

Trabalho Prático 3:

Mutações

Hugo Richard Amaral, Luís Eduardo O. Lizardo

Universidade Federal de Minas Gerais

{hucharal, lizardo}@dcc.ufmg.br

Resumo: Neste trabalho estamos interessados em identificar mutações severas que possam ter alterado a função de uma enzima triose-fosfato isomerase. Para isso alinhamos a enzima mutada com a selvagem e também a selvagem com a família para identificar as mutações. Para identificar entre as mutações encontradas àquelas mais severas, calculamos os contatos intramoleculares existentes e aplicamos algumas métricas de redes complexas para analisar padrões e anomalias.

1. INTRODUÇÃO

Proteínas são compostos orgânicos bioquímicos que desempenham funções específicas nos organismos. Elas são formadas por sequências de resíduos de aminoácidos. Cada aminoácido possui características distintas e tem seu papel na função proteica. A troca de um aminoácido por outro, decorrente de uma mutação, pode provocar alterações no funcionamento da proteína.

O objetivo deste trabalho é identificar possíveis mutações que alteraram a função proteica de uma enzima triose-fosfato isomerase (TIM). Essa identificação é feita, primeiramente, alinhando a enzima que sofreu a mutação (dTIM) com a enzima selvagem (2YPI¹) para descobrir todas as mutações. Posteriormente são filtradas, dentre as mutações, as mais importantes e possíveis causadoras de danos a função proteica, com base na conservação do resíduo na família, propriedades físico-químicas e topológicas. O cálculo de contatos e o uso de grafos e algoritmos de redes complexas são usados para ajudar a analisar padrões ou anomalias encontradas.

¹ <http://www.pdb.org/pdb/explore/explore.do?structureId=2ypi>

Com base nos experimentos realizados, identificamos 10 possíveis mutações severas que podem alterar a funcionalidade da enzima triose-fosfato isomerase.

2. CONCEITOS BÁSICOS

2.1. Alinhamento de sequências

O alinhamento de sequências é uma técnica da bioinformática para comparar sequências de DNA, RNA ou proteína para identificar regiões similares. A similaridade pode indicar a ocorrência de relações funcionais, estruturais ou evolucionárias entre as sequências. O alinhamento pode apresentar a ocorrência de **matches**, quando há casamento exato entre dois nucleotídeos (no caso de DNA e RNA) ou resíduos de aminoácidos (no caso de proteínas), de **mismatches**, interpretados como mutações pontuais, e **gaps**, como inserções ou deleções introduzidas em uma ou ambas as sequências.

2.2. Contatos

Neste trabalho, um contato é definido entre qualquer par de resíduo $\{i, j\}$ se a distância Euclidiana entre eles é menor ou igual a um limiar. Este limiar representa então uma distância limite e são considerados contatos apenas as interações que estiverem em um raio menor ou igual a esse limiar. A distância dos contatos é calculada utilizando as coordenadas atômicas dos ligantes e dos Carbonos Alfa dos aminoácidos das proteínas.

2.3. Medidas de centralidade

Ao estudar mutações, é importante não só analisar a mutação isoladamente, como também os efeitos da interação da posição mutada com outros resíduos. Para ajudar a definir a importância da posição dos resíduos mutantes no contexto das interações, podemos utilizar algumas métricas de redes complexas [1]. São elas:

Betweenness: É igual ao número de menores caminhos de todos os vértices para quaisquer outros vértices que passam por um nó. Para calcular o *betweenness* é necessário computar os menores caminhos entre todos os pares de vértices do grafo. A ordem de complexidade desse cálculo é $\Theta(|V|^3)$, onde V é o número de arestas no grafo.

No contexto de interações, um *betweenness* alto é esperado nos casos de resíduos que atuam como pontes na estrutura da proteína [2], como exemplo, aqueles resíduos que reúnem duas estruturas secundárias diferentes.

Closeness: É a distância geodésica média entre um vértice e todos os outros vértices do grafo. Assim, quanto mais central for um nó, menor é a sua distância para todos os outros vértices (porque ele está próximo de todo mundo). Na biologia, mutações estão associadas frequentemente a doenças quando acontecem num ponto com alto valor de centralidade (closeness) e/ou alto grau na rede [3].

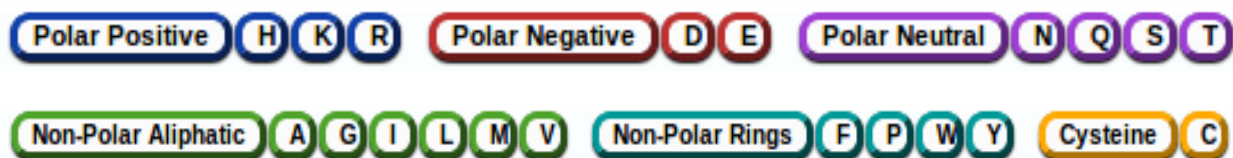
Grau: O grau de um vértice é simplesmente o número de outros vértices que este vértice em referência faz conexões.

3. METODOLOGIA

3.1. Alinhamento

A fim de determinar as mutações ocorridas na selvagem e presentes na dTIM, o primeiro passo foi realizar o alinhamento da sequência de resíduos. Para execução dessa etapa, foi utilizada a ferramenta Biopython (<http://biopython.org/>), com a matriz de pontuação PAM60 e desconto para o gap de -10 e para o gap estendido de -0,5.

Além disso, para caracterizar as mutações conservativas, utilizamos a categorização dos aminoácidos feita pela ferramenta VERMONT, conforme podemos ver na Figura.



Assim, para as trocas de resíduos envolvendo o mesmo grupo, consideramo-as conservativas. Com isso, levando em consideração que mudanças conservativas não são drásticas, todas as mutações que se encaixem nesse critério são descartadas para qualquer análise posterior.

3.1.1. 2YPI e dTIM

O alinhamento entre a proteína selvagem e a mutante serviu como base para identificarmos todas as diferenças existentes entre dois resíduos em uma mesma posição. Assim, essa etapa definiu todas as possíveis mutações ocorridas que serviram como base para as próximas etapas de análise.

3.1.2. 2YPI e Família

Esse alinhamento envolveu a proteína selvagem e todas as de sua família. O alinhamento foi realizado par-a-par entre a 2YPI e cada uma das 133 proteínas da família, com o intuito de identificarmos todas as mutações que ocorreram tanto na mutada como também em alguma de sua família. Essas mutações foram consideradas não severas e, portanto, descartadas de futuras análises.

3.2. Cálculo das distâncias

Para o cálculo das distâncias, conforme discutido anteriormente, consideramos a distância euclidiana entre os resíduos e três valores limiares 5Å, 7Å e 9Å. Inicialmente, deveríamos estabelecer qual a posição do espaço do resíduo. Para isso, determinamos como a posição do carbono alfa como aquela que localiza o resíduo no espaço. Com isso, calculamos as distâncias entre todos os pares de resíduos da proteína 2YPI.

3.3. Contatos

A partir do cálculo das distâncias, consideramos aquelas que estavam a valores menores que 5Å, 7Å e 9Å como os contatos entre os aminoácidos. Com essas relações de pares de resíduos, geramos os arquivos com a representação do grafo em formato .net. A Figura abaixo ilustra um dessas representações:

*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
```

3.4. Métricas de Redes

A última etapa envolveu calcular as métricas de redes sobre os grafos representados na etapa anterior. Foram considerados apenas os resíduos determinados na seção 3.1.2 e para cada um deles calculamos o valor do grau, betweenness e closeness para o grafo que representa os contatos estabelecidos na proteína 2YPI. O intuito dessa etapa foi determinar um valor que represente como são as relações desses resíduos e qual a influência que eles possuem sobre os vizinhos.

4. RESULTADOS

Nesta seção serão apresentados os resultados obtidos que nos ajudaram a escolher as 10 mutações mais propícias a alterarem a função da proteína. Na subseção 4.1 é explicado como foi feito o alinhamento e seus resultados. Nas subseções 4.2, 4.3 e 4.4 são apresentados gráficos com os valores de grau, closeness e betweenness para cada mutação encontrada no alinhamento. Os resultados são apresentados separados pelo valor do Cutoff.

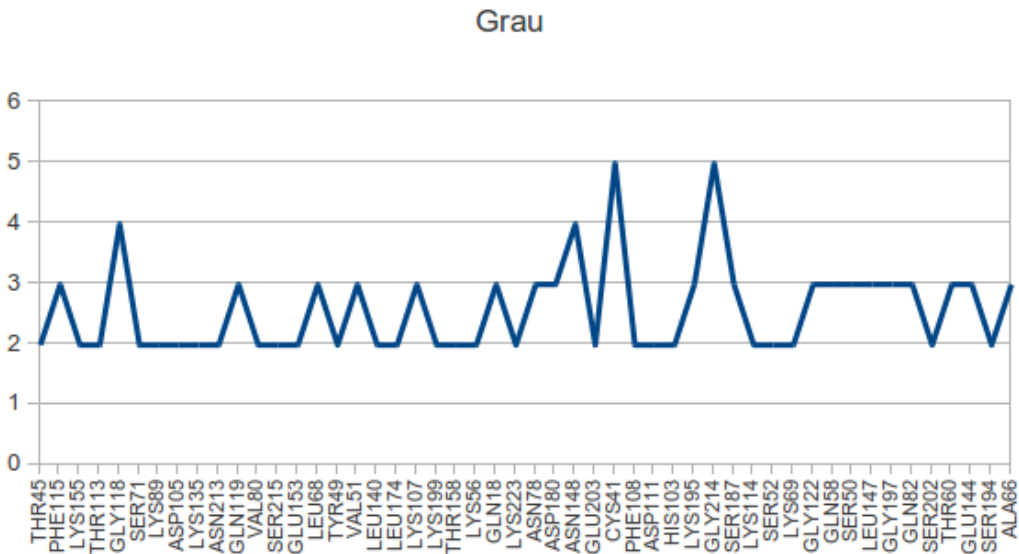
4.1. Alinhamento

O alinhamento entre a proteína selvagem e mutada juntamente com o alinhamento entre a selvagem e todas da sua família, possibilitou chegarmos a um resultado de 50 mutações ocorridas na 2YPI e presentes na dTIM. A Figura abaixo ilustra o alinhamento entre essas duas

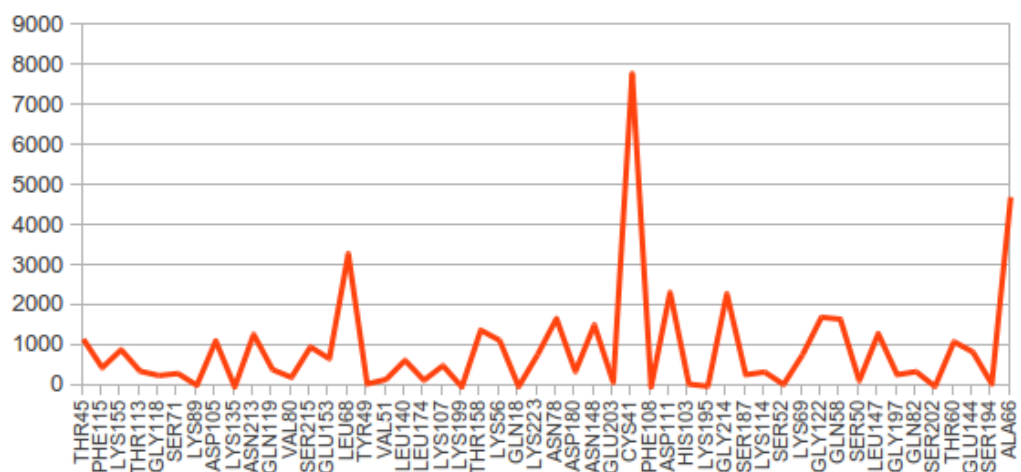
proteínas.

2YPIA	MARTFFVGGN	FKLNGSKQSI	KEIVERLNTA	SIPENVEVVI	CPPATYLDYS
dtIM	MARTPFVGGN	WKMNGTKAEA	KELVEALK-A	KLPDDVEVVV	APPAVYLDTA
	VSLVKKPQVT	VGAQNAYLKA	SGAFTGENSV	DQIKDVGAKW	VILGHSERRS
	REALKGSKIK	VAAQNCYKEA	KGFTTGEISP	EMLKDLGADY	VILGHSERRH
	YFHEDDKFIA	DKTKFALGQG	VGVLICIGET	LEEKKAGKTL	DVVERQLNAV
	YFGETDELVA	KKVAHALEHG	LKVIACIGET	LEEREAGKTE	EVVFRQTKAL
	LEEVKD-WTN	VVVAYEPVWA	IGTGLAATPE	DAQDIHASIR	KFLASKLGDK
	LAGLGDEWKN	VVIAYEPVWA	IGTGKTATPE	QAQEVHAFIR	KWLAENVSAA
	AASELRILYG	GSANGSNAV	FKDKA---DV	DGFLVGGASL	KPEFVDIINS
	VAESVRILYG	GSVKPANA--	-KELAAQPDI	DGFLVGGASL	KPEFLDIINS
	RN				
	RN				

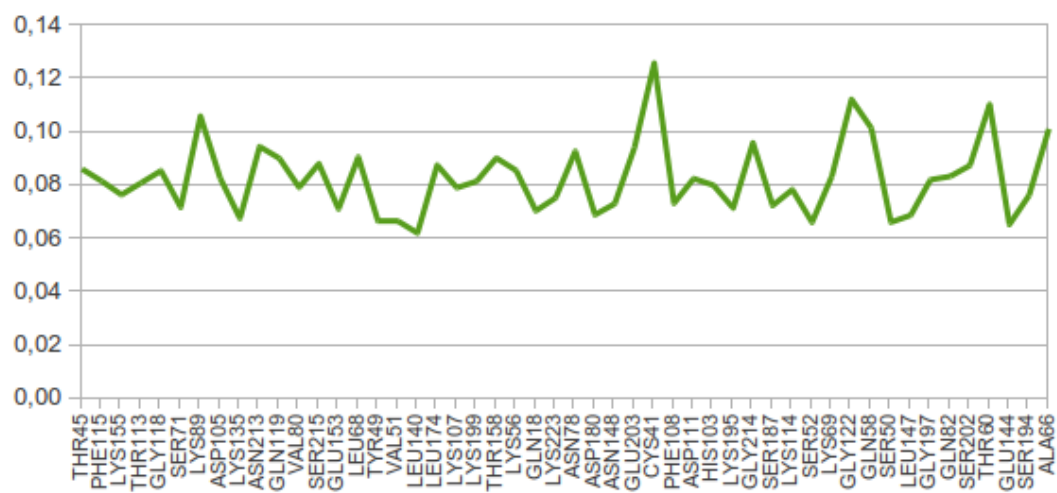
4.2. Cutoff 5Å



Betweenness

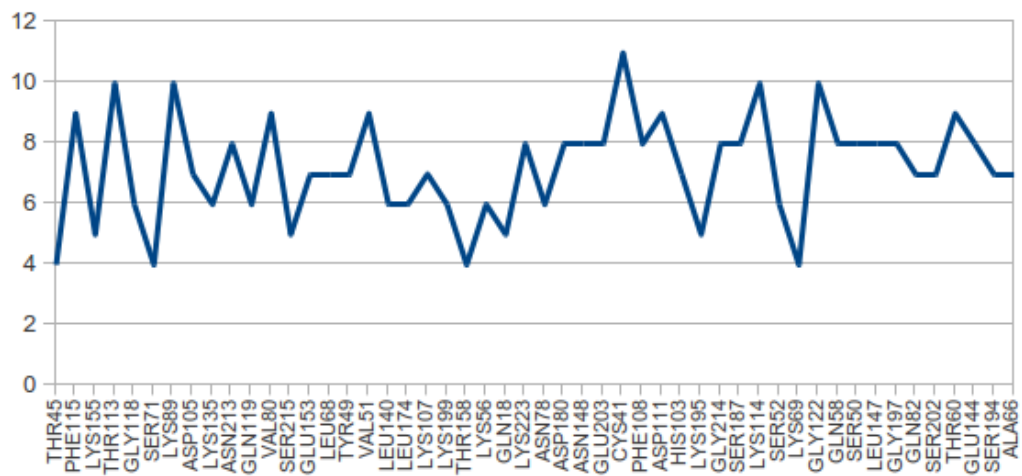


Closeness

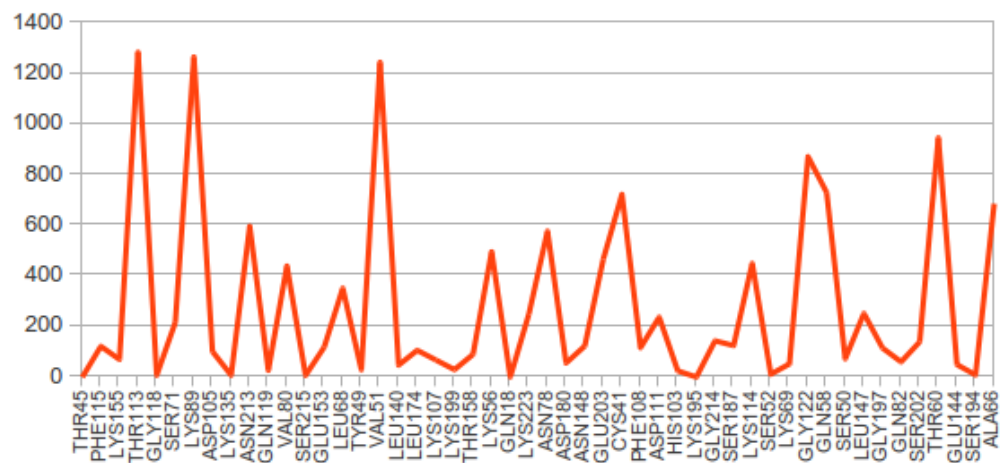


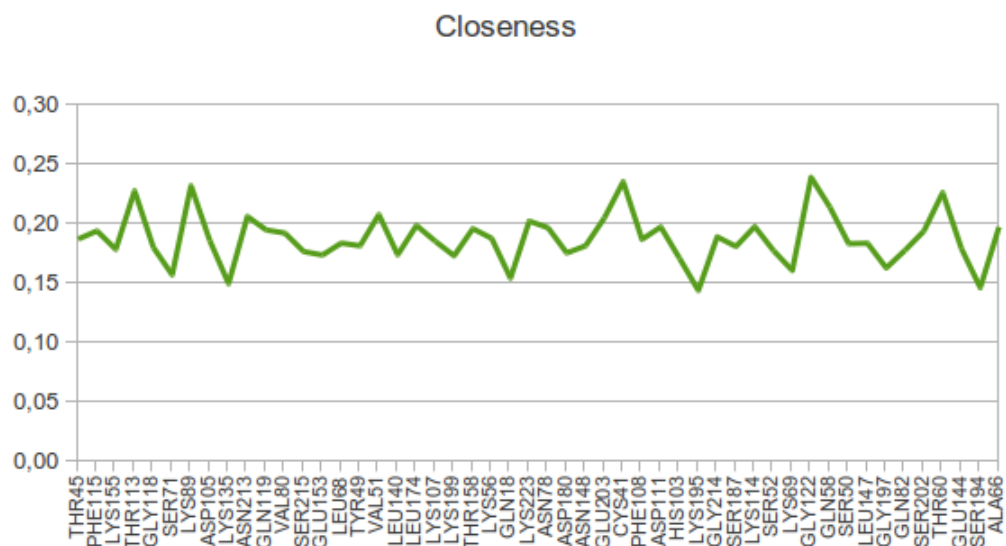
4.3. Cutoff 7Å

Grau

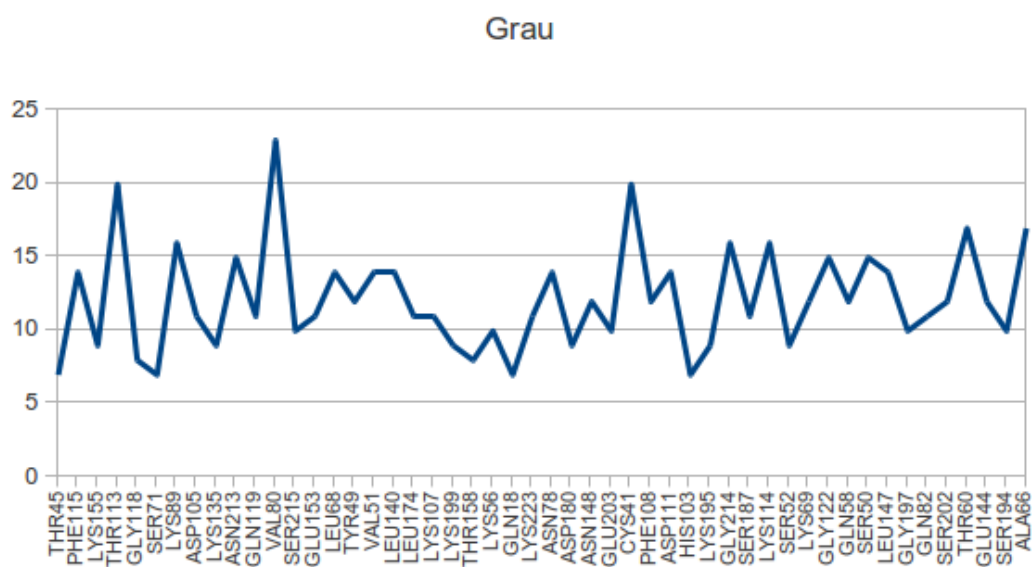


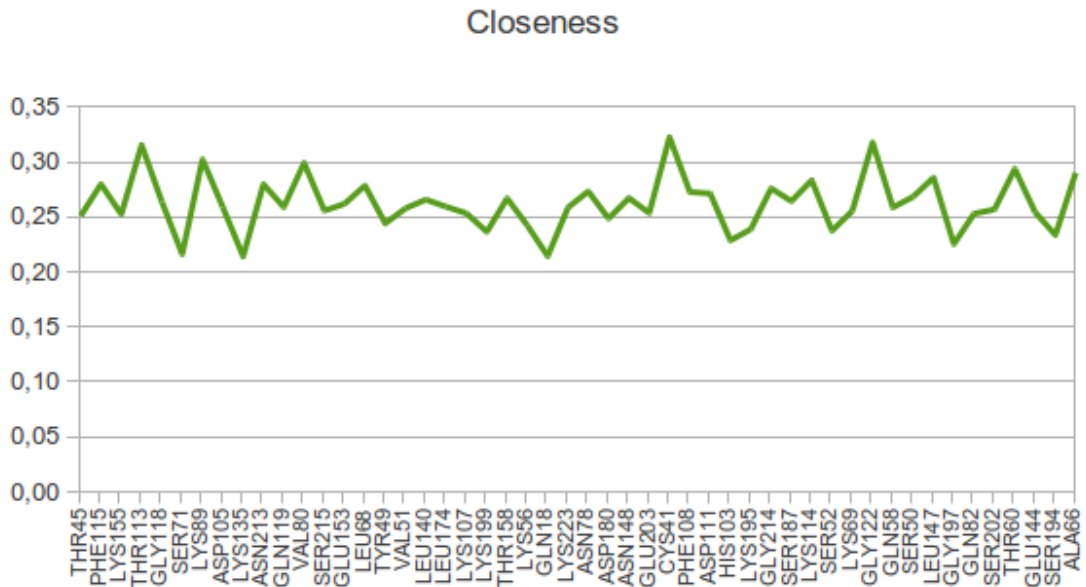
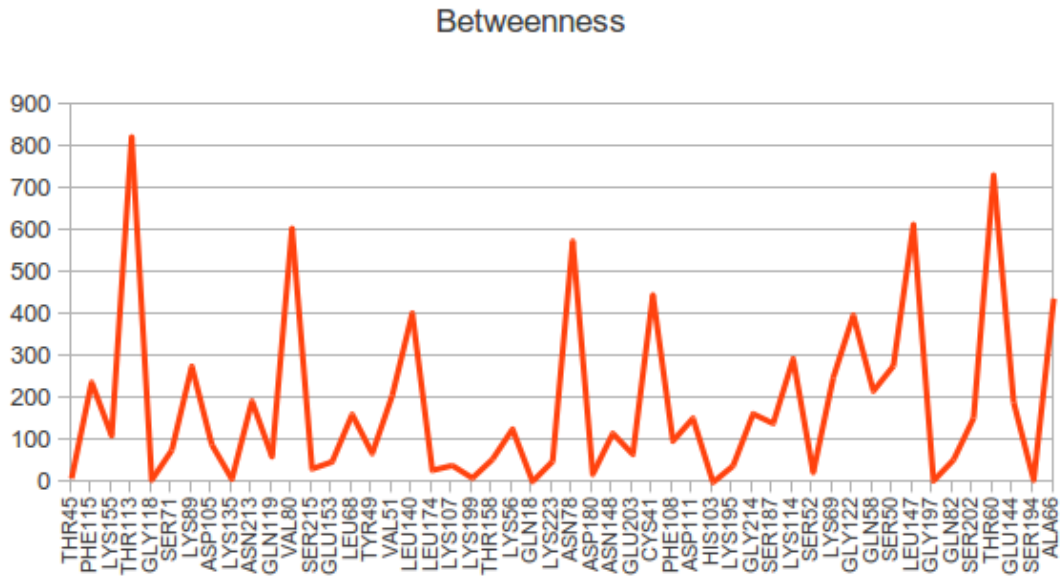
Betweenness





4.4. Cutoff 9Å





4.5. Mutações Severas

Com os valores de grau, closeness e betweenness obtidos, fizemos três análises diferentes. Na primeira análise somamos os três valores (grau, closeness e betweenness) com pesos iguais. Na segunda análise utilizamos os pesos: 0.50, 0.33 e 0.17 respectivamente para grau, betweenness e closeness. Na terceira análise os pesos foram: 0.50, 0.25 e 0.25. Como as métricas de redes (grau, closeness e betweenness) estão em domínios diferentes, tivemos que

normalizar todos os valores entre 0 e 1. Para isso, consideramos que, para as métricas grau e betweenness, o maior valor encontrado representasse 1 e para a métrica closeness, o menor valor. Para os demais valores, calculamos o valor proporcional dele sobre o “melhor”, ou seja, o resultado nos deu um valor entre 0 e 1. Com isso, foi possível realizarmos uma análise comparativa envolvendo as três métricas simultaneamente.

Nas três subseções seguintes são apresentados os resultados obtidos das análises para os Cutoffs de 5Å, 7Å e 9Å. Os resultados mostram as 10 mutações que obteram os maiores valores de cada análise, ordenadas do maior para o menor valor.

4.5.1. *Cutoff* 5Å

	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

4.5.2. *Cutoff* 7Å

	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

4.5.3. Cutoff 9Å

	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

5. CONCLUSÃO

Como conclusão do trabalho, apresentamos as 10 mutações que consideramos mais severas. Elas foram retiradas dos resultados obtidos para todas as três métricas. Consideramos aquelas que obtiveram um maior valor de cada um dos critérios apresentados nas seções anteriores.

C41A
G214P
N148K
T113V

K89D
V51E
V80P
T60K
A66C
L147T

REFERÊNCIAS

[1] Strogatz, Steven H. "Exploring complex networks." *Nature* 410.6825 (2001): 268–276.

[2] del Sol, Antonio, and Paul O'Meara. "Small-world network approach to identify key residues in protein–protein interaction." *Proteins: Structure, Function, and Bioinformatics* 58.3 (2005): 672–682.

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