

Sting_RDB: A Relational Database of Structural Parameters for Protein Analysis

Stanley R. M. Oliveira, Diego N. Rodrigues, Kassyus R. R. Souza, Gustavo V. Almeida, Goran Neshich, Paula R. Kuser-Falcão, Michel E. B. Yamagishi, Edgard H. Santos, Fábio D. Vieira, José G. Jardine

Embrapa Agricultural Informatics, Av. André Tosello 209, Campinas - SP – Brazil

Abstract

Sting_RDB is a relational database composed of structural parameters for protein analysis operating with a collection of both publicly available data (e.g., PDB [1], HSSP [2], Prosite [3], and UniProt [4]) and its proprietary protein sequence and structure (PSS) descriptors, such as geometric parameters (e.g., cavity, curvature), physical-chemical parameters (e.g., electrostatic potential), and conservation related parameters (e.g., HSSP, evolutionary pressure). The data consolidated and integrated into Sting_RDB makes this database one of the most comprehensive databases available for analysis of protein structure and sequence. The main features of Sting database can be summarized as follows:

- It is based on indices, which speeds up the search for information and, consequently, improves the response time in the PSS analysis process.
- It is available for different platforms. Currently, it is implemented and available in the database MySQL. However, it could be easily ported to other platforms, such as ORACLE and Postgres.
- It greatly reduces storage requirements since it was designed to generate a set of relationships that allow one to store information without unnecessary redundancy.
- It allows users to compare different protein structures, at the same time, which was not possible with the previous version (flat files).
- Its update is much simpler, since it was built on relational database features.

Apart from the features mentioned above, Sting_RDB is now going to be more accessible and readily addressable for data warehousing and mining. Most importantly, some effort has been made to make the Sting_RDB unique in terms of quality assessment when compared with other counterparts in the Bioinformatics domain. To the best of our knowledge, Sting_RDB is one of the most comprehensive data repositories for PSS analysis capable of providing its users with a data quality indicator. This database is connected to Sting, a Web-based suite of programs for comprehensive and simultaneous analysis of structure and sequence [5].

References

- [1] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. (2000). The protein data bank. *Nucleic Acids Res.* 28(1): 235–42.
- [2] Schneider, R., de Daruvar, A. and Sander, C. (1997). The HSSP database of protein structure-sequence alignments. *Nucleic Acids Res.*, 25, 226–230.
- [3] Hulo N., Bairoch A., Bulliard V., Cerutti L., De Castro E., Langendijk-Genevaux P.S., Pagni M., Sigrist C.J.A. (2006). The PROSITE database. *Nucleic Acids Res.* 34:D227-D230.
- [4] Apweiler, Rolf, Bairoch, Amos, Wu, Cathy H., Barker, Winona C., Boeckmann, Brigitte, Ferro, Serenella, Gasteiger, Elisabeth, Huang, Hongzhan, Lopez, Rodrigo, Magrane, Michele, Martin, Maria J., Natale, Darren A., O'Donovan, Claire, Redaschi, Nicole, Yeh, Lai-Su L. (2004). UniProt: the Universal Protein knowledgebase. *Nucl. Acids Res.* 2004 32: D115-119.
- [5] NESHICH, G.; BORRO, L. C.; HIGA, R. H.; KUSER, P. R.; YAMAGISHI, M. E. B.; FRANCO, E. H.; KRAUCHENCO, J. N.; FILETO, R.; RIBEIRO, A. A.; BEZERRA, G. B. P. VELLUDO, T. M.; JIMENEZ, T. S.; FURUKAWA, N.; TESHIMA, H.; KITAJIMA, K.; BAVA, A.; SARAI, A.; TOGAWA, R. C.; MANCINI, A. L. (2005). The Diamond Sting server. *Nucleic Acids Research*, v. 33, p. 29-35.