

The Proteasix Ontology

Mercedes Arguello Casteleiro¹, Julie Klein² and Robert Stevens¹

¹School of Computer Science, University of Manchester, Oxford road, Manchester, United Kingdom.

²Institut National de la Santé et de la Recherche Médicale (INSERM), U1048, Toulouse, France. And Université Toulouse III Paul-Sabatier, Toulouse, France.

We describe the Proteasix Ontology (PxO) a new application ontology that supports the Proteasix tool (<http://www.proteasix.org/>). Proteasix is an open-source peptide-centric tool that can be used to predict automatically and in a large-scale fashion in silico the proteases involved in the generation proteolytic cleavage fragments (peptides). Body fluids (e.g. serum, urine, cerebrospinal fluid) contain thousands of peptides. However, there is a lack of knowledge about the proteolytic mechanisms leading to the generation of these fragments, which may hold the key to a better understanding of molecular mechanisms of disease, define new biomarkers, and therapeutic targets for a variety of disorders such as kidney, cardiovascular, autoimmune and infectious diseases as well as certain types of cancer. Proteasix addresses this by using the N- and C-terminal sequences of peptides which are reconstructed using information from UniprotKB and knowledge about cleavage sites to predict the possible proteases that were involved in their generation. This is a knowledge intensive task and the new version of Proteasix uses an ontology to provide this knowledge.

PxO re-uses parts of the Protein Ontology, the three Gene Ontology sub-ontologies, the Chemical Entities of Biological Interest Ontology, the Sequence Ontology, the NCBI Taxonomy Ontology and bespoke extensions in support of a series of roles:

1. To describe the known proteases, their target cleavage sites, observed and predicted proteolytic cleavage fragments resulting from proteolysis.
2. To enable the description of data within and produced by Proteasix to facilitate analysis.
3. To use knowledge about the function and location of a protease to support the prioritisation of proteases in the prediction.
4. To facilitate the visualisation of data in Proteasix.

In this paper we will describe the motivation for the PxO and how it supports and improves Proteasix in its analysis. In doing so we will describe the competencies PxO will need to support Proteasix; the axiom patterns that capture the appropriate knowledge and then the use of the ontology in satisfying those competencies. At the core of PxO is the modelling of the proteolytic process and its various participants, such as the protease and proteins together with their cleavage sites (observed and predicted), and then the output proteolytic

cleavage fragments.

The Proteasix Ontology may be found at: <http://swproteasix.cs.man.ac.uk> . This ontology is free and open for use by everyone.