

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
DSEELIPTTVMAKNEVFVSSPTARKPARLIVYPDLVPRACEKRAMYDLFQKLPYAVMGKAYGFQ
YTPRQRVNRLLDMWRHFKNPMGFSYDTKCFDSTVTPHDIDTERDIFLSATLPDEAKTVIKNLTS
RLYRGSPMYNSRGDLVGKRECRASGVFPTSMGNTLTNFIKATAAACAAGLSDPQFLICGDDLVC
ITSSKGVEEDEQALREFTSAMTKYSAIPGDLPKPYYDLEQITSCSSNVTVAQDRNGRPYYFLTR
DPTTPLARASWETISHSPVNSWLGNIIAFAPTVMVRLVFLTHFFGLLLQQDAVDRNYEFEMYGS
TYSVNPLDLPALIIYKLGPEAFDLTNYSPEYQQRVAAALQKLGSPPLRAWKRRAKLDRSKLKVR
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).

BLASTP programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

```
SLSYSITGALVLTATREERRHPIGPLSNTLITKHNLYQTTTASASARHTKVTIDREQLDKHYFDTVTAVKK
KASEVTADLLTNDDEVARLTPKNTARSKSGLSGSDVRQLTRAARRELNISMQDLLSDSEELIPTTVMAKNEVFV
SSPTARKPARLLIVYPDLVPRACEKRAMYDLFQKLPYAVMGKAYGFQYTPRQRVWRLDMIRHFKINPMGFSYDT
KCFDSTVTPHDIDTERDIFLSATLPDEAKTVIKNLTSLRYRGSPIYNSRGDLVGKRECRASGVFFPTSMGNTLT
NFIIKATAAAKAAGLSDPQFLICGDDLVLCITSSKGVEEDEQALREFTSAMTKYSAIPGDLPKPYDLEQITSCS
```

Clear Query subrange [?](#)

From

To


Or, upload file Nenhum arquivo selecionado [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

We are beta testing a New Results page

Click here if you would like to see your results in the new format. You can always switch back to the Traditional Results page. 

Choose Search Set

Database [?](#)

Organism [Optional](#) exclude
 Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude [Optional](#) Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm

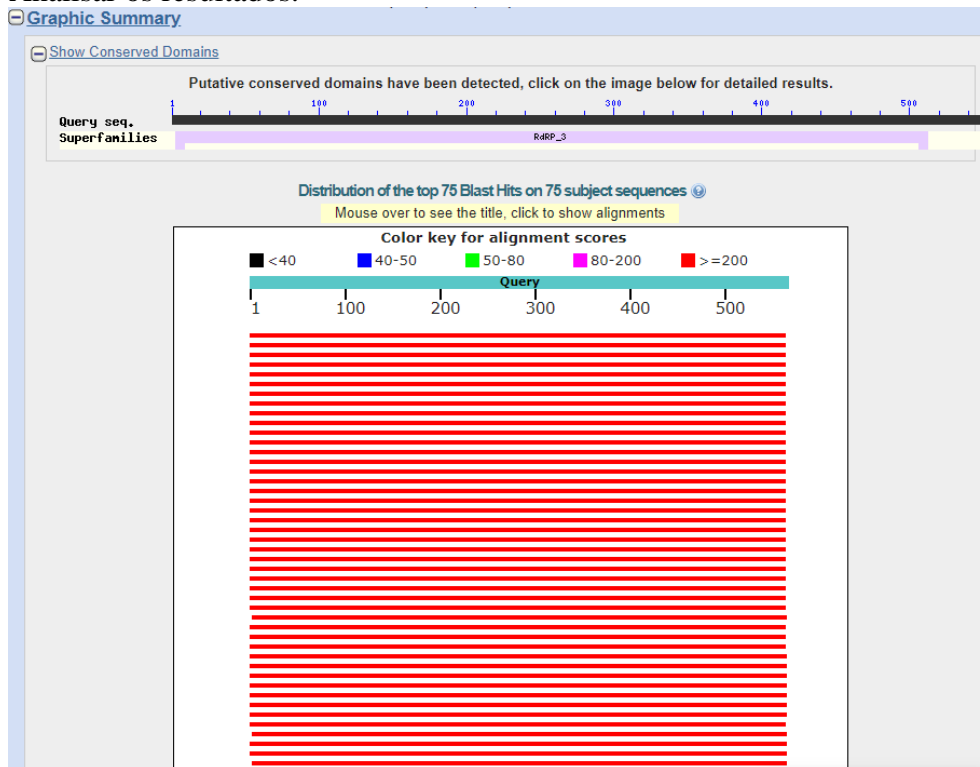
- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST Search database pdb using Blastp (protein-protein BLAST)

Show results in a new window

5) Analisar os resultados.



Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

| | Description | Max Score | Total Score | Query Cover | E value | Per. Ident | Accession |
|--------------------------|--|-----------|-------------|-------------|---------|------------|------------------------|
| <input type="checkbox"/> | Chain A, Structure Of The Genotype 2b Hcv Polymerase [Hepatitis C virus isolate HC-J8] | 690 | 690 | 99% | 0.0 | 59.79% | 3GSZ_A |
| <input type="checkbox"/> | Chain A, Rna Dependent Rna Polymerase [Hepatitis C virus isolate HC-J6] | 684 | 684 | 99% | 0.0 | 59.43% | 2XWH_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepacivirus C] | 683 | 683 | 99% | 0.0 | 59.25% | 4ADP_A |
| <input type="checkbox"/> | Chain A, Rna-dependent Rna Polymerase [Hepacivirus C] | 682 | 682 | 99% | 0.0 | 59.25% | 1YUY_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepacivirus C] | 681 | 681 | 99% | 0.0 | 58.72% | 4AEP_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepatitis C virus JFH-1] | 681 | 681 | 99% | 0.0 | 58.72% | 3I5K_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepacivirus C] | 681 | 681 | 99% | 0.0 | 58.72% | 2XXD_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepatitis C virus JFH-1] | 679 | 679 | 99% | 0.0 | 58.54% | 2XYM_A |
| <input type="checkbox"/> | Chain A, RNA-dependent RNA polymerase [Hepacivirus C] | 678 | 678 | 99% | 0.0 | 58.54% | 5QJ0_A |
| <input type="checkbox"/> | Chain A, RNA-directed RNA polymerase [Hepatitis C virus JFH-1] | 672 | 672 | 99% | 0.0 | 58.19% | 4QBC_A |
| <input type="checkbox"/> | Chain A, NS5B RNA-dependent RNA polymerase [Hepatitis C virus subtype 1a] | 655 | 655 | 99% | 0.0 | 57.30% | 3HKW_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepacivirus C] | 653 | 653 | 99% | 0.0 | 57.30% | 2XJ2_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepatitis C virus JFH-1] | 653 | 653 | 99% | 0.0 | 57.12% | 4E76_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepatitis C virus subtype 1a] | 652 | 652 | 99% | 0.0 | 57.30% | 3QGH_A |
| <input type="checkbox"/> | Chain A, Genome Polyprotein [Hepacivirus C] | 650 | 650 | 99% | 0.0 | 56.94% | 5UJ2_A |
| <input type="checkbox"/> | Chain A, Hcv Polymerase [Hepatitis C virus subtype 1a] | 649 | 649 | 99% | 0.0 | 56.94% | 4KHM_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepatitis C virus JFH-1] | 647 | 647 | 99% | 0.0 | 56.76% | 4WTR_A |
| <input type="checkbox"/> | Chain A, NS5B RNA-dependent RNA polymerase [Hepatitis C virus subtype 1a] | 643 | 643 | 99% | 0.0 | 56.76% | 4KHR_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepatitis C virus isolate HC-J4] | 639 | 639 | 99% | 0.0 | 55.34% | 4QOW_A |
| <input type="checkbox"/> | Chain A, Genome polyprotein [Hepatitis C virus isolate HC-J4] | 639 | 639 | 99% | 0.0 | 55.34% | 3MWV_A |
| <input type="checkbox"/> | Chain A, Rna-directed Rna Polymerase [Hepatitis C virus isolate HC-J4] | 639 | 639 | 99% | 0.0 | 55.34% | 4DRU_A |
| <input type="checkbox"/> | Chain A, Hc-J4 Rna Polymerase Apo-Form [Hepacivirus C] | 639 | 639 | 99% | 0.0 | 55.34% | 4DRU_A |
| <input type="checkbox"/> | Chain A, NS5B RNA-dependent RNA polymerase [Hepatitis C virus subtype 1a] | 630 | 630 | 99% | 0.0 | 55.34% | 4DRU_A |

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

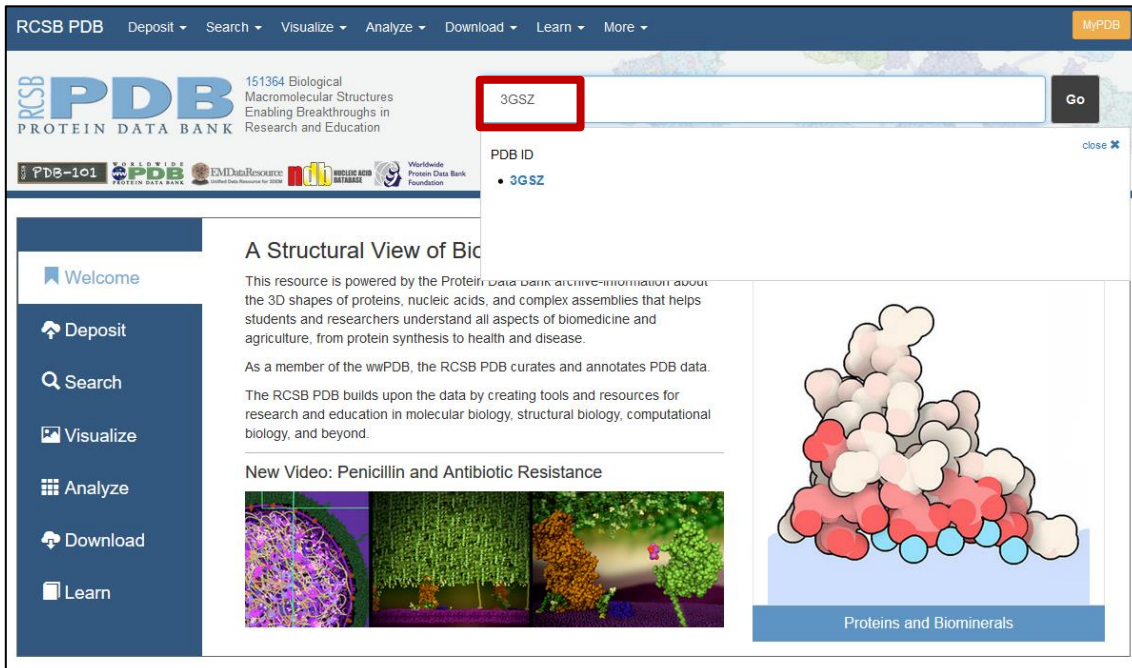
| Score | Expect | Method | Identities | Positives | Gaps |
|-----------------|-------------------------|------------------------------------|-------------------------|--------------------------|--------------------------------------|
| 690 bits (1781) | 0.0 | Compositional matrix adjust. | 336/562 (60%) | 413/562 (73%) | 3/562 (0%) |
| Query 1 | SLSYSWTGALVTATRREERRHP | I | GPLSNTLITKHNLYO | TTASASARMTKVTIDREQIL | 60 |
| Sbjct 1 | SMSYWTGAL+T EE + PI | PLSN+L+ HN VY TT+ SAS R KVT DR Q+L | SMSYWTGALITFCGPEEEKL | FINPLSNLMRPHNKVYSTTSR | SASLRAKKEVTFDRVQVL |
| Query 61 | DKHYFDTVAKKEASEVTADLL | TWDEVARLTPKNTARSKS | GLSGSDVRQLTRAARRELN | D HY + VK+ AS+V+A LLT +E | LTP ++A+S+G +VR L+R A + |
| Sbjct 61 | DAHYSVLQDVKRAASKVSARLL | TVEEACALTPPHSAKSRYP | GFAKEVRSLSRRAVNHIR | | |
| Query 121 | SMWQDLSDSEELIPTTVMARNE | VVSSPT--ARKPARLIVY | PDLPVRACEKRAMYDLFQ | S+W+DL D I TT+MARNEVF | PT +KPARLIVYDPL VR CEK A+YD+ Q |
| Sbjct 121 | SWWEDLLEDQHTPIDTTIMARNE | VFCIDPDKGGKPARLIVY | PDLPVRCMKMALYDIAQ | | |
| Query 179 | KLPYAVMGKAYGFYTPFRQVRN | LLDMRHRFKNPMGFSYD | TKCFDSTVTPHDIDTERDIF | KLP A+MG +YGFQY+P +RV+ | LL W K+PMGFSYD+CFDSTVT DI TE I+ |
| Sbjct 181 | KLPKAIMGFSYGFQYSPAERVD | FLKAWGSKDPMGFSYDTRC | FDSTVTERDIRTEESYI | | |
| Query 239 | LSATLPDEAKTVIKNLSRLYRGS | SPMYNSRGLVGRKCRASG | VFPPTSMGNTLITNFIKAT | + +LP EA+TVI +LT RLY G | PM NS+G G R CRASGVF TSMGNT+T +IKA |
| Sbjct 241 | QACSLPQEARTVIHSALTERLY | VGGPMTNSKQSGCYRRCR | ASGVFTSMGNTMTCIYKAL | | |
| Query 299 | AAAKAAGSDPQFLICGDDLVCI | TSSKGVVEDEQALREPT | SAMTRYSAIPGDLPPFYD | AA KAAG+ DP L+CGDDLV I+ | S+G EDE+ LR FT AMT+YSA PGDLP+P YDL |
| Sbjct 301 | AACKAAGIVDFVMLVCGDDLV | ISESQNEEDERNLRAPT | AMTRYSAIPGDLPPFYD | | |
| Query 359 | EQITSCSSNVTVAQDRNGR | FYFLTRDPTPLARASWETI | ISHSPVNSWLGNIIFAFTW | E ITSCSSN+VA D GR YFLTRD | PTTE+ RA+WET+ HSPVNSWLGNI I +AFT+W |
| Sbjct 361 | ELITSCSSNVSVALDSRGR | RYFLTRDPTTITRAAET | TVRHSFPVNSWLGNI IQAFTW | | |
| Query 419 | VRLVLTHFFGLLQDQDAVDRN | YEFEMYGSTYSVNPFLD | LPAIYKLGPEAFDLNYSY | VR+V +THFF +LL QD +++N | FEMYG+ YSVNPLDLP AI I +LHGPEAFDLNYSY |
| Sbjct 421 | VRMVIMTHFFSILLAQDTLN | QNLNFEMYGAVYSVNPFLD | LPAI IERLHGLEAFSLHTYSPH | | |
| Query 479 | EVORVAALQKLGSPFLRAWK | RRAKLRDRSKLVGRG | RYAVADYLFGEFASRYRPP | E+ RVAA L+KLG+PFLRAWK | RA+ R+ L +G R A+ YLF +A + K P |
| Sbjct 481 | ELSRVAA+LRKLGAPFLRAWK | SRARAVRASLIAQ | GARAAICGRYLFNVAVETK | LKLTPLP | |
| Query 539 | GVNSIDVSGWFSIGDSDS | IGDIY 560 | | | |
| Sbjct 541 | EASRLDSGWFTVAGG | -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.



RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

151364 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

3GSZ Go

PDB ID

- 3GSZ

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Bic

This resource is powered by the Protein Data Bank archive information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

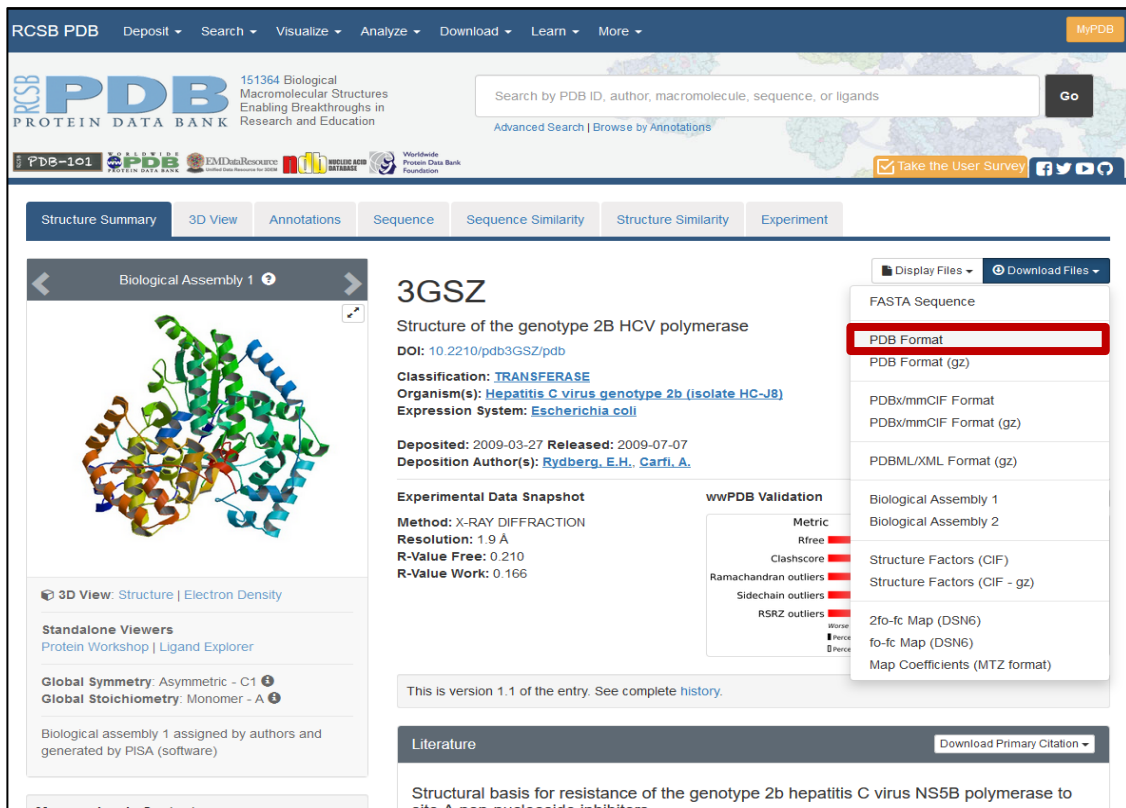
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

New Video: Penicillin and Antibiotic Resistance

Proteins and Biomaterials

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



RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

151364 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

Take the User Survey

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment

3GSZ

Structure of the genotype 2B HCV polymerase

DOI: [10.2210/pdb3GSZ/pdb](https://doi.org/10.2210/pdb3GSZ/pdb)

Classification: [TRANSFERASE](#)

Organism(s): [Hepatitis C virus genotype 2b \(isolate HC-J8\)](#)

Expression System: [Escherichia coli](#)

Deposited: 2009-03-27 Released: 2009-07-07

Deposition Author(s): [Rydberg, E.H., Carfi, A.](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.9 Å

R-Value Free: 0.210

R-Value Work: 0.166

wwPDB Validation

Metric

- Rfree
- Clashscore
- Ramachandran outliers
- Sidechain outliers
- RSRZ outliers

Biological Assembly 1

Biological Assembly 2

Structure Factors (CIF)

Structure Factors (CIF - gz)

2fo-fc Map (DSN6)

fo-fc Map (DSN6)

Map Coefficients (MTZ format)

This is version 1.1 of the entry. See complete [history](#).

Literature

Download Primary Citation

Structural basis for resistance of the genotype 2b hepatitis C virus NS5B polymerase to site A non-nucleoside inhibitors

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**).

```
1 >P1;MODEL
2 sequence:MODEL:::::0.00: 0.00
3 SLSYSWTGALVTATRREERRHPIGPLSNTLITKHNLVYQTTTASASARMTKVTIDREQIILDKHYFDTVTAVKKKASEVTA
4 DLLTWDEVARLTPKNTARSKSGLSGSDVQRQLTRAARRELNMSWQDLLSDSEELIPTVMKNEVFVSSPTARKPARLIVY
5 PDLEVRACEKRAMYDLFQKLPYAVMGKAYGFQYTPRQRVNRLLDMWRHFKNPMGFSDTKCFDSTVTPHDIDTERDIFLS
6 ATLPDEAKTVIKNLTSRLYRGSPMYNSRGDLVGKRECRASGVFPTSMGNTLTNFIKATAAAKAAGLSDPQFLICGDDLVC
7 ITSSKGVVEDEQALREFTSAMTKYSAIPGDLPKPYDLEQITSCSSNVTVAQDRNGRPPYYFLTRDPTTPLARASWETISH
8 SPVNSWLGNI IAFAPT VVRLVFLTHFFGLLLQQDAVDRNYEFEMYGSTYSVNPDLPAI IYKLGHEAFDLTNYSPEV
9 QRVAAALQKLGSPPLRAWKRRAKLDRSKLKVRRGGRYAVVADYLFGFASAYRPKRPAPPVNSIDVSGWFSIGDSDIY
10 RQ*
```

- b. Arquivo PDB da estrutura molde (**3gsz.pdb**).
 - c. Script em linguagem python 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 SMSYTWGALITPCGPPEEKLPINPLNSLMRFHNKVYSTTSRSASLRAKKVTFDRVQVLDHAHYDSVLQDVKRAA
5 SKVSARLLTVEEACALTPPHSAKSRYGFGAKEVRSLSRRVNHIRSVWEDLLEDQHTPIDTTIMAKNEVFCT---
6 DKKPARLIVYPDLGVRVCEKMALYDIAQKLPKAIMGSPSYGFQYSPAERVDFLLKAWGSKKDPMGFSYDTRCFDST
7 VTERDIRTEESIQACSLPQEARTVIHSLTERLYVGGPMTNSKGQSCGYRRCRASGVFTTSMGNTMTCYIKALAA
8 CKAAGIVDPVMLVCGDDLVIISQSQNEEEDERNLRAFTEAMTRYSAPPGDLPPEYDLELITSCSSNVSVALDSR
9 GRRRYFLTRDPTTPIITRAAWETVRHSPVNSWLGNI IQYAPT IWVRMVMIMTHFFSILLAQDTLNQNLNFEMYGAVY
10 SVNPLDLPATIIERLHGLEAFSLHTYSPHELRSVAATLRKLGAPPLRAWKSRARAVRASLIAQGARAICGRYLFN
11 WAVKTKLKLTPLEASRLDLSGWFTVGAGG-GDIYHS*
12
13 >P1;MODEL
14 sequence:MODEL: : : : :: 0.00: 0.00
15 SLSYSWTGALVTATRREERRHPIGPLSNTLITKHNLYQTTTASASARMTKVTIDREQILDKHYFDTVTAVKKA
16 SEVTADLLTWDEVARLTPKNTARSKSGLSGSDVRQLTRAARRELNMSWQDLLSDSEELIPTVMMAKNEVFVSSPT
17 ARKPARLIVYPDLVPRACEKRAMYDLFQKLPYAVMGKAYGFQYTPRQRVNRLDMWRHFKNPMGFSYDTKCFDST
18 VTPHDI DTERDIFLSATLPDEAKTVIKNLTSRLYRGSMPYNSRGLDVGKRECRASGVFPTSMGNTLTNFIKATAA
19 AKAAGLSDPQFLICGDDLVCITSSKGVEEDEQALREFTSAMTKYSATPGDLPKPYDLEQITSCSSNVTVAQDRN
20 GRPYYFLTRDPTTPLARASWETISHSPVNSWLGNI IAFAPT VVVRVLFVLFTHFFGLLLQDDAVDRNYEFEMYGSTY
21 SVNPLDLPATIIYKLGPEAFDLTNYSPEVQRVAAALQKLGSPPLRAWKRRAKLDRSKLKVGRGRYAVVADYLFQ
22 FASAYRPKRPAPPGVNSIDVSGWFSIGDDSIGDIYRQ*
23
```

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


```

1  _aln.pos      10      20      30      40      50      60
2  3gsz          SMSYTWGTGALITPCGP EEEKLPINPLSNLSLMRFHNKVYSTTSRSASLRAKKVT FDRVQVLD AHDYDVL
3  MODEL        SLSYSWTGALVTATRREERRHP IGPLSNTLITKHNLVYQT TTSASARMTKVT IDREQILD KH YFDTV
4  _consrvd     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
5
6  _aln.p       70      80      90      100     110     120     130
7  3gsz          QDVKRAASKVSARLLTVEEACALTP PHSAKSRYGFGAKEVRSLRRAVNHIRSVWEDLLEDQHTPIDT
8  MODEL        TAVKKKASEVTADLLTWDEVARLT PKNTARSKSGLSGSDVRQLTRAARRELSNMWQDLLSDSEELIPT
9  _consrvd     ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
10
11 _aln.pos     140     150     160     170     180     190     200
12 3gsz          TIMAKNEVFCCI--DKKPARLIVY PDLGVRVCEKMALYDIAQKLPKAIMGPSYGFQYSPAERVD FLLK
13 MODEL        TVMAKNEVFVSSPTARKPARLIVY PDLVPRACEKRAMYDLFQKLPYAVMGKAYGFQYTPRQRVNRLLD
14 _consrvd     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
15
16 _aln.pos     210     220     230     240     250     260     270
17 3gsz          AWGSKKDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPQEARTVIHSLTERLYVGGPMTNSKQGQSC
18 MODEL        MWRHFKNPMGFSYDTRCFDSTVTPHDIDTERDIFLSATLPDEAKTVIKNLTSLRYSRSPMYSRGLV
19 _consrvd     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
20
21 _aln.pos     280     290     300     310     320     330     340
22 3gsz          GYRRCRASGVFTTSMGNTMTCYIKALAAACKAAGIVDPVMLVCGDDL VVISESQGNEEDERNLRAFTEA
23 MODEL        GKRECRASGVFTTSMGNTLTFNIKATAA AAGLSDPQFLICGDDLVCITSSKGV EEDQALREFTSA
24 _consrvd     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
25
26 _aln.pos     350     360     370     380     390     400
27 3gsz          MTRYSAPPGLPRPEYDLELITSCSSNVSV ALDSRGRRRYFLTRDPTTPI TRAAWETVRHSPVNSWL G
28 MODEL        MTKYSALPGDLKPKYYDLEQITSCSSNVTVAQDRNGR PYYFLTRDPTTPI LARASWETISHSPVNSWL G
29 _consrvd     ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
30
31 _aln.p       410     420     430     440     450     460     470
32 3gsz          NIIQYAPTIVWRMVMIMTHFFSILLAQDTL NQNLFEMYGAVYSVNPLDLP AI IERLHGLEAFSLHTYS
33 MODEL        NIIAFAPTIVVRLVFLTHFFGLLQQDAVDRNYEFEMYGSTYSVNPLDLP AI IYKLGPEAFDLTNY S
34 _consrvd     *** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
35
36 _aln.pos     480     490     500     510     520     530     540
37 3gsz          PHEL SRVAATLRKLGAPPLRAWKSRARAVRASLI AQGARAAICGRYLFNWAVKTKLKLTP LPEASRLD
38 MODEL        PYEVQRVAAALQKLGSPPLRAWKRRAKLD RSKLKVRRYAVVADYLF GFASAYRPKRPAPPGVNSID
39 _consrvd     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
40
41 _aln.pos     550     560
42 3gsz          LSGWFTVGAGG-GDIYHS
43 MODEL        VSGWFSIGDDSIGDIYRQ
44 _consrvd     **** * * * * *
45

```

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 python 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.

```

3208 >> Summary of successfully produced models:
3209 Filename                molpdf          DOPE score      GA341 score
3210 -----
3211 MODEL.B99990001.pdb      2874.98608     -66463.61719    1.00000
3212 MODEL.B99990002.pdb      2787.88892     -66913.05469    1.00000
3213 MODEL.B99990003.pdb      2843.51221     -66461.59375    1.00000
3214 MODEL.B99990004.pdb      2725.74146     -66682.02344    1.00000
3215 MODEL.B99990005.pdb      3027.27954     -66237.45312    1.00000
3216 MODEL.B99990006.pdb      2714.04736     -66634.33594    1.00000
3217 MODEL.B99990007.pdb      2993.76465     -66044.28906    1.00000
3218 MODEL.B99990008.pdb      2895.72876     -66338.40625    1.00000
3219 MODEL.B99990009.pdb      2792.26758     -67115.99219    1.00000
3220 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.

Index of /bioufmg/olhe

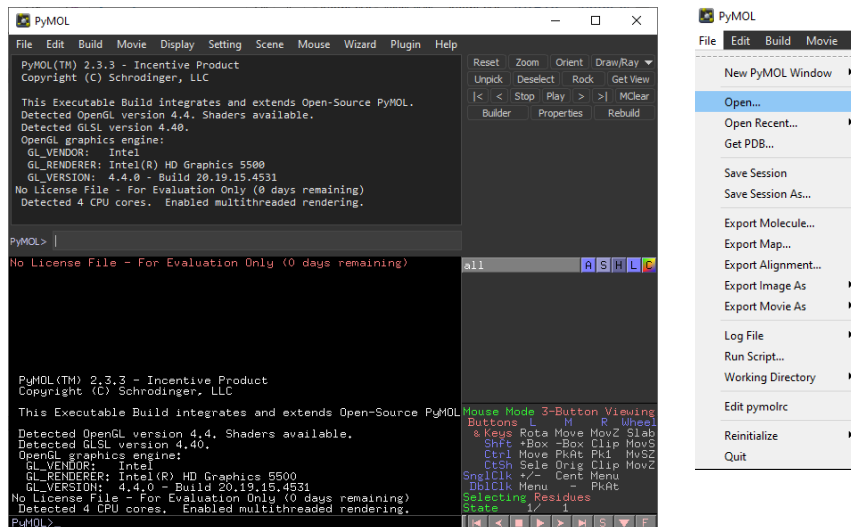
| Name | Last modified | Size | Description |
|------------------|------------------|------|-------------|
| Parent Directory | | - | |
| INTRO/ | 2020-09-17 14:27 | - | |
| ace222/ | 2020-08-27 10:15 | - | |
| amandacpa/ | 2020-10-05 14:39 | - | |
| anna_menezes/ | 2020-10-11 19:17 | - | |
| arthurtrcf/ | 2020-10-13 13:10 | - | |
| barbaramas/ | 2020-10-12 12:37 | - | |
| beatrizappgava/ | 2020-10-13 15:50 | - | |
| caizin/ | 2020-10-09 14:15 | - | |
| camilabd/ | 2020-09-28 16:25 | - | |
| camilaclic/ | 2020-08-13 10:49 | - | |
| ceciliahorta/ | 2020-10-12 16:54 | - | |
| clecio_alonso/ | 2020-10-13 11:43 | - | |
| danikayali/ | 2020-10-13 12:52 | - | |
| dg_sousa/ | 2020-10-07 17:55 | - | |
| elisaamancio/ | 2020-10-12 16:56 | - | |
| eusoujacu/ | 2020-10-14 14:15 | - | |
| g.../ | 2020-10-14 13:11 | - | |

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.

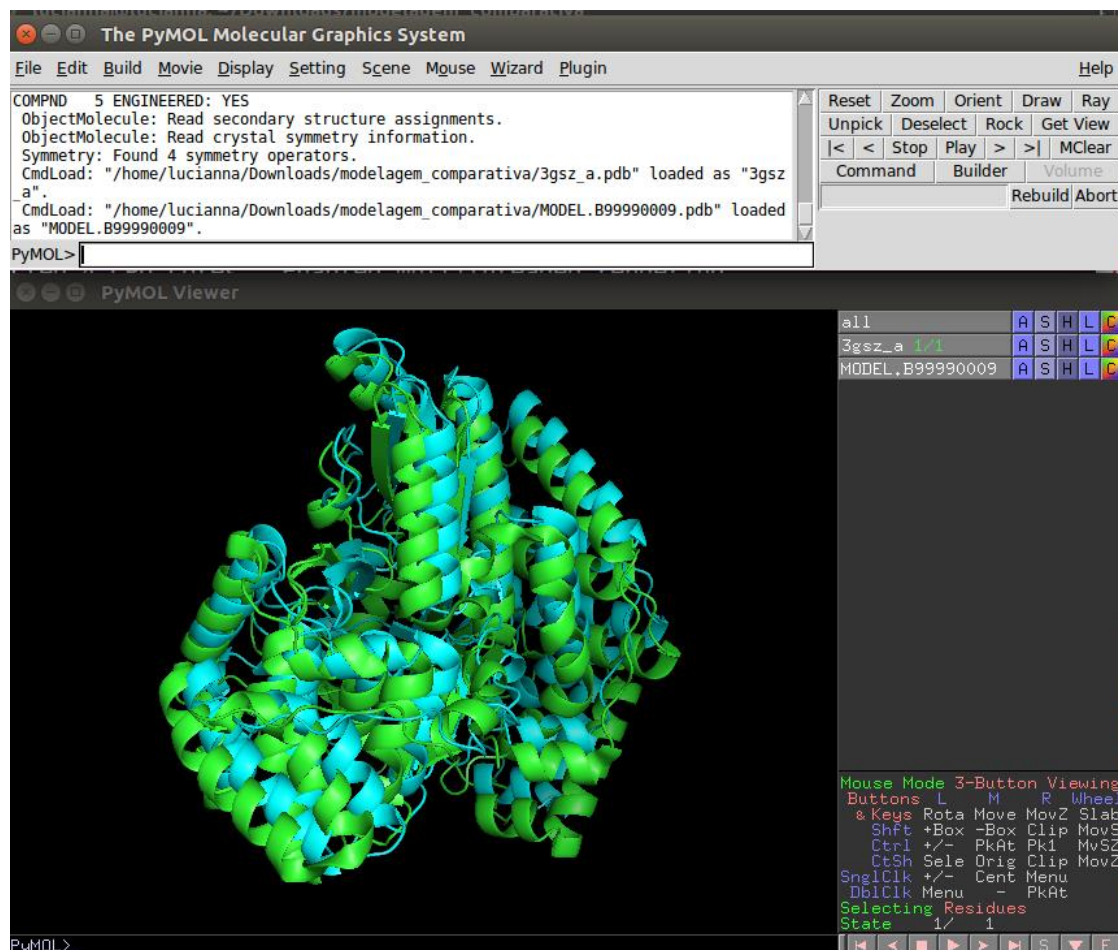
d. Abrir o pymol (após instalação): 3gsz_a.pdb e MODEL.B99990009.pdb





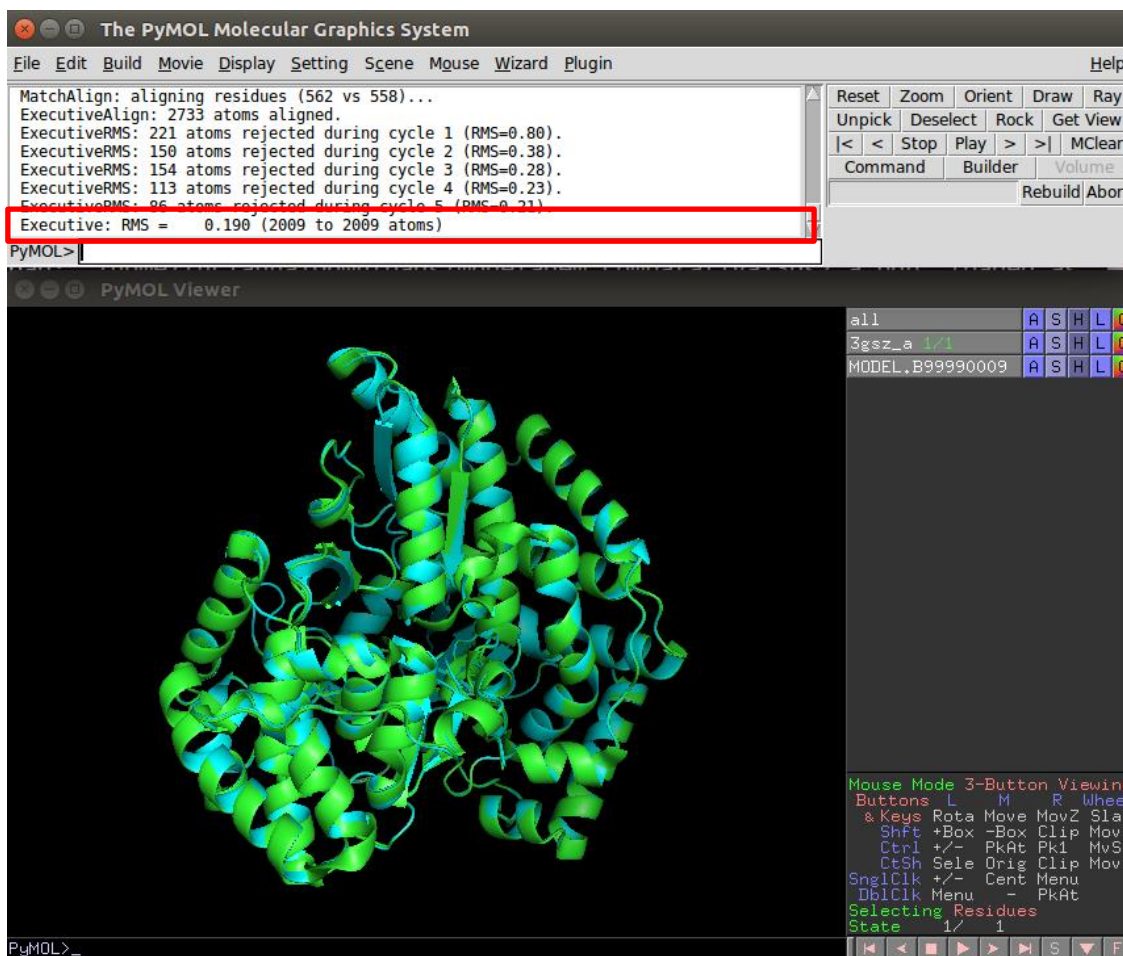
No Linux (após instalação):

`$ 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.

Structure Assessment

Help Examples ▾

Start a new Structure Assessment Project

Structure File:

Project Title (Optional):

Email (Optional):

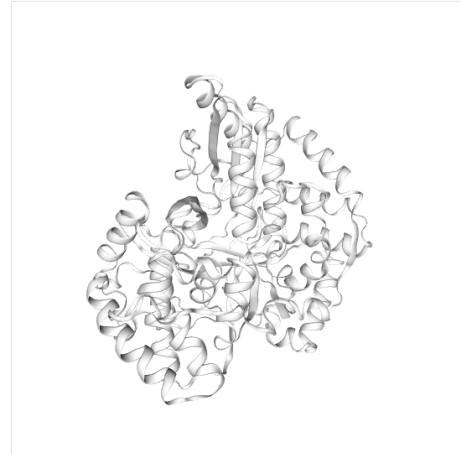
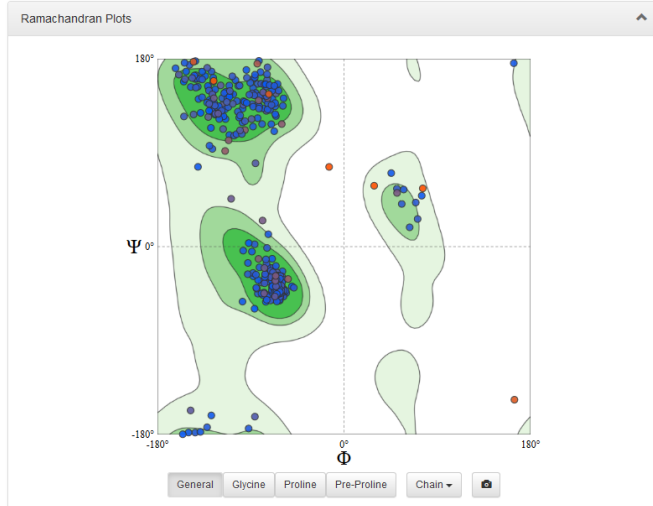
c) Analisar Resultados.

Structure Assessment

Help Examples ▾

Structure Assessment of MODEL.B99990009.pdb; ⏏ 📄 ✕

Created: Sun 28 Apr 2019, 01:24;



📷 NGL 📷 Cartoon 📷 ▶ 📷 ↺