Repeats versus Mass-windows – optimization strategies for Mudpit type experiments to increase proteome coverage

Jürgen Schäfer, Christian Baumann and Josef Schwarz
Proteome Sciences R&D; Industriepark Höchst Geb.G865; 65929 Frankfurt Germany

Overview

- **Gelfree Proteomics techniques have to deal with the analysis of high complex mixtures.**
- **Sample separations with SCX and Reversed phase HPLC are needed to reduce the complexity in MS [1] [2].**
- **Even after SCX-RP HPLC the complexity is still high in ESI-MS.**
- **The number of precursors that can be fragmented in MS/MS is limited by the MS instrument, it is not possible to fragment every potential peptide ion.**
- **Simple repeats can increase the number of identified proteins. The use of mass windows can also reduce the complexity for MS.**
- **Here we present results of both optimization strategies.**

Introduction

For the separation and visualization of complex protein mixtures two dimensional gel electrophoresis combined with MALDI MS is a well established technique. However, some of the limitations associated with the 2D/MS approach (e.g. difficulties in the handling of membrane proteins) have prompted the development of gelfree strategies. In the so-called shot gun approaches, the whole proteome is usually enzymatically digested. The analysis of the resulting digested protein mixture has high demands on the separation system (SCX, RP HPLC) and the data acquiring system (MS and MS/MS). Here we discuss the optimization of repeated injections which enhance the performance of a MUDPIT type analysis.

Methods

SCX-HPLC ESI-MS/MS and data analysis.
A tryptic yeast digest separated manually on a SCX column into 9 fractions, was used for analysis. All the MS measurement were done on a Q-Tof 2 Instrument equipped with a Cap LC HPLC system. The SCX fractions were trapped onto a 0,3 mm i.d. x 5 mm precolumn and eluted trough a 75 µm i.d. RP18 column in an 120 min run. The Q-Tof 2 settings were optimized with regard to scan times and precursor selection. All MS/MS data were processed with SEQUEST and analyzed according to well established criteria:

Z	2	3	4	(delta Cn > 0,1)
Xcorr	> 1,8	> 2,5	> 3,0	

Only results that fulfil these criteria were included in this study.

Results

Repeats

# Runs	# dta	# pos id	eff%	# pep	# prot
1	1475	832	56	**548**	**264**
1	1486	860	58	**551**	**264**
1	1467	841	57	**547**	**266**
1	1438	812	56	**517**	**249**
1	1457	798	55	**517**	**245**

# Runs	# dta	# pos id	eff%	# pep			# prot		
					Increase	%		Increase	%
1	1475	832	56	**548**			**264**		
2	2961	1692	57	**664**	**116**	***21,2***	**332**	**68**	***25,8***
3	4428	2533	57	**751**	**87**	***13,1***	**381**	**49**	***14,8***
4	5866	3345	57	**805**	**54**	***7,2***	**413**	**32**	***8,4***
5	7323	4143	57	**848**	**43**	***5,3***	**438**	**25**	***6,1***

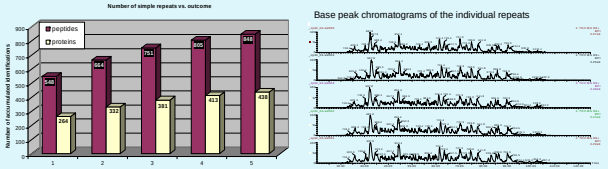

Figure 1. Table, chart and Base peak chromatogram for the result of repeated injections of the sample

Mass windows

MW	mw	# dta	# pos id	eff%	# pep	# prot
1	150-553	733	182	25	**145**	**109**
2	547-703	1254	525	42	**373**	**255**
3	697-853	1361	680	50	**332**	**237**
4	847-1003	1051	521	50	**243**	**198**
5	997-2000	473	226	48	**110**	**93**
1	1	733	182	25	**145**	**109**
1-2	2	1987	707	36	**511**	**304**
1-3	3	3348	1387	41	**827**	**436**
1-4	4	4399	1908	43	**1053**	**530**
1-5	5	4872	2134	44	**1146**	**572**

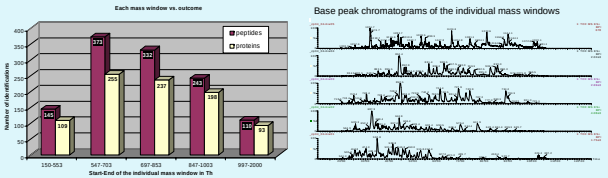

Figure 2. Table, chart and Base peak chromatogram for the result of the usage of mass windows for the analysis.

Repeats of one mass window

Rep.	# Runs	# dta	# pos id	eff%	# pep	# prot
1	1	1254	525	42	**373**	**255**
2	1	1157	512	44	**352**	**233**
3	1	1200	527	44	**364**	**246**
4	1	1161	487	42	**333**	**232**
5	1	1173	495	42	**341**	**238**

Rep.	# Runs	# dta	# pos id	eff%	# pep			# prot		
						Increase	%		Increase	%
1	1	1254	525	42	**373**			**255**		
1-2	2	2411	1037	43	**466**	**87**	***23,3***	**314**	**59**	***23,1***
1-3	3	3611	1564	43	**523**	**63**	***13,7***	**354**	**40**	***12,7***
1-4	4	4772	2051	43	**544**	**21**	***4,0***	**372**	**18**	***5,1***
1-5	5	5945	2546	43	**566**	**22**	***4,0***	**391**	**19**	***5,1***

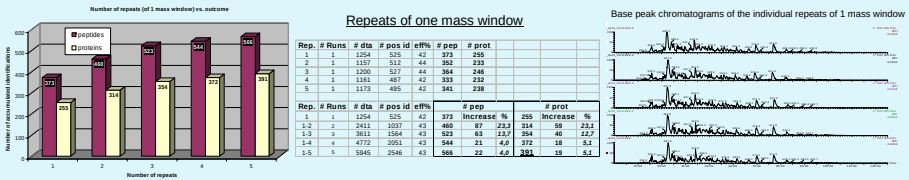

Figure 3. Chart, table and Base peak chromatogram for the repeated analyses using the mass window from 547 to 703 Th.

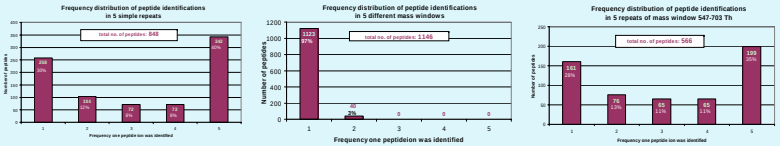

Figure 4. Identification frequency distribution chart for simple repeats, 5 mass windows and repeating mass window from 547 to 703 Th.

I. Repeated sample injection within a MUDPIT experiment show a high reproducibility for the number of identified proteins of each run (245 to 266 proteins out of the same SCX fraction). - In each run of a series of repeats new proteins can be identified. - The number of overall identified proteins behaves like a saturation curve.
II. Repeated sample injection analyzing different slightly overlapping mass windows results in a higher number of identifications. - There is a sample and protease dependent inhomogeneous distribution of the peptide ions over the whole m/z range with a main focus in the region 550 to 1000 Th. With the use of mass windows there is no need to work with exclude lists. Adapting the mass windows to the peptide distribution results in higher overall number of 572 identified proteins for 5 replicate injections out of one SCX fraction.
III. Repeats of a selected mass window show a similar saturation behavior for the identification number.
IV. The usage of mass windows reduces the frequency a peptide ion is identified without the use of exclude lists.

Conclusions & Perspectives

- **Repeats using different mass windows results in a higher number of identified proteins than simple repeats.**
- **Repeats of one mass window show a saturation behavior for the number of identified proteins like simple repeats.**
- **The use of mass windows reduces the redundancy.**
- **Mass windows allow to fragment more precursors than in simple repeats and enables a deeper look into a proteome.**
- **Combination of mass windows and repeats can be used together for high identification numbers.**

References

[1] Washburn, M.P., D. Wolters, et al.
"Large-scale analysis of the yeast proteome by multidimensional protein identification technology"
Nature Biotechnology **2001**, 19 (3): 242-247

[2] Washburn, M.P., D. Wolters, et al.
"An automated multidimensional protein identification technology for shotgun proteomics"
Anal. Chem. **2001**, 73 (23): 5683-90