



## RESEARCH ARTICLE

# Improving Proteome Coverage on a LTQ-Orbitrap Using Design of Experiments

Genna L. Andrews,<sup>1</sup> Ralph A. Dean,<sup>2</sup> Adam M. Hawkriged,<sup>1</sup> David C. Muddiman<sup>1</sup>

<sup>1</sup>W. M. Keck FT-ICR Mass Spectrometry Laboratory, Department of Chemistry, North Carolina State University, Raleigh, NC 27695, USA

<sup>2</sup>Center for Integrated Fungal Research, North Carolina State University, Raleigh, NC 27695, USA

### Abstract

Design of experiments (DOE) was used to determine improved settings for a LTQ-Orbitrap XL to maximize proteome coverage of *Saccharomyces cerevisiae*. A total of nine instrument parameters were evaluated with the best values affording an increase of approximately 60% in proteome coverage. Utilizing JMP software, 2 DOE screening design tables were generated and used to specify parameter values for instrument methods. DOE 1, a fractional factorial design, required 32 methods fully resolving the investigation of six instrument parameters involving only half the time necessary for a full factorial design of the same resolution. It was advantageous to complete a full factorial design for the analysis of three additional instrument parameters. Measured with a maximum of 1% false discovery rate, protein groups, unique peptides, and spectral counts gauged instrument performance. Randomized triplicate nanoLC-LTQ-Orbitrap XL MS/MS analysis of the *S. cerevisiae* digest demonstrated that the following five parameters significantly influenced proteome coverage of the sample: (1) maximum ion trap ionization time; (2) monoisotopic precursor selection; (3) number of MS/MS events; (4) capillary temperature; and (5) tube lens voltage. Minimal influence on the proteome coverage was observed for the remaining four parameters (dynamic exclusion duration, resolving power, minimum count threshold to trigger a MS/MS event, and normalized collision energy). The DOE approach represents a time- and cost-effective method for empirically optimizing MS-based proteomics workflows including sample preparation, LC conditions, and multiple instrument platforms.

**Key words:** LTQ-Orbitrap XL MS/MS instrument parameters, Design of experiments, Fractional-factorial design, *Saccharomyces cerevisiae*, Proteomics

## Introduction

Mass spectrometry is at the intersection of several proteomics workflows and the diversity of its user base continues to expand. A consequence of the rising popularity and importance of mass spectrometry (MS) in

biological research has been increasing demands on instrument time and performance. Because the time and cost of MS-based proteomics experiments are significant, the efficient optimization and set-up of instrument parameters remain of paramount importance when pushing the qualitative and quantitative limits of proteome analysis. Although data quality in a proteomics experiment can be defined multiple ways (e.g., proteome depth, biological relevance, quantitative accuracy), experimental outcomes are often dictated to varying degrees by common factors that include sample preparation, sample fractionation and separation, instrument settings, and post-acquisition bioinfor-

**Electronic supplementary material** The online version of this article (doi:10.1007/s13361-011-0075-2) contains supplementary material, which is available to authorized users.

Correspondence to: David C. Muddiman; e-mail: david\_muddiman@ncsu.edu

Received: 24 September 2010  
Revised: 3 January 2011  
Accepted: 3 January 2011  
Published online: 15 February 2011

matic platforms. Many MS proteomics laboratories, including our own, have a preferred method(s) for MS interrogation, but lack systematic studies to justify the overall optimization of the instrument method. Initiatives from the Human Proteome Organization (HUPO) Proteomics Standards Initiative and Clinical Proteomic Technology Assessment for Cancer (CPTAC) have focused on improving the reproducibility of proteomics measurements within and between laboratories by advocating the use of benchmark proteome standards. Recent studies from researchers directly involved with these initiatives have shown interesting results for measuring the performance of liquid chromatography (LC) and MS instrumentation [1], evaluating LC-MS interlaboratory performance [2], and reproducibility in generating protein identifications by LC-MS [3]. However, thus far these initiatives have not focused on MS instrument parameter optimization; rather, they have allowed each laboratory to use a “favorite method” or a standard operating protocol method. Limited investigations of high resolving power MS instrument parameters exist in demonstrating maximum instrument response [4–7] and, furthermore, a large-scale investigation of MS instrument parameters for increased proteome coverage is absent. Herein, we report results from our efforts to systematically and efficiently explore nine MS instrument parameters on a LTQ-Orbitrap XL gauging instrument performance using several proteomic metrics for the analysis of *Saccharomyces cerevisiae*.

An efficient method for investigating the effects of several MS parameters is fractional-factorial design (FracFD), which generates an experimental framework for evaluating several variables (>3) in less time than more conventional approaches such as full-factorial design (FullFD) [8, 9]. FullFD investigates one variable at a time and usually at several different values/levels to accomplish experimental objectives. A recent example of this approach was reported by Raji et al. [10] for optimizing the response of three synthetic peptides on two electrospray ionization (ESI) MS instruments. However, as the number of experimental variables increase, the process becomes more time consuming and costly. For example, an experiment with  $n$  variables/factors at two different levels requires  $2^n$  experiments. As previously proposed by Riter et al. [9] as an effective tool for mass spectrometrists, FracFD provides a more efficient experimental approach or design of experiments (DOE) in which a carefully chosen subset of experiments is performed simultaneously evaluating variables at two levels. These two levels, a maximum and minimum for continuous variables or two categories for categorical variables, are most beneficial to the DOE platform analysis if they are selected based upon experimental or literature reference. It is common to perform one-fourth to one-eighth the number of full-factorial experiments in a FracFD significantly reducing the time of analysis. Recently our group successfully employed the FracFD DOE platform reducing experimental time and cost for the development of an air amplifier to increase MS-ion abundance [11] and for

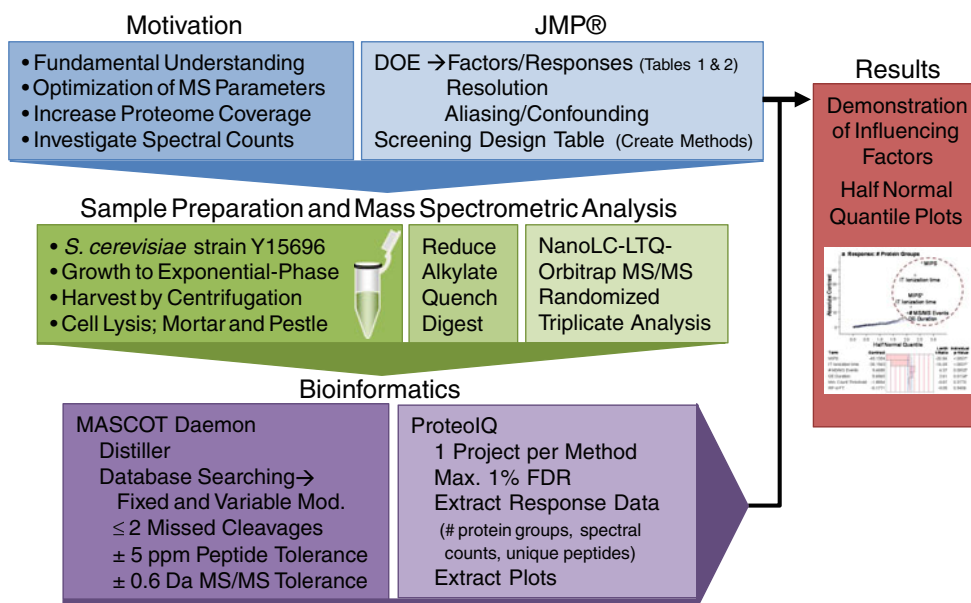
the optimization of sample preparation conditions to improve the MS detection of glycans [12].

In an effort to empirically justify the settings for several MS parameters in a standard shotgun LC-MS/MS experiment, we examined the responses of a total protein digest of *S. cerevisiae* as a function of nine LTQ-Orbitrap MS/MS instrument parameters in two DOE experiments. The proteomic metrics (i.e., responses) used to assess the significance of each parameter were: (1) total number of protein groups (one or more proteins identified with the same peptides and unable to be distinguished as unique); (2) unique peptides; and (3) spectral counts, and these offer quantitative feedback for the analysis of a tryptic digest of *S. cerevisiae*. The mass accuracy of the resultant peptides was not employed as a metric due to the outcome of database searching with different MS tolerances ( $\pm 1$ –10 ppm). It was demonstrated that as the MS tolerance was increased, there was an initial gain in the number of proteins identified followed by very little variation in the number as the tolerance was opened (see Supplementary Figure 1). The first two responses demonstrate the proteome and protein coverage and consider database redundancy. Regarding protein quantification, label-free spectral counting affords a relative measurement of protein concentration by comparing the number of resultant MS/MS spectra from peptides associated with a specific protein [13–17]. The most advantageous instrument method would afford the highest number of total spectral counts with reproducibility maintained over *S. cerevisiae* replicate analyses. *S. cerevisiae* has a sufficiently complex and highly characterized proteome, and was the first organism with a complete annotated database of the complete proteome [18]. It is one of the most extensively analyzed organisms in proteomics research spanning the analysis of MS instrument technologies [19–26] to efforts evaluating global protein expression [24, 27–31]. While this investigation utilizes the entire *S. cerevisiae* proteome, we were only concerned with the relative changes in proteome coverage and the sensitivity of the measurements to detect change and reveal significant variables. The DOE method described herein represents a viable strategy for moving forward with establishing proteomics as a robust, reproducible, and translatable technique for researchers spanning multiple disciplines of biological research and technology development.

## Experimental

### *Saccharomyces cerevisiae* Sample Preparation

*S. cerevisiae* strain Y15696 (BY4742; MaT $\alpha$ ; his3D1; leu2D0; lys2D0; ura3D0; YIR034c::kanMX4), an auxotroph for lysine due to the *lys1* gene deletion, was acquired from EuroScarf (Frankfurt, Germany). The experimental design analyzing the yeast sample is illustrated in Figure 1 and described here in more detail. The yeast was grown for 24 h in yeast peptone dextrose broth at 30 °C to exponential-phase. The culture was



**Figure 1.** The experimental workflow investigating LTQ-Orbitrap MS/MS instrument parameters anticipating the increase of proteome coverage. JMP software afforded the generation of experiments for nanoLC-LTQ-Orbitrap MS/MS investigation of the tryptic digest of *S. cerevisiae*. Following data processing with the MASCOT and ProteolQ platform, the results were provided for DOE analysis with JMP demonstrating significant variables

harvested by centrifugation at 5000 rpm for 10 min at 4 °C. The cell pellet was washed in 50 mL of 50 mM Tris-HCl (Sigma-Aldrich, St. Louis, MO, USA) and again subjected to centrifugation as described above. The yeast cell pellet was flash-frozen prior to lysis with mortar and pestle. The resulting powder was reconstituted in 50 mM Tris-HCl. Following centrifugation, cellular debris was removed as the supernatant was collected. A modified Bradford Assay (Coomassie Plus Assay) and a bicinchoninic acid assay, both products of Pierce (Thermo Scientific, Rockford, IL, USA), were used to approximate the total protein concentration. An in-solution tryptic digestion was completed on ~1 mg of protein and is briefly described here. Urea (Sigma-Aldrich) was added to the yeast protein solution such that the final concentration was 8 M. The denatured proteins were reduced by adding a 100 mM dithiothreitol (DTT) (BioRad, Hercules, CA, USA) solution to a final concentration of 5 mM followed by a 30 min incubation at 56 °C. The solution was then allowed to cool to room temperature followed with alkylation by adding a 200 mM iodoacetamide (Sigma-Aldrich) solution to a final concentration of 20 mM and incubated for 30 min at room temperature in the dark. The reaction was quenched with 100 mM DTT for 30 min and then diluted with 50 mM Tris-HCl, such that the urea concentration was 2 M. Proteins were digestion with trypsin (Sigma-Aldrich) at a 1:50 enzyme:substrate ratio and allowed to proceed overnight at 37 °C. Formic acid (Sigma-Aldrich) was added to the peptide solution present as 1% of the volume. The sample was aliquoted into multiple identical fractions (by volume) and dried under reduced pressure prior to storage at -20 °C.

### NanoLC-LTQ-Orbitrap MS/MS Analysis

A nanoLC-1D (Eksigent Technologies, Dublin, CA, USA) was coupled to a LTQ-Orbitrap XL (Thermo Scientific, San Jose, CA, USA) via a continuous, vented column configuration described previously [32]. Both the trap and analytical columns were self-packed with Magic C<sub>18</sub>AQ stationary phase (5 µm particle, 200 Å pore) (Michrom Bioresources, Auburn, CA, USA) utilizing a pressurized cell. Mobile phase A and B were composed of water/ acetonitrile/formic acid (98/2/0.2% and 2/98/0.2%, respectively). The solvents (Burdick and Jackson, Muskegon, MI, USA) were HPLC-grade and the formic acid (Sigma-Aldrich, St Louis, MO, USA) was MS-grade. Two µL of yeast digest (100 ng/µL in 50 mM Tris-HCl pH 8.0) was loaded onto the trap column followed by washing with approximately 10 column washes with 2% B from Channel 1 flowing at 1.5 µL/min. The following gradient was applied at a flow-rate of 350 nL/min: 2% B (0–5 min), 2%–10% B (5–7 min), 10%–40% B (7–67 min), 40%–90% B (67–68 min), 90% B (68–78 min), 90%–2% B (78–80 min), 2% B (80–85 min). A new reconstituted sample was loaded every eight sample injections for analysis. Details of the LTQ-Orbitrap XL instrument settings and pertinent comparisons will be provided in Results in Discussion (vide infra) and Supplementary Tables 1 and 2.

### Data Analysis

Shotgun proteomics data generated during this study was searched against a concatenated target-reverse *S. cerevisiae*

database (Uniprot ver. 4932) created with *Bos taurus* trypsin sequence, and *Homo sapiens* keratin and keratin related proteins using Mascot Daemon (ver. 2.2.2, Matrix Science, Boston, MA, USA) to batch process files, Mascot Distiller (ver. 2.2.1.0, Matrix Science, Boston, MA, USA) to generate peak lists, and then Mascot (ver. 2.3.01, Matrix Science, Boston, MA, USA) to perform the searches. Carbamidomethyl (C) was set as a fixed modification and deamidation (N and Q) and oxidation (M) were set as variable modifications. Additional search settings included a maximum of 2 missed cleavages, peptide tolerance of  $\pm 5$  ppm, and MS/MS tolerance of  $\pm 0.6$  Da. Protein grouping, statistical filtering, and quantification (spectral counts) of the Mascot DAT files were accomplished using ProteoIQ (ver. 1.5.05, BioInquire, Athens, GA, USA) that utilizes a combination of Peptide/Protein Prophet [33, 34] and ProValT [35]. One ProteoIQ project was created for each DOE FracFD or FullFD method (32 projects for DOE 1 and 11 projects for DOE 2) and the data was filtered based on a maximum 1% protein false discovery rate (FDR).

The number of protein groups, total spectral counts, and unique peptides for each replicate as a function of the 9 LTQ-Orbitrap XL instrument parameters are shown in Supplementary Tables 3 and 4. These measurements (i.e., responses) were used to generate the outcome for the DOE screening data analysis in JMP 8.0.2 (SAS Institute, Inc., Cary, NC, USA) as illustrated in Figure 1. Half normal quantile probability plots, the complementary bar graphs, and statistical measurements afforded presentation of influencing factors.

## Results and Discussion

### *Previous S. cerevisiae* LTQ-Orbitrap MS/MS Analysis

*S. cerevisiae* has been used as a MS performance standard evaluating the performance of several laboratories with equivalent instrument platforms and bioinformatics [2]. Paulovich et al. [2] provides a reference *S. cerevisiae* dataset to the MS community for opportunity to evaluate the performance of LTQ and LTQ-Orbitrap MS/MS instruments with the *S. cerevisiae* NIST performance standard. The laboratories included in this study were requested to employ a “favorite” instrument method as well as a standard operating protocol with both applying a 2 h gradient for instrument performance analysis. While the instrument methods were not optimized, Paulovich et al. [2] describes that this investigation allows for laboratories to compare instrument performance and expand upon the development of optimized methods for analysis. Resultant data was processed with MyriMatch [36] as the bioinformatic platform and the absolute number of proteins and peptides identified were used as a measure of performance.

Although we requested but were unable to acquire the NIST performance standard for use in our LTQ-Orbitrap

MS/MS instrument parameter investigation, their RAW data files were attainable through ProteomeCommons.org for data sharing. As a rough comparison of the analysis to our own, a dataset was randomly selected in which 120 ng of the yeast sample was loaded on the column (Orbi2\_study8\_W080923\_yeast\_120\_ft8\_pc in triplicate analysis). Processing the RAW data with the more commonly employed Mascot bioinformatic platform combined with ProteoIQ, as described *vide supra*, resulted in 1088 protein groups, 7707 unique peptides identified, and 19,790 spectral counts. While this output exceeded our best results by approximately 2-fold, several differences are apparent between the studies, and instrument method deviations are detailed in Supplementary Table 2.

One of the more significant differences in experimental conditions in comparison to Paulovich et al. [2] is analytical separation. Paulovich et al. [2] employs a 2 h gradient whereas our methods employ a 1 h gradient. The reason for a reduced gradient length in our study was attributed to the nature and size of the DOE studies and the fact that we were primarily interested in relative changes in proteome coverage, not in setting new records in numbers of proteins identified. This difference in gradient length is evidenced to influence the peak capacity and consequently analyte separation and detection [37, 38]. An extended gradient decreases the probability of species overlapping, and therefore reduces complexity as the analyte assumes MS detection supporting an increase in protein identifications. Second, a direct comparison of methods would require access to the NIST yeast standard.

### DOE 1

Requiring only half the number of experiments and consequentially half the time (32 in triplicate versus 64 in triplicate), DOE 1 afforded the analysis of six factors (see Table 1) demonstrating effective variables by a FracFD platform. These six factors were of great interest to our group attributable to the curiosity in MS data acquisition and empirically demonstrating factors contributing to the data quality and quantity. Half normal quantile probability plots and the corresponding bar graphs were generated by JMP affording a demonstration of influencing factors as a function of 3 responses (see Figure 2). It is clearly evidenced that the monoisotopic precursor selection (MIPS) function and the ion trap (IT) maximum ionization time, also known as maximum injection time, are significant variables; the large absolute value of the contrast and the Lenth *t*-ratio, and the almost zero p-Value exhibit the influence of these parameters. The negative contrast values indicate that the first item specified for the categorical factor (on for MIPS) and the minimum value for the continuous factor (80 ms for IT maximum ionization time) afford the greatest influence in proteome coverage and spectral counts. MIPS affords the selection of only the monoisotopic peak, while excluding all other peaks in the same isotopic distribution, and as



**Table 1.** Six factors were included in DOE 1 for the analysis of improved proteome coverage by nanoLC-LTQ-Orbitrap MS/MS. The settings used for previous MS analysis are included followed by the minimum (Min.) and maximum (Max.) factor values included in the DOE FracFD, and rationale behind the selection of the instrument parameters

Factor	Previous Settings	Min.	Max.	Motivation
Ion Trap Ionization Time (ms)	400	80	400	More proteins identified with 80 ms and $8 \times 10^3$ AGC <sub>Target</sub> than 100 ms and $1 \times 10^4$ AGC <sub>Target</sub> (unpublished data) [37].
Dynamic Exclusion Duration (s)	30-180	30	180	This duration effects spectral counting [38]. What length produces the best data and least variation?
# Tandem MS Events	3-8	3	8	How many events can occur at the specified resolving power? Does resolving power or # MS/MS events control the sequence of events?
Min. Count Threshold	500	500	2500	At a higher value, will the ion be selected for MS/MS when it is eluding closer to the apex of the chromatographic peak? At a lower value will less abundant species be selected?
Resolving Power in FT	60,000	30,000	60,000	Does a lower resolving power (i.e., faster duty cycle) provide equivalent protein identifications?
Monoisotopic Precursor Selection	Off	On	Off	Is this an important factor; will this parameter increase or reduce the number of proteins identified?

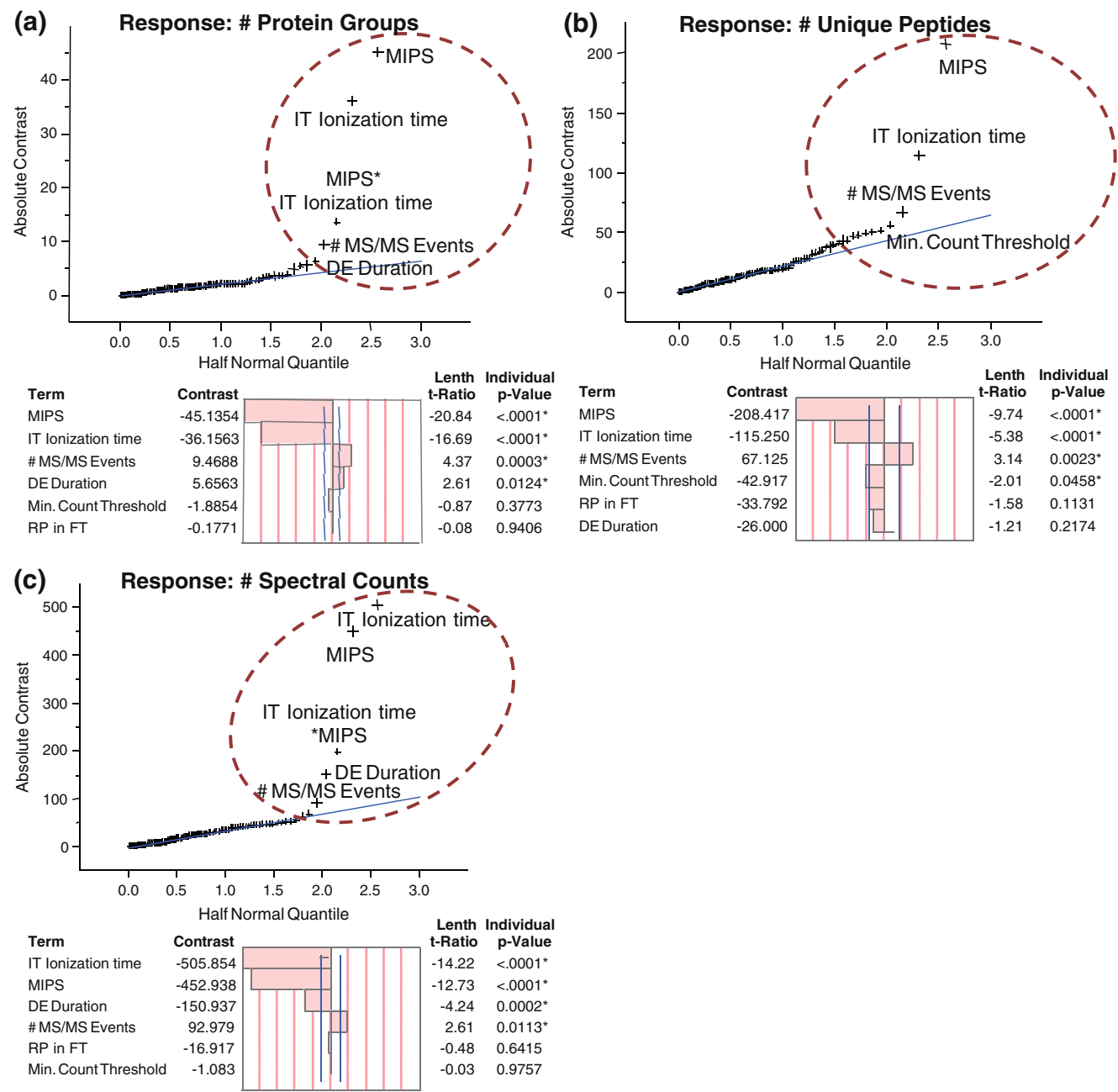
evidenced, significantly effects proteome coverage. As expected, IT maximum ionization time greatly influences the responses as the interplay between shorter maximum ionization times and the automatic gain control target (AGC<sub>Target</sub>) allows for more MS/MS events to occur between precursor ion scans, and accordingly a greater number of available peptide sequences subjected for identification. Longer maximum IT ionization times may time out when little to no signal is present in the analysis not reaching the AGC<sub>Target</sub> and wasting time between the precursor ion scans. Consequently with the shorter IT maximum ionization time favored, the number of MS/MS events, while considered not significant for number of proteins identified, was most effective when set to 8 events. As a result of a personal communication, the AGC<sub>Target</sub> for both the ion trap and Orbitrap were not evaluated in this DOE investigation, but maintained at  $8 \times 10^3$  and  $1 \times 10^6$ , respectively [39].

It is evidenced that dynamic exclusion (DE) duration influences the number of spectral counts [40], and can affect the proteome coverage of investigation. More abundant peptides eluting off the column over a broader chromatographic peak will inherently have more opportunities for MS interrogation depending on the exclusion time period. DE duration, as displayed in Figure 2a and b, appears to be a significant variable; however, both the minimum and maximum values are favored depending on the response. The maximum factor value for DE duration, 180 s, generates an increased number of protein groups identified. The MS selects ions for interrogation by abundance, excluding ions for 180 s gives rise to MS interrogation of lesser abundant ions over a 3 min period versus that of a shorter time period and, consequently, a greater variety of species have the opportunity for MS/MS analysis. Spectral counts as a response (see Figure 2c), the minimum factor value, 30 s, affords a greater output due to shorter exclusion periods of highly abundant species and not as demanding of the interrogation of lower abundant species. Normalization efforts will facilitate direct comparisons in quantification,

but it is important to acknowledge these results when investigating a sample with a large dynamic range.

Two factors with minimal if any contributions towards the responses, minimum count threshold and resolving power (RP) of the precursor survey scan, appear to fall closely to the limit of significance in response towards the number of unique peptides identified (Figure 2b). In the instance of minimum count threshold, the absolute value of the Lenth *t*-ratio is just outside the commonly significant value of two. Also, the individual *p*-value falls close to the 0.05 significance limit. This factor, minimum count threshold to trigger a MS/MS event, establishes the minimum amount of signal required for an ion to be selected for a MS/MS event. In principle, a larger value would instigate MS/MS interrogation of more abundant ions potentially generating higher quality mass spectra. When deemed a significant variable (see Figure 2b), a value of 500 counts is most effective for the minimum count threshold, still the factor does not greatly influence the proteome coverage or spectral counts. Yates and coworkers extensively evaluated the minimum count threshold and demonstrated similar results generating no significant difference in the number of protein identifications at comparable threshold values [41]. The last factor, RP, did not significantly increase proteome coverage; however, the bar graphs suggest that a resolving power of 30,000<sub>fwhm</sub> at *m/z* 400 may contribute to increased proteome coverage as opposed to 60,000<sub>fwhm</sub>. The instrument method from DOE 1 generating the best response data employed a RP of 30,000<sub>fwhm</sub> at *m/z* 400 (see Table 2a). The maximum RP, and as a consequence the potential for increased mass accuracy, does not necessarily contribute to increased protein identifications and Kim et al., who systematic evaluated resolving power in shotgun proteomic experiments, also demonstrated limited gain in protein identifications when comparing maximum RP [42].

The resolution of FracFD, or degree of confounding, is specified prior to creation of the screening design table influencing the number of experiments required for analysis and the possible number of aliasing effects. For our purposes



**Figure 2.** Half normal quantile probability plots and corresponding bar graphs for each response for the determination of significant variables for DOE 1 FracFD. The half normal quantile curves (blue line) in each plot represent the normal distribution, and data points greatly deviating from the curves and with the appropriate statistical measurements indicate a significant variable. Those factors in the red hashed circle are significant variables. The contrast, Lenth t-ratio, and individual p-Value are different statistical measurements concluding the significance of each factor. The Lenth t-ratio is calculated by dividing the contrast value by the Lenth pseudo-standard error which is an estimate of error based on the inactive effects and generated with each half normal plot [43]. The vertical blue lines on the bar graphs represent the default cutoff p-Value (0.05) for indication of the degree of significance of each parameter. The plots and bar graphs are organized by response as follows: **(a)** # protein groups; **(b)** # unique peptides; and **(c)** # spectral counts

we selected a resolution of five, which afforded no confounding effects, equivalent to the resolution of a FullFD, though requiring only half the number of experiments and, hence, half the time. This type of resolution

affords the realization of significance of each variable on the data whether or not the variables interact with each other [9]. As displayed in Figure 2, confounding factors are specified in the analysis and recognized as significant variables.

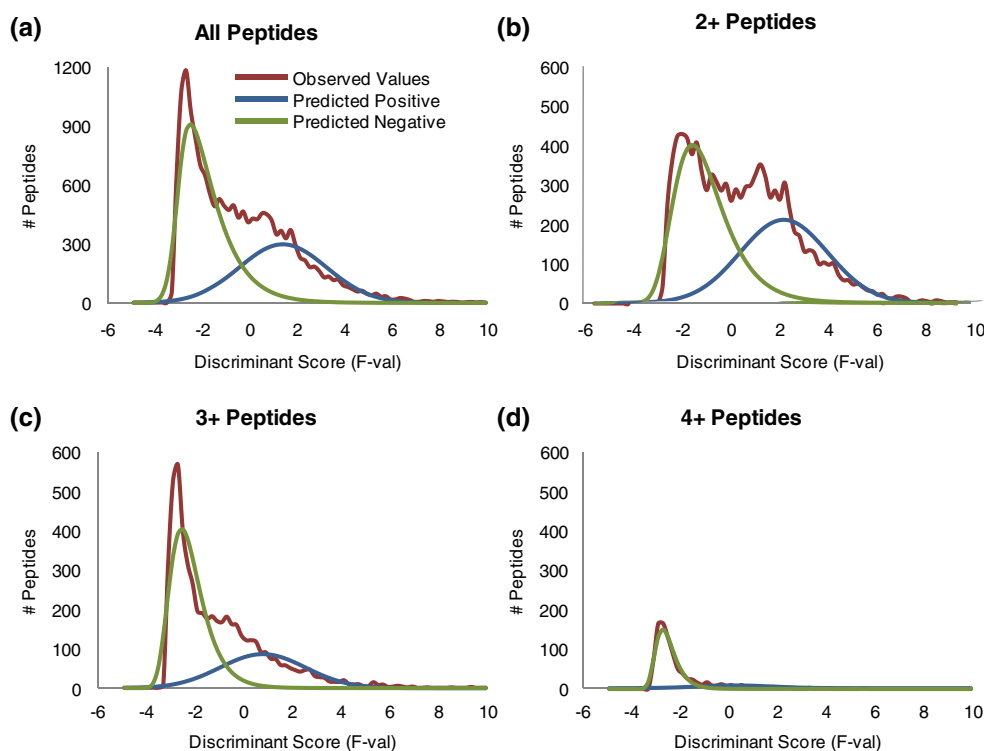
**Table 2.** Factor and response values for DOE 1 method generating the most protein identifications with ProteoIQ. **(a)** As demonstrated through the half normal plots and bar graphs for DOE 1 (Figure 2), this instrument method contains the appropriate settings for the significant variables and was set to the maximum factor value for the insignificant variables. **(b)** The response data for each replicate that was entered into the JMP screening design table

<b>a Factors</b>				
Instrument Parameter		Setting		
IT Ionization Time (ms)		80		
DE Duration (s)		180		
# MS/MS Events		8		
Min. Count Threshold		500		
RP in FT		30,000		
MIPS		On		
<b>b Responses</b>				
Result	Rep 1	Rep 2	Rep 3	Total
# Protein Groups	364	414	396	490
# Spectral Counts	3259	3934	3726	10919
# Unique Peptides	1165	1457	1352	3974

However, due to the resolution specified for DOE 1, confounding factors can be confirmed as significant or insignificant based on the results of the individual factors. This confounding provides a glimpse of possible significance of interacting factors of DOE 1 analysis had a resolution of five not been performed. Most half normal plots in Figure 2 contain IT maximum ionization time

confounded with MIPS, but it is clear that each individual factor and not just the confounding nature cause the variables to be significant towards the response.

While six instrument parameters were included in DOE 1, the setting for a seventh instrument parameter was also suggested from the investigation. ProteoIQ affords output of peptide discriminant value distributions gauging probability [33] as illustrated in Figure 3. Figure 3 represents discriminant value plots for the DOE 1 instrument method producing the most protein identifications (see Tables 2a and b for factor and response data), and Supplementary Figure 2 represents discriminant value plots for the DOE 1 instrument method producing the least protein identifications (see Tables 3a and b for factor and response data). Both figures conclude that peptides in the 2+ charge-state yield more positive peptide identifications versus 3+ or 4+ charge-state peptides. As mentioned in Supplementary Table 1, 1+ and unassigned charge-states are rejected from MS/MS analysis. Attributable to discriminant value distribution plots, peptides with charge-states >3+ appear to be consuming available MS/MS interrogation without giving rise to peptide identification, and accordingly charge-state 4+ may also be rejected. Although Figure 3 and Supplementary Figure 2 suggest that 2+ charge-state peptides are predominately observed and identified, further investigations are



**Figure 3.** ProteoIQ output of peptide discriminant value distributions for the instrument method that generated the most protein identifications. The number of peptides is plotted versus the discriminant value or the measurement of peptide assignment accuracy. The observed values, predicted positive, and predicted negative data are included in each plot. **(a)** All peptides, **(b)** 2+ charge-state peptides, **(c)** 3+ charge-state peptides, and **(d)** 4+ charge-state peptides are included for comparison as a function of confident peptide identification. It is desirable to have the greatest area overlap of the observed and predicted positive values

**Table 3.** Factor and response values for DOE 1 method generating the least protein identifications with ProteoIQ. **(a)** Comparing to the instrument method that generated the most protein identifications, the opposite parameter settings yielded the least response. Of the total proteins identified in DOE 1, less than half were identified using this set of parameters. **(b)** The response output for the replicate data which was used for DOE 1 analysis

a Factors				
Instrument Parameter		Setting		
IT Ionization Time (ms)		400		
DE Duration (s)		30		
# MS/MS Events		3		
Min. Count Threshold		2,500		
RP in FT		60,000		
MIPS		Off		
b Responses				
Result	Rep 1	Rep 2	Rep 3	Total
# Protein Groups	170	190	180	238
# Spectral Counts	1547	1958	1720	5225
# Unique Peptides	458	564	514	1536

necessary to evidence if only 2+ charge-state peptides should be selected opposed to 2+ and 3+ peptides. Overall, the instrument method resulting in the most protein identifications for DOE 1 (Table 2a) resulted in 490 confident (maximum 1% FDR) protein groups and 3187 unique peptides from triplicate analysis (see Table 2b), while the method resulting in the least protein identifications for DOE 1 (Table 3a) resulted in 238 protein groups and 1694 unique peptides (see Table 3b). Evaluating LTQ-Orbitrap MS/MS instrument parameters afforded the improvement in instrument response by roughly 2-fold.

DOE 2

An additional DOE investigation (DOE 2) was initiated following DOE 1 data processing in order to further the investigation of instrument parameters. DOE 2 evaluated normalized collision energy (NCE), tube lens voltage, and capillary temperature (see Table 4) for increased proteome coverage using the *S. cerevisiae* tryptic digest. Vide infra, our curiosity in the interplay of these factors with the resultant number of proteins identified, lead to the selection of parameters. NCE provides a level of energy for peptide fragmentation in the LTQ, and it is crucial to select a value in which the species is sufficiently fragmented; too little

NCE will result in no fragmentation, while too much NCE may over-fragment the peptide limiting sequence information and complicating the MS/MS spectra through generation of  $w_n$  and  $d_n$  side chain fragment ions and internal fragment ions. Within the Xcalibur software, the default setting for NCE is 35, however Paulovich et al. [2] employed a NCE of 28 and limited to no empirical evidence in this selection contribute to our curiosity in altering the NCE. The tube lens voltage directs ions into the ion guide which is offset from the orifice of the detector. The redirection of ions prevents neutral species from accumulating in the MS. This voltage may be a function of the molecular weight and charge as our group has assessed a tube lens value of 120 V for N-linked glycans (unpublished data) and variation of tube lens voltage for intact proteins influenced by the molecular mass (unpublished data). The capillary temperature influences the desolvation and other associated properties of the electrospray droplets as they travel from the ESI emitter towards the orifice of the MS and form gas-phase ions.

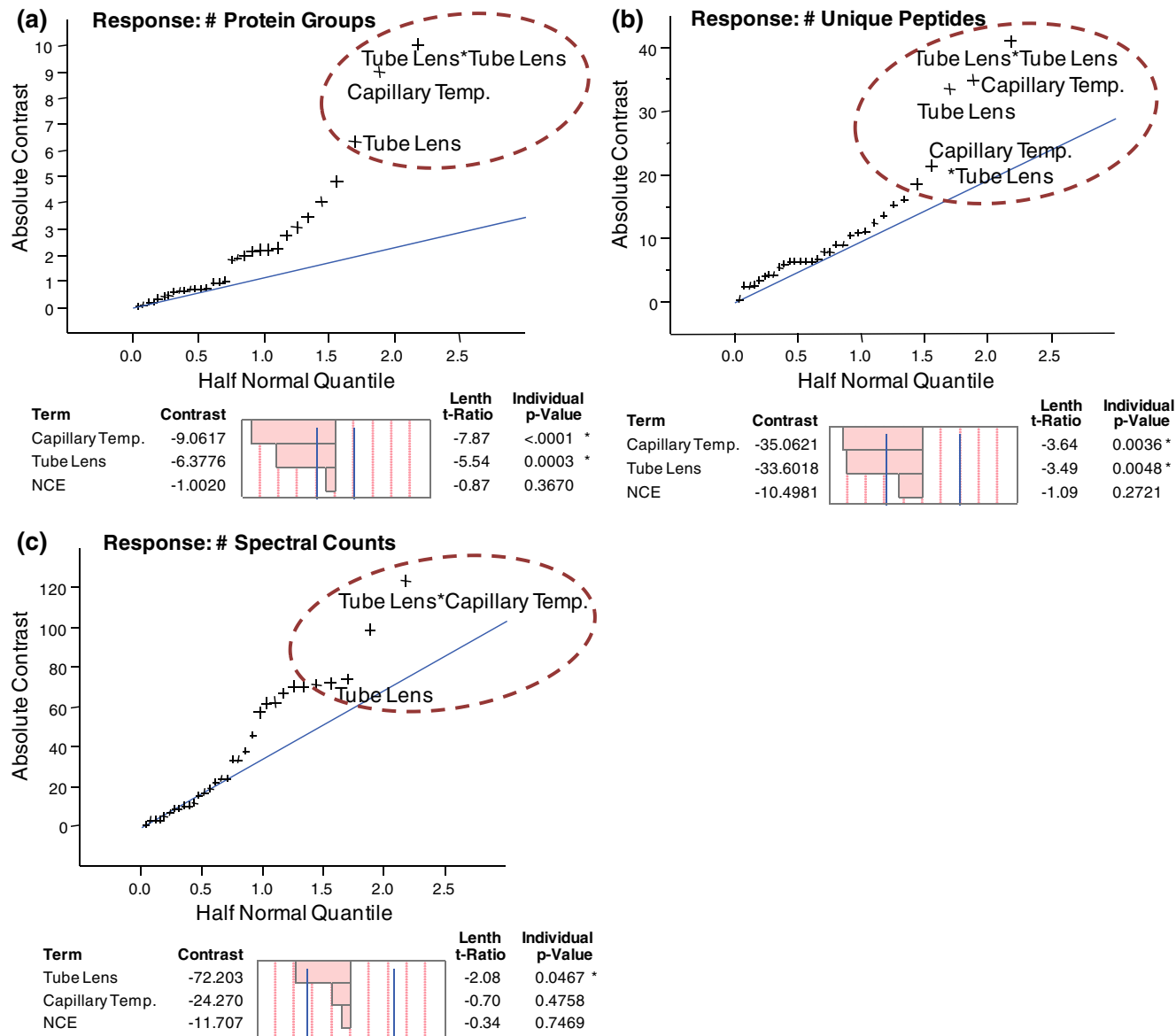
The equivalent motivation and experimental workflow was followed as illustrated in Figure 1, and the parameters producing the greatest number of protein identifications from DOE 1 were used (see Table 2 and Supplementary Table 1). Whereas a resolution of five for DOE 1 was accomplished in half the number of experiments as a FullFD, DOE 2 required a FullFD to accomplish the same resolution. To make for a more complete experimental design, three additional experiments were included in DOE 2 reflecting the instrument parameter settings employed for the best responses from DOE 1, as well as midpoints for tube lens (120 V) and capillary temperature (187 °C) such that time permitted (see Table 4).

Eleven experiments in triplicate were completed resulting in half normal probability plots and bar graphs produced by JMP. Figure 4 exhibits that the tube lens voltage is a significant variable for all responses (Figure 4a, b, and c) and capillary temperature is a significant variable for two responses (Figure 4a and b). The bar graphs reveal that the minimum value for tube lens voltage (100 V) and capillary temperature (150 °C) are preferred for increased response. Tube lens voltage contributes to the identity of the species allowed to be directed towards the MS detector and this analysis reveals a lower voltage than previously employed is favored. Capillary temperature alters the droplet desolvation

**Table 4.** Three factors were included in DOE 2 for a more complete nanoLC-LTQ-Orbitrap MS/MS parameter analysis. Similar to Table 1, the previous instrument setting, minimum and maximum factor values, and motivation driving the investigation are included. Also, there is indication of a midpoint value if one was used

Factor	Previous Settings	Min.	Max.	Motivation
Normalized Collision Energy (%)	35	28	35	The instrument default value was set to 35%. The SOP for the interlaboratory yeast study set NCE to 28%. It is possible that the species were too fragmented resulting in reduced identifications.
Capillary Temperature (°C)	150-250	150	225	Does this parameter significantly alter the number of confident protein identifications? Midpoint=175 °C
Tube Lens (V)	80-150	100	140	What is the “sweet spot” value for the LTQ Orbitrap with this yeast sample? Midpoint=120 V





**Figure 4.** Half normal quantile probability plots and corresponding bar graphs for each response demonstrating the significance of each factor in DOE 2 FullFD organized equivalent to Figure 2. Those factors that deviate from the half normal quantile curve (blue line) and within the red hashed circle are significant variables

rate, and the minimum temperature favored suggests a rate limiting thermal degradation and charge stripping which would be exist if the temperature was too high. As with DOE 1, confounding factors are represented as significant factors; however, attributable to the resolution, any aliasing can be evaluated based on the individual factors. The instrument method investigated in DOE 2 providing the most protein identifications is presented in Table 5a while the response data from the triplicate analysis is presented in Table 5b. As demonstrated in the systematic characterization of LTQ-Orbitrap MS/MS instrument parameters in DOE 1, DOE 2 resulted in increased responses. A total of 570 protein groups were confidently identified in DOE 2, which is an increase of 80 protein groups over the best results from DOE 1 affording roughly 20% more proteome coverage.

**Table 5.** Factor and response values for DOE 2 instrument method generating the greatest number of protein identifications. (a) Based on the half normal plots and bar graphs for DOE 2 (Figure 4), the capillary temperature and tube lens in this method reflect the largest protein identification response as described by the triplicate analysis (b)

a Factors				
Instrument Parameter	Setting			
Normalized Collision Energy (%)	35			
Tube Lens (V)	120			
Capillary Temperature (°C)	175			
b Responses				
Result	Rep 1	Rep 2	Rep 3	Total
# Protein Groups	436	424	436	570
# Spectral Counts	3430	3180	3313	9923
# Unique Peptides	1358	1242	1332	3932

## Conclusions

The DOE platform afforded a systematic approach investigating large experimental space for the analysis of 9 LTQ-Orbitrap MS/MS instrument parameters. Variables and their settings of significant influence to most instrument responses in DOE 1 included 80 ms IT maximum ionization, MIPS on, and 8 MS/MS events. In DOE 2, a capillary temperature of 175 °C and a tube lens value of 120 V afforded the best instrument response. Overall improvement to the instrument method afforded 570 protein groups with the best DOE 2 method employed versus 238 protein groups with the worst DOE 1 method. The proteome coverage increased approximately 60%, performing approximately 75% of the total experiments required for a FullFD.

Here it is evidenced that LTQ-Orbitrap MS/MS parameters influence the resultant data (see Supplementary Table 1 for full detailed parameter settings). Significant improvement was realized from this evaluation, and optimization for individual instruments and conditions may be required. The objective of these initial DOE studies was to demonstrate the significance of each variable for improved proteome analysis. Whereas the minimum or maximum value was determined as an improvement, depending on the condition or type of high resolution MS, each MS instrument is unique, and this investigation will provide a proven foundation with which to begin optimization for increased proteome coverage. Modifications to the nanoLC and bioinformatic platforms also merit investigation and may contribute to increased proteome coverage.

## Supporting Information Available

The data associated with this manuscript may be downloaded from the ProteomeCommons.org Tranche network using the following hash:

44kPwxvjy9zCSFSirSGCbFUQGMRyjtvmv0a547DeJ5g+vNN2OK4IUPGloA/LhxsLLOfmPuVkiMROcijsdRCA-h7AW8hcAAAAAAAFpw==

The hash may be used to prove exactly what files were published as part of this manuscript's dataset, and the hash may also be used to check that the data has not changed since publication.

## Acknowledgments

The authors acknowledge the financial support of the National Institutes of Health (grant 5T32GM00-8776-08), which supports G.L.A. in the North Carolina State University Molecular Biotechnology Training Program, the National Science Foundation (grant MCB-0918611), and the W. M. Keck Foundation. The authors also thank Hunter Walker and Tim Collier for their helpful discussions.

## References

- Rudnick, P.A., Clauser, K.R., Kilpatrick, L.E., Tchekhovskoi, D.V., Neta, P., Blonder, N., Billheimer, D.D., Blackman, R.K., Bunk, D.M., Cardasis, H.L., Ham, A.J.L., Jaffe, J.D., Kinsinger, C.R., Mesri, M., Neubert, T.A., Schilling, B., Tabb, D.L., Tegeler, T.J., Vega-Montoto, L., Variyath, A.M., Wang, M., Wang, P., Whiteaker, J.R., Zimmerman, L.J., Carr, S.A., Fisher, S.J., Gibson, B.W., Paulovich, A.G., Regnier, F.E., Rodriguez, H., Spiegelman, C., Tempst, P., Liebler, D.C., Stein, S.E.: Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses. *Mol. Cell Proteom.* **9**(2), 225–241 (2010)
- Paulovich, A.G., Billheimer, D., Ham, A.J.L., Vega-Montoto, L., Rudnick, P.A., Tabb, D.L., Wang, P., Blackman, R.K., Bunk, D.M., Cardasis, H.L., Clauser, K.R., Kinsinger, C.R., Schilling, B., Tegeler, T.J., Variyath, A.M., Wang, M., Whiteaker, J.R., Zimmerman, L.J., Fenyo, D., Carr, S.A., Fisher, S.J., Gibson, B.W., Mesri, M., Neubert, T.A., Regnier, F.E., Rodriguez, H., Spiegelman, C., Stein, S.E., Tempst, P., Liebler, D.C.: Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS platform performance. *Mol. Cell Proteom.* **9**(2), 242–254 (2010)
- Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.J.L., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L., Carr, S.A., Clauser, K.R., Jaffe, J.D., Kowalski, K.A., Neubert, T.A., Regnier, F.E., Schilling, B., Tegeler, T.J., Wang, M., Wang, P., Whiteaker, J.R., Zimmerman, L.J., Fisher, S.J., Gibson, B.W., Kinsinger, C.R., Mesri, M., Rodriguez, H., Stein, S.E., Tempst, P., Paulovich, A.G., Liebler, D.C., Spiegelman, C.: Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **9**(2), 761–776 (2010)
- Oberacher, H., Walcher, W., Huber, C.G.: Effect of instrument tuning on the detectability of biopolymers in electrospray ionization mass spectrometry. *J. Mass Spectrom.* **38**(1), 10–116 (2003)
- Soule, M.C.K., Longnecker, K., Giovannoni, S.J., Kujawinski, E.B.: Impact of instrument and experiment parameters on reproducibility of ultrahigh resolution ESI FT-ICR mass spectra of natural organic matter. *Org. Geochem.* **41**(8), 725–733 (2010)
- Zhou, Y., Song, J.Z., Choi, F.F.K., Wu, H.F., Qiao, C.F., Ding, L.S., Gesang, S.L., Xu, H.X.: An experimental design approach using response surface techniques to obtain optimal liquid chromatography and mass spectrometry conditions to determine the alkaloids in *Meconopsis* species. *J. Chromatogr. A* **1216**(42), 7013–7023 (2009)
- Wenner, B.R., Lynn, B.C.: Factors that affect ion trap data-dependent MS/MS in proteomics. *J. Am. Soc. Mass Spectrom.* **15**(2), 150–157 (2004)
- Louvar, J.F.: Simplify, experimental design. *Chemical Eng. Prog.* **106**(1), 35–40 (2010)
- Riter, L.S., Vitek, O., Gooding, K.M., Hodge, B.D., Julian Jr., R.K.: Statistical design of experiments as a tool in mass spectrometry. *J. Mass Spectrom.* **40**(5), 565–579 (2005)
- Raji, M.A., Schug, K.A.: Chemometric study of the influence of instrumental parameters on ESI-MS analyte response using full factorial design. *Int. J. Mass Spectrom.* **279**(2/3), 100–106 (2009)
- Robichaud, G., Dixon, R.B., Potturi, A.S., Cassidy, D., Edwards, J.R., Dow, T.A., Muddiman, D.C.: Design, modeling, fabrication, and evaluation of the air amplifier for improved detection of biomolecules by electrospray ionization mass spectrometry. *Int. J. Mass Spectrom.* (2010), in press.
- Walker, S.H., Papas, B.N., Comins, D.L., Muddiman, D.C.: Interplay of permanent charge and hydrophobicity in the electrospray ionization of glycans. *Anal. Chem.* **82**(15), 6636–6642 (2010)
- Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B.: Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**(4), 1017–1031 (2007)
- Gao, J., Friedrichs, M.S., Dongre, A.R., Opiteck, G.J.: Guidelines for the routine application of the peptide hits technique. *J. Am. Soc. Mass Spectrom.* **16**(8), 1231–1238 (2005)
- Liu, H.B., Sadygov, R.G., Yates, J.R.: A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**(14), 4193–4201 (2004)
- Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., Ahn, N.G.: Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell Proteom.* **4**(10), 1487–1502 (2005)
- Zybailov, B., Coleman, M.K., Florens, L., Washburn, M.P.: Correlation of relative abundance ratios derived from peptide ion chromatograms

- and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal. Chem.* **77**(19), 6218–6224 (2005)
18. Payne, W.E., Garrels, J.I.: Yeast protein database (YPD): a database for the complete proteome of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **25**(1), 5–62 (1997)
  19. Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M.: Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* **93**(25), 14440–14445 (1996)
  20. Washburn, M.P., Wolters, D., Yates, J.R.: Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**(3), 242–247 (2001)
  21. Peng, J.M., Elias, J.E., Thoreen, C.C., Licklider, L.J., Gygi, S.P.: Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2**(1), 43–50 (2003)
  22. Nägele, E., Vollmer, M., Hörth, P.: Improved 2D nano-LC/MS for proteomics applications: a comparative analysis using yeast proteome. *J. Biomol. Tech.* **15**, 15134–15143 (2004)
  23. Wei, J., Sun, J., Yu, W., Jones, A., Oeller, P., Keller, M., Woodnutt, G., Short, J.M.: Global proteome discovery using an online three-dimensional LC-MS/MS. *J. Proteome Res.* **4**(3), 801–808 (2005)
  24. de Godoy, L.M.F., Olsen, J.V., de Souza, G.A., Li, G.Q., Mortensen, P., Mann, M.: Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **7**(6), R50 (2006)
  25. Piening, B.D., Wang, P., Bangur, C.S., Whiteaker, J., Zhang, H.D., Feng, L.C., Keane, J.F., Eng, J.K., Tang, H., Prakash, A., McIntosh, M. W., Paulovich, A.: Quality control metrics for LC-MS feature detection tools demonstrated on *Saccharomyces cerevisiae* proteomic profiles. *J. Proteome Res.* **5**(7), 1527–1534 (2006)
  26. Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B., Aebersold, R.: Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138**(4), 795–806 (2009)
  27. de Godoy, L.M.F., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Frohlich, F., Walther, T.C., Mann, M.: Comprehensive mass spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**(7217), 1251–1260 (2008)
  28. Fletcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., Garrels, J.I.: A sampling of the yeast proteome. *Mol. Cell Biol.* **19**(11), 7357–7368 (1999)
  29. Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R.: Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19** (3), 1720–1730 (1999)
  30. Usaita, R., Wohlschlegel, J., Venable, J.D., Park, S.K., Nielsen, J., Olsson, L., Yates, J.R.: Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression *Saccharomyces cerevisiae* strains: the comparison of two quantitative methods. *J. Proteome Res.* **7**(1), 266–275 (2008)
  31. Washburn, M.P., Koller, A., Oshiro, G., Ulaszek, R.R., Plouffe, D., Deciu, C., Winzler, E., Yates, J.R.: Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **100**(6), 3107–3112 (2003)
  32. Andrews, G.L., Shuford, C.M., Burnett, J.C., Hawkrige, A.M., Muddiman, D.C.: Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **877**(10), 948–954 (2009)
  33. Keller, A., Nesvizhskii, A.I., Kolker, E., Aebersold, R.: Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**(20), 5383–5392 (2002)
  34. Nesvizhskii, A.I., Keller, A., Kolker, E., Aebersold, R.: A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**(17), 4646–4658 (2003)
  35. Weatherly, D.B., Atwood, J.A., Minning, T.A., Cavola, C., Tarleton, R. L., Orlando, R.: A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol. Cell Proteom.* **4**(6), 762–772 (2005)
  36. Tabb, D.L., Fernando, C.G., Chambers, M.C.: MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J. Proteome Res.* **6**(2), 654–661 (2007)
  37. Liu, H.J., Finch, J.W., Lavalley, M.J., Collamati, R.A., Benevides, C.C., Gebler, J.C.: Effects of column length, particle size, gradient length, and flow rate on peak capacity of nanoscale liquid chromatography for peptide separations. *J. Chromatogr. A* **1147**(1), 30–36 (2007)
  38. Wang, X.L., Stoll, D.R., Schellinger, A.P., Carr, P.W.: Peak capacity optimization of peptide separations in reversed-phase gradient elution chromatography: fixed column format. *Anal. Chem.* **78**(10), 3406–3416 (2006)
  39. Johnson, K.L.: In (2010)
  40. Zhang, Y., Wen, Z.H., Washburn, M.P., Florens, L.: Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Anal. Chem.* **81**(15), 6317–6326 (2009)
  41. Wong, C.C.L., Cociorva, D., Venable, J.D., Xu, T., Yates, J.R.: Comparison of different signal thresholds on data dependent sampling in Orbitrap and LTQ mass spectrometry for the identification of peptides and proteins in complex mixtures. *J. Am. Soc. Mass Spectrom.* **20**(8), 1405–1414 (2009)
  42. Kim, M.S., Kandasamy, K., Chaerkady, R., Pandey, A.: Assessment of resolution parameters for CID-based shotgun proteomic experiments on the LTQ-Orbitrap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **21** (9), 1606–1611 (2010)
  43. Schlotzhauer, S.D.: Elementary statistics using JMP. SAS Institute Inc, Cary (2007)