

# Tutorial para predição de estrutura de proteínas

## VII Escola de Modelagem Molecular em Sistemas Biológicos

*Priscila V. Z. Capriles Goliatt - capriles@Incc.br*

*Fábio Lima Custódio- flc@Incc.br*

*Marx G. van der Linden - marxgomes@gmail.com*

### Comandos Básicos do Linux:

cd diretório/	Entra na pasta diretório/
ls diretório/	Lista as subpastas e arquivos existentes em diretório/
cd ..	Retorna para a pasta anterior
mkdir diretório/	Cria a pasta diretório/
rm arquivo.txt	Remove arquivo.txt (.txt ou outro formato de arquivo)
rmdir diretório/	Remove a pasta diretório/
cp diretório1/arquivo.txt diretório2/	Copia arquivo.txt no diretório1/ para diretório2/

**Atenção:** veja também o tutorial básico de linux em <http://www.emmsb.Incc.br/index.php?pagina=26>

## Parte 3: Predição via método híbrido: modelagem comparativa + *ab initio*

### Exemplo 1: Modelagem com baixa taxa de identidade

#### **Passo 1: Construindo o modelo via MHOLline**

- 1 Acessar o site do MHOLline: <http://www.mholline.Incc.br>
- 2 Clicar em “Submit Your Job Now” e em “Submit Now”.
- 3 Selecionar todos os programas
- 4 Fazer o upload do arquivo FASTA (**Atenção: não submeter o job!!!!**)
- 5 Acesse o resultado através do link:

<http://www.mholline.Incc.br/result.php?id=7365816a219fc310e058a0260d299fa7>

**Atenção:** Os resultados do MHOLline já estão salvos no diretório:

~/modelagem/sodef/jobSODEF17

### Passo 4: Analisar o modelo construído via MHOLline

- 1 Visualizar o resultado do BLAST: abrir com um navegador web (e.g. firefox, iceweasel, konqueror, chrome) o arquivo ~/modelagem/sodef/jobSODEF17/jobSODEF17.html
- 2 Visualizar o resultado do HMMTOP: abrir com um editor de texto (e.g. gedit, ooffice, libreoffice) o arquivo ~/modelagem/sodef/jobSODEF17/hmmtop.out
- 3 Visualizar o resultado do BATS: abrir com um editor de texto (e.g. gedit, ooffice, libreoffice) o arquivo ~/modelagem/sodef/jobSODEF17/bats\_results.txt
- 4 Visualizar o resumo do resultado do BATS: abrir com um editor de texto (e.g. gedit, ooffice, libreoffice) o arquivo ~/modelagem/sodef/jobSODEF17/bats\_log.txt
- 5 Visualizar as estruturas classificadas em G1 no site do PDB (acesse o site: <http://www.pdb.org/pdb/home/home.do>)
- 6 Visualizar a estrutura classificadas em G1 no site do PDB (acesse o site: <http://www.pdb.org/pdb/home/home.do>)
- 7 Fazer o download das estruturas no formato FASTA e PDB.

**Atenção:** Os arquivos FASTA e PDB já estão salvos no diretório:

~/modelagem/sodef/sequences

### Passo 3: Construindo o modelo via Modeller

### 5.1. Modelagem baseada em mais de um molde (template):

## Preparando os arquivos de entrada:

- 1 Em um terminal, entre no diretório ~/modelagem/sodef/
- 2 Digite na linha de comando do terminal: mod9.14 readseq.py

**Atenção:** irá aparecer a mensagem: 'import site' failed; use -v for traceback.

## Não se preocupe!

- 3 Realizar o alinhamento múltiplo entre os possíveis templates obtidos com o MHOLline e a sequência alvo através do site: <http://www.ebi.ac.uk/Tools/msa/clustalo/>

```
>P1; ./sequences/SODEF17
```

```
sequence:../sequences/SODEF17::::::::::
```

HTPTPTPICKSRSEYKGRCIQDMDCNAACVKESSESYTGGFCNGRPPFKQCFCTKPCRERAAATLRWPGL\*

- 4 Na opção “OUTPUT FORMAT”, selecionar o formato “PEARSON” (formato usado pelo MODELLER) e em “OUTPUT ORDER”, selecionar input
- 5 Abaixo, colar as seqüências FASTA dos templates e da seqüência alvo contidas no arquivo:  
`/home/vemmsb/modelagem/sodef/sequences/multi.fasta`
- 6 Clicar em “RUN”, visualizar o resultado e salvar o arquivo de resultado no formato ALN.

- 7 Edite o arquivo 4AAZ.pdb e 4AAZ.seg removendo todas as informações referentes a cadeia B. Salve como 4AAZ\_A.pdb e 4AAZ\_A.seg. Completar o cabeçalho de cada sequência como o obtido no formato.seg
- 8 Replique os heteroátomos (.) e as águas (w) para a entrada da SODEF17 que passará a ter o seguinte formato:

- 9 Salve o arquivo como SODEF17.ali

### Gerando o modelo da proteína alvo 3A9J com o Modeller:

- Atenção:** irá aparecer a mensagem: 'import site' failed; use -v for traceback. Não se preocupe!
- 2 O *script* model-multi.py lê o arquivo SODEF17.ali, gera três modelos para a proteína alvo SODEF17 através da função automodel(), lista dentro do arquivo model-multi.log a pontuação dos modelos gerados de acordo com a função de energia molpdf e com os métodos DOPE (*Discrete Optimized Protein Energy*) e GA341, e retorna o nome da melhor estrutura gerada de acordo com a pontuação do DOPE. Através da função cluster(cluster\_cut=1.00), clusteriza os modelos gerados e gera uma estrutura representativa já otimizada via Gradiente Conjugado (esta função é interessante para quando se deseja gerar uma grande quantidade de modelos). O resumo dos resultados é gerado dentro do arquivo model-multi.out.

## Avaliando os modelos 3D gerados:

- 4 Observe os resultados do DOPE, DOPE normalizado e GA341 para os modelos e para a estrutura representativa (cluster.opt).

- 5 Via terminal, entre no diretório ./sequences/multiplt/
- 6 Digite na linha de comando do terminal:

vmd -m SODEF17.B9999000\*.pdb cluster/cluster.opt ../\*.pdb

7 Alinhe todas as estruturas.

8 Compare todos os modelos com a estrutura dos templates

#### **Validando os modelos gerados via Molprobability:**

9 Avaliar os modelos gerados no diretório ./sequences/multiopt/ e no diretório ./sequences/multiopt/

10 Entre no diretório ./sequences/

11 Digite na linha de comando do terminal:  
evince ./multiopt/molprobabilityresults/\*.pdf &

12 Compare os resultados dos modelos gerados com as estruturas usadas como templates.

### **OPS! AS REGIÕES C- E N-TERMINAL NÃO ESTÃO MODELADAS. O QUE EU FAÇO AGORA?**

#### **Passo 6: Procurando por região de peptídeo sinal e transmembrana:**

1 Acessar o site do SignalP: <http://www.cbs.dtu.dk/services/SignalP/>

2 Colar a sequência FASTA da SODEF17 e definir os parâmetros de busca ( Output format = All / N-terminal truncation of input sequence = 0).

3 Analisar os resultados no link: <http://www.cbs.dtu.dk/cgi-bin/webface2.fcgi?jobid=53F4D1EF00006A8329CC1409&wait=20> (ou veja o arquivo ~/modelagem/sodef/sequences/SODEF17\_signalp.pdf)

4 Acessar o site do TMHMM: <http://www.cbs.dtu.dk/services/TMHMM/>

5 Colar a sequência FASTA da SODEF17

6 Analisar os resultados no link: <http://www.cbs.dtu.dk/cgi-bin/webface2.fcgi?jobid=53F4D8380000083D6D8C8687&wait=20> (ou veja o arquivo ~/modelagem/sodef/sequences/SODEF17\_tmhmm.pdf)

### **OK! NÃO SÃO PEPTÍDEOS SINAIS NEM PORÇÕES TRANSMEMBRANARES. O QUE EU FAÇO AGORA?**

#### **Passo 7: Procurando por conservação de estrutura secundária**

1 Acessar os sites do Psipred, Juford, Netsurf (e/ou outro preditor de estrutura secundária) e faça a predição da sequência da SODEF17

2 Veja os resultados nos arquivos salvos: gedit ~/modelagem/sodef/ssprediction/\*

3 Observe que a região de consenso predita na porção terminal da proteína é para uma hélice na região 58-64.

#### **Passo 7: Impondo restrições no modeller**

1 Abrir o arquivo ~/modelagem/sodef/model-multiss.py

- 2 Observe a sintaxe da classe MyModel
- 3 Observe a imposição de restrição de hélice na região 58-64.
- 4 Digite na linha de comando do terminal: mod9.10 model-multi.py

**Atenção:** irá aparecer a mensagem: 'import site' failed; use -v for traceback. Não se preocupe!

**Atenção:** Os modelos gerados (SODEF17.B99990001.pdb, SODEF17.B99990002.pdb e SODEF17.B99990003.pdb) já encontram-se no diretório ./sequences/multiptss/

**Atenção:** A estrutura representativa inicial (cluster.ini) e otimizada (cluster.opt) encontram-se no diretório ./sequences/multiptss/cluster/

- 5 Via terminal, entre no diretório ./sequences/multiptss/
- 6 Digite na linha de comando do terminal:
- 7 vmd -m SODEF17.B9999000\*.pdb cluster/cluster.opt ../\*.pdb
- 8 Alinhe todas as estruturas.
- 9 Compare todos os modelos com a estrutura dos templates

Observe que ainda é necessário observar restrições de contato para determinar a posição da hélice. Veja o resultado de um exemplo de preditor (NNCon) com o comando gedit ~/modelagem/sodef/ssprediction/nncon.txt

Para inserir restrições de contato veja exemplo do arquivo ~/modelagem/outrosexemplos/exemplo\_ssrestraint.py