



Mining Amphibian and Insect Transcriptomes for Antimicrobial Peptides

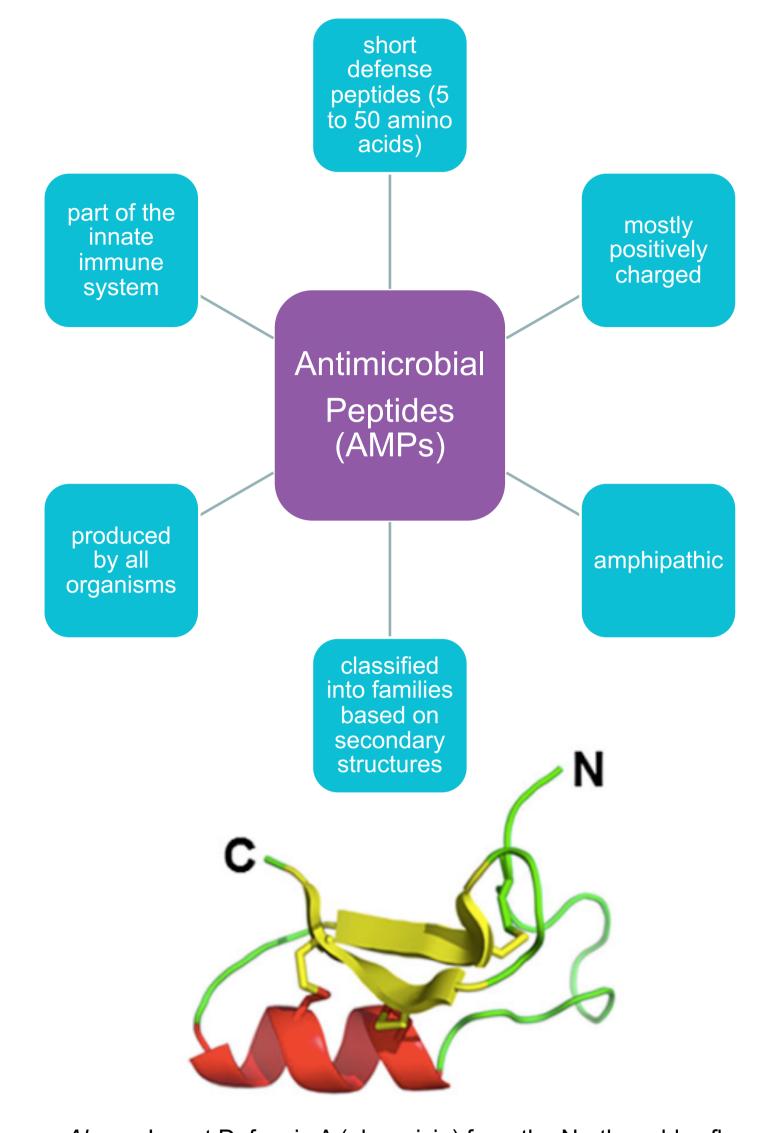
Diana Lin^{1,2}, Ka Ming Nip^{1,2}, Chenkai Li^{1,2}, René L. Warren¹, Caren Helbing³, Linda Hoang^{4,5}, Inanc Birol^{1,6}

¹Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada ⁵Public Health Laboratory, British Columbia Centre for Disease Control, Vancouver, BC, Canada ²Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC, Canada ⁶Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada ³Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada ⁴Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

Introduction

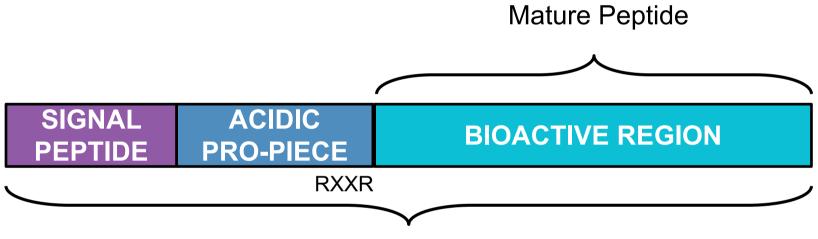
Antimicrobial Peptides (AMPs)

Characteristics¹



Above: Insect Defensin A (phormicin) from the Northern blowfly https://www.rcsb.org/structure/1ICA https://www.uniprot.org/uniprot/P10891

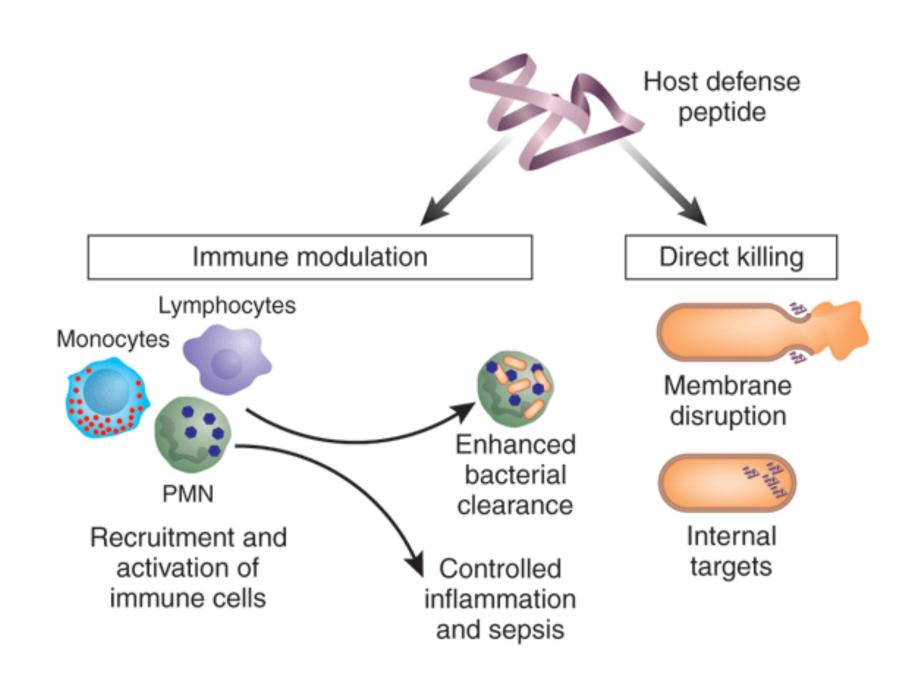
Peptide Structure¹



Precursor

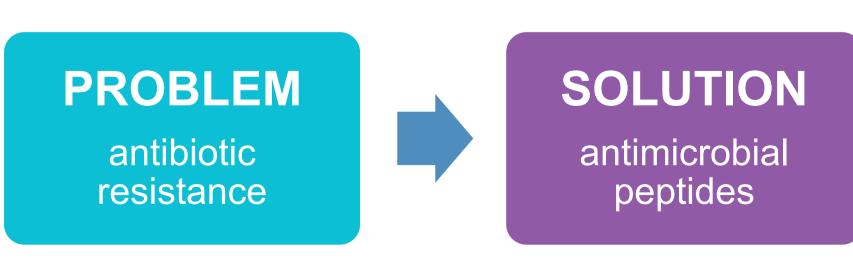
- AMPs are activated by cleavage at the RXXR motif (acidic propiece inhibits basic bioactive region)
- Cleavage separates the signal peptide and acidic propiece from the bioactive region yielding the mature peptide

Mechanisms of Action¹



Source: Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat. Biotechnol. 24, 1551–1557 (2006) doi: 10.1038/nbt1267

Motivation



Problem

- The rise of antibiotic resistance¹
- The antibiotic "discovery" void²: few new antibiotics, but old antibiotics less effective
- The need for novel methods to fight pathogen

Solution

- AMPs do not confer resistance as easily as antibiotics, due to co-evolution with the human microbiome¹
- AMPs are a potential alternative to antibiotics³
- AMPs can be mined from organisms of rich AMP diversity, such as the North American bullfrog⁴

Objectives

- To develop and execute a scalable bioinformatics-based AMP discovery pipeline to mine for AMP sequences in publicly available genomic resources
- To package a fully functional bioinformatics pipeline
- To obtain a list of potential AMP sequences for
 - Downstream analysis
 - In vitro bioactivity testing
 - Drug development

INPUT Publicly available RNA-seg reads

PIPELINE

OUTPUT Potential AMP protein sequences

Methods



- Download bulk RNA-seq reads from the <u>Sequence Read Archive</u>⁵ (SRA) using SRA toolkit using SRA accession numbers
- 45 amphibian and 39 insect bulk RNAseq datasets from skin, venom gland, and other tissues

Pre-process Reads

- Trim to remove adapter sequences using <u>fastp</u>⁶
- Filter out poor-quality bases and sequences
- fastp finds adapter sequences using sequence overlap

Assemble **Transcripts**

- Multi-sample pooled de novo or reference-guided assembly of reads into transcripts using RNA-Bloom⁷, for both single and paired-end reads
- Lowly-expressed transcripts are filtered out using quantification from Salmon⁸

Translate Transcripts

- *In silico* translation of transcripts into amino acid space
- Six-frame translation and open reading frame (ORF) prediction using Transdecoder⁹

Homology Search

- Homology search with jackhmmer from the <u>HMMER</u>¹⁰ package
- HMMs built using amphibian and insect precursor and mature AMP sequences from NCBI

 Additional mature AMPs collated from the APD3¹¹ and DADP¹²



- Predict signal peptide and propeptide cleavage sites using prediction tool ProP¹³
- Separate signal peptide and acidic propieces from the bioactive region

Characterize **AMPs**

- Predict secondary structures and solvent accessibility using **SABLE**¹⁴
- Using SABLE, alpha helix and beta strand structures can be predicted

Prioritize AMPs

- Obtain the probability that each sequence is an AMP using AMPlify¹⁵
- Rank potential AMPs and prioritize which sequences are synthesized and tested in vitro for bioactivity

Results Total # AMPs (HMMs and AMPlify) to date: 105,357 1e+06 Toads Ants Wasps Frogs Bees # Filtered Transcripts # AMPs (HMMs) # ORFs # AMPs (HMMs and AMPlify) # Transcripts ☐ Skin or Venom Gland * Reference Available В Total # short and 1e+08confident AMPs to date: 2,692

Short AMPs # Confident AMPs # Short and (score >= 0.99) # Confident AMPs A. Breakdown of data in each step in the pipeline

Ants

* * * * **

Wasps

B. Further breakdown of AMPs found using HMMs and AMPlify (pink in Fig. A), by length and AMPlify's confidence score*

*A sequence is predicted to be an AMP if its score >=0.50

Conclusions

Frogs

- In each of the assembled transcriptomes, a large set of potential AMPs using HMMs were found
- Those AMPs can be further filtered by AMPlify score (>= 0.99) and length (<= 50) to obtain shorter and higher-confidence peptides (ideal for synthesis)

References

1e+02

- 1. Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat. Biotechnol. **24**, 1551–1557 (2006) doi: <u>10.1038/nbt1267</u>
- Peptide Science vol. **111** (2019) doi: <u>10.1002/pep2.24122</u> 3. Reinhardt, A. & Neundorf, I. Design and Application of Antimicrobial Peptide Conjugates. Int. J. Mol. Sci. 17, (2016) doi:

2. Koo, H. B. & Seo, J. Antimicrobial peptides under clinical investigation.

- 10.3390/ijms17050701 4. Helbing, C. C. et al. Antimicrobial peptides from Rana [Lithobates] catesbeiana: Gene structure and bioinformatic identification of novel forms from tadpoles. Sci. Rep. 9, 1529 (2019) doi: 10.1038/s41598-018-
- 38442-1 Leinonen, R., Sugawara, H., Shumway, M. & International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic* Acids Res. 39, D19–21 (2011) doi: 10.1093/nar/gkq1019
- 6. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018) doi: 10.1093/bioinformatics/bty560
- 7. Nip, K. M. et al. RNA-Bloom enables reference-free and referenceguided sequence assembly for single-cell transcriptomes. Genome Res. **30**, 1191–1200 (2020) doi: <u>10.1101/gr.260174.119</u>
- 8. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419 (2017) doi: 10.1038/nmeth.4197
- 9. Haas, B. J. et al. De novo transcript sequence reconstruction from RNAseq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512 (2013) doi: 10.1038/nprot.2013.084
- 10. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 11, 431 (2010) doi: 10.1186/1471-2105-11-431
- 11. Wang, G., Li, X. & Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Res. 44, D1087–93 (2016) doi: <u>10.1093/nar/gkv1278</u>
- 12. Novković, M., Simunić, J., Bojović, V., Tossi, A. & Juretić, D. DADP: the database of anuran defense peptides. Bioinformatics 28, 1406–1407
- (2012) doi: 10.1093/bioinformatics/bts141 13. Duckert, P., Brunak, S. & Blom, N. Prediction of proprotein convertase cleavage sites. Protein Eng. Des. Sel. 17, 107-112 (2004) doi:
- 10.1093/protein/gzh013 14. Adamczak, R., Porollo, A. & Meller, J. Combining prediction of secondary structure and solvent accessibility in proteins. Proteins 59,
- 467–475 (2005) doi: 10.1002/prot.20441 15. Li, C. et al. AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens. bioRxiv **2020.06.16.155705** (2020) doi: <u>10.1101/2020.06.16.155705</u>

Funding & Acknowledgements



GenomeCanada



Foundation

of British Columbia









https://github.com/dy-lin