

Modelagem de Proteínas

Para essa prática, usaremos a proteína N-myristoyltransferase de *Leishmania donovani*, cuja sequência, no formato FASTA é:

>XP_003863427.1 N-myristoyltransferase [Leishmania donovani]

```
MSRNPSNSDAAHAFWSTQPVPQTEDETEKIVFAGPMDEPKTVADIPEEPYPIASTFEWWTPNME  
AADDIHAIYELLRDNYVEDDDSMFRFNYSSEFLQWALCPPSYIPDWHVAVRRKADKKLLAFIAGVP  
VTLRMGTPKYMKVKAQEKGEQEEAAKYDAPRHICEINFLCVHKQLREKRLAPILIKEVTRRVNRTNV  
WQAVYTAGVLLPTPYASGQYFHRSLNPEKLVEIRFSGIPAQYQKFQNPAMMLKRNQYQLPNAPKNS  
GLREMKPSDVPQVRRILMNYLDNFDVGPVFSDAEISHYLLPRDGVVFTYVVENDKKVTDFFSFYRIP  
STVIGNSNYNINLAAYVHYAATSMPLHQLLDLLIVAHSRGFDVCNMVEILDNRSFVEQLKFGAGD  
GHLRYFFYNWGYPKIKPSRVALVML
```

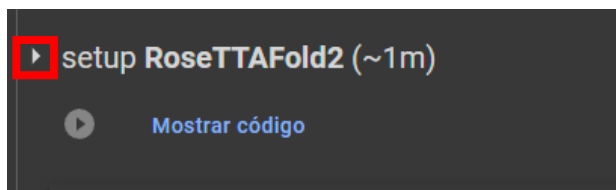
Essa sequência já tem estrutura experimentalmente resolvido (PDB: 2WUU), porém sem uma das regiões.

Para essa prática, usaremos o programa RosettaFold, no notebook disponibilizado pela iniciativa ColabFold (todos os notebooks estão disponíveis em <https://github.com/sokrypton/ColabFold>).

O notebook em que trabalharemos e está disponível na página <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/RoseTTAFold2.ipynb#scrollTo=oJTZGgdeKkO>

1º Passo – Setup

Aperte **play** no item setup RoseTTAFold2 (~1m):



2º Passo – Configuração e Execução

No item run RoseTTAFold2:

Adicione a **sequência** e **nomeie o seu projeto** (essa função não está consistente, recomendo não mudar)

```
sequence: "ARMNRPAPVEVTYKNMRFLLTHNPNTATLNKFIELKKYGVTTIVRVCEATYDTTLVEKEGHHVLDWPFDDGAPPSNQIVDDWLSLVKIKFREEPGCCIAVHCVAGLGRAPVLVALALIEGGMKYEDAVQFIRQKRRGAFNSKQLLYLEKYRPMRLRF"  
jobname: "Human protein tyrosine phosphatase 4A1"
```

Não faça alterações nessa parte

```
symmetry settings  
sym: X  
order: 1  
msa_concat_mode: diag  
msa settings  
msa_method: mmseqs2  
pair_mode: unpaired_paired  
collapse_identical: ☐  
RoseTTAFold2 settings  
num_recycles: 6
```

Mude o **número de modelos** para 4

```
stochastic settings  
use_mlm: ☐  
use_dropout: ☐  
max_msa: 256  
random_seed: 0  
num_models: 4
```

Aperte **play**

```
run RoseTTAFold2
sequence: "ARMNRPAPVEVTYKNMRLITHNPTNATLNKFIIEELKKYGVTTIVRVCEATYDTLVEKEGHHVLDWPFDDGAPPNSQIVDDWL SLVKIKFREEPGCCIAVHCVAGLGRAPVLVALALIEGGMKYEDAVQFIRQKRRGAFNSKQLLYLEKYRPMRLRF"
jobname: "test"

symmetry settings
sym: X
order: 1
msa_concat_mode: diag

msa settings
msa_method: mmseqs2
pair_mode: unpaired_paired
collapse_identical: ☐

RoseTTAFold2 settings
num_recycles: 6

stochastic settings
use_nla: ☐
use_dropout: ☐
max_msa: 256
random_seed: 0
num_models: 1

Mostrar código
```

3º Passo – Visualização

Aperte **play**

```
Display 3D structure
color: plddt
Mostrar código
```

Explore seus modelos

2. Validação

O modelo está colorido de acordo com o plddt. Como está em beta só o primeiro será analisado assim.

Você nota algo?

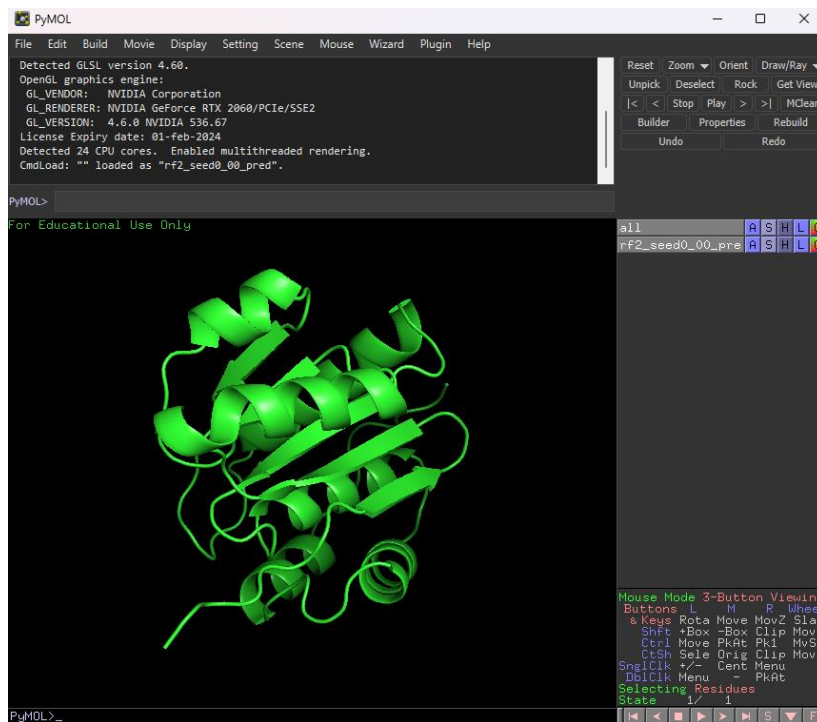
Vamos analisar todos os modelos usando o pymol.

Abra cada um dos arquivos.

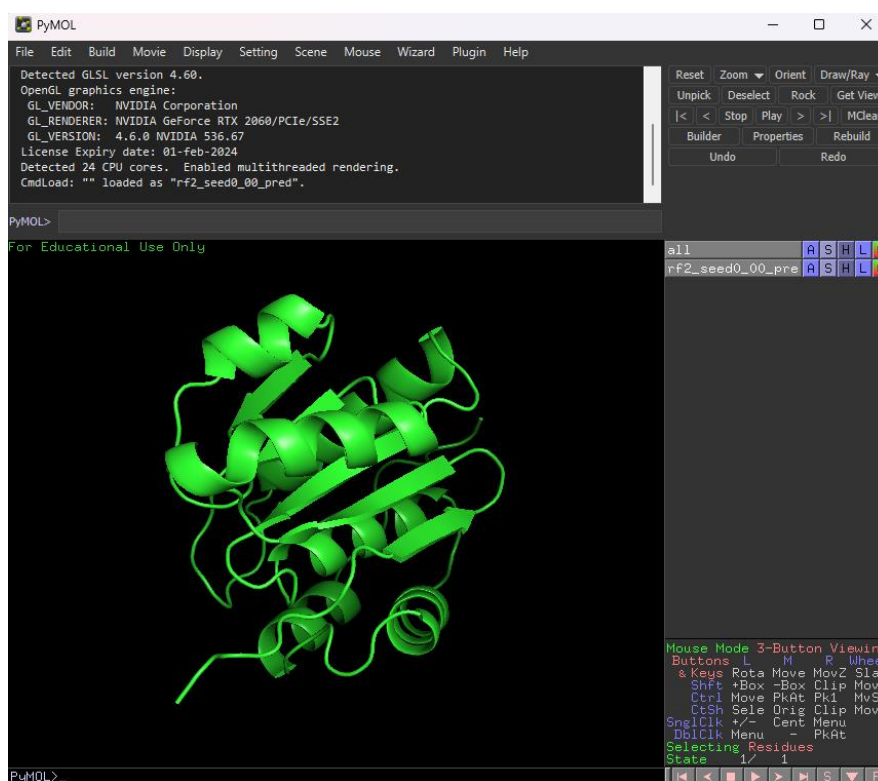
A tela parecerá isso:



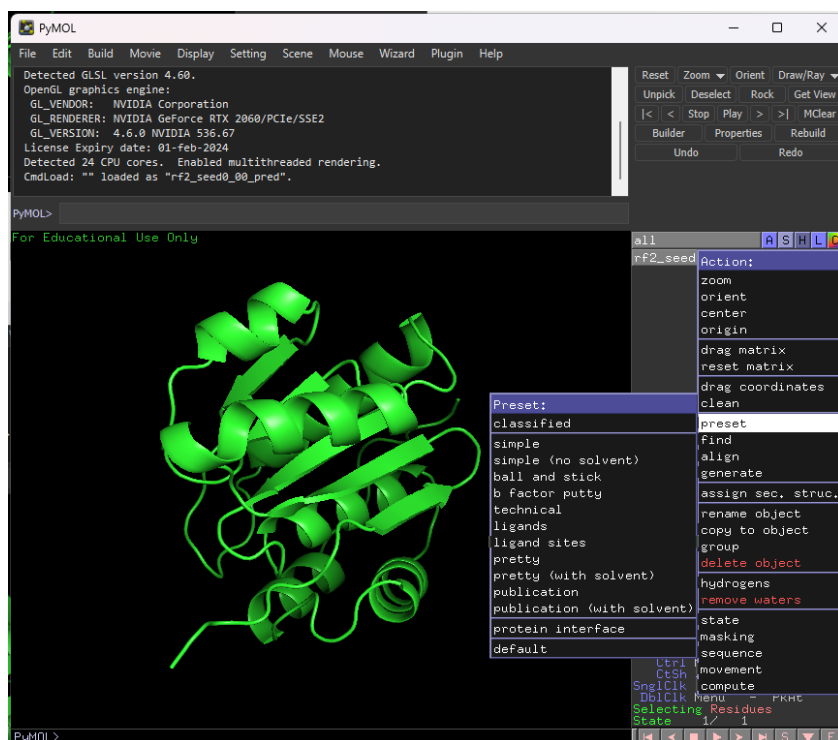
Introdução a modelagem, docking e dinâmica molecular - Vicente Salgado Pires e RSG-Brazil



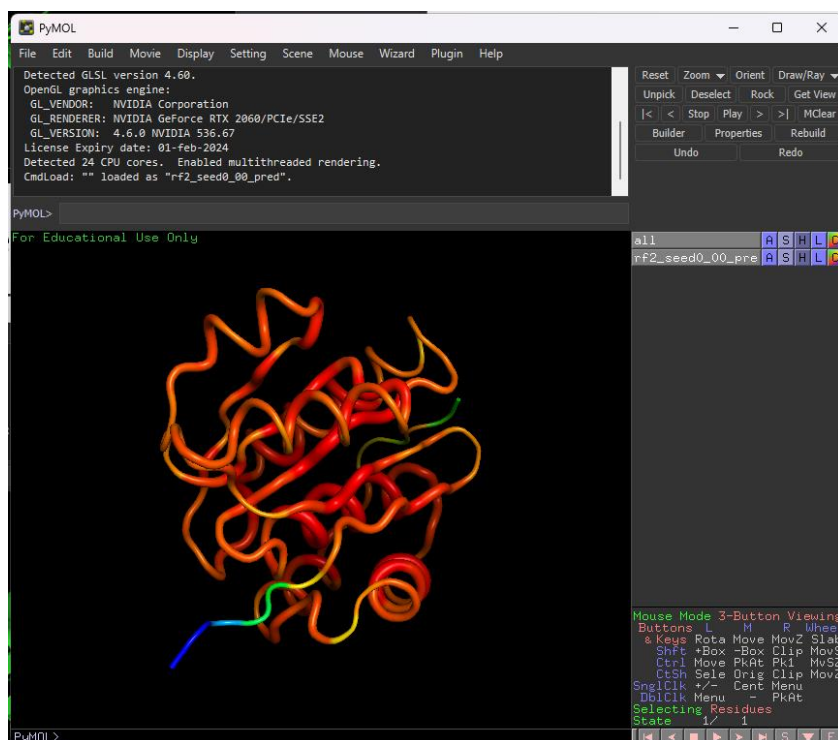
Selecione a opção A, na barra à direita:



Selecione preset e depois b-factor putty:



A estrutura refletirá o plddt, resíduos mais espessos e vermelhos têm maiores scores:

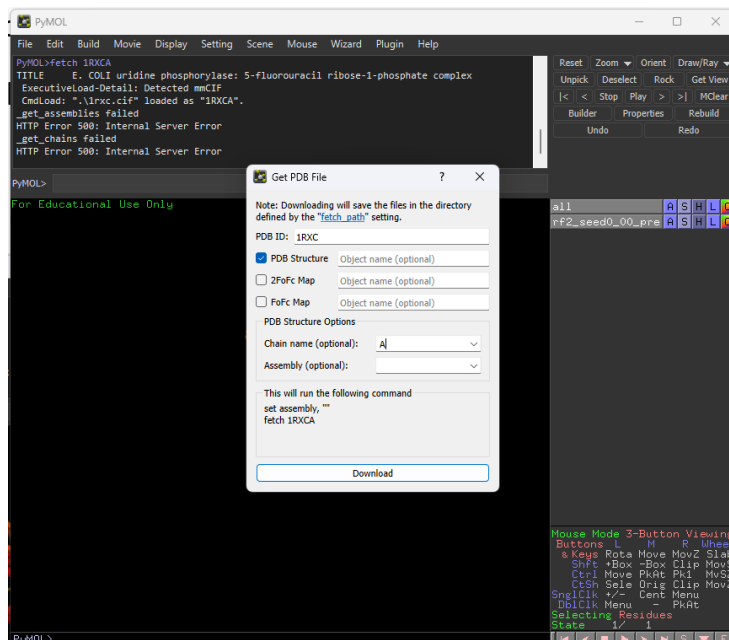
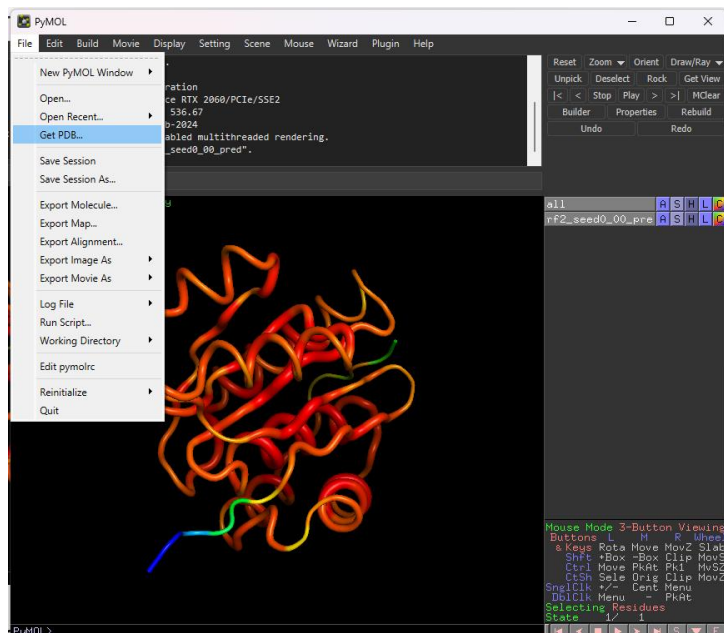


Faça o mesmo para todos os seus modelos e anote seus resultados.

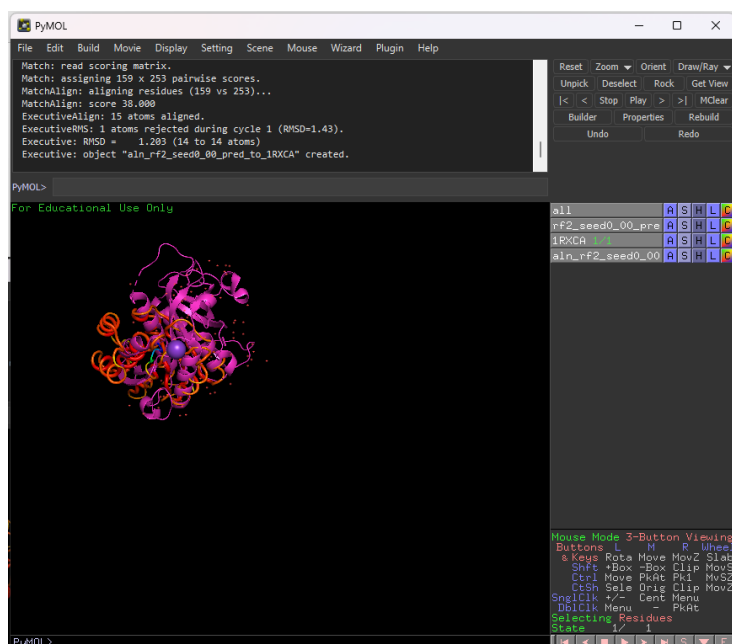
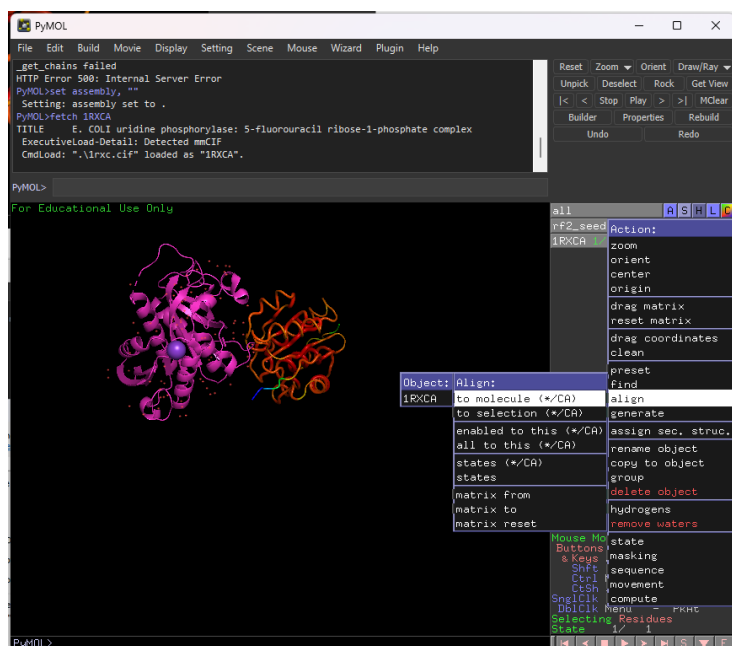
Veremos agora a qualidade da estrutura de acordo com uma referência.

A estrutura da cadeia A da entrada 1RXC do PDB representa a mesma proteína, ainda que incompleta.

Siga os passos para baixar a estrutura de referência:



Agora alinhe a estrutura em relação à estrutura de referência:



Faça isso com todos os seus modelos.

Qual o RMSD?

As estruturas são parecidas?