

Trabalho Prático 1

Tópicos em Ciência da Computação: Introdução à Bioinformática

Sabrina de Azevedo Silveira
DCC030 DCC049

1 O problema

Neste trabalho estamos interessados no estudo da função protéica e, mais especificamente, como a sequência de resíduos de aminoácidos de uma proteína se relaciona com sua função. A partir de uma série de mutações feitas na sequência de uma proteína, queremos saber quais delas causam alteração na função. O objetivo é desenvolver um algoritmo de alinhamento de sequências de caracteres que possa identificar as mutações que alteram a função e sugerir modificações adicionais para a proteína que permitam resgatar a funcionalidade.

É importante pontuar que os caracteres nesse caso representam aminoácidos e cada um deles possui características físico-químicas distintas, de modo que determinadas trocas de aminoácidos podem causar impacto mais severo na função da proteína. Na ferramenta de visualização VERMONT¹ é possível ver as características de cada um dos aminoácidos.

Predizer mudanças na função protéica com base em mudanças na sequência de aminoácidos é um desafio para a ciência. Apesar da enorme quantidade de dados de sequências disponíveis, a resposta para a pergunta "a mutação no resíduo X é capaz de afetar a função da proteína Y?" continua em aberto.

Proteínas funcionais são coleções de átomos no espaço tridimensional e sua funcionalidade é determinada por como os substratos e outras moléculas que formam complexo com proteínas interagem com as mesmas em termos eletrostáticos, topológicos e da topografia da superfície. Os resíduos de aminoácido não funcionam isoladamente, de modo que as propriedades biofísicas do conjunto de resíduos deve ser analisada para fazer boas previsões a respeito das mutações que afetam a função.

Um algoritmo de alinhamento que permita comparar as propriedades biofísicas e estruturais de uma proteína mutante com a selvagem e, mais ainda, com proteínas funcionais da mesma família (que possuem a mesma função) e verificar se a proteína mutada e a selvagem se assemelham ou são diferentes do consenso da família poderia apoiar os especialistas do domínio na compreensão da

¹<http://homepages.dcc.ufmg.br/sabrinassilveira/vermont/>

relação entre sequência e função e ajudar a apontar quais resíduos mutados são os causadores de impacto na função protéica e o que pode ser feito para resgatá-la.

Nesse trabalho será fornecida a sequência de uma triose fosfato isomerase (TIM) que teve alteração na função devido a mutações nos seus resíduos. O objetivo do trabalho consiste em responder à pergunta: Quais dessas mutações são capazes de impactar na função e o que pode ser feito para restaurar a funcionalidade?

2 Dados

Serão fornecidos os seguintes dados:

- dTIM: sequência de uma triose fosfato isomerase com mutações em diversos de seus aminoácidos.
- 2YPIA: sequência da triose fosfato isomerase selvagem correspondente à dTIM, ou seja, a proteína original sem mutações.
- family: conjunto de sequências protéicas que são da mesma família (possuem a mesma função) da 2YPIA. Esse conjunto de sequências protéicas possui estrutura resolvida no *Protein Data Bank* (PDB)².

3 Entrega do trabalho

O trabalho poderá ser realizado em dupla ou individual. Cada dupla terá 15 minutos para apresentar o trabalho no dia 23 de setembro de 2013, descrevendo a solução proposta para o problema, os resultados e justificativa para os mesmos. Devem ser entregues no dia 19 de setembro 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, lembrando que a utilização de funcionalidades do Bio Perl e coisas do gênero são expressamente proibidas.
- Apresentação: deve ser entregue em formato pdf.

²<http://www.rcsb.org/pdb/home/home.do>