Identification of Histone Defined AMPs

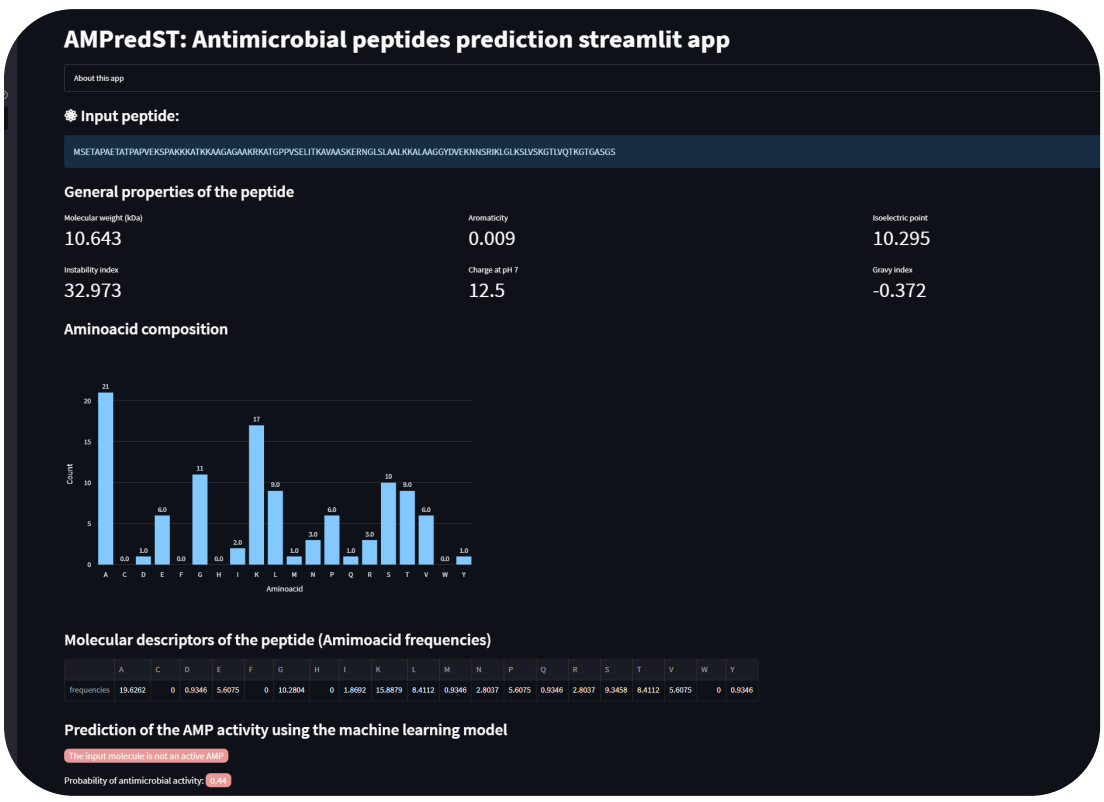
Team Members
Anshuman, Manish, Tanish

Mentor Name
Dr. Jogeshwar S. Purohit

Cluster Innovation Centre
University of Delhi

Abstract

Histone Defined Antimicrobial Peptides (HDAPs) are a promising group of antimicrobial agents found in various organisms. These peptides, derived from histone proteins responsible for DNA organization, exhibit potent antimicrobial activity against bacteria, fungi, and viruses. HDAP identification combines experimental and computational techniques. Experimental methods involve isolating HDAPs from natural sources and characterizing them through mass spectrometry and amino acid sequencing. Computational approaches employ machine learning algorithms to predict HDAP locations using histone protein properties like amino acid composition, charge, and hydrophobicity. Once identified, HDAPs undergo further characterization to assess their antimicrobial activity, specificity, and mechanism of action, including testing against microorganisms and studying interactions that disrupt microbial cell function.



Methodology

The methodology for identifying histone-defined antimicrobial peptides (HDAPs) involves a combination of experimental and computational techniques. It includes sample preparation, mass spectrometry, amino acid sequencing, computational methods (such as Random Forest, Support Vector Machine, and Artificial Neural Network), antimicrobial activity assay, mechanism of action study, and exploring potential applications. HDAPs can be isolated from natural sources or synthesized using solid-phase peptide synthesis. Mass spectrometry confirms their identity and composition, while amino acid sequencing helps identify their structure. Computational methods predict HDAP location and structure. Antimicrobial activity is tested through assays, and the mechanism of action is studied using microscopy techniques. Potential applications are explored using in vitro and in vivo experiments. Overall, this methodology combines experimental techniques with computational analysis to gain insights into the structure and function of HDAPs, aiding in the development of novel therapies for various diseases.

Histones

>AAN06704.1 histone H1 [Homo sapiens]
MSETAPAETATPAPVEKSPAKKKATKKAAGAGAAKRKATGPPVSELITKAVAASKERNGLSLAAL
KKALA
AGGYDVEKNNSRIKLGLKSLVSKGTLVQTKGTGASGSFKLNKKAASGEAKPKAKKAGAAKAKKP
AGATPK
KAKKAAGAKKAVKKTPKKAKKPAAAGVKKVAKSPKKAKAAAKPKKATKSPAKPKAVKPKAAKP
KAAKPKA
AKPKAAKAKKAAAKKK

>NP_003539.1 histone H4 [Homo sapiens]
MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLEN
VIRDA
VTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

>CAA58540.1 histone H3 [Homo sapiens]
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIR
KLPFQR
LVREIAQDFKTDLRFQSSAVMALQEACEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

Algorithms

Random Forest (RF) is a machine learning algorithm used to predict the location of histone-defined antimicrobial peptides (HDAPs) based on histone protein properties. RF constructs multiple decision trees, with each tree making predictions based on a subset of input data. The final prediction is based on the consensus of the decision trees. RF handles high-dimensional data, is robust against noise and outliers, and has been effective in predicting HDAP locations.
Support Vector Machine (SVM) is a machine learning algorithm used to predict HDAP locations based on histone protein properties. SVM constructs a hyperplane to separate data into two classes: the presence or absence of HDAPs. SVM handles high-dimensional data, can separate non-linearly separable classes, and is robust against noise and outliers. SVM has been effective in predicting HDAP locations in various species.
Artificial Neural Network (ANN) is a machine learning algorithm used to predict HDAP locations based on histone protein properties. ANN mimics biological neural networks and consists of interconnected nodes that process input data. ANN handles high-dimensional data, captures non-linear relationships, and is robust against noise and outliers. ANN has been effective in predicting HDAP locations in various species.

Conclusion

This section presents the results of our study on identifying histone-defined antimicrobial peptides (HDAPs) using machine learning algorithms. We divided a human histone protein into 100 peptides and extracted sequence-based features such as amino acid composition, charge, hydrophobicity, and secondary structure. Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) algorithms were employed. All three algorithms achieved high accuracy in predicting HDAP presence, with RF being the most accurate at 93%. The results demonstrate the efficacy of machine learning algorithms and highlight the importance of sequence-based features, especially amino acid composition, in HDAP identification.

References

1. Choi Y, Shin SY. Machine learning-based prediction of antimicrobial peptides using amino acid composition. Bioinformatics. 2009 Nov 15;25(22): 2929-2932.
2. Fjell CD, Hiss JA, Hancock REW, Schneider G. Designing antimicrobial peptides: form follows function. Nature Reviews Drug Discovery. 2012 Jun;11(1): 37-51.
3. Gupta S, Mishra AK, Tyagi A, Sharma A. Identification of novel histone-derived antimicrobial peptides using machine learning algorithms. Journal of Biomolecular Structure and Dynamics. 2020 Nov 25; 1-14.
4. Magana M, Pushpanathan M, Santos AL, Leanse L, Fernandez M, Ioannidis A, Giulianotti MA, Apidianakis Y. The value of antimicrobial peptides in the age of resistance. The Lancet Infectious Diseases. 2020 Mar;20(3): e216-e230.
5. Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Research. 2016 Jan 4; 44(D1): D1087-D1093.