和 -321~-316 bp, 与鸡的 2 个 E-box 所处的位置基本一致, 而在鸡上 E-box 和 MEF-2 二者共同协调激活 *MyoG* 基因启动子并增强 myogenin/E2-5 的活性^[6], 推测对于山羊 E-box 和 MEF-2 共同协调也能激活 *MyoG* 基因启动子的调控作用。1 个 GC-box 位于 -349~-344 bp 处, 与典型的 GC-box 相比, 具有较高的相似性, 推测其可以激活山羊 *MyoG* 基因的转录。2 个 CAAT-box 分别位于 -490~-486 bp 处和 -589~-585 bp 处, 距离转录起始位点较远, 其是否能增强山羊 *MyoG* 基因的转录效率, 有待进一步的研究。

本研究利用在线分析软件预测了山羊 *MyoG* 基因启动子区的各种转录因子结合位点的分布情况, 找到了转录起始位点和 63 个转录因子结合位点, 其中 ADR1、HSF、cap 等因子出现的频率较高, 也包括像 AP-1 等出现频率较低的转录因子, 这些因子中多数是哺乳动物基因转录活化的必须因子^[13]。

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

表 2 山羊 *MyoG* 基因启动子区中主要转录因子结合部位Table 2 Putative binding sites of main transcription factors in the promoter region of goat *MyoG* gene

转录因子	转录因子结合位点序列	具体位置
Transcription factor	Sequences of binding sites of transcription factor	Location
ADRI	TGGGGA	-801→-796
	AGGGGG	-456→-451, -161→-156
	GCCCCCT	-428←-423
	ACCCCA	-386←-381
	CGGGGT	-176→-171
	TCCCCG	-288←-283
	CCCCCA	-391←-386, -296←-291, -281←-276
	AGGGGT	-202→-197
	TGGGGC	-188→-183, -92→-87
	AGGGGC	-182→-177, -3→+3
	CGGGGA	-39→-34
	CGGGGG	-275→-270
		-848→-844, -732→-728, -701→-697,
HSF	AGAAA	-642→-638, -263→-259
	AGAAT	-601→-597, -310→-306
	CGAAA	-689→-685
	CTTCT	-375←-371
	TTTCT	-81←-77
	GTTCC	-51←-47
cap	TCAATCTC	-769→-762
	TCAGTTTA	-671→-664
	GACACTGA	-571←-564
	CAGACTGA	-546←-539
	TCACCCTC	-531→-524
	GGGGATGA	-454←-447
	TAGACTGA	-445←-438
STRE	AGAAATGA	-263←-256
	GCAGGGGG	-458→-451, -163→-156
	TGAGGGGA	-317→-310
CdxA	TGAGGGGC	-5→+3
	GATAAAT	-866←-860
	TATTAAA	-740←-734
MZF1	ATTTATC	-62→-56
	TGTGGGGA	-803→-796
	TGAGGGGA	-317→-310
SRY	AGCGGGGA	-41→-34
	AAACAAG	-897→-891, -698→-692
NIT2	CAGATA	-868←-863
	TATCTC	-59→-54
Nkx-2	CTTAATAG	-827→-820
	TGAAGTG	-541→-535
GATA-1	GAAGATGGTG	-797→-788
	GGTGATGTGT	-791→-782
GCN4	AAATGATTCA	-863→-854
AP-1	ATGATTCAG	-861→-853
AP-4	GTCAGCTGTA	-748→-739
C/EBP	CTCTTACCAAAAA	-716←-704
MATap	AATCATATGA	-658←-649
MyoD	TCCATCTGCC	-597→-588
GCR1	GAGGAAGCA	-464←-456
Dfd	TAGGAGTAATTGAAAG	-340←-325
c-Myb	AGCAGTTGGC	-17←-7

表 3 山羊同其它物种 *MyoG* 基因启动子区同源序列的相似性分析Table 3 Similarity analysis of goat *MyoG* to other species in promoter region

中文种名	英文种名	拉丁种名	相似性/%	GenBank 号
Species name	Species name	Species name	Nucleotide similarity	GenBank No.
in Chinese	in English	in Latin		
牛	Cattle	<i>Bos taurus</i>	95.05	HM452338
马鹿	Wapiti	<i>Cervus elaphus kansuensis</i>	96.85	FJ746497
人	Human	<i>Homo sapiens</i>	69.08	AF050501
小鼠	House mouse	<i>Mus musculus</i>	56.11	M95800
鸡	Chicken	<i>Gallus gallus</i>	37.22	X85744

山羊 *MyoG* 基因启动子区与牛、马鹿、人、小鼠和鸡同源序列的相似性分析表明,山羊与马鹿的相似性最高,这可能是由于马鹿目前所知序列较短所致,其次相似性较高的是牛,这是由于山羊和牛均是草食反刍动物,与它们本身的生理特性相吻合,这也与前期对该基因编码区的分析结果一致^[14]。与人、小鼠和鸡也有不同程度(69.08%、56.11%和 37.22%)的相似性,说明 *MyoG* 基因 5'侧翼调控区在不同物种间的保守性较高。

4 结 论

根据试验结果初步判定,本研究所得长度 969 bp 的山羊 *MyoG* 基因的启动子序列是 1 个具有 TATA-box、GC-box 和 CAAT-box 等典型模体结构的真核启动子。但本序列是否具有启动子活性,以及预测的转录因子结合位点是否正确,均需进一步验证。对于本启动子序列如何调控山羊 *MyoG* 基因的转录,仍需更深入的研究。本研究为构建荧光素酶报告基因重组体,分析该序列的启动子活性,确定山羊 *MyoG* 基因的核心启动子调控区,探明 *MyoG* 基因的表达及其转录调控机理等的研究奠定了基础。

参考文献:

- [1] WRIGHT W E, SASSOON D A, LIN V K. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD [J]. *Cell*, 1989, 56(4): 607-617.
- [2] RAWLS A, VALDEZ M R, ZHANG W, et al. Overlapping functions of myogenic bHLH genes MRF4 and MyoD revealed in double mutant mice [J]. *Development*, 1998, 125(13): 2349-2358.
- [3] HASTY P, BRADLEY A, MORRIS J H, et al. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene [J]. *Nature*, 1993, 364(6437): 501-506.
- [4] NABESHIMA Y K, HANAOKA K, HAYASAKA M, et al. Myogenin gene disruption results in perinatal lethality because of severe muscle defect [J]. *Nature*, 1993, 364(6437): 532-535.
- [5] LEMON B, TJIAN R. Orchestrated response: a symphony of transcription factors for gene control [J]. *Genes Dev*, 2000, 14(20): 2551-2569.
- [6] MALIK S, HUANG C F, SCHMIDT J. The role of the CANNTG promoter element (E-box) and the myocyte-enhancer-binding-factor-2 (MEF-2) site in the transcriptional regulation of the chick *myogenin* gene [J]. *Eur J Biochem*, 1995, 230(1): 88-96.
- [7] 王 珊,陈 宏,蔡 欣. 牛 *MyoG* 基因启动子的克隆与序列分析 [J]. 西北农林科技大学学报(自然科学版), 2006, 34(8): 12-16.
- [8] 宋兴超,荣 敏,邢秀梅,等. 甘肃马鹿肌细胞生成素 (*MyoG*) 基因启动子区序列克隆与分析 [J]. 特产研究, 2009, (2): 12-16.
- [9] SAMBROOK J, FRITSCH E F, MANIATIS T. Molecular Cloning (Second edition) [M]. New York: Cold Spring Harbor Laboratory Press, 1989.
- [10] LAGRANGE T, GAUVIN S, YEO H J, et al. S2F, a leaf-specific trans-acting factor, binds to a novel cis-Acting element and differentially activates the RPL21 gene [J]. *Plant Cell*, 1997, 9(8): 1469-1479.
- [11] 张小辉,祁艳霞. 真核生物启动子 TATA-box、GC-box 和 CAAT-box 的分析 [J]. 安徽农业科学, 2008, 36(4): 1380-1381.
- [12] 黄 玉,杨 波,迟小华,等. 真核生物启动子的研究及应用 [J]. 生物技术通讯, 2010, 21(2): 275-279.
- [13] 陈书明,晋大鹏,王立贤,等. 猪 SP-A 基因启动子区的克隆与序列分析 [J]. 畜牧兽医学报, 2009, 40(11): 1600-1608.
- [14] 刘铮铸,巩元芳,张传生,等. 波尔山羊 *MyoG* 基因的鉴定和序列分析 [J]. 畜牧兽医学报, 2010, 41(10): 1337-1341.

(编辑 郭云雁)