•ARTICLES•

February 2021 Vol.64 No.2: 313–321
https://doi.org/10.1007/s11426-020-9878-8

Fully automated sample treatment method for high throughput proteome analysis

Huiming Yuan[1*], Zhongpeng Dai[1], Xiaodan Zhang[1], Baofeng Zhao[1], Hongwei Chu[1,2], Lihua Zhang[1*] & Yukui Zhang[1]

[1]*CAS Key Laboratory of Separation Sciences for Analytical Chemistry, National Chromatographic R & A Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China;*
[2]*Zhang Dayu School, Dalian University of Technology, Dalian 116024, China*

Received July 14, 2020; accepted September 18, 2020; published online December 23, 2020

The bottom-up strategy for proteome analysis typically employs a multistep sample preparation workflow that suffers from being time-consuming and sample loss or contamination caused by the off-line manual operation. Herein, we developed a hollow fibre membrane (HFM)-aided fully automated sample treatment (FAST) method. Due to the confinement effects of HFMs and the immobilized enzymatic reactor, the proteome samples could be denatured, reduced, desalted and digested within 8–20 min *via* the one-stop service. This method also showed superiority in trace sample analysis. In one and half hours, we could identify about 1,600 protein groups for 500 HeLa cells as the starting materials, 1.5–8 times more than those obtained by previously reported methods. Through the on-line combination of FAST with nano-liquid chromatography-electrospray ionization tandem mass spectrometry (nanoLC-ESI-MS/MS), we further established a fully integrated platform for label-free quantification of proteome with high reproducibility and precision. Collectively, FAST presented here represents a major advance in the high throughput sample treatment and quantitative analysis of proteomes.

proteomics, fully automated sample treatment, hollow fibre membrane, immobilized enzymatic reactor, label-free quantification

Citation: Yuan H, Dai Z, Zhang X, Zhao B, Chu H, Zhang L, Zhang Y. Fully automated sample treatment method for high throughput proteome analysis. *Sci China Chem*, 2021, 64: 313–321, https://doi.org/10.1007/s11426-020-9878-8

1 Introduction

Currently, the bottom-up strategy is widely used for proteome profiling [1–3]. By conventional protocols, multi-step off-line sample preparation, including denaturation, reduction, alkylation, digestion and desalting, has become one of the main bottlenecks that might affect the accuracy, precision, sensitivity and throughput of proteome analysis. Therefore, considerable efforts have been made to improve the sample preparation efficiency [4–6].

Till now, the most popular method is filter-aided sample preparation (FASP) by which the removal of low-molecular weight interfering substances, protein reduction, alkylation and digestion are performed in a single ultra-filtration device [7–10]. However, it takes approximately 1–1.5 days to complete the entire process. Furthermore, to further accerelate the proteins extraction, sample preparation by easy extraction and digestion (SPEED) was developed, which consists of three mandatory steps, acidification, neutralization and digestion [11]. However, since traditional in-solution digestion was employed in this protocol, it still requires more than 20 h to complete the whole sample preparation. Recently, *in-situ* sample processing techniques such as the nanodroplet processing platform (nanoPOTS) [12–15], sin-

*Corresponding authors (email: huimingyuan@dicp.ac.cn; lihuazhang@dicp.ac.cn)

gle-pot solid-phase-enhanced sample preparation (SP3) [16–18] and the integrated proteome analysis device (iPAD) [19,20] have been developed to accelerate the treatment of proteomic samples through enhancing the collision possibility of reactive molecules in a micro-scale chamber or on beads, which enable the preparation of sub-nanogram biological materials in 1–18 h. Furthermore, in-Stage Tip digestion (iST) [21–23] and simple and integrated spin tip-based proteomics technology (SISPROT) [24–26] have also been developed, by which protein concentration, reduction, alkylation, clean-up and digestion can be completed in tips within 2–6 h. However, the above-mentioned methods are difficult to online integrate with LC-ESI/MS/MS to achieve automated proteome quantitation, which might affect the quantitation accuracy, precision, coverage and throughput, especially for label-free quantification.

Herein, we developed a fully automated sample treatment (FAST) method to achieve a "one-stop" proteomic sample preparation with an on-line combination of protein denaturation and reduction at 90 °C within a hollow fibre membrane interface (HFMI), desalting by the second HFMI, and digestion by an inorganic-organic hybrid silica monolith-based immobilized enzymatic microreactor (IMER). Not only was the sample preparation time greatly shortened to within 20 min, but the required sample amount was also decreased to 20 ng. Importantly, FAST could be integrated with a nano-LC-ESI/MS/MS system to achieve large-scale label-free-based proteome quantification.

2 Materials and methods

2.1 Reagents

Myoglobin (horse heart, Myo), trypsin (bovine pancreas) and transferrin (bovine, Tref) were purchased from Sigma (USA). Bovine serum albumin (BSA) was obtained from Sino-American Biotechnology (China). HPLC-grade acetonitrile (ACN) was purchased from Merck (Germany). Tetrathoxysilane (TEOS, 95%), 3-amino-propyltriethoxysilane (ATPES, 99%) and sodium cyanoborohydride were purchased from Acros Organics (Geel, Belgium). α-Cyan-4-hydroxycinnamic acid (CHCA) and sinapinic acid (SA) were purchased from Bruker Daltonik (Germany). Rabbit polyclonal antibodies to SSB and RNPS1 were purchased from GeneTex Company (USA).

Water was purified using a Milli-Q system (Millipore, USA). All other chemicals and solvents were of analytical grade.

2.2 Cell culture and metabolic labeling

HeLa cells were cultured in minimum essential medium (MEM) containing 10% (*v/v*) foetal bovine serum (FBS) and were maintained in 95% humidified air and 5% CO_2 atmosphere at 37 °C. The cell lines MHCC97H and MHCC97L, which are hepatocellular carcinoma (HCC) cells with high and low metastatic potential, respectively, were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% (*v/v*) FBS, and penicillin/streptomycin, and left at 37 °C in a humidified atmosphere of 5% CO_2. The adherent cell layer was washed with PBS and then harvested by treatment with 0.05% trypsin-EDTA solution for 5 min at 37 °C. After cell counting, the cells were centrifuged at 250×*g* for 5 min to remove trypsin and then washed 3 times with PBS.

A549 cells were labelled using stable isotope labeling with amino acids in cell culture (SILAC) media (Thermo, USA) at 37 °C with 5% CO_2. For the "medium" labelling media, L-lysine- and L-arginine-depleted SILAC RPMI 1640 media were supplemented with [4,4,5,5-D4] L-lysine (100 μg/mL) and [13C6] L-arginine (100 μg/mL), 10% dialyzed FBS, and 1% penicillin/streptomycin mixture. For the "heavy" labeling media, only [4,4,5,5-D4] L-lysine and [13C6] L-arginine were replaced with [13C6, 15 N2] L-lysine and [13C6, 15N4] L-arginine. The SILAC-labelled cells were cultured in SILAC meida for at least six doubling times to ensure the complete incorporation of the isotope amino acids.

2.3 Protein extraction

The HeLa cells (1×10^7) were divided into two portions, and then lysed with 8 M urea containing 0.2% NP-40 and C12Im-Cl containing 50 mM NH_4HCO_3 (pH 8.0), according to previous protocols [27,28], followed by ultrasonication for 180 s at 130 W and centrifugation at 20,000×*g* for 20 min. The supernatant was collected, and the protein concentration was determined using a BCA assay.

Proteins were extracted from SILAC-labelled A549 cells and the hepatocellular carcinoma (HCC) cells with high and low metastatic potentials (MHCC97H/L) according to the above-mentioned protocol. In brief, the grown cells were suspended in the extraction buffer composed of 8 M urea and 50 mM NH_4HCO_3 (pH 8.0) with subsequent ultrasonication for 180 s at 130 W and centrifugation at 20,000×*g* for 20 min. The supernatants were collected as the soluble fraction of the extracted whole cell lysate proteins, and the protein concentration was determined by BCA assay.

We employed the cell sorter (SH800S, Sony, Japan) to directly isolate 100 or 500 HeLa cells into 0.2 mL skirted 96-well PCR plates (Thermo Scientific, UK). After cell collection, the 96-well PCR plates were directly submitted for protein exaction. The cells were first incubated with 2 μL NP-40 (1%, *w/v*, Beyotime) for 15 min. After the addition of 2 μL urea (8 M), the cell lysates were then heated at 95 °C for 30 min. Finally, the supernatant was collected after centrifugation at 20,000×*g* for 20 min.

2.4 Hollow fibre membrane interface

The hollow fibre membrane interface (HFMI) was prepared according to the following procedures. The PEG coating of fused silica capillaries (90 μm i.d.× 180 μm o.d.× 5 cm length, Zhengzhou Innosep Biosciences Co,. Ltd, China) was inserted into a 200 μm-i.d. hollow fibre membrane (molecular cut-off weight of 3,000 Da, donated by Professor Jiuyang Zhao at the Second Affiliated Hospital of Dalian Medical University) and two pieces of peek tubes (250 μm i.d.). Then, they were threaded through quartz tubes and fixed together with two unions. Finally, two holes were drilled on opposite sides of the quartz tube to introduce reaction buffers or an exchange buffer.

2.5 Organic-inorganic hybrid silica monolith-based microreactor

The organic-inorganic hybrid silica monolith-based immobilized trypsin microreactor (IMER) was prepared according to our previous procedure [29]. In brief, the capillary with an inner diameter of 250 μm was filled with a polymerization solution containing TEOS (112 μL), APTES (118 μL), anhydrous ethanol (215 μL), cetyltrimethyl ammonium bromide (8 mg) and water (32 μL). Then, the capillary was placed into a 40 °C water bath for 24 h to form an organic-inorganic monolith. Subsequently, the monolithic support was activated by flushing a solution of 10% (*v*/*v*) glutaraldehyde for 6 h at room temperature. Next, 2 mg/mL trypsin dissolved in 100 mM phosphate buffer (pH 8.0) containing 50 mM benzamidine, and 5 mg/mL sodium cyanoborohydride was pumped continuously for 24 h at 4 °C to immobilize trypsin. After this, the microreactor was purged with 1 M Tris-HCl (pH 8.0) and 20% acetonitrile (ACN) (*v*/*v*) for 4 h. Then, the microreactor was filled with 0.02% (*w*/*v*) NaN_3 solution and stored at 4 °C before use.

2.6 On-line protein pre-treatment by FAST

2.6.1 Standard proteins

With a flow rate of 5 μL/min, 130 μg/mL protein mixture (50 μg/mL myoglobin, 50 μg/mL transferrin, 30 μg/mL BSA) was pumped through the first HFMI (0.2 mm i.d. ×80 mm), placed into a 90 °C heater with 6 M guanidine hydrochloride containing 50 mM dithiothreitol (DTT) as reaction buffers, and then passed through the second HFMI (0.2 mm i.d.×40 mm) with 20 mM ammonium bicarbonate (pH 8.0) as an exchange buffer at a flow rate of 3 mL/min for desalting. This was followed by on-line digestion by an IMER (0.25 mm i.d. ×50 mm) at ambient temperature with a flow rate of about 1 μL/min. The digests were collected, and 1 μL of the collected peptides was directly deposited on a MALDI plate for further analysis. After each sample preparation, 50 mM ABC containing 50% ACN (*v*/*v*) was pumped into FAST for system clean-up for 10 min.

2.6.2 HeLa cell lysates

A total of 0.1 mg/mL HeLa cell lysates (8 M urea) were pumped directly into the FAST, and the experimental conditions were the same as those for standard proteins. A total of 20 μL of digests were collected, and 500 ng peptides (5 μL) were analysed by a nano-liquid chromatography-electrospray ionization-tandem mass spectrometry system (nano-LC-ESI/MS/MS).

For 100 and 500 HeLa cells, no more than 5 μL cell lysates were injected into the FAST and pretreated with the same experimental conditions as those described above. The resulting peptides were then captured by a C18 precolumn (0.15 mm i.d. ×20 mm, 120 Å) and analysed by nano-LC-ESI/MS/MS.

2.6.3 HCC cells with high and low metastatic potentials

For label-free quantification, the same aliquots of protein extracts from the MHCC97H and MHCC97L cell lines were processed by the FAST with the same conditions as those described for the HeLa cell lysates. The resulting peptides with the same quantities (each sample: 500 ng) were directly analysed by the nano-LC-ESI/MS/MS system.

The whole analysis process consisted of three steps. At step 1, the valve 1 was at Position A (3−4 connected), and 500 ng (5 μL) from cell lysates was firstly introduced into FAST device, and performed in sequence protein denaturation and reduction (3 min), desalting (2 min) and protein digestion (3 min), and then the protein digests were captured by a sample-loop (50 μL) on the valve 1. After that, to ensure all samples could enter into the subsequent nanoLC-MS system for analysis, additional 50 mM ABC (12 μL, about 12 min) was pumped into FAST, and the residuals were collected by the same sample loop. At step 2, after valve 1 switch (Positon B, 1−6 connected), the sample loop was connected with LC system, and the protein digests were then flushed into the C18 precolumn, followed by LC-MS/MS analysis. At step 3, after each sample preparation, 50 mM ABC containing 50% ACN (*v*/*v*) was pumped into FAST for system clean-up for 10 min.

2.6.4 The carryover evaluation

The carryover of proteins/peptides on the FAST was evaluated by the following steps. The light and heavy SILAC-labelled proteins from A549 (60 μg/mL) were in sequence pumped into the FAST at a flow rate of 5 μL/min for 20 min. The light fraction was discarded, whereas the heavy fraction was collected and subjected to nano-LC-MS/MS analysis.

2.7 In-solution sample preparation

For comparison, the same aliquots of the 3-protein mixture were processed by the conventional in-solution protocol described elsewhere [28]. In brief, Myoglobin, BSA and Transferrin were first denatured in 20 mM NH_4HCO_3 buffer (pH 8.0) containing 8 M urea for 1 h in a 37 °C water bath, and then reduced by 20 μM DTT at 56 °C for 1.5 h. After being cooled to room temperature, 40 μM iodoacetamide (IAA) was added to the solution and reacted at room temperature in the dark for 40 min, followed by dilution with 20 mM NH_4HCO_3 (pH 8.0) until the final concentration of myoglobin, BSA and transferrin was 50, 30 and 50 μg/mL, respectively. The in-solution digestion of proteins was performed by adding trypsin to 0.13 mg/mL protein solution with a substrate-to-enzyme ratio of 25:1 (*m*/*m*). The solution was incubated at 37 °C for 24 h, and the digests were stored at −20 °C before use.

For HeLa cell lysate (C12Im-Cl), 100 μg proteins were transferred to a filter device (molecular cut-off weight of 10,000 Da). Proteins were denatured and reduced with 100 mM DTT at 95 °C for 5 min and washed with 8 M urea in 50 mM NH_4HCO_3. Subsequently, proteins were subjected to alkylation with 20 mM IAA containing 8 M urea in 50 mM NH_4HCO_3 in the dark at the room temperature for 20 min, followed by centrifugation at 14,000×*g* for 15 min. Then, the proteins were washed three times with 50 mM NH_4HCO_3. Finally, proteins were digested with a trypsin/protein ratio (*m*/*m*) of 1:25 at 37 °C for 12 h, and the peptides were collected by centrifugation at 14,000×*g* for 15 min.

2.8 Nano-LC MS/MS analysis

For 500-cell, HeLa cell, MHCC97H and MHCC97L cell sample analysis, the peptides were separated by a C18 capillary column (150 μm i.d. ×150 mm, 1.9 μm, 120 Å) at a flow rate of 600 nL/min. The mobile phases included H_2O containing 2% ACN and 0.1% formic acid (A) and ACN containing 80% ACN and 0.1% formic acid (B). For the separation of HeLa cell digests, the linear gradients were set as follows: 9% (0 min)→25% (34 min) →50% (76 min) →95% (77min)→95% (80 min). For the separation of protein digests from the MHCC97H and MHCC97L cells, the linear gradients were set as follows: 9% (0 min)→25% (90 min) →50% (110 min)→95% (115 min)→95% (120 min). After each nano-RPLC separation, the column was equilibrated with the initial mobile phase for 20 min.

Lumos Fusion mass spectrometer (Thermo Fisher, USA) was hyphenated with an EASY-nLC 1200 LC system for protein identification. Full MS scans were performed in the Orbitrap mass analyzer over *m*/*z* range of 350−1500 with a mass resolution of 60,000. The MS/MS spectra were acquired in data-dependent mode with a Top Speed method (3 s). Tandem MS was performed in the ion trap mass analyzer using an isolation window of 1.6 Da by quadrupole mass analyzer and HCD fragmentation with normalized collision energy of 30. The dynamic exclusion time was set to 18 s.

For 100-cell sample analysis, the peptides were separated by a C18 capillary column (75 μm i.d. ×150 mm, 1.9 μm, 120 Å) at a flow rate of 300 nL/min. The mobile phases included H_2O containing 2% ACN and 0.1% formic acid (A) and ACN containing 98% ACN and 0.1% formic acid (B). The linear gradients were set as follows: 2% (0 min)→7% (0.1 min) →23% (50.1 min)→40% (70.1 min)→80% (72.1 min) →80% (85.1 min). After each nano-RPLC separation, the column was equilibrated with the initial mobile phase for 20 min.

Q Exactive mass spectrometer (Thermo Fisher) was coupled with an EASY-nLC 1000 LC system for protein identification. The electrospray voltage was set at 2.4 kV and the temperature of ion transfer capillary was 275 °C. A full MS/MS cycle consisted of one full MS scan (resolution, 70,000; automatic gain control (AGC) value, $3e^6$; maximum injection time, 60 ms) in profile mode over a mass range between *m*/*z* 300 and 1,800, followed by fragmentation of the top 12 most intense ions by high energy collisional dissociation (HCD) with normalized collision energy at 28% in centroid mode (resolution, 17,500; AGC value: $5e^5$, maximum injection time: 200 ms). The dynamic exclusion window was set at 18 s.

2.9 Data analysis

The raw data were searched against the Swissport human. fasta database (42,164 entries, downloaded on July 27, 2016) using the Mascot node integrated within the Proteome Discoverer (PD) software (Version 2.1 SP1, Thermo Fisher Scientific). The precursor and fragment mass tolerances were set to 10 and 20 ppm, respectively. A maximum of two missed cleavages was allowed for trypsin digestion. Oxidation (M), Acetyl (Protein N-term), Carbamyl (N-term, K) and Deamidate (N-term, Q) were set as variable modifications. For conventional in-solution sample preparation, Cysteine carbamidomethylation was set as the fixed modification. False discovery rate (FDR) of peptide spectrum matches (PSMs) and identified peptides were validated by the Percolator algorithm at 1% based on *q*-values.

The MaxLFQ algorithm integrated within MaxQuant (version 1.5.1.0) was used for the label-free quantification (LFQ) analysis of proteins with default parameters [30,31]. The human database and the same maximum missed cleavage were set the same as in the PD search. Oxidation (M) and acetyl (Protein N-term) were set as the variable modifications. FDR based on posterior error probability (PEP) was determined by searching a reverse database and was set

to 0.01 for proteins and peptides.

Absolute quantification of protein abundances (copy numbers) were computed using peptide label-free quantification values, sequence length and molecular weight as described before based on a total protein per cell value of 100 pg for HeLa cells [19,21]. For label-free quantitative analysis of proteomes from MHCC97H and MHCC97L cell lines, proteins quantified in all six replicates were processed using Benjamini-Hochberg (BH) FDR estimations, and those passed the 1% BH-FDR threshold were retained in the volcano plot. Using two-fold differences as the cutoffs, proteins with ratio <0.5 or >2 were considered to exhibit significant changes in expression. Violin plots show the kernel density distribution of protein ratio log2 (heavy/light), which is plotted using the JMP 10 software.

3 Results and discussion

3.1 Development and validation of FAST

To achieve integrated sample preparation, we designed a device based on the FAST method (Figure 1(a)) by which on-line protein denaturation and reduction (unit A), buffer exchange (unit B) and digestion (unit C) were automatically performed.

In unit A, proteins were propelled through the hollow fibre membrane (HFM) with a cut-off molecular weight of 3,000 Da by a syringe pump at a flow rate of 5 μL/min (Figure 1 (b)). The buffer for denaturation and reduction, composed of 6 M guanidine hydrochloride (GdCL) and 50 mM dithiothreitol (DTT), was introduced with tangential flow into the HFM by a piezoelectricity pump. Due to the microscale confinement effect, intermolecular interactions between proteins and reactive molecules were greatly enhanced. Simultaneously, the first HFMI was heated to 90 °C to further accelerate the denaturation and reduction by avoiding protein aggregation. The efficiency of protein denaturation and reduction was evaluated with transferrin, a protein with 19 disulfide bonds. Without treatment, the sequence coverage of transferrin digested by IMER was 29%±2% (n=3) (Figure S1 (a), Supporting Information online). However, after treatment by unit A, the sequence coverage was increased to 65% ±5% (n=3) (Figure S1(b)). Furthermore, the recovery of proteins in the first HFMI was calculated as 99% by dividing the concentration of HeLa proteins after and before they entered unit A with BCA assay. The high protein recovery could be attributed to the excellent hydrophilicity of the cellulose acetate membrane.

Although by traditional in-solution sample preparation protocols, cysteine alkylation after protein reduction is an indispensable step, we ignored the cysteine alkylation step in FAST system since all sample preparation procedures were operated in an enclosed chamber. By FAST, without IAA alkylation, BSA could be identified with the sequence coverage of 93%, comparable to that (92%) obtained with IAA alkylation, as shown in Table S1 (Supporting Information online).

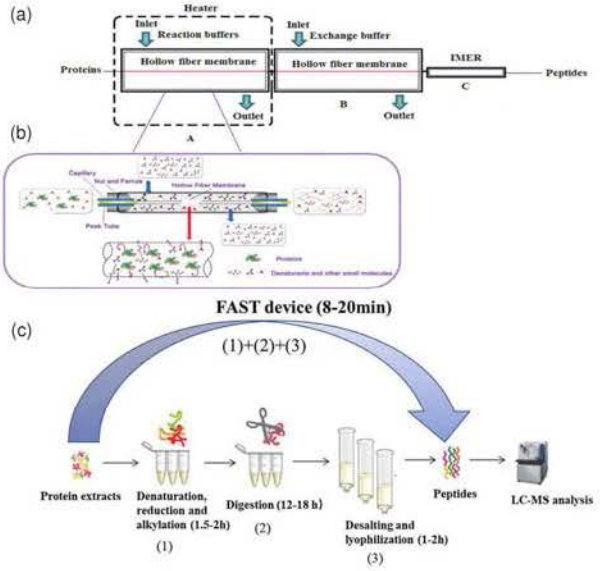

Figure 1 The hollow fibre membrane-aided fully automated sample treatment (FAST) method. (a) Scheme of FAST, involving (A) on-line protein denaturation and reduction, (B) buffer exchange, and (C) on-column digestion. (b) Principle of HFMI for protein denaturation and reduction. (c) Workflow of FAST for proteomic sample preparation (color online).

To avoid inhibition of enzymatic activity by GdCL and DTT residuals in the denatured and reduced proteome samples, the second HFMI (unit B) was used to replace them with 20 mM ammonium bicarbonate (ABC, pH 8.0), which were also introduced by forced convection with a tangential flow rate of 3 mL/min. As shown in Figure S2(a, b), after buffer exchange *via* unit B, transferrin was effectively digested by IMER with a sequence coverage of 70%±0.5% (n=3), demonstrating the high desalting efficiency of the second HFMI.

To prevent the back-pressure generated by the IMER from destroying the pore structure of the HFM (maximum 6.6 MPa), a previously developed IMER [17] with an organic-inorganic hybrid silica monolith as the support was used for unit C by considering the extremely low back pressure provided (<1 MPa even with a sample flow rate of 5 μL/min). Furthermore, the high enzyme/substrate ratio in a limited volume could enhance the effective collision probability of the protein and enzyme molecules, thereby enabling fast digestion. With a flow rate of 3 μL/min, 2 μg BSA was introduced into the IMER. BSA was efficiently digested within 48 s with a sequence coverage of 74%±3% (n=9), as shown in Figure S3.

Due to the good compatibility of the peptide buffer with MS detection, the resulting peptides from FAST were entered

directly into an LC-MS/MS for further analysis without any additional peptide clean-up.

As a proof-of-principle, a protein mixture of bovine serum albumin, myoglobin and transferrin ranging from 30 to 50 μg/mL was treated by FAST, followed by MALDI-TOF MS analysis. Due to the low sample loss and high digestion efficiency, sequence coverages of 88.1%±3.2%, 40.6% ±0.8% and 60.1%±1.5% (n=6) were obtained, 2–3 times higher than those obtained by the conventional in-solution protocol (Figure S4). Furthermore, the total sample treatment time was shortened from more than 20 h to 8 min, mainly attributed to the on-line fast digestion by IMER.

We further applied FAST to the treatment of proteomic samples. Compared to the widely used detergent-based FASP methods [8,28,32], the use of strong detergents for cell lysis was avoided since they were incompatible with proteolytic digestion and LC-MS/MS analysis. For comparison, we performed the quantitative proteome analysis of HeLa cell lysates obtained by urea and detergent solubilization by FAST and FASP, respectively. It should be noted that although less starting materials (2 μg *vs.* 100 μg) and shorter sample treatment time (20 min *vs.* 15 h) were employed by FAST compared to FASP, higher sample preparation recovery (95% *vs.* 50%–80% [32–35]), comparable identified protein number (3128±58 *vs.* 3403±54, n=3, Figure 2(a, b), Table S2) and similar dynamic range in protein abundance (both with 6 orders of magnitudes, Figure S5) were achieved due to the integrated on-line sample preparation.

Furthermore, we investigated the carryover of proteins/peptides on the FAST. It could be observed from Table S3 that the ignorable light labelled peptides (314/16179) were found in the heavy labelled fraction, and the average ratio of L/H labelled peptides was 7.8%±0.09% (n=4), demonstrating the low protein/peptide carryover of FAST.

3.2 High throughput analysis of a trace proteome sample

Since FAST could dramatically reduce sample contamination and loss risks, it is favourable for use in the treatment of trace proteomic samples. A total of 500 HeLa cells were prepared by FAST within 20 min, followed by a 1-h gradient single-shot LC-MS/MS analysis. From this, 1673±4 (n=3) protein groups were identified, 1.5–8 times more than those identified by previously reported results [8,19,36,37] (Figure 3(a)), among which 1312 proteins were detected in all experiments done in triplicate (Figure S6) with a median coefficient variation of label-free quantitative data (CV) of 11.7% (protein level, Figure 3(b)) and a dynamic range in protein abundance of 5 orders of magnitude (Figure 3(c)). Moreover, we further decreased the sample amount to 100 HeLa cells, and 644±8 (n=3) protein groups were also identified (Table S4). These results demonstrate that FAST ensures coverage, reproducibility and throughput for profiling trace proteome samples.

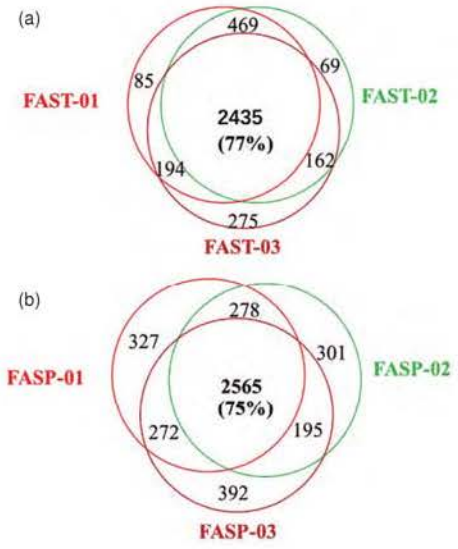

Figure 2 Comparative analysis of HeLa cell lysates using the FAST (a) and FASP (b) methods. Venn diagrams of protein identification in triplicate experiments (color online).

Additionally, we estimated the copy number of each identified protein from 500 HeLa cells according to a previously reported method [19,21]. The most abundant protein was Histone H4, with 7.0×10^7 copies/cell, and the median abundance value was ~7.8×10^5 copies/cell. The abundance of nearly 90% of the identified proteins was within 10^4–10^6 copies/cell. Furthermore, we found the abundance of 181 proteins was below 10^4 copies/cell (Figure 4(d) and Table S5), which were significantly enriched for the GO terms "metabolic process" and "mRNA processing" (P<2.57×10^{-8} and P<4.37×10^{-5}, respectively). The protein with the lowest abundance was the Zinc finger CCCH domain-containing protein 18 with 344 copies/cell, demonstrating the high identification sensitivity obtained by FAST, especially for the analysis of trace proteome samples.

3.3 Automated platform for label-free quantification of proteomes

Due to the excellent compatibility of FAST with LC-MS, we established a fully automated platform for label-free quantification of proteomes (Figure 4(a)). As shown in Figure 4 (b) and Table S6, the Pearson correlation values of the quantified protein signal from hepatocellular carcinoma cells between six replicate runs were 0.99, demonstrating the excellent reproducibility and precision of the label-free proteome quantification due, in part, to the automated "single-shot" analysis that includes sample preparation, separation and quantitation.

We further quantified the proteins from human hepato-

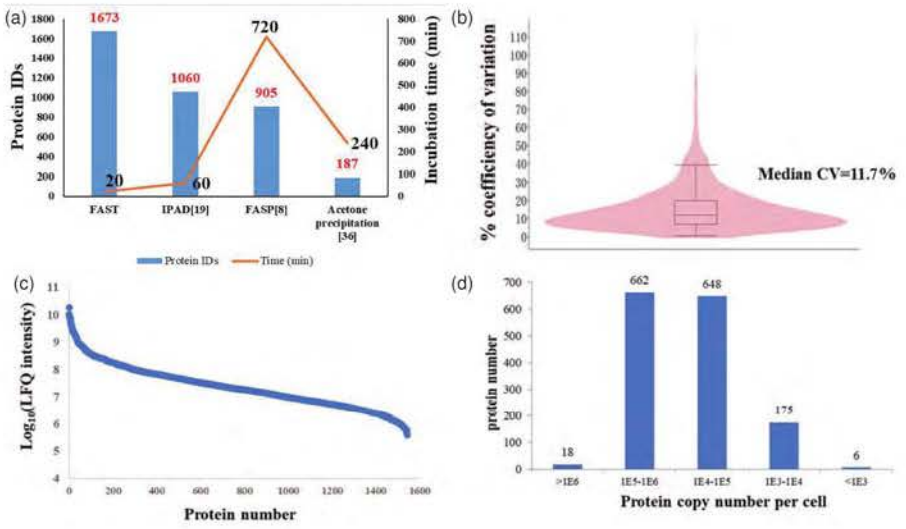

Figure 3 Proteome analysis of 500 HeLa cells. (a) Comparison of sample preparation using the FAST method and other reported methods. (b) Violin plot showing the distributions of coefficients of variation of protein LFQ intensities. (c) Dynamic range of quantified proteins achieved using the FAST method. (d) Distribution of estimated copy number (color online).

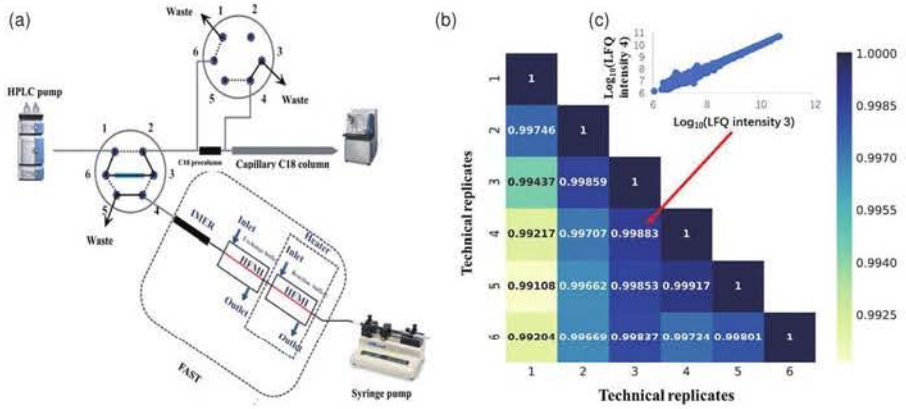

Figure 4 Fully automated proteome quantification. (a) Schematic diagram of the automated proteome quantification workflow including sample preparation, "single-shot" LC separation and MS detection. (b) Distribution of proteins quantified in the 2-h analysis of the biological triplicates. (c) Reproducibility of the LFQ intensities of the quantified proteins for 6 workflow replicates (color online).

cellular carcinoma cells with high and low metastatic potentials (MHCC97H/L). By this platform, 2,111 proteins were co-quantified in the six replicate runs (Figure 5(a) and Table S7(a)), and 37 proteins showed differences larger than 2 times (*t* test, $P<0.01$, Table S7(b)), among which as many as 59.4% of the differentially expressed proteins (22 proteins) were reported as cancer metastasis-associated proteins, and 40.6% of the proteins (15 proteins) were not yet reported in cancer metastasis. Among them, we found that the RNA binding protein RNPS1, a general activator of the pre-mRNA splicing involved in preventing the formation of an R-loop, dramatically decreased in the MHCC97H cell quantified by LC-MS/MS and was further confirmed by western-blot (Figure 5(b)). RNPS1 was reported to be a crucial subunit of the apoptosis-and splicing-associated protein (ASAP) complex, which was thought to link RNA processing with apoptosis [38]. Our results suggest that the ASAP complex might play a crucial role in linking RNA processing with tumour metastasis. Another RNA binding protein SSB, which can bind to the 3′ poly(U) terminus of nascent mRNA to protect them from exonuclease digestion and facilitate their folding and maturation, increased by more than two times in MHCC97H cells, which was also confirmed by western-blot (Figure 5(b)). Though the SSB protein function in metastasis was unclear, it has been reported to stabilize one signal transducer and activator of the transcription protein STAT3 [39,40], activation of which is important for tumour metastasis by affecting cell adhesion. Therefore, we postulate that the SSB protein might promote cancer metastasis through the STAT3 mediate signal pathway. Further biological study of the relationship between such proteins with cancer metastasis is underway in our lab.

Recent studies have shown TNF-α promoted invasion and metastasis in hepatocellular carcinoma *via* the NF-Kappa B

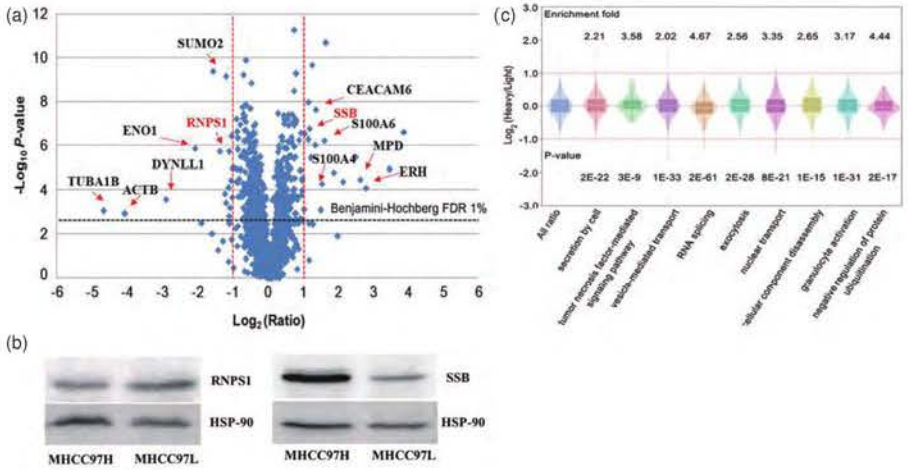

Figure 5 (a) Quantitative analysis of hepatocarcinoma ascites syngeneic cells with high (MHCC97H) and low (MHCC97L) lymph node metastasis rates. The proteins quantified in all six replicates were processed using Benjamini-Hochberg (BH) FDR estimations, and those that passed the 1% BH-FDR threshold were retained in the volcano plot. (b) Two differentially expressed proteins (RNPS1 and SSB) validated by Western blot assays. (c) Gene ontology analysis of quantified proteins from hepatocellular carcinoma (HCC) cells with high and low metastatic potential (MHCC97H/L). Violin plots of the distributions are shown with the corresponding p values (color online).

signalling pathway [41] and lymphatic metastasis of gallbladder cancer through the ERK1/2/AP-1/VEGF-D pathway [42]. Furthermore, RNA splicing and misregulated expression of members of the ubiquitin cascade represent a novel mechanism for controlling EMT, which is the critical early step for cancer cell metastasis [43,44]. In our results, gene ontology (GO) analysis (Figure 5(c), Table S8) suggests that proteins involved in the tumour necrosis factor (TNF)-mediated signal pathway were significantly increased in MHCC97H cells, and proteins involved in RNA splicing and negative regulation of protein ubiquitination were significantly decreased in MHCC97H cells. All these results demonstrate that our developed approach would provide a promising tool for screening more candidate target proteins, allowing for the development of novel drugs to suppress cancer progression and metastasis.

Except for those, we also expect its potential applications in personalized medicine and translational studies. Fast response time is frequently important such as in the case of myocardial infarction. Our workflow could use a very short LC-MS time (60 min), which could be reduced by further optimization, so that the entire procedure could conceivably be performed in less than 1 h, therefore it might provide timely data for clinical decision making. Furthermore, such a method could be also extended to automatically measure the absolute abundance of multiple known clinically diagnostic biomarkers [45,46].

4 Conclusions

In summary, we developed a FAST method, by which the whole sample preparation procedures including protein denaturation, reduction, desalting and digestion were integrated seamlessly without any sample transfer, permitting analysis of a minimal sample amount with high throughput and technical reproducibility. In contrast to the existing methods of sample preparation, such a simple and fast method required only 8–20 min, much shorter than several hours or days typically required using more complex workflows. Moreover, with the seamless conjugation of the FAST and "single-shot" LC-MS workflow, the unprecedented speed and precision of this proteomic analysis are achieved due to minimized sample handling and easy-to-prepare chromatography, making it easily transferable to other laboratories given established standard operating procedures.

Acknowledgements This work was supported by the National Key Research and Development Program of China (YS2019YFE020015, 2018YFC0910202, 2017YFA0505002) and the National Natural Science Foundation of China (21974136, 21725506, 91543201).

Conflict of interest The authors declare no conflict of interest.

Supporting information The supporting information is available online at http://chem.scichina.com and http://link.springer.com/journal/11426. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

1 Mertins P, Tang LC, Krug K, Clark DJ, Gritsenko MA, Chen L, Clauser KR, Clauss TR, Shah P, Gillette MA, Petyuk VA, Thomas SN, Mani DR, Mundt F, Moore RJ, Hu Y, Zhao R, Schnaubelt M, Keshishian H, Monroe ME, Zhang Z, Udeshi ND, Mani D, Davies SR, Townsend RR, Chan DW, Smith RD, Zhang H, Liu T, Carr SA. *Nat Protoc*, 2018, 13: 1632–1661
2 Chen N, Liu J, Qiao Z, Liu Y, Yang Y, Jiang C, Wang X, Wang C. *Chem Sci*, 2018, 9: 2826–2830
3 Zhang Z, Hebert AS, Westphall MS, Qu Y, Coon JJ, Dovichi NJ. *Anal*

Chem, 2018, 90: 12090–12093
4 Shen S, An B, Wang X, Hilchey SP, Li J, Cao J, Tian Y, Hu C, Jin L, Ng A, Tu C, Qu M, Zand MS, Qu J. *Anal Chem*, 2018, 90: 10350–10359
5 Holfeld A, Valdés A, Malmström PU, Segersten U, Lind SB. *Anal Chem*, 2018, 90: 5841–5849
6 Wu X, Xiong E, Wang W, Scali M, Cresti M. *Nat Protoc*, 2014, 9: 362–374
7 Wiśniewski JR, Zougman A, Nagaraj N, Mann M. *Nat Methods*, 2009, 6: 359–362
8 Wisniewski JR, Ostasiewicz P, Mann M. *J Proteome Res*, 2011, 10: 3040–3049
9 Olkowicz M, Jablonska P, Rogowski J, Smolenski RT. *Talanta*, 2018, 182: 492–499
10 Yu Y, Bekele S, Pieper R. *J Proteom*, 2017, 166: 1–7
11 Doellinger J, Schneider A, Hoeller M, Lasch P. *Mol Cell Proteom*, 2020, 19: 209–222
12 Zhu Y, Piehowski PD, Zhao R, Chen J, Shen Y, Moore RJ, Shukla AK, Petyuk VA, Campbell-Thompson M, Mathews CE, Smith RD, Qian WJ, Kelly RT. *Nat Commun*, 2018, 9: 882
13 Zhu Y, Dou M, Piehowski PD, Liang Y, Wang F, Chu RK, Chrisler WB, Smith JN, Schwarz KC, Shen Y, Shukla AK, Moore RJ, Smith RD, Qian WJ, Kelly RT. *Mol Cell Proteomics*, 2018, 17: 1864–1874
14 Zhu Y, Podolak J, Zhao R, Shukla AK, Moore RJ, Thomas GV, Kelly RT. *Anal Chem*, 2018, 90: 11756–11759
15 Zhu Y, Clair G, Chrisler WB, Shen Y, Zhao R, Shukla AK, Moore RJ, Misra RS, Pryhuber GS, Smith RD, Ansong C, Kelly RT. *Angew Chem Int Ed*, 2018, 57: 12370–12374
16 Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB, Krijgsveld J. *Nat Protoc*, 2019, 14: 68–85
17 Cleland TP. *J Proteome Res*, 2018, 17: 3976–3983
18 Moggridge S, Sorensen PH, Morin GB, Hughes CS. *J Proteome Res*, 2018, 17: 1730–1740
19 Chen Q, Yan G, Gao M, Zhang X. *Anal Chem*, 2015, 87: 6674–6680
20 Shao X, Wang X, Guan S, Lin H, Yan G, Gao M, Deng C, Zhang X. *Anal Chem*, 2018, 90: 14003–14010
21 Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. *Nat Meth,* 2014, 11: 319–324
22 Rieckmann JC, Geiger R, Hornburg D, Wolf T, Kveler K, Jarrossay D, Sallusto F, Shen-Orr SS, Lanzavecchia A, Mann M, Meissner F. *Nat Immunol*, 2017, 18: 583–593
23 Sielaff M, Kuharev J, Bohn T, Hahlbrock J, Bopp T, Tenzer S, Distler U. *J Proteome Res*, 2017, 16: 4060–4072
24 Ke M, Liu J, Chen W, Chen L, Gao W, Qin Y, He A, Chu B, Tang J, Xu R, Deng Y, Tian R. *Anal Chem*, 2018, 90: 12574–12583
25 Xue L, Lin L, Zhou W, Chen W, Tang J, Sun X, Huang P, Tian R. *J Chromatogr A*, 2018, 1564: 76–84
26 Chen W, Wang S, Adhikari S, Deng Z, Wang L, Chen L, Ke M, Yang P, Tian R. *Anal Chem*, 2016, 88: 4864–4871
27 Yuan H, Zhang S, Zhao B, Weng Y, Zhu X, Li S, Zhang L, Zhang Y. *Anal Chem*, 2017, 89: 6324–6329
28 Zhao Q, Fang F, Shan Y, Sui Z, Zhao B, Liang Z, Zhang L, Zhang Y. *Anal Chem*, 2017, 89: 5179–5185
29 Yuan H, Zhou Y, Xia S, Zhang L, Zhang X, Wu Q, Liang Z, Zhang Y. *Anal Chem*, 2012, 84: 5124–5132
30 Cox J, Mann M. *Nat Biotechnol*, 2008, 26: 1367–1372
31 Sharma K, Schmitt S, Bergner CG, Tyanova S, Kannaiyan N, Manrique-Hoyos N, Kongi K, Cantuti L, Hanisch UK, Philips MA, Rossner MJ, Mann M, Simons M. *Nat Neurosci*, 2015, 18: 1819–1831
32 Wiśniewski JR, Wegler C, Artursson P. *J Proteome Res,* 2019, 18: 217–224
33 Wiśniewski JR, Zielinska DF, Mann M. *Anal Biochem*, 2011, 410: 307–309
34 Erde J, Loo RRO, Loo JA. *J Proteome Res*, 2014, 13: 1885–1895
35 Yang Z, Shen X, Chen D, Sun L. *J Proteome Res*, 2020, 19: 3315–3325
36 Wang N, Xu M, Wang P, Li L. *Anal Chem*, 2010, 82: 2262–2271
37 Li S, Plouffe BD, Belov AM, Ray S, Wang X, Murthy SK, Karger BL, Ivanov AR. *Mol Cell Proteom*, 2015, 14: 1672–1683
38 Deka B, Singh KK. *Int J Biol Sci*, 2017, 13: 545–560
39 Kota V, Sommer G, Hazard ES, Hardiman G, Twiss JL, Heise T. *Mol Cell Biol*, 2017, 38: e00129
40 Kim KJ, Kwon SH, Yun JH, Jeong HS, Kim HR, Lee EH, Ye SK, Cho CH. *Oncogene*, 2017, 36: 5445–5459
41 Xiao Y, Huang S, Qiu F, Ding X, Sun Y, Wei C, Hu X, Wei K, Long S, Xie L, Xun Y, Chen W, Zhang Z, Liu N, Xiang S. *Ebiomedicine*, 2020, 51: 102603
42 Hong HJ, Jiang L, Lin YF, He CL, Zhu GW, Du Q, Wang XQ, She FF, Chen YL. *BMC Cancer*, 2016, 16: 240
43 Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh LA, Yang J, Cheng C. *J Clin Invest*, 2011, 121: 1064–1074
44 Gallo LH, Ko J, Donoghue DJ. *Cell Cycle*, 2017, 16: 634–648
45 Li Q, Jin J, Lou F, Tang D. *Sci China Chem*, 2018, 61: 750–756
46 Liu G, Zhang T, Tang Y, Miao P. *Sci China Chem*, 2018, 61: 393–396