

Readme for using PRESCOTT Online

PRESCOTT Online comprises the ESCOTT webserver, the PRESCOTT webserver and the comprehensive ESCOTT Human Protein Database.

Example files in ESCOTT webserver

Example files allow you to test the BLAT protein.

There are three input files, one mandatory and two optional:

- MSA file of the query sequence (mandatory), in fasta format -> aliBLAT.fasta. Your query protein must be the first sequence in the fasta file. In addition, the query sequence should not contain any gaps.
- pdb file for the query sequence (optional) -> blat-af2.pdb. ESCOTT functions without it, but with less precision; in this case, it computes mutational effects with iGEMME, an improved version of the GEMME program.
- a file listing specific mutations of interest (optional).

The ESCOTT output comprises a text file corresponding to the predictions reported with raw scores (BLAT_normPred_EvolCombi.txt). A folder containing csv files with raw (BLAT_normPred_EvolCombi.txt) and ranksorted scores (BLAT_normPred_EvolCombiTransposedRanksorted.csv) together with evolutionary conservation, physico-chemical, circular variance scores (see columns "trace", "pc", "cv" in BLAT_jet.res; column "sstjetormaxtwocomponent" reports ESCOTT maxscore) and a help file (ESCOTT_Outputs_Readme.txt). Both raw and ranksorted scores are visualised online through dedicated graphical interactive heatmaps.

If a file with specific mutations of interest was submitted, the output is an online table reporting ESCOTT raw scores for each specified mutation. The table is downloadable (BLAT_normPred_EvolCombi.txt).

From the graphical interface, you can download png images of the matrix display by using the panel of available buttons (on Safari you need to do it twice and we apologize for the inconvenience).

Example files in PRESCOTT webserver

Example files allow you to test the MLH1 protein.

PRESCOTT uses three input files:

- Output of the ESCOTT prediction in txt format -> MLH1_normPred_evolCombi.txt
- Query sequence in fasta format -> MLH1.fasta
- Gnomad v4.0.0 csv file related to the query -> gnomAD_v4.0.0_MLH1_HUMAN_ENSG00000076242.csv

PRESCOTT output comprises a csv file containing the predictions reported with ranksorted scores (MLH1.csv), a csv file containing extra information on each mutation (MLH1-details.csv) and a help file (PRESCOTT_Outputs_Readme.txt) with the descriptions of the output files.

Both PRESCOTT and a comparative visualisation of PRESCOTT vs ESCOTT are accessible online through dedicated graphical interactive heatmaps. Differences between PRESCOTT and ESCOTT score values can be also visualised through available buttons.

From the graphical interface, you can download png images of the matrix displays by using the panel of available buttons (on Safari you need to do it twice and we apologize for the inconvenience).

ESCOTT Human Protein Database

The database requires as input a protein name (Uniprot ID or gene name). It outputs ESCOTT predictions. From the graphical interface, you can download png images of the matrix display by using the panel of available buttons.

Predictions for the ESCOTT Human Proteome Database of 19,432 proteins are downloadable at <https://doi.org/10.5281/zenodo.10577421>.

Command lines for a local usage of the PRESCOTT software

PRESCOTT source code is available at <http://gitlab.lcqb.upmc.fr/tekpinar/PRESCOTT/>. ESCOTT and iGEMME are a part of the PRESCOTT software. We provide a docker image of PRESCOTT at <https://hub.docker.com/repository/docker/tekpinar/PRESCOTT-docker/general>.

For ESCOTT:

```
`docker run --rm -v $PWD:/home/tekpinar/research/lcqb
tekpinar/prescott-docker:v1.6.0 escott aliBLAT.fasta -f aliBLAT.fasta
--pdbfile blat-af2.pdb`
```

For PRESCOTT:

```
`docker run --rm -v $PWD:/home/tekpinar/research/lcqb
tekpinar/prescott-docker:v1.6.0 prescott -e
MLH1_normPred_evolCombi.txt -g
gnomAD_v4.0.0_MLH1_HUMAN_ENSG00000076242.csv -s MLH1.fasta`
```