P20

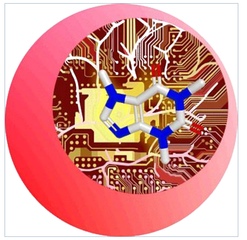
MACHINE LEARNING GUIDED GENETIC ALGORITHM FOR THE DISCOVERY OF NOVEL ANTIMICROBIAL PEPTIDES

MARKO NJIRJAK [1], ERIK OTOVIĆ [1], DANIELA KALAFATOVIĆ [2,3], GORAN MAUŠA [1,3]

[1] FACULTY OF ENGINEERING, UNIVERSITY OF RIJEKA; *MNJIRJAK@RITEH.HR, EOTOVIC@RITEH.HR, GMAUSA@RITEH.HR*

[2] DEPARTMENT OF BIOTECHNOLOGY, UNIVERSITY OF RIJEKA; *DANIELA.KALAFATOVIC@UNIRI.HR*

[3] CENTER FOR ARTIFICIAL INTELLIGENCE AND CYBERSECURITY, UNIVERSITY OF RIJEKA



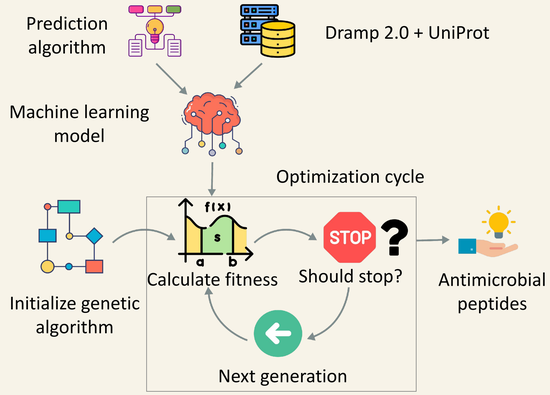
INTRODUCTION

- Chemical search space grows exponentially with peptide length
- Principles that govern the activity of short peptides at the sequence level are unknown
- Advancements in artificial intelligence represent a novel way to explore chemical space
- Machine learning is increasingly used to address various challenges of peptide chemistry

GOAL

- **General goal:** In silico discovery of de novo peptides with antimicrobial activity
- **First research question:** How does change in amino acid composition affect peptide activity?
- **Second research question:** Can a genetic algorithm construct novel compounds with desired properties? If so, in what timespan?

METHODOLOGY



1. Train a machine learning model to detect antimicrobial peptides
1. Use the model as a fitness function of the genetic algorithm
1. Genetic algorithm performs a directed search of chemical space and finds novel antimicrobial peptides

RESULTS

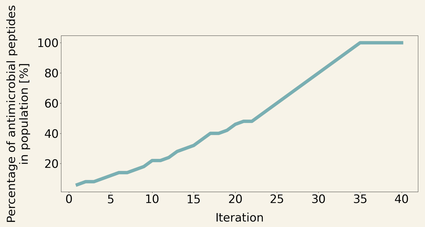
- Machine learning model for the detection of antimicrobial peptides

Model	Accuracy	Precision	Recall	F_1	AUC
Random Forest	88.9%	88.3%	87.0%	87.6%	88.8%

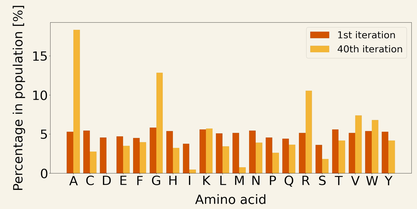
- The algorithm finds novel antimicrobial peptides in approximately 30 minutes

Peptide	Probability of AM activity	$CAMP_{R3}$ RF classifier
ITIVPKKCKLLL	90%	96%
GTWAMDYGGPNLCRYFLNAKCWKCKRKHV	83%	81%
LQVLCEKKCAKKVRTCSK	82%	87%

- As the algorithm progresses, an increasing percentage of the population is made up of antimicrobial peptides



- Changes in peptide composition could indicate which amino acids are important for antimicrobial activity



FUTURE WORK

- Discovery of features important for activity
- Development of a decision support system for the design of new active peptides
- Experimental evaluation of the results
- Enhance algorithms to efficiently process sequences larger than 50 amino acids, and thus enable exploration of broader chemical spaces

PUBLICATIONS

- E. Otović, M. Njirjak, I. Žužić, D. Kalafatović, G. Mauša; *Genetic Algorithm Parametrization for Informed Exploration of Short Peptides Chemical Space*, Proceedings of SoftCOM 2020, pp. 1–3
- D. Kalafatović, G. Mauša, T. Todorovski, E. Giralt; *Algorithm-supported, mass and sequence diversity-oriented random peptide library design*, Journal of cheminformatics, Vol. 11 (25), pp. 1–15, 2019

ACKNOWLEDGEMENTS

- **Design of short catalytic peptides and peptide assemblies** (Deshpet, grant no. UIP-2019-04-7999)
- **SARS-CoV-2 supramolecular mimetics for discovery of peptides that induce viral entrapment**
- **Applying Machine Learning for the Discovery of Peptides with Catalytic Activity** (grant no. uniri-pr-tehnic-19-10)
- Funded by:

HRZZ Croatian Science Foundation

This cover has been designed using resources from Flaticon.com

RSC INTEREST GROUP
CHEMICAL INFORMATION AND COMPUTER APPLICATIONS

RSC INTEREST GROUP
BIOLOGICAL AND MEDICINAL CHEMISTRY SECTOR