

neXtProt: a knowledgebase on human proteins

Authors: Pascale Gaudet¹, Pierre-André Michel¹, Monique Z. Zahn¹, Aurore Britan¹, Valérie Hinard¹, Paula Duek¹, Alain Gateau¹, Anne Gleizes¹, Daniel Teixeira¹, Frederic Nikitin¹, Valentine Rech De Laval^{1,2}, Mathieu Schaeffer^{1,2}, Lydie Lane^{1,2}, Amos Bairoch^{1,2}

¹ Swiss Institute of Bioinformatics (SIB)

² University of Geneva (UNIGE)

Corresponding author: Valentine Rech de Laval

Keywords: knowledgebase ; human ; protein

neXtProt (<http://www.nextprot.org>) is a comprehensive human protein-centric knowledgebase. Its web-based platform offers to its users a seamless integration of and navigation through protein-related data. Focused on human proteins, neXtProt aims to provide a state-of-the-art resource for the representation of human biology by capturing a wide range of data, with precise and fully traceable annotations. neXtProt integrates content relevant to human proteins such as Bgee, Human Protein Atlas (HPA), Peptide Atlas, SRMATlas, GOA, COSMIC, and IntAct.

neXtProt provides multiple data query tools to access entries, publications and terminology: 1) a simple search system using Solr technology; 2) an advanced search system using the SPARQL Protocol and RDF Query Language; 3) a REST API (<http://api.nextprot.org>) that allow users to develop their own applications/tools. neXtProt data are downloadable in several formats through: 1) a web interface from the search results page; 2) a FTP site, for current and past releases, controlled vocabularies and ontologies developed for neXtProt, etc. (<ftp://ftp.nextprot.org>). We develop viewers for the display of our data as modular generic components to make them reusable by the community independently of neXtProt. The source code of our components is available on Github (<http://github.com/calipho-sib>).

neXtProt data can be analyzed using tools such as BLAST, to find entries using an user sequence or a peptide uniqueness checker, to define which peptides can be used to validate the existence of human proteins.

The big challenge is to be able to understand how a given set of proteins share or differ in their various features. Indeed data is useful, but we want neXtProt to be much more than a well-organized comprehensive data repository. In addition to enhanced search capabilities, we offer tools that help to make sense of the contents.