Staphylococcus aureus Biofilm Dispersion: Computationally Analyzing Interactions between Nattokinase Binding-Partners via a Novel Stacked Generalization Machine Learning Approach

Sriram, Raghav (School: Carmel High School)

Nattokinase is a serine protease that is used to treat Staphyloccocus aureus biofilm formation on surgical tools. Despite preliminary knowledge on the chemical structure of nattokinase, there is a desperate need to understand the specific interactions occurring during nattokinase facilitated biofilm dispersion. Thus, determining specific residues involved in interactions between nattokinase and its binding partners provides further insights behind the protein-mediated interactions during nattokinase-aided biofilm dispersion. This project utilized a variety of technologies including RGN-Protein Modeling software developed by AQ Laboratories to model protein interactions. Through Stacked Generalization this software coupled with 15 individual machine learning algorithms, each utilizing either Random Forest or SVM regression models established a higher-level model with a confidence interval of 93.2%. Afterward, parameters for nattokinase binding sites were determined through a Matplotlib linear regression model. Results indicate nattokinase as a heparin-binding protein with an affinity of ~217 nM. Nattokinase binding interactions occurred for binding partners who displayed the presence of N-sulfo groups, 3-O-sulfo, and 6-O-desulfo groups, but not 2-O-sulfo groups (which was previously believed to be involved), present. Hydrogen bonds formed between residues D60, S33 S62, and T220 which stabilized the binding site, and G127, L126, and S125 often served as the substrate binding sites. Cross-referencing with surface proteins yielded no matches, however, a peptidoglycan-repeat unit consisting of the GlcNAc-MurNAc disaccharide showed promise. Further experimentation will involve modeling lumbrokinase, another commonly used enzyme for biofilm dispersion.