Original article

In silico annotation of unreviewed acetylcholinesterase (AChE) in some lepidopteran insect pest species reveals the causes of insecticide resistance

Qudsia Yousafi [a,*], Ayesha Sarfaraz [a], Muhammad Saad Khan [a], Shahzad Saleem [a], Umbreen Shahzad [b], Azhar Abbas Khan [b], Mazhar Sadiq [a], Allah Ditta Abid [c], Muhammad Sohail Shahzad [d], Najam ul Hassan [e]

[a] *COMSATS University Islamabad, Sahiwal Campus, Sahiwal, Punjab, Pakistan*
[b] *College of Agriculture, Bahauddin Zakariya University, Bahadur Campus, Layyah, Pakistan*
[c] *Director Technical MNSFR, Islamabad, Pakistan*
[d] *Director Technical Plant Protection Pakistan*
[e] *Deputy Director Pest Warning, Bahawalpur, Pakistan*

ARTICLE INFO

ABSTRACT

Lepidoptera is the second most diverse insect order outnumbered only by the Coeleptera. Acetylcholinesterase (AChE) is the major target site for insecticides. Extensive use of insecticides, to inhibit the function of this enzyme, have resulted in the development of insecticide resistance. Complete knowledge of the target proteins is very important to know the cause of resistance. Computational annotation of insect acetylcholinesterase can be helpful for the characterization of this important protein. Acetylcholinesterase of fourteen lepidopteran insect pest species was annotated by using different bioinformatics tools. AChE in all the species was hydrophilic and thermostable. All the species showed lower values for instability index except *L. orbonalis, S. exigua* and *T. absoluta*. Highest percentage of Arg, Asp, Asn, Gln and Cys were recorded in *P. rapae*. High percentage of Cys and Gln might be reason for insecticide resistance development in *P. rapae*. Phylogenetic analysis revealed the AChE in *T. absoluta*, *L. orbonalis* and *S. exigua* are closely related and emerged from same primary branch. Three functional motifs were predicted in eleven species while only two were found in *L. orbonalis, S. exigua* and *T. absoluta*. AChE in eleven species followed secretory pathway and have signal peptides. No signal peptides were predicted for *S. exigua*, *L. orbonalis* and *T. absoluta* and follow non secretory pathway. Arginine methylation and cysteine palmotylation was found in all species except *S. exigua*, *L. orbonalis* and *T. absoluta*. Glycosylphosphatidylinositol (GPI) anchor was predicted in only nine species.

1. Introduction

The order Lepidoptera, encompasses butterflies and moths (Wahlberg et al., 2013), is the second most diverse insect order outnumbered only by the Coeleptera. There is hardly any cultivated plant that is not attacked by at least one lepidopteran pest. As pollinators of many plants, adult moths and butterflies are usually beneficial insects that help in pollination and suck flower nectar, but the larvae are chewing and feed on various parts of plant. Almost all the caterpillars are harmful for the plant by causing damage as defoliators, miners, bollworms or borers. These insects conjoined with angiosperms' and lie in the second highest harmful pest category of insect at the larval stage (Scoble 1992). Their devastating nature of damage make use of pesticides inevitable (Kulye et al., 2007; Firake et al., 2012). The nervous system is the best target site for effective and instant ontrol these insect pests.

Acetylcholinesterase (AChE) is a major enzyme of the nervous system as it regulates the acetylcholine level and catalyzes hydrolysis of acetylcholine to terminate nerve impulses. In the case of

* Corresponding author.
E-mail addresses: qudsia@cuisahiwal.edu.pk (Q. Yousafi), ayeshasarfraz39@gmail.com (A. Sarfaraz), saad.khan@cuisahiwal.edu.pk (M. Saad Khan), shahzadsaleem@cuisahiwal.edu.pk (S. Saleem), umbreenshahzad@bzu.edu.pk (U. Shahzad), drkupchani@bzu.edu.pk (A. Abbas Khan), masadiq@cuisahiwal.edu.pk (M. Sadiq), adabid5@gmail.com (A. Ditta Abid).

insects, AChE is a glycosylated dimer attached to the membrane through a glycolipid anchor (Chaabihi et al. 1994; Fournier et al. 1988). As AChE is the most important enzyme of the nervous system its inhibition results in death of insects (Rajashekar et al., 2014). Most commonly used insecticides for rapid control of these insect pests are organophosphates and carbamates (Fournier and Mutero, 1994). Acetylcholinesterase (AChE) is the main target site of organophosphate and carbamate pesticides. These insecticides act similar to acetylcholine but are hemi substrates because they phosphorylate or carbamoylate the active-site serine leading to irreversible inhibition of the enzyme (Corbet et al., 1974). This inhibition causes the accumulation of the acetylcholine in the synaptic region which leads to permanently opened acetylcholine receptors, resulting in the death of the target insect (Aldrige, 1950). First incidence of insensitivity of AChE to pesticides was reported 60 years ago (Smissaert, 1964) and several cases of insecticide resistance due to modification in AChE has been reported in many species of insect pests (Fournier and Mutero, 1994; Guedes et al., 1997). Important nervous system proteins like GABA receptor, Acetylcholinesterase and voltage-gated sodium channels have been reported to be involved in increased target site insensitivity to insecticides (Feyereisen, 1995)

Functional annotation of protein sequence with high accuracy has become one of the most important issues in understanding the molecular mechanism of life and has great biological significance (Goldstrohm et al., 2018). With the rapid increase of a wealth of protein sequences, the functional annotation of proteins has become increasingly challenging (Zhao et al. 2016). Less than 1% of entire protein sequences available in UniProt (Uniprot Consortium, 2018) have experimentally verified functions (Das et al., 2016; You et al., 2018; Tang et al., 2019) and it is estimated that about 90% of the annotated proteins in the ontology of biological process (BP) / molecular function (MF) (Ashburner et al., 2000) come from only nine species (Clark and Radivojac, 2011). Even for these nine model organisms, more than 60% of their proteins have not had any experimentally determined BP/MF term (Clark and Radivojac, 2011). Traditional methods for protein function annotation are mainly based on the experiments such as mass spectrometry, microscopy and RNA interference, which are reported as very time-consuming and resource-demanding because of the low throughput and restricted scope of methodology (Cao and Cheng 2016; Frasca and Cesa-Bianchi, 2017; Li et al., 2018). The computational approaches of significantly accelerated analysis process and enhanced accuracy are greatly desired for proteome studies (Cao et al., 2017; Fu et al., 2018; Zhu et al., 2018). Computational methods popular in current protein function prediction (Wan et al., 2017; Zhang et al., 2017; Jain and Kihara, 2019; Yousafi et al., 2019) can be roughly divided into three main classes: sequence, information and structure based (Cruz et al., 2017). Insect proteins are not very extensively studied. The characterization of the protein is important to have better knowledge about its function. Computational annotation of proteins will help us to know about the structure and function of a large number of unreviewed proteins in a short time and inexpensive manner.

2. Materials and methods

2.1. Retrieval of protein sequences

The amino acid sequences of AChE of fourteen lepidopteran insect pest species ,(Pieris rapae, Papilio machaon, Heliothis assulta, Heliothis armigera, Spodoptera frugiperda, Spodoptera exigua, Cydia pomonella, Leucinodes orbonalis, Tuta absoluta, Scirpophaga incertulas, Chilo suppressalis, Chilo auricilius, Papilio xuthus, Spodoptera litura), were retrieved from UniProtkb (http://www.uniprot.org/).

The retrieved sequences were saved in the FASTA format for computational study (Table 1).

2.2. Phylogenetic analysis

Phylogenetic study is an important approach to determine the evolutionary distances among species (Roy et al., 2014). The phylogram for AChE of selected insect pest species was generated through built-in neighbor joining method of MEGA X (Kumar et al., 2008).

2.3. Physiochemical properties

The amino acids primary sequence predicts various important physiochemical properties of protein, such as theoretical isoelectric point, aliphatic index, instability index, molecular weight, GRAVY and amino acid composition (Friedberg, 2005; Gasteiger et al., 2005). An online available server, Expassy ProtParam, was used to analyze the physiochemical properties of proteins.

2.4. Motif Prediction

Motif is a short conserved sequence of amino acids which stimulates a common function in a large number of proteins, such as to localize a protein within a cell toward the sub-cellular organelles. Motif prediction was done by using MEME suite (Bailey et al., 2009; Bailey and Elkan, 1994).

2.5. Signal Peptide prediction

Signal peptide is referred as approximately 16 to 30 amino acid long sequences located at the N-terminal of most of the nascent proteins. A freely available web tool "PrediSi (www.predisi.de) was used for signal peptide cleavage site prediction of AChE in the lepidopteran insect pests.

2.6. Subcellular localization

Subcellular localization of protein defines location of a protein within the cell. The subcellular localization of eukaryotic proteins was predicted by using a freely available online tool, DeepLoc-1 (www.cbs.dtu.dk › CBS Prediction Servers)

Table 1
UniProt IDs of acetylcholinesterase of some lepidopteran insect pest.

No.	Technical name	Common name	UniProt ID
1	Pieris rapae	Small white butterfly	A0A1U9X1Z7
2	Papilio Machaon	Old World swallowtail butterfly	A0A194RTQ3
3	Heliothis assulta	Oriental tobacco budworm	Q5RLH9
4	Heliothis armigera	Cotton bollworm	G1CRK3
5	Spodoptera frugiperda	Fall armyworm	A0A2H1VS32
6	Spodoptera exigua	Beet armyworm	S4VD29
7	Cydia pomonella	Codling moth	Q2XQ04
8	Leucinodes orbonalis	Brinjal shoot and fruit borer	J7FSX9
9	Tuta absoluta	Tomato leafminer	A0A218KFK3
10	Scirpophaga incertulas	Rice yellow stem borer	A0A1D8QQG3
11	Chilo suppressalis	Asiatic rice borer moth	B2BD82
12	Chilo auricilius	Gold-fringed rice borer	A0A076VJ71
13	Papilio xuthus	Asian swallowtail butterfly	A0A194PJR4
14	Spodoptera litura	Asian cotton leafworm	A0A1S5YCX9

2.7. Posttranslational modifications

The final tagging by the golgi bodies or the attachment of additional groups to C or N- terminus of the nascent protein is called Post-translational modification (PTMs) (Voet et al., 2006). Four PTMs on AChE of lepidopteran insect pest species have been predicted in the current study i.e., Palmotylation, by CSS-Palm 5.0 tool (csspalm.biocuckoo.org). ii. Methylation by GPS 3.0 (http://msp.biocuckoo.org/), iii. Lipidation by using GPS Lipid(http://lipid.biocuckoo.org/) and Myristoylator (https://web.expasy.org/myristoylator/) which predicts the N-terminus myristoylation of proteins by using neural networks (Casey and Seabra 1996) iv. Glycosylphosphatidylinositol (GPI) by a server Big-PI Predictor (http://mendel.imp.ac.at/sat/gpi/gpi_server.html)

2.8. Protein 3D model prediction, refinement and validation

Three dimensional protein models were predicted by a freely available online server I- TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy et al., 2010). ModRefiner a web based server was used to refine the predicted 3D protein models. The ModRefiner server refines terminus regions or loops by using *ab initio* modelling (Ko et al., 2012). Different online available tools were used to check and validate the quality of refined structures. SAVES server is a platform to access all these validation tools at once and is freely available at (http://services.mbi.ucla.edu/SAVES/.).

3. Results

3.1. Amino acid sequences retrieval and physiochemical properties analysis

The amino acid sequences of acetylcholinesterase of fourteen species of lepidoptera were retrieved from UniProt database (Table 1). The physiochemical properties of acetylcholinesterase of selected species were predicted (Table 2). High isoelectric point (pI) was found in *P. xhthus* (8.12) indicates its basic nature while the lowest pI value (4.94) in *P. machaon* showed acidic nature of protein. AChE of all other species exhibited the acidic nature with pI values between 5.01 and 5.65. GRAVY (Grand average of hydropathicity) reveals the average of hydropathicity of a given protein. The highest GRAVY value was recorded for *L. orbonalis* (-0.156) and that lowest for *T. absoluta* (-0.425). The instability index (II) depicts protein stability. The values of eleven species were less than 40 while *L. orbonalis, T. absoluta* and *S. exigua,* showed higher II values i.e., 41.00, 41.28 and 45.31, respectively. Similarly identifying aliphatic index (AI) is an important parameter to determine the thermostability of a given protein. In the current study, AChE in two species, *S. exigua* and *L. orbonalis,* had more thermostability having high AI's, 81.65 and 82.91, respectively. AChE of the remaining twelve species had AI values varying between 73 and 79 (Table 2). AChE in *P. rapae* is made up the highest percentage of five amino acids i.e., Arg, Asp, Asn, Gln and Cys which was 7.1% for each (Table 3).

Table 2
Physiochemical Properties of acetylcholinesterase of lepidopteran insect pest species.

Organism	UniProt ID	No of A.A	Molecular weight	GRAVY	Negatively charged residues	Positively charged residues	Theoretical isoelectric point(Pi)	Instability index	Aliphatic index
Pieris rapae	A0A1U9X1Z7	638	71737.49	−0.208	72	49	5.26	39.02	75.69
Papilio machaon	A0A194RTQ3	530	59824.93	−0.218	66	43	4.94	40.65	79.83
Heliothis assulta	Q5RLH9	647	72605.57	−0.217	72	53	5.42	36.21	74.59
Heliothis armigera	G1CRK3	647	72665.63	−0.227	72	54	5.47	37.05	73.99
Spodoptera frugiperda	A0A2H1VS32	638	71669.39	−0.243	71	52	5.46	36.26	74.12
Spodoptera exigua	S4VD29	418	47095.38	−0.275	54	38	5.01	41.28	81.65
Cydia pomonella	Q2XQ04	638	71412.17	−0.216	70	53	5.44	38.71	75.96
Leucinodes orbonalis	J7FSX9	361	40586.33	−0.156	38	27	5.3	45.31	82.91
Tuta absoluta	A0A218KFK3	607	68792.8	−0.425	75	59	5.41	41	76.26
Scirpophaga incertulas	A0A1D8QQG3	638	71776.63	−0.240	72	54	5.43	38.05	75.5
Chilo suppressalis	B2BD82	638	71698.62	−0.225	70	55	5.66	38.31	75.82
Chilo auricilius	A0A076VJ71	638	71653.49	−0.229	71	54	5.54	39.1	75.36
Papilio xuthus	A0A194PJR4	499	54735.17	−0.326	47	49	8.12	41.4	74.83
Spodoptera litura	A0A1S5YCX9	638	71669.39	−0.243	71	52	5.46	36.26	74.12

Table 3
Percentage of amino acids in acetylcholinesterase of different insect pest species of Lepidoptera order.

Organism	Amino acids																			
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Pieris rapae	7.1	7.1	7.1	7.1	7.1	7.1	5.8	7.7	3.0	4.7	8.9	4.2	3.3	5.5	6.4	6.4	7.2	2.5	4.7	5.3
Papilio machaon	7.1	3.2	4.0	6.0	1.1	3.6	6.4	7.7	2.3	5.3	9.1	4.9	3.0	5.7	5.8	6.4	6.4	2.3	4.7	6.2
Heliothis assulta	6.2	3.7	3.9	5.9	1.5	2.6	5.6	8.0	2.8	4.3	8.5	4.5	3.6	5.4	6.5	6.3	7.7	2.3	4.6	6.3
Heliothis armigera	6.2	3.9	3.7	5.9	1.5	2.6	5.3	7.9	2.8	4.2	8.5	4.5	3.6	5.4	6.5	6.5	7.9	2.3	4.6	6.3
Spodoptera frugiperda	6.1	3.6	3.8	5.8	1.6	2.8	5.3	8.2	3.0	4.4	8.6	4.5	3.4	5.2	6.3	6.6	7.7	2.5	4.7	6.0
Spodoptera exigua	7.2	5.5	5.3	6.0	1.2	3.3	6.9	7.2	1.9	5.3	8.9	3.6	2.6	5.3	6.7	5.5	5.3	1.9	3.8	6.7
Cydia pomonella	6.3	3.9	4.1	5.6	1.6	2.7	5.3	8.2	2.5	4.5	8.8	4.4	3.3	5.2	6.4	6.9	7.4	2.4	4.5	6.1
Leucinodes orbonalis	6.9	4.2	6.6	4.7	1.4	3.3	5.8	7.2	2.5	5.3	8.9	3.3	3.3	5.3	5.8	6.6	4.7	1.7	5.0	7.2
Tuta absoluta	5.8	4.9	6.8	4.9	1.3	2.5	7.4	6.8	2.1	4.8	7.9	4.8	2.6	4.1	7.2	6.4	5.3	2.1	4.9	7.2
Scirpophaga incertulas	5.8	3.8	3.8	6.0	1.7	2.7	5.3	8.0	2.7	4.9	8.9	4.7	3.3	5.2	6.4	6.9	7.4	2.5	4.7	5.5
Chilo suppressalis	6.3	3.8	3.8	5.5	1.4	2.7	5.5	8.2	2.8	4.7	8.9	4.9	3.3	5.5	6.3	6.7	7.2	2.4	4.7	5.6
Chilo auricilius	6.4	3.8	3.8	5.6	1.4	2.7	5.5	8.2	2.8	4.7	8.8	4.7	3.3	5.5	6.4	6.6	7.2	2.4	4.7	5.6
Papilio xuthus	6.6	5.2	4.4	5.2	1.2	3.2	4.2	9.6	2.8	4.2	8.2	4.6	2.8	3.8	7.4	7.2	7.4	2.2	2.8	6.8
Spodoptera litura	6.1	3.6	3.8	5.8	1.6	2.8	5.3	8.2	3.0	4.4	8.6	4.5	3.4	5.2	6.3	6.6	7.7	2.5	4.7	6.0

3.2. Phylogenetic analysis

The relatedness and differences of AChE among different species was studied with phylogenetic analysis. Phylogenetic tree was generated with the help of MEGAX tool (Fig. 1). The range of data coverage for analysis found more than 80%. All the ambiguous positions were removed for each sequence pair and the refined data set consisted of 647 positions. The phylogenetic tree was initially bifurcated in two primary branches. One of these two branches (distance = 1.246375) further divided into three, S. exigua, T. absoluta and L. orbonalis. Other primary branch (distance = 1.567907) gave rise to remaining eleven species. The AChE amino acid sequences of two species, S. exigua and T. absoluta, were found to be closely related and with that of L. orbonalis, but different from the other eleven species.

3.3. Functional motif prediction

Three AChE motifs were predicted for eleven species but only two were found in three species S. exigua, T. absoluta and L. orbonalis. Motif 1 (Fig. 2a, Table 4a) had an E value of 4.4 e^{-540}. Closely related motif 1 was found in eleven species. Eight out of eleven species, S. litura, P. xuthus, C. auricilius, S. incertulas, S. frugiperda, H. armigera, H. assulta, and P. machaon had exactly similar motifs with a p-value of 8.67 e^{-64}. P. rapae ($2.77e^{-62}$) and C. suppressalis ($4.18e^{-62}$) were slightly different from those with replacement of R19 with Q and D16 with E, respectively. C. pomonella ($4.15e^{-61}$) was slightly different from ten species with the replacement of A20 by V and E26 by D. Among the rest of the three species, L. orbonalis ($8.68e^{-46}$) and T. absoluta ($1.17e^{-46}$), were different from S. exigua ($1.13e^{-47}$) with replacement of D16 with E and I I28 with V, respectively. Motif 2 (Fig. 2b) predicted in all the species having an E value of 9.2e^{-515}, started from 14th positon and 50 amino acids long. Ten species had exactly similar motif with p-value 3.45e^{-63} (Table 4b). P. rapae with p-value 2.14e^{-61} was closely related to those ten species by replacement of only one amino acid S27 by A. AChE Motif 2 in remaining three species were found distantly related with those of the former ten species. L. orbonalis and T. absoluta were similar to each other with p-value 1.74e^{-44}, but different from S. exigua ($3.54e^{-45}$) with replacement of A22 with P. AChE Motif 3 (Fig. 2c, Table 4c) was predicted having E value 6.2e^{-524} started at the 11th amino acid and fifty amino acids wide. It was found in eleven species among which nine species showed exactly the same motif with p-value 9.68e^{-68}. Two species, C. auricilius ($1.94e^{-67}$) and C. suppressalis ($1.08e^{-66}$), were slightly different with replacement of D50 with K and E, respectively.

3.4. Signal peptide prediction

The signal peptide cleavage site along with their score were predicted, for AChE of fourteen lepidopteran insect pest species, by PrediSi (Table 5). Eleven species were found to have strong signal peptides because they scored greater than 0.5. Last four amino acids in c-region of signal peptides were consensus sequence 'RSWA'. Fig. 3 shows the cleavage sites in which the true cleavage site is marked with an arrow (Nielsen et al. 1997). The signal peptides were not predicted in three species, L. orbonalis, S. exigua and T. absoluta.

3.5. Subcellular localization prediction

Subcellular localization was predicted through DeepLoc 1.0 (Table 6). It predicts the localization within eight different sub cellular locations as well as the extracellular parts of the protein i.e., endoplasmic reticulum, nucleus, cytoplasm, mitochondrion, cell membrane, peroxisome, golgi apparatus and lysosome/vacuole. The scores indicate that the AChE of S. exigua, L. orbonalis and T. absoluta follow a non-secretory pathway and lack a signal peptide. The remaining eleven species followed a secretory pathway with a probability score of 1. Out of the eleven species following a
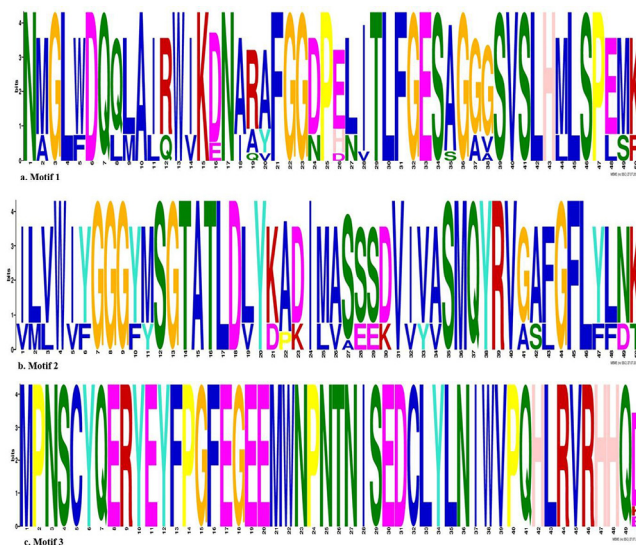

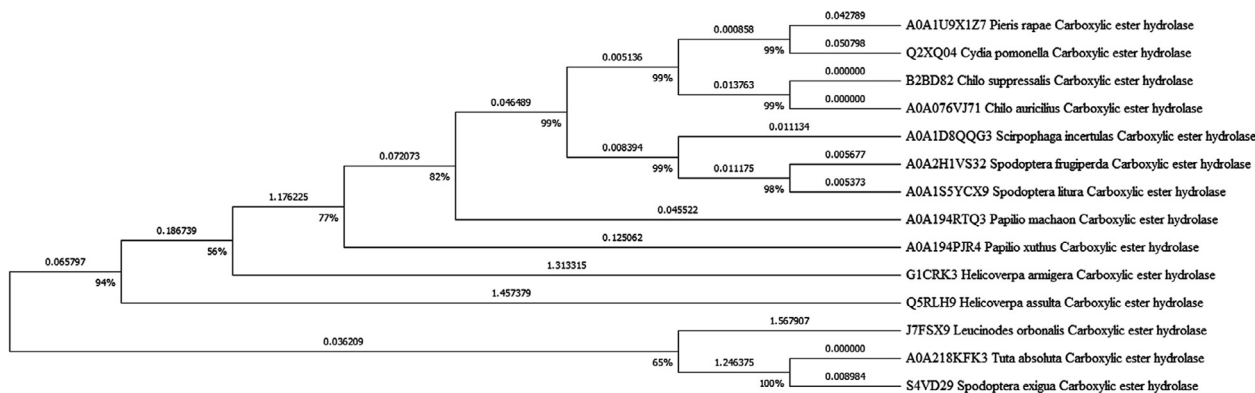**Fig. 2.** Consensus logo for motifs in acetylcholinesterase (AChE) of lepidopteran insect pest species.

Fig. 1. Phylogenetic tree of acetylcholinesterase (AChE) of lepidopteran insect pest species.

Table 4a
Curated Motif 1 location and probabilities in different lepidopteran insect pest species.

Organism	Start	p-value	Start Flanking site	Motif site	End Flanking site
Spodoptera litura	233	$8.67e^{-64}$	FSPGSEEAPG	NMGLWDQQLAIRWIKDNARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Papilio xuthus	233	$8.67e^{-64}$	FSPGSEEAPG	NMGLWDQQLAIRWIKDNARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Chilo auricilius	233	$8.67e^{-64}$	FSAGSEEAPG	NMGLWDQQLAIRWIKDNARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Scirpophaga incertulas	233	$8.67e^{-64}$	FSSGSEEAPG	NMGLWDQQLAIRWIKDNARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	DLFKRGILQS
Spodoptera frugiperda	233	$8.67e^{-64}$	FSPGSEEAPG	NMGLWDQQLAIRWIKDNARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Helicoverpa armigera	242	$8.67e^{-64}$	FSPGSEEAPG	NMGLWDQQLAIRWIKDNARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Helicoverpa assulta	242	$8.67e^{-64}$	FSPGSEEAPG	NMGLWDQQLAIRWIKDNARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Papilio machaon	233	$8.67e^{-64}$	FSPGSEEAPG	NMGLWDQQLAIRWIKDNARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Pieris rapae	233	$2.76e^{-62}$	FSPGSEEAPG	NMGLWDQQLAIRWIKDNAQAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Chilo suppressalis	233	$4.18e^{-62}$	FSAGSEEAPG	NMGLWDQQLAIRWIKENARAFGGDPELITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Cydia pomonella	233	$4.15e^{-61}$	FSPGSEEAPG	NMGLWDQQLAIRWIKDNARVFGGDPDLITLFGESAGGGSVSLHMLSPEMK	GLFKRGILQS
Spodoptera exigua	200	$1.13e^{-47}$	LFFDTPDVPG	NAGLFDQLMALQWVKDNIAYFGGNPHNITLFGESAGAVSVSLHLLSPLSR	NLFSQAIMQS
Leucinodes orbonalis	85	$1.17e^{-46}$	LFFDTPDVPG	NAGLFDQLMALQWVKENIAYFGGNPHNITLFGESAGAVSVSLHLLSPLSR	NLFSQAIMQS
Tuta absoluta	223	$8.68e^{-46}$	LFFDTPDVPG	NAGLFDQLMALQWVKDNIAYFGGNPHNVTLFGESSGAASVSLHLLSPLSR	NYFSQAIMQS

Table 4b
Curated Motif 2 location and probabilities in different lepidopteran insect pest species.

Organism	Start	p-value	Start Flanking site	Motif site	End Flanking site
Spodoptera litura	172	$3.45e^{-63}$	KPLTERPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	YFSPGSEEAP
Papilio xuthus	172	$3.45e^{-63}$	KPLTERPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	FFSPGSEEAP
Chilo auricilius	172	$3.45e^{-63}$	KPLTEKPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	YFSAGSEEAP
Chilo suppressalis	172	$3.45e^{-63}$	KPLTEKPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	YFSAGSEEAP
Scirpophaga incertulas	172	$3.45e^{-63}$	KPLTEKPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	YFSSGSEEAP
Cydia pomonella	172	$3.45e^{-63}$	KPLAERPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	FFSPGSEEAP
Spodoptera frugiperda	172	$3.45e^{-63}$	KPLTERPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	YFSPGSEEAP
Helicoverpa armigera	181	$3.45e^{-63}$	KPLAERPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	YFSPGSEEAP
Helicoverpa assulta	181	$3.45e^{-63}$	KPLAERPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	YFSPGSEEAP
Papilio machaon	172	$3.45e^{-63}$	KPLTERPKVP	ILVWIYGGGYMSGTATLDLYKADIMASSSDVIVASMQYRVGAFGFLYLNK	YFSPGSEEAP
Pieris rapae	172	$2.14e^{-61}$	KPLAERPNVP	ILVWIYGGGYMSGTATLDLYKADIMAASSDVIVASMQYRVGAFGFLYLNK	FFSPGSEEAP
Spodoptera exigua	145	$3.54e^{-45}$	TPRPRPKNAA	VMLWVFGGGFYSGTATLDVYDAKILVSEEKVVYVSMQYRVASLGFLFFDT	PDVPGNAGLF
Tuta absoluta	168	$1.74e^{-44}$	VPKPRPKNAA	VMLWVFGGGFYSGTATLDVYDPKILVSEEKVVYVSMQYRVASLGFLFFDT	PDVPGNAGLF
Leucinodes orbonalis	30	$1.74e^{-44}$	SPRPRPQNAA	VMLWVFGGGFYSGTATLDVYDPKILVSEEKVVYVSMQYRVASLGFLFFDT	PDVPGNAGLF

Table 4c
Curated Motif 3 location and probabilities in different lepidopteran insect pest species.

Organism	Start	p-value	Start Flanking site	Motif site	End Flanking site
Spodoptera litura	112	9.68e⁻⁶⁸	PWHGVLEATA	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLTERPKVP
Papilio xuthus	112	9.68e⁻⁶⁸	PWHGVLEATA	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLTERPKVP
Scirpophaga incertulas	112	9.68e⁻⁶⁸	ʾWHGVLEATS	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLTEKPKVP
Cydia pomonella	112	9.68e⁻⁶⁸	PWHGVLDAT	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLAERPKVP
Spodoptera frugiperda	112	9.68e⁻⁶⁸	PWHGVLEATA	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLTERPKVP
Helicoverpa armigera	121	9.68e⁻⁶⁸	PWHGVLEATA	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLAERPKVP
Helicoverpa assulta	121	9.68e⁻⁶⁸	PWHGVLEATA	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLAERPKVP
Papilio machaon	112	9.68e⁻⁶⁸	PWHGVLEATA	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLTERPKVP
Pieris rapae	112	9.68e⁻⁶⁸	PWHGILEATS	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQD	KPLAERPNVP
Chilo auricilius	112	1.94e⁻⁶⁷	PWHGVLEATA	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQE	KPLTEKPKVP
Chilo suppressalis	112	1.08e⁻⁶⁶	PWHGVLEATA	MPNSCYQERYEYFPGFEGEEMWNPNTNISEDCLYLNIWVPQLRVRHHQK	KPLTEKPKVP

secretory pathway, in nine species it is localized in the cell membrane with a high probability score of 0.9463 to 0.9953. In two species, *P. xuthus* and *P. machaon,* although it followed the secretory pathway but did not localize in the cell membrane and remained in the extracellular region.

3.6. Posttranslational modification

Three posttranslational modifications i.e., palmotylation, methylation and Glycosylphosphatidylinositol (GPI) were predict in AChE of fourteen lepidopteran insect pest species while lipidation was missing among all the species. The cutoff values for palmotylation, methylation prediction were kept 3.717 and 4.11 respectively, the results with higher scores were selected. The score shows the probability of predicted amino acid position as a true PTM site for a particular thresholds (Xue et al., 2008). Methylation is referred as addition of a methyl group to lysine side chain (Saraswathy and Ramalingam, 2011). In selected insect pest species the AChE showed arginine (R: bold, highlighted and red) methylation in eleven species (Table 7). The sequence '**YYYYFTHRTSTSLWG**' of one methylation position was the same among all species while the sequence '**EKWPLYSRSSPHYYT**' for the other position was the same in all species except *H. assulta* and *H. armigera* i.e., '**EKWPLYSRTSPHYYT**' only S after R was replaced by T. Highest score (4.86) was found in *P. rapae, S. frugiperda, S. incertulas, C. suppressalis, C. auricilius* and *S. litura* at position 563 for sequence '**EKWPLYSRSSPHYYT**'. *H. assulta* and *H. armigera* showed the second highest score (4.74) at position 572 for sequence '**EKWPLYSRTSPHYYT**'. This may be considered as strong evidence for methylation sites compared to positions with lower scores. Palmotylation is a covalent attachment of palmitic acid to cysteine (S-palmitoylation) and rarely to threonine and serine (O-palmitoylation) residues of proteins (Linder, 2001). Cysteine palmitoylation was found in AChE of nine insect pest species (Table 8). The selected cutoff value was equal (3.717) for all species and the predicted residue is underlined highlighted in red. *P. machaon* and *P. xuthus* had only one palmotylation site '**DPSLVMDCMRGVDAK**' at position 325. *P. rapae* had two palmotylation sites '**FTKLLLCCFVAAAWA**' and '**DPSLVMDCMRGVDAK**' at positions 16 and 325, respectively. The remaining six species had three palmotylation sites. The sequence '**DPSLVMDCMRGVDAK**' was a common palmotylation site among all species. Highest score (37.22) was shown for *P. rapae* at position 16 followed by those of *S. litura* (33.925) at position 16, *S. frugiperda* (33.925) at position 16, *C. auricilius* (25.397) at position 15 and *C. pomonella* (24.966) at position 15. Which indicates that these five species have strong evidence for palmotylation among all the species. Big-PI Predictor predicted GPI anchors on AChE in nine insect pest species (Table 9). The results presented for GPI anchor prediction are given by expected ω position with a particular score. The higher score is a preferred site and the amino acid colored red. An alternative option is also given with a relatively lower score and amino acid colored yellow. It is obvious from the results that the preferred site is with amino acid residue **C**. Two positions have been predicted for GPI anchor in all the species except *S. incertulas* and *C. Suppressalis* Two species, *S. incertulas* and *C. suppressalis* which showed only one position, **S,** with a lower score.

3.7. Tertiary structure prediction, evaluation, refinement

The 3D structures and suitable templates for homology modeling of AChE of selected lepidopteran species, were not available. I-TASSER server was used for 3D model prediction, as it uses multiple templates to predict protein structure (Zhang, 2008). For quality estimation of I-TASSER generated five models, against each query sequence, with different confidence score (C-score). Good

Table 5
Signal Peptide prediction in acetylcholinesterase of lepidopteran insect pest species.

Organism	UniProt ID	Signal peptide sequence	C-site	C-site score	Mode of movement
Pieris rapae	A0A1U9X1Z7	MTSQNRIVFTKLLLCCFVAAAWARSWA	27	1.0000	Secretory
Papilio machaon	A0A194RTQ3	MSGHDKIVYTKLLLLLFVSSAWSRSWA	27	1.0000	Secretory
Helicoverpa assulta	Q5RLH9	MPVRTTIFEMISNTKIVFTKLLLCCFVSGAVARSWA	36	0.8968	Secretory
Helicoverpa armigera	G1CRK3	MPVRTTTFEMISSTKIVFTKLLLCCFVSGAVARSWA	36	0.8968	Secretory
Spodoptera frugiperda	A0A2H1VS32	MISNTKIVFTKLLLCCIVSGAWTRSWA	27	0.8907	Secretory
Spodoptera exigua	S4VD29	No signal	62	0.1689	non Secretory
Cydia pomonella	Q2XQ04	MTCNTKIVITKLLVCFLLSGVRGRSWA	27	0.8027	Secretory
Leucinodes orbonalis	J7FSX9	No signal	45	0.4364	non Secretory
Tuta absoluta	A0A218KFK3	No signal	0	0.0000	non Secretory
Scirpophaga incertulas	A0A1D8QQG3	MCCNRKIVFTKLLLCIFVSGAWGRSWA	27	0.9523	Secretory
Chilo suppressalis	B2BD82	MSRNIKIVFTKLLLCFFVSGAFGRRSWA	27	0.8910	Secretory
Chilo auricilius	A0A076VJ71	MSRNIKIVFTKLLPCFFVSGAFGRSWA	27	0.7430	Secretory
Papilio xuthus	A0A194PJR4	MSGHDKIVYTKLLLLLFVSGAWSRSWA	27	1.0000	Secretory
Spodoptera litura	A0A1S5YCX9	MSGHDKIVYTKLLLLLFVSGAWSRSWA	27	0.8907	Secretory

quality models with high C score were selected for each species. The evaluation and refinement of selected models were performed by QMEAN and SAVES, to further improve the quality and reliability. SAVES (Structure Analysis and Verification Server) consists of different programs i.e., ERRAT, Verify3D, Prove, RAMPAGE, for structural evaluation and refinement. The structures were selected which have ERRAT value above 60%. Percentage of amino acids in favored and allowed regions were shown by RAMPAGE values. The models having amino acids with more than 80% in allowed and favored regions were selected (Table 10).

4. Discussion

Insecticide resistance is very challenging problem to be addressed in agricultural sector worldwide. Most of the insecticide have been developed to target the nervous system (GABA receptor and acetylcholinesterase) and interfere nerve impulse transmission to get quick control (Keane et al., 1999; Tong and Coats 2012). *In silico* genes and protein analysis have been receiving attention and are popular among scientific community, because it plays an important role in getting base line information in a swift and reliable manner. In the present study we have used computational approaches to find out possible reasons of insecticide resistance at protein level using AChE as a target enzyme. In protein study the primary structure depicts the most important information about nature and function of protein. Different properties i.e., GRAVY, instability index, aliphatic index and amino acid composition of protein have been studied using bioinformatics tools. GRAVY (Grand average of hydropathicity) reveals the average of hydropathicity of a given protein. High positive score of GRAVY makes a peptide more hydrophobic and negative score makes a protein hydrophilic (Kyte and Doolittle, 1983; Roy et al., 2011). All the selected species contain negative values of GRAVY which indicate their hydrophilic nature. The instability index (II) depicts protein stability. If its II value is lower than 40 the protein is considered stable (Roy et al., 2011). The values of eleven species were less than 40 which indicated their stability while *L. orbonalis, T. absoluta* and *S. exigua,* showed II values 41.00, 41.28 and 45.31, respectively, which indicated an unstable nature of AChE in these species. Similarly identifying aliphatic index (AI) is an important parameter to determine the thermostability of a given protein. The higher the aliphatic index value the greater the thermostability of proteins (Ikai, 1980). In the current study, two species, *S. exigua* and *L. orbonalis,* had more thermostability having high AI's, 81.65 and 82.91, respectively. AChE in *P. rapae* made up the highest percentage of five amino acids i.e., Arg, Asp, Asn, Gln and Cys which was 7.1% for each. Cysteine and Glutamine might be the reason for insecticide resistance in *P. rapae*. Cysteine plays significant role

in protein structure stability and might be in insecticide resistance, because it is involved in esterase -organophosphate molecules interaction (Krejci et al., 1991). Cysteine is required for a unique posttranslational modification induced by oxidative stress, in factor Keap1 (Zhang and Hannink, 2003). Expression level of Keap1 was found to be higher in pyrethroid and OP resistant *Aedes aegypti* population (Bottino-Rojas et al., 2018). Gln is an important amino acid playing crucial role in imidacloprid sensitivity of nicotinic acetylcholine receptor (Shimomura et al., 2002). Amino acid composition of the protein can be helpful to get an idea about the protein function and especially insecticide resistance in insect pests. Two knock down resistance mutations i.e., valine 419 to leucine and leucine 925 to isoleucine, were found linked with pyrethroid resistance in *Cimex lectularius* (Yoon et al., 2008).Three point mutations, phenylalanine 115 to serine (F115S), isoleucine 199 to threonine/valine (I199T/V) and glycine 303 to alanine (G303A) in the AChE of *D. melanogaster* were identified and found associated with OPs and carbamate resistance (Mutero et al., 1994).

The phylogenetic analysis was done to find out relatedness and differences of AChE among different species. Phylogenetic tree was generated with the help of MEGAX tool. Multiple sequence alignment in MEGAX is done by ClustalW and MUSCLE (Kumar et al., 2018). The phylogenetic analysis was generated by using aligned regions while nonaligned regions were referred as deletions (Horiike, 2016). The sum of branch length for optimally generated phylogenetic tree was 7.50317185. The branch length depicts the evolutionary time, in terms of unit number of amino acid substitutions per site, calculated through Poisson correction method (Zuckerkandl and Pauling, 1965). The data coverage for tree construction was above 80% which reflects the accuracy of input query sequence (Filipski et al., 2014). The closely related species originated from the same primary branch which reflects the reliability of the predicted phylogenetic relationship (Horiike, 2016). In the current study almost related genera originated from common branch e.g. *C. suppressalis, H. armigera* and *S. litura* this shows accuracy of predicted results for species relatedness. *Leucinodes orbonalis, S. exigua* and *T. absoluta* were originated as outgroup.

A three dimensional conserved structure of a protein that stimulates the distinct biological function among different types of proteins is called a motif (Xiong, 2006). The MEME suit was used for functional motifs prediction in AChE of fourteen lepidopteran species. The MEME algorithm (Bailey and Elkan, 1994) has been widely used for the discovery of DNA and protein sequence motifs and MEME continues to be the starting point for most analyses using the MEME Suite. Most of the motif prediction tools failed to predict the gapped motifs but a recent algorithm GLAM2 (Frith et al., 2008) has been incorporated into the MEME suite for gapped motif discovery. MEME usually predicts the most

Table 6
Likelihood of subcellular localization acetylcholinesterase in some lepidopteran insect pest species.

Organism name	Uniprot ID	Extracellular localization	Localization inside the cell								Pathway followed	
			Endoplasmic reticulum	Lysosome	Cell membrane	Cyto plasm	Golgi apparatus	Mitochondrion	Peroxisome	Nucleus	Secretory	Non Secretory
Papilio machaon	A0A194RTQ3	0.4867	0.3765	0.2101	0.0224	0	0.0011	0.0007	0.0002	0	0.998	0.002
Pieris rapae	A0A1U9X1Z7	0	0.0135	0.0065	0.9781	0	0.0013	0.0002	0.0003	0	1	0.00
Helicoverpa assulta	Q5RLH9	0	0.0054	0.0053	0.9886	0	0.0005	0	0	0	1	0.00
Helicoverpa armigera	G1CRK3	0	0.0114	0.0186	0.9689	0	0.0008	0.0001	0.0002	0	1	0.00
Spodoptera frugiperda	A0A2H1VS32	0.0016	0.0152	0.0197	0.961	0	0.0018	0.0002	0.0004	0	1	0.00
Spodoptera exigua	S4VD29	0.1214	0.1108	0.1148	0.0473	0.314	0.0368	0.0458	0.1316	0.1527	0.26	0.74
Cydia pomonella	Q2XQ04	0.0018	0.0433	0.0085	0.9436	0	0.0013	0.0004	0.0009	1	1	0.00
Leucinodes orbonalis	J7FSX9	0.1617	0.0616	0.0604	0.0171	0.2843	0.0131	0.0699	0.2208	0.0100	0.23	0.77
Tuta absoluta	A0A218KFK3	0.0478	0.143	0.195	0.0296	0.295	0.0265	0.0201	0.0424	0.0481	0.484	0.516
Scirpophaga incertulas	A0A1D8QQG3	0	0.0019	0.0025	0.9953	0	0.0002	0	0.0001	0	1	0.00
Chilo suppressalis	B2BD82	0.0001	0.0034	0.0016	0.9945	0	0.0003	0.0001	0.0001	0	1	0.00
Chilo auricilius	A0A076VJ71	0	0.008	0.0128	0.9774	0	0.0012	0.0002	0.0003	0	1	0.00
Papilio xuthus	A0A194PJR4	0.9875	0.1259	0.2033	0.0113	0	0.0003	0.0001	0	0	1	0.00
Spodoptera litura	A0A1S5YCX9	0.0016	0.0152	0.0197	0.961	0	0.0018	0.0002	0.0004	0	1	0.00

statistically significant (low E-value) motifs first. E value shows the statistical significance of motif prediction *i.e.*, lower the E value, higher the significance (Bailey et al., 2009). The results are produced in the form of consensus logo and a table of motif alignment. The stacked amino acid letter codes represent their relative frequencies among the compared protein sequences. The height of letters in consensus logo shows the degree of conservation of each sequence (Schneider and Stephens, 1990). Similar colors have been assigned to amino acids from the same class (Thomsen and Nielsen, 2012). Motifs are arranged in ascending order of the p-value in the table. The p-value shows the statistical significance of similarities among the compared sequences. Two species, *C. auricilius* ($1.94e^{-67}$) and *C. suppressalis* ($1.08e^{-66}$), were slightly different with replacement of D50 with K and E, respectively. Only one amino acid has been replaced in both species but p-value for *C. suppressalis* was greater than that of *C. auricilius* because D and E are from different classes. The lower the p-value the greater the sequence similarity (Bailey et al., 1998). The predicted motifs confirmed the phylogenetic analysis results in current study. Three motifs have been predicted in eleven species while outgroup species *L. orbonalis*, *S. exigua* and *T. absoluta* showed only two.

Signal peptide is usually a 16–30 amino acid long sequence (Voss et al., 2013). It is located at N-terminal of most of the nascent proteins that are directed towards the secretory pathway (Rapoport, 1992) to endomembrane system i.e., endoplasmic reticulum and golgi apparatus, in which proteins undergo folding and posttranslational modifications (Rapoport, 2007). A general structure of the signal peptide is composed of three main parts: n-region (N-terminus), h- region (hydrophobic) and c-region (cleavage) (Voss et al., 2013). Cleavage site (c-region) prediction of signal peptide is crucial to confirm the presence or absence of signal peptide (Nielsen et al., 1997). Most commonly used method is weight matrix in the prediction method to discriminate between signal peptides and non-signal peptide sequence by using the maximum cleavage site score. The signal peptide cleavage site along with their score were predicted, for AChE of fourteen lepidopteran insect pest species, by PrediSi. PrediSi provides a normalized score on a scale between 0 and 1. A score greater than 0.5 means that the amino acid sequence very likely contains a signal peptide (Hiller et al., 2004). The benefit of this simple score method is that it is comparable between different weight matrices. Eleven species were found to have strong signal peptides because they scored greater than 0.5. Last four amino acids in c-region of signal peptides were consensus sequence 'RSWA'. The amino acid on the cleavage site was Alanine (A) which is a nonpolar and uncharged or neutral amino acid (Hundal et al., 1989). This also supports our results of signal peptide prediction because the amino acid in cleavage region/site must be small and neutral for ease of correct cleavage (von Heijne, 1985). Three outgroup species, *L. orbonalis*, *S. exigua* and *T. absoluta,* showed no signal peptides and are supposed to use unconventional mechanisms, other than secretory, for translocation (Nickel and Seedorf, 2008).

Function of a protein is reflected by its subcellular localization (Hung and Link 2011) and its prediction is very important for functional study of protein (Wan and Mak, 2015). Acetylcholinesterase is mainly responsible for synaptic transmission in the cells by catalyzing hydrolysis of acetylcholine. Its localization in non-cholinergic and non-neural cells and tissues (Soreq and Seidman 2001), and the variations in its molecular forms (Meshorer et al., 2002) are evident of its non-classical and diversified functions. Deeploc 0.1 server was used for AChE subcellular localization prediction. It predicted the localization within eight different sub cellular locations as well as the extracellular parts of the protein i.e., endoplasmic reticulum, nucleus, cytoplasm,

Table 7
Methylation sites in acetylcholinesterase of lepidopteran insect pest species.

Organism	UniProt ID	Methylation			
		Position	Peptide	Score	Cutoff
Pieris rapae	A0A1U9X1Z7	495	YYYYFTH**R**TSTSLWG	4.36	4.11
		563	EKWPLYS**R**SSPHYYT	4.86	4.11
Helicoverpa assulta	Q5RLH9	504	YYYYFTH**R**TSTSLWG	4.36	4.11
		572	EKWPLYS**R**TSPHYYT	4.74	4.11
Helicoverpa armigera	G1CRK3	504	YYYYFTH**R**TSTSLWG	4.36	4.11
		572	EKWPLYS**R**TSPHYYT	4.74	4.11
Spodoptera frugiperda	A0A2H1VS32	495	YYYYFTH**R**TSTSLWG	4.36	4.11
		563	EKWPLYS**R**SSPHYYT	4.86	4.11
Cydia pomonella	Q2XQ04	495	YYYYFTH**R**TSTSLWG	4.36	4.11
Scirpophaga incertulas	A0A1D8QQG3	495	YYYYFTH**R**TSTSLWG	4.36	4.11
		563	EKWPLYS**R**SSPHYYT	4.86	4.11
Chilo suppressalis	B2BD82	495	YYYYFTH**R**TSTSLWG	4.36	4.11
		563	EKWPLYS**R**SSPHYYT	4.86	4.11
Chilo auricilius	A0A076VJ71	495	YYYYFTH**R**TSTSLWG	4.36	4.11
		563	EKWPLYS**R**SSPHYYT	4.86	4.11
Spodoptera litura	A0A1S5YCX9	495	YYYYFTH**R**TSTSLWG	4.36	4.11
		563	EKWPLYS**R**SSPHYYT	4.86	4.11

Table 8
Palmotylation sites in acetylcholinesterase of lepidopteran insect pest species.

Organism	UniProt ID	Palmotylation			
		Position	Peptide	Score	Cutoff
Pieris rapae	A0A1U9X1Z7	16	FTKLLLC**C**FVAAAWA	37.223	3.717
		325	DPSLVMD**C**MRGVDAK	6.242	3.717
Papilio machaon	A0A194RTQ3	325	DPSLVMD**C**MRGVDAK	5.801	3.717
Helicoverpa assulta	Q5RLH9	344	DPSLVMD**C**MRGVDAK	5.873	3.717
		25	FTKLLLC**C**FVSGAVA	6.888	3.717
		329	VLVDD**C**NCNSSLLAA	4.356	3.717
Helicoverpa armigera	G1CRK3	344	DPSLVMD**C**MRGVDAK	5.873	3.717
		25	FTKLLLC**C**FVSGAVA	6.888	3.717
		329	VLVDDCN**C**NSSLLAA	4.356	3.717
Spodoptera frugiperda	A0A2H1VS32	16	FTKLLLC**C**IVSGAWT	33.925	3.717
		320	VLVDDCN**C**NSSLLAA	4.356	3.717
		335	DPSLVMD**C**MRGVDAK	5.873	3.717
Cydia pomonella	Q2XQ04	15	VITKLLV**C**FLLSGVR	24.966	3.717
		320	VLVDDCN**C**NSSLLSA	4.49	3.717
		335	DPSLVMD**C**MRGVDAK	5.547	3.717
Chilo auricilius	A0A076VJ71	15	VFTKLLP**C**FFVSGAF	25.397	3.717
		320	VLVDDCN**C**NSSLLAA	4.356	3.717
		335	DPSLVMD**C**MRGVDAK	5.873	3.717
Papilio xuthus	A0A194PJR4	325	DPSLVMD**C**MRGVDAK	5.801	3.717
Spodoptera litura	A0A1S5YCX9	16	FTKLLLC**C**IVSGAWT	33.925	3.717
		320	VLVDDCN**C**NSSLLAA	4.356	3.717
		335	DPSLVMD**C**MRGVDAK	5.873	3.717

mitochondrion, cell membrane, peroxisome, golgi apparatus and lysosome/vacuole. The results are generated in terms of probability of likelihood to be present in a location. The pathway of protein is defined by first bifurcation of secretory and non-secretory pathway then further scoring is done for compartmentalization. The joint probability of decisions, on the whole, is used to calculate a likelihood score i.e., 0–1 (Almagro et al., 2017). The scores indicate that the AChE of outgroup species *S. exigua*, *L. orbonalis* and *T. absoluta* follow a non-secretory pathway and lack a signal peptide, though it might contain a transit peptide to reach and localize in cytoplasm. The remaining eleven species followed a secretory pathway with a

probability score of 1. Out of the eleven species following a secretory pathway, in nine species it is localized in the cell membrane the extended N-terminal may allow this protein to localize in the cell membrane of these species (Mor et al., 2008). In two species, *P. xuthus* and *P. machaon*, although it followed the secretory pathway it did not localize in the cell membrane but remained in the extracellular region. It might act as a cell membrane adhesion receptor (Lodish et al., 2000) and involve a protein interaction (Soreq and Seidman, 2001). Among the nine species, in which AChE is localized in the cell membrane and follow secretory pathway, insecticide resistance is a common and major problem. It shows

Table 9
GPI Anchor Prediction in acetylcholinesterase of lepidopteran insect pest species.

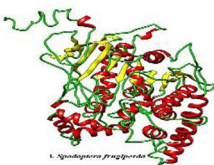| Organism	UniProt ID	GPI anchor		
		Predicted Residue	ω Position	Score
Pieris rapae	A0A1U9X1Z7	DELEHVP**C**DG	608	−1.31
		AVTGPY**S**SVA	617	−3.21
Helicoverpa assulta	Q5RLH9	ELEHMP**C**DGA	617	−0.84
		VTGPY**S**SVAG	626	−3.23
Helicoverpa armigera	G1CRK3	ELEHMP**C**DGA	617	−0.84
		VTGPY**S**SVAG	626	−3.23
Spodoptera frugiperda	A0A2H1VS32	NELEHMP**C**DG	608	−0.84
		AVTGPY**S**SVA	617	−3.23
Cydia pomonella	Q2XQ04	NELERVP**C** **D**G	608	−1.25
			609	−1.8
Scirpophaga incertulas	A0A1D8QQG3	AVTGPY**S**SVA	617	−3.91
Chilo suppressalis	B2BD82	AVTGPY**S**SVA	617	−2.46
Chilo auricilius	A0A076VJ71	AVTGPY**C** **D**VA	618	−2.19
			617	−4.6
Spodoptera litura	A0A1S5YCX9	NELEHMP**C**DG	608	−0.84
		AVTGPY**S**SVA	617	−3.23
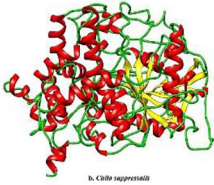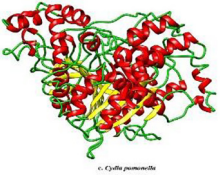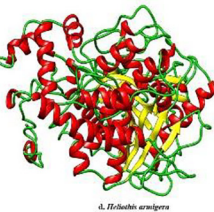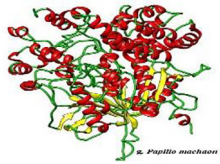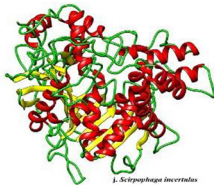
that similar localization site and mode of movement of AChE in these insects might be the reason of development of insecticide resistance.

Three posttranslational modifications i.e., palmotylation, methylation and Glycosylphosphatidylinositol (GPI) were found in AChE of fourteen lepidopteran insect pest species while lipidation was missing among all the species. The cutoff values for prediction of palmotylation (3.717), methylation (4.11) were kept the same for all the species and the results with higher scores were selected. The score shows whether the predicted amino acid position is a true PTM site for selected thresholds (Xue et al., 2008). Methylation is referred as addition of a methyl group to lysine side chain (Saraswathy and Ramalingam, 2011). Arginine (R) methylation is a predominant PTM in both nuclear and cytoplasmic proteins (Bedford, 2007). In selected insect pest species the AChE showed arginine (R bold, highlighted and red) methylation in eleven species. Palmotylation is a covalent attachment of palmitic acid to cysteine (S-palmitoylation) and rarely to threonine and serine (O-palmitoylation) in proteins (Linder, 2001). Cysteine palmitoylation was found in AChE of nine insect pest species. Five species *P. rapae*, *S. litura, S. frugiperda, C. auricilius* and *C. pomonella* showed a relatively higher score compared to remainder species. Which indicates that these five species have strong evidence for palmotylation among all the species. The species that showed palmotylation sites are involved in regulation of cellular functions such as protein attachment to the membranous portion that makes them more hydrophobic (Rocks et al., 2005). Glycosylphosphatidylinositol (GPI) is a glycolipid attached to the C terminus of newly synthesized proteins and plays an important role in biological processes (Paulick and Bertozzi, 2008). The GPI is attached through transamidation on the C terminus residue, and that position is denoted by ω, of newly synthesized protein (Fournier et al., 1988; Toutant, 1989). Big-PI Predictor predicted GPI anchors on AChE in nine insect pest species. The results presented for GPI anchor prediction are given by expected ω position with a score. The higher score is a preferred site and the amino acid colored red. An alternative option is also given with a relatively lower score and amino acid colored yellow. Insects have cysteine (C) as C-terminal peptide which forms a disulfide linked dimer and induces a signal for addition of glycophosphatidylinositol (GPI) as post translational modification (Kakani et al., 2011). It is obvious from the results that the preferred site is with amino acid residue **C**. Two species, *S. incertulas* and *C. Suppressalis,*

showed only one position, **S,** with a lower score which predicts that these might not be a potential position for GPI modification. Increasing the number of GPI-anchored molecules in the synaptic cleft reduces the sensitivity to insecticides (Kakani and Mathiopoulos, 2008) especially organophosphates (Kakani et al., 2011). We can predict that seven, *P. rapae, H. armigera, H. assulta, S. frugiperda C. pomonella, C. auricilius* and *S. litura* having two GPI anchor positions may have more ability to develop insecticide resistance. *S. incertulas* and *C. suppressalis* may be less prone to insecticide resistance and *L. orbonalis, S. exigua* and *T. absoluta* may tend not to develop insecticide resistance due to no GPI anchor modification found.

The 3D structures and suitable templates for homology modeling of AChE of selected lepidopteran species, were not available. I-TASSER server was used for 3D model prediction, as it uses multiple templates to predict protein structure (Zhang, 2008). Good quality models with high confidence score (C score) were selected from five models generated against each query sequence. C-score is calculated on the basis of the importance of threading template alignments and the convergence parameters of the structure assembly simulations. Generally the range of C-score is between −5 to 2. High confidence level on predicted structure is reflected by high value of C- score (Roy et al., 2010; Yang et al., 2015). The evaluation and refinement of selected models were performed by QMEAN and SAVES, to further improve the quality and reliability. The good quality structures were generated by using refinement methods utilizing multiple templates and by rebuilding unreliable loops using an optimization-based refinement method (Ko et al., 2012). Refinement protocol employs iterative optimization of hydrogen bonding network assisted with energy minimization at the atomic level on the optimized model by means of a composite physics and knowledge-based force fields (Bhattacharya et al., 2016). For quality estimation of models QMEAN is an online server. QMEAN score which shows the predicted global model reliability ranging from 0 to 1 is used for the ranking of the models mentioned in the results (Benkert et al., 2008). SAVES (Structure Analysis and Verification Server) consists of different programs i.e., ERRAT, Verify3D, Prove, RAMPAGE, for structural evaluation and refinement. The structures were selected which have ERRAT value above 60%. Percentage of amino acids in favored and allowed regions were shown by RAMPAGE values. The models having amino acids with more than 80% in allowed and favored regions were selected

Table 10
Predicted three dimensional structures of acetylcholinesterase in different lepidopteran insect pest species.

Insect pest	3D protein structure	Insect pest	3D protein structure
Chilo auricilius		*Papilio xuthus*	
Chilo suppressalis		*Pieris rapae*	
Cydia pomonella		*Scirpophaga incertulas*	
Heliothis armigera		*Spodoptera exigua*	
Heliothis assulta		*Spodoptera frugiperda*	
Leucinodes orbonalis		*Spodoptera litura*	
Papilio Machaon		*Tuta absoluta*	

5. Conclusion

Acetylcholinesterase is an important enzyme involved in insecticide resistance., The knowledge about the structure and function of the protein is necessary to know the exact reason of resistance, but most of the insect proteins are unreviewed. Computational annotation of acetylcholinesterase of fourteen insect pest species *Pieris rapae*, *Papilio machaon*, *Heliothis assulta*, *Heliothis armigera*, *Spodoptera frugiperda*, *Spodoptera exigua*, *Cydia pomonella*, *Leucinodes orbonalis*, *Tuta absoluta*, *Scirpophaga incertulas*, *Chilo suppressalis*, *Chilo auricilius*, *Papilio xuthus* and *Spodoptera litura* revealed that AChE in three species, *T. absoluta, S. exigua* and *L. orbonalis,* is closely related. These insect pests seem to have the least chances of insecticide resistance. We can predict that *P. rapae, H. armigera, H. assulta, S. frugiperda C. pomonella, C. auricilius* and *S. litura,* having two GPI anchor positions, may have more ability to develop insecticide resistance, moreover they have common localization for AChE i.e., in cell membrane. High percentage of Cys and

Gln amino acid residues in *P. rapae* strongly added the cause of insecticide resistance.

Funding

NA

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

NA

References

Aldrige, W.N., 1950. Some properties of specific cholinesterase with particular reference to the mechanism of inhibition by diethyl p-nitrophenyl thiophosphate (E 605) and analogues. Biochem J. 46, 451–460.

Almagro, A.J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H.W.O., 2017. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics. 33, 3387–3395.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al., 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

Bailey, T., Gribskov, M., Bailey, T.L., 1998. Methods and Statistics for Combining Motif Match Scores: The development of hydrocephalus in mice lacking the transcription factor NFIX View project Methods and statistics for combining motif match scores. Artic. J. Comput. Biol. 5, 211–221.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208.

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings /International Conference on Intelligent Systems for Molecular Biology; ISMB International Conference on Intelligent Systems for Molecular Biology.

Bedford, M.T., 2007. Arginine methylation at a glance. J. Cell Sci. 120, 4243–4246.

Benkert, P., Tosatto, S.C.E., Schomburg, D., 2008. QMEAN: a comprehensive scoring function for model quality assessment. Proteins: Struct Funct. Bioinform. 71, 261–277.

Bhattacharya, D., Nowotny, J., Cao, R., Cheng, J., 2016. 3Drefne: an interactive web server for efficient protein structure Refinement. Nucleic Acids Res. 44, W406–W409.

Bottino-Rojas, V., Talyuli, O.A.C., Carrara, L., Martins, A.J., James, A.A., Oliveira, P.L., Paiva-Silva, G.O., 2018. The redox-sensing gene Nrf2 affects intestinal homeostasis, insecticide resistance, and Zika virus susceptibility in the mosquito *Aedes aegypti*. J. Biol. Chem. 293, 9053–9063.

Cao, R., Cheng, J., 2016. Integrated protein function prediction by mining function associations, sequences, and protein protein and gene-gene interaction networks. Methods 93, 84–91.

Cao, R., Freitas, C., Chan, L., Sun, M., Jiang, H., Chen, Z., 2017. ProLanGO: Protein Function Prediction Using Neural Machine Translation Based on a Recurrent Neural Network. Molecules 22, 1732–1746.

Casey, P.J., Seabra, M.C., 1996. Protein prenyltransferases. J. Biol. Chem. 271, 5289–5292.

Chaabihi, H., Fournier, D., Fedon, Y., Bossy, J.P., Ravellec, M., Devauchelle, G., Cérutti, M., 1994. Biochemical characterization of *Drosophila melanogaster* acetylcholinesterase expressed by recombinant baculoviruses. Biochem. Biophys. Res. Commun. 203, 734–742.

Clark, W.T., Radivojac, P., 2011. Analysis of protein function and its prediction from amino acid sequence. Proteins Struct. Funct. Bioinforma. 79, 2086–2096.

Corbett, J., Wright, K., Baillie, A.C., 1974. Insecticides inhibiting acetylcholinesterase. Biochem Mode Action Pesticides. Academic Press, NewYork, pp. 107–163.

Cruz, L.M., Trefflich, S., Weiss, V.A., Castro, M.A.A., 2017. Protein Function Prediction. In: Methods in Molecular Biology. Humana Press Inc., pp. 55–75.

Das, S., Orengo, C.A., 2016. Protein function annotation using protein domain family resources. Methods 93, 24–34.

Feyereisen, R., 1995. Molecular biology of insecticide resistance. Toxicol Lett. 82, 83–90.

Filipski, A., Murillo, O., Freydenzon, A., Tamura, K., Kumar, S., 2014. Prospects for building large timetrees using molecular data with incomplete gene coverage among species. Mol. Biol. Evol. 31, 2542–2550.

Firake, D.M., Lytan, D., Behere, G.T., 2012. Bio-diversity and Seasonal Activity of Arthropod Fauna in Brassicaceous Crop Ecosystems of Meghalaya. Mol. Entomol. 3, 18–22.

Fournier, D., Bergé, J.B., de Almeida, M.C., Bordier, C., 1988. Acetylcholinesterase from *Musca domestica* and *Drosophila melanogaster* Brain are linked to membranes by a Glycophospholipid anchor sensitive to an endogenous Phospholipase. J. Neurochem 50, 1158–1163.

Fournier, D., Mutero, A., 1994. Modification of acetylcholinesterase as a mechanism of resistance to insecticides. Comp Biochem. Physiol. Part C Pharmacol. 108, 19–31.

Frasca, M., Cesa-Bianchi, N., 2017. Multitask protein function prediction through task dissimilarity. IEEE/ACM Trans Comput. Biol. Bioinforma. 16, 1550–1560.

Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., 2005. DNA repair and mutagenesis. American Society for Microbiology Press.

Frith, M.C., Saunders, N.F.W., Kobe, B., Bailey, T.L., 2008. Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput. Biol. 4, (5) e1000071.

Fu, J., Tang, J., Wang, Y., Cui, X., Yang, Q., Hong, J., Li, X., Li, S., Chen, Y., Xue, W., Zhu, F., 2018. Discovery of the consistently well-performed analysis chain for swath-ms based pharmacoproteomic quantification. Front. Pharmacol. 9, 681.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. Humana Press, 571–607.

Goldstrohm, A.C., Hall, T.M.T., McKenney, K.M., 2018. Post-transcriptional regulatory functions of mammalian pumilio proteins. Trends Genet. 34, 972–990.

Guedes, R.N.C., Kambhampati, S., Dover, B.A., Zhu, K.Y., 1997. Biochemical mechanisms of organophosphate resistance in *Rhyzopertha dominica* (Coleoptera: Bostrichidae) populations from the United States and Brazil. Bull. Entomol. Res. 38687, 581–586.

Hiller, K., Grote, A., Scheer, M., Münch, R., Jahn, D., 2004. PrediSi: Prediction of signal peptides and their cleavage positions. Nucleic Acids Res. 32, 375–437.

Horiike, T., 2016. Invited Mini-Review an Introduction To Molecular. Rev. Agric. Sci. 4, 36–45.

Hundal, H.S., Rennie, M.J., Watt, P.W., 1989. Characteristics of acidic, basic and neutral amino acid transport in theperfused rat hindlimb. J. Physiol. 408, 93–114.

Hung, M.C., Link, W., 2011. Protein localization in disease and therapy. J. Cell Sci. 124, 3381–3392.

Ikai, A., 1980. Thermostability and Aliphatic Index of Globular Proteins. J. Biochem. 88, 1895–1898.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: Protein structure and function prediction. Nature Methods 12, 7–8.

Jain, A., Kihara, D., 2019. Phylo-PFP: improved automated protein function prediction using phylogenetic distance of distantly related sequences. Bioinformatics. 35 (5), 753–759.

Kakani, E.G., Bon, S., Massoulié, J., Mathiopoulos, K.D., 2011. Altered GPI modification of insect AChE improves tolerance to organophosphate insecticides. Insect Biochem. Mol. Biol. 41, 150–158.

Kakani, E.G., Mathiopoulos, K.D., 2008. Organophosphosphate resistance-related mutations in the acetylcholinesterase gene of Tephritidae. J. Appl. Entomol. 132, 762–771.

Keane, S., Ryan, M.F., 1999. Purification, characterisation, and inhibition by monoterpenes of acetylcholinesterase from the waxmoth, *Galleria mellonella* (L.). Insect Biochem. Mol. Biol. 29, 1097–1104.

Ko, J., Park, H., Heo, L., Seok, C., 2012. GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res. 40, 294–297.

Krejci, E., Duval, N., Chatonnett, A., Vincensf, P., Massouliv, J., 1991. Cholinesterase-like domains in enzymes and structural proteins: Functional and evolutionary relationships and identification of a catalytically essential aspartic acid. Proc. Natl. Acad. Sci. 88, 6647–6651.

Kulye, M., Gondhalekar, A.D., Chaudhari, C.S., Chandele, C.A., 2007. Comparative toxicity of some benzoylphenyl urea moltinhibiting insecticides against diamondback moth (*Plutella xylostella*). Pestology 31, 45–46.

Kumar, S., Nei, M., Dudley, J., Tamura, K., 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. 9, 299–306.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.

Kyte, J., Doolittle, R.F., 1983. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157 (414), 105–132.

Li, Y.H., Yu, C.Y., Li, X.X., Zhang, P., Tang, J., Yang, Q., Fu, T., Zhang, X., Cui, X., Tu, G., et al., 2018. Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. Nucleic Acids Res. 46, D1121–D1127.

Linder, M.E., 2001. Reversible modification of proteins with thioester-linked fatty acids. Enzym. 21, 215–240.

Lodish, H., Berk, A. and Zipursky, S.L., et al., 2000. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; Section 22.4, Noncollagen Components of the Extracellular Matrix..

Meshorer, E., Erb, C., Gazit, R., Pavlovsky, L., Kaufer, D., Friedman, A., Glick, D., Ben-Arie, N., Soreq, H., 2002. Alternative splicing and neuritic mRNA translocation under long-term neuronal hypersensitivity. Science. 295, 508–512.

Mor, I., Bruck, T., Greenberg, D., Berson, A., Schreiber, L., Grisaru, D., Soreq, H., 2008. Alternate AChE-R variants facilitate cellular metabolic activity and resistance to genotoxic stress through enolase and RACK1 interactions. Chemico-biological interactions 175 (1–3), 11–21.

Mutero, A., Pralavorio, M., Bride, J.M., Fournier, D., 1994. Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase. Proc. Natl. Acad. Sci. USA 91, 5922–5926.

Nickel, W., Seedorf, M., 2008. Unconventional Mechanisms of Protein Transport to the Cell Surface of Eukaryotic Cells. Annu. Rev. Cell Dev. Biol. 24, 287–308.

Nielsen, H., Engelbrecht, J., Brunak, S., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites; Artificial neural networks have been used for many biological. Prot. Engin. 10, 1–6.

Paulick, M.G., Bertozzi, C.R., 2008. The glycosylphosphatidylinositol anchor: A complex membrane-anchoring structure for proteins. Biochem. 47, 6991–7000.

Rajashekar, Y., Raghavendra, A., Bakthavatsalam, N., 2014. Acetylcholinesterase inhibition by biofumigant Coumaran) from leaves of Lantana camara in stored grain and household insect pests. BioMed Research. International.

Rapoport, T.A., 1992. Transport of proteins across the endoplasmic reticulum membrane. Science 258, 931–936.

Rapoport, T.A., 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature 450:663–669-7000.

Rocks, O., Peyker, A., Kahms, M., Verveer, P.J., Koerner, C., Lumbierres, M., Kuhlmann, J., Waldmann, H., Wittinghofer, A., Bastiaens, P.I.H., 2005. An acylation cycle regulates localization and activity of palmitoylated Ras isoforms. Science 307, 1746–1752.

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738.

Roy, S., Maheshwari, N., Chauhan, R., Sen, N.K., Sharma, A., 2011. Structure prediction and functional characterization of secondary metabolite proteins of Ocimum. Bioinformation 6, 315–319.

Roy, S.S., Dasgupta, R., Bagchi, A., 2014. A Review on Phylogenetic Analysis : A Journey through Modern Era. Comput. Mol. Biosci. 4, 39–45.

Saraswathy, N., Ramalingam, P., 2011. Concepts and techniques in genomics and proteomics. Elsevier.

Schneider, T.D., Stephens, R.M., 1990. Sequence logos: A new way to display consensus sequences. Nucleic Acids, Res 18 (20), 6097–6100.

Scoble, M.J., 1992. The Lepidoptera. Form, function and diversity. Lepid. Form, Funct. Divers. Natural History Museum Publications, Oxford University Press London, UK.

Shimomura, M., Okuda, H., Matsuda, K., Komai, K., Akamatsu, M., Sattelle, D.B., 2002. Effects of mutations of a glutamine residue in loop D of the α7 nicotinic acetylcholine receptor on agonist profiles for neonicotinoid insecticides and related ligands. Br. J. Pharmacol. 137, 162–169.

Smissaert, H.R., 1964. Cholinesterase inhibition in spider mites susceptible and resistant to organophosphate. Science 143, 129–131.

Soreq, H., Seidman, S., 2001. Acetylcholinesterase – new roles for an old actor. Nat. Rev. Neurosci. 2, 294–302.

Tang, J., Fu, J., Wang, Y., Li, B., Li, Y., Yang, Q., Cui, X., Hong, J., Li, X., Chen, Y., et al., 2019. ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief.* Bioinform. bby127.

Thomsen, M.C.F., Nielsen, M., 2012. Seq2Logo: A method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res. 40, W281–W287.

Tong, F., Coats, J.R., 2012. Quantitative structure–activity relationships of monoterpenoid binding activities to the housefly GABA receptor. Pest Manage. Sci. 68, 1122–1129.

Toutant, J.P., 1989. Insect acetylcholinesterase: Catalytic properties, tissue distribution and molecular forms. Prog. Neurobiol. 32, 423–446.

UniProt Consortium, 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 46. 2699-2699.

Voet, D., Voet, J.G., Pratt, C.W., 2006. Fundamentals of biochemistry: life at the molecular level. Wiley p1264, Hoboken.

von Heijne, G., 1985. Signal sequences. The limits of variation. J. Mol. Biol. 184, 99–105.

Voss, M., Schröder, B., Fluhrer, R., 2013. Mechanism, specificity, and physiology of signal peptide peptidase (SPP) and SPP-like proteases. Biochimica Et Biophysica Acta (BBA)-Biomembranes 1828, 2828–2839.

Wahlberg, N., Wheat, C.W., Peña, C., 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLoS One 8, e80875.

Wan, S., Duan, Y., Zou, Q., 2017. HPSLPred: An Ensemble Multi-Label Classifier for Human Protein Subcellular Location Prediction with Imbalanced Source. Proteomics 17, 17–18.

Wan, S., Mak, M.W., 2015. Machine Learning for Protein Subcellular Localization Prediction. Walter de Gruyter GmbH & Co KG.

Xiong, J., 2006. Essential bioinformatics. Cambridge University Press.

Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., Yao, X., 2008. GPS 2.0, a Tool to Predict Kinase-specific Phosphorylation Sites in Hierarchy. Mol. Cell Proteom. 7 (9):1598–1608.

Yoon, K.S., Kwon, D.H., Strycharz, J.P., Hollingsworth, C.S., Lee, S.H., Clark, J.M., 2008. Biochemical and Molecular Analysis of Deltamethrin Resistance in the Common Bed Bug (Hemiptera: Cimicidae). J. Med. Entomol. 45, 1092–1101.

You, R., Zhang, Z., Xiong, Y., Sun, F., Mamitsuka, H., Zhu, S., 2018. GOLabeler: improving sequence-based large-scale protein function prediction by learning to rank. Bioinformatics. 34, 2465–2473.

Yousafi, Q., Kanwal, S., Rashid, H., Khan, M.S., Saleem, S., Aslam, M., 2019. *In silico* structural and functional characterization and hylogenetic study of alkaline phosphatase in bacterium, *Rhizobium leguminosarum* (Frank 1879). Comput. Biol. Chem. 83, 107142.

Zhang, C., Freddolino, P.L., Zhang, Y., 2017. COFACTOR: Improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Res. 45, W291–W299.

Zhang, D.D., Hannink, M., 2003. Distinct Cysteine Residues in Keap1 are required for Keap1-Dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. Mol. Cell Biol. 23, 8137–8151.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9 (1), 40.

Zhao, B., Hu, S., Li, X., Zhang, F., Tian, Q., Ni, W., 2016. An efficient method for protein function annotation based on multilayer protein networks. Hum. Genomics. 10, 33.

Zhu, F., Li, X.X., Yang, S.Y., Chen, Y.Z., 2018. Clinical success of drug targets prospectively predicted by in silico study. Trends Pharmacol. Sci. 39, 229–231.

Zuckerkandl, E., Pauling, L., 1965. Evolutionary Divergence and Convergence in Proteins. Evol. Genes. Proteins. 97, 166.