# Identification of Novel Antimicrobial Peptides with Designed Activity through a QSAR Based Machine Learning Model

Luo, Fiona (School: Monta Vista High School)

Antimicrobial peptides (AMPs) are a growing field of therapeutics with lower risks of antimicrobial resistance compared to conventional antibiotics. However, experimentally testing peptides is costly and inefficient, making a preliminary in-silico screening more efficient. Our goal is to design a machine-learning program to predict the activity of AMPs, and to determine novel AMP drug leads. We present a novel model structure for predicting AMP activity by combining a QSAR (quantitative structure-activity relationship) model and an LSTM (long short-term memory) model. We built a QSAR neural network to predict probability of peptide activity using 29 calculated physicochemical descriptors as a representation of each sequence. We then used a generative LSTM network to sample 10,000 promising de novo sequences and validated the samples' activity with the QSAR model. We chose samples predicted to have over 99% probability of activity and ran them through a protein secondary structure prediction server (CABS-fold) to analyze structure. Finally, we applied our program to anti-HIV sequences to generate novel anti-HIV drug leads. Our QSAR model achieved an accuracy of 92.60% on AMP prediction and 81.67% on anti-HIV prediction. We determined 69 novel anti-HIV drug leads and 707 antimicrobial leads, largely predicted to be alpha-helices, that can be further tested in vitro. Overall, our project improves on prior research in several ways. Our model is flexible and can be further applied to AMPs of designed activities (e.g. antiviral, antifungal, antimalarial), and we provide the specific example of anti-HIV peptides. To our best knowledge, we are the first machine learning program to predict anti-HIV activity, and the first QSAR and LSTM combined model which analyzes AMPs.

**Awards Won:**