

In-Depth, Comprehensive Mapping of the Human Seminal Plasma Proteome by a Novel, Iterative LC-MS/MS/Database Search Workflow

Claire Dauly (1) Antoine D. Rolland (2), Martin Hornshaw (2), Régis Lavigne (3), and Charles Pineau (2,3)

¹Thermo Fisher Scientific, 16 Avenue du Québec, 91963 Courtaboeuf cedex, France; ²Inserm U625, Campus de Beaulieu, 35042 Rennes cedex, France;

³Proteomics Core Facility OUEST-genopole, Inserm U625, Campus de Beaulieu, 35042 Rennes cedex, France



Overview

Purpose: Maximize sequence coverage and protein identification for the analysis of non-liquefied human seminal plasma proteome using iterative LC-MS/MS and exclusion lists.

Methods: A list of peptides which were confidently identified from the first LC-MS analyses was generated with Proteome Discoverer software and used as a dynamic exclusion list during a second analyses of the sample. Peptides identified from the first and the second acquisition were further excluded during a third analyses of the sample.

Results: This strategy allowed the identification of additional peptides which were not characterized in the first LC-MS/MS analyses. Low copy number proteins were identified and overall protein coverage was increased, leading to more than 800 proteins confidentially identified (5% FDR).

Introduction

Apart from its obvious role in transporting male gametes, the seminal plasma provides a protective environment for ejaculated spermatozoa and improves their fertilizing potential. This fluid is also a potential source of biomarkers for male reproductive disorders. Here we describe a method for enhanced protein identification in complex samples. A combination of iterative LC-MS/MS, exclusion list generation and iterative database searching with Proteome Discoverer software was used to analyze in depth the seminal plasma proteome.

Methods

Non-liquefied seminal plasma (500 mg proteins) from a healthy donor was loaded onto two sequential hexapeptide ligand libraries (1 mL each; ProteoMiner™ - primary amino terminal peptides - and a carboxylated form, Bio-Rad Laboratories). Proteins bound to both libraries were desorbed through 4 different elution buffers, generating 8 complementary and treated sub-proteomes. Each fraction was then digested and analysed by nano-LC-MS/MS with a Thermo Scientific LTQ Orbitrap XL hybrid mass spectrometer. Peptides were preconcentrated with a peptide Captrap™ cartridge (Michrom Bioresources, Inc.) and separated onto a 15 cm x 100 µm capillary column. Detected peptides were selected for CID fragmentation using data dependent criteria and MS/MS spectra were searched against a SwissProt human database (April 2008) for peptide characterisation and protein identification. A list of peptide masses resulting from database searching was generated with Thermo Scientific Proteome Discoverer software for each sample and these identified peptide masses used to generate an exclusion list, including retention times, for subsequent LC-MS analysis of the same sample. A second database search was performed. Identified peptide masses from the first and second LC-MS experiments were combined to generate a second, longer exclusion list to be used for a third round of mass spectrometry analysis with subsequent database search yielding peptide and protein identification.

LC/MS:

HPLC System: Thermo Scientific Accela MS pump with a flow splitter
 Autosampler: Accela™ autosampler
 Columns: C18 analytical column (0.1 mm x 15 cm) with C18 trapping cartridge (Captrap, Michrom)
 Mobile Phase: A: Water/ 0.1% formic acid;
 B: Acetonitrile/ 0.1% formic acid
 Flow Rate: 300 nL/min, post split
 Mass Spectrometer: LTQ Orbitrap XL™

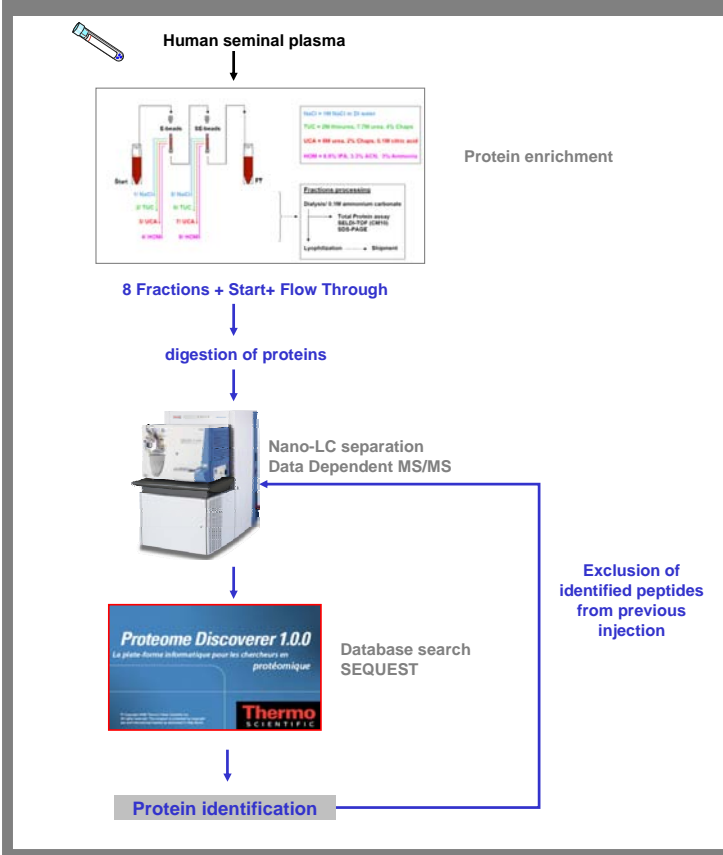
Spray Voltage: 2.2 kV
 Capillary Temp: 200 °C
 Dynamic Exclusion™: Repeat count: 1
 Exclusion list size: 500
 Exclusion duration: 120 s

Full MS Scans: MS/MS Scans:
 Detector: Orbitrap Detector: linear ion trap (LTQ)
 Resolution: 60000 AGC Target: 30000
 Reagent AGC Target: 1E6 Collision Energy: 35%

Data Dependent selection of Top 7 peaks for MSMS

Data Processing and Exclusion lists generation: Proteome Discoverer software using SEQUEST® and SwissProt human database (April 2008).

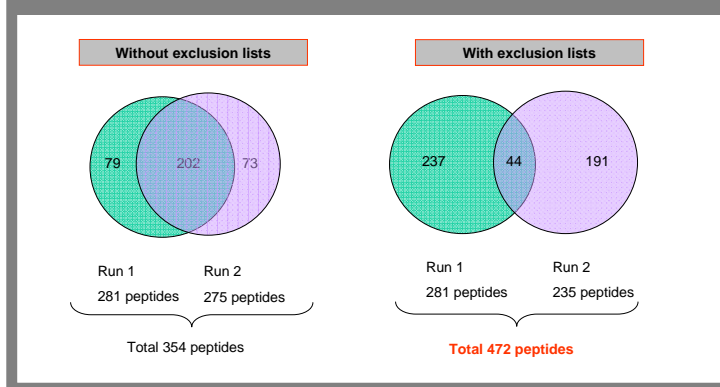
FIGURE 1. Experimental workflow



Results

Although the enrichment and depletion strategies have proven to be efficient at increasing the number of identified proteins in seminal plasma the complexity of the sample remains an issue. Only a fraction of the peptides present can be efficiently fragmented in a given period of time. To increase this number a new strategy was tested which consists of a re-injection of the sample for re-analysis by LC-MS/MS with exclusion of the peptides previously identified. A database search of the first MS/MS spectra acquired was performed to identify peptides and generate a list of peptide masses and retention times associated with highly confident identifications. To test this strategy one elution sample was analyzed a second time with and without using an exclusion list. The second analysis without exclusion list allowed the identification of a total of 275 peptides of which 73 had not been seen in the first run. However for the LC-MS experiment utilising an exclusion list, a total of 235 peptides were identified of which 191 had not been observed in the first experiment (Figure 2).

FIGURE 2. Multiple injections of fraction E1 with and without exclusion lists.



The same results were observed for all samples with an average of 78% of total peptides being newly identified in each acquisition with exclusion lists (Figure 3). This procedure was repeated 2 times for a total of 3 acquisitions per sample. A total of 864 proteins were identified of which 488 were identified with at least 2 distinct peptides (Figure 4). Low copy number proteins were identified like Annexin A5 as well as known markers of the seminal plasma: PSA, beta-defensins or prostatic acid phosphatase (Figure 5). New proteins were identified from the second or third injections. This strategy also increased the sequence coverage of many proteins. 35% of "one-hit wonders" from the first injection of elution E3, for example, could be confirmed by the identification of additional peptides.

FIGURE 3. Number of identified peptides through multiple injections of each fraction.

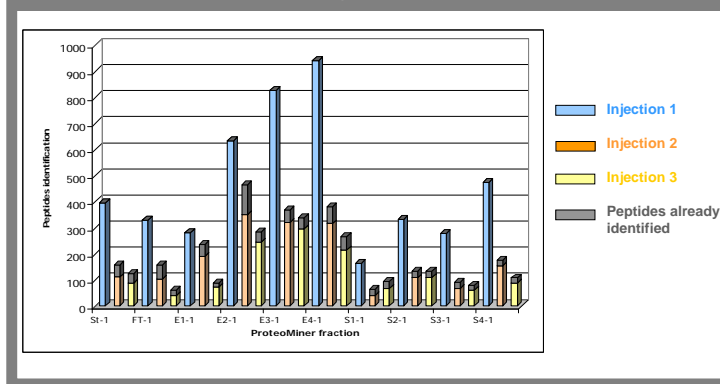


FIGURE 4. Number of identified proteins through multiple injections of each fraction.

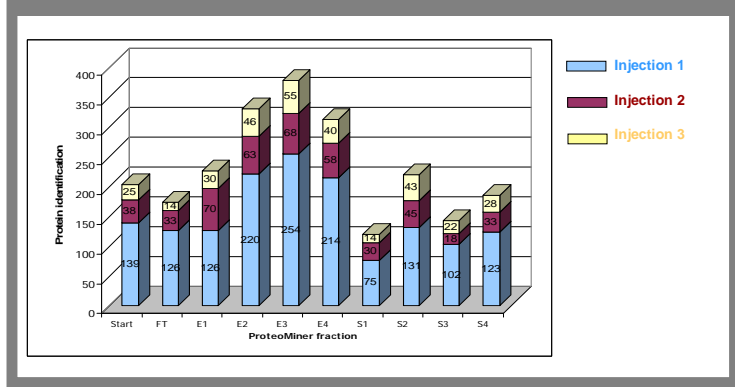
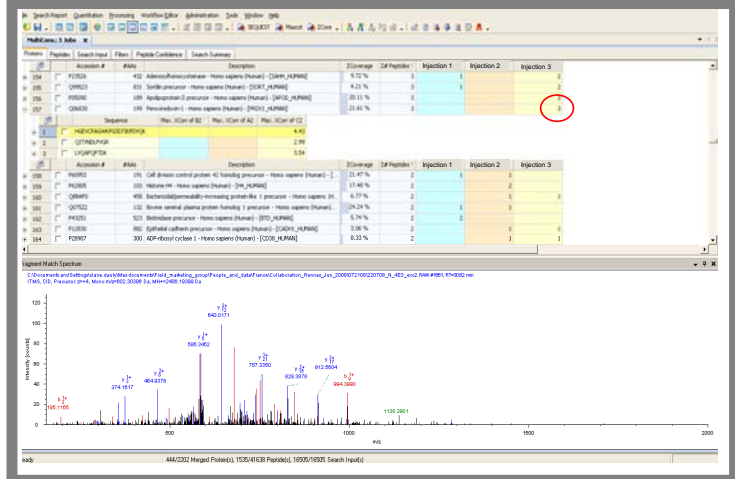


FIGURE 5. Peroxiredoxin-1 was identified from injection 3 only.



Conclusions

- Iterative analysis through multiple injections of the sample and exclusion of previously identified peptides enabled the in-depth analysis of seminal plasma identifying 864 proteins at 5% FDR.
- Proteome Discoverer software and the application of exclusion list for the data-dependent data acquisition on an LTQ Orbitrap XL provided more than 75% new peptides in each injection.
- The workflow also increased confidence in the results by eliminating 35% of "one-hit wonders" that were confirmed by the identification of a second peptide (or more).

SEQUEST is a registered trademark of the University of Washington. Captrap as a trademark of Michrom Bioresources. Proteominer is a trademark of Biorad Laboratories. All trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries. This information is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others.