AJAB

Asian J Agric & Biol. 2024(2).
DOI: 10.35495/ajab.2023.200

Original Article

In silico identification of Cry1Ac Hot Spot amino acid residues to improve toxicity against mutated cadherin receptor of *Bt* resistant *Helicoverpa armigera*

Sadam Munawar[1], Muhammad Qasim[2], Muhammad Sarwar Khan[1], Muhammad Amjad Ali[3], Faiz Ahmad Joyia[1*]
[1]Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad-38040, Pakistan
[2]Department of Bioinformatics and Biotechnology, Government College University, Faisalabad-38040, Pakistan
[3]Department of Plant Pathology, University of Agriculture, Faisalabad-38040, Pakistan

Received:
September 06, 2023
Accepted:
September 21, 2023
Published Online:
October 15, 2023

**Corresponding author email:*
faizahmad1980@uaf.edu.pk

Abstract

Bacillus thuringiensis Cry proteins are used for biological control of insect pests; however, insect pests are developing resistance against these proteins especially Cry1Ac. We hypothesized that the problem of resistance development can be overcome by mutating hot spot amino acid residues of Cry1Ac protein conferring enhanced toxicity against resistant insects. Cadherin protein region CD7-CD8 is known to act as primary insect midgut receptors involved in Cry1Ac binding. Amino acid mutations in this region may evolve resistance against *Bt* toxins. Here we have identified three key amino acid residues A1264, H1436 and L1461 of cadherin receptor protein from *Helicoverpa armigera* genome and *in silico* protein-protein interaction studies revealed their role in insect resistance against Cry1Ac. Three mutations viz A1264P, H1436L and L1461V showed significantly high ΔΔG values as 9.3, 6.0 and 5.9 respectively indicating destabilization of cadherin protein which reduced its binding with Cry1Ac resulting in resistance development. Further, molecular docking of these mutated amino acid residues revealed lack of interaction with amino acid residues of Cry1Ac viz Q509, Y513, W544, N547 and I585 essential for cadherin-Cry1Ac binding in susceptible insects. In second part of our study, we identified two hot spot amino acid residues of Cry1Ac viz S548, I586 whose mutation viz S548H or S548W and I586Y brought about strong interaction with midgut receptors of resistant insects having mutated cadherin. Based on these results, we sugest, reported Cry1Ac hot spot amino acids if mutated can help to overcome resistance mechanism. Here we laid a foundation for further experiments to modify the Cry1Ac hot spot residues which bind with the resistant receptor binding protein (cadherin) more strongly to perform efficient insecticidal activity against resistant strains of *Helicoverpa armigera*.

Keywords: Multiple sequence alignment, Cadherin, Aminopeptidase-N, Alkaline phosphatase

How to cite this:

Munawar S, Qasim M, Khan MS, Ali MA and Joyia FA. *In silico* identification of Cry1Ac Hot Spot amino acid residues to improve toxicity against mutated cadherin receptor of *Bt* resistant *Helicoverpa armigera*. Asian J. Agric. Biol. 2024(2): 2023200. DOI: https://doi.org/10.35495/ajab.2023.200

Introduction

The "International Cotton Advisory Committee" called its 79th plenary meeting in 2021 and documented that *Bt* cotton aided farmers more by reducing insecticides and producing higher yields without raising the cost of fertilizers and agronomic applications (ICAC, 2022). Genetically modified insect resistant *Bt* crops provide numerous benefits, hence, the area under cultivation of *Bt* crops is increasing rapidly. Farmers in 29 countries planted 190.4 Mha of GM crops in 2019, an increase of 3% or 4.5Mha from 2015 (ISAAA, 2021). According to estimates, insect pests and diseases reduce agricultural production by 30 and 50% per year (Tabashnik et al., 2023). However, insect pests have developed resistance to the previous insect resistant *Bt* varieties in major *Bt* growing counties, including Pakistan.

In the last five years, the cotton bollworm (*Helicoverpa armigera*) has ranked first among the top ten insect pests in terms of scientific publications in the "Centre for Agriculture and Biosciences International" abstract (CABI, 2017). It is still spreading and resistant to many pesticides. There have been numerous cases of genetically engineered *Bt* crops expressing *Bacillus thuringiensis Bt* proteins developing *Helicoverpa armigera* field-evolved resistance. *Helicoverpa armigera* resistance has been reported in *Bt* cotton (expressing Cry1Ac) in Australia, China, Pakistan, China, India, and the United States in 2004, 2009, 2012, 2014, 2018, and 2021, respectively (Gunning et al., 2005; Liu et al., 2010; Alvi et al., 2012; Xiao et al., 2017; Zhang et al., 2017; Kumar et al., 2020). The amino acid sequences of Cry proteins share about 40% of their amino acid sequences with each other or with the other group. Cry1 is effective against lepidopteran insects, while Cry2 is effective against both lepidopteran and dipteran insects. Three domains that made Cry proteins are Domain I (responsible for pore formation), Domain II (receptor binding and recognition), and Domain III, which is responsible for proteolytic protection, receptor binding and recognition (Yu et al., 2022). They trigger the formation of pores in insect pests by attaching to midgut receptors binding proteins for instance CAD, APN, and ALP.

The insecticidal activity of cry proteins is entirely dependent on the interactions of amino acid residues with receptor proteins, as any change in amino acid residues will result in a shift in toxicity (Kurgan et al., 2019). Minor changes in amino acid sequence can result in less or more toxicity. Previous findings revealed that Cry1Ac bound with cadherin fragment at amino acids 1217 to 1461 and involved in toxin binding and cytotoxicity (Peng et al., 2010). Susceptible field pests have developed resistance to exposed *Bt* proteins by changing essential amino-acid residues in the midgut receptor proteins. Similarly, we need to modify Cry proteins that can be expressed in crops to deal with the insect pest resistance development mechanism (Reinoso-Pozo et al., 2018). Modifying protein in plants is extremely difficult and time-consuming, therefore, bioinformatics helps to save time and labor. In order to predict protein intrinsic functional characterization in biology, computational biology using *in silico* analysis is an important research area (Munawar et al., 2021). *In silico* protein-protein interactions (PPIs) provide crucial information for cellular pathways. Before beginning extensive *in vitro* lab experiments, bioinformatics can help in save time, money, and labor (Ding and Kihara, 2018).

Hence, in this study hot spot amino acid residues of Cry1Ac proteins were identified by *in silico* protein-protein docking interactions and mutated with 20 amino acids using molecular simulation and alanine scanning to find the best amino acids which can improve the Cry1Ac toxicity to the resistant insect pest in comparison to the existing Cry1Ac. We characterized mutated Cry1Ac protein *in silico* through protein-protein interactions (PPIs/Docking) using bioinformatics tools. Mutated Cry proteins will be expected more toxic to the insect pest of cotton and will also be assumed to delay field developed resistance against pests to the earlier expressed Cry proteins in the transgenic *Bt* crops. Bioinformatics results ultimately help scientists to work with *in vitro* lab experiments and benefit scientists to work under *in vitro* lab conditions.

Material and Methods

Crystal structure and domain identification of Cry1Ac

The crystal structure of Cry1Ac protein was retrieved from the RCSB PDB (Protein Data Bank) https://www.rcsb.org/ accessed 6 Oct. 2020. A single chain (4w8j) PDB model structure was discovered, which has three domains containing 620 amino acids. The CDD (Conserved Domain Database) was used to

find conserved domains of Cry1Ac. The InterPro tool was used to determine the function of domains https://www.ebi.ac.uk/interpro/ accessed 10 Oct. 2020.

Cadherin receptor protein sequence retrieval

The full amino acid sequence of the susceptible cadherin midgut receptor (Accession# AFB74168) and twenty-one (Accession# AFB74172.1, ABF69362.1, ABI55359.1, ACZ06064.1, ACZ06065.1, ABI55356.1, AAT67416.1, AFQ60152.1, AFB74170.1, ABI55358.1, ACF94775.1, AFQ60151.1, AFB74173.1, AWJ76614.1, ACZ06062.1, AFB74169.1, ACZ06063.1, AFC17899.1, AFB74174.1, AFB74171.1, ABI55357.1) resistant cadherin midgut receptor proteins of *Helicoverpa armigera* were retrieved from "National Center for Biotechnology Information" (NCBI https://www.ncbi.nlm.nih.gov/ accessed 10 Nov. 2020.

Multiple sequence alignment and identification of mutations

MSA (Multiple sequence alignment) of one susceptimultiplmultble and twenty-one resistant cadherin receptor proteins was performed in order to identify mutated amino acids by using ClustalW https://www.genome.jp/tools-bin/clustalw/ accessed 10 Nov. 2020.

Cadherin receptor protein structure prediction, evaluation, validation, and confirmation

Homology modelling was used to predict the 3-D structure of susceptible and resistant cadherin (CAD) protein of the *Helicoverpa armigera*. Different bioinformatics tools MODLLER (v10.2) https://salilab.org/modeller/ accessed 10 Dec. 2020, MOE (2018.01), Phyre^2 http://www.sbg.bio.ic.ac.uk/phyre2/ accessed 10 Dec. 2020, SWISS-MODLE https://swissmodel.expasy.org/ accessed 10 Dec. 2020, I-TASSER https://zhanggroup.org/I-TASSER/ accessed 10 Dec. 2020, Z-server were used. To select the best protein model, all predicted protein structures were further evaluated using 'Ramachandran plot' and 'ProtParam' tool https://saves.mbi.ucla.edu/ accessed 10 Dec. 2020.

Quality assessment and superimposition of cadherin protein models

To assess the quality and overall statistics of receptor protein models bioinformatics tools "ERRAT (Quality Factor), PROCHECK (Overall G-Factor), Qmean4 (Score), and Verify 3D (3D-1D>0.20)," were used http://servicesn.mbi.ucla.edu/ accessed 16 Jan. 2021. Protein models will be fine-tune using online server 'ModRefiner' https://zhanggroup.org/ModRefiner/ accessed 16 Jan. 2021.

Docking analysis of Cry1Ac with CAD receptor binding protein

Protein-protein interactions of Cry1Ac protein with susceptible/resistant cadherin midgut binding receptor proteins of *H. armigera* were studied using the lzerd server (Kihara lab, USA) https://lzerd.kiharalab.org accessed 20 Feb. 2021. Pymol was used to visualize the docking results https://pymol.org/2/ accessed 20 Feb. 2021. We identified interacting pairs involved in receptor-toxin binding complex (Cry1Ac-CAD). Multiple sequence alignment was also used to predict key mutant cadherin receptor amino acid residues involved in resistance.

Rosetta common for ΔΔG monomer calculation

The energy change (ΔΔG) of key mutant amino acid residues of cadherin was revealed using the 'Rosetta common benchmark' https://www.rosettacommons.org/ accessed 25 Mar. 2021. Based on these findings, hot spot amino acid residues of Cry1Ac were predicted using computer-aided alanine scanning mutation to improve the Cry1Ac (with increased activity) protein to overcome insect resistance.

NMSim approach for simulation trajectories of cadherin mutations

Bioinformatics tool "NMSim" was used to perform molecular dynamics simulation http://www.nmsim.de/ accessed 10 Aug. 2021. The variables for stiff cluster decomposition were as "energy limit for hydrogen bonding (1.0 kcal/mol), the cutoff for adding hydrophobic constraints (0.35) and placing hydrophobic constraints (3). In addition, the following parameters were used during the simulation: step size (0.5), normal mode (1-50), side chain distortions (0.3), simulation cycles (500), and number of trajectories (CABI) to calculate the structural diversity of ensembles.

Hot spot scanning at protein-protein interface

The 'PIIMS' web server was used to identify hot spot amino acids at protein-protein interface of Cry1Ac and cadherin (mutated) receptor protein http://chemyang.ccnu.edu.cn/ccb/server/PIIM/ accessed 29 Sep. 2021. It combines molecular dynamics modeling and one-step free energy perturbation to explore the mutational consequences of hot spot residues. Important hot spot amino acids were found at protein-protein interfaces, and these residues are more significant for protein-protein binding than other residues.

Results

Structure of Cry1Ac

Cry1Ac protein crystal structures (4w8j, 4arx, 4ary, and 6dj4) were found using (RCSB:PDB) tool. The PDB (Protein Data Bank) entry 4w8j was used as insecticidal protein model with structural resolution 2.78Å (Fig. 1A). Future, model was visualized in Pymol to check missing residues and the correct position of the alfa-helix and beta-sheets. Selected model (4w8j) had some missing residues (shown in red rectangle in Fig. 1B), which were filled with the help of the 'MODELLER' software.

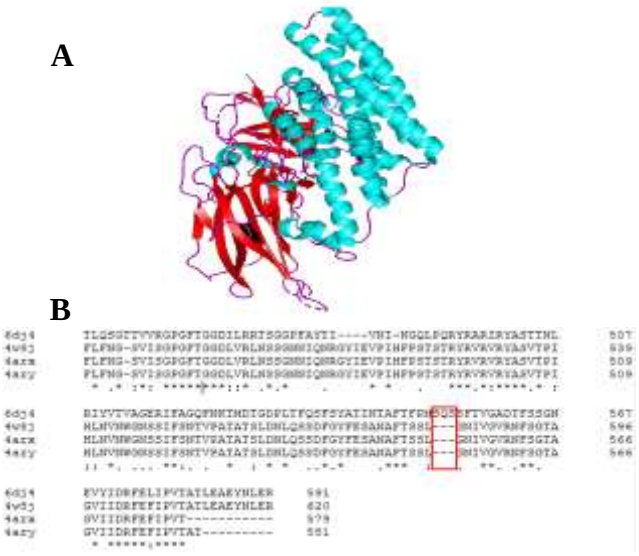

Figure-1. (A) Crystal structure of Cry1Ac *Bt* protein modle (B) Missing residues of *Bt* protein modles (6dj4,4w8j,4arx and 4ary) were indicated in red color

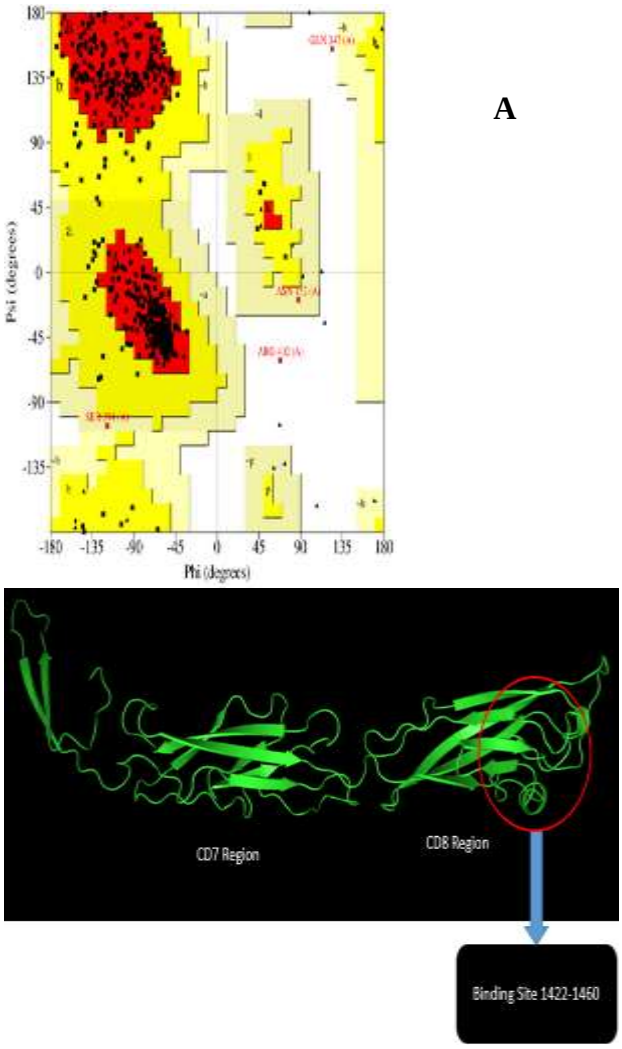

Figure-2. (A) Ramachandran plot to identify allowed and disallowed residues (B) CD7-CD8 region of cadherin protein. Binding site of cadherin protein that is important for protein-protein interaction (indicated in red circle)

Domain identification of Cry1Ac

Three functional domains (I, II, III) containing 620 amino acids were taken (from full-length 1184bp) for the physiochemical properties and protein-protein docking analysis. It had molecular weight of 68217.57(Da) and theoretical isoelectric point to be 5.79 (indicating the acidic nature of protein). The values of stability index, aliphatic index and GRAVY were found to be 38.00, 87.54, and -0.177 respectively. Stability index below 40 revealed that protein was stable, and negative GRAVY index indicated that this Cry1Ac molecule was hydrophilic. The ERRAT and Verify-3D were applied to check whether 3-D structure was compatible with its amino

acid sequence. The ERRAT score was 91.489 while Verify-3D score of 97.05 percent amino acid residues had an average 3D-1D score of 0.2 that is evidence of high-quality protein. Hence, we concluded here that the 3-D model was well-complemented with its sequence. Local similarity to target was predicted to be -0.88. A 'Ramachandran plot' was also drawn to determine percentage of allowed and disallowed amino acid residues in the model that revealed 87.3% residues to lie in the most favored region of plot (Fig. 2A).

Cadherin receptor sequence retrieval and constraints identification

One cadherin sequence from susceptible while twenty-one cadherin sequences from resistant insects were retrieved (Accession# given in materials and methods). All sequences (Susceptible & Resistant) were confirmed from literature. Multiple sequence alignment resulted was carried for 260 amino acids out of 1730 amino acids (Fig. 2B) of full-length cadherin protein as CD7-CD8 region. The amino acid sequence of 1200-1460 is most important region for interaction with Cry1Ac protein. Hence, any change or mutation in this region may bring about significant alteration(s) in protein structure resulting in weaker Cry1Ac-cadherin interaction(s) culminating in insect resistance development.

Table-1. Mutated Amino acid nature and type in the resistant insect protein sequence.

Sr. #	*AA Seq. no	Susceptible AA Abbreviations/Name		AA Type	Mutated AA Abbreviations/Name		AA Type	Repeat/21
1.	1205	R	Arginine	Positive	Q	Glutamine	Negative	1
2.	1238	V	Valine	Hydrophobic	A	Alanine	Hydrophobic	1
3.	1252	S	Serine	Polar	R	Arginine	Positive	1
4.	1260	E	Glutamate	Negative	G	Glycine	Hydrophobic	1
5.	1266	E	Glutamate	Negative	L	Leucine	Hydrophobic	1
6.	1268	R	Arginine	Positive	E	Glutamate	Negative	1
7.	1270	E	Glutamate	Negative	V	Valine	Hydrophobic	1
8.	1272	P	Proline	Hydrophobic	L	Leucine	Hydrophobic	1
9.	1289	T	Threonine	Polar	A	Alanine	Hydrophobic	1
10.	1298	S	Serine	Polar	P	Proline	Hydrophobic	4
11.	1315	K	Lysine	Positive	R	Arginine	Positive	9
12.	1316	P	Proline	Hydrophobic	L	Leucine	Hydrophobic	1
13.	1320	S	Serine	Polar	A	Alanine	Hydrophobic	1
14.	1323	E	Glutamate	Negative	D	Aspartate	Negative	3
15.	1328	I	Isoleucine	Hydrophobic	L	Leucine	Hydrophobic	1
16.	1338	A	Alanine	Hydrophobic	V	Valine	Hydrophobic	1
17.	1364	A	Alanine	Hydrophobic	P	Proline	Hydrophobic	1
18.	1386	T	Threonine	Polar	S	Serine	Polar	1
19.	1390	G	Glycine	Hydrophobic	D	Aspartate	Negative	2
20.	1398	I	Isoleucine	Hydrophobic	V	Valine	Hydrophobic	1
21.	1401	S	Serine	Polar	T	Threonine	Polar	1
22.	1406	P	Proline	Hydrophobic	H	Histidine	Positive	1
23.	1407	T	Threonine	Polar	P	Proline	Hydrophobic	1
24.	1436	H	Histidine	Positive	L	Leucine	Hydrophobic	1
25.	1449	R	Arginine	Positive	S	Serine	Polar	1
26.	1450	R	Arginine	Positive	T	Alanine	Hydrophobic	1
27.	1452	T	Threonine	Polar	A	Alanine	Hydrophobic	1
28.	1461	L	Leucine	Hydrophobic	V	Valine	Hydrophobic	1

Structure prediction of susceptible cadherin receptor protein

Structure of susceptible cadherin (CD7-CD8) receptor protein was not available; therefore, 3-D structure was predicted using computer-aided prediction modeling tools. The structure predicted through Z-server (Kihara lab, USA) was found the best among all. The active part of cadherin involved in an interaction was correctly predicted, where beta-sheets and alpha-helix were found in correct position as evident in Fig. 2B). The template identities of models predicted through Swiss-Model and Phyre[2] were only 26% and 20%, respectively, while I-TASSER model lacked correct amino acid presentation.

Multiple sequence alignment of one susceptible *AA (amino acid) sequence of cadherin receptor protein with the twenty-one resistant protein sequences resulted 28 random mutations. Amino acids sequence number, abbreviations, type, nature and repetition are given in detail.

Hence, we used only Z-server model in further study. 'Ramachandran plot' was used to evaluate the predicted models which also confirmed the highest favorability of Z-server model (Table 2).

Table-2. Cadherin protein structure evaluation using Ramachandran plot.

Sr. #	Proteins	Favoured Region	Allowed Region	Generously Allowed Region	Outline Region
1.	Moddeller	202 (87.8%)	21 (9.1%)	6 (2.6%)	1 (0.4%)
2.	Swiss	172 (70.2%)	56 (22.9%)	115 (4.8%)	6 (2.4%)
3.	Phyre[2]	194 (83.6%)	28 (12.1%)	6 (2.6%)	4 (0.7%)
4.	I-tasser	187 (82.7%)	34 (15.0%)	4 (1.8%)	1 (0.4%)
5.	Z-Server	209 (90.93%)	14 (6.1%)	6 (2.6%)	1 (0.4%)

Further, 'ProtParam tool' was used to check the physiochemical properties (Amino acid, Molecular weight, Iso-electric point, Half life, Extinction coefficient, Nega./Pos. amino acids, GRAVY, instability index, aliphatic index) of Z-server model before docking (details are given in Table 3).

Docking analysis of Cry1Ac protein with susceptible cadherin receptor protein

For docking ('lzerd sever') parameters were fixed as; fraction 80%, min distance 3 and max distance 14. The values of 10 best docking complexes are shown in Fig. 3. We selected all docking complexes with the best rank sum score (DFIRE+GOAP+ITScore) and found common residues repeated in all docking poses (supplementary data Fig. 2). For the visualization of docking results 'Pymol' visualization software was used. We identified 14 interacting pairs involved in receptor-toxin binding complexes of Cry1Ac with the cadherin (Table 4).

Table-3. Physiochemical properties of cadherin receptor protein (Z-server model) structure.

Parameters	Score
Amino acid	263
Molecular weight	29020.44Da
Iso-electric point	4.46
Half life	5.5
Extinction coefficient	14565 $M^{-1}cm^{-1}$
Nega./Pos. amino acids	40/18
GRAVY	-0.198
instability index	36.53
aliphatic index	83.00

Table-4. Interacting pairs of susceptible cadherin receptor protein with Cry1Ac protein.

Sr no.	Cadherin AA residue Interact with Cry AA residue in the form of pair	Sr no.	Cadherin AA residue Interact with Cry AA residue in the form of pair
1	A1364-W544	8	H1436-N547
2	T1367-W545	9	M1438-N547
3	A1368-N547	10	V1459-R590
4	G1369-G546	11	Y1460-W544
5	I1370-I585	12	Y1460-I550
6	L1380-R590	13	L1461-S549
7	L1381-Y513	14	S1463-Q509

Protein-protein interaction of Cry1Ac (susceptible) protein with the susceptible cadherin protein of *H. armigera* resulted 14 *AA (amino acid) pairs from both protein interfaces.

*A1364-W544, H1436-N547, L1461-S549 key amino acid pairs from both proteins interfaces.

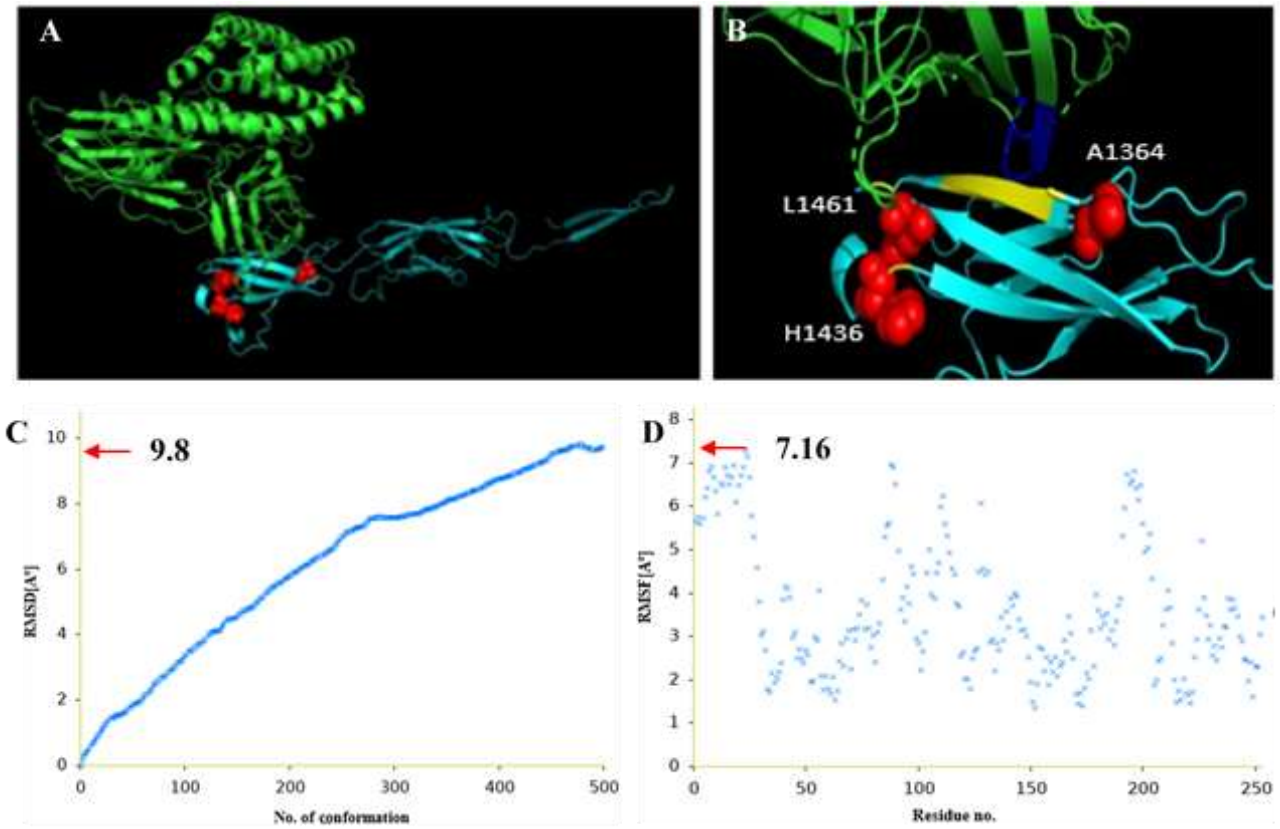

Figure-3. (A) Pairwise Docking prediction by 'lzerd server' for a protein complex of ligand protein (Cry1Ac: Green) and receptor protein (Cadherin: Cyan). Interacting pairs are indicated in yellow color for cadherin protein and blue color for Cry1Ac protein (B) Red color indicate the mutated amino acid of the resistant insects which also identified in multiple sequence alignment results (C) Graph displaying the RMSD of the Cα atom to the initial structure along the trajectory (D) A graph displaying the RMSF of the Cα atom over the trajectory.

Here we noticed three key amino acids A, H and L at 1364, 1436 and 1461 positions of susceptible cadherin were strongly interacting with Cry1Ac. The amino acids at same positions of resistant cadherin proteins were found mutated as depicted in (Fig. 3A). Hence, these could be involved in resistance development in insects against Cry1Ac (Fig. 3) and if we mutate their interacting amino acids of Cry1Ac (highlighted yellow in Table 4), it can improve binding of Cry1Ac to cadherin receptor of resistant insects and may help in breaking Cry1Ac resistance in insects.

Rostetta common benchmark to find unstable mutations in cadherin receptor

We used "Rostetta" software to validate the effect of mutations in previously identified amino acids of cadherin on its structural stability and found that point mutation of wild type amino acids Alanine 1364 to Proline, Histidine 1436 to Leucine, and Leucine 1461 to Valine cause change in structure and folding-free energy for cadherin receptor binding protein.

Table-5. Multiple sequence alignment of mutated amino acid of the cadherin protein.

# MSA of resistant insect	Wild Type Amino Acid to Mutated	Nature of wild A.A to Mutated	ΔΔG_ Monomer	Results
Mutation 1	Alanine 1364 Proline	*Hydro. To Hydro.	9.35346	Highly Unstable
Mutation 2	Histidine 1436 Leucine	Positive To Hydro.	6.0099	Highly Unstable
Mutation 3	Leucine 1461 Valine	Hydro. To Hydro.	5.9003	Highly Unstable

***Hydrophobic**

To predict the change in stability of a monomeric protein, ΔΔG of point mutation was calculated and shown in Table 5. Results indicated that these mutations are destabilizing cadherin structure and masking it inaccessible for Cry1Ac.

Stabilizing mutations = ΔΔG ≤ -1.0kcal/mol
Neutral = ΔΔG between -1.0 and 1.0kcal/mol
Destabilizing ΔΔG ≥ 1.0kcal/mol

Comparison of susceptible and mutated cadherin protein structure

The susceptible and mutated protein models were aligned using Pymol software to identify structural differences in the loop region. Results indicated clear structural changes closer to the β-sheet and loop regions as depicted by red circled in supplementary Fig. 3 and values are given in (Table 5).

NMSim for dynamic simulations and evaluation

Our findings revealed that β-sheets were primarily responsible for the assembly of the CR7–CR8 region, therefore, it should be in correct position for docking analysis. The 3-D structure of mutated cadherin protein model was analyzed by model assessment method, following MD-simulation and energy minimization. The average RMSD of the structures after 200 simulation iterations was discovered to be 9.8. (Fig. 3C). The total flexibility of the residues was examined using RMSF (root mean-square fluctuations). The mean RMSF of the CR7–CR8 region of mutated cadherin protein was 7.16 (Fig. 3D), which was in range to other structures. After molecular dynamics, the model's parameters were calculated. The results demonstrated that model is reliable and can be used in future docking studies.

Mutated cadherin receptor protein docking with Cry1Ac (susceptible) protein

Docking results were 75% different from those of the susceptible cadherin protein. In the docking analysis, the mutated amino acids (P1264, L1436, V1461) of cadherin were not involved in the interacting pairs. Supplementary Fig. 4 shows the change in structure and shift in interacting pairs (Table 6) because of mutation in cadherin protein to make it resistant.

Table-6. Interacting pairs of mutant cadherin (resistant) receptor with Cry1Ac protein.

Sr no.	Mutated cadherin AA residue interact with susceptible Cry1Ac AA residue in the form of pair	Sr no.	Mutated cadherin AA residue interact with susceptible Cry1Ac AA residue in the form of pair
1	Y1366-T525	**9**	L1380-S581
2	T1367-V529	**10**	L1381-S581
3	A1368-S548	**11**	V1382-L583
4	G1369-W545	**12**	T1382-S548
5	I1370-S548	**13**	V1383-I 585
6	S1371-S549	**14**	V1424-V586
7	R1378-S548	**15**	G1437-V586
8	N1379-I550	**16**	I1462-R590

Protein-protein interaction of Cry1Ac (susceptible) protein with the mutated cadherin protein of *H. armigera* resulted 16 *AA (amino acid) pairs from both proteins interfaces.

Pymol-based visualization of hot spot amino acid in Cry1Ac at protein-protein Interface: Based on the destabilizing cadherin receptor mutations, 5 hot spot residues within the 8Å distance at the protein-protein interface were selected by 'Pymol' as depicted by Blue color in Fig. 4. These hot spot amino acids can be used to overcome the resistance mechanism of cadherin receptors. If cadherin protein mutations remain constant, then we have an option to mutate Cry1Ac hot spot amino acids to cope the insect resistance. These hot spot amino acids were further verified and analyzed by an online server "PIIMS; Server for Scanning hot spot Mutation at the Protein–Protein Interface" as given in next section.

#	Name and number of Amino Acid	Nature of Amino Acid
1	Phenylalanine 481	Aromatic
2	Glutamine 509	Negative
3	Serine 548	Polar
4	Valine 586	Non-Polar
5	Arginine 590	Positive

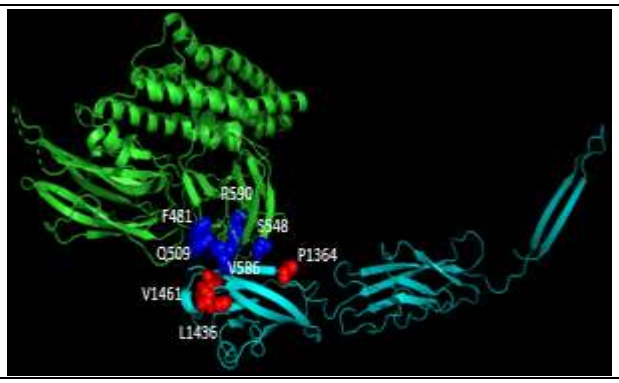

Figure-4. Cry1Ac hot spot residues within the 8Å distance using Pymol. Cry1Ac hot spot residues (shown in BLUE) within the 8Å distance with the 3 mutated amino acid of cadherin receptor protein (shown in RED) using Pymol

Alanine scanning of hot spot amino acid using PIIMS

Web server 'PIIMS' employs computational alanine scanning analysis to locate hot spots and comprehensive mutation scanning analysis to assess the consequences of hot spot mutations. Furthermore, hot spot scanning at the PPi (protein-protein interface) with the other 20 amino acids was also used to identify the most potent amino acids that could be mutated to overcome the resistant in insects and improve the existing toxicity of the Cry1Ac against the mutated cadherin receptor.

Heatmap analysis for binding affinity of the PPI after mutation

A "Heatmap" illustrating the changes in binding free energy of mutated hot spot residues was generated for an intuitive understanding of the predictions. The rectangle on the map is colored according to the PPI binding affinity: the greater the binding affinity, the redder the rectangle; the lower the binding affinity, the bluer the rectangle (shown in Fig. 5B). The results revealed that if Cry1Ac Ser548 be replaced with His or Trp and Val 586 be replaced with Tyr, the resultant Cry1Ac will have higher binding affinities with cadherin protein that has been mutated in *Helicoverpa armigera* overtime. This is expected to break the Cry1Ac resistance prevailing in field.

The primary step in Cry toxicity is binding of Cry protein to cadherin receptor in the mid gut of target pests (Gómez et al., 2014; Stalinski et al., 2016). The transmembrane domain of the cadherin receptor of different target pests often contains one or more cadherin repeats (CRs), upon which Cry protein bound and causes pest death. According to many studies, any mutations in these transmembrane domains may lead to Cry resistance in insects (Xu et al., 2016; Tabashnik and Carrière, 2017).

The present study findings demonstrated that the cadherin domain of *Helicoverpa armigera* has a distinctive Cry1Ac-toxin binding motif CD7-CD8 (residues of 1364–1463) that is crucial for both toxin binding and toxicity. (Fig. 2B, 5). Similarly, Xie et al. (2005) proposed Cry1Ab-binding sites in the *Manduca sexta* cadherin region, AA 1363–1464, which are involved in toxin binding and cytotoxicity, and a 219-amino-acid residue (1245–1464) of the cadherin in *Bombyx mori* is responsible for Cry1Aa binding (Nagamatsu et al., 1999). Peng et al. (2010) also found that Cry1Ac might be bound by a cadherin fragment of the amino acids 1217 to 1461.

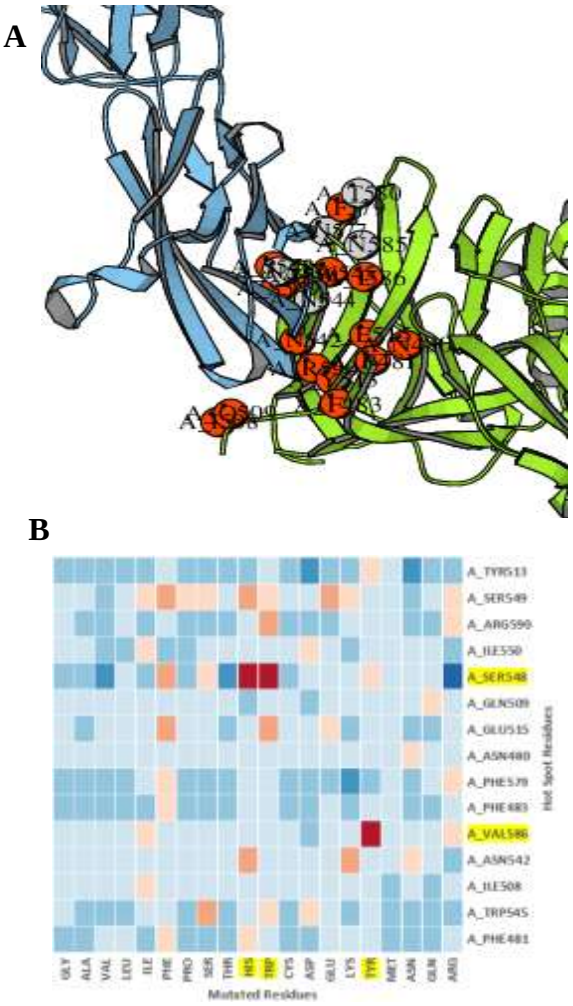

Figure-5. (A) Full mutation scanning analysis for hot spot residues (B) Heatmap indicating that, the greater the binding affinity, the redder the rectangle of the PPI for mutation

Discussion

Furthermore, our findings revealed that Cry1Ac failed to bind to mutated cadherin region CR7–CR8 (A1264P, H1436L, L1461V) of *Helicoverpa armigera*. It can be predicted that the activated toxin Cry1Ac will become less toxic to insects with such mutations in cadherin (supplementary Fig. 3). We have suggested that toxin-binding epitopes in the CR7– CR8 region of cadherin are likely to bend and change shape. On the other hand, distinct CR7-CR8 folding scenarios could lead to various binding locations for Cry1Ac; this has to be further verified.

The interaction/binding of two molecules can be studied effectively using computer-assisted molecular docking. On the other hand, the "Protein Data Bank"

(PDB) lacks a crystal structure of cadherin protein for *Helicoverpa armigera*. The amino acid sequence of cadherin region (CR7-CR8) was uploaded to Z-server (Kihara lab, USA) in order to build a 3D structure of CAD receptor. It was found that the 3D structure of CR7–CR8 was mostly made up of β-sheets (Fig. 2B), which is consistent with the cadherin crystal structures published in the PDB database for other insects. Following MD (molecular dynamics) modelling, the final 3-dimensional structures of CR7-CR8 showed improved assessment parameters (Fig. 3, 6), and they were employed for additional docking investigations. The Cry1Ac toxin domains II/III are important in binding to the CR7-CR8 region of CAD receptor. According to the molecular docking data, "Loop2 and Loop3" in domain III contribute more for interaction with receptor (Table 1, 2). These claims were supported by several studies (Gómez et al., 2006; Bravo et al., 2011).

We used multiple sequence alignment to compare the sequences of all known resistant (twenty-one) *Helicoverpa armigera* insects from 'NCBI' to identify the mutated amino acids that may be involved in resistance (Supplementary Fig. 1). From these 42 mutated amino acids of 'Multiple sequence alignment', we identified three key mutated amino acids A1264P, H1436L, L1461V which cause structural changes in β-sheets of the cadherin receptor protein (supplementary Fig. 3). Protein stability was altered as a result of these mutations as the interacting amino acids of the susceptible cadherin protein got altered, resulting in a reduction of toxicity with Cry1Ac (Supplementary Fig. 4). The change of "Positive/Hydrophobic amino acid to Hydrophobic" is more important to change the structure of the cadherin receptor. Additionally, it was thought that hydrophobic interactions and the hydrogen bonding in toxin-receptor interaction were crucial for preserving the stability of the complex (Table 5).

From this perspective, mutations A1264P and L1461V were found conservative, preserving the hydrophobic nature of these positions, which are correlated with the conserved toxicity level. Contrarily, the mutation H1436L introduced a hydrophobic residue into the suitable environment, perhaps altering the loop structure, which accounts for the significant decline in relative toxicity. According to recent research by Xiao et al. (2017), a point mutation in cadherin receptor of *Helicoverpa* provides resistance to Cry1Ac by decreasing toxin binding. In *Heliothis virescens*, *Pectinophora gossypiella*, and *Helicoverpa armigera*, there are several mutations linked with Cry1Ac resistance, which shows that not all mutations impart in toxicity but instead alter other downstream toxicity processes (Zhao et al., 2010; Fabrick and Tabashnik, 2012; Xu et al., 2016).

Xie et al. (2005) discovered a crucial toxin-binding site of cadherin receptor in *Heliothis virescens* cadherin, demonstrating that point mutations in cadherin at position (Leu 1425 and Phe 1429) are required for toxicity of Cry1Ac and interaction. Toxin binding was lost when charged amino acids were substituted, resulting in a significant reduction in toxicity. They also discovered that Cry1Ab and Cry1Ac toxins (loop 3) bind to this site with the G439 and F440 and loop 3 of Cry1Ac causing the most toxicity loss. Hydropathic complementarity predicted that these amino acids would interact with 1422-1440 region in the toxin-binding receptor. They mutated Leu1425, Asn1428, Phe1429, and Gln1430, and concluded that change at position 1425 and 1429 caused loss in toxicity.

The alanine substitution mutants S581A, I585A, F536A, and N546A of the Cry1Ac5 showed increased toxicity towards *Helicoverpa armigera* larvae, according to many reports (Lv et al., 2011; Reinoso-Pozo et al., 2018). Liu et al. (2010) did site-directed mutagenesis to investigate the role of W544F residue in Cry1Ac stability. Reinoso-Pozo et al. (2018), found that two introduced mutations N547Y and R602G in the Cry1Ac gene played a significant role in toxicity. The N547Y mutant contributed more towards toxicity, while the R602G mutant contributed less. When compared to our results, the amino acids at position W544, N547 are vital, and they play significant role in toxicity activity of Cry1Ac. It can be deduced that strong binding of Cry1Ac with cadherin from susceptible insects is due to the interaction of W544, N547 with amino acids A1364, H1436 of susceptible cadherin. However, in mutant cadherin amino acids at both positions were altered hence, not interacting as shown in our results (Fig. 3A, 7).

Xia et al. (2008) investigated the effects of various mutations in Cry1Ac on its insecticidal activity against *Helicoverpa armigera*. These mutations N543A, W544A, G545A, S547A, S548A, and I549A reduced its toxicity while N546A on the other hand, doubled the toxicity of Cry1Ac. Similarly, heat map analysis in our study predicted two mutations at S548 (S548H and S548W) to increase toxicity of Cry1Ac

protein against resistant *Helicoverpa armigera* (Fig. 5B). Our *in-silico* homology modeling results for Cry1Ac revealed that residue S548 in the protein's surface is oriented towards the solvent (Fig. 5A), implying that it may play a role in interaction. The side chains of residues W544, G545, S547, and I549, on the other hand, may play a structural role in local stability. We predict that increased toxicity of Cry1Ac could be caused by exposing other binding sites like S548, V586 and an easy conformational change due to increased loop flexibility via "Histidine/Tryptophan and Tyrosine" mutations, respectively (Table 6). Xia et al. (2008) confirmed that the side chains of S547 and S548 were crucial for Cry1Ac toxicity. Furthermore, mutation of W544F was significantly more toxic as compared to W544Y and representing that the hydrophobic nature of this amino acid is critical for Cry1Ac protein stability and activity demonstrating that the hydrophobic property of this amino acid was essential for the stability and proper function of the Cry1Ac protein.

Lv et al. (2011) found convincing evidence that the mutants of Cry1Ac5 (S580A, L582A, G583A, N584A, and V586A) had less toxicity and S581A and I585A had higher toxicity against cotton bollworm. The V586 was buried deep within the molecule, and its side chains made close contact with the mutated cadherin, possibly contributing to local stability structurally (Espinosa et al., 2001; Colombo et al., 2003). As a result of our findings, the change in toxicity of the V586Y mutant could be due to effects in interacting with mutated cadherin or preserving the protein stability. According to study by Xia et al. (2008), the aromatic rings of F578, F604, and W544 were organized in a propeller-like manner, which preserved protein structural stability through hydrophobic contact. Any disruption to the hydrophobic interaction's balance could result in a loss of protein stability. As a result, we predict that the increased insecticidal activity of mutation S548, V586 was due to the removal of some unstable factors, resulting in an improved hydrophobic force balance among interacting residues. However, more research is needed to see if the mutants S548H/W and V586Y have improved Cry1Ac structural stability (Fig. 5B). According to our findings, the insecticidal activity of the Cry1Ac protein is significantly influenced by the residues in this loop region.

The hot spot amino acids of the toxin-receptor complex were predicted in this work, and their significance in the interaction was confirmed. In general, amino acid residues at the protein-protein complex interface contribute various binding energies (Schreiber, 2020). The majority of the binding energy is produced by important residues S548H/W and V586Y (Ovek et al., 2022). Due to their size and shape, these are more suitable as hot spot in protein-protein interaction, as evidenced by the presence in the complexes. We hypothesized that the increased toxicity of hot spot amino acids facilitates conformational changes in the protein's binding region due to an increase in the loop's flexibility, allowing Cry1Ac to bind with the mutated receptor cadherin more easily or firmly. To date, no *in silico* study has been published on the impact of these amino acid residues of *B. thuringiensis* toxin in toxicity of *Helicoverpa armigera*.

Conclusion

We confirmed the Cry1Ac binding with midgut cadherin receptor region CD7-CD8 of *Helicoverpa armigera* and identified the important interacting amino acids A1264, H1436 and L1461. Hence, mutations of these interacting amino acids A1264P, H1436L and L1461V cause resistant development in *Helicoverpa armigera*. Furthermore, hot spot amino acid residues of Cry1Ac S548 and I586 in the interaction of the toxin-receptor complex were predicted. These results will lay the groundwork for additional investigation of receptor interaction and enhancement of Cry protein efficacy against common *Bt*-resistant strains of cotton bollworm.

Disclaimer: None.
Conflict of Interest: None.
Source of Funding: None.

Contribution of Authors

Munawar S: Performed the experiment, collected & analyzed data and wrote the initial draft
Qasim M: Helped in docking analysis & literature review
Khan MS & Ali MA: Did literature review and edited the manuscript
Joyia FA: Designed the study, supervised the experiments, interpreted data and approved final draft of manuscript

References

Alvi AH, Sayyed AH, Naeem M and Ali M, 2012. Field evolved resistance in *Helicoverpa armigera* (Lepidoptera: Noctuidae) to *Bacillus thuringiensis* toxin Cry1Ac in Pakistan. PlosOne. 2012. 7(10): e47309.

Bravo A, Likitvivatanavong S, Gill SS and Soberón M, 2011. *Bacillus thuringiensis*: a story of a successful bioinsecticide. Insect Biochem. Mol. Biol. 41(7):423-31.

CABI, 2017. State of the worlds plants/plant health-state of research. 64–71.

Colombo G, De Mori GM and Roccatano D, 2003. Interplay between hydrophobic cluster and loop propensity in β-hairpin formation: A mechanistic study. Prot. Sci. 12:538-550.

Ding Z and Kihara D, 2018. Computational methods for predicting protein-protein interactions using various protein features. Curr.Prot. Prot. Sci. 93:e62.

Espinosa JF, Muñoz V and Gellman SH, 2001. Interplay between hydrophobic cluster and loop propensity in β-hairpin formation. J. Mol. Biol. 306:397-402.

Fabrick JA and Tabashnik B, 2012. Similar genetic basis of resistance to *Bt* toxin Cry1Ac in boll-selected and diet-selected strains of pink bollworm. PlosOne 7:e35658.

Gómez I, Arenas I, Benitez I, Miranda-Ríos J, Becerril B, Grande R, Almagro JC, Bravo A and Soberón M, 2006. Specific epitopes of domains II and III of *Bacillus thuringiensis* Cry1Ab toxin involved in the sequential interaction with cadherin and aminopeptidase-N receptors in *Manduca sexta*. J. Biol. Chem. 281:34032-34039.

Gómez I, Sánchez J, Muñoz-Garay C, Matus V, Gill SS, Soberón M and Bravo A, 2014. *Bacillus thuringiensis* Cry1A toxins are versatile proteins with multiple modes of action: two distinct pre-pores are involved in toxicity. Biochem. J. 459:383-396.

Gunning RV, Dang HT, Kemp FC, Nicholson IC and Moores GD, 2005. New resistance mechanism in *Helicoverpa armigera* threatens transgenic crops expressing *Bacillus thuringiensis* Cry1Ac toxin. Appl. Envir. Microbiol. 71:2558-2563.

ICAC, 2022. Statement of the 80th Plenary Meeting for Cotton: Review of the World Situation, International Cotton Advisory Committee (ICAC). 76:1-18.

ISAAA, 2021. Breaking Barriers with Breeding: A Primer on New Breeding Innovations for Food security, ISAAA Breif 56-2021

Kumar K, Gambhir G, Dass A, Tripathi AK, Singh A, Jha AK, Yadava P, Choudhary M and Rakshit S, 2020. Genetically modified crops: current status and future prospects. Planta 251:1-27.

Kurgan L, Radivojac P, Sussman JL and Dunker AK, 2019. On the importance of computational biology and bioinformatics to the origins and rapid progression of the intrinsically disordered proteins field. In: Pacific Symposium on Biocomputing 2020. p 149-158.

Liu F, Xu Z, Zhu YC, Huang F, Wang Y, Li H, Li H, Gao C, Zhou W and Shen J, 2010. Evidence of field-evolved resistance to Cry1Ac-expressing *si* cotton in *Helicoverpa armigera* (Lepidoptera: Noctuidae) in northern China. Pest Manag. Sci. 66:155-161.

Lv Y, Tang Y, Zhang Y, Xia L, Wang F, Ding X, Yi S, Li W and Yin J, 2011. The Role of β20–β21 Loop Structure in Insecticidal Activity of Cry1Ac Toxin from *Bacillus thuringiensis*. Curr. Microbiol. 62:665-670.

Munawar S, Sagir M, Mustafa G, Ali MA, Niazi AK, Parvaiz A, Yasmin F, Mansoor F, Kanwal S and Rasheed M, 2021. *In silico* analyses of predicted substitutions in fibrinolytic protein 'Lumbrokinase-6'suggest enhanced activity. Proc. Biochem. 110:292-301.

Nagamatsu Y, Koike T, Sasaki K, Yoshimoto A and Furukawa Y, 1999. The cadherin-like protein is essential to specificity determination and cytotoxic action of the *Bacillus thuringiensis* insecticidal CryIAa toxin. FEBS Lett. 460:385-390.

Ovek D, Abali Z, Zeylan ME, Keskin O, Gursoy A and Tuncbag N, 2022. Artificial intelligence based methods for hot spot prediction. Curr. Opin. Struc. Biol. 72:209-218.

Peng D, Xu X, Ye W, Yu Z and Sun M, 2010. *Helicoverpa armigera* cadherin fragment enhances Cry1Ac insecticidal activity by facilitating toxin-oligomer formation. Appl. Microbiol. Biotechnol. 85: 1033–1040.

Reinoso-Pozo Y, Del Rincón-Castro MC and Ibarra J, 2018. Significant increase in toxicity of the Cry1Ac1 protein of *Bacillus thuringiensis* by the exchange of two amino acids in its domain III. Biol. Cont. 125:1-6.

Schreiber G, 2020. Protein–protein interaction interfaces and their functional implications. (pp.1-24) In Siddhartha R. and Haian F (Eds) Protein–Protein Interaction Regulators. Royal Society of Chemsitry, Cambridge, UK. DOI:10.1039/9781788016544-00001

Stalinski R, Laporte F, Tetreau G and Després L, 2016. Receptors are affected by selection with each *Bacillus thuringiensis* israelensis Cry toxin but not with the full *Bt*i mixture in Aedes aegypti. Genet. Evol.. 44:218-227.

Tabashnik BE and Carrière Y, 2017. Surge in insect resistance to transgenic crops and prospects for sustainability. Nat. Biotechnol. 35:926-935.

Tabashnik BE, Fabrick JA and Carrière Y, 2023. Global patterns of insect resistance to transgenic *Bt* crops: the first 25 years. J. Econ. Entomol. 116(2):297-309.

Xia L, Wang F, Ding X, Zhao X, Fu Z, Quan M and Yu Z, 2008. The role of β18-β19 loop structure in insecticidal activity of Cry1Ac toxin from *Bacillus thuringiensis*. Chin. Sci. Bull. 53: 3178-3184

Xiao Y, Dai Q, Hu R, Pacheco S, Yang Y, Liang G, Soberón M, Bravo A, Liu K and Wu K, 2017. A single point mutation resulting in cadherin mislocalization underpins resistance against *Bacillus thuringiensis* toxin in cotton bollworm. J. Biol. Chem. 292:2933-2943.

Xie R, Zhuang M, Ross LS, Gomez I, Oltean DI, Bravo A, Soberon M and Gill SS, 2005. Single Amino Acid Mutations in the Cadherin Receptor from Heliothis virescens Affect Its Toxin Binding Ability to Cry1A Toxins. J. Biol. Chem. 280(9): 8416-8425. https://doi.org/10.1074/jbc.M408403200.

Xu P, Islam M, Xiao Y, He F, Li Y, Peng J, Hong H, Liu C and Liu KC, 2016. Expression of recombinant and mosaic Cry1Ac receptors from *Helicoverpa armigera* and their influences on the cytotoxicity of activated Cry1Ac to Spodoptera litura Sl-HP cells. Cytotechnology 68:481-496.

Yu S, Wang C, Li K, Yang Y, He YZ and Wu Y, 2022. Transcriptional Analysis of Cotton Bollworm Strains with Different Genetic Mechanisms of Resistance and Their Response to *Bacillus thuringiensis* Cry1Ac Toxin. Toxins 14:366.

Zhang H, Yu S, Shi Y, Yang Y, Fabrick JA and Wu Y, 2017. Intra-and extracellular domains of the *Helicoverpa armigera* cadherin mediate Cry1Ac cytotoxicity. Insect Biochem. Mol. Biol. 86:41-49.

Zhao J, Jin L, Yang Y and Wu Y, 2010. Diverse cadherin mutations conferring resistance to *Bacillus thuringiensis* toxin Cry1Ac in *Helicoverpa armigera*. Insect Biochem. Mol. Biol. 40:113-118.