# Introduction

* ML has been used extensively to answer the protein classification problem
* Discriminative approach introduction supervised learning paradigm
* Different feature extraction methods
* Kernel methods – spectrum kernel
* SVM with kernel previous examples

# Methods

We have used three different feature representation techniques in our formulation which resulted in a diverse set of individual models required by the ensemble to capture various trends in our dataset. Each of the techniques is described as follows:

## k-mer motif builder

## GAAC based k-mer motif builder

## Binding pocket specific positional model

Our base model creates a PCA based encoding of the feature space and then uses SVM to predict the enzyme classes.