

文章编号: 1674-8085(2016)05-0029-07

大鼠胰岛素增强子结合蛋白 ISL1 的生物信息学分析

石姚黄, 王程媛, 梁晶莹, 李帅洪, *杨献光

(河南师范大学生命科学学院, 河南, 新乡 453000)

摘 要: 以 NCBI 中收录的不同物种 ISL1 蛋白的相关序列信息为基础, 进行生物信息学分析和功能预测。其中大鼠的 *ISL1* 基因长 12307 bp, 编码 349 个氨基酸。所编码蛋白质不含信号肽, 可能是存在于内质网的亲水性蛋白。结构预测分析发现 ISL1 蛋白二级结构存在较多无规则卷曲和双锌指的空间结构, 证实其发挥转录因子功能的结构基础。同源性比对分析显示, 大鼠 ISL1 基因编码蛋白与人、奶牛、小鼠、金仓鼠、灰狼的 ISL1 蛋白高度同源, 达到 100%; 与原鸡、斑马鱼、非洲爪蟾同源性也较高, 分别达到 99%、98% 及 97%, 这说明了 ISL1 蛋白序列的稳定性和保守性。通过对大鼠 ISL1 蛋白的生物信息学分析和功能预测, 为进一步探讨其分子调控机制提供重要的参考依据。

关键词: 大鼠; ISL1; 生物信息学分析

中图分类号: Q781/Q811.4

文献标识码: A

DOI:10.3969/j.issn.1674-8085.2016.05.007

CHARACTERIZATION OF ISL1 PROTEIN FROM RATTUS NORVEGICUS USING BIOINFORMATICS TOOLS

SHI Yao-huang, WANG Cheng-yuan, LIANG Jing-ying, LI Shuai-hong, *YANG Xian-guang

(College of Life Science, Henan Normal University, Xinxiang, Henan 453000, China)

Abstract: On the basis of the sequence information about ISL1 gene of different animals from NCBI, a series of bioinformatics analysis and function predictions was done about ISL1. Results revealed that the full-length of rat ISL1 gene was 12307 bp, encoding 349 amino acids. It was predicted that the ISL1 protein didn't have signal peptide and was a hydrophilic protein which was found in the endoplasmic reticulum. Structural prediction analysis revealed that the secondary structure of ISL1 protein had more random coil. The spatial structure resembled the double zinc finger that proved it could play a function as transcription factor. Homology comparison and phylogenetic analysis showed that the amino acid encoded by ISL1 gene in rat has highest homologies of 100% with those in human being, cow, mice, golden hamster and grey wolf. It was also highly homologous with those in *Gullas gullas*, *Danio rerio*, *Xenopus laevis*, with homologies of 99%, 98% and 97% respectively. The results show the stability and conservation of ISL1 protein sequence, and it could provide the theoretical foundation and reference for the subsequent experiment.

Key words: *Rattus norvegicus*; ISL1; bioinformatics analysis

收稿日期: 2016-04-03; 修改日期: 2016-05-16

基金项目: 国家自然科学基金项目(U1404312); 河南师范大学博士启动基金项目(31201093); 青年骨干教师项目(81200317);

国家级大学生创新创业训练计划项目(201410476067)

作者简介: 石姚黄(1996-), 女, 河南洛阳人, 河南师范大学生命学院生物科学专业 2013 级本科生(E-mail: 2690754332@qq.com);

王程媛(1994-), 女, 河南洛阳人, 河南师范大学生命学院生物科学专业 2013 级本科生(E-mail: 532755908@qq.com);

梁晶莹(1995-), 女, 河南洛阳人, 河南师范大学生命学院生物科学专业 2013 级本科生(E-mail: 2622411088@qq.com);

李帅洪(1992-), 男, 河南商丘人, 硕士生, 主要从事分子生物学研究(E-mail: 19920217@163.com);

*杨献光(1980-), 男, 河北邯郸人, 副教授, 博士, 硕士生导师, 主要从事肝再生机理研究(Email: yangxg@htu.cn).

胰岛因子 1(Islet 1, ISL1),又称胰岛素增强子结合蛋白 1(insulin gene enhancer binding protein 1),最初由瑞典 Karlsson 等^[1]在大鼠胰岛素瘤细胞系 RIN 中发现。*ISL1* 基因存在于大鼠的 2 号染色体上,基因全长约 12307 bp,含有 6 个外显子,共编码 349 个氨基酸,编码蛋白质属于 LIM 的同源异形结构域转录因子家族。作为 LIM 家族最早发现的成员之一,ISL1 蛋白最初被发现作为一种转录因子结合在胰岛素基因增强子上发挥作用。

LIM 结构域是最早发现于 Lin-11, Isl-1, Mec-3 三种同形域(homeodomain)蛋白质中,并以它们的首字母命名的具有特异结构和独立功能的蛋白质区域^[2]。ISL1 属于 LIM-HD 因子,LIM 结构域介导的蛋白质-蛋白质相互作用是 LIM 蛋白调节很多生物学过程的重要方式,可看做是更高层次的蛋白质复合物的脚手架,由 LIM 结构域介导形成多蛋白复合物的方式适应生命现象的多种联系,而且对于一些独立调节通路的功能发挥非常重要^[3]。

国内外许多研究发现,ISL1 在很多器官或组织的细胞分化发育过程中扮演重要角色。例如,参与调控胰腺内分泌细胞的分化^[4]、神经发育^[5-6]、动物

肢体发育^[7],对动物性腺睾丸的发育也有一定的调控作用^[8]。在对大鼠肝再生基因芯片表达谱分析中发现,ISL1 出现时期性的表达异常上调现象^[9]。为此,本研究对大鼠 ISL1 蛋白作了系统的生物信息学分析及功能预测,为进一步探讨其在器官或组织的细胞分化发育过程中的分子调控机制提供重要的参考依据。

1 材料与方法

利用 NCBI 数据库检索大鼠 *ISL1* 基因相关信息,获得其相关信息如下(DNA NCBI Reference Sequence: NC_005101.4, Gene ID: 64444, NM_017339)。MAP viewer 定位 *ISL1* 基因位于大鼠 2 号染色体上 43383525~43408665 位置之间,含 6 个外显子,编码含 349 个氨基酸的蛋白质。

运用一系列生物信息学的方法,对 ISL1 蛋白的相关理化性质,物理结构,及同源性进化等方面进行了预测分析。相关同源性分析运用生物类软件 DNASTar 中的 Megalign 进行分析,其他相关分析数据库及在线预测工具如表 1 所示。

表 1 分析所使用的生物学软件

Table 1 The resources and tools we used

编号	网址	主要功能
1	http://www.ncbi.nlm.nih.gov/	NCBI 数据库查询序列信息
3	http://expasy.org/tools/protparam.html	计算氨基酸含量,预测分子量和等电点
4	http://expasy.org/tools/protscale.html	计算蛋白质疏水图谱
5	http://www.cbs.dtu.dk/services/SignalP/	蛋白质的信号肽分析
6	http://www.cbs.dtu.dk/services/TMHMM4.1/	跨膜区预测
7	http://expasy.org/tools/SOPMA.html	二级结构分析
8	http://psort.nibb.ac.jp	亚细胞定位
9	http://www.cbs.dtu.dk/services/NetPhos/	磷酸化分析
10	http://www.ncbi.nlm.nih.gov/Structure/MMDb/mmdb.shtml	三级结构预测

2 结果与分析

2.1 ISL1 蛋白的保守结构域预测

利用 NCBI 中 CDD (Conserved Domain Database) 在线预测工具对 ISL1 蛋白的保守结构域

进行预测分析,结果如图 1 所示。可知由于 ISL1 蛋白 LIM1_Isl(17-71), LIM2_Isl(79-133), homeodomain(183-240)三个保守结构域的一级结构基础的存在,为 ISL1 蛋白能够表现 LIM 家族特性提供了可能。

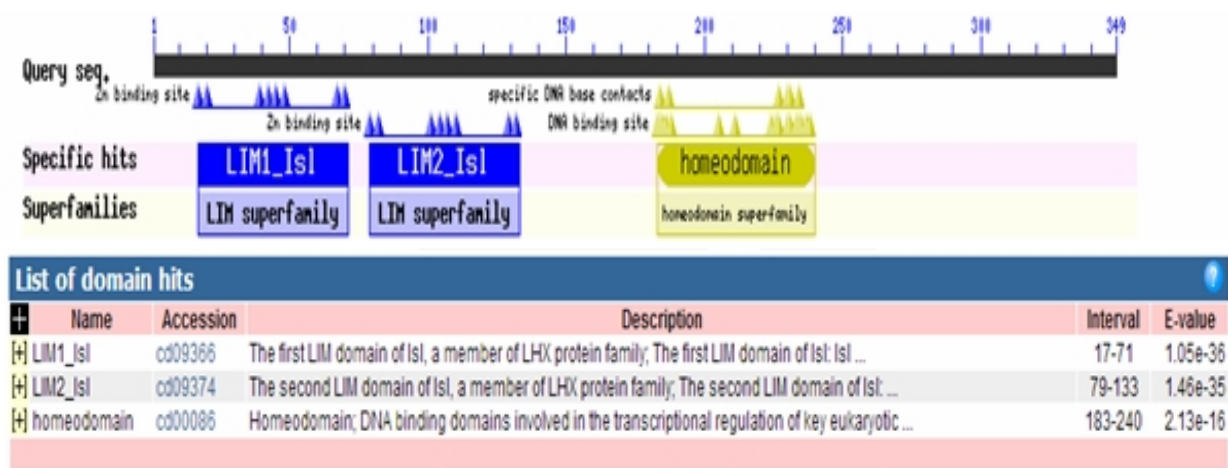


图 1 ISL1 蛋白氨基酸序列的开放阅读框预测
Fig. 1 Predicted ORF of the amino acid of ISL1 protein

2.2 ISL1 蛋白理化性质的预测

利用 EXPASY 数据库中的 protparam 和 compute pi/wm 在线分析工具，预测 ISL1 蛋白质一级结构的氨基酸组成、分子量、等电点、平均疏水性、不稳定系数及脂肪系数等理化性质^[10]。分析结果显示，大鼠中该 ISL1 基因所编码的蛋白质是分

子量为 39035.7 Da，等电点为 8.64，含 349 氨基酸数的不稳定蛋白。其负电荷残基与正电荷残基分别为 38 和 45，表示该蛋白在动物机体正常稳态条件下存在时带正电荷。蛋白质的不稳定系数为 59.72，大于 40，可知该蛋白较不稳定。

表 2 对 12 个物种的 ISL1 蛋白理化性质的预测

Table 2 Predicted physcial and chemical properties of twelve species ISL1 protein

物种名称	氨基酸残基数(个)	分子量(Da)	等电点(PI)	负电荷残基数(个)	正电荷残基数(个)	分子式	蛋白质不稳定系数
Rattus norvegicus	349	39035.7	8.64	38	45	C ₁₆₉₀ H ₂₇₀₁ N ₄₉₉ O ₅₀₈ S ₂₈	59.72
Mus musculus	349	39035.7	8.64	38	45	C ₁₆₉₀ H ₂₇₀₁ N ₄₉₉ O ₅₀₈ S ₂₈	59.72
Danio rerio	349	39180.8	8.64	38	45	C ₁₆₉₄ H ₂₇₀₈ N ₅₀₂ O ₅₁₁ S ₂₈	60.03
Xenopus laevis	354	39614.3	8.58	38	45	C ₁₇₁₆ H ₂₇₃₉ N ₅₀₅ O ₅₁₇ S ₂₈	61.08
Homo sapiens	349	39035.7	8.64	38	45	C ₁₆₉₀ H ₂₇₀₁ N ₄₉₉ O ₅₀₈ S ₂₈	59.72
Notophthalmus viridescens	349	38993.6	8.54	38	44	C ₁₆₈₉ H ₂₆₉₉ N ₄₉₇ O ₅₁₀ S ₂₇	60.34
Xenopus tropicalis	335	38669.1	9.28	29	49	C ₁₇₂₇ H ₂₇₂₈ N ₄₈₆ O ₄₇₁ S ₂₆	52.82
Schmidtea mediterranea	517	56861.3	8.11	55	59	C ₂₄₂₄ H ₃₈₉₅ N ₇₃₃ O ₇₇₁ S ₃₈	50.61
Bos Taurus	349	39035.7	8.64	38	45	C ₁₆₉₀ H ₂₇₀₁ N ₄₉₉ O ₅₀₈ S ₂₈	59.72
Gullasgullas	349	39033.7	8.64	38	45	C ₁₆₉₁ H ₂₇₀₃ N ₄₉₉ O ₅₀₇ S ₂₈	56.97
Mesocricetus auratus	349	39035.7	8.64	38	45	C ₁₆₉₀ H ₂₇₀₁ N ₄₉₉ O ₅₀₈ S ₂₈	59.72
Cains lupus	349	39035.7	8.64	38	45	C ₁₆₉₀ H ₂₇₀₁ N ₄₉₉ O ₅₀₈ S ₂₈	59.72

脂肪系数，一般用于球蛋白热稳定性增加的阳性因素分析，指的是一个蛋白质中脂肪侧链所占的相对值，检测得 ISL1 蛋白的脂肪系数值为 70.72。N 端残基在蛋白质稳定性判断方面起着十分重要的作用，ISL1 蛋白的 N 端氨基酸为甲硫氨酸，相关半衰期测定可知，ISL1 蛋白的半衰期为 30 h。

2.3 ISL1 蛋白亲水性/疏水性的鉴定

疏/亲水性是蛋白质的重要属性，其对蛋白质二

级结构的预测和分析具有重要的参考价值。蛋白质的疏水性通常依据蛋白的 GRAVY 值来预测^[11]（GRAVY 值通常分布在-2 至 2 之间，如果该值为负值，则判定为亲水蛋白）。利用基于 kyte&dodlittle 标度计算准则的 protscale 对 ISL1 蛋白进行疏水/亲水性分析，结果显示 ISL1 蛋白 GRAVY 值为 -0.510，为亲水性蛋白(图 2)。

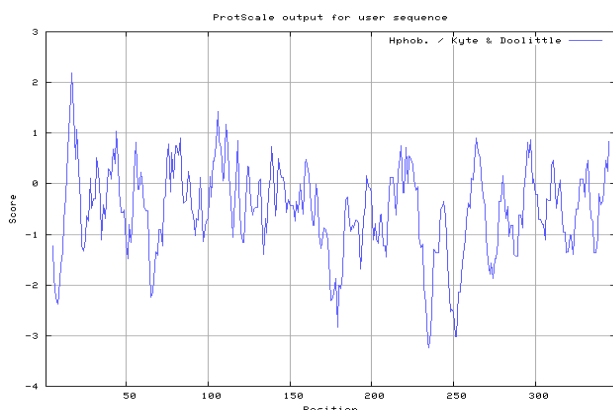


图2 ISL1 蛋白质氨基酸序列的亲水/疏水性分析
Fig. 2 Predicted hydrophobicity/hydrophilicity of the amino acid sequence of rat ISL1 protein

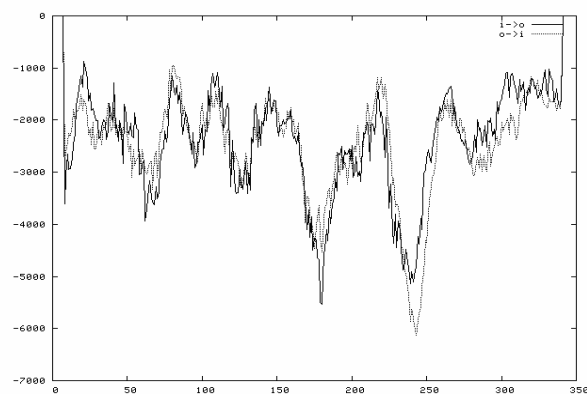


图3 ISL1 蛋白的跨膜区预测
Fig. 3 Predicted transmembrane domain of rat ISL1 protein

2.4 ISL1 蛋白跨膜区预测

利用 ExPASy 中的 TMHMM Server4.1 在线工具对 ISL1 蛋白的跨膜螺旋区进行分析预测, 表明其不存在跨膜序列, 即 ISL1 蛋白为可能在胞内或胞间发挥生理活性的非跨膜蛋白(图 3)。

2.5 ISL1 蛋白二级结构的预测

运用 ExPASy 软件中的 SOPAM 在线工具预测 ISL1 蛋白的二级结构^[12], 发现在 ISL1 蛋白可能的二级结构中, 无规则卷曲占比例最高, 占约 44.41%,

其次为 α -螺旋和延伸链, 分别占 24.93% 和 22.06%; β -折叠最少, 只占约 8.6%(表 3 和图 4)。

表 3 SOPMA 分析预测结果

Table 3 SOPMA analytical prediction of ISL1 protein

二级结构类型	氨基酸残基数目	所占百分比(%)
无规则卷曲(Cc)	155	44.41
α -螺旋(Hh)	87	24.93
延伸链(Ee)	77	22.06
β -转向(Tt)	30	8.6

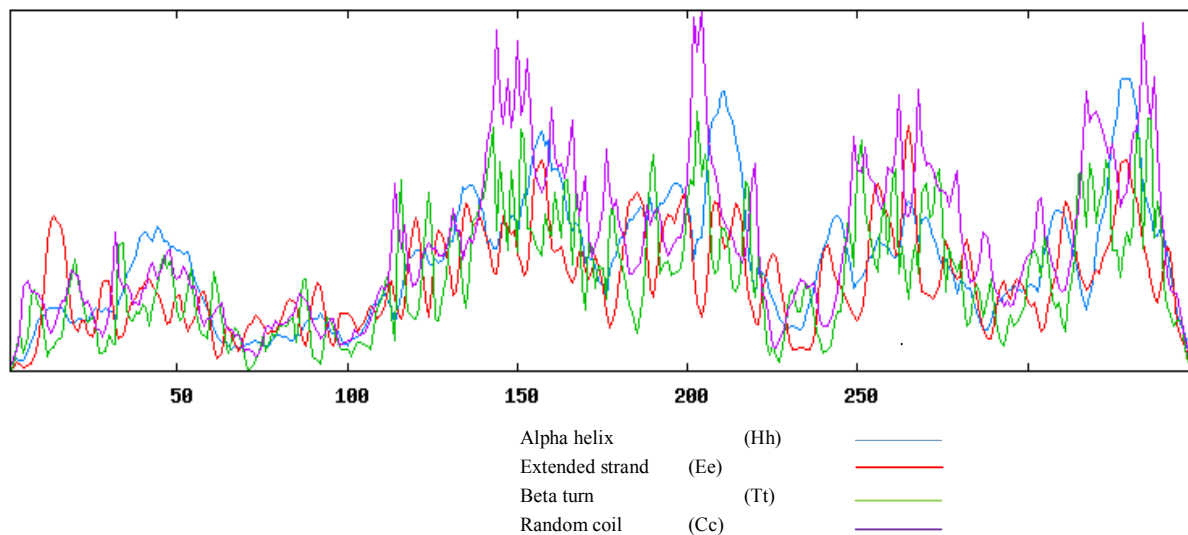


图4 ISL1 蛋白二级结构预测图
Fig. 4 Figure of ISL1 protein secondary structure prediction

2.6 ISL1 蛋白亚细胞定位

运用在线 PSORT 工具进行 ISL1 蛋白的亚细胞定位^[13], 预测结果如下: ISL1 蛋白有 0.550 的概率

存在于内质网膜, 分布于内质网管腔内的概率为 0.190, 分布于溶酶体管腔内部及分布于胞质内的概率均为 0.100(表 4)。结合信号肽序列分析结果, 证

明 ISL1 蛋白为非分泌蛋白，在细胞内发挥作用。

表 4 ISL1 蛋白亚细胞定位

Table 4 Subcellular location of ISL1 protein

亚细胞定位	概率
内质网(膜上)endoplasmic reticulum(membrane)	0.550
溶酶体(管腔)lysosome(lumen)	0.190
内质网(管腔)endoplasmic reticulum(lumen)	0.100
胞质(outside)	0.100

2.7 ISL1 蛋白质磷酸化分析

蛋白质磷酸化是调节和控制蛋白质活力和功能的最基本、最普遍，也是最重要的机制。蛋白质磷酸化主要发生在两种氨基酸上，一种是丝氨酸(包括苏氨酸)，另一种是酪氨酸。这两类酸磷酸化的酶不一样，功能也不一样。利用在线工具 NetPhos 2.0 Server 对 ISL1 蛋白进行磷酸化分析，结果显示，该蛋白具有丝氨酸(Ser)磷酸化位点 15 个，苏氨酸(Thr)磷酸化位点 3 个，酪氨酸(Tyr)磷酸化位点 2 个，ISL1 蛋白的多磷酸化位点可能与发挥转录因子调控功能密切相关(图 5)。

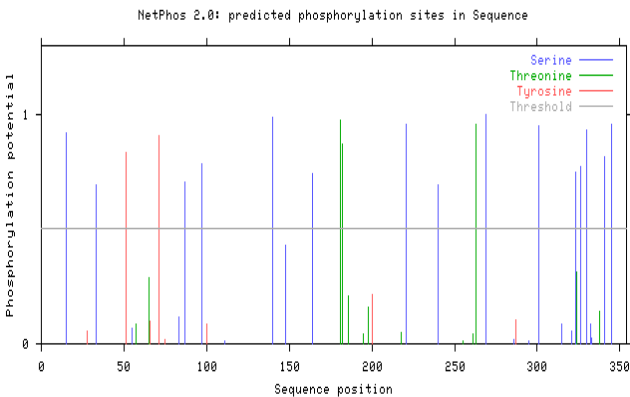


图 5 ISL1 蛋白的磷酸化分析

Fig. 5 Predicted phosphorylation sites of rat ISL1 protein

2.8 ISL1 蛋白信号肽预测分析

运用在线工具 SignalP 3.0 Server 进行 ISL1 蛋白的信号肽序列预测分析，结果如图 6 所示。用神经网络法(neural network, NN)分析(图 6A)，显示在 5 号甘氨酸处出现分值为 0.358 的最高信号肽分值，在第 24 位异亮氨酸处出现最高值为 0.171 的原始剪切位点值及在第 27 位谷氨酰胺出现分值为 0.035 的最高综合剪切位点。1-26 位的氨基酸各氨基酸的综合平均信号肽分值为 0.103。预测结果显示 ISL1 蛋白为非分泌蛋白。结合隐马可夫模型^[10](hidden markov models, 结果中简写为 HMM)的

分析结果(图 6B)，可断定 ISL1 基因所编码的 ISL1 蛋白不存在信号肽，为非分泌蛋白，进一步说明 ISL1 蛋白是在胞内发挥作用。

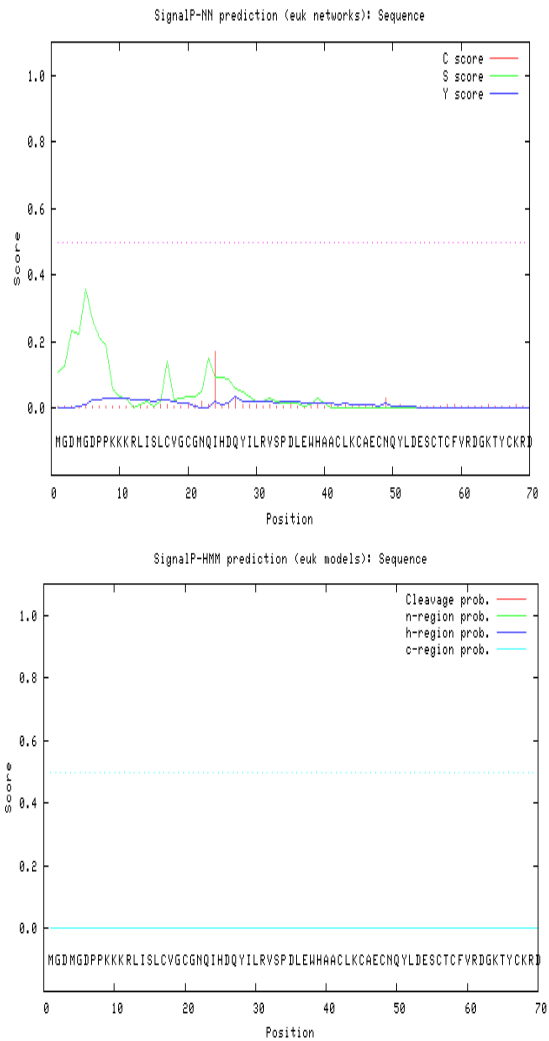


图 6 (A)NN 法预测蛋白质磷酸化和(B)HMM 法预测 ISL1 蛋白质磷酸化

Fig. 6 (A) SignalP-NN prediction of ISL1 protein; (B) SignalP-HMM prediction of ISL1 protein

2.9 ISL1 蛋白三级结构预测

运用 NCBI 数据库在线工具 Structure Summary MMDB 对 ISL1 蛋白进行蛋白质三维结构预测，预测结果如图所示(图 7)。有相关研究显示，ISL1 的独特性在于其所具有 3 部分高度保守的结构域，包括 2 个串联的 LIM 结构域和 1 个 LIM 同源域。而 LIM 结构域序列在不同的物种不同组织中高度保守，其基元序列为：[(C-X₂-C-X₁₇₋₁₉-H)-X₂-(C-X₂-C-X₇₋₁₁-(C)-X₈-C)]。由于结构域中含有保守的精氨酸、组氨酸和半胱氨酸残基，且空间构象类似于一个双

锌指结构。它的主要功能是介导与其他蛋白质之间的相互作用,主要有 POU 同源域蛋白、bHLH 蛋白、还有其它 LIM 蛋白^[2]。

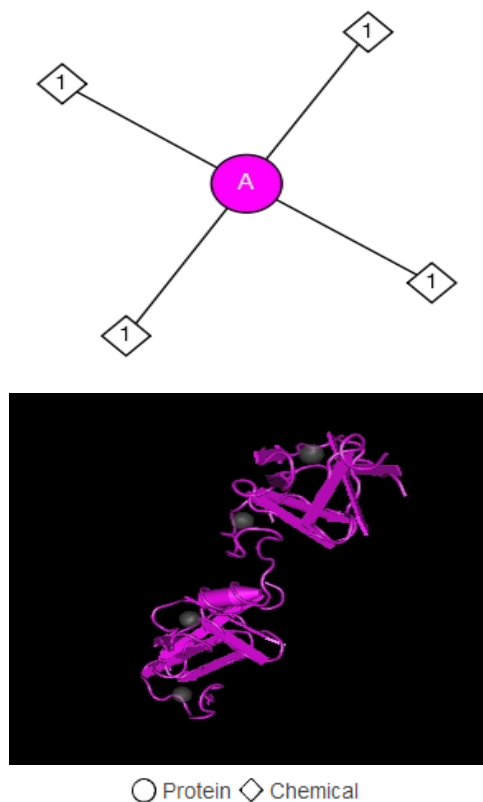


图 7 ISL1 蛋白三级结构预测

Fig. 7 Predicted the third structure of ISL1 protein

2.10 ISL1 蛋白的同源性分析

利用生物学分析软件 DNASTAR 中的 Megalign 软件,对 ISL1 蛋白氨基酸进行同源性分析,分析显示大鼠 ISL1 蛋白与人(*Homo sapiens*)、灰狼(*Canis lupus familiaris*)、奶牛(*Bos Taurus*)及金仓鼠(*Mesocricetus auratus*)氨基酸序列相似性高达 100%,与原鸡(*Gullasgullas*)、斑马鱼(*Daniorerio*)、非洲爪蟾(*Xenopus laevis*)、象鲨(*Callorhinchus milii*)等物种的氨基酸序列相似性达到了 90%以上,而与涡虫(*Schmidteamediterranea*)及绿红东美螈(*Notophthalmus viridescens*)的氨基酸序列相似性相对就比较低。利用上述物种的 ISL1 蛋白的氨基酸序列的相关信息构建了进化树图谱(图 8),更直观地反映了 ISL1 蛋白在不同物种间的亲缘关系远近,该结果与传统意义上的哺乳动物间相对亲缘关系较近,与两栖类动物、鱼类、鸟类等亲缘关系较远的传统分类结果基本一致。

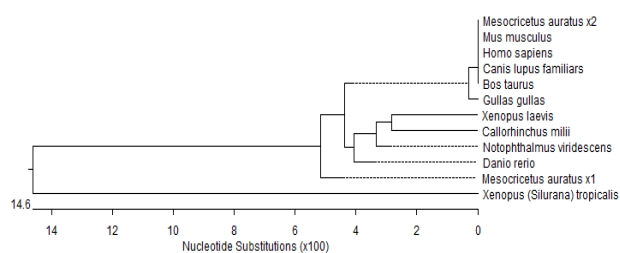


图 8 ISL1 蛋白的物种同源性分析比较以及进化树

Fig. 8 Homology comparison and phylogenetic analysis of ISL1 protein

3 讨论

运用生物信息学的相关软件,进行蛋白质的多种理化性质预测,对蛋白质功能和机理的阐述有重要意义^[14]。例如结合蛋白质等电点、跨膜预测、亚细胞定位以及对其信号肽序列存在与否的判断等结果,可推测出 ISL1 蛋白在胞内发挥作用,具生理活性的 ISL1 蛋白携带正电荷,且较多分布在内质网区域;结合该蛋白结构预测结果,发现一级结构中存在三个保守结构域,二级结构中无规则卷曲、 α -螺旋、 β -折叠及延伸链的含量,与三级结构中类似双锌指的空间结构形成存在相互吻合;由于 LIM 结构域和同源域的存在可以与其他分子特异性的结合,来介导蛋白质之间的相互作用;以及含有较多的磷酸化位点,亲水性强,半衰期较长等都与 ISL1 蛋白承担其生理作用有着相应的关系。

通过进一步与其他物种 ISL1 蛋白的同源性分析,揭示 ISL1 蛋白在不同物种间的进化程度及同源关系远近,发现 ISL1 蛋白在大鼠与人、奶牛等物种中具有高度同源性,推测其在相似的生理活动中发挥着类似的功能。随着对小鼠心脏发育研究的不断深入,发现 ISL1 蛋白在心肌细胞的分化过程中起重要作用,其作用途径之一就是调控组蛋白的乙酰化^[15]。而组蛋白的乙酰化则有利各种转录因子和协同转录因子与 DNA 特定位点的特异性结合,激活基因的转录。由于 ISL1 蛋白独特的三级空间结构,其能够通过 LIM 结构域的结合介导多种蛋白质的相互作用,如结构蛋白、激酶、转录因子等;与乙酰化酶相互结合时,其复合体有乙酰化酶活性^[16],且含有 DNA 结合结构域,能够调控基因转录进程,达到对细胞分化等生理过程的调控。由于大鼠与小

鼠的亲源关系较近,对蛋白质同源性分析也说明 ISL1 蛋白在两物种之间的高度相似性,说明其可能在大鼠类似生理过程发挥相同的作用。而对 ISL1 蛋白的相关理化性质及结构分析预测的生物信息学分析结果将为后期实验的开展奠定坚实的理论基础,大鼠 ISL1 相关方面功能研究正在进行中。

参考文献:

- [1] Karlsson O, Thor S, Norberg T, et al. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo-and a Cyc-His domain [J]. *Nature*, 1990, 344(6269): 879-882.
- [2] 刘耀波,汪家政,范明. LIM 同源盒转录因子在发育中的作用机制[J]. *生物化学与生物物理进展*, 2001, 28(3): 295-298.
- [3] 孙佳,鹿培源,贾弘禔. LIM 家族的研究进展[J]. *生理科学进展*, 2003, 34(2): 187-189.
- [4] 郑扬,周作红,侯玲玲,等. 胚胎干细胞分化为胰腺内分泌细胞的主要转录因子及其相关机制[J]. *中国组织工程研究与临床康复*, 2010, 14(19): 3573-3577.
- [5] 陈彪,范明,周长满,等. 同源框基因 islet-1 表达产物在大鼠脑中分布的免疫组织化学研究[J]. *解剖学报*, 1997, 28(2): 161-164.
- [6] Praff S L, Mendelsohn M, Stewart C I, et al. Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation [J]. *Cell*, 1996, 84(2): 309-320.
- [7] 庄涛,张清泉,孙云甫. Isl1 在动物后肢发育过程中的表达及作用[J]. *同济大学学报:医学版*, 2013, 34(5): 1-6.
- [8] 陈麟岫,王路,于卓,等. islet1 在小鼠睾丸发育过程中的表达研究[J]. *复旦大学学报:医学版*, 2014, 34(3): 263-267.
- [9] 徐存拴,章静波. 大鼠肝再生的功能基因组学研究[M]. 北京: 高等教育出版社, 2009.
- [10] 陈姗姗,郭晋隆,李国印,等. 甘蔗过氧化氢酶基因的电子克隆及生物学信息分析[J]. *生物信息学*, 2012, 10(1): 65-70.
- [11] 张云鹤,曹翠岩,邢慧清,等. 甘蔗 ATP 合酶基因的电子克隆及生物信息学分析[J]. *生物技术* 2013, 23(2): 34-39.
- [12] 闫秋良,翟洪军,马惠海,等. 绵羊 LDHAL6A 基因的电子克隆与生物信息学预测[J]. *广东农业科学*, 2013, 10: 154-158.
- [13] 王鹏良,韦翠娥,覃柳辉,等. 赤桉 EcWRKY50 转录因子基因的电子克隆及生物信息学分析[J]. *农业研究与应用*, 2014, 4: 9-15.
- [14] 熊伟,张晓娟,孙美涛,等. 人线粒体转录终止因子 2 基因启动子区的生物信息学分析[J]. *井冈山大学学报:自然科学版*, 2016, 37(1): 51-56.
- [15] 林建萍,田杰,刘官信,等. Islet1 在乙酰化调控网络中特意现行辅助 C3H10T1/2 细胞向心肌细胞分化[J]. *基础医学与临床*, 2012, 32(4): 363-368.
- [16] 周娜,朱静,田杰,等. 组蛋白乙酰化酶调控 MSCs 经 5-azaC 诱导后的细胞周期和增殖特性[J]. *基础医学与临床*, 2010, 30: 689-697.