

大豆抗胞囊线虫基因 *Rhg1* 和 *Rhg4* 产物基本性质分析与结构预测

高亚梅¹, 韩毅强¹, 杜吉到², 刘金文¹

(1. 黑龙江八一农垦大学生命科学技术学院, 大庆 163319; 2. 黑龙江八一农垦大学植物科学技术学院, 大庆 163319)

摘要 抗胞囊线虫病基因是大豆抗病育种的重要基因资源。目前已经在抗胞囊线虫品种中发现了两个对于抗性具有重要贡献的 *Rhg1* 和 *Rhg4* 基因, 但是关于蛋白质结构和功能的信息还很少。本研究注释了 RHG1 和 RHG4 完整的结构域特征, 二级结构分析显示两蛋白质均采用 α/β 折叠, 蛋白激酶结构域的空间结构采用与所有真核生物的丝氨酸/苏氨酸及酪氨酸特异性蛋白激酶的核心相同的折叠方式, 具有类似结构特点。

关键词 胞囊线虫; *Rhg1*; *Rhg4*; 结构域; 结构预测

中图分类号 S565.1 文献标识码 A 文章编号 1000-9841(2007)05-0663-06

PRIMARY STRUCTURE ANALYSIS AND PREDICTION OF *rhg1* AND *rhg4* GENE PRODUCTS FOR CYST NEMATODE RESISTANCE IN SOYBEAN

GAO Ya-mei¹, HAN Yi-qiang¹, DU Ji-dao², LIU Jin-wen¹

(1. College of Life Science and Technology, Heilongjiang August First Land Reclamation University, Daqing 163319; 2. College of Plant Science and Technology, Heilongjiang August First Land Reclamation University, Daqing, 163319)

Abstract Disease resistant gene to soybean cyst nematode is important gene resource for soybean breeding. Two genes *Rhg1* and *Rhg4* related to cyst nematode resistance in soybean had been cloned in disease resistant cultivar, but the structure and role of RHG protein are not well understood. This paper explained the detailed domain feature of RHG protein, secondary structure analysis indicated RHG protein has an alpha beta fold, tertiary structure of protein kinase domain adopted the same fold as core structure in almost all eukaryotic serine/threonine and tyrosine kinase, and had similar structure feature according to structure prediction.

Key words *Heterodera glycines Ichinohe*; *Rhg1*; *Rhg4*; Domain; Structure prediction

大豆胞囊线虫(soybean cyst nematode, SCN)病是大豆生产上流行的一种重要毁灭性病害, 给大豆生产造成了严重的损失。国内外的研究和实践证明, 选育优良的抗病品种是控制大豆胞囊线虫病最经济、有效

的措施。目前, 借助分子标记构建遗传图谱来定位抗 SCN 病基因已经取得了很大的进展。2 个重要的基因位点 *Rhg1* 和 *Rhg4* 的定位已经在多个群体中得到验证, 美国南伊利诺大学(Southern Illinois University)的

D. A. Lihgtfoot 博士的实验室已从 Forrest 的 BAC 克隆 73 P6 和 100B10 中,获得了 *Rhg1* 和 *Rhg4* 的全长 cDNA 序列,在公共核酸数据库 GeneBank 的登录号分别为 AF506517 和 AF506518^[1]。*Rhg1* 和 *Rhg4* 基因产物具有类似的结构,均具有三个重要的结构域:细胞外的亮氨酸富集区(extracellular Leucine rich repeat,LRR)、跨膜结构域(transmembrane spanning domain,TM)、激酶结构域(kinase domain,PK)。根据其结构特征可以将 RHG1 和 RHG4 归为富含亮氨酸重复序列型类受体激酶(receptor-like protein kinase,RLK),与水稻抗白叶枯病抗性基因 *Xa21* 和拟南芥类受体激酶家族有很高的同源性,类受体蛋白激酶在植物抗逆性反应、信号传导和发育调控等过程发挥重要生物学功能。然而,对于 RHG1 和 RHG4 蛋白质及其结构了解还很少。作者利用生物信息学的方法和工具,由基因 *Rhgl* 和 *Rhg4* 产物的氨基酸序列分析蛋白质的基本性质、结构域,并预测其可能的空间结构,试图寻找其与抗病相关的结构和功能信息。

1 材料与方法

1.1 序列数据

基因 *Rhgl* 和 *Rhg4* 产物分别编码 855 和 893 个氨基酸,序列来自 Swiss-Prot 蛋白质序列数据库,登录号分别为:Q8L3Y5 和 Q8LKR3。

1.2 蛋白质特征分析与二级数据库搜索

1.2.1 编码蛋白质的理化性质分析 应用 <http://www.expasy.ch> 蛋白质分析专家系统的 ProtParam 在线分析软件分析该基因所编码蛋白质的氨基酸组成、等电点、电荷分布等理化性质^[2]。

1.2.2 蛋白质亲疏水性分析 跨膜区的预测分别利用 <http://www.expasy.ch> 蛋白质分析专家系统的 ProtScale 在线分析软件^[2]和 TMPRED 服务器(http://www.ch.embnet.org/software/TMPRED_form.html)完成。

1.2.3 结构域分析与二级数据库搜索 利用简单模块构架搜索工具(Simple Modular Architecture Research Tool,SMART)服务器(<http://smart.embl-heidelberg.de/>)分析序列中的结构域^[3];在 Pfam 数据库中搜索相应的结构模块,同时结合 Prosite 数据库中的 Motif 与 Profile 搜索分析序列中的功能位点。

1.3 编码蛋白二级结构预测

蛋白质二级结构的预测采用 PHD 程序在线服务器进行(<http://www.embl-heidelberg.de/predict-protein/predictprotein.html>)。

1.4 蛋白激酶结构域三维结构同源建模

利用三级结构同源建模在线服务器 SWISS - MODEL 对蛋白激酶结构域部分进行具体建模工作;三级结构观察软件为 PdbViewer 3.7 和 Raswin 软件;抓图软件 HyperSnap 6。

2 结果与分析

2.1 RHG1 和 RHG4 蛋白质理化性质

Rhgl 和 *Rhg4* 产物的基本理化性质分析结果显示,RHG1、RHG4 的理论等电点分别为 8.64 和 5.68,分子量 93157.9 道尔顿和 96564.4 道尔顿。稳定性系数表明 RHG1 蛋白质为稳定蛋白,而 RHG4 为不稳定蛋白。平均亲水性 RHG1 为 -0.025,RHG4 为 0.049。氨基酸组成分析表明在两蛋白质中亮氨酸(Leu,L)含量均较高,分别为 15.0%和 12.5%。另外,丝氨酸(Ser,S)含量比其在蛋白质中出现的频率高,见表 1。由 RHG1 和 RHG4 蛋白质序列发现,亮氨酸主要集中于蛋白质的 LRR 结构域,而大量的丝氨酸可能对于蛋白激酶结构域具有重要意义。

表 1 RHG1 和 RHG4 蛋白质的氨基酸组成

Table 1 Amino acid composition of RHG1and RHG4 protein(%)

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
	(A)	(R)	(N)	(D)	(C)	(Q)	(E)	(G)	(H)	(I)	(L)	(K)	(M)	(F)	(P)	(S)	(T)	(W)	(Y)	(V)
RHG1	5.7	4.4	7.1	4.0	2.2	3.5	4.2	7.6	2.1	4.8	15.0	4.8	1.2	3.9	4.7	10.9	4.7	1.4	1.4	6.4
RHG4	5.6	3.2	5.9	5.2	1.2	4.8	3.9	8.6	1.7	4.3	12.5	4.4	1.3	4.1	5.6	10.6	6.3	1.6	1.9	7.2
f/pr	8.3	5.7	4.4	5.3	1.7	4.0	6.2	7.2	2.2	5.2	9.0	5.7	2.4	3.9	5.1	6.9	5.8	1.3	3.2	6.6

f/pr 代表氨基酸在蛋白质中出现的频率,其中在 RHG1 和 RHG4 蛋白质中高于该频率的氨基酸的百分比加粗显示。

f/pr indicates the frequency of amino acid appeared in protein,and the bold data stands for the ratio of amino acid higher than f/pr.

对 RHG1 和 RHG4 进行亲/疏水性分析,结果提示

RHG1 有两个区域具有明显的疏水性,大约位于 30 ~50

区域和 480 ~510 区域(图 1),RHG4 只有一个明显的疏水性区域,大约位于 450 ~470 区域(图 2)。

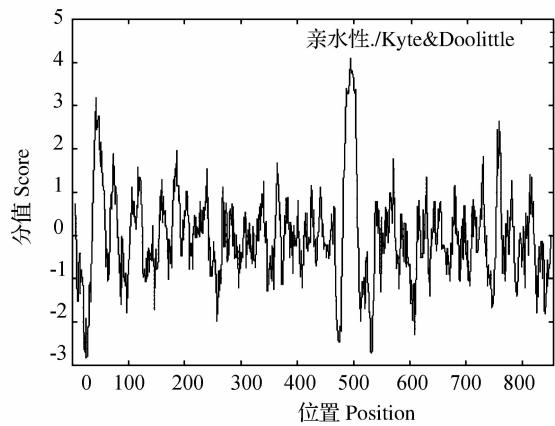


图 1 RHG1 蛋白质的亲/疏水性

Fig. 1 Hydrophobicity profile of RHG1 protein

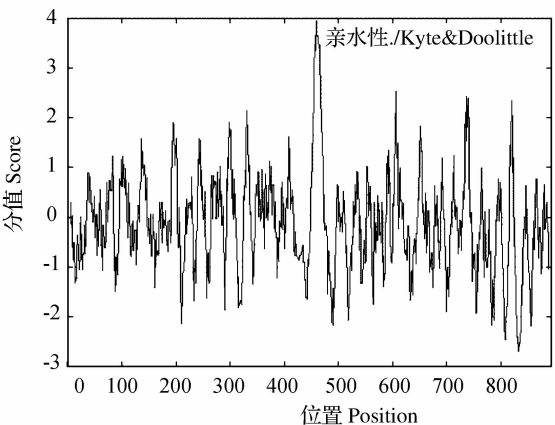


图 2 RHG4 蛋白质的亲/疏水性

Fig. 2 Hydrophobicity profile of RHG4 protein

在 TMPRED 服务器进行两蛋白质的跨膜区域分析,RHG1 在 483(486) ~ 507(505) 区域具有高的分值 3752(o ~ i),在 37 ~ 56(20) 区域具有高的分值 2284(i ~ o),而 RHG4 在 452(452) ~ 468(468) 区域具有高的分值 3601(o ~ i)。跨膜结构预测的结果与蛋白质亲/疏水性分析结果的区域基本一致,预测的跨膜结构是这两个类受体蛋白激酶产物重要的结构域之一,同时也提示两蛋白质产物为膜蛋白,在抗病反应的细胞内外信号传导过程中发挥重要作用。

2.2 RHG1 和 RHG4 蛋白质结构域及功能位点

据报道,基因 *Rhg1* 和 *Rhg4* 产物具有三个重要的结构域:细胞外的亮氨酸富集区、跨膜结构域和激酶结构域,数据库注释其为类受体激酶家族,但缺少具体结构域位置特征的注释信息。利用简单模块构架搜索工具 SMART 对 RHG1 和 RHG4 的结构域

进行分析,结果 RHG1 胞外亮氨酸富集区位于 139 ~ 409 之间,RHG4 胞外亮氨酸富集区位于 56 ~ 401 之间。RHG1 在 N-端 32 ~ 51 区域、LRR 与蛋白激酶结构域之间的 485 ~ 507 区域预测了两个跨膜结构疏水性区域,而 RHG4 只在 LRR 与蛋白激酶结构域之间的 449 ~ 471 区域为跨膜结构域,SMART 关于跨膜结构的预测结果与蛋白质基本性质分析中的亲/疏水性分析和 TMPRED 跨膜区域分析结果基本一致。SMART 结果显示 RHG1 的膜内 568 ~ 837 区域和 RHG4 膜内 544 ~ 820 区域为其蛋白激酶结构域。这种相似的结构域特征预示两蛋白质可能通过二聚体发挥功能。

在多数情况下,R 基因的多样性由 LRR 结构域控制,通过适应性选择产生小种特异性^[4]。因此对 LRR 结构域的具体结构做了进一步的分析。在 Pfam 数据库中搜索显示 RHG1 和 RHG4 的 LRR 结构域均包括一个 LRRNT_2 (LRR N-terminal domain, LRR N 末端的结构域) 和 12 个 LRR_1 重复单元, LRR 重复单元含有 24 个左右的氨基酸,两者的 LRR 结构域的保守序列分别为 LXXLXLXXNX-LXGXIP 与 LXXLXLXXNXLXGXXP。其中在 RHG4 蛋白质 LRR 结构域第 9 个和第 10 个 LRR_1 重复之间有一段由 62 个氨基酸组成的大的间隔区域,该区域是否具有某种特殊的意义有待于进一步的实验证明。

在 Prosite 数据库中的 Motif 与 Profile 搜索蛋白激酶结构域功能位点模式,结果见表 2,两蛋白质都具有 NP_BIND 和 BINDING 模体,并且两模体序列具有一定的保守性: NP_BIND 模体序列 GxG/SxxGxV 和 BINDING 模体序列 AVK。将 RHG1 和 RHG4 与其他植物来源的 LRR 型类受体蛋白激酶进行多序列比较,发现 RHG1 和 RHG4 同样具有激酶域的 11 个高度保守的亚结构域,详见图 3。拟南芥 BRI1 研究表明其 AIK 的 K₉₁₁ 位点为 ATP 结合位点,该位点突变导致蛋白激酶活性丧失,根据多序列比对结果和模体预测结果,RHG1 和 RHG4 的 AVK 的赖氨酸同样为 ATP 结合位点,对蛋白激酶活性具有重要意义。另外 RHG4 模体预测的 Fi-HrDLKpsNILL 与 BRI1 的 HRDMKSSNVLL 同属于蛋白激酶结构域 VIb, BRI1 的 D 位点预测为质子受体,对于激酶活性有重要影响,而 RHG4 模体预测其为丝氨酸/苏氨酸蛋白激酶的活性位点,同样具有重要作用。

表 2 RHG1 和 RHG4 蛋白质 motif 分析结果
Table 2 Motif in RHG1 and RHG4 protein

蛋白质 Protein	模体 Motif	氨基酸位点 Amino acid site
RHG4	NP_BIND (Protein kinases ATP-binding region)	550 ~ 572 LGRGGFGVVYkGqlhdgtk IAVK
	BINDING(ATP)	572 AVK
	Serine/Threonine protein kinases active-site signature	669 ~ 681 FiHrDLKpsNILL
RHG1	NP_BIND(Protein kinases ATP-binding region)	575 ~ 583 MGKSTNGTV
	BINDING(ATP)	597 AVK

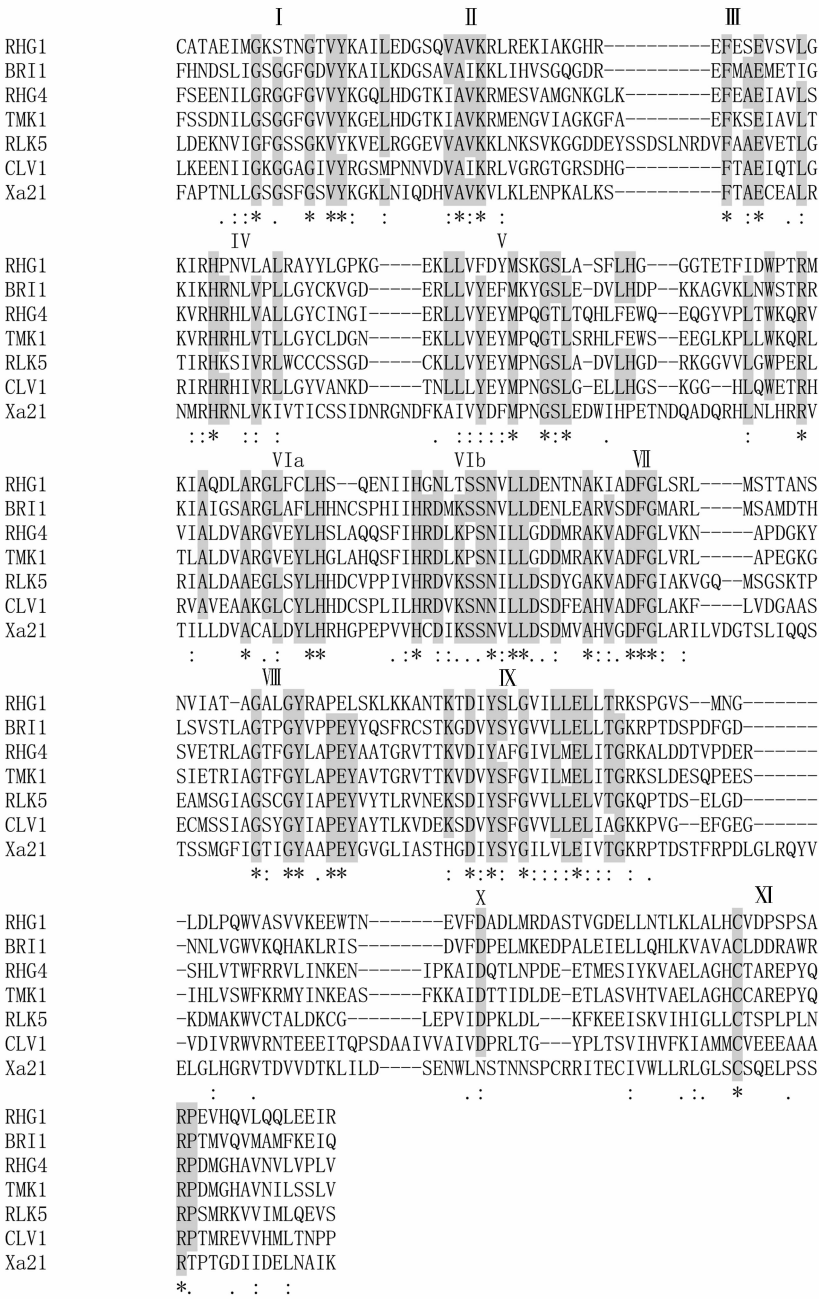


图 3 植物 LRR 型受体激酶的蛋白激酶结构域的多序列比较

Fig. 3 Multiple sequence alignment of protein kinase domain among putative LRR receptor kinases in plants

表 3 RHG1 与 RHG4 蛋白激酶结构域内的磷酸化位点

Table 3 Phosphorylation site in RHG1 and RHG4 protein kinase domain

蛋白质 Protein	Pattern-ID	Pattern-DE	模板 Pattern			
RHG1	PKC_PHOSPHO_SITE	Protein kinase C	479	STK,512	TSK,764	TRK
	PS00005 DOC00005	phosphorylation site	812	TLK,825	SAR	
RHG4	PKC_PHOSPHO_SITE	Protein kinase C	643	TWK,723	TGR,727	TTK
	PS00005 PDOC00005	phosphorylation site	744	TGR,803	TAR,877	SSK

受体蛋白激酶结构域的重要特征之一是带有磷酸化位点,激酶域的磷酸化能够激活激酶或进一步提高酶的活性。通过 Prosite 数据库搜索对 RHG1 与 RHG4 蛋白激酶结构域内可能的磷酸化位点进行预测分析,发现了多个磷酸化位点,详见表 3。激酶结构域的亚结构域Ⅶ和Ⅷ被称为活性环,许多激酶在激活时,这个区域被磷酸化,从而提供一个与底物结合的稳定结构。在拟南芥 BRI1 活化环上的磷酸化位点的突变导致多肽磷酸化和自我磷酸化的丧失,而其他的磷酸化位点的突变降低多肽磷酸化,但不影响自我磷酸化。在 RHG1 与 RHG4 相对应的活化环位置可能进行磷酸化的氨基酸残基分别为 LSRLMSTTANSNVIATAGAL, LVKNAPDGKYSVETR-LAGTF。

2.3 RHG1 与 RHG4 蛋白质二级结构

应用 PHD 对蛋白质进行二级结构预测,在 RHG1 与 RHG4 之间相同的结构域区间其二级结构单元的组成基本一致,在 LRR 结构域 LRR 重复主要由不规则卷曲、折叠和 α 螺旋组成,折叠-不规则卷曲- α 螺旋作为基本单元重复出现。跨膜结构由 α 螺旋组成。激酶结构域主要由 α 螺旋和折叠组成, α 螺旋、折叠分布相对较集中,其间伴随有大量的随机卷曲,N 端相对折叠含量高,含有部分 α 螺旋结构单元,而 C 末端主要由 α 螺旋组成,少量的甚至缺少折叠结构单元。由预测结果可知 RHG1 与 RHG4 均采用 α/β 折叠。

2.4 RHG1 与 RHG4 蛋白激酶结构域空间建模

将 RHG1 与 RHG4 提交给 SWISS-MODEL 进行自动的蛋白质建模,在全长序列范围内没有发现合适的模型,与 1OGQ_A 植物防御反应 Pgip (Polygalacturonase Inhibiting Protein) 在 LRR 结构域有约 25% 左右的一致性,在蛋白激酶结构域内与已知结构的蛋白激酶有 30% 左右的一致性。同源建模结果的可靠性主要依赖于选取模板与蛋白质的相似性程度,因此要求模板与蛋白质的一致性高于 30%。根据前面的分析,RHG1 与 RHG4 蛋白质均属于跨膜蛋白,LRR 结构域位于细胞外,而蛋白激酶结构

域位于细胞内,两者之间由于空间的限制,蛋白质结构上没有相互作用,可以采用分别预测的策略,利用局部序列进行相应结构的预测。针对具有较高一致性的蛋白激酶结构区段进行了同源建模预测。对于 LRR 结构域,几种 LRR 型蛋白亚家族的三维结构已被确定,均为类似马蹄状的弯曲开放结构。根据 RHG 蛋白质二级结构分析结果和与 LRR 结构域具有相似结构,可以推测 RHG 蛋白的胞外 LRR 结构域形成马蹄状的空间结构。

所有的真核生物丝氨酸/苏氨酸及酪氨酸特异性蛋白激酶的核心采用相同的折叠方式。预测获得的结构也具有相同的结构特点。具有由铰链区连接的两个叶状结构区,N 端的叶状结构包含若干 β 片层和一个 α 螺旋,C 端的叶状结构主要由 α 螺旋组成。蛋白激酶的核心结构包括富含甘氨酸的 P 环 GXGXXGXV (G₇RGGFGVV₁₄-RHG4, G₃₇KSTNGTV₄₄-RHG1) 是 ATP 的磷酸根锚定所必需的结构,与 prosite 数据库搜索认为该区域为蛋白激酶 ATP-结合区域的结果一致,该区域在模拟的结构中由不规则卷曲形成环状,见图 4,图 5。N 端叶状结构内的 C 螺旋在 RHG4 中由 E₄₄AEIAVLSK₅₂ 组成,RHG1 的 C 螺旋不明显,推测由 E₇₀FESEVSVLG₇₉ 组成,形成松散的结构。在大多数活性状态的蛋白激酶构象中,C 螺旋形成一个盐桥(图中利用点表示出分子表面),连接到 N 端叶状结构内保守的赖氨酸残基上,即 RHG1 AVK₅₈与 RHG4 AVK₂₈,使得 ATP 的磷酸盐获得最理想的定位。在模拟的结构中 C 螺旋与赖氨酸残基在空间比邻具有形成盐桥的有利条件,在活性状态两者可能进一步靠近,从而形成盐桥。催化环 HRDLKXXN(RHG4) HXXLXXSN(RHG1) 主要由不规则卷曲组成,在这个结构中 RHG4 具有保守的天门冬氨酸 D₁₂₉,根据其他蛋白激酶的结构特点,其为与镁离子-ATP 结合的区域,RHG1 相应的位置为 N₁₅₄残基。活化区形成了保守的 DFG 基序,其中 RHG1 D₁₇₂和 RHG4 D₁₄₇残基能与金属离子-ATP 相互结合。在许多蛋白激酶中,活化环上的丝氨酸/苏氨酸发生磷酸化以响应外界激活信号,这种磷酸化

反应促使激酶转入有活性状态的构象。对于可能的磷酸化位点的突变分析,将进一步确定其对于激酶活性状态转变的作用。两蛋白在蛋白激酶结构域区域共同的空间结构特征同样可以说明两者有可能单独或彼此形成二聚体行使功能。

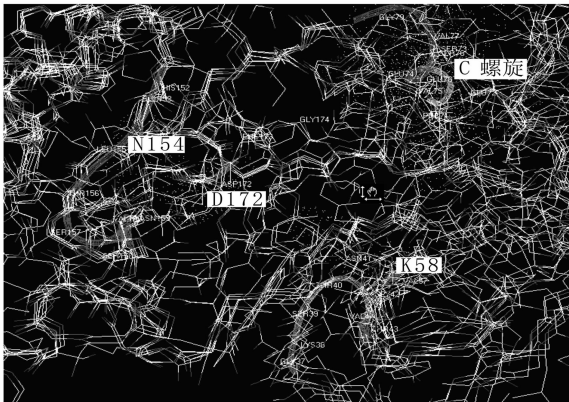


图4 RHG1 蛋白激酶结构域重要的核心结构
Fig.4 Core part of RHG1 protein kinase domain structure

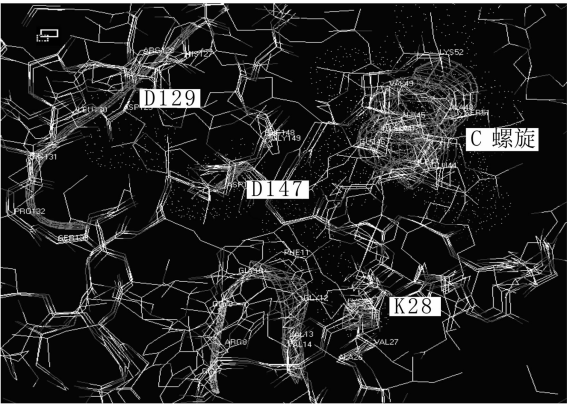


图5 RHG4 蛋白激酶结构域重要的核心结构
Fig.5 Core part of RHG4 protein kinase domain structure

3 讨论

目前,在已经鉴定的植物富含亮氨酸重复序列型类受体蛋白激酶中与抗病相关的 LRR 型类受体蛋白激酶受到人们的关注,例如拟南芥的抗性基因 *FLS2*、水稻的 *Xa21* 等。本研究对大豆抗病相关的类受体蛋白激酶 RHG1 和 RHG4 蛋白进行了详细的生物信息学分析,尤其是对其功能具有重要意义的三个重要结构域。在 LRR 结构域,Pfam 预测结果表明两蛋白的 LRR 重复单元均为 12 个(不包括 LRR N 末端的结构域),保守序列分别为 LXXLXLXXNXLXXXIP (RHG1)与 LXX-LXLXXNXLXGXXP (RHG4),两者具有高度的相似性。LRR 结构域参与蛋白质-蛋白质互作和与配体结合,

也可促进基因产物与参与防卫反应信号的其他蛋白质作用,但对其中的绝大多数尚未鉴定出其配基。蛋白激酶结构域具有 11 个高度保守的亚结构域,空间结构与所有真核生物的丝氨酸/苏氨酸及酪氨酸特异性蛋白激酶的核心相同的折叠方式。类受体蛋白激酶可以通过胞外结构与胞外的信号分子的特异结合来激活胞内的激酶域的自磷酸化和互磷酸化活性,完成跨膜传递信号的功能。对这两个蛋白质的磷酸化位点、活性位点以及蛋白激酶结构域空间结构的预测将有助于理解 RHG 蛋白在抗病反应中的作用。

蛋白质空间结构对于蛋白质发挥功能和蛋白质活性部位必须氨基酸的确定具有重要意义。同源建模方法是目前蛋白质结构预测的重要方法,本研究运用 SWISS - MODEL 对 RHG 蛋白质进行了蛋白激酶结构域区域的同源建模,预测的结构具有明显的激酶结构的保守特征,两蛋白在空间结构上同样具有很高的相似性。同时基于结构对功能位点进行了详细的分析。对于胞外的 LRR 结构域,虽然没有进行同源建模,但是基于二级结构分析的结果和已知的具有 LRR 结构域的空间结构,认为 RHG 蛋白的三维结构可能为马蹄型,该结构模体是 LRR 类蛋白质普遍的结构特征。

通过生物信息学方法对蛋白质的分析预测,对 RHG 蛋白质从基本性质、结构域到空间结构,包括可能的功能位点,有了一个较全面的认识,为进一步对 RHG 蛋白质生物功能的深入研究提供了线索。

参 考 文 献

[1] Jmui A,Ruben E,Shihama H,et al. Resistance to soybean cyst nematode;An *Xa21* like gene family that requires possible dimerization for signal transduction[C]. 9TH Biennial conference of the cellular and molecular biology of the soybean,2002;11 -14.

[2] Gasteiger E,Hoogland C,Gattiker A,et al. Protein identification and analysis tools on the ExPASy Server [M]. The Proteomics Protocols Handbook, Humana Press, N J,2005;571 -607.

[3] Schultz J,Milpetz F,Bork P. SMART, a simple modular architecture research tool;Identification of signaling domains[J]. Proceedings National Acadeing Science U S A,1998,95(11):5857 -5864.

[4] Dangl J L,Jones J D. Plant pathogens and integrated defence responses to infection[J]. Nature,2001,411:826 -833.