



DOI:10.13364/j.issn.1672-6510.20220108

芝麻过敏原的生物信息学分析

令狐晓盼, 王莎莎, 邱金平, 陆 旻
(天津科技大学食品科学与工程学院, 天津 300457)

摘要: 通过 DNASTAR 和 IEDB 等生物信息学软件对 7 种芝麻过敏原的生物学信息进行分析, 采用多种方案对芝麻过敏原的抗原指数、亲水性、表面可及性、柔韧性等参数及其二级结构进行预测, 并对预测的抗原表位综合分析. 利用 SWISS-MODEL 进行同源建模并用拉氏图评价其结构稳定性. 结果表明, 7 种芝麻过敏原蛋白均存在多个可能的抗原表位, 采用同源建模的方式成功构建了芝麻过敏原蛋白的三级结构模型, 拉氏图显示该模型的构象稳定. 本研究为制备特异性抗体肽段和过敏原检测等提供了依据.

关键词: 芝麻过敏原; 生物信息学; DNASTAR; SWISS-MODEL; 线性表位预测; 同源建模
中图分类号: TS201.6 **文献标志码:** A **文章编号:** 1672-6510(2023)02-0019-09

Bioinformatics Analysis of Sesame Allergens

LINGHU Xiaopan, WANG Shasha, QIU Jinping, LU Yang

(College of Food Science and Engineering, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: In this study the biological information of seven sesame allergens was analyzed via bioinformatics software such as DNASTAR and IEDB. The antigenic index, hydrophilicity, surface accessibility, flexibility and secondary structure of sesame allergens were predicted using (or with) various approaches (or methods). A comprehensive analysis of the predicted antigenic epitopes was performed. SWISS-MODEL was used for homology modeling and Ramachandran Plot was used to evaluate structural stability. The results showed that there were multiple possible epitopes in the seven sesame allergen proteins. The tertiary structure model of the sesame allergen protein was successfully constructed by homology modeling, and the conformational stability of the model was evaluated by Ramachandran Plot. The results could be useful for the preparation of specific antibody peptides and allergens detection.

Key words: sesame allergens; bioinformatics; DNASTAR; SWISS-MODEL; linear epitope prediction; homology modelling

芝麻 (*Sesamum indicum*) 是一种重要的油料作物, 在香料化工、医药保健等领域常被用作加工原料, 也通常被添加在烘焙食品、凉拌菜肴和饮品中增加食品的香气和风味. 同时, 芝麻也是常见的食物过敏原之一, 因其致敏的发病率逐年升高, 故引起越来越多科研人员的关注^[1-3]. 芝麻过敏是免疫球蛋白 (immunoglobulin, Ig) E 诱导的过敏反应, 在婴幼儿时期就有可能产生严重的过敏反应, 并且这种反应可能会伴随终身^[4-6]. 对芝麻过敏原阈值的相关研究发现, 芝麻过敏原的引发剂量 (eliciting dose, ED)₀₅ 在

芝麻蛋白含量为 1.2 ~ 4.0 mg 之间, ED₁₀ 在芝麻蛋白含量为 4.2 ~ 6.2 mg 之间^[7-11]. 正确标识过敏原是预防过敏患者误食相关食品的唯一有效途径, 许多西方国家要求对含有芝麻成分的食品必须进行标识, 日本在 2014 年也要求对此类食品进行标识.

芝麻过敏原有 7 种: 2 种 2S 白蛋白 Ses i 1 (相对分子质量为 $9 \times 10^3 \sim 1.4 \times 10^4$) 和 Ses i 2 (相对分子质量为 7×10^3), 1 种 7S 豌豆球蛋白 Ses i 3 (相对分子质量为 4.5×10^4), 2 种油质蛋白 Ses i 4 (相对分子质量为 1.7×10^4) 和 Ses i 5 (相对分子质量为 $1.5 \times$

收稿日期: 2022-04-23; 修回日期: 2022-08-15

作者简介: 令狐晓盼 (1996—), 女, 山西运城人, 硕士研究生; 通信作者: 陆 旻, 副教授, luyang@tust.edu.cn

10^4), 2种 11S 球蛋白 Ses i 6(相对分子质量为 5.1×10^4) 和 Ses i 7(相对分子质量为 5.6×10^4)^[12-16]. 其中 2S 白蛋白、7S 豌豆球蛋白和 11S 球蛋白是主要的过敏原蛋白, 其致敏活性由抗原表位决定. 明确芝麻过敏原蛋白的氨基酸序列、晶体结构及其 IgE 结合表位等生物信息, 对于预防芝麻过敏和开展致敏性消减研究等都具有重要意义. 马秀丽等^[17]对芝麻过敏原 Ses i 3 的生物信息学进行分析和预测, 其预测结果对过敏原识别和检测具有重要参考作用.

生物信息学对过敏原的分析越来越重要. 通过公开的过敏原网站可以检索各类过敏原的序列, 借助预测模拟系统预测蛋白质的生物学信息, 然后再借助核磁共振、冷冻电镜进行更深一步的分析^[18-22]. 当前有两种预测蛋白质结构的方法: 一是从头开始法, 不使用模板, 此类方法需要丰富的经验及物理知识的积累; 二是同源建模法, 根据相似度达到 30% 以上的已知蛋白质结构为参考模板进行分析预测^[23]. Alpha-Fold 算法应用于蛋白质结构预测, 其原理是根据蛋白质的氨基酸序列确定其三级结构, 算法准确度突破了 96%^[24]. 此算法可以与冷冻电镜等标准方法媲美. 常用的数据库有蛋白质结构数据库(PDB、BioMagResBank 等)、蛋白质序列数据库(the BepiPred、Uniprot 等)、蛋白质定位数据库(LOCAT、PSORTdb、MiCrokkit 等).

蛋白质属于生物大分子, 其结构的复杂性也决定了抗原表位的多样性. 由于表位是指一段特定的氨基酸序列, 并不是整个蛋白质的氨基酸序列, 因此即使去除其他相关序列、破坏蛋白质的三级结构, 由线性表位引起的特异结合也会引起人们的过敏反应, 即抗原和抗体的特异性结合. 抗原和抗体的结合点一般是几个氨基酸组成的结合位点, 最长的结合位点也不会超过 20 个氨基酸^[25-26]. 人体免疫细胞对过敏原的识别, 通常不是识别整个过敏原, 而是与蛋白质上的特定表位结合, 这些表位包括线性表位和空间表位. 线性表位是由一些氨基酸组成的一小段特定序列, 也称为顺序表位. 空间表位是由蛋白质的高度卷曲引起的不连续的氨基酸序列间结构, 也称为构象表位. 破坏蛋白质的三级结构, 空间表位就会失去特异性结合, 从而减轻过敏反应, 并且三级结构的破坏是不可恢复的^[27-28].

本研究将分子生物学技术与生物信息学技术相结合, 使用多参数软件预测芝麻过敏原的抗原表位, 并对其三级结构进行模拟, 旨在深入了解芝麻过敏

原, 进而为芝麻过敏原的精准检测和诊断提供参考.

1 材料与方法

1.1 芝麻过敏原的氨基酸序列

确定芝麻已知的致敏蛋白在国际免疫学联合会(IUIS)过敏原命名小组委员会中的名称, 在 NCBI 蛋白质数据库(<http://www.ncbi.nlm.nih.gov/protein/>)中查找其对应的氨基酸序列. 根据致敏蛋白结构数据库(SDAP), 确定获取的过敏原所属的蛋白家族.

1.2 芝麻过敏原序列与其他过敏原的同源性分析

将芝麻蛋白的氨基酸序列导入 SDAP 中进行同源查询(BLAST), 然后将具有高度同源的序列使用 DNASTar 中的 MegAlign 软件进行序列比对.

1.3 芝麻过敏原二级结构分析

使用 DNASTar 中 Protein 模块(采用 Chou-Fasman 和 Garnier-Robson 两种方案)进行芝麻过敏原的二级结构预测.

1.4 芝麻过敏原抗原性质分析

蛋白质的抗原性可以用 DNASTar 预测, 使用 DNASTar 中 Protein 模块的 Hoop-woods 的氨基酸亲水性分析方案、Emini-Surface Probability 的溶剂可及性算法、Kparlus-Schuzl 的可塑性分析和 Jameson-Wolf 的抗原指数分析方案, 对芝麻蛋白的一级序列性质进行分析, 再对这些参数预测结果的重叠部分进行综合分析, 最后确定芝麻蛋白的线性抗原表位序列.

1.5 芝麻过敏原抗原表位的分析预测

使用免疫表位数据库(IEDB)对芝麻过敏原的抗原表位进行预测. 该数据库是由美国国际过敏与感染性疾病研究院(NIAID)资助的免费资源. 该数据库记录了研究的抗体和 T 细胞表位的实验数据, 包括传染病、食物过敏和一些自身免疫的疾病等方面的研究. 将预测结果与 DNASTar 的预测结构进行比对, 最终预测分析其潜在的抗原表位.

1.6 同源建模

采用同源建模法对芝麻过敏原的三级结构进行预测, 选用的软件为 SWISS-MODEL^[29], 其原理是首先匹配具有 30% 以上同源的已知蛋白质结构, 然后以此为模板进行预测和分析. 所构建的模型会对其质量进行打分, 评价标准是分子动力学模拟和能量最小化两种处理方式结合的最佳结果. 评估结果通过拉氏图呈现, 拉氏图可以体现模型的准确度以及可接受性. 步骤如下:

(1) 首先找到目标序列的同源性蛋白,并且该同源性蛋白要满足两个条件:一是序列已知,二是同源性 $\geq 30\%$ 。

(2) 为目标序列与模板序列(可以多条)创建序列比对。通常比对软件自动创建的序列,同时还需要进一步的人工校正。

(3) 根据第二步创建的序列比对,用同源建模软件预测结构模型。

(4) 评估模型质量,并根据评估结果重复以上过程,直至模型质量合格。

如果目标序列与模板序列一致度极高,那么同源建模的准确性也会很高。

2 结果与讨论

2.1 芝麻过敏原的基本信息

芝麻过敏原的基本信息见表1,采用SignalP 5.0软件分析芝麻蛋白的信号肽,主要原理是先从数据库(NCBI)获得芝麻蛋白的序列,输入序列进行分析后得到信号肽。

表1 芝麻过敏原的基本信息
Tab. 1 Basic information of sesame allergens

过敏原	过敏原编号	氨基酸序列
Ses i 1	> AAK15088.1 2S albumin [<i>Sesamum indicum</i>]	<u>MAKKLALAAVLLVAMVALASA</u> TTYTTTTTTTAAIDDEANQQSQQCRRQLQGRQFRSCQRYLSQGRSPYGGEEDEVLEMSTGNQQSEQLRDCCQQLRNVDERCCEAIRQAVRQQQEGGYQEGQSQQVYQRARDLPRRCNMRPQQCQFRVIFV
Ses i 2	> AAD42943.1 2S albumin precursor [<i>Sesamum indicum</i>]	<u>MARFTIVLAVLFAAALVSASAHK</u> TVVTTTSVAEEGEEENQRGCEWESRQCQMRHMQWMRSMRGQYEEFLRSAEANQGQFEHFRECCNELRDVKSHCRCEALRCMMRQMQQEYGMEMQMMQMMQYLPRMCGMSYPTECRMRPIFA
Ses i 3	> AAK15089.1 7S globulin [<i>Sesamum indicum</i>]	MSCGGRLCLVLFALLASAVVASESKDPELKQCKHQCKAQQQISKEQKEACIQACKEYIRKQHQGEHGRGGGDILEEEVWNRKSPIERLRECSRGCEQQHGQEQRREECLRRQEEYQREKGRQDDNPTDPEKQYQQCRLQCRQEGEGGFSREHCERRREEKYREQQGREGGRGEMYEGREEREEQEEQGRGRIPYVFEDQHFITGFRTQHGRMRVLQKFTDRSELLRGIENYRVAILEAEPQTFIVPNHWDAESVVFVAKGRGTISLVRQDRRESLNKQGDILKINAGTTAYLINRDNNERLVLAKLLQPVSTPGEFELFFGAGGENPESFFKSFSDEILEAAFNTRRDRLQRIFGQQRQGVIVKASEEQVRAMSRHEEGGIWPFGGESKGTINIYQRPHTSNQYQGLHEVDAQSYRQLRDLTLVSLANITQGAMTAPHYNSKATKIALVVDGEGYFEMACPHMSRSRGSYQGETRGRPSYQRVASRLTRGTVIIIPAGHPFVAVASSNQLQVLCFEVNANNNEKFLPAGRRNVMMNQLEREAKELAFAFMPAREVEEVSRSQQEEFFKGPQQQGRADA
Ses i 4	> AAG23840.1 oleosin [<i>Sesamum indicum</i>]	MADDRDRPHPHQIQVHPQHPHYEGGVKSLLPQKGPSTTQILAIITLLPISGTLCLAGITLVGTLIGLAVATPVFVIFSPVLVPAAILIAGAVTAFITSGAFGLTGLSLSWVLSNFRFRATGQGPLEYAKRGVQEGTLYVGEKTKQAGEAIKSTAKEGGREGTART
Ses i 5	> AAD42942.1 15 oleosin [<i>Sesamum indicum</i>]	<u>MAEHYGOQQQTRAPHLOLOPRA</u> QRVVKAAATAVTAGGSLLVLSGLTLAGTVIALTIATPLLIFSPVLVPAVITIFLLGAGFLASGGFVAALSVLWYRYLTGKHPPGADQLES AKTKLASKAREMKDRAEQFSQPVA GSQTS
Ses i 6	> AAD42944.1 11S globulin precursor [<i>Sesamum indicum</i>]	<u>MVAFKFLALSLSLVSAIAQ</u> TREPRLTQGQCRFORISGAQPSLRIQSEGGTTELWDERQEQQFCAGIVAMRSTIRPNGLSLPNYHPSRPLVYIERGQGLISIMVPGCAETYQVHRSQRTMERTEASEQQDRGSVRDLHQKVHRLRQGDIVAIPSGAAHWCYNDGSEDLVAVSINDVNHLNQLDQKFRAFYLGGVPRSGEQEQQA RQTFHNIFRAFDAELLSEAFNVQETIRRMQSEEEERGLIVMARERMTFVRPDEEEGEQEHRGRQLDNGLEETFCTMKFRTNVESRREADIFSRQAGRVHVVDNRNKLPIKYMDSLAEKGNLYSNALVSPDWSMTGHTIVYVTRGDAQVQVVDHNGQALMNDRVNQGEMFVVPQYYTSTARAGNNGFEWVAFKTTGSPMRSPLAGYTSVIRAMPLQVITNSYQISPNQAQALKMNRGSQSFLSPGGRRS
Ses i 7	> AAK15087.1 11S globulin [<i>Sesamum indicum</i>]	<u>MALTSLSFFIVVTLIRGLSAQ</u> LAGEQDFYQDLQSQQHQKQARTDCRVERLTAQEPTIRFESEAGLTFWDRNNQFECAGVAAVRNVIQPRGLLLPHYNNAPQLLYVVRGRGIQGTVIPGCAETFERDTQPRQDRRRRFRMDRHQKVRQFRQGDILALPAGLTLWFYNNNGEPLITVALLDTGNAANQLDQTFRHFFLAGNPQGGRQSYFGRPQTEKQQGETKNIFNGFDDEILADAFGVDVQTARRLKGQDDLGRIVRAERLDIVLPGEEEERWERDPYSGANGLEETLCTAKLRENLDDEPARADVYNPHGGRISSLSLTLPLVLSWLRLSAEKGVLYRNLVAPHWNLNHAHSIYITRSGRFQVVGHTGRSVFDGVVREGQLIIVPQNYVAKRASQDEGLEWISFKTNDNAMTSQLAGRLSAIRAMPEEVVMTAYQVSRDEARRLKYNREESRVFSSTSRYSWPRSSRPMSPYMPKPFYVLDVIKSM

注:下划线标注为预测所得的信号肽部分

由于翻译后的蛋白质序列不包含信号肽部分,所

以去除信号肽部分进行蛋白质等电点和相对分子质

量的预测. 非信号肽部分的序列使用 DNASTar 中的 Editsep 和 ExPASy 中的 ProtParam 两种预测方式进行预测分析, 得出理论等电点和相对分子质量. 芝麻过敏原等电点与相对分子质量的预测结果见表 2. 两种方法的预测结果相差不大, 可为芝麻过敏原的分离与纯化提供参考.

表 2 芝麻过敏原等电点与相对分子质量的预测

Tab. 2 Prediction of isoelectric point and relative molecular weight of sesame allergens

过敏原	ExPASy		DNASTar	
	等电点	相对分子质量	等电点	相对分子质量
Ses i 1	7.42	9 364.86	7.52	9 262.83
Ses i 2	5.68	7 282.45	5.78	7 281.61
Ses i 3	7.34	45 792.10	7.37	45 791.27
Ses i 4	9.84	17 373.19	9.88	17 372.20
Ses i 5	10.08	15 193.68	10.12	15 192.69
Ses i 6	7.15	51 541.47	7.48	51 540.65
Ses i 7	8.29	56 111.85	8.14	56 110.99

2.2 芝麻过敏原与其他过敏原的同源性

将芝麻蛋白的氨基酸序列在 SDAP 过敏原数据库中进行同源查询 (BLAST), 获得类似蛋白质, 使用 DNASTar 中的 MegAlign 进行蛋白质的氨基酸序列比对.

根据同源性绘制系统发育树, 将氨基酸序列比对结果同系统发育树比较可以看出, 蛋白质的氨基酸序列相似性越高, 物种之间亲缘关系越近 (图 1). 从图 1 中的系统发育树也可以进一步看出芝麻 Ses i 1、Ses i 2 与腰果 Ana o 3、榛子 Cor a 14.010 1、Jug r 1 同源性最高, 其次是花生 Ara h 2、Ara h 6、Ara h 7, 再次是芝麻 Ses i 3、羽扇豆 Lup an 1、酪蛋白 Bos d 11 等. 由此可见, 亲缘关系较远的一些物种间, 也有序列相似性很高的蛋白质, 它们之间可能会存在一定程度的交叉反应.

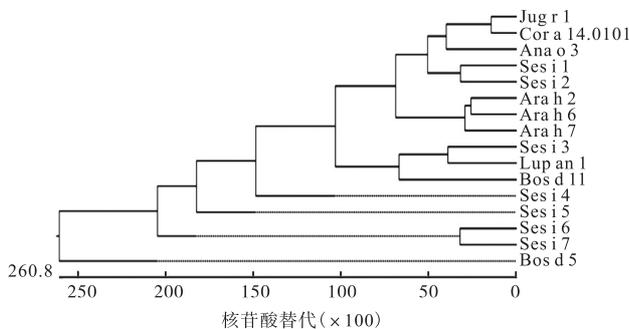


图 1 DNASTar 序列比对结果

Fig. 1 Results of the DNASTar sequence alignment

2.3 芝麻过敏原蛋白的二级结构

通过 DNASTar 软件中 Chou-Fasman 方案和

Garnier-Robson 方案对芝麻过敏原的二级结构进行预测, 结果如图 2 所示, 其中红色表示 α -螺旋结构, 蓝色表示 β -折叠结构, 绿色表示 β -转角结构, 粉色表示 Garnier-Robson 方案的无规则卷曲结构. Chou-Fasman 方案没有对无规则卷曲结构的预测功能.

由图 2 可知: 两种方案分析结果均显示在芝麻过敏原的二级结构中, 除了 Ses i 4 和 Ses i 5 外, 其余芝麻蛋白的 β -转角结构和 α -螺旋结构所占比例较大, β -折叠结构和无规则卷曲结构所占比例较小. β -转角结构和 α -螺旋结构与蛋白质的稳定性相关, 其占比高说明了这些芝麻蛋白的结构较紧密, 不易被降解, 空间结构稳定, 因此空间表位也可以稳定存在. 芝麻蛋白的紧密三级结构可以与人体的 IgE 抗体紧密结合, 从而导致过敏. 通常情况下, 经过深加工的食品还是具有较高的致敏性, 因为处理方式不会使蛋白质的三级结构被完全破坏, 所以加工后空间表位和线性表位还会共存. 因此, 由芝麻蛋白的二级结构分析可知, 这些高紧密性的蛋白质存在较高的潜在致敏性.

2.4 芝麻过敏原的序列性质分析

DNASTar 软件分为多个模块, 每个模块有自己特定的功能与算法. Kyte-Doolittle 方法 (蛋白质的亲水性)、Emini 方法 (蛋白质的溶剂可及性)、Jameson-Wolf 方法 (抗原指数) 和 Karplus-Schultz 方法 (序列可塑性/柔韧性) 具有不同的预测功能. 芝麻过敏原 (Ses i 1—Ses i 7) 序列性质分析结果如图 3 所示, 其中粉色表示蛋白质的亲水性分析, 红色表示序列可塑性分析, 绿色表示抗原指数分析, 蓝色表示蛋白质的溶剂可及性分析. 整体来看, 所有芝麻过敏原蛋白 N 端信号肽部分的疏水区域比较集中, 而亲水性区域分布很广, 几乎全部序列都有分布. 此外, 芝麻过敏原蛋白的溶剂可及性区域以及序列可塑性较高区域在序列分子中分布较广泛, 且在哪些区域中, 这些氨基酸残基可能分布在分子表面, 因此形成抗原表位的可能性比较大. 综合分析 DNASTar 中的 4 个方案: 在大于 0 的区域, 4 个部分的氨基酸序列几乎都是相同的, 其大于 0 的区域分布也比较广, 因此推测可能存在多个抗原表位, 从而提高了刺激机体产生免疫应答的概率. 序列可塑性高表明柔韧性强, 所以在空间结构上更容易折叠弯曲, 因此可以有更高的概率形成结构域和一些二硫键之间的功能区, 这些区域存在空间表位的可能性较高.

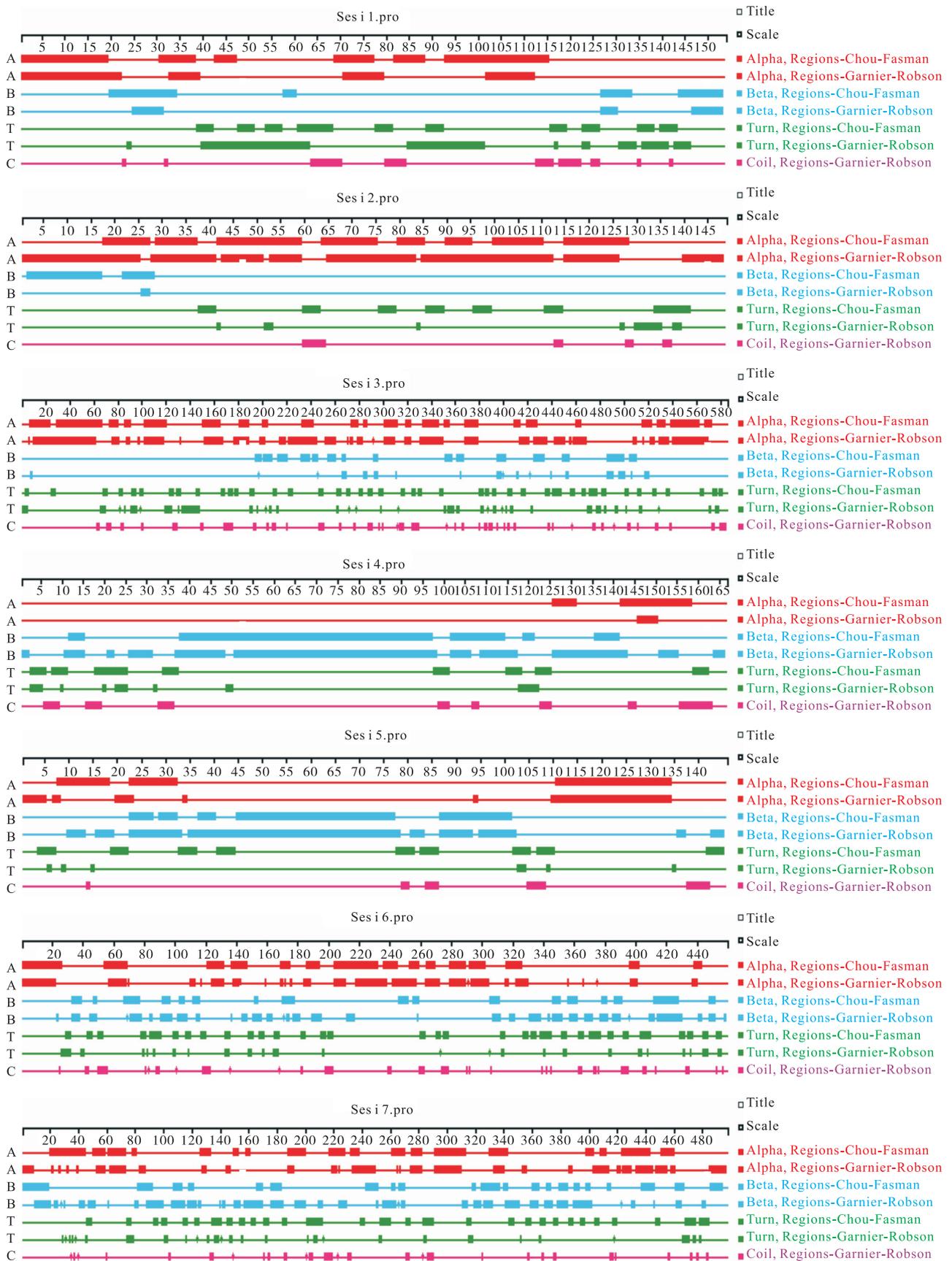


图 2 芝麻过敏原的二级结构预测

Fig. 2 Secondary structure prediction of sesame allergens



图3 芝麻过敏原的序列性质分析
Fig. 3 Sequence analysis of sesame allergens

2.5 芝麻过敏原抗原表位的综合分析预测

综合分析 DNASTar Protean 和 IEDB 两个软件对蛋白质序列性质的预测结果, 选择序列中溶剂可及性高(区域大于 0)、抗原指数高(区域大于 0)、亲水性强(区域大于 0)和可塑性好(区域大于 0)的区域作为可能存在的抗原表位. 表 3 中列出了两个服务器各自所预测的芝麻过敏原抗原表位的结果. 由表 3 可知, DNASTar Protean 与 IEDB 两者预测所得的抗原表位结果较为一致. 两种方法的表位重叠区域作为最终的抗原表位结果: Ses i 1 的线性表位为 33~77、79~102、112~141, Ses i 2 的线性表位为 30~49、51~53、81~93、110~121、133~143, Ses i 3 的线性表位为 29~191、211~233、467~487、526~582, Ses i 4 的线性表位为 4~11、117~131、155~163, Ses i 5 的线性表位为 2~15、101~114、132~141, Ses i 6 的线性表位为 23~65、180~190、405~413、430~444, Ses i 7 的线性表位为 26~51、115~159、187~196、204~232、418~425、467~491.

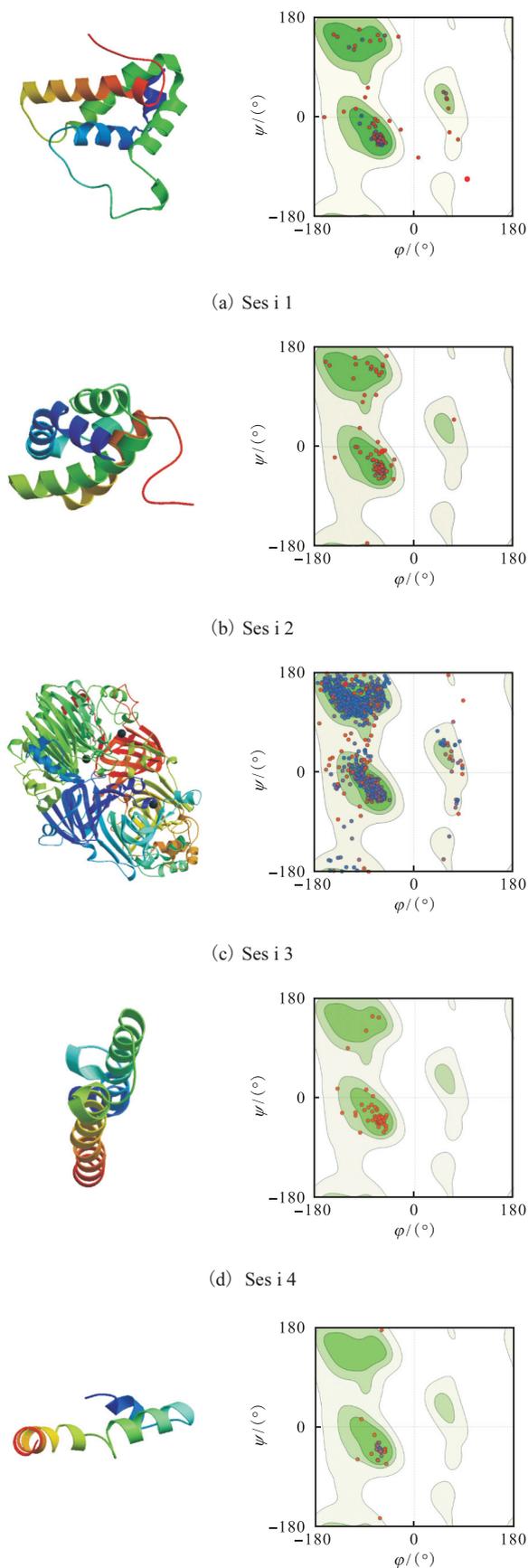
表 3 芝麻过敏原抗原表位预测

Tab. 3 Prediction of sesame allergen epitope

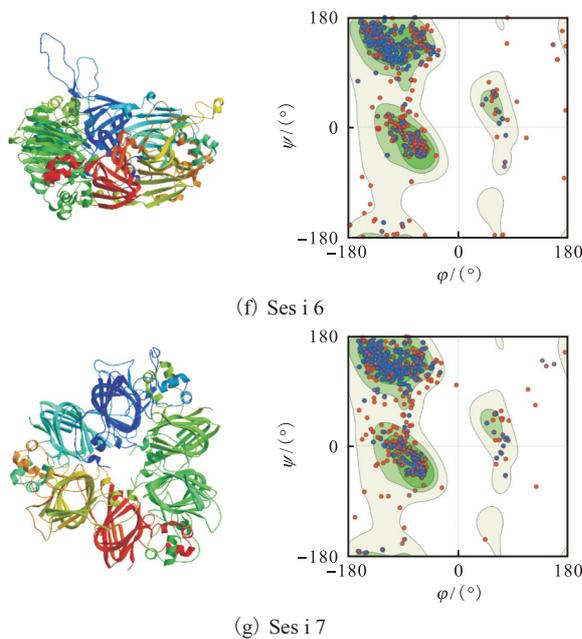
预测工具	过敏原	预测表位的序列位置
DNASTar Protean	Ses i 1	33 ~ 77, 79 ~ 128, 131 ~ 141
	Ses i 2	30 ~ 51, 59 ~ 80, 81 ~ 106, 110 ~ 121, 133 ~ 143
	Ses i 3	29 ~ 191, 211 ~ 235, 264 ~ 288, 470 ~ 497, 529 ~ 582
	Ses i 4	4 ~ 11, 17 ~ 37, 115 ~ 137, 141 ~ 166
	Ses i 5	2 ~ 15, 19 ~ 28, 103 ~ 136, 139 ~ 144
	Ses i 6	23 ~ 65, 78 ~ 103, 182 ~ 213, 235 ~ 306, 390 ~ 417, 439 ~ 458
	Ses i 7	27 ~ 70, 115 ~ 159, 172 ~ 198, 206 ~ 323, 369 ~ 422, 453 ~ 486
IEDB	Ses i 1	32 ~ 89, 97 ~ 102, 112 ~ 149
	Ses i 2	30 ~ 49, 51 ~ 53, 59 ~ 93, 91 ~ 95, 110 ~ 124, 127 ~ 144
	Ses i 3	27 ~ 233, 315 ~ 340, 368 ~ 424, 467 ~ 487, 526 ~ 583
	Ses i 4	4 ~ 37, 117 ~ 131, 134 ~ 143, 155 ~ 163
	Ses i 5	4 ~ 24, 101 ~ 114, 132 ~ 141
	Ses i 6	23 ~ 66, 111 ~ 148, 180 ~ 190, 198 ~ 228, 230 ~ 306, 317 ~ 319, 405 ~ 413, 417 ~ 422, 430 ~ 444
	Ses i 7	26 ~ 51, 53 ~ 80, 126 ~ 154, 187 ~ 196, 204 ~ 232, 328 ~ 331, 418 ~ 425, 430 ~ 437, 445 ~ 458, 467 ~ 491

2.6 芝麻过敏原的三级结构模型构建

芝麻过敏原的三级结构预测如图 4 所示.



(e) Ses i 5



左图为三级结构图,右图为拉氏图

图4 芝麻过敏原的三级结构预测

Fig. 4 Prediction of tertiary structure of sesame allergens

由图4可知, Ses i 1、Ses i 2、Ses i 4、Ses i 5 二级结构中 α -螺旋结构在整个序列中占据主导地位. Ses i 1、Ses i 2 与软件的预测结果相符, Ses i 4、Ses i 5 与预测结果有一定差异. 在之前的二级结构分析中, Ses i 4、Ses i 5 的二级结构中还含有 β -折叠结构, 导致这一结果可能是因为模型选择不当, 但如果选用其他模型, 序列同源性又达不到30%以上, 所以不能进行建模. 在此基础上, Ses i 1、Ses i 2 形成了致密、稳固的球状分子, 具有较强的结构稳定性, 这也为其成为主要的芝麻过敏原提供了潜在的结构基础. Ses i 3、Ses i 6、Ses i 7 主要以 α -螺旋结构和 β -折叠结构为主, 相对分子质量大, 空间结构复杂, 具体的结构特性还需进一步的实验说明.

在模型预测结束后, SWISS-MODEL 服务器会自动生成拉氏图, 以反映预测模型的质量. 拉氏图作为一种平面交角图, 由 α -碳与酰胺组成, 横坐标表示 φ 角, 纵坐标表示 ψ 角. 等高线上的不同颜色区分为允许区和不允许区, 也因此会在拉氏图中形成两种构象, 允许构象和不允许构象. 如图4中拉氏图所示, 其中一个氨基酸残基表示为一个点, 颜色深的区域表明稳定区(也就是允许区), 落在此区域的点越多, 说明该模型越可靠. 拉氏图封闭的最内部区域(深绿色)为允许区(这里的预测结果也最准确和稳定), 越往外即颜色越浅说明越不允许和不稳定. 在最内部区域的任何成对二面角(φ , ψ)所模拟出来的构象都

是立体化学所允许和接受的. 在构象中不存在斥力, 构象能量也就最低, 这是因为非共价键键合原子之间的距离大于等于最小接触距离, 所以两者间的斥力也就消失了, 因此构象是稳定存在的, 这部分的模型评分也最高. 浅绿色区域的任何二面角所规定的主链构象虽然是立体化学允许的, 但是不够稳定(也叫作临界区), 因为在此构象中非共价键键合原子之间的距离虽然小于允许距离, 但仍大于所规定的极限距离^[30]. 除此之外是不允许区, 也就是不稳定区. 由图4可以看出, Ses i 1—Ses i 7 大部分氨基酸残基都落在允许区和临界区, 在一定程度上说明芝麻过敏原 Ses i 1—Ses i 7 的空间构象比较合理, 结合分子势能和动力学分析表明其建模比较成功, 空间构象具有很强的稳定性与合理性.

3 结论

根据蛋白质的氨基酸序列, 利用生物信息学软件, 分析了芝麻过敏原的二级结构、亲水性、表面可及性、抗原性以及柔韧性等. 其中, 亲水性和表面可及性在形成表位时十分关键. 运用 DNASTAR 和 IEDB 方法结合蛋白质特性和二级结构, 预测分析了芝麻7种过敏原的线性表位. 利用 SWISS-MODEL 进行同源建模并预测了三级结构, 对预测结构的稳定性进行评估, 提高了预测的准确性. 结果表明, 7种过敏原均存在多个可能的抗原表位, 在整个氨基酸序列中分布均匀, 并且与同家族已知的过敏原线性表位存在着部分区域氨基酸序列高度相似的情况. 本研究的表位预测结果增加了对芝麻过敏原的认识, 为定位表位缩小了范围, 为抗体制备和免疫检测提供了一定的理论基础.

参考文献:

- [1] GOTO S, TAKAOKA Y, FUTAMURA M, et al. Characteristics of sesame seed allergy and sesame-specific IgE test[J]. Journal of allergy and clinical immunology, 2008, 121(2): S247.
- [2] ALEXIS A, SAFIER B, ROSENTHAL D W, et al. Sesame food allergy and sensitization in children: a growing concern in the USA[J]. Journal of allergy and clinical immunology, 2009, 123(2): S32.
- [3] COHEN A, GOLDBERG M, LEVY B, et al. Sesame food allergy and sensitization in children: the natural history and long-term follow-up[J]. Pediatric allergy and immunology, 2007, 18(3): 217-223.

- [4] SICHERER S H, MUÑOZ-FURLONG A, GODBOLD J H, et al. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up[J]. *Journal of allergy and clinical immunology*, 2010, 125(6) : 1322–1326.
- [5] OSBORNE N J, KOPLIN J J, MARTIN P E, et al. Prevalence of challenge-proven IgE-mediated food allergy using population-based sampling and predetermined challenge criteria in infants[J]. *Journal of allergy and clinical immunology*, 2011, 127(3) : 668–676.
- [6] DALAL I, BINSON I, REIFEN R, et al. Food allergy is a matter of geography after all: sesame as a major cause of severe IgE-mediated food allergic reactions among infants and young children in Israel[J]. *Allergy*, 2002, 57(4) : 362–365.
- [7] 闫瑞. 消费者食品过敏原标签的认知现状与对策研究[D]. 太原:山西医科大学, 2016.
- [8] TUANO K T S, DILLARD K H, GUFFEY D, et al. Development of sesame tolerance and cosensitization of sesame allergy with peanut and tree nut allergy in children[J]. *Annals of allergy asthma & immunology*, 2016, 117(6) : 708–710.
- [9] GUPTA R S, LAU C H, SITA E E, et al. Factors associated with reported food allergy tolerance among US children[J]. *Annals of allergy asthma & immunology*, 2013, 111(3) : 194–198.
- [10] ENTINK R K, REMINGTON B C, BLOM W M, et al. Food allergy population thresholds: an evaluation of the number of oral food challenges and dosing schemes on the accuracy of threshold dose distribution modeling[J]. *Food and chemical toxicology*, 2014, 70: 134–143.
- [11] DANO D, REMINGTON B C, ASTIER C, et al. Sesame allergy threshold dose distribution[J]. *Food and chemical toxicology*, 2015, 83: 48–53.
- [12] WOLFF N, COGAN U, ADMON A, et al. Allergy to sesame in humans is associated primarily with IgE antibody to a 14 kDa 2S albumin precursor[J]. *Food and chemical toxicology*, 2003, 41(8) : 1165–1174.
- [13] PASTORELLO E A V E, FARIOLI L. The major allergen of sesame seeds (*Sesamum indicum*) is a 2S albumin[J]. *Journal of chromatography B: biomedical sciences and applications*, 2001, 756(1) : 85–93.
- [14] BEYER K, BARDINA L, GRISHINA G, et al. Identification of sesame seed allergens by 2 dimensional proteomics and Edman sequencing: seed storage proteins as common food allergens[J]. *Journal of allergy and clinical immunology*, 2002, 110(1) : 154–159.
- [15] LEDUC V, MONERET VAUTRIN D A, TZEN J T, et al. Identification of oleosins as major allergens in sesame seed allergic patients[J]. *Allergy*, 2006, 61(3) : 349–356.
- [16] BEYER K, GRISHINA G, BARDINA L, et al. Identification of 2 new sesame seed allergens: Ses i 6 and Ses i 7[J]. *Journal of allergy and clinical immunology*, 2007, 119(6) : 1554–1556.
- [17] 马秀丽, 张久凯, 孙劲旅, 等. 芝麻过敏原 Ses i 3 同源建模及 B 细胞线性抗原表位预测[J]. *中国食品学报*, 2021, 21(4) : 9–18.
- [18] 单晓红. 大豆和牛奶致敏原表位预测及加工对其抗原性和结构的影响[D]. 无锡:江南大学, 2013.
- [19] 黄影. 隐马尔可夫模型在生物信息学中的应用[J]. *电子科技*, 2015, 28(8) : 185–189.
- [20] YANG X, YU X. An introduction to epitope prediction methods and software[J]. *Reviews in medical virology*, 2009, 19(2) : 77–96.
- [21] VEREDA A, ANDREAE D A, LIN J, et al. Identification of IgE sequential epitopes of lentil (Len c 1) by means of peptide microarray immunoassay[J]. *Journal of allergy and clinical immunology*, 2010, 126(3) : 596–601.
- [22] SHIN D S, COMPADRE C M, MALEKI S J, et al. Biochemical and structural analysis of the IgE binding sites on Ara h 1, an abundant and highly allergenic peanut protein[J]. *Journal of biological chemistry*, 1998, 273(22) : 13753–13759.
- [23] 黄韧, 薛成. 生物信息学:网络资源与应用[M]. 中山:中山大学出版社, 2003: 87–102.
- [24] 蒋彦, 王晓行, 曹毅, 等. 基础生物信息学及应用[M]. 北京:清华大学出版社, 2003: 4–5.
- [25] SÖLLNER J, MAYER B. Machine learning approaches for prediction of linear B-cell epitopes on proteins[J]. *Journal molecular recognition*, 2006, 19(3) : 200–208.
- [26] GREENBAUM J A, ANDERSEN P H, BLYTHE M, et al. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools[J]. *Journal molecular recognition*, 2007, 20(2) : 75–82.
- [27] KYTE J, DOOLITTLE R F. A simple method for displaying the hydropathic character of a protein[J]. *Journal molecular biology*, 1982, 157(1) : 105–132.
- [28] KARPLUS P A, SCHULZ G E. Prediction of chain flexibility in proteins: a tool for the selection of peptide anti-

- [2] MAZARÉ L. Using unification for opacity properties[R/OL]. (2004-10-28) [2022-10-12]. <https://www-verimag.imag.fr/TR/TR-2004-24.pdf>.
- [3] SABOORI A, HADJICOSTIS C N. Notions of security and opacity in discrete event systems[C]//2007 46th IEEE Conference on Decision and Control. New Orleans: IEEE, 2007: 5056-5061.
- [4] SABOORI A, HADJICOSTIS C N. Verification of initial-state opacity in security applications of discrete event systems[J]. Information sciences, 2013, 246: 115-132.
- [5] SABOORI A, HADJICOSTIS C N. Verification of K -step opacity and analysis of its complexity[J]. IEEE Transactions on automation science and engineering, 2011, 8(3): 549-559.
- [6] SABOORI A, HADJICOSTIS C N. Verification of infinite-step opacity and complexity considerations[J]. IEEE Transactions on automatic control, 2011, 57(5): 1265-1269.
- [7] BRYANS J W, KOUTNY M, MAZARÉ L, et al. Opacity generalized to transition systems[J]. International journal of information security, 2008, 7(6): 421-435.
- [8] LIN F. Opacity of discrete event systems and its applications[J]. Automatica, 2011, 47(3): 496-503.
- [9] FALCONE Y, MARCHAND H. Enforcement and validation (at runtime) of various notions of opacity[J]. Discrete event dynamic systems, 2015, 25(4): 531-570.
- [10] MA Z Y, YIN X, LI Z W. Verification and enforcement of strong infinite- and k -step opacity using state recognizers[J]. Automatica, 2021, 133: 109838.
- [11] HAN X G, ZHANG K Z, ZHANG J H, et al. Strong current-state and initial-state opacity of discrete-event systems[J]. Automatica, 2023, 148: 110756.
- [12] HAN X G, ZHANG K Z, LI Z W. Verification of strong K -step opacity for discrete-event systems[EB/OL]. (2022-04-10) [2022-10-20]. <https://arxiv.org/pdf/2204.04698.pdf>.
- [13] CASSEZ F, DUBREIL J, MARCHAND H. Synthesis of opaque systems with static and dynamic masks[J]. Formal methods in system design, 2012, 40(1): 88-115.
- [14] WU Y C, LAFORTUNE S. Comparative analysis of related notions of opacity in centralized and coordinated architectures[J]. Discrete event dynamic systems, 2013, 23(3): 307-339.
- [15] BALUN J, MASOPUST T. Comparing the notions of opacity for discrete-event systems[J]. Discrete event dynamic systems, 2021, 31(4): 553-582.
- [16] CASSANDRAS C G, LAFORTUNE S. Introduction to discrete event systems[M]. Boston: Springer, 2008.

责任编辑: 周建军

(上接第 27 页)

- gens[J]. Naturwissenschaften, 1985, 72: 212-213.
- [29] BIASINI M, BIENERT S, WATERHOUSE A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information[J]. Nucleic acids research, 2014, 42(1): 252-258.
- [30] 郑礼娜. 虾类过敏原的活性分析及其抗原表位的研究[D]. 青岛: 中国海洋大学, 2011.

责任编辑: 郎婧