# Venomous gland transcriptome and venom proteomic analysis of the scorpion *Androctonus amoreuxi* reveal new peptides with anti-SARS-CoV-2 activity

Ahmad Ghazal [a,b,1], David Clarke [c,2], Mohamed A. Abdel-Rahman [d,3], Antonio Ribeiro [e,4], Elaina Collie-Duguid [e,5], Craig Pattinson [f], Kate Burgoyne [f,6], Taj Muhammad [g,7], Sanad Alfadhel [a,8], Zeynab Heidari [e], Reham Samir [d], Mariam M. Gerges [d], Istifanus Nkene [f], Rosa A. Colamarino [f], Karolin Hijazi [f,9], Wael E. Houssen [a,b,*,10]

[a] *Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK*
[b] *Department of Chemistry, University of Aberdeen, Aberdeen AB24 3UE, UK*
[c] *School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, UK*
[d] *Zoology Department, Faculty of Science, Suez Canal University, Ismailia 41522, Egypt*
[e] *Centre for Genome-Enabled Biology and Medicine, University of Aberdeen, Aberdeen AB24 3RY, UK*
[f] *School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen AB25 2ZD, UK*
[g] *Pharmacognosy, Department of Pharmaceutical Biosciences, Uppsala University, Biomedical Centre, Box 591 SE-75124 Uppsala, Sweden*

ABSTRACT

The recent COVID-19 pandemic shows the critical need for novel broad spectrum antiviral agents. Scorpion venoms are known to contain highly bioactive peptides, several of which have demonstrated strong antiviral activity against a range of viruses. We have generated the first annotated reference transcriptome for the *Androctonus amoreuxi* venom gland and used high performance liquid chromatography, transcriptome mining, circular dichroism and mass spectrometric analysis to purify and characterize twelve previously undescribed venom peptides. Selected peptides were tested for binding to the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein and inhibition of the spike RBD – human angiotensin-converting enzyme 2 (hACE2) interaction using surface plasmon resonance-based assays. Seven peptides showed dose-dependent inhibitory effects, albeit with $IC_{50}$ in the high micromolar range (117–1202 $\mu$M). The most active peptide was synthesized using solid phase peptide synthesis and tested for its antiviral activity against SARS-CoV-2 (Lineage B.1.1.7). On exposure to the synthetic peptide of a human lung cell line infected with replication-competent SARS-CoV-2, we observed an $IC_{50}$ of 200 nM, which was nearly 600-fold lower than that observed in the RBD – hACE2 binding inhibition assay. Our results show that scorpion venom peptides can inhibit the SARS-CoV-2 replication although unlikely through inhibition of spike RBD – hACE2 interaction as the primary mode of action. Scorpion venom peptides represent excellent scaffolds for design of novel anti-SARS-CoV-2 constrained peptides. Future studies should fully explore their antiviral mode of action as well as the structural dynamics of inhibition of target virus-host interactions.

* Corresponding author at: Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK.
  *E-mail address:* w.houssen@abdn.ac.uk (W.E. Houssen).
[1] orcid.org/0009-0009-2125-0811
[2] orcid.org/0000-0002-3741-2952
[3] orcid.org/0000-0003-3862-5535
[4] orcid.org/0000-0001-8680-7224
[5] orcid.org/0000-0002-7511-4139
[6] orcid.org/0000-0003-3986-0975
[7] orcid.org/0000-0002-6636-5809
[8] orcid.org/0000-0001-8092-8607
[9] orcid.org/0000-0002-8589-2129
[10] orcid.org/0000-0001-9653-5111

## 1. Introduction

The COVID-19 pandemic, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), has led to millions of confirmed cases and deaths worldwide (https://covid19.who.int/). Despite the accelerated development and deployment of vaccines, there is an urgent need for development of new therapeutics to combat the threat of new evolving SARS-CoV-2 variants against which current vaccines may be less effective. Currently approved small molecule antiviral drugs include remdesivir [1] and molnupiravir [2], which target the RNA-dependent RNA polymerase and nirmatrelvir which inhibits the main protease of the virus [3]. As these targets are involved in viral replication, these drugs will only work after the virus infects the cell. Structural studies showed that the higher transmissibility of SARS-CoV-2 compared to previous coronaviruses is due to tighter binding of the viral spike RBD to hACE2 [4,5]. This critical interaction precedes the viral entry into the cell and its inhibition is a key therapeutic target. Protein-protein interactions (PPIs) are generally very challenging therapeutic targets for traditional small molecule drugs due to their highly dynamic and large interfacial surface [6]. However, there is growing evidence that conformationally constrained and macrocyclic peptides can modulate these interactions and be a valuable springboard for the design of effective small molecule inhibitors for PPIs [7,8]. Scorpion venoms are known to be rich in these conformationally constrained peptides of which several have been proven to possess antiviral activity against a wide range of viruses [9–18]. An example is the broad-spectrum antiviral venom peptide, Ev37 from the scorpion *Euscorpiops validus* that was found to be able to inhibit dengue virus type 2 (DENV-2), hepatitis C (HCV), Zika virus (ZIKV) and herpes simplex type I (HSV-1) in a dose-dependent manner at noncytotoxic concentrations [14]. Ev37 exhibits its action via alkalizing acidic organelles to prevent low pH-dependent fusion of the viral membrane-endosomal membrane, which mainly blocks the release of the viral genome from the endosome to the cytoplasm and then restricts viral late entry. SARS-CoV viruses are considered pH-sensitive viruses, and their intracellular trafficking requires an acidic environment [19]. Several other mechanisms have been reported for venom peptides and include direct virucidal effect, disturbance of the attachment of virus particles to the cell membrane surface, or interference with the virus replication [12].

In this manuscript, we used mass spectrometry (MS) and venomous gland transcriptomic analysis to identify the sequence of previously undescribed purified peptides from the venom of the north African scorpion *Androctonus amoreuxi* collected from Egypt. The chemistry of *A. amoreuxi* venom and its transcriptome have not been extensively studied as only ten venom peptides have been identified and characterized to date (Table S1). We have generated an annotated reference transcriptome for the *A. amoreuxi* venom gland. We also used a surface plasmon resonance-based assay to test the ability of the newly identified peptides to bind to the SARS-CoV-2 spike protein and to inhibit the spike RBD – hACE2 interaction. We synthesized the most active peptide and tested its antiviral activity against SARS-CoV-2 (Lineage B.1.1.7).

## 2. Materials and methods

### 2.1. Compliance with the Nagoya protocol on access and benefit sharing

This research project was conducted in accordance with the obligations under the Nagoya Protocol on Access and Benefit Sharing. We obtained Prior Informed Consent (PIC) no. 00308021010800/4 from the Egyptian Convention on Biological Diversity (CBD) National Focal Point to conduct this research using the scorpion, *A. amoreuxi* collected from Egypt.

### 2.2. Collection of scorpions and venom preparation

Adult *A. amoreuxi* scorpions were purchased from El Marwa Office

for Export (Egypt) and were housed individually (to avoid cannibalism) in clear plastic containers in the laboratory of Professor Abdel-Rahman at Suez Canal University. The scorpion specimens were identified according to the key of El-Hennawy [20]. Scorpions were fed on small insects (mealworms) and received water ad libitum. Crude venom was extracted using electrical stimulation (12–16 V; 10 s) and the milked venom was collected and centrifuged for 20 min at 13,000 rpm / 4 °C. Clear supernatants were pooled, freeze-dried and stored at − 20 °C until use [21]. All methods were performed following the guidelines of the Suez Canal University's Faculty of Science Research Ethics Committee and the experimental protocol was approved by this committee with reference number REC/01–02-01–2022.

### 2.3. Illumina sequencing

RNA was extracted from telsons dissected from five adult scorpions and stored in RNAlater® at −80 °C. Telsons were shipped on dry-ice and stored on receipt at −80 °C until processing. Scorpions were anesthetized before telson dissection by putting them in a freezer at −20 °C for ~4–5 min. Venom was milked 4 days before dissection as recommended in previous studies [21–23]. Telsons were combined, washed in ice-cold PBS to remove RNAlater® and homogenized in 1 mL TRIzol reagent (Thermo Scientific, UK) with $2 \times 3$ mm tungsten carbide beads in a TissueLyser II (Qiagen, CA). Total RNA was extracted in TRIzol (Thermo Scientific, UK) and purified on RNeasy micro columns with on column DNase digestion (Qiagen, CA), according to the manufacturer's instructions. Total RNA was quantified by fluorimetry (Qubit, Thermo Scientific, UK) and quality (RIN 9.1) was assessed on a Tapestation 4200 (Agilent, CA). Unique dual indexed stranded TruSeq mRNAseq libraries were prepared from 500 ng total RNA, according to the manufacturer's instructions (Illumina, CA). Libraries were quantified by qPCR with SYBR green (Kapa Library Quantification Complete Universal, Roche, CH) on the QuantStudio 6 Flex (ThermoFisher Scientific, UK) and average library size of 291 bp determined on the Tapestation 4200 (Agilent, CA) prior to sequencing and base calling on an Illumina NextSeq500 with v2.5 chemistry and 75 bp paired-end reads, with 14.3 Gb sequence output.

### 2.4. Transcriptome assembly and annotation

Raw read quality was assessed before and after pre-processing with FastQC (version 0.11.8) [24] and MultiQC (version 1.7) [25]. Raw reads were pre-processed using Trim Galore! (Version 0.6.4) [26] to trim low-quality bases (<Q30) and adapters and exclude short-trimmed reads (<20 bp), before de novo transcriptome assembly with Trinity (version 2.8.5) [27,28], with unique accession Trinity_DN prefix identifying each feature in the output assembly. Assembly quality was assessed with BUSCO (version 4.0.6) with the arthropod lineage dataset, arthropoda_odb10 [29], rnaQUAST (version 2.2.0) [30] and GeneMarkS-T (version 5.1) gene prediction [31]. Transcript abundances were determined with embedded Trinity scripts using the salmon method [32]. For evaluation of RNA-seq read representation in the Trinity assembly as a measure of quality, input raw reads used for the venom gland transcriptome assembly were mapped against the assembly using Bowtie 2 (version 2.3.5) [33]. The alignment file was post-processed with SAMtools (version 1.9) and coverage and allelic composition of transcripts of interest were inspected with SAMtools mpileup (version 1.9) [34,35]. Translated transcripts were searched for peptide sequence tags from the MS identified peptides, to identify the encoding transcript, and the sequence of the full-length precursor peptide. Sequence tags were produced with R (version 3.6.1) [36], Rstudio (version 1.1.456) [37], ggplot2 (version 3.3.6) [38], and ggseqlogo (version 0.1) [39].

The assembly was annotated using Trinotate software (version 3.2.1) [40] with residual background rRNA transcripts ($n = 6$) identified with Trinotate RNNAMER. AMPFinder with Diamond (version 1.1.0) was

used to identify putative *A. amoreuxi* orthologues of known antimicrobial peptides, with the *A. amoreuxi* contig nucleotide sequences as input queries against translated RNA sequences of known AMPs in the manually curated dbAMP version 2.0 database, which includes anti-bacterial, anti-viral and other AMPs [41].

## 2.5. Isolation of venom peptides

Reversed-phase high performance liquid chromatography (RP-HPLC) was performed using an Agilent 1260 LC System. Forty milligrams of the crude venom were dissolved in 4 mL (50:50 water: acetonitrile v/v), cleared by centrifugation at 7000 g and the supernatant was injected into a C4 column (ACE 5 C4–300, 5 μm, 250 × 10 mm, 300 Å). Peptides were eluted using a starting linear gradient at 0%– 55% of solution B (0.1% TFA in acetonitrile) in solution A (0.1% TFA in water) over 40 min with a flow rate of 3 mL/min. The gradient was increased linearly to 100% solution B over the following 5 min and the composition of the mobile phase was kept at 100% solution B until the end of the 50 min-run. Individual fractions were manually collected based on the peak absorbances at 220 nm, 254 nm and 280 nm. The fractions were further purified using a C18 column (ACE 5 C18–300, 5 μm, 250 × 10 mm, 300 Å). The collected samples where then dried using rotary evaporator to remove the acetonitrile, followed by lyophilization using a LaboGene CoolSafe freeze dryer.

## 2.6. Mass spectrometric analysis

MS data was acquired on a 12 T SolariX 2XR Fourier transform – Ion cyclotron resonance (FT-ICR) instrument equipped with electrospray (ESI) ionization (Bruker Daltonics). RP-HPLC fractionated samples were infused at 2 μL/min and spectra were acquired between 200 and 4000 *m/z*. For top-down fragmentation, individual ions (single proteoforms) were isolated using the quadrupole and tandem MS was performed using either collision induced dissociation (CID) or electron capture dissociation (ECD). For ECD, typical cathode conditions were a bias of 1.5 V, lens voltage of 15 V, and pulse length was varied between 5 and 15 ms. Collision induced dissociation (CID) was typically conducted using a voltage between 15–25 V. The resulting top-down fragmentation spectra were processed using Data Analysis v4.2 (Bruker Daltonics) and the sophisticated numerical annotation procedure (SNAP) was used to produce a monoisotopic mass list. These mass lists were manually searched for sequence tags. Translated transcripts were searched for these sequence tags to identify the encoding transcript of each peptide and the sequence of the precursor peptide was confirmed for existence, in the transcriptome, using the sequence logo method described above.

## 2.7. Template-based 3D modelling

The homology modelling of the identified peptides with more than 30 amino acid residues was performed using SWISS-MODEL [42]. The templates were selected based on the global model quality estimate (GMQE) which is a quality estimate that combines properties from the target-template alignment and the template structure. For large peptides (> 50 aa), we used GMQE cut-off value of 0.6, coverage = 0.9 and identity = 40% while the corresponding values for short peptides (<50 aa) are GMQE = 0.8, coverage = 0.9 and identity = 80%. Visualization and drawing of selected models were executed with PyMOL [43]. The presence of disulfide (DS) bridges was confirmed through MS analysis, and the configuration of these DS bridges was predicted through homology with previously identified scorpion toxins.

## 2.8. Far-UV circular dichroism (CD)

CD spectra of purified peptides (50 μM in 20 mM sodium phosphate buffer [pH 7.4] and 20 mM sodium fluoride at 25 °C) were acquired between wavelengths of 260 nm and 190 nm on a MOS-500 CD

Spectrometer (BioLogic). Each CD spectrum represent the average of 3 scans which was baseline-corrected by subtracting of the buffer spectrum. Mean residue differential extinction coefficient $\Delta\varepsilon res$ of each spectrum was calculated from the observed ellipticity θ and plotted against wavelength. The CD analysis was carried out only on peptides with sufficient amounts. Bestsel server (https://bestsel.elte.hu/index.php) was employed for determining the secondary elements from CD spectra [44].

## 2.9. Peptide synthesis

Reagents: The majority of the Fmoc amino acid derivatives were purchased from Novabiochem, Merck Biosciences, UK. Fmoc-Leu-OH, Fmoc-Gln(Trt)-OH, *N,N*′-diisopropylcarbodiimide (DIC), trifluoroacetic acid (TFA) and triisopropylsilane (TIS) were acquired from Fluorochem, UK. Fmoc-L-Arg(Pbf)-OH, Fmoc-Tyr(tBu)-OH, ethyl cyano(hydroxyimino)acetate (Oxyma), Fmoc-Rink Amide ProTide (LL) resin was obtained from the CEM corporation, USA. *N,N*-dimethylformamide (DMF), dichloromethane (DCM), methanol (MeOH) and diethyl ether were purchased from VWR, Avantor, USA. Piperidine was purchased from Merck Life Sciences, UK. Sodium diethyldithiocarbamate trihydrate, 2,2′-(ethylenedioxy)diethanethiol (DODT) and pyridine were acquired from Sigma Aldrich.

Synthesis: Peptide synthesis was performed on a Liberty Blue™ Automated Microwave Peptide Synthesizer (CEM Corporation, North Carolina, USA) using the standard solid-phase peptide synthesis (SPPS)-Fmoc/tBu chemistry with piperidine (20% v/v) as the Fmoc-deprotecting agent. The AM29A5-syn peptide was assembled on a 0.1 mmol scale using Fmoc-Rink Amide ProTide (LL) resin with a substitution value of 0.18 mmol/g using 0.2 M Fmoc-protected amino acids and activated by DIC (1 M) and in the presence of 1 M Oxyma prepared in DMF. All the arginine residues were double coupled, and rest of the amino acids were single coupled. On synthesis completion, the resin was washed with dry DCM to remove DMF and subsequently dried under $N_2$. The dried-resin peptide was cleaved with a cocktail solution of TFA/TIS/DODT/$H_2O$ (92.5:2.5:2.5:2.5) for 4 h followed by filtration to remove resin beads. The flow through was then concentrated under $N_2$ gas. The peptides were precipitated using cold diethyl ether, placed into a −20 ºC freezer overnight, washed with ether (3x) and dried under vacuum to give a crude solid. The crude peptides were purified using reversed-phase HPLC on an Agilent Technologies 1260 Infinity using a C18 column (ACE 5 C18-HL, 5 μm, 250 × 10 mm, 100 Å) through an acetonitrile (+0.1% TFA)/Water (+0.1% TFA) gradient. Fractions containing the pure peptide were subsequently lyophilized on a LaboGene CoolSafe Freeze dryer to give the pure solid. Peptide identity was confirmed by MS analysis.

Folding: Four buffers (Table S2) were used to allow the correct oxidative folding of the linear synthetic peptide (AM29A5-syn). Aliquots of each reaction were collected at different time intervals and immediately quenched using 0.4% TFA in MilliQ water. The quenched aliquots were then injected in a C18 column (ACE 5 C18–300, 5 μm, 250 × 10 mm, 300 Å) and the retention time of the folded product was compared with that of the native peptide. MS analysis was also used to confirm the formation of DS bridges.

## 2.10. Surface plasmon resonance

Recombinant proteins: The receptor binding domain (RBD) of the spike protein of SARS-CoV-2 (aa 319–541) and hACE2 (aa 19–615) were recombinantly produced in HEK293 cells and provided by Peak Proteins Ltd (Macclesfield, UK). QC sheets are provided (Supplementary Info File 6 & Supplementary Info File 7).

Surface plasmon resonance assays: Venom peptide binding to RBD and competition with hACE2 was assayed in surface plasmon resonance experiments carried out using Biacore X100 (Cytiva, Uppsala, Sweden).

RBD binding measurements: RBD in PBS-P + pH 7.4 was captured

and immobilized (approximately 6300 RU) on the surface of Flow cell 2 of a nitrilotriacetic acid (NTA) sensor chip using the standard nickel activation procedure (Cytiva). This was followed by activation with a mixture of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and *N*-hydroxysuccinimide (NHS) for 7 min and inactivation with ethanolamine for 7 min for covalent linking of the 6 ×His-tag-RBD to the nickel-activated surface. Flow cell 1 (activated and blocked) served as a reference (blank) cell. Binding of fluid phase venom-derived peptides was determined over a range of concentrations determined by peptide quantity. Peptides were dissolved in PBS-P+ buffer pH 7.4 (0.2 M phosphate buffer, 27 mM KCl, 1.37 M NaCl, 0.5% surfactant P20; Cytiva), supplemented with 71 µL of 350 mM EDTA. Injection volume was 60 µL and flow rate was 5 µL/min. The surface was regenerated with 350 mM EDTA and 0.5% w/v SDS. Sensorgrams of RBD binding measurements were obtained by subtraction of the sensorgram of buffer alone from that generated by injection of peptides. Equilibrium dissociation constants ($K_D$) as well as association ($k_a$) and dissociation constant ($k_d$) rates were calculated for the peptides that showed dose-dependent binding using the Biacore X100 Evaluation Software version 2.0.2 (Cytiva). Curves were fitted to a 1:1 binding model as judged by the Chi2 value and distribution of residuals.

Inhibition studies: Binding of fluid phase hACE2 at a concentration of 150 nM was determined in the presence of venom peptides at the same concentrations listed above and compared to binding activity of hACE2 alone. For peptides that showed direct binding activity to RBD, final inhibition sensorgrams were obtained by subtraction of the sensorgram of injection of peptides alone from that generated by injection of peptides in presence of hACE2. Fluid phase compounds were dissolved in PBS-P+ pH 7.4 buffer supplemented with 500 µM EDTA. Flow rate was 5 µL/min, and the injection volume was 60 µL. After each binding measurement, the surface was regenerated with 250 mM EDTA and 0.5% w/v SDS. Single measurements were carried out for each condition. $IC_{50}$ values were calculated by linear regression using Microsoft Excel with percentage inhibition values (relative to hACE2 binding without peptides) plotted against peptide concentration.

### 2.11. Antiviral assay

Virus propagation and antiviral neutralization assays were performed at Containment Level 3.

Cells, virus and reagents: Vero E6 cells for SARS-CoV-2 propagation were obtained from BEI Resources. Cells were grown and maintained in grown in Dulbecco's Modified Eagles Medium (DMEM) (Gibco, UK) supplemented with 10% foetal calf serum (FCS), 1 mM sodium pyruvate and 2 mM L-glutamine. Human Lung Carcinoma Cells (A549) Expressing Human Angiotensin-Converting Enzyme 2 (BEI NR-53821) were used for virus titration, neutralization assays and peptide cytotoxicity assays. A549 cells were grown and maintained in grown in Dulbecco's Modified Eagles Medium (DMEM) (Gibco, UK) supplemented with 10% foetal calf serum (FCS), 1 mM sodium pyruvate and 2 mM L-glutamine,1% Non-Essential Amino acids solution (Sigma) and 100 µg/mL Blasticidin (InvivoGen, France). Cells were grown in an environment enriched with $CO_2$ (5%) at 37 °C and passaged every 2–3 days or at approximately 80% confluence; the number of passages did not exceed 6.

SARS-CoV-2, isolate hCov-19/England/204820464/2020 (Lineage B.1.1.7) was kindly made available by Public Health England through the BEI resources repository National Institute of Allergy and Infectious Diseases and managed by the American Type Culture Collection (ATCC). The virus was propagated in Vero E6 cells for 72 h at 37 °C and 5% $CO_2$ in DMEM supplemented with 2% FCS before recovery and storage (−70 °C) of cell-free virus. Viruses from second and third cell passages were used in all experiments. Infectivity in A549 cells was estimated by measurement of 50% tissue culture infective dose ($TCID_{50}$) values of the cell-free viral supernatants using the Reed-Muench method [45,46].

The anti-RBD human neutralising antibody CV30 [47], used as reference compound in antiviral assays, was purchased from Absolute Antibody (Cleveland, UK).

Cytotoxicity assays: Peptide cytotoxicity was assessed using the Vybrant® MTT Cell Proliferation Assay Kit (Thermofisher Scientific) by comparison of cell viability in the presence vs absence of peptide. Vero cells were seeded at $3 \times 10^4$ per well for 24 h. Cells were then washed with PBS and challenged with serial dilutions of the AM29A5-syn peptide in DMEM for 24 h at the range of concentrations tested in antiviral assays. After 24 h, the supernatant was removed and cells were treated with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) in 100 µL of DMEM (free of phenol red) for 4 h at 37 °C in 5% $CO_2$. Finally, 100% DMSO was added for formazan dissolution at 37 °C before absorbance assessment at 540 nm. Viability was estimated by comparison with cells exposed to medium alone and assays were performed in triplicate.

Virus neutralization assays: A549-hACE2 cells were seeded at $3 \times 10^4$ per well in DMEM supplemented with 10% FCS and incubated for 24 h at 37 °C and 5% $CO_2$. On the day of the experiment A549-hACE2 cells were challenged with 10-fold serial dilutions of cell-free SARS-CoV-2 (starting concentration $10^8$ PFU/mL) pre-incubated with serial dilutions of AM29A5-syn peptide (0.008 µM – 25 µM) in DMEM with 2% FCS for 1 h. A cytotoxicity control condition with peptide alone at the range of test concentrations (without virus) was tested alongside the conditions challenged with virus. A virus alone condition (without peptide) was used as positive control and reference for estimation of infectivity inhibition. The neutralizing antibody CV30 at 2 µg/mL was used as internal reference compound. After the 1 h pre-incubation period, 1.25% Avicell (Sigma) in DMEM+ 2% FCS was added to all wells before the main incubation for 72 h at 37 °C and 5% $CO_2$ to allow for viral replication. After this time, infected cells were fixed with 10% formalin (Sigma) for 3 h, before staining with crystal violet in 20% ethanol (Sigma) for 30 min. Finally, wells were washed two times with water and left to dry before plaque counting under light microscopy for quantification of plaque-forming units (PFU). Each condition was performed in quadruplicate. Two independent experiments were carried out. Infectivity was estimated by $TCID_{50}$ measurements on the basis of the four repeats for each condition using the Reed-Muench method [45,46]. A dose-response curve was generated by plotting the percentage reduction of PFU/mL values in the presence of peptide (relative to virus infectivity without peptide) against peptide concentration. The $IC_{50}$ value was calculated using the log-logistic regression function of the *drc* package in R (https://www.r-project.org/).

### 3. Results

#### 3.1. Androctonus amoreuxi transcriptome

We have assembled an *A. amoreuxi* venom gland transcriptome using Illumina shotgun whole transcriptome sequencing. The assembly is 95.4 Mb with a E90N50 of 2309 bp, with high completeness (>96% complete evolutionary conserved arthropod BUSCO marker genes identified, with 72.2% complete & single copy), and a high proportion of input reads mapping back to the assembly (> 98% all reads; > 96% strand-specific concordant pairs), indicative of a high-quality assembly (Table S3).

A total of 102336 transcripts encoded by 75175 coding and non-coding genes were identified in the *A. amoreuxi* venom gland transcriptome. Many transcripts were rare (TPM < 1), with 32172 transcripts (23663 genes) expressed at TPM > 1 (mean TPM 30; median TPM 2), including 3819, 536 and 106 transcripts expressed at low (>10 & <100 TPM), moderate (>100 & <1000 TPM) or high (> 1000 TPM) levels, respectively (Supplementary Info File 2). Predicted peptides and proteins range in size from 99 to 5196 amino acids, with a mean and median length of 379 and 248 amino acids, respectively.

A total of 10509 *A. amoreuxi* transcripts were assigned biological functions from sequence identity with database homologues, and over half of these genes have highest identity with an arthropod gene (Table S3; Supplementary Info File 3). Predicted proteins include 56

transposable element related proteins. The protein products of over 2000 *A. amoreuxi* venom gland transcripts are predicted to have ion channel, transporter or toxin activity (Supplementary Info File 3). Many putative novel proteins were identified in the *A. amoreuxi* transcriptome with 4205 unannotated genes containing an open reading frame (Table S3; Supplementary Info File 3). Structural and functional domains, and signal peptide sequences were identified in transcripts with or without functional annotation (Table S3; Supplementary Info File 3).

Sixteen *A. amoreuxi* transcripts encode proteins that are putative orthologues of previously reported antimicrobial peptides (Supplementary Info file 4).

### 3.2. Characterization of selected A. amoreuxi venom peptides

The transcriptome analysis demonstrates *A. amoreuxi* venom is rich in peptides with putative important functional properties. Reversed phase high performance liquid chromatography (RP-HPLC) was used to fractionate the crude *A. amoreuxi* venom and to purify individual peptides (Fig. 1). The amino acid sequences of the purified peptides were identified based on their MS fragmentation pattern and the obtained transcriptomic data of *A. amoreuxi*. A total of twelve previously undescribed peptides with chain length spanning 29–65 amino acid residues were isolated and identified. Table 1 shows the sequences of these peptides, and the encoding transcripts with their abundances. The monoisotopic mass peaks of these peptides are shown in Fig. 2 while the MS fragmentation analyses are shown in Figs. S2, S4, S6, S8, S11, S13, S17, S21 and the list of MS data is provided in the Supplementary Info File 5. Minor levels of deamidation were noted in samples AM42A5–1, AM42A5–2 and AM51A4. In most cases, the signal peptide was deduced from the alignment of the precursor peptides with homologous peptides (Figs. S1, S3, S5, S7, S9, S10, S12, S14, S15, S16, S18, S20) and using SignalP [48]. Sufficient quantities of three peptides allowed analysis by

Circular dichroism (CD) spectroscopy. CD-UV spectra were recorded for peptides AM42A5–1, AM44A9, AM53A6 (Fig. 3). The obtained spectra showed that these peptides contain both α-helical and β-sheets secondary structure elements. This was further shown by generating 3D homology models for these peptides using SWISS-MODEL [42]. The homology models suggested that the conserved disulfide bridges are critical for the folding of these peptides (Fig. 3). Unfortunately, the limited quantities (<0.1 mg) available for the other peptides were insufficient for CD-UV analysis.

A *C*-terminus amidated derivative of peptide AM29A5 was synthesized and correctly folded in buffer condition 4 (Table S2, Figs. S22-S23). The CD-UV data of the synthetic peptide AM29A5-syn was acquired and shown in Fig. S24.

### 3.3. Venom peptides as inhibitors of RBD binding to hACE2

Surface plasmon resonance competition assays were used to evaluate the venom-derived peptides as inhibitors of RBD binding to hACE2. Eight peptides were selected for testing based on the quantities available and the findings of our initial screening of binding of partially purified samples to RBD (data not shown), namely AM28, AM29A3, AM29A4, AM29A5 (and the respective synthetic peptide AM29A5-syn), AM42A5–2, AM44A9, AM53A6 and AM53A7. Inhibition assays were carried out by immobilization of RBD on the sensor chip surface and binding measurement of fluid phase hACE2 at a fixed concentration in presence of pure peptides at varying concentrations.

Seven out of eight peptides showed some level of dose-dependent inhibition of hACE2 binding to RBD: AM29A3, AM29A4, AM29A5/ AM29A5-syn, AM42A5–2, AM44A9, AM53A6 and AM53A7. The inhibitory activity of two representative peptides (AM53A6 & AM29A5) is shown in Fig. 4. The values of the half-maximal inhibitory concentration ($IC_{50}$) which is the concentration of peptide that reduced hACE2
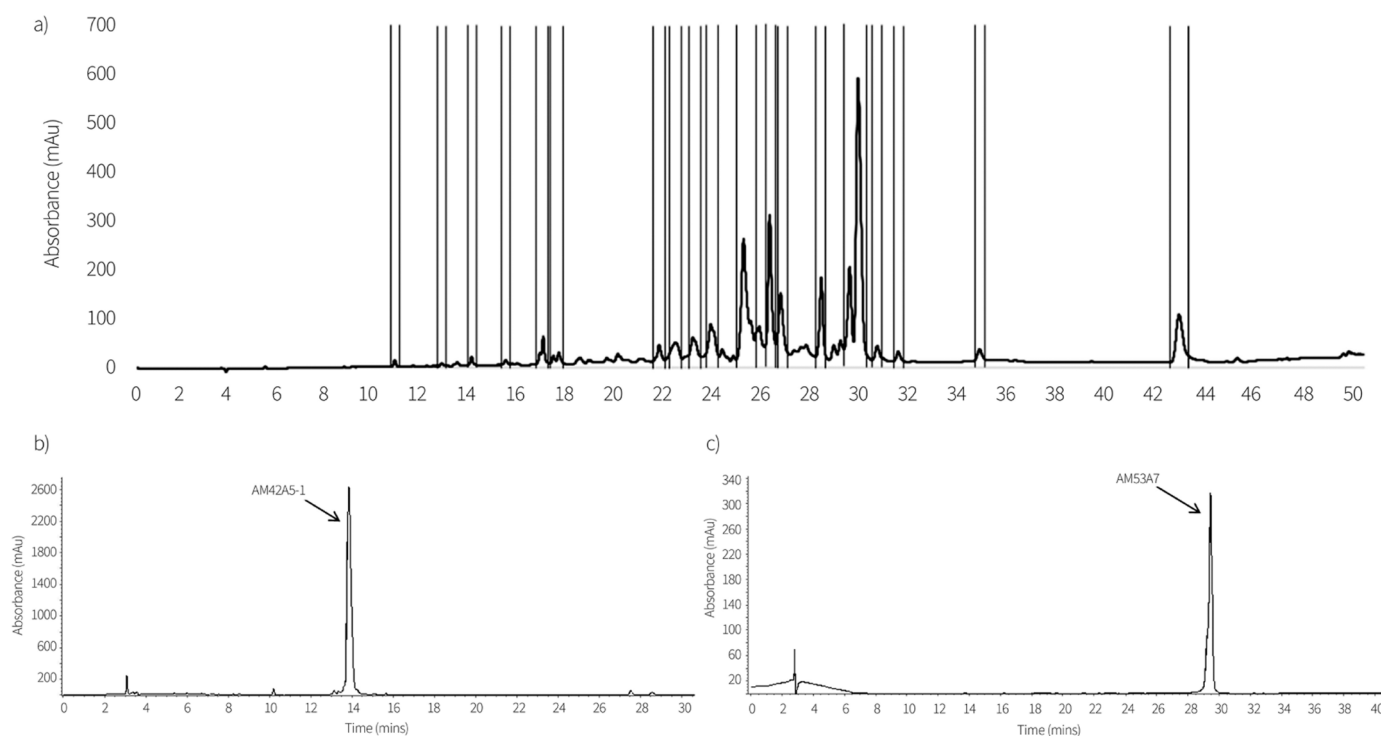


**Fig. 1.** a) HPLC chromatogram as recorded at 220 nm for the total *A. amoreuxi* venom using the C4 column (ACE 5 C4–300, 5 μm, 250 × 10 mm, 300 Å) and a starting linear gradient from 0% to 55% of solution B (0.1% TFA in acetonitrile) in solution A (0.1% TFA in water) over 40 min which was increased linearly to 100% solution B over the following 5 min and the composition was kept at 100% solution B until the end of the 50 min-run. The mobile phase was running at a flow rate of 3 mL/min and the column temperature was at 25ºC. Vertical lines are drawn around the peaks collected. Collected fractions were further purified using the C18 column (ACE 5 C18-300, 5 μm, 250 × 10 mm, 300 Å). b) HPLC chromatogram of the purified peptide AM42A5–1. c) HPLC chromatogram of the purified peptide AM53A7.

**Table 1**

Sequences of the identified pure peptides from *A. amoreuxi*.

| Peptide | Exact Molecular mass (Da) | Transcript isoform | peptide* | Figure |
|---|---|---|---|---|
| AM28 | 3176.24 | TRINITY_DN2961_c0_g1_i1 | VSCEDCPEHCSTQKAQAKCDNDKCVCEPI | S1-S2 |
| AM29A3 | 4142.97 | TRINITY_DN587_c0_g2_i1 | GVEINVKCTGSPQCIKPCKDAGMRFGKCINRKCHCTPK | S3-S4 |
| AM29A4 | 4018.85 | TRINITY_DN587_c0_g1_i2 | VGIPVSCKHSGQCLKPCKDAGMRFGKCMNRKCDCTPK | S5-S6 |
| AM29A5 | 3993.71 | TRINITY_DN24_c0_g1_i3 | GLIDVKCYATSQCWAPCKKETGSGQSKCQNNQCRCY | S7-S8 |
| AM42 | 7153.12 | TRINITY_DN1455_c0_g1_i2 | NDVGYLKAEDGCPIICSRENDSCDSLCKSAGGSSGNCM FFVHVCWCKDLPEDRIWKLKTDTCGRS | S9 |
| AM42A5-1 | 6717.00 | TRINITY_DN290_c0_g1_i3 | DDVPGNYPVNSSGKRYFCPTLGDSEYCIDICKKHGVKY GYCYAFQCWCEYLEDKNVKI# | S10-S11 |
| AM42A5-2 | 6801.01 | TRINITY_DN290_c0_g1_i2 | NDVPGNYPLDSSGKRYFCETLGEHEDCIHICKKHGVQY GYCYAFQCWCEYLKDENVKI# | S12-S13 |
| AM44A8-1 | 6641.78 | TRINITY_DN409_c0_g1_i2 | DGYPMGSDRCKIGCVINHKFCDIECKSQGGSYGYCYFP ACYCEGLPDEKAWKPETNKCGS | S14 |
| AM44A9 | 6758.09 | TRINITY_DN621_c0_g1_i4 | VRDGYIVYPNNCVYHCVPACDGLCKKNGGSSGSCSFLI PSGLACWCKDLPDNVPIKDTSRKCT | S15 |
| AM51A4 | 6890.15 | TRINITY_DN290_c0_g1_i4 | NDVPGNYPLKPSRKRYFCPTLGDSDYCVHICRKHGVQ YGYCYLFTCWCEYLEDKNVQI# | S16-S17 |
| AM53A6 | 6572.94 | TRINITY_DN290_c0_g1_i12 | ADVPGNYPLDSSDNTYLCAPLGDNPSCVEICRKHGVQ YGYCYAFQCWCEYLKDKNIKI | S18-S19 |
| AM53A7 | 5452.42 | TRINITY_DN290_c0_g1_i4 | KRYFCPTLGDSDYCVHICRKHGVQYGYCYLFTCWCEY LEDKNVQI | S20-S21 |

* Cysteine residues that are connected via disulfide bridge have the same color.

# deamidation.
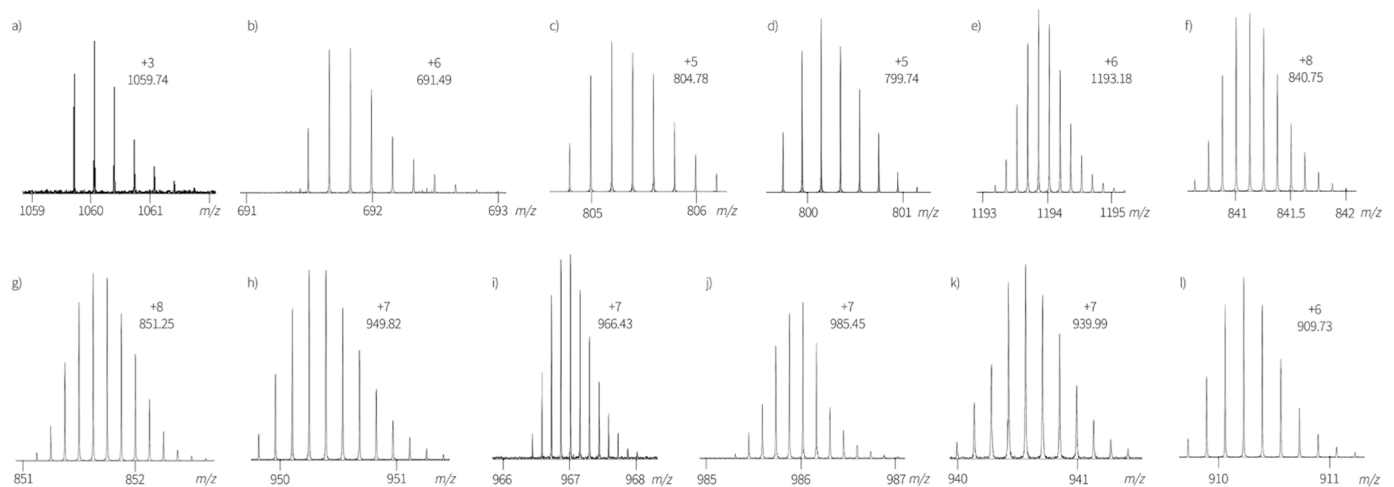


**Fig. 2.** Monoisotopic mass peak for peptide a) AM28, b) AM29A3, c) AM29A4, d) AM29A5, e) AM42, f) AM42A5–1, g) AM42A5–2, h) AM44A8–1, i) AM44A9, j) AM51A4, k) AM53A6, and l) AM53A7.

binding to RBD to 50% when compared with binding of hACE2 alone ranged from 117 μM for AM29A5/AM29A5-syn to 1202 μM for AM44A9, although the $IC_{50}$ could not be determined for AM29A3 and AM53A7 due to insufficient data points (Table 2). AM28 did not show dose-dependent inhibition in this assay at concentration as high as 288 μM (data not shown). To measure the direct binding activity of the peptides to RBD, fluid phase peptides alone were injected at varying concentrations on immobilized RBD. Peptides AM29A5 / AM29A5-syn and AM53A6 showed dose-dependent binding to RBD with affinity of 280 μM and 3.3 mM, respectively (Table 2) as determined by sensor-gram fitting to the 1:1 binding model. Peptides AM29A3, AM29A4, AM42A5–2, AM44A9 and AM53A7 demonstrated binding to RBD only at the higher end of the range of concentrations tested (Table 2). In these instances, binding kinetics could not be evaluated due to limitations on
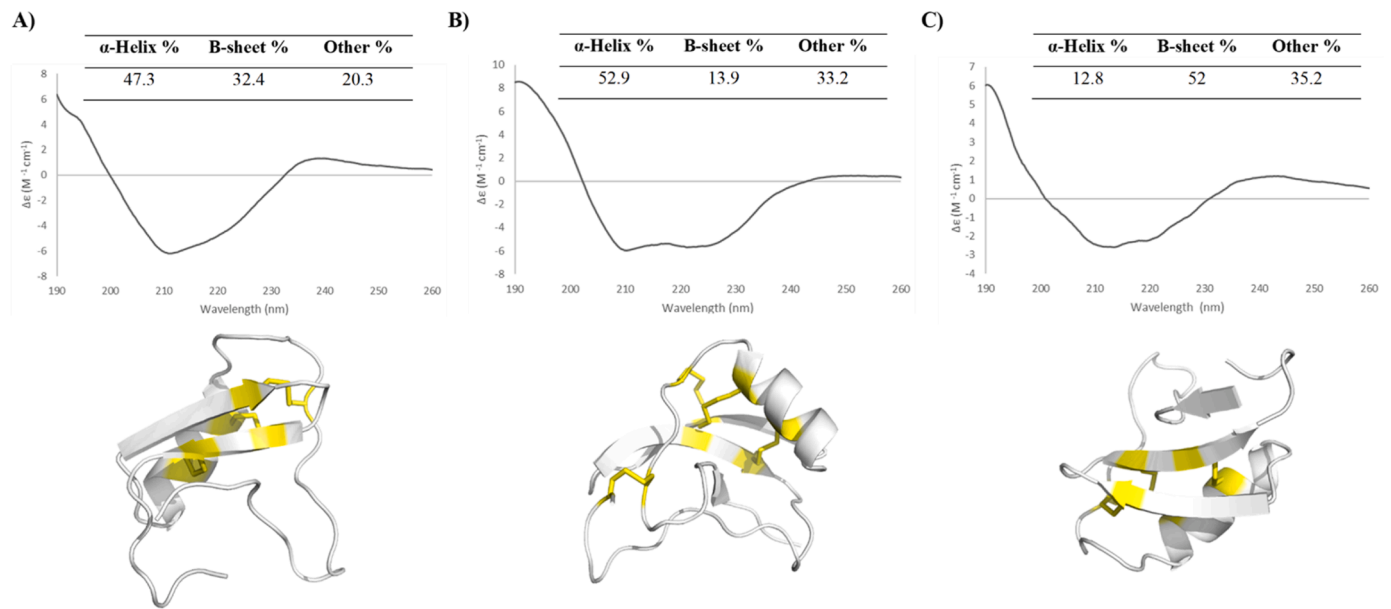
**Fig. 3.** Top panel: CD-UV data for peptide A) AM42A5–1; B) AM44A9 and C) AM53A6. Bottom panel: The 3D structure of A) peptide AM42A5–1 as modelled using scorpion neurotoxin (variant 3) (PDB: 2SN3; Xray crystal structure- Resolution: 1.20 Å) as a homology template; QMEAN = –1.56; B) peptide AM44A9 as modelled using Lqh3 (PDB: 1BMR; NMR solution structure) as a homology template; QMEAN = –2.58 and C) The 3D structure of the peptide AM53A6 as modelled using CN12 neurotoxin (PDB: 1PE4; NMR solution structure) as a homology template; QMEAN = –4.27.
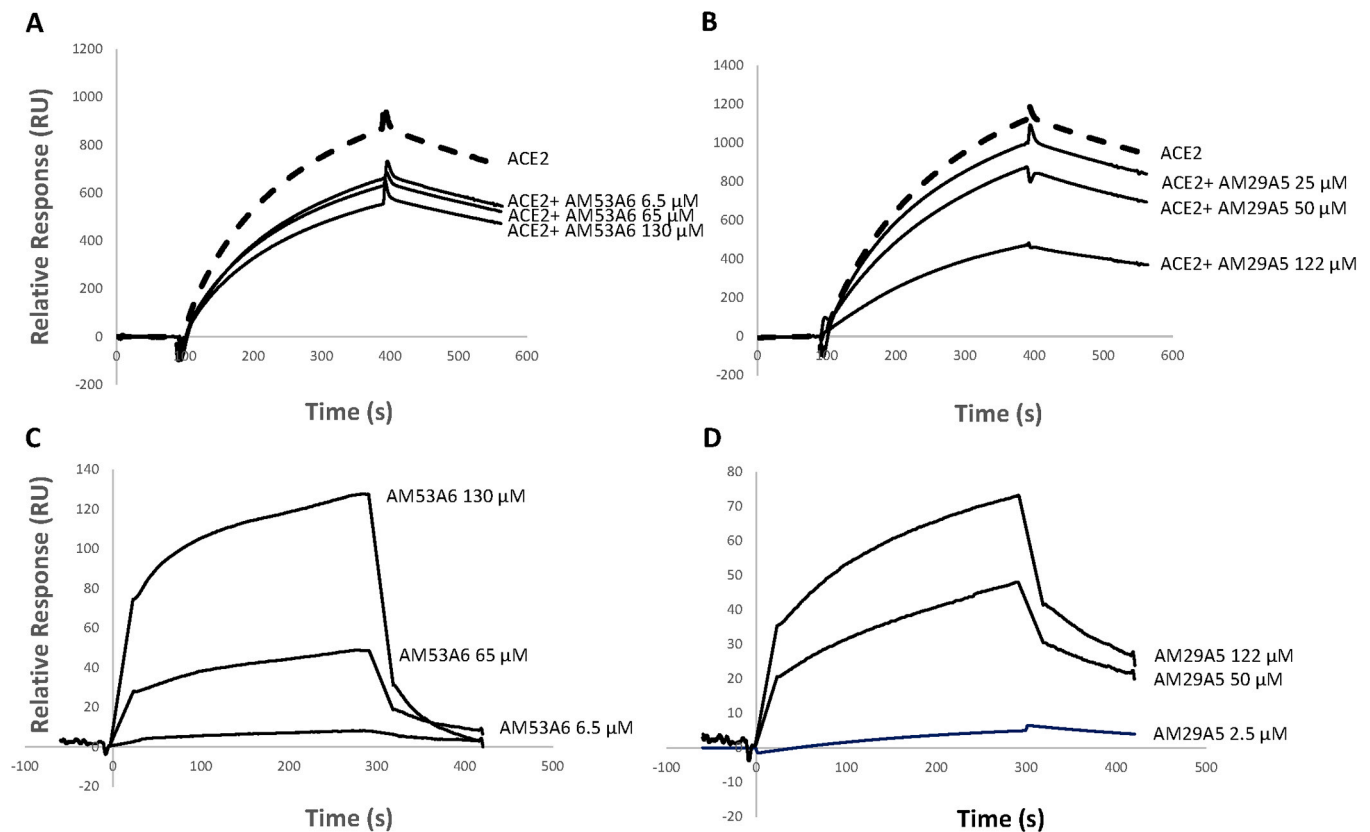


**Fig. 4.** Inhibitory activity of venom peptides on RBD: hACE2 interaction and direct binding activity of peptides to RBD. RBD was immobilized on the sensorchip by nickel affinity and subsequent covalent linking. A-B) To assess inhibition of RBD: hACE2 interaction by the peptides, binding activity of fluid phase hACE2 in presence of peptides was determined and compared to hACE2 alone. Superimposed sensorgrams representing hACE2 (150 nM) binding activity in absence (dashed curve) and presence (solid curves) of (A) peptide AM53A6 (6.5 – 130 μM) and (B) peptide AM29A5 (25 – 122 μM). C-D) Superimposed sensorgrams representing dose-dependent binding activity of (C) AM53A6 (6.5 – 130 μM) and (D) AM29A5 (2.5 – 122 μM) to immobilized RBD. Sensorgrams were fitted to a 1:1 binding model for affinity measurements to RBD.

**Table 2**
RBD - hACE2 binding inhibition and RBD binding measurements of venom-derived pure peptides.

| Peptide | Concentration range (μM) | RBD - hACE2 binding inhibition | Binding to RBD |
|---|---|---|---|
| AM28 | 28.8 – 288.0 | No | No |
| AM29A3 | 43.4 – 434.0 | IC$_{50}$ not determined (14% at 43.4 μM, 27% at 434 μM) | Yes (detected only at 434 μM) |
| AM29A4 | 13.2 – 264.0 | IC$_{50}$ = 175 μM | Yes (detected at 264 μM) |
| AM29A5/AM29A5-syn | 2.5 – 245.0 | IC$_{50}$ = 117 μM | Dose-dependent binding $k_a$ (M$^{-1}$ s$^{-1}$) = 14.8 $k_d$ (s$^{-1}$) = 4.1E$^{-3}$ $K_D$ (M) = 2.8 × 10$^{-4}$ Chi2 = 8.45 |
| AM42A5-2 | 36.7 – 550.5 | IC$_{50}$ = 838 μM | Yes (at ≥ 367 μM) |
| AM44A9 | 17 – 340 | IC$_{50}$ = 1202 μM | Yes (detected at 340 μM) |
| AM53A6 | 6.5 – 130 | IC$_{50}$ = 328 μM | Dose-dependent binding $k_a$ (M$^{-1}$ s$^{-1}$) = 4.8 $k_d$ (s$^{-1}$) = 1.5 × 10$^{-3}$ $K_D$ (M) = 3.3 × 10$^{-3}$ Chi2 = 12.2 |
| AM53A7 | 6.3 – 63 | IC$_{50}$ not determined (16% at 6.3 μM, 30% at 63 μM) | Yes (detected at 63 μM) |

sample quantity. The binding activity of two representative peptides is shown in Fig. 4. Peptide AM28 did not bind RBD in keeping with the lack of activity of this peptide in inhibition assays (data not shown).

### 3.4. Antiviral activity of peptide AM29A5syn

We undertook SARS-CoV-2 neutralization assays in A549-hACE2 cells to assess the inhibitory activity of peptide AM29A5-syn against replication-competent virus. On virus exposure to AM29A5-syn (0.008 μM – 25 μM), we observed dose-dependent inhibition of viral replication (compared to virus alone) with IC$_{50}$ of 200 nM (Fig. 5A). Infectivity of virus exposed to AM29A5-syn at ≥ 2.5 μM was below the detection limit (>3 log relative to non-exposed virus) as measured by TCID$_{50}$ (Fig. 5B). In MTT dye reduction assays, there was no evidence of reduced cell viability on exposure to the AM29A5-syn peptide at concentrations as high as 1.6 mM (data not shown).

### 4. Discussion

In this manuscript, we characterized twelve previously undescribed peptides from the venom of the scorpion, *A. amoreuxi* (Table 1) using MS fragmentation and venomous gland transcriptome mining. The identified peptides can be categorized into two groups; the short chain peptides with disulfide bridges (AM28, AM29A3, AM29A4 and AM29A5) and the long chain peptides with disulphide bridges (AM42, AM42A5–1,

AM42A5–2, AM44A8–1, AM44A9, AM51A4, AM53A6 and AM53A7) (Table 1). To the best of our knowledge, only ten peptides have been previously identified from the venom of *A. amoreuxi* (Table S1) and the transcriptome or genome of this species has not yet been published. We tested the ability of these peptides to inhibit the critical interaction between the SARS-CoV-2 spike RBD and hACE2 that precedes virus entry into the cell [5] using a surface plasmon resonance-based assay. Several scorpion venom peptides are conformationally constrained with a set of disulfide bridges between Cys residues [49]. These disulfide bridges allow the peptide to fold into the correct 3D conformation that can interact with the receptor. Most disulfide bridge-containing peptides interact specifically with neuronal ion channels, and they are further classified into four families based on the ion channels they target (Na$^+$, K$^+$, Cl$^-$ or Ca$^{2+}$ channels). Sodium channel-specific toxins are usually long-chain peptides with 60–76 amino acid residues in length and are cross-linked with three or four disulfide bridges [49,50]. Potassium channel-specific toxins are made up of peptides with 23–64 residues with three or four disulfide bridges [51] while chloride ion channel-specific toxins are composed of fewer than 40 amino acid residues and contain four disulfide bonds [52]. However, recent research has demonstrated that constrained peptides e.g. macrocyclic or stapled peptides can also inhibit disease-relevant PPIs such as SARS-CoV-2 Spike – hACE2, which are known to be very challenging to small molecule drugs [53,54]. Moreover, the antiviral activity of several scorpion venom peptides against a wide range of viruses including
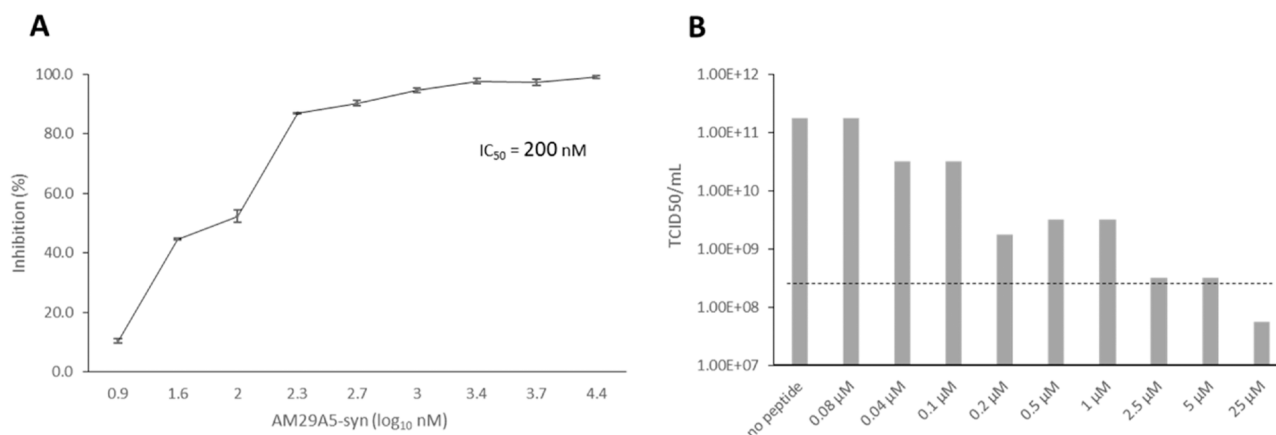


**Fig. 5.** Peptide AM29A5-syn inhibits SARS-CoV-2 replication in A549-hACE2 cells. **A)** Dose-response curve showing % reduction of viral infectivity (PFU/mL) in A549-hACE2 cells challenged with SARS-Cov-2 B.1.1.7 at 10$^8$ PFU/mL in the presence of peptide AM29A5-syn (0.008 μM – 25 μM). Standard deviation is shown ($n = 4$). IC$_{50}$ estimated by log-logistic regression is inset. **B)** Infectivity of virus exposed to peptide AM29A5-syn (0.008 μM – 25 μM) compared to non-exposed virus (control) represented as TCID$_{50}$/mL. Reduction of viral infectivity below the detection limit (>3 log relative to the control in which the peptide was absent) is represented by the horizontal dashed line. One representative dataset of two independent experiments is shown.

SARS-coronavirus have been confirmed and several mechanisms have been suggested [9–17]. A recent molecular docking study [55] has revealed the ability of some known antiviral scorpion venom peptides to interact with the SARS-CoV-2 spike RBD. However, no supporting experimental biological data were provided. These studies and the lack of knowledge on the venom peptides from *A. amoreuxi* were the stimulus for this research.

We report the first annotated reference transcriptome for the *A. amoreuxi* venom gland. Transcriptome annotation and AMP mining have identified a wealth of diverse *A. amoreuxi* venom gland peptides with putative ion channel, transporter or toxin activity, as well as putative orthologues of known AMPs, including some with antiviral activity. This catalogue of potentially medically important gene products provides rich data to support antimicrobial drug development. Further curation and functional analyses of the identified *A. amoreuxi* venom transcripts and encoded peptides with putative AMP activity, along with further exploration and refinement of the full diversity of *A. amoreuxi* toxins, will be the focus of future work [56,57]. Importantly, our *A. amoreuxi* transcriptome has enabled a proteo-transcriptomic approach to antiviral drug discovery, through identification of full-length transcripts encoding SARS-CoV-2 spike protein inhibitors, that were partially resolved by MS fragment sequences.

Out of the twelve venom peptides isolated and identified in this study, eight were tested for their ability to bind to SARS-CoV-2 spike protein and to inhibit its interaction with hACE2 (Table 2, Fig. 4). Peptides AM29A4 and AM29A5 displayed the highest inhibition and lowest $IC_{50}$ whilst AM28 showed no evidence of inhibition. Peptide AM53A7 also showed notable inhibition at relatively low concentrations whilst the effect of this peptide at higher concentrations could not be determined due to limited amount of sample.

This data aligned with findings of direct binding to RBD, notwithstanding that full binding kinetics data could only be determined for peptides AM29A5 and AM53A6. The properties of most highly active peptide AM29A5 were confirmed by testing the chemically synthesized analogous peptide which displayed the same $IC_{50}$ and RBD-binding affinity values as AM29A5. Altogether these findings provide evidence of specific and competitive inhibition of RBD – hACE2 binding by seven of the peptides tested likely occurring, at least in part, through interaction with the hACE2 binding site of the SARS-CoV-2 spike RBD. However, the relatively high $IC_{50}$ in RBD – hACE2 binding assays suggests that inhibition of RBD – hACE2 binding may not be the primary mode of antiviral action.

SARS- CoV-2 RBD – hACE2 is a very challenging target for drug design. For example, in a previous study [58], a series of thirteen different stapled peptidomimetics based on the hACE2 interaction motif were designed to bind the coronavirus S-protein RBD and inhibit binding to the hACE2 receptor. The peptidomimetics were assessed for antiviral activity in an array of assays including a neutralization pseudovirus assay, immunofluorescence assay and in vitro fluorescence polarization assay. However, none of the peptidomimetics showed activity in these assays. Importantly, this highlights the importance of developing an effective strategy to screen biological resources, like scorpion venoms, with proven therapeutic potential in combating microbial and viral pathogens, to identify and test putative antiviral peptides, thus providing evidence-based lead candidates for further exploration and development. The antiviral activity of AM29A5-syn was seen at concentrations well below the concentration range used in the RBD – hACE2 inhibition assay. Inhibition of replication of SARS-CoV-2 B.1.1.7 in hACE2-overexpressing A459 cells was observed at an $IC_{50}$ nearly 600-fold lower than that observed in the RBD – hACE2 surface plasmon resonance assay. This suggests that the anti-SARS-CoV-2 property of this peptide is mediated by additional antiviral mode(s) of action beyond RBD – hACE2 binding inhibition.

Our results show that scorpion venom peptides can inhibit SARS-CoV-2 replication although unlikely through inhibition of the SARS-CoV-2 spike RBD – hACE2 interaction. Scorpion venom peptides represent excellent scaffolds for design of novel anti-SARS-CoV-2 constrained peptides; future studies should explore their antiviral mode of action as well as the structural dynamics of inhibition of target virus-host interactions. Future studies must also assess the biocompatibility of these peptides before pursual into drug development pipelines.

## 5. Conclusion

Collectively, our results show that scorpion venom peptides inhibit SARS-CoV-2 spike protein interaction with hACE2 receptor while exhibiting anti-SARS-CoV-2 activity via additional unexplored modes of action. These peptides can be used as a blueprint to design natural product analogues with higher antiviral potency against SARS-CoV-2 and other coronaviruses. This study also shows the value of using proteo-transcriptomic approaches to structurally identify new venom peptides.

## CRediT authorship contribution statement

**Pattinson Craig:** Data curation, Formal analysis. **Burgoyne Kate:** Data curation, Formal analysis. **Nkene Istifanus:** Formal analysis. **Collie-Duguid Elaina:** Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition. **Heidari Zeynab:** Data curation. **Samir Reham:** Methodology, Resources. **Houssen Wael E.:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Muhammad Taj:** Data curation, Writing – review & editing. **Alfadhel Sanad:** Data curation, Formal analysis, Writing – review & editing. **Hijazi Karolin:** Conceptualization, Data curation, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. **Abdel-Rahman Mohamed A.:** Resources, Supervision, Writing – review & editing, Funding acquisition. **Ribeiro Antonio:** Data curation, Formal analysis, Methodology, Writing – review & editing. **Gerges Mariam M.:** Methodology, Resources. **Ghazal Ahmad:** Data curation, Formal analysis, Writing – review & editing. **Colamarino Rosa A.:** Data curation, Formal analysis. **Clarke David:** Data curation, Formal analysis, Methodology, Writing – review & editing.

## Declaration of Competing Interest

The authors declare no competing financial interest.

## Data availability

The *A. amoreuxi* transcriptome data is available at GEO. Accession number: GSE221799.

support of Digital Research and the Maxwell high performance compute cluster, University of Aberdeen.

## Author contributions

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.peptides.2023.171139.

## References

[1] R.T. Eastman, J.S. Roth, K.R. Brimacombe, A. Simeonov, M. Shen, S. Patnaik, et al., Remdesivir: a review of its discovery and development leading to emergency use authorization for treatment of COVID-19, ACS Cent. Sci. 6 (2020) 672–683.

[2] G.R. Painter, M.G. Natchus, O. Cohen, W. Holman, W.P. Painter, Developing a direct acting, orally available antiviral agent in a pandemic: The evolution of molnupiravir as a potential treatment for COVID-19, Curr. Opin. Virol. 50 (2021) 17–22.

[3] R.P. Joyce, V.W. Hu, J. Wang, The history, mechanism, and perspectives of nirmatrelvir (PF-07321332): an orally bioavailable main protease inhibitor used in combination with ritonavir to reduce COVID-19-related hospitalizations, Med. Chem. Res. 31 (2022) 1637–1646.

[4] J. Huo, A. Le Bas, R.R. Ruza, H.M.E. Duyvesteyn, H. Mikolajek, T. Malinauskas, et al., Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2, Nat. Struct. Mol. Biol. 27 (2020) 846–854.

[5] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C. Hsieh, O. Abiona, et al., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science 367 (2020) 1260–1263.

[6] M. Gao, K. Cheng, H. Yin, Targeting protein-protein interfaces using macrocyclic peptides, Biopolymers 104 (2015) 310–316.

[7] E.A. Villar, D. Beglov, S. Chennamadhavuni, J.A. Porco, D. Kozakov, S. Vajda, et al., How proteins bind macrocycles, Nat. Chem. Biol. 10 (2014) 723–731.

[8] D.M. Krüger, A. Glas, D. Bier, N. Pospiech, K. Wallraven, L. Dietrich, et al., Structure-based design of non-natural macrocyclic peptides that inhibit protein−protein interactions, J. Med. Chem. 60 (2017) 8982–8988.

[9] Q. Li, Z. Zhao, D. Zhou, Y. Chen, W. Hong, L. Cao, et al., Virucidal activity of a scorpion venom peptide variant mucroporin-M1 against measles, SARS-CoV and influenza H5N1 viruses, Peptides 32 (2011) 1518–1525.

[10] P.I. Harrison, M.A. Abdel-Rahman, K. Miller, P.N. Strong, Antimicrobial peptides from scorpion venoms, Toxicon 88 (2014) 115–137.

[11] A.M.H. El-Bitar, M.M.H. Sarhan, C. Aoki, Y. Takahara, M. Komoto, L. Deng, et al., Virol. J. 12 (2015) 47.

[12] É.C.G. da Mata, C.B.F. Mourão, M. Rangel, E.F. Schwartz, Antiviral activity of animal venom peptides and related compounds, J. Venom. Anim. Toxins Incl. Trop. Dis. 23 (2017) 3.

[13] Z. Ji, F. Li, Z. Xia, X. Guo, M. Gao, F. Sun, et al., The scorpion venom peptide Smp76 inhibits viral infection by regulating Type-I interferon response, Virol. Sin. 33 (2018) 545–556.

[14] F. Li, Y. Lang, Z. Ji, Z. Xia, Y. Han, Y. Cheng, et al., A scorpion venom peptide Ev37 restricts viral late entry by alkalizing acidic organelles, J. Biol. Chem. 294 (2019) 182–194.

[15] Y. Cheng, F. Sun, S. Li, M. Gao, L. Wang, M. Sarhan, et al., Inhibitory activity of a scorpion defensin BmKDfsin3 against Hepatitis C Virus, Antibiotics 9 (2020) 33.

[16] A.M.H. El-Bitar, M. Sarhan, M.A. Abdel-Rahman, V. Quintero-Hernandez, C. Aoki-Utsubo, M.A. Moustafa, et al., Smp76, a scorpine-like peptide isolated from the venom of the scorpion *Scorpio maurus palmatus*, with a potent antiviral activity against hepatitis C virus and dengue virus, Int. J. Pept. Res. Ther. 26 (2020) 811–821.

[17] M.A. El Hidan, M.A. Laaradia, O. El Hiba, A. Draoui, A. Aimrane, K. Kahime, Scorpion-derived antiviral peptides with a special focus on medically important viruses: An update, BioMed. Res. Int. 2021 (2021) 9998420.

[18] E.M. Saied, Y.A. El-Maradny, A.A. Osman, A.M.G. Darwish, H.H. Abo Nahas, G. Niedbała, et al., A comprehensive review about the molecular structure of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Insights into natural products against COVID-19, Pharmaceutics 13 (2021) 1759.

[19] T. Liu, S. Luo, P. Libby, G. Shi, Cathepsin L-selective inhibitors: A potentially promising treatment for COVID-19 patients, Pharmacol. Ther. 213 (2020) 107587.

[20] H.K. El-Hennawy, A simplified key to Egyptian scorpions, Serket 1 (1987) 15–17.

[21] M.A. Abdel-Rahman, V. Quintero-Hernandez, L.D. Possani, Venom proteomic and venomous glands transcriptomic analysis of the Egyptian scorpion *Scorpio maurus palmatus* (Arachnida: Scorpionidae), Toxicon 74 (2013) 193–207.

[22] E.F. Schwartz, E. Diego-García, R.C. Rodríguez de laVega, L.D. Possani, Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones), BMC Genom. 8 (2007) 119.

[23] J.A. Guerrero-Vargas, C.B.F. Mourão, V. Quintero-Hernandez, L.D. Possani, E.F. Schwartz, Identification and phylogenetic analysis of *Tityus pachyurus* and *Tityus obscurus* novel putative Na$^+$-channel scorpion toxins, PLoS One 7 (2012) e30478.

[24] S.R. Andrews, et al., FastQC: a quality control tool for high throughput sequence data. Retrieved from Babraham Bioinformatics: Available at: ⟨https://www.bioinformatics.babraham.ac.uk/projects/fastqc/⟩ (2010).

[25] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics 32 (2016) 3047–3048.

[26] F. Krueger, F. James, P. Ewels, E. Afyounian, B. Schuster-Boeckler, Zenodo (2021) (https://doi.org/10.5281/zenodo.5127899).

[27] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Full-length transcriptome assembly from RNA-seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652.

[28] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, De novo transcript sequence reconstruction from RNA-seq: reference generation and analysis with Trinity, Nat. Protoc. 8 (2013) 1494–1512.

[29] M. Manni, M.R. Berkeley, M. Seppey, E.M. Zdobnov, BUSCO: Assessing genomic data quality and beyond, Curr. Protoc. 1 (2021) e323.

[30] E. Bushmanova, D. Antipov, A. Lapidus, V. Suvorov, A. Prjibelski, rnaQUAST: a quality assessment tool for de novo transcriptome assemblies, Bioinformatics 32 (2016) 2210–2212.

[31] S. Tang, A. Lomsadze, M. Borodovsky, Identification of protein coding regions in RNA transcripts, Nucleic Acids Res 43 (2015) e78.

[32] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, *Salmon*: fast and bias-aware quantification of transcript expression using dual-phase inference, Nat. Methods 14 (2017) 417–419.

[33] B. Langmead, S. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359.

[34] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., The Sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.

[35] H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, Bioinformatics 27 (2011) 2987–2993.

[36] R Core Team, R: A language and environment for statistical computing. (R Foundation for Statistical Computing) Retrieved March 15, 2022, from R Foundation for Statistical Computing: ⟨http://www.R-project.org/⟩ (2019).

[37] RStudio Team, RStudio: Integrated Development for R. (RStudio, Inc.) Retrieved May 02, 2018, from RStudio, Inc.: ⟨http://www.rstudio.com/⟩ (2015).

[38] H. Wickham, ggplot2: elegant graphics for data analysis, Springer-Verlag New York, New York, 2009.

[39] O. Wagih, ggseqlogo: a versatile R package for drawing sequence logos, Bioinformatics 33 (2017) 3645–3647.

[40] D.M. Bryant, K. Johnson, T. DiTommaso, T. Tickle, M.B. Couger, D. Payzin-Dogru, et al., A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors, Cell Rep. 18 (2017) 762–776.

[41] J. Jhong, L. Yao, Y. Pang, Z. Li, C. Chung, R. Wang, et al., dbAMP 2.0: updated resource for antimicrobial peptides with an enhanced scanning method for genomic and proteomic data, Nucleic Acids Res 50 (2022) D460–D470.

[42] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, et al., SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res 46 (2018) W296–W303.

[43] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

[44] A. Micsonai, É. Moussong, F. Wien, E. Boros, H. Vadászi, N. Murvai, et al., BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy, Nucleic Acids Res 50 (2022) W90–W98.

[45] L.J. Reed, H. Muench, A simple method of estimating fifty percent endpoints, Am. J. Hyg. 27 (1938) 493–497.

[46] C. Lei, J. Yang, J. Hu, X. Sun, On the calculation of TCID$_{50}$ for quantitation of virus infectivity, Virol. Sin. 36 (2021) 141–144.

[47] N.K. Hurlburt, E. Seydoux, Y. Wan, V.V. Edara, A.B. Stuart, J. Feng, et al., Structural basis for potent neutralization of SARS-CoV-2 and role of antibody affinity maturation, Nat. Commun. 11 (2020) 5413.

[48] F. Teufel, J.J.A. Armenteros, A.R. Johansen, M.H. Gíslason, S.I. Pihl, K.D. Tsirigos, et al., SignalP 6.0 predicts all five types of signal peptides using protein language models, Nat. Biotechnol. 40 (2022) 1023–1025.

[49] S. Ahmadi, J.M. Knerr, L. Argemi, K.C.F. Bordon, M.B. Pucca, F.A. Cerni, et al., Scorpion venom: detriments and benefits, Biomedicines 8 (2020) 118.

[50] E. Ortiz, G.B. Gurrola, E.F. Schwartz, L.D. Possani, Scorpion venom components as potential candidates for drug development, Toxicon 93 (2015) 125–135.

[51] R.C. Rodriguez de la Vega, L.D. Possani, Current views on scorpion toxins specific for K$^+$-channels, Toxicon 43 (2004) 865–875.

[52] J.A. DeBin, J.E. Maggio, G.R. Strichartz, Purification and characterization of chlorotoxin, a chloride channel ligand from the venom of the scorpion, Am. J. Physiol. 264 (1993) C361–C369.

[53] A. Henninot, J.C. Collins, J.M. Nuss, The current state of peptide drug discovery: Back to the future? J. Med. Chem. 61 (2018) 1382–1414.

[54] H. Wang, R.S. Dawber, P. Zhang, M. Walko, A.J. Wilson, X. Wang, Peptide-based inhibitors of protein–protein interactions: biophysical, structural and cellular consequences of introducing a constraint, Chem. Sci. 12 (2021) 5977–5993.

[55] K. Mahnam, M. Lotfi, F.A. Shapoorabadi, Examining the interactions scorpion venom peptides (HP1090, Meucin-13, and Meucin-18) with the receptor binding domain of the coronavirus spike protein to design a mutated therapeutic peptide, J. Mol. Graph. Model 107 (2021) 107952.

[56] M.L. Holding, M.J. Margres, A.J. Mason, C.L. Parkinson, D.R. Rokyta, Evaluating the performance of de novo assembly methods for venom-gland transcriptomics, Toxicon 10 (2018) 249.

[57] K. Sunagar, D. Morgenstern, A.M. Reitzel, Y. Moran, Ecological venomics: How genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom, J. Proteom. 135 (2016) 62–72.

[58] D.C. Morgan, C. Morris, A. Mahindra, C.M. Blair, G. Tejeda, I. Herbert, et al., Stapled ACE2 peptidomimetics designed to target the SARSCoV-2 spike protein do not prevent virus internalization, Pept. Sci. 113 (2021) e24217.