

## 棉铃虫 CCE001a 蛋白结构与功能的生物信息学分析

张莉, 王海亮, 马雪儿, 牛俊丽, 陈程\*, 张文举\* (石河子大学动物科技学院, 新疆石河子 832000)

**摘要** [目的]深入预测 CCE001a 蛋白结构与生物学功能。[方法]使用生物信息学方法预测 CCE001a 基因所编码的蛋白质, 预测其功能与结构。[结果]该蛋白的总氨基酸残基数为 556 个, 分子式为  $C_{2865}H_{4321}N_{737}O_{811}S_{23}$ , 蛋白的等电点 pI 为 5.32, 带正电的氨基酸残基 (Arg+Lys) 为 58 个, 带负电的氨基酸残基 (Asp+Glu) 为 75 个, 蛋白不稳定系数为 35.22, 脂肪系数为 75.91, 总平均亲水性系数为 -0.262; CCE001a 蛋白有一个脱氢酶超家族 (cl21494) 结构域; 经分析显示,  $\alpha$ -螺旋、无规律的卷曲是这种蛋白质的主要二次结构, 分别占 31.35% 和 46.49%, 还包括 3.96% 的  $\beta$ -转角和 18.20% 的扩展延伸链。经过三级结构学预测表明, 该蛋白包括  $\alpha$ -螺旋、 $\beta$ -转角; 该蛋白亚细胞被确认为定位在内质网内, 存在着信号肽的序列, 初步可以认为是具有分泌力的亲水蛋白, 且没有出现跨膜蛋白, 共有 37 个磷酸化位点, 1 个 N-糖基化位点。[结论]该研究可为研究 CCE001a 的解毒机制提供理论基础。

**关键词** CCE001a; 羧酸酯酶; 解毒; 蛋白质修饰位点; 生物信息

中图分类号 S435.622+.3 文献标识码 A

文章编号 0517-6611(2022)01-0102-04

doi:10.3969/j.issn.0517-6611.2022.01.026

开放科学(资源服务)标识码(OSID):



### Bioinformatics Analysis of the Protein Structure and Function of *Helicoverpa armigera* CCE001a

ZHANG Li, WANG Hai-liang, MA Xue-er et al (College of Animal Science and Technology, Shihezi University, Shihezi, Xinjiang 832000)

**Abstract** [Objective] In order to further predict the protein structure and biological function of CCE001a. [Method] This study used a bioinformatics method to predict the protein encoded by this gene, and predict its function and structure. [Result] The prediction results show that the total number of amino acid residues of the protein is 556, its molecular formula is  $C_{2865}H_{4321}N_{737}O_{811}S_{23}$ , the isoelectric point pI of the protein is 5.32, and the positively charged amino acid residues (Arg+Lys) are 58, which are negatively charged. There are 75 amino acid residues (Asp+Glu), protein instability coefficient is 35.22, fat coefficient is 75.91, total average hydrophilicity coefficient is -0.262; CCE001a protein has a dehydrogenase superfamily (cl21494) domain; analysis shows that  $\alpha$ -helix and irregular coils are the main secondary structures of this protein. They account for 31.35% and 46.49%, respectively, including 3.96% of  $\beta$ -turns and 18.20% of extended stretched chains. After tertiary structural prediction, it is shown that the protein includes  $\alpha$ -helix and  $\beta$ -turn; the protein subcellular is confirmed to be located in the endoplasmic reticulum, and there is a signal peptide sequence, which can be initially considered as secretory Hydrophilic protein, and no transmembrane protein, a total of 37 phosphorylation sites, one N-glycosylation site. [Conclusion] This study can provide a theoretical basis for studying the detoxification mechanism of CCE001a.

**Key words** CCE001a; Carboxylesterase; Detoxification; Protein modification sites; Biological information

羧酸酯酶 (CarEs) 是一类主要的解毒酶, 与各种昆虫的抗药性有关<sup>[1]</sup>。羧酸酯酶普遍存在于动物、植物和微生物中, 在包括羧酸酯在内的异生化合物水解成相应的醇和羧酸的过程中发挥着关键作用<sup>[2]</sup>。昆虫中参与多种类型杀虫剂解毒作用的物质可能是 1 个或多个酶系, 这些杀虫剂包括拟除虫菊酯、有机磷和氨基甲酸酯。在人类大量使用有机磷、拟除虫菊酯杀虫剂的形势下, 对棉花危害较大的害虫棉铃虫 (*Helicoverpa armigera*) 产生了较严重的抗药性<sup>[3]</sup>。CarEs 是一类主要负责异生素化学物质的解毒酶, 因此在棉铃虫抗药性中起关键作用<sup>[4]</sup>。次生代谢物是植食性昆虫及其寄主植物共同进化中的一组重要介质。由于其幼虫在棉花上觅食和生长, 棉铃虫可通过解毒和螯合机制耐受棉花植物产生的主要防御代谢物棉酚<sup>[5]</sup>。当受到昆虫攻击时, 宿主植物会产生次生代谢物作为防御的一部分, 而昆虫在进食时会产生解毒酶来抵消次生代谢物。通常情况下,

代谢排毒途径一般可以分为 3 个阶段。第 1 阶段的酶, 细胞色素 P450 单加氧酶或羧酸酯酶直接作用于毒素分子, 增加其官能团的亲水性和反应性。第 2 阶段, 疏水性和毒性化合物通过谷胱甘肽 S-转移酶 (GSTs) 的作用转化为亲水性产物, 促进其排泄。最后, 依靠 ATP 结合盒 (ABC) 和主要的膜转运蛋白主动将结合的异源物质泵出细胞<sup>[6-7]</sup>。数据表明, 在棉铃虫中至少发现了 40 种羧酸酯酶<sup>[7]</sup>。目前, 杀虫剂代谢产生活性 CarEs 的另一种方法是利用异源表达系统。迄今为止, 已经用杆状病毒表达系统在昆虫细胞 (Sf9) 中表达了来自棉铃虫 GR 菌株的若干 CarEs<sup>[8]</sup>, 然而, 由于昆虫细胞内细胞生长缓慢和表达水平低, 只能产生有限数量的酶。一些昆虫基因已经在大肠杆菌系统中进行了功能性表达。例如, 利用这种异源表达系统, 来自库蚊的突变 B1 酯酶显示出对 4 种有机磷酸酯杀虫剂的水解酶活性<sup>[9]</sup>。苹果蠹蛾的 CpCE-1 能够代谢乙酰甲胺磷<sup>[10]</sup>。靳明辉<sup>[11]</sup> 采用 qPCR 技术筛选 6 个候选基因, 采用 qPCR 技术对棉铃虫取食棉酚后 6 个候选基因的表达变化进行了研究, CCE001a 基因表达量显著上调 ( $P < 0.05$ ), 由此可知, 基因 CCE001a 与棉铃虫代谢棉酚有关。因此, 了解棉铃虫羧酸酯酶 CCE001a 蛋白质生化特性和生物活性功能, 可能会大大促进对棉铃虫羧酸酯酶解毒棉酚机制的理解。

**基金项目** 国家自然科学基金“棉铃虫棉酚降解酶的高活性筛选及其酶学特性与降解机理”(31860660); 石河子大学高层次人才科研启动资金专项“基于棉酚降解相关基因的热带假丝酵母遗传转化及应用项目”(RCZK2019)。

**作者简介** 张莉 (1997—), 女, 甘肃天水人, 硕士研究生, 研究方向: 饲料资源的开发与利用。\* 通信作者: 陈程, 讲师, 从事饲料资源的开发与利用研究; 张文举, 教授, 从事饲料资源的开发与利用研究。

**收稿日期** 2021-05-24; **修回日期** 2021-06-11

## 1 材料与与方法

棉铃虫 CCE001a 蛋白的生物信息学分析内容和使用的软件及网站见表 1<sup>[12]</sup>。

表 1 分析内容及工具软件

Table 1 Analysis content and tool software

序号 No.	内容 Content	分析软件及网站 Analysis software and website
1	理化性质分析	ExPASy 在线工具 ProtParam
2	保守区域预测	https://www.genome.jp/tools/motif//
3	二级结构及三级结构的预测	SOPMA
4	亚细胞定位预测	PSOR
5	氨基酸信号肽	SignalP4.0 Server
6	亲疏水性分析	ExPASy 在线工具 ProtScale
7	跨膜区域分析	http://www.cbs.dtu.dk/services/TMHMM-2.0/
8	磷酸化位点/糖基化位点分析	http://kinasephos.mbc.nctu.edu.tw/index.php/NetNGlyc 1.0

## 2 结果与分析

**2.1 CCE001a 蛋白理化性质** CCE001a 蛋白酶的理化特性是通过 ExPASy 在线工具 ProtParam 来分析,通过分析表明,棉铃虫 CCE001a 的相应分子质量为 62 802.66。由表 2 可知,棉铃虫 CCE001a 共含 20 种氨基酸。该基因所编码的蛋白质氨基酸总计 556 个,分子式为 C<sub>2865</sub> H

4321N7370811S23,蛋白等电点 PI 为 5.32,不稳定系数计算为 35.22,故将其归类为稳定性蛋白;带正电荷的氨基酸残基总数(Arg + Lys)为 58 个,带负电荷的氨基酸残基总数(Asp + Glu)为 75 个,脂肪系数为 75.91;亲水性的平均值(GRAVY)为 -0.262,为亲水性蛋白。

表 2 棉铃虫 CCE001a 蛋白的氨基酸组成

Table 2 Amino acid composition of *Helicoverpa armigera* CCE001a protein

名称 Name	数目 Number 个	含量 Content %	名称 Name	数目 Number 个	含量 Content %
Ala	45	8.1	Leu	4	7.6
Arg	26	4.7	Lys	32	5.8
Asn	23	4.1	Met	16	2.9
Asp	38	6.8	Phe	38	6.8
Cys	7	1.3	Pro	41	7.4
Gln	11	2.0	Ser	21	3.8
Glu	37	6.7	Thr	30	5.4
Gly	38	6.8	Trp	12	2.2
His	13	2.3	Tyr	20	3.6
Ile	24	4.3	Val	41	7.4

**2.2 CCE001a 蛋白保守区域预测** 该蛋白的保守序列如图 1 所示,发现 22~530 有 1 个共酯酶家族(PF00135)结构域,在 118~206 有 1 个脱氢酶家族(PF07859)结构域。

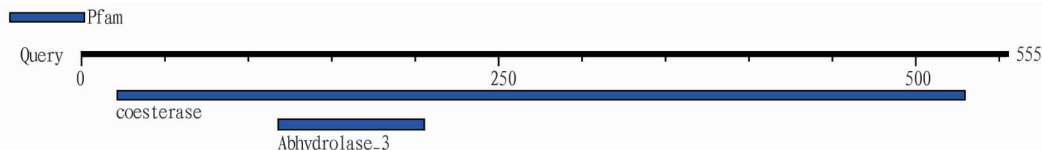
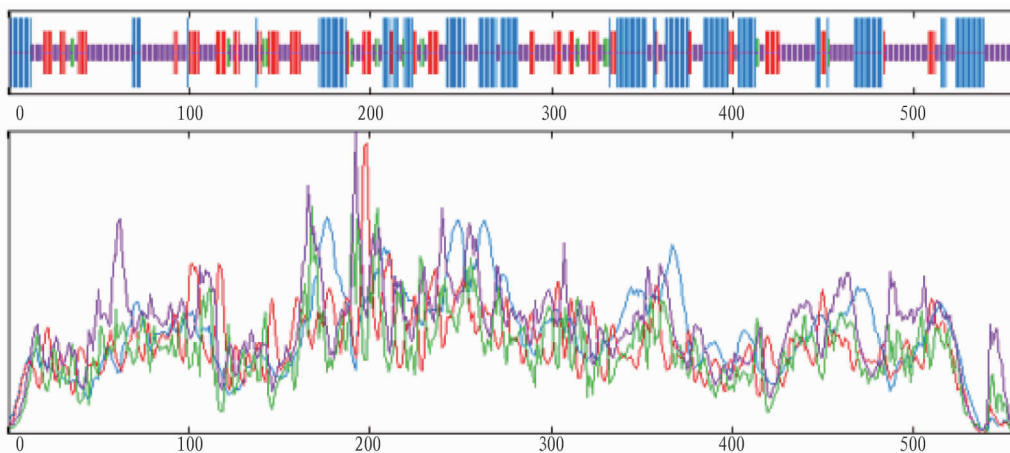


图 1 CCE001a 结构域预测

Fig. 1 CCE001a domain prediction

**2.3 CCE001a 蛋白二级结构及三级结构的预测** CCE001a 蛋白二级组成体系运用 SOPMA 网络软件工具进行分析和预测,该蛋白二级结构中, $\alpha$ -螺旋(alpha helix)占 31.35%, $\beta$ -转角(beta turn)占 3.96%,无规则卷曲(random coil)占

46.49%,扩展链(extended strand)占 18.20%(图 2)。该蛋白的三级结构(图 3)由 Phyre2 在线预测,该预测模型的蛋白覆盖率为 100%,可信度为 92%。



注:蓝色、绿色、红色、紫色分别代表  $\alpha$  螺旋、 $\beta$  转角、延伸链、无规则卷曲

Note: Blue, green, red, and purple respectively represent  $\alpha$  helix,  $\beta$  turn, extended chain and random coil

图 2 CCE001a 蛋白的二级结构

Fig. 2 The secondary structure diagram of CCE001a protein

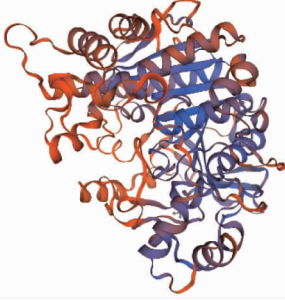


图3 CCE001a蛋白的三级结构

Fig.3 Tertiary structure of CCE001a protein

**2.4 CCE001a 蛋白亚细胞定位预测** 使用 PSOR 在线工具分析该蛋白的亚细胞定位,结果显示,CCE001a 基因编码蛋白主要定位在内质网中,内质网中含有很多参与不同类型解毒作用的酶系,从结合棉铃虫对杀虫剂及植物次生代谢产物来看,CCE001a 可能会参与解毒作用,通过氧化还原、水解或结合等反应来代谢有毒的外源物质,或将其转化成为无毒物质排出体外。

**2.5 CCE001a 蛋白氨基酸序列信号肽预测** 使用 SignalP4.0 Server 在线分析,发现该蛋白存在一个信号肽,存在位置为第 17 个氨基酸处(图 4),初步认为该蛋白为分泌型蛋白。

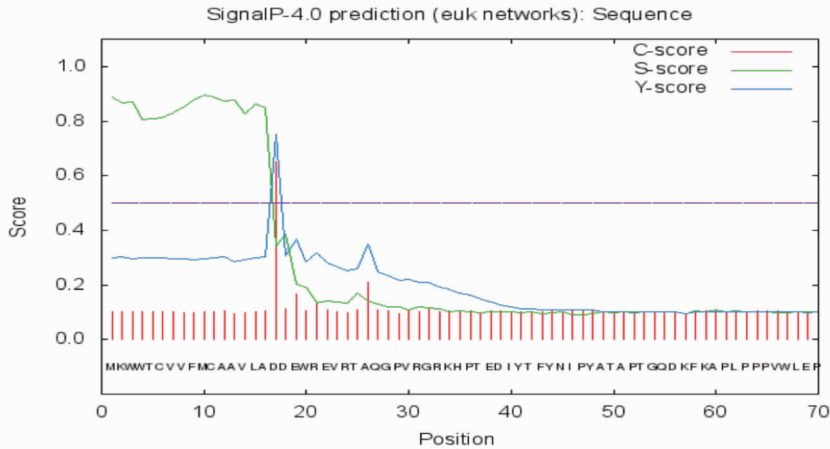


图4 信号肽预测

Fig.4 Signal peptide prediction map

**2.6 CCE001a 的蛋白质亲疏水性** 经在线分析,预测结果见图 5。CCE001a 蛋白的亲水性均值(GRAVY)为-0.262,疏水指数最大值为 3.011,位于第 11 位氨基酸上,最小为-2.922,位于第 35 位氨基酸上。CCE001a 蛋白的亲水性氨基酸范围大于疏水性氨基酸范围,故将该预测蛋白归属为亲水性蛋白。

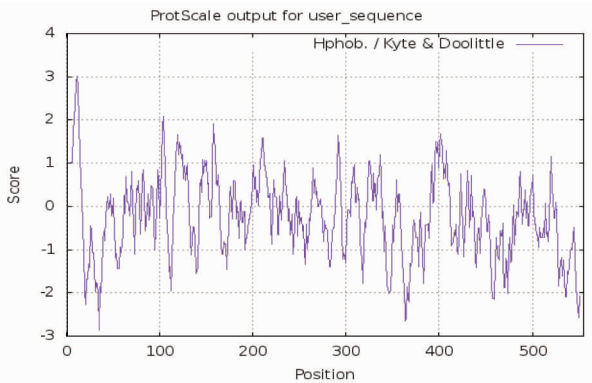


图5 蛋白的亲疏水性预测结果

Fig.5 Prediction results of protein affinity and hydrophobicity

**2.7 CCE001a 的蛋白质跨膜区的预测** 通过 TMHMM 在线软件对该蛋白质的跨膜螺旋区进行分析,结果如图 6 所示,CCE001a 蛋白的跨膜螺旋区间数为 0,且其跨膜螺旋内氨基酸的数目很小,故该蛋白质未形成一定的跨膜结构区,但由于 CCE001a 蛋白的特殊结构,该预测结果结合亲水性分析结

果,表明棉铃虫 CCE001a 基因编码蛋白为亲水性蛋白,不含有跨膜区。

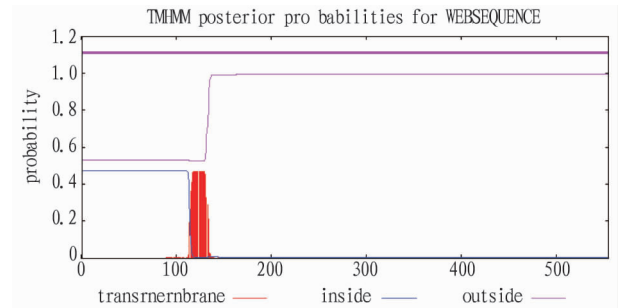


图6 CCE001a蛋白的跨膜区结果

Fig.6 Transmembrane region results of CCE001a protein

**2.8 CCE001a 蛋白磷酸化位点** 这一编码蛋白包括 15 处丝氨酸 (serine, s)、16 处苏氨酸 (threonine, t)、6 处酪氨酸 (tyrosine, y) 磷酸化定位(图 7)。蛋白 N-糖基化位点经 NetNGlyc 1.0 在线分析软件预测,结果显示评分高于 0.50 的位点仅有 1 个,位于 114 位的 N-Gly(图 8)。

### 3 讨论

棉铃虫羧酸酯酶是重要的昆虫解毒蛋白。目前,学者已证实了羧酸酯酶基因 CCE001a 与棉铃虫代谢棉酚有关。羧基酯酶多基因家族是  $\alpha/\beta$ -水解酶折叠超家族。这个超家族的成员在一级序列中有一个高度保守的非邻接残基催化三联体。三联体通常是 Ser、Glu 和 His,也可以是 Ser、Asp 和



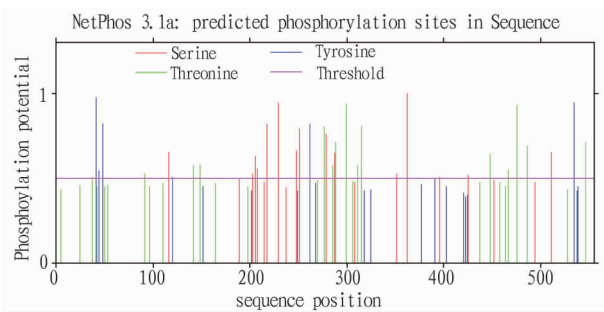


图7 CCE001a 蛋白序列的磷酸化位点分析结果

Fig. 7 Phosphorylation site analysis results of CCE001a protein sequence

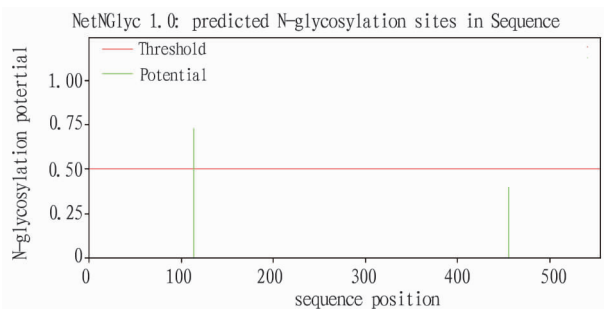


图8 CCE001a 蛋白的 N-糖基化位点

Fig. 8 N-glycosylation site of CCE001a protein

His 或 Cys、Asp 和 His。CarE 属于丝氨酸水解酶,其活性中心具有丝氨酸残基,多数生物体中有多种 CarE 同工酶。这些同工酶参与药物代谢、解毒作用和脂类物质代谢。该研究利用生物信息技术发现 CCE001a 是一个含有信号肽的亲水性蛋白,无跨膜区。分析表明 CCE001a 在蛋白中存在多个磷酸化位点,蛋白质磷酸化是一种重要和广泛应用的翻译后调节方法。蛋白磷酸化在真核生物中普遍对蛋白进行翻译修饰(磷酸化、糖基化、甲基化、乙酰乙酰化等)有重要作用,对生物细胞内的信号传递起着主导作用<sup>[13]</sup>。磷酸化前后蛋白的三维结构发生变化会影响蛋白活性的改变,进而影响信号通路中下游蛋白的活性,从而对生理生化过程进行调控<sup>[14-15]</sup>。近年来,棉铃虫具有较强的抗药性,这在国内外研究中得到了越来越多的认识,而其体内的解毒酶含量和活力程度对于耐药以及解毒效果都起着重要作用。而棉花色素腺体和棉酚能诱导棉铃虫体内的解毒酶活性增加。据文献报道,棉酚对棉铃虫解毒酶的活力诱导作用中,多功能氧化酶具有主导作用,且羧酸酯酶也能帮助多功能氧化酶在新陈代谢过程中进行水解<sup>[14]</sup>。刘杰<sup>[14]</sup>研究发现,适量的棉酚会在一定范围内刺激引起棉铃虫体内的不同种类羧酸酯酶,且这些物质还可以增强棉铃虫体内羧酸酯酶的存在,但是这种新型的诱导酶类作用于外源物质的解毒及其效果,仍有待于进一步研究才能确定。CCE001a 蛋白的三级结构结合二级结构分析显示,该蛋白主要由31.35%的 $\alpha$ -螺旋、3.96%的

$\beta$ -转角、46.49%的无规则卷曲和 18.20%的延伸链组成,而其中无规则卷曲结构有助于 CCE001a 识别底物,因为 $\alpha$ 螺旋和 $\beta$ 折叠空间结构不易发生改变而无规则卷曲空间结构易于发生改变<sup>[12]</sup>。CCE001a 存在 1 个高可信度的潜在 N-糖基化位点,真核细胞中,大部分蛋白内质网(ER)通过 N-糖基化修饰,蛋白质的折叠和分选也与糖基化息息相关<sup>[16]</sup>,但单个糖基化位点的破坏不会影响 CCE001a 的正确加工和运输功能。

该研究从 CCE001a 蛋白序列出发,运用不同在线分析软件进行分析,深入了解该蛋白的性质及功能,可为棉铃虫抗药和植物次生代谢产物提供参考。此外,仍需从多个角度来验证预测分析结果,以弥补生物信息学分析的局限性。

#### 参考文献

- [1] BAI L S,ZHAO C X,XU J J,et al. Identification and biochemical characterization of carboxylesterase O01G associated with insecticide detoxification in *Helicoverpa armigera* [J]. Pestic Biochem Physiol,2019,157:69-79.
- [2] HATFIELD M J,UMANS R A,HYATT J L,et al. Carboxylesterases: General detoxifying enzymes [J]. Chem Biol Interact,2016,259:327-331.
- [3] 鲁梅. 棉铃虫羧酸酯酶 CarE001C 原核表达及突变酶活性研究[D]. 杨凌:西北农林科技大学,2016.
- [4] OAKESHOTT J G,FARNSWORTH C A,EAST P D,et al. How many genetic options for evolving insecticide resistance in heliothine and spodopteran pests? [J]. Pest Manag Sci,2013,69(8):889-896.
- [5] JIN M H,CHENG Y,GUO X Q,et al. Down-regulation of lysosomal protein ABCB6 increases gossypol susceptibility in *Helicoverpa armigera* [J/OL]. Insect Biochem Mol Biol,2020,122[2021-02-13]. https://doi.org/10.1016/j.ibmb.2020.103387.
- [6] CRAVA C M,BRÜTTING C,BALDWIN I T. Transcriptome profiling reveals differential gene expression of detoxification enzymes in a hemimetabolous tobacco pest after feeding on jasmonate-silenced *Nicotiana attenuata* plants [J]. BMC Genom,2016,17(1):1-15.
- [7] PEARCE S L,CLARKE D F,EAST P D,et al. Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species [J]. BMC Biol,2017,15(1):1-30.
- [8] TEESE M G,FARNSWORTH C A,LI Y Q,et al. Heterologous expression and biochemical characterisation of fourteen esterases from *Helicoverpa armigera* [J]. PLoS One,2013,8(6):1-9.
- [9] CUI F,QU H,CONG J,et al. Do mosquitoes acquire organophosphate resistance by functional changes in carboxylesterases? [J]. FASEB J,2007,21(13):3584-3591.
- [10] YANG X Q,LIU J Y,LI X C,et al. Key amino acid associated with acephate detoxification by *Cydia pomonella* carboxylesterase based on molecular dynamics with alanine scanning and site-directed mutagenesis [J]. J Chem Inf Model,2014,54(5):1356-1370.
- [11] 靳明辉. 棉铃虫对棉酚的代谢机制研究[D]. 重庆:西南大学,2018.
- [12] 刘嘉伟. OATP1B1 蛋白结构与功能的生物信息学分析[J/OL]. 分子植物育种,2021-04-30[2021-05-02]. https://kns.cnki.net/kcms/detail/46.1068.S.20210430.0926.002.html.
- [13] 谢洁. 蛋白磷酸酶 PP1 介导的组蛋白甲基转移酶 EZH2 去磷酸化的机制与功能探究[D]. 长沙:湖南师范大学,2020.
- [14] 刘杰. 杀虫剂对松墨天牛几种酶和蛋白质磷酸化的影响[D]. 北京:中国农业大学,2005.
- [15] 祝水金,季道藩,刘胜安,等. 棉花色素腺体和棉酚对棉铃虫抗药性的诱导作用及其酯酶同工酶谱分析[J]. 棉花学报,2000,12(1):12-16.
- [16] 林艳梅,罗湘,李瑞杰,等. 纤维二糖水解酶 N-糖基化对其在草酸青霉中的分泌和酶活影响[J]. 中国生物工程杂志,2021,41(4):18-29.