Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1080 (2005) 117–123
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Fast on-column protein digestion with subsequent peptide mapping using tandem mass spectrometry with information dependent acquisition

Martin Hedström, Maria Andersson, Igor Yu. Galaev, Bo Mattiasson *

Department of Biotechnology, Center for Chemistry and Chemical Engineering, Lund University, P.O. Box 124, SE-221 00 Lund, Sweden

Received 13 December 2004; received in revised form 18 April 2005; accepted 26 April 2005

Abstract

A platform for rapid on-line protein digestion of protein mixtures for direct infusion to a mass spectrometer is presented. A mixture of protein A, staphylococcal enterotoxin B and cytochrome *c* was used as a model mixture injected on a gel filtration column and a trypsin reactor which were connected in series to a micro liquid chromatography (μLC) system. The peptides in the column eluate were analyzed with ESI tandem mass spectrometry, utilizing information dependent acquisition (IDA). In one step, the proteins in the mixture (μM concentrations) were concomitantly desalted, separated, digested and identified with an overall analysis time of less than 40 min. Protein sequence coverage of 78–95% for the involved substances was achieved.

Keywords: Mass spectrometry; q-TOF; On-line digestion; Trypsin; Sequence coverage; Protein characterization

1. Introduction

In the field of biotechnology, fast and generic methods for the initial identification and characterization of key proteins are of great importance. The remarkable progress within mass spectrometric (MS) analysis during the last decades has solved many critical problems concerning this task [1,2]. The method most frequently used for protein identification and characterization is based on retrieving the amino acid sequence subsequent to trypsin digestion of the highly purified target protein. The generated tryptic peptides are typically separated via liquid chromatography or capillary electrophoresis prior to the detection using tandem mass spectrometry. However, for conventional protein digestion methods (e.g. in-solution-, in-gel digestion) limitations are posed by time-consuming sample pre-treatments (i.e. incubation times typically between 4 and 16 h). For the existing on-line digestion systems, the set-up arrangements are often complex with crucial intermediate unit operations (e.g. buffer exchange, solid phase extraction, RP chromatography). This usually results in properly separated peptides, whereas, the overall analysis time tends to exceed the requirements imposed by high-throughput methods. Indeed, several successful on-line techniques based on mass spectrometric detection for the construction of peptide maps from pre-treated pure protein substances have been reported [3–6]. Also tryptic peptide mapping using capillary electrophoresis (CE) has become a well established technique [7–11] where the combination CE-MS has been shown to be a rapid and powerful tool for the separation and detection of pre-digested protein samples [12].

Typically, in the field of biotechnology, a target protein is isolated through a series of down-stream processing steps. In the later part of this route, the qualitative determination of contaminating proteins is of great importance, especially for pharmaceutically related products where high demands are applied to the quality control of the product as well as to the manufacturing process utilized. In fact, the access to a digestion method applicable to complex sample preparations that concomitantly has intrinsically high sequence coverage would make it possible not only to detect the presence of contaminants but also to verify the location of minor changes in the target protein primary structure (e.g. mutations or post-

* Corresponding author. Tel.: +46 46 2228264; fax: +46 46 2224713.
E-mail address: bo.mattiasson@biotek.lu.se (B. Mattiasson).

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.04.069

translational modifications). Hence, a further addition to the conceptual requirements stated for good manufacturing practice (GMP) could be achieved.

The increasing demand for pure antibody preparations in large amounts for clinical applications [13] have made protein A (PA) from *Staphylococcus aureus* a desirable product to the biotechnology industry dealing with protein purification. The specific affinity to the Fc-part of immunoglobulin G molecules makes PA an excellent tool for affinity chromatography. Staphylococcal enterotoxin B (SEB), a key contaminant in PA purification processes, is an extracellular protein (28.4 kDa) classified as a superantigen due to the mitogenic activity exerted on T-cells via a ternary molecular complex formed with the major histocompatibility complex class II molecule (MHC II) on healthy eucaryotic cell surfaces [14,15]. The resulting T-cell proliferation leads to a massive cytokine release with epithelial damage, capillary leakage and hypotension. The clinical importance related to a number of induced syndromes in humans, including food poisoning and toxic shock syndrome have made SEB one of the more well studied proteinaceous agents over the recent decades. In this study SEB was chosen because of its notable tightly folded protein structure and thus inherent resistance to protease degradation [16]. Cytochrome *c* was included in the sample to add some complexity to the mixture and was chosen because of its well-characterized nature.

The separation of proteins according to their size in parallel with an efficient desalting step could be an interesting approach when aiming for a generic on-line digestion method. Furthermore, if stable proteins with an intrinsic resistance to proteolytic degradation (e.g. SEB) should be subjected to the digestion system, typical pre-treatment procedures using high concentrations of chaotropic agents (e.g. urea, guanidine) would be necessary. The problem of removing these low-molecular weight substances in order to avoid suppression of the MS signal could then be equally solved with the gel filtration unit.

A system for rapid on-line protein digestion of protein mixtures for direct infusion to a ESI mass spectrometer is presented. A size-exclusion chromatography (SEC) unit and a protease reactor were coupled in series and a joint flow of acetonitrile containing 1% (v/v) formic acid was introduced post-column. A mixture of PA, SEB and cytochrome *c* was used as a model protein mixture.

2. Experimental

2.1. Chemicals

Trypsin (Novo 6.0 S, 1250 tryp-USP/mg), SEB and PA (*S. aureus*) were kind gifts from Novozymes Biopharma AB (Lund, Sweden). Cytochrome *c* (horse heart) was purchased from Sigma (St. Louis, USA). Eupergit C (particle size 150 μm) was from Rohm Pharma (Darmstadt, Germany) and Sephadex G-75 was purchased from Amersham Biosciences (Uppsala, Sweden). Water and other chemicals were of analytical grade.

2.2. Preparation of trypsin and gel filtration columns

Trypsin (17500 tryp-USP units) was dissolved in a suspension of 0.1 g (d.w.) of Eupergit C in 10 ml 1 M potassium phosphate buffer (PPB), pH 7.2 and the mixture was gently degassed for 45 s and incubated on an end-to-end turntable for 48 h (25 °C) [17,18]. Thereafter, 5% (v/v) β-mercaptoethanol was added to the suspension to block any unreacted oxirane groups on the Eupergit C matrix [19]. The solution was then incubated for another 16 h under otherwise the same conditions and thereafter washed twice with 100 mM NH_4HCO_3, pH 8.0. The protease-matrix was then packed in HPLC type columns (100 mm × 1 mm, 150 mm × 2 mm) at a maximum pressure of 50 bar for 10 min using a Haskel (Burbank, USA) DSTV-100 high-pressure air driven liquid pump with 100 mM NH_4HCO_3, pH 8.0 as packing buffer. Concomitantly, pre-swollen Sephadex G-75 matrix was low-pressure packed (i.e. manually) in a HPLC-column (100 mm × 4.6 mm).

2.3. On-line system

The SEC and protease columns were connected in series in a μLC system from Perkin-Elmer (Boston, USA) according to Fig. 1. To maximize trypsin activity the temperature over the enzyme reactor was controlled using a water bath. The flow rate of digestion buffer through the column system (100 mM NH_4HCO_3, pH 8.0) was held constant and before the entrance to the TurboIonSpray™ source on the QSTAR® pulsar-*i*-Q-TOF mass spectrometer (PE Sciex, Toronto, Canada), a joint flow (i.e. continuously held at 1:1 ratio with fragmentation buffer) of acetonitrile containing 1% (v/v) formic acid was introduced via a Harvard syringe pump, controlled by the MS software. The TurboIonSpray™ source was set to positive ion mode with a source voltage of +5500 V. The quadrupole system was adjusted to scan between *m*/*z* 500 and 2500 in TOF-MS mode whereas for product ion mode (i.e. MS/MS) a range of *m*/*z* 50–2000 was chosen. The *m*/*z* value of individual precursor ions was automatically selected in the information dependent acquisition (IDA) software feature for fragmentation and the ions were collided under argon pressure using rolling collision energies ranging from 12 to 60 eV (i.e. collision induced dissociation, CID).

2.4. Sample preparations

Pure PA in 100 mM NH_4HCO_3, pH 8.0 with no prior treatment steps was used for maximizing the digestion power of the protease column. For the SEC/protease column system, a sample containing PA, SEB and cytochrome *c* (0.33 mg/ml of each protein in 100 mM NH_4HCO_3, pH 8.0) was processed by adding guanidine–HCl and β-mercaptoethanol to the final concentrations of 8 M and 0.2 mM, respectively. In addition,

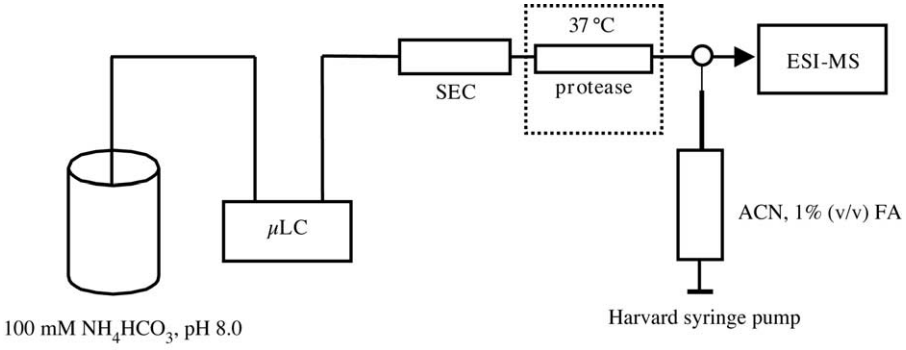

Fig. 1. On-column digestion system set-up. SEC (Sephadex G-75) and protease columns were coupled in series through a μLC system. A temperature of 37 °C was maintained over the protease reactor. Before the entrance to the ESI-Q-TOF MS, a flow (1:1 ratio with fragmentation buffer) of acetonitrile containing 1% (v/v) formic acid was added to enhance ionization/visualization of the eluted peptides.

before injection, the protein mixture was incubated 10 min in a water bath set to 75 °C.

2.5. Peptide processing

The IDA feature of the QSTAR® software (Analyst QS) was used for the automated retrieval of MS/MS spectra of eluted peptides through the formulation of criteria stipulated by the user. By fine-tuning of the involved IDA parameters in concomitant use of a well-defined exclusion list, where peaks related to background noise were filtered off, it was possible to achieve high yields in target-specific peak processing. The fragmentation spectra were investigated through the MASCOT "MS/MS Ions Search" database (www.matrixscience.com). The time-scale, however, where MS/MS processing of the peptide map of a certain protein is possible will be dependent both on the amount of injected analyte and on the flow rate of digestion buffer, i.e. analyte residence time in the column. Also the resolution power of the gel filtration column is dependent on these parameters, which implies that an optimisation could be necessary for any new protein system applied.

3. Results and discussion

Given the value of having a column digestion system for high-throughput analysis of protein samples, the benefits of a continuous on-line system, requiring no intermediate off-line trapping or desalting steps, are several. Indeed, the flexibility of such a method regarding applicability to different target analytes increases due to the relatively few parameters in need of optimization (i.e. flow rate, injection volume and SEC matrix type).

To investigate the efficiency of the enzyme reactor for untreated protein substances, pure PA (2 μl, 7.1 μM) was applied (column dimensions 100 mm × 1 mm). Factors that generally affect the catalytic efficacy (temperature, substrate contact time) were optimised to achieve the most complete digestion (i.e. maximal sequence coverage). A temperature of 37 °C over the protease column was chosen whereas flow rates between 10 and 30 μl/min of the running buffer were investigated, with the lowest value resulting in the highest digestibility. From the maximized PA mass spectrum (Fig. 2B) several peaks related to chymotrypsin activity were found. In fact, the used trypsin (Novo 6.0 S) is a comparatively crude preparation intended primarily for use as an additive in detergents. In this study, the crude enzyme preparation was used in order to approach the full potential of the matrix. As Eupergit C has a specified binding capacity of 140 mg protein per g of matrix (d.w.), the use of pure enzyme preparations would, for the column-dimensions investigated here, be economically unfeasible. Still, a relatively high protein digestibility was achieved in the reactor under the current conditions, i.e. ∼10 min analyte residence time in the protease column, with no evidence of undigested PA present in the eluate. One reason for this effective digestion should be the mentioned high enzyme content in the column with >100-fold higher load compared to conventional in-solution protease digestion. In addition, the large interconnected pore structure in the Eupergit C beads imply that the mass transfer resistance related to analyte diffusion is minimal. Hence, in the Eupergit C matrix, protein molecules will essentially be transported convectively to the immobilized enzyme molecules yielding a large contact surface over time. In contrast, the large size (150 μm) together with the porous structure of the Eupergit C beads give rise to a significant band broadening (i.e. 30 times) as can be seen in Fig. 2A. To minimize this analyte dilution one needs to use a matrix with smaller particle size and pack the column even harder. However, to be able to process the sufficient amount of generated peptides, it should be taken into account that a certain elution time is needed.

To further investigate the influence of residence time on the digestion efficiency, the protease reactor was scaled up to 150 mm × 2 mm. PA solution (12 μl, 7.1 μM) was injected (i.e. the injection volume was adjusted to constitute 2.5% of the packed matrix volume) in the flow of digestion buffer maintained at 10 μl/min. Compared to the smaller column set-up the scaled-up system generated a 33% increase in number of peptides detected using a *signal-to-noise* ratio of 5

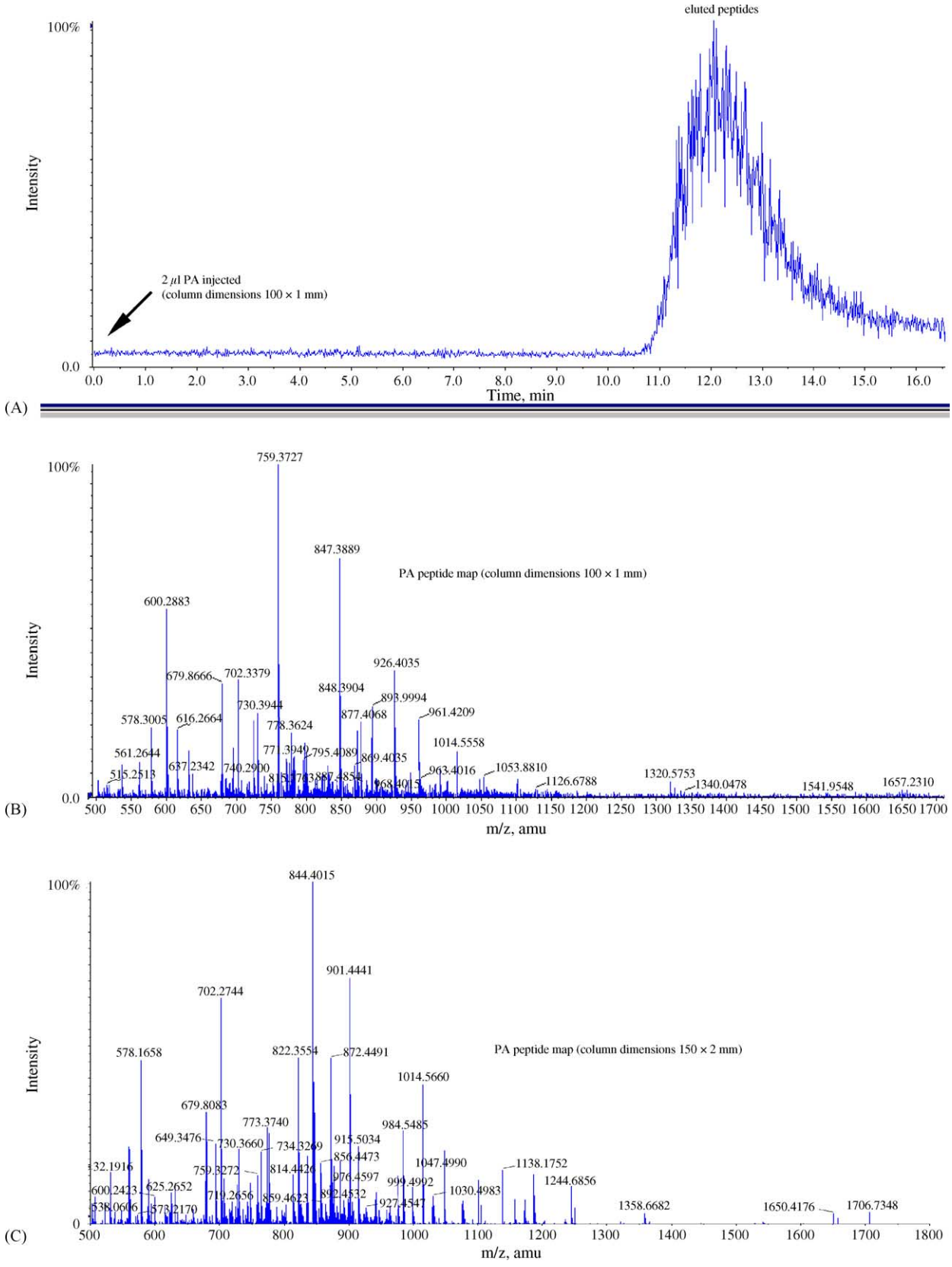

Fig. 2. The total ion count (TIC) over time (A) after injection of 2 μl PA (7.1 μM) on the protease column (100 mm × 1 mm). The corresponding PA mass spectrum of eluted peptides is shown in (B). The digestion efficiency was maximized with regard to the flow rate through (10 μl/min) and temperature over (37 °C) the enzyme reactor. In (C) 12 μl of PA (7.1 μM) was injected on a 150 mm × 2 mm column under otherwise the same conditions. A 33% increase in number of detected peptides was achieved.

(Fig. 2C). Hence, improved digestibility with increased contact time for PA (i.e. from 10 to 21 min) implied that the impeded digestion performance of the smaller protease column was due to limitations connected to the performance of the utilized μLC system rather than to the actual digestion efficiency of the protease matrix (i.e. minimal reproducible flow rate generated by the HPLC pump was 10 μl/min). However, taking into account the effect of increased protein digestion for the larger enzyme reactor along with the opportunity of increasing the amount of analyte injected, the analytical window where eluted peptides could be processed by tandem-MS was increased from 5 to 15 min. When analysing large proteins with numerous eluted peptides, increased elution time would be beneficial allowing the IDA tandem-MS processing of the peptide map to approach completion.

To explore the possibility of a concomitant analysis of proteins in a mixture, a low-pressure packed Sephadex G-75 gel filtration column with dimensions 100 mm × 4.6 mm was connected in series before the protease column (150 mm × 2 mm). In order to achieve a practical over-all analysis time, the flow rate of digestion buffer was increased to 25 μl/min, i.e. yielding a mean residence time for the analytes in the system of 28 min. A sample containing PA, SEB and cytochrome *c* was then applied to the serial column system (Fig. 3). Prior to injection the sample protein mixture was heat-treated in the presence of guanidine-HCl and β-mercaptoethanol. Theoretically, the high concentrations of salt (8 M) in the sample preparation could cause interferences in terms of suppressed MS ionisation because of the strong association between denaturant and protein. However, the peptide maps presented in Fig. 4 indicated that no or an insignificantly small amount of guanidine was co-eluted with the peptide fractions (i.e. the absence of +95 *m*/*z* guanidine adducts). In contrast, from the corresponding TIC-spectrum (Fig. 3) it can be seen that a substantial co-elution of peptides from different proteins occur. Even so, with this relatively low resolution, resulting in somewhat mixed peptide maps, the selective capacity of the MS software yielded an efficient target peptide processing.

The assumed high trypsin load in the digestion column achieved with the polyacrylate matrix (>100 mg trypsin per g (d.w.) of Eupergit C) offered high yields in terms of protein sequence coverage for the proteins investigated in this study. Average sequence coverage of 50–80% for moderately sized proteins (30–60 kDa) is reported for batch performed trypsin digestion [20,21], whereas for smaller proteins (10–30 kDa) average sequence coverage well above 90% is typically reached [22,23]. For on-column digestion methods this number has been reported to be considerably lower [4,9]. The sequence coverage achieved in this work was 78–95% for the model proteins examined. Notable is the rather high sequence coverage for SEB (Fig. 5), which is an extremely stable protein due to the tight-folded structure. SEB has been reported to rapidly recover its three-dimensional structure even after being exposed to harsh conditions (e.g. pH 11, >100 °C) providing an efficient protection against protease digestion [24,25].

The generic feature of the digestion system together with an inherently high reproducibility implies that a vast array of different protein substances can be processed with sufficient degree of fragmentation. In addition, to have more than one proteolytic activity in the column could turn out to be advantageous since the enzymes may be able to work in a synergistic fashion. Also the identification of small changes in the protein primary structure of a target substance should be possible. In order to avoid complicated cleavages, the use of enzyme preparations of higher purity would in this case be recommended.

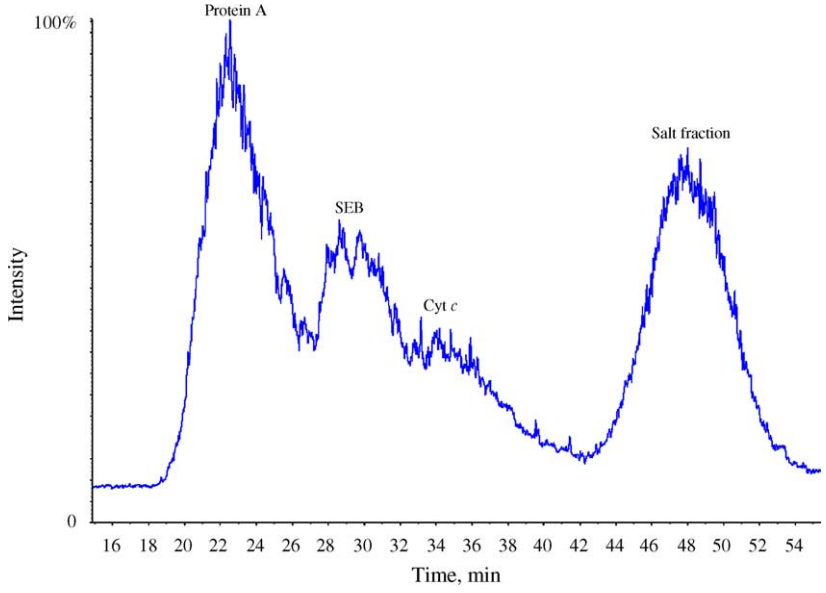

Fig. 3. Total ion count (TIC) over time after injection of a sample (12 μl) containing PA, SEB and cytochrome *c* (0.33 mg/ml of each protein) on the SEC (Sephadex G-75)/protease column system (100 mm × 4.6 mm and 150 mm × 2 mm, respectively).

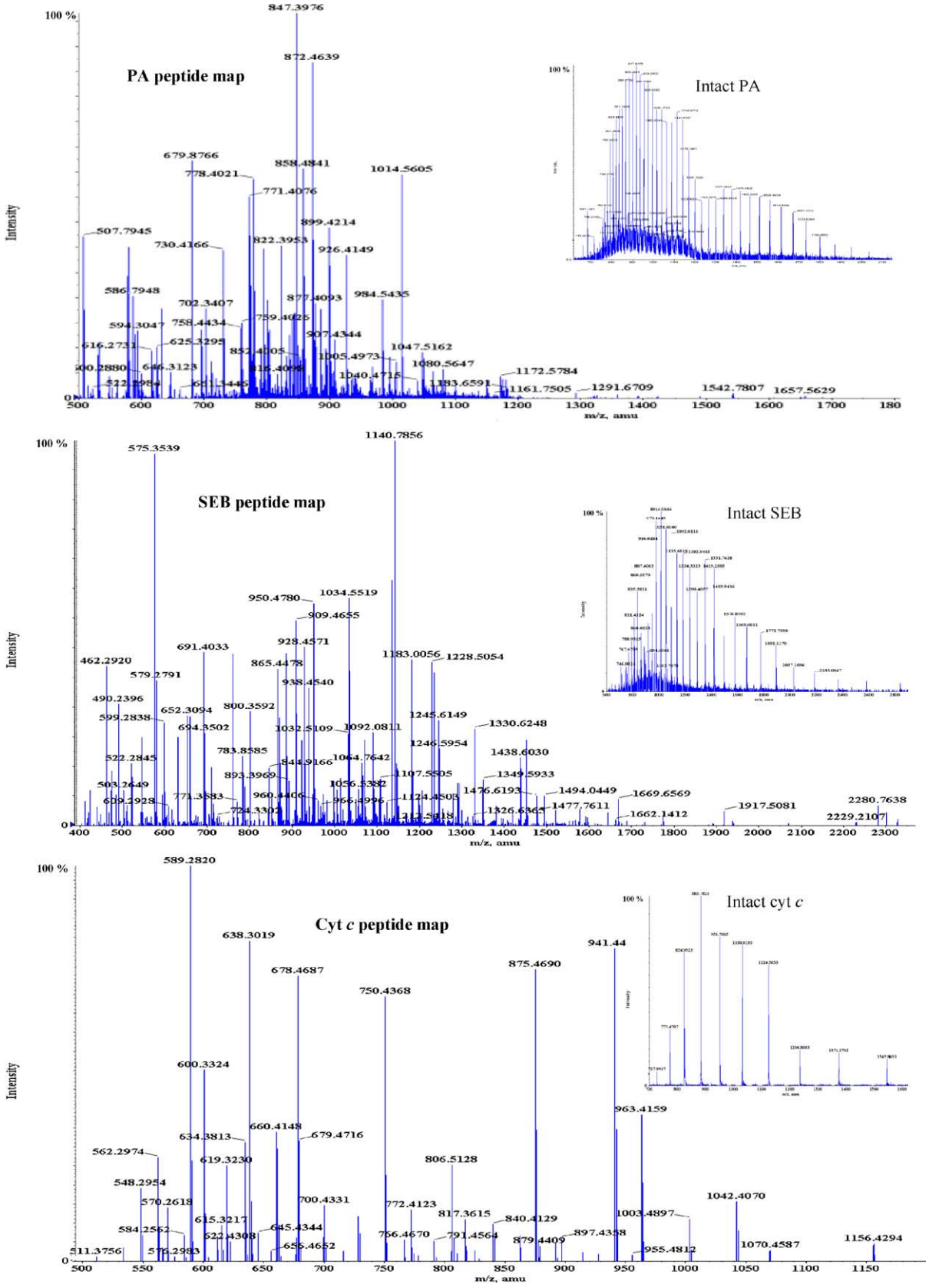

Fig. 4. Generated peptide maps of PA, SEB and cytochrome *c* (retention times 22, 29 and 34 min, respectively) after injection of a sample (12 μl, 0.33 mg/ml of each protein) on the SEC (Sephadex G-75)/protease column system (100 mm × 4.6 mm and 150 mm × 2 mm, respectively). Envelopes derived from the intact proteins (ESI-TOF-MS) is shown.

Fig. 5. SEB amino acid sequence with identified peptides underlined and the cleavage-sites marked with arrows.

Hence, for biotechnological purposes this on-line digestion approach could be a useful tool in the quality control of partially or fully purified proteins derived from the fermentation in a biotechnology production process. The linear column set-up also offers advantages when moving towards a miniaturization. The benefits in approaching nano-scale flow-rates, utilizing capillary-sized columns (e.g. 25–75 μm Ø) would be a further exploitation of the catalytic efficiency of the protease-matrix. Consequently, as high sensitivity and efficient sample utilization are important features when developing high throughput based analysis systems, we believe that the method described here could be a platform suitable for proteomic routine work.

Acknowledgements

The work was supported by the Swedish Competence Center for Bioseparation (CBioSep). Maria Andersson wishes to acknowledge the Swedish International Development Agency-Department for Research Cooperation, SIDA/SAREC for financial support.

References

[1] K.S. Nair, A. Jaleel, Y.W. Asmann, K.R. Short, S. Raghavakaimal, Am. J. Phys. 286 (2004) 863.
[2] J.A. Loo, Adv. Protein Chem. 65 (2003) 25.
[3] G. Vecchione, B. Casetta, R. Santacroce, M.A. Margaglione, Rapid Commun. Mass Spectrom. 15 (2001) 1383.
[4] J. Samskog, D. Bylund, S.P. Jacobsson, K.E. Markides, J. Chromatogr. A 998 (2003) 83.
[5] J. Yun, S.L. Cheng, J. Chromatogr. A 924 (2001) 315.
[6] E. Calleri, C. Temporini, E. Perani, C. Stella, S. Rudaz, D. Lubda, G. Mellerio, J.L. Veuthey, G. Caccialanza, G. Massolini, J. Chromatogr. A 1045 (2004) 99.
[7] S.K. Kumiko, K. Masaru, T. Toshimasa, Anal. Chem. 74 (2002) 2943.
[8] E. Bonneil, K.C. Waldron, Talanta 53 (2000) 687.
[9] M. Ye, S. Hu, R.M. Schoenherr, N.J. Dovichi, Electrophoresis 25 (2004) 1319.
[10] Y. He, W. Zhong, E.S. Yeung, J. Chromatog. B 782 (2002) 331.
[11] I. Migneault, C. Dartiguenave, J. Vinh, M.J. Bertrand, K.C. Waldron, Electrophoresis 25 (2004) 1367.
[12] W. Tong, A. Link, J.K. Eng, J.R. Yates, Anal. Chem. 71 (1999) 2270.
[13] N.E. Labrou, J. Chromatog. B 790 (2003) 67.
[14] A. Seth, L.J. Stern, T.H. Ottenhoff, I. Engel, M.J. Owen, J.R. Lamb, R.D. Klausner, D.C. Wiley, Nature 369 (1994) 324.
[15] M.D. Baker, K.R. Acharya, Int. J. Med. Microb. 293 (2004) 529.
[16] R.K. Tweten, J.J. Iandolo, J. Bacteriol. 153 (1983) 297.
[17] C. Mateo, G. Fernandez-Lorente, O. Abian, R. Fernandez-Lafuente, J.M. Guisan, Biomacromolecules 1 (2000) 739.
[18] C. Mateo, G. Fernandez-Lorente, E. Cortes, J.L. Garcia, R. Fernandez-Lafuente, J.M. Guisan, Biotech. Bioeng. 76 (2001) 269.
[19] C. Mateo, O. Abian, R. Fernandez-Lafuente, J.M. Guisan, Enz. Microb. Tech. 26 (2000) 509.
[20] E.J. Finehout, K.H. Lee, Electrophoresis 24 (2003) 3508.
[21] V. Kadlcik, M. Strohalm, M. Kodicek, Biochem. Biophys. Research Commun. 305 (2003) 1091.
[22] D.A. Barnett, B. Ells, R. Guevremont, R.W. Purves, J. Am. Soc. Mass Spectrom. 13 (2002) 1282.
[23] K. Bernardo, S. Fleer, N. Pakulat, O. Krut, F. Hunger, M. Kronke, Proteomics 2 (2002) 740.
[24] M. Schwabe, S. Notermans, R. Boot, S.R. Tatini, J. Kramer, Int. J. Foodmicrobiol. 10 (1990) 33.
[25] F.S. Fluer, Zhurnal mikrobiologii, epidemiologii, i immunobiologii 11 (1983) 36 (in Russian, PubMed ID 6362299).