



Predição e o COVID-19

Conseguimos descobrir se
uma variante vai crescer
antes dela espalhar?

Predição e o COVID-19

Conseguimos descobrir se uma variante vai crescer antes dela espalhar?

Conseguimos entender os fatores que afetam a frequência de uma mutação?

Análise de Frequência de Mutações no SARS-CoV2

Francisco Galuppo

Mathias Batista

Pedro Souza

Ramon Durães

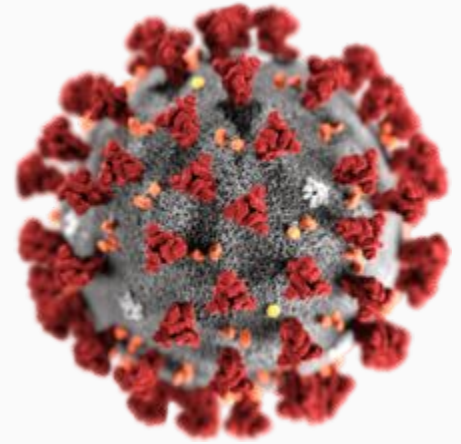
Thaís Soares



Coleta dos dados



Swiss Institute of
Bioinformatics



“Servidor de modelagem automática de estruturas de proteínas com propósito de tornar a modelagem acessível a todos os pesquisadores das áreas biológicas no mundo todo.”

P0DTC2	17	17	#ffdfdf	-	0.000030 (N17K, N17T, N17Y)	Yes
P0DTC2	18	18	#ffaeae	-	0.051930 (L18F, L18I)	Yes
P0DTC2	19	19	#fff9f9	-	0.000320 (T19A, T19I, T19K, T19P, T19R, T19S)	Yes
P0DTC2	20	20	#fff7f7	-	0.000550 (T20A, T20I, T20N, T20S)	Yes
P0DTC2	21	21	#ffebeb	-	0.002930 (R21G, R21I, R21K, R21S, R21T)	Yes
P0DTC2	22	22		-	0.000550 (T22A, T22I, T22N, T22P)	Yes
P0DTC2	23	23				Yes
P0DTC2	24	24				Yes
P0DTC2	25	25	#fff5f5			Yes
P0DTC2	26	26	#fff1f1	-	0.001470 (P26F, P26H, P26L, P26R, P26S, P26T, P26Y)	Yes
P0DTC2	27	27	#fff5f5	-	0.000810 (A27E, A27K, A27S, A27T, A27V)	Yes
P0DTC2	28	28	#fffafa	-	0.000190 (Y28D, Y28F, Y28H, Y28N)	Yes
P0DTC2	29	29	#fff5f5	-	0.000780 (T29A, T29I, T29S)	Yes
P0DTC2	30	30	#fffcfc	-	0.000050 (N30*, N30D, N30H, N30K, N30S, N30T)	Yes

Desenvolvimento de ferramentas de scraping

Dados de aminoácidos



Artigos

Hidropatia - Kyte, J. e
Doolittle (1982)

Volume - Zamyatnin, A.A.
(1972)

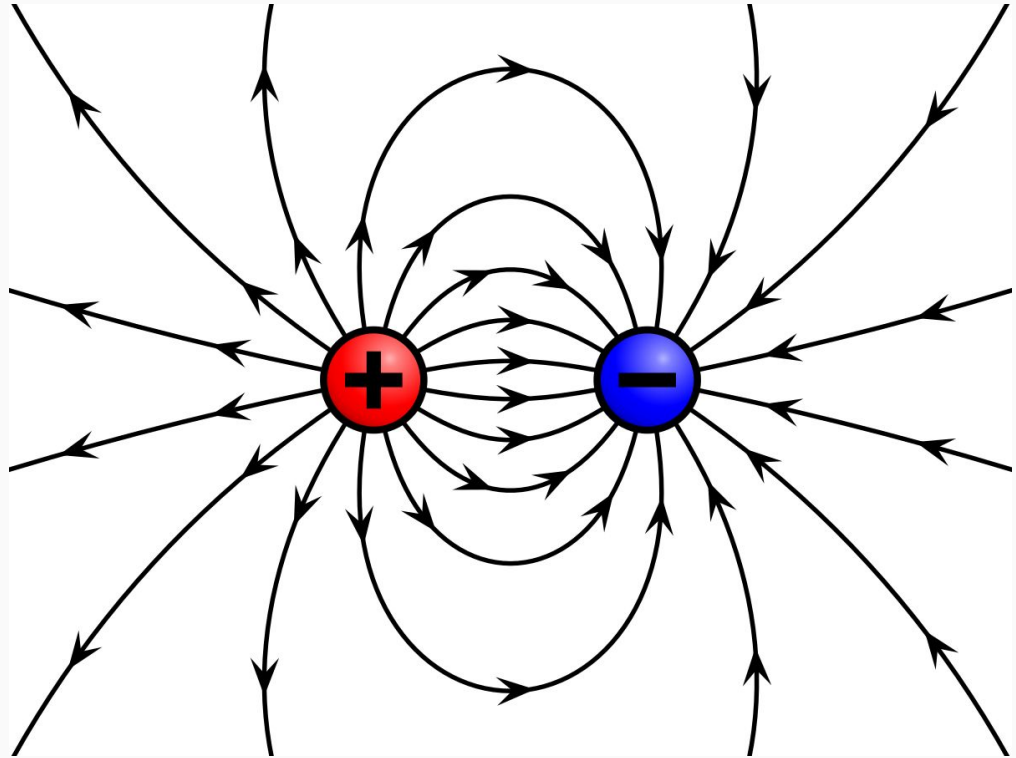
- https://swissmodel.expasy.org/repository/user_annotation_project/BdWKmZ
- JSONS <https://swissmodel.expasy.org/repository/uniprot/P0DTC1.json>
(sequencias)
- Aminoácidos
 - Hidropatia - Kyte, J. e Doolittle (1982)
 - Volume - Zamyatnin, A.A. (1972)
 - Ponto isoelétrico - Biopython
 - Carga em pH 7 - Biopython

Extração de características

Carga elétrica na água

Propriedade física que faz a matéria sofrer força quando em um campo eletromagnético. Cargas similares se repelem, e opostas se atraem.

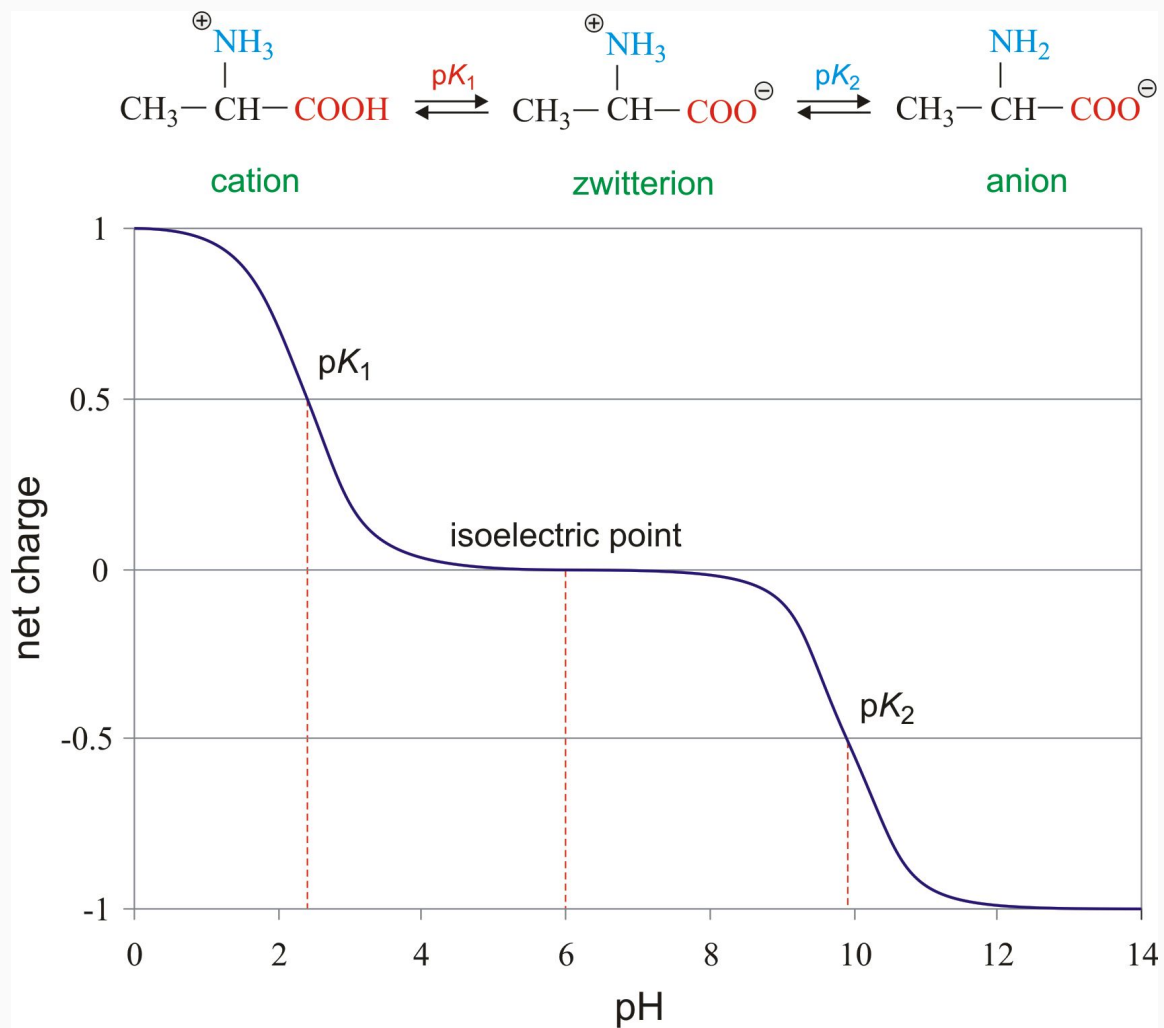
Muitas interações acontecem em meio aquoso, a carga dos aminoácidos nesse meio afeta a estrutura resultante.



Ponto Isoelétrico

Ponto isoelétrico é o valor de pH onde um aminoácido apresenta carga elétrica líquida igual a zero.

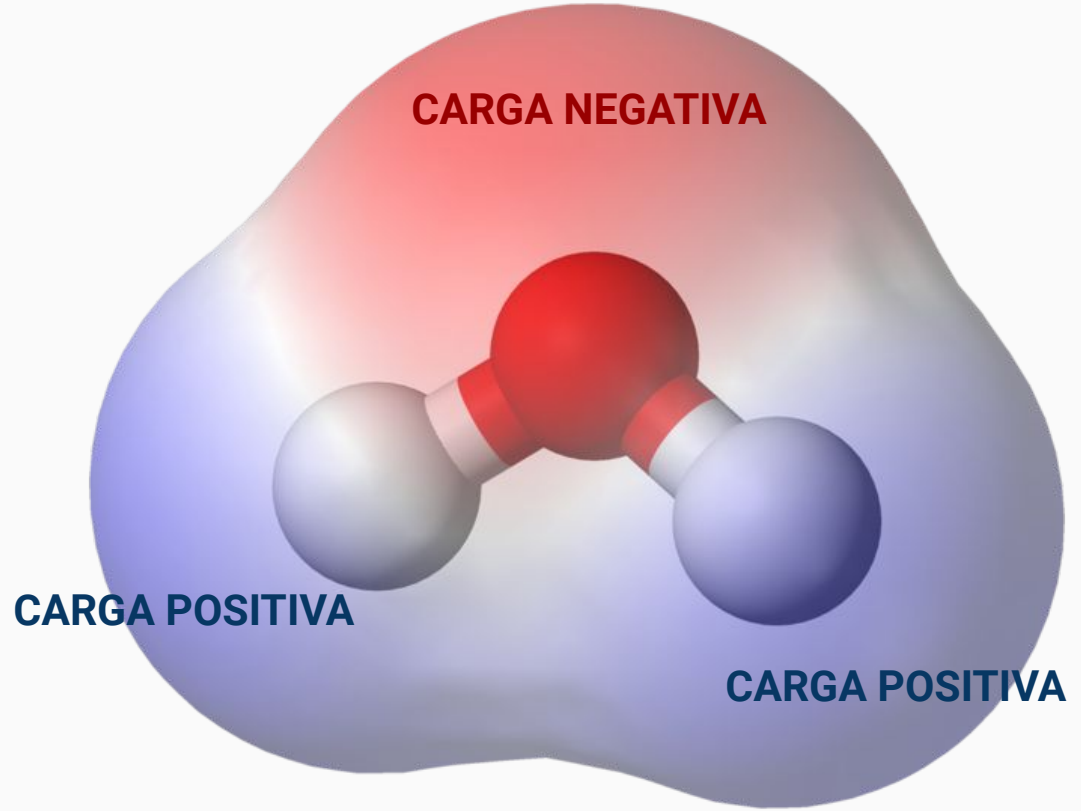
Como vimos, carga é importante para as interações. Essa métrica dá um referencial de como o meio afeta a carga.



Polaridade

Separação das cargas elétricas levando a molécula a ter uma região com carga negativa e outra com carga positiva.

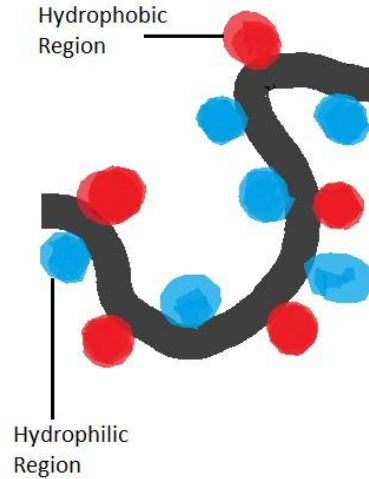
Essa polaridade pode interagir com a polaridade de outras moléculas, podendo criar ligações dipolo-dipolo. Importante para a estrutura da proteína.



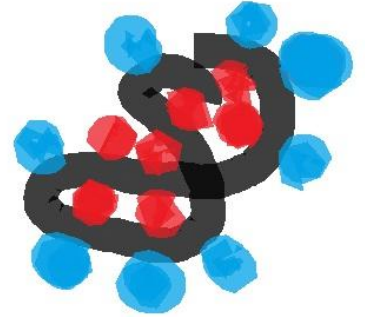
Hidropatia

É uma escala de hidrofilia/hidrofobia de aminoácidos. Quanto maior, mais hidrofóbica é a molécula.

Essa medida é comum de ser utilizada para prever características de proteínas que interagem com membranas biológicas.



Isolated Protein

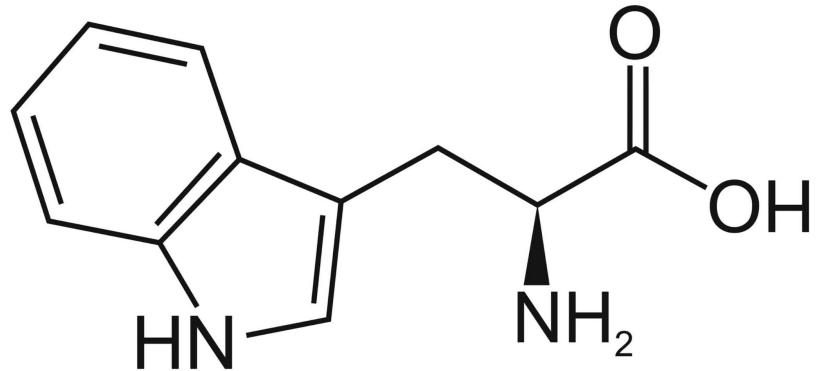
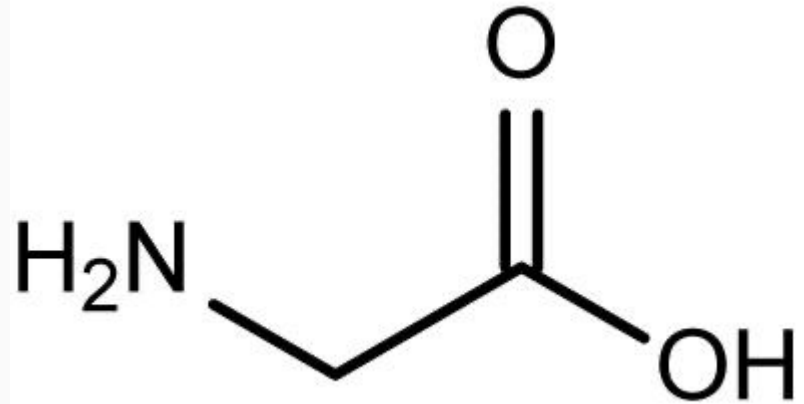


Protein in aqueous solution

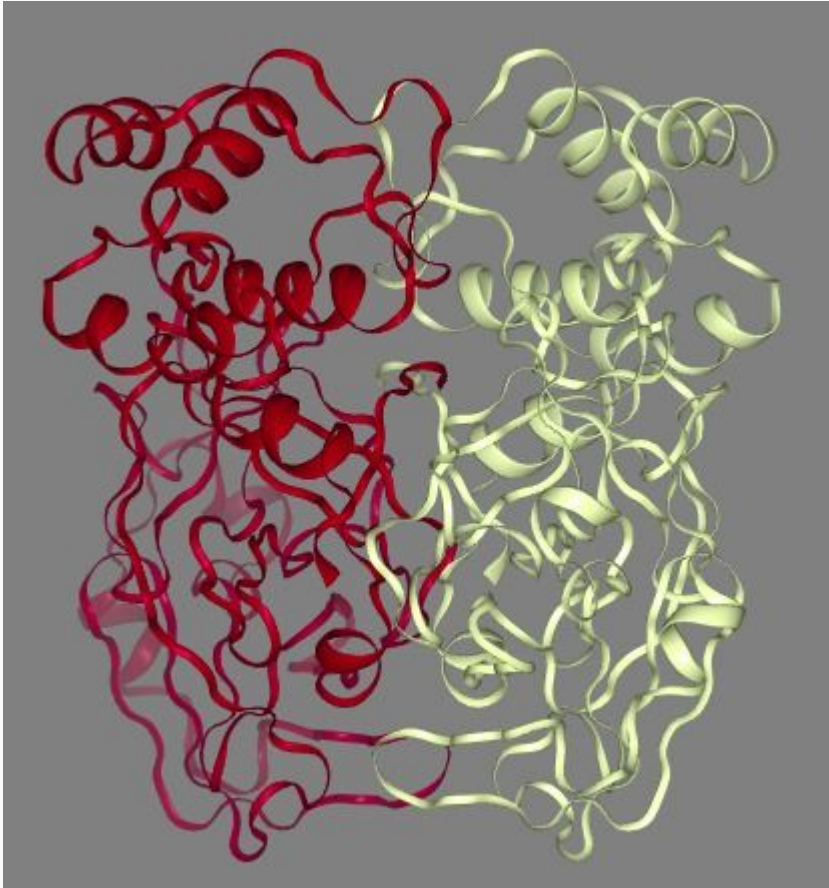
Volume

Calculado a partir da cadeia lateral do aminoácido.

Macromoléculas adquirem determinada conformação e ocupam certo volume dependendo das moléculas que a compõem, as interações entre elas e as interações com o solvente. O volume dos aminoácidos é uma forma de estimar o volume da proteína.



Limitações

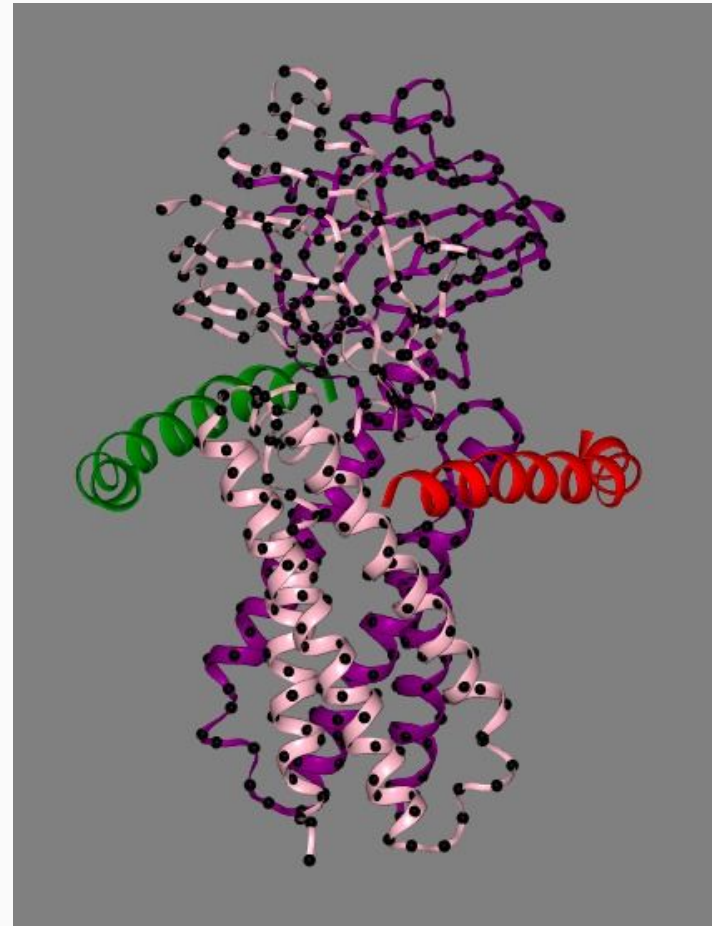


- As frequências são agregadas, não temos os valores reais
- Poucas estruturas reais ou modeladas para usarmos como entrada dos modelos
- As posições das mutações não são condizentes em todas as fontes de dados

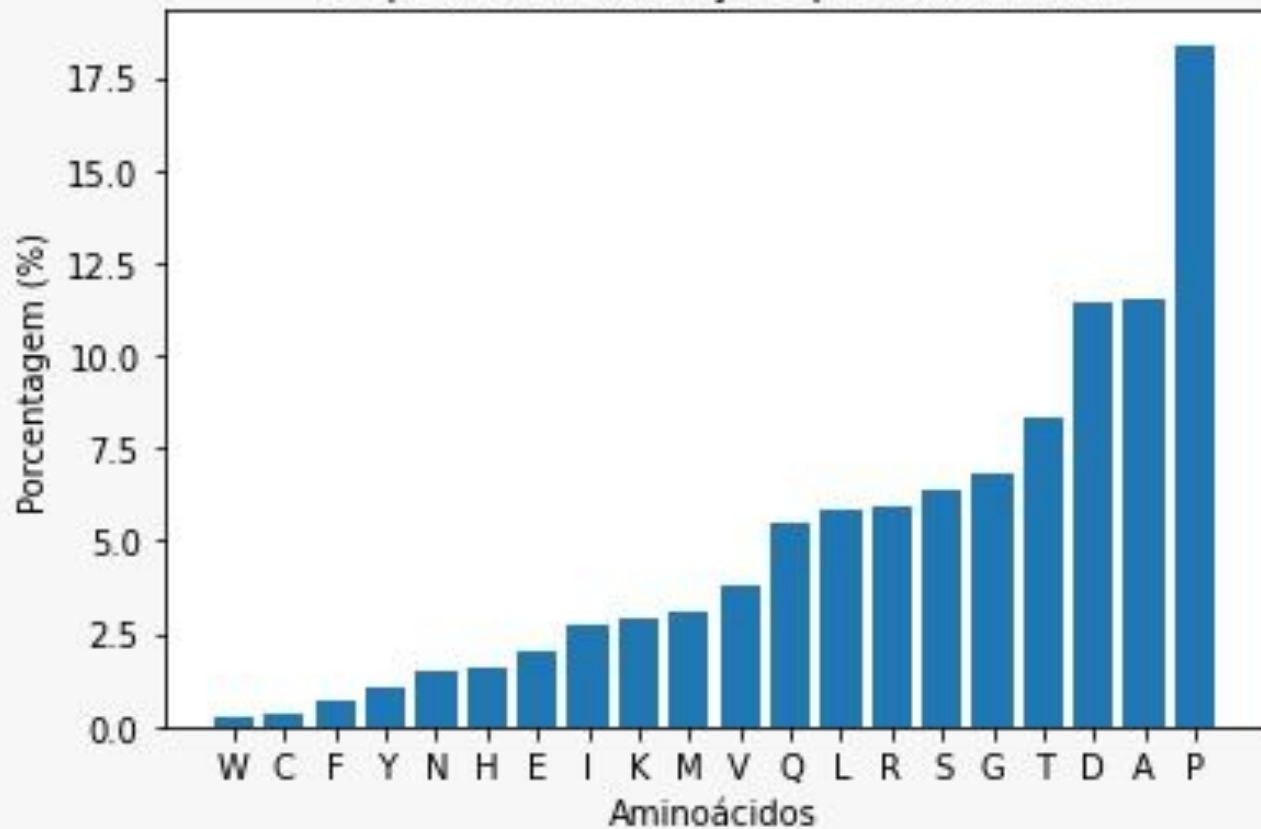
Análises

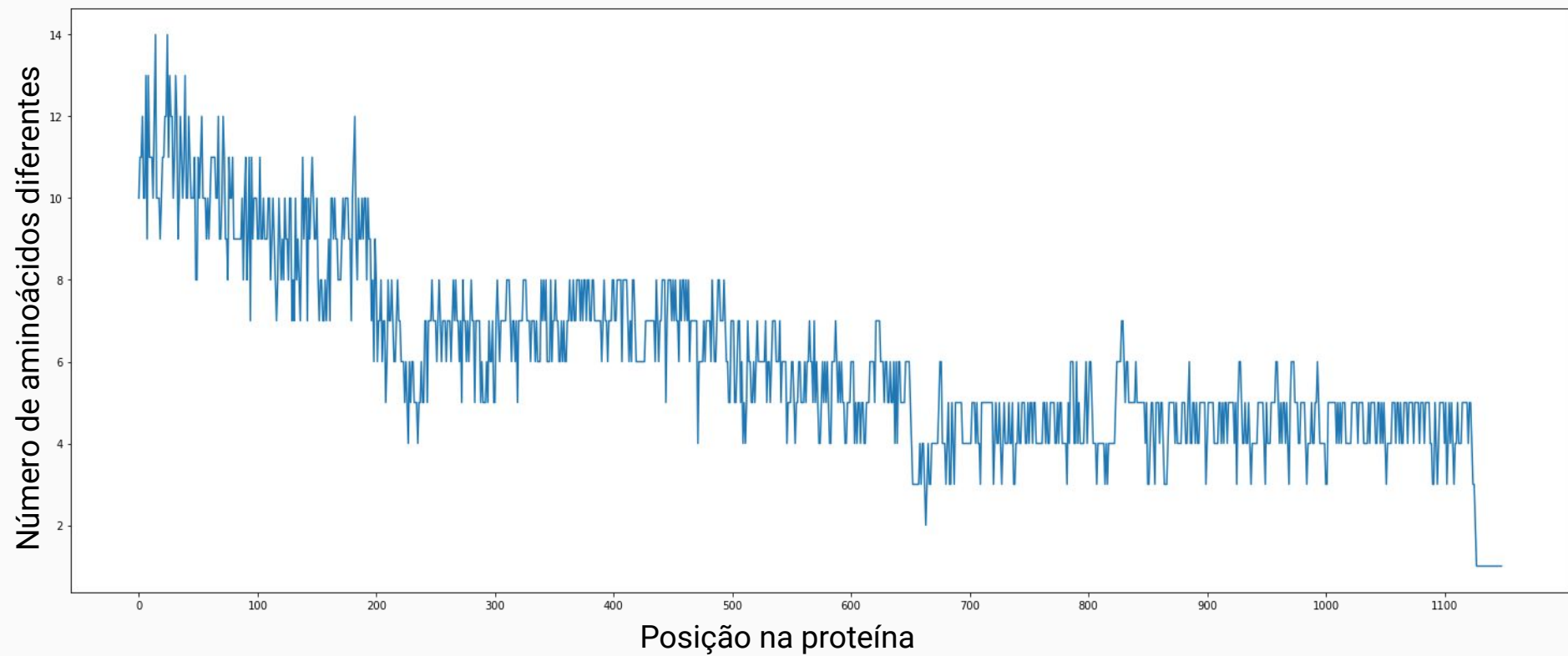
P0DTC3 (AP3A_SARS2)

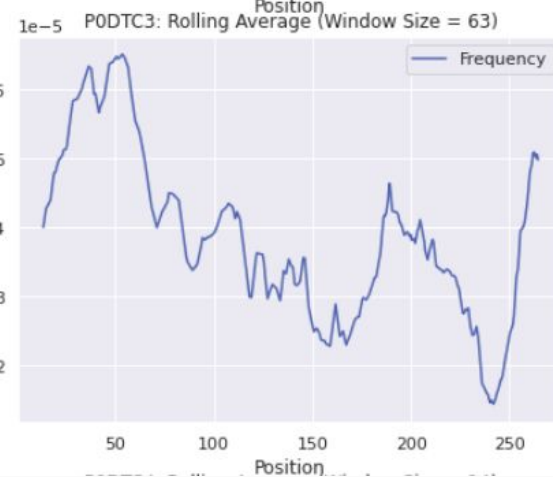
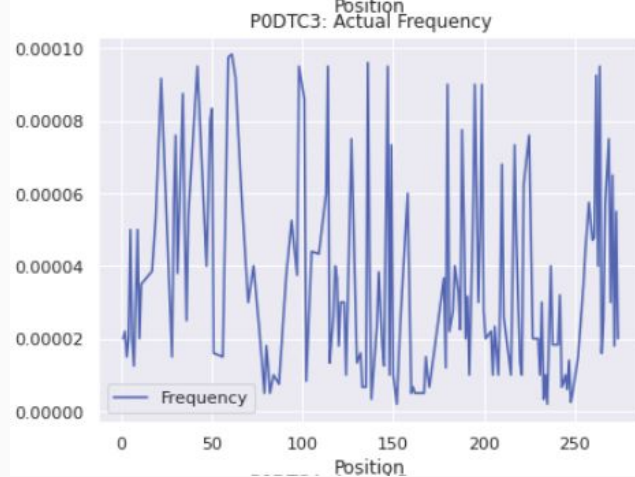
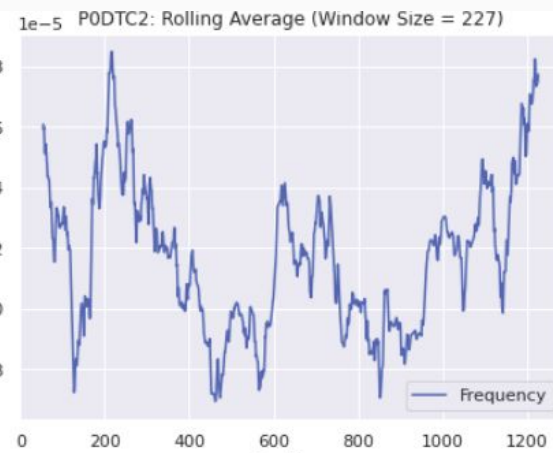
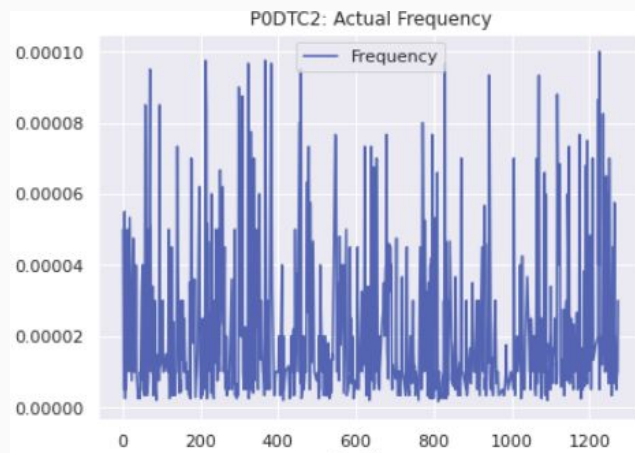
ORF3a protein

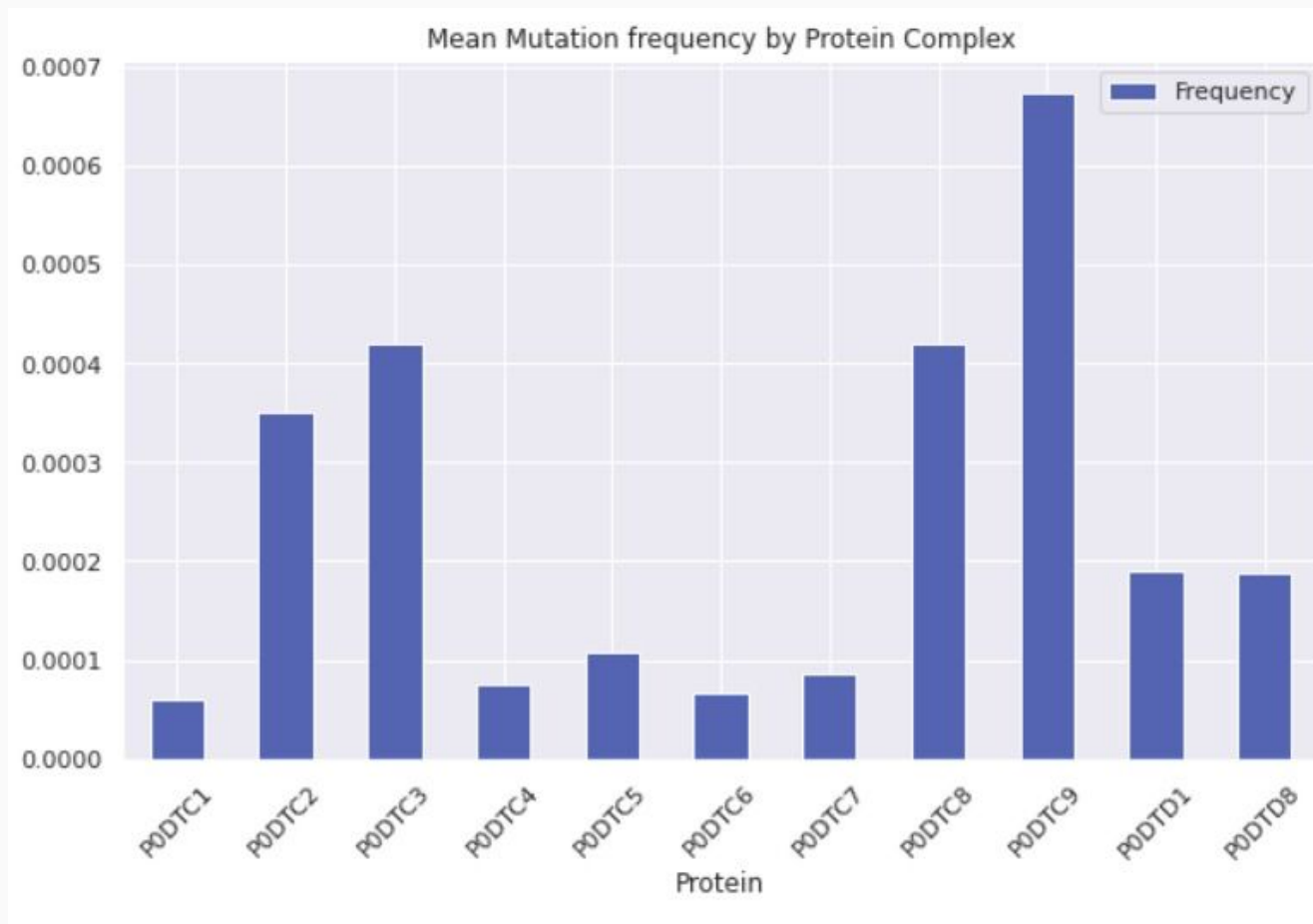


Frequência de mutações por aminoácido



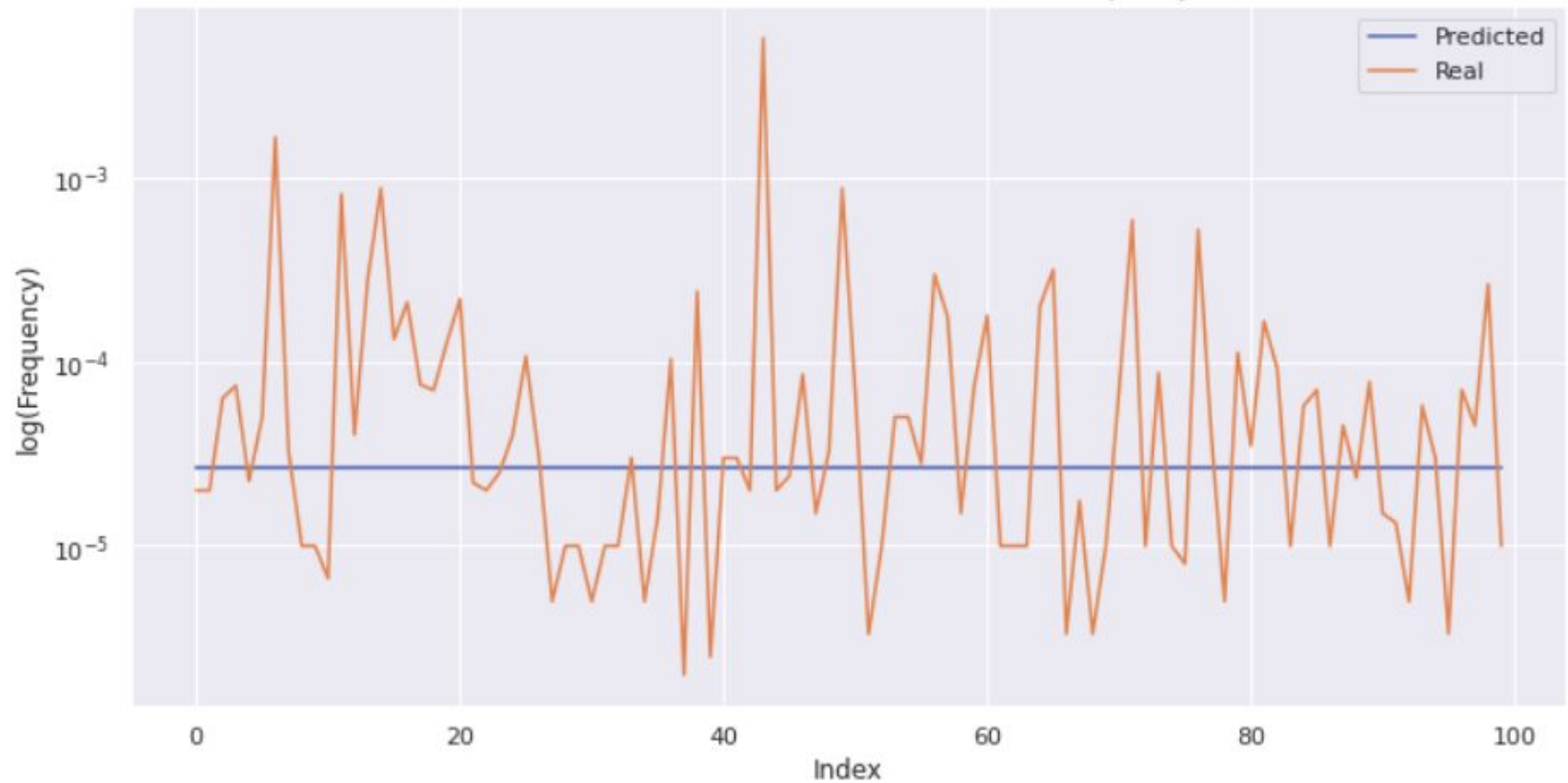




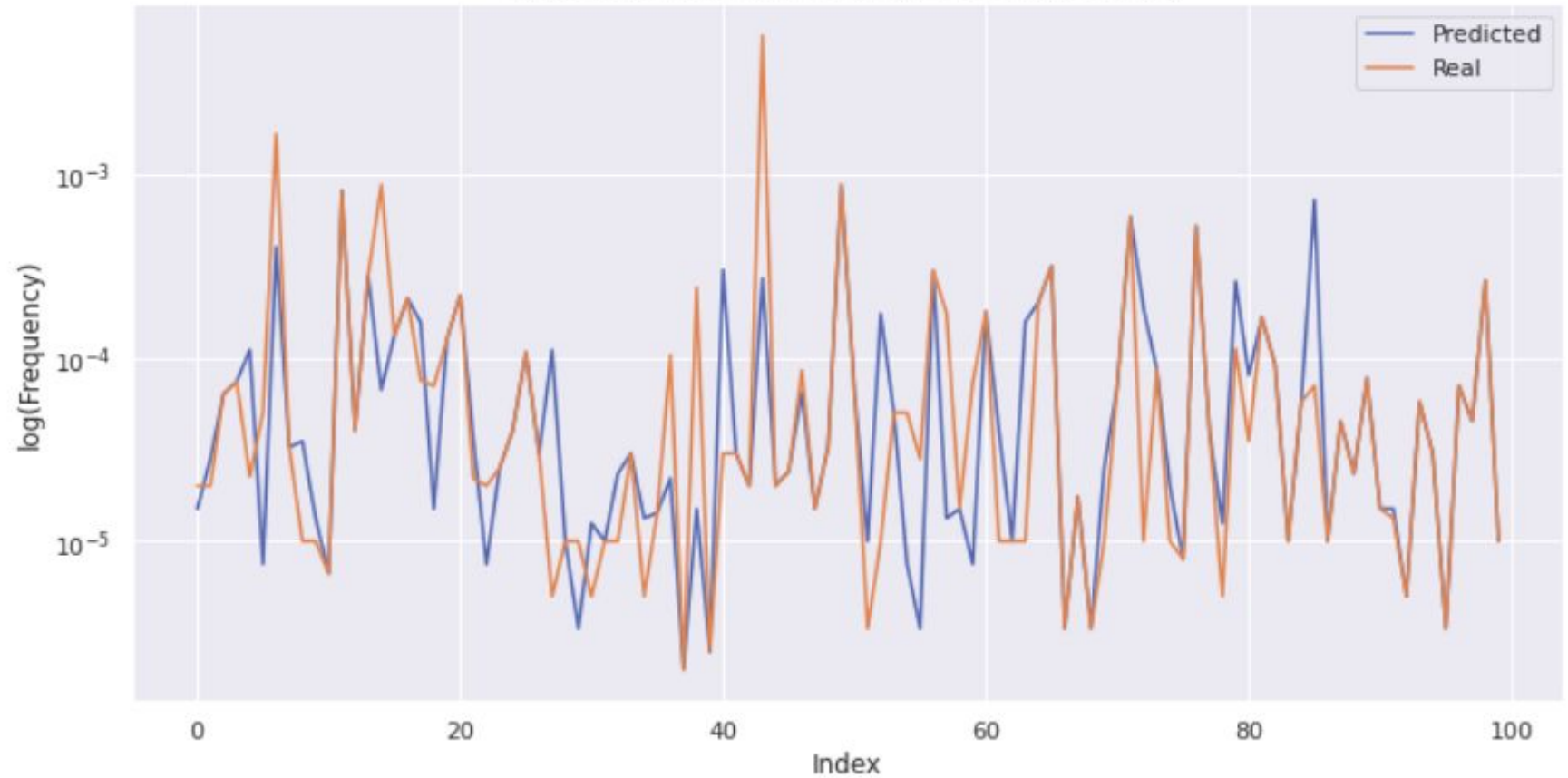


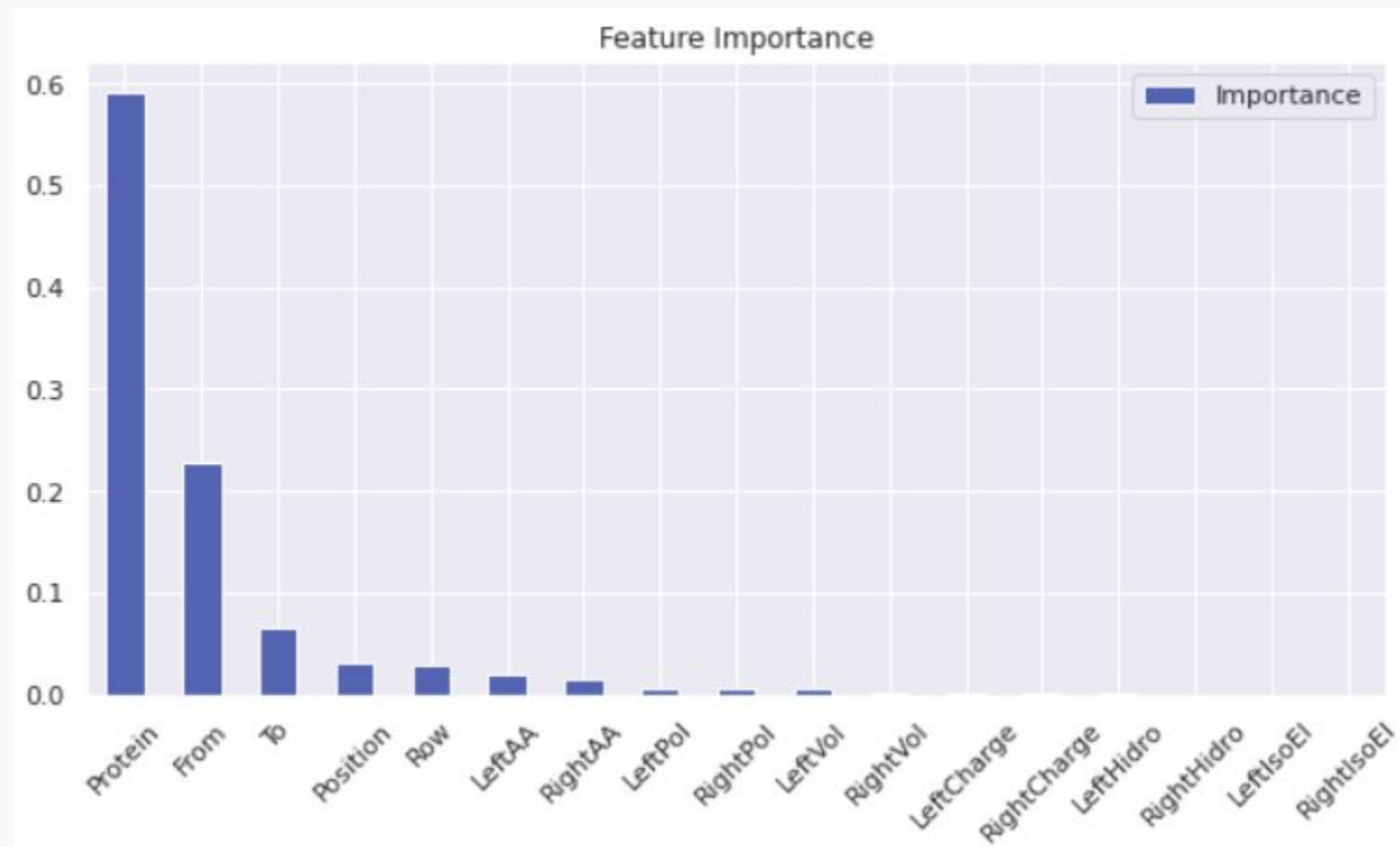
Modelos e Resultados

MEAN MODEL: Predicted vs Read Mutation Frequency



CART MODEL: Predicted vs Read Mutation Frequency





Métricas

	RMSE	MAE	Relative_RMSE	Relative_MAE
Mean	0.002286	0.000231	1.00	1.00
Linear	0.002258	0.000418	0.99	1.81
CART	0.000359	0.000047	0.16	0.20
RF	0.000700	0.000141	0.31	0.61

Análise de Frequência de Mutações no SARS-CoV2

https://github.com/ramongduraes/bioinfo_final

Francisco Galuppo

Mathias Batista

Pedro Souza

Ramon Durães

Thaís Soares

