Comparison of Hydrophobic, Lipophilic and Immunodepletion Pre-Fractionation Methods for Label-Free LC-MS/MS Identification of Biomarkers in Human Cerebrospinal Fluid

Sylvain Lehmann1*, Jérôme Vialaret1*, Martial Seveno2, Laurent Tiers1, Audrey Gabelle1,3 and Christophe Hirtz*4

1CHU Montpellier, Institute of Research in Biotherapy, Hôpital Saint Eloi, Laboratory of Biochemistry and Clinical Proteomics CCBHM, Montpellier, INSERM U1040-UM1, Montpellier, F-34000 France
2Platform Functional Proteomics, IGF, CNRS UMR 5203, INSERM U661, Université Montpellier I and II, Montpellier, France
3Memory Resources Research Centre, CHU Montpellier, Gui de Chauliac Hospital, Montpellier, Montpellier I University, Montpellier, F-34000 France
4These authors contributed equally to the manuscript

Abstract

Background: Proteomics analysis of human cerebrospinal fluid (CSF) is a major tool for identifying novel biomarkers for neurological diseases. However, the complexity and wide dynamic range of CSF represent a major challenge for detecting specific low-abundance biomarkers. One way to overcome this problem is to rely on different pre-fractionation techniques. However, the most relevant technique remains to be determined.

Methods: This study compared three different well-known pre-fractionation methods: immune-depletion of major proteins (Seppro® IgY14), hydrophobic solid phase extraction (Oasis® HLB), and lipophilic sorbent concentration (Liposorb®™). Unfractionated and pre-fractionated CSF was digested with trypsin and analyzed by RP-LC-MS/MS with an Orbitrap™ mass spectrometer. We documented the number of peptides detected and sets of proteins identified. Experiments were repeated to minimize pre-analytical and analytical variability.

Results: Compared to unfractionated CSF, the OASIS® HLB fractionated CSF method showed a significant 28% increase in the total number of proteins identified, while the Liposorb®™ capture resulted in a significant 46% decrease. Interestingly, results based on the number of peptides detected were different. We also evaluated the capacity of these pre-fractionation methods to detect different proteins in terms of their molecular weight, isoelectrophoretic point (IEP) or nature. Each of these pre-fractionation methods identified a specific subset of proteins, when compared to unfractionated CSF, and/or other methods. This was particularly obvious for the lipophilic sorbent, which allowed the detection of many lipoproteins.

Conclusion: Direct analysis of digested CSF led to the identification of several proteins despite matrix complexity. As expected, single pre-fractionation methods that can be included in simple and cost-effective workflows, yielded significant differences in terms of number, or range of proteins identified. This suggests that a single pre-fractionation method cannot cover the full range of protein species present in a complex sample.

Keywords: Mass spectrometry; Cerebrospinal fluid; Biomarkers; Pre-fractionation

Introduction

Proteomics studies using mass spectrometry (MS) are well-suited to the identification and quantitative detection of protein/peptide biomarkers in complex biological fluids such as blood or cerebrospinal fluid (CSF) [1,2]. However, this approach is quite challenging for identifying low-abundant targets in light of the important protein/peptide dynamic range (which spans 6 to 7 orders of magnitude) and diversity (both in terms of identity and post-translational changes) of these biological fluids [1,3]. At peptide level (i.e. after trypsin digestion of the samples) this diversity amounted to an important intrinsic variability among the different clinical samples and the most abundant peptides that co-eluted in the matrix suppressed the signal by inhibiting the ionization of less-abundant peptides [4]. Also, multiple interactions between analytes of interest, within the wide range of proteins present in the sample during its preparation may generate an intrinsic variability [5]. To reduce these problems, pre-fractionation of CSF samples is the method of choice to detect less-abundant proteins [6,7]. The most classic proteomic workflow is the one combining bidimensional gel electrophoresis (2DE) followed by MALDI-TOF-MS or LC-ESI-MS protein identification [8,9]. However, this approach has serious limits linked to the poor electrophoresis migration of low- and high-molecular weight proteins (>150 kDa and <15 kDa) or basic, acid or hydrophobic proteins [10]. Thanks to the improvement in sensitivity and specificity of mass spectrometers, the most commonly used approach nowadays for biomarker discovery, instead of 2DE, is a bottom up approach using targeted proteomics and pre-fractionation samples. Several pre-fractionation methods have already been tested on CSF samples, including bi-dimensional liquid chromatography.

*Corresponding authors: Sylvain Lehmann, CHU Montpellier, Institute of Research in Biotherapy, Hôpital Saint Eloi, Laboratory of Biochemistry and Clinical Proteomics CCBHM, Montpellier, INSERM U1040-UM1, Montpellier, F-34000 France, Tel: +33 603422349, E-mail: s-lehmann@chu-montpellier.fr
Christophe Hirtz, CHU Montpellier, Institute of Research in Biotherapy, Hôpital Saint Eloi, Laboratory of Biochemistry and Clinical Proteomics CCBHM, Montpellier, INSERM U1040-UM1, Montpellier, F-34000 France, Tel: +33 603422349, E-mail: c-hirtz@chu-montpellier.fr

Received March 09, 2014; Accepted April 17, 2014; Published April 21, 2014


Copyright: © 2014 Lehmann S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
techniques [11,12], low-molecular weight protein enrichment [13], binding to solid-phase libraries or depletion of the most abundant proteins such as albumin [14,15]. The latter is commonly used and relies on immunodepletion using 1, 6, 12 and 20 proteins depletion columns [16].

One important medical and Public Health challenge is the diagnosis and treatment of neurodegenerative diseases, in particular Alzheimer’s disease, as the worldwide population age 65 and older continues to increase. There is therefore a lot of proteomics research in this field, especially on CSF biomarkers [17,18]. This biological fluid of the Central Nervous System essential to the brain and spinal cord is generally collected by lumbar puncture and its components originate from blood, following active and passive transports in the choroid plexuses, and from the drainage of interstitial fluid from the nervous tissues [3]. This explains why CSF is a good “mirror” of the brain’s physiological and pathological status and is the object of many proteomics studies on disease-related biomarkers [2].

In this work, we tested three different well known pre-fractionation methods for CSF: immunodepletion of major proteins (Seppro® IgY14), hydrophobic solid phase extraction (Oasis® HLB), and lipophilic sorbent concentration (Liposorb™). The objective was to test their feasibility and capacity to access different subproteomes. Protein identification was performed using high-resolution Orbitrap™ mass spectrometry. We observed differences in terms of the number and nature of proteins identified under the different conditions. This element is relevant to select an appropriate workflow depending on the type of biomarker targeted.

Material and Methods

Equipment

Corning CentriStar™ Centrifuge Tubes: Corning® 15 mL PP Centrifuge Tubes, Rack Packed with CentriStar™ Cap, Sterile (Product #430790). 1.5 mL Protein LoBind tube, ref. 022431081 from Eppendorf (Le Pecq, France). HLB: Oasis® HLB µElution Plate, reference 186001828BA, Waters. Lipo: PHM-L LIPOSORB ABSORBENT, reference 524371-1, VWR. D14: Seppro® IgY14, reference SEP010, Sigma-Aldrich. C18 Sep-Pak: Sep-Pak Vac 1cc (100 mg) C18 cartridges, Waters, ref. VAT023590, CentriVap Acid-Resistant System from labconco (ref: 7983013). Analytical balance CPA224S-0CE from Sartorius. LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, USA) coupled with an Ultimate 3000 HPLC (Dionex, Amsterdam, Netherland).

Clinical samples

To compare the different methods without inducing a biological bias, we used a 1.5 mL CSF pool named SPE1. CSF samples were obtained from control patients (i.e. patient who had a lumbar puncture to investigate headaches or memory complaints but for whom the etiological research was negative). These samples had normal values for cytology, protein, glucose and amyloid peptide 1-42, tau and phospho-tau proteins. CSF was collected in polypropylene tubes, sent to the laboratory within 4 hours and was centrifuged at 900 g for 10 minutes at 4°C. CSF was aliquoted in proteins was determined with the BCA Assay (Pierce™) and 40 mM Iodoacetamide (Sigma-Aldrich) and 40 mM Dithiothreitol (Sigma-Aldrich) and digested with sequencing-grade trypsin (Promega). The concentration in proteins was determined with the BCA Assay (Pierce™). Classic in-solution tryptic digestion was performed. Samples were briefly digested with 1:50 w/w ratio. The mixes were incubated overnight at 37°C and the digestion was stopped by adding 15 µL of pure formic acid (pH<4).

The generated tryptic peptides were desalted using the C18 Sep-Pak column. Columns were conditioned twice with 200 µL of 70% acetonitrile/0.1% TFA; equilibrated twice with 0.1% TFA (300 µL).
Samples were loaded onto the C18 Sep-Pak column and washed three times with 200 µL of 0.1% TFA and eluted with 200 µL of 70% acetonitrile/0.1% TFA. Samples were dried on vacuum concentrator (Labconco, Kansas city, USA) and resuspended using 0.1% TFA before LC-MS/MS injection.

**Mass spectrometric analysis**

All MS analyses were performed in duplicates.

Generated peptides were analyzed online by nano-flow HPLC-nanoelectrospray ionization using a LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, USA) coupled with an Ultimate 3000 HPLC (Dionex, Amsterdam, Netherland). Desalting and pre-concentration of samples were performed on-line on a Pepmap™ precolumn (0.3 mm × 10 mm, Dionex). A gradient consisting of 0-40% B for 120 min and 80% B for 15 min (A=0.1% formic acid, 2% acetonitrile in water; B=0.1% formic acid in acetonitrile) at 300 nL/min was used to elute peptides from the reverse-phase capillary (0.075 mm × 150 mm) column (Pepmap™, Dionex), fitted with an uncoated silica PicoTip Emitter (New Objective, Woburn, USA). Spectra were recorded using the Xcalibur software (v 2.0.7, Thermo Fisher Scientific) and acquired in data-dependent acquisition mode throughout the HPLC gradient. The mass scanning range (m/z) was 400-2000 and capillary temperature was 200°C. Survey scans were acquired in the Orbitrap system with the resolution set at 60000. For all full-scan measurements with the Orbitrap detector a lock-mass ion from ambient air (m/z 445.120024) was used as internal calibrant as previously described [19]. Up to five of the most intense ions per cycle were fragmented and analyzed in the linear trap. Peptide fragmentation was conducted with nitrogen gas on the most abundant and at least doubly-charged ions detected in the initial MS Scan. Normalized collision energy of 35 eV and activation time of 30 ms was used for CID.

**Peptide identification**

All MS/MS spectra were searched against the Homo sapiens Complete Proteome Set database (70101 entries, release July 2012 http://www.uniprot.org/) by using the Proteome Discoverer software (v 1.2, Thermo Fisher Scientific) and the Mascot v 2.3 algorithm (Matrix Science, http://www.matrixscience.com/) with trypsin enzyme specificity and one trypsin missed cleavage. Carbamidomethylation of cysteine was set as a fixed modification. The search was also performed to allow the following variable modifications: Oxidation (M) and Deamidated (NQ). Mass tolerances in MS and MS/MS were set to 5 ppm and 0.5 Da respectively, and instrument settings were specified as “ESI-TRAP” for identification. Management and validation of mass spectrometry data were performed using the Proteome Discoverer software (Mascot significance threshold p<0.01, with a minimum of one peptide per protein). The protein identified by each pre-fractionation method was compared directly with the Proteome Discoverer software and more refined comparisons were performed, at protein level, by PSM (Peptide Spectral Match, corresponding to a sum of correct interpretation of single spectrum) comparison.

**Results and Discussion**

The goal of this study was to compare the relevance of 3 single pre-fractionation methods on CSF, which were originally designed to isolate low-abundant biomarkers. Such pre-analytical phases are generally necessary when working with complex biological fluids such as blood or CSF. In fact, by reducing the complexity and dynamic range of the proteins, pre-fractionation reduces the ion suppression phenomenon occurring in LC/MS analysis of complex samples. The complete experimental workflow of the different pre-fractionation methods is presented in Figure 1. CSF pools were used to cover the maximum proteome diversity and have enough material for all the experiments. To reach a good level of proteome coverage, we used the combination of a mass spectrometer with high mass accuracy and sensitivity (LTQ Orbitrap XL, Thermo Fisher Scientific) and a 120 min high-performance liquid chromatography gradient. The three pre-fractionation methods tested are well known and have intrinsic properties: immunodepletion is considered as one of the most effective approaches for the removal of highly-abundant proteins in complex matrix. In this study we used the Seppro® IgY14 [20] which removes the top 14 most abundant plasma proteins: HSA, IgG, Fibrinogen, Transferrin, IgA, IgM, Haptoglobin, alpha2-Macroglobulin, alpha1-Acid Glycoprotein, alpha1-Antitrypsin, Apo A-I, Apo A-II, Complement C3 and ApoB. The Solid Phase Extraction (SPE) [21] approach is commonly used for both cleanup and analyte enrichment. Here we used the Oasis® HLB µ Elution plates that contain a universal polymeric reversed-phase sorbent (Particle Size 30 µm, pore Size 80Å) for extraction of a wide range of compounds from various matrices [22-24]. The third pre-fractionation approach, PHM-L LIPOSORB™, a sorbent composed of polyhydroxymethylene substituted by fat oxidized alcohol, was originally designed to selectively remove lipids from plasma or serum. It can also be used to purify lipoproteins [25] with a binding capacity of 50 mg per g of Liposorb™.

It is interesting to note that the Oasis® HLB and Seppro® IgY14 procedures are quite similar in terms of technical difficulties, whereas Liposorb™ is easier and faster to use and involves less technical steps. Adding pre-analytical steps generates a risk of contaminating the samples. However, as validated by the percentage of keratin peptides identified, we did not observe any additional keratin contamination, in spite of sample manipulations in the pre-fractionation method. The contamination percentage ranged between 4 and 7% after
pre-fractionation, comparable to the percentage obtained from unfractonated CSF (not shown). One major issue when adding pre-analytical steps lies in introducing some variability into the experiments. To address this issue, we performed the pre-fractionation experiments in triplicates and analyzed each sample in duplicate by LC-MS/MS. Crude, unfractonated CSF was also included in all analyses. To standardize the amount of peptides injected on the LC, each sample was initially analyzed in full-scan acquisition mode. We adapted the volume of loaded samples (ranging from 1 to 6.25 μL) to obtain chromatograms with a relative intensity similar to the Total Ion Chromatogram (TIC) in order to conduct a comparison of all chromatograms across the different methods. The main evaluation criterion was the number of proteins identified with at least one peptide and 99% confidence with Mascot (Table 1). The MS reproducibility was estimated by the coefficient of variation (CV) between the two LC/MS runs. This CV ranged in average between 1 and 5% which was expected for this type of analyzers. Then, the crucial point was for the different methods, to estimate the reproducibility of the entire workflow and evaluate the capacity of these methods to identify proteins. To be more confident with the results, for each experimental condition we only computed the proteins identified by the two LC/MS runs (table 1). Over the three experiments, the reproducibility in terms of number of protein identified in unfractonated CSF was 1.2% vs. 13.2%, 3% and 4% respectively for the IgY14, SPE HLB and Liposorb™ methods. This result was in line with other studies on CSF were the variability was as high as 20% [26]. We then compared the total number of proteins identified by the different methods (Table 1 and figure 2A). Compared with unfractonated CSF and in terms of number of proteins identified, we only observed a significant difference for the OASIS® HLB and Liposorb™ methods with 28% more and 48% fewer proteins identified respectively, the other method showed results similar to those obtain with unfractonated CSF. Surprisingly, the total number of proteins identified using the IgY14 method was not significantly better than proteins identified with unfractonated CSF. One explanation might be the removal of several proteins associated with the 14 immunocaptured proteins [16,27]. Interestingly, when the number of PSMs (Peptide Spectral Match, corresponding to a sum of correct interpretation of single spectrum) was compared for the different methods (Figure 2B), both OASIS® HLB and Liposorb™ resulted in significantly higher values than for unfractonated CSF. The apparent discrepancy between the number of protein identifications (IDs) and PSMs for Liposorb™ might be explained by the fact that only a small sub-proteome was retained and consequently a better sequencing occurred by the MS system. For this method, the identified proteins have shown a higher sequence coverage than the other preparations.

To investigate if the different pre-fractionation methods target specific protein subsets, we first plotted the proteins IDs and PSMs according to their molecular weight (Figure 3). Overall, the distributions were very similar, except that Liposorb™ apparently promoted the identification of low- vs. high-molecular weight proteins. PSM distribution also revealed a lower number of identification, in the 60-100 kDa range, for the IgY14 method. As a matter of fact this difference could be explained by albumin-related decrease in PSMs [28], since albumin is a protein specifically removed by immunocapture. We also looked at the distribution of protein IDs depending on their isolectric points, but no differences were unveiled between the different methods (not shown).

To further investigate the differences between pre-fractionation methods, we performed a Venn diagram of the identified proteins (Figure 4A). This distribution revealed that each method (including unfractonated CSF) had a different profile with 35, 18, 71 and 18 proteins respectively for unfractonated CSF, IgY14, SPE HLB and Liposorb™. Interestingly, only 46 proteins were common to all methods. This difference in protein profiles was also apparent when using hierarchical clustering (Figure 4B) which differentiated unfractonated CSF from

---

**Table 1:** Number of protein identifications (IDs) by LC-MS/MS. Protein IDs in each pre-fractionation scheme were obtained with Proteome Discoverer and a Mascot significance threshold p<0.01, with a minimum of one peptide per protein.

<table>
<thead>
<tr>
<th>ID of proteins</th>
<th>Replicate</th>
<th>LC-MS run #1</th>
<th>LC-MS run #2</th>
<th>Run #1 and 2 common IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfractionated CSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1</td>
<td>131</td>
<td>125</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>126</td>
<td>127</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>125</td>
<td>124</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>IgY14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1</td>
<td>95</td>
<td>83</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>106</td>
<td>113</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>117</td>
<td>113</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>SPE HLB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1</td>
<td>158</td>
<td>159</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>161</td>
<td>168</td>
<td>143</td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>165</td>
<td>162</td>
<td>144</td>
<td></td>
</tr>
<tr>
<td>Liposorb™</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1</td>
<td>65</td>
<td>69</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>76</td>
<td>69</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>65</td>
<td>65</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3: Distribution of the identified proteins according to their molecular weight. The number of protein in each situation (Seppro® IgY14, HLB Oasis®, PHM-L LIPOSORB™, unfractionated CSF) was reported according to the number of identifications and number of PSMs. For IgY14, the lower number of PSMs, in the 60-100 kDa range, corresponded to an albumin-related decrease in PSMs.

Figure 4: A: Venn diagram with the number of unique proteins identified in the different conditions (Seppro® IgY14, HLB Oasis®, PHM-L LIPOSORB™ and unfractionated CSF). Note that the number of proteins detected by only one method was as high as the number of proteins detected by several methods. 4B: Hierarchical clustering according to the number of unique protein identifications for each pre-² method (Seppro® IgY14, HLB Oasis®, PHM-L LIPOSORB™) and for unfractionated CSF. Unfractionated CSF is well separated from the different pre-fractionations, while IgY14 and PHM-L LIPOSORB™ were closer together.
**Pre-Fractionation Methods**

With SPE HLB being apart from the others, we understand the differences between these methods we looked at the top ten proteins that had the highest number of PSMs (Figure 5). In unfractinated CSF, albumin, transferrin, immunoglobulin and complement were present in the top list as expected since these proteins are the ones present in higher concentrations in this biological fluid [29]. With the IgY14 pre-fractionation method, one would expect to have the 14 immunocaptured proteins removed from this list yet we observed that albumin, serotransferrin or immunoglobulin were still present but not at the top of the list. This could be explained by the fact that these proteins are present in very high concentrations in CSF, and since IgY14 immunocapture is not optimized for CSF, the fact that these proteins are present in higher concentrations [29]. We also reported it is not 100% efficient and might not be able to remove all protein isoforms (e.g. truncated) eventually detectable in MS. We also reported the greatest number of PSMs for the OASIS® HLB method which is a reversed-phase sorbent specifically developed for the purification of a wide range of small-size acid, basic and neutral compounds. The SPE HLB showed a decrease of the most abundant proteins identifications which is very interesting for in-depth analysis. Although, we observed only a partial depletion of these abundant proteins, which may also be considered as a normalization of the sample, because it reduced the dynamic range of proteins (the most abundant ones are the less represented). Interestingly, the depletion IgY14 kit also showed this phenomenon. In any case, we observed that albumin or transferrin were still at the top of the list, but at the same time so was secretogranin, a protein usually found in low concentrations in CSF (10⁻⁶ g/L). The top list for the Liposorb™ pre-fractionation method was also very interesting. Albumin was still in the first place, but as expected [30] the list validated a real enrichment in apolipoproteins. Notably, the fact that the PHM-L LIPOSORB™ permits to recover preferentially a protein like Apolipoprotein J, which is involved specifically in AD [31] and that Seppro® IgY14 enables to retrieve Angiotensinogen which is a biomarker of multiple sclerosis [32] illustrates that one of the pre-fractionation method might be better suited depending on the focus of the research.

**Conclusion**

Many elements can be deducted from comparing these different pre-fractionation methods. First of all, and as reported before, these approaches do not add a significant variability and are compatible with biomarker-discovery programs. OASIS® HLB was identified as the most efficient method to increase the number of protein identifications. IgY14 probably suffers from a co-depletion phenomenon that reduced its performance. The Liposorb™ approach was identified as the method with the fewer number of proteins identified. However, it presented a real enrichment of specific protein (lipoproteins). This result makes sense as this method works by capturing a subset of proteins instead of depleting them. One very important observation was also that

---

**Figure 5**: Pie chart of the top 10 proteins with the highest number of PSMs under each condition (Seppro® IgY14, HLB Oasis®, PHM-L LIPOSORB™ and unfractinated CSF). Note that albumin was at the top of the list for all conditions but IgY14. The LIPOSORB™ distribution was remarkable due to the presence of lipoproteins in the top 10, while the HLB Oasis® distribution detected a protein usually found in low concentrations in CSF (secretogranin).
Acknowledgments

This work was in part supported by the “Programme hospitalier de recherche Clinique” (PHRC) ProMarA: Use of targeted quantitative proteomics and metabolic labelling with stable isotopes for the diagnosis and the investigation of neurological disorders and in particular Alzheimer’s disease” and the European FP7 Joint Programming on Neurodegenerative Diseases (JPND) BiomarkAPD. The authors would like to specially thank Bénédicte Clément for editing the manuscript.

References


