


Santiago Codesido<sup>1,2</sup>   
 Nicolas Drouin<sup>1,2†</sup>   
 Sabrina Ferré<sup>1,2</sup>   
 Julie Schappler<sup>1,2</sup>   
 Serge Rudaz<sup>1,2,3</sup>   
 Víctor González-Ruiz<sup>1,2,3</sup> 

<sup>1</sup>Department of Analytical Sciences, Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland

<sup>2</sup>School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland

<sup>3</sup>Swiss Centre for Applied Human Toxicology, Basel, Switzerland

Received November 12, 2020

Revised June 18, 2021

Accepted June 24, 2021

## Research Article

# New insights into the conversion of electropherograms to the effective electrophoretic mobility scale

CE–MS is increasingly gaining momentum as an analytical tool in metabolomics, due to its ability to obtain information about the most polar elements in biological samples. This has been helped by improvements of robustness in peak identification by means of mobility-scale representations of the electropherograms (mobilograms). As a necessary step toward facilitating the use of CE–MS for untargeted metabolomics data, the authors previously developed and introduced ROMANCE, a software automating mobilogram generation for large untargeted datasets through a simple and self-contained user interface. Herein, we introduce a new version of ROMANCE including new features such as compatibility with other types of data (targeted MS data and 2D UV-Vis absorption-like electropherograms), and the much needed additional flexibility in the transformation parameters (including field ramping and the use of secondary markers), more measurement conditions (depending on detection and integration modes), and most importantly tackling the issue of quantitative peak conversion. First, we present a review of the current theoretical framework with regard to peak characterization, and we develop new formulas for multiple marker peak area corrections, for anticipating peak position precision, and for assessing peak shape distortion. Then, the new version of the software is presented and validated experimentally. We contrast the multiple marker mobility transformations with previous results, finding increased peak position precision, and finally we showcase an application to actual untargeted metabolomics data.

### Keywords:

Area correction / Compound identification / Electrophoretic mobility / Fundamentals / Quantitation  
 DOI 10.1002/elps.202000333



Additional supporting information may be found online in the Supporting Information section at the end of the article.

## 1 Introduction

Due to the different nature of analytes present in biological systems, a panel of complementary separation and detection techniques is needed to increase the amount of information retrieved in metabolomic studies. In the case of CZE, the mechanism driving the separation makes it particularly well suited for the analysis of polar and charged molecules [1]. Even though a significant fraction of the most commonly studied metabolites falls within this category, the use of CE still remains quite limited in the field of metabolomics [2, 3].

**Correspondence:** Dr. Víctor González-Ruiz, Department of Analytical Sciences, Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, 1205 Geneva, Switzerland. E-mail: victor.gonzalez@unige.ch

**Abbreviations:** IL-1, interleukin 1; LPS, lipopolysaccharide; PCA, principal component analysis; PQN, probabilistic quotient normalization; QC, quality control; TNF, tumor necrosis factor

Compared to other techniques such as LC or GC, where the retention of the analyte can be used to identify the unknowns along with other properties such as the accurate mass or the fragmentation pattern, the use of migration times in CE presents the disadvantage of being a much less robust parameter. Relative migration times can be useful, but they are unable to adequately cope with the effect of constant parameters such as the application of a pressure to assist the EOF during the separation. Effective electrophoretic mobility is a much more robust option, and it constitutes a better alternative for the identification of metabolites since it is a molecular property, which depends only on the nature of the chosen BGE and separation temperature [4–6].

In spite of its convenience, calculation of electrophoretic mobilities is a tedious process, and so we previously

**Color online:** See article online to view Figs. 1–4 in color.

<sup>†</sup>Current address: Leiden Academic Centre for Drug Research, University of Leiden, 2311 EZ Leiden, The Netherlands.

developed ROMANCE, a software allowing the automation of the process of converting batches of CE–MS files into effective electrophoretic mobility scale [7]. As it has been already described, files using this  $x$ -axis instead of time scale, present several advantages such as constant peak position for a given compound, no need for alignment and better shareability of experimental and reference data between different labs. Nevertheless, our previous article was focused on how the use of the electrophoretic mobility scale could improve metabolite annotation, while paying little attention to the influence of the time-to-electrophoretic mobility conversion on the peak area.

We herein present ROMANCE v2.0, a new version of the software developed to take into account new operational scenarios such as nonconstant electric fields, the use of more than one reference compound, and the influence of the conversion or the detection type on the peak area. These and other relevant fundamental considerations are described before illustrating their utility and performance using a panel of reference compounds, and a metabolomics study conducted on a set of cell culture samples.

## 2 Theory

In this section, we will present formulas to transform migration times to effective electrophoretic mobilities, taking into account nonconstant fields and the possibility of using more than one marker in the spirit of [8, 9]. The effect of these transformations on peak areas will be reviewed generalizing the results in [10]. Finally, we will study their impact on peak shapes giving quantitative measures of peak displacement and deformation, and from them we will propose *a priori* rules for the optimization of precision of peak position.

Only the final formulas will be discussed in the manuscript, since the interested reader can find their detailed derivation in the Supporting Information (Section E) and elsewhere [11].

### 2.1 Effective electrophoretic mobility

#### 2.1.1 Migration under nonconstant fields

Let us assume the presence of a *ramp time*  $t_R$  over which the electric field will change up to a final constant value. This ramp time will always be smaller than the migration times of the analytes,  $t_M$ .

The whole contribution of the shape can be summed up in a *shape parameter*,  $\lambda$ , that can be computed from  $\tau = t/t_R$  as:

$$\lambda = 1 - \int_0^1 S_R(\tau) d\tau, \quad (1)$$

where  $S_R(\tau)$  is a function defined on the  $[0, 1] \times [0, 1]$  square describing the dimensionless shape of the ramp, such that the integral is just the area of the square above the ramp.

The electroosmotic mobility  $\mu_{BGE}$  is highly variable, and needs to be determined run by run. Being just an offset, by measuring also the migration time  $t_{\text{marker}}$  of a substance of known effective mobility  $\mu_{\text{marker}}$  we arrive at:

$$\mu(t_M) = \mu_{\text{marker}} + \frac{L}{E_m} \left( 1 - \frac{\lambda t_R v_p}{L} \right) \times \left( \frac{1}{t_M - \lambda t_R} - \frac{1}{t_{\text{marker}} - \lambda t_R} \right), \quad (2)$$

where  $\mu$  is the effective electrophoretic mobility of the analyte,  $E_m$  the final value (i.e., after the ramp) of the electric field,  $v_{BGE}$  the speed of the BGE flow,  $\mu_{BGE}$  is the electroosmotic mobility of the BGE, and  $v_p$  is a constant term generated by the application of a pressure gradient along the capillary.

A detailed derivation of 2 can be found in Supporting Information Section E.1.

Solving for the effective mobilities, taking  $t_{\text{marker}}$  as the migration time of the EOF,  $t_{\text{EOF}}$ , which has effective mobility  $\mu_{\text{EOF}} = 0$ , and if there is no ramp time, we have just the usual formula:

$$\mu(t_M) = \frac{L}{E_m} \left( \frac{1}{t_M} - \frac{1}{t_{\text{EOF}}} \right). \quad (3)$$

However, when a ramp is present there is another unknown parameter,  $v_p$ , which depends on the applied pressure, temperature, or viscosity of the fluid. Just as for  $\mu_{BGE}$  (and unlike  $L$  or  $E_m$ ) we cannot expect to know its value on a run by run basis. There is an intuitive reason for the appearance of this parameter in the equation. In the absence of a ramp, the time it takes for an analyte to traverse the capillary is directly proportional to the sum of the speeds induced by the (constant) electric field and the (constant) pressure. This single combined speed parameter can be rewritten in terms of single experimental value—in particular, the observed time of the marker. In the presence of a ramp, the speeds of the analytes change over time in a nonlinear fashion, and in particular the contributions to overall speed from pressure and electric field are not constant with respect to each other. This means that we can use a single marker to remove the parameter from a single contribution—electric field or pressure, but not both at the same time.

#### 2.1.2 Two-marker formulas

We can apply the same reasoning as for  $\mu_{BGE}$ : eliminate an unknown instrumental parameter with a measurable marker property. The idea goes back to [9, 12], where it is applied to other sources of time shift, although without considering the implications on peak area and always assuming one of the markers to be the EOF.

In short, take a second marker ( $t_B, \mu_B$ ) and if one of the two markers is the EOF, say the second marker  $\mu_B = \mu_{\text{EOF}} = 0$ :

$$\mu(t_M) = \frac{(t_M - t_{\text{EOF}})(t_A - \lambda t_R)}{(t_A - t_{\text{EOF}})(t_M - \lambda t_R)} \mu_A. \quad (4)$$

In this form, it interpolates between zero mobility, for analytes arriving with the EOF ( $t_M \simeq t_{\text{EOF}}$ ), and mobility  $\mu_A$ , for analytes arriving with A. As in the general two-marker formula, the interpolation is not linear with respect to time.

The two-marker formula in Eq. (4) has the additional advantage, already seen in [9] that it does not depend on  $v_p$ ,  $E$ , or  $L$ .

As a summary, it can be said that the mobility of a peak with migration time  $t_M$ , in a capillary of length  $L$ , with a maximal electrical field  $E_m$  and a *linear* field ramp of length  $t_R$  is:

$$\mu = \frac{L}{E_m} \left( \frac{1}{t_M - t_R/2} - \frac{1}{t_{\text{EOF}} - t_R/2} \right), \quad (5)$$

as a function of the EOF time  $t_{\text{EOF}}$ , and neglecting contributions from the backpressure. If another marker with mobility  $\mu_A$  and time  $t_A$  is used, the formula is:

$$\mu = \frac{(t_M - t_{\text{EOF}})(t_A - t_R/2)}{(t_A - t_{\text{EOF}})(t_M - t_R/2)} \mu_A. \quad (6)$$

## 2.2 Electropherogram peak shapes

### 2.2.1 Area-preserving transformation

With an electropherogram measured as a list of couples  $(t_i, I_i)$  of times  $t_i$  and intensities/counts  $I_i$ , we can map  $t \mapsto \mu(t)$  as described in [7] being the mobilogram computed by:

$$(t_i, I_i) \mapsto (\mu(t_i), I_i). \quad (7)$$

In [7] we observed that this performs well for the *annotation* of features. However, if the intensities  $I_i$  represent a concentration profile that should be *integrated* over time, this change poses a problem. Because of the nonlinearity of the  $t \mapsto \mu$  transformation, the peak areas are completely changed because peak *widths* are changed notably. This was already observed in [8], and the necessary correction was derived in [10] for the simple case of (3).

Mobilograms should take into account this factor, giving a *corrected mobilogram intensity*  $I_i^{\text{mob}}$ :

$$I_i^{\text{mob}} = (t_i - \lambda t_R)^2 I_i \times \frac{t_B - t_A}{(\mu_A - \mu_B)(t_A - \lambda t_R)(t_B - \lambda t_R)}. \quad (8)$$

The detailed derivation of this formula can be found in Supporting Information Section E.1. The areas obtained by integrating the electropherogram  $(t_i, I_i)$  and the corrected mobilogram  $(\mu(t_i), I_i^{\text{mob}})$  will therefore be the same.

The right factors in (8) do not depend on the specific point of the mobilogram, and it is a function of only  $L$ ,  $E_m$ ,  $t_R$ , and  $v_p$ . This means that when comparing runs performed under the same instrumental conditions (field magnitude and ramp, capillary length, and pressure), we can do without such overall factor, simply using:

$$I_i^{\text{mob}} \propto (t_i - \lambda t_R)^2 I_i. \quad (9)$$

### 2.2.2 Effects on peak shape

Both correction factor and transformation itself may introduce distortions to the peak shape. Peaks closer to the EOF show a greater increase in relative width in the mobilogram, but their shape remains largely Gaussian. Peaks of very short migration times present little broadening, but may display asymmetry when the peak width is comparable to the migration time.

We can define a dimensionless relative asymmetry between the left and right widths of the peak in the mobility scale, which in terms of time-scale variables is:

$$\begin{aligned} & \frac{\text{Difference in left and right peak widths (mobility scale)}}{\text{Total peak width (mobility scale)}} \\ &= \frac{\text{Total peak width (time scale)}}{t_M - \lambda t_R}. \end{aligned} \quad (10)$$

Equation (10) shows that the *relative asymmetry* of a peak in the mobilogram equals precisely its electropherogram *relative width* with respect to its migration time. Thus, except for compounds of extremely short migration times, the impact of the transformation on peak shape should not affect peak asymmetry.

This asymmetry induced by the mobility transformation will also affect the computation of the center-of-mass position of the peak. This effect can be quantified and the center of mass  $\bar{\mu}$  in the mobilogram calculated as a function of the center of mass  $\bar{t}$  in the electropherogram:

$$\bar{\mu} = \mu(\bar{t}) \cdot \left[ 1 - \frac{t_{\text{EOF}} - \lambda t_R}{t_{\text{EOF}} - \bar{t}} \left( \frac{\sigma}{\bar{t} - \lambda t_R} \right)^2 + \dots \right]. \quad (11)$$

This means that the center of mass in the mobilogram,  $\bar{\mu}$  is *shifted* with respect to the mobility that corresponds to the center of mass in the electropherogram  $\mu(\bar{t})$ . In conclusion, as long as relative peak widths are kept below reasonable limits (<5% of the migration time), the transformation and correction should not negatively affect peak detection, and if center-of-mass integration is used, it will also not affect peak position whatsoever.

### 2.2.3 Detection mode

The electrospraying and detection of the analytes in the interface between the CE and the MS is a complex process that can take place in the so-called *mass-sensitive* or *concentration-sensitive* modes [13]. In mass mode, if the capillary flow is increased, more substance will be ionized per unit time and the MS will simply register more counts. By contrast, when in concentration mode, the signal is proportional to the volumetric concentration of substance arriving from the capillary, independent of time. Crucially, increasing the flow in the capillary will not change the signal.

When an electropherogram is obtained in concentration mode (or from UV measurements), given by couples  $(t_i, I_i^{\text{conc}})$ , the equivalent mass-mode electropherogram can be

**Table 1.** Electropherogram correction summary, showing the possible combinations of 9 and 13

		Peak integration mode	
		Area integration	Counts summation
Detection mode	Mass	$\times t^2$	$\times 1$
	concentration	$\times t$	$\times t^1$

computed multiplying by the analyte's output speed, which results in:

$$I_i^{\text{mass eq.}} = \frac{I_i^{\text{conc.}}}{t_i - \lambda t_R} \times \frac{E_m \cdot (\mu_A - \mu_B)(t_A - \lambda t_R)(t_B - \lambda t_R)}{t_B - t_A} \quad (12)$$

The detailed derivation of this formula can be found in Supporting Information Section E.1. As for the mobility transformation correction (9), when comparing runs under the same experimental conditions one may as well ignore the constant factor and use:

$$I_i^{\text{mass eq.}} \propto \frac{I_i^{\text{conc.}}}{t_i - \lambda t_R} \quad (13)$$

## 2.2.4 Summary

The correction to apply as shown in the previous sections can be selected by following these guidelines:

1. Integration: When the peak-picking software integrates the area under the curve, correction according to (9) is required due to the distortion of the mobility transformation on the time scale. On the contrary, if each point in the electropherogram is understood as the number of counts *since the previous acquisition*, no correction needs to be made, since the timing between points becomes irrelevant.
2. Detection: Mass-dependent detection requires no correction, while concentration-dependent mode requires (13).

The combined correction to be applied in each case to the electropherogram, prior to its conversion to mobilities, is summarized in Table 1. Of course, if there are nonnegligible contributions from the ramp, the full formulas (8) and (12) should be used.

## 2.3 Marker influence on precision

The choice of EOF as a marker is quite natural, but this begs the question of which criterion to follow when choosing a secondary marker in the two-marker formula.

Starting from the formula for an analyte's mobility using the EOF and a marker (4), if there is some error  $\delta t_A$  in the measurement of the marker migration time, it will induce a variation in the transformed mobility,  $\mu \rightarrow \mu + \delta\mu$ . We can

prove that the relative uncertainty in mobility caused by an uncertainty  $\delta t_A$  in the marker time is:

$$\left. \frac{\delta\mu}{\mu} \right|_A \simeq \frac{t_{\text{EOF}} - \lambda t_R}{(t_{\text{EOF}} - t_A)(t_A - \lambda t_R)} \cdot \delta t_A \quad \text{if } \delta t_A \ll t_{\text{EOF}} - t_A \quad (14)$$

The approximation we just made is valid for as long as the uncertainty is smaller than the distance between the secondary marker and the EOF. In such case, the relative mobility uncertainty is directly proportional to the marker position uncertainty. When  $t_A$  is too close to either  $t_A \rightarrow t_{\text{EOF}}$  or  $t_A \rightarrow \lambda t_R$  the relative error induced in the computed mobility becomes very large.

We can also compute the mobility uncertainty induced by that of  $t_{\text{EOF}}$ , which we denote  $\delta t_{\text{EOF}}$ . A similar computation yields:

$$\left. \frac{\delta\mu}{\mu} \right|_{\text{EOF}} \simeq \frac{t_M - t_A}{(t_{\text{EOF}} - t_A)(t_{\text{EOF}} - t_M)} \cdot \delta t_{\text{EOF}} \quad \text{if } \delta t_{\text{EOF}} \ll t_{\text{EOF}} - t_A \quad (15)$$

Unlike the uncertainty from  $t_A$ , this depends also on the migration time of the analyte  $t_M$ —analytes too close to the EOF will suffer more from uncertainty in its determination.

Assuming that both  $\delta t_A$  and  $\delta t_{\text{EOF}}$  are independent random variables, the variance of their combined effects will just be the sum of the variances. The expected total uncertainty is:

$$\left. \frac{\delta\mu}{\mu} \right|_{A+\text{EOF}} = \sqrt{\left. \frac{\delta\mu}{\mu} \right|_{\text{EOF}}^2 + \left. \frac{\delta\mu}{\mu} \right|_A^2} \quad (16)$$

Overall, the recommendation is to choose the secondary marker as close as possible to the midpoint toward the EOF, going on the side of slightly lower mobilities only if necessary.

## 3 Software

We have hitherto presented a theoretical derivation of a two-marker mobility formula, the possibility to handle field ramping, and the necessary area corrections to ensure that mobilograms represent consistently the amount of analyte in the samples. All these new features were not available in the original version of the ROMANCE software <https://ispso.unige.ch/labs/fanal/romance> and still developed in the Scala language, to take advantage of parallelization and multiplatform support. Summarizing the main changes and additions, it now offers:

- possibility to choose between providing the instrumental parameters ( $E$ ,  $L$ ) or using a secondary marker;
- ramp time correction;
- visual peak assessment windows for both markers, if applicable;
- mobilogram area correction, ionization mode area correction, and intersample area normalization;
- support for untargeted MS (spectra-based) mzML files, targeted MS (electropherogram-based) mzML files, and plain comma separated values (CSV) files for UV-style data.

This updated version has been used to perform the conversions of experimental data studied in the following sections.

## 4 Experimental validation

### 4.1 Materials and methods

#### 4.1.1 Chemicals, cell cultures, and instruments

To validate the formulas derived in the previous section, we have worked with a panel of 15 reference compounds. Furthermore, a metabolomics case study is presented using 2D-cell cultures of astrocytes grown in the presence of different natural neuroinflammatory triggers at different concentrations.

A comprehensive description of the experimental details can be found in Supporting Information A.

#### 4.1.2 Data processing

The raw data files were converted to the mzML format [14] using ProteoWizard msConvert [15]. They were subsequently converted to the effective electrophoretic mobility scale with ROMANCE v2.0, when applicable. Finally, the peak positions and areas were extracted with Skyline [16] for the targeted analyses on standards, and with Progenesis QI v2.4 for the untargeted metabolomics data (Nonlinear Dynamics, Newcastle upon Tyne, UK).

### 4.2 Ramp and two-marker formulas

In this section, we will study the effect of the ramp correction and two-marker formula (4) on the determination of the mobilities of the analytes. We will compare the obtained mobilities to a set of previously measured reference values [7], and their precision within our set of experiences.

The samples were separated with a linear ramp of  $t_R = 60$  s, and run in triplicate once applying a 0 mbar pressure, and once applying a 50 mbar pressure. This had the purpose of ensuring a considerable spread of the migration times. Each run was transformed with ROMANCE v2.0 to the mobility scale under each of the following four modes:

1. One marker, no ramp: using the classical formula (3), neglecting the ramp.
2. One marker, with ramp: using (2), neglecting the pressure-induced speed  $v_p$ .
3. Two markers, no ramp: using (4), neglecting the ramp.
4. Two markers, with ramp: using (4).

We chose as a secondary marker choline, the compound with migration times closest to mid-point toward the EOF, following the conclusions of section 2.3.

Figure 1A shows the results of the comparison against the previously measured values for the mobilities of these compounds. For each of them, the mobility was computed for each run, and the maximum relative deviation with respect to the known value amongst all six replicates (three at each pressure) was taken as an indicator of the maximal potential deviation and gathered in the shown box plots. In the case of a single marker ignoring the ramp, the variation reaches 18%. Including the ramp correction reduces the median to around 4% even while neglecting the correction due to pressure. Finally, using two markers and the ramp correction lowers the median to a maximal 2% deviation.

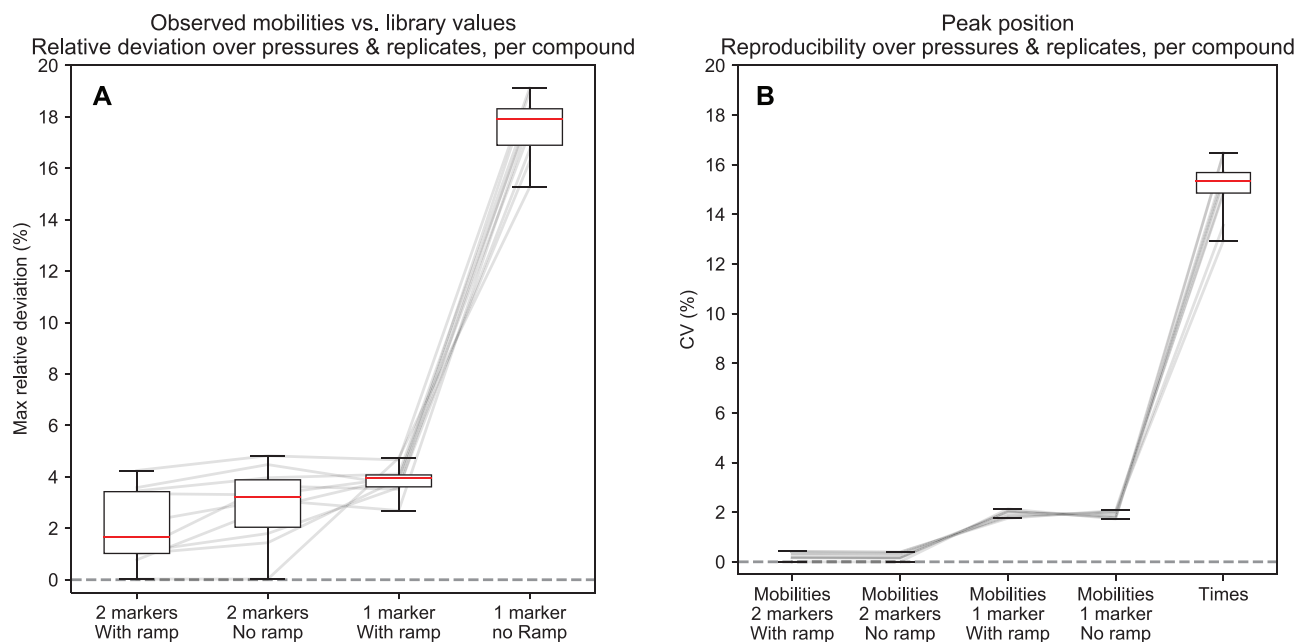
At this point, the deviation may come as much from inaccuracies in the present determination as from inaccuracies in the reference values, the latter derived with the traditional single-marker formula. In Fig. 1B, we plot the coefficient of variation (standard deviation over average) of the mobility of each compound over the six replicates (again, three at each pressure). The spread in migration times is high, which is to be expected from running each half of the replicates at different pressures. The conversion to mobilities, even with the single marker formula and no ramp correction reduces the variability to little more than 2%. This does not change with the addition of the ramp correction, meaning that the large deviation in Fig. 1A is caused by a systematic shift, as one would expect. But the addition of a second marker reduces the variability further to less than 0.5% between the six runs at different pressures. In this light, the deviations of  $\sim 2\%$  of the two-marker formulas in Fig. 1A are most likely due to variability in the original determination. These new mobility values for the chosen standards are available in Supporting Information B.

As expected from the elimination of all instrumental parameters from the formula, the use of the second marker improves the precision of the mobilities of the compounds by a factor of  $\sim 4$ . For this reason, if a reliable second marker is present in the sample, we strongly recommend using two markers to determine the mobility of compounds separated with CE.

### 4.3 Peak area precision

Our second aim is to assess the suitability of the mobility transformation for quantitative CE. All experiences under this section used the 15 selected compounds, and a milder ramp of  $t_R = 6$  s.

First, to make the choice of correction as per Table 1, we determined the detection mode. The mix of 15 compounds was analyzed under four different pressures (30, 50, 70, and 90 mbar), each run in triplicate, resulting in an array of migration times  $t_{c,p,i}$  for each compound  $c$ , pressure  $p$  and replicate  $i$ , and another one of peak areas  $A_{c,p,i}$ . To follow each compound along the different pressures, the replicates were averaged out, and in order to compare the compounds against



**Figure 1.** (A) Relative deviation of observed mobilities measured under a ramp with respect to the reference values. (B) Precision of observed mobilities measured under a ramp.

each other, normalized by their own mean over all pressures, as detailed in Supporting Information E.3. The same transformations were applied to obtain an array of normalized areas per compound and pressure,  $A_{c,p}^{(\text{normalized})}$ . These normalized values track only the variation between the free parameter (in this case, the pressure) relative to the compound's overall mean, so that if for some compound  $t_{c,p}^{(\text{average})}$  would not change at all between pressures,  $t_{c,p}^{(\text{normalized})} = 1$  for all  $p$ .

In Fig. 2A, we show the box plot for these normalized migration times. Note that the low variability at each pressure is induced by the mean over replicates. We observe the expected effect: migration times decrease with pressure with perfect consistency. Figure 2B shows the same but for peak areas. The variability is logically higher than it was for the peak positions, but clearly no trend is present on the data. The results are compatible with the detection responding to the amount of mass rather than to the concentration, since peak areas are independent of the applied pressure, and therefore of the flow in the capillary. This is indeed the preferred situation for this study, to ensure that we observe only the effects of the mobility transformation on peak areas, without mixing in flow-related effects.

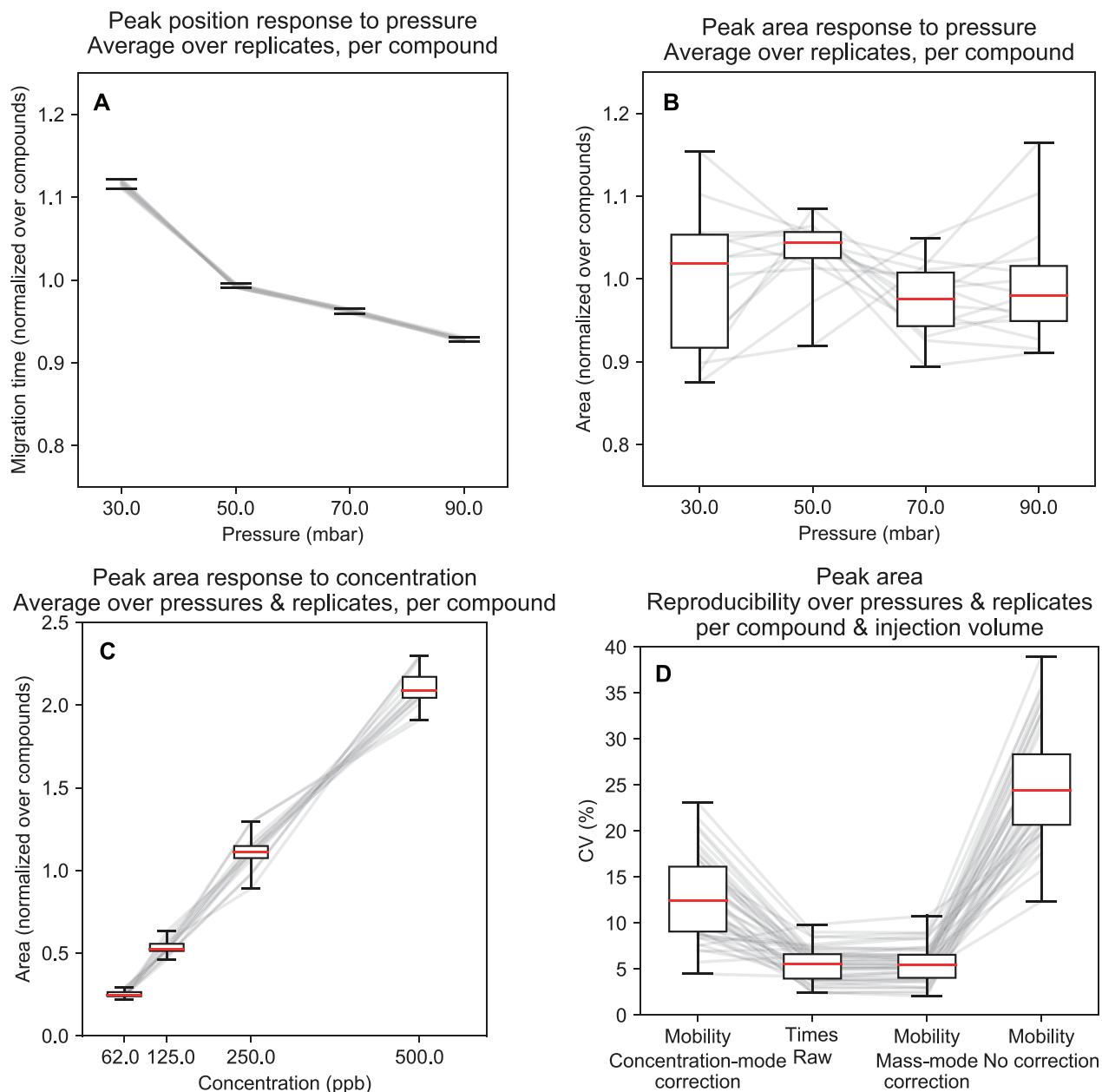
To ensure that these areas respond linearly to the amount of substance in the mix, it was analyzed from preparations at four concentration levels (62.5, 125, 250, and 500 ppb), each under three pressures (10, 30, and 50 mbar), which in turn were also run in triplicate. Since we have observed that different pressures provided the same areas, we can average both over replicates and pressures, and normalize over the free variable, the concentration  $\rho$  as depicted in Supporting Information E.3. In Fig. 2C, we can see the linear response of peak areas to concentrations.

We can finally move to our main interest, the integration of areas in the mobilogram. A mixture of standards was prepared at a constant concentration, but run with three different injection volumes (1, 2, and 3% of the capillary length) to induce peaks of different width. Each was analyzed at two pressures (10 and 50 mbar), again in triplicate. This gives a total of six replicates per combination of compound and injection volume.

The peaks were then extracted from the raw electropherograms, and from the mobilograms under different types of area correction. The conversion to mobilities was done using the full two-marker formula as in the previous subsection. For each compound and injection volume, we computed the coefficient of variation over the six replicates of the corresponding peak areas. Figure 2D displays the results. When applying the correct transformation for our case ( $\times t^2$ , following Table 1), the CVs remain the same between the peaks in the electropherograms and the peaks in the mobilograms ( $\sim 5\%$ ). If the wrong correction is made, by assuming that the ESI operates in concentration mode, leading to a factor of  $\times t$ , the peak areas show about three times more variability. It is only made worse by making no correction whatsoever.

Figure 3 shows the electropherogram for one of the runs at 10 mbar, and the corrected mobilogram obtained after conversion by ROMANCE, showcasing how the relative changes in width are compensated with the peaks' height.

In summary, choosing the right correction is critical to obtain repeatable areas, in which case the transformed mobilograms will perform just as well as the electropherograms while having the additional advantage of permitting the identification of peaks by their position, using libraries of known mobilities determined using the BGE [5].



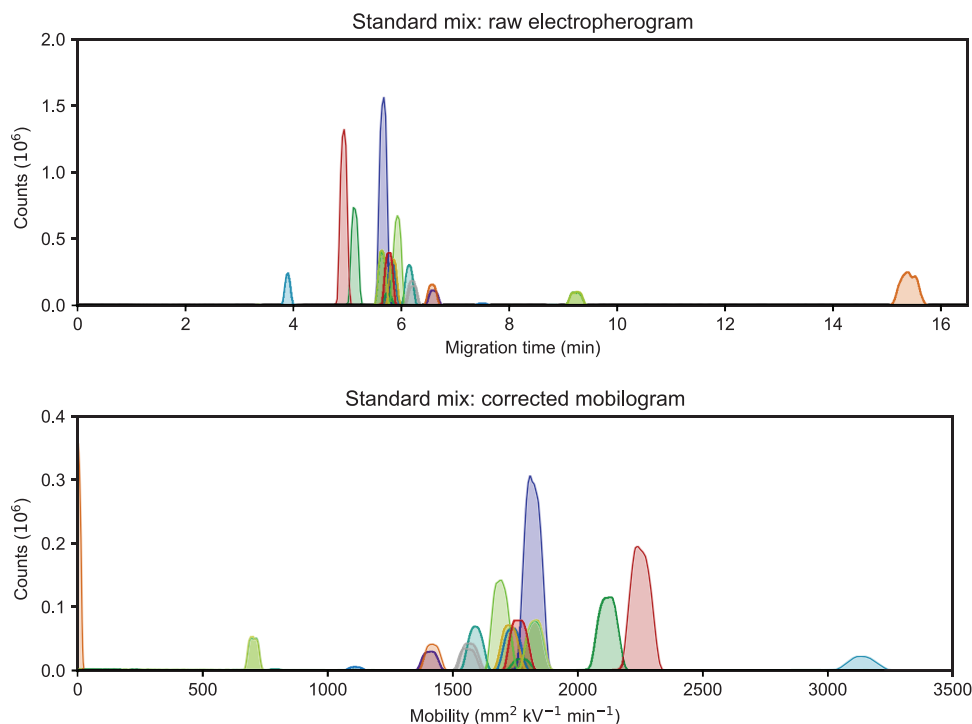
**Figure 2.** (A) Migration times and (B) peak areas as a function of pressure. (C) Electropherogram peak areas as a function of concentration. (A)–(C) were taken from raw time-scale electropherograms to study the phenomena discussed in the text. (D) Precision of converted peak areas using different approaches.

#### 4.4 Metabolomics application

The development of simplified assays for safety assessment (such as assays performed on cell cultures) is a key element within the changes taking place over the last decade in the field of chemical toxicology testing [17]. When it comes to toxicity assessment of molecules with neurotoxic potential, astrocytes are an appealing model system, since their activation upon exposure to different inflammatory triggers can drive them into either neurotoxic or neurotrophic states [18]. To check this approach, and as a first proof-of-concept,

2D-cultures of astrocytes were exposed to different natural neuroinflammatory triggers, namely interleukin  $1\beta$ , tumor necrosis factor  $\alpha$ , and lipopolysaccharide. In order to study polar metabolites involved in these processes, and as a showcase of the full ROMANCE workflow on actual metabolomics data, we have analyzed the astrocyte samples using a CE–MS platform and the data treatment pipeline introduced in this paper.

In our previous paper [7], we demonstrated how the use of the electrophoretic mobility scale could simplify several steps of the CE–MS-based untargeted metabolomics



**Figure 3.** Example of electropherogram to mobilogram conversion.

workflows—in particular, alignment, peak-picking, and annotation. In the present work, we have focused on how the integration process and the resulting calculated areas could be affected by this conversion. Therefore, we have used multivariate analysis (principal component analysis) to check whether the clustering of the experimental groups and the discriminant capacity of the technique was affected by mobility-based data processing of the CE–MS data with regard to the conventional time-based processing.

mzML-converted files were either directly imported into Progenesis QI, or transformed (and area-corrected) into the effective mobility scale by ROMANCE and then imported into Progenesis QI for peak-picking and peak grouping. Peaks were manually reviewed to ensure correct identifications against an in-house library, producing a set of 38 identified features in ESI+ mode and 28 in ESI– mode, common to both electropherogram and mobilogram peak extraction. To compensate sample amount variability in the samples, probabilistic quotient normalization was applied to both datasets [19]. Drift and other analytical effects were in turn corrected with the inclusion of quality control samples [20, 21].

Running a principal component analysis on the peaks extracted from the electropherograms, we obtained the score plots shown in Fig. 4A and B. We can observe that samples from each experimental group are well clustered, and that the first component captures the largest part of the inflammatory triggers' effect on the metabolome of the astrocytes, distinguishing all three treatments from the control group. Additionally, the second component finds an effect separat-

ing the tumor necrosis factor  $\alpha$  from the other two groups (interleukin  $1\beta$  and lipopolysaccharide), which remain clustered together.

Figure 4C and D shows the scores plots of the same samples after conversion by ROMANCE, using mass-mode area correction. Following the expectations from the results of Fig. 2D on standards, the score plots are fundamentally equivalent before and after conversion. Of course, mobilograms offer the advantage of allowing reliable identification based on external libraries, enlarging the amount of metabolites that can be identified without needing to resort to the evaluation of in-house libraries of standards.

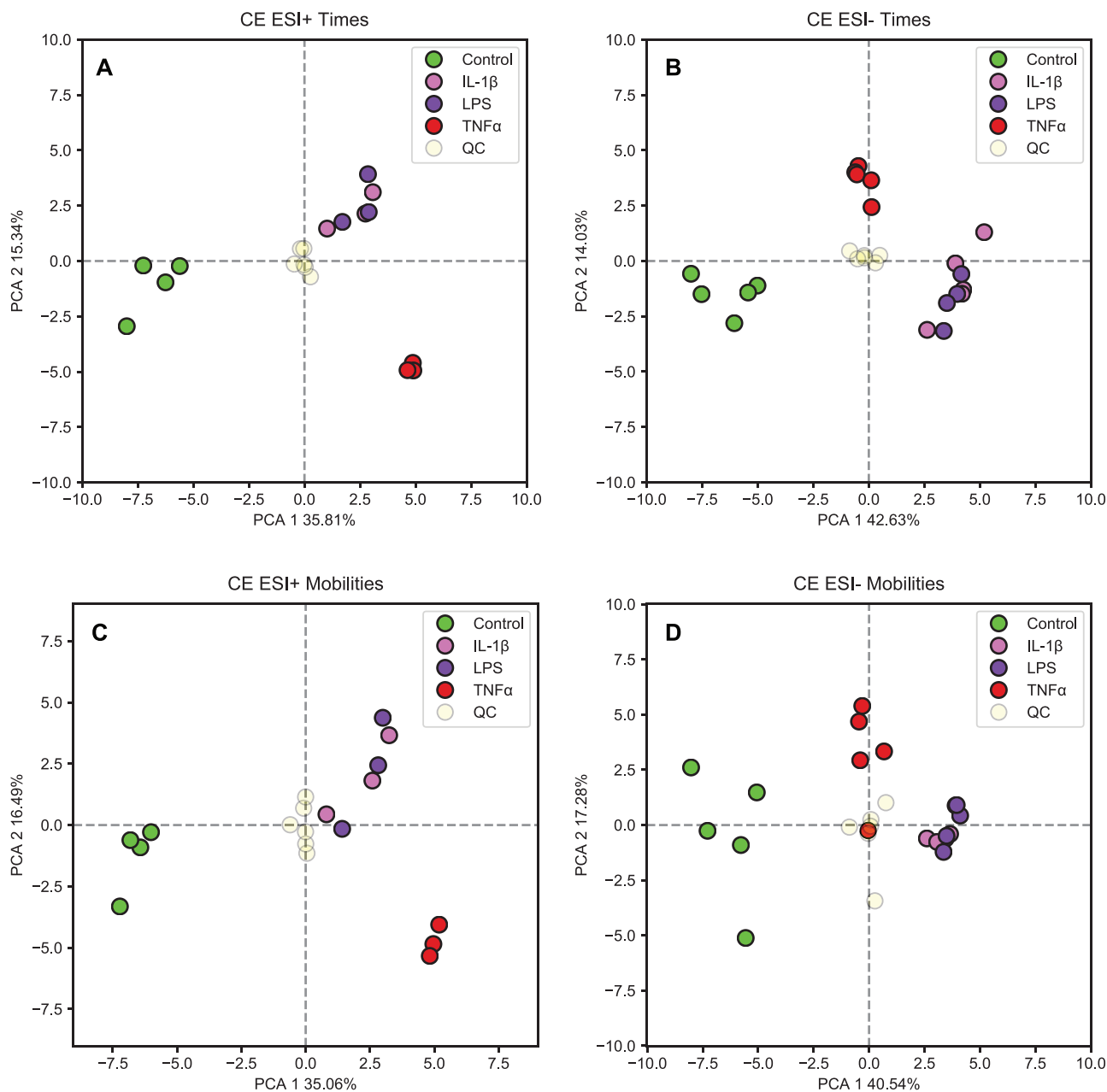
The list of identified peaks and their corresponding loadings from the mobility data are available in Supporting Information Table D.

## 5 Concluding remarks

We have seen that the transformation of CE data to the electrophoretic mobility scale not only improves peak identification, as was already known [7], but it also allows quantitative information to be extracted from mobilograms.

ROMANCE v2 has been introduced to perform these corrections, and also give more control to the user over the transformation parameters including the possibility of using multiple markers, field ramps, and selecting different detection regimes. We have validated the theoretical framework by studying peak position and area precision under the several transformation formulas shown in the article, showing the need to use the right area transformation to





**Figure 4.** Principal component analysis scores plots showing the clustering of the samples in the metabolomics application, and highlighting the analogue topology in the case of the raw electropherogram data (A and B) versus corrected mobilogram data (C and D). Two different combinations of separation polarities and ESI ionization modes are shown: direct with positive ESI (A and C) and reverse with negative ESI (B and D).

have reliable quantitative data. Finally, we have seen that with the current version of ROMANCE the workflow is ready for multivariate analysis of real metabolomics data, achieving a significant milestone in the path to make CE-MS part of the metabolomics toolkit.

Yoric Gagnebin is acknowledged for his contribution to the creation of ROMANCE. Marie-Gabrielle Zurich and David Pamies (Department of Physiology, University of Lausanne, Switzerland) are gratefully acknowledged for providing the as-

trocyte samples. V. G.-R. gratefully acknowledges the Martín-Carbajo family for supporting his scientific development through the IX<sup>th</sup> Mario Martín Velamazán award. This research was funded by the Swiss National Science Foundation through grant 31003A\_166658 for support through a fellowship to S. C. and by the Swiss Centre for Applied Human Toxicology through grants from the Research Programme 2017–2020 (Core Project 3: Neurotoxicity).

The authors have declared no conflict of interest.

## Data availability statement

Raw data and code depicted in this manuscript can be made available by the authors under request.

## 6 References

- [1] Zhang, W., Ramautar, R., *Electrophoresis* 2021, 42, 381–401.
- [2] García, A., Godzien, J., López-González, Á., Barbas, C., *Bioanalysis* 2017, 9, 99–130.
- [3] Zhang, W., Hankemeier, T., Ramautar, R., *Curr. Opin. Biotechnol.* 2017, 43, 1–7.
- [4] Schmitt-Kopplin, P., Fekete, A., *Capillary Electrophoresis*, Springer, Berlin, 2008, pp. 611–629.
- [5] Drouin, N., Pezzatti, J., Gagnebin, Y., Gonzalez-Ruiz, V., Schappler, J., Rudaz, S., *Anal. Chim. Acta* 2018, 1032, 178–187.
- [6] Drouin, N., vanMeyer, M., Zhang, W., Tobolkina, E., Ferre, S., Servais, A.-C., Gou, M.-J., Nyssen, L., Fillet, M., Lageveen-Kammeijer, G. S. M., Nouta, J., Chetwynd, A. J., Lynch, I., Thorn, J. A., Meixner, J., Lößner, C., Taverna, M., Liu, S., Tran, N. T., Francois, Y., Lechner, A., Nehmé, R., Al Hamoui Dit Banni, G., Nasreddine, R., Colas, C., Lindner, H. H., Faserl, K., Neusüß, C., Nelke, M., Lämmerer, S., Perrin, C., Bich-Muracciole, C., Barbas, C., González, Á. L., Guttman, A., Szigeti, M., Britz-McKibbin, P., Kroezen, Z., Shanmuganathan, M., Nemes, P., Portero, E. P., Hankemeier, T., Codesido, S., González-Ruiz, V., Rudaz, S., Ramautar, R., *Anal. Chem.* 2020, 92, 14103–14112.
- [7] González-Ruiz, V., Gagnebin, Y., Drouin, N., Codesido, S., Rudaz, S., Schappler, J., *Electrophoresis* 2018, 39, 1222–1232.
- [8] Mammen, M., Colton, I. J., Carbeck, J. D., Bradley, R., Whitesides, G. M., *Anal. Chem.* 1997, 69, 2165–2170.
- [9] Ikuta, N., Yamada, Y., Yoshiyama, T., Hirokawa, T., *Chin. J. Chromatogr. A* 2000, 894, 11–17.
- [10] Chamieh, J., Martin, M., Cottet, H., *Anal. Chem.* 2015, 87, 1050–1057.
- [11] Codesido, S., Drouin, N., Ferre, S., Schappler, J., Rudaz, S., González-Ruiz, V., ChemRxiv preprint 2020. <https://doi.org/10.26434/chemrxiv.12287048.v1>.
- [12] Ikuta, N., Yamada, Y., Hirokawa, T., *Electrophoresis* 2000, 21, 360–366.
- [13] Bonvin, G., Schappler, J., Rudaz, S., *J. Chromatogr. A* 2012, 1267, 17–31.
- [14] Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W. H., Römpp, A., Neumann, S., Pizarro, A. D., Montecchi-Palazzi, L., Tasman, N., Coleman, M., Reisinger, F., Souda, P., Hermjakob, H., Binz, P.-A., Deutsch, E. W., *Mol. Cell. Proteomics* 2011, 10, R110–000133.
- [15] Adusumilli, R., Mallick, P., *Proteomics*, Springer, Berlin, 2017, pp. 339–368.
- [16] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., MacCoss, M. J., *Bioinformatics* 2010, 26, 966–968.
- [17] Krewski, D., Acosta, Jr, D., Andersen, M., Anderson, H., Bailar, III, J. C., Boekelheide, K., Brent, R., Charnley, G., Cheung, V. G., Green, Jr, S., Kelsey, K. T., Kerkvliet, N. I., Li, A. A., McCray, L., Meyer, O., Patterson, R. D., Pennie, W., Scala, R. A., Solomon, G. M., Stephens, M., Yager, J., Zeise, L., Staff of Committee on Toxicity Testing and Assessment of Environmental Agents, *J. Toxicol. Environ. Health Part B* 2010, 13, 51–138.
- [18] Parpura, V., Heneka, M. T., Montana, V., Oliet, S. H., Schousboe, A., Haydon, P. G., Stout, Jr, R. F., Spray, D. C., Reichenbach, A., Pannicke, T., Pekny, M., Pekna, M., Zorec, R., Verkhratsky, A., *J. Neurochem.* 2012, 121, 4–27.
- [19] Dieterle, F., Ross, A., Schlotterbeck, G., Senn, H., *Anal. Chem.* 2006, 78, 4281–4290.
- [20] Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., Dunn, W. B., *Metabolomics* 2018, 14, 72.
- [21] Sun, J., *J. Chemom.* 1997, 11, 525–532. [https://doi.org/10.1002/\(SICI\)1099-128X\(199711/12\)11:6%3C525::AID-CEM489%3E3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-128X(199711/12)11:6%3C525::AID-CEM489%3E3.0.CO;2-G)