Supplementary Data

1.1 SS composition

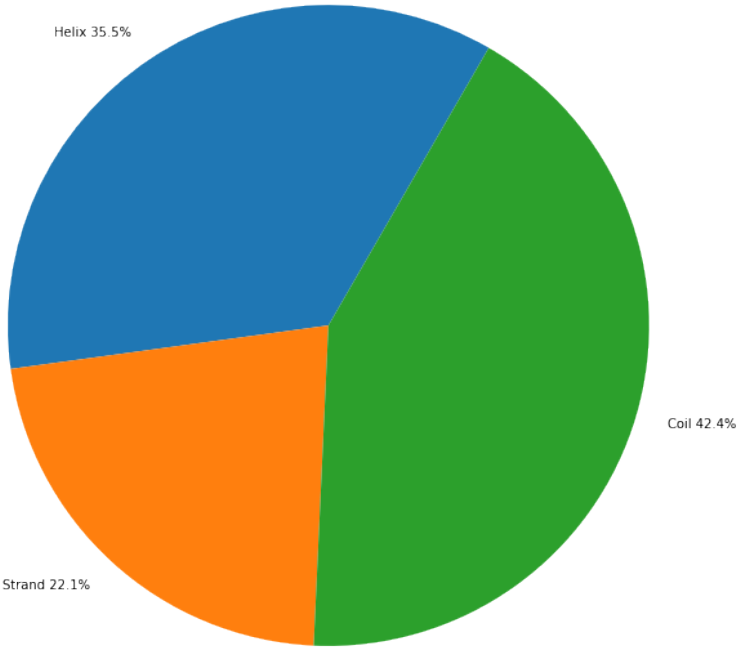

Fig. S 2. Secondary structure composition of the training set. Abundances of helix, strand and coil are shown in blue, orange and green, respectively.

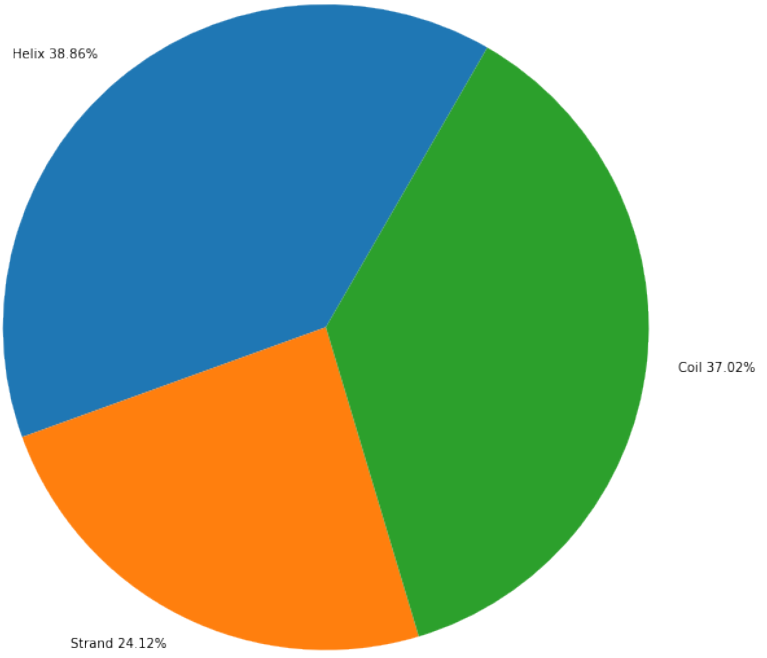

Fig. S 3. Secondary structure composition of the blind test set. Abundances of helix, strand and coil are shown in blue, orange and green, respectively.

1.2 Residue composition

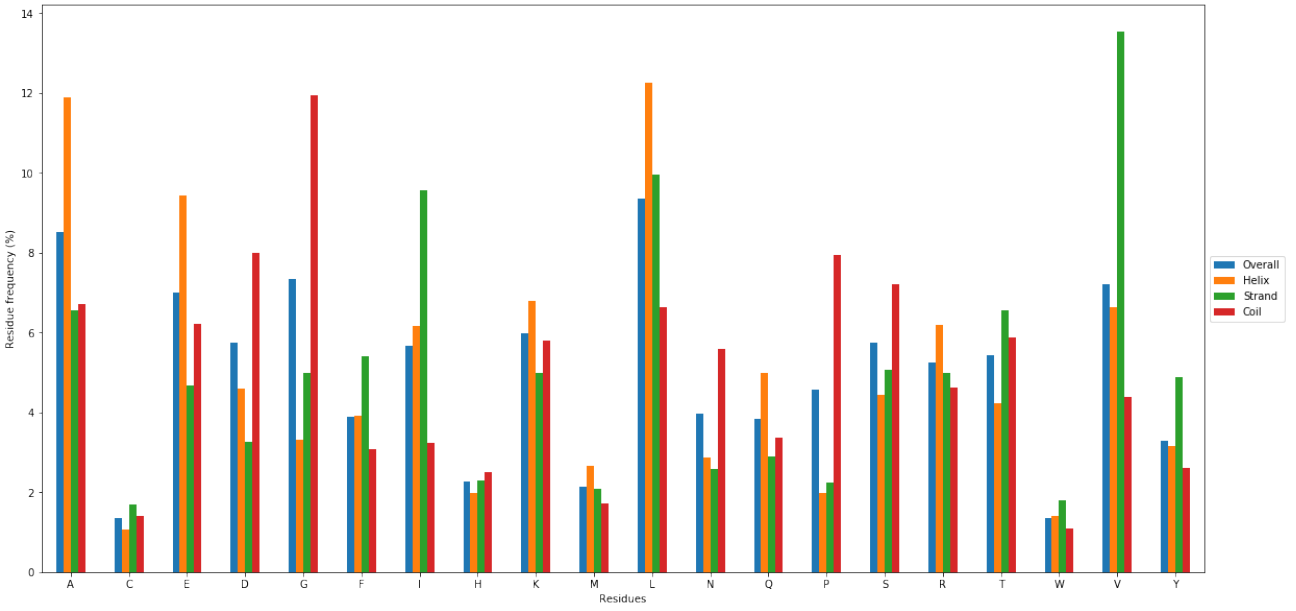

Fig. S 4. Residue composition of the training set. The overall composition is shown in blue. Compositions of helix, strand and coil are shown in orange, green and red, respectively.

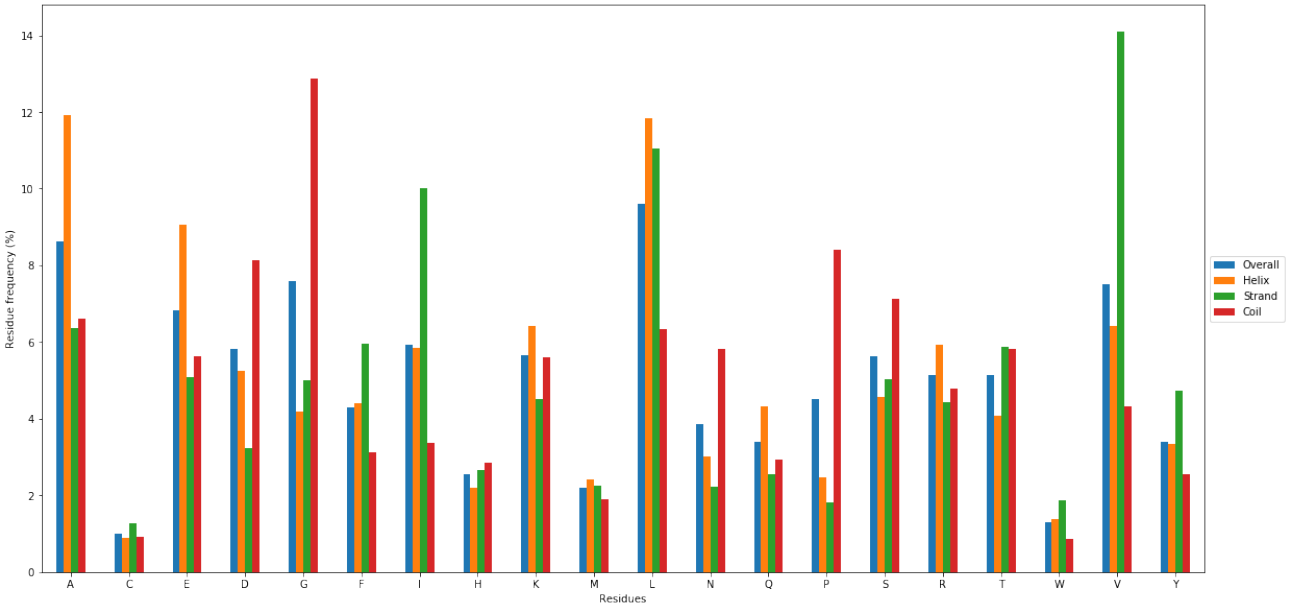

Fig. S 5. Residue composition of the blind test set. The overall composition is shown in blue. Compositions of helix, strand and coil are shown in orange, green and red, respectively.

1.3 Residue composition: windows (helix vs strand)

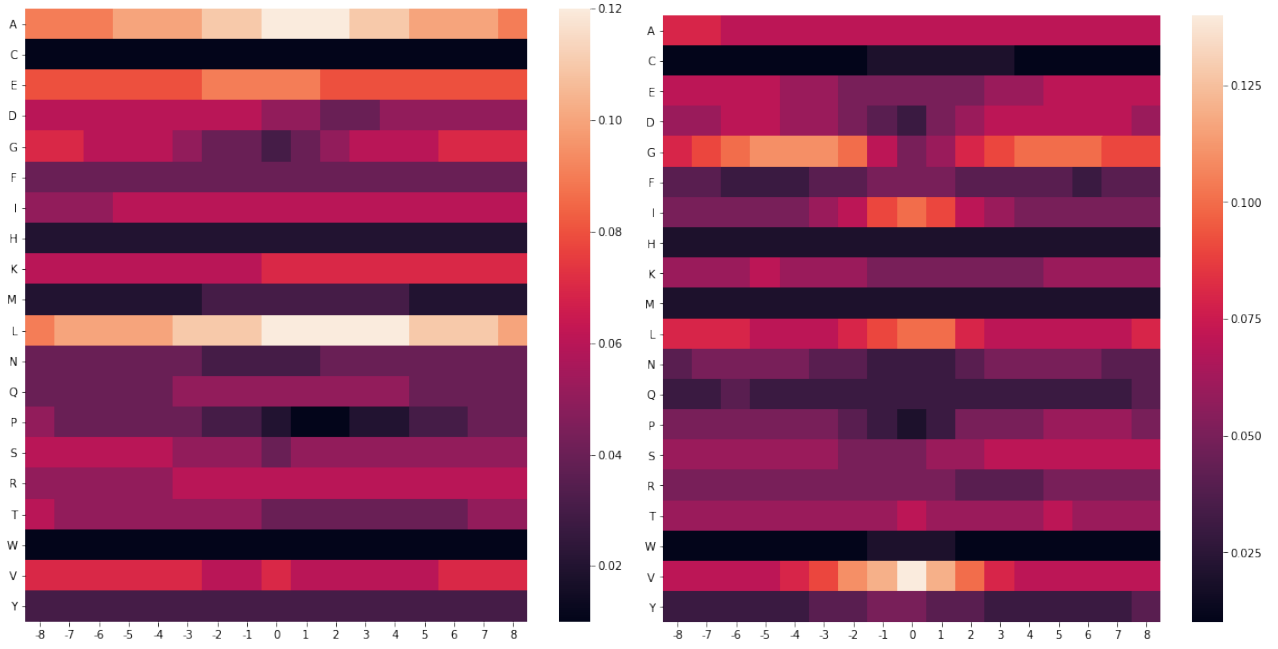

Fig. S 6. Window residue composition of the training set. On the left the composition of a 17-residue window whose central residue's conformation is helix. On the right the composition of a 17-residue window whose central residue's conformation is strand. Lighter colors indicate higher percentages while darker colors indicate lower percentages.

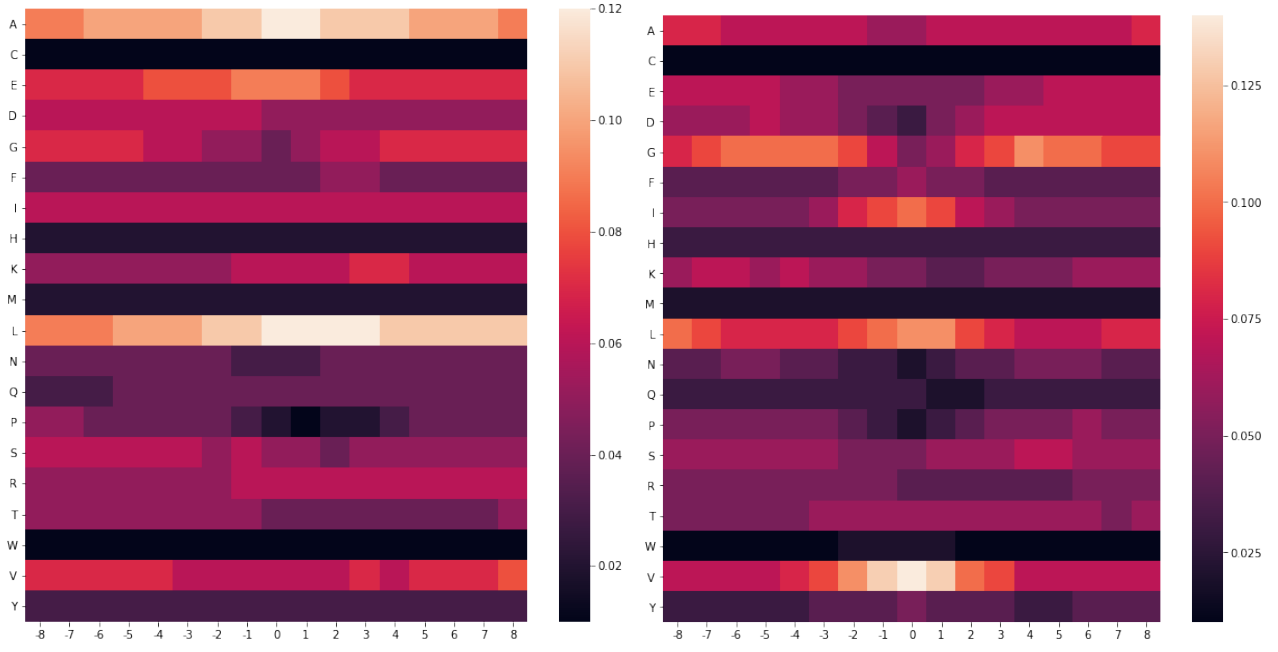

Fig. S 7. Window residue composition of the blind test set. On the left the composition of a 17-residue window whose central residue's conformation is helix. On the right the composition of a 17-residue window whose central residue's conformation is strand. Lighter colors indicate higher percentages while darker colors indicate lower percentages.

1.4 Taxonomic classification: kingdom

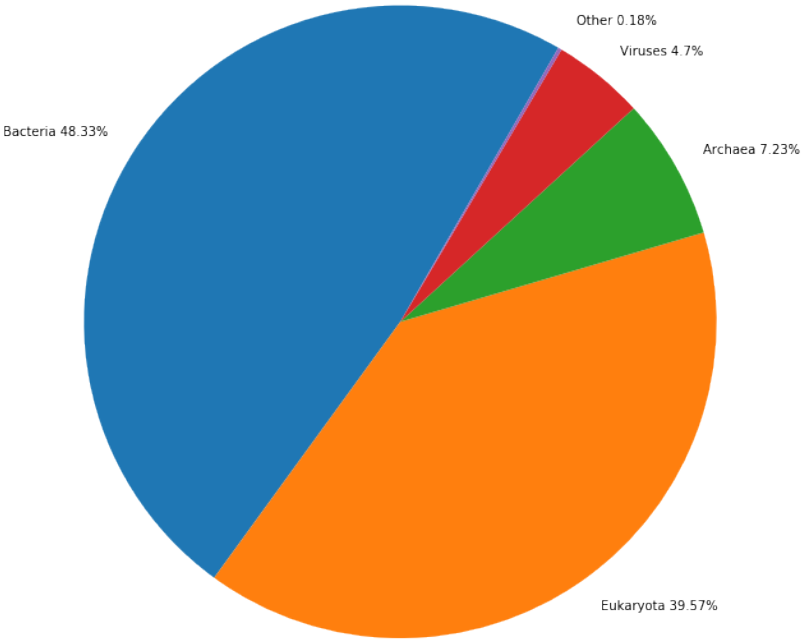

Fig. S 8. Taxonomic classification for sequences of the training set grouped by kingdom. Abundance of sequences coming from Bacteria, Eukaryota, Archaea, Viruses and other are shown in blue, orange, green, red and purple, respectively.

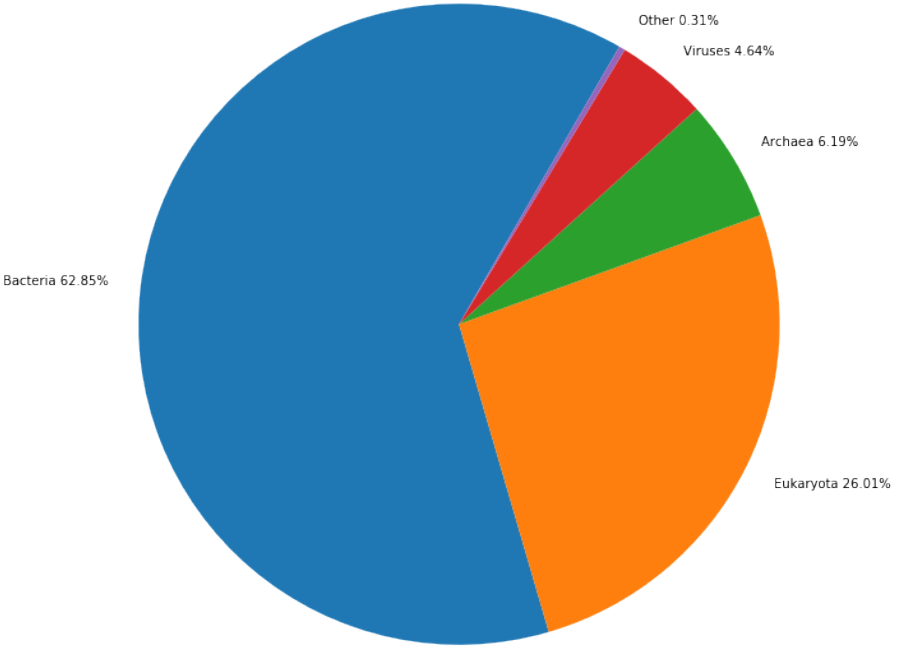

Fig. S 9. Taxonomic classification for sequences of the blind test set grouped by kingdom. Abundance of sequences coming from Bacteria, Eukaryota, Archaea, Viruses and other are shown in blue, orange, green, red and purple, respectively.

1.5 Taxonomic classification: species

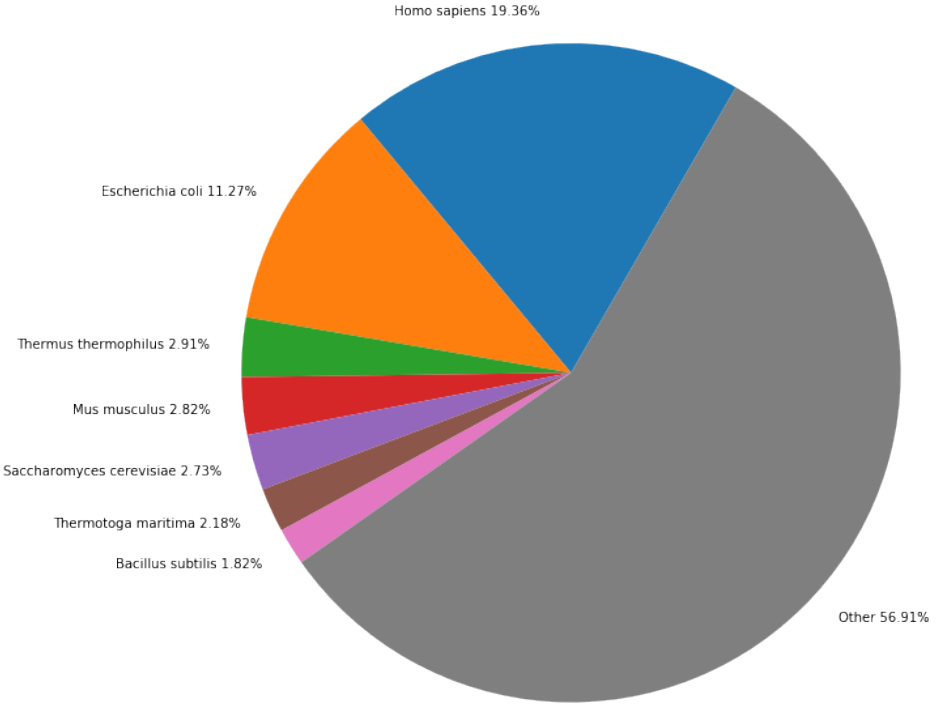

Fig. S 10. Taxonomic classification for sequences of the training set grouped by species.

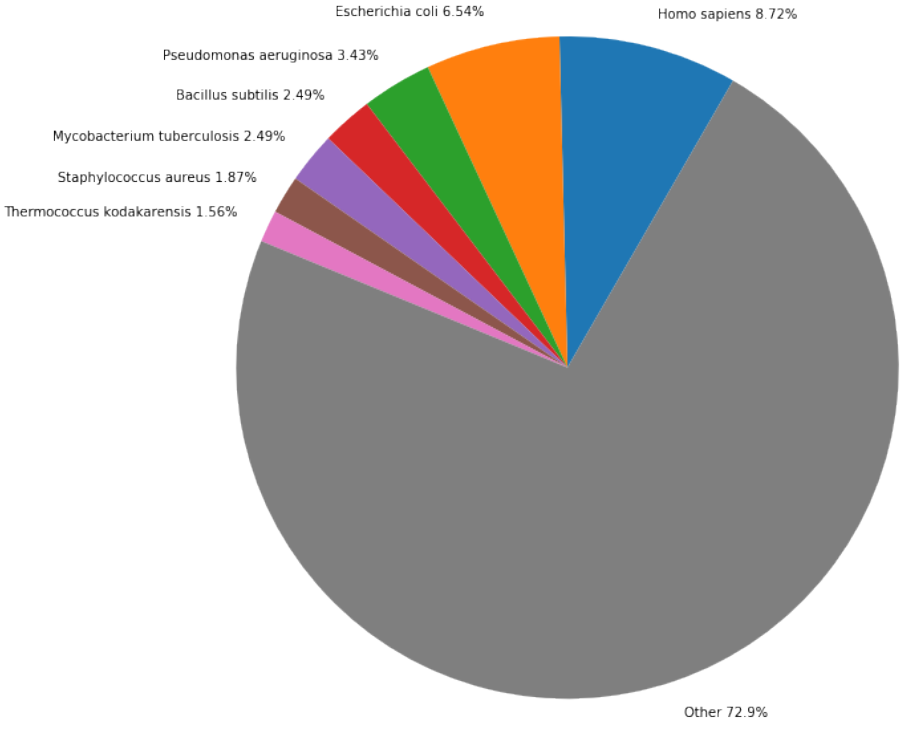

Fig. S 11. Taxonomic classification for sequences of the blind test set grouped by species.