Evaluation of sample preparation methods for label-free quantitative profiling of salivary proteome

Xi Zhang¹, Pawel Sadowski², Chamindie Punyadeera^{1*}

1. Saliva and Liquid Biopsy Translational Research Team, Institute of Health and Biomedical Innovation, School of Biomedical Science, Queensland University of Technology, Australia.
2. Central Analytical Research Facility, Queensland University of Technology, Australia

*Correspondence:

Chamindie Punyadeera

Saliva and Liquid Biopsy Translational Research Team, The School of Biomedical Sciences, Institute of Health and Biomedical Innovations, Queensland University of Technology, 60 Musk Avenue, GPO Box 2434, Brisbane QLD 4001, Australia.

Abstract

Saliva has become one of the more attractive body fluids for protein biomarker discovery studies because of its ease of collection, non-invasiveness and multiple samples can be collected from an individual at a single time point. With the development of modern data acquisition strategies, such as Sequential Window Acquisition of all Theoretical Mass Spectrometry (SWATH-MS), the quality of saliva sample preparation has become a crucial factor for a successful protein identification and quantification. Several sample preparation methods have been proposed, but there has been no systematic evaluation conducted to date that compared each of these methods. We have therefore, performed an extensive assessment using technical and biological repeats to evaluate the number of protein IDs and repeatability of three most commonly used techniques, in-solution digestion, filter-aided sample preparation (FASP) and in stage-tip (iST) digestion. We discovered that in the case of human saliva sample, FASP provided the highest number of proteins (human and microbial) identifiable from a pool saliva sample, and there were no significant differences in terms of repeatability among the three methods investigated.

Introduction

Human saliva contains a wealth of proteins/peptides and as such is gaining attention as an alternative diagnostics medium compared to more traditional methods of sampling, blood. Collecting saliva is easy, non-invasive and safe for the collector [1-4]. The evidence for protein transport across salivary acini cells and blood vessel endothelium cells, suggests that saliva can be used for monitoring systemic events [1].

Data-independent acquisition (DIA) mass spectrometry (MS) approaches, such as Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH-MS) available on Sciex instruments, enable unbiased biomarker discovery [5] and have been previously successfully integrated into basic and clinical proteomic research [6-8]. However, being label-free, sample preparation is the key to a successful quantitation. Various sample preparation methods have been utilized for salivary proteomics studies in the past. These include a) in-solution digestion which utilises chemical and

43 thermal denaturation and trypsin digestion using organic or inorganic solvents [9], b) filter-aided
44 sample preparation (FASP) [10], that relies on filtration of detergent that would interfere with trypsin
45 digestion and c) in-StageTip digestion (iST), [11] in which all sample preparation steps are carried out
46 in a single, enclosed volume to minimize sample losses. However, no systematic evaluation of the
47 efficacy of each of these methods for saliva sample preparation has been made. The aim of this study
48 was to evaluate the performance of three most common sample processing techniques (FASP [10],
49 iST [11] and in-solution digestion [9]) prior to protein quantitation using label-free quantitative
50 proteomics profiling of human saliva.

51 **Materials and Methods**

52 The study design is summarised in Fig 1. Human saliva samples were collected from 12 healthy
53 volunteers using signed consent form approved by The University Human Research Ethics Committee
54 (approval number: 1400000617). Participants were asked to refrain from eating and drinking (except
55 for water) for two hours before saliva collection. Resting whole mouth saliva was collected as
56 described previously [2, 12-14]. In brief, participants were asked to sit in a comfortable position and
57 to rinse their mouths with water. They were asked to tilt their heads down, pool saliva in the mouth
58 for 1 minute and expectorate into a 50 mL Falcon tubes kept on ice. Saliva samples were transported
59 on ice, aliquoted into 1.5 mL micro centrifuge tubes and stored at -80 °C within 30 minutes until
60 analysis. Upon analysis, 12 saliva samples were randomly divided into three groups and combined
61 together to form three pooled saliva biological replicates. Total protein concentrations of pooled saliva
62 samples were quantified by Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific Inc, Waltham, MA,
63 USA) as per the manufacturer's protocol. In brief, 2 µL of the samples were added into 23 µL of 1x
64 phosphate buffered saline (PBS), mixed with 200 µL of working BCA reagent in a 96 wells microplate,
65 and incubated for 30 minutes at 37°C. The absorbance was read at 562nm on a plate reader and the
66 concentration was calculated against a standard curve ranging from 20- 2,000 µg/mL.

67 The three pooled saliva samples each containing 12.5 µg of total protein were then processed using
68 the methods listed below:

69 1. Filter-Aided Sample Preparation (FASP) method

70 Samples were mixed with 15 µL of SDS-Tris buffer (4% sodium dodecyl sulphate (SDS), 100mM Tris-
71 HCl pH 8.5, 100 mM Dithiothreitol (DTT)) and 200 µL of DTT-Urea buffer (25 mM DTT, 8M urea in 100
72 mM Tris-HCl pH 8.5) within a 30 kDa Microcon YM-30 centrifugal filter device (Millipore) and incubated
73 at RT for 60 minutes on agitator. The buffers were then washed away by centrifuging the filters at
74 14,000 g at 21°C for 15 minutes. The filters were washed with 200 µL Urea buffer (8M Urea in 100
75 mM Tris-HCl pH 8.5) by centrifuging the filters at 14,000 g at 21°C for 15 minutes. For cysteine
76 alkylation, 100 µL of 50 mM iodoacetamide (IAM) in Urea buffer was added to each sample and
77 incubated in the dark for 20 minutes. The filters were then washed with 200 µL Urea buffer (8M Urea
78 in 100 mM Tris-HCl pH 8.5) by centrifuging the filters at 14,000 g at 21°C for 15 minutes. Two additional
79 wash steps with 100 µL of 100 mM ammonium bicarbonate were performed before adding proteomics
80 grade trypsin (Sigma-Aldrich, St. Louis, MI, USA) at 1:50 enzyme to protein ratio and digesting at 37 °C
81 overnight. The digested peptides were collected into a clean 1.5 ml microcentrifuge tubes by
82 centrifugation at 14,000 g at 21 °C for 15 minutes. The samples were dried using a Speed Vac vacuum
83 concentrator and reconstituted in 10 µL of 2% ACN/0.1% trifluoroacetic acid (TFA) and transferred
84 into STAGE tips for sample clean up.

85 2. In-StageTip (iST) method

86 Samples were mixed with 5x their volume of buffer A consisting of 6M Guanidinium chloride, 10 mM
87 TCEP, 40 mM 2-Chloroacetamide, 10% ACN in 100 mM Tris pH 8.5 and incubated at 21°C for 20
88 minutes. To guanidinium chloride concentration down, the sample was diluted 10 x using 10% ACN in
89 100 mM Tris-HCl pH 8.5. Proteomics grade trypsin (Sigma-Aldrich) at 1:50 enzyme to protein ratio was
90 added and digested at 37 °C. After digestion was complete, 330 µL of 4% TFA was added to the sample

for acidification. Insoluble debris was pelleted by centrifuging at 21 °C at 16,000 g for 10 minutes and the supernatant was transferred into STAGE tips for sample clean up.

3. In-solution method

Proteins in the samples were denatured and reduced by the addition of 100 µL of 1% SDS, 100 mM Tris-HCl buffer pH 8.0 and 10 mM DTT, and incubation at 95 °C for 10 min. Cysteine in the samples were then alkylated by addition of 30 mM IAM and incubating at RT for 30 min. Excess IAM was quenched with the addition of DTT to a final concentration of 10 mM. Proteins were then precipitated by addition of 800 µL of 1:1 methanol/acetone, incubation at -20 °C overnight and centrifugation at 18,000 × g for 10 min. The protein pellet was resuspended in 50 µL of 50 mM ammonium acetate and digested with 0.5 µg of proteomics grade trypsin (Sigma-Aldrich) at 37 °C for 16 h. The samples were dried using a concentrator and reconstituted in 10 µL of 2% ACN/0.1% TFA.

Clean up

All the samples from the above mentioned methods were subjected to clean up procedure using STAGE tips with double SCX membrane (Empore, part no: 2251, 3M, Maplewood, MI, US) [15]. In short, the STAGE tips were activated by passing through 30 µL of 100% ACN and equilibrated with 30 µL of 0.2% TFA. The acidified samples were loaded to the STAGE tips, followed by three subsequent washes with 30 µL of 0.2% TFA. The samples were eluted with 30 µL of 5% Ammonium hydroxide/80% ACN. The eluted peptides were dried using a concentrator and reconstituted with 10 µL 2% ACN/0.1% formic acid and transferred into polypropylene autosampler vial (Thermo Scientific).

DDA data acquisition and data processing

Peptide spectral data from approximately 400 ng – 1 µg of injected tryptic peptides per sample were generated using nanoLC-nanoESI-MS/MS instrument (TripleTOF® 5600+, SCIEX). Peptide spectral data were generated using nanoLC-nanoESI-MS/MS on a TripleTOF® 5600+ instrument (SCIEX). Peptides

were separated by performing reversed-phase chromatography using an Eksigent ekspert™ nanoLC 400 System directly interfaced to the MS/MS instrument. The LC platform was set up in a trap and elute configuration with a 10 mm × 0.3 mm trap cartridge packed with ChromXP C18CL 5 µm 120 Å material and a 150 mm × 75 µm analytical column packed with ChromXP C18 3 µm 120 Å (Eksigent Technologies, Dublin, CA). The mobile phase solvents were composed of mobile phase A: water/0.1 % FA; mobile phase B: ACN/0.1% FA; and mobile phase C: water/2% ACN/0.1% FA. Trapping was performed in mobile phase C for 5 min at 5 uL/min. Separation involved mobile phases A and B running at a conserved flowrate of 300 nL/min. and adjusting the proportions of solvents (98, 60, 35, 10, 10, 98 and 98 % of solvent A) at specified time-points (0, 30, 35, 40, 49, 50 and 60 min). To minimise retention time drift, the analytical column was maintained at 40°C.

After chromatography, peptides were analysed by data-dependent acquisition (DDA). The DDA mode of the instrument was set to obtain high resolution (30,000) TOF-MS scans over a range of 350-1350 m/z, followed by 40 high sensitivity MS/MS scans of the most abundant peptide ions per cycle over the range of 100- 2000 m/z. The selection criteria for the peptide ions included intensity greater than 150 cps and charge state of 2-5. The dynamic exclusion duration was set at 9 s. Each survey (TOF-MS) scan lasted 250 ms and the product ion (MS/MS) scan was acquired for 50 ms resulting in a total cycle time of either 2.3 s. The ions were fragmented in the collision cell using rolling collision energy, and collision energy spread (CES) was set to 5. An overlay of Total Ion Chromatograms are shown in Fig 2.

The collected peptide ion fragmentation spectra were stored in .wiff and .wiff.scan formats (SCIEX). Peptide identification was performed by the Protein Pilot 5.0 software (SCIEX) using the UniProt/Swiss-Prot database with the following setting: Sample Type, identification; cysteine alkylation, iodoacetamide or Iodoacetic acid; instrument, TripleTof 5600; Species, human; ID focus, Biological modification; enzyme, trypsin; Search effort, thorough ID. We downloaded the LudwigNR database (downloaded from <http://apcf.edu.au> as at 27 January 2012; 16 818 973 sequences; 5 891 363 821 residues) for human protein identification and the Human Oral Microbiome Database [16] for oral

microbial protein identification. Protein Pilot was used to perform false discovery rate (FDR) analysis for all searches. The search result with a confidence interval of 80% -99%, including False Discovery Rate (FDR) analysis were shown in Table 1. The lower limit of peptide identification confidence and cut off of global FDR was set at 99% and 1% respectively for further analysis.

SWATH-MS data acquisition and label-free quantitation

The LC conditions were the same as for DDA acquisition. Mass spectrometry acquisition followed the strategy described in Gillet *et al.* [17] with the following modifications. High-resolution TOF-MS scans were collected over a range of m/z of 350–1500 for 0.08 s followed by high-sensitivity TOF-MS/MS scans collected over a range of m/z 100-1800 on all peptide ions whose mass fell within a particular SWATH window and for 0.08 s per window. There were 36 overlapping windows, each 26 m/z units wide, that covered a range of m/z 350–1500 which resulted in a cycle time of 3s. Collision energy was automatically assigned by the Analyst software (SCIEX) based on m/z window ranges Protein library was generated with the DDA data as described above for the targeted data extraction and peptide quantification.. The abundance of peptides was measured using PeakView Software with standard settings essentially as described [18], summing the integrated areas of up to six fragment ions per peptide. Protein abundance was measured using the sum of the abundances of two to six peptides per protein. The accuracy of peak selection by PeakView was manually confirmed for proteins of interest in each sample.

Results

Proteome coverage

A Venn diagram depicts the unique proteins identified per sample preparation method as well as the impact of each sample preparation method on the human proteome coverage (Fig 3A). Grassi *et al* reported significant number of microbial proteins in saliva [19]. Therefore, on a separate Venn

162 diagram we are reporting the number of oral microbial proteins after grouping proteins together from
163 each species for the three sample preparations used (Fig 3B).

164 **Label-free quantitation**

165 Next, we evaluated the reproducibility of label-free quantitation resulting from these three methods
166 by calculating the variability of the protein intensity across three pooled saliva samples. Comparing
167 the reproducibility of protein intensities generated by the three methods, there was no statistically
168 significant differences (as evaluated using repeated measures ANOVA, Fig 4)

169 **Discussion**

170 The total number of human proteins detected by FASP, iST and in-solution methods with 99%
171 confidence were 488, 230 and 130 respectively. The proteins covered by FASP, iST method and in
172 solution methods were 94%, 92% and 92% respectively. In addition, we determined FASP, iST and in
173 solution methods identified 206, 153 and 81 oral microbial protein with 99% confidence respectively.
174 With respect to protein identification, the results demonstrated that the FASP saliva sample
175 preparation method yielded a higher number of both human and microbial proteins. The
176 reproducibility of FASP protocol was also comparable to other methods, making it an ideal sample
177 processing method for protein biomarker discovery research using saliva samples. Strong detergent
178 like SDS is a powerful solution to solubilise proteins, but can negatively influence protein digestion
179 and mass-spectrometric measurement even in very small concentration. With the filter device used in
180 this method, SDS is efficiently removed and poses no threat to the downstream procedure and
181 analysis. FASP protocol also allows uniform representation of the samples' proteome, without the
182 need to separate it into insoluble and supernatant fractions. This potentially contributes to the larger
183 number of proteins detected in our samples. Even though the filter device used in the FASP method
184 could increase the total cost of the proteomics studies, the reliability and higher performance of the
185 FASP method still makes it an ideal method for processing saliva samples. Recently, plate-based FASP

protocol has been introduced [20] which has a potential to significantly reduce the cost of FASP digestion. In conclusion, saliva samples processed using the FASP method provides a much higher number of identifiable proteins compared to both iST method and in solution method. Both iST and in solution methods only provide a small number of unique identifiable proteins compared to FASP method. Our results suggest that FASP method of saliva sample processing is superior label-free quantitative technique to profile the saliva proteome. Recently, new sample processing techniques have been developed with the focus of efficient SDS removal for label free quantitative mass spectrometry. Both Suspension trapping (S-trap) [21] and Single-Pot Solid -Phase-enhanced Sample Preparation (SP3) [22] were developed and evaluated against FASP for their efficiency of sample processing respectively [23, 24]. It was found that both of these techniques provide better proteomics coverage and quantification consistency in comparison to FASP. However currently these techniques have not been applied to saliva sample processing. Sequential window acquisition of all theoretical fragment ion spectra mass spectrometry (SWATH-MS) is recently developed and a popular label-free quantitative strategy that provides fast and reliable quantitative information of sample's proteome. Due to its advantage over the traditional methods, SWATH-MS has attracted interest from the saliva research community. The finding in this study are directly translatable to SWATH-based quantitation.

Acknowledgements

This study was supported by the Heart Foundation Vanguard Grant Scheme and the Royal Brisbane Women Hospital (RBWH) Research Foundation.

References

- [1] T. Pfaffe, J. Cooper-White, P. Beyerlein, K. Kostner, C. Punyadeera, Diagnostic potential of saliva: current state and future applications, *Clin. Chem.* 57(5) (2011) 675-87.
- [2] D.A. Ovchinnikov, M.A. Cooper, P. Pandit, W.B. Coman, J.J. Cooper-White, P. Keith, E.J. Wolvetang, P.D. Slowey, C. Punyadeera, Tumor-suppressor Gene Promoter Hypermethylation in Saliva of Head and Neck Cancer Patients, *Transl. Oncol.* 5(5) (2012) 321-6.
- [3] X. Zhang, Y. Wan, J. Cooper-White, G. Dimeski, J. Atherton, C. Punyadeera, Quantification of D-dimer levels in human saliva, *Bioanalysis* 5(18) (2013) 2249-56.
- [4] X. Zhang, G. Dimeski, C. Punyadeera, Validation of an immunoassay to measure plasminogen-activator inhibitor-1 concentrations in human saliva, *Biochemia medica* 24(2) (2014) 258-65.
- [5] A. Mebazaa, G. Vanpoucke, G. Thomas, K. Verleysen, A. Cohen-Solal, M. Vanderheyden, J. Bartunek, C. Mueller, J.M. Launay, N. Van Landuyt, F. D'Hondt, E. Verschuere, C. Vanhaute, R. Tuytten, L. Vanneste, K. De Cremer, J. Wuyts, H. Davies, P. Moerman, D. Logeart, C. Collet, B. Lortat-Jacob, M. Tavares, W. Laroy, J.L. Januzzi, J.L. Samuel, K. Kas, Unbiased plasma proteomics for novel diagnostic biomarkers in cardiovascular disease: identification of quiescin Q6 as a candidate biomarker of acutely decompensated heart failure, *Eur. Heart J.* 33(18) (2012) 2317-24.
- [6] R.Y. Chang, N. Etheridge, A.S. Nouwens, P.R. Dodd, SWATH analysis of the synaptic proteome in Alzheimer's disease, *Neurochem. Int.* 87 (2015) 1-12.
- [7] I. Ortea, A. Rodriguez-Ariza, E. Chicano-Galvez, M.S. Arenas Vacas, B. Jurado Gamez, Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction, *J. Proteomics.* 138 (2016) 106-14.
- [8] G. Hou, X. Lou, Y. Sun, S. Xu, J. Zi, Q. Wang, B. Zhou, B. Han, L. Wu, X. Zhao, L. Lin, S. Liu, Biomarker Discovery and Verification of Esophageal Squamous Cell Carcinoma Using Integration of SWATH/MRM, *J. Proteome Res.* 14(9) (2015) 3793-803.
- [9] Y. Xu, U.M. Bailey, C. Punyadeera, B.L. Schulz, Identification of salivary N-glycoproteins and measurement of glycosylation site occupancy by boronate glycoprotein enrichment and liquid chromatography/electrospray ionization tandem mass spectrometry, *Rapid Commun. Mass Spectrom.* 28(5) (2014) 471-82.
- [10] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nat. Methods* 6(5) (2009) 359-62.
- [11] N.A. Kulak, G. Pichler, I. Paron, N. Nagaraj, M. Mann, Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells, *Nat. Methods* 11(3) (2014) 319-24.
- [12] M. Navazesh, Methods for collecting saliva, *Ann. N. Y. Acad. Sci.* 694 (1993) 72-7.
- [13] E. Topkas, P. Keith, G. Dimeski, J. Cooper-White, C. Punyadeera, Evaluation of saliva collection devices for the analysis of proteins, *Clin. Chim. Acta* 413(13-14) (2012) 1066-70.
- [14] C. Punyadeera, G. Dimeski, K. Kostner, P. Beyerlein, J. Cooper-White, One-step homogeneous C-reactive protein assay for saliva, *J. Immunol. Methods* 373(1-2) (2011) 19-25.
- [15] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, *Nat. Protoc.* 2(8) (2007) 1896-906.
- [16] T. Chen, W.H. Yu, J. Izard, O.V. Baranova, A. Lakshmanan, F.E. Dewhirst, The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information, *Database : the journal of biological databases and curation* 2010 (2010) baq013.
- [17] L.C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, *Mol. Cell. Proteomics* 11(6) (2012) O111.016717.
- [18] L.F. Zacchi, B.L. Schulz, SWATH-MS Glycoproteomics Reveals Consequences of Defects in the Glycosylation Machinery, *Mol. Cell. Proteomics* 15(7) (2016) 2435-47.

- [19] N. Grassl, N.A. Kulak, G. Pichler, P.E. Geyer, J. Jung, S. Schubert, P. Sinitcyn, J. Cox, M. Mann, Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome, *Genome Med.* 8(1) (2016) 44.
- [20] J. Potriquet, M. Laohaviroj, J.M. Bethony, J. Mulvenna, A modified FASP protocol for high-throughput preparation of protein samples for mass spectrometry, *PLoS ONE* 12(7) (2017) e0175967.
- [21] A. Zougman, P.J. Selby, R.E. Banks, Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis, *Proteomics* 14(9) (2014) 1006-0.
- [22] C.S. Hughes, S. Foehr, D.A. Garfield, E.E. Furlong, L.M. Steinmetz, J. Krijgsveld, Ultrasensitive proteome analysis using paramagnetic bead technology, *Mol. Syst. Biol.* 10 (2014) 757.
- [23] K.R. Ludwig, M.M. Schroll, A.B. Hummon, Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies, *J. Proteome Res.* 17(7) (2018) 2480-2490.
- [24] M. Sielaff, J. Kuharev, T. Bohn, J. Hahlbrock, T. Bopp, S. Tenzer, U. Distler, Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range, *J. Proteome Res.* 16(11) (2017) 4060-4072.

Figure legends

Figure 1 Schematic workflow for the experiment.

Figure 2 Overlay of Total Ion Current chromatograms generated for samples processed by (A) FASP, (B) iST and (C) in solution method

Figure 3 Venn Diagram showing the overlap of proteins identified by FASP, iST and in solution method in (A) human and (B) oral microbiome

Figure 4 Percentage coefficient of variations of shared protein intensities between the three different processing methods

Table 1 Summary of the number of human Protein Identified using three methods

	FASP		iST		in solution	
Confidence	Protein ID		Protein ID		Protein ID	
> 99%	488		230		133	
> 95%	583		285		167	
> 90%	598		305		173	
> 80%	615		326		187	
FDR analysis	Protein	Peptide	Protein	Peptide	Protein	Peptide
1%	488	7538	219	4747	120	3253
5%	488	9312	230	5900	133	4032
10%	488	10560	230	6666	133	4572

A

B