Predicting secondary structure of proteins: a comparison between GOR method and Support Vector Machines

Supplementary materials

Matteo Bolner

1) Blind set statistics

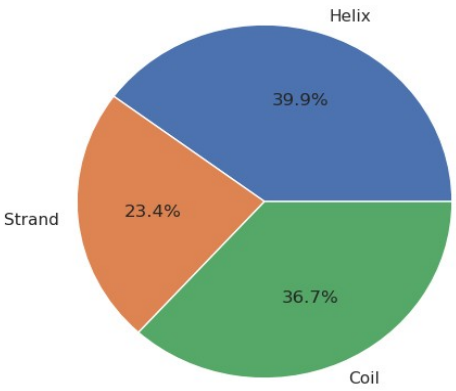

Figure 1 : Distribution of secondary structure conformations

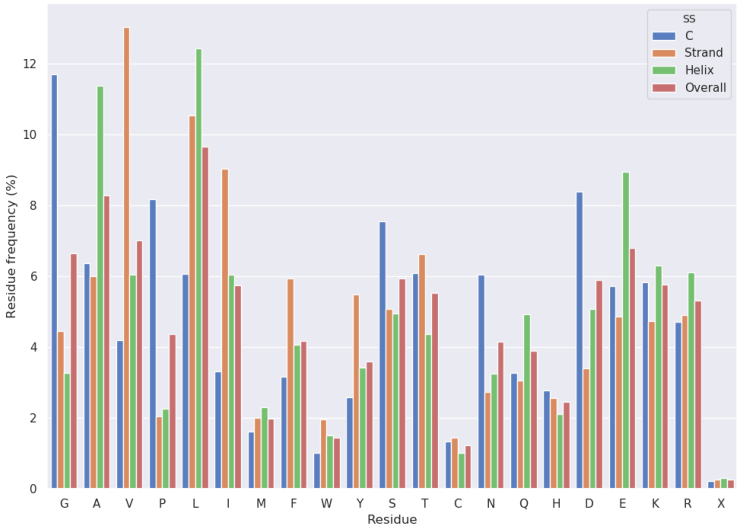

Figure 3 : Distribution of secondary structure conformations given the residues

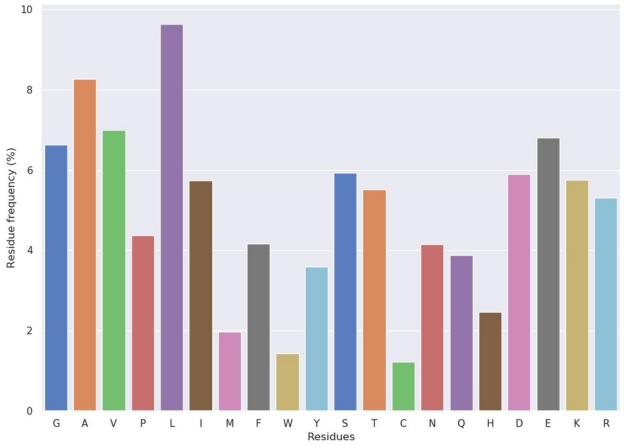

Figure 2 : Distribution of residues

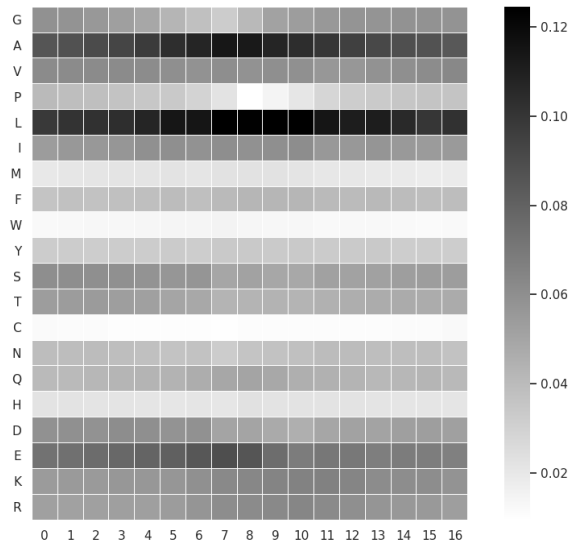

Figure 4A : Heatmap for the helix conformation

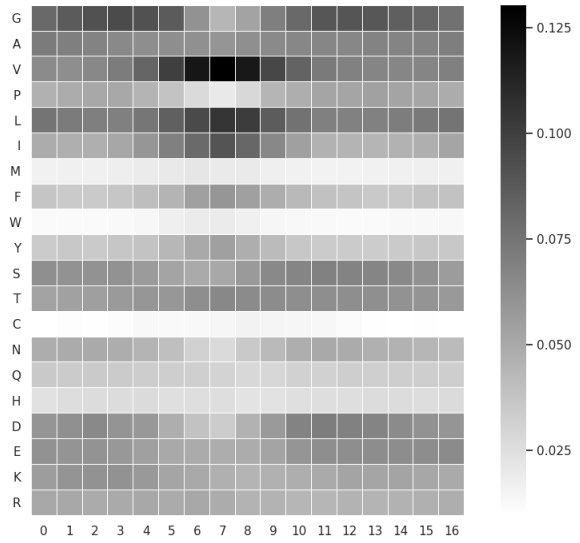

Figure 4B : Heatmap for the strand conformation

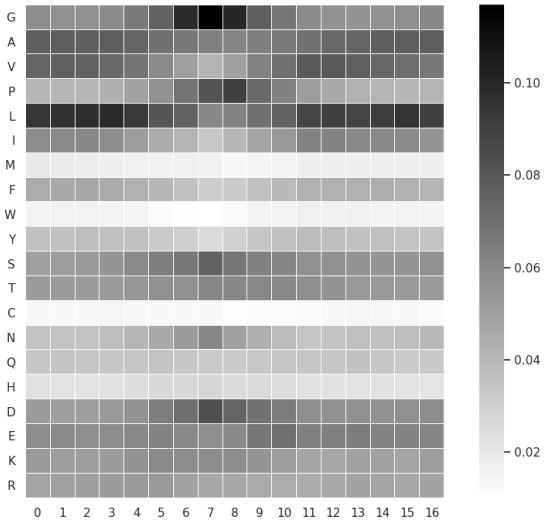

Figure 4C : Heatmap for the coil conformation

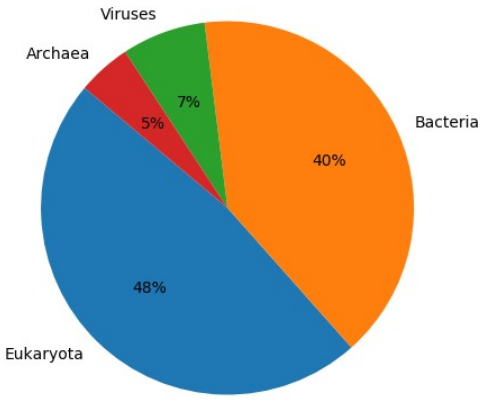

Figure 5A : Taxonomic distribution by kingdom

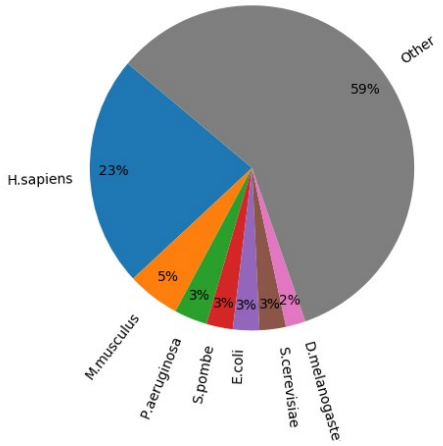

Figure 5B : Taxonomic distribution by species

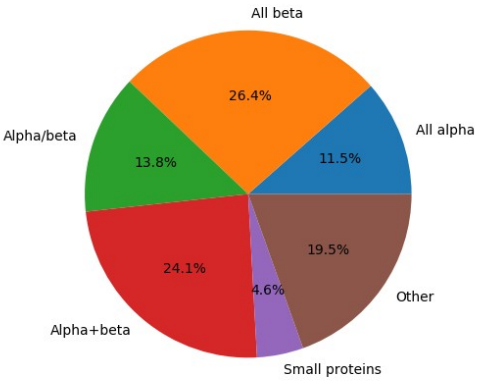

Figure 6 : Distribution of SCOP classes

2) Training set additional statistics

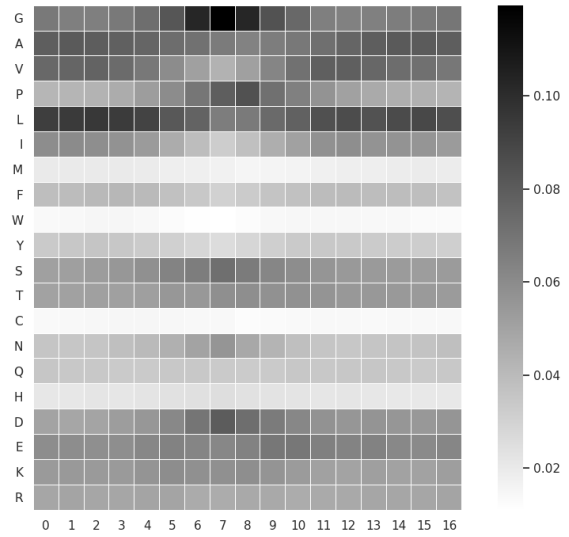

Figure 7 : Heatmap for the coil conformation of the training set

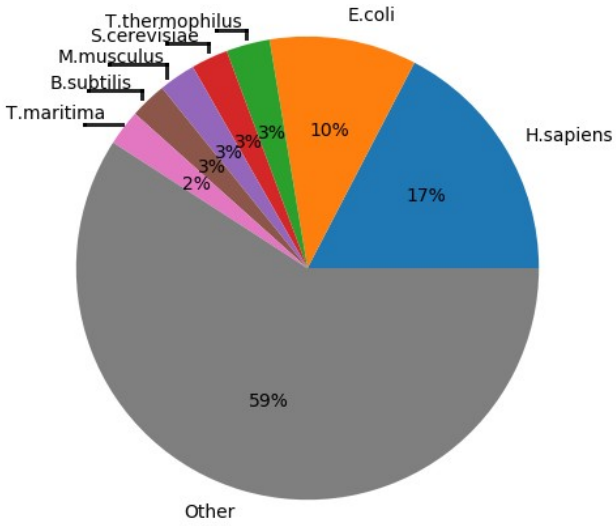

Figure 8: Taxonomic distribution of the training set by species

3) SVM alternative models

-g 0.5 , -c 4	SEN	PPV	MCC	TCA
H	68.2 ± 0.91	84.41 ± 0.33	64.81 ± 0.73	
E	39.63 ± 0.94	78.74 ± 0.8	48.28 ± 0.51	70.1 ± 0.26
C	87.6 ± 0.28	61.69 ± 0.37	48.0 ± 0.33	

-g 2 , -c 4	SEN	PPV	MCC	TCA
H	12.16 ± 0.43	87.44 ± 0.96	24.72 ± 0.64	
E	1.93 ± 0.31	81.4 ± 2.76	10.21 ± 0.88	46.53 ± 0.31
C	98.54 ± 0.09	44.2 ± 0.23	15.08 ± 0.46	

-g 2 , -c 2	SEN	PPV	MCC	TCA
H	12.34 ± 0.45	87.47 ± 0.94	24.92 ± 0.68	
E	1.96 ± 0.31	82.24 ± 2.93	10.39 ± 0.88	46.61±0.32
C	98.54 ± 0.09	44.24 ± 0.23	15.27 ± 0.48	