

# 绵羊 *RPS20* 基因的生物信息学分析

陈倩玲, 何亚鹏, 张引弟, 赵航, 王雪怡, 张小雪

(甘肃农业大学动物科学技术学院, 甘肃 兰州 730070)

**摘要:** 核糖体蛋白 20 基因 (ribosomal protein S20, *RPS20*) 是一类参与蛋白质生物合成及对细胞增殖、分裂、分化、凋亡具有调控作用的基因, 广泛存在于多种动物体内。为探究 *RPS20* 基因在绵羊体内的功能, 运用生物信息学数据库及其软件分析了该基因及其编码的产物。结果发现, 绵羊 *RPS20* 基因总共编码了 119 个氨基酸残基, 所编码的蛋白质分子式为  $C_{587}H_{995}N_{173}O_{171}S_5$ , 分子质量为 13.372 71 KDa, 等电点 pI 为 9.95。通过基本理化性质的分析可知, 绵羊 *RPS20* 蛋白是稳定性强、具亲水性的非分泌蛋白, 不含信号肽序列且无跨膜结构。亚细胞定位结果表明, 其编码产物主要存在于细胞质中 (65.2%)。绵羊与牛、马、黑猩猩、野生双峰驼等哺乳动物的 *RPS20* 蛋白相似度均为 100%。该蛋白的二级结构和三级结构主要以无规卷曲、 $\alpha$  螺旋和  $\beta$  折叠组成。

**关键词:** 绵羊; *RPS20* 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(2022)08-0039-05

[doi:10.3969/j.issn.1001-1463.2022.08.010](https://doi.org/10.3969/j.issn.1001-1463.2022.08.010)

## Bioinformatic Analysis of Ovine *RPS20* Gene

CHEN Qianling, HE Yapeng, ZHANG Yindi, ZHAO Hang, WANG Xueyi, ZHANG Xiaoxue

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

**Abstract:** Ribosomal protein S20 (*RPS20*) gene is a class of genes involved in protein biosynthesis and regulating cell proliferation, division, differentiation and apoptosis, which is widely present in a variety of animals. In order to explore the function of *RPS20* gene in sheep, bioinformatic analysis of *RPS20* gene and its encoded products was conducted by using bioinformatic databases and software. The results showed that *RPS20* gene encoded 119 amino acid residues, the protein molecular formula was  $C_{587}H_{995}N_{173}O_{171}S_5$ , the molecular weight was 13.37 271 KDa, isoelectric point pI was 9.95. The *RPS20* protein was a stable, hydrophilic, non-secretory protein and that contained no signal peptide sequence and no transmembrane structure. Subcellular localization mainly existed in the cytoplasm (65.2%). The similarity of *RPS20* proteins among sheep and mammals such as cattle, horses, chimpanzees, and wild bactrian camels was 100%. Its secondary and tertiary structure were mainly composed of random curl,  $\alpha$  helix and  $\beta$  fold.

**Key words:** Sheep; *RPS20* gene; Bioinformatics; Analysis

核糖体是催化蛋白质合成的细胞器, 由 1 个小的 40S 亚基和 1 个大的 60S 亚基组成。这些亚基由 4 种 RNA 和大约 80 种结构不同的蛋白质组成, 核糖体蛋白 (ribosomal protein, RP) 即所有核糖体蛋白质组成部分的统称。*RPS20* 基因是一类参与蛋白质生物合成及对细胞增殖、分裂、分化、凋亡具有调控作用的基因<sup>[1]</sup>, *RPS20* 基因编码的核糖体蛋白是 40S 亚基的组成部分, 该蛋白属于核糖体蛋白的 S10P 家族, 它位于细胞质中。

*RPS20* 基因作为核糖体蛋白编码基因的典型, 有多个加工过的假基因分散在基因组中<sup>[2]</sup>。有研究表明, *RPS20* 基因 RNA 干扰序列成功构建包装干扰慢病毒转染后能有效抑制小鼠结肠腺癌细胞 (CT26) 的生长<sup>[3-4]</sup>。*RPS20* 基因在猪繁殖与呼吸综合征病毒复制中有一定作用<sup>[5]</sup>。此外, 在筛选与猪脑心肌炎病毒 (EMCV) VP1 蛋白相互作用的宿主蛋白时发现, VP1 和 *RPS20* 蛋白之间存在特异性相互作用<sup>[6]</sup>。对大鼠 *RPS20* 基因的体外功能研

收稿日期: 2022-07-14

基金项目: 甘肃农业大学学生科研训练计划项目 (202204059)。

作者简介: 陈倩玲 (2001—), 女, 甘肃定西人, 本科在读, 研究方向为动物科学之畜牧兽医方向。联系电话: (0)15709321535。Email: 1179103558@qq.com。

通信作者: 张小雪 (1984—), 女, 湖北武汉人, 副教授, 主要从事动物遗传育种与繁殖研究及其教学工作。Email: zhangxx@gsau.edu.cn。

究结果初步表明,核糖体蛋白基因在脾虚证发生发展过程中可能具有重要作用<sup>[7]</sup>,对慢性浅表性胃炎脾气虚证等患者进行基因表达谱的分析结果表明,脾气虚证具有蛋白质合成相关基因下调的趋势,其中以核糖体蛋白基因下调比较显著,选择重复性表达下调的 *RPS20* 对其进行生物功能鉴定, *RPS20* 基因 RNA 干扰后使得大鼠小肠上皮细胞(IEC-6)的形态结构变化,细胞的增殖、分化能力, DNA 修复功能受到抑制,证明 *RPS20* 基因对大鼠小肠上皮细胞的消化吸收和粘膜损伤修复功能起重要作用<sup>[7-11]</sup>,与脾气虚证患者消化吸收功能下降、黏膜损伤修复能力下降等表现相符合。有关 *RPS20* 基因及其编码产物的研究主要是在人类和小鼠上,而在绵羊上的研究比较少。我们通过检索 NCBI 数据库中绵羊的 *RPS20* 基因序列,利用生物信息学的方法研究了绵羊 *RPS20* 基因及其编码产物的序列、基本理化性质、蛋白质结构和生物学功能等,以期为进一步探究该基因的结构和生物学功能提供参考。

## 1 材料与方法

### 1.1 序列来源

序列均来源于 NCBI 网站 GenBank 数据库,包括绵羊(XM\_004011685.4, XP\_004011734.1)、人(NM\_001023.4, NP\_001014.1)、兔子(NM\_001253734.1, NP\_001240663.1)、鸡(XM\_040664730.2, XP\_040520664.1)、牛(NM\_001034438.1, NP\_001029610.1)、猪(NM\_001129954.1, NP\_001123426.1)、马(XM\_005613124.3, XP\_005613181.1)、鹌鹑(XM\_015855656.1, XP\_015711142.1)、黑猩猩(XM\_016959501.1, XP\_016814990.1)、狮子(XM\_042924174.1, XP\_042780108.1)、野生双峰驼(XM\_032470290.1, XP\_032326181.1)11个物种的 mRNA 序列和氨基酸序列,括号内为 GenBank 的

登录号。

### 1.2 方法

绵羊 *RPS20* 基因开放阅读框(Open reading frame, ORF)分析采用 NCBI 的 ORF Finder 程序,基本理化性质分析采用 Bioedit 分析软件。绵羊 *RPS20* 蛋白亚细胞定位采用 PSORT II 软件,蛋白潜在信号肽剪切位点预测采用 SignalP 3.0 软件,蛋白跨膜螺旋区域预测采用 TMHMM 程序,蛋白保守结构域分析采用 Smart 软件,蛋白亲疏水性分析采用 Prot Scale 软件。蛋白二级结构预测采用 Jpred 软件,蛋白三级结构预测采用 Swiss-model 软件。多序列比对及同源性分析采用 DNAMAN 软件。

## 2 结果与分析

### 2.1 绵羊 *RPS20* 基因开放阅读框分析

开放阅读框(Open Reading Frame, ORF)是 mRNA 上的一段碱基序列,起始于起始密码子,终止于终止密码子,1个 ORF 对应1个蛋白质<sup>[12]</sup>,通过 ORF 分析可以证明新的 DNA 序列是否能够成为特定的蛋白质编码基因。根据图1可知,该序列最大的开放阅读框长度为360 bp(起始密码子位于121 bp处,终止密码子位于480 bp处),编码了119个氨基酸残基。

### 2.2 绵羊 *RPS20* 蛋白理化性质分析

蛋白质的基本性质包括相对分子量、等电点 pI、氨基酸组成、基因编码产物的不稳定指数和半衰期等<sup>[13]</sup>。利用 Prot Param 在线工具和 Bioedit 软件对绵羊 *RPS20* 基因编码产物的理化性质进行分析,表明其编码产物的分子式为  $C_{587}H_{995}N_{173}O_{171}S_5$ ,共含有1931个原子。该编码产物有119个氨基酸残基数,分子质量为13.37271 KDa,理论等电点 pI 为9.95,可知绵羊 *RPS20* 蛋白呈碱性。由图2可以看出绵羊 *RPS20* 基因编码

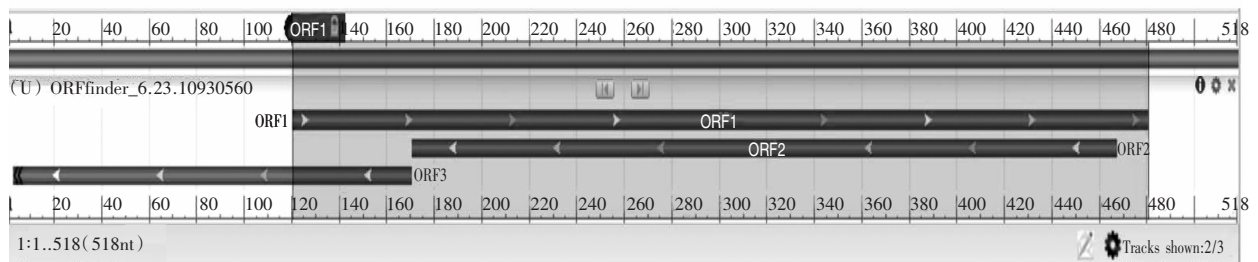


图1 绵羊 *RPS20* 基因序列的 ORF 分析

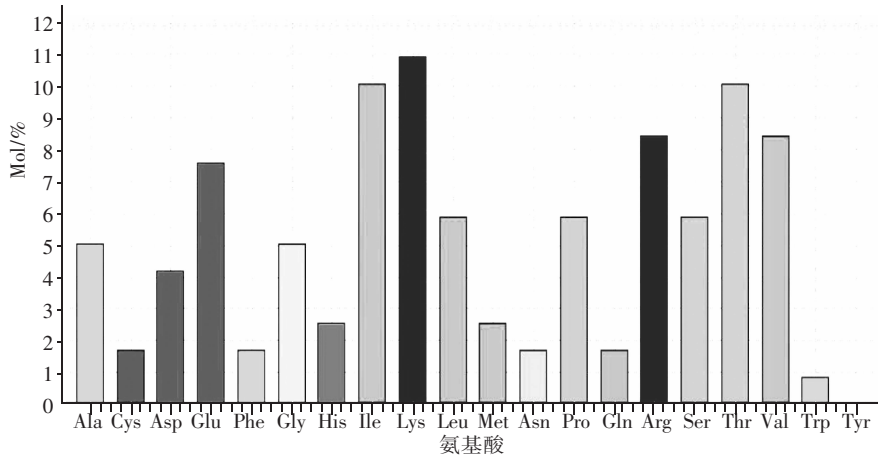


图2 绵羊 RPS20 基因编码的蛋白氨基酸组成



图3 11个物种的 RPS20 基因编码产物序列的同源性分析

的蛋白氨基酸组成。绵羊 RPS20 基因编码的蛋白中所含氨基酸种类丰富，其中含量最多的是 Lys (赖氨酸)，所占比例为 10.92%；Ile(异亮氨酸)、Thr(苏氨酸)含量较高，均为 10.08%。绵羊 RPS20 基因编码产物中不含 Tyr(酪氨酸)，负电荷残基总数(Asp+Glu)为 14，正电荷残基总数(Arg+ Lys)为 23。所编码产物在哺乳动物体外的半衰期为 30 h、不稳定指数为 32.28。根据不稳定指数指标(>40 为不稳定；<40 为稳定)，由于 32.28<40.00 可以确定该蛋白属于稳定蛋白。

### 2.3 绵羊 RPS20 蛋白亚细胞定位

从绵羊 RPS20 基因蛋白亚细胞定位结果可知，绵羊 RPS20 蛋白分布在细胞质的可能性最大，为 65.2%；分布在线粒体的可能性次之，为 17.4%；分布在细胞核的可能性居第 3 位，为 13.0%；分布在过氧化物酶体的可能性最小，仅为 4.3%。由此可以推断，绵羊 RPS20 基因编码的产物主要是在细胞质中发挥生物学作用。

### 2.4 不同物种 RPS20 蛋白的同源性分析

采用 DNAMAN 软件将绵羊 RPS20 蛋白序列与其他一些已发表的动物如人、绵羊、兔子、鸡、牛、鹤鹑、黑猩猩、狮子、马、猪和野生双峰驼 11

种动物的氨基酸序列进行多序列比对的结果(图3、图4)表明，RPS20 氨基酸序列同源性均为 100%。由于 RPS20 基因的氨基酸序列较短，绵羊与人、牛、黑猩猩、野生双峰驼等哺乳动物的相似性较高。

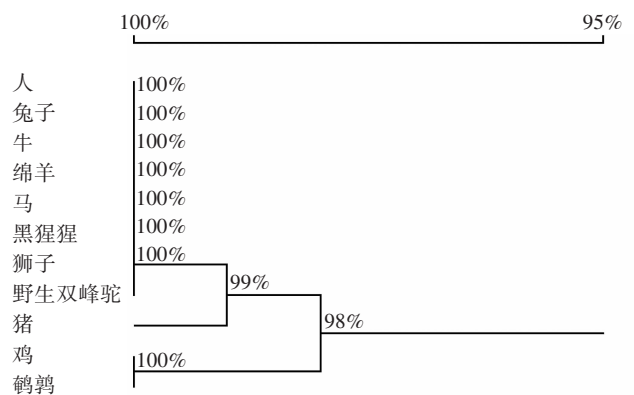


图4 11个物种的 RPS20 基因编码产物序列的同源树

### 2.5 绵羊 RPS20 蛋白潜在信号肽剪切位点预测

信号肽序列本质上是蛋白质序列中起始于 N 端的 1 段特定氨基酸序列，在引导和转运跨膜蛋白和分泌蛋白方面发挥着重要作用。核糖体是蛋白质合成的场所，信号肽在此合成过程中发挥重要作用 [14-18]。从对绵羊 RPS20 蛋白潜在信号肽剪切位点的预测可知，此基因编码的产物是否在引

导和转运跨膜蛋白和分泌蛋白方面发挥一定作用。从图5可以看出，绵羊 *RPS20* 基因编码产物的 C 值为 0.084、Y 值为 0.064、S 值为 0.128。由此可知，*RPS20* 基因的编码产物无跨膜区，也不存在信号肽序列，即该蛋白不是分泌性蛋白，也不是跨膜蛋白，主要位于细胞质内。

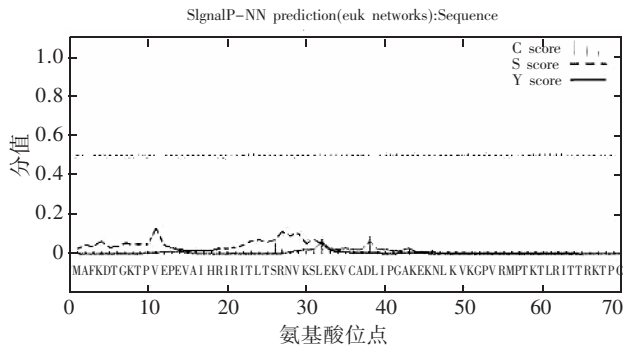


图5 绵羊 *RPS20* 基因蛋白潜在信号肽剪切位点分析

2.6 绵羊 *RPS20* 蛋白跨膜螺旋结构预测

采用 TMHMM 在线软件对于绵羊 *RPS20* 蛋白跨膜螺旋结构进行分析预测，结果表明绵羊 *RPS20* 基因编码的蛋白质是非跨膜蛋白，即无跨膜结构(图6)。

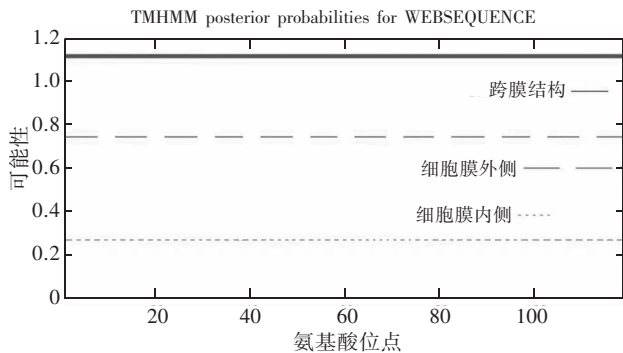


图6 绵羊 *RPS20* 基因蛋白跨膜螺旋结构分析

2.7 绵羊 *RPS20* 蛋白保守结构域分析

结构域是蛋白质中具有独立三级结构的部分，具有特定功能，因此也是蛋白质功能单元。保守结构域具有重要的功能，不能被改变，是基因的核心。由绵羊 *RPS20* 蛋白潜在信号肽剪切位点预测可知，绵羊 *RPS20* 基因编码的蛋白是非跨膜蛋白。进而通过 Smart 软件分析，如图7所示，绵羊 *RPS20* 蛋白家族包括原核生物中的小核糖体亚基

S10 和真核生物中的小核糖体亚基 S20，由于序列较短，没有跨膜结构，也没有复杂的跨膜区域。

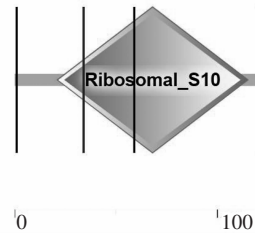


图7 绵羊 *RPS20* 蛋白保守结构域分析

2.8 绵羊 *RPS20* 蛋白亲疏水性分析

利用 Prot Scale 软件对绵羊 *RPS20* 蛋白亲疏水性分析，结果(图8)显示，该基因编码蛋白疏水性的最大值为 1.200，位于第 113 位；最小值为 -2.067，位于第 75 位。该基因编码的绵羊 *RPS20* 蛋白属于亲水蛋白。

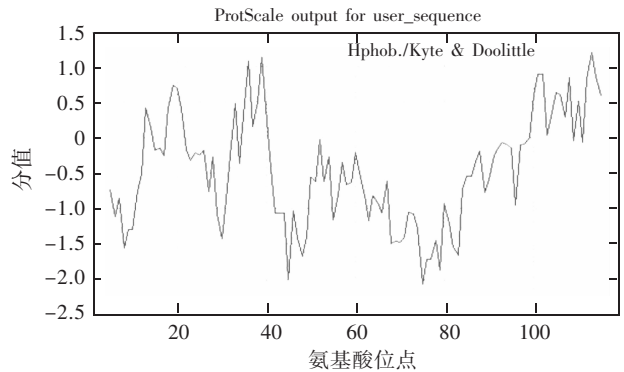


图8 绵羊 *RPS20* 基因编码蛋白质的疏水性/亲水性预测分析

2.9 绵羊 *RPS20* 蛋白二级结构预测

蛋白质的高级结构，特别是二、三级结构决定其生物学功能，其二级结构(secondary structure)指的是在它的多肽链中具有规则重复的构象<sup>[19]</sup>，二级结构是指蛋白质第1个水平的折叠，即部分蛋白质链折叠形成一些通用结构，这些通用结构在所有蛋白质中都能找到<sup>[13]</sup>。通过 Jpred 软件分析可知(图9)，绵羊 *RPS20* 蛋白二级结构主要由无规卷曲结构、α螺旋和β折叠三部分构成，三者所占比例分别为 44.54%、22.69%和 32.77%。

2.10 绵羊 *RPS20* 蛋白三级结构预测与分析

三级结构(tertiary structure)是指蛋白质的多肽



图中：- 为无规卷曲；H 为 α 螺旋；E 为 β 折叠

图9 绵羊 *RPS20* 基因编码蛋白质二级结构预测

链在二级结构甚至结构域基础上的进一步盘绕、折叠,依靠次级键的维系固定形成的特定空间结构<sup>[15,20]</sup>。通过 Swiss-model 在线工具得到的空间构象(图10)可以看出,绵羊 *RPS20* 基因编码蛋白的三级结构主要由无规卷曲、 $\alpha$  螺旋和  $\beta$  折叠盘曲折叠缠绕形成,与二级结构的预测结果相符。

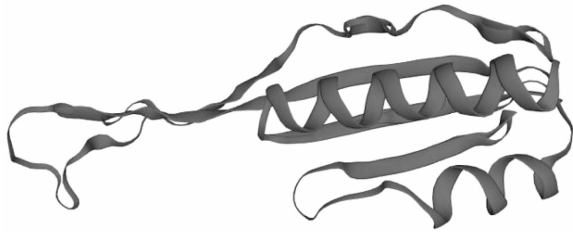


图10 绵羊 *RPS20* 基因编码蛋白的三级结构的分析结果

### 3 小结

绵羊 *RPS20* 基因的最长 ORF 长度是 360 bp, 编码的氨基酸残基数为 119。必需氨基酸种类丰富,赖氨酸所占比例最多 16.67%; 相对分子质量为 13.372 71 KDa, 理论等电点(pI)为 9.95。*RPS20* 基因编码的产物是稳定性强、富含亲水氨基酸的非分泌蛋白,也是非跨膜蛋白,无跨膜区且不存在信号肽序列。亚细胞定位的结果显示,绵羊 *RPS20* 蛋白在细胞质中发挥生物学作用的可能性最大,为 65.2%; 其次是在线粒体中,为 17.4%。绵羊 *RPS20* 蛋白在多种物种中氨基酸序列高度相似。绵羊 *RPS20* 基因编码产物的二级结构由无规卷曲、 $\alpha$  螺旋和  $\beta$  折叠三部分组成,三级结构亦由此三部分盘曲折叠缠绕形成。

### 参考文献:

- [1] FREED EF, BLEICHERT F, DUTCA LM, et al. When ribosomes go bad: diseases of ribosome biogenesis[J]. Mol Biosyst, 2010, 6(3): 481-493.
- [2] EBRIGHT RY, LEE S, WITTNER BS, et al. Dereglulation of ribosomal protein expression and translation promotes breast cancer metastasis[J]. Science. 2020. doi: 10.1126/science.aay0939.
- [3] 高小玲, 陈蔚文, 李茹柳, 等. 小鼠 *RPS20* 基因 miRNA 干扰载体的构建及鉴定[J]. 中药新药与临床药理, 2014, 25(5): 639-642.
- [4] 高小玲, 李茹柳, 郭文峰, 等. *RPS20* mRNA 质粒载体的 CT26 细胞转染条件优化研究[J]. 中药新药与临床药理, 2010, 21(3): 110-114.
- [5] 殷杰. 核糖体 L12、S20 蛋白在猪繁殖与呼吸综合征病毒复制中的作用研究[D]. 呼和浩特: 内蒙古民族大学, 2021.
- [6] 张元峰, 张昆丽, 薛娟, 等. 猪脑心肌炎病毒 VP1 蛋白与宿主 *RPS20* 蛋白相互作用研究[J]. 中国预防兽医学报, 2018, 40(8): 715-719.
- [7] 杨泽民, 陈蔚文, 王颖芳. 慢性浅表性胃炎脾虚与脾胃湿热证患者物质能量代谢基因差异表达研究[J]. 中国中西医结合杂志, 2012, 32(9): 1180-1187.
- [8] 陈蔚文, 王颖芳, 劳绍贤, 等. 脾气虚证患者基因差异表达研究 I[J]. 中国病理生理杂志, 2008, 24(1): 148-152.
- [9] 王静. 脾气虚证相关基因 *RPS20* 在 IEC-6 细胞的生物功能鉴定[D]. 广州: 广州中医药大学, 2009.
- [10] 王静, 李茹柳, 郭文峰, 等. 基于脾虚证相关基因功能鉴定的小肠上皮细胞 RNA 干扰转染条件研究[J]. 广州中医药大学学报, 2009, 26(6): 598-602; 616.
- [11] 陈炜璇. 基于肠细胞生长及健脾中药效应的 *RPS20* 基因功能鉴定[D]. 广州: 广州中医药大学, 2012.
- [12] SLIGHTOM J L, DURAND-TARDIF M, JOUANIN L, et al. Nucleotide sequence analysis of TL-DNA of *Agrobacterium rhizogenes* agropine type plasmid. Identification of open reading frames[J]. The Journal of biological chemistry, 1986, 261(1): 108-121.
- [13] 张小雪, 李发弟, 王维民. 绵羊 *ANXA10* 基因生物信息学分析[J]. 甘肃农业科技, 2016(6): 1-4.
- [14] 张维洵. 基于领域规则与混合深度模型的蛋白质信号肽预测研究[D]. 上海: 上海交通大学, 2020.
- [15] 刘佳, 王祯, 代友超, 等. 绵羊 *HTR4* 基因的生物信息学分析[J]. 甘肃农业科技, 2020(10): 35-40.
- [16] 靳皓宇, 武顺, 权洁, 等. 绵羊 *HMGAI* 基因的生物信息学分析[J]. 甘肃农业科技, 2022, 53(2): 78-83.
- [17] 张小雪, 赵利明, 刘佳, 等. 绵羊 *ELOVL5* 基因的生物信息学分析[J]. 甘肃农业科技, 2022, 53(4): 24-29.
- [18] 张小雪, 赵利明, 刘佳, 等. 绵羊 *KCNHI* 基因的生物信息学分析[J]. 甘肃农业科技, 2022, 53(3): 39-45.
- [19] 陈珂, 张君, 刘嘉斐, 等. 绿豆 *AP2/ERF* 转录因子家族的生物信息学鉴定与特征分析[J]. 分子植物育种, 2020, 18(20): 6605-6617.
- [20] 孔繁良. 基于二级结构的蛋白质三级结构预测[D]. 济南: 济南大学, 2016.