

Incorporating Digestion Efficiency Analysis to Improve Shotgun Proteomics Analytical Pipelines

Michael R. Hoopmann, Vidur Kailash, Luis Mendoza, and Robert L. Moritz
Institute for Systems Biology, Seattle WA



Overview

- SPACEPro (Shotgun Proteomic Analysis of Cleavage Efficiency of Proteins) is a software tool for evaluating digestion efficiency.
- SPACEPro is seamlessly integrated into the Trans-Proteomic Pipeline.
- SPACEPro identifies digestion deficiencies to optimize sample preparation conditions.

Introduction

Shotgun proteomics relies on efficient and predictable cleavage of proteins in a complex mixture for analysis by liquid chromatography mass spectrometry (LC-MS). Most shotgun proteomics analyses are performed following digestion with proteolytic enzymes, such as trypsin, that have been well characterized for their specificity and efficiency. Protein digestion efficiency is rarely evaluated or reported in proteomics data analyses, perhaps because ubiquitous use of digestion enzymes has led to the assumption that the digestion consistently runs to completion. However, poor protein cleavage can have adverse effects on protein identification and protein quantification, and therefore should be monitored in every experiment. Here, we show how digestion efficiency can be integrated into a shotgun proteomics pipeline to improve proteomics analyses.

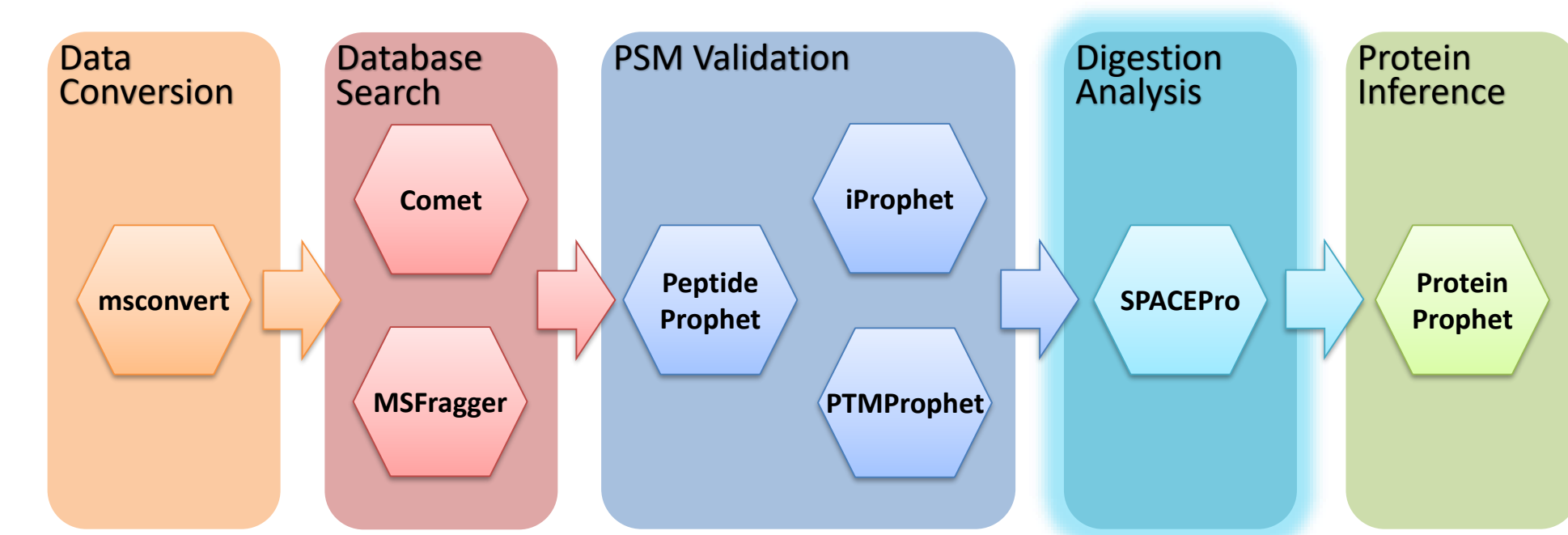


Figure 1: Integration of SPACEPro into the Trans-Proteomic Pipeline. SPACEPro seamlessly resides in the workflow following PSM validation with the various tools provided in the pipeline. Runtime is typically less than one minute, and therefore recommended to assure digestion quality for every analysis.

Features

Seamless Integration Into the Trans-Proteomic Pipeline Shotgun Proteomics Workflows

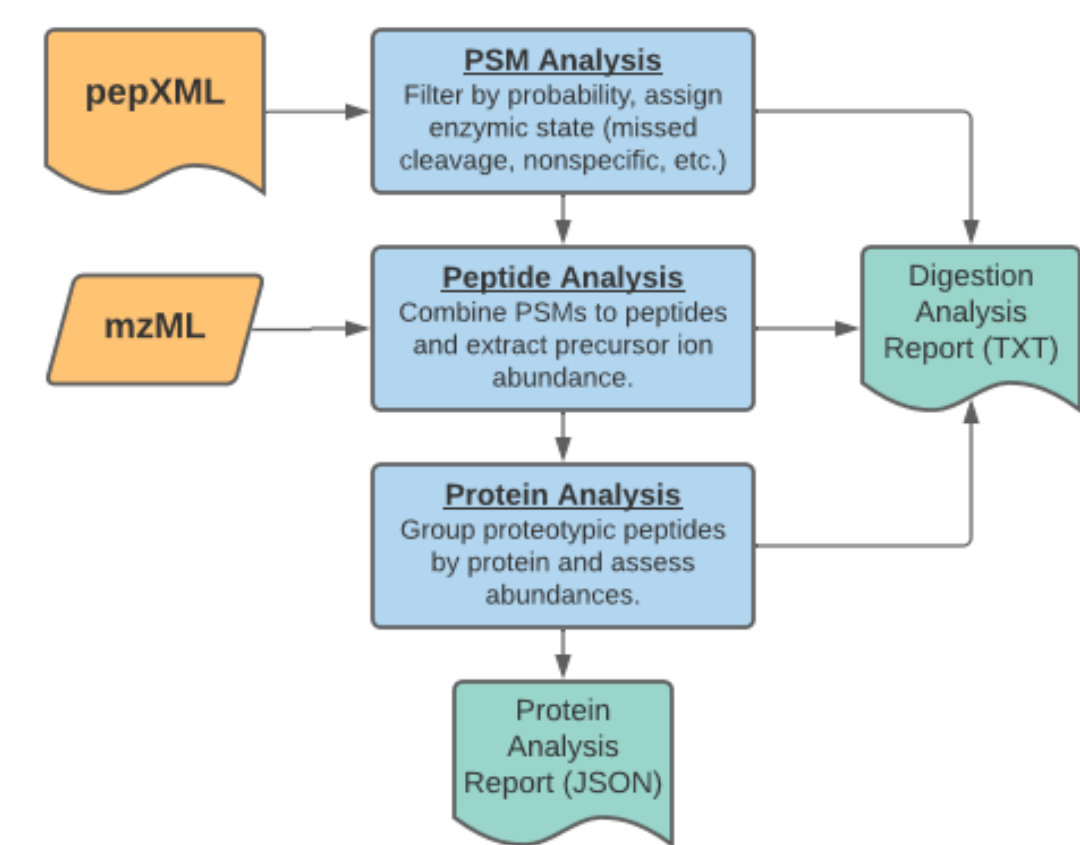


Figure 2: Flow diagram for SPACEPro. Input files (yellow) are shotgun search results in pepXML format and spectral data in mzML format, both widely-used open formats. Output consists of two files, a simple text summary of results, and an in-depth report of all proteotypic peptides arranged by protein identifier in JSON format, for further processing.

Simple, Browser-based User Interface (Command Line, too)

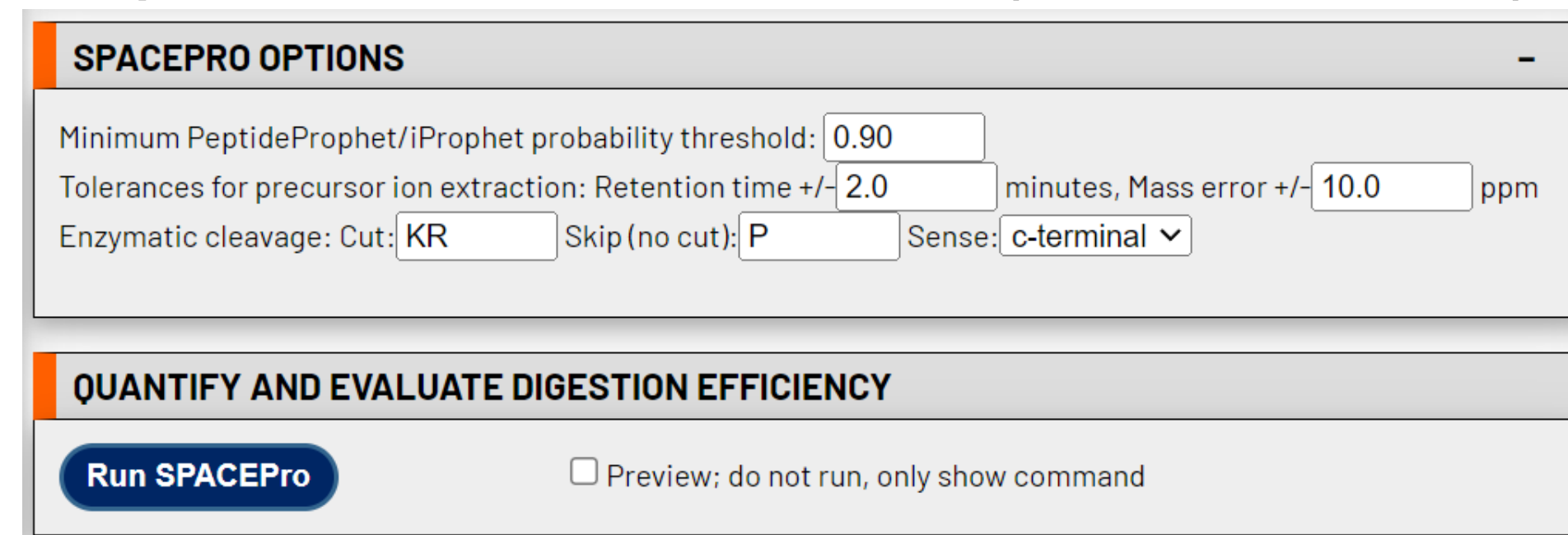
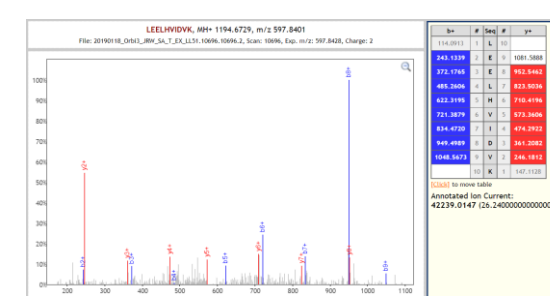


Figure 3: Browser-based user interface for SPACEPro. Minimal number of parameters, customizable to any digestion enzyme or instrument capabilities.

Peptide-based Quantitation of Digestion Efficiency



R.LEELHVIDK.F

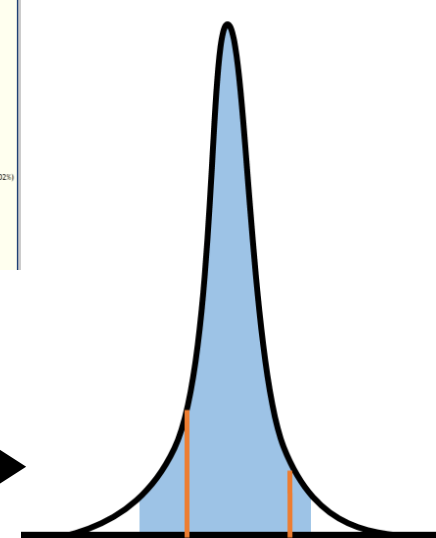


Figure 4: Quantitation of digestion efficiency is performed at the peptide-level. MS/MS-based peptide sequence matches (PSMs) are mapped back to their precursor extracted ion chromatogram (XIC), and the area under the XIC represents the abundance of a peptide. Multiple PSMs to the same peptide signal (orange lines) are represented as a single quantity.

Output

Clear Summaries of Analysis at PSM, Peptide, and Protein Level

```
----- PSM STATS -----
Total PSMs in file: 38657
PSMs above probability threshold: 13306
Enzymatic PSMs: 9455, 71.06%
Mis-cleaved PSMs: 2216, 16.65%
Nonspecific PSMs: 1822, 13.69%
```

```
----- PEPTIDE STATS -----
Total Unique Peptides: 4071
Total Peptides Quantified: 4066
Enzymatic Peptide Signal: 75.65%
Mis-cleaved Peptide Signal: 20.17%
Nonspecific Peptide Signal: 4.44%
Average number of mis-cleavages among mis-cleaved peptides: 1.09
```

Peptide-level analysis is best for evaluating digestion efficiency

```
----- PROTEIN STATS -----
Total Proteins from Proteotypic Peptides: 310
Average Enzymatic Protein Signal: 78.77%
Average Mis-cleaved Protein Signal: 9.36%
Average Nonspecific Protein Signal: 13.94%
```

Figure 5: SPACEPro exports a simple summary of analysis to the user. The summary contrasts typical PSM count-based results with the more refined, peptide quantity-based analysis recommended in SPACEPro.

All Proteotypic Peptide Results in JSON Format, Organized by Protein

```
"Protein Name": "sp|O43157|PLXB1_HUMAN",
"Total Intensity": 9589.31787109375,
"Total Enzymatic Peptide Intensity": 9589.31787109375,
"Total Mis-cleaved Peptide Intensity": 0.0,
"Total Nonspecific Peptide Intensity": 0.0,
"Percent Enzymatic": 100.0,
"Percent Mis-Cleaved": 0.0,
"Percent Nonspecific": 0.0,
"Peptides": [
  {
    "Sequence": "FGAVVIK",
    "Abundance": 3780.80810546875,
    "PSM Count": 1,
    "Mis-cleaved": false,
    "Nonspecific": false
  },
  {
    "Sequence": "YGLIQAAAVTSR",
    "Abundance": 5808.509765625,
    "PSM Count": 1,
    "Mis-cleaved": false,
    "Nonspecific": false
  }
]
```

Figure 6: Proteotypic peptide results are provided in JSON format, organized by protein identifier. Users can inspect the digestion efficiency and observed sequences of any particular peptide, or explore trends in digestion deficiencies shared by multiple peptides.

Digestion Optimization

Exploring Peptides with Missed Cleavage: Sequence Analysis and Efficiency Recalculation

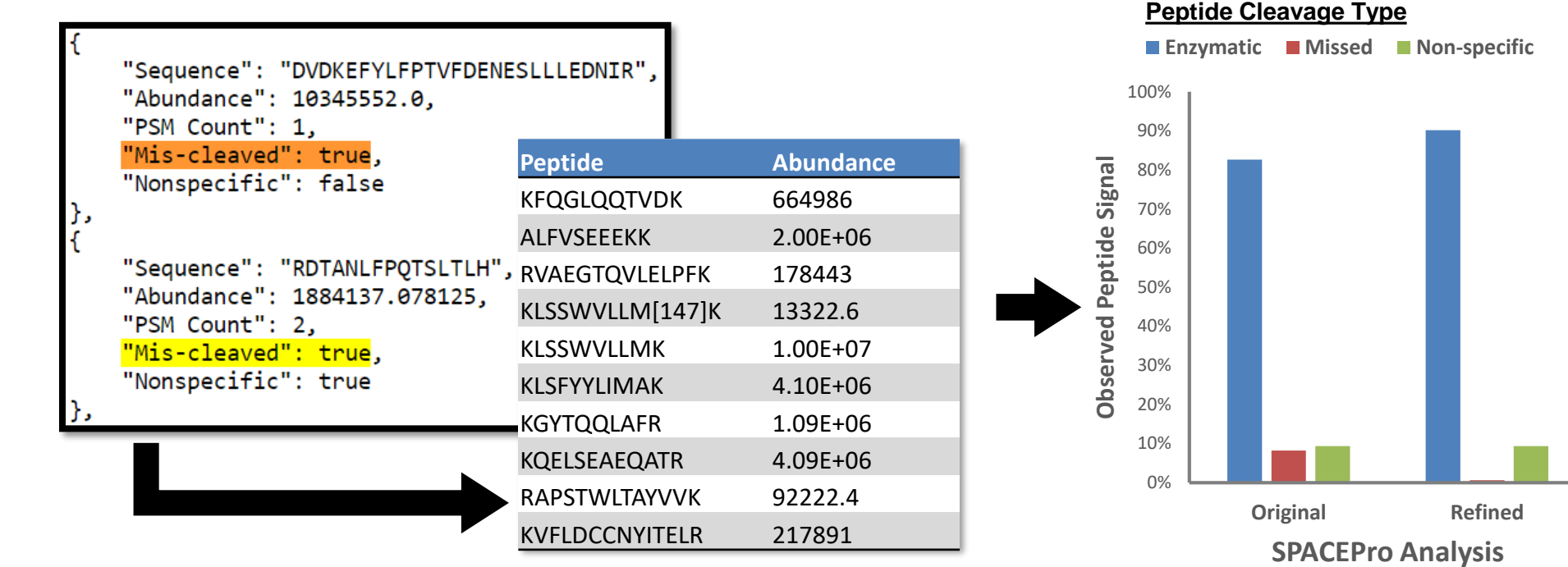


Figure 7: Analysis of missed cleavages from the JSON results showed that most occurred at dibasic sites (KK, RK/KR, RR), a known artifact of trypsin digestion. Accounting for these cases nearly eliminated the observation of missed cleavage and improved the digestion efficiency estimates.

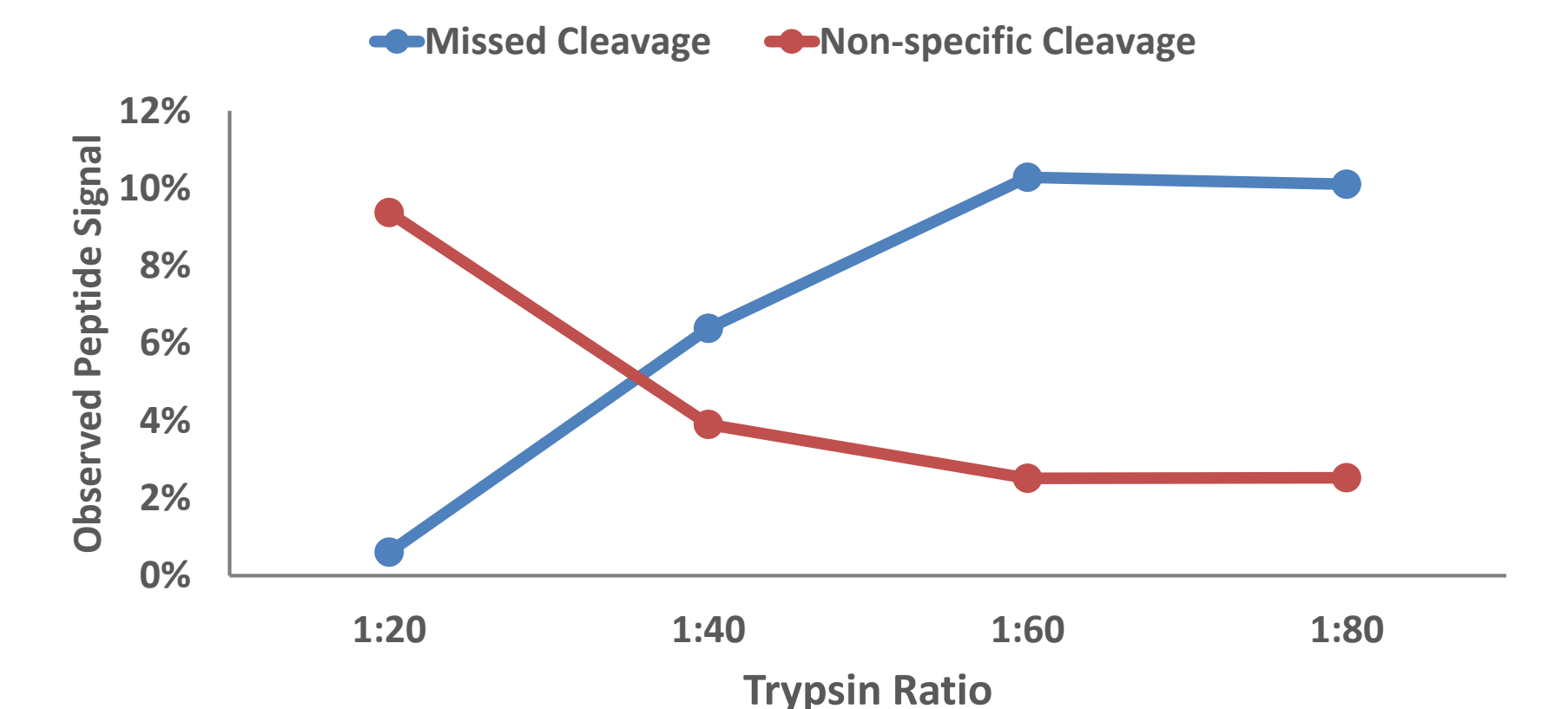


Figure 8: Optimization of the digestion of human serum samples (S-Trap, ProtiFi) following SPACEPro. Different enzyme-to-protein ratios were tested to find the optimum balance that minimized both missed and non-specific cleavage events.

Support & Information

Support provided by:

- R01GM087221
- NIA U19AG023122

<http://tppms.org>

