



大豆 Argonaute4 的生物信息学分析

曹乐生, 王晗慧, 张天旭, 解莉楠

(东北林业大学 生命科学院, 黑龙江 哈尔滨 150040)

摘要:在 RNA 介导的 DNA 甲基化途径(RNA-directed DNA methylation pathway)中, Argonaute4(AGO4)蛋白是 DNA 被甲基化修饰的关键蛋白。研究大豆中 Argonaute4 蛋白能够为大豆甲基化研究奠定基础, 本研究参考 NCBI、Phytozome、UniProt KB 等网站及数据库, 对拟南芥的 *AGO4* 基因进行同源性分析, 确定了大豆中 AGO4 蛋白及其基因序列, 并对其进行分析预测。研究发现, 大豆中 *AGO4* 包括 3 个基因 *GmAGO4A*、*GmAGO4B*、*GmAGO4C*, 它们的一级结构、二级结构及理化性质都很相似, 三级结构有差异, 大豆 AGO4C 与拟南芥 AGO4 结构更相似。在 302 个野生大豆品系中对 *GmAGO4* 的 3 个基因的外显子进行核酸及蛋白序列的分析, 发现一些大豆品系的蛋白质由于突变三维结构发生了变化。对大豆中 AGO4 及其参与的甲基化途径的研究具有指导意义。

关键词:大豆; AGO4; 生物信息学; DNA 甲基化

Bioinformatics Analysis of Soybean Argonaute4

CAO Le-sheng, WANG Han-hui, ZHANG Tian-xu, XIE Li-nan

(School of Life Science, Northeast Forestry University, Harbin 150040, China)

Abstract: In the RNA-directed DNA methylation pathway, the Argonaute4 (AGO4) protein plays a key role in DNA methylation. In order to study Argonaute4 protein in soybean and lay a foundation for the study of methylation in soybean, in this paper, homology analysis of *AGO4* gene in *Arabidopsis thaliana* was performed with NCBI, Phytozome, UniProt KB and other websites, moreover, AGO4 protein and its gene sequence in soybean were confirmed and analyzed. The results indicate that there are three homologous genes of *AtAGO4* in soybean, *GmAGO4A*, *GmAGO4B*, *GmAGO4C* respectively, which primary structures, secondary structures and physicochemical property are similar, but the tertiary structures are different, the tertiary structures of *GmAGO4C* and *Arabidopsis* AGO4 are more similar. The nucleic acid and protein sequences of the three genes' exons were analyzed in 302 wild soybean lines, and it was found that the tertiary structures of some soybean lines changed due to mutation. The results of this paper will be helpful for further study on the function of *AGO4* in the methylation pathway in soybean.

Keywords: Soybean; *AGO4*; Bioinformatics; DNA methylation

大豆[*Glycine max* (Linn.) Merr.]是我国重要粮食和油料作物, 富含植物蛋白、脂肪、异黄酮等, 具有极高的营养价值和经济价值, 因此在世界各地广泛种植。由于大豆从古四倍体向二倍体的不完全进化, 造成基因中存在串联重复的序列, 结构上类似于转座子^[1], 许多基因被甲基化遮蔽, 导致表达程度下降。有研究表明, 大豆基因存在 DNA 甲基化多样性和遗传多样性^[2-3]。DNA 甲基化是表观遗传学的重要调控机制之一, 甲基化水平通过多种调控因子的共同作用影响基因的表达^[4]。在高等植物中, DNA 甲基化对基因组印记、转座子沉默和 X 染色体失活具有重要意义^[5]。

RNA 介导的 DNA 甲基化途径依赖于 24nt siRNA 的存在, 在一系列酶和蛋白的作用下, 在 DNA 胞嘧啶第五位碳原子上添加甲基基团, 导致 DNA 的表达水平改变^[6]。AGO4 是 RdDM(RNA-directed DNA methylation pathway)途径中的关键因子^[7-10], 其在进化上高度保守^[11], 对建立与维持 DNA 甲基化、调控基因表达、调节植物的逆境生理、提高适应性具有重要作用。近年来, 关于 AGO4 功能的研究已经有了很大的进展。在细胞质中, AGO4 蛋白结合单链的 24nt siRNA 形成沉默复合体, 随后入核招募 DNA 甲基转移酶, 催化实现 DNA 甲基化^[12-17], 导致基因沉默。在生物发育的不同阶段基因组甲基化水平

收稿日期: 2018-08-03

基金项目: 中央高校基本科研业务费专项资金项目(2572017DA06, 2572014CA21)。

第一作者简介: 曹乐生(1997-), 男, 学士, 主要从事大豆表观遗传学调控因子研究。E-mail: caolesheng@nefu.edu.cn。

通讯作者: 解莉楠(1978-), 女, 博士, 副教授, 主要从事植物抗逆生物学研究。E-mail: linanxie@nefu.edu.cn。

是呈现动态的,AGO4 能与其它调控因子相互作用来改变基因组的甲基化水平^[18]。在拟南芥中发现,AGO4 在早期胚珠发育过程中有防止减数分裂前配子体异常的作用^[19-20]。此外,已知拟南芥的 NB-LRR(富含亮氨酸的重复序列)基因家族在抗病性方面起着重要的作用,而野生型中大部分 NB-LRR 编码基因都被甲基化,AGO4 的突变则会导致此基因家族的胞嘧啶甲基化减少^[21]。研究发现在芸苔属植物中 AGO4 相较拟南芥有显著扩增,并且在抗链球菌方面起作用^[22]。水稻的 AGO4 对基因的甲基化模式的调节有重要作用^[23]。AGO4 可以与 DNA 损伤结合蛋白 2 (Damage-specific DNA-binding protein 2, DDB2) 作为重要的调控因子相互作用参与到 DNA 损伤修复中^[24-25]。一些疾病的发生也跟 AGO4 有密切的联系,一些病毒细菌等病原体的感染会引起 AGO4 及其调控的基因表达水平的改变^[26-28]。

目前已经有研究者对拟南芥、烟草、小麦的 AGO4 进行了相对详细的研究,但大豆 AGO4 的研究尚未报道。本文利用生物信息学方法,以拟南芥 AGO4 基因序列为参照,在大豆基因组中鉴定 *GmA-GO4* 基因,进行序列同源性分析、系统发生分析、蛋白质序列理化性质分析、二级结构分析及结构域分析,并以 302 种野生大豆品系作为数据库分析 AGO4 的多态性变化,为大豆 AGO4 的进一步研究提供参考。

1 材料与方法

1.1 材料

拟南芥 (*Arabidopsis thaliana*) AGO 家族基因及氨基酸序列、大豆 (*Glycine max*) 全基因组和氨基酸序列、9 种双子叶植物和 10 种单子叶植物中与拟南芥 AGO4 氨基酸序列相似度高的氨基酸序列。

1.2 方法

1.2.1 大豆 AGO4 基因的筛选和结构分析 利用 TAIR 网站查找到的 9 个拟南芥 AGO 家族基因 (AGO10 为丧失正常功能的假基因,因此选择拟南芥的 AGO1-9 基因) 确定大豆 AGO 基因,在 Phytozome (<http://www.phytozome.net/index.php/>) 网站上进行同源比对,筛选与拟南芥 AGO4 蛋白同源性高的大豆候选蛋白。利用 NCBI 数据库进行 Blastp 序列验证,筛选出同源性高的大豆 AGO4 基因并命

名。利用 GSDS 网站在线分析基因的结构特点。

1.2.2 多序列比对 在 Phytozome 网站查找 9 种双子叶植物和 10 种单子叶植物中与拟南芥 AGO4 蛋白序列相似度最高的 19 条序列。利用 DNAMAN 软件对拟南芥 AGO4 蛋白序列、大豆中筛选到的 AGO4 基因的蛋白序列以及上述植物的蛋白序列进行多序列比对。

1.2.3 基因进化分析 利用 MEGA 7.0 软件,Bootstrap method 进行 1 000 次重复取样,采用最大似然法构建系统进化树分析拟南芥与大豆 AGO 的亲缘关系。

1.2.4 基因特性分析 利用 Protparam (<https://web.expasy.org/protparam/>) 分析新命名蛋白质的分子量、等电点、稳定性、总亲水性等理化性质。利用 SignalP 4.1 Server 分析新命名蛋白质有无信号肽序列。利用 PSORT II Prediction 进行新命名蛋白质的亚细胞定位。利用 NetPhos 3.1 Server 分析新命名蛋白质的磷酸化位点。

1.2.5 基因二级和三级结构分析 采用 SOPMA 软件对新命名蛋白质的二级结构进行预测,分析蛋白质的 α -螺旋、 β -折叠、 β -转角和无规则卷曲所占比例。利用 Smart 网站分析新命名的 3 个蛋白质的结构域。利用 Swiss-Model 预测新命名蛋白质的三级结构。

1.2.6 基因多态性分析 利用 ANNOVAR 软件建立本地数据库,以 *Glycine max* Wm82 (W82) 作为对照,在拥有 302 个野生大豆品系的本地数据库中对大豆新命名 AGO4 基因的外显子进行核酸及蛋白序列的分析,寻找发生突变的大豆品系。利用 SWISS-MODEL 建模分析基因的突变类型。

2 结果与分析

2.1 大豆 AGO4 的基因鉴定与命名

根据同源比对得到的大豆中与拟南芥相似性高的基因如表 1 所示。按照同源性高低顺序,将大豆基因命名为 *GmAGO4A*、*GmAGO4B* 和 *GmAGO4C* (表 2)。GSDS 网站分析示意基因结构如图 1。结果显示,3 个基因分别位于第 2、14、20 号染色体上,编码区全长分别为 2 721、2 721、2 844 bp,编码蛋白质序列长度几乎一致。*GmAGO4A* 和 *GmAGO4B* 基因有 23 个外显子,*GmAGO4C* 则有 24 个外显子,基因结构相似。

表 1 大豆与拟南芥 *AGO* 家族序列相似性比对

Table 1 Comparison of <i>AGO</i> sequence similarity between soybean and <i>Arabidopsis thaliana</i> (%)									
基因 Gene	<i>AGO1</i> (AT1G48410. 2)	<i>AGO2</i> (AT1G31280)	<i>AGO3</i> (AT1G31290)	<i>AGO4</i> (AT2G27040)	<i>AGO5</i> (AT2G27880)	<i>AGO6</i> (AT2G32940)	<i>AGO7</i> (AT1G69440)	<i>AGO8</i> (AT5G21030)	<i>AGO9</i> (AT5G21150)
Glyma. 09G167100	79. 1	45. 1	37. 7	53. 8	62. 9	52. 7	51. 3	54. 2	54. 2
Glyma. 16G217300	80. 3	45. 1	37. 2	53. 5	63. 3	52. 7	51. 1	54. 2	53. 8
Glyma. 02G002500	72. 0	45. 3	39. 3	54. 1	62. 6	54. 1	52. 6	54. 7	54. 4
Glyma. 10G242800	71. 8	44. 4	38. 9	53. 7	62. 5	54. 1	52. 7	55. 1	54. 1
Glyma. 20G151400	71. 2	44. 0	38. 4	52. 9	62. 3	53. 2	52. 1	54. 2	54. 4
Glyma. 10G002700	72. 4	44. 9	38. 9	51. 4	62. 3	53. 2	50. 7	53. 1	52. 9
Glyma. 04G150900	67. 0	44. 0	36. 3	52. 7	59. 5	54. 2	49. 9	55. 1	53. 1
Glyma. 05G011300	67. 2	43. 2	36. 4	52. 8	60. 2	53. 1	49. 9	55. 9	54. 2
Glyma. 06G213300	66. 5	42. 9	36. 5	51. 9	60. 4	53. 6	50. 1	55. 3	53. 0
Glyma. 17G119400	65. 5	43. 1	35. 0	51. 7	58. 9	51. 8	49. 0	54. 2	54. 1
Glyma. 12G083500	61. 2	45. 1	37. 4	51. 7	65. 3	53. 3	51. 0	51. 4	53. 0
Glyma. 11G190900	55. 6	42. 0	34. 9	—	59. 8	49. 5	46. 8	—	49. 0
Glyma. 02G111600	50. 7	47. 9	41. 5	48. 7	49. 3	48. 2	73. 7	50. 0	52. 1
Glyma. 01g053100	49. 0	47. 7	41. 1	48. 4	49. 1	44. 8	73. 8	50. 7	52. 0
Glyma. 20G052500	47. 2	40. 5	35. 3	79. 5	49. 0	72. 2	45. 8	76. 9	78. 1
Glyma. 02G274900	45. 6	41. 4	35. 2	81. 9	47. 2	73. 3	47. 2	78	80. 5
Glyma. 15G126700	43. 5	60. 0	49. 2	40. 6	43. 3	46. 0	48. 5	46. 1	45. 2
Glyma. 20G022900	44. 2	59. 9	49. 6	47. 2	45. 6	45. 4	48. 0	46. 5	47. 8
Glyma. 09G020200	—	30. 7	25. 5	—	—	—	23. 9	—	—
Glyma. 14G041100	45. 0	40. 6	35. 2	80. 5	46. 9	71. 9	46. 2	76. 6	79. 5
Glyma. 13G193200	43. 8	39. 3	35. 7	68. 6	45. 8	74. 8	45. 3	69. 2	70. 9
Glyma. 06G314500	35. 8	32. 4	27. 6	58. 0	37. 9	55. 5	36. 7	57. 6	58. 6

加粗字体表示大豆中与拟南芥相似性高的序列。
The font bold show the sequence of deeper homology with *Arabidopsis thaliana* of soybean.

表 2 大豆 *AGO4* 基因筛选、命名与定位

Table 2 Genetic screening, naming and localization of soybean <i>AGO4</i>					
命名 Name	基因 Gene	长度 Length/bp	染色体定位 Chromosomal localization	编码蛋白长度 Protein length/aa	同源性 Homology/%
<i>GmAGO4A</i>	Glyma. 02G274900	8185	Chromosome 2-NC_016089. 2(45798303. . 45806487)	906	81. 9
<i>GmAGO4B</i>	Glyma. 14G041100	7814	Chromosome 14- NC_016101. 2(3087012. . 3094825 , complement)	906	80. 5
<i>GmAGO4C</i>	Glyma. 20G052500	8762	Chromosome 20- NC_016107. 2(11866317. . 11875078)	947	79. 5

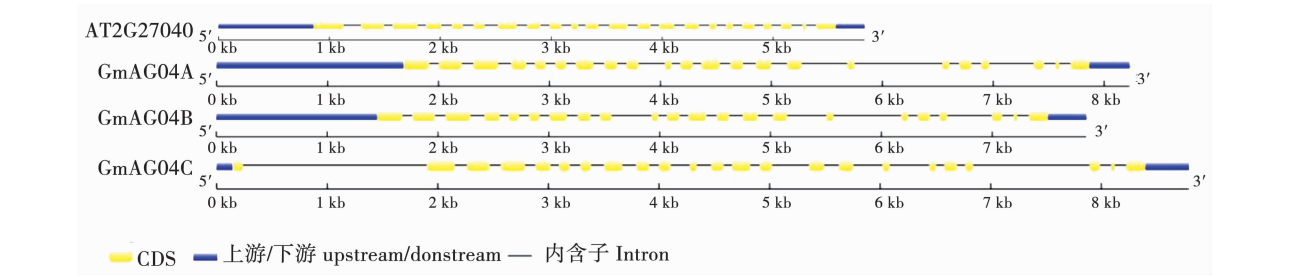


图 1 基因结构分布

Fig. 1 Gene structure distribution

2.2 大豆 AGO4 蛋白的多序列比对

拟南芥 AGO4 蛋白序列、大豆中 *GmAGO4A*、*GmAGO4B* 和 *GmAGO4C* 基因的蛋白序列以及其它植物的蛋白序列多序列比对结果如图 2 所示,所有序列的一致性高达 71. 07% ,可见 AGO4 作为 RdDM 途径中非要重要的成分在植物界中是非常保守的。



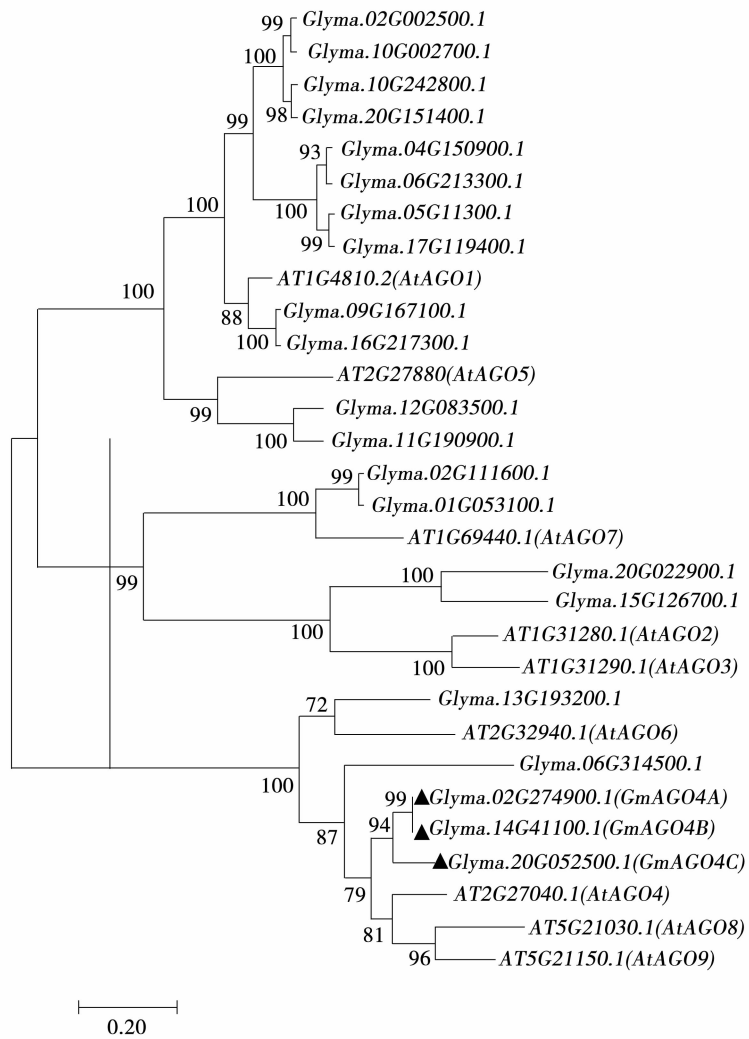
图2 AGO4 蛋白序列多序列比对

Fig. 2 Multiple sequence comparison of AGO4 protein sequences

2.3 大豆 AGO4 基因的进化分析

蛋白质序列分析结果显示基因间随机检验得分均大于70,说明进化树可信度较高,可以直观地反映出大豆 *GmAGO4A*、*GmAGO4B*、*GmAGO4C* 和拟南芥 *AtAGO4* 的序列在进化过程中的亲缘远近关系。拟南芥中 *AtAGO4*、*AtAGO8*、*AtAGO9* 与大豆基因 *Glyma. 02G274900*、*Glyma. 14G041100*、*Glyma. 20G052500*

距离近,且 *AtAGO4* 距离3个基因更近,说明亲缘关系更近,进一步验证了命名的准确性(图3)。在大豆的3个基因中, *Glyma. 02G274900* 和 *Glyma. 14G041100* 亲缘较近,且与拟南芥 *AtAGO4* 的亲缘距离都较近,说明 *GmAGO4A* 与 *GmAGO4B* 可能由同一祖先序列进化而来; *Glyma. 20G052500* 与拟南芥 *AtAGO4* 的亲缘距离则相对较远。



三角表示 3 个大豆 *AGO4* 在进化树中的位置。
The triangle represents the location of three soybean *AGO4* species in the evolutionary tree.

图 3 大豆与拟南芥 *AGO* 基因系统进化树
Fig. 3 Evolution tree of *AGO* gene system of soybean and *Arabidopsis thaliana*

2.4 大豆 AGO4 的基本性质、亚细胞定位及磷酸化位点分析

GmAGO4A、*GmAGO4B* 和 *GmAGO4C* 与拟南芥的 *AGO4* 基因编码的蛋白质相比,分子量相差不大,等电点相似为碱性,性质不稳定,为亲水性蛋白,且不

含信号肽序列(表 3),在细胞核内出现可能性最大(图 4)。拟南芥和大豆中 *AGO4* 蛋白的磷酸化位点在数量和类型上相似,其中丝氨酸最多,苏氨酸次之,酪氨酸最少,可能存在相同的蛋白激酶磷酸化位点(表 5)。

表 3 大豆 AGO4 蛋白的理化性质
Table 3 Physical and chemical properties of soybean AGO4

蛋白 Protein	相对分子质量 Relative molecular mass	等电点 Isoelectric point	稳定性 Stability	总平均亲水性 Total average hydrophilicity	信号序列 Signal sequence
AtAGO4	102840.44	9.05	不稳定	亲水	—
GmAGO4A	101152.72	9.04	不稳定	亲水	—
GmAGO4B	101417.06	9.07	不稳定	亲水	—
GmAGO4C	106128.38	8.93	不稳定	亲水	—

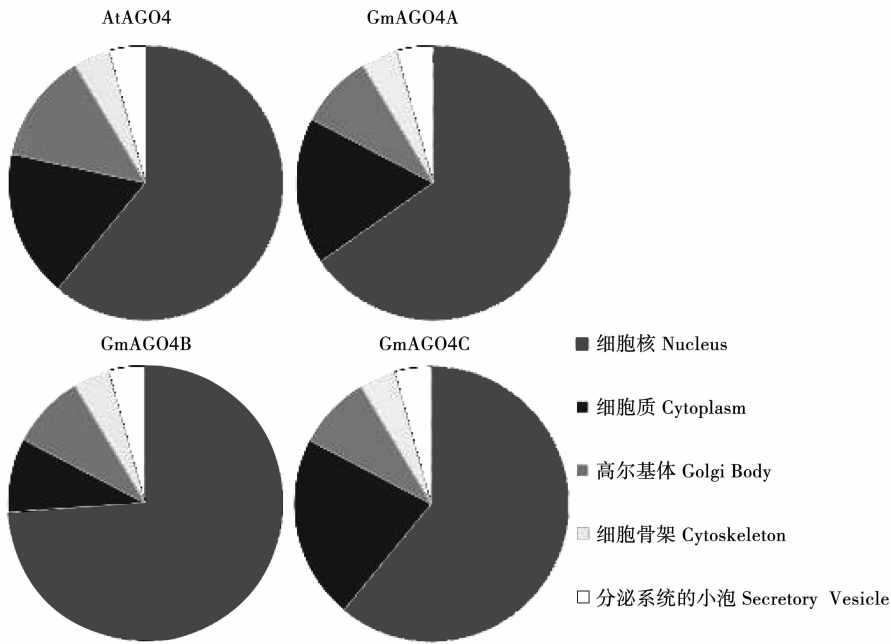


图4 大豆 AGO4 蛋白的亚细胞定位

Fig. 4 Subcellular localization of soybean AGO4 protein

表 4 大豆 AGO4 蛋白磷酸化位点及数量

Table 4 Phosphorylation sites and number of AGO4 protein in soybean

蛋白 Protein	丝氨酸 Ser	苏氨酸 Thr	酪氨酸 Tyr
AtAGO4	47	29	8
GmAGO4A	47	25	7
GmAGO4B	48	24	7
GmAGO4C	48	24	4

2.5 大豆 AGO4 的二级结构分析

大豆 AGO4 的二级结构预测结果如图 5。3 个

基因的 α -螺旋、 β -折叠所占比例相差不大,且相加和接近 50%,二级结构的相似性极高。

2.6 大豆 AGO4 的结构域及三级结构预测

GmAGO4A、*GmAGO4B* 和 *GmAGO4C* 具有 5 个共同结构域:ArgoN、DUF1785、PAZ、ArgoL2 和 Piwi,且分布大致相同(图 6 和表 5)。SWISSMODEL 对拟南芥 AGO4 与大豆中 3 个基因编码蛋白质三级结构建模表明大豆中 3 个基因的蛋白的三级结构与拟南芥 AGO4 的三级结构有差异,但是 *GmAGO4C* 与 *AtAGO4* 更相似(图 7)。

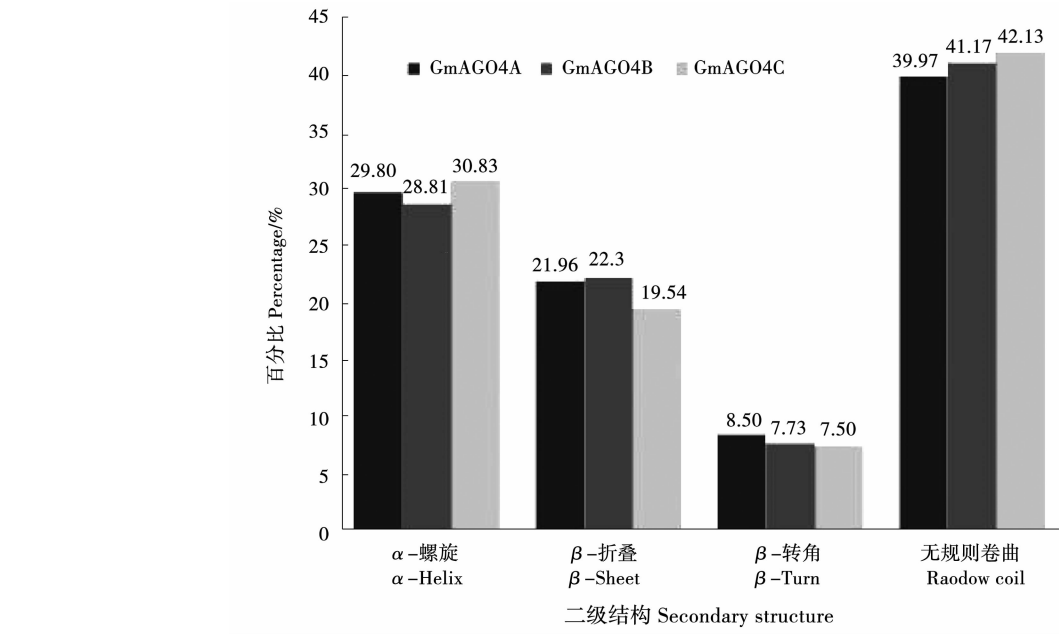


图5 大豆 AGO4 蛋白二级结构分析

Fig. 5 Secondary structure analysis of soybean AGO4 protein

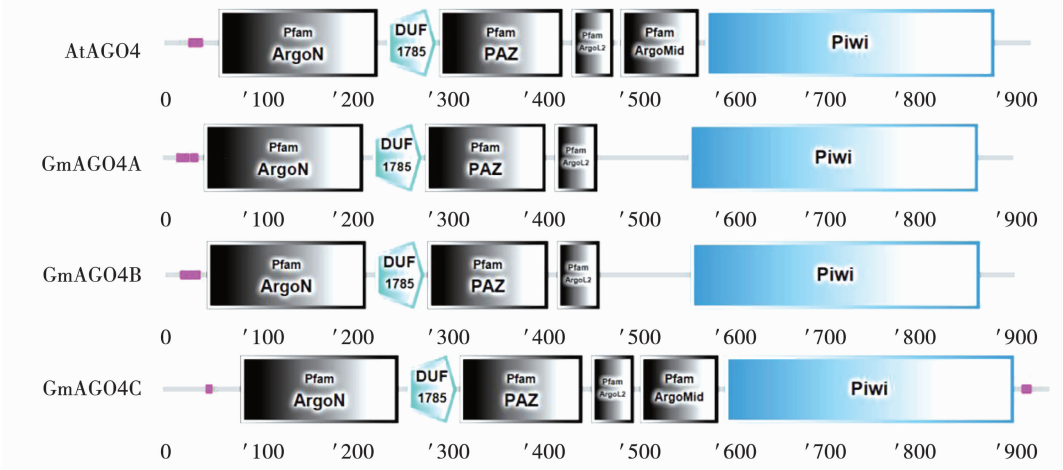


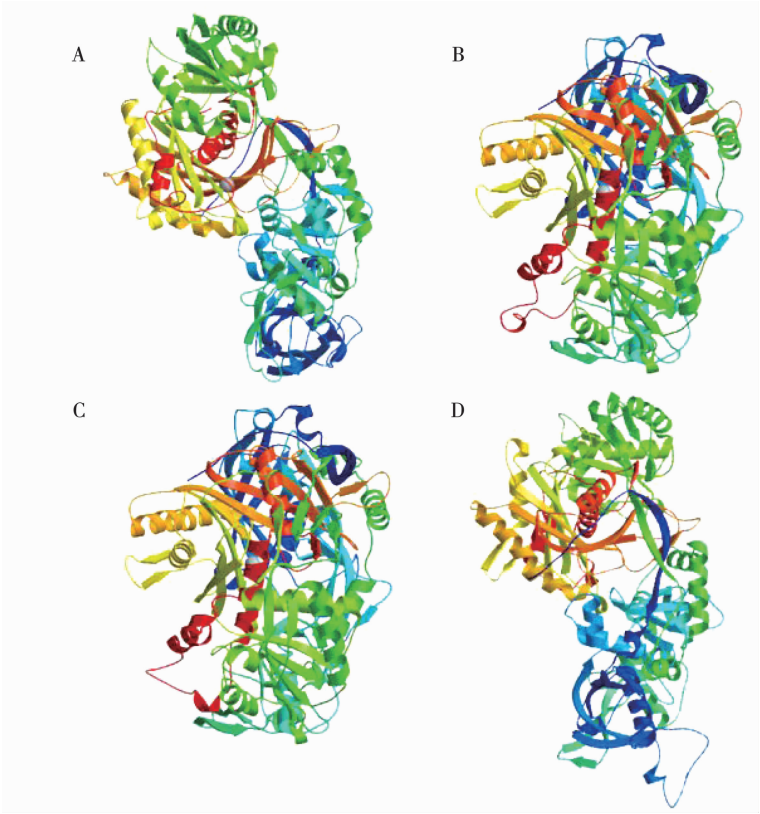
图 6 大豆 AGO4 蛋白的结构域预测

Fig. 6 Structural domain prediction of soybean AGO4 protein

表 5 大豆 AGO4 蛋白的结构域在肽链的位置

Table 5 The structural domain of soybean AGO4 protein in the peptide chain (aa)

蛋白 Protein	Pfam : ArgoN	DUF1785	Pfam : PAZ	Pfam : ArgoL2	Pfam : ArgoMid	Piwi
GmAGO4A	45 ~ 214	224 ~ 276	280 ~ 408	418 ~ 463	—	560 ~ 868
GmAGO4B	45 ~ 214	224 ~ 276	280 ~ 408	418 ~ 463	—	560 ~ 868
GmAGO4C	84 ~ 252	262 ~ 314	318 ~ 448	458 ~ 503	510 ~ 593	600 ~ 908



A: AtAGO4; B:GmAGO4A;C: GmAGO4B; D:GmAGO4C.

图 7 大豆 AGO4 3D 建模分析

Fig. 7 3D modeling analysis of soybean AGO4

2.7 大豆 AGO4 多态性分析

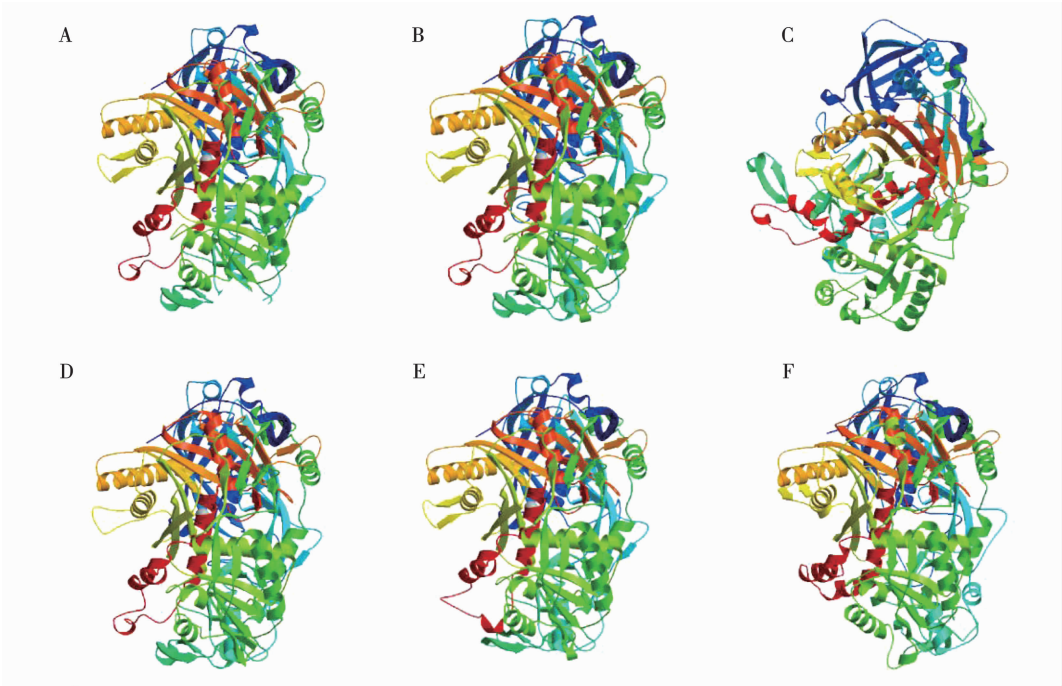
本地数据库中对大豆 *AGO4* 的外显子核酸及编码蛋白序列的分析发现了非同义突变,碱基和氨基酸序列的变化如表 6 所示。发现 302 个大豆品系的 *AGO4* 可产生非同义突变的 SNP 数目少,可见在这 302 个大豆品系中该亚家族基因和蛋白质序列相对保守。

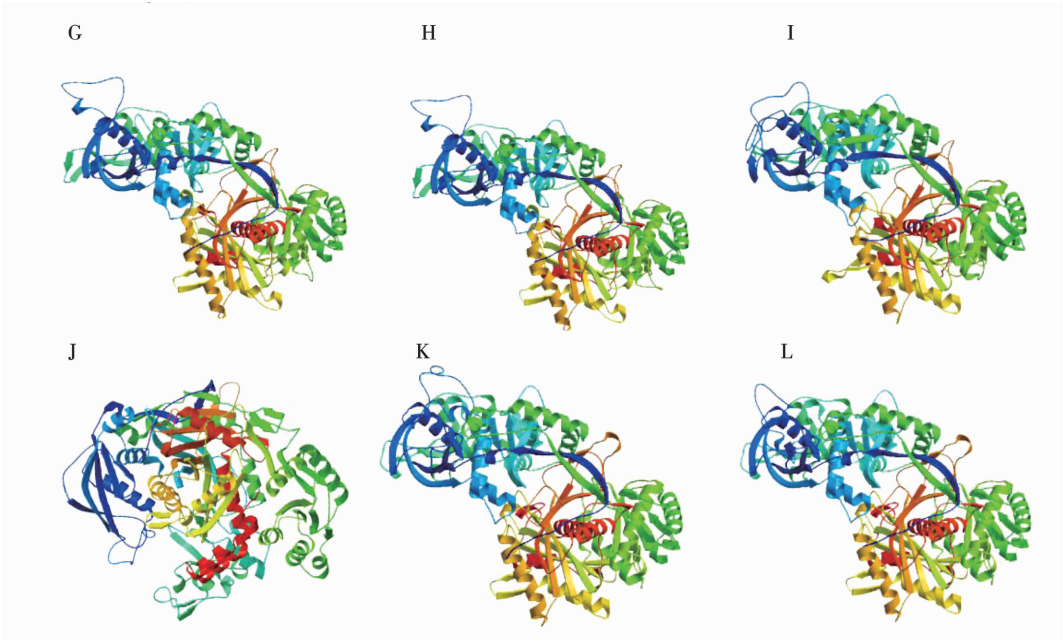
突变类型分析结果显示:*GmAGO4A* 中第 516 位的 T 突变为 A 和第 976 位的 T 突变为 C 时,导致了 Pfam:ArgoN 和 Pfam:PAZ 两个结构域各一个氨基酸

的改变(类型 2)进而大大改变了蛋白质的三维结构。而这两个位点单独的突变(类型 1 和 3)则不会发生三维结构的改变,可见突变的累计可能会引起蛋白结构与功能的变化,而有超过 2/3 的大豆品系含有类型 2 的 *GmAGO4A* 蛋白。*GmAGO4C* 中 122 位的 A 突变为 T,导致 41 位的 N 变成 I,但是此突变并不在结构域里,却会引起蛋白三维结构的巨大改变,推测此位点是氨基酸盘曲折叠为正确蛋白质过程中发挥非常重要作用的位点,仅有 21 个大豆品系含有此类型的突变蛋白(表 6 和图 8)。

表 6 大豆 AGO4 多态性分析
Table 6 Polymorphism analysis of soybean AGO4

蛋白 Protein	突变类型 Mutation type	碱基变化 Nucleotide variation			氨基酸序列变化 Amino acid sequence changes			大豆品系数量 Number of soybean varieties
		位点 Site	原碱基 Original base	突变碱基 Mutant bases	位点 Site	原氨基酸 Original amino acid	突变氨基酸 Mutant amino acids	
GmAGO4A	1	516	T	A	172	S	R	9
	2	516	T	A	172	S	R	211
		976	T	C	326	F	L	
	3	976	T	C	326	F	L	5
GmAGO4B	1	523	G	A	175	A	T	22
GmAGO4C	1	277	C	T	93	R	C	42
	2	218	A	T	73	N	I	22
	3	122	A	T	41	N	I	21
	4	373	C	T	125	R	C	31
	5	91	G	A	31	E	K	5





A: GmAGO4A; B: GmAGO4A 突变类型 1; C: GmAGO4A 突变类型 2; D: GmAGO4A 突变类型 3; E: GmAGO4B; F: GmAGO4A 突变类型 1; G: GmAGO4C; H: GmAGO4C 突变类型 1; I: GmAGO4C 突变类型 2; J: GmAGO4C 突变类型 3; K: GmAGO4C 突变类型 4; L: GmAGO4C 突变类型 5。

A: GmAGO4A; B: Mutation type 1 of GmAGO4A; C: Mutation type 2 of GmAGO4A; D: Mutation type 3 of GmAGO4A; E: GmAGO4B; F: Mutation type 1 of GmAGO4A; G: GmAGO4C; H: Mutation type 1 of GmAGO4C; I: Mutation type 2 of GmAGO4C; J: Mutation type 3 of GmAGO4C; K: Mutation type 4 of GmAGO4C; L: Mutation type 5 of GmAGO4C.

图 8 大豆 AGO4 突变类型 3D 建模分析

Fig. 8 3D modeling analysis of soybean AGO4 mutation type

3 讨论

GmAGO4A、*GmAGO4B* 和 *GmAGO4C* 3 个基因与拟南芥 *AGO4* 有很高的同源性。*GmAGO4A*、*GmAGO4B* 和 *GmAGO4C* 有 5 个共同的结构域, DUF1785 常与 PIWI 共同存在, 具体功能未知; *ArgoN* 与 PIWI 结构域是 PAZ 存在的基础; PAZ 结构域位于 3 个基因的中间区域, 具有结合双链 siRNA 的能力, 可能有助于 siRNA 和 miRNAs 特异性进入 RNAi 途径^[29]; PIWI 结构域首先在 PIWI 蛋白和大量相关核酸结合蛋白中发现, 可结合和切割 RNA^[30]。由此可见, *GmAGO4A*、*GmAGO4B* 和 *GmAGO4C* 都具有 AGO4 蛋白行使功能的关键元件, 但是 *GmAGO4C* 的 *ArgoMid* 结构域在 *GmAGO4A* 和 *GmAGO4B* 中没有, *GmAGO4C* 与拟南芥 *AGO4* 结构更相似。蛋白三级结构模型分析结果表明, *GmAGO4A* 和 *GmAGO4B* 为同一种蛋白三级结构类型, *GmAGO4C* 为另外一种结构类型, 这两种蛋白结构类型都与拟南芥 *AGO4* 的结构有差异, 但是 *GmAGO4C* 的结构与 *AtAGO4* 的结构更相似。所以这两方面结果表明 *GmAGO4C* 更可能是大豆中真正的有功能的 *AGO4* 基因。但是同源比对结果显示 *GmAGO4A* 和 *GmAGO4B* 与拟南芥 *AGO4* 相似度更高, 此结果更倾向于 *GmAGO4A* 和 *GmAGO4B* 为大豆中行使 AGO4 功

能的蛋白。关于这个矛盾本研究现无法给出合理的解释, 然而基于蛋白行使特定功能严重依赖蛋白的三维结构这一被广泛认同的观点, 本研究更倾向于 *GmAGO4C* 为大豆中行使 AGO4 功能的蛋白。本研究结果表明 *GmAGO4A*、*GmAGO4B* 与 *GmAGO4C* 是功能不同的两类蛋白, 但要想确定这 3 个基因是否能在大豆中发挥作用以及发挥什么样的作用还需要进一步研究。

大豆 *AGO4* 多态性分析的结果显示: 302 个大豆野生品种中 3 个基因的外显子都产生了非同义突变, *GmAGO4A* 有 3 种类型的突变, 有 211 种野生大豆发生了同一类突变, 该突变类型会导致其蛋白三级结构明显的变化, 此种普遍突变类型的具体原因需要进一步研究; *GmAGO4C* 有 5 种突变类型, 其中 1 种突变类型会导致蛋白三级结构明显变化; *GmAGO4B* 只有 1 个突变类型, 且其蛋白三级结构变化不大。对这些不同突变类型野生大豆的进一步研究将有助于对大豆 AGO4 的深入研究。

4 结论

本研究利用多种生物信息学的方法找到并命名大豆的 3 个 *GmAGO4* 基因, 并且对其进化地位、基本性质、亚细胞定位、磷酸化位点、二级结构、三级结构进行生物信息学的分析, 预测 *GmAGO4A*、

GmAGO4B 与 GmAGO4C 是两类不同功能的蛋白, 且 GmAGO4C 更有可能是有 AGO4 功能的蛋白。在 302 种野生大豆品系信息的本地数据库中对大豆 *AGO4* 进行多态性分析, 找出由于突变导致 GmAGO4 的三级结构发生变化进而导致蛋白功能可能发生变化的大豆品系。本研究结果对大豆中 *GmAGO4* 基因的进一步研究及大豆 DNA 甲基化的研究具有一定的指导意义。

参考文献

- [1] Song Q, Lu X, Li Q, et al. Genome-wide analysis of DNA methylation in soybean[J]. *Molecular Plant*, 2013, 6(6):1961-1974.
- [2] Rodríguez-Leal D, Castillo-Cobián A, Rodríguez-Arévalo I, et al. A primary sequence analysis of the Argonaute protein family in plants[J]. *Frontiers in Plant Science*, 2016, 7:1347.
- [3] Cho Y B, Jones S I, Vodkin L O. Mutations in Argonaute5 illuminate epistatic interactions of the K1 and I loci leading to saddle seed color patterns in *Glycine max*[J]. *Plant Cell*, 2017, 29(4):708-725.
- [4] Bender J. DNA methylation and epigenetics[J]. *Annual Review of Plant Biology*, 2004, 55(1):41-68.
- [5] Zhong X. Comparative epigenomics: A powerful tool to understand the evolution of DNA methylation[J]. *New Phytologist*, 2016, 210(1):76-80.
- [6] Movahedi A, Sun W, Zhang J, et al. RNA-directed DNA methylation in plants [J]. *Plant Cell Reports*, 2015, 34(11):1857-1862.
- [7] Pck A, Dennis E S, Wang M B. Analysis of Argonaute 4-associated long non-coding RNA in *Arabidopsis thaliana* sheds novel insights into gene regulation through RNA-directed DNA methylation [J]. *Genes*, 2017, 8(8):198.
- [8] Wang F, Axtell M J. AGO4 is specifically required for heterochromatic siRNA accumulation at Pol V-dependent loci in *Arabidopsis thaliana*[J]. *Plant Journal*, 2017, 90(1):37-47.
- [9] Lahmy S, Pontier D, Bies-Etheve N, et al. Evidence for Argonaute4-DNA interactions in RNA-directed DNA methylation in plants[J]. *Genes Development*, 2016, 30(23):2565-2570.
- [10] Xie M, Yu B. siRNA-directed DNA methylation in plants[J]. *Current Genomics*, 2015, 16(1):23-31.
- [11] Rodríguez-Leal D, Castillo-Cobián A, Rodríguez-Arévalo I, et al. A primary sequence analysis of the ARGONAUTE protein family in plants[J]. *Frontiers in Plant Science*, 2016, 7:1347.
- [12] Jacobsen S E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals[J]. *Nature Reviews Genetics*, 2010, 11(3):204-220.
- [13] Zilberman D, Cao X, Jacobsen S E. *Argonaute4* control of locus-specific siRNA accumulation and DNA and histone methylation [J]. *Science*, 2003, 299(5607):716-719.
- [14] Pck A, Dennis E S, Wang M B. Analysis of Argonaute 4-associated long non-coding RNA in *Arabidopsis thaliana* sheds novel insights into gene regulation through RNA-directed DNA methylation [J]. *Genes*, 2017, 8(8):198.
- [15] Liu W, Duttke S H, Hetzel J, et al. RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in *Arabidopsis*[J]. *Nature Plants*, 2018, 4(3):181-188.
- [16] Parida A P, Sharma A, Sharma A K. AtMBD6, a methyl CpG binding domain protein, maintains gene silencing in *Arabidopsis* by interacting with RNA binding proteins[J]. *Journal of Bioscience*, 2017, 42(1):57-68.
- [17] Córdoba-Cañero D, Cognat V, Ariza R R, et al. Dual control of ROS1-mediated active DNA demethylation by DNA damage-binding protein 2 (DDB2)[J]. *Plant Journal*, 2017, 92(6):1181.
- [18] Hernándezlagana E, Rodríguezleal D, Lúa J, et al. A multigenic network of ARGONAUTE4 clade members controls early megaspore formation in *Arabidopsis* [J]. *Genetics*, 2016, 204(3):1045-1056.
- [19] Oliver C, Santos J L, Pradillo M. Accurate chromosome segregation at first meiotic division requires AGO4, a protein involved in RNA-dependent DNA methylation in *Arabidopsis thaliana*[J]. *Genetics*, 2016, 204(2):543-553.
- [20] Kong W, Li B, Wang Q, et al. Analysis of the DNA methylation patterns and transcriptional regulation of the NB-LRR-encoding gene family in *Arabidopsis thaliana*[J]. *Plant Molecular Biology*, 2018, 96(6):563-575.
- [21] Cao J Y, Xu Y P, Li W, et al. Genome-wide identification of Dicer-like, Argonaute, and RNA-dependent RNA polymerase gene families in brassic species and functional analyses of their *Arabidopsis* homologs in resistance to *sclerotinia sclerotiorum*[J]. *Frontiers in Plant Science*, 2016, 7:1614.
- [22] Li S, Xia Q, Wang F, et al. Laser irradiation-induced DNA methylation changes are heritable and accompanied with transpositional activation of mPing in rice[J]. *Frontiers in Plant Science*, 2017, 8:363.
- [23] Schalk C, Molinier J. Global genome repair factors control DNA methylation patterns in *Arabidopsis*[J]. *Plant Signaling & Behavior*, 2016, 11(12):e1253648.
- [24] Schalk C, Drevenssek S, Kramdi A, et al. DNA DAMAGE BINDING PROTEIN 2 (DDB2) shapes the DNA methylation landscape [J]. *Plant Cell*, 2016, 28(9):2043-2059.
- [25] Jin X, Guo X, Zhu D, et al. miRNA profiling in the mice in response to *Echinococcus multilocularis* infection[J]. *Acta Tropica*, 2017, 166:39-44.
- [26] Brosseau C, El O M, Adurogbangba A, et al. Antiviral defense involves AGO4 in an *Arabidopsis*-potyvirus interaction[J]. *Molecular plant-microbe interactions*, 2016, 29(11):878-888.
- [27] Hamera S, Yan Y, Song X, et al. Expression of Cucumber mosaic virus suppressor 2b alters FWA methylation and its siRNA accumulation in *Arabidopsis thaliana* [J]. *Biology Open*, 2016, 5(11):1727-1734.
- [28] 刘宇原, 付爱根, 徐敏. 大豆转座子的研究现状及应用前景[J]. *大豆科学*, 2016, 35(3):512-518. (Liu Y Y, Fu A G, Xu M. Research status and application prospects of soybean transposon[J]. *Soybean Science*, 2016, 35(3):512-518.)
- [29] Song J J, Liu J, Tolia N H, et al. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes [J]. *Nature Structural Biology*, 2003, 10(12):1026-1032.
- [30] Cerutti L, Mian N, Bateman A. Domains in gene silencing and cell differentiation proteins: The novel PAZ domain and redefinition of the Piwi domain [J]. *Trends in Biochemical Sciences*, 2000, 25(10):481.