ItPS Seminars

17th May 2024
3pm CEST

DANIELA KALAFATOVIC
Assistant Professor
University of Rijeka
Rijeka Croatia

Machine Learning-guided peptide discovery

Join the Seminar

Italian Peptide Society

www.italianpeptidesociety.it

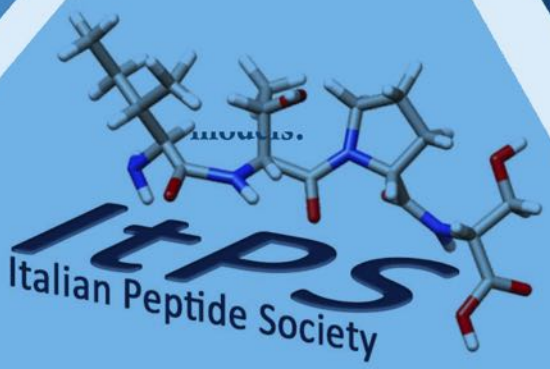
ItPS Seminars

17th May 2024 3pm CEST

Daniela Kalafatovic
Assistant Professor
University of Rijeka

Machine Learning-guided peptide discovery

The discovery of new peptides (i.e., antimicrobial, antiviral, catalytic) is challenging, as they are part of a very large search space and the principles responsible for the desired activities at the sequence level are not yet fully understood. To avoid expensive and time-consuming guesswork and experimental failure, our strategy is to apply soft computing techniques to accelerate peptide discovery. Search-based algorithms allow for a faster exploration of peptide permutation space which grows exponentially with peptide length and whose amount and dimensionality is too overwhelming to rationally comprehend. Machine learning can find patterns or regularities in data, build mathematical models based on the theory of statistics and make up for the lack of knowledge. To date, both strategies have been applied to a variety of chemical problems to maximize the chance of successful and rapid solving of complex issues. One of the challenges in applying ML to peptides is the lack of large, balanced, and well-structured datasets. Another challenge is a variety of representation schemes with different levels of information and complexity that can be applied and that yield different prediction results. For this purpose, we developed a new sequential properties representation scheme that combines physico-chemical properties with the amino acid order within the sequence. We also used SMILES with optimized data preprocessing step and both representation schemes led to successful prediction models. Furthermore, we developed a flexible and adaptive model based on ML-driven genetic algorithm that allows for a directed search of the sequence space by promoting a desired property. This enabled the discovery of sequences in unexplored regions of the peptide space, which were confirmed using MD simulations or experimentally creating new knowledge that can be used to amplify the existing datasets and prediction models.



www.italianpeptidesociety.it

Join the Seminar