

# 猪 *Musclin* 基因编码蛋白的生物信息学分析

孙超, 王力, 齐仁立, 姜东凤

(西北农林科技大学 动物科技学院, 陕西 杨凌 712100)

**[摘要]** 【目的】克隆猪 *Musclin* 基因, 分析和预测其编码蛋白的结构与功能。【方法】从猪肌肉组织中提取总 RNA, RT-PCR 扩增获得 *Musclin* 基因, 对其进行克隆和测序, 并根据生物信息学分析方法, 利用 Tmpred、SignalP 3.0、TargetP 1.1 等分析软件, 对猪 *Musclin* 蛋白的跨膜结构、信号肽、细胞定位、结构特征等进行分析 and 预测。【结果】猪 *Musclin* 蛋白主要集中在分泌途径上, 存在于细胞外, 其核苷酸序列与家兔的具有较高的相似性, 氨基酸序列 32~45 位和 55~63 位存在 2 个可能的强跨膜螺旋区, 该蛋白的空间结构以  $\alpha$  螺旋、 $\beta$  转角和无规则卷曲为主; 其 1~26 位氨基酸序列是一段信号肽, 在 26 与 27 位氨基酸之间存在 1 个酶切位点; 其羧基端含有较多的转角结构, 同时该区域的抗原指数较高, 亲水性指数和呈现在蛋白表面的可能性均较大。【结论】*Musclin* 是一种含有信号肽序列的跨膜蛋白。

**[关键词]** *Musclin*; 生物信息学; 结构与功能; 序列分析

**[中图分类号]** S828.2

**[文献标识码]** A

**[文章编号]** 1671-9387(2010)01-0030-05

## Analysis of the structure and function of porcine *Musclin* protein by bioinformatics

SUN Chao, WANG Li, QI Ren-li, JIANG Dong-feng

(College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China)

**Abstract:** 【Objective】 This study was to clone the gene of porcine *Musclin* protein as well as analyze and predict structures of this protein. 【Method】 Total RNA was extracted from porcine muscle tissue. Reverse transcription-polymerase chain reaction (RT-PCR) was used to detect *Musclin* gene. According to the analysis methods of bioinformatics, the research applied many analysis programs, e. g. Tmpred, SignalP 3.0, TargetP 1.1 to analyze and predict the structural characteristics of porcine *Musclin* protein, its transmembrane helices and signal peptide. 【Result】 Most of *Musclin* exist in extracellular space. There is a high similarity of *Musclin* nucleotide sequence between swine and rabbit, and the main structures of *Musclin* protein are  $\alpha$ -spirals,  $\beta$ -turn and random coil. 1-26 amino acid is a signal peptide and there is a enzyme cutting site between 26 and 27. Moreover,  $\beta$ -turn is abundant in the C-terminal region, and this domain is supposed to be the epitopes of the B cell. This results would be helpful for the research of immunological function of *Musclin*. 【Conclusion】 *Musclin* is a transmembrane protein which comprises signal peptides.

**Key words:** *Musclin*; bioinformatic; structure and function; sequence analysis

*Musclin* 是 Nishizawa<sup>[1]</sup> 于 2004 年用信号序列捕捉技术在小鼠中率先发现的, 其由动物肌肉组织特异性分泌, 可减弱肌细胞的葡萄糖摄取和糖元合

成作用, 并参与胰岛素的负反馈调节作用。Banzet<sup>[2]</sup> 研究发现, *Musclin* 表达与快肌的糖酵解过程存在密切关系。此外, *Musclin* 在脂肪代谢过程

\* [收稿日期] 2009-04-22

[基金项目] 国家自然科学基金项目(30871785); 教育部重点项目(105167); 教育部新世纪优秀人才计划项目(NCET-06-0865); 国家“十一五”科技支撑计划项目(2006BAD04A11)

[作者简介] 孙超(1968-), 男, 陕西商南人, 教授, 博士, 主要从事动物遗传育种研究。E-mail: sunchao2775@163.com

中也可能具有重要的调控作用。Yasui 等<sup>[3]</sup>发现,脂肪代谢中的一个重要转录因子 Foxo1 可抑制 *Musclin* 表达。王伟杰等<sup>[4]</sup>研究了猪 *Musclin* 基因在 BL21 大肠杆菌中的表达,并获得融合蛋白 GST-*Musclin*。姜东风等<sup>[5]</sup>认为,猪 *Musclin* 基因除在肌肉中表达外,在脂肪等其他组织中也能高度表达;且该基因的表达与猪的品种和生长阶段及一些生脂基因(如 FAS、PPAR $\gamma$  等)有密切关系。Staiger 等<sup>[6]</sup>发现,PPAR $\gamma$  激动剂 troglitazone 可诱导人肌细胞 *Musclin* 表达。本试验结合他人常用的、具有较高准确性和可靠性的分析手段<sup>[7-9]</sup>,利用生物信息学的分析方法,对猪 *Musclin* 蛋白的序列结构进行预测分析,以期为进一步研究 *Musclin* 基因的结构、表达与调控及生物学功能奠定基础。

## 1 材料与方法

### 1.1 材料

1.1.1 试验动物 10 月龄健康关中黑猪,由西北农林科技大学畜牧场提供,屠宰后取其骨骼肌作为试验材料。

1.1.2 试剂 RNA 提取试剂盒 TRI pure Reagent,百泰克公司产品;反转录试剂盒 Revert Aid TM First Strand cDNA Synthesis Kit, TaKaRa 公司产品;Taq 酶, Fermentas 公司产品;Marker II,购自北京天为时代科技有限公司;DNA 回收试剂盒,购自安徽优品生物工程公司;pMD-18T vector, TaKaRa 公司产品;E. coli DH5 $\alpha$ ,由西北农林科技大学蛋白质与酶工程实验室保存;质粒微量提取试剂盒,购自北京天为时代科技有限公司。

### 1.2 方法

1.2.1 引物设计 根据 GenBank 中已发表的人、小鼠和大鼠 *Musclin* 基因序列,应用 Primer Premier 5.0 软件,设计了 *Musclin* 基因引物,预计扩增片段长度 374 bp,由上海生物工程技术有限公司合成。上游引物序列:5'-ATGGACTGGAGACTGGCAAG-3',下游引物序列:5'-CGGTTTCTACCAATCCGATC-3'。

1.2.2 总 RNA 提取及 cDNA 第一链的合成 总 RNA 提取按照 TRI pure Reagent 试剂盒说明进行,用 10 g/L 琼脂糖凝胶检测。按照 Revert Aid TM First Strand cDNA Synthesis Kit 反转录试剂盒推荐的方法,以 Random hexamers primer 为引物合成 cDNA 第一链。

1.2.3 PCR 扩增 PCR 反应体系:14.3  $\mu$ L 灭菌

的双蒸水,2.5  $\mu$ L 10 $\times$  PCR Buffer,2.5  $\mu$ L 2.5 mmol/L dNTP Mix,1  $\mu$ L 10  $\mu$ mol/L Primer I,1  $\mu$ L 10  $\mu$ mol/L Primer II,0.2  $\mu$ L 5 U/ $\mu$ L Taq 酶,1  $\mu$ L cDNA 模板。PCR 反应条件:95  $^{\circ}$ C 预变性 10 min;94  $^{\circ}$ C 变性 40 s,56  $^{\circ}$ C 退火 40 s,72  $^{\circ}$ C 延伸 40 s,34 个循环;最后 70  $^{\circ}$ C 延伸 10 min;4  $^{\circ}$ C 保存。用 10 g/L 琼脂糖凝胶检测 PCR 扩增产物。

1.2.4 猪 *Musclin* 基因的克隆和测序 按 DNA 回收试剂盒说明回收特异性条带,连接 pMD-18T,转化感受态细胞 E. coli DH5 $\alpha$ ,进行蓝白斑筛选,挑取阳性克隆接种于 LB 培养基,过夜培养后用质粒微量提取试剂盒抽提质粒 DNA,经 PCR 扩增初步鉴定重组质粒,并将鉴定为阳性的质粒送上海生物工程技术有限公司测序。

1.2.5 猪 *Musclin* 蛋白的跨膜分析 用 Tmpred 跨膜结构分析软件,分析 *Musclin* 蛋白跨膜区域及方向。

1.2.6 猪 *Musclin* 蛋白的信号肽分析 用蛋白质序列信号肽在线分析工具 SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>),预测 *Musclin* 蛋白的信号肽序列。

1.2.7 猪 *Musclin* 蛋白的细胞定位分析 分别利用在线亚细胞定位工具 TargetP 1.1 (<http://www.cbs.dtu.dk/services/TargetP/>)和 PSORT II prediction (<http://psort.nibb.ac.jp/form2.html>)进行 *Musclin* 蛋白亚细胞定位。

1.2.8 猪 *Musclin* 蛋白的结构特征分析 用 Gamier-Robson(Gamier)、Chou-Fasman(Chou)和 Karplus-Schulz 方法预测 *Musclin* 蛋白的结构特征;用 Kyte-Doolittle 方法对 *Musclin* 蛋白的亲水性进行分析;采用 Emini 方法预测 *Musclin* 蛋白的表面可能性;用 Jameson-Wolf 方法预测该蛋白的抗原指数。

## 2 结果与分析

### 2.1 猪 *Musclin* 基因的克隆与序列分析

2.1.1 目标基因的克隆 猪的肌肉组织总 RNA 提取结果见图 1。图 1 显示,5 S、18 S、28 S 3 条带清晰可见,且无污染,符合分子生物学试验的要求。PCR 产物连接至克隆载体 pMD-18T 后,转化感受态 E. coli DH5 $\alpha$  细胞,抽提质粒 DNA 后经 PCR 扩增鉴定为阳性克隆。由图 2 可知,扩增产物条带特异,且大小与预期目标条带相同。

2.1.2 基因序列分析 将克隆产物经测序,得到 374 bp 核苷酸序列,该序列与小鼠骨骼肌的 *Musclin* cDNA 结构一致,均有丝氨酸蛋白酶切割位点

“KKKR”,可确定为目标序列,提交 GenBank(登陆号为 EU122441)。所得基因序列通过 DNAMAN 软件推导出一段长 122 个氨基酸的序列。经比对分析,猪 *Musclin* 的核苷酸序列与人、大鼠、小鼠、牛、绵羊、家兔和鸡的相似性依次为 86%,78%,72%,

88%,88%,98%和 59%;氨基酸序列相似性分别为 87%,72%,72%,84%,83%,98%和 58%,表明猪 *Musclin* 与家兔、牛、绵羊和人的序列有较高的相似性,而与鸡的序列相似性最低。

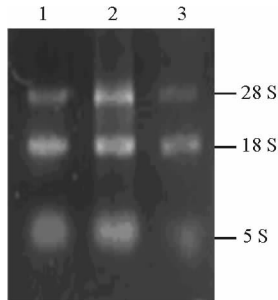


图 1 猪肌肉组织总 RNA 的提取结果  
1~3. 猪骨骼肌总 RNA

Fig. 1 Result of total RNA extracted from pig muscle  
1-3. Total RNA extracted from pig muscle

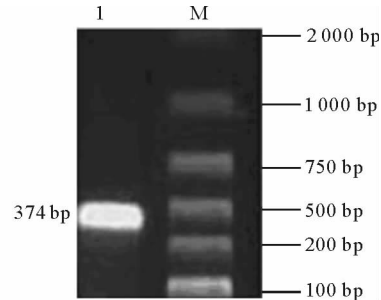


图 2 猪 *Musclin* 基因重组质粒的 PCR 鉴定结果  
1. *Musclin* 基因;M. DNA Marker

Fig. 2 Result of PCR identification of recombinant plasmid of porcine *Musclin* gene  
1. *Musclin* gene ;M. DNA Marker

### 2.2 猪 *Musclin* 蛋白的跨膜分析

利用 Tmpred 工具分析猪 *Musclin* 蛋白的亲水区域,预测其跨膜结构及方向,结果(图 3)表明,猪 *Musclin* 蛋白可能是膜蛋白,且从内向外和从外向内分别含有 1 个强跨膜螺旋区,其中由内向外的螺旋位置在 32~45 位氨基酸处,由外向内的螺旋位置

在 55~63 位氨基酸处。

### 2.3 猪 *Musclin* 蛋白的信号肽分析

经 SignalP 3.0 预测猪 *Musclin* 蛋白的信号肽序列,结果见图 4。由图 4 可知,猪 *Musclin* 蛋白第 1~26 位氨基酸序列是一段信号肽,26 与 27 位氨基酸之间存在 1 个酶切位点。

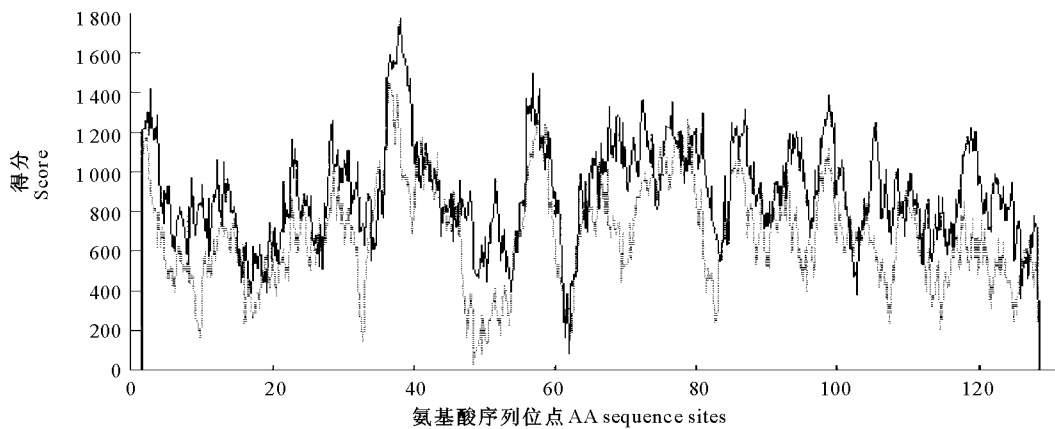


图 3 猪 *Musclin* 蛋白的跨膜螺旋分析

Fig. 3 Prediction of transmembrane helices in porcine *Musclin*

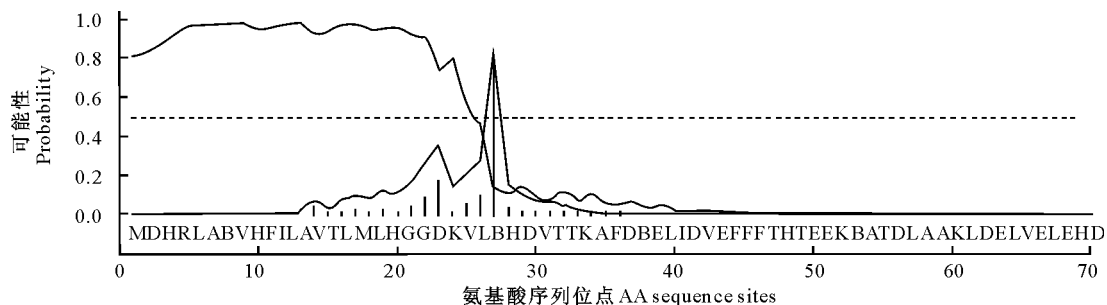


图 4 猪 *Musclin* 蛋白的信号肽分析

Fig. 4 Analysis of the signal peptide in porcine *Musclin*

2.4 猪 *Musclin* 蛋白的细胞定位分析

利用 TargetP 1.1 和 PSORT II prediction 分析猪 *Musclin* 蛋白的细胞定位,结果(表 1)表明,猪 *Musclin* 蛋白主要集中在分泌途径上,线粒体中很

少,说明该蛋白应属于分泌型蛋白。猪 *Musclin* 蛋白主要分布在细胞外部,包括细胞壁,只有少数分布在胞内胞质中。

表 1 猪 *Musclin* 蛋白亚细胞定位分析

Table 1 Predicts the subcellular location of porcine *Musclin*

软件 Predict software	亚细胞定位 Subcellular location	比例/% Proportion	软件 Predict software	亚细胞定位 Subcellular location	比例/% Proportion
TargetP 1.1	分泌通路信号位点 (SP) Secretory pathway signal peptide	95.6	PSORT II prediction	内质网 Endoplasmic reticulum	11.1
	线粒体作用位点 (MTP) Mitochondrial targeting peptide	2.0		细胞外(包括细胞壁) Extracellular (including cell wall)	77.8
	其他 Other	2.4		液泡 Vacuolar	11.1

2.5 猪 *Musclin* 蛋白的结构特征分析

通过不同软件和算法对猪 *Musclin* 蛋白的结构和特性进行预测分析,结果见图 5。图 5-A 显示,猪 *Musclin* 蛋白主要以  $\alpha$  螺旋为主,很少以  $\beta$  折叠存在。用 Chou-Fasman 和 Gamier-Robson 2 种方法

分析均表明,猪 *Musclin* 蛋白  $\alpha$  螺旋位置主要在 20~55 区段。用 Kyte-Doolittle 方法分析结果显示,该蛋白存在众多亲水性区域,而且亲水性指数较高,其中 C 端亲水性最高,所跨区段也最大,提示该区域暴露于表面的几率较大(图 5-B)。

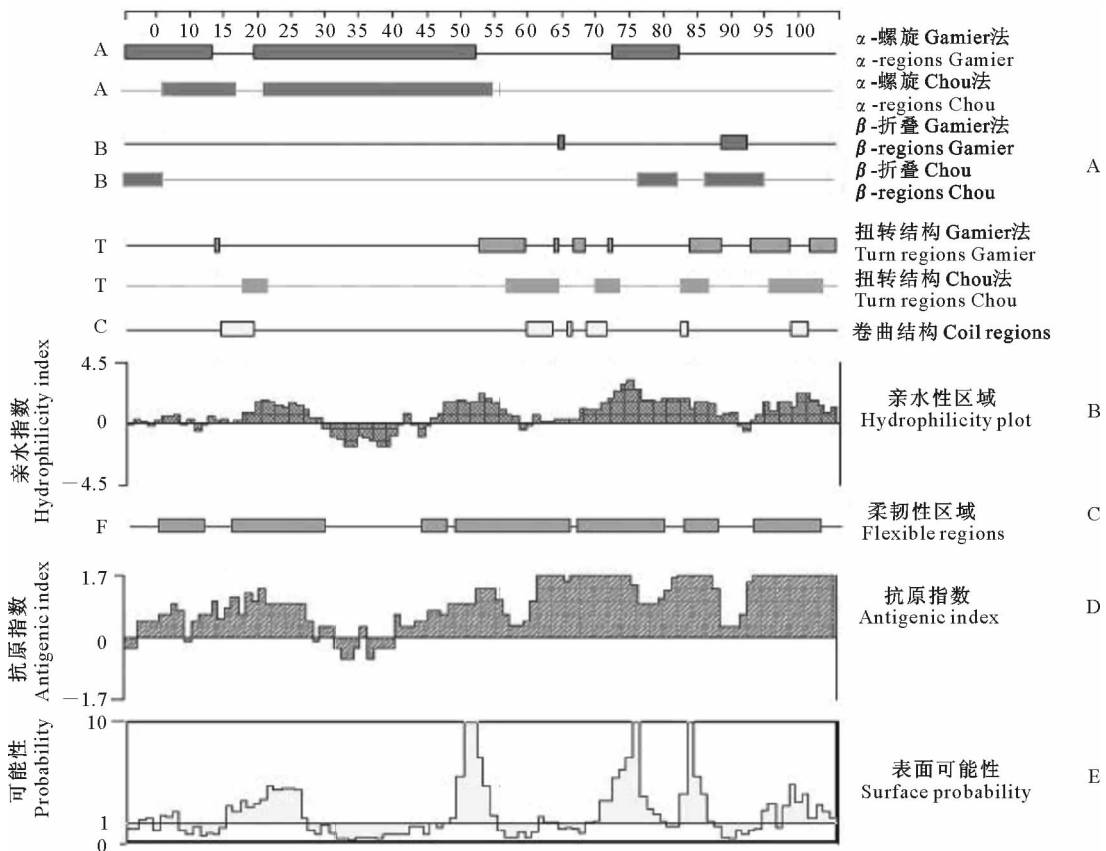


图 5 采用不同方法预测的猪 *Musclin* 蛋白结构和特性

Fig. 5 Secondary structure and characteristics of porcine *Musclin* predicted by different methods

用 Karplus-Schulz 方法对猪 *Musclin* 蛋白骨架区的柔韧性进行分析发现,该蛋白骨架区含有较多的柔韧性区域,说明该蛋白肽段的柔韧性较大,发生扭曲、折叠几率较高,能形成丰富的二级结构(图 5-C)。用 Jameson-Wolf 方法分析猪 *Musclin* 蛋白潜在的抗原决定簇发现,该蛋白含有大量抗原指数较

高的区域,其中 C 端的最高,其所跨区域也最大(图 5-D)。利用 Emini 方法分析猪 *Musclin* 蛋白的表面可能性区域,发现其氨基酸呈现在表面可能性较大的区域,主要集中在 50~55,70~78 和 82~87 区段,其他部位可能性较小(图 5-E)。

### 3 讨 论

目前,在蛋白质结构测定中存在分析难度大、费用高等问题,且某些蛋白质因难以形成晶体而无法进行结构分析,因此依据蛋白质序列预测蛋白质可能的结构就显得非常重要<sup>[10]</sup>。PHD 系统被认为是目前最好的蛋白二级结构预测程序,其平均准确率超过 72%,最佳者达 90%以上<sup>[11-12]</sup>。本试验为了提高预测的准确性,对蛋白相关性分析都选用了多个不同的软件进行综合比较。由于不同预测程序的测序原理和算法的侧重点不同,所以在其预测结果中出现的一致性结果具有较高的可信度。本研究发现,猪 Musclin 蛋白 1~26 位氨基酸序列是一段信号肽,这与王伟杰等<sup>[4]</sup>的研究结果基本一致。猪 Musclin 蛋白可通过信号肽引导从合成部位移到细胞膜壁,最后分泌到胞外,这与 Nishizawa<sup>[1]</sup>将 Musclin 定义为骨骼肌分泌因子的结果相符。

预测蛋白序列中跨膜螺旋的位置和数目是当前生物信息学中重要的研究课题,因为通过对膜蛋白跨膜结构的预测能为该蛋白的功能研究提供重要线索。本研究发现,在猪 Musclin 的氨基酸序列 32~45 以及 55~63 位存在 2 个可能的强跨膜螺旋区,其具体生物学功能有待进一步研究。Musclin 蛋白二级结构以  $\alpha$  螺旋、 $\beta$  转角和无规则卷曲结构为主,属混合型蛋白<sup>[13]</sup>。通常,转角和无规则卷曲区域决定蛋白质的功能,其中  $\beta$  转角结构为凸出结构,多出现在蛋白质抗原表面,利于与抗体嵌合,可能与抗原表位有关<sup>[14]</sup>。转角及无规则卷曲等二级结构是比较松散的结构,易于发生扭曲、盘旋并展示在蛋白的表面,因此该区域常含 B 细胞的优势抗原表位<sup>[15]</sup>。本研究表明,猪 Musclin 蛋白羧基端含有较多的转角结构,同时该区域抗原指数较高,其亲水性指数及呈现在蛋白表面的可能性也较大,因此该区段的抗原表位应该是优势抗原表位所在。此结果对进行 Musclin 蛋白的免疫功能研究具有一定的指导意义。

#### [参考文献]

- [1] Nishizawa H. Musclin, a novel skeletal muscle-derived secretory factor [J]. *Journal of Biological Chemistry*, 2004, 279(19): 19391-19395.
- [2] Banzet S. Musclin gene expression is strongly related to fast-glycolytic phenotype [J]. *Biochemical and Biophysical Research Communications*, 2007, 353(3): 713-718.
- [3] Yasui A, Hitoshi N, Yosuke O, et al. Foxo1 represses expression of Musclin, a skeletal muscle-derived secretory factor [J]. *Biochemical and Biophysical Research Communications*, 2007, 364(2): 358-365.
- [4] 王伟杰,李宏基,韩立强,等.猪肌肉素基因的 cDNA 克隆与表达 [J]. *生物工程学报*, 2008, 24(7): 1248-1252.  
Wang W J, Li H J, Han L Q, et al. Cloning and expression of the pig skeletal muscle *Musclin* gene [J]. *Chinese Journal of Biotechnology*, 2008, 24(7): 1248-1252. (in Chinese)
- [5] 姜东风,孙超,彭永佳.猪 *Musclin* 基因的组织表达特征及其与生脂基因的相关性分析 [J]. *动物学报*, 2008, 54(3): 453-459.  
Jiang D F, Sun C, Peng Y J. Expression characterization of *Musclin* gene in porcine tissues and its correlation with lipogenic genes [J]. *Acta Zoologica Sinica*, 2008, 54(3): 453-459. (in Chinese)
- [6] Staiger H, Haas C, Machicao F, et al. The PPARgamma agonist troglitazone induces musclin mRNA expression in human myotubes [J]. *Horm Metab Res*, 2006, 38(9): 614-616.
- [7] Enright A J, Iliopoulos I, Kyrpides N C, et al. Protein interaction maps for complete genomes based on gene fusion events [J]. *Nature*, 1999, 40(2): 86-90.
- [8] Ikeda M, Arai M, Lao D M, et al. Transmembrane topology prediction methods: A reassessment and improvement by a consensus method using a dataset of experimentally characterized transmembrane topologies [J]. *In Silico Biol*, 2002, 2(1): 19-33.
- [9] 李明,赵巧辉,陈其新,等.家兔 BMP7 基因的克隆及其生物信息学分析 [J]. *遗传*, 2008, 30(7): 885-892.  
Li M, Zhao Q H, Chen Q X, et al. Cloning and bioinformatic analysis of bone morphological protein 7 (BMP7) in rabbit [J]. *Hereditas*, 2008, 30(7): 885-892. (in Chinese)
- [10] 郝柏林,张淑誉.生物信息学手册 [M]. 上海:上海科学技术出版社, 2002: 250-256.  
Hao B L, Zhang S Y. Handbook of bioinformatics [M]. Shanghai: Shanghai Sci-Tech Press, 2002: 250-256. (in Chinese)
- [11] Rost B, Sander C. Combining evolutionary information and neural networks to predict protein secondary structure [J]. *Proteins*, 1994, 19(1): 55-72.
- [12] Rost B, Yachdav G, Liu J. The predict protein server [J]. *Nucleic Acids Res*, 2004, 32: 321-326.
- [13] Skolnick J, Fetrow J S. From genes to protein structure and function: novel applications of computational approaches in the genomic era [J]. *Trends in Biotechnology*, 2000, 18(1): 34-39.
- [14] 宋建勋,朱锡华,陈克敏.人 Fas 抗原表位预测 [J]. *免疫学*, 1999, 15(1): 14.  
Song J X, Zhu X H, Chen K M. Prediction of B cell epitopes of human FAS (H FAS) antigen [J]. *Immunological Journal*, 1999, 15(1): 14. (in Chinese)
- [15] Blythe M, Flower D. Benchmarking B cell epitope prediction: underperformance of existing methods [J]. *Protein Sci*, 2005, 14: 1246-1248.