Staphylococcus *aureus* Biofilm Dispersion: Computationally Analyzing Interactions between Nattokinase Binding-Partners via Stacked Generalization

Raghav Sriram, Carmel High School, Carmel, IN USA

Background and Research Question

- **Staphylococcus *aureus*** is a gram-positive round-shaped bacterium that commonly forms on surgical devices and is the leading cause of soft tissue infection
- **Nattokinase,** an enzyme produced by nattōkin, has been used as a treatment option for S. *aureus* biofilms. However, the mechanism remains largely unknown
- **Stacked Generalization** allows for the creation of higher-level models from low-level models
- By determining **residues** necessary for nattokinase binding we can determine surface protein structure of biofilms
- Through this we can determine the n**attokinase-aided biofilm dispersion pathway**

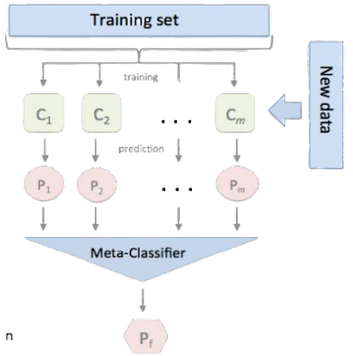

Figure 1 | Diagram describing the components and process involved in a Stacked Generalization ensemble algorithm

Data Analysis and Results

	Accuracy	Precision
STKD-1	0.897	0.912
STKD-7	0.832	0.789
STKD-12	0.913	0.925
STKD-15	0.883	0.893
MSTR	0.932	0.915

Table 1 | Precession and accuracy of selected low-level (STKD) and higher level (MSTR) models

- Higher-level model aggregated via stack generalization had an accuracy of 0.932 and precession of 0.915
- Lower-level **Random Forest** models proved to be more effective than lower-level S**VM models**
- 90% of nattokinase binding sites had N-sulfo groups and 80% had 6-*O*-desulfo groups present

Methodology

1.) **Data Concatenation -** Combine publicly accessible data on binding interactions between known nattokinase binding partners from KLIFS, BioGRID, and RCSB datasets

2.) **Training -** Nattokinase binding interactions are fed as training data to lower-level Random Forest and SVM models

3.) **Generate Interactions -** Binding site interactions are modeled via AQ Laboratory RGN-Protein Modeling Software. Top ten most probable interactions from each lower-level modem move on to the meta-classifier to allow Stacked Generalization

4.) **Binding Parameters -** After higher-level model achieved an accuracy greater than 90%, necessary residues and chemical groups for nattokinase binding sites were established via a linear regression Matplot

5.) **Surface Protein Modeling -** Potential non-experimentally determined binding partners with known protein structures found at the surface of biofilms following nattokinase binding parameters were generated through logistic regression

Interpretation and Conclusions

- Nattokinase is a **heparin-binding protein** with an affinity of ~217 nM
- NK binding percentages suggest nattokinase binding partners must have *N*-sulfo groups and 6-*O*-desulfo groups, but not 2-*O*-desulfo groups
- Hydrogen bonds formed between D^{60}, S^{33} S^{62}, and T^{220} to **stabilize binding site** and G^{127}, L^{126}, and S^{125} served as **substrate binding sites**
- No cross referencing matches with known structures of biofilm surface proteins
- **Surface Peptidoglycan-repeat unit** could serve as a potential binding site
  - Contains S^{125} residue and 6-*O*-desulfo chemical group
  - 79.3% confidence in NK Binding
- Further experimental research required to determine relevance

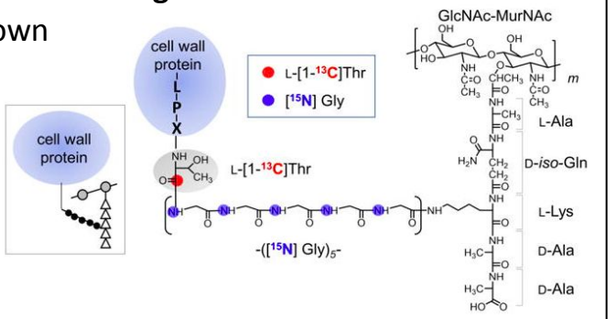

Figure 2 | Chemical structure of S. *aureus* surface peptidoglycan-repeat unit. Potential initiator of the nattokinase biofilm dispersion cascade