Challenges in relational and probabilistic data mining: a peptidomics case study (invited talk)

Walter Luyten

Abstract

Peptidomics is (by analogy with other -omics like genomics and proteomics) the systematic study of the complete set of (endogenous) peptides of an organism, tissue, cell, or organelle and its changes in space and time under different conditions. Endogenous peptides play a critical role as signaling molecules in most biological systems and their disturbance underlies many disease processes.

Mass spectrometry is the most common and powerful experimental analytical tool for peptidomics. However, the data it generates (mass of a peptide and of a number of its fragments) are usually insufficient to identify a peptide uniquely. Probabilistic methods as well as background knowledge can help to solve this problem.

Apart from this experimental approach, data mining has also been used to discover novel candidate peptides based on (sub)sets of known ones.

The identification of novel (candidate) peptides leads to the question of their function. Various data mining approaches have been used to assign function to peptides; we will review the literature and illustrate some of our own approaches.