

Trabalho Prático 2

Tópicos em Ciência da Computação: Introdução à Bioinformática

Sabrina de Azevedo Silveira
DCC030 DCC049

1 O problema

As interações, também chamadas de contatos, que os resíduos de uma proteína estabelecem entre si são de grande importância para o enovelamento e manutenção da estrutura protéica. Tais interações também são muito relevantes para a compreensão dos complexos proteína-proteína e proteína-ligante. É importante ressaltar que esses contatos não são estabelecidos devido a ligações covalentes, mas interações chamadas fracas que, tomadas isoladamente, são muito mais fracas que as covalentes. Entretanto, como em geral ocorrem em grande número, seu resultado é significativo. Por exemplo, uma ligação covalente entre carbono e hidrogênio tem energia de 100Kcal/mol enquanto a energia de uma ponte de hidrogênio varia de 1 a 3 Kcal/mol.

Neste trabalho estamos interessados nos contatos proteína-ligante das enzimas conhecidas como *Cyclin-dependent kinases* (CDKs). As quinases são transferases (EC number classe 2) que fazem fosforilação, ou seja, transferem grupos fosfato de doadores de alta energia para substratos. Um exemplo pode ser visto na Figura 1.

Reaction catalyzed by non-specific serine/threonine protein kinase (2.7.11.1), cAMP-dependent protein kinase (2.7.11.11), cGMP-dependent protein kinase (2.7.11.12), protein kinase C (2.7.11.13), Ca²⁺/calmodulin-dependent protein kinase (2.7.11.17), polo kinase (2.7.11.21), cyclin-dependent kinase (2.7.11.22), mitogen-activated protein kinase (2.7.11.24), mitogen-activated protein kinase kinase (2.7.11.25), dual-specificity kinase (2.7.12.1), mitogen-activated protein kinase kinase (2.7.12.2)

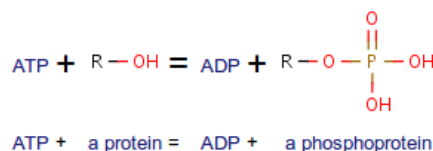


Figura 1: Reação catalisada pela quinase.

Dentre os vários papéis das CDKs, está o de serem enzimas relacionadas à regulação do ciclo celular. Além disso, defeitos no ciclo celular associados a tumor são frequentemente mediados por alterações na atividade de CDKs. CDKs desreguladas induzem proliferação celular em momentos inadequados bem como problemas em genes e cromossomos [2].

O objetivo desse trabalho é, dado um conjunto de CDKs para as quais existe estrutura tridimensional resolvida no Protein Data Bank (PDB)¹:

- Calcular os contatos estabelecidos entre os resíduos de cada CDK e seu respectivo ligante;
- Verificar a existência de tendências e exceções nesses contatos;
- Analisar os contatos encontrados (tipos, distância, dentre outros), apontando o significado dos padrões encontrados no item anterior.

Um ponto de partida para esse trabalho são os exercícios realizados em sala (aula 9) nos quais vários tipos de interações fracas foram analisados nos complexos proteína-aspirina (PDB ids 1TGM, 1OXR e 2QQT) utilizando a ferramenta *Ligand Explorer* do PDB. Outro recurso interessante é a ferramenta de análise de contatos do Blue Star Sting² da Embrapa (*Protein Ligand Contact*).

Com relação aos contatos, uma boa representação para os mesmos é uma matriz na qual as linhas representam os átomos do ligante e as colunas representam resíduos da proteína estudada, o que pode ser visto na Figura 2.

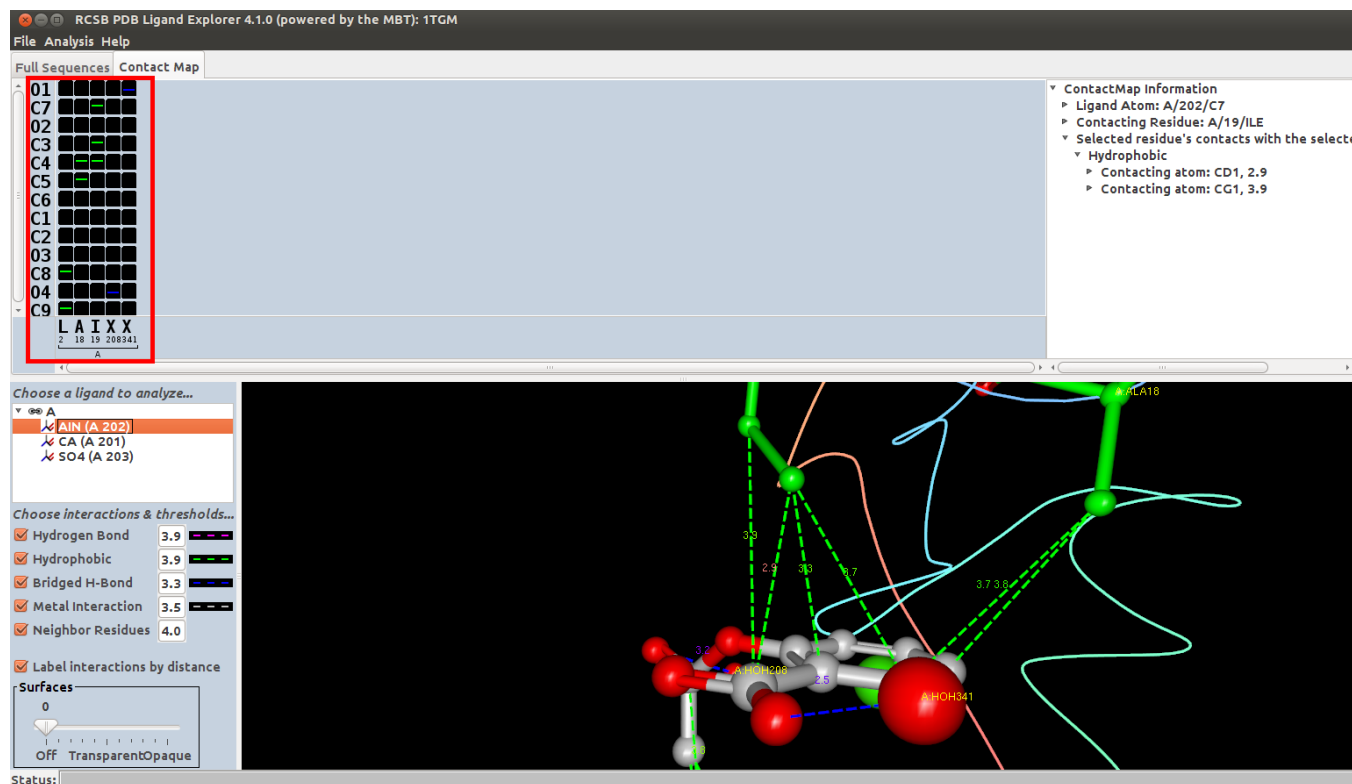


Figura 2: Matriz de contatos (destacada em vermelho) da ferramenta *Ligand Explorer* do PDB para o id 1TGM

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

²<http://www.cbi.cnptia.embrapa.br/SMS/>

2 Cálculo de contatos

Conforme estudamos, os contatos podem ser calculados utilizando-se um limiar de distância (normalmente chamado *cutoff*) ou critério geométrico. No caso do limiar de distância, dado um átomo ou resíduo, seus contatos são todos os átomos ou resíduos que se encontram num raio menor ou igual ao limiar. No caso do critério geométrico, os contatos são definidos através do cálculo do diagrama de Voronoi seguido da triangulação de Delaunay [3, 4]. Por simplicidade, nesse trabalho adotaremos o cálculo de contatos através do limiar de distância. Uma análise dos diversos valores para limiares de distância pode ser encontrada em [1].

3 Dados

Serão fornecidos os seguintes dados:

- Conjunto de CDKs: arquivos em formato pdb referentes às CDKs de interesse nesse trabalho.

4 Entrega do trabalho

O trabalho poderá ser realizado em dupla ou individual. Cada dupla terá 15 minutos para apresentar o trabalho no dia 28 de outubro de 2013, descrevendo a solução proposta para o problema de cálculo de contatos, as decisões de implementação, os resultados (padrões e exceções encontrados nas matrizes de contatos) e análise dos mesmos. Devem ser entregues até dia 28 de outubro 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.

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.
- [2] Marcos Malumbres and Mariano Barbacid. Cell cycle, cdks and cancer: a changing paradigm. *Nature Reviews Cancer*, 9(3):153–166, 2009.
- [3] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial tessellations: concepts and applications of Voronoi diagrams*, volume 501. Wiley, 2009.
- [4] Anne Poupon. Voronoi and voronoi-related tessellations in studies of protein structure and interaction. *Current opinion in structural biology*, 14(2):233–241, 2004.