ELSEVIER

International Journal of Mass Spectrometry 219 (2002) 245–251

International Journal of Mass Spectrometry

www.elsevier.com/locate/ijms

Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT

W. Hayes McDonald[a], Ryoma Ohi[b], David T. Miyamoto[b], Timothy J. Mitchison[b,c], John R. Yates, III[a,d,*]

[a] *Department of Cell Biology, SR-11, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92122, USA*
[b] *Department of Cell Biology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA*
[c] *Institute of Chemistry and Cell Biology (ICCB), Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA*
[d] *Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121, USA*

Received 17 August 2001; accepted 10 January 2002

Abstract

One of the most effective methods for the direct identification of proteins from complex mixtures without first having to resolve them by polyacrylamide gel electrophoresis is to separate proteolytically generated peptides by microcapillary HPLC and then collect data directly on the eluent using a tandem mass spectrometer. Multidimensional HPLC separation techniques provide access to even more complex mixtures of proteins. A set of techniques for multidimensional analysis was developed in our lab; collectively they are known as multidimensional protein identification technology (MudPIT). These strategies employ a biphasic column with a section of reversed phase (RP) material flanked by strong cation exchange (SCX) resin and allow for multidimensional separation of peptides. A variation on MudPIT adds an additional section of RP material behind the SCX and RP. This 3-phase column can be used for "online" desalting of the sample. We compare the analysis of a complex mixture of proteins purified by their association with bovine brain microtubules using a single-dimension LC-MS/MS column, a 2-phase (standard) MudPIT column, and a 3-phase MudPIT column. We find that the 3-phase MudPIT column yields a greater number of protein identifications for this test sample and allows data to be collected on a set of hydrophilic peptides not sampled using the 2-phase MudPIT column. (Int J Mass Spectrom 219 (2002) 245–251)

Keywords: Protein identification; MudPIT; Tandem mass spectrometry; MS/MS; HPLC

1. Introduction

The use of HPLC directly coupled to a tandem mass spectrometer can identify proteins in mixtures and circumvent the use of gel electrophoresis to separate proteins prior to identification by mass spectrometry [1–5]. To identify proteins in mixtures, the proteins are first proteolytically digested. The resulting complex peptide mixture is separated by reversed phase (RP) chromatography, the eluting peptides directly ionized via electrospray, and then analyzed by tandem mass spectrometry. LC-MS/MS can be applied to peptides recovered from digested bands/spots resolved by single-dimension (1D) and two-dimensional

* Corresponding author. E-mail: jyates@scripps.edu

1387-3806/02/$ – see front matter
PII S1387-3806(02)00563-8

(2D) polyacrylamide gel electrophoresis, but the true power of the technique becomes evident for mixture analysis. The tandem mass spectrometry data can be correlated to the database sequences and peptides identified by their amino acid sequence [1,6]. LC-MS/MS has proven successful for the analysis of numerous proteins and protein complexes [3,7–11]; however, more complicated protein mixtures produce much greater numbers of peptides which require added separation capacity and greater sampling time to acquire data on these peptides.

The multidimensional protein identification technology (MudPIT) strategy was developed to address both of these problems [7]. For MudPIT, strong cation exchange (SCX) resin is placed upstream of the RP portion of the column. Peptides at low pH are bound to the SCX phase and subsequently "stepped" off using multiple salt pulses followed by reversed phase separation of each subset of peptides. The initial application of this strategy was known as direct analysis of large protein complexes (DALPC). DALPC was validated on the *S. cerevisiae* ribosome, a complex that was particularly difficult to resolve by 2D gel electrophoresis due to both high pIs of the proteins and their small size. This analysis revealed two previously undescribed ribosomal components and demonstrated that the amount of starting material required for DALPC was significantly less than the amounts necessary for 2D gel-based analysis. Others groups have also found multidimensional separation coupled directly to a tandem mass spectrometer to be an effective substitute for gel-based protein analysis strategies [11,12].

For analysis of even more complex protein mixtures, such as whole cell extracts, greater separation and data acquisition capacity is required. Further optimizations in HPLC materials, buffers, and general strategy have resulted in the current form of MudPIT [13,19]. This strategy coupled with sample preparation modifications demonstrated the most complete description of the *S. cerevisiae* proteome reported to date, with the positive identification of 1498 proteins [13].

Because of inefficiencies of peptide interaction with the SCX caused by competing salts, peptide solutions must first be desalted before loading onto the standard 2-phase MudPIT column. This desalting step is usually performed "offline" using a solid phase extraction column; the peptides in the organic eluent are then vacuum concentrated down to near dryness before being diluted back into aqueous buffer for loading onto the MudPIT column. In order to minimize these extra steps, and thus manipulations of the sample, it is possible to place an additional phase of RP material in the microcapillary upstream of the SCX. Samples can then be loaded directly onto this third phase and desalted "online" during the first cycle of the MudPIT. While 1D, 2-phase MudPIT and 3-phase MudPIT columns are routinely used for different purposes in the lab, an actual head-to-head comparison was needed. We present here a comparison of the relative efficacies of LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT on a complex mixture of proteins associated with bovine brain microtubules (Fig. 1).

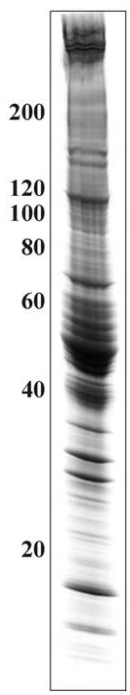

Fig. 1. Proteins associated with bovine brain microtubules. 50 μg of proteins associated with bovine brain microtubules (see Section 2) were TCA precipitated, resuspended in SDS–PAGE loading buffer, resolved on a 6–20% SDS–PAGE gradient gel, and visualized via Coomassie staining.

2. Materials and methods

Bovine microtubule associated proteins (MAPs) were produced using standard procedures as described in [14,15] except that following phosphocellulose chromatography to remove MAPs from the depolymerized tubulin, MAPs were eluted with three column volumes (1200 mL) of column buffer (CB; 50 mM K-Pipes, pH 6.8, 1 mM EGTA, 0.2 mM $MgCl_2$) containing 1 M KCl. We collected 6 mL fractions; the MAPs came out in one peak between fractions 38–70. MAPs were concentrated about two-fold (from 0.5 to 1.2 g/L) using YM10 Centriprep centrifugal filtration devices and dialyzed overnight against 20 mM Na-HEPES, pH 6.8, 5 mM $MgCl_2$, 1 mM Na-EGTA, 150 mM NaCl at 4 °C. Prior to digestion, proteins were precipitated using tricloroacetic acids (TCA) [16].

Protein digestion was performed as described in [7] with modifications. Briefly, 1 mg of precipitated proteins were resuspended in 200 μL of 8.0 M urea, 100 mM tris pH 8.5, reduced and alkylated using 3 mM TCEP (Sigma) for 20 min and 10 mM IAA (Sigma) for 30 min, and then digested for 4 h at 37 °C using 5 μg of endonuclease Lys-C (Roche). The mixture was then diluted to 2.0 M urea, 100 mM tris pH 8.5, 100 μL of Porozyme trypsin beads (PE Biosystems) were added, and the proteins were digested with shaking overnight at 37 °C. The peptide mixture was acidified by adding formic acid to 5% final volume.

1D LC-MS/MS was performed as in [3] with changes to buffers and RP material as previously described for the MudPIT columns [13]. Using a pressure cell, 6.5 cm of 5 μm Zorbax-XDB C-18 (Agilent) was packed into a 100 μm internal diameter fused silica column whose tip had been pulled to approximately 5 μm internal diameter (Fig. 2A) using a Sutter P-2000 (Sutter Manufacturing). Peptides from 15 μg of MAPs were loaded directly onto the column using a pressure cell. For analysis the column was placed in front of the heated capillary opening of an LCQ ion trap mass spectrometer (Thermo-finnigan) and voltage was applied using a liquid junction to facilitate electrospray (Fig. 2) [17]. Peptides were

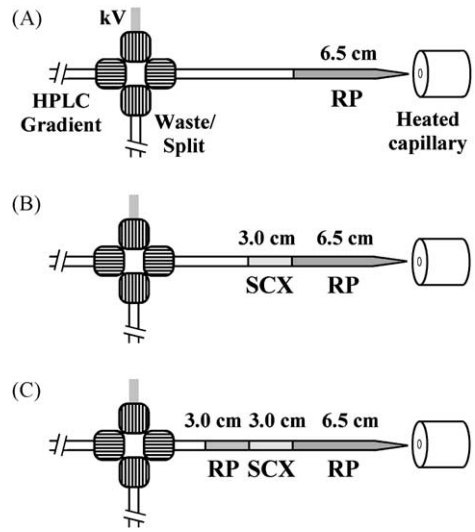

Fig. 2. General diagrams of 1D LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. (A) 1D LC-MS/MS [15]. An HPLC gradient is delivered to a junction where the voltage is applied and a proportion of the flow is split off to reduce the flow rate through column to approximately 300 nL/min. 6.5 cm of 5 μm C-18 material (RP) were packed into a 100 μm internal diameter fused silica cappilary whose tip had been pulled down to approximately 5 μm. (B) 2-Phase MudPIT column [11]. Configuration is identical to 1D column except that 3.0 cm of strong cation exchange (SCX) was placed upstream of the RP phase. (C) 3-Phase MudPIT column. The column is identical to the 2-phase MudPIT column except that an additional 3.0 cm of RP material is placed upstream of the SCX phase for the purpose of "online" desalting.

resolved with a linear gradient from 100% buffer A (5% ACN, 0.05% HFBA) to 80% buffer B (80% ACN, 0.05% HFBA) for 2 h (Fig. 3A). A flow rate at the tip of the column of approximately 200 nL/min was accomplished by adjusting the restriction from the flow splitter. MS/MS data were collected in a data dependent manner using settings described in [13].

2-phase MudPIT analysis was performed as described previously [13] with slight modifications. The column was identical to the 1D column except that 3 cm of SCX resin (Paritsphere SCX, Waters) was placed upstream of the RP section (Fig. 2B). Before loading, 15 μg of peptides were desalted using a C-18 solid phase extraction cartridge (C-18 SPEC, Ansys) [13]. For the 3-phase MudPIT column an additional 3 cm of RP material was placed upstream of the SCX

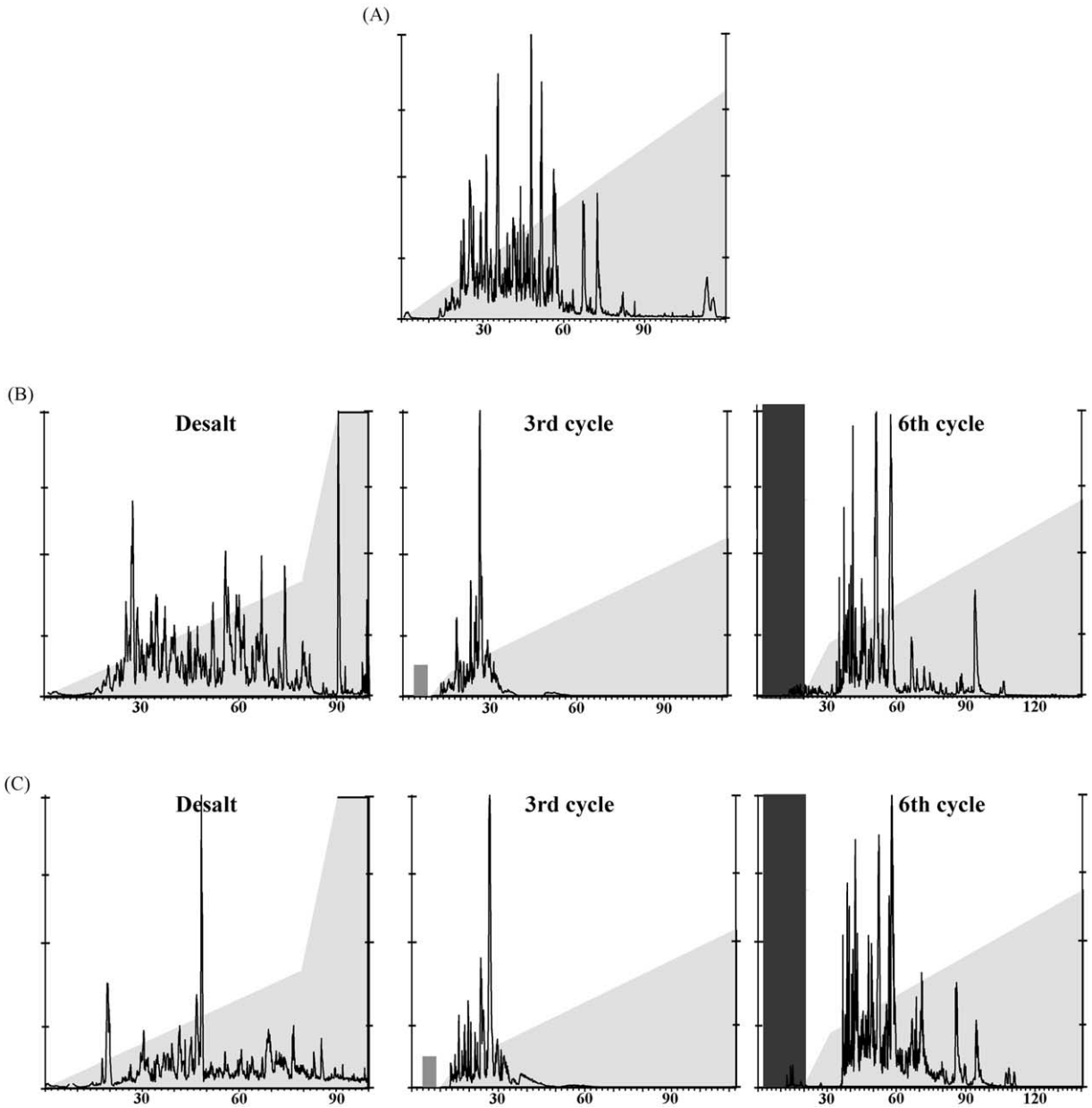

Fig. 3. Peptide chromatograms and diagrams of chromatographic conditions. (A) 1D LC-MS/MS. Shows a base-peak peptide chromatogram and the gradient profile with the percentage of buffer B (Section 2) indicated by grey shading. (B) 2-Phase MudPIT. As in A except that profiles for three out of the six steps are included. The darker grey box indicated percentage of buffer C (third cycle—Section 2) and black shaded box indicated percentage of buffer D (sixth cycle—Section 2). Elution cycles for second, fourth, and fifth steps were as for the third cycle except with differing percentages of buffer C. (C) 3-Phase MudPIT. As in B except for 3-phase MudPIT analysis.

(Fig. 2C). Both MudPIT experiments used six elution steps (Fig. 3B and C). The first was a gradient from 0 to 70% buffer B (see above) which served as a "desalting" step. The next four cycles were as described in [13] with 2 min salt steps of 10, 25, 50, and 80% buffer C (250 mM ammonium acetate, 5% ACN, 0.05% HFBA) prior to each cycle of RP separation. The final cycle includes a longer (20 min) salt bump of 100% buffer D (500 mM ammonium acetate, 5% ACN, 0.5% HFBA). Electrospray setup and data collection settings were as in the 1D LC-MS/MS analysis.

MS/MS data were searched against a human subset database from NCBI using SEQUEST [1,6]. Search results were filtered, combined by locus, and redundant locus information removed using the DTASelect program [20]. A PERL script, HPLCpH34, was written to calculate predicted hydrophobic retention of peptides from the various experiments based on values for individual amino acids from [16]. Statistical comparison was performed using the two-tail *t*-test function in Excel (Microsoft).

3. Results and discussion

To test the relative performance of 1D LC-MS/MS, 2- and 3-phase MudPIT we needed a sample that was sufficiently complex and had enough of a range of relative protein amounts to "challenge" each of the systems without unduly biasing towards a specific technique (see discussion below). For this we chose to use a set of proteins that were purified based on their association with bovine brain microtubules [14,15]. Extensive analysis of this sample has shown that more than 200 proteins are present at various concentrations in the mixture (Fig. 1 and data not shown).

For the comparison approximately 15 μg of digested protein were loaded directly onto a 1D LC-MS/MS column, offline desalted and loaded onto a 2-phase MudPIT column, or loaded directly onto a 3-phase MudPIT column (see Section 2 and Fig. 2). For the 1D run, peptides were resolved and data collected for 120 min (Fig. 2A). Both MudPIT experiments were identical six cycle runs of approximately 12 h (Fig. 2B and C). For the purpose of comparison, MS/MS data were searched against the human subset database from NCBI using SEQUEST, and these results filtered and compared using a uniform set of criteria with the DTASelect program [20].

The most important comparison among the three techniques was the total number of proteins that each returns. For this purpose it was necessary to generate a list of proteins that represents the minimum number of proteins consistent with the data. Depending on the database being searched additional numbers of proteins may be returned for a variety of reasons such as multiple entries for that protein in the database, partial coding sequences that are subsets of larger proteins, and closely related proteins for which the peptide data is inadequate for differentiation. Thus, even though there could be many sets of protein paralogs present within the sample, they were not counted separately for our analysis if there were not peptides specific for that particular protein. After the minimization filtering, the results were 26, 55, and 62 proteins for the 1D, 2-phase, and 3-phase experiments, respectively (Table 1).

Dramatic differences were observed between the 1D separation and the MudPIT runs, where over twice as many proteins are identified. These results were to be expected because of the greater resolving power of the MudPIT columns and the fact that data were collected over a six-fold greater time period. More differences were seen in the total number of peptides matching to these proteins. For 1D LC-MS/MS 147 individual peptides were found and for 2-phase and 3-phase MudPITs 341 and 431 peptides were found (Table 1). The analyses did not vary dramatically in the percentages of "true-tryptic" (both cleavages specific for trypsin) and "at least half-tryptic" (at least one cleavage specific for trypsin and the other potentially non-specific) peptides (Table 1). These data are consistent with other results from the lab showing that while approximately 80–90% of matched peptides contain at least one tryptic cleavage site, only 55–70% of matched peptides are fully tryptic peptides (data not shown) and illustrate the risk that one could throw away over one-third of the data from an experiment by limiting the search

Table 1
Proteins, peptides, and total spectra returned after filtering

	Number of proteins	Number of peptides	Total spectra matched
1D	26	147 (82%[a], 64%[b])	186
2-Phase MudPIT	55	341 (91%[a], 64%[b])	634
3-Phase MudPIT	62	431 (85%[a], 57%[b])	996

[a] Percent of "at least half-tryptic" peptides (one or both of the cleavages specific for trypsin).
[b] Percent of "true-tryptic" peptides (both cleavages specific for trypsin).

strategy to only peptides specific for the enzyme used. The "unexpected" proteolytic specificity could come from a variety of sources such as endogenous proteases that were active during the cell lysis/purification procedure or even from specific processing of these proteins as part of their functional context in the cell.

One possible reason for the differences in results between the two types of MudPIT columns was that a subset of peptides was lost from the mixture during the offline cleanup step. The peptides most likely to be lost were those not sufficiently hydrophobic to be retained on the C-18 SPEC cartridge. To test this possibility we compared the predicted reverse phase retention indices at pH 3.4 [18] for both the total peptides returned from each run and for the peptides that were found only in the 2-phase analysis. If more hydrophilic peptides were lost during the offline desalting, then the average hydrophobicity would be higher for peptides matched in the 2-phase experiment than for 3-phase experiment. If there were no bias for peptide loss during the offline cleanup then one would not expect a statistically significant difference between the groups of peptides from the various experiments.

The average hydrophobicity for peptides from the 3-phase run of 1.92 was markedly lower than either the average of the total peptides from the 2-phase run, 2.94, or those peptides only found in the 2-phase column, 3.01 (Table 2). Consistent with the hypothesis, the peptides that were only observed in the 3-phase experiment (possibly representing those peptides which were lost in the offline desalting) showed an even lower average hydrophobicity of 0.82 (Table 2). All of these differences were significant with $P \leq 0.01$ for two-tail *t*-tests between the data sets (Table 3).

Table 2
Average predicted hydrophobic retention index for the peptides at pH 3.4 [16]

	Average hydrophobicity	*n*
1D	3.11	147
Unique to 1D	1.45	13
2-Phase MudPIT	2.94	341
Unique to 2-phase MudPIT	3.01	128
3-Phase MudPIT	1.92	431
Unique to 3-phase MudPIT	0.82	188

None of these analyses was sufficient to return more than 200 proteins that previously have been shown to be present in the sample by analyses using more sample, a longer analysis time, and more comprehensive database searches (data not shown). We purposely did not load additional sample because for this comparison we did not wish to "stack the deck" in favor of one strategy. For this particular sample and amount, the

Table 3
P-values from two-tail *t*-test comparisons of predicted peptide hydrophobicity

	2-Phase MudPIT	Unique to 2-phase	3-Phase MudPIT
2-Phase MudPIT	–	–	–
Unique to 2-phase	0.868	–	–
3-Phase MudPIT	0.001	0.009	–
Unique to 3-phase	<0.001	<0.001	0.003

3-phase MudPIT column was superior both in number of proteins and number of peptides that were sampled. However, the choice of strategy depends on the complexity of the sample, the amount of starting material available, and the level of analysis needed. If it were only necessary to identify the most abundant proteins within this sample or the sample was significantly less complex, then 1D LC-MS/MS would be preferred because it is faster and more easily automated. The amount of sample, which can be loaded onto a 3-phase MudPIT column without offline desalting, is limited to around 100 μg without substantial risk of clogging the column (Yates lab, unpublished data). In many cases, especially when dealing with extremely complex mixtures, such as cell lysates, it is desirable to load more protein onto the column following the offline cleanup and extend the analysis for 12–18 cycles [13]. In this case the additional total protein and analysis time can more than make up for any subset of peptides that might be lost during the cleanup step. The opposite is true when sample quantity is limiting. Based on results from this study and our experience with a wide variety of protein mixtures, fewer peptide manipulation steps afford less chance for both general and specific peptide loss and makes the 3-phase MudPIT an excellent choice when a robust analysis is required on very small quantities of a protein mixture.

Acknowledgements

We would like to thank Claire Delahunty and Mike MacCoss for critical reading of the manuscript and other members of the Yates laboratory for fruitful discussion. W.H.M is supported by MERCK-MGRI-241 and UOW/RR11823-05. J.R.Y. is supported by RO1 EY1328801, MERCK-MGRI-241, CA81665, and RR11823. R.O is supported by NIH/NIGMS F32GM20309, and D.T.M is supported by a Howard Hughes Medical Institute predoctoral fellowship.

References

[1] J.K. Eng, A.L. McCormack, J.R. Yates III, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[2] J.R. Yates III, A.L. McCormack, A.J. Link, D. Schieltz, J. Eng, L. Hays, Analyst 121 (1996) 65R.
[3] A.L. McCormack, D.M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin, J.R. Yates III, Anal. Chem. 69 (1997) 767.
[4] A.J. Link, L.G. Hays, E.B. Carmack, J.R. Yates III, Electrophoresis 18 (1997) 1314.
[5] A.J. Link, E. Carmack, J.R. Yates III, Int. J. Mass Spectrom. 160 (1997) 303.
[6] J.R. Yates III, J.K. Eng, A.L. McCormack, D. Schieltz, Anal. Chem. 67 (1995) 1426.
[7] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates III, Nat. Biotechnol. 17 (1999) 676.
[8] B.L. Goode, J.J. Wong, A.C. Butty, M. Peter, A.L. McCormack, J.R. Yates III, D.G. Drubin, G. Barnes, J. Cell Biol. 144 (1999) 83.
[9] R. Verma, S. Chen, R. Feldman, D. Schieltz, J.R. Yates III, J. Dohmen, R.J. Deshaies, Mol. Biol. Cell 11 (2000) 3425.
[10] S. Meeusen, Q. Tieu, E. Wong, E. Weiss, D. Schieltz, J.R. Yates III, J. Nunnari, J. Cell Biol. 145 (1999) 291.
[11] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Proc. Natl. Acad. Sci. USA 97 (2000) 9390.
[12] M.T. Davis, J. Beierle, E.T. Bures, M.D. McGinley, J. Mort, J.H. Robinson, C.S. Spahr, W. Yu, R. Luethy, S.D. Patterson, J. Chromatogr. B Biomed. Sci. Appl. 752 (2001) 281.
[13] M.P. Washburn, D. Wolters, J.R. Yates III, Nat. Biotechnol. 19 (2001) 242.
[14] A.A. Hyman, T.J. Mitchison, Nature 351 (1991) 206.
[15] A. Hyman, D. Drechsel, D. Kellogg, S. Salser, K. Sawin, P. Steffen, L. Wordeman, T. Mitchison, Meth. Enzymol. 196 (1991) 478.
[16] J.J. Tasto, R.H. Carnahan, W.H. McDonald, K.L. Gould, Yeast 18 (2001) 657.
[17] C.L. Gatlin, G.R. Kleemann, L.G. Hays, A.J. Link, J.R. Yates III, Anal. Biochem. 263 (1998) 93.
[18] R. Cowan, R.G. Whittaker, Pept. Res. 3 (1990) 75.
[19] D.A. Wolters, M.P. Washburn, J.R. Yates III, Anal. Chem. 73 (2001) 5683.
[20] D.L. Tabb, W.H. McDonald, J.R. Yates III, J. Proteome Research 1 (2002) 21.