Team 4 (Downstream)
Paris Diderot University

Authors : Gabriel Cretin, Hélène Kabbech, Tom Gutman, Flora Mikaeloff & Franz-Arnold Ake

We have developed the Fold-U program, a protein structure prediction project which is part of the 2018-2019 Meet-U competition.

After running the SALUT program (Upstream) developed by the Team 1, the query sequence is threaded on the template for each alignments. Then, the program MODELLER generates a new refined 3D model by homology. Several other scores are also calculated in order to take into account different parameters describing the alignment. Each scores are normalized using the min-max scaling method (values between 0 and 1). We use machine learning (logistic regression) to calculate a weighted combined score and obtain a better ranking. Finally, the top N models are generated. (See **Figure 1**).

The Fold-U program was run for 20 tests queries and benchmarked using enrichment style plot (See **Figure 2**) and top N information to evaluate the power and the relevance of the different scores. The machine learning improves the ranking and makes the program less specific and more sensitive.

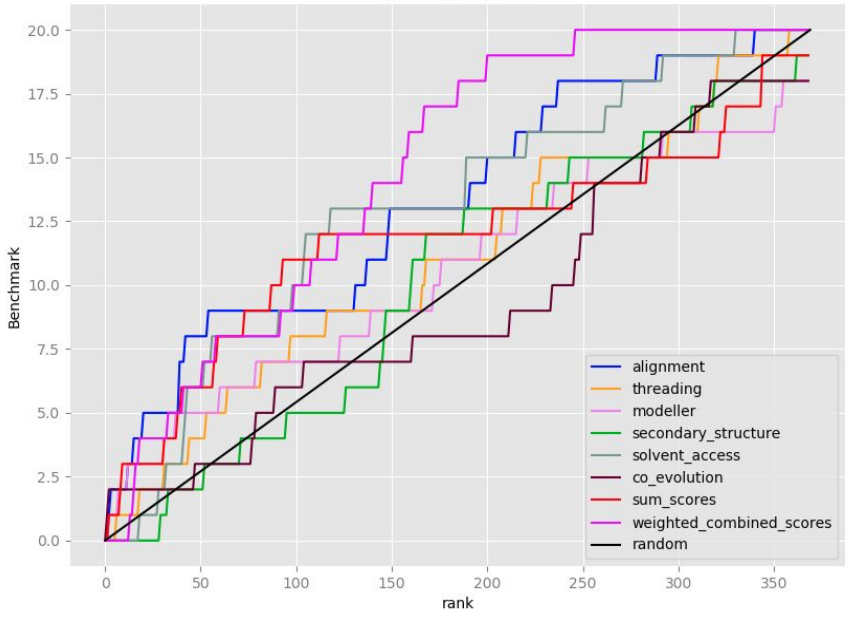

Figure 2 : Enrichment plot : Cumulative sum of benchmarks encountered along the ranking (from rank 1 to rank 397).

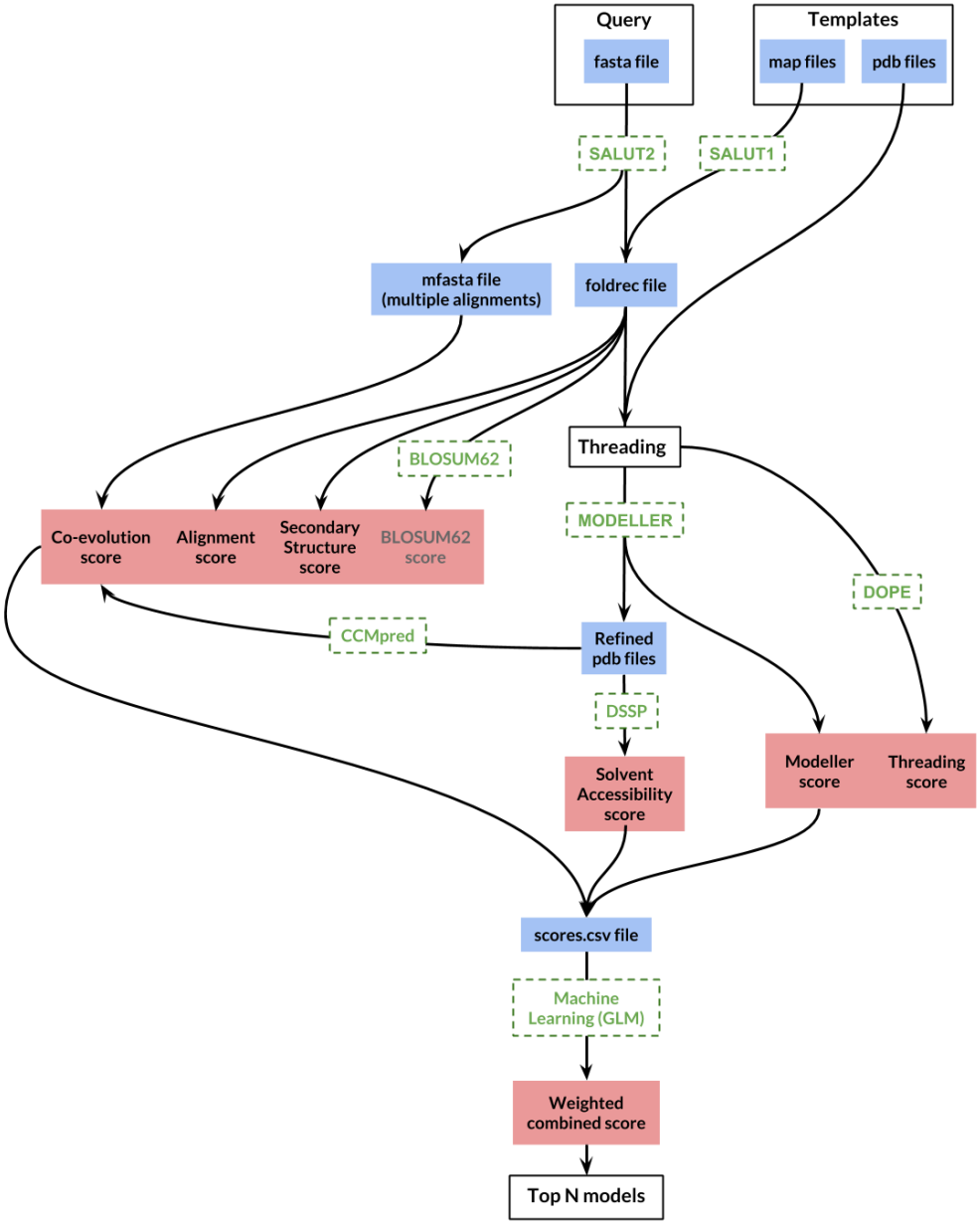

Figure 1 : Pipeline summarizing the strategy of the Fold-U program