Conclusions & Outlook

We have shown that high-level secretion of homologous protein can be accurately predicted using amino acid composition. SVMs trained on both homologous and heterologous proteins were found to be similar, showing that the same amino acids are important in both cases, i.e. general properties are defining for successful high-level secretion. However, compared to the homologous data set, a larger overlap between the positives and negatives in the heterologous data set causes a lower prediction performance.

As a next step we aim to redesign proteins to improve secretion by adjusting the amino acid composition, thereby moving a protein to the other side of the classifier boundary, while maintaining the same protein structure (Fig.3).