

ORAL PRESENTATION

Open Access

Integration of chemical information with protein sequences and 3D structures

Adel Golovin*, K Henrick, G Kleywegt

From 5th German Conference on Cheminformatics: 23. CIC-Workshop
Goslar, Germany. 8-10 November 2009

The Protein Data Bank (PDB) contains a wealth of small molecule - macro molecule complexes the study of which contribute enormously to our understanding of the interactions. However, exploiting and mining this treasure trove of data requires advanced analysis and retrieval methods that take into account both types of molecules. One such method is PDBeMotif, that has been developed by the Protein Data Bank in Europe (PDBe) at EMBL-EBI. Utilizing a relational database model at the back-end, the data structure represents a network of molecule, residue and motif interactions as well as their relative positions in the sequence and in 3D. The loader applies a number of algorithms to analyse PDB and derive necessary information, such as planarity and aromaticity of the chemical compounds, hydrogen-bonds network, coordination geometry, bond types (including pi electron interactions), 3D structural motifs, sequence domains and families. It collects information about sequence features, motifs and catalytic sites from available Distributed Annotation System (DAS) resources. The web application allows for a wide variety of searches and data analysis including protein motifs with chemical fragments association, protein sites characterisation, correlating properties, hits multiple sequence and 3D alignments. The whole system is released under GPL and available with the source code from <http://sourceforge.net/projects/pdbsam> and on line at <http://www.ebi.ac.uk/pdbe-site/PDBeMotif/>

Published: 4 May 2010

References

1. Golovin A, Henrick K: Chemical Substructure Search in SQL. *J Chem Inf Model* 2009, **49**(1):22-27.

* Correspondence: golovin@ebi.ac.uk

EMBL-EBI/PDBe, Hinxton Hall, Genome Campus, Cambridge, Cams, CB10 1SD, UK

2. Golovin A, Henrick K: MSDmotif: exploring protein sites and motifs. *BMC Bioinformatics* 2008, **9**:312.
3. Golovin A, Dimitropoulos D, Oldfield T, Rachedi A, Henrick K: MSDsite: A Database Search and Retrieval System for the Analysis and Viewing of Bound Ligands and Active Sites. *PROTEINS: Structure, Function, and Bioinformatics* 2005, **58**(1):190-9.
4. Golovin A, Oldfield TJ, Tate JG, Velankar S, Barton GJ, Boutselakis H, Dimitropoulos D, Fillon J, Hussain I, Ionides JMC, John M, Keller PA, Krissinel E, McNeil P, Naim A, Newman R, Pajon A, Pineda J, Rachedi A, Copeland J, Sitnov A, Sobhany S, Suarez-Uruena A, Swaminathan J, Tagari M, Tromm S, Vranken W, Henrick K: E-MSD: an integrated data resource for bioinformatics. *Nucleic Acids Research* 2004, **32** Database: D211-D216.

doi:10.1186/1758-2946-2-S1-O17

Cite this article as: Golovin et al: Integration of chemical information with protein sequences and 3D structures. *Journal of Cheminformatics* 2010 **2**(Suppl 1):O17.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>



ChemistryCentral



ChemistryCentral

© 2010 Golovin et al; licensee BioMed Central Ltd.