

# Tutorial para predição de estrutura de proteínas

## VII Escola de Modelagem Molecular em Sistemas Biológicos

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

*Fábio Lima Custódio- flc@lncc.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.lncc.br/index.php?pagina=26>

## 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>**

## Exemplo 1: Modelagem com alta taxa de identidade

### Passo 1: Identificando a sequência via NCBI

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

```
>TESTE  
MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQORLIFAGKQLEDGRTLSDYNIQKESTLHLVLRRLGG
```

**Atenção:** Neste curso o arquivo FASTA já está salvo em:

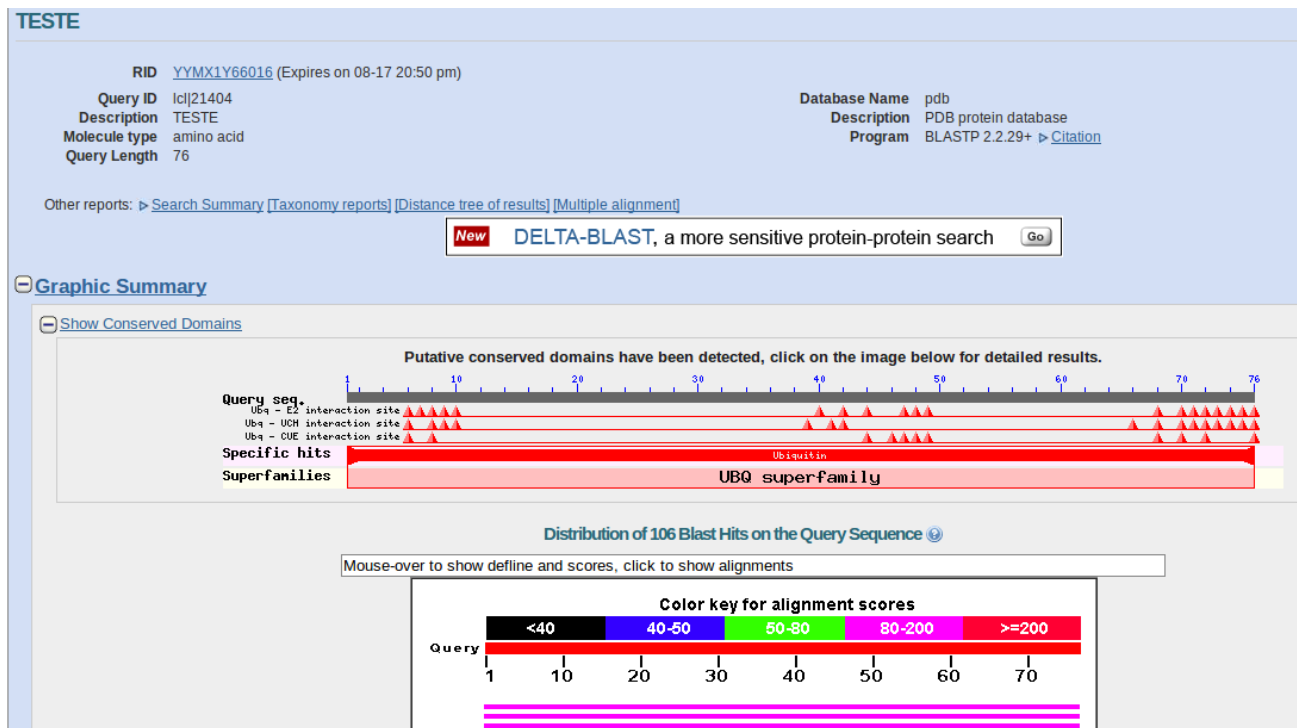
***./modelagem/basico/sequences/teste\_emmsb.fasta***

- 4 Buscar pela sequência usando os seguintes parâmetros:
  - 4.1 “Database”: Protein Data Bank proteins (pdb)
  - 4.2 “Algorithm”: blastp
- 5 Clicar em “Algorithm parameters”
  - 5.1 “Expected Treshold”: 0.0001
  - 5.2 “Matrix”: BLOSUM62
  - 5.3 “Filter”: Low complexity regions

The screenshot shows the NCBI BLAST web interface. The 'Database' is set to 'Protein Data Bank proteins(pdb)'. The 'Algorithm' is set to 'blastp (protein-protein BLAST)'. The 'Expect threshold' is set to '0.0001'. The 'Matrix' is set to 'BLOSUM62'. The 'Filter' is set to 'Low complexity regions'.

- 1 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.

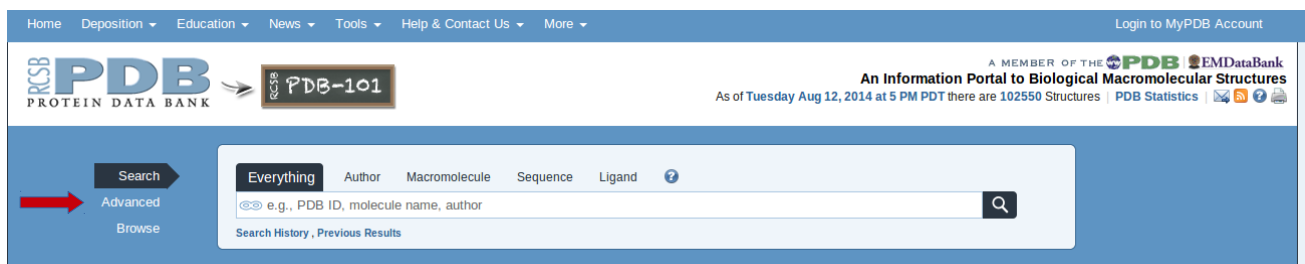
## 6 Analisar os resultados.



## Passo 2: Identificando a sequência via PDB

7 Acessar o site do PDB: <http://www.pdb.org/pdb/home/home.do>

8 Selecionar a ferramenta “Advanced”, para realizar uma busca avançada.



9 No campo “Choose a Query Type”, selecionar a opção “Sequence (BLAST/FASTA/PSI-BLAST)” e no campo “Sequence” colar a sequência que se deseja modelar (no formato FASTA)<sup>1</sup>.

10 Search Tool: BLAST

11 Mask Low Complexity: Yes

12 E-Value Cutoff: 0.0001

Advanced Search Interface

Sequence (BLAST/FASTA/PSI-BLAST)

Sequence search (BLAST or FASTA)

Structure Id

Chain Id

Sequence

Search Tool

Mask Low Complexity

E-Value Cutoff

Sequence Identity Cutoff (%)

Result Count

MQIFVKLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLI  
FAGKQLEDGRITLSDYNIQKESTLHLVLRGG

BLAST

Yes

0.0001

0

13 Compare o resultado obtido no PDB com o obtido pelo NCBI.

### Passo 3: Construindo o modelo via MHOLline

- 1 Acessar o site do MHOLline: <http://www.mholline.lncc.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!!!!**)

Automatic 3D Comparative Modelling

# MHOLLINE

Home New Job Submission Manual Team Contact Us

In "New Job Submission" area, you can submit at most 50 sequences per FASTA file. [Submit Now!](#)

In "My MHOLline" area, you can submit more than 50 sequences per FASTA file. [Access Now!](#)

More details in how to submit a new job see "MHOLline Tutorial". [Watch Video Now!](#)

**ATTENTION: MHOLline is not yet working on Google Chrome!!!**

<input type="checkbox"/>	Program	Program description
<input checked="" type="checkbox"/>	BLAST	A tool to find regions of similarity between biological sequences.
<input checked="" type="checkbox"/>	HMMTOP	Prediction of protein transmembrane regions using hidden Markov model.
<input checked="" type="checkbox"/>	BATS	A method to analyse BLAST results, defining the best template to be used in the comparative modelling.
<input checked="" type="checkbox"/>	FILTERS	A tool to group sequences according to BATS score getting information about model quality.
<input checked="" type="checkbox"/>	ECNGET	A tool that associates at least one Enzyme Commission (EC) number to each sequence that can be modelled. The EC number is assigned according to the PDB template.
<input checked="" type="checkbox"/>	MODELLER	Production of 3D homology models of proteins structures.
<input checked="" type="checkbox"/>	PROCHECK	Protein structure validation program.

E-mail Address (optional):

Upload your FASTA file  Nenhum arquivo selecionado

5 Acesse o resultado através do link:

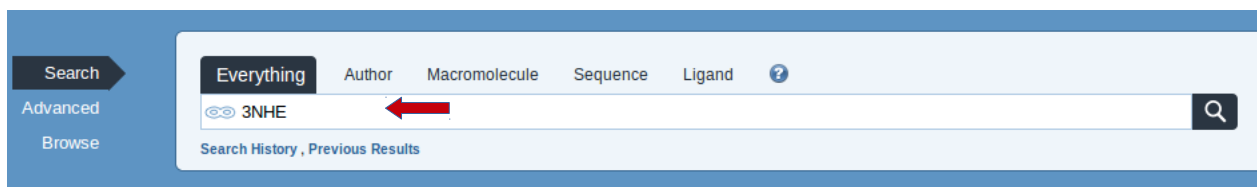
<http://www.mholline.lncc.br/result.php?id=9f75b17eb38fc3ca0396efa677ed22f0>

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

./modelagem/basico/jobteste\_emmsb

#### 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/basico/jobteste\_emmsb/jobteste\_emmsb.html
- 2 Visualizar o resultado do HMMTOP: abrir com um editor de texto (e.g. gedit, ooffice, libreoffice) o arquivo ./modelagem/basico/jobteste\_emmsb/hmmtop.out
- 3 Visualizar o resultado do BATS: abrir com um editor de texto (e.g. gedit, ooffice, libreoffice) o arquivo ./modelagem/basico/jobteste\_emmsb/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/basico/jobteste\_emmsb/bats\_log.txt
  - 4.1 Compare o molde escolhido pelo MHOLline com os escolhidos pelo NCBI e PDB.
- 5 Visualizar o resultado do FILTERS: abrir com um editor de texto (e.g. gedit, ooffice, libreoffice) o arquivo ./modelagem/basico/jobteste\_emmsb/filters\_log.txt
- 6 Visualizar a estrutura do molde (*template*) encontrado pelo MHOLline (acesse o site: <http://www.pdb.org/pdb/home/home.do>)
  - 6.1 Em “Search” digite o nome da estrutura 3NHE.



- 6.2 Analisar os dados da estrutura (e.g. Espécie, Método Experimental, Resolução, Ligantes).
- 6.3 Clicar em “Sequence”: Analisar a estrutura secundária e verificar a existência ou não de ligações particulares (e.g. ligações dissulfídicas).

Summary 3D View **Sequence** Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Links

High Resolution Structure (1.26Å) of USP2a in Complex with Ubiquitin

DOI:10.2210/pdb3nhe/pdb

Primary Citation

High resolution structure of USP2a in complex with ubiquitin

Schormann, N., Powell McCombs, D., DeLucas, L.J.

Journal: To be Published

PubMed ID is not available

Molecular Description

Classification: Hydrolase/protein Binding

Structure Weight: 48857.29

Molecule: Ubiquitin carboxyl-terminal hydrolase 2

Biological Assembly

- 7 Visualizar a estrutura do modelo gerado via MHOLline: abrir com um visualizador 3D o arquivo ./modelagem/basico/jobteste\_emmsb/Modeller/1\_teste76letters.B99990001.pdb:
  - 7.1 Digite na linha de comando de um terminal:

```
vmd -m 1_teste76letters.B99990001.pdb 3NHE_B.atm
```

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

- 7.2 Para mudar a visualização no vmd: Clique em “Graphics → Representations”, em “Coloring Method” use preferencialmente a opção “Secondary Structure” e em “Drawing Method” use preferencialmente a opção “New Cartoon”.
- 7.3 Alinhar as estruturas 3D:
- 7.3.1 Clique em: “Extensions” → “Analysis” → “MultiSeq”.
  - 7.3.2 Selecionar: “VMD Protein Structures”.
  - 7.3.3 Clicar em: “Tools” → “Stamp Structural Alignment” → “Marked Structures” → “OK”.
- 7.4 Use também outros visualizadores como rasmol ou pymol.
- 8 Visualizar a qualidade da estrutura avaliada com o programa Procheck: abrir a figura ./modelagem/basico/jobteste\_emmsb/Procheck/1\_teste76letters.B99990001\_01.ps
- 8.1 Para abrir o gráfico de Ramachandran gerado, digite na linha de comando de um terminal: okular 1\_teste76letters.B99990001\_01.ps&
- 9 Avaliar a qualidade da estrutura com o programa Molprobit (acesse o site: <http://molprobit.biochem.duke.edu/>):
- 9.1 Na página principal, na seção “File Upload”, carregue o arquivo do modelo gerado pelo MHOLline e salve em:  
./modelagem/basico/jobteste\_emmsb/Modeller/1\_teste76letters.B99990001.pdb
  - 9.2 Na nova janela, clique em “Continue”.
  - 9.3 Na seção “Suggested Tools”, clique em “Analyze geometry without all-atom contacts”.
  - 9.4 Dentro da opção “3-D kineimage graphics”, selecione as saídas do tipo “Clashes”, “Hydrogen bonds” e “van der Waals contacts”. Em “Charts, plots, and tables” selecione a opção “Clashes & clashscore”.
  - 9.5 Clique em “Run programs to perform these analysis”. Clique em OK.
  - 9.6 No final da nova página clique em “Continue”.
  - 9.7 Na nova página, desça até “Recently Generated Results”, e na opção “POPULAR DOWNLOADS” clique em “(ALL DOWNLOADS)”.
  - 9.8 Na nova página, clique em “Check all” e faça o download.  
**Atenção:** Os resultados do Molprobit já estão salvos no diretório:  
./modelagem/basico/jobteste\_emmsb/Molprobit
  - 9.9 Para abrir o gráfico de Ramachandran gerado, digite na linha de comando de um terminal: okular alvo.B99990001.pdf&

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

**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 5: Construindo o modelo via Modeller

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

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

```
>P1;3NHE
MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQORLIFAGKQLEDGRTLSDYNIQKESTLHLVLRRLGG*
```

Exemplo de arquivo no formato PIR.

```
from modeller import *

log.verbose()

env = environ()

env.io.hetatm = env.io.water = True

code = './sequences/3NHE'

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:

- 1 Em um terminal, entre no diretório ./modelagem/basico/modeller
- 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 O *script* readseq.py lê a estrutura PDB e gera arquivo.seg no formato PIR.
- 4 Abra com um editor de texto o arquivo ./sequences/3NHE.seg e observe a formatação.
- 5 Copie o arquivo ./sequences/teste\_emmsb.fasta para ./sequences/alvo.seg.
- 6 Editar o arquivo ./sequences/alvo.seg para ficar no formato de entrada do Modeller.

```
>P1;./sequences/alvo
sequence:teste_emmsb:::::::::
MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQORLIFAGKQLEDGRTLSDYNIQKESTLHLVLRRLGG*
```

Exemplo de arquivo no formato.seg do Modeller.



### Gerando o alinhamento entre 3NHE.seg e alvo.seg com o Modeller:

- 1 Digite na linha de comando do terminal: `mod9.14 align2d.py`

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

**Não se preocupe!**

- 2 O *script* `align2d.py` lê os arquivos `3NHE.pdb`, `3NHE.seg` e `alvo.seg` e gera os arquivos `3NHE_alvo.pap` e `3NHE_alvo.ali`. O primeiro contém o alinhamento entre as duas sequências no formato ALN e o segundo contém o alinhamento no formato de entrada do Modeller. Observe que nesse *script* nós definimos que a cadeia a ser usada na modelagem será a cadeia B (como selecionada pelo MHOLline).
- 3 Abra esses dois arquivos com um editor de texto e observe as diferenças.

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

- 4 Digite na linha de comando do terminal: `mod9.14 model-single.py`

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

**Não se preocupe!**

- 5 O *script* `model-single.py` lê o arquivo `3NHE_alvo.ali`, gera três modelos para a proteína alvo 3NHE através da função `automodel()`, lista dentro do arquivo `model-single.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-single.out`.

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

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

### Avaliando os modelos 3D gerados:

- 6 Abra com um editor de texto o arquivo `./modelagem/basico/modeller/model-single.out`
- 7 Observe os resultados do DOPE, DOPE normalizado e GA341 para os modelos e para a estrutura representativa (`cluster.opt`).
- 8 Via terminal, entre no diretório `./sequences/singleinit/`
- 9 Digite na linha de comando do terminal:  
`vmd -m alvo.B9999000*.pdb cluster/cluster.opt ../3NHE.pdb`
- 10 Alinhe todas as estruturas.



## Gerando os modelos otimizados da proteína alvo 3NHE com o Modeller

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

11 Digite na linha de comando do terminal: `mod9.14 model-single-opt.py`

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

**Não se preocupe!**

12 O *script* `model-single-opt.py` lê o arquivo `3NHE_alvo.ali` e gera três modelos otimizados para a proteína alvo 3NHE através da função `automodel()`. O resumo dos resultados é gerado dentro do arquivo `model-single-opt.out`.

**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:

13 Abra com um editor de texto o arquivo `./modelagem/basico/modeller/model-single-opt.out`

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

15 Via terminal, entre no diretório `./sequences/singleopt/`

16 Digite na linha de comando do terminal:

```
vmd -m alvo.B9999000*.pdb cluster/cluster.opt ../3NHE.pdb
```

17 Alinhe todas as estruturas.

### Validando os modelos gerados via Molprobit:

- Avaliar os modelos gerados no diretório `./sequences/singleinit/` e no diretório `./sequences/singleopt/`
- Entre no diretório `./sequences/`
- Digite na linha de comando do terminal:

```
okular ./singleinit/molprobitresults/*.pdf ./singleopt/molprobitresults/*.pdf &
```

### Validando os modelos gerados via Procheck:

- Avaliar os modelos gerados no diretório `./sequences/singleinit/` e no diretório `./sequences/singleopt/`
- Entre no diretório `./sequences/`
- Digite na linha de comando do terminal:

```
okular ./singleinit/procheckresults/*.pdf ./singleopt/procheckresults/*.pdf &
```

### ***Passo 6: Analisando todos os modelos gerados***

1 Via terminal, entre no diretório './modelagem/basico/'

2 Digite na linha de comando do terminal:

```
vmd -m modeller/sequences/3NHE.pdb modeller/sequences/singleinit/alvo.B99990002.pdb  
modeller/sequences/singleopt/alvo.B99990002.pdb jobteste_emmsb/Modeller/1_teste76letters.B99990001.pdb
```

3 Alinhe todas as estruturas.