Journal of Advanced Zoology

ISSN: 0253-7214

*Volume **45** Issue 6 Year **2024** Page 322-347*

An *In Silico* Comparative Study of Eleven Wasp Venom Allergens: An Arena for Therapeutic Approach

Satarupa Das[1*], Dr. Neeta Shweta Kerketta[2], Jayanta Sinha[3]

[1]Ph.D. Research Scholar, Jharkhand Rai University, Ranchi, Jharkhand, India.
[2]Associate Professor, Department of Agriculture, Jharkhand Rai University, Ranchi, Jharkhand, India.
[3]Ex-Associate Professor, P.G. Department of Zoology, Vivekananda College, Thakurpukur, Kolkata, West Bengal, India.

*[*]**CORRESPONDING AUTHOR:** Satarupa Das*
[*]Aikatan Housing Society, AA-5, Prafulla Kanan(W), Block-C, Flat No.-19, Kolkata-700101, West Bengal, India. e-mail Id: sdbiology0508@gmail.com

Article History

***Abstract*:**

Peptide toxins in animals as part of the chemical arsenal for predation and/or protection that can also safeguard the host from pathogenic infections. Insects including the hymenopterans generate a battery of toxic bactericidal or bacteriostatic molecules which are small multifunctional, linear peptides that cause pain, have antimicrobial effects, and inflammatory processes. The toxins under the experiment are active against gram-positive, gram-negative bacteria and fungi. The present *in silico* study aims to predict the physicochemical attributes like molecular weight, theoretical pI, amino acid composition, extinction co-efficient, estimated half-life, instability index, aliphatic index and grand average of hydropathy (gravy) of 11 wasps' (class: Insecta; order: Hymenoptera) venom allergens through Expasy Protparam & Pepstat software tools. The secondary structures of the toxins were predicted using psi-Blast-based secondary structure prediction GOR4 tools revealing the % α helix, extended β strand, random coil and ambiguous state reflecting a comparative picture of physicochemical parameters of these venom allergens. 3D Homology modelling of these toxins was accomplished through Swiss-model webserver tool and validated through various *in silico* tools like ANOLEA, ProSA-web, QMean4 determining Z score, PROCHEK establishing the 3D models of these toxins. Use of Inter Pro, CDD, PROSITE, P fam, Tox DL, PrDOS software predicted the protein family, protein toxicity, protein disorder respectively. Scratch Protein Predictor software tools predicted cysteine – cysteine bonds. Docking of the eleven (11) wasp venom allergens individually with bacterial cell wall component N-acetylglucosamine was done by CB DOCK webserver resulting negative affinity scores reflecting towards the strong binding between the mentioned 11 toxins and N-acetylglucosamine indicating that the mentioned wasps' toxin molecules might be used as potential antibacterial therapeutic molecules binding to N-acetylglucosamine leading to an avenue to the probable bacterial drug discovery.

KEY WORDS: *Wasp, venom allergen toxin, in silico, Docking, Anti-bacterial*

2. MATERIALS AND METHODS:

Various bioinformatic software tools that were employed for the *insilico* work are as follows:

a. Sequences of the eleven wasps' toxins were retrieved from Uniport database (www.expasy.org/sprot) [18].
b. Expasy ProtParam tool and Pepstat tool to compute the physicochemical properties of wasp venom allergens [19].
c. PSI-blast-based secondary structure prediction GOR4 software tool adopted to characterize and predict secondary structure of wasp venom allergens [20, 21].
d. 3D Homology modelling of these Wasp Venom Allergens were computed through SWISS-MODEL tool [22].
e. 3D Model Validation using ANOLEA, ProSA , QMEAN4 programs determining Z score and PROCHECK software tools [23-25].
f. Cysteine Bonding state prediction through scratch protein predictor software tool [26,27].
g. CDD,InterPro,Prosite,ToxDL, and PrDOS software were used for protein family prediction, protein toxicity prediction, protein disorder prediction respectively [28-31].
h. The molecular structure of N-acetylglucosamine is being mined from the compound repository Pubchem [32].
i. Docking of the said eleven (11) wasp venom allergens individually with bacterial cell wall component N-acetylglucosamine was done through CB DOCK docking webservers [33].
j. Toxin Codes were obtained from Uniport database (www.expasy.org/sprot) for each Uniprot ID for each of the eleven (11) wasp venom allergens (Table 1).

3. RESULTS AND DISCUSSION:

3.1. RESULTS:

Table 1a. Sequences of eleven wasp venom allergens retrieved from Uniport database (www.expasy.org/sprot).

Sl. No.	Name of venom	Source	Uniprot Accession No.	Toxin Code	Peptide Sequence	Sequence length (No. of amino acid)
1.	Venom allergen 5 (Antigen5) (Cysteine-rich venom protein)	*Microctonus hyperodae*	A9YME1	VA5_MICHY	MEQIKYLLIGIIFSSAISSSLQCAYE NCADWANPGQSQKHTMCLYPTT ALGNKCNEGRIIQLTEADKQYILQ LHNELRAKVASGGESQGSNGPQP AGKIGPLKWDNEIAEIAQRWVNQ CTFEHDKCRNTKANSVGQNLYM MGSSEKSENTHDILTASVNSWYSE VKDFDNRSVREYKFEFTTGHYSQ VVWGDTTHVGCGLVQYKDSGFY TTMVACNYSPAGNLIGGTVYPTL	232
2.	Venom allergen 5 (Antigen 5)	*Vespula maculifrons*	Q2L6Z1	VAL5_VESMC	MEISGLVYLIIIVTIIDLPYGKANNY CKIKCLKGGVHTACKYGSLKPNC GNKVVVSYGLTKQEKQDILKEHN DFRQKVARGLETRGNPGPQPPAKN MKNLVWNDELAYVAQVWANQCQ YGHDTCRDVAKYQVGQNVALTGS TAAKYENPVNLVKMWENEVKDY NPKKKFSENNFIKIGHYTQMVWA NTKEIGCGSMKYTENKWHYHYLV CNYGPSGNFGNEELYQTK	227
3.	Venom allergen 5 (Allergen Ves m V)	*Vespula*	P35760	VA5_VESMC	NNYCKIKCLKGGVHTACKYGSLK PNCGNKKVVSYGLTKQEKQDILK EHNDFRQKIARGLETRGNPGPQPP AKNMKNLVWSDELAYIAQVWAN QCQYGHDTCRDVAKYQVGQNVA LTGSTAAVYNDPVKLVKMWEDEV KDYNPKKKFSENNFLKIGHYTQM VWANTKEVGCGSIKYIQENWHKH YLVCNYGPSGNFQNEELYQTK	203

4.	Venom allergen 5 (Antigen 5) (Ag5)	*Polybia paulista*	P86686	VA5_POLPI	NKYCNIKCSKVAHTVCQTGESTKP SSKNCAKVSITSVGVTEEEKKLIV DEHNRFRQKVAQGLETRGNPGPQ PAASDMNNLVWNDELAYIAQVWA SQCQFFVHDKCRNTAQYQVGQNI AYSASTAAYPGVVKLIVLWENEVK DFNYNTGITKENFAKVGHYTQVV WAKTKEVGCGSIKYIEKGMKSHY LVCNYGPAGNVLGAQIYEIK	207
5.	Venom allergen 52 (Antigen 5)	*Polybia paulista*	D4P2Y4	VA52_POLPI	NKYCNIKCSKVAHTVCQYGESTK PSSKCNKVSITSVGVTEEEKKLIVD EHNRFRQKVAQGLETRGNPGPQP AASDMNNLVWNDELAYIAQVWA SQCQFFVHDKCRNTAQYQVGQNI AYSASTAAYPGIVSLIVLWENEVK DFNYSQGITKENFSKVGHYTQVV WAKTKEVGCGSIKYIEKGMKSHY LVCNYGPAGNYMGQPIYTKK	206
6.	Venom allergen 5 (Allergen Pol a V) (Antigen 5) (Ag5)	*Polistes annularis*	Q05109	VA5_POLAN	SSQGVDYCKIKCPSGIHTVCQYGE STKPSKNCAGKVIKSVGPTEEEKK LIVSEHNRFRQKVAQGLETRGNPG PQPAASDMNDLVWNDELAHIAQV WASQCQFLVHDKCRNTAKYPVGQ NIAYAGGSNLPDVVSLIKLWENEV KDFNYNTGITKQNFAKIGHYTQM VWGKTKEIGCGSLKYMENNMQN HYLICNYGPAGNYLGQLPYTKK	209
7.	Venom allergen 5 (Allergen Ves s V) (Antigen 5)	*Vespula squamosa*	P35786	VA5_VESSQ	VDYCKIKCLKGGVHTACKYGTST KPNCGNMVVKSYGVTQAEKQEIL KIHNDFRNKVARGLETRGNPGPQP PAKNMNNLVWNNELANIAQIWAS QCKYGHDTCKDTTKYNVGQNIAV SSSTAAVYENVGNLVKAWENEVK DFNPTISWEQNEFKKIGHYTQMV WAKTKEIGCGSIKYVDNNWYTHY LVCNYGPAGNFGNQEVYERK	205
8.	Venom allergen 5 (Allergen Pol f V) (Antigen 5)	*Polistes fuscatus*	P35780	VA5_POLFU	VDYCKIKCSSGIHTVCQYGESTKP SKNCADKVIKSVGPTEEEKKLIVN EHNRFRQKVAQGLETRGNPGPQP AASDMNNLVWNDELAHIAQVWA SQCQILVHDKCRNTAKYQVGQNI AYAGGSKLPDVVSLIKLWENEVK DFNYNKGITKQNFGKVGHYTQMI WAKTKEIGCGSLKYMKNNMQHH YLICNYGPAGNYLGQLPYTKK	205
9.	Venom allergen 5	*Vespula flavopilosa*	P35783	VA5_VESFL	NNYCKIKCLKGGVHTACKYGSLK PNCGNKVVVSYGLTKQEKQDILK EHNDFRQKIARGLE TRGNPGPQPPAKNMKNLVWNDEL AYVAQVWANQCQYGHDTCRDIA KYQVGQNVALTGSTA AKYDDPVKLVKMWEDEVKDYNP KKKFSGNNFLKTGHYTQMVWAN TKEVGCGSIKFIQEKW HKHYLVCNYGPSGNFQNEELYQT K	204
10.	Venom allergen 5		Q7Z156	VA5_POLSR	NKYCNIKCSKVAHTVCQYGESTK PSSKNCNKVSITSVGVTEEEKKLIV DEHNRFRQKVAQ GLETRGNPGPQPAASDMNNLVWN DELAYIAQVWANQCQFFVHDKCR NTAQYQVGQNIAYS ASTAAYPGIVSLIVLWENEVKDFN YSQGITKENFAKVGHYTQVVWAK TKEVGCGSIKYIE	207

		Polybia scutellaris rioplatensis			KGMKSHYLVCNYGPAGNYMGQP IYTKK	
11.	Venom allergen 3 homolog	*Dinoponera quadriceps*	P0DSI3	VA3_DINQU	NNYCKIKCRSGIHTLCKYGTSTKP NCGRSVVKASGLTKAEKLEILKQH NEFRQKVARGLE TRGNPGPQPPAKSMNTLVWNDEL AQIAQVWASQCKYGHDNCRNTA KYLVGQNIAEQSTTA ASFEPVSNMVKMWSDEVKDYQY GSSKNKLNDVGHYTQMVWAKTK EIGCGNIKYIENGWHH HYLVCNYGPAGNIGNEPIYEKK	231

Table 1b. Structure of N-acetylglucosamine (https://pubchem.ncbi.nlm.nih.gov).

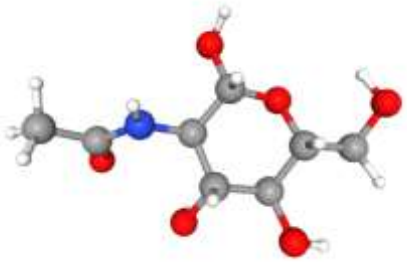
PubChem CID	Molecular Formula (Molecular weight)	2D Structure	3D Structure
24139	C8H15NO6 (**221.21g/mol**)		

Table 2. Physicochemical properties of eleven wasp venom allergens.

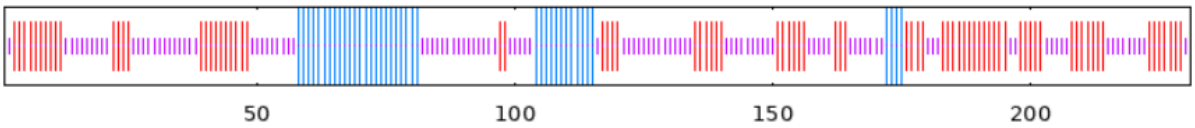
Sl. No.	Name of Toxin	Source	Toxin Code	Residues	Charge	Molecular weight	Isoelectric point	Extinction coefficient	Instability iIndex	Aliphatic Index	GRAVY
1.	Venom allergen 5 (Antigen5) (Cysteine-rich venom protein)	*Microctonus hyperodae*	VA5_MICHY	232	-1	25764.8	6.2178	1.761	42.01	68.53	-0.473
2.	Venom allergen 5 (Antigen 5)	*Vespula maculifrons*	VAL5_VESMC	227	12	25831.53	9.0579	1.987	22.80	71.23	-0.618
3.	Venom allergen 5 (Allergen Ves m V)	*Vespula maculifrons*	VA5_VESMC	204	13	23332.51	9.2415	2.009	20.10	63.04	-0.828

4.	Venom allergen 5 (Antigen 5) (Ag5)	*Polybia paulista*	VA5_POLPI	207	8.5	23009.12	8.6328	1.668	17.94	73.43	-0.425
5.	Venom allergen 5 2 (Antigen 5)	*Polybia paulista*	VA52_POLPI	206	9.5	23084.17	8.7875	1.792	23.30	67.62	-0.515
6.	Venom allergen 5 (Allergen Pol a V) (Antigen 5) (Ag5)	*Polistes annularis*	VA5_POLAN	209	9	23293.43	8.634	1.648	27.76	68.09	-0.608
7.	Venom allergen 5 (Allergen Ves s V) (Antigen 5)	*Vespula squamosa*	VA5_VESSQ	205	9.5	23114.07	8.7936	2.201	18.24	62.73	-0.693
8.	Venom allergen 5 (Allergen Pol f V) (Antigen 5)	*Polistes fuscatus*	VA5_POLFU	205	13.5	23068.37	9.2865	1.664	23.84	71.32	-0.636
9.	Venom allergen 5	*Vespula flavopilosa*	VA5_VESFL	204	14	23274.47	9.3898	1.95	21.38	61.13	-0.833
10.	Venom allergen 5	*Polybia scutellaris rioplatensis*	VA5_POLSR	207	9.5	23209.3	8.7875	1.782	21.80	67.78	-0.53
11.	Venom allergen 3 homolog	*Dinoponera quadriceps*	VA3_DINQU	202	13.5	22717.74	9.2953	1.932	36.79	62.28	-0.765
Mean				209.81	21.13	230609.04	8.7385	1.854	25.08	67.01	-0.636
Range	**Min**			202	-1	22717.74	6.2178	1.668	17.94	61.13	-0.833
	Max			232	14	25831.53	9.3898	2.201	42.01	73.43	-0.425

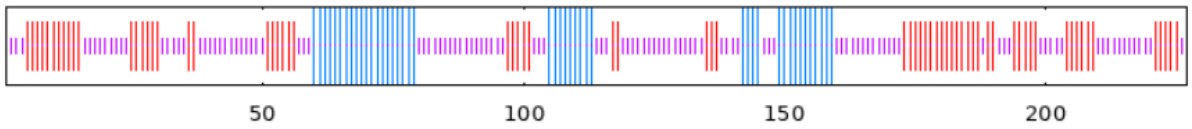
Table 3. Prediction of Secondary structures of eleven wasp venom allergens.

Sl. No.	Name of venom	Source	Toxin Code	α-Helix (%)	Extended β-strand (%)	Random coil (%)	Ambiguous state (%)
1.	Venom allergen 5 (Antigen5) (Cysteine-rich venom protein)	*Microctonus hyperodae*	VA5_MICHY	17.24	34.05	48.71	0

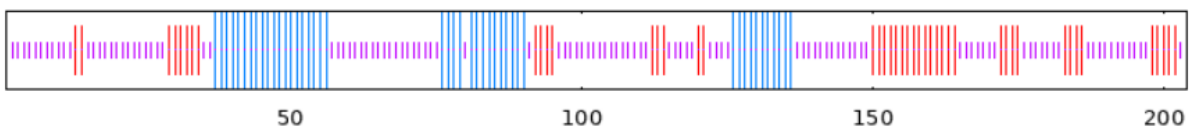
2.	Venom allergen 5 (Antigen 5)	*Vespula maculifrons*	VAL5_VESMC	19.38	29.96	50.66	0
3.	Venom allergen 5 (Allergen Ves m V)	*Vespula maculifrons*	VA5_VESMC	22.06	22.06	55.88	0
4.	Venom allergen 5 (Antigen 5) (Ag5)	*Polybia paulista*	VA5_POLPI	16.91	23.67	59.42	0
5.	Venom allergen 5 2 (Antigen 5)	*Polybia paulista*	VA52_POLPI	16.99	21.84	61.17	0
6.	Venom allergen 5 (Allergen Pol a V) (Antigen 5) (Ag5)	*Polistes annularis*	VA5_POLAN	22.49	26.79	50.72	0
7.	Venom allergen 5 (Allergen Ves s V) (Antigen 5)	*Vespula squamosa*	VA5_VESSQ	24.88	24.39	50.73	0
8.	Venom allergen 5 (Allergen Pol f V) (Antigen 5)	*Polistes fuscatus*	VA5_POLFU	24.88	23.41	51.71	0
9.	Venom allergen 5	*Vespula flavopilosa*	VA5_VESFL	21.57	22.06	56.37	0
10.	Venom allergen 5	*Polybia scutellaris rioplatensis*	VA5_POLSR	20.29	22.22	57.49	0
11.	Venom allergen 3 homolog	*Dinoponera quadriceps*	VA3_DINQU	22.77	18.32	58.91	0
Total				229.46	268.77	601.77	0
Range		**Min**		16.91	18.32	48.71	0
		Max		24.88	34.05	61.17	0



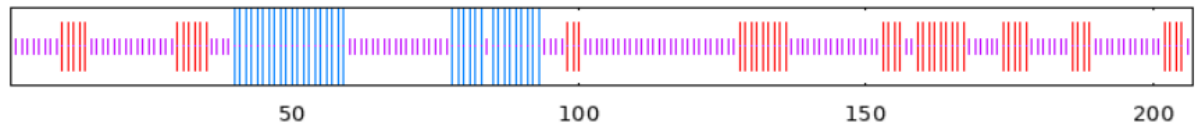
VA5_MICHY



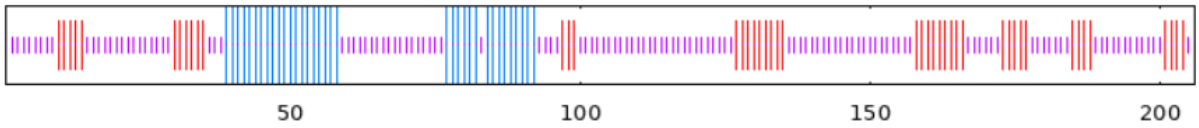
VAL5_VESMC



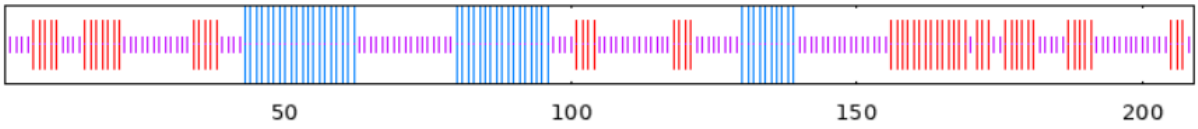
VA5_VESMC



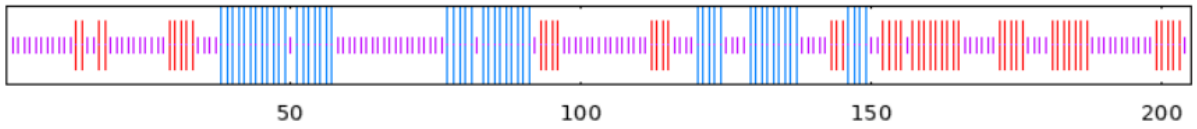
VA5_POLPI



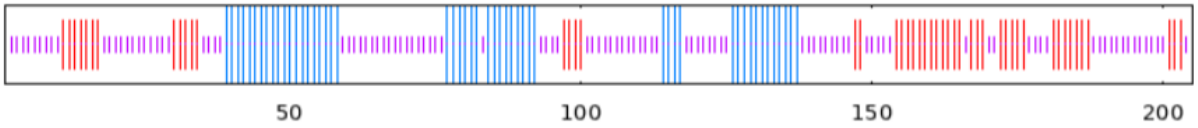
VA52_POLPI



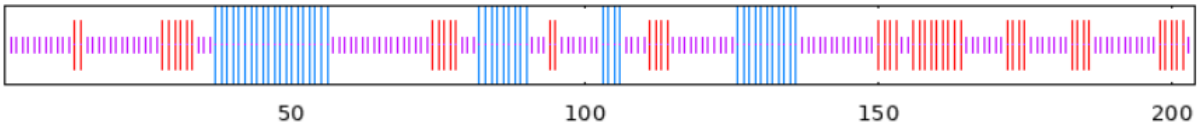
VA5_POLAN



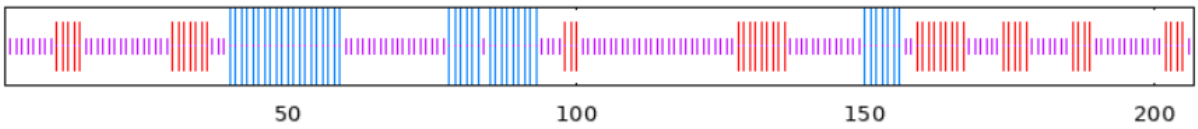
VA5_VESSQ



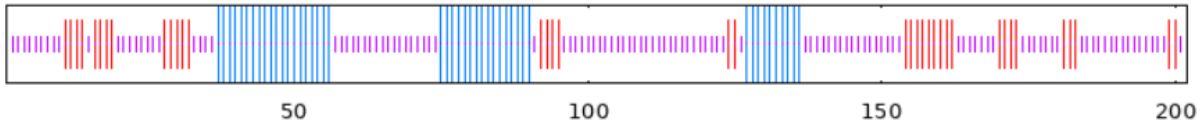
VA5_POLFU



VA5_VESFL



VA5_POLSR



VA5_POLSR

BLUE: Helix

RED: Sheet

PURPLE: Coil

Fig. 1. Secondary Structure (cartoon diagram) of eleven wasp venom allergens.

Table 4. Cysteine-Cysteine bonding state prediction of eleven wasp venom allergens.

Sl. No.	Name of Toxins	Source	Toxin code	Cysteine sequence position	Predicted Bonds
1.	Venom allergen 5 (Antigen5) (Cysteine-rich venom protein)	*Microctonus hyperodae*	VA5_MICHY	[198-215] [119-126] [23-28]	3
2.	Venom allergen 5 (Antigen 5)	*Vespula maculifrons*	VAL5_VESMC	[193-210] [27-40] [117-124]	3
3.	Venom allergen 5 (Allergen Ves m V)	*Vespula maculifrons*	VA5_VESMC	[170-187] [4-17] [94-101]	3
4.	Venom allergen 5 (Antigen 5) (Ag5)	*Polybia paulista*	VA5_POLPI	[173-190] [97-105] [4-16]	3
5.	Venom allergen 5 2 (Antigen 5)	*Polybia paulista*	VA52_POLPI	[172-189] [96-104] [4-16]	3
6.	Venom allergen 5 (Allergen Pol a V) (Antigen 5) (Ag5)	*Polistes annularis*	VA5_POLAN	[175-192] [100-108] [8-20]	3
7.	Venom allergen 5 (Allergen Ves s V) (Antigen 5)	*Vespula squamosa*	VA5_VESSQ	[171-188] [4-17] [95-102]	3
8.	Venom allergen 5 (Allergen Pol f V) (Antigen 5)		VA5_POLFU	[171-188] [96-104] [4-16]	3

		Polistes fuscatus			
9.	Venom allergen 5	*Vespula flavopilosa*	VA5_VESFL	[170-187] [4-17] [94-101]	3
10.	Venom allergen 5	*Polybia scutellaris rioplatensis*	VA5_POLSR	[173-190] [97-105] [4-16]	3
11.	Venom allergen 3 homolog	*Dinoponera quadriceps*	VA3_DINQU	[168-185] [4-16] [94-101]	3

Table 5. Protein family prediction of eleven wasp venom allergens.

Sl. No.	Toxin Code	CDD	InterPro	PROSITE	Pfam	Toxin family
1.	VA5_MICHY	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family.
2.	VAL5_VESMC	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family
3.	VA5_VESMC	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family. Venom allergen 5 like sub-family
4.	VA5_POLPI	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family
5.	VA5_POLPI	cd05380	IPR018244 IPR014044 IPR035940 IPR001283	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family

			IPR002413			
6.	VA5_POLAN	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family
7.	VA5_VESSQ	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family
8.	VA5_POLFU	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family
9.	VA5_VESFL	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family
10.	VA5_POLSR	cd05380	IPR018244 IPR014044 IPR035940 IPR001283 IPR002413	PS01009 PS01010	PF00188	Crisp family.Venom allergen 5 like sub-family
11.	VA3_DINQU	cd05380	IPR018244 IPR014044 IPR035940 IPR001283	PS01010	PF00188	Crisp family

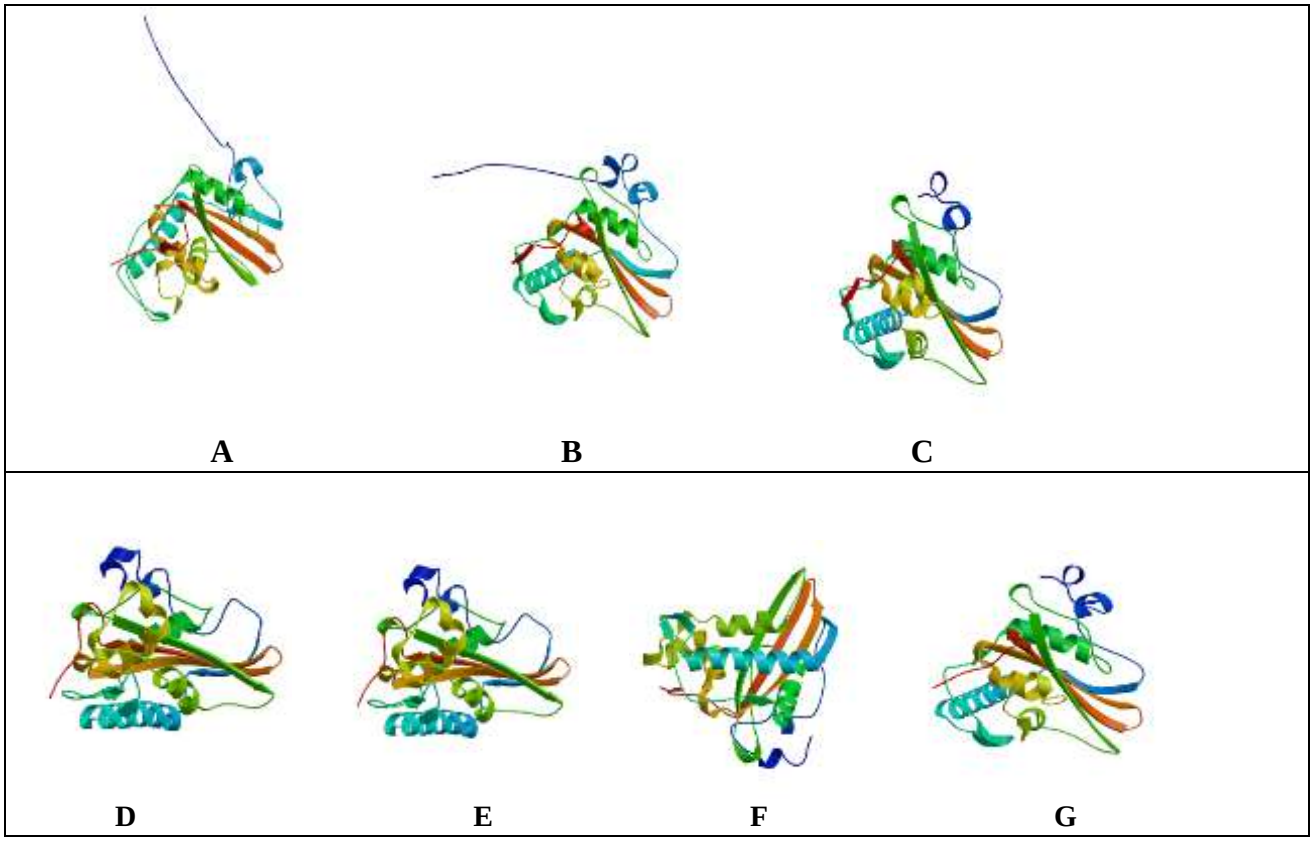
Table 6. Protein Disorder Prediction of eleven wasp venom allergens.

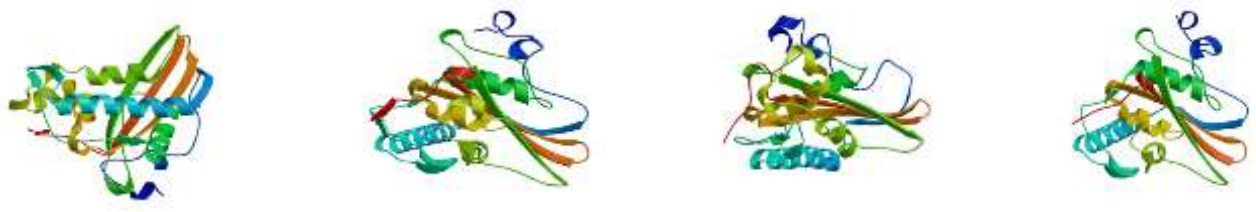
Sl. No.	Name of Toxin	Toxin Code	No. of predicted residue	No. of residues disordered	Numbered Disordered region	Longest Disordered Region	Overall percent disorder	Prediction disordered segment
1.	M-ectatotoxin-Eb2a (M-ECTX-Eb2a)	VA5_MICHY	232	4	1	4	1.72	[1]-[4]
2.	M-ectatotoxin-Eb2b (M-ECTX-Eb2b)	VAL5_VESMC	227	4	1	4	1.76	[1]-[4]
3.	M-ectatotoxin-Eb2c (M-ECTX-Eb2c)	VA5_VESMC	204	0	0	0	0	0

4.	U1-poneritoxin-Na1a (U1-PONTX-Na1a)	VA5_POLPI	207	0	0	0	0	0
5.	U1-poneritoxin-Na1b (U1-PONTX-Na1b)	VA52_POLPI	206	0	0	0	0	0
6.	U1-poneritoxin-Na3a (U1-PONTX-Na3a)	VA5_POLAN	209	2	1	2	0.96	[1]-[2]
7.	U1-poneritoxin-Na3b (U1-PONTX-Na3b)	VA5_VESSQ	205	0	0	0	0	0
8.	M-poneritoxin-Ng1a (M-PONTX-Ng1a)	VA5_POLFU	205	0	0	0	0	0
9.	M-poneritoxin-Ng1c (M-PONTX-Ng1c)	VA5_VESFL	204	0	0	0	0	0
10.	M-poneritoxin-Ng1d (M-PONTX-Ng1d)	VA5_POLSR	207	0	0	0	0	0
11.	M-poneritoxin-Ng1e (M-PONTX-Ng1e)	VA3_DINQU	2	2	2	0.99	[201]-[202]	2

Table 7. Protein Toxicity Prediction of eleven wasp venom allergens.

Sl. No.	Toxin Code	Score	Toxic domain name	Toxic domain position	ToxDL remarks
1.	VA5_MICHY	0.016260436	No toxic domain is predicted	NA	Being toxic protein
2.	VAL5_VESMC	0.01084075	No toxic domain is predicted	NA	Being toxic protein
3.	VA5_VESMC	0.0054884884	No toxic domain is predicted	NA	Being toxic protein
4.	VA5_POLPI	0.0051484983	No toxic domain is predicted	NA	Being toxic protein
5.	VA5_POLPI	0.0039935564	No toxic domain is predicted	NA	Being toxic protein
6.	VA5_POLAN	0.006213264	No toxic domain is predicted	NA	Being toxic protein
7.	VA5_VESSQ	0.020850359	No toxic domain is predicted	NA	Being toxic protein
8.	VA5_POLFU	0.009197847	No toxic domain is predicted	NA	Being toxic protein
9.	VA5_VESFL	0.011403729	No toxic domain is predicted	NA	Being toxic protein
10.	VA5_POLSR	0.0043708426	No toxic domain is predicted	NA	Being toxic protein
11.	VA3_DINQU	0.0039935564	No toxic domain is predicted	NA	Being toxic protein





H **I** **J** **K**

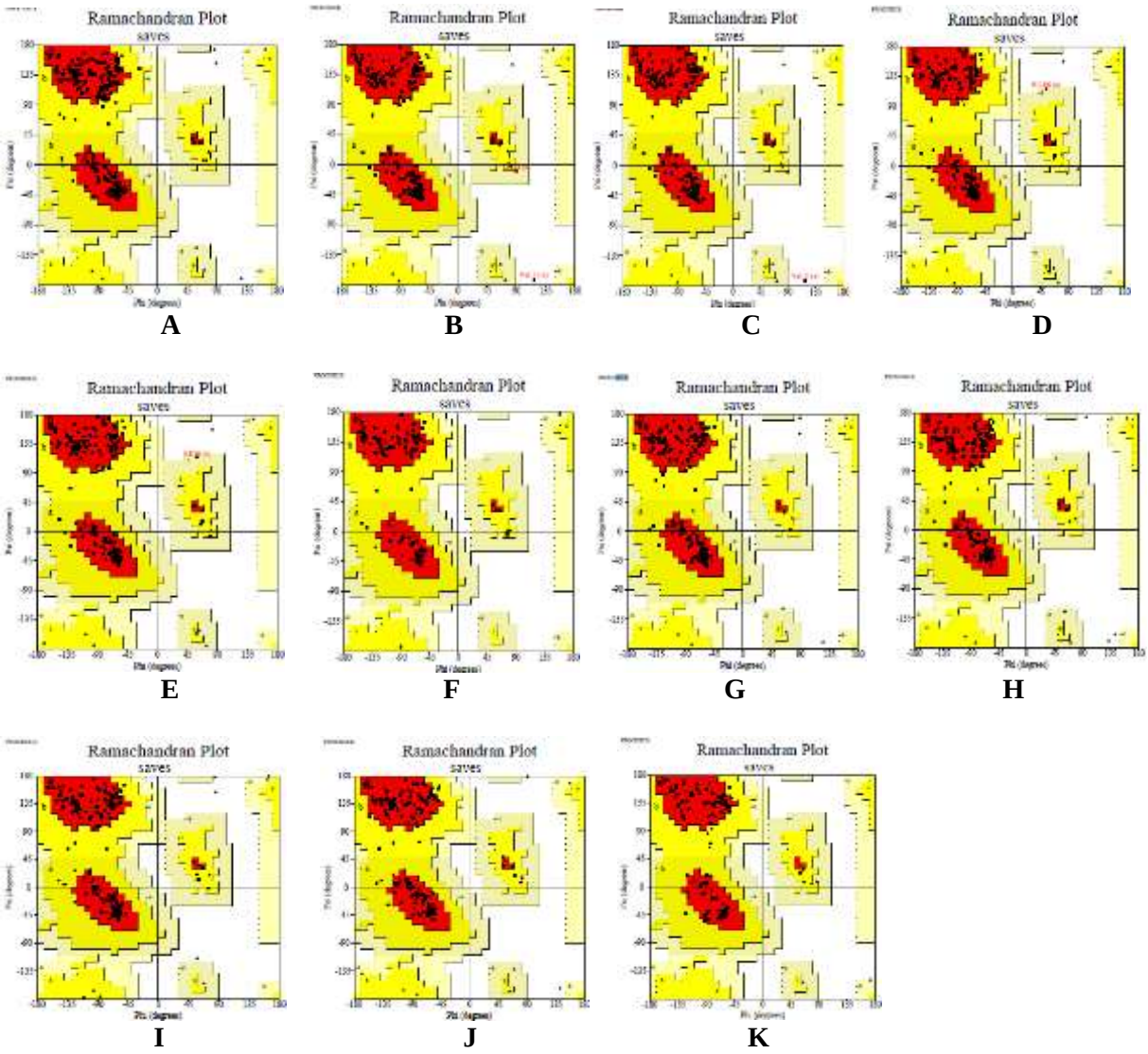
A: VA5_MICHY **B: VAL5_VESMC** **C: VA5_VESMC**

D: VA5_POLPI **E: VA52_POLPI** **F: VA5_POLAN**

G: VA5_VESSQ **H: VA5_POLFU** **I: VA5_VESFL**

J: VA5_POLSR **K: VA3_DINQU**

Fig. 2. Homology Models of eleven wasp venom allergens.



A: VA5_MICHY **B: VAL5_VESMC** **C: VA5_VESMC**

D: VA5_POLPI **E: VA52_POLPI** **F: VA5_POLAN**

G: VA5_VESSQ H: VA5_POLFU I: VA5_VESFL

J: VA5_POLSR K: VA3_DINQU

Fig 3. Ramachandran plot of eleven wasp venom allergens.

Table 7a. Homology Model validation of eleven wasp venom allergens by ANOLEA,ProSA, QMEAN & VERIFY 3D

Sl. No.	Source	Name of Toxin	Toxin code	ANOLEA Z-Score	ProSA Z-Score	QMEAN Z-Score	VERIFY 3D	
							Percentage of residues had an average 3D-ID>=0.1	Verify Status
1.	Venom allergen 5 (Antigen5) (Cysteine-rich venom protein)	*Microctonus hyperodae*	VA5_MICHY	0.26	-6.07	-1.49	83.19	Pass
2.	Venom allergen 5 (Antigen 5)	*Vespula maculifrons*	VAL5_VESMC	0.61	-6.46	-1.04	86.78	Pass
3.	Venom allergen 5 (Allergen Ves m V)	*Vespula maculifrons*	VA5_VESMC	0.06	-6.86	-0.52	98.04	Pass
4.	Venom allergen 5 (Antigen 5) (Ag5)	*Polybia paulista*	VA5_POLPI	0.61	-6.86	-0.75	81.64	Pass
5.	Venom allergen 5 2 (Antigen 5)	*Polybia paulista*	VA52_POLPI	0.61	-6.86	-0.75	81.64	Pass
6.	Venom allergen 5 (Allergen Pol a V) (Antigen 5) (Ag5)	*Polistes annularis*	VA5_POLAN	0.19	-6.9	0.28	80.38	Pass

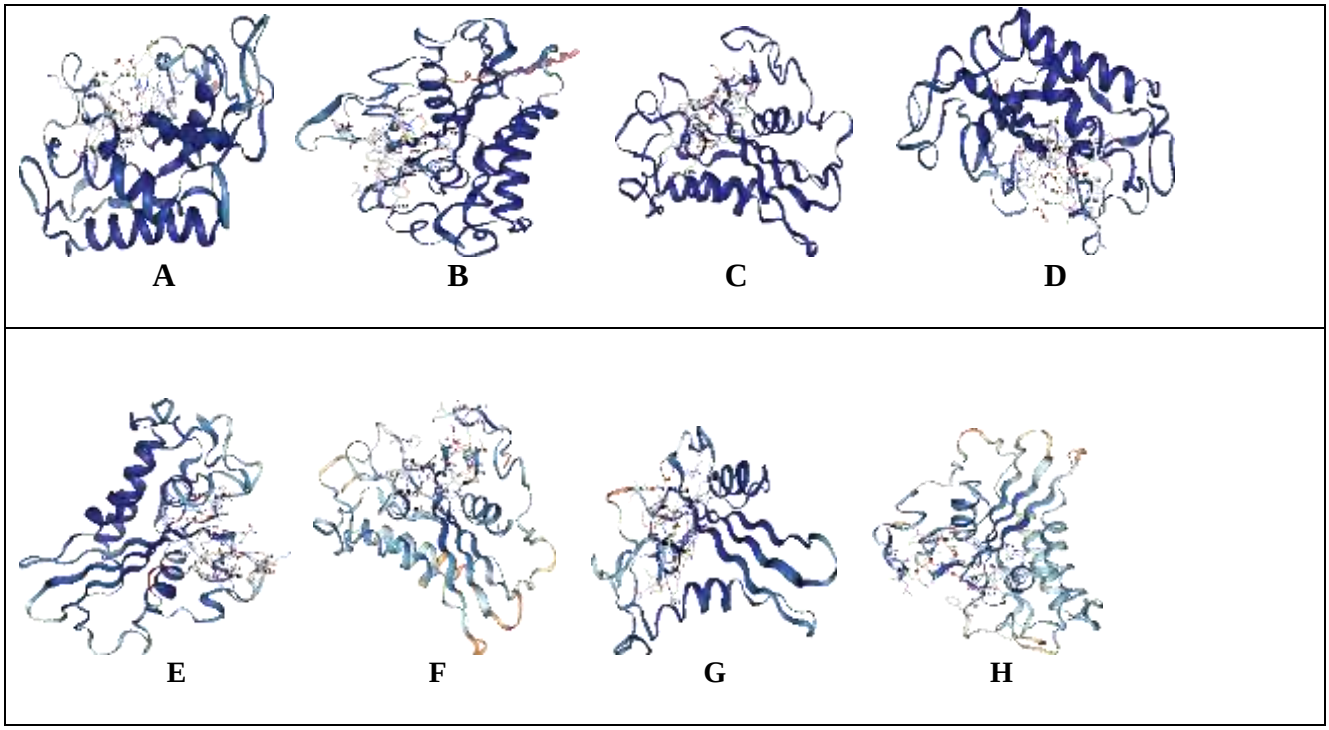
7.	Venom allergen 5 (Allergen Ves s V) (Antigen 5)	*Vespula squamosa*	VA5_VESSQ	0.23	-6.48	-0.8	96.59	Pass
8.	Venom allergen 5 (Allergen Pol f V) (Antigen 5)	*Polistes fuscatus*	VA5_POLFU	0.29	-6.64	-0.17	80	Pass
9.	Venom allergen 5	*Vespula flavopilosa*	VA5_VESFL	0.01	-6.85	-0.7	97.55	Pass
10.	Venom allergen 5	*Polybia scutellaris rioplatensis*	VA5_POLSR	0.6	-6.94	-0.61	84.06	Pass
11.	Venom allergen 3 homolog	*Dinoponera quadriceps*	VA3_DINQU	0.61	-6.73	-0.74	93.53	Pass

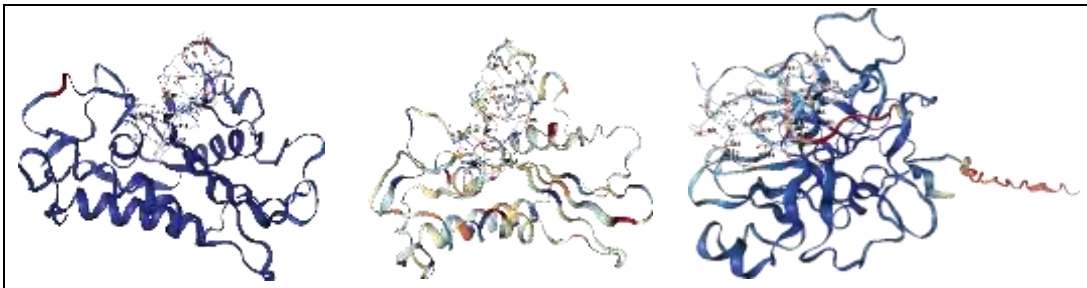
Table 7b. Homology Model validation of eleven wasp venom allergens by Ramachandran Plot.

PROCHECK RESULTS								
Sl. No.	**Source**	**Toxin name**	**Toxin code**	**Residues in most Favored Region (%)**	**Residues in additional allowed region (%)**	**Residues in generously allowed region (%)**	**Residues in disallowed region (%)**	**Model quality**
1.	Venom allergen 5 (Antigen5) (Cysteine-rich venom protein)	*Microctonus hyperodae*	VA5_MICHY	89.2	10.8	0	0	Good
2.	Venom allergen 5 (Antigen 5)	*Vespula maculifrons*	VAL5_VESMC	93.3	6.7	0	0	Very Good

3.	Venom allergen 5 (Allergen Ves m V)	*Vespula maculifrons*	VA5_VESMC	94.0	5.5	0	0.5	Very Good
4.	Venom allergen 5 (Antigen 5) (Ag5)	*Polybia paulista*	VA5_POLPI	91.6	8.4	0	0	Very Good
5.	Venom allergen 5 2 (Antigen 5)	*Polybia paulista*	VA52_POLPI	92.7	7.3	0	0	Very Good
6.	Venom allergen 5 (Allergen Pol a V) (Antigen 5) (Ag5)	*Polistes annularis*	VA5_POLAN	90.3	9.7	0	0	Very Good
7.	Venom allergen 5 (Allergen Ves s V) (Antigen 5)	*Vespula squamosa*	VA5_VESSQ	93.3	6.7	0	0	Very Good
8.	Venom allergen 5 (Allergen Pol f V) (Antigen 5)	*Polistes fuscatus*	VA5_POLFU	89.9	9.6	0	0.6	Good
9.	Venom allergen 5			90.1	9.6	0.3	0	Very Good

		Vespula flavopilosa	VA5_VESFL					
10.	Venom allergen 5	*Polybia scutellaris rioplatensis*	VA5_POLSR	90.2	9.2	0.6	0	Very Good
11.	Venom allergen 3 homolog	*Dinoponera quadriceps*	VA3_DINQU	81.8	16.9	0.8	0.4	Good





I **J** **K**

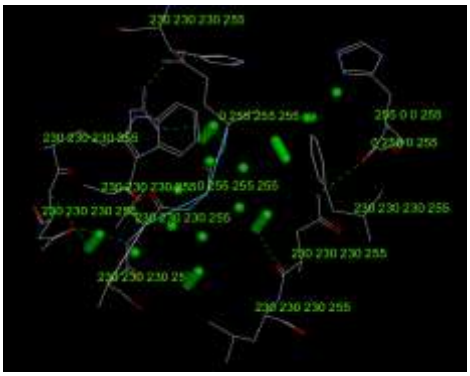
A: VA5_MICHY **B: VAL5_VESMC** **C: VA5_VESMC**

D: VA5_POLPI **E: VA52_POLPI** **F: VA5_POLAN**

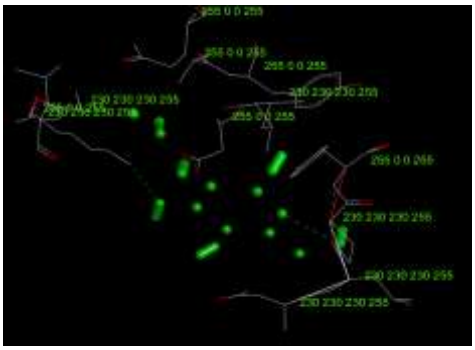
G: VA5_VESSQ **H: VA5_POLFU** **I: VA5_VESFL**

J: VA5_POLSR **K: VA3_DINQU**

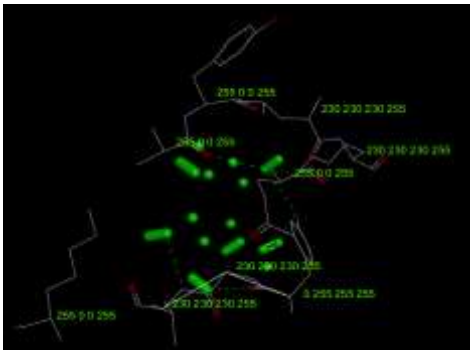
Fig 4a. Docking Models of eleven wasp venom allergens with N-acetylglucosamine docked through CB DOCK software tool.



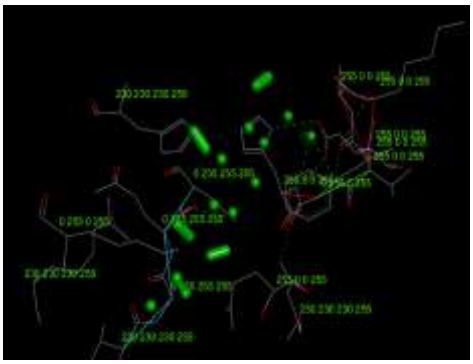
A



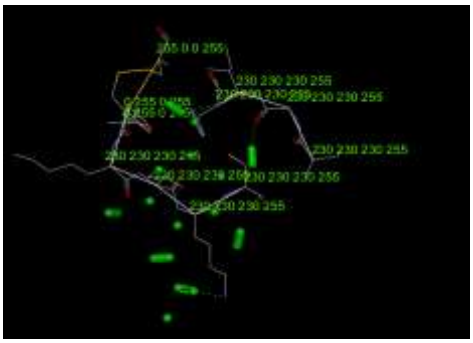
B



C



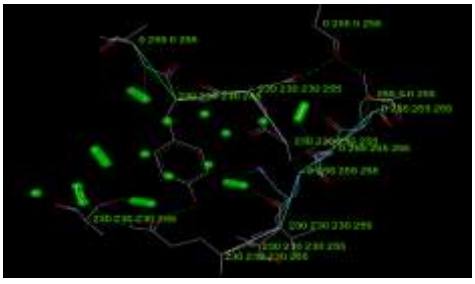
D



E

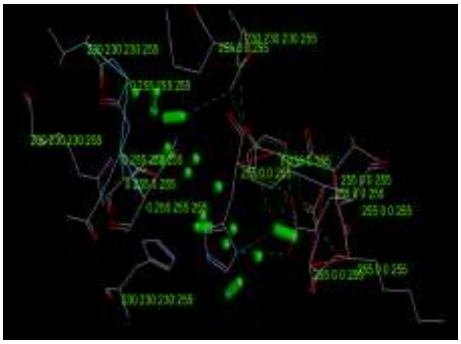
F

G

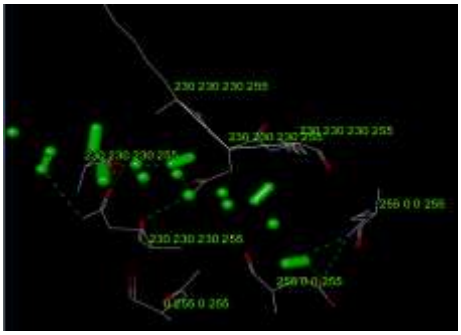


H

I



J



K

A: VA5_MICHY **B: VAL5_VESMC** **C: VA5_VESMC**

D: VA5_POLPI **E: VA52_POLPI** **F: VA5_POLAN**

G: VA5_VESSQ **H: VA5_POLFU** **I: VA5_VESFL**

J: VA5_POLSR **K: VA3_DINQU**

Fig 4b. Docking models visualised by Bio via Discovery Studio.

Table 8. Results of docking of eleven wasp venom allergens with N-acetylglucosamine by CB DOCK webserver.

Sl. No.	Name of Toxin	Source	Toxin Code	Ligand	Vina Score	Cavity Size	Binding Affinity (kcal/mol)	Remarks
1.	Venom allergen 5 (Antigen5) (Cysteine-rich venom protein)	*Microctonus hyperodae*	VA5_MICHY	N-acetylglucosamine	-8	596	-8	Docking Valid

2.	Venom allergen 5 (Antigen 5)	*Vespula maculifrons*	VAL5_VESMC	N-acetylglucosamine	-7.6	245	-7.6	Docking Valid
3.	Venom allergen 5 (Allergen Ves m V)	*Vespula maculifrons*	VA5_VESMC	N-acetylglucosamine	-7.2	210	-7.2	Docking Valid
4.	Venom allergen 5 (Antigen 5) (Ag5)	*Polybia paulista*	VA5_POLPI	N-acetylglucosamine	-7.7	526	-7.7	Docking Valid
5.	Venom allergen 5 2 (Antigen 5)	*Polybia paulista*	VA52_POLPI	N-acetylglucosamine	-7.7	630	-7.7	Docking Valid
6.	Venom allergen 5 (Allergen Pol a V) (Antigen 5) (Ag5)	*Polistes annularis*	VA5_POLAN	N-acetylglucosamine	-7.3	202	-7.3	Docking Valid
7.	Venom allergen 5 (Allergen Ves s V) (Antigen 5)	*Vespula squamosa*	VA5_VESSQ	N-acetylglucosamine	-7.4	638	-7.4	Docking Valid
8.	Venom allergen 5 (Allergen Pol f V) (Antigen 5)	*Polistes fuscatus*	VA5_POLFU	N-acetylglucosamine	-7.4	207	-7.4	Docking Valid
9.	Venom allergen 5		VA5_VESFL	N-acetylglucosamine	-7.3	249	-7.3	Docking Valid

		Vespula flavopilosa						
10.	Venom allergen 5	*Polybia scutellaris rioplatensis*	VA5_POLSR	N-acetylglucosamine	-7.7	621	-7.7	Docking Valid
11.	Venom allergen 3 homolog	*Dinoponera quadriceps*	VA3_DINQU	N-acetylglucosamine	-7.4	418	-74	Docking Valid

3.2. DISCUSSION:

3.2.1 Basic description of wasps' eleven venom allergens and bacterial cell wall component N-acetylglucosamine:

Table 1a, provides the details of different wasp venom allergen proteins, their source from various wasp species, primary sequences, Uniprot ID (www.expasy.org/sprot), a code for each venom allergen and peptide length. It *is* observed that, minimum peptide sequence length 203 was observed in VA5_MICHY while maximum peptide sequence length 232 was observed for VA5_VESMC.**Table 1b,** shows the 2D and 3D structures of N-acetylglucosamine (https://pubchem.ncbi.nlm.nih.gov).

3.2.2 Physico-chemical parameters of wasps'eleven venom allergens:

Table 2, shows various physico-chemical data of these venom allergen molecules that are predicted through ProtParam tool. These properties include peptide sequence length, molecular weight, isoelectric point (pI), total number of negatively (-R) and positively charged residues (+R), Extinction coefficient, Instability Index (II), Aliphatic Index (AI) and Grand Average of Hydropathicity (GRAVY) for the 11 wasp species. Theoretical pI values range from 6.2178 to 9.3898 indicating that all the venom allergen molecules are on the basic side of the scale except VA5_MICHY which is in the acidic side. Local subcellular formation, interaction and melting depends on the isoelectric zone and the number of well-charged and negatively charged residues [34]. PI is the pH value at which the protein is free or the amount of negative and positive costs are equal. Negatively charged residues are found in VA5_MICHY while highly charged negative residues are found inVA5_VESFL. Proteins with an index of instability (II) <40 are considered stable and those with a value of II > 40 are called unstable [35]. Here the Instability Index values range from 17.94 to 42.01 which means that all the proteins may be stable except VA5_VESFL would be unstable in the test tube as a value of less than 40 is stable in the test tube. Here the low levels of coagulation in the extinction are indicators of the low absorption capacity of the protein under study. Almost all of the global proteins contain large numbers of α-helices and β-sheets / fibres folded into a composite structure stabilized by both polar and lightweight interactions [36]. Aliphatic index plays a role in the thermal stability of proteins. Proteins having high Aliphatic index are stable in temperature. The Aliphatic index of cytotoxins in the range 61.13 to 73.43 showed that these proteins are stable in temperature. With the disruption of the biological membrane this is an important indicator of the molecule [37]. High values of Aliphatic index indicate that the protein is thermo-stable over a wide range of temperatures. Here it ranges from 61.13 to 73.43 indicating a wide range of temperature. Good GRAVY values refer to hydrophobicity; negative values refer to hydrophilicity. Here it ranges from -0.833 to -0.425, which means that decreasing negative points indicates the range of hydrophilicity of the proteins being studied.

Table 3, portrays *in silico* predictions of the secondary structures of 11 venom allergens from various wasp species with the help of GOR4 tools depicting the highest α-helix, and random coils followed by β-strand displays limited amount of dynamic stability. Secondary structure predicts that approximately or more than 30% of the secondary structure is made up of alpha helix. This points to the fact that protein toxins are naturally occurring globular [38].
Fig. 1, reveals a cartoon diagram of the comparative secondary structure of 11 venom allergen molecules from wasp. α helix extensions were found at the beginning of the venom allergen protein sequences. Homology modelling is useful when the model protein (in the known sequence and unknown structure) is related to at least one single protein with both known sequence and known structure [39].
Fig. 2, shows homology models predicted using SWISS-MODEL software tool. Z-score of all modelled proteins obtained within acceptable scores that point towards a native protein structure pointing towards a good model quality. In the QMEAN Z-score represents a measure of how the model can be compared to test-based structures of the same size (https://tshi.page/ox/notes/techniques.html). Ramachandran plots generated for all venom allergen molecules through PROCHECK tool also pointed towards a validated and acceptable Homology model produced through Swiss-Model tool.

3.2.3 Cysteine-Cysteine bond prediction of eleven wasp venom allergens:

Table 4, shows the cysteine-cysteine binding prediction using Scratch Protein Predictor webserver. This demonstrates the stability of the binding site with any ligand or other type of molecule, structural relationships - protein activity. The toxins under study, most of them carry 03 pairs of Cysteine in bonded cases, which is a signature of wasps' venom allergen.

3.2.4 Protein Family/Domain Prediction of eleven wasp venom allergens:

Table 5, showing 11 wasp venom allergen proteins family prediction using PROSITE, InterPro, Pfam and CDD software tools. Results reveals that all of the eleven wasp venom allergens. Protein intrinsic disorder is increasingly being identified in proteomics studies. While lack of structure, many disturbance regions have been associated with biological activity.

3.2.5 Protein Disorder Prediction of eleven wasp venom allergens:

Table 6, reveals the disordered regions predicted by the PrDos software tool for the 11 venom allergen molecules. Of the 11 wasp venom allergens the highest disorderedness is found in VA5_MICHY & VA5_VESSQ with 04 numbers of residues in the disordered state [40].

3.2.6 Predicted Protein Toxicity of eleven wasps venom allergens:

Table 7, represents Protein Toxicity of 11 wasp venom allergen proteins molecules Prediction by Tox DL Scores and Toxic domain positions obtained are as follows: Highest toxicity score is 0.20850359 in VA5_VESSQ and lowest toxicity score is 0.0039935564 for VA52_POLPI. All venom allergen molecules exhibit a predicted Toxicity score less than 01. This is indicative of low toxicity level. No toxic domain was predicted. Highest toxicity marked VA5_VESSQ with 0.20850359 which is less than 01.

3.2.7 Homology models of eleven wasps' venom allergen toxins:

Fig. 2, shows Homology Models of 11 wasp venom allergen proteins by SWISS-MODEL webserver where the 3D protein model is automatically generated by first transferring conserved atom coordinates as defined by the target-template alignment [41].
Table 8, shows Homology model validation scores predicted through ANOLEA, Pro SA, QMEAN 4 and VERIFY 3D tools with Z-score analysis showing that 3D models are viable and acceptable.
Fig. 3 (A-K), exhibits the Ramachandran plots of 11 wasps' venom allergen molecules generated through PROCHECK software. In Ramachandran plot core or allowed regions are the areas that show preferred regions for Ramachandran plot displays the ψ and φ angle pairs for residues in the protein. PROCHECK results clearly showed that around 90 to 94 percent of the amino acid residues for 11 wasp venom allergen molecules are in the core i.e., mostly favoured regions. This confirmed the reliability of the model [42].

3.2.8 Docking results of 11 wasp venom allergen proteins with bacterial cell wall component N-acetylglucosamine:

Fig. 4a & 4b as well as **Table 9**, portrays Docking results reflecting Vina Score, Cavity Size, Binding Affinity generated through the CB DOCK webserver showing the docked models and the results of high negative affinity values obtained from docking of 11 wasp venom allergens with bacterial cell wall component N-

acetylglucosamine indicate strong binding interaction as well as stable target-venom allergen binding interaction. Highest negative affinity score of -8 and with lowest vina score -8 was observed with the venom allergen 5 with Toxin Code VA5_MICHY.

4. CONCLUSION:

In silico analysis of 11 wasp venom allergen proteins were carried out in details. The homology modelled venom allergen molecules docked with bacterial cell wall component N-acetylglucosamine strongly. These wasp venom allergens thus can have a potential use as an anti-bacterial agent.

5. ACKNOWLEDGEMENT:

The Authors acknowledge Jharkhand Rai University, Ranchi Authority for support provided to the them for the present research. SD acknowledges the Education Department, Govt. of West Bengal for providing moral support and encouragement to carry out the work.

6. CONFLICT OF INTEREST STATEMENT:

The authors declared that there is no conflict of any sort in relation to the paper.

7. REFERENCES:

1. King G. F, 2011.Venoms as a platform for human drugs: Translating toxins into therapeutics. Expert Opin Biol Ther.,11(11):1469–1484.
2. Bazon M.L.,A.Perez Riverol, J.R.A Dos Santos Pinto, L.G.R. Fernandes, A.M. Lasa, D.L. Justo Jacomini, *et al*., 2017. Heterologous expression, purification and immunoreactivity of the antigen-5 from *Polybia paulista* wasp venom. Toxins (Basel),9pii: E259.

3.Cociancich S, O.A. Ghazi, C. Hetru, A. Jules, S.N. Hoffmann & L. Letelliers, 1993. Insect Defensin, an Inducible Antibacterial Peptide, Voltage-dependent Channels in Micrococcus luteus. *The Journal of Biochemistry*, 19239-19245.

4.Bulet, P., C. Hetru, D. Jean-Luc & D. Hoffmann, 1999. Antimicrobial peptides in insects, structure and function. *Developmental & Comparative Immunology*,23(4–5):329-344.

5.Palma, M.S., 2011. Peptides as toxins/defensins. *Amino Acids*, 40:1-4.

6.Bazon, M.L., L.H.Silverid, P.U.Simioni, M.R.Brochetto-Braga,2018. Current Advances in Immunological Studies on the Vespidae Venom Antigen 5: Therapeutic and Prophylaxis to Hypersensitivity Responses. *Toxins*, *10*(8), 305.

7.Casteels-Josson, K., W. Zhang, T. Capaci, P. Casteels & P. Tempst, 1994. Acute transcriptional response of the honey bees peptide-*antibiotics gene repertoire, required posttranslational conversion of the precursor structures. J. Biol. Chem*, 269: 28569-28575.

8.Bulet, P. & R. Stocklin,2005. Insect antimicrobial peptides: structure, properties and gene regulation. *Protein & Peptide Letters*,12: 3-11.

9.Chapot-Chartier, M. & S.Kulakauskas,2013. Cell wall structure and function in lactic acid bacteria**.***Microbial Cell Factories* volume 13, Article number: S9 (2014).

9.Henriksen, A., T.P.King, O.Mirza et al,2001.Major venom allergen of yellow jackets, Ves v 5: Structural characterization of a pathogenesis-related protein superfamily. Proteins,45:438–448.

10.Mohammad, R.F.,2015. 'In silico structural analysis, physicochemical characterization and homology modelling of arabidopsis thaliana Na+/H+ exchanger 1 (ATNHX1) protein. B.S. Thesis. BRAC University, p.96.

11.Prajapati,K.K. & R.K. Upadhyay, 2021.Bees and wasps venom toxins, its immune-allergic responses, diagnosis and therapeutics. *International Journal of Pharmacy and Pharmaceutical Sciences*, vol. 13, no. 1, pp. 1-9.

12.Obando, M.A., T.dorr ,2023.Novel role for peptidoglycan carboxypeptidases in maintaining the balance between bacterial cell wall synthesis and degradation. bioRxiv[Preprint],2023.07.12.548665. [Version 1].

13.Lu, G., M**.** Villalba**,**M. R. Coscia,D.R. Hoffman,T.P. King,1993. Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets. J Immunol, 150 (7): 2823–2830.

14.Yang Liu, M., Grimm, +3 authors C., Yang,2019. CB-Dock: a web server for cavity detection-guided protein–ligand blind docking.Acta Pharmacologica Sinica. 41(Suppl 14), DOI:10.1038/s41401-019-0228-6
15.de Magalhães, C.S., M. Almeida, H.J.C. Barbosa & L.E. Dardenne,2014 . A dynamic niching genetic algorithm strategy for docking highly flexible ligands. *Information Sciences* ,289: 206–224.
16.Guedes, I.A., A.M.S. Barreto, D. Marinho, E. Krempser, M.A. Kuenemann & O. Sperandio, et al,2021. New Machine Learning and Physics-Based Scoring Functions for Drug Discovery. *Sci Rep*., 11(3198).
17.Santos Ricardo, N., L.G.L. Ferreira & D.A. Andricopulo,2018. Practices in Molecular Docking and Structure-Based Virtual Screening. *Methods in molecular biology*, 1762:31-50.
18.Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel & A. Bairoch,2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. ,31:3784-3788.
19.Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel & A. Bairoch,2005. *Protein Identification and Analysis Tools on the ExPASy Server*. In: Walker JM, editor: The Proteomics Protocols Handbook, Humana Press,p. 571-607.
20.Moffat, L. & D.T. Jones,2021. Increasing the Accuracy of Single Sequence Prediction Methods Using a Deep Semi-Supervised Learning Framework. Bioinformatics,37(21):3744–51.
21.Garnier, J., D.J. Osguthorpe & Robson, B. Robson,1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120, 97-120.
22.Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello & R. Gumienny, et al.,2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*., 46: W296-W303.
23.Wiederstein, S.,2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids* Research,35: W407-W410.
24.Benkert, P., M. Biasini & T. Schwede,2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27: 343-350.
25.Jain, C.K., M. Gupta, Y. Prasad, G. Wadhwa & S. Sharma,2014. Homology modeling and protein engineering of alkane monooxygenase in *Burkholderia thailandensis* MSMB121: In silico insights. *Journal of molecular modelling*, 20: 2340.
26.Fariselli, P., P. Riccobelli & R. Casadio, 1999. The role of evolutionary information in predicting the disulfide bonding state of cysteines in proteins. *Proteins*,36: 340-346.
27.Yang, J., Bao-Ji. He, R. Jang, Y. Zhang & H. Shen, 2015. Accurate disulfide-bonding network predictions improve ab initio structure prediction of cysteine-rich proteins, *Bioinformatics*,31: 3773-3781.
28.Blum, M., H.Chang, S. Chuguransky, T. Grego, S. Kandasaamy & A. Mitchell, et al, 2020. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*,49(D1): D344–D354.
29.Marchler-Bauer, A., J.B. Anderson, P.F. Cherukuri, C. DeWeese-Scott, L.Y. Geer & M. Gwadz, et al,2005. *Nucleic Acids Res*,33(Database issue).
30.Pan ,X., J. Zuallaert, X. Wang, H.B. Shen, E.P. Campos & D.O. Marushchak ,et al, 2021. ToxDL: deep learning using primary structure and domain embeddings for assessing protein toxicity. *Bioinformatics*,36(21):5159-5168.
31.Ishida ,T. & K. Kinoshita, 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res*,35(2): W460–W464.
32.National Center for Biotechnology Information. PubChem Compound Summary for CID 1738118,1,4-beta-D-Glucan. https://pubchem.ncbi.nlm.nih.gov/compound/N-acwtylglucosamine. Accessed Mar. 5, 2024.
33.Guedes, A.I., C.S. de Magalhães & L.E. Dardenne,2014. Receptor–ligand molecular docking, *Biophysical Reviews*,6(1): 75–87.
34.Kaur, A., P.K. Pati, A.M. Pati & A.K. Nagpal, 2020. Physico-chemical characterization and topological analysis of pathogenesis-related proteins from *Arabidopsis thaliana* and *Oryza sativa* using *in-silico* approaches. *PLoS* ONE,15(9): e0239836.
35.Panda,S. & G. Chandra, 2012. Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates. *Bioinformatio*n,*8*(18):891–896.
36.Guruprasad, K., B.V.B. Reddy & M.W. Pandit, 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Enginee*ring, 4: 155–161.
37.Shen, C.H., 2019. Chapter 4 - Gene Expression: Translation of the Genetic Code,In: Shen CH, editor. Diagnostic Molecular Biology, Academic Press,87-116.
38.Gupte, T.M., M. Ritt & S. Sivaramakrishnan, 2021. Chapter Seven - ER/K-link—Leveraging a native protein linker to probe dynamic cellular interactions, In: Maarten M, editor. Methods in Enzymology, Academic Press, 64:173-208.

39.Robinson, S.W., A.M. Afzal & D.P. Leader, 2014. Chapter 13 - Bioinformatics: Concepts, Methods, and Data,In: Padmanabhan S, editor. Handbook of Pharmacogenomics and Stratified Medicine, Academic Press, 259-287.
40.Xue, B., A.K. Dunker & V.N. Uversky, 2010. Retro-MoRFs: identifying protein binding sites by normal and reverse alignment and intrinsic disorder prediction. *Int J Mol Sci*., 11(10):3725-3747.
41.Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello & R. Gumienny, 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46: W296–W303.
42.Ramachandran, G.N., C. Ramakrishnan & V. Sasisekharan, 1963. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*,**7**: 95–9.