Application Note: 469

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

Claire Dauly, Thermo Fisher Scientific, Courtaboeuf, France
Régis Lavigne, Charles Pineau, Proteomics Core Facility OUEST-genopole®, Inserm U625, Campus de Beaulieu, Rennes, France
Antoine D. Rolland, Inserm U625, Campus de Beaulieu, Rennes, France
Martin Hornshaw, Thermo Fisher Scientific, Hemel Hempstead, UK

Key Words

- Proteomics Workflow
- Exclusion Lists
- Seminal plasma
- Proteome Discoverer Software
- LTQ Orbitrap XL

Introduction

Apart from its obvious role in transporting male gametes, seminal plasma provides a protective environment for ejaculated spermatozoa and improves their fertilizing potential. Compositional changes can alter these properties and cause reproductive disorders. Thus, characterization of seminal plasma has the potential to provide biomarkers of reproductive disorders in males. Seminal plasma is a highly complex fluid that contains proteins secreted from several glands of the male reproductive tract: prostate, seminal vesicle, epididymis, and testis (Figure 1). However, like most body fluids, seminal plasma presents a wide range of protein concentrations that renders low-abundance proteins very difficult to identify and quantify.

Few studies have been performed to investigate the protein composition of seminal plasma. Several proteins were identified in the seminal plasma of healthy donors using a 2D gel-mass spectrometry approach.[1, 2] More recent studies were performed using a combination of 1D gel electrophoresis and LC-MS/MS analysis. In 2006, 923 proteins were identified by Mann and co-workers using this strategy with mass analysis performed by an LTQ FT instrument.[3] However, this study failed to identify known markers in the seminal plasma such as epididymis-specific defensins that may have been lost during the fractionation step, demonstrating the complexity of this biological fluid.

Mass spectrometers have a limited dynamic range where peptide analyte ions of interest can be fragmented efficiently for confident identification by database search. Due to sample preparation, chromatographic, and mass spectrometric constraints, a limited number of peptides can be analyzed in any given LC-MS/MS experiment. On-the-fly strategies involve the selection of the most intense peptides for fragmentation in a data-dependent manner. Very complex samples like seminal plasma thus require an optimized data acquisition strategy to achieve a thorough analysis of the sample.

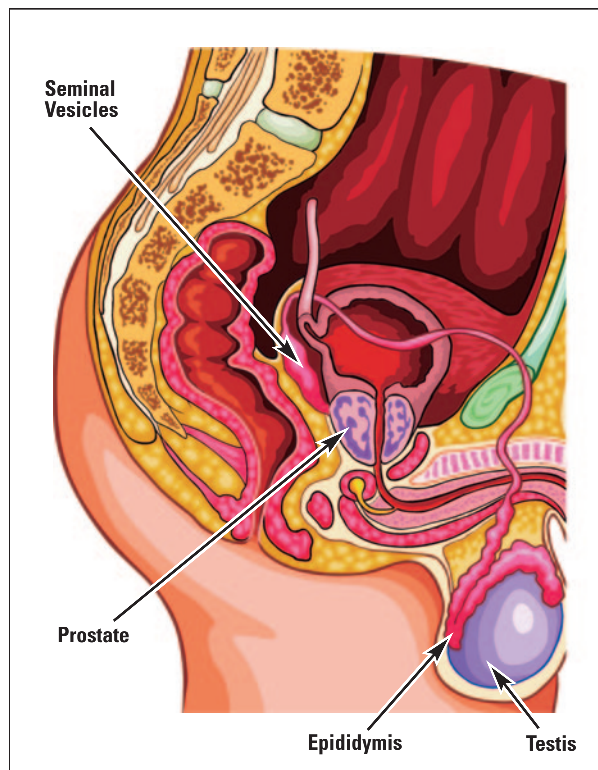


Figure 1: Organs secreting proteins found in seminal plasma

Here we describe a method for extensive protein identification in complex samples that involves iterative nanoflow LC-MS/MS analysis, exclusion list generation, and iterative database searching with Thermo Scientific Proteome Discoverer software, leading to the identification of numerous low-copy-number proteins in the seminal plasma proteome.

Goal

Develop a nano-LC-MS/MS method to deeply investigate complex protein samples and increase the number of low-copy-number proteins that can be identified.

Thermo SCIENTIFIC

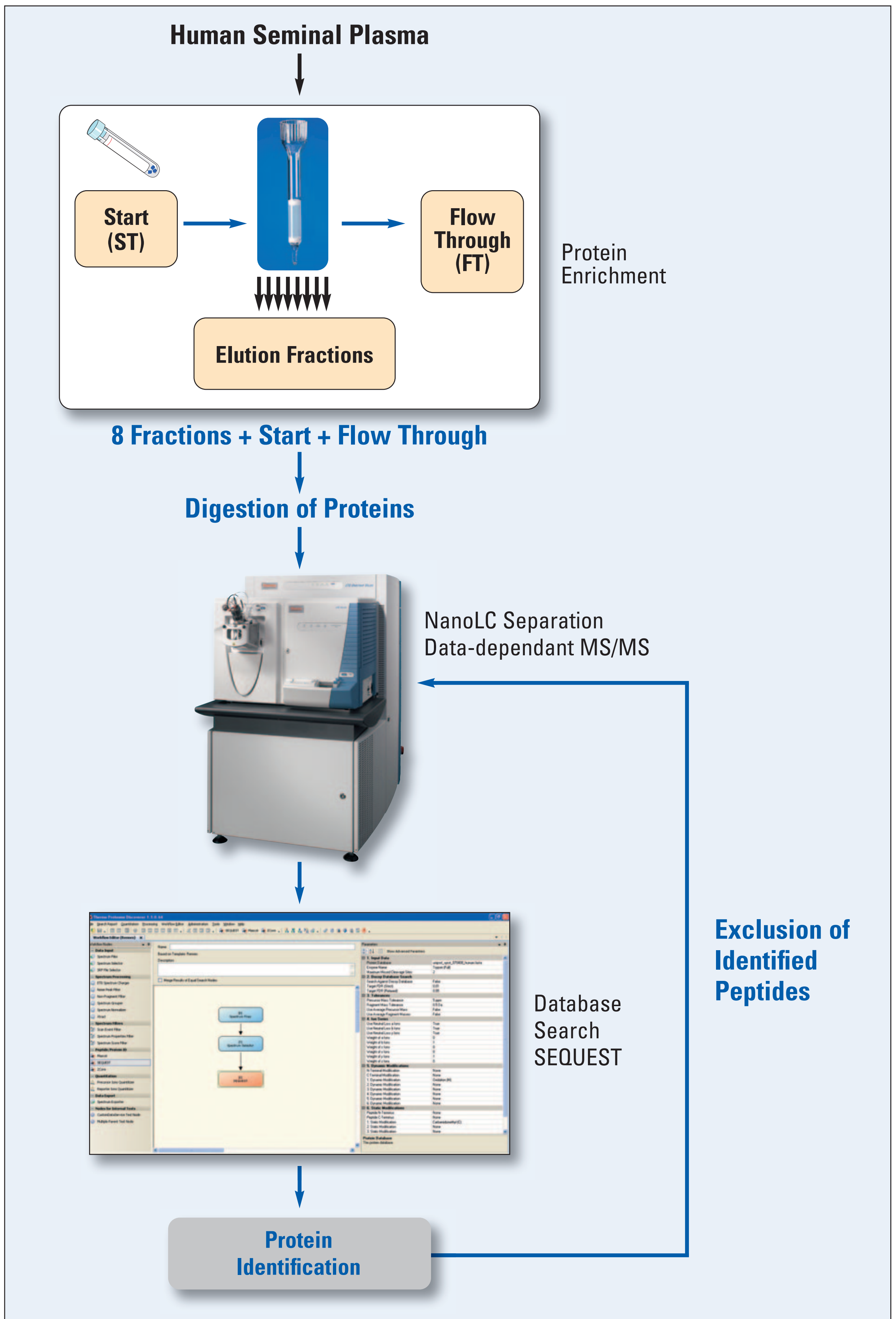


Figure 2: Experimental strategy

Experimental

The experimental strategy is described in Figure 2. Seminal plasma was fractionated into 8 complementary sub-proteomes. Each fraction was enzymatically digested and analyzed by nano-LC-MS/MS with an Thermo Scientific LTQ Orbitrap XL mass spectrometer. Proteome Discoverer™ software with the SEQUEST® search algorithm was used to identify proteins and produce a list of identified peptides containing *m/z* values and retention times for subsequent LC-MS/MS runs with exclusion lists containing all previously identified peptides.

Sample Digestion

A 500 mg sample of non-liquefied seminal plasma from a healthy donor was fractionated into 8 complementary sub-proteomes using the ProteoMiner™ system (BioRad Laboratories).[4] Each fraction was reduced with 65 mM DTT and alkylated with 135 mM iodoacetamide prior to enzymatic digestion and analysis by nano-LC-MS/MS with a LTQ Orbitrap XL mass spectrometer.

Liquid Chromatography

HPLC system: Thermo Scientific Accela MS pump with a flow splitter
Autosampler: Accela™
Column: C18 (Reprosil) 100 µm id x 15 cm packed tip column (Nikkyo Technos Co. Ltd) with C18 trapping cartridge (CapTrap™, MICHROM Bioresources, Inc.)
Mobile phase: A: Water / 0.1% formic acid;
B: Acetonitrile / 0.1% formic acid
Gradient: 5% to 40% of buffer B in 70 min
Flow rate: 300 nL/min, post split

Mass Spectrometry

Mass spectrometer: LTQ Orbitrap XL™
Spray voltage: 2.2 kV
Capillary temp: 200 °C

Data Dependent Acquisition

Full-scan mass spectra were acquired with the LTQ Orbitrap XL mass spectrometer at 60000 resolving power. The seven most intense peptides were selected for CID MS/MS scans in parallel in the linear ion trap.

Full MS scans:

Detector: Orbitrap™
Resolution: 60,000 at *m/z* 400 (FWHM)
Reagent AGC Target: 1E6

MS/MS scans:

Detector: linear ion trap
AGC target: 1E4
Collision energy: 35%

Dynamic exclusion: active

Repeat count: 1
Exclusion list size: 500
Exclusion duration: 120 s
Exclusion lists:
Retention time tolerance: ±1 min
Mass tolerance: ±10 ppm

Data Processing and Exclusion List Generation

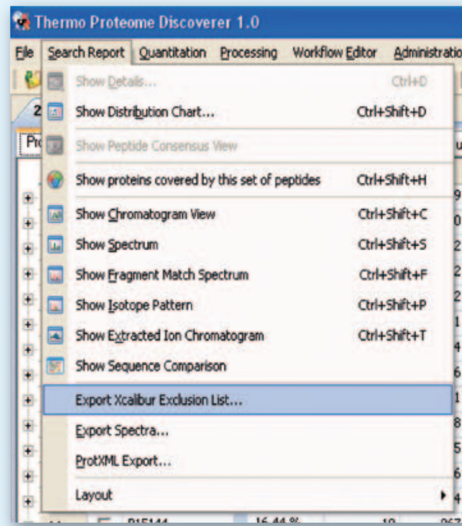
Proteome Discoverer software with the SEQUEST search algorithm was used for protein identification. MS/MS spectra were searched against a Swiss-Prot human database (release 55) for peptide characterization and protein identification. Modifications selected were set as fixed carbamidomethylation of cysteines, variable oxidation of methionine, and variable deamidation of asparagine and glutamine. Identified peptides were filtered based on a probability value of 10 and an Xcore value greater than 2.2 for doubly-charged peptides, 2.5 for triply-charged, 2.75 for quadruply-charged peptide ions, and 3 for charge states quintuple or higher.

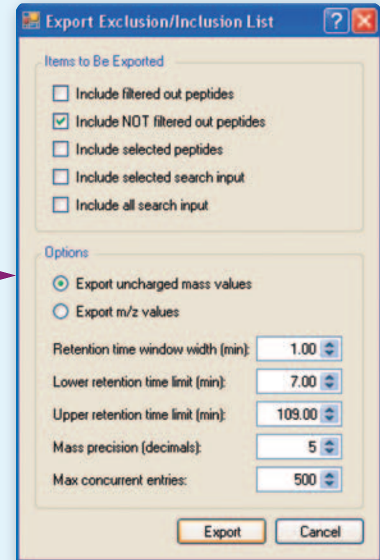
Proteome Discoverer software was used to generate lists of identified peptides to exclude them for subsequent LC-MS/MS analyses (Figure 3). Lists of unidentified peptides were exported as a text file containing uncharged mass values with 5 decimals mass precision and a retention time window of ±1 minute. The lists can be opened and edited with an editor. Thermo Scientific Xcalibur software does not allow inputting masses >4000 so they were deleted from this list.

The exclusion lists were then imported in the Xcalibur™ method file under the *Mass lists* tab. Note that the *Use global mass lists* option was checked in the *Reject mass list* data dependent settings tab to activate the *Mass lists* tab. Other data dependent settings were crucial for the success of the strategy (Figure 4). Under *Global* parameters, *Use m/z values as masses* was checked to work with uncharged masses. A mass tolerance of ±10 ppm was used under *Mass Widths* for *Reject mass width*. Using lower values can result in the reselection of masses because, when using a parallel mode of operation on an LTQ Orbitrap XL, the parent ion selection for an ion trap MS/MS is based on an Orbitrap preview scan. This preview scan is acquired at a lower resolution (RP 15,000) than the final Orbitrap full scan, so the masses are less accurate. Finally, *Enable monoisotopic precursor selection* was checked to avoid the selection of several isotopes from the same precursor.

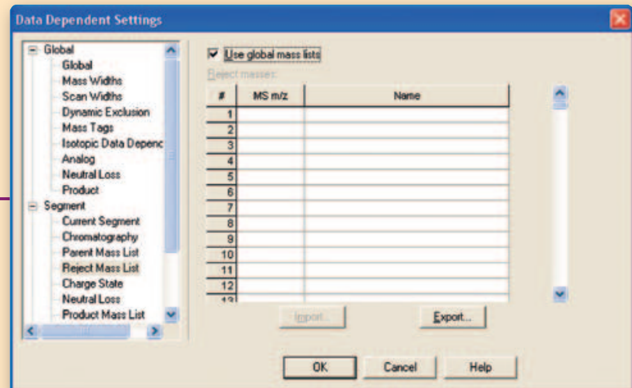
A total of 3 acquisitions per sample were performed following this strategy. For the third analysis, results files from the first two runs were combined in Proteome Discoverer software to generate a second exclusion list that contained peptides identified from the first and second acquisitions.

Proteome Discoverer





Xcalibur Method



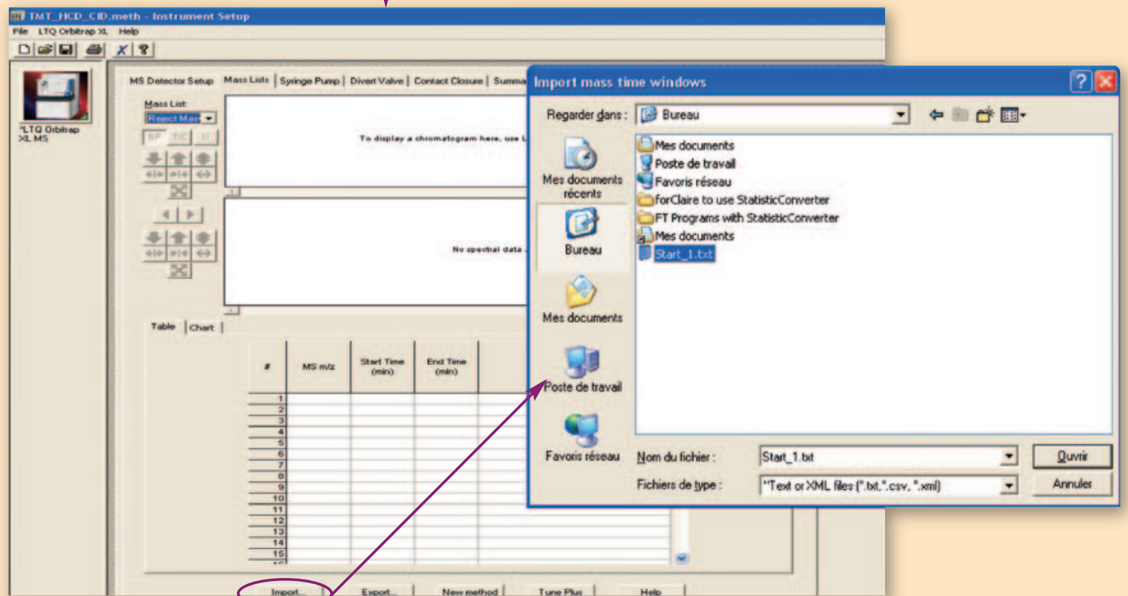


Figure 3: Exclusion list generation with Proteome Discoverer software and importation into the Xcalibur instrument method

Xcalibur Method

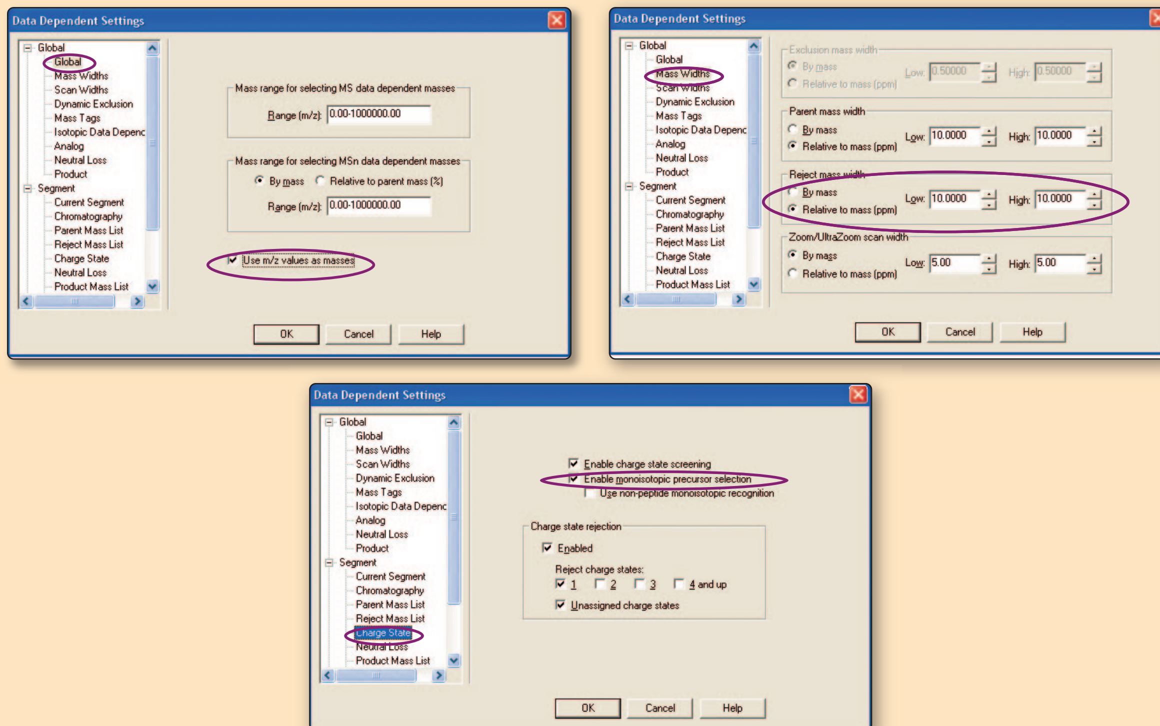


Figure 4: Optimum data dependant parameters for a successful use of exclusion lists

Results and Discussion

Proteins from seminal plasma were affinity-bound to an enrichment column, and naturally high-abundance species were reduced in concentration by washing the excess of unbound species.[4] Enriched proteins were eluted in 8 different fractions. After digestion and LC-MS/MS analysis using the LTQ Orbitrap XL, 395 unique peptides were identified in the Start (ST) sample, allowing the identification of 139 proteins. The same experiment was performed with the Flow Through (FT) and all different elution fractions of the chromatographic supports (Figure 5). There were 126 proteins identified in the FT and a large overlap was observed with the ST fraction, confirming that most of the abundant proteins bind only partially to the beads. More interestingly, 828 peptides were identified under the same LC-MS conditions in fraction E3, resulting in the identification of 252 proteins. When looking at the overlap between E3, ST, and FT, more than 70% of proteins identified in E3 could not be found in either ST or FT.

Although the enrichment strategy leads to an increase in the number of proteins identified, the complexity of the sample remains an issue. Even if peptides ionize well and can be detected, not all of them will be selected for CID in the linear ion trap due to the nature of data dependency. To increase the rate of peptide identification, another strategy was tested. This involved the re-injection of sample for re-analysis by LC-MS/MS with exclusion of the peptides

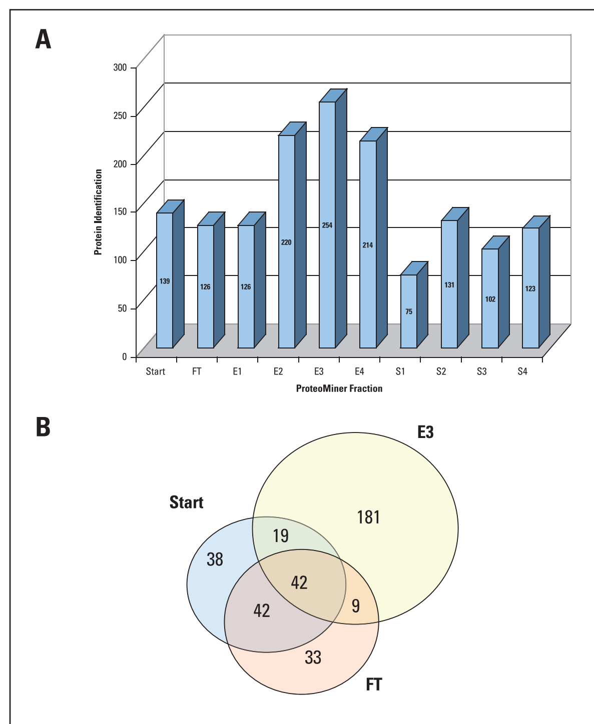


Figure 5: Protein identification after enrichment. A) Number of proteins identified in each fraction after LC-MS/MS analysis and SEQUEST search with Proteome Discoverer software. B) Venn diagram of protein identified in non-enriched seminal plasma Start, Flow Through (FT), and best elution E3.

previously identified. A database search of the first set of MS/MS spectra was performed to identify peptides and generate a list of masses corresponding to high-confidence identifications. Indeed, some MS/MS spectra were not informative enough to result in a positive identification in a first round of analysis, and better fragmentation may be obtained during subsequent analysis if the corresponding mass is not excluded. The database search, results parsing, and exclusion lists export can be done automatically with Proteome Discoverer software. The exported mass list also contains the retention time information, allowing the restriction of the dynamic exclusion to a short duration of 2 minutes. Moreover, the high mass accuracy provided by the LTQ Orbitrap XL mass spectrometer enables the specific exclusion of all peptide masses within a very narrow mass range (10 ppm).

To test this strategy, the first elution sample, E1, was analyzed a second time with and without applying an exclusion list (Figure 6). It is known that 2 consecutive injections of the same sample will not result in a complete overlap of the identified peptides. This phenomenon occurs due to small variations in the chromatographic conditions leading to differences in the full-scan mass spectra at a given retention time and to a different selection of peptides for fragmentation. In this case, the repeat analysis of E1 without an exclusion list allowed the identification of 275 peptides, of which 73 had not been identified in the first run. However, for the LC-MS experiment using an exclusion list, a total of 235 peptides were identified, of which 191 had not been observed in the first experiment. 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 7). This procedure was performed twice to make a total of 3 LC-MS acquisitions per sample. Database search results were combined to determine the total number of peptides and proteins identified in each sample and in the

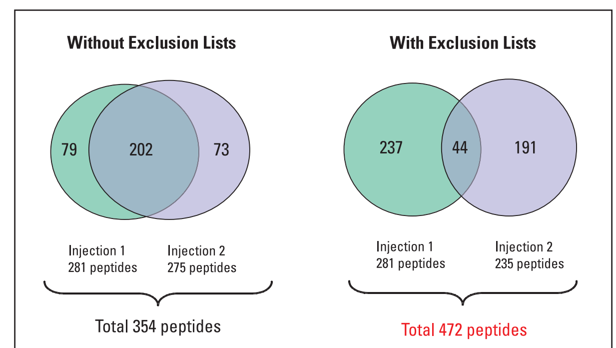


Figure 6: Multiple injection of fraction E1 with and without exclusion lists

whole analysis (Figure 7). A total of 864 proteins were identified, of which 488 had at least 2 distinct peptides. New proteins could be identified preferentially from the second or third injections, peroxiredoxin-1 (Figure 8) and histone H4 (Figure 9B) for example. PSA and prostatic acid phosphatase were also identified as well as low-copy number proteins, among them annexin A5 (Figure 9A) and beta-defensins. This strategy also increased the sequence coverage of many proteins. In total, 35% of "one hit wonder" proteins from the first injection of elution E3 were confirmed by the identification of additional peptides of the same protein in subsequent injections (Figure 9A).

Parallel analysis of the transcription levels of identified proteins was carried out using the Annotation, Mapping, Expression and Network (AMEN) software and Affymetrix transcriptome datasets of the different organs contributing to the production of the seminal plasma.[5] Identified proteins were further classified by subcellular localization and biological function using the advanced Inforsense capabilities within the Proteome Discoverer software, which allows retrieving GO annotations from Swissprot or NCBI databases (Figure 10).

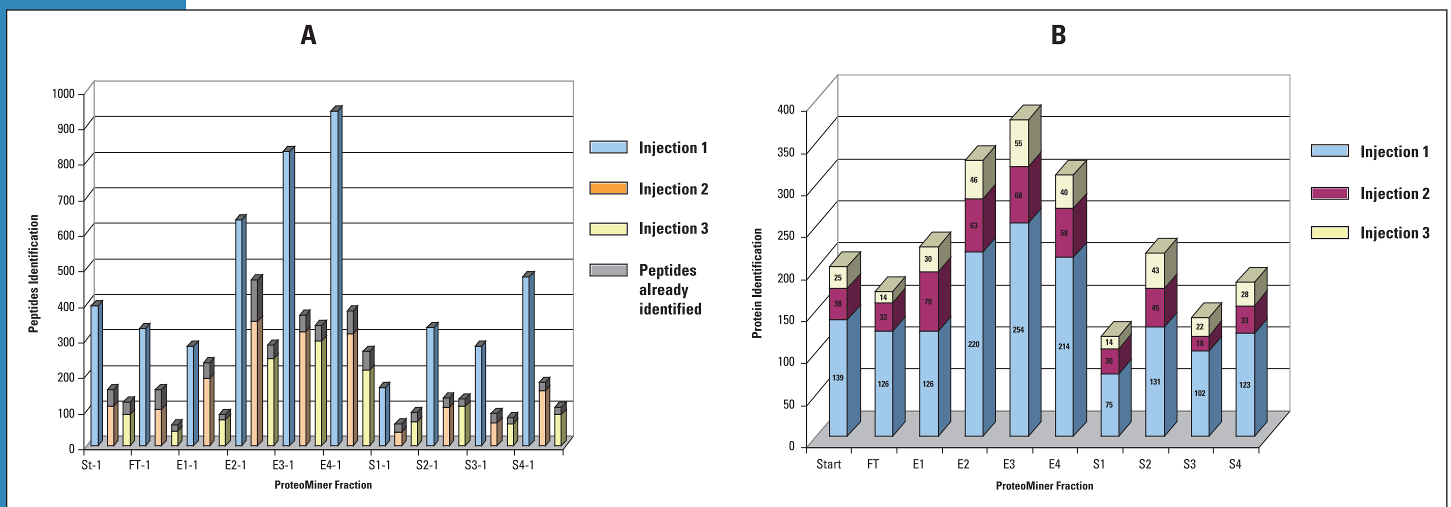


Figure 7: Number of identified peptides (A) and proteins (B) from multiple injections of each fraction and dynamic exclusion of previously identified peptides

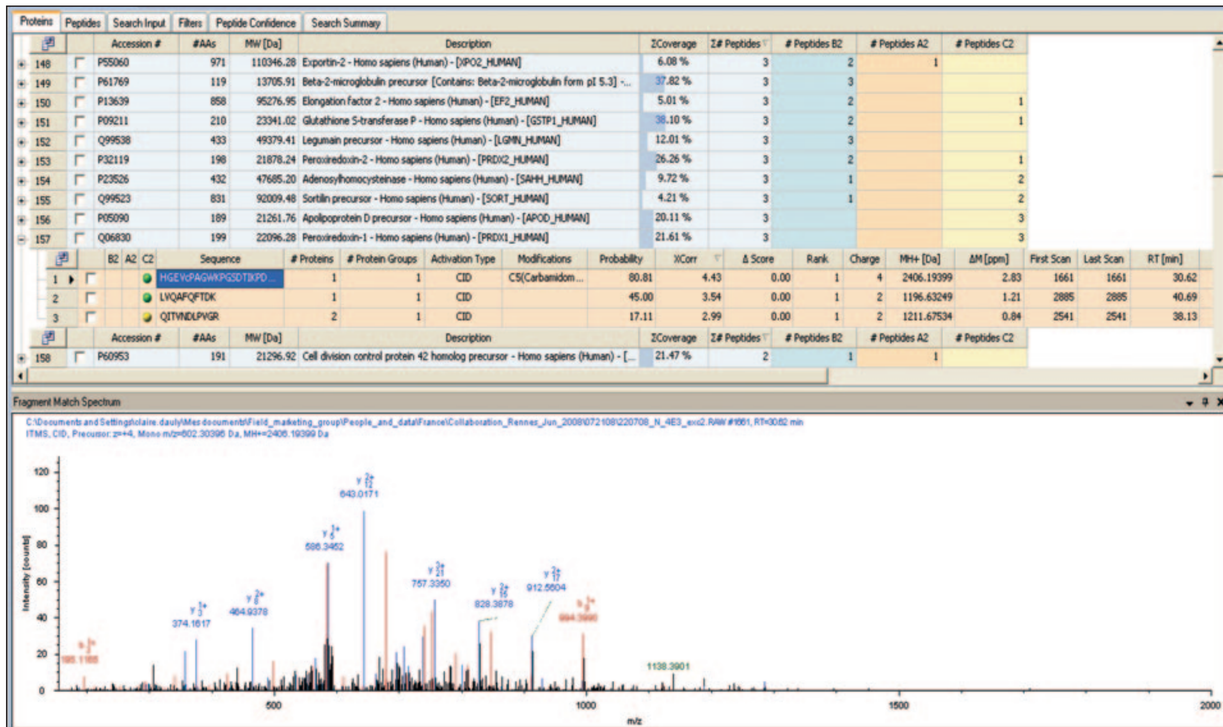


Figure 8: Peroxiredoxin-1 was identified from Injection 3 with 3 different peptides

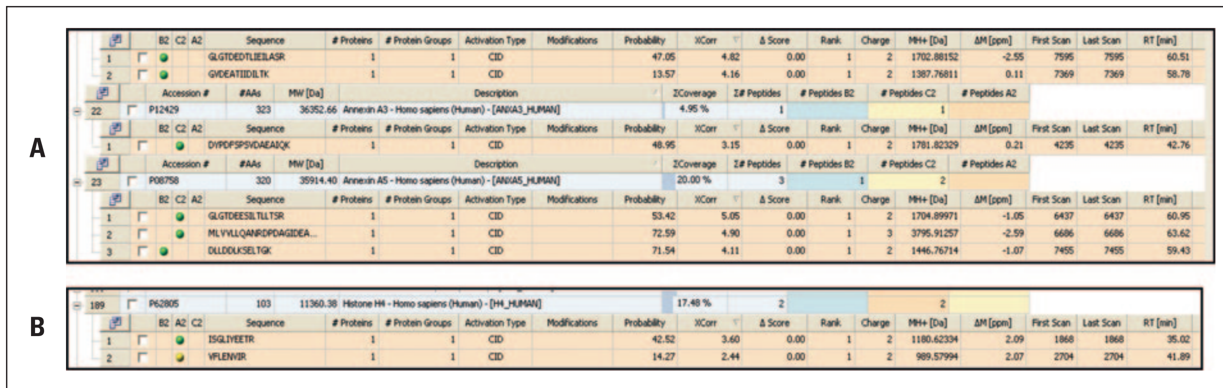


Figure 9: A) Annexin A3 was identified from injection 2, whereas the presence of annexin A5 was confirmed by the identification of 2 additional peptides in injection 3. B) Histone H4 was identified with 2 peptides from injection 2

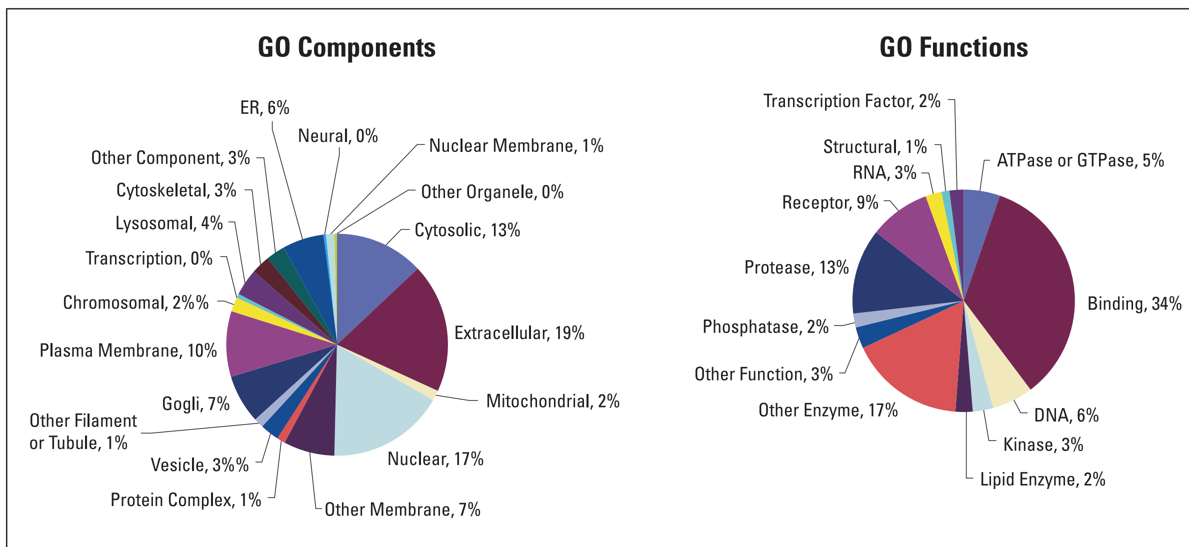


Figure 10: Gene ontology (GO) annotations of the identified proteins retrieved from Swissprot database by the Inforsense capabilities within the Proteome Discoverer software

Conclusion

The combination of protein enrichment and iterative injections with exclusion of previously identified peptides resulted in the identification of many more peptides and proteins in the seminal plasma sample. Re-injection of the same sample allowed the identification of about 80% new peptides in each run compared to about 25% without exclusion lists. New proteins were identified, "one hit wonder" proteins could be confirmed by additional peptides, and overall the protein coverage was increased, demonstrating the benefit of this strategy for deep proteomic characterization of complex biological samples such as seminal plasma.

References and Acknowledgements

1. Utleg AG, Yi EC, Xie T, Shannon P, White JT, Goodlett DR, Hood L, Lin B. Proteomic analysis of human prostasomes. Prostate 2003; 56: 150-161.
2. Fung KY, Glode LM, Green S, Duncan MW. A comprehensive characterization of the peptide and protein constituents of human seminal fluid. Prostate 2004; 61: 171-181.
3. Pilch B, Mann M. Large-scale and high-confidence proteomic analysis of human seminal plasma. Genome Biol 2006; 7: R40.
4. Boschetti E, Righetti PG. The ProteoMiner in the proteomic arena: a non-depleting tool for discovering low-abundance species. J Proteomics 2008; 71: 255-264.
5. Chalmel F, Primig M. The Annotation, Mapping, Expression and Network (AMEN) suite of tools for molecular systems biology. BMC Bioinformatics 2008; 9: 86.

In addition to these offices, Thermo Fisher Scientific maintains a network of representative organizations throughout the world.

Africa-Other +27 11 570 1840
Australia +61 2 8844 9500
Austria +43 1 333 50 34 0
Belgium +32 2 482 30 30
Canada +1 800 530 8447
China +86 10 8419 3588
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland/Norway/Sweden +46 8 556 468 00
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9434
Italy +39 02 950 591
Japan +81 45 453 9100
Latin America +1 608 276 5659
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
South Africa +27 11 570 1840
Spain +34 914 845 965
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

www.thermo.com

Legal Notices

©2016 Thermo Fisher Scientific Inc. All rights reserved. CapTrap is a trademark of MICHROM Bioresources, Inc. ProteoMiner is a trademark of Bio-Rad Laboratories, Inc. SEQUEST is a registered trademark of the University of Washington. OUEST-genopole is a registered trademark of CRITT Sante Bretagne Association. All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries. This information is presented as an example of the capabilities of Thermo Fisher Scientific Inc. products. It is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others. Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details.

***View additional Thermo Scientific LC/MS application notes at:* www.thermo.com/appnotes**

AN63100_E 1016M