

Uma proposta de ferramenta para a manipulação de conjuntos de proteínas do Protein Data Bank

Seus, Vinicius Rosa
Machado, Karina dos Santos
Werhli, Adriano Velasque
viniciusseus@gmail.com

Evento: XVII Encontro de Pós-Graduação

Área do conhecimento: Ciência Exatas e da Terra / Ciência da Computação

Palavras-chave: Dados proteicos; PDB; ferramenta

1 INTRODUÇÃO

PDB (do inglês *Protein Data Bank*) é um banco de dados público disponível na *web* com mais de 100 mil estruturas macromoleculares biológicas (Berman, et al., 2000). Com esta grande quantidade de estruturas de proteínas disponíveis no *PDB*, torna-se necessário o uso de ferramentas para aquisição e análise de conjuntos específicos de estruturas biológicas. Assim, neste trabalho, propomos o desenvolvimento de uma ferramenta para a aquisição, armazenamento e análises de conjuntos específicos de proteínas a partir do *PDB*. A ferramenta proposta é executada no ambiente *desktop* permitindo ao usuário adquirir estruturas a partir do *web-service RESTful* fornecido pelo servidor do *PDB*. Após a aquisição de um conjunto de proteínas interessantes ao usuário, o mesmo poderá manipular esses dados em um ambiente sem a necessidade de conexão com a internet. Esses dados armazenados são informações referentes a características das estruturas, por exemplo, ligantes, mutações, resíduos, seqüência secundária e dados de docagem molecular.

2 REFERENCIAL TEÓRICO

Segundo (Verli, 2014), as proteínas são consideradas polímeros sintetizados pelas células a partir de aminoácidos. São também classificadas como as biomoléculas mais versáteis da natureza, sendo capazes de adotar uma grande possibilidade de arranjos tridimensionais, não encontrada nos demais polímeros.

Segundo (Berman et al., 2000) um arquivo *PDB* é constituído de um conjunto de coordenadas atômicas e outras informações que descrevem proteínas e outras importantes macromoléculas biológicas. Biólogos estruturais utilizam métodos experimentais para determinar a localização de cada átomo em relação a outros átomos na mesma molécula como, por exemplo, cristalografia por difração de raio-X e ressonância magnética nuclear. Assim, esses pesquisadores depositam essas informações, as quais são anotadas e lançadas publicamente no arquivo *wwPDB5*.

3 MATERIAIS E MÉTODOS

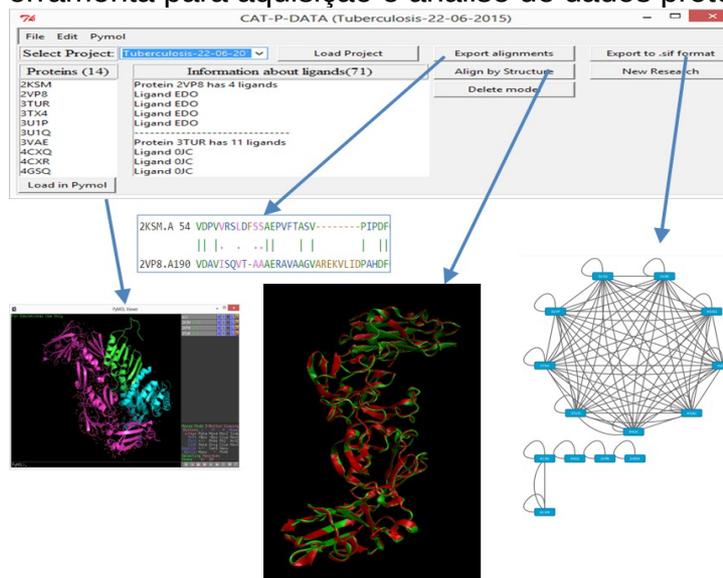
A ferramenta neste trabalho está sendo desenvolvida utilizando a linguagem de programação *Python*. Para a implementação do código em *Python* foi utilizada a ferramenta *PyCharm*. O banco de dados que acompanha a ferramenta foi implementado utilizando *MySQL*. Para a criação e manipulação do banco foi utilizada a ferramenta *Navicat*. A biblioteca *biopython* foi utilizada para a extração de informação dos arquivos *PDB*. Segundo (Cock et al., 2009) *Biopython* é uma ferramenta madura, de código livre que provê bibliotecas em *Python* que ajudam em

uma ampla gama de problemas encontrados na bioinformática.

4 RESULTADOS e DISCUSSÃO

A Figura 1 a seguir apresenta um resumo sobre o ferramental que o software desenvolvido neste trabalho possui.

Figura 1 – Ferramenta para aquisição e análise de dados proteicos do PDB



Fonte: Elaborada pelo autor

A Figura 1 apresenta as funcionalidades desenvolvidas até o atual momento. Como pode ser visualizada na Figura 1, a ferramenta apresenta um ambiente de visualização das estruturas terciárias das moléculas, além de uma ferramenta de alinhamento sequencial e alinhamento estrutural. Outra possibilidade desse *software* é a extração de proteínas que possuem certa similaridade entre elas, possibilitando o usuário identificar quais moléculas possuem uma estrutura sequencial semelhante e quais não possuem.

5 CONSIDERAÇÕES FINAIS

Acredita-se que a ferramenta em desenvolvimento nesse trabalho ajudará pesquisadores na descoberta de conhecimento em proteínas por meio do conjunto de funcionalidades que a ferramenta oferece.

REFERÊNCIAS

- BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. *The protein Data Bank. Nucleic Acids Research*, p. 235-242, 2000.
- VERLI, H. *Bioinformática: da Biologia à Flexibilidade Molecular*. <http://www.ufrgs.br/bioinfo/ebook>, v.1, 2014.
- COCK, P. J. A.; ANTAO, T.; CHANG, J. T.; CHAPMAN, A. B.; COX, C. J.; DALKE, A.; FRIEDBERG I.; HAMELRYCK, T.; KAUFF F.; WILCZYNSKI, B.; de HOON, M. J. L. *Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics*, v.25, n.11, p.1422-1423, 2009.