A mature ROMANCE: a matter of quantity and how two can be better than one

Santiago Codesido[1]   Nicolas Drouin[1,3]   Sabrina Ferre[1]

Julie Schappler[1,4]   Serge Rudaz[1,2]   Víctor González-Ruiz[1,2] *

[1]Analytical Sciences, Institute of Pharmaceutical Sciences of Western Switzerland, and School of Pharmaceutical Sciences, University of Geneva, Switzerland

[2]Swiss Centre for Applied Human Toxicology, Switzerland

[3]Present address: Leiden Academic Centre for Drug Research, University of Leiden, The Netherlands

[4]Present address: JULS life science, Singapore

ORCID identifiers: SC: 0000-0003-3669-6313, ND: 0000-0003-1568-7944, JS: 0000-0002-6281-3002, SR: 0000-0002-4180-5417, VGR: 0000-0001-7204-2363

Abstract

Capillary electrophoresis coupled to mass spectrometry (CE-MS) is increasingly gaining momentum as an analytical tool in metabolomics, thanks to its ability to obtain information about the most polar elements in biological samples. This has been helped by improvements in peak robustness by means of mobility-scale representations of the electropherograms (mobilograms). As a necessary step towards the use of CE-MS for untargeted metabolomics data, the authors previously developed and introduced the ROMANCE software, with the purpose of automating mobilogram generation for large untargeted datasets while offering a simple and self-contained user interface. In natural continuation ROMANCE has been upgraded to its v2 to read other types of data (targeted MS data and 2D UV-like electropherograms), offer more flexibility in the transformation parameters (including field ramping delays and the use of secondary markers), more measurement conditions (depending on detection and ionization modes), and most importantly tackle the issue of quantitative CE-MS. To prepare the ground for such an upgrade, we present a review of the current theoretical framework with regards to peak reproducibility and quantification, and we develop new formulas for multiple marker peak area corrections, for anticipating peak position precision, and for assessing peak shape distortion. We then present the new version of the software, and validate it experimentally. We contrast the multiple marker mobility transformations with previous results, finding increased precision, and finally we showcase an application to actual untargeted metabolomics data.

---

*victor.gonzalez@unige.ch

1 Introduction

There is a vast diversity of molecules presenting potential interest as the subject of metabolomic analyses. Due to the different nature of their chemical properties, it is unfeasible to explore their behaviours in biological systems except when using a panel of complementary separation and detection techniques. Each analytical platform shall be able to grasp the most relevant information about one of more families of chemically-related compounds. In the case of capillary zone electrophoresis (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].

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. Many factors account for this lack of robustness, such as different experimental setups, matrix adsorption on the capillary walls, or changes in the suction caused by ESI sources at the capillary outlet when hyphenated to MS. To mitigate this issue, some alternatives to migration times have been proposed. Relative migration times expressed as ratios to a reference compound 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 electroosmotic flow (EOF) during the separation. Effective electrophoretic mobility is a much more robust parameter, and it constitutes a better option for the identification of metabolites since it is a molecular property which only depends on the nature of the chosen BGE and separation temperature [4, 5].

In spite of its convenience, calculation of electrophoretic mobilities is a tedious process, and so we recently developed ROMANCE, a software allowing to automate the process of converting batches of CE-MS files into effective electrophoretic mobility scale [6]. 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.

As a natural continuation to the introduction of ROMANCE in our previous article, we herein present ROMANCE v2. This new version of the software has been developed to take into account new operational scenarios such as non-constant electric fields, the use of more than one reference compound, and the influence of the conversion or the ionization regime on the peak area. These and other relevant fundamental considerations are described in detail in the provided theoretical framework before illustrating their utility and performance using a panel of reference compounds used for validation and then, a metabolomics study conducted on a small set of cell culture samples.

2 Theory

In this section we will derive formulas to transform migration times to effective electrophoretic mobilities, taking into account non-constant fields and the possibility of using more than one marker in the spirit of [7, 8]. We will review the effect of these transformations on peak areas generalizing the results in [9]. 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.

Each derivation is accompanied by a summary subsection recollecting simplified versions of the most important formulas, for readers that may wish to skip the corresponding mathematical details or just have a quick reference.

2.1 Effective electrophoretic mobility

2.1.1 Migration under non-constant fields

To make the connection between the kinetics of electrophoretic migration and the parameters of the system, start by recalling that the migration speed v of a particular analyte in CE is

$$v = v_{BGE} + \mu\, E \tag{1}$$

where μ is the effective electrophoretic mobility of the analyte, E is the applied electric field, and v_{BGE} the speed of the background electrolyte (BGE) flow. We assume this can be written as

$$v_{BGE} = v_p + \mu_{BGE}\, E, \tag{2}$$

where μ_{BGE} is the electroosmotic mobility of the BGE and v_p a possible constant term generated typically by the application of a pressure gradient along the capillary.

Our first modification to the standard story is 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,

$$t_R < t_M. \tag{3}$$

On the one hand, it is crucial to have formulas that can be written in closed form, without relying on numerical integration for every transformation. Without such closed formulas it would be impossible to derive the correction factors for the peak areas that will be presented later in the article. On the other, this is already the case in virtually every CE setup. Contrast with LC, where gradients of retentivity have an intrinsic positive effect on separation. In CE, a high electric field will normally just induce a better separation, as it linearly affects the migration speeds. The "ramp phase" of the electric field is often required for mostly instrumental reasons.

This allows us to parametrize an arbitrary electric field profile $E(t)$ by

$$E(t) = \begin{cases} E_m \cdot S_R\left(\frac{t}{t_R}\right) & \text{if } \ 0 \leq t \leq t_R, \\ E_m & \text{if } \ t > t_R, \end{cases} \tag{4}$$

where t_R is the time over which the field is allowed to vary, and E_m the final, constant applied field. The function

$$S_R : [0,1] \rightarrow [0,1] \tag{5}$$

describes the *shape* of the ramp, giving the *fraction* of the maximum applied field E_m as a function of the *fraction* of time between 0 and t_R. For example, a simple linear ramp has a shape function given by

$$S_{linear}(\tau) = \tau, \tag{6}$$

resulting in a linear ramp field profile

$$E_{linear}(t) = \begin{cases} \frac{t}{t_R} E_m & \text{if} \quad 0 \leq t \leq t_R, \\ E_m & \text{if} \quad t > t_R, \end{cases} \tag{7}$$
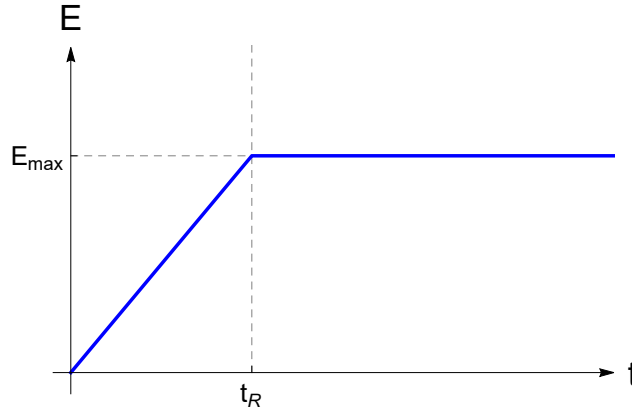
as shown in Figure 1.

Figure 1: Linear field ramp example.

During the separation, the analyte must travel the length of the capillary, L. This length is obtained from integrating the speed between the start of the separation at $t = 0$ and the migration time of the analyte, t_M.

$$L = \int_0^{t_M} v \, dt = \int_0^{t_M} \left(v_p + (\mu_{BGE} + \mu) \, E(t) \right) dt \tag{8}$$

Since we assumed that $t_M > t_R$, we can split the integral into three terms: the contributions of pressure, of the ramping field, and the constant field

$$
\begin{aligned}
L = v_p \cdot \int_0^{t_M} dt + \\
+ (\mu_{BGE} + \mu) \cdot E_m \cdot \int_0^{t_R} S_R \left(t/t_R \right) dt + \\
+ (\mu_{BGE} + \mu) \cdot E_m \cdot \int_{t_R}^{t_M} dt.
\end{aligned} \tag{9}
$$

The only dependence on the shape of the ramp is in the second term, which in turn does not depend on the migration time. We get

$$
\begin{aligned}
L = & \left(v_p + (\mu_{BGE} + \mu) \cdot E_m \right) \cdot t_M - \\
& - \left(1 - \int_0^1 S_R(\tau) \, d\tau \right) \cdot (\mu_{BGE} + \mu) \cdot E_m \cdot t_R.
\end{aligned}
\tag{10}
$$

The whole contribution of the shape can be summed up in a *shape parameter*, defined by

$$
\lambda = 1 - \int_0^1 S_R(\tau) \, d\tau.
\tag{11}
$$

This is just the area *over* the ramp shape curve. As a sanity check, a constant, maximal shape function ($S_R(\tau) = 1$) produces a shape factor of $\lambda = 0$, or that is, no contribution from the ramp. A linear ramp like in Figure 1 on the other hand will have

$$
\lambda_{linear} = \frac{1}{2}.
\tag{12}
$$

The shape factor for other ramp profiles can be easily calculated with formula (11).

Our objective is to transform the migration times t_M into effective mobilities μ, so we solve for the latter in (10),

$$
\mu(t_M) = \frac{L}{E_m} \frac{1 - \lambda t_R v_p / L}{t_M - \lambda t_R} - \frac{v_p}{E_m} - \mu_{BGE}.
\tag{13}
$$

Here we face the well known problem of CE – the electroosmotic mobility μ_{BGE} is highly variable, and needs to be determined run by run. Being just an offset, by measuring also the migration time t_A of a substance of known effective mobility μ_A we arrive at

$$
\mu(t_M) = \mu_A + \frac{L}{E_m} \left(1 - \frac{\lambda t_R v_p}{L} \right) \left(\frac{1}{t_M - \lambda t_R} - \frac{1}{t_A - \lambda t_R} \right).
\tag{14}
$$

A convenient choice for A is the migration time of the electroosmotic flux, t_{EOF}, which has effective mobility $\mu_{EOF} = 0$ (this should not be mistaken for the *electroosmotic* mobility μ_{BGE}, which is not zero). If additionally there is no ramp time, we have just the usual formula

$$
\mu(t_M) = \frac{L}{E_m} \left(\frac{1}{t_M} - \frac{1}{t_{EOF}} \right).
\tag{15}
$$

However, when a ramp is present there is another unknown parameter, v_p. This depends in a non-trivial manner on the applied pressure, but also indirectly on harder to control variables such as temperature that influence, e.g., the viscosity of the fluid. Just as for μ_{BGE} (and unlike L or E_m) we cannot expect to know its value on a run by run basis.

2.1.2 Two-marker formulas

Instead we can apply the same reasoning as for μ_{BGE}: eliminate an unknown instrumental parameter with a measurable marker property. The idea goes back to [8, 10], 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, μ_B) and consider the quotient

$$\frac{\mu(t_M) - \mu_A}{\mu(t_M) - \mu_B} = \frac{\frac{1}{t_M - \lambda t_R} - \frac{1}{t_A - \lambda t_R}}{\frac{1}{t_M - \lambda t_R} - \frac{1}{t_B - \lambda t_R}}. \tag{16}$$

The only instrumental parameter this relation depends is the ramp profile (through λ and t_R). Solving for μ, we arrive at our *two-marker effective mobility formula*,

$$\mu(t_M) = \frac{(t_M - t_B)(t_A - \lambda t_R)\,\mu_A - (t_M - t_A)(t_B - \lambda t_R)\,\mu_B}{(t_A - t_B)(t_M - \lambda t_R)}. \tag{17}$$

Intuitively, this is an interpolation between the mobilities of the two markers, weighted (non-linearly) by their respective distances to an analyte with migration time t. The formula simplifies considerably if one of the two markers is the EOF, say the second one $\mu_B = \mu_{EOF} = 0$,

$$\mu(t_M) = \frac{(t_M - t_{EOF})(t_A - \lambda t_R)}{(t_A - t_{EOF})(t_M - \lambda t_R)}\,\mu_A. \tag{18}$$

In this form, it interpolates between zero mobility, for analytes arriving with the EOF ($t_M \simeq t_{EOF}$), and mobility μ_A, for analytes arriving with A. As in the general two-marker formula, the interpolation is not linear. We can see an example of the shape of μ as a function of t in Figure 2.
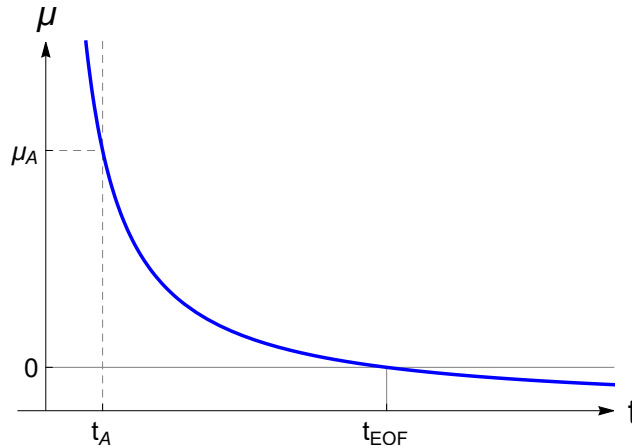

Figure 2: Two-marker effective mobility formula time dependence.

The two-marker formula in (18) has an additional advantage, already seen in [8]. In the search to eliminate the instrumental parameter v_p that appeared as a consequence of the field ramp, we have also gotten rid of the field and capillary length, E and L. Because of this, even in absence of a field ramp, the two-marker formula can actually be an improvement in terms of precision with respect to the known one-marker formula (due to inexact values of the effective capillary length, parasitic resistances affecting the actual value of the field applied to the capillary, etc.).

2.1.3 Summary

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

$$\mu = \frac{L}{E_m} \left(\frac{1}{t_M - t_R/2} - \frac{1}{t_{EOF} - t_R/2} \right), \tag{19}$$

as a function of the electroosmotic flow time t_{EOF}, and neglecting contributions from the backpressure. If another marker with mobility μ_A and time t_A is used, the formula is

$$\mu = \frac{(t_M - t_{EOF})(t_A - t_R/2)}{(t_A - t_{EOF})(t_M - t_R/2)} \mu_A. \tag{20}$$

2.2 Electropherogram peak shapes

2.2.1 Area-preserving transformation

Recall that the purpose of transforming migration times into effective mobilities is to allow peak-picking software to have a robust identifying parameter (the mobility of the compound) to annotate features extracted from electropherograms. The obvious approach, as implemented in [6], is to use formula (14) or (18) to map $t \mapsto \mu(t)$. An electropherogram is measured as a list of couples (t_i, I_i) of times t_i and intensities/counts I_i. Therefore the mobilogram should be computed by

$$(t_i, I_i) \mapsto (\mu(t_i), I_i). \tag{21}$$

In [6] 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. In Figure 3 we show an example. We suppose an EOF arriving at $t_{EOF} = 10$, a marker with $t_A = 1$ and effective mobility $\mu_A = 100$, and no ramp time. We then place two peaks with similar widths $\sigma_1 = 0.2$ and $\sigma_2 = 0.25$ and exactly the same area $A_1 = A_2 = 1$, at times $t_1 = 2.5$ and $t_2 = 7$. Because of the nonlinearity of the $t \mapsto \mu$ transformation, the peak areas are completely changed. Although gaussian-like peaks will remain essentially gaussian as long as they do not overlap the EOF (whatever distortion may be introduced by the transformation is negligible next to the inherent variations in CE peak shapes, as we will see), peak *widths* are changed notably.

This was already observed in [7], and the necessary correction was derived in [9] for the simple case of (15). We will redo the calculation to include the ramp factor and the secondary marker. Essentially, having a closed formula such as (17) for $\mu(t)$ allows us to compute a correction factor. In the limit of a continuous intensity profile $I(t)$, the area under the curve of a peak in the electropherogram is given by

$$A = \int I(t) \, dt. \tag{22}$$

Changing variables from $t \mapsto \mu(t)$, the integration measure changes by

$$dt = \left| \frac{\partial \mu}{\partial t} \right| d\mu \implies A = \int I(\mu) \cdot \left| \frac{\partial \mu}{\partial t} \right|^{-1} d\mu. \tag{23}$$
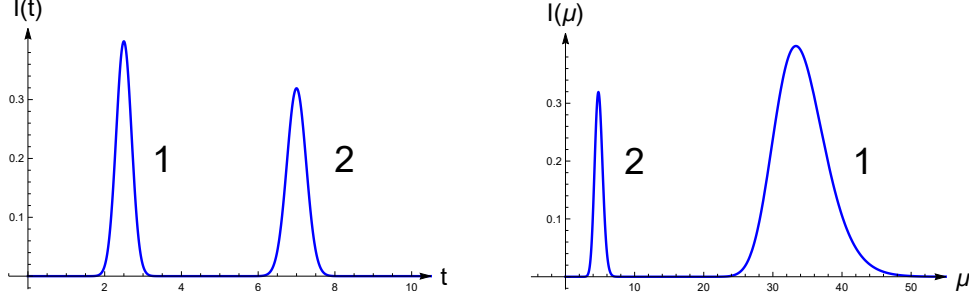
Figure 3: Continuous electropherogram before and after conversion to mobility scale.

This has a simple expression as a function of t, which can be obtained from (17),

$$\frac{\partial \mu}{\partial t} = -\frac{(\mu_A - \mu_B)(t_A - \lambda t_R)(t_B - \lambda t_R)}{t_B - t_A} \cdot \frac{1}{(t - \lambda t_R)^2}. \tag{24}$$

Mobilograms should take into account this factor, giving a *corrected mobilogram intensity* I^{mob},

$$I_i^{mob} = (t_i - \lambda t_R)^2 \, I_i \, \times \, \frac{t_B - t_A}{(\mu_A - \mu_B)(t_A - \lambda t_R)(t_B - \lambda t_R)}. \tag{25}$$

The areas obtained by integrating the electropherogram (t_i, I_i) and the corrected mobilogram $(\mu(t_i), I_i^{mob})$ will therefore be the same.

The right factors in (25) 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

$$I_i^{mob} \propto (t_i - \lambda t_R)^2 \, I_i. \tag{26}$$

Of course, in doing so we ignore potential variability arising from the instrumental parameters. This may be important for the precision of migration times vs. effective mobilities. But the intrinsic variability in both the detection's response to actual concentration and the numerics of peak integration makes instrumental parameter variability barely relevant for peak areas.

2.2.2 Effects on peak shape

With (26) –or more correctly (25)– we can ensure that peak areas in the mobilogram stay faithful to those in the electropherogram. While our previous derivation, or that of [9], focus on this aspect, there is another issue that can severely affect peak integration: peak shape. Both the correction factor and the transformation itself may introduce distortions. Figure 4 provides a few examples. The larger peak widths have been intentionally chosen to exaggerate the effects of the transformation, and real electropherograms will display milder effects. We can see that peaks closer to the EOF (case B) show, as expected, a greater increase in relative width in the mobilogram (corrected through their reduced height), but their shape remains largely Gaussian. Peaks of very short migration times (high mobilities, case A) present little

broadening, but may display asymmetry if the peak width is comparable to the migration time. This asymmetry will complicate peak detection by peak-picking software. Nonetheless, note that case A2 is quite extreme, with its relative peak width being 40% of its migration time.
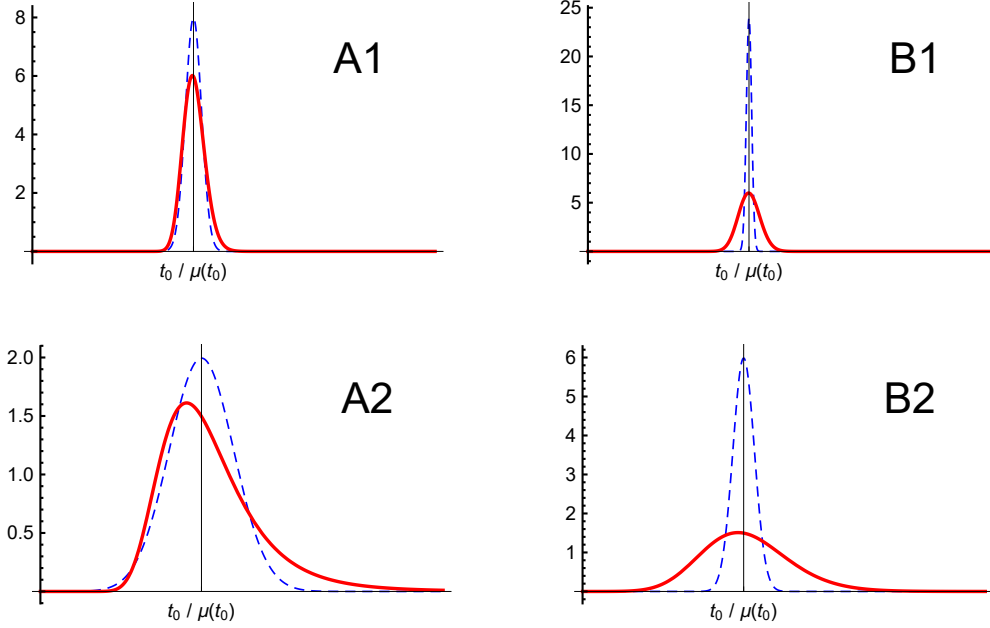

Figure 4: Mobility transformation effect on peak shapes, for $t_{EOF} = 10$ and $t_A = 5$. The dashed blue line represents a Gaussian peak in an electropherogram, and the solid red line the corresponding *corrected* mobilogram. Every curve has area $A = 1$. The x-axes are respectively normalized to the center of the Gaussian and its value in mobility. (**A**): Electropherogram peak at $t_0 = 2.5$. (**B**): Electropherogram peak at $t_0 = 7.5$. (**1**): Peak width $\Delta t = 0.25$. (**2**): Peak width $\Delta t = 1$.

Let us first focus on the asymmetry introduced only by the transformation. Suppose we have a peak centered in the electropherogram at t_0, spanning from $t_- = t_0 - \Delta t$ to $t_+ = t_0 + \Delta t$. By solving

$$\mu_+ = \mu(t_-), \quad \mu_0 = \mu(t_0), \quad \mu_- = \mu(t_+) \tag{27}$$

with equation (13) one can easily arrive at

$$
\begin{aligned}
\Delta\mu_+ = \mu_+ - \mu_0 = \left(1 - \frac{\Delta t}{t_0 - \lambda t_R}\right)^{-1} \cdot \frac{L - \lambda t_R v_p}{E_m} \cdot \frac{\Delta t}{(t_0 - \lambda t_R)^2}, \\
\Delta\mu_- = \mu_0 - \mu_- = \left(1 + \frac{\Delta t}{t_0 - \lambda t_R}\right)^{-1} \cdot \frac{L - \lambda t_R v_p}{E_m} \cdot \frac{\Delta t}{(t_0 - \lambda t_R)^2}.
\end{aligned}
\tag{28}
$$

These would be the (asymmetric) widths of the peak in the mobilogram. The terms on the right side of the expression are essentially (25), and correspond to the width change

9

introduced by the transformation we have already studied. We can define a dimensionless relative asymmetry by

$$\gamma = \frac{\Delta\mu_+ - \Delta\mu_-}{\Delta\mu_+ + \Delta\mu_-}, \tag{29}$$

which from (28) turns out to be equal to

$$\gamma = \frac{\Delta t}{t_0 - \lambda t_R}, \tag{30}$$

having a remarkably simple form. In short, equation (30) says that the *relative asymmetry* of a peak in the mobilogram equals precisely its electropherogram *relative width* with respect to its migration time. The latter will normally be low anyway, if we are to have a separation between the different analytes at all. Except for compounds of extremely high mobility (and thus short migration times), the impact of the transformation on peak shape should not affect peak integration.

However, as we can see in case A2 in Figure 4, the asymmetry will also shift the position of the peak's maximum away from the correct value. While a 5% asymmetry originating from a 5% relative peak width will not be a problem for the detection of the peak shape, a potential 5% shift in peak position would definitely render the advantages of the transformation useless. On the other hand, the centre of mass of the peak seems to be roughly in the same place, as the asymmetry gives a slightly heavier tail on the right side of the mobilogram. Let us quantify this effect.

In the electropherogram, the center of mass position $\bar{t}$ is given by

$$\bar{t} = \frac{\int t\, I(t)\, \mathrm{d}t}{\int I(t)\, \mathrm{d}t}. \tag{31}$$

The mobility corresponding to this center of mass is $\mu(\bar{t})$. In the case of an infinitely thin peak, this would also be the mobilogram's center of mass. In general, it will be

$$\bar{\mu} = \frac{\int \mu\, I^{mob}(\mu)\, \mathrm{d}\mu}{\int I^{mob}(\mu)\, \mathrm{d}\mu}. \tag{32}$$

We defined I^{mob} in (25) so that

$$I^{mob}(\mu)\, \mathrm{d}\mu = I(t)\, \mathrm{d}t, \tag{33}$$

and so

$$\bar{\mu} = \frac{\int \mu(t)\, I(t)\, \mathrm{d}t}{\int I(t)\, \mathrm{d}t}. \tag{34}$$

We can expand $\mu(t)$ in a Taylor series around the electropherogram's center of mass,

$$\mu(t) = \mu(\bar{t}) + (t - \bar{t}) \cdot \left.\frac{\partial\mu}{\partial t}\right|_{t=\bar{t}} + \frac{1}{2}(t - \bar{t})^2 \cdot \left.\frac{\partial^2\mu}{\partial t^2}\right|_{t=\bar{t}} + \dots \tag{35}$$

and inserting it into the expression for the center of mass of the mobilogram,

$$\bar{\mu} = \mu(\bar{t}) + \left(\left.\frac{\partial\mu}{\partial t}\right|_{t=\bar{t}} \cdot \frac{\int(t - \bar{t})\, I(t)\, \mathrm{d}t}{\int I(t)\, \mathrm{d}t}\right) + \left(\frac{1}{2}\left.\frac{\partial^2\mu}{\partial t^2}\right|_{t=\bar{t}} \cdot \frac{\int(t - \bar{t})^2\, I(t)\, \mathrm{d}t}{\int I(t)\, \mathrm{d}t}\right) + \dots = $$
$$= \mu(\bar{t}) + 0 + \frac{1}{2}\left.\frac{\partial^2\mu}{\partial t^2}\right|_{t=\bar{t}} \cdot \sigma^2 + \dots \tag{36}$$

Critically, the linear term vanishes, and the quadratic one is simply the variance of the electropherogram's peak, σ^2. Using (18) to find the second derivative of μ with respec to t, we end with

$$\bar{\mu} = \mu(\bar{t}) \cdot \left[1 - \frac{t_{EOF} - \lambda t_R}{t_{EOF} - \bar{t}} \left(\frac{\sigma}{\bar{t} - \lambda t_R} \right)^2 + \dots \right]. \tag{37}$$

This means that the center of mass in the mobilogram, $\bar{\mu}$ is *shifted* with respect to the true mobility of the center of mass in the electropherogram. The term depending on t_{EOF} is just an artefact of considering the relative shift – compounds close to the EOF have near zero effective mobility, so any variation will induce a large relative shift. The important part is the dependence on σ – *quadratic* on the ratio between peak width and peak position.

As an example, for a peak mid-way through the electropherogram, $\bar{t} = 0.5 \, t_{EOF}$, and a relatively large peak width, $\sigma = 5\% \cdot \bar{t}$, the relative shift between the mobilogram's center of mass and the true mobility is

$$\frac{\bar{\mu} - \mu(\bar{t})}{\mu(\bar{t})} \simeq \frac{t_{EOF}}{0.5 \, t_{EOF}} \left(\frac{0.05 \, \bar{t}}{\bar{t}} \right)^2 = 0.5\%. \tag{38}$$

If the peak width is $\sigma = 1\% \cdot \bar{t}$, this quickly descends to

$$\frac{\bar{\mu} - \mu(\bar{t})}{\mu(\bar{t})} = 0.02\%, \tag{39}$$

much less than the variability on effective mobility that will stem from other sources.

The lesson is that 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 Ionization mode

The previous section has been devoted to ensuring that areas in mobilograms map faithfully to areas in electropherograms. Without further treatment, this assumes that in turn electropherogram areas represent faithfully the amount of substance in the sample.

This is highly dependent on the detection type. In the context of CE-MS, the critical element is the electrospray ionization (ESI) that nebulizes the output of the capillary to be fed into the MS. This ionization has two well known regimes of operation, the so-called *mass* and *concentration* modes [11]. In mass mode, the ESI is able to ionize (nearly) all the substance coming from the capillary. 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 ESI is saturated. It outputs an amount of ions proportional to the volumetric concentration of substance arriving from the capillary, independently of time. Crucially, increasing the flow in the capillary *does not* increase the number of counts.

The increase in width between the two flow regimes is strictly proportional to the exit speed of the analyte from the capillary – this does not depend on the ionization mode, simply on the fact that a fixed length in the capillary spends twice as much time exiting it if it is moving at half the speed. One can emulate the height reduction observed in mass

mode on data measured in concentration mode by simply multiplying each peak by its speed. Equation (1) relates the speed of an analyte to its mobility, which in turn can be derived from its migration time from (13) or (17). This provides the exit speed v_{exit} as a function of migration time t_M,

$$
\begin{aligned}
v_{exit}(t_M) &= (L - \lambda t_R v_p) \cdot \frac{1}{t_M - \lambda t_R} = \\
&= \frac{E_m(\mu_A - \mu_B)(t_A - \lambda t_R)(t_B - \lambda t_R)}{t_B - t_A} \cdot \frac{1}{t_M - \lambda t_R}.
\end{aligned}
\tag{40}
$$

As a side note, detection from UV absorbance and similar methods work exactly in the same way concentration mode ESI does in CE-MS. The absorbance is not a function of the flow rate, so total peak areas will depend on it. The same correction applies to them.

If one has an electropherogram measured in concentration mode (or from UV measurements), given by couples $(t_i, I_i^{conc.})$, the equivalent mass-mode electropherogram can be computed with

$$
I_i^{mass\ eq.} = \frac{I_i^{conc.}}{t_i - \lambda t_R} \times \frac{E_m \cdot (\mu_A - \mu_B)(t_A - \lambda t_R)(t_B - \lambda t_R)}{t_B - t_A}.
\tag{41}
$$

Just like for the mobility transformation correction (26), when comparing runs under the same experimental conditions one may as well ignore the constant factor and use

$$
I_i^{mass\ eq.} \propto \frac{I_i^{conc.}}{t_i - \lambda t_R}.
\tag{42}
$$

2.2.4   Summary

The need to apply the corrections shown in the previous sections is selected by the following two questions:

1. Integration: does the peak-picking software integrate the area under the curve, or does it sum the number of counts?
   The former scenario requires correction (26) due to the distortion of the mobility transformation on the time scale. On the other hand, 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: is the signal proportional to the total mass of substance, or to its concentration?
   Mass mode ESI requires no correction, while concentration mode ESI or UV absorbance measurements require (42).

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 non-negligible contributions from the ramp, the full formulas (25) and (41) should be used.

		Peak integration	
		Curve	Counts
ESI mode	Mass	$\times t^2$	$\times 1$
	Concentration	$\times t$	$\times t^{-1}$

Table 1: Electropherogram correction summary

2.3 Marker influence on precision

The choice of EOF as a marker is quite natural, since it is normally easily identifiable. This begs the question of which criterion to optimize when choosing a secondary marker in the two-marker formula. Intuitively, one can already see that choosing a marker too close to the EOF will yield bad quality results, since any small error in the determination of the marker's position would have important effects when extrapolated away.

Let us start with the formula for an analyte's mobility using the EOF and a marker (18). Assume there is some variation δt_A in the determination of the marker time. This can reflect either instrumental variability, or uncertainty in peak integration. Similar to the calculations for peak asymmetry, we consider the variation this would induce in the transformed mobility, $\mu \to \mu + \delta\mu$. From

$$\delta\mu = \mu|_{t_A \to t_A + \delta t_A} - \mu \tag{43}$$

and equation (18) we get that the relative uncertainty in mobility caused by an uncertainty δt_A in the marker time is

$$\left.\frac{\delta\mu}{\mu}\right|_A = \left(1 - \frac{\delta t_A}{t_{EOF} - t_A}\right)^{-1} \frac{t_{EOF} - \lambda t_R}{(t_{EOF} - t_A)(t_A - \lambda t_R)} \cdot \delta t_A \simeq$$
$$\simeq \frac{t_{EOF} - \lambda t_R}{(t_{EOF} - t_A)(t_A - \lambda t_R)} \cdot \delta t_A \quad \text{if} \quad \delta t_A \ll t_{EOF} - t_A. \tag{44}$$

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 $t_A \to t_{EOF}$ or $t_A \to \lambda t_R$ the relative error induced in the computed mobility becomes very large.

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

$$\left.\frac{\delta\mu}{\mu}\right|_{EOF} = \left(1 + \frac{\delta t_{EOF}}{t_{EOF} - t_A}\right)^{-1} \frac{t_M - t_A}{(t_{EOF} - t_A)(t_{EOF} - t_M)} \cdot \delta t_{EOF} \simeq$$
$$\simeq \frac{t_M - t_A}{(t_{EOF} - t_A)(t_{EOF} - t_M)} \cdot \delta t_{EOF} \quad \text{if} \quad \delta t_{EOF} \ll t_{EOF} - t_A. \tag{45}$$

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

13

Assuming that both δt_A and δt_{EOF} are independent random variables, the variance of their combined effects will just be the sum of the variances. The expected total uncertainty is

$$\left.\frac{\delta\mu}{\mu}\right|_{A+EOF} = \sqrt{\left.\frac{\delta\mu}{\mu}\right|_{EOF}^2 + \left.\frac{\delta\mu}{\mu}\right|_A^2}. \tag{46}$$
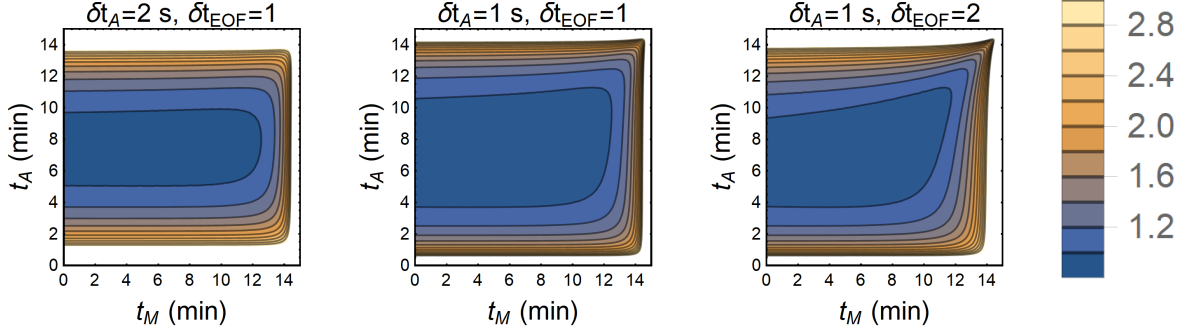

Figure 5: Relative mobility uncertainty $\delta\mu/\mu$ (in %) for $t_{EOF} = 15$ min, as a function of the analyte's migration time t_M and the secondary marker's migration time t_A.

In Figure 5 we plot this combined effect, as a function of t_A and t_M. We take $t_{EOF} = 15$ min, and consider uncertainties of 1 s and 2 s. From the influence of t_A, the lesson is that the secondary marker should be chosen at the center between $t = 0$ and t_{EOF}. However, if uncertainty in the determination of the EOF's time is larger (this can happen if the EOF marker is a relatively wide or non-Gaussian peak), compounds with low mobilities –close to the EOF– will be highly affected by a choice of t_A with high mobility, further away from the EOF. Overall, the recommendation is to choose the secondary marker as close as possible to the midpoint towards the EOF, erring 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.

These features were not available in the version of the ROMANCE software introduced in [6]. With the publication of this article, we have updated ROMANCE to its v2, to include them. The software is still publicly available at

https://ispso.unige.ch/labs/fanal/romance

and still developed in the Scala language, to take advantage of parallelization and multi-platform support. Summarizing the main changes and additions, it now offers:

- Possibility to choose between the old instrumental parameters (E, L) and a secondary marker as in (17).

- Ramp time correction.

- Visual peak assessment windows for both markers, if applicable.

- Mobilogram area correction, ionization mode area correction, and inter-sample area normalization.

- Support for untargeted MS (spectra-based) mzML files, targeted MS (electropherogram-based) mzML files, and plain 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   Material and methods

To validate the formulas derived in the previous section, we have selected 15 compounds, which can be found in the supplementary material, Table A. They were chosen first for having medium to high mobilities, to focus on those that can in principle suffer the highest distortions due to the transformation, as seen in section 2.2.2. In second place, by displaying good (i.e. relatively Gaussian) peak shapes in the electropherogram, to reduce as much as possible the variability in peak area stemming from peak integration, and again focus on the effects of the transformation. The compounds were prepared as a mixture of standards.

4.1.1   Chemicals and reagents

Acetic acid and formic acid were purchased from Biosolve (Dieuze, France). Water, methanol, isopropanol (iPrOH), and acetonitrile (ACN) were purchased from Fisher Scientific (Loughborough, United Kingdom). Standard compounds were purchased from Sigma Aldrich (Buchs, Switzerland).

4.1.2   Standard solution preparation

Individual stock solutions of compounds were prepared in 5% v/v ACN and 0.1 % v/v FA at a final concentration of 1 mg/mL and stored at -80°C. Mix stock solutions were prepared in 5% v/v ACN and 0.1% v/v FA at 10 μg/mL and stored at -80°C. Mix stock was extemporaneously diluted to 500, 250, 125 and 62.5 ng/mL with water.

4.1.3   Cell culture preparation

Four replicates of 2D-cell cultures of astrocytes were grown in the presence of different natural neuro-inflammatory triggers at different concentrations, namely interleukin 1β (IL-1β) at 30 ng/mL, tumoral necrosis factor α (TNFα) at 30 ng/mL, and lipopolysaccharide (LPS) at 10 μg/mL. The control cell culture was grown in parallel in absence of any inflammatory trigger. After two weeks of growth, cell cultures were snap-frozen in liquid nitrogen. Protein precipitation was achieved by adding 1 mL of a cold solution (-20°C) of MeOH:H$_2$O (80:20

v/v), scraping, and by vortexing during 1 minute. Samples were centrifuged at 14000 g during 15 minutes at 4°C, the supernatant was collected and then evaporated to dryness before resuspension in 100 μL of a solution made of MeCN:H$_2$O (50:50 v/v). Quality control (QC) samples were prepared by pooling the same volume from each sample after reconstitution. Volumes of 10 μL of individual cell culture extracts, QCs and diluted QCs were evaporated to dryness using a SpeedVac (ThermoFisher, Langenselbold, Germany). Before injection, samples were reconstituted with 10 μL of an aqueous solution containing paracetamol, procaine and ethyl-sulfate at a concentration of 50, 5 and 5 μg/mL respectively.

4.1.4   BGE and sheath-liquid preparation

Through the study, 10% v/v acetic acid in water was used as BGE. The sheath-liquid was composed of isopropanol-water-acetic acid (50:50:1). To ensure a large metabolome coverage for comprehensive metabolomics profiling, the sheath-liquid was composed of acetic acid (5 mM) in an isopropanol-water (50:50) solution. Purine and HP-0921 were purchased from Agilent technologies (Santa Clara, CA, USA, P/N: G1969-8001) and used as lock masses after being spiked into the sheath-liquid to yield final concentrations of 50 and 25 nM, respectively.

4.1.5   Validation analyses

A triple quadrupole platform was used for the study and validation of the formulas derived in the theory section. The separation was carried out with a G7100 capillary electrophoresis (CE) system from Agilent Technologies (Waldbronn, Germany). Separations were performed using a fused silica capillary purchased from BGB technologies (Boeckten, Switzerland) with a total length of 70 cm and an internal diameter of 50 μm. Prior to its first use, the capillary was conditioned with MeOH, H$_2$O, NaOH 1M, H$_2$O, HCl 1M, H$_2$O, HCl 0.1M, H$_2$O, and BGE at 5 bar during 1 minute each. Injections were performed hydrodynamically by application of 50 mbar during 12 s, using $\sim$1% of the capillary total length, circa 14 nL. Injected volumes were calculated with Zeecalc v1.0b (https://ispso.unige.ch/labs/fanal/zeecalc). Separation was performed by application of $+$ 30 kV. Before each analysis, the capillary was washed with MeOH and BGE at 5 bar during 1 minute. To avoid temperature inhomogeneities between the capillary parts inside and outside the CE instrument, the CE thermostat was set at room temperature ($\sim$23 °C).

The CE system was hyphenated with an Agilent 6490 triple quadrupole mass spectrometer (QqQ MS, Agilent Technologies, Santa Clara, CA, US) equipped with an ESI source via a coaxial sheath-flow interface with a standard triple-tube sprayer (P/N G1607B) from Agilent Technologies. The sheath liquid was delivered at a flow rate of 3 μL/min, using a 2300 Series isocratic pump purchased from Agilent Technologies (Waldbronn, Germany) equipped with a 1:100 split. Electrospray ionization was operated in positive mode, and spectra were acquired via SRM measurements. The pressures and injection volumes used during the validation are described in section 4.2. The precursor and productions monitored for each compound and the collision energies are reported in Table B in the supplementary information. The following source parameters were used: the nebulizing gas pressure was set at 0 psi and the sheath gas at 11 L/min and 150°C. The capillary voltage was adjusted to 5500 V. The ion funnel voltages were set at 150 V for the high-pressure funnel and 60 V for the low-pressure one. The EMV

voltage was set at 400 V and the cell accelerator voltage at 5 V. For all transitions, precursor and product ion selection was performed with a resolution of 1.2 and 0.7 m/z, respectively. Data acquisition and instrument control were performed using MassHunter version B.08.00 (Agilent, Santa Clara, US).

4.1.6 Untargeted metabolomics

The CE setup for the untargeted metabolomics profiling was the same as described in section 4.1.5 for the validation. The CE system was in turn hyphenated to a maXis-3G QTOF MS from Bruker (Bremen, Germany), equipped with an ESI source via a coaxial sheath-flow ESI interface with a standard triple-tube sprayer (Agilent P/N G1607A) and a platinum needle. The sheath liquid was delivered at a flow rate of 3 μL/min, using a 2300 Series isocratic pump purchased from Agilent Technologies (Waldbronn, Germany) equipped with a 1:100 split.

For cationic profiling, ESI was operated in positive mode with the following MS parameters: nebulizer and sheath gas were set to 0 bar and 10 L/min, 100°C, respectively. Capillary, end-plate and funnel voltages were respectively adjusted to 6000, 400 and 300 V. For anionic profiling, ESI was performed in negative mode with the following source parameters: nebulizing gas and sheath gas were set to 0.3 bar and 4 L/min, 150°C respectively. Capillary, end-plate and funnel voltages were adjusted to 4000, 400 and 300 V, respectively. MS acquisitions were performed at a frequency of 1 Hz, with a mass range going from 50 to 1000m/z.

4.1.7 Data processing

The raw data files were converted to the mzML format [12] using ProteoWizard msConvert [13]. They were subsequently converted to the effective electrophoretic mobility scale with ROMANCE v2, when applicable. Finally, the peak positions and areas were extracted with Skyline [14] for the targeted analyses on standards, and with Progenesis QI v.2.4 for the untargetd metabolomics data (Nonlinear Dynamics, Newcastle upon Tyne, UK).

4.2 Ramp and 2-marker formulas

In this section we will study the effect of the ramp correction and 2-marker formula (18) on the determination of the mobilities of the analytes. We will compare the obtained mobilities with values from a previous library [6] (restricting ourselves to the 10 substances for which a mobility was there given), 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 to the mobility scale under each of the following four modes:

1. One marker, no ramp: using the classical formula (15), neglecting the ramp.
2. One marker, with ramp: using (14), neglecting the pressure-induced speed v_p.
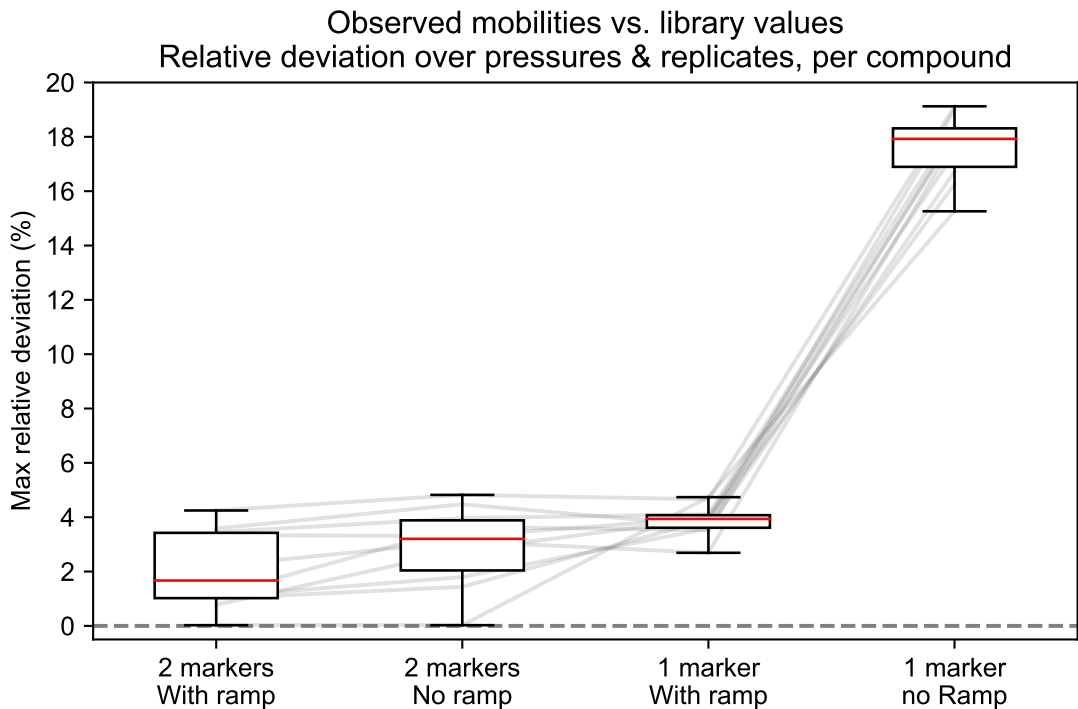3. Two markers, no ramp: using (18), neglecting the ramp.

Figure 6: Relative deviation of observed mobilities measured under a ramp with respect to the reference values.

4. Two markers, with ramp: using (18).

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

In Figure 6 plot the results of the comparison against the previously known 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. These maximum deviations per compound were then gathered in the shown box-plots. In the case of a single marker ignoring the ramp, the variation reaches 20%. Simply 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 with respect to the known values.

At that 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 Figure 7 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 Figure 6 is caused by a systematic shift, as one would expect. But the addition of a second
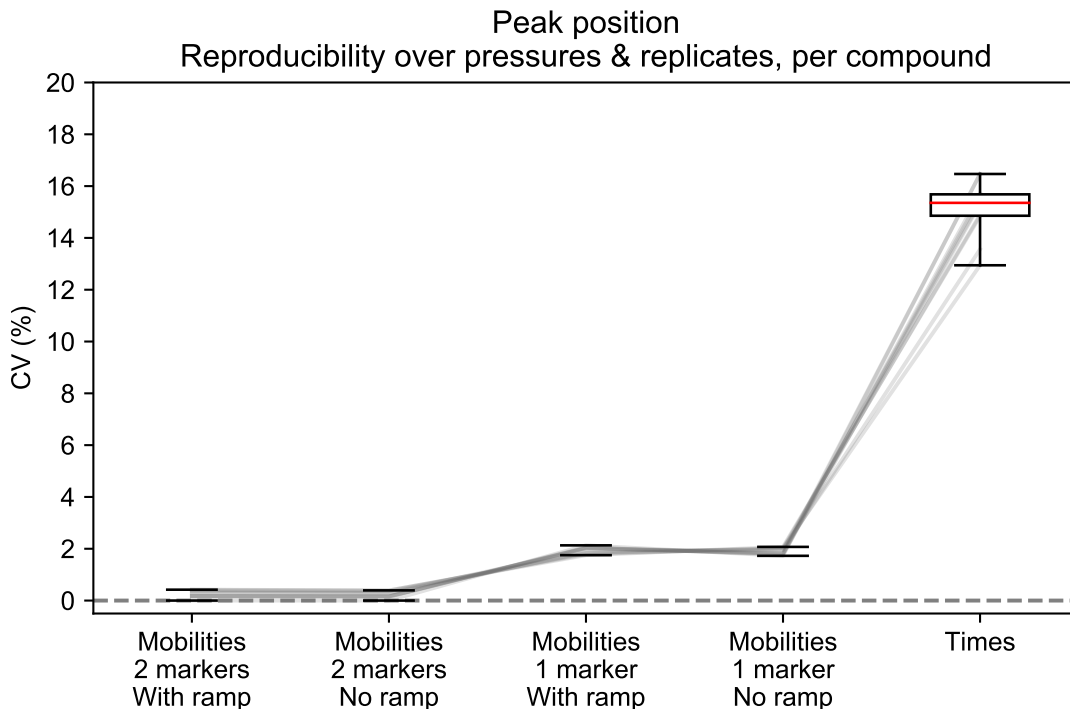
18

Figure 7: Precision of observed mobilities measured under a ramp.

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 Figure 6 are most likely due to variability *in the original* determination. These new mobility values for the chosen standards are available in the supplementary material. Notice still that this variability happens *within* the same set of experiences, and somewhat higher variability should be expected in inter-laboratory comparisons.

As one would hope 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 about a factor of ~ 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 regime of the ESI source. The mix of 15 compounds was analysed under four different pressures (30, 50, 70, and 90 mbar), each run in triplicate, resulting in a array of migration times $t_{c,p,i}$ of migration times for each compound c, pressure p and replicate i, and another one of peak areas $A_{c,p,i}$.
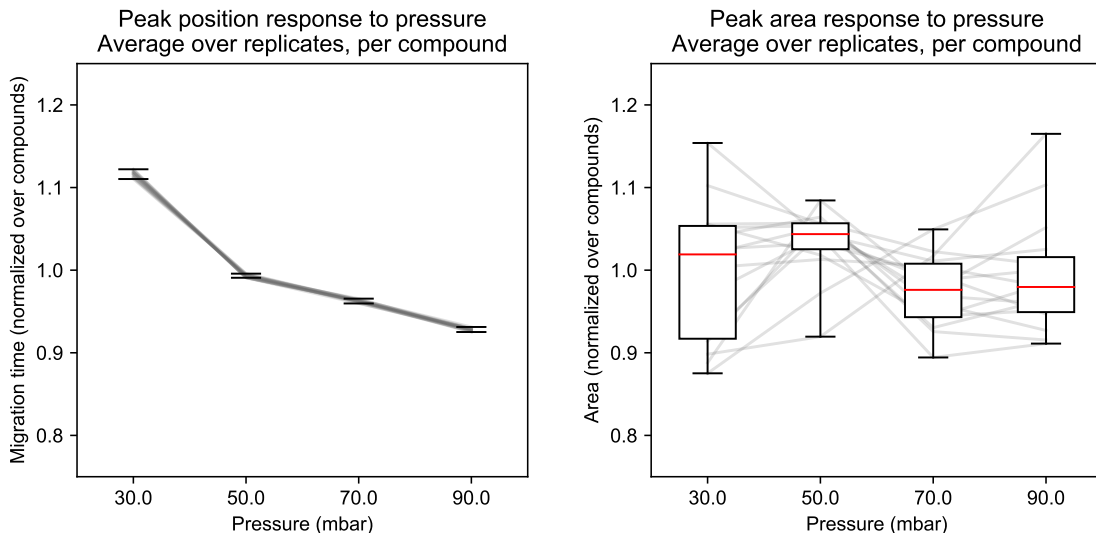
Figure 8: Migration times and peak areas as a function of pressure.

To follow each compound along the different pressure, the replicates were averaged out,

$$t_{c,p}^{(\text{average})} = \sum_{i}^{\text{replicates}} t_{c,p,i}/N_{\text{replicates}}. \tag{47}$$

and in order to compare the compounds against each other, normalized by their own mean over all pressures,

$$t_{c,p}^{(\text{normalized})} = \frac{t_{c,p}^{(\text{average})}}{\sum_{p'}^{\text{pressures}} t_{c,p'}^{(\text{average})}/N_{\text{pressures}}}. \tag{48}$$

The same transformations were applied to obtain an array of normalized areas per compound and pressure, $A_{c,p}^{(normalized)}$. 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.

On the left of Figure 8 we show the box-plot for these normalized migration times. Notice 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. The right-hand plot in Figure 8 shows the same but for peak areas. The variability is logically higher than it was for the peak positions, but clearly not trend is present on the data. The results are compatible with the ESI operating in mass mode, the peak areas being 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. As by now we know that different pressures should provide the same areas, we average both over
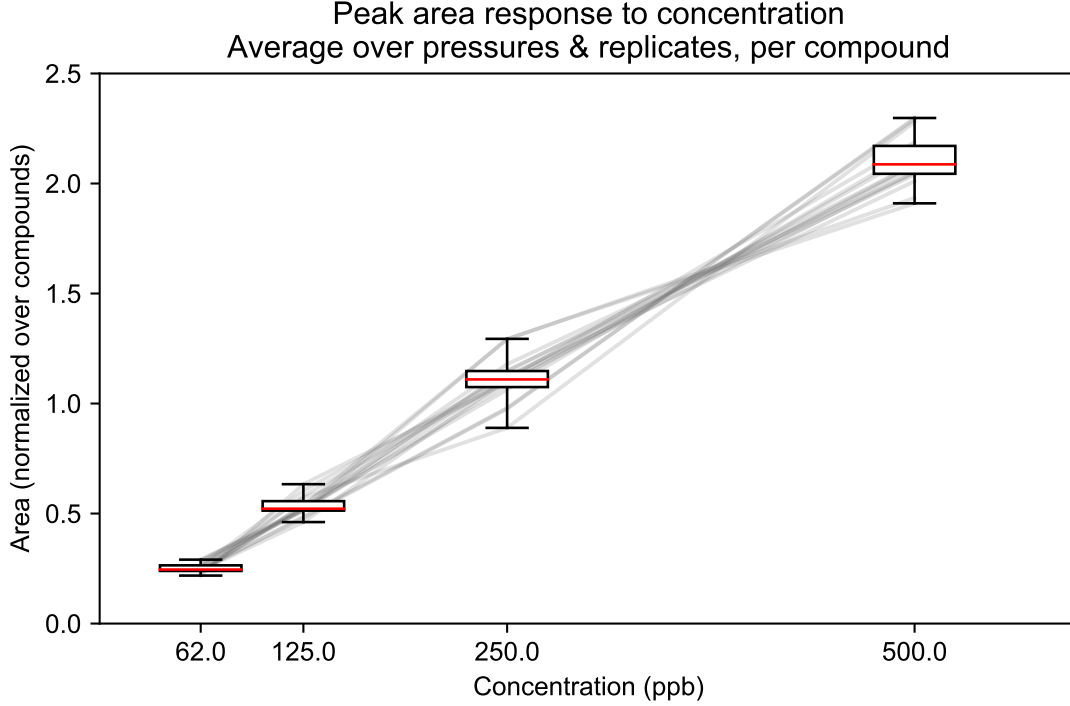
20

Figure 9: Electropherogram peak areas as a function of concentration.

replicates and pressures,

$$A_{c,\rho}^{(\text{average})} = \sum_{p}^{\text{pressures}} \sum_{i}^{\text{replicates}} A_{c,\rho,p,i} / \left(N_{\text{replicates}} \cdot N_{\text{pressures}} \right), \tag{49}$$

and normalize over the free variable, the concentration ρ,

$$A_{c,\rho}^{(\text{normalized})} = \frac{A_{c,\rho}^{(\text{average})}}{\sum_{\rho'}^{\text{concentrations}} A_{c,\rho'}^{(\text{average})} / N_{\text{concentrations}}}. \tag{50}$$

In Figure 9 we can see the linear response of peak areas to concentrations.

With the characterization of the ESI and detection in our hand, we finally move to our main interest, the integration of areas in the mobilogram. The mixture 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 analysed 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 10 displays the results. When applying the correct transformation for
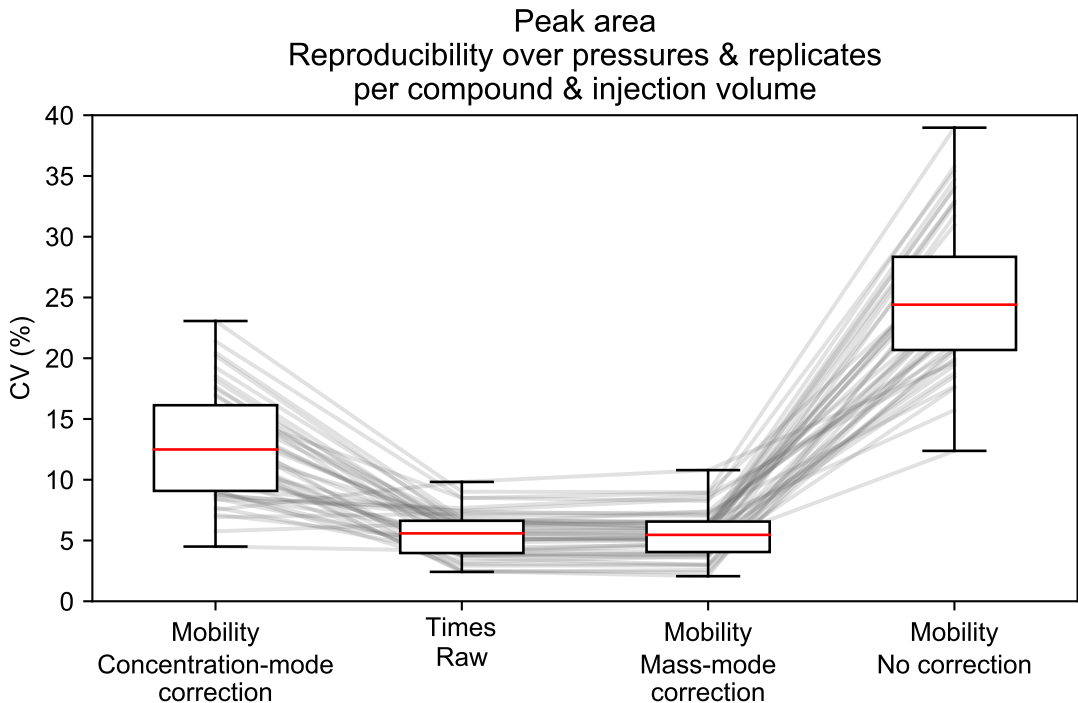
21

Figure 10: Converted peak area precision.

our case ($\times t^2$, following Table 1) the CVs remain the same between the peaks in the electropherograms and the ones 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 11 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 reproducible areas, in which case the transformed mobilograms will perform just as well as the electropherograms. Of course, while having the advantage of permitting the identification of peaks by their position, using libraries of known mobilities.

4.4 Metabolomics application

The development of simplified assays for safety assessment is a key element within the changes taking place over the last decade in the field of chemical toxicology testing. By moving from the classical observation of apical endpoints in animals towards cheaper and faster assays performed on cell cultures, it is possible to cut costs and save time, paving the way to increased-throughput testing [15]. When it comes to toxicity assessment of molecules with neurotoxic potential, astrocytes are an appealing model system, since their activation upon exposure to different neuroinflammatory triggers can take them to either neurotoxic
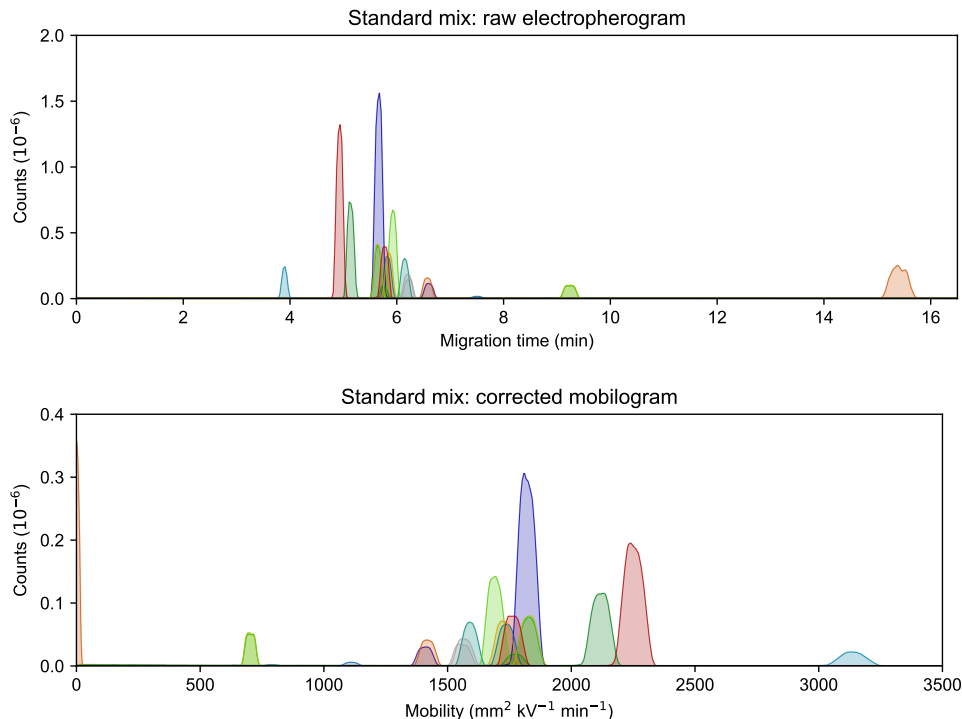
Figure 11: Electropherogram to mobilogram.

or neurotrophic states [16]. To check this approach, and as a first proof-of-concept, 2D-cultures of astrocytes were exposed to different natural neuroinflammatory triggers, namely interleukin 1β (IL-1β), tumoral necrosis factor α (TNFα), and lipopolysaccharide (LPS). 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 the previously described CE-MS approach.

First, raw Agilent .d files were transformed to the open format mzML. Then the files were either directly imported into the Progenesis QI peak-picking software, or transformed (and area-corrected) into the effective mobility scale by ROMANCE prior to the peak-picking step. 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 (PQN) was applied to both datasets, a widely used normalization technique in metabolomics [17]. Drift and other analytical effects were in turn corrected with the inclusion of quality control (QC) samples [18] to ensure analytical consistency, used to apply a principal component based correction to cancel out the sources of variability between the QCs [19].

Running a principal component analysis on the peaks extracted from the electropherograms, we obtain the score plot shown in the top half of Figure 12. We can observe that each sample set is well clustered, and that the first component captures the largest part of the inflammatory triggers' effect on the metabolic status of the astrocytes, distinguishing all three treatments from the control group. Additionally, the second component finds an effect
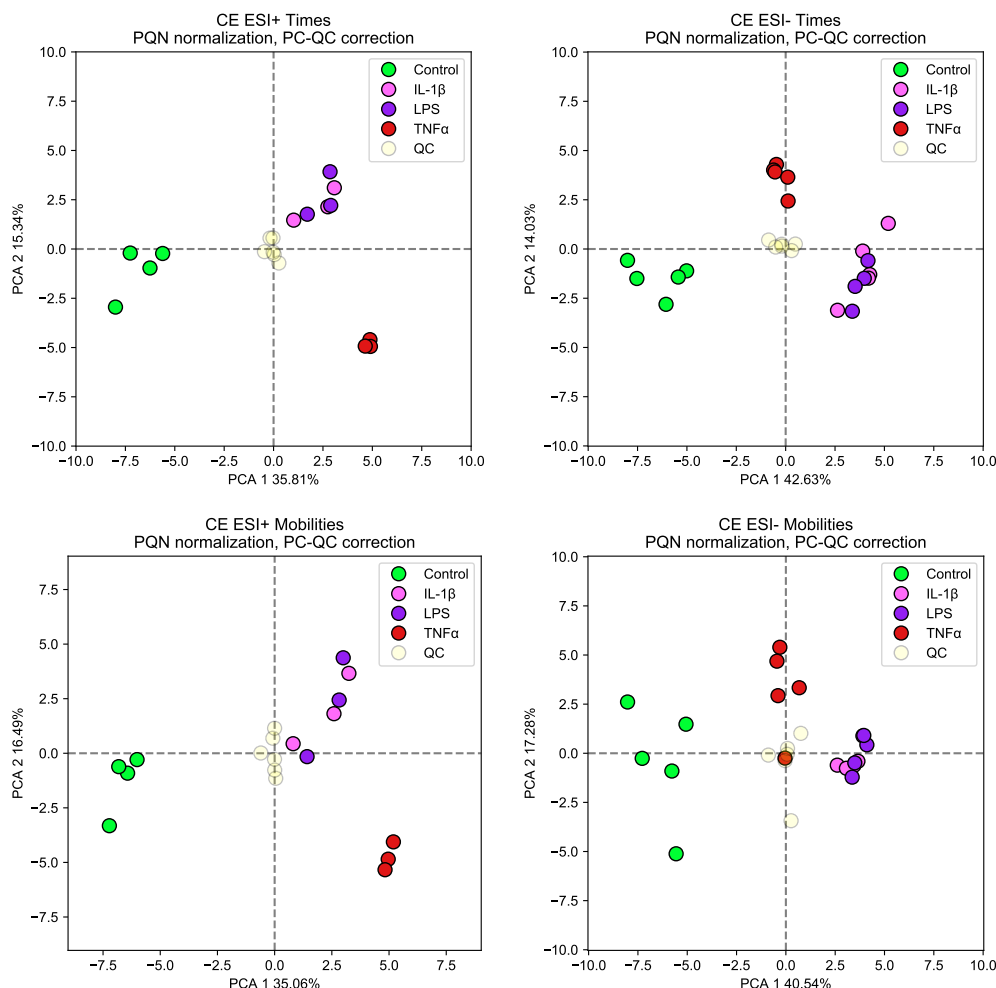
Figure 12: Metabolomics study PCA, raw electropherogram data vs. corrected mobilogram data.

separating the TNFα from the other two groups (IL-1β and LPS), which remain clustered together.

The lower half of Figure 12 shows a PCA of the same samples after conversion by RO-MANCE, using mass-mode area correction. Following the expectations from the results of Figure 10 on standards, the score plots are fundamentally equivalent before and after conversion. The corresponding identified metabolites, together with their loadings, are available in the supplementary material. 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 Table C in the supplementary information.

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 [6], 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 ionization and 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 have reliable quantitative data. Finally, we have seen that with the current version of ROMANCE the worfklow 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.

Acknowledgements

Funding

References

[1] R. Ramautar, G. W. Somsen, and G. J. de Jong, "CE–MS for metabolomics: developments and applications in the period 2014–2016," *Electrophoresis*, vol. 38, no. 1, pp. 190–202, 2017.

[2] A. García, J. Godzien, Á. López-Gonzálvez, and C. Barbas, "Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics," *Bioanalysis*, vol. 9, no. 1, pp. 99–130, 2017.

[3] W. Zhang, T. Hankemeier, and R. Ramautar, "Next-generation capillary electrophoresis–mass spectrometry approaches in metabolomics," *Current opinion in biotechnology*, vol. 43, pp. 1–7, 2017.

[4] P. Schmitt-Kopplin and A. Fekete, "The CE way of thinking," in *Capillary electrophoresis*, pp. 611–629, Springer, 2008.

[5] N. Drouin, J. Pezzatti, Y. Gagnebin, V. Gonzalez-Ruiz, J. Schappler, and S. Rudaz, "Effective mobility as a robust criterion for compound annotation and identification in metabolomics: Toward a mobility-based library," *Analytica chimica acta*, vol. 1032, pp. 178–187, 2018.

[6] V. González-Ruiz, Y. Gagnebin, N. Drouin, S. Codesido, S. Rudaz, and J. Schappler, "ROMANCE: A new software tool to improve data robustness and feature identification in CE-MS metabolomics," *Electrophoresis*, vol. 39, no. 9-10, pp. 1222–1232, 2018.

[7] M. Mammen, I. J. Colton, J. D. Carbeck, R. Bradley, and G. M. Whitesides, "Representing primary electrophoretic data in the 1/time domain: comparison to representations in the time domain," *Analytical chemistry*, vol. 69, no. 11, pp. 2165–2170, 1997.

[8] N. Ikuta, Y. Yamada, T. Yoshiyama, and T. Hirokawa, "New method for standardization of electropherograms obtained in capillary zone electrophoresis," *Journal of Chromatography A*, vol. 894, no. 1-2, pp. 11–17, 2000.

[9] J. Chamieh, M. Martin, and H. Cottet, "Quantitative analysis in capillary electrophoresis: transformation of raw electropherograms into continuous distributions," *Analytical chemistry*, vol. 87, no. 2, pp. 1050–1057, 2015.

[10] N. Ikuta, Y. Yamada, and T. Hirokawa, "Electropherogram of capillary zone electrophoresis with effective mobility axis as a transverse axis and its analytical utility I. transformation applying the hypothetical electroosmotic flow," *ELECTROPHORESIS*, vol. 21, no. 2, pp. 360–366, 2000.

[11] G. Bonvin, J. Schappler, and S. Rudaz, "Capillary electrophoresis–electrospray ionization-mass spectrometry interfaces: Fundamental concepts and technical developments," *Journal of Chromatography A*, vol. 1267, pp. 17–31, 2012.

[12] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W. H. Tang, A. Römpp, S. Neumann, A. D. Pizarro, *et al.*, "mzML – a community standard for mass spectrometry data," *Molecular & Cellular Proteomics*, vol. 10, no. 1, pp. R110–000133, 2011.

[13] R. Adusumilli and P. Mallick, "Data conversion with ProteoWizard msConvert," in *Proteomics*, pp. 339–368, Springer, 2017.

[14] B. MacLean, D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, and M. J. MacCoss, "Skyline: an open source document editor for creating and analyzing targeted proteomics experiments," *Bioinformatics*, vol. 26, pp. 966–968, 02 2010.

[15] D. Krewski, D. Acosta Jr, M. Andersen, H. Anderson, J. C. Bailar III, K. Boekelheide, R. Brent, G. Charnley, V. G. Cheung, S. Green Jr, *et al.*, "Toxicity testing in the 21st

century: a vision and a strategy," *Journal of Toxicology and Environmental Health, Part B*, vol. 13, no. 2-4, pp. 51–138, 2010.

[16] V. Parpura, M. T. Heneka, V. Montana, S. H. Oliet, A. Schousboe, P. G. Haydon, R. F. Stout Jr, D. C. Spray, A. Reichenbach, T. Pannicke, *et al.*, "Glial cells in (patho) physiology," *Journal of neurochemistry*, vol. 121, no. 1, pp. 4–27, 2012.

[17] F. Dieterle, A. Ross, G. Schlotterbeck, and H. Senn, "Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics," *Analytical chemistry*, vol. 78, no. 13, pp. 4281–4290, 2006.

[18] D. Broadhurst, R. Goodacre, S. N. Reinke, J. Kuligowski, I. D. Wilson, M. R. Lewis, and W. B. Dunn, "Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies," *Metabolomics*, vol. 14, no. 6, p. 72, 2018.

[19] J. Sun, "Statistical analysis of NIR data: data pretreatment," *Journal of Chemometrics*, vol. 11, no. 6, pp. 525–532, 1997.

Supplementary information

A   Table of standard compound mobilities

ROMANCE v1 values taken from [6]. ROMANCE v2 values taken from 2-marker formula results in section 4.2.

Compound	Mobility ROMANCE V1 ($mm^2\,kV^{-1}\,min^{-1}$)	Mobility ROMANCE V2 ($mm^2\,kV^{-1}\,min^{-1}$)
Acetylcholine	1913	1879
Adenine	1865	1795
Choline	2184	2175
L-histidine	1760	1747
L-lysine	1941	1881
Nicotinamide	1830	1773
Phenethylamine	1839	1824
Thiamine	2333	2321
Tryptamine	1669	1616
Tyramine	1626	1626
3-methoxytyramine	N/A	1411
Agmatine	N/A	3119
Amphetamine	N/A	1598
GABA	N/A	1774
L-glutamine	N/A	1833

B Table of standard compound precursors

Compound	Precursor ion (m/z)	Product ion (m/z)	Collision energy (V)	Dwell time (ms)
Acetylcholine	146.1	87.2	13	14
Adenine	136.2	118.9	25	14
Choline	104.1	60.1	17	14
L-histidine	156.2	110.0	13	14
L-lysine	147.1	84.1	17	14
Nicotinamide	123.1	80.1	17	14
Phenethylamine	122.1	105.1	9	14
Thiamine	265.1	122.0	13	14
Tryptamine	161.1	144.2	9	14
Tyramine	138.1	121.1	9	14
3-methoxytyramine	168.1	151.2	9	14
Agmatine	131.1	60.1	9	14
Amphetamine	136.1	91.0	17	14
GABA	104.2	87.0	9	14
L-glutamine	147.1	84.1	17	14

C Tables of PCA loadings

List of identified metabolites and mobility-based PCA loadings (see Figure 12) from the untargeted metabolomics analysis (on astrocyte inflammation).

C.1 ESI− mode

Metabolites	Component 1	Component 2
D-pantothenic acid	−0.08	+0.46
5-oxo-D-proline	−0.24	−0.33
N-acetyl-DL-serine	+0.18	−0.33
N-acetylneuraminic acid	−0.17	−0.31
SN-glycerol 3-phosphate	−0.20	+0.27
Gluconic acid	+0.01	−0.33
6-phosphogluconic acid	+0.29	−0.14
N-acetyl-L-aspartic acid	−0.31	−0.07
Dihydroxyacetone phosphate	+0.30	+0.11
CDP-ethanolamine	+0.18	+0.24
D-glucose 6-phosphate	+0.29	+0.02
Fructose 1,6-biphosphate	+0.29	−0.06
(S)-(-)-L-malic acid	+0.27	+0.04
3-methyl-2-oxovaleric acid	−0.17	−0.19
Citric acid	+0.21	−0.14
D-ribose 5-phosphate	+0.23	−0.07
Guanosine 5'-diphosphate	−0.07	−0.20
D-(+)-galactose	−0.14	+0.12
Creatine phosphate	+0.15	+0.12
(R,R)-L-(+)-tartaric acid	+0.16	−0.07
3-(4-hydroxyphenyl)lactic acid	−0.15	+0.05
Phenylacetic acid	+0.05	−0.15
Adenosine 3',5'-diphosphate	−0.11	−0.10
Cytidine 2',3'-cyclic mono-phosphate	+0.15	−0.04
Adenosine 5'-monophosphate	+0.13	−0.04
Uridine 5'-diphosphate	+0.10	−0.08
N-acetyl-DL-glutamic acid	−0.01	−0.10
Uridine-5-monophosphate	+0.06	−0.03

C.2 ESI+ mode

Metabolites	Component 1	Component 2
5'-deoxyadenosine	+0.22	−0.33
Alpha-aminoadipic acid	+0.16	+0.35
L-tyrosine	−0.31	−0.12
Thiamine pyrophosphate	+0.05	−0.33
L-anserine	+0.30	−0.12
L-histidine	−0.31	−0.08
L-asparagine	−0.23	+0.21
Folic acid	−0.28	−0.02
D-mannosamine	+0.15	+0.24
Neopterin	+0.27	−0.02
Carnosine	+0.25	−0.09
L-phenylalanine	+0.16	+0.21
L-tryptophan	−0.24	−0.11
L-arginine	−0.06	+0.26
L-lysine	−0.06	+0.24
N,N-dimethylarginine	−0.06	+0.23
O-acetyl-L-carnitine	+0.21	−0.08
Biopterin	+0.18	−0.14
Nicotinic acid	−0.13	−0.18
Citrulline	+0.03	+0.22
5'-methylthioadenosine	+0.21	−0.05
D-(+)-galactosamine	+0.06	+0.20
S-(5'-adenosyl)-L-homocysteine	−0.02	−0.18
N-alpha-acetyl-L-lysine	−0.15	−0.08
Glutathione	+0.16	−0.02
Spermidine	+0.04	+0.15
Adenosine	−0.04	−0.15
N-amidino-L-aspartic acid	−0.13	+0.01
Theobromine	−0.07	−0.11
Deoxycarnitine	+0.13	+0.01
Cytidine	−0.02	−0.12
Pyridoxal	−0.09	−0.00
Dopamine	+0.02	−0.08
Tyramine	−0.07	−0.04
Thiamine	−0.03	−0.07
Pyridoxamine	−0.05	−0.03
L-kynurenine	−0.03	−0.04
Isobutyryl-L-carnitine	+0.00	+0.02