

大豆过敏原 Gly m 4 蛋白抗原表位特征预测

郭晓晓<sup>1</sup>, 李文<sup>2</sup>, 王永飞<sup>1</sup>, 何颖<sup>2</sup>, 黄于艺<sup>2</sup>, 邹泽红<sup>2</sup>

(1. 暨南大学 生物工程学系, 广东 广州 510632; 2. 广州医科大学 附属第二医院/呼吸疾病国家重点实验室/中法霍夫曼免疫研究所/广东省过敏反应与免疫重点实验室/变态反应国家临床重点专科/广州市过敏反应临床医学研究与转化中心, 广东 广州 510250)

**摘要:**通过生物信息学相关软件分析大豆过敏原 Gly m 4 蛋白理化性质(ProtParam)、信号肽(SignalP 4.1 Server)、跨膜区(TMhmm Server V 2.0)、B 细胞表位(DNAStar)、MHC-II 类分子的结合能力(NetMHC II 2.2)。结果发现:Gly m 4 蛋白稳定性较好,无信号肽与跨膜区,转角结构丰富;B 细胞抗原表位预测表明,Gly m 4 蛋白 61~64、93~94、122~125、127~130、134~137 区域是潜在 B 细胞抗原表位;MHC-II 类分子的结合力分析表明 Gly m 4 蛋白 144~153 区域及 82~96 区域是潜在 T 细胞抗原表位,同时发现 HLA-DRB10701、HLA-DRB10101 等位基因型人群对 Gly m 4 蛋白较敏感。Gly m 4 蛋白抗原表位分析为大豆过敏原的低过敏性改造提供参考依据。

**关键词:**过敏性;Gly m 4 蛋白;B 细胞表位;T 细胞表位

**中图分类号:**R392.8      **文献标识码:**A      **DOI:**10.11861/j.issn.1000-9841.2017.03.0365

Prediction on Antigenic Epitope Characteristics of Gly m 4 Protein in Soybean

GUO Xiao-xiao<sup>1</sup>, LI Wen<sup>2</sup>, WANG Yong-fei<sup>1</sup>, HE Ying<sup>2</sup>, HUANG Yu-yi<sup>2</sup>, ZOU Ze-hong<sup>2</sup>

(1. Department of Biotechnology, Jinan University, Guangzhou 510632, China; 2. The Second Affiliated Hospital of Guangzhou Medical University/The State Key Laboratory of Respiratory Disease/Sino-French Hoffmann Institute/Guangdong Provincial Key Laboratory of Allergy & Clinical Immunology/The State Key Clinical Specialty in Allergy/Guangzhou Municipal Research Center of Clinical and Translational Allergy, Guangzhou 510632, China)

**Abstract:** To analyze the physicochemical properties and structure of major *Glycine max*(Linn.) Merr. allergen Gly m 4 protein using bioinformatics software and provide a reference for modifying the allergen Gly m 4 protein experimentally. The amino acid sequence of Gly m 4 protein was searched from NCBI database. The physicochemical properties were analyzed by ProtParam, the signal peptide of Gly m 4 protein was analyzed by SignalP 4.1 Server, the transmembrane helix was analyzed by TMhmm Server V 2.0, the B cell epitopes were predicted with DNAStar. The binding affinity between Gly m 4 protein and MHC-II molecules was analyzed with NetMHCII 2.2 Server to predict the T cell epitopes. The results showed that Gly m 4 protein was stable and doesn't possess any signal peptide and transmembrane helix, most secondary structures of Gly m 4 were turn regions. Prediction result suggested the potential B cell epitope of Gly m 4 were located in the region of 61-64, 93-94, 122-125, 127-130 and 134-137. Analysis of the binding affinity between Gly m 4 and MHC-II molecules suggested the regions of 144-153 and 82-96 were the potential T cell epitopes. Human with HLA-DRB10701 alleles and HLA-DRB10101 alleles maybe more sensitive to Gly m 4 protein. This study facilitates to understand the antigen epitope of Gly m 4 protein and provides a reference to reduce the allergenicity of Gly m 4.

**Keywords:** Allergenicity; Gly m 4; B cell epitopes; T cell epitope

随着人们食物多样性日益增加,食物过敏的发病率逐年升高,已经成为世界范围内备受关注的公共卫生问题<sup>[1-2]</sup>。目前已有 170 多种食物能引起食物过敏反应,其中 IgE 介导的速发型过敏反应最为常见<sup>[3]</sup>。

大豆[*Glycine max*(Linn.) Merr.]是八大食物过敏原之一,其主要过敏原为 Gly m 1、Gly m 2、Gly m 3、Gly m 4、Gly m 5、Gly m 6、Gly m 7、Gly m 8,其中除 Gly m 1、Gly m 2 外均为食物过敏原(<http://www.allergen.org/index.php>)。Gly m 5、Gly m 6、Gly m 8 常见于儿童过敏反应,Gly m 5 研究较为广泛。

Gly m 4 是一种病程蛋白,属于 Bet v 1 蛋白家族,常见于成人食物过敏反应中。近年来受到人们的广泛关注,研究多集中于 Gly m 4 在过敏反应中的灵敏性以及其特异性 IgE 结合位点<sup>[4-6]</sup>,鲜有 Gly m 4 抗原表位分析相关研究。Berneder 等<sup>[7]</sup>研究发现 Gly m 4 在大豆过敏检测中具有较高灵敏度,可作为检测大豆食物过敏反应程度的指标。Havenith 等<sup>[8]</sup>通过分析 Gly m 4 与 Bet v 1 相关的 IgE 特异性肽段结合位点,对其进行突变获得低过敏性突变体。

抗原表位在免疫活动发生过程中发挥关键作用,预测抗原表位可为过敏原相关研究奠定基础。

收稿日期:2017-01-11  
基金项目:国家科技重大专项重点课题(2014ZX08011-05B)。  
第一作者简介:郭晓晓(1990-),女,硕士,主要从事蛋白过敏性评价研究。E-mail:1214508482@qq.com。  
通讯作者:王永飞(1972-),男,博士,教授,主要从事植物育种研究。E-mail:wyfmsm@163.com。

抗原表位因其受体细胞不同,分为 T 细胞表位和 B 细胞表位。针对 T 细胞表位进行免疫治疗,使得过敏原特异性免疫治疗更为安全<sup>[9]</sup>。Hosoyama 等<sup>[10]</sup>研究发现了 Gly m Bd 30K 的主要抗原表位,邹菊等、林苏霞等<sup>[11-12]</sup>在此基础上对 Gly m Bd 30K 进行克隆表达及免疫原性鉴定研究,并成功制备 Gly m Bd 30K 单克隆抗体。

为推进 Gly m 4 在过敏反应中的下游研究,亟需对其过敏原表位进行分析。本研究利用生物信息学技术分析 Gly m 4 蛋白理化性质和空间结构,并预测其 B 细胞和 T 细胞抗原表位,为 Gly m 4 蛋白的单克隆抗体制备及低过敏原性改造奠定基础,同时为其它食物过敏原抗原表位预测提供参考。

1 材料与方法

1.1 材料

Gly m 4 蛋白氨基酸序列(P26987)来自 Unipro (<http://www.uniprot.org/>)数据库。

1.2 方法

1.2.1 Gly m 4 蛋白理化性质分析 通过 ProtParam (<http://web.expasy.org/protparam/>),分析 Gly m 4 蛋白分子量、等电点、氨基酸组成、半衰期、不稳定指数、亲水性等理化性质。

1.2.2 Gly m 4 蛋白信号肽预测 通过 SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>)预测 Gly m 4 蛋白信号肽<sup>[13]</sup>。D 值为 S 值平均值与 Y 值最高峰的加权平均数,用以区别是否为分泌蛋白。

1.2.3 Gly m 4 蛋白跨膜区预测 通过 TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>)在线软件<sup>[14]</sup>预测 Gly m 4 蛋白的跨膜区。

1.2.4 Gly m 4 蛋白 B 细胞表位预测 利用 DNASTar 软件 Protean 模块,分别采用 Chou-Fasman 法、

Karplus-Schulz 法、Kyte-Doolittle 法、Emini 法、Jame-son-Wolf 法分析 Gly m 4 蛋白二级结构、柔韧性、亲水性、表面可及性、抗原性指数。

1.2.5 Gly m 4 蛋白三级结构预测 通过在线服务器 SWISS-MODEL (<https://swissmodel.expasy.org/interactive>),采用自动建模的方式构建蛋白三维结构。

1.2.6 Gly m 4 蛋白 T 细胞表位预测 通过在线软件 NetMHC II 2.2 Server (<http://www.cbs.dtu.dk/services/NetMHC/>)预测 Gly m 4 蛋白与 MHC II 类分子的结合能力<sup>[15]</sup>。每 9 个相邻氨基酸为一段待测多肽,预测其亲和力。结合能力以“1 - log50k (aff)”的数值表示。

2 结果与分析

2.1 Gly m 4 蛋白理化性质分析

Gly m 4 蛋白等电点为 4.69,由 158 个氨基酸组成,含有 25 个带负电荷氨基酸(天冬氨酸和谷氨酸)、16 个带正电荷氨基酸(精氨酸和赖氨酸),丙氨酸含量最高(18 个,占 11.4%)。分子式为 C<sub>753</sub>H<sub>1179</sub>N<sub>187</sub>O<sub>243</sub>S<sub>1</sub>,分子量为 16 771.81 Da; Gly m 4 蛋白不稳定指数为 21.44,表明 Gly m 4 蛋白较为稳定;Gly m 4 蛋白脂肪酸系数为 83.42,提示其不耐热;平均亲水性为 -0.272,表明 Gly m 4 蛋白易溶于水。

2.2 Gly m 4 蛋白信号肽预测

图 1 为 Gly m 4 信号肽预测结果,第 28 位天冬氨酸(D)C 值和 Y 值数值最大,分别为 0.131、0.114;第 16 个位点的脯氨酸(P)S 值最大,为 0.120;第 1~27 个位点的 S 值平均为 0.100;第 1~27 个位点 D 值为 0.107,低于 D 值阈值 0.45,推断 Gly m 4 无信号肽。

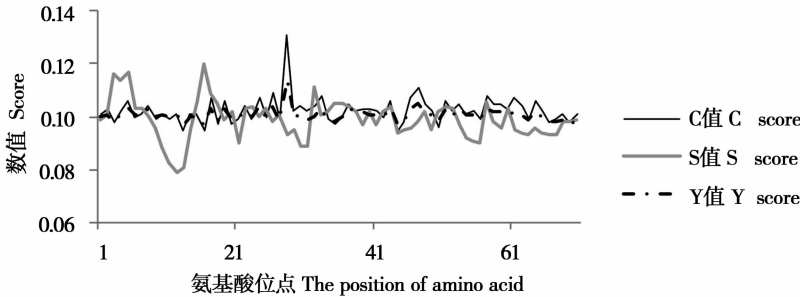


图 1 Gly m 4 信号肽预测

Fig. 1 Prediction of Gly m 4 signal peptide

2.3 Gly m 4 蛋白跨膜区预测

Gly m 4 蛋白跨膜区预测结果如图 2 所示,该蛋白被标记为 outside,即该蛋白位于膜外,跨膜区氨基酸数以及 N 端前 60 个氨基酸位于膜一侧的数目为

0,据此推测该蛋白不含跨膜区。

综上所述,推测 Gly m 4 蛋白稳定性较好,不耐热,亲水性强;无信号肽、无跨膜区,可能存在于细胞质中。

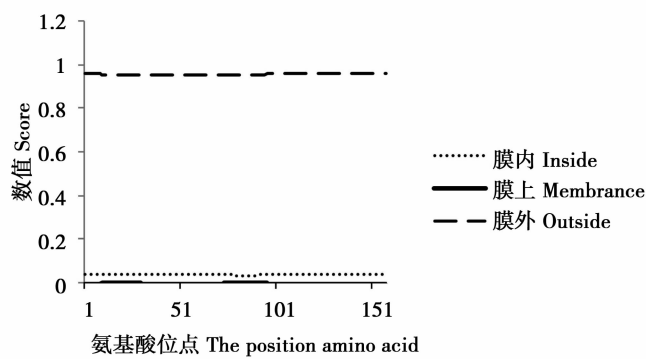


图2 Gly m 4 蛋白跨膜区预测

Fig. 2 Prediction of Gly m 4 transmembrane helix

2.4 Gly m 4 蛋白 B 细胞抗原表位预测

2.4.1 Gly m 4 蛋白二级结构分析 Gly m 4 蛋白二级结构分析结果显示:α 螺旋、β 折叠之间转角结构丰富,主要位于 26 ~ 29、36 ~ 39、45 ~ 52、61 ~ 64、81 ~ 84、91 ~ 94、101 ~ 104、107 ~ 110、112 ~ 115、122 ~

125、127 ~ 130、134 ~ 137、154 ~ 157 区段(图 3)。蛋白质二级结构中 α 螺旋、β 折叠不易发生形变,通常不作为抗原表位参与抗体结合,转角结构易于形变,利于结合抗体,符合抗原表位条件,抗原表位可及性较高。

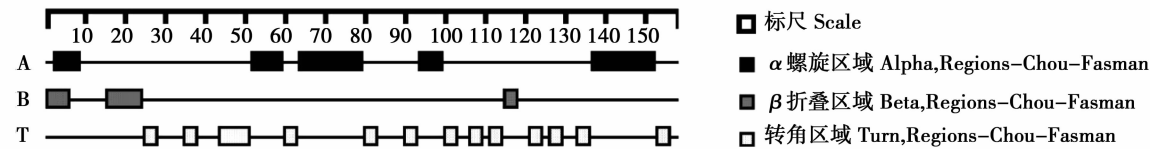


图3 Chou-Fasman 法分析 Gly m 4 蛋白二级结构

Fig. 3 Chou-Fasman method to analyze the secondary structure of Gly m 4 protein

2.4.2 Gly m 4 蛋白柔韧性分析 分析结果显示, Gly m 4 蛋白柔韧性范围较为集中,区段范围较大,主要位于 8 ~ 14、25 ~ 29、33 ~ 56、60 ~ 65、71 ~ 77、

92 ~ 104、107 ~ 116、119 ~ 140 区段(图 4)。蛋白柔韧性较大的区域利于结合抗体,可能出现在抗原表面。

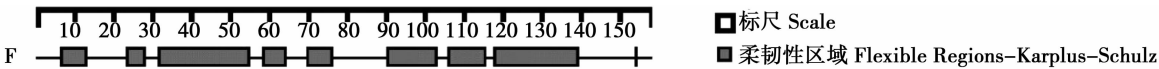


图4 Karplus-Schulz 法分析 Gly m 4 蛋白柔韧性

Fig. 4 Karplus-Schulz method to analyze the suppleness structure of Gly m 4 protein

2.4.3 Gly m 4 蛋白亲水性分析 Gly m 4 蛋白亲水性分析结果如图 5 所示,其亲水性区段分布较广,主要分布在 40 ~ 53、58 ~ 66、72 ~ 80、93 ~ 100、117 ~

139 等区段,结果表明 Gly m 4 蛋白亲水性较大,易溶于水,与 Gly m 4 蛋白理化性质分析结果相符,亲水性大的区段位于抗原表位的可能性较大。

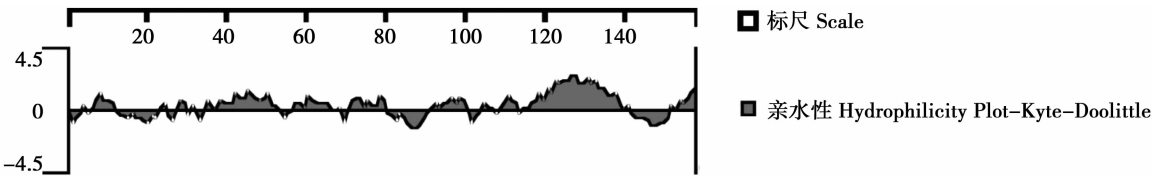


图5 Kyte-Doolittle 法分析 Gly m 4 蛋白亲水性

Fig. 5 Kyte-Doolittle method to analyze the hydrophilicity of Gly m 4 protein

2.4.4 Gly m 4 蛋白表面可及性分析

蛋白表面可及性较大的区段易于接触溶剂分子,出现在抗原表位的可能性较大。Gly m 4 蛋白表面可及性结果如图 6 所示,蛋白表面可及性大的位

置主要位于 61 ~ 63、93 ~ 95、119 ~ 139、155 ~ 158 区段。

2.4.5 Gly m 4 蛋白抗原性指数分析

Gly m 4 蛋白抗原性指数预测结果如图 7 所显,

Gly m 4 蛋白存在大量潜在抗原表位点,可能的抗原表位区域为 6 ~ 15、24 ~ 31、33 ~ 58、60 ~ 64、70 ~ 79、90 ~ 104、108 ~ 143、154 ~ 158 区段。

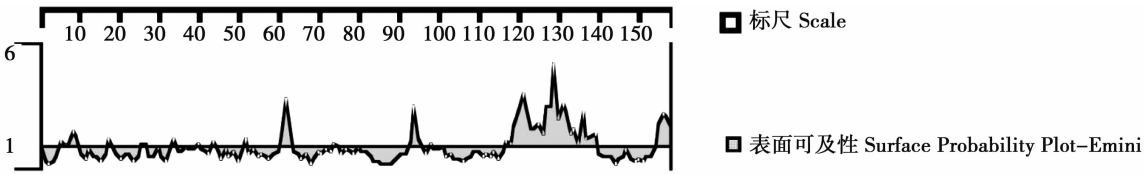


图 6 Emini 法分析 Gly m 4 蛋白表面可及性  
Fig. 6 Emini method to analyze the surface accessibility of Gly m 4 protein

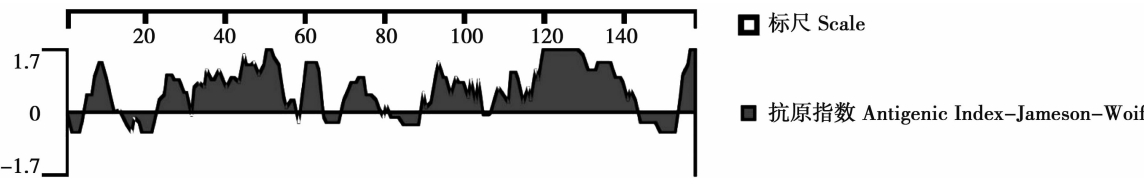


图 7 Jameson-Wolf 法分析 Gly m 4 蛋白抗原性指数

Fig. 7 Jameson-Wolf method to analyze the antigenic index of Gly m 4 protein

综合以上分析, 61 ~ 64、93 ~ 94、122 ~ 125、127 ~ 130、134 ~ 137 区域是潜在 B 细胞抗原表位。< 500 即  $0.425 < 1 - \log_{50}k(\text{aff}) < 0.6384$  ]。其中, 第 144 位九肽结合力最强, 87 位、85 位、81 位九肽次之, 潜在 MHC- II 类抗原表位为 144 ~ 153 区域和 82 ~ 96 区域附近。

2.5 Gly m 4 蛋白三维结构预测

蛋白功能与其空间结构密不可分, 对 Gly m 4 蛋白的高级结构预测, 有助于研究其过敏反应过程。Gly m 4 蛋白三维模型如图 8 所示, Gly m 4 蛋白主要由  $\alpha$  螺旋、 $\beta$  折叠构成, 转角结构松散, 数量较多, 与 Chou-Fasman 法二级结构分析相符。松散的空间结构易于形变, 利于结合免疫细胞, 参与下游免疫反应, 推测转角部位为抗原结合区域。

2.6 Gly m 4 蛋白 T 细胞表位预测

MHC- II 类抗原表位预测结果如图 9 所示, Gly m 4 蛋白有多个位点与 MHC- II 类分子具有较强结合力。其中有 2 个核心九肽具有强亲和力 (SB) [ $\text{affinity}(\text{nM}) < 50$  即  $0.6384 < 1 - \log_{50}k(\text{aff})$  ], 35 个核心九肽具有弱结合力 (WB) [ $50 < \text{affinity}(\text{nM})$  ]。

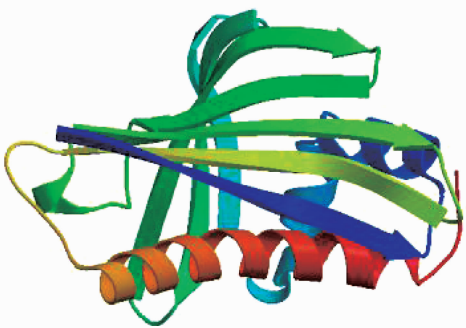


图 8 Gly m 4 蛋白三维结构模型  
Fig. 8 Gly m 4 protein three-dimensional structure model

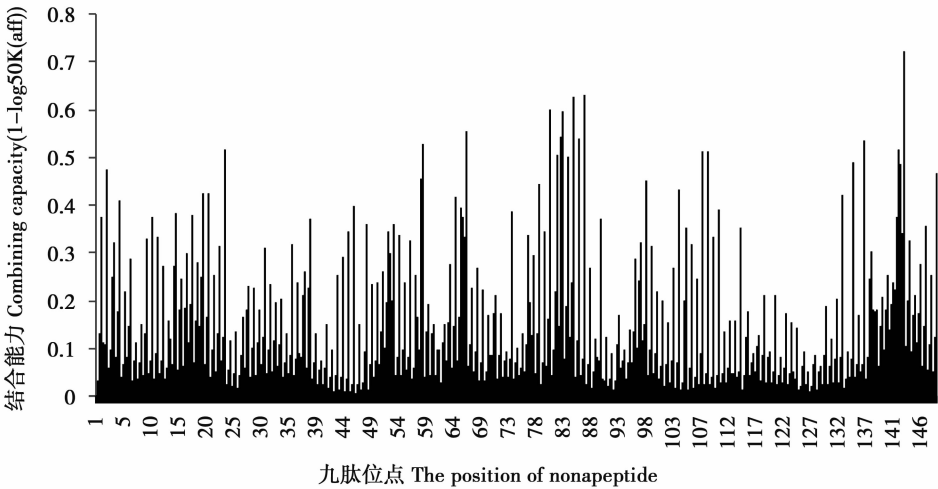


图 9 Gly m 4 蛋白 MHC- II 类分子的结合能力预测分析  
Fig. 9 Predication on MHC- II epitopes of Gly m 4

HLA 等位基因客观存在的多态性使得不同 HLA 基因型的个体可能会对某些特异性过敏原就有不同程度的免疫反应。本研究发现,与 Gly m 4 的九肽结合较强的 MHC-II 类分子集中在 HLA-

DRB10701、HLA-DRB10101 两种基因型(图 10),推断此两种基因型群体较其它基因型群体对 Gly m 4 蛋白易发生过敏反应。

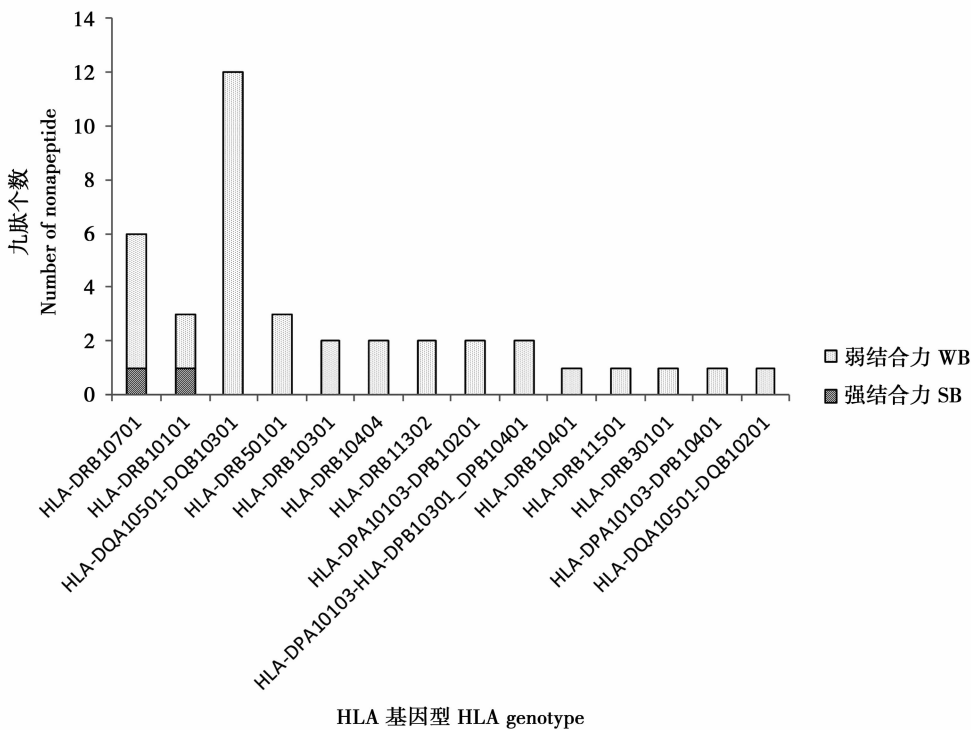


图 10 Gly m 4 与 HLA 基因型的相关性分析  
Fig. 10 Analysis of correlation between HLA genotype and Gly m 4

3 讨 论

近年来,由于环境改变以及物质生活日益丰富,人们接触致敏物质日益增多,过敏性疾病逐年增加。作为唯一针对过敏性疾病病因的治疗方法,过敏原特异性免疫治疗能够有效改变过敏性疾病自然进程,其治疗效果取决于过敏原的质量,皮下注射及舌下含服是其主要治疗方式<sup>[16]</sup>。基于过敏原组分的过敏原成分诊断能够快速识别过敏原,根据患者对不同组分的反应情况,对患者进行个体化治疗,提高了过敏原特异性免疫治疗疗效和安全性<sup>[17]</sup>。

过敏原表位分析可为过敏原成分诊断提供参考,研究者多采用蛋白结构分析、杂交肽裂解、肽芯片高通量筛选等方法分析过敏原表位,但此类方法存在工作量大、效率低下、成本高等缺点。通过生物信息学方法对过敏原表位进行预测分析在很大程度上降低研究成本、减少工作量、提高工作效率。

过敏原蛋白二级结构、柔韧性、亲水性、表面可及性以及抗原指数是 B 细胞抗原表位分析中常用指标<sup>[18-19]</sup>。本文经分析发现 61~64、93~94、122~125、127~130、134~137 区域位于 Gly m 4 蛋白转

角部位,柔韧性、亲水性、表面可及性大,抗原指数高,推断该区域为潜在 B 细胞抗原表位。

T 细胞在免疫反应中无法直接识别游离抗原,需经 MHC(主要组织相容性复合体)提呈抗原。因此,可通过 MHC-II 类分子与外源肽结合力强弱预测 T 细胞抗原表位。李胜涛等<sup>[20]</sup>综合分析多种预测软件,发现 NetMHC II 与 NetMHC II pan 预测准确度最高。本文选用 NetMHC II 2.2 Server 在线服务器,分析发现 Gly m 4 蛋白 144~153 区域及 82~96 区域 MHC-II 类分子结合力较高,推测该区域为潜在 T 细胞抗原表位,同时发现 HLA-DRB10701、HLA-DRB10101 等位基因型人群对 Gly m 4 蛋白较敏感。

本研究通过生物信息学软件分析大豆主要过敏原 Gly m 4 蛋白理化性质、结构,为 Gly m 4 蛋白表达纯化提供理论支持,抗原表位预测为单克隆抗体设计、过敏原疫苗设计提供参考。后续研究可对 Gly m 4 MHC II 抗原表位进行定点突变,降低其过敏原性,为食物过敏性疾病提供安全有效的候选疫苗。此外,分析获得的抗原 T 细胞表位可用于大豆过敏性疾病治疗,快速精确定位过敏原组分。

4 结 论

本研究发现 Gly m 4 蛋白稳定性较好,无信号肽与跨膜区,转角结构丰富; 61 ~ 64、93 ~ 94、122 ~ 125、127 ~ 130、134 ~ 137 区域是潜在 B 细胞抗原表位,144 ~ 153 区域及 82 ~ 96 区域是潜在 T 细胞抗原表位; HLA-DRB10701、HLA-DRB10101 等位基因型人群对 Gly m 4 蛋白较敏感。

参考文献

[1] Warner J O, Kaliner M A, Crisci C D, et al. Allergy practice worldwide: A report by the world allergy organization specialty and training council[J]. International Archives of Allergy and Immunology, 2006, 139(2): 166-174.

[2] Katz Y, Goldberg M R. Natural history of food protein-induced enterocolitis syndrome[J]. Current Opinion in Allergy and Clinical Immunology, 2014, 14(3): 229-239.

[3] Boyce J A, Assa'ad A, Burks A W, et al. Guidelines for the diagnosis and management of food allergy in the united states: Summary of the niaid-sponsored expert panel report[J]. Nutrition Research, 2011, 31(1): 61-75.

[4] Ito K. Grain and legume allergy[J]. Chemical Immunology and Allergy, 2015, 101: 145-151.

[5] Mittag D, Vieths S, Vogel L, et al. Soybean allergy in patients allergic to birch pollen: Clinical investigation and molecular characterization of allergens[J]. The Journal of Allergy and Clinical Immunology, 2004, 113(1): 148-154.

[6] Berkner H, Neudecker P, Mittag D, et al. Cross-reactivity of pollen and food allergens: Soybean gly m 4 is a member of the bet v 1 superfamily and closely resembles yellow lupine proteins[J]. Bio-science Reports, 2009, 29(3): 183-192.

[7] Berneder M, Bublin M, Hoffmann-Sommergruber K, et al. Allergen chip diagnosis for soy-allergic patients: Gly m 4 as a marker for severe food-allergic reactions to soy[J]. International Archives of Allergy and Immunology, 2013, 161(3): 229-233.

[8] Havenith H, Kern K, Rautenberger P, et al. Combination of two epitope identification techniques enables the rational design of soy allergen Gly m 4 mutants[J]. Biotechnology Journal, 2017, 12(2):1600441.

[9] Prickett S R, Rolland J M, O'Hehir R E. Immunoregulatory t cell epitope peptides: The new frontier in allergy therapy[J]. Clinical and Experimental Allergy, 2015, 45(6): 1015-1026.

[10] Hosoyama H, Obata A, Bando N, et al. Epitope analysis of soybean major allergen gly m bd 30k recognized by the mouse monoclonal antibody using overlapping peptides[J]. Bioscience, Biotechnology, and Biochemistry, 1996, 60(7): 1181-1182.

[11] 邹菊, 刘志刚. 大豆主要过敏原 Gly m Bd 30k 蛋白单克隆抗体的制备与应用[J]. 大豆科学, 2011, 30(5): 723-726. (Zou J, Liu Z G. Preparation and application of monoclonal antibodies against allergen Gly m Bd 30K[J]. Soybean Science, 2011, 30(5): 723-726. )

[12] 林苏霞, 王晓梅, 刘志刚, 等. 大豆主要过敏原 Gly m Bd 30k 的抗原表位区基因的克隆表达、纯化及免疫原性鉴定[J]. 大豆科学, 2010, 29(2): 186-190. (Lin S X, Wang X M, Liu Z G, et al. Cloning and expression of the antigenic epitope of Gly m Bd 30K protein from soybean and purification and identification of expressed product[J]. Soybean Science, 2010, 29(2): 186-190. )

[13] Petersen T N, Brunak S, von Heijne G, et al. Signalp 4.0: Discriminating signal peptides from transmembrane regions[J]. Nature Methods, 2011, 8(10): 785-786.

[14] Combet C, Blanchet C, Geourjon C, et al. Nps@ : Network protein sequence analysis [J]. Trends in Biochemical Sciences, 2000, 25(3): 147-150.

[15] Nielsen M, Lundegaard C, Worning P, et al. Reliable prediction of t-cell epitopes using neural networks with novel sequence representations[J]. Protein Science, 2003, 12(5): 1007-1017.

[16] Smarr C B, Bryce P J, Miller S D. Antigen-specific tolerance in immunotherapy of th2-associated allergic diseases [J]. Critical Reviews in Immunology, 2013, 33(5): 389-414.

[17] Bronnert M, Mancini J, Birnbaum J, et al. Component-resolved diagnosis with commercially available d. Pteronyssinus der p 1, der p 2 and der p 10: Relevant markers for house dust mite allergy [J]. Clinical and Experimental Allergy, 2012, 42(9): 1406-1415).

[18] Zhang H, Kurgan L. Improved prediction of residue flexibility by embedding optimized amino acid grouping into rsa-based linear models[J]. Amino Acids, 2014, 46(12): 2665-2680.

[19] Welling G W, Weijer W J, van der Zee R, et al. Prediction of sequential antigenic regions in proteins[J]. FEBS Letters, 1985, 188(2): 215-218.

[20] 李胜涛, 刘昌文, 邹泽红, 等. MHC-II 类抗原表位预测软件的对比评价[J]. 生物医学工程研究, 2010(2): 128-132. (Li S T, Liu C W, Zou Z H, et al. Comparative assessment of different softwares for MHC-II antigen epitope prediction[J]. Journal of Biomedical Engineering Research, 2010(2): 128-132. )