

牛 *ANGPTL4* 基因结构和功能的生物信息学分析

马云¹ 骆莹莹² 李芬¹ 王军芳¹ 徐源¹ 姚瑾¹ 李荣荣³ 陈宏³

(1. 信阳师范学院 生命科学学院 河南 信阳 464000; 2. 驻马店市平舆县第一高级中学 河南 驻马店 463300; 3. 西北农林科技大学 动物科技学院 陕西 杨凌 712100)

摘要:以牛 *ANGPTL4* 基因为研究对象 运用生物信息学方法对其编码蛋白的结构、理化性质、信号肽、跨膜结构、亚细胞定位、二级结构以及高级结构进行生物信息学分析 并推测与其他物种的生物进化关系。结果表明,牛 *ANGPTL4* 蛋白是一个疏水性不稳定蛋白,定位于细胞外。含有 12 个 α -螺旋 8 个 β -折叠 30 个转角 3 个跨膜结构区域。通过分子进化树研究发现牛的 *ANGPTL4* 基因在人、小鼠、大鼠、猪、黑猩猩、狗、鸡等物种中,与猪的亲缘关系最近。

关键词:牛;血管生成素样蛋白-4(*ANGPTL4*);序列分析;生物信息学

中图分类号:S823.2 文献标识码:A 文章编号:1000-2286(2010)01-0135-06

Bioinformatics Analysis of Bovine *ANGPTL4* Gene Structure and Function

MA Yun¹ ,LUO Ying-ying² ,LI Fen¹ ,WANG Jun-fang¹ ,
XU Yuan¹ ,YAO Jin¹ ,LI Rong-rong³ ,CHEN Hong³

(1. College of Life Science ,Xinyang Normal University ,Xinyang 464000 ,China;2. The First Senior Middle School of Pingyu County ,Zhumadian 463300 ,China;3. College of Animal Science and Technology ,North-west Agricultural & Forestry University ,Yangling 712100 ,China)

Abstract:*ANGPTL4* gene in cattle was investigated. By means of bioinformatics analysis its encoding protein was analyzed including its structure , physico - chemical property , signal peptide , transmembrane structure , subcellular localization , secondary structure and higher structure , and its evolution relationship with other species was inferred. The results indicated that *ANGPTL4* protein of cattle was a kind of hydrophobicity labile protein , locating cellular exterior , which contained 12 α - helices , 8 β - strands , 30 turns and 3 transmembrane domains. It might be concluded that *ANGPTL4* gene of cattle has close affinity with that of pig , compared with that of human , mus , rat , pan , dog and fowl based on the analysis of molecular evolution.

Key words: cattle ; angiotensin - like 4 (*ANGPTL4*) ; sequence analysis ; bioinformatics

类血管生成素 4 (*ANGPTL4*) 最早发现于 1999 年,又称为禁食诱导脂肪因子(fasting induced adipose factor ,FIAF)^[1]。Yoon 等研究发现 *ANGPTL4* 基因为过氧化物增殖体激活受体 α (PPAR α) 和 γ (PPAR γ) 的下游靶基因,将其编码产物称为过氧化物增殖体激活受体血管生成素相关蛋白(peroxisome proliferator - receptor (PPAR) - gamma angiotensin - related ,PGAR)^[2]。2002 年,国际人类基因组组织(HUGO)

收稿日期:2009-09-18 修回日期:2009-11-24

基金项目:河南省自然科学基金项目(092300410010)、信阳师范学院青年骨干教师资助计划(2006BAD04A16)和河南省高等学校青年骨干教师资助计划及现代农业(肉牛)产业技术体系专项共同资助

作者简介:马云(1974-)男,副教授,博士,主要从事动物遗传育种及分子生物学研究,E-mail:tmlf74@126.com。

据该基因和蛋白的结构和功能,命名为 *ANGPTL4* (angiopoietin-like 4)^[3]。

现有研究表明,类血管生成素 4 除了在血管发生方面同血管生成素 (*ANGPT*) 相似的作用外,它还可以抑制脂蛋白脂肪酶的活性同时正调控血浆甘油三酯水平^[4-5],促进脂肪动员,抑制脂肪贮存^[6]。Le 和 Gbaguidi 研究认为 *ANGPTL4* 是 PPAR γ 的一个下游靶基因, *ANGPTL4* 基因的表达受过氧化物增殖物激活体受体所有配体的激活^[7-8]。既然 PPAR 的靶基因与脂肪酸代谢密切相关,它是调控脂类分化和葡萄糖代谢动态平衡的核受体,因此 *ANGPTL4* 基因可能在脂类代谢和葡萄糖代谢动态平衡的调控方面发挥作用^[9-10]。

目前,在人和鼠上对 *ANGPTL4* 基因的研究报道较多,但是在家畜该基因的研究方面报道很少。马云^[11]克隆了牛的 *ANGPTL4* 基因的 CDS 序列和部分 DNA 序列,运用测序和 PCR-SSCP 结合的方法对该基因的部分基因组 DNA 片段进行了单核苷酸多态检测,同时运用一般线性模型研究了 *ANGPTL4* 基因与 8 个品种牛肉质性状的相关性,初步研究结果表明,牛 *ANGPTL4* 基因的遗传变异与牛育肥后肌内脂肪含量存在密切的关系。研究 *ANGPTL4* 基因的结构和功能,剖析其影响牛的肌间脂肪含量的分子机制,有助于实现肌间脂肪含量性状的遗传改良。为此,本研究利用生物信息学方法对 *ANGPTL4* 基因的功能进行了分析,旨在为牛肉质性状遗传改良提供参考依据。

1 材料与方法

1.1 生物信息学数据库和软件

NCBI(美国生物技术研究中心):<http://www.ncbi.nlm.nih.gov/blast>;EXPASy:<http://www.us.expasy.ch>(蛋白质数据库);<http://www.cbs.dtu.dk/services/SignalP>(信号肽预测);<http://www.ch.embnet.org/software/tepred-form.html>(跨膜结构预测);www.expasy.org/seissmod/swiss-model.html(高级结构预测);DNASar 软件。

1.2 方法

利用 NCBI 搜索牛 *ANGPTL4* 基因的 cDNA 序列,利用 DNASar 软件的 Editseq 程序,对其开放阅读框进行翻译,并分析其翻译后的氨基酸序列;利用 <http://www.expasy.ch/tools/prot-param.html> 网站对 *ANGPTL4* 蛋白质的各种理化性质进行分析;利用 DNASar 软件中的 Megalign 程序进行同源比对;利用 DNASar 软件中的 Protean 程序和 Chou-Fasman 方法分析牛的 *ANGPTL4* 蛋白质的二级结构;利用 <http://www.cbs.dtu.dk/services/SignalP> 网站对 *ANGPTL4* 蛋白的信号肽进行预测;利用网上跨膜结构分析程序 Tampered (www.ch.embnet.org/software/tepred-form.html) 对 *ANGPTL4* 蛋白进行跨膜结构分析;利用 WOLFPsort 网站预测 *ANGPTL4* 蛋白在真核细胞的亚细胞位置;利用 <http://swissmodel.expasy.org/> 网站,预测 *ANGPTL4* 蛋白的高级结构。

2 结果与分析

2.1 牛 *ANGPTL4* 基因基本信息分析

将牛 *ANGPTL4* 基因编码的 410 个氨基酸通过 DNASar 软件的 Editseq 程序进行翻译并分析,其结果表明:牛 *ANGPTL4* 蛋白的相对分子质量约为 45 552.47 Da,且这些氨基酸的组成情况如下:47 个碱性氨基酸(K,R),占 11.46%;40 个酸性氨基酸(D,E),占 9.76%;128 个疏水性氨基酸(A,I,L,F,W,V),占 31.22%;106 个极性氨基酸(N,C,Q,S,T,Y),占 25.85%;该蛋白的等电点为 8.758。程序显示碱基使用情况:A=19.71% [243],G=31.39% [387],T=15.65% [193],C=33.25% [410],Ambiguous=0.00% [0],(A+T)=35.36% [436],(C+G)=64.64% [797]。由此可见,牛 *ANGPTL4* 基因的碱性氨基酸和疏水性氨基酸所占比例较大;且密码子的 4 种碱基使用频率不平衡。CG 含量明显高于 AT 含量,并且无不明确碱基的使用。

2.2 牛 *ANGPTL4* 蛋白的理化性质

ANGPTL4 基因序列编码 410 个氨基酸。利用 <http://www.expasy.ch/tools/prot-param.html> 网站对蛋白质的各种理化性质进行分析。结果显示:牛 *ANGPTL4* 蛋白的原子组成是 C₁₉₉₇ H₃₁₂₆ N₆₁₀ O₅₈₉ S₁₄。其消光系数:在 280 nm 时为 62 950,不稳定系数是 53.12,说明该蛋白是一个不稳定蛋白;脂肪系数为 70.54,总平均亲水性为 -0.649,说明该蛋白是 1 个疏水蛋白。该蛋白质由 20 种氨基酸组成,其中 *Leu*、

Gly、*Gln*、*Ala* 含量最为丰富, *Tyr* 含量最少。

2.3 牛 *ANGPTL4* 蛋白的分子进化分析

将人、猪、小鼠、大鼠、鸡、黑猩猩、狗的 *ANGPTL4* 基因编码蛋白及牛 *ANGPTL4* 蛋白序列调入 DNAS-tar 软件中的 Megalign 程序, 进行系统进化分析。

其结果见图 1, 进化树总体分为两支, 鸡自己成为独立的一支, 剩余 7 个物种成为独立的大分支。在这一大分支中人与黑猩猩聚在一起成为独立的分支, 而猪、牛、狗、小鼠和大鼠另为一分支。猪与牛聚在一起, 再与狗成为一类, 小鼠和大鼠聚为一类。由此表明, 牛的 *ANGPTL4* 基因与猪 *ANGPTL4* 基因分子进化距离最近, 亲缘关系最为密切。

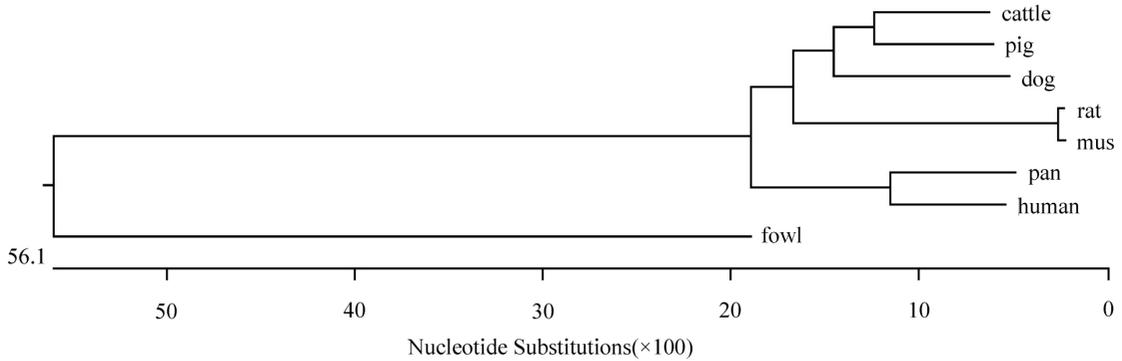


图 1 牛 *ANGPTL4* 氨基酸序列与人及其它的 *ANGPTL4* 蛋白质氨基酸序列分子进化树

Fig.1 Phylogenetic tree of the amino acid sequences of cattle *ANGPTL4* and other *ANGPTL4* proteins in human and other animals

2.4 牛 *ANGPTL4* 蛋白的二级结构预测结果

蛋白质结构、功能及其关系的分析是蛋白质组研究中的一个重要组成部分。通过分析未知功能或新发现的蛋白质分子的结构, 确认其功能单位或结构域, 可以为遗传操作提供目标, 为设计新的蛋白质或改造已有蛋白质提供可靠的依据, 为预测新基因的结构和功能提供参照^[12-14]。图 2 为利用 Protean 软件和在线蛋白质分析系统得到的牛 *ANGPTL4* 蛋白质的结构预测图。

由图 2 利用 Chou - Fasman 方法分析, 得知牛 *ANGPTL4* 蛋白质具有 12 个 α 螺旋, 主要分布在 6 ~ 25, 33 ~ 81, 93 ~ 97, 109 ~ 204, 261 ~ 289, 376 ~ 380 位氨基酸处; 8 个 β 折叠, 主要分布在 209 ~ 212, 232 ~ 236, 294 ~ 324, 368 ~ 374, 385 ~ 391, 396 ~ 404 位氨基酸处; 30 个转角, 主要分布于 29 ~ 32, 62 ~ 158, 192 ~ 220, 229 ~ 260, 275 ~ 302, 317 ~ 366, 381 ~ 395, 405 ~ 408 位氨基酸处。

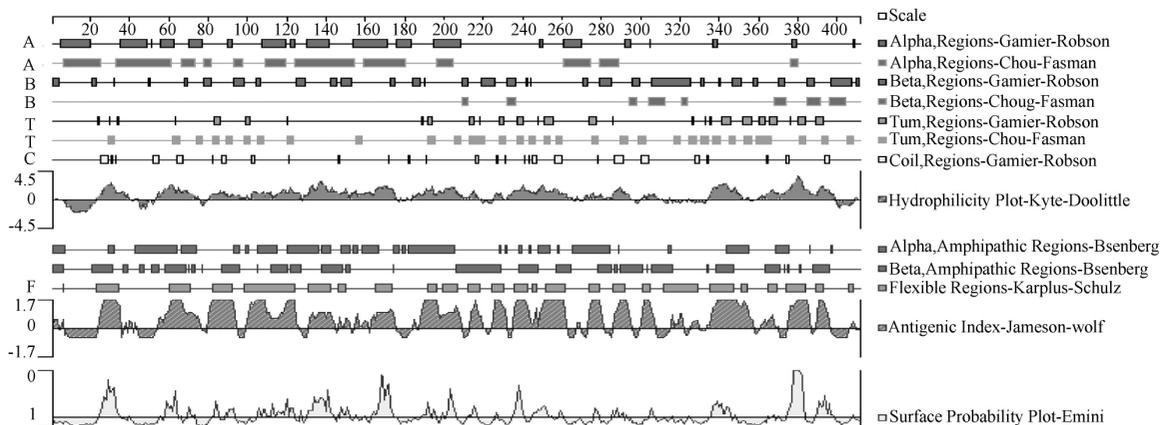


图 2 牛 *ANGPTL4* 蛋白质二级结构图

Fig. 2 Secondary structure of *ANGPTL4* proteins

2.5 牛 *ANGPTL4* 蛋白的信号肽分析

信号肽本身由氨基酸组成, 是蛋白质的一部分, 常位于蛋白质的一端。它将蛋白质导向细胞的正确位置并使蛋白质越过细胞器的膜。预测蛋白质结构的同时对其进行信号肽分析有助于蛋白质功能域的区分及蛋白质细胞定位。利用 <http://www.cbs.dtu.dk/services/SignalP> 网站对编码蛋白信号肽进行预测。

从图 3 中可看出, 标记为 C 的列是剪切位点打分, 标记为 S 的列是信号肽打分, Y 列是综合剪切点打分。只有 C、Y、S 值的计算结论为 YES 的预测结果较好。具有高 C 分值同时又是 S 分值由高转低的位点才是信号肽预测的最可能的剪切点。编码蛋白信号肽预测结果显示, 第 23 ~ 24 位点是最可能的剪切点。

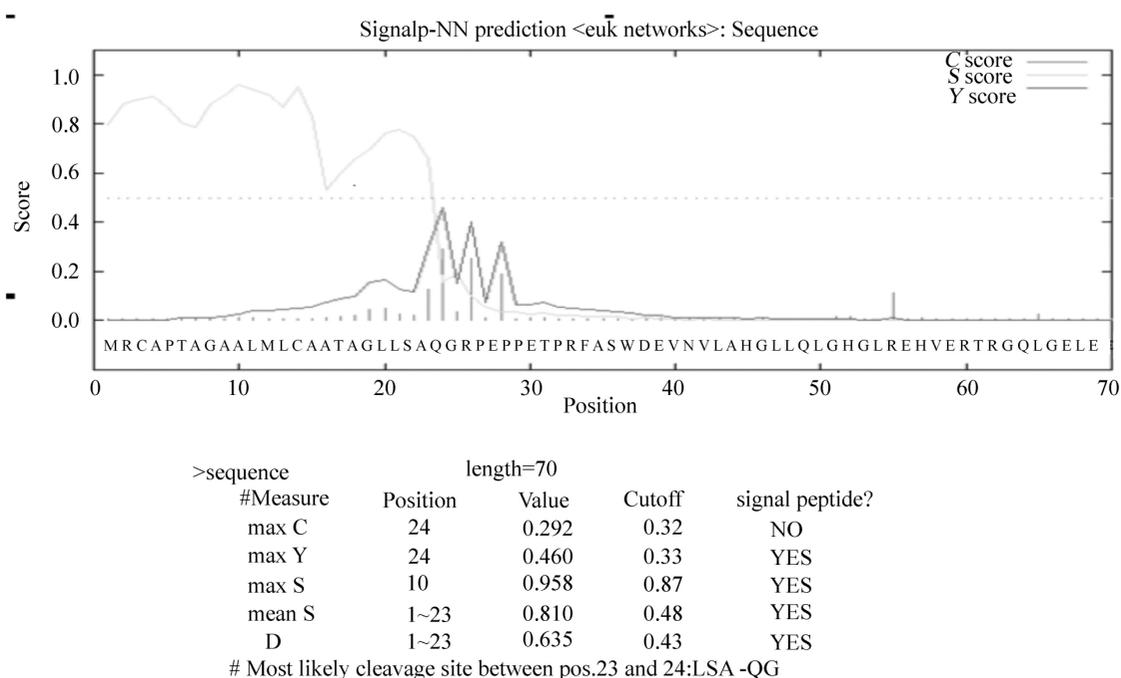


图 3 ANGPTL4 蛋白的信号肽分析

Fig. 3 Analyzed signal peptide of ANGPTL4 proteins

2.6 牛 ANGPTL4 蛋白的跨膜结构分析

跨膜结构域是膜内在蛋白与膜脂相结合的主要部位, 一般由 20 个左右的疏水氨基酸残基组成, 形成一螺旋, 它固着于细胞膜上起“锚定”作用。跨膜结构域的预测和分析, 对正确认识和理解蛋白质的功能、结构、方位及细胞中作用部位均有着重要的指示意义。利用网上跨膜结构分析程序 Tampered (www.ch.embnet.org/software/tepred-form.html) 对编码蛋白进行跨膜结构分析, 结果显示见图 4。

编码蛋白分别在 3 ~ 25, 18 ~ 36, 351 ~ 371 位氨基酸处形成了 3 次高峰, 这 3 次高峰就是它的 3 个跨膜区域。

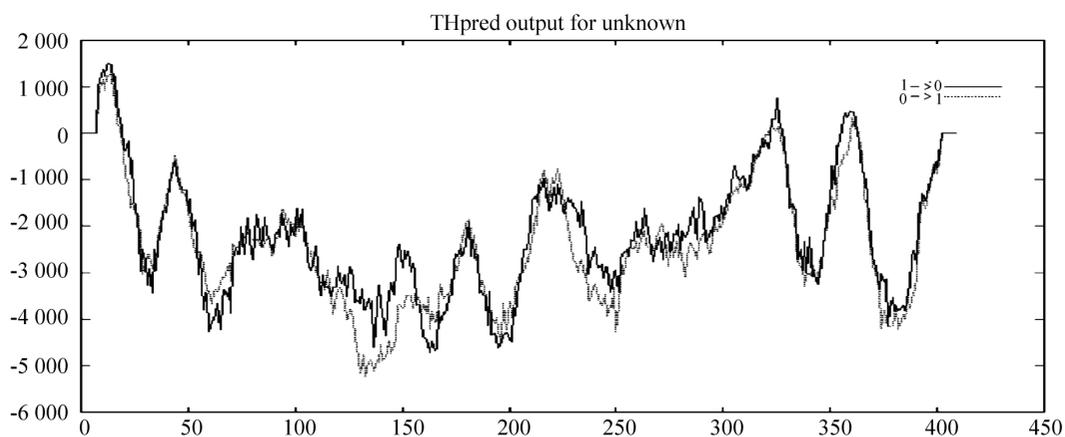


图 4 ANGPTL4 蛋白的跨膜区域分析

Fig. 4 Analyzed transmembrane domains of ANGPTL4 proteins

2.7 牛 ANGPTL4 蛋白的亚细胞定位

蛋白质要参与正常的生命活动, 必须处于特定的亚细胞区域内 (如细胞核、线粒体、细胞质等) 才能发挥特定的功能。现已有很多生物信息工具可以进行蛋白亚细胞定位的预测, WOLFPsort 主要针对蛋

白在真核细胞不同的亚细胞位置进行预测^[15] ,这些亚细胞位置是细胞质、叶绿体、细胞外、线粒体、细胞核以及质膜等。预测的平均准确性达到 93.1%。牛的 *ANGPTL4* 蛋白预测数据如下:K used for kNN is: 32 queryProtein details extr:14.0 pero: 7.0 mito:4.0 , E. R:3.0 lyso:2.0。

K 近邻(K - Nearest Neighbors kNN) 算法是 Keller 等^[16] 在 1985 年提出的。它是一种比较简单的分类算法,广泛应用于机器学习的各个领域。其基本思想是:对于一个没有给定类标签(待分类)的测试数据样本,首先从训练数据集中找出与测试数据最接近(通常定义为距离最近,如欧氏距离等)的 *K* 个近邻,然后通过各个近邻的分类标签来确定测试数据的类别。同时决策算法的实现也有许多方法,比如可以通过统计 *K* 近邻中数量最多的类标签作为测试数据的类别,也可以通过赋给 *K* 个近邻不同的加权系数,然后用模糊决策方法算出系数最大的类标签作为测试数据的类别。从牛的分析结果看出,存在细胞外的数量最多,因此判断该细胞位于细胞外。

2.8 牛 *ANGPTL4* 蛋白的高级结构

蛋白质高级结构的预测和分析,对理解蛋白质结构和功能之间的关系有着极其重要的意义。为了获得编码蛋白的高级结构模拟图,先把编码的氨基酸序列通过网上在线服务器,网址为 <http://swissmodel.expasy.org/>,*ANGPTL4* 蛋白的高级结构见图 5。



图 5 *ANGPTL4* 蛋白的跨膜区域分析

Fig.5 Three - dimensional structure of *ANGPTL4* proteins

3 讨 论

在本研究中,通过蛋白质的亚细胞定位、信号肽、跨膜结构预测,分析牛 *ANGPTL4* 基因编码蛋白功能。并且通过进化比对,发现牛的 *ANGPTL4* 基因与猪 *ANGPTL4* 基因的分子进化距离最近,亲缘关系最为密切。这与动物学分类结果一致,这种同源性在一定程度上代表着物种亲缘关系的远近,同时也反映了 *ANGPTL4* 蛋白在不同物种上结构上的稳定性对生物体的功能重要性^[17]。

从分析结果中可知,牛的 *ANGPTL4* 蛋白位于细胞外,并且预测出信号肽的剪切位点,已知信号肽位于分泌蛋白的 N 端,它对于外分泌蛋白的分泌起主导作用^[18-19]。Antonio 等^[20]发现,改造了的信号肽可大大提高了外源蛋白表达量。因此,在我们预测的信号肽位置进行修饰,可以提高编码蛋白在牛体内的表达。

蛋白质的结构决定其功能,因此可以通过亲缘关系,从其它物种 *ANGPTL4* 基因的功能来推测牛的 *ANGPTL4* 基因所具有的生物学功能。Le Jan^[7]等研究发现,重组 *ANGPTL4* 通过 CHO 细胞移植入鸡胚尿囊膜,它被 CHO/*ANGPTL4* 细胞释放,引起明显的新生血管形成及血管的重新分布。通过静脉注射重组 *ANGPTL4* 蛋白,Yoshida 等^[21]发现 *ANGPTL4* 可有效的升高血浆甘油三酯的浓度。用重组 *ANGPTL4* 蛋白和脂蛋白脂肪酶做体外实验,证实 *ANGPTL4* 蛋白可抑制脂蛋白脂肪酶(LPL)活性,这为它的高甘油三酯血症效应提供了比较合理的解释。腺病毒介导 *ANGPTL4* 基因的过度表达证实了这些,并显示病毒入侵若干天后,血浆甘油三酯水平急剧升高^[10-22]。除了调节脂类代谢之外,现有资料表明 *ANGPTL4* 还能调节血浆葡萄糖水平。Xu 等^[10]研究发现腺病毒介导的 *ANGPTL4* 基因超表达小鼠表现为血糖水平的降低(可能是通过抑制肝脏葡萄糖的产生)、葡萄糖耐受性的增强、高脂血症、脂肪肝和肝肿大。丰胜求^[23]以梅山猪为材料克隆了包含 *ANGPTL4* 在内的 *ANGPTL* 家族 5 个基因,利用辐射杂种细胞板技术将 *ANGPTL4* 基因定位在猪 2 号染色体上;利用半定量 RT - PCR 方法对该基因在梅山猪十个组织的表达谱进行了分析,结果发现猪 *ANGPTL4* 基因为广谱表达,在脂肪组织和肝脏表达最丰富;Xu 等^[10]以人肝癌细胞系 HepG2 为材料,研究了葡萄糖、胰岛素和地塞米松对 *ANGPTL4* 的调控作用,结果发现,高浓度葡萄糖处理可以显著增加 *ANGPTL4* 基因在 mRNA 水平的表达量,而胰岛素处理能显著抑制该基因的表达,地塞米松处理肝癌细胞能显著增加 *ANGPTL4* 基因的表达。据此可以在牛上探讨 *ANGPTL4* 基因及其编码的蛋白与牛相应的生物经济学性状的关系。

本研究是对牛 *ANGPTL4* 基因的初步探索,本研究的结果将有利于对牛 *ANGPTL4* 基因和其编码的 *ANGPTL4* 蛋白的功能进行深入研究,对进一步了解 *ANGPTL4* 基因对牛脂肪蓄积、肌肉生长的基因效应,以及利用该基因改良牛经济性状具有重要的意义。

参考文献:

- [1] Kersten S, Mandart S, Tan N S, et al. Characterization of the fasting - induced adipose factor FIAF, A novel peroxisome proliferator - activated receptor target gene [J]. *J Biol Chem* 2000, 275(37):28488 - 28493.
- [2] Yoon J C, Chickering T W, Rosen E D, et al. Peroxisome proliferator - activated receptor gamma target gene encoding a novel angiopoietin - related protein associated with adipose differentiation [J]. *Mol Cell Biol* 2000, 20(14):5343 - 5349.
- [3] Wan D F, Gong Y, Wen X Q, et al. Large - scale cDNA transfection screening for genes related to cancer development and progression [J]. *Proc Natl Acad Sci, USA* 2004, 101(44):15724 - 15729.
- [4] Sukonina V, Lookene A, Olivercrona T, et al. Angiopoietin - like protein 4 converts lipoprotein lipase to inactive monomers and modulates lipase activity in adipose tissue [J]. *Proc Natl Acad Sci, USA* 2006, 103(46):17450 - 17455.
- [5] Lichtenstein L, Berbee J F, Van Dijk S J, et al. *ANGPTL4* upregulates cholesterol synthesis in liver via inhibition of LPL and HL - dependent hepatic cholesterol uptake [J]. *Arterioscler Thromb Vasc Biol* 2007, 27(11):2420 - 2427.
- [6] Zhu H, Li J, Qin W, et al. Cloning of a novel gene, *ANGPTL4* and the functional study in angiogenesis [J]. *Zhonghua Yi Xue Za Zhi* 2002, 82(2):94 - 99.
- [7] Le Jan S, Amy C, Cazes A, et al. Angiopoietin - like 4 is a proangiogenic factor produced during ischemia and in conventional renal cell carcinoma [J]. *Am J Pathol* 2003, 162(5):1521 - 1528.
- [8] Gbaguidi F G, Chinetti G, Milosavljevic D, et al. Peroxisome proliferator - activated receptor (PPAR) agonists decrease lipoprotein lipase secretion and glycated LDL uptake by human macrophages [J]. *FEBS Lett* 2002, 512(113):85 - 90.
- [9] Koster A, Chao Y B, Mosior M, et al. Transgenic angiopoietin - like (angptl)4 overexpression and targeted disruption of *ANGPTL4* and *ANGPTL3*: regulation of triglyceride metabolism [J]. *Endocrinology* 2005, 146(11):43 - 50.
- [10] Xu A, Lam M C, Chan K W, et al. Angiopoietin - like protein 4 decreases blood glucose and improves glucose tolerance but induces hyperlipidemia and hepatic steatosis in mice [J]. *Proc Natl Acad Sci USA* 2005, 102(9):6086 - 6091.
- [11] 马云. 牛三个基因的分离、定位及其与牛经济性状的关联研究 [D]. 杨凌:西北农林科技大学, 2006.
- [12] Simon K O, Nutt E M D G, Abraham G A, et al. The V_3 integrin regulates 51 - mediated cell migration toward fibronectin [J]. *Biol Chem*, 1997, 272(46):29380 - 29389.
- [13] Kiosses W B, Shattil S J, Pampori N, et al. Rac recruits high - affinity integrin V_3 to lamellipodia in endothelial cell migration [J]. *Nat Cell Biol* 2001, 3(3):316 - 320.
- [14] Bader B L, Rayburn H, Crowley D, et al. Extensive vasculogenesis, angiogenesis and organogenesis precede lethality in mice lacking all α_v integrins [J]. *Cell*, 1998, 95(4):507 - 519.
- [15] Lu Z, Szafron D, Greiner R, et al. Predicting subcellular localization of proteins using machine - learned classifiers [J]. *Bioinformatics* 2004, 20(3):547 - 556.
- [16] Keller J M, Gray M R, Givens J A. A fuzzy k - nearest neighbor algorithm [J]. *IEEE Trans Syst Man Cybern*, 1985, 15(2):580 - 585.
- [17] Walter J, High K A. Gene therapy for the hemophiliac [J]. *Adv Vet Med*, 1997, 23(12):119 - 134.
- [18] Izard J W, Doughty M B, Kendall D A. Physical and conformational properties of synthetic idealized signal sequences parallel their biological function L [J]. *Biochemistry*, 1995, 34(31):9904 - 9906.
- [19] 杨运桂, 徐京宁, 胡泰山. 信号肽疏水性的提高促进青霉素 G 酰化酶分泌 [J]. *生物化学与生物物理学报*, 2000, 32(2):163 - 168.
- [20] Antonio M R, Martinez A. Secretion of recombinant pro - and mature fungal - sarenin ribo toxin by the methylotrophic yeast *Pichia pastoris*: The Lys - arg motif is required for maturation [J]. *Protein Expr Purif*, 1998, 12(6):315 - 322.
- [21] Yoshida K, Shimizugawa T, Ono M. Angiopoietin - like protein 4 is a potent hyperlipidemia - inducing factor in mice and inhibitor of lipoprotein lipase [J]. *J Lipid Res* 2002, 43(11):1770 - 1772.
- [22] Ge H, Yang G, Yu X. Oligomerization state dependent hyperlipidemic effect of angiopoietin - like protein 4 [J]. *J Lipid Res* 2004, 45(12):2071 - 2079.
- [23] 丰胜求. 猪 *Angptl* 家族基因的克隆、表达及调控研究 [D]. 武汉:华中农业大学, 2006.