Journal of proteome research

Article

pubs.acs.org/jpr

Large Scale Discovery and De Novo-Assisted Sequencing of Cationic Antimicrobial Peptides (CAMPs) by Microparticle Capture and Electron-Transfer Dissociation (ETD) Mass Spectrometry

Melanie L. Juba,[†] Paul S. Russo,[§] Megan Devine,[†] Stephanie Barksdale,[⊥] Carlos Rodriguez,[†] Kent A. Vliet,[#] Joel M. Schnur,[‡] Monique L. van Hoek,[⊥,¶] and Barney M. Bishop*[,†]

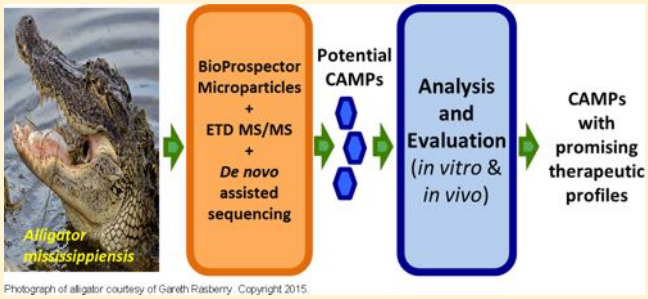
[†]Department of Chemistry and Biochemistry and [‡]College of Science, George Mason University, Fairfax, Virginia 22030, United States

[§]Center for Applied Proteomics and Molecular Medicine, [⊥]School of Systems Biology, and [¶]National Center for Biodefense and Infectious Diseases, George Mason University, Manassas, Virginia 20110, United States

[#]Department of Biology, University of Florida, Gainesville, Florida 32611, United States

Supporting Information

ABSTRACT: The identification and sequencing of novel cationic antimicrobial peptides (CAMPs) have proven challenging due to the limitations associated with traditional proteomics methods and difficulties sequencing peptides present in complex biomolecular mixtures. We present here a process for large-scale identification and de novo-assisted sequencing of newly discovered CAMPs using microparticle capture followed by tandem mass spectrometry equipped with electron-transfer dissociation (ETD). This process was initially evaluated and verified using known CAMPs with varying physicochemical properties. The effective parameters were then applied in the analysis of a complex mixture of peptides harvested from American alligator plasma using custom-made (Bioprospector) functionalized hydrogel particles. Here, we report the successful sequencing process for CAMPs that has led to the identification of 340 unique peptides and the discovery of five novel CAMPs from American alligator plasma.



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

INTRODUCTION

Cationic antimicrobial peptides (CAMPs) are produced by nearly all living organisms and are an essential part of the innate immune defense against invading pathogens in higher organisms.[1−3] CAMPs tend to be low molecular weight amphipathic peptides, combining both cationic and hydrophobic elements. These physicochemical properties complement those of bacterial membranes, which tend to be rich in anionic lipids, allowing these peptides to directly interact with bacteria in a nonreceptor mediated pathway and exert broad spectrum antimicrobial effectiveness. Despite their pervasive expression in nature, the occurrence of bacterial resistance to CAMPs has been limited and not as widespread as has been the case for conventional antibiotics.[4] Thus, CAMPs and their unique ability to exert direct broad spectrum antimicrobial effects have the potential to dramatically impact the development of future antibacterial therapeutics.[5] The discovery and cataloging of these remarkable peptides could unlock the key to overcoming antibiotic resistance. Unfortunately, the peptide diversity captured in extant CAMP databases is limited due to the technical challenges inherent to current CAMP discovery strategies and methods.

The approaches that have been used to discover and identify native CAMPs from biological samples have proven slow and low-yielding. Traditionally, new CAMPs have been discovered through fractionation methods or through the identification of conserved sequences in cDNA. However, it has been found that related CAMPs are not conserved at the amino acid level across species, (for example, SMAP-29, LL-37, and indolicidin are CAMPs of the cathelicidin class from three mammalian species, but they are not highly homologous) making this type of identification difficult. Current proteomic methods for CAMP discovery usually require large sample volumes of biological materials (frequently upward of 1 L), involve time-consuming HPLC or electrophoretic fractionation, and rely on enzymatic digestions coupled with collision induced dissociation (CID) mass spectrometry or MALDI-TOF coupled with Edman degredation to determine peptide sequences.[6−10] Large sample volumes and HPLC fractionation can lead to loss of low abundance peptides such as CAMPs as well as low activity in

Received: May 21, 2015
Published: September 1, 2015

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

biological assays. Enzymatic digestion of samples prior to mass spectrometry is problematic since information regarding the native form of peptides can be lost, which could result in incomplete peptide sequences and erroneous interpretation of active antimicrobial peptides. In addition, since CAMPs are known to contain an abundance of lysine and arginine residues, enzymatic digestion with commonly used trypsin can produce very small hydrophilic peptide fragments that are not retained on an HPLC column. While CAMPs are comparatively small peptides relative to other proteins and peptides present in plasma and other biological environments, they are large when compared to the peptide fragments typically generated by proteolytic digestion for analysis by mass spectrometry. Unfortunately, CID fragmentation efficiency drops off significantly for peptides with gas phase charge states of +4 or greater. Therefore, this type of fragmentation is not well suited for larger, more highly charged peptides such as intact CAMPs,[11] which can demonstrate gas phase charge states of up to +9. To overcome these issues, a new approach to CAMP discovery must be employed.

Recent advances in protein mass spectrometry have greatly improved mass accuracy resolving power, sensitivity limits, and data acquisition speed. Another recent advance comes in the form of new peptide fragmentation chemistries. One of these advances, electron-transfer dissociation (ETD), developed by Donald Hunt in 2004, is of particular relevance as it allows for the efficient fragmentation of larger, more highly charged peptides.[11−13] By combining ETD with high-resolution and high mass accuracy mass spectrometry, it becomes possible to effectively and efficiently sequence the larger, more highly charged CAMPs, and in a de novo manner when necessary. This negates the need for enzymatic digestion and allows for identification of intact, full-length native peptides.

Recently, we developed a new and effective method for CAMP identification that allows for the rapid extraction and analysis of the native, functional peptidome.[14] This method uses our Bioprospector microparticles to enrich intact, functional small peptides with CAMP-like characteristics from biological samples coupled with analysis of the captured peptides using ETD mass spectrometry (Figure 1). The Bioprospector microparticles consist of a 1:1 combination of two types of cross-linked poly(*N*-isopropylacrylamide)-based particles, one incorporating acrylic acid as its affinity bait and the other combining acrylic acid and 2-acrylamido-2-methylpropanesulfonic acid. The carboxylic and sulfonic acids in the particle matrices are deprotonated and electrostatically complement the positive charges associated with CAMPs. The cross-linking of the polymer framework helps to exclude larger proteins/peptides from binding effectively to the anionic baits. Initial analysis of the harvested peptides by mass spectrometry reveals a complex mixture of peptides. De novo sequencing of peptides from the highly complex mixture can be achieved with the assistance of PEAKS, a de novo sequencing software package.[15−17] PEAKS then uses sequence tags from the de novo sequences to search against any available database (de novo-assisted). In this study, we compare the ability of two mass spectrometers (LTQ-Orbitrap Elite and the LTQ XL) both equipped with ETD, in conjunction with PEAKS, to de novo sequence known full-length, functional CAMPs. This technique was then in turn implemented in the de novo and de novo-assisted sequencing of multiple novel CAMPs harvested from American alligator plasma.[14]

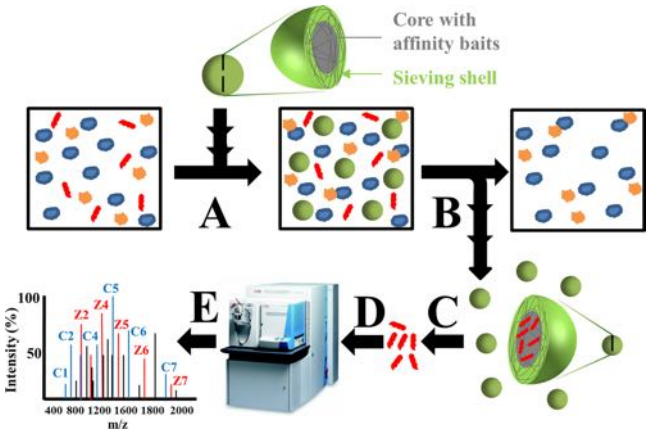


Figure 1. Bioprospecting approach to CAMP discovery: (A) Bioprospector hydrogel microparticles are introduced into the plasma sample. (B) The particles capture small cationic peptides, which are present in the sample, while excluding high molecular weight proteins. (C) The particles are then recovered, and (D) captured low molecular weight peptides are eluted from the particles and (E) analyzed by high-resolution MS/MS.

MATERIALS AND METHODS

Peptide Harvest and Elution

Peptides were harvested and eluted as previously described.[14] Briefly, plasma (100 μL) from American alligator blood that had been stimulated with ionomycin (1 μM, 30 min, 30 °C) was diluted into 1.6 mL of Bioprospector hydrogel particles[14] (40 mg) suspended in 10 mM Tris-Cl buffer (particle suspension = pH 5) for a final volume of ∼1.7 mL. After incubating for approximately 18 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 resuspended in 10 mM Tris-Cl buffer (pH 7.4). This centrifugation and resuspension process was repeated two more times to ensure removal of excluded proteins and peptides. Following the final wash with Tris-Cl buffer, the pelleted particles were suspended in an elution buffer consisting of 1:1 trifluoroethanol (TFE)/0.1% TFA in water. The particles were gently agitated for 1 h at room temperature before pelleting (as described earlier). The supernatant, containing eluted captured peptides, was collected and set aside. The elution process was repeated three more times with suspended particles being incubated 20 min in elution buffer each time prior to pelleting. The elution supernatants were combined and dried via vacuum centrifugation before being desalted using a Zip-Tip (Millipore, Billerica, MA) solid phase extraction for mass spectrometry analysis.

LC−MS/MS

LC−MS/MS experiments carried out on the LTQ XL (ThermoFisher, Waltham, MA, USA) used nanospray equipped with an Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA). Peptides were separated using a reversed-phase manually packed 75 μm i.d. × 10 cm long with 5 μm, 200 Å pore size, C18 resin LC column (Michrom Bioresources, Auburn, CA). The mobile phase was a gradient prepared from 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 10 min with A; the peptides were eluted by using a linear gradient from 0 to 50% B over 45 min and ramping to 100% B for an additional 2 min; the flow rate was 300 nL/min. The LTQ-XL was operated in a data-

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

dependent mode in which each full MS scan was followed by five MS/MS scans in which the five most abundant molecular ions were dynamically selected and fragmented by electron transfer dissociation (ETD) using fluoranthene as the electron transfer reagent and a reaction time of 100 ms. Supplemental activation was also applied.

LC−MS/MS experiments carried out on the LTQ-Orbitrap Elite (ThermoFisher Scientific, Waltham, MA, USA) used nanospray equipped with an EASY-nLC 1000 HPLC system (Thermo Fisher Scientific, Waltham, MA, USA). 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, USA). The mobile phase was a gradient prepared from 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 45 min and ramping to 100% B for an additional 2 min; the flow rate was 300 nL/min. The LTQ-Orbitrap Elite was operated in a data-dependent mode in which each full MS scan (120 000 resolving power) was followed by five MS/MS scans (120 000 resolving power) in which the five most abundant molecular ions were dynamically selected and fragmented by ETD using fluoranthene as the electron transfer reagent, using either one or three MS/MS microscans, using 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 ≥3+ ions are selected and fragmented by ETD. Supplemental activation was also applied.

Spectral Analysis

Mass spectra were directly imported as .RAW files and analyzed by SIEVE software (Thermo Scientific, Waltham, MA, USA) to determine how many ions were present along with their corresponding charge states. Analysis was done using non-differential single class analysis of the raw data using the following parameters: full m/z range (300−2000), 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.

Sequencing Analysis

Mass spectra were imported directly as .RAW files and analyzed by PEAKS de novo sequencing software version 6 (Bioinformatics Solutions Inc., Waterloo, ON Canada). PEAKS first performs a de novo sequence analysis of the ETD MS/MS data. Mass tolerance for precursor ions was 2 Da (LTQ-XL) or 10 ppm (LTQ-Orbitrap), and mass tolerance for fragment ions was 0.5 Da (LTQ-XL) or 0.05 Da (LTQ-Orbitrap). 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 sequence tags from the confident de novo sequences to search against two separate databases. The first database was an expressed sequence tag (EST) database obtained by searching the EST database at NCBI (http://www.ncbi.nlm.nih.gov) for all known alligator EST sequences. A total of 5469 alligator EST sequences were found from a number of sources including the Adult American Alligator Testis Library (University of Florida, Department of Zoology, Gainesville, FL), the Juvenile American Alligator Liver Library (NIBB, Japan), and the Adult American Alligator Liver Library (University of Florida, Department of Zoology, Gainesville, FL). The second database was an *Alligator mississippiensis* transcriptome obtained from the International Crocodilian Genome Working Group (www.crocgenomes.org).[18] A 1% false discovery rate (FDR) was used as a cutoff value for reporting peptide spectrum matches (PSM) from either database. For peptides of interest that were identified against at least one of the databases utilized, the sequences were 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. 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 since it is indistinguishable from isoleucine (I) by ETD fragmentation.

CAMP Prediction

A two-pronged approach was taken for predicting potential CAMPs, with potential CAMPs being determined based on the results from CAMP prediction algorithms as well as their calculated physicochemical properties. In the first approach, verified sequences were submitted for analysis by web-based CAMP prediction sites (*CAMP* database, *AntiBP2*, and *APD2*[19−21]), where each peptide was scored, and the likelihood that they were antimicrobial was predicted. Peptides that were predicted to have antimicrobial activity were selected for synthesis. In the other approach, the physicochemical properties (length, molecular weight, nominal solution charge, pI, and hydrophobicity) of all verified sequences were calculated. Peptide sequences that showed physicochemical properties similar to those of known CAMPs were also selected for synthesis to be evaluated for antimicrobial activity.

Antimicrobial Performance Assay

The antibacterial activity of the peptides was previously assessed using a resazurin-based assay.[14] Briefly, frozen enumerated bacterial aliquots were thawed on ice and mixed to uniformly suspend the cells. For each strain, bacteria were diluted to 2×10^6 CFU/mL in sterile 10 mM sodium phosphate (pH 7.4), and the bacterial suspensions were then added in 50 μL aliquots to the wells of a 96-well black microtiter plate that had already been charged with 50 μL volumes of various dilutions of CAMP dissolved in the same phosphate buffer. Control wells contain bacteria with no peptide. The microtiter plate was incubated for 3 h at 30 °C (*B. cereus*) or 37 °C (all other strains). After 3 h, 100 μL of PBS solution with dissolved resazurin and MHB was added to each well. Following addition of resazurin/MHB buffer, the plate was immediately placed in plate-reading fluorimeter for incubation overnight at either 30 °C (*B. cereus*) or 37 °C (all other strains) while fluorescence was monitored for each well. Bacterial survival results were determined and evaluated for each CAMP.[22]

RESULTS AND DISCUSSION

De Novo Sequencing of Known CAMPs

Since the de novo sequencing of CAMPs using mass spectrometry has only recently become an area of interest, the best methods for achieving this objective have yet to be

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

Table 1. Known CAMP Properties. Sequences, Length, MW, Charge, pI, and Hydrophobicity of Each Known CAMP Used for Validating De Novo Sequencing Methods

peptide	actual sequence	length (res)	molecular weight (Da)	net charge[a]	pI	hydrophobicity[b]
SMAP-29	RGLRRLGRKIAHGVKKYGPTVLRIIRIAG	29	3254.03	9	12.31	−0.21
Buforin	TRSSRAGLQFPVGRVHRLLRK	21	2433.43	6	12.60	−0.64
Indolicidin	ILPWKWPWWPWRR-NH2	13	1906.03	3	12.01	−1.07

[a]Net solution charge at pH 7. [b]Calculated using the GRAVY hydrophobicity scale.

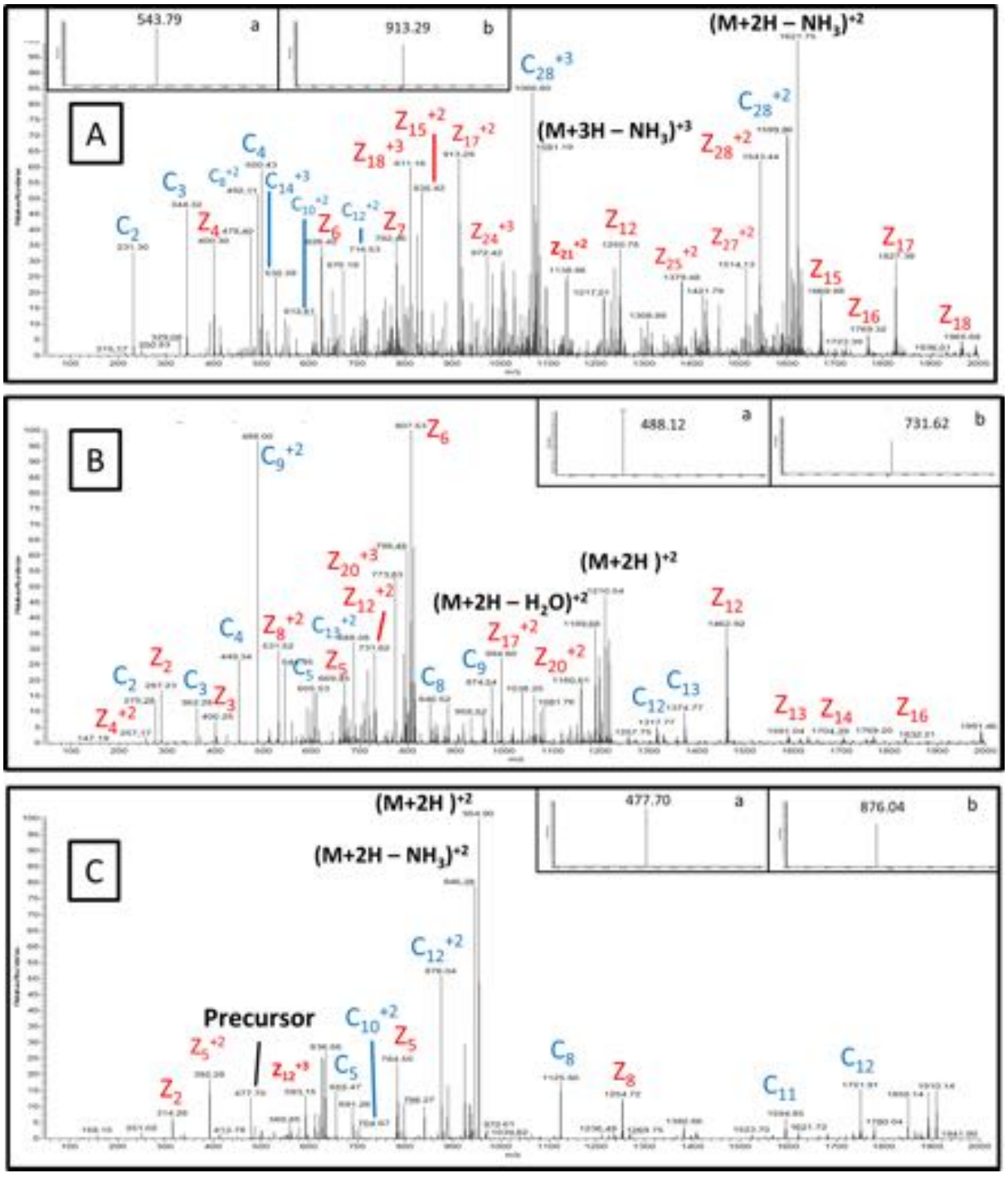


Figure 2. MS/MS spectra of SMAP-29, Buforin, and Indolicidin on LTQ-ETD. The MS/MS spectrum of (A) SMAP-29 is presented along with insets showing the (a) precursor (M+6H)+6 ion at m/z 543.79 and (b) a doubly charged ETD z-ion at m/z 913.29. The MS/MS spectrum of (B) Buforin is presented along with insets showing the (a) precursor (M+H)+5 ion at m/z 488.12 and (b) a doubly charged ETD z-ion at m/z 731.62. The MS/MS spectrum of (C) Indolicidin is presented along with insets showing the (a) precursor (M+4H)+4 ion at m/z 477.75 and (b) a doubly charged ETD c-ion at m/z 876.04. The c-ions are denoted in blue and z-ions in red.

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

Table 2. Comparison of PEAKS Identification with LTQ-ETD versus Orbitrap-ETD Data. Three Known CAMPs Were Run on Both Thermo LTQ-ETD and Orbitrap-ETD Instruments and Analyzed Using PEAKS De Novo Sequencing Software. De Novo Sequences Determined by PEAKS Are Compared to the Actual CAMP Sequences

		% correct
	SMAP-29	
actual sequence	RGLRRLGRKIAHGVKKYGPTVLRIIRIAG	
LTQ-ETD[a]	RGLRRLGSLAFGMCGSSALKLFHLWLRLV	24.1
Orbitrap-ETD	RGLRLRGRRRHGVKKYPGTVLRLLARVA	82.3
	Buforin	
actual sequence	TRSSRAGLQFPVGRVHRLLRK	
LTQ-ETD[a]	TRSSRAGLKSKKGSSSNLRAAHH	38.1
Orbitrap-ETD	TRSSRAGLQKTNGRVHRLLRK	85.7
	Indolicidin	
actual sequence	ILPWKWPWWPWRR-NH2	
LTQ-ETD[a]	LPLWKEAERELWPLK(−0.98)	38.4
Orbitrap-ETD	LPLWKEGPWQRWRR(−0.98)	76.9

[a]Precursor charge state could not be determined by PEAKS, so charge state was manually corrected to obtain a de novo sequence.

established. Therefore, initial studies focused on three known CAMPs (SMAP-29, Buforin, and Indolicidin)[24−29] to establish and validate a workflow and parameters for the successful de novo sequencing of CAMPs. These three model CAMPs are well characterized and vary in their length, molecular weight, net charge, pI, and hydrophobicity (Table 1). The three known model CAMPs were selected for de novo sequencing method development based on their varied physicochemical properties to ensure that different classes of CAMPs were represented.

Initial sequencing of the three known CAMPs was performed using a Thermo LTQ XL mass spectrometer equipped with ETD (LTQ-ETD). Because of the lower resolution of the LTQ-ETD, it was not possible to determine the charge states of either the highly charged peptide precursor ions or of the multiply charged ETD fragment ions (Figure 2). Without the correct precursor charge states available, PEAKS was not able to calculate the correct precursor peptide masses. For PEAKS to perform a de novo analysis of the three known CAMPs, the precursor charge states had to be manually adjusted. Once this was done, however, PEAKS still produces low confidence de novo sequences (Table 2). PEAKS was able to correctly identify only seven of the 29 amino acids for SMAP-29 (24%), eight of the 21 amino acids for buforin (38%), and five of the 13 amino acids for indolicidin (38%), despite all three spectra containing almost a complete c- and z-ion series. The low sequence accuracy was most likely due to the fact that the LTQ-ETD provides low-mass accuracy for both the precursor and the fragment ions.

Since de novo sequencing of these known highly charged CAMPs using the LTQ-ETD and PEAKS yielded poor results, subsequent analyses were done on an LTQ-Orbitrap Elite, which offers very high resolution and mass accuracy. The increased mass accuracy and resolution allow for the correct determination of the charge states and the monoisotopic masses for both the precursor and fragment ions. This is illustrated in the spectra generated for the three model CAMPs, which have been provided in Figure 3. For these highly charged peptides, the de novo sequencing ability of PEAKS was greatly enhanced when analyzing the higher resolution and accuracy spectra generated using the LTQ-Orbitrap Elite (Table 2). This time, PEAKS was able to correctly assign 23 of the 29 amino acids for SMAP-29 (79%), 18 of the 21 amino acids for buforin (86%), and 10 of the 13 amino acids for indolicidin (77%). Although the de novo sequences generated by PEAKS were not 100% correct, they provide excellent templates, and discrepancies with the actual peptide sequences can be detected with manual verification and corrected since the ETD spectra for each peptide contain nearly complete c- and z-ion series.

Sequencing of Novel Alligator CAMPs

The high resolution and accuracy of the Orbitrap were needed in combination with PEAKS for the effective de novo sequencing of the three known CAMPs. Therefore, the Orbitrap alone was used as part of the Bioprospector hydrogel particle-based CAMP discovery process for the identification of novel CAMPs from alligator plasma.[14] LC−MS/MS analysis of the microparticle captured alligator peptides was performed using the LTQ-Orbitrap Elite using one MS/MS microscan. However, in an effort to obtain better quality ETD spectra by increasing the signal-to-noise ratio, a second analysis was performed using three MS/MS miocroscans. Table 3 shows the total number of ions with charges from +3 to +8 (calculated using SIEVE) along with the total number of MS/MS spectra generated, the total number of de novo sequences above 30% ALC, and the total number of peptide spectrum matches and unique peptides from the Alligator EST and transcriptome databases identified using PEAKS acquired with both one and three MS/MS microscans. As expected, the increase in scan cycle time using three microscans resulted in a decrease in the total number of ions detected with charge states between +3 and +8 by approximately 20−25%, while the total number of MS/MS spectra acquired decreased by over 60%.

However, increasing the number of microscans from one to three did not significantly improve the overall quality of the resulting ETD spectra. Examples are provided in Figure 4, where ETD spectra collected using one and three MS/MS microscans for the same precursor peptide are compared. No enhancement of spectral quality is apparent when acquiring the data with three microscans, with the spectra appearing extremely similar in both the c- and z-ions that are present. Switching from one MS/MS microscan to three does not appear to significantly improve signal-to-noise or relative abundance in the spectra. In fact, the PEAKS de novo score for the one microscan sequence (65% ALC) was actually higher than the one for the three microscan sequence (44% ALC). Since there was only about a 30% difference in the abundance of precursor peptide immediately preceding each ETD spectrum (as seen in the insets), this should not be a major contributing factor in the quality of the MS/MS spectra.

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

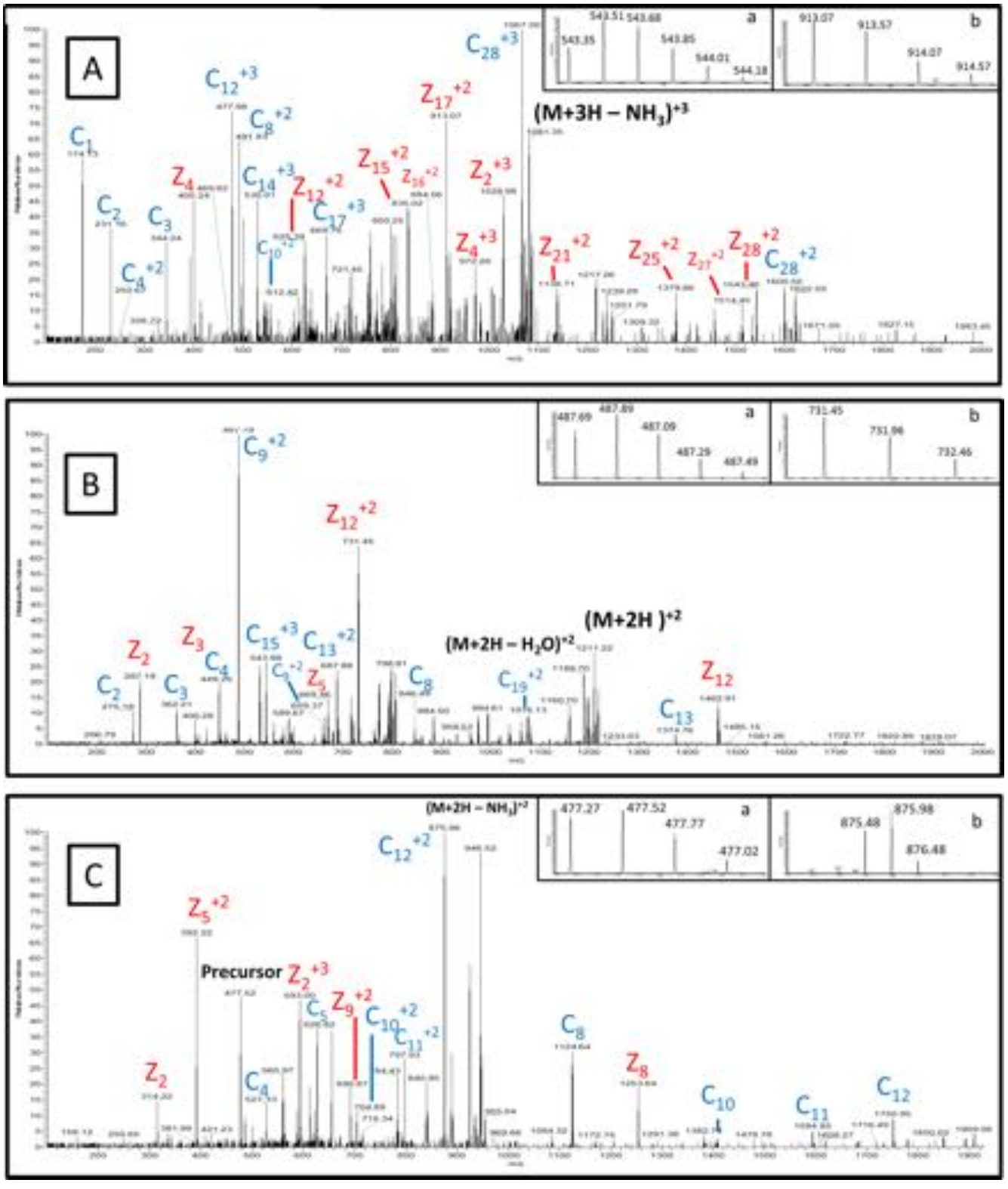


Figure 3. MS/MS spectra of SMAP-29, Buforin, and Indolicidin on Orbitrap-ETD. The MS/MS spectrum of (A) SMAP-29 is presented along with insets showing the (a) precursor (M+6H)+6 ion at m/z 543.35 and (b) a doubly charged ETD z-ion at m/z 913.07. The MS/MS spectrum of (B) Buforin is presented along with insets showing the (a) precursor (M+H)+5 ion at m/z 487.69 and (b) a doubly charged ETD z-ion at m/z 731.45. The MS/MS spectrum of (C) Indolicidin is presented along with insets showing the (a) precursor (M+4H)+4 ion at m/z 477.27 and (b) a doubly charged ETD c-ion at m/z 875.48. The c-ions are denoted in blue and z-ions in red.

The percentages of high-confidence de novo sequences or database PSMs from the total number of MS/MS spectra acquired provide additional evidence suggesting that acquiring data with three MS/MS microscans did not significantly improve overall data quality. The percentages of de novo sequences from the number of MS/MS spectra for one microscan and three

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282–4295

Table 3. Comparison of LC−MS/MS Analyses of Alligator Peptides for One versus Three Microscans. MS and MS/MS Data of One and Three Microscans Were Compared To Identify Changes in Data Collection Based on Total Ions, Total MS/MS Spectra Collected, and Total De Novo Sequence Count

number of microscans	total number of ions with charge +3 to +8	total number of MS/MS spectra	total number of de novo sequences with charge +3 to +8	total number of EST database PSMs with charge +3 to +8	total number of EST database unique peptides with charge +3 to +8	total number of transcriptome database PSMs with charge +3 to +8	total number of transcriptome database unique peptides with charge +3 to +8
1	4585	2900	1349	545	269	526	272
3	3553	1137	562	223	131	202	123

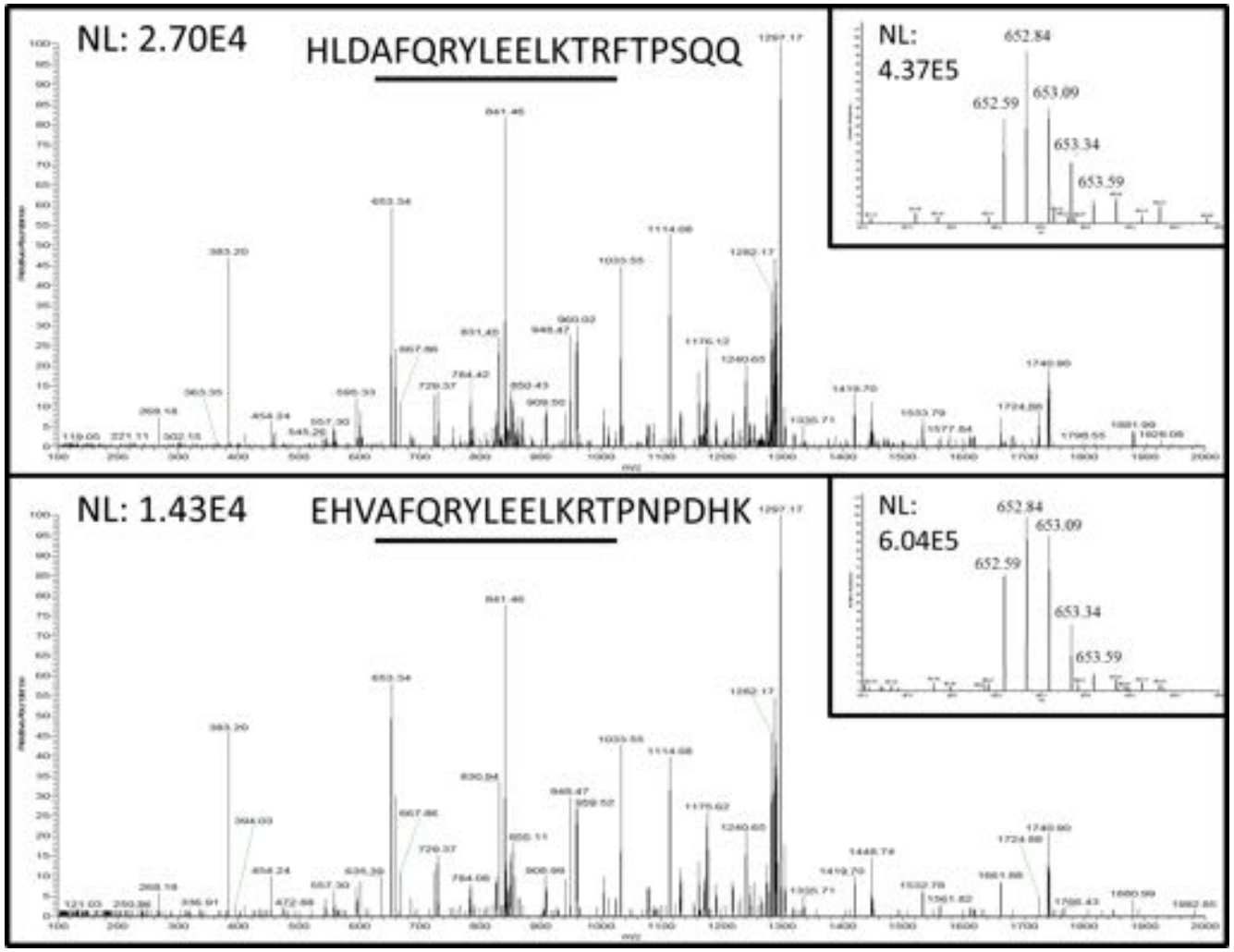


Figure 4. Comparison of one versus three microscan ETD spectral quality. Top panel: spectra collected using one microscan with the precursor ion shown in the inset. Bottom panel: spectra collected using three microscans with the precursor ion shown in the inset. De novo sequences identified by PEAKS are shown at the top of each spectra with the matching portion of the sequences underlined.

microscans were 46%, and 49%, respectively. The percentages of PSMs from the number of MS/MS spectra using the EST database for one microscan and three microscans were 19% and 20%, respectively. In addition, the percentages of PSMs from the number of MS/MS spectra using the transcriptome database for one microscan and three microscans were both 18%. If acquiring spectra with three microscans produced higher quality data, one would expect a larger number of de novo sequences and PSMs relative to the total number of MS/MS spectra generated. However, this was not the case.

Finally, the last piece of evidence suggesting that acquiring data with three MS/MS microscans did not improve the overall data quality is provided by de novo sequence and database search scores. The average ALC for all of the de novo sequences was 48% for both the one microscan and three microscans data. The average −10LogP score for all PSMs was 92 for one microscan and 98 for three microscans when searching against the EST database, and the average −10LogP score for all PSMs was 71 for one microscan and 59 for three microscans when searching against the alligator transcriptome database. One would expect higher −10LogP score values for data acquired with three microscans versus one microscan. However, this has proven not the case. While the average EST database −10LogP score was slightly higher with three microscans, the average transcriptome database PSM score for one microscan data was significantly higher than that afforded by three microscans data. Furthermore,

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

the average de novo sequence scores attained using either one or three microscans data are identical. The Venn diagrams in Figure 5 compare the number of unique peptide identifications from each database determined using one and three MS/MS microscans data. For the EST database search, 87% of the peptides determined based on three microscans data were also identified using one microscan data, while 88% of the peptides identified from the alligator transcriptome database were common to analyses using both three microscans and one microscan data.

Although we had expected that using three MS/MS microscans would result in higher quality ETD spectra, thereby yielding higher peptide sequence scores and probabilities, this did not occur. Using one MS/MS microscan resulted not only in a much greater number of MS/MS spectra as expected, but also it also afforded de novo sequences and database peptide identifications with equal or near equal scores and probabilities relative to analyses using data acquired from three microscans.

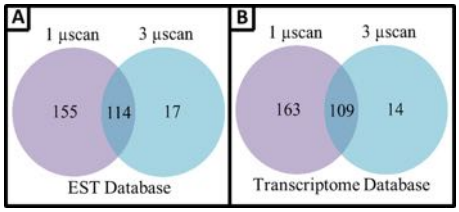


Figure 5. Comparison of the number of unique peptide identifications using one MS/MS microscan versus three MS/MS microscans searching against the (A) Alligator EST database and the (B) Alligator Transcriptome database.

Table 4. Physico-Chemical Properties of Potential CAMPs Identified from Alligator Plasma Analyzed by ETD LC–MS/MS. Source Protein, Sequences, Length, MW, Charge, pI, and Hydrophobicity of Identified Peptide Sequences. Peptides Highlighted in Yellow Were Chosen for Synthesis and Antimicrobial Analysis. Peptides Highlighted in Blue Were De Novo-Only Sequences. Peptides Highlighted in Green Were De Novo-Only Sequences Chosen for Synthesis and Antimicrobial Analysis

Peptide Sequence	Length (res)	Molecular Weight (Da)	Net Charge[†]	pI	Hydrophobicity[‡]	Source Protein
FVLKSFAQARRY	12	1485.75	3	11.00	-0.11	Alpha-2-macroglobulin precursor
RESIKPYTESIKTHL	15	1802.06	1	8.50	-1.04	Apolipoprotein A-I precursor
VKDLSRQKLEL	11	1328.57	1	8.56	-0.73	Apolipoprotein A-I precursor
KSRVNRMKQNL	11	1372.77	4	12.02	-1.65	Apolipoprotein B-I
KSRVNRM(+15.99)KQNL	11	1388.76	4	12.02	-1.65	Apolipoprotein B-I
EHFKKVKEKLK	11	1413.73	3	9.83	-1.72	Apolipoprotein C-I precursor
FKKVKEKLKDTFA	13	1581.92	3	9.83	-0.91	Apolipoprotein C-I precursor
KVFADNIGEKTKA	13	1420.63	1	8.50	-0.63	Apolipoprotein C-I precursor
EHFKKVKEKLKDTFA	15	1848.18	2	9.41	-1.23	Apolipoprotein C-I precursor
FSEHFKKVKEKLKDTFA	16	2082.43	2	9.40	-0.97	Apolipoprotein C-I precursor
KTRNWFSEHFKKVKEKLKDTFA	22	2766.49	4	10.00	-1.36	Apolipoprotein C-I precursor
STKTRNWFSEHFKKVKEKLKDTFA	24	2954.57	4	10.00	-1.31	Apolipoprotein C-I precursor
FSTKTRNWFSEHFKKVKEKLKDTFA	25	3103.57	4	10.00	-1.17	Apolipoprotein C-I precursor
ALRDQGQRLREQL	13	1582.78	1	9.56	-1.4	Apolipoprotein E precursor
NKGKIVQAGRQLRQAGQNL	19	2078.18	4	12.02	-1.01	Complement 3
ILNKGKIVQAGRQLRQAGQNL	21	2304.35	4	12.02	-0.51	Complement 3
DRLEELREFANFDFDIWRKKYMRWMNHKKSRVMDFFRRI	39	5224.63	4	10.14	-1.11	Dystonin
MHLKKPVAFSNF	12	1418.72	2	10.00	0.03	F2 Prothrombin precursor
YSLKKTSMKIIPFTRL	16	1926.39	4	10.46	-0.05	Fibrinogen
YSLKKTSM(+15.99)KIIPFTRL	16	1942.38	4	10.46	-0.05	Fibrinogen
KKTSMKIIPFTRL	13	1562.92	4	11.26	-0.19	Fibrinogen
KKTSM(+15.99)KIIPFTRL	13	1578.97	4	11.26	-0.19	Fibrinogen
WLQNEKIHL	9	1109.29	0	6.75	-0.37	Fibrinogen
WKQSWYSLK	9	1154.33	2	9.70	-1.01	Fibrinogen
YSLKKMSMKIRPFFPQ	16	2000.07	4	10.46	-0.51	Fibrinogen
KKMSMKIRPFFPQ	13	1636.89	4	11.26	-0.75	Fibrinogen
PALKYVRPGGGFAPNFQL	18	1932.26	2	10.01	-0.09	Glutathione peroxidase 3 precursor
DMSHNSAQIRAHQKKVFSAL	20	2197.5	2	9.99	-0.47	Hemoglobin subunit alpha
NSAQIRAHQKKVFSAL	16	1727	3	11.27	-0.24	Hemoglobin subunit alpha
ASFQEAVKHLDNIKQHFANL	20	2168.44	0	6.96	-0.13	Hemoglobin subunit epsilon
KRTFTPSQAG	10	1092.22	2	11.00	-1.15	Hypothetical Protein
KFIQKSVQKQPG	12	1387.65	3	10.30	-1.13	Immunoglobulin
KFTQRSIQKTAG	12	1364.57	3	11.17	-1.07	Immunoglobulin
KFSQRSVQKSPGN	13	1462.63	3	11.17	-1.55	Immunoglobulin
ALPMKFIQKSVQKQPG	16	1800.19	3	10.30	-0.48	Immunoglobulin
VERIPLVRFKSIKKQLHERGDL	22	2656.17	3	10.27	-0.62	NOTS1
RPQASPRKKPRKQ	13	1445.85	5	12.02	-2.31	RAP 130
PPPVIKFNRPFLMWIVERDTRSILFMGKIVNPKAP	35	4106.28	4	11.00	0.02	Alpha-1-antiproteinase
YRFGKELVQSRKYR	14	1829	4	10.43	-1.52	Titin
EQQTRFGR	8	1022.1	1	9.70	-2.23	Transferrin
LQTKLKKLLQLESVF	15	1717.11	2	9.70	0.36	Vitellogenin
KVPRVKEHSKQK	12	1391.83	4	10.46	-1.77	Vitellogenin
MFPLRSKYNRLTK	13	1658.93	4	11.10	-1.35	Unknown
TPVFPQRRRQSTNLRASPQ	19	2025.09	4	12.48	-0.94	Unknown

[†]Net solution charge at pH 7. [‡]Calculated using the GRAVY hydrophobicity scale.

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282–4295

Table 5. Antibacterial Performance Data for Alligator CAMPs (Adapted from Bishop et al.[14]). Antibacterial Performances against *E. coli*, *B. cereus*, *P. aeruginosa*, and *S. aureus* Are Expressed Qualitatively as Very Effective with EC50 within 1 log of LL-37 EC50 (++), Effective with EC50 between 1 and 2 logs of LL-37 EC50 (+), Slightly Effective (∓) with EC50 between 2 and 3 logs of LL-37 EC50, and Ineffective (−). LL-37 Is a Human CAMP That Was Used as a Reference Standard and a Basis for Comparing the Antibacterial Performances of the Alligator Peptides[23]

Peptide	Bacterium				*CAMP* database Prediction Score			*AntiBP2* Prediction Score	*APD2* Prediction Probability
	E.coli	*B.cereus*	*P.aeruginosa*	*S.aureus*	SVM	RF	DA	SVM	Qualitative
LL-37	++	++	++	++	NA	NA	NA	NA	NA
$APOC1_{64-88}$	+	+	++	-/+	-	-	-	-	+
$APOC1_{67-88}$	+	+	++	-/+	-	-	-	-	+
$FGG_{398-413}$	+	-/+	+	+	+	-	+	-	-
$FGG_{401-413}$	+	-/+	-/+	-	+	+	+	ND*	-
$A1P_{394-428}$	++	+	+	+	-	-	-	-	+
AVTG2LP	-	-	-	-	+	-	+	+	+
ASAP130LP	-	-	-	-	-	-	-	ND*	+
$NOTS_{17-38}$	-	-	-	-	+	+	-	+	+

*Could not be analyzed using *AntiBP2* because it requires the queried peptide sequence be ≥15 amino acids in length to generate a prediction score.

The reason for the lack of difference between one microscan and three microscans is not entirely clear. Some possible reasons, however, could be having too low of fill target values on the linear ion trap or orbitrap, nonoptimal thresholds for precursors to be chosen for MS/MS, or too low of a precursor isolation width. These parameters will be adjusted in future experiments. Between the EST and transcriptome databases, a total of 340 unique peptides were identified using one MS/MS microscan. (See Supplemental Tables S1−S6 for the complete list of de novo sequences and database peptide identifications.)

Adjusting the number of microscans was only one method of potentially increasing MS/MS spectral quality and was meant only to be a first pass attempt to do so. Other possible means of improving spectral quality include increasing fill times, increasing until target values on the ion trap and the Orbitrap, and adjusting precursor isolation widths. Future experiments will investigate how effective these types of adjustments are at improving MS/MS spectral quality and the resulting sequence scores and probabilities that are generated. Additionally, incorporation of HCD fragmentation in addition to ETD could yield improvements in the sequencing of large, highly charged peptides. For this work, ETD alone was chosen as a first pass attempt for sequencing as many peptides as possible since there was very large variation in the peptide gas-phase charge states (from +3 to +9). At the outset of this work, ETD appeared to allow efficient fragmentation of most of the population of peptides. Since both MS and MS/MS were being acquired in the Orbitrap, the ETD fragmentation scans alone had a cycle time of approximately 1 s. While HCD could prove helpful, we were initially concerned about the increased cycle time that adding separate HCD scans to the analysis would incur. As we continue to refine the process in the future, we will investigate using longer LC gradient times and incorporating separate HCD scans along with ETD scans. Spectral quality could also be enhanced with the use of the technique known as EThcD by Heck et al.[30,31] EThcD uses a combination of ETD and HCD in a single MS/MS spectrum and has been used to sequence MHC peptides that have similar sequence characteristics to those associated with CAMPs. Unfortunately, the Orbitrap Elite used for this work did not have this capability.

Antimicrobial Prediction and Performance

After PEAKS had generated both de novo and database aligned sequences for the alligator peptides, the next step was to predict which sequences would likely have antimicrobial properties. A combination of web-based CAMP prediction algorithms and rational peptide sequence analysis based on known CAMP physicochemical properties was used to identify probable CAMP sequences for synthesis and testing. The sequences are submitted for analysis by three different CAMP prediction Web sites (*APD2*,[21] *CAMP* database,[20] and *AntiBP2*[19]) and their respective prediction models. Each prediction program employs its own set of models using various known CAMP properties as the basis for classification algorithms and machine learning algorithms to predict antimicrobial potential for each sequence, as we previously described.[14]

The rational analysis approach focuses on physicochemical properties that can be calculated based on the peptide sequences such as molecular weight, length, theoretical charge at physiological pH, peptide isoelectric point, and hydrophobicity (calculated using the GRAVY hydrophobicity scale). Only sequences corresponding to peptides with molecular weights of less than 5.5 kDa were considered because this is consistent with the molecular weights of the majority of known vertebrate CAMPs. Sequences predicted to have charges of +4 or higher at physiological pH were considered because high cationic character is believed to be linked with activity. Since peptide isoelectric point and hydrophobicity can vary among known CAMPs, these properties were not used as a primary consideration in determining antimicrobial potential.

These two approaches to CAMP identification (web-based CAMP-prediction algorithms and rational analysis) were used in combination so as to capture the greatest number of credible peptides with the potential for antimicrobial activity. This afforded two lists of potential CAMPs, one consisting of peptide sequences that showed positive CAMP prediction scores (Tables S7−S9) and another of peptides exhibited physicochemical properties associated with known CAMPs (Tables S10−S12).

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

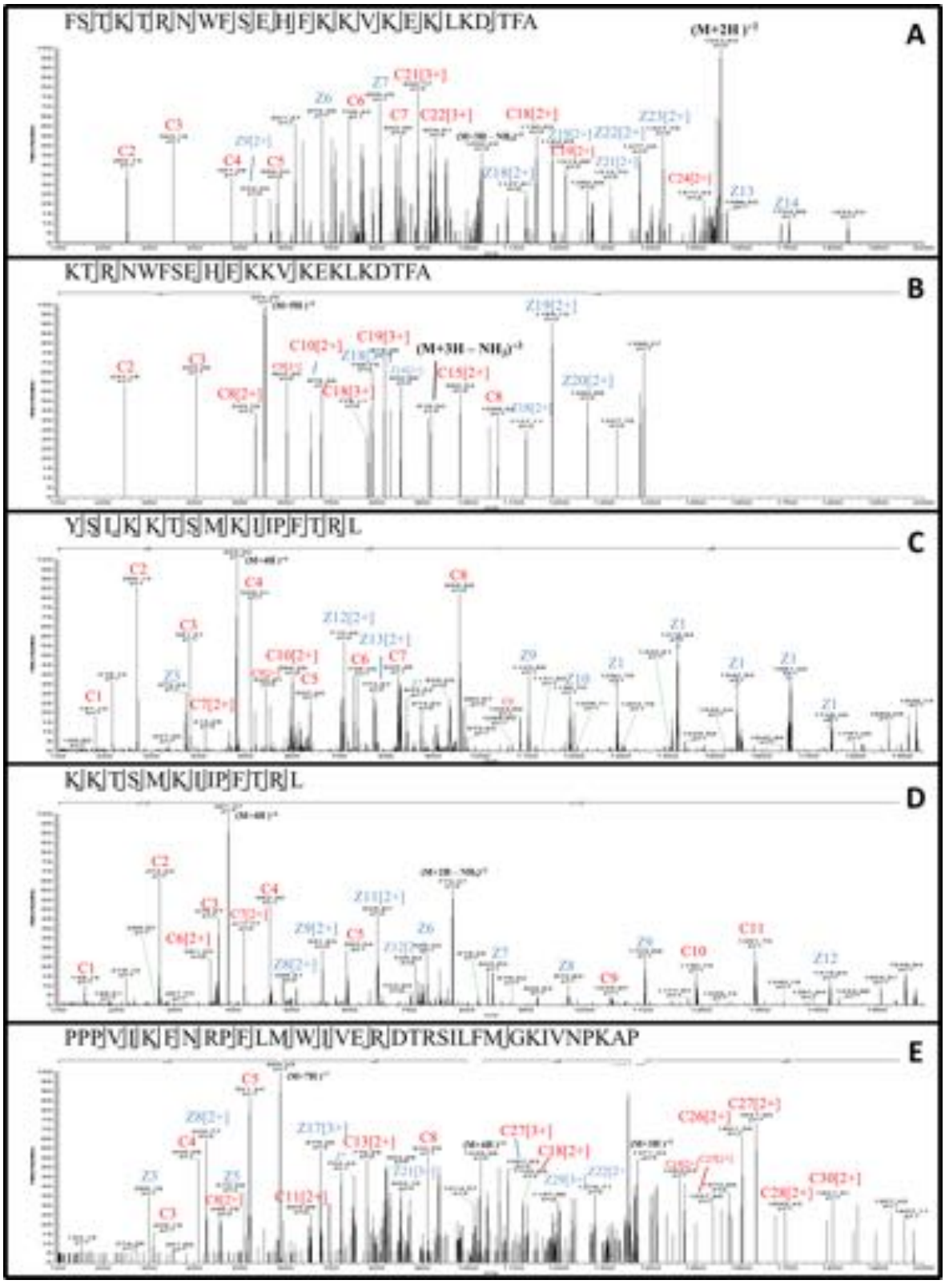


Figure 6. ETD spectra of five alligator peptides determined to have antimicrobial activity. The ETD spectra for (A) $APOC_{164-88}$, (B) $APOC_{167-88}$, (C) $FGG_{401-413}$, (D) $FGG_{398-413}$, and (E) $A1P_{394-428}$ are presented. Observed singly, doubly, and triply charged c- (blue) and z- (red) ions are indicated on the spectra. The underlined ions in the sequence indicate ions that are present in the spectrum.

The lists of potential CAMPs generated using the two prediction approaches were consolidated and duplicates eliminated. This combination of computational and rational analysis yielded 44 potential CAMPs (Table 4). Four of the peptides in the consolidated list of CAMP candidates were de novo-only sequences (highlighted in green and blue). To determine the source proteins for de novo sequenced peptides, BLAST searches of the peptide sequences were performed. The closest

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

probable parental source proteins were determined based on homology.

From this list, eight peptides [$APOC1_{64-88}$ (FSTKTRNWFSEHFKKVKEKLKDTFA), $APOC1_{67-88}$ (KTRNWFSEHFKKVKEKLKDTFA), $A1P_{394-428}$ (PPPVIKFNRPFLMWIVERDTRSILFMGKIVNPKAP), $FGG_{398-413}$ (YSLKKTSMKIIPFTRL), $FGG_{401-413}$ (KKTSMKIIPFTRL), AVTG2LP (LQTKLKKLLGLESVF), ASAP130LP (PPGASPRKKPRKQ), and $NOTS_{17-38}$ (VERIPLVRFKSIKKQLHERGDL)] were chosen for synthesis so their antimicrobial activities could be evaluated (Table 5). Peptides AVTG2LP, ASAP130LP, and $NOTS_{17-38}$ were de novo-only sequences, while the other five peptides were identified from the transcriptome and EST databases. Four of the peptides ($APOC1_{64-88}$, $APOC1_{67-88}$, $A1P_{394-428}$, and ASAP130LP) were selected because they exhibit relatively high theoretical net positive charges (ranging from +4 to +5) at pH 7. Peptides AVTG2LP and $NOTS_{17-38}$ were chosen due to the prediction algorithms' general (4 out of 5) agreement that they were likely CAMPs. The final two peptides ($FGG_{398-413}$ and $FGG_{401-413}$) were selected because of their overlap between both prediction methods.

As previously described,[14] the eight synthetic peptides were tested against a panel of Gram-positive and Gram-negative bacteria, which included *Bacillus cereus*, *Staphylococcus aureus*, *Escherichia coli*, and *Pseudomonas aeruginosa*. Results were previously reported in our work Bishop et al.[14] and are summarized qualitatively in this work (Table 5). The well-characterized human CAMP LL-37 was used in these assays as a reference standard for measuring antibacterial performance.[32−38] On the basis of the results of these assays, five of the eight peptides demonstrated significant antibacterial activity against one or more of the bacteria in the panel.[14] These five peptides were $APOC1_{64-88}$, $APOC1_{67-88}$, $A1P_{394-428}$, $FGG_{398-413}$, and $FGG_{401-413}$.

The five active CAMPs showed varied sequences and MS/MS spectral quality (Figure 6). By comparing the PEAKS de novo assigned sequences to the actual sequences, as determined by alignment with the alligator databases and manual verification, it is evident that PEAKS de novo sequencing works very well with some spectra and not as well with others (Table 6). As expected, this was highly dependent upon the spectral quality and the level of c- and z-ions present in each spectrum.

PEAKS de novo sequences for $APOC1_{64-88}$, $FGG_{398-413}$, and $FGG_{401-413}$ showed very good correlation to the actual sequences of these peptides. Notably, the sequence obtained for $APOC1_{64-88}$ had 92% correct identity with the actual sequence, with only the two N-terminal amino acids being incorrectly assigned. PEAKS gave 100% correct de novo sequence information for both $FGG_{398-413}$ and $FGG_{401-413}$. Because leucine and isoleucine are indistinguishable by ETD, and leucine occurs with greater frequency than isoleucine in soluble and membrane proteins,[39] leucine residues were used where the de novo sequences indicated the presence of an isoleucine/leucine. Upon inspection of the ETD spectra for these peptides, it is evident that they are high quality, and all three peptides have a nearly complete c- and z-ion series (Figure 6A,C,D).

The PEAKS de novo sequence for $APOC1_{67-88}$ had very little in common with the correct sequence, correctly identifying only 13% of the amino acids, despite a relatively high ALC score of 68%. Only the first three N-terminal amino acids were correctly assigned, although not in the correct order. The ETD spectrum for $APOC1_{67-88}$ (Figure 6B) presents very few fragment ions to

Table 6. Comparison of PEAKS De Novo Sequences versus PEAKS Database Sequences of the Five Active Alligator CAMPS. De Novo Sequences Determined by PEAKS Are Compared to the Manually Verified CAMP Sequences, and the Identity Correct Was Recorded. PEAKS De Novo and Database Scores for Each Sequence Are Also Reported

peptide	actual sequence	PEAKS de novo sequence	% identity	de novo ALC (%)	EST database score (−logP)	transcriptome database score (−logP)
$APOC1_{64-88}$	FSTKTRNWFSEHFKKVKEKLKDTFA	PHTKTRNFWSEHFKKVKEKLKDTFA	92.0	31	138.92	108.19
$APOC1_{67-88}$	KTRNWFSEHFKKVKEKLKDTFA	TKRKM(+15.99)HPPGLPVMPVPGPHVRTRY	13.6	68	51.17	41.83
$FGG_{398-413}$	YSLKKTSMKIIPFTRL	SYLKKTSMKLPLFTRL	100.0	72	117.87	94.39
$FGG_{401-413}$	KKTSMKIIPFTRL	KKTSMKLLPFTLR	100.0	72	101.02	76.90
$A1P_{394-428}$	PPPVIKFNRPFLMWIVERDTRSILFMGKIVNPKAP	YKVLQTVGMFKWGVSKRGRQSVGSLLGMFKLVVLTHG	31.4	74	78.44	63.33

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

use to assign the peptide sequence in a de novo manner. However, the few fragment ions present, along with other de novo sequence tags (data not shown), were enough to correctly identify the peptide with −logP scores of 51.17 and 41.83 when the spectrum was searched against the EST and transcriptome databases, respectively.

For $A1P_{394-428}$, PEAKS produced a de novo sequence that was only 31% correct relative to the actual sequence. Examination of the ETD spectrum (Figure 6E) reveals few high molecular weight fragment ions and a limited number of multiply charged ions, which makes the C-terminal amino acids difficult to identify. However, with the fragment ions that are present, along with other de novo sequence tags (data not shown), the same spectrum yielded −logP scores of 78.44 and 63.33 when searched against the EST and transcriptome databases, respectively, allowing proper identification of the peptide.

Although $NOTS_{17-38}$ and AVTG2LP were predicted likely to be effective antimicrobial peptides based on the CAMP prediction algorithms (Table 5), neither peptide demonstrated significant antibacterial activity against the bacteria panel (Table 5). These results suggest limitations in the reliability of currently available CAMP prediction models.[19−21] Although ASAP130LP appeared to have physicochemical properties consistent with known CAMPs, it similarly proved ineffective against the bacterial panel.

Sharing a nominal net charge of +4 at physiological pH, $APOC1_{64-88}$ (25aa) and $APOC1_{67-88}$ (22aa) are a nested pair of peptide fragments derived from apolipoprotein C1.

Both $APOC1_{64-88}$ and $APOC1_{67-88}$ showed significant antimicrobial activity against *E. coli*, *P. aeruginosa*, and *B. cereus* but were not as effective against *S. aureus*.[14] Relative to the performance of the known CAMP LL-37, $APOC1_{64-88}$ showed substantial activity against *E. coli*, *P. aeruginosa*, and *B. cereus*.[14] The shorter of the two peptides, $APOC1_{67-88}$, demonstrated increased potency, relative to the longer peptide $APOC1_{64-88}$, against *E. coli*, *P. aeruginosa*, *B. cereus*, and *S. aureus*.[14] The performance of $APOC1_{67-88}$ was similar to the effectiveness exhibited by LL-37. These results are very interesting since neither of these peptides were predicted to be strongly antimicrobial by the CAMP prediction algorithms. Only the APD2 database correctly predicted that these peptides would have antimicrobial activity. This again highlights the difficulties associated with predicting whether a peptide will be antimicrobial based on our current understanding of the factors governing peptide antimicrobial activity.

$A1P_{394-428}$, a 35-residue peptide fragment of alpha-1-antiproteinase, has a predicted +4 charge at neutral pH. Present in human plasma, alpha-1-antiproteinase, also known as alpha-1-antitrypsin, is a serine protease inhibitor (Serpin) and is a major protease inhibitor that has been linked to anti-inflammatory immune response.[40] Although predicted by four of the five CAMP-predictor algorithms to not show antimicrobial activity, $A1P_{394-428}$ demonstrated significant activity against *E. coli*, *B. cereus*, *S. aureus*, and *P. aeruginosa*.[14] Compared to the performance of LL-37 against the panel of bacteria, $A1P_{394-428}$ appears to show good broad-spectrum antimicrobial effectiveness.

Although the protein fibrinogen is associated with coagulation, which is initiated through proteolytic processing via thrombin, it has recently been suggested that treating fibrinogen with thrombin may also generate peptides that exhibit antimicrobial activity.[41] The two nested peptides derived from fibrinogen, $FGG_{398-413}$ (16aa) and $FGG_{401-413}$ (11aa), both carry a nominal charge of +4 at physiological pH. All but one of the prediction algorithms predicted $FGG_{401-413}$ to have antimicrobial activity while only two out of five algorithms predicted $FGG_{398-413}$ to be a CAMP. Interestingly, both $FGG_{398-413}$ and $FGG_{401-413}$ were found to have strong antimicrobial activity against *E. coli*. $FGG_{398-413}$ proved to be moderately effective against *S. aureus* and *P. aeruginosa*, while the shorter peptide $FGG_{401-413}$ proved less effective against both bacteria. Compared to the performance of LL-37, peptide $FGG_{398-413}$ presented mixed antimicrobial effectiveness, while peptide $FGG_{401-413}$ showed relatively poor antimicrobial activity against all of the bacteria tested, except for *E. coli*. The two FGG peptides seem to demonstrate a counter-example to the APOC1 peptides, which showed activity but were predicted to be inactive by the prediction algorithms with the sole exception being the APD2 database. In this case, the FGG peptides were correctly predicted to be inactive only by APD2.

The antimicrobial effectiveness exhibited by the eight selected peptides against *E. coli* (ATCC 25922) suggests that this bacterial strain is particularly sensitive to CAMP activity (Table 5).[14] In these studies, antimicrobial effectiveness against *E. coli* did not predict potency against the other bacterial strains tested, emphasizing the importance of testing potential CAMPs against multiple strains and Gram-types of bacteria to provide a more comprehensive view of their potential activity.

CONCLUSION

We have developed a new approach to identifying novel and potentially useful antimicrobial peptides, which aims to avoid the technical limitations associated with conventional approaches to CAMP discovery. In this process, Bioprospector hydrogel microparticles capture and enrich intact CAMPs based on their physicochemical properties. Subsequent analysis of the captured peptides using high-accuracy and high-resolution mass spectrometry allows identification of the sequences of these intact captured peptides. This approach to CAMP identification minimizes the number of steps separating the sample and mass spectral analysis, and it eliminates labor-intensive, low-yield purification and isolation processes associated with conventional proteomics approaches to CAMP discovery. Known CAMPs were used as models to establish the proper instrumentation, fragmentation techniques, and instrument parameters required for effective de novo sequencing of highly charged, high molecular weight peptides. Once the parameters and instrumental requirements for effective de novo sequencing of CAMPs were determined, they were applied in the identification of CAMPs from alligator plasma. In the process of identifying novel CAMP sequences from 100 μL of alligator plasma, the ETD MS/MS spectra for harvested peptides were analyzed using PEAKS de novo sequencing software, which greatly enhanced de novo and de novo-assisted sequencing throughput and sensitivity. The PEAKS software provided the ability to analyze thousands of both high and low quality spectra and accurately determine peptide sequences either purely in a de novo manner or with the aid of database searches.

Utilization of this particle-based CAMP discovery process with only 100 μL of American alligator plasma resulted in the capture and de novo and de novo-assisted sequencing of 340 unique peptides. A combination of web-based predictor algorithms and rational analysis resulted in the identification of 44 potential CAMPs, eight of which were synthesized and their antimicrobial effectiveness evaluated. A total of five peptides ($APOC1_{64-88}$, $APOC1_{67-88}$, $A1P_{394-428}$, $FGG_{398-413}$, and $FGG_{401-413}$) exhibited antimicrobial activity against one or more of the Gram-positive

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

or Gram-negative bacteria tested.[14] The ability to rapidly and precisely sequence naturally occurring native peptides allows the identification of numerous new potential CAMP species and potentially greatly increases the repertoire of known CAMPs.

Improving CAMP harvesting selectivity with the development of new microparticle chemistries and designs as well as modifying mass spectrometry parameters (as discussed previously) will allow more efficient capture and identification of CAMP candidates. We believe that the sequence and performance data from newly discovered CAMPS and CAMP-like peptides can be used to improve the versatility and reliability of current CAMP-prediction algorithms as well as to help develop new ones. Furthermore, because of the relatively small sample volume required, as well as the sample agnostic nature of the CAMP discovery process, it is possible to apply this to a wide variety of species that were thought to be inaccessible such as smaller-bodied or endangered species. This will help drastically expand the list of currently known CAMPs and possibly reveal novel strategies for battling antibiotic resistance with the discovery of new antimicrobial peptides.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00447.

De novo sequences determined by PEAKS software; EST database identifications determined by PEAKS software; transriptome database identifications determined by PEAKS software; CAMP Prediction for all de novo sequenced peptides; CAMP prediction for all EST identified peptides; CAMP prediction for all transcriptome identified peptides; physico-chemical properties of all de novo sequenced peptides; physico-chemical properties for American alligator EST database peptides; physico-chemical properties for American alligator transcriptome database peptides (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bbishop1@gmu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to the Defense Threat Reduction Agency for funding: HDTRA1-12-C-0039. We are grateful to Shaylyn Scott for technical assistance in preparing the Supporting Information. We gratefully acknowledge the International Crocodilian Genome Working Group for providing the Alligator mississippiensis transcriptome (www.crocgenomes.org), and the St. Augustine Alligator Farm Zoological Park and its staff for providing alligator blood samples. We would also like to thank Dr. Jennifer Van Eyk and Irina Chernysheva at Johns Hopkins University for use of their Thermo Scientific Orbitrap Elite and helpful discussions.

■ REFERENCES

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

(2) Hof, W. V. T.; Veerman, E. C.; Helmerhorst, E. J.; Amerongen, A. V. Antimicrobial peptides: properties and applicability. *Biol. Chem.* **2001**, *382*, 597−619.

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

(4) Yeaman, M.; Yount, N. Mechanisms of Antimicrobial Peptide Action and Resistance. *Pharmacol. Rev.* **2003**, *55*, 27−55.

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

(6) 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*, 308−316.

(7) Stegemann, C.; Kolobov, A.; 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*, 1364−1373.

(8) Mor, A.; Van Huong, N.; 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*, 8824−8830.

(9) 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*, 7806−7810.

(10) 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*, 5030−5039.

(11) 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*, 1811−1822.

(12) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528−9533.

(13) Udeshi, N. D.; Compton, P. D.; Shabanowitz, J.; Hunt, D. F.; Rose, K. L. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. *Nat. Protoc.* **2008**, *3*, 1709−1717.

(14) 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.

(15) 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*, M111.010587.

(16) 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.

(17) Rusmili, M. R. A.; Yee, T. T.; Mustafa, M. R.; Hodgson, W. C.; Othman, I. Proteomic characterization and comparison of Malaysian Bungarus candidus and Bungarus fasciatus venoms. *J. Proteomics* **2014**, *110C*, 129−144.

(18) St John, J. a; Braun, E. L.; Isberg, S. R.; Miles, L. G.; Chong, A. Y.; Gongora, J.; Dalzell, P.; Moran, C.; Bed'hom, B.; Abzhanov, A.; et al. Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biol.* **2012**, *13*, 415.

(19) Lata, S.; Sharma, B.; Raghava, G. Analysis and prediction of antibacterial peptides. *BMC Bioinf.* **2007**, *8*, 263.

(20) 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*, D774−80.

(21) Wang, G.; Li, X.; Wang, Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.* **2009**, *37*, D933−D937.

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295

(22) Shiloh, M. U.; Ruan, J.; Nathan, C. Evaluation of bacterial survival and phagocyte function with a fluorescence-based microplate assay. *Infect. Immun.* **1997**, *65*, 3193−3198.
(23) 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*, 114.
(24) Bagella, L.; Scocchi, M.; Zanetti, M. cDNA sequences of three sheep myeloid cathelicidins. *FEBS Lett.* **1995**, *376*, 225−228.
(25) Skerlavaj, B.; Benincasa, M.; Risso, A.; Zanetti, M.; Gennaro, R. SMAP-29: a potent antibacterial and antifungal peptide from sheep leukocytes. *FEBS Lett.* **1999**, *463*, 58−62.
(26) Park, C. B.; Kim, M. S.; Kim, S. C. A novel antimicrobial peptide from Bufo bufogargarizans. *Biochem. Biophys. Res. Commun.* **1996**, *218*, 408−413.
(27) Giacometti, A.; Cirioni, O.; Del Prete, M. S.; Barchiesi, F.; Fortuna, M.; Drenaggi, D.; Scalise, G. In vitro activities of membrane-active peptides alone and in combination with clinically used antimicrobial agents against Stenotrophomonas maltophilia. *Antimicrob. Agents Chemother.* **2000**, *44*, 1716−1719.
(28) Selsted, M. E.; Novotny, M. J.; Morris, W. L.; Tang, Y. Q.; Smith, W.; Cullor, J. S. Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. *J. Biol. Chem.* **1992**, *267* (7), 4292−5.
(29) Giacometti, A.; Cirioni, O.; Barchiesi, F.; Del Prete, M. S.; Scalise, G. Antimicrobial activity of polycationic peptides. *Peptides* **1999**, *20*, 1265−1273.
(30) Frese, C. K.; Altelaar, A. F.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668−73.
(31) Mommen, G. P.; Frese, C. K.; Meiring, H. D.; van Gaans-van den Brink, J.; de Jong, A. P.; van Els, C. A.; Heck, A. J. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (12), 4507−12.
(32) Cowland, J. B.; Johnsen, A. H.; Borregaard, N. hCAP-18, a cathelin/pro-bactenecin-like protein of human neutrophil specific granules. *FEBS Lett.* **1995**, *368*, 173−176.
(33) Larrick, J. W.; Hirata, M.; Balint, R. F.; Lee, J.; Zhong, J.; Wright, S. C. Human CAP18: a novel antimicrobial lipopolysaccharide-binding protein. *Infect. Immun.* **1995**, *63*, 1291−1297.
(34) Agerberth, B.; Gunne, H.; Odeberg, J.; Kogner, P.; Boman, H. G.; Gudmundsson, G. H. FALL-39, a putative human peptide antibiotic, is cysteine-free and expressed in bone marrow and testis. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 195−199.
(35) Turner, J.; Cho, Y.; Dinh, N. N.; Waring, A. J.; Lehrer, R. I. Activities of LL-37, a cathelin-associated antimicrobial peptide of human neutrophils. *Antimicrob. Agents Chemother.* **1998**, *42*, 2206−2214.
(36) 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*, 128.
(37) Pompilio, A.; Scocchi, M.; Pomponio, S.; Guida, F.; Di Primio, A.; Fiscarelli, E.; Gennaro, R.; Di Bonaventura, G. Antibacterial and anti-biofilm effects of cathelicidin peptides against pathogens isolated from cystic fibrosis patients. *Peptides* **2011**, *32*, 1807−1814.
(38) Blower, R. J.; Barksdale, S. M.; van Hoek, M. L. Snake Cathelicidin NA-CATH and Smaller Helical Antimicrobial Peptides Are Effective against *Burkholderia thailandensis*. *PLoS Neglected Trop. Dis.* **2015**, *9*, e0003862.
(39) Cedano, J.; Aloy, P.; Perez-Pons, J.; Querol, E. Relation between amino acid composition and cellular location of proteins. *J. Mol. Biol.* **1997**, *266*, 594−600.
(40) Churg, A.; Dai, J.; Zay, K.; Karsan, A.; Hendricks, R.; Yee, C.; Martin, R.; MacKenzie, R.; Xie, C.; Zhang, L.; et al. Alpha-1-antitrypsin and a broad spectrum metalloprotease inhibitor, RS113456, have similar acute anti-inflammatory effects. *Lab. Invest.* **2001**, *81*, 1119−1131.
(41) Påhlman, L. I.; Mörgelin, M.; Kasetty, G.; Olin, a I.; Schmidtchen, a; Herwald, H. Antimicrobial activity of fibrinogen and fibrinogen-derived peptides–a novel link between coagulation and innate immunity. *Thromb. Haemostasis* **2013**, *109*, 930−939.

DOI: 10.1021/acs.jproteome.5b00447
J. Proteome Res. 2015, 14, 4282−4295