Structural, functional, and phylogenetic studies of chymotrypsin enzyme genes in insects: a bioinformatics approach

S. A. Hemmati

Corresponding Author: Assistant professor, Department of Plant Protection, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran (sa.hemmati@scu.ac.ir)

(DOI): 10.22055/PPR.2020.16417

Received: 30 November 2020 Accepted: 12 January 2021

Abstract

Chymotrypsin-like enzymes are involved in many physiological processes of insects including digestion, development, survival, and immunity. The enzyme cleaves the peptide bonds on C-terminal of the aromatic amino acids in proteins and release active peptides as well as amino acids required for growth, reproduction, and development of the insect. Due to the importance of food proteins on insect survival and development, more fundamental entomologists are researching the structures, functions, mechanisms, and interactions of the involved digestive enzymes. Discovering biochemical and structural properties of insect proteases can help develop rational control strategy based on specific protease inhibitors. Accordingly, the present study focuses on protein structural analysis, protein sequence alignment, phylogenetic analysis as well as conserved motif assessment from various insect species using different bioinformatics tools. The multiple sequence alignment revealed different conserved stretches of amino acids along with highly conserved catalytic sites (His, Asp, Ser) and ten conserved motifs were also discovered by MEME and MAST tools. The phylogenetic data suggest that the insect chymotrypsins might share a common ancestor. The three-dimensional structures of chymotrypsins were generated by I-TASSER, and the 3D model was further verified using PROCHECK, ERRAT, and Verify-3D. The protein – protein interactions network constructed by STRING 11 provided ten enriched pathways in *Aedes aegypti* (L.) chymotrypsin. Totally, the present work can provide new insights for designing alternative pesticides based on specific digestive enzyme inhibitors.

Keywords: Insect chymotrypsin; Peptide Motif; Phylogenetic study; Proteinase Inhibitors; Pests Control

Associate editor: M. Mehrabadi (Ph.D)

Introduction

The damage caused by herbivorous insects reduces crop production around the world (van der Meijden, 2015). Ideally, insecticides are lethal to the targeted pests, but they can also cause damage to non-target species, including human, as well as to the environment (Kalia and Gosal, 2011). Currently, one promising pest management strategy involves the use of digestive enzyme (mainly trypsin and chymotrypsin) inhibitors (Hemmati and Mehrabadi, 2020). Proteolysis plays critical roles in insect physiology including development and immune defense as well as in food digestion processes. Proteases hydrolyze the peptide bonds in the proteins and release active peptides plus amino acids required for the insects' growth, reproduction, and development (Yang et al., 2017a).

Serine proteases (SPs) are a dominant class of active proteases in the digestive lumen of insects, which are involved in various physiological processes such as digestion, development, and immunity (Yang et al., 2017b). Chymotrypsin (EC 3.4.21.1) and trypsin (EC 3.4.21.4) are the most important digestive SPs of insects specifically Lepidoptera, which are among the most serious pests of most crops (Dunse et al., 2010). The lepidopteran larvae need the proteolytic enzymes for digestion of proteins, and many serine proteases are frequently present in the larval midgut (Mahdavi et al., 2013). Trypsin cleaves peptides on the carboxyl side of lysine and arginine amino acids, but the aromatic residues are the preferential cleavage site for chymotrypsins (Lazarević and Janković-Tomanić, 2015). Based on previous structural analysis, chymotrypsin and trypsin enzymes showed some similarities in their tertiary structures, but their substrates, as well as substrate-binding sites are specific (Haginaka, 2012). The catalytic triad of serine proteinases involves serine, histidine, and aspartic acid residues (Hemmati and Mehrabadi, 2020).

Chymotrypsin preferentially cleaves the amino acid chain selectively on the C terminus of the amino acids tryptophan, tyrosine, phenylalanine, and methionine. Insect chymotrypsins have usually molecular weights within the range of 20 to 31kDa and optimal pH within the range of 8 to 11 (Terra, 2005). Chymotrypsin is initially synthesized as inactive precursors containing 245 amino acids (as chymotrypsinogen). Then, activation of the precursor to chymotrypsin is triggered by the proteolytic cleavage at two sites (Parekh et al., 2011).

Due to two main reasons, i) the SPs roles in the digestive system of insects as well as their development and survival, and ii) significant variations across the biochemical characteristics of insects' digestive serine proteases, the characterization of SPs is required for designing new safe pesticides based on proteinaceous plant protease inhibitors. Accordingly, in the present study, we have characterized the structural and biochemical properties of chymotrypsin enzyme in different insect pests using various bioinformatics analyses. The results would provide new insights for designing and developing new alternative pesticides based on specific digestive enzyme inhibitors.

Materials and Methods

Data collection

A total of 21 protein sequences for chymotrypsins from different insect species available in GenBank were obtained by NCBI (http://www.ncbi.nlm.nih.gov/) and the Universal Protein resource (Uniprot) website (www.uniprot.org/).

Identification of conserved peptide motifs

Chymotrypsin sequence motifs were generated by using MEME (http://meme-

suite.org/). The parameters were as follows; number of repetitions was 0 or 1 per sequence, the maximum width for each motif was 50, and maximum number of motifs was 10.

Phylogenetic study

The neighbor-joining method was utilized to construct a phylogenetic tree from the aligned sequences of insect chymotrypsin amino acids by MEGA 5.10 (https://www.megasoftware.net) (Kumar et al., 2016) with 1000 bootstrap replicates. Initial tree(s) for the heuristic search were generated automatically by using Neighbor-Join (NJ) and BioNJ algorithms to a matrix of pairwise distances estimated applying a JTT model, and then selecting the topology with a superior log likelihood value (Jones et al., 1992).

***In silico* analysis of chymotrypsin: primary and secondary structures**

Expasy ProtParam server was used for physicochemical characterization (molecular weight, theoretical pI (isoelectric point), instability index, and aliphatic index) of the protein (https://web.expasy.org/protparam/) (Gasteiger et al., 2005). A multiple sequence alignment program, MAFFT (Multiple Alignment using Fast Fourier Transform), was used to discover LINSI as the most accurate MSA (Multiple protein Sequence Alignment) method. SOPMA (the improved self-optimized prediction method) was further used for the prediction of the secondary structure of the protein (http://npsapbil.ibcp.fr/cgibin/npsa_automat.pl?page=npsa_sopma.html) (Geourjon and Deleage, 1995). Moreover, analysis of transmembrane motifs was done through the Transmembrane Helices; Hidden Markov Model (TMHMM) Server (http://www.cbs.dtu.dk/services/TMHMM) based on the default settings of the server.

Homology modeling and identification of amino acid sequences

The chymotrypsin sequences of insects including *Spodoptera exigua* (Hübner), *Tenebrio molitor* L., *Lygus hesperus* (Knight), *Periplaneta americana* (L.) and *Aedes aegypti* (L.) were aligned with the enzyme from other species including *Mus musculus* (L.). The multiple sequence alignments were performed with CLUSTALW (Thompson et al., 1994) and Easy Sequencing in PostScript (ESPript), version 2.2 (Gouet et al., 1999).

Furthermore, the structural model of chymotrypsin were generated by Iterative Threading ASSEmbly Refinement (I-TASSER) (Yang et al., 2015). The best models were selected according to the I-TASSER parameters including C-score, TM-score, and RMSD. C-score is a confidence score for the quality of predicted models, and RMSD and TM-score are known standards for comparing the structural similarity between structures. Moreover, the validity of the predicted models was checked by using several structural analysis programs including PROCHECK, ERRAT, and Verify-3D (https://servicesn.mbi.ucla.edu/SAVES/). ERRAT validates the model by statistical analysis of non-bonded interactions between different atom types depending on their characteristic atomic interaction. Verify_3D assesses the compatibility of the predicated 3D model based on the local environment of each amino acid, and PROCHECK evaluates the overall stereo-chemical quality of the protein.

STRING: functional protein association networks

The search tool for retrieval of interacting Genes/Proteins (STRING) (https://string-db.org/) was applied to construct the chymotrypsin interactions network by considering individual proteins as nodes, and devoting an edge among proteins if they have

some interactions. The interactions contain direct and indirect interactions, providing from various experimental or computational methods.

Result and Discussion

To date, 1103 sequences of insect chymotrypsins have been registered in the GenBank (http://www.ncbi.nlm.nih.gov/). The identified sequences generally contained a signal peptide, three pairs of conserved cysteine residue, a catalytic triad of His 57, Asp 102, and Ser 195 residues, and a conserved N-terminal sequence IVGG and Ser/Gly/Tyr 189, which confers specificity to chymotrypsin-like enzymes (Aluko, 2015). To perform homology modeling, multiple sequence alignment, phylogenetic and motif analysis, the insect chymotrypsin protein sequences were retrieved from different databases (Table 1).

Table 1. Chymotrypsins originated from various insects and mouse as a mammalian in this study

No.	Scientific name	Latin name	Family	Accession no.	No. sequence
1	*Periplaneta Americana*	American cockroach	Blattidae	AIA09349.1	251
2	*Anoplophora glabripennis*	Asian longhorn beetle	Cerambycidae	JAB63100.1	295
3	*Hermetia illucens*	Black soldier fly	Stratiomyidae	ADP88998.1	267
4	*Helicoverpa armigera*	Cotton bollworm	Noctuidae	ADI32882.1	295
5	*Phlebotomus papatasi*	Sand fly	Psychodidae	AAM96939.1	260
6	*Lucilia cuprina*	Green bottle fly	Calliphoridae	AAA68986.1	255
7	*Lygus Hesperus*	Western plant bug	Miridae	JAG42770.1	281
8	*Harpegnathos saltator*	Jerdon's jumping ant	Formicidae	EFN84779.1	306
9	*Papilio xuthus*	Asian swallowtail butterfly	Papilionidae	KPJ01456.1	232
10	*Operophtera brumata*	winter moth	Geometridae	KOB58154.1	311
11	*Mythimna separate*	Oriental armyworm	Noctuidae	AKR06192.1	296
12	*Heliothis virescens*	Tobacco budworm moth	Noctuidae	AFM28261.1	307
13	*Danaus plexippus*	monarch butterfly	Nymphalidae	OWR53227.1	276
14	*Phaedon cochleariae*	Mustard beetle	Chrysomelidae	CAA76928.1	276
15	*Teleogryllus emma*	Emma field cricket	Gryllidae	ABV32556.1	290
16	*Anopheles gambiae*	African malaria mosquito	Culicidae	CAA79325.1	259
17	*Spodoptera exigua*	Beet armyworm	Noctuidae	AAX35812.1	281
18	*Ostrinia nubilalis*	European corn borer	Crambidae	AFM77767.1	298
19	*Aedes aegypti*	Yellow fever mosquito	Culicidae	Q8T4T5	268
20	*Tenebrio molitor*	*Yellow mealworm beetle*	Tenebrionidae	ABC88746.1	275
21	*Mus musculus*	House mouse	Muridae	AAI15518.1	268

Identification of novel conserved functional motifs

MEME and MAST programs were applied to search conserved peptide motifs among the insect chymotrypsin proteins. A conserved motif is an amino acid sequence pattern, which is repeatedly present among a group of related proteins. Functional or structural motifs that are inherent to a protein act as markers to trace the evolution of proteins and the inter-relationships between their various specificities (Krem and Di Cera, 2001). MEME detects motifs using position-dependent letter-probability matrices, which specify the possibility of each probable letter at each position in the pattern. Here, ten conserved peptide motifs of chymotrypsin were identified among the different insect species by MEME and MAST (Figure 1 and Table 2). Furthermore, the findings revealed that the conserved motifs were common in the studied insects (Figure 2).

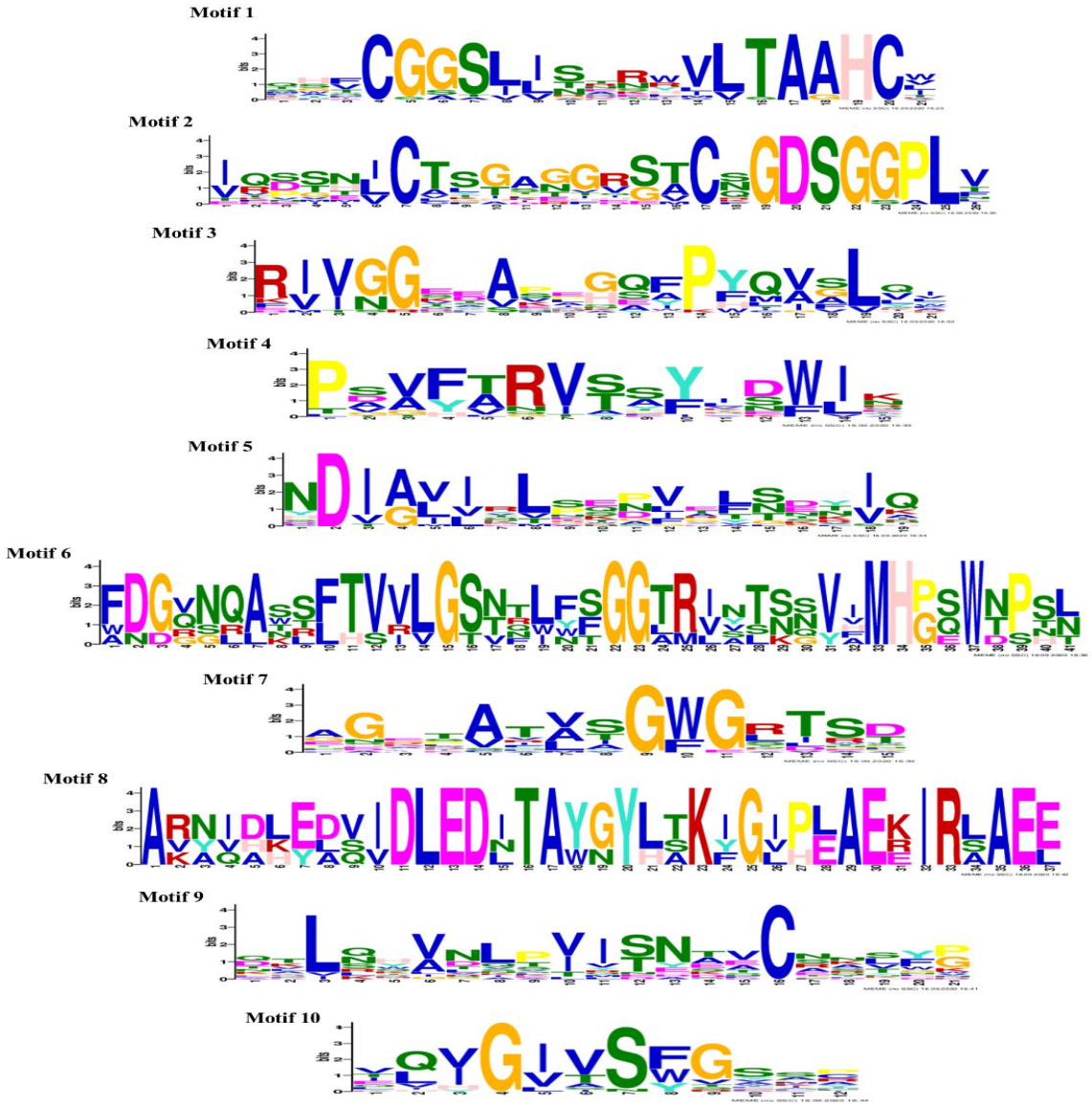

Figure 1. Sequence-specific MEME conserved peptide motifs obtained from various chymotrypsins. Motif sequences are color-coded through the MEME default (Red – positively charged (Lys, Arg); pink – negatively charged (Asp, Glu); blue – hydrophobic (Ala, Val, Leu, Ile, Phe, Met), green – polar (Gln, Ser, Asn, Thr), orange (Gly), yellow (Pro), cyan (Tyr) and peach (His). The heights of the amino acids related to their conservation at that position.

Table 2. The chymotrypsin conserved peptide motifs generated by the MEME web server in the various insects

Motif	E-value[a]	Sites	Width	Sequence
1)	4.6e-182	21	21	[QTSM][HSGW][FVT]CGGSLIS[TRNEA][RN][WR]VLTAAHC[WVLI]
2)	5.9e-109	14	26	[IV][QR][SDP][STG][NH][ILV]CTSGA[GN]G[RV][SG][TA]C[SNQ]GDSGGPLV
3)	1.6e-104	20	21	R[IV]VGG[ESK][ED]A[PSVK]E[GH]Q[FA]P[YFHW]Q[VA][SGA]L[QV][IVL]
4)	3.6e-088	20	15	P[SDAV][VA]F[TA]RV[ST][SA][YF][IL][DSN]WI[KN]
5)	2.3e-067	20	19	NDI[AG][VL]IRL[SP][EQ][PND]V[ETA][LF][SNT][DEN][YTN][IV]Q
6)	2.1e-047	6	41	FDG[VQRG]NQA[SWTNK]SFTVVLGS[NT][TRQN]LF[SF]GGTR[IV]NT[SNK][SNQG]V[IVHFA]MH[PG][SQ]W[NT]PS[LN]
7)	1.1e-044	21	15	AGE[TS]AT[VAL][ST]GWGRTS[DT]
8)	4.9e-037	18	21	[QT][TKV]L[QN][HYW][VA]NL[PT][VI]I[ST]N[TAE][VET]C[NSRA][NKA][SLVT][YF][PG]
9)	2.1e-031	4	37	ARN[IVQ][DHA][LKH]EDVIDLED[INL]TAYGYL[TSA]K[IYF]G[IVL]P[LE]AE[KRE]IR[LSA]AEE
10)	7.8e-031	20	12	[VT]QYG[IV]VS[FW]GS[SA][PCRA]

[a]The statistical significance of the motif. MEME usually discovers the most statistically significant (low E-value) motifs first. The E-value of a motif is relying on its log likelihood ratio, width, sites, the background letter frequencies, and the size of the training set (e.g. to demonstrate variable positions in motifs, the symbol of [XY] specifies that either X or Y may be present).

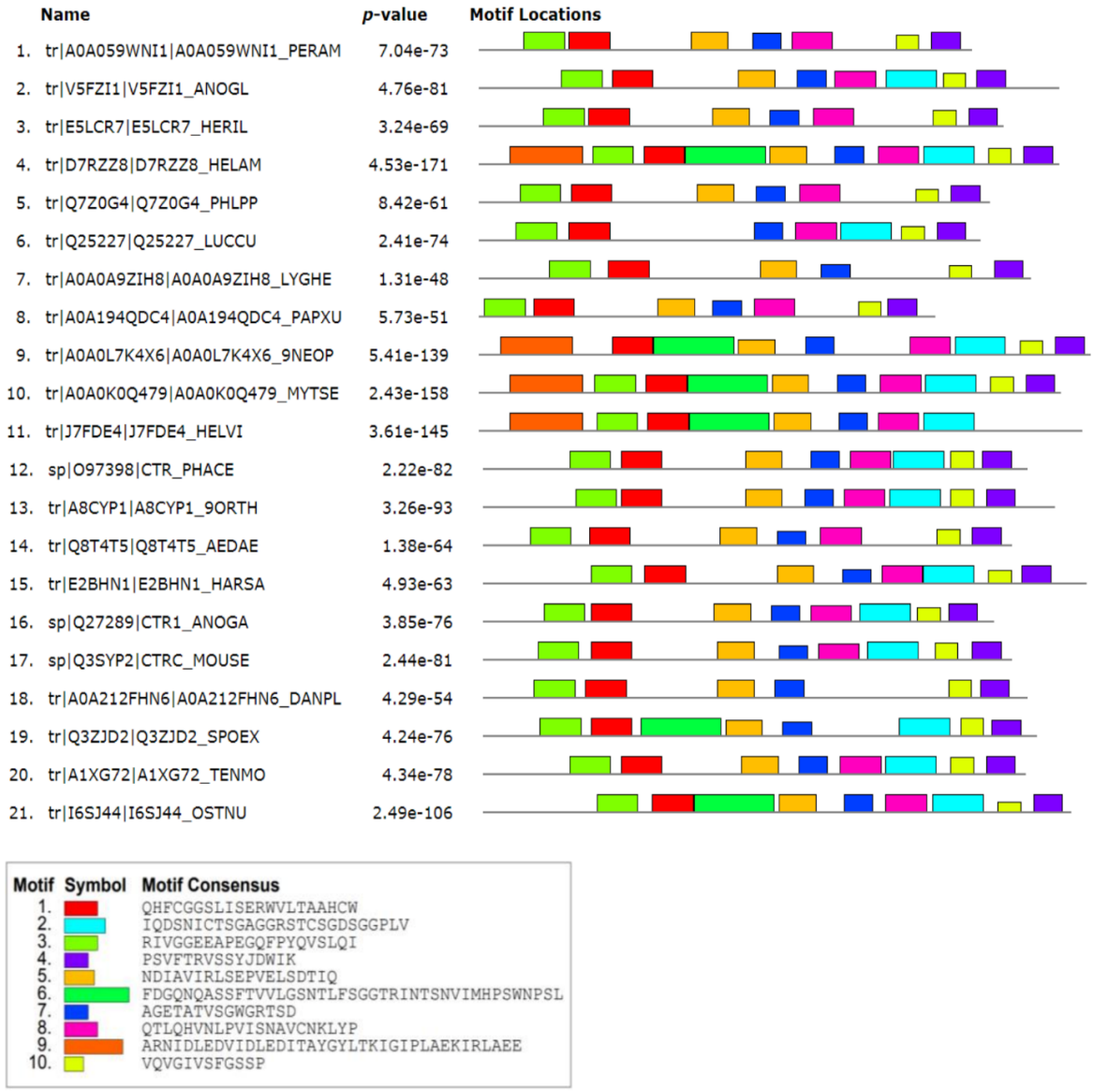

Figure 2. The conserved peptide motifs for chymotrypsin revealed via the MEME web server. The motifs are shown as various-colored boxes.

Phylogenetic tree construction

The evolutionary relationship (phylogeny) between chymotrypsin of insects and the enzyme amino acid sequence in animals (mouse) is shown in Figure 3. Phylogenetic relationships can provide important information regarding biological diversity, genetic or structural classifications, and also understanding the events that occurred during evolution (Baum, 2008). The phylogenetic tree constructed using NJ method showed two main evolutionary groups and further sub-clusters were observed in the first group based on the closeness of chymotrypsin protein sequences among insects (Figure 3). The first group comprises fifteenth insect orders and the second group related to *M. musculus* as a mammalian. Chymotrypsin of *S. exigua*, *T. molitor*, *L. hesperus*, *P. americana* and *A. aegypti* from the first group, were chosen and compared with the enzyme from *M. musculus* in details.

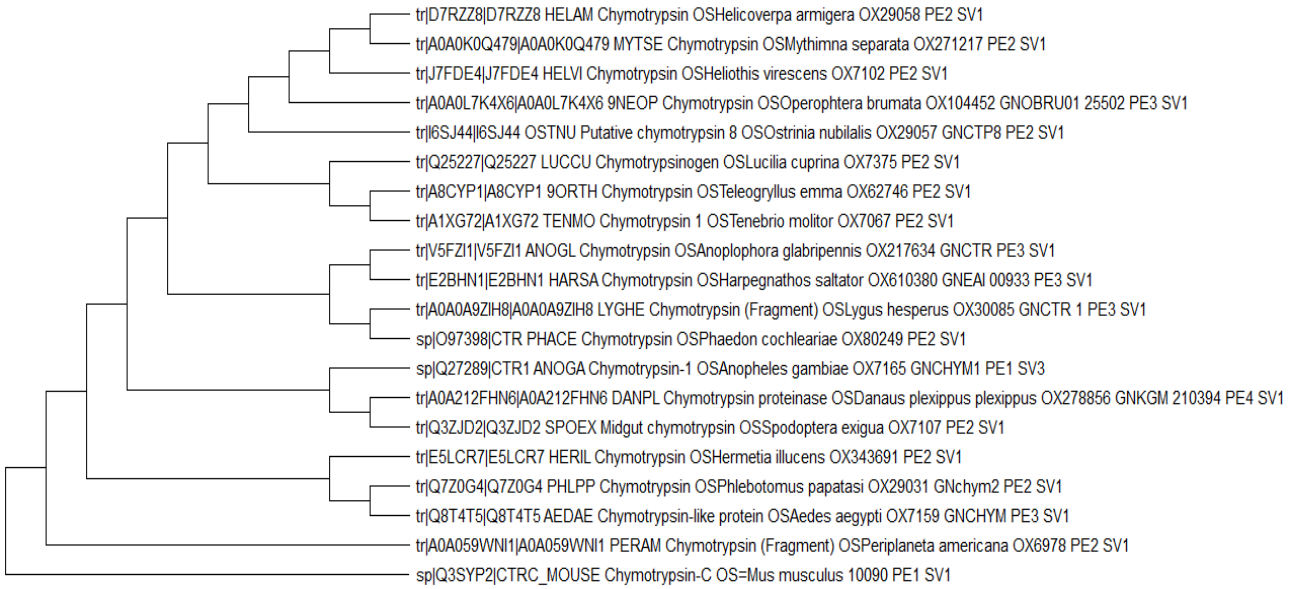

Figure 3. Phylogenetic tree study of insect chymotrypsin sequences using Maximum Likelihood technique based on the JTT matrix-based model. The Neighbor Joining (NJ) method and BioNJ algorithms was utilized to constructing the phylogenetic tree. Evolutionary analyses were performed in MEGA-X.

Chymotrypsin's primary and secondary structure

The structural analysis of insect chymotrypsins of *S. exigua*, *T. molitor*, *L. hesperus*, *P. americana* and *A. aegypti* sequences were performed in comparison with chymotrypsin from *M. musculus*. Our analysis of the amino acid sequences of different insect chymostrypsins by ProtParam program indicated that the enzyme composed of a sequence ranged from 232 to 311 amino acids and an average of 278.9, depending on the species. Chymostrypsins varied in molecular weights; usually from about 20–40 kDa (Table 3). For example, chymotrypsin from the midgut of *Helicoverpa armigera* (*Hübner*) (Lepidoptera: Noctuidae) with a molecular weight of 23.7 and 24.6 kDa (Bown et al., 1997), a 23.7 kDa chymotrypsin from the midgut of *Rhyzopertha dominica* (F.) (Coleoptera: Bostrichidae) (Zhu and Baker, 2000) and a 23 kDa chymotrypsin from *T. molitor* (Elpidina et al., 2005) were identified. The differences between chymotrypsin molecular masses may be as a result of the insect divergence. Moreover, the isoelectric point of Chymostrypsins ranges from 4.5 to 9.5 (Table 3). The pI is the pH at which an enzyme's overall electric charge is zero or is electrically neutral. An enzymes' isoelectric point value is utilized for electro focusing (Avwioroko et al., 2018).

In addition, the secondary structures were predicted by the SOPMA. It predicts four conformational states including helix, β-sheet and bridges, turns and coils for each structure. The secondary structure analysis of chymotrypsins using SOPMA indicated that the enzymes contain random coil (ranged from 43% to 53%) and β-sheet (ranged from 23% to 31%) as the predominant components among the secondary structure elements, followed by α-helix (ranged from 12% to 23%) and β-turns (ranged from 3% to 9%) (Figure 4, Table 3). Furthermore, *in silico* determinations of transmembrane helices of the proteins were

performed via the TMHMM server and the results revealed that there was no any transmembrane helix in the chymotrypsin structures (Table 3).

According to the Uniprot database, chymotrypsin has active binding sites in positions 21 – 311. There is a catalytic triad charge-relay system in chymotrypsins as commonly found in proteases.

It consists of three amino acids in the active site including aspartic acid, histidine, and serine. Serine and histidine residues are well conserved catalytic triad sequences in the serine proteases (Polgár, 2005).

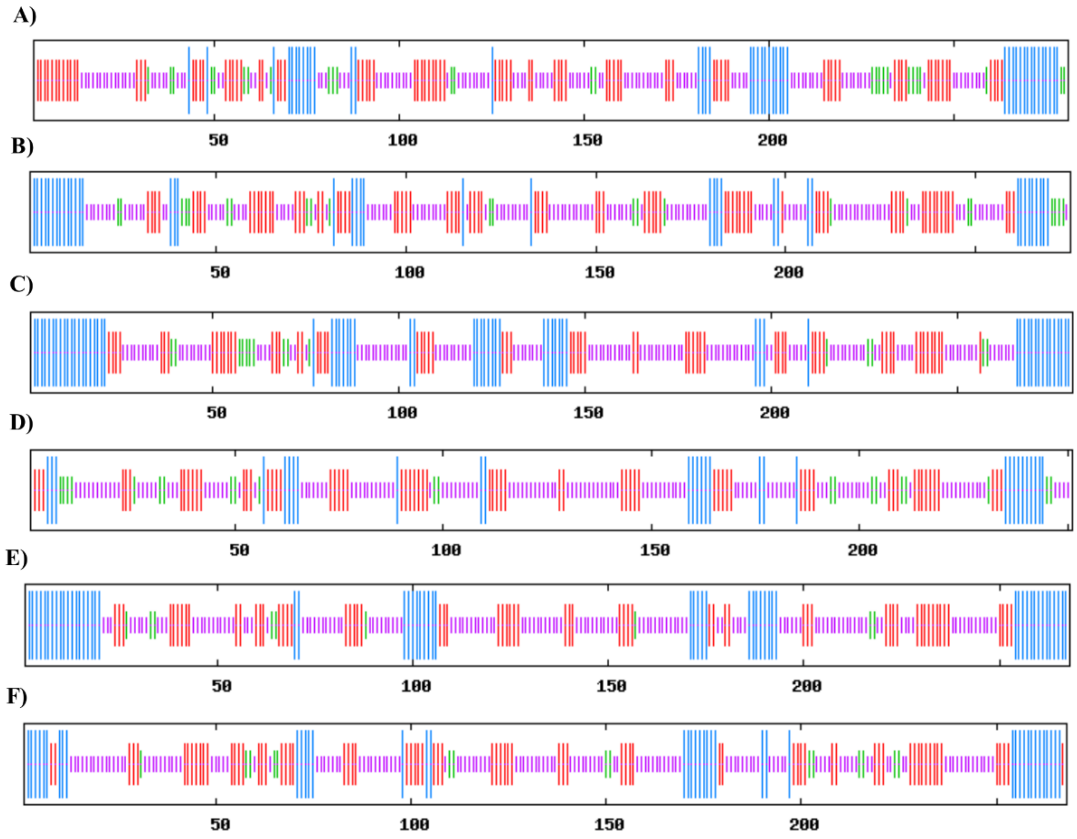

Figure 4. Comparison of the secondary structure of chymotrypsin from (A) *Spodoptera exigua*, (B) *Tenebrio molitor*, (C) *Lygus hesperus*, (D) *Periplaneta americana*, (E) *Aedes aegypti* and (F) *Mus musculus*. Amino acids were recognized by colors presented on the default of SOPMA (red: sheet, blue: alpha helix, purple: coil, green: turn)

Table 3. The results of primary and secondary structure analysis of chymotrypsin amino acid sequences among insects and mouse as a mammalian through different tools

Tools	Parameters	*Spodoptera exigua*	*Tenebrio molitor*	*Lygus hesperus*	*Periplaneta americana*	*Aedes aegypti*	*Mus musculus*
ProtParam	Number of amino acids (aa)	281	275	281	251	268	268
	Molecular weight (Mw) (Da)	31329.59	27617.07	31471.17	26179.41	29017.61	29469.57
	Theoretical isoelectric point (pI)	5.21	8.85	9.50	4.57	4.78	4.98
	Total number of negatively charged residues (Asp+Glu)	32	6	26	23	36	28
	Total number of positively charged residues (Arg+Lys)	20	11	39	10	21	18
	Instability index	39.40	31.13	24.86	31.41	36.44	31.09
	Aliphatic index	83.59	89.82	91.78	90.84	86.90	95.60
	GRAVY[a]	-0.082	0.372	-0.286	0.018	-0.077	0.018
SOPMA							
	Alpha helix (%)	16.37	15.64	23.13	12.35	22.01	16.42
	Extended strand (%)	30.96	28.36	23.49	26.29	23.51	27.24
	Beta turn (%)	9.61	8.36	5.34	8.37	3.36	5.60
	Random coil (%)	43.06	47.64	48.04	52.99	51.12	50.75
TMHMM							
	Number of predicted TMHs[b]	0	0	0	0	0	0
	Exp number of AAs in TMHs[c]	5.04384	20.89352	4.53238	0.21876	0.40887	0.04957
	Exp number, first 60 AAs[d]	4.84264	10.60235	4.52981	0.00377	0.40811	0.03674
	Total prob of N-in[e]	0.23798	0.43051	0.19309	0.00600	0.03012	0.01402

[a]Grand average of hydropathicity index.
[b]The number of predicted transmembrane helice.
[c]The expected number of amino acids intransmembrane helices.
[d]The expected number of amino acids in transmembrane helices in the first 60 amino acids of the protein.
[e]The total probability that the N-term is on the cytoplasmic side of the membrane.

The presence of signal peptide cleavage sites in chymostrypsin amino acid sequences was checked by SignalP 4.1 server. The results indicated a signal peptide between the positions 1-19 and a cleavage site between the positions 30 and 31 in *S. exigua* chymotrypsin protein. Moreover, a signal peptide was identified at positions 1-31 in *L. hesperus*, 1-22 in *A. aegypti*, and 1-16 in *T. molitor, P. americana* and *M. musculus* (Figure 5), respectively.

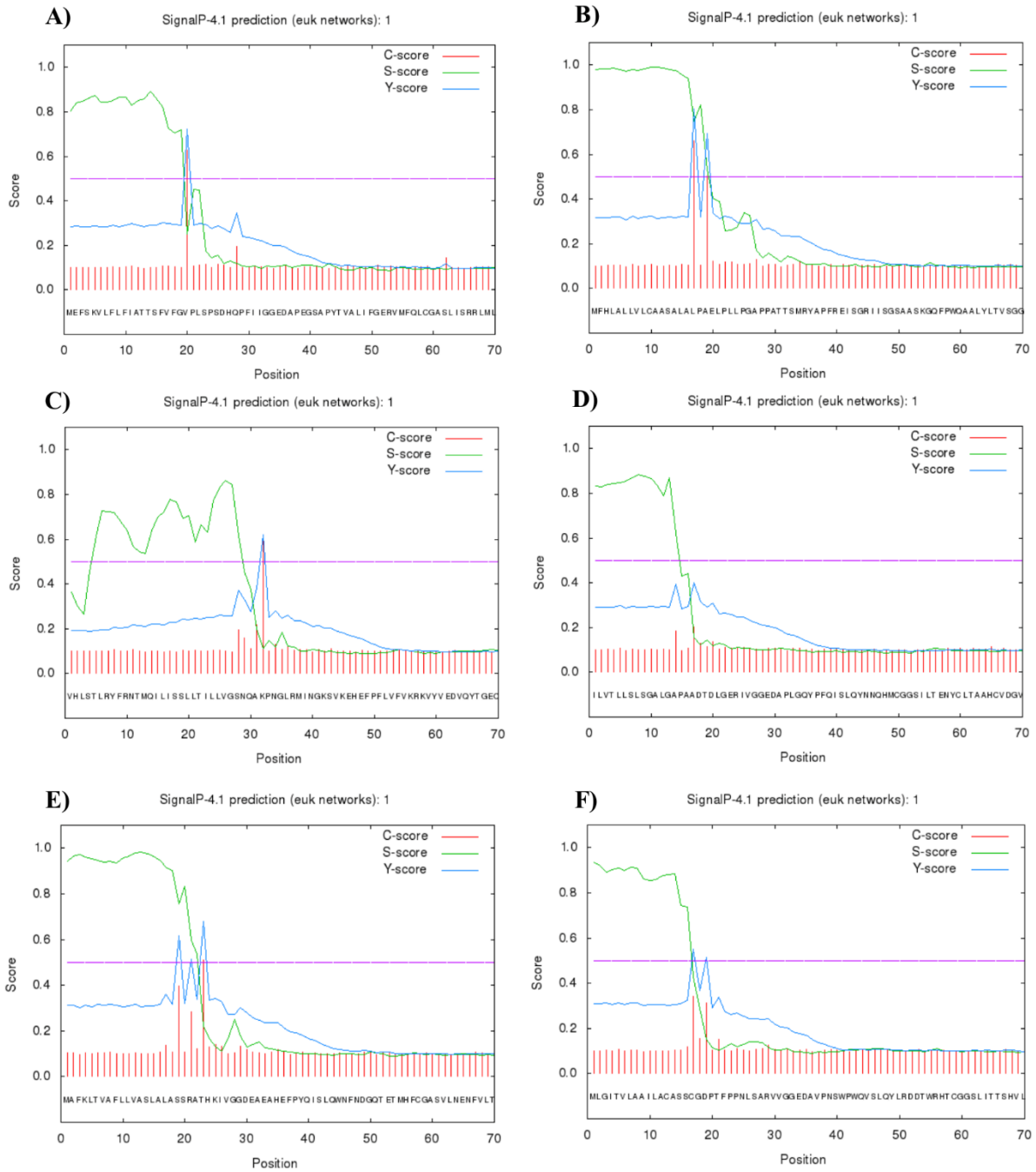

Figure 5. Signal peptide and cleavage sites prediction of chymotrypsin for (A) *Spodoptera exigua*, (B) *Tenebrio molitor*, (C) *Lygus hesperus*, (D) *Periplaneta americana*, (E) *Aedes aegypti* and (F) *Mus musculus* using the SignalP 4.1 server. S-score (signal peptide score), C-score (raw cleavage site score), and Y-score (combined cleavage site score) are revealed.

3D structure prediction and sequences alignment

Alignment of the amino acid sequences of chymotrypsins from *S. exigua*, *T. molitor*, *L. hesperus*, *P. americana*, *A. aegypti* and *M. musculus* indicated that *A. aegypti* chymotrypsin shared 21.7%, 23.4%, 23.1%, 34.5% and 30.8% amino acid sequence identity with *S. exigua*, *T. molitor*, *L. hesperus*, *P. americana* and *M. musculus*, respectively (Figure 6). Besides, the conserved regions at the catalytic sites (His, Asp, Ser), associated with the serine protease family, were identified for all sequences in the alignment (Figure 6). According to the results, the selected sequences probably belong to the same serine protease family of chymotrypsins (Peterson et al., 1994; Wang et al., 1993).

Then, the top-ranked structure models of chymotrypsin from the selected species were selected based on the I-TASSER parameters including C-score, TM-score, and RMSD (Figure 7 and Table 4). Furthermore, the quality of the models was verified by PROCHECK, ERRAT, and Verify-3D (Table 4). PROCHECK represent the overall stereochemical quality of backbone conformation of the model. ERRAT tool was also used to check the overall quality factor of nonbonded interactions in the protein structure, and Verify-3D compares the compatibility of proteins with its amino acid sequence (Colovos and Yeates, 1993).

Functional interaction network of chymotrypsin

The STRING bank data supply a serious evaluation of protein–protein interactions (Szklarczyk et al., 2014). The protein-protein interaction network (PIN) is a valuable tool, which can provide a deeper understanding of the molecular machinery involved in the cellular biological activities (De Las Rivas and Fontanillo, 2012). It refers to the physical binding and interactions that occur between two or more proteins in responses to various disturbances and conditions (Sarkar et al., 2016).

To determine the chymotrypsin interactions, *A. aegypti* chymotrypsin (as the only model of insect chymotrypsin whose interactions have been considered) was mapped utilizing the STRING tool. Based on the results, ten functional partners were exposed in the network analysis (Figure 8).

AAEL011230 has chymotrypsin activity from *A. aegypti* and belongs to the peptidase S1 family (246 aa). AAEL015103 is also a chymotrypsin with predicted serine-type endopeptidase activity (243 aa). Moreover, AAEL009558 and AAEL003219 act as serine proteases and are associated with the peptidase S1 family (316 and 358 aa, respectively). AAEL001703 acts as a Juvenile hormone-regulated chymotrypsin-like serine protease and originated from the peptidase S1 family (252 aa). Furthermore, AAEL006376, AAEL015104, and AAEL011882 indicate trypsin activity involving in the biological process as can be seen in proteolysis (279, 279 and 270, respectively). AAEL008378 is a putative preproacrosin, which belongs to the peptidase S1 family (284 aa). CLIPA3 and AAEL003076 are clip-domain serine protease family (397 aa) and glucosyl/glucuronosyl transferases (524 aa), respectively.

The STRING database collects and integrates data from multiple sources such as experimental data, computational modeling methods and public text collections (Schwartz et al., 2009). Identifying the predicted networks can provide new pathways for future experimental researches and exhibit cross-species predictions for effective interaction mapping (Snel et al., 2000).

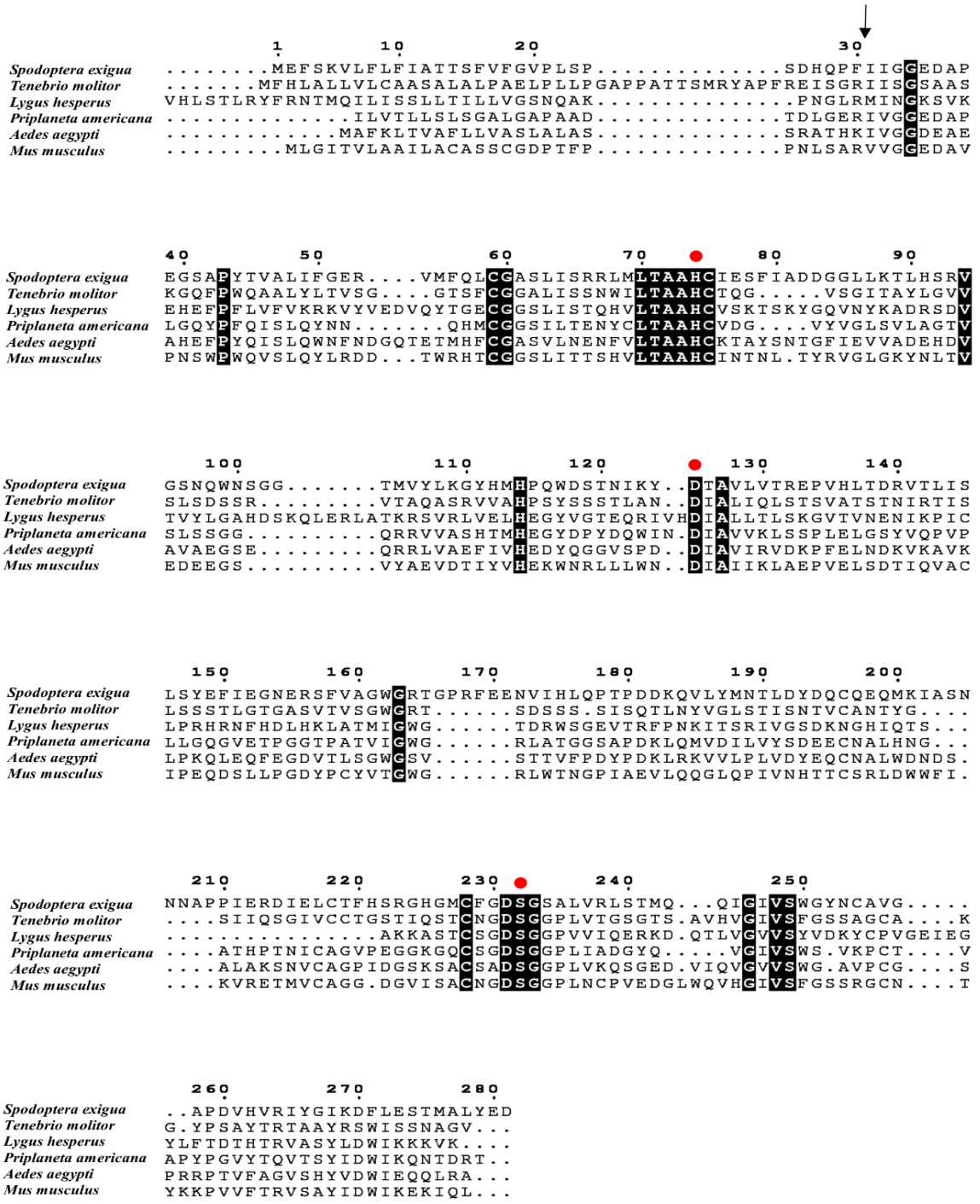

Figure 6. Multiple sequence alignment of chymotrypsin from *Spodoptera exigua* (accession number: AAX35812.1), *Tenebrio molitor* (accession number: ABC88746.1), *Lygus hesperus* (accession number: JAG42770.1), *Periplaneta americana* (accession number: AIA09349.1), *Aedes aegypti* (accession number: Q8T4T5) and *Mus musculus* (accession number: AAI15518.1). Mature enzyme begins at Ile residue (↓). Functionally important residues are indicated using red circles on the top of sequences. Equal residues among all sequences are exhibited by black box. Dots (…) signify sequence alignment gaps. The illustration was generated by ESPript 2.2.

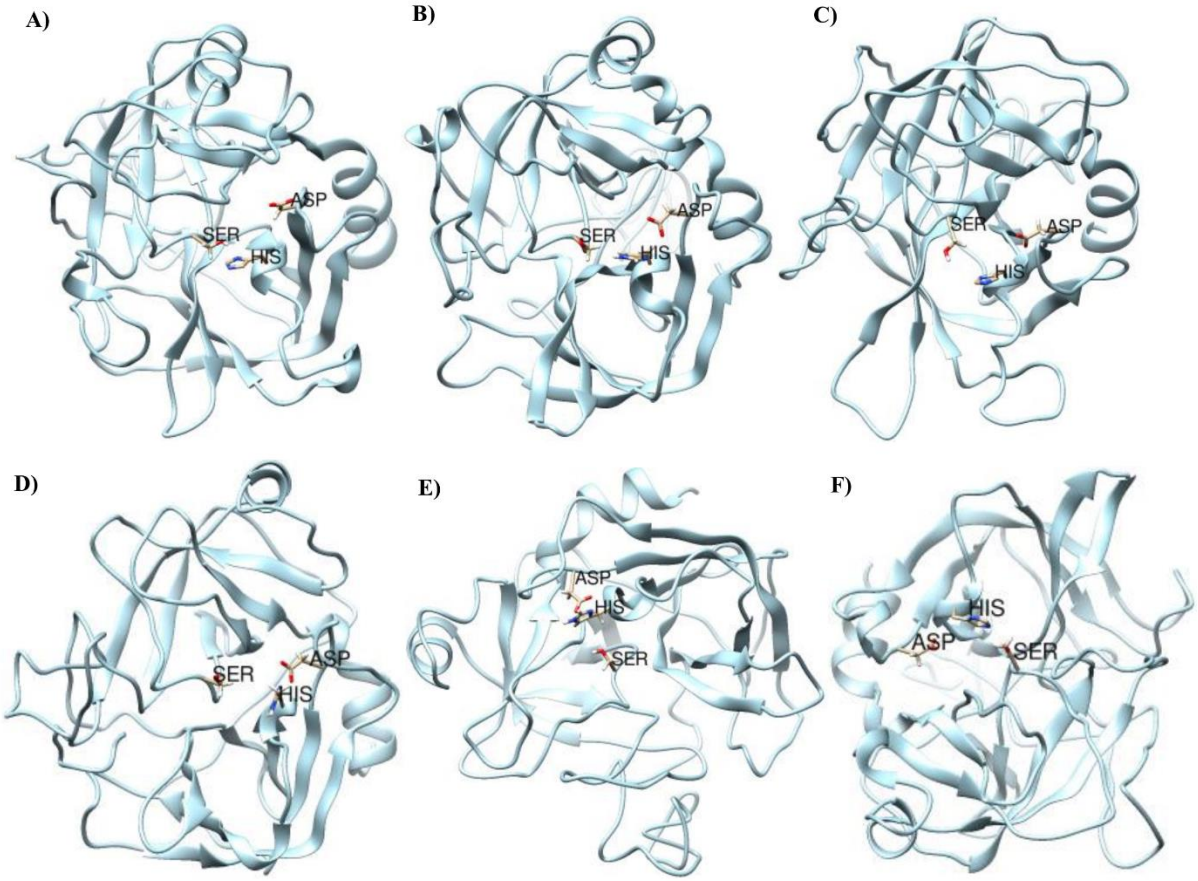

Figure 7. The ribbon illustration of the predicted structural models of chymotrypsin. (A) *Spodoptera exigua*, (B) *Tenebrio molitor*, (C) *Lygus hesperus*, (D) *Periplaneta americana*, (E) *Aedes aegypti* and (F) *Mus musculus*

Table 4. The quality analysis of the predicted structural models of chymotrypsin from various insect species and mouse as a mammalian

Organism	Parameters					
	RMSD	TM-score	C-score	PROCHECK	ERRAT	Verify-3D
Spodoptera exigua	7.5±4.2	0.63±0.14	-0.67	63.6	73.6	90.7
Tenebrio molitor	8.2±4.5	0.59±0.14	-1.0	67.5	74.3	86.5
Lygus hesperus	6.9±4.1	0.67±0.13	-0.37	62.4	78.3	85.4
Periplaneta americana	7.0±4.1	0.64±0.13	-0.55	64.8	77.7	93.2
Aedes aegypti	7.3±4.2	0.64±0.13	-0.60	64.1	73.0	81.2
Mus musculus	4.8±3.1	0.79±0.09	-0.55	73.2	90.8	86.4

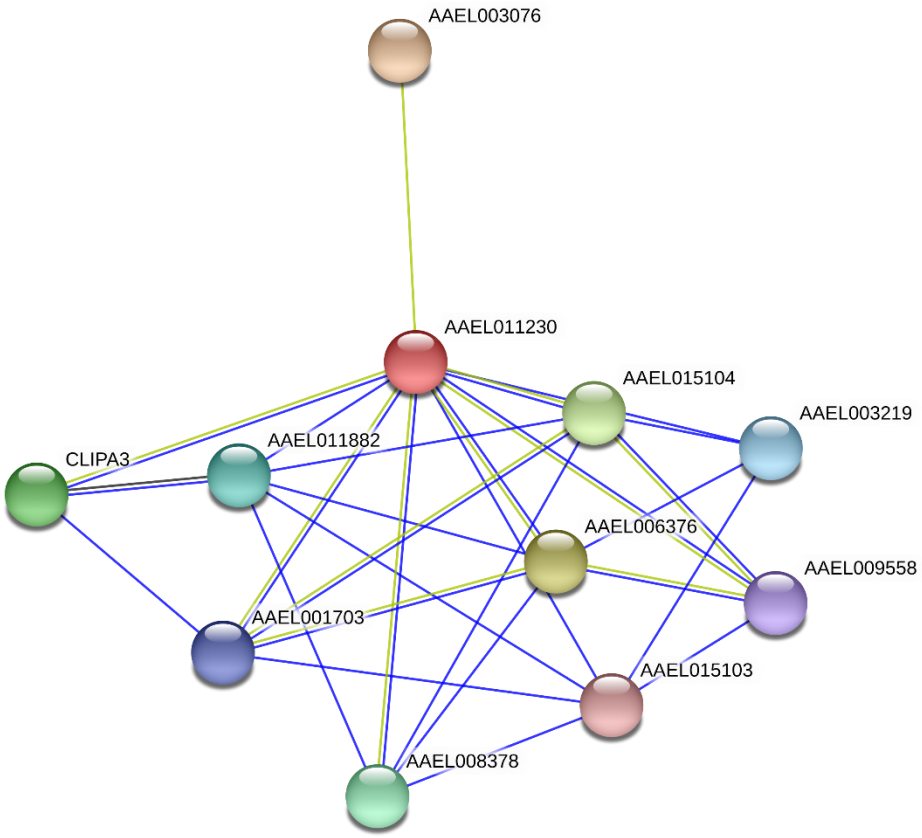

Figure 8. Estimated interactions network view of chymotrypsin from *Aedes aegypti* through the STRING ver. 11. Network nodes denote proteins; colored nodes: white nodes: second shell of interactors; query proteins and first shell of interactors; empty nodes: proteins of unknown tertiary structure, and filled nodes: some tertiary structure is identified. Edges indicate protein-protein relations, which contacts in blue, black, pale green and lavender edges obtained from gene co-occurrence, text mining, co-expression and protein homology.

Conclusion

Chymotrypsin enzymes are involved in many physiological processes in insects, mainly in digestion and reproduction. Accordingly, they are considered as the most promising targets for pest management. Providing biochemical and structural properties of critical insect proteases enables scientists and entomologists to rationally design specific inhibitors against the target enzymes. To date, there is no report on the crystal structure of insects' chymotrypsin. Hence, here we tried to provide confidential information on the structure and function of chymotrypsins through structural assessment, multiple sequence alignment, homology modeling, as well as phylogenetic and motif analysis from various insect species using different bioinformatics tools. Our results can develop our understanding about the structural and functional features of insect chymotrypsins, which is important for selective inhibitors designing against the target enzyme as a promising tool in combating invasion by insect pests.

Acknowledgments

The author is grateful to Shahid Chamran University of Ahvaz for providing financial support for this research (Grant no. SCU.AP99.39134).

REFERENCES

Aluko, R.E. 2015. Amino acids, peptides, and proteins as antioxidants for food preservation. In Handbook of antioxidants for food preservation. Woodhead Publishing. pp. 105-140.

Avwioroko, O.J., Anigboro, A.A., Unachukwu, N.N., and Tonukari, N.J. 2018. Isolation, identification and in silico analysis of alpha-amylase gene of *Aspergillus niger* strain CSA35 obtained from cassava undergoing spoilage. Biochemistry and Biophysics Reports, 14: 35-42.

Baum, D. 2008. Reading a phylogenetic tree: the meaning of monophyletic groups. Nature Education, 1 (1): 190.

Bown, D.P., Wilkinson, H.S., and Gatehouse, J.A. 1997. Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. Insect Biochemistry and Molecular Biology, 27 (7): 625-638.

Colovos, C., and Yeates, T.O. 1993. ERRAT: an empirical atom-based method for validating protein structures. Protein Science, 2 (9): 1511-1519.

De Las Rivas, J., and Fontanillo, C. 2012. Protein–protein interaction networks: unraveling the wiring of molecular machines within the cell. Briefings in Functional Genomics, 11 (6): 489-496.

Dunse, K.M., Stevens, J.A., Lay, F.T., Gaspar, Y.M., Heath, R.L., and Anderson, M.A. 2010. Coexpression of potato type I and II proteinase inhibitors gives cotton plants protection against insect damage in the field. Proceedings of the National Academy of Sciences, 107 (34): 15011-15015.

Elpidina, E.N., Tsybina, T.A., Dunaevsky, Y.E., Belozersky, M.A., Zhuzhikov, D.P., and Oppert, B. 2005. A chymotrypsin-like proteinase from the midgut of *Tenebrio molitor* larvae. Biochimie, 87 (8): 771-779.

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. Springer. pp. 571-607.

Geourjon, C., and Deleage, G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer Applications in the Biosciences, 11 (6): 681-684.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics, 15 (4): 305-308.

Haginaka, J. 2012. 8.9 Chromatographic Separations and Analysis: Protein and Glycoprotein Stationary Phases. Elsevier. pp. 153-176.

Hemmati, S.A., and Mehrabadi, M. 2020. Structural ensemble-based computational analysis of trypsin enzyme genes discovered highly conserved peptide motifs in insects. Archives of Phytopathology and Plant Protection, 53 (7-8): 335-354.

Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics, 8 (3): 275-282.

Kalia, A., and Gosal, S.K. 2011. Effect of pesticide application on soil microorganisms. Archives of Agronomy and Soil Science, 57 (6): 569-596.

Krem, M.M., and Di Cera, E. 2001. Molecular markers of serine protease evolution. The EMBO Journal, 20 (12): 3036-3045.

Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution, 33 (7): 1870-1874.

Lazarević, J., and Janković-Tomanić, M. 2015. Dietary and phylogenetic correlates of digestive trypsin activity in insect pests. Entomologia Experimentalis et Applicata, 157 (2): 123-151.

Mahdavi, A., Ghadamyari, M., Sajedi, R.H., Sharifi, M., Kouchaki, B., and Greenplate, J. 2013. Identification and partial characterization of midgut proteas-es in the lesser mulberry pyralid, *Glyphodes pyloalis*. Journal of insect science, 13 (1): 1-11.

Parekh, V.J., Rathod, V.K., and Pandit, A.B. 2011. Substrate hydrolysis: Methods, mechanism, and industrial applications of substrate hydrolysis. Elsevier. pp. 103-118.

Peterson, A.M., Barillas-Mury, C.V., and Wells, M.A. 1994. Sequence of three cDNAs encoding an alkaline midgut trypsin from *Manduca sexta*. Insect Biochemistry and Molecular Biology 24 (5): 463-471.

Polgár, L. 2005. The catalytic triad of serine peptidases. Cellular and Molecular Life Sciences, 62 (19-20): 2161-2172.

Sarkar, S., Raymick, J., and Imam, S. 2016. Neuroprotective and therapeutic strategies against Parkinson's disease: recent perspectives. International Journal of Molecular Sciences, 17 (6): 904.

Schwartz, A.S., Yu, J., Gardenour, K.R., Finley, Jr.R.L., and Ideker, T. 2009. Cost-effective strategies for completing the interactome. Nature Methods, 6 (1): 55.

Snel, B., Lehmann, G., Bork, P., and Huynen, M.A. 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighborhood of a gene. Nucleic Acids Research, 28 (18): 3442-3444.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., and Mering, C. 2014. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research, 43 (D1): D447-D452.

Terra, W.R. 2005. Biochemistry of digestion. Comprehensive molecular insect science, 171-224.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22 (22): 4673-4680.

van der Meijden, E. 2015. Herbivorous Insects—A Threat for Crop Production. In Lugtenberg, B. (Eds.). Principles of plant-microbe interactions. Springer, Cham. pp. 103-114.

Wang, S., Magoulas, C., and Hickey, D.A. 1993. Isolation and characterization of a full-length trypsin-encoding cDNA clone from the lepidopteran insect, *Choristoneura fumiferana*. Gene, 136 (1-2): 375-376.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. 2015. The I-TASSER Suite: protein structure and function prediction. Nature Methods, 12 (1): 7.

Yang, L., Mei, Y., Fang, Q., Wang, J., Yan, Z., Song, Q., Lin, Z., and Ye, G. 2017a. Identification and characterization of serine protease inhibitors in a parasitic wasp, *Pteromalus puparum*. Scientific Reports, 7 (1): 1-13.

Yang, L., Lin, Z., Fang, Q., Wang, J., Yan, Z., Zou, Z., Song, Q., and Ye, G. 2017b. The genomic and transcriptomic analyses of serine proteases and their homologs in an endoparasitoid, *Pteromalus puparum.* Developmental & Comparative Immunology, 77: 56-68.

Zhu, Y.C., and Baker, J.E. 2000. Molecular cloning and characterization of a midgut chymotrypsin-like enzyme from the lesser grain borer, *Rhyzopertha dominica*. Archives of Insect Biochemistry and Physiology, 43 (4): 173-184.

مطالعات ساختاری، عملکردی و فیلوژنتیکی ژن‌های آنزیم کیموتریپسین در حشرات: با رویکرد بیوانفورماتیک

سید علی همتی

*نویسنده مسوول: استادیار، گروه گیاه‌پزشکی، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران (sa.hemmati@scu.ac.ir)

تاریخ دریافت: ۹۹/۰۹/۱۰ تاریخ پذیرش: ۹۹/۱۰/۲۳

چکیده

آنزیم‌های شبه کیموتریپسین در بسیاری از فرآیندهای فیزیولوژیکی حشرات از جمله هضم، رشد، بقاء و ایمنی نقش دارند. این آنزیم‌ها، پیوندهای پپتیدی را از انتهای کربوکسیلی اسیدهای آمینه حلقوی موجود در پروتئین‌ها هیدرولیز نموده و پپتیدهای فعال و اسیدهای آمینه مورد نیاز برای رشد و تولید مثل حشره را آزاد می‌نمایند. نظر به اهمیت نقش پروتئین‌های غذایی در بقاء و رشد حشرات، توجه بسیاری از حشره شناسان به ساختار، عملکرد، مکانیسم و میان‌کنش‌های آنزیم‌های گوارشی جلب می‌شود. آگاهی از خصوصیات بیوشیمیایی و ساختاری پروتئازهای حشرات می‌تواند به توسعه استراتژی کنترل آفات مبتنی بر مهار کننده‌های پروتئازی اختصاصی کمک نماید. بر این اساس، مطالعهٔ حاضر روی تجزیه و تحلیل ساختار و هم‌ترازی توالی پروتئین، آنالیز فیلوژنتیک و همچنین ارزیابی موتیف‌های محافظت شده در گونه‌های مختلف حشرات با استفاده از ابزارهای مختلف بیوانفورماتیک متمرکز می‌باشد. هم‌ترازی چندگانهٔ توالی‌ها، نواحی مختلف محافظت شده و به طور ویژه حفاظت اسیدهای آمینهٔ دخیل در جایگاه فعال آنزیم (هیستیدین، آسپارتیک اسید، سرین) را نشان داد و همچنین ده موتیف محافظت شده نیز با استفاده از برنامه‌های MEME و MAST، بدست آمد. نتایج مطالعات فیلوژنی نشان می‌دهد که کیموتریپسین‌های حشرات مورد مطالعه احتمالا دارای یک جد مشترک هستند. ساختارهای سه بعدی کیموتریپسین با استفاده از سرور I-TASSER پیش‌بینی شد و کیفیت مدل‌ها با استفاده از برنامه‌های PROCHECK، ERRAT و Verify-3D مورد تأیید قرار گرفت. همچنین شبکه میان‌کنش‌های پروتئین – پروتئین بدست آمده از برنامهٔ STRING 11، ده مسیر ارتباطی برای کیموتریپسین *Aedes aegypti* (L.) را ارائه نمود. به‌طور کلی، مطالعه حاضر می‌تواند چشم‌اندازی جدید جهت طراحی آفت‌کش‌های آتی بر اساس مهارکننده‌های اختصاصی آنزیم‌های گوارشی فراهم نماید.

کلیدواژه‌ها: کیموتریپسین حشرات، موتیف پپتیدی، مطالعه فیلوژنتیکی، مهارکننده‌های پروتئازی، کنترل آفات