

# Tutorial para predição de estrutura de proteínas

---

## V Escola de Modelagem Molecular em Sistemas Biológicos

Priscila V. Z. Capriles Goliatt – [capriles@lncc.br](mailto:capriles@lncc.br)

Fábio Lima Custódio- [flc@lncc.br](mailto:flc@lncc.br)

### Comandos Básicos do Linux:

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

**\* Neste curso os arquivos já estão salvos pastas mencionadas no Tutorial.**

## Parte 1: Predição via modelagem comparativa usando MHOLline e Modeller

### Introdução

Este tutorial está orientado para usuários do sistema operacional Linux e tem por objetivo exemplificar a aplicação do portal MHOLline (<http://www.mholline.lncc.br>) e do software Modeller (<http://www.salilab.org/modeller/>), para predição de estruturas de proteínas (PSP) via modelagem comparativa (MC). O MHOLline é um *workflow* que combina uma série de programas usados para análise de função de proteínas e PSP via MC, em grande escala. Neste tutorial ele será usado na obtenção de modelos tridimensionais (3D) de proteínas, que servirão como ponto de partida para refinamentos das mesmas via Modeller.

*Para maiores detalhes sobre o funcionamento do MHOLline acesse a página:  
<http://www.mholline.lncc.br/index.php?pag=14>*

*Para maiores detalhes sobre o funcionamento do Modeller acesse a página:  
<http://www.salilab.org/modeller/documentation.html>*

---

## Organização do tutorial

O tutorial está organizado em exemplos de aplicação e é importante que seja seguida a ordem para melhor compreensão. Os textos que estiverem escritos com a seguinte formatação:

**\$ cd modelagem/**

São comandos para serem digitados em uma janela de terminal (e.g. konsole, xterm).

## Exemplo 1: Modelagem com alta taxa de identidade

### Passo 1: Identificando a sequência

- 1.1 Acessar o site do NCBI: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- 1.2 Selecionar a ferramenta “protein blast”.
- 1.3 No campo “Enter Query Sequence”, colar a sequência que se deseja modelar (no formato FASTA)<sup>1</sup>:

```
>TESTE
```

```
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLLGG
```

**Atenção:** Neste curso o arquivo FASTA já está salvo em:  
‘/home/vemmsb/modelagem/basico/sequences/1UBQ.fasta’

- 1.4 Buscar pela sequência usando os seguintes parâmetros:
  - 1.4.1 “Database”: Protein Data Bank proteins (pdb)
  - 1.4.2 “Algorithm”: blastp
  - 1.4.3 “Expected Threshold”: 0.0001
  - 1.4.4 “Matrix”: BLOSUM62
- 1.5 Analisar os resultados.

#### <sup>1</sup>Arquivo fasta

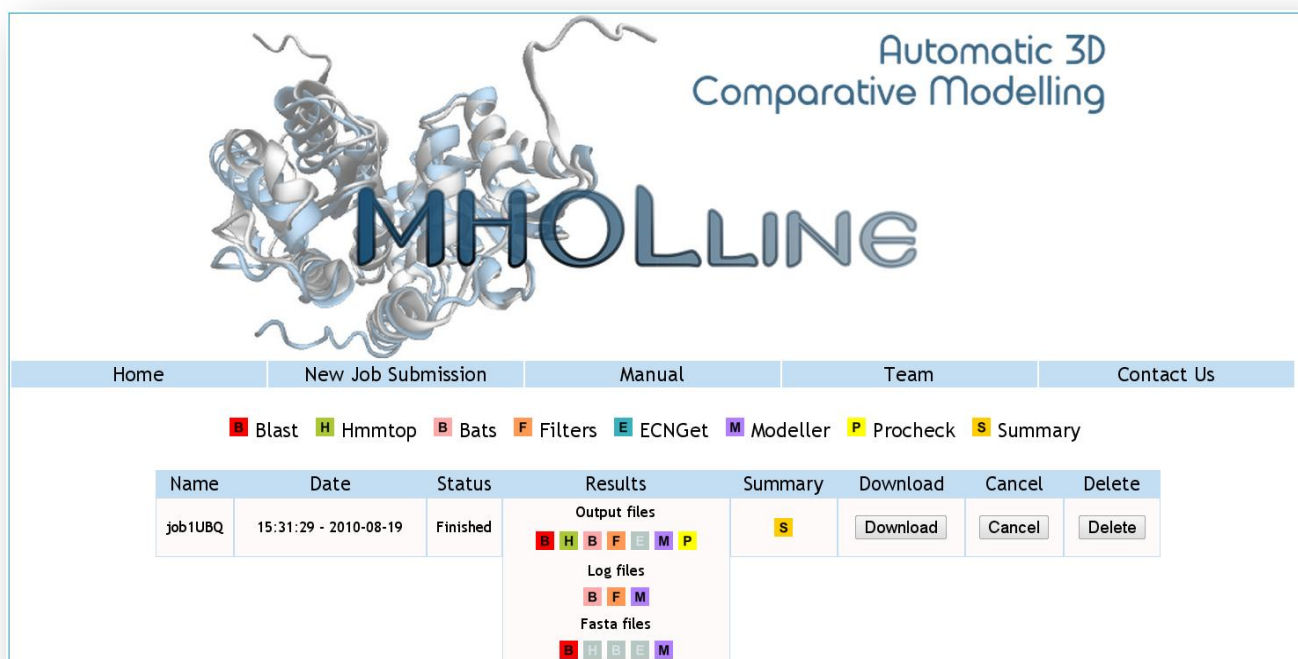
*O arquivo FASTA contém a sequência de resíduos de aminoácidos, e pode ser obtido de diversos bancos de sequências, ou mesmo editado pelo usuário. É composto por uma primeira linha sempre iniciada por “>” seguida do nome (identificação) da sequência e por uma segunda linha contendo a sequência propriamente dita.*

---

## Passo 2: Construindo o modelo via MHOLline

- 2.1 Acessar o site do MHOLline: <http://www.mholline.lncc.br>
- 2.2 Clicar em “Submit Your Job Now” e em “Submit Now”.
- 2.3 Selecionar todos os programas
- 2.4 Fazer o upload do arquivo FASTA (**Atenção: não submeter o job!!!!**)
- 2.5 Acesse o resultado através do link:

<http://www.mholline.lncc.br/result.php?id=1b5bc280afca10d53ad46a4deb6732af>



The screenshot displays the MHOLline web interface. At the top, there is a 3D protein structure and the text "Automatic 3D Comparative Modelling" and "MHOLLINE". Below this is a navigation bar with links: Home, New Job Submission, Manual, Team, and Contact Us. A row of colored buttons represents different tools: Blast (red), Hmmtop (green), Bats (pink), Filters (orange), ECNGet (teal), Modeller (purple), Procheck (yellow), and Summary (light blue). Below the navigation bar is a table showing job submissions. The table has columns: Name, Date, Status, Results, Summary, Download, Cancel, and Delete. The first row shows a job named "job1UBQ" with a date of "15:31:29 - 2010-08-19" and a status of "Finished". The "Results" column for this job contains a list of output files categorized by tool: Output files (B, H, B, F, E, M, P), Log files (B, F, M), and Fasta files (B, H, B, E, M). The "Summary" column shows a yellow "S" button. The "Download" column has a "Download" button, and the "Cancel" column has a "Cancel" button. The "Delete" column has a "Delete" button.

**Atenção:** Os resultados do MHOLline já estão salvos no diretório:  
'/home/vemmsb/modelagem/basico/job1UBQ'

## Passo 3: Analisar o modelo construído via MHOLline

- 3.1 Visualizar o resultado do BLAST: abrir com um navegador web (e.g. firefox, iceweasel, konqueror, chrome) o arquivo /home/vemmsb/modelagem/basico/job1UBQ/job1UBQaa.html
- 3.2 Visualizar o resultado do HMMTOP: abrir com um editor de texto (e.g. gedit, ooffice) o arquivo /home/vemmsb/modelagem/basico/job1UBQ/hmmtop.out
- 3.3 Visualizar o resultado do BATS: abrir com um editor de texto (e.g. gedit, ooffice) o arquivo /home/vemmsb/modelagem/basico/job1UBQ/bats\_results.txt
- 3.4 Visualizar o resumo do resultado do BATS: abrir com um editor de texto (e.g. gedit, ooffice) o arquivo /home/vemmsb/modelagem/basico/job1UBQ/bats\_log.txt

- 3.5 Visualizar o resultado do FILTERS: abrir com um editor de texto (e.g. gedit, ooffice) o arquivo /home/vemmsb/modelagem/basico/job1UBQ/filters\_log.txt
- 3.6 Visualizar a estrutura do molde (*template*) encontrado pelo MHOline (acesse o site: <http://www.pdb.org/pdb/home/home.do>)
- 3.6.1 Em “Search” digite o nome da estrutura 1OGW.
- 3.6.2 Analisar os dados da estrutura (e.g. Espécie, Método Experimental, Resolução, Ligantes).
- 3.6.3 Clicar em “Sequence”: Analisar a estrutura secundária e verificar a existência ou não de ligações particulares (e.g. ligações dissulfídicas).
- 3.7 Visualizar a estrutura do modelo gerado via MHOline: abrir com um visualizador 3D o arquivo /home/vemmsb/modelagem/basico/job1UBQ/Modeller/1\_1UBQAPDBIDCH.B99990001.pdb:
- 3.7.1 Digite na linha de comando de um terminal:
- ```
vmd -m 1_1UBQAPDBIDCH.B99990001.pdb 1OGW_A.atm
```
- 3.7.2 Para mudar a visualização: Clique em “Graphics → Representations”, em “Coloring Method” use preferencialmente a opção “Secondary Structure” e em “Drawing Method” use preferencialmente a opção “New Cartoon”.
- 3.7.3 Alinhar as estruturas 3D:
- Clique em: “Extensions” → “Analysis” → “MultiSeq”.
  - Selecionar: “VMD Protein Structures”.
  - Clicar em: “Tools” → “Stamp Structural Alignment” → “Marked Structures” → “OK”.
- 3.7.4 Use também outros visualizadores como rasmol e pymol.
- 3.8 Visualizar a qualidade da estrutura avaliada com o programa Procheck: abrir a figura /home/vemmsb/modelagem/basico/job1UBQ/Procheck/1\_1UBQAPDBIDCH.B99990001\_01.ps
- 3.8.1 Para abrir o gráfico de Ramachandran gerado, digite na linha de comando de um terminal:
- ```
gv 1_1UBQAPDBIDCH.B99990001_01.ps&
```

**Observação:** Para analisar a qualidade do modelo gerado, experimente também os seguintes programas online:

**MolProbity:** <http://molprobity.biochem.duke.edu/index.php>

**Qmean Server:** <http://swissmodel.expasy.org/qmean/cgi/index.cgi?>

**ProSA-Web:** <https://prosa.services.came.sbg.ac.at/prosa.php>

**Verify3D:** [http://nihserver.mbi.ucla.edu/Verify\\_3D/](http://nihserver.mbi.ucla.edu/Verify_3D/)

---

## Passo 4: Construindo o modelo via Modeller

### *Modelagem baseada em um único molde (template):*

O modeller trabalha basicamente com arquivos de entrada no formato PIR e *scripts* em linguagem python.

```
>P1;1_1UBQ
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYNIQKESTLHLVLRRLGG*
```

Exemplo de arquivo no formato PIR.

```
from modeller import *

log.verbose()
env = environ()
env.io.hetatm = env.io.water = True

code = './sequences/1UBQ'
mdl = model(env, file=code)
aln = alignment(env)
aln.append_model(mdl, align_codes=code)
aln.write(file=code+'.seg')
```

Exemplo de script em linguagem python.

### Preparando os arquivos de entrada:

- 4.1 Em um terminal, entre no diretório /home/vemmsb/modelagem/basico/modeller
- 4.2 Digite na linha de comando do terminal: mod9v8 readseq.py  
**Atenção:** irá aparecer a mensagem: 'import site' failed; use -v for traceback. Não se preocupe!
- 4.3 O *script* readseq.py lê a estrutura PDB e gera arquivo.seg no formato PIR.
- 4.4 Abra com um editor de texto o arquivo ./sequences/1UBQ.seg e observe a formatação.
- 4.5 Copie o arquivo ./sequences/1UBQ.fasta e para ./sequences/alvo.seg.
- 4.6 Editar o arquivo ./sequences/alvo.seg para ficar no formato de entrada do Modeller.

```
>P1;./sequences/alvo
sequence:1UBQ:::::::::
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYNIQKESTLHLVLRRLGG*
```

Exemplo de arquivo no formato.seg do Modeller.

### Gerando o alinhamento entre 1UBQ.seg e alvo.seg com o Modeller:

4.7 Digite na linha de comando do terminal: `mod9v8 align2d.py`

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

4.8 O *script* `align2d.py` lê os arquivos `1UBQ.seg` e `alvo.seg` e gera os arquivos `1UBQ_alvo.pap` e `1UBQ_alvo.ali`. O primeiro contém o alinhamento entre as duas seqüências no formato ALN e o segundo contém o alinhamento no formato de entrada do Modeller.

4.9 Abra esses dois arquivos com um editor de texto e observe as diferenças.

### Gerando o modelo da proteína alvo 1UBQ com o Modeller:

4.10 Digite na linha de comando do terminal: `mod9v8 model-single.py`

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

4.11 O *script* `model-single.py` lê o arquivo `1UBQ_alvo.ali`, gera três modelos para a proteína alvo 1UBQ através da função `automodel()`, lista dentro do arquivo `model-single.log` a pontuação dos modelos gerados de acordo com os métodos DOPE (*Discrete Optimized Protein Energy*) e GA34, e retorna o nome da melhor estrutura gerada de acordo com a pontuação do DOPE.

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

### Avaliando os modelos 3D gerados:

4.12 Abra com um editor de texto o arquivo `/home/vemmsb/modelagem/basico/modeller/model-single.log`

4.13 Vá até o final do arquivo e observe os resultados do DOPE.

4.14 Via terminal, entre no diretório `./sequences/singleinit/`

4.15 Digite na linha de comando do terminal: `vmd -m alvo.B9999000*.pdb. ../1UBQ.pdb`

4.16 Alinhe todas as estruturas: (proceda como no item 3.7.3).

### Gerando os modelos otimizados da proteína alvo 1UBQ com o Modeller

(Usando o método *Variable Target Function Method - VTFM*):

4.17 Digite na linha de comando do terminal: `mod9v8 model-single-opt.py`

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

4.18 O *script* `model-single-opt.py` lê o arquivo `1UBQ_alvo.ali` e gera três modelos otimizados para a proteína alvo 1UBQ através da função `automodel()`.

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

#### Avaliando os novos modelos 3D gerados:

- 4.19 Abra com um editor de texto o arquivo /home/vemmsb/modelagem/basico/modeller/model-single-opt.log
- 4.20 Vá até o final do arquivo e observe os resultados do DOPE.
- 4.21 Via terminal, entre no diretório ./sequences/singleopt/
- 4.22 Digite na linha de comando do terminal: vmd -m alvo.B9999000\*.pdb. ../1UBQ.pdb
- 4.23 Alinhe todas as estruturas: (proceda como no item 3.7.3).

#### Validando os modelos gerados via Procheck:

- 4.24 Avaliar os modelos gerados no diretório ./sequences/singleinit/ e no diretório ./sequences/singleopt/
- 4.25 Entre no diretório ./sequences/singleinit/procheckresult/
- 4.26 Digite na linha de comando do terminal: procheck ../alvo.B99990001.pdb 1.5

**Observação:** o número 1.5 refere-se à resolução para a análise do modelo.

- 4.27 Este comando irá gerar o gráfico de Ramachandran em alvo.B99990001\_01.ps
- 4.28 Agora entre no diretório ./sequences/singleopt/procheckresult/
- 4.29 Digite na linha de comando do terminal: procheck ../alvo.B99990003.pdb 1.5
- 4.30 Este comando irá gerar o gráfico de Ramachandran em alvo.B99990003\_01.ps

**Atenção:** Os resultados já encontram-se nos diretórios:  
./sequences/singleinit/procheckresult e  
./sequences/singleopt/procheckresult

#### Analisando o gráfico de Ramachandran:

- 4.31 Abra os gráficos de Ramachandran gerados para a análise dos resultados.
- 4.32 Digite na linha de comando do terminal:  
  
gv ./sequences/singleinit/procheckresult/ alvo.B99990001\_01.ps&
- 4.33 Digite na linha de comando do terminal:





- 6.11 Vá até o final do arquivo e observe os resultados do DOPE.
- 6.12 Via terminal, entre no diretório `./sequences/singleinit/`
- 6.13 Digite na linha de comando do terminal: `vmd -m ../1UBQ.pdb ../1OGW.pdb alvo.B99990003.pdb`
- 6.14 Alinhe todas as estruturas: (proceda como no item 3.7.3).

#### Validando os modelos gerados via Procheck:

- 6.15 Avaliar os modelos gerados no diretório `./sequences/singleinit/`
- 6.16 Entre no diretório `./sequences/singleinit/procheckresult/`
- 6.17 Digite na linha de comando do terminal: `procheck ../alvo.B99990003.pdb 1.5`
- 6.18 Este comando irá gerar o gráfico de Ramachandran em `alvo.B99990003_01.ps`

**Atenção:** Os resultados já encontram-se no diretório `./sequences/singleinit/procheckresult`

#### Analisando o gráfico de Ramachandran:

- 6.19 Abra os gráficos de Ramachandran gerados para a análise dos resultados.
- 6.20 Digite na linha de comando do terminal:  
  
`gv ./sequences/singleinit/procheckresult/alvo.B99990003_01.ps&`
- 6.21 Digite na linha de comando do terminal:

`gv /home/vemmsb/modelagem/basico/job1UBQ/Procheck/1_1UBQAPDBIDCH.B99990001_01.ps&`

Repare que após a modelagem considerando as águas do molde (*template*) o resíduo LEU73 antes em região energeticamente não permitida foi incluído na região energeticamente mais favorável (aumentando de 93,9% para 95,5%).