Final version published as: Sun, B., Kovatch, J. R., Badiong, A., & Merbouh, N. (2017). Optimization and Modeling of Quadrupole Orbitrap Parameters for Sensitive Analysis toward Single-Cell Proteomics. Journal of Proteome Research, 16(10), 3711–3721. https://doi.org/10.1021/acs.jproteome.7b00416.

Optimization and Modeling of Quadrupole Orbitrap Parameters for Sensitive Analysis towards Single-Cell Proteomics

Bingyun Sun[1,2,*], Jessica Rae Kovatch[1,¶], Albert Badiong[3 ¶], Nabyl Merbouh[1]

1. Department of Chemistry, Simon Fraser University, Burnaby, British Columbia, Canada

2. Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada

3. Faculty of Health Science, Simon Fraser University, Burnaby, British Columbia, Canada

¶ Equal contribution

* Corresponding author, Bingyun Sun, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6, Email: bingyun_sun@sfu.ca

Abstract

Single-cell proteomics represents a field of extremely sensitive proteomic analysis, owning to the minute amount of yet complex proteins in a single cell. Without amplification potential as of nucleic acids, single-cell mass spectrometry (MS) analysis demands special instrumentation running with optimized parameters to maximize the sensitivity and throughput for comprehensive proteomic discovery. To facilitate such analysis, we here investigated two factors critical to peptide sequencing and protein detection in shotgun proteomics, i.e. precursor ion isolation window (IW) and maximum precursor ion injection time (ITmax), on an ultra-high-field quadrupole Orbitrap (Q-Exactive HF). Counter intuitive to the frequently used proteomic parameters for bulk samples (> 100 ng), our experimental data and subsequent modeling suggested a universally optimal IW of 4.0 Th for sample quantity raging from 100 ng to 1 ng, and a sample-quantity dependent ITmax of more than 250ms for 1-ng samples. Comparing with the benchmark condition of IW = 2.0 Th and ITmax = 50 ms, our optimization generated up to 300% increase to the detected protein groups for 1-ng samples. The additionally identified proteins allowed deeper penetration of proteome for better revealing crucial cellular functions such as signaling and cell adhesion. We hope this effort can prompt single-cell and trace proteomic analysis, and enable a rational selection of MS parameters.

Keywords:

shotgun proteomics; tandem mass spectrometry; LC-MS/MS; Orbitrap; ion injection time; mass isolation window, single-cell proteomics, trace proteomics

Introduction

Tandem mass spectrometry (MS) plays a critical role in proteomic analysis of complex biological systems. The advancement of contemporary MS instrumentation has greatly improved the accuracy, resolution, sensitivity, and throughput of the proteomic analysis[1-3]. Various methods developed on these instruments have enabled the complete proteome study first in lower organisms such as yeast and bacteria[4-6], and lately in mammals such as human [7-9] and mouse [10]. Sensitive ion trap designs, such as linear ion trap and Orbitrap, have also allowed reliable detection of attomole and subattomole of proteins from minute amount of sample at MS2 level [11-15], which fostered single-cell proteomics.

The need of understanding rare events and tissue complexity in biological systems cultivated the recent blossom of single-cell studies. Unprecedented information offered from single-cell analysis has been demonstrated by investigations of single-cell genomics, transcriptomics[16-19], and metabolomics[20,21]. Fluorescence and MS based imaging techniques are capable of offering spatial and sometimes temporal information of targeted molecules within a single cell across a large cell population. The direct analysis of thousands of proteins in a single MS run demonstrated by modern shotgun proteomics will be a powerful tool for single-cell studies. However, most of these proteomic analyses were carried out in bulk with more than micrograms of protein samples.

To effectively detect thousands of proteins estimated in a single cell, is technically challenging. The lack of amplification potential of proteins as compared to nucleic acids has limited the study of miniscule amount of proteins (> 0.05ng), from a single mammalian cell. MS has been considered as the most suitable approach to characterize the chemical composition of single cells by a recent insightful review [22]. The frequently used time-of-flight (TOF) based MS

analysis, including mass cytometry, MALDI-MS imaging, and secondary ion MS (SIMS) had provided exclusive information on single-cell biology. Yet, the identification of proteins in these techniques is constrained and indirect, such as the use of antibodies, which does not offer the same accuracy, specificity, and flexibility as sequencing based tandem MS (i.e. MS/MS) and shotgun proteomics. Using integrated technologies such as capillary electrophoresis coupled to MS/MS (CE-MS/MS), a few labs now are capable of detecting proteomes from individual special cells such as neurons and oocytes[23,24] with micrograms of total proteins. The front-end sample preparation, including protein extraction, denaturation, digestion, and desalting, operated at single cell level is extremely challenging and often suffers severe sample loss [25-27]. To study regular somatic cells of 10-20 um in diameter, researchers have pooled a few hundreds or thousands of cells together for manageable protein quantity [11,28,29].

With many single-cell proteomics efforts devoted to front-end sample preparation, less attention has been focused on MS parameters. The importance of MS parameters to successful peptide detection and protein inference cannot be overstated [30-33]. The optimization of MS parameters, however, has been focused primarily on bulk analysis, in which sample quantity is abundant and complex. Single-cell represented trace proteomics poses unique challenges that are different from bulk studies. It has been clear that sample quantity greatly influence the suitable MS parameters [32,34,35]. In trace proteomics, peptide signals can often be successfully registered in MS1 (survey) scan but failed to be sequenced in MS2 scan. The reason is mostly due to the insufficient number of precursor ions for high quality MS2 spectra. In ion traps, the number of precursor ions can be controlled both by the width of mass isolation window (IW), i.e. the ion flux, and the accumulation time of the target ions, i.e. ion injection time (IT). Widening the IW will increase the ion flux but reduce the ion purity, whereas extending the maximum IT

allowance (ITmax) will sacrifice the throughput, manifested by a slower analysis speed (i.e duty cycle) and the reduced total number of MS2 scans. In shotgun proteomics, the number of MS2 scans and the quality of these scans together determine the efficiency of peptide and protein identification. A balance between the MS2 quality and duty cycle (i.e. MS2 quantity) has been pinpointed by Randall et al. in their systematic optimization of quadrupole Oribtrap parameters using fractional factorial design [36]. Even though this large scale effort was carried out in bulk analysis using 200 ng of yeast digest, it set the foundation for follow-up studies including ours.

To facilitate single-cell MS proteomics, we decided to examine parameters that can improve MS2 spectra quality and peptide sequencing efficiency in trace samples, i.e. MS2 ITmax and IW. In recent years, the benefit of longer ITmax to low abundance samples has been identified, even though in bulk shortened ITmax (50-100ms) has increased the depth of proteome coverage [30,32]. Both ITmax of 250 ms in dynamic IT scheme[15] for < 5 ng samples and IT of 120 ms in fixed IT scheme [11] for 125-ng samples, have been implemented on quadrupole Orbitrap instrument. Conversely, the optimized IW value has been controversial. In isobaric labeling centered quantitative proteomics, such as TMT and iTRAQ, narrowing IW is the key to increase the quantification accuracy [37-39]. Yet, when the IW is narrower than 0.6 Da, the ion transmission efficiency is reduced significantly with a loss up to 95% in the first generation quadrupole Orbitrap (Q-Exactive) [11]. Michalski et al using LTQ-Velos discovered that IW of 4.0 Th offered the best protein identifications even though the interference to the precursors was as high as 50% [40]. In the second-generation ultra-high-field quadrupole Orbitrap (Q-Exactive HF), a segmented quadrupole situated immediately after a prefiltering flatapole can achieve a 2-fold increase of the ion transmission at narrow IW [33]. Scheltema et al [33] used Q-Exactive HF on the bulk analysis of

1 ug of yeast and HeLa cell digests, and identified an optimal IW of 1.4 Th with 70% ion transmission efficiency.

We are interested in the plausibility of widening IW on Q-Exactive HF for trace proteomics. The rationale is that even though widening IW risks the purity of precursor ions and increases the noise in MS2 scans, the Q-Exactive HF is an ultra-high-field Orbitrap that has much improved resolution and better mass accuracy to regular Orbitrap. The more compact Orbitrap cell[33] and the higher ion oscillation frequency in Q-Exactive HF should in theory dampen noise level in MS2 spectra[11] and effectively separate any confounding signals. A wider IW increases ion flux and ion transmission efficiency[11], which is crucial in trace proteomics. We therefore examined the potential of Q-Exactive HF for a widened IW in conjunction with the extended ITmax in trace proteomics. We used various quantities of Chinese hamster ovary (CHO) cell protein digest as the testing sample, and systematically varied ITmax and IW separately and synergistically in searching of their suitable values for trace proteomics.

Material and methods

Chemicals

Sequence-grade bovine pancreas TPCK-treated trypsin (EC 3.4.21.4) was obtained from Worthington Biochemical Co. (Lakewood, NJ, USA). Tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), iodoacetamide (IAA), protease inhibitor cocktail, Bradford reagent were obtained from Sigma-Aldrich (St. Louis, MO, USA). Rapigest and Sep-Pak C18 columns were from Waters (Milford, MA, USA), Amicon ultra-0.5mL centrifugal filters (10K m.w. cutoff) were from Millipore (Bedford, MA, USA). Dulbecco's modified Eagle's medium (DMEM), L-glutamine, and fetal bovine serum (FBS) were purchased from Lonza (Basel, Switzerland).

Stable isotope labeling by amino acids in cell culture (SILAC) suitable dialyzed FBS was from Life Technology (Burlington, ON, Canada). Lysines labeled with ^{13}C and ^{15}N were purchased from Cambridge Isotope Laboratory, Inc. (Tewksbury, MA, USA). SILAC compatible Dulbecco's modified Eagle's medium (DMEM), and the rest of chemicals were obtained from Fisher Scientific (Pittsburgh, PA, USA).

Cell culture and harvest

Regular CHO-K1 cells were cultured at 37 ℃ and 5% CO_2 in DMEM supplemented with L-glutamine and 10% FBS. SILAC labeled CHO-K1 cells were cultured under the similar condition as published [41,42], in which SILAC suitable DMEM was supplemented with ^{13}C and ^{15}N-labeled lysines, regular arginines, and 10% dialyzed FBS. Cells were doubled five times in SILAC media before checking the labeling efficiency. Both types of cells were harvested by washing the culture plate with cold phosphate-buffered saline (PBS), following scraping the plate in cold PBS with protease inhibitor cocktail, and pelleting by centrifugation. The SILAC labeled cells were mixed with regular cells at 5:1 ratio before down-stream proteomic processing.

Cell lysate and filtration aided sample preparation (FASP)

Cell pellet was lysed in denaturation buffer (5 mM EDTA, 10 mM TCEP, 0.5% rapigest, 8 M urea, and 40 mM tris at pH 8) and sonicated further to break down the chromosomal DNA. Protein quantity was checked by Bradford assay. The sample was then processed based on the reported FASP protocol [43] with minor modifications. In brief, the sample solution was transferred to an Amicon 10K ultra-0.5mL centrifugal filter vial, and spun at 14, 000 rpm by a

bench-top centrifuge for ~ 30 min to reduce the volume below 20uL. IAA solution of 15 mM in denaturation buffer was added to the filter and incubated with the protein sample at room temperature away from light for 15 min, before being removed by centrifuge to below 20 uL. Then the sample buffer was exchanged to digestion solution, which was 10 times diluted denaturation buffer, by 3 consecutive resuspensions and centrifugations. After final spin, 0.2 mL of the digestion solution together with sequence grade trypsin at 1:20 enzyme to protein ratio was added to the vial, and incubated at 37 ℃ overnight. The following day, the solution received final spin and the filtrate including the digested peptides was acidified and incubated at 37℃ to degrade the rapigest. The precipitant was removed by spin and the supernatant was desalted by a Sep-Pak C18 cartridge. The dried and cleaned peptides were reconstituted in MS loading buffer (0.1% formic acid, FA, and 1% acetonitrile in HPLC water).

UHPLC-MS/MS analysis and data search

An EASY-nLC 1000 system was coupled to the Q-Exactive HF through the nano-EASY spray source (all from Thermo Scientific, Mississauga, ON, Canada) for all of the LC-MS/MS analyses. Peptides of 100, 10, or 1 ng was loaded first onto a Pepmap 100 trapping column (Thermo Scientific, 20 mm x 75 μm ID, 3 μm C18 resin with 100 Å pore size) at constant pressure of 450 bar using buffer A (0.1% FA in HPLC water) before they were eluted on a Pepmap EASY-spray analytical column (Thermo Scientific, 150 mm x 75 μm ID, 3 μm C18 resin with 100 Å pore size) with an integrated spray tip. A 60 min-gradient of 2-35% buffer B (0.1% FA in acetonitrile) in buffer A at a constant flow rate of 200 nL/min was used to elute peptides off the analytical column. An voltage of 2.0 kV was used for electrospray, and the ion transfer tube that guided ions from spray into MS had a temperature of 250 ℃.

For Q-Exactive HF MS2 scans, data-dependent acquisition was employed to select top 10 most abundant precursors in MS 1 scans for higher energy collisional dissociation (HCD) with a dynamic exclusion of 10 sec. The MS1 scans were acquired at m/z range of 400-2000 with mass resolution of 60,000 and automatic gain control (AGC) target of 1E6. The threshold to trigger MS2 scans was 2.0E4; the normalized HCD energy was 22%; the resolved fragments were scanned at mass resolution of 15,000 and AGC target value of 1E5. Precursors with changes of 1, 5-7, >8, and unassigned were excluded from MS2 scans. The ITmax for MS1 was 100 ms. For MS2 scans, the ITmax was varied between 50-250 ms, and the IW was varied between 1.0-8.0 Th.

Data analysis

The raw files from Q-Exactive HF were processed by Proteome Discoverer 2.1 (Thermo Scientific, Mississauga, ON, Canada) and searched using Sequest HT. A default signal to noise ratio of 1.5 was applied to MS2 spectra without specification. The search sequence database was ipi.MOUSE.v3.82.fasta appended with common contaminating sequences (trypsin and keratin) [44]. The parameters for search were: 10 ppm precursor mass tolerance (monoisotopic mass), full trypic ends with 2 miss cleavages, minimum 6 amino acids, 0.05 maximum delta Cn, 0.05 Da fragment mass tolerance (monoisotopic mass). Cysteine carbamidomethylation was treated as a fixed modification, and methionine oxidation, SILAC labeling, and N-acetylation of protein were treated as dynamic modifications. A decoy database comprised of the reversed sequences in the target database was used for FDR estimation (strict FDR cutoff: 0.01), and Percolator was used to evaluate false positives. The final result was filtered against 0.01 FDR for high peptide

confidence level, and protein groups were considered as the protein identification in which congeneric proteins identified by the same peptides were grouped into one.

Gene ontology (GO) enrichment analysis was performed using DAVID (Database for Annotation, Visualization and Integrated Discovery, v6.8, https://david.ncifcrf.gov/) [45,46]. The comparison was carried out on proteins detected from 10-ng samples using optimized parameters (ITmax of 100 ms and IW of 4.0 Th), as well as those from benchmark parameters (ITmax of 50 ms and IW of 2.0 Th). We first obtained commonly and uniquely identified proteins from the two results, and separately performed the GO biological process (GO_BP) enrichment analysis on these proteins. Enrichment p value of 0.05 was used as the cutoff.

Results and Discussion

Contemporary quadrupole Orbitrap (Q-Exactive HF) has embraced several recent technological advancements[47] such as an enhanced Fourier transform[48], a segmented quadrupole[33], a compact ultra-high field Oribrap cell[33] together with a flatapole prefiltering design[33] and a higher energy collisional dissociation (HCD) cell[49]. These features enable better sensitivity, resolution, and speed than the earlier model (Q-Exactive). Taking advantage of these newly acquired instrumental features, we examined two closely related parameters of MS2 scans, i.e. ITmax and IW, for effective proteomic analysis towards single-cell represented trace proteomics. Successful trace proteomics requires to balance the tradeoff between sensitivity and throughput, which is in part determined by the MS2-scan quality and quantity, for limited sample quantity and high complexity. A conservative estimation by Schmid et al. [50] indicated 10^9 total number of proteins per 10 um mammalian cell with 100,000 average copy number (0.17 attomole) for 1000 protein species. Based on the detected dynamic range from bulk, 10^6-10^7

concentration difference exists in cellular proteins. It is therefore reasonable to estimate the abundant proteins in a single cell having attomole and subattomole of quantity that should be detected by sensitive MS if parameters are optimal.

To maximize MS capacity in trace proteomics, we focused on the MS2 ITmax and IW that are key to the sensitivity and throughput of the peptide detection. To avoid front-end sample preparation caused bias but target to MS performance, we diluted a CHO cell digest to 1-100 ng range, with an internal dilution down to 0.2 ng introduced by SILAC labeling. These sample quantities can roughly represent proteins from 200-2000 cells (100 ng), 2-20 cells (1 ng) and 0.4-4 cells (0.2ng) based on various reported single-cell protein quantities [50, 26, 27].

MS2 maximum Ion injection time (ITmax). MS2 ITmax has been a frequently varied parameter in the dynamic IT scheme for analyses of low abundance samples [32,35]. The dynamic IT in quadrupole Orbitrap is to ensure the relatively constant ion abundance, i.e. AGC (automatic gain control) target, when the number of precursors fluctuate in complex samples with high dynamic range. Exceeding the AGC target will trigger the space-to-charge effect and reduce the accuracy to the measured m/z ratio. Substantially below the AGC target such as in the case of single-cell proteomics, the MS2 scan quality will be compromised. To avoid prolonged injection in trace sample analysis, an ITmax is required in dynamic IT scheme. Extending ITmax allows low abundance ions to accumulate in the C trap for high quality MS2 spectra, but reduces sampling frequency and the quantity of MS2 scans.

To seek the optimal ITmax suitable for single-cell proteomics, we explored the sample-quantity dependent peptide identification as a function of ITmax. To do so, we systematically varied the ITmax from 50 to 250 ms, and examined the CHO cell peptide detection from total

proteins of 1, 10, and 100 ng. First we evaluated the actual MS2 IT that resolved peptide identification (peptide spectra matching, PSMs) as a function of ITmax shown in Fig. 1A. In the figure, the 100-ng samples exhibited obvious difference from the rest, in which the ITavg of 100-ng shortened from its ITmax, whereas the ITavg of both 1 and 10 ng samples was almost identical as the ITmax in the entire range of 50 to 250 ms. This observation suggested that for low abundance samples (< 10 ng), the ITmax determined the IT, and most MS2 scans did not reach the AGC target. This dependency is important to explain many observations below, and can be used as a hallmark to distinguish trace from bulk shotgun proteomics.

To evaluate the sample-quantity dependent balance of MS2 scan quality and quantity for optimal peptide detection, we compared the MS2 quantity and the detected PSMs as a function of ITmax for the three tested sample quantities as shown in Fig. 1B. The results unveiled a complex relationship: from a strong MS2 scan quantity dependent peptide identification in bulk (100 ng), to a mixed model at 10 ng, and a strong MS2 scan quality dependence at 1 ng.

In detail at 100-ng quantity, the number of PSMs dropped together with the decrease of MS2 scans when the ITmax was increased. The Pearson correlation between PSMs and MS2 quantity had a correlation coefficient, r, of 0.99 and p of 0.001. A result suggested a strong MS2 quantity depended peptide identification, which was in congruent with the previous bulk studies[30,31]. Even though some of the previous studies were carried out in the LTQ-Orbitrap hybrid systems, but they all represented a MS2-quantity dominated peptide identification.

At 10-ng protein quantity, however, the trend was different. The MS2 scans still showed a consistent decrease to the increase of ITmax, but a local maximum of PSMs at ITmax of 100-ms disrupted the MS2 quantity controlled PSM detection in 100-ng results. Longer ITmax can increase precursor population in the trap and the MS2 scan quality as discussed above. The

observed PSM peak indicated a transition from quantity to quality of peptide detection at 10-ng sample level, which we named it as a mixed model. This complex response suggested that for samples with a quantity ~10ng, efforts were needed to optimize the ITmax, which as shown in Fig. 1B can lead to 50% detection difference, and further verified in Fig. 4B to be discussed later.

Interestingly, at 1-ng sample quantity, a continuous rise of PSMs as a function of ITmax was observed as opposed to the descending trend in the 100-ng sample. This result suggested a complete reversal of MS2 quantity-control in peptide identification at this sample quantity. The trend, however, did not indicate the existence of an optimal ITmax for the best PSMs.

To better understand our observation, we quantitatively evaluated the gain of MS2 quality as the function of ITmax based on the model below. Firstly, as the quality of MS2 scans depends on the number of target ions, which can be computed by the total number of ions in the trap and the percentage of precursors, we used the following equation to describe,

$$MS2\ scan\ quality \propto N_{total\ ions} * (1 - interference). \quad (1)$$

The total number of ions ($N_{total\ ions}$) in the C-trap can be measured by the percentage of AGC (AGC%), and its relationship to the ITmax can be learned from in Fig. 1C. At 1 ng, a linear correlation between ITmax and average AGC% was observed with a correlation coefficient of 1, in which AGC% = 0.0004*ITmax. In addition, the percentage of interference (interference%) as a function of ITmax was also obtained in Fig. 1D. At 1 ng, a constant interference was independent from ITmax with an average value of 47 ± 1%. Therefore,

$$MS2\ scan\ quality \propto AGC\% \quad (2).$$

Secondly, we examined the MS2 scan quantity, i.e. the number of MS2 scans in Fig. 1B. In the figure, an initial increase of MS2 quantity was uniquely observed for 1-ng samples as

compared to the constant decrease of MS2 scans with the increase of ITmax for 10 and 100-ng samples. We noted that it was caused by a complex relationship between the MS2 trigger threshold and duty cycle. In Q-Exactive HF, to maintain a fixed AGC, an increase of ITmax is accompanied by a decrease of the MS2 trigger threshold adjusted automatically by Xcalibur, the controlling software. At ITmax of 50 ms, the threshold of MS2 fragmentation was the highest (2E4). When the ITmax raised from 50ms to 100ms, the threshold was lowered 100% (1E4). Once the sample was relatively abundant (100ng and 10ng) and our predefined top10 MS2 scan frequency per duty cycle was mostly saturated, the decrease of MS2 trigger threshold did not affect the proportion of the peptide precursor ions being fragmented by MS2 scans (< 3% change).

For 1 ng sample, however, due to the ubiquitously low abundance of the precursor ions, not enough precursor ions met the threshold for fragmentation, and the duty cycle only had an average of 1.6 MS2 scans/cycle at ITmax of 50 ms. Lowering the threshold from 2E4 to 1E4 had drastically increased the average MS2 scans from 1.6 to 5.8 MS2 scans /cycle, and tripled the number of peptide precursors being fragmented, which compensated the decrease in sampling frequency ($N_{MS1\ scans}$) that was due to the increase of the duty cycle. Therefore, the change of MS2 scan quantity in Fig. 1B is the net result of the increased MS2 scans/cycle and the decreased sampling frequency.

To combine the MS2 quality and quantity effect together for a better understanding of the PSM values, we proposed the following mathematical model,

$$PSM \propto MS2\ scan\ quality * MS2\ scan\ quantity \quad (3).$$

The product of the two factors in the model ensures zero PSM value if any one of the factors goes to zero. For 1-ng samples, when replace MS2 scan quantity in equation (3) with equation (2),

$$PSM \propto AGC\% * N_{MS2\ scans} \quad (4).$$

Based on the relationship observed in Fig. 1C for 1-ng samples, AGC% can be expressed as:

$$AGC\% = k * ITmax\ of\ MS2 \quad (5),$$

in which k is a protein quantity-dependent constant. In addition, the total analysis time (T), which is the gradient time used for LC separation, determines the number of MS1 and MS2 scans with following relationship:

$$T = N_{MS2\ scans} * IT\ of\ MS2 + N_{MS1\ scans} * IT\ of\ MS1 \quad (6).$$

Based on the observation in Fig. 1A for 1 ng sample, the ITavg of MS2 scan can be replaced by ITmax, equation (6) can then be expressed as:

$$T = N_{MS2\ scans} * ITmax\ of\ MS2 + N_{MS1\ scans} * IT\ of\ MS1 \quad (7).$$

Then, the number of total MS2 scans can be computed as following

$$N_{MS2\ scans} = \frac{T - N_{MS1\ scans} * IT\ of\ MS1}{ITmax\ of\ MS2} \quad (8).$$

When replacing AGC% and $N_{MS2\ scans}$ in equation (4) with equation (5) and (8), the number of PSMs is expressed as below:

$$PSM \propto k * T - k * N_{MS1\ scans} * IT\ of\ MS1 \quad (9).$$

In equation (9), IT of MS1 is relatively constant and small regardless of ITmax of MS2, as most AGC of MS1 can be reached even at 1-ng protein quantity. Therefore, when the ITmax of MS2 increases, the MS1 scan number decreases, and the PSMs converge towards a maximum that is independent of ITmax but associates with k (ratio of AGC% and ITmax) and the total analysis time. Further, using equation (9) and for 1ng sample at ITmax 250 ms, with learned k value of

0.0004 AGC%/ms, and measured T value of 60 min, $N_{MS1\ scans}$ of 1887, and ITavg of MS1 as 19.6 ms, we computed that 90% of PSMmax had been reached.

To verify the existence of PSMmax and the reaching of 90% of PSMmax at ITmax of ~250ms, we repeated the 1-ng experiments and extended ITmax to 400ms as shown in Fig. S1. The result verified our modeling, with saturated PSMs over 250 ms and the maximum detected protein groups of 233 and unique peptides of 559, at 400 ms ITmax. Therefore in our experimental condition, an ITmax of > 250 ms would be the optimal value for trace proteomics. The ITmax of 250ms had also been chosen by Sun et al. on their characterization of the proteome of 1 and 5 ng of RAW 264.7 cell lysate by Q-Exactive[35].

MS2 isolation window (IW). Besides ITmax, to effectively increase ion abundance in the C-trap for high quality MS2 scans, we were also interested in optimizing IW. A wide IW of 4.0 Th had been employed in the Orbitrap Velos for bulk analysis by Michalski et al. [40] after evaluating IW from 1.0 to 32 Th, in which MS2 scans were conducted in linear ion trap, LTQ. The results showed high sensitivity to low abundance ions, even though the interference can be as high as at 50%. The benefit of widened IW is not only for a better transmission of the monoisotopic precursor ions, but also to allow the subsequent isotopic ions of the precursor to enter the trap. The tradeoff for a wider IW is the risk of the reduced precursor purity and increased noise. Yet Oribtrap is known to possess higher resolution than LTQ as previously reported [51], which can reduce the noise, increase the mass accuracy, and provide a better separation of the target fragment signals from those of the co-isolates. The ultra-high-field Q-Exactive HF has close to 2 fold higher resolution than regular Orbitrap Q-Exactive. If LTQ can tolerate wider IW for the successful sensitive detection, Q-Exactive HF should in theory do better.

Widened IW in general is not preferred in bulk analysis, particularly due to the signal suppression experienced in the low-mass reporter ion region in iTRAQ and TMT, with which narrow IW less than 2.0 Th is necessary [38,39]. Therefore, the effort of IW study has been focused on the extremely narrow value. Scheltema et al. had evaluated Q-Exactive HF with IW ranging from 0.4 to 2.4 Th, and discovered an optimal IW of 1.4 Th with the co-isolates ("second peptide") of 6.5% [33].

Nevertheless, the study of sample-quantity dependent IW effect is elusive. Previous studies examining the LC separation condition to MS detection had reported that lowering the loading amount from 4 ug to 10 ng can half the elution peak width [52]. Narrower peaks and more resolved LC separation in trace samples can then minimize the presence of confounding peptides. Indeed, reports studying coisolates also noticed a 50% reduction of the presence of chimeric peptides when the loading quantity was reduced from 2 to 0.2 ug[53]. With the increased separation efficiency of LC in trace samples, and the increased resolution of the Orbitrap cell, we asked whether Q-Exactive HF can behave similarly as Velos under widened IW for trace proteomics.

Towards this end, we evaluated the PSMs as the function of IW with varied total protein quantities. Fig. 2A summarizes the results of the MS2 scans and PSMs as a function of IW for 1, 10, and 100 ng samples. Interestingly, Fig. 2A exhibited a unified trend irrespective of the varied total quantity of proteins, in which a local maximum of both PSMs and the number of MS2 scans was observed when IW varied from 1.0 – 8.0 Th. The maximum value of PSMs was shown at IW of 4.0 Th, which agreed with that of Michalski et al. [40] but not Scheltema et al. [33]. As Scheltema and colleagues used percentage of unique peptides instead of the number of PSMs as the evaluation criteria, we also analyzed the percentage of unique peptides (i.e. peptide groups)

as a function of IW as shown in Fig. S2. In the figure, we observed a similar trend as of Scheltema et al, in which IW of 1.0 Th had the highest percentage of unique peptide detection in all three protein quantities examined. The seemingly contradictory results were due to the low peptide identifications made at narrow IW.

To better understand the cause of an optimal IW of 4.0 Th, and to further evaluate our model, we examined the AGC% and the interference% as a function of IW as shown in Fig. 2B and 2C. As expected, Fig. 2B shows an increase of AGC% with widened IW, and a positive linear correlation ($p < 0.0006$) when the sample quantity was 1 and 10 ng. At 100ng, a positive linear correlation ($P < 0.0001$) to natural logarithm transformed IW was observed suggesting a more frequent saturation of AGC target compared to those of 1 and 10 ng. For interferences in Fig. 2C, independent to sample quantity, a positive linear correlation ($P < 0.004$) was observed between the percentage of interferences of MS2 scans and IW, suggesting an increased noise when widening IW that also agreed with previous studies[40].

To evaluate the MS2 scan quality model described by equation (1), we used the learned correlation in Figs. 2B & 2C to replace the $N_{total\ ions}$ (AGC%) and the interference value in equation (1). The simulated MS2 scan quality was shown in Figs. 3A-3C for all three protein quantities. Interestingly in these figures, even though the value of AGC% altered, the trend of AGC% as a function of IW was similar in all three sample quantities, and all peaked between 4-6 Th, with 100-ng leaning more towards 4 Th, and 1 ng more towards 6 Th. Both the trend and the optimal IW from simulation agreed with the experimental observation shown in Figs. 2A-2C, supporting the contribution of MS2 spectra quality to the peptide detection.

Subsequently, we examined the quantitative influence of MS2 scans to PSMs shown in Fig. 3D. The results peaked also at ~4 Th for all three sample quantities. As a result, both the

MS2 scan quality and the MS2 scan quantity lead to an optimal IW ~4 Th, particularly for 1-ng samples, which explained the unified PSM detection peaking at IW of 4 Th in Fig. 2A.

In Fig. 3D, the initial rise of MS2 scan number was due to the decreased MS2 trigger threshold on the targeted precursor, abbreviated as MS2 target trigger threshold. MS2 trigger threshold is a parameter predefined in the MS acquiring method as a ratio relative to AGC target. This threshold is used to gauge the total number of ions in the C-trap. With a constant MS2 threshold, widening the IW increased noise as shown in Fig. 2C, which led to a decrease of the need of precursor ions. We considered this indirect effect as a decreased MS2 "target" trigger threshold. So the end result of wider IW is similar to the longer ITmax that the number of MS2 scans was increased. The subsequent slight decrease of MS2 scans as IW wider than 4.0 Th was accompanied by the slight increase of MS1 scans, which was the combined effect of the shortened duty cycles as ion flux was increased and the dynamic exclusion in MS2 scans.

Effect of MS2 threshold. In previous bulk studies[36], lowering MS2 isolation threshold had improved protein identification. Our investigation here on both ITmax and IW effect also directly or indirectly lowered the MS2 trigger threshold, we therefore examined this factor. Without changing the ITmax and IW, lowering the trigger threshold 10 fold did not show improved protein detection as shown in Supplementary Fig. S3, even though the total number of MS2 scans was increased. The reason was due to the decreased MS2 scan quality represented by a lowered AGC%, which diminished the gain of MS2 scan quantity.

Robustness of the observation. To further test the robustness of our observation, we selected sample quantity of 10 ng, in which ITmax evaluation showed a mixed model (Fig. 1B), and

performed three repeated analysis on the ITmax and IW effects. Fig. 4 shows the detected peptide and protein groups as opposed to the PSMs in Figs. 1 and 2. Previously, detected protein and peptide groups had reported for less effect than the PSMs when optimizing MS parameters in bulk[30]. In Fig. 4 we observed a consistent maximal detection of both protein and peptide groups at IW of 4.0 Th and ITmax of 100-150 ms, which were similar to those of PSMs shown in Figs. 2A and Fig. 1B, respectively. The improvement gained at optimal condition compared with our benchmark, i.e. IW of 2.0 Th and ITmax of 50 ms, was significant ($p < 0.05$). The repeated experiments further verified the trend of our observation. Additional experimental verifications were also carried out in the synergistic optimization described below.

Synergistic effect between MS2 ITmax and IW. To evaluate the synergistic effect, we performed a set of combinatorial experiments with two ITmax values of 50 and 100 ms and two IW values of 2.0 and 4.0 Th, in which we chose IW of 2.0 Th and ITmax of 50 ms as the benchmark values. Under each condition, three technical replicates were obtained, and the numbers of the searched MS2 spectra, the PSMs, and the peptide and protein groups were plotted as histogram in Fig. 5. As anticipated when separately changing IW and ITmax, the results reproduced the observations in Figs. 1 & 2. The synergistic optimization of both ITmax and IW showed the largest improvement on the detection of PSMs, peptides and proteins, as compared to the individual factors, and more prominent when sample quantity was low (1 and 10 ng), and less for 100-ng samples. Specifically, at 100-ng quantity, only a slight increase of 2.9% protein groups was observed to the benchmark; yet at 10 and 1 ng quantity level, the increase was 35% and 150% respectively. To evaluate the impact of ITmax and IW to the detection reliability, we also separately analyzed the results using elevated S/N ratio cutoff of 3.0 instead of the default 1.5,

and the results were summarized in the Supplementary Fig. S3. A similar trend as of Fig. 5 was shown in Fig. S4.

Impact to biological information. To evaluate the impact of such optimization to biological insights, we performed the gene ontology (GO) enrichment analysis on the detected protein groups from 10-ng protein samples using the optimal (ITmax of 100 ms and IW of 4.0 Th) and the benchmark sets. Fig. 6 summarizes the results. Fig. 6A shows the Venn diagram comparison of the overall detected protein groups under the two conditions. The details of these proteins were summarized in Supplementary Table S1. Fig. 6B shows the pie charts on the –log(p-value) of GO_BP terms performed on the top ranked commonly and uniquely identified proteins in the two conditions. The number of corresponding proteins were listed the brackets in Fig. 6B next to the pie chart. The complete list of the enriched GO terms is in supplementary Table S2. The optimized parameters outperformed the benchmark in all enriched categories. The commonly detected proteins were enriched of known and abundant proteins such as "translation", "metabolic process", and "protein transport". Yet the optimal parameter set also identified many additional proteins in GO terms related to membrane proteins and signaling, such as "positive regulation of substrate adhesion-dependent cell spreading", "regulation of actin cytoskeleton organization", and "regulation of Rac protein signal transduction". No enrichment was obtained from proteins uniquely identified in benchmark results. These results demonstrated the importance and necessity to optimize MS parameters for deeper penetration in trace proteomics.

Sensitivity beyond 1 ng of the total proteins. To evaluate the possibility to detect proteins with less than 1 ng quantity, we had introduced a 1:5 protein quantity difference using lysine labeling,

in which light lysines were 5 fold less than heavy lysines. In 1 ng of sample, the light labeled proteins were about 0.2 ng. Under such low abundance, only 28 protein groups and 40 peptide groups were detected, accounted for 10% and 6% of detected protein and peptide groups from 1 ng, respectively. Because the identification was few, we did not pursue further analysis on these results.

Conclusion

In shotgun proteomics, protein identification relies on the obtained MS2 spectra. The quality and quantity of MS2 scans directly affect the sensitivity and throughput of proteomics detection, thereof the comprehensiveness and the reliability of the derived proteome. In trace proteomics in which scarce amount of sample is available and non-specific adsorption caused sample loss in front-end sample preparation is severe and challenging to overcome, the optimization of the MS parameters designated for trace samples is cost-benefit to improve proteomics outcome.

Considering the unique characteristics in trace proteomics, we evaluated the optimal value of Q-Exactive HF MS2 ITmax and IW for better peptide and protein detection. Our results demonstrated that more than 300% gain to over 200 protein groups can be achieved at 1-ng sample quantity after optimization. Our mathematical modeling for the first time quantified the tradeoff between MS2 quality and quantity, and guided us to discover an ITmax > 250ms and IW of 4.0 Th for the most effective analysis of 1 ng samples. Our study also revealed a complex transition for optimal MS parameters when the sample quantity changed from bulk (e.g., 100 ng) to trace (e.g., 1 ng) proteomics.

Since the advent of shotgun proteomics after the successful development of the first automatic search engine — Sequest— in 1994 [54], the optimization of MS parameters has been largely empirical. Numerous parameters to define the run condition of MS analysis render the optimization a daunting task, and many strategies have been developed to evaluate a subset of these parameters, such as the fractional factorial analysis [36] and genetic search [55]. Even though every peptide in the complex sample has its own optimal detection parameters, it does not exclude that a universal parameter set can yield the best protein detection of the entire sample. Our successful modeling of a subset of MS parameters marks an important step towards a rational design of instrumental parameters for trace proteomics. We hope our effort can promote future advancement in single-cell proteomics.

Acknowledgements:

This research was funded by Stem Cell Network of Canada, Canada Foundation for Innovation, Compute Canada, British Columbia Knowledge Development Fund, Natural Sciences and Engineering Research Console of Canada, and Simon Fraser University. J. R. K. was supported in part by the training fellowship of British Columbia Proteomics Network. We acknowledge the technical contribution of Prabhpreet K. Dhillon.

Supporting information:

The following files are available free of charge at ACS website http://pubs.acs.org:

Table S1. Protein identification for 10-ng CHO lysate using benchmark and optimal instrumental parameters.

Table S2. GO enrichment analysis of proteins identified in Supplementary Table S1.

Fig S1. Numbers of MS2 scans and PSMs as a function of ITmax for trypsin digest of 1-ng CHO-cell lysate acquired by Q-Exactive HF.

Fig. S2. Percentage of uniquely detected peptides in CHO-cell lysate digests of 1, 10, and 100 ng as a function of IW.

Fig. S3. Protein groups detected from 1, 10, and 100 ng of trypsin digest of CHO-cell lysate at regular (2E4) and low (2E3) threshold of MS2 fragmentation defined in Q-Exactive HF at MS2 ITmax of 50 ms and IW of 2.0 Th.

Fig. S4. Number of MS2 scans, PSMs, peptide and protein groups obtained from 1, 10, and 100 ng of trypsin digest of CHO-cell lysate with a minimum S/N of 3.0 in the Orbitrap for more reliable detection compared to the default S/N of 1.5.

References:

(1) Cravatt, B. F.; Simon, G. M.; Yates, J. R., 3rd. The biological impact of mass-spectrometry-based proteomics, *Nature* **2007**, *450*, 991-1000.

(2) Choudhary, C.; Mann, M. Decoding signalling networks by mass spectrometry-based proteomics, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 427-439.

(3) Lamond, A. I.; Uhlen, M.; Horning, S., et al. Advancing cell biology through proteomics in space and time, *Mol. Cell Proteomics* **2012**, *11*, 1-12.

(4) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The one hour yeast proteome, *Mol. Cell Proteomics* **2014**, *13*, 339-347.

(5) Nagaraj, N.; Kulak, N. A.; Cox, J.; Neuhauser, N.; Mayr, K.; Hoerning, O.; Vorm, O.; Mann, M. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top orbitrap, *Mol. Cell Proteomics* **2012**, *11*, 1-11.

(6) Picotti, P.; Clement-Ziza, M.; Lam, H., et al. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis, *Nature* **2013**, *494*, 266-270.

(7) Kim, M. S.; Pinto, S. M.; Getnet, D., et al. A draft map of the human proteome, *Nature* **2014**, *509*, 575-581.

(8) Uhlen, M.; Fagerberg, L.; Hallstrom, B. M., et al. Proteomics. Tissue-based map of the human proteome, *Science* **2015**, *347*, 1260419.

(9) Wilhelm, M.; Schlegl, J.; Hahne, H., et al. Mass-spectrometry-based draft of the human proteome, *Nature* **2014**, *509*, 582-587.

(10) Geiger, T.; Velic, A.; Macek, B.; Lundberg, E.; Kampf, C.; Nagaraj, N.; Uhlen, M.; Cox, J.; Mann, M. Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse, *Molecular & cellular proteomics : MCP* **2013**, *12*, 1709-1722.

(11) Chen, Q.; Yan, G.; Gao, M.; Zhang, X. Ultrasensitive Proteome Profiling for 100 Living Cells by Direct Cell Injection, Online Digestion and Nano-LC-MS/MS Analysis, *Anal. Chem.* **2015**, *87*, 6674-6680.

(12) Choi, S. B.; Zamarbide, M.; Manzini, M. C.; Nemes, P. Tapered-Tip Capillary Electrophoresis Nano-Electrospray Ionization Mass Spectrometry for Ultrasensitive Proteomics: the Mouse Cortex, *J. Am. Soc. Mass Spectrom.* **2016**, *28*, 597-607.

(13) Hanke, S.; Besir, H.; Oesterhelt, D.; Mann, M. Absolute SILAC for accurate quantification of proteins in complex mixtures down to the attomole level, *J. Proteome Res.* **2008**, *7*, 1118-1130.

(14) Shen, Y.; Tolic, N.; Masselon, C.; Pasa-Tolic, L.; Camp II, D. G.; Hixon, K. K.; Zhao, R.; Anderson, G. A.; Smith, R. D. Ultrasensitive proteomics using high-efficiency on-line micro-SPE-nanoLC-nanoESI MS and MS/MS, *Anal. Chem.* **2004**, *76*, 144-154.

(15) Sun, L.; Zhu, G.; Zhao, Y.; Yan, X.; Mou, S.; Dovichi, N. J. Ultrasensitive and fast bottom-up analysis of femtogram amounts of complex proteome digests, *Angew Chem. Int. Ed. Engl.* **2013**, *52*, 13661-13664.

(16) Ciuffi, A.; Rato, S.; Telenti, A. Single-Cell Genomics for Virology, *Viruses* **2016**, *8*, doi: 10.3390/v8050123.

(17) Saadatpour, A.; Lai, S.; Guo, G.; Yuan, G. C. Single-Cell Analysis in Cancer Genomics, *Trends Genet.* **2015**, *31*, 576-586.

(18) Saliba, A. E.; Westermann, A. J.; Gorski, S. A.; Vogel, J. Single-cell RNA-seq: advances and future challenges, *Nucleic Acids Res.* **2014**, *42*, 8845-8860.

(19) Wang, Y.; Navin, N. E. Advances and applications of single-cell sequencing technologies, *Mol. Cell* **2015**, *58*, 598-609.

(20) Onjiko, R. M.; Moody, S. A.; Nemes, P. Single-cell mass spectrometry reveals small molecules that affect cell fates in the 16-cell embryo, *Proc. Natl. Acad. Sci. U S A* **2015**, *112*, 6545-6550.

(21) Onjiko, R. M.; Morris, S. E.; Moody, S. A.; Nemes, P. Single-cell mass spectrometry with multi-solvent extraction identifies metabolic differences between left and right blastomeres in the 8-cell frog (Xenopus) embryo, *Analyst.* **2016**, *141*, 3648-3656.

(22) Comi, T. J.; Do, T. D.; Rubakhin, S. S.; Sweedler, J. V. Categorizing Cells on the Basis of their Chemical Profiles: Progress in Single-Cell Mass Spectrometry, *J. Am. Chem. Soc.* **2017**, *139*, 3920-3929.

(23) Lombard-Banek, C.; Moody, S. A.; Nemes, P. High-Sensitivity Mass Spectrometry for Probing Gene Translation in Single Embryonic Cells in the Early Frog (Xenopus) Embryo, *Frontiers in cell and developmental biology* **2016**, *4*, 100.

(24) Sun, L.; Dubiak, K. M.; Peuchen, E. H.; Zhang, Z.; Zhu, G.; Huber, P. W.; Dovichi, N. J. Single Cell Proteomics Using Frog (Xenopus laevis) Blastomeres Isolated from Early Stage Embryos, Which Form a Geometric Progression in Protein Content, *Anal Chem* **2016**, *88*, 6653-6657.

(25) Lee, M. C.; Wu, K. S.; Nguyen, T. N.; Sun, B. Sodium dodecyl sulfate polyacrylamide gel electrophoresis for direct quantitation of protein adsorption, *Anal. Biochem.* **2014**, *465*, 102-104.

(26) Rubakhin, S. S.; Romanova, E. V.; Nemes, P.; Sweedler, J. V. Profiling metabolites and peptides in single cells, *Nat. Methods* **2011**, *8*, S20-29.

(27) Zhou, H.; Ning, Z.; Wang, F.; Seebun, D.; Figeys, D. Proteomic reactors and their applications in biology, *FEBS J.* **2011**, *278*, 3796-3806.

(28) Ethier, M.; Hou, W.; Duewel, H. S.; Figeys, D. The proteomic reactor: A microfluidic device for processing minute amounts of protein prior to mass spectrometry analysis, *Journal of Proteome Research* **2006**, *5*, 2754-2759.

(29) Wang, N.; Xu, M.; Wang, P.; Li, L. Development of Mass Spectrometry-Based Shotgun Method for Proteome Analysis of 500 to 5000 Cancer Cells, *Analytical Chemistry* **2010**, *82*, 2262-2271.

(30) Kalli, A.; Hess, S. Effect of mass spectrometric parameters on peptide and protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer, *Proteomics* **2012**, *12*, 21-31.

(31) Kalli, A.; Smith, G. T.; Sweredoski, M. J.; Hess, S. Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers, *J. Proteome Res.* **2013**, *12*, 3071-3086.

(32) Kelstrup, C. D.; Young, C.; Lavallee, R.; Nielsen, M. L.; Olsen, J. V. Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer, *J. Proteome Res.* **2012**, *11*, 3487-3497.

(33) Scheltema, R. A.; Hauschild, J. P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a benchtop mass spectrometer with a pre-filer, high performace quadrupole and an ultra-high-field orbitrap analyzer, *Mol. Cell Proteomics* **2014**, *13*, 3698-3708.

(34) Andrews, G. L.; Dean, R. A.; Hawkridge, A. M.; Muddiman, D. C. Improving proteome coverage on a LTQ-Orbitrap using design of experiments, *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 773-783.

(35) Sun, L.; Zhu, G.; Dovichi, N. J. Comparison of the LTQ-Orbitrap Velos and the Q-Exactive for proteomic analysis of 1-1000 ng RAW 264.7 cell lysate digests, *Rapid. Commun. Mass Spectrom.* **2013**, *27*, 157-162.

(36) Randall, S. M.; Cardasis, H. L.; Muddiman, D. C. Factorial experimental designs elucidate significant variables affecting data acquisition on a quadrupole orbitrap mass spectrometer, *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1501-1512.

(37) Kocher, T.; Pichler, P.; Schutzbier, M.; Stingl, C.; Kaul, A.; Teucher, N.; Hasenfuss, G.; Penninger, J. M.; Mechtler, K. High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all, *J. Proteome Res.* **2009**, *8*, 4743-4752.

(38) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C. iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly", *J. Proteome Res.* **2009**, *8*, 5347-5355.

(39) Savitski, M. M.; Sweetman, G.; Askenazi, M.; Marto, J. A.; Lang, M.; Zinn, N.; Bantscheff, M. Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers, *Anal. Chem.* **2011**, *83*, 8959-8967.

(40) Michalski, A.; Cox, J.; Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS, *J. Proteome Res.* **2011**, *10*, 1785-1793.

(41) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Mol. Cell Proteomics* **2002**, *1*, 376-386.

(42) Ong, S.-E.; Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC), *Nat. Protoc.* **2006**, *6*, 2650-2660.

(43) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis, *Nat. Methods* **2008**, *6*, 359-362.

(44) Sun, B.; Ma, L.; Yan, X., et al. N-Glycoproteome of E14.Tg2a mouse embryonic stem cells, *PLOS ONE.* **2013**, *8*, e55722.

(45) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics, *Nat. Protoc.* **2009**, *4*, 44-57.

(46) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Res.* **2009**, *37*, 1-13.

(47) Zubarev, R. A.; Makarov, A. Orbitrap mass spectrometry, *Anal. Chem.* **2013**, *85*, 5288-5296.

(48) Lange, O.; Damoc, E.; Wieghaus, A.; Makarov, A. Enhanced Fourier transform for Orbitrap mass spectrometry, *Int. J. Mass Spectrom.* **2014**, *369*, 16-22.

(49) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis, *Nat. Methods* **2007**, *4*, 709-712.

(50) Schmid, A.; Kortmann, H.; Dittrich, P. S.; Blank, L. M. Chemical and biological single cell analysis, *Curr. Opin. Biotechnol.* **2010**, *21*, 12-20.

(51) Michalski, A.; Damoc, E.; Hauschild, J. P., et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer, *Mol. Cell Proteomics* **2011**, *10*, 1-11.

(52) Xu, P.; Duong, D. M.; Peng, J. Systematical optimization of reverse-phase chromatography for shotgun proteomics, *J. Proteome Res.* **2009**, *8*, 3944-3950.

(53) Houel, S.; Abernathy, R.; Renganathan, K.; Meyer-Arendt, K.; Ahn, N. G.; Old, W. M. Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies, *J. Proteome Res.* **2010**, *9*, 4152-4160.

(54) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, *Am. Soc. Mass Spectr.* **1994**, *5*, 976-989.

(55) Vaidyanathan, S.; Broadhurst, D. I.; Kell, D. B.; Goodacre, R. Explanatory optimization of protein mass spectrometry via genetic search., *Anal. Chem.* **2003**, *75*, 6679-6686.

Figures

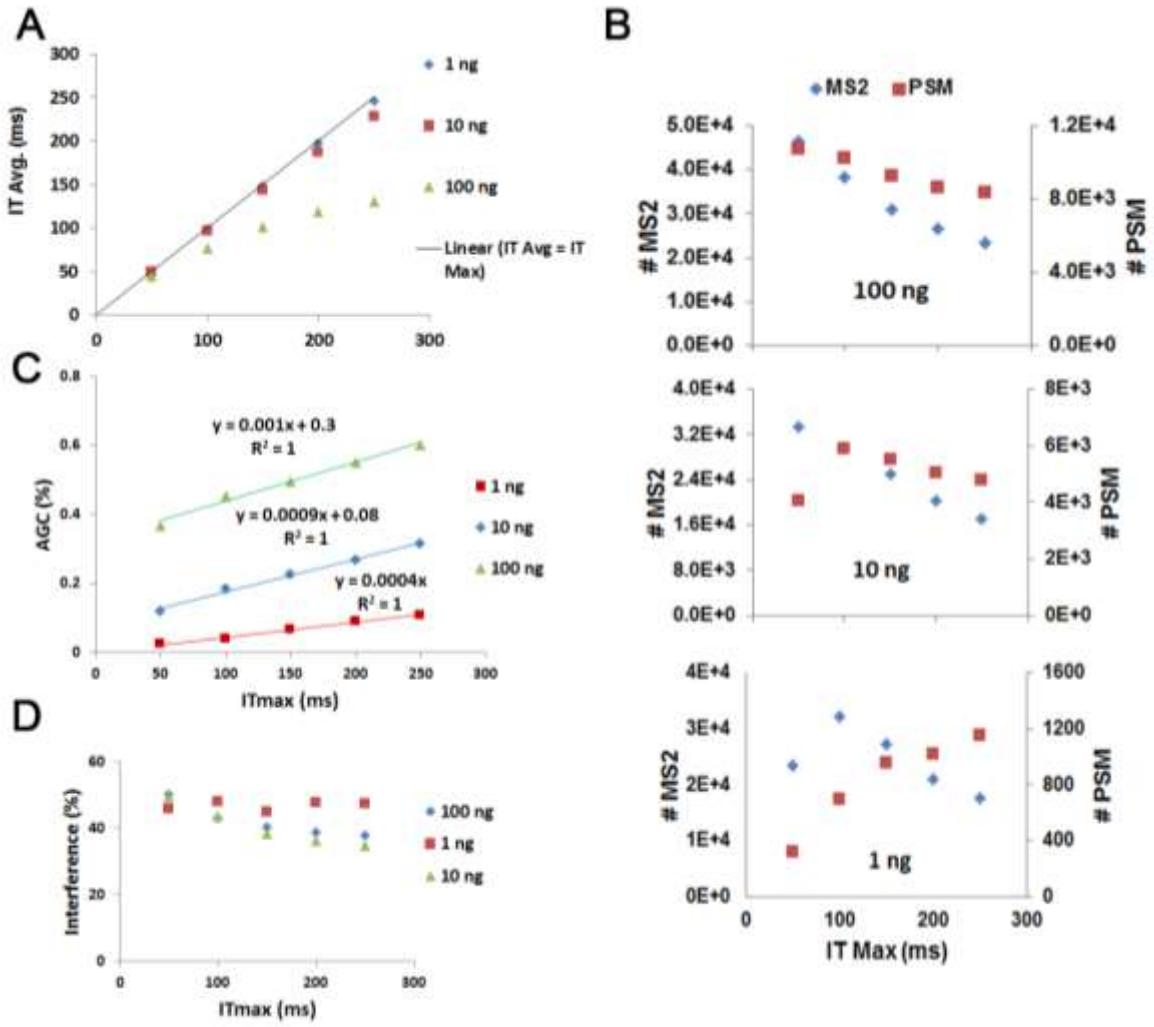

Fig. 1. Evaluation of the maximum MS2 ion injection time (ITmax). A) Relationship between detected average IT of PSMs and the ITmax B) Number of MS2 scans and PSMs as a function of ITmax for 100, 10, and 1-ng samples. C) Relationship of percentage of AGC filling and the ITmax at different total quantity of proteins. D) Changes of MS2 scan interference as a function of ITmax for different total quantity of proteins.

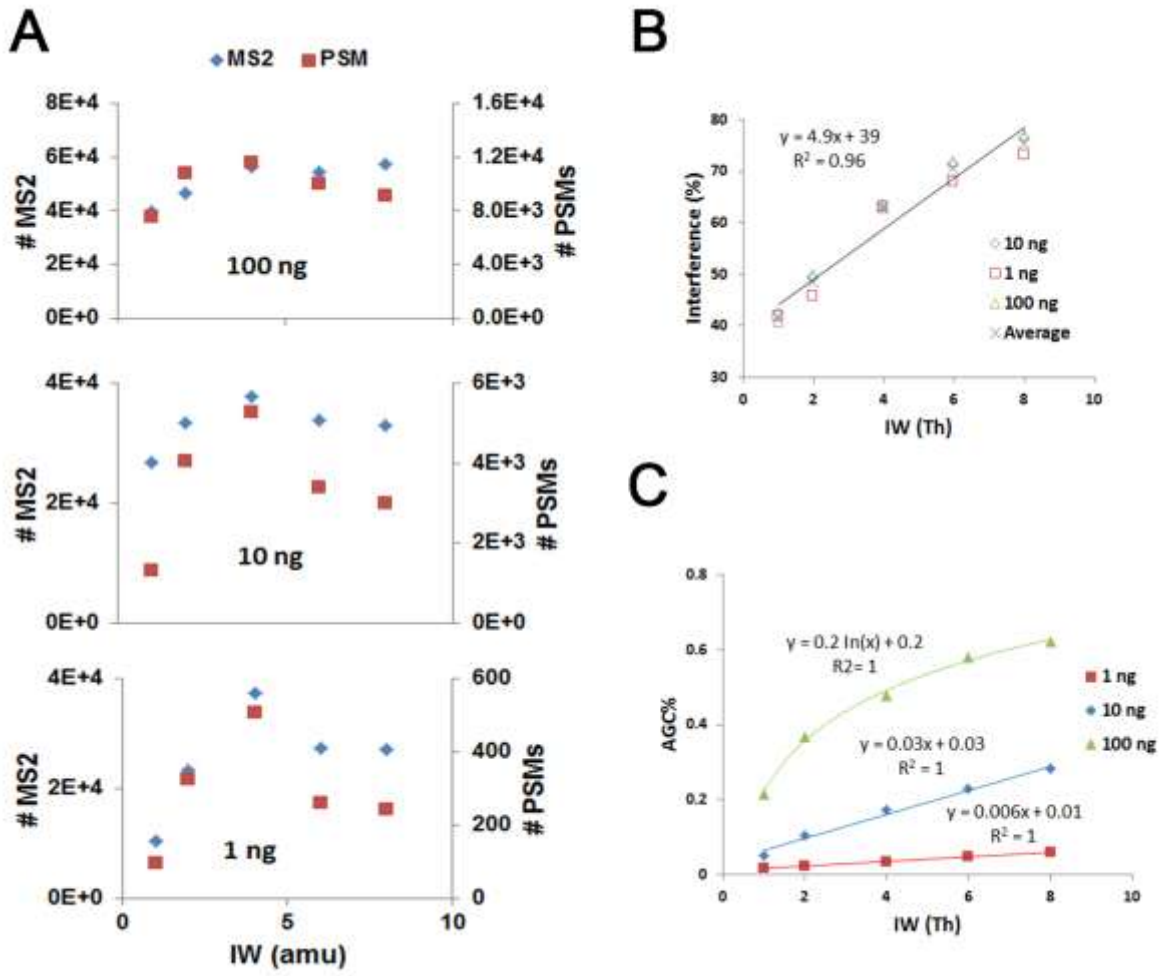

Fig. 2. Evaluation of MS2 mass isolation window (IW). A) Number of MS2 scans and PSMs as a function of IW for 100, 10, and 1-ng samples. B) Changes of MS2 scan interference as a function of IW for different total quantity of proteins. C) Relationship of percentage of AGC filling and different IW values at varied total quantity of proteins.

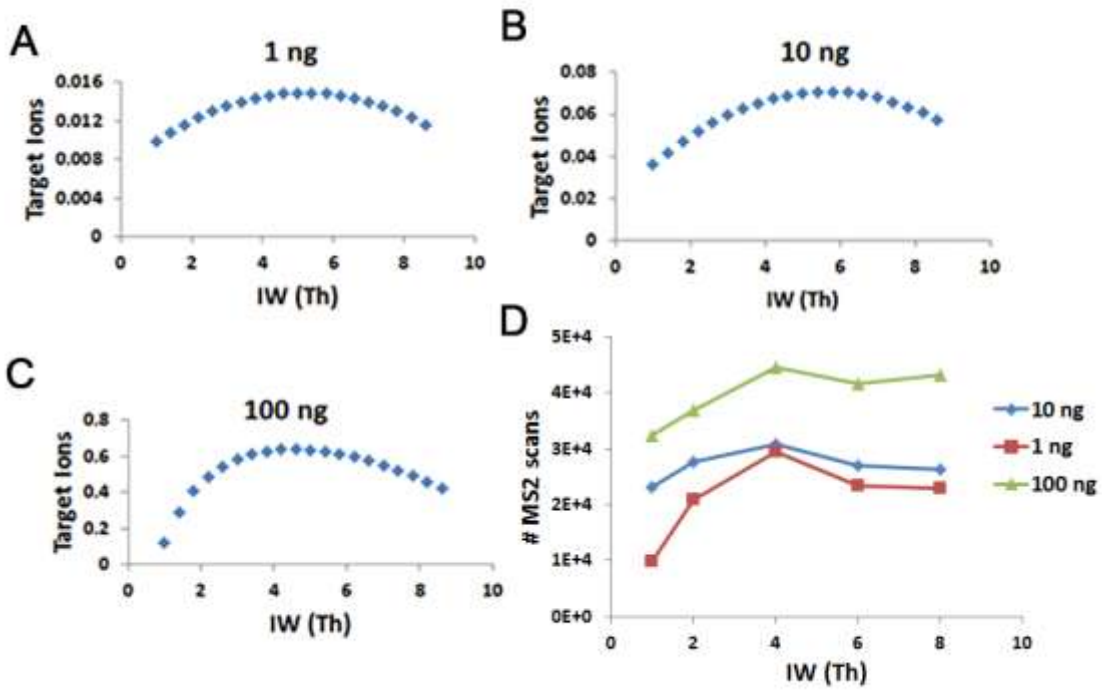

Fig. 3 A-C) Simulation of target ions in C-trap during MS2 scan at total protein quantity of 1 ng, 10 ng, and 100 ng, respectively. D) Relationship of the number of MS2 scans and the IW values at varied total protein quantities.

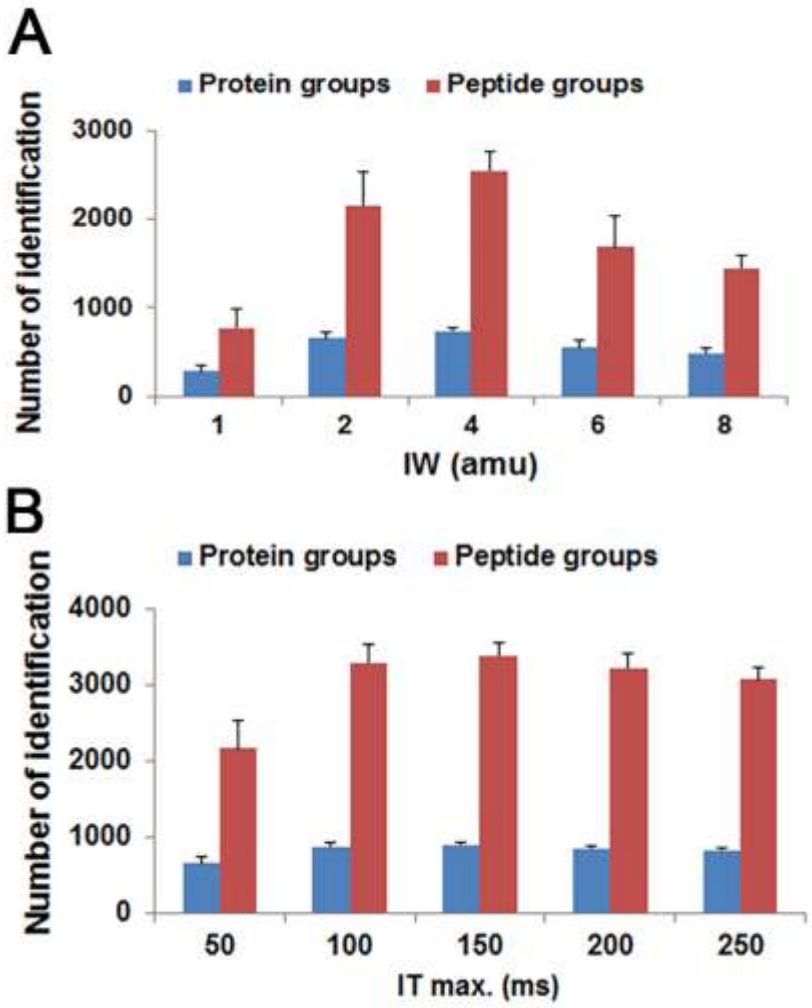

Fig. 4. Protein and peptide group detection at 10-ng total protein level as a function of IW (A) and ITmax (B). The error bar is the standard deviation of the mean.

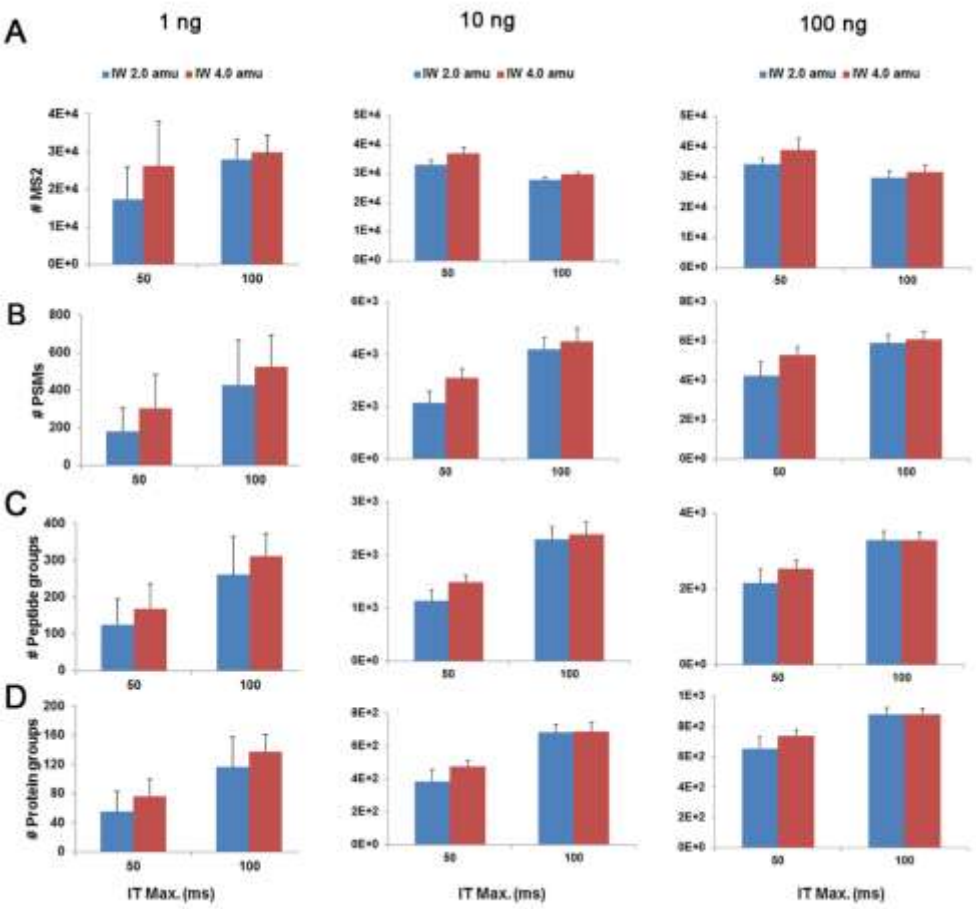

Fig. 5. Synergistic effect of co-varying ITmax and IW from the benchmark parameter setting of ITmax of 50 ms, and IW of 2 Th to ITmax of 100 ms and IW of 4 Th for total protein quantity of 100, 10, and 1 ng. A) Number of MS2 scans. B) Number of PSMs. C) Number of peptide groups. D) Number of protein groups. The error bar is the standard deviation of the mean.

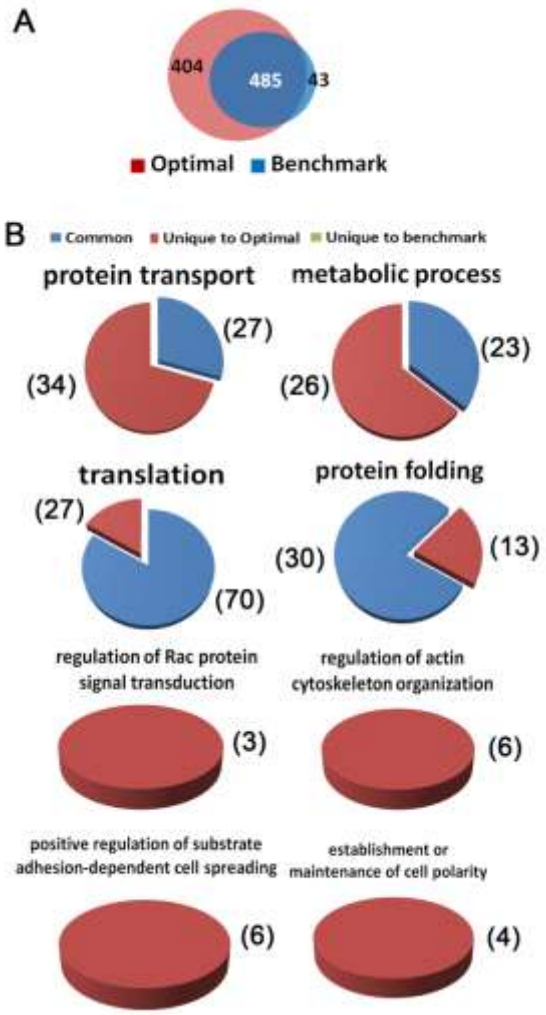

Fig. 6. Biological impact comparison between proteins detected using optimal and benchmark parameter sets on 10 ng of CHO proteins. A) Venn diagram of identified proteins by the two parameter sets. B) Gene ontology enrichment of biological process among proteins detected in common or unique to the two parameter sets. The pie chart is generated on the –log(p), whereas the number of detected proteins are in the bracket.

For TOC only.

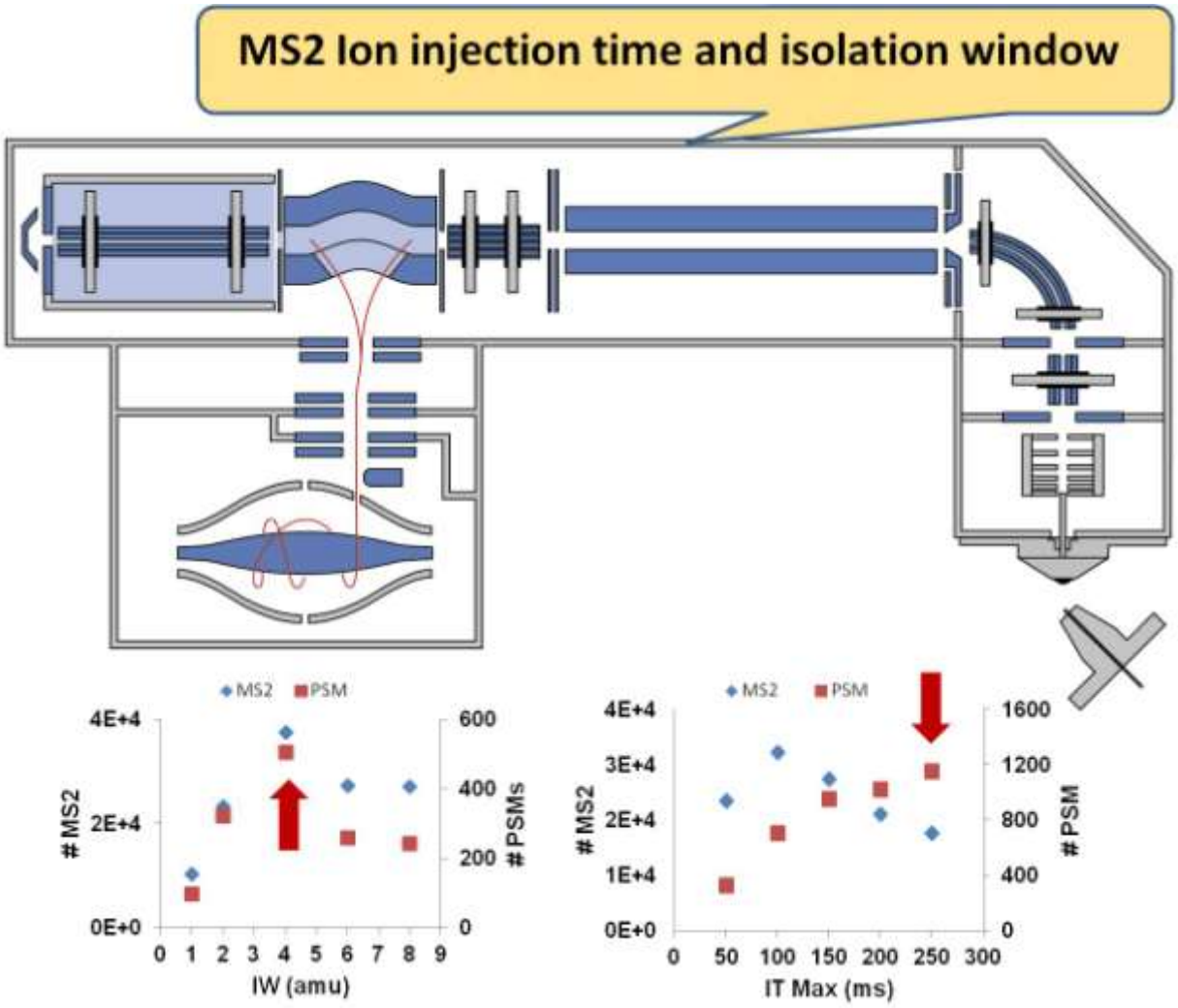

Photo courtesy of Bingyun Sun.