APPLYING MACHINE LEARNING FOR THE DISCOVERY OF PEPTIDES WITH CATALYTIC ACTIVITY

FACULTY OF ENGINEERING, UNIVERSITY OF RIJEKA

INTRODUCTION

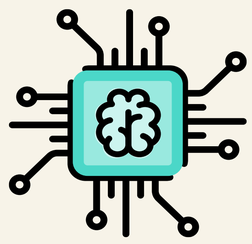
- Chemical search space grows exponentially with peptide length
- Principles that govern the activity of short peptides at the sequence level are unknown
- Advancements in optimization algorithms represent a novel way to conduct chemical space exploration
- Machine learning is increasingly used to address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

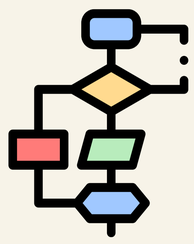
OBJECTIVES

1. Reliable dataset of active peptides and their physicochemical properties
2. In silico design of multiple peptide libraries that cover greater area of chemical space
3. Prediction model for the discovery of de novo peptides with high activity level

METHODS

- **Machine learning**
  - Artificial neural networks
  - Random forest
  - SVM
- **Metaheuristic optimization**
  - Simulated annealing
  - Evolutionary algorithms
    - NSGA-II
- **Peptide features calculation**
  - Hydrophobicity
  - Molecular weight
  - Isoelectric charge



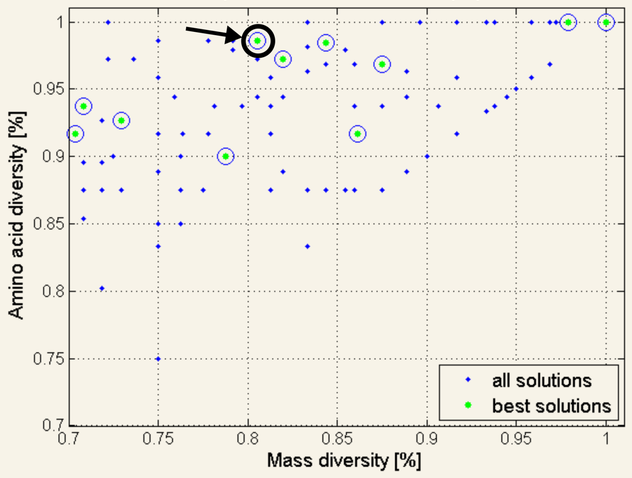


DESHPET WEB

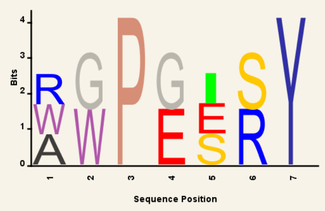
- **Online web service**
  - Genetic algorithm supported search of short peptides chemical space
  - Single and multi-library options

RESULTS

- Multi-objective evolutionary approach for the design of mass and sequence diversity-oriented random peptide libraries

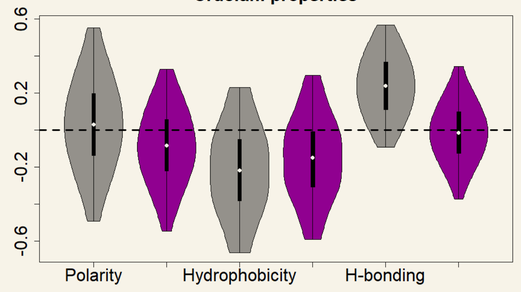


Amino acid diversity = 98%
Mass diversity = 81%



All = 72
Unique by sequence = 71
Unique by mass = 58

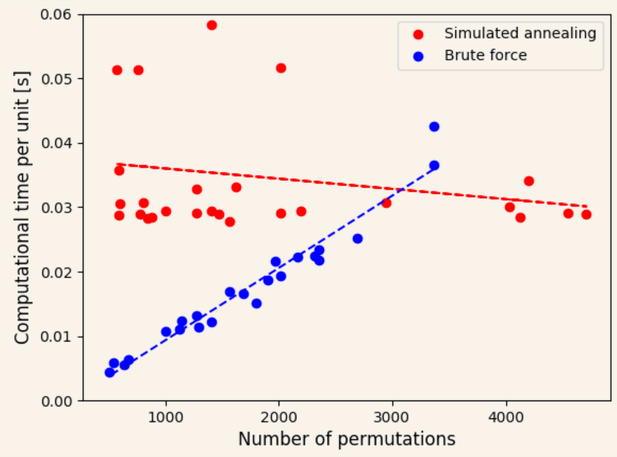
- Feature computation and visualization



- Prediction of peptide antiviral activity

Model	Acc	Prec	Sens	F_1	AUC
Random Forest	82.4%	82.9%	88.7%	82.4%	88.7%

- Simulated annealing was used when the number of combinations was too large for brute force approach



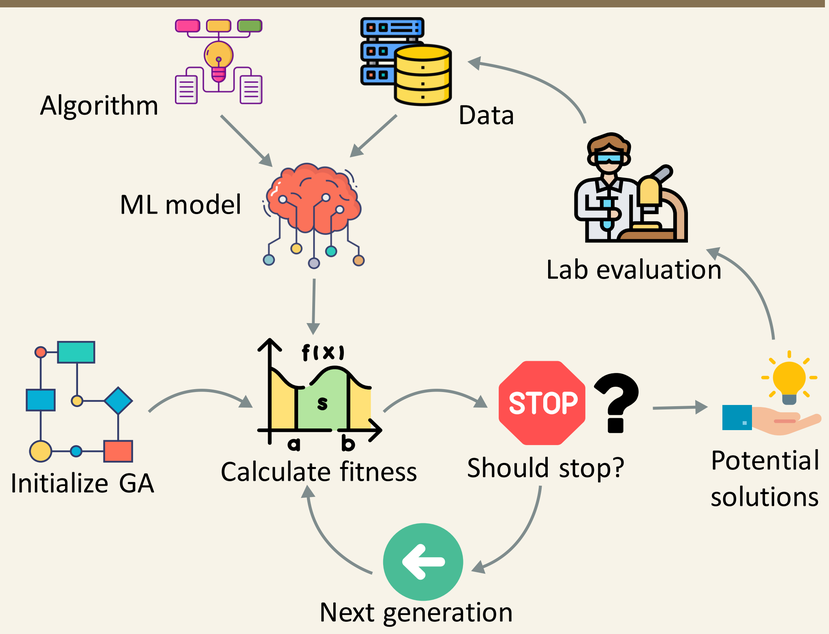
RELATED PROJECTS

- **Design of short catalytic peptides and peptide assemblies** (Deshpet, grant no. UIP-2019-04-7999)
  - PI: Daniela Kalafatović
  - Funded by: Croatian Science Foundation

- **SARS-CoV-2 supramolecular mimetics for discovery of peptides that induce viral entrapment**
  - PI: Daniela Kalafatović
  - Substitute PI: Goran Mauša
  - Funded by: University of Rijeka

FUTURE WORK

- Creation of a catalytic peptides database
- Building of a knowledge-based prediction model
- Discovery of features important for activity
- Development of a decision support system for the design of new active peptides



PUBLICATIONS

- D. Kalafatović, G. Mauša, D. Rešetar Maslov, E. Giralt; *Bottom-Up Design Approach for OBOC Peptide Libraries*, **Molecules**, Vol. 25 (15), pp. 1–15, 2020
- 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. Kalafatovic, 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

TEAM

- Dr. Goran Mauša (PI)
- Dr. Daniela Kalafatović
- Dr. Damir Arbula
- Erik Otović
- Marko Njirjak

FUNDED BY