An ontology-supported approach to predict automatically the proteases involved in the generation of peptides

Mercedes Arguello Casteleiro[1], Julie Klein[2] and Robert Stevens[1]

[1]School of Computer Science, University of Manchester, Oxford Road,Manchester, UK M13 9PL.
[2]Institut National de la Sante et de la Recherche Medicale (INSERM), U1048, Toulouse, France 24105.

ABSTRACT

The Proteasix Ontology (PxO) is an application ontology that supports the Proteasix tool, an open-source peptide-centric tool that can be used to predict automatically the proteases involved in the generation of proteolytic cleavage fragments (peptides). Body fluids (e.g. serum, urine, cerebrospinal fluid) contain thousands of protein fragments and disease-associated peptides. Information about proteases and their substrates and cleavage sites is scattered across publications and databases, different formats. This knowledge needs to be brought together and exploited to gain insights into the proteases implicated in peptide generation.

PxO seeks to support this task; it is built in OWL and JSON-LD following the W3C's vision of the Web of linked data. The PxO core is a selection of concepts names; concept expressions; concept definitions; and annotation assertions from the Protein Ontology (PRO), the three Gene Ontology (GO) sub-ontologies, Chemical Entities of Biological Interest Ontology (ChEBI), Phenotypic Quality Ontology (PATO), the Sequence Ontology (SO), NCBI Taxonomy Ontology, GALEN ontology, and Relationship Ontology (RO). The axioms extracted are designed to describe the biological underpinnings of the generation of peptides. These axioms together with axiom patterns, which capture the appropriate peptide-centric knowledge, enable the PxO ontology to support the Proteasix tool.

PxO addresses the following competency questions (CQ):
1. What are the known protease and their target cleavage sites (observed and/or predicted)?
2. For a given peptide, what are the cleavage sites that led to its production and is it the product of observed or predicted proteolysis?
3. What are the function, species and cellular location for both proteases and their substrate proteins?
4. For a given protease, what is the cleavage site specificity?

PxO was developed for the sysVASC project, which takes a systems medicine approach to understanding disese. Thus the emphasis is on Metazoa and more concretely, the organisms human, mouse, and rat. The Proteasix Ontology may be found at:
http://bioportal.bioontology.org/ontologies/PXO This ontology is free and open for use by everyone.