Journal of Chromatography A, 1445 (2016) 27–35

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Impact of injection solvent composition on protein identification in column-switching chip-liquid chromatography/mass spectrometry

V. Houbart[a], G. Cobraiville[a], G. Nys[a], M-P. Merville[b], M. Fillet[a,*]

[a] Laboratory for the Analysis of Medicines, Department of Pharmacy, CIRM, University of Liege, Belgium
[b] Department of Clinical Chemistry, CIRM, CHU, University of Liege, Belgium

ARTICLE INFO

Article history:
Received 24 December 2015
Received in revised form 19 March 2016
Accepted 21 March 2016
Available online 24 March 2016

Keywords:
Injection solvent
RP-LC–MS
Protein identification
Sensitivity

ABSTRACT

In shotgun proteomics, the gold standard technique is reversed-phase liquid chromatography coupled to mass spectrometry. Many researches have been carried out to study the effects on identification performances of chromatographic parameters such as the stationary phase and column dimensions, mobile phase composition and flow rate, as well as the gradient slope and length. However, little attention is usually paid to the injection solvent composition.

In this study, we investigated the effect of the injection solvent on protein identification parameters (number of distinct peptides, amino acid coverage and MS/MS search score) as well as sensitivity. Tryptic peptides from six different proteins, covering a wide range of physicochemical properties, were employed as training set. Design of experiments was employed as a tool to highlight the factors related to the composition of the injection solvent that significantly influenced the obtained results. Optimal results for the training set were applied to analysis of more complex samples. The experiments pointed out optimising the composition of the injection solvent had a strong beneficial effect on all the considered responses. On the basis of these results, an approach to determine optimal conditions was proposed to maximise the protein identification performances and detection sensitivity.

1. Introduction

Mass spectrometry-based proteomics has turned out to be increasingly popular over the past decade and has become the strategy of choice for the study of the proteins in biological systems [1–4]. Among the different approaches for protein identification, shotgun proteomics consists of using a site-specific proteolytic enzyme, commonly trypsin, to digest the proteins and analysing the resulting peptide mixture by a separation technique (*e.g.* liquid chromatography) coupled to a mass spectrometer. Finally, the raw experimental data are computed using powerful bioinformatics tools.

In shotgun proteomics, the gold standard separation technique is reversed-phase liquid chromatography (RP-LC), due to its high peak capacity and compatibility with mass spectrometry (MS) [5]. Nanoflow LC is the chromatographic system of choice for protein identification since chromatographic performances improved with the square of the reduction of the inner diameter of the column [6]. In the same time, MS ionisation performances are also increased since lower flow rates produce smaller droplets in the MS source that improves desolvation efficiency, providing a higher amount of ions available for detection in the gas phase [7–9]. The efficiency of the chromatographic separation is a determinant parameter in the quality of the data obtained from a set of experiments. Therefore, many researches have been carried out studying the influence of chromatographic parameters such as the stationary phase and column dimensions [10–12], mobile phase composition and flow rate [13,14], and the gradient slope and length [14,15] on identification performances. However, little attention is usually paid to the sample dissolution medium, also called the injection solvent. During analytical method development, peptides are still frequently considered in the same way as small molecules, although differences between their chromatographic behaviour have already been described in the eighties [16].

Previously, our group demonstrated that peptide analysis in RP-LC coupled to mass spectrometry was strongly influenced by the composition of the injection solvent in terms of sensitivity, carry-over and chromatographic behaviour [17,18]. In the present study, the effect of injection solvent composition was studied on

Abbreviations: AA, amino acid; DoE, design of experiments; FA, formic acid; FFD, full factorial design; TFA, trifluoroacetic acid.

* Corresponding author at: Laboratory for the Analysis of Medicines, LAM, Department of Pharmacy, CIRM, University of Liege, CHU, B36, B-4000 Liege, Belgium.
E-mail address: marianne.fillet@ulg.ac.be (M. Fillet).

http://dx.doi.org/10.1016/j.chroma.2016.03.056
0021-9673/

protein identification performances (number of distinct peptides, amino acid coverage, MS/MS search score and mean peptide spectral intensity) using a training set consisting of a tryptic digest from six proteins. The chromatographic separation was performed on a nanoflow LC-chip system that integrates in one component an enrichment column and an analytical column containing the same C18 stationary phase. The injection solvents studied are widely employed for peptide analysis by RP-LC–MS, *i.e.* acetonitrile (ACN) and water as solvents, and formic acid (FA) and trifluoroacetic acid (TFA) as ion-pairing agents. Design of experiments (DoE) was used as a tool to highlight the factors, related to the injection solvent composition (presence/absence of ACN, presence/absence of ion pairing reagent at different percentage), that significantly influence protein identification performances. An optimal composition was determined and significant parameters are discussed in regard to their impact on the protein identification efficiency of the training set. Finally, the proposed condition was challenged with the analysis of a very complex sample, *i.e.* the *Escherichia Coli* proteome.

2. Experimental

2.1. Chemicals and standards

Water, acetonitrile (ACN) and formic acid (FA) 99% (all LC/MS grade) were obtained from Biosolve (Valkenswaard, Netherlands). Analytical grade trifluoroacetic acid (TFA) was purchased from Fluka (St. Louis, MO, USA). Helium and nitrogen (Alphagaz 1) were obtained from Air Liquide (Milmort, Belgium). MassPREP digestion standards, containing bovine serum albumin (BSA), hemoglobin subunit alpha (HBA), enolase (ENO-1), alcohol dehydrogenase (ADH-1), hemoglobin subunit beta (HBB) and phosphorylase b (PYGM), as well as tryptic digest of *E. Coli*, were purchased from Waters (Milford, United States).

2.2. Instruments

Chromatographic separation was achieved on a 1200 series LC-chip system consisting of a nanoflow pump, a capillary pump, a wellplate sampler and a LC-chip/MS interface (Agilent Technologies, Waldbronn, Germany). Chromatographic separation was performed on a chip including a 160 nL trapping column and a 150 mm × 75 μm analytical column, both packed with a Zorbax 300SB 5 μm C_{18} phase (Agilent Technologies). The mobile phase was composed of H_2O/FA (100:0.1, v/v) (A) and ACN/H_2O/FA (90:10:0.1, v/v/v) (B) degassed by ultrasonication for 15 min before use. Analytical process was performed in two steps: first, the sample was loaded on the trapping column during an isocratic enrichment phase using the capillary pump delivering a mobile phase in isocratic mode composed of H_2O/ACN/FA (97:3:0.1, v/v/v) at a flow rate of 4 μL/min. A flush volume of 6 μL was used to remove unretained components. Then, after valve switching, a gradient elution phase in backflush mode was performed through the enrichment and analytical columns using the nanopump. The analysis was performed using a gradient starting at 3% B that linearly ramped up to 45% B in 30 min at a flow rate of 300 nL/min; then up to 95% B in 5 min. Column was then rinsed with 95% B during 5 min before returning to 3% B. Ten column volumes were used for reequilibration prior to the next injection. The total analysis time was 43 min for each run. All the experiments were carried out with a 1 μL sample injection volume. During the analysis, the injection needle was thoroughly rinsed three times from the inside and the outside with a mix of ACN/H_2O/TFA (60:40:0.1, v/v/v) commanded by an injection program set in the injector parameters.

The identifications were performed with electrospray ionisation MS/MS, using a 6300 series ion trap mass spectrometer (Agilent Technologies). The collision energy was set automatically depending on the mass of the precursor ion. Each MS full scan was followed by MS/MS scans of the six most intense precursor ions detected in the MS scan (exclusion time: 1 min). The results were subsequently introduced into the database for protein identification searches using Spectrum Mill (Agilent Technologies).

2.3. Preparation of solutions

Digestion protein standards were dissolved in 0.1% FA to reach a concentration of 1 pmol/μl. The different solutions were separately aliquoted and stored at −80 °C. Depending of the experimental conditions, dilutions were made with ACN, water and an ion-pairing reagent in the appropriate proportions to reach a concentration of 10 fmol/μl. According to the supplier's documentation, the digested protein solutions were stable at least for one week at 4 °C.

In some experiments digestion standards of the six proteins were spiked to reach a final concentration of 10 fmol/μl to an *E. Coli* digest at 50, 100 and 200 ng/μl.

Sample handling and storage were made in Protein LoBind polypropylene tubes (Eppendorf, Hamburg, Germany) using Maxymum Recovery pipette tips (Axygen Scientific, Union City, United States).

2.4. Design of experiments

Response surface methodology (RSM) is a multivariate approach for optimisation that establishes a relationship between selected factors and chosen response(s) through a second-degree polynomial relationship, and that allows to find optimal values for the selected factors to minimise, maximise or reach a target value for the response(s) [19].

RSM may be achieved using a three-level full factorial design. This kind of design is useful when the number of factors is low ($k < 3$)[20], since the experimental number N grows rapidly with the number of factors. The final result is a response surface, which is a three-dimensional view that provides a graphical representation of the relationship between responses and variables. If the regression model displays only main effects (*i.e.* first order model), the surface response will be planar. However, if the model contains interaction and/or quadratic effects, contour lines will be curved.

A full factorial design of experiments (FFD) was created to determine the optimal injection solvent for protein identification. Three factors were considered: the nature and proportion of the ion-pairing agent, and the proportion of ACN. The FFD defined 12 experimental conditions that were carried out twice, and the central point was performed in quadruplicate, leading to a total of 28 experiments.

To study the identification performances provided by the different injection solvents, four parameters to maximise were monitored: the number of distinct peptides (number of distinct peptides detected for each protein, with each peptide counted once, regardless of the charge state or number of acquired spectra for the same peptide), the distinct summed MS/MS search scores (total score for all the distinct peptides), the percentage of amino acid coverage (percentage of amino acids in the protein hit covered by the spectral data) and the mean peptide spectral intensity (mean intensity of all the peptides assigned to a protein, calculated from extracted ion chromatograms from the precursor ions).

The raw MS data were treated with Spectrum Mill. Spectrum Mill Workbench software (version A.03, Agilent Technologies, Santa Clara, CA, USA) was used for data processing. Peptides and proteins were identified using parameters as follows: fixed modifications: carbamidomethylation of cysteine; scans with the same precursor mass and spectral similarity were merged within tolerances (retention time ± 15 s, mass ± 1.4 m/z); precur-