Journal of proteome research

Article

pubs.acs.org/jpr

Discovery of Novel Antimicrobial Peptides from *Varanus komodoensis* (Komodo Dragon) by Large-Scale Analyses and De-Novo-Assisted Sequencing Using Electron-Transfer Dissociation Mass Spectrometry

Barney M. Bishop,*[,†] Melanie L. Juba,[†] Paul S. Russo,[‡] Megan Devine,[†] Stephanie M. Barksdale,[§] Shaylyn Scott,[†] Robert Settlage,[∥] Pawel Michalak,[⊥] Kajal Gupta,[#] Kent Vliet,[▽] Joel M. Schnur,[#] and Monique L. van Hoek[§]

[†]Department of Chemistry and Biochemistry, George Mason University, 10920 George Mason Circle, 4C7, Manassas, Virginia, 20110, United States

[‡]Center for Applied Proteomics and Molecular Medicine, George Mason University, 10920 George Mason Circle, 1A9, Manassas, Virginia 20110, United States

[§]School of Systems Biology, George Mason University, 10920 George Mason Circle, 1H8, Manassas, Virginia 20110, United States

[∥]Advanced Research Computing, Virginia Polytechnic Institute and State University, 620 Drillfield Drive, Blacksburg, Virginia 24061, United States

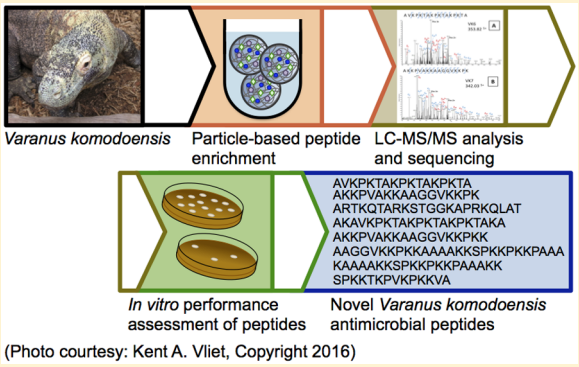
[⊥]Biocomplexity Institute, Virginia Polytechnic Institute and State University, 1015 Life Science Circle, Blacksburg, Virginia 24061, United States

[#]College of Science, George Mason University, 4400 University Drive, 5C3, Fairfax, Virginia 22030, United States

[▽]Department of Biology, University of Florida, 876 Newell Drive, PO Box 118525, Gainesville, Florida 32511, United States

Supporting Information

ABSTRACT: Komodo dragons are the largest living lizards and are the apex predators in their environs. They endure numerous strains of pathogenic bacteria in their saliva and recover from wounds inflicted by other dragons, reflecting the inherent robustness of their innate immune defense. We have employed a custom bioprospecting approach combining partial de novo peptide sequencing with transcriptome assembly to identify cationic antimicrobial peptides from Komodo dragon plasma. Through these analyses, we identified 48 novel potential cationic antimicrobial peptides. All but one of the identified peptides were derived from histone proteins. The antimicrobial effectiveness of eight of these peptides was evaluated against *Pseudomonas aeruginosa* (ATCC 9027) and *Staphylococcus aureus* (ATCC 25923), with seven peptides exhibiting antimicrobial activity against both microbes and one only showing significant potency against *P. aeruginosa*. This study demonstrates the power and promise of our bioprospecting approach to cationic antimicrobial peptide discovery, and it reveals the presence of a plethora of novel histone-derived antimicrobial peptides in the plasma of the Komodo dragon. These findings may have broader implications regarding the role that intact histones and histone-derived peptides play in defending the host from infection. Data are available via ProteomeXChange with identifier PXD005043.



(Photo courtesy: Kent A. Vliet, Copyright 2016)

KEYWORDS: *antimicrobial, Komodo dragon, peptides, microparticles, mass spectrometry, electron transfer dissociation, PEAKS db, de novo sequencing, de-novo-assisted sequencing*

INTRODUCTION

The Komodo dragon (*Varanus komodoensis*) is the largest living lizard and inhabits five small islands in southeast Indonesia. These large, carnivorous lizards dominate their ecosystem, hunting invertebrates, birds, and mammals. Although Komodo dragons will ambush and kill their prey directly, they will also feed on carrion or carcasses killed by other dragons and animals. In addition to direct physical attack, bacteria or venom within the dragons' bite could contribute to the ultimate demise of its prey.[1−4] In a study analyzing the saliva and oral cavities of multiple dragons in the wild, 57 species of bacteria were identified,

Received: September 27, 2016
Published: February 6, 2017

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

with over 90% being classified as potentially pathogenic.[2] The ability of Komodo dragons to endure numerous strains of pathogenic bacteria in their saliva yet show no signs of infection and furthermore recover from wounds inflicted by other dragons reflects the inherent advantage of their innate immune defense and demands further investigation.[1,2] Moreover, Komodo dragon serum has been demonstrated to have in vitro antibacterial properties against both Gram-negative and Gram-positive bacteria.[5]

Cationic antimicrobial peptides (CAMPs) are produced by nearly all living organisms and are an essential part of the innate immune defense in higher organisms.[6−8] Recently, interest in CAMPs has grown as antimicrobial resistance continues to spread and associated risks become greater. These peptides have the ability to exert direct broad-spectrum antimicrobial, antiviral, and antifungal properties.[9] CAMPs tend to be low-molecular-weight peptides that are both highly cationic and amphipathic in nature. These physicochemical properties allow CAMPs to directly interact with pathogens and exert broad-spectrum effectiveness. Although these peptides are used pervasively in nature and are evolutionarily ancient, limited bacterial resistance has been observed.[4] However, the known CAMP library is limited, and peptides identified from domesticated and laboratory animals and primates are over-represented relative to the potential species diversity that is present in the animal kingdom. Thus identifying novel CAMPs from unexplored species is essential to further their therapeutic potential. The role that CAMPs play in the innate immunity of the Komodo dragon is potentially very informative, and the newly identified Komodo dragon CAMPs may lend themselves to the development of new antimicrobial therapeutics.

The current approaches that have been used to discover and identify native CAMPs from biological samples have proven to be slow and low-yielding.[10−14] Recently, we developed a new and effective method for CAMP identification that allows for the rapid extraction and analysis of the native, functional peptidome.[15,16] This method uses custom-made microparticle harvesting of intact, functional peptides from biological samples coupled to de novo sequencing of the harvested peptides using electron-transfer dissociation (ETD) mass spectrometry. In this approach, we have been able to more effectively leverage the high sensitivity of mass spectrometry and the ability to sequence native peptides based on their MS/MS fragmentation than is possible using conventional chromatography, electrophoresis and fractionation methods.

Here we report the application of our bioprospecting peptide discovery approach to harvest, extract, and elucidate the sequences of novel Komodo dragon CAMPs. Initial analyses of the harvested peptides by mass spectrometry revealed a highly complex mixture of peptides. Manual and PEAKS de-novo-assisted sequencing of this complicated mixture was accomplished based on high-resolution ETD MS/MS data. Using the automated PEAKS sequencing software allowed for rapid determination of peptide sequences that can then be manually verified. Access to a transcriptome, which can be used to assist PEAKS in sequence determination, is very helpful in cases like Komodo, where a full genome sequence is not available. Through analysis of aliquots of stimulated and unstimulated Komodo dragon plasma (100 μL of each), we were able to identify eight potential novel Komodo dragon CAMPs. We subsequently assessed the antimicrobial effectiveness of these eight novel peptides against Gram-negative *Pseudomonas aeruginosa* and Gram-positive *Staphylococcus aureus*.

■ MATERIALS AND METHODS

RNA Preparation

RNA was extracted from leukocyte-enriched (2 h settling) fresh Komodo dragon blood (St. Augustine Alligator Farm Zoological Park, St. Augustine, FL) using the QIAamp RNA Blood Mini KitQiagen (Qiagen) and submitted to the Biocomplexity Institute at Virginia Polytechnic Institute and State University (Blacksburg, VA) for sequencing. RNA quality was checked by examining the RNA preparation on an agarose gel and determining the OD260/280 nm following manufacturer's instructions.

Stranded RNA-Seq Library Construction using Wafergen Apollo 324

Sequencing library preparations were performed on the Apollo 324 Robot (Wafergen, CA). The quality of total RNA was checked on an Agilent BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA). One μg of total RNA was depleted of globin and rRNAs using Illumina's Globin-Zero Gold Kit (Illumina, P/N GZG 1224). Poly-A RNA was then converted into a library of template molecules using PrepX RNA-Seq for Illumina Library Kit (Wafergen, Fremont, CA) for subsequent cluster generation and sequencing by Illumina HiSeq. In brief, poly-A mRNA was fragmented into smaller pieces (~140 nt). 3′ and 5′ adapters were ligated to the cleaved RNA fragments and converted to first-strand cDNA using reverse transcriptase, followed by second-strand synthesis. The products were then purified and enriched with 13 cycles of PCR to create the final cDNA library. The 280−300 bp libraries (160−180 bp insert) generated were validated using Agilent 2100 Bioanalyzer and quantitated using a Quant-iT dsDNA HS Kit (Invitrogen) and qPCR.

Cluster Generation and HiSeq HighOutput Sequencing Paired End Read

Libraries were clustered onto a flow cell using Illumina's TruSeq PE Cluster Kit v3-cBOT-HS and sequenced using TruSeq SBS Kit v3-HS (300 cycles, 2 × 150 cycle paired-end) on an Illumina HiSeq 2500.

Transcriptome Assembly

Following sequencing, data were trimmed for both adaptor and quality using a combination of ea-utils and Btrim.[17,18] Data were then assembled using Trinity following the published protocol.[19,20] TransDecoder within Trinity was used to identify candidate coding sequences within assembled transcripts, and transcripts with open reading frame (ORF) lengths less than 300 base pairs (or 100 amino acids) were filtered out (31 791 left).

Peptide Harvest and Elution

Harvests were performed using a combination of poly-*N*-isopropylacrylamide-based (NIPAm) particles that we had successfully employed in the identification of novel CAMPs from the American alligator, one being a core−shell particle incorporating acrylic acid as its affinity bait and the other a shell-less particle with acrylic acid and 2-acrylamido-2-methylpropanesulfonic acid as baits.[15,16] These particles were designed to complement the physicochemical properties associated with CAMPs, which are generally small, cationic, and amphipathic. Accordingly, the particles capture the peptides primarily through electrostatic interactions, while the cross-linking of the polymer helps to exclude larger proteins. To enhance the harvesting preference for CAMP-like peptides, which generally exhibit pI values > pH 9, harvests were

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

performed at pH 9. Following harvest, the particles were recovered and subsequently washed to remove unbound proteins and peptides. Captured peptides were then eluted using a mixed organic/aqueous elution solution (1:1 trifluoroethanol/water with 0.1% trifluoroacetic acid), and the recovered peptides were then desalted in preparation for LC−MS/MS analyses.

Plasma was collected from Komodo dragon blood that had been stimulated in vitro by incubating it with LPS (1 μg/mL, 30 min, 30 °C) as well as blood that had not been stimulated. Harvests were performed from both the stimulated and unstimulated plasma. In each case, plasma (100 μL) was diluted into 1.6 mL of hydrogel particles (20 mg of AAc core−shell particles and 20 mg of the AAc/MPs core particles) suspended in 10 mM Tris-Cl buffer (particle suspension = pH 9) for a final volume of ∼1.7 mL. After incubating approximately 18−24 h at room temperature, the plasma−particle harvest mixture was centrifuged at 16.1×10^3 rcf to pellet the particles, and the pelleted particles were then resuspended in 10 mM Tris-Cl buffer (pH 9.0). This centrifugation and resuspension process was repeated to ensure removal of excluded proteins and peptides. Following the final wash with Tris-Cl buffer, the pelleted particles were suspended in an elution solution that consisted of 1:1 (v/v) trifluoroethanol (TFE)/0.1% trifluoroacetic acid (TFA) in water. The particles were gently agitated for 1 h at room temperature before pelleting (as described above). The supernatant layer, containing eluted captured peptides, was set aside for later use. To ensure that all peptides had been removed from the particle interior, the elution process was repeated three more times with incubation times of 20 min. All elution supernatants were combined, dried via vacuum centrifugation, then desalted via C-18 Zip-Tip (Millipore, Billerica, MA) solid-phase extraction for mass spectrometry analysis.

LC−MS/MS

LC−MS/MS experiments were performed on an LTQ-Orbitrap Elite (ThermoFisher Scientific, Waltham, MA) equipped with a nanospray EASY-nLC 1000 HPLC system (Thermo Fisher Scientific). Peptides were separated using a reversed-phase PepMap 50 μm i.d. × 15 cm long with 3 μm, 100 Å pore size, C18 resin LC column (ThermoFisher Scientific, Waltham, MA). The mobile phase consisted of 0.1% aqueous formic acid (mobile phase component A) and 0.1% formic acid in acetonitrile (mobile phase component B). After sample injection, the column was washed for 5 min with A; the peptides were eluted by using a linear gradient from 0 to 50% B over 120 min and ramping to 100% B for an additional 2 min. The flow rate was set at 300 nL/min. The LTQ-Orbitrap Elite was operated in a data-dependent mode in which one full MS scan (60 000 resolving power) from 300 to 2000 Da was followed by a second full MS scan (60 000 resolving power) from 300 to 900 Da, and was followed by eight MS/MS scans (60 000 resolving power) in which the eight most abundant molecular ions were dynamically selected (based on the second full scan from 300 to 900 Da) and fragmented by ETD. Fluoranthene was used as the electron-transfer reagent with a charge-state-dependent ETD reaction time based on a reaction time of 100 ms for doubly charged ions. "FT master scan preview mode", "Charge state screening", "Monoisotopic precursor selection", and "Charge state rejection" were enabled so that only the ≥4+ ions are selected and fragmented by ETD. Supplemental activation was also applied. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD005043.

Spectral Analysis

Mass spectra were directly imported as .RAW files and analyzed by SIEVE software (Thermo Scientific) to determine how many ions were present, along with their corresponding charge states. Analysis was done using nondifferential single class analysis of the raw data using the following parameters: full m/z range (300−900), full retention time range, full frame time width (2.5 min) and m/z width (10 ppm), 20 000 maximum frames, peak intensity threshold of 200 000, and maximum charge state of +8. Results were filtered so that only monoisotopic isotopes were considered.

Sequencing Analysis

Mass spectra were imported directly as .RAW files and analyzed by PEAKS de novo sequencing software version 6 (Bioinformatics Solutions, Waterloo, ON, Canada).[21,22] PEAKS first performs a de novo sequence analysis of the ETD MS/MS data. Mass tolerance for precursor ions was 10 ppm, and mass tolerance for fragment ions was 0.05 Da. Data were analyzed with no enzyme specificity, along with oxidation (+15.9949 Da) on methionine as a variable post-translation modification and a quality spectral filter of 0.65. Confident de novo peptide sequences were achieved by filtering average local confidence (ALC) to ≥30%. PEAKS then used the high-confidence de novo generated sequence tags to search against a database consisting of our transcriptome assembly. A 1% false discovery rate (FDR) was used as a cutoff value for reporting peptide spectrum matches (PSMs) from the database. Peptides of interest that were identified against the transcriptome database had their sequences manually verified by matching the experimental spectral ions to a fragment prediction program (MS-Product; http://prospector.ucsf.edu). For peptides of interest with de-novo-only sequences, the sequences were manually verified by matching the experimental spectral ions to a fragment prediction program (MS-Product; http://prospector.ucsf.edu). In addition, the peptides synthesized for the antimicrobial activity assay based on the de-novo-only sequences were analyzed by ETD-MS, and those spectra were matched against the experimentally derived sequences. For de-novo-only sequences, only leucine (L) was denoted because it is indistinguishable from isoleucine (I) by ETD fragmentation.

Cationic Antimicrobial Peptide Prediction

Potential CAMPs were selected based on results from CAMP prediction algorithms or physicochemical properties. We developed a Python script to aid in the identification of potential antimicrobial peptides. The Python script generates a FASTA file of the sequences determined by PEAKS de novo sequencing software. The script then sends post requests to the Collection of Anti-Microbial Peptides (CAMP) database that allows evaluation using three different machine learning algorithms (Support Vector Machine (SVM), Random Forest (RF), and Discriminant Analysis (DA)) to predict whether a specific peptide sequence will likely correspond to a peptide with antimicrobial properties.[23,24] The script then calculates the physicochemical properties (length, molecular weight, nominal solution charge, pI, and hydrophobicity) of each sequence. Finally, all of the data are automatically exported to an Excel file and sorted by the various physicochemical parameters, including hydrophobicity and net charge, or by the number of algorithms

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. Comparison of LC−MS/MS Analyses of Komodo Peptides for Unstimulated and LPS-Stimulated Plasma[a]

Komodo sample	total no. of ions with charges 4−8	total no. of MS/MS spectra	total no. of de novo sequences with charges 4−8	total no. of transcriptome db PSMs with charges 4−8	total no. of transcriptome db unique peptides with charges 4−8
unstimulated	2120	3121	193	79	71
LPS-stimulated	3156	3813	259	110	99

[a]MS and MS/MS data of unstimulated and LPS-stimulated plasma were compared to identify changes in total ions, total MS/MS spectra collected, total de novo sequence count, and total transcriptome database (db) identifications.

that predicted a sequence to be antimicrobial. The script is available upon request.

Antimicrobial Performance Assessment

Bacterial Stocks. *Staphylococcus aureus* (ATCC 25923) and *Pseudomonas aeruginosa* (ATCC 9027) were purchased from the American Type Culture Collection (Manassas, VA). Bacteria were grown in Nutrient Broth (Difco 234000) overnight in a shaking incubator (37 °C). Bacteria were aliquoted and frozen at −80 °C and enumerated via serial dilution and plating prior to experimentation.

Antibacterial Activity. The antimicrobial (EC_{50}) activity of the peptides against *Pseudomonas aeruginosa* and *Staphylococcus aureus* was determined as previously described.[25,26] The EC_{50} activity for each peptide was determined after a 3 h incubation time. The appropriate dilutions of each well were plated in triplicate, and the killing kinetics for the peptides were determined. In brief, 1×10^5 CFU (Colony Forming Unit) per well of bacteria were incubated with different peptide concentrations in 50 μL of low salt buffer (10 mM potassium phosphate buffer, pH 7.2, and 0.1% Nutrient broth) and incubated for 3 h at 37 °C.[26] Serial dilutions were then prepared in low salt buffer and plated in triplicate on Nutrient agar plates, which were then incubated for 24h at 37 °C. Bacterial survival at each peptide concentration was calculated based on the ratio of the number of colonies on each experimental plate and the average number of colonies that were observed in assay cultures lacking peptide. The peptide concentration required to kill 50% of the viable bacteria in the assay cultures (EC_{50}) was determined by plotting percent survival as a function of the log of peptide concentration (log μg/mL) and fitting the data to an equation describing a standard sigmoidal dose−response curve (eq 1) using GraphPad Prism 5 (GraphPad Software, San Diego, CA). In eq 1, Y corresponds to bacterial killing (%) at a given peptide concentration (μg/mL), with X being the logarithm of that concentration (log μg/mL), and the terms "Top" and "Bottom" refer to the upper and lower boundaries, which were constrained to values <100% and >0%, respectively. For the purpose of graphing, samples that had no peptide are plotted at 10^{-9} μg/mL. EC_{50} errors were reported based on the 95% confidence interval calculated on the log EC_{50} values to represent $p < 0.05$.

$$Y = \text{bottom} + (\text{top} - \text{bottom})/(1 + 10^{[(\log EC_{50} - X) * \text{hill slope}]}) \quad (1)$$

RESULTS AND DISCUSSION

Harvesting from Komodo Dragon Plasma

Cationic antimicrobial peptides are essential elements of innate immunity and are generated by a variety of tissues, with leukocytes producing an ensemble of CAMPs both constitutively and in response to pathogen exposure. Our previous investigations into the host-defense peptidome in plasma from the American alligator revealed the presence of a complex mixture of potential antimicrobial peptides.[15,16] In the present study, we have employed custom-made hydrogel particles to capture and enrich peptides with CAMP physicochemical characteristics for subsequent analysis and identification using LC−MS/MS. Here we have harvested and analyzed peptides in plasma from both Komodo dragon blood that had been stimulated using bacterial LPS and blood that had not, providing a broader analysis of Komodo dragon CAMPs and CAMP-like peptides. In addition to peptide harvesting, mRNA was extracted from Komodo dragon leukocytes, which was used to generate a transcriptome library, which provided a database to facilitate the sequencing of harvested peptides.

Identification of Novel Potential Komodo Dragon CAMPs

Our previous work on the discovery of CAMPs from the American alligator demonstrated that most of the peptides of interest in those harvests exhibited gas-phase charge states ranging from +4 to +6 and a molecular-weight range of 1200−4000 Da.[15,16] Therefore, for this work, data-dependent LC−MS/MS analysis was conducted on ions with gas-phase charge states of +4 or higher based on ions selected from the full MS m/z range of 300 to 900 Da. Table 1 shows the total number of ions with gas-phase charges of +4 to +8 (calculated using SIEVE), along with the total number of MS/MS spectra acquired, the total number of de novo sequences above 30% ALC, and the total number of PSMs and unique peptides from the Komodo transcriptome database identified using PEAKS for both the unstimulated and LPS-stimulated plasma samples. The results reveal a larger number of peptides as well as de novo sequences and peptide identifications for the LPS-stimulated plasma sample compared with the harvests from unstimulated plasma sample. The total number of ions increases from 2120 for the unstimulated plasma to 3156 for the LPS-stimulated plasma. This is an increase of 49%. The total number of MS/MS spectra goes from 3121 for the unstimulated plasma to 3813 for the LPS-stimulated plasma, an increase of 22%. The number of de novo sequences increased from 193 in unstimulated plasma to 259 in LPS-stimulated plasma (+34%). The unstimulated plasma sample yielded 79 transcriptome PSMs and 71 unique peptide identifications, while the LPS-stimulated sample yielded 110 transcriptome PSMs and 99 unique peptide identifications. In both cases, this represents a 39% increase. See Supplemental Tables S1−S4 for all de novo sequences and peptides identified against the transcriptome using PEAKS for the unstimulated and LPS-stimulated samples. There were 21 transcriptome-identified peptides in common between the LPS-stimulated and unstimulated plasma samples, while the LPS-stimulated sample showed 78 unique peptides and the unstimulated sample showed 50 unique peptides (Figure 1). Of the 21 peptides shown to be in common, all but one (NDSYLAERKQDLQK, a fragment of band 3 ion transport protein) are derived from histone proteins. All of the peptides identified against the

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

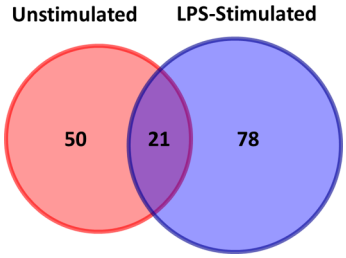


Figure 1. Venn diagram showing the overlap of peptides between the LPS-stimulated and unstimulated Komodo dragon plasma samples identified from the transcriptome database using PEAKS db software.

transcriptome, along with the proteins from which they are derived, are provided in Supplemental Tables S3 and S4. Note that the observed increases in the number of peptides, de novo sequences, PSMs, and peptide identifications for plasma from LPS-stimulated blood were expected. Pathogen-associated molecules such as LPS have been shown to stimulate host–cell responses including the release of peptides and proteins associated with chemotaxis, inflammation, immunomodulation, and host defense.[27,28]

Selecting Peptides for Antimicrobial Assessment

Once PEAKS had determined de novo sequences and peptide identifications against the transcriptome from the unstimulated and LPS-stimulated Komodo plasma samples, the next step in the process was to determine which peptide sequences would most likely demonstrate antibacterial properties. See Supplemental Tables S1–S4 for the complete list of de novo sequences and database peptide identifications. Both rational peptide sequence analysis based on known CAMP physicochemical properties and web-based CAMP prediction algorithms were used to select probable CAMP sequences for synthesis and testing. We developed a Python script to aid in the identification of potential antimicrobial peptides. The script employs three different web-based machine learning CAMP prediction algorithms (SVM, RF, and DA) provided at the Collection of Anti-Microbial Peptides (CAMP) database site to analyze the submitted sequences.[29] Each prediction algorithm utilizes models based on numerous known CAMP properties as the basis to predict antimicrobial activity for each sequence. See Supplemental Tables S5–S8 for the CAMP prediction scores for the de novo sequences and the database identified peptides. The script also calculates the physicochemical properties (length, molecular weight, nominal solution charge, pI, and hydrophobicity based on the GRAVY scale) for each sequence. The Python script dramatically reduced the time to process peptide sequence data from weeks to minutes to analyze each data set.

Physicochemical properties, such as theoretical charge at physiological pH, molecular weight, length, hydrophobicity, and peptide isoelectric point, were the focus for the rational analysis approach. Only sequences with molecular weights of <5 kDa were considered because this is consistent with the size of the majority of known vertebrate CAMPs. Also, because high cationic character is believed to be associated with activity, sequences that have predicted net solution charges of +4 or higher at pH 7 were chosen. Isoelectric point and hydrophobicity were not utilized as a primary consideration in determining antimicrobial potential because these properties are known to vary among CAMPs.

Using these two approaches for CAMP selection afforded two lists of potential CAMPs, one consisting of peptides possessing physicochemical properties associated with known CAMPs and another of peptide sequences assigned positive CAMP prediction scores in analyses using web-based CAMP prediction algorithms. The two lists of potential CAMPs were consolidated, and duplicate sequences were eliminated. This combination of computational and rational analysis yielded 48 potential CAMPs (Table 2). Because the transcriptome database used in these analyses was unannotated, information regarding potential peptide identifications was not provided in the PEAKS analyses. Therefore, to determine the identities and source proteins for the sequenced peptides, BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) searches were performed against the nonredundant protein database at NCBI. The closest probable parental source proteins were determined based on homology. All but one of the identified potential CAMPs were histone-derived peptides. In addition, several of the peptides can be sorted based on their amino acid sequences into three groups of closely related nested peptides (Table 3). These observed patterns suggest that histones may be cleaved by proteases in vivo at various locations on the amino acid chain, resulting in the generation of the observed peptides.

Eight peptides were selected from this list of 48 potential CAMPs for synthesis and evaluation [VK6 (AVKPKTAKPKTAKPKTA), VK7 (AKKPVAKKAAGGVKKPK), VK10 (ARTKQTARKSTGGKAPRKQLAT), VK11 (AKAVKPKTAKPKTAKPKTAKA), VK12 (AKKPVAKKAAGGVKKPKK), VK13 (AAGGVKKPKKAAAAKKSPKKPKKPAAA), VK14 (KAAAAKKSPKKPKKPAAAKK), and VK25 (SPKKTKPVKPKKVA)]. One of the peptides (VK25) was a de-novo-only sequence, while the other seven peptides were identified in the transcriptome database. All seven of the transcriptome-identified peptides were predicted to be CAMPs by all three web-based algorithms and demonstrated relatively high net positive charges (+6 to +10) at physiological pH. The de-novo-only peptide (VK25) was selected because of its relatively small molecular weight (1536 Da) coupled to a high net positive charge (+6) at pH 7. Additionally, VK25 was also predicted to be antimicrobial by all three of the algorithms. VK6 and VK11 are a nested set of peptides, where VK6 shares the same internal sequence as VK11, with VK11 containing two additional amino acids both on the C- and N-termini. VK7, VK12, VK13, and VK14 reside within a continuous 41-residue sequence, with VK7, VK12, and VK13, which come from the N-terminal portion of the sequence, being a nested set of peptides that share the core sequence AAGGVKKPK. The sequences of VK7 and VK12 are nearly identical, with the only difference being that VK12 contains an additional C-terminal Lys that is absent in VK7. The sequence of VK13 partially overlaps with VK7 and VK12 but extends another 17 amino acids its C-terminus. While VK14 lacks the nine-residue core sequence, its sequence shares 18 amino acid residues with that of VK13 and contains an additional two C-terminal amino acids beyond the sequence of VK13. BLAST searches for these eight peptides against the nonredundant protein database at NCBI revealed that all of the peptides, except VK10, show homology with histone H1 from various species. A similar BLAST analysis indicated that the sequence of VK10 shows homology with histone H3. These results suggest that these peptides originate from histone proteins.

Peptide Sequence Determination

As can be seen in Table 4, which compares PEAKS de-novo-assigned sequences for the eight peptides selected for

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX–XXX

Table 2. Physicochemical Properties of Potential CAMPs Identified from Komodo Plasma Analyzed by ETD LC−MS/MS[a]

peptide name[b]	peptide sequence	MW (Da)	pI	length (res)	net charge[c]	hydropathy[d]	source protein
	GTGASGSFKLNKKQQEGKAAAAAKKKPAV	2870.61	10.40	29	6	−0.81	histone H1
VK13	**AAGGVKKPKKAAAAKKSPKKPKKPAAA**	**2626.65**	**10.95**	**27**	**10**	**−0.99**	**histone H1**
	PEPAKSAPVPKKGSKKAITKTQK	2418.43	10.40	23	6	−1.30	histone H2B
	SLVSKGTLVQTKGTGASGSFKLNK	2407.34	10.48	24	4	−0.21	histone H1
	GIGASGSFKIGKKPGEGKEKAPKK	2398.37	10.13	24	5	−1.09	histone H1
	PEPAKSAPAPKKGSKKAVTKTQK	2376.39	10.40	23	6	−1.42	histone H2B
VK10	**ARTKQTARKSTGGKAPRKQLAT**	**2354.36**	**12.31**	**22**	**7**	**−1.41**	**histone H3**
	GIGASGSFKIGKKPGEGKEKAPK	2270.27	10.00	23	4	−0.97	histone H1
	ARTKQTARKSTGGKAPRKQLA	2253.31	12.31	21	7	−1.45	histone H3
	IKRLVTTGVLKQTKGVGASGSF	2246.31	11.26	22	4	0.20	histone H1
VK11	**AKAVKPKTAKPKTAKPKTAKA**	**2162.36**	**10.85**	**21**	**8**	**−1.10**	**histone H1**
	AKKPVAKKAAGGVKKPKKAAAA	2117.35	10.85	22	8	−0.56	histone H1
	AVKPKTAKPKTAKPKTAKAK	2091.33	10.85	20	8	−1.25	histone H1
	PEPAKSAPIPKKGSKKAITK	2075.25	10.30	20	5	−1.08	histone H2B
	IGKKPGEGKEKAPKKKAAAK	2063.26	10.22	20	6	−1.55	histone H1
	PEPAKSAPVPKKGSKKAITK	2061.23	10.30	20	5	−1.10	histone H2B
	SAPATGGVKKPHRYRPGTVA	2049.12	11.10	20	4	−0.79	histone H2B
VK14	**KAAAAKKSPKKPKKPAAAKK**	**2046.32**	**10.90**	**20**	**9**	**−1.41**	**histone H1**
	AAAKAKKPVAKKAAGGVKKPK	2046.32	10.85	21	8	−0.68	histone H1
	PEPAKSAPAPKKGSKKAVTK	2019.18	10.30	20	5	−1.23	histone H2B
	PATGGVKKPHRYRPGTVAL	2004.14	11.10	19	4	−0.68	histone H3
	AVKPKTAKPKTAKPKTAKA	1963.23	10.78	19	7	−1.11	histone H1
	PEPAKSAPIPKKGSKKAIT	1947.15	10.18	19	4	−0.93	histone H2B
	AAAAKKSPKKPKKPAAAKK	1918.22	10.85	19	8	−1.27	histone H1
	AVKPKTAKPKTAKPKTAK	1892.19	10.78	18	7	−1.27	histone H1
	PEPAKSAPAPKKGSKKAVT	1891.09	10.18	19	4	−1.09	histone H2B
	KKPTPAPKKAKKPKTVK	1874.22	10.85	17	8	−1.84	histone H1
	TAKPKTAKAKKAAPKKK	1794.16	10.85	17	8	−1.58	histone H1
VK7	**AKKPVAKKAAGGVKKPK**	**1705.11**	**10.78**	**17**	**7**	**−0.92**	**histone H1**
	AVKPKTAKPKTAKPKT	1693.06	10.70	16	6	−1.29	histone H1
	PEPAKSAPVPKKGSKK	1647.97	10.18	16	4	−1.48	histone H2B
	FAGRNFRNPRVK	1460.81	12.30	12	4	−1.23	alpha enolase
	KTKPVKPKKVAK	1350.91	10.70	12	6	−1.43	histone H1
	LNKKQQEGKAAAAAKKKPAVK	2206.33	10.40	21	6	−1.17	histone H1
	PEPAKSAPVPKKGSKKAIT	1933.14	10.18	19	4	−0.95	histone H2B
	AKAVKPKTAKPKTAKPKT	1892.19	10.78	18	7	−1.27	histone H1
VK12	**AKKPVAKKAAGGVKKPKK**	**1833.20**	**10.85**	**18**	**8**	**−1.09**	**histone H1**
	AKAVKPKTAKPKTAKPK	1791.15	10.78	17	7	−1.30	histone H1
VK6	**AVKPKTAKPKTAKPKTA**	**1764.10**	**10.70**	**17**	**6**	**−1.11**	**histone H1**
	IGKKPGEGKEKAPKKK	1722.05	10.13	16	5	−2.03	histone H1
	TKPVKPKKVAKSPAKA	1677.07	10.70	16	6	−0.99	histone H1
	AAAAKKSPKKPKKPAAA	1662.03	10.70	17	6	−0.96	histone H1
	SPAKAKAVKPKTAKPK	1649.04	10.70	16	6	−1.14	histone H1
	AVKPKTAKPKTAKPK	1592.01	10.70	15	6	−1.33	histone H1
	TAKPKTAKAKKAAPK	1537.97	10.70	15	6	−1.27	histone H1
VK25	*SPKKTKPVKPKKVA*	*1534.99*	*10.70*	*14*	*6*	*−1.39*	*histone H1*
	AKPVKASKPKKAK	1379.90	10.70	13	6	−1.37	histone H1
	KTKPVKPKKVA	1222.81	10.60	11	5	−1.20	histone H1

[a]Source protein, sequences, length, MW, charge, pI, and hydrophobicity of identified peptide sequences are listed. The peptides in bold were chosen for synthesis and antimicrobial analysis. The peptide in italics was a de-novo-only sequence chosen for synthesis and antimicrobial analysis. [b]Names of peptides selected for synthesis and analysis. [c]Net solution charge at pH 7. [d]Calculated using the GRAVY hydrophobicity scale.

evaluation to their correct sequences (determined by alignment with the Komodo dragon transcriptome or manual verification), PEAKS de novo sequencing worked better with some spectra than with others. This was expected because the MS/MS spectral quality and the number of c and z ions present in each spectrum varied (Figure 2).

The PEAKS de novo sequence for VK7 had the highest possible correlation to the actual sequence with 100% correct identity, the only discrepancy being that the order of the third and fourth amino acids was reversed. The ETD spectrum is of high quality, showing a nearly complete c and z ion series (Figure 2B).

Peptide VK10 demonstrated excellent correlation between the PEAKS de novo sequence and the actual sequence with 77% correct identity. The de novo sequence was able to correctly identify the first 17 of the 22 total amino acids, where

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 3. Peptides from the List of 48 Potential CAMPS Showing Which Are Nested Sets of One Another[a]

Group 1	PEPAKSAP**V**PKKGSKK
	PEPAKSAP**V**PKKGSKKAIT
	PEPAKSAP**V**PKKGSKKAITK
	PEPAKSAP**V**PKKGSKKAITKTQK
	PEPAKSAP**A**PKKGSKKAVT
	PEPAKSAP**A**PKKGSKKAVTK
	PEPAKSAP**A**PKKGSKKAVTKTQK
	PEPAKSAP**I**PKKGSKKAIT
	PEPAKSAP**I**PKKGSKKAITK
Group 2	AVKPKTAKPKTAKPK
	AVKPKTAKPKTAKPKT
	AVKPKTAKPKTAKPKTA[1]
	AVKPKTAKPKTAKPKTAK
	AVKPKTAKPKTAKPKTAKA
	AVKPKTAKPKTAKPKTAKAK
	AKAVKPKTAKPKTAKPKTAKA[2]
Group 3	AAAKAKKPVAKKAAGGVKKPK
	AKKPVAKKAAGGVKKPK[3]
	AKKPVAKKAAGGVKKPKK[4]
	AKKPVAKKAAGGVKKPKKAAAA
	AAGGVKKPKKAAAAKKSPKKPKKPAAA[5]
	KAAAAKKSPKKPKKPAAAKK[6]
	AAAAKKSPKKPKKPAAAKK

[a]Peptides in Group 1 show a core sequence of PEPAKSAP**X**PKKGSKK, where **X** is Val, Ala, or Ile (highlighted in yellow). Group 2 consists of a nested set of peptides with a core sequence of AVKPKTAKPKTAKPK (highlighted in light blue). Group 3 shows peptides with two predominant segments—one with a core sequence of AKKPVAKKAAGGVKKPK (highlighted in gray) and the other with a core sequence of AAAAKKSPKKPKKPAAA (highlighted in purple), bridged together with a Lys residue (highlighted in green). Six of the eight peptides chosen for synthesis and antimicrobial analysis are included in this list ([1]VK6, [2]VK11, [3]VK7, [4]VK12, [5]VK13, [6]VK14).

Table 4. Comparison of PEAKS De Novo Sequences versus PEAKS Database Sequences of the Eight Komodo Peptides Chosen for Synthesis and Antimicrobial Analysis[a]

peptide	actual sequence	PEAKS de novo sequence	% correct identity	de novo ALC (%)	transcriptome db score (−logP)
VK6	AVKPKTAKPKTAKPKTA	GLKPKTAPKKTALLKGN	59	49	85.53
VK7	AKKPVAKKAAGGVKKPK	AKPKVAKKAAGGVKKPK	100	73	102.22
VK10	ARTKQTARKSTGGKAPRKQLAT	ARTKQTARKSTGGKPARAKTLQ	77	61	97.25
VK11	AKAVKPKTAKPKTAKPKTAKA	AKVAPKADTKLLALNLKVAAK	43	16	48.97
VK12	AKKPVAKKAAGGVKKPKK	AKPKVAKKAAGKRALALL	61	50	63.10
VK13	AAGGVKKPKKAAAAKKSPKKPKKPAAA	PTGARLLKKAAAAQPVALKLLKRGGV	23	23	89.62
VK14	KAAAAKKSPKKPKKPAAAKK	KAAAAKKALKPKKGVKPGKK	45	60	63.81
VK25	SPKKTKPVKPKKVA	PSKKTKVPKPKKVA	100	45	N/A

[a]De novo sequences determined by PEAKS are compared with the manually-verified CAMP sequences and the percent correct identity was recorded. The PEAKS de novo and database (db) scores for each sequence are also reported. VK25 was a de-novo-only sequence and did not have a transcriptome database identification. ALC: average local confidence.

only the 15th and 16th amino acids were assigned in the opposite order. PEAKS was able to determine the correct mass tag for the missing C-terminal residues and assigned the correct five amino acids but in the incorrect order. The ETD spectrum of VK10 (Figure 2C) shows that many of the c and z ions are present but lacks sufficient low mass z ions and high mass c ions that are necessary to accurately sequence the C-terminal end of the peptide in a de novo manner.

Peptides VK6 and VK12 showed moderate correlation between their PEAKS de novo sequences and their actual sequences, with 59 and 61% correct identity, respectively. VK6 de novo sequence was able to correctly identify ten consecutive amino acids, beginning with the Lys residue in position three and ending with the Ala residue in position 12. However, the first two N-terminal and the last five C-terminal amino acids were assigned incorrectly. The MS/MS spectrum of VK6 (Figure 2A) exhibits too few low-mass z ions and high-mass c ions necessary to accurately sequence the C-terminal end of the peptide in a de novo manner. In the case of VK12, the de novo sequence correctly identified the first 11 consecutive amino acid residues, with the exception of the third and fourth residues, which were in the incorrect order. As with VK10 and VK6, the ETD spectrum of VK12 (Figure 2E) affords insufficient low-mass z ions and high-mass c ions necessary to accurately sequence the C-terminal end of the peptide in a de novo manner, leading to an incorrect assignment of the seven C-terminal amino acids.

The next three peptides, VK11, VK13, and VK14, all showed the lowest correlations between their PEAKS de novo sequences and their true sequences (43, 23, and 45% correct identity, respectively). The ETD spectrum for VK11 (Figure 2D) presented few fragment ions, which made it more difficult to assign a sequence in a de novo manner. However, on the basis of the few fragment ions that were present in the spectrum,

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

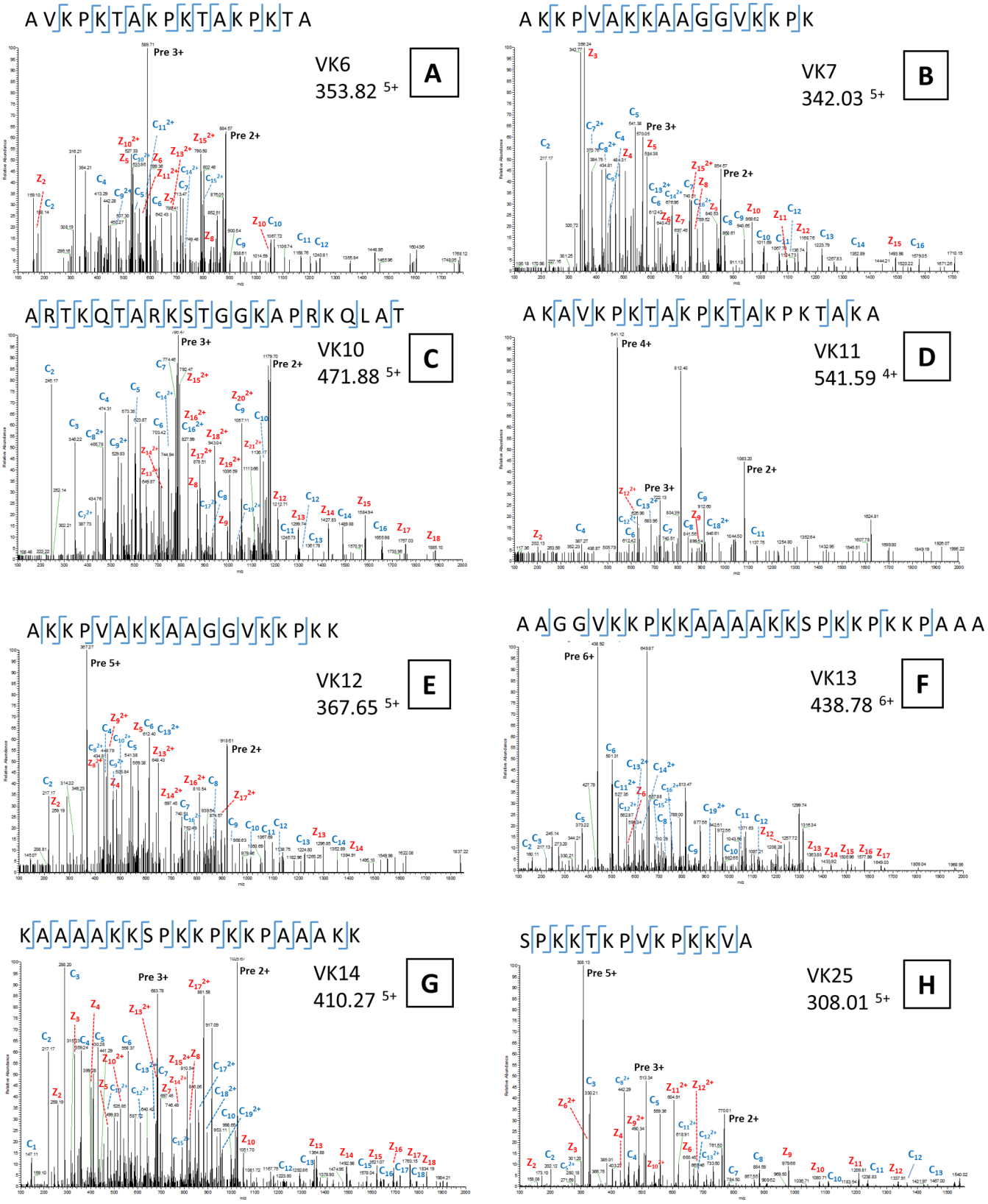


Figure 2. ETD MS/MS spectra of the 8 peptides chosen for synthesis and antimicrobial analysis. (A) VK6−353.82^{5+}, (B) VK7−342.03^{5+}, (C) VK10−471.88^{5+}, (D) VK11−541.59^{4+}, (E) VK12−367.65^{5+}, (F) VK13−438.78^{6+}, (G) VK14−410.27^{5+}, and (H) VK25−308.01^{5+}. Observed c ions (blue) and z ions (red) are indicated on the spectra. Underlined amino acid residues in the sequence indicate ions that are present in the spectra.

together with other de novo sequence tags (data not shown), PEAKS was able to correctly identify the peptide with a −logP score of 45.22 when the spectrum was searched against the transcriptome database. While the MS/MS spectrum for VK13 (Figure 2F) appeared to contain an adequate supply of fragment ions for PEAKS to determine a reliable de novo sequence,

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

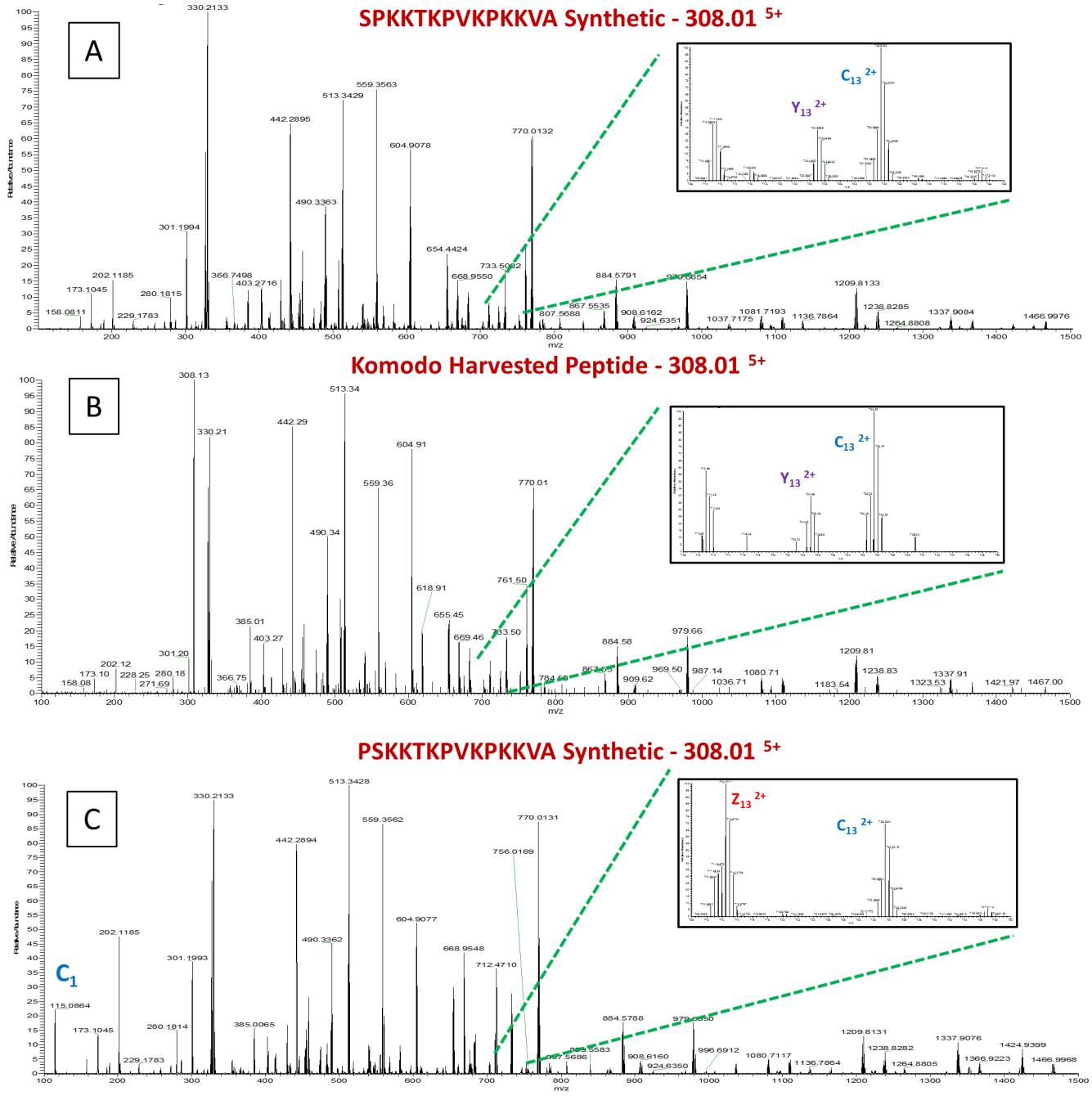


Figure 3. Comparison of the ETD MS/MS spectra for precursor 308.01^{5+} from (A) VK25 synthetic peptide SPKKTKPVKPKKVA, (B) the natural Komodo harvested peptide, and (C) VK25 synthetic peptide PSKKTKPVKPKKVA. For panels A and B, the insets show the presence of the $Y_{13}{}^{2+}$ ion at m/z 724.98 and the absence of the $Z_{13}{}^{2+}$ ion at m/z 711.96. For panel C, the inset shows the presence of the $Z_{13}{}^{2+}$ ion at m/z 711.96 and the absence of the $Y_{13}{}^{2+}$ ion at m/z 724.98. Also, the C1 ion at m/z 115.08 is present in panel C but absent in panels A and B. These spectra demonstrate that the natural VK25 peptide harvested from Komodo blood has the sequence SPKKTKPVKPKKVA.

only six consecutive residues in the middle of the peptide were correctly assigned. The reason for this is not clearly understood. Despite the failure to deduce a reasonable de novo sequence, the peptide was accurately identified against the transcriptome database with a −logP score of 89.62. As was the case with VK13, the MS/MS spectrum for VK14 also seemed to contain an adequate supply of c and z ions (Figure 2G) to allow PEAKS to determine a reliable de novo sequence. However, only the first seven consecutive residues and the last two residues in the peptide sequence were correctly assigned. Again, the reason for the PEAKS software's failure to accurately assign the complete sequence is not clear. However, when PEAKS searched against the transcriptome, a correct sequence was determined with a −logP score of 63.81.

The sequence of the peptide VK25 was not identified from the Komodo transcriptome database, and its de novo sequence had to therefore be manually verified. The de novo sequence determined by PEAKS for VK25 was PSKKTKVPKPKKVA, with an ALC of 55%. By comparing the actual spectrum (Figure 2H) to the experimental spectral ions (MS-Product; http://prospector.ucsf.edu), it was determined that seventh and eighth residues (Val-Pro) were in the incorrect order. This was

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 5. Antibacterial Performance Data of Komodo Dragon CAMPs[a]

peptide	*S. aureus* 25923			*P. aeruginosa* 9027		
	EC_{50} (μg/mL)	95% CI	EC_{50} (μM)	EC_{50} (μg/mL)	95% CI	EC_{50} (μM)
VK6	6.24	4.35 to 8.95	3.57	5.80	4.04 to 8.33	3.31
VK7	65.15	43.3 to 87.9	38.4	2.17	0.914 to 5.17	1.28
VK10	3.14	1.41 to 7.00	1.33	5.99	4.65 to 7.70	2.54
VK11	5.99	4.65 to 7.70	2.79	10.0	7.07 to 14.27	4.68
VK12	1.29	1.07 to 1.55	0.70	1.46	1.10 to 1.94	0.80
VK13	2.34	2.02 to 2.72	0.89	2.65	1.65 to 4.25	1.01
VK14	1.25	1.02 to 1.53	0.61	2.52	2.17 to 2.93	1.24
VK25	4.92	4.28 to 5.64	3.24	3.81	2.50 to 5.80	2.51
LL-37	3.96	2.75 to 3.70	0.88	1.78	1.361 to 2.33	0.39

[a]Eight selected peptides were tested for antibacterial properties against Gram-negative and Gram-positive bacteria, *Pseudomonas aeruginosa* and *Staphylococcus aureus*, respectively. The human CAMP LL-37, whose antimicrobial properties have been well characterized, is included as a reference.

verified by the presence of the C7 and C72+ ions at m/z 784.50 and m/z 392.75, respectively, as well as the absence of the Z8 and Z82+ ions at m/z 850.56 and m/z 425.78, respectively. The C7 and C72+ ions were expected to be present due to fragmentation on the C-terminal side of proline residues during ETD, while the Z8 and Z82+ ions were not expected to be present due to lack of fragmentation on the N-terminal side of proline residues.[30] The remaining parts of the sequence were determined to be correct, and the peptide PSKKTKPVKPKKVA was chosen to be synthesized for subsequent analysis.

Subsequent to initial sequence determination, it was noticed that the first and second amino acids may have been assigned in the incorrect order. If the correct order was Pro-Ser (as originally thought), the Z_{13}^{2+} ion at m/z 711.97 would be expected to be present, while the peak at m/z 724.99 (Y_{13}^{2+}) would be expected to be absent. However, if the actual order of the first and second residues was Ser-Pro, then the Z_{13}^{2+} ion at m/z 711.97 would be expected to be absent, while the Y_{13}^{2+} ion at m/z 724.99 would be expected to be present. Because the MS/MS spectrum shows the presence of the Y_{13}^{2+} ion at m/z 724.99 and the absence of the Z_{13}^{2+} ion at m/z 711.97, this indicated that the true sequence was SPKKTKPVKPKKVA. As an additional means to verify the correct sequence, a BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was conducted against nonredundant protein sequences from all species using both PSKKTKPVKPKKVA and SPKKTKPVKPKKVA. The search results for SPKKTKPVKPKKVA (Table S9) returned all 100 alignments from histone H1, with as many as 13/14 (93%) identities, the closest match being SPKKAKPVKPKKVA from *Thamnophis sirtalis* (Common Garter Snake), *Gekko japonicas* (Japanese Gecko), and *Ophiophagus Hannah* (King Cobra). All 100 alignments returned the first and second residues as Ser-Pro. The results for PSKKTKPVKPKKVA (Table S10) showed similar results with a mixture of alignments from histone H1 and hypothetical proteins. The closest three matches were also from *Thamnophis sirtalis*, *Gekko japonicas*, and *Ophiophagus Hannah*, with a sequence of KKAKPVKPKKVA. These BLAST results, along with the evidence from the MS/MS spectrum, indicated that the correct sequence of peptide VK25 was actually SPKKTKPVKPKKVA.

Once this new peptide was synthesized, ETD MS/MS was conducted on both versions of VK25 to verify that the correct sequence was indeed SPKKTKPVKPKKVA. Figure 3 compares the ETD spectra of both synthesized versions of VK25 and the natural peptide harvested from the Komodo plasma using the microparticles. As can be seen in these ETD spectra, the synthetic peptide SPKKTKPVKPKKVA shows the presence of the Y_{13}^{2+} ion at m/z 724.99 and the absence of the Z_{13}^{2+} ion at m/z 711.97, while the synthetic peptide PSKKTKPVKPKKVA shows the presence of the Z_{13}^{2+} ion at m/z 711.97 and the absence of Y_{13}^{2+} ion at m/z 724.99. Furthermore, the spectrum for the synthetic peptide PSKKTKPVKPKKVA contains the C1 ion at m/z 115.08, while this ion is absent in both the synthetic peptide SPKKTKPVKPKKVA and the naturally harvested peptide. These results clearly verified that the sequence of the natural VK25 peptide harvested from the Komodo plasma was SPKKTKPVKPKKVA.

Antimicrobial Performance

The antimicrobial activities of the eight selected peptides and a reference human CAMP LL-37, whose antimicrobial properties have been well characterized, were determined in low salt buffer (Table 5). None of the peptides were found to exhibit activity in minimum inhibitory concentration (MIC) experiments performed using cation-adjusted Mueller Hinton broth and following Clinical and Laboratory Standards Institute (CLSI) protocols, where peptides were added to bacteria in broth (data not shown).[31] However, it has been demonstrated previously that traditional MIC experiments may not accurately reflect potential antimicrobial potency, and the activity of physiologically important CAMPs can be evaluated in low salt buffer with a short incubation and subsequent plating or addition of media.[26,32−34] Under these conditions, the eight synthesized peptides displayed a range of activities against the model Gram-negative and Gram-positive bacteria, *Pseudomonas aeruginosa* and *Staphylococcus aureus*, respectively. The various peptides represented nested pairs (VK6 and VK11, VK7 and VK12, and VK13 and VK14 (Table 3), with the sequence of one peptide being completely or partially represented in the amino acid sequence of another.

At 21 residues, VK11 is four amino acids longer than the shorter nested peptide VK6. Moreover, the charge of VK11 (+8) at physiological pH is estimated to be two greater than that of the shorter VK6 (+6). Despite the differences in length and charge, VK6 and VK11 exhibited similar potencies against both *P. aeruginosa* (2.26 and 3.15 μM, respectively) and *S. aureus* (2.44 and 1.88 μM, respectively).

The 27-residue peptide VK13 is seven amino acids longer than the nested peptide VK14 and has a predicted charge of +10 at physiological pH, one greater than that predicted for VK14 (+9). As was the case with VK6 and VK11, the nested VK13 and VK14 peptides also exhibited similar potencies against *P. aeruginosa* (0.683 and 0.791 μM, respectively) and *S. aureus* (0.603 and 0.392 μM, respectively).

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

By contrast, VK7 and VK12 demonstrated similar potencies against *P. aeruginosa* (0.829 and 0.511 μM, respectively) but differed significantly in their effectiveness against *S. aureus* (23.6 and 0.451 μM, respectively). Notably, the peptide VK7, with a length of 17 residues, is only one amino acid shorter than VK12, whose sequence includes an additional C-terminal Lys residue that is absent in VK7.

The two remaining Komodo dragon peptides VK10 and VK25, whose sequences are not related to each other nor to the other six VK peptides tested, were similarly effective against *P. aeruginosa* (1.83 and 1.63 μM, respectively) and *S. aureus* (0.961 and 2.11 μM, respectively). The predicted charges of these peptides at physiological pH (VK10 = +7 and VK25 = +6) are similar to those of the other six Komodo dragon peptides. At 14 residues, VK 25 is the shortest of the peptides tested. All of the peptides demonstrated antimicrobial effectiveness approaching or exceeding that of the known well-characterized reference human peptide LL-37 (*P. aeruginosa*: 0.304 μM and *S. aureus*: 0.676 μM), with the performance of VK7 against *S. aureus* being a notable exception.

Histone-Derived Peptides

The identified Komodo dragon peptides are not represented in the known repertoire of antimicrobial peptides in reptiles.[8] All but one of the 48 identified potential CAMP peptides appear to be derived from Komodo dragon histone proteins; specifically, they show homology to portions of histones H2B, H1, and H3 (Table 2). Besides their role in gene regulation, histones have been known to have antibacterial activity.[35] Shrimp, fish, frogs, and murine cell lines have all been found to produce antimicrobial histones.[36−39] In humans, variations of histone H1 isolated from the gut have been shown to have activity against bacteria, particularly Gram-negative bacteria,[40,41] while histones H2A and H2B that exhibit activity against *E. coli* have been found in placenta.[42]

Peptide fragments of histones have also been found to exhibit antimicrobial properties. The most well known are buforins I and II, which are fragments from the N-terminus of histone H2A and were originally isolated from the stomach lining of the toad *Bufo gargarizans*.[43] Homologous peptides have been discovered in catfish and Atlantic halibut.[44,45] Despite being histone-derived, these Komodo dragon VK peptides all differ significantly from the buforin peptides.

It is notable that all of the VK peptides reported here are all derived from histone H1. While H1 histones have been generated and studied in the laboratory, few H1-derived peptides have previously been observed in nature. The bioprospector approach to CAMP discovery focuses on identifying intact native peptides, and thus the VK histone peptides that we have identified from Komodo plasma represent naturally occurring H1-derived CAMPs. Other H1-derived antimicrobial peptides that have been reported in the literature include oncorhyncin II, which was identified from skin secretions of rainbow trout, *Oncorhynchus mykiss*. Oncorhyncin II was reported to be a broad-spectrum antimicrobial peptide that exhibited antimicrobial effectiveness against a wide range of bacteria, including *Micrococcus luteus*, *Planococcus citreus*, *Escherichia coli*, and *Listonella anguillarum*.[46] Fragments of histone H1 have also been found in portions of the human gut, and these peptides show in vitro antimicrobial activity against *Salmonella typhimurium*.[41] Various fragments of histone H1 have also been found in the wound fluid of human patients,[47] indicating that histones and histone fragments may actually play a role in the human immune response against bacteria.

CONCLUSIONS

We have successfully applied our bioprospecting approach to CAMP discovery in the analysis of plasma from the Komodo dragon and discovered the presence of a complex milieu of novel potential CAMPs. As indicated in Table 2, all but one of the 48 potential CAMPs that we identified here were fragments of histone proteins, and their detection would have been a daunting challenge using more traditional proteomic and genomic methods. By capturing and analyzing the intact native peptides, our bioprospecting approach elucidated the sequences of the peptides as they existed in the plasma. The complex and diverse array of potential CAMPs reported here resembles the large number of potential CAMPs that we previously identified from American alligator plasma.[15,16] As with the alligator peptides, the identified Komodo dragon peptides are fragments of larger proteins.

As previously mentioned, there is evidence that histones play an important role in immunity and antimicrobial histone-derived peptides, such as buforin, are known. However, the peptides that we identified from Komodo dragon are distinct from known histone-derived peptides. The majority of these peptides are from histone H1; however, histones H2B and H3 are also represented among the identified Komodo dragon peptides. Seven of the eight peptides selected for evaluation (VK6, 10, 11, 12, 13, 14, and 25) exhibited potent antimicrobial effectiveness against both Gram-negative *P. aeruginosa* and Gram-positive *S. aureus*, while the remaining peptide (VK7) was effective against the former microbe but not as potent against the latter.

The results of these studies suggest that histone-derived peptides may play a larger role in innate immunity, at least in the Komodo dragon, than previously believed. While most of the identified peptides were present in both unstimulated and LPS-stimulated plasma, a greater number of histone-derived potential antimicrobial peptides were identified from stimulated plasma than from plasma that had not been stimulated. While our bioprospecting approach establishes sequences of the intact native peptides that are present in the sample, it does not provide information regarding the mechanisms by which they are produced or their regulation. Future efforts will focus on determining whether peptides are constitutively produced or the result of pathogen detection as well as whether this phenomenon is limited to Komodo dragons or if it occurs in other species, including humans.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00857.

Table S1: De novo sequences determined by PEAKS software for the unstimulated Komodo plasma. Table S2: De novo sequences determined by PEAKS software for the LPS-stimulated Komodo plasma. Table S3: Transcriptome database identifications determined by PEAKS software for the Komodo unstimulated plasma. Table S4: Transcriptome database identifications determined by PEAKS software for the Komodo LPS-stimulated plasma. Table S5: CAMP Prediction for the de novo sequences from unstimulated Komodo plasma. Table S6: CAMP

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

Prediction for the de novo sequences from LPS-stimulated Komodo plasma. Table S7: CAMP Prediction for the transcriptome database identified peptides from unstimulated Komodo plasma. Table S8: CAMP Prediction for the transcriptome database identified peptides from LPS-stimulated Komodo plasma. Table S9: BLAST search results for sequence SPKKTKPVKPKKVA against nonredundant protein sequences from all species. Table S10: BLAST search results for sequence PSKKTKPVKPKKVA against nonredundant protein sequences from all species. (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: bbishop1@gmu.edu.

ORCID

Barney M. Bishop: 0000-0002-6626-9251

Notes

The authors declare no competing financial interest.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD005043.

ACKNOWLEDGMENTS

We are grateful to the Defense Threat Reduction Agency (DTRA), HDTRA1-12-C-0039, for supporting this work. The funding agency had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are also very grateful to the St. Augustine Alligator Farm Zoological Park and its staff for their support and for providing the Komodo dragon blood samples used in these studies. We also thank Dr. Jennifer Van Eyk (Johns Hopkins University) for use of their Thermo Scientific Orbitrap Elite and helpful discussions.

REFERENCES

(1) Montgomery, J. M.; Gillespie, D.; Sastrawan, P.; Fredeking, T. M.; Stewart, G. L. Aerobic salivary bacteria in wild and captive Komodo dragons. *J. Wildl. Dis.* **2002**, *38* (3), 545−551.

(2) Bull, J. J.; Jessop, T. S.; Whiteley, M. Deathly Drool: Evolutionary and Ecological Basis of Septic Bacteria in Komodo Dragon Mouths. *PLoS One* **2010**, *5* (6), e11097−e11097.

(3) Goldstein, E. J. C.; Tyrrell, K. L.; Citron, D. M.; Cox, C. R.; Recchio, I. M.; Okimoto, B.; Bryja, J.; Fry, B. G. ANAEROBIC AND AEROBIC BACTERIOLOGY OF THE SALIVA AND GINGIVA FROM 16 CAPTIVE KOMODO DRAGONS (VARANUS KOMODOENSIS): NEW IMPLICATIONS FOR THE "BACTERIA AS VENOM" MODEL. *Journal of Zoo and Wildlife Medicine* **2013**, *44* (2), 262−272.

(4) Fry, B. G.; Wroe, S.; Teeuwisse, W.; van Osch, M. J. P.; Moreno, K.; Ingle, J.; McHenry, C.; Ferrara, T.; Clausen, P.; Scheib, H.; et al. A central role for venom in predation by Varanus komodoensis (Komodo Dragon) and the extinct giant Varanus (Megalania) priscus. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (22), 8969−8974.

(5) MERCHANT, M.; Henry, D.; Falconi, R.; Muscher, B.; Bryja, J. Antibacterial activities of serum from the Komodo Dragon (Varanus komodoensis). *Microbiol. Res.* **2013**, *4* (1), 16−20.

(6) Hancock, R. E.; Scott, M. G. The role of antimicrobial peptides in animal defenses. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (16), 8856−8861.

(7) Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **2002**, *415* (6870), 389−395.

(8) van Hoek, M. L. Antimicrobial peptides in reptiles. *Pharmaceuticals* **2014**, *7* (6), 723−753.

(9) Brogden, K. A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* **2005**, *3* (3), 238−250.

(10) Darville, L. N. F.; Merchant, M. E.; Hasan, A.; Murray, K. K. Proteome analysis of the leukocytes from the American alligator (Alligator mississippiensis) using mass spectrometry. *Comp. Biochem. Physiol., Part D: Genomics Proteomics* **2010**, *5* (4), 308−316.

(11) Stegemann, C.; Kolobov, A., Jr.; Leonova, Y. F.; Knappe, D.; Shamova, O.; Ovchinnikova, T. V.; Kokryakov, V. N.; Hoffmann, R. Isolation, purification and de novo sequencing of TBD-1, the first beta-defensin from leukocytes of reptiles. *Proteomics* **2009**, *9* (5), 1364−1373.

(12) Mor, A.; Nguyen, V. H.; Delfour, A.; Migliore-Samour, D.; Nicolas, P. Isolation, amino acid sequence, and synthesis of dermaseptin, a novel antimicrobial peptide of amphibian skin. *Biochemistry* **1991**, *30* (36), 8824−8830.

(13) Park, C. H.; Valore, E. V.; Waring, A. J.; Ganz, T. Hepcidin, a Urinary Antimicrobial Peptide Synthesized in the Liver. *J. Biol. Chem.* **2001**, *276* (11), 7806−7810.

(14) Lauth, X.; Shike, H.; Burns, J. C.; Westerman, M. E.; Ostland, V. E.; Carlberg, J. M.; Van Olst, J. C.; Nizet, V.; Taylor, S. W.; Shimizu, C.; et al. Discovery and Characterization of Two Isoforms of Moronecidin, a Novel Antimicrobial Peptide from Hybrid Striped Bass. *J. Biol. Chem.* **2002**, *277* (7), 5030−5039.

(15) Bishop, B. M.; Juba, M. L.; Devine, M. C.; Barksdale, S. M.; Rodriguez, C. A.; Chung, M. C.; Russo, P. S.; Vliet, K. A.; Schnur, J. M.; van Hoek, M. L. Bioprospecting the American Alligator (Alligator mississippiensis) Host Defense Peptidome. *PLoS One* **2015**, *10* (2), e0117394.

(16) Juba, M. L.; Russo, P. S.; Devine, M.; Barksdale, S.; Rodriguez, C.; Vliet, K. A.; Schnur, J. M.; van Hoek, M. L.; Bishop, B. M. Large Scale Discovery and De Novo-Assisted Sequencing of Cationic Antimicrobial Peptides (CAMPs) by Microparticle Capture and Electron-Transfer Dissociation (ETD) Mass Spectrometry. *J. Proteome Res.* **2015**, *14* (10), 4282−4295.

(17) Aronesty, E. Comparison of sequencing utility programs. *Open Bioinf. J.* **2013**, *7*, 1−8.

(18) Kong, Y. Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* **2011**, *98* (2), 152−153.

(19) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29* (7), 644−652.

(20) Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8* (8), 1494−1512.

(21) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Mol. Cell. Proteomics* **2012**, *11* (4), M111.010587−M111.010587.

(22) Liu, X.; Shan, B.; Xin, L.; Ma, B. Better score function for peptide identification with ETD MS/MS spectra. *BMC Bioinf.* **2010**, *11* (Suppl 1), S4−S8.

(23) Thomas, S.; Karnik, S.; Barai, R. S.; Jayaraman, V. K.; Idicula-Thomas, S. CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Res.* **2010**, *38* (Database), D774−D780.

(24) Waghu, F. H.; Gopi, L.; Barai, R. S.; Ramteke, P.; Nizami, B.; Idicula-Thomas, S. CAMP: Collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Res.* **2014**, *42* (D1), D1154−D1158.

(25) Amer, L. S.; Bishop, B. M.; van Hoek, M. L. Antimicrobial and antibiofilm activity of cathelicidins and short, synthetic peptides

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX

against Francisella. *Biochem. Biophys. Res. Commun.* **2010**, *396* (2), 246−251.
(26) Han, S.; Bishop, B. M.; van Hoek, M. L. Antimicrobial activity of human beta-defensins and induction by Francisella. *Biochem. Biophys. Res. Commun.* **2008**, *371* (4), 670−674.
(27) Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **2007**, *449* (7164), 819−826.
(28) Tosi, M. F. Innate immune responses to infection. *J. Allergy Clin. Immunol.* **2005**, *116* (2), 241−249.
(29) Waghu, F. H.; Gopi, L.; Barai, R. S.; Ramteke, P.; Nizami, B.; Idicula-Thomas, S. CAMP: Collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Res.* **2014**, *42* (D1), D1154−D1158.
(30) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764* (12), 1811−1822.
(31) *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically; Approved Standard*, 9th ed.; CLSI document M07-A9; CLSI: Wayne, PA, 2012; pp 1−88.
(32) Fazakerley, J.; Crossley, J.; McEwan, N.; Carter, S.; Nuttall, T. In vitro antimicrobial efficacy of β-defensin 3 against Staphylococcus pseudintermedius isolates from healthy and atopic canine skin. *Veterinary Dermatology* **2010**, *21* (5), 463−468.
(33) Dean, S. N.; Bishop, B. M.; van Hoek, M. L. Susceptibility of Pseudomonas aeruginosa Biofilm to Alpha-Helical Peptides: D-enantiomer of LL-37. *Front. Microbiol.* **2011**, *2*, 1−11.
(34) Dean, S. N.; Bishop, B. M.; van Hoek, M. L. Natural and synthetic cathelicidin peptides with anti-microbial and anti-biofilm activity against Staphylococcus aureus. *BMC Microbiol.* **2011**, *11* (1), 114.
(35) Hirsch, J. G. BACTERICIDAL ACTION OF HISTONE. *J. Exp. Med.* **1958**, *108* (6), 925−944.
(36) Hiemstra, P. S.; Eisenhauer, P. B.; Harwig, S. S.; van den Barselaar, M. T.; van Furth, R.; Lehrer, R. I. Antimicrobial proteins of murine macrophages. *Infect. Immun.* **1993**, *61* (7), 3038−3046.
(37) Kawasaki, H.; Isaacson, T.; Iwamuro, S.; Conlon, J. M. A protein with antimicrobial activity in the skin of Schlegel's green tree frog Rhacophorus schlegelii (Rhacophoridae) identified as histone H2B. *Biochem. Biophys. Res. Commun.* **2003**, *312* (4), 1082−1086.
(38) Patat, S. A.; Carnegie, R. B.; Kingsbury, C.; Gross, P. S.; Chapman, R.; SCHEY, K. L. Antimicrobial activity of histones from hemocytes of the Pacific white shrimp. *Eur. J. Biochem.* **2004**, *271* (23−24), 4825−4833.
(39) Richards, R. C.; O'Neil, D. B.; Thibault, P.; Ewart, K. V. Histone H1: An Antimicrobial Protein of Atlantic Salmon (Salmo salar). *Biochem. Biophys. Res. Commun.* **2001**, *284* (3), 549−555.
(40) Howell, S. J.; Wilk, D.; Yadav, S. P.; Bevins, C. L. Antimicrobial polypeptides of the human colonic epithelium. *Peptides* **2003**, *24* (11), 1763−1770.
(41) Rose, F. R.; Bailey, K.; Keyte, J. W.; Chan, W. C.; Greenwood, D.; Mahida, Y. R. Potential role of epithelial cell-derived histone H1 proteins in innate antimicrobial defense in the human gastrointestinal tract. *Infect. Immun.* **1998**, *66* (7), 3255−3263.
(42) Kim, H. S.; Cho, J. H.; Park, H. W.; Yoon, H.; Kim, M. S.; Kim, S. C. Endotoxin-Neutralizing Antimicrobial Proteins of the Human Placenta. *J. Immunol.* **2002**, *168* (5), 2356−2364.
(43) Park, C. B.; Kim, M. S.; Kim, S. C. A novel antimicrobial peptide from Bufo bufo gargarizans. *Biochem. Biophys. Res. Commun.* **1996**, *218* (1), 408−413.
(44) Birkemo, G. A.; Lüders, T.; Andersen, Ø.; Nes, I. F.; Nissen-Meyer, J. Hipposin, a histone-derived antimicrobial peptide in Atlantic halibut (Hippoglossus hippoglossus L.). *Biochim. Biophys. Acta, Proteins Proteomics* **2003**, *1646* (1−2), 207−215.
(45) Park, I. Y.; Park, C. B.; Kim, M. S.; Kim, S. C. Parasin I, an antimicrobial peptide derived from histone H2A in the catfish, Parasilurus asotus. *FEBS Lett.* **1998**, *437* (3), 258−262.
(46) Fernandes, J. Isolation and characterisation of oncorhyncin II, a histone H1-derived antimicrobial peptide from skin secretions of rainbow trout, Oncorhynchus mykiss. *Dev. Comp. Immunol.* **2004**, *28* (2), 127−138.
(47) Frohm, M.; Gunne, H.; Bergman, A. C.; Agerberth, B.; Bergman, T.; Boman, A.; Lidén, S.; Jörnvall, H.; Boman, H. G. Biochemical and antibacterial analysis of human wound and blister fluid. *Eur. J. Biochem.* **1996**, *237* (1), 86−92.

DOI: 10.1021/acs.jproteome.6b00857
J. Proteome Res. XXXX, XXX, XXX−XXX