



A Proposal Tool for Manipulation of a Set of Protein Structures from PDB

Vinicius R. Seus ^{1,*}, Adriano V. Werhli ² and Karina S. Machado ²

¹ Universidade Federal do Rio Grande—FURG, Campus Carreiros: Av. Itália km 8 Bairro Carreiros

² Universidade Federal do Rio Grande—FURG, Campus Carreiros: Av. Itália km 8 Bairro Carreiros;
E-Mails: werhli@furg.br (A.V.W.); karina.machado@furg.br (K.S.M.)

* Author to whom correspondence should be addressed; E-Mail: viniciusseus@furg.br;
Tel.: +(55)-(53)-32-33-65-00.

Published: 4 December 2015

Abstract: Protein Data Bank (PDB) is a public web database with more than 100,000 biological macromolecular structures. With this large amount of protein structures available on PDB the use of tools for acquisition and analysis of specific sets of biological macromolecules is a necessity. Hence, in this work we propose the development of a tool for acquiring, storing and analyzing specific sets of proteins from the PDB database. The proposed tool runs on desktop environment allowing the user to acquire the structures from the RESTful web-service provided by PDB server. After the acquisition of a set of interesting PDBs the user can manipulate these data in an off-line environment through a local database that stores the information about the characteristics of the structures, for example, ligands, mutations, residues, sequences and docking results. The protein files are locally stored in the users' computer and can be used, for instance, for molecular docking simulations and alignment of sequences and structures. Having a set of proteins of interest available locally and using our proposed tool the user can perform analysis related to alignments and visualize important proteins characteristics improving the knowledge about specific target. Besides, the user can select PDB files to be visualized on a graphical environment that is integrated in our tool. Other features are related to the exporting of sequence alignments results in csv (comma separated value) format or exporting sequences that have a similar identity in a format that can be easily loaded on graph tools. These alignments allow the user to visualize which proteins are similar and discard those that are not.

Keywords: Protein Data Bank; Sequence alignment; Structural alignment

1. Introduction

With the large growth of protein data stored in the databases available on the web, appears the necessity to create different computer applications that help researchers in knowledge

discovery. Protein Data Bank (PDB) [1] is a public web database containing over 100.000 biological macromolecular structures. So, with this large amount of protein structures available

in PDB and other global servers that store various information of macromolecular structures, it becomes clear that the use of tools for the acquisition and analysis of specific sets of biological macromolecules is a need to facilitate the search for a specific target protein. In this work, we proposed the development of a tool for acquisition, storage and analysis of specific sets

2. Results and Discussion

In this section we present the proposed tool showing all modules and functionalities. Figure 1 shows all modules of our proposed tool. In the following each module is discussed:

- Proposed tool – This module is the application itself, where the local database and the interface are located in this scheme;
- PDB module - This module represents the public web database PDB (Protein Data Bank). The proposed tool connects with this database for acquiring molecular three-dimensional structures through the web-service RESTfull provided by the PDB. After the data acquisition, the user relates these data to a unique project that can be edited or deleted later;
- Sequence alignment module – This module performs sequence alignments. Having this feature, the user can align the sequences of all structure proteins of one project. As a result we have a matrix $n \times n$ where n is the number of proteins of a project and each cell is the identity between two sequence of proteins. Thus, it is possible to visualize proteins that have higher sequence similarity;
- Protein visualization module – This module opens an external protein visualization tool called PyMol[3].

of proteins from PDB in an environment that integrates different functionalities. The proposed tool is executed on the desktop environment allowing the user to perform acquisition of the molecular structures from RESTful web-service provided by the PDB own server.

Having this module the user can visualize the proteins of a project in a three-dimensional environment;

- Ligand visualization module – This module presents a list with all ligands presented in a protein structure discretized by chain linked to an user project;
- Structural alignment module - This module is for alignment of the tertiary structures. With respect to this operation, the user selects one of the proteins of his project to be the reference structure for all other structures of one project. Then, an algorithm is performed to align the tertiary structure of all proteins of a project with the reference structure. This functionality is important for Virtual Screening (VS) process, because to consider a set of receptor proteins with an equal grid box for docking all the structures need to have the same cartesian coordinate system.

The proposed tool was developed for the desktop environment. Using its interface, the user starts searching for a target receptor through a keyword in the search field. Next, a set of possible protein structures related to the specific target is listed in the tool showing all their PDB id's. Thus, the user can perform the acquisition

of these proteins structures. This acquisition process is possible using the web-service Restful provided by PDB itself together with the local database provided by our tool. The purpose of this database is to provide access to features of the molecular structures of a specific project in an offline environment. This database is populated after each time the user performs the

acquisition of a set of specific data from the proposed tool. Then, with the inclusion of molecular structures in the local database completed, the user can use all the modules presented in the figure 1.

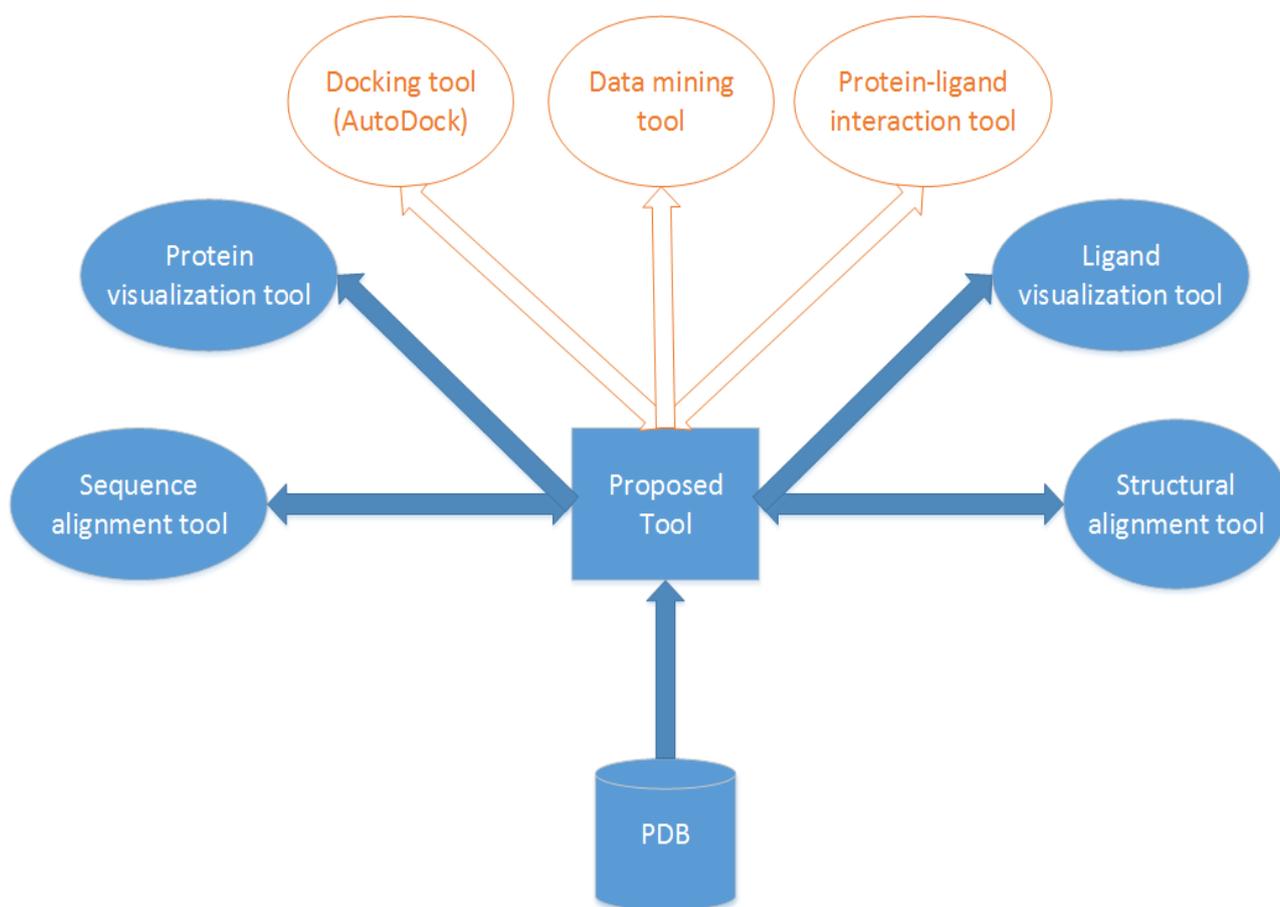


Figure 1. System's modules.

3. Materials and Methods

This section presents the materials and methods applied in the development of the proposed tool.

Our proposed tool was developed using Python language programming. Python [5] is a high level interpreted and interactive language

that provides to the developers the use of a strong and dynamic typing. Furthermore, the language provides an easy syntax to understand, turning the programming faster and more productive. Besides, Python presents another relevant feature that is the use of the virtual machine bytecode, what makes the code portable. This means that the program can be compiled in one platform to be executed on other platforms.

The local database of the proposed tool was implemented using MySQL. MySQL [4] is an open-source database management system that is the most popular in the world. Its uses SQL (Structured Query Language) as interface allowing an easy handling, excellent performance and stability.

For development of the tool it was used the IDE (Integrated Development Environment) PyCharm [6] and to develop the local database the Navicat [7] software was employed.

The Alignment sequence and alignment structural tools were developed using Biopython libraries.

According to Cock et al [2], Biopython is a mature and open source tool that provides libraries in Python that help in a wide range of problems commonly found in Bioinformatics. Moreover, Biopython has modules for reading and writing files with different formats and multiple sequence alignments. It also has modules that deal with tertiary macromolecular

structures, perform access to the PDB database available on the web and also provide numerical methods for statistical learning. Since its founding in 1999, Biopython has grown to currently having a large collection of modules. This tool is intended for developers of computational biology that need to incorporate in their scripts or their own software modules that help in their specific problems.

For the development of molecular visualization module the Pymol library was used.

According to Schrödinger [3] Pymol is a molecular visualization system that provides high quality three-dimensional images of small and larger macromolecules such as proteins. Pymol is one of the few viewing open source tools available for use in computational biology [3].

4. Conclusions

This paper presents a tool developed to allow the investigation of a set of different macromolecular structures related to a specific target.

Having our proposed tool, the researcher can perform a number of manipulations in a set of protein structures, working in a unique environment with modules for performing the alignment of sequences and structures, for visualization of these structures, to list the associate ligands and so on.

Acknowledgments

K. Machado acknowledges the support from CNPq (Process number 477462/2013-8). Vinicius Rosa Seus receives a CAPES Master Degree fellowship.

Conflicts of Interest

“The authors declare no conflict of interest”.

References and Notes

1. Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov I., Bourne, P.. The protein Data Bank. *Nucleic Acids Research* **2000**, *28*, 235–242.
2. Cock, P. J. A., Antao, T., Chang, J. T., Chapman, A. B., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, T., Wilczynski, B., de Hoon M. J.. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *11*, 1422–1423.
3. Schrödinger, L.. The PyMOL Molecular Graphics System, Version 1.7.4, LLC **2010**.
4. MySQL - The world's most popular open source database. Available online: <https://www.mysql.com> (accessed on 14 October 2015).
5. Python. Available online: <https://www.python.org/> (accessed on 14 October 2015).
6. PyCharm – The Most intelligent Python IDE. <https://www.jetbrains.com/pycharm> (accessed on 15 October 2015).
7. Navicat. <http://www.navicat.com> (accessed on 15 October 2015).

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions defined by MDPI AG, the publisher of the Sciforum.net platform. Sciforum papers authors the copyright to their scholarly works. Hence, by submitting a paper to this conference, you retain the copyright, but you grant MDPI AG the non-exclusive and unrevocable license right to publish this paper online on the Sciforum.net platform. This means you can easily submit your paper to any scientific journal at a later stage and transfer the copyright to its publisher (if required by that publisher). (<http://sciforum.net/about>).