

鸡 RB1 基因的结构与功能分析

杨莉莉^{1,2,3}, 李 辉^{1,2,3}, 张 慧^{1,2,3}

(1. 农业部 鸡遗传育种重点实验室, 哈尔滨 150030; 2. 黑龙江省普通高等学校动物遗传育种与繁殖重点实验室, 哈尔滨 150030; 3. 东北农业大学 动物科学技术学院, 哈尔滨 150030)

中图分类号: S831; Q78

文献标识码: A

文章编号: 1004-7034(2017)05-0016-06

关键词: 鸡; 视网膜母细胞瘤基因 1 (RB1 基因); 生物信息学; 结构与功能; 分析; 预测

摘 要: 为了研究鸡视网膜母细胞瘤基因 1 (RB1 基因) 的结构和功能, 试验采用生物信息学的方法, 选择 NCBI 数据库上 RB1 基因的氨基酸序列, 从物种间的蛋白理化性质、信号肽、蛋白磷酸化、结构域、Motif 和三维结构等方面对鸡 RB1 基因的结构和功能进行了分析和预测。结果表明: 鸡与哺乳动物之间序列的同源性很高。鸡 RB1 基因具有结构域和磷酸化等位点, 没有信号肽和跨膜区。成熟肽与模板 4elj. 1. A 有 73.16% 的氨基酸序列一致。说明 RB1 基因可以调控细胞的生长发育。

Analysis of the structure and function of the RB1 gene in chicken

YANG Lili^{1,2,3}, LI Hui^{1,2,3}, ZHANG Hui^{1,2,3}

(1. Key Laboratory of Chicken Genetics and Breeding, Ministry of Agriculture, Harbin 150030, China; 2. Heilongjiang Province Key Laboratory of Animal Genetics, Breeding and Reproduction, Harbin 150030, China; 3. College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China)

Keywords: chicken; retinoblastoma gene 1 (RB1 gene); bioinformatics; structure and function; analysis; forecast

Abstract: To explore the structure and function of retinoblastoma gene 1 (RB1 gene) of chicken, the prediction and analysis of bioinformatics of the RB1 gene were carried out according to amino acid sequence of RB1 gene in the NCBI database, involving in species of the physicochemical nature, signal peptide, protein phosphorylation, domain structure, motif and three dimensional structure etc. The results showed that the sequence homology between chicken and mammal is very high. Chicken RB1 gene has the structure and phosphorylation sites, no signal peptide and transmembrane domain. The mature peptide has 73.16% amino acid similarity with 4elj. 1. A. These data demonstrated that the RB1 gene will has a potential role in regulation of the growth and development of cells in chicken.

视网膜母细胞瘤基因 1 (RB1 基因) 是人类第一个分离克隆的抑癌基因^[1], 因为该基因与视网膜母细胞瘤的发生密切相关, 所以被命名为视网膜母细胞瘤基因。哺乳动物中 RB1 基因家族包括 3 个成员: RB1/p105、RBL1/p107、RB2/p130^[2-4]。

RB1 基因由 27 个外显子和 26 个内含子组成。鸡 RB1 基因定位在 1 号染色体上, DNA 序列全长为 78 217 bp, mRNA 全长为 4 464 bp, CDS 全长为 2 766 bp, 编码 921 个氨基酸^[5]。W. H. Lee 等^[6]对

这段序列的 RNA 进行测序并预测蛋白质序列, 发现此段序列的 mRNA 长约 4.6 kb, 蛋白质大约包含 816 个氨基酸残基, 而且发现此氨基酸序列有与核酸结合蛋白类似的结构域。Y. K. Fung 等^[7]利用 H3-8 探针在 13q14 区域上也分离获得了两段相似序列的 cDNA 克隆: RB1 和 RB2。R. Feinstein 等^[8]克隆了鸡 RB1 的 cDNA 序列, 发现鸡 RB1 基因编码 104 ku 的蛋白质, 比哺乳动物的稍小。

大量研究表明, RB1 可以抑制多种肿瘤的发生, 如视网膜母细胞瘤、乳房瘤、骨肉瘤、胰腺瘤等。RB1 的抑癌作用与其对细胞周期、细胞分化、细胞衰老、细胞凋亡和生长抑制等的调控是密切相关的。RB1 通过对细胞生长发育的调节来维持生物体生长发育的平衡, 对生物体的正常生长发育起着重要的调控作用^[5]。

本研究以 NCBI 数据库上 8 个物种的 RB1 mRNA 序列为基础, 对其编码的氨基酸序列进行比对, 并进行同源性和进化关系分析, 还应用生物信息学方法对其编码蛋白质进行理化性质、信号肽、蛋白磷酸化、结

收稿日期: 2016-08-30; 修回日期: 2017-03-28

基金项目: 国家自然科学基金项目 (31301960); 高等学校博士学科点专项科研基金新教师类资助课题 (20132325120016); 黑龙江省普通本科高等学校青年创新人才培养计划项目 (UNPY-SCT-2015007); 黑龙江省教育厅面上项目 (12541019); 东北农业大学“青年才俊”项目 (14QC15)

作者简介: 杨莉莉 (1991-), 女, 硕士研究生, 研究方向为动物遗传育种, 992413642@qq.com.

通信作者: 张 慧 (1982-), 女, 副教授, 博士, 研究方向为动物遗传育种, huizhang@neau.edu.cn.



构域、Motif 及三维结构等方面的预测,并分析其生物学功能,以期为鸡 RB1 功能研究奠定基础。

1 材料与方法

1.1 各物种 RB1 基因序列

数据资料来自美国国立信息技术中心(NCBI)数据库,从中获取 8 个物种的 RB1 基因的蛋白质序列(鸡 NM_205237、牛 NM_001045877、小鼠 NM_007554.2、人 NM_001202、猪 NM_001101031、非洲爪蟾 NM_001101793、斑马鱼 NM_131342、虹鳟 FJ591154)用于研究。

1.2 物种间 RB1 氨基酸同源性及进化关系分析

利用 DNAMAN 软件对这 8 个物种的 RB1 mRNA 序列进行比对,并进行同源性和进化关系分析。

1.3 RB1 结构与功能分析

利用 PSORT 等在线工具或软件分析 RB1 的结构与功能,对其理化性质、蛋白质功能、信号肽、结构域、亚细胞定位及二级、三级结构等进行预测与分析。主要软件相关信息见表 1。

表 1 预测基因结构和功能应用的软件

Table 1 The software for prediction of the gene structure and function

编号	名称	功能
1	ProtParam&ProtScale	蛋白质理化性质分析
2	PSORT	亚细胞定位预测
3	SignalP4.1 Server&TMHMM	信号肽和跨膜区预测
4	NetPhos2.0Server	蛋白磷酸化分析
5	ProtFun	蛋白功能预测
6	SMART	结构域预测
7	ScanProsite	Motif 搜索
8	SOPMA	蛋白二级结构预测
9	SWISS - MODEL Workspace	同源建模和三维结构预测

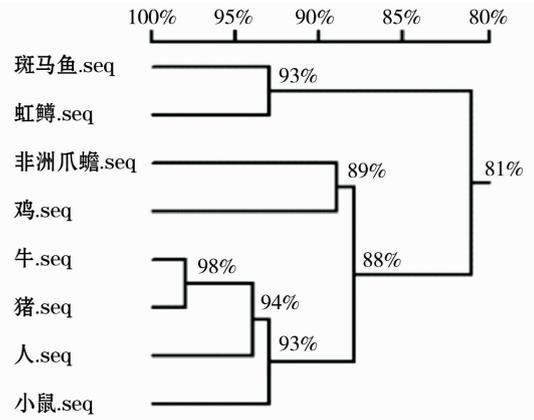
2 结果与分析

2.1 物种间 RB1 氨基酸序列同源性及进化关系分析

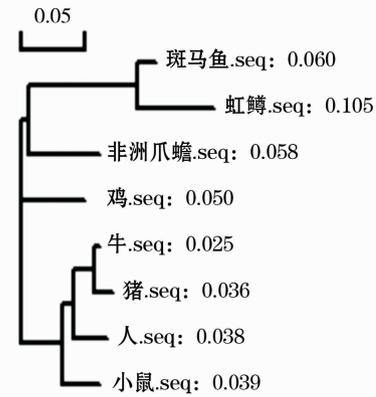
将鸡、非洲爪蟾、虹鳟、斑马鱼、小鼠、人、猪、牛的 RB1 氨基酸序列进行比对分析,发现鸡与牛、猪、人和小鼠之间序列的同源性很高,鸡与非洲爪蟾处于同一个小分支,在系统进化树中距离最近(见图 1A)。进化聚类分析结果表明,RB1 基因构建的生物进化关系图谱与已知物种进化关系基本一致(见图 1B)。

2.2 RB1 理化性质分析

利用鸡的 RB1 mRNA 序列,用 ProtParam 工具对 RB1 的氨基酸组成、分子质量和理论等电点进行预测。RB1 各种氨基酸残基的含量:半胱氨酸含量为 1.6%,亮氨酸含量最高(10.6%),丝氨酸次之(8.9%),色氨酸最低(0.7%),见表 2。亮氨酸和丝氨酸在促进机体的正常生长发育中发挥重要作用。



A 物种间RB1氨基酸序列同源性分析



B 物种间RB1氨基酸序列进化关系分析

图 1 RB1 氨基酸序列聚类分析结果

Fig. 1 Cluster analysis result of amino acid sequences of the RB1

表 2 RB1 各种氨基酸残基含量

Table 2 The content of a variety of amino acid residues of the RB1

氨基酸	数目/ 含量/ 个 %	氨基酸	数目/ 含量/ 个 %
丙氨酸(Ala ,A)	55 6.0	赖氨酸(Lys ,K)	69 7.5
精氨酸(Arg ,R)	47 5.1	蛋氨酸(Met ,M)	23 2.5
天冬酰胺(Asn ,N)	41 4.5	苯丙氨酸(Phe ,F)	42 4.6
天冬氨酸(Asp ,D)	50 5.4	脯氨酸(Pro ,P)	52 5.6
半胱氨酸(Cys ,C)	15 1.6	丝氨酸(Ser ,S)	82 8.9
谷氨酰胺(Gln ,Q)	44 4.8	苏氨酸(Thr ,T)	58 6.3
谷氨酸(Glu ,E)	60 6.5	色氨酸(Trp ,W)	6 0.7
甘氨酸(Gly ,G)	34 3.7	酪氨酸(Tyr ,Y)	28 3.0
组氨酸(His ,H)	20 2.2	缬氨酸(Val ,V)	52 5.6
异亮氨酸(Ile ,I)	45 4.9	吡咯赖氨酸(Pyl ,O)	0 0
亮氨酸(Leu ,L)	98 10.6	硒半胱氨酸(Sec ,U)	0 0

RB1 的分子质量为 104 ku,理论等电点为 8.37。说明 RB1 为碱性氨基酸。

利用 ProtScale 工具中的 Kyte - Doolittle 法进行疏水性分析,分值代表其疏水性大小,负值越大(绝

对值)表示亲水性越强,正值越大表示疏水性越强。结果表明,RB1 的总平均疏水指数为 -0.413 ,整条链中亲水性氨基酸残基多于疏水性氨基酸(见图 2),因此 RB1 为亲水性蛋白,可推测其为可溶性蛋白。分析结果表明 RB1 的不稳定系数为 47.56 ,表明该蛋白质性质不稳定。

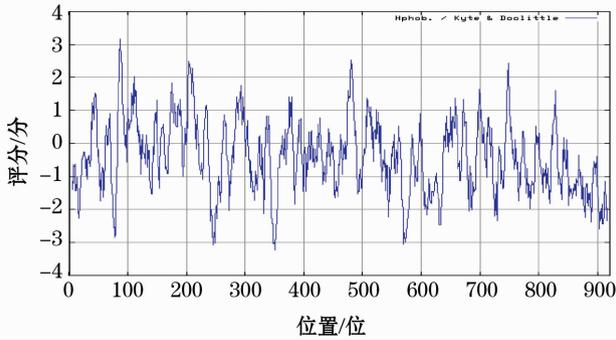


图 2 RB1 疏水性分析结果

Fig. 2 Hydrophobicity analysis result of the RB1

2.3 RB1 亚细胞定位预测

运用 PSORT II 工具预测 RB1 亚细胞定位,根据 $k-NN$ 算法预测结果可知:RB1 位于细胞核的概率最大,占整体的 69.6% ;其次是细胞质和细胞膜,概率为 8.7% ;位于囊泡分泌系统、高尔基体和内质网的概率均是 4.3% 。说明 RB1 可能在细胞核内发挥生物学作用,在细胞质和细胞膜中也有存在的可能。

2.4 RB1 信号肽预测和跨膜区分析

利用鸡的 RB1 mRNA 序列,分别采用 SignalP4.1 Server 和 TMHMM 程序对 RB1 信号肽和蛋白质跨膜区进行分析预测。C 值(C-score)代表原始信号肽裂解位点记分,值越大表示该点出现裂解位点可能性越大;S 值(S-score)代表信号肽评分,值越大表示该氨基酸位于信号肽区域概率越大,越低表示相应氨基酸不含信号肽或位于成熟蛋白部分;Y 值(Y-score)是最有可能的信号肽裂解位点。D 值(D-score)是 S 均值和最大 Y 值加权平均值,用于区分信号肽和非信号肽。D 值为 0.450 是区分信号肽和非信号肽临界值,RB1 蛋白预测 D 值为 0.111 ,因此分析发现 RB1 蛋白不含有信号肽,不属于分泌性蛋白(见 290 页彩图 3)。

使用 TMHMM 工具进行跨膜区分析发现 RB1 没有跨膜区(见 290 页彩图 4),非跨膜螺旋区(ExpAA)预测值为 0.2908 ,该值 >18 时说明预测蛋白为跨膜蛋白,蛋白 N 端位于膜内一侧的可能性是 0.01381% ,总体来看该序列位于膜外,即 RB1 基因所编码蛋白为非跨膜蛋白。

2.5 蛋白磷酸化分析

蛋白磷酸化是一种最普遍的翻译后修饰,对生命

活动的整个过程起调节作用,因此蛋白质的磷酸化分析对了解蛋白质可能的功能具有重要意义。利用 NetPhos 2.0 Server 对 RB1 蛋白磷酸化作用位点进行预测,结果表明,该蛋白具有 5 个酪氨酸磷酸化作用位点、14 个苏氨酸磷酸化作用位点和 40 个丝氨酸磷酸化作用位点,即具有多个不同的磷酸化作用位点(见 290 页彩图 5)。

2.6 RB1 蛋白功能预测

利用 ProtFun 2.2 预测 RB1 蛋白细胞水平功能类别、基因存在类别和酶类别,结果表明:该蛋白细胞水平功能是组成细胞外被膜(可能性为 0.804% ,概率为 13.180%),可能参与转运结合和氨基酸生物合成;属于非酶可能性为 0.766% 。在基因存在类别预测中荷尔蒙可能性为 0.114% ,概率最大(为 17.538%);生长因子可能性为 0.040% ,概率为 2.857% 。见表 3。

表 3 RB1 蛋白功能预测结果

Table 3 Function prediction result of the RB1 protein

项目	分类	可能性/%	概率/%
功能类别	细胞外被膜 (cell envelope)	0.804	13.180
	转运结合 (transport and binding)	0.820	2.000
	氨基酸生物合成 (amino acid biosynthesis)	0.038	1.727
	生物合成的辅助因子 (biosynthesis of cofactors)	0.118	1.639
	脂肪酸代谢 (fatty acid metabolism)	0.017	1.308
酶类	酶 (enzyme)	0.234	0.817
	非酶 (nonenzyme)	0.766	1.073
基因存在类别	荷尔蒙 (hormone)	0.114	17.538
	生长因子 (growth factor)	0.040	2.857
	胁迫应答 (stress response)	0.223	2.534
在类别	免疫应答 (immune response)	0.123	1.447
	离子通道 (ion channel)	0.079	1.386
	信号转导 (signal transducer)	0.272	1.271

2.7 RB1 结构域预测

结构域是蛋白质中具有特异结构和独立功能的区域,结构域与蛋白质完成生理功能有着密切的关系。利用鸡的 RB1 mRNA 序列,使用 EMBL 网站的 SMART 工具对 RB1 的结构域进行分析预测,结果表明,该蛋白具有两个保守结构域和一个周期性结构域(见 290 页彩图 6)。它们分别是位于氨基酸 $363 \sim 565$ 位的 RB_A 结构域,位于氨基酸 $760 \sim 920$ 位的 C-末端结构域和位于氨基酸 $652 \sim 753$ 位的域细胞周期蛋白。RB_A 域是 RB1 的成熟肽区,是细胞周期蛋白折叠区。域细胞周期蛋白是一种细胞周期和转录调控中的蛋白质识别功能域。这个周期性结构域被发现于细胞周期蛋白,但也发现于转录因子 IIB 核心领域 (TFIIB) 和视网膜母细胞瘤肿瘤抑制基因。Rb 的 C-末端结构域是高亲和力结合的复合物,与



E2F 启动子结合后抑制它的活性,从而阻断细胞周期 G1 - S 过渡,抑制细胞生长。这个结构域有一股环螺旋结构,与 E2F1 和 DP1 直接交互作用。

2.8 RB1 Motif 搜索

Motif 是蛋白质分子中具有特定空间构象和特定功能的结构成分,它对蛋白家族起决定性作用。利用鸡的 RB1mRNA 序列,用 ExPASy 网站上的 ScanProsite 工具对 RB1 中的 Motif 进行搜索,结果表明:共找到 48 个 Prosite 位点。其中包括 8 个 N - 肉豆蔻酸化位点,位于氨基酸 10 ~ 15 位 (GAarSQ)、21 ~ 26 位 (GGagTA)、22 ~ 27 位 (GAgTAS)、246 ~ 251 位 (GQnrNA)、294 ~ 299 位 (GVvaSN)、425 ~ 430 位 (GQgcAE)、874 ~ 879 位 (GQdeAD)、912 ~ 917 位 (GNdtSA); 15 个酪氨酸激酶 II 磷酸化位点,位于氨基酸 17 ~ 20 位 (TspE)、32 ~ 35 位 (TrIE)、53 ~ 56 位 (SvrE)、99 ~ 102 位 (TftE)、130 ~ 133 位 (TkvD)、341 ~ 344 位 (SqlE)、381 ~ 384 位 (SatD)、453 ~ 456 位 (SeeE)、493 ~ 496 位 (Sqsd)、498 ~ 501 位 (TsaE)、593 ~ 596 位 (TaaD)、618 ~ 621 位 (SnpD)、726 ~ 729 位 (TnqE)、827 ~ 830 位 (SigE)、916 ~ 919 位 (SanE); 16 个 PKC 磷酸化位点,位于氨基酸 14 ~ 16 位 (SqR)、53 ~ 55 位 (SvR)、129 ~ 131 位 (StK)、239 ~ 241 位 (StR)、242 ~ 244 位 (TpR)、346 ~ 348 位 (TpR)、383 ~ 385 位 (Tdk)、433 ~ 435 位 (SqR)、604 ~ 606 位 (SpK)、630 ~ 632 位 (TqK)、730 ~ 732 位 (TK)、765 ~ 767 位 (SnR)、787 ~ 789 位 (SpR)、819 ~ 821 位 (TpR)、835 ~ 837 位 (SeK)、899 ~ 901 位 (StR); 2 个 cAMP 和 cGMP 依赖的蛋白激酶磷酸化位点,位于氨基酸 78 ~ 81 位 (KKeT)、607 ~ 610 位 (KKaS); 6 个 N - 糖基化位点,位于氨基酸 237 ~ 240 位 (NGST)、396 ~ 399 位 (NCTV)、462 ~ 465 位 (NF-SK)、591 ~ 594 位 (NHTA)、682 ~ 685 位 (NESE)、913 ~ 916 位 (NDTS); 1 个酪氨酸激酶磷酸化位点,位于氨基酸 309 ~ 315 位; 1 个亮氨酸拉链模式结构,位于氨基酸 654 ~ 675 位。

2.9 RB1 蛋白二级结构预测

通过 SOPMA 程序预测 RB1 蛋白二级结构,发现由 429 个氨基酸残基组成的 α - 螺旋,占整个二级结构的 46.58%,是 RB1 二级结构的主要构成;无规则卷曲由 332 个氨基酸构成,占整体的 36.05%;延伸带和 β - 转角分别由 107,53 个氨基酸残基构成,在整体中所占比例分别为 11.62% 和 5.75%。见 290 页彩图 7。

2.10 RB1 同源建模和三维结构预测

蛋白质的三维结构能够提供丰富的功能信息。利用鸡和人的 RB1mRNA 序列,应用 SWISS - MODEL Workspace 蛋白质建模平台对 RB1 进行同源建模和三维结构预测^[9-12],结果表明:RB1 成熟肽和 4elj. 1.

A(2.70A)(37 ~ 763 个)有 73.16% 的氨基酸序列一致,人的成熟肽和 4elj. 1. A(2.70A)(75 ~ 928 个)有 99.23% 的氨基酸序列一致。一般同源性超过 30% 的蛋白质序列均可建立精确结构模型,序列同源性越高则建立模型结构准确性越高^[13-15]。SWISS - MODEL Workspace 平台以 4elj. 1. A 为模板推测出了 RB1 的三维结构(见 290 页彩图 8)。

3 讨论

3.1 鸡 RB1 基因基本生物信息学分析

研究对氨基酸序列进行进化聚类分析,结果表明鸡与哺乳动物之间序列的同源性很高,并且该基因在物种进化过程中发生一定程度的突变。理化性质分析结果表明,亮氨酸和丝氨酸可以促进机体的正常生长发育,RB1 为不稳定的碱性亲水蛋白,还可能为可溶性蛋白。

3.2 鸡 RB1 基因功能预测

亚细胞定位分析结果表明 RB1 可能在细胞核内合成;信号肽和跨膜区分析结果表明鸡 RB1 不具有信号肽,为非跨膜蛋白,证实该蛋白不是分泌性蛋白;蛋白磷酸化分析结果表明 RB1 分子中存在多个磷酸化作用位点,可以促进细胞生长;蛋白功能预测结果表明该蛋白细胞水平功能是组成细胞外被膜,可能参与转运结合和氨基酸生物合成,属于非酶类,可能以荷尔蒙和生长因子等形式存在于基因中,可调控细胞的生长发育;进行结构域预测和 Motif 搜索发现该蛋白成熟肽具有 3 个结构域,维持蛋白构象并调控细胞周期和转录,而且 RB1 分子中存在 N - 肉豆蔻酸化、酪氨酸激酶 II 磷酸化、PKC 磷酸化以及 N - 糖基化等位点,可参与调控细胞生长、增殖、凋亡等过程。

3.3 鸡 RB1 基因结构预测

SOPMA 预测 RB1 蛋白二级结构主要由 α - 螺旋构成,体现结构域的多种生物学作用。对蛋白三维结构进行分析后发现,鸡 RB1 成熟肽序列与模板 4elj. 1. A 相似性很高,可以根据该模板的结构建立 RB1 的蛋白质结构模型。比较鸡与人的三维结构,发现两者之间空间构型差别较小,说明鸡与人的氨基酸序列极其相似。

4 结论

研究通过鸡 RB1 基因基本生物信息学分析预测编码氨基酸序列的结构与功能,可为研究 RB1 基因对细胞的调控功能提供依据。

参考文献:

- [1] EWEN M E, XING Y G, LAWRENCE J B, et al. Molecular cloning, chromosomal mapping, and expression of the cDNA for p107, a retinoblastoma gene product - related protein [J]. Cell, 1991, 66 (6): 1155 - 1164.
- [2] HANNON G J, DEMETRICK D, BEACH D. Isolation of the Rb - related p130 through its interaction with CDK2 and cyclins [J]. Genes Dev, 1993, 7(12A): 2378 - 2391.

(下转第 29 页)