TD5: Prediction of protein 3D structures using molecular modelisation

You will need the following programs:

- *a sequence editor*
- *3D structure viewer (PyMOL)*

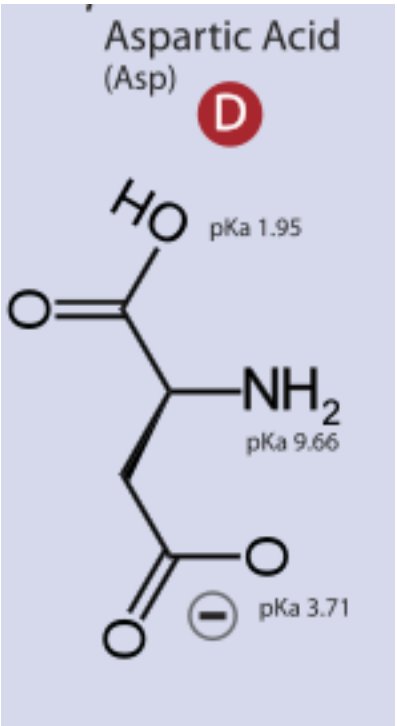
We will test an hypothesis about therole of a SNP on a gene of interest.

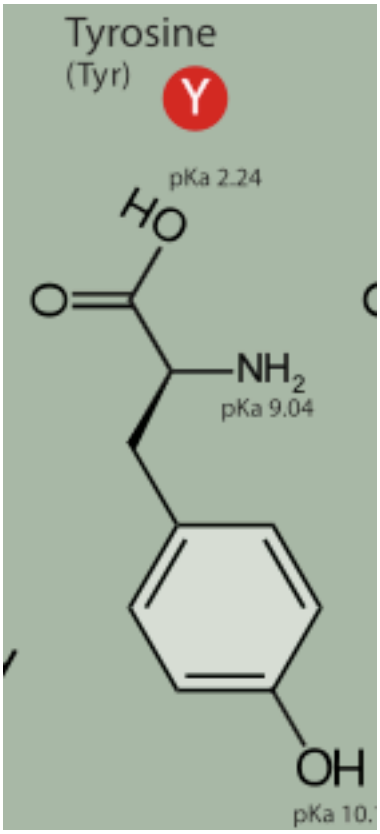
Here is the gene sequence:

```
>TDR
MGRGDHLLMKNSNAAAAAAAVNGGGTSLDAALRPLVGSDGWDYCIYWRLS
PDQRFLEMTGFCCSSELEAQVSALLDLPSSIPLDSSSIGMHAQALLSNQP
IWQSSSEEEEADGGGGAKTRLLVPVAGGLVELFASRYMAEEQQMAELVMA
QCGGGGAGDDGGGQAWPPPETPSFQWDGGADAQRLMYGGSSLNLFDAAAA
DDDPFLGGGGGDAVGDEAAAAGAWPYAGMAVSEPSVAVAQEQMQHAAGGG
VAESGSEGRKLHGGDPEDDGDGEGRSGGAKRQQCKNLEAERKRRKKLNGH
LYKLRSLVPNITKMDRASILGDAIDYIVGLQKQVKELQDELEDNHVHHKP
PDVLIDHPPPASLVGLDNDDASPPNSHQQQPPLAVSGSSSRRSNKDPAMT
DDKVGGGGGGGHRMEPQLEVRQVQGNELFVQVLWEHKPGGFVRLMDAMNA
LGLEVINVNVTTYKTLVLNVFRVMVRDSEVAVQADRVRDSLLEVTRETYP
GVWPSPQEEDDAKFDGGDGGQAAAAAAAAGGEHYHDEVGGGYHQHLHYLA
FD
```

in red, look at the SNP: D in the wild crop becomes Y in the mutant crop.
ITKMDRASIL becomes ITKYRASIL

A residue <u>charged</u> and <u>hydrophilic</u> is replaced by a <u>neutral</u> and <u>hydrophobic</u> one.
What could be the impact of this punctual mutation on the protein function:





<u>How to get the sequence if interest</u>

A simple copy and paste from this document or on the website of the training Sequence_TD5

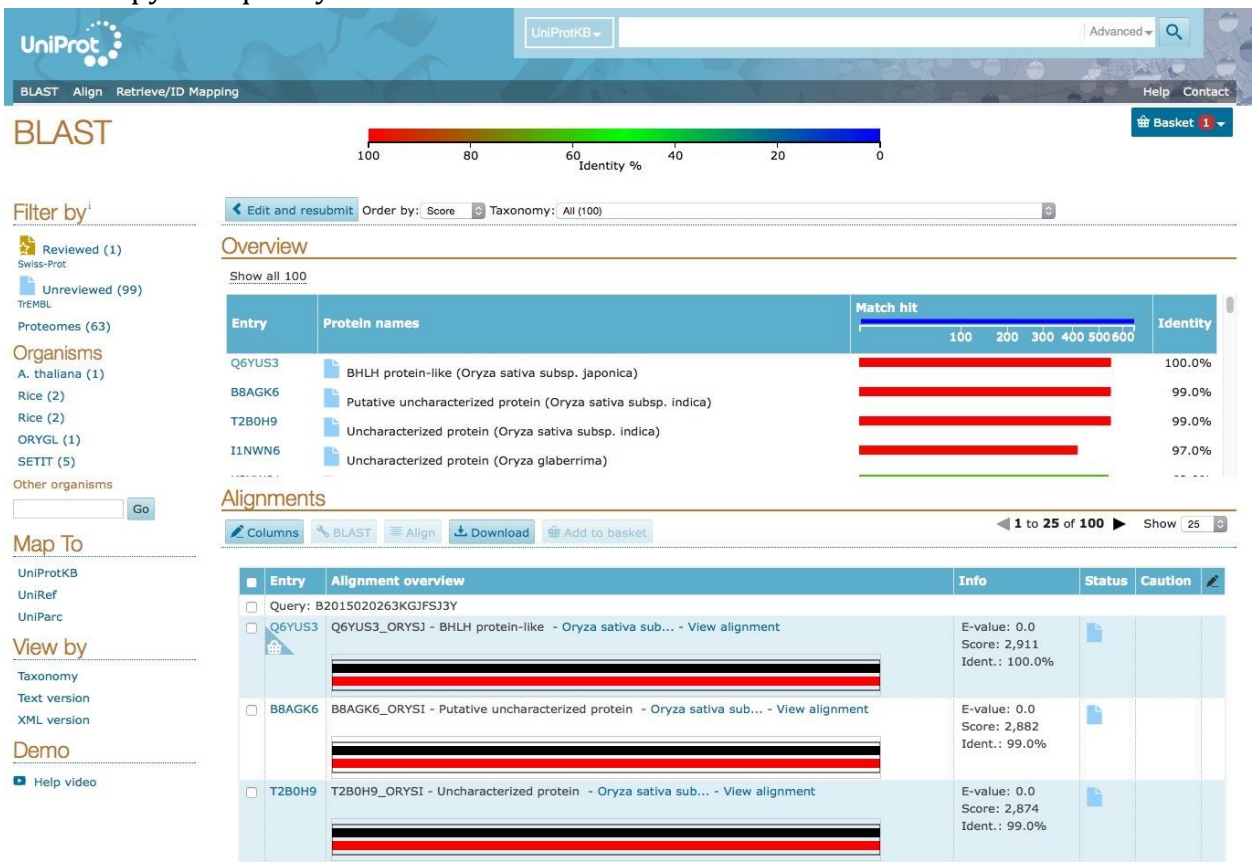
<u>Searching data about this sequenc:</u>

Search similar proteins in the Expasy database Swiss-Prot (http://www.uniprot.org).

<u>Sequence similarity analysis using BLASTP</u>

Select the tab *Blast* (BLASTP)

Copysefrequency or load the file and slee



ctionner

It is possible to select target database (use standard settings)

Launch Blast (can last from 2 to 20 minutes ...) then look at resulting *hits*:

What are resulting proteins (function, molecular weight ..? .)

What is the largest alignment coverage and the largest identity percentage obtained? Is this the same hit ?

Example of request result:

Select Q6YUS3

Explore different types of information

Open pages in new tabs

Look Particularly at

Protein Attributes

General Annotations

In the item "Cross-references":

Launch a request ModBase

Take care of DataSet Information
Open links PFAM and Protein Model Portal

PFAM
See information

Protein Model Portal
Open HLH link (new tab)
Appreciate the quality of the model

Is there a structure of this protein =Is there a PDB access number for this protein ? If so, look for the reference on the PDB (http://rcsb.org/pdb/).

Is it an experimentaly determined structure (X-ray crystallography...) if yes, you can go home.
If it is a theoretical model, you can try to do a better one.

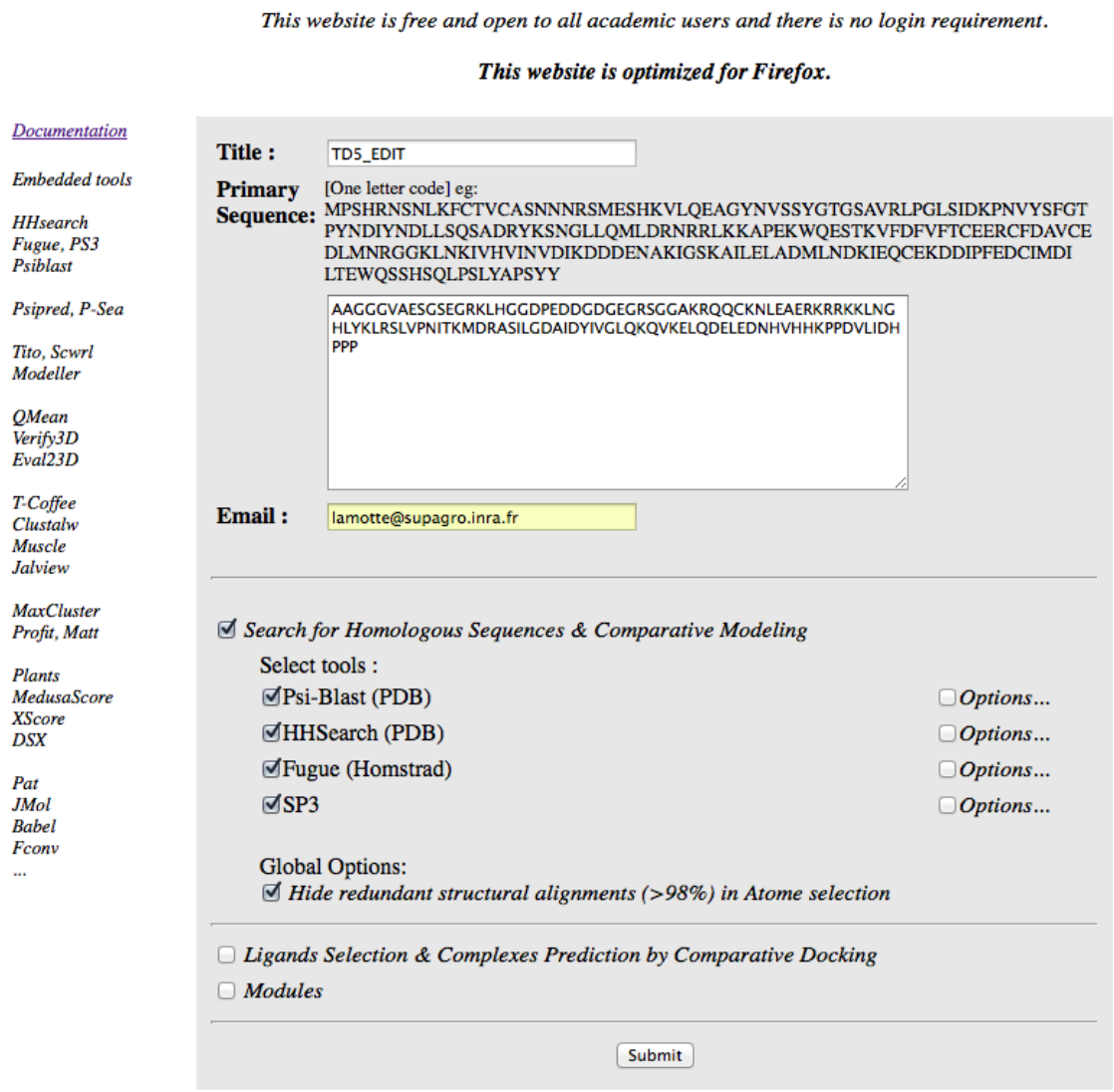
<u>Modelisation</u>

Edit the sequence to only conserve the spoted part : region of 250-360
Sequence_TD5_eDIT

atome2
http: //atome.cbs.cnrs.fr/AT2B/meta.html

enter a title
Paste the sequence
Leave email
Click all boxes of Search for homologous ...
Click on submit ... take a cafe

The next step is to validate one or more templates to start the



Ref: *Pons & Labesse - Nucleic Acids Research, Web Server Issue- 2009*

Generation of the 3D model

Download the model and the template

<u>Visualization:</u>

Start PyMOL and load model:

Is the general structure coherent ?
Try several types of representation
Rename the model

Load the template (1HLO.pdb)

Align *align PROT1, prot2*

explore PyMOL functionalities, and if necessary, consult PyMOL Wiki (http://www.pymolwiki.org/index.php/Main_Page)

Applications:.

Regarding the result model, what applications can, or can't, be tried ?

PMB
10.1007 / s11103-013-0166-5

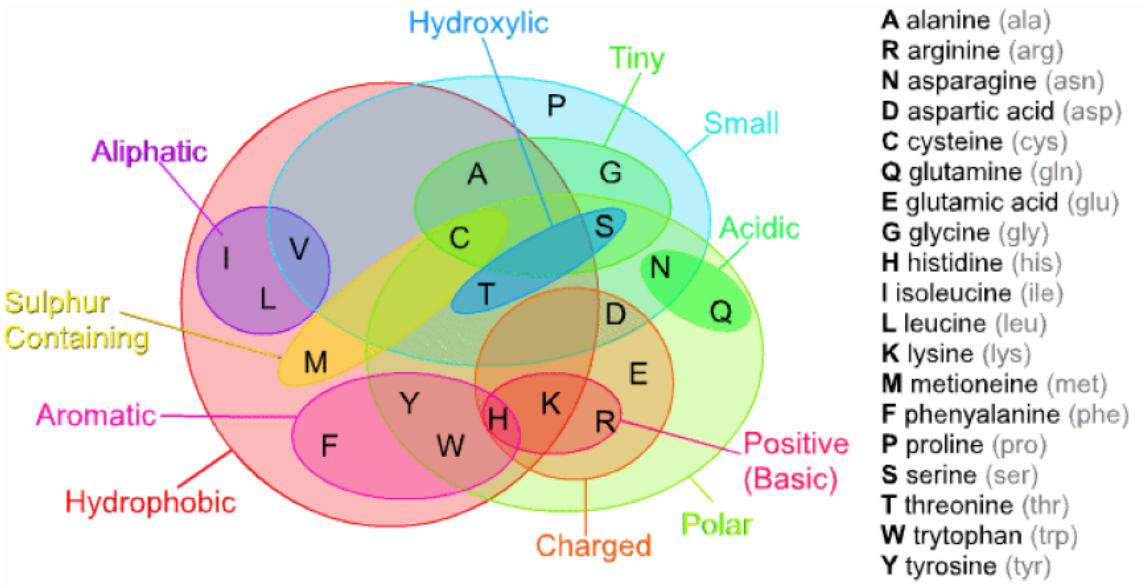
Molecular Breeding
10.1007 / s11032-013-9972-3

If you have any questions concerning this TD or thatthematic research contact lamotte@supagro.inra.fr us.

TD predbyed by Frederic de Lamotte

NB

It is important to keep in mind that the degree of similaritiess / differences between amino acids are much more complex



than