



(12) 发明专利

(10) 授权公告号 CN 116768990 B

(45) 授权公告日 2023. 11. 07

(21) 申请号 202311033251.5

C12N 15/82 (2006.01)

(22) 申请日 2023.08.16

A01N 37/46 (2006.01)

(65) 同一申请的已公布的文献号

A01P 7/04 (2006.01)

申请公布号 CN 116768990 A

A01H 5/00 (2018.01)

C12R 1/19 (2006.01)

(43) 申请公布日 2023.09.19

(56) 对比文件

(73) 专利权人 莱肯生物科技(海南)有限公司

WO 2020064408 A1, 2020.04.02

地址 572000 海南省三亚市崖州区崖州湾

CN 101225114 A, 2008.07.23

科技城雅布伦产业园3号楼3306室

CN 101508725 A, 2009.08.19

专利权人 武汉莱肯博奥科技有限公司

CN 107109418 A, 2017.08.29

(72) 发明人 张素洁 张玉静 李晨

WO 2016184387 A1, 2016.11.24

CN 110062579 A, 2019.07.26

(51) Int. Cl.

审查员 段珊

C07K 14/325 (2006.01)

C12N 15/31 (2006.01)

C12N 15/70 (2006.01)

C12N 1/21 (2006.01)

权利要求书1页 说明书4页

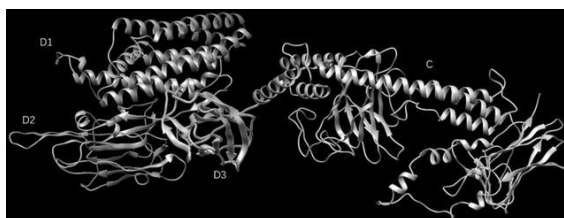
序列表(电子公布) 附图2页

(54) 发明名称

一种人工智能辅助生成的杀虫蛋白

(57) 摘要

本发明公开了一种人工智能辅助生成的杀虫蛋白及其应用,属于基因工程领域。本发明基于人工智能算法的结构预测,结合批量活性检测,筛选获得一批新型的杀虫蛋白。这些杀虫蛋白在制备杀虫剂或培育基因工程植物领域具有应用价值。



1. 一种蛋白,其特征在于,该蛋白的氨基酸序列如SEQ ID NO. 1或SEQ ID NO. 7所示。
2. 一种核酸分子,其特征在于,所述的核酸分子编码权利要求1所述的蛋白。
3. 根据权利要求2所述的核酸分子,其特征在于,所述核酸分子的核苷酸序列如SEQ ID NO. 11或SEQ ID NO. 17所示。
4. 一种载体,其特征在于,所述载体包含权利要求2~3任一项所述的核酸分子。
5. 一种重组细胞,其特征在于,所述重组细胞含有权利要求2~3任一项所述的核酸分子或权利要求4所述的载体。
6. 根据权利要求5所述的重组细胞,其特征在于,所述重组细胞为原核生物细胞。
7. 根据权利要求6所述的重组细胞,其特征在于,所述重组细胞为大肠杆菌细胞。
8. 权利要求1所述的蛋白,或权利要求2~3任一项所述的核酸分子,或权利要求4所述的载体,或权利要求5~7任一项所述的重组细胞在抗草地贪夜蛾或制备抗草地贪夜蛾制剂或培育抗草地贪夜蛾植物中的应用。

一种人工智能辅助生成的杀虫蛋白

技术领域

[0001] 本发明公开了一种人工智能辅助生成的杀虫蛋白及其应用,属于基因工程领域。

背景技术

[0002] 源于苏云金芽孢杆菌中的杀虫蛋白是目前转基因作物和生物农药开发过程中普遍使用的目的蛋白。然而随着这些商品化的杀虫剂和抗虫作物的大规模应用,害虫普遍产生了对已有杀虫蛋白的抗药性,迫切需要找到或创制出新型的杀虫蛋白。

[0003] 近年来,农业农村部根据《农作物病虫害防治条例》有关规定制定的《一类农作物病虫害名录》中,草地贪夜蛾被定为影响作物种植的一类农作物虫害。草地贪夜蛾(学名: *Spodoptera frugiperda*)属于夜蛾科灰翅夜蛾属,其幼虫可大量啃食水稻、甘蔗和玉米之类禾本科以及菊科、十字花科等多种农作物,造成严重的经济损失。该物种原产于美洲热带地区,具有很强的迁徙能力,2016年起,草地贪夜蛾散播至非洲、亚洲各国,并于2019年出现在中国,已造成巨大的农业损失。美国等转基因作物种植大国主要使用Cry1Fa、Vip3Aa蛋白培育玉米和大豆新品种防治草地贪夜蛾,然而由于该昆虫强迁徙、繁殖快的特点,越来越多的针对Cry1Fa、Vip3Aa的抗性品系已经被发现(Fatoretto J C, Michel A P, Silva Filho M C, Silva N. Adaptive potential of fall armyworm (Lepidoptera: Noctuidae) limits Bt trait durability in Brazil[J]. J. Integr. Pest Manag. 2017, 8, 17)。因此,未来的抗草地贪夜蛾作物品种需要获得与Cry1Fa和Vip3Aa完全不同的新型杀虫蛋白。

[0004] 蛋白质在细胞中的功能是由其三维结构决定的。但通过实验测定蛋白质结构费时费力。随着人工智能技术的发展,蛋白结构预测已经取得了突破性的进展。这为海量研究和创制蛋白结构提供了新的机会。谷歌DeepMind研发出的AlphaFold人工智能网络已经给出了几十万蛋白质的三维结构,包括人体自身能制造的每一种蛋白质,这些成果有望为医学和药物设计领域带来更大惊喜。虽然更加精准的蛋白折叠、配体结合等细节上还有很多需要优化的空间,但通过深度学习、神经网络的人工智能算法,结合实验层面的功能验证和筛选的方式,能够实现从蛋白结构预测、实验测试、优化改进,到最终生成新的工程化功能性蛋白。

发明内容

[0005] 为了解决上述问题,本发明采用如下技术方案:

[0006] 本发明提供一种人工智能辅助生成的杀虫蛋白,其特征在于,该蛋白的氨基酸序列如SEQ ID NO. 1或SEQ ID NO. 7任意一个所示。

[0007] 本发明还提供一种核酸分子,其特征在于,所述的核酸分子编码上述的蛋白。

[0008] 在一些实施方案中,编码SEQ ID NO. 1蛋白的核酸分子的核苷酸序列如SEQ ID NO. 11所示,编码SEQ ID NO. 7蛋白的核酸分子的核苷酸序列如SEQ ID NO. 17所示。

[0009] 本发明还提供一种载体,其特征在于,所述载体包含上述的核酸分子。

[0010] 本发明还提供一种重组细胞,其特征在于,所述重组细胞含有上述的核酸分子或上述的载体。

[0011] 在一些实施方案中,上述重组细胞为原核生物细胞。

[0012] 在一些实施方案中,上述重组细胞为大肠杆菌细胞。

[0013] 本发明还提供上述蛋白、核酸分子、载体、重组细胞在抗虫或制备抗虫制剂或培育抗虫植物中的应用。

[0014] 在一些实施方案中,上述抗虫为对草地贪夜蛾具有杀虫活性。

[0015] 本发明的有益效果在于:本发明基于人工智能算法的结构预测,结合批量活性检测筛选,本发明获得了2种具有更高杀虫活性的新型草地贪夜蛾杀虫蛋白,可用于开发生物农药或培育新型抗虫植物。

附图说明

[0016] 图1 eCry1Gb.1Ig蛋白的结构生成图。D1、D2、D3分别标识蛋白核心杀虫区的三个结构域;C标识蛋白C端结构域。

[0017] 图2 WBY-1~WBY-10与eCry1Gb.1Ig氨基酸序列相似性分析。

[0018] 图3 WBY-1蛋白的结构生成图。D1、D2、D3分别标识蛋白核心杀虫区的三个结构域;C标识蛋白C端结构域。

[0019] 图4 WBY-7蛋白的结构生成图。D1、D2、D3分别标识蛋白核心杀虫区的三个结构域;C标识蛋白C端结构域。

具体实施方式

[0020] 提供以下定义和方法用以更好地界定本申请以及在本申请实践中指导本领域普通技术人员。除非另作说明,术语按照相关领域普通技术人员的常规用法理解。本文所引用的所有专利文献、学术论文、行业标准及其他公开出版物等,其中的全部内容整体并入本文作为参考。

[0021] 以下实施例用于说明本发明,但不用来限制本发明的范围。在不背离本发明精神和实质的情况下,对本发明方法、步骤或条件所作的修改或替换,均属于本申请的范围。若无特别指明,实施例按照常规实验条件,如Sambrook等人的分子克隆实验手册(Sambrook J&Russell DW, Molecular cloning: a laboratory manual, 2001),或按照制造厂商说明书建议的条件。若未特别指明,实施例中所用的化学试剂均为常规市售试剂,实施例中所用的技术手段为本领域技术人员所熟知的常规手段。

[0022] 实施例1新型蛋白的结构预测和序列生成

[0023] eCry1Gb.1Ig是一种嵌合的Cry类草地贪夜蛾杀虫蛋白,由Cry1Gb作为骨架,并将第三结构域替换成Cry1Ig而获得,对已商业化的草地贪夜蛾杀虫蛋白Cry1Fa、Vip3Aa、Cry1A.105/Cry2Ab均没有交互抗性(Chae H, Wen Z, Hootman T, et al. eCry1Gb.1Ig, A Novel Chimeric Cry Protein with High Efficacy against Multiple Fall Armyworm (*Spodoptera frugiperda*) Strains Resistant to Different GM Traits[J]. *Toxins* (Basel), 2022, 14(12).)。

[0024] 本发明使用eCry1Gb.1Ig蛋白作为模板进行结构分析和拟合,以形成新的杀虫蛋

白。首先使用AlphaFold2、RoseTTAFold在线工具预分域组装生成eCry1Gb.1Ig蛋白结构(见图1)。

[0025] 接着基于深度学习、神经网络,利用数据库和PROTEINGAN算法预测工具(Repecka, D., Jauniskis, V., Karpus, L. et al. Expanding functional protein sequence spaces using generative adversarial networks. Nat Mach Intell 3, 324-333 (2021)),建立模型,生成和设计蛋白质序列,并继续使用AlphaFold2、RoseTTAFold在线工具生成蛋白结构。由于eCry1Gb.1Ig的核心杀虫区的第三结构域(Domain 3, D3)在识别草地贪夜蛾肠道受体中具有重要的作用,因此在生成新的蛋白序列和结构时保持eCry1Gb.1Ig蛋白的D3结构域序列不变,只将D1和D2结构域以及C端结构域进行拟合生成。

[0026] 计算生成序列和参考序列的TM-score(Chengxin Zhang, Morgan Shine, Anna Marie Pyle, Yang Zhang. US-align: Universal Structure Alignment of Proteins, Nucleic Acids and Macromolecular Complexes. Nature Methods, 19: 1109-1115 (2022)),选择得分高的序列。然后使用Chimera 1.17可视化比较生成蛋白结构。

[0027] 最终,获得10个生成的新蛋白,命名为WBY-1~WBY-10,序列如SEQ ID NO. 1~ SEQ ID NO. 10所示。从序列相似性上看,这10个蛋白与eCry1Gb.1Ig的相似度很低,最高的也不超过42%(图2)。这为从根本上创制出全新的蛋白提供了基础。

[0028] 实施例2测试新生成蛋白的杀虫效果

[0029] 使用蛋白表达实验系统合成这些蛋白实物,并测试它们对草地贪夜蛾的杀虫效果。

[0030] 首先根据氨基酸序列设计编码该序列的核酸序列(在如下在线工具中开展<http://www.friendbio.com/codon.html>),其中密码子设置为大肠杆菌(K12菌株)偏爱性,并避免XhoI和HindIII酶切位点。获得编码WBY-1~WBY-10的核酸序列(分别如SEQ ID NO. 11~SEQ ID NO. 20所示)。

[0031] 人工合成上述序列的核酸核酸分子,将其与编码eCry1Gb.1Ig蛋白的核酸分子一起,分别克隆到载体pET28a表达载体中限制性内切酶XhoI和HindIII的位点之间,获得蛋白表达载体。将该载体转入大肠杆菌BL21细胞系,并进行蛋白表达。具体步骤如下:

[0032] 单个菌落接种到0.5 mL LB液体培养基,在37°C下培养4 h至培养基浑浊,取100 uL菌液加入 IPTG(Isopropyl-β-D-thiogalactoside)至终浓度为0.8 mM,同时取100 uL菌液作为阴性对照,继续培养4 h,在100 uL菌液中加入25 uL 上样缓冲液制样电泳,根据阴性对照和加入IPTG诱导的结果对比,判断是否有表达。有表达的取剩余20 uL,接种到2 mL LB液体培养基中37°C下培养12-16 h作为种子液,种子液再接种到250 mL LB液体培养基至OD600=0.5-0.6,然后加IPTG(Isopropyl-β-D-thiogalactoside)至浓度为0.8 mM,并继续在这样的条件下培养4小时。培养液经过5000 g离心10分钟沉淀大肠杆菌细胞,然后弃上清收集沉淀。沉淀中加30mL 20mM Tris-50mM NaCl缓冲液,超声破碎。离心后检测上清液是否含有重组蛋白。

[0033] 进一步将上述实验获得的重组蛋白进行杀虫活性测试。具体为:

[0034] 采用表面涂抹方法进行生物测定,先在24孔板中先加入约1 mL未凝固的人工饲料(约0.5 g),轻微晃动使饲料铺满孔板底部,待饲料凝固后,再加入不同浓度的蛋白溶液(10 μL/孔),加入后轻轻晃动使药液均匀地平铺在饲料表面上,在通风橱内自然风干1 h。实验

设置6个梯度浓度(0.01、0.1、0.5、1、2、5 $\mu\text{g/g}$)和一个空白对照(缓冲液),每个处理接24头人工饲养的草地贪夜蛾(*Spodoptera frugiperda*)初孵幼虫(孵化时间为2~12 h),设3次重复,放置在温度 $25 \pm 2^\circ\text{C}$,光周期14:10(L:D)h,相对湿度50-70%的养虫室培养,7天后调查死亡率。以用毛笔轻触幼虫尾部,幼虫不动视为死亡,幼虫未发育2龄的也视为死亡。

[0035] 根据如下公式计算死亡率和校正死亡率,并利用graphpad计算LC50值。

$$[0036] \quad \text{死亡率}(\%) = \frac{\text{接虫数} - \text{存活虫数} + \text{虫龄仍处于2龄以下的虫数}}{\text{接虫数}} \times 100 \quad (\text{公式1})$$

$$[0037] \quad \text{校正死亡率}(\%) = \frac{\text{处理死亡率} - \text{对照死亡率}}{100 - \text{对照死亡率}} \times 100 \quad (\text{公式2})$$

[0038] 测试结果表明,这10个新生成的蛋白中,WBY-1和WBY-7的杀虫活性超过了eCry1Gb.1Ig,可以作为一种活性更加的新型杀虫蛋白用于生物农药和抗虫转基因植物的开发。蛋白的杀虫活性数据如表1所示。WBY-1和WBY-7的蛋白结构图如图3和图4所示。

[0039] 表1蛋白的杀虫活性

[0040]

蛋白名称	氨基酸序列	核酸序列	杀虫活性(LC50 ¹)
eCry1Gb.1Ig	SEQ ID NO. 21	SEQ ID NO. 22	0.09
WBY-1	SEQ ID NO. 1	SEQ ID NO. 11	0.06*
WBY-2	SEQ ID NO. 2	SEQ ID NO. 12	0.69
WBY-3	SEQ ID NO. 3	SEQ ID NO. 13	1.67
WBY-4	SEQ ID NO. 4	SEQ ID NO. 14	1.07
WBY-5	SEQ ID NO. 5	SEQ ID NO. 15	3.35
WBY-6	SEQ ID NO. 6	SEQ ID NO. 16	1.13
WBY-7	SEQ ID NO. 7	SEQ ID NO. 17	0.07*
WBY-8	SEQ ID NO. 8	SEQ ID NO. 18	0.35
WBY-9	SEQ ID NO. 9	SEQ ID NO. 19	0.76
WBY-10	SEQ ID NO. 10	SEQ ID NO. 20	0.89

[0041] “*”表示相对于eCry1Gb.1Ig蛋白有显著差异($\alpha=0.05$)。1:单位 $\mu\text{g/g}$ 。

[0042] 虽然,上文中已经用一般性说明及具体实施方案对本发明作了详尽的描述,但在本发明基础上,可以对之作一些修改或改进,这对本领域技术人员而言是显而易见的。因此,在不偏离本发明精神的基础上所做的这些修改或改进,均属于本发明要求保护的范围。



图 1

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	WRKY_7		720	720	100%	0.0	39.35%	1169	Query_55106
<input checked="" type="checkbox"/>	WRKY_3		719	719	100%	0.0	38.53%	1169	Query_55102
<input checked="" type="checkbox"/>	WRKY_1		711	711	87%	0.0	41.68%	1169	Query_55100
<input checked="" type="checkbox"/>	WRKY_6		707	707	100%	0.0	38.35%	1169	Query_55105
<input checked="" type="checkbox"/>	WRKY_5		700	700	89%	0.0	41.32%	1169	Query_55104
<input checked="" type="checkbox"/>	WRKY_4		689	689	100%	0.0	39.00%	1169	Query_55103
<input checked="" type="checkbox"/>	WRKY_8		671	714	92%	0.0	40.06%	1169	Query_55107
<input checked="" type="checkbox"/>	WRKY_9		666	666	100%	0.0	37.72%	1169	Query_55108
<input checked="" type="checkbox"/>	WRKY_2		655	655	100%	0.0	38.42%	1169	Query_55101
<input checked="" type="checkbox"/>	WRKY_10		649	649	87%	0.0	39.34%	1169	Query_55109

图 2

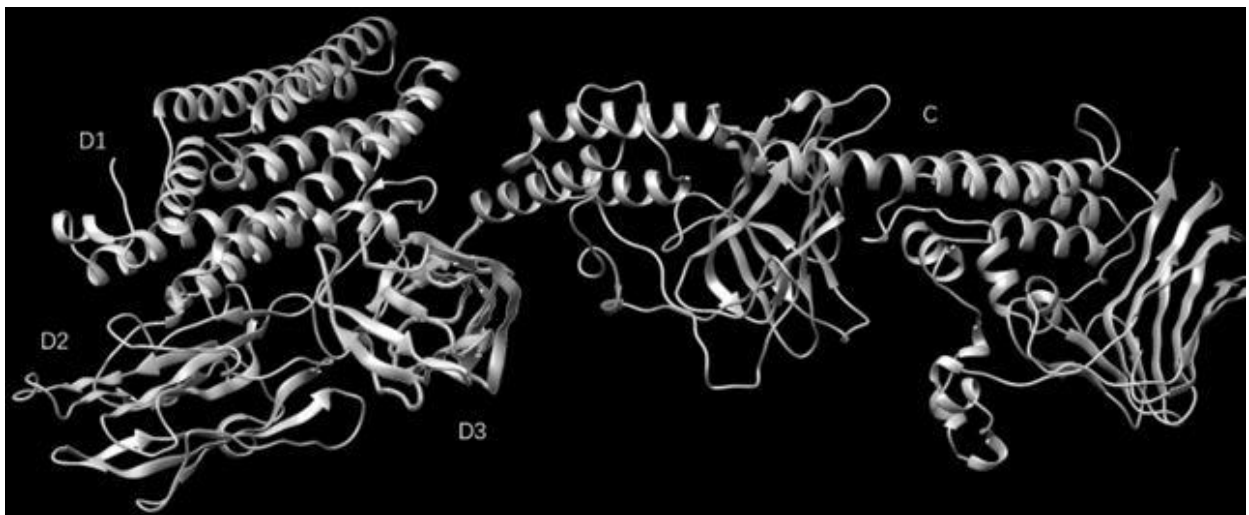


图 3

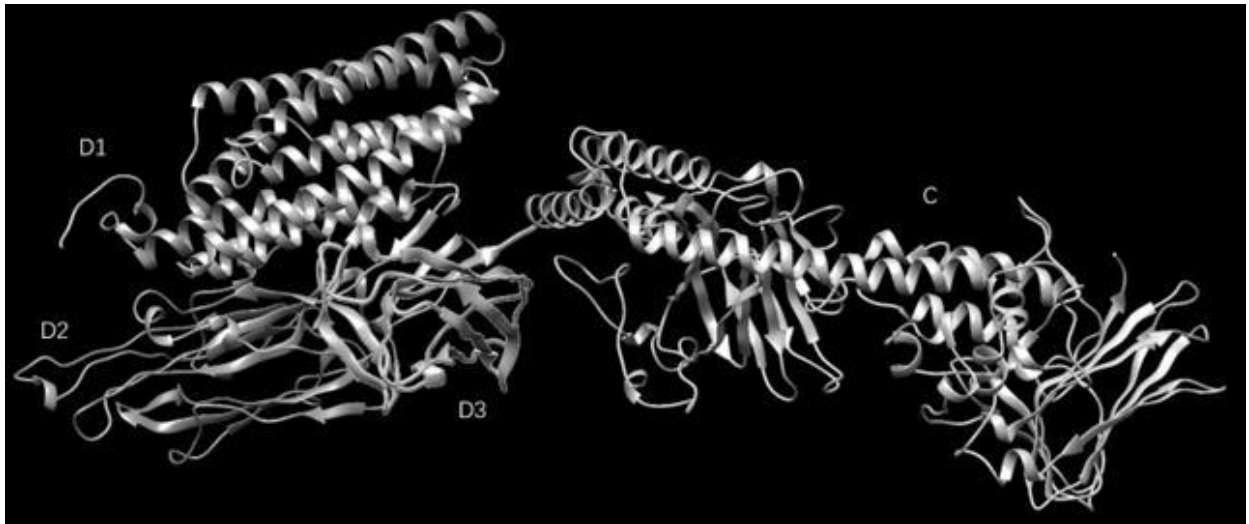


图 4