

Team 9 - Downstream

Team members : Apolline BRULEY, Clara CHARON, Jules RODRIGUES, Fahmi SATOURI et Alexis TRICOT

Context

This year's goal of the Meet-U initiative is to predict proteins folding from their primary sequence. To achieve this goal two sets of teams were formed : upstream teams and downstream teams. The upstream teams have to align the query sequences on a database in order to find the proteins' closest families and the downstream teams have to predict the proteins' 3D structure based on the alignments obtained by the upstream teams.

Strategy

We decided to be a downstream team, to predict the 3D structure of proteins we threaded them over the alignments obtained by team 2, we paired with them and used their program to get the alignments we would work with. To assess the predictive 3D models we obtained we computed a personalized version of the H-factor which is a metric developed by *di Luccio & Koehl (2011)*, an indicator of the quality of a protein model.

The H-factor combines 4 of the model's characteristics:

- alignment: with a score inversely related to the identity
- secondary structure: with a score depending on whether the predicted secondary structure of a query residue is coherent with the secondary structure of the aligned residue of the template
- structural coherence: score based on statistical distance (DOPE matrice)
- heterogeneity of the models: distance between a model and the average model.

Results

We tested our program on a test dataset, and obtained results that are represented in the ROC curves in *figure 1*.

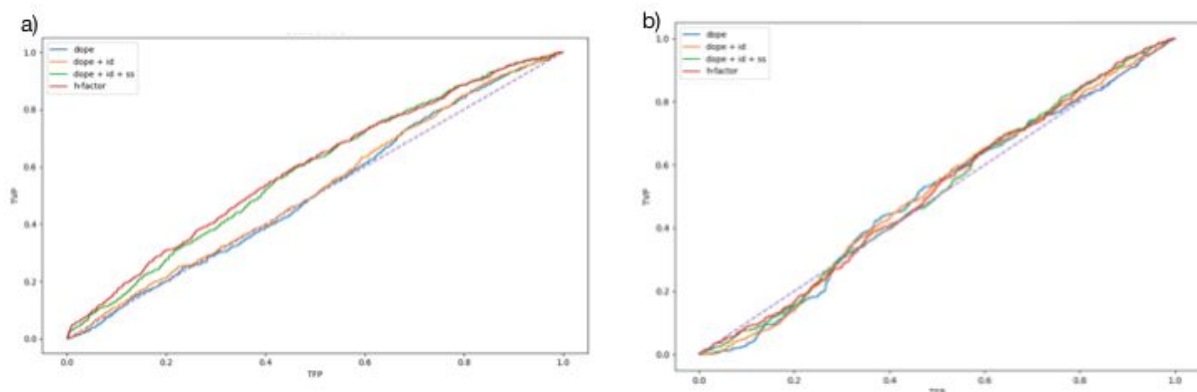


Figure 1: ROC curves combining False Positive Rates according to True Positive Rate. Each curve a) Models are considered as True positives if the template they are based on is of the same scope Class as the query one. b) Models are considered as True positives if the template they are based on is of the same scope Superfamily as the query one.