Tutorial – Modelagem Comparativa com o Modeller

Gentilmente cedido por Deborah Antunes

Sobre o Modeller

O programa MODELLER é usado para a modelagem por homologia ou comparativa de estruturas tridimensionais de proteínas. O usuário fornece um alinhamento de uma sequência a ser modelada com estruturas relacionadas conhecidas e o programa calcula automaticamente um modelo contendo todos os átomos não-hidrogênio. O MODELLER implementa modelagem comparativa de estrutura proteica por meio da satisfação de restrições espaciais e pode realizar muitas tarefas adicionais, incluindo modelagem *de novo* de loops em estruturas de proteínas, otimização de vários modelos de estrutura proteica em relação a uma função objetiva, alinhamento múltiplo de sequências de proteínas e/ou estruturas, agrupamento, pesquisa em bases de dados de sequências, comparação de estruturas de proteínas, etc. O programa está disponível para download na maioria dos sistemas Unix / Linux, Windows e Mac.

Download do Modeller

https://salilab.org/modeller/

MODELLER está disponível gratuitamente para instituições acadêmicas sem fins lucrativos. No entanto, é necessário registrar-se para obter uma licença para usar o software.

Este tutorial está orientado para usuários do **sistema operacional Linux** e tem por objetivo exemplificar a aplicação do software Modeller para predição de estruturas de proteínas via modelagem comparativa utilizando a técnica de restrições espaciais.

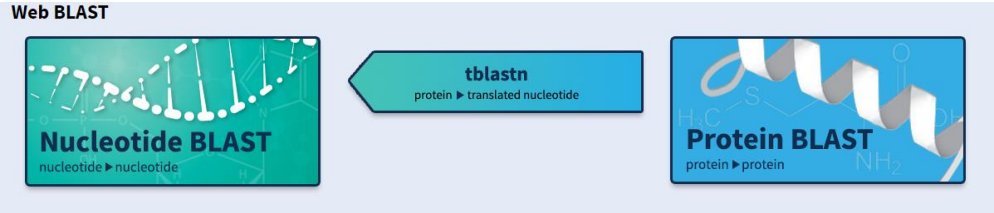
Sequência de interesse: NS5B protease Hepacivirus C

```
>P1;MODEL
sequence:MODEL:::::::0.00: 0.00
SLSYSWTGALVTATRREERRHPIGPLSNTLITKHNLVYQTTTASASARMTKVTIDREQILDKHY
FDTVTAVKKKASEVTADLLTWDEVARLTPKNTARSKSGLSGSDVRQLTRAARRELNSMWQDLLS
DSEELIPTTVMAKNEVFVSSPTARKPARLIVYPDLPVRACEKRAMYDLFQKLPYAVMGKAYGFQ
YTPRQRVNRLLDMWRHFKNPMGFSYDTKCFDSTVTPHDIDTERDIFLSATLPDEAKTVIKNLTS
RLYRGSPMYNSRGDLVGKRECRASGVFPTSMGNTLTNFIKATAAAKAAGLSDPQFLICGDDLVC
ITSSKGVEEDEQALREFTSAMTKYSAIPGDLPKPYYDLEQITSCSSNVTVAQDRNGRPYYFLTR
DPTTPLARASWETISHSPVNSWLGNIIAFAPTVWVRLVFLTHFFGLLLQQDAVDRNYEFEMYGS
TYSVNPLDLPAIIYKLHGPEAFDLTNYSPYEVQRVAAALQKLGSPPLRAWKRRAKLDRSKLKVR
GGRYAVVADYLFGFASAYRPKRPAPPGVNSIDVSGWFSIGDDSIGDIYRQ*
```

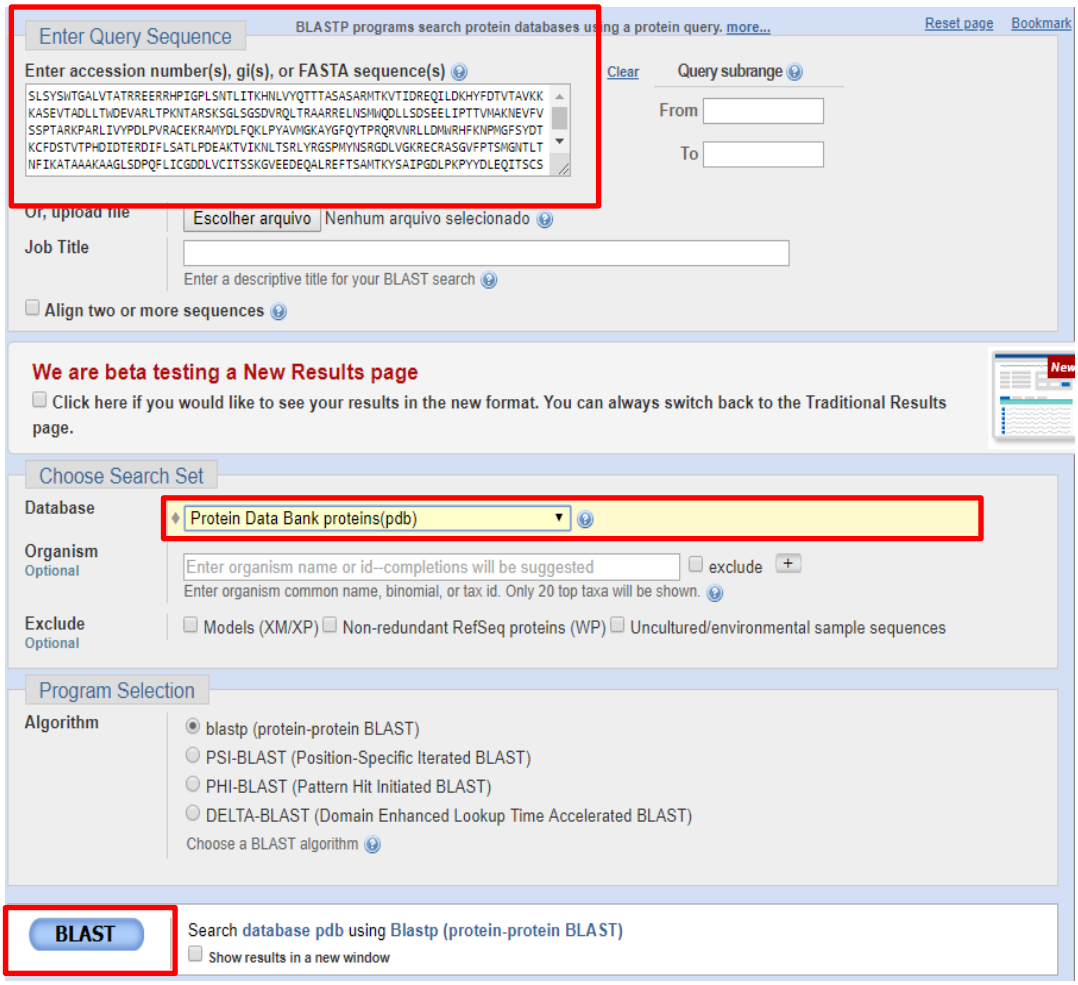
I. IDENTIFICAÇÃO DOS MOLDES

Submeter a sequência ao servidor BLAST para selecionar o(s) melhor(es) moldes (templates) baseados na identidade, cobertura e e-value.

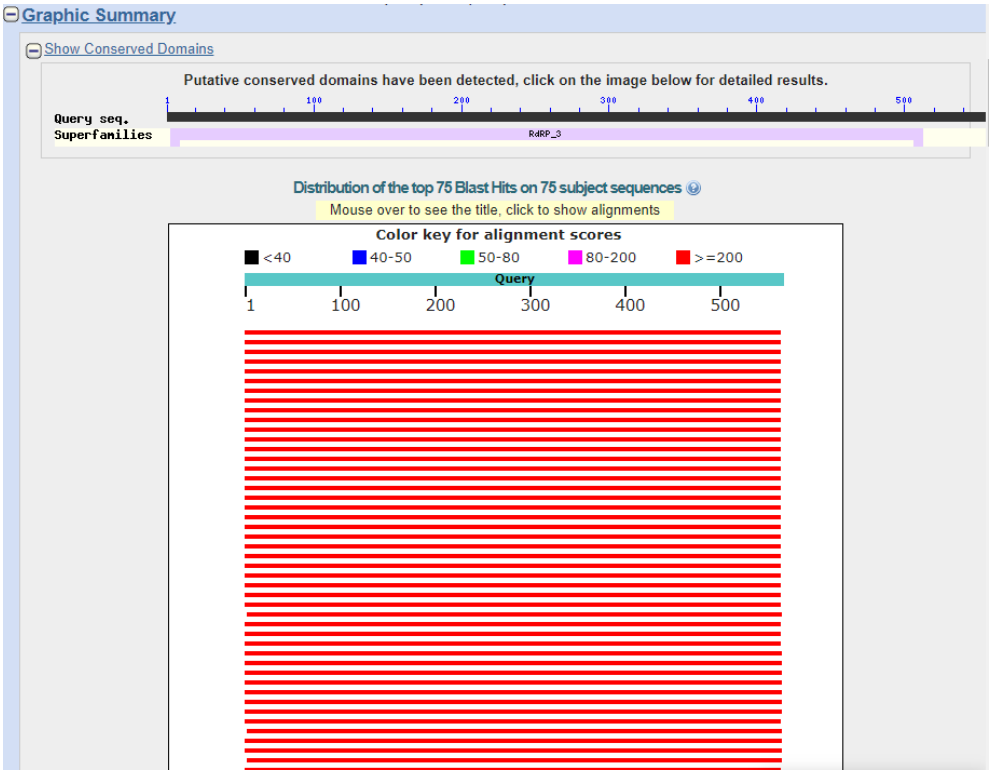
1) Acessar o site do NCBI: https://blast.ncbi.nlm.nih.gov/Blast.cgi
2) Selecionar a ferramenta “Protein BLAST”.



3) No campo “Enter Query Sequence”, colar a sequência que se deseja modelar (no formato FASTA).
4) Buscar pela sequência usando o parâmetro: “Database”: Protein Data Bank proteins (pdb).



5) Analisar os resultados.



Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☐	Chain A, Structure Of The Genotype 2b Hcv Polymerase [Hepatitis C virus isolate HC-J8]	690	690	99%	0.0	59.79%	3GSZ_A
☐	Chain A, Rna Dependent Rna Polymerase [Hepatitis C virus isolate HC-J6]	684	684	99%	0.0	59.43%	2XWH_A
☐	Chain A, Rna-directed Rna Polymerase [Hepacivirus C]	683	683	99%	0.0	59.25%	4ADP_A
☐	Chain A, Rna-dependent Rna Polymerase [Hepacivirus C]	682	682	99%	0.0	59.25%	1YUY_A
☐	Chain A, Rna-directed Rna Polymerase [Hepacivirus C]	681	681	99%	0.0	58.72%	4AEP_A
☐	Chain A, Rna-directed Rna Polymerase [Hepatitis C virus JFH-1]	681	681	99%	0.0	58.72%	3I5K_A
☐	Chain A, Rna-directed Rna Polymerase [Hepacivirus C]	681	681	99%	0.0	58.72%	2XXD_A
☐	Chain A, Rna-directed Rna Polymerase [Hepatitis C virus JFH-1]	679	679	99%	0.0	58.54%	2XYM_A
☐	Chain A, RNA-dependent RNA polymerase [Hepacivirus C]	678	678	99%	0.0	58.54%	5QJ0_A
☐	Chain A, RNA-directed RNA polymerase [Hepatitis C virus JFH-1]	672	672	99%	0.0	58.19%	4OBC_A
☐	Chain A, NS5B RNA-dependent RNA polymerase [Hepatitis C virus subtype 1a]	655	655	99%	0.0	57.30%	3HKW_A
☐	Chain A, Rna-directed Rna Polymerase [Hepacivirus C]	653	653	99%	0.0	57.30%	2XI2_A
☐	Chain A, Rna-directed Rna Polymerase [Hepatitis C virus JFH-1]	653	653	99%	0.0	57.12%	4E76_A
☐	Chain A, Rna-directed Rna Polymerase [Hepatitis C virus subtype 1a]	652	652	99%	0.0	57.30%	3QGH_A
☐	Chain A, Genome Polyprotein [Hepacivirus C]	650	650	99%	0.0	56.94%	5UJ2_A
☐	Chain A, Hcv Polymerase [Hepatitis C virus subtype 1a]	649	649	99%	0.0	56.94%	4KHM_A
☐	Chain A, Rna-directed Rna Polymerase [Hepatitis C virus JFH-1]	647	647	99%	0.0	56.76%	4WT9_A
☐	Chain A, NS5B RNA-dependent RNA polymerase [Hepatitis C virus subtype 1a]	643	643	99%	0.0	56.76%	4KHR_A
☐	Chain A, Rna-directed Rna Polymerase [Hepatitis C virus isolate HC-J4]	639	639	99%	0.0	55.34%	4OOW_A
☐	Chain A, Genome polyprotein [Hepatitis C virus isolate HC-J4]	639	639	99%	0.0	55.34%	3MWV_A
☐	Chain A, Rna-directed Rna Polymerase [Hepatitis C virus isolate HC-J4]	639	639	99%	0.0	55.34%	4DRU_A
☐	Chain A, Hc-J4 Rna Polymerase Apo-Form [Hepacivirus C]	639	639	99%			
☐	Chain A, Ns5b [Hepatitis C virus subtype 1b]	639	639	99%			

Questions/comments

Download GenPept Graphics ▼ Next ▲ Previous ◂◂Descriptions

Chain A, Structure Of The Genotype 2b Hcv Polymerase

Sequence ID: 3GSZ_A **Length:** 563 **Number of Matches:** 1

See 3 more title(s)

Range 1: 1 to 561 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
690 bits(1781)	0.0	Compositional matrix adjust.	336/562(60%)	413/562(73%)	3/562(0%)

```
Query  1    SLSYSWTGALVTATRREERRHPIGPLSNTLITKHNLVYQTTTASASARMTKVTIDREQIL  60
            S+SY+WTGAL+T     EE + PI PLSN+L+  HN VY TT+ SAS R  KVT DR Q+L
Sbjct  1    SMSYTWTGALITPCGPEEEKLPINPLSNSLMRFHNKVYSTTSRSASLRAKKVTFDRVQVL  60

Query  61   DKHYFDTVTAVKKKASEVTADLLTWDEVARLTPKNTARSKSGLSGSDVRQLTRAARRELN  120
            D HY    +  VK+ AS+V+A LLT +E   LTP ++A+S+ G    +VR L+R A   +
Sbjct  61   DAHYDSVLQDVKRAASKVSARLLTVEEACALTPPHSAKSRYGFGAKEVRSLSRRAVNHIR  120

Query  121  SMWQDLLSDSEELIPTTVMAKNEVFVSSPT--ARKPARLIVYPDLPVRACEKRAMYDLFQ  178
            S+W+DLL D    I TT+MAKNEVF   PT   +KPARLIVYPDL VR CEK A+YD+ Q
Sbjct  121  SVWEDLLEDQHTPIDTTIMAKNEVFCIDPTKGGKKPARLIVYPDLGVRVCEKMALYDIAQ  180

Query  179  KLPYAVMGKAYGFQYTPRQRVNRLLDMWRHFKNPMGFSYDTKCFDSTVTPHDIDTERDIF  238
            KLP A+MG +YGFQY+P +RV+ LL  W   K+PMGFSYDT+CFDSTVT  DI TE  I+
Sbjct  181  KLPKAIMGPSYGFQYSPAERVDFLLKAWGSKKDPMGFSYDTRCFDSTVTERDIRTEESIY  240

Query  239  LSATLPDEAKTVIKNLTSRLYRGSPMYNSRGDLVGKRECRASGVFPTSMGNTLTNFIKAT  298
             + +LP EA+TVI +LT RLY G PM NS+G   G R CRASGVF TSMGNT+T +IKA
Sbjct  241  QACSLPQEARTVIHSLTERLYVGGPMTNSKGQSCGYRRCRASGVFTTSMGNTMTCYIKAL  300

Query  299  AAAKAAGLSDPQFLICGDDLVCITSSKGVEEDEQALREFTSAMTKYSAIPGDLPKPYYDL  358
            AA KAAG+ DP  L+CGDDLV I+ S+G EEDE+ LR FT AMT+YSA PGDLP+P YDL
Sbjct  301  AACKAAGIVDPVMLVCGDDLVVISESQGNEEDERNLRAFTEAMTRYSAPPGDLPRPEYDL  360

Query  359  EQITSCSSNVTVAQDRNGRPYYFLTRDPTTPLARASWETISHSPVNSWLGNIIAFAPTVW  418
            E ITSCSSNV+VA D  GR  YFLTRDPTTP+ RA+WET+ HSPVNSWLGNII +APT+W
Sbjct  361  ELITSCSSNVSVALDSRGRRRYFLTRDPTTPITRAAWETVRHSPVNSWLGNIIQYAPTIW  420

Query  419  VRLVFLTHFFGLLLQQDAVDRNYEFEMYGSTYSVNPLDLPAIIYKLHGPEAFDLTNYSPY  478
            VR+V +THFF +LL QD +++N  FEMYG+ YSVNPLDLPAII +LHG EAF L  YSP+
Sbjct  421  VRMVIMTHFFSILLAQDTLNQNLNFEMYGAVYSVNPLDLPAIIERLHGLEAFSLHTYSPH  480

Query  479  EVQRVAAALQKLGSPPLRAWKRRAKLDRSKLKVRGGRYAVVADYLFGFASAYRPKRPAPP  538
            E+ RVAA L+KLG+PPLRAWK RA+  R+ L  +G R A+   YLF +A   + K    P
Sbjct  481  ELSRVAATLRKLGAPPLRAWKSRARAVRASLIAQGARAAICGRYLFNWAVKTKLKLTPLP  540

Query  539  GVNSIDVSGWFSIGDDSIGDIY  560
              + +D+SGWF++G    GDIY
Sbjct  541  EASRLDLSGWFTVGAGG-GDIY  561
```

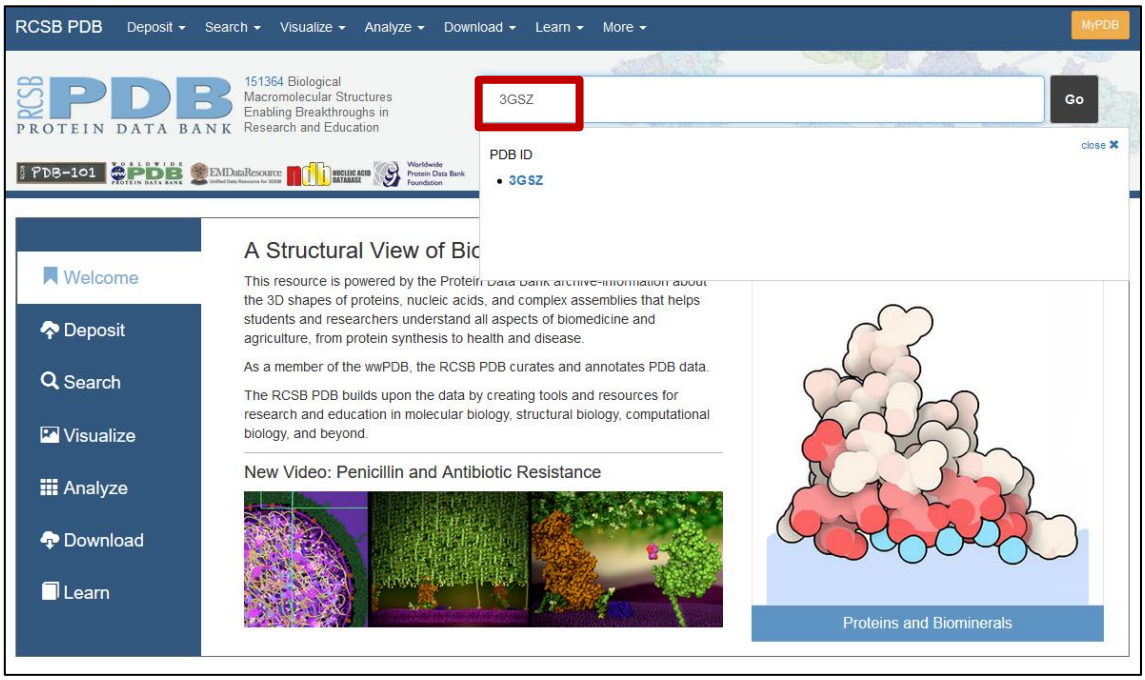
Related Information

Structure - 3D structure displays

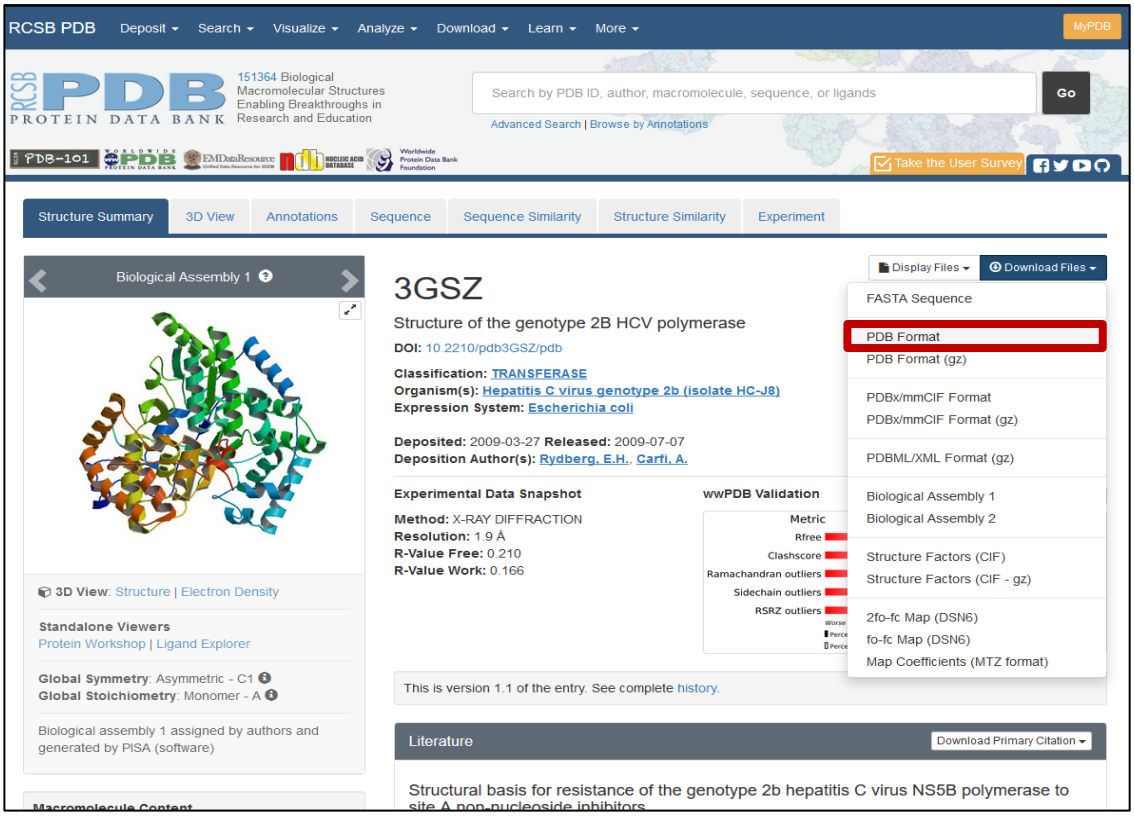
Identical Proteins - Identical proteins to 3GSZ_A

Buscar o template escolhido no Protein Data Bank.

6) Acessar o site do PDB: https://www.rcsb.org/. Buscar o template através do código PDB encontrado na etapa anterior: 3GSZ.



7) Fazer o download do arquivo PDB: “Download Files” – PDB Format.



Para realizar o Alinhamento temos que ACESSAR O SERVIDOR no terminal de vocês:

```
$ ssh bioufmg@bioinfo.icb.ufmg.br
```

DIGITE A SENHA.
Encontre a sua pasta (com seu nome):

```
lucianna@DESKTOP-JQF7JLS:~$ ssh bioufmg@bioinfo.icb.ufmg.br
bioufmg@bioinfo.icb.ufmg.br's password:
Last login: Wed Oct 14 17:09:15 2020 from 201.17.140.105
[bioufmg@bioinfo ~]$ ls
INTRO          elisaamancio      larama           milenap
ace222         eusoujacu         larissa_cler     monic_lops
amandacpa      flavia            laryhenriques    moysesmn
anna_menezes   flaviam           letbarbosa       nathalia_alv
arthurtrcf     franciscoacarmo   letchris         pedrobala
barbaramas     fredericogabriel  lucasduque       rayssa_soares
beatrizapgaua  gabijager         luizacaixeta     renan
caizin         gabriel_ferrari   macl3y           tarefa1
camilabd       gabrielrod        madumascarenhas  tarefa2
camilacllc     gapmo             mari_1305        teste
ceciliahorta   jessica_alves     marinak          thiagoac
clecio_alonso  jorge             markko
danikayali     karenkanda        mbarcelos
dg_sousa       key               michelle_157
[bioufmg@bioinfo ~]$
```

***se não possuir pasta. Crie uma com o comando mkdir <seunome>**

Acesse sua pasta com

```
$ cd minhapasta
```

Confira se você está realmente na pasta com o seu nome:

```
$ pwd
```

Dentro da sua pasta, crie uma nova pasta chamada modeller_pratica:

```
$ mkdir modeller_pratica
```

Acesse a pasta

```
$ cd modeller_pratica
```

Copie o conteúdo da pasta /home/treinamento/modeller_pratica para cá:

```
$ cp /home/treinamento/modeller_pratica/* .
```

Pronto. Temos um diretório para rodar o modeller:

```
[bioufmg@bioinfo modeller_pratica]$ ls
3gsz.pdb    5uj2.pdb    evaluate_model.py  model-3gsz.pse  sequencia.txt
3gsz_a.pdb  align2d.py  genmodel.py        model.ali
[bioufmg@bioinfo modeller_pratica]$
```

II. ALINHAMENTO ENTRE O ALVO E O MOLDE

Realizar o alinhamento da sequência alvo com o molde correspondente utilizando o Modeller.

8) Três arquivos são necessários:
    a. Sequência alvo em formato PIR (**model.ali**).

```
>P1;MODEL
sequence:MODEL:::::::0.00: 0.00
SLSYSWTGALVTATRREERRHPIGPLSNTLITKHNLVYQTTTASASARMTKVTIDREQILDKHYFDTVTAVKKKASEVTA
DLLTWDEVARLTPKNTARSKSGLSGSDVRQLTRAARRELNSMWQDLLSDSEELIPTTVMAKNEVFVSSPTARKPARLIVY
PDLPVRACEKRAMYDLFQKLPYAVMGKAYGFQYTPRQRVNRLLDMWRHFKNPMGFSYDTKCFDSTVTPHDIDTERDIFLS
ATLPDEAKTVIKNLTSRLYRGSPMYNSRGDLVGKRECRASGVFPTSMGNTLTNFIKATAAAKAAGLSDPQFLICGDDLVC
ITSSKGVEEDEQALREFTSAMTKYSAIPGDLPKPYYDLEQITSCSSNVTVAQDRNGRPYYFLTRDPTTPLARASWETISH
SPVNSWLGNIIAFAPTVWVRLVFLTHFFGLLLQQDAVDRNYEFEMYGSTYSVNPLDLPAIIYKLHGPEAFDLTNYSPYEV
QRVAAALQKLGSPPLRAWKRRAKLDRSKLKVRGGRYAVVADYLFGFASAYRPKRPAPPGVNSIDVSGWFSIGDDSIGDIY
RQ*
```

    b. Arquivo PDB da estrutura molde (**3gsz.pdb**).
    c. Script em linguagem phyton para alinhamento (**align2d.py**).

    ##

```
from modeller import *

env = environ()
aln = alignment(env)
mdl = model(env, file='3gsz', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='3gsz', atom_files='3gsz.pdb')
aln.append(file='model.ali', align_codes='MODEL')
aln.align2d()
aln.write(file='model-3gsz.ali', alignment_format='PIR')
aln.write(file='model-3gsz.pap', alignment_format='PAP')
```

    ##

    ***Caso escolhêssemos outro PDB teríamos que mudar os valores em negrito com o PDB escolhido.**

9) Através do Terminal rodar o script align2d.py no Modeller:

```
$ python align2d.py
```

    *O comando normalmente é mod e a versão do modeller. Ex.: mod9.25 align2d.py

10) Dois arquivos são gerados:

a. model-**3gsz**.ali – arquivo com o alinhamento para próxima etapa.

```
1
2  >P1;3gsz
3  structureX:3gsz.pdb:   1 :A:+558 :A:MOL_ID  1; MOLECULE  RNA-DIRECTED RNA POLYMERASE;
4  SMSYTWTGALITPCGPEEEKLPINPLSNSLMRFHNKVYSTTSRSASLRAKKVTFDRVQVLDAHYDSVLQDVKRAA
5  SKVSARLLTVEEACALTPPHSAKSRYGFGAKEVRSLSRRAVNHIRSVWEDLLEDQHTPIDTTIMAKNEVFCI---
6  DKKPARLIVYPDLGVRVCEKMALYDIAQKLPKAIMGPSYGFQYSPAERVDFLLKAWGSKKDPMGFSYDTRCFDST
7  VTERDIRTEESIYQACSLPQEARTVIHSLTERLYVGGPMTNSKGQSCGYRRCRASGVFTTSMGNTMTCYIKALAA
8  CKAAGIVDPVMLVCGDDLVVISESQGNEEDERNLRAFTEAMTRYSAPPGDLPRPEYDLELITSCSSNVSVALDSR
9  GRRRYFLTRDPTTPITRAAWETVRHSPVNSWLGNIIQYAPTIWVRMVIMTHFFSILLAQDTLNQNLNFEMYGAVY
10 SVNPLDLPAIIERLHGLEAFSLHTYSPHELSRVAATLRKLGAPPLRAWKSRARAVRASLIAQGARAAICGRYLFN
11 WAVKTKLKLTPLPEASRLDLSGWFTVGAGG-GDIYHS*
12
13 >P1;MODEL
14 sequence:MODEL:     : :     : ::: 0.00: 0.00
15 SLSYSWTGALVTATRREERRHPIGPLSNTLITKHNLVYQTTTASASARMTKVTIDREQILDKHYFDTVTAVKKKA
16 SEVTADLLTWDEVARLTPKNTARSKSGLSGSDVRQLTRAARRELNSMWQDLLSDSEELIPTTVMAKNEVFVSSPT
17 ARKPARLIVYPDLPVRACEKRAMYDLFQKLPYAVMGKAYGFQYTPRQRVNRLLDMWRHFKNPMGFSYDTKCFDST
18 VTPHDIDTERDIFLSATLPDEAKTVIKNLTSRLYRGSPMYNSRGDLVGKRECRASGVFPTSMGNTLTNFIKATAA
19 AKAAGLSDPQFLICGDDLVCITSSKGVEEDEQALREFTSAMTKYSAIPGDLPKPYYDLEQITSCSSNVTVAQDRN
20 GRPYYFLTRDPTTPLARASWETISHSPVNSWLGNIIAFAPTVWVRLVFLTHFFGLLLQQDAVDRNYEFEMYGSTY
21 SVNPLDLPAIIYKLHGPEAFDLTNYSPYEVQRVAAALQKLGSPPLRAWKRRAKLDRSKLKVRGGRYAVVADYLFG
22 FASAYRPKRPAPPGVNSIDVSGWFSIGDDSIGDIYRQ*
23
```

b. model-**3gsz**.pap – arquivo com alinhamento em que mostra os resíduos conservados.

```
 _aln.pos         10        20        30        40        50        60
3gsz      SMSYTWTGALITPCGPEEEKLPINPLSNSLMRFHNKVYSTTSRSASLRAKKVTFDRVQVLDAHYDSVL
MODEL     SLSYSWTGALVTATRREERRHPIGPLSNTLITKHNLVYQTTTASASARMTKVTIDREQILDKHYFDTV
 _consrvd * ** ***** *    **   ** **** *   ** ** **  *** *  *** ** * ** **

 _aln.p   70        80        90       100       110       120       130
3gsz      QDVKRAASKVSARLLTVEEACALTPPHSAKSRYGFGAKEVRSLSRRAVNHIRSVWEDLLEDQHTPIDT
MODEL     TAVKKKASEVTADLLTWDEVARLTPKNTARSKSGLSGSDVRQLTRAARRELNSMWQDLLSDSEELIPT
 _consrvd   **  ** * * ***  *   ***   * *  *     ** * * *     * * *** *    * *

 _aln.pos  140       150       160       170       180       190       200
3gsz      TIMAKNEVFCI---DKKPARLIVYPDLGVRVCEKMALYDIAQKLPKAIMGPSYGFQYSPAERVDFLLK
MODEL     TVMAKNEVFVSSPTARKPARLIVYPDLPVRACEKRAMYDLFQKLPYAVMGKAYGFQYTPRQRVNRLLD
 _consrvd * *******       *********** ** *** * **  **** * **  ***** *  **  **

 _aln.pos    210       220       230       240       250       260       270
3gsz      AWGSKKDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPQEARTVIHSLTERLYVGGPMTNSKGQSC
MODEL     MWRHFKNPMGFSYDTKCFDSTVTPHDIDTERDIFLSATLPDEAKTVIKNLTSRLYRGSPMYNSRGDLV
 _consrvd  *   * ******** *******  ** **  *     ** ** ***  ** *** * ** ** * *

 _aln.pos      280       290       300       310       320       330       340
3gsz      GYRRCRASGVFTTSMGNTMTCYIKALAACKAAGIVDPVMLVCGDDLVVISESQGNEEDERNLRAFTEA
MODEL     GKRECRASGVFPTSMGNTLTNFIKATAAAKAAGLSDPQFLICGDDLVCITSSKGVEEDEQALREFTSA
 _consrvd * * ******* ****** *  *** ** ****  **  * ****** *  * * ****  ** ** *

 _aln.pos        350       360       370       380       390       400
3gsz      MTRYSAPPGDLPRPEYDLELITSCSSNVSVALDSRGRRRYFLTRDPTTPITRAAWETVRHSPVNSWLG
MODEL     MTKYSAIPGDLPKPYYDLEQITSCSSNVTVAQDRNGRPYYFLTRDPTTPLARASWETISHSPVNSWLG
 _consrvd ** *** ***** * **** ******** ** *  **  **********  ** ***  *********

 _aln.p  410       420       430       440       450       460       470
3gsz      NIIQYAPTIWVRMVIMTHFFSILLAQDTLNQNLNFEMYGAVYSVNPLDLPAIIERLHGLEAFSLHTYS
MODEL     NIIAFAPTVWVRLVFLTHFFGLLLQQDAVDRNYEFEMYGSTYSVNPLDLPAIIYKLHGPEAFDLTNYS
 _consrvd ***  *** *** *  ****  ** **    *  *****  ************  *** *** *  **

 _aln.pos  480       490       500       510       520       530       540
3gsz      PHELSRVAATLRKLGAPPLRAWKSRARAVRASLIAQGARAAICGRYLFNWAVKTKLKLTPLPEASRLD
MODEL     PYEVQRVAAALQKLGSPPLRAWKRRAKLDRSKLKVRGGRYAVVADYLFGFASAYRPKRPAPPGVNSID
 _consrvd * *  **** * *** ******* **   *  *   * * *    ***  *     *    *     *

 _aln.pos    550       560
3gsz      LSGWFTVGAGG-GDIYHS
MODEL     VSGWFSIGDDSIGDIYRQ
 _consrvd  ****  *    ****

```

III. CONSTRUÇÃO DO MODELO

Construção do modelo utilizando o resultado do alinhamento (model-**3gsz**.ali).

11) Três arquivos são necessários:
    a. Alinhamento entre sequência de interesse e estrutura molde (model-**3gsz**.ali).
    b. Arquivo PDB da estrutura molde (**3gsz**.pdb).
    c. Script em linguagem phyton para construção do modelo (genmodel.py).

```
##############################################################################

from modeller import *
from modeller.automodel import *              # Load the automodel class

log.verbose()
env = environ()
```

```
env.io.hetatm = False       # Read in HETATM records from template PDBs

a = automodel(env,
              alnfile  = 'model-3gsz.ali',          # alignment filename
              knowns   = ('3gsz'),                  # codes of the templates
              sequence = 'MODEL',
              assess_methods = (assess.DOPE, assess.GA341))    # code of the target

a.starting_model= 1                                 # index of the first model
a.ending_model  = 10                                # index of the last model
                                                    # (determines how many models to calculate)

a.make()                                            # do homology modeling
```

##

*Como estamos apenas fazendo um tutorial, colocamos apenas 10 modelos. Em uma modelagem normalmente se pede 100 modelos ou mais.

12) Através do Terminal rodar o script genmodel.py no Modeller:

```
$ python genmodel.py >&1 | tee genmodel.log
```

13) Dez modelos foram gerados.

```
>> Summary of successfully produced models:
Filename                          molpdf     DOPE score    GA341 score
----------------------------------------------------------------------
MODEL.B99990001.pdb           2874.98608   -66463.61719        1.00000
MODEL.B99990002.pdb           2787.88892   -66913.05469        1.00000
MODEL.B99990003.pdb           2843.51221   -66461.59375        1.00000
MODEL.B99990004.pdb           2725.74146   -66682.02344        1.00000
MODEL.B99990005.pdb           3027.27954   -66237.45312        1.00000
MODEL.B99990006.pdb           2714.04736   -66634.33594        1.00000
MODEL.B99990007.pdb           2993.76465   -66044.28906        1.00000
MODEL.B99990008.pdb           2895.72876   -66338.40625        1.00000
MODEL.B99990009.pdb           2792.26758   -67115.99219        1.00000
MODEL.B99990010.pdb           2763.75220   -66281.07812        1.00000
```

Deve-se escolher aquele com **menor DOPE score**. Ou seja, o valor mais negativo. Este resultado encontra-se ao final do arquivo ***genmodel.log***. No terminal escrever:

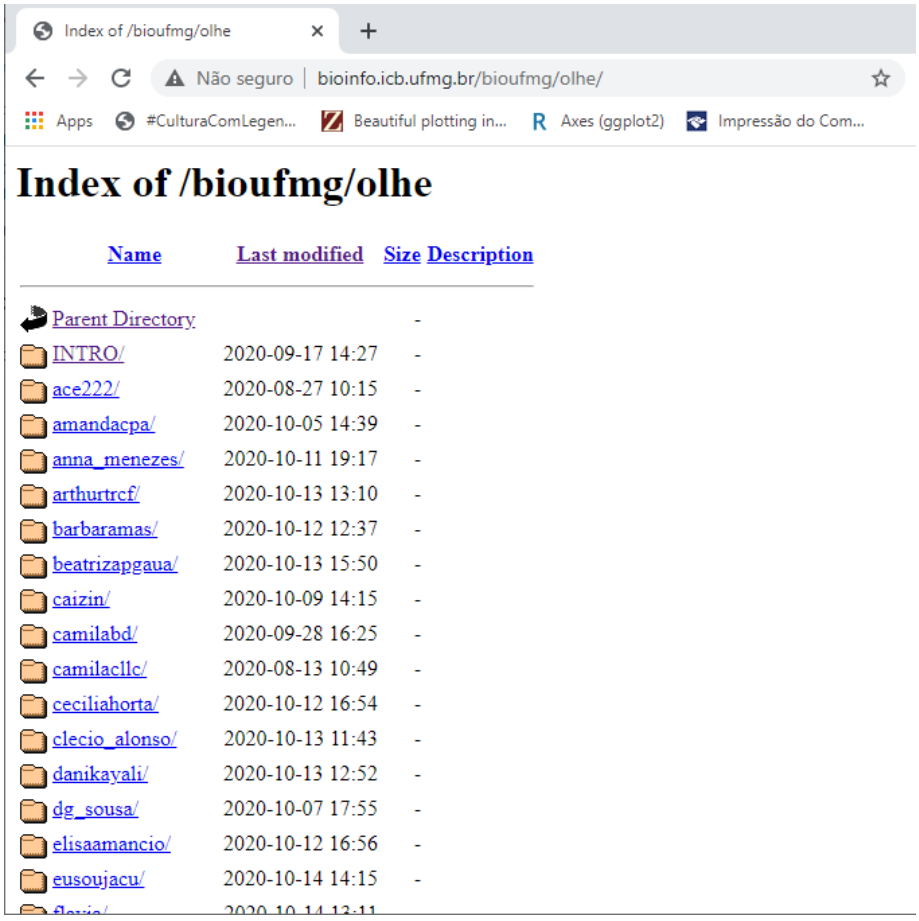
```
$ sed -e '1,/Summary of successfully produced models/d' *.log | grep pdb| sort -nk 3
```

```
1  MODEL.B99990009.pdb          2792.26758   -67115.99219          1.00000
2  MODEL.B99990002.pdb          2787.88892   -66913.05469          1.00000
3  MODEL.B99990004.pdb          2725.74146   -66682.02344          1.00000
4  MODEL.B99990006.pdb          2714.04736   -66634.33594          1.00000
5  MODEL.B99990001.pdb          2874.98608   -66463.61719          1.00000
6  MODEL.B99990003.pdb          2843.51221   -66461.59375          1.00000
7  MODEL.B99990008.pdb          2895.72876   -66338.40625          1.00000
8  MODEL.B99990010.pdb          2763.75220   -66281.07812          1.00000
9  MODEL.B99990005.pdb          3027.27954   -66237.45312          1.00000
10 MODEL.B99990007.pdb          2993.76465   -66044.28906          1.00000
```

14) Após determinar qual o modelo de menor energia, vamos transferir para o nosso computador através do site

bioinfo.icb.ufmg.br/bioufmg

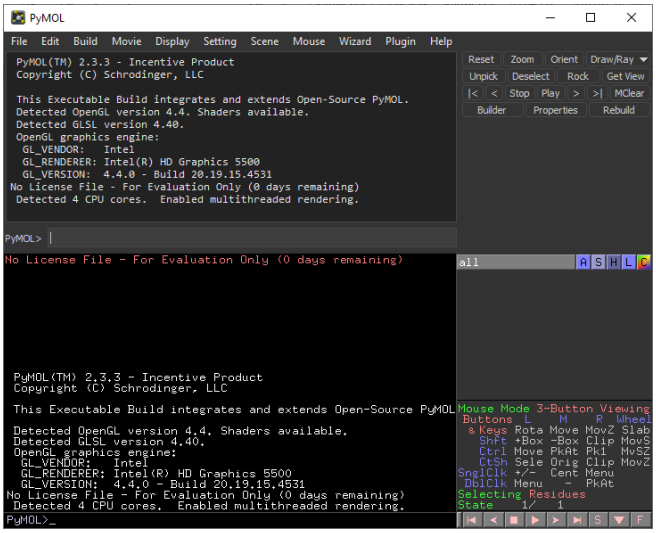
Digite usuário e senha.

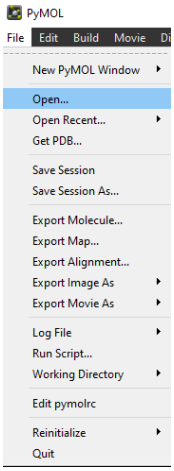


Baixe os arquivos 3gsz_a.pdb e MODEL.B99990009.pdb de sua pasta para seu computador.

15) Verificar o modelo gerado com **menor DOPE score** e visualizar a estrutura pelo PyMOL. Podemos verificar o distanciamento entre os dois objetos por RMSD.

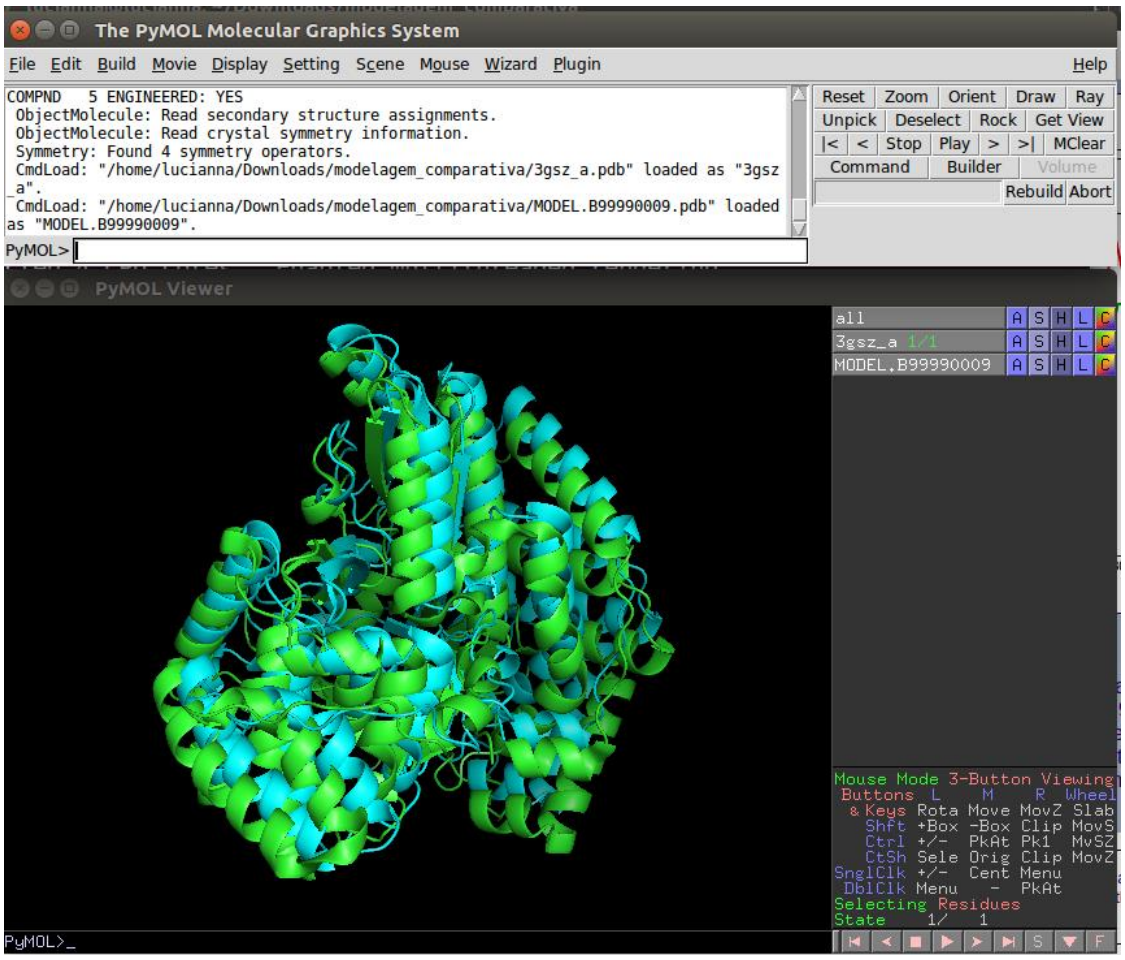
d. Abrir o pymol (após instalação): 3gsz_a.pdb e MODEL.B99990009.pdb





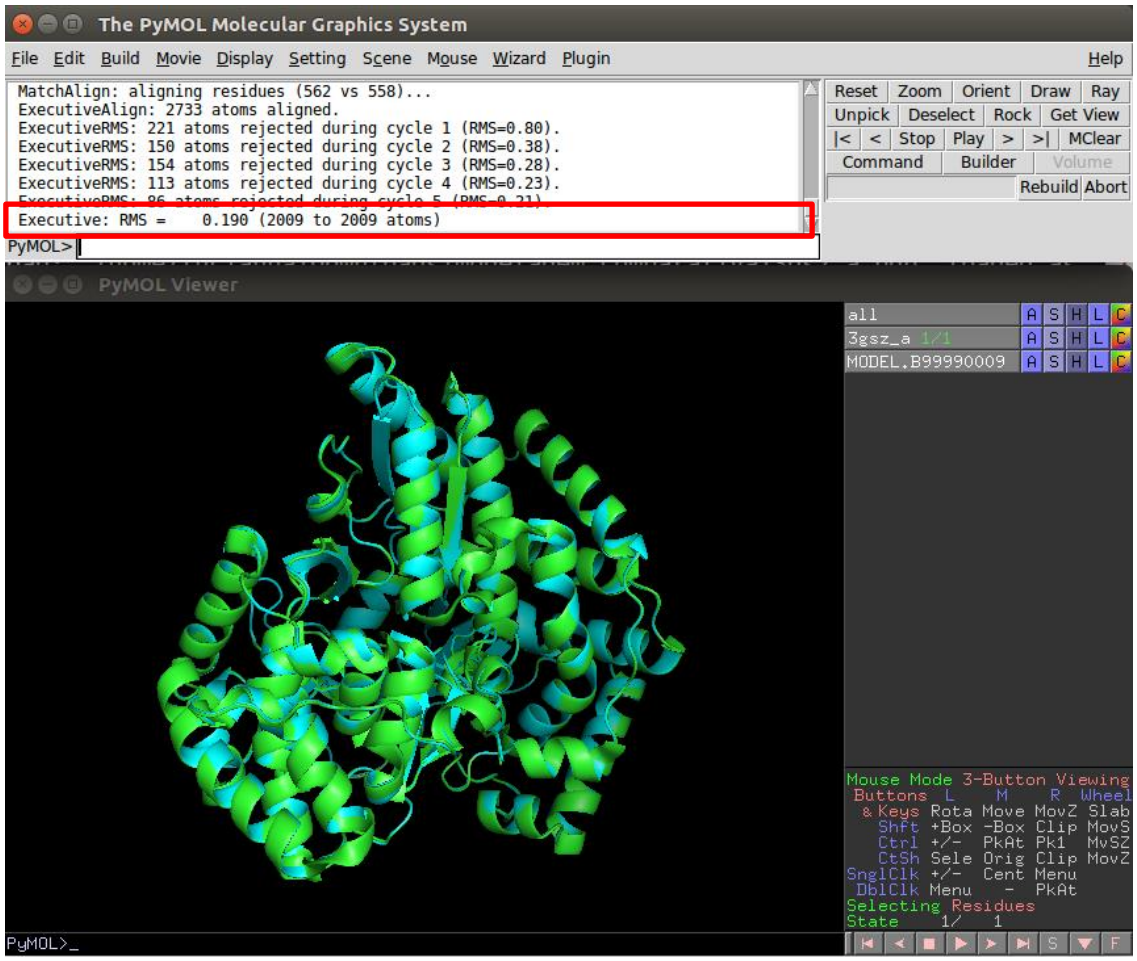
No Linux (após instalação):

```
$ pymol 3gsz_a.pdb MODEL.B99990009.pdb
```



Utilize o comando de alinhamento do pymol:

align MODEL.B99990009, 3gsz_a

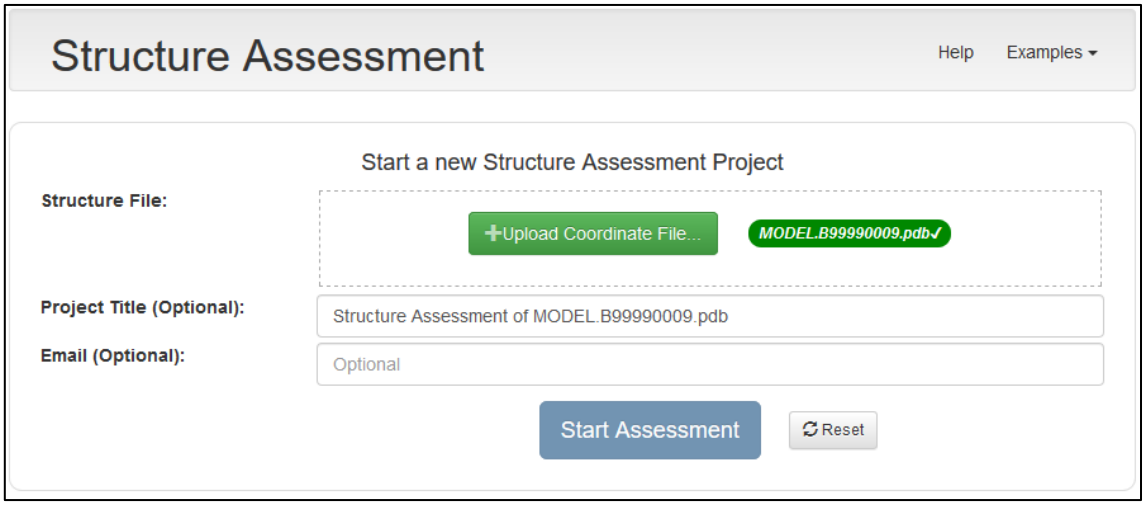


IV. VALIDAÇÃO DO MODELO

16) Avaliar a qualidade do modelo gerado utilizando o servidor SWISS-MODEL - Structure Assessment.

a) Acessar o servidor: https://swissmodel.expasy.org/assess

b) No campo “Structure File” adicionar o arquivo PDF com menor DOPE score.



c) Analisar Resultados.