

绵羊 *HSPA1L* 基因的生物信息学分析

陈占玉, 黄永亮, 王维民, 李冲, 王国秀, 李明月

(甘肃农业大学动物科学技术学院, 甘肃 兰州 730070)

摘要: 为了解绵羊热休克蛋白1L (heat shock protein family A member 1 like, *HSPA1L*) 基因及其编码产物的基本结构和生物学特征, 给绵羊的抗热应激和免疫等方面的研究提供依据, 利用生物信息学数据库及软件对绵羊 *HSPA1L* 基因进行生物信息学分析。结果表明, 该基因最大长度的 ORF 有 1 926 bp, 编码 641 个氨基酸序列。其编码的蛋白质分子式为 $C_{3094}H_{4985}N_{859}O_{969}S_{20}$, 理论等电点为 5.89, 不稳定指数为 32.56。该蛋白无信号肽和跨膜结构, 为非分泌性蛋白, 主要在细胞质中发挥生物学作用。无规则卷曲是构成该蛋白二级结构和三级结构的主要方式。

关键词: 绵羊; *HSPA1L* 基因; 生物信息学分析

中图分类号: S565.9

文献标志码: A

文章编号: 1001-1463(2022)07-0083-06

[doi:10.3969/j.issn.1001-1463.2022.07.019](https://doi.org/10.3969/j.issn.1001-1463.2022.07.019)

Bioinformatic Analysis of *HSPA1L* Gene in Sheep

CHEN Zhanyu, HUANG Yongliang, WANG Weimin, LI Chong, WANG Guoxiu, LI Mingyue

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

Abstract: In order to understand the basic structure and biological characteristics of the heat shock protein family A member 1 like (*HSPA1L*) gene and its coding products in sheep, the bioinformatic database and software were used to conduct bioinformatic analysis of the *HSPA1L* gene in sheep. The results showed that the ORF of the maximum length of the gene was 1 926 bp, encoding 641 amino acid sequences. The protein molecular formula it encoded was $C_{3094}H_{4985}N_{859}O_{969}S_{20}$, the theoretical isoelectric point was 5.89, and the instability index was 32.56. This protein with no signal peptides and transmembrane structure, was a non-secretory protein, and plays a biological role mainly in the cytoplasm. Irregular curling was the main way of constituting the secondary and tertiary structure of the protein.

Key words: Sheep; *HSPA1L* gene; Bioinformatic analysis

热休克蛋白 (Heat Shock Proteins, HSPs) 是由机体受损、受寒、紫外线暴露、组织修复等各种应激诱导表达的蛋白^[1]。该蛋白由 Ritossa 于 1962 年在果蝇体内发现, 直至 1974 年经 Tissieres 证实并命名。HSPs 按大小可分为五大类, 分别为 HSP110(90–110KDa)、HSP90(83–90KDa)、HSP70(68–78KDa)、HSP60(40–60KDa) 以及 sHSP(12–34 KDa)^[2]。HSPs 广泛存在于从低等生物到高等动物

的机体中, 当生物有机体暴露于高温时, 就会由热激发合成 HSPs 来保护机体细胞免受热损伤^[3]。除高温环境以外, 其他应激源如缺氧、严寒、中毒等恶劣环境也能诱发机体细胞生成 HSPs, 从而帮助每个细胞维持正常的生理活动, 阻截影响细胞健康的蛋白质相互作用。也有研究发现, HSPs 对维持绵羊心脏的正常功能也有重要作用^[4]。

HSPA1L (heat shock protein family A member 1

收稿日期: 2022-05-05; 修订日期: 2022-06-03

基金项目: 甘肃农业大学学生科研训练计划项目(202104027)。

作者简介: 陈占玉(2000—), 女, 甘肃白银人, 本科在读, 研究方向为动物科学。联系电话: (0)18198057439。Email: Chenzy_gsau@163.com。

通信作者: 王国秀(1987—), 女, 甘肃天水人, 实验师, 主要从事绵羊生产研究与教学工作。联系电话: (0931)7631225。Email: wanggx@gsau.edu.cn。

[24] EFSA AHAW Panel (EFSA Panel on Animal Health and Welfare), 2016. Statement: Urgent advice on lumpy skin disease [M]. EFSA Journal 2016; 14(8): 4573, 27 pp. doi: 10.2903/j.efsa. 2016.4573

[25] JIHANE HAMD, ZAHRA BAMOUH, MOHAMMED JAZOULI, et al. Experimental evaluation of the cross-protection between Sheeppox and bovine Lumpy skin vaccines[J]. Scientific reports, 2020, 10: 8888.

like)是HSP70蛋白中的一员,HSP70是HSP家族中稳定性最高含量最多的一类蛋白^[5]。HSPA1L与HSPA1A、HSPA1B都属于HSP70家族中的编码蛋白质基因,HSPA1L基因编码蛋白有90%的氨基酸序列与HSP70家族一致。HSP70作为重要的应激蛋白,在保护细胞免于凋亡、氧化损伤和基因损伤中起着关键作用^[6]。由于HSPA1L基因是编码HSP70的3个基因之一,所以HSPA1L基因对绵羊抗应激和免疫功能也极为重要。但HSP70增强耐热力的作用机制模糊,如何通过技术措施诱导HSP70基因在局部合理表达以提高免疫效率仍需进一步研究^[7]。

目前针对HSPA1L基因的研究集中在人类医学方面,有研究表明HSPA1L编码蛋白与细胞损伤、肿瘤干细胞样特性和精子质量等有关^[8]。绵羊虽然有较强的耐受性和抗逆能力,但热适应性较差,温度过高会对其健康和生长造成影响,如诱导机体发生氧化应激、降低干物质的采食量、损害肠道组织等^[9],目前,国内外对绵羊HSPA1L基因的理论研究和生物信息学分析极少,缺乏实质性的参考信息。我们用生物信息学方法和程序软件对绵羊HSPA1L基因的开放阅读框、蛋白质亲/疏水性、基因基本理化性质进行分析,预测HSPA1L基因的初级和高级结构、跨膜区和亚细胞定位,旨在对绵羊的抗热应激和免疫等方面的研究提供理论基础和依据。

1 材料与方法

1.1 序列来源

从美国国立生物技术信息中心(NCBI)数据库中查询了绵羊HSPA1L基因的mRNA序列(登录号:XM_027958817.2),该序列来自爱荷华大学绵羊全基因组鸟枪测序(WGS)项目及绵羊基因组功能注释计划项目,绵羊品种为兰布莱绵羊(Rambouillet)。同时下载了除绵羊以外的其他11个物种的HSPA1L基因编码蛋白质序列(表1)。

1.2 研究方法

绵羊HSPA1L基因开放阅读框(Open reading frame, ORF)采用NCBI的ORF Finder程序分析(<https://www.ncbi.nlm.nih.gov/orffinder/>)。采用Bioedit软件预测HSPA1L基因编码产物的基本理化性质(<https://www.expasy.org/resources/protparam>)。蛋白

表1 12个物种的HSPA1L基因编码蛋白质序列

物种	登录号
绵羊	XP_027814618.1
水牛	NP_005518.3
人	NP_005518.3
普通牛	NP_001161367.1
猪	NP_001116600.1
狗	XP_038538972.1
猴	XP_028703546.1
马	XP_001492044.1
兔	XP_008260823.1
小鼠	NP_038586.2
山羊	NP_001301135.1
猫	XP_003993206.1

亲疏水性采用ProtScale工具进行分析(<http://ca.expasy.org/tools/protscale.html>)。蛋白潜在信号肽预测采用SignalP方法分析(<http://www.cbs.dtu.dk/services/SignalP-3.0/>)。蛋白质亚细胞定位预测采用PSORT方法预测(<https://psort.hgc.jp/form2.html>)。跨膜螺旋区域的预测采用TMHMM Server v. 2.0程序分析(<http://www.cbs.dtu.dk/services/TMHMM>)。采用DNAMAN软件对多序列比对和同源性分析进行预测。HSPA1L基因的保守结构域采用网站分析<http://smart.embl-heidelberg.de/>。蛋白二级结构的预测采用Jpred软件分析,三级结构的预测使用SWISS-MODEL在线工具进行分析(<https://www.swissmodel.expasy.org/>)。

2 结果与分析

2.1 绵羊HSPA1L基因开放阅读框分析

开放阅读框(Open reading frame, ORF)是一段可编码蛋白质中间无终止序列打断,从启动子到终止子的碱基序列,通过识别ORF可以快速识别出基因序列中的编码区域^[10]。从ORF分析结果可以看出(图1),序列中共识别的绵羊HSPA1L基因有20个ORF,其中最大长度的ORF有1926bp,推测其编码641个氨基酸残基。

2.2 绵羊HSPA1L基因蛋白基本理化性质分析

蛋白质基本性质主要包括氨基酸组成、相对分子质量和等电点^[11-13]。本文采用Bioedit软件预测绵羊HSPA1L基因编码蛋白质的理化特性并分析该基因编码产物的氨基酸组成,结果显示(表2),绵羊HSPA1L基因可编码641个氨基酸,其中Ala的比例最高,为9.0%;含量最少的是Trp,为0.3%。负电荷残基总数(Asp+Glu)为

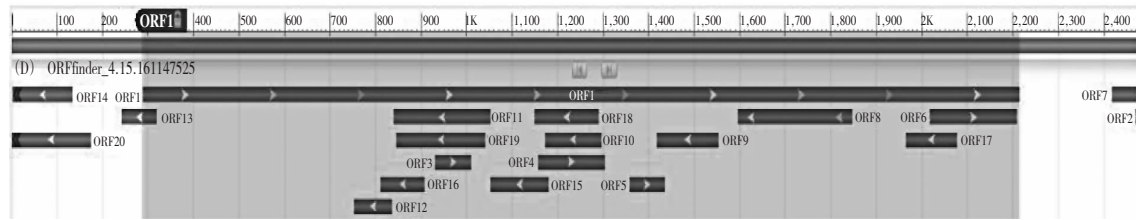


图1 绵羊 HSPA1L 基因序列的 ORF 分析

89, 正电荷残基总数(Arg+Lys)为 83, 分子式为 C₃₀₉₄H₄₉₈₅N₈₅₉O₉₆₉S₂₀, 理论等电点 pI 为 5.89, 不稳定指数为 32.56。

表 2 绵羊 HSPA1L 基因编码产物的氨基酸组成

氨基酸	比例 %	氨基酸	比例 %	氨基酸	比例 %
Ala	9.0	Gly	8.0	Pro	3.4
Arg	5.0	His	1.1	Ser	4.8
Asn	4.7	Ile	6.6	Thr	7.2
Asp	6.9	Leu	7.8	Trp	0.3
Cys	1.1	Lys	8.0	Tyr	2.7
Gln	3.9	Met	2.0	Val	7.0
Glu	7.0	Phe	3.6		

2.3 绵羊 HSPA1L 基因蛋白质亲/疏水性分析

蛋白质的亲疏水性取决于氨基酸序列, 一般高度易溶的蛋白质袒露在最外侧的都为亲水蛋白, 最内侧的都为疏水蛋白^[14]。采用 ProtScale 程序分析绵羊 HSPA1L 基因编码蛋白质的亲/疏水性分析, (图2)表明, 绵羊 HSPA1L 基因编码蛋白的疏水性最大值为 2.211(第 397 位), 最小值为 -3.056(第250 位), 故绵羊 HSPA1L 基因编码的蛋白是亲水蛋白。

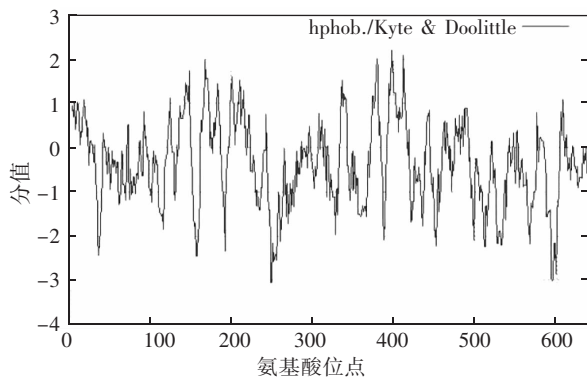


图 2 绵羊 HSPA1L 基因蛋白质亲/疏水性分析

2.4 编码蛋白质的信号肽预测

信号肽序列是分布在分泌蛋白基因的编码序列中的位于起始密码子之后的一段含有较多疏水氨基酸多肽的序列^[15]。绵羊 HSPA1L 基因编码产物的 C 值为 0.088, Y 值为 0.114, S 值为 0.515(图3)。

表明绵羊 HSPA1L 基因的编码产物不存在信号肽。

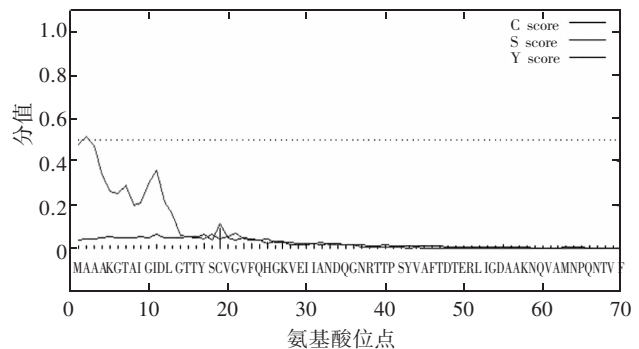


图 3 绵羊 HSPA1L 基因蛋白潜在信号肽预测结果

2.5 绵羊 HSPA1L 蛋白质亚细胞定位预测

采用 PSORT 方法预测绵羊 HSPA1L 蛋白质亚细胞定位。绵羊 HSPA1L 蛋白分布在细胞质中的可能性是 69.6%, 在细胞核内的可能性是 26.1%, 在线粒体中为 4.3%(表 3)。表明绵羊 HSPA1L 基因在细胞核中发挥的作用较小, 主要是在细胞质中发挥生物学作用。

表 3 绵羊 HSPA1L 蛋白质亚细胞定位预测分析

亚细胞分布	概率 %
细胞核	26.1
细胞质	69.6
线粒体	4.3

2.6 绵羊 HSPA1L 基因跨膜区预测

应用 TMHMM Server v. 2.0 程序分析了绵羊 HSPA1L 基因的信号肽, 结果显示(图4), 绵羊 HSPA1L 基因编码产物没有跨膜区。

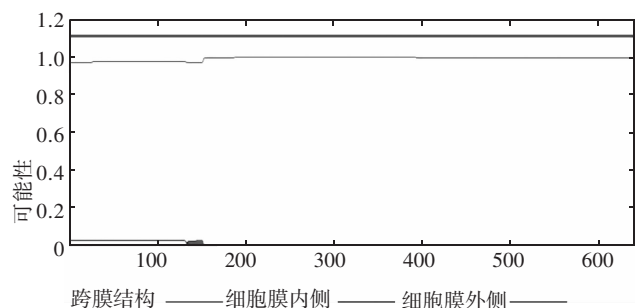


图 4 绵羊 HSPA1L 基因跨膜螺旋结构分析结果

2.7 多序列比对和同源性分析

HSPA1L 基因在很多物种中均有表达。本文采用 DNAMAN 软件对水牛、人、普通牛、绵羊、猪、狗、猴、马、兔子、小鼠、山羊和猫的同源性分析...

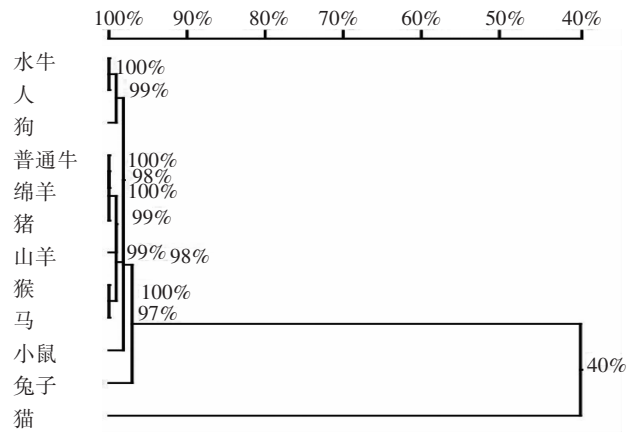


图 5 12 个物种的 HSPA1L 基因编码产物的同源树

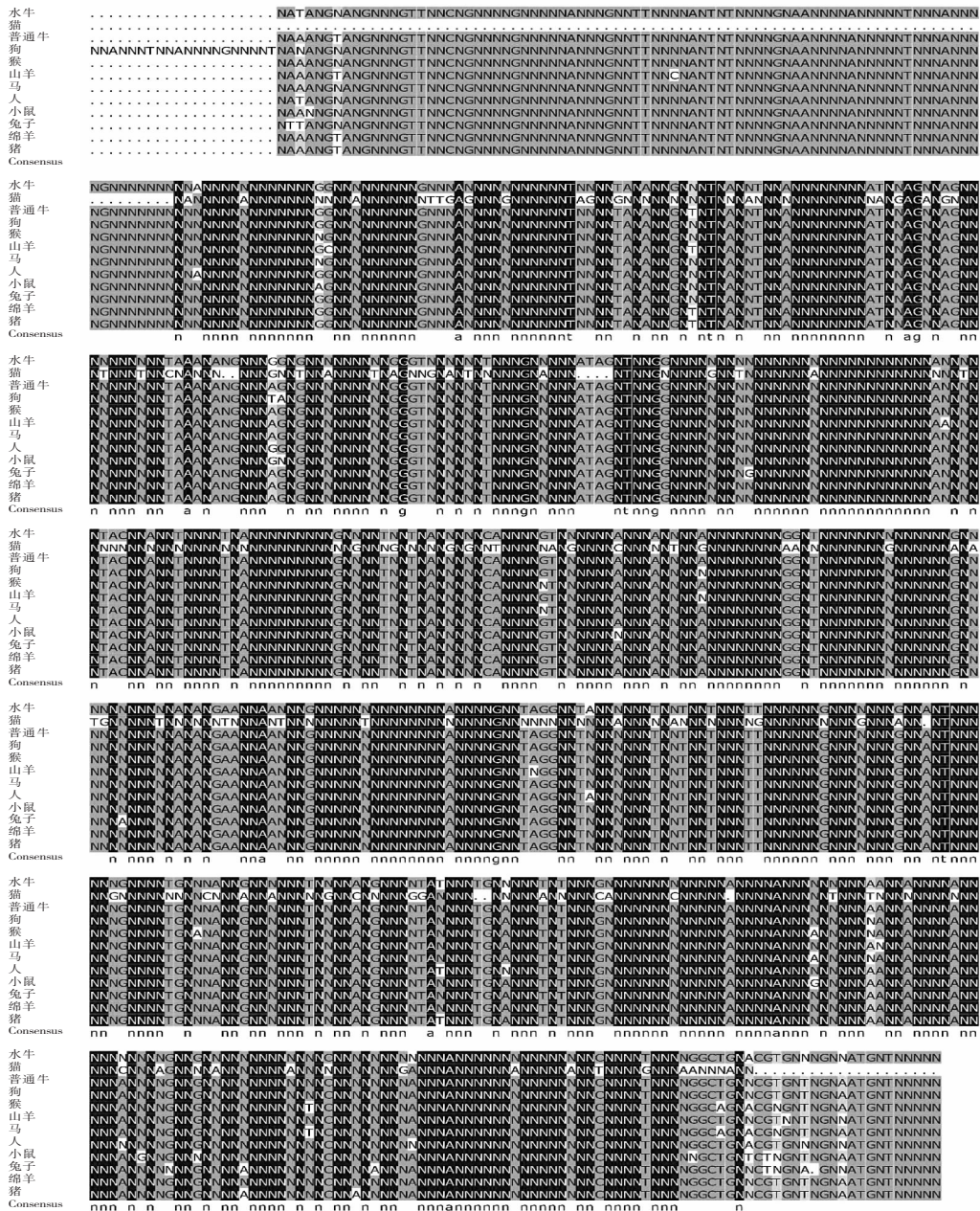


图 6 12 个物种的 HSPA1L 基因编码序列的同源性分析

2.8 绵羊 HSPA1L 蛋白保守结构域分析

采用 Smart 软件分析 HSPA1L 蛋白保守结构域，绵羊 HSPA1L 蛋白包含 2 个结构域，分别位于 370~380 和 393~405 位，均为低复杂性结构域(low complexity domain)(图7)。

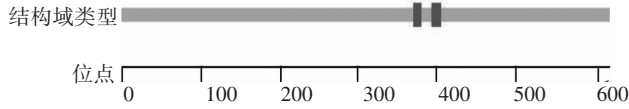
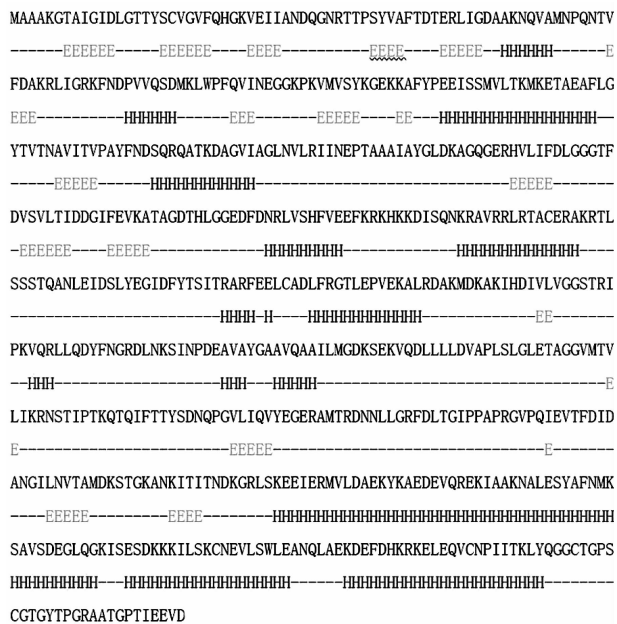


图 7 绵羊 HSPA1L 蛋白保守结构域分析

2.9 绵羊 HSPA1L 蛋白二级结构的预测分析

蛋白质的二级结构是 2 条核苷酸链反向平行盘绕所生成的双螺旋结构。采用 Jpred 软件对其进行预测，绵羊 HSPA1L 蛋白二级结构中无规则卷曲 (Cc)占 58.81%，为绵羊 HSPA1L 蛋白二级结构的主要组成部分；其次是 α -螺旋 (Hh)，占比 28.86%； β -折叠(Ee)占比最少，为 12.32%(图8)。



(—：无规则卷曲；E： β -折叠；H： α -螺旋)

图 8 绵羊 HSPA1L 基因编码蛋白质二级结构预测

2.10 绵羊 HSPA1L 蛋白三级结构的预测分析

蛋白质的三级结构是在二级结构的基础上进一步盘绕折叠所形成的特定空间结构。使用 SWISS-MODEL 在线工具对其进行预测分析结果显示(图9)，HSPA1L 基因编码蛋白的三级结构中主要由无规则卷曲构成，结构较为复杂。

3 结论与讨论

本研究分析结果表明，绵羊 HSPA1L 基因

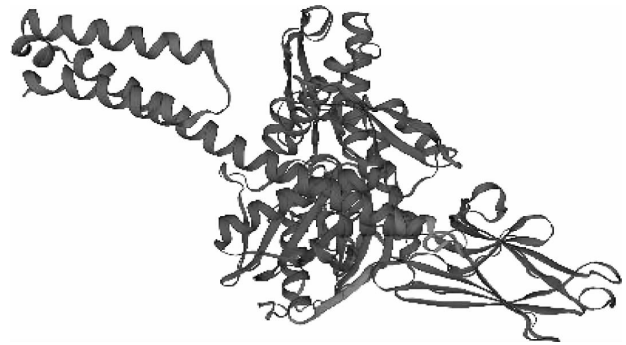


图 9 绵羊 HSPA1L 蛋白三级结构的预测分析

最长 ORF 有 1 926 个 bp，其编码产物稳定，编码的蛋白以亲水性区域为主，是一种易溶的弱酸性蛋白质，不存在信号肽，无跨膜结构，主要在细胞质中发挥生物学作用。与普通牛、和猪的同源性最高，该蛋白二级结构和三级结构中占主导地位的都是无规则卷曲，结构较为复杂。

有研究表明，热休克蛋白 1L(HSPA1L)是一种免疫性和保守性较强的蛋白^[16]。通过生物信息学分析可知，绵羊 HSPA1L 基因序列中最大长度的 ORF 有 1 926 个 bp，可编码 641 个氨基酸残基，呈弱酸性，较稳定，属于亲水蛋白且不存在信号肽。哺乳动物在热应激条件下，热休克细胞中的许多核蛋白会变的不可溶并成为基质的一部分，而在应激过程中合成的 HSP70 蛋白可向细胞核迁移，加速细胞核热休克的应激恢复^[17]。本研究表明，HSPA1L 基因编码蛋白稳定性较强，二级结构和三级结构中占主导地位的都是无规则卷曲，结构较复杂。较强的稳定性和在结构上较高的保守性，是其在热应激条件下保持结构和功能稳定的分子基础。本研究还证实，HSPA1L 编码蛋白属于亲水蛋白。有研究表明热应激条件下 HSP70 蛋白以亲水的方式结合到基质上在细胞核集中，以修复基质上变性的核蛋白^[18]。

HSPA1L 基因主要在细胞质中发挥作用，基因编码产物没有跨膜区。有研究表明，正常生理状态下 HSPs 主要存在于细胞质中^[19]，这与本研究一致。虽然细胞核对热更加敏感，但本文研究显示，绵羊 HSPA1L 蛋白分布于细胞核的可能性仅为 26.1%，而在细胞质中可能性高达 69.6%，原因在于 HSP70 中的 HSPA9 蛋白是一个主要的线粒体

伴侣蛋白, 可通过线粒体的双层膜将细胞核中的 HSPA1L 转移到细胞质中, 从而使细胞质中 HSPA1L 基因表达更加明显^[20]。由于热休克蛋白基因具有高度应激性和高表达性, 在应激源刺激下 HSPA1L 基因有可能在细胞质中快速表达合成。应激条件下, HSP70 蛋白会迅速从细胞质进入到细胞核和核仁, 而应激恢复期, 又迅速从其他组织返回细胞质^[17]。

在不同物种间绵羊与普通牛和猪的 HSPA1L 基因同源性最高为 100%, 这说明在进化过程中它们在该基因上具有较近的亲缘关系。但绵羊与山羊 HSPA1L 基因同源性相对较低。绵羊与山羊的物种亲缘关系较近, 但生活环境和适应性有较大差异, 山羊的热适应性更强。这有可能是环境差异导致的自然选择过程中 HSPA1L 基因在进化中出现了变异, 但具体的进化机制, 以及差异位点对该基因在抗热应激中发挥的作用和机理的影响仍需研究。

HSPA1L 基因作为编码 HSP70 蛋白的基因, 有可能对绵羊在热耐力形成过程中维持细胞内环境的稳定具有重要作用, 并对绵羊在各种环境应激条件下的适应性具有重要影响^[10]。以 HSPA1L 基因作为绵羊抗应激能力的分子标记选育开展分子标记辅助选育, 有可能加快绵羊育种进程, 改善绵羊的适应性, 减少生产中降低应激, 提高效益, 节约成本, 从而推动养羊业的发展。然而, HSPA1L 基因编码蛋白的生物学功能、表达调控以及 HSPA1L 基因多态性与绵羊健康与生长发育的关系仍有待进一步研究。

参考文献:

- [1] PARK CHANG-JIN, SEO YOUNG-SU. Heat shock proteins: A review of the molecular chaperones for plant immunity[J]. The plant pathology journal, 2015, 31(4): 323-333.
- [2] 张建鑫, 陈辉, 张竞乾. 动物热应激蛋白研究进展[J]. 饲料研究, 2008(3): 26-27; 30.
- [3] 李俄然, 严达伟, 赵宝, 等. 哺乳动物热休克蛋白研究进展[J]. 畜牧与兽医, 2017(11): 141-144.
- [4] MAGDALENA WISNIEWSKA, TOBIAS KARLBERG, LARI LEHTIÖ, et al. Crystal structures of the ATPase domains of four human Hsp70 isoforms: HSPA1L/Hsp70-hom, HSPA2/Hsp70-2, HSPA6/Hsp70B', and HSPA5/BiP/GRP78.[J]. PLoS ONE, 2010, 5(1): e8625.
- [5] 隋景巍. 绵羊 PTGER3 基因生物信息学分析[J]. 乡村科技, 2017(11): 52-55.
- [6] 唐妹, 徐蛟, 狄良娇, 等. 热休克蛋白的功能及其在畜牧业中的应用[J]. 南京农业大学学报, 2018(4): 605-615.
- [7] 赵志远, 徐刚. 热应激蛋白 70 与热应激反应[J]. 第一军医大学分校学报, 2003(2): 151-152.
- [8] TAKAHASHI S, ANDREOLETTI G, CHEN R, et al. De novo and rare mutations in the HSPA1L heat shock gene associated with inflammatory bowel disease[J]. Genome Medicine. 2017, 9(1): 8.
- [9] 石璐璐, 徐元庆, 王哲奇, 等. 热应激诱发的氧化应激对羊的影响及其作用机制[J]. 动物营养学报, 2019(7): 3016-3022.
- [10] 彭乔峰, 陈朗, 马小梅, 等. 热休克蛋白 HSP70 家族功能研究进展[J]. 甘肃畜牧兽医, 2019(5): 18-20.
- [11] 张小雪, 李发弟, 王维民. 绵羊 STMN2 基因生物信息学分析[J]. 甘肃农业科技, 2016(7): 58-61.
- [12] 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.
- [13] 靳皓宇, 武顺, 权洁, 等. 绵羊 HMGA1 基因的生物信息学分析[J]. 甘肃农业科技, 2022, 53(2): 78-83.
- [14] 米小梅, 李子元, 李耀东. 牛 TLR4 基因生物信息学分析[J]. 中国奶牛, 2021(4): 1-5.
- [15] 靳泽希, 冯芬, 邓晓银, 等. 绵羊 NRCAM 基因的生物信息学分析[J]. 甘肃农业科技, 2020(12): 19-24.
- [16] 陈曦. 热休克蛋白及热休克蛋白 70 家族进化保守性的系统分析[D]. 大连: 辽宁师范大学, 2011.
- [17] 陈洪博, 段滇宁, 鲍恩东. 热应激致 Hsc70 出入核与细胞凋亡的关系[J]. 畜牧兽医学报, 2017, 48(12): 2414-2420.
- [18] MOSSER D D, CARON A W, BOURGET L, et al. Role of the human heat shock protein hsp70 in protection against stress-induced apoptosis[J]. Mol Cell Biol. 1997, 17(9): 5317-5327.
- [19] 高晓霞, 冯义朝. 热休克蛋白 70 及其疫苗与胃癌的相关性研究[J]. 现代临床医学, 2014, 40(2): 93-95.
- [20] 王佳. 小鼠热应激蛋白家族在愈伤组织中表达的研究[D]. 长春: 吉林大学, 2012.