MULTIDIMENSIONAL PEPTIDE/PROTEIN ANALYSIS AND IDENTIFICATION BY SEQUENCE DATABASE SEARCH USING MASS SPECTROMETRIC DATA

Christian Schley[1], Matthias Altmeyer[2], Rolf Müller[2], Christian G. Huber[1]

[1]Department of Chemistry, Instrumental Analysis and Bioanalysis, Saarland University, 66123 Saarbrücken, Germany

[2]Department of Pharmacy, Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany

Abstract. In order to generate proteomics data that are suitable to validate protein identification in complex mixtures using multidimensional liquid-chromatography-mass spectrometry approaches, we implemented an offline two-dimensional liquid chromatography method combining strong cation-exchange- and ion-pair reversed-phase chromatography followed by electrospray ionization tandem mass spectrormetry (ESI-MS/MS) for the analysis of a bovine serum albumin digest. The fragment ion spectra generated by ESI-MS/MS were subsequently analyzed via MASCOT database search. The obtained identification data were evaluated in terms of quality of protein/peptide identification by means of score values, reproducibility of identification in replicate measurements, distribution of tryptic peptides among different fractions, and overall number of unique identified proteins/peptides. Finally, we improved the trapping conditions in the second dimension by using a more hydrophobic amphiphile in the loading buffer. The improvement was demonstrated by comparison of the obtained identification data, such as number of identified peptides, cumulative mowse scores and reproducibility of identification.

1 Introduction

Protein samples of biological origin are by nature highly complex and therefore, their analysis requires separation techniques with high resolving power and high peak capacity, respectively. During the past few years, the separation of digests of whole-protein lysates by multidimensional liquid chromatography, which can be readily interfaced on-line to mass spectrometry [1-3], has become a real alternative to two-dimensional polyacrylamide gel electrophoresis for high-throughput protein identifications.

Dagstuhl Seminar Proceedings 05471
Computational Proteomics
http://drops.dagstuhl.de/opus/volltexte/2006/538

For proteins and peptides, ion-exchange chromatography has been proven very useful as a first-dimension fractionation method. Reversed-phase chromatography is the method of choice for separation in the second dimension due to its high degree of orthogonality with respect to ion exchange chromatography and the compability with on-line or off-line analyte detection and ionization by electrospray ionization-(ESI) or matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS).
Monolithic columns based on poly-styrene-divinylbenzene copolymer (PS/DVB) have been shown to be highly suitable for high-resolution separation of biopolymers such as proteins [4,5] and peptides [6,7]. Therefore we developed a two-dimensional HPLC system using monolithic PS/DVB columns in the second dimension for the analysis of complex protein/peptide mixtures. The first dimension comprised a strong cation exchange column (250 x 4.0 mm i.d.) and in the second dimension we applied a preconcentration column (10 x 0.2 mm i.d.) and separation column (60 x 0.2 mm i.d.) based on poly-styrene-divinylbenzene.

2 Characterization of the off-line 2D-HPLC system coupled to MS and MS/MS using a bovine serum albumin digest

The developed 2D-HPLC system using monolithic columns was characterized with identification data obtained by Mascot [8,9] database search for a bovine serum albumin digest. For this purpose we separated 200 pmol of a bovine serum albumin digest with strong cation exchange chromatography (1^{st} dimension) and took 19 one-minute fractions for subsequent reinjection in the second dimension. Additionally the strong cation exchange separation and fractionation could be controlled via UV detection.
Figure 1 shows the obtained UV chromatogram with the outlined one-minute fractions. The collected SCX-fractions were concentrated and desalted on a monolithic trap column (10 x 0.2 mm i.d.) and subsequently transferred to the monolithic separation column in backflush mode. The ion-pair reversed-phase separation of 5 representative SCX-fractions on a 200 µm i.d. PS/DVB monolith is illustrated in **Figure 2**.

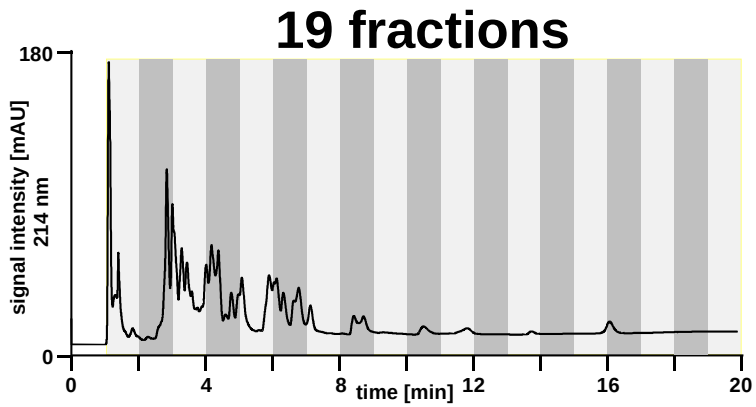

Figure 1. Cation-exchange separation of tryptic peptides of serum albumin bovine digest. Column, ProPac SCX-10, 250 x 4 mm i.d.; mobile phase, (A) 5 mM aqueous NaH_2PO_4 buffer in 5 % acetonitrile, pH 3.0, (B) as eluent (A) + 0.5 M NaCl; gradient, 0-100 % B in 45 min; flow rate, 1.0 ml/min; temperature, 25°C; detection, UV, 214 nm; sample, serum albumin bovine digest, 200 pmol; fractionation, every minute.

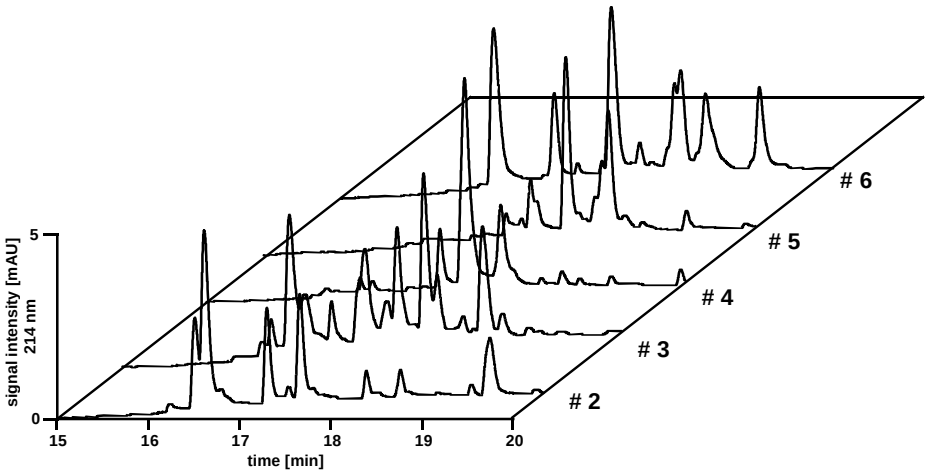

Figure 2. Reversed-phase chromatograms of tryptic peptides of bovine serum albumin after pre-fractionation by cation-exchange chromatography. Columns, monolithic poly-styrene/divinylbenzene, 10 and 60 x 0.2 mm i.d.; loading phase, 0.05 % aqueous trifluoroacetic acid; trap time, 8 min; loading flow rate, 10 µl/min; mobile phase, (A) 0.05 % aqueous trifluoroacetic acid, (B) 0.05 % trifluoroacetic acid in acetonitrile; gradient, 0-50 % B in 15 min; flow rate, 2.0 µl/min; temperature, 25°C; detection, UV, 214 nm; sample, tryptic peptides of bovine serum albumin, fractions 2-6 from cation-exchange chromatography, 10 µl injected.

These chromatograms clearly prove that the separation system is truly two-dimensional, which means that the coupled separation mechanisms of each dimension are orthogonal to each other, since the whole separation window in the second dimension is covered with tryptic peptides from the first dimension. This property enables us to get an increased peak capacity compared to an one-dimensional approach, what is especially important for the analysis of complex biological samples. This additional level of separation facilitates an improvement in the relative amounts of peptides/proteins that can be detected. Moreover the use of off-line fractionation in a relatively large SCX separation column in the first dimension offers a means of increasing the dynamic range of the analytical method due to the considerably higher loadability in the first dimension as compared to an on-line 2D system. The increased dynamic range gives us the opportunity to detect even low abundant peptides/proteins in the presence of a large excess of peptides/proteins coming from high abundant proteins.
Examplarily one reconstructed total ion current chromatogram of ion-exchange fraction 6 is depicted in **Figure 3**. The peptides of serum albumin identified by tandem mass spectrometry and Mascot database search are labelled in the chromatogram. All 19 SCX-fractions were analyzed in triplicates by tandem mass spectrometry and subsequently evaluated via Mascot database search. The identified peptides of serum albumin are summarized in Figure 4. The parameters of the Mascot search were the following:

Database:	MSDB
Taxonomy:	chordata(vertebrates and relatives)
Stringency:	medium
Fixed modification:	carboxymethylated
Variable modification:	none
Enzyme:	tryspin
Peptide tolerance:	+/- 1.3 Da
MS/MS tolerance:	+/- 0.3 Da
Max. no. of missed cleavages:	2

Each identified peptide gets a certain mowse (**mo**lecular **w**eight **se**arch) score assigned by the probability based Mascot scoring algorithm, which indicates how reliable the proposed peptides are. In this analysis we used a mowse score threshold of 15. This limit is relatively low, but

because we know the content of the sample, the possibility of a random hit is very little. With these settings we obtained a sequence coverage of 63% for serum albumin, with 33 individual identified tryptic peptides.

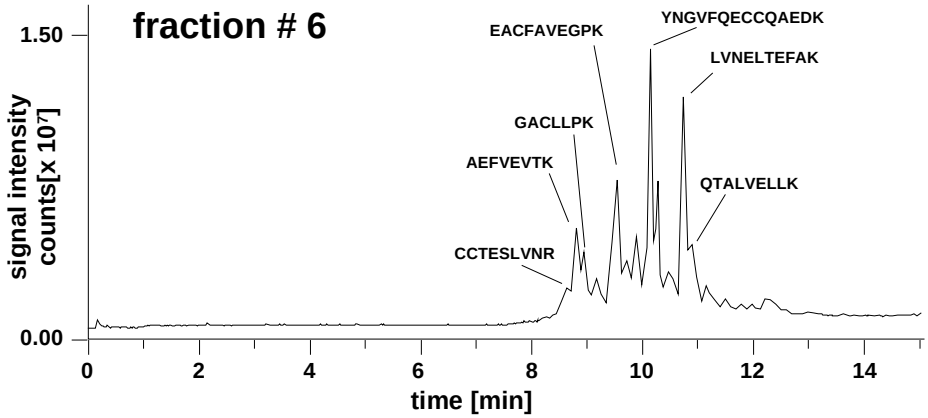

Figure 3. RP-HPLC-ESI-MS/MS peptide identifications of the tryptic digested serum albumin, exemplarily for the SCX fraction 6. Detection, ESI-MS/MS.Other conditions as in Figure 2.

DTHKSEIAHR FKDLGEEHFK GLVLIAFSQY LQQCPFDEHV KLVNELTEFA
KTCVADESHA GCEKSLHTLF GDELCKVASL RETYGDMADC CEKQEPERNE
CFLSHKDDSP DLPKLKPDPN TLCDEFKADE KKFWGKYLYE IARRHPYFYA
PELLYYANKY NGVFQECCQA EDKGACLLPK IETMREKVLA SSARQRLRCA
SIQKFGERAL KAWSVARLSQ KFPKAEFVEV TKLVTDLTKV HKECCHGDLL
ECADDRADLA KYICDNQDTI SSKLKECCDK PLLEKSHCIA EVEKDAIPEN
LPPLTADFAE DKDVCKNYQE AKDAFLGSFL YEYSRRHPEY AVSVLLRLAK
EYEATLEECC AKDDPHACYS TVFDKLKHLV DEPQNLIKQN CDQFEKLGEY
GFQNALIVRY TRKVPQVSTP TLVEVSRSLG KVGTRCCTKP ESERMPCTED
YLSLILNRLC VLHEKTPVSE KVTKCCTESL VNRRPCFSAL TPDETYVPKA
FDEKLFTFHA DICTLPDTEK QIKKQTALVE LLKHKPKATE EQLKTVMENF
VAFVDKCCAA DDKEACFAVE GPKLVVSTQT ALA

Figure 4. Sequence coverage of tryptic peptides in two-dimensional analysis of bovine serum albumin. The identified peptides are labelled red.

3 Improvement of trapping

Evaluation of the identification data of the two-dimensional serum albumin digest analysis, revealed that some small and/or hydrophilic peptides are lost during the preconcentration step. One strategy, which we pursued to increase the retentivity of these peptides, was to use more hydrophobic amphiphiles in the solvents used for trapping of the peptides. Long-chain carboxylic acids are known to be more strongly adsorbed to hydrophobic surfaces, resulting in increased surface potential and stronger retention of peptides in ion-pair reversed-phase chromatography. Therefore we compared trifluoroacetic acid and heptafluorobutyric acid as ion-pair reagents in the loading solvent. 0.5 and 2.0 pmol of a bovine serum albumin digest were preconcentrated at the monolithic trap column, subsequently separated on the monolithic separation column and finally analyzed via electrospray tandem mass spectrometry. Each analysis was done in triplicates. The identification data obtained by Mascot database search proved the assumption that the use of heptafluorobutyric acid increases the retentivity of peptides, because the overall number of identified peptides was higher than with trifluorobutyric acid (Figure 5). The same is confirmed by the cumulative mowse score and the reproducibility of peptide identification (number of 3-fold identified peptides). In all cases we obtained the best results when we applied heptafluorobutyric acid instead of trifluoroacetic acid. But due to the fact that the serum albumin digest offers only a relative limited set of tryptic peptides, the results between both ion-pair reagents differ not that much. For this reason it would be preferable to evaluate the effect of different loading solvent additives with a more complex sample.

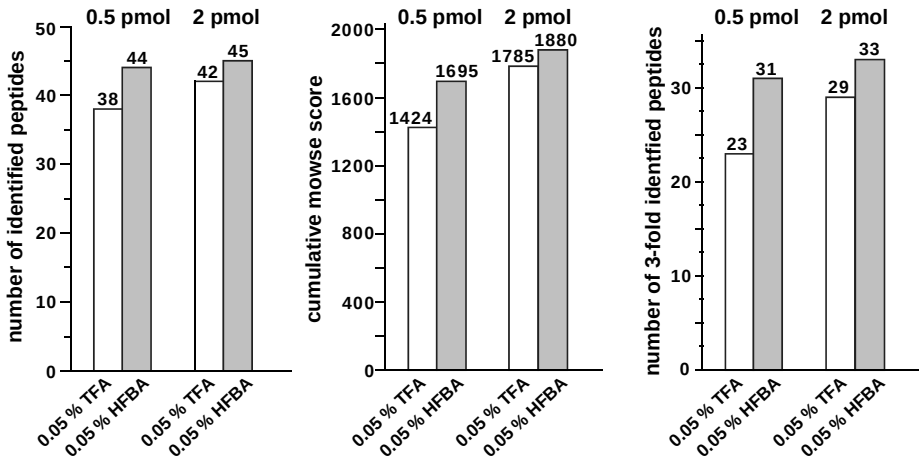

Figure 5. Comparison of the number of identified peptides, the cumulative mowse scores and the number of 3-fold identified peptides for serum albumin using trifluoroacetic acid and heptfluorobutyric acid as loading solvent additives. Ion-pair reagents, trifluoroacetic acid and heptafluorobutyric acid; detection, ESI-MS/MS; sample, tryptic peptides of bovine serum albumin, 0.5 and 2.0 pmol; other conditions as in Figure 2.

4 Conclusion

The analysis of the serum albumin digest using our two-dimensional HPLC system coupled to an electrospray mass spectrometer gave us the opportunity to characterize and improve the system especially with the obtained identification data from Mascot database search. But although we only used a digest of serum albumin, which is by far not comparable with a whole proteome digest, as evaluation sample, manual interpretation of the data was very time consuming and challenging, not only because data export was tedious but also because the commercial software packages are usually not very flexible in terms of parameter selection and settings, as well as selected quality criteria. So one can imagine how labor-intensive the evaluation of such an analysis would be. Therefore the development of new powerful algorithms and data evaluation software (e.g. MS data analysis software, database search software) for proteome analyses, which is able to deal with huge amounts of protein/peptide identification data and to minimize false positive hits in a fully automated manner would be desirable. Our evaluated data set, which is relatively clear, could be suitable to develop such software tools.

5 References

[1] M. Gilar, P. Olivova, A. E. Daly and J. C. Gebler, *J. Sep. Sci.*, 28 (2005) 1694.

[2] W. J. Qian, J. M. Jacobs, D. G. Camp 2nd, M. E. Monroe, R. J. Moore, M. A. Gritsenko, S. E. Calvano, S. F. Lowry, W. Xiao, L. L. Moldawer, R. W. Davis, R. G. Tompkins and R. D. Smith, *Proteomics*, 5 (2005) 572.

[3] D. A. Wolters, M. P. Washburn and J. R. Yates, *Anal. Chem.*, 73 (2001) 5683.

[4] S. Xie, R. W. Allington, F. Svec and J. M. J. Fréchet, *J. Chromatogr. A*, 865 (1999) 169.

[5] I. Gusev, X. Huang and C. Horváth, *J. Chromatogr. A*, 855 (1999) 273.

[6] A. Premstaller, H. Oberacher, W. Walcher, A.-M. Timperio, L. Zolla, J.-P. Chervet, N. Cavusoglu, A. Van Dorsselaer and C. G. Huber, *Anal. Chem.*, 73 (2001) 2390.

[7] A. R. Ivanov, L. Zang and B. L. Karger, *Anal. Chem.*, 75 (2003) 5306.

[8] D. N. Perkins, D. J. Pappin, D. M. Creasy and J. S. Cottrell, *Electrophoresis*, 20 (1999) 3551.

[9] D. J. Pappin, P. Hojrup and A. J. Bleasby, *Curr. Biol.*, 3 (1993) 327.