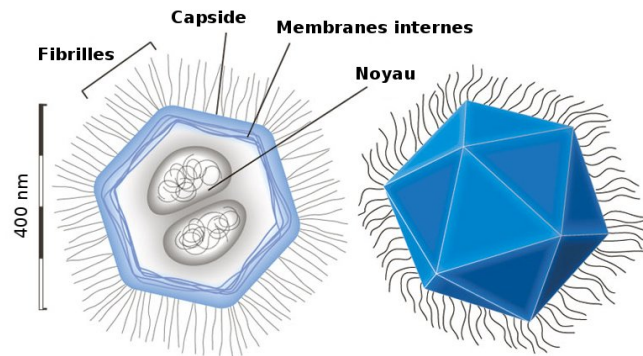# Homology modeling
# of mimivirus tyrosyl-tRNA synthetase

Mimivirus is a giant DNA virus. It is larger than many bacteria, and can itself be infected by other viruses. It was discovered that mimivirus possesses certain genes for proteins involved in translation, absent in other viruses that use the host cell's machinery to multiply. These discoveries have fuelled debate about the boundary between living and inert matter.

Homology modeling aims at building a model of the unknown structure of a target protein, knowing its sequence and the structure of another template protein of homologous sequence. The method can be decomposed into four steps :

A. **template selection**

B. **target-template alignment**

C. **model building**

D. **model evaluation**

The aim of this work is to propose the best possible structural model (criteria to be defined) of mimivirus tyrosyl-tRNA synthetase (its structure is assumed to be unknown) by homology modelling with the Modeller program.

1. Retrieving the sequence

   Retrieve the sequence of mimivirus tyrosyl-tRNA synthetase in FASTA format in UniProt (`http://www.uniprot.org`) database.

2. Template selection

   Carefully select a structure to be used as a guide for homology modeling (of course, we won't use the structure of mimivirus tyrosyl-tRNA synthetase, which we assume to be unknown). Retrieve this structure in PDB format.

3. Conversion of the sequence format

Convert query sequence from FASTA to PIR format (`http://salilab.org/modeller/manual`, File formats, Alignment file (PIR)) with which Modeller works. Example of a sequence in PIR format :

```
>P1;TvLDH
sequence:TvLDH:::::::::
MSEAAHVLITGAAGQIGYILSHWIASGELYGDRQVYLHLLDIPPAMNRLTALTMELEDCAFPHLAGFVATTDPKA
AFKDIDCAFLVASMPLKPGQVRADLISSNSVIFKNTGEYLSKWAKPSVKVLVIGNPDNTNCEIAMLHAKNLKPEN
FSSLSMLDQNRAYYEVASKLGVDVKDVHDIIVWGNHGESMVADLTQATFTKEGKTQKVVDVLDHDYVFDTFFKKI
GHRAWDILEHRGFTSAASPTKAAIQHMKAWLFGTAPGEVLSMGIPVPEGNPYGIKPGVVFSFPCNVDKEGKIHVV
EGFKVNDWLREKLDFTEKDLFHEKEIALNHLAQGG*
```

4. Inspect the files available

The Modeller program is launched as follows :

`modeller file.py`

5. Target-template alignment

Align the query sequence with the sequence of the selected guide structure by adapting the align2d.py script.

`modeller align2d.py`

The alignment produced is written in PIR, PAP, and FASTA formats. Examine these files.

6. Model building

Model by homology the target structure by adapting the model-single.py script.

`modeller model-single.py`

A summary including the PDB file names of the models produced as well as the value of the Modeller energy function and the DOPE score for each model can be found at the end of the output file (model-single.log). Examine the models produced with PyMOL.

7. Model evaluation

Evaluate the models generated in the previous step by adapting the evaluate_model.py script.

`modeller evaluate_model.py`

This script allows a more detailed evaluation of the models produced by calculating the DOPE score for each position of the alignment. Plot the DOPE score as a function of position (columns 1 and 42) for the different models.

8. How to define the best model produced ?

For further information :
Modeller tutorial : `http://salilab.org/modeller/tutorial/basic.html`
Modeller manual : `http://salilab.org/modeller/manual`