

Aula 20 – Utilizando grafos para estudar proteínas

Introdução

Nesse tutorial veremos um exemplo de como a teoria e algoritmos de grafos podem ser utilizados no estudo das proteínas. Utilizaremos uma proteína (que é uma enzima) já bem conhecida pelos alunos de Introdução à Bioinformática, a 2YPI. No primeiro TP dessa disciplina, estudamos a proteína selvagem (2YPI) e sua respectiva mutante (dTIM) através do alinhamento de sequências. Naquele trabalho os alunos apontaram, dentre as mutações que a enzima havia sofrido, aquelas que eram potenciais causadoras de perda de função.

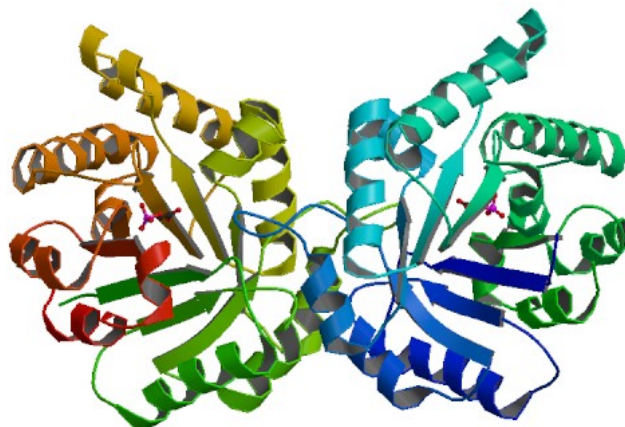


Figura 1: Estrutura tridimensional da 2YPI.
CRYSTALLOGRAPHIC ANALYSIS OF THE COMPLEX BETWEEN
TRIOSEPHOSPHATE ISOMERASE AND 2-PHOSPHOGLYCOLATE AT
2.5-ANGSTROMS RESOLUTION. IMPLICATIONS FOR CATALYSIS

A enzima triose fosfato isomerase é importante no processo de glicólise (quebra de açúcar, a glucose) para produção de energia. Mais especificamente, é responsável pela quinta etapa da glicólise e é considerada uma enzima perfeita pois a reação que catalisa acontece da

ordem de bilhões de vezes mais rápido do que aconteceria sem a enzima.

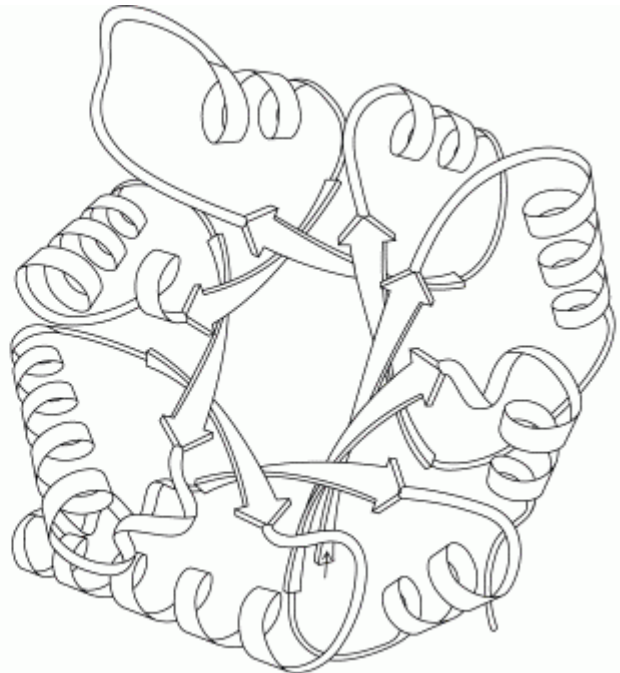
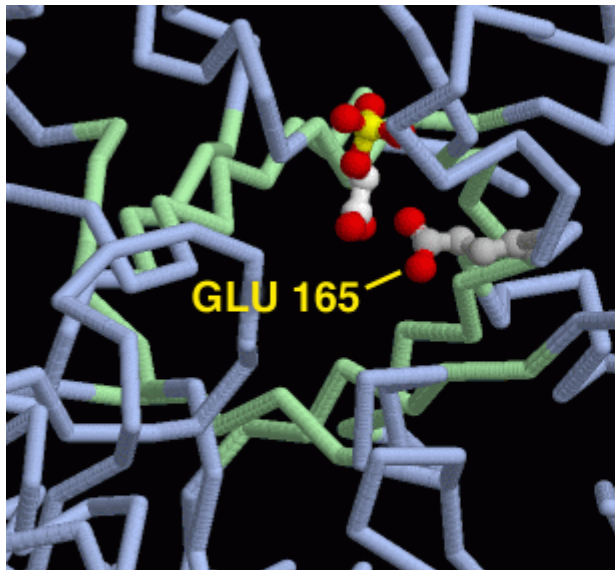


Figura 2: Glutamato 165 destacado (esse aminoácido é parte do sítio ativo) na figura da esquerda. Na direita um esquema do fold beta barril da triose fosfato isomerase.

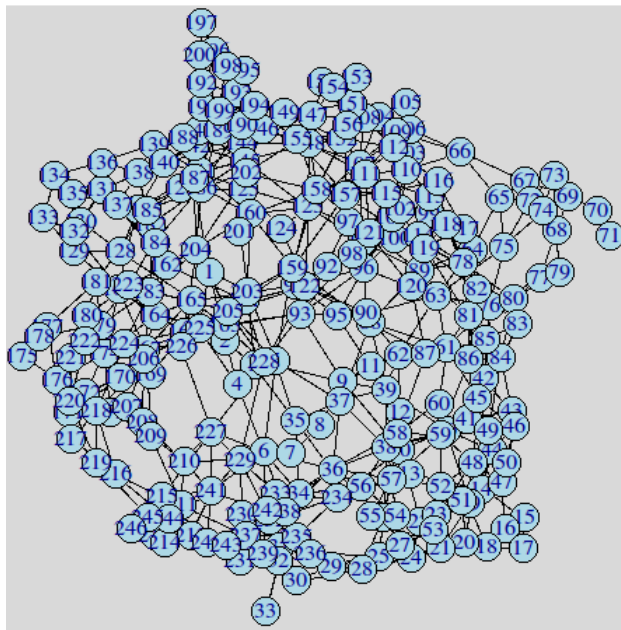
Note que o sítio ativo está num beta barril, esse arranjo cilíndrico (destacado em verde). O esquema mostra o envovelamento da cadeia, com um anel interno de folhas beta envolvido num anel externo de alfa hélices. As hélices aparentemente conectam duas folhas vizinhas no barril. Esse *fold* (beta barril) também está presente em outras enzimas glicolíticas.

Na aula de hoje utilizaremos o software R (<http://www.r-project.org/>) e seu pacote *igraph* (<http://igraph.sourceforge.net/>) para analisar o gráfico de contatos da enzima 2YPI cadeia A.

Tutorial

Acesse a página do Minha UFMG (<https://sistemas.ufmg.br/idp/login.jsp>) para acessar os arquivos que utilizaremos na aula de hoje. Começaremos com o arquivo 2YPIA.G.net, que é a cadeia A da proteína 2YPI modelada como grafo utilizando critério geométrico.

- Lendo o grafo a partir do arquivo:
 > library(igraph)
 > grafo = read.graph("2YPIA.G.net", format="pajek")
- Plotando o grafo:
 > tkplot(grafo, vertex.size = 10, vertex.color = "lightblue", edge.color = "black",
 vertex.label.font = 1)



- Número de nós e arestas:
 > vcount(grafo)

```
[1] 246
```

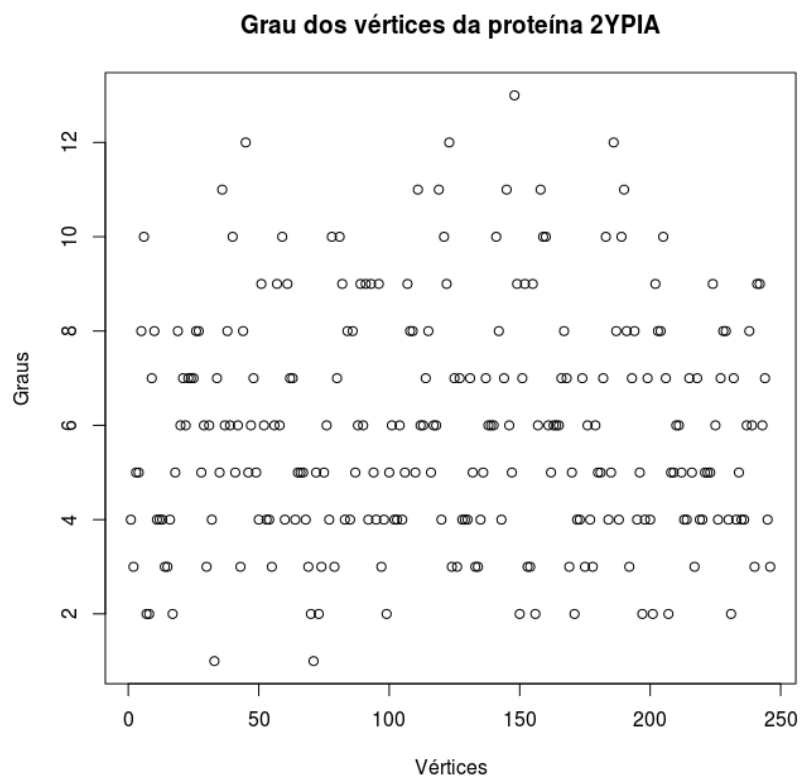
```
> ecount(grafo)
```

```
[1] 723
```

- Grau dos nós:

```
> grau = degree(grafo)
```

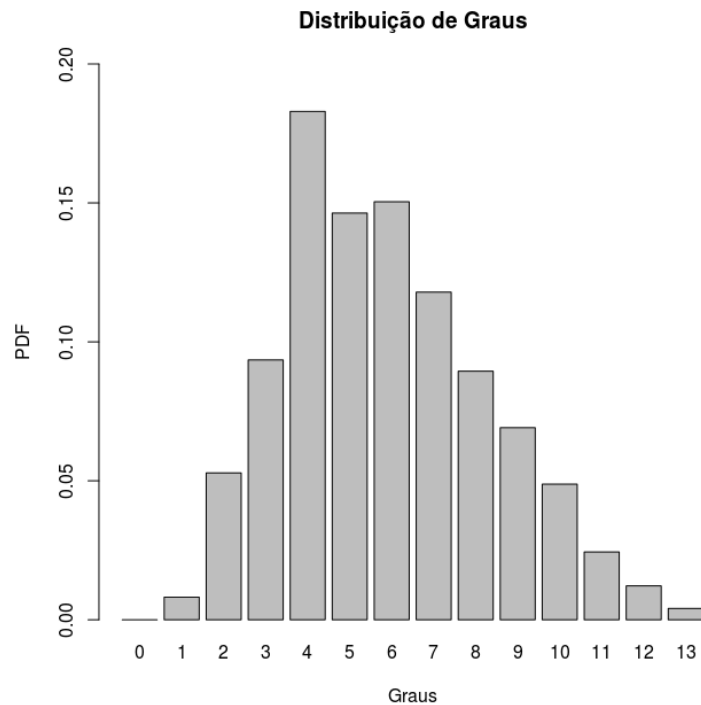
```
> plot(c(1:246), grau, xlab="Vértices", ylab="Graus", main="Grau dos vértices da  
proteína 2YPIA");
```



- Distribuição de graus dos nós:

```
> distGrau = degree.distribution(grafo, cumulative=FALSE)
```

```
> barplot(distGrau, main="Distribuição de Graus", xlab="Graus", ylab="PDF",  
names.arg=c(0:(length(distGrau)-1)), ylim = c(0, 0.2));
```



- Betweenness dos nós:
 - > `b = betweenness(grafo)`
 - > `bmax=max(b)`
 - > `bindex_max = which(betweenness(grafo) == bmax) #vértice com max b.`
 - > `bmaxLabel = V(grafo)[bindex_max]$id #label do vértice com max b.`
- Closeness dos nós:
 - > `c = closeness(grafo)`
 - > `cmax = max(c)`
 - > `vindex_max = which(closeness(grafo) == cmax) #vértice com max c.`
 - > `vmaxLabel=V(grafo)[vindex_max]$id #label do vértice com max c.`

Essas são apenas algumas análises simples que podemos realizar utilizando o pacote *igraph* da ferramenta R. Calculamos a distribuição dos graus dos nós, betweenness e closeness para o grafo 2YPIA.G.net, que representa todos os contatos da enzima 2YPI cadeia A, independente do tipo dos mesmos. Esse é apenas um dos grafos do nosso conjunto de dados. Vejamos o significado dos demais grafos:

- 2YPIA.**A**.net: representa os contatos do tipo **atrativo**;
- 2YPIA.**H**.net: contém os contatos **hidrofóbicos**;
- 2YPIA.**L**.net: contatos do tipo **ligação de hidrogênio**;
- 2YPIA.**M**.net: representa os contatos **aromáticos**.
- 2YPIA.**R**.net: contatos do tipo **repulsivo**;

Com esses dados, estamos aptos a calcular as métricas de redes para os diferentes tipos de contatos, podendo fazer uma comparação entre seus valores.

Exercícios

1. Calcule o grau dos nós, distribuição dos graus, betweenness e closeness para o grafo 2YPIA.G.net e registre os resultados. Gere a representação visual do grafo e um gráfico da distribuição de seus graus.
2. Repita esse procedimentos para os demais gráficos, que representam os contatos atrativos, repulsivos, hidrofóbicos, ligação de hidrogênio e aromáticos.
3. Compare as métricas, grafos e representações visuais dos grafos analisados. Descreva as semelhanças ou diferenças para os diferentes tipos de contatos.
4. Acesse o Catalytic Site Atlas (CSA) (<http://www.ebi.ac.uk/thornton-srv/databases/CSA/>) e obtenha a informação de sítio ativo para a proteína 2YPI. Analise o grau, betweenness e closeness dos resíduos do sítio ativo. Compare os valores dessas métricas para os resíduos do sítio com os valores calculados no exercício 1.
Sugestão: apresente os valores numa tabela de resíduos x medidas.
5. Os resíduos do sítio ativo estão envolvidos em quais tipos de contatos?