Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

Cite this: *Anal. Methods*, 2023, **15**, 2512

Factorial design applied to LC-ESI-QTOF mass spectrometer parameters for untargeted metabolomics

Olívia Brito de Oliveira Moreira,[a] Jéssica Cordeiro Queiroz de Souza,[a] João Marcos Beraldo Candido,[a] Maria Patrícia do Nascimento,[a] Paula Rocha Chellini,[ab] Lúcio Marco de Lemos[d] and Marcone Augusto Leal de Oliveira*[ac]

Investigations of untargeted metabolomics are based on high-quality data acquisition usually from multiplatform systems that include high-resolution mass spectrometry equipment. The comprehensive set of results is used as data entry for bioinformatics and machine learning sciences to access reliable metabolic and biochemical information for clinical, forensic, environmental, and endless applications. In this context, design of experiments is a powerful tool for optimizing data acquisition procedures, using a multivariate approach, which enables the maximization of a high-quality amount of information with reduced number of tests. In this study, we applied a 3^3 Box–Behnken factorial design with central point triplicate for optimizing the ionization of an HPLC-ESI-QTOF method used for screening urine samples. Nozzle voltage (V), fragmentor voltage (V) and nebulizer pressure (psig) were the factors selected for variation. The response surface methodology was applied in the molecular features extracted at each level, resulting in a statistical model that helps evaluating the synergic interaction between these factors. Together with the qualitative analysis of the resulting total ion chromatograms, we came across a reproducible (6.14% RSD) and highly efficient method for untargeted metabolomics of human urine samples. The proposed method can be useful for applications in several urine-based metabolomics-driven studies, as the factorial design can be applied in the development of any analytical protocol considering different LC-MS setups.

Received 17th January 2023
Accepted 28th April 2023

DOI: 10.1039/d3ay00094j

rsc.li/methods

1. Introduction

Untargeted metabolomics is now a consolidated and valuable proposition for healthcare related studies when applied to access perturbations in biological systems affected by health anomalies, environmental conditions and a handful of factors that affect human homeostasis. Metabolomics-based studies usually exploit multiplatform systems including high resolution mass spectrometry aiming at a comprehensive set of results to enhance the access of possible biomarkers and other metabolic/biochemical insights, thereby relying on high-quality data acquisition.[1–4]

In this context, design of experiments (DOE) is an efficient process of optimizing procedures in an objective and controlled way with reduced number of tests. By selecting and testing key experimental conditions, a quadratic polynomial model that considers the Response Surface Methodology (RSM) can be accessed to evaluate the interaction between the variables and their influence on the final result, which gives an overview of experimental possibilities so the best condition for any experiment to be performed can be chosen.[1–4]

Box–Behnken factorial design was applied to the optimization of the ionization step of an HPLC-ESI-QTOF method used for screening urine samples for untargeted metabolomics purposes. The 3^3 Box–Behnken is a three-level incomplete factorial design that results in 15 experiments, including a triplicate on the central point[2] (Fig. 1). According to Ferreira *et al.*,[2] it is an advantageous design for RSM application since it allows the estimation of parameters of a quadratic model with the possibility of evaluating its precision and lack of fit with reduced number of experiments.

Nozzle voltage (V), fragmentor voltage (V) and nebulizer pressure (psig) were the electrospray ionization source (ESI) selected parameters for adjustments to perform this

[a]*Grupo de Química Analítica e Quimiometria - GQAQ, Chemistry Department, Institute of Exact Sciences, Federal University of Juiz de Fora (UFJF), 36026-900, Juiz de Fora, MG, Brazil. E-mail: marcone.oliveira@ufjf.br*

[b]*Faculdade de Farmácia, Federal University of Juiz de Fora (UFJF), 36026-900, Juiz de Fora, MG, Brazil*

[c]*National Institute of Science and Technology for Bioanalytics – INCTBio, Institute of Chemistry, University of Campinas (UNICAMP), 13083-970, Campinas, SP, Brazil*

[d]*Lemos Laboratório de Análises Clínicas, 36025-290, Juiz de Fora, MG, Brazil*

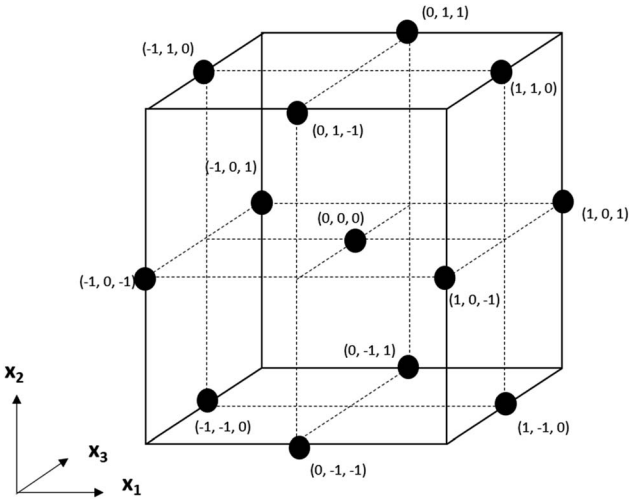

Fig. 1 Cube representation of the 3^3 Box–Behnken factorial design with coded factor levels.

experiment, given that they are one of the relevant factors that have a direct effect on the formation process of all charged molecules that go into the MS analyzer. We performed a qualitative and quantitative evaluation of the outcomes of this experiment, considering both the total ion chromatogram (TIC) profile and the molecular feature count and their chem-informatics applications.[5] The aim of this work was to study and discuss the importance of the analytical phase in metabolomics-associated data science. The proposed method can be applied to the study of any health anomaly desired considering the analysis of urine. Additionally, it can function as a guide for other optimization protocols that may be pursued for other purposes.

2. Materials and methods

2.1 Chemicals and solutions

Acetonitrile, isopropanol and water (HPLC/MS grade) along with formic acid (HPLC grade) and the creatinine (Crn) analytical standard were purchased from Sigma Aldrich (Missouri, USA). Solvents used during sample preparation (acetonitrile, acetone, and methanol ACS grade) were purchased from Sigma Aldrich (Missouri, USA), Neon (Sao Paulo, Brazil), and Dinamica (Sao Paulo, Brazil), respectively.

2.2 Instrumentation

MS experiments were performed in a 1260 Infinity II HPLC equipped with a quaternary pump and a 6530 Accurate-Mass QTOF mass analyzer connected to a Dual Jet Stream Technology Electrospray Ion Source (AJS-ESI) system controlled by MassHunter LC/MS Data Acquisition Software B.08.00, all from Agilent Technologies (Palo Alto, CA, USA). The reversed-phase column Infinity Lab Poroshell 120 EC-C18 column 4.6 × 100 mm 2.7 micron was kept at 30 °C while the multisampler was kept at 10 °C. 5 μL of each sample was injected into the system with solvents A (water and formic acid 99.9 : 0.1, v/v) and B (acetonitrile and formic acid 99.9 : 0.1, v/v) as mobile phases at 0.5 mL min^{-1} in an elution gradient set as follows: 0% of solvent B from 0 to 0.5 min, increasing to 9% up to 2 min, 20% at 5 min, 45% at 8 min, and reaching 100% at 9.5 min where it was held until 11 min. The re-equilibration step was held until 20 min for returning to the initial condition.[6]

Data were initially collected in ESI positive mode operated in full scan at Standard (3200 *m*/*z*) Extended Dynamic Range (2 GHz) with *m*/*z* from 50 to 1500 at 1.02 spectra/s scan rate. Drying gas was kept at 325 °C with 11 L min^{-1} flow rate[6] and sheath gas at 350 °C with 11 L min^{-1} flow rate as well. Capillary voltage was fitted at 4000 V. Nebulizer pressure, fragmentor, and nozzle voltage were the chosen parameters for the experimental design and will be described in the following sections. The reference *m*/*z* enabled were 121.0509 (purine, $C_5H_4N_4$) and 922.0098 (hexakis(1*H*,1*H*,3*H*-tetrafluoropropoxy)phosphazine, $C_{18}H_{18}O_6N_3P_3F_{24}$) for ESI (+); 119.0363 (purine, $C_5H_3N_4$) and 301.9981 (tris(2,4,6-trifluoromethyl)-1,3,5-triazine, $C_6N_3F_9OH$) for ESI (−). Data acquisition on negative mode was performed by following the same optimized condition set for the positive mode.

2.3 Samples

A pool of 10 human urine samples was used for this optimization study. The samples were collected upon authorization of the Ethics Committee from the University Hospital of the Federal University of Juiz de Fora (acceptance codes: 4.473.404; 4.566.092; 5.039.371) and the Ethics Committee from the Federal University of Juiz de Fora (acceptance code 5.267.924). The participants were assigned as random people above 18 years old that formally agreed to participate in this study by signing the Free and Informed Consent Form and filling up the anamnesis questionnaire. The donations were done with no preparation requirements at random periods of the day.

Once collected, the samples went through a triage process at which 10% (v/v) of acetone/methanol (6 : 4, v/v)[7] was added for metabolic quenching. After sealing and encoding, the samples were kept in an ultrafreezer at −50 °C. On the day of analysis, the samples were thawed at room temperature for approximately 20 min. 10% (v/v) of ice-cold acetonitrile was added, followed by vortex agitation for 10 s and centrifugation at 7500 rpm for 10 minutes to complete the protein precipitation protocol. The supernatant of each sample was mixed together and the resulting pool was transferred to the vial without further processing.

2.4 Experimental design

The 15 experiments required in the 3^3 Box–Behnken design were performed in a randomly organized sequence applied to the analysis of a urine pool. The ESI (+) parameters used as variables of this study were set in three levels: nozzle voltage 0 (−), 250 (0), 500 (+); nebulizer pressure 30 (−), 40 (0), 50 (+); fragmentor voltage 100 (−), 175 (0), 325 (+). Before each run, mobile phase-only runs were applied as a system cleanup. The ESI source was cleaned with isopropanol HPLC/MS grade before analysis.

2.5 Data analysis and statistics

The resulting Total Ion Chromatograms (TICs) were submitted to the extraction workflow considering compound discovery by molecular features followed by Base Peak Chromatogram (BPC) extraction. Information on creatinine peak extraction was obtained with automatic integration of Extracted Ion Chromatograms (EICs) considering the m/z of $[M + H]^+$ 144.0669 at a retention time (Rt) of 1.8 min previously accessed by the analysis of a creatinine standard aqueous solution at 10 mg L^{-1}. Total areas of TIC in arbitrary units (AUs) were calculated through automatic integration by using the Agile 2 algorithm. All aforementioned calculations were performed using MassHunter Qualitative Analysis B.10.00 software (Agilent Technologies). No filters or other preprocessing protocols were applied.

The data of extracted molecular features were then analyzed by RSM considering a second-order polynomial function (eqn (1)):

$$\hat{y}_i = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j=2,\ i<j<k}^{k} \beta_{ij} x_i x_j + \varepsilon_i \quad (1)$$

where k is the number of variables, x_i and x_j are the factors (independent variables), β represents the coefficient of each term, those being, β_0 the constant term, β_i the first order term, β_{ij} the quadratic ($i = j$) and the interaction ($i < j < k$, and $k = 3$) terms. The residual associated with the experiment is expressed by ε_i. Microsoft Excel® 2019 was used for the polynomial model fitting, lack of fit test, and for building the response surface graphic.

3. Results and discussion

3.1 RSM and method optimization

Two outcomes were considered for a complete evaluation of the factorial design proposed. At first, a qualitative response of the TIC resulting from each experiment (Fig. 2) was accessed. This approach has been considered since TIC or Extracted Ion Chromatogram (EIC) is nowadays being used as a promising image data entry for pattern recognition analysis as artificial neural networks (ANNs) and other deep learning (DLNs) and Machine Learning (MLs) possibilities.[8–12]

Experiments 7, 8, 11, and 12 were immediately discarded as final method candidates given their low intensity and poor peak resolution, as a consequence of a high fragmentor voltage. We have chosen a wide range of this particular ESI parameter aiming at a comprehensive overview of its influence on the final result, since it has to be set in a value that fragments the molecule sufficiently to be analyzed but not so much that it destroys it, as it probably happened with the creatinine molecule under the condition of the highest fragmentor voltage, for example (Table 1). In this context, the opposite level of this factor, applied in experiments 5, 6, 9, and 10 results in an enhanced chromatogram regarding intensity, however, not better than the mid-factor, 175 V, described in experiments 1–4 and the central point (Fig. 2).

Among the pre-selected promising outcomes, the experiments where nozzle voltage is applied have slight variations from the ones that did not, considering peak resolution, count intensity, and the amount of information registered within the elution region of the organic phase increasing, between minutes 5 and 10. The same goes for the variations in the nebulizer pressure factor (Table 1). We observe that the MFE within the selected 5 minute window, where there are visual variations on the TIC profile, gave better results with fragmentor 175 V and no nozzle voltage.

Response surface methodology was applied considering the number of extracted molecular features of each experiment (Table 1). Since the number of molecular features strictly express the number of distinct compounds found, higher the amount of information, better the contribution for enhancing further acquisitions, considering the chemometrics involved in the metabolomics workflow. At the first glance, the number of

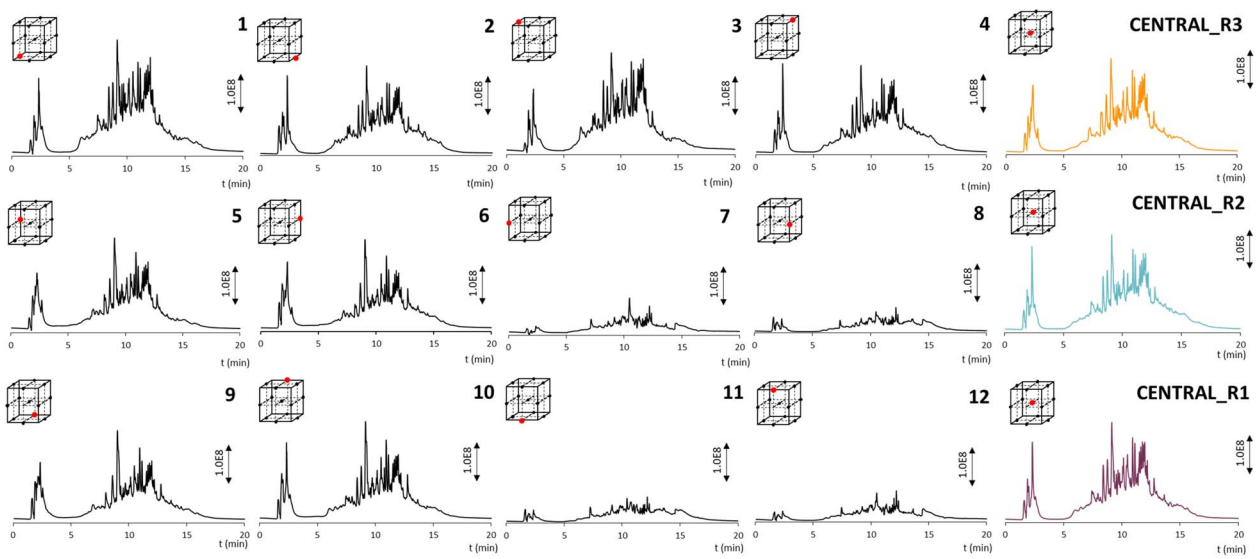

Fig. 2 TIC results from each experiment of the Box–Behnken factorial design.

Table 1 Number of molecular features extracted (MFE) and considerations of intensity/area by creatinine (Crn) peak measurements from results of each experiment of the 3^3 Box–Behnken factorial design[a]

Exp.	X_1	X_2	X_3	MFE total	MFE (Rt 5–10 min)	TIC total area (AU)	Creatinine (Crn) Area (AU)	Creatinine (Crn) Peak height	Creatinine (Crn) Rt (min)
1	0	30	175	3917	1388	1.27×10^{10}	0.56×10^{7}	1.64×10^{6}	1.866
2	500	30	175	4002	1266	0.49×10^{10}	1.30×10^{7}	3.70×10^{6}	1.881
3	0	50	175	3905	1481	1.23×10^{10}	0.58×10^{7}	1.55×10^{6}	1.885
4	500	50	175	3346	1008	0.48×10^{10}	1.15×10^{7}	3.67×10^{6}	1.886
5	0	40	100	3488	1122	0.81×10^{10}	0.41×10^{7}	1.22×10^{6}	1.850
6	500	40	100	2974	1029	0.37×10^{10}	1.29×10^{7}	3.64×10^{6}	1.848
7	0	40	325	823	269	0.11×10^{10}	—	—	—
8	500	40	325	640	190	0.13×10^{10}	—	—	—
9	250	30	100	3013	1022	0.44×10^{10}	0.83×10^{7}	2.59×10^{6}	1.847
10	250	50	100	3693	1065	0.70×10^{10}	1.05×10^{7}	2.59×10^{6}	1.863
11	250	30	325	801	240	0.10×10^{10}	—	—	—
12	250	50	325	751	228	0.19×10^{10}	—	—	—
0_R1	250	40	175	3482	1187	0.79×10^{10}	0.84×10^{7}	2.98×10^{6}	1.866
0_R2	250	40	175	3773	1067	0.73×10^{10}	0.81×10^{7}	2.68×10^{6}	1.860
0_R3	250	40	175	3973	1284	0.93×10^{10}	0.94×10^{7}	2.99×10^{6}	1.875

[a] X_1: nozzle voltage (V); X_2: nebulizer pressure (psig); X_3: fragmentor voltage (V); Rt = retention time. Statistics of the central point: MFE total – mean value: 3742 ± 246 RSD: 6.59%; MFE 5–10 min – mean value: 1179 ± 109. RSD 9.22%; TIC total – mean value: $0.82 \times 10^{10} \pm 0.10 \times 10^{10}$ RSD: 12.5%; Crn area (AU) – mean value: $0.86 \times 10^{7} \pm 0.065 \times 10^{7}$ RSD: 7.48%; Crn height – mean value: $2.86 \times 10^{6} \pm 0.158 \times 10^{6}$ RSD: 5.52%. Rt Crn – mean value: 1.87 ± 0.0075 RSD 0.40%.

molecular features found by the software for each experiment agreed with the qualitative observations aforementioned.

Based on the resulting coefficient values (Table 2), fragmentor voltage (V) was the one factor that had the most influence on the final result (p-value < 0.05), followed by the first factor, nozzle voltage, that influenced the system approximately 10 times less than the fragmentor voltage. According to the resulting coefficients, the nebulizer pressure is not influencing the system as much as the other factors within the levels considering for the DOE. Therefore, the nebulizer pressure was not considered as relevant in further evaluations. RSM was applied as an overview approach to access the interaction between the nozzle and the fragmentor voltage.

Even though any of the combinatory factors that consider two distinct variables were mathematically significant, given that voltage application is a crucial factor for direct molecular ionization, we have decided to build the RSM model for the graphical analysis on the slight contributions of using nozzle voltage for urine analysis. For a quantitative evaluation of the TIC profile, the total area was calculated together with the EIC information of one of the main components of urine, creatinine.[13] According to the data described in Table 1, there is an interaction resemblance in the values of TIC area and peak area/intensity of creatinine according to the combination of nozzle and fragmentor voltage. Besides, the mathematical approach can be interesting for optimizing the ionization of samples that present a greater amount of non-polar chemicals in their constitution, such as blood parts for example, as will be discussed in further sections.

Table 2 Coefficients and statistics resulting from the quadratic model

Coefficients		Standard error	Stat t	p-Value
β_0	3742.7	142.5	26.2	0.001448
β_1	−146.4	87.3	−1.68	0.235572
β_2	−4.75	87.3	−0.05	0.961552
β_3[a]	−1269.1	87.3	−14.5	0.004698
β_{11}	−16.7	128.5	−0.13	0.908438
β_{22}	66.5	128.5	0.52	0.656142
β_{33}[a]	−1744.7	128.5	−13.6	0.00538
β_{12}	−161.0	123.4	−1.30	0.322075
β_{13}	82.7	123.4	0.67	0.571696
β_{23}	−182.5	123.4	−1.48	0.277399

[a] Significant coefficients (p-value <0.05).

The built model was adjusted, considering a 95% confidence interval, presented no lack of fit (p-value = 0.48, higher than 0.05) and a significant regression (p-value = 7.82×10^{-6}, lower than 0.05) with 99.9% of the maximum of variation explained (Table 3). The calculated precision of the DOE resulted in an RSD of 6.14% considering the number of molecular features of the central point triplicate.

According to the RS (Fig. 3), experiments that consider the (−1) and (0) levels of fragmentor voltage are within the optimal region of the ESI conditions, considering the enhanced extraction of molecular features. Taking this into consideration, experiments 1, 3, 5, 9, 10, and the central point appear among the optimal ones. However, along with our previous qualitative evaluation of the results, experiments 1, 3 and central present as a better fit considering the TIC profile. Altogether, among these options, experiment 3 was selected as the optimized method. All values extracted are statistically the same as experiment 1, considering the associated RSD; but given the previous visual evaluations of better peak definition and resolution, especially over the beginning of the elution gradient, in addition to good area and intensity values of EIC/TIC, it was defined finally as the better fit. It is important to highlight that not using nozzle

Table 3 Variance analysis associated with the adjustments of the quadratic model[a]

Variation factors	Sum of squares	Number of degrees of freedom	Mean of squares
Regression	120 790 957.07	9	13 421 217.45
Residues	341 740.42	5	68 348.08
Lack of fit	219 819.75	3	73 273.25
Pure error	121 920.67	2	60 960.33
Total	121 132 697.5	14	—

[a] Variation explained = 99.7%; maximum of variation explained = 99.9%.

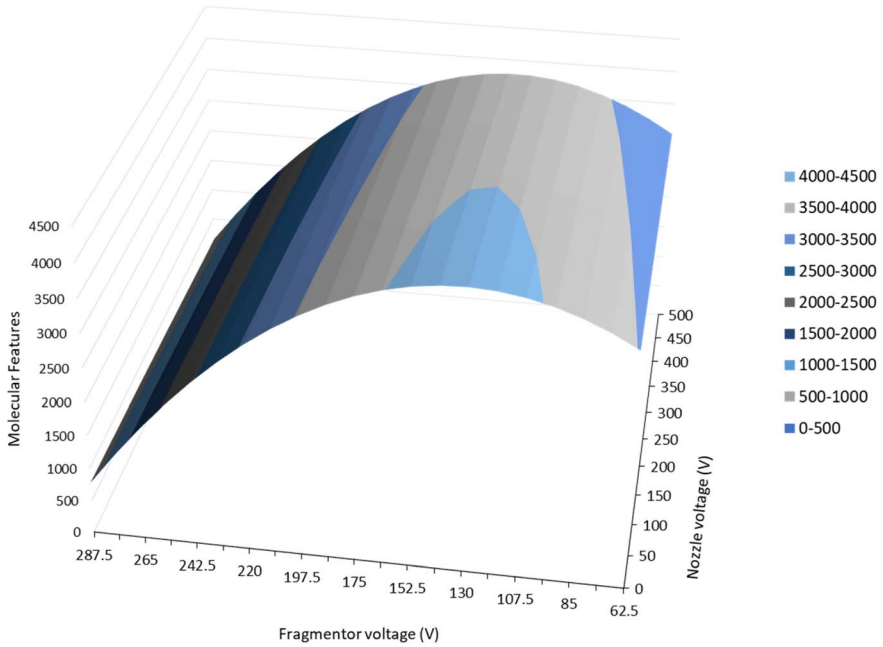

Fig. 3 Response surface of the interaction between the nozzle and fragmentor voltage considering the number of molecular features generated by the application of each combination of factors.

voltage is an interesting advantage, so any equipment can be used to reproduce this condition since not all MS systems inherit this accessory.

3.2 Method application

The optimized method was further applied as untargeted metabolomics analytical protocol on 100 urine samples. During the analysis, 10 quality control (QC) runs were performed considering the pool of these 100 urine samples (Fig. 4A). The same method was also applied in ESI negative mode (Fig. 4B). By submitting the results to a molecular extraction workflow, mean values of 3809 and 3868 of molecular features were found, with 2.27% and 7.29% RSD for positive and negative modes, respectively. The variation in the negative mode is expected, since the ionization in this mode can be troublesome; however, a 7.29% RSD when biological matrices are in focus is not doomed. According to the ICH (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use), chromatographic-based bioanalytical methods that present variations up until 20% are completely acceptable.[14]

This method can be applied to any urinary sample for whatever investigation it may be pursued, which means that these adjustments can be used for metabolomics-driven studies that consider any disease or healthy anomaly. The proposed protocol was also applied to human blood serum and saliva samples (Fig. 5) with promising outcomes; however, in these cases the new application of the DOE might be on hand. Saliva and blood serum are samples with higher viscosity, so a dilution step shall be studied; besides, these samples have a significant amount of non-polar components as hormones, steroids, fatty acids, vitamins, and more. So, in this case we might find out that the use of a high nozzle voltage (Fig. 5) and other levels of fragmentor voltage together with adjustments in temperature, pressure and other parameters is more adequate, since those molecules are harder to be ionized.

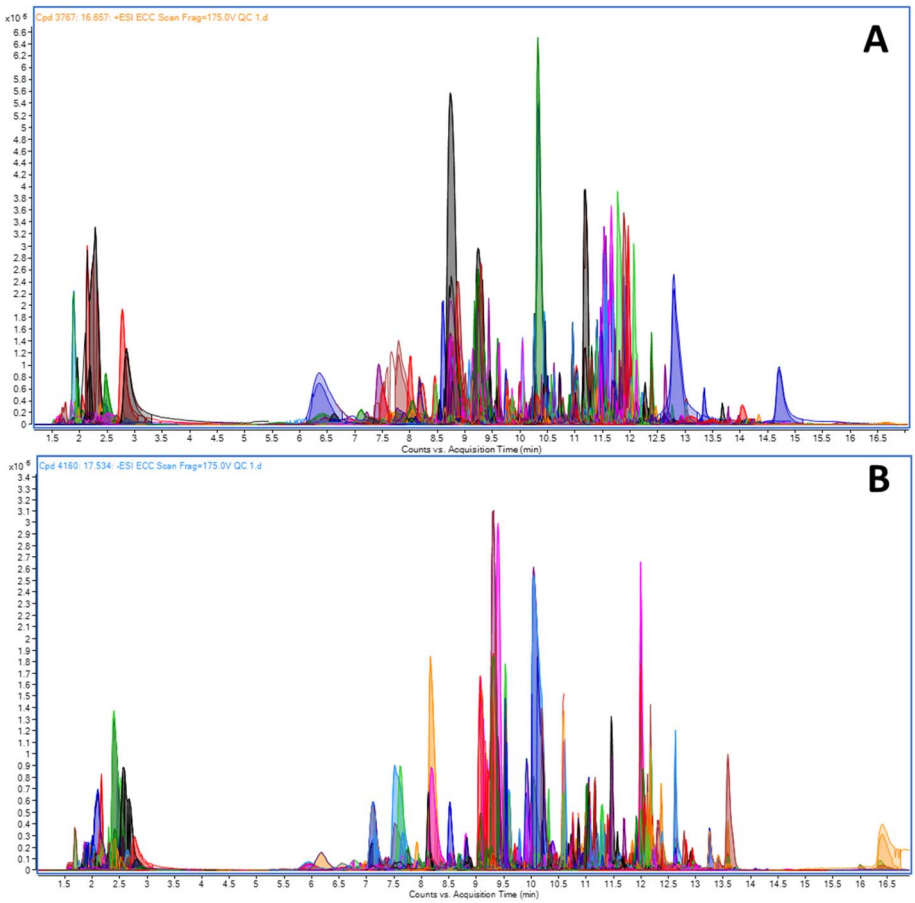

Fig. 4 Base peak chromatogram extracted from a QC of 100 urine samples submitted to the optimized method in ESI (+) mode (A) and ESI (−) mode (B).

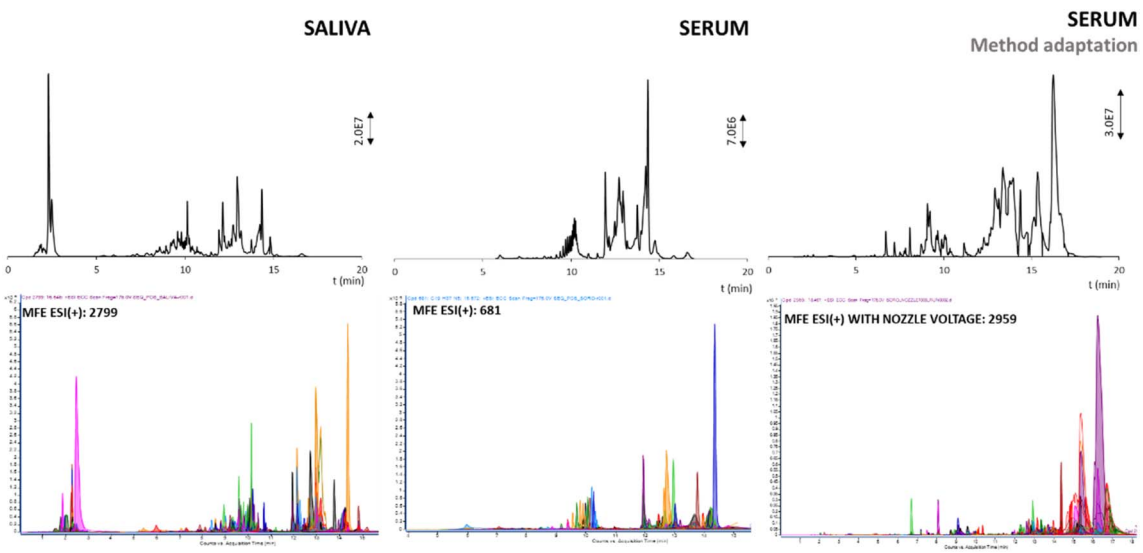

Fig. 5 TIC and BPC data from application of the method on saliva and serum samples. Method adaptation to the serum sample: use of 1000 V nozzle voltage.

Table 4 Compilation of optimization-driven studies towards protocol enhancement for metabolomics applied to human urine samples[a]

Equipment	Investigation	Urine sample preparation	HPLC parameters	ESI parameters	MS parameters	MFE	Ref.
nUHPLC-nESI-TOF	Pre-analysis osmolality optimization	Dilution, SPE, reconstitution in methanol : water (10 : 90, v/v)	RP (C18) column: Waters nanoAcquity HSS-T3 (100 mm × 100 m × 2.8 m, 100 Å); mobile phases: A (water : formic acid 99.99 : 0.01 v/v) and B (ACN : formic acid 99.99 : 0.01 v/v) at 700 nL min^{-1}. 55 min multistep gradient	Nano (+) ESI	Mass resolution of 20 000. MS mode at 10 eV followed by an energy ramp of 20–40 eV	n.m.	15
UHPLC-LTQ-Orbritap	2D-LC/MS/MS and LC-MS method optimization	Protein precipitation, vacuum drying, resuspension in ACN, filtration	RP (C18) column: Waters HSS. Dimensions not mentioned. Mobile phases: A (water : formic acid 99.99 : 0.01 v/v) and B (ACN at 0.3 mL min^{-1}). 29 min multistep gradient. 50 °C	n.m	Full scan mode: 100–1000 *m/z* at 60 000 resolution	2357	16
HPLC-microQTOF	Optimization of data processing	Centrifugation and dilution in water	RP (C18) column: Agilent ZORBAX 300SB (0.5 × 150 mm, 3.5 μm). Mobile phase: ACN with 0.1% formic acid. 15 minute gradient	(−) ESI. Capillary: + 3400 V; nebulizer: 0.4 bar; drying gas 4.0 L min^{-1}, 180 °C	Full scan mode: 40–1000 *m/z*	n.m.	17 and 18
UHPLC-ESI-HRMS	Optimization of sample preparation	Centrifugation, extraction, vacuum drying	Method A – RP (C18) column Phenomenex (1.7 μm × 100 Å × 150 × 2.1 mm). Mobile phases: A (water : formic acid 99.9 : 0.1 v/v) and B (methanol : formic acid 99.9 : 0.1 v/v) at 0.4 mL min^{-1}. 24 min gradient. 55 °C Method B – HILIC column: Waters Cortecs (unbonded silica; 1.6 μm × 100 Å × 150 × 2.1 mm). Mobile phases: A (water) and B (ACN with 5% water) both with 10 nM ammonium formate and 0.5% of formic acid at 0.3 mL min^{-1}. 27 min multistep gradient. 40 °C	(+)(−) ESI. Spray voltage: + 3500 V; capillary temperature 325 °C; heater 350 °C; sheath gas 50 AU, auxiliary gas 13 AU, sweep gas 3 AU.	Full scan mode: 58–870 *m/z*. 70 000 resolution	n.m.	19
UHPLC-DualESI-QTOF	Optimization of sample preparation	Centrifugation and dilution in water	RP (C18) column: Agilent Zorbax Eclipse-C18 (2.1 × 50 mm × 1.8 μm). Mobile phases: A (water : formic acid 99.9 : 0.1 v/v) and B (methanol : formic acid 99.9 : 0.1 v/v) at 0.3 mL min^{-1}. 12 min multistep gradient. 35 °C	(+)(−) ESI. Capillary: + 3000 V; drying gas: 10 L min^{-1}, 300 °C; 40 psig; fragmentor 140 V	Full scan mode: 70–1050 *m/z*. 1.2 spectra per s	286 (+) 231 (−)	20
HPLC-ESI-QTOF	Method validation	Centrifugation, extraction, and dilution	RP (C18) column: Phenomenex (3 μm × 30 × 2 mm). Mobile phases: A (water : formic acid 99.9 : 0.1 v/v) and B (ACN : formic acid 99.9 : 0.1 v/v) at 0.3 mL min^{-1}. 18 min multistep gradient	(+) ESI. Capillary: + 2500 V; drying gas: 5 L min^{-1}, 325 °C; 60 psig; sheath gas: 300 °C, 12 L min^{-1}. Nozzle voltage: 1200 V; fragmentor 170 V	Full scan mode: 100–900 *m/z*. 5 spectra per s	n.m.	21
HPLC-ESI-QTOF	Method application	Centrifugation and dilution in water	RP (C18) column: Supelco Ascentis Express (5 cm × 2.1 mm, 2.7 m, 90 Å). Mobile phases: A (water : formic acid 99.5 : 0.5 v/v) and B (acetonitrile : formic acid 99.5 : 0.5 v/v) at 0.5 mL min^{-1}. 20 min multistep gradient. 30 °C	(+) ESI. Capillary: + 4000 V; drying gas: 11 L min^{-1}, 325 °C; 50 psig	Full scan mode: 50–1000 *m/z*. 1.02 spectra per s	684[b]	6

Table 4 (*Contd.*)

Equipment	Investigation	Urine sample preparation	HPLC parameters	ESI parameters	MS parameters	MFE	Ref.
HPLC-ESI-QTOF	Optimization of ionization conditions	Protein precipitation	RP (C18) column: Infinity Lab Poroshell 120 EC (4.6 × 100 mm × 2.7 micron). Mobile phases: A (water and formic acid 99.9 : 0.1, v/v) and B (acetonitrile and formic acid 99.9 : 0.1, v/v) at 0.5 mL min^{-1}. 20 min multistep gradient. 30 °C	(−)(+) ESI. Capillary: + 4000 V; drying gas: 11 L min^{-1}, 325 °C; 50 psig. Sheath gas: 350 °C, 11 L min^{-1}. Fragmentor 175 V	Full scan mode: 50–1500 *m/z*. 1.02 spectra per s	3809 (+) 3868 (−)	This work

[a] ACN: acetonitrile; AU: arbitrary units; MFE: molecular features extracted; n.m: not mentioned. SPE: solid phase extraction. [b] Results after entire statistical processing.

4. Conclusions

This study shows the importance of the analytical phase to obtain reliable and substantial amount of information that will be further explored through chemometric tools to access a biological response of the anomaly under investigation. An effective way to improve the outcomes of any analysis is applying design of experiments together with the Response Surface Methodology, since it allows an overview of how the factors chosen influence the system, thereby helping the decision-making process. For ESI studies, we considered three of the several factors that influence the ionization and the molecular flow that goes into the MS system. Among them, fragmentor voltage was the one factor that made a difference in the final result. This parameter is crucial to the formation of the ionized molecule that goes into the MS analyzer, so it makes sense that it needs optimization. For any other application, regarding the aim of the analysis or the sample that is being analyzed, adjustments of this parameter should be considered, in a way that the molecules are ionized enough to be analyzed but not so much that they lose their characteristics. Altogether, DOE and RSM can be applied to a combination of several other factors such as temperature and flow rate of drying, sheath or auxiliary gas, capillary voltage, and more, as needed.

Metabolomics is today a highly explored workflow to access molecular responses of clinical conditions; therefore, the publications are mainly focused on applications rather than method optimization. Nevertheless, some studies were performed over the last 10 years to optimize the basic steps of metabolomics, *i.e.*, sample preparation, mass spectra acquisition, separation technique method optimization and more (Table 4). All the optimization procedures can be studied and combined aiming at a full enhanced protocol. In our work, we have focused on the optimization of the parameter of ionization, a step of the analytical process that will define the data quality for further statistical processing. If the analytical phase fails in gathering comprehensive information, further procedures can be impaired.

In comparison with the other optimization-driven studies, our protocol provides an easy sample preparation procedure together with an enhancement of the raw collected dataset (Table 4). For analyses that were performed in similar equipment to ours, the proposed method provided improved information. Besides, this protocol can be adapted to other systems. The novelty of our work was to apply experimental design to the optimization of ESI parameters that result in a method with better outcome regarding the molecular feature extraction and TIC profile. Overall, our aim was to demonstrate that slight variations can produce a better set of data, even with high-resolution and high-technology mass spectrometers and show that DOE is an excellent mathematical tool for method optimization for any LC-MS system.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

The authors acknowledge the Coordination of Superior Level Staff Improvement – Brazil (CAPES) – Finance Code 001; Brazilian National Council for Scientific and Technological Development (CNPQ: 424032/2018-0, 303867/2020-5, Scholarship: 142502/2020-0); Funding Authority for Studies and Projects (Finep) (CT-INFRA 01/2013-REF 0633/13); RQ-MG (CEX.RED- 00010-14); National Institute of Science and Technology for Bioanalytics (INCTBio) (FAPESP 2014/50867-3, CNPq 465389/2014-7, 401256/2020-0); State of Minas Gerais Research Support Foundation (FAPEMIG) (APQ-00585-21). The authors also thank the team of collaborators from *Lemos Laboratório de Análises Clínicas* for the collection and donation of samples.

References

1 R. Bruns, I. Scarminio and B. de B. Neto, *Statistical Design – Chemometrics*, Elsevier, 1st edn, 2006.

2 S. L. C. Ferreira, R. E. Bruns, H. S. Ferreira, G. D. Matos, J. M. David, G. C. Brandão, E. G. P. da Silva, L. A. Portugal, P. S. dos Reis, A. S. Souza and W. N. L. dos Santos, Box-Behnken design: An alternative for the optimization of analytical methods, *Anal. Chim. Acta*, 2007, **597**, 179–186, DOI: **10.1016/j.aca.2007.07.011**.

3 Y. Zhou, J. Z. Song, F. F. K. Choi, H. F. Wu, C. F. Qiao, L. S. Ding, S. L. Gesang and H. X. Xu, An experimental design approach using response surface techniques to obtain optimal liquid chromatography and mass spectrometry conditions to determine the alkaloids in Meconopsi species, *J. Chromatogr. A*, 2009, **1216**, 7013–7023, DOI: **10.1016/j.chroma.2009.08.058**.

4 D. Veličković, G. Zhang, D. Bezbradica, A. Bhattacharjee, L. Paša-Tolić, K. Sharma, T. Alexandrov and C. R. Anderton, Response Surface Methodology As a New Approach for Finding Optimal MALDI Matrix Spraying Parameters for Mass Spectrometry Imaging, *J. Am. Soc. Mass Spectrom.*, 2020, **31**, 508–516, DOI: **10.1021/jasms.9b00074**.

5 M. Vaškevičius, J. Kapoči, L. Šlepikas, J. Lucas Bao and J.-B. Tristan, Prediction of Chromatography Conditions for Purification in Organic Synthesis Using Deep Learning, *Molecules*, 2021, **26**, 2474, DOI: **10.3390/molecules26092474**.

6 J. V. Alberice, A. F. S. Amaral, E. G. Armitage, J. A. Lorente, F. Algaba, E. Carrilho, M. Márquez, A. García, N. Malats and C. Barbas, Searching for urine biomarkers of bladder cancer recurrence using a liquid chromatography-mass spectrometry and capillary electrophoresis-mass spectrometry metabolomics approach, *J. Chromatogr. A*, 2013, **1318**, 163–170, DOI: **10.1016/j.chroma.2013.10.002**.

7 M. Widera, S. Westhaus, H. F. Rabenau, S. Hoehl, D. Bojkova, J. Cinatl and S. Ciesek, Evaluation of stability and inactivation methods of SARS-CoV-2 in context of laboratory settings, *Med. Microbiol. Immunol.*, 2021, **210**, 235–244, DOI: **10.1007/s00430-021-00716-3**.

8 A. B. Risum and R. Bro, Using deep learning to evaluate peaks in chromatographic data, *Talanta*, 2019, **204**, 255–260, DOI: **10.1016/J.TALANTA.2019.05.053**.

9 A. Mandalapu, S. Calzadilla and S. Mondesire, Image Classification of High-Performance Liquid Chromatography Chromatograms with Neural Networks, *Proc. - 2020 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2020*, 2020, pp. 1647–1652, DOI: **10.1109/CSCI51800.2020.00303**.

10 S. E. Reichenbach, C. A. Zini, K. P. Nicolli, J. E. Welke, C. Cordero and Q. Tao, Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition, *J. Chromatogr. A*, 2019, **1595**, 158–167, DOI: **10.1016/j.chroma.2019.02.027**.

11 C. Jirayupat, K. Nagashima, T. Hosomi, T. Takahashi, W. Tanaka, B. Samransuksamer, G. Zhang, J. Liu, M. Kanai and T. Yanagida, Image Processing and Machine Learning for Automated Identification of Chemo-/Biomarkers in Chromatography-Mass Spectrometry, *Anal. Chem.*, 2021, **93**, 14708–14715, DOI: **10.1021/acs.analchem.1c03163**.

12 J. Guo, S. Shen, S. Xing, Y. Chen, F. Chen, E. M. Porter, H. Yu and T. Huan, EVA: Evaluation of Metabolic Feature Fidelity Using a Deep Learning Model Trained with over 25000 Extracted Ion Chromatograms, *Anal. Chem.*, 2021, **93**, 12181–12186, DOI: **10.1021/acs.analchem.1c01309**.

13 O. B. O. Moreira, J. C. Q. de Souza, J. M. B. Candido, M. P. do Nascimento, P. R. Chellini, L. M. de Lemos and M. A. L. de Oliveira, Determination of creatinine in urine and blood serum human samples by CZE-UV using on-column internal standard injection, *Talanta*, 2023, **258**, 124465, DOI: **10.1016/j.talanta.2023.124465**.

14 European Medicines Agency, ICH Guideline M10 on Bioanalytical Method Validation, *Sci. Med. Health*, 2019, **44**, 57, **https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-m10-bioanalytical-method-validation-step-5_en.pdf%0A**, **https://www.ema.europa.eu/en/ich-m10-bioanalytical-method-validation#current-version-section**.

15 A. J. Chetwynd, A. Abdul-Sada, S. G. Holt and E. M. Hill, Use of a pre-analysis osmolality normalisation method to correct for variable urine concentrations and for improved metabolomic analyses, *J. Chromatogr. A*, 2016, **1431**, 103–110, DOI: **10.1016/j.chroma.2015.12.056**.

16 J. Xu, S. Zheng, M. Li, X. Liu, H. Sun, Z. Guo, J. Wei, L. Jia and W. Sun, A Comprehensive 2D-LC/MS/MS Profile of the Normal Human Urinary Metabolome, *Diagnostics*, 2022, **12**, 2184, DOI: **10.3390/diagnostics12092184**.

17 X. Zhang, M. R. Clausen, X. Zhao, H. Zheng and H. C. Bertram, Enhancing the power of liquid chromatography-mass spectrometry-based urine metabolomics in negative ion mode by optimization of the additive, *Anal. Chem.*, 2012, **84**, 7785–7792, DOI: **10.1021/ac3013835**.

18 H. Zheng, M. R. Clausen, T. K. Dalsgaard, G. Mortensen and H. C. Bertram, Time-saving design of experiment protocol for optimization of LC-MS data processing in metabolomic

approaches, *Anal. Chem.*, 2013, **85**, 7109–7116, DOI: **10.1021/ac4020325**.

19 C. Martias, N. Baroukh, S. Mavel, H. Blasco, A. Lefèvre, L. Roch, F. Montigny, J. Gatien, L. Schibler, D. Dufour-Rainfray, L. Nadal-Desbarats and P. Emond, Optimization of sample preparation for metabolomics exploration of urine, feces, blood and saliva in humans using combined nmr and uhplc-hrms platforms, *Molecules*, 2021, **26**, 4111, DOI: **10.3390/molecules26144111**.

20 R. Skibiński and Ł. Komsta, Optimization of data acquisition and sample preparation methods for LC-MS urine metabolomic analysis, *Open Chem.*, 2015, **13**, 763–768, DOI: **10.1515/chem-2015-0096**.

21 N. Dasí-Navarro, M. Lozano, S. Llop, A. Esplugues, A. Cimbalo, G. Font, L. Manyes, J. Mañes and P. Vila-Donat, Development and Validation of LC-Q-TOF-MS Methodology to Determine Mycotoxin Biomarkers in Human Urine, *Toxins*, 2022, **14**, 1–14, DOI: **10.3390/toxins14100651**.