

Introdução à Bioinformática

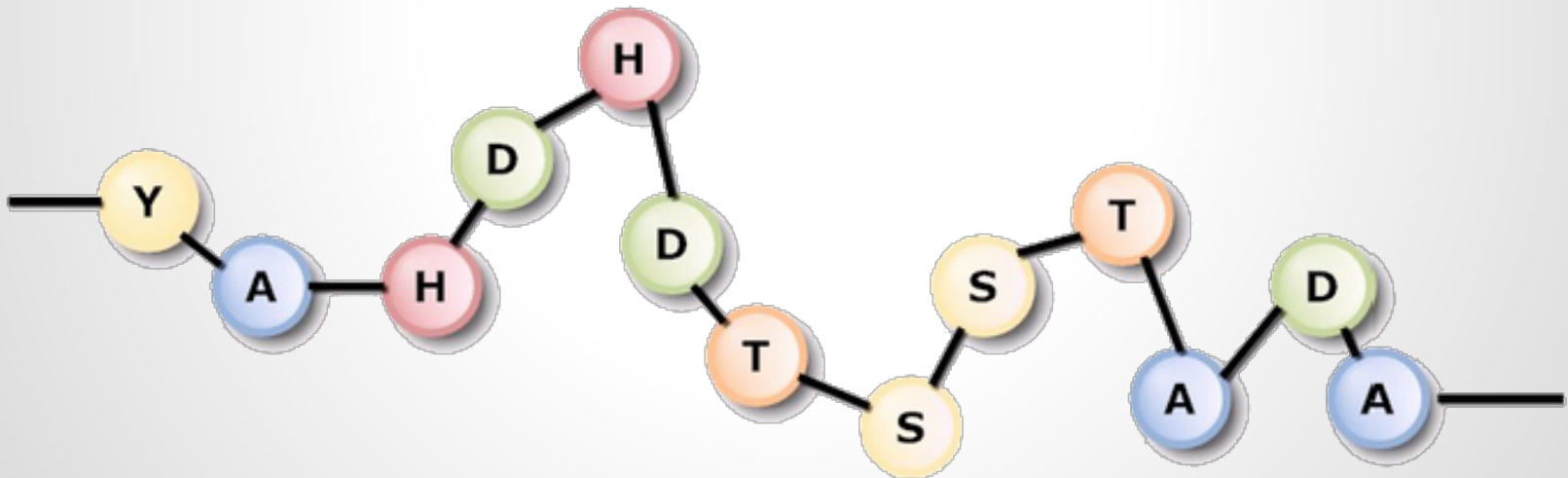
TP3 - Mutações

Hugo Richard Amaral

Luís Eduardo Oliveira Lizardo

Problema

Identificar quais as mutações em uma proteína causam alteração de sua função.

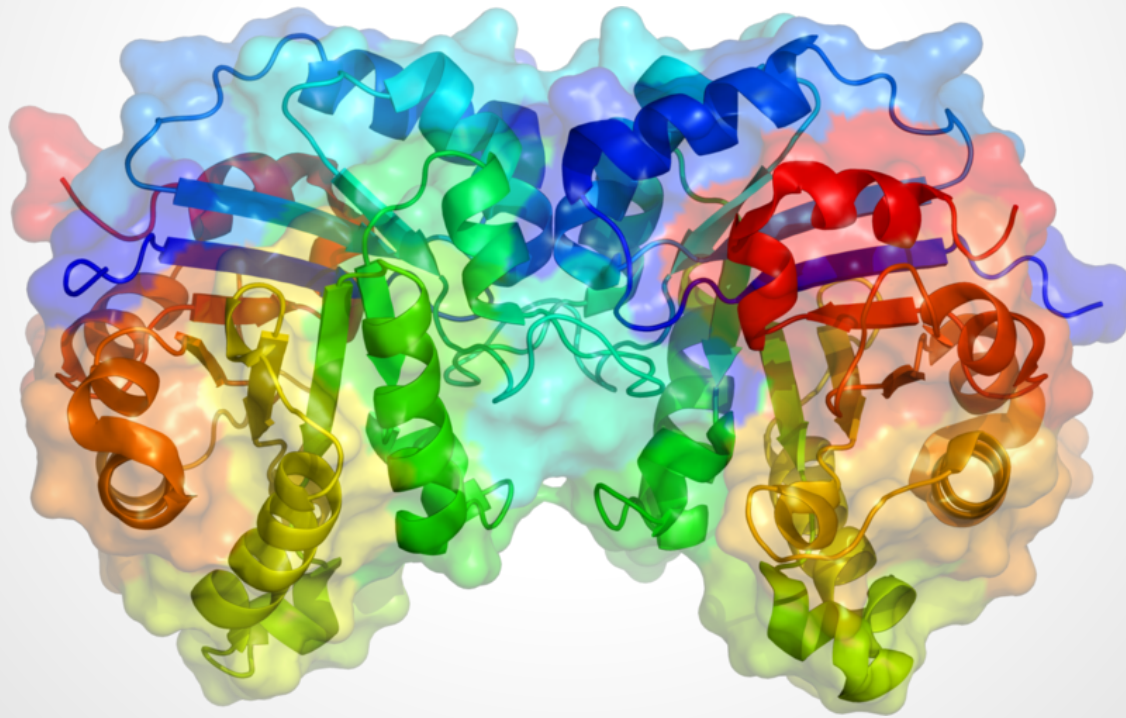


Objetivo

- Utilizar um algoritmo de alinhamento para identificar as mutações;
- Modelar todo o conjunto de proteínas como grafos de contatos intramoleculares;
- Analisar as mutações identificadas de acordo com o potencial impacto que podem causar na função da 2YPI, utilizando o alinhamento de sequências em conjunto com a análise dos grafos.
- Fornecer como resultado as 10 mutações consideradas mais severas, com todos os valores das métricas calculadas para tais mutações e todos os dados que justifiquem o resultado.

Triose-fosfato isomerase - 2YPIA

248 aminoácidos



Fonte: Wikimedia Commons

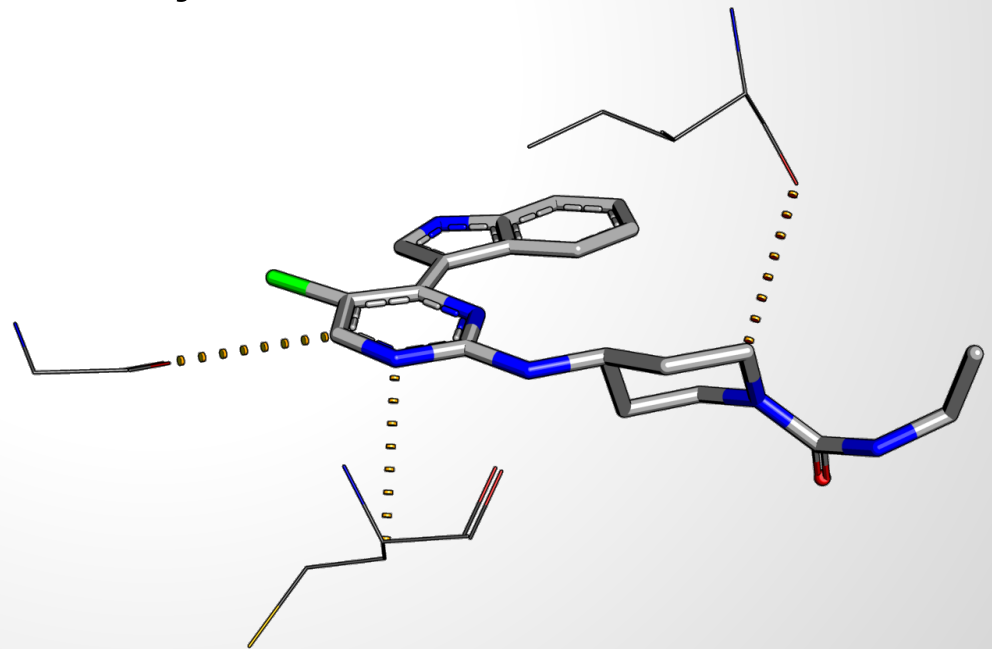
Alinhamento

- Algoritmo de Needleman-Wunsch:
 - Alinhamento global de duas sequências;
 - Mede a similaridade entre as sequências com base em uma tabela de pontuação;
- Quanto maior o *score* de alinhamento, mais similares são as sequências.

Fonte: Needleman, Saul B.; and Wunsch, Christian D. (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins"

Contatos

- Determinar as interações que ocorrem entre os resíduos da proteína;
- Envolve determinar um conjunto de vizinhos de cada um dos resíduos

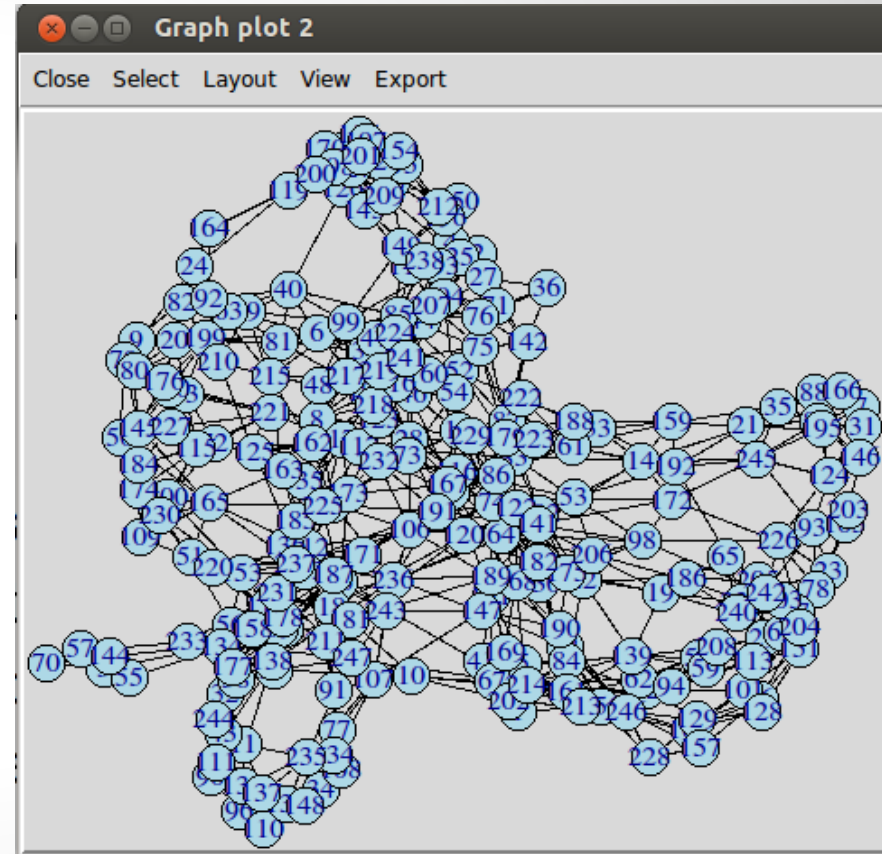


Contatos

- Utilizamos o limiar de distância;
- O limiar define quando um contato é realizado;
- Por simplicidade, adotamos o limiar de distância euclidiana com valores em 5Å, 7Å e 9Å.

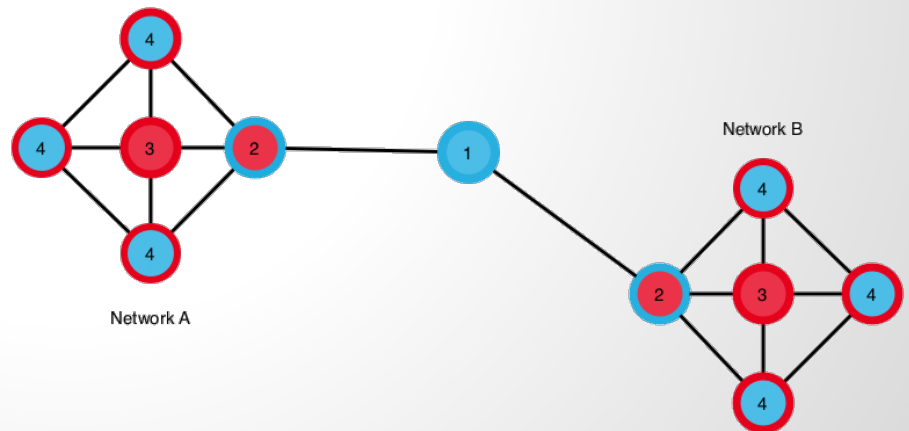
Contatos como Grafos

- De posse do conjunto de resíduos e das interações entre os mesmos, é necessária uma modelagem que auxilie na caracterização do problema;
- Modelagem natural: Grafos



Métricas de Centralidade

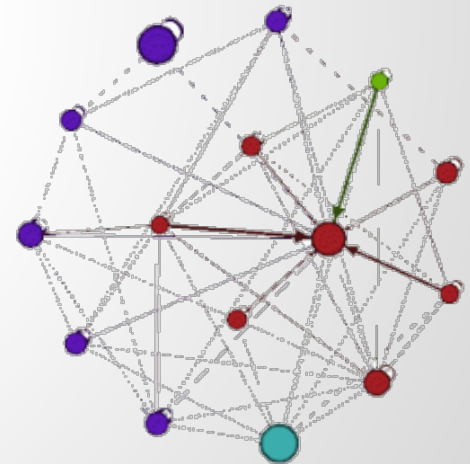
- **Betweenness** - Número de menores caminhos de todos os vértices para quaisquer outros vértices que passam por um nó.
- No contexto de interações, um *betweenness* alto é esperado nos casos de resíduos que atuam como pontes na estrutura da proteína.



Li, Yizhou, et al. "Predicting disease-associated substitution of a single amino acid by analyzing residue interactions." BMC bioinformatics 12.1 (2011): 14.

Métricas de Centralidade

- **Closeness** - (Centralidade) Distância geodésica média entre um vértice e todos os outros vértices do grafo.
- **Grau** - Número de outros vértices que fazem conexão com o vértice em referência.
- Mutações estão associadas frequentemente a doenças quando acontecem num ponto com baixo valor de *closeness* e/ou alto grau na rede



Li, Yizhou, et al. "Predicting disease-associated substitution of a single amino acid by analyzing residue interactions." BMC bioinformatics 12.1 (2011): 14.

Metodologia - Alinhamento 2YPI e dTIM

- **Objetivo:** Identificar todas as mutações que ocorreram na dTIM;
- Ferramenta: Biopython (<http://biopython.org/>)
- Matriz de pontuação: PAM60
- Gap: -10.0, Gap extension: -0.5

Metodologia - Alinhamento 2YPI e dTIM

- **Objetivo:** Identificar aminoácidos com características semelhantes e agrupá-los.
- A troca desses aminoácidos causam pouco impacto na função da proteína.

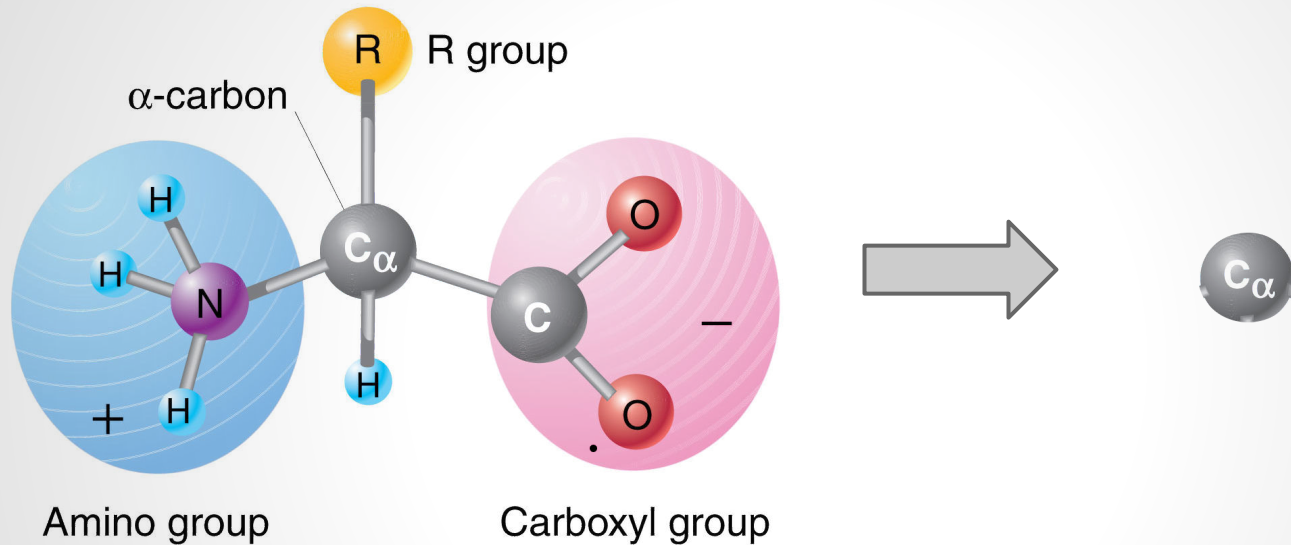


Fonte: Vermont ViewER MutatiON Tool

Metodologia - Alinhamento: Família

- Alinhamos a 2YPI com todas as 133 proteínas da família;
- **Objetivo:** Identificar se as mutações que ocorrem na dTIM também ocorrem nas famílias;
- Filtramos apenas as mutações que não ocorreram nas famílias;

Metodologia - Cálculo das distâncias



© 2010 Pearson Education, Inc.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Metodologia - Contatos

- Modelagem dos contatos como grafos: geração dos arquivos .net

```
*Vertices 247
```

```
1 "GLY118"  
2 "SER50"  
3 "PHE74"  
4 "SER52"  
5 "LEU117"  
6 "SER235"  
7 "ALA136"  
8 "VAL7"  
9 "VAL241"  
10 "LEU68"  
11 "ARG189"  
12 "ASN159"  
13 "GLY214"  
14 "PRO166"  
15 "GLY62"  
16 "VAL80"  
17 "VAL86"  
18 "ASN78"  
19 "GLY94"  
20 "GLU239"  
21 "THR130"  
22 "LEU93"  
23 "GLU144"  
24 "ILE32"  
25 "TYR49"  
26 "ARG26"
```

```
*Edges
```

```
211 183 3.722  
183 125 3.835  
125 163 3.800  
163 162 3.819  
162 8 3.834  
162 152 4.653  
8 218 3.835  
8 217 4.485  
8 232 4.648  
218 219 3.835  
218 28 4.337  
219 241 3.775  
219 180 4.815  
219 6 4.575  
241 85 3.810  
241 52 4.743  
241 39 4.503  
85 207 3.827  
207 118 3.903  
118 238 3.772  
238 143 3.718  
143 200 3.735  
200 209 3.811  
200 170 4.737  
200 198 4.663  
209 170 3.750
```

Metodologia - Métricas de redes

- Utilização do pacote *igraph* Python;
- Geração dos seguintes dados:
 - Grau;
 - Betweenness;
 - Closeness
- Calculado para cada uma das mutações encontradas no alinhamento

Resultados - Alinhamento

- Total: 50 mutações

```
2YPIA      MARTFFVGGN FKLNGSKQSI KEIVERLNTA SIPENVEVVI CPPATYLDYS
dTIM       MARTPFVGGN WKMNGTKAEA KELVEALK-A KLPDDVEVVV APPAVYLDTA

          VSLVKKPQVT VGAQNAYLKA SGAFTGENSV DQIKDVGAKW VILGHSERRS
          REALKGSKIK VAAQNCYKEA KGFTGEISP EMLKDLGADY VILGHSERRH

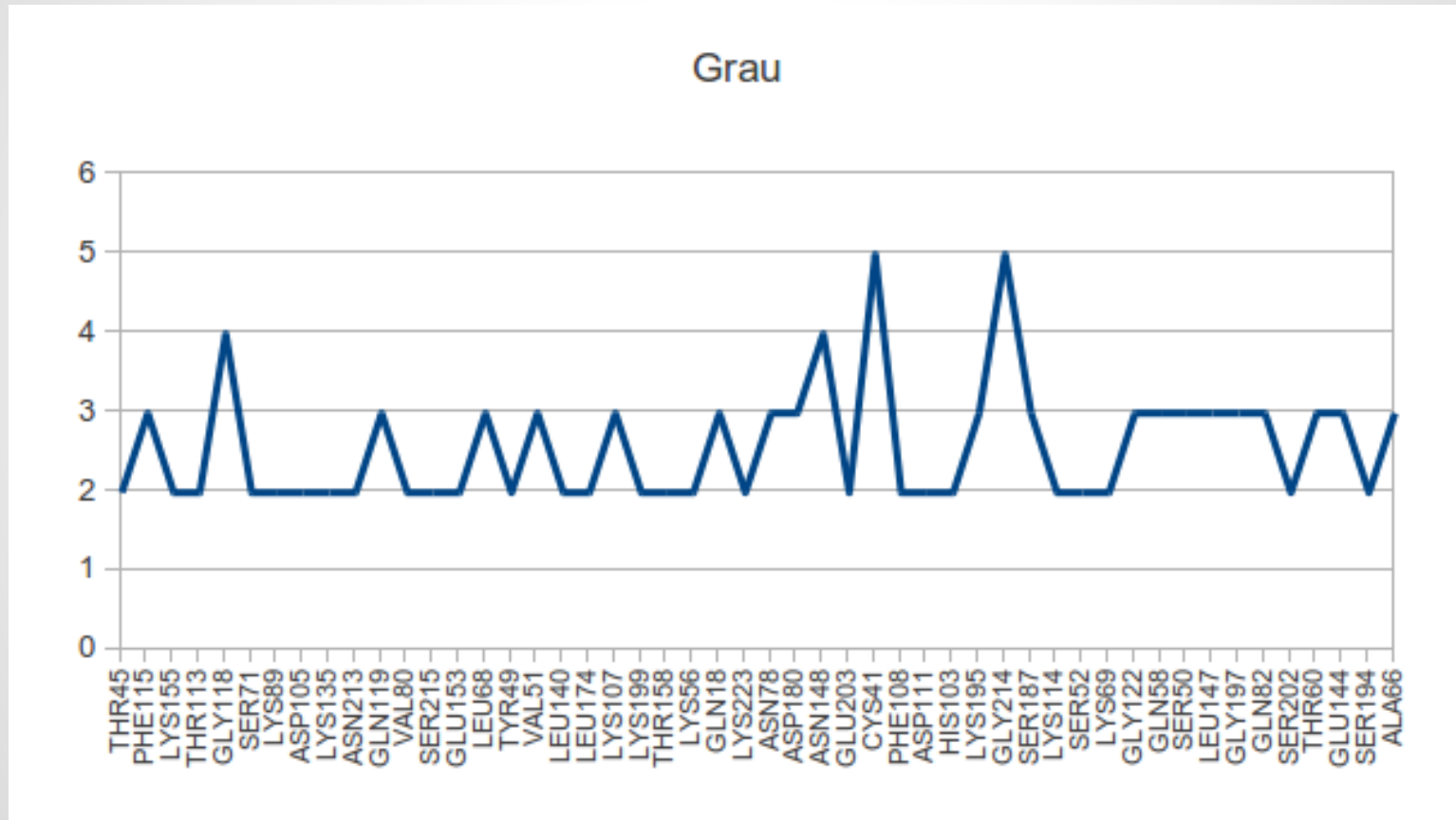
          YFHEDDKFIA DKTKFALGQG VGVILCIGET LEEKKAGKTL DVVERQLNAV
          YFGETDELVA KKVAAHALEHG LKVIACIGET LEEREAGKTE EVVFRQTKAL

          LEEVKD-WTN VVVAYEPVWA IGTGLAATPE DAQDIHASIR KFLASKLGDK
          LAGLGDEWKN VVIAYEPVWA IGTGKTATPE QAQEVHAFIR KWLAENVSAE

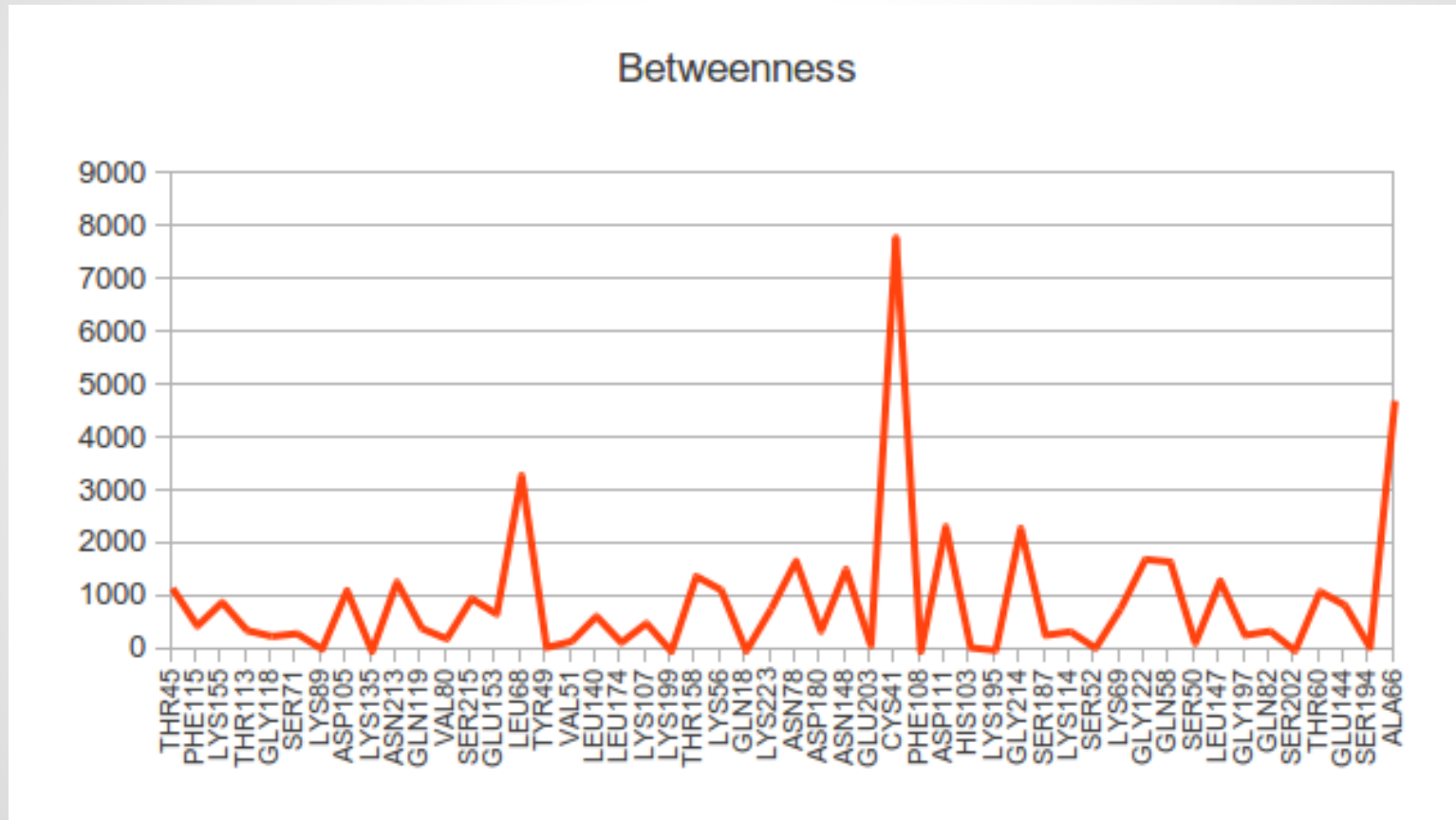
          AASELRILYG GSANGSNAV FDKA-- -DV DGFLVGGASL KPEFVDIINS
          VAESVRILYG GSVKPANA-- -KELAAQPD I DGFLVGGASL KPEFLDIINS

          RN
          RN
```

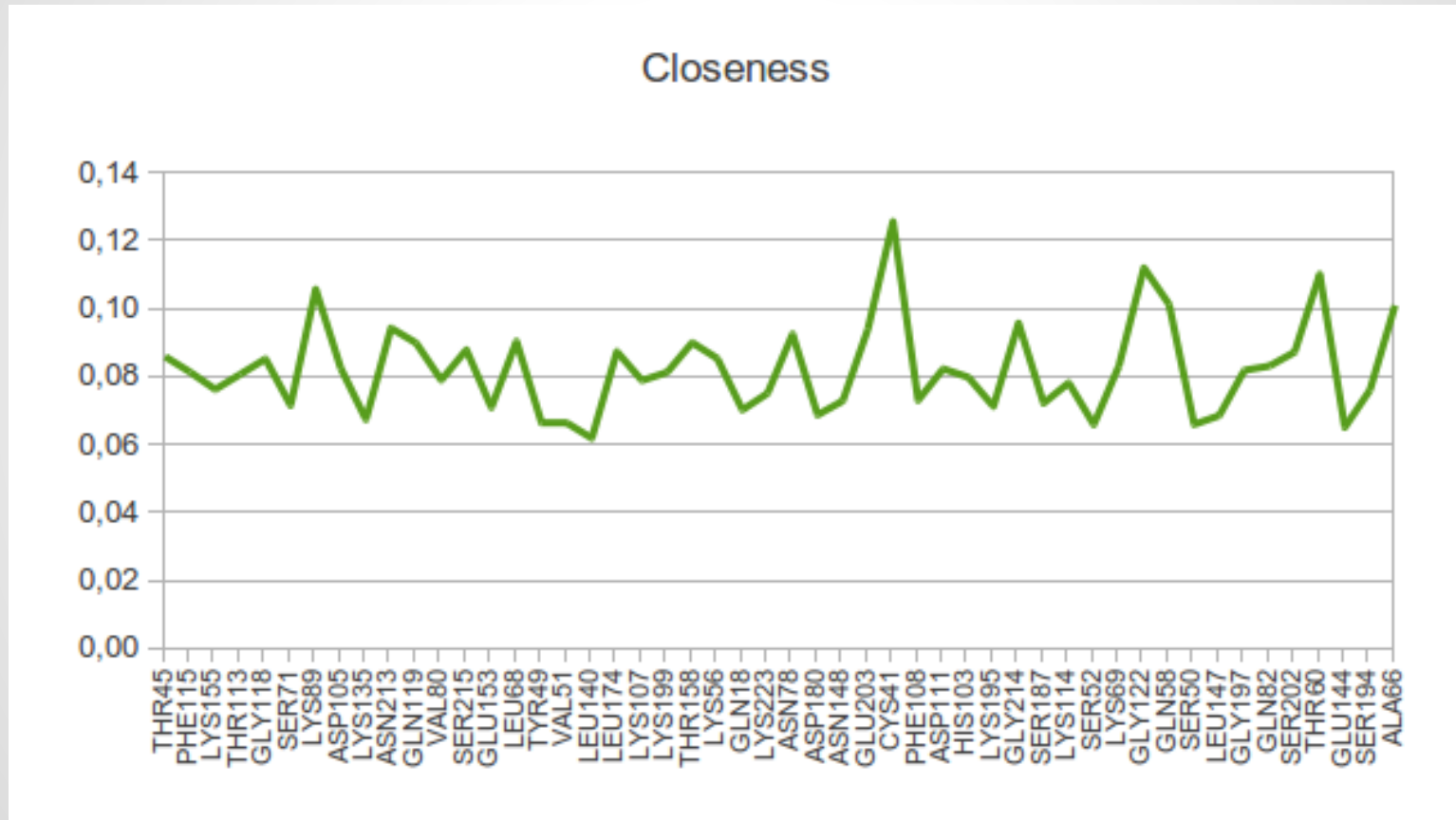
Resultados - Cutoff 5



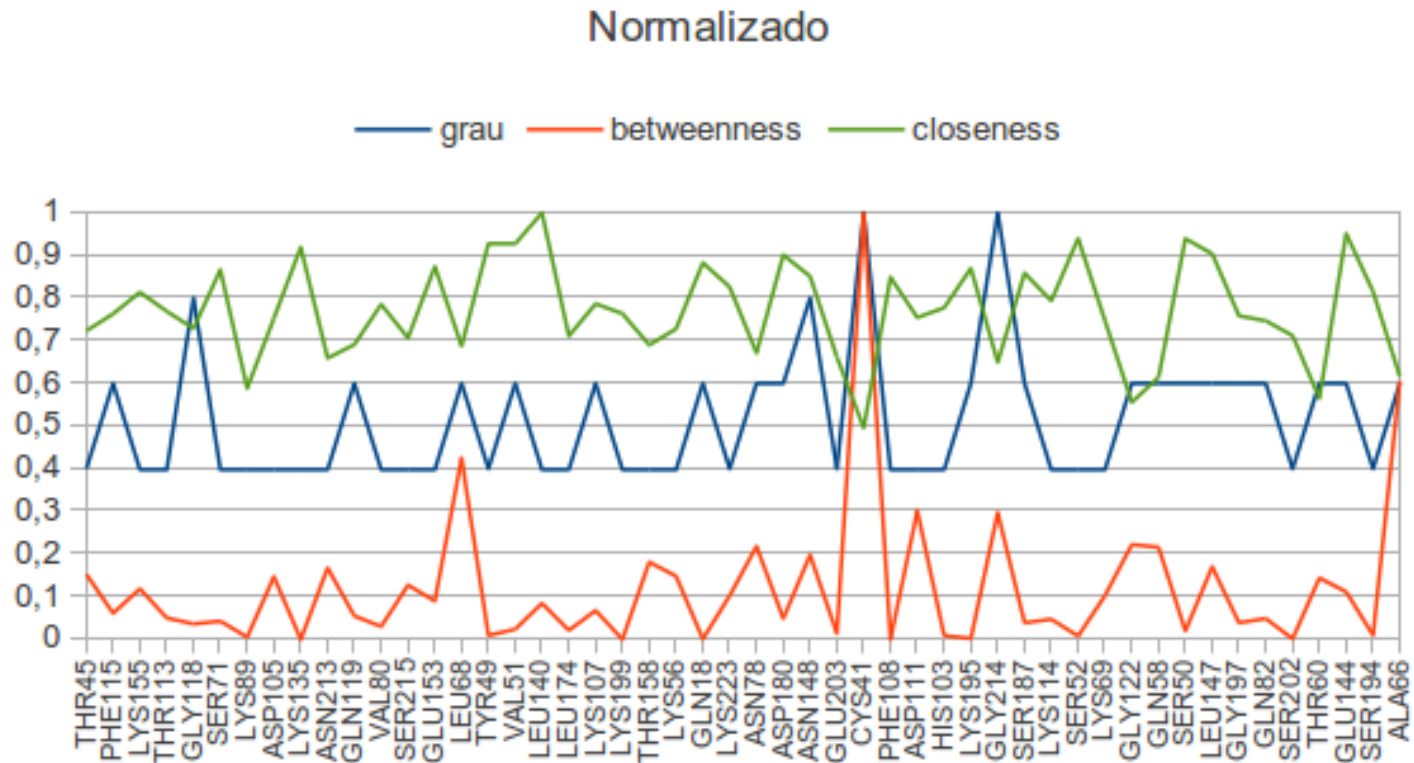
Resultados - Cutoff 5



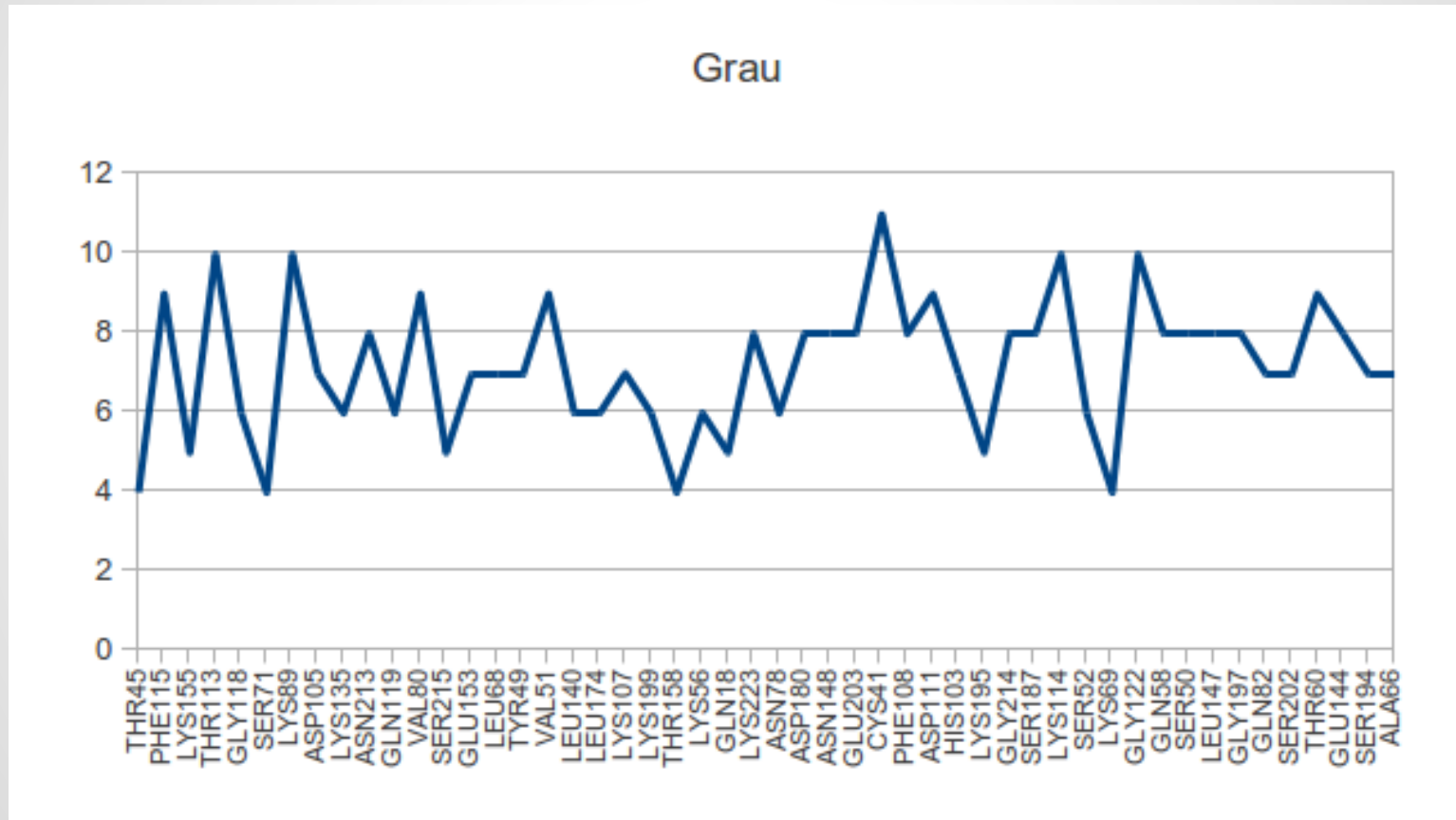
Resultados - Cutoff 5



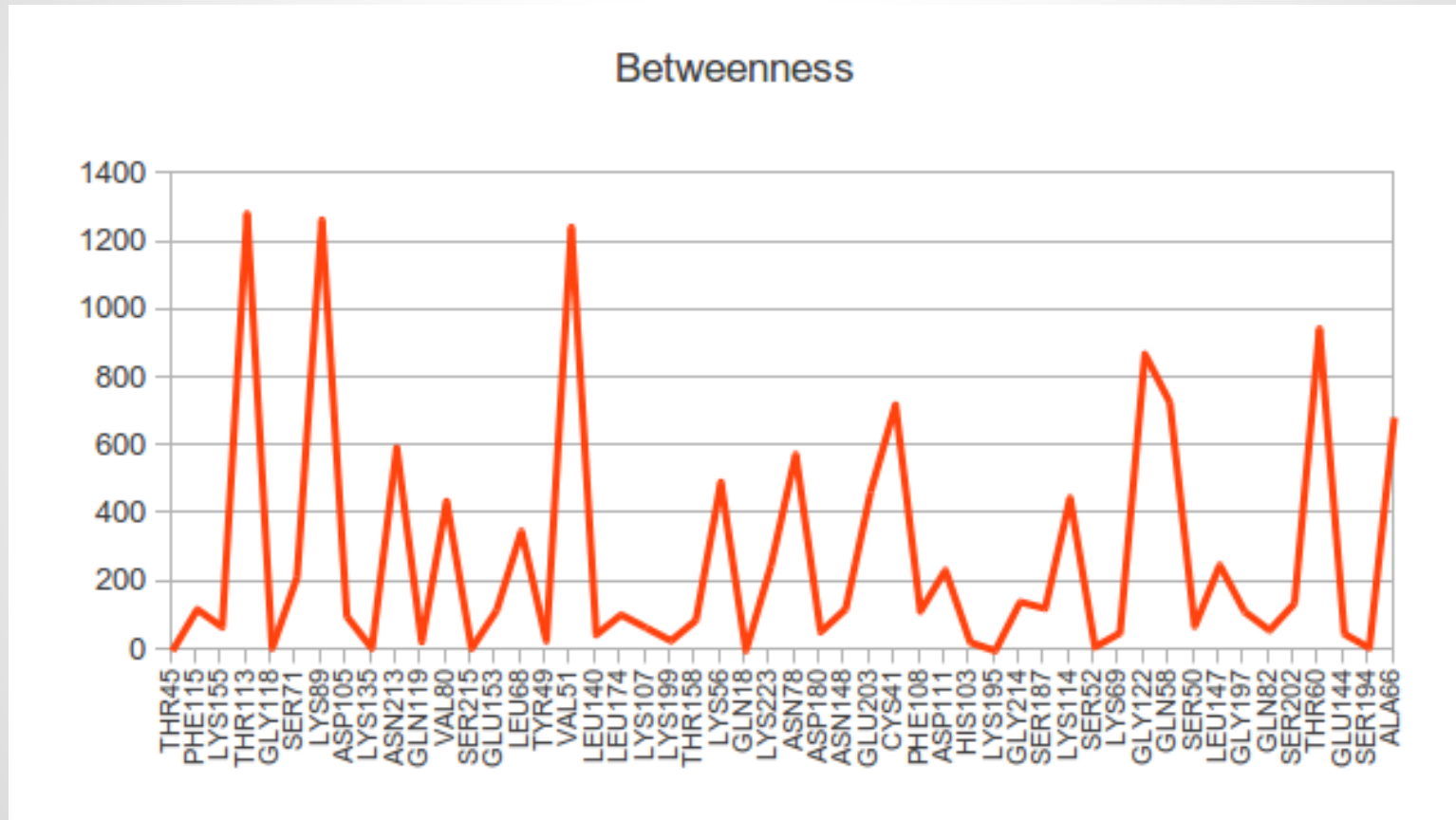
Resultados - Cutoff 5



Resultados - Cutoff 7



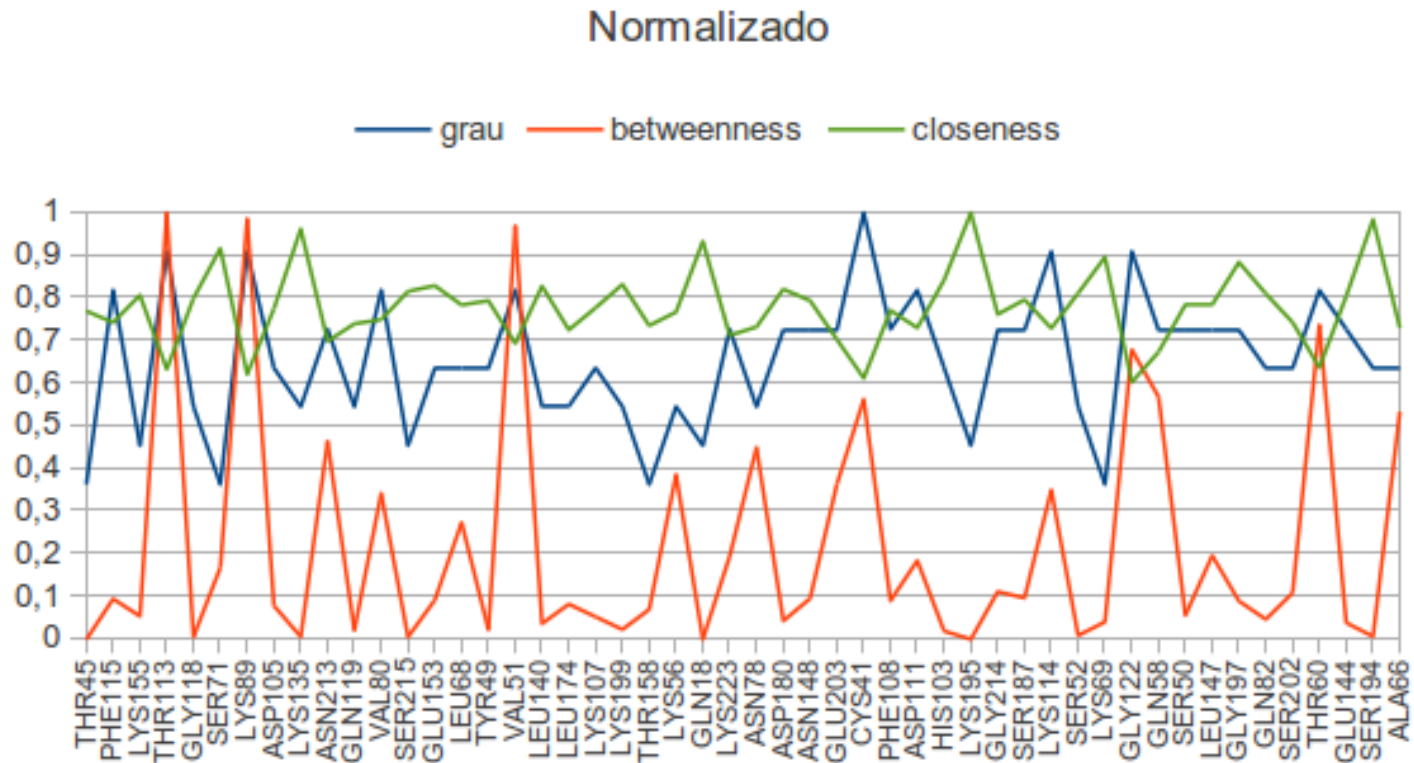
Resultados - Cutoff 7



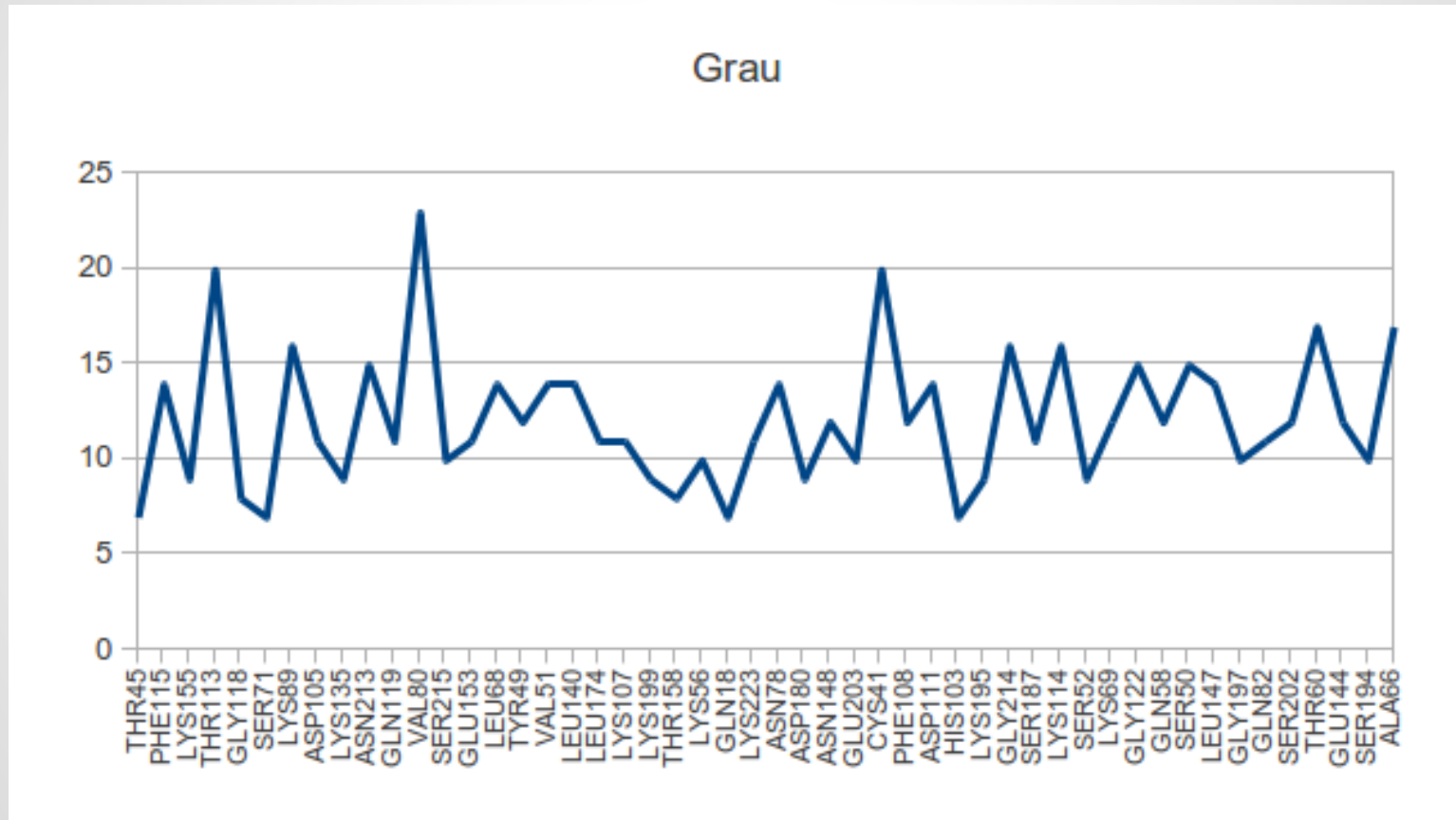
Resultados - Cutoff 7



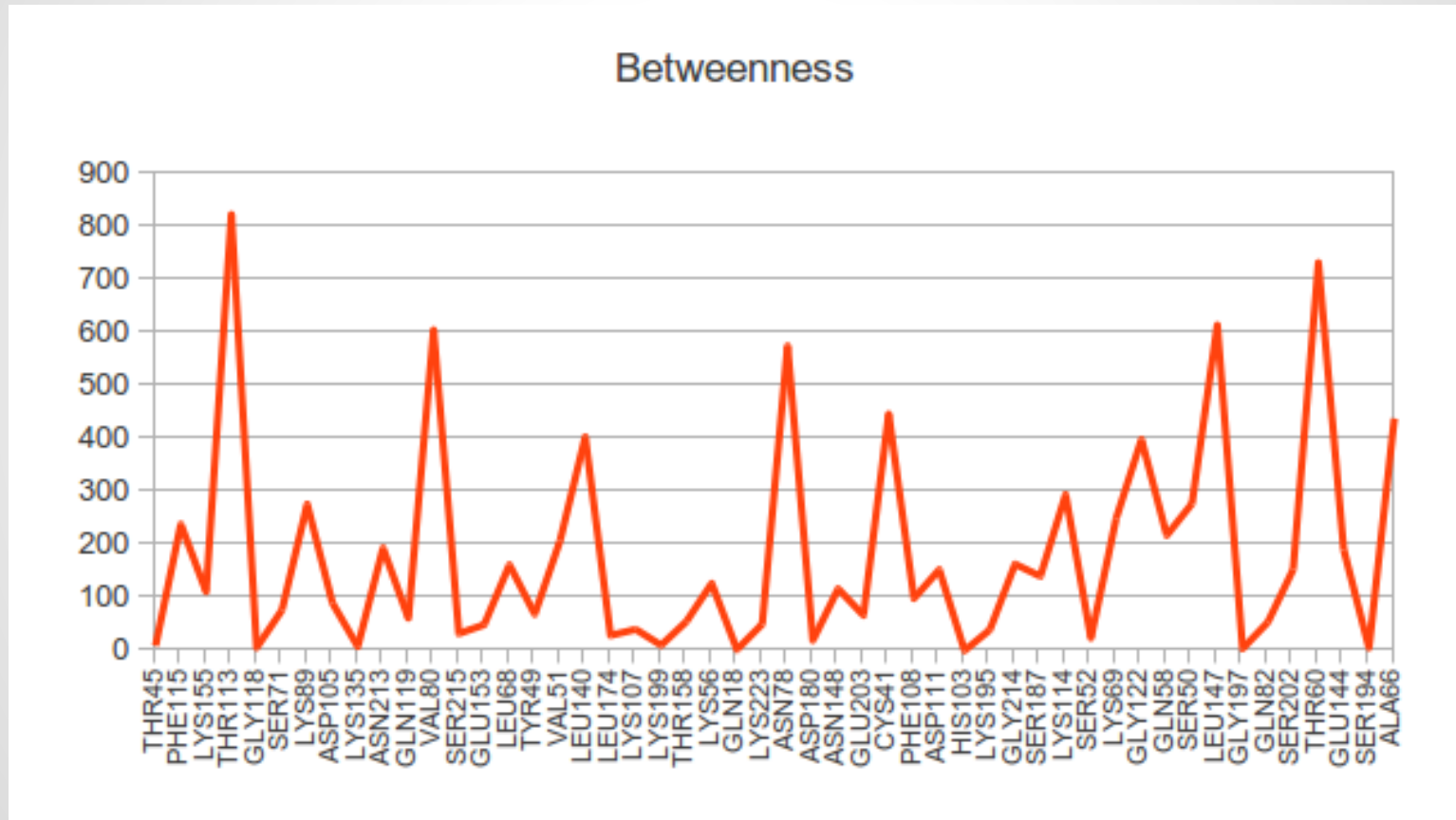
Resultados - Cutoff 7



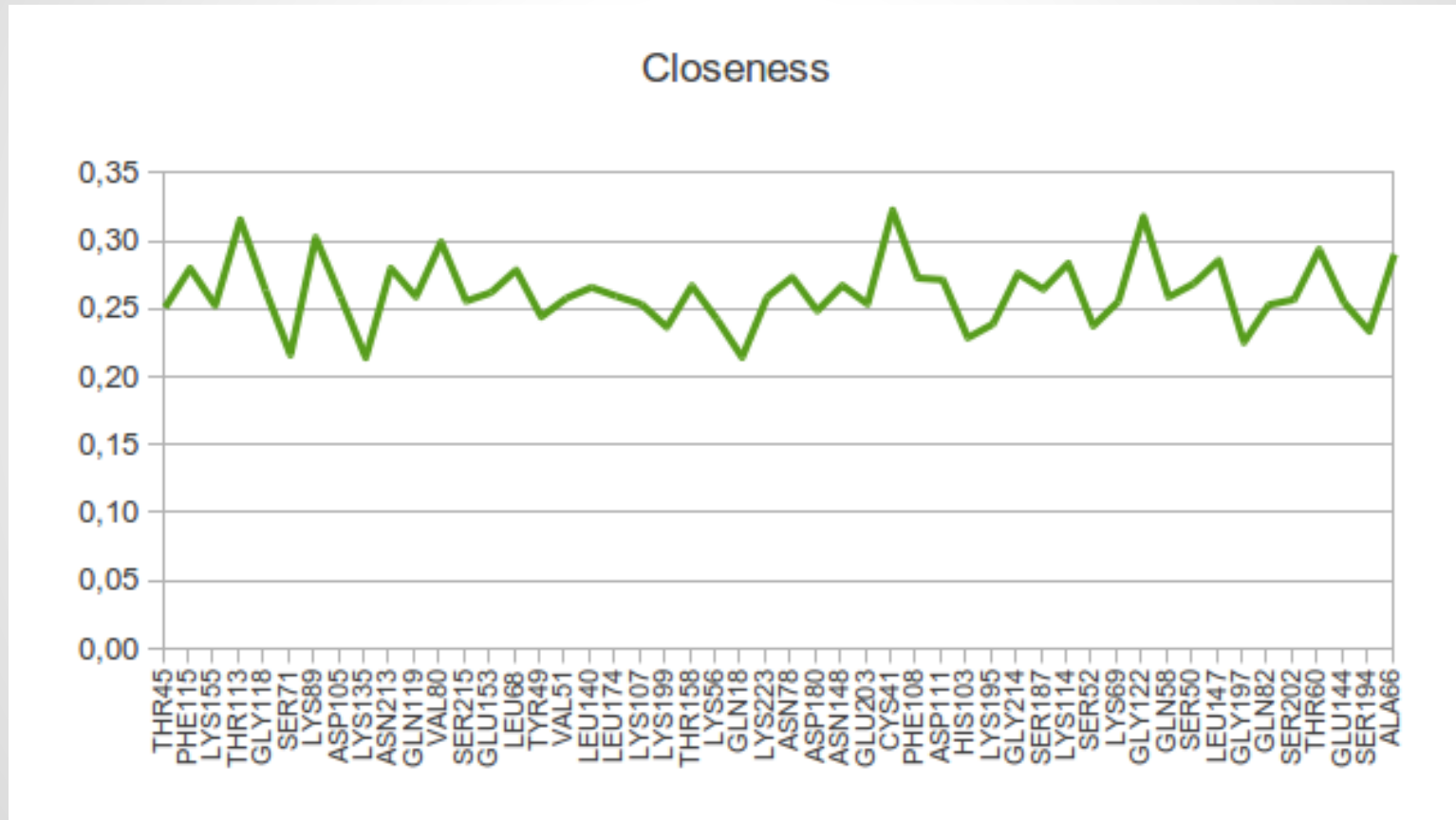
Resultados - Cutoff 9



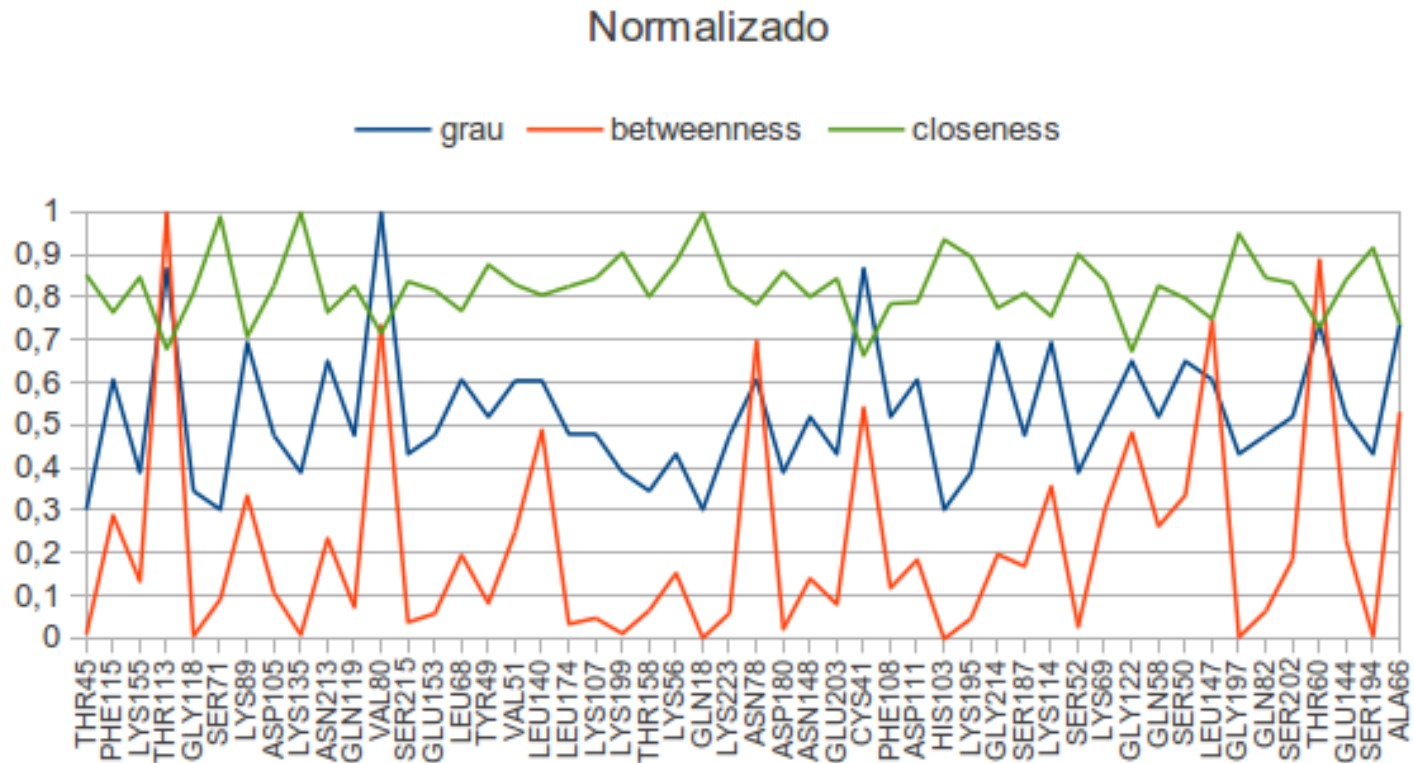
Resultados - Cutoff 9



Resultados - Cutoff 9



Resultados - Cutoff 9



Resultados - Cutoff 5

	Resíduo	Soma
1º	C41A	2,496
2º	G214P	1,947
3º	N148K	1,849
4º	A66C	1,822
5º	L68K	1,713
6º	L147T	1,674
7º	E144F	1,663
8º	G118E	1,565
9º	S50A	1,559
10º	V51R	1,553

Resíduo	Soma Ponderada (3 2 1)
C41A	0,916
G214P	0,707
N148K	0,608
A66C	0,605
L68K	0,556
G118E	0,534
L147T	0,507
E144F	0,496
N78I	0,485
Q58K	0,474

Resíduo	Soma Ponderada (2 1 1)
C41A	0,874
G214P	0,737
N148K	0,662
A66C	0,605
G118E	0,591
L68K	0,578
L147T	0,569
E144F	0,566
S50A	0,540
V51R	0,538

Resultados - Cutoff 7

	Resíduo	Soma
1º	T113V	2,542
2º	K89D	2,516
3º	V51R	2,481
4º	T60K	2,192
5º	G122K	2,191
6º	C41A	2,175
7º	K114A	1,988
8º	Q58K	1,970
9º	V80P	1,910
10º	A66C	1,901

Resíduo	Soma Ponderada (3 2 1)
T113V	0,893
K89D	0,887
V51R	0,848
C41A	0,790
G122K	0,781
T60K	0,761
K114A	0,693
Q58K	0,665
V80P	0,648
N213K	0,635

Resíduo	Soma Ponderada (2 1 1)
T113V	0,863
K89D	0,856
V51R	0,825
C41A	0,794
G122K	0,775
T60K	0,753
K114A	0,724
V80P	0,682
Q58K	0,674
N213K	0,654

Resultados - Cutoff 9

	Resíduo	Soma
1º	T113V	2,550
2º	V80P	2,453
3º	T60K	2,359
4º	L147T	2,106
5º	N78I	2,093
6º	C41A	2,078
7º	A66C	2,012
8º	L140E	1,906
9º	G122K	1,812
10º	K114A	1,811

Resíduo	Soma Ponderada (3 2 1)
T113V	0,882
V80P	0,865
T60K	0,788
C41A	0,727
L147T	0,678
A66C	0,670
N78I	0,668
L140E	0,602
G122K	0,600
K114A	0,593

Resíduo	Soma Ponderada (2 1 1)
V80P	0,863
T113V	0,855
T60K	0,775
C41A	0,737
A66C	0,688
L147T	0,679
N78I	0,675
L140E	0,629
K114A	0,627
G122K	0,616

Conclusão

- 10 mutações mais severas

C41A
G214P
N148K
T113V
K89D
V51E
V80P
T60K
A66C
L147T



Perguntas



Obrigado