

Comparison of gel-aided sample preparation (GASP) and two in-solution digestion workflows for proteomic analysis using *Saccharomyces cerevisiae* lysate

Keywords: GASP, in-solution digestion, proteomic analysis, *Saccharomyces cerevisiae*.

Nashriq Mohammad Jailani¹, Asmahani Azira Abd Sani², Roman Fischer³, Honglei Huang³, Wan Mohd Aizat¹, Sheila Nathan⁴, Holger B. Kramer⁵, Benedikt M. Kessler³, Mukram Mohamed Mackeen^{1,6*}

¹*Institute of Systems Biology, ⁴School of Biosciences and Biotechnology, ⁶School of Chemical Sciences and Food Technology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*

²*Malaysian Genome Institute (MGI), Jalan Bangi, 4300, Bangi, Selangor, Malaysia*

³*Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7FZ United Kingdom*

⁵*MRC Clinical Sciences Centre, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, W12 0NN, United Kingdom*

¹Nashriq Mohammad Jailani: E-mail: nashriqjailani@yahoo.com, Tel: +6011-10566400

¹Wan Mohd Aizat Wan Kamaruddin: E-mail: wma@ukm.edu.my, Tel: +603-89214554

²Asmahani Azira Abdu Sani: E-mail: zira1304@gmail.com, Tel: +603-8926 7446 Ext: 1435

³Roman Fischer: E-mail: roman.fischer@ndm.ox.ac.uk, Tel: +44 (0) 1865612935

³Honglei Huang: E-mail: honglei.huang@ndm.ox.ac.uk, Tel: +44 (0) 1865612900

³Benedikt M. Kessler: Email: benedikt.kessler@ndm.ox.ac.uk, Tel: +44 (0) 1865612921

⁴Sheila Nathan: E-mail: sheila@ukm.edu.my, Tel: +603-89213862

⁵Holger B. Kramer: E-mail: h.kramer@lms.mrc.ac.uk, Tel: 020-8383-8411

* Corresponding author: E-mail: mukram.mackeen@ukm.edu.my, Tel/fax; +603-8921 5410

Comparison of Gel-Aided Sample Preparation (GASP) and two in-solution digestion workflows for proteomic analysis using *Saccharomyces cerevisiae* lysate

Abstract

Using *Saccharomyces cerevisiae* lysate, two in-solution trypsin digestions (chloroform-methanol-water precipitation and RapiGest) were compared to the recently reported gel-aided sample preparation (GASP) workflow. Our proteomic results showed that GASP afforded the highest number of overall protein identifications and peptide spectrum matches without systematic bias towards peptide or protein size. (49 words)

INTRODUCTION

The entire set of proteins in cells or organisms, termed the proteome (derived from the combination of “protein” and “genome”), can be analysed by mass spectrometry (MS) using a systems biology approach (Wilkins, et al., 1996; Wilkins, et al., 1996). Besides the identification of proteins as biomarkers (Hanif et al, 2018; Crutchfield, et al., 2016) and those involved in cellular responses (Walther & Mann, 2010), MS has also found application in the functional characterisation of proteins via activity-based profiling (Zweierink et al., 2017).

“Bottom-up” shotgun proteomics is based on the analysis of a mixture of proteins through the identification of peptides released via proteolysis. The overall shotgun proteomics strategy comprises the following sequence of steps: (1) generation of peptide fragments from a mixture of proteins, (2) chromatographic or gel-based separation of proteins, (3) mass spectrometric analysis and (4) identification of proteins using bioinformatics. The experimental and theoretical (obtained from the *in silico* digestion of a protein database) mass spectra of the acquired peptide fragments are compared and scored using software such Mascot (Perkins et al., 1999).

The enzymatic digestion of proteins for proteomic analysis commonly involves in-solution or in-gel methods. The in-gel workflow may include destaining of the gel, reduction, alkylation of cysteines, cleavage of proteins into peptides by enzymatic reaction, and finally the extraction of peptides from the gel (Rosenfeld et al., 1992). However, in-solution digestion is preferable to in-gel digestion because of its simpler workflow. Additionally, protein/peptide loss has been reported with in-gel digestion particularly during the extraction process (Speicher, et al., 2000).

Gel-aided sample preparation (GASP) was first developed by Fischer & Kessler (2015) as an alternative method that combined in-solution and in-gel digestion methods. Recently, it has also been used for “deep proteome” analysis using a multistage approach called CHOPIN (Davis et al., 2017). The GASP method mainly involves the use of DTT as the reducing agent, copolymerization of proteins with monomeric acrylamide, shredding of the resulting gel plug into small pieces to increase surface area, proteolysis and peptide recovery. The alkylating step in GASP using monomeric acrylamide is quick and involves less processing steps compared to the alkylating reagent iodoacetamide that is typically used in standard digestion methods. This results in the expedient formation of cys-S- β -propionamide (PAM-cys) from cysteine residues. This substitution step reduces contamination by minimising contact with the sample making this method facile, sensitive and reproducible.

In this paper, we compare three different sample preparation techniques using a *Saccharomyces cerevisiae* lysate sample for LC-MS/MS analysis. The sample preparation methods used are (i) the gel-assisted method GASP, and two in-solution based methods i.e. (ii) chloroform/methanol/water (CMW) precipitation and (iii) RapiGest™ (Waters protocol). GASP and CMW precipitation uses the chaotropic agent urea as the denaturing agent to disrupt protein structure and increase the accessibility of proteins to trypsin for more efficient proteolysis. The RapiGest™ method (commercially available reagent) was introduced to replace the use of SDS to avoid interference and problems with the MS analysis by: (1) obviating the difficulty associated with SDS

removal prior to MS analysis and (2) improving the efficiency of digestion. Based on our results, all three techniques showed no systematic bias towards peptide and protein mass. The percentage of missed cleavages in all the methods was low to indicate good trypsinolysis. The GASP method showed the highest protein number and peptide recovery, therefore, suggesting it to be the better than the in-solution methods (CMW precipitation and RapiGest).

MATERIALS AND METHODS

Preparation of Microbial Lysate

S. cerevisiae wild type strain was cultured on agar medium YPD (Difco) and incubated overnight at 37 °C. A single colony was then inoculated in 5 mL of YPD broth and was grown at 30 °C and 49 x *g* for 2 days. Cells were harvested through centrifugation at 2040 x *g* for 20 minutes and the cell pellet was collected. The cell pellet was washed with ice-cold phosphate-buffered saline (PBS) solution three times before snap-freezing with liquid nitrogen. The snap-freezing step was repeated 3 times before the pellet was lysed with glass beads in buffer containing 50 mM Tris-HCl (pH 7.5) and 100 mM NaCl (glass beads to buffer = 1:1). The solution was vortexed 6 times for 30 seconds with 1 minute intervals on ice followed by centrifugation at 8200 x *g* for 30 minutes. The supernatant was collected and the total protein concentration was determined using the bicinchoninic assay (BCA).

In-solution Digestion Using Chloroform-Methanol-Water (CMW) Precipitation

This commonly used standard procedure was carried out as described previously (Wessel & Flugge, 1984). Briefly, *S. cerevisiae* lysate (100 µg) was added with DTT to a final concentration of 5 mM and vortexed. The sample was incubated for 30 minutes at room temperature. Iodoacetamide was added to a final concentration of 20 mM. The resulting solution was vortexed and incubated for 30 minutes at room temperature. Proteins in the samples were precipitated via methanol/chloroform extraction. Methanol (600 µL) was added to a sample volume of 200 µL followed by 150 µL

chloroform and the solution was vortexed. Milli-Q water (450 μ L) was added and the solution was centrifuged at 12000 x g for 1 minute at room temperature. The upper aqueous phase was removed using a pipette without disrupting the precipitate at the interface. Methanol (450 μ L) was added, vortexed and then centrifuged for 2 minutes at room temperature. The supernatant was carefully removed and the protein pellet was resuspended in 6 M urea buffer. The solution was sonicated and vortexed. The concentration of urea was reduced to a final concentration of < 1M by diluting the reaction mixture with 250 μ L Milli-Q water and then vortexed again. Trypsin was added at the enzyme: total protein ratio of 1:50 (w/w). The solution was mixed carefully and digested overnight at 37 °C. The next day, sample was subjected to a standard C18 zip tip clean-up step.

In-solution Digestion Using RapiGest™ (Waters Protocol)

This procedure was carried out according to the manufacturer's instructions (Yu & Gilar, 2002) (<http://www.waters.com/webassets/cms/library/docs/720003102en.pdf>). RapiGest™ (Waters) (1 mg) was resuspended in 500 μ L of 50 mM ammonium bicarbonate resulting 0.2 % (w/v) RapiGest solution. *S. cerevisiae* lysate (100 μ g) was dissolved in 50 μ L of 0.2 % RapiGest solution and was vortexed. DTT was added to the protein sample solution to a final concentration of 5 mM. The sample was boiled at 60 °C for 30 minutes and was left to cool at room temperature. Iodoacetamide was added to the sample to a final concentration of 15 mM and the sample was then placed in the dark for 30 minutes. The trypsin digestion step was the same as described above followed by a C18 zip tip clean-up step carried out the next day.

Digestion using Gel-Aided Sample Preparation (GASP)

This procedure was carried as described previously by Fischer and Kessler (2015). DTT (50 mM) was added to *S. cerevisiae* lysate (100 μ g) and left at room temperature for 20 minutes. An equal volume of Protogel (40% w/v, acrylamide:bisacrylamide solution) was added, mixed by gentle

pipetting and left at room temperature for 20 minutes. Solutions of 10 % APS (5 µl) and TEMED (5 µl) were then added. The sample was left to polymerise for 10 minutes until the gel was solid and the resulting gel plug transferred to a filter centrifuge tube. The filter support was used to cut the gel pieces into small cubes by pulse centrifugation. The gel pieces were then fixed by adding 1 ml of methanol/acetic acid/water (50/40/10) for 10 minutes and vortexed. After pulse centrifugation, the supernatant was discarded, 500 µl of 6 M urea was added and the gel pieces were washed for 10 minutes using a rotator. Acetonitrile (1 ml) was added to dehydrate the gel pieces. The dehydration step was followed by rehydration in 50 mM triethylammonium bicarbonate (TEAB) (500 µl) and rotated for 10 minutes. The gel pieces were dehydrated by the addition of 1 ml acetonitrile. After discarding the supernatant, the gel pieces were dried further by adding 500 µl acetonitrile until the gel pieces aggregated at the bottom of the tube. Trypsin solution (enzyme: protein concentration = 1:50) of the same volume as the original gel plug was added to the dried gel pieces for overnight digestion at 37 °C. The next day, peptide extraction was started with the addition of 1 ml of acetonitrile. The supernatant was transferred into a new tube and the remaining gel pieces were rehydrated in 5 % formic acid, followed by dehydration with 1 ml of acetonitrile. The supernatants were combined after further dehydration in 1 ml of acetonitrile. The samples were dried using a vacuum concentrator and resuspended in 0.1 % formic acid in 2 % acetonitrile (an optional C18 zip tip clean-up step was also included for the purpose of comparison with original GASP method).

Mass Spectrometric Analysis

Dry samples were reconstituted in 0.1 % formic acid in 2 % acetonitrile and desalted online using a trapping column. Samples were separated on an EASY-Spray Column (Acclaim PepMapTM C₁₈ 100 Å⁰, 2 µm particle size, 75 µm id x 25 cm) over 60 minutes using a gradient of 5 – 40% 0.1 % formic acid in acetonitrile at 300 nl/min. Survey scans were acquired with the Thermo ScientificTM Orbitrap FusionTM TribridTM Mass Spectrometer at a resolution of 120000 @ 310-1800 m/z in the 3s Top

Speed Mode where precursors were selected for a maximum 3 second cycle. Only 20 of the most abundant precursors were selected for CID fragmentation.

Data Analysis

For relative label-free data analysis, raw data files were analysed using the LC Progenesis software (Non-Linear Dynamics). In brief, raw data files were imported and aligned based on chromatography profiles. Molecular features defined by accurate mass and retention times were detected and their corresponding MS/MS spectra submitted to a Mascot search. Peptide identifications were reimported into the LC Progenesis software and filtered for unique matches. Data from the in-solution methods and GASP were compared using LC Progenesis at a false discovery rate of 1 %, mass deviation of 10 ppm for MS1 and 0.6 Da for MS2 spectra.

RESULTS AND DISCUSSION

S. cerevisiae lysate (100 µg) was prepared for proteomic analysis using three different methods: (1) in-solution CMW precipitation: (2) in-solution RapiGestTM (Waters protocol), and with (3) a gel assisted method called gel-aided sample preparation (GASP). Samples from all three methods were prepared in triplicate and analysed using LC-MS/MS. Signal intensities of identified peptides and proteins of the MS data summed in mass bins were extracted using LC Progenesis.

Table 1 and Figure 1 show the results of sample quality and Figure 2 indicates sample peptide abundance based on signal intensity. All identified peptide ions found in the samples were plotted in Figure 1 as the mass error against the highest scoring peptide ion m/z that was identified (Figure 1). Our results showed that the CMW and GASP methods gave low peptide ion SD values (most peptide ions < 1 SD) while peptide ions from the RapiGest method had higher SDs.

Mascot analysis of our data indicated differences in total peptides (based on peptide spectrum matches, PSMs) and the proteins identified for the three methods. GASP showed the highest total

number of PSMs (10027 ± 9.7) than the in-solution methods CMW (10676 ± 6.1) and RapiGest (10027 ± 9.7) that were both comparable. However, there were marked differences in the number of unique PSMs for all three methods in the following order: GASP (6135 ± 4.0), CMW (2242 ± 5.1) and RapiGest (1627 ± 7.9). Differences between the three methods were more noticeable for the number of identified proteins with GASP and RapiGest showing the highest and lowest values, respectively. The proteolysis step appears to be effective for all the three methods as reflected by the high percentage of non-missed cleavages observed for all the methods. If the percentage of 0 or 1 missed cleavages was high, this would suggest a problem with trypsinisation which was not observed in our data set. Fewer missed cleavage sites were obtained with GASP compared to the in-solution digestion methods used in this study. This suggested that the GASP processed samples were better solubilised and protein cleavage sites were more accessible to trypsin. The proteomic results of the slightly modified GASP workflow that included an extra C18 clean-up step at the end of the protocol showed no marked difference to the unmodified GASP method (data not shown). This suggested that the original GASP protocol reported by Fischer and Kessler (2015) did not require this extra step that was used in the CMW and RapiGest protocols. The time taken to process all three digestion methods is similar for RapiGest and CMW (time before overnight trypsin digestion) i.e. ca.75 min whereas GASP is slightly longer (ca. 90 min with 1 wash cycle). Although both RapiGest and GASP are more suitable for the processing of a larger number of samples because both methods lack a precipitation step and involve less pipetting, the latter method will afford more peptide and protein identifications.

Based on Figure 2, we observed that the distribution of identified peptides did not show systematic bias towards peptide mass between GASP and the in-solution workflows (CMW and RapiGest) for the different groups of peptide masses. The summed ion counts of different peptides based on mass bins from the GASP and CMW methods showed better peptide recovery compared to the RapiGest method. The latter is less prone to introducing bias because of the absence of protein

immobilisation that is a problem associated with in-gel based digestion. The highest number of proteins acquired and peptides recovered in descending order was GASP followed by the in-solution methods (CMW precipitation and RapiGest). Therefore, none of the methods introduced a systematic bias towards peptide and protein size despite in-solution digestion being the most unrestricted (no filtering or immobilisation step involved). Figure 3 shows that GASP afforded the highest number of unique protein identifications compared to the in-solution methods. This is consistent with the results previously reported by Fischer and Kessler (2015). Our comparison of GASP and two common in-solution methods (CMW and RapiGest) showed GASP to be the best of the three sample preparation methods (especially with reference to the commercial RapiGest method) in terms of PSMs, overall protein identification, as well as peptide recovery and protein number based on different peptide mass bins.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Higher Education under Grant FRGS/1/2014/ST01/UKM/01/1. The authors thank the School of Chemical Sciences and Food Technology, School of Bioscience and Biotechnology, Institute of Systems Biology at Universiti Kebangsaan Malaysia and the Malaysia Genome Institute, Ministry of Science, Technology and Innovation Malaysia (MOSTI) for access to their facilities and technical support.

REFERENCES

- Crutchfield, C. A., Thomas, S. N., Sokoll, L. J. and Chan, D. W. (2016).** Advances in mass spectrometry-based clinical biomarker discovery. *Clinical Proteomics* **13**, 1.
- Davis, S., Charles, P. D., He, L., Mowlds, P., Kessler, B. M. and Fischer, R. (2017).** Expanding

proteome coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) combined with broad specificity proteolysis. *Journal of Proteome Research* **16(3)**, 1288-1299.

Fischer, R. and Kessler, B. M. (2015). Gel-aided sample preparation (GASP)-A simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. *Proteomics* **15(7)**, 1224–1229.

Hanif, Z., Sufiyan, N., Patel, M. and Akhtar, M. Z. (2018). Role of biobanks in transplantation. *Annals of Medicine and Surgery* **28**, 30-33.

Perkins, D. N., Pappin, D. J., Creasy, D. M. and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20(18)**, 3551–67.

Rosenfeld, J., Capdevielle, J., Guillemot, J. C. and Ferrara, P. (1992). Gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Analytical Biochemistry* **203**, 173–179.

Speicher, K. D., Kolbas, O., Harper, S. and Speicher, D. W. (2000). Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *Journal of Biomolecular Techniques* **11(2)**, 74–86.

Walther, T. C. and Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. *The Journal of Cell Biology* **190(4)**, 491–500.

Wessel, D. and Flügge, U. I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical Biochemistry* **138(1)**, 141–143.

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C. and Hochstrasser, D. F. (1996a). From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/Technology* **14(1)**, 61–65.

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F. and Williams, K. L. (1996b). Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews* **13(1)**, 19–50.

Yu, Y. Q. and Gilar, M. (2002). RapiGest SF Surfactant: An enabling tool for in-solution enzymatic protein digestions. *Waters Technical Note*. [Online]. Available from: <http://www.waters.com/webassets/cms/library/docs/720003102en.pdf>. (Accessed on March. 2018).

Zweerink, S., Kallnik, V., Ninck, S., Nickel, S., Verheyen, J., Blum, M. and Kaiser, M. (2017). Activity-based protein profiling as a robust method for enzyme identification and screening in extremophilic Archaea. *Nature Communications* **8(5)**.

FIGURES AND TABLE

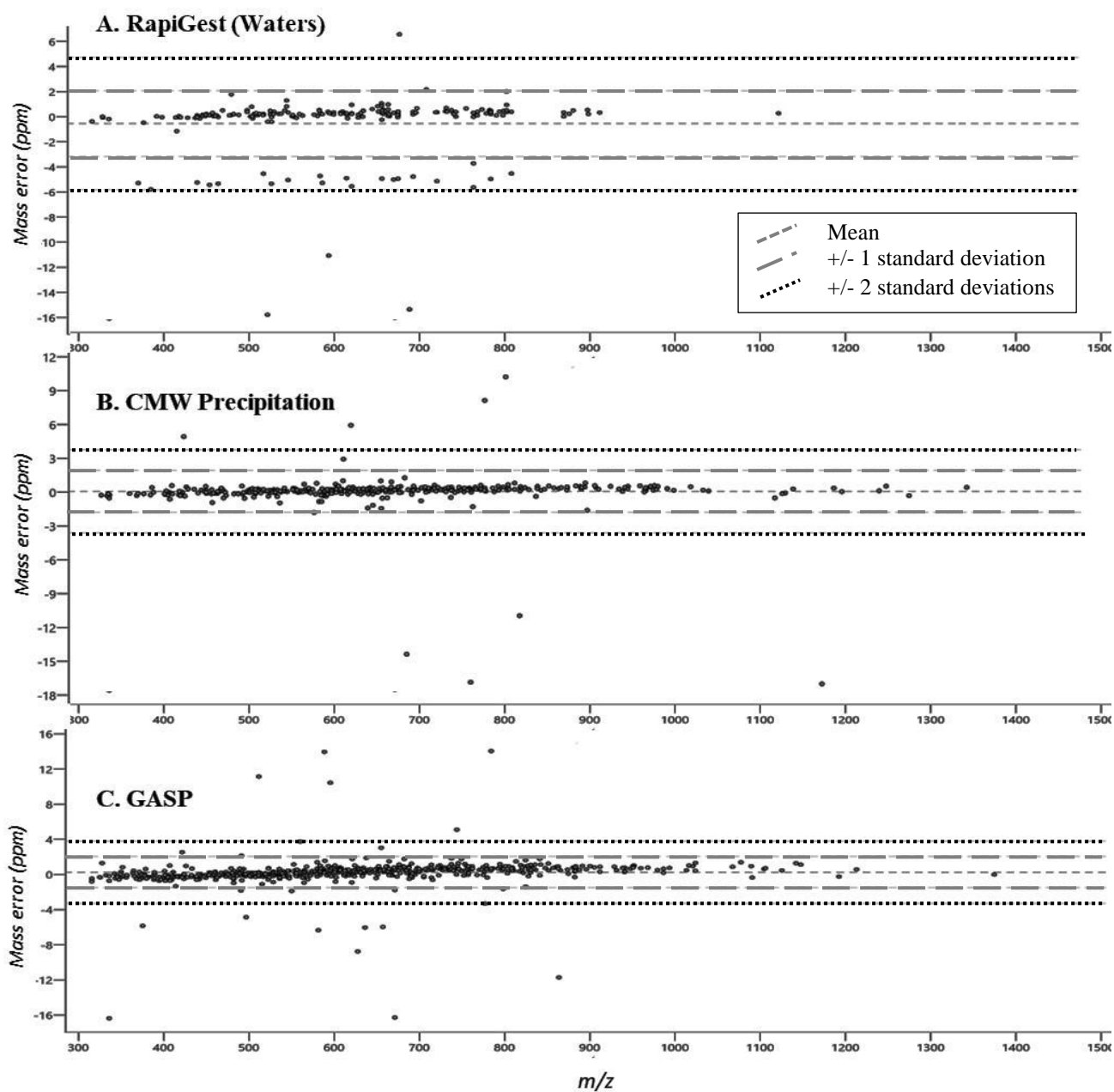
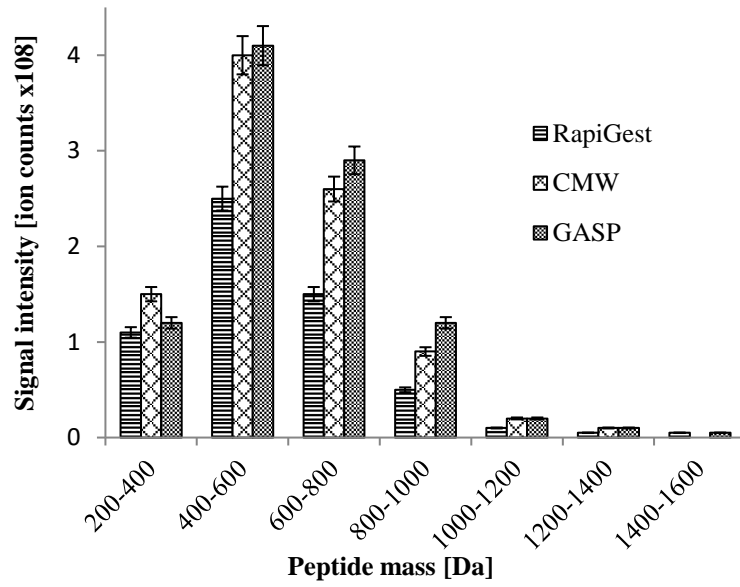


Figure 1: Mass accuracy of identified peptide ions of the (A) RapiGest (B) CMW and (C) GASP methods. Mass error (ppm): Identified peptides are represented as points and the lines represent mean ($-$) and standard deviations (SD) values, ($-$): ± 1 SD; (\cdots): ± 2 SD.

A.



B.

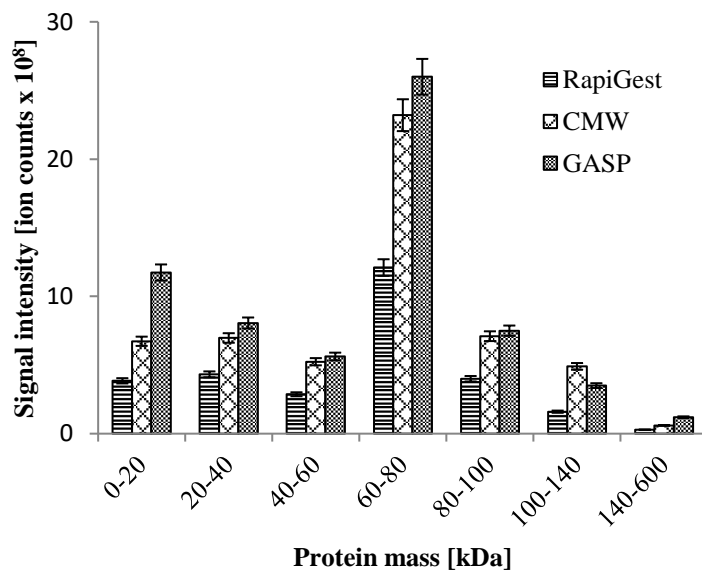


Figure 2: Comparison of (A) peptides identified and (B) protein identified after GASP and in-solution digestions (CMW precipitation and RapiGest).

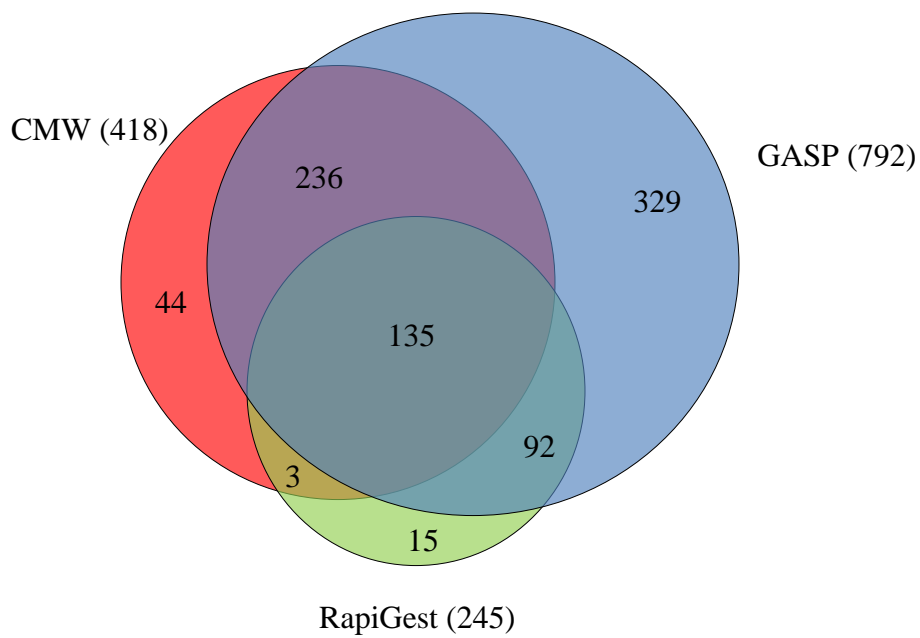


Figure 3: Number of identified proteins and their overlap using GASP, CMW and RapiGest digestions.

Table 1: Comparison of data characteristics after LC-MS/MS analysis of *S. cerevisiae* lysate using different sample preparations

	Sample preparation		
	RapiGest	CMW	GASP
# of MS/MS	22544 \pm 8.8	28957 \pm 6.3	30744 \pm 6.2
PSMs (Mascot, Score >20)	10027 \pm 9.7	10676 \pm 6.1	13339 \pm 3.0
Unique PSMs (Mascot)	1627 \pm 7.9	2242 \pm 5.1	6135 \pm 4.0
Proteins (grouped, Mascot)	245 \pm 4.1	418 \pm 3.5	792 \pm 2.5
0 missed cleavage	82.1 \pm 0.5 %	87.2 \pm 0.3 %	90.7 \pm 0.1%
1 missed cleavage	14.4 \pm 0.2 %	10.2 \pm 0.1 %	5.1 \pm 0.1 %