

不同植物防御素的生物信息学分析

王汉屏*

陕西教育学院生命科学系, 西安 710061

提要: 采用生物信息学的方法和工具对已在 GenBank 上注册的花生、鹰嘴豆、合欢、车前草、沙梨、辣椒等植物的防御素的核酸和氨基酸序列作了分析, 并对其组成成分、翻译后修饰、导肽、跨膜结构域、疏水性/亲水性、蛋白质二级结构和功能结构域进行预测和推断的结果表明: 此类蛋白可能是一具有信号肽的分泌型蛋白, α -螺旋和不规则卷曲是蛋白质二级结构中量最大的结构元件, β -转角和延伸链散布于整个蛋白质中, 包含一个 gamma-thionin 功能结构域。

关键词: 植物防御素; 生物信息学

Bioinformatics Analysis of Defensins in Different Plants

WANG Han-Ping*

Department of Life Sciences, Shannxi College of Education, Xi'an 710061, China

Abstract: By bioinformatic methods and tools, defensin genes as well as the composition of amino acid sequences, post-translational modification, hydrophobicity or hydrophilicity, transmembrane ability, subcellular localization, secondary structure and functional domains of sucrose synthase in *Arachis diogeni*, *Cicer arietinum*, *Tephrosia platycarpa*, *Plantago major*, *Pyrus pyrifolia*, and *Capsicum annuum* were analyzed and predicted. The results demonstrated that defensins is a sort of secretory proteins containing signal peptide. The main motifs of predicted secondary structure of defensin are alpha helix and random coil, beta turn and extended strand are spreaded in the whole secondary structure. Defensins contain a functional domain—gamma-thionin.

Key words: phytodefensin; bioinformatics

植物具有内在的防御系统, 在外部病原菌感染之下, 自身会产生一系列拮抗物质, 如一些低分子量的次生代谢产物橡胶素(dambonite)、植保素(phytoalexin)、凤仙花素(anthocyanin)等, 以阻止病原菌的繁殖扩展和进一步侵入寄主植物(Berrocal-Lobo 等 2002)。植物防御素正是这类低分子量短肽中的一个家族, 它广泛分布于植物体内, 其可以拮抗细菌和真菌等多种植物病原菌, 在植物中是极为重要的一种抗病因子, 参与植物的防卫反应。

植物防御素是一类分子量小(约 5 kDa, 由 45~54 个氨基酸组成)、碱性、富含半胱氨酸、具复杂三维结构的短肽(廖乾生等 2003)。在微摩尔浓度下就能表现出广谱的抗菌活性, 但对人、畜、植物细胞无害。有的植物防御素是植物的内源性产物, 也有的是在病原物感染后诱导其基因表达而产生的。在双子叶植物和单子叶植物中均发现有植物防御素的存在, 目前已经在苜蓿、绿豆、菠菜、白菜、向日葵、欧洲云杉、甘蓝型油菜、

小麦、辣椒、沙梨、烟草、白芥等植物中克隆到(Hanks 等 2005; 缪建锟等 2006; Seggura 等 1998; Park 等 2002; Urdangarin 等 2000; Sharma 和 Lönneborg 1996; 刘胜毅等 2004; Koike 等 2002; Do 等 2004; Zhou 等 2003; Li 和 Gray 1999; Volpicella 等 2000)。防御素主要作用于病原微生物的细胞膜, 致使病菌对该物质不易产生抗性, 因此, 它有着其他拮抗多肽所无可比拟的优势(张宏等 2006)。植物防御素基因工程的研究取得一定进展, 人们已成功得到一些经济作物和农作物的防御素转基因植物, 并证明它们能提高转基因植物的抗病能力, 如烟草、马铃薯、水稻、拟南芥和哈密瓜等(Rancois 等 2002; Gao 等 2000; Mitter 等 1998; 王轶等 2005)。有人还把防御素基因转

收稿 2007-06-22 修订 2007-10-16

资助 陕西教育学院重点科研项目(06KJ013)。

* E-mail: hanpingwang@126.com; Tel: 029-85370063

入番茄、毛白杨、小麦、菊花和棉花等植物中以期获得抗病强的优良品种(余兴邦等 2006)。

生物信息学是一门由生物、数学、计算机交叉的学科。人类基因组计划完成后, 又进入后基因组时代, 并已从基因组测序和表达分析, 转向对蛋白质组结构与功能的预测和分析。采用各类分子生物学数据库和分析软件对已知的核酸和蛋白质序列进行比对、分析、建立计算模型, 从而对其结构、功能进行推断及预测, 已成为生物信息学研究的一种趋势。

本文用生物信息学的方法, 以花生(*Arachis diogeni*)为研究重点, 对鹰嘴豆(*Cicer arietinum*)、合欢(*Tephrosia platycarpa*)、车前草(*Plantago major*)、沙梨(*Pyrus pyrifolia*)和辣椒(*Capsicum annuum*)等植物的防御素基因及相应氨基酸序列的理化性质、结构特征、功能及系统演化关系等进行预测和分析, 以期能为它们的深入研究提供参考。以往的防御素研究多为实验性的研究, 而用已有的生物信息进行研究还少见报道, 为此, 本文对此进行了尝试。

材料与方 法

数据资料来源于 National Center for Biotechnology Information (NCBI)核酸及蛋白质数据库中已注册的防御素的核苷酸序列及其对应的氨基酸序

列: 花生(A Y288448、A A P92330)、鹰嘴豆(A Y206395、A A O38756)、合欢(A Y907349、A A X86993)、车前草(A J843264、C A H58740)、沙梨(A B052688、B A B64930)、辣椒(A F442388、A A L35366)。

依据 DNAS tar 软件和 <http://www.ncbi.nlm.nih.gov/>、<http://www.ebi.ac.uk/>、<http://www.cbs.dtu.dk/>、<http://www.expasy.org/> 等网站提供的各类生物信息学软件进行在线分析。核酸及氨基酸序列的组成成分分析、理化性质分析、开放阅读框(open reading frame, ORF)的查找和翻译均用 DNAS tar 软件进行; 核苷酸和氨基酸序列的同源性比对用 Blast 在线工具完成; 氨基酸翻译后修饰、跨膜结构域和亲水性/疏水性的分析用在线工具 NetPhos 2.0 Server、TargetP 1.1 Server、TMHMM 和 ProtScale 完成; 蛋白质的二级结构和三级结构用 SOPMA 和 SWISS-MODEL 在线工具预测。

结果与讨论

1 核苷酸序列及其推导的核苷酸序列的组成成分和理化性质

用 DNAS tar 软件的 ORF Finder、Editseq、Protean 程序分析 NCBI 上注册的花生、鹰嘴豆、合欢、车前草、沙梨、辣椒防御素的核苷酸序列及其推导的氨基酸序列的结果(表 1)表明, ORF 为

表 1 不同植物防御素基因的核苷酸序列以及推导的氨基酸序列组成成分和理化性质分析

Table 1 Analysis of composition and physical and chemical characters of nucleic acid sequence and deduced amino acid sequence of defensin among different plants

植物	ORF/bp	终止 密码子	推导的 氨基酸残基数	分子量 /kDa	pI	氨基酸比例(%)					主要氨基酸
						碱性 AA	酸性 AA	亲水 AA	疏水 AA	带电 AA	
花生	225	TAA	74	8.45486	7.89	13.51	10.81	31.08	35.14	37.84	Cys Leu Phe Thr Ala Glu Lys Arg
鹰嘴豆	219	TAA	72	9.2252	8.13	15.28	12.50	29.17	31.94	44.44	Cys Lys Leu Thr Phe Ala Glu Arg
合欢	228	TAA	75	8.55594	7.52	14.67	13.33	30.67	30.67	44.00	Cys Lys Leu Asp Arg Thr
车前草	219	TGA	72	7.94941	8.71	15.28	6.94	26.39	31.94	36.11	Cys Gly Arg Phe Leu Thr Val Ser
沙梨	264	TAA	87	9.51733	8.69	14.94	8.05	29.89	32.18	34.48	Cys Leu Trp Glu Lys Arg Gly Ser
辣椒	237	TAA	78	8.68831	8.35	12.82	7.69	28.21	30.77	33.33	Arg Cys Ala Met Ser Phe Gly Glu Thr Val

200 bp 左右, 分子量为 9 kDa 左右。理论等电点均高于 7, 说明防御素呈碱性。终止密码子大都用 TAA。在所分析的植物中, 只有车前草防御素基因采用 TAG 作密码子。在鹰嘴豆和车前草防御素氨基酸组成中, 碱性氨基酸的比例最高, 均为 15.28%; 在合欢防御素氨基酸中酸性氨基酸的比例最高(13.33%); 在花生防御素总氨基酸中, 亲水性氨基酸和疏水性氨基酸的比例均最高, 分别为 31.08% 和 35.14%。Cys、Leu 和 Arg 是这些植物共有的主要氨基酸。

2 氨基酸序列比对分析

用 Clustal W 在线程序对拟南芥、甘蓝、银杏、小麦、花生、鹰嘴豆、合欢、车前草、沙梨、辣椒的防御素氨基酸序列进行多序列比对

的结果(图 1)表明, 拟南芥、甘蓝、银杏、小麦、花生、鹰嘴豆、合欢、车前草、沙梨、辣椒的防御素氨基酸序列的 8 个半胱氨酸是完全保守的, 在相当多的一些植物防御素中 Ser、Gly 和 Glu 氨基酸残基也是非常保守的氨基酸位点 (Uradangarin 等 2000)。

3 翻译后磷酸化修饰的预测和分析

多肽链在核糖体上合成释放后, 通常要经过翻译后修饰如糖基化、甲基化、磷酸化等才能正确折叠形成有效的三维构象, 并运输到特定场所发挥功能。翻译后修饰的预测和分析, 对正确认识和理解蛋白质功能构象的形成很重要。根据 NetPhos 2.0 Server 软件(Blom 等 2004), 磷酸化分值超过阈值的氨基酸位点被认为是可能的磷酸化作

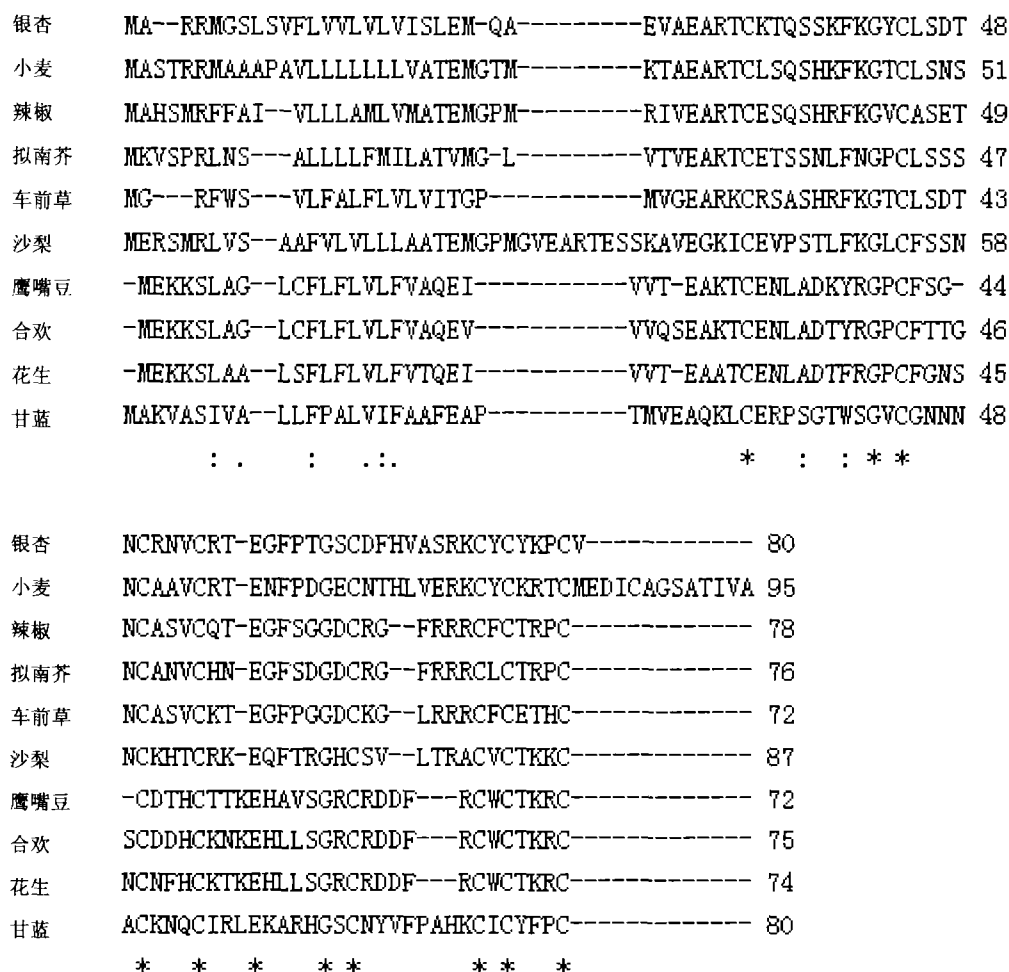


图 1 不同植物防御素氨基酸序列多重比对

Fig.1 Multiple alignment of amino acid sequences among different plant defensins
符号“*”、“:”、“.”分别表示氨基酸完全相同、保守和半保守 3 种情况。

用位点, 可知花生的第59个氨基酸位点上发生Ser的磷酸化(图2), 第36和71个氨基酸位点上发生Thr的磷酸化; 鹰嘴豆在第43和57个氨基酸位点发生Ser的磷酸化, 第47、50和69个氨基酸位点上发生Thr的磷酸化; 合欢的第60个氨基酸位点发生Ser的磷酸化, 第37和72个氨基酸位点上发生Thr的磷酸化; 车前草的第32、47个氨基酸发生Ser的磷酸化; 沙梨的第34个氨基酸位点发生Ser的磷酸化; 辣椒的38和61个氨基酸位点上发生Ser的磷酸化。用NetPhosK 1.0 Server软件分析这几种植物防御素氨基酸序列的翻译后激酶磷酸化修饰, 观察到这几种植物的防御素氨基酸序列均存在蛋白激酶C (PKC)的Ser磷酸化位点。

4 导肽的预测和分析

在核糖体上合成的蛋白质或者留在细胞质中, 或者运送到其他细胞器中, 并形成有功能的构象后才能发挥其功能。导肽是一段引导新合成的肽链进入细胞器的识别序列。因此, 预测和分析导肽, 对正确认识蛋白质的亚细胞定位和功能作用的途径及机制有一定的意义。为此, 用TargetP 1.1 Server预测花生、鹰嘴豆、合欢、车前草、沙梨和辣椒防御素氨基酸序列的导肽(Emanuelsson等2000)的结果表明, 这些植物的防御素均含有分泌途径信号肽, 在花生、鹰嘴豆、合欢、车前草、沙梨和辣椒防御素氨基酸序列中的信号肽分值分别为0.995、0.994、0.996、0.826、0.963和0.940, 预测的可靠性等级均为

1。据此认为防御素蛋白均属于分泌型蛋白, 在核糖体合成后, 由信号肽引导其前体蛋白到达指定部位, 切除信号肽而成为成熟蛋白质发挥其抗菌功能。

5 信号肽的预测和分析

分泌蛋白及细胞膜蛋白均以前体物质多肽的形式合成, 其N末端含有作为通过膜时之信号的氨基酸序列, 这种氨基酸序列称信号肽或信号序列。由15~25个氨基酸所组成。用SignalP 3.0 Server分析这几种植物防御素氨基酸序列的信号肽的存在位置及序列(Bendtsen等2004), 结果表明不同植物的氨基酸序列上信号肽的剪切位点分值和位置分别为: 花生在第27~28个氨基酸之间(分值为0.675; TEA-AT) (图3); 鹰嘴豆为第27~28个氨基酸之间(分值为0.775; TEA-KT); 合欢在第28~29个氨基酸之间(分值为0.499; SEA-KT); 车前草在第25~26个氨基酸之间(分值为0.816; GEA-RK); 沙梨在第23~24个氨基酸之间(分值为0.354; TEM-GP); 辣椒在第20~21个氨基酸之间(分值为0.913; VMA-TE)。总之, 花生和鹰嘴豆信号肽序列为整体氨基酸序列的N端27个氨基酸, 合欢的为N端28个氨基酸, 车前草信号肽序列为N端25个氨基酸, 沙梨的为N端23个氨基酸, 辣椒的为N端20个氨基酸。因此, 可以认为植物防御素的基因在植物体内表达的是一个前体蛋白, 此蛋白通过翻译后加工并借助信号肽运送到胞质外的空间发挥植物防御素

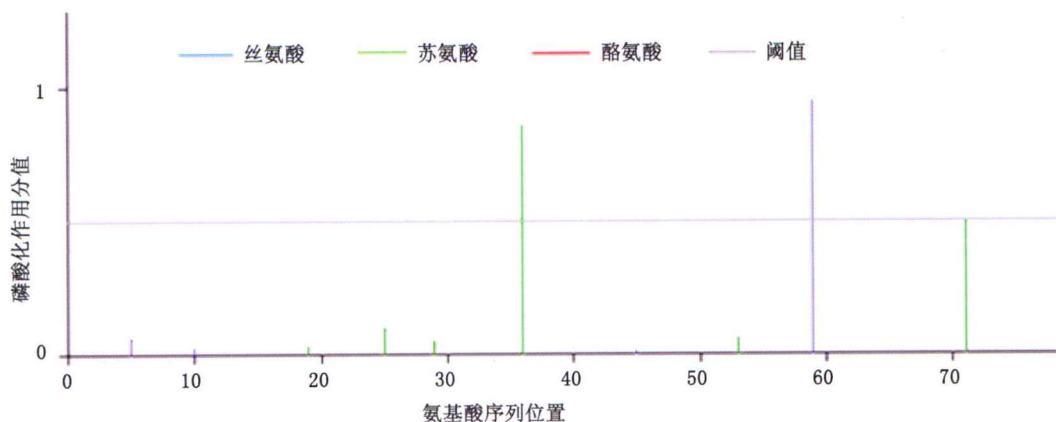


图2 花生防御素氨基酸序列翻译后的磷酸化修饰预测

Fig.2 Predicted phosphorylation sites in amino acid sequence of defensin in peanut

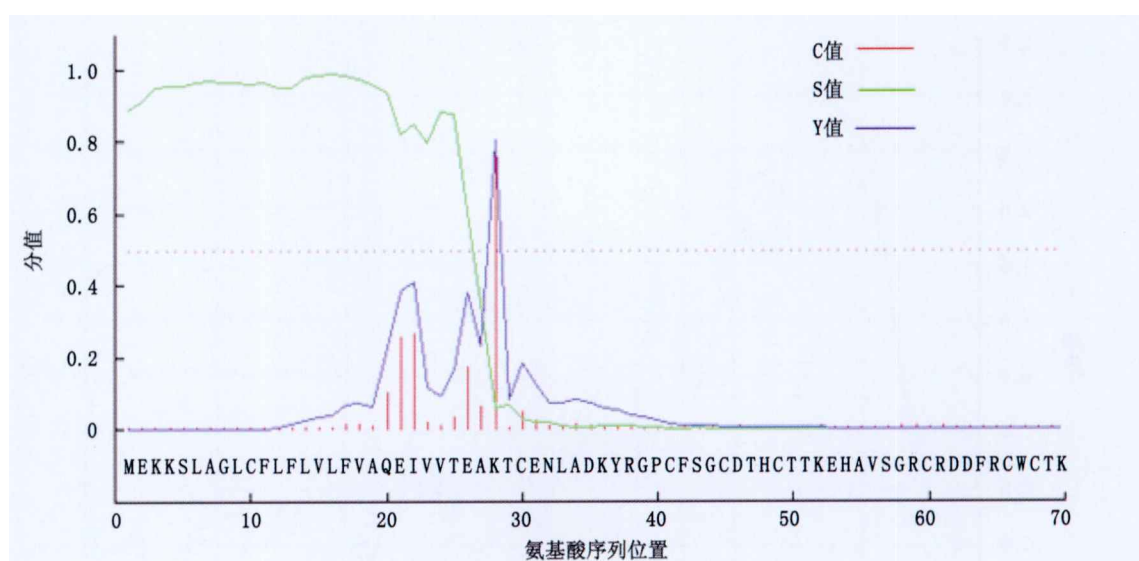


图3 花生防御素氨基酸序列信号肽的预测分析

Fig.3 Predicted signal peptide of defensin in peanut

C 值: 原始剪切位点的分值; S 值: 信号肽的分值; Y 值: 综合剪切位点的分值。

的抗菌功能。

将信号肽按预测位置切除后分析这些植物防御素编码序列的组分和性质的结果(表2)表明, 切除信号肽后的防御素氨基酸序列分子量(5 kDa左右)比切除前(9 kDa左右)小很多(3~4 kDa)(张宏等2006)。

6 不同植物防御素氨基酸序列疏水性/亲水性的预测和分析

用ProtScale分析花生、鹰嘴豆、合欢、车前草、沙梨、辣椒防御素氨基酸序列的疏水性/亲水性(Kyte和Doolittle 1982)的结果表明, 预测结果均与花生一致(图4), 整个多肽链N端20个左

表2 去除信号肽后的防御素氨基酸序列组分

Table 2 Analysis of amino acid sequence composition of defensin after signal peptide were cut off

植物	信号肽序列	成熟肽序列 氨基酸数	分子量/kDa	pI	氨基酸比例/%				
					碱性AA	酸性AA	亲水AA	疏水AA	带电AA
花生	MEKKS LAGLCFLFLVLFV TQE- IVVTEA	47	5.44522	8.14	17.02	10.64	38.30	21.28	48.94
鹰嘴豆	MEKKS LAGLCFLFLVLFV AQE- IVVTEA	45	5.1889	8.13	20.00	13.33	37.78	15.56	57.78
合欢	MEKKS LAGLCFLFLVLFV AQE- VVVQSEA	47	5.4742	7.89	19.15	14.89	38.30	14.89	57.45
车前草	MGRFWSVLFA- LFLVLVITGPM- VGEA	47	5.21302	8.71	21.28	8.51	36.17	17.02	51.06
沙梨	MERSMRLVSA- AFVLVLLLA- TEM	64	6.98317	8.69	17.19	7.81	35.94	23.44	40.62
辣椒	MAHSMRFFAI- VLLLAMLVMA	58	6.44037	8.14	15.02	10.34	36.21	18.97	41.38

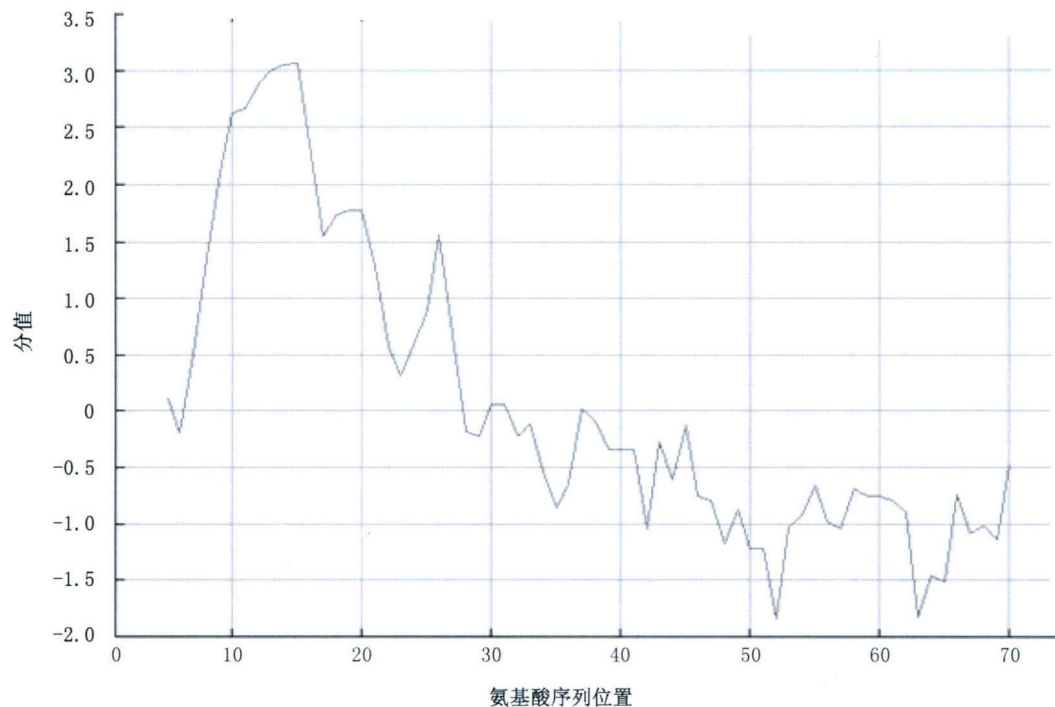


图4 花生防御素氨基酸序列疏水性/亲水性的预测

Fig.4 Predicted hydrophobicity/hydrophilicity for the defensin in peanut

右氨基酸形成一个明显的疏水区域,即为疏水的头部,第30个氨基酸后均表现为亲水性,结合上述分析可知组成疏水性头部的氨基酸均是组成信号肽的氨基酸,这种疏水性的氨基酸组成对信号肽引导防御素前体蛋白与膜脂结合通过膜是有利的。

7 不同植物防御素跨膜结构域的预测和分析

跨膜结构域是膜中蛋白与膜脂结合的主要部位,一般由20个左右的疏水氨基酸残基组成,形成 α 螺旋,与膜脂相结合。预测和分析跨膜结构域,对认识蛋白质的结构、功能、分类以及在细胞中的作用部位均有一定的意义。用TMHMM 2.0 Server 预测(Ikeda 等2002)花生的可能跨膜结构域的结果表明,花生防御素在第7~29个氨基酸(图5),鹰嘴豆在第5~24个氨基酸、合欢在第5~24个氨基酸、车前草在第4~26个氨基酸、沙梨在第7~26个氨基酸、辣椒在第7~29个氨基酸形成跨膜结构域,结合信号肽的预测分析结果,说明跨膜结构域均在信号肽位置。

8 二级结构的预测和分析

多肽链借助氢键排列成沿一维方向而呈现有规则的重复构象的二级结构,是氨基酸顺序与三

维构象之间的桥梁。二级结构借助范德华力、氢键、静电和疏水等相互作用形成蛋白质的三级结构,从而发挥正常的生物学功能。据此,用SOPMA对花生、鹰嘴豆、合欢、车前草、沙梨、辣椒防御素氨基酸序列的二维结构进行预测(Geourjon 和 Deleage 1995)的结果表明,预测结果均与花生一致(图6)。 α -螺旋和不规则卷曲是防御素整体结构中的主要组成结构元件, β -转角和延伸链散布于整个蛋白质中。

9 功能结构域分析

蛋白质由多个模体组成,组成这些模体的氨基酸区段行使特异的功能,同时还蕴含着各自的遗传进化信息。利用NCBI链接的CDD分析花生、鹰嘴豆、合欢、车前草、沙梨、辣椒防御素氨基酸序列的功能结构域,得到了相一致的结果(图7),即切掉信号肽,剩下的组成成熟肽的氨基酸序列均包含一个起杀菌作用的毒素 γ -thionin功能域,属于 γ -thionin家族蛋白。此功能域是防御素发挥其功能的单位。

上述结论虽属推断和预测,但为进一步研究防御素打下了基础。

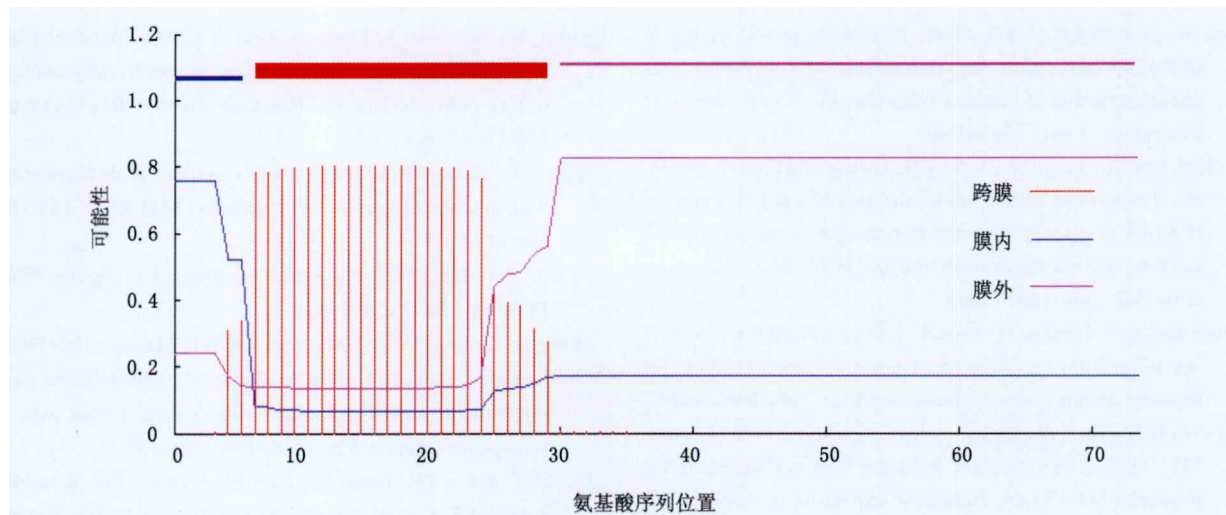


图5 TMHMM 2.0 Server 对花生防御素氨基酸序列跨膜结构域的预测分析

Fig.5 Predicted transmembrane segments for defensin in peanut

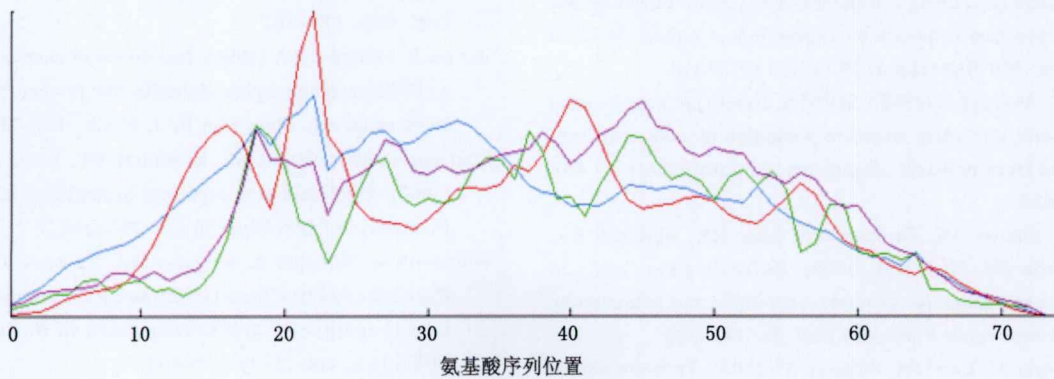


图6 花生防御素二级结构的预测

Fig.6 Secondary structure analysis of defensin in peanut

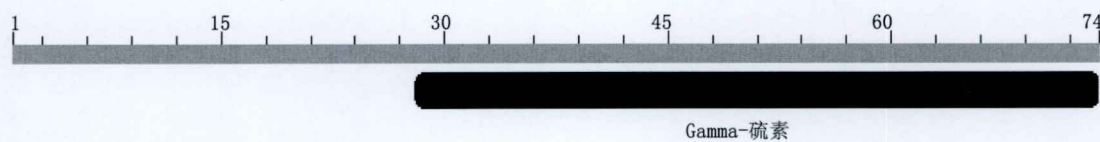


图7 花生防御素氨基酸序列功能域分析

Fig.7 Function domain analysis of defensin in peanut

参考文献

- 廖乾生, 林福呈, 李德葆(2003). 植物防御素及其研究进展. 浙江大学学报(农业与生命科学版), 29 (1): 113~118
- 刘胜毅, Fitt B, 刘仁虎, Evans N, 董彩华, 黄永菊(2004). 油菜防御素和草酸氧化酶基因的克隆与诱导表达水平研究. 中国油料作物学报, 26 (3): 43~49
- 缪建锟, 孙黎, 彭晓明, 祝建波(2006). 绿豆防御素基因的克隆、序列分析和植物表达载体的构建. 石河子大学学报, 24 (1): 112~115
- 王轶, 罗勤, 吴俊, 郭经宇, 杨毅, 李旭锋(2005). 甘蓝型油菜植物防御素基因的克隆与原核表达. 四川大学学报(自然科学版), 42 (4): 831~834
- 余兴邦, 郭锁链, 乌翠兰, 李大竣, 呼格吉勒图, 鄂优(2006). 防御素研究进展. 动物医学进展, 27 (8): 47~51
- 张宏, 胡春香, 张德禄, 刘永定(2006). 植物防御素研究进展. 西北师范大学学报(自然科学版), 42 (5): 112~117
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004). Improved prediction of signal peptides: Signal P 3.0. J Mol Biol, 340: 783~795
- Berrocal-Lobo M, Segura A, Moreno M, López G, García-Olmedo F, Molina A (2002). Snakin-2, a peptide from potato whose gene is locally induced by wounding and responds to pathogen infection. Plant Physiol, 128 (3): 951~961

- Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomic*, 4 (6): 1633~1649
- Do HM, Lee SC, Jung HW, Sohn KH, Hwang BK (2004). Differential expression and *in situ* localization of a pepper defensin (CADEF1) gene in response to pathogen infection, abiotic elicitors and environmental stresses in *Capsicum annuum*. *Plant Sci*, 166: 1297~1305
- Emanuelsson O, Nielsen H, Brunak S, Heijne G (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Mol Biol*, 300: 1005~1016
- Francois IEJA, De Bolle MFC, Dwyer G, Goderis IJWM, Wouters PFJ, Verhaert PD, Proost P, Schaaper WMM, Cammue BPA, Broekaert WF (2002). Transgenic expression in *Arabidopsis* of a polyprotein construct leading to production of two different antimicrobial proteins. *Plant Physiol*, 128 (4): 1346~1358
- Gao AG, Hakimi SM, Mittanck CA, Wu Y, Woerner BM, Stark DM, Shah DM, Liang J, Rommens CM (2000). Fungal pathogen protection in potato by expression of a plant defensin peptide. *Nat Biotechnol*, 18 (12): 1307~1310
- Geourjon C, Deleage G (1995). SOPMA: significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, 11 (6): 681~684
- Hanks JN, Snyder AK, Graham MA, Shah RK, Blaylock LA, Harrison MJ, Shah DM (2005). Defensin gene family in *Medicago truncatula*: structure, expression and induction by signal molecules. *Plant Mol Biol*, 58: 385~399
- Ikeda M, Arai M, Lao DM, Shimizu T (2002). Transmembrane topology prediction methods: a reassessment and improvement by a consensus method using a dataset of experimentally characterized transmembrane topologies. *Silico Biol*, 2 (1): 19~33
- Koike M, Okamoto T, Tsuda S, Imai R (2002). A novel plant defensin-like gene of winter wheat is specifically induced during cold acclimation. *Biochem Biophys Res Commun*, 298 (1): 46~53
- Kyte J, Doolittle RF (1982). A simple method for displaying the hydropathic character of a protein. *Mol Biol*, 157 (6): 105~132
- Li HY, Gray JE (1999). The electronic plant gene register. *Plant Physiol*, 120 (2): 633~635
- Mitter N, Kazan K, Way HM, Broekaert WF, Manners JM (1998). Systemic induction of an *Arabidopsis* plant defensin gene promoter by tobacco mosaic virus and jasmonic acid in transgenic tobacco. *Plant Sci*, 136: 169~180
- Park HC, Kang YH, Chun HJ, Koo JC, Cheong YH, Kim MC, Chung WS, Kim JC, Yoo JH, Young YC et al (2002). Characterization of a stamen-specific cDNA encoding a novel plant defensin in Chinese cabbage. *Plant Mol Biol*, 50: 59~69
- Seggura A, Moreno M, Molina A, Garcia-Olmedo F (1998). Novel defensin subfamily from spinach (*Spinacia oleracea*). *FEBS Lett*, 435: 159~162
- Sharma P, Lönneborg A (1996). Isolation and characterization of a cDNA encoding a plant defensin-like protein from roots of Norway spruce. *Plant Mol Biol*, 31 (3): 707~712
- Uradangarin MC, Norero NS, Broekaert WF, Laura de la Canal (2000). A defensin gene expressed in sunflower inflorescence. *Plant Physiol Biochem*, 38 (3): 253~258
- Volpicella M, Schipper A, Jongsma MA, Spoto N, Gallerani R, Ceci LR (2000). Characterization of recombinant mustard trypsin inhibitor 2 (MTI2) expressed in *Pichia pastoris*. *FEBS Lett*, 468 (2~3): 137~141
- Zhou YG, Norioka N, Li SL, Norioka S (2003). Nucleotide sequence of *pdn3* gene encoding a plant defensin-like protein in pollen grains of Japanese pear (*Pyrus pyrifolia*). *J Plant Physiol Mol Biol*, 29: 360