Research Article ISSN 2689-1050

Chemical & Pharmaceutical Research

Tarantula Toxin Transcripts and Toxin Structure Prediction by AlphaFold

Christine Vega, Rebekkah Sepulveda and Ying Jia*

Department of Biology, The University of Texas Rio Grande Valley, Brownsville, Texas 78520, USA.

*Correspondence:
Ying Jia, Department of Biology, The University of Texas Rio Grande Valley, Brownsville, Texas 78520, USA.

Received: 13 Jan 2024; Accepted: 01 Mar 2024; Published: 07 Mar 2024

Citation: Christine Vega, Rebekkah Sepulveda, Ying Jia. Tarantula Toxin Transcripts and Toxin Structure Prediction by AlphaFold. Chem Pharm Res. 2024; 6(2): 1-7.

ABSTRACT

Tarantula venom predominantly comprises peptides and small proteins (venom toxins). To explore this further, we generated a combined gland cDNA library from five different tarantula species. Subsequently we randomly sequenced 752 clones, which led to the identification of sixteen distinct toxin groups, including one novel group. Each group consists of a minimum of three identical sequences. A comprehensive analysis of these toxin groups was carried out, and we simulated the 3D structures of certain predicted toxins. These transcripts can be used to the production of venom toxins through recombinant techniques, offering promising applications in various biomedical applications.

Keywords

Tarantula, cDNA library, Toxin transcript, AlphaFold.

Introduction

After undergoing eons of natural selection, animal venom toxins (proteins and peptides) have evolved to target specific "channel" proteins in the nervous system of prey organisms. This enables venomous animals, such as tarantulas, to use their venom for rapid paralysis and prey capture. Due to their specificity and selectivity, animal venom has been utilized as a tool to study the functions of specific proteins, such as the gating cycles of "channel" proteins. They have also served as templates for developing drugs to combat human diseases such as chronic pain. One of the most successful examples is a snake venom component termed three-finger toxin (3FTx) (e, g., α-Bungarotoxin) that was discovered from the venom of *Bungarus multicinctus* over sixty years ago [1] and is still the most widely used and valuable inhibitor in deciphering the different gating cycles of nicotinic acetylcholine receptors (nAChRs) [2-4]. Additionally, Agwa, et al. [5,6] developed an HwTx-IV analogue that improved *in vitro* selectivity for the pain target $Na_v1.7$ and with an *in vivo* efficacy similar to that of native HwTx-IV purified from tarantula (*Haplopelma schmidti*). It is noteworthy that the venom of a single tarantula species may contain hundreds of active toxins [7,8], and with approximately 1,067 tarantula species in the world [9], there is an almost limitless supply of tarantula toxins for biological research and biomedical applications. However, compared to snakes, tarantulas have relatively small bodies that makes extracting sufficient crude venom, especially the low-expressed venom components challenging when investigating their potential biomedical applications. Fortunately, cloning venom toxin transcripts offers a solution. Upon cloning, these transcripts can be used indefinitely to recombinantly produce a large quantity of a specific toxin, making it a promising avenue for further research and drug development. In this work, we generated a cDNA library by pooling glands from five tarantula species, sequenced 752 cDNA clones and identified 16 distinct toxin groups. The 3D structures of selected toxins were also simulated. Additionally, to compare with tarantula toxin transcripts, we also generated a snake (*Crotalus atrox*) gland cDNA library. The cloned toxin transcripts and their predicted toxin structures provide a valuable resource for future biomedical applications.

Materials and Methods

Five tarantulas (*Aphonopelma hentzi*; *Brachypelma smithi*; *Chromatopelma cyaneopubescens*; *Nhandu coloratovillosus*; *Grammostola rosea*), with a starting length of 2 cm, were purchased from a tarantula farmer and raised for 2 years until they reached a length of 4-7 cm (Figure 1). They were then euthanized in accordance with the University of Texas Rio Grande Valley (UTRGV) Institutional Animal Care and Use Committee (IACUC)

protocol (ref.no. AUP-18-08). Additionally, one adult snake (*Crotalus atrox*) (Figure 1) was purchased from a snake farmer and sacrificed in accordance with the same protocol (AUP-18-08). The venom glands were carefully dissected, placed in nuclease-free tubes, and promptly flash-frozen in liquid nitrogen to preserve their RNA integrity. Total RNA was then extracted using TRIzol reagents (ThermoFisher Scientific, CA) and subsequently reverse-transcribed into cDNA. The resulting total venom gland cDNA was utilized to construct cDNA libraries with the In-Fusion SMARTer Directional cDNA library construction kit (Takara Bio USA, Inc. CA). We selected 752 clones at random from tarantula gland cDNA library and extracted plasmid DNAs from each clone. Plasmid DNAs were sequenced at MCL (Molecular Cloning Laboratory, CA), and readable sequences were individually subjected to Protein-BLAST search against non-redundant protein sequences (nr) in the NCBI database. Only mature transcripts, which contain start and stop codons and polyadenylation signal sequence, were considered for further analysis; these mature transcripts were then translated into amino acid sequences using Expasy-translate. Multiple sequence alignment (MSA) and phylogenetic analysis were carried out to identify unique toxin groups. A unique toxin group was defined as one that contained at least three identical sequences with over 95% coverage and 70% identity with other members. Clustal Omega [10] was used to perform MSA, while Bayesian Evolutionary Analysis Sampling Trees (BEAST) [11] was used to generate the phylogenetic tree of Bayesian inference (BI). The 3D structures of toxins were simulated using AlphaFold2 [12-13] by accessing the Texas Advanced Computing Center (TACC, UT-Austin, TX), and further assessed by analyzing their Ramachandran angles (φ and ψ) using SAVESv6.0 (https://saves.mbi.ucla.edu/).

Results and Discussion

cDNA library of mixed glands of tarantulas

A directional full-length cDNA library with a titer of 1.1 x 10^7 cfu/mL was successfully constructed using mixed venom glands from five tarantulas. We randomly selected 46 clones for examining their inserts and their lengths using primers located on library vector and PCR technique; of the 46 clones examined, 45 clones possessed the inserts. The cDNA had an average length of approximately 750 bp, ranging from 250 bp to 2 kb (Figure 2). We further selected 752 cDNA clones at random from the library and individually isolated their recombinant DNAs for sequencing. Additionally, we generated a venom gland cDNA library of snake

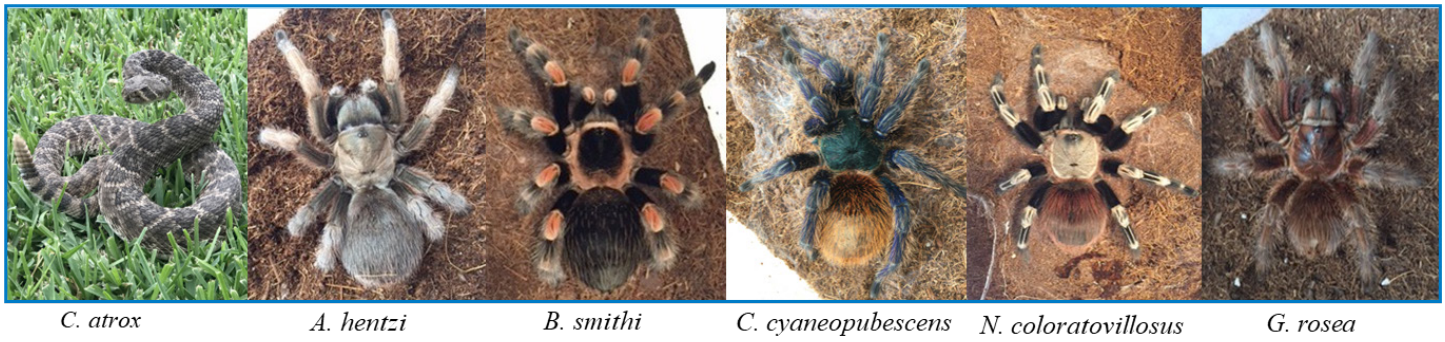

Figure 1: Snake and tarantula species used for venom gland cDNA library construction.

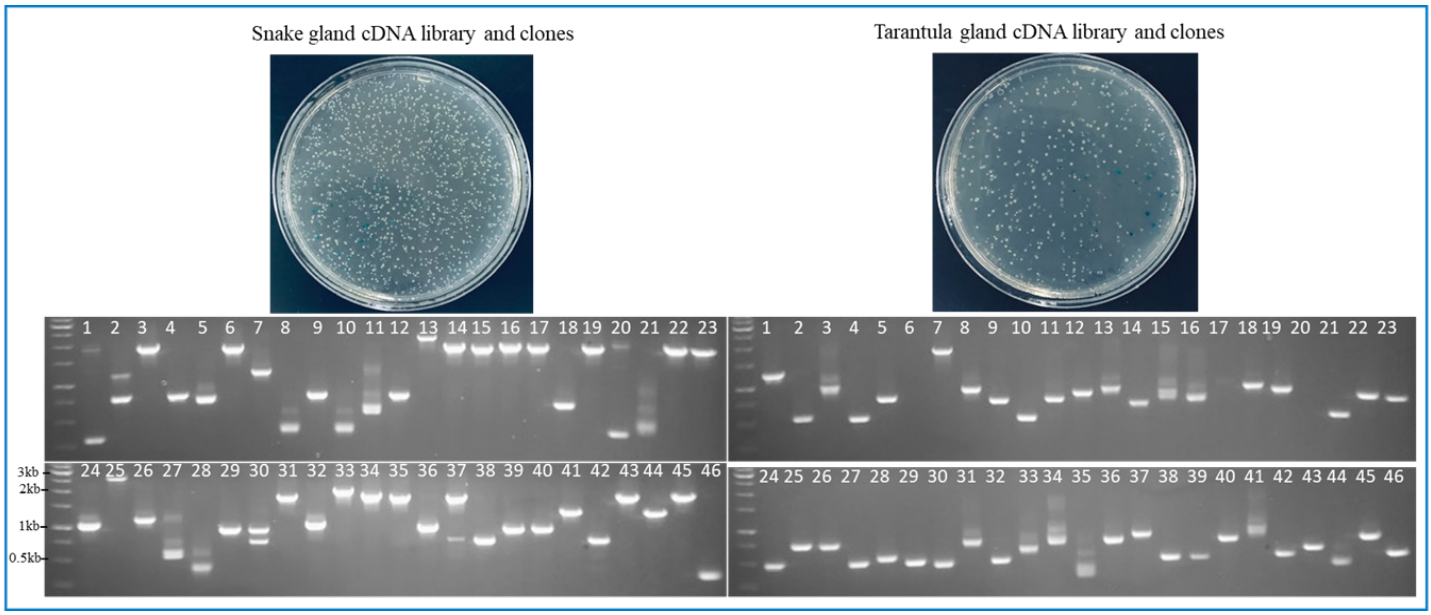

Figure 2: Venom gland cDNA libraries of snake and tarantulas as well as clones checked for insert availability and length.

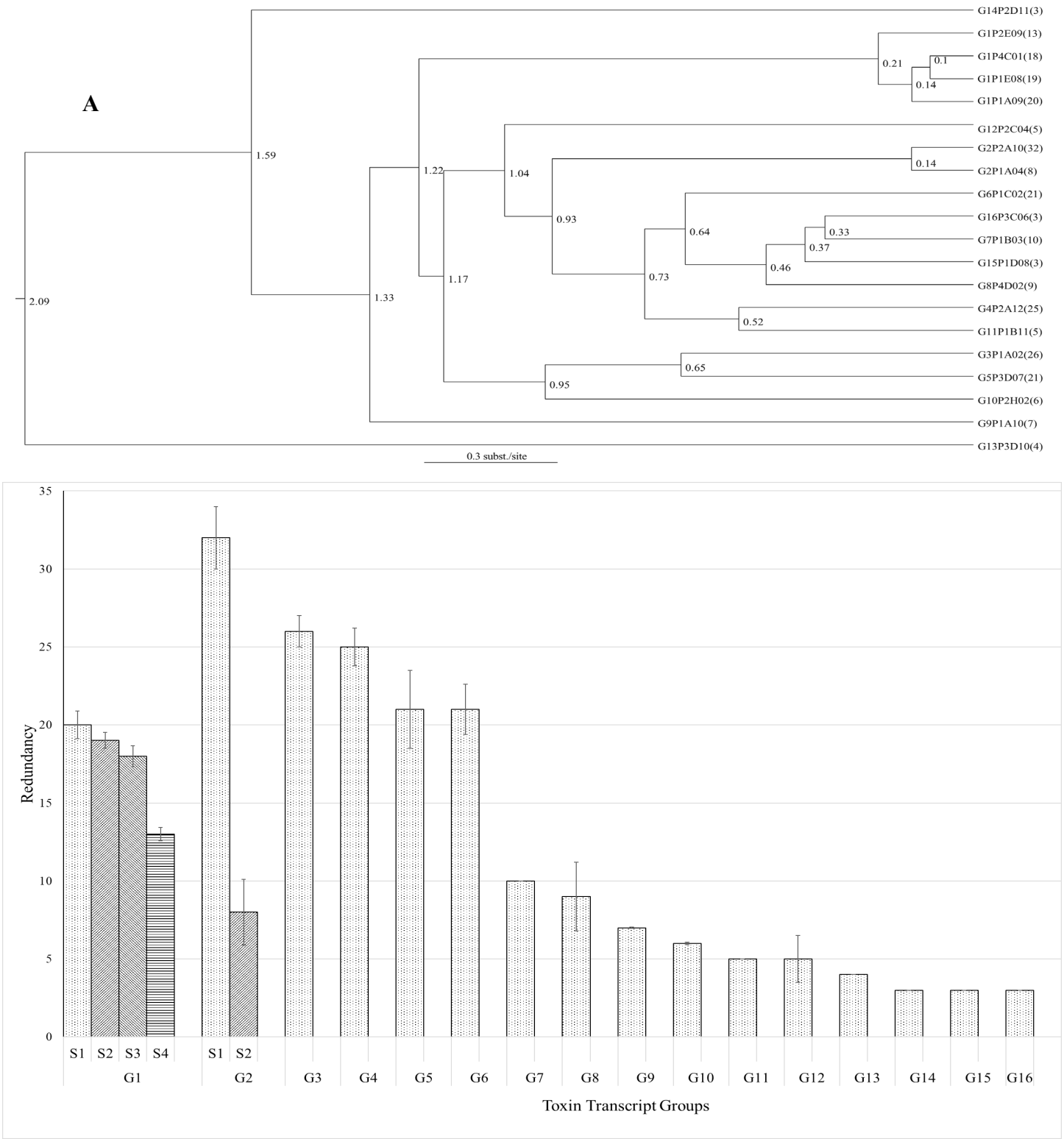

Figure 3: Tarantula venom toxin classification. (A) Phylogenetic analysis, based on the amino acid sequence alignment using MEGA11, the tree was generated by the Bayesian inference (BI) method using BEAST v1.8.4. The 258 manure transcripts were clustered into 16 toxin groups, and groups 1 and 2 were further sorted into 4 and 2 subgroups, respectively. Each group (subgroup) with a representative transcript such as G14P2D11(3): G14 (group14), P2D11 (transcript name), (3) (3 sequences in this group). (B) The distribution of tarantula toxin transcripts. The 16 groups including 4 and 2 subgroups in group 1 and 2, respectively. The "error bars" represent the diverse ranges (70-100%).

```
G1P2E09(9)   EELEAQEGHLMIPGDTDSALETLDDERGLFECVMSCEIEKDGAYVNNNKP
G1P1A09(15) EELEAQEGHLMKPGDIDTALETVDDER-IFECFLSCEIEKDGKP-KEGKP
G1P1E08(14) EELEAQEGHLMIPGDTDTALETVDDER-IFECSFECDIKKEG------KP
G1P4C01(5)   EEYEAQEGYLMNPGDTDTALQTVDDERTIFECVFSCDIKKEG------KP
G1P2E09(9)   CKPKK---------EKKCTGGWRCKFNICLKV 73
G1P1A09(15) CKPKGGK--DKEKDPKKCSGGWRCKFKICLKV 78
G1P1E08(14) CKPKGCKCDKSDKDHKKCSGGWRCKLKLCLKI 75
G1P4C01(5)   CKPKG---------EKKCTGGWRCKIKMCLKI 67
```

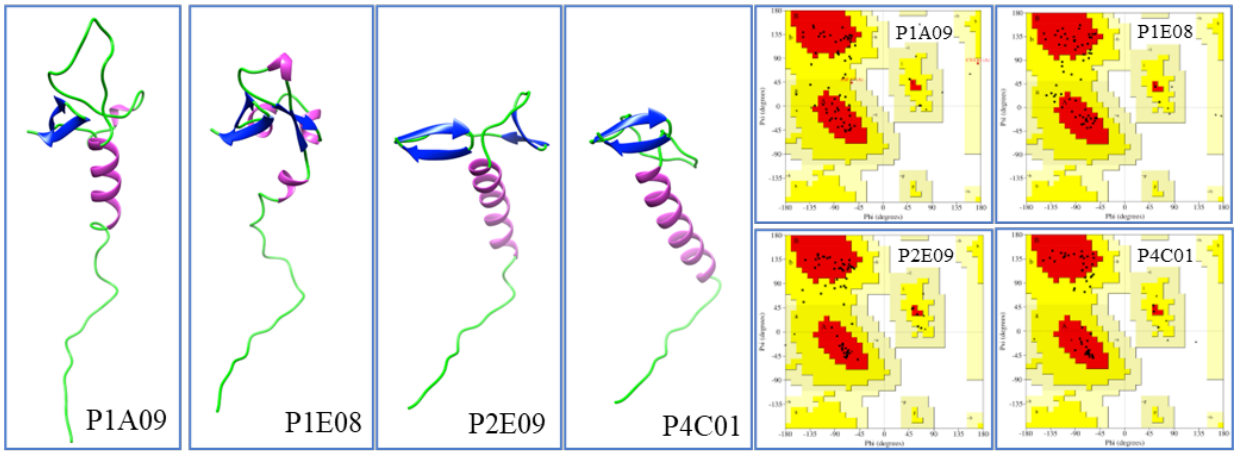

Figure 4: Sequence and structure analysis for the most diverse group 1 transcript. Upper panel, multiple sequence alignment: mature amino acid sequences of four representatives (G1P2E09, G1P1A09, G1P1E08, G1P4C01) in each subgroup transcript were aligned by Clustal Omega. Each subgroup contains 9, 15, 14 and 5 identical sequences. Lower panel, the 3D structures of four representatives were simulated by AlphaFold2. Secondary structures, α-helix (Magenta), β-pleated sheets (blue) and coli(green). The reliability of 3D structures was assessed by Ramachandran plot, the residue distribution (in "most favored and favored") was 96.8%, 100%, 100% and 100% for P1A09, P1E08, P2E09, and P4C01, respectively.

(*C. atrox*) using the same methods. The average length of the snake toxin transcripts was approximately 1 kb, ranging from 200 bp to 3 kb, whereas tarantula toxin transcripts were relatively shorter, implying that tarantula venom toxins are smaller than those of snakes (Figure 2). The venom toxin transcripts and toxins of *C. atrox* have been well characterized [14,15], leading us to direct our focus solely on the tarantula toxin transcripts in this study.

Tarantula Toxin Transcripts

Of the sequenced 752 cDNA clones, 342 were identified as full-length transcripts possessing start and stop codons as well as polyadenylation sequences. Among the translated 342 amino acid sequences, 258 were further clustered into 16 toxin groups including one novel group, with each group containing at least 3 identical sequences (Figure. 3A&B, Table S1). The remaining 84 full-length transcripts include singletons and those with two identical sequences, which encode various enzymes and proteins, such as acetoacetyl-CoA synthetase, phospholipase A_2, Arginine kinase, metalloproteinase, serine endopeptidase, calreticulin-like protein, glutathione peroxidase-like protein, C-type lectin, and ADP/ATP translocase, etc.

Group 1 transcript

Containing 70 sequences, group 1 is the largest and most diverse group, and was further clustered into 4 subgroups consisting of 20, 19, 18 and 13 sequences each (Figure 3B, Table 1S). The sequence identity within each subgroup covers 100% of the sequence and ranges from 91.09% to 100%, but distinct differences exist between each subgroup based on sequence alignment (Figure 4, upper panel). Each subgroup, represented by G1P1A09, G1P1E08, G2P2E09 and G1P4C01, contains 15, 14, 9 and 5 identical sequences, respectively. The representative sequences encode venom toxins: 1) U1-theraphotoxin-Ap1a, a member of HWTX-II family and isolated from Tarantula (*Acanthoscurria paulensis*) [17], 2) U-theraphotoxin-Pv7a isolated from Tarantula (*Pamphobeteus verdolaga*) (Salinas-Restrepo, 2022, direct submission), 3) Omega-theraphotoxin-Ba1 translated from venom gland transcript of Tarantula (*Brachypelma ruhnaui*) [18], and 4) Toxin-like peptide translated from venom gland transcript of tarantula (*Grammostola rosea*) (Kimura et al. 2011, direct submission). It has been demonstrated that theraphotoxins possess antimicrobial and anti-insect activities [19,20], as well as the ability to inhibit "channel" proteins [21,22]. Thus, due to their higher identity, the group 1 toxins might have similar functions. Since we used mixed glands of 5 different tarantula species, it is impossible to reliably predict toxin transcript origins, but subgroup 4 seems from tarantula (*Grammostola rosea*) that was included in cDNA library construction. We further predicted the 3D structures of these four representative toxins (Figure 4, lower panel) using the latest and most accurate program, AlphaFold2 [12,13]. Since there is only one experimentally determined structure available (PDB ID 2KGH, as of February 2023), and it shares less than 50% coverage with these four toxins, it is impossible to evaluate the

accuracy of the predicted toxin structures. We therefore assessed the reliability of the predicted structures by Ramachandran angles (φ and ψ) (Figure 4, lower panel), resulting that the residue distribution (in "most favored and favored") reached 96.8%, 100%, 100% and 100% for P1A09, P1E08, P2E09, and P4C01, respectively (Table 1).

Table 1: Evaluation of Predicted 3D Structures by Ramchandran angles.

Toxin	Most favored (%)	Favored (%)	Allowed (%)	Disallowed (%)
G1P1A09	69.8	27.0	3.2	0.0
G1P1E08	73.4	26.6	0.0	0.0
G1P2E09	77.4	22.6	0.0	0.0
G1P4C01	83.9	16.1	0.0	0.0

Group 2 transcript

The group 2 transcript consists of 40 sequences, which were further classified into two subgroups: one with 32 sequences and the other with 8 sequences (Figure 2B, Table 1S). In addition to 9 individual sequences, subgroup 1 comprises 23 identical sequences, making it the most abundantly expressed single transcript. The representative sequence (G2P2A10) has over 80.25% homology with the other 9 individual sequences. Subgroup 2 contains 5 identical sequences (representative G2P1A04) that have at least 78.75% sequence identity with three other sequences. Two representative sequences, G2P2A10 and G2P1A04, are composed of 81 and 80 amino acids, respectively, and share 70.14% identity. They encode beta-theraphotoxin, which is translated from the venom gland cDNA of tarantula (*Grammostola rosea*) [23]). Beta-theraphotoxins have been detected to have inhibitory effects on voltage-gated sodium and potassium channels [24], suggesting that group 2 toxins may have similar inhibitory effects.

Group 3 transcript

Group 3 consists of 26 sequences, 22 of which are identical. The representative (G3P1A02) shares over 89.86 sequence identity with four other sequences and encodes U5-theraphotoxin-Cg1a, which is translated from the venom gland transcript of the tarantula species (*Chilobrachys jingzhao*) [25].

Group 4 transcript

Group 4 transcript is composed of 25 sequences, 13 of which are identical, and share a sequence identity ranging from 87.65% to 100%. Of the 25 transcripts, 24 sequences contain a "PQER" motif, which is believed to be the cleavage site of the propeptide [26]. The representative (G4P2A12) shares 100% coverage and 82.93% identity with a translated toxin protein (GTx1-12) from the venom gland cDNA of tarantula (*Grammostola rosea*), as reported in the NCBI database (Kimura et al. 2005, direct submission). The transcripts in group 4, as well as those in groups 7, 8, 9 and 16, encode GTx toxin that exhibit higher sequence homology with GTx-15 from the venom of *Grammostola rosea* [27]. GTx1-15 has been shown to preferentially inhibit T-type voltage-dependent calcium channels (Cav3.1), implying that the toxins encoded by these transcripts may possess similar functions.

Group 5 transcript

Group 5 contains 21 sequences, of which 5 are identical. The percent identity among these sequences ranges from 75.38-100%. With 100% coverage, the representative (G5P3D07) is 61.54% identical with an antimalarial peptide isolated from tarantula (*Psalmopoeus cambridgei*) [28].

Group 6 transcript

Twenty-one sequences constitute the group 6 transcript, out of which 17 are identical, and they share at least 84.62% homology with the remaining 4 sequences. The representative (G6P1C02), with 100% coverage, is 72.09% identical to toxin-like peptides in database. Toxin-like peptides are short chains of amino acids with structural similarities, and they are known for their diverse biological activities such as interacting with ion channels or other proteins, as well as possessing antimicrobial, anticancer, and analgesic effects.

Group 7 transcript

The group 7 transcript, the only group containing all identical transcripts (10 sequences), encodes 83 amino acids, and hits in the database with a GTx1-2 toxin protein that was translated from the gland cDNA library of tarantula (*Grammostola rosea*) (Kimura et al. 2005, unpublished). These transcripts also contain a "PQER" cleavage site for propeptide.

Group 8 transcript

There are 9 sequences in group 8 transcript. With 100% coverage, the three identical sequences show at least 77.91% identity with the other 6 sequences. Similar to the transcripts in group 7, all sequences in group 8 encode GTx1-2, but the representative sequences of group 7 and 8 only share 36.5 % identity.

Group 9 transcript

The group 9 transcript is composed of 7 sequences, and they share at least 70.03% amino acid sequence identity. The representative sequence (G9P1A10) has a 98% coverage and shows 57.27% identity to toxin protein (GTx5-1) in the database.

Group 10 transcript

The 5 identical sequences within the group 10 transcript exhibit 99.14% identity with the sixth sequence. With 100% coverage, the representative (G10P2H02) is 96.55% identical with a HWTX protein in the database, which was translated from the venom gland transcript of tarantulas (*Cyriopagopus schmidti*) [29].

Toxin groups 11and 12

Groups 11 and 12 each consist of 5 sequences. All sequences in-group 11 are identical, whereas in group 12, 4 identical sequences share 85.54% identity with the fifth one. Group 11 transcripts encode toxin-like peptide, while group 12 encodes U11-theraphotoxin-Agm2a.

Toxin group 13

Consisting of 4 identical sequences, as of March 2023, group 13 is a novel group with no significant similarity found in the NCBI database.

Toxin groups 14, 15 and 16

Groups 14, 15, and 16 each consist of 3 identical sequences. Group 14 encodes the shortest toxin protein, Hainantoxin-XV2, consisting

of 48 amino acids, while groups 15 and 16 hit Kppa-theraphotoxin-Gr1a and GTx1-1, respectively, in the NCBI database.

Conclusions

We isolated 342 full-length tarantula toxin transcripts from a mixed gland of cDNA library of five tarantula species and categorized them into 16 distinct groups. These groups were then discussed in further detail. Additionally, we used AlphaFold2 to predict the 3D structures of four representative toxins from the most diverse group (group 1). In future work, we plan to use these transcripts to recombinantly produce venom toxins to explore their biological functions and potential biomedical applications.

Funding

This research was funded by NIH-NINDS, grant number R15NS128563.

References

1. Chang CC, Lee CY. Isolation of neurotoxins from the venom of Bungarus multicinctus and their modes of neuromuscular blocking action. Arch Int Pharmacodyn Ther. 1963; 144: 241-257.
2. Dellisanti CD, Yao Y, Stroud JC, et al. Crystal structure of the extracellular domain of nAChR a1 bound to α-bungarotoxin at 1.94 Å resolution. Nat Neurosci. 2007; 10: 953-962.
3. Rahman MM, Teng JF, Worrell BT, et al. Structure of the native muscle-type nicotinic receptor and inhibition by snake venom toxins. Neuron. 2020; 106: 952-962.
4. Noviello CM, Gharpure A, Mukhtasimova N, et al. Structure and gating mechanisms of the α7 nicotinic acetycholine receptor. Cell. 2021; 184: 2121-2134.
5. Agwa AJ, Peigneur S, Chow CY, et al. Gating modifier toxins isolated from spider venom: modulation of voltage-gated sodium channels and the role of lipid membranes. J. Biol. Chem. 2018; 293: 9041-9052.
6. Agwa AJ, Tran P, Mueller A, et al. Manipulation of a spider peptide toxin alters its affinity for lipid bilayers and potency and selectivity for voltage-gated sodium channel subtype 1.7. J. Biol. Chem. 2020; 296: 5067-5080.
7. Vetter I, Davis JL, Rash LD, et al. Venomisc: a new paradigm for natural products-based drug discovery. Amino acids. 2011; 40: 15-28.
8. Utkin YN, Animal venom studies: current benefits and future developments. World J Biol Chem. 2015; 26: 28-33.
9. Kambas D. Tarantupedia: an online taxonomic database for the worlds largest spiders. www.tarantupedia.com (accessed on 2023-6-13).
10. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 2011; 7: 539.
11. Drummond AJ, Suchard MA, Xie D, et al. Bayesian phylogenetics with BEAUTi and BEAST 1.7. Mol. Biol. Evol. 2012; 29: 1969-1973.
12. Jumper J, Evans R, Pritzel A. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596: 583-589.
13. Marx V. Method of the year: protein structure prediction. Nat Methods. 2022; 19: 5-10.
14. Jia Y, Lopez I, Kowalski P. Toxin transcripts in Crotalus atrox venom and in silico structures of toxins. J Venom Res. 2020; 10: 18-22.
15. Calvete JJ, Fasoli E, Sanz L, et al. Exploring the Venom Proteome of the Western Diamondback Rattlesnake, Crotalus atrox, via Snake Venomics and Combinatorial Peptide Ligand Library Approaches. J. Proteome Res. 2009; 8: 3055-3067.
16. Teufel F, Armenteros JJA, Johansen AR, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022; 40: 1023-1025.
17. Mourao CB, Heghinian MD, Barbosa EA, et al. Characterization of a novel peptide toxin from Acanthoscurria paulensis spider venom: a distinct cysteine assignment to the HWTX-II family. Biochemistry. 2013; 52: 2440-2452.
18. Ardisson-Araújo DM, Morgado FDS, Schwartz EF, et al. A new theraphosid spider toxin causes early insect cell death by necrosis when expressed in vitro during recombinant baculovirus infection. PLoS One. 2013; 8: e84404.
19. Rates B, Prates MV, Verano-Braga T, et al. µ-Theraphotoxin-An1a: Primary structure determination and assessment of the pharmacological activity of a promiscuous anti-insect toxin from the venom of the tarantula Acanthoscurria natalensis (Mygalomorphae, Theraphosidae). Toxicon. 2013; 70: 123-134.
20. Abreu TF, Sumitomo BN, Nishiyama Jr MY, et al. Peptidomics of Acanthoscurria gomesiana spider venom reveals new toxins with potential antimicrobial activity. J Proteomics. 2017; 151: 232-242.
21. Murray JK, Ligutti J, Liu D, et al. Engineering potent and selective analogues of GpTx-1, a tarantula venom peptide antagonist of the Na(V)1.7 sodium channel. J Med Chem. 2015; 58: 2299-2314.
22. McArthur JR, Wen J, Hung A, et al. µ-Theraphotoxin P13a inhibition of Cav3.3 channels reveals a novel isoform-selective drug binding site. eLife. 2022; 11: e74040.
23. Kimura T, Ono S, Kubo T. Molecular Cloning and Sequence Analysis of the cDNAs Encoding Toxin-Like Peptides from the Venom Glands of Tarantula Grammostola rosea. Int J pept. 2012; 731293.
24. Luo J, Zhang Y, Gong M, et al. Molecular surface of JZTX-V (β-Theraphotoxin-Cj2a) interacting with voltage-gated sodium channel subtype Nav1.4. Toxins. 2014; 6: 2177-2193.
25. Chen J, Deng M, He Q, et al. Molecular diversity and evolution of cystine knot toxins of the tarantula Chilobrachys jingzhao. Cell Mol Life Sci. 2008; 65: 2431-2444.
26. Hu Z, Chen B, Xiao Z, et al. Transcriptomic analysis of the spider venom gland reveals venom diversity and species consanguinity. Toxins. 2019; 11: 68.

27. Ono S, Kimura T, Kubo T. Characterization of voltage-dependent calcium channel blocking peptides from the venom of the tarantula Grammostola rosea. Toxicon. 2011; 58: 265-276.

28. Choi SJ, Parent R, Guillaume C, et al. Isolation and characterization of Psalmopeotoxin I and II: two novel antimalarial peptides from the venom of the tarantula Psalmopoeus cambridgei. FEBS Lett. 2004; 572: 109-117.

29. Jiang L, Peng L, Chen J, et al. Molecular diversification based on analysis of expressed sequence tags from the venom glands of the Chinese bird spider Ornithoctonus huwena. Toxicon. 2008; 51: 1479-1489.

Supplementary Materials

Table S1: Toxin transcript groups and their amino acid sequences.

Transcript group	Clone #	Rep. Tx	Deduced amino acid sequence (Signal peptide in bold)
Group 1	20	G1P1A09	**MRSLTLAAVLACSLLLVFHTSAA**EELEAQEGHLMKPGDIDTALETVDDERIFECFLSCEIEKDGKPKEGKPCKPKG-GKDKEKDPKKCSGGW RCKFKIC LKV
Group 1	19	G1P1E08	**MRSLTLAALLACSLLLVFHTSAA**EELEAQEGHLMIPGDTDTALETVDDERIFECSFECDIKKEGKPCKPKGCKC-DKSDKDHKKCSGGWRCKL KLCLKI
Group 1	18	G1P4C01	**MRSLTLVAILACSLLLVLHTSAA**EEYEAQEGYLMNPGDTDTALQTVDDERTIFECVFSCDIKKEGKPCKPKGEK-KCTGGWRCKIKMCLKI
Group 1	13	G1P2E09	**MRSLTLAAIFACSLLLVFHTAA**EELEAQEGHLMIPGDTDSALETLDDERGLFECVMSCEIEKDGAYVNNNKPCKP-KKEKKCTGGWRCKFNICLKV
Group 2	32	G2P2A10	**MKASVFFAVLGLALCAYSFAL**EEQDQLSLRNDLVSLMFADNTELIPEAEGRYCQKWMWTCDQERKCCEDM-VCELWCKIRLG
Group 2	8	G2P1A04	**MKAFFVILGLALLCAYSFAL**EEQDHLSLRNDLLTIMFAENNELTPETEERGCQKWMWTCDSKRPCCDDMACELW-CKVRLG
Group 3	26	G3P1A02	**MNGKIFVFLVVLNLVICNLA**ERKSETDIENAPLIQENFGRFCFPEGRPCTTSARCCIPMVCKQKKCLRS
Group 4	25	G4P2A12	**MKTSLVLVIAGLALLSVCYA**SEMKEQSSINEVLSAIFHVEQPQERDDCLGFFKSCNPDNDKCCENYKCNRRDKW-CKYVIGK
Group 5	21	G5P3D07	**MGMKTIIFLVFLTLVVCSNA**AINAEIDTGDSPMIQERRCLPAGKPCAGVTQKIPCCGKCSRNKCT
Group 6	21	G6P1C02	**MKTLGLLLLLGLAVLYCSA**SELANKELVKEVLRAMVVQPEERECKYYLGSCTKDDDCCPHLQCHSIHEWCL-WDGSFGK
Group 7	10	G7P1B03	**MRTSVLVGFLGLTIFAVICSA**SETQGNDLPVKLLAAFLEATAKVEPQERGECHWLLGYCKKDSECCSDLTCSKLN-FCQWIAFA
Group 8	9	G8P4D02	**MRTAVLAAVLGVVLLVCFCSA**SELQKNGVPEEVVSAIMGEILGMRPTERADCRMMFGGCAKDSDCCAHLG-CKPTAKYCAWDGTVGK
Group 9	7	G9P1A10	**MKLILGVIAVFLVIAAVA**LPSGNLRDGFDPSELLGQPMEEKRTETARACSKQVGEKCKRNCECCGAYTVC-GYYYVGSTTVYECMNKTSNNV ILNTMGHGMNAVTNAFSFCWS
Group 10	6	G10P2H02	**MNTVRVTFLMVFVLAVSLGQA**DKDENRMEMQEKTEKTEEDKSYLDFAENLLLQKLEELEAKLLEEDSEESRN-SRQKRCIGEGVPCDENDPRC CSGLVCLKPPLHGIWYKSYYCYKK
Group 11	5	G11P1B11	**MKAFVLLAIAALSLLSVVCYA**SESKDQDSIDEMLSAILSEQPQQRGDCHKFWGWCRGEPDPCCEHLTCSTKHGW-CVWDGSFGK
Group 12	5	G12P2C04	**MKLATLLGLSVLLLTLCVLSCTS**QHPGLEKSRVSYENMGDEENAEERFCVDERETCSKIKGPLCCTGECICPIYGD-CFCYGS
Group 13	4	G13P3D10	MKFSLLILLIEIIMKVRRLSFFRGTTRPCRALLMFSWGGKNLHLYFSKSPIILFFDKSSYRRDNSIIFFKRPYQKKRL
Group 14	3	G14P2D11	MSTLCTASWVDGNQIKLCRNKGGKLKKVLHFIQKSFSKIKSCKKKKKN
Group 15	3	G15P1D08	**MKSSVLAVFLGLTLLVVLCSA**SESEENDLYEDMFGAAIEISAEAKPQESGRECRYMFGGCKKDSECCKHLGCT-TRAPKYCAWDGTFSK
Group 16	3	G16P3C06	**MKTSVLVVFLGVTFFAVLCSA**SDSSENDLSEEVIRAVFEADAEPKERGECRWFMGGCDSTLDCCKHLSCK-MGLYYCAWDGTFGK

* Signal peptides in bold were predicted by SignalP-6.0 [16]. No signal sequence for Group 13 and 14 was detected by SignalP-6.0. Amino acid sequences were translated using toxin transcripts by Expasy (https://web.expasy.org/translate/). The representative toxin (Rep. Tx) and their amino acid sequences were listed.