microbial cell

In a nutshell

www.microbialcell.com

Paving the way for new antimicrobial peptides through molecular de-extinction

Karen O. Osiro[1], Abel Gil-Ley[2], Fabiano C. Fernandes[1,3], Kamila B. S. de Oliveira[2], Cesar de la Fuente-Nunez[4-7,*], Octavio L. Franco[1,2,*]

[1] Centro de Análises Proteômicas e Bioquímicas, Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília 70790-160, Brazil. [2] S-Inova Biotech, Pós-graduação em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande, Mato Grosso do Sul, Brazil. [3] Departamento de Ciência da Computação, Instituto Federal de Brasília, Campus Taguatinga, Brasília, Brazil. [4] Machine Biology Group, Departments of Psychiatry and Microbiology, Institute for Biomedical Informatics, Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania; Philadelphia, Pennsylvania, United States of America. [5] Departments of Bioengineering and Chemical and Biomolecular Engineering, School of Engineering and Applied Science, University of Pennsylvania; Philadelphia, Pennsylvania, United States of America. [6] Department of Chemistry, School of Arts and Sciences, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America. [7] Penn Institute for Computational Science, University of Pennsylvania; Philadelphia, Pennsylvania, United States of America.

* Corresponding Authors:
Octavio Luís Franco, E-mail: ocfranco@gmail.com
Cesar de la Fuente-Nunez, E-mail: cfuente@upenn.edu

ABSTRACT Molecular de-extinction has emerged as a novel strategy for studying biological molecules throughout evolutionary history. Among the myriad possibilities offered by ancient genomes and proteomes, antimicrobial peptides (AMPs) stand out as particularly promising alternatives to traditional antibiotics. Various strategies, including software tools and advanced deep learning models, have been used to mine these host defense peptides. For example, computational analysis of disulfide bond patterns has led to the identification of six previously uncharacterized β-defensins in extinct and critically endangered species. Additionally, artificial intelligence and machine learning have been utilized to uncover ancient antibiotics, revealing numerous candidates, including mammuthusin, and elephasin, which display inhibitory effects toward pathogens *in vitro* and *in vivo*. These innovations promise to discover novel antibiotics and deepen our insight into evolutionary processes.

doi: 10.15698/mic2025.02.841
Received originally: 06.10.2024;
in revised form: 02.01.2025,
Accepted 07.01.2025,
Published 20.02.2025.

Keywords: molecular de-extinction, AMPs, encrypted peptides, defensins, bioinformatics, machine learning, deep learning.

Abbreviations:
AEP – archaic encrypted peptide,
AMP – antimicrobial peptides,
EP – encrypted peptide,
MEP – modern encrypted peptide,
MIC – minimal inhibitory concentration,
ML – machine learning,
NCBI – National Center for Biotechnology Information.

INTRODUCTION

Molecular de-extinction represents a promising area of scientific research that enables the identification, synthesis, and understanding of molecules' biological functions throughout evolution [1]. However, it also raises interesting bioethical and philosophical debates within the scientific community [2]. Advances in ancient DNA sequencing methods [3] have increasingly allowed us to access biological data from the past. Ancient DNA sheds light on historical protein-coding sequences that may not exist in our current time or have been changed throughout evolution [4, 5]. Furthermore, advancements in computational biology and artificial intelligence [6–9] have shifted the discovery of promising molecules from a chance-based approach to a more intentional and data-driven methodology.

Among the extensive range of molecules that can be found through proteomics or genomics, antimicrobial peptides (AMPs) stand out. AMPs have played crucial roles in the defense mechanisms of animals, evolving over millions of years to protect hosts against various pathogens, thereby ensuring survival in ancient environments [10]. These molecules continue to function in the innate immune systems of various organisms today, fighting mul-

tiple microorganisms. They can modulate the immune system, disrupt cell membranes, target intracellular processes, and inhibit biofilm formation [11–14]. Despite considerable variability in AMP structures [15], identifying specific features is critical for mining these potential antibiotics, particularly when leveraging computational tools.

However, it is important to highlight the gap between the amount of discovered antimicrobial peptides and those that successfully advance the clinical trials and, finally, reach the market [16]. Issues related to stability and toxicity frequently hinder the development cycle, which, in the best-case scenarios, still averages 13 years to be launched [17]. The advancement of artificial intelligence and molecular de-extinction offers a valuable opportunity not only to discover new antimicrobials but also to provide accurate *in silico* predictions, thereby shortening the path to addressing the global antibiotic resistance crisis.

BIOINFORMATIC TOOLS

Molecular de-extinction employs various methodologies to mine compounds such as AMPs and other peptide antibiotics from extinct organisms (**Figure 1**). These approaches utilize diverse bioinformatics tools to discover novel AMPs from the genomes and proteomes [1, 18] of extinct organisms. Digital repositories like the National Center for Biotechnology Information (NCBI) and the Protein Data Bank (PDB) provide access to genomic and proteomic data [19, 20]. One current methodology involves identifying AMPs with specific characteristics of β-defensins from the genomes of these organisms [21]. Programs such as AUGUSTUS can locate protein-coding genes within genomic sequences obtained from NCBI [21, 22]. To refine the identification process, tools such as HMMER [23] and InterPro [24] determine which protein sequences within the selected genomic data belong to the β-defensin protein family. Subsequent structural and physicochemical analyses are conducted to evaluate the potential antimicrobial properties of these proteins. Advanced tools such as AlphaFold 2 [25] or AlphaFold 3 [26] can accurately predict protein structures, while features such as cationicity and amphipathicity can be analyzed using ExPASy to complement the AMP characterization [21].

A different strategy for molecular de-extinction employs machine learning (ML) models to identify and classify encrypted peptides (EPs) — protein fragments with antimicrobial properties — from the proteomes of extant and extinct organisms [1, 18, 27]. The panCleave ML model, for example, applies a pan-protease cleavage site classifier to conduct computational proteolysis, identifying potential EPs within protein sequences [1]. This open-source model achieved over 80% accuracy for proteases, with at least 100 observations in the test set. Specifically, for cysteine catalytic types, the average accuracy was 81.3%, based on 1,858 correct predictions out of 2,286 observations, whereas for threonine catalytic types, the accuracy was 34.6%, with 18 out of 52 observations predicted correctly [1]. Other genome mining tools, such as ThioFinder [28], RODEO [29], and RiPPMiner-Genome [30], have also been widely used to discover new AMPs in extant organisms [31].

Progress in this field is further demonstrated by the development of the deep learning model APEX (**Figure 2**), which has been used to mine all extinct organisms as sources of antibiotics [18]. This state-of-the-art model consists of a peptide sequence encoder coupled with neural networks for predicting antimicrobial activity, enabling the extraction and classification of peptide antibiotics based on their potential Minimal Inhibitory Concentration (MIC) [18]. APEX achieved a significant Pearson correlation (>0.3) for predicting species-specific antimicrobial activity, showing correlations between predicted and experimentally validated activities for several strains, including *Escherichia coli* strains AIC221, AIC222, ATCC 11775; *Acinetobacter baumannii* ATCC 19606; *Pseudomonas aeruginosa* strains PAO1 and PA14; and *Enterococcus faecium* ATCC 700221. Predicting antibiotic activity with advanced artificial intelligence like APEX (**Figure 2**) brings us closer to mining effective candidates for novel antibiotic alternatives. The antimicrobial activity of these ancient compounds can potentially also be predicted using various deep learning models, such as AMP-Bert [32] or AMPlify [33].

UNVEILING ANTIMICROBIAL PEPTIDES FROM EXTINCT ORGANISMS

Natural AMPs and other peptide antibiotics can originate from four main processes: (i) genome-encoded peptides [21, 34], (ii) cleavage via proteolysis [27, 35], (iii) synthesis by non-ribosomal means [36], and (iv) small open reading frames (smORFs) [37, 38]. Due to the vast nature of genomic and proteomic databases, molecular de-extinction of AMPs has thus far been driven by strategic exploration to unveil antimicrobials encoded in the genome [21] or encrypted within proteins [1, 18]. For instance, six β-defensins were predicted through computational analysis of genomes from extinct and critically endangered species. This prediction was based on intrinsic disulfide bonding patterns (Cys1–Cys5, Cys2–Cys4, and Cys3–Cys6), as well as the characteristic structural features of β-defensins [21], including three antiparallel β-strands and a right-handed α-helix. Two defensins, *Ad*-AvBD5 and *Ad*-AvBD10, were identified from *Anomalopteryx didiformis*, the New Zealand moa that became extinct approximately 600 years ago [39]. Three β-defensins (*Cs*-AvBD1, *Cs*-AvBD9, and *Cs*-AvBD10) were derived from *Cyanopsitta spixii*, Spix's macaw, which is endemic to Brazil. Additionally, one β-defensin was identified from *Diceros bicornis minor*, a critically endangered subspecies of the black

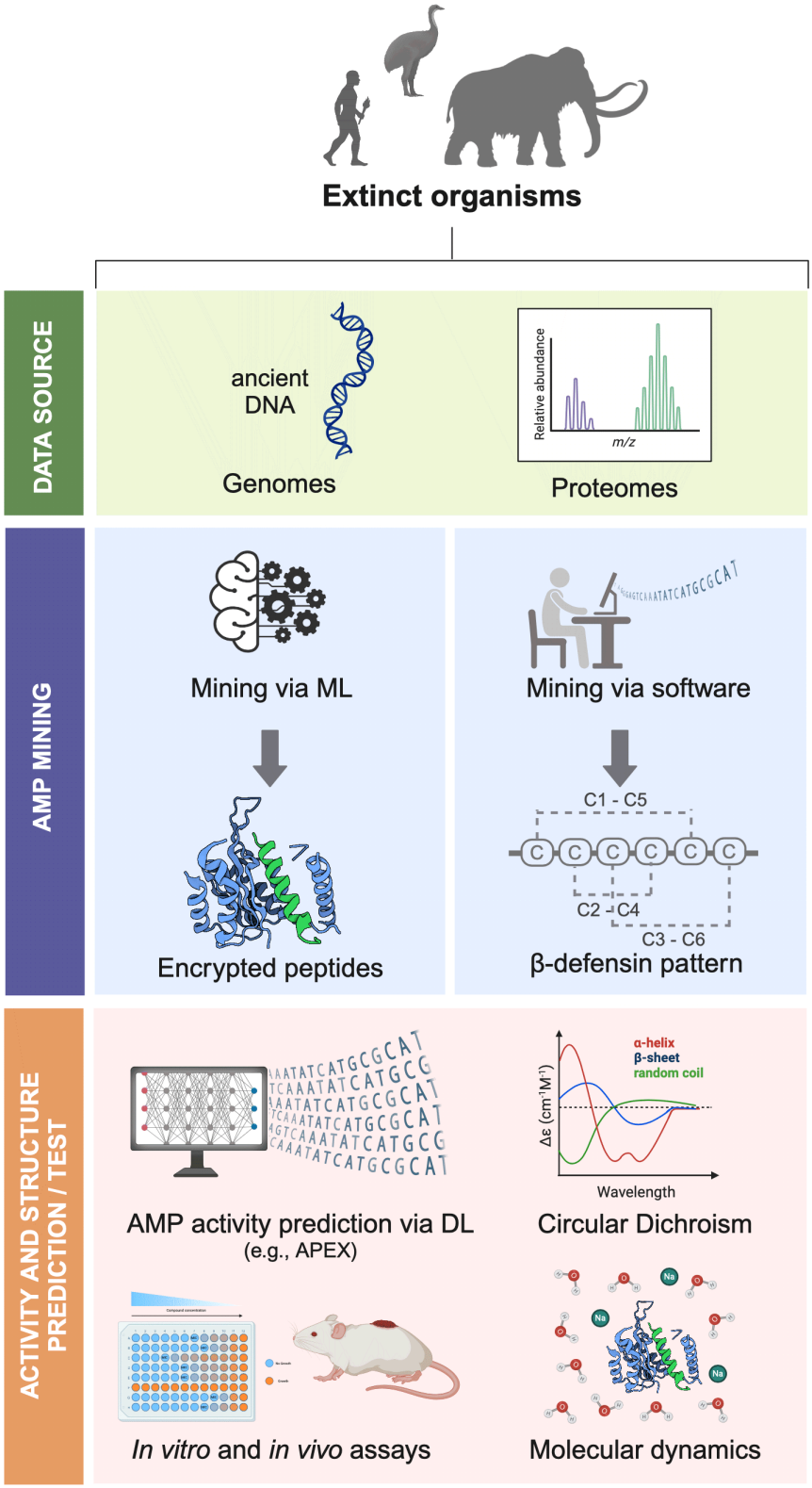

FIGURE 1 ● **Workflow for identifying ancient antimicrobial peptides.** Genome and proteome data serve as sources for mining ancient AMPs. This mining can be performed using (i) machine learning (ML) methods to generate encrypted peptides (EPs), or (ii) computational tools that identify defensins by analyzing disulfide bond patterns. The activity of these ancient AMPs can be predicted using deep learning (DL) models, and their structures can be elucidated through molecular dynamics simulations or circular dichroism. The predicted AMP activity can be validated through *in vitro* and *in vivo* assays.-Figure created with BioRender.com.

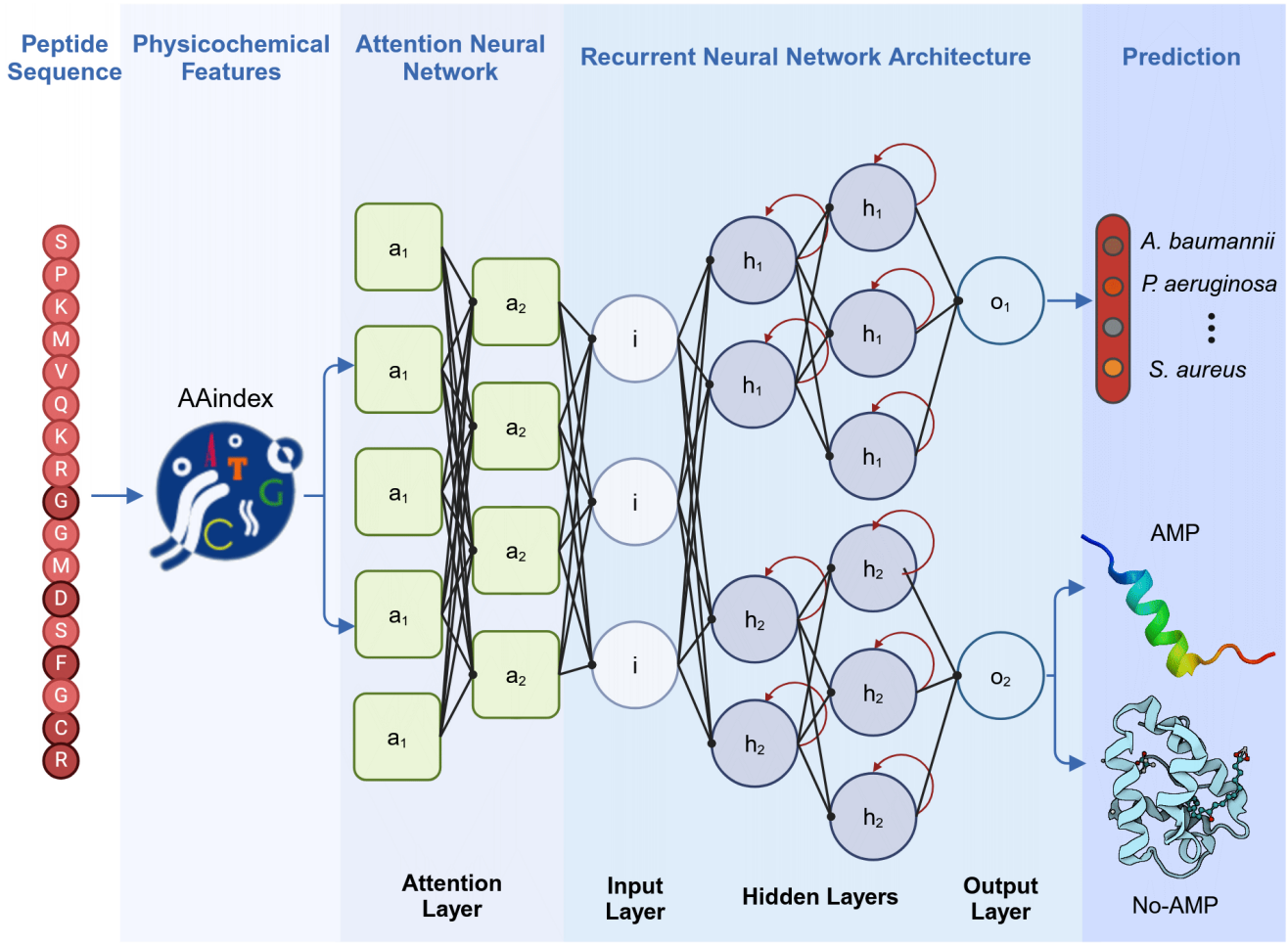

FIGURE 2 ● APEX Model Architecture. The APEX model combines recurrent and attention neural networks to analyze peptide sequences for antimicrobial prediction. The model first extracts physicochemical features of peptide sequences using the AAindex library. These features are processed through a two-layer attention neural network (a_1 and a_2), enhancing global feature interactions and compressing the representation into a lower-dimensional format. The output of this attention network is then processed by two separate recurrent neural networks (RNNs) (h_1 and h_2): one predicts species-specific antimicrobial activity (o_1), and the other classifies the peptide as antimicrobial (AMP) or non-antimicrobial (non-AMP) (o_2). Figure created with BioRender.com.

rhinoceros. *Ad*-AvBD5, *Cs*-AvBD1, and *Cs*-AvBD10 were noted for their high stability in molecular dynamics analyses, displaying cationic charges of +3, +7, and +2, respectively [21].

Using machine learning to predict patterns of protein cleavage into peptide fragments, we have detected encrypted peptides resulting from proteolysis in *Homo sapiens neanderthalensis* (Neanderthals) and *Homo sapiens* subsp. *Denisova* (Denisovans) [1]. Among the 69 archaic protein fragments identified, six showed *in vitro* antimicrobial activity, four from Neanderthals and two from Denisovans [1]. The molecule PDB6I34D-ALQ29, a fragment from Chain D of Neanderthal glycine decarboxylase, displayed broad-spectrum antimicrobial activity against both *Pseudomonas aeruginosa* and *Escherichia coli* strains, with MIC values ranging from 32 to 128 mmol.L^{-1}. This archaic encrypted peptide (AEP) possesses a net charge of +5 and an amphiphilicity index of 0.99. Conversely, compound A0A343EQH4-LAM11 (also known as neanderthalin), an AEP with a net charge of 0 and an amphiphilicity index of 0.63, demonstrated significant efficacy in preclinical animal models by reducing bacterial loads by several orders of magnitude against *Acinetobacter baumannii*. It was observed that AEPs have lower net charge and normalized hydrophobicity, more basic residues, and fewer acidic residues and polar residues compared to modern encrypted peptides (MEP) [1]. The differences in amino acid composition led to distinct physicochemical traits in AEPs, including lower amphiphilicity, a greater tendency toward disordered conformations, and reduced aggregation.

Moreover, the deep learning model, APEX, has facilitated the identification of several ancient encrypted peptides within the proteomes of extinct animals and plants [18]. Species such as the New Zealand moa (*Anomalopteryx didiformis*), the South American giant sloth (*Mylodon*

TABLE 1 ● Extinct peptides identified through molecular de-extinction.

Extinction	Species	Name	Peptides sequence	Amino acids	Net charge	Ref.
~600 years ago	*Anomalopteryx didiformis*	*Ad*-AvBD5	TRQDCESRGGFCSRGSCPLGITRIGICSLQDFCCRRKMGE	40	3	[21][a]
		Ad-AvBD10	VSFADTEECRSQGNFCRPVSCPPVFSVSGSCYGGAMKCCKKEYGQ	45	1	
Extinct in the wild in 2000	*Cyanopsitta spixii*	*Cs*-AvBD1	NKAQCHREKGFCALLKCPFPYVISGRCTKFTFCCKKGA	38	7	
		Cs-AvBD10	DPLFPDTTECKNQGNFCRAGTCPPTFAISGSCHGGLLRCCSKKISS	46	2	
		Cs-AvBD9	PAYSQVDADTAACRQNRGSCSFVECSSPMVNIGTCRSGKLKCCKXYV	47	3	
~40,000 years ago	*Homo sapiens neanderthalensis*	PDB6I34D-ALQ29	ALQLCYRHNKRRKFFVDPRCHPQTIAVVQ	29	5	[1][b]
		A0A384E0N4-DLI09	DLIERIQAD	9	-2	
		A0A343EQH4-LAM11	LAMVIPLWAGA	11	0	
		A0A343EQH0-NVK38	NVKMKWQFEHTKPTPFLPTLITLTTLLLPISPFMLMIL	38	2	
~50,000 years ago	*Homo sapiens subsp. 'Denisova'*	A0A343AZS4-FMA25	FMAEYTNIIMMNTLTTTIFLGTTYN	25	-1	
		A0A0S2IB02-AYT38	AYTTWNILSSAGSFISLTAVMLMIFMIWEAFASKRKVL	38	2	
1889	*Ara tricolor*	AWH62785.1-RLA27	RLATLQLWTINKITKQLMIPLNKPGHK	27	5	[8][c]
~781,000–30,000 years ago	*Elephas antiquus*	AQU14158.1-LHL12	LHLKILKIIRLL	12	3	
		AQU14158.1-IFL14	IFLHLKILKIIRLL	14	3	
1883	*Equus quagga boehmi*	ABN79624.1-CVL25	CVLLFSQLPAVKARGTKHRIKWNRK	25	7	
		ADN88909.1-RAY26	RAYICRKKFLSLRKASIKLQSLVRMK	26	9	
1875	*Hesperelaea palmeri*	CED79820.1-KLL26	KLLRKVLKETKKWVIKSVVFFKKIRK	26	10	
1768	*Hydrodamalis gigas*	AKN52354.1-LYC24	LYCRIYSLVRARGRRLTFRKNISK	24	8	
~103,000 to 42,000 years ago	*Lophiomys imhausi maremortum*	QYC36821.1-HWI16	HWITINTIKLSISLKI	16	2	
~6050–5050 BCE	*Mammut americanum*	ABQ86189.1-WMT15	WMTIHALKLSLSFKL	15	2	
~1.8 million to 12,000 years ago	*Mylodon darwinii*	AWK29290.1-WFH14	WFHFNSKILLLTGL	14	1	
		SMQ11516.1-KRK18	KRKRGLKLATALSLNNKF	18	6	
		SMQ11516.1-KIY25	KIYKKLSTPPFTLNIRTLPKVKFPK	25	7	
1952	*Pinguinus impennis*	ASB29243.1-KFI13	KFILNFKIPISFK	13	3	

[a] List of extinct β-defensins identified by software but not validated its activity experimentally.
[b] List of active AEPs identified by panCleave and validated experimentally.
[c] List of active AEPs identified by APEX and validated experimentally.

darwinii), the giant elk (*Megaloceros sp.*), Grant's zebra (*Equus quagga boehmi*), the woolly mammoth (*Mammuthus primigenius*), the straight-tusked elephant (*Elephas antiquus*), the ancient sea cow (*Hydrodamalis gigas*), as well as extinct plant species like *Magnolia latahensis* and *Hesperelaea palmeri*, revealed peptide molecules with excellent antimicrobial traits. These peptides exhibited low MIC values against some ESKAPE pathogens: *Enterococcus spp., Staphylococcus aureus, Klebsiella pneumoniae, A. baumannii, P. aeruginosa,* and *E. coli* [18]. Testing these encrypted peptides in two different preclinical mouse models indicated that mammuthusin-2 (MEP), elephasin-2 (AEP), and mylodonin-2 (AEP) possess high potential for antibiotic and anti-infective efficacy. Additionally, mammuthusin-2 exhibited slower degradation kinetics. Notably, although the AEPs and MEPs identified through APEX showed an atypical prevalence of low amphiphilicity and uncharged polar residues, they were primarily characterized by helical structures, resulting in more effective membrane disruption [18]. These findings highlight the potential of computational and artificial intelligence tools in uncovering extinct peptides with antimicrobial properties (**Table 1**).

CONCLUSION

Just as the evolutionary loss of AMPs is evident [40], microbial resistance genes also impose significant fitness costs on organisms due to their energy burdens, leading to the eventual loss of some of these genes over time [41]. Molecular de-extinction thus emerges as an innovative concept in drug discovery, offering the possibility of uncovering ancient molecules that may exhibit unique mechanisms of action or target sites not addressed by contemporary antibiotics. Reintroducing these ancient antimicrobials paves the way for exploring alternative therapeutic approaches to combat contemporary drug-resistant pathogens. Moreover, AI-driven molecular de-extinction has the potential to enrich our understanding of evolution, ecology, and biodiversity.

ACKNOWLEDGEMENTS

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF) and Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (Fundect). Cesar de la Fuente-Nunez holds a Presidential Professorship at the University of Pennsylvania, is a recipient of the Langer Prize by the AIChE Foundation, and acknowledges funding from the IADR Innovation in Oral Care Award, the Procter & Gamble Company, United Therapeutics, a BBRF Young Investigator Grant, the Nemirovsky Prize, Penn Health-Tech Accelerator Award, the Dean's Innovation Fund from the Perelman School of Medicine at the University of Pennsylvania, the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM138201, and the Defense Threat Reduction Agency (DTRA; HDTRA1-22-10031, HDTRA1-21-1-0014, and HDTRA1-23-1-0001). All figures were prepared in BioRender.

CONFLICT OF INTEREST

Cesar de la Fuente-Nunez provides consulting services to Invaio Sciences and is a member of the Scientific Advisory Boards of Nowture S.L., Peptidus, European Biotech Venture Builder, and Phare Bio. He is also a member of the Advisory Board for the Peptide Drug Hunting Consortium (PDHC). All other authors have no conflicts of interest to declare.

Please cite this article as: Karen O. Osiro, Abel G. Ley, Fabiano C. Fernandes, Kamila B. S. de Oliveira, Cesar de la Fuente-Nunez, Octavio L. Franco (**2025**). Paving the way for new antimicrobial peptides through molecular de-extinction. **Microbial Cell** 12: 1-8. doi: 10.15698/mic2025.02.841.

REFERENCES

1. Maasch JRMA, Torres MDT, Melo MCR, and de la Fuente-Nunez C (**2023**). Molecular de-extinction of ancient antimicrobial peptides enabled by machine learning. **Cell Host Microbe** 31(8): 1260-1274.e6. doi: 10.1016/j.chom.2023.07.001

2. Torrance AW, and de la Fuente-Nunez C (**2024**). The patentability and bioethics of molecular de-extinction. **Nat Biotechnol** 42(8): 1179–1180. doi: 10.1038/s41587-024-02332-x

3. Essel E, Zavala EI, Schulz-Kornas E, Kozlikin MB, Fewlass H, Vernot B, Shunkov M V., Derevianko AP, Douka K, Barnes I, Soulier M-C, Schmidt A, Szymanski M, Tsanova T, Sirakov N, Endarova E, McPherron SP, Hublin J-J, Kelso J, Pääbo S, Hajdinjak M, Soressi M, and Meyer M (**2023**). Ancient human DNA recovered from a Palaeolithic pendant. **Nature** 618(7964): 328–332. doi: 10.1038/s41586-023-06035-2

4. Hunter P (**2022**). Molecular archaeology and machine learning. **EMBO Rep** 23(6): e55315. doi: 10.15252/embr.202255315

5. Chen N, and Nedoluzhko A (**2023**). Ancient DNA: The past for the future. **BMC Genomics** 24(1): 23–25. doi: 10.1186/s12864-023-09396-0

6. Wong F, de la Fuente-Nunez C, and Collins JJ (**2023**). Leveraging artificial intelligence in the fight against infectious diseases. **Science** 381(6654): 164–170. doi: 10.1126/science.adh1114

7. Porto WF, Irazazabal L, Alves ESF, Ribeiro SM, Matos CO, Pires ÁS, Fensterseifer ICM, Miranda VJ, Haney EF, Humblot V, Torres MDT, Hancock REW, Liao LM, Ladram A, Lu TK, de la Fuente-Nunez C, and Franco OL (**2018**). In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design. **Nat Commun** 9(1): 1490. doi: 10.1038/s41467-018-03746-3

8. Wan F, Wong F, Collins JJ, and de la Fuente-Nunez C (**2024**). Machine learning for antimicrobial peptide identification and design. **Nat Rev Bioeng** 2(5): 392–407. doi: 10.1038/s44222-024-00152-x

9. Torres MDT, and de la Fuente-Nunez C (**2019**). Toward computer-made artificial antibiotics. **Curr Opin Microbiol** 51: 30–38. doi: 10.1016/j.mib.2019.03.004

10. Islam S, Akhand MRN, and Hasan M (**2023**). Evolutionary trend of bovine β-defensin proteins toward functionality prediction: A domain-based bioinformatics study. **Heliyon** 9(3): e14158. doi: 10.1016/j.heliyon.2023.e14158

11. Browne K, Chakraborty S, Chen R, Willcox MD, Black DS, Walsh WR, and Kumar N (**2020**). A New Era of antibiotics: The clinical potential of antimicrobial peptides. **Int J Mol Sci** 21(19): 7047. doi: 10.3390/ijms21197047

12. Cardoso P, Glossop H, Meikle TG, Aburto-Medina A, Conn CE, Sarojini V, and Valery C (**2021**). Molecular engineering of antimicrobial peptides: Microbial targets, peptide motifs and translation opportunities. **Biophys Rev** 13(1): 35–69. doi: 10.1007/s12551-021-00784-y

13. Catte A, Wilson MR, Walker M, and Oganesyan VS (**2018**). Antimicrobial action of the cationic peptide, chrysophsin-3: A coarse-grained molecular dynamics study. **Soft Matter** 14(15): 2796–2807. doi: 10.1039/C7SM02152F

14. de la Fuente-Nunez C, Cesaro A, and Hancock REW (**2023**). Antibiotic failure: Beyond antimicrobial resistance. **Drug Resist Updat** 71: 101012. doi: 10.1016/j.drup.2023.101012

15. Koehbach J, and Craik DJ (**2019**). The vast structural diversity of antimicrobial peptides. **Trends Pharmacol Sci** 40(7): 517–528. doi: 10.1016/j.tips.2019.04.012

16. Osiro KO, Hashemi N, Brango-Vanegas J, Oliveira SMD, and Franco OL (**2024**). Emerging peptide-based technology for biofilm control. **Expert Opin Biol Ther** 24(12): 1311–1315. doi: 10.1080/14712598.2024.2430623

17. Cresti L, Cappello G, and Pini A (**2024**). Antimicrobial peptides towards clinical application—A long history to be concluded. **Int J Mol Sci** 25(9): 4870. doi: 10.3390/ijms25094870

18. Wan F, Torres MDT, Peng J, and de la Fuente-Nunez C (**2024**). Deep-learning-enabled antibiotic discovery through molecular de-extinction. **Nat Biomed Eng** 8(7): 854–871. doi: 10.1038/s41551-024-01201-x

19. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, and Karsch-Mizrachi I (**2020**). NCBI Taxonomy: A comprehensive update on curation, resources and tools. **Database** 2020: baaa062. doi: 10.1093/database/baaa062

20. de la Fuente-Nunez C (**2024**). AI in infectious diseases: The role of datasets. **Drug Resist Updat** 73: 101067. doi: 10.1016/j.drup.2024.101067

21. Ferreira AFL, Osiro KO, de Oliveira KBS, Cardoso MH, de Lima LR, Duque HM, Macedo MLR, Landon C, de la Fuente-Nunez C, and Franco OL (**2024**). Defensins identified through molecular de-extinction. **Cell Rep Phys Sci** 5(9): 102193. doi: 10.1016/j.xcrp.2024.102193

22. Hoff KJ, and Stanke M (**2013**). WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. **Nucleic Acids Res** 41(W1): W123–W128. doi: 10.1093/nar/gkt418

23. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, and Finn RD (**2018**). HMMER web server: 2018 update. **Nucleic Acids Res** 46(W1): W200–W204. doi: 10.1093/nar/gky448

24. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, and Bateman A (**2023**). InterPro in 2022. **Nucleic Acids Res** 51(D1): D418–D427. doi: 10.1093/nar/gkac993

25. Jumper J et al. (**2021**). Highly accurate protein structure prediction with AlphaFold. **Nature** 596(7873): 583–589. doi: 10.1038/s41586-021-03819-2

26. Abramson J et al. (**2024**). Accurate structure prediction of biomolecular interactions with AlphaFold 3. **Nature** 630(8016): 493–500. doi: 10.1038/s41586-024-07487-w

27. Torres MDT, Melo MCR, Flowers L, Crescenzi O, Notomista E, and de la Fuente-Nunez C (**2022**). Mining for encrypted peptide antibiotics in the human proteome. **Nat Biomed Eng** 6(1): 67–75. doi: 10.1038/s41551-021-00801-1

28. Li J, Qu X, He X, Duan L, Wu G, Bi D, Deng Z, Liu W, and Ou H-Y (**2012**). ThioFinder: A web-based tool for the identification of thiopeptide gene clusters in DNA sequences. **PLoS One** 7(9): e45878. doi: 10.1371/journal.pone.0045878

29. Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai H-C, Zakai UI, and Mitchell DA (**2017**). A new genome-mining tool redefines the lasso peptide biosynthetic landscape. **Nat Chem Biol** 13(5): 470–478. doi: 10.1038/nchembio.2319

30. Agrawal P, Amir S, Deepak, Barua D, and Mohanty D (**2021**). RiPP-Miner-Genome: A web resource for automated prediction of crosslinked chemical structures of RiPPs by genome mining. **J Mol Biol** 433(11): 166887. doi: 10.1016/j.jmb.2021.166887

31. Kumar N, Bhagwat P, Singh S, and Pillai S (**2024**). A review on the diversity of antimicrobial peptides and genome mining strategies for their prediction. **Biochimie** 227(Pt A): 99-15. doi: 10.1016/j.biochi.2024.06.013

32. Lee H, Lee S, Lee I, and Nam H (**2023**). AMP-BERT: Prediction of antimicrobial peptide function based on a BERT model. **Protein Sci** 32(1): e4529. doi: 10.1002/pro.4529

33. Li C, Sutherland D, Hammond SA, Yang C, Taho F, Bergman L, Houston S, Warren RL, Wong T, Hoang LMN, Cameron CE, Helbing CC, and Birol I (**2022**). AMPlify: Attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens. **BMC Genomics** 23(1): 77. doi: 10.1186/s12864-022-08310-4

34. Hazlett L, and Wu M (**2011**). Defensins in innate immunity. **Cell Tissue Res** 343(1): 175–188. doi: 10.1007/s00441-010-1022-4

35. Pizzo E, Cafaro V, Di Donato A, and Notomista E (**2018**). Cryptic antimicrobial peptides: Identification methods and current knowledge of their immunomodulatory properties. **Curr Pharm Des** 24(10): 1054–1066. doi: 10.2174/1381612824666180327165012

36. Izoré T, Candace Ho YT, Kaczmarski JA, Gavriilidou A, Chow KH, Steer DL, Goode RJA, Schittenhelm RB, Tailhades J, Tosin M, Challis GL, Krenske EH, Ziemert N, Jackson CJ, and Cryle MJ (**2021**). Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. **Nat Commun** 12(1): 2511. doi: 10.1038/s41467-021-22623-0

37. Torres MDT, Brooks EF, Cesaro A, Sberro H, Gill MO, Nicolaou C, Bhatt AS, and de la Fuente-Nunez C (**2024**). Mining human microbiomes reveals

an untapped source of peptide antibiotics. **Cell** 187(19): 5453-5467.e15. doi: 10.1016/j.cell.2024.07.027

38. Santos-Júnior CD, Torres MDT, Duan Y, Rodríguez del Río Á, Schmidt TSB, Chong H, Fullam A, Kuhn M, Zhu C, Houseman A, Somborski J, Vines A, Zhao X-M, Bork P, Huerta-Cepas J, de la Fuente-Nunez C, and Coelho LP (**2024**). Discovery of antimicrobial peptides in the global microbiome with machine learning. **Cell** 187(14): 3761-3778.e16. doi: 10.1016/j.cell.2024.05.013

39. Bunce M, Worthy TH, Phillips MJ, Holdaway RN, Willerslev E, Haile J, Shapiro B, Scofield RP, Drummond A, Kamp PJJ, and Cooper A (**2009**). The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. **Proc Natl Acad Sci USA** 106(49): 20646–20651. doi: 10.1073/pnas.0906660106

40. Hanson MA, Lemaitre B, and Unckless RL (**2019**). Dynamic evolution of antimicrobial peptides underscores trade-offs between immunity and ecological fitness. **Front Immunol** 10: 2620. doi: 10.3389/fimmu.2019.02620

41. Kim M, Park J, Kang M, Yang J, and Park W (**2021**). Gain and loss of antibiotic resistant genes in multidrug resistant bacteria: One Health perspective. **J Microbiol** 59(6): 535–545. doi: 10.1007/s12275-021-1085-9