

# Practical 1

## 01 August 2022

Discriminate  $\alpha$ -helical and  $\beta$ -barrel membrane proteins using amino acid composition

### Instructions

1. Upload your program files, output files and the report (naming: “roll no.pdf”) in a zipped folder (naming: “roll no.zip”).
2. Do not copy codes, it will be checked for plagiarism.

### Steps:

1. Go to PDBTM database (<http://pdbtm.enzim.hu/>)
2. Download alpha helical membrane protein sequences (TMH) in FASTA format
3. Obtain non-redundant sequences using CD-HIT software (40%) and online (30%)
4. Repeat steps 2 and 3 for beta barrel membrane proteins (TMB)
5. Compute the overall amino acid composition in TMH and TMB (20 values each)
6. For each sequence in TMH
  - (a) Compute the composition
  - (b) Compare with overall composition of TMH and compute the absolute deviation and total for the 20 residues
$$\sigma(\text{TMH}) = \sum |\text{comp}(x) - \text{comp}(\text{TMH})|$$
  - (c) Compare with overall composition of TMB and compute the absolute deviation and total for the 20 residues
$$\sigma(\text{TMB}) = \sum |\text{comp}(x) - \text{comp}(\text{TMB})|$$
  - (d) If  $\sigma(\text{TMH}) < \sigma(\text{TMB})$ , the protein is TMH  
Otherwise, it is TMB
  - (e) Correctly predicted TMH are True Positives (TP)
  - (f) Wrongly predicted as TMB are False Negatives (FN)
7. Repeat the same with all TMB proteins. In this case,
  - (e) Correctly predicted TMB are True Negatives (TN)
  - (f) Wrongly predicted as TMH are False Positives (FP)
8. Compute sensitivity, specificity and accuracy
$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$
$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$
$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

9. Take 50% of TMH and 50% of TMB to compute the composition (step 5). For the remaining set of proteins follow steps 6 to 8 to assess the performance.
10. Change the split in question 9 to 30%, 40%, 60% and 70% and repeat the computation. Tabulate the data. (Optional)
11. In 6d include a deviation  $\delta$  (E.g.,  $\sigma(\text{TMH}) + 0.5$ ) estimate the sensitivity, specificity and accuracy. (Optional)

**Deadline: 07 August 2022**