Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 824 (2005) 189–200
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Optimization of protein identification from digests as analyzed by capillary isoelectric focusing-mass spectrometry

Henricus F. Storms *, Robert van der Heijden, Ubbo R. Tjaden, Jan van der Greef

Leiden University, Amsterdam Center for Drug Research, Division of Analytical Biosciences, P.O. Box 9502, 2300 RA Leiden, The Netherlands

Received 29 April 2005; accepted 3 July 2005
Available online 3 August 2005

Abstract

Capillary isoelectric focusing (CIEF) is a high-resolution separation technique for the analysis of peptides and protein digests. When coupled to ion trap-mass spectrometry (CIEF-MS) the unique separation mechanism is combined with a highly efficient detection system. In an earlier report, we described aspects of separation and interfacing in connection to the analysis of a digest of set of standard proteins. Now, we report on different aspects of the process of protein identification. Sequest software parameters were optimized by using a standard protein digest. These settings were used for the analysis of periplasmic proteins from *Escherichia coli*. Since in CIEF peptides are focused according to their p*I* values, the mobilization time of a particular peptide is dependent on its p*I* value. Based on this relation, the identification of some peptides was facilitated. Furthermore, the Sequest settings that were used could be evaluated. In total, 159 proteins were identified in a single run.

Keywords: Capillary isoelectric focusing; Mass spectrometry; Protein analysis; Data analysis; Sequest

1. Introduction

For the understanding of cellular processes, the study of the proteomes has taken an increasingly important role. The use of the so-called shotgun approach has become increasingly popular in the last decade. In this approach, the protein extract is digested at the beginning of the work flow, and the resulting peptides are then separated by liquid chromatography (LC) or capillary electrophoresis (CE) [1,2]. MS/MS analysis is used for the identification of the individual peptides and subsequently the corresponding proteins.

Earlier, we have reported on the use of capillary isoelectric focusing as a separation technique for the shotgun approach. CIEF-MS of complex peptide mixtures was performed without the use of carrier ampholytes, and with the use of low amounts of carrier ampholytes as mere spacers [3]. CIEF is a high-resolution separation technique that can be applied for amphoteric compounds, such as proteins or peptides. These are separated according to their p*I*-values, in a pH gradient formed under the influence of an electric field [4].

Though normally carrier ampholytes are added to the sample in order to establish a linear pH gradient, in the case of peptides this is not needed for focusing to occur; this process is called autofocusing [5–7]. It should be noted though that the peptides lack the high buffering capacity of the carrier ampholytes, but compared to carrier ampholytes the peptides are fully compatible with mass spectrometry.

The described approach, despite a rather limited resolution, allowed the identification of the eight components of a mixture of standard proteins. The addition of a relatively low concentration of carrier ampholytes (0.2%) resulted in increased separation efficiency, although already ion suppression was observed. Furthermore, the use of higher sample concentrations also resulted in improved separation efficiency.

* Corresponding author.
E-mail address: harriestorms@zonnet.nl (H.F. Storms).

1570-0232/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2005.07.020

In continuation of that research we have studied several aspects of data acquisition, data processing and protein identification aiming to obtain clean data, in which the number of correctly identified proteins is maximized while the need for manual checking for false positives is minimized.

In this study, a 10-protein mix was analyzed using CIEF followed by linear ion trap mass spectrometry, which has both superior sensitivity and a higher scanning speed as compared to the conventional 3D ion trap [8]. For identification, Sequest software was used. Sequest software compares the acquired MS/MS spectra with theoretical MS/MS spectra of (tryptic) peptides; these theoretical spectra are based on their sequences which are taken up in a database.

Several earlier studies have made it clear that the parameters and settings used in this software are very important for distinguishing true hits from false positives. Various improvements have been suggested, like the use of discriminant function analysis for optimization of Sequest parameters [9,10] and the use of a machine-learning algorithm [11].

In general, what these studies have made clear is that several parameters contribute significantly to the discrimination of correct and incorrect peptide assignments. The most important one is the cross correlation value (Xcor), which is a correlation coefficient for the match. A minimum threshold for this parameter is commonly being used. Link et al. [14] used 1.5 for singly charged peptides, and 2.0 for multiply charged peptides. Some later reports have used higher thresholds for the Xcor, and on top of that also used a threshold of 0.1 for the ΔC_N delta correlation value (ΔC_n), which is a measure for the difference in the correlation value between a particular hit and the first following hit [15,16].

Other parameters that can be taken into account are the preliminary score based on the number of ions in the MS/MS spectrum that match with the experimental data (Sp), the ranking of the peptide match in the resulting preliminary scoring list (RSp) and finally the coverage of *y*- and *b*-ions. Anderson et al. [11] have shown that these parameters can also contribute in the discrimination of correct and incorrect peptide assignments. Furthermore, they showed that all parameters, including the Xcor, are dependent on the type of mass spectrometer used, resulting from differences in accuracy and amount of noise in the MS/MS spectra. Taking this into account, it is expected that protein identification will benefit from optimization of all mentioned parameters.

When CIEF is performed with protein digests, ordering according to the p*I* values of the peptides is expected. This means that the mobilization times of the peptides are dependent on their p*I* values. Thus, the use of CIEF provides an extra criterion for the identification of the individual peptides. In this study, it is investigated whether this could be used for data analysis.

Finally, the optimized procedures were applied to the analysis of a biological sample, the periplasmatic proteins from *Escherichia coli* cells.

2. Materials and methods

2.1. Chemicals

DTT and iodoacetamide were purchased from Sigma–Aldrich Chemie (Steinheim, Germany). Myoglobin from horse heart, bovine serum albumin, cytochrome *C* from horse heart, human insulin, human serum albumin, carbonic anhydrase I from human erythrocytes, lactoglobulin B from bovine milk, bovine ribonuclease A, lysozyme C from chicken egg and ovalbumin from chicken egg were all purchased from Sigma–Aldrich Chemie (Steinheim, Germany).

2.2. Standard protein mix

Proteins were digested at a concentration of 1 mg/mL with sequencing grade trypsin (Roche Diagnostics Boehringer Mannheim B.V, Mannheim, Germany) according to Matsudaira [12]. The proteins were solved in 50 mM NH_4HCO_3, treated with DTT (2.25 mM) for the reduction of disulfide bonds and were then carboxymethylated with iodoacetamide (5.0 mM) to prevent reoxidation. Digestion was performed by incubating the proteins for 24 h at 37 °C with trypsin (at a 1:30 enzyme:protein ratio).

2.3. Periplasmatic protein extract

E. coli cells (K12 strain) were grown in LB medium at 37 °C. Periplasmatic proteins were isolated by osmotic shock [13]. Harvested cells (20 mL cell suspension) were centrifuged at 1000 × *g* for 10 min and suspended in 50 mM Tris/HCl, pH 7.5, containing 0.5 M sucrose. This was repeated twice, but the second time the cells were suspended in 50 mM Tris/HCl, pH 7.5, containing 0.5 M sucrose and 1 mM EDTA. After equilibration for 20 min cells were centrifuged again and suspended in 0.20 mL of 50 mM Tris/HCl, pH 7.5, containing 0.5 M sucrose and 1 mM EDTA. Osmotic shock is performed by suspending the cells in 5 mL water at 4 °C. The resulting suspension was centrifuged at 10,000 × *g* and the supernatant was recovered as the periplasmatic fraction. The protein concentration was estimated according to Bradford [14].

By ultrafiltration (Microcon centrifugal filter devices, MWCO 10 kDa, Millipore, Amsterdam, the Netherlands) the protein extract of *E. coli* cells was washed and the buffer was replaced with digestion buffer (50 mM NH_4HCO_3). The lysate was treated with DTT and iodoacetamide, and digested with sequencing grade trypsin (Roche Diagnostics Boehringer Mannheim B.V., Mannheim, Germany) according to Matsudaira [12], as described for the 10-protein mix.

2.4. CIEF

Fused-silica capillaries (75 μm i.d., 375 μm o.d.) were obtained from Bester (Amstelveen, the Netherlands). In order to prevent adsorption at the capillary walls and to reduce the