

Lysine Phosphoglycerylation

The table below shows the comparison of PhoglyStruct with simpler set of features for the same 10-fold cross-validation set. Simpler_Features are features for each lysine taking into account the occurrence of amino acids in the ± 2 residue window. The matrix representing each lysine is of the size 20x5 where 20 is the different types of amino acid in the genome and 5 is the 2 upstream, the lysine and 2 downstream amino acid. Out of the 20 amino acids, a value of 1 is given to the amino acid present at that position otherwise the rest are given a value of 0. Therefore, the feature representing each lysine is a 100-dimensional vector. As it can be seen, PhoglyStruct, which uses the structural properties of amino acids, outperforms this simpler set of features.

Method	Sensitivity	Specificity	G-Mean	Accuracy	MCC	F-Measure	AUC
Simpler_Features	0.6470	0.7490	0.6887	0.7233	0.3612	0.5326	0.6982
PhoglyStruct	0.8726	0.7331	0.7976	0.7678	0.5343	0.6513	0.8034