MolTarPred Tutorial

MolTarPred, freely available at http://moltarpred.marseille.inserm.fr, is a web server for protein target prediction exploiting ChEMBL22 data. The underlying method was investigated in a recent publication [1].

To run a prediction of protein targets, the user must provide the chemical structure of the query molecule in SMILES format (length within 1KB) and click on the *Submit* button. These two steps are illustrated in Figure 1, with the SMILES of the anticancer drug vorinostat displayed, which acts as an example of query molecule.

Once the query is submitted, Tanimoto similarities between the Morgan fingerprints of the query molecule and that of each of the 607,659 ChEMBL compounds to be screened are calculated (these compounds have 4,553 single-protein targets with a activity annotated better or equal than 10 µM). Predicted targets will be those retrieved from the top 10 most similar molecules to the query (the confidence in a predicted target will be estimated as the number of these top hits known to have activity against the predicted target).

On completion of this run (typical wall time is one minute), the user will be redirected to the result webpage (Figure 2) with a unique URL that is only available to the user. Users are advised to bookmark this webpage if they want to browse the result at a later time. In this webpage, the three blue links at the top provide downloads for the query molecule, the hit targets and the hit compounds. The table in the middle lists the ChEMBL_ID of the predicted target, target name, organism and confidence score (the number of top 10 hit compounds with the predicted target). The higher the confidence score, the more likely the predicted target to be the target of the query molecule. In retrospective analysis, we have observed that on average the predicted target is an actual target of the query molecule 50% of the times when the confidence score is 2 or more.

If the confidence score is clicked, the hit compounds of the selected target will be shown and visualised at the bottom. Users can visually compare the chemical structures of the query molecule to those of the hit compounds. The query molecule *vorinostat* is one of the 607,659 ChEMBL compounds that are screened in MolTarPred, so it is listed as the first hit compound with a similarity of 100% (Figure 2). Results can be interpreted as follows:

- Those predicted targets with the query molecule appearing as the only top10 hit are actually not predictions, but one of the known targets for the query molecule (disregard the confidence score of 1 as this is only meaningful for predicted targets).
- Those predicted targets with the query molecule as a top10 hit and a confidence score of 2 or more represent correct predictions of MolTarPred because in this case the predicted target is already a known target of the query molecule shows (confidence score minus 1 is a more accurate estimation of conference). For instance, histone deacetylase 1 is predicted with confidence score 9 because 9 out of the top 10 most similar molecules to the query are hitting this target, including *vorinostat*.
- Those predicted targets from a set of top hits not including the query molecule are still to be validated prospectively (use the confidence score as given). None of the predicted targets are of this type with *vorinostat*. So we will be looking at a second example.

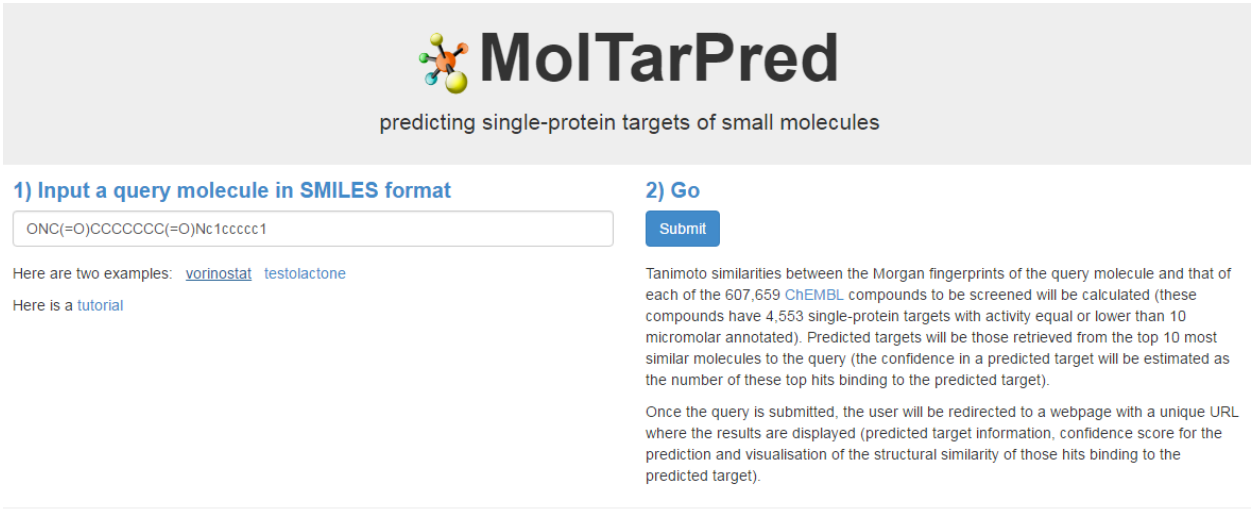

Figure 1. Submission webpage with vorinostat as the query molecule.

MolTarPred

predicting single-protein targets of small molecules

Completed query.smi (tutorial)
Download hit_targets.tsv and hit_compounds.tsv

Predicted Target	Target Name	Organism	Confidence
CHEMBL325	Histone deacetylase 1	Homo sapiens	9
CHEMBL1865	Histone deacetylase 6	Homo sapiens	8
CHEMBL1829	Histone deacetylase 3	Homo sapiens	7
CHEMBL1937	Histone deacetylase 2	Homo sapiens	7
CHEMBL3192	Histone deacetylase 8	Homo sapiens	5
CHEMBL3310	Histone deacetylase 11	Homo sapiens	5
CHEMBL5103	Histone deacetylase 10	Homo sapiens	5
CHEMBL2111383	Histone deacetylase 3/Nuclear receptor corepressor 2 (HDAC3/NCoR2)	Homo sapiens	2
CHEMBL1293190	Histone deacetylase	Plasmodium falciparum 3D7	1
CHEMBL1795094	Nuclear receptor subfamily 0 group B member 1	Homo sapiens	1
CHEMBL2563	Histone deacetylase 5	Homo sapiens	1
CHEMBL2716	Histone deacetylase 7	Homo sapiens	1
CHEMBL3038484	Histone deacetylase 3/NCoR1	Homo sapiens	1
CHEMBL3524	Histone deacetylase 4	Homo sapiens	1
CHEMBL3541	Histone deacetylase HD1B	Zea mays	1
CHEMBL4001	Histone deacetylase 1	Mus musculus	1
CHEMBL4145	Histone deacetylase 9	Homo sapiens	1
CHEMBL4919	Histone deacetylase HD2	Zea mays	1
CHEMBL5541	Histone deacetylase	Plasmodium falciparum	1
CHEMBL6017	Histone deacetylase-like amidohydrolase	Alcaligenes sp. (strain DSM 11172) (Bordetella sp. (strain FB188))	1

Hit Compound	SMILES	Similarity
CHEMBL98	O=C(CCCCCCC(=O)Nc1ccccc1)NO	100%
CHEMBL512644	O=C(CCCCCCC(=O)Nc1ccc(-c2ccccc2)cc1)NO	83%
CHEMBL3098605	Cc1ccc(NC(=O)CCCCCCC(=O)NO)cc1	74%
CHEMBL1923808	O=C(CCCCCCC(=O)Nc1ccc([18F])cc1)NO	74%
CHEMBL473270	O=C(CCCCCCC(=O)Nc1ccc(-c2ccccc2)c1)NO	73%
CHEMBL1092762	Cc1ccc(NC(=O)CCCCC(=O)NO)cc1	71%
CHEMBL325676	O=C(CCCCCCS)Nc1ccccc1	70%
CHEMBL326433	O=C(CCCCCCCS)Nc1ccccc1	70%
CHEMBL109654	O=C(CCCCCS)Nc1ccccc1	70%

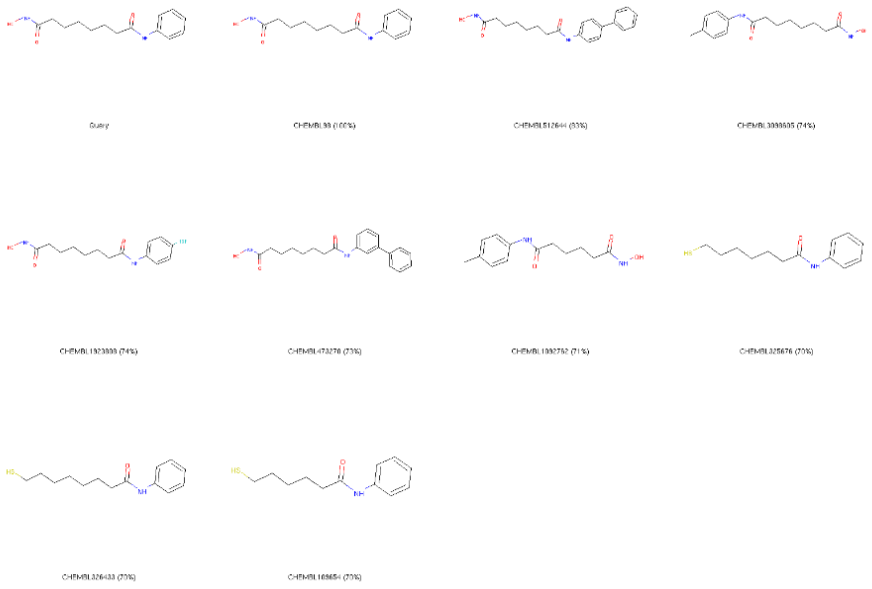

Figure 2. Result webpage showing a list of predicted targets of vorinostat and visualising the structures of the query against those top10 hit compounds annotated with the inspected predicted target (histone deacetylase 1).

Figure 3 below shows the result webpage of using another anticancer drug *testolactone* as a query molecule. This drug is included in ChEMBL, but it is not among the 607,659 ChEMBL compounds screened in MolTarPred because it has no single-protein targets annotated with activity equal or lower than 10 micromolar annotated (only targets with ChEMBL's confidence score = 7,8 and 9, which is different from our confidence score for target predictions and indicate direct assignment to protein complexes or single proteins). Therefore, none of the predicted targets in Figure 3 include the query molecule and thus they are still to be validated prospectively (use the confidence score as given). For instance, Cytochrome P450 19A1 is predicted with confidence score 6 because 6 out of the top 10 most similar molecules to the query are hitting this target (these three hits do not include the query molecule, as it can be seen at the bottom of Figure 3).

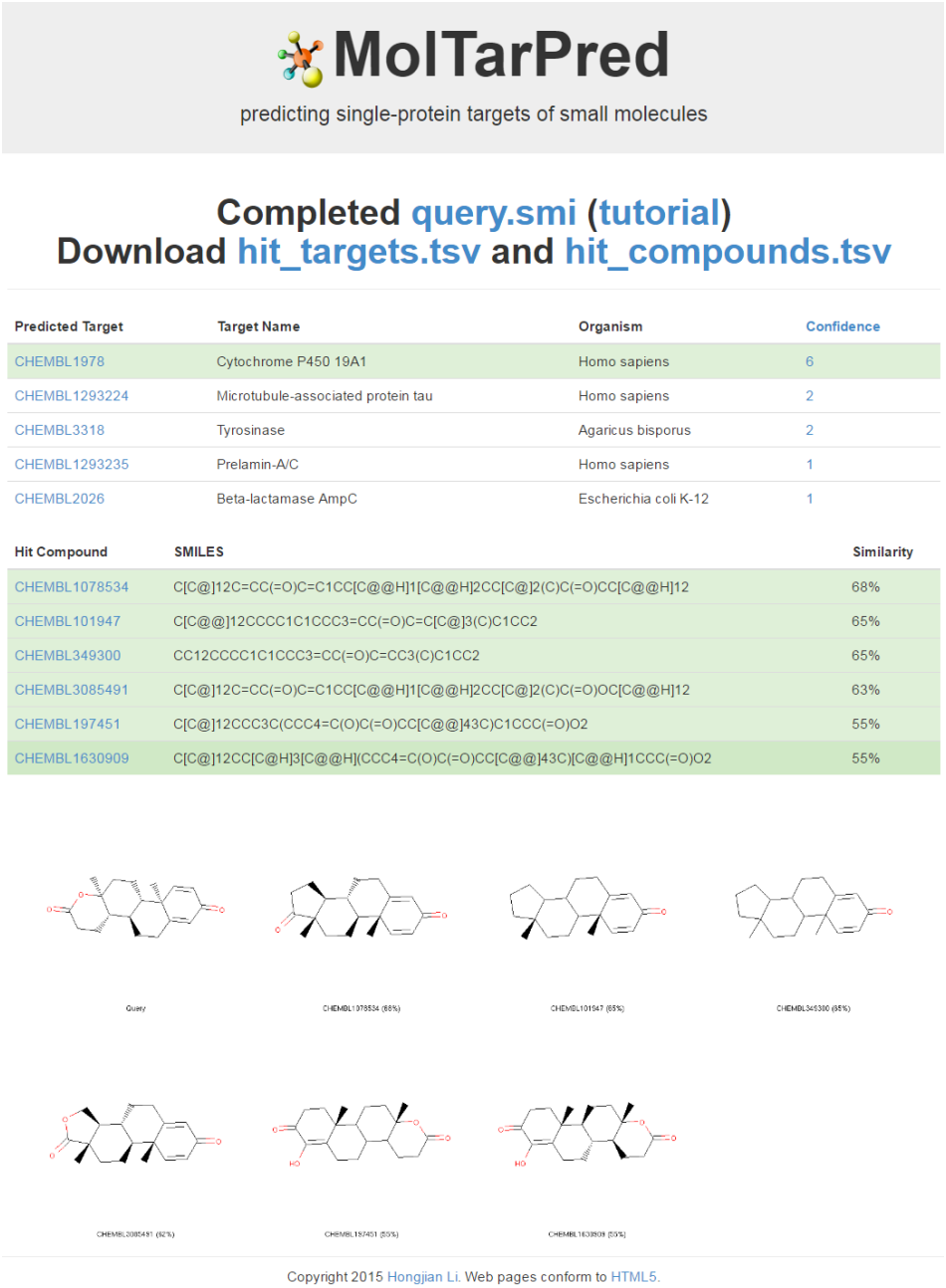

Figure 3. Result webpage showing a list of predicted targets of testolactone and visualizing the query and hit compounds of the selected target cytochrome P450 19A1.

References

1. Peón, A.; Naulaerts, S.; Ballester, P. J. Predicting the Reliability of Drug-target Interaction Predictions with Maximum Coverage of Target Space. *Sci. Rep.* **2017**, *7*, 3820.