AWARD NUMBER: W81XWH-18-1-0801

TITLE: Bioprospecting for Antiviral Peptides

PRINCIPAL INVESTIGATOR: Barney Bishop

CONTRACTING ORGANIZATION: George Mason University, Fairfax, VA

REPORT DATE: October 2020

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2020	Annual	30Sep2019-29Sep2020

4. TITLE AND SUBTITLE
Bioprospecting for Antiviral Peptides

5a. CONTRACT NUMBER
W81XWH-18-1-0801

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Barney Bishop
Kylene Kehn-Hall

E-Mail: bbishop1@gmu.edu and kkehnhall@vt.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

GEORGE MASON UNIVERSITY
Karen Cohn
4400 University Dr.
Fairfax, VA 22030-4422

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Mosquito-borne alphaviruses (e.g. chikungunya virus and Venezuelan equine encephalitis virus-VEEV) are of great concern due to their potential to cause severe acute human disease. The absence of available vaccines or therapeutic agents for almost all arboviruses makes development of effective treatment options a significant national security and public health issue. Here, we focus on the study of VEEV as a model arbovirus. Cationic antimicrobial peptides (CAMPs) of innate immunity represent a promising resource for the development of new antiviral therapeutics. However, there is need for a system specifically designed for purpose of discovering AMPs with antiviral activity. We aim to develop a bioprospecting-inspired process for the identification of antiviral peptides against VEEV, using an integrated workflow that combines novel antiviral peptide harvesting microparticles, advanced mass spectrometry, and data analysis tools in order to establish peptide sequences and identify those with potential antiviral properties. We will focus on alligators and snakes CAMPs, since both have been suggested as potential arbovirus reservoir species.

15. SUBJECT TERMS
Reptile, Venezuelan equine encephalitis virus, innate immunity, cytokine, chemokines, host defense peptides, bioprospecting, antiviral, and antiviral peptide discovery.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	33	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. INTRODUCTION:

The rapid, worldwide spread of arthropod-borne viruses (arboviruses) that were previously limited to remote tropical and sub-tropical regions, including mosquito-borne alphaviruses (e.g. chikungunya virus and Venezuelan equine encephalitis virus-VEEV), is of great concern due to their potential to cause severe acute human disease. The absence of FDA approved vaccines or therapeutic agents for almost all arboviruses makes development of effective treatment options for this class of pathogens a significant national security and public health issue. Here we focus on the study of VEEV as a model arbovirus. Cationic antimicrobial peptides (CAMPs) of innate immunity represent a promising resource for the development of new antiviral therapeutics. There is a need to develop a system for the specific purpose of discovering peptides with antiviral activity. We have aimed to develop a bioprospecting-inspired process for the identification of antiviral peptides against VEEV, using novel antiviral peptide harvesting microparticles. These particles incorporate elements from healthy host cells, infected host cells and virions in order to enable the identification of peptides that preferentially target infected cells and virus over healthy host cells. We have focused our efforts on identifying antiviral CAMPs from alligators and snakes, since these species have been suggested as potential arbovirus reservoir species. Our process employs an integrated workflow that begins with the preferential enrichment of CAMPs that target viral elements from reptile serum or plasma, followed by advanced mass spectrometry and data analysis in order to establish their sequences and identify those peptides with potential antiviral properties.

2. KEYWORDS: *Provide a brief list of keywords (limit to 20 words).*

Reptile, Venezuelan equine encephalitis virus, innate immunity, cytokine, chemokines, host defense peptides, bioprospecting, machine learning, antiviral, and antiviral peptide discovery.

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Specific Aim 1: Develop microparticles for capturing antiviral peptides.

Major Task 1: Develop VLP-based capture particles (12 months: 09/30/18-09/29/19)

Milestones:

1.) Reproducible protocol for synthesizing and purifying PP2MA or similar click-chemistry-compatible monomer. (delivery date: 03/29/19 – completed)

2.) Preparations of VEEV virus like particles. (delivery date: 03/29/19 – completed)

3.) Protocol for preparing VLP-modified hydrogel particles using click chemistry and preliminary harvesting data for the particles. (delivery date: 09/29/19 – 90% complete)

4.) GMU IACUC Approval (delivery date: 03/29/19 – completed)

5.) ACURO Approval (delivery date: 05/29/19 – completed)

Major Task 2: Develop cell-membrane based capture particles (9 months: 09/30/18-06/29/19)

Milestones:

1.) Protocols for preparing and assessing hydrogel particles encapsulated in membranes from healthy host cells, with preliminary harvesting data. (delivery date: 06/29/19 – completed)

2.) Protocols for preparing and assessing hydrogel particles encapsulated in membranes from infected host cells, with preliminary harvest data. (delivery date: 06/29/19 – completed)

Specific Aim 2: Capture, analysis and identification of peptides from reptile plasma.

Major Task 3: Perform plasma harvests (12 months: 03/30/19-03/29/20)

Milestones:

1.) VEEV challenge conditions for stimulating release/production of peptides by reptile blood cells. (delivery date: 09/29/19 - completed)

2.) Harvest protocol and conditions that afford efficient capture and identification of known model antiviral CAMPs from plasma. (delivery date: 09/29/19 – completed)

3.) Harvests performed from stimulated and unstimulated plasma using panel of particles developed in Aim 1 for the purpose of identifying antiviral peptides. (delivery date: 03/29/20 – completed)

Major Task 4: Analyze harvested peptides (12 months: 03/30/19-03/29/20)

Milestones:

1.) Workflow for efficiently analyzing harvest samples via tandem mass spectrometry and determination of the peptide sequences. (delivery date: 09/29/19 – completed)

2.) Mass spectrometry data and sequences from peptides harvested for the purpose of identifying novel antiviral peptides. (delivery date: 03/29/20 – 95% complete)

3.) Statistical methods for analyzing harvested peptide sequences to afford statistical significance values of potential antiviral peptides. (delivery date: 03/29/20 – completed)

4.) Predictive model(s) based on machine learning for analyzing harvested peptides to aid in identification of peptides likely to have antiviral properties. (delivery date: 03/29/20 – completed)

Major Task 5: Assess performance of likely antiviral peptides (6 months: 09/30/19-03/29/20)

Milestones:

1.) One or more novel alligator or snake peptides that exhibit antiviral properties against VEEV. (delivery date: 03/29/20 – 90% complete).

What was accomplished under these goals?

Specific Aim 1: Develop microparticles for capturing antiviral peptides.

Major Task 1: Develop VLP-based capture particles (12 months: 09/30/18-09/29/19)

1.) Preparation of particles containing propargyl-PEG-monomer (PP4ME)

The first generation of VLP-based capture particles utilized core-shell hydrogel particles that combined an inert core consisting of cross-linked tert-butylmethacrylamide (TBMAm) and shells based on cross-linked copolymers of *N*-methacryloyl-6-aminohexanoic acid (MA6AHA) and PP4ME. Descriptions and analyses of these hydrogel particles were provided in the previous annual report.

A second generation of VLP-based particles were prepared based on core-shell hydrogel particles that combined inert cross-linked tert-butylmethacrylamide (TBMAm) cores and shells consisting of cross-linked copolymers combining hydroxypropylmethacrylamide (HPMA) with 10-20% PP4ME (with some also incorporating low amounts of MA6AHA in the shell formulation). These hydrogel particles were developed in order to reduce background peptide capture associated with the hydrogel particle component of the VLP-modified particles. The presence of peptides captured by the hydrogel particle of the VLP-modified particles could complicate the detection of peptides captured specifically by the VLP baits. Literature reports describe the use of HPMA in anti-fouling coatings (Langmuir, 2010

Nov 16;26(22):17375-82.). Furthermore, HPMA is a neutral polar monomer in contrast to MA6AHA, which is anionic, and thus HPMA-based hydrogel particles should exhibit different harvesting properties from the general harvest particles.

Particle		N4.3	N4.4	N4.5	N4.6	N4.9	N4.7	N4.8	N4.10
Monomers	**PP4ME**	10.0%	9.0%	10.0%	10.0%	10.0%	X	10.0%	25.0%
	MA6AHA	70.0%	81.0%	85.0%	90.0%	70.0%	X	X	10.0%
	BIS	20.0%	10.0%	5.0%	0.0%	20.0%	10.0%	10.0%	5.0%
	HPMA	X	X	X	X	X	90.0%	80.0%	60.0%

Table 1. Series of core-shell particles generated for development of virus-modified particles.

2.) Virus-modified particles were generated from 1st and 2nd (N4.9 and N4.8 particles respectively) generation core-shell hydrogel particles and sucrose-cushion purified VEEV TC83 via click-chemistry using the method described in the previous annual report.

3.) General harvests from healthy alligator plasma were performed in preparation for harvests using cell-membrane based capture particles in Major Task 2, as described in the previous annual report. However, here the eluents from the general harvests were used for harvests with virus-modified particles. Harvests were performed using both the virus modified and unmodified N4.9 and N4.8 particles, with harvests using each particle type being done in duplicate. Harvests and post-harvest processing were performed using protocols developed using the cell-membrane based capture particles. Similarly, the captured peptides were analyzed via mass spectrometry using the Thermo Fusion instrument at GMU using our standard parameters for peptidomic analyses, and resulting mass spectrometry data (focusing on MS/MS data) were processed using PEAKS de novo sequencing software and our Python script in order to determine and organize the sequences of peptides captured in each harvest. The sequence and additional mass spectrometry data were transferred to the team at PNNL for statistical analyses.

4.) We prepared a new generation of core-shell particles as the framework for constructing virus-modified particles. The formulations for these particles were designed to maximize virus loading. Both of these particles are built upon the same cross-linked N-*tert*-butylmethacrylamide polymer core. They differ in the architecture of the outer harvesting layer(s). The outer harvesting layers of both consist of cross-linked HPMA-based copolymers that include MA6AHA and PP4ME in their formulations. Where they differ is the distribution of the co-monomers. In one case, N5.1 particles, the outer harvesting layer consists of a random co-mixture of the different monomers. The N5.2 particles have a more complex architecture where PP4ME monomer is added later in the synthesis of the harvesting layer, resulting in a higher concentration of the monomer near the outer surface of the particle. This arrangement should make the PP4ME more accessible for cross-linking with virus. These particles were transferred to the Kehn-Hall lab for coupling to virus. These studies are underway and nearing completion.

Major Task 2: Develop cell-membrane based capture particles (9 months: 09/30/18-06/29/19)
Work on Major Task 2 deliverables was completed in Q3 of the project.

Specific Aim 2: Capture, analysis and identification of peptides from reptile plasma.

Major Task 3: Perform plasma harvests (12 months: 03/30/19-03/29/20)
Multiple harvests had been performed and were at various stages of analysis at the time that GMU shut down non COVID-related research in spring 2020. Upon our return to the lab we completed

analysis of the data that we had collected. We have completed our analyses of alligator plasma and PBMC-secreted peptides. These studies, were primarily intended to investigate how different factors impacted peptide discovery. We are presently wrapping up analyses of snake (*Boa constrictor*) plasma peptides. Information regarding these harvests and their advancement to date are indicated in Table 2 below.

Exp. ID #	Species	Sample	Harvests	Post-Harvest Processing	MS/MS Analysis	Data Analysis	Purpose/Results of study
8	*A. mississippiensis*	Media Supernatant – Control					Purpose: Identify potential alligator AMPs. Outcome: Afforded varied results. Due to limited access to the mass spectrometer, these studies were not repeated.
		Media Supernatant – 2h post VEEV					
		Media Supernatant – 8h post VEEV					
9	*A. mississippiensis*	Unstimulated Plasma - EDTA					Purpose: Evaluate impact of anticoagulant. Outcome: Afforded few peptide identifications complicating analysis. Study was not central to project – opted to focus on other harvests.
		Unstimulated Plasma - Heparin					
10[1]	*A. mississippiensis*	Unstimulated Plasma –Crazy					Purpose: Analyze alligator peptides from individual with inflammation. Outcome: Poor peptide identification – Repeated.
11	*A. mississippiensis*	Media Supernatant – 8h post VEEV					Purpose: assess parameters and their impact on peptide harvesting. Outcome: data is being analyzed.
12	*B. constrictor*	Unstimulated – healthy membranes					Purpose: Identify potential snake AMPs. Outcome: Afforded few high confidence peptide identifications –repeated.
		Unstimulated – infected membranes					
13[1]	*A. mississippiensis*	Unstimulated Plasma –Crazy					Purpose: Analyze alligator peptides from individual with inflammation. Outcome: Analyses are nearing completion.
14[2]	*B. constrictor*	Unstimulated – healthy membranes					Purpose: Identify potential snake AMPs and assess direct harvest and MS/MS parameters. Outcome: Analyses are nearing completion.
		Unstimulated – infected membranes					

Table 2. Harvests and analyses from third and fourth quarters this year, and their current status/progress. Initial analyses of mass spectrometry data have been performed. Blue bars indicate that harvests, post-harvest processing and analyses have been completed or are nearly complete. Purple bars indicate that harvests and post-harvest processing have been completed, and analyses of data are underway. The nature of these studies does not require their analysis by PNNL.

1. Blood was collected from a large adult alligator named Crazy that had a large growth on its leg that was being removed by the veterinarian. Studies explore how altering MS/MS parameters on the mass spectrometer influences the number and size of peptides for whom sequences are determined. These studies are ongoing.
2. Experiment designed to explore direct harvesting from snake plasma using membrane-encapsulated particles and to assess the effect of MS/MS fragmentation parameters on peptide identification. These studies are still in process.

Results from analyses of first round Boa constrictor harvests: Due to the low volume of plasma provided (~800 µL), harvests were performed at half normal scale (50µL of plasma) in order to allow experimental replicates for each harvest. Two rounds of harvest were performed. In the second round, harvests were performed directly from plasma in buffer using particles encapsulated in membranes from healthy and VEEV-infected cell membranes. Peptides captured in the first round of harvests were analyzed via LC-MS/MS immediately before they were shut down in compliance with GMU precautionary measures in response to COVID-19. The second round of snake harvests were analyzed by LC-MS/MS in September of this year, after the instrument was brought back online and the proteomic MS lab returned to operations. The sequences of the peptides from the first round of harvests were determined using PEAKS *de novo* sequencing software version X+ using a reference a database consisting of predicted protein sequences from a *B. constrictor* genome assembly and

annotation (Card et al., 2019). An additional homology match search was performed using PEAKS to identify peptides similar to database peptides that may have been initially missed in the more stringent database search stage. The results of these initial analyses of Boa constrictor peptides are given below in Table 3.

Particle Type	Rep. #	# Features	# DB Search PSMs	# DB Search Unique Peptides > 7 aa	# *de novo* Spectra	# *de novo* Unique Peptides > 7 aa
MA6AHA	1	4994	5	3	213	25
Healthy	1	25335			105	33
Healthy	2	4581	3	1	48	11
Healthy	3	4155			44	14
Infected	1	12049	7	3	493	126
Infected	2	13753			400	81
Infected	3	10435			480	114

Table 3. Peptide identification summary for all samples and replicates: Particle type indicates the particle type used to perform the harvest: "MA6AHA" = general harvest particle, "Healthy" = particles encapsulated in membranes form healthy leukocytes, and "Infected" refers to particles prepared using membrane from VEEV-infected leukocytes. The low number of peptides with database matches (# DB Search PSMs) could reflect gaps in our sequence database or poor fragmentation on the mass spectrometer.

In addition to analyzing the MS/MS spectra and peptide sequences, we have analyzed the mass spectra and the features that were present, focusing on features we predict correspond to peptides. The results of these feature analyses are given in Figure 1. We have determined that we do not have high quality MS/MS spectra for the many of these features. We are investigating modifications to the MS/MS parameters in order to capture better data for large highly charged features/peptides.

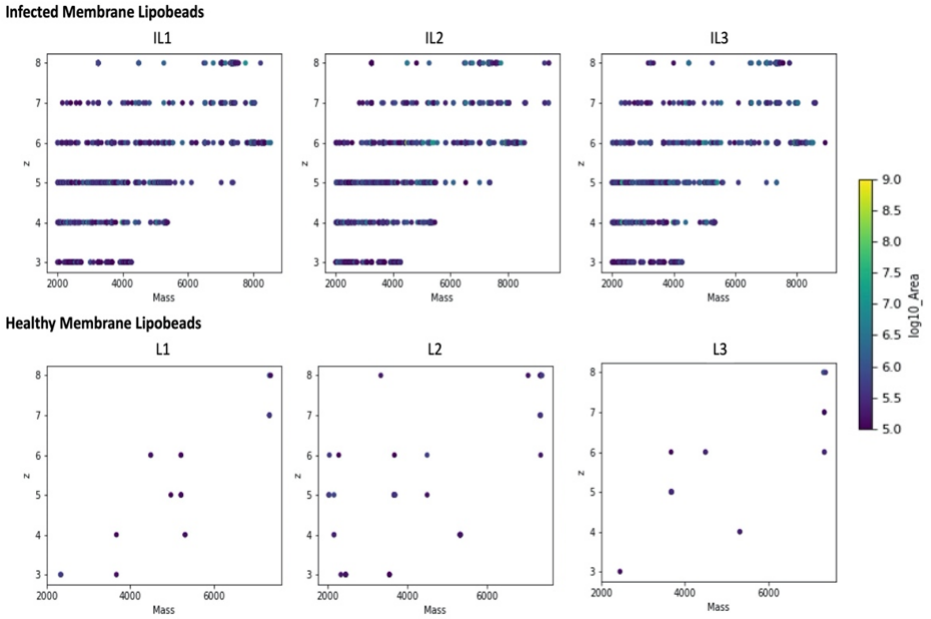

Figure 1. Peptide feature analyses of mass spectrometry data from snake samples: These analyses have focused on features with peak areas ≥ 1E5, and corresponding to peptides with masses of ≥ 2000 Da and charges of +3 - +8. **IL1**, **IL2** and **IL3** present the features identified from harvests performed using membranes from infected leukocytes**. L1**, **L2** and **L3** present the features identified from harvests performed using membranes from healthy leukocytes.

These are encouraging results for a first harvest of *Boa constrictor* host defense peptides; however further analysis of the first-round data is complicated by the limited number of confirmed peptide identifications from the database. Hence, we have focused more on the second round of harvests, and these analyses are currently underway.

Major Task 4: Analyze harvested peptides (12 months: 03/30/19-03/29/20)

1. Statistical Analyses of Harvests:

Supernatant from media with VEEV-infected alligator PBMCs (8 hours time point)

In this experiment, 20 samples were analyzed, 13 stimulated with VEEV and 7 that had not been - Table 4, with the various particles and membranes as shown. The short descriptor names for each harvest are given in the first column. Analysis of mass spectrometry data resulted in peptide sequences that were sorted into two groups based on whether they could be matched to sequences in a reference database (database) or whether were strictly determined de novo from the mass spectra and did not have a match in the reference database (de novo). These harvests yielded 83 database peptides, as well as 1903 de novo peptides. The total number of peptides per sample is shown in Figure 2. We first compared the peptides identified from the general harvest to harvests using membrane-encapsulated particles. No overlaps were observed in the database peptides identified from the general harvest (GH) and harvests using membrane-encapsulated particles (Membrane). While for de novo peptides, one peptide was identified from both GH and Membrane harvests, Figure 3.

Name	Sample Description	Particle	Membrane
8H-GH	Supernatant from media with VEEV-infected gator	MA6AHA	N/A
8H-o-H1	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Healthy
8H-o-H2	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Healthy
8H-o-H3	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Healthy
8H-o-I1	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Infected
8H-o-I2	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Infected
8H-o-I3	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Infected
8H-t-H1	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Trypsinized Healthy
8H-t-H2	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Trypsinized Healthy
8H-t-H3	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Trypsinized Healthy
8H-t-I1	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Trypsinized Infected
8H-t-I2	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Trypsinized Infected
8H-t-I3	Supernatant from media with VEEV-infected gator	NIPMAM Hydro Mod	Trypsinized Infected
MK-GH	Supernatant from media	MA6AHA	N/A
MK-o-H1	Supernatant from media	NIPMAM Hydro Mod	Healthy
MK -o-H2	Supernatant from media	NIPMAM Hydro Mod	Healthy
MK -o-H3	Supernatant from media	NIPMAM Hydro Mod	Healthy
MK -o-I1	Supernatant from media	NIPMAM Hydro Mod	Infected
MK -o-I2	Supernatant from media	NIPMAM Hydro Mod	Infected
MK -o-I3	Supernatant from media	NIPMAM Hydro Mod	Infected

Table 4. Description of experimental samples

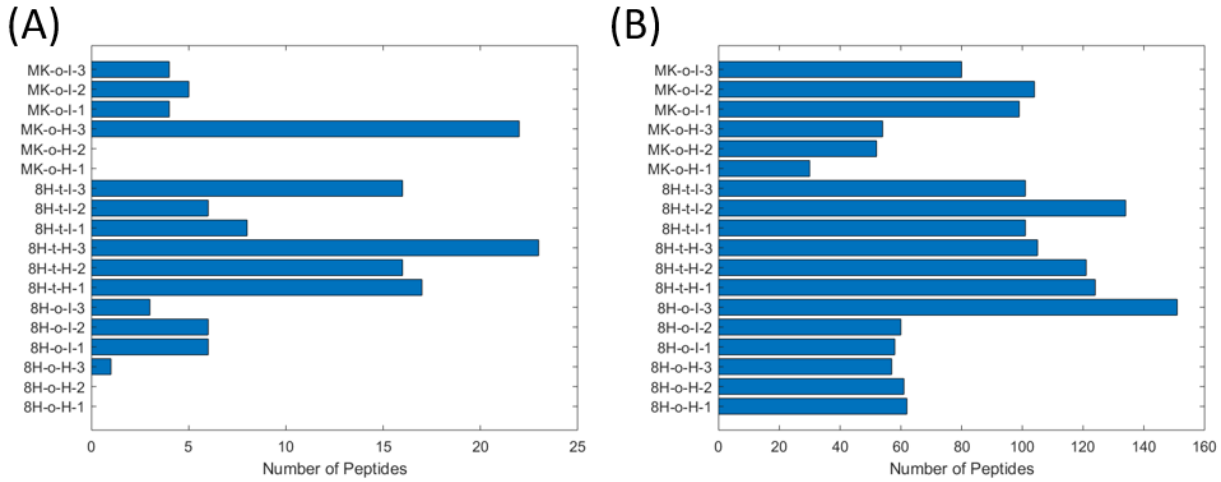

Figure 2. Total number of peptides per sample for the (A) database and (B) de novo datasets.

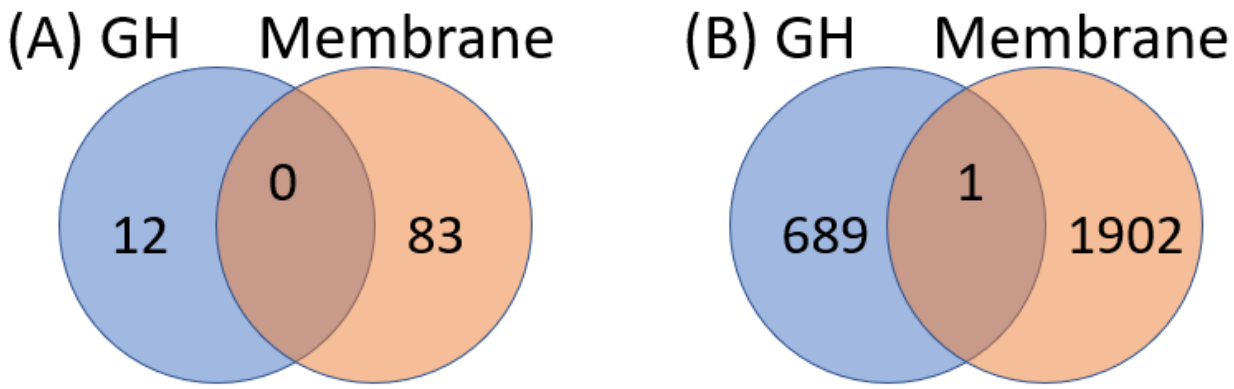

Figure 3. Overlap between peptide sequences from general and membrane harvests: (A) database and (B) de novo sequences.

We evaluated the similarity between each sample using the Jaccard Index (1), which is the number of peptides that two samples have in common scaled to the total number of peptides in the two samples. In Figure 4, the JA is shown in the bottom left triangle and the specific counts are shown in the remainder of the figure where the diagonal is the number of peptides per sample, also seen in Figure 2. Two samples with perfect identity will have a JA of 1. Figure 5 further demonstrates how the samples cluster based on the JA. In particular, they cluster on healthy versus infected, as well as a separation based on whether cell membrane proteins had been trypsinized prior to their use in generating the membrane-encapsulated particles (these particles have been referred to as "trypsinized", and this term is also used to denote harvests using these particles).

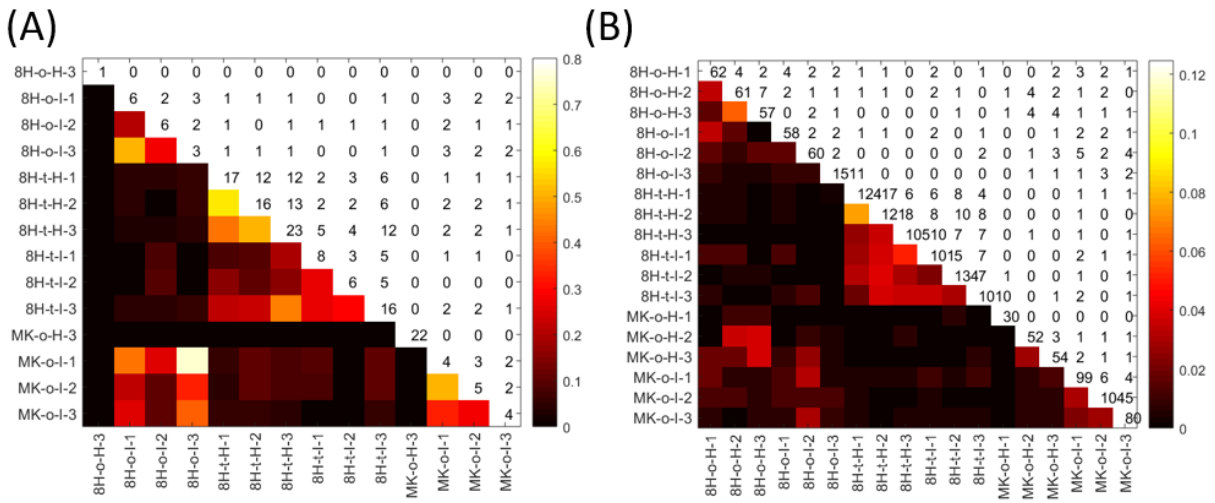

Figure 4. Jaccard Index for the (A) database and (B) de novo datasets with the peptide counts per sample given on the diagonal and the overlap in the upper triangle cells.

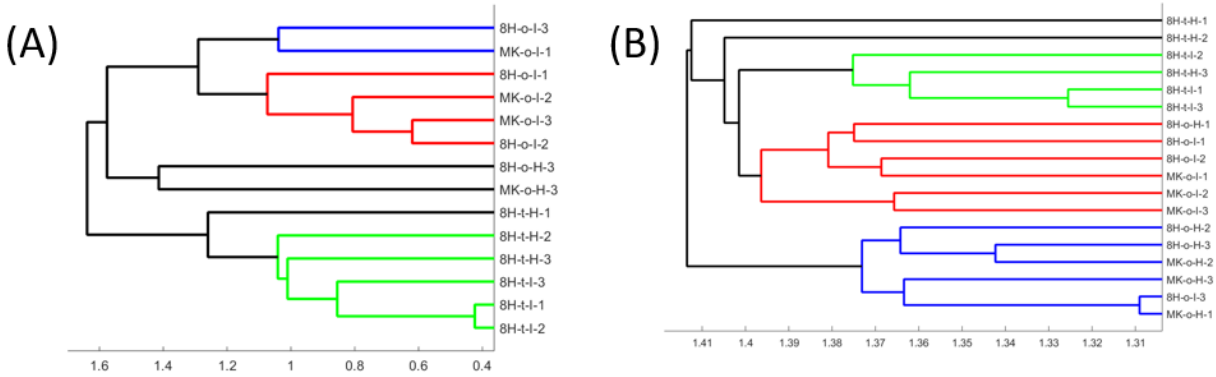

Figure 5. Hierarchal clustering based on Jaccard Index similarity: (A) database and (B) de novo datasets.

For statistical analyses, we used a Fisher's Exact test due to the small sample size. Similarly, significant p-values via the Fisher's Exact test are difficult to achieve due to the small sample sizes, however we can use them to sort the results. There were 10 and 7 peptides with p-values that can be utilized for prioritization from the database and de novo searches, respectively. ST and US are used represent harvests from media collected from stimulated and unstimulated alligator PBMCs, respectively, and results are in Figures 6 and 7.

	FISHER EXACT TEST			COUNT				
	HLvsIN	ST(HLvsIN	US(HLvsIN	US-HL	ST-HL	US-IN	ST-IN	
FPSIVGRPR	0.027473	1	0.033333	0	3	0	0	
KVVNPLFEKRPK	0.027473	1	0.033333	0	3	0	0	
NKRSTITSREIQTAVR	0.027473	1	0.033333	0	3	0	0	
RDNIQGITKPAIRRL	0.027473	1	0.033333	0	3	0	0	
SSRAGLQFPVGRVH	0.027473	1	0.033333	0	3	0	0	
AVGVIKAVDKKAAGAGKVTK	0.030969	0.25	0.2	0	0	3	3	> INFECTED
ALKRQGRTLYGFGG	0.094905	1	0.190476	0	3	0	1	
FRPAGAAPRPPPKP	0.094905	1	0.190476	0	3	0	1	
KIKIIAPPERK	0.094905	1	0.190476	0	3	0	1	
SRSSRAGLQFPVGRVH	0.094905	1	0.190476	0	3	0	1	

Figure 6. Top antiviral peptide candidates for peptides from the database dataset (labeled ">INFECTED").

	FISHER EXACT TEST			COUNT				
	HLvsIN	ST(HLvsIN	US(HLvsIN)	US-HL	ST-HL	US-IN	ST-IN	
KVVVSPTKKVAV	0.00905	0.1	0.1818182	0	0	3	3	> INFECTED
GPLLLNPPAP	0.082353	0.4	0.4545455	2	2	0	0	
CMSDQCCMHN	0.205882	0.4	1	0	0	2	1	
KAALKVVA	0.205882	1	0.1818182	0	3	0	0	
KALLKVVG	0.205882	1	0.1818182	0	3	0	0	
PPKNNLLP	0.205882	1	0.1818182	0	3	0	0	
APLSRSHK	0.205882	1	0.4545455	1	2	0	0	

Figure 7. Top antiviral peptide candidates for the de novo peptides (labeled ">INFECTED").

<u>Harvests from plasma isolated from healthy and VEEV-stimulated alligator blood (8 hours time point)</u>
In this experiment, 10 samples were analyzed, 5 stimulated with VEEV and 5 that had not been - Table 5, with the various particles and membranes as shown. The short descriptor names for each harvest are given in the first column. These harvests yielded identification of 137 database peptides, as well as 1840 de novo peptides. The total number of peptides per sample is shown in Figure 8. We first compared the peptides identified from the general harvest to those from harvests using membrane-encapsulated particles. For the database there were 5 peptides that overlap between GH and Membrane harvests, while the de novo data exhibited no overlap in peptide sequences between the two groups, Figure 9.

Name	Sample Description	Particle	Membrane
ST-GH	Gator plasma, incubated 8 hours, stimulated VEEV	MA6AHA	N/A
ST-o-HL	Gator plasma, incubated 8 hours, stimulated VEEV	NIPMAM Hydro Mod	Healthy
ST-o-IN	Gator plasma, incubated 8 hours, stimulated VEEV	NIPMAM Hydro Mod	Infected
ST-t-HL	Gator plasma, incubated 8 hours, stimulated VEEV	NIPMAM Hydro Mod	Trypsinized Healthy
ST-t-IN	Gator plasma, incubated 8 hours, stimulated VEEV	NIPMAM Hydro Mod	Trypsinized Infected
US-GH	Gator plasma, incubated 8 hours, no VEEV	MA6AHA	N/A
US-o-HL	Gator plasma, incubated 8 hours, no VEEV	NIPMAM Hydro Mod	Healthy
US-o-IN	Gator plasma, incubated 8 hours, no VEEV	NIPMAM Hydro Mod	Infected
US-t-HL	Gator plasma, incubated 8 hours, no VEEV	NIPMAM Hydro Mod	Trypsinized Healthy
US-t-IN	Gator plasma, incubated 8 hours, no VEEV	NIPMAM Hydro Mod	Trypsinized Infected

Table 5. Description of experimental samples

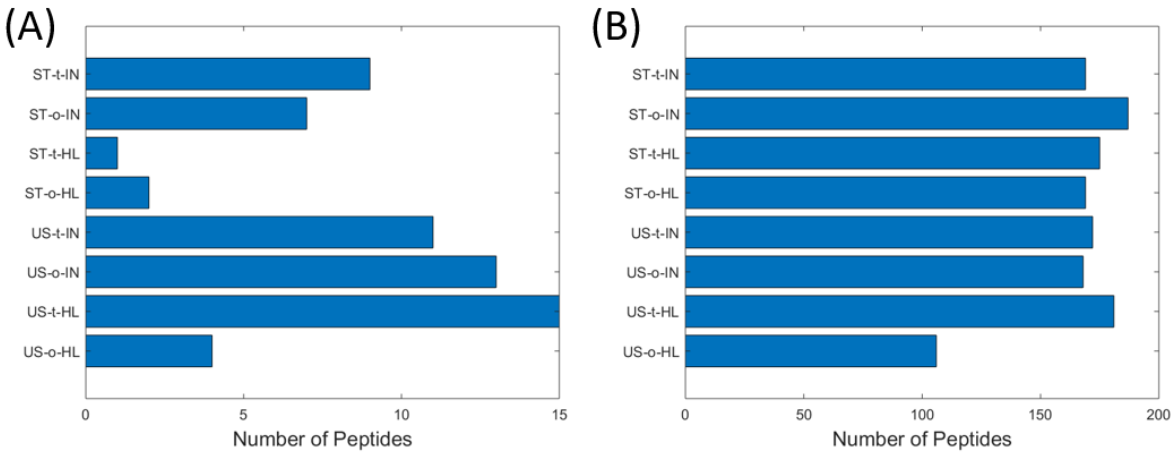

Figure 8. Total number of peptides per sample for the (A) database and (B) de novo datasets.

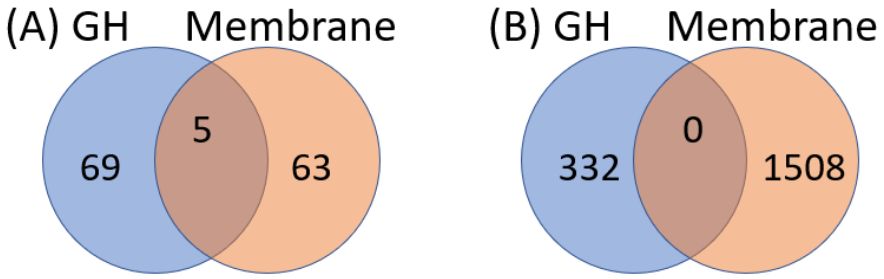

Figure 9. Overlap between peptide sequences from general and membrane harvests: (A) database and (B) de novo sequences.

We evaluated the similarity between each sample using the Jaccard Index (1), which is the number of peptides that two samples have in common scaled to the total number of peptides in the two samples. In Figure 10, the JA is shown in the bottom left triangle and the specific counts are shown in the remainder of the figure where the diagonal is the number of peptides per sample, also seen in Figure 8. Two samples with perfect identity will have a JA of 1. Figure 11 further demonstrates how the samples cluster based on the JA. In particular, they cluster on healthy versus infected, as well as a separation based on whether the cell membrane proteins had been trypsinization.

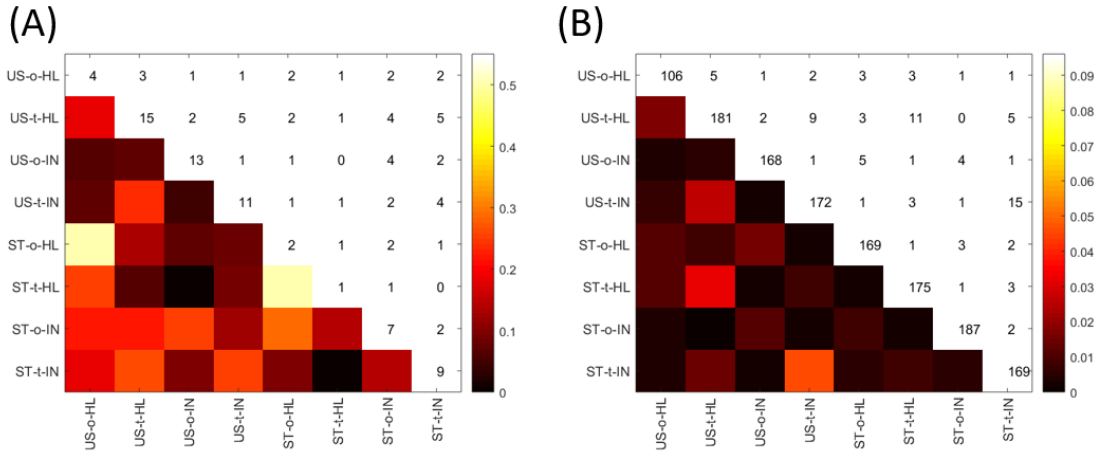

Figure 10. Jaccard Index for the (A) database and (B) de novo datasets with the peptide counts per sample given on the diagonal and the overlap in the upper triangle cells.

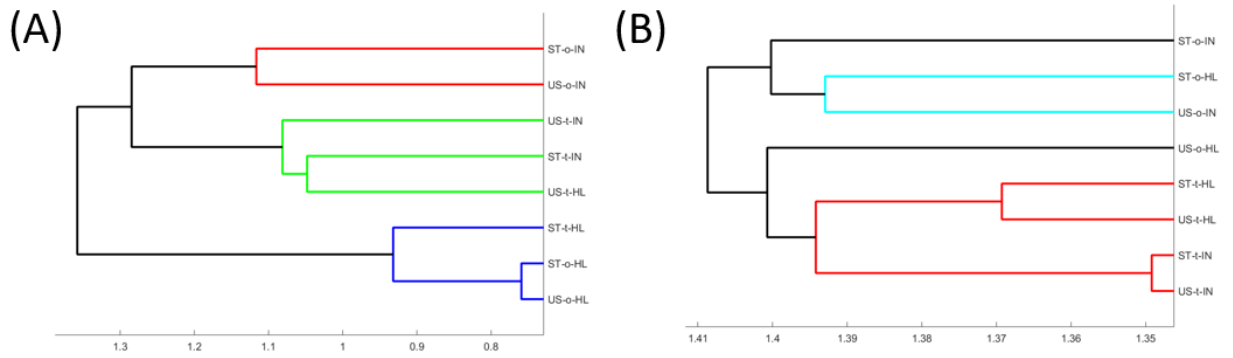

Figure 11. Hierarchal clustering based on Jaccard Index similarity: (A) database and (B) de novo datasets.

For statistical analyses, we used a Fisher's Exact test due to the small sample size. Similarly, significant p-values via the Fisher's Exact test are difficult to achieve due to the small sample sizes, however we can use them to sort the results. There were 6 and 5 peptides with p-values that can be utilized for prioritization from the database and de novo searches, respectively. ST and US are used to represented harvests from plasma derived from stimulated and unstimulated alligator blood, respectively, and results are in Figures 12 and 13.

	FISHER EXACT TEST			COUNT				
	HLvsIN	ST(HLvsIN	US(HLvsIN	US-HL	ST-HL	US-IN	ST-IN	
ALKRQGRTLYGFGG	0.142857	0.333333	1	1	0	2	2	> INFECTED
ANGTTVHVGIHPSKVVITR	0.428571	1	1	0	0	1	1	> INFECTED
ASEHLDAFQRYLEELKRTFTPS(	0.428571	1	1	2	2	1	1	
AVGVIKAVDKKAAGAGKVTK	0.428571	1	1	0	0	1	1	> INFECTED
IVDVKANKHQIKQAVKKL	0.428571	1	1	0	0	1	1	> INFECTED
KVLKQVHPDTGISSK	0.428571	1	1	0	0	1	1	> INFECTED

Figure 12. Top antiviral peptide candidates for peptides from the database search (labeled ">INFECTED").

	FISHER EXACT TEST			COUNT				
	HLvsIN	ST(HLvsIN	US(HLvsIN	US-HL	ST-HL	US-IN	ST-IN	
LGKKPAKL	0.142857	0.333333	1	0	0	1	2	> INFECTED
AKKVLAPK	0.142857	0.333333	1	1	2	0	0	
HAKSTYCKL	0.142857	1	0.333333	2	1	0	0	
LSSAAAAPLT	0.142857	1	0.333333	2	1	0	0	
LKKGACCC	0.428571	0.333333	1	0	0	0	2	> INFECTED

Figure 13. Top antiviral peptide candidates for the de novo peptides (labeled ">INFECTED").

<u>Harvests comparing general harvest with ones using VEEV-modified particles</u>

In this experiment, 9 samples were analyzed, 4 stimulated with VEEV and 5 that had not been - Table 6, with the various particle formulations and with/without covalently-bound VEEV as shown. The short descriptor names for each harvest are given in the first column. These harvests yielded 138 database peptides, as well as 1399 de novo peptides. The total number of peptides per sample is shown in Figure 14. We first compared the peptides identified from the general harvest to those from

the harvests using VEEV-modified particles (VLP). For database peptides, 26 peptides were shared between general harvest (GH) and harvests using virus-modified particles (VLP), while de novo peptides afforded 119 shared peptides, Figure 15.

Name	Sample Description	Particle	Bound Virus
HP-HL-1	Harvest with HPMA	HPMA	No
HP-HL-2	Harvest with HPMA	HPMA	No
HP-IN-1	Harvest with HPMA coupled to virus	HPMA	Yes
HP-IN-2	Harvest with HPMA coupled to virus	HPMA	Yes
N4-HL-1	Harvest with MA6aHA	N4	No
N4-HL-2	Harvest with MA6aHA	N4	No
N4-IN-1	Harvest with MA6aHA coupled to virus	N4	Yes
N4-IN-2	Harvest with MA6aHA coupled to virus	N4	Yes

Table 6. Description of experimental samples

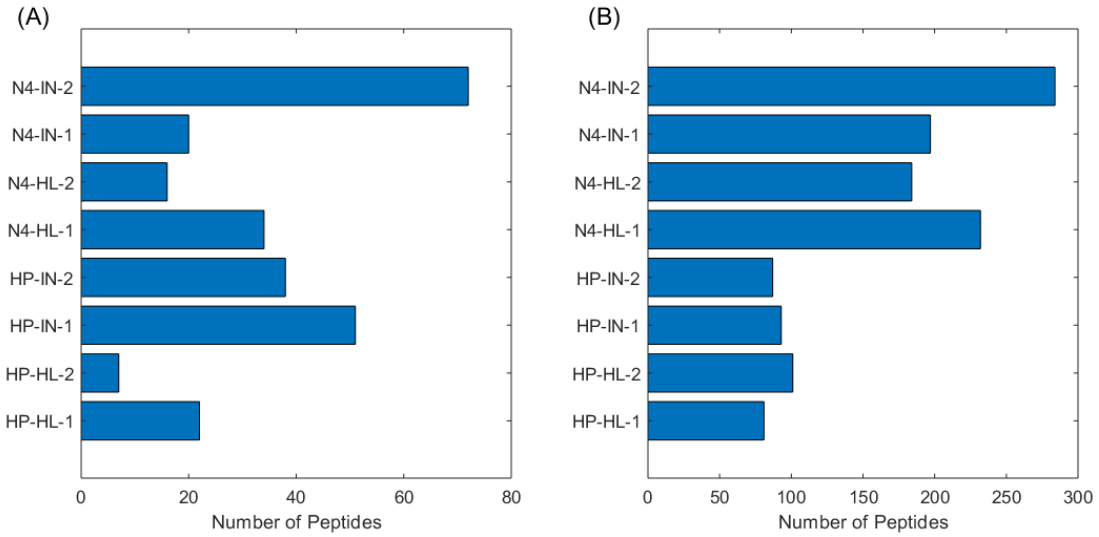

Figure 14. Total number of peptides per sample for the (A) database and (B) de novo datasets.

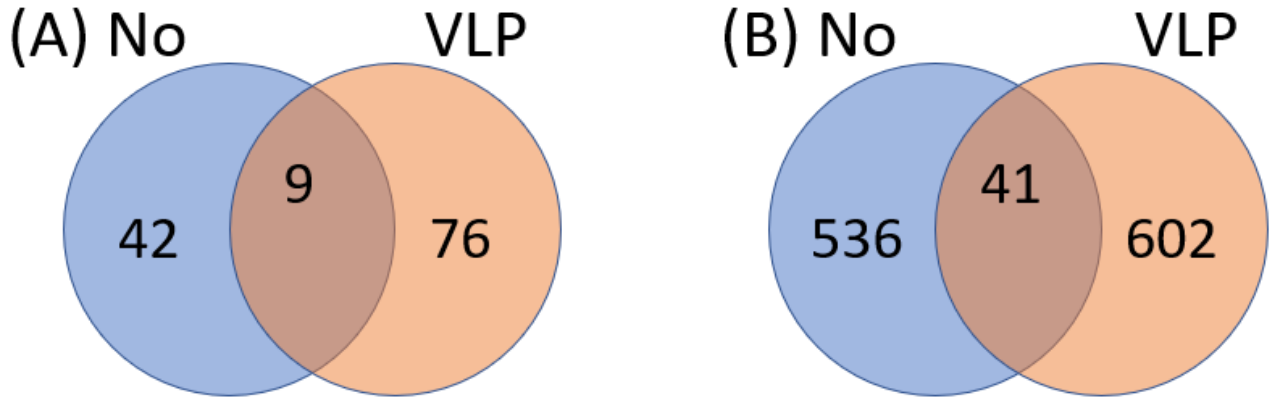

Figure 16. Overlap between peptide sequences from general and VLP harvests: (A) database and (B) de novo sequences.

We evaluated the similarity between each sample using the Jaccard Index (1), which is the number of peptides that two samples have in common scaled to the total number of peptides in the two samples. In Figure 16, the JA is shown in the bottom left triangle and the specific counts are shown in the remainder of the figure where the diagonal is the number of peptides per sample, also seen in Figure

14. Two samples with perfect identity will have a JA of 1. Figure 17 further demonstrates how the samples cluster based on the JA. In particular, they cluster on core particle type (HPMA vs N4) and unmodified versus virus-modified particles.

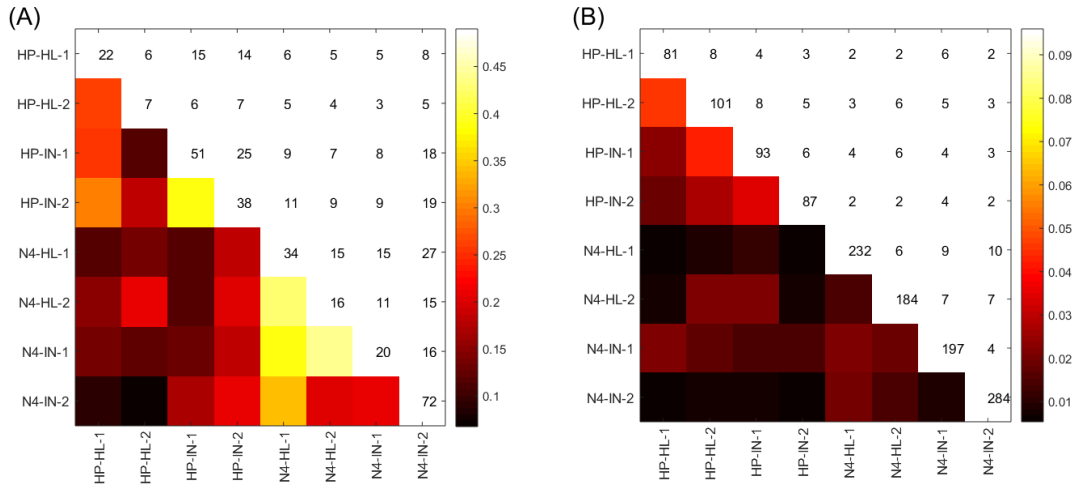

Figure 16. Jaccard Index for the (A) database and (B) de novo datasets with the peptide counts per sample given on the diagonal and the overlap in the upper triangle cells.

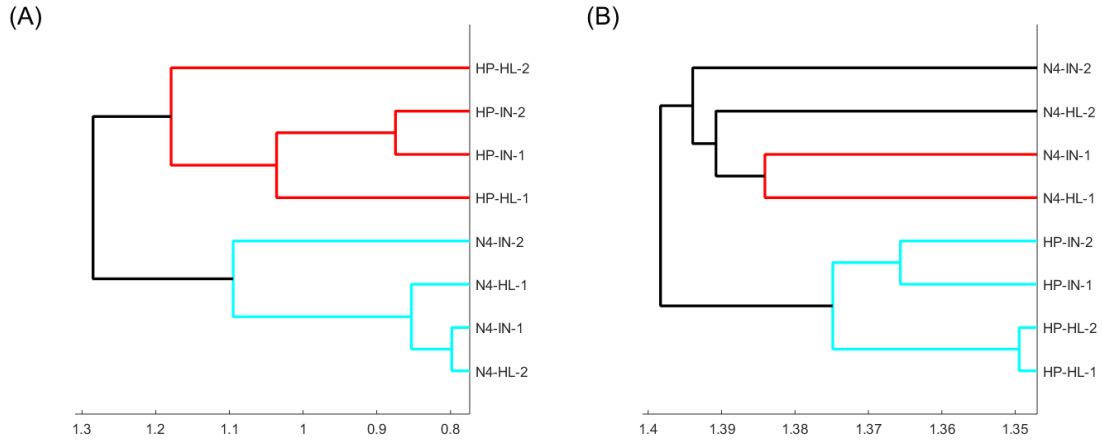

Figure 17. Hierarchal clustering based on Jaccard Index similarity: (A) database and (B) de novo sequences.

For statistical analyses, we used a Fisher's Exact test due to the small sample size. Similarly, significant p-values via the Fisher's Exact test are difficult to achieve due to the small sample sizes, however we can use them to sort the results. There were 15 and 12 peptides with p-values that can be utilized for prioritization from the database and de novo sequences, respectively; results are in Figures 18 and 19.

	FISHER EXACT TEST			COUNT				
	HLvsIN	HPMA(HLv	N4(HLvsIN	HP-HL	N4-HL	HP-IN	N4-IN	
ADGTTITSGVETTKPAK	0.142857	0.333333	1	0	1	2	2	> BOUND VIRUS
ALVLRRTNAVLFEKR	0.142857	0.333333	1	0	0	2	1	> BOUND VIRUS
LKETGFPVIENK	0.142857	0.333333	1	0	0	2	1	> BOUND VIRUS
SRGPRLISRDWR	0.142857	0.333333	1	0	0	2	1	> BOUND VIRUS
SGGKRILLDRTQNHIQEELKFP	0.142857	1	0.333333	2	2	1	0	
ASPTVHLFPPSSEEVSSK	0.428571	0.333333	1	0	0	2	0	> BOUND VIRUS
IAHGVKKYGPTVLRIIRIAG	0.428571	0.333333	1	0	0	2	0	> BOUND VIRUS
IKPMKDSTVLPHFK	0.428571	0.333333	1	0	2	2	2	> BOUND VIRUS
NFPSVLDLSK	0.428571	0.333333	1	0	0	2	0	> BOUND VIRUS
NQKPVNPKPLELEEKFDGSFTA	0.428571	0.333333	1	0	0	2	0	> BOUND VIRUS
RTNAVLFEKR	0.428571	0.333333	1	0	0	2	0	> BOUND VIRUS
TIPGAGTAPTVR	0.428571	0.333333	1	0	2	2	2	> BOUND VIRUS
TLKTPLVKSFK	0.428571	0.333333	1	0	2	2	2	> BOUND VIRUS
TNAVLFEKR	0.428571	0.333333	1	0	0	2	0	> BOUND VIRUS
TSLKSAFSASR	0.428571	1	0.333333	0	0	0	2	> BOUND VIRUS
SFSGKKGQVSFKPTLNQQR	0.428571	1	1	0	0	1	1	> BOUND VIRUS
SRITISKDNSKNQFSLQLR	0.428571	1	1	0	0	1	1	> BOUND VIRUS
SSMEHEIGPGQANEDAQGTG	0.428571	1	1	0	0	1	1	> BOUND VIRUS
TKAHITKTIPGAGTAPTVR	0.428571	1	1	0	0	1	1	> BOUND VIRUS
VKEKLKDTFA	0.428571	1	1	0	0	1	1	> BOUND VIRUS
YRTLLQPAKFK	0.428571	1	1	0	0	1	1	> BOUND VIRUS
ALPIFTPDWEKR	0.485714	1	1	1	0	2	1	> BOUND VIRUS
KDSGTISTDVKNFPSVLDLSK	0.485714	1	1	1	0	2	1	> BOUND VIRUS

Figure 18. Top antiviral peptide candidates for peptides from the database search.

	FISHER EXACT TEST			COUNT				
	HLvsIN	HPMA(HLvsIN)	N4(HLvsIN)	HP-HL	N4-HL	HP-IN	N4-IN	
KELVQFALLP	0.428571	0.333333333	1	2	0	0	0	
KELVQFVVPV	0.428571	0.333333333	1	0	0	2	0	> BOUND VIRUS
MHARACGDM	0.428571	0.333333333	1	0	0	2	0	> BOUND VIRUS
MHMPVDCGC	0.428571	0.333333333	1	0	0	2	0	> BOUND VIRUS
SQFGPVFVV	0.428571	0.333333333	1	0	0	2	0	> BOUND VIRUS
LGKKPAKL	0.428571	1	0.33333333	0	2	0	0	
LGSAFGRRDNLRQPS	0.428571	1	0.33333333	0	0	0	2	> BOUND VIRUS
SGLKSHVLQLTNHQVHR	0.428571	1	0.33333333	0	0	0	2	> BOUND VIRUS
ALKKAFLLP	0.428571	1	1	1	1	0	0	
KELVQFLLAP	0.428571	1	1	0	0	1	1	> BOUND VIRUS
SQFSLQLR	0.428571	1	1	0	0	1	1	> BOUND VIRUS
VDPKFNKPPV	0.428571	1	1	0	0	1	1	> BOUND VIRUS
PPVTDRLTQR	0.485714	0.333333333	1	2	1	0	1	> BOUND VIRUS
MHDDDGSM	0.485714	1	1	2	1	1	0	
TLKTLPVK	0.485714	1	1	0	1	1	2	> BOUND VIRUS

Figure 19. Top antiviral peptide candidates for the de novo peptides.

2. Develop Machine Learning Algorithms for AVP Prediction:

The series of machine learning developments in AVP have to date focused on increasing the features that characterize a peptide and modifications of the machine learning algorithm. They have not included feature reduction techniques that would determine the most relevant and non-redundant features from the input features used. The performance of a machine learning model relies heavily on

using most informative features and non-informative features can degrade classifier performance. We generated candidate features from the physicochemical and secondary structure properties of known AVP and non-AVP sequences, Table 7. We identified the most important features by estimating Pearson's correlation coefficient and mean decrease of Gini index (MDGI) for all candidate features. We then applied a recursive feature elimination (RFE) algorithm to determine the most important features to aide in interpretability of the model. Finally, we developed a machine learning predictive model based on a Support Vector Machine (SVM) to predict antiviral activity based on sequence.

Peptide Feature	Feature dimension
Amino acid composition	20D
Dipeptide composition	400D
Pseudo-amino acid composition	25D
Amphiphilic pseudo-amino acid composition	30D
Composition/transition/distribution	168D
Secondary structure sequence	6D

Table 7. List of 649 Peptide Features

The performance of the SVM classifiers with our best feature sets is evaluated by estimating sensitivity, specificity, accuracy and Mathew's Correlation Coefficient (MCC) values using Eqs.1-4, where TP, TN, FP, and FN are true positives (positives accurately classified), true negatives (negatives accurately classified), false positives (negatives classified as positives), and false negatives (positives classified as negatives), respectively. Table 8 shows that the new SVM approach achieved better prediction accuracy as compared to the state-of-the-art in all four core metrics reported by previous algorithms.

The Feature-Informed Machine Learning approach to AVP prediction (FIRM-AVP) has been accepted to the journal Scientific Reports and is In Press. In addition to improving overall accuracy as seen in Table 8, mostly using the Support Vector Machine (SVM) approach versus a Random Forest (RF) or Deep Learning (DL) methods, we have also developed a web-service to enable users to evaluate AVP peptides (Figure 20).

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)	MCC
FIRM-AVP (SVM)	93.3	91.1	92.4	0.84
FIRM-AVP (RF)	95.0	82.2	89.5	0.79
FIRM-AVP (DL)	91.7	80.0	86.7	0.73
AVP-649D (SVM)	95.0	82.2	89.5	0.79
AVP-649D (RF)	90.0	82.2	86.7	0.73
AVPcompo	83.3	88.9	85.7	0.72
AVPphysico	88.3	82.2	85.7	0.71
RFcompo+structure+agg	91.7	86.7	89.5	0.79

Table 8. Performance comparison of our models with existing models on validation data where MCC is the Matthew's Correlation Coefficient, for which the higher the value the more overall accuracy of the model.

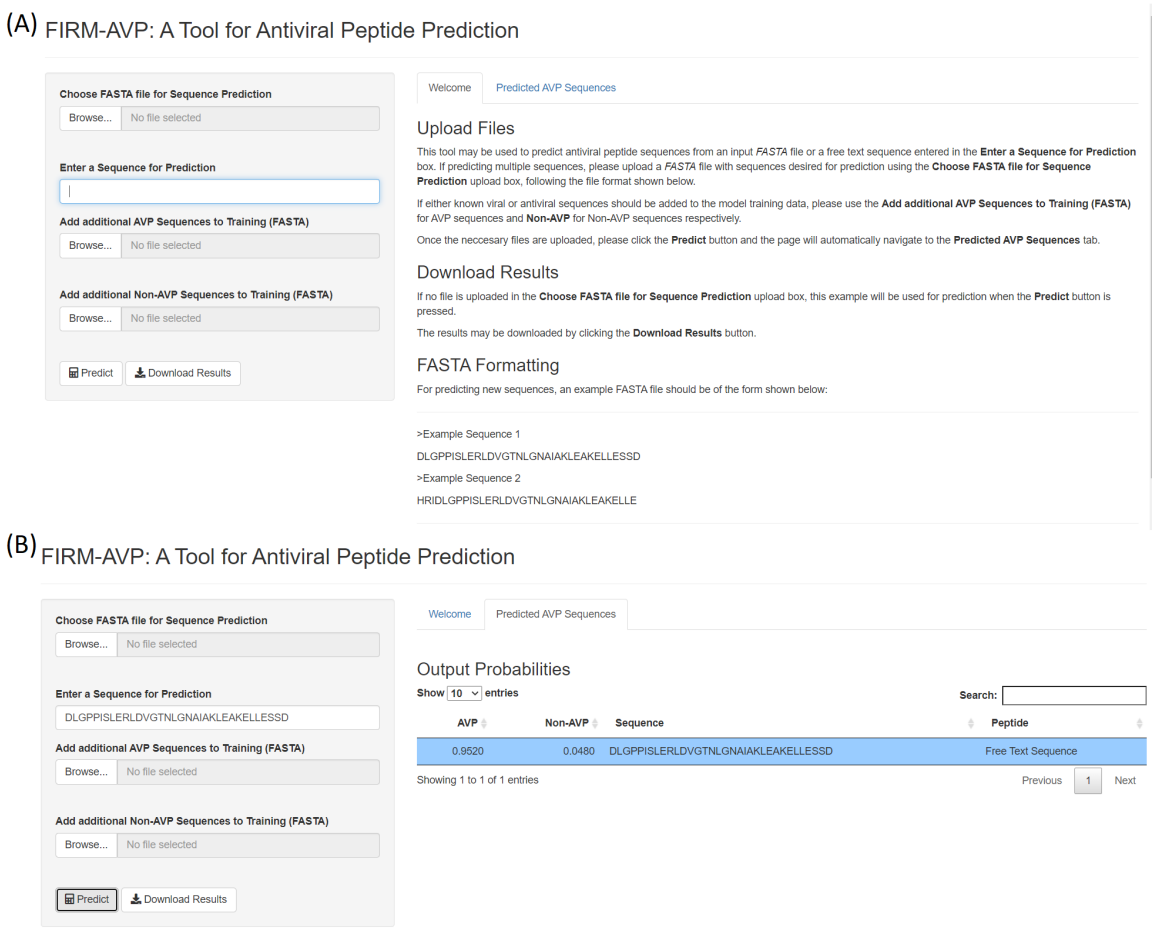

Figure 20. Online FIRM-AVP software interface (https://msc-viz.emsl.gov/AVPR/). Where (A) is the starting page that allows users to either paste in a single peptide sequence or upload a FASTA file containing a collection of peptide sequences. Example sequences and files are given. (B) The probability of AVP versus non-AVP is returned for each sequence based on the pasted peptide sequence or the uploaded FASTA file.

Major Task 5: Assess performance of likely antiviral peptides (6 months: 09/30/19-03/29/20)

1.) One or more novel alligator or snake peptides that exhibit antiviral properties against VEEV.

Five peptides were identified as binding specifically to the cell-membrane particles from infected cells as compared to cell-membrane particles from uninfected cells. These five peptides were purchased and tested for in vitro cytotoxicity and anti-viral activity using Vero cells. LL37 was included as a positive control as it has been shown to inhibit VEEV replication (Ahmed et al., 2019, PMID: 30738837). We also included a scrambled LL-37 as a negative control. Cell viability was determined using CellTiter-Glo (Promega), which measures ATP within cells. Minimal cell toxicity was observed in Vero cells treated with up to 100 ug/mL of the novel peptides (Figure 21A). In particular, peptides KF, AR, and RS did not induce any cell death even at 100 ug/mL However, there

was some toxicity (~25%) observed in cells treated with 100 ug/mL of LL-37, but not scrambled LL-37.

Antiviral activity was determined by pre-treating cells for 1 hour with peptides, infecting cells with a firefly luciferase reporter virus (VEEV-luc) for 1 hour, followed by post-treatment with peptides. Luminescent was measured at 16 hours post-infection as an indicator of viral replication. Peptide GK showed no antiviral activity (Figure 21B). Peptide PK only inhibited VEEV at 100 ug/mL Peptides KF, AR, and RS showed dose dependent inhibition of VEEV, with peptide AR showing the most potent inhibition. LL-37 potently inhibit VEEV replication down to background levels, but surprisingly the scrambled LL-37 also inhibited VEEV. The most potent novel peptide is AR, which is derived from Alligator complement-3.

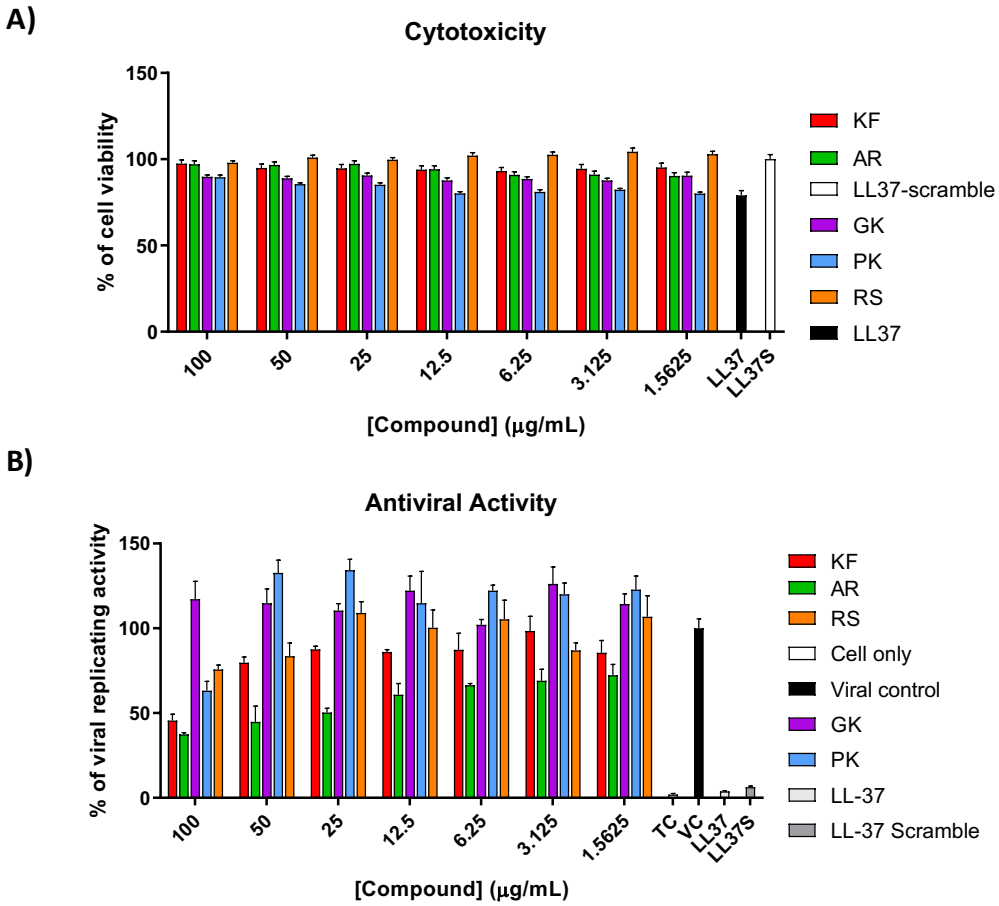

Figure 21: Cell toxicity and antiviral activity of novel peptides. A) Vero cells were treated with various concentrations of peptides (as indicated) for 16 hours. LL-37 and scrambled LL-37 were used at 100 ug/mL Cell viability was measured using CellTiter-Glo. Untreated cells were set to 100% and all treatments normalized to these values. Values graphed are averages plus standard deviations, N=3 **B)** Vero cells were pre-treated with the indicated concentrations of peptides for 1 hour, infected with VEEV-luciferase reporter virus (MOI 1) for 1 hour, and then post-treated with peptides. Luminescence was measured at 16 hours post-infection. LL-37 and scrambled LL-37 were used at 100 ug/mL Untreated infected cells (viral control) were set to 100% and all treatments normalized to these values. Cells only are uninfected cells which display only background luminescence signal. Values graphed are averages plus standard deviations, N=3.

Additional assays were performed to confirm the antiviral activity of AR-17. First, we tested the ability of AR-17 to suppress replication of VEEV-GFP reporter virus. As can be seen in Figure 8, treatment with 100 ug/mL and 50 ug/mL of AR-17 suppressed the expression of GFP as compared to the untreated virally infected samples (**Figure 22**).

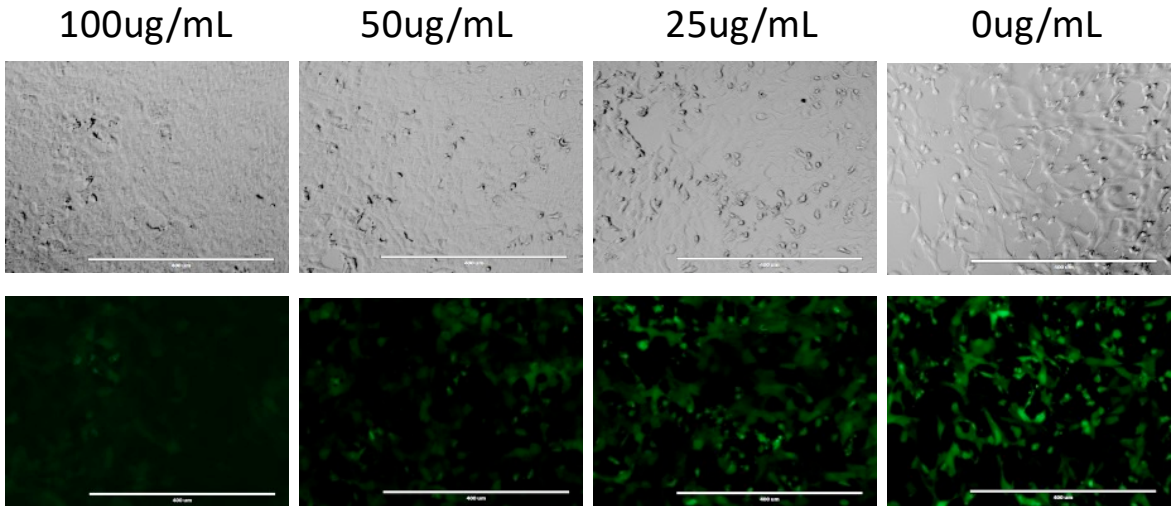

Figure 22. Peptide AR-17 inhibits VEEV replication. Vero cells were pre-treated with the indicated concentrations of peptide for 1 hour, infected with VEEV-GFP reporter virus (MOI 1) for 1 hour, and then post-treated with peptides.

Second, we tested the impact of AR-17 on VEEV replication through plaque assays which measure the amount of infectious virus being released from cells. At 8 hours post infection (hpi) AR-17 decreased VEEV titers by ~2 logs and by ~ 1 log at 16 hpi (**Figure 23**). These data confirm that AR-17 inhibits VEEV replication.

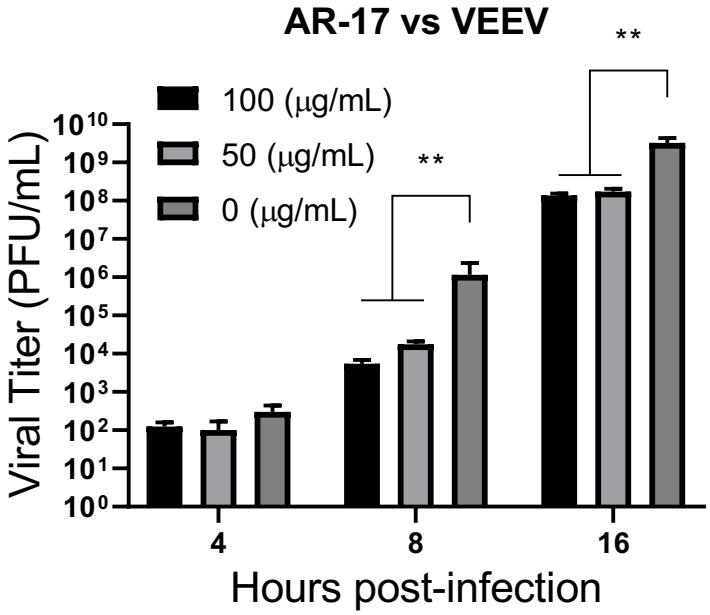

Figure 23. AR-17 inhibits VEEV infectious titers. Vero cells were pre-treated with the indicated concentrations of peptide for 1 hour, infected with VEEV (MOI 1) for 1 hour, and then post-treated with peptides. At 4, 8, and 16 hpi viral supernatants were collected and tittered by plaque assay.

A second set of nine peptides were tested (**Table 9**). These peptides were harvested preferentially using particles bearing either membranes from infected human leukocytes or covalently bound VEEV as baits. In laboratory testing, none of the peptides displayed any toxicity at up to 100 μg/mL towards Vero cells (Figure 24). Antiviral activity testing for these peptides is currently in progress.

Name	Sequence	Molecular Weight	Charge	Origin
AM-EF	AVGVIKAVDKKAAGAGKVTK	1911.3	+4	Elongation Factor
AM-VIT	SRGPRLISRDWR	1498.7	+3	Vitronectin
AM-Plasmin	ALVLRRTNAVLFEKR	1786.2	+3	Plasminogen Isoform
AM-Comp	YRTLLQPAKFK	1364.6	+3	Complement component C8
AM-SerIg	TKAHITKTIPGAGTAPTVR	1920.2	+3	Immunoglobulin
AM-IgG	SRITISKDNSKNQFSLQLR	2235.5	+3	Immunoglobulin
AM-Rib1	IVDVKANKHQIKQAVKKL	2060.5	+4	Ribosomal protein
AM-Rib2	ANGTTVHVGIHPSKVVITR	1985.3	+2	Ribosomal protein
AM-His	KVLKQVHPDTGISSK	1636.9	+2	Histone

Table 9. Second round of peptides selected from alligator plasma harvests. The peptide AM-EF was identified from two separate harvest studies and was strongly associated with harvests using membranes form infected cells, making it a promising candidate, and one of the few peptides that we identified that appears in its entirety in multiple harvests. The peptides AM-VIT, AM-Rib1, AM-Rib2 and AM-His were identified in harvests from plasma and stimulated PBMC's using particle bearing membranes from VEEV-infected human leukocytes. The remaining peptides, AM-Plasmin, AM-Comp, AM-SerIg and AM-IgG were identified specifically from harvests using particles bearing covalently bound VEEV as baits.

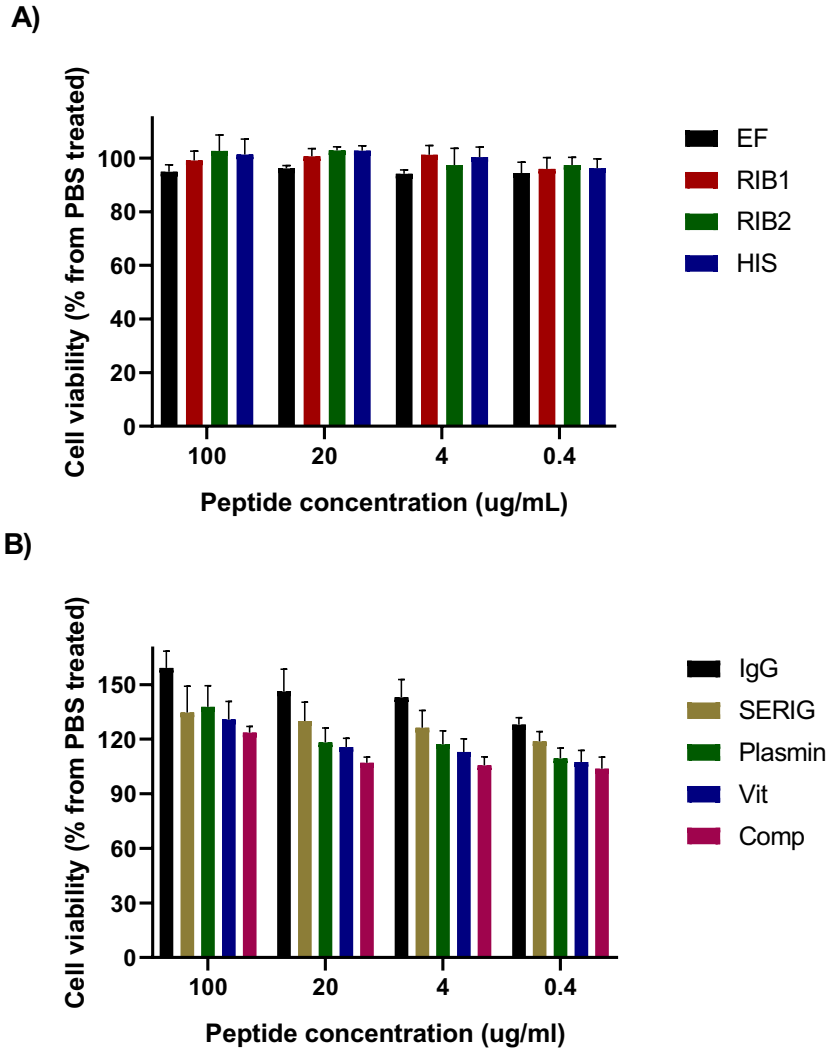

Figure 24: Cell toxicity of 2nd set of novel peptides. Vero cells were treated with various concentrations of peptides (as indicated) for 16 hours. Cell viability was measured using CellTiter-Glo. Untreated cells were set to 100% and all treatments normalized to these values. Values graphed are averages plus standard deviations, N=3

What opportunities for training and professional development has the project provided?

Over the course of the past year, one student (Victoria Callahan) and two research associates (Shih-Chao Lin and Ivan Akhrymuk) have been involved in this project from the Kehn-Hall Lab. Over the same period, two graduate students (Amy Carfagno and Samuel Garvey) and a research associate (Liana Soares) have been involved on this project from the Bishop Lab. Dr. Bishop, Dr. Kehn-Hall and researchers from both labs meet frequently to plan and coordinate their efforts on the project.

This project provides a unique training environment because it is a multidisciplinary project, which allows researchers from the biology/virology side to be exposed to the application of chemistry to address a biological question. They are also getting experience stimulating reptile blood cells with

virus, which is outside the scope of the typically human focused research in the Kehn-Hall Lab. Similarly, the project provides researchers with backgrounds in organic chemistry, materials science and proteomic mass spectrometry the opportunity to work with virologists and biologists to gain experience and understanding of the biology that is at the heart of the project.

Researchers on the project contribute in the preparation of the reports and will also be included in the preparation of manuscripts produced by the project. We also anticipate they will present their data future conferences.

How were the results disseminated to communities of interest?

Our strategy for disseminating the results and findings of the project to communities of interest focus primarily on publication of technical manuscripts in peer-reviewed journals and presentations and national and other professional meetings. This year:

- We presented a poster on the project and our findings at the Defense Threat Reduction Agency CBD S&T conference in November of 2019 (see below under **"Other publications, conference papers, and presentations"**).
- We also presented on the project and our findings at ASM Biothreats in January, 2020. (see below under **"Other publications, conference papers, and presentations"**).
- Data from the project was also included as part of a DTRA Tech Watch Seminar presented by Dr. Bishop in April, 2020.
- Our manuscript reporting development of improved machine learning algorithms for predicting antiviral peptides has been accepted and is in press in the scientific journal Nature Scientific Reports.

What do you plan to do during the next reporting period to accomplish the goals?

In the remaining month of our funded project, we will:

Major Task 1: Develop VLP-based capture particles. We will complete preparation and analysis of the current iteration of the virus-modified particles. Concluding with assessment of viral loading and stability of bound virus particles for the new generation of virus-modified particles. This will complete Major Task 1.

Major Task 4: Analyze harvested peptides. We will complete analysis of the mass spectrometry data from the final round of harvests from last quarter, with a particular focus on data generated from snake plasma harvests. We will also complete statistical analyses of the harvested peptides to identify any potential antiviral peptide candidates that may be present in the harvests. Conclusion of these analyses in October will complete Major Task 4.

Major Task 5: Assess performance of likely antiviral peptides. We will assess the antiviral effectiveness and cytotoxicity of the second round of antiviral peptide candidates identified in the last quarter. We will then compare their performance against the antiviral peptides identified in the first rounds. The conclusion of these studies, leading to identification of promising antiviral peptides, will complete Major Task 5 and the overall objectives of the project.

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

The ultimate goal of this project is to identify novel peptides that have antiviral activity against Venezuelan equine encephalitis virus (VEEV). VEEV is an important human and veterinary pathogen with no current treatment options. Therefore, discovery of antivirals is an urgent unmet need. Due to the nature of antiviral peptides, the discovered peptides will likely have broad-spectrum activity, being able to inhibit multiple viruses within the alphavirus family that cause significant disease, such as chikungunya virus (CHIKV) and eastern equine encephalitis virus (EEEV). The peptide discovery process developed here, one that tightly integrates advanced peptididomic analytical technologies, rigorous performance evaluation and machine learning, can in the future be adapted to include other viruses or classes of virus. In expanding the scope of the process and tools to other viruses, we anticipate that it can be developed into a platform for identifying peptides with broad-spectrum antiviral efficacy and basis for developing broad-spectrum antiviral therapeutics for use against emerging viral pathogens and future pandemic outbreaks. Additionally, this work will provide basic immunological information about the peptide response to VEEV infection in reptile species.

What was the impact on other disciplines?

We are developing novel particles that can be used to isolate peptides associate with the innate immune response. These particles can be used as discovery tools for multiple other disciplines including cancer biology, neurobiology and immunology. There are also implications for using these particles as sensors.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

The rapid spread of SARS-CoV-2, and its evolution into the COVID-19 pandemic that currently grips the United States and the rest of the world has revealed the inadequacy of our preparations for such emerging viral threats and the need for new strategies new therapeutics. Our successful identification of novel peptides that exhibit antiviral properties against the target pathogen Venezuelan equine encephalitis virus (VEEV), demonstrates the potential of this strategy for antiviral peptide discovery. While our current efforts are focused on identifying antiviral peptides that are effective against alpha viruses, specifically targeting VEEV, the peptide discovery platform and tools that we are developing can be readily adapted and extended to include a broader spectrum of viruses, including SARS-CoV-2. Hence, the present research provides the foundation for the development of a powerful antiviral peptide discovery and design platform and an arsenal of novel effective antiviral peptides to combat emerging viruses and future outbreaks in their early stages. Moreover, the antiviral peptide discovery platform and tools that we are developing should be amenable to being adapted to target other pathogens including fungal and parasites. Thus, this research has potential broad biomedical

applications and the potential to have a significant impact on the health of both warfighters and civilians.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

Nothing to Report

Actual or anticipated problems or delays and actions or plans to resolve them

The SARS-CoV-2 pandemic and its impact on GMU operations became the biggest challenge to us completing project tasks and achieving deliverables within the remaining time allocated for the project (reflecting original requested and approved no-cost-extension to May). In mid- March, the university shifted to remote learning/teaching for all Spring semester classes, and the preparations and changes in schedule impacted faculty, staff and student's work on the project. In late March, GMU effectively temporarily shut down all research that was not determined to be critical SARS-COV-2 research, which included our AIMM project. At the time of the shutdown, harvests from alligator plasma and stimulated media along with snake plasma were at various stages of processing and analysis. Many of these samples had not been analyzed on the mass spectrometer at the time of the shutdown and were placed in cold storage (-80C) until our return. However, we were able to remotely analyze mass spectrometry data from those samples that had been analyzed on the instrument. We identified promising peptides in the third quarter based on analyses performed by the team at PNNL.

At the time that GMU shut down non-SARS-CoV-2 research, we requested and were granted an additional no cost extension through August 30 to provide us added time to compensate for the lost time and effort associated with the shut-down. At the time, we believed that if we were able to return to the lab in early summer, we are positioned to wrap up these studies and deliver on our project deliverables by that time. However, when we were allowed to return to lab in June, the restrictions and time constraints put in place as part of GMU's safety precautions in response to the pandemic slowed our progress on the project. Accordingly, we requested and were granted an additional no cost extension to October 30. Based on our experience working under the new rules and conditions, we believed that this would allow us to complete the projects objectives and deliverables.

In August, we arranged for the synthesis of the identified antiviral peptide candidates from New England Peptides. These peptides arrived later that month, and we have begun assessing their properties. We believe that we are positioned to wrap up these studies and deliver on our project deliverables by that time.

Changes that had a significant impact on expenditures

Nothing to Report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to Report

Significant changes in use or care of human subjects

Not applicable

Significant changes in use or care of vertebrate animals

Protocol [ACURO Assigned Number]: ACURO Log Number = DM171301

Title: Translational Peptide Research for Personnel Protection

Target required for statistical significance: Proposed 20 crocodilians and 15 snakes

Target approved for statistical significance: Approved

Protocol Modification (1 OF 1)

Protocol modified to include cardiocentesis as means of blood collection in snakes.

SUBMITTED TO AND APPROVED BY:

Protocol originally submitted 01/15/2019 – Approved 03/20/2019

Modification Submitted 01/24/20 – Approved 02/06/2020

STATUS:

• Protocol modification has been approved.

This year:

• 50 ml of American alligator blood was collected from 4 animals at the St. Augustine Alligator Farm Zoological Park on 10/23/2019 and received at GMU on 10/24/2019.

• 32 ml of American alligator blood was collected from two animals at the St. Augustine Alligator Farm Zoological Park on 12/12/2019 and received at GMU on 12/13/2019.

• 2 ml of Boa constrictor blood was collected from one animal at the St. Augustine Alligator Farm Zoological Park on 02/20/2020 and received at GMU on 02/21/2020.

Rationale for modification: The resident veterinarian at the St. Augustine Alligator Farm Zoological Park advised us that an alternative method was preferred for collecting large volumes of blood from constrictor snakes than that which was included in our approved IACUC and ACURO protocol. We requested a modification to our IACUC protocol to include this method, which was approved by GMU IACUC. We then submitted our request for approval of this modification to ACURO, who in turn approved the modification.

Significant changes in use of biohazards and/or select agents

Nothing to report

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

Chowdhury, A., Reehl, S., Kehn-Hall, K., Bishop, B., Webb-Robertson, B. "Better Understanding and Prediction of Antiviral Peptides through Primary and Secondary Structure Feature Importance", *Nature Scientific Reports*, 2020, accepted and in press. (yes - acknowledges federal support)

Books or other non-periodical, one-time publications.

Nothing to report

Other publications, conference papers and presentations.

Poster Presentation at DTRA CBD S&T 2019 conference:
Bishop, B., Carfagno, A., Chafran, L., Callahan, V., Lin, S., Akhrymuk, I., Tsogtbayar, D., Po, M., Lamont, S., Webb-Robertson, B., and Kehn-Hall, K. "Reptilian Host Defense and New Antiviral Agents", presented at DTRA CBD S&T conference, Cincinnati, Ohio, November 18-21, 2020

Webb-Robertson, B., Bishop, B., Kehn-Hall, K. "A Biosprospecting Approach to Identify Host Defense Peptides that Target Viruses", presented at ASM Biotreats, Arlington, Virginia, January 28-20, 2020

- **Website(s) or other Internet site(s)**

The antiviral peptide prediction tool that we developed can be accessed at https://msc-viz.emsl.gov/AVPR/.

- **Technologies or techniques**

We have developed a novel Feature-Informed Machine Learning approach to AVP prediction (FIRM-AVP) and a manuscript describing our algorithms and their accuracy has been submitted to the journal Scientific Reports. In addition to improving overall accuracy using the Support Vector Machine (SVM) approach versus a Random Forest (RF) or Deep Learning (DL) methods, we have also developed a web-service (https://msc-viz.emsl.gov/AVPR/) to

allow users to submit peptide sequences for analysis and predict whether they are likely to have antiviral properties.

- **Inventions, patent applications, and/or licenses**

  FIRM-AVP has been released open source on GitHub (https://github.com/pmartR/FIRM-AVP) and is licensed up the BSD 2-Clause "Simplified" License

- **Other Products**

  Nothing additional to report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: *Barney Bishop*
Project Role: *Principle Investigator*
Researcher Identifier (e.g. ORCID ID): *0000-0002-6626-9251*
Nearest person month worked: *7*
Contribution to Project: *As principle investigator, Dr. Bishop manages the project and coordinates research efforts between the GMU teams and the team at PNNL. He is also directly involved with particle and peptide harvest development as well as analysis of harvested peptides via mass spectrometry.*

Name: *Kylene Kehn-Hall*
Project Role: *co-Principle Investigator*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *5*
Contribution to Project: *As co-principle investigator, Dr. Kehn-Hall has worked closely with Dr Bishop in coordinating research efforts. She is directly involved with microbiology/virology aspects of the project, including cell culture, preparation of virus like particles, blood stimulation and infection studies.*

Name: Bobbie-Jo Webb-Robertson
Project Role: *PNNL co-Principle Investigator*
Researcher Identifier (e.g. ORCID ID): *0000-0002-4744-2397*
Nearest person month worked: *2*
Contribution to Project: *Dr. Webb-Robertson is responsible for performing the statistical aspects of the projects and leads the PNNL efforts. She has been analyzing the technical replicate data.*

Name: Abu Chowdhury
Project Role: *PNNL post-doc*

Researcher Identifier (e.g. ORCID ID): *0000-0003-3454-5861*
Nearest person month worked: *2*
Contribution to Project: *Dr. Chowdhury is responsible for coding and validating the machine learning AVP prediction methods.*

Name: Sarah Reehl
Project Role: *PNNL data scientist*
Researcher Identifier (e.g. ORCID ID): *0000-0003-3727-5801*
Nearest person month worked: *1*
Contribution to Project: *Ms. Reehl was responsible for continuing code development after Dr. Chowdhury finished his post-doc at PNNL and generated the web-service and public respositories.*

Name: *Paul Russo*
Project Role: *Affiliated Faculty*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *3*
Contribution to Project: *Dr. Russo's area of expertise and contribution is in protein/peptide mass spectrometry. He is responsible for analyzing the harvested peptides via LC-MS/MS and developing methods/protocols for that purpose. He is also working with students in Dr. Bishop's lab in analyzing the mass spectrometry data for the harvested peptides.*

Name: *Sabrina Lamont*
Project Role: *Research Assistant*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *6*
Contribution to Project: *Ms. Lamont has been supporting research efforts by Dr. Bishop and his graduate students in the area of particle development and evaluation. She has helped with coordinating research efforts in the lab.*

Name: *Liana Soares*
Project Role: *Research Associate*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *11*
Contribution to Project: *Dr. Soares is an experienced polymer chemist with experience in materials science and nanotechnology. Dr. Soares has been working with Dr. Bishop and his graduate students to improve particle production and develop new particle technologies for the targeted capture of peptides and proteins of interest from plasma. She has also contributed in performing harvests from Alligator plasma using the hydrogel particles and processing samples for mass spectrometry analysis. She has helped to provide much needed bandwidth in this area. Simultaneously, she is introducing multiple advancements in to the particle harvest technology, enhancing their performance and versatility.*

Name: *Shih-Chao Lin*
Project Role: *Research Associate*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *7*

Contribution to Project: *Dr. Lin is responsible for performing the virological aspects of the projects. He developed the VLP production, plasma membrane isolations and blood stimulation protocols. He has also been responsible for preparing the cell membranes for incorporation in particles. He recently began evaluating the antiviral properties of peptides identified from healthy alligator plasma.*

Name: *Amy Carfagno*
Project Role: *Graduate Student*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *3*
Funding/Support: *GMU Presidential Graduate Fellowship (over the summer received some wages support from the grant for efforts on the project).*
Contribution to Project: *Ms. Carfagno has been supporting research efforts by Dr. Bishop and his graduate students in the area of mass spectrometry data processing and analysis.*

Name: *Victoria Callahan*
Project Role: *Graduate Student*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *2*
Funding/Support: *Volunteer*
Contribution to Project: *Ms. Callahan has been assisting with development of the VLP-based particles, including virus preparation, modification, coupling to particles, and performing harvests.*

Name: *Ivan Akhrymuk*
Project Role: *Post-doctoral Researcher*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *3*
Contribution to Project: *Dr. Akhrymuk has been assisting with development of the VLP-based particles, including virus preparation, modification, coupling to particles, and performing harvests.*

Name: *Samuel Garvey*
Project Role: *Graduate Student*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *3*
Funding/Support: *GMU Graduate Teaching Assistantship.*
Contribution to Project: *Mr. Garvey developed the protocol for synthesizing the PP4ME monomer and is involved with developing hydrogel particles for use in generating virus-modified particles.*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Dr. Barney Bishop has received the following active support since the time of the last reporting period:

GMU Grant # 223632
Role: PI
Project Title: Collaborative Research using BioProsector Approach
Sponsor: Regents of the University of California Davis
Total Award Amount: **Total Award Period:** 4/1/2020 – 03/31/2021
Commitment in Person-Months per Year: 0.00 Sumr. and 0.9 Acad.

Dr. Bobbie-Jo Webb-Robertson has received the following active support since the time of the last reporting period:

PNNL Grant # DOE 76764
Role: PI
Project Title: Molecular Design and Analysis to Inform Therapeutics Releated COVID-19
Sponsor: Department of Energy Office of Biological and Environemental Research
Total Award Amount: **Total Award Period:** 4/27/2020 – 03/31/2021
Commitment in Person-Months per Year: 1.0

What other organizations were involved as partners?

Organization Name: Pacific Northwest National Laboratory (PNNL)

Location of Organization: 902 Battelle Blvd, Richland, WA 99354

Partner's Contribution: Dr. Bobbie-Jo Webb-Robertson, of the Biological Sciences Division at PNNL, is collaborating with us on the project via a subcontract. She is primarily contributing in Major Task 4 efforts and her focus is on statistical analyses of the peptides identified from the harvests in Major Task 3 in order to identify those peptides uniquely associated with harvests performed using virion-modified particles as well as particles encapsulated in membranes from VEEV-infected Leukocytes. These peptides are expected to be potential antiviral and will by synthesized and evaluated.

Organization Name: Virginia Tech

Location of Organization: 1981 Kraft Drive, Blacksburg, VA 24060

Partner's Contribution: Dr. Kylene Kehn-Hall, Co-PI on this project, left George Mason University in August to join the faculty in the Department of Biomedical Sciences and Pathobiology at Virginia Tech. Her laboratory continues to contribute to Major Task 1 and 5 including final optimization studies on the VLP-particles and analysis of peptides for antiviral activity.

8. **SPECIAL REPORTING REQUIREMENTS**

   COLLABORATIVE AWARDS:

   QUAD CHARTS:

9. **APPENDICES:** *Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.*