

# Plasma proteome coverage is increased by unique peptide recovery from sodium deoxycholate precipitate

Aida Serra<sup>1</sup> · Hongbin Zhu<sup>1</sup> · Xavier Gallart-Palau<sup>1</sup> · Jung Eun Park<sup>1</sup> · Hee Haw Ho<sup>2</sup> · James P. Tam<sup>1</sup> · Siu Kwan Sze<sup>1</sup> 

Received: 25 November 2015 / Revised: 19 December 2015 / Accepted: 4 January 2016 / Published online: 25 January 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** The ionic detergent sodium deoxycholate (SDC) is compatible with in-solution tryptic digestion and LC-MS/MS-based shotgun proteomics by virtue of being easy to separate from the peptide products via precipitation in acidic buffers. However, it remains unclear whether unique human peptides co-precipitate with SDC during acid treatment of complex biological samples. In this study, we demonstrate for the first time that a large quantity of unique peptides in human blood plasma can be co-precipitated with SDC using an optimized sample preparation method prior to shotgun proteomic analysis. We show that the plasma peptides co-precipitated with SDC can be successfully recovered using a sequential resolubilization and precipitation procedure, and that this approach is particularly efficient at the extraction of long peptides. Recovery of peptides from the SDC pellet dramatically increased overall proteome coverage (>60 %), thereby improving the identification of low-abundance proteins and enhancing the identification of protein components of membrane-bound organelles. In addition, when we analyzed the physiochemical properties of the co-precipitated peptides, we observed that SDC-based sample preparation improved the identification of mildly hydrophilic/hydrophobic proteins that would otherwise be lost upon discarding the pellet. These data

demonstrate that the optimized SDC protocol is superior to sodium dodecyl sulfate (SDS)/urea treatment for identifying plasma biomarkers by shotgun proteomics.

**Keywords** Plasma proteome · Mass spectrometry · Shotgun proteomics · Sodium deoxycholate · Trypsin digestion

## Introduction

Characterizing the proteome of complex biological samples by LC-MS/MS-based shotgun proteomic analysis depends on a number of critical factors, including effective protein solubilization and denaturation, thorough protein to peptide digestion, and efficient recovery of tryptic peptides from the digestion buffer [1]. Together, these key steps determine the overall extent of proteome coverage, and each process is known to be critically influenced by the buffer used during sample preparation. The enzyme most commonly used for sample digestion is trypsin, which cleaves proteins at the C-terminal of lysine and arginine residues [2]; hence, buffer selection is typically restricted to reagents that are compatible with this enzyme. While several different detergents, chaotropes, and surfactants have been tested as protein-solubilizing agents, very few of these are fully compatible with shotgun proteomic sample preparation. Ideally, the proteome-solubilizing agent should also be capable of improving protein digestion and enhancing tryptic peptide recovery, thereby enabling high-confidence identification of proteins and post-translational modifications (PTM), as well as enhancing the performance of multiple reaction monitoring in targeted quantitative proteomic analysis [3–6].

The protein-solubilizing agent sodium deoxycholate (SDC) may be capable of significantly improving proteome coverage due to being fully compatible with the sample

**Electronic supplementary material** The online version of this article (doi:10.1007/s00216-016-9312-7) contains supplementary material, which is available to authorized users.

✉ Siu Kwan Sze  
SKSZE@NTU.EDU.SG

<sup>1</sup> School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore

<sup>2</sup> Department of Cardiology, Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, Singapore 308433, Singapore

preparation process as well as easy to remove from the digestion reaction via acid precipitation (AP) [7–9]. More importantly, SDC is able to enhance trypsin activity up to five-fold when used at concentrations ranging from 0.01 to 1 % [7], thereby obtaining an overall yield of tryptic peptides comparable with conventional urea-based methods [10, 11]. Removal of SDC from in-solution digestion (ISD) preparations can be achieved not only by AP but also by phase transfer (PT) to a non-miscible organic solvent phase added to the tryptic-digested sample [7]. When these techniques were directly compared during the analysis of rat liver mitochondria-enriched fractions isolated by ISD:SDC-based processing/extraction, the authors observed that the AP and PT methods identified comparable numbers of unique proteins, but the PT protocol identified ~11 % more unique peptides [10].

SDC has previously been reported to offer excellent denaturing and solubilizing performance when compared with other protease digestion enhancers including detergents, surfactants, bile salts, chaotropes, and various organic solvents [7, 10]. Indeed, in previous comparisons, SDC and RapiGest were determined to exhibit the highest solubilizing capacities of all MS-compatible reagents tested [7, 10]. While the strong ionic detergent sodium dodecyl sulfate (SDS) displays very high solubilizing capacity, this reagent is not compatible with MS [7]. The filter-aided sample preparation (FASP) method introduced by Mann and colleagues represents a universal strategy for sample processing that combines the capacity to remove low molecular weight components, as one of the main virtues of in-gel digestion (IGD), and the robustness of the ISD. The FASP method is a versatile strategy as it allows the ISD processing of special and generally MS-incompatible buffers, such as SDS, making their use possible in ESI-MS systems. SDC inclusion in FASP strategies has also been proposed [10, 12], showing better performance when SDC is coupled to SDS rather than urea [10]. Additionally, PT removal of deoxycholic acid in an enhanced FASP (eFASP) method was also introduced showing increased tryptic digestion efficiency for cytosolic and membrane proteins as compared to urea [12].

The potential utility of SDC for proteomic sample preparation has been studied intensively with respect to the analysis of moderately complex biological samples such as membrane protein-enriched fractions. However, the composition of the acid-precipitated SDC pellet for complex biological samples has yet to be determined in detail, despite the fact that this often-discarded material may contain co-precipitated tryptic peptides with potential to substantially increase proteome coverage. We therefore sought to determine whether unique tryptic peptides could be successfully recovered from the acid-precipitated SDC pellet of human plasma samples and assessed the potential impact of this approach on total proteome coverage. We now report an optimized alternative to urea-based ISD/sodium dodecyl sulfate polyacrylamide gel

electrophoresis (SDS-PAGE)-based in-gel digestion (IGD) methods that represents a highly efficient novel approach for the characterization of complex proteomic samples by LC-MS/MS.

## Materials and methods

### Reagents and chemicals

All reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA) unless otherwise indicated. Proteomics sequencing-grade modified trypsin was from Promega (Madison, WI, USA). Water (HPLC grade) and acetonitrile (ACN, HPLC grade) were purchased from Thermo Scientific Inc. (Bremen, Germany).

### Human plasma samples

Plasma samples were from anonymous human subjects recruited via Tan Tock Seng Hospital, Singapore, and stored at  $-150\text{ }^{\circ}\text{C}$  until use. Seven plasma samples (100  $\mu\text{L}$  each) were pooled (to reduce biological variation) and then used for optimization of the SDC-assisted ISD protocol. Twenty-seven additional plasma samples (25  $\mu\text{L}$  each) were then randomly divided into three groups (nine subjects per group) for assessment of the optimized method. The study was approved by the Research Ethics Committee of Tan Tock Seng Hospital and Nanyang Technological University, Singapore. Written informed consent was obtained from all study participants ( $n = 34$ ).

### Standard SDC-assisted and urea-assisted in-solution tryptic digestions

Conditions used for trypsin digestion of the plasma samples are summarized in Table 1. Briefly, plasma samples were processed according to the method described by Proc [11] except for minor modifications. Urea-based sample processing was carried out according to the protocol described by León [10], with minor modifications.

### Standard SDS in-gel tryptic digestion

SDS denaturation of proteins was coupled with IGD and performed as previously described (protocol workflow described in Table 1) [13, 14]. Proteins were resolved on a 12 % SDS-PAGE gel which was subsequently stained with Coomassie Brilliant Blue. Proteins were excised from the gel and bands were diced into 1-mm<sup>2</sup> cubes prior to reduction, alkylation, and digestion.

**Table 1** Summary of the reduction, alkylation, and digestion conditions used for each of the methods tested

Standard SDC-ISD	Standard urea-ISD	Standard SDS-IGD	Optimized SDC-ISD	
25 $\mu$ L human plasma				Sample
10-fold with ABB <sup>a</sup>	10-fold with ABB or AA <sup>b</sup>	–	10-fold with AA	Dilution
1 % SDC	8 M urea	1 % SDS <sup>c</sup> 5 min, 95 °C	1 % SDC	Denaturing agent
10 mM DTT	10 mM DTT	100 mM DTT	10 mM DTT	Reducing agent
30 min, 60 °C	3 h, 30 °C	1 h, 56 °C	30 min, 60 °C	Incubation
20 mM IAA	20 mM IAA	10 mM IAA	20 mM IAA	Alkylating agent
45 min, RT dark	45 min, RT dark	45 min, RT dark	45 min, RT dark	Incubation
2-fold with ABB, 10 mM DTT	<1 M urea	–	2-fold with AA, 10 mM DTT	Dilution
1:50	1:50	10 ng/ $\mu$ L	1:50	Trypsin ratio ( <i>w/w</i> )
30 °C, overnight				Incubation
0.5 % FA	0.5 % FA	Peptide extraction with 50:45:5 ( <i>v/v/v</i> ) ACN/water/acetic acid	0.5 % FA	Stop digestion

<sup>a</sup> ABB, 25 mM ammonium bicarbonate

<sup>b</sup> AA, 100 mM ammonium acetate

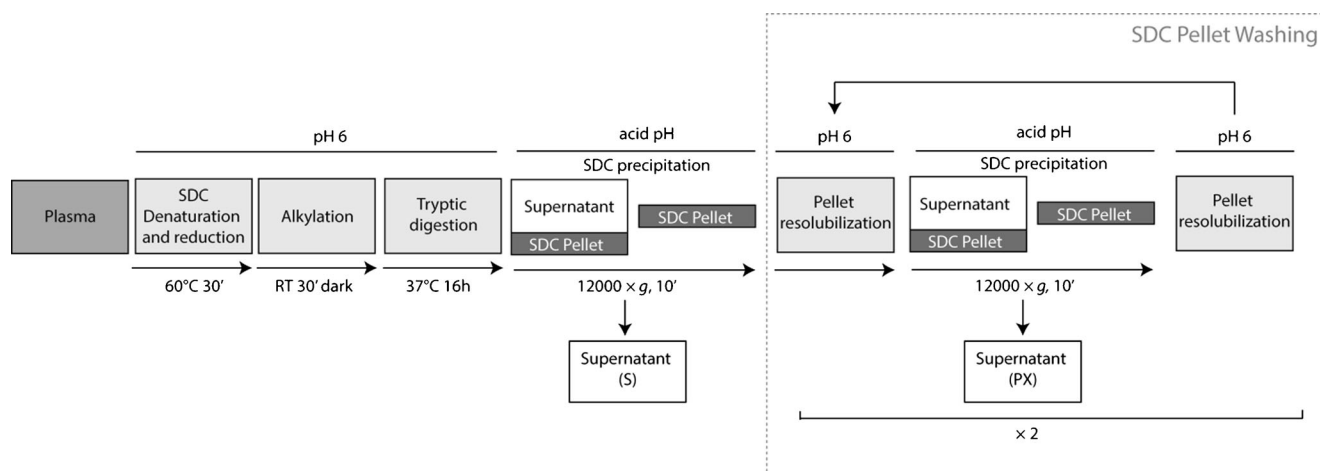
<sup>c</sup> 1 % SDS was prepared in 2 $\times$  Laemmli sample buffer

### Optimized SDC-assisted in-solution tryptic digestion

Reduction, alkylation, and tryptic digestion of proteins were performed as previously described [10], except for minor modifications (Table 1). Briefly, pooled plasma (25  $\mu$ L) was diluted 10-fold with 100 mM ammonium acetate (AA) and then denatured under reducing conditions using 1 % *w/v* SDC and 10 mM dithiothreitol (DTT) for 30 min at 60 °C. Samples were then alkylated with 20 mM IAA in the dark for 45 min. Prior to tryptic digestion, samples were diluted two-fold with 100 mM AA and 10 mM DTT and incubated at 37 °C for 30 min. Proteins were digested enzymatically overnight at 30 °C using sequencing-grade-modified trypsin at trypsin-to-substrate ratio 1:50 (*w/w*). The enzymatic reaction was quenched and the SDC precipitated simultaneously by addition of formic acid (FA) to a final concentration of 0.5 %. To further optimize SDC precipitation and peptide recovery, a sequential extraction of peptides from the SDC pellet was included (Fig. 1). The SDC precipitate was pelleted by centrifugation at 12,000  $\times$  *g* for 10 min, and the resultant supernatant (S) was collected into a new tube. The SDC pellet was then treated as follows: the pellet was re-dissolved in 600  $\mu$ L 0.5 % ammonium hydroxide, precipitated by addition of 0.5 % FA, and pelleted by centrifugation at 12,000  $\times$  *g* for 10 min. The washing process was repeated twice, resulting in two additional supernatants P1 and P2 (in reference to their pellets of origin). Samples were then desalted using Sep-Pak C18 cartridges (Waters, UK), and the eluted peptides were then dried using a vacuum concentrator. P1 and P2 fractions were analyzed separately during the protocol optimization process and subsequently combined to test the final optimized version of the protocol.

### LC-MS/MS

Peptides were desalted using a Sep-Pak 50-mg C18 cartridge (Waters, Milford, MA, USA) and eluted with 75 % ACN and 0.1 % FA prior to analysis by LC-MS/MS analysis. Eluted peptides were first concentrated using a vacuum concentrator (Eppendorf, Hamburg, Germany) and reconstituted with 3 % ACN and 0.1 % FA. Peptides (0.5  $\mu$ g) were separated and analyzed with a Dionex UltiMate 3000 UHPLC coupled to a linear quadrupole ion trap-Fourier transform Ultra apparatus (LTQ-FT Ultra, Thermo Scientific Inc.). Separation was performed using a reversed-phase Acclaim PepMap RSL column (75  $\mu$ m, ID  $\times$  15 cm, 2  $\mu$ m particle size, Thermo Scientific Inc.). Mobile phases were 0.1 % FA (phase A) and 80 % acetonitrile 0.1 % FA (phase B). Separation of samples was performed in a 240-min gradient of 3–6 % B for 2 min, 6–30 % for 208 min, 30–54 % for 14 min, 54–72 % for 1 min, 72 % for 5 min, 72–5 % for 2 min, and then initial isocratic conditions for 8 min. On-line ionization was performed using a Michrom CaptiveSpray ion source (Bruker-Michrom Inc., Auburn, USA) at an electrospray potential of 1.5 kV and an ion transfer tube temperature of 180 °C. Data acquisition was conducted in centroid and positive mode (350–1600 *m/z* range) in the FT-ICR cell at a resolution of 100,000 and maximum injection time of 1000 ms using Xcalibur version 2.0 SR2 (Thermo Scientific Inc., Bremen, Germany). The automatic gain control (AGC) target for FT was set to 1.0e+06, and precursor ion charge state screening was enabled. The 10 most intense ions with a 500-count threshold were fragmented by collision-induced dissociation, with a normalized collision energy of 35 %, activation *Q* of 0.25, isolation width for precursor ions of 2 Da, and activation time of 30 ms. MS/MS data



**Fig. 1** Flow chart outlining the novel SDC-assisted in-solution tryptic digestion method. *PX* represents the fraction obtained after completing either one (P1) or two (P2) washes of the SDC pellet

were acquired in the linear ion trap with AGC target for full MS of  $3.0 \times 10^4$  and  $MS^n$  of  $1.0 \times 10^4$  with maximum injection time of 200 ms. Peptide ions with 1+ charge were excluded for fragmentation. The dynamic exclusion list was enabled with a repeat count of 1 and exclusion duration of 40 s.

### Bioinformatics and data analysis

Raw MS/MS data were converted into Mascot generic format files using Thermo Proteome Discoverer software (version 1.4.1.14, Thermo Fisher Scientific Inc.). The concatenated target-decoy Uniprot human database (downloaded on 29th October 2013, 176,946 sequences and 70,141,034 residues) was used for data searching. Database searching and carbamylation analysis were performed using an in-house Mascot server (version 2.3.02, Matrix Science, Boston, MA, USA) with precursor MS and MS/MS tolerances of 10 ppm and 0.8 Da, respectively. Two missed trypsin cleavage sites of peptides were tolerated unless specified otherwise. Deamidation (Asn) and oxidation (Met) were set as variable modifications. The peptide/protein list obtained by Mascot was processed for further analysis using an in-house script together with Excel. Peptides with a Mascot score  $>20$  were used to generate the peptide/protein list for determination of false discovery rates ( $FDR = 2.0 \times \text{decoy hits}/\text{total hits}$ ). Peptides with same sequence were considered as unique peptides. The grand average of hydropathy (GRAVY), charge, and isoelectric point ( $pI$ ) of the identified peptides were calculated using in-house software. Venn diagrams were constructed using Venny 2.0.2.

ANOVA with Bonferroni correction for multiple comparisons was performed using GraphPad Prism 6 (GraphPad Software, Palo Alto, USA), and only  $p$  values  $<0.01$  were

considered significant. Results are reported as mean  $\pm$  standard error of the mean (SEM).

Gene enrichment analysis was performed with FunRich 2.1.2 [15].

### Proteomics data deposition

The mass spectrometry proteomics data have been submitted to the Proteome Xchange Consortium [16] via the PRIDE partner repository under the dataset identifier PXD002811.

## Results and discussion

### Comparison of classical proteomics methods for digestion of plasma samples

To evaluate the efficiency of the most common buffers used for shotgun proteomic sample preparation, we performed a systematic comparative analysis of human plasma sample processing using urea, SDS, or SDC. To do this, we used either 8 M urea or 1 % SDC for protein denaturation/ISD or performed a conventional SDS-based IGD protocol. In accordance with previous studies [17], we observed that the standard urea-assisted tryptic digestion method provided the highest proteome coverage, closely followed by the ionic detergent SDC (see Electronic Supplementary Material (ESM) Fig. S1).

Although several surfactants have been proposed as effective reagents for membrane protein denaturation and solubilization, the most suitable detergents SDS and CHAPS (3-[[3-cholamidopropyl] dimethylammonium]-1-propanesulfonate) are not compatible with most shotgun proteomic protocols

[18]. However, since SDS is one of the most efficient known solubilizing agents [7], there have been numerous attempts to remove this detergent from ISD preparations in order to render these suitable for MS analysis, most notably filter-aided sample preparation [19, 20] and off-line cation-exchange chromatography [21]. More commonly, SDS is used for polyacrylamide gel electrophoresis of proteins (SDS-PAGE), so we included this technique in our comparison of classical sample processing methods and sought to determine the relative efficiency of IGD versus that of ISD. The number of unique proteins identified using SDS-assisted IGD was comparable to the ISD protocols, but the number of unique peptides identified was substantially lower (47.4 % of the total unique peptides identified using the standard urea-based protocol), likely due to extensive peptide loss during sample extraction from the polyacrylamide gel [17, 22].

The routine use of SDC in MS sample preparation provides several major advantages over other more common techniques, and reportedly offers an efficient alternative to urea/SDS-based approaches to the analysis of moderately complex biological samples, such as cell culture-derived mitochondrial proteomes [10]. However, since current SDC-assisted ISD protocols deliver proteome coverage only similar to or lower than that obtained by urea, these techniques have been regarded as ill-suited to the analysis of highly complex human samples.

### Optimization of SDC-assisted in-solution tryptic digestion

When combined with a routine AP procedure, the use of SDC to prepare human plasma samples for shotgun proteomic analysis resulted in the identification of fewer unique proteins and peptides than did the standard urea-assisted ISD method. While this could be taken to indicate poor protein/peptide recovery following SDC-based processing, Masuda [7] has previously conjectured that unique peptides could in fact be co-precipitated with SDC during the AP process, and Lin et al. found co-precipitation of peptides in membrane protein samples [23]. We therefore hypothesized that SDC used in complex samples for shotgun proteomics involves loss of a substantial proportion of novel peptides that remain bound to the surfactant during SDC precipitation at low pH and could potentially be recovered from the SDC pellet by performing a sequential re-solubilization and precipitation procedure.

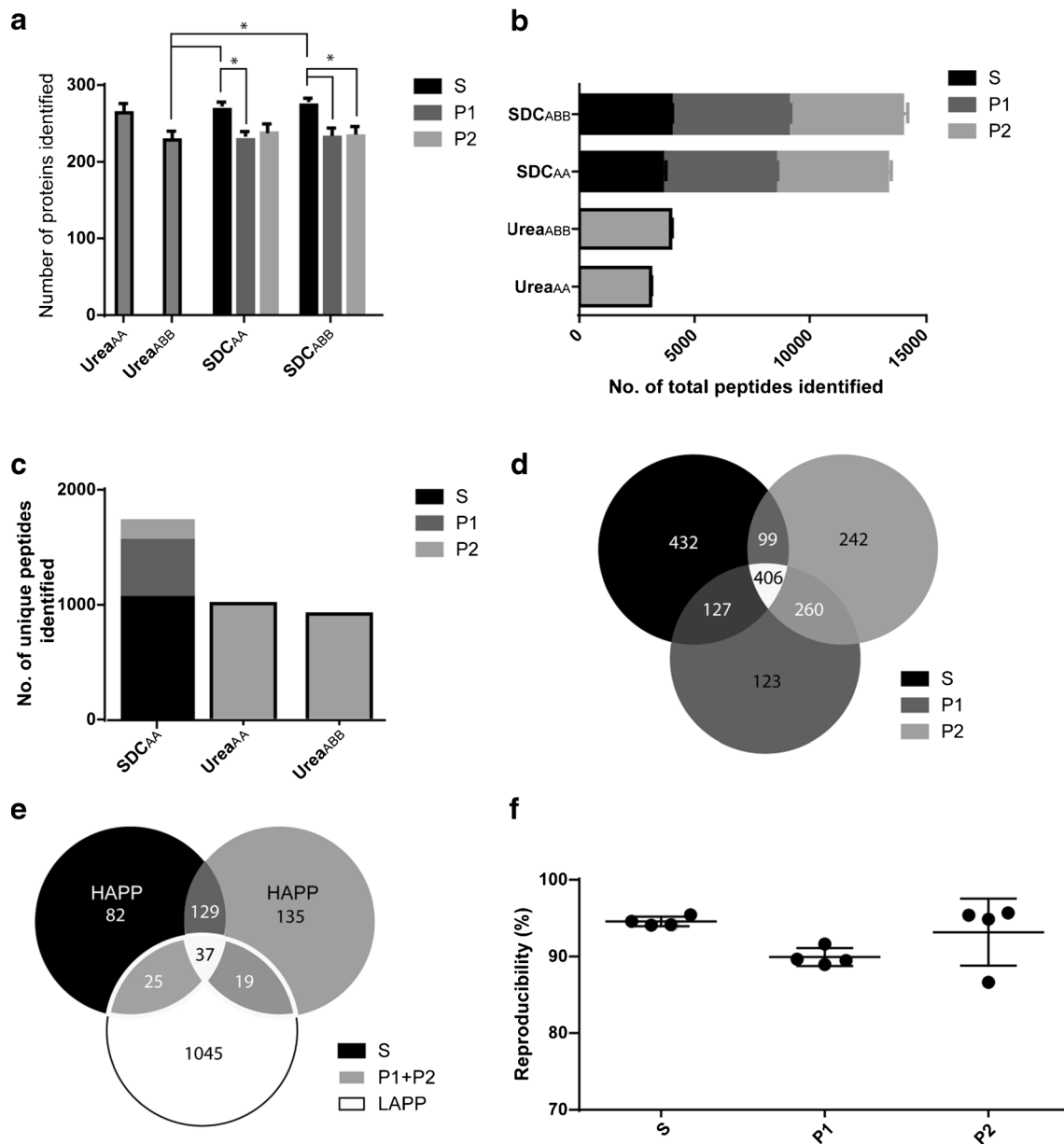
In order to study the co-precipitation of peptides with SDC, we tested whether this detergent could increase peptide recovery from ISD/AP-processed samples when using two different reaction buffers (including four replicates per condition): 25 mM ammonium bicarbonate (ABB) buffer (SDC<sub>ABB</sub>) which was used according to the methodology reported by Proc [11] and 100 mM AA buffer (SDC<sub>AA</sub>), which we have previously shown to reduce experimentally induced

deamidation and achieve optimal identification of unique peptides [24]. Additional comparisons were made with urea-assisted digestions performed in ABB (urea<sub>ABB</sub>) and AA (urea<sub>AA</sub>) buffers using two replicates per condition. For all experiments performed in this study, FDR distribution was plotted against pep\_expect distribution (ESM Figs. S2 and S3). Pep\_expect < 0.05 (as generated by Mascot) was set as the threshold for peptide identification since this was found to be more stringent than the application of a 1 % FDR threshold.

In standard protocols, SDC is only involved during analysis of the first supernatant obtained after acidic precipitation [10, 25, 26], which in our experiments generated an average of 276 identified proteins. No significant differences were detected between the two buffers tested (ABB and AA), and the number of proteins identified in the supernatant fraction (S) was comparable to the urea-assisted ISD performed in urea<sub>AA</sub> buffer (257 proteins identified; Fig. 2A). Consistent with the findings of previous studies [24], the number of proteins recovered from the urea<sub>ABB</sub>-processed sample was significantly lower, with just 228 proteins being identified. The total number of peptides identified in fraction S was comparable between the urea<sub>AA</sub> and urea<sub>ABB</sub> buffers (~4000 peptides each).

We next analyzed the composition of the SDC pellet in order to test our hypothesis that a substantial subset of peptides binds and interacts with the ionic detergent and can therefore be co-precipitated during AP treatment. Indeed, using a sequential re-solubilization and precipitation procedure, we were able to successfully identify more than 230 proteins from the peptides that co-precipitated with SDC (Fig. 2A). Independent analysis of fractions P1 and P2, when combined with the analysis of fraction S, significantly increased the number of peptides identified to a total of 13,703 (averaged from SDC<sub>AA</sub> and SDC<sub>ABB</sub>) (Fig. 2B). No significant differences were detected between the two buffers tested with respect to the number of peptides identified in fractions P1 and P2. Although the buffer used did not affect the proteome coverage achieved by SDC-assisted ISD, we considered SDC<sub>AA</sub> to be optimal due to the reduced incidence of experimentally induced deamidation in AA buffer [24].

When we proceeded to test the optimized SDC<sub>AA</sub> protocol, we identified a total of 13,376 peptides, of which 1728 were unique. Of the total peptides recovered, 1062 were derived from the supernatant (S), 495 were recovered during the initial washing steps (P1), and 171 were obtained during the second washing step (P2) (Fig. 2C). ESM Data Set 1 and 2 contain the lists of total peptides identified from SDC<sub>AA</sub> and SDC<sub>ABB</sub>, respectively, and ESM Data Set 3 contains the list of total peptides identified from urea<sub>AA</sub> and urea<sub>ABB</sub> for each replicate. By comparing peptidome composition between fractions, we confirmed that 406 unique peptides were present in all three fractions analyzed (S, P1, and P2) (Fig. 2D).



**Fig. 2** Optimization of SDC-assisted in-solution tryptic digestion. SDC-assisted in-solution tryptic digestion was performed using either ammonium acetate ( $SDC_{AA}$ ) or ammonium bicarbonate ( $SDC_{ABB}$ ) and compared with urea-assisted in-solution tryptic digestions performed under the same conditions ( $Urea_{AA}$  and  $Urea_{ABB}$ , respectively). **A** Number of proteins identified using the tested methodologies. For SDC-assisted protocols, proteins in fractions *S*, *P1*, and *P2* were identified separately. No significant differences were detected between *S* and  $Urea_{AA}$ . **B** Number of total peptides identified using the tested methodologies. There were no significant differences in the total number of peptides identified when using buffers *AA* and *ABB*, but SDC significantly increased peptide identification relative to urea ( $p < 0.01$ ). **C** Number of unique peptides identified by the  $SDC_{AA}$  protocol (total obtained by combining replicates). Unique peptides

identified in fractions *S*, *P1*, and *P2* were filtered to remove peptides commonly found in more than one fraction. **D** Venn diagram showing overlap between unique peptides identified in fractions *S*, *P1*, and *P2* (totals obtained by combining replicates). **E** Analysis of low-abundance plasma proteins (*LAPP*; ng/mL concentration or lower) identified in *S* and *P1 + P2* fractions. Identification of *LAPP* was carried out by comparison with the low-abundance plasma proteins (<ng/mL) present in the Plasma Proteome Database. High-abundance plasma proteins are indicated as *HAPP*. Four replicates were considered in this analysis. The Venn diagram shows the averaged number of proteins. **F** Estimation of technique reproducibility as calculated by evaluating peptide overlap between four replicates for the supernatant fraction (*S*) and pellets (*P1 + P2*)

The unique peptides recovered from the SDC pellets represented a 63.5 % increase in the total number derived from

fraction *S* (Table 2). Similarly, the total number of proteins identified was increased by 67.7 % following sequential re-

solubilization and precipitation of the SDC pellets. This level of efficiency compared favorably with the reported improvement in protein/peptide recovery attained using PT-assisted techniques; Masuda [7] achieved 57 % improvement in total proteins recovered and 48.3 % improvement in total peptides recovered from a membrane protein-enriched fraction derived from *Escherichia coli* (SDC-ISD<sub>PT</sub> vs. SDC-ISD<sub>AP</sub>, shown in Table 2). Given the membrane-bound origin of the sample tested by Masuda, this likely included many hydrophobic peptides that were susceptible to being co-precipitated with SDC using a conventional AP method. In contrast, when the same methodology was applied to a rat mitochondrial fraction more closely resembling the normal cellular proteome, the difference between AP and PT was less significant, with just 0.38 % increase in the total number of proteins identified and 13.0 % improvement in total peptide recovery (León [10]; SF-SDC-ISD<sub>PT</sub> vs. SF-SDC-ISD<sub>AP</sub>, shown in Table 2). Again, when PT removal of SDC was coupled to spin filter-aided ISD, there was only modest improvement in the recovery of either proteins (11.6 %) or peptides (14.3 %), likely due to the intensive sample handling required to apply this methodology.

We next evaluated the reproducibility of the SDC<sub>AA</sub> sample processing method by assessing peptide overlap between four independent sample preparations (Fig. 2F). Our optimized SDC-based method exhibited peptide overlaps of 94,

89, and 93 % for the fractions S, P1, and P2, respectively, thus displaying an excellent level of reproducibility comparable to that reported for PT-assisted detergent removal methods [7]. We therefore proceeded to assess the ability of our SDC-based method to recover low-abundance plasma proteins (LAPP) from the combined fractions P1 + P2. To do this, the list of all unique proteins identified in S and P1 + P2 across four technical replicates was assessed for the presence of proteins reported in the Plasma Proteome Database [27] at the nanogram-per-milliliter level or lower. Using this approach, we determined that fraction S contained 76 % of the total of LAPP identified by SDC<sub>AA</sub> and the extraction of peptides co-precipitated with SDC allowed us to identify an additional 30 % of LAPP. Both fractions shared 45 % of the total of LAPP (Fig. 2E and ESM Data Set 4). From the recovery of co-precipitated peptides with SDC, 135 unique proteins were identified in front of 82 identified in fraction S. All together demonstrate the importance of the extraction of co-precipitated peptides with SDC as that simple procedure allows the identification of a significant subset of proteins that are not considered in AP and expands significantly the proteome coverage obtained by the use of SDC in plasma biomarkers research by shotgun proteomics. Although MS technology has evolved rapidly in recent decades, biomarker research is still limited by the wide dynamic range of the human

**Table 2** Comparison of standard and optimized SDC-assisted ISD protocols (standard SDC-ISD, optimized SDC-ISD) with the detergent removal methodologies reported by León (SDC-ISD (AP), SDC-ISD

(PT)). The spin filter-aided versions of the detergent removal protocols are included for comparison (SF-ISD:SDC (AP), SF-ISD:SDC (PT))

Method	Sample	Proteins		Unique peptides		Instrument	Ref
		Average (sd)	Percentage	Average (sd)	Percentage		
Urea <sub>AA</sub> -ISD	0.5 µg from human plasma	257 (3)	–	869 (31)	–	LTQ-FT	
Standard SDC-ISD <sup>a</sup>	0.5 µg from human plasma	276 (7) <sup>c</sup>	67.7 % <sup>c</sup>	1062 (33) <sup>c</sup>	63.5 % <sup>c</sup>	LTQ-FT	
Optimized SDC-ISD <sup>b</sup>	0.5 µg from human plasma	463 (13) <sup>c</sup>		1736 (75) <sup>c</sup>		LTQ-FT	
SDC-ISD <sub>AP</sub>	0.4 µg from <i>E. coli</i> membrane proteins	≈49 <sup>d</sup>	57.1 % <sup>f</sup>	≈472 <sup>d, g</sup>	48.3 % <sup>f</sup>	QSTAR	Masuda [7]
SDC-ISD <sub>PT</sub>	0.4 µg from <i>E. coli</i> membrane proteins	≈77 <sup>d</sup>		≈700 <sup>d, g</sup>		QSTAR	Masuda [7]
SDC-ISD <sub>AP</sub>	0.25 µg from rat mitochondrial fraction	261 (9) <sup>d</sup>	0.38 % <sup>f</sup>	2782 (129) <sup>d</sup>	13.0 % <sup>f</sup>	LTQ Orbitrap XL	León [10]
SDC-ISD <sub>PT</sub>	0.25 µg from rat mitochondrial fraction	262 (15) <sup>d</sup>		3144 (295) <sup>d</sup>		LTQ Orbitrap XL	León [10]
SF-SDC-ISD <sub>AP</sub>	0.25 µg from rat mitochondrial fraction	257 (13) <sup>d</sup>	11.6 % <sup>f</sup>	3020 (119) <sup>d</sup>	14.3 % <sup>f</sup>	LTQ Orbitrap XL	León [10]
SF-SDC-ISD <sub>PT</sub>	0.25 µg from rat mitochondrial fraction	287 (5) <sup>d</sup>		3453 (63) <sup>d</sup>		LTQ Orbitrap XL	León [10]

<sup>a</sup> Standard SDC-ISD corresponds to fraction S

<sup>b</sup> Optimized SDC-ISD corresponds to the combined fractions S + P1 + P2

<sup>c</sup> Average number of unique peptides or proteins obtained from four replicates

<sup>d</sup> Average number of unique peptides or proteins obtained from three replicates

<sup>e</sup> Percent enrichment as calculated by comparing the number of unique peptides or proteins identified by the standard SDC-ISD versus the optimized SDC-ISD method

<sup>f</sup> Percent enrichment as calculated by comparing the number of unique peptides or proteins identified by PT versus AP methodologies for the ISD and SF-ISD approaches

<sup>g</sup> Total number of peptides as reported by Masuda [7]

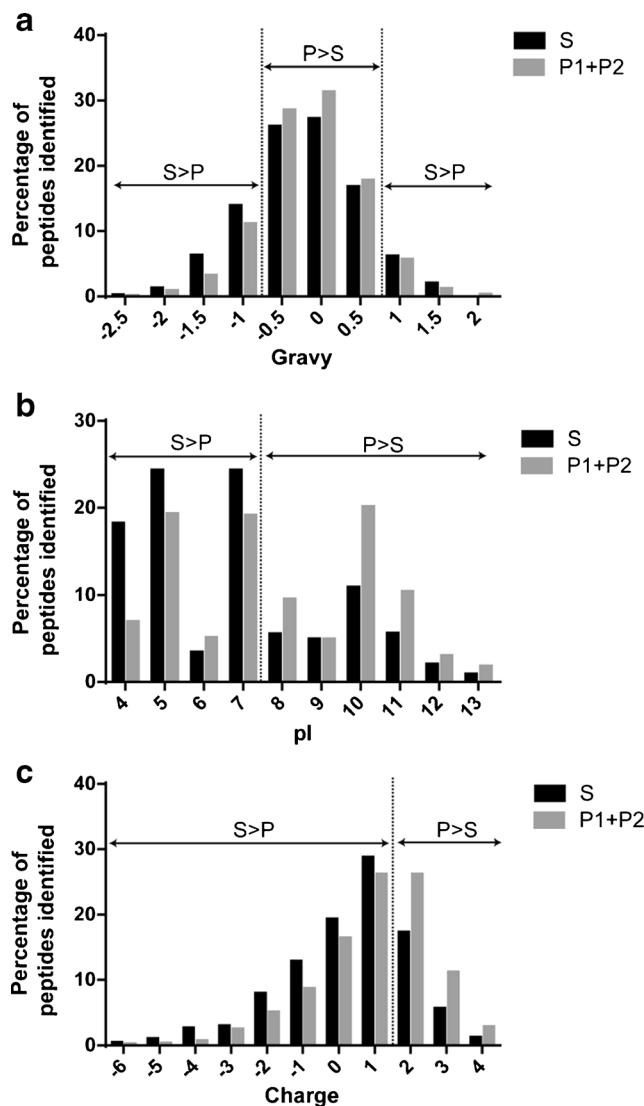
plasma proteome, which contains proteins varying in concentration by an estimated 10 orders of magnitude [28, 29]. Since the top 10 most abundant plasma proteins represent ~90–95 % of total protein content, it has become commonplace in MS-based studies to reduce the dynamic range of the samples by first fractionating these using methods including chromatography, gel-based electrophoresis, and immunoaffinity depletion of high-abundance plasma proteins (HAPP) [30]. It is important to note, therefore, that the SDC-based ISD method we report not only enhances trypsin performance and maximizes proteome coverage but also acts as a fractionation strategy that can help to reduce the dynamic range of the blood plasma proteome, thereby improving the identification of low-abundance plasma proteins of potential interest to biomarker researchers.

### Physicochemical properties of identified peptides derived from SDC-assisted in-solution tryptic digestion

We next sought to better characterize the peptides obtained using our optimized SDC method by analyzing their physicochemical properties, including GRAVY, *pI*, and peptide charge (Fig. 3). Supernatant S contained a high abundance of hydrophobic proteins (GRAVY score > 0.5), which was consistent with the conventional use of SDC for the analysis of membrane-derived proteins (Fig. 3A) [7, 8, 31]. However, hydrophilic proteins (GRAVY score < -0.5) were also enriched in this fraction. According to the *pI* distribution, peptides derived from proteins with low *pI* (< 7) were easily extracted in fraction S (Fig. 3B), which also contained a high number of low- or negatively charged peptides (charge < +1).

The peptides obtained from combined fractions P1 + P2 would typically be disregarded when conducting a single SDC/AP protocol [10, 25, 26], but our physicochemical analyses revealed the presence of multiple mildly hydrophilic/hydrophobic proteins in that fraction (GRAVY value from -0.5 to 0.5). Proteins with high *pI* values (8–13) that were not successfully recovered in the first supernatant (S) were instead efficiently enriched in the P1 + P2 fraction. According to the distribution of peptide charges shown in Fig. 3C, peptide recovery from the SDC pellet increased the percentage of highly charged peptides obtained (including +2-charged peptides, which correspond to the average net charge of complex tryptic samples generated at pH 2) [32].

When comparing all the parameters assessed here (GRAVY values, *pI*, and protein charge), the optimized SDC-assisted ISD was found to be well-suited to the analysis of all major protein types, thus offering excellent overall proteome coverage (rather enriching for membrane proteins only, as suggested by previous reports). Indeed, the protein compositions of SDC fractions S, P1, and P2 were complementary, and the optimized protocol acted as an effective fractionation strategy for improving the identification of low-abundance



**Fig. 3** Physicochemical properties of peptides identified using the SDC<sub>AA</sub> protocol. **A** Hydrophobic analysis of peptides identified in fractions S and P1 + P2. **B** The isoelectric point distribution of all peptides identified in fractions S and P1 + P2. **C** Peptide charge distribution of all peptides identified in fractions S and P1 + P2. Peptide net charge was calculated based on the number of constituent lysine (+), arginine (+), aspartic acid (-), and glutamic acid (-) residues, as well as the presence of a N-terminal amino group (+) and C-terminal carboxylic group (-)

proteins and long peptides by efficiently separating them from highly abundant soluble peptides.

### Miscleavage of peptides

The efficiency of trypsin cleavage and generation of nonstandard protein products can be influenced by many intrinsic properties of the peptides themselves, particularly the frequency/distribution of lysine and arginine residues along the backbone [33]. Consequently, proteolysis of proteins during sample processing is often incomplete, and peptides with missed cleavage



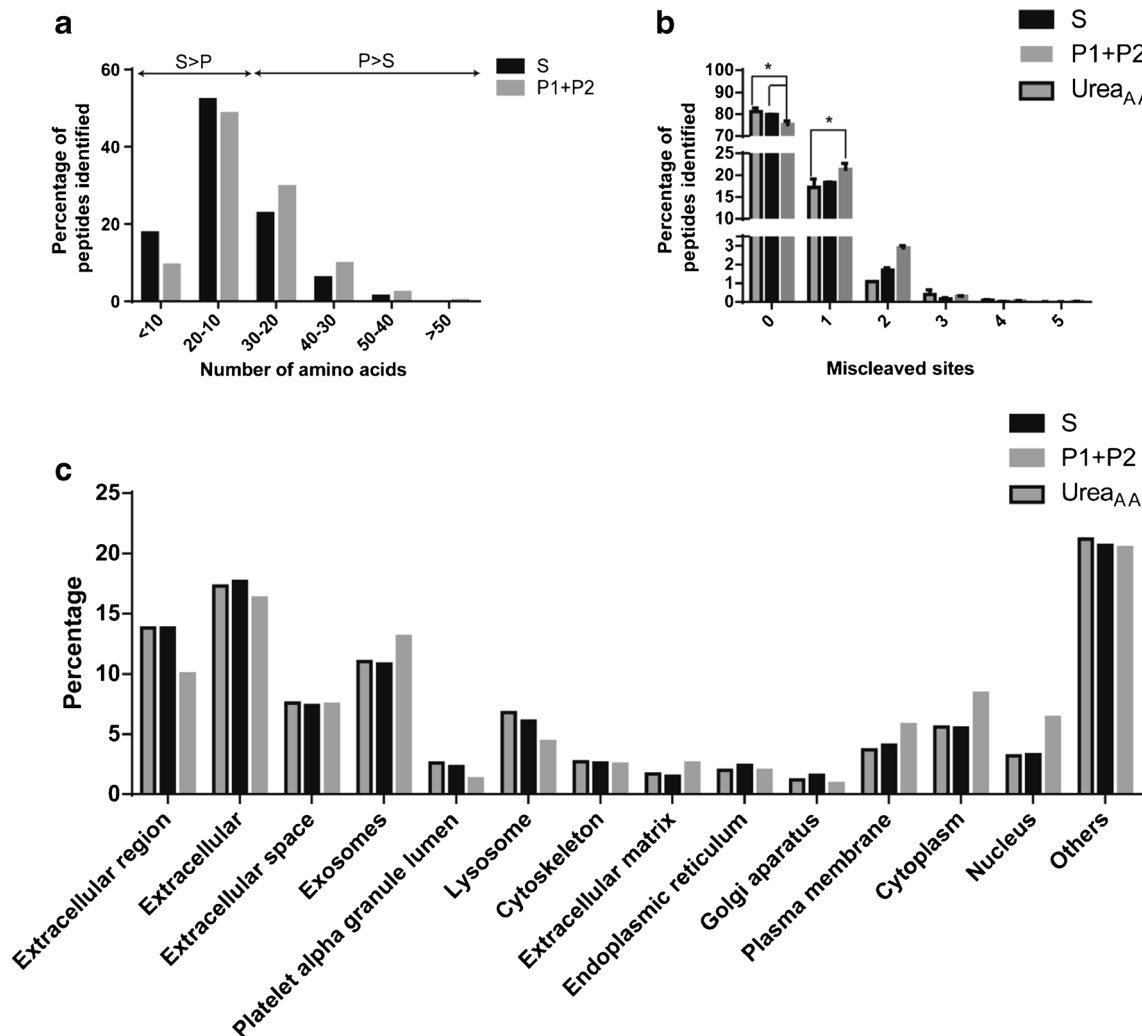
sites are common in tryptic samples. We therefore assessed the extent of peptide miscleavage and the length variability associated with our optimized SDC approach (Fig. 4). Miscleavage rates for the urea- and SDC-assisted ISDs were both ~20 %, in line with the findings of León [10]. According to our data, the majority of tryptic miscleaved peptides were co-precipitated during the first AP process, possibly because the longer peptides remained partially bound to the surfactant and readily pelleted together with SDC. This is consistent with our observation that longer peptides were more frequently identified in the combined P1 + P2 fraction (Fig. 4A), and may explain why previous investigations detected only low levels of peptide miscleavage when using SDC, restricting their analyses to the first supernatant S [10].

When the recovered peptides were analyzed and classified according to the number of missed cleavage sites (up to five max. per peptide), we observed no significant differences

between the urea-assisted ISD and fraction S data (Fig. 4B). However, the combined P1 + P2 fraction exhibited a higher incidence of peptides with one or two miscleaved sites, as well as a significantly lower number of fully cleaved peptides. These data indicated that analysis of miscleaved and long peptides in the fractions P1 and P2 is required in order to perform robust structural characterization of the plasma proteome.

### Gene enrichment study

We next performed a gene enrichment analysis in order to categorize the proteins identified in the pellet fraction. The subset of peptides obtained by re-precipitation and solubilization of the SDC pellet was observed to improve the identification of membrane and cytoplasmic proteins (Fig. 4C). In contrast, extracellular proteins were more abundant in fraction



**Fig. 4** Characterization of peptide infractions *S* and *P1+P2* using the SDC<sub>AA</sub> protocol and gene enrichment analysis. **A** Peptide length distribution for all peptides identified in fractions *S* and *P1+P2*. **B** Graph showing incidence of miscleaved sites in the SDC fractions (*S*

and *P1+P2*) compared with urea (*Urea<sub>AA</sub>*). Peptides with up to five miscleaved sites were considered. **C** Functional gene enrichment analysis based on cellular components

S and urea<sub>AA</sub>, consistent with the high solubility of these proteins in common digestion buffers [34]. Additionally, analysis of the combined fraction P1 + P2 enhanced the identification of proteins derived from membrane-bound organelles, including both nuclear and exosomal proteins.

## Concluding remarks

In the current study, we confirmed that unique peptides co-segregate with the detergent SDC during acid precipitation of digested plasma samples and that these peptides can be efficiently recovered from the pellet by sequential resolubilization and precipitation. Analysis of the peptides obtained from the SDC pellet significantly increased the total number of unique peptides detected in human plasma samples; hence, optimization of this SDC-based approach represents a substantial methodological advance in the field of shotgun proteomics for complex biological samples. Indeed, by acting as a de facto fractionation strategy, our SDC protocol also efficiently reduced the dynamic range of the human plasma proteins under study, and is therefore well-suited to the preparation of complex biological samples for biomarker discovery studies.

**Acknowledgments** This work was supported in part by the Singapore Ministry of Education (Tier 2: ARC9/15), NTU-NHG Ageing Research Grant (ARG/14017), and Singapore Ministry of Education (Tier 1: RGT15/13).

**Compliance with ethical standards** The study was approved by the Research Ethics Committee of Tan Tock Seng Hospital and Nanyang Technological University, Singapore. Written informed consent was obtained from all study participants ( $n = 34$ ).

**Conflict of interest** The authors declare that they have no competing interests.

## References

- Zhang Y, Fonslow BR, Shan B, Baek MC, Yates 3rd JR. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*. 2013;113(4):2343–94. doi:10.1021/cr3003533.
- Saveliev S, Bratz M, Zubarev R, Szapacs M, Budamgunta H, Urh M. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nat Meth*. 2013;10(11). doi:10.1038/nmeth.f.371.
- Leimgruber R. Extraction and solubilization of proteins for proteomic studies. In: Walker J (ed) *The proteomics protocols handbook*. Humana Press; 2005. pp 1–18. doi:10.1385/1-59259-890-0:001.
- Choudhary C, Weinert BT, Nishida Y, Verdin E, Mann M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol*. 2014;15(8):536–50. doi:10.1038/nrm3841.
- Choudhary J, Grant SG. Proteomics in postgenomic neuroscience: the end of the beginning. *Nat Neurosci*. 2004;7(5):440–5. doi:10.1038/nn1240.
- Fonslow BR, Stein BD, Webb KJ, Xu T, Choi J, Park SK, et al. Digestion and depletion of abundant proteins improves proteomic coverage. *Nat Methods*. 2013;10(1):54–6. doi:10.1038/nmeth.2250.
- Masuda T, Tomita M, Ishihama Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res*. 2008;7(2):731–40. doi:10.1021/pr700658q.
- Zhou J, Zhou T, Cao R, Liu Z, Shen J, Chen P, et al. Evaluation of the application of sodium deoxycholate to proteomic analysis of rat hippocampal plasma membrane. *J Proteome Res*. 2006;5(10):2547–53. doi:10.1021/pr060112a.
- Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW, et al. Analysis of the plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature*. 2002;419(6906):537–42. doi:10.1038/nature01111.
- Leon IR, Schwammle V, Jensen ON, Sprenger RR. Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Mol Cell Proteomics*. 2013;12(10):2992–3005. doi:10.1074/mcp.M112.025585.
- Proc JL, Kuzyk MA, Hardie DB, Yang J, Smith DS, Jackson AM, et al. A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *J Proteome Res*. 2010;9(10):5422–37. doi:10.1021/pr100656u.
- Erde J, Loo RR, Loo JA. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J Proteome Res*. 2014;13(4):1885–95. doi:10.1021/pr4010019.
- Molloy MP, Herbert BR, Walsh BJ, Tyler MI, Traini M, Sanchez JC, et al. Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. *Electrophoresis*. 1998;19(5):837–44. doi:10.1002/elps.1150190539.
- Datta A, Chen CP, Sze SK. Discovery of prognostic biomarker candidates of lacunar infarction by quantitative proteomics of microvesicles enriched plasma. *PLoS One*. 2014;9(4), e94663. doi:10.1371/journal.pone.0094663.
- Pathan M, Keerthikumar S, Ang C-S, Gangoda L, Quek CYJ, Williamson NA, et al. FunRich: an open access standalone functional enrichment and interaction network analysis tool. *Proteomics*. 2015;15(15):2596–601. doi:10.1002/pmic.201400515.
- Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol*. 2014;32(3):223–6. doi:10.1038/nbt.2839.
- Wisniewski JR, Ostasiewicz P, Mann M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J Proteome Res*. 2011;10(7):3040–9. doi:10.1021/pr200019m.
- Yu Y, Leng T, Yun D, Liu N, Yao J, Dai Y, et al. Global analysis of the rat and human platelet proteome—the molecular blueprint for illustrating multi-functional platelets and cross-species function evolution. *Proteomics*. 2010;10(13):2444–57. doi:10.1002/pmic.200900271.
- Manza LL, Stamer SL, Ham A-JL, Codreanu SG, Liebler DC. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics*. 2005;5(7):1742–5. doi:10.1002/pmic.200401063.
- Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Meth*. 2009;6

- (5):359–362. doi:[http://www.nature.com/nmeth/journal/v6/n5/suppinfo/nmeth.1322\\_S1.html](http://www.nature.com/nmeth/journal/v6/n5/suppinfo/nmeth.1322_S1.html).
21. Weber K, Kuter DJ. Reversible denaturation of enzymes by sodium dodecyl sulfate. *J Biol Chem*. 1971;246(14):4504–9.
  22. Meng W, Zhang H, Guo T, Pandey C, Zhu Y, Kon OL, et al. One-step procedure for peptide extraction from in-gel digestion sample for mass spectrometric analysis. *Anal Chem*. 2008;80(24):9797–805. doi:[10.1021/ac801344z](https://doi.org/10.1021/ac801344z).
  23. Lin Y, Liu Y, Li J, Zhao Y, He Q, Han W, et al. Evaluation and optimization of removal of an acid-insoluble surfactant for shotgun analysis of membrane proteome. *Electrophoresis*. 2010;31(16):2705–13. doi:[10.1002/elps.201000161](https://doi.org/10.1002/elps.201000161).
  24. Hao P, Ren Y, Datta A, Tam JP, Sze SK. Evaluation of the effect of trypsin digestion buffers on artificial deamidation. *J Proteome Res*. 2014;14(2):1308–14. doi:[10.1021/pr500903b](https://doi.org/10.1021/pr500903b).
  25. Lin Y, Zhou J, Bi D, Chen P, Wang X, Liang S. Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal Biochem*. 2008;377(2):259–66. doi:[10.1016/j.ab.2008.03.009](https://doi.org/10.1016/j.ab.2008.03.009).
  26. Lin Y, Lin H, Liu Z, Wang K, Yan Y. Improvement of a sample preparation method assisted by sodium deoxycholate for mass-spectrometry-based shotgun membrane proteomics. *J Sep Sci*. 2014;37(22):3321–9. doi:[10.1002/jssc.201400569](https://doi.org/10.1002/jssc.201400569).
  27. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res*. 2014;42(Database issue):D959–65. doi:[10.1093/nar/gkt1251](https://doi.org/10.1093/nar/gkt1251).
  28. Mitchell P. Proteomics retrenches. *Nat Biotechnol*. 2010;28(7):665–70.
  29. Ernoult E, Bourreau A, Gamelin E, Guette C. A proteomic approach for plasma biomarker discovery with iTRAQ labelling and OFFGEL fractionation. *J Biomed Biotechnol*. 2010;2010:8. doi:[10.1155/2010/927917](https://doi.org/10.1155/2010/927917).
  30. Smejkal GB, Lazarev A. Separation methods in proteomics. CRC Press, 2005.
  31. Lin Y, Wang K, Yan Y, Lin H, Peng B, Liu Z. Evaluation of the combinative application of SDS and sodium deoxycholate to the LC-MS-based shotgun analysis of membrane proteomes. *J Sep Sci*. 2013;36(18):3026–34. doi:[10.1002/jssc.201300413](https://doi.org/10.1002/jssc.201300413).
  32. Ballif BA, Villen J, Beausoleil SA, Schwartz D, Gygi SP. Phosphoproteomic analysis of the developing mouse brain. *Mol Cell Proteomics*. 2004;3(11):1093–101. doi:[10.1074/mcp.M400085-MCP200](https://doi.org/10.1074/mcp.M400085-MCP200).
  33. Thiede B, Lamer S, Mattow J, Siejak F, Dimmler C, Rudel T, et al. Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rap Comm Mass Spec*. 2000;14(6):496–502. doi:[10.1002/\(sici\)1097-0231\(20000331\)14:6<496::aid-rcm899>3.0.co;2-1](https://doi.org/10.1002/(sici)1097-0231(20000331)14:6<496::aid-rcm899>3.0.co;2-1).
  34. Ngoka LC. Sample prep for proteomics of breast cancer: proteomics and gene ontology reveal dramatic differences in protein solubilization preferences of radioimmunoprecipitation assay and urea lysis buffers. *Proteome Sci*. 2008;6:30. doi:[10.1186/1477-5956-6-30](https://doi.org/10.1186/1477-5956-6-30).