

 Open access • Journal Article • DOI:10.1021/PR4001114

Predicting Tryptic Cleavage from Proteomics Data Using Decision Tree Ensembles

— [Source link](#) 

Thomas Fannes, Elien Vandermarliere, Leander Schietgat, Sven Degroeve ...+2 more authors

Institutions: Ghent University, Katholieke Universiteit Leuven

Published on: 04 Apr 2013 - Journal of Proteome Research (American Chemical Society)

Related papers:

- [Prediction of Missed Cleavage Sites in Tryptic Peptides Aids Protein Identification in Proteomics](#)
- [Prediction of missed proteolytic cleavages for the selection of surrogate peptides for quantitative proteomics.](#)
- [Getting intimate with trypsin, the leading protease in proteomics](#)
- [Chromatographic retention time prediction for posttranslationally modified peptides.](#)
- [Prediction of high-responding peptides for targeted protein assays by mass spectrometry.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/predicting-tryptic-cleavage-from-proteomics-data-using-3vqxq0n8se>

Supplementary table 1: Window size parameter tuning.

Window size	AUROC (%)
0	48.223
1	76.784
2	81.689
3	82.870
4	83.010
5	83.119
6	84.421
7	84.111
8	83.818
9	84.398
10	83.960
11	83.813
12	84.445
13	84.214
14	83.719
15	83.817
16	83.214

From each of the three test sets, iPRG, CPTAC and MS_Lims, 10% of the instances were removed and used to create a secondary test set. The secondary test set was used to tune the window size parameter. Different CP-DT models were learned on PRIDE for window sizes from 1 up to 16. Subsequently, these models were evaluated on the secondary test set, containing 10% of the three test sets. From these tests, a window size of 6 was selected, as it achieves the highest value on the secondary test set. Note that these test data are removed from the iPRG, CPTAC and MS_Lims, such that no overtuning of CP-DT occurs.