



HAL
open science

A sDOE (simple design-of-experiment) approach for parameter optimization in mass spectrometry. Part 1. Parameter selection and interference effects in topdown ETD fragmentation of proteins in a UHR-QTOF instrument.

Guillaume Gabant, Martin Stekovic, Matej Nemcic, Justine Pinêtre, Martine Cadene

► **To cite this version:**

Guillaume Gabant, Martin Stekovic, Matej Nemcic, Justine Pinêtre, Martine Cadene. A sDOE (simple design-of-experiment) approach for parameter optimization in mass spectrometry. Part 1. Parameter selection and interference effects in topdown ETD fragmentation of proteins in a UHR-QTOF instrument.. *Journal of The American Society for Mass Spectrometry*, 2023, 34 (1), pp.27-35. 10.1021/jasms.2c00215 . hal-03871364

HAL Id: hal-03871364

<https://hal.science/hal-03871364v1>

Submitted on 20 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A sDOE (simple design-of-experiment) approach for parameter optimization in mass spectrometry. Part 1. Parameter selection and interference effects in top-down ETD fragmentation of proteins in a UHR-QTOF instrument

Guillaume Gabant¹, Martin Stekovic¹, Matej Nemcic¹, Justine Pinêtre¹, Martine Cadene¹

¹ Centre de Biophysique Moléculaire, UPR4301, CNRS, affiliated to Université d'Orléans, Rue Charles Sadron, 45071, Orléans Cedex 2, France.

Running title

Simple DOE for parameter optimization of mass spectrometry instruments

Keywords: Design of Experiments, sDOE, mass spectrometry, UHR-QTOF, ETD, optimization

Address reprint requests to Martine Cadene, CBM CNRS, rue Charles Sadron, 45071 Orleans cedex 2, France, email: martine.cadene@cnrs-orleans.fr

Abstract

Design-of-experiment (DOE) approaches, originally conceived by J. Fischer, are widely applied in industry, particularly in the context of production for which have been greatly expended. In a research and development context,

DOE can be of great use for method development. Specifically, DOE can greatly speed up instrument parameter optimization by first identifying parameters that are critical to a given outcome, showing parameter interdependency where it occurs, and accelerating optimization of said parameters using matrices of experimental conditions.

While DOE approaches have been applied in mass spectrometry experiments, they have so far failed to gain widespread adoption. This could be attributed to the fact that DOE can get quite complex and daunting to the everyday user. Here we make the case that a subset of DOE tools, here after called SimpleDOE (sDOE), can make DOE accessible and useful to the Mass Spectrometry community at large. We illustrate the progressive gains from a purely manual approach to sDOE through a stepwise optimization of parameters affecting the efficiency of top-down ETD fragmentation of proteins on a high-resolution Q-TOF mass spectrometer, where the aim is to maximize sequence coverage of fragmentation events.



191 words

Introduction:

To carry out their functions, proteins are expressed in many flavors *in vivo* and *in vitro*. These flavors, called proteoforms, are isoforms resulting from sequence variability, post-translational modifications (PTMs), and/or modifications induced by exogenous agents such as radiations or chemicals. The precise localization of modifications and their quantification is becoming a major quest of biology. While the bottom-up proteomics approach based on proteolysis and peptide MS fragmentation can reliably be used to address this challenge [1], it presents some drawbacks, such as the loss of filiation between a modified peptide and its proteoform of origin [2]. The top-down approach based on the fragmentation of intact protein molecules leads not only to primary sequence information but also to positional identification of modifications [3,4], and potentially, to their quantification [5]. Since sequence coverage is paramount to a successful top-down characterization of proteoforms, extending fragmentation coverage (FC) becomes central to top-down strategies.

In terms of sequencing, several fragmentation methods are readily available on commercial instruments such as collision-induced dissociation (CID), photodissociation (PD), electron-transfer dissociation (ETD), electron-capture dissociation (ECD), and (ExD). In CID, multiple collisions increase the internal energy of precursor ions, which can cause both internal fragmentation, a re-fragmentation of fragment ions, and loss of side-chain PTMs [6]. In ETD [7], a singly charged anion reagent known as the "reactant" transfers an electron to multiply protonated peptides/proteins, inducing fragmentation along the peptide backbone. In contrast to CID, this method achieves nonergodic fragmentation without prior redistribution of energy [6] and leads to a rapid bond cleavage predominantly at the sites where the electron was trapped [8], thereby promoting random N-C α (alpha carbon) bond breakage while leaving side-chains and PTMs intact. Unlike CID, the outcome of ETD is affected by a number of parameters including but not limited to the amount of reagent and precursor ions, reaction time, and selected charge state [9]. Top-down, middle-down and bottom-up ETD fragmentation in a Q-TOF instrument can be used to characterize the structural heterogeneity of antibodies [10], to read out HDX results while minimizing scrambling [11] and to discriminate leucine from isoleucine, in a data-independent analysis on peptides [12]. Top-down fragmentation presents a particular challenge as fragmentation of peptide bonds in the middle of the protein tends to be energetically unfavorable, resulting in a typical "middle-gap" coverage pattern, also known as the "dark zone" in fragmentation coverage along the

sequence. Achieving the high fragmentation coverage necessary to localize modifications anywhere for any kind of proteoform thus requires the careful optimization of ETD fragmentation conditions [4].

The principle of a rational design of experiments to give a sound statistical basis to experimental results interpretation was initially conceived by Fischer as exposed in his foundational book [13]. Design-of-experiment (DOE) approaches *per se* were formalized by statisticians [14] and saw exponential growth in the 1950s as a means to provide for fast, reliable, and objective optimization of production processes [15]. These methods are regularly used in the pharmaceutical and food industry [16]. Complex production processes depend on multiple parameters, some of which may be interdependent. The motivation for DOE in this context is based on the simple observation that the parameters which are critical to the performance of a product cannot necessarily be intuited based on past experience in production.

To reach a certain level of performance to match a specific research aim, mass spectrometrists likewise need to optimize instrument parameters and quickly face a multiplicity of parameters, complexity, and interdependency. ETD fragmentation, alone or in combination with other fragmentation modes, is a good illustration of this type of situation. Figure 1 shows the main components a typical ETD fragmentation cycle. Hereafter, we will refer to precursor intensity as PI, reagent intensity as RI, and reaction time as RT.

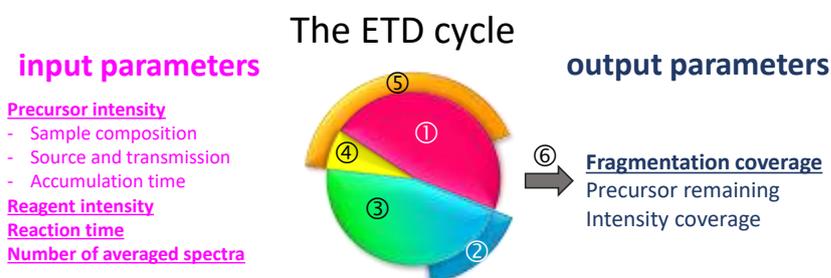


Figure 1. Overview of input and output parameters. The ETD cycle includes the 4 following steps: 1) Accumulation of precursor ions, 2) Accumulation of reagent ions, 3) ETD Reaction, 4) Fragment transfer and 5) Detection of fragment ions. An additional step was considered: 6) Data processing. Image inspired from Bruker Daltonics illustration, with permission.

Different strategies can be applied to optimize multiple parameters for a given desired outcome ("response" in DOE terms), such as maximum protein fragmentation coverage. The simplest and most frequent strategy is sequential optimization of parameters or "change one at a time", which can be described as the "manual" approach. This approach obviously consumes large amounts of resources, including sample quantities and time. After the first round of optimization, each parameter may need to be readjusted because its optimum has

shifted in response to the change in at least one other parameter which interferes with it. One way to attempt circumvention of this problem is to plan a "co-optimization" experiment, by making a 2D table of values for two parameters to be tested jointly. Co-optimization has its own limits: first, the choice of parameters to co-optimize relies on assumptions about which co-dependencies are most relevant, and second, it is still time-consuming and sample-intensive. Faced with this problem, some authors have proposed to use a "brute-force" approach to scan for all possible combinations of parameters in the instrument, or a large subset thereof [17]. Automatizing the optimization process requires scripting of instrument parameter setups. This approach, however, still requires both large amounts of time and sample.

In a research context, samples are often scarce, so we looked for an alternative to make the process less time-consuming and sample-intensive. DOE, by cleverly combining a subset of conditions to test parameters effects, offers just such an alternative. Here we show how it can be used to optimize the efficiency of ETD fragmentation, in comparison with the "manual" and "co-optimization" parameter optimization approaches. We give concrete examples of building matrices, performing experiments, and calculating effects of different parameters. This allows us to identify parameters that are critical to the outcomes of interest, mainly fragmentation coverage, and to show parameter interdependency. In this first paper we show how a selection of DOE tools, which we call sDOE, can offer the user a large part of the benefits of DOE, without necessarily delving into the more complex aspects of the method.

Methods

Reagents

Acetonitrile and formic acid were from Biosolve (Dieuze, France). 18 M Ω purified water (MilliQ reagent grade system, Millipore) was used to prepare the ionization solution. Ubiquitin and apomyoglobin were purchased from Sigma (St. Louis, MO). Recombinant human Phosphatidyl Ethanolamine Binding Protein 1 (hPEBP1) was cloned and overexpressed in BL21DE3 *E.coli* as previously described by Schoentgen and coworkers [18]. The protein was purified without tag using the protocol originally developed by Bernier and Jolles [19]. HU α -beta protein was a kind gift from Dr Bertrand Castaing at CBM.

Top-down mass spectrometry fragmentation

Top-down MS sequencing of proteins was performed using ESI-QTOF ETD fragmentation method. Samples were diluted at 0.5 μM in a solution of 50% acetonitrile, 49.8% water and 0.2% formic acid and were analyzed by direct infusion in a 4-GHz MaXis ultra high resolution quadrupole-TOF mass spectrometer (Bruker Daltonics) equipped with an electrospray ion source. Spectra were acquired in positive ion MS mode over a 115–3000 m/z range and were externally calibrated with the ESI-L Low Concentration Tuning Mix (Agilent Technologies). Mass spectra were processed using DataAnalysis 3.1 software (Bruker Daltonics). When applied, the smoothing parameters were Savitzky Golay algorithm, width 0.05 Da and 1 cycle. Peaks were annotated using the SNAP algorithm with a quality factor of 0.5. Fragmentation coverage (FC) was obtained using Biotoools software (Bruker). The c, c-1, c+1, z, z+1 and z+2 fragment ions were considered.

Design of experiments

The Design Of Experiments (DOE) method is an efficient and structured way to maximize information while minimizing resources. It is used to develop a process that will 1) identify the factors that have the largest effect on the studied output (also known as the response) and 2) adjust them to achieve an optimal response. In the present work, we focus on the first step and the output of interest is the fragmentation coverage.

In the first step, screening experiments are designed to gauge the relative effects of the factors studied. This allows for a first coarse selection of the most interesting factors, characterized as "crucial" or "critical" [20], in other words, factors which strongly affect the studied process. In mass spectrometry, factors could be instrumental parameters, sample conditions, or any other factor thought to influence the outcome of the experiment. To set up factor selection experiments, we can use a) a full factorial design or b) a fractional factorial design. A full factorial design also gives access to interference effects between parameters, while a fractional factorial design only gives access screening of individual factors, albeit with fewer experiments. When measuring a response, the general principle of a factorial design is to plan to vary 2 or more factors at the same time. As shown in Table 1, a matrix is built by choosing for each factor two numerical or alphanumeric values called levels -1 and +1. Level -1 can be the "base level", i.e. a starting quantitative or qualitative value based on prior data. If the design is orthogonal, i.e. if factors are varied independently, then

the effect of a factor can be calculated independently of the effect of other factors. Mathematically, two factors are orthogonal if the sum of their columns in standard format (+1, -1) equals 0. In the present work we used two full factorial designs of 4 and 5 factors to quickly highlight important parameters for FC. The building of experimental runs for a full 4-factor design is shown in Table 1.

Table 1. Full factorial matrix for 4 factors with 2 levels.

Experiment #	X1	X2	X3	X4	Response y
1	-1	-1	-1	-1	y1
2	-1	-1	-1	1	y2
3	-1	-1	1	-1	y3
4	-1	-1	1	1	y4
5	-1	1	-1	-1	y5
6	-1	1	-1	1	y6
7	-1	1	1	-1	y7
8	-1	1	1	1	y8
9	1	-1	-1	-1	y9
10	1	-1	-1	1	y10
11	1	-1	1	-1	y11
12	1	-1	1	1	y12
13	1	1	-1	-1	y13
14	1	1	-1	1	y14
15	1	1	1	-1	y15
16	1	1	1	1	y16

In the 4 factors matrix, factors were protein identity, reaction time, isCID and charge state (CS), while in the 5 factors matrix, we picked PI, number of precursor ions (NPI), CS, RT, and number of accumulated spectra (NAS). The rationale for picking these factors is described in the results section.

For the DOE experiments presented herein, the effect of a parameter is calculated as the sum of the products of the level (-1 or +1) and the outcome (FC) of a given experiment, divided by the number of the experiments (n) as represented by the following equation:

$$Effect = \frac{\sum_{i=1}^n (level_i \times FC_i)}{n}$$

Once key factors have been identified, in the second step finer experiments focusing on these key factors are needed to find the set of factor values that will maximize the response, i.e. the optimal combination. These optimization experiment designs can rely on full factorial design if one the factors is qualitative, or on response surface design, if all the factors are quantitative. In these experiments, it may be advisable to test more than two levels for certain factors, for example when the response to a given factor is suspected of being non-linear.

Results

To get a useable ETD fragmentation reaction, one basically needs to optimize precursor intensity (PI), reagent intensity (RI) and reaction time (RT). PI is however directly dependent on the precursor accumulation time (AT), just as RI is dependent on the reactant reagent injection time (RgT). Since our instrument is not equipped with automatic ion gain control, the accumulation time needs to be adjusted manually to reach a given intensity. We started by optimizing AT using two calibrant proteins, namely ubiquitin ($M_{\text{mono}} = 8\,560$ Da) and apomyoglobin ($M_{\text{mono}} = 16\,941$ Da) (Figure 2A). The dependency of PI upon AT values can be seen to be protein-dependent, which is the first clue that the ETD optimization process may have to be performed anew for each type of protein sample.

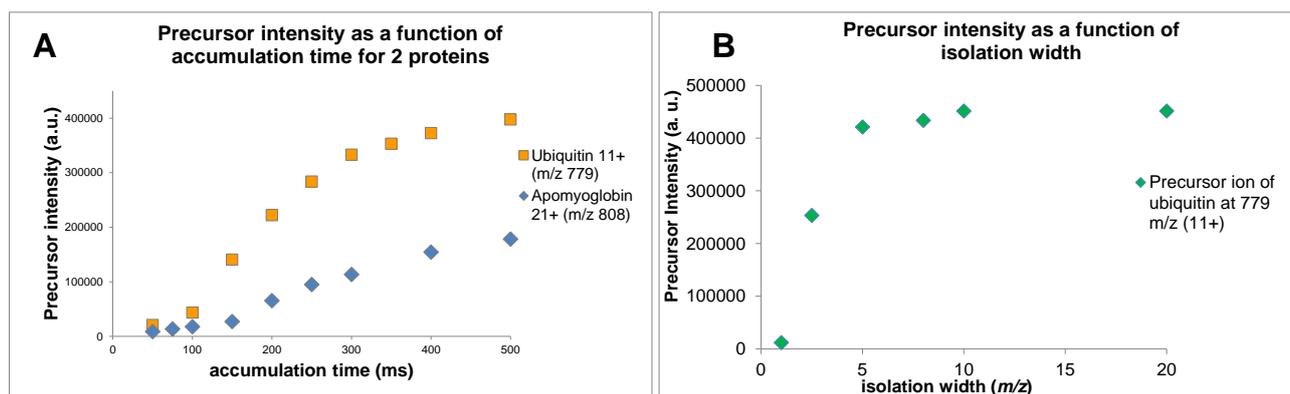


Figure 2. Manual optimization of precursor accumulation time and isolation width. A. Precursor intensity as a function of accumulation time for ubiquitin and apomyoglobin, B. Precursor intensity as a function of isolation width for ubiquitin. The accumulation time used in the isolation width study was 300 ms.

Precursor isolation width is commonly optimized and shows a saturation effect (Figure 2B). After further optimization based on reagent time, a FC of 55.9% was obtained on the hPEBP1 protein. From this number, it is immediately apparent that in this example, "manual" optimization of parameters yields insufficient FC for most PTM localization applications. Apart from favorable cases such as ubiquitin and from thorough, time consuming manual optimization of the parameters, in ETD such manual approach may frequently yield a FC that is indeed too low for these applications. Thus, we looked into parameters that may benefit from a "co-optimization" approach.

Using ubiquitin as an "expendable" model protein, we set up experiments varying both PI and RT. For this model system, the FC varies between 60 and 98%, while a co-optimum can be seen at PI = 74 000 and RT = 8 ms (Figure 3A).

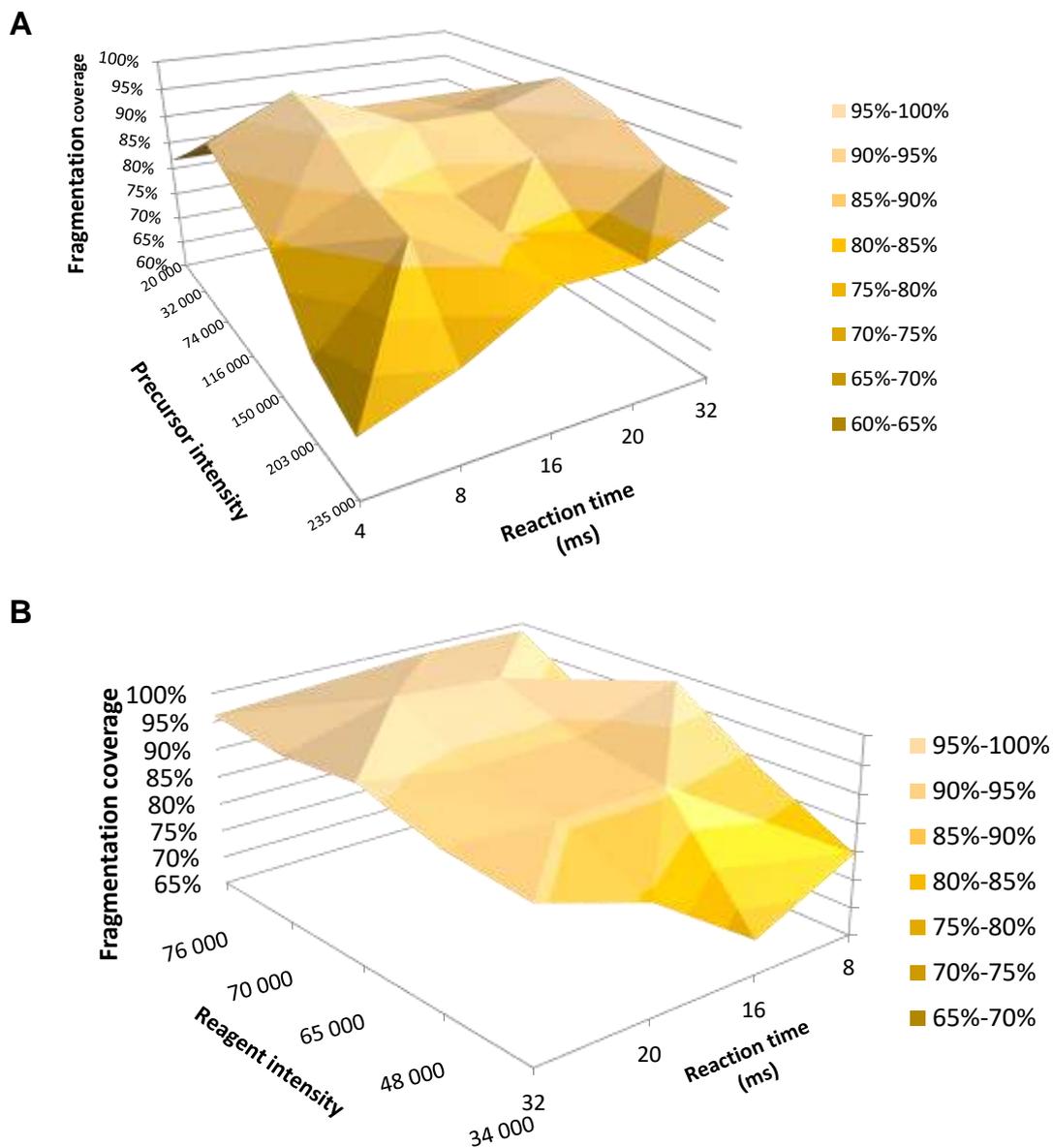


Figure 3. 2D plots of co-optimization of parameters for maximizing FC. A. Fragmentation coverage as a function of pairs of (PI,RT) values. B. Fragmentation coverage as a function of pairs (RI,RT) of values.

To evaluate the repeatability of measurements, we performed a decuplicate experiment. The distributions of FC for ubiquitin and apomyoglobin are shown as a box plot (Figure 4).

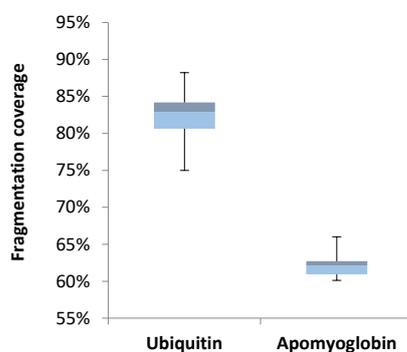


Figure 4. Distribution of FC for ETD decuplicate experiments on ubiquitin and apomyoglobin.

Standard deviations for FC were 3.9% for ubiquitin and 1.8% for apomyoglobin, respectively. These values can give an order of magnitude for the degree of variability for FC determination, which can give some context when evaluating the result of manual or co-optimization. We also noted that reagent accumulation time and reaction time cannot be completely independent because, as can be seen in Figure 1, the reaction of proteins/peptides with the ETD reactant can start as soon as the reactant is introduced into the fragmentation chamber. Co-optimization of these parameters proved useful as well. The plotted surface shows a co-optimum area from 70 000 to 76 000 for the reagent intensity and from 8 ms to 20 ms for the reaction time (Figure 3B).

To further our understanding of which parameters have the largest effect on the output and/or interfere with one another, we next set up a DOE experiment to start exploring this tool. A set of 4 parameters can be tested with a relatively simple full factorial matrix as shown in Table 1. For each parameter, some prior knowledge of the system is preferable to assist in choosing two different levels per factor, hereafter labeled as the "+1" and "-1" conditions.

We included as parameters the identity of the protein, reaction time, the voltage for in-source collision-induced dissociation (isCID), and the charge state, all of which seemed to affect FC during manual optimizations (Table 2).

Table 2. Determination of the main effects and interactions. Full factorial matrix with two levels per factor. Reagent intensity was set at 70 000 and precursor intensity at 75 000. Proteins were HU alpha chain (M = 9535 Da) and apomyoglobin at 0.5 μ M. For apomyoglobin the charge states used were 21+ and 25+, while for HU they were 11+ and 14+.

	-1	+1
Protein identity	HU alpha	ApoMyo
Reaction Time (ms)	5	50
isCID (eV)	10	50
Charge state	Z_n	$Z_{n \times 1.2}$

For the value of Z_n , the charge state with the highest intensity in the charge state distribution of a previous experiment was chosen. Since by increasing the number of charges, we can expect to improve the yield of the ETD reaction, for the second charge state value we chose to test a value corresponding to a 20% increase in z , called $Z_{n \times 1.2}$. The criterion for choosing 20% as the increase is a compromise between a higher charge state and a usable intensity. The 25+ and 14+ charges states of apomyoglobin and HU respectively, fit well with these specifications.

Individually, the identity of the protein (which contains its sequence data and physicochemical profile) was found to have the largest effect, with over twice the effect observed for the next-most important factor, namely RT (Table 3). Factors that can make ETD fragmentation protein-dependent, may include protein size, sequence and residual gas phase structure. Protein sequence affects the distribution of polar, apolar, positively and negatively charged amino acids along the sequence, which may influence the transfer of electrons in the ETD process. A fully unfolded structure should allow full access to peptide bonds for electron transfer. However, the sequence of a protein affects its folding in solution, and unfolding kinetics in solvents and during transfer to the fragmentation cell is also protein-dependent. As a consequence, any structure that survives denaturing solvent conditions as well as transfer to the gas phase may influence ETD fragmentation, as was shown for example for ECD [21].

Factor	Effect (%)
Protein identity	-10.8
Reaction time	4.7
isCID	-2.5
Charge state	4.4
Protein identity x Reaction time	-0.1
Protein identity x isCID	0.3
Protein identity x Charge state	-2.0
Reaction time x isCID	-1.7
Reaction time x Charge state	-2.3
isCID x Charge state	0.5

Table 3. Effect of individual parameters and their interactions on fragmentation coverage.

Interestingly, the largest interference effects were seen for the charge state x RT effects and the protein identity x charge state effects, so that both of the largest observed interference effects involve the charge state of the precursor. While the interpretation of interference effects can be challenging, some hypotheses based on the distribution of positively charged sites on the protein can be formulated. For the interference of the charge state with RT, since the reactant brings an electron for transfer, electrostatic effects between positively charged sites and the reactant affect the transfer of electrons to the backbone, and the charge state affects the distribution of positively charged sites. As a consequence, for a given efficiency of transfer for a given charge state, reaction time may need to be adjusted to compensate for these electrostatic effects, and the adjustment would then depend on the charge state. From the point of view of the sequence, protein identity may also affect the efficiency of transfer of electrons to the backbone, as the identity of amino acid residues has a direct bearing on the ionization of residues. This may explain the interplay between charge state and protein identity.

Next we moved to optimizations for a more scarce protein, the hPEBP1 metastasis suppressor. For this single protein and 5 different parameters, a full factorial matrix consisting of $2^5 = 32$ partially randomized top-down MS experiments was constructed (Table 4). After co-optimization, the best FC values were obtained with a RI of 70000-76000 which correspond to a reagent accumulation time of approximately 10 ms. Since the maximum reagent intensity and the length of reagent injection time are dependent on the amount of reagent in the cartridge, and the reagent gets consumed over time, keeping reagent intensity at 70000 makes it possible to do multiple experiments without being limited in reachable reagent intensity. Increasing the reagent injection time

would also have had an adverse impact on the length of the duty cycle. Thanks to the co-optimization results, the RI parameter was set at 70000 for subsequent experiments.

Table 4. Full factorial matrix with two levels per factor. Reagent intensity was set at 70 000. Protein was hPEBP1 (M = 20926 Da) at 0.5 μ M.

Factors		-1	1
X1	Precursor intensity	75 000	135 000
X2	Charge state	26+	30+
X3	Number of precursor ion peaks	1	3
X4	Reaction Time (ms)	20	60
X5	Number of averaged spectra	200	800

The first 4 parameters are randomized, meaning fragmentation sequences are programmed in random order in the analysis. Upon closer examination of experimental data, we noticed that the fragmentation segments we programmed in the instrument had a variable number of spectra. This is due to the effect of variations in protein accumulation times and reaction time on duty cycle length. To test the effect of this parameter (number of averaged spectra), we constructed a 5-parameter matrix by picking either 200 spectra, or the following 800 spectra within the same segment, which means that the NAS parameter is not randomized with respect to the others. In DOE, mixing randomized and non-randomized parameters is allowed. As can be seen in Table 5, PI and NAS are in a draw as the main parameter to optimize. In this representation, each X factor (parameter) interference with another X factor is shown. The major interference, PI with RT, is quite far ahead of all others.

Table 5. Effect of individual parameters and their interactions on hPEBP1 fragmentation coverage

Factor	Effect (%)
Precursor intensity	6.6%
Charge state	-1.9%
Number of precursor ion peaks	-0.9%
Reaction Time (ms)	0.7%
Number of averaged spectra	6.4%
Precursor intensity x Charge state	-0.4%
Precursor intensity x Nbr of precursor ion	0.8%
Precursor intensity x Reaction Time	2.3%
Precursor intensity x Nbr of av. spectra	-1.2%
Charge state x Nbr of precursor ion	0.4%
Charge state x Reaction Time	-0.2%
Charge state x Nbr of av. spectra	-1.0%
Nbr of precursor ion x Reaction Time	-0.4%
Nbr of precursor ion x Nbr of av. spectra	-0.6%
Reaction Time x Nbr of av. spectra	-0.8%

Figure 5 shows the extent of progress that could be achieved with manual optimization (upper panel) versus co-optimization (middle panel) versus this first DOE step (lower panel). A fragmentation coverage of 76.3% is reached with the following parameter setting: PI = 135 000, CS = 26+, NPI = 1, RT = 60 ms, NAS = 800.

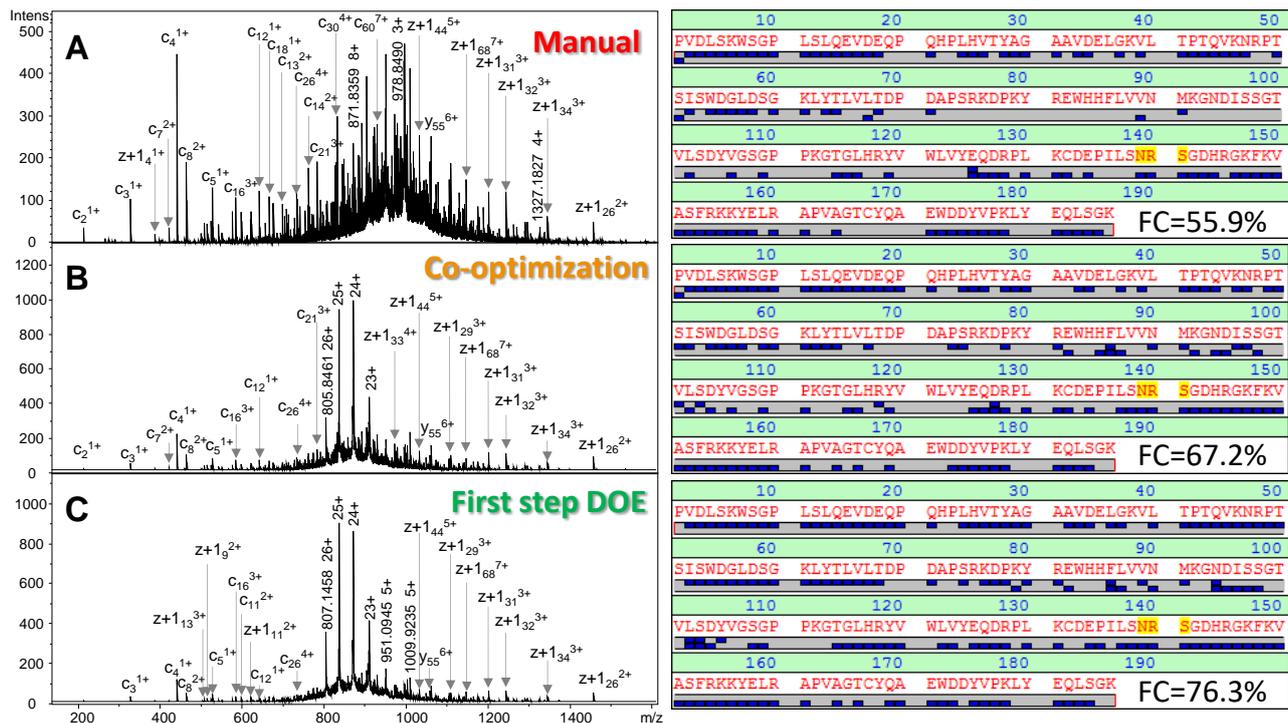


Figure 5. Effect of the optimization strategy on optimum fragmentation coverage of hPEBP1 protein. A. ETD spectrum and fragmentation coverage after manual optimization. The spectrum was not smoothed. The parameters are CS = 26+, NPI = 1, PI = 145 000, RI = 100 000, RT = 80ms, NAS = 440. B. ETD spectrum and fragmentation coverage after co-optimization. The spectrum was smoothed. The parameters are CS = 26+, NPI = 1, PI = 150 000, RI = 70 000, RT = 64ms, NAS = 140. C. ETD spectrum and fragmentation coverage after first step DOE optimization. The spectrum was smoothed. The parameters are CS = 26+, NPI = 1, PI = 135 000, RI = 70 000, RT = 20ms, NAS = 800. For all the spectra, the MS/MS tolerance was 0.1 Da for the fragmentation coverage determination.

Discussion

If we define fragmentation coverage (FC), as the percentage of amino acids bonds in the sequence where at least one fragmentation event is observed through fragment ions, the difficulty of achieving close to 100% FC in top-down sequencing increases with respect to protein size [22]. Fragmentation occurs predominantly at the ends of the protein, leaving an area at the center without corresponding fragments, which is termed "dark zone" to indicate the lack of sequence information. Dark zones can also occur outside the center area in a sequence. For an exhaustive census of modified residues, dark zones are an obvious impediment to reaching the stated goal. When faced with incomplete fragmentation coverage, several responses are possible. The first is to abandon the study of a protein showing imperfect coverage and to focus on systems that behave perfectly on the first try. This "low-hanging fruit" approach has serious limitations, as the mass spectrometry community will quickly run out of such ideal systems. Moreover, this analysis ethics runs counter to the goal of science,

which is to go beyond the tip of the iceberg to explore both emerged and immersed parts of the factual berg, and counter to the social compact, which is to provide information that is useful for human health, wellbeing, the environment, and economics. Therefore, to achieve high FC for any kind of protein and proteoform, ETD fragmentation requires careful optimization. In the present paper, we compare several strategies for optimization on an UHR-QTOF MS instrument with ETD fragmentation, with a final focus on a simple version of DOE. Ultimately, the reasoning could also apply to any number of MS techniques where multiple input parameters affect the outcome of the analysis. To improve the efficiency of top-down ETD fragmentation of proteins on a high-resolution Q-TOF mass spectrometer with the aim of maximizing FC, we started from single-parameter optimization, then moved to co-optimization of parameters. Through a peak-and-valley visualization of the optimization landscape, we could see the importance of interdependency (Figure 3). Yet co-optimization still only deals with 2 parameters at a time.

Other authors have made the case that sequential, manual optimization or co-optimization of MS/MS experiments is too slow and incomplete. Jensen *et al* had to deal with multiple parameters when using a combination of CID, HCD, ETD, ETciD, EThcD, and UVPD for top-down protein fragmentation on an Orbitrap instrument [17]. The solution the authors proposed and successfully applied was to first narrow down the set of parameters they thought would most affect the outcome (FC), then to use a "brute-force" approach to scan for all possible combinations of these parameters in the instrument. This required the development of the "topdownr" R-package to automate multimodal MS/MS analyses as well as data treatment, which basically scripts instrument parameter setups to allow for automation of the data acquisition and filtering of thousands of MS/MS datasets [17]. This approach, however, requires both time and large amounts of sample, which are often scarce in a research context, prompting us to search for an alternative.

DOE, by cleverly combining a subset of conditions to test parameters effects, offers just such an alternative. Experimental plans are designed as matrices of conditions to run experiments. DOE can be used to maximize the efficiency of ETD fragmentation, or in principle any combination of MS/MS fragmentation methods. In addition to parameters that are selected and randomized prior to the experiment, parameters can be added to the analysis *a posteriori* as non-randomized parameters. This provides some flexibility in the exploitation of data, for a limited cost in data independence. Different sets of DOE tools have been successfully applied for optimization in an industrial [23] or academic [24,25,26,27,28,29,30,31] context, and/or for method validation [32,33].

A number of reviews [20,34] have sought to make full sets of DOE tools accessible and attractive to the analytical sciences community. Hibbert produced a detailed and instructive guide on how to choose and apply tools in DOE [21]. Hecht *et al.* wrote an extensive review with case studies to "enable a post-doctorate mass spectrometrist to utilize DOE to design and execute an optimization study independently and without a statistician consultant" [34]. Despite the clarity and pedagogy displayed in these reviews, DOE as a general set of tools may still seem quite complex and daunting to the everyday user. Yet, DOE should attract a much larger audience since a subset of these tools potentially yields a large portion of useful information for a fraction of the time and sample costs.

The present work furnishes concrete examples of building matrices, performing experiments, and calculating the relative effects of different parameters. This allows us to identify parameters that are critical to FC, the outcome of interest, and to show parameter interdependency for the purpose of adjusting critical parameters for the best possible outcome.

In comparison to manual optimization methods, DOE enables a relatively fast screening of crucial parameters and interactions thereof. Although DOE can be made into quite a complex and powerful tool, in its simplest form, shown here as simpleDOE (sDOE), it can be leveraged quickly, with a lower learning curve, to solve the type of problems faced in the mass spectrometry lab.

In its simplest form, sDOE consists in creating a full-factorial matrix to determine the most relevant parameters and interference between them. In this work for example, some general knowledge of relevant parameters was acquired through 16 experiments, followed by a protein-specific optimization using 10 experiments.

In principle, this sDOE step can be followed by co-optimization of the top two parameters, without necessarily going to the modeling step. For experiments considered as critical or in the context of regulation, additional statistical tests on replicates may be needed to evaluate the significance of the result, although replication on a single set of conditions can be enough to derive measurement errors [34]. Where necessary, further FC optimization towards maximal coverage could be reached by multi-level optimization of relevant parameters followed by modeling of the equation for each parameter's influence. These sDOE-2 steps will be covered in a future publication.

In many cases, sDOE shows great potential for rapid and efficient optimization of experimental conditions and instrumental setup in mass spectrometry and associated techniques.

Acknowledgements

Financial support of European FEDER funds (grant #2699-33931, SyMBioMS), Région Centre Val de Loire (Biocontrol), Université d'Orléans and CNRS for high resolution mass spectrometry is gratefully acknowledged. We thank Dr Bertrand Castaing for the kind gift of HU alpha protein and Dr Kathrin Breuker for stimulating discussion.

References

1. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., & Yates, J. R., 3rd (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*, 113, 2343–2394. doi:10.1021/cr3003533
2. Cui W, Rohrs HW, Gross ML (2011). Top-down mass spectrometry: recent developments, applications and perspectives. *The Analyst*. 136, 3854-3864.
3. Smith, L. M., Agar, J. N., Chamot-Rooke, J., Danis, P. O., Ge, Y., Loo, J. A., Paša-Tolić, L., Tsybin, Y. O., Kelleher, N. L., & Consortium for Top-Down Proteomics (2021). The Human Proteoform Project: Defining the human proteome. *Science Adv.* 7, eabk0734. <https://doi.org/10.1126/sciadv.abk0734>
4. Srzentić K, Fornelli L, Tsybin YO, Loo JA, Seckler H, Agar JN, Anderson LC, Bai DL, Beck A, Brodbelt JS, van der Burgt YEM, Chamot-Rooke J, Chatterjee S, Chen Y, Clarke DJ, Danis PO, Diedrich JK, D'Ippolito RA, Dupré M, Gasilova N, Ge Y, Goo YA, Goodlett DR, Greer S, Haselmann KF, He L, Hendrickson CL, Hinkle JD, Holt MV, Hughes S, Hunt DF, Kelleher NL, Kozhinov AN, Lin Z, Malosse C, Marshall AG, Menin L, Millikin RJ, Nagornov KO, Nicolardi S, Paša-Tolić L, Pengelley S, Quebbemann NR, Resemann A, Sandoval W, Sarin R, Schmitt ND, Shabanowitz J, Shaw JB, Shortreed MR, Smith LM, Sobott F, Suckau D, Toby T, Weisbrod CR, Wildburger NC, Yates JR 3rd, Yoon SH, Young NL, Zhou M. (2020) Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *J Am Soc Mass Spectrom.*, 31, 1783-1802. doi: 10.1021/jasms.0c00036.
5. Gabant G, Boyer A, Cadene M (2016). SSPaQ: A Subtractive segmentation approach for the exhaustive parallel quantification of the extent of protein modification at every possible site. *J. Am. Soc. Mass Spectrom.*, 27, 1328-1343. doi:10.1007/s13361-016-1416-y
6. Sobott F, Watt SJ, Smith J, Edelman MJ, Kramer HB, Kessler BM. (2009) Comparison of CID versus ETD based MS/MS fragmentation for the analysis of protein ubiquitination. *J Am Soc Mass Spectrom.* 20, 1652-1659. doi:10.1016/j.jasms.2009.04.023
7. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA.*, 101, 9528-9533. doi:10.1073/pnas.0402700101
8. Li X, Cournoyer JJ, Lin C, O'Connor PB. (2008) The effect of fixed charge modifications on electron capture dissociation. *J Am Soc Mass Spectrom.* 19, 1514-1526. doi:10.1016/j.jasms.2008.06.014
9. Rose CM, Rush MJ, Riley NM, Merrill AE, Kwiecien NW, Holden DD, Mullen C, Westphall MS, Coon JJ (2015) A calibration routine for efficient ETD in large-scale proteomics. *J Am Soc Mass Spectrom.*, 26, 1848-1857. doi:10.1007/s13361-015-1183-1
10. Zhu W, Li M, Zhang J. (2021) Integrating Intact Mass Analysis and Middle-Down Mass Spectrometry Approaches to Effectively Characterize Trastuzumab and Adalimumab Structural Heterogeneity. *J Proteome Res.*, 20(1):270-278. doi:10.1021/acs.jproteome.0c00373
11. Wollenberg DTW, Pengelley S, Mouritsen JC, Suckau D, Jørgensen CI, Jørgensen TJD (2020) Avoiding H/D Scrambling with Minimal Ion Transmission Loss for HDX-MS/MS-ETD Analysis on a High-Resolution Q-TOF Mass Spectrometer. *Analyt Chem.* 92, 7453-7461. doi:10.1021/acs.analchem.9b05208
12. Cramer CN, Brown JM, Tomczyk N, Nielsen PK, Haselmann KF (2017) Electron Transfer Dissociation of All Ions at All Times, MS^{ETD}, in a Quadrupole Time-of-Flight (Q-ToF) Mass Spectrometer. *J Am Soc Mass Spectrom.* 28, 384-388. doi:10.1007/s13361-016-1538-2
13. Fischer JOI (1935) *The Design of Experiments*. Oliver & Boyd, Edinburgh & London.
14. Cochran WG and Cox GM (1957) *Experimental design*. John Wiley & Sons, New York.
15. Taguchi G and Konishi S (1987) *Taguchi Methods Orthogonal Arrays and Linear Graphs: Tools for Quality Engineering*. American Supplier Institute, Dearborn, Michigan.
16. Myers RH, Montgomery, DC, Anderson-Cook CM (2016) *Response Surface Methodology* (4th ed.). John Wiley & Sons, New York.
17. Shliaha PV, Gibb S, Gorshkov V, Jespersen MS, Andersen GR, Bailey D, Schwartz J, Eliuk S, Schwämmel V, Jensen, ON (2018) Maximizing Sequence Coverage in Top-Down Proteomics By Automated Multimodal Gas-Phase Protein Fragmentation. *Analyt Chem.* 90, 12519-12526. doi:10.1021/acs.analchem.8b02344

18. Atmanene C, Laux A, Glattard E, Muller A, Schoentgen F, Metz-Boutigue MH, Aunis D, Van Dorsselaer A, Stefano GB, Sanglier-Cianféran S, Goumon Y (2009) Characterization of human and bovine phosphatidylethanolamine-binding protein (PEBP/RKIP) interactions with morphine and morphine-glucuronides determined by noncovalent mass spectrometry. *Med Sci Monit.* 15(7),BR178-BR187.
19. Bernier I, Jollès P (1984) Purification and characterization of a basic 23 kDa cytosolic protein from bovine brain. *Biochim Biophys Acta.* 790, 174-181. doi:10.1016/0167-4838(84)90221-8.
20. Hibbert DB (2012) Experimental design in chromatography: a tutorial review. *J Chromatogr B Analyt Technol Biomed Life Sci.* 910, 2-13. doi:10.1016/j.jchromb.2012.01.020
21. Skinner OS, McLafferty FW, Breuker K (2012) How ubiquitin unfolds after transfer into the gas phase. *J Am Soc Mass Spectrom.* 23, 1011-10143. DOI: 10.1007/s13361-012-0370-6
22. Pesavento J, Mizzen CA, Kelleher NL (2006). Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. *Analyt. Chem.* 78, 4271–4280.
23. Seto C, Bateman KP, Gunter B. (2002) Development of generic liquid chromatography-mass spectrometry methods using experimental design. *J Am Soc Mass Spectrom.* 13, 2-9. doi:10.1016/S1044-0305(01)00334-8
24. Switzer L, Giera M, Lingeman H, Irth H, Niessen WM (2011) Protein digestion optimization for characterization of drug-protein adducts using response surface modeling. *J Chromatogr A.* 1218, 1715-1723. doi:10.1016/j.chroma.2010.12.043
25. Zheng H, Clausen MR, Dalsgaard TK, Mortensen G, Bertram HC (2013) Time-saving design of experiment protocol for optimization of LC-MS data processing in metabolomic approaches. *Analyt. Chem.* 85(15):7109-7116. doi:10.1021/ac4020325
26. Zhang L, Borrer CM, Sandrin TR (2014) A designed experiments approach to optimization of automated data acquisition during characterization of bacteria with MALDI-TOF mass spectrometry. *PLoS One.* 9, e92720. doi: 10.1371/journal.pone.0092720.
27. Zhang Z, Dovichi NJ (2018) Optimization of mass spectrometric parameters improve the identification performance of capillary zone electrophoresis for single-shot bottom-up proteomics analysis. *Anal Chim Acta* 1001, 93-99. doi:10.1016/j.aca.2017.11.023
28. Bheemanapally K, Ibrahim MMH, Briski KP (2020) Optimization of Ultra-High-Performance Liquid Chromatography-Electrospray Ionization-Mass Spectrometry Detection of Glutamine-FMOC Ad-Hoc Derivative by Central Composite Design. *Sci Rep.* 10, 7134. doi:10.1038/s41598-020-64099-w
29. Serrano-Blesa E, Porter A, Lendrem DW, et al. (2021) Robust optimization of SWATH-MS workflow for human blood serum proteome analysis using a quality by design approach. *Clin Proteomics.* 18, 20. doi:10.1186/s12014-021-09323-z
30. Pizzala H, Claeys-Bruno M, Monnier V, Sergent M, Charles L (2021) Design of Experiments for Matrix-Assisted Laser Desorption/Ionization of Amphiphilic Poly(Ethylene Oxide)-b-Polystyrene Block Copolymers. *Front Chem.* 9, 740495. doi:10.3389/fchem.2021.740495
31. Afoullouss S, Balsam A, Allcock AL, Thomas OP (2022) Optimization of LC-MS² Data Acquisition Parameters for Molecular Networking Applied to Marine Natural Products. *Metabolites.* 12, 245. doi:10.3390/metabo12030245
32. Frankenfield AM, Fernandopulle MS, Hasan S, Ward ME, Hao L. (2020) Development and Comparative Evaluation of Endolysosomal Proximity Labeling-Based Proteomic Methods in Human iPSC-Derived Neurons. *Analyt Chem.* 92, 15437-15444. doi:10.1021/acs.analchem.0c03107
33. Tibon J, Silva M, Sloth JJ, Amlund H, Sele V (2021) Speciation analysis of organoarsenic species in marine samples: method optimization using fractional factorial design and method validation. *Anal Bioanal Chem.* 413, 3909-3923. doi: 10.1007/s00216-021-03341-4.
34. Hecht ES, Oberg AL, Muddiman DC. (2016) Optimizing mass spectrometry analyses: A tailored review on the utility of design of experiments. *J Am Soc Mass Spectrom.* 27, 767-785. doi:10.1007/s13361-016-1344-x