مجله دنیای میکروب‌ها
سال هفدهم شماره چهارم (پیاپی ٦١) زمستان ١٤٠٣
صفحات ٢٧٤ –٢٦٠

Original Article

تعیین برخی از ویژگی‌های توالی اسیدآمینه پروتئین انهانسین در باکولوویروس‌ها

مریم راشکی [1*]، مجتبی مرتضوی [2]، ما.کریستینا دل رینکو-کاسترو [3]

[1] دانشیار، گروه تنوع زیستی، پژوهشکده علوم محیطی، پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران. [2] دانشیار، گروه بیوتکنولوژی، پژوهشکده علوم محیطی، پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران. [3] استاد، گروه علوم غذایی و کارشناسی ارشد علوم زیستی، بخش علوم زیستی، دانشگاه گواناخواتو، ایراپواتو، گوآناخوآتو، مکزیک.

چکیده

سابقه و هدف: در میان تمام ویروس‌های شناخته‌شده‌ی بیماری‌زای حشرات، بیشترین مطالعات مربوط به باکولوویریدها است. یکی از پروتین‌های موجود در باکولوویروس‌ها، انهانسین نام دارد. این پروتین تجزیه پروتین‌ها، به‌ویژه موسین، را در غشای اطراف غذای حشرات تسهیل می‌کند و نفوذپذیری باکولوویروس‌ها را افزایش می‌دهد.

مواد و روش‌ها: بر اساس درخت فیلوژنتیکی ٦٧ توالی اسید آمینه‌ای آنزیم انهانسین، نه توالی انتخاب شد. ویژگی‌های فیزیکوشیمیایی توالی‌ها با استفاده از سرور پروت پارام اکس‌پنسی بررسی شد. ویژگی‌هایی مانند پایداری، نقطه ایزوالکتریک، گراوی، زنجیره‌های آلیفاتیک و ضریب جذب نور مورد بررسی قرار گرفت.

یافته‌ها: در این تحقیق، شاخص ناپایداری پروتین‌های انتخاب‌شده کمتر از ٤٠ بود که پیش‌بینی‌کننده پایداری پروتین است. نقطه ایزوالکتریک بزرگتر از هفت، ماهیت قلیایی پروتین را اثبات می‌کند. در این مطالعه، نقاط ایزوالکتریک پروتین‌های انتخاب‌شده کمتر از هفت بود. حداقل و حداکثر مقادیر گراوی در توالی‌های اسید آمینه HaGV مشاهده شد. شاخص آلیفاتیک با پایداری حرارتی پروتین‌ها مرتبط است و محدوده شاخص آلیفاتیک برای پروتین‌های انتخاب شده بین ٨٢/٦٧ تا ١٠٠/٧٧است. ضریب جذب نور، میزان نوری را که یک پروتین در طول موج خاصی جذب می‌کند، نشان می‌دهد.

نتیجه‌گیری: این اندازه‌گیری، زمان خالص‌سازی را با نظارت بر پروتین با اسپکتروفتومتر تخمین می‌زند. با توجه به نقش انهانسین‌ها در افزایش اثربخشی باکولوویروس‌ها به‌عنوان عوامل کنترل بیولوژیکی، پیش‌بینی ساختارهای سه‌بعدی هر نه پروتین انانسین برای شناسایی پروتین‌هایی که ممکن است بیشترین فعالیت را علیه حشرات نشان دهند، ضروری است.

واژگان کلیدی: توالی اسید آمینه، باکولوویرید، انهانسین، آماره‌های فیزیکو شیمیایی.

دریافت مقاله: ١٤٠٣/٩/٩ ویرایش مقاله: ١٤٠٣/١٠/١٧ پذیرش مقاله: ١٤٠٣/١١/١١

*) آدرس برای مکاتبه: گروه تنوع زیستی، پژوهشکده علوم محیطی، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران.
تلفن: ٠٣٤٣٣٧٧٦٦١١ پست الکترونیک: ma_rashkigh@yahoo.com

Journal of Microbial World
Volume 17, No. 4, March 2025

Original Article

Determining some characteristics of the amino acid sequences of the enhancin protein in baculoviruses

Maryam Rashki [1], **Mojtaba Mortazavi** [2], **Ma. Cristina Del Rincón-Castro** [3]

[1] Associate Professor, Department of Biodiversity, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran. [2] Associate Professor, Department of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran. [3] Professor, Food Department and Postgraduate of Biosciences, Life Sciences Division, University of Guanajuato, Irapuato, Gto, México.

Abstract

Background & Objectives: Among all known entomopathogenic viruses, the most extensively studied ones are related to *Baculoviridae*. One protein found in *baculoviruses* is called enhancin. This protein facilitates the breakdown of proteins, particularly mucin, in the peritrophic membrane of insects and enhances the permeability of *baculoviruses*.

Materials & Methods: Based on the phylogenetic tree of the 67 amino acid sequences of the enhancin enzyme, nine sequences were selected. The physicochemical characteristics of the sequences were studied using the ProtParam Expasy server. Features such as stability, isoelectric point, Gravy, aliphatic chains, and light absorption coefficient were studied.

Results: In this research, the instability index of the selected proteins was less than 40, a predictor of protein stability. An isoelectric point greater than seven proves the alkaline nature of the protein. In this study, the isoelectric points of the selected proteins were found to be below seven. The minimum and maximum GRAVY values were observed in the amino acid sequences of HaGV. The aliphatic index is associated with the thermal stability of proteins, with the aliphatic index range for the selected proteins being between 82.67 and 100.77. The light absorption coefficient indicates the amount of light a protein absorbs at a specific wavelength.

Conclusion: This measurement estimates purification time by monitoring protein with a spectrophotometer. Considering the role of enhancins in boosting the effectiveness of baculoviruses as biological control agents, predicting the three-dimensional structures of all nine enhancin proteins is essential for identifying those that may exhibit the highest activity against insects.

Keywords: Amino acid sequence, *Baculoviridae*, enhancin, physicochemical parameters.

Received: 29 November 2024 Revised: 6 January 2025 Accepted: 30 January 2025

Introduction

Baculoviruses are a family of insect pathogenic viruses and are divided into four different genus Alphabaculovirus, Betabaculovirus, Deltabaculovirus and Gammabaculovirus (1) and two phenotypic groups including nucleopolyhedroviruses and granuloviruses. Both groups contain circular double-stranded

Correspondence to: Maryam Rashki
Tel: +98 3433776611
E-mail: ma_rashkigh@yahoo.com
Journal of Microbial World 2025, 17 (4): 260 - 274

DNA genomes of approximately 80-180 kb within nucleocapsids and are predicted to encode approximately 90-180 genes (2). Baculoviruses are the most studied and applied viral biological control agents, and more than 700 species of insects are naturally infected with these viruses, and 90% of them have been isolated from the order of Lepidoptera (3). The study of insect viruses is important because they are not harmful to humans or domestic animals; they are harmless to the environment and do not generate insect resistance (3).

One of the baculovirus proteins is called enhancin, which causes the degradation of the mucin proteins in the peritrophic membrane of insect midguts and increases the permeability of baculoviruses within the columnar cells of the midgut (4). Further, Hoover et al. demonstrated a new mode of action for enhancin. This protein is present in the occlusion bodies of baculovirus and is related to the host infection (5). Moreover, the protein related to alphabaculoviruses, as a part of the occlusion-derived virion (ODV), proved for Lymantria dispar multiple nucleopolyhedrovirus (LdMNPV) (6).

The genome sequence analysis of nucleopolyhedroviruses and granuloviruses showed that these viruses contain many gene homologues that most have been conserved (7). Several gene groups are conserved in some nucleopolyhedroviruses and granuloviruses and one of them is enhancin genes that the largest number of enhancin homologues has been reported for Helicoverpa armigera granulovirus (HearGV) (8).

Nevertheless, little knowledge exists about the characteristics of enhancin amino acids. Pathogenicity may vary in viruses harboring the different amino acid sequences of enzymes. For all NPV enhancin sequences, it was predicted that the domain of the alpha-helix transmembrane was close to the carboxyl terminal, and no alpha-helix transmembrane was conversely expected for the related amino acid sequences of GVs (8).

In this study, Lymantria dispar multiple nucleopolyhedrovirus (LdMNPV) (Accession No. NC-001973), Agrotis segtum nucleopolyhedrovirus -A (AgseNPV-A) (Accession No. NC_007921) and Helicoverpa armigera granulovirus (HearGV) (Accession No. NC_010240) were selected based on phylogenetic relationships among the 67 baculovirus nucleotide and amino acid sequences of enhancin genes (9). The genes include nine nucleotide sequences that encode enhancin enzyme (data from NCBI). The amino acid sequences were analyzed for physicochemical characteristics such as stability, isoelectric point, Grand Average of Hydropathy (Gravy), aliphatic chains, and light absorption coefficient. Moreover, the three-dimensional structures of the nine sequences of peptides belonging to different enhancin genes were predicted.

Materials and Methods

A. Selected amino acid sequences of enhancin proteins in baculoviruses: Based on the phylogenetic analysis of 67 different baculovirus nucleotide and amino acid sequences from enhancin genes (9), three representative baculoviruses including Lymantria dispar multiple nucleopolyhedrovirus (LdMNPV), Agrotis segtum nucleopolyhedrovirus-A (AgseNPV-A) and Helicoverpa armigera granulovirus (HearGV) were selected (one gene per main clade). Nine amino acid sequences of enhancin proteins from these three baculovirus genomes were selected (Table 1).

Table 1: Some information on enhancin sequences of baculoviruses from NCBI.

Virus	Accession No. (Genome)	Protein	Accession No. (Protein)	Length (aa)	Length (nt)	Position (at genome)
LdMNPV	NC_001973	VEF1	NP047702	783	2,331	62,619-60,288
		VEF2	NP047797	788	2,366	158,178-155,812
HaGV	NC_010240	VEF1	YP001649133	823	2,471	143,697-141,226
		VEF2	YP001649134	865	2,597	146,449-143,852
		VEF3	YP001649135	902	2,708	149,181-146,473
		VEF4	YP001649146	856	2,570	160,585-158,015
AsNPV-A	NC_007921	VEF1	YP529745	877	2,633	71,598-68,965
		VEF2	YP529746	883	2,651	74,372-71,721
		VEF3	YP529798	862	2,588	127,866-125,278

B. Identity percentage, frequency and motif structure of enhancin amino acid sequences in baculoviruses: The identity percentage of sequences was calculated using the BLASTP program (http:// www.ncbi.nlm.com), and the frequency of amino acid sequences was determined using ProtParam program (10). The Pfam program was employed to identify any significant motifs in each protein structure (https://pfam.xfam.org), and the detected motifs were visualized using the HM weblogo (11).

C. Physicochemical properties of the enhancin proteins: The physicochemical properties of the enhancin proteins including theoretical isoelectric point (pI), extinction coefficients (assuming all pairs of Cys residues form cysteine), aliphatic index (AI), instability index (II) and grand average hydropathy (GRAVY) were analyzed using ProtParam program (10). An instability index below 40 indicates a stable protein, while a value above 40 suggests instability. The computed isoelectric point (PI) helps develop buffer systems for purifying recombinant proteins via isoelectric focusing; a PI above seven indicates the protein's alkaline nature. The GRAVY value is the total hydropathy of all amino acids divided by the number of residues in the sequence.

D. Prediction of transmembrane helices of the enhancin proteins: The TMHMM-2.0 program was employed to predict protein transmembrane helices (http://www.cbs.dtu.dk/services/TMHMM). The predicted number of transmembrane helices and the expected number of amino acids within them were assessed. A prediction exceeding 18 helices suggests a higher likelihood of the protein being transmembrane or having a signal peptide. Additionally, the expected amino acid count within transmembrane helices for the first 60 amino acids was analysed; a higher count indicates a potential signal peptide in the N-terminal region. The total likelihood of the N-terminal being cytoplasmic and the locations of the transmembrane sites were also evaluated.

E. Determination of enhancin protein types: The SOSUI program was used to detect the type of proteins (soluble or membrane) (https:// harrier.nagahama-i-bio.ac.jp/sosui).

F. Determination of enhancin protein Hydrophobicity: The ProtScale program (https://web.expasy.org/protscale/) was used to assess protein hydrophobicity by generating

profiles based on various amino acid scales. Kyte and Doolittle's method was employed to identify hydrophobic amino acids (12). Additionally, the hydrophilicity scale was utilised to predict potential protein structures and domains.

*G. Prediction of enhancin protein secondary and three-dimensional structure:*An online server, CFSSP, was used to predict protein secondary structure (http://www.biogem.org/tool/chou-fasman/index.php).

This program identifies regions such as alpha helices, beta sheets, and turns in the amino acid sequence and utilises the Chou-Fasman algorithm, which analyses the relative frequencies of amino acids in these structures based on X-ray crystallography data from known proteins.

The enhancin proteins were modelled using the I-TASSER server (http://zhanglab.ccmb.med.umich. edu/I-TASSER/). Protein sequences were modelled based on the closest available three-dimensional structures. The best model was selected based on the lowest e-value, highest coverage, suitable TM-Score, overall C-score, and Expected RMSD.

Results

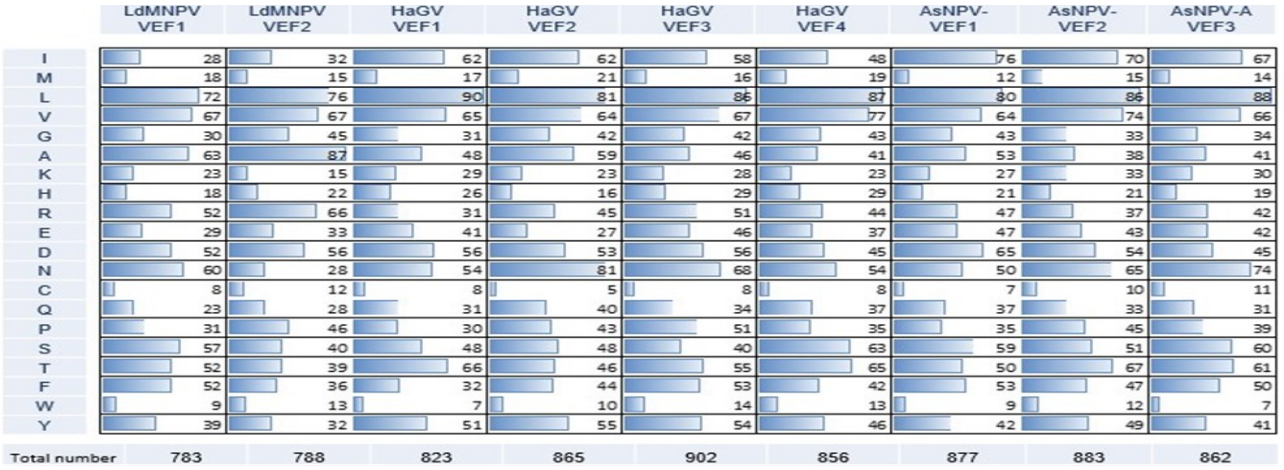
A. Identity percentage, frequency and motif structure of enhancin amino acid sequences in baculoviruses: The identity percentage results for nine different amino acid sequences of enhancin proteins have been shown in Table 2. The highest identity percentage was between AsNPV-A-VEF1 and AsNPV-A-VEF2 with 100%, and the lowest was between HaGV-VEF2 and AsNPV-A-VEF2, LdMNPV-VEF2 and AsNPV-A-VEF2, HaGV-VEF4 and AsNPV-A-VEF1 with 20.42, 20, 69 and 21.99 %, respectively.

Table 2: The identity percentage of amino acid sequences calculated using BLASTP.

Protein	LdMNPV-VEF1	LdMNPV-VEF2	HaGV-VEF1	HaGV-VEF2	HaGV-VEF3	HaGV-VEF4	AsNPV-A-VEF1	AsNPV-A-VEF2	AsNPV-A-VEF3
LdMNPV-VEF1	100	30.11	26.42	32.66	31.34	32.35	24.86	23.07	25.25
LdMNPV-VEF2	30.33	100	25.67	29.28	28.52	24.34	23.84	20.69	20.46
HaGV-VEF1	26.42	25.67	100	29.86	30.76	25.56	22.24	26.92	24.48
HaGV-VEF2	32.66	29.28	29.86	100	34.13	28.66	25.46	20.42	21.68
HaGV-VEF3	31.34	28.52	30.76	34.13	100	29.71	27.36	25.00	21.21
HaGV-VEF4	32.35	24.34	25.56	28.66	29.71	100	21.99	22.41	24.49
AsNPV-A-VEF1	24.86	23.84	22.24	25.46	27.36	21.99	100	46.70	36.34
AsNPV-A-VEF2	23.07	20.69	26.92	20.42	25.00	22.41	46.70	100	35.73
AsNPV-A-VEF3	25.25	20.46	24.48	21.68	21.21	24.49	36.34	35.73	100

The lowest and highest number of amino acids was found in the specific sequence of LdMNP-VEF1 and HaGV-VEF3 protein,respectively (Fig.1). The leucine amino acid (nonpolar) had the highest frequency among others in all enhancin sequences. In the HaGV-VEF2 sequence, asparagine (polar) frequency equalled with leucine.

In the next step, the Pfam program identified significant motifs in each protein structure, with results including domain names, positions, and predicted e-values presented in Table 3. No significant Pfam matches were found for LdMNPV-VEF2, HaGV-VEF1, and AsNPV-A-VEF3. The peptidase M60 motif was detected in LdMNPV-VEF2 and HaGV-VEF2/VEF3, while the mucin bdg motif was found in other sequences.



	LdMNPV VEF1	LdMNPV VEF2	HaGV VEF1	HaGV VEF2	HaGV VEF3	HaGV VEF4	AsNPV-VEF1	AsNPV-VEF2	AsNPV-A VEF3
I	28	32	62	62	58	48	76	70	67
M	18	15	17	21	16	19	12	15	14
L	72	76	90	81	86	87	80	86	88
V	67	67	65	64	67	77	64	74	66
G	30	45	31	42	42	43	43	33	34
A	63	87	48	59	46	41	53	38	41
K	23	15	29	23	28	23	27	33	30
H	18	22	26	16	29	29	21	21	19
R	52	66	31	45	51	44	47	37	42
E	29	33	41	27	46	37	47	43	42
D	52	56	56	53	56	45	65	54	45
N	60	28	54	81	68	54	50	65	74
C	8	12	8	5	8	8	7	10	11
Q	23	28	31	40	34	37	37	33	31
P	31	46	30	43	51	35	35	45	39
S	57	40	48	48	40	63	59	51	60
T	52	39	66	46	55	65	50	67	61
F	52	36	32	44	53	42	53	47	50
W	9	13	7	10	14	13	9	12	7
Y	39	32	51	55	54	46	42	49	41
Total number	783	788	823	865	902	856	877	883	862

Figure 1: The frequency of enhancin amino acid using the ProtParam program.

Table 3: The significant match results from the Pfam search.

Virus	Proteins	Domain-Family	Position (Start-end)	E-value
LdMNPV	VEF1	Peptidase M60	111-365	3.9 e-10
	VEF2	-	-	-
HaGV	VEF1	-	-	-
	VEF2	Peptidase M60	118-381	8.3 e-12
	VEF3	1) Peptidase M60	745-862	5.6 e-08
		2)Mucin bdg	112-366	2.7 e-07
	VEF4	Mucin bdg	731-844	7.2 e-09
AsNPV-A	VEF1	Mucin bdg	506-610	1.4 e-10
	VEF2	Mucin bdg	513-622	7.5 e-14
	VEF3	-	-	-

B. Physicochemical properties of the enhancin proteins: The physicochemical properties of enhancin proteins were analysed using the ProtParam program (Table 4). An instability index below 40 indicates protein stability, while an index above 40 suggests possible instability. The isoelectric point (PI) greater than seven indicates an alkaline nature. The isoelectric point (PI) is the pH at which a molecule, such as a protein or amino acid, has no net electrical charge. A PI below seven indicates an acidic nature, meaning the molecule carries a net negative charge at a neutral pH (around seven). Molecules with a PI

Table 4: The physicochemical properties of the enhancin proteins.

Sequence Property	LdMNPV VEF1	LdMNPV VEF2	HaGV VEF1	HaGV VEF2	HaGV VEF3	HaGV VEF4	AsNPV-A VEF1	AsNPV-A VEF2	AsNPV-A VEF3
MW	89336.93	88408.65	93995.83	98848.20	104791.86	97834.15	100130.37	101622.90	98486.24
PI	6.33	6.28	5.04	5.77	5.68	6.03	4.94	5.24	5.74
Ext.C.	108110	119930	114990	137200	157960	140540	112455	139635	100215
Ext.C. (reducing Cys)	107610	119180	114490	136950	157460	140040	112080	139010	99590
HT-M	30 h	30 h	30 h	30 h	30 h	30 h	30 h	30 h	30 h
HT-Y	>20 h	>20 h	>20 h	>20 h	>20 h	>20 h	>20 h	>20 h	>20 h
HT-C	> 10 h	> 10 h	> 10 h	> 10 h	> 10 h	> 10 h	> 10 h	> 10 h	> 10 h
II	36.64	39.41	38.08	33.53	39.33	36.96	37.83	31.85	37.53
AI	82.67	89.15	100.77	92.75	88.90	92.38	96.58	97.51	97.09
GRAVY	-0.209	-0.136	-0.083	-0.169	-0.273	-0.127	-0.114	-0.098	-0.098

MW: Molecular weight (Da), PI: Theoretical isoelectric point, Ext. C: Extinction coefficients, assuming all pairs of Cys residues form cysteine, Ext. C (reducing cys): Extinction coefficients, assuming all Cys residues are reduced, HT-M: Estimated half-life (mammalian reticulocytes, in vitro), HT-Y: Estimated half-life (yeast, in vivo), HT-C: Estimated half-life (*Escherichia coli*, in vivo), II: Instability index, AI: Aliphatic index, GRAVY: Grand average of hydropathicity.

under seven are more soluble in acidic environments and may behave differently in biological contexts, influencing their interactions with other molecules. All VEF values in Table 4 are below seven, signifying that the associated molecules or proteins are predominantly acidic. They likely contain carboxyl or similar functional groups that affect their overall charge and behaviour in biological systems. This acidity may also affect how these molecules interact with proteins, enzymes, or substrates, potentially leading to specific biological functions.

In this study, the instability index of enhancin proteins exceeded 40, and their PIs were all above seven. The aliphatic index, representing the volume of aliphatic side chains (alanine, valine, isoleucine, and leucine) that contribute to thermostability, ranged from 82.67 to 100.77 for the selected proteins. The extinction coefficient measures the light absorption of a protein at a specific wavelength, which is useful for monitoring during purification. The extinction coefficient (ε) is a key biochemistry parameter that measures how much light a protein absorbs at a specific wavelength, expressed in units of M^{-1} cm^{-1}.

It is influenced by the protein's amino acid composition, notably aromatic residues such as tryptophan, tyrosine, and phenylalanine; a higher extinction coefficient suggests a greater abundance of these residues, offering insights into the protein's structure. Additionally, the extinction coefficient can indicate protein conformation, as changes in the absorbance spectrum, particularly in the UV range, may signal denaturation or conformational changes that affect the functional capabilities of protein. Among the enhancin proteins from the baculovirus species, LdMNPV exhibited the lowest molecular weight relative to VEF1 and VEF2. According to GRAVY values, HaGV contained enhancin proteins with the highest (VEF3) and the lowest (VEF1) hydrophilic amino acids.

C. Prediction of transmembrane helices of the enhancin proteins: Figure 2 displays graphs for the predicted helix positions. Additionally, results in Table 5 provide the number of predicted helices and their locations.Notably, HaGV proteins lack transmembrane helices, with the positions of TM sites for both inside and outside listed at the end of the table. Transmembrane Helices Hidden Markov Model (TMHMM) is a computational tool designed to predict the presence and location of transmembrane helices in proteins. It employs a Hidden Markov Model, which excels in sequence analysis and pattern recognition. This statistical approach effectively captures the probabilistic nature of biological sequences, making it well suited for interpreting complex protein data.

The results showed that the number of predicted transmembrane helices varied between zero and one. The enhancin proteins of the HaGV had no transmembrane helices. If a single transmembrane helix is predicted for AsNPV-A-VEF1, it could indicate that this protein may be partially membrane-associated. The region with the helix might play a role in anchoring the protein to the membrane, which could be important for its function, perhaps in retaining activity near cell membranes or facilitating interactions with membrane-bound receptors. The presence of a potential helix could suggest that AsNPV-A-VEF1 may interact with lipid bilayers in a specific context, like possibly aiding in the viral entry mechanism or enhancing interaction with host cell membranes.

The absence of predicted transmembrane

helices for HaGV-VEF4 indicates that this protein is likely soluble and interacts with cellular components through cytoplasmic or extracellular mechanisms rather than through membrane anchoring. This suggests a more likely role in modulating processes that do not require direct membrane association. Given its soluble nature, HaGV-VEF4 might play a role in enzyme activity or signaling pathways that occur away from membranes, thus engaging with other proteins or substrates in the cytosol or extracellular matrix.Similar to HaGV-VEF4 and the absence of LdMNPV-VEF1 transmembrane helices means that this protein is likely not membrane-bound and may interact with other proteins in a soluble state. This characteristic may indicate that LdMNPV-VEF1 is involved in systemic processes, such as immune modulation or cellular signaling, without needing to interface directly with cellular membranes.

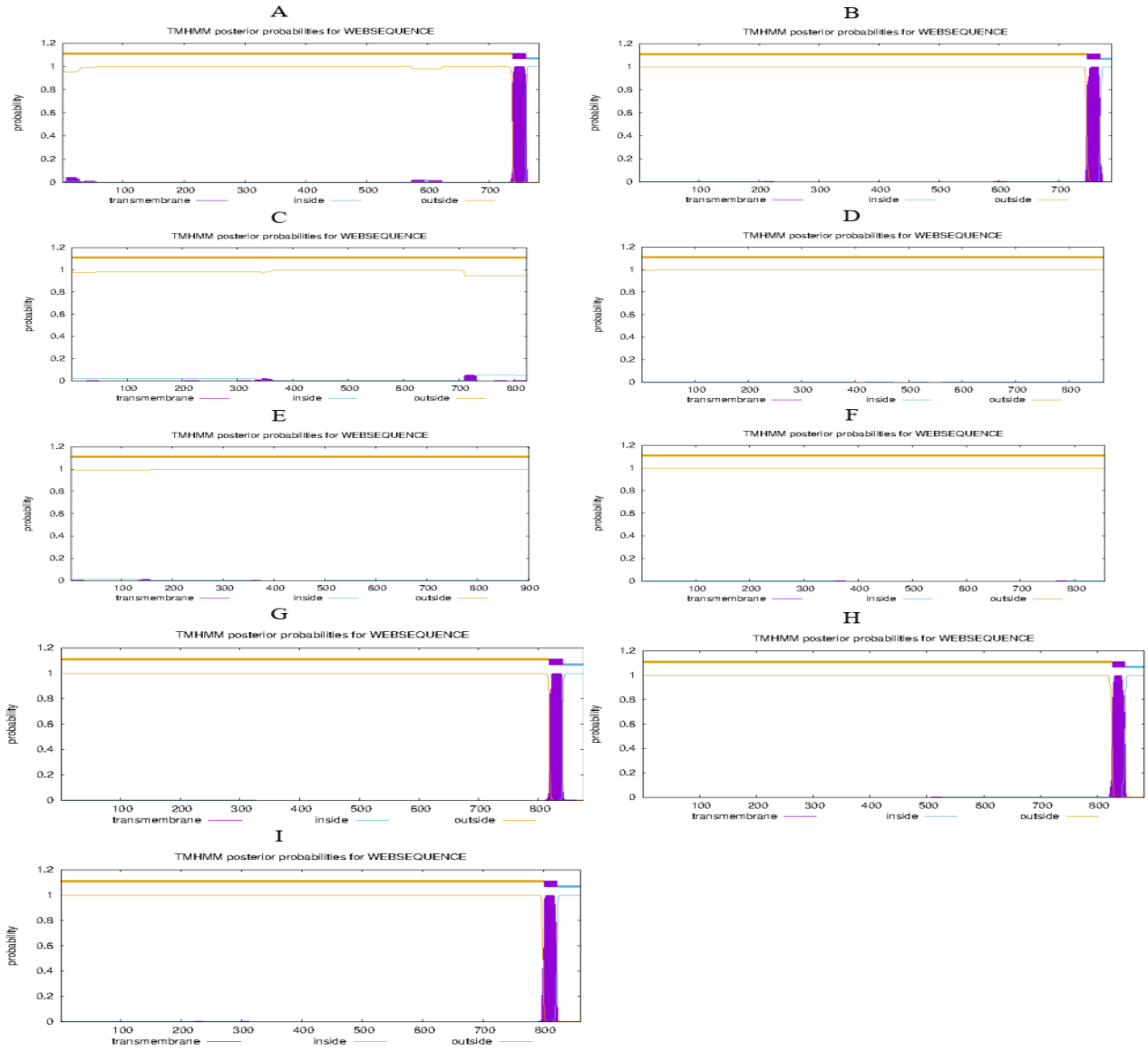


Figure 2: The obtained graphs using TMHMM-2.0 related to the position of the enhancin predicted helices in, A and B: LdMNPV-VEF1 and VEF2, respectively; C, D, E and F: HaGV-VEF1, VEF2, VEF3 and VEF4, respectively and G, H and I: AsNPV-A-VEF1, VEF2 and VEF3, respectively.

Table 5: Prediction of transmembrane helices (TMHs) of the enhancin proteins using TMHMM-2.0.

Protein Parameter		LdMNPV VEF1	LdMNPV VEF2	HaGV VEF1	HaGV VEF2	HaGV VEF3	HaGV VEF4	AsNPV-A VEF1	AsNPV-A VEF2	AsNPV-A VEF3
No. of predicted TMHs		1	1	0	0	0	0	1	1	1
Expected no. of AAs in TMHs		24.99	23.0	1.69	0.10	0.35	0.12	22.55	22.82	22.94
Expected no., first 60 Amino Acids		1.16	0.0005	0.008	0.0059	0.053	0.0006	0.00045	4e-05	0.023
Total probability of N-in		0.04	0.0005	0.022	0.004	0.013	0.0027	0.00018	0.00006	0.0012
Positions (Amino Acid)	Outside	1-738	1-746	1-823	1-865	1-902	1-856	1-818	1-826	1-800
	TMH	739-761	747-769	-	-	-	-	819-841	827-849	801-823
	Inside	762-783	770-788	-	-	-	-	842-877	850-883	824-862

D. Determination of enhancin protein types: According to Table 6, all enhancin proteins of HaGV were soluble in water except VEF3. The rest of the studied proteins were membrane types. Membrane-crossing sequences in proteins suggest their potential interaction with the membrane, likely positioning different protein regions on either side. This suggests that these proteins likely differ significantly in function and role. The function of proteins pertains to their specific actions, while their role encompasses a broader view of how these functions contribute to overall biological processes within cells or organisms. Understanding both aspects is essential for clarifying the significance of membrane-crossing sequences and their effects on protein behaviour and interactions with cellular membranes.

Table 6: SOSUI prediction of enhancin protein types.

Virus	Protein	Predicted type of protein	Region	Transmembrane sequence
LdMNPV	VEF1	Membrane	739-761	IMMAVVIFCFVLVALVIVFILVF
	VEF2	Membrane	749-771	SVLSLVAVGVVCLLLLFFVIATI
HaGV	VEF1	Soluble	-	-
	VEF2	Soluble	-	-
	VEF3	Membrane	137-159	SSGYCFLYLDLVCILVPPASKNV
	VEF4	Soluble	-	-
AsNPV-A	VEF1	Membrane	817-839	GPAIPLIVVIGAIVFFIAIILFL
	VEF2	Membrane	827-849	LFVVGVTVLVIIIIIIIVKFLVG
	VEF3	Membrane	794-816	DLRPLALVAAVAVLIIIIVVIYK

E. Determination of enhancin protein Hydrophobicity: The shape and scale of the plots (Fig. 3) provide insight into the proteins' secondary structures of enhancin. For example, a stretch of about 20 hydrophobic amino acids may indicate an alpha helix crossing a lipid bilayer made of hydrophobic fatty acids. In contrast, highly hydrophilic amino acids suggest these residues interact with solvents or water, likely placing them on the protein's outer surface. The variations in water repellency suggest corresponding structural differences and consequently affect their performance.

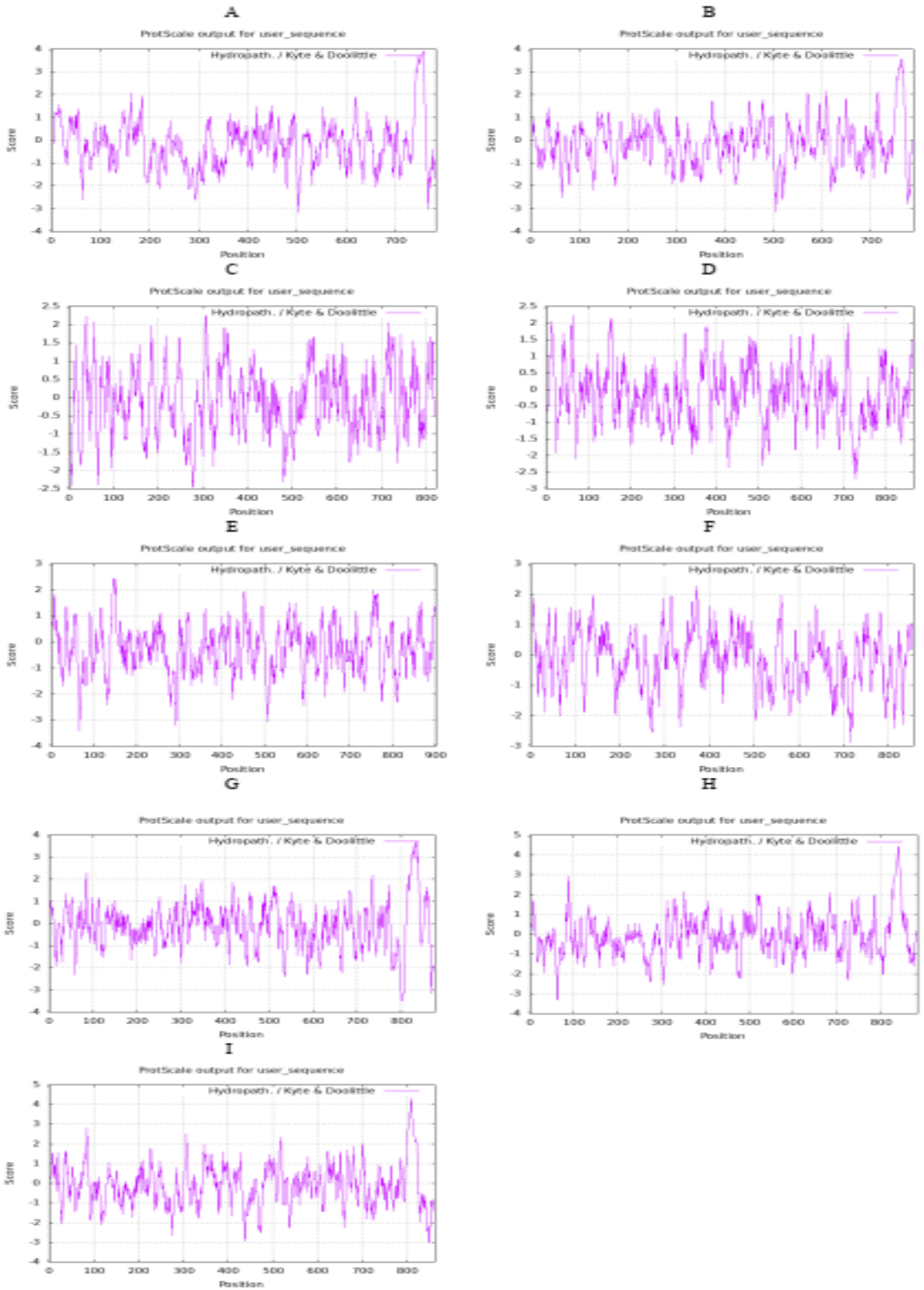


Figure 3: The obtained graphs using ProtScale program for detecting the hydrophobicity of the enhancin protein sequences in, A and B: LdMNPV-VEF1 and VEF2, respectively; C, D, E and F: HaGV-VEF1, VEF2, VEF3 and VEF4, respectively and G, H and I: AsNPV-A-VEF1, VEF2 and VEF3, respectively.

F. Prediction of enhancin protein secondary and three-dimensional structures: The Chou-Fasman algorithm results showed that AsNPV-A-VEF1 had the highest percentage of helices, HaGV-VEF4 had the most beta-sheets, and LdMNPV-VEF1 exhibited the most turns(Table 7).

Table 7: Percentage of helices turns and beta-sheets in the predicted secondary structure of enhancin proteins using the Chou-Fasman algorithm.

Baculovirus species	Protein	Helices (%)	Beta-sheet (%)	Turn (%)
LdMNPV	VEF1	60.3	48.1	11.7
	VEF2	69.4	62.7	11.4
HaGV	VEF1	68.8	67	10
	VEF2	66	54.6	10.4
	VEF3	58.8	53.5	11.2
	VEF4	66.1	82.5	11.6
AsNPV-A	VEF1	70.6	71.3	11.6
	VEF2	66.5	61.5	10.9
	VEF3	61.3	77.8	11.1

Additionally, the three-dimensional structures of all nine enhancin proteins were predicted (Fig.4). Interpreting the results of the Chou-Fasman algorithm and the predicted 3D structures of enhancin proteins is essential to understanding their functions and roles. The 3D structure of AsNPV-A-VEF1 (with the highest Percentage of helices) is expected to show a significant number of alpha helices, represented by coiled or spiral regions, which likely contribute to the protein's stability and potential interactions with other molecules or membranes. This model belonging to HaGV-VEF4 (with the most beta-sheets) should exhibit substantial beta-sheet structures, characterized by pleated sheets arranged in parallel or anti-parallel formations, likely providing stability through hydrogen bonding between strands. The 3D structure of LdMNPV-VEF1 (with the most turns) should display a higher prevalence of turns (or loops), which are short connections between secondary structure elements, providing flexibility and allowing the protein to adopt varying conformations.In all three proteins, the number of turns was nearly equal, while alpha helices ranged from 60 to 70 percent. The most significant difference lay in the beta-sheet, which varied by thirty four percent between LdMNPV and HaGV.

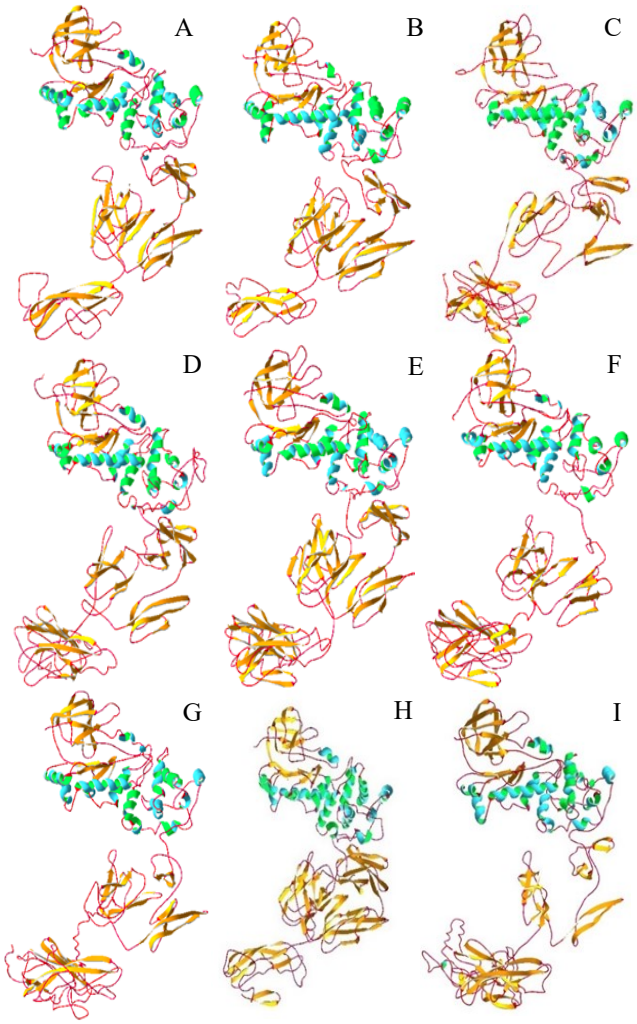


Figure 4: Predicted three-dimensional structures of the enhancin proteins using the I-TASSER server for, A and B: LdMNPV-VEF1 and VEF2, respectively; C, D, E and F: HaGV-VEF1, VEF2, VEF3 and VEF4, respectively and G, H and I: AsNPV-A-VEF1, VEF2 and VEF3, respectively.

Discussion

In this study, the biochemical properties and the prediction of the three-dimensional structures of all nine enhancin proteins were determined. LdMNPV enhancins VEF 1 and VEF2, HaGV VEF1, VEF2, VEF3 and VEF4, and AsNPV VEF1, VEF2 and VEF3. The highest identity percentage was between AsNPV-A-VEF1 and AsNPV-A-VEF2 with 100% and the lowest was between HaGV-VEF2 and AsNPV-A-VEF2, LdMNPV-VEF2 and AsNPV-A-VEF2, HaGV-VEF4 and AsNPV-A-VEF1 with 20.42, 20, 69 and 21.99 %, respectively. In another research, the comparison of the LdMNPV-VEF-1 enhancins amino acid sequence revealed the presence of a signature pattern characteristic of a zinc-binding domain found within metalloproteases (13, 8). The signature pattern, HEXXH, is sufficient to group a protein into the metalloprotease superfamily. Most *baculovirus* enhancins have this conserved metalloprotease zinc binding (14). However, in this work, a comparison of the number of amino acids in baculovirus enhancins identifies a great deal of heterogeneity. For example, AsNPV-A-VEF1 and AsNPV-A-VEF2 is 100% identical. In contrast, HaGV-VEF2 and AsNPV-A-VEF2 is only 20.4 % identical, this is understandable since both enhancins belong to two very different groups of baculoviruses, the NPVs and the GVs. However, this does not explain why the differences between the two enhancins of NPVs studied also present very low levels of percentages of identity, also being around 20% (AsNPV and LdMNPV).

Conversely, in this work, no significant Pfam matches were found for LdMNPV-VEF2, HaGV-VEF1, and AsNPV-A-VEF3. The peptidase M60 motif was detected in LdMNPV-VEF2 and HaGV-VEF2/VEF3, while the mucin bdg motif was found in other sequences. The size of baculovirus enhancins is from 758 amino acids to 1004 amino acids. All GV enhancins, except AgseGV, form a group; all NPV enhancins form another group. The high level of heterogeneity exhibited by the baculovirus enhancins may suggest that these genes arose in viral genomes from independent sources, as could be observed in this study since the motifs present in each amino acid sequence were similar among the enhancins of the GV but different from those of the enhancins of the NPV.

A study on peritrophic membranes from *T. ni* and *P. unipuncta* larvae showed that the enhancins of both granuloviruses made the peritrophic membrane of the virus AcMNPV compared to controls (15). These results show that this structure is a barrier to virus movement and facilitates infection (16). Other findings probed that the target substrate for GV enhancins is insect intestinal mucin and the degradation of intestinal mucin increased access of virions to the midgut epithelial cells (17). In this study, no experiments were performed to analyze the direct effect of the nine enhancins studied. However, the determination of their possible secondary structures indicated that the enhancins of the granuloviruses (GVs) are soluble and do not bind to cell membranes. In contrast, the enhancins of the nucleopolyhedroviruses (NPVs) presented alpha-helix structures, suggesting that they are membrane-binding proteins. This difference implies that the enhancins from these two virus groups may operate through distinct mechanisms (18).

Granulovirus enhancins may be classified as soluble proteins, which could explain their presence inside the GV granules instead of being fused to the virion envelopes like the

NPV proteins (8). The NPV enhancins were found within ODV envelopes, and it is a common characteristic among NPV enhancins, in contrast to GV enhancins being located within granules (6). The difference indicates that they use different mechanisms to degrade the peritrophic membrane.

Conclusion

In this study, clear differences were found between the physical and biochemical properties between the enhancins of the GV and the NPV and between the NPV themselves, which indicates that they must have a very different mechanism of action (19). The biotechnological potential of each of them must be tested *in vivo* to determine which of them provides improved bioinsecticidal activity to the virus that has it and if it is possible to exchange information between them to obtain *baculoviruses* with highly improved virulence properties.

Conflict of interest

There is no conflict of interest.

References

1. Jehle JA, Blissard GW, Bonning BC, Cory JS, Herniou EA, Rohrmann GF, Theilmann DA, Thiem SM, Vlak JM. On the classification and nomenclature of baculoviruses: a proposal for revision. Arch Virol. 2006; 151: 1257-1266.
2. van Oers MM, Vlak JM. Baculovirus Genomics. *Current Drug Targets.* 2007; 8: 1051-1068.
3. Del Rincón-Castro, MaC, Ibarra, JE. Entomopathogenic viruses. In: Biological control of insect pests, Rosas N, editor, 1st edition, Studium Press LLC, USA; 2011.
4. Wang M, Hu Z. Cross-talking between baculoviruses and host insects towards a successful infection. Philosophical Transactions of the Royal Society B: Biological Sciences, 2019; 374: 20180324.
5. Hoover K, Humphries MA, Gendron AR, Slavicek JM. Impact of viral enhancin genes on potency of *Lymantria dispar* nucleopolyhedrovirus in *L. dispar* following disruption of the peritrophic matrix. J Inverteb Pathol. 2010; 104:150-152.
6. Slavicek JM, Popham HJ. The Lymantria dispar nucleopolyhedrovirus enhancins are components of occlusion-derived virus. J Virol. 2005; 79: 10578-88.
7. Herniou EA, Jehle JA. Baculovirus phylogeny and evolution. Curr Drug Targets. 2007; 8: 1043-1050.
8. Slavicek JM. Baculovirus enhancins and their role in viral pathogenicity. In: Adoga M, editors. Molecular virology. Rijeka, Croatia: InTech; 2012: 147-155.
9. Rashki M, Mortazavi M. Examination of nucleotide and amino acid sequences of enhancin enzyme in baculoviruses. Journal of Microbial World, 2023; 16 (3): 219 - 231.
10. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: Walker M, editor. The proteomics protocols Handbook. Humana press; 2005: 571-607.
11. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004:14(6): 1188-1190.
12. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology, 1982; 157 (1): 105-132.
13. Murphy GJP, Murphy G, Reynolds JJ. The origin of matrix metalloproteinases and their familial relationships. FEBS Lett. 1991; 289: 4-7.
14. Bischoff DS, Slavicek, JM. Molecular analysis of an enhancin gene in the Lymantria dispar nuclear polyhedrosis virus. J. Virol. 1997; 71: 8133-8140.
15. Del Rincón-Castro, MaC, Ibarra JE. Effect of a nucleopolyhedrovirus of *Autographa californica* expressing the enhancin gene of *Trichoplusia ni* granulovirus on T. ni larvae. Biocontrol Science and Technology, 2005; 15(7): 701-710.

16. Peng J, Zhong J, Granados RR. A baculovirus enhancin alters the permeability of a mucosal midgut peritrophic matrix from lepidopteran larvae. J Insect Physiol. 1999; 45:159-166.
17. Wang P, Granados RR. An intestinal mucin is the target substrate for a baculovirus enhancin. Proc Natl Acad Sci, USA, 1997; 94: 6977-6982.
18. Lepore LS, Roelvink PR, Granados RR. Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease, 1996; 68(2):131-40.
19. Ferrelli ML, Salvador R. Effects of mixed baculovirus infections in biological control: a comprehensive historical and technical analysis. 2023; 15(9): 1838.