

绵羊GP5基因的生物信息学分析

宋雅萍, 李彦霞, 郭文婧, 张小雪

(甘肃农业大学动物科学技术学院, 甘肃 兰州 730070)

摘要: 为探讨绵羊血小板糖蛋白 V 基因(Glycoprotein V, Platelet, GP5)编码蛋白的结构和功能, 以 NCBI 网站的 GenBank 数据库中发布的绵羊 GP5 基因序列(登录号为 XM_027965957.1)为基础, 采用生物信息学在线工具和软件对绵羊的 GP5 基因进行生物信息学分析。结果表明, 绵羊 GP5 基因序列中包含 1 个最大长度为 1 620 bp 的开放阅读框, 推测其编码 539 个氨基酸, 其中亮氨酸含量最多, 为 22.6%。GP5 基因所编码的蛋白其相对分子质量约为 59 305.38 KDa, 理论等电点为 9.40; 亚细胞定位结果表明其主要位于内质网(44.4%)。GP5 蛋白含有信号肽和跨膜螺旋结构, 在二级结构预测中, 该蛋白以无规卷曲为主, 为 70.14%。GP5 蛋白三级结构主要由无规卷曲折叠缠绕形成, 与二级结构预测结构一致。

关键词: 绵羊; GP5 基因; 生物信息学分析

中图分类号: S826 **文献标志码:** A **文章编号:** 1001-1463(2020)09-0054-06

[doi:10.3969/j.issn.1001-1463.2020.10.012](https://doi.org/10.3969/j.issn.1001-1463.2020.10.012)

Bioinformatics Analysis of Sheep GP5 Gene

SONG Yaping, LI Yanxia, GUO Wenjing, ZHANG Xiaoxue

(College of Animal Science and Technology, Gansu Agricultural University, Lanzhou Gansu 730070, China)

Abstract: In order to investigate the structure and function of the protein encoded by Glycoprotein V, Platelet (GP5), this paper uses bioinformatics online tools and software to analyze the GP5 gene of sheep based on the GP5 gene sequence of sheep published in GenBank database of NCBI website (registration number: XM_027965957.1). The results showed that GP5 gene sequence of sheep contained an open reading frame with a maximum length of 1 620 bp, and it was speculated that it encoded 539 amino acids, of which leucine content was the largest (22.6%). The relative molecular weight of the protein encoded by GP5 gene is about 59 305.38 KDa, and the theoretical isoelectric point is 9.40. The results of subcellular localization showed that it was mainly located in the endoplasmic reticulum (44.4%). GP5 protein contains signal peptide and transmembrane helix structure. In the prediction of secondary structure, random crimp was dominant, which was 70.14%. The tertiary structure of GP5 protein is mainly formed by random crimp, winding and folding, which is consistent with the predicted structure of the secondary structure.

Key words: Sheep; GP5 gene; Bioinformatics

血小板糖蛋白V(Glycoprotein V, Platelet, GP5)是血小板膜表面含量最丰富的糖蛋白之一, 它是血小板 vWF 和凝血酶受体 GPIb-IX-V 复合物中的一个亚单位。GP5 可

以促进 GPIb-IX 在血小板表面的表达, 增强血小板与 vWF 的结合能力。用凝血酶处理血小板后会导致血小板膜表面 GP5 的缺乏, 而缺乏 GP5 的血小板可增加血小板对凝血

收稿日期: 2020-04-27

基金项目: 甘肃农业大学学生科研训练计划项目(202004029)。

作者简介: 宋雅萍(1999—), 女, 甘肃靖远人, 本科在读, 研究方向为动物科学。联系电话: (0)18393413447。Email: 2973835198@qq.com。

通信作者: 张小雪(1984—), 女, 湖北武汉人, 讲师, 主要从事动物遗传育种与繁殖研究工作。联系电话: (0931)7631225。Email: zhangxx@gsau.edu.cn。

酶的应答反应,从而促进血小板纤维蛋白原结合和血小板聚集。人 GP5 的分子量为 83.3 ku,由 469 个氨基酸组成,且存在一种由 16 个氨基酸组成的信号肽。成熟 GP5 是由 544 个氨基酸组成的跨膜糖蛋白,包括 1 个跨膜区,1 个胞浆区(16 个氨基酸)和 1 个胞外区(504 个氨基酸)。胞外区包含 15 个富含亮氨酸的重复序列和 1 个凝血酶水解识别序列和敏感位点^[1]。人 GP5 是凝血酶的表面糖蛋白和底物^[2]。目前,GP5 基因在人、猪、鼠等中都有研究,对绵羊 GP5 基因却鲜有报道。我们利用生物信息学相关软件和在线工具,对绵羊 GP5 基因及其编码蛋白的理化性质、结构和功能等进行分析,旨在为今后绵羊 GP5 基因的进一步研究提供信息学参考。

1 材料与方法

1.1 序列来源

基因序列来源于 NCBI 网站 GenBank 数据库,包括绵羊(XM_027965957.1)的 mRNA 序列及 8 个物种的蛋白质序列:绵羊(XP_027821758.1)、人(NP_004479.1)、牛(XP_024850825.1)、家鼠(NP_036927.1)、猪(XP_020925727.1)、兔子(XP_008265039.1)、鸡(XP_004943385.1)、马(XP_023479286.1)。括号内为 GenBank 登录号。

1.2 方法

绵羊 GP5 基因开放阅读框(Open reading frame, ORF)采用 NCBI 的 ORF Finder 程序(<https://www.ncbi.nlm.nih.gov/orffinder/>)分析,参照 Kozak 法则。绵羊 GP5 蛋白潜在信号肽剪切位点预测采用在线工具 SignalP 3.0(<http://www.cbs.dtu.dk/services/SignalP-3.0/>)。

各物种 GP5 编码产物序列同源性比对采用 DNAMAN 软件分析。绵羊 GP5 蛋白保守结构域分析采用 Smart 软件(<http://smart.embl-heidelberg.de/>)^[3]。采用 Bioedit 软件及 ExPASy 的 ProtParam(<http://www.expasy.org/tools/protparam.html>)在线工具对绵羊 GP5 蛋白理化性质进行分析^[4-5]。采用 ExPASy 的 ProtScale(<http://ca.expasy.org/tools/protscale.html>)在线工具对绵羊 GP5 基因编码蛋白的亲疏水性进行预测^[5]。绵羊 GP5 基因编码产物亚细胞定位采用在线工具 PSORT II(<https://psort.hgc.jp/form2.html>)预测^[4]。绵羊 GP5 蛋白跨膜螺旋结构预测采用 TMHMM 在线软件(<http://www.cbs.dtu.dk/services/TMHMM>)。绵羊 GP5 蛋白二级结构采用 Jpred4(<http://www.compbio.dundee.ac.uk/jpred4/index.html>)分析预测^[6],三级结构采用 Swiss-model(<https://www.swissmodel.expasy.org/>)在线工具预测^[3]。

2 结果与分析

2.1 绵羊 GP5 基因开放阅读框分析

开放阅读框(Open reading frame, ORF)是一段从启动子到终止子,中间无终止序列打断的能编码蛋白质的碱基序列。通过识别 ORF 可以快速识别基因系列中的编码区域^[7]。由绵羊 GP5 基因 ORF 分析结果(图 1)可知,该序列中共识别有 6 个开放阅读框,其中最大开放阅读框的长度为 1 620 bp(起始密码子和终止密码子分别位于 1 bp 处和 1 620 bp 处),推测其编码 539 个氨基酸残基。

2.2 绵羊 GP5 编码产物理化性质分析

蛋白基本理化性质主要包括相对分子质

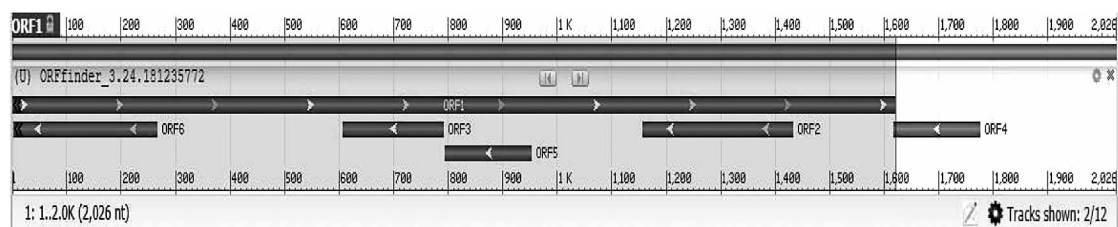


图 1 绵羊 GP5 基因序列开放阅读框分析

量、理论等电点(pI)值、氨基酸组成等^[3]。利用 ProtParam 在线工具和 Bioedit 软件对绵羊 GP5 基因编码产物的理化性质进行分析,结果表明,绵羊 GP5 基因编码 539 个氨基酸残基,其分子式为 $C_{2681}H_{4320}N_{756}O_{728}S_{16}$,相对分子质量为 59 305.38 KDa, pI 为 9.40。氨基酸组成如图 2 所示,其中亮氨酸(Leu)含量最多,所占比例为 22.6%;含量最少为酪氨酸(Tyr),所占比例为 0.2%。正电荷残基总数(Arg+Lys)和负电荷残基总数(Asp+Glu)分别为 52、40。基因编码产物半衰期为 30 h。不稳定指数为 42.49,大于 40.00,由此可知,GP5 基因编码的产物属于不稳定蛋白。

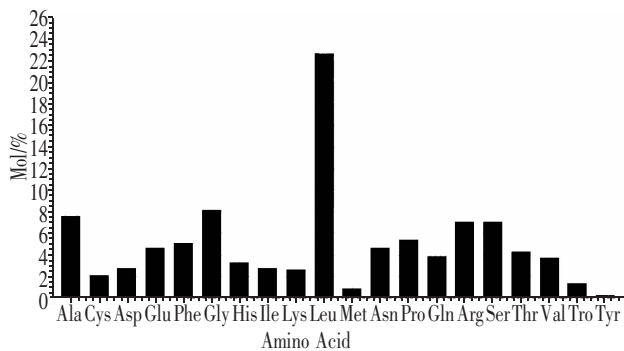


图 2 绵羊 GP5 基因编码蛋白的氨基酸组成

2.3 绵羊 GP5 蛋白亲疏水性分析

利用 ExPASy 的 ProtScale 在线工具对绵羊 GP5 蛋白的亲疏水性分析的结果(图3)表明,绵羊 GP5 基因编码蛋白疏水性最大值为 2.778(509~510位),疏水性最强;最小值为 -1.800(394位),亲水性最强。从亲水性/疏水性总体来看,亲水性氨基酸明显多于疏水性氨基酸,由此推断,GP5 蛋白属于亲水性蛋白。

2.4 绵羊 GP5 蛋白潜在信号肽剪切位点预测

通过 SignalP 3.0 软件对绵羊 GP5 蛋白潜在信号肽剪切位点进行预测,以此推断绵羊 GP5 蛋白是否存在信号肽序列及蛋白编码产物是否为分泌蛋白。由图 4 可知,在 16 和 17 位氨基酸之间存在剪切位点的可能性较大,可以推断出 GP5 蛋白存在信号肽序列,

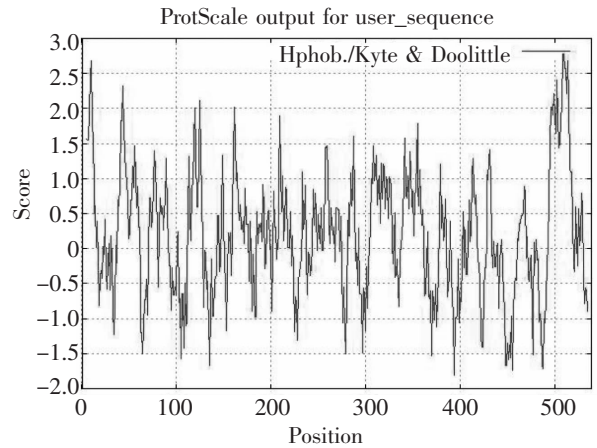


图 3 绵羊 GP5 基因编码蛋白疏水性/亲水性预测分析

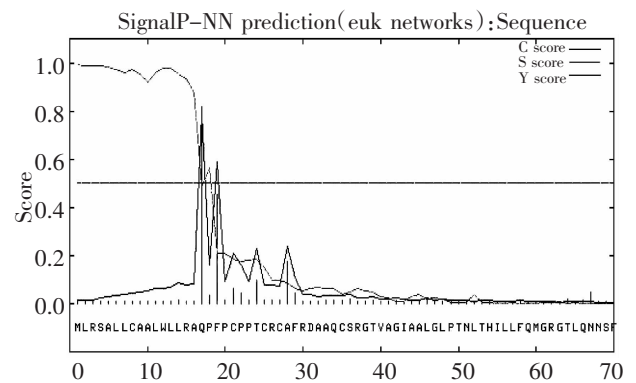


图 4 绵羊 GP5 基因蛋白潜在信号肽剪切位点分析

且属于分泌蛋白。绵羊 GP5 基因编码产物的 C 值为 0.817, Y 值为 0.790, S 值为 0.992。

2.5 绵羊 GP5 基因编码产物亚细胞定位分析

蛋白质参与正常生命活动,必须位于特定的亚细胞区域内(如细胞核、线粒体、细胞质、内质网、高尔基体等)^[8]。由在线工具 PSORT II 预测的绵羊 GP5 基因编码蛋白的亚细胞定位结果见表 1。绵羊 GP5 蛋白在内质网分布的概率为 44.4%,在质膜和高尔基体分布的概率分别为 22.2%、33.3%。可以看出,绵羊 GP5 蛋白存在于内质网的概

表 1 GP5 基因编码产物的亚细胞定位预测结果

亚细胞定位	概率 /%
内质网	44.4
高尔基体	33.3
质膜	22.2

率最大,由此可推测出绵羊 GP5 基因编码蛋白主要在内质网中发挥生物学作用,其次在高尔基体和质膜中发挥生物学作用。

2.6 绵羊 GP5 蛋白跨膜螺旋结构预测

利用 TMHMM 在线软件对绵羊 GP5 蛋白跨膜螺旋结构进行分析预测的结果(图5)表明,在 494~516 位氨基酸之间存在 1 段跨膜结构,其中第 1~493 位氨基酸在细胞膜外,第 517~539 位氨基酸在细胞质内。

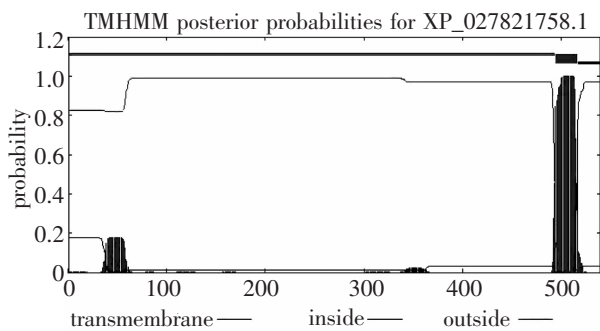


图5 绵羊 GP5 基因蛋白跨膜螺旋结构分析

2.7 绵羊及其他物种 GP5 基因编码产物序列同源性比对

通过 DNAMAN 软件对人、兔子、家鼠、牛、猪、绵羊、马、鸡等 8 个物种的 GP5 蛋白进行多序列比对(图6)及同源树分析(图7)的结果表明,该基因在各个物种中都有表达。绵羊与牛、猪等物种同源性较高,其中与牛的同源性高达 95%,与鸡同源性较低。

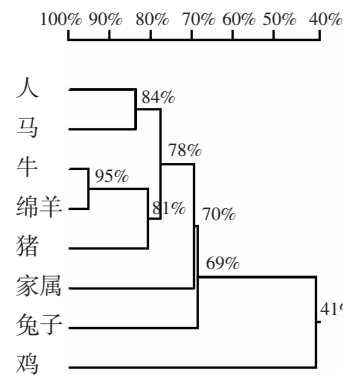


图6 8个物种(人、兔子、家鼠、牛、猪、绵羊、马、鸡)GP5 基因编码产物序列的同源性分析

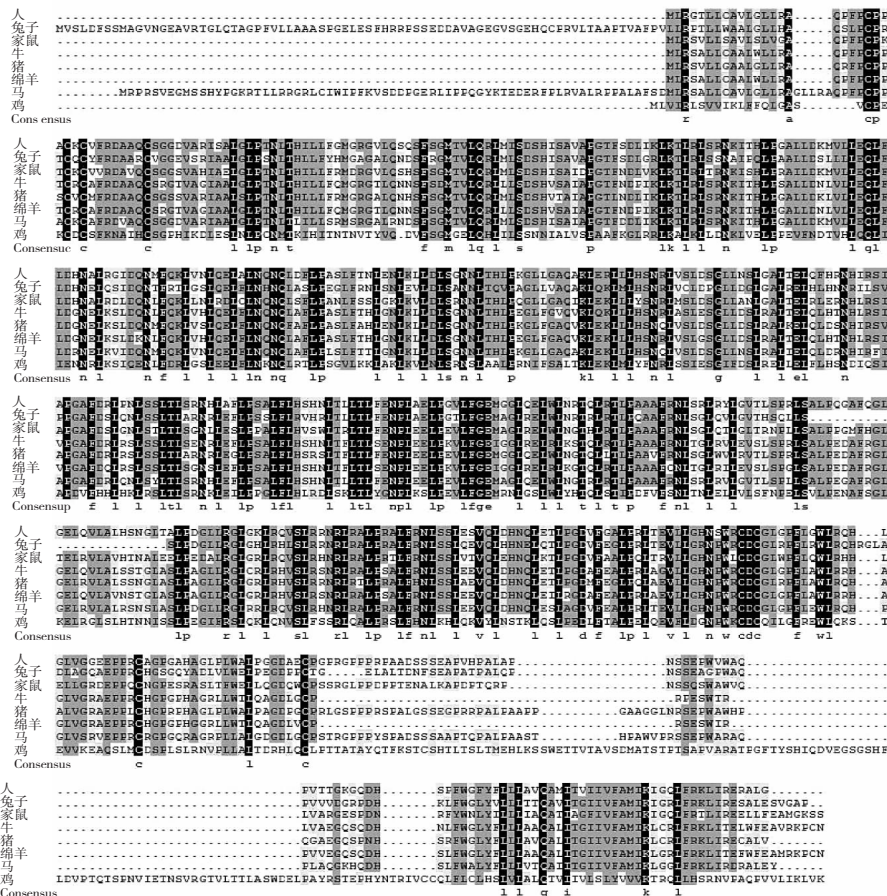


图7 8个物种(人、兔子、家鼠、牛、猪、绵羊、马、鸡)GP5 基因编码产物序列的同源树

2.8 绵羊 GP5 蛋白保守结构域分析

保守结构域是在生物进化或一个蛋白家族中具有不变或相同的结构域，而且保守结构域具有重要的功能，不可改变，是基因的核心。通过 Smart 软件对绵羊 GP5 基因编码蛋白的保守结构域进行分析，从图 8、表 2 可以看出，绵羊 GP5 蛋白第 494 ~ 516 位氨基酸残基存在跨膜区，第 2 ~ 16 位氨基酸残基为低复杂性区域，第 421 ~ 473 位氨基酸残基有 1 个 LLRCT 的结构域。

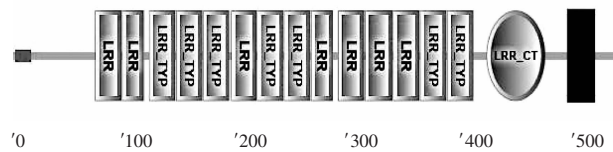


图 8 绵羊 GP5 蛋白保守结构域

表 2 绵羊 GP5 蛋白保守结构域分析数据

名称	起始位点	终止位点	E 值
LLRCT	421	473	3.24e-10
低复杂性区域	2	16	N/A
跨膜区	494	516	N/A

2.9 绵羊 GP5 蛋白二级结构的预测

蛋白质结构分析是蛋白质组学研究的重要组成部分，通过对新发现的蛋白质分子结构等的分析，可为预测新基因的结构和功能提供参照^[8]。通过在线工具 Jpred4 分析绵羊 GP5 蛋白二级结构的结果见图 9。绵羊 GP5 蛋白有 45 个 α 螺旋 (alphahelix, Hh)，所占比例为 8.35%；β 折叠 (extendedstrand, Ee) 79 个，占 14.66%；无规卷曲 (Random



图 9 绵羊 GP5 基因编码蛋白质二级结构预测

coil, -)415 个，占 70.14%。可以看出，绵羊 GP5 蛋白的二级结构以无规卷曲为主。

2.10 绵羊 GP5 蛋白三级结构的预测与分析

蛋白质的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重要意义^[8]。通过在线工具 Swiss-model 对绵羊 GP5 蛋白三级结构的预测和分析 (图 10) 可知，GP5 基因编码蛋白的三级结构主要是无规卷曲折叠缠绕形成的，与二级结构的预测结果一致。



图 10 绵羊 GP5 蛋白三级结构分析

3 结论

选取 NCBI 数据库中人、兔子、家鼠、牛、猪、绵羊、马、鸡等 8 个物种的 GP5 基因的蛋白质序列及绵羊 GP5 基因的 mRNA 序列，采用生物信息学分析方法对绵羊 GP5 基因的功能和编码蛋白结构等进行分析。结果表明，绵羊 GP5 基因包含一个最大的开放阅读框，其长度为 1 620 bp；编码 539 个氨基酸残基，其中，亮氨酸含量最多 (22.6%)，酪氨酸含量最少 (0.2%)。相对分子质量为 59 305.38 KDa，理论等电点为 9.40。绵羊 GP5 基因编码产物属不稳定蛋白，其富含亲水氨基酸，属于亲水性蛋白。绵羊 GP5 基因编码产物中存在信号肽序列，属于分泌蛋白，且存在 1 段跨膜区，属跨膜蛋白。绵羊 GP5 基因主要在内质网 (44.4%) 中发挥生物学作用，其次在高尔基体 (33.3%) 和质膜 (22.2%) 中发挥生物学作用，且编码的蛋白有 1 段跨膜结构。GP5 基因在牛、猪、家鼠、马等多个物种中都有表达，其中与牛的同源性高达 95%，与鸡的同源性较低。绵羊 GP5 基因编

8个菜用型马铃薯品种在榆中县的引种试验初报

刘世海^{1,2}, 丁述森³, 张海林², 张永福⁴, 王小燕⁵, 王燕革¹

(1. 榆中百稼汇农业种植农民专业合作社, 甘肃 榆中 730105; 2. 榆中县农业技术推广中心, 甘肃 榆中 730100; 3. 兰州市种业监督中心, 甘肃 兰州 730010; 4. 榆中县种子管理站, 甘肃 榆中 730100; 5. 兰州邦夫达农业科技有限公司, 甘肃 榆中 730100)

摘要: 在全黑膜垄作栽培条件下, 对引进的8个菜用型马铃薯品种进行比较试验。结果表明: 参试马铃薯各品种折合产量以华颂11号最高, 为81 000.0 kg/hm², 较对照品种陇薯7号增产60.44%; 其次是青薯9号, 折合产量为59 555.6 kg/hm², 较对照陇薯7号增产17.86%; 华颂7号居第3位, 折合产量为51 949.1 kg/hm², 较对照陇薯7号增产2.90%; 其余品种均较对照减产, 减幅为5.36%~37.69%。各参试马铃薯品种的薯形、皮色、肉色、芽眼深浅等商品性状及经济性状均适合商品市场。中早熟马铃薯品种华颂34号、中薯5号、中薯3号、华颂7号、华颂11号均适合在兰州地区川水地种植, 可弥补6—9月商品马铃薯的市场短缺。中早熟黄皮黄肉品种华颂34号、中薯5号、中薯3号、华颂7号宜作为扩繁种薯的品种供应南方市场; 晚熟品种陇薯10号、青薯9号、陇薯7号等适宜在兰州地区旱作区及山旱地种植。

关键词: 马铃薯; 菜用型; 品种; 引种; 榆中县

中图分类号: S532 **文献标志码:** A **文章编号:** 1001-1463(2020)10-0059-05

doi: 10.3969/j.issn.1001-1463.2020.10.013

兰州地处西北黄土高原腹地, 土层深厚, 气候冷凉, 昼夜温差大, 海拔1 500~

收稿日期: 2020-05-22; 修订日期: 2020-08-25

基金项目: 兰州市人才创新创业项目(2018-RC-64); 兰州市科技计划项目(2019-1-130)。

作者简介: 刘世海(1972—), 男, 甘肃榆中人, 高级农艺师, 主要从事马铃薯种薯繁育及栽培技术研究。联系电话: (0)13893144645。Email: liushihai2005@163.com。

通信作者: 丁述森(1968—), 男, 甘肃榆中人, 高级农艺师, 主要从事农作物品种管理工作。联系电话: (0)13893224476。Email: 609001944@qq.com。

码产物的二级结构主要为无规则卷曲, 三级结构主要由无规卷曲折叠缠绕而成。

参考文献:

- [1] 赵益明, 阮长耿. 活化血小板分子标志物: P-选择素、糖蛋白和糖蛋白 V 的研究进展[J]. 血栓与止血学, 2005(2): 85-87.
- [2] ROTH GERALD J, CHURCH TODD A, MC-MULLEN BRAD A, et al. Human platelet glycoprotein V: A surface leucine-rich glycoprotein related to adhesion[J]. Academic Press, 1990, 170(1): 153-161.
- [3] 李娜, 王维民, 张德印, 等. 绵羊 *ADRA1B* 基因生物信息学分析[J]. 甘肃农业科技, 2019(9): 42-48.
- [4] 张小雪, 潘香羽, 李发弟, 等. 绵羊 *GDF9* 基因生物信息学分析[J]. 甘肃农业科技, 2014(8): 18-21.
- [5] 胡斐, 吕慎金, 金一, 等. 绵羊 *FSHR* 基因生物信息学分析及其器官表达规律研究[J]. 黑龙江畜牧兽医, 2019(3): 24-27; 31; 177.
- [6] 臧胜芹, 陈晓勇, 杨凌, 等. 绵羊 *Cytb* 基因生物信息学分析[J]. 江苏农业科学, 2018, 46(18): 24-27; 37.
- [7] 隋景巍. 绵羊 *PTGER3* 基因生物信息学分析[J]. 乡村科技, 2017(11): 52-55.
- [8] 杨文, 沈文, 郭海英, 等. 新疆巴什拜羊 *BPI* 基因序列的生物信息学分析[J]. 畜牧与兽医, 2014, 46(12): 19-24.

(本文责编: 杨杰)