

Trabalho Prático 3

Tópicos em Ciência da Computação: Introdução à Bioinformática

Sabrina de Azevedo Silveira
DCC030 DCC049

1 O problema

A teoria e algoritmos referentes a grafos podem ser utilizados para modelar e estudar uma grande variedade de problemas. Nesse trabalho, os grafos serão utilizados para modelar contatos intramoleculares, ou seja, entre os resíduos de aminoácido de proteínas. Já vimos nas aulas teóricas que os contatos estabelecidos entre os resíduos de uma proteína indicam a existência de interações químicas entre os mesmos e que tais interações estão fortemente relacionadas com a estrutura tridimensional da proteína e, conseqüentemente, com a função. Estudamos também que os contatos podem ser calculados utilizando-se critério geométrico ou limiar de distâncias.

No TP1 foram fornecidas: a sequência de uma triose fosfato isomerase (TIM, PDBid 2YPI) selvagem, tal como é encontrada na natureza; uma sequência da mesma enzima que sofreu alteração na função devido a mutações nos seus resíduos, chamada dTIM; foi fornecido ainda um conjunto de sequências protéicas da mesma família da 2YPI com estruturas resolvidas no *Protein Data Bank* (PDB)¹. Ainda no TP1, o objetivo era identificar quais mutações são capazes de impactar na função e o que poderia ser feito para restaurar a funcionalidade da proteína com base na implementação de um algoritmo de alinhamento de sequências.

O TP2 tinha como objetivo encontrar tendências ou exceções nas interações realizadas entre proteínas *Cyclin-dependent kinases* (CDKs) e seus ligantes para um conjunto contendo apenas a CDK-2 humana formando complexo com diversos ligantes distintos. Tal trabalho envolveu o cálculo de contatos entre os resíduos das proteínas e os átomos dos ligantes.

Neste trabalho, os dados fornecidos no TP1 deverão ser analisados novamente. Os contatos intramoleculares do conjunto de proteínas deverá ser calculado através de critério de distância (de modo semelhante ao que foi realizado no TP2) e cada uma das proteínas deverá ser modelada como um grafo (também chamado rede) de contatos. Esse conjunto de grafos deverá ser analisado de modo a revelar padrões ou anomalias em termos de contatos que possam apoiar na escolha das mutações que causam impacto na função. Os grafos devem ser analisados com base nas métricas de

¹<http://www.rcsb.org/pdb/home/home.do>

redes complexas estudadas, como *betweenness*, *closeness* e grau dos nós. Outras análises também podem ser realizadas.

O objetivo desse trabalho é, dada a proteína 2YPI, sua respectiva mutante e o conjunto de proteínas da mesma família:

- 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.

2 Cálculo de contatos

Nesse trabalho, adotaremos o cálculo de contatos com base num limiar de distância. Uma análise dos diversos valores para limiares de distância pode ser encontrada em [1]. É importante variar o limiar de distância para o cálculo dos contatos, o que impacta nos grafos e suas análises. Uma sugestão são os valores 5\AA , 7\AA e 9\AA .

3 Dados

Serão fornecidos os seguintes dados:

- 2YPI cadeia A: sequência da triose fosfato isomerase selvagem.
- dTIM: sequência mutante correspondente à 2YPI.
- family: conjunto de sequências protéicas que são da mesma família da 2YPIA.

4 Entrega do trabalho

O trabalho poderá ser realizado em dupla ou individual. Cada dupla terá 15 minutos para apresentar o trabalho no dia 27 de novembro de 2013, descrevendo todo o procedimento para cálculo de contatos e geração dos grafos, bem como todas as análises realizadas que justifiquem a seleção das 10 mutações consideradas mais severas para a função da proteína 2YPI. Devem ser entregues, também no dia 27 de novembro de 2013:

- Documentação: descrevendo o problema, a abordagem dada ao mesmo e todas as decisões de projeto/implementação tomadas.
- Código fonte: o código fonte completo, mais uma vez lembrando que a utilização de funcionalidades do Bio Perl e coisas do gênero são expressamente proibidas. É livre a utilização do R e seu pacote *igraph*.
- Apresentação: deve ser entregue em formato pdf.

Referências

- [1] Carlos H da Silveira, Douglas EV Pires, Raquel C Minardi, Cristina Ribeiro, Caio JM Veloso, Julio CD Lopes, Wagner Meira, Goran Neshich, Carlos HI Ramos, Raul Habesch, et al. Protein cutoff scanning: A comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. *Proteins: Structure, Function, and Bioinformatics*, 74(3):727–743, 2009.