New methods for the characterization of essential distributions

Niezen, L.E.

Publication date
2023
Document Version
Final published version

New methods for the characterization of essential distributions

Leon E. Niezen

New methods for the characterization of essential distributions

ACADEMISCH PROEFSCHRIFT

Ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op woensdag 10 mei 2023, te 14.00 uur

door
Leon Ewoud Niezen
geboren te Aalsmeer

Promotiecommissie

Promotores:	prof. dr. ir. P.J. Schoenmakers *Universiteit van Amsterdam* prof. dr. G.W. Somsen *Vrije Universiteit Amsterdam*
Copromotores:	dr. B.W.J. Pirok *Universiteit van Amsterdam* dr. B.B.P. Staal *BASF SE*
Overige leden:	dr. A.M. Striegel *NIST* dr. W. Radke *PSS / Agilent* prof. dr. R.A.H. Peters *Universiteit van Amsterdam* dr. A. Gargano *Universiteit van Amsterdam* prof. dr. S. Woutersen *Universiteit van Amsterdam* prof. dr. A.K. Smilde *Universiteit van Amsterdam* prof. dr. K.U. Loos *Rijksuniversiteit Groningen*

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The work in this thesis was financially supported by the UNMATCHED project, which was supported by BASF, DSM and Nouryon and received funding from the Dutch Research Council (NWO) in the framework of the Innovation Fund for Chemistry (CHIPP Project 731.017.303) and from the Ministry of Economic Affairs in the framework of the "TKI-toeslagregeling".

This work was performed in the context of the Chemometrics and Advanced Separations Team (CAST) within the Centre for Analytical Sciences Amsterdam (CASA). The valuable contributions of the CAST members are gratefully acknowledged.

Industrial partners:

Academic partners:

Table of Contents

Chapter 1

Introduction

Prelude

Preparative liquid chromatography was introduced around the start of the 20th century. A few decades after its introduction it was realized that chromatography could offer a separation selectivity unmatched by any other technique. This selectivity meant the technique could be used to obtain pure substances from mixtures of components. Previously, such pure substances were acquired by extraction and crystallization, which was an arduous task, as it could take years to obtain the amount (several grams) of purified substance required for the reactions that were to be performed afterwards. Luckily, this changed with the introduction of chromatography, which could be performed significantly faster. Many developments occurred in the latter half of the century, aided by the introduction of the first microprocessors and the transfer from analogue to digital. As a result, the present-day situation is significantly different. Whilst the ability to perform large-scale purification, or elimination, of particular compounds is still desirable, many separations are carried out at much smaller, so-called "analytical" scale and are combined with statistics and chemometrics, so that the information gathered from a small amount of sample is sufficient to answer contemporary questions. Chromatography is invaluable for resolving questions regarding the quality of water, the degradation of paintings, the health of an individual, the cause of a disease, the development of drugs and vaccines, the properties of a material, and many other important aspects of science and society.

Introduction

In the field of separation science no process is more prevalent than chromatography. At the heart of any chromatographic separation is the distribution of compounds over two phases, usually one stationary (the stationary phase) and the other mobile (the mobile phase). In case a liquid mobile phase is used, the method is referred to as liquid chromatography (LC). The stationary phase may be either entirely a liquid, a solid, or a liquid immobilized on a solid support. Of these the latter two are currently most common. In its simplest form LC is realized by applying sample at the bottom of a sheet of inert material coated with a layer of adsorbent material (the stationary phase) and then allowing a solvent (the mobile phase) to be drawn through the sample and up the sheet through capillary action. A separation may be achieved because analytes in the mixture ascend the plate at different rates based on the type of stationary and mobile phase and the affinity of the analyte towards one phase relative to the other. This implementation of LC is termed thin-layer chromatography (TLC). TLC is still often employed, for example in organic-chemistry laboratories, due to its simplicity and low costs, allowing for quick and simple reaction monitoring. However, because TLC is challenging to automate and cannot[1] separate highly complex mixtures, LC is more commonly implemented as a technique in which the mobile phase and the unknown mixture are flushed through a steel or glass tube, termed a column, either by gravity or, preferably, by using a pump. Irrespective of the type of column used, this format is termed column LC, henceforth referred to simply as LC. It was the first documented implementation of chromatography, invented around the start of the previous century by Tswett [1].

In LC the stationary phase can take many different forms [2–4]. Most commonly it is a layer of adsorbent material coated onto fine porous or non-porous inert particles that are packed into the column. In other cases the particles themselves are the stationary phase, or monolithic structures synthesized directly inside the column act as the stationary phase [2,3]. Sometimes no particles are used at all and the adsorbent layer is created on the inner wall of the column [4].

[1] At least not unless very specific adaptations are made that make the instrumental set-up significantly more complex

1.1. Basics of migration in (isocratic) chromatography

Irrespective of the LC format (plate, column) or the type of stationary phase (particles, layer, monolith), as long as there is a constant homogenous flow of mobile phase moving in a constant direction through a homogenous medium (and if the separation conditions do not change) then the mean velocity at which the analyte migrates (u_i) through the column remains constant and is a fraction (R) of the velocity at which the solvent migrates (u_m), or

$$u_i = Ru_m \tag{1.1}$$

Hence, R is a direct measure of the migration rate of the analyte. If R does not depend on u_m, then the time it takes to obtain a separation is, logically, shortened when u_m is increased, *i.e.* when the volumetric flowrate of the mobile phase through the column (F) is increased. For an empty column with a volume V, given by the cross-sectional area (A) and the length of the column (L) as $V = AL$, F is the volume pumped over a given time (t) and is related to the average linear velocity of the mobile phase through the empty column (u_m) as $F = \frac{V}{t} = Au_m$. When the column contains a solid medium the volume that is available is reduced by the volume of the solid. This is captured in the "porosity" (ε). The velocity through the column is reduced because of ε as, $F = \frac{V\varepsilon}{t_0} = Au_m\varepsilon$. This reduced volume is referred to as the void volume of the column ($V_0 = V\varepsilon$, or $V_m = V\varepsilon$, note that both V_0 and V_m are used interchangeably throughout this thesis). The time it takes for the mobile phase to move through this volume is referred to as the void time (t_0), given by,

$$t_0 = \frac{V_0}{F} = \frac{L}{u} \tag{1.2}$$

An analyte moves through the column at a velocity that is a fraction R of the mobile phase and takes a different time to move through the column, which is commonly referred to as the retention time (t_R) and given by

$$t_R = \frac{V_0}{RF} = \frac{V_R}{F} = \frac{L}{u_i} \tag{1.3}$$

The times t_0 and t_R are the experimental parameters determined from a separation. For a single component, and in the absence of peak dispersion, or if dispersion results in symmetrical peaks, t_R corresponds to the top of the peak. In reality peaks always

show some asymmetry and t_R is better described as the first (uncorrected or raw) moment of the distribution, *i.e.* the *centre of mass* of the peak. Typically, t_0 is measured by injecting a low-molecular-weight unretained tracer component. Assuming a homogenous column packing, *i.e.* a constant column permeability over the length of the column, t_0 is representative of V_0 or V_m. However, if there are zones within the packing that cannot be entered by the tracer, as might be the case for a porous stationary phase, then t_0 is influenced by the experimental conditions [5–7]. In such a case t_0 may be a measure for the entire mobile phase volume in the column, *i.e.* the sum of the stagnant volume inside the pores (V_p) and the moving interparticle volume (V_i), or it may correspond more closely to only V_i. Throughout this thesis we assume that t_0 is representative of the entire volume V_0.

1.1.1. Size-exclusion chromatography

Particularly for high-molecular-weight analytes, *i.e.* macromolecules such as proteins or polymers, V_p may not be entirely accessible due to the size of the analyte. This is the basis for a technique called size-exclusion chromatography (SEC) [8–12]. In SEC the analyte elutes at a time given by,

$$t_e = \frac{V_i + K_{\mathrm{SEC}} V_p}{F} \tag{1.4}$$

where K_{SEC} is the SEC distribution coefficient ($K_{\mathrm{SEC}} = \frac{V_e - V_i}{V_p}$), which describes how much of V_p is accessible to the analyte (or to what extent an analyte is excluded from V_p). In SEC the terms elution time (t_e) and elution volume ($V_e = V_i + K_{\mathrm{SEC}} V_p$) are used, instead of "retention" time or volume, because there is no retention in SEC. For SEC, the mobile and stationary phase are chosen such as to minimize interactions of the analyte with the stationary phase. Because analytes cannot enter all of V_p, they elute before t_0. This corresponds to an R in **Equation 1.1** that is larger than one. Because the exclusion of the analyte depends on its hydrodynamic volume or its size, SEC can provide information on the molecular weight of the analyte. To obtain this information a calibration must be performed by using standards of known molecular weight, or the SEC separation must be coupled with light-scattering or viscometric techniques. Typically SEC yields information on a molecular-weight distribution (MWD), rather than a single value for the molecular weight. This is because most (synthetic) polymers consist of chains of different length, and because in SEC a

packing is used with a broad pore-size distribution. The MWD obtained from SEC is often translated into averages. Most commonly used are the number average and weight average molecular weights. These are defined as,

$$M_n = \frac{\sum_{i=1}^{\infty} N_i M_i}{\sum_{i=1}^{\infty} N_i} \tag{1.5}$$

$$M_w = \frac{\sum_{i=1}^{\infty} N_i M_i^2}{\sum_{i=1}^{\infty} N_i M_i} \tag{1.6}$$

Where N is the number of molecules and M is the molecular mass of said molecules. The ratio of these two averages ($\frac{M_w}{M_n}$) is indicative of the width of the distribution of molecular weight in the sample. This ratio used to be, and still is commonly, referred to as the polydispersity index (PDI), although nowadays also the term dispersity ($Đ_M$) is used. In most cases the obtainable PDI depends on the polymerization technique. For free-radical and condensation polymerization relatively large values are obtained (larger than 1.5 in both cases), while with other methods, such as ionic polymerization, narrow distributions can be obtained.

1.1.2. Interaction liquid chromatography

For small[2] compounds exclusion effects usually do not play a significant role. This is a simpler situation, because t_R now reflects the analyte distribution between the mobile and stationary phase on a specific column. In this case R corresponds to the equilibrium fraction of the analyte in the mobile phase and $1 - R$ corresponds to the fraction in the stationary phase. Nowadays, the ratio $\frac{1-R}{R}$ is commonly referred to as the retention factor (k), given by:

$$k = \frac{1-R}{R} = \frac{t_R - t_0}{t_0} = \frac{V_R - V_0}{V_0} \tag{1.7}$$

For small components the values of k, t_R and R depend on how much stationary-phase surface area there is available for the analyte to interact with (adsorb onto, or partition into) relative to V_0. As previously mentioned, the stationary phase usually corresponds to a layer of adsorbent material grafted onto an inert packing. Often this layer is accessible to the analyte and is of a finite thickness, so that it is better described as a volume or a volume equivalent, V_s, rather than a surface, although

[2] Relative to the average pore diameter of the packing.

both are used interchangeably. The ratio $\frac{V_s}{V_0}$ can change based on the packing inside a column and, for example, on how much of the stationary phase was grafted onto the inert support. To account for such differences, k may be converted to the distribution coefficient (K_{LC}), if two conditions are met. Namely, *i*) the exact volumes or volume equivalents of the mobile and stationary phase must be known, and *ii*) the separation must be performed on an analytical scale, using low concentrations of injected sample, so that it can be assumed that no competition occurs between analytes for interaction with the stationary phase. This latter condition is often referred to as "linear chromatography". Under these conditions the concentration of the analyte in one phase is proportional to its concentration in the other phase, so that the equilibrium isotherm is a straight line with slope $k\beta^{-1}$ or,

$$k = \frac{N_S}{N_m} = \frac{V_S}{V_m} K_{\mathrm{LC}} = \beta K_{\mathrm{LC}} \tag{1.8}$$

Where the subscripts s and m denote the stationary and mobile phase, respectively; N is the number of molecules in the respective phase, V is the volume (or volume equivalent) of the respective phase and β is the volumetric phase ratio of the column (sometimes also defined in literature as $\Phi = \beta^{-1}$). In practice, β and t_0 are difficult to determine precisely, and may change based on experimental conditions [13–15].

1.1.3. Interaction polymer chromatography

So far two extremes have been discussed. Namely, *i*) large analytes that do not interact with the stationary phase and undergo a separation solely by being excluded from (part of) V_p, and *ii)* small compounds that can enter all of V_p and are separated by means of differing types or strengths of interactions with the stationary phase. However, depending on the experimental conditions and the analyte, other situations can be imagined. For example, if the pore diameter becomes sufficiently small, then exclusion will start to play a role even for small analytes. *Vice versa*, if the mobile and stationary phase and temperature are chosen such that interaction with the stationary phase is promoted, then large analytes can experience both exclusion and interaction. Of these two examples the former is rare, because most columns that are in use today feature (relatively) large pore sizes. The latter situation is more common and may be referred to as liquid adsorption chromatography (LAC) or interaction polymer chromatography (IPC). Because there is also a stationary phase

layer on the outside of the porous packing material, a macromolecule can be simultaneously excluded from V_p, while also being delayed due to interactions with the fraction of V_s that is in contact with the interparticle volume. Because not all analytes are fully excluded from all pores, each analyte will have a different fraction of V_s available to interact with. However, the effect of a difference in accessible V_s on t_R is not usually observed, because the strength of interaction with the stationary phase is magnified by the size of the analyte. In other words, those analytes that should be most affected by a lower amount of accessible stationary phase also tend to be fully retained. The differences in accessible V_s (V_s') likely only matters when interactions are very weak. Experimental examples that might correspond to this situation are included in Chapter 4. To account for both exclusion and interaction, **Equations 1.4**, **1.7** and **1.8** can be combined, and the elution volume is given by

$$V_e = V_i + K_{\mathrm{SEC}}V_p + K_{\mathrm{LC}}V_s \tag{1.9}$$

The above equation assumes that the probability that an analyte is excluded from the pore volume is separate from the probability that the analyte interacts with the stationary phase. When K_{SEC} is considered to be not only representative of the amount of exclusion from pores, but also representative of potential differences in V_s', then both events can be described by a single distribution coefficient as $K = K_{\mathrm{SEC}}K_{\mathrm{LC}}$. The potential situations described by **Equation 1.9** are illustrated in **Figure 1.1**

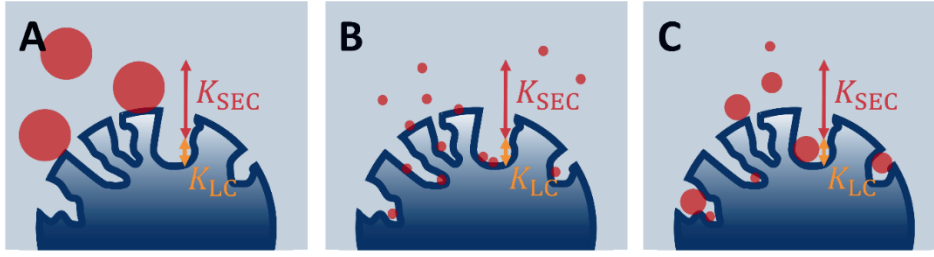

Figure 1.1: Situations corresponding to $\boldsymbol{K_{\mathrm{LC}} > 1}$ and A) $\boldsymbol{K_{\mathrm{SEC}} = 0}$, B) $\boldsymbol{K_{\mathrm{SEC}} = 1}$, and C) $\boldsymbol{0 < K_{\mathrm{SEC}} < 1}$.

The distribution coefficients (K, K_{SEC} and K_{LC}) may be described in terms of the partial molar change in Gibbs free energy, through the van 't Hoff equation:

$$K = e^{-\frac{\Delta G^\circ}{N_A k_B T}} = e^{-\frac{\Delta H^\circ}{N_A k_B T}\frac{1}{T} + \frac{\Delta S^\circ}{N_A k_B}} \quad (1.10)$$

where ΔG°, ΔH° and ΔS° are the partial molar changes in Gibbs free energy, enthalpy, and entropy, respectively, for transfer of the analyte from the mobile phase to the stationary phase. N_A and k_B are Avogadro's and Boltzmann's constant, and T is the temperature in Kelvin. ΔH°, and ΔS° (and V_s) are often considered invariant with T, and pressure (P), which is not always correct [16]. The logarithmic form of the van 't Hoff equation ($\ln k = -\frac{\Delta H^\circ}{N_A k_B}\frac{1}{T} + \frac{\Delta S^\circ}{N_A k_B} + \ln \beta$), or a mathematically equivalent form ($\log k = a + b\frac{1}{T}$), have also been used to model the effect of T on k. Provided that V_s, ΔH°, and ΔS° are invariant with T, and assuming the resulting change in pressure does not significantly affect these parameters, there is a linear relationship between $\ln k$ and $\frac{1}{T}$, with a slope of $-\frac{\Delta H^\circ}{N_A k_B}$ and a *y*-intercept of $\frac{\Delta S^\circ}{N_A k_B} + \ln \beta$ These assumptions are often reasonable over the range of temperatures used in LC (5 to 80°C), but are not always valid [16]. Typically the slope increases with the molecular weight of the analyte, so that small temperature differences result in large changes in k for macromolecules. This effect is used in Chapter 2 of this thesis to demonstrate that polymers can relatively easily be trapped by creating only small differences in temperature.

When the statistical definition of entropy is used, *i.e.* $S = k_b \ln \Omega$, where Ω is the number of microstates of the system, it follows that, according to our definition of ΔS°, in SEC $\Delta S^\circ < 0$, since there is a reduction in the number of conformational states a polymer can have upon entering a pore. On the other hand, a change in ΔH° upon entering a pore is considered negligible ($\Delta H^\circ = 0$)[3]. When an analyte instead interacts with (*i.e.* adsorbs on, or partitions into) the stationary phase, it is also expected that $\Delta S^\circ < 0$, as the number of conformations near the surface of the stationary phase is reduced relative to the number of possible conformations in solution. In some literature this effect is considered negligible. However, to obtain reasonable retention factors ($1 < k < 10$), experimental parameters are implicitly tuned such that the chromatographic separation is performed under conditions where ΔH° and ΔS° are (nearly) compensated. This is the case irrespective of the size of the analyte, indicating that the latter change in ΔS° is not negligible. Finally, for

[3] Note that these changes are not indicative of the sign of ΔG°.

interactions with the stationary phase $\Delta H° < 0$. With these assumptions K can be written as:

$$K = K_{\mathrm{SEC}} K_{\mathrm{LC}} \cong e^{\frac{\Delta S°_{\mathrm{SEC}}}{N_A k_B}} e^{-\frac{\Delta H°}{N_A k_B T} + \frac{\Delta S°_{\mathrm{LC}}}{N_A k_B}} \quad (1.11)$$

Where the changes in $\Delta S°$ due to exclusion, and due to adsorption or partitioning into the stationary phase are separated into $\Delta S°_{\mathrm{SEC}}$ and $\Delta S°_{\mathrm{LC}}$. For a column packed with non-porous particles, or for small analytes that do not experience exclusion, $K_{\mathrm{SEC}} = 1$ and so the equation (and situation) are easier to describe

$$K = K_{\mathrm{LC}} \cong e^{-\frac{\Delta H°}{N_A k_B T} + \frac{\Delta S°_{\mathrm{LC}}}{N_A k_B}} \quad (1.12)$$

Furthermore, in such a packing differences in V_s with analyte size do not affect the retention and t_0 should be more representative of V_0 (V_i) than in a porous packing.

1.1.4. Liquid chromatography at critical conditions

For certain experimental conditions it can be observed that the elution order of a homopolymer featuring non-interacting end groups becomes independent of molecular weight. These conditions are often referred to as "critical", and the method is thus called liquid chromatography at critical conditions (LCCC) [17–21]. These conditions can be found by fine-tuning the mobile-phase composition (*e.g.* by varying the volume fraction of strong mobile phase modifier, φ, in a binary solvent mixture), T or P. Usually, φ is adjusted while T and P are constant. For a specific binary solvent mixture critical conditions then correspond to φ_{crit}. However, in principle it should be possible to find critical conditions at multiple combinations of mobile-phase composition, temperature and pressure. For example, a larger φ_{crit} will be found at lower temperatures, if retention increases with a decrease in temperature, and *vice versa*. This also means that it is possible to obtain critical conditions while simultaneously changing solvent polarity through the use of ternary solvents [22,23].

At critical conditions the free energy of transfer of a particular type of monomer in the polymer backbone is zero ($\Delta G^o_{\mathrm{mon}} = 0$). In other words, at these conditions $\Delta H°$ scales with analyte molecular weight in the same way that $T\Delta S°$ does. Often LCCC is portrayed as an intermediate mode of chromatography, between exclusion (elution before t_0) and interaction chromatography (elution after t_0). As a result, in some

cases critical conditions are referred to as those conditions where the entropic term from exclusion ($\Delta S°_{\mathrm{SEC}}$) counteracts the enthalpic term ($\Delta H°$) of the interaction with the stationary phase. This can only be the case if interaction with the stationary phase increases the same way as does exclusion. However, in Chapter 3 it is shown that critical conditions can also be obtained on columns packed with non-porous particles, while on a porous packing no conditions could be found that were entirely critical. On a non-porous packing exclusion is absent and so to reach critical conditions $\Delta H^{o}_{\mathrm{LC}}$ must equal $T\Delta S^{o}_{\mathrm{LC}}$, or $K_{\mathrm{LC}} = 1$. Since it is unlikely that $T\Delta S^{o}_{\mathrm{LC}}$ would be absent on a porous packing, it is likely that partial exclusion from the pores actually hinders obtaining fully critical conditions on a porous packing.

One of the primary applications of LCCC for homopolymers is a separation based on end groups [20,24–26]. However, critical conditions do not solely exist for homopolymers that have a backbone consisting of an identical repeating unit. For statistical copolymers critical conditions can also be established [27–30]. For such copolymers the critical conditions will be determined by the chemical composition of the copolymer. However, because most copolymers feature a chemical composition distribution (CCD) the sample will also feature a distribution in φ_{crit}, and so LCCC can only be used to analyse a small (compositional) fraction of such a copolymer. For (large) block copolymers critical conditions do not exist for the entire copolymer [28,31]. However, critical conditions for individual blocks may be established. Therefore, in case of di-block copolymers it is possible to obtain separations based on the block-length of one of the blocks by performing an isocratic separation at φ_{crit} of the other block [31]. For this to be practical, φ_{crit} for one of the blocks must correspond to "exclusion" conditions of the other block, as otherwise an isocratic interaction-based separation will occur for the block that is not at critical conditions and retention would increase too strongly with molecular weight to elute a polydisperse sample. If the block is too large and starts to interact with the stationary phase very broad peaks will be obtained, or the copolymer will not elute from the column. Usually, which of the two modes the non-critical block elutes in can be adjusted based on whether critical conditions are established in normal-phase LC, or the reverse mode, RPLC. LCCC usually works best for a di-block copolymer, because for tri-block or even larger number of segments, the order of the blocks begins to matter. For example, while the middle block in a tri-block copolymer can

be made to not interact with the stationary phase, it still connects the two outside blocks physically, which will affect the separation that is obtained.

1.1.5. Modelling the effect of mobile-phase composition on migration

After selection of the stationary phase, the most common parameter to be adjusted for method development is the mobile-phase composition. A change in k, or the extent of interaction of the analyte with the stationary phase, may be described in terms of φ. It is valuable to know the relationship between k and φ, as this allows one to estimate t_R when φ is adjusted without actually performing the experiment, which can greatly reduce method development time. Several strategies exist [32]. One of these is to model the influence of φ on k empirically. Depending on the mode of chromatography and the analyte, different models may be applicable [33,34]. Of course any model is only applicable if the analyte actually elutes from the column after experiencing interaction with the stationary phase. A wealth of different models exists for specific situations [16,33–41]. For an isocratic RPLC separation of uncharged analytes the influence of φ on k has been shown to be effectively described using a quadratic model (QM) [42],

$$\ln k = \ln k_0 - S_1\varphi - S_2\varphi^2 \tag{1.13}$$

Where $\ln k_0$ is the y-intercept, or $\ln k$ extrapolated to 100% weak solvent ($\varphi = 0$), S_1 is a slope parameter, and S_2 is a curvature parameter. Because reasonable k values (*e.g.* $1 < k < 10$) for polymers are only encountered across a (very) narrow range of φ, **Equation 1.13** can often be simplified to

$$\ln k = \ln k_0 - S\varphi \tag{1.14}$$

$$k = k_0\, e^{-S\varphi} \tag{1.15}$$

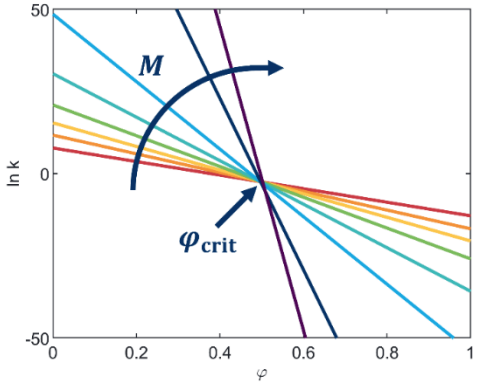

Figure 1.2: Natural logarithm of the retention factor *vs.* mobile-phase composition for a series of polystyrene standards of increasing molecular weight (increasing from red to purple traces).

which is a log-linear model often referred to as the linear-solvent strength (LSS) model [43]. When determined on different columns of the same type (*e.g.* C_{18}), **Equation 1.8** suggests that larger values of k_0 will be encountered for columns featuring larger values of β. In practice this corresponds to columns featuring smaller pore sizes. For a homologues series $\ln k_0$ and S are correlated, resulting in a bundle of lines, where the largest analyte has the steepest slope and *vice versa* [27,30]. This is illustrated in **Figure 1.2**.

When the effect of exclusion can be neglected (such as on non-porous packings), the critical composition (φ_{crit}) corresponds to the intersection point of this bundle. Even when the individual lines do not perfectly intersect at one point φ_{crit} may still be estimated by assuming a linear correlation between S and $\ln k_0$, *i.e.* $S = p + q \ln k_0$ [27,30]. In that case a plot of S versus $\ln k_0$ features a slope of $q \cong \frac{1}{\varphi_{\mathrm{crit}}}$ and a *y*-intercept of $p \cong \frac{\ln(\beta^{-1})}{\varphi_{\mathrm{crit}}}$. The retention factor at φ_{crit}, in the absence of end-group effects, should be $k_{\mathrm{crit}} = \beta$.

For large analytes the above model can only be correct when the contribution of exclusion is absent ($K_{\mathrm{SEC}} \cong 1$) or if exclusion effects cannot occur during the experiment (as in a solvent gradient experiment). In reality, when analytes are eluted from porous packings ($K_{\mathrm{SEC}} < 1$) and so the measured elution volume would be smaller than predicted. Naturally, this will be more noticeable when K_{LC} is very small, *i.e.* when φ is very close to φ_{crit}. At such conditions ($K_{LC} \cong 1$) it may be necessary to account for the contribution of exclusion. Assuming K_{SEC} and V_p are independent of φ, the predicted retention time can be corrected by adjusting V_0 based on analyte size. The correction factor can be determined with a prior SEC experiment.

In certain cases a different, so-called polymer model (PM), is used to account for the contribution of SEC. This model is based on a statistical theory of polymers in large slit-like pores [35,44]. In that model K_{PM} is given by

$$K_{PM} = 1 - \frac{4}{\sqrt{\pi}}\frac{R_g}{D} + 2\frac{R_g}{D}\frac{Y(-cR_g)-1}{cR_g} \tag{1.16a}$$

$$Y(-cR_g) = e^{cR_g^2}\left[1 - \mathrm{erf}\left(-cR_g\right)\right] = e^{cR_g^2}\left[\mathrm{erfc}\left(-cR_g\right)\right] \tag{1.16b}$$

where R_g is the radius of gyration of the polymer molecule, D is the (average) pore diameter, and c is an interaction parameter that varies from $c < 0$ to $c > 0$ during the transition from SEC to IPC. In **Equation 1.16b** erf denotes the error function and erfc the complimentary error function ($\mathrm{erfc(x)} = 1 - \mathrm{erf\,(x)}$). To calculate V_R the relationship between c and φ must be known. Often c is considered to vary linear with φ over a small range of φ, which is the same assumption made in LSS theory (*i.e.* $\Delta H°$ decreases linearly with φ). It is further assumed that both R_g and D are invariant with φ, or that those effects are negligible.

If all assumptions hold V_R is calculated (if V_i and V_p are known) from **Equation 1.4** by determining three parameters: $R_g\frac{dc}{d\varphi}$, $\frac{R_g}{D}$, and φ_{crit}, from at least three (isocratic) experiments. It is assumed that the equation for SEC (**Equation 1.4**) can also describe retention in IPC, and that the influence of V_s is represented by V_p, which is likely not correct. The value of K_{PM} varies between $K_{PM} < 1$ for $c < 0$ to $K_{PM} > 1$ for $c > 0$. At $c \cong 0$, the last term in **Equation 1.16a** approaches $\frac{4}{\sqrt{\pi}}\frac{R_g}{D}$ so that $K_{PM} = 1$. In this model the SEC contribution is (for the most part) captured in the first part of the equation $\left(-\frac{4}{\sqrt{\pi}}\frac{R_g}{D}\right)$, since the last part, which captures the contribution of interaction, $\left(+2\frac{R_g}{D}\frac{Y(-cR_g)-1}{cR_g}\right)$ quickly becomes negligible when $c \ll 0$. This model has the benefit of being able to describe SEC, LCCC and IPC, but it also implies that a (negative) SEC contribution is required to cancel out the last term, to reach critical conditions, which may not be realistic.

Within the vicinity of φ_{crit} a simplified model has been described [28], *i.e.*

$$K = e^{\left(2\frac{R_g^2(c_{crit}-c)}{Da\ k_bT}\right)} \tag{1.17}$$

where k_b is the Boltzmann constant, and a is the thickness of a "monomolecular" adsorption layer. This model is essentially equivalent to the LSS model (**Equation 1.15**) if $\ln k_0 = \ln\left(\beta e^{\frac{2R_g^2}{Dak_bT}\varphi_{\mathrm{crit}}}\right)$ and $S = \frac{2R_g^2}{Dak_bT}$, so that $\ln k = \ln\left(\beta e^{\frac{2R_g^2}{Dak_bT}\varphi_{\mathrm{crit}}}\right) - \frac{2R_g^2}{Dak_bT}\varphi$, as long as it is assumed that c varies linearly with φ over a small range of φ.

1.1.6. Modelling migration in solvent-gradients

To expand upon the range of analytes that may be analysed within one experiment and within a reasonable time, gradient-elution can be used. In the case of polymers gradients are typically required, as retention scales approximately exponentially with molecular weight. Mobile-phase (or sometimes temperature) gradients are abundantly used, especially to assess a copolymer's CCD. As copolymers feature a wealth of critical compositions, based on their CCD, it is not possible to obtain critical conditions for all analytes in a copolymer sample. However, with gradient-elution it is possible to obtain pseudo-critical behaviour for all analytes, if the influence of molecular weight in the gradient can be sufficiently supressed [29].

While the principle of solvent-gradient-elution LC appears straightforward, it is significantly more complex than isocratic LC, because a change in mobile-phase composition is accompanied by changes in, for example, solvent viscosity, analyte solubility and size (in solution) [45], the possibility of system or column-induced gradient deformation [46], and potential changes in the stationary phase [47]. Additionally, mobile-phase components may be (partly) retained during the gradient [13,48–50]. Typically, solvents are chosen that do not have large differences in viscosity. However, there is still a change in pressure during the gradient that is often not taken into account. Pressure effects are usually small for low-molecular-weight analytes [51–55], but they tend to increase with increasing analyte molecular weight. The other parameters (analyte solubility, size, gradient deformation, changes in stationary phase) are not commonly accounted for. The effects are expected to be

small, but this will again depend on the method, system hardware, and analyte. Equations have been derived that allow for reasonable predictions of retention time in gradient-elution LC, which ignore these secondary effects [36–39,42]. Accurate predictions require that *i*) the initial mobile-phase composition of the gradient is sufficiently weak, *ii*) the actual gradient shape, or the change in φ with time (t), is known, and *iii*) it is known (or an assumption is made on) how k changes with φ. For polymers an additional requirement is that the analyte is soluble in a relatively weak mobile phase, so that the analytes can be retained on the column, before the analyte elutes under "SEC-like" elution in a phenomenon called "breakthrough" [56]. In case of a linear gradient program, and neglecting potential gradient deformation, the change in φ with t is considered equivalent to a simple linear equation:

$$\varphi(t) = \varphi_{\text{init}} + \frac{d\varphi}{dt}t = \varphi_{\text{init}} + \varphi' t \tag{1.18}$$

In which φ_{init} equals φ at the start of the gradient ($t = 0$), and $\frac{d\varphi}{dt}$ is the slope of the gradient or the gradient rate (φ'). For the dependence of k on φ it is assumed that one of the above models (QM, LSS, PM) can describe a change in the local retention factor of the analyte due to the gradient. For a polymer different situations can be imagined in gradient-elution (disregarding the situations where elution occurs by breakthrough or does not occur at all), such as

i) The polymer is soluble only in combinations of mobile phases that are stronger than those in which the polymer has significant interactions with the stationary phase. The analyte elutes during or after the gradient around the mobile-phase composition where it is first soluble (φ_{sol}) if the solubilization kinetics are sufficiently fast. In this case column length is not expected to influence the elution composition.

ii) The polymer is soluble in a relatively weak mobile phase, so that interaction with the stationary phase may determine retention. In this case the polymer will elute before or around the critical composition (φ_{crit}) if the final mobile phase-composition of the gradient is sufficiently strong. If the latter is not the case, only lower-molecular-weight analytes may elute from the column with $k > 0$.

These two cases can be referred to as precipitation/redissolution chromatography and interaction chromatography, respectively, and are illustrated in **Figure 1.3** [45,57,58].

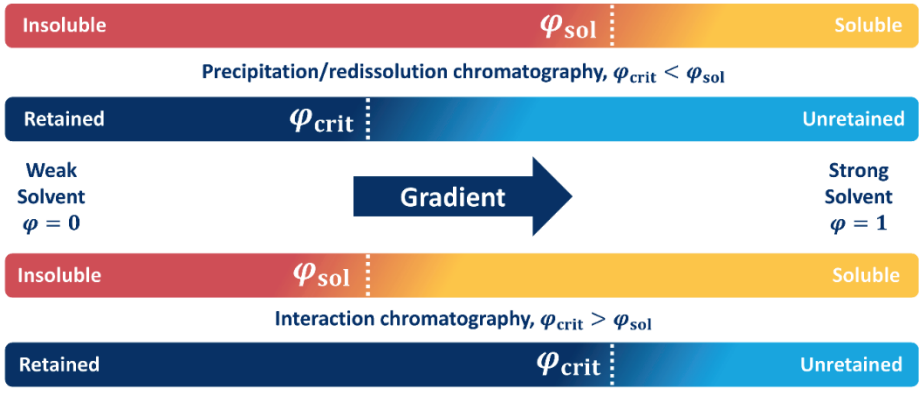

Figure 1.3: Two types of gradient-elution liquid chromatography (precipitation/redissolution, and interaction chromatography) for polymers.

Note that case *i)* can be shown to occur even without a column (providing a separation based on molecular weight), but that this results in bad peak shapes and significant breakthrough, since the analyte cannot be sufficiently focused or separated from the injection solvent. Provided that the polymer is soluble in a weak mobile-phase, two different situations may occur depending on the initial mobile-phase composition of the gradient. Namely,

i) When $\varphi_{init} > \varphi_{crit}$, so that the gradient will not play a role and elution occurs with a SEC-like mechanism before t_0.

ii) When $\varphi_{init} < \varphi_{crit}$. In this case elution will depend primarily on the retention factor of the analyte at the initial mobile-phase composition and on $\frac{dk}{d\varphi}$.

For case *ii)*, $\varphi_{init} < \varphi_{crit}$, elution will occur either isocratically before the gradient catches up with the analyte zone, during the gradient, or after the gradient. When an analyte elutes during the gradient the elution mechanism cannot change from IPC to SEC because if the analyte would experience exclusion it would nearly immediately experience a weaker solvent and a strong interaction with the stationary phase.

Therefore, the ability to describe a transition from IPC to SEC, as offered by the PM, loses its main advertised advantage when isocratic elution is replaced by mobile-phase gradients. The retention time of an analyte during a gradient can be calculated by solving a differential equation, given by **Equation 1.19** [37–42].

$$\int_0^{t_R} \frac{1}{t_0} \frac{dt}{k} = 1 \tag{1.19}$$

For the LSS model (**Equation 1.14**) and a linear gradient (**Equation 1.20**) the solution for an analyte eluting within the gradient, after accounting for a delay (t_{dwell}) due to the dwell volume of the instrument (the volume required for a change in mobile-phase composition to reach the start of the column), is

$$t_R = \frac{1}{\varphi' S} \ln\left[1 + S\varphi' k_{\mathrm{init}} \left(t_0 - \frac{t_{\mathrm{dwell}}}{k_{\mathrm{init}}}\right)\right] + t_{\mathrm{dwell}} + t_0 \tag{1.20}$$

in which k_{init} is the retention factor at the initial conditions. In case an initial delay is programmed into the gradient, this can be added to t_{dwell}. For polymers eluting within the gradient, within a narrow range of φ (*i.e.* for large-molecular-weight polymers) the LSS model has been shown to allow for reasonable predictions of the retention time, as long as gradient experiments are also used to establish k_0 (or k_{init}) and S. Note that there will likely be differences in k_0 and S measured on different systems, since a change in pressure occurs during the mobile phase gradient, and potentially because of system-induced gradient deformation [46]. As mentioned above **Equation 1.19** can only be solved for linear gradients, and not for the more-complex gradient profiles obtained due to insufficient mixing and/or other system imperfections [46,59]. A correction procedure for these cases is described in Chapter 5. For such, more-complex gradients shapes a combination of multiple linear gradient segments can be used to approximate the gradient profile. To do so **Equation 1.20** can be extended, as has been done in various works [39,46].

1.1.7. Challenges when modelling the retention of (co)polymers

Irrespective of the validity of the above retention models, for polymers a retention modelling approach is much-more challenging than for low-molecular-weight analytes, because the elution time is a function of multiple properties. The most-important of these are the molecular weight and the chemical composition. As a rule, we are dealing with distributions, rather than with uniquely defined properties. The

first problem with modelling the retention of polymers is the extreme change in k with φ for large analytes. As a result, high-molecular-weight analytes, such as most synthetic polymers, experience measurable retention only within a very narrow range of φ, and will not elute from the column at weaker mobile-phase compositions. This makes it often difficult, or practically impossible, to achieve isocratic IPC separations of high-molecular-weight (co)polymers. Therefore, gradient-elution LC must be used. However, this does not solve all problems. For small analytes retention modelling can be challenging, primarily when analytes cannot be properly tracked between measurements, for example due to changing elution orders and analyte co-elution. In those cases mass spectrometry is very helpful to discriminate between various analytes [60,61]. For polymers this problem is largely the same, but exacerbated by the molecular weight of the analytes and the much larger number of individual species, especially in the case of copolymers. Additionally, aspects such as crystallinity or solubility may hinder the retention modelling of (co)polymers [45,62].

If online LC-MS is not feasible, then to apply the above retention models either an off- or on-line fractionation is required so that the polymer distribution may be deconvoluted. A different strategy is to predict the retention of copolymers consisting of two different monomers based on the retention characteristics of individual homopolymers. Such an approach is significantly more feasible, since standards featuring a narrow MWD are more likely available for homopolymers than for copolymers. Some work on this has been performed by Brun *et al.* and Fitzpatrick *et al.* [27,28]. The results from this work show that for high-molecular-weight copolymers, that elute close or at the (pseudo) critical composition, that feature a random distribution of the two (and likely also more) monomers in the chain, retention under gradient conditions can be considered equivalent to the weighted average of the natural logarithms of the retention factors of both homopolymers. Accounting for the mass fractions of each monomer, X_I and $X_{II} = (1 - X_I)$, this corresponds to:

$$\ln k_{CP} = X_I \ln k_I + (1 - X_I) \ln k_{II} \tag{1.21}$$

For copolymers that feature larger sections of one repeat unit in the chain, *i.e.* more blocky copolymers, this is generally believed to be no longer true and the retention

of the polymer is expected to depend on the "randomness" (or "blockiness") of the copolymer [28], as

$$\ln k_{CP} = \frac{1}{2}\Big[\ln k_I\,(1-\nu X_{II}) + \ln k_{II}\,(1-\nu X_I) + \{[\ln k_I\,(1-\nu X_{II}) + \ln k_{II}\,(1-\nu X_I)]^2 - 4\ln k_I \ln k_{II}(1-\nu)\}^{\frac{1}{2}}\Big] \quad (1.22)$$

where ν corresponds to the randomness of the copolymer, and varies from 0 to 1 to 2 for a block, random and alternating copolymer, respectively. This model returns **Equation 1.21** for $\nu = 1$ and predicts increased retention for copolymers featuring $\nu < 1$, *i.e.* copolymers that feature a more blocky structure. The idea that a block copolymer elutes later than a statistical copolymer of equivalent overall chemical composition corresponds fairly well to experimental results [28]. However, so far no work has been performed to validate **Equation 1.22** experimentally. This is most likely because samples that feature a well-defined randomness are very difficult to obtain. In part this is because no methods are available that can separate based purely on polymer sequence. In fact, no experimental methods exist that can provide detailed information on the underlying sequence distribution. According to **Equation 1.22** information on sequence distributions may, in principle, be obtained by using (gradient-elution) IPC.

1.2. Peak dispersion, resolution and peak capacity

1.2.1. Plate model

The degree of separation between a pair of compounds, in isocratic elution, can be characterized by the ratio of the net retention times of the two compounds, known as the selectivity.

$$\alpha = \frac{t_{R,2} - t_0}{t_{R,1} - t_0} = \frac{k_2}{k_1} \quad (1.23)$$

Of course, the actual performance of a separation depends not only on the positions of peaks in the chromatogram, but also on their width. The peak width that is recorded at the end of a separation is the result of multiple sources of variance from within the system. Examples include broadening in the volume of the detector cell or connection capillaries or during the injection procedure, broadening during the separation, and broadening as a result of stagnant volumes in the system caused by,

for example, bad connections between the different parts of the system. Furthermore, in gradients a small sharpening of the analyte zone may occur due to a difference in retention between the front and rear of the peak. However, this effect is usually small. Additional sources of variance can be differences in viscosity, or temperature gradients in the system, most typically caused by viscous heating within the column [63]. The measured peak width will be the sum of all variances [64]. These different sources of band broadening or sharpening also often induce changes in the peak shape. Since peaks are often described by a (symmetrical) Gaussian distribution, most of the equations that were previously derived to describe the performance of a separation are not entirely correct. Instead, peaks are better described by, for example, exponentially modified gaussian (EMG) or modified Pearson VII distributions, allowing one to take into account the peak asymmetry [65,66]. Nevertheless, since most of the equations derived for symmetrical peaks are still commonly used we have chosen to describe those in this thesis.

Usually peak width is directly related to the plate number (N), which is a property adapted to chromatography from industrial-scale distillation columns by Martin and Synge in 1941 [67]. In this plate theory, a column is considered (incorrectly) to consist of a number of locations or plates of finite volume. At each plate an equilibrium is achieved between the mobile and stationary phases and transfer from one plate to the next occurs when a certain amount of mobile phase is added. Accordingly, peak broadening is essentially only caused by the distribution coefficient and the number of plates. The resulting predictions of this theory were that peak broadening increased with increasing retention time and that it is minimized by decreasing the (theoretical) plate height (H), which is a measure for the distance between successive plates. The results of this theory can be summarized in **Equations 1.24 to 1.26**,

$$\sigma_t = \frac{t_R}{\sqrt{N}} = \frac{1}{\sqrt{N}} t_0(1+k) \tag{1.24a}$$

$$\sigma_V = \frac{V_R}{\sqrt{N}} = \frac{1}{\sqrt{N}} V_0(1+k) \tag{1.24b}$$

$$\sigma_x = \sqrt{Hx} \tag{1.24c}$$

$$H = \frac{L*\sigma_t^2}{t_R^2} = \frac{L}{N} \tag{1.25}$$

$$N = 16\left(\frac{t_R}{w_b}\right)^2 = 5.54\left(\frac{t_R}{w_{1/2}}\right)^2 = \left(\frac{t_R}{\sigma_t}\right)^2 \quad (1.26)$$

With σ_t and σ_V are a measure of the peak width in time and volume units, σ_x the same measure but in terms of distance travelled, x the path length travelled by the analyte, $w_{1/2}$ the peak width at half height, w_b the peak width at the "base" of the peak (determined from the two intersection points of tangent lines drawn from the inflection points of the peak with the baseline on either side of the peak). **Equation 1.26** is derived by assuming that each peak can be described as a (symmetrical) Gaussian distribution, implying $w_{1/2}$ is equal to $2.354\sigma_t$, or w_b equal to $4\sigma_t$. This equation can be used to obtain an indication of separation efficiency, in the case of isocratic separations. With gradient-elution k changes during the gradient and an additional (usually small) peak compression occurs as a result of a (small) difference in k between the front and end of the peak [68,69]. The peak width is described by

$$\sigma_t = \frac{t_R}{\sqrt{N}}G = \frac{1}{\sqrt{N}}t_0(1 + k_e)G \quad (1.27)$$

Where G is a gradient compression factor (smaller than one), and k_e is the retention factor at the time of elution. Based on this simplified view one may be inclined to consider N to be a property of a specific column. In reality peak broadening does not only depend on retention, but on a combination of analyte-, solvent- and column-specific parameters. These effects are not accounted for by the plate model. The most well-known of many (more-or-less) comprehensive models that do incorporate such contributions is the "van Deemter" equation [70].

1.2.2. Van Deemter equation

Plate theory does not provide a full description of the peak broadening that occurs during separation. A more-comprehensive theory was developed by van Deemter in 1956 [70]. The result of this work is the well-known van Deemter equation, the non-expanded version of which is given by

$$H = A + \frac{B}{u_m} + Cu_m \quad (1.28)$$

This equation relates the (theoretical) plate height to the linear flow velocity (u_m) through three parameters, *viz.* (A) eddy dispersion, (B) longitudinal diffusion and (C) resistance to mass transfer. **Equation 1.28** is visualized in **Figure 1.4**.

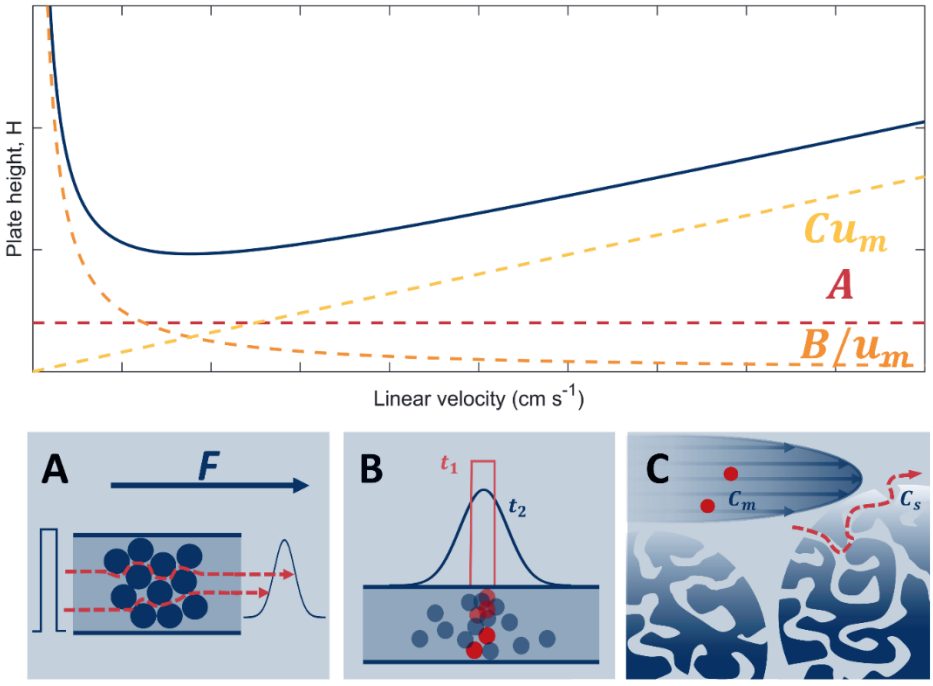

Figure 1.4: Van Deemter plot, illustrating the A-, B- and C-term contributions to peak broadening. A) Eddie dispersion or diffusion, B), Longitudinal diffusion, and C) Resistance to mass transfer.

The A-term describes the dispersion of an analyte band due to differing path lengths through the column. To decrease its effect, a packing consisting of small particles with a narrow particle-size distribution or a pillar-array column with an ordered structure [71] should be used.

The B-term describes the band spreading due to diffusion, caused by analyte molecules moving from high-concentration to low-concentration areas. Entropy increases and the steepness of the concentration gradient that exists when the analyte is injected in the mobile-phase is gradually reduced, before it entirely

disappears. The effect of longitudinal diffusion can be reduced by increasing the flow velocity or viscosity of the mobile phase, or by reducing the temperature. This term is usually not significant due to the nature of HPLC (low D_m) and the flow velocities that are employed (reduced linear velocity, $\nu = \frac{u_m d_p}{D_m} \gg 5$, where d_p is the particle size), but it may become important at low flow velocities, especially for small analytes.

The third and in many cases dominant term related to peak broadening in HPLC is the resistance to mass transfer (C-term), which consists of a C_s term that describes the movement into and out of the stationary phase, and a C_m term comprising velocity differences in the column. The effects of mass transfer can be reduced by using *i*) smaller particles, to decrease the distance analytes must travel when moving in and out of the particles or between fast and slow streamlines between the particles, *ii*) higher temperatures, since more-rapid diffusion will increase the movement of analyte molecules, and *iii*) using lower flowrates. Typically the C-term is the main contributor to peak width for large analytes.

Various improvements in describing peak width have been made since the work of van Deemter, and are still ongoing. The description of these models is outside the scope of this thesis. The reader is instead referred to some of the literature which discusses these models more in-depth [68,69,72–78].

1.2.3. Resolution and peak capacity

For Gaussian peaks of equal peak area, knowledge of the position and variance is sufficient to fully describe the extent to which two compounds can be resolved. This is commonly referred to as the resolution (R_s) [79].

$$R_s = \frac{1}{2}\frac{t_{R,2} - t_{R,1}}{\sigma_{t,1} + \sigma_{t,2}} \tag{1.29}$$

$$R_s = \frac{\sqrt{N}}{2(\alpha+1)}\left(\frac{k_{\text{mean}}}{k_{\text{mean}}+1}\right)(\alpha - 1) \tag{1.30}$$

In which k_{mean} is the mean retention factor between the two peaks ($k_{\text{mean}} = \frac{k_1 + k_2}{2}$). The plate number in **Equation 1.30** is assumed to be equal for both analytes ($N = N_1 = N_2$). This equation illustrates the degree to which each parameter improves resolution. For example, an increase in selectivity will have more of an effect than an

increase in the number of plates. This is the primary reason why multidimensional chromatography (described in the next section) allows for the separation of more-complex samples. As already described above **Equation 1.30** is only valid given a number of assumptions, and is restricted to isocratic elution. Many forms exist, given different assumptions [79].

Most chromatograms contain more than two peaks. In that case the resolution of all peaks pairs must be determined to assess how well the separation performs, ideally taking into account also the measurement duration. In practice, there are a plethora of objectives against which a separation may be optimized, apart from "simply" maximizing the lowest observed R_s value. Perhaps the most common metric to compare the performance of separation methods is the peak capacity (n_c). In the case of isocratic separations Giddings defined the peak capacity as an approximation of the "maximum number of peaks to be separated on a given column" [80] and he derived the following equation:

$$n_c = 1 + \frac{\sqrt{N}}{4R_s} \ln \frac{t_{R,n}}{t_{R,1}} = 1 + \frac{\sqrt{N}}{4R_s} \ln \frac{1+k_n}{1+k_1} \tag{1.32}$$

In which n_c depends on the retention of both the first ($t_{R,1}$) and last ($t_{R,n}$) eluting component as well as on N and R_s. Given a particular mobile-phase composition at which all analytes can elute from the column, the optimal peak capacity will be obtained when N is maximized for a given R_s. Assuming that N is independent of the retention factor, this involves using long columns and large particles. In practice, small particle diameters and relatively short columns are used, resulting in slightly lower peak capacities, but in significantly shorter analysis times. For gradient separations the peak capacity can be approximated as,

$$n_c = \frac{t_g}{w_{b,avg}} + 1 \tag{1.33}$$

where $w_{b,avg}$ is the average peak width measured from the base of the peak and t_g is the gradient duration. When n_c is sufficiently large, adding one to the outcome (latter part of **Equation 1.33**) is insignificant. Again, care should be taken that N is analyte specific. Hence the use of $w_{b,avg}$ in **Equation 1.33**, which should give a reasonable estimation of n_c.

The caveat to both equations are that they assume every peak to be perfectly spaced apart with sufficient resolution (~1). This implies an ordered distribution of peaks throughout the chromatogram, which in reality will almost never be the case. The concept was further improved by Davis and Giddings in 1983 [81], who instead assumed a random distribution of peaks, which is much closer to reality. For a saturated chromatogram, where the number of peaks is equal to the peak capacity, the number of completely resolved peaks that will then be observed will be at most 37% of the peaks expected to be in the chromatogram had they been uniformly spaced [81]. In fact, as a rule-of-thumb, if 90% of expected peaks are to be resolved a peak capacity approximately 20 times larger than the number of peaks present is required.

In the case of polymers it is not completely straightforward to use metrics such as n_c, since these samples contain many different, often very similar, analytes, resulting in low resolution. This is generally the case for all homologues, except the smallest oligomers. As a result it may be better to evaluate the quality of polymer separations by the type (for example the MWD, CCD, or FTD), rather than the amount (in terms of the number of peaks) of information that is obtained.

1.3. Two-dimensional chromatography

The performance of a one-dimensional (1D) LC separation depends on many factors, including the chemical nature of the analytes, the particle size and particle-size distribution, flow velocity, column parameters (*i.e.* diameter, length), operating pressure, temperature, the type and viscosity of the mobile phase, the type of stationary phase, the column permeability, particle porosity, the effective gradient steepness, the gradient shape, system dwell volume, and sources of extra-column band broadening that cannot be avoided [82]. Often it is desirable to obtain conditions that result in the highest possible overall peak capacity in the shortest possible time, *i.e.* maximizing separation efficiency. This has been the subject of research by, among others, Giddings [83], Knox [77], Guiochon [82], Poppe [84], and Desmet *et al.* [76,78]. Based on this research 1D-LC separations are most attractive for analysis times of about one hour or less, resulting in peak capacities of a few hundred. For longer analysis times two-dimensional LC (2D-LC) is deemed more appropriate [85].

2D-LC separations involve the coupling of two or more separations that ideally provide vastly different ("orthogonal") selectivity towards the sample [86–88]. The simplest implementation of 2D-LC involves taking fractions from a first separation and reinjecting these in a second separation, *i.e.* off-line 2D-LC [85]. This approach is challenging to automate, and even when automation were to succeed, there is a significant risk of sample losses and contamination. Therefore, a more common implementation of the 2D-LC involves the coupling of two one-dimensional LC separations using a so-called modulator [86]. The latter is typically some type of switching valve that allows the system to temporarily store a (small) fraction of the effluent from the first-dimension (^{1}D) separation, which is then transferred to the second-dimension (^{2}D) separation. Different modes of operation have also been envisioned, such as spatial 2D-LC [89,90]. However, as of now these still require significant development. Depending on how much of the first-dimension effluent is sampled, different terms are used to describe 2D-LC approaches, such as heart-cut (LC-LC), where only one selected fraction is transferred; multiple-heart-cut (mLC-LC), where multiple fractions are transferred; selective comprehensive (sLC×LC), where series of fractions are transferred from multiple regions; and fully comprehensive (LC×LC), where the entire effluent is sampled. In the case of polymers sometimes the name "cross-fractionation" is used [91].

For 2D-LC separations selectivity can be optimized based on at least two parameters *e.g.* size and charge, or size and polarity. This dramatically increases the total separation "space", or peak capacity. This is true even if the peak capacity for one of the separations is low, since an increase in resolution is most easily achieved by tweaking the selectivity. Therefore, multidimensional separations are often the prime choice for separations of samples that feature high dimensionality [92]. The actual peak capacity for a 2D-LC separation can be calculated in different ways. If the selectivity of both separations is completely uncorrelated, *i.e.* orthogonal, the peak capacity is calculated from

$$n_c = n_{c,1} \times n_{c,2} \tag{1.34}$$

with $n_{c,1}$ and $n_{c,2}$ the peak capacity of the first and second dimension, respectively. However, often the two dimensions are somewhat correlated. For example, analyte

polarity often scales with analyte size. This results in a lower effective peak capacity ($n_{c,\mathrm{eff}}$), *i.e.*

$$n_{c,\mathrm{eff}} = n_{c,1} \times n_{c,2} f_{\mathrm{coverage}} \tag{1.35}$$

where f_{coverage} is a measure of the orthogonality between the two separations, *i.e.* a measure of incomplete surface coverage. Additionally, it is further assumed that the entire effluent is sampled sufficiently. As the effluent is transferred by means of a switching valve that is equipped with two sample loops, a loss in resolution may occur during the transfer from the first to the second-dimension separation due to re-mixing within the loop. This is referred to as under-sampling [93,94]. Whether this occurs will depend on the modulation time (t_{mod}), *i.e.* how quickly each loop is emptied, and the (average) first dimension peak width ($^1\sigma_t$). To account for the possibility of under-sampling another correction can be added,

$$n_{c,\mathrm{eff}} = n_{C,1} \times n_{C,2} \frac{f_{\mathrm{coverage}}}{f_{\mathrm{undersampling}}} = n_{C,1} \times n_{C,2} \frac{f_{\mathrm{coverage}}}{\sqrt{1+0.21\left(\frac{t_{\mathrm{mod}}}{^1\sigma_t}\right)^2}} \tag{1.36}$$

Where $f_{\mathrm{undersampling}}$ is an under-sampling correction factor. According to this equation it may be concluded that each peak should be sampled 4 to 5 times to reduce the loss in peak capacity by under-sampling to below 1% [94].

1.3.1. Challenges and recent developments in LC×LC

The potential of 2D-LC has often been said to be hindered by challenges related to, for example, the aforementioned under-sampling, the increased development time, potential incompatibility between the two separations, and the inaccessibility of commercial tools for data analysis. However, despite these challenges the use of 2D-LC in industry is becoming more and more common. Over the last ten years significant developments have been made with respect to instrumentation [88,94,95]. Additionally, the growing usage of chemometrics aimed specifically at 2D-LC makes the technique more accessible than ever before [96].

The most common challenges for the on-line coupling of two methods by means of a switching valve or "modulator" with two or more sample loops are mobile-phase incompatibility, under-sampling, and increased dilution or injection band broadening due to overloading of the ^{2}D column. The two most-common solutions

to mitigate these issues are *i*) the use of small "trap" columns instead of sample loops, commonly referred to as stationary-phase assisted modulation (SPAM) [97], and *ii*) in-loop dilution, referred to as active-solvent modulation (ASM) [98]. The passive and SPAM modulation strategies are illustrated in **Figure 1.5**.

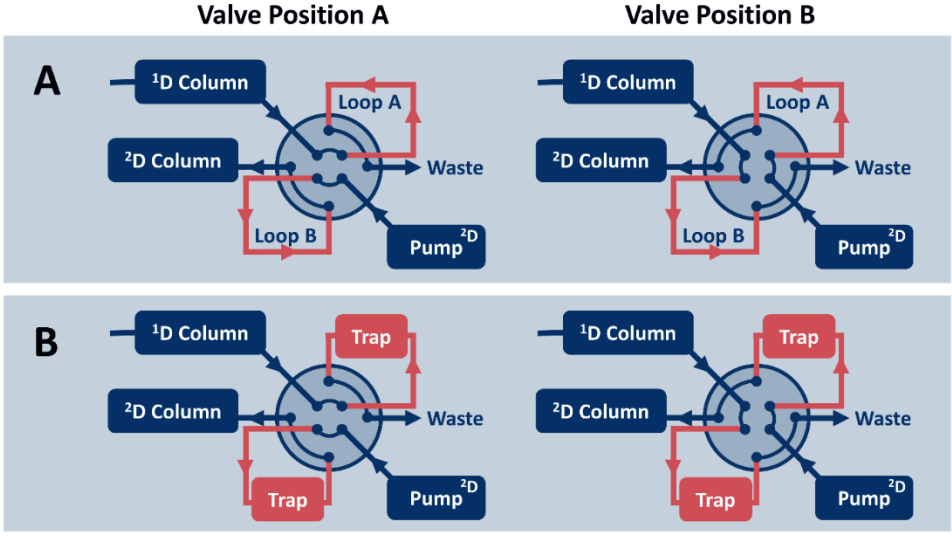

Figure 1.5: A) Conventional 2D-LC configuration (passive modulation), and B) SPAM configuration, lefthand side and righthand side figures indicate valve positions A and B.

The use of SPAM allows for a re-focusing of analytes within the trap column and a large reduction in the amount of ^{1}D solvent entering the ^{2}D column [97]. It consequently allows for a reduction in analysis time and an increase in sensitivity, by increasing the ^{1}D flowrate and decreasing the ^{2}D column diameter, respectively. However, to perform SPAM successfully the trapping columns should retain all analytes irrespective of the ^{1}D mobile-phase composition in which they elute. As a result, SPAM cannot be used if SEC is applied as a first-dimension separation, unless large dilution flows are used after the ^{1}D column. In the case of a ^{1}D-RPLC separation either a dilution flow should be used prior to the effluent entering the trap column, or the stationary phase in the trap should be more retentive than that of the ^{1}D column. In the specific case where SEC is used as a ^{2}D separation, neither solution is ideal. Therefore, in certain cases a combination of SPAM with temperature modulation may be more feasible and easier to implement [99,100]. Such a strategy

is discussed in Chapter 2 of this thesis. The alternative strategy, ASM, allows for a potential re-focusing at the head of the ^{2}D column by ensuring that the analyte is transferred in a weak(er) solvent. Since no trap columns are used ASM does not suffer from the same issues as SPAM, and is also more repeatable, since the lifetime of the trap columns is limited. However, ASM does ideally require a refocusing on the ^{2}D column, to counteract the dilution that occurs. In practice this limits ASM to systems in which RPLC is used in the second dimension.

1.4. Understanding materials by characterizing essential distributions

Nowadays many materials used in a wide range of fields consist of polymers that embody great structural complexity. Examples include amphiphilic block copolymers that are employed as drug delivery systems, or acid-functionalized polymers used in waterborne coatings. Often these materials must feature specific functional properties that are required for the application in which they will be used. Often synthesis conditions and post-processing are varied to obtain the desired material properties, and the underlying structure-property relationships are often not completely understood. Detailed analysis of these materials is one of the steps necessary to aid in elucidating this connection between on the one hand the distributions of the polymer in terms of, for example, microstructure, chemical composition or molecular weight, and on the other hand the functional properties. Within the UNMATCHED project (UNderstanding MATerials by CHaracterizing Essential Distributions) many techniques have been investigated, or further developed, to aid in the analysis of such materials. One set of techniques that is particularly suitable for the analysis of essential distributions is liquid chromatography and two-dimensional liquid chromatography. Several new ways to implement these techniques for the separation of synthetic polymers by LC are presented in this thesis.

Scope of the thesis

Challenges in analytical chemistry must often be tackled by developing new experimental methods or techniques, or by smartly applying chemometrics. Often a combination of these is required. Both subjects are represented in this thesis. While every chapter includes both, the focus in the first chapters is primarily on the development of new LC methods and techniques. In later chapters this gradually shifts towards a focus on chemometric strategies.

Chapter 1 provides a general introduction to the subjects discussed in this thesis. Relevant characteristics of synthetic polymers are briefly discussed and the foundations of the liquid-chromatographic techniques used in later chapters are reviewed.

In **Chapter 2** the effect of temperature on retention is utilized for trapping high-molecular-weight analytes on small (so-called) "trap" columns. This is done to improve the two-dimensional liquid chromatography (2D-LC) separations of polymers. The relatively stronger effect of temperature on the retention of higher-molecular-weight polymers is shown to allow for their successful trapping and subsequent elution, without requiring rapid temperature (heating and cooling) cycles. This allows one to remove solvents and to simultaneously refocus the analytes prior to a subsequent second-dimension analysis by size-exclusion chromatography (SEC). Such an LC×SEC combination is significantly more challenging when utilizing conventional trapping or dilution strategies. As part of this work a new modulation strategy was also developed that reduces the pressure pulses during valve switching. This potentially enhances the lifetime of the used trap columns.

In **Chapter 3** a new method termed recycling gradient-elution liquid chromatography (LC↻LC) is described. The aim is to increase the effective gradient steepness, so as to suppress the underlying influence of the molecular-weight distribution typically observed for lower-molecular-weight polymers. The method is successfully applied to measure the chemical-composition distributions of various copolymers.

In **Chapter 4** the use of SEC gradients (or gradient-SEC, gSEC) as an alternative to RPLC for the chemical-composition analysis of polymers is investigated. As a part of this work the influence of the mobile-phase composition on the elution volume in

SEC was investigated. The method was then applied in 1D-LC and 2D-LC separations, as well as in the recycling mode described in chapter 3. It is shown that, compared to conventional gradient-elution LC, the use of SEC-gradients is primarily advantageous when breakthrough is an issue in reversed-phase LC (RPLC). For example, in the case of LC×LC, with RPLC in the second dimension.

In **Chapter 5** the effect of system-induced gradient deformation on retention-modelling approaches is highlighted. Capacitively coupled contactless conductivity detection is demonstrated as a straightforward method to measure gradient deformation. The approach allows for nearly identical retention parameters to be determined on systems that vary greatly in terms of gradient deformation.

In **Chapter 6** an overview is given of strategies used for the pre-processing of one-dimensional data. Different facets of such pre-processing are reviewed, including smoothing, drift correction, peak detection and deconvolution.

In **Chapter 7** a critical comparison of some of the drift-correction and smoothing algorithms discussed in chapter 6 is made. To do so, simulated data are generated that are representative of complex real data based on three different components, *i.e.* high-frequency noise, low-frequency background drift, and relevant chemical signals, or peaks, with a frequency varying between that of drift and noise. The individual components were either directly taken from or based on experimental data. A data set consisting of 500 chromatograms was generated and corrected with different combinations of drift-correction and smoothing algorithms (35 combinations) to identify the best-performing combinations.

In **Chapter 8** the conclusions from the work in this thesis are reviewed and recommendations for future studies are given. Additionally, preliminary work and challenges that were not yet solved are discussed. Some strategies that appear feasible are envisioned.

References

[1] M. Tswett, Adsorption analysis and chromatographic method. Application to the chemistry of chlorophyll, Ber. Deut. Bot. Ges. 24 (1906).

[2] S. Eeltink, P. Gzil, W.T. Kok, P.J. Schoenmakers, G. Desmet, Selection of comparison criteria and experimental conditions to evaluate the kinetic performance of monolithic and packed-bed columns, J Chromatogr A. 1130 (2006). https://doi.org/10.1016/j.chroma.2006.05.068.

[3] G. Guiochon, Monolithic columns in high-performance liquid chromatography, J Chromatogr A. 1168 (2007). https://doi.org/10.1016/j.chroma.2007.05.090.

[4] G. Gulochon, Conventional Packed Columns vs. Packed or Open Tubular Microcolumns in Liquid Chromatography, Anal Chem. 53 (1981). https://doi.org/10.1021/ac00232a006.

[5] C.A. Rimmer, C.R. Simmons, J.G. Dorsey, The measurement and meaning of void volumes in reversed-phase liquid chromatography, J Chromatogr A. 965 (2002). https://doi.org/10.1016/S0021-9673(02)00730-6.

[6] A.M. Krstulovlć, H. Colin, G. Guiochon, Comparison of Methods Used for the Determination of Void Volume in Reversed-Phase Liquid Chromatography, Anal Chem. 54 (1982). https://doi.org/10.1021/ac00251a009.

[7] J.H. Knox, R. Kaliszan, Theory of solvent disturbance peaks and experimentaldetermination of thermodynamic dead-volume in column liquid chromatography, J Chromatogr A. 349 (1985). https://doi.org/10.1016/S0021-9673(01)83779-1.

[8] A.M. Striegel, Size-Exclusion Chromatography: A Twenty-First Century Perspective, Chromatographia. 85 (2022). https://doi.org/10.1007/s10337-022-04143-1.

[9] A.M. Striegel, Multiple Detection in Size-Exclusion Chromatography of Macromolecules, Anal Chem. 77 (2005). https://doi.org/10.1021/ac053345e.

[10] S.T. Popovici, P.J. Schoenmakers, Fast size-exclusion chromatography - Theoretical and practical considerations, J Chromatogr A. 1099 (2005). https://doi.org/10.1016/j.chroma.2005.08.071.

[11] E. Uliyanchenko, Size-exclusion chromatography - From high-performance to ultra-performance, Anal Bioanal Chem. 406 (2014) 6087–6094. https://doi.org/10.1007/s00216-014-8041-z.

[12] A.M. Striegel, W.W. Yau, J.J. Kirkland, D.D. Bly, Modern Size-Exclusion Liquid Chromatography: Practice of Gel Permeation and Gel Filtration Chromatography: Second Edition, 2009. https://doi.org/10.1002/9780470442876.

[13] S. Bocian, P. Vajda, A. Felinger, B. Buszewski, Solvent excess adsorption on the stationary phases for reversed-phase liquid chromatography with polar functional groups, J Chromatogr A. 1204 (2008). https://doi.org/10.1016/j.chroma.2008.07.056.

[14] S. Moldoveanu, V. David, Estimation of the phase ratio in reversed-phase high-performance liquid chromatography, J Chromatogr A. 1381 (2015). https://doi.org/10.1016/j.chroma.2015.01.034.

[15] A.C. Soare, V. David, S.C. Moldoveanu, Does phase ratio in reversed phase high performance liquid chromatography vary with temperature?, J Chromatogr A. 1620 (2020). https://doi.org/10.1016/j.chroma.2020.461023.

[16] A.R. Horner, R.E. Wilson, S.R. Groskreutz, B.E. Murray, S.G. Weber, Evaluation of three temperature- and mobile phase-dependent retention models for reversed-phase liquid chromatographic retention and apparent retention enthalpy, J Chromatogr A. 1589 (2019). https://doi.org/10.1016/j.chroma.2018.12.055.

[17] S.M. Graef, A.J.P. van Zyl, R.D. Sanderson, B. Klumperman, H. Pasch, Use of gradient, critical, and two-dimensional chromatography in the analysis of styrene- and methyl methacrylate-grafted epoxidized natural rubber, J Appl Polym Sci. 88 (2003) 2530–2538. https://doi.org/10.1002/app.12060.

[18] M.I. Malik, Critical parameters of liquid chromatography at critical conditions in context of poloxamers: Pore diameter, mobile phase composition, temperature and gradients, J Chromatogr A. 1609 (2020). https://doi.org/10.1016/j.chroma.2019.460440.

[19] M.A. Bashir, A. Brüll, W. Radke, Fast determination of critical eluent composition for polymers by gradient chromatography, Polymer (Guildf). 46 (2005) 3223–3229. https://doi.org/10.1016/j.polymer.2005.02.080.

[20] A.M. Skvortsov, A.A. Gorbunov, Achievements and uses of critical conditions in the chromatography of polymers, J Chromatogr A. 507 (1990) 487–496. https://doi.org/10.1016/S0021-9673(01)84228-X.

[21] A.M. Skvortsov, A.A. Gorbunov, D. Berek, B. Trathnigg, Liquid chromatography of macromolecules at the critical adsorption point: Behaviour of a polymer chain inside pores, Polymer (Guildf). 39 (1998) 423–429. https://doi.org/10.1016/S0032-3861(97)00279-6.

[22] S. Abrar, B. Trathnigg, Analysis of polyethyleneoxide macromonomers by liquid chromatography along the critical adsorption line, Anal Bioanal Chem. 400 (2011) 2577–2586. https://doi.org/10.1007/s00216-010-4554-2.

[23] M. Mlynek, W. Radke, Critical chromatography in ternary solvents, J Chromatogr A. 1284 (2013) 112–117. https://doi.org/10.1016/J.CHROMA.2013.02.005.

[24] A. van der Horst, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography of polymers, J Chromatogr A. 1000 (2003). https://doi.org/10.1016/S0021-9673(03)00495-3.

[25] X. Jiang, A. van der Horst, V. Lima, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography for the characterization of functional acrylate polymers, J Chromatogr A. 1076 (2005) 51–61. https://doi.org/10.1016/j.chroma.2005.03.135.

[26] S.R.A. Molenaar, B. van de Put, J.S. Desport, S. Samanipour, R.A.H. Peters, B.W.J. Pirok, Automated Feature Mining for Two-Dimensional Liquid Chromatography Applied to Polymers Enabled by Mass Remainder Analysis, Anal Chem. 94 (2022). https://doi.org/10.1021/acs.analchem.1c05336.

[27] F. Fitzpatrick, R. Edam, P. Schoenmakers, Application of the reversed-phase liquid chromatographic model to describe the retention behaviour of polydisperse macromolecules in gradient and isocratic liquid chromatography, J Chromatogr A. 988 (2003) 53–67. https://doi.org/10.1016/S0021-9673(02)02050-2.

[28] Y. Brun, P. Foster, Characterization of synthetic copolymers by interaction polymer chromatography: Separation by microstructure, J Sep Sci. 33 (2010) 3501–3510. https://doi.org/10.1002/jssc.201000572.

[29] L.E. Niezen, B.B.P. Staal, C. Lang, H.J.A. Philipsen, B.W.J. Pirok, G.W. Somsen, P.J. Schoenmakers, Recycling gradient-elution liquid chromatography for the analysis of chemical-composition distributions of polymers, J Chromatogr A. 1679 (2022) 463386. https://doi.org/10.1016/J.CHROMA.2022.463386.

[30] P. Schoenmakers, F. Fitzpatrick, R. Grothey, Predicting the behaviour of polydisperse polymers in liquid chromatography under isocratic and gradient conditions, J Chromatogr A. 965 (2002) 93–107. https://doi.org/10.1016/S0021-9673(01)01322-X.

[31] A.A. Gorbunov, A. v. Vakhrushev, Two-dimensional liquid chromatography of diblock copolymers: Simulation at various adsorption interaction conditions, J Chromatogr A. 1217 (2010). https://doi.org/10.1016/j.chroma.2010.05.045.

[32] F. Gritti, Perspective on the Future Approaches to Predict Retention in Liquid Chromatography, Anal Chem. 93 (2021). https://doi.org/10.1021/acs.analchem.0c05078.

[33] M.J. den Uijl, P.J. Schoenmakers, B.W.J. Pirok, M.R. van Bommel, Recent applications of retention modelling in liquid chromatography, J Sep Sci. 44 (2021) 88–114. https://doi.org/10.1002/jssc.202000905.

[34] P. Agrafiotou, C. Ràfols, C. Castells, E. Bosch, M. Rosés, Simultaneous effect of pH, temperature and mobile phase composition in the chromatographic retention of ionizable compounds, J Chromatogr A. 1218 (2011). https://doi.org/10.1016/j.chroma.2010.12.119.

[35] A.A. Gorbunov, A.M. Skvortsov, Statistical properties of confined macromolecules, Adv Colloid Interface Sci. 62 (1995) 31–108. https://doi.org/10.1016/0001-8686(95)00270-Z.

[36] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, J Chromatogr A. 1217 (2010) 3794–3803. https://doi.org/10.1016/j.chroma.2010.04.023.

[37] L.M. Blumberg, Migration and elution equations in gradient liquid chromatography, J Chromatogr A. 1599 (2019) 35–45. https://doi.org/10.1016/j.chroma.2019.03.057.

[38] L.M. Blumberg, Theory of gradient elution liquid chromatography with linear solvent strength: Part 1. migration and elution parameters of a solute band, Chromatographia. 77 (2014) 179–188. https://doi.org/10.1007/s10337-013-2555-y.

[39] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, Anal Chem. 77 (2005) 5670–5677. https://doi.org/10.1021/ac0506783.

[40] P. Jandera, J. Churáček, Gradient elution in liquid chromatography. II. Retention characteristics (retention volume, band width, resolution, plate number) in solvent-programmed chromatography - theoretical considerations, J Chromatogr A. 91 (1974) 223–235. https://doi.org/10.1016/S0021-9673(01)97902-6.

[41] P. Jandera, M. Holčapek, L. Kolářová, Retention mechanism, isocratic and gradient-elution separation and characterization of (co)polymers in normal-phase and reversed-phase high-performance liquid chromatography, in: J Chromatogr A, 2000: pp. 65–84. https://doi.org/10.1016/S0021-9673(99)01216-9.

[42] P.J. Schoenmakers, H.A.H. Billiet, R. Tussen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, J Chromatogr A. 149 (1978) 519–537. https://doi.org/10.1016/S0021-9673(00)81008-0.

[43] L.R. Snyder, J.W. Dolan, High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model, 2006. https://doi.org/10.1002/0470055529.

[44] M.A. Bashir, W. Radke, Comparison of retention models for polymers. 1. Poly(ethylene glycol)s, J Chromatogr A. 1131 (2006) 130–141. https://doi.org/10.1016/j.chroma.2006.07.089.

[45] A.M. Striegel, Method development in interaction polymer chromatography, TrAC - Trends in Analytical Chemistry. 130 (2020). https://doi.org/10.1016/j.trac.2020.115990.

[46] T.S. Bos, L.E. Niezen, M.J. den Uijl, S.R.A. Molenaar, S. Lege, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Reducing the influence of geometry-induced gradient deformation in liquid chromatographic retention modelling, J Chromatogr A. 1635 (2021). https://doi.org/10.1016/j.chroma.2020.461714.

[47] R. Swart, S. Brouwer, J.C. Kraak, H. Poppe, Swelling behaviour and kinetic performance of polyacrylate stationary phases for reversed-phase and normal-phase open-tubular liquid chromatography, J Chromatogr A. 732 (1996). https://doi.org/10.1016/0021-9673(95)01289-3.

[48] A. Velayudhan, M.R. Ladisch, Effect of modulator sorption in gradient elution chromatography: gradient deformation, Chem Eng Sci. 47 (1992). https://doi.org/10.1016/0009-2509(92)80217-Z.

[49] F. Gritti, G. Guiochon, Separations by gradient elution: Why are steep gradient profiles distorted and what is their impact on resolution in reversed-phase liquid chromatography, J Chromatogr A. 1344 (2014). https://doi.org/10.1016/j.chroma.2014.04.010.

[50] F. Gritti, G. Guiochon, The distortion of gradient profiles in reversed-phase liquid chromatography, J Chromatogr A. 1340 (2014) 50–58. https://doi.org/10.1016/j.chroma.2014.03.004.

[51] S. Fekete, M. Fogwill, M.A. Lauber, Pressure-Enhanced Liquid Chromatography, a Proof of Concept: Tuning Selectivity with Pressure Changes and Gradients, Anal Chem. 94 (2022) 7877–7884. https://doi.org/10.1021/acs.analchem.2c00464.

[52] M. Martin, G. Guiochon, Effects of high pressure in liquid chromatography, J Chromatogr A. 1090 (2005) 16–38. https://doi.org/10.1016/j.chroma.2005.06.005.

[53] G. Guiochon, M.J. Sepaniak, Influence of pressure on solute retention in liquid chromatography, J Chromatogr A. 606 (1992) 248–250. https://doi.org/10.1016/0021-9673(92)87031-3.

[54] X. Liu, D. Zhou, P. Szabelski, G. Guiochon, Influence of pressure on the retention and separation of insulin variants under linear conditions, Anal Chem. 75 (2003) 3999–4009. https://doi.org/10.1021/ac0205964.

[55] M. Martin, G. Blu, G. Guiochon, The effect of pressure on the retention time and the retention volume of an inert compound in liquid chromatography, J Chromatogr Sci. 11 (1973) 641–654. https://doi.org/10.1093/chromsci/11.12.641.

[56] X. Jiang, A. van der Horst, P.J. Schoenmakers, Breakthrough of polymers in interactive liquid chromatography, J Chromatogr A. 982 (2002) 55–68. https://doi.org/10.1016/S0021-9673(02)01483-8.

[57] G. Glöckner, High-performance precipitation liquid chromatography, Trends in Analytical Chemistry. 4 (1985). https://doi.org/10.1016/0165-9936(85)87062-X.

[58] G. Glöckner, D. Wolf, H. Engelhardt, Separation of copoly(styrene/acrylonitrile) samples according to composition under reversed phase conditions, Chromatographia. 32 (1991) 107–112. https://doi.org/10.1007/BF02325011.

[59] S. Nawada, F. Gritti, Theoretical framework for mixer design for noise reduction and gradient fidelity, J Chromatogr A. 1653 (2021). https://doi.org/10.1016/j.chroma.2021.462357.

[60] B.W.J. Pirok, S.R.A. Molenaar, L.S. Roca, P.J. Schoenmakers, Peak-Tracking Algorithm for Use in Automated Interpretive Method-Development Tools in Liquid Chromatography, Anal Chem. 90 (2018). https://doi.org/10.1021/acs.analchem.8b03929.

[61] S.R.A. Molenaar, T.A. Dahlseid, G.M. Leme, D.R. Stoll, P.J. Schoenmakers, B.W.J. Pirok, Peak-tracking algorithm for use in comprehensive two-dimensional liquid chromatography – Application to monoclonal-antibody peptides, J Chromatogr A. 1639 (2021). https://doi.org/10.1016/j.chroma.2021.461922.

[62] H.J.A. Philipsen, M. Oestreich, B. Klumperman, A.L. German, Characterization of low-molar-mass polymers by gradient polymer elution chromatography. III. Behaviour of crystalline polyesters under reversed-phase conditions, J Chromatogr A. 775 (1997). https://doi.org/10.1016/S0021-9673(97)00208-2.

[63] G. Desmet, Theoretical calculation of the retention enthalpy effect on the viscous heat dissipation band broadening in high performance liquid chromatography columns with a fixed wall temperature, J Chromatogr A. 1116 (2006) 89–96. https://doi.org/10.1016/j.chroma.2006.03.024.

[64] K. Vanderlinden, K. Broeckhoven, Y. Vanderheyden, G. Desmet, Effect of pre- and post-column band broadening on the performance of high-speed chromatography columns under isocratic and gradient conditions, J Chromatogr A. 1442 (2016). https://doi.org/10.1016/j.chroma.2016.03.016.

[65] P.J. Schoenmakers, J.K. Strasters, Á. Bartha, Correction of the resolution function for non-ideal peaks, J Chromatogr A. 458 (1988). https://doi.org/10.1016/S0021-9673(00)90578-8.

[66] J.A. Navarro-Huerta, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Estimation of peak capacity based on peak simulation, J Chromatogr A. 1574 (2018). https://doi.org/10.1016/j.chroma.2018.09.009.

[67] A.J.P. Martin, R.L.M. Synge, A new form of chromatogram employing two liquid phases. 1. A theory of chromatography 2. Application to the micro-determination of the higher monoamino-acids in proteins, Biochem Journal. 35 (1941) 1358–1368. https://doi.org/10.1042/bj0351358.

[68] U.D. Neue, D.H. Marchand, L.R. Snyder, Peak compression in reversed-phase gradient elution, J Chromatogr A. 1111 (2006) 32–39. https://doi.org/10.1016/j.chroma.2006.01.104.

[69] H. Poppe, J. Paanakker, M. Bronckhorst, Peak width in solvent-programmed chromatography. I. General description of peak broadening in solvent-programmed elution, J Chromatogr A. 204 (1981) 77–84. https://doi.org/10.1016/S0021-9673(00)81641-6.

[70] J.J. van Deemter, Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. J. J. van Deemter, F. J. Zuiderweg and A. Klinkenberg, Chem. Engng Sci. 5 271-289, 1956, Chem Eng Sci. 50 (1995). https://doi.org/10.1016/0009-2509(96)81812-4.

[71] W. de Malsche, H. Eghbali, D. Clicq, J. Vangelooven, H. Gardeniers, G. Desmet, Pressure-driven reverse-phase liquid chromatography separations in ordered nonporous pillar array columns, Anal Chem. 79 (2007). https://doi.org/10.1021/ac070352p.

[72] F. Gritti, General theory of peak compression in liquid chromatography, J Chromatogr A. 1433 (2016) 114–122. https://doi.org/10.1016/j.chroma.2016.01.032.

[73] A. Felinger, Molecular dynamic theories in chromatography, J Chromatogr A. 1184 (2008). https://doi.org/10.1016/j.chroma.2007.12.066.

[74] K.M. Usher, C.R. Simmons, J.G. Dorsey, Modeling chromatographic dispersion: A comparison of popular equations, J Chromatogr A. 1200 (2008). https://doi.org/10.1016/j.chroma.2008.05.073.

[75] F. Gritti, G. Guiochon, The van Deemter equation: Assumptions, limits, and adjustment to modern high performance liquid chromatography, J Chromatogr A. 1302 (2013). https://doi.org/10.1016/j.chroma.2013.06.032.

[76] K. Broeckhoven, G. Desmet, Methods to determine the kinetic performance limit of contemporary chromatographic techniques, J Sep Sci. 44 (2021). https://doi.org/10.1002/jssc.202000779.

[77] J.H. Knox, M. Saleem, Kinetic conditions for optimum speed and resolution in column chromatography, J Chromatogr Sci. 7 (1969). https://doi.org/10.1093/chromsci/7.10.614.

[78] G. Desmet, D. Clicq, P. Gzil, Geometry-independent plate height representation methods for the direct comparison of the kinetic performance of LC supports with a different size or morphology, Anal Chem. 77 (2005). https://doi.org/10.1021/ac050160z.

[79] J.P. Foley, Resolution Equations, Analyst. 116 (1991).

[80] J. Calvin Giddings, Maximum Number of Components Resolvable by Gel Filtration and Other Elution Chromatographic Methods, Anal Chem. 39 (1967). https://doi.org/10.1021/ac60252a025.

[81] J.M. Davis, J.C. Giddings, Statistical Theory of Component Overlap in Multicomponent Chromatograms, Anal Chem. 55 (1983). https://doi.org/10.1021/ac00254a003.

[82] G. Guiochon, The limits of the separation power of unidimensional column liquid chromatography, J Chromatogr A. 1126 (2006) 6–49. https://doi.org/10.1016/j.chroma.2006.07.032.

[83] J.C. Giddings, Dynamics of chromatography: Principles and theory, 2017. https://doi.org/10.1201/9781315275871.

[84] H. Poppe, Some reflections on speed and efficiency of modern chromatographic methods, in: J Chromatogr A, 1997. https://doi.org/10.1016/S0021-9673(97)00376-2.

[85] S. Eeltink, S. Dolman, G. Vivo-Truyols, P. Schoenmakers, R. Swart, M. Ursem, G. Desmet, Selection of column dimensions and gradient conditions to maximize the peak-production rate in comprehensive off-line two-dimensional liquid chromatography using monolithic columns, Anal Chem. 82 (2010). https://doi.org/10.1021/ac101514d.

[86] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E.G. Porter, S.C. Rutan, Fast, comprehensive two-dimensional liquid chromatography, J Chromatogr A. 1168 (2007). https://doi.org/10.1016/j.chroma.2007.08.054.

[87] G. Vanhoenacker, I. Vandenheede, F. David, P. Sandra, K. Sandra, Comprehensive two-dimensional liquid chromatography of therapeutic monoclonal antibody digests, Anal Bioanal Chem. 407 (2015). https://doi.org/10.1007/s00216-014-8299-1.

[88] B.W.J. Pirok, A.F.G. Gargano, P.J. Schoenmakers, Optimizing separations in online comprehensive two-dimensional liquid chromatography, J Sep Sci. 41 (2018). https://doi.org/10.1002/jssc.201700863.

[89] G. Guiochon, N. Marchetti, K. Mriziq, R.A. Shalliker, Implementations of two-dimensional liquid chromatography, J Chromatogr A. 1189 (2008). https://doi.org/10.1016/j.chroma.2008.01.086.

[90] J.A. Abia, J. Putnam, K. Mriziq, G.A. Guiochon, Design and implementation of an array of micro-electrochemical detectors for two-dimensional liquid chromatography-Proof of principle, J Chromatogr A. 1217 (2010). https://doi.org/10.1016/j.chroma.2010.01.021.

[91] G. Glöckner, Gradient HPLC of Copolymers and Chromatographic Cross-Fractionation, 1991. https://doi.org/10.1007/978-3-642-75799-0.

[92] J.C. Giddings, Sample dimensionality: A predictor of order-disorder in component peak distribution in multidimensional separation, J Chromatogr A. 703 (1995). https://doi.org/10.1016/0021-9673(95)00249-M.

[93] R.E. Murphy, M.R. Schure, J.P. Foley, Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography, Anal Chem. 70 (1998). https://doi.org/10.1021/ac971184b.

[94] D.R. Stoll, P.W. Carr, Two-Dimensional Liquid Chromatography: A State of the Art Tutorial, Anal Chem. 89 (2017). https://doi.org/10.1021/acs.analchem.6b03506.

[95] B.W.J. Pirok, D.R. Stoll, P.J. Schoenmakers, Recent Developments in Two-Dimensional Liquid Chromatography: Fundamental Improvements for Practical Applications, Anal Chem. 91 (2019) 240–263. https://doi.org/10.1021/acs.analchem.8b04841.

[96] T.S. Bos, W.C. Knol, S.R.A. Molenaar, L.E. Niezen, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Recent applications of chemometrics in one- and two-dimensional chromatography, J Sep Sci. 43 (2020). https://doi.org/10.1002/jssc.202000011.

[97] R.J. Vonk, A.F.G. Gargano, E. Davydova, H.L. Dekker, S. Eeltink, L.J. de Koning, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography with stationary-phase-assisted modulation coupled to high-resolution mass spectrometry applied to proteome analysis of saccharomyces cerevisiae, Anal Chem. 87 (2015). https://doi.org/10.1021/acs.analchem.5b00708.

[98] D.R. Stoll, K. Shoykhet, P. Petersson, S. Buckenmaier, Active Solvent Modulation: A Valve-Based Approach to Improve Separation Compatibility in Two-Dimensional Liquid Chromatography, Anal Chem. 89 (2017). https://doi.org/10.1021/acs.analchem.7b02046.

[99] M. Verstraeten, M. Pursch, P. Eckerle, J. Luong, G. Desmet, Thermal Modulation for Multidimensional Liquid Chromatography Separations Using Low-Thermal-Mass Liquid Chromatography (LC), Anal Chem. 83 (2011) 7053–7060. https://doi.org/10.1021/ac201207t.

[100] L.E. Niezen, B.B.P. Staal, C. Lang, B.W.J. Pirok, P.J. Schoenmakers, Thermal modulation to enhance two-dimensional liquid chromatography separations of polymers, J Chromatogr A. 1653 (2021). https://doi.org/10.1016/j.chroma.2021.462429.

Chapter 2

Thermal modulation to enhance 2D-LC separations of polymers

Abstract

Many materials used in a wide range of fields consist of polymers that feature great structural complexity. One particularly suitable technique for characterizing these complex polymers, that often feature correlated distributions in *e.g.* microstructure, chemical composition, or molecular weight, is comprehensive two-dimensional liquid chromatography (LC×LC). For example, using a combination of reversed-phase LC and size-exclusion chromatography (RPLC×SEC). Efficient and sensitive LC×LC often requires focussing of the analytes between the two stages. For the analysis of large-molecule analytes, such as synthetic polymers thermal modulation (or cold trapping) may be feasible and this approach is studied for the analysis of a styrene/butadiene "star" block copolymer. Trapping efficiency is evaluated qualitatively by monitoring the effluent of the trap with an evaporative light-scattering detector and quantitatively by determining the recovery of polystyrene standards from RPLC×SEC experiments. The recovery was dependent on the molecular weight and the temperatures of the first-dimension column and of the trap, and ranged from 46% for a molecular weight of 2.78 kDa to 86% (or up to 94.5% using an optimized set-up) for a molecular weight of 29.15 kDa, all at a first-dimension-column temperature of 80 °C and a trap temperature of 5 °C. Additionally a strategy to reduce the pressure pulse from the modulation has been developed, bringing it down from several tens of bars to only a few bar.

Publication: L.E. Niezen, B.B.P. Staal, C. Lang, B.W.J. Pirok, P.J. Schoenmakers, Thermal modulation to enhance two-dimensional liquid chromatography separations of polymers, J Chromatogr A. 1653 (2021). https://doi.org/10.1016/j.chroma.2021.462429.

2.1. Introduction

High-performance liquid chromatography (HPLC) is one of the most prevalent techniques for the analysis of soluble samples. Both practice and theory have proven that LC is limited in terms of the separation power that can be achieved within a given timespan, depending on the operating pressure [1]. Ultra-high-pressure liquid chromatography (UHPLC) allows for faster or more-efficient separations, but the gain of about a factor of four in maximum pressure (and achievable number of theoretical plates) in moving from HPLC to UHPLC only results in a factor of two increase in separation power (resolution). To gain more information on complex samples, LC is oftentimes hyphenated to mass spectrometry (MS) or even high-resolution mass-spectrometry (HRMS), typically by utilizing an electrospray (ESI) interface. It is well-known, however, that such an approach is rarely feasible for polymer analysis [2], as it is limited to relatively small and polar polymers unless supercharging is utilized [3,4]. Larger (sufficiently polar and narrowly distributed) polymers can be analysed by matrix-assisted laser-desorption/ionization (MALDI) MS. However, MALDI cannot easily be interfaced with LC and is ultimately still molecular-weight limited, even after pre-fractionation with LC. For relatively high-molecular-weight polymers multidimensional chromatography offers additional selectivity, separation power and, thus, information. For example, combined chemical-composition and molecular-weight distributions can be obtained from the structured chromatograms generated by comprehensive two-dimensional liquid chromatography (LC×LC) [5,6]. Two-dimensional LC (2D-LC) may be applied in one of three modes, *viz.* heart-cutting (LC-LC), multiple-heart cutting (mLC-LC) or comprehensive (LC×LC) [5]. During an LC×LC separation, the entire effluent from the first dimension is subjected to an additional separation in many small fractions, leading much higher peak capacities and peak production rates (peak capacity per unit time) than 1D-LC. LC×LC has seen several significant developments in recent years, many of which focused on the interface ("modulator") between the first and second dimension. Examples include the use of active-solvent modulation (ASM) [7] and stationary-phase-assisted modulation (SPAM) [8]. A reaction chamber may be incorporated between the two separations [9] so that additional structural information may be obtained. Both ASM and SPAM aim to alleviate incompatibility issues between the first and second dimensions, primarily focusing on solvent incompatibility, but also allowing narrow second-dimension (^{2}D) columns and low ^{2}D flowrates to be used, reducing analyte

dilution and improving compatibility with MS. Briefly, in the case of ASM this is achieved by diluting the fraction collected in the loop, while SPAM achieves focussing and a switch of solvents by replacing the conventional sample loops by short, so-called "trap" columns containing a suitable stationary phase. Both ASM and SPAM can allow for a focusing or reconcentration of the analyte, in the case of ASM this may be achieved at the inlet of the ^{2}D column, while in SPAM it occurs in the trap column. One of the most significant advantages of SPAM when compared to ASM is that the ^{1}D eluent can be completely eliminated from the system, not just diluted. Disadvantages of SPAM include the need to develop methods for specific applications (depending on the ^{1}D eluent, the ^{2}D eluent and the analytes) and the limited life-time of the trap columns, which may be related to pressure pulses [10]. One strategy to improve the life-time of the trap columns may be to synchronize the modulation with the pump-frequency (pump-frequency-synchronized modulation, PFSM; *vide infra*).

Trapping or focusing may also be achieved by means of a difference in temperature [11–18] rather than eluent strength. This was first demonstrated for off-line 2D-LC by Verstraeten *et al.* [11] using capillary columns packed with porous graphitic carbon (PGC) as a trapping device. By first cooling and then rapidly heating (1200 °C/min) this column, neutral analytes could be successfully trapped and a concentration enhancement factor of 18 could be achieved. A form of thermal modulation called temperature-assisted on-column solute focusing (TASF) was also demonstrated, initially for parabens as analytes, in capillary 1D-LC by Groskreutz *et al.* [12,13] In their approach analytes were focused by cooling the column inlet using Peltier devices, after which the inlet was rapidly heated to "inject" the analytes as a narrow band. Another thermal approach to allow for focussing of the analytes and solvent switching was developed by van de Ven *et al.* [18]. In this "in-column focusing" approach the analytes were first loaded into a modulation column in the initial mobile phase at a relatively high temperature, after which the modulation column was cooled down and the analytes were eluted in the backflush mode with a stronger solvent, which allowed for the analytes to leave the zone of initial mobile phase if their retention increased with the decrease in temperature, and subsequent refocusing into a more narrow band.

Most of the work described above has been carried out using 1D-LC, either to allow for better sensitivity in capillary LC or with the eventual aim of applying the method in LC×LC. Thermal focussing in 1D-LC may be practically useful, as a relatively straightforward way to help concentrate the analytes if other means of focusing, such as injection in a weak eluent, cannot be effectively applied. However, when thermal focusing is to be applied for modulation in 2D-LC, the cooling and heating must be performed repeatedly and much-more rapidly, which make the concept much-more challenging. Typically, trap columns have a very small internal volume and contain a more-hydrophobic stationary phase than used in the ^{1}D column [5,11] In the case of polymers many of these issues are avoided simply due to their retention characteristics. Because retention varies much-more strongly with mobile-phase composition or temperature for polymers than for small-molecule analytes, thermal-modulation strategies may be feasible for their separation by 2D-LC. For the 2D RPLC×SEC analysis of polymers there are obvious benefits of using a trapping strategy. Thanks to a lowered ^{2}D injection volume, efficient small-particle SEC columns can be used that facilitate fast, highly sensitive, and high-resolution separations [19]. Also, the ^{2}D column may be narrower than the ^{1}D column, further enhancing the mass sensitivity of the analysis and greatly reducing the amount of eluent required. However, thermal strategies may exacerbate issues around the lifetime of the traps and the switching-induced pressure pulses, since cooling down the trap column will locally increase the viscosity of the mobile phase.

The objective of the present work is to demonstrate thermal modulation as an easy-to-implement means to achieve fast and efficient two-dimensional polymer separations. We first aim to demonstrate that the cold-trapping principle can be applied to polystyrene standards in simple 1D-LC experiments and we set out to study the applicable range of molecular weights. Subsequently, we aim to extend the approach to LC×LC separations of a polystyrene/polybutadiene star block copolymer. Our final objective is to create a robust system that can be used for a large number of LC×LC analysis without intervention.

2.2. Theory

In all cases the principle underlying the focusing of the analyte may be described by known retention models [20–22]. In reversed-phase (RP) LC it is generally accepted that the retention of an analyte may be approximately described by a log-linear relationship between retention factor and solvent composition. This is often termed the linear-solvent-strength (LSS) model and it is described by **Equation 2.1:**

$$\ln k = \ln k_0 - S\varphi \tag{2.1}$$

With k_0 the retention factor extrapolated to a composition of 100% weak solvent, S is the slope which determines how quickly k changes with φ, and φ the volume fraction of strong solvent in the mobile phase. Hence reducing the fraction of strong solvent, increases retention, as long as S is positive. Generally, the higher the slope in the LSS curve, the easier it will be to trap the analyte, for example by dilution of the eluent with weak solvent. Typically, solvent-based focusing occurs more readily at ambient or sub-ambient temperatures, because for most analytes retention decreases with increasing temperature, implying that a lower solvent strength (*i.e.* a lower fraction of strong solvent) will be required to achieve the same retention. However, typically the effect of solvent composition will be much greater than the effect of temperature, which is the primary reason why thermal modulation for small analytes requires highly retentive stationary phases (such as PGC in the RPLC mode). In those cases the temperature is mainly utilized to decrease the time it takes for the analytes to elute from the trap (*i.e.* reduced peak width). In case of typical gradient separations analytes are expected to be less focused at a particular composition when temperature is increased, unless the starting composition of the gradient is altered (to lower fraction of strong solvent) concomitantly. This effect of temperature on retention implies that thermal modulation can be applied for focusing or trapping. The effectiveness of this strategy depends on the analytes' retention as a function of temperature, which can be described by the van 't Hoff equation, **Equation 2.2:**

$$\ln k = -\frac{\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R} - \ln \beta \tag{2.2}$$

With ΔH° the molar enthalpy of solute transfer between phases, ΔS° the corresponding entropy change, R the universal gas constant, T the absolute temperature (in Kelvin) and β the volumetric phase ratio. The plot of $\ln k$ versus $\frac{1}{T}$ is

called a van 't Hoff plot. In most cases linear van 't Hoff behaviour is observed, and the slope of the plot allows $\Delta H°$ to be determined across a certain temperature range. Differences in $\Delta H°$ for different components then result in varying selectivity of an LC separation with temperature. Thermal modulation can be achieved more easily with a given temperature difference if the slope of the van 't Hoff plot is larger (*i.e.* at larger $\Delta H°$). However, the effect of temperature on retention is much smaller than the effect of mobile-phase composition. As a rule-of-thumb, a change of 5 to 10 °C corresponds to a change of only about 1% mobile-phase composition for small compounds [23]. In many of the examples in literature a reasonably large change in temperature was therefore required to focus the analytes [11]. For most compounds a lower recovery is experienced when using thermal modulation, as the large temperature differences required for trapping and the rigorous cooling and heating cycles to achieve proper transfer from trap column to ^{2}D column can be difficult to realize. Apart from the large temperature differences, highly retentive stationary phases, such as porous graphitic carbon (PGC), have proven to be required. However, for compounds with high molecular weights thermal modulation may be more attractive, because the enthalpy of transfer (the slope of the van 't Hoff plot) typically increases with increasing molecular weight [22,24,25].

The high slope in both the LSS and the van 't Hoff plot means that higher molecular-weight polymers generally require only a very small change in either mobile phase composition or temperature to achieve trapping compared to most small, uncharged, analytes, at their time of elution from the ^{1}D column. A combined use of a gradient ^{1}D separation operated at high temperature and the use of thermal modulation prior to the ^{2}D separation therefore benefits in two ways. Firstly, due to the high LSS slope polymers will elute at or close to a specific mobile phase composition, unlike small analytes which may be more strongly affected by the gradient slope due to the changing equilibrium while moving through the column. Simultaneously, these analytes will also have a high slope in the van' t Hoff plot, which means that the composition at which the analyte elutes will be more greatly influenced by the temperature than a small analyte. Both of these aspects suggest that a small change in temperature will be sufficient to retain the analyte within the trap. Of course, it is expected that this will become increasingly more challenging the higher the gradient rate and the smaller the polymer. In both cases the elution

composition of the polymer at the trap temperature may already be reached by the mobile phase before the analyte reaches the trap, resulting in an insufficient difference in retention at the trap.

2.3. Materials and Methods

2.3.1. Chemicals and Materials

A 10 port 2-position UHPLC valve (MXT715-102) was purchased from Rheodyne, IDEX (Lake Forest, IL, USA). An Arduino Uno Rev 3 was purchased from a local electronics supplier. Acetronitrile (ACN, ≥99.9%, LC-MS Grade) was purchased from Honeywell Research Chemicals (Seelze, Germany), Tetrahydrofuran (THF, 99.9%, Isocratic grade, non-stabilized) was purchased from Bernd Kraft (Oberhausen, Germany), MilliQ Water was obtained using a purification system purchased from MilliPore (Burlington, MA, USA). An EasiCal polystyrene-standards kit was purchased from Agilent (Waldbronn, Germany), while the Styrolux 693D sample was obtained from BASF (Ludwigshafen am Rhein, Germany).

Columns used during testing included two 150 mm length × 2.1 mm I.D. APC SEC columns packed with 2.5 µm ethylene bridged-hybrid (BEH) particles with 450-Å pore size, and a single 50 × 4.6 mm XBridge BEH Shield RP18 XP column containing 2.5-µm particles with 130-Å pore size, all purchased from Waters (Milford, MA, USA). For the trapping columns two 2.1 × 5.0 mm, XBridge BEH C18 XP VanGuard Cartridges were used containing 2.5-µm particles with 130-Å pores, also purchased from Waters.

2.3.2. Equipment and software

The system used for testing included a (G1322A) 1260 degasser, a (G1311A) 1100 quaternary pump, a (G5667A) 1260 HiP auto-sampler, a (G4260B) 1260 Infinity evaporative light-scattering detector (ELSD), a (G1314D) variable-wavelength detector (VWD), and a (G1316A) 1100 column oven, all purchased from Agilent, as well as an Acquity system, including a p-isocratic solvent manager (isocratic pump), sample manager pFTN (autosampler), column manager S (column oven), photodiode-array detector with taper slit and refractive-index detector; purchased from Waters. Cooling was performed using a Huber ministat v3.03 purchased from HUBER SE (Berching, Germany).

Data acquisition was performed using WinGPC software purchased from PSS Polymer Standards Service (Mainz, Germany). The Acquity system was controlled using Empower-3 software purchased from Waters. Data analysis was performed in MATLAB R2020a (Mathworks, Woodshole, MA, USA).

2.3.3. Introducing cold trapping

The 2D-LC cold-trap set-up used is illustrated in **Figure 2.1.** A Huber ministat v3.03 was utilized to cool and circulate a mixture of isopropyl alcohol (IPA) and mineral oil through an aluminium block, in which holes were drilled to hold the trapping columns in place. The columns themselves were chosen based on their small volume (approximately 10 µL) and contained the same C18 silica-based stationary phase as used in the ^{1}D column. The aluminium block was cooled to approximately 5 °C (unless otherwise specified) by continuously flushing a cold mixture of IPA and mineral oil through the inside of the holder, a thermocouple was utilized to measure the temperature. The first-dimension column was held at 80 °C, resulting in a temperature difference of 75 °C between the column and the aluminium block. In the current experiments solvents were not preheated before entering the column and were not precooled before entering the trap.

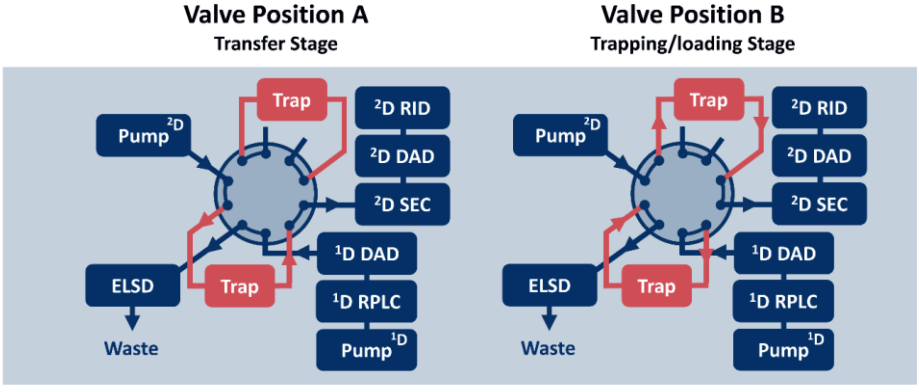

Figure 2.1: Schematic illustrating the 2D-LC cold-trap set-up, with on the lefthand side valve position A, and the righthand side valve position B. Colours are not indicative of temperature.

In case of the 1D-LC experiments, a DAD was placed directly after the RPLC column. The trap was placed after the first DAD and its outlet was connected to a second DAD. This allowed us to clearly monitor the effect of the trap on polymer retention

and compare the modulation set-up to conventional RPLC experiments. In the current work a single trap was used for the trapping, while a secondary trap was used to ensure that the backpressure between valve position A and B remained similar when the ^{2}D SEC pump was not transferring the contents from trap A to the SEC column. The modulations consisted of two phases: a loading phase, and a transfer phase. Unless otherwise specified the duration of the loading phase was 74.8 s, while the duration of the transfer phase was 4.4 s. The decision to use a single trap in this case was made to ensure that solely the effects of temperature on the trapping were studied. Any effects that may result from differences between the two trap columns are excluded from the observations.

2.4. Results and discussion

2.4.1 Pump-frequency synchronized modulation

It is known that many columns may suffer from a sharp increase in pressure that either occurs when switching the modulator valve between positions A and B or as a result from the very steep gradients that may be used in the second dimension. This seems to be especially the case for very low-volume columns, such as the guard columns used for trapping in this study. Even in the case of an isocratic second dimension, as used in the present work, LC×LC cannot generally be carried out without performing modulations (with the exception of spatial two-dimensional separations [26–28]), and hence this issue affects any LC×LC system. Such sharp pressure pulses may have a negative impact on the lifetime of the second-dimension column and they may cause variations in the flow, resulting in a worse repeatability of LC×LC measurements [10]. To reduce the pressure pulses resulting the modulations a strategy was designed in which the modulation time was adjusted to the pump frequency. As the isocratic pump used had accessible pressure sensors in both the accumulator and primary pump heads, the read outs could be fed to the WinGPC software used to control the LC×LC experiments. This allowed monitoring the positions of the pistons inside the pump head and the frequency at which these moved. The trace obtained from such measurements is illustrated in blue in **Figure 2.2-A**, which corresponds to the piston movement inside the primary pump**.**

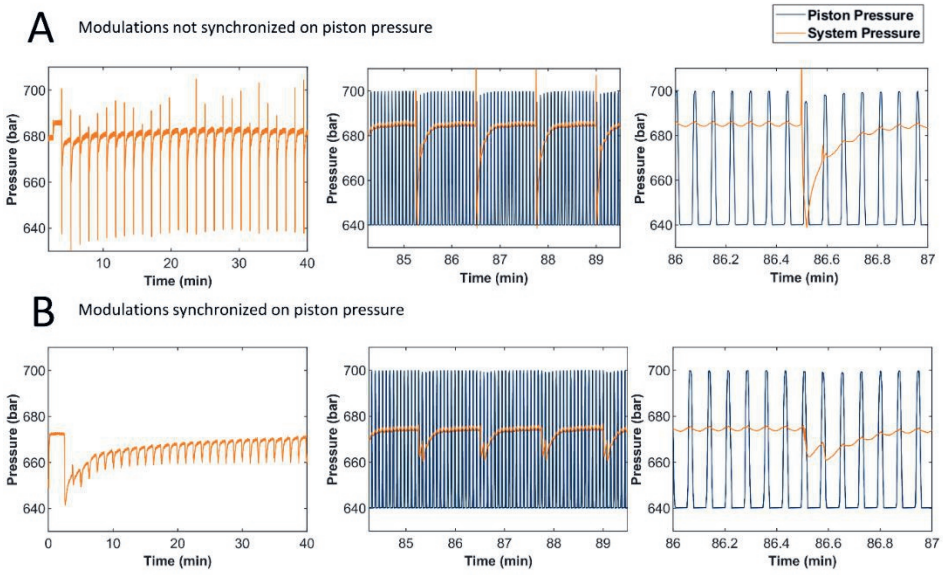

Figure 2.2: A) Pressure profile in case of normal, unsynchronized, modulation, B) Pressure profiles when synchronizing piston movement and modulation. Left: overview of the pressure during the first 40 minutes of the separation; middle: system and piston pressure during the final modulations; right: expansion of the middle figures.

In our case we are performing SEC in the 2nd dimension, where we are using an isocratic pump, consisting of a combination of a primary pump and an accumulator pump (dual-piston in-series, see Supplementary Material **Figure S-1**). The modulations are synchronized with the piston movement by reading out the pressure sensor using an Arduino-Uno microcontroller, which directs the modulations at a frequency corresponding to that of the piston movement. The latter will remain constant at constant flow. The resulting traces are shown in **Figure 2.2-B.** The results show that the magnitude of the pressure spikes in the second dimension due to the modulation (orange signal) can be significantly reduced using this strategy. Furthermore, when comparing the traces of the pressure inside the primary pump head (blue signals) it can be seen that without synchronization (**Figure 2.2-A**, middle/right) the pump responds to an increase in the system pressure (orange signal) by reducing its movement (lower pressure), as is evident from the small decrease in the tops of the blue trace after the modulation. This can be a source of flowrate inaccuracies. The effect is reduced when synchronizing the modulation with

the piston stroke (**Figure 2.2-B**, middle/right). At this stage there is insufficient evidence to proof that the lifetime of the trap columns increases, but based on experience elsewhere [10], it is reasonable to assume this to be the case. The synchronization method also allows operation closer to the pressure limit of the system while avoiding a pump shutdown, so that UHPLC systems can be used to their full potential.

2.4.2. Cold-trap set-up and 1D experiments

2.4.2.1Illustrating the principle by 1D-LC experiments

To quickly assess whether a particular compound can be focused in the cold-trap, 1D-LC experiments were performed. In this case two DAD detectors were installed, one before and one after the trap, to monitor the change in retention times and peak profiles. A linear gradient from 0 to 100% ACN to THF was run in 10 minutes. This resulted in the following chromatograms shown in **Figure 2.3** for a selection of polystyrene standards.

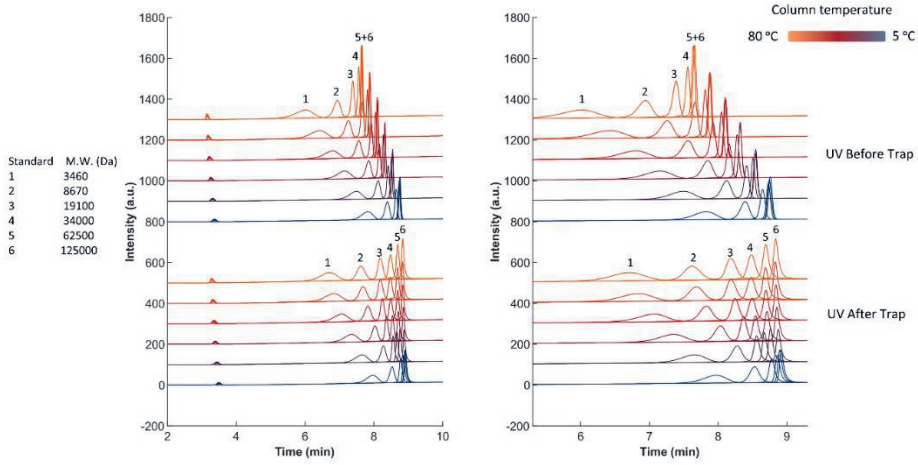

Figure 2.3: Gradient-elution chromatograms recorded at 254 nm with a cold-trap installed after the column, with uninterrupted flow and trap temperature of 5 °C throughout. Line colour indicates column temperature. Left: full chromatograms; right: expansion of 5 to 9 min range. Injection of individual polystyrenes of different molecular weight ranging from 3.5 to 125 kDa.

From the first set (upper) traces in **Figure 3** it is clear that the low-molecular-weight standards elute before the higher molecular-weight standards. The latter elute

increasingly close together, approaching the pseudo-critical point for polystyrene for this combination of stationary and mobile phases, *i.e.* the composition at which retention becomes independent of molecular weight in this gradient. This pseudo-critical point is seen to shift towards longer elution times (higher fractions of strong solvent) at lower column temperatures. When inspecting the second set of traces (bottom), recorded using the detector located after the trap, it can be seen that a significant gain in resolution (from $R_s = 0.0842$ to $R_s = 0.995$ for standard 5 and 6, for a column temperature of 80 °C) could be achieved for the highest molecular weight standards. This additional resolution indicated that a separation was occurring within the trap. Our current explanation for this additional separation occurring in the very small trap (volume of about 10 µL) is based on three effects. Firstly, it is assumed that the high-molecular-weight polystyrenes are adsorbed at the start of the 1D-LC column and only start moving with the mobile phase once a composition close to the critical composition is approached. This is consistent with prior observations and explanations [29]. Secondly, all these polystyrenes reach the trap nearly simultaneously where, due to the lower temperature, the polystyrene standards are significantly more retained (*i.e.* "trapped"). Thirdly and finally, in the trap column the standards then essentially experience a second gradient step. Due to the very small volume of the trap this second gradient is extremely shallow, since the effective slope of a (LSS) gradient can be defined as:

$$b = V_m \Delta\varphi S / (t_g F) \tag{2.3}$$

In which V_m is the column void volume, $\Delta\varphi$ is the change in mobile phase composition such that for a 0-100%B gradient $\Delta\varphi = 1$, t_g is the gradient duration, F is the mobile phase flowrate and S is a compound-specific parameter that describes the variation of retention ($\ln k$) with a change in mobile phase composition (φ). Such a shallow gradient enhances the influence of the molecular weight on the retention of polystyrenes. Once again, this is consistent with previous results and it is also in accordance with the idea that the optimal gradient for an RPLC separation of a homologous series or a homopolymer is convex in shape [30] or uses a convex temperature gradient [31] if a separation based on molecular weight is desired. In our case the separation is simply achieved by using two different column volumes, which is conceptually much simpler. The lower-weight-standards are seen not to be retained on the trap column, because for these analytes the effect of temperature on

retention is much smaller. Achieving increased resolution for high-molecular-weight standards was not the objective of the cold-trap experiments, but it was an interesting side effect. The original objective was to investigate the shift in elution composition resulting from the trapping for the standards of different molecular weight (**Figure 2.4**).

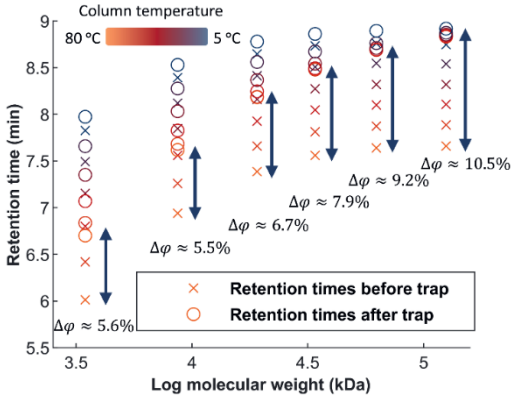

Figure 2.4: Retention time as function of molecular weight before and after the trap, including difference in composition of elution ($\Delta\boldsymbol{\varphi}$) for the largest temperature difference.

From this it can be observed that the low-molecular-weight standards are only trapped to a limited extent. The delay caused by the trap increases with increasing molecular weight, indicating that high-molecular-weight standards are trapped during at least some fraction of the 1D-LC gradient. This will be an important factor in 2D-LC, where we aim to trap analytes for a certain (modulation) time. As long as the increased elution composition that is observed in these experiments is not reached during the trapping time, one would expect that the analyte will be successfully trapped prior to injection in the second dimension. This means that the gradient rate in the ^{1}D separation and the temperature difference between the ^{1}D column and the trapping column determine the maximum modulation time and that the latter will be larger for high-molecular-weight analytes. Larger temperature differences will be required between the ^{1}D column and the trap to successfully trap analytes when using faster gradients. In our LC×LC experiments the gradient was much shallower (0.09%/min and 0.25%/min in most cases) than the one used in the 1D experiments (10%/min). Therefore, no problems with trapping were anticipated, except for the lowest-molecular-weight standards (≤ 10 kDa), which experienced limited trapping. However, for low-molecular-weight polymers other options exist, including different retention mechanisms and the use of mass-spectrometric detection [32].

2.4.3. LC×LC experiments

Several LC×LC measurements were performed to illustrate the application of the cold-trap strategy in practice. To demonstrate the performance and feasibility of the developed trapping strategy a separation of a Styrolux 693D sample was performed. Separation could be achieved within 1.5 h based on the number and length of polystyrene arms. In the schematic illustration on the right-hand side of **Figure 2.5** [33] polystyrene (PS) arms are indicated in red and polybutadiene (PB) blocks are indicated in blue. PS arms may be either long (L; 98 kDa) or short (S; 18 kDa). Up to seven PB chains can be connected using a coupling agent. The separation of this sample, using the cold-trap, is illustrated in **Figure 2.5**.

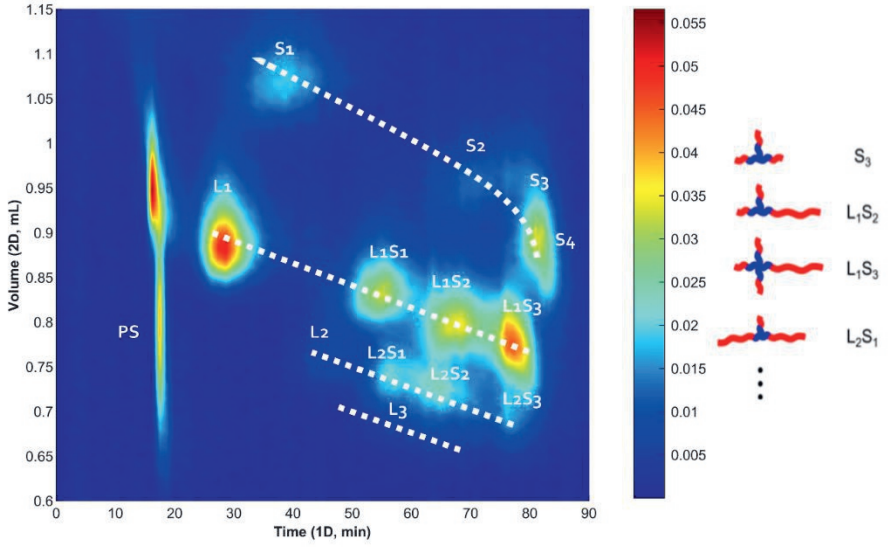

Figure 2.5: RPLC×SEC separation of Styrolux based on number and length of polystyrene arms (indicated in red in right-hand schematic). L denotes long polystyrene arms of 98 kDa, S indicates short arms of 18 kDa.

Note that the individual "peaks" or distributions were in this case assigned manually, based on the work by Chang *et al.* [33] who analysed this sample by a combination of reversed-phase temperature-gradient interaction chromatography and SEC (RP-TGIC×SEC). The separation achieved in the present work (using solvent-programmed RPLC in the first dimension instead of TGIC) is comparable, but the analysis time is

four times shorter, thanks largely to the thermal modulation. Thermal modulation allowed narrower columns to be used in the second dimension (2.1 mm i.d. as compared to 7.5 and 8 mm used in [33]). By using a volumetric flow rate that was about four times lower (0.6 mL/min instead of 2.5 mL/min) and columns that were a factor two shorter (300 *vs.* 600 mm), ^{2}D separations could be about six times faster, while reducing the amount of eluent required per analysis (^{2}D flow rate × analysis time) by a factor of about 14 and increasing the mass sensitivity (detected concentration / injected concentration) by at least a factor 14 (volume effect only; effective trapping will increase this factor further).

2.4.3.1. Investigating the effect of transfer time and flow direction

One of the critical parameters for accurate quantification is the possible loss of analyte during the trapping/loading stage or during transfer from the trap to the second dimension (*i.e.* the transfer stage). To ensure that no such losses were incurred, an ELSD was placed in the waste line, using the setup illustrated schematically in Supplementary Material (**Figure S-2**). Signals were observed at times corresponding with the moment the modulation occurs (*i.e.* when switching from the trapping stage to the transfer stage), the intensities of which corresponded with the DAD trace of the 1D-LC separation. Backflushing the trap led to much lower pulses than forward flushing (see **Figure S-3).** The exact origin of these modulation pulses is not known, but they are thought to be related to this particular set-up with a single loop and a ten-port valve. No signal was observed on the ELSD during the trapping phase. The signal between the evenly spread "modulation" peaks showed a completely flat baseline, indicating that there are no detectable losses during the trapping.

Several different (pump-frequency synchronized) flush times were investigated, namely about 4.4, 8.8 and 13.2 s. These times were selected because the period between piston strokes determined in the section above was approximately 4.4 s. Longer transfer times led to lower pulses in the ELSD signal. To determine whether any significant losses occurred we compared the resulting LC×LC chromatograms directly. These are shown in **Figure 2.6.** In **Figures 2.6-A-C** only the transfer duration is varied. Longer transfer times are seen to lead to slightly less-intense peaks, which may be explained by the analyte sent to waste during the transfer phase in the current single-trap set-up. Losses corresponding to the transfer time divided by the

cycle time are anticipated. With a constant cycle time of 79.2 s this would result in losses of about 5.5, 11, and 17% (=4.4/79.2), for the 4.4, 8.8 and 13.2 s transfer times, respectively. This is reflected in the peak intensities in the LC×LC chromatograms of **Figures 2.6-A, 2.6-B** and **2.6-C**, respectively. A comparison of **Figure 2.6-B** and **2.6-D** shows much lower peak intensities in case of forward-flushing of the trap during the transfer, which is in line with the observations in **Figure S-3**. Backflushing of the trap resulted in the smallest loss of analyte. Based on **Figure 2.6**, we selected a transfer time of 4.4 s with back-flushing of the trap to the second-dimension for further experiments.

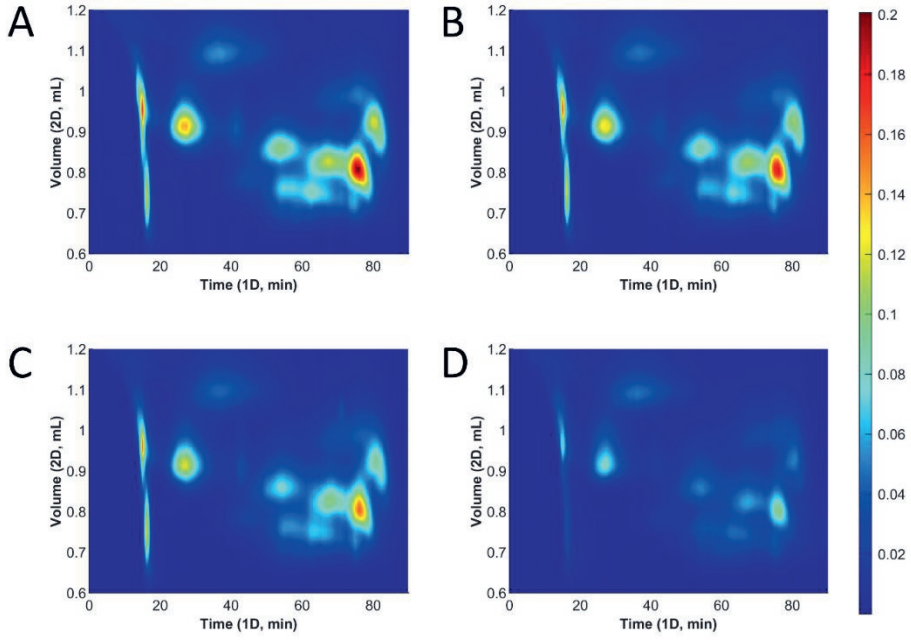

Figure 2.6: 2D-LC chromatogram obtained as a function of transfer duration, A) Duration of 4.4 s, B) Duration of 8.8 s, C) Duration of 13.2 s and D) Duration of 8.8 s with a forward's flush direction.

2.4.3.2. Investigating the effect of trap temperature on trapping efficiency

To investigate the trapping efficiency as a function of temperature, several 2D-LC measurements were performed, for both the Styrolux sample and polystyrene standards, with the cold trap set at different temperatures. The recovery of polystyrene standards with molecular weights within the range of 10 to 300 kDa was investigated, which was the separation range of the APC SEC columns.

The recoveries of two sets of polystyrene standards were measured at trap temperatures of 5, 40 and 70 °C, all at a first-dimension-column temperature of 80 °C. Quantification was performed by first correcting for the drift using asymmetrically-reweighted penalized least-squares (arPLS) [34], after which a deconvolution was performed using the modified Pearson VII distribution [35]. Finally, the peak areas were obtained using a trapezoidal approximation on the individual peaks. Chromatograms before and after baseline correction are illustrated in **Figure 2.7-A.** An example of the results of peak deconvolution is illustrated in **Figure 2.7-B**.

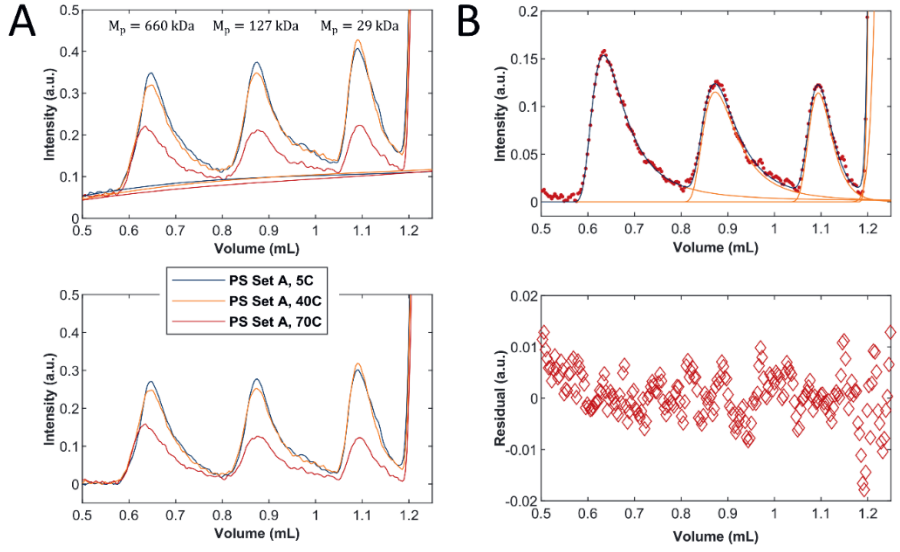

Figure 2.7: Approach for peak area determination, A) Top: background correction with arPLS; bottom: corrected chromatograms, B) Top: peak deconvolution of the different polystyrene standards; bottom: Residuals between data and peak fit.

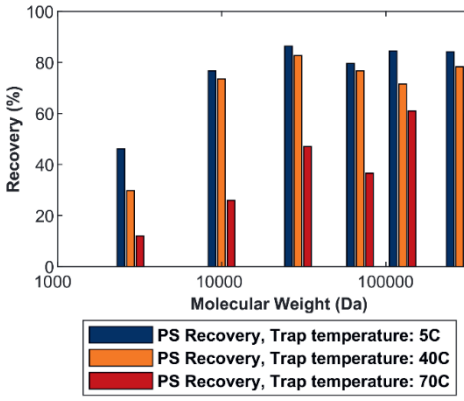

Figure 2.8: Recovery for polystyrene standards of different molecular weight at different trap temperatures. The peaks eluting at the exclusion limit of the SEC columns (molecular weights above 600 kDa) were not considered.

After determining the peak areas in this way, the recovery was determined for the different polystyrene standards. The 1D experiments (areas of eluting peaks without a trap installed) were used as reference. The results are illustrated in **Figure 2.8.**

The recovery is seen to clearly improve with an increase in molecular weight of the analytes and with a decrease in trapping temperature (*i.e.* an increase in the temperature difference between the 1D-LC column and the trap). The losses observed may be due to the single-trap configuration (anticipated loss of 5.5% in the present case) or to incomplete desorption of the analytes from the trap. Also, errors in the curve fitting and, especially, the background filtering may have resulted in lower calculated recoveries. In the case of a trapping temperature of 5 °C recoveries approached the maximum attainable value of 94.5%.

A similar procedure as described above was used to investigate the recovery for the Styrolux sample as a function of the trapping temperature. In this case curve fitting was not performed since there were few individual peaks visible, instead only the overall recovery was determined. The same trap temperatures of 5, 40 and 70 °C were used and the same first-dimension-column temperature of 80 °C. The LC×LC chromatograms and overall recoveries obtained from these experiments are shown in **Figure 2.9.**

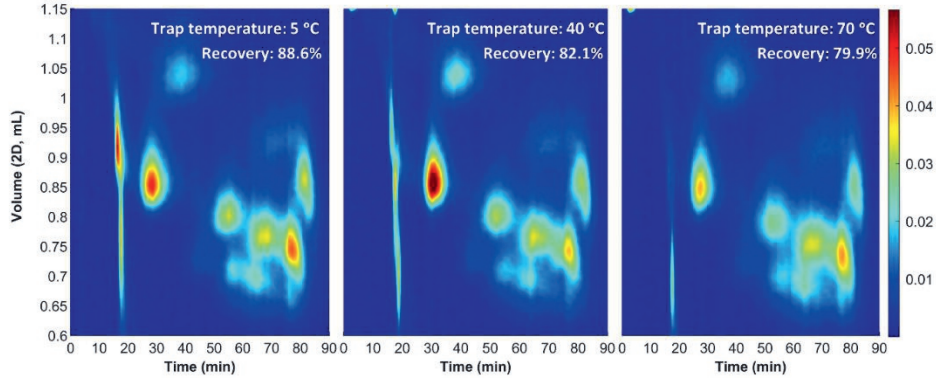

Figure 2.9: LC×LC chromatograms and calculated overall recoveries for Styrolux with different trap temperatures as indicated and a first-dimension-column temperature of 80 °C.

The peaks showing the greatest losses in recovery in the LC×LC chromatograms elute during the steepest step in the gradient used in in the ^{1}D separation (elution times 10 to 20 min). This corresponds to the results and conclusions that were already drawn from the 1D-LC experiments (section 4.2.1) and illustrates that a larger temperature difference will be required especially for low-molecular-weight analytes that are transferred to the trap in a steep 1D-LC gradient. At the same time, it is quite remarkable that even with a temperature difference of only 10 °C most of the polymer seems to be successfully retained on the trap-column. This further supports the conclusion that the combination of the typically shallow gradients used in the first dimension of LC×LC experiments and the retention characteristics of high-molecular-weight analytes creates conditions for successful thermal modulation. However, in the present paper predictions were not made regarding the conditions required to trap a polymer of a specific polarity and molecular weight. When knowing the actual gradient shape [36] and retention-temperature relationships [37–39] it should be possible to, based on only a few 1D experiments, predict whether a particular polymer or statistical copolymer can be effectively focussed using the cold-trapping method. An in-depth investigation regarding such an approach is warranted.

2.5. Conclusions

A new trapping strategy termed cold-trapping has been developed, which is applicable to all analytes that show sufficient increase in retention with decreasing temperature. This is expected to include all high-molecular-weight compounds. In the current work polystyrene and Styrolux were used to assess the applicability of the strategy. A single trap was used to assess the strategy, however, for further use in LC×LC applications two trapping columns should be utilized rather than one as the increase in duration of the "transfer" phase should result in higher recoveries. Possible limitations in terms of analyte polarity will be a subject of further study. Additionally, the pressure pulse observed during modulation was minimized. Pump-frequency synchronized modulation was demonstrated as a simple and effective means to consistently reduce the observed pressure pulses arising from valve switching, as compared to regular operation of the switching valve. This may lead to extended life time of the trapping columns, but this must be confirmed in future research. Also, the long-term repeatability and precision of thermally modulated LC×LC warrants further investigation.

Supplementary material

References

[1] G. Guiochon, The limits of the separation power of unidimensional column liquid chromatography, Journal of Chromatography A. 1126 (2006). https://doi.org/10.1016/j.chroma.2006.07.032.

[2] T. Gruendling, S. Weidner, J. Falkenhagen, C. Barner-Kowollik, Mass spectrometry in polymer chemistry: A state-of-the-art up-date, Polymer Chemistry. 1 (2010). https://doi.org/10.1039/b9py00347a.

[3] A.T. Iavarone, E.R. Williams, Supercharging in electrospray ionization: Effects on signal and charge, International Journal of Mass Spectrometry. 219 (2002). https://doi.org/10.1016/S1387-3806(02)00587-0.

[4] T. Gruendling, G. Hart-Smith, T.P. Davis, M.H. Stenzel, C. Barner-Kowollik, Enhanced ionization in electrospray ionization mass spectrometry of labile end-group-containing polystyrenes using silver(I) tetrafluoroborate as doping salt, Macromolecules. 41 (2008). https://doi.org/10.1021/ma702163v.

[5] B.W.J. Pirok, D.R. Stoll, P.J. Schoenmakers, Recent Developments in Two-Dimensional Liquid Chromatography: Fundamental Improvements for Practical Applications, Analytical Chemistry. 91 (2019). https://doi.org/10.1021/acs.analchem.8b04841.

[6] G. Groeneveld, B.W.J. Pirok, P.J. Schoenmakers, Perspectives on the future of multi-dimensional platforms, Faraday Discussions. 218 (2019). https://doi.org/10.1039/c8fd00233a.

[7] D.R. Stoll, K. Shoykhet, P. Petersson, S. Buckenmaier, Active Solvent Modulation: A Valve-Based Approach to Improve Separation Compatibility in Two-Dimensional Liquid Chromatography, Analytical Chemistry. 89 (2017). https://doi.org/10.1021/acs.analchem.7b02046.

[8] R.J. Vonk, A.F.G. Gargano, E. Davydova, H.L. Dekker, S. Eeltink, L.J. de Koning, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography with stationary-phase-assisted modulation coupled to high-resolution mass spectrometry applied to proteome analysis of saccharomyces cerevisiae, Analytical Chemistry. 87 (2015). https://doi.org/10.1021/acs.analchem.5b00708.

[9] B. Wouters, B.W.J. Pirok, D. Soulis, R.C. Garmendia Perticarini, S. Fokker, R.S. van den Hurk, M. Skolimowski, R.A.H. Peters, P.J. Schoenmakers, On-line microfluidic immobilized-enzyme reactors: A new tool for characterizing synthetic polymers, Analytica Chimica Acta. 1053 (2019). https://doi.org/10.1016/j.aca.2018.12.002.

[10] E.S. Talus, K.E. Witt, D.R. Stoll, Effect of pressure pulses at the interface valve on the stability of second dimension columns in online comprehensive two-dimensional liquid chromatography, Journal of Chromatography A. 1378 (2015). https://doi.org/10.1016/j.chroma.2014.12.019.

[11] M. Verstraeten, M. Pursch, P. Eckerle, J. Luong, G. Desmet, Thermal modulation for multidimensional liquid chromatography separations using low-thermal-mass liquid chromatography (LC), Analytical Chemistry. 83 (2011). https://doi.org/10.1021/ac201207t.

[12] S.R. Groskreutz, S.G. Weber, Temperature-assisted on-column solute focusing: A general method to reduce pre-column dispersion in capillary high performance liquid chromatography, Journal of Chromatography A. 1354 (2014). https://doi.org/10.1016/j.chroma.2014.05.056.

[13] S.R. Groskreutz, A.R. Horner, S.G. Weber, Temperature-based on-column solute focusing in capillary liquid chromatography reduces peak broadening from pre-column dispersion and volume overload when used alone or with solvent-based focusing, Journal of Chromatography A. 1405 (2015). https://doi.org/10.1016/j.chroma.2015.05.071.

[14] S.R. Groskreutz, S.G. Weber, Temperature-assisted solute focusing with sequential trap/release zones in isocratic and gradient capillary liquid chromatography: Simulation and experiment, Journal of Chromatography A. 1474 (2016). https://doi.org/10.1016/j.chroma.2016.10.062.

[15] R.E. Wilson, S.R. Groskreutz, S.G. Weber, Improving the Sensitivity, Resolution, and Peak Capacity of Gradient Elution in Capillary Liquid Chromatography with Large-Volume Injections by Using Temperature-Assisted On-Column Solute Focusing, Analytical Chemistry. 88 (2016). https://doi.org/10.1021/acs.analchem.5b04793.

[16] S.R. Groskreutz, A.R. Horner, S.G. Weber, Development of a 1.0 mm inside diameter temperature-assisted focusing precolumn for use with 2.1 mm inside diameter columns, Journal of Chromatography A. 1523 (2017). https://doi.org/10.1016/j.chroma.2017.07.015.

[17] M.T. Rerick, S.R. Groskreutz, S.G. Weber, Multiplicative On-Column Solute Focusing Using Spatially Dependent Temperature Programming for Capillary HPLC, Analytical Chemistry. 91 (2019). https://doi.org/10.1021/acs.analchem.8b04826.

[18] H.C. van de Ven, A.F.G. Gargano, S.J. van der Wal, P.J. Schoenmakers, Switching solvent and enhancing analyte concentrations in small effluent fractions using in-column focusing, Journal of Chromatography A. 1427 (2016). https://doi.org/10.1016/j.chroma.2015.11.082.

[19] B.W.J. Pirok, P. Breuer, S.J.M. Hoppe, M. Chitty, E. Welch, T. Farkas, S. van der Wal, R. Peters, P.J. Schoenmakers, Size-exclusion chromatography using core-shell particles, Journal of Chromatography A. 1486 (2017). https://doi.org/10.1016/j.chroma.2016.12.015.

[20] L.M. Blumberg, Theory of gradient elution liquid chromatography with linear solvent strength: Part 1. migration and elution parameters of a solute band, Chromatographia. 77 (2014). https://doi.org/10.1007/s10337-013-2555-y.

[21] M.J. den Uijl, P.J. Schoenmakers, B.W.J. Pirok, M.R. Bommel, Recent applications of retention modelling in liquid chromatography, Journal of Separation Science. (2020). https://doi.org/10.1002/jssc.202000905.

[22] F. Fitzpatrick, R. Edam, P. Schoenmakers, Application of the reversed-phase liquid chromatographic model to describe the retention behaviour of polydisperse macromolecules in gradient and isocratic liquid chromatography, Journal of Chromatography A. 988 (2003). https://doi.org/10.1016/S0021-9673(02)02050-2.

[23] T. Greibrokk, T. Andersen, Temperature programming in liquid chromatography, Journal of Separation Science. 24 (2001). https://doi.org/10.1002/1615-9314(20011201)24:12<899::AID-JSSC899>3.0.CO;2-K.

[24] J. Ryu, T. Chang, Thermodynamic prediction of polymer retention in temperature-programmed HPLC, Analytical Chemistry. 77 (2005). https://doi.org/10.1021/ac0507486.

[25] F. Fitzpatrick, B. Staal, P. Schoenmakers, Molar mass distributions by gradient liquid chromatography: Predicting and tailoring selectivity, Journal of Chromatography A. 1065 (2005) 219–229. https://doi.org/10.1016/j.chroma.2004.12.030.

[26] G. Guiochon, M.F. Gonnord, M. Zakaria, L.A. Beaver, A.M. Siouffi, Chromatography with a two-dimensional column, Chromatographia. 17 (1983). https://doi.org/10.1007/BF02271033.

[27] B. Wouters, J. de Vos, G. Desmet, H. Terryn, P.J. Schoenmakers, S. Eeltink, Design of a microfluidic device for comprehensive spatial two-dimensional liquid chromatography, Journal of Separation Science. 38 (2015). https://doi.org/10.1002/jssc.201401192.

[28] T. Adamopoulou, S. Nawada, S. Deridder, B. Wouters, G. Desmet, P.J. Schoenmakers, Experimental and numerical study of band-broadening effects associated with analyte transfer in microfluidic devices for spatial two-dimensional liquid chromatography created by additive manufacturing, Journal of Chromatography A. 1598 (2019). https://doi.org/10.1016/j.chroma.2019.03.041.

[29] A.M. Striegel, Method development in interaction polymer chromatography, TrAC - Trends in Analytical Chemistry. 130 (2020). https://doi.org/10.1016/j.trac.2020.115990.

[30] P.J. Shoenmakers, H.A.H. Billiet, L. de Galan, Influence of organic modifiers on the rentention behaviour in reversed-phase liquid chromatography and its consequences for gradient elution, Journal of Chromatography A. 185 (1979). https://doi.org/10.1016/S0021-9673(00)85604-6.

[31] J. Ryu, S. Park, T. Chang, Rapid molecular weight analysis of polymers by temperature gradient interaction chromatography, Journal of Chromatography A. 1075 (2005). https://doi.org/10.1016/j.chroma.2005.04.012.

[32] G. Groeneveld, M.N. Dunkle, M. Rinken, A.F.G. Gargano, A. de Niet, M. Pursch, E.P.C. Mes, P.J. Schoenmakers, Characterization of complex polyether polyols using comprehensive two-dimensional liquid chromatography hyphenated to high-resolution mass spectrometry, Journal of Chromatography A. 1569 (2018). https://doi.org/10.1016/j.chroma.2018.07.054.

[33] S. Lee, H. Choi, T. Chang, B. Staal, Two-Dimensional Liquid Chromatography Analysis of Polystyrene/Polybutadiene Block Copolymers, Analytical Chemistry. 90 (2018). https://doi.org/10.1021/acs.analchem.8b00913.

[34] S.-J. Baek, A. Park, Y.-J. Ahn, J. Choo, Baseline correction using asymmetrically reweighted penalized least squares smoothing, The Analyst. 140 (2015). https://doi.org/10.1039/c4an01061b.

[35] G.R. McGowan, M.A. Langhorst, Development and application of an integrated, high-speed, computerized hydrodynamic chromatograph, Journal of Colloid And Interface Science. 89 (1982). https://doi.org/10.1016/0021-9797(82)90124-2.

[36] T.S. Bos, L.E. Niezen, M.J. den Uijl, S.R.A. Molenaar, S. Lege, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Reducing the influence of geometry-induced gradient deformation in liquid chromatographic retention modelling, Journal of Chromatography A. 1635 (2021). https://doi.org/10.1016/j.chroma.2020.461714.

[37] P. Jandera, K. Krupczyńska, K. Vyňuchalová, B. Buszewski, Combined effects of mobile phase composition and temperature on the retention of homologous and polar test compounds on polydentate C8 column, Journal of Chromatography A. 1217 (2010). https://doi.org/10.1016/j.chroma.2010.07.019.

[38] P.L. Zhu, L.R. Snyder, J.W. Dolan, N.M. Djordjevic, D.W. Hill, L.C. Sander, T.J. Waeghe, Combined use of temperature and solvent strength in reversed-phase gradient elution. I. Predicting separation as a function of temperature and gradient conditions, Journal of Chromatography A. 756 (1996). https://doi.org/10.1016/S0021-9673(96)00721-2.

[39] P.L. Zhu, J.W. Dolan, L.R. Snyder, Combined use of temperature and solvent strength in reversed-phase gradient elution. II. Comparing
selectivity for different samples and systems, Journal of Chromatography A. 756 (1996). https://doi.org/10.1016/S0021-9673(96)00722-4.

Chapter 3

LC↻LC for the analysis of chemical-composition distributions of polymers

Abstract

Synthetic polymers typically show dispersity in molecular weight and potentially in chemical composition. For the analysis of the chemical-composition distribution (CCD) gradient liquid chromatography may be used. The CCD obtained using this method is often convoluted with an underlying molecular-weight distribution (MWD). In this paper we demonstrate that the influence of the MWD can be reduced using very steep gradients and that such gradients are best realized utilizing recycling gradient liquid chromatography (LC↻LC). This method allows for a more-accurate determination of the CCD and the assessment of (approximate) critical conditions (if these exist), even when high-molecular-weight standards of narrow dispersity are not readily available. The performance and usefulness of the approach is demonstrated for several polystyrene standards, and for the separation of statistical copolymers consisting of styrene/methyl methacrylate and methyl methacrylate/butyl methacrylate. For the latter case, approximate critical compositions of the copolymers were calculated from the critical compositions of two homopolymers and one copolymer of known chemical composition, allowing for a determination of the CCD of unknown samples. Using this approach it is shown that the copolymers elute significantly closer to the predicted critical compositions after recycling of the gradient. This is most clear for the lowest-molecular-weight copolymer (M_w = 4.2 kDa), for which the difference between measured and predicted elution composition decreases from 7.9% without recycling to 1.4% after recycling.

Publication: L.E. Niezen, B.B.P. Staal, C. Lang, H.J.A. Philipsen, B.W.J. Pirok, G.W. Somsen, P.J. Schoenmakers, Recycling gradient-elution liquid chromatography for the analysis of chemical-composition distributions of polymers, J Chromatogr A. 1679 (2022) 463386. https://doi.org/10.1016/J.CHROMA.2022.463386.

3.1. Introduction

Synthetic polymers play an important role in our current society. The use and applications of these materials is widespread; examples include polyurethane foam cushions, use of aramid in optical fiber cables and jet engine enclosures, the use of polytetrafluoroethylene in low friction bearings or non-stick pans, and many more. To continue to develop new products tailored towards specific applications, the analysis of these materials and their underlying distributions is vital. For homopolymers these include distributions in size or molecular weight (MWD), degree of branching (DBD), functionality-type/end-group (FTD), or molecular architecture (MAD). For copolymers additional distributions in terms of chemical composition (CCD) and sequence or block length (BLD) exist and specific distributions, such as on degree-of-substitution and/or tacticity are important characteristics of specific types of polymers. To analyze and understand the relationship between these distributions and the resulting material properties, typically some form of liquid chromatography (LC) is utilized [1–5]. One example is size-exclusion chromatography (SEC), which is the current benchmark for the analysis of the MWD and is often coupled to various detectors to provide additional information such as on the change in average chemical composition across the molecular weight distribution [6,7] or to assess the degree of branching [8]. To determine the CCD there is not a single, generally accepted method. Gradient-elution LC methods, including reversed-phase liquid chromatography (RPLC) and normal-phase liquid chromatography (NPLC) are most common, but isocratic LC methods such as temperature-gradient interaction chromatography (TGIC) [9–11], barrier methods such as SEC-gradients (or gradient SEC, gSEC) [12,13], and thermal field-flow-fractionation (ThFFF) [14] are also used.

To properly determine the MWD or the CCD, both distributions must not simultaneously influence the separation. Typically this is not the case since the retention of a polymer increases approximately exponentially with molecular weight in the case of isocratic LC separations [15–17]. Both the MWD and CCD may be determined by using two-dimensional liquid chromatography (2D-LC) or comprehensive 2D-LC (LC×LC), which can simultaneously provide information on molar mass and chemical composition distributions if a method such as RPLC is coupled with SEC. However, in certain cases it can be desirable to have a one-

dimensional method available that can provide information on solely the CCD, as this avoids the practical complexity of 2D-LC. Currently there are no easy-to-implement methods that do so, although examples of such separations exist [18–20]. One approach which may potentially be applied for this is recycling liquid chromatography (LC↻LC). This method, which was introduced several decades ago [21,22], aims to improve column performance by artificially increasing the column length. Nowadays the method is primarily used for specific (preparative) purification purposes, but has otherwise mostly been abandoned as a result of improvements in column and system performance [23–26]. However, the combination of gradient-elution and LC↻LC may prove especially beneficial to obtain a separation less affected by the MWD. This is because it allows for a reduction of the molecular weight influence through an increase in the gradient steepness, which should reduce the influence of molar mass, by virtually increasing the column hold-up volume (V_0) without being limited by pressure or requiring an increase in column diameter.

Our objective in the present work was to investigate the applicability of gradient elution LC↻LC for achieving a separation that is dominated by the CCD, while minimizing the effect of the molecular weight. To lay the foundation for such an approach, several practical aspects of column selection first needed to be considered and the approach was tested for narrow polystyrene standards, which were considered an ideal model system. The ultimate objective was to obtain high-resolution separations of copolymers with very similar average composition and broad MWD and to clearly distinguish effects of the CCD and the MWD in the chromatogram. Challenging samples consisted of two (statistical) styrene/methyl methacrylate (S/MMA) copolymers and statistical copolymers of methyl methacrylate and butyl methacrylate (MMA/BMA). With this work we aim to explore the benefits of LC↻LC, and to establish when and how the method may be used for the analysis of synthetic (co-)polymers.

3.1.1. Theory

To reduce the influence of a polymer's molar mass in RPLC, one must have an indication of how the retention time (t_R) of a polymer is influenced by its chemical composition and molecular weight. Under isocratic conditions the retention time increases linearly with the analyte retention factor (k), which is governed by the distribution equilibrium of the analyte between the stationary and the mobile phase.

k varies with the (volume) fraction of strong solvent in the mobile phase (φ). When the solubility of the analyte polymer in the mobile phase is not a limiting factor, one of four situations can occur, namely *i*) the polymer elutes in order of high to low molecular weight before the void volume of the column without experiencing any interaction with the stationary phase, and thus eluting primarily based on its hydrodynamic volume (i.e. size exclusion chromatography (SEC)); *ii*) the polymer elutes in order of low to high molecular weight at a volume larger than the void volume of the column, due to differential adsorption on (or partitioning into) the stationary phase (i.e. liquid adsorption chromatography (LAC)); *iii*) the polymer elutes without a significant molecular-weight dependence, often attributed to a balance between enthalpic adsorption and entropic exclusion (but more accurately solely the balance between enthalpy and entropy) and termed liquid chromatography at critical conditions (LCCC) [27–29]; *iv*) the polymer does not elute at all. For a homopolymer subjected to LAC the retention factor (k) increases approximately exponentially with molar mass, so that Case *ii* can easily turn into Case *iv.* To avoid this, gradient-elution is generally preferred for the LAC analysis of high-molecular-weight analytes. In case of a gradient, φ increases with time, which typically (if the initial k is sufficiently large) leads to a decrease in k with time [15–17,30–33]. When the initial mobile-phase composition is chosen such that k is large ($k_{\text{init}} > 10$) for all analytes and the injection solvent is not significantly stronger than the starting eluent [34], sample focusing will occur at the top of the column. As the gradient progresses, k decreases and the analyte's velocity will increase as it is caught up by the gradient, until it leaves the column. At the time of leaving the column the local retention factor of the analyte has become (much) smaller compared to the starting conditions. This is the main reason why peaks in gradient-elution chromatograms are much narrower than well-retained peaks in isocratic LC. In addition, peaks may be compressed thanks to the gradient, which causes the rear of the peak to travel faster than the front [35–37].

However, retention in LAC is also strongly affected by analyte molecular weight. This causes broad and typically fronting peaks for polymers with a broad MWD. The ultimate elution pattern of the polymer depends on the actual gradient program and on the MWD. To understand the influence of the MWD during gradient elution, it must be known how the distribution of (local) retention factors vary with the (local) mobile-phase composition. With this knowledge one can describe the elution

behaviour of the polymer distribution in a similar way as for small molecules by solving the differential gradient equation [15–17,28,30–33,38–42]. Many different models have been proposed to describe the variation of the retention factor with mobile-phase composition [43]. Examples include models that are generally used for small molecules, such as the log-linear model, commonly referred to as the linear-solvent strength (LSS) model [16,17,44], quadratic model (QM) [40] and the Neue-Kuss [45] model, but also polymer-specific models that aim to incorporate entropic exclusion effects [28,39]. As has previously been shown by multiple authors [16,17,39], simpler models such as the LSS model can often adequately describe the retention of a polymer in gradient-LC, most likely as a result of the typically (very) small range in φ across which high-molecular-weight analytes elute with reasonable retention factors (*e.g.* $1 < k < 10$). When using the log-linear (LSS) model it is assumed that the logarithm of the retention factor varies linearly with mobile-phase composition,

$$\ln k = \ln k_0 - S\varphi \tag{3.1}$$

in which k_0 is the retention factor extrapolated to $\varphi = 0$ and S is a parameter that captures the change in retention with mobile phase composition. Assuming a linear gradient and taking the above approach to determine the dependence of t_R on φ' (with $\varphi' = \frac{d\varphi}{dt}$), one may define the intrinsic gradient steepness (b, defined as the rate of change in k during the gradient per volume of mobile phase passing through the column for a specific analyte). According to the linear-solvent-strength (LSS) concept of Snyder [44] b is defined as

$$b = -\frac{d(\ln k)}{d\varphi}\frac{d\varphi}{dt}t_0 = S\Delta\varphi\frac{V_0}{V_G} = S\Delta\varphi\frac{t_0}{t_G} = S\Delta\varphi\frac{V_0}{t_G F} \tag{3.2}$$

where V_0 and t_0 are the column hold-up/void volume and time, respectively, $\Delta\varphi$ is the composition range spanned by the gradient, F is the volumetric flowrate, and t_G and V_G are the duration and the volume of the gradient, respectively. Time and volume are related by the flow rate, *i.e.,* $t_0 = V_0/F$ and $V_G = t_G F$. Therefore, b does not vary with F at constant V_G, but does vary with F at constant t_G. In **Equation 3.2** S depends on the molecular weight and the chemical composition of the analyte. It has been shown that S increases with molecular weight for a homologues series [15] and, hence, for polymers of similar structure/composition.

From isocratic experiments performed on narrow polymer standards it is known that at some particular φ (the so-called "critical composition", φ_{crit}) the influence of the molecular weight may vanish. At this mobile-phase composition the retention factor k is identical for all members of a homopolymeric series, irrespective of molecular weight [27–29]. Unless specific interactions occur, for example with end groups, the value of k at this critical composition tends to be very small, resulting in elution close to t_0. Performing an isocratic separation at this composition can give insights in end-group and block-length distributions. However, isocratic separations at the critical conditions are difficult to perform and virtually impossible for separations of (high molecular weight) copolymers, because φ_{crit} strongly depends on the composition of the copolymer. For statistical copolymers without strongly adsorbing end groups k varies due to chemical composition and molecular weight. For high-molecular-weight molecules S is very large, so that analyte molecules do not migrate at φ values below the critical composition (*i.e.* weaker solvents). In case of gradient elution, large analytes are completely retained on the column until the critical composition is reached. If an analyte molecule falls behind, it will catch up due to SEC effects; if it were to run ahead, it would immediately stop migrating, because of the weaker solvent composition. Hence, all high-molecular-weight components of a series tend to be focussed at the critical composition.

The LSS model yields a simple approximation for the retention factor at the moment of elution (k_e),

$$k_e = \frac{k_0}{bk_0+1} \tag{3.3}$$

which for very large values of k_0, and not extremely shallow gradients, simplifies to $k_e = \frac{1}{b}$. Because S values are large for high-molecular-weight analytes, b values are also large (**Equation 3.2**) and each analyte has a similarly small retention factor at the point of elution (k_e). In contrast, the low-molecular-weight (oligomeric) members have much smaller S values and larger values of $\Delta\varphi\frac{V_0}{V_G}$ (*i.e.* steeper gradients) are needed to minimize the effect of molecular weight on the elution composition (and, thus, on the elution time). For steep gradients (large values of b) the elution time depends solely on the chemical composition of the analyte and the selectivity depends primarily on $\Delta\varphi_{\mathrm{crit}}$. All copolymers created from monomers I and II are expected to elute between the respective critical compositions of the two

homopolymers, *i.e.* $\varphi_{\mathrm{crit},I}$ to $\varphi_{\mathrm{crit},II}$. The highest chemical selectivity for copolymers with a narrow chemical-composition distribution is obtained with steep gradients that span a narrow range in mobile phase composition ($\Delta\varphi$) around the critical point of the copolymer $\varphi_{\mathrm{crit},CP}$. To compensate for the narrow range (small $\Delta\varphi$), $\frac{V_0}{V_G}$ must be made high, either by reducing the gradient volume (*e.g.* by reducing the flow rate, while keeping t_G constant, or by shortening t_G), or by increasing the column volume (V_0). Reducing the flow rate whilst keeping t_G constant implies a reduction of the linear velocity, and an increase in analysis time. A lower gradient volume also increases the risk of system-induced gradient deformation, depending on the ratio of the gradient volume to the system's dwell volume ($\frac{V_G}{V_{\mathrm{dwell}}}$) [46,47]. It is generally recommended that this ratio ($\frac{V_G}{V_{\mathrm{dwell}}}$) should remain around or above unity. Reducing t_G would reduce the analysis time, but would lead to a decrease in peak capacity. An increase in column length to increase V_0 would cause an increase in the plate number and the peak capacity, but is limited by restrictions on the pressure and the analysis time. The above discussion suggests that it would be highly attractive to achieve the required high (effective) gradient steepness by increasing V_0 through lengthening the column, without increasing the pressure drop. This is exactly what can be achieved by repeatedly recycling the gradient.

3.1.2. Summary of potential advantages and disadvantages

In the present work such an LC↻LC setup is realized by using a single ten-port valve, which allows for the initially created gradient to be alternated between two columns, increasing the gradient steepness by virtually increasing the column length. LC↻LC seems to be an effective method to achieve very small k_e values for analytes of divergent molecular weights, while potentially maintaining a high selectivity with regard to the chemical composition. Furthermore, in LC↻LC the flow rate does not have to be reduced, since the increase in (effective) column length does not result in an increase in pressure. Maintaining a high flow rate reduces system-induced deformation of a low-volume gradient caused by the mixer and avoids an increase in the dwell time [46,47]. LC↻LC is, therefore, expected to be considerably faster than a non-recycling approach where a low flow rate must be used. However, LC↻LC is possibly not without disadvantages. Column-induced gradient deformation caused by adsorption or absorption of mobile-phase components ("solvent de-mixing") may

play a larger role [48,49], as may a possible build-up of impurities (depending on their retention characteristics). LC↻LC requires fast column equilibration. This is not expected to be a problem for RPLC, but it may be for other methods, such as hydrophilic-interaction liquid chromatography (HILIC) and ion-exchange chromatography (IEC). To remedy this, a larger initial ratio of $\frac{V_0}{V_G}$, so that the gradient fills a smaller % of the column and allows for longer equilibration of the stationary phase, would be required. Finally, because very small values of k_e are reached at the moment of elution, extra-column band broadening may become more significant.

3.2. Experimental

Two different systems (A and B), in two different laboratories (referred to below as laboratory A and laboratory B), were used for different parts of this work for comparison and to demonstrate the transferability of the method. In case the utilized system is not indicated, system A was used.

3.2.1. Laboratory A

3.2.1.1. Equipment and software

System A, located in Germany, consisted of an Acquity Quaternary Solvent Manager, an Acquity Column Heater, an Acquity PDA Detector, equipped with a pressure-resistant UV cell (up to 413 bar), and an Acquity Sample Manager with flow-through needle (FTN), all purchased from Waters (Milford, MA, USA). System control and data acquisition was performed using WinGPC software purchased from PSS Polymer Standards Service (Mainz, Germany).

3.2.1.2. Chemicals and Materials

Acetonitrile (ACN, ≥99.9%, LC-MS Grade) was purchased from Honeywell Research Chemicals (Seelze, Germany) and tetrahydrofuran (THF, 99.9%, Isocratic grade, unstabilized) from Bernd Kraft (Oberhausen, Germany). Narrow polystyrene standards were obtained from Polymer Standards Service.

3.2.2. Laboratory B

3.2.2.1. Equipment and software

System B, located in The Netherlands, included a (G1322A) 1100 degasser, (G1311A) 1100 quaternary pump, an (G1329A) 1100 auto-sampler, and an (G1316A) 1100 column oven, all purchased from Agilent (Waldbronn, Germany). An LC-10 AVvp UV detector, equipped with a pressure-resistant UV cell (up to 80 bar) was purchased from Shimadzu (Kyoto, Japan).

System control was performed using Agilent ChemStation. Data acquisition was performed using Shimadzu LabSolutions software.

3.2.2.2. Chemicals and Materials

THF and non-stabilized THF (99.9%, LC-MS Grade, unstabilized) were obtained from VWR Chemicals (Darmstadt, Germany), ACN (≥99.9%, LC-MS Grade) and methanol (MeOH, 99.9%, LC-MS Grade) were obtained from Biosolve (Valkenswaard, the Netherlands). 2,2'-Azodi(2-MethylButyroNitrile) (AMBN, 98%) and Methyl-methacrylate monomers (MMA, 99%) were obtained from Sigma Aldrich (Steinheim, Germany). Styrene monomers (ST, 99%) was obtained from Fluka (Seelze, Germany). 1-Butanon (MEK, 99%) was obtained from Acros (Geel, Belgium). All water was purified in-house using a Satorius Arium 611VF at a resistivity of 18.2 MΩ·cm obtained from Sartorius (Göttingen, Germany). A polystyrene (PS) standards kit was obtained from Polymer Standards Service.

3.2.3. Material and methods common to Laboratory A and B

Certain equipment and chemicals, as well as procedures, were transferred and therefore identical in both laboratories. These are included in this section.

3.2.3.1. Equipment and procedure

For the recycling experiments two sets of two 250 × 4.6 mm Nucleosil columns (C_{18} and bare silica), both containing 5-µm particles with a pore size of 4000 Å were obtained from Macherey Nagel (Düren, Germany). Two 250 × 4.6 mm C_{18} columns containing 5-µm particles with a pore size of 100 Å were obtained from YMC (Kyoto, Japan). Additionally, two 250 × 4.6 mm Imtakt Presto FF-C_{18} columns from Imtakt (Kyoto, Japan), containing non-porous 2-µm particles, were also evaluated.

For the SEC experiments three 150 × 4.6 mm Acquity APC XT columns containing 1.7-µm particles with a pore size of 45 Å were used. Non-stabilized THF was used as eluent.

A 10-port 2-position UHPLC valve (MXT715-102) was purchased from Rheodyne, IDEX Corporation (Lake Forest, IL, USA). An Arduino Uno Rev 3 was purchased from a local electronics supplier and was used to control the timing of the 10-port valve, irrespective of the system used.

In all cases the approximate cycle timing was determined from a blank THF injection and a 0-100% gradient of THF in ACN was run to determine the dwell volume. Unless otherwise mentioned, the temperature of the column oven was set to 30 °C.

3.2.3.2. Chemicals

Five (statistical) copolymer samples consisting of styrene and methyl methacrylate (S/MMA), with average compositions of: 84/16; 71/29; 57/43; 42/58; 25/75, were synthesized in-house in laboratory B using thermally-initiated free-radical polymerization. The full procedure is included in the supplementary information (section S1).

Six different (statistical) copolymer samples consisting of methyl methacrylate and butyl methacrylate (MMA/BMA) were obtained from DSM (Waalwijk, The Netherlands). A block copolymer from MMA/BMA was obtained from Polymer Standards Service GmbH.

3.2.4. Data analysis

All data analysis (*e.g.* alignment, background correction, chromatogram reshaping and peak analysis) was performed in MATLAB R2021a, purchased from Mathworks (Natick, MA, USA).

3.3. Results & Discussion

3.3.1. Design and initial experiments

3.3.1.1. Design of the LC↻LC set-up

To perform the recycling gradient experiments a ten-port valve and two identical columns were utilized. A scheme of the set-up is shown as **Figure 3.1-A.** For the experiment the gradient is only created a single time and is continuously recycled between two columns. Because it is not possible to recycle a gradient that exceeds a single column volume without losing part of the gradient to waste, the gradient volume was always kept below the void volume of one column. A pressure-resistant UV-detector was installed in-line to allow monitoring of the separation and the gradient during each cycle. **Figure 3.1-B** shows an example of the data obtained from this in-line UV detector when running of a test compound. A recurring signal is obtained that may be "folded" in a similar manner as is commonly done for modulations in LC×LC or comprehensive two-dimensional gas chromatography (GC×GC) (**Figure 3.1-C)**. The folded data can then be visualized as either a stacked plot (left) or as a surface plot (right).

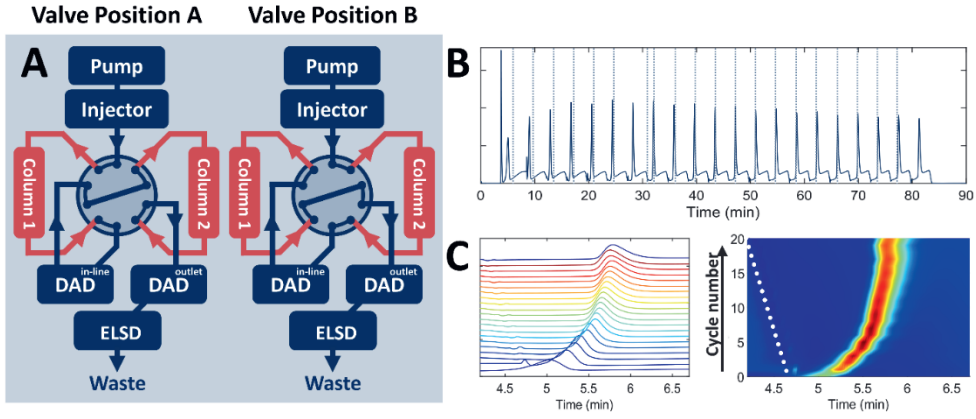

Figure 3.1: A) Schematic illustration of the recycling-gradient set-up, B) Trace from the in-line DAD resulting from the recycling gradient with the switching moments of the valve indicated by the dotted lines, C) Data folded and aligned, displayed as stacked individual cycles (left) or as a surface plot (right).

The duration of the first cycle was $\frac{(V_{0,1}+V_{0,2})+V_{\mathrm{dwell}}}{F} \approx \frac{2V_0+V_{\mathrm{dwell}}}{F}$. In the present case two columns of (nearly) equal volume were used ($V_0 \approx V_{0,1} \approx V_{0,2}$). However, in principle any combination of columns (packed with the same particles) may be used when unequal switching times are used, provided that the gradient volume remains below the smallest of the two column volumes ($V_G \leq \min\{V_{0,1},\ V_{0,2}\}$). After the first cycle, the gradient (with the analytes positioned in it) was redirected to the first column. The gradient was then alternated between columns for a number of *n* cycles with a constant recycle time of $\frac{V_0}{F}$. Folding the individual cycles (**Figure 3.1-C**) reveals a few important aspects of LC↻LC. Firstly, it is possible to track the progression of an analyte within the gradient. Secondly, it shows that selecting the correct recycle timing is critical, especially when a very large number of cycles is to be performed. When the timing of each cycle is off, the gradient and the position of the analytes are not aligned in each run. In **Figure 3.1-C** the selected cycle time was about 1.2 s too short. The dotted line in **Figure 3.1-C** corresponds to a benchmark point (signal disturbance around the moment the valve is switched) in the chromatograms from each cycle. If the correct cycle time is used such a line becomes vertical. In most cases the correct cycle timing could be accurately determined by aligning each cycle based on characteristic features in the background signal.

3.3.1.2. Experimental evaluation of gradient deformation

From previous work it is known that steep gradients come with a higher risk of strong column-induced gradient deformation [49]. To practically assess the magnitude of this effect and its consequences for LC↻LC, several initial tests were performed on a variety of columns. A reasonably large PS standard (113 kDa, PS6) was followed during a number of cycles. For all experiments the same gradient from 0-100% THF in ACN in 3 min was used. For the different columns the flowrate was adjusted so that the gradient volume remained below V_0. For the 120 and 4000 Å columns V_0 was about 3.1 mL, so a flowrate of 1 mL·min^{-1} was used. For the non-porous C_{18} columns V_0 was about 1.2 mL so a flowrate of 0.4 mL·min^{-1} was used. The results of these initial experiments are illustrated in **Figure 3.2** for several sets of columns with different stationary-phase chemistries, pore sizes, and particle sizes. The decision to recycle the entirety of the gradient ($\Delta\varphi$ = 1, $V_G = V_0$) was based on the desire to cover a wide range of possible critical compositions (φ_{crit}). This is especially relevant when little or

no information is available on the retention characteristics of the sample (*i.e.* no known information on the distributions of $\ln k_0$ and S, or on φ_{crit}). This will often be the case when analysing (co-)polymers.

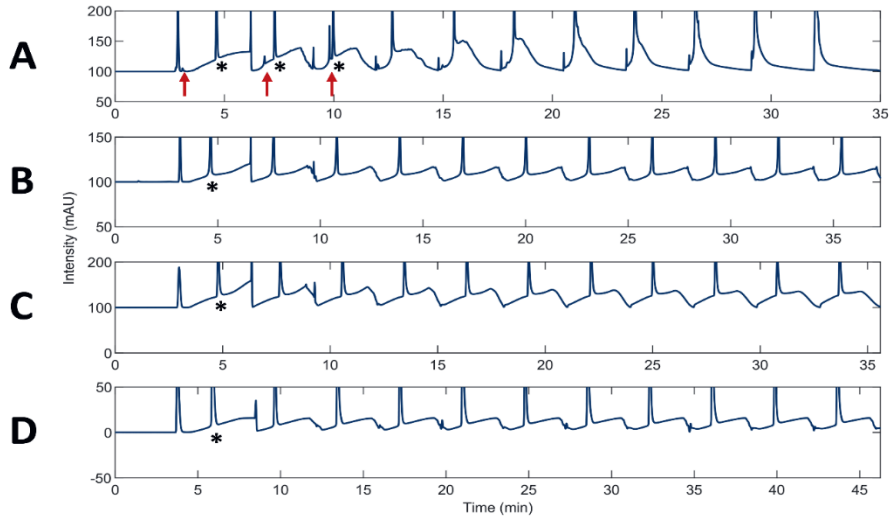

Figure 3.2: LC↻LC of PS6 (113 kDa) using recycling of a 3-min 0-100% THF in ACN gradient for a couple of A) 120 Å, 5-µm C_{18} columns, B) 4000 Å, 5-µm C_{18} columns, C) 4000 Å, 5-µm bare silica columns and D) non-porous 2-µm C_{18} columns

From **Figure 3.2** it may be concluded that the worst result was obtained for the 120 Å C_{18} columns. The shape of the background absorbance signal due to the gradient is seen to drastically change and the PS6 peak (indicated by the asterisk) in the gradient becomes eventually obscured (**Figure 3.2-A**). Apparently, the column is not sufficiently equilibrated between cycles. Also, a spurious peak appears in the first cycle, and can be more clearly seen in the second cycle (indicated by the red arrow). A convex shape of the leading part of the gradient is indicative of solvent de-mixing caused by the preferential adsorption of the more-UV-active and most non-polar solvent (THF) on the column. Due to the inadequate equilibration of the column and an apparent saturation of the stationary phase with THF, no useful results were obtained. After only three cycles the peak corresponding to PS6 completely overlaps with a "breakthrough peak" of THF. In contrast, for both the columns containing 4000 Å particles (**Figure 3.2-B** for C_{18} particles and **Figure 3.2-C** for bare-silica particles),

as well as the columns containing non-porous C_{18} particles (**Figure 3.2-D**) the traces for each cycle are much more consistent and the PS6 standard readily assumes its position around the critical composition for polystyrene in the gradient (which is expected considering its relatively large molecular weight). For all columns other than the 120 Å C_{18} columns, a gradual increase in the pressure was consistently observed during each cycle, due to an increase in the fraction of the more-viscous THF. In conclusion, successful recycling of the full gradient ($\Delta\varphi$ = 1, $V_G = V_0$) could not be achieved in columns that contained particles with small pores (120 Å), likely because the required equilibration time for these columns was much longer than for the wide pore packings [50]. However, if an application is run across a narrower range of compositions (smaller $\Delta\varphi$), small-pore particles with large available surface areas may still feasibly be used. In the present study all further experiments were conducted using the stationary phases with 4000 Å pores and the non-porous particles.

3.3.1.3. LC↻LC of PS standards on various columns

To investigate the applicability of the method for reducing the molecular-weight influence on retention, PS standards of different molecular weight were used as a model system. Peak molecular weights (M_p) and polydispersity indices (PDI, in brackets) were 4.29 kDa (1.05), 10.4 kDa (1.03), 19.6 kDa (1.03), 43.3 kDa (1.03), 70.9 kDa (1.03), and 113 kDa (1.03), respectively, henceforth referred to as PS1 through PS6. The separation obtained for these standards on the non-porous C_{18}, the 4000-Å C_{18}, and the 4000-Å bare-silica columns is illustrated in **Figure 3.3**. Examples of the non-aligned signals are included in the supplementary material (Figure S-1, section S2).

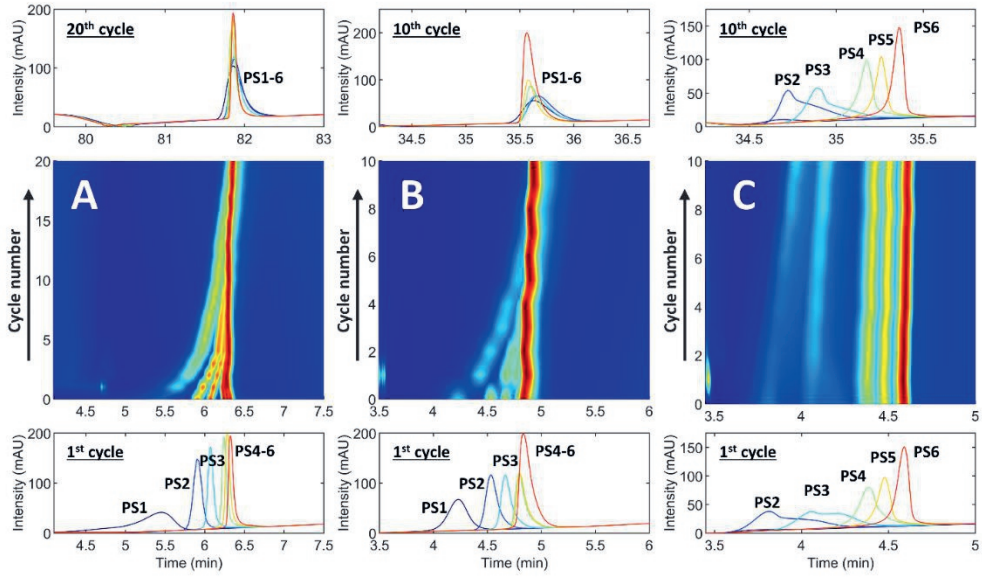

Figure 3.3: LCxLC of PS1-6. A) non-porous C_{18} columns using a 3-min gradient of 20-80% THF in ACN at a flow rate of 0.4 mL·min^{-1}; B) 4000 Å C_{18} columns using a 3-min gradient of 20-80% THF in ACN at a flow rate of 1 mL·min^{-1}; C) 4000 Å bare-silica columns using a 3-min gradient of 0-100% THF in ACN at a flow rate of 1 mL·min^{-1}. The first-cycle chromatograms are shown in the bottom panel; the last (20th or 10th) cycle chromatograms are shown in the top panel. The central panel displays the surface plots for all cycles.

These experiments confirm that the influence of the molecular weight is progressively reduced with an increasing number of cycles in case of the C_{18} columns (for both the non-porous particles, **Figure 3.3-A**, and the 4000 Å particles, **Figure 3.3-B**). The mitigation of the molecular-weight effect concurs with an increase in the effective gradient steepness (b). On the non-porous columns (**Figure 3.3-A**), the difference in elution composition between PS1 (4.29 kDa) and PS6 (113 kDa) is reduced from $\Delta\varphi = 17\%$ (first cycle, *i.e.* no recycling) to $\Delta\varphi < 0.1\%$ (20 cycles). Evidently, when the gradient steepness is sufficiently large, the elution order becomes essentially independent of molecular weight. A comparison of **Figure 3.3-A** and **Figure 3.3-B** also demonstrates that, in case of gradient elution, the presence of pores does not determine whether a (pseudo) critical composition exists. For the bare-silica columns (**Figure 3.3-C**), only a marginal reduction in the molecular-weight influence was observed, which indicates the absence of critical conditions on these columns and with this combination of solvents. The separation obtained using

the bare-silica columns (**Figure 3.3-C**) is nearly independent of the effective column length and there is little or no variation in the retention factor at the moment of elution (k_e) with b. This demonstrates that LC↻LC may, within one experiment, also provide information on the underlying elution behaviour, as the minor influence of an increase in column length indicates that elution is governed more so by solubility (ACN to THF corresponding to a non-solvent to solvent gradient) than by interaction with the column. This results in another potential practical application of LC↻LC, namely the ability to determine approximate critical conditions when narrow standards are not available, as is very often the case (*e.g.* for copolymers).

For all analytes the changes in peak width and shape as a function of cycle number were assessed for both the non-porous and 4000-Å C_{18} packings (**Figure 3.4**).

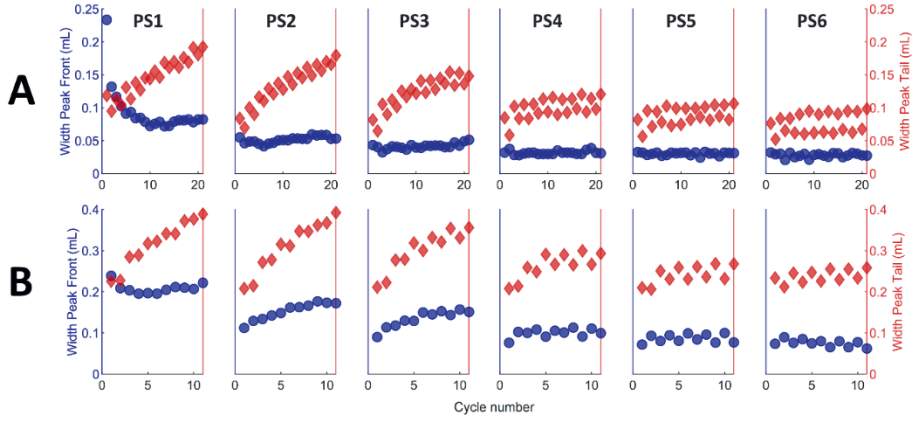

Figure 3.4: Front and tail peak widths (in mL) obtained during LC↻LC of PS1-6; widths are measured to the peak center line at 10% of the maximum peak height, and depicted as function of cycle number. Blue circles: front peak widths; red diamonds: tail peak widths. Gradient: 3-min 20-80% THF in ACN. A) non-porous C_{18} particles; flow rate, 0.4 mL·min^{-1}; B) 4000 Å C_{18} particles; flow rate, 1 mL·min^{-1}.

The obtained peak-width parameters on the columns packed with non-porous particles was, in most cases, a factor two to three smaller than those obtained for the 4000 Å C_{18} columns, likely thanks to faster mass-transfer in these columns, because of the smaller particle size (2-µm *vs.* 5-µm) and the absence of pores. Additionally, irrespective of the column used, the shape of the peak depends on the molecular weight of the analyte and small differences can be observed in the peak widths

between successive cycles ("zig-zag" effect). Apparently, the chromatogram depends slightly on which of the two columns the gradient has passed through before entering the in-line DAD. This may be explained by differences in the packing, the stationary phase itself, or small differences in the pressure for the two columns. The latter effect is a less likely explanation, because LC↻LC requires only moderate pressures. An eventual pressure effect may be expected to be more pronounced for high-molecular-weight analytes, which from previous studies are known to experience relatively large changes in partial molar volume with a change in pressure compared to small analytes [51–53], which cannot be discerned from **Figure 3.4**. Concerning the shape of the peak, two processes can be observed. Firstly, the peak fronting decreased significantly with cycle number, most noticeably for the low-molecular-weight analytes and marginally for PS5 and PS6. Secondly, the peak tailing increased with cycle number, again less strongly for the high-molecular-weight standards. The first process is likely a result of the selectivity with respect to molecular weight, which is much larger for PS1 than for PS6, as a result of the much shallower effective gradient that this standard experiences (*i.e.* lower value of b, because of smaller S values). The second process may be a result of either chromatographic peak broadening or an inversion of the molecular weight dependence around the "pseudo" critical composition. Using gradient elution the peak width (in volume units, σ_V) may be described using **Equation 3.4**:

$$\sigma_V = G\frac{V_0}{\sqrt{N}}(1 + k_e) \tag{3.4}$$

In which G is a band compression factor, which for very steep gradients (large b) and an unretained mobile-phase modifier should reach a (supposedly limiting) value of about 0.58 [36,37]. Because in our case large b values can likely be reached and the resulting k_e values are small (and likely similar) for all analytes, the peak width after a given number of cycles should depend primarily on N and V_0. When such conditions are reached σ_V is expected to increase with the square root of the number of cycles. Given the small k_e values, extra-column band broadening is also a point of concern.

In this work the peak broadening seemed to manifest itself primarily in the form of peak tailing, rather than as an increase in overall peak width. This effect was largest for PS1. To investigate this effect, an LC↻LC analysis of PS1 on the non-porous

column was ended after the 10th cycle. Fractions of the effluent were collected and subsequently measured with SEC. The results of these experiments, as performed on the non-porous-particle C_{18} columns, are illustrated in **Figure 3.5**.

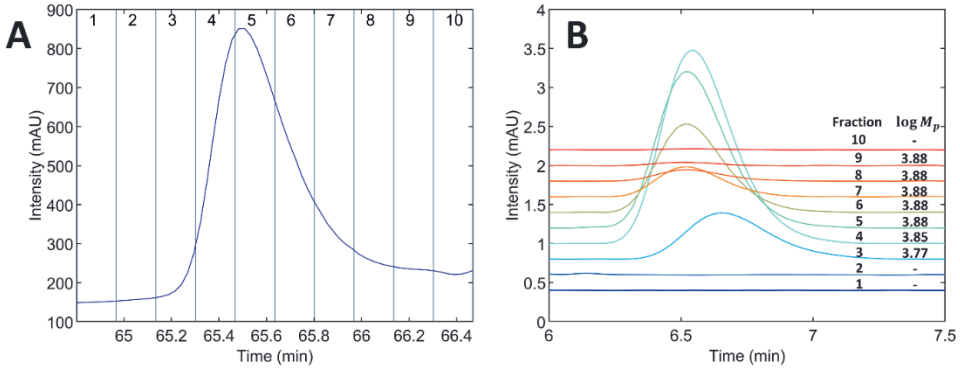

Figure 3.5: A) Fractionation of PS1 after analysis by LC↻LC (10 cycles) using non-porous C_{18} particles with a 3-min 20-80% THF gradient in ACN at a flowrate of 0.4 mL·min^{-1}; fraction numbers are indicated. B) SEC chromatograms of the fractions indicated in A, measured using Acquity APC XT columns, with unstabilized THF at a flowrate of 0.5 mL·min^{-1} and a column oven temperature of 60°C.

Small differences in elution time (and thus molecular weight) are found to remain after 10 cycles, especially for fractions 3 and 4 ($\Delta M_p \approx 1.1$ kDa). Additionally, the average M_p (as determined by calibration relative to a different set of PS standards) differed slightly from the listed value. Irrespective of these differences, all later fractions showed nearly consistent peak molecular weights. This confirms that the observed peak tailing is a result of chromatographic and extra-column dispersion, rather than selectivity. Chromatographic peak broadening occurs predominantly at the trailing edge of the peak. This can be explained by the fact that, after the molecular-weight effect on retention is fully diminished (no remaining selectivity as observed in **Figure 3.5**), a peak-sharpening effect due to the gradient likely prevails at the front of the peaks. Molecules that run ahead of the peak (and thus the gradient) will slow down due to the increase in weak solvent and get back in line. Such gradient-sharpening is absent at the back side of the peaks, where all k values are low. Such an explanation is in agreement with the observation that the broadening is greatest for low-molecular-weight standards, while higher-molecular-weight standards show less broadening. Contrarily, extra-column band broadening

is expected to be more severe for high-molecular-weight standards, as a result of their much smaller diffusion coefficients. However, SEC or hydrodynamic effects could help sharpen the peaks, as this would allow large molecules that have fallen behind to catch up. For the 4000 Å columns a brief assessment of the influence of flowrate and the range of mobile-phase composition covered by the gradient ($\Delta\varphi$) on peak width was performed across 10 cycles for a narrow and broad PS standard. The results of these experiments are included in the supplementary material (Figure S-2, section S3) and indicated that broad and narrow standards reach nearly equal peak width at high number of cycles for the same gradient. Gradients spanning smaller $\Delta\varphi$ and higher flow rates generally resulted in broader peaks.

3.3.2. LC↻LC for the analysis of chemical-composition distributions

3.3.2.1. Separations of S/MMA copolymers

Because LC↻LC could successfully suppress the influence of the molecular weight in case of PS, it was deemed to be a good technique for determining chemical-composition distributions (CCD), without a confounding effect of molecular weight. Experiments were performed on five statistical copolymers consisting of S/MMA (SM1-5), as well as on seven MMA/BMA copolymers (MB1-7), to assess whether the approach could be applied to achieve higher resolution between samples differing only slightly in composition. For SM1-2 a gradient spanning a narrow range in composition (small $\Delta\varphi$) was used. This caused a pronounced influence of the underlying broad MWD (M_w = 54 kDa (PDI = 2.3) and 64 kDa (PDI = 2.1) for copolymer SM1 and SM2, respectively) of these samples on the elution profile obtained with conventional gradient-elution LC, as is clear from the first-cycle trace in **Figure 3.6-A** where distinctly fronting peaks are obtained.

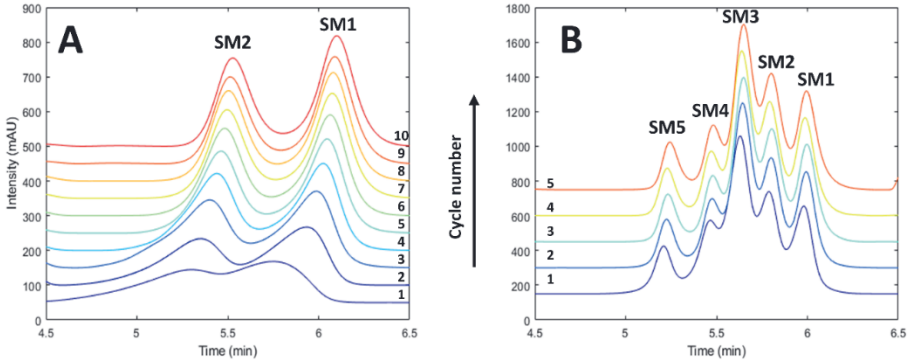

Figure 3.6: LCCLC of S/MMA copolymers SM1-2 (A) and SM1-5 (B) performed on two 4000 Å C_{18} columns using a flow rate of 1 mL·min^{-1}. Gradient, A) 30-50% THF in ACN in 2.5 min, B) 0-60% THF in ACN in 2.5 min. Average S/MMA compositions: SM1, 84/16; SM2, 71/29; SM3, 57/43; SM4, 42/58; SM5, 25/75. Experiments were performed on System B.

The underlying MWD jeopardizes the determination of the CCD when a shallow gradient is used. In subsequent cycles the effective gradient slope (b) gradually increases causing the profile to reflect the CCD, with little or no influence of the broad MWD. Much sharper peaks were obtained after ten cycles, as a result of the narrow CCD of both copolymers. The signal-to-noise ratio improved by more than a factor of three for both distributions and the their resolution improved from 0.66 to 1.5 (determined after deconvoluting the two distributions). If a broader range of polymer compositions (broad CCD) is considered (SM1-5), a gradient with a larger $\Delta\varphi$ is required (**Figure 3.6-B).** This increases the value of b and reduces the influence of the MWD for all copolymers, even in the first cycle. Because the difference in the critical compositions of SM1 and SM2 ($\Delta\varphi_{\mathrm{crit}} = \varphi_{\mathrm{crit,SM2}} - \varphi_{\mathrm{crit,SM1}}$) is about 4.8%, and is independent of the slope of the gradient, a higher resolution in terms of chemical composition is obtained when the gradient covers a smaller range of eluent compositions, within the same time frame. This confirms that the retention of these copolymers follows the same basic rules as the PS homopolymers, with a strong correlation between the molecular-weight dependent slope (S) and intercept ($\ln k_0$) of **Equation 3.1**. Peaks are seen to remain broader in time units at smaller $\Delta\varphi$ even after recycling of the gradient. In terms of volume-fraction units (at the elution composition) peaks are narrower for narrow range gradients. This may be the best reflection of the actual CCD, because the chemical-composition selectivity of the

separation is maximized and overshadows the contribution of the chromatographic dispersion.

3.3.2.2. Separations of MMA/BMA copolymers

To further illustrate the effect of gradient recycling the method was also applied to a separation of MMA/BMA copolymers (MB1-7), using both the columns containing non-porous and 4000 Å C_{18} particles (**Figure 3.7**).

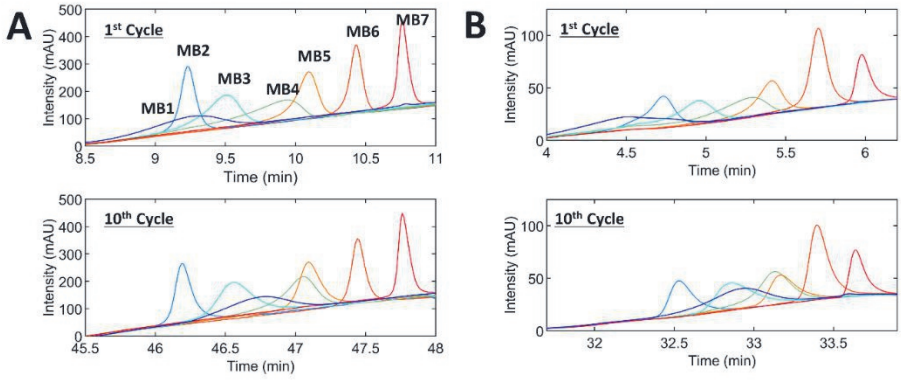

Figure 3.7: LC↻LC of MMA/BMA copolymers MB1-7 performed on A) non-porous C_{18} particles using a gradient of 0-60% THF in ACN in 3 min at a flowrate of 0.4 mL·min^{-1}, and B) 4000-Å C_{18} particles using a gradient of 0-60% THF in ACN in 2.5 min at a flowrate of 1 mL·min^{-1}. Average MMA/BMA compositions (as determined by ^{1}H-NMR) and $\boldsymbol{M_w}$: MB1, 50/50 (4.2 kDa); MB2, 76/24 (80 kDa); MB3, 58/42 (20 kDa); MB4, 32/68 (15 kDa); MB5, 30/70 (50 kDa); MB6, 85/15 (100 kDa); MB7, 0/100 (160 kDa).

In this case a broader range of composition ($\Delta\varphi$) was used. Again we observed that the separation with respect to polymer composition, once obtained, can be maintained in subsequent cycles. Unlike the above example of the S/MMA copolymers, most peaks show the characteristic fronting due to the confounding MWD in the first cycle (upper panels in **Figure 3.7**). The fronting is reduced or disappears for many peaks with an increasing number of cycles, as the effect of the MWD is increasingly suppressed. An additional method to illustrate the effect of the recycling is to predict the approximate critical compositions of the copolymers and comparing these with the obtained elution compositions before and after a recycling of the gradient. Previous work has shown that the approximate critical composition

of a statistical copolymer, which consists of two types of monomers (I and II), can be calculated using data obtained for the corresponding homopolymers [16], by using **Equation 3.5**

$$\varphi_{\mathrm{crit},CP} = \frac{p_I(1-X_{II})+p_{II}X_{II}}{q_I p_I(1-X_{II})+q_{II}p_{II}X_{II}} \tag{3.5}$$

in which the subscripts I and II indicate monomer type I and II, respectively, X is the mass fraction of the respective monomer in the copolymer (CP), q is the slope obtained by assuming a linear correlation between S and $\ln k_0$, and corresponds to the approximate critical composition as $\varphi_{\mathrm{crit}} = \frac{1}{q}$, p is the slope obtained by assuming a linear correlation between $\ln k_0$ and molecular weight, and $\varphi_{\mathrm{crit},CP}$ is the approximate critical composition of copolymer CP with mass fraction X_{II}. Determining p_I and p_{II} individually for both homopolymers may require multiple experiments and can be tedious. However, since $\varphi_{\mathrm{crit},CP}$ can be shown to depend on $\frac{p_I}{p_{II}}$ by dividing **Equation 3.5** by p_{II} it can be easier to rewrite **Equation 3.5** to:

$$\frac{p_I}{p_{II}} = \frac{X_{II}\left(1-\frac{\varphi_{\mathrm{crit},CP}}{\varphi_{\mathrm{crit},II}}\right)}{(1-X_{II})\left(\frac{\varphi_{\mathrm{crit},CP}}{\varphi_{\mathrm{crit},I}}-1\right)} \tag{3.6}$$

This equation allows one to determine $\frac{p_I}{p_{II}}$ provided that the approximate critical conditions are determined for two high-molecular-weight homopolymers (each consisting of monomers I and II, respectively), and one high-molecular-weight copolymer CP of known average composition, given by X_{II}. In our case recycling of the gradient promotes elution at the approximate critical composition. Therefore, it is expected that the difference between the measured elution composition (φ_e) and the predicted critical composition ($\varphi_{\mathrm{crit},CP}$) is minimized with an increase in the number of cycles (or gradient steepness), especially for the lowest-molecular-weight analytes (MB1 and MB4). The approximate critical compositions were calculated in this way using $\varphi_{\mathrm{crit,PMMA}} = 0.09$, $\varphi_{\mathrm{crit,PBMA}} = 0.47$, and $\varphi_{\mathrm{crit,MB5}} = 0.34$ (with $X_{\mathrm{BMA}} = 0.70$, as determined from ^{1}H-NMR). The differences between the measured elution compositions and the elution compositions predicted in this way (calculated as: $\left|\varphi_e - \varphi_{\mathrm{crit},CP}\right| * 100$) for MB1 and MB4 decreased from 7.9% and 2.0% in the first cycle, to 1.4% and 0.092% after the final cycle, respectively. Assuming instead that $\varphi_{\mathrm{crit},CP}$ varied linearly with X_{BMA} between $\varphi_{\mathrm{crit,PMMA}}$ and $\varphi_{\mathrm{crit,PBMA}}$ led to an

overestimation in all cases. A full overview is given in the supplementary information (Figure S-3, section S4). The largest shift in elution composition after recycling of the gradient occurred for copolymer MB1. This is not surprising, since this is a low-molecular-weight copolymer (M_w = 4.2 kDa). Additionally, because it is a block copolymer, the peak remains broad even after recycling. Block copolymers tend to have a much broader CCD than statistical copolymers, due to the block-length distributions of the two blocks. The peak of copolymer MB4 showed significant fronting, even after 10 cycles. To evaluate whether this fronting occurred due to the remaining influence of the MWD or was the result of the underlying CCD, peak fractions were taken after 1 and 20 cycles. The MWD of each fraction was subsequently determined using SEC and also the change in peak asymmetry during the recycling experiment was evaluated (**Figure 3.8**).

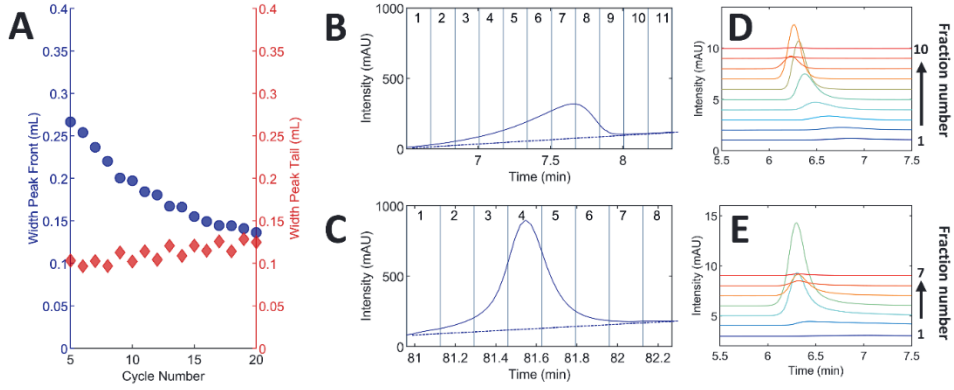

Figure 3.8: LC↻LC of copolymer MB4 using non-porous C_{18} particles with a 3-min 0-60% THF gradient in ACN at a flowrate of 0.4 mL·min^{-1}. A) Front (blue) and tail (red) peak widths (in mL) as function of cycle number (calculation, see Figure 4). B and C) Peak profiles after 1st and 20th cycle, respectively, with fractions taken indicated; dashed line under the peak indicates the background signal of the gradient. D and E) SEC chromatograms of the fractions indicated in B and C, respectively, measured using Acquity APC XT columns at a flowrate of 0.5 mL·min^{-1} and a column oven temperature of 60 ºC.

As seen in **Figure 3.8-A**, the peak fronting decreases during the cycles, until it seems to converge after 20 cycles, indicating that the confounding effect of the underlying MWD has been diminished. However, significant fronting remains, even after 20 cycles (**Figure 3.8-C**), the underlying gradient is indicated in the figure to better highlight the remaining extent of peak fronting. An analysis of the fractions taken

from the 20th cycle (**Figure 3.8-D**) shows that the underlying MWD within all fractions after the first two is the same, indicating that even for a relatively low molecular weight polymer (M_p = 15 kDa) a good reflection of the true CCD of the polymer can be obtained. This case underlines the value of LC↻LC. Without recycling there is a strong confounding effect of the MWD and the CCD, which prevents correct interpretation of the results.

3.4. Conclusion

In this work the use of LC↻LC for the analysis of the CCD of copolymers is introduced and demonstrated. The entirety of the gradient is continuously recycled to achieve extremely steep gradients, so as to minimize the effect of the MWD on the elution profile. Conventionally, very fast gradients require short durations, in combination with long columns and low flow rates, resulting in decreased peak capacities, long analysis times, and an increased risk of system-induced gradient deformation. Such issues can be avoided with LC↻LC. It is demonstrated that a set of polystyrene standards of greatly different molecular weights can be made to (nearly) completely co-elute. LC↻LC was used to determine the CCD of two sets of copolymers (S/MMA and MMA/BMA), with the confounding effect of the MWD being successfully suppressed. Based on the results presented, LC↻LC appears suitable for the accurate determination of the CCD of a wide range of copolymers with narrow or broad CCDs and MWDs. No prior information on the critical conditions is required, greatly reducing the effort required and eliminating the need for (narrow) standards.

Chromatographic dispersion remains, but gradient conditions and column dimensions may be chosen such that the chemical-composition selectivity is dominant. Columns packed with large-pore particles or non-porous particles can be used for LC↻LC, but small-pore particles give rise to column-induced gradient deformation. This was ascribed to adsorption of mobile-phase components on packing**s** with large surface areas.

An LC↻LC experiment may be ended after any number of cycles and combined with any detector suitable for gradient LC. Also, LC↻LC may be coupled on-line with other methods, such as size-exclusion chromatography, to better highlight potential differences between samples. A comprehensive coupling of LC↻LC and SEC may provide clearly interpretable results, and the orthogonality between RPLC or NPLC

and SEC will be increased. Even without addition of another method LC↻LC was shown to be capable of a more direct determination of the CCD.

Supplementary material

References

[1] A.M. Striegel, Method development in interaction polymer chromatography, TrAC - Trends in Analytical Chemistry. 130 (2020). https://doi.org/10.1016/j.trac.2020.115990.

[2] A.M. Striegel, W.W. Yau, J.J. Kirkland, D.D. Bly, Modern Size-Exclusion Liquid Chromatography: Practice of Gel Permeation and Gel Filtration Chromatography: Second Edition, 2009. https://doi.org/10.1002/9780470442876.

[3] W. Radke, Polymer separations by liquid interaction chromatography: Principles - prospects - limitations, Journal of Chromatography A. 1335 (2014) 62–79. https://doi.org/10.1016/j.chroma.2013.12.010.

[4] B. Trathnigg, Determination of MWD and chemical composition of polymers by chromatographic techniques, Progress in Polymer Science. 20 (1995) 615–650. https://doi.org/10.1016/0079-6700(95)00005-Z.

[5] A. Baumgaertel, E. Altuntaş, U.S. Schubert, Recent developments in the detailed characterization of polymers by multidimensional chromatography, Journal of Chromatography A. 1240 (2012) 1–20. https://doi.org/10.1016/j.chroma.2012.03.038.

[6] I.A. Haidar Ahmad, A.M. Striegel, Determining the absolute, chemical-heterogeneity-corrected molar mass averages, distribution, and solution conformation of random copolymers, Analytical and Bioanalytical Chemistry. 396 (2010) 1589–1598. https://doi.org/10.1007/s00216-009-3320-9.

[7] W.C. Knol, B.W.J. Pirok, R.A.H. Peters, Detection challenges in quantitative polymer analysis by liquid chromatography, Journal of Separation Science. 44 (2021) 63–87. https://doi.org/10.1002/jssc.202000768.

[8] P. Castignolles, R. Graf, M. Parkinson, M. Wilhelm, M. Gaborieau, Detection and quantification of branching in polyacrylates by size-exclusion chromatography (SEC) and melt-state 13C NMR spectroscopy, Polymer (Guildf). 50 (2009) 2373–2383. https://doi.org/10.1016/j.polymer.2009.03.021.

[9] T. Chang, H.C. Lee, W. Lee, S. Park, C. Ko, Polymer characterization by temperature gradient interaction chromatography, Macromolecular Chemistry and Physics. 200 (1999) 2188–2204. https://doi.org/10.1002/(sici)1521-3935(19991001)200:10<2188::aid-macp2188>3.3.co;2-6.

[10] W. Lee, D. Cho, B.O. Chun, T. Chang, M. Ree, Characterization of polystyrene and polyisoprene by normal-phase temperature gradient interaction chromatography, Journal of Chromatography A. 910 (2001) 51–60. https://doi.org/10.1016/S0021-9673(00)01163-8.

[11] W. Radke, S. Lee, T. Chang, Temperature gradient interaction chromatography of polymers: A molecular statistical model, Journal of Separation Science. 33 (2010) 3578–3583. https://doi.org/10.1002/jssc.201000462.

[12] M. Schollenberger, W. Radke, SEC-Gradients, an alternative approach to polymer gradient chromatography: 1. Proof of the concept, Polymer (Guildf). 52 (2011) 3259–3262. https://doi.org/10.1016/j.polymer.2011.05.047.

[13] M. Schollenberger, W. Radke, Size exclusion chromatography-gradients, an alternative approach to polymer gradient chromatography: 2. Separation of poly(meth)acrylates using a size exclusion chromatography-solvent/non-solvent gradient, Journal of Chromatography A. 1218 (2011) 7827–7831. https://doi.org/10.1016/j.chroma.2011.08.090.

[14] F.A. Messaud, R.D. Sanderson, J.R. Runyon, T. Otte, H. Pasch, S.K.R. Williams, An overview on field-flow fractionation techniques and their applications in the separation and characterization of polymers, Progress in Polymer Science (Oxford). 34 (2009) 351–368. https://doi.org/10.1016/j.progpolymsci.2008.11.001.

[15] P. Jandera, M. Holčapek, L. Kolářová, Retention mechanism, isocratic and gradient-elution separation and characterization of (co)polymers in normal-phase and reversed-phase high-performance liquid chromatography, in: Journal of Chromatography A, 2000: pp. 65–84. https://doi.org/10.1016/S0021-9673(99)01216-9.

[16] F. Fitzpatrick, R. Edam, P. Schoenmakers, Application of the reversed-phase liquid chromatographic model to describe the retention behaviour of polydisperse macromolecules in gradient and isocratic liquid chromatography, Journal of Chromatography A. 988 (2003) 53–67. https://doi.org/10.1016/S0021-9673(02)02050-2.

[17] P. Schoenmakers, F. Fitzpatrick, R. Grothey, Predicting the behaviour of polydisperse polymers in liquid chromatography under isocratic and gradient conditions, Journal of Chromatography A. 965 (2002) 93–107. https://doi.org/10.1016/S0021-9673(01)01322-X.

[18] T. Brooijmans, P. Breuer, A. Schreuders, M. van Tilburg, P.J. Schoenmakers, R.A.H. Peters, Charge-Based Separation of Acid-Functional Polymers by Non-aqueous Capillary Electrophoresis Employing Deprotonation and Heteroconjugation Approaches, Analytical Chemistry. 93 (2021) 5924–5930. https://doi.org/10.1021/acs.analchem.1c00311.

[19] S. Abrar, B. Trathnigg, Analysis of polyethyleneoxide macromonomers by liquid chromatography along the critical adsorption line, Analytical and Bioanalytical Chemistry. 400 (2011) 2577–2586. https://doi.org/10.1007/s00216-010-4554-2.

[20] M. Mlynek, W. Radke, Critical chromatography in ternary solvents, Journal of Chromatography A. 1284 (2013) 112–117. https://doi.org/10.1016/J.CHROMA.2013.02.005.

[21] K.J. Bombaugh, R.F. Levangie, High Resolution Gel Permeation Chromatography-Using Recycle, Separation Science. 5 (1970) 751–763. https://doi.org/10.1080/00372367008055537.

[22] J. Porath, P. Flodin, Gel Filtration: A method for desalting and group separation, Nature. 183 (1959) 1657–1659. https://doi.org/10.1038/1831657a0.

[23] F. Gritti, S. Besner, S. Cormier, M. Gilar, Applications of high-resolution recycling liquid chromatography: From small to large molecules, Journal of Chromatography A. 1524 (2017) 108–120. https://doi.org/10.1016/j.chroma.2017.09.054.

[24] F. Gritti, Rebirth of recycling liquid chromatography with modern chromatographic columns: Extension to gradient elution, Journal of Chromatography A. 1653 (2021). https://doi.org/10.1016/j.chroma.2021.462424.

[25] F. Gritti, S. Cormier, Performance optimization of ultra high-resolution recycling liquid chromatography, Journal of Chromatography A. 1532 (2018) 74–88. https://doi.org/10.1016/j.chroma.2017.11.047.

[26] L.W. Lim, H. Uzu, T. Takeuchi, Separation of benzene and deuterated benzenes by reversed-phase and recycle liquid chromatography using monolithic capillary columns, Journal of Separation Science. 27 (2004) 1339–1344. https://doi.org/10.1002/jssc.200401882.

[27] A.M. Skvortsov, A.A. Gorbunov, D. Berek, B. Trathnigg, Liquid chromatography of macromolecules at the critical adsorption point: Behaviour of a polymer chain inside pores, Polymer (Guildf). 39 (1998) 423–429. https://doi.org/10.1016/S0032-3861(97)00279-6.

[28] Y. Brun, P. Alden, Gradient separation of polymers at critical point of adsorption, Journal of Chromatography A. 966 (2002) 25–40. https://doi.org/10.1016/S0021-9673(02)00705-7.

[29] A.M. Skvortsov, A.A. Gorbunov, Achievements and uses of critical conditions in the chromatography of polymers, Journal of Chromatography A. 507 (1990) 487–496. https://doi.org/10.1016/S0021-9673(01)84228-X.

[30] L.R. Snyder, M.A. Stadalius, M.A. Quarry, L.R. Snyder, Gradient Elution in Reversed-Phase HPLC Separation of Macromolecules, Analytical Chemistry. 55 (1983) 1412A-1430A. https://doi.org/10.1021/ac00264a001.

[31] L.R. Snyder, Linear elution adsorption chromatography. VII. gradient elution theory, Journal of Chromatography A. 13 (1964) 415–434. https://doi.org/10.1016/s0021-9673(01)95138-6.

[32] L.M. Blumberg, Theory of gradient elution liquid chromatography with linear solvent strength: Part 1. migration and elution parameters of a solute band, Chromatographia. 77 (2014) 179–188. https://doi.org/10.1007/s10337-013-2555-y.

[33] P. Jandera, J. Churáček, Gradient elution in liquid chromatography. II. Retention characteristics (retention volume, band width, resolution, plate number) in solvent-programmed chromatography - theoretical considerations, Journal of Chromatography A. 91 (1974) 223–235. https://doi.org/10.1016/S0021-9673(01)97902-6.

[34] X. Jiang, A. van der Horst, P.J. Schoenmakers, Breakthrough of polymers in interactive liquid chromatography, Journal of Chromatography A. 982 (2002) 55–68. https://doi.org/10.1016/S0021-9673(02)01483-8.

[35] H. Poppe, J. Paanakker, M. Bronckhorst, Peak width in solvent-programmed chromatography. I. General description of peak broadening in solvent-programmed elution, Journal of Chromatography A. 204 (1981) 77–84. https://doi.org/10.1016/S0021-9673(00)81641-6.

[36] U.D. Neue, D.H. Marchand, L.R. Snyder, Peak compression in reversed-phase gradient elution, Journal of Chromatography A. 1111 (2006) 32–39. https://doi.org/10.1016/j.chroma.2006.01.104.

[37] F. Gritti, General theory of peak compression in liquid chromatography, Journal of Chromatography A. 1433 (2016) 114–122. https://doi.org/10.1016/j.chroma.2016.01.032.

[38] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, Analytical Chemistry. 77 (2005) 5670–5677. https://doi.org/10.1021/ac0506783.

[39] M.A. Bashir, W. Radke, Comparison of retention models for polymers. 1. Poly(ethylene glycol)s, Journal of Chromatography A. 1131 (2006) 130–141. https://doi.org/10.1016/j.chroma.2006.07.089.

[40] P.J. Schoenmakers, H.A.H. Billiet, R. Tussen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, Journal of Chromatography A. 149 (1978) 519–537. https://doi.org/10.1016/S0021-9673(00)81008-0.

[41] E.C. Freiling, Ion Exchange as a Separations Method. IX. Gradient Elution Theory, J Am Chem Soc. 77 (1955) 2067–2071. https://doi.org/10.1021/ja01613a010.

[42] L.M. Blumberg, Migration and elution equations in gradient liquid chromatography, Journal of Chromatography A. 1599 (2019) 35–45. https://doi.org/10.1016/j.chroma.2019.03.057.

[43] M.J. den Uijl, P.J. Schoenmakers, B.W.J. Pirok, M.R. van Bommel, Recent applications of retention modelling in liquid chromatography, Journal of Separation Science. 44 (2021) 88–114. https://doi.org/10.1002/jssc.202000905.

[44] L.R. Snyder, J.W. Dolan, High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model, 2006. https://doi.org/10.1002/0470055529.

[45] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, Journal of Chromatography A. 1217 (2010) 3794–3803. https://doi.org/10.1016/j.chroma.2010.04.023.

[46] L.R. Snyder, J.J. Kirkland, J.W. Dolan, Introduction to Modern Liquid Chromatography, 2010. https://doi.org/10.1002/9780470508183.

[47] T.S. Bos, L.E. Niezen, M.J. den Uijl, S.R.A. Molenaar, S. Lege, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Reducing the influence of geometry-induced gradient deformation in liquid chromatographic retention modelling, Journal of Chromatography A. 1635 (2021). https://doi.org/10.1016/j.chroma.2020.461714.

[48] M.A. Quarry, R.L. Grob, L.R. Snyder, Measurement and use of retention data from high-performance gradient elution. Correction for "non-ideal" processes originating within the column, Journal of Chromatography A. 285 (1984) 19–51. https://doi.org/10.1016/S0021-9673(01)87733-5.

[49] F. Gritti, G. Guiochon, The distortion of gradient profiles in reversed-phase liquid chromatography, Journal of Chromatography A. 1340 (2014) 50–58. https://doi.org/10.1016/j.chroma.2014.03.004.

[50] A.P. Schellinger, D.R. Stoll, P.W. Carr, High speed gradient elution reversed-phase liquid chromatography, Journal of Chromatography A. 1064 (2005) 143–156. https://doi.org/10.1016/j.chroma.2004.12.017.

[51] M. Martin, G. Guiochon, Effects of high pressure in liquid chromatography, Journal of Chromatography A. 1090 (2005) 16–38. https://doi.org/10.1016/j.chroma.2005.06.005.

[52] X. Liu, D. Zhou, P. Szabelski, G. Guiochon, Influence of pressure on the retention and separation of insulin variants under linear conditions, Analytical Chemistry. 75 (2003) 3999–4009. https://doi.org/10.1021/ac0205964.

[53] G. Guiochon, M.J. Sepaniak, Influence of pressure on solute retention in liquid chromatography, Journal of Chromatography A. 606 (1992) 248–250. https://doi.org/10.1016/0021-9673(92)87031-3.

Chapter 4

Principles and potential of solvent gradient SEC for polymer analysis

Abstract

The properties of a polymeric material are influenced by its underlying molecular distributions, including the molecular-weight (MWD), chemical composition (CCD), and/or block length (BLD) distributions. Gradient-elution liquid chromatography (LC) is commonly used to determine the CCD. Due to the limited solubility of polymers, samples are often injected in strong solvents. Such solvents may lead to broadened or poorly shaped peaks and, in unfavourable cases, to "breakthrough" phenomena, where a part of the sample travels through the column unretained. To remedy this, a technique called size-exclusion-chromatography gradients or gradient size-exclusion chromatography (gSEC) was developed in 2011. In this work, we aim to further explore the potential of gSEC for the analysis of the CCD, also in comparison with conventional gradient-elution reversed-phase LC, which in this work corresponded to gradient-elution reversed-phase liquid chromatography (RPLC). The influence of the mobile-phase composition, the pore size of the stationary-phase particles, and the column temperature were investigated. The separation of five styrene/ethyl acrylate copolymers was studied with one-dimensional RPLC and gSEC. RPLC was shown to lead to a more-accurate CCD in shorter analysis time. The separation of five styrene/methyl methacrylate copolymers was also explored using comprehensive two-dimensional (2D) LC involving gSEC, *i.e.* SEC×gSEC and SEC×RPLC. In 2D-LC, the use of gSEC was especially advantageous as no breakthrough could occur.

Publication: L.E. Niezen, J.D. Kruijswijk, G. van Henten, B.W.J. Pirok, B.B.P. Staal, W. Radke, H.J.A. Philipsen, G.W. Somsen, and P.J. Schoenmakers, Principles and potential of solvent gradient size-exclusion chromatography for polymer analysis, Anal Chim Acta, *accepted*

4.1. Introduction

Polymers are among the most important building blocks of materials. The applications of polymers are nearly limitless, from packaging applications to electronics, paints, clothing, and drug delivery systems. To continuously improve polymeric materials, it is vital to understand how their molecular structure and composition relate to their physical properties. Several techniques can be used to characterize the chemical structure of polymers, including pyrolysis – gas chromatography coupled to mass spectrometry, and spectroscopic methods such as infrared, ultraviolet absorbance, Raman or nuclear magnetic resonance spectroscopy [1–6]. However, most of these methods merely provide average polymer composition, and do not yield information on the distributions present, such as molecular-weight (MWD), chemical-composition (CCD), and/or block-length (BLD) distributions. To assess these distributions, liquid chromatography (LC) is more generally used [7–16]. A common LC technique for the analysis of polymers is size-exclusion chromatography (SEC), which is a well-established benchmark for determining molecular-weight distributions (MWD) [1,10,13,14]. There is no such benchmark method for the analysis of the CCD, although gradient-elution LC is probably most commonly used [7–11,15,16].

The greatest challenges for the application of solvent-gradient LC are analyte detection and polymer solubility. Due to the changing mobile-phase composition, many typical detectors used for SEC, such as refractive-index, viscometric, or light-scattering detectors cannot be used. Efforts to make such detectors work with gradients [17–20] have been moderately successful at best. Many polymers are difficult to dissolve. Strong solvents and patience are typically required and injecting the resulting sample at the starting conditions of a gradient separation, in a weak eluent, can be problematic. This may be exacerbated for crystalline (co)polymers, due to issues with slow redissolution [21]. Even if crystallinity is not an issue, an injection solvent that is a strong eluent can still lead to broad or deformed peaks and, ultimately, to a phenomenon called breakthrough [22]. When this occurs a part of the sample stays in the solvent plug and passes (nearly) unretained through the column. Several methods have been developed to address this issue [23,24]. These include sandwich injection [23] and solvent-mixing strategies [24]. Additionally, so-called barrier methods may be applied [25,26]. However, the latter fundamentally

result in very low peak capacities (one per barrier). SEC-gradients, henceforth referred to in this work as gradient-SEC (gSEC), can be seen as a much-improved implementation of a barrier method, with a gradual rather than stepwise change in mobile-phase composition. Originally introduced by Schollenberger *et al.* [27,28], gSEC eliminates the risk of breakthrough and offers a much larger peak capacity than conventional barrier methods (the principles and practise of gSEC are outlined below). However, reports on the use of gSEC for the separation of synthetic polymers have so far been scarce [27–30]. While some work has been performed illustrating different elution mechanisms [27,28], many aspects of gSEC have not yet been studied. These include the influence of the solvent composition and of the column packing material (*e.g.* the particle pore size). Importantly, gSEC has never been directly compared with conventional gradient-elution LC, neither in one-dimensional LC nor as a second-dimension separation in two-dimensional LC.

Our objectives in the present work were to critically evaluate the influence of the mobile phase, the temperature, and the pore size in gSEC, and to compare RPLC and gSEC, so as to highlight the advantages and shortcomings of both methods. SEC separations of polystyrene (PS) and polymethyl methacrylate (PMMA) were performed on various columns of different chemistries, containing particles of different pore sizes, and using various mobile-phase mixtures of acetonitrile (ACN) and tetrahydrofuran (THF), at two different temperatures (25 and 60 °C). Following these experiments, gSEC measurements were performed using small-pore stationary-phase particles. We set out to compare the separation of five styrene/ethyl acrylate (S/EA) copolymers and five styrene/methyl methacrylate (S/MMA) copolymers, each characterized by a broad MWD, but a narrow CCD. Both RPLC and gSEC were used for the separation of the S/EA copolymers. For the S/MMA copolymers comprehensive two-dimensional liquid chromatography (LC×LC) in either SEC×RPLC or SEC×gSEC mode were applied. To perform SEC×RPLC successfully, we needed to avoid breakthrough in the second dimension. This is notoriously difficult, because a strong first-dimension solvent is combined with a large second-dimension injection volume. In SEC×gSEC breakthrough did not occur, as the strong first-dimension solvent matches the starting conditions of the gSEC separation.

4.2. Theory

The retention of an analyte may be described using the retention factor (k), given by

$$k = \frac{t_R - t_0}{t_0} \quad (4.1)$$

where t_R is the retention time of the analyte and t_0 is the void time of the column (given by the void volume V_0, divided by the flowrate, F), or the time it takes for an unretained marker to move through the column. On a particular column, k is commonly adjusted by changing the mobile phase composition and/or the temperature. Solvents are generally classified as either weak (causing large k) or strong (causing small k). For high-molecular-weight analytes, such as polymers, the size (hydrodynamic radius) of the analyte molecules in relation with the pore diameter of the packing material also plays a role for retention. When the solvent strength is sufficiently high, no interaction occurs between the analytes and the packing material. The elution order is then (ideally) determined solely by the hydrodynamic volume of the polymer, as the accessible pore volume will be different for polymers of different sizes (size-exclusion conditions). In this case, elution occurs before t_0 and the analytes travel faster than the surrounding mobile phase. At low solvent strength the situation is more complex, as the polymer may not be soluble, or k may be very large. In either case, the polymer will not elute from the column. When increasing the fraction of strong solvent (φ) in the case of gradient elution, solubility generally improves and retention (*i.e.* the local retention factor) decreases with time. The migration of the polymer through the column now depends on how quickly solubility improves with φ or how quickly k decreases with φ. If there is no interaction with the column at the mobile phase composition where the polymer is first soluble, the elution composition will be solely determined by the polymer's solubility [12,31]. In all other cases the polymer will elute "normally", *i.e.* primarily based on the strength of interaction with the column material [32–34]. Size-exclusion conditions will not be encountered, because as soon as polymer molecules move faster than the mobile phase they will experience a weaker solvent and be retained again until the eluent strength is sufficient.

To describe the gradient-elution process quantitatively it must be known how φ changes with time, and how k changes with φ. In previous work, it has been shown that the retention of a polymer in gradient-elution RPLC is reasonably well described

by a log-linear model, often referred to as the linear-solvent strength (LSS) model [32–36]. It is assumed that the logarithm or natural logarithm of the retention factor (k), varies linearly with φ. Based on this model an "effective" gradient steepness parameter can be defined. Using this concept, it was shown that effectively steep gradients minimize the influence of the MWD on retention, while the influence of the CCD is enhanced [37]. In such a steep gradient all molecules of a specific homopolymer, or of a co-polymer of a given composition, are found to elute at a specific ("critical") composition (φ_{crit}), independent of their molecular weight [33,34,37–43].

Effectively steep gradients are most easily realized by using a step gradient, changing instantaneously from an initial composition (φ_{init}, weak eluent, high k) to a final composition (φ_{final}, strong eluent, low k) higher than φ_{crit}. However, this results in co-elution of all analytes (homopolymers and copolymers) with critical compositions that fall within the range of φ that is covered by the gradient (*i.e.* $\varphi_{\mathrm{init}} \leq \varphi_{\mathrm{crit}} \leq \varphi_{\mathrm{final}}$). This implies that a finite $\Delta\varphi$ results in co-elution of a (large) faction of the CCD of a copolymer. Therefore, a step gradient cannot be used to determine the CCD. For this purpose, a continuous gradient must be used, of which linear gradients are most common. However, a step gradient may be used to separate two different homo- or co-polymers with sufficiently different φ_{crit}.

Normally, analytes are injected before the gradient arrives at the column inlet. It is also possible to inject the analyte after the end of the step gradient has passed the injector. If $\varphi_{\mathrm{crit}} \leq \varphi_{\mathrm{final}}$, the injected analyte molecules experience SEC conditions, and large molecules travel through the column faster than the gradient. This results in so-called "barrier" methods, where a step gradient ($\varphi_{\mathrm{init}} \leq \varphi_{\mathrm{crit}} \leq \varphi_{\mathrm{final}}$) is established within the column before injection [25,26]. If the analyte travels much faster than the solvent and the time of injection is not too far after the step gradient has passed the injector, the analyte molecules may catch up with the gradient step. They cannot move past the gradient, as they would experience a weaker solvent and slow down. Again, all analytes with $\varphi_{\mathrm{init}} \leq \varphi_{\mathrm{crit}} \leq \varphi_{\mathrm{final}}$ will co-elute at the moment the step gradient reaches the end of the column. Once again, the use of a continuous (*e.g.* a linear) gradient will provide a better characterization of the CCD. This latter type of method is referred to as gSEC [27–30].

As with conventional gradient-elution RPLC, a gSEC separation that is based primarily on the CCD is desirable. Such a separation can be combined with SEC to achieve high orthogonality in LC×LC. With a conventional gradient, the influence of the MWD is reduced when using effectively steep gradients [37]. In the case of gSEC this is also true, as whether analyte molecules can reach φ_{crit} before the gradient elutes from the column depends on the extent of exclusion they experience and on how far the analyte molecules must travel through the column to reach φ_{crit}. As with a conventional gradient, the effective gradient steepness depends on the ratio of the column volume to the gradient volume. At constant column volume, a smaller gradient volume (steeper gradient) implies that the analyte has to travel less far to reach φ_{crit}. A larger column volume (relative to the volume of the gradient) implies that analyte molecules have a greater chance (*i.e.* a greater part of the column) to reach φ_{crit}. The maximum gradient volume ($V_{G,\mathrm{max}}$), such that analyte molecules can reach φ_{crit}, can be calculated if a set of criteria are met, *viz.* (*i*) a critical composition exists for a particular combination of mobile phase, stationary phase, and analyte (*i.e.* $\varphi_{\mathrm{init}} \leq \varphi_{\mathrm{crit}} \leq \varphi_{\mathrm{final}}$, where the initial and final compositions may cover the entire range from 0 to 1), (*ii*) the analyte molecules are unretained at $\varphi > \varphi_{\mathrm{crit}}$, (*iii*) the injection takes place at the moment the end of the gradient arrives at the column inlet, and (*iv*) the calibration curve for the column is known. $V_{G,\mathrm{max}}$ may then be calculated using

$$V_{G,\mathrm{max}} = \frac{V_0 - \left(V_i + K_{\mathrm{SEC}} V_p\right)}{1 - \frac{\varphi_{\mathrm{crit}} - \varphi_{\mathrm{init}}}{\varphi_{\mathrm{final}} - \varphi_{\mathrm{init}}}} = \frac{\Delta\varphi}{\varphi_{\mathrm{final}} - \varphi_{\mathrm{crit}}} \left(V_0 - \left(V_i + K_{\mathrm{SEC}} V_p\right)\right) \tag{4.2}$$

where V_i and V_p are the interstitial, and pore volume, respectively, and K_{SEC} is the SEC distribution coefficient. It is assumed that $V_0 = V_i + V_p$. Note that any smaller gradient volume would lead to a steeper gradient and causes elution of the analyte at $\varphi = \varphi_{\mathrm{crit}}$; hence, our use of the term "maximum gradient volume". For gradient volumes larger than $V_{G,max}$ the analytes cannot catch up with φ_{crit} and elute unretained. In such situations, elution is based on hydrodynamic size, but not on chemical composition. In gSEC, each analyte will have to first move some distance through the column to reach φ_{crit}. This is in contrast with conventional gradient-elution LC, where migration of the analyte polymers will, for relatively large molecular-weight analytes, occur within the vicinity of φ_{crit} as the analyte is first retained at the head of the column. As a result, a steeper gradient will be required

using gSEC compared to conventional gradient-elution LC to separate high-molecular-weight polymers with different φ_{crit}. Some other general conclusions can also be drawn from **Equation 4.2**, namely: *i*) to avoid the use of very small gradient volumes, analyte exclusion should be promoted (K_{SEC} should be minimized) by choosing small-pore-size packings, *ii*) columns with a large pore volume (V_p) are preferred, since this maximizes the migration velocity of the analyte relative to the mobile-phase velocity, and *iii*) V_G will be determined by the polymer with the smallest φ_{crit} in the sample. Under the assumptions underlying **Equation 4.2,** it is clear that gSEC will perform better when the difference between φ_{crit} and φ_{final} is reduced. However, this would hinder the ability of the method to analyse samples of widely different chemical composition, since these samples will feature large differences in φ_{crit}. Once again, conventional gradient-elution will be less affected by this, since analytes will only move within the vicinity of φ_{crit}. In this study we set out to evaluate the points raised above and to verify the conclusions drawn from **Equation 4.2**.

4.3. Experimental

Two different systems (A and B), in two different laboratories (referred to below as laboratory A, located at the University of Amsterdam and laboratory B, located at the Vrije Universiteit Amsterdam), were used for different parts of this work. Certain samples were common to both laboratories, these are described here.

A polystyrene (PS) standards kit was obtained from PSS (Mainz, Germany) and polymethylmethacrylate (PMMA) standards were obtained from Polymer Laboratories (Church Stretton, UK). Styrene/methyl methacrylate (S/MMA) and styrene/ethyl acrylate (S/EA) copolymers were synthesized in-house using thermally initiated free-radical polymerization, the procedure, chemicals, along with their molar masses, polydispersity, and approximate average chemical composition are included in the supplementary information (Table S-1, Section S-1).

4.3.1. Laboratory A

4.3.1.1. Chemicals and Materials

Non-stabilized tetrahydrofuran (THF, 99.9%, LC-MS Grade) and toluene (99%, LC-MS grade) were obtained from VWR Chemicals (Darmstadt, Germany), acetonitrile (ACN, ≥99.9%, LC-MS Grade) was obtained from Biosolve (Valkenswaard, The Netherlands).

4.3.1.2. Systems and Equipment

The experiments in this study were carried out on an Agilent system, combining components from both an Agilent 1100 and a 1290 Infinity 2D-LC system, all obtained from Agilent (Waldbronn, Germany). The system included an 1100 autosampler (G1313A), an 1100 capillary pump (G1376A), a 1290 binary pump (G7120A) equipped with a jet weaver V35 mixer, and a 1290 column compartment (G1316C) equipped with a 2-position/8-port valve (model 5067-4214). For detection, the system comprised a 1290 diode-array detector (DAD, G4214A) with a max-light cartridge cell (model G4212-6008, 10 mm path length, 1 µL cell volume), and a 1260 evaporative light-scattering detector (ELSD, G4260B). The capillary pump was used for the ^{1}D SEC measurements and the binary pump for the gSEC experiments. The system was controlled using Agilent OpenLAB CDS ChemStation software (rev. c.01.10).

The ^{1}D-gSEC and SEC-recovery experiments were carried out on an Agilent 1290 Infinity LC system. This system included a 1290 Infinity II autosampler (G7129B), a 1290 binary pump (G7120A) equipped with a jet weaver V35 mixer, and a 1290 MCT column oven (G7116B). For detection, the system comprised a 1290 DAD (G7117B) with a max-light cartridge cell (Model G4212-6008, 10 mm path length, 1 µL cell volume). The system was controlled using Agilent OpenLAB CDS ChemStation software (rev. c.01.10).

Conventional (isocratic) SEC measurements were performed at various ACN/THF eluent compositions, at 25 and 60°C. The flow rate was 0.5 mL·min^{-1}, the injection volume 3.0 µL, and the sample concentration was 0.5 mg·mL^{-1}. Columns used for these experiments included a 50 × 4.6 mm XBridge bridged-ethylene hybrid (BEH) Shield RP18 XP column, packed with 130-Å, 3.5-µm particles, two Nova-Pak 150 ×

3.9 mm C_{18} columns packed with 60-Å, 4-µm particles, all purchased from Waters (Milford, MA, USA), as well as a 150 × 4.6 mm Dionex Acclaim Polar-Advantage C_{18} column packed with 300-Å, 3-µm particles purchased from Dionex (Sunnyvale, CA, USA), and two sets of two 250 × 4.6 mm Nucleosil columns (C_{18} and bare silica), both packed with 4000-Å, 5-µm particles, purchased from Macherey Nagel (Düren, Germany).

For the 1D-LC gSEC experiments, three different silica-based C_{18} columns were used; often multiple columns of the same type were coupled in series. These columns included two 50 × 4.6 mm and one 100 × 4.6 mm XBridge BEH Shield RP18 XP columns, packed with 130-Å, 3.5-µm particles, the two Novapak columns, and the Dionex Acclaim Polar-Advantage C_{18} column. For these experiments an injector program was used to realize the required delayed injection. Injection was timed to occur one minute after the start of the experiment, *i.e.* at the end of the gradient. A variable initial hold-up time was programmed into the gradient so that gradient volumes could be varied based on the volume of each column, while simultaneously keeping the injection time constant. The measured retention times were not adjusted for the injection time and are relative to the start of the experiment.

For the 2D-LC experiments, two 150 × 2.1 mm APC SEC columns packed with 2.5-µm BEH particles with 450-Å pore size were used in the first dimension, while for the second-dimension separation two of the 50 × 4.6 mm XBridge BEH Shield RP18 XP columns were used. All columns were obtained from Waters. For SEC×RPLC, the gradient program was as follows: 0-0.1 min linear gradient 80/20% ACN/THF, 0.1-0.65 min linear gradient 50/50%, 0.65-0.75 linear gradient 45/55%. For SEC×gSEC, it was the following: 0-0.1 min isocratic 40/60% ACN/THF, 0.1-0.2 min linear gradient to 95/5%, 0.2-0.35 min linear gradient 20/80%, 0.35-0.9 min linear gradient 50/50%, 0.9-0.95 min linear gradient 45/55%, 0.95 min isocratic 40/60%.

The data of φ_{crit} *vs.* temperature was obtained from gradient experiments performed at different temperatures on the XBridge column. The critical composition was assumed to be equivalent to the elution composition of a high-molecular-weight standard.

Data analysis was performed in MATLAB R2020a (Mathworks, Woodshole, MA, USA).

4.3.2. Laboratory B

4.3.2.1. Chemicals and Materials

Acetonitrile (ACN, ≥99.9%, HPLC Grade) was obtained from Biosolve, non-stabilized tetrahydrofuran (THF, HPLC grade) was obtained from VWR Chemicals.

4.3.2.2. Systems and Equipment

The experiments were performed on a 1290 Infinity II Agilent system. The system included an autosampler (G7167B), a quaternary pump (G7104A), a multi-column thermostat (MCT, G7116B). A variable-wavelength detector (VWD, G7114B) and an ELSD (G7102A) were used for detection. The VWD was set to record UV absorption at 210 and 260 nm.

Conventional (isocratic) SEC measurements were performed at various ACN/THF eluent compositions, at 25 and 60°C. The flow rate was 0.5 $mL \cdot min^{-1}$, the injection volume 3.0 µL, and the sample concentration was 0.5 $mg \cdot mL^{-1}$. Columns used for these experiments were all Zorbax columns purchased from Agilent. These included a 100 × 4.6 mm Rapid Resolution StableBond C_{18} column, packed with 300-Å, 3.5-µm particles, a 250 × 9.4 mm semi-preparative StableBond C_{18} column packed with 300-Å, 5-µm particles, and a 250 × 9.4 mm semi-preparative StableBond C_{18} column packed with 80-Å, 5-µm particles.

For the ^{1}D-gSEC and RPLC experiments in sections 4.3 and 4.4, the gradient program is described in the respective figures. The temperature was 25°C, the flow rate 0.5 $mL \cdot min^{-1}$, the injection volume 3.0 µL, and the sample concentration 0.5 $mg \cdot mL^{-1}$. Once again, a delayed injection was used for the gSEC experiments. The injection took place after a delay of 0.4 mL (48 s) to account for the system dwell volume. These experiments were carried out on Rapid Resolution StableBond column.

4.4. Results and Discussion

4.4.1. Influence of mobile-phase composition on elution volume in SEC

The elution volume in SEC is generally thought to be independent of the nature and composition of the solvent, provided that interactions with the stationary phase are absent. However, the size (hydrodynamic radius) of molecules in solution may be affected by the nature of the solvent. Because the solvent composition is varied in gSEC, it is relevant to investigate the effects of solvent composition on elution volumes and calibration curves in SEC. Similar work, albeit using overall stronger mobile phase compositions, has previously been performed by Caltabiano *et al.* [44].

Conventional (isocratic) SEC experiments were performed using different fractions of THF in ACN, all of which were above the φ_{crit} of PS and PMMA, to ensure SEC behaviour. The results of these SEC experiments for homopolymer PS and PMMA on various columns with different pore sizes and chemistries are provided in **Figure 4.1**.

The elution volume of PS increases with an increase of the concentration of weak solvent (ACN) on all C_{18} stationary phases, whereas the effect of the THF fraction on the elution volume of PS is small on the one bare-silica column tested (**Figure 4.1-H**). The possible explanations for the increase in elution volume with a decrease in strong solvent when using the C_{18} columns include the following:

i) the polymeric coil size/hydrodynamic volume decreases with an increase of the weak solvent. This effect should be independent of the column used;

ii) the pore volume of the column changes with the mobile phase composition, due to a swelling or shrinking of the packing particles;

iii) non-SEC interactions occur with the stationary phase as the mobile-phase composition approaches φ_{crit};

iv) a change in the ratio of the interstitial to pore volume occurs due to a change in the thickness of the C_{18} layer with mobile-phase composition.

Several of these effects may occur concomitantly. Explanation (*i*), the addition of ACN (non-solvent for PS) is expected to result in a contraction of the PS chain. Precipitation of PS occurs at a composition of about 50/50 in ACN/THF. For the bare

silica columns (**Figure 4.1-H**) ACN is a stronger (more polar) solvent than THF. However, because solvents are strong, retention on these columns should not be strongly affected by the ratio of ACN/THF. Thus, the small increase in elution volume observed for the larger standards on the bare silica column might be mainly attributed to the change of coil size with eluent composition. The effect should increase with increasing polymer size and be most prominent in the linear (shallow) part of the calibration curve. This is indeed observed in **Figure 4.1-F**. For similar slopes of the calibration curves this effect is expected to occur to similar extend also on C_{18} columns, but there it is likely overshadowed by other effects.

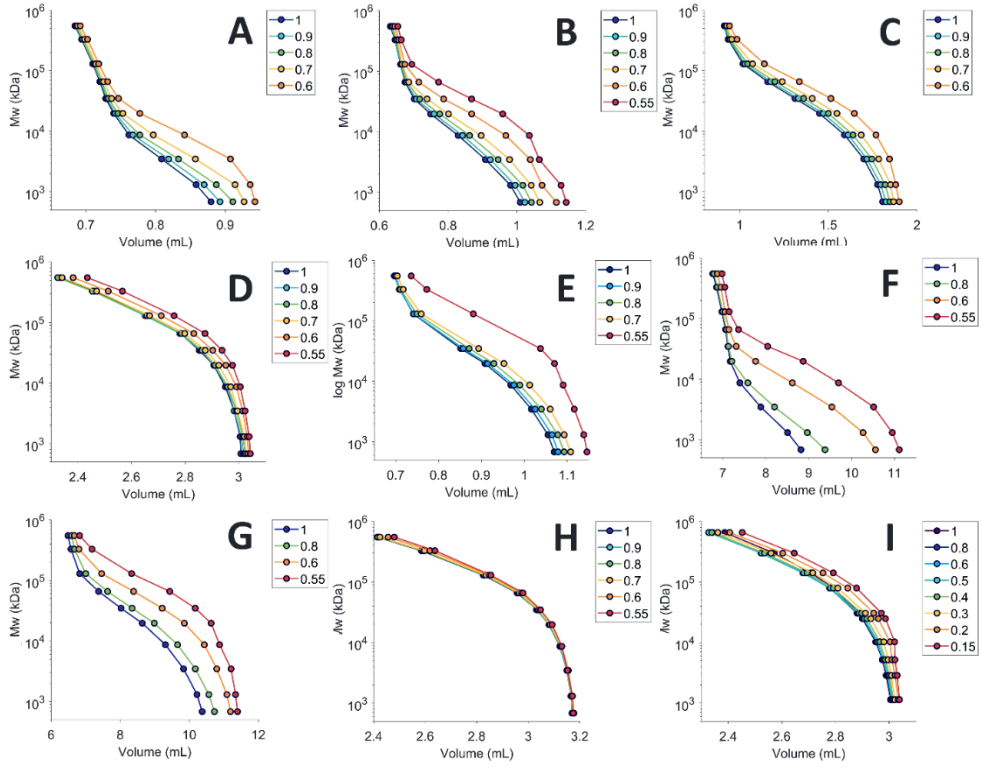

Figure 4.1: SEC calibration curves of ten PS standards (A-G) and ten PMMA standards (H-I) obtained for different mobile phase compositions at 25°C. The fraction of THF (in ACN) is indicated in the legend. Columns: A) Novapak C_{18} (150 × 3.9 mm, 60Å), B) Two XBridge C_{18} (50 × 4.6 mm, 130Å), C) Dionex C_{18} (150 × 4.6 mm, 300Å), D and I) Nucleosil C_{18} (250 × 4.6 mm, 4000Å), E), Zorbax SB C_{18} (100 × 4.6 mm, 300Å), F) Zorbax SB C_{18} (250 × 9.4 mm, 80Å), G) Zorbax SB C_{18} (250 × 9.4 mm, 300Å), and H) Nucleosil Bare Silica (250 × 4.6 mm, 4000Å).

Explanation (*ii*), an effect of the mobile-phase composition on the volume occupied by the solid packing material in the column is considered highly unlikely for the silica particles used in the present study. The effect may be quite significant when other types of particles, such as PS cross-linked with divinylbenzene (PS-DVB), are used.

This leaves an increase in interaction of the analyte molecules with the C_{18} layer (Explanation *iii*) or a change in the stationary phase volume (Explanation *iv*) as the more-probable causes of the significant increase in elution volume on the C_{18} columns. Explanation *iii* is an analyte-specific effect, whereas Explanation *iv* is column specific. Therefore, it may be possible to distinguish between the two effects by performing the same experiments with a series of analyte polymers with similar hydrodynamic volumes, but a different chemical structure. Hence, the experiments on the 4000Å Nucleosil C_{18} column (**Figure 4.1-D**) were repeated for a set of PMMA standards (**Figure 4.1-I**). At first sight, the observed variations in the calibration curves seem to be similar for the PS and PMMA standards. However, a closer inspection of the data reveals that at low concentrations of ACN the elution volume of the PMMA standards slightly increases with an increase in ACN content, whereas an increase in elution volume is observed at high concentrations of ACN. The maximum effect for PMMA is reached at 85% ACN in **Figure 4.1-F**, whereas in case of PS it is already reached at 45% ACN (see supplementary information, Figure S-1, section S-2). For both PS and PMMA it seems that the effect is strongest when the mobile-phase composition is close to φ_{crit}. Clearly, the observed effects depend on the analyte polymer, which is a strong indicator that interactions with the stationary phase (Explanation *iii*) account for the largest changes in elution volume, with changes in C_{18} layer thickness (Explanation *iv*) and polymer hydrodynamic volume (Explanation *i*) playing secondary roles. The same conclusions can be drawn from experiments with PMMA on the columns used in **Figures 4.1-E/F/G** (see supplementary information, Figure S-2, section S-2).

If the elution volume increases because of retention, then it should be possible to estimate φ_{crit} by fitting a retention model to the elution volume *vs.* φ data for the analyte polymers of various molecular weights. According to the LSS model, the changes in elution volume can be fitted to an exponential equation for analytes that show reasonably large changes with φ (*e.g.* those that elute in the linear part of the calibration curve). The intersection point of the fitted lines for analytes of different

molar mass should correspond to a reasonable estimate for φ_{crit} [33,34,37]. This approach was performed for the data in **Figures 4.1-A/B/C**. However, $t_R < t_0$, k is negative (and hence $\ln k$ undefined) for our data. In that case, it is not possible to use the LSS model. Instead, we chose to fit a different equation ($V_e = k_0 e^{-S\varphi} - c$). Here, k_0 and S still correspond to k at $\varphi = 0$ and the change in k with φ, respectively. The third parameter, c, accounts for a reduction in V_0 with molecular weight, due to the limited accessibility of the pore volume. Fitting this equation only gave a reasonable estimate for $\varphi_{\mathrm{crit,PS}}$ (approximately 50%) on the XBridge column (**Figure 4.1-B**). This may be because the analyte's interaction with the column is (likely) very weak when the mobile phase composition is stronger than φ_{crit}. The effect of a change in analyte hydrodynamic volume with φ is then also relatively enhanced. Both effects will complicate the extrapolation that is required to determine φ_{crit}, implying that measurements must be performed close to the critical composition to yield good estimates of φ_{crit}, which means such predictions are not very useful.

Similar experiments, as those underlying **Figure 4.1,** were also performed at a temperature of 60 °C for some of the columns (supplementary information, Figure S-3, section S-2). Compared to 25 °C, a shift towards a lower elution volume was observed for all analytes (including the t_0 marker, toluene) and columns. Based on the thermal expansion coefficients of ACN and THF this shift is likely a result of a thermal expansion of the mobile phase (from the pump temperature to the column temperature) and the concomitant increase in velocity. Apart from this general shift, at higher temperatures the variation of the elution volume with mobile-phase composition is seen to be smaller. This is consistent with the above explanation, as a higher temperature generally leads to a reduction in enthalpic interactions. Moreover, in the present system an increased temperature leads to a shift of φ_{crit} to lower φ, implying that the experiments across the same range of φ are performed further from φ_{crit}, and, hence, the increase in elution volume will be smaller at higher temperatures (supplementary information, Figure S-4, section S-2).

One aspect that has not been discussed is the diminishing effect of the solvent composition for analyte polymers that approach the exclusion limit of the column (see for example **Figure 4.1-A**). Two competing effects determine the overall variation observed. Larger molecules are expected to exhibit greater interaction with the column, because in adsorption or partition LC retention increases exponentially

with molecular weight [33,34]. The opposing effect is a reduction in available surface area for larger polymers, which depends on the ratio of the size of the polymer in solution (hydrodynamic radius) and the pore-size distribution of the packing material. The result of the two competing effects may be that the effect of mobile phase composition is largest for analytes eluting within the most-selective range of the calibration curve. When comparing different columns, it is seen that an increase in pore size (see for example **Figure 4.1-D**) results in an increase in the effect of φ on the high-molecular-weight analytes, as these are no longer fully excluded. For low-molecular-weight polymers ($M_w \approx 10^3$ Da) the effect of composition is much greater in **Figure 4.1-A** than in **Figure 4.1-D**. Such analytes fully permeate all pores in the latter case. However, the total available surface area is much smaller on the 4000-Å column used to record **Figure 4.1-D** than on the 60-Å column used to record **Figure 4.1-A**.

Irrespective of the cause of a change in elution volume, the above results imply that **Equation 4.2** only provides an approximate value for the required gradient steepness in case of gSEC when C_{18} columns are used. Any shifts towards higher elution volumes (*i.e.* less exclusion) implies that steeper gradients than predicted by **Equation 4.2** will be required in gSEC to achieve a separation dominated by the CCD. While in principle it should be possible to estimate the required gradient steepness based on experiments where the slope of the gradient is varied, this will be challenging because the change in retention with gradient steepness is likely too small in gSEC. Hence, this was not investigated.

4.4.2. Influence of pore size in gSEC

To investigate the influence of the pore size in gSEC, the separation of a series of polystyrenes was performed using several different columns (**Figure 4.2**). Error in elution time, as determined from triplicate measurements was, in nearly all cases, below 0.002 min.

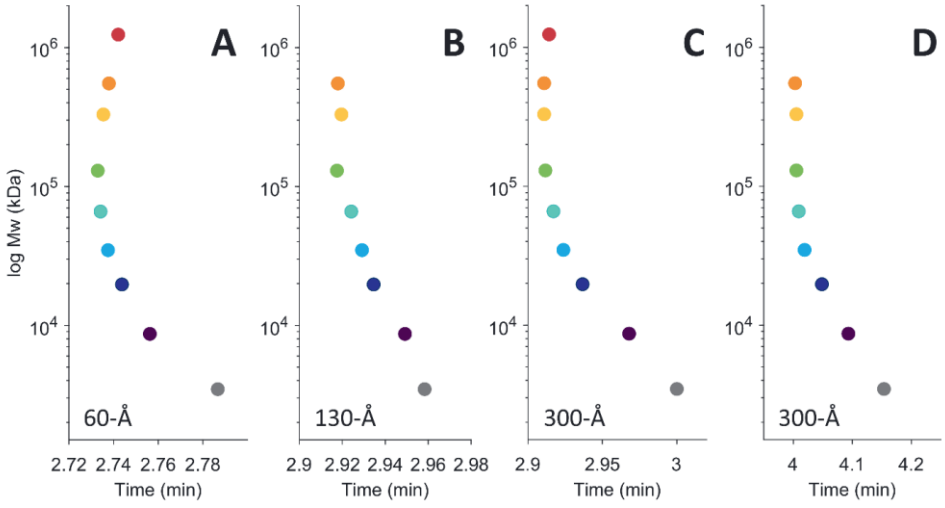

Figure 4.2: Elution time of PS standards of different molecular in gSEC on four different columns. Linear gradient from 5-60% THF in ACN; the gradient duration (t_G) and flowrate were adjusted based on the column. Injection was timed to occur at the end of the gradient. A) Two Novapak columns (150 × 3.9 mm, 60Å; t_G = 0.50 min, flow: 1.0 mL·min^{-1}, $V_G/V_0 \cong 0.30$); B) Two XBridge columns (50 × 4.6 mm, 130Å; t_G = 0.80 min, flow: 0.5 mL·min^{-1}, $V_G/V_0 \cong 0.40$); C) Dionex column (150 × 4.6 mm, 300Å; t_G = 0.65 min, flow: 0.9 mL·min^{-1}, $V_G/V_0 \cong 0.35$); D) Zorbax column (100 × 4.6 mm, 300Å; t_G = 0.65 min, flow: 0.5 mL·min^{-1}, $V_G/V_0 \cong 0.30$).

Figure 4.2 shows that for all columns the elution time still varies with analyte molecular weight. Although the elution time differences are small, these variations are consistent and significant. Under the applied conditions, it was not possible to achieve a fully molecular-weight-independent elution on any of the columns. This was quite disappointing, considering that the difference between the eluent composition at the injection point (φ_{final}) and the approximate φ_{crit}, as determined from earlier gradient-elution LC experiments of large standards, was only about 0.1. This difference corresponds to only about 5-7% of the column volume, depending on the exact column used. Apparently, to achieve molecular-weight-independent

elution, even steeper gradients would be required. On the column with the smallest pores (**Figure 4.2-A**), an inversion of the molecular-weight dependence of the elution time can be observed for the highest molecular-weight standards. Confounding SEC and interaction-LC mechanisms, or potentially confounding interaction and precipitation effects, probably prohibit genuine critical behaviour. For the different columns, a loss in recovery was also observed for specific standards (supplementary information, Table S-2, section S-3). For the 60-Å columns, low recoveries of 11% and 19% were obtained for the standards of 130 and 330 kDa, respectively. The largest standards (552 and 1270 kDa) showed higher recoveries (80-90%, or higher). On the 130-Å column, the standards of 330 and 552 kDa showed low recoveries (11% and 7%, respectively). For the 300-Å column (Dionex), the loss in recovery was significantly smaller, with 79% as the lowest recovery for the 330 kDa standard. The loss in recovery seems related to the pore size of the stationary phase particles. Possibly, as a result of small changes in hydrodynamic volume with mobile-phase composition, analytes may get trapped when their size is similar to the pore. This explanation is supported by the data and discussion of **Figure 4.1**, and will be especially important to consider for the quantitation of copolymers that feature a broad MWD. The very fast gradients applied in gSEC may give rise to differences in mobile-phase composition within and outside the pores. Small analytes have ample room to move in and out of the pores, whereas the largest analytes do not enter the pores at all. Since in gSEC the gradient must always pass through the column before the analytes, insufficient equilibration in small pores is difficult to avoid. In columns with larger pores (and smaller surface areas), column equilibration will be faster and differences in eluent composition in and outside the particles will be smaller [45]. However, such columns will diminish the extent of exclusion, and, consequently, steeper gradients will be required to attain elution independent of molecular weight.

4.4.3. Influence of gradient steepness

To further investigate the effects of the gradient steepness on elution behaviour and recovery, a series of PMMA and PS standards were subjected to different gradient volumes on the Zorbax 300-Å column. In **Figure 4.3**, the influence of gradient steepness is illustrated for PS and PMMA standards when analysed with both gSEC and RPLC.

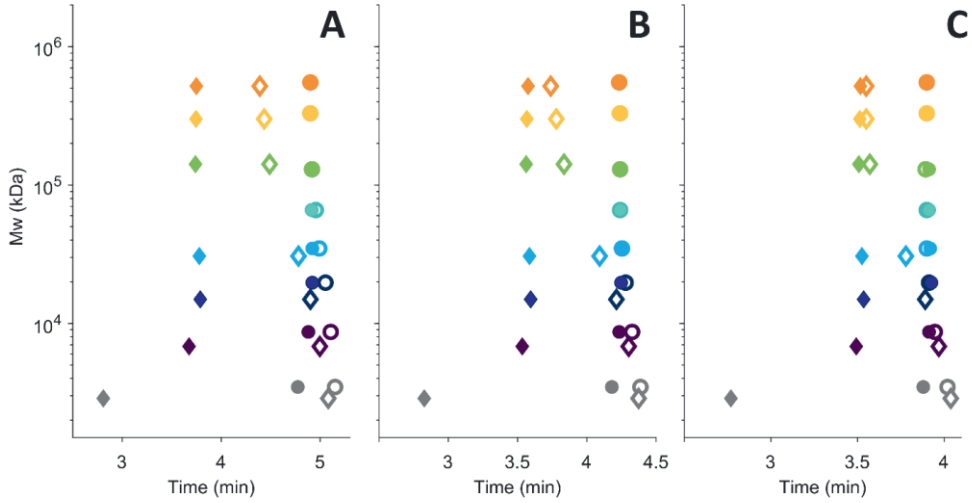

Figure 4.3: Elution time of different (known) molecular weight PS (circles) and PMMA (diamonds) standards in gSEC (open markers) and RPLC (filled markers) obtained at different gradient steepnesses using the Zorbax (4.6 × 100 mm), 300-Å column. In all cases, a linear gradient from 0-60% THF in ACN was used at a flow of 0.5 mL·min^{-1}; the gradient duration was varied as follows: A) 2.0 min ($\boldsymbol{V_G}$ = 1 mL, $\boldsymbol{V_G/V_0} \cong 0.83$), B) 1.2 min ($\boldsymbol{V_G}$ = 0.6 mL, $\boldsymbol{V_G/V_0} \cong 0.50$), and C) 0.8 min ($\boldsymbol{V_G}$ = 0.4 mL, $\boldsymbol{V_G/V_0} \cong 0.33$).

For the gSEC experiments (**Figure 4.3**, open markers), an increase in gradient steepness (moving from **Figure 4.3-A** to **C**) did result in the expected decrease in molecular-weight dependence for both the PS and PMMA standards. Note that retention times are relative to the experiment start and not relative to the injection time. However, for the PMMA standards, which must travel significantly further through the gradient to reach their critical composition (expected $\varphi_{\mathrm{crit,PMMA}} \approx 0.09 - 0.10$ *vs.* $\varphi_{\mathrm{crit,PS}} \approx 0.5$), even a SEC-gradient that occupied only 33% of the column volume (**Figure 4.3-C**) was not sufficiently steep to eliminate the molecular-weight dependence. Only the three largest PMMA standards eluted nearly

unaffected by their molecular weight, *i.e.* close to $\varphi_{\mathrm{crit,PMMA}}$. The smallest PMMA standards eluted unretained (at $t = t_0 + t_G$) in all cases. Equivalent experiments were also performed using RPLC (**Figure 4.3**, filled markers). In this case, the same gradient steepnesses are found to consistently lead to a smaller influence of the molecular weight, as compared with gSEC. We envisage three possible reasons. *i*) Fundamentally, in gSEC analytes need to catch up with the gradient front, and the main mechanism contributing to a higher migration velocity is the extent of exclusion from the pores. The largest possible difference in velocities is about a factor of two, which is only achieved in strong solvents. In contrast, in a conventional gradient experiment the difference in migration velocity between the analytes and the gradient front relies on interaction (adsorption, partition) effects, allowing for an "infinite" ratio of velocities. *ii*) Exclusion effects are strongly reduced in the vicinity of the critical composition (see **Figure 4.1** and the accompanying discussion). This implies that before the analyte polymers arrive at their appropriate position in the gradient (at $\varphi = \varphi_{\mathrm{crit}}$) their progression slows down. *iii*) For analytes for which $\varphi_{\mathrm{final}} \gg \varphi_{\mathrm{crit}}$, *e.g.* PMMA, the analytes need to cover a wide range of mobile-phase composition. The final composition of the gradient is determined by the last-eluting analytes, in this case PS. Hence, the first-eluting analytes must travel the furthest in the least amount of time.

4.4.4. 1D-LC separations of S/EA copolymers by RPLC and gSEC

An important advantage of gSEC is that injection occurs in a strong solvent and that breakthrough can be avoided. To evaluate this, one- and two-dimensional separations (SEC×RPLC and SEC×gSEC) of S/EA copolymers were performed. Representative results of the 1D-LC RPLC and gSEC experiments performed on the Zorbax 300-Å column are shown in **Figure 4.4**.

All copolymers have high molecular weights (> 70 kDa), so that molecular-weight-independent elution can be expected in RPLC gradients. For RPLC (**Figure 4.4,** upper traces in each frame), the separation based on chemical composition improves when progressively less-steep gradients are used (**Figures 4.4-A to C**). When the gradient is steep enough to allow for a copolymer to elute at their (approximate) critical composition, the chromatographic peak gives an adequate impression of the chemical-composition distribution. At this point, even steeper gradients will result in

a reduced selectivity. The optimal gradient will be based on the lowest molecular-weight fraction in the sample. Low-molecular-weight analytes will elute closer to the critical composition when steeper gradients are applied. Therefore, a relatively straightforward method to estimate the required gradient steepness for a sample that features an unknown CCD and MWD will be to perform several gradient experiments and to find the gradient steepness that minimizes the change in peak fronting.

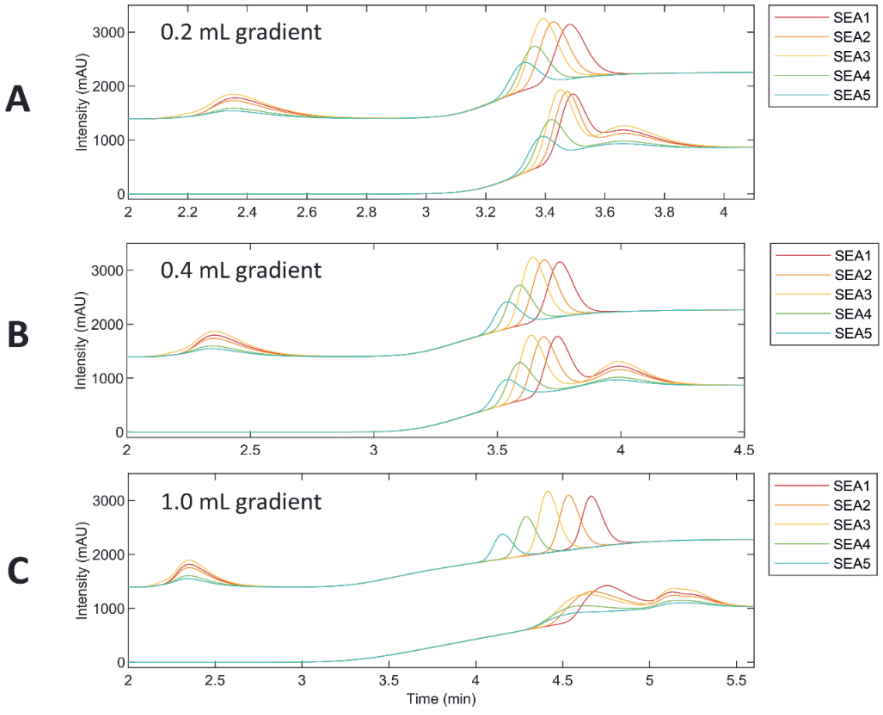

Figure 4.4: RPLC (top traces) and gSEC (bottom traces) of S/EA copolymers on the Rapid Resolution SB C_{18} column (100 × 4.6 mm, 300-Å). Approximate average copolymer compositions of S/EA1: 80/20, S/EA2: 65/35, S/EA3: 50/50, S/EA4: 35/65 and S/EA5: 20/80). Gradients from 0 to 60% THF in ACN in A) 0.4 min ($\boldsymbol{V_G}$ = 0.2 mL, $\boldsymbol{V_G/V_0} \cong 0.16$), B) 0.8 min ($\boldsymbol{V_G}$ = 0.4 mL, $\boldsymbol{V_G/V_0} \cong 0.33$), and C) 2.0 min ($\boldsymbol{V_G}$ = 0.4 mL, $\boldsymbol{V_G/V_0} \cong 0.83$); flowrate 0.5 mL·min^{-1}; UV absorbance detection at 210 nm.

For gSEC (**Figure 4.4,** bottom traces in each frame), the results are different, since in this case the analytes cannot reach their adsorption threshold before eluting from the column, so elution is dominated by molecular weight rather than chemical composition. Such a molecular-weight-dependent elution results in very broad peaks (**Figure 4.4-C**). The highest chemical-composition selectivity can be achieved with gradients that are just steep enough to suppress the molecular-weight effect on elution. For RPLC, a (much) lower gradient steepness suffices. Consequently, RPLC offers greater chemical-composition selectivity than gSEC, when using the 300-Å column. Smaller pore packings might provide improved gSEC separations but will likely lead to reduced recovery and a skewed view of the CCD. The best separation is achieved with the 1-mL gradient in RPLC (upper traces in **Figure 4.4-C)**.

4.4.5. SEC×RPLC and SEC×gSEC separations of S/MMA copolymers

One possible attractive application of gSEC is as a second-dimension separation in a 2D-LC system, where SEC is used in the first dimension using a mobile phase that is a very strong solvent in the second dimension. In case of SEC×RPLC, breakthrough may be an issue. Circumvention would require an additional modulation step, *e.g.* by dilution of the transferred fractions with weak solvent before re-injection in the second dimension. A separation of five different S/MMA copolymers, and a homopolymer PS, was performed using SEC×RPLC and SEC×gSEC. The results of these experiments are provided in **Figure 4.5.**

In both cases, a separation based on both the CCD and the MWD is achieved within approximately 35 min. To perform SEC×gSEC (**Figure 4.5-B**), the gradient was delayed to ensure that injection occurred just after the gradient. This resulted in an offset in the second dimension, which was corrected for in **Figure 4.5-B** to allow for easier comparison. After correction, the elution composition of the different copolymers and the PS standard are nearly equivalent. Some molecular-weight influence is observed in SEC×gSEC as seen from the upward curving of the elution profiles towards the right in **Figure 4.5-B**. When using RPLC as the second dimension (**Figure 4.5-A**), some breakthrough is observed around 3 min. This results in lower signals for the more-polar copolymers that contain a greater fraction of MMA (SM4 and SM5; signals with 2D retention times of about 20 and 15 min, respectively). Breakthrough is absent in SEC×gSEC (**Figure 4.5-B**). However, due to the molecular-

weight effect, SEC×gSEC is not fully orthogonal and careful calibration will be required to obtain quantitative MWD×CCD information.

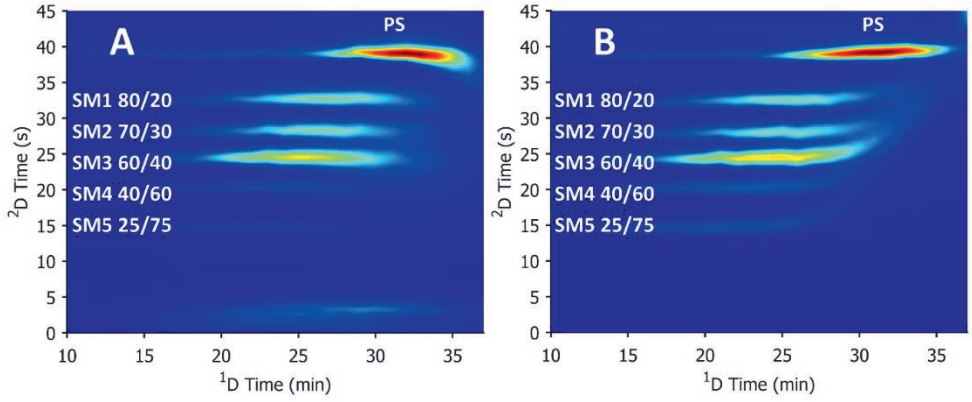

Figure 4.5: A) SEC×RPLC and B) SEC×gSEC separation of five S/MMA copolymers and a PS homopolymer. First dimension: SEC; two 150 × 2.1 mm APC SEC columns (2.5-µm particles, 450-Å pore size), mobile phase 100% THF; flowrate 15 µL·min^{-1}. Second dimension: A) RPLC or B) gSEC; two 50 × 4.6 mm XBridge BEH Shield RP18 XP columns (3.5-µm particles, 130-Å pore size), flowrate 0.9 mL·min^{-1}.

4.5. Conclusion

In this work, the applicability of gSEC is investigated for the analysis of the CCD, as an alternative to conventional gradient-elution RPLC. It was shown that gSEC can be advantageous, as it is not susceptible to the breakthrough phenomenon commonly observed in RPLC. For both gradient-elution RPLC and gSEC the application of steep gradients resulted in a reduced influence of the MWD on the separation. Hence, a better impression of the CCD of copolymers could be obtained in such gradients. However, molecular-weight-independent elution was shown to be much more challenging to achieve in gSEC. Because the difference in migration velocity in gSEC is restricted to approximately a factor of two (*i.e.* total exclusion *vs.* total permeation for very large and very small analytes, respectively), it takes long for (relatively) low-molecular-weight analytes to reach their final position in the gradient, which is around their critical composition. This problem is aggravated by a limited choice of columns, as the small pore-size packings that should ideally be used resulted in reduced recovery. This is important to consider for the quantitation of polymers that feature broad molecular-weight distributions since the reduction in recovery seemed

to depend on the pore size of the packing relative to the hydrodynamic volume of the polymer. Molecular-weight information and chemical-composition information are likely to be confounded in gSEC. We also implemented gSEC as a second-dimension separation technique for comprehensive LC×LC characterization of polymers. Comprehensive two-dimensional distributions (MWD×CCD) could be obtained by SEC×RPLC, as well as by SEC×gSEC. In the latter case breakthrough in the second dimension was avoided. However, in SEC×gSEC the residual molecular-weight dependence complicates quantitative analysis.

Supplementary material

References

[1] A.M. Striegel, Multiple Detection in Size-Exclusion Chromatography of Macromolecules, Anal Chem. 77 (2005). https://doi.org/10.1021/ac053345e.

[2] W.C. Knol, T. Gruendling, P.J. Schoenmakers, B.W.J. Pirok, R.A.H. Peters, Co-Polymer sequence determination over the molar mass distribution by size-exclusion chromatography combined with pyrolysis - gas chromatography, J Chromatogr A. 1670 (2022) 462973. https://doi.org/10.1016/j.chroma.2022.462973.

[3] H.W. Siesler, Infrared Spectroscopy of Polymers, Appl Spectrosc Rev. 11 (1976). https://doi.org/10.1080/05704927608081704.

[4] J.F. Kennedy, A.D. Suett, NMR spectroscopy of polymers, Carbohydr Polym. 27 (1995). https://doi.org/10.1016/0144-8617(95)90065-9.

[5] T. Górecki, J. Poerschmann, In-column pyrolysis: A new approach to an old problem, Anal Chem. 73 (2001) 2012–2017. https://doi.org/10.1021/ac000913b.

[6] E. Kaal, H.G. Janssen, Extending the molecular application range of gas chromatography, J Chromatogr A. 1184 (2008) 43–60. https://doi.org/10.1016/j.chroma.2007.11.114.

[7] P. Schoenmakers, P. Aarnoutse, Multi-dimensional separations of polymers, Anal Chem. 86 (2014) 6172–6179. https://doi.org/10.1021/ac301162b.

[8] H.J.A. Philipsen, Determination of chemical composition distributions in synthetic polymers, J Chromatogr A. 1037 (2004) 329–350. https://doi.org/10.1016/j.chroma.2003.12.047.

[9] G. Glöckner, Gradient HPLC of Copolymers and Chromatographic Cross-Fractionation, 1991. https://doi.org/10.1007/978-3-642-75799-0.

[10] G. Glöckner, H.G. Barth, Use of high-performance liquid chromatography for the characterization of synthetic copolymers, J Chromatogr A. 499 (1990). https://doi.org/10.1016/S0021-9673(00)97009-2.

[11] G. Glöckner, J.H.M. van den Berg, Copolymer fractionation by gradient high-performance liquid chromatography, J Chromatogr A. 384 (1987). https://doi.org/10.1016/S0021-9673(01)94665-5.

[12] G. Glöckner, High-performance precipitation liquid chromatography, Trends in Analytical Chemistry. 4 (1985). https://doi.org/10.1016/0165-9936(85)87062-X.

[13] E. Uliyanchenko, Size-exclusion chromatography - From high-performance to ultra-performance, Anal Bioanal Chem. 406 (2014) 6087–6094. https://doi.org/10.1007/s00216-014-8041-z.

[14] S.T. Popovici, P.J. Schoenmakers, Fast size-exclusion chromatography - Theoretical and practical considerations, J Chromatogr A. 1099 (2005). https://doi.org/10.1016/j.chroma.2005.08.071.

[15] A.M. Striegel, Method development in interaction polymer chromatography, TrAC - Trends in Analytical Chemistry. 130 (2020). https://doi.org/10.1016/j.trac.2020.115990.

[16] W. Radke, Polymer separations by liquid interaction chromatography: Principles - prospects - limitations, J Chromatogr A. 1335 (2014) 62–79. https://doi.org/10.1016/j.chroma.2013.12.010.

[17] E.F. Casassa, Remarks on Light Scattering from Polymers in Mixed Solvents: Effects of Polymer Molecular Weight, Molecular Weight Distribution, and Solvent Composition, Polym J. 3 (1972) 517–525. https://doi.org/10.1295/polymj.3.517.

[18] R. Mhatre, I.S. Krull, Determination of On-Line Differential Refractive Index and Molecular Weight via Gradient HPLC Interfaced with Low-Angle Laser Light Scattering, Ultraviolet, and Refractive Index Detection, Anal Chem. 65 (1993) 283–286. https://doi.org/10.1021/ac00051a016.

[19] A.M. Striegel, P. Sinha, Absolute molar mass determination in mixed solvents. 1. Solving for the SEC/MALS/DRI "trivial" case, Anal Chim Acta. 1053 (2019) 186–195. https://doi.org/10.1016/j.aca.2018.11.051.

[20] K.J. Bombaugh, R.N. King, A.J. Cohen, A new gradient elution chromatograph, J Chromatogr A. 43 (1969) 332–338. https://doi.org/10.1016/s0021-9673(00)99209-4.

[21] H.J.A. Philipsen, M. Oestreich, B. Klumperman, A.L. German, Characterization of low-molar-mass polymers by gradient polymer elution chromatography. III. Behaviour of crystalline polyesters under reversed-phase conditions, J Chromatogr A. 775 (1997). https://doi.org/10.1016/S0021-9673(97)00208-2.

[22] X. Jiang, A. van der Horst, P.J. Schoenmakers, Breakthrough of polymers in interactive liquid chromatography, J Chromatogr A. 982 (2002) 55–68. https://doi.org/10.1016/S0021-9673(02)01483-8.

[23] Y. Mengerink, R. Peters, M. Kerkhoff, J. Hellenbrand, H. Omloo, J. Andrien, M. Vestjens, S. van der Wal, Analysis of linear and cyclic oligomers in polyamide-6 without sample preparation by liquid chromatography using the sandwich injection method. I. Injection procedure and column stability, J Chromatogr A. 876 (2000) 37–50. https://doi.org/10.1016/S0021-9673(00)00179-5.

[24] E. Reingruber, F. Bedani, W. Buchberger, P. Schoenmakers, Alternative sample-introduction technique to avoid breakthrough in gradient-elution liquid chromatography of polymers, J Chromatogr A. 1217 (2010) 6595–6598. https://doi.org/10.1016/j.chroma.2010.07.073.

[25] D. Berek, Liquid chromatography of macromolecules under limiting conditions of desorption. 1. Principles of the method, Macromolecules. 31 (1998) 8517–8521. https://doi.org/10.1021/ma980533g.

[26] D. Berek, Separation of parent homopolymers from diblock copolymers by liquid chromatography under limiting conditions desorption, 1 - Principle of the method, Macromol Chem Phys. 209 (2008) 695–706. https://doi.org/10.1002/macp.200700540.

[27] M. Schollenberger, W. Radke, SEC-Gradients, an alternative approach to polymer gradient chromatography: 1. Proof of the concept, Polymer (Guildf). 52 (2011) 3259–3262. https://doi.org/10.1016/j.polymer.2011.05.047.

[28] M. Schollenberger, W. Radke, Size exclusion chromatography-gradients, an alternative approach to polymer gradient chromatography: 2. Separation of poly(meth)acrylates using a size exclusion chromatography-solvent/non-solvent gradient, J Chromatogr A. 1218 (2011) 7827–7831. https://doi.org/10.1016/j.chroma.2011.08.090.

[29] H. Maier, F. Malz, W. Radke, Characterization of the Chemical Composition Distribution of Poly(n-butyl acrylate-stat-acrylic acid)s, Macromol Chem Phys. 216 (2015) 228–234. https://doi.org/10.1002/macp.201400399.

[30] H. Maier, F. Malz, G. Reinhold, W. Radke, SEC gradients: An alternative approach to polymer gradient chromatography. Separation of poly(methyl methacrylate-stat-methacrylic acid) by chemical composition, Macromolecules. 46 (2013) 1119–1123. https://doi.org/10.1021/ma3023553.

[31] G. Glöckner, D. Wolf, H. Engelhardt, Separation of copoly(styrene/acrylonitrile) samples according to composition under reversed phase conditions, Chromatographia. 32 (1991) 107–112. https://doi.org/10.1007/BF02325011.

[32] L.R. Snyder, M.A. Stadalius, M.A. Quarry, L.R. Snyder, Gradient Elution in Reversed-Phase HPLC Separation of Macromolecules, Anal Chem. 55 (1983) 1412A-1430A. https://doi.org/10.1021/ac00264a001.

[33] P. Schoenmakers, F. Fitzpatrick, R. Grothey, Predicting the behaviour of polydisperse polymers in liquid chromatography under isocratic and gradient conditions, J Chromatogr A. 965 (2002) 93–107. https://doi.org/10.1016/S0021-9673(01)01322-X.

[34] F. Fitzpatrick, R. Edam, P. Schoenmakers, Application of the reversed-phase liquid chromatographic model to describe the retention behaviour of polydisperse macromolecules in gradient and isocratic liquid chromatography, J Chromatogr A. 988 (2003) 53–67. https://doi.org/10.1016/S0021-9673(02)02050-2.

[35] L.R. Snyder, Linear elution adsorption chromatography. VII. gradient elution theory, J Chromatogr A. 13 (1964) 415–434. https://doi.org/10.1016/s0021-9673(01)95138-6.

[36] L.R. Snyder, J.W. Dolan, High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model, 2006. https://doi.org/10.1002/0470055529.

[37] L.E. Niezen, B.B.P. Staal, C. Lang, H.J.A. Philipsen, B.W.J. Pirok, G.W. Somsen, P.J. Schoenmakers, Recycling gradient-elution liquid chromatography for the analysis of chemical-composition distributions of polymers, J Chromatogr A. 1679 (2022) 463386. https://doi.org/10.1016/J.CHROMA.2022.463386.

[38] M. van Hulst, A. van der Horst, W.T. Kok, P.J. Schoenmakers, Comprehensive 2-D chromatography of random and block methacrylate copolymers, J Sep Sci. 33 (2010) 1414–1420. https://doi.org/10.1002/jssc.200900737.

[39] X. Jiang, A. van der Horst, V. Lima, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography for the characterization of functional acrylate polymers, J Chromatogr A. 1076 (2005) 51–61. https://doi.org/10.1016/j.chroma.2005.03.135.

[40] S.M. Graef, A.J.P. van Zyl, R.D. Sanderson, B. Klumperman, H. Pasch, Use of gradient, critical, and two-dimensional chromatography in the analysis of styrene- and methyl methacrylate-grafted epoxidized natural rubber, J Appl Polym Sci. 88 (2003) 2530–2538. https://doi.org/10.1002/app.12060.

[41] M.I. Malik, Critical parameters of liquid chromatography at critical conditions in context of poloxamers: Pore diameter, mobile phase composition, temperature and gradients, J Chromatogr A. 1609 (2020). https://doi.org/10.1016/j.chroma.2019.460440.

[42] Y. Brun, P. Alden, Gradient separation of polymers at critical point of adsorption, J Chromatogr A. 966 (2002) 25–40. https://doi.org/10.1016/S0021-9673(02)00705-7.

[43] Y. Brun, P. Foster, Characterization of synthetic copolymers by interaction polymer chromatography: Separation by microstructure, J Sep Sci. 33 (2010) 3501–3510. https://doi.org/10.1002/jssc.201000572.

[44] A.M. Caltabiano, J.P. Foley, A.M. Striegel, Organic solvent modifier and temperature effects in non-aqueous size-exclusion chromatography on reversed-phase columns, J Chromatogr A. 1531 (2018) 83–103. https://doi.org/10.1016/j.chroma.2017.11.027.

[45] A.P. Schellinger, D.R. Stoll, P.W. Carr, High speed gradient elution reversed-phase liquid chromatography, J Chromatogr A. 1064 (2005) 143–156. https://doi.org/10.1016/j.chroma.2004.12.017.

Chapter 5

C^4D to account for system-induced gradient deformation in LC

Abstract

The time required for method development in gradient-elution liquid chromatography (LC) may be reduced by using an empirical modelling approach to describe and predict analyte retention and peak width. However, prediction accuracy is impaired by system-induced gradient deformation, which can be especially prominent for steep gradients. As the deformation is unique to each LC instrument, it needs to be corrected for if retention modelling for optimization and method transfer is to become generally applicable. Such a correction requires knowledge of the actual gradient profile. The latter has been measured using capacitively coupled "contactless" conductivity detection (C^4D), featuring a low detection volume (approximately 0.05 µL) and compatibility with very high pressures (80 MPa or more). Several different solvent gradients, from water to acetonitrile, water to methanol, and acetonitrile to tetrahydrofuran, could be measured directly without the addition of a tracer component to the mobile phase, exemplifying the universal nature of the approach. Gradient profiles were found to be unique for each solvent combination, flowrate, and gradient duration. The profiles could be accurately described by convoluting the programmed gradient with a weighted sum of two distribution functions. Knowledge of the exact profiles was used to improve the inter-system transferability of retention models for polystyrene standards.

Publication: L.E. Niezen, T.S. Bos, P.J. Schoenmakers G.W. Somsen and B.W.J. Pirok, Capacitively coupled contactless conductivity detection to account for system-induced gradient deformation in liquid chromatography, Anal Chim Acta, *submitted*

5.1. Introduction

The separation of soluble analytes is typically achieved by means of liquid chromatography (LC). For the vast majority of applications reversed-phase liquid chromatography (RPLC) is used. LC is performed either isocratically (*i.e.* constant mobile-phase composition) or, more commonly, by using gradient elution. Gradients allow for decreasing retention during the experiment and are indispensable for analysing complex samples that contain many compounds with large differences in hydrophobicity. To handle extremely complex samples the use of two-dimensional LC (2D-LC), specifically comprehensive 2D-LC (LC×LC), is becoming more common. To keep the overall analysis time within reasonable limits (*e.g.* an hour or less) the second dimension of an LC×LC separation has to be very short (typically less than one minute), and the gradient duration must be even shorter, to allow for column re-equilibration. In the most-common implementations of LC×LC, relatively high flowrates (1 mL·min^{-1}or higher) are used in the second dimension [1], but nowadays contemporary modulation techniques allow a combination of narrow columns and low flowrates to be used for fast second-dimension (^{2}D) separations [2–7].

The use of steep solvent gradients offers a plethora of benefits, but also introduces certain complications, one of which is gradient deformation [8–13]. Typically the term "gradient" is used for the programmed variation of the solvent composition with time. The actual composition will not equal the programmed value if the gradient is distorted when it is formed, or if it deforms when it travels through the chromatographic system. The former is primarily a result of the type, model, configuration, and condition of the pump that is used to create the gradient [8,10,12,13], while several mechanisms contribute to the latter. Examples include selective adsorption of mobile-phase components on parts of the system or the stationary phase in the column, or inadvertent stagnant zones in the system, for instance caused by poor connections. Especially the adsorption of the mobile phase on the stationary phase is well-described and is known to result in severe deformation, in particular when large and rapid changes in solvent polarity are imposed [9,11]. Previously, we have shown that correcting for (system-induced) gradient distortion allowed for great improvements when transferring retention time predictions between different LC systems [8].

In this work the focus is on the system-induced deformation that is introduced when the gradient is formed by the pump. Previously, it has been shown that the actual gradient profile delivered is unique for each system, and that it depends on the type of pump (*e.g.* quaternary or binary), the flowrate (F), the gradient duration (t_G), the dwell volume of the pump, and any purposeful or accidental mixing [8,12]. The system-induced gradient deformation is related to, for instance, the pump geometry, the mixing volume, the specific (mixing) properties of the solvents, and the type of mixer. The smaller the solvent volume occupied by the gradient, the more difficult it is to minimize the deformation. Knowledge of the actual gradient profile that enters the column is useful for troubleshooting purposes, but also essential for improving retention-time prediction (*e.g.* during selectivity optimization using gradient-scanning experiments), and for inter-system comparisons and method transfer [8].

The actual gradient profile can be obtained by direct or indirect measurements [13–17]. A typical example of the former is the measurement of the dwell curve, performed by adding a tracer compound (such as acetone) to one of the mobile-phase components and installing a detector (often a UV/Vis absorbance or diode-array detector) immediately after the pump [13]. Such a measurement is often used to obtain a value for the dwell time (t_{dwell}) of the system, which is related to the dwell volume (V_{dwell}) through the volumetric flowrate (F) by $V_{\mathrm{dwell}} = F t_{\mathrm{dwell}}$. However, dwell curves also contain information on gradient distortion.

Measuring accurate gradient profiles is not trivial. Solvatochromic effects can disturb the measurement and must be avoided or corrected for by accounting for shifts in the absorbance band of the tracer. Therefore, water and water with tracer are often used as the quasi gradient-forming solvents, however, this implies that viscosity and non-ideal-mixing effects are not accounted for [8,13]. If a volatile tracer, such as acetone, is used, its concentration and thus the measurement may be strongly affected by the degasser, which is incorporated in most contemporary pumps. Ideally, the deformation should be measured during genuine separation conditions (LC solvents, high pressure), as this is much-more representative of the actual distortion and accounts for possible variations in the profile with time (*e.g.* related to pump functioning) and temperature. A direct measurement of the gradient profile in principle can be performed if a (UV) detector is installed before the column. However, in this case the presence of a tracer is undesirable, as it may affect the separation

and the subsequent detection of the analytes. Moreover, the detector cell should be high-pressure resistant and installing a pre-column detector also adds to the dwell volume and the gradient distortion. For these reasons, a direct measurement of the actual gradient profile as it enters the column with a UV absorbance detector is quite cumbersome.

Alternatively, an indirect method for measuring the gradient profile may be considered [15–17]. If we for any arbitrary gradient profile can accurately calculate the elution time of analytes for which an accurate retention model and model parameters are known, the gradient profile may be computed based on retention data. However, this entails many measurements for a series of analytes that elute through the entire gradient, and it requires prior determination of accurate retention parameters for all these analytes [15–17]. Moreover, dealing with arbitrary gradient profiles mathematically involves several assumptions. Clearly, due to the large experimental effort needed, this approach is not all that feasible. Consequently, neither the direct nor the indirect approach are practical.

One method that could potentially fill this gap was described in a publication by Zhang *et al.* [14]. They presented capacitively coupled contactless conductivity detection (C^4D) as a tool to measure the gradient for troubleshooting purposes. The authors used formic acid or trifluoro acetic acid as an additive to the mobile phase to facilitate detection. C^4D has several attractive features that could improve the ease-of-use and correction capability of our previously developed approach for determining gradient deformation [8]. Firstly, the ability to measure changes in solvent composition by monitoring a bulk property of the solvent, *i.e.*, the conductivity (or the dielectric constant), instead of using tracer compounds. Secondly, the detector does not make contact with the mobile phase (hence "contactless"), but only with the outside wall of a (fused-silica) capillary. Therefore, this detector has no pressure limitation beyond those of the LC instrument, and is compatible with all solvents. Finally, the mobile-phase volume inside the detector can be very small (detection volume of about 0.05 µL in case of a capillary i.d. of 75 µm). As a result, C^4D shows potential for direct measurement of the gradient at different points in a chromatographic system during a separation without contributing to dispersion or dwell volume.

In this work, we investigate whether C^4D can be used to directly measure solvent gradients without any additives in the mobile phase. Subsequently, we extend our previously developed method to correct retention parameters for gradient deformation to the use of the organic modifiers methanol (MeOH) acetonitrile (ACN) and tetrahydrofuran (THF), and assess its applicability for improving method transfer between systems. Gradients from water to ACN, water to MeOH, and ACN to THF were directly measured using C^4D. No additives were added to the solvents. To explore the feasibility of using C^4D for quantitative monitoring of the gradient, the detector linearity was evaluated as a function of mobile phase composition at different input frequencies and voltages. A possible change in detector response with a change in pressure was also investigated, within the range of 15 to 65 MPa (at a frequency of 150 kHz and a voltage gain of 0 dB). In our previous work, gradient deformation was accurately described by convoluting the set profile with a single distribution function that characterized the response of the system. In this work, we aimed to assess whether the use of a weighted combination of two distributions could be used to account for potential differences caused by the two solvents. We also investigated whether correction of the deformed gradient allowed more accurate retention-time predictions across different systems.

5.2. Theory

5.2.1. Capacitively coupled contactless conductivity detection

Contactless conductivity detection is used to non-invasively and non-destructively measure the resistivity, or conductivity, of a solution. So-called capacitively coupled contactless conductivity detection (C^4D), has proven useful in combination with electrokinetic separations, such as capillary isotachophoresis and capillary electrophoresis [18–23]. The detector consists of two consecutive tubular electrodes, *i.e.* the actuator and sensing electrodes, through which a fused-capillary is placed concentrically (**Figure 5.1A**). Unlike conventional conductivity detectors that feature two parallel electrodes that are placed on either side of the analysed solution, C^4D measures the conductivity of a solution longitudinally (along the capillary), rather than transversally (across the capillary). This allows for a higher detection sensitivity, based on the axial distance between the electrodes. Consequently, narrow capillaries can be used with good sensitivity [23]. An alternating current (AC) in the form of a sine-wave with a frequency (f) is applied to the actuator electrode. After passing

through the solution in the capillary, the sine wave can be measured at the sensing electrode and the signal can be converted to an output voltage ($V_{\text{out}} = R_F I$) by means of a transimpedance amplifier that contains a feedback resistor with resistance (R_{F}), and a feedback capacitor with capacitance (C_{F}). The current passing through the cell (I) is determined by the solution in the capillary and the components of the detector cell. The equivalent electronic circuit for the detector cell in its simplest form is depicted in **Figure 5.1B**. ([18]; amplifier not shown).

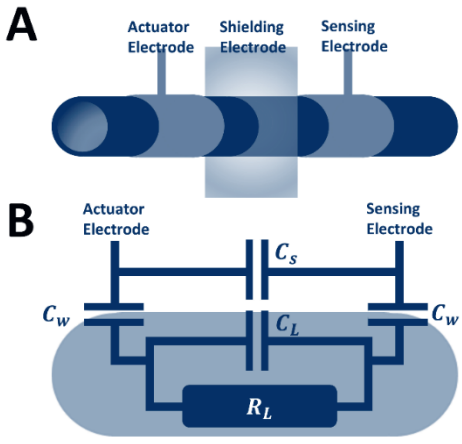

Figure 5.1: A) Schematic overview of the C^{4}D probe; B) Equivalent electronic circuit for the configuration shown in A.

As shown in **Figure 5.1** several components of the C^{4}D cell act as capacitors. The measured voltage (V_{out}) depends on the impedance (Z), rather than solely on the resistance of the solution (R_L) as $V_{\text{out}} = R_F I = R_F \frac{V_{\text{in}}}{Z_{\text{cell}}}$. Hence, the response of the detector is influenced by the capacitance of the capillary walls (C_w), a stray capacitance (C_S), and the capacitance of the solution in the capillary (C_L). In the absence of an inductor, Z is the opposition to current given by the resistance, R, and the capacitive reactance (X_C), the resistance to a change in current over time for a capacitor (C), and is given by **Equation 5.1**.

$$Z = R - iX_C = R - i\frac{1}{2\pi fC} = R + \frac{1}{i2\pi fC} \tag{5.1}$$

The measured response (V_{out}) will be a function of f. The C^{4}D cell would consequently better be called an impedance detector or, more aptly, an admittance detector, rather than a conductivity detector. The conductance (G) depends solely on R as $G = \frac{1}{R}$, while the admittance corresponds to $A = \frac{1}{Z}$. A C^{4}D cell can be presented as the simplified parallel circuit shown in **Figure 5.1-B**. For such a circuit the impedance can be calculated from the individual impedances in the circuit.

$$Z_{\text{cell}} = \frac{Z_1 Z_2}{Z_1 + Z_2} = \frac{Z_S(Z_L + Z_W)}{Z_S + (Z_L + Z_W)} \tag{5.2}$$

Where the stray impedance $Z_1 = Z_S = \frac{1}{i2\pi f C_S}$, and the impedance over the detection window $Z_2 = Z_L + Z_W = \frac{\frac{R_L}{i2\pi f C_L}}{R_L + \frac{1}{i2\pi f C_L}} + \frac{1}{i2\pi f C_W}$. Note that the capillary walls are treated here as a single capacitance. The impedance of the entire circuit is then given by **Equation 5.3**.

$$Z_{\text{cell}} = \frac{1 + R_L i2\pi f C_W + R_L i2\pi f C_L}{i2\pi f (R_L i2\pi f C_W C_S + R_L i2\pi f C_L C_S + R_L i2\pi f C_W C_L + C_S + C_W)} \tag{5.3}$$

This equation shows that the response of the C^{4}D cell depends on the capacitance of the liquid in the capillary relative to its resistance. When $R_L \gg C_L$, the C^{4}D cell is more akin to a dielectometric detector, since the capacitance (C_L) depends linearly on the dielectric constant (ε) of the solution. On the other hand, if the resistance of the solution is much smaller ($R_L \ll C_L$), then it acts as a conductivity detector. In the latter case Z_2 simplifies to $R_L + \frac{1}{i2\pi f C_W}$ and the impedance of the circuit is given by

$$Z_{\text{cell}} = \frac{1}{(C_W + C_S) i2\pi f} \frac{1 + R_L i2\pi f C_W}{1 + R_L i2\pi f \frac{C_W C_S}{C_W + C_S)}} \tag{5.4}$$

If $C_S \approx 0$ (assumed to be the case when a shielding electrode is placed between the actuator and sensing electrodes) this equation further simplifies to $Z_{\text{cell}} = \frac{1 + R_L i2\pi f C_W}{i2\pi f C_W}$.

Equation 5.4 shows that the influence of the wall capacitance increases at lower AC frequencies, while at higher frequencies the stray capacitance starts to influence the detector response. The impedance decreases at higher frequencies. Hence, the detector response increases with frequency. There is only a limited region of frequencies where the detector response does not depend on the frequency. In case of shielded electrodes this region is significantly increased.

5.2.2. Empirical retention modelling

The objective of empirical retention modelling is to describe the retention of an analyte as a function of the mobile-phase composition, pressure and/or temperature. Once a suitable model has been found, t_R can be predicted at any set of experimental conditions, without additional experiments. Because the number of initial measurements to establish the model is small, method-development time can be

drastically reduced. Software is available for this purpose, even for the case of LC×LC [24]. We briefly describe the approach here.

5.2.2.1. Linear-solvent strength model

Under isocratic conditions, *i.e.* a mobile phase composition that remains constant over the measurement duration, it is prudent to convert the retention time (t_R) to the (dimensionless) retention factor (k), which is a measure of the distribution of analyte between stationary and mobile phase. k is the ratio of the adjusted retention time (t_R') to the void time of the column (t_0).

$$k = \frac{t_R - t_0}{t_0} = \frac{t_R'}{t_0} \tag{5.5}$$

For an uncharged analyte in RPLC k decreases exponentially with an increase in the volume fraction of the strong solvent (φ) [25]. The empirical relationship between $\ln k$ and φ is approximately linear (with slope S_1), but more accurately described by including a curvature parameter (S_2) [26].

$$k(\varphi) = k_0 e^{-S_1\varphi} e^{S_2\varphi^2} \tag{5.6}$$

Where k_0 is the retention factor extrapolated to $\varphi = 0$. Because S_2 is often small, **Equation 5.6** can be suitable approximated as:

$$k(\varphi) = k_0 e^{-S\varphi} \tag{5.7}$$

Which is a log-linear model, commonly referred to as the linear-solvent strength (LSS) model [25]. Since k varies exponentially with φ, the use of gradient-elution LC is common for samples featuring analytes with large differences in retention. In case of gradient-elution LC, φ, and consequently k, change with time during the run. If it is known how φ changes with time, and an assumption is made on how the (local) retention factor changes with φ, the retention factor at the end of the column (*i.e.* the retention factor at the moment of elution, k_e) may be estimated. **Equation 5.8** can be solved to calculate t_R:

$$\frac{1}{\varphi'} \int_{\varphi_{\text{init}}}^{\varphi_e} \frac{d\varphi}{k(\varphi)} = t_0 - \frac{t_{\text{init}} + t_D}{k_{\text{init}}} \tag{5.8}$$

Where k_{init} is the retention factor at the initial conditions ($k(\varphi_{\text{init}})$). In case of a simple linear gradient, φ' is the change in the volume fraction φ ($\Delta\varphi = \varphi_{\text{final}} - \varphi_{\text{init}}$) over

the gradient duration (t_G), *i.e.* $\varphi' = \frac{\Delta\varphi}{t_G}$ for a linear gradient and $\varphi_e = (t_R - t_0 - t_D - t_{\text{init}})\frac{\Delta\varphi}{t_G} + \varphi_{\text{init}}$.

5.2.3. Describing system-induced gradient deformation

The system-induced gradient deformation can be parametrically described by means of a response function [27]. The approach is summarized in **Equation 5.9**.

$$\varphi_d = \mathbf{M}'\varphi_p \tag{5.9}$$

Where φ_d is the actual volume fraction obtained with a deformed gradient, $\mathbf{M}$ is a square matrix of uniform basis functions (and **M'** is its transpose), and φ_p is the volume fraction according the programmed gradient. In previous work the deformed gradient was found to be accurately described by using a stable distribution (**Figure 5.2A**) as the basis or response function [8]. The stable distribution is a function of four parameters $RF(\delta, \gamma, \beta, \alpha)$, which roughly correspond to the first four statistical moments of a distribution [28]. Parameter δ corresponds to the first moment or mean of the distribution. Parameter γ corresponds to the second centralized moment, which reflects the variance or width of the distribution. Parameters β and α correspond to the standardized and centralized third and fourth moment, respectively, which represent the skewness (asymmetry) and kurtosis ('tailedness') of the distribution. A part of matrix $\mathbf{M}$ (with, for clarity, a reduced number of functions), the programmed gradient (providing φ_p) and the deformed gradient (providing φ_d), are illustrated in **Figure 5.2**.

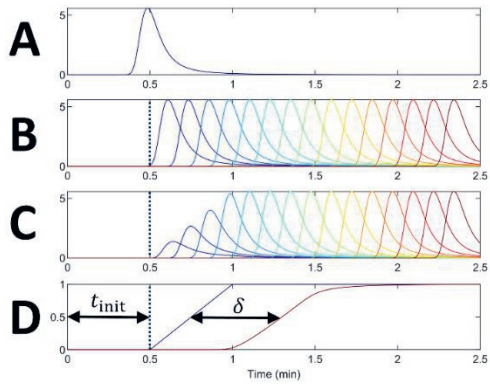

Figure 5.2: A) Stable distribution with $\boldsymbol{\delta = 0.5}$, $\boldsymbol{\gamma = 0.05}$, $\boldsymbol{\beta = 1.2}$, $\boldsymbol{\alpha = 1.2}$; B) Subset of stable distribution basis functions in matrix $\mathbf{M}$; C) Element-wise multiplication of the functions in B with the programmed gradient providing $\boldsymbol{\varphi_p}$; D) Programmed gradient (blue line) and deformed gradient (red line) obtained by **Equation 5.9**.

As shown by **Figures 5.2B** and **C**, φ_p essentially acts as a set of weights for the functions in $\mathbf{M}$. The form of the stable function and the gradient-program vector φ_p are assumed

invariable over the gradient program. Therefore, the approach cannot account for additional deformation related to particular mobile phase compositions, for instance when non-ideal functioning of the pump is correlated to the nature and/or percentage of mobile-phase components, or when solvent-specific effects of volume contraction or expansion upon mixing occur. To incorporate potential effects of the mobile-phase composition, in this work a weighted combination of two stable distributions is used, where the weights are assigned based on φ_p. This approach is summarized in **Equation 5.10**.

$$\varphi_d = \mathbf{M_1'}{\varphi_p}^2 + \mathbf{M_2'}(\varphi_p - {\varphi_p}^2) \tag{5.10}$$

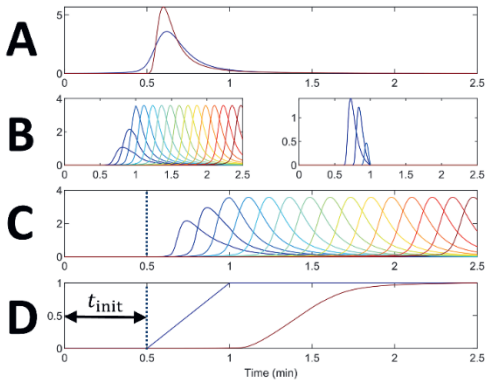

Figure 5.3: A) Stable distributions $\boldsymbol{S_1(0.64, 0.08, 0.75, 1.2)}$ and $\boldsymbol{S_2(0.62, 0.05, 1, 1)}$, B) Weighted representation of (left) $\boldsymbol{RF_1}$ and (right) $\boldsymbol{RF_2}$ with weights $\boldsymbol{\varphi_p^2}$ and $(\boldsymbol{\varphi_p - \varphi_p^2})$, respectively, C) Sum of the functions illustrated in B, D) Programmed gradient (blue line) and deformed gradient (red line) obtained by **Equation 5.10**.

Where $\mathbf{M_1'}$ and $\mathbf{M_2'}$ are both square matrices containing RF_1 and RF_2, respectively.

As illustrated, the weights are assigned such that the first response function, RF_1, still describes most of the gradient distortion, contributing across the entire gradient program. The second response function, RF_2, is expected to contribute relatively more at the beginning of the gradient. At $\varphi_p = 0.5$, RF_1 and RF_2 contribute equally, albeit also depending (to a certain extent) on the parameters of the response function. For example, for RF_2 a shift towards higher α (and larger mean) implies a larger contribution at a later point in the gradient, and *vice versa*.

5.3. Experimental

5.3.1. Instrumental

Experiments were carried out on two Agilent LC instruments (Agilent Technologies, Waldbronn, Germany). Instrument I was an Agilent 1290 Infinity II series equipped with a binary pump (G7120A) equipped with a 35-µL JetWeaver mixer, an autosampler (G7129B), a column oven (G7116B) and a diode-array detector (DAD; G7117B). Instrument II was an Agilent 1100 system equipped with a degasser (G1322A), a quaternary pump (G1311A), an autosampler (G1329A), a column oven (G1316A), and a multiple-wavelength detector (G1365B). For all measurements involving LC, an InfinityLab Poroshell 120 SB-C18 column (50 mm × 4.6 mm i.d.) containing 2.7-µm core-shell particles with 120-Å pore size was used, purchased from Agilent. For both instruments the same TraceDec conductivity detector was used, which was purchased from Innovative Sensor Technologies (Strasshof an der Nordbahn, Austria).

5.3.2. Chemicals

The eluent was prepared using deionized water (resistivity 18.2 MΩ cm; Arium 611UV, Sartorius, Germany). Non-stabilized tetrahydrofuran (THF; 99.9%, LC-MS Grade) was obtained from VWR Chemicals (Darmstadt, Germany). Acetonitrile (ACN; ≥99.9%, LC-MS Grade) was obtained from Biosolve (Valkenswaard, The Netherlands). Narrow polystyrene standards (PS1-3; polydispersity 1.05) with molecular weights of 682, 1800 and 3400 Da, respectively, were obtained from PSS Polymer Standards Service (Mainz, Germany). Test solutions of PS1, 2 and 3 were prepared in non-stabilized THF at a concentration of 1.0 $mg \cdot mL^{-1}$. Thiourea (ACS reagent grade, ≥ 99%) was obtained from Sigma-Aldrich (St. Louis, MO, United States)

5.3.3. Analytical procedures

5.3.3.1. Gradient formation and measurement

The gradient shape was measured on both LC instruments using two different approaches. For the first approach, the gradient profiles were measured without the column installed using water as solvent A, and water containing 0.001 g/L thiourea as solvent B. A total of 27 measurements was performed, namely three gradients from 0 to 100% B in 0.5, 1.0 or 1.5 min, each at three different flow rates (0.25, 0.50

and 0.75 mL·min^{-1}), all performed in triplicate. An initial isocratic delay period (100% A) of 0.25 min was used. UV absorbance detection was performed at 220 and 254 nm with a bandwidth of 4 nm. The sampling rate for Instrument I was 160 Hz, while for Instrument II it was 20 Hz. Column ovens were set to 30°C.

In the second approach, the gradient shape was measured before and during the actual LC analyses of the standards (details included below) using the TraceDec C^{4}D probe with a frequency of 150 kHz, a gain of 50% and a voltage gain of 0 dB, unless otherwise specified. Voltage gain in dB is given by $20 \log \frac{V_{out}}{V_{in}}$. All built-in smoothing was turned off. For these measurements, a fused-silica capillary was used with an inner diameter (i.d.) of 75 µm and an outer diameter (o.d.) of 375 µm. The capillary was 10 cm long and was positioned within the C^{4}D probe by sliding it into a circular slot (diameter of approximately 375 µm) in the probe. This capillary and the C^{4}D probe were present for all measurements. The C^{4}D probe was kept at ambient temperature without additional temperature control. Next to the water-to-water gradients as mentioned in the first approach, for Instrument I the following gradients were performed: water to ACN, water to MeOH, and ACN to THF. For Instrument II the evaluated gradients were water to ACN and ACN to THF.

The PS test compounds were analysed with the LC column installed. On Instrument I measurements of the gradients (ACN to THF) were performed at three flow rates (0.25, 0.50 and 0.75 mL·min^{-1}), whereas on Instrument II only 0.25 mL·min^{-1} was used due to pressure limitations. An initial isocratic delay period (100% A) of 0.25 min was followed by a gradient from 0 to 60% B in 0.5, 1.0 or 1.5 min, followed by a 5-min isocratic hold, with subsequently 100% ACN and 100% non-stabilized THF. All measurements were performed in triplicate. The injection volume was 2 µL. UV absorbance detection was performed at 220 and 254 nm with a bandwidth of 4 nm.

5.3.3.2. Data treatment

Algorithms were written in MATLAB 2022a (Mathworks, Natick, MA, USA). To determine the response functions, all measurements for each solvent combination (*i.e.* different gradient times and flow rates) were fitted simultaneously. The recorded gradient measurements were truncated to 6 and 10 min for Instruments I and II, respectively. The TraceDec conductivity detector was linearized based on calibration

with step gradients as described in section 5.4.1.1. For each solvent system, the response as a function of mobile-phase composition (φ) was described through a fourth-order polynomial with intercept zero. Based on the polynomial, the gradients were normalized between the expected signal values of the start and end composition of the programmed gradient. The normalized signals were linearized using the inversed polynomial. These linearized gradients were reduced to 500 data points to reduce computation time.

Two stable distributions were fitted as response functions to describe the deformed gradient. **Equation 5.10** was used to describe the contribution of the response functions. An interior-point algorithm was utilized for fitting. The actual shape of gradients outside the training set were predicted by multiplying the response matrices for the specified flow by the programmed gradient.

Retention times and k values were manually determined from the chromatograms. The average values from the triplicate measurements were used for fitting the LSS model. Retention-time correction accounting for gradient deformation was done using the corresponding established response functions and approximating the distorted gradient profile as a series of 500 linear steps. Corrected retention parameters were compared with those obtained conventionally, *i.e.*, correcting only for dwell volume and initial-hold volume.

5.4. Results and Discussion

5.4.1. Investigating the use of C^4D for gradient measurements

5.4.1.1. Linearity of C^4D response

The response of the C^4D as a function of the mobile-phase composition was evaluated by running a step gradient featuring increments of 10 vol% for different solvent combinations, *i.e.*, water/(water with thiourea), water/ACN, water/MeOH and ACN/THF. No additives were used to increase the solvent conductivity. The experiments were performed for different detector settings by varying detector frequency and applied voltage gain. The results are illustrated in **Figure 5.4**.

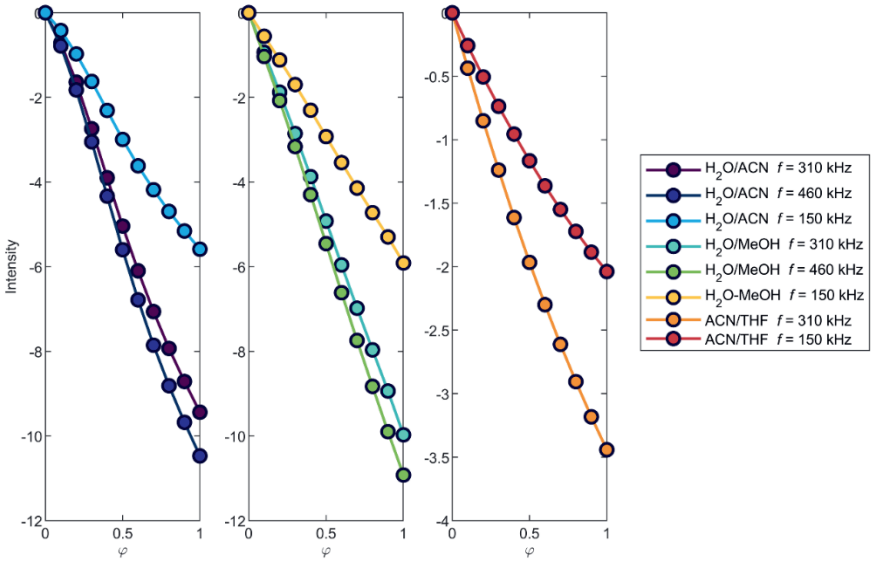

Figure 5.4: Influence of the mobile-phase composition (volume fraction φ of the strong solvent) and detector frequency on the C^4D response; Left, water/ACN; Middle, water/MeOH; Right, ACN/THF; Voltage gain, 0 dB; AC frequency *f*, 150, 310 or 460 kHz (as indicated in legend).

The detector voltage (V_{in}) proved an important variable. Irrespective of the solvent combination, no measurable signal could be obtained when the voltage gain was set lower than 0 dB (*e.g.* -24 dB, which implies a 16 times lower V_{out} than V_{in}). This suggests that the impedance Z_{cell} is relatively high compared to more-conventional applications of this detector. According to **Equation 5.3**, the absolute response at a particular mobile-phase composition is expected to increase non-linearly with the frequency and this is confirmed in **Figure 5.4**. The degree of linearity of the response *vs.* mobile-phase composition was not affected by the frequency. The curves in **Figure 5.4** are seen to be predominantly linear for water/MeOH and ACN/THF. The response was somewhat less linear for water/ACN. This non-linearity was corrected for using a fourth-order polynomial as described in the Experimental section. The detector response is expected to increase (i.e. become more negative) with decreasing solution conductivity and dielectric constant (ε). Considering the insignificant conductivity of the used solvents, the response likely depends primarily on ε. The dielectric constants of the respective solvents are $\varepsilon_{THF}(7.58) <$

$\varepsilon_{MeOH}(32.7) < \varepsilon_{ACN}(37.5) < \varepsilon_{water}(80.1)$, and the detector response roughly follows this order indeed. If it is assumed that the ε of a binary solvent mixture is described by $\varepsilon_{AB} = \varphi_A\varepsilon_A + \varphi_B\varepsilon_B$ (where φ_A and φ_B are the volumetric fractions of solvents A and B, respectively) one would certainly expect the response of the detector to vary linearly with composition. However, multiple expressions or mixing rules for ε_{AB} have been proposed, and it appeared that experimental data cannot be accurately described by a single expression for ε for all binary solvent mixtures [29]. Thus, it cannot be concluded from the present data whether the response is solely a function of ε, or whether additional effects play a role.

When a frequency of 150 kHz and a voltage gain of 0 dB were used, changes in binary solvent compositions studied in this work could be detected adequately without requiring additives. A lower frequency improves the dynamic range and increases the universality of the approach, while a higher frequency leads to an increase in sensitivity. In this case it was chosen to use 150 kHz, as this was sufficiently sensitive to measure the gradient profile. These settings were used for all further measurements. When solvents A and B were identical, *i.e.* water/(water with thiourea), no change in the C^4D signal was observed. This was expected, since thiourea will not significantly affect the dielectric constant, nor the conductivity of water. The measurements of **Figure 5.4** were repeated using a DAD at 254 nm instead of the C^4D. Due to the solvatochromic behaviour of thiourea, no combination of solvents, except for the combination of water/(water with thiourea), showed an absorbance that varied linearly with mobile-phase composition at the measured wavelength (Supplementary information, Section S1, Figure S1). While the linearity correction performed for C^4D could also enable one to use a DAD for this approach, the use of C^4D is still more practical due to the significantly lower detection volume and the resistance to high pressure or ultra-high pressure conditions.

5.4.1.2. Influence of pressure on C^4D response

To be sure we can use C^4D adequately for measuring actual gradients, we assessed whether a change in pressure caused by the solvent gradient affected the response of the C^4D cell when positioned before the column. The step gradient schedules described above were measured for water/ACN at seven different initial gradient pressures ranging from about 15 to 65 MPa. The different pressures were achieved by adapting a recently described set-up [30] (Supplementary information, Section

S2). This allowed us to increase the pressure at the inlet of the C^4D (and thus in the C^4D cell) without changing the flowrate. In these experiments no significant changes in detector response with varying pressure was observed (Supplementary information, Section S2, Figure S2). Previously, Zhang *et al.* have reported an effect of pressure on the C^4D response [14]. This may potentially be due to the presence of TFA in the solvent assessed in their study. We did not use any mobile-phase additives in our pressure experiments. The effect of pressure observed by Zhang *et al.* [14] may be related to a change in conductivity, whereas in our work the detector response mainly determines on ε, which does not significantly change with pressure.

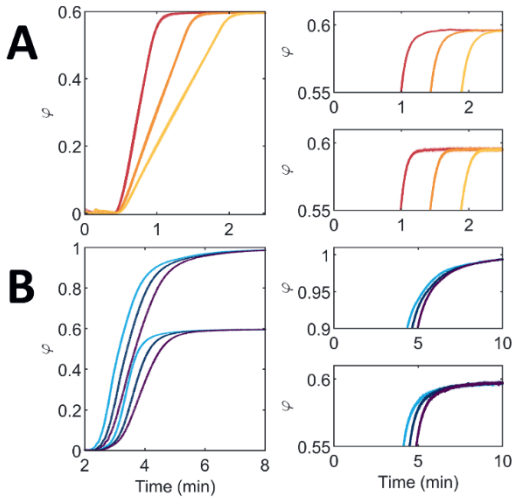

Figure 5.5: C^4D-recorded gradient profiles (φ *vs.* time) generated by A) Instrument I, and B) Instrument II. Gradient conditions: flow rate, 0.25 mL·min^{-1}; gradient time, 0.5, 1.0 and 1.5 min (respectively from red to yellow for A), Instrument I and from light blue to purple for B); solvent combinations, A) water/ACN and ACN/THF (both 0 to 60% solvent B), and B) water/ACN (0 to 100% solvent B), and ACN/THF (0 to 60% solvent B). Smaller frames at the right show expansions of the recorded profiles at the end of the gradient for water/ACN (top A and top B) and ACN/THF (bottom A and bottom B). C^4D settings: frequency, 150 kHz; voltage gain: 0 dB.

5.4.2. Effect of mobile-phase composition on system-induced gradient distortion

The gradients delivered by the LC pumping system were directly measured using the C^4D cell. To investigate whether system properties, such as mixing volume or pump geometry, affect both solvent channels (A and B) equally, different gradients were monitored with C^4D, *viz.* from water to water containing thiourea as a tracer, from water to MeOH, from water to ACN, and from ACN to THF. Ideally, differences in type of solvent should not affect the capacity of the LC system to consistently form gradients, and for the same gradient scheme all solvent combinations should yield identical gradient profiles when measured

before entering the column. The gradient profiles obtained on both LC systems for three different gradient durations using water/ACN and ACN/THF are depicted in **Figures 5.5-A** (Instrument I) and **5.5-B** (Instrument II). Right frames show vertical expansions of the profiles near the end of the gradient.

In **Figure 5.5** the gradient profiles obtained for the various solvents combinations and gradient durations are shown. It can be seen that the differences, with respect to solvents, were largest at the end of the gradient. Typically differences at the start of the gradient were minor, although if there were issues with repeatability, they caused differences at the start of the gradient (not shown). Usually such measurements were the first in a series but it is not entirely clear what caused these issues. The greatest effect seems to be the much larger dwell volume in Instrument II relative to that of Instrument I. However, a closer inspection of the right frames of **Figure 5.5** reveals that different solvent combinations also result in different gradient profiles. Generally, these differences manifested themselves as an increase in tailing (relative to a perfectly linear gradient profile). In most cases deviations for water/ACN (**Figures 5.5A and B**, top right frames) were larger than for ACN/THF (**Figures 5.5A and B**, bottom right frames), which may be related to the larger differences in polarity, viscosity, and/or compressibility of the former two solvents.

When gradients times were very short (0.5 min) and gradients were ran from 0 to 100% A to B, small distortions, which appeared as a negative dip at the end of the gradient, could be observed for water/ACN and water/MeOH (Supplementary information, section S3, Figure S3). For ACN/THF, or for 0 to 60% gradients, these distortions were not observed. Such distortions increased for higher flowrates and shorter gradients. As the detector response is linear with the solvent composition change and not affected by pressure (see above), these distortions were ascribed to incomplete mixing of the A and B solvents. Such mixing issues are expected when $\Delta\varphi$ is large and t_G is small.

5.4.3. Describing deformed gradients using response functions

As outlined in the previous sections, gradient distortions can be reliably measured with C^4D and also appear to depend on mobile-phase composition. To correctly describe gradients with deformations, a weighted combination of two response functions (details in Section 5.2.3) was applied to negotiate with distortions which

vary in magnitude along the gradient program. This approach was used to more accurately describe actual water/ACN, water/MeOH and ACN/THF gradient profiles as monitored by C^{4}D. Several measured gradient profiles, established response functions, and resulting fits with corresponding residuals, are shown in **Figure 5.6** for Instruments I and II.

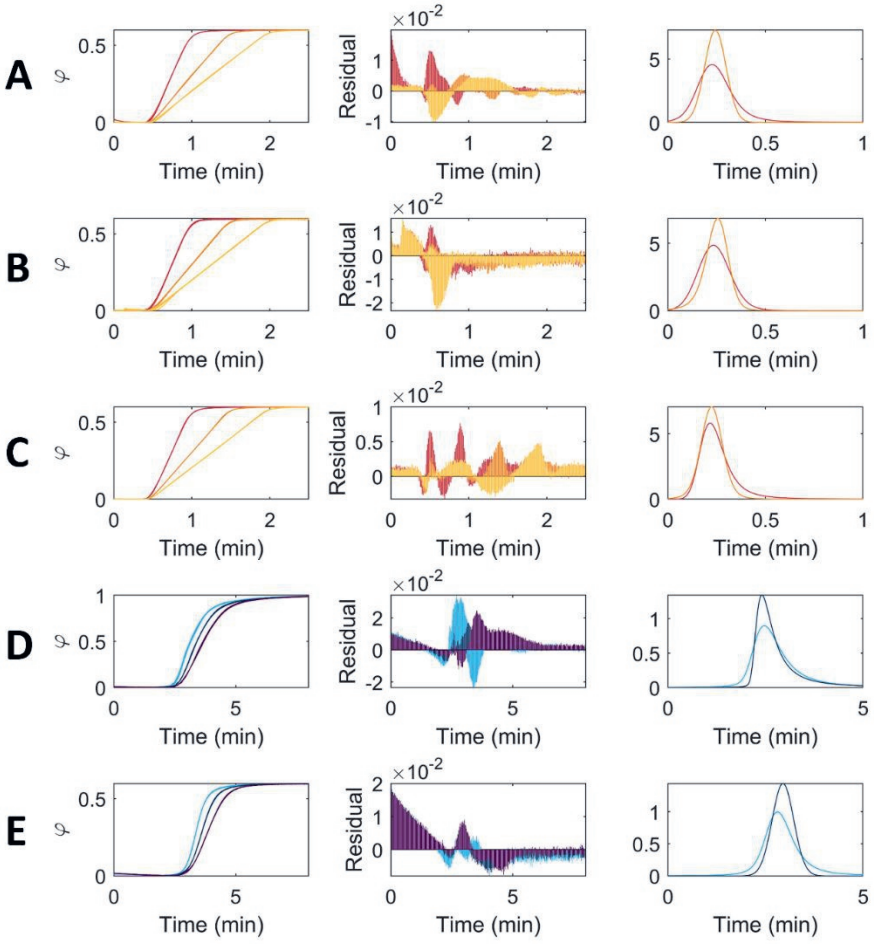

Figure 5.6: (left) C^{4}D-recorded and fitted gradient profiles, (middle) corresponding residuals, and (right) response functions $\boldsymbol{RF_1}$ and $\boldsymbol{RF_2}$ used for profile fitting by Equation 10. A-C) Instrument I; D-E) Instrument II. Gradient conditions: flow rate, 0.25 mL·min^{-1}; gradient time, 0.5 min (red and light blue traces), 1.0 min (orange and dark blue traces) and 1.5 min (yellow and purple traces); solvent combinations, A and D) water/ACN, B) water/MeOH, C and E) ACN/THF; A-C) 0 to 60% solvent B, and D-E) 0 to 100% solvent B.

The fitted gradient profiles nicely matched the recorded profiles (**Figure 5.6**, left). As indeed can be concluded from the squared residual plots (squared differences between the fit and the measurement points, **Figure 5.6**, middle), good profile fits were obtained for all solvent combinations on both LC instruments. To place the obtained fits into context we also used our previous approach. Namely, the use of a single stable distribution response function rather than two. The obtained fits (in terms of sum-squared errors, SSE) using two response functions rather than one were (average for all solvents combinations) 26.8 and 35.2% better than when using only a single response function for Instrument I and II, respectively. The full information on these fits is included in the Supplementary information, section S4, Table S1.

The overall shape of the gradient distortions is indicated by the response functions shown in the right frames of **Figure 5.6**. Especially on Instrument II, the combination of water/ACN (**Figure 5.6-D**, right) led to a more-tailing curve as compared to the response function for ACN/THF (**Figure 5.6-E**, right), which is more symmetric. Indeed, the water/ACN gradient showed more distortion at the end of the gradient as compared to ACN/THF. For Instrument I, the shape of the response functions was quite similar for all solvent combinations, indicating that for this instrument solvent-specific distortions are small. The overall distortion, *i.e.* resulting from instrument and solvents, is described by the combined response functions (RF_1 and RF_2), which complicates assignment of effects to particular parameters. However, certain trends can be observed. In most cases for a given instrument and gradient, the mean of RF_1 and RF_2 (δ) did not differ significantly. This could be expected since irrespective of the weighing, δ is predominantly determined by the dwell volume of the LC instrument. Response functions with a larger δ (*i.e.* dwell volume, as for Instrument II) also had a relatively larger variance (γ, or width). Indeed, a larger dwell volume causes more gradient deformation, as for a constant gradient volume a relatively larger portion of that gradient volume is present in the mixer, the pump head and connecting tubing, leading to dispersion. This results effectively in a decrease of the gradient slope, which is the most important parameter for building retention models from scanning gradients and use these for retention predictions. The response-function parameters which are related to skew and kurtosis (β and α, respectively) describe the function's asymmetry and are indicative for the severity of the gradient deformation at the start and end of the gradient. When $\alpha = 2$, the stable distribution

corresponds to a normal distribution and hence, has no asymmetry (β has no effect). For other cases, for example $\alpha = 1$ and $\beta = 0$ or 1 the distribution changes from a Cauchy to a Landau distribution. Finally, for $\alpha = 0.5$ and $\beta = 1$ the distribution equals a Levy distribution. In our case β and α values varied widely (Supplementary information, section S4, Table S2) and it was challenging to attribute the magnitude of these parameters to particular solvent attributes or mixing effects. To correctly interpret these parameters a significantly more in-depth study of the gradient deformation using different solvents, potentially incorporating computational fluid dynamics, is required. Since the change in asymmetry is the most significant difference from the standard correction approach for gradient deformation. A study looking at each part of the pump and how it contributes to this asymmetry might be interesting to eventually improve pump design.

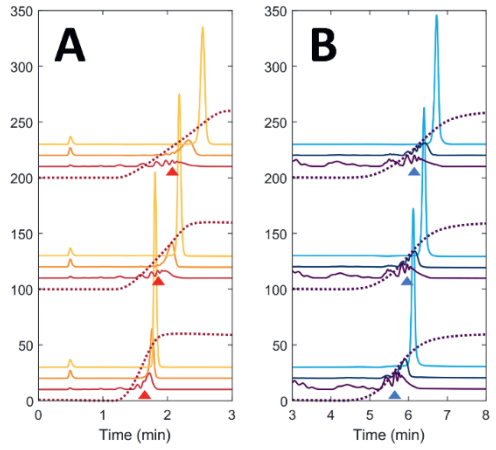

Figure 5.7: Chromatograms obtained for PS1 (red and purple traces), PS2 (orange and dark blue traces) and PS3 (yellow and light blue traces) using a (from top to bottom) 1.5, 1.0 and 0.5 min gradient on A) Instrument I, and B) Instrument II. Flow rate: 0.25 mL·min^{-1}. Signals were manually offset and scaled to aid visual comparison. Dotted lines indicate the approximate shape of the gradient (from 0 to 60% THF), obtained from blank measurements obtained using UV detection at 254 nm. Triangles indicate which oligomer peak from PS1 was used for modelling.

5.4.4. Retention modelling of polymers and method transfer

We investigated the possibility to obtain identical retention parameters from two very different LC instruments (I and II) by correcting for the instrument-specific gradient deformation. Three (scanning) gradients from 0 to 60% THF in ACN in 0.5, 1.0 and 1.5 min at a flowrate of 0.25 mL·min^{-1} were performed for three polystyrene standards of different molecular weight (M_w; 682, 1800 and 3400 Da, henceforth referred to as PS1, 2 and 3, respectively). The standards were well separated under the gradient conditions used. PS2 and 3 yielded a single peak on the RPLC column, whereas for PS1 consistently the

same peak of one of the separated oligomers was selected. The same set of gradient experiments was performed on both Instruments I and II. Since different gradient profiles were observed on these two instruments, different retention parameters will be obtained when the gradient deformation is not taken into account. The elution patterns for the three standards are shown in **Figure 5.7**.

As expected, the oligomers from PS1 (red and purple traces) elute first, partially before the gradient (*i.e.* in 100% ACN), followed by PS2 (orange and dark blue traces) and PS3 (yellow and light blue traces), which elute entirely within the gradient. An increase in dwell volume, as for Instrument II, improved the resolution for the lowest-M_w oligomers of PS1 that elute early in the gradient. In principle, if retention at φ_{init} is sufficiently large, the dwell volume should not affect the retention of an analyte. This is not the case for PS1, but may be the case for PS2 and, especially, PS3. Consequently, if gradient deformation is absent the same gradient program should result in identical retention times on the two instruments (if corrected for the dwell volume) and identical elution compositions. However, the gradient profiles measured with the C^4D cell clearly show that such is not the case. As previously described, the response-function parameter γ is significantly larger for Instrument II than for Instrument I, resulting in a shallower gradient. Shallow gradients enhance the influence of M_w in gradient separations of homologues series [31,32]. For the 0.5-min gradient (**Figures 5.7A and B**, bottom sets of traces) this is especially noticeable when comparing the retention times and peak shapes for PS2 and PS3 on both systems. For the response function of Instrument I, γ is smaller and the gradient steeper, resulting in near co-elution of the standards. This effect is expected to increase when analytes elute closer towards the end of the gradient.

Without considering the gradient deformation that occurs on Instrument II, the retention times of PS1-3 all seem to correspond to elution after the gradient (i.e. at $\varphi_e = \varphi_{\mathrm{final}}$) rather than during the gradient. However, the C^4D measurements show that the gradient deforms to such an extent that a 0.5-min gradient time actually results in a gradient duration closer to 1.5 min, which is further increased by gradient deformation. The post-column volumetric fraction of THF could be tracked as well using its absorbance at 220 or 230 nm (**Figures 5.7A and B**, dotted lines). This confirms that the gradients measured with C^4D are more indicative for the gradients entering the column than the programmed gradient. However, additional distortion

seems to occur within the column, indicating our current correction could still be improved by taking into account the distortion that occurs from the injector to the detector, for example by additionally accounting for adsorption of one or more of the mobile-phase components in the column [9,11] after first accounting for the pump-related distortions.

The retention times obtained from the measurements in **Figure 5.7** were converted to retention parameters, as done in our previous work [8]. Either a straightforward single-step linear gradient were used to describe the gradient profile (the conventional approach) or a series of 500 linear segments was used to approximate the distorted profile. The errors in prediction of the retention times using a 1.0-min gradient (using models based on data from the 0.5 and 1.5 min gradients), and the obtained LSS parameters ($\ln k_0$ and S), before and after accounting for the gradient deformation, are included in **Table 5.1**. In all cases the absolute error in measured retention times (n =3) was smaller than 0.0035 min.

Table 5.1: Predicted retention parameters ($\boldsymbol{\ln k_0}$ and $\boldsymbol{S}$) of PS1-3 obtained using models based on the measurements for 0.5 and 1.5 min gradients in Figure 6 before and after gradient correction, average prediction errors (squared differences) of the retention times ($\Delta \boldsymbol{t_e}$; n =3) obtained for the 1.0-min gradient, and percentual differences in retention parameters between Instrument I and II before and after correction; $\boldsymbol{\delta_{AB}(S) = \frac{|S_A - S_B|}{\left(\frac{S_A + S_B}{2}\right)} * 100}$ and analogous for $\boldsymbol{\ln k_0}$. Colours indicate the severity of the differences, with red being worse and green being better.

	Instrument I			Instrument II			% Differences	
Uncorrected	$\boldsymbol{\ln k_0}$	$\boldsymbol{S}$	$\boldsymbol{\Delta t_e}$ **(min)**	$\boldsymbol{\ln k_0}$	$\boldsymbol{S}$	$\boldsymbol{\Delta t_e}$ **(min)**	$\boldsymbol{\delta_{AB}(\ln k_0)}$	$\boldsymbol{\delta_{AB}(S)}$
PS1	1.70	2.97	0.027	2.23	1.34	0.009	27.0	75.4
PS2	2.48	4.44	0.017	2.74	2.75	0.001	9.9	47.1
PS3	3.69	6.47	0.009	3.27	3.79	0.010	12.2	52.2
Corrected								
PS1	1.71	3.05	0.026	2.21	2.99	0.002	25.5	2.0
PS2	2.50	4.58	0.016	2.78	4.88	0.002	10.4	6.4
PS3	3.77	6.75	0.008	3.77	6.96	0.006	0.1	3.1

Overall, the largest changes in $\ln k_0$ and S for PS1-3 before and after correction, occurred for Instrument II. For Instrument I on average much smaller changes were observed. More specifically, for Instrument II the S values of PS1-3 are significantly larger after correction. and much more in line with the values obtained with Instrument I. This can be explained considering the differences in retention times of PS1-3 observed between the different gradients. When $\ln k_0$ is reasonably large ($k_0 \geq 10$), small values for S indicate that the elution composition of the analyte may be strongly affected by the gradient slope. For the 0.5-min gradient, the selected oligomer peak of PS1 appears to elute later than 5.4 min (the sum of the instrument dwell time, the column dead time, the initial hold-time, and the gradient duration). This implies that without paying attention to gradient deformation, the peak appears to elute after the gradient, whereas with the 1.5 min gradient it appears to elute within the gradient. Thus, without correction for the actual gradient profile, the change in elution composition with gradient duration appears larger than it actually is and variations in the (programmed) gradient slope appear to have much greater effect. The actual slopes of the 0.5 and 1.5 min gradients are more comparable than the programmed slopes, since the 0.5-min gradient deforms to a greater extent than the 1.5-min gradient. When accounting for this, the same retention times would correspond to smaller differences in elution composition and a larger value for S is expected, which is also what we observe. On Instrument I the retention parameters only changed slightly, because for this system the deformation of the gradient was much smaller, and because the analytes did not elute within the region where deformation was most significant.

Interestingly, the error in the retention-time predictions did not significantly change before and after correction, and was similarly small for all compounds. Furthermore, the errors were smaller for Instrument II than for Instrument I. Since there is less deformation on Instrument I the same gradient programs/durations (0.5 and 1.5 min) correspond to actual gradient slopes that are more different on Instrument II, which are all more shallow than intended. Therefore, a greater range in retention is covered on Instrument I than on Instrument II and interpolation to the 1.0-min gradient is less accurate. One source of prediction error may be non-linearity of the $\ln k$ *vs.* φ relationship and as a consequence greater errors may result if a greater range in retention is covered (Instrument I). In addition, if non-linearity effects indeed play a

significant role, prediction errors will be larger for analytes featuring a smaller S parameter. This is indeed observed, as the error decreases in order of analyte M_w. However, Bashir *et al.* have observed an increase in prediction error with analyte M_w when using the LSS model [33], suggesting that other factors also contributed to the error. Notably, Bashir *et al.* studied much larger polymers (M_w up to 40 kDa). Likely, effects of pore size, such as differences in accessible stationary-phase surface area or differences in mobile-phase composition inside and outside of the pores, also play a role for high-molecular-weight polymers.

While the effect of gradient deformation on retention time predictions was most evident on (the older) Instrument II, some solvent-specific deformation of the gradient could still be observed on Instrument I. The effect that such small deviations have on retention (and on retention-time prediction) will depend on the specific analyte and on the gradient program. Especially at low flowrates and for complex, multi-segment gradients the effect may be significant.

5.5. Conclusion

In this work the possibility was investigated to directly measure gradient profiles using capacitively coupled contactless conductivity detection (C^4D) on-line. C^4D was capable of measuring gradients generated using various commonly-used LC solvents. These included water and the organic solvents ACN, MeOH and THF, to which no ionic tracers were added. In most cases, the detector response changed linearly with mobile-phase composition, and if needed slight non-linearities were corrected for. The C^4D response was used to accurately monitor various gradients created by two different LC instruments, one equipped with a quaternary pump and one with a binary pump. The formed gradients were seen to be uniquely distorted depending on the solvent combination. Large distortions were especially observed for gradients that encompassed an elution volume that was low in comparison with the dwell volume of the pump.

Gradient distortions could be accurately described by a weighted combination of two stable functions. This allowed us to predict the actual gradient profile for a given set of solvents as a function of the applied gradient program and the flow rate. When empirical retention modelling is used to obtain retention parameters for a given analyte and a given model, the distorted gradient profile must be taking into

account. In this work the distorted gradient profile was approximated by a segmented gradient consisting of 500 linear segments. Using this approach, retention parameters obtained when using the linear-solvent strength model on the two different systems were much-more similar than without correction for distorted gradients. This suggests that the proposed correction approach can also be used for transferring gradient-LC methods between instruments. We envision that our approach is a step towards the creation of a library of retention parameters that does not depend on the instrument used.

Our approach relied on measuring the actual gradient profile without adding extra distortion and without pressure limitations using C^4D. Our approach does not account for column-induced distortion of the gradient, for example, by selective ad- or absorption of one of the mobile-phase components on the stationary phase. Such column-induced distortions will be larger for low-volume (short duration, low flow rate) gradients. To obtain retention parameters that are completely independent of the column and of the instrument, all contributions to gradient distortion must be taken into account. Without such a rigorous correction, we found that conclusions drawn regarding which retention model yields the most-accurate description for a given set of scanning gradients are influenced by the experimental conditions, due to changes in the actual gradient slope.

Supplementary material

References

[1] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E.G. Porter, S.C. Rutan, Fast, comprehensive two-dimensional liquid chromatography, J Chromatogr A. 1168 (2007). https://doi.org/10.1016/j.chroma.2007.08.054.

[2] D.R. Stoll, P.W. Carr, Two-Dimensional Liquid Chromatography: A State of the Art Tutorial, Anal Chem. 89 (2017). https://doi.org/10.1021/acs.analchem.6b03506.

[3] B.W.J. Pirok, A.F.G. Gargano, P.J. Schoenmakers, Optimizing separations in online comprehensive two-dimensional liquid chromatography, J Sep Sci. 41 (2018). https://doi.org/10.1002/jssc.201700863.

[4] R.J. Vonk, A.F.G. Gargano, E. Davydova, H.L. Dekker, S. Eeltink, L.J. de Koning, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography with stationary-phase-assisted modulation coupled to high-resolution mass spectrometry applied to proteome analysis of saccharomyces cerevisiae, Anal Chem. 87 (2015). https://doi.org/10.1021/acs.analchem.5b00708.

[5] D.R. Stoll, K. Shoykhet, P. Petersson, S. Buckenmaier, Active Solvent Modulation: A Valve-Based Approach to Improve Separation Compatibility in Two-Dimensional Liquid Chromatography, Anal Chem. 89 (2017). https://doi.org/10.1021/acs.analchem.7b02046.

[6] L.S. Roca, A.F.G. Gargano, P.J. Schoenmakers, Development of comprehensive two-dimensional low-flow liquid-chromatography setup coupled to high-resolution mass spectrometry for shotgun proteomics, Anal Chim Acta. 1156 (2021). https://doi.org/10.1016/j.aca.2021.338349.

[7] L.E. Niezen, B.B.P. Staal, C. Lang, B.W.J. Pirok, P.J. Schoenmakers, Thermal modulation to enhance two-dimensional liquid chromatography separations of polymers, J Chromatogr A. 1653 (2021). https://doi.org/10.1016/j.chroma.2021.462429.

[8] T.S. Bos, L.E. Niezen, M.J. den Uijl, S.R.A. Molenaar, S. Lege, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Reducing the influence of geometry-induced gradient deformation in liquid chromatographic retention modelling, J Chromatogr A. 1635 (2021). https://doi.org/10.1016/j.chroma.2020.461714.

[9] F. Gritti, G. Guiochon, The distortion of gradient profiles in reversed-phase liquid chromatography, J Chromatogr A. 1340 (2014) 50–58. https://doi.org/10.1016/j.chroma.2014.03.004.

[10] G. Hendriks, J.P. Franke, D.R.A. Uges, New practical algorithm for modelling retention times in gradient reversed-phase high-performance liquid chromatography, J Chromatogr A. 1089 (2005). https://doi.org/10.1016/j.chroma.2005.07.003.

[11] F. Gritti, G. Guiochon, Separations by gradient elution: Why are steep gradient profiles distorted and what is their impact on resolution in reversed-phase liquid chromatography, J Chromatogr A. 1344 (2014). https://doi.org/10.1016/j.chroma.2014.04.010.

[12] S. Nawada, F. Gritti, Theoretical framework for mixer design for noise reduction and gradient fidelity, J Chromatogr A. 1653 (2021). https://doi.org/10.1016/j.chroma.2021.462357.

[13] M.A. Quarry, R.L. Grob, L.R. Snyder, Measurement and use of retention data from high-performance gradient elution. Correction for "non-ideal" processes originating within the column, J Chromatogr A. 285 (1984) 19–51. https://doi.org/10.1016/S0021-9673(01)87733-5.

[14] M. Zhang, A. Chen, J.J. Lu, C. Cao, S. Liu, Monitoring gradient profile on-line in micro- and nano-high performance liquid chromatography using conductivity detection, J Chromatogr A. 1460 (2016). https://doi.org/10.1016/j.chroma.2016.07.005.

[15] P.G. Boswell, J.R. Schellenberg, P.W. Carr, J.D. Cohen, A.D. Hegeman, Easy and accurate high-performance liquid chromatography retention prediction with different gradients, flow rates, and instruments by back-calculation of gradient and flow rate profiles, J Chromatogr A. 1218 (2011). https://doi.org/10.1016/j.chroma.2011.07.070.

[16] M.H. Magee, J.C. Manulik, B.B. Barnes, D. Abate-Pella, J.T. Hewitt, P.G. Boswell, "Measure Your Gradient": A new way to measure gradients in high performance liquid chromatography by mass spectrometric or absorbance detection, J Chromatogr A. 1369 (2014). https://doi.org/10.1016/j.chroma.2014.09.084.

[17] N. Wang, P.G. Boswell, Accurate prediction of retention in hydrophilic interaction chromatography by back calculation of high pressure liquid chromatography gradient profiles, J Chromatogr A. 1520 (2017). https://doi.org/10.1016/j.chroma.2017.08.050.

[18] F. Opekar, P. Tůma, K. Štulík, Contactless impedance sensors and their application to flow measurements, Sensors (Switzerland). 13 (2013). https://doi.org/10.3390/s130302786.

[19] A.J. Zemann, Capacitively coupled contactless conductivity detection in capillary electrophoresis, Electrophoresis. 24 (2003). https://doi.org/10.1002/elps.200305476.

[20] P. Kubáň, P.C. Hauser, Fundamental aspects of contactless conductivity detection for capillary electrophoresis. Part I: Frequency behavior and cell geometry, Electrophoresis. 25 (2004). https://doi.org/10.1002/elps.200406059.

[21] P. Kubáň, P.C. Hauser, Fundamental aspects of contactless conductivity detection for capillary electrophoresis. Part II: Signal-to-noise ratio and stray capacitance, Electrophoresis. 25 (2004). https://doi.org/10.1002/elps.200406060.

[22] E. Baltussen, R.M. Guijt, G. van der Steen, F. Laugere, S. Baltussen, G.W.K. van Dedem, Considerations on contactless conductivity detection in capillary electrophoresis, Electrophoresis. 23 (2002). https://doi.org/10.1002/1522-2683(200209)23:17<2888::AID-ELPS2888>3.0.CO;2-4.

[23] P. Kubáň, P.C. Hauser, Ten years of axial capacitively coupled contactless conductivity detection for CZE - A review, Electrophoresis. 30 (2009). https://doi.org/10.1002/elps.200800478.

[24] S.R.A. Molenaar, P.J. Schoenmakers, B.W.J. Pirok, Multivariate Optimization and Refinement Program for Efficient Analysis of Key Separations (MOREPEAKS), (2021). https://doi.org/doi:10.5281/zenodo.5710442.

[25] L.R. Snyder, J.W. Dolan, J.R. Gant, Gradient elution in high-performance liquid chromatography. I. Theoretical basis for reversed-phase systems, J Chromatogr A. 165 (1979). https://doi.org/10.1016/S0021-9673(00)85726-X.

[26] P.J. Schoenmakers, H.A.H. Billiet, R. Tussen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, J Chromatogr A. 149 (1978) 519–537. https://doi.org/10.1016/S0021-9673(00)81008-0.

[27] E. Tyteca, G. Desmet, A universal comparison study of chromatographic response functions, J Chromatogr A. 1361 (2014). https://doi.org/10.1016/j.chroma.2014.08.014.

[28] D.W. Morton, C.L. Young, Analysis of peak profiles using statistical moments, 210421-001518. 33 (1995). https://doi.org/10.1093/chromsci/33.9.514.

[29] B. Zhuang, G. Ramanauskaite, Z.Y. Koa, Z.G. Wang, Like dissolves like: A first-principles theory for predicting liquid miscibility and mixture dielectric constant, Sci Adv. 7 (2021). https://doi.org/10.1126/sciadv.abe7275.

[30] S. Fekete, M. Fogwill, M.A. Lauber, Pressure-Enhanced Liquid Chromatography, a Proof of Concept: Tuning Selectivity with Pressure Changes and Gradients, Anal Chem. 94 (2022) 7877–7884. https://doi.org/10.1021/acs.analchem.2c00464.

[31] F. Fitzpatrick, R. Edam, P. Schoenmakers, Application of the reversed-phase liquid chromatographic model to describe the retention behaviour of polydisperse macromolecules in gradient and isocratic liquid chromatography, J Chromatogr A. 988 (2003) 53–67. https://doi.org/10.1016/S0021-9673(02)02050-2.

[32] L.E. Niezen, B.B.P. Staal, C. Lang, H.J.A. Philipsen, B.W.J. Pirok, G.W. Somsen, P.J. Schoenmakers, Recycling gradient-elution liquid chromatography for the analysis of chemical-composition distributions of polymers, J Chromatogr A. 1679 (2022) 463386. https://doi.org/10.1016/J.CHROMA.2022.463386.

[33] M.A. Bashir, W. Radke, Comparison of retention models for polymers. 1. Poly(ethylene glycol)s, J Chromatogr A. 1131 (2006) 130–141. https://doi.org/10.1016/j.chroma.2006.07.089.

Chapter 6

Recent data pre-processing strategies in one- and two-dimensional chromatography

Abstract

The proliferation of increasingly more sophisticated analytical separation systems, often incorporating increasingly more powerful detection techniques, such as high-resolution mass spectrometry, causes an urgent need for highly efficient data-analysis and optimization strategies. This is especially true for comprehensive two-dimensional chromatography applied to the separation of very complex samples. In this contribution the latest developments in approaches for (pre-)processing are reviewed.

Publication: T.S. Bos, W.C. Knol, S.R.A. Molenaar, L.E. Niezen, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Recent applications of chemometrics in one- and two-dimensional chromatography, J Sep Sci. 43 (2020). https://doi.org/10.1002/jssc.202000011

6.1. Introduction

Analytical instruments are indispensable for modern society. To keep pace with the growing needs of society to obtain extended and reliable information on an increasing number of sample characteristics, analytical methods are continuously improved [1]. New analytical tools typically are able to generate more and more complex data, from which it is increasingly difficult to extract useful information and deduce simple and correct answers, especially when multi-component samples are analysed. To extract all valuable information from what has been referred to by some as "a tsunami of data" or, more generally, "Big Data", efficient data-analysis strategies are evidently needed [2].

One frequently applied analytical tool is chromatography, where the separation of analytes in a mixture may be obtained by exploiting differences in their partitioning between the employed stationary and mobile phases. The employed detection techniques can detect one signal as a function of time, often referred to as single-channel data, or spectrum at every point in time. This multi-channel data may facilitate identification or quantification of the analyte represented by the chromatographic signal. Although co-elution of multiple analytes upon chromatographic analysis may significantly complicate quantification and identification [3,4].

The quest for more separation power led to the development of comprehensive two-dimensional (2D) chromatography where the entire first-dimension (^{1}D) effluent is divided in many fractions, each of which is subjected to a second-dimension (^{2}D) separation [5,6]. The result is illustrated for a comprehensive two-dimensional liquid chromatography (LC×LC) separation in **Figure 6.1**, where a mixed-mode ion-exchange LC separation (A) is combined with a reversed-phase LC separation (B) leading to a 2D chromatogram (C) [7]. Qualitative information may be obtained from the position of the spots (potentially supported by data obtained from MS detection) and quantitative information from the spot intensities [8].

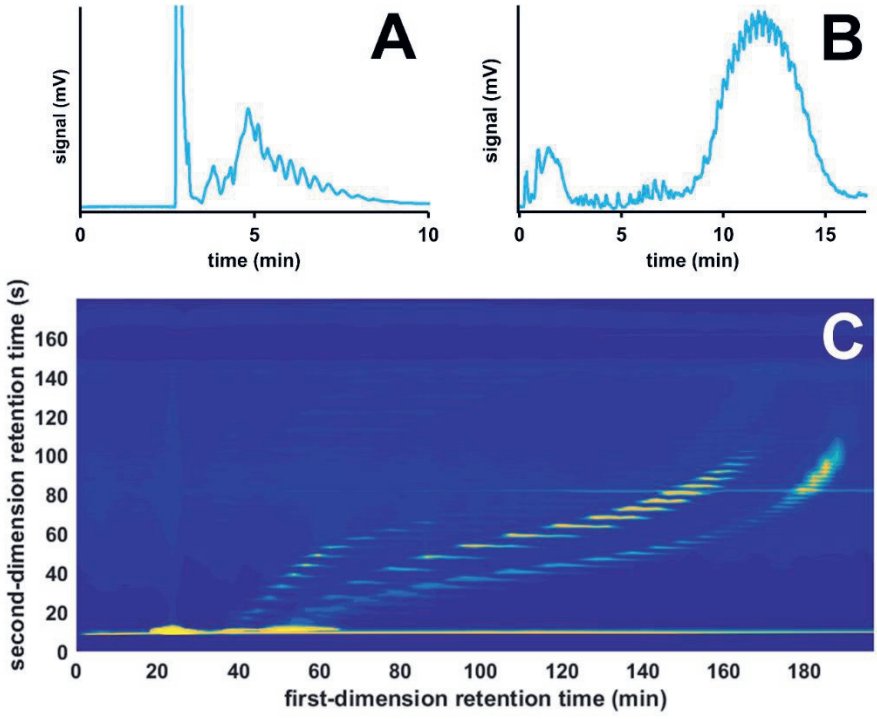

Figure 6.1: Separation of a mixture of industrial surfactants using A) mixed-mode ion-exchange LC, B) reversed-phase LC, and C) a comprehensive combination of mixed-mode ion-exchange LC and reversed-phase LC. Adapted with permission from [7].

However, when applied to highly complex samples even with 2D chromatography it can still be difficult to extract accurate and correct information from the obtained results. Indeed, samples such as copolymer formulations [9,10], food [11,12], protein digests [13,14], metabolic mixtures [15] and oil mixtures [16–18] may easily contain thousands of different components. To resolve these, powerful separation systems are needed, often equipped with sophisticated detectors such as high-resolution mass spectrometers which are able to generate huge amounts of higher-order data [19]. A large amount of information is contained in the resulting datasets, with a mass spectrum (and sometimes multiple fragmentation spectra) at each point in time in the 2D separation space. Arguably, extracting all relevant information is the biggest challenge we currently face in high-resolution chromatography. Fortunately, many

researchers are devoting their time to developing efficient chemometric data-processing strategies.

In this work, the latest pre-processing methods will be discussed, in which we address post-analysis corrections to resolve baseline drift, undesired background signals, shifting retention times, and unresolved peaks.

It is worth mentioning that, ultimately, two-dimensional chromatographic datasets comprise a collection of 1D separations. Consequently, many of the chemometric strategies used in 2D chromatography are based on the analysis of one-dimensional chromatograms.

6.2. Pre-processing

6.2.1. Aim

The main data pre-processing strategies are generally considered to be (*i*) denoising and smoothing (*ii*) baseline (drift) correction, (*iii*) retention time alignment, (*iv*) peak deconvolution and resolution enhancement, and (*v*) data compression. Steps (*i*) and (*ii*) together are generally termed "background correction" and are required for the accurate identification and, especially, quantification of analytes. This has been a long-standing issue, with the first reports having been published in the 1960's [20,21]. During the denoising and smoothing procedures, low-amplitude signals are first removed, irrespective of their frequency spectrum, after which high-frequency signals are removed, irrespective of their amplitude. Next, baseline (drift) correction can be performed, with the aim to determine the baseline shape and subtract it from the measurement. Step (*iii*), retention time alignment, is used to correct shifts in retention time that occur between experiments. This is required to compare series of chromatograms and to allow one to discern the real differences between similar samples. Peak deconvolution and resolution enhancement (*iv*) are utilized to resolve two or more (partially) overlapping signals. Finally, data compression (*v*) is generally required for large datasets to both reduce the computational resources required and to speed up data analysis.

Important to note here is that all pre-processing strategies tend to rely on assumptions or premises, which, in some cases, may lead to incorrect conclusions. A case in point is background correction, which may lead to removal of true signals.

This is likely to occur when real peaks cannot easily be distinguished from the background signal. Another example is inaccurate alignment, which may occur due to the incorrect identification of landmark peaks (or anchor points) used for the alignment. This can subsequently lead to errors during data analysis when assessing the differences between chromatograms. It should also be stressed that, while a pre-processing method may yield correct results in a specific situation, its usefulness should always be critically assessed for any other application, lest incorrect conclusions are drawn. This section reviews recent developments regarding the pre-processing of chromatographic data, with a focus on recent strategies for background correction and retention-time alignment. Where useful, less-recent methods are also briefly explained.

6.2.2. Baseline correction

As described previously, the first pre-processing step involves denoising, smoothing and baseline-drift correction to reduce baseline disturbances. In liquid chromatography (LC), noise mainly results from small fluctuations in the flow rate, the mobile-phase composition and the temperature. Drift results primarily from a variation in the mobile-phase composition (gradients). In GC electronic noise may dominate and drift arises from the variations in the flow rate and temperature-induced "bleeding" of the stationary phase. Certain derivative-based peak detection methods (see section 3.1) may struggle when such noise is present, illustrating the necessity for noise removal. In this paper well-known noise removal strategies, such as Savitsky-Golay [22] or Kalman filtering [23], are not specifically discussed. However, many of the recent background correction procedures either perform such noise removal prior to base-line drift correction or utilize subsequent peak detection methods that do not require noise removal. The baseline-drift correction is often performed by either a curve-fitting or a smoothing strategy [24]. The aim in both approaches is to fit a curve through the presumed background data points, by utilizing a loss function, such as the well-known least-squares, or by polynomial fitting [24]. Background correction methods can be roughly categorized as parametric or non-parametric. Parametric models are defined as those models that assume the background is of a certain form which can be described by a constant number of parameters, *e.g.* linear, quadratic or polynomial regression. Non-parametric methods on the other hand make no prior assumptions regarding the

shape of the baseline and allow for a flexible number of parameters, the exact number of which depends on the data. Many background correction methods are non-parametric, these include adaptive iteratively reweighted penalized least squares (airPLS), asymmetrical least squares (asLS) and corner cutting (CC) [25]. Interpolation may in some cases also be required when the actual shape of the background signal under the peaks must be determined. When a large number of peak clusters are present, baseline correction can become increasingly difficult, as the data points that contain information on the background become scarce. However, as stated above, such baseline-less data sets are becoming increasingly common with the ever-increasing complexity of the samples analytical chemists are asked to deal with. Certain techniques perform especially well in these cases (see section 2.2.5) [26].

6.2.2.1. Penalized least squares approaches

Many background correction algorithms are based around the use of penalized least squares, which is a smoothing method based on the Whittaker smoothing function [27]. Such methods include adaptive iteratively reweighted penalized least squares (airPLS), modified airPLS (MairPLS), asymmetrical least squares (asLS), asymmetrically reweighted penalized least squares (arPLS) and morphologically weighted penalized least squares (MPLS) [24,28,29].

The penalized least squares algorithm relies on balancing the fit of a model to the data, F, given by the sum of squares error (SSE), against its roughness (R) by adjusting a smoothing parameter, λ. This is given by:

$$\varphi = F + \lambda R = \sum_{i=1}^{m}(x_i - z_i)^2 + \lambda \sum_{i=2}^{m}(\Delta z_i)^2 = \|\mathrm{x} - \mathrm{z}\|^2 + \lambda \|\mathrm{Dz}\|^2 \quad (6.1)$$

Where x_i is the ith data point in the signal (x), $\mathbf{D}$ is the derivative of the identity matrix ($\mathbf{I}$), and z_i is the ith point of the fitted data, z. The difference between adjacent fitted data points is given by Δz_i Solving for $\frac{\partial \varphi}{\partial z} = 0$ returns a set of linear equations that can be solved to determine the fit, z:

$$(\mathbf{I} + \lambda \mathbf{D}'\mathbf{D})\mathrm{z} = \mathrm{x} \quad (6.2)$$

To utilize this smoothing function for baseline correction, one must first establish the location of peaks in the chromatogram. Once these peak points are known, a binary

mask or "weighted matrix" can be created, the points of which correspond to either one or zero, depending on whether the data point in the chromatogram corresponds to background or to a peak, respectively. This is the approach taken by both Cobas [30] and Zhang *et al.* [31].

$$(\mathbf{W} + \lambda \mathbf{D}'\mathbf{D})\mathrm{z} = \mathbf{W}\mathrm{x} \tag{6.3}$$

With $\mathbf{W}$ the weighted matrix or binary mask indicating the location of peaks. The disadvantage of this weighted-least-squares method is that it requires peak detection, which may in itself be affected by the correct definition of the baseline. The asymmetrical least squares (asLS) method developed by Eilers *et al.* [24] aims to solve this issue by introducing an asymmetry parameter. This parameter allows for the weights that are placed on positive and negative deviations from the baseline to be smaller and larger, respectively. However, in the case of asLS, this asymmetry parameter is constant, irrespective of the position on the baseline. For this reason, airPLS was introduced [29], which allows for certain regions of the baseline to be penalized more than other regions. In airPLS a weight vector is obtained by iteratively solving a weighted penalized least squares problem. An accurate weight vector is thought to be established once the difference between the signal and the fitted vector $|d_t|$ falls below one thousandth of the original signal.

$$|d_t| < 0.001|\mathrm{x}| \tag{6.4}$$

Both asLS and airPLS overestimate the baseline in the presence of additive noise. Therefore, the asymmetrically reweighted penalizes least squares (arPLS) approach was developed by Baek *et al.* [32]. Additional methods based around the same principles are MairPLS, in which the chromatogram (x) is pre-treated prior to performing airPLS (see section 2.2.4) [29], and MPLS, developed by Li *et al.* For MPLS a morphological strategy is used for the initial determination of the weight vector [28,33]. Background drift is ultimately accounted for by using the previously described weighted penalized least squares.

While the penalized least squares approaches are not considered computationally intensive, it should be noted that all of them require finding the correct smoothing (λ) parameter to fit the baseline. This may make these methods more time-consuming than some of the other methods.

6.2.2.2. Multivariate curve resolution and orthogonal subspace projection for background correction

Multivariate curve resolution (MCR-ALS) is one of the best-known 2-way data analysis methods. It allows recovering the number of components in a mixture, their response profiles, and their estimated concentrations [34,35]. Therefore, MCR is often applied for quantitative purposes. However, it may also be used for background correction. MCR requires the data to satisfy the condition of bilinearity. Examples of its application include LC-DAD and LC-MS data [36,37]. MCR decomposes a matrix into pure chromatographic and spectral profiles, plus noise or error**,** as in **Equation 6.5**.

$$X = CS^T + E \tag{6.5}$$

in which X represents the recorded data, and C and S the pure chromatographic and spectral profiles of the components in the sample, respectively. E is the error matrix, (ideally) containing only instrumental noise. Often initial estimates are made by singular value decomposition (SVD) [38] or PCA, but sometimes alternative methods are used [39]. Then constraints are set in place and the equation is iteratively optimized by means of alternating least-squares (ALS). The signal X does not only contain information on analytes but also on background drift:

$$X = X_{\text{analyte}} + X_{\text{background}} \tag{6.6a}$$

$$X_{\text{analyte}} = c_1 {s_1}^T + c_2 {s_2}^T \cdots c_N {s_N}^T \tag{6.6b}$$

$$X_{\text{background}} = c_{bk,1} {s_{bk,1}}^T + c_{bk,2} {s_{bk,2}}^T \cdots c_{bk,M} {s_{bk,M}}^T \tag{6.6c}$$

By considering that the spectra of the analytes s_N also contain background data, a subspace projection can be created that is orthogonal to the original data. Multiplication of the original data with this subspace will cause the background drift to be cancelled out. Which is called orthogonal subspace projection (OSP) or orthogonal spectral signal projection (OSSP). For more information regarding this technique and its use in background correction please refer to the literature [40,41].

6.2.2.3. Corner cutting with Bezier smoothing

One example of non-parametric background correction is the corner cutting (CC) method that has been developed by Liu *et al.* [25]. In CC, a smooth baseline is generated by fitting a Bezier curve [42] through the points that remain after corner points are removed from the signal vector. These corner points are defined as those points that lie above a straight line created between the previous and subsequent points in the data. This results in the automatic removal of peaks as these, by definition, will be corner points. However, a disadvantage of the approach is that it results in increasingly concave baselines as the algorithm progresses. This has been addressed by the authors by introducing a terminal condition related to the average area reduction that occurs during the iterations. The baseline is obtained after the iteration at which the average reduction in area is maximal. The approach was evaluated by comparing it to airPLS and various software packages [43–45], as well as by employing Support Vector Machines (see Section 3.3.4) classification. Since improved baseline correction should lead to better classification results, this may be one criterion to decide which method performs best. By correcting the background in Raman, X-ray diffraction (XRD), LC-MS and matrix-assisted laser desorption/ionization – time-of-flight MS (MALDI-ToF MS) data, the CC method was shown to yield the best results, without requiring additional parameters to be determined.

6.2.2.4. Local minimum value approach

Another approach to baseline correction is by utilizing the concept of local minimum values (LMVs) [46]. The approach consists of three stages, namely: *i*) initialization, *ii*) iterative optimization, and *iii*) an estimation of background drift. In the first stage a set of data points are assigned as local minimum values if the following set of conditions are satisfied:

$$x_{i-1} > x_i \tag{6.7a}$$

$$x_i < x_{i+1} \tag{6.7b}$$

In which x_i is the ith data point in the chromatogram, while x_{i-1} and x_{i+1} are the data points before and after x_i. A chromatogram with LMVs selected is illustrated in **Figure 6.2**.

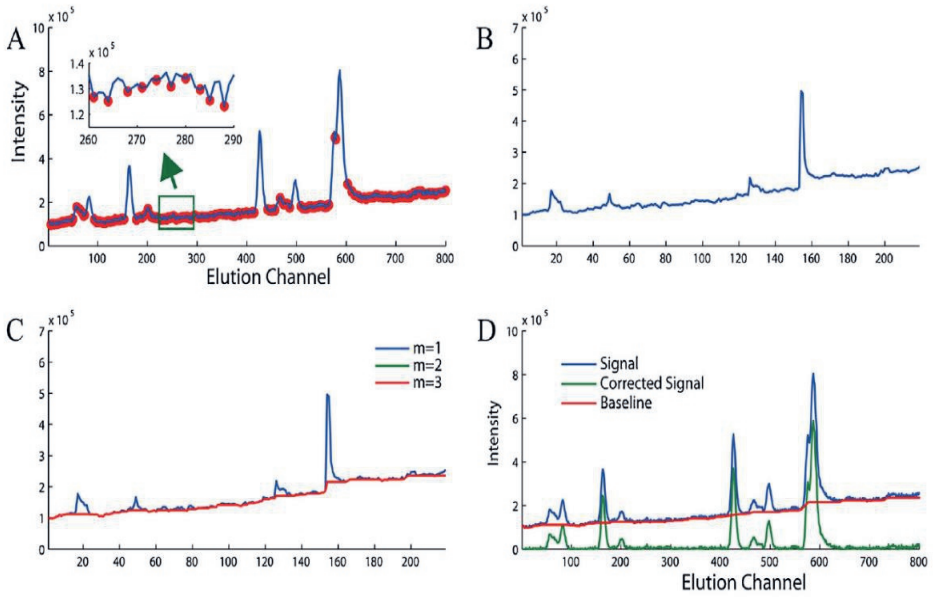

Figure 6.2: Background correction using LMVs, A) The selection of LMV's by the criteria of Eqns. 7a and 7b, B) The resulting minimum vector, C) Removal of outliers by a moving-window strategy, with m the respective iteration, and D) The original signal, the baseline and the signal corrected for background. Reproduced with permission from [46].

The complete set of LMVs is stored in a "minimum vector" and consists of chromatographic peak points and noise. This minimum vector is shown in **Figure 6.2-B**. Any of the peak points that may have been included in the minimum vector are treated as outliers and removed by utilizing a moving-window strategy. This requires an *a priori* estimation of the appropriate width of the moving window. The initial minimum vector that still contains outliers and the corrected minimum vector are shown in **Figure 6.2-B** and **6.2-C**, respectively. Any point with a signal-to-noise ratio (S/N) larger than 2.5 is considered a peak point and replaced with the median value of an extracted vector from the window in which that point occurs. This strategy is then repeated until convergence. After the iterative optimization stage, the baseline is estimated by linear interpolation. The corrected chromatogram is ultimately obtained by subtracting the estimated baseline from the original data, as is illustrated in **Figure 6.2-D**.

The LMV method was compared with morphologically-penalized-least-squares (MPLS) [28] and moving-window-minimum-value (MWMV) methods [47] using both

simulated and GC data. The simulated data consisted of both single and overlapping peaks, with the latter being comprised of contributions of two, three or four peaks. Using the simulated data, peak areas and standard deviations were determined after background correction by local minimum values-robust statistical analysis (LMV-RSA), MWMV and MPLS at different levels of noise. It was demonstrated that the LMV-RSA approach yielded the most-accurate peak areas and the lowest standard deviations, with recoveries close to 100% in all cases and standard deviations below 4.5% at all but the highest noise level. MWMV performed slightly worse, while MPLS generally resulted in significantly lower peak areas, especially in the case of overlapping peaks, with recoveries of around 53% and 74% for the peak clusters containing three and four peaks, respectively. The influence of the moving-window width (in the range between 20 and 80 data points) was found negligible for the GC data set.

Additionally, the LMV approach was compared to the "background drift correction by orthogonal subspace projection" (BD-OSP) method, which was utilized for the LC-QTOF-MS data [41]. In this case the differences were only assessed qualitatively. It was shown that after correction with BD-OSP, total-ion-current (TIC) data still contained background drift, whereas data corrected with LMV-RSA did not contain background drift but had lost part of the information contained in the TIC [44]. The comparisons showed that LMV-RSA performed comparably or better than the MPLS, MWMV and OSP approaches. However, as also stated by the authors, it is important to note that the technique can only be applied if local minimum values can be assigned.

6.2.2.5. Automatic peak detection and background drift correction

Another approach to background correction and detection is the automatic peak detection and background drift correction (ACPD-BDC) method of Yu *et al.* [48] Firstly, peak start points (x_i) and end points (x_j) were determined. A data point was defined as a start point if the following condition was satisfied:

$$x_i < x_{i+1} < x_{i+2} < x_{i+3} \tag{6.8}$$

i.e. starting position of a peak x_i must be smaller than the next three data points, x_{i+1} to x_{i+3}. Similarly, any data point is defined as an end point of the peak if the following condition is satisfied:

$$x_j > x_{j+1} > x_{j+2} > x_{j+3} \tag{6.9}$$

which similarly states that a peak's end point x_j must be larger than the next three points, x_{j+1} to x_{j+3}. While not stated explicitly by the authors it is assumed by us that in condition (8) only the first point in an increasing series is taken as a peak starting point, as this condition will lead to multiple points of increasing intensity being detected while the signal is rising, depending on peak width and detector frequency. Similarly, for **Equation 6.9** only the last point in a decreasing series should be taken as the end point of a peak. These start and end points were then contained in two vectors ($a = [a_1\ a_2\ \dots\ a_p]$ and $b = [b_1\ b_2\ \dots\ b_q]$). A combination of a starting and ending point, $[a_m\ b_n]$, was considered a peak's elution range as long as the following logical condition is met: $b_{n-1} < a_m < b_n < a_{m+1}$. All detected peaks were subsequently subtracted from the original signal, $\boldsymbol{x}$. In this way an initial estimate of the background was made (x_{new}). Threshold values were established using the first-order derivative of this initial estimate (dx_{new}) and outliers were iteratively removed by condition (**Equation 6.10**), with noise thresholds being defined as 3σ.

$$\frac{|d_i - \overline{dx_{\text{new}}}|}{\sigma} > 3 \tag{6.10}$$

In which σ is the standard deviation within dx_{new} and d_i is the ith element of dx_{new}. This condition estimates the noise level, by iteratively removing elements in dx_{new}. It is important to obtain a correct dx_{new} vector, as its first-order derivative is subsequently used as a threshold to selectively remove pseudo peaks from the original signal (x). This was carried out by evaluating the first-order and second-order derivatives of the original signal. Pseudo peaks were removed based on two conditions, *i.e.* (*i*) the absolute value of the first-order derivative of the original signal, relative to the threshold value previously established using **Equation 6.10**, and (*ii*) the number of times the second-order derivative crosses the zero-line. The authors accepted a signal as a true peak if the absolute value of the first-order derivative was five times larger than the noise threshold, and if the second-order derivative crossed the zero-line fewer than eight times. Background drift was ultimately corrected for

by first replacing the previously detected regions containing peaks by linear base lines and was denoised using three-point moving-window averaging. This resulted in a modified signal vector ($x_{\text{background}}$), which is now assumed to accurately describe the background. Baseline correction is then performed by subtracting this background from the original signal. The developed background correction procedure was then evaluated and compared to the use of airPLS [29] and MairPLS, in which the background signal ($x_{\text{background}}$) is used rather than the original chromatogram signal (x) as in airPLS. These three methods were applied for the background correction of simulated data, experimental LC data on a sample containing 11 antibiotics in tap water, and GC data on plant-based flavour extracts. MairPLS and ACPD-BDC performed similarly for all data sets evaluated, while airPLS performed considerably worse. This is illustrated in **Figure 6.3** where for each method the uncorrected and background-corrected LC chromatograms are shown.

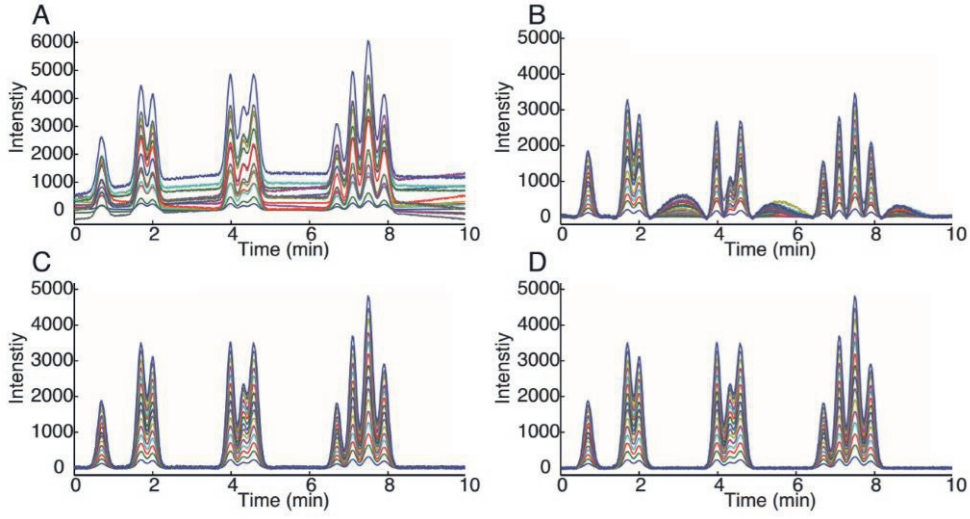

Figure 6.3: Comparison of background drift correction in 15 LC samples, containing 11 antibiotics in tap water. A) Original chromatograms, B) and C) background correction by airPLS and MairPLS, respectively (smoothing factor, $\lambda = 10^4$), D) correction by ACPD-BDC. Reproduced with permission from [48].

MairPLS and ACPD-BDC were further evaluated by means of PCA. In this study, the variance explained by the first principal component before and after background correction are used as figures of merit. For the LC data, this increased from 36.9% before background correction to 43.5% after background correction by MairPLS and to 44.4% when ACPD-BDC was used. For the GC data set almost no change was observed in the percentage of variance explained, which remained close to 95.0% in all cases.

6.2.2.6. Bayesian approaches to background correction

As previously stated, baseline correction is often hindered by crowded chromatograms and low signal-to-noise (S/N) ratios. One approach aimed specifically at facilitating baseline correction even under these conditions has been developed by Lopatka et al. [26]. In this approach, a probabilistic peak-detection algorithm is used to determine the probability of a point in the chromatogram belonging to a peak or to the baseline. It is hence termed the peak-weighted (PW) method. The algorithm operates by fitting several different models across a set window of data using a least-squares approach. Then, a likelihood is assigned to each model and from this, the probability of the data point belonging to a peak is calculated. User-defined parameters include the number of overlapping peaks allowed in each section and the window width, which directly depends on the peak width. This approach was compared to the mixture model (MM) and asymmetrical least-squares (asLS) [24,49] approaches and was shown to perform especially well in case of crowded chromatograms. This is illustrated for simulated data in **Figure 6.4**.

The PW method was also applied for background correction of a comprehensive two-dimensional GC-FID chromatogram of fire debris. However, with suitable benchmarks unavailable the authors found it impossible to objectively assess the performance of the PW method in this situation.

A different approach based on Bayesian regularized artificial neural networks (BRANN) [50] was developed by Mani-Varnosfaderani *et al.*. The iterative BRANN algorithm was compared to airPLS, MPLS, iterative polynomial fitting (iPF), and CC (see Sections 2.3.2) methods using the projected-difference-resolution (PDR) criterion.

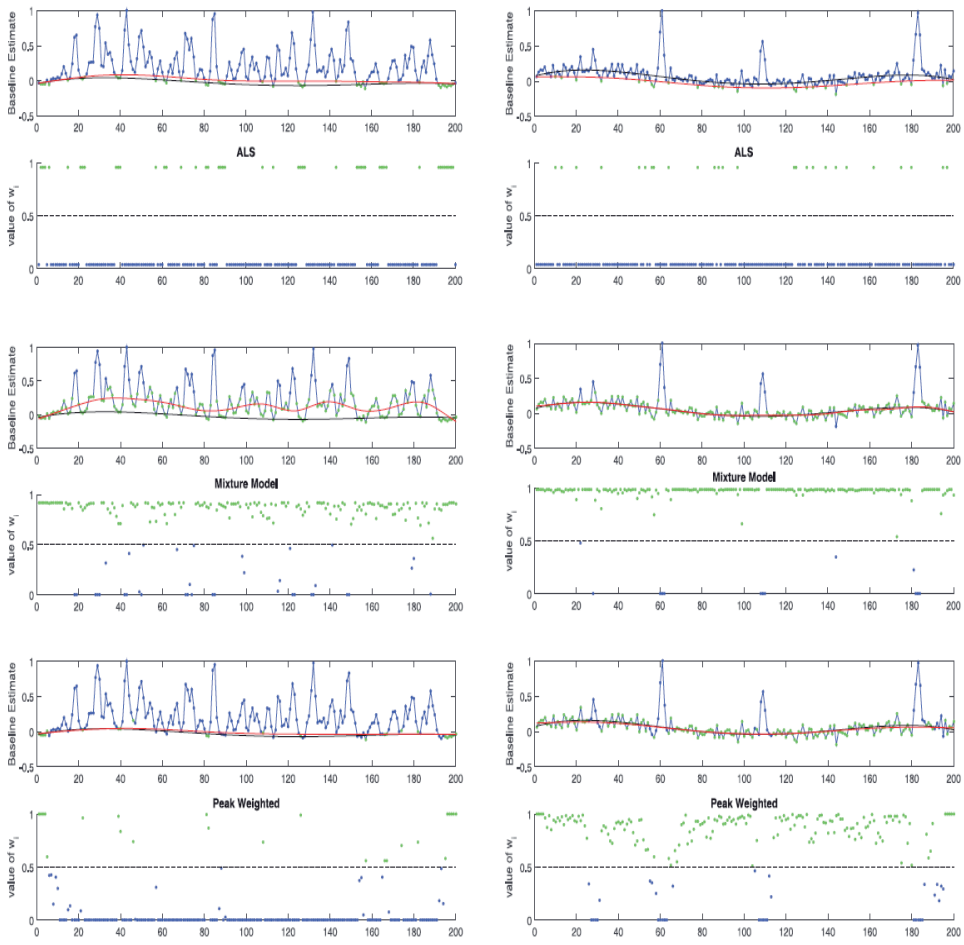

Figure 6.4: Comparison of background-drift correction by asLS, MM and PW methods for crowded (left) and sparse (right) simulated chromatograms [26]. The green points are those points that have been given high weight by the PW model and are primarily used to describe the background, while the blue points have been given low weights.

6.2.2.7. Baseline estimation and denoising using sparsity

When a signal can be described reasonably accurately using only a few non-zero parameters it can be classified as sparse. For a typical chromatogram, consisting of peaks, noise, and background, this assumption may also be applied if it features relatively few peaks compared to the number of baseline points. One algorithm that utilizes this concept of sparsity, and has been developed recently, is called baseline estimation and denoising using sparsity (BEADS) [51]. It was later further improved to create the "assisted BEADS" algorithm [52]. BEADS specifically aims to model the signal, background, and noise, without employing the use of overly restrictive parametric models. As the background is considered a low-pass signal, depending on the cut-off frequency, low-pass filters may allow this background to be removed. In mathematical terms the approach is based on modelling the chromatographic signal as:

$$x = s + w = x_p + f + w \tag{6.11}$$

With x the input data or chromatogram containing peaks x_p, baseline f, and white Gaussian noise w. Thus s describes the noiseless input chromatogram ($x_p + f$). It is assumed that in the absence of peaks the baseline can be estimated by utilizing a low-pass filter. Thus, from an estimate of the peak vector ($\hat{x}_p$) an estimate of the baseline ($\hat{f}$) can be obtained by filtering the chromatogram.

$$\hat{f} = \mathrm{L}(y - \hat{x}) \tag{6.12}$$

Once the baseline is estimated, the noiseless input chromatogram ($\hat{s}$) can also be obtained as this is simply $\hat{x} + \hat{f}$. This means $\hat{s}$ can be estimated by using both a low-pass filter L and a high-pass filter H.

$$\hat{s} = \mathrm{L}y + \mathrm{H}\hat{x} \tag{6.13}$$

The task is then to obtain an accurate estimate of the peak vector and to establish suitable filters. To achieve this, the authors investigated two different cost functions and employed an algorithm to minimize these. For a more extensive overview of the cost functions and algorithm employed, please refer to [51].

The performance of BEADS was compared to airPLS [29] and backcor [53] strategies for baseline correction of simulated and real chromatographic data. The results are illustrated in Figure 6.5, using chromatographic data from [29]

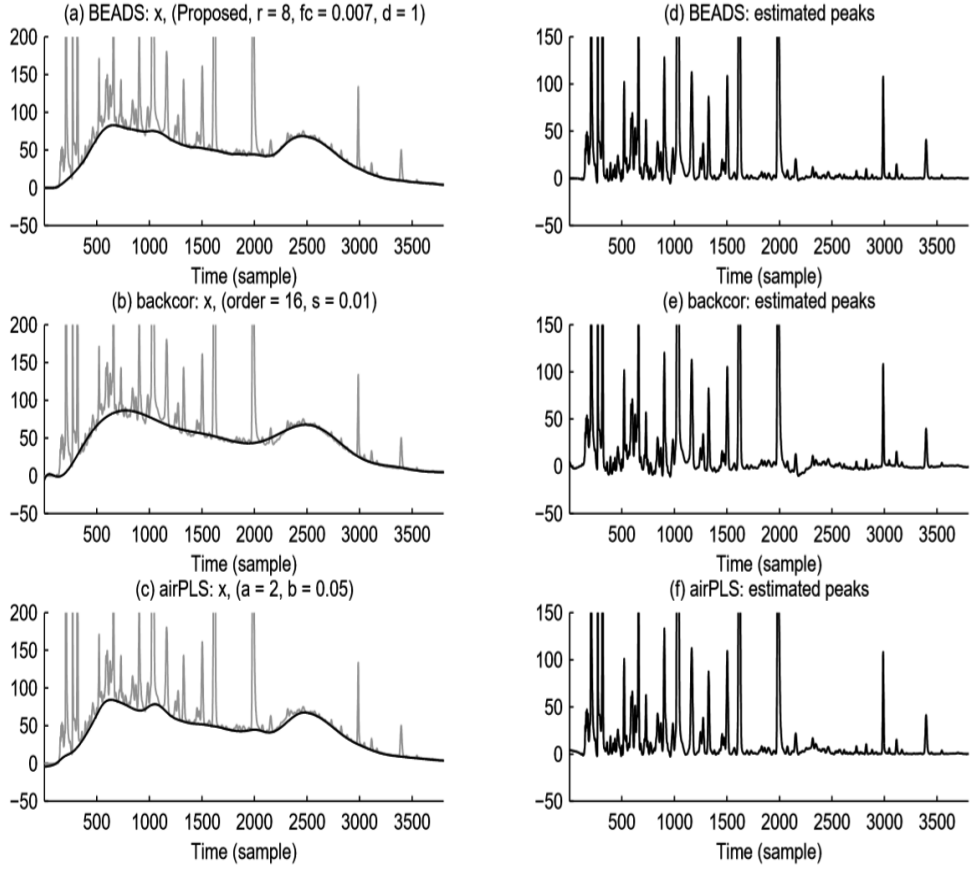

Figure 6.5: Comparison of background drift correction using BEADS (top), backcor (middle) and airPLS (bottom). Reproduced with permission from [51].

BEADS was found to have performed favourably in comparison with airPLS and backcor, with the former underestimating the baseline in the range from datapoint (sample in **Figure 5**) 2200 to 2500 and the latter overestimating the baseline in this region. However, while BEADS performed well, the baseline was required to be periodic, *i.e.* the signal at the start of the chromatogram should be equal to that at the end of the chromatogram. If the above requirement is not fulfilled, for example

due to changes in mobile-phase composition or temperature, the modelled baseline will show end-point transient artefacts. This would manifest in a decrease (or increase) of the baseline towards the starting value at the end point of the chromatogram. An additional limitation includes the need to manually adjust parameters, such as the order of the filter employed and its cut-off frequency, the penalty function utilized in the optimization and its asymmetry, and the regularization parameters, which should be set in accordance with the expected sparsity of the data. Small changes in these parameters can result in very different baselines. However, as noted by the authors, these parameters do allow the approach to be used for diverse signals, including, for example, baseline estimation in electrocardiography (ECG). Furthermore, BEADS cannot correctly handle negative signals, such as those observed in for example refractive-index detection (RID).

To summarize, the following difficulties arise when using BEADS for baseline correction: (*i*) parameter adjustment and selection (*ii*) the signal intensity for the first and last points in the chromatogram should be equal, and (*iii*) difficulties with assessing data that may contain negative peaks. Most of these limitations have been addressed by Navarro-Huerta *et al.* [52] who have developed the assisted-BEADS algorithm, and by Selesnick, who has proposed a solution for the endpoint artefacts [54]. Parameter selection may be facilitated by auxiliary autocorrelation plots. In such plots the correlation between consecutive data points is measured. By determining the autocorrelation of the noise after background correction and by plotting this as a function of one (or, ideally, all) of the adjustable parameters, the optimal value of the parameter(s) can be established from the location in the plot where autocorrelation is minimized. To address the sensitivity of BEADS to negative peaks an additional algorithm has been applied, which discards sporadic negative signals [52].

6.2.2.8. Background correction in GC-MS and LC-MS using recorded profile spectra

An MS-based approach to baseline correction and noise removal in GC-MS and LC-MS data has been developed by Erny et al. [55]. In this work, the recorded profile (full) spectra were used rather than conventional centroid mass spectra. The latter is obtained by retaining only the peak centres at discrete m/z values (*i.e.* zero-line width), and the corresponding intensity while discarding any other information.

However, it has recently been shown that errors may result from the use of centroided spectra in subsequent data analysis, the most prominent being the merging of overlapping peaks. As the number of profile spectra to be analysed was 60,000 and 141,000 for CE-ToF-MS and UHPLC-QToF-MS respectively, a selection of profile spectra was first made based on their relative length. This relative length is defined as the number of non-zero values divided by the total number of values in the MS profile. As a zero value means no ion is detected at the given time and m/z interval, the relative length is an indicator of what type of information is contained within the profile. By generating a base peak profile from a selection of profiles that differ in relative length, the information in these profiles can be visualized. Using this approach, the authors selected the profile spectra with a relative range of 75-100% as the data to use for background correction, along with profiles containing more than 50% of non-zero values. This resulted in 3,909 and 37,000 profiles for background correction in CE-ToF-MS and UHPLC-QToF-MS, respectively. The same strategy was also applied for noise estimation, using the profile spectra in the relative range from 0-25%. Both airPLS and arPLS were then investigated for baseline correction, while a moving-window strategy was employed for noise removal using the noise estimated from the base-peak profile as a threshold value. The use of a higher noise threshold was also investigated, however, this ultimately resulted in the removal of low-intensity peaks. The background correction itself, performed with arPLS, did not result in significant alterations of the total-ion profile. As a final step the spectra are converted back to conventional MS-centroid spectra. The computation time was approximately 2 and 20 mins for the CE-ToF-MS (0.7 GB) and UHPLC-Q/ToF-MS (2.9 GB) data sets, respectively. The primary difference with other approaches is that baseline correction and noise removal are primarily based on the profile spectra, which are first selected based on their relative length, so as to improve the accuracy of the correction. This allowed the authors to reliably obtain base-peak ions that were previously obscured by background ions. It also allowed for a substantial reduction in data size.

6.2.2.9. Methods for 2D chromatography

A number of research groups have specifically investigated methods for 2D chromatography. One example exploits the trait of visualizing LC×LC and GC×GC separations as 2D image. In their work, Reichenbach et al. utilized a number of

statistical and structural characteristics of the background signal in 2D chromatograms, including the white noise properties of noise in chromatographic signals to correct for the background [56]. Their algorithm has been applied to both GC×GC and LC×LC data using the GC Image and LC Image software tools [57,58]. Other approaches have consulted the data from the 1D perspective. Zeng et al. used the linear least-squares curve fitting approach combined with moving-average smoothening to correct all 1D peaks within the 2D chromatograms [59]. Zhang et al. employed alternating trilinear decomposition (ATLD) to correct the analytical signal for the background drift of LC×LC-DAD data [60]. Self-weighted alternating trilinear decomposition (SWATLD) and parallel factor analysis (PARAFAC) were also applied for this function.

6.2.3. Retention-time-alignment strategies

After the data have been corrected for the background signal alignment may be required. This is especially the case in LC, where retention-time shifts between analyses are not uncommon. This alignment is generally performed either based on integrated peak tables or on pixel-level chromatograms. In the latter case, the entire chromatogram is used for the alignment. When using integrated peak tables, peaks are aligned by assigning a unique identifier to each peak and assuming this to be consistent across all chromatograms being aligned. Therefore, such alignment strategies are often closely linked with other chemometrics methods that allow for both peak detection and tracking. The algorithms vary in complexity from simple scalar shift alignment, alignment to selected target peaks, local alignment, to globally optimized alignment, which automatically optimizes the alignment in multiple regions of the chromatogram. Some of the best-known globally optimized alignment approaches are correlation-optimized warping (COW), dynamic time warping (DTW), parametric time warping (PTW), and correlation-optimized shifting (COSHIFT) [61–63]. Many of these algorithms have been applied in various fields, such as forensic profiling and metabolic fingerprinting [64,65].

6.2.3.1. Correlation-optimized warping

In COW, the chromatogram is first divided in several local regions, which are iteratively stretched and compressed until the Pearson correlation coefficient (PCC)

between the sample and the reference chromatogram is maximized. The PCC is calculated from **Equation 6.14**.

$$\mathrm{PCC} = \frac{(r-\bar{r})^T(x-\bar{x})}{\sqrt{(r-\bar{r})^T(r-\bar{r})(x-\bar{x})^T(x-\bar{x})}} \tag{6.14}$$

in which r is a vector describing the reference chromatogram, while x is the test chromatogram. Their means are given by $\bar{r}$ and $\bar{x}$, respectively. Several input parameters are required, such as the segment length and the slack length. Adaptations to COW have also been developed, including the 2D-COW algorithm by Zhang *et al.* [66] and an alternative method by Gros *et al.* [67], which has recently been applied for alignment of GC×GC-HRMS data [68].

6.2.3.2. Automatic time-shift alignment

An additional approach to time-shift alignment, automatic time-shift alignment (ATSA), was developed by Zheng et al. [69]. This method comprises three different steps, *viz.* (*i*) automatic baseline correction and peak detection, (*ii*) preliminary alignment through adaptive segment partition, and (*iii*) a precise alignment. Baseline correction was performed by LMV-RSA (see Section 2.2.1) and peak detection was carried out by a multi-scale Gaussian smoothing-based strategy (see section 3.3.2) [70]. Then the chromatogram was divided into a number of short segments, the time shifts within which were expected to be similar. A preliminary alignment of the chromatograms was performed by first establishing a reference chromatogram. However, as noted by the authors, relying solely on maximizing PCC values can lead to misalignments, as the magnitude of the PCC value is influenced strongly by large peaks. Therefore, the preliminary alignment was performed by using the total peak correlation coefficient (TPC) instead, which is calculated from:

$$\mathrm{TPC} = \left(\frac{\sum_{i=1}^{I} w_i \mathrm{PCC}_i}{\sum_{i=1}^{I} w_i}\right)\frac{N_{\mathrm{test}}}{N_{\mathrm{ref}}} \tag{6.15}$$

in which w_i is the weight of the ith-matched peak, defined as the ratio between peak area and peak length, and N_{test} and N_{ref} are the number of peaks in the test and reference chromatograms, respectively. Peak length describes the width of the peak, but in number of data points, rather than time units. Segments that were not correctly aligned were treated as outliers and were re-aligned if they did not fall within the 99% confidence interval. For re-alignment PCC values were used and the

coefficient closest to the expected time-shift value was selected as optimal. After preliminary alignment, overlapping and disconnected segments may be present in the chromatogram. These were corrected by using a warping strategy and adjusting the boundaries between segments. To ensure that the chromatogram retains the same start and endpoints after time-alignment, a linear interpolation strategy was used. The PCC values obtained after preliminary alignment already showed significant improvement, increasing from 0.72 to 0.96.

After the preliminary alignment, the final precise alignment was carried out by first segmenting the aligned test chromatogram based on the number of chromatographic peaks present. Boundaries set in the middle between the end position of a peak and the starting position of a subsequent peak. Each segment was then aligned to the nearest reference peak based on retention time. For segments that did not contain a reference peak, the time shift was taken as the average of that of neighbouring segments. Then once again warping was used to properly align the segment boundaries, as the time-shifts caused disconnected and overlapping segments. After performing the entire retention-time alignment procedure the correlation coefficient improved further, from about 0.96 to about 0.99.

The authors then evaluated their approach. The influence of the two pre-estimated parameters, *i.e.* initial segment size and initial time shift were investigated. Several different settings were tested, and the obtained PCC values were compared. Initial segment size was varied incrementally from 1 to 10 min and was found to result in nearly constant PCC values of approximately 0.993. However, the authors noted that larger segment sizes (> 10 min) would reduce the required computing power but resulted in drastic time-shift changes. The initial time-shift estimate was varied from 0.1 to 1 min and resulted in constant PCC values. The ATSA method was also evaluated by analysing the eventual peak areas. This is especially important because a warping strategy was used, which may influence quantification. Once again, the peak areas before and after the entire alignment strategy were compared by using the obtained PCC values. The approach was shown to have a negligible effect on the determined peak area (PCC = 0.9998). However, as stated by the authors, the relative deviation increased for very small peaks. ATSA was applied in a study concerning the storage of essential oils and it was compared with COW. The experimental data suggested degradation of the essential oils during storage. However, after alignment

using either COW or ATSA, the obtained correlation coefficients suggested that no degradation had taken place. This demonstrates clearly that the use of retention-time alignment may lead to incorrect conclusions. Thus, whether such a strategy can be applied must be critically assessed for each application.

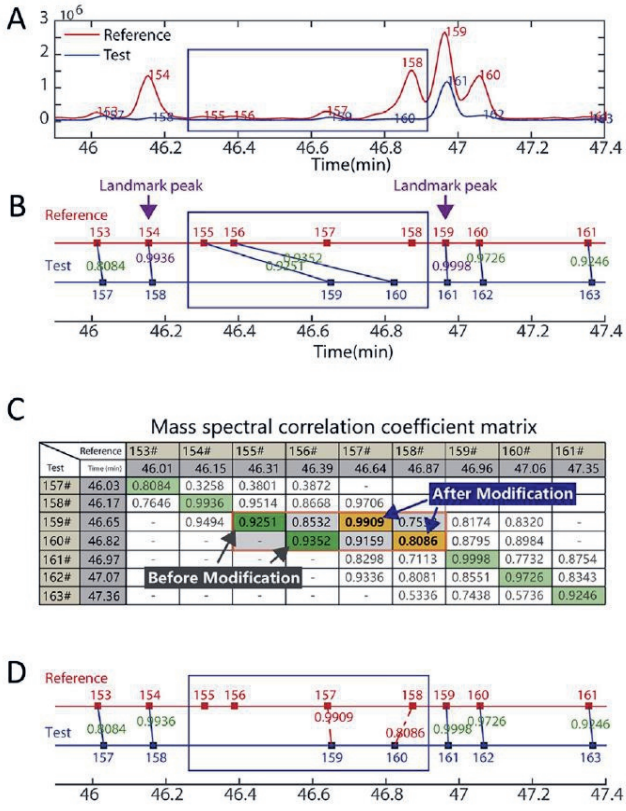

Figure 6.6: Peak alignment based on maximum-correlation path and the additional use of landmark peaks. A) selected range of the chromatogram; (B) misaligned peaks when only mass-spectral information is utilized; (C) The locations of the misaligned peaks in the maximum correlation coefficient path and the modified result after utilizing landmark peaks. In green the highest obtained PCC values are shown prior to the correction using landmark peaks; higher PCC values could be obtained by ignoring peaks 155 and 156 however, the new PCC values are shown in yellow; (D) Aligned chromatogram after correction. Reproduced with permission from [71].

6.2.3.3. MS-based peak alignment

Several alignment algorithms have been developed that are based on the use of MS [71,72]. In the approach of Fu *et al.* [71] baseline correction was first carried out by a LMV approach (see Section 2.2.4). The actual time-shift alignment consists of four steps: (*i*) extraction of the path of maximum MS-correlation, (*ii*) peak-alignment modification using landmark peaks, (*iii*) grouping and registration of missing peaks, and (*iv*) peak-alignment refinement. The first step required an initial estimate of the time shift (0.5 min in the described case), after which PCCs (see Section 2.3.1, **Equation 6.14)** were calculated based on mass spectra for each test and

reference chromatographic peak that fell within this initial time-shift window. All PCCs were collected in a correlation matrix that was used to determine the maximum-correlation path. The correlation matrix and the determined maximum-correlation path are illustrated in **Figure 6.6**.

The approach is based on the assumption that peak elution order is consistent between samples. This may not always the case. Therefore, alignment based on landmark peaks has also been incorporated. In this approach, landmark peaks are first defined as those peaks showing PCCs above 0.99. The time shifts of these landmark peaks are then stored in a vector and outliers are removed based on the median and the standard deviations of the landmark peaks' time shifts. Time shifts between two landmark peaks are linearly interpolated and an expected time shift can be calculated. This is then compared to the original time shift resulting from step (*i*) and the peak is realigned to the nearest reference peak in case the expected time shift is significantly different from the original time shift. However, as noted by the authors, while the time shift can also be approximated using non-linear interpolation, it cannot be employed in situations where the elution order has changed. In steps (*iii*) and (*iv*), certain peaks may be not be present in the reference chromatogram. These missing peaks are grouped based on their retention time with a maximum time shift window of 0.1 min, after which the chromatogram is realigned one final time.

The developed MS-based alignment was validated by applying it to a GC-MS data set including 12 growth and 18 maturation plant samples. Peak-alignment results of these 30 samples are illustrated in **Figure 6.7** for a selection of 15 closely eluting peaks.

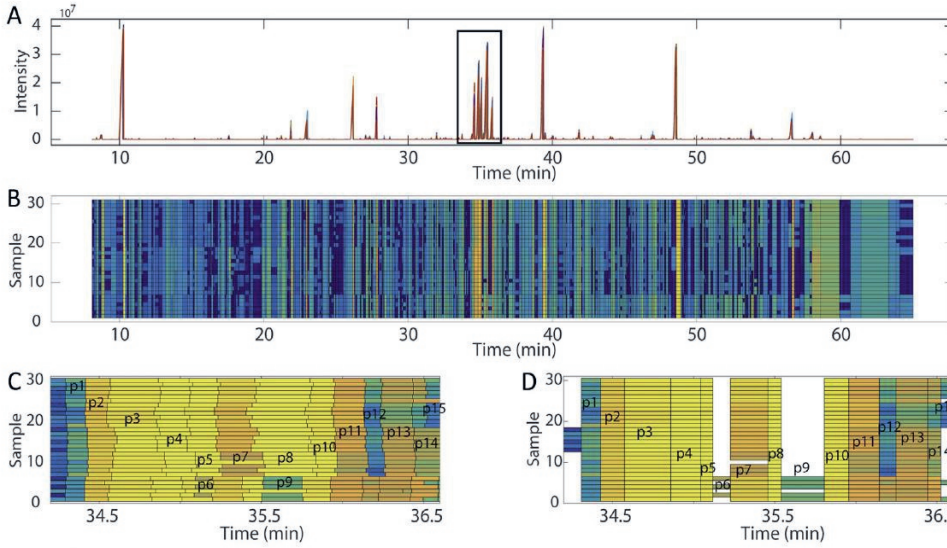

Figure 6.7: Peak alignment results with A) the original chromatogram; highlighted are 15 closely eluting peaks, (B) alignment results, and C) and D) original and aligned peaks within the region containing the 15 closely eluting peaks. Reproduced with permission from [71].

Another method incorporating baseline correction, peak detection, and time-shift alignment was proposed by Yu *et al.* for metabolic profiling analysis of 30 plant samples [73]. The method uses ACPD for peak detection and baseline correction, after which time shifts are corrected for based on the TIC data. After this pre-treatment PCA, ANOVA and partial least-squares discriminant analysis (PLSDA) were applied to further analyse the data. Peak alignment required first choosing a reference chromatogram, which in this case was the chromatogram containing the highest number of peaks. After peak detection and background correction, segments from both the chromatogram to be aligned and the reference chromatogram were selected based on an initial time-shift estimate (0.5 min was chosen). Initially, a rough alignment was performed using a similar approach as described in Section 2.3.1. In this case the cosine correlation was calculated rather than the PCC. Note that both are related, with the difference being that the PCC is the centred cosine correlation, which itself is the normalized inner product. The sum of the weighted individual cosine values (COS) was then used to obtain the initially aligned chromatogram.

After initial alignment, a precise alignment was carried out by accounting for the relative distances, cosine values, and real distances between a chromatographic peak

in the reference and each of the peaks to be aligned in the sample chromatogram, within the respective segment. This yielded an alignment table. In those cases where two of the reference peaks were aligned to the same sample peak, the peak with the smallest cosine correlation would be removed, the roles of reference and test chromatogram inverted, and the two alignment tables would be combined. For all other cases this approach was not applied.

Although the time-shift-alignment procedure was validated by aligning the data from the plant samples, the procedure was not compared with other approaches. As also stated by the authors, one of the disadvantages of this peak alignment approach is that the elution order must remain unchanged between samples. This assumption is actually inherent to many of the peak-alignment methodologies currently available.

6.2.3.4. Approaches for 2D chromatography

In addition to the approaches above, a number of less-recent studies have focused on retention-time alignment in 2D chromatography where in particular second-dimension modulations must be aligned to facilitate further data analysis. PARAFAC was applied to correct such retention-time shifts between neighbouring modulations [74]. Johnson et al. applied a windowed-rank minimization with interpolative stretching to the separations of naphthalenes in jet fuel by GC×GC [75]. Another method applied to GC×GC data used indexing schemes for warping in both dimensions [76]. Similar to background correction, other developed methods for retention-time alignment approached the data from an imagine perspective [59,77,78]. With most developed approaches generally exclusively adoptable to three-way data structures, Allen and Rutan developed an approach which allowed processing of four-way data structures and applied this to LC×LC-DAD data [79].

6.2.3.5. Correction for wrap around

In some cases, analytes may not elute within the modulation time and appear in following modulations. This is known as wrap around and is rather common in GC×GC. One method to resolve this treats the 2D chromatogram as a continuous three-dimensional cylinder where the end of one modulation is the beginning of the next [80]. Alternatively, absolute retention times may be determined by using an integer fraction of the original modulation to detect occurrences of wrap around [81].

6.3. Conclusions and outlook

Robust data-analysis strategies are needed to obtain useful information on complex samples using the increasingly advanced analytical tools. Pre-processing of the data is indispensable to remove irrelevant anomalies, which otherwise may induce significant errors in, for example, quantification or classification. For background correction, BEADS (baseline estimation and denoising using sparsity) and assisted BEADs are highly promising recent developments, as these approaches seems capable of handling many different types of background distortions and are fast. The main downside is that these are parametric methods that require prior optimization [63,64]. An important development that may lead to more-accurate information is the use of profile spectra instead of centroid spectra in the correction of GC/LC-MS data, which is especially important considering the prevalence of these hyphenated MS methods [55]. Along similar lines, the most-noteworthy strategies for peak alignment in two-dimensional chromatography are those that operate not just in one-dimension but in both. Methods developed for such pixel-level alignment are still quite scarce, especially for application in LC×LC. One such method has, however, been recently been developed for GC×GC-HRMS data by Zushi *et al.* [68].

Although there have been many additional developments, it is difficult to judge which methods truly perform best. What has become abundantly clear is that a two-dimensional chromatogram is still very often treated as a series of individual 1D chromatograms, with the pre-processing methods being applied separately to all of these. This is most likely because many of the existing methods have been developed for LC-MS data sets, rather than for two-dimensional data. In terms of background correction, improvements can quite possibly be made by focusing on a series of modulations. The surface of the chromatogram may then be corrected, instead of applying a 1D-method iteratively (row or column-wise) to the data.

Data-analysis strategies, aimed to extract relevant information, are also rather difficult to compare, because the results greatly depend on the quality of the data. Most reported methods were developed to tackle a specific challenge in a data set and comparisons with other approaches supported by numerical data have rarely been reported. A comprehensive study of different types of data and data-analysis techniques would allow a better overview of which techniques can be best used in which situation.

References

[1] Olivieri, A. C., Analytical Advantages of Multivariate Data Processing. One, Two, Three, Infinity? *Anal. Chem.* 2008, 80, 5713–5720.

[2] Qin, S. J., Process data analytics in the era of big data. *AIChE J.* 2014, 60, 3092–3100.

[3] Cook, D. W., Burnham, M. L., Harmes, D. C., Stoll, D. R., Rutan, S. C., Comparison of multivariate curve resolution strategies in quantitative LCxLC: Application to the quantification of furanocoumarins in apiaceous vegetables. *Anal. Chim. Acta* 2017, 961, 49–58.

[4] Peters, S., Vivó-Truyols, G., Marriott, P. J., Schoenmakers, P. J., Development of an algorithm for peak detection in comprehensive two-dimensional chromatography. *J. Chromatogr. A* 2007, 1156, 14–24.

[5] Pirok, B. W. J., Stoll, D. R., Schoenmakers, P. J., Recent Developments in Two-Dimensional Liquid Chromatography: Fundamental Improvements for Practical Applications. *Anal. Chem.* 2019, 91, 240–263.

[6] Pirok, B. W. J., Gargano, A. F. G., Schoenmakers, P. J., Optimizing separations in online comprehensive two-dimensional liquid chromatography. *J. Sep. Sci.* 2018, 41, 68–98.

[7] Pirok, B. W. J., Making Analytical Incompatible Approaches Compatible. University of Amsterdam, Amsterdam 2019.

[8] Venter, P., Muller, M., Vestner, J., Stander, M. A., Tredoux, A. G. J., Pasch, H., de Villiers, A., Comprehensive Three-Dimensional LC × LC × Ion Mobility Spectrometry Separation Combined with High-Resolution MS for the Analysis of Complex Samples. *Anal. Chem.* 2018, 90, 11643–11650.

[9] Uliyanchenko, E., Cools, P. J. C. H., van der Wal, S., Schoenmakers, P. J., Comprehensive Two-Dimensional Ultrahigh-Pressure Liquid Chromatography for Separations of Polymers. *Anal. Chem.* 2012, 84, 7802–7809.

[10] Pirok, B. W. J., Abdulhussain, N., Aalbers, T., Wouters, B., Peters, R. A. H., Schoenmakers, P. J., Nanoparticle Analysis by Online Comprehensive Two-Dimensional Liquid Chromatography combining Hydrodynamic Chromatography and Size-Exclusion Chromatography with Intermediate Sample Transformation. *Anal. Chem.* 2017, 89, 9167–9174.

[11] Muller, M., Tredoux, A. G. J., de Villiers, A., Predictive kinetic optimisation of hydrophilic interaction chromatography × reversed phase liquid chromatography separations: Experimental verification and application to phenolic analysis. *J. Chromatogr. A* 2018, 1571, 107–120.

[12] Cacciola, F., Giuffrida, D., Utczas, M., Mangraviti, D., Dugo, P., Menchaca, D., Murillo, E., Mondello,

L., Application of Comprehensive Two-Dimensional Liquid Chromatography for Carotenoid Analysis in Red Mamey (Pouteria sapote) Fruit. *Food Anal. Methods* 2016, 9, 2335–2341.

[13] Sorensen, M., Harmes, D. C., Stoll, D. R., Staples, G. O., Fekete, S., Guillarme, D., Beck, A., Comparison of originator and biosimilar therapeutic monoclonal antibodies using comprehensive two-dimensional liquid chromatography coupled with time-of-flight mass spectrometry. *MAbs* 2016, 8, 1224–1234.

[14] Wang, X., Stoll, D. R., Schellinger, A. P., Carr, P. W., Peak Capacity Optimization of Peptide Separations in Reversed-Phase Gradient Elution Chromatography: Fixed Column Format. *Anal. Chem.* 2006, 78, 3406–3416.

[15] Montero, L., Ibáñez, E., Russo, M., di Sanzo, R., Rastrelli, L., Piccinelli, A. L., Celano, R., Cifuentes, A., Herrero, M., Metabolite profiling of licorice (Glycyrrhiza glabra) from different locations using comprehensive two-dimensional liquid chromatography coupled to diode array and tandem mass spectrometry detection. *Anal. Chim. Acta* 2016, 913, 145–159.

[16] Beens, J., Blomberg, J., Schoenmakers, P. J., Proper Tuning of Comprehensive Two-Dimensional Gas Chromatography (GC×GC) to Optimize the Separation of Complex Oil Fractions. *J. High Resolut. Chromatogr.* 2000, 23, 182–188.

[17] Blomberg, J., Schoenmakers, P. J., Beens, J., Tijssen, R., Compehensive two-dimensional gas chromatography (GC×GC) and its applicability to the characterization of complex (petrochemical) mixtures. *J. High Resolut. Chromatogr.* 1997, 20, 539–544.

[18] van Beek, F. T., Edam, R., Pirok, B. W. J., Genuit, W. J. L., Schoenmakers, P. J., Comprehensive two-dimensional liquid chromatography of heavy oil. *J. Chromatogr. A* 2018, 1564, 110–119.

[19] Groeneveld, G., Pirok, B. W. J., Schoenmakers, P. J., Perspectives on the future of multi-dimensional platforms. *Faraday Discuss.* 2019, 218, 72–100.

[20] Wilson, J. D., McInnes, C. A. J., The elimination of errors due to baseline drift in the measurement of peak areas in gas chromatography. *J. Chromatogr. A* 1965, 19, 486–494.

[21] Pearson, G. A., A general baseline-recognition and baseline-flattening algorithm. *J. Magn. Reson.* 1977, 27, 265–272.

[22] Savitzky, A., Golay, M. J. E., Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 1964, 36, 1627–1639.

[23] Kalman, R. E., A new approach to linear filtering and prediction problems. *J. Fluids Eng. Trans. ASME* 1960, DOI: 10.1115/1.3662552.

[24] Eilers, P. H. C. C., A Perfect Smoother. *Anal. Chem.* 2003, 75, 3631–3636.

[25] Liu, Y., Zhou, X., Yu, Y., A concise iterative method using the Bezier technique for baseline

construction. *Analyst* 2015, 140, 7984–7996.

[26] Lopatka, M., Barcaru, A., Sjerps, M. J., Vivó-Truyols, G., Leveraging probabilistic peak detection to estimate baseline drift in complex chromatographic samples. *J. Chromatogr. A* 2016, 1431, 122–130.

[27] Whittaker, E. T., On a New Method of Graduation. *Proc. Edinburgh Math. Soc.* 1922, 41, 63–75.

[28] Li, Z., Zhan, D.-J., Wang, J.-J., Huang, J., Xu, Q.-S., Zhang, Z.-M., Zheng, Y.-B., Liang, Y.-Z., Wang, H., Morphological weighted penalized least squares for background correction. *Analyst* 2013, 138, 4483.

[29] Zhang, Z.-M., Chen, S., Liang, Y.-Z., Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst* 2010, 135, 1138.

[30] Carlos Cobas, J., Bernstein, M. A., Martín-Pastor, M., Tahoces, P. G., A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data. *J. Magn. Reson.* 2006, 183, 145–151.

[31] Zhang, Z.-M., Chen, S., Liang, Y.-Z., Liu, Z.-X., Zhang, Q.-M., Ding, L.-X., Ye, F., Zhou, H., An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy. *J. Raman Spectrosc.* 2009, 41, 659–669.

[32] Baek, S.-J., Park, A., Ahn, Y.-J., Choo, J., Baseline correction using asymmetrically reweighted penalized least squares smoothing. *Analyst* 2015, 140, 250–257.

[33] Perez-Pueyo, R., Soneira, M. J., Ruiz-Moreno, S., Morphology-Based Automated Baseline Removal for Raman Spectra of Artistic Pigments. *Appl. Spectrosc.* 2010, 64, 595–600.

[34] Tauler, R., Smilde, A., Kowalski, B., Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.* 1995, 9, 31–58.

[35] de Juan, A., Tauler, R., Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Crit. Rev. Anal. Chem.* 2006, 36, 163–176.

[36] Allen, R. C., John, M. G., Rutan, S. C., Filgueira, M. R., Carr, P. W., Effect of background correction on peak detection and quantification in online comprehensive two-dimensional liquid chromatography using diode array detection. *J. Chromatogr. A* 2012, 1254, 51–61.

[37] Kuligowski, J., Quintás, G., Tauler, R., Lendl, B., de la Guardia, M., Background Correction and Multivariate Curve Resolution of Online Liquid Chromatography with Infrared Spectrometric Detection. *Anal. Chem.* 2011, 83, 4855–4862.

[38] Hendler, R. W., Shrager, R. I., Deconvolutions based on singular value decomposition and the pseudoinverse: a guide for beginners. *J. Biochem. Biophys. Methods* 1994, 28, 1–33.

[39] Nagai, Y., Sohn, W. Y., Katayama, K., An initial estimation method using cosine similarity for

multivariate curve resolution: application to NMR spectra of chemical mixtures. *Analyst* 2019, 144, 5986–5995.

[40] Liang, Y., Kvalheim, O. M., Unique resolution of hidden minor peaks in multidetection chromatography by first-order differentiation and orthogonal projections. *Anal. Chim. Acta* 1993, 276, 425–440.

[41] Yu, Y.-J., Wu, H.-L., Fu, H.-Y., Zhao, J., Li, Y.-N., Li, S.-F., Kang, C., Yu, R.-Q., Chromatographic background drift correction coupled with parallel factor analysis to resolve coelution problems in three-dimensional chromatographic data: Quantification of eleven antibiotics in tap water samples by high-performance liquid chromatography . *J. Chromatogr. A* 2013, 1302, 72–80.

[42] Koch, A., Weber, J.-V., Baseline Correction of Spectra in Fourier Transform Infrared: Interactive Drawing with Bézier Curves. *Appl. Spectrosc.* 1998, 52, 970–973.

[43] Prakash, B. D., Wei, Y. C., A fully automated iterative moving averaging (AIMA) technique for baseline correction. *Analyst* 2011, 136, 3130.

[44] Coombes, K. R., Tsavachidis, S., Morris, J. S., Baggerly, K. A., Hung, M.-C., Kuerer, H. M., Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics* 2005, 5, 4107–4117.

[45] Mantini, D., Petrucci, F., Pieragostino, D., Del Boccio, P., Di Nicola, M., Di Ilio, C., Federici, G., Sacchetta, P., Comani, S., Urbani, A., LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* 2007, 8, 101.

[46] Fu, H.-Y., Li, H.-D., Yu, Y.-J., Wang, B., Lu, P., Cui, H.-P., Liu, P.-P., She, Y.-B., Simple automatic strategy for background drift correction in chromatographic data analysis. *J. Chromatogr. A* 2016, 1449, 89–99.

[47] Yaroshchyk, P., Eberhardt, J. E., Automatic correction of continuum background in Laser-induced Breakdown Spectroscopy using a model-free algorithm. *Spectrochim. Acta Part B At. Spectrosc.* 2014, 99, 138–149.

[48] Yu, Y.-J., Xia, Q.-L., Wang, S., Wang, B., Xie, F.-W., Zhang, X.-B., Ma, Y.-M., Wu, H.-L., Chemometric strategy for automatic chromatographic peak detection and background drift correction in chromatographic data. *J. Chromatogr. A* 2014, 1359, 262–270.

[49] de Rooi, J. J., Eilers, P. H. C., Mixture models for baseline estimation. *Chemom. Intell. Lab. Syst.* 2012, 117, 56–60.

[50] Burden, F., Winkler, D., Methods in Molecular Biology. 2008, pp. 23–42.

[51] Ning, X., Selesnick, I. W., Duval, L., Chromatogram baseline estimation and denoising using

sparsity (BEADS). *Chemom. Intell. Lab. Syst.* 2014, 139, 156–167.

[52] Navarro-Huerta, J. A., Torres-Lapasió, J. R., López-Ureña, S., García-Alvarez-Coque, M. C., Assisted baseline subtraction in complex chromatograms using the BEADS algorithm. *J. Chromatogr. A* 2017, 1507, 1–10.

[53] Mazet, V., Carteret, C., Brie, D., Idier, J., Humbert, B., Background removal from spectra by designing and minimising a non-quadratic cost function. *Chemom. Intell. Lab. Syst.* 2005, DOI: 10.1016/j.chemolab.2004.10.003.

[54] Selesnick, I., 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE 2017, pp. 4546–4550.

[55] Erny, G. L., Acunha, T., Simó, C., Cifuentes, A., Alves, A., Background correction in separation techniques hyphenated to high-resolution mass spectrometry – Thorough correction with mass spectrometry scans recorded as profile spectra. *J. Chromatogr. A* 2017, 1492, 98–105.

[56] Reichenbach, S. E., Ni, M., Zhang, D., Ledford, E. B., Image background removal in comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* 2003, 985, 47–56.

[57] Reichenbach, S. E., Carr, P. W., Stoll, D. R., Tao, Q., Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography. *J. Chromatogr. A* 2009, 1216, 3458–3466.

[58] Reichenbach, S. E., Kottapalli, V., Ni, M., Visvanathan, A., Computer language for identifying chemicals with comprehensive two-dimensional gas chromatography and mass spectrometry. *J. Chromatogr. A* 2005, 1071, 263–269.

[59] Zeng, Z.-D., Chin, S.-T., Hugel, H. M., Marriott, P. J., Simultaneous deconvolution and reconstruction of primary and secondary overlapping peak clusters in comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* 2011, 1218, 2301–2310.

[60] Zhang, Y., Wu, H.-L., Xia, A.-L., Hu, L.-H., Zou, H.-F., Yu, R.-Q., Trilinear decomposition method applied to removal of three-dimensional background drift in comprehensive two-dimensional separation data. *J. Chromatogr. A* 2007, 1167, 178–183.

[61] Tomasi, G., van den Berg, F., Andersson, C., Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemom.* 2004, 18, 231–241.

[62] Eilers, P. H. C., Parametric Time Warping. *Anal. Chem.* 2004, 76, 404–411.

[63] Parastar, H., Jalali-Heravi, M., Tauler, R., Comprehensive two-dimensional gas chromatography (GC×GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution. *Chemom. Intell. Lab. Syst.* 2012, 117, 80–91.

[64] Gröger, T., Schäffer, M., Pütz, M., Ahrens, B., Drew, K., Eschner, M., Zimmermann, R., Application

of two-dimensional gas chromatography combined with pixel-based chemometric processing for the chemical profiling of illicit drug samples. *J. Chromatogr. A* 2008, 1200, 8–16.

[65] Gröger, T., Zimmermann, R., Application of parallel computing to speed up chemometrics for GC×GC–TOFMS based metabolic fingerprinting. *Talanta* 2011, 83, 1289–1294.

[66] Zhang, D., Huang, X., Regnier, F. E., Zhang, M., Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GC×GC−MS Data. *Anal. Chem.* 2008, 80, 2664–2671.

[67] Gros, J., Nabi, D., Dimitriou-Christidis, P., Rutler, R., Arey, J. S., Robust Algorithm for Aligning Two-Dimensional Chromatograms. *Anal. Chem.* 2012, 84, 9033–9040.

[68] Zushi, Y., Gros, J., Tao, Q., Reichenbach, S. E., Hashimoto, S., Arey, J. S., Pixel-by-pixel correction of retention time shifts in chromatograms from comprehensive two-dimensional gas chromatography coupled to high resolution time-of-flight mass spectrometry. *J. Chromatogr. A* 2017, 1508, 121–129.

[69] Zheng, Q.-X., Fu, H.-Y., Li, H.-D., Wang, B., Peng, C.-H., Wang, S., Cai, J.-L., Liu, S.-F., Zhang, X.-B., Yu, Y.-J., Automatic time-shift alignment method for chromatographic data analysis. *Sci. Rep.* 2017, 7, 256.

[70] Fu, H.-Y., Guo, J.-W., Yu, Y.-J., Li, H.-D., Cui, H.-P., Liu, P.-P., Wang, B., Wang, S., Lu, P., A simple multi-scale Gaussian smoothing-based strategy for automatic chromatographic peak extraction. *J. Chromatogr. A* 2016, 1452, 1–9.

[71] Fu, H.-Y., Hu, O., Zhang, Y.-M., Zhang, L., Song, J.-J., Lu, P., Zheng, Q.-X., Liu, P.-P., Chen, Q.-S., Wang, B., Wang, X.-Y., Han, L., Yu, Y.-J., Mass-spectra-based peak alignment for automatic nontargeted metabolic profiling analysis for biomarker screening in plant samples. *J. Chromatogr. A* 2017, 1513, 201–209.

[72] Yu, Y.-J., Zheng, Q.-X., Zhang, Y.-M., Zhang, Q., Zhang, Y.-Y., Liu, P.-P., Lu, P., Fan, M.-J., Chen, Q.-S., Bai, C.-C., Fu, H.-Y., She, Y., Automatic data analysis workflow for ultra-high performance liquid chromatography-high resolution mass spectrometry-based metabolomics. *J. Chromatogr. A* 2019, 1585, 172–181.

[73] Yu, Y.-J., Fu, H.-Y., Zhang, L., Wang, X.-Y., Sun, P.-J., Zhang, X.-B., Xie, F.-W., A chemometric-assisted method based on gas chromatography–mass spectrometry for metabolic profiling analysis. *J. Chromatogr. A* 2015, 1399, 65–73.

[74] Skov, T., Hoggard, J. C., Bro, R., Synovec, R. E., Handling within run retention time shifts in two-dimensional chromatography data using shift correction and modeling. *J. Chromatogr. A* 2009, 1216, 4020–4029.

[75] Johnson, K. J., Prazen, B. J., Young, D. C., Synovec, R. E., Quantification of naphthalenes in jet fuel

with GC×GC/Tri-PLS and windowed rank minimization retention time alignment. *J. Sep. Sci.* 2004, 27, 410–416.

[76] Pierce, K. M., Wood, L. F., Wright, B. W., Synovec, R. E., A Comprehensive Two-Dimensional Retention Time Alignment Algorithm To Enhance Chemometric Analysis of Comprehensive Two-Dimensional Separation Data. *Anal. Chem.* 2005, 77, 7735–7743.

[77] Nelson, R. K., Kile, B. M., Plata, D. L., Sylva, S. P., Xu, L., Reddy, C. M., Gaines, R. B., Frysinger, G. S., Reichenbach, S. E., Tracking the Weathering of an Oil Spill with Comprehensive Two-Dimensional Gas Chromatography. *Environ. Forensics* 2006, 7, 33–44.

[78] Cordero, C., Liberto, E., Bicchi, C., Rubiolo, P., Reichenbach, S. E., Tian, X., Tao, Q., Targeted and Non-Targeted Approaches for Complex Natural Sample Profiling by GCxGC-qMS. *J. Chromatogr. Sci.* 2010, 48, 251–261.

[79] Allen, R. C., Rutan, S. C., Investigation of interpolation techniques for the reconstruction of the first dimension of comprehensive two-dimensional liquid chromatography–diode array detector data. *Anal. Chim. Acta* 2011, 705, 253–260.

[80] Weusten, J. J. A. M., Derks, E. P. P. A., Mommers, J. H. M., van der Wal, S., Alignment and clustering strategies for GC×GC–MS features using a cylindrical mapping. *Anal. Chim. Acta* 2012, 726, 9–21.

[81] Micyus, N. J., Seeley, S. K., Seeley, J. V., Method for reducing the ambiguity of comprehensive two-dimensional chromatography retention times. *J. Chromatogr. A* 2005, 1086, 171–174.

Chapter 7

Comparison of background correction algorithms used in chromatography

Abstract

The objective of the present work was to make a quantitative and critical comparison of a number of drift and noise-removal algorithms, which were proven useful by other researchers, but which had never been compared on an equal basis. To make a rigorous and fair comparison, a data generation tool is developed in this work, which utilizes a library of experimental backgrounds, as well as peak shapes obtained from curve fitting on experimental data. Several different distribution functions are used, such as the log-normal, bi-Gaussian, exponentially convoluted Gaussian, exponentially modified Gaussian and modified Pearson VII distributions. The tool was used to create a set of hybrid (part experimental, part simulated) data, in which the background and all peak profiles and areas are known. This large data set (500 chromatograms) was analysed using seven different drift-correction and five different noise-removal algorithms (35 combinations). Root-mean square errors and absolute errors in peak area were determined and it was shown that in most cases the combination of sparsity-assisted signal smoothing and asymmetrically reweighted penalized least-squares resulted in the smallest errors for relatively low-noise signals. However, for noisier signals the combination of sparsity-assisted signal smoothing and a local minimum value approach to background correction resulted in lower absolute errors in peak area. The performance of correction algorithms was studied as a function of the density and coverage of peaks in the chromatogram, shape of the background signal, and noise levels. The developed data-generation tool is published along with this article, so as to allow similar studies with other simulated data sets and possibly other algorithms. The rigorous assessment of correction algorithms in this work may facilitate further automation of data-analysis workflows.

Publication: L.E. Niezen, P.J. Schoenmakers, B.W.J. Pirok, Critical comparison of background correction algorithms used in chromatography, Anal Chim Acta. 1201 (2022). https://doi.org/10.1016/j.aca.2022.339605.

7.1. Introduction

Spectroscopic or chromatographic data can generally be assumed to consist of three components, (*i*) low-frequency baseline drift, (*ii*) high-frequency noise and (*iii*) relevant chemical information, typically with a frequency between that of drift and noise. The latter two contributions together are also commonly described as "background". Often, there is more background than chemical information present in a signal, as each data point contains a background contribution. In such a case, or if the background is of a frequency very similar to that of the relevant signals, problems may occur with the interpretation of the data. For example, peak detection may be hindered, and errors in classification, discrimination, and, especially, quantification, may occur [1–7]. It is, therefore, desirable to perform baseline-drift correction and noise removal to ensure a correct interpretation of the data, unless peak detection can be performed in such a way that it is not hindered by the presence of noise and drift.

A large number of background-correction algorithms have been developed [8–25]. Examples of baseline-drift-correction algorithms include many of the penalized least-squares (PLS) methods, including asymmetrical least squares (asLS) [16], asymmetrically reweighted penalized least squares (arPLS) [24], adaptive iteratively reweighted penalized least squares (airPLS) [19], modified airPLS (MairPLS), and morphologically weighted penalized least squares (MPLS) [10] as well as other techniques, such as iterative polynomial fitting [26,27], Corner-Cutting [9], Backcor [11,12] and baseline estimation and denoising using sparsity (BEADS) [14]. Additionally, methods based on Fourier filtering and on wavelets have also been developed [22,28,29] as well as less conventional methods based on the use of neural networks [30,31]. Although many background-correction methods have been proposed, comparisons between the performance of these are scarce and often inadequate.

Firstly, in many cases the background-correction methodologies developed for use on spectroscopic or chromatographic data are compared only qualitatively to two or three other methods using experimental data, while quantitative comparisons tend to be limited to small sets of simple simulated data [19,24]. Consequently, it is not clear which background-correction methods perform best. Instead, a trial-and-error

approach is routinely taken, in which three or four methods are arbitrarily selected and applied to a small test set of data. The (qualitatively) best performing one is used for the correction of all further measurements. If the test set is representative for all data and good methods are selected, such an approach can work reasonably well. However, this is by no means guaranteed and when correction is required for large numbers of measurements automation of background correction in data-analysis workflows is susceptible to errors. This is especially relevant when data-analysis methods, such as classification, discrimination or clustering are employed. In such cases, incorrect background correction can lead to erroneous results and incorrect conclusions.

Secondly, most approaches have been developed for specific datasets, such as Backcor, which was originally intended for the background correction of optical spectra [11]. While this is understandable, it induces the risk of a data-dependent bias in performance when evaluating the different methods. However, since quantitative comparisons are virtually non-existent, the magnitude of this risk cannot be assessed.

Thirdly, there are no data sets available for an objective comparison of background-correction approaches. Authors have generally employed specific datasets or simulated data. The latter is a pragmatic solution, which has the advantage that the ground-truth values for peak characteristics (*e.g.* peak area, shape) and background are known. This allows quantification of the extent of information loss as a result of the correction. A common criticism against the use of simulated data is that it is thought to be less representative than real data. In many cases this can be deemed true, as simple polynomial, sinusoidal or linear baselines are used, along with Gaussian peak shapes. Ideally, a large set of generated data that is sufficiently varied should be used against which all methods can be benchmarked.

In this work we aim to rigorously compare a number of recently developed background correction methodologies (i.e. baseline-drift correction as well as noise removal) in a comprehensive and critical manner. For this purpose, we used experimental data on backgrounds and peaks to create large sets of hybrid (part experimental, part simulated) data. Methods that took very long computation times (*e.g.* several minutes or more for a one-dimensional signal of approximately 20000

data points) were discarded in the present work after an initial evaluation, as our eventual objective is to apply the most-appropriate algorithms to two-dimensional liquid chromatography (2D-LC). For the same reason, only methods with no more than three different input parameters were assessed, to avoid significant manual tuning of the parameters to obtain satisfactory results.

7.2. Theory

A variety of different background-correction methods have been compared in this work. Here a brief overview of the theory behind each method is given. The noise-removal methods that were used prior to drift correction include well-known smoothing methods, such as Savitsky-Golay smoothing (SG) [8], Whittaker smoothing [32], finite-impulse-response (FIR) low-pass filtering [33], and wavelet filtering [34], as well as the more novel sparsity-assisted signal smoothing (SASS) [21]. After smoothing, drift-correction methods, *viz.* asymmetric least-squares (asLS) [16] and two conceptually similar methods, including adaptive iteratively reweighted penalized least-squares (airPLS) [19], asymmetric reweighted penalized least-squares (arPLS) [24] were applied. Additionally, the mixture model (MM) [18], a method based on iterative polynomial fitting with an asymmetric cost function (Backcor [11]), a method based on local minimum values (LMV) [15] and a recent method based on the use of an artificial neural network (ANN), henceforth referred to as the Autoencoder [31], were also included.

7.2.1. Drift-Correction Methods

7.2.1.1. Backcor

Many drift-correction methods are based on a polynomial-fitting approach, with the signal drift being described by a polynomial of a certain order. Such approaches cannot easily be automated, as this would require automatic detection of peaks and selection of background regions in the raw data. The correction method should itself be capable of determining which points belong to the drifted baseline and which do not. In Backcor this is achieved by iteratively fitting a polynomial through the entire signal and utilizing asymmetric forms of typically symmetric cost functions, such as the Huber or truncated quadratic cost functions, to penalize data points falling above the fit less harshly than those that fall below the fit [11,12]. As a result, minimizing such a cost function results in positive peaks being automatically filtered out during

the fitting procedure, since they have a lower cost. Because the noise around the drift is assumed to be normally distributed, this method still relies on a user-defined threshold to distinguish between noise and peaks.

The condition for the asymmetric truncated quadratic cost function is as follows.

$$\varphi = \begin{cases} d^2, & d < s \\ s^2, & d \geq s \end{cases} \tag{7.1}$$

With $d = x_i - z_i$, x the original data, z the fitted data, d the difference between the fit and the data for the i-th datapoint, φ the cost, s the user-defined threshold. The user additionally has to choose the cost function and the degree of the polynomial. Note that when the asymmetrical truncated quadratic cost function is used together with a threshold s of zero the approach is equivalent to the iterative polynomial fitting approaches [26,27]. It can be seen that d^2 corresponds to conventional least squares, unless the difference exceeds the set threshold, above which d^2 becomes constant (and equal to s^2). Minimizing the sum of the φ for all datapoints allows determining the polynomial coefficients and hence the drift.

7.2.1.2. Local-minimum value

The local-minimum-value approach relies on the presence of local-minimum values (LMV's), which are data points lower in intensity than adjacent data points [15]. The approach first establishes an "initial background" consisting of local minimum values, and then removes any points above this initial estimate by using a moving window and median-based outlier detection. The latter relies on a threshold based on the amount of noise in the signal. The drift is then obtained by linear interpolation between the areas that were considered as outliers (peak regions). In this approach the peak regions are therefore detected based primarily on the noise in the signal and the chosen window width, which should be chosen based on the peak width.

7.2.1.3. Asymmetric least-squares

Many well-known background-correction algorithms are based on the use of penalized least squares (PLS). The PLS algorithm relies on balancing the fit of a model to the data, F, given by the sum of squares error (SSE), and its roughness (R) by adjusting a smoothing parameter, λ:

$$\varphi = F + \lambda R = \sum_{i=1}^{m}(x_i - z_i)^2 + \lambda \sum_{i=2}^{m}(\Delta z_i)^2 = \|x - z\|^2 + \lambda\|\mathbf{D}z\|^2 \quad (7.2)$$

Where x_i is the i^{th} data point in the signal, x, and z_i is the i^{th} point of the fitted data, z. The difference between adjacent fitted data points is therefore given by Δz_i. This method as such cannot be used for background-drift correction, as it requires prior information on the locations of peaks in the signal. If these locations are known a binary mask or "weighted matrix" can be created, which ensures that only the background drift is modelled [35,36].

$$(\mathbf{W} + \lambda\mathbf{D}'\mathbf{D})z = \mathbf{W}x \quad (7.3)$$

$$z = (\mathbf{W} + \lambda\mathbf{D}'\mathbf{D})^{-1}\mathbf{W}x \quad (7.4)$$

Where $\mathbf{W}$ is a diagonal matrix with weight vector w_i on its diagonal, λ is the smoothing parameter and $\mathbf{D}$ is a difference matrix such that $\mathbf{D}z = \Delta z$. In case a binary mask is used w_i consists of solely ones and zeroes to differentiate between peaks and baseline, respectively. However, in principle the weights may be any value between zero and one depending on how the weights are established. Furthermore, in case of the asLS, arPLS, and airPLS algorithms the determination of these weights is based on an iterative process where weights are selected based on the difference from the fitted baseline. For the initial fit no penalty is given (weights are all equal to one). Points far away from this initially determined baseline are then given smaller weight and hence will have less influence on the fit. These weights are then used to solve **Equation 7.4** once again, and new weights are established. This process is continued until the weights become invariable. In asLS, developed by Eilers *et al.* [16], the weights are established using an asymmetry parameter (p), which allows for the weights associated with positive and negative deviations from the baseline to be different (smaller and larger, respectively). This approach is summarized in **Equation 7.5.**

$$w_i = \begin{cases} p, & x_i > z_i \\ 1 - p, & x_i \leq z_i \end{cases} \quad (7.5)$$

The asymmetry parameter p can vary between 0 and 0.5, with 0.5 resulting in a conventional fit, while anything smaller than 0.5 will result in the peaks being taken into account less.

7.2.1.4. Adaptive iteratively reweighted penalized least-squares

In the case of airPLS weights are selected based on an exponential function (**Equation 7.6a**) [19]. The algorithm is terminated once the difference between the signal and the fitted vector $|d_t|$ falls below a user-selected threshold, *i.e.* when condition (**Equation 7.6b**) is met.

$$w_i = \begin{cases} 0, & x_i \geq z_i \\ e^{n(x_i - z_i)/|d|}, & x_i < z_i \end{cases} \tag{7.6a}$$

$$|d_t| < 0.001|\mathrm{x}| \tag{7.6b}$$

With n being the iteration index and $d = x_i - z_i$, as earlier defined.

7.2.1.5. Asymmetric reweighted penalized least-squares

In the asymmetrically reweighted penalized least-squares (arPLS) algorithm, developed by Baek *et al.* [24] the weights are established based on a logistic function, as shown in **Equation 7.7.** This method functions in essentially the same way as asLS and airPLS, but is claimed to be better at establishing the drift in the presence of noise, due to how the weights are selected.

$$w_i = \begin{cases} logistic\left(d, \mu_{d,i}, \sigma_{d,i}\right), & x_i \geq z_i \\ 1, & x_i \leq z_i \end{cases} \tag{7.7}$$

$$logistic\left(d, \mu_{d,i}, \sigma_{d,i}\right) = \frac{1}{1 + e^{2(d - (-\mu_d + 2\sigma_d))/\sigma_d}} \tag{7.8}$$

With $d = x_i - z_i$, μ_d and σ_d are the mean and standard deviation of d^-, which is the part of d where the condition $x_i < z_i$ is met. This allows for weights above and below the signal to be the same, while any signal higher than the noise mean will receive a progressively lower weight. The baseline is established once the weights become invariable, once again depending on a set threshold.

7.2.1.6. Mixture model

The mixture model estimates the baseline by calculating the posterior probability that a point belongs to the baseline [18]. The entire signal is assumed to be constructed from a mixture of two probability densities, one of which is normal (and corresponds to the baseline) and one of which is unknown, corresponding to the peaks. To estimate both components of the signal/mixture, a so-called Expectation-

Maximization algorithm is used. In the first step of this algorithm the posterior probabilities are calculated, after which the baseline is modelled using P-splines (penalized B-splines). The coefficients (or penalties α) of these P-splines are determined by minimizing the following objective function:

$$\varphi = (x - \mathbf{B}\alpha)^T \mathbf{P}(x - \mathbf{B}\alpha) + \lambda \|\mathbf{D}\alpha\|^2 \quad (7.9)$$

$$\hat{\alpha} = (\mathbf{B}'\mathbf{P}\mathbf{B} + \lambda \mathbf{D}'\mathbf{D})^{-1}\mathbf{B}'\mathbf{R}x \quad (7.10)$$

In which x corresponds to the data, $\mathbf{B}$ corresponds to an n × m cubic spline basis of m number of splines, λ is the smoothing parameter, $\mathbf{P} = diag(p_i)$ and p_i is the posterior probability for the i^{th} data point to belong to the baseline. These posterior probabilities are calculated from

$$p_i = \frac{\pi g(x|\mu,\sigma)}{\pi g(x|\mu,\sigma) + (1-\pi)h(x-\mu)} \quad (7.11)$$

where $g(x|\mu,\sigma)$ is the normal density function (baseline + normally distributed noise with background level, μ, and standard deviation, σ), $h(x-\mu)$ the unknown density function (peaks), and π an unknown mixing ratio. This approach is conceptually similar to the previous three methods, with the posterior probabilities used as the weights. Once again, the method differs in how these weights are determined.

7.2.1.7. Autoencoder

The Autoencoder method is based on the use of deep learning algorithms and aims to concomitantly denoise and drift correct the input data [31]. The method achieves this by using a large number (in the order of several thousands) of differentiable or adaptable filters, which can be fine-tuned as long as a representative and large data set is available on which to train the method. Naturally this method is therefore limited by the data on which it was trained. However, recently Kensert *et al.*[31] have shown that by using a model trained on a large set of simulated data the successful drift correction and smoothing of experimental data may be achieved. In the present study we have included their pre-trained model to further assess how well this method can perform without performing extensive initial training. This is interesting because if the method is sufficiently flexible it could allow for unsupervised background correction, which is typically very difficult to achieve.

7.3. Experimental

Experimental backgrounds were obtained from various sources. Background 5 was measured on an Acquity system purchased from Waters (Milford, MA, USA) using refractive-index detection, while all other backgrounds (1 to 4) were measured on an Agilent 1260 system using diode-array UV detection, purchased from Agilent (Waldbronn, Germany). Backgrounds 1, 3, and 5 were obtained from empty modulations in comprehensive two-dimensional liquid chromatography (LC×LC) runs, while backgrounds 2 and 4 were blank measurements in one-dimensional LC.

Signal simulation has been performed using MATLAB 2018a purchased from MathWorks (Natick, MA, USA), on a Dell XPS13 Laptop purchased from Dell (Round Rock, TX, USA). Background correction and automatic parameter determination were performed using MATLAB 2020a on a Dell Alienware Area 51-9829 R2 PC.

The developed tool is available as a downloadable application on: https://cast-amsterdam.org/software/

7.4. Results & Discussion

7.4.1. Establishing experimental data for use in simulation

The major disadvantage of simulated data is that the complexity of experimental data may be oversimplified. Conversely, using experimental data may complicate the comparison of background-correction methods, as the ground-truth values (*e.g.* true peak areas) are not known. Therefore, we developed a library of simulated data which was based on experimental data. The workflow comprised three steps (see **Figure 7.1**). (*i*) A background was selected from a pool of blank experiments; (*ii*) Varying degrees of white noise were drawn from a Gaussian distribution; (*iii*) A number of peaks were added, with a shape extracted from experimental data by curve-fitting.

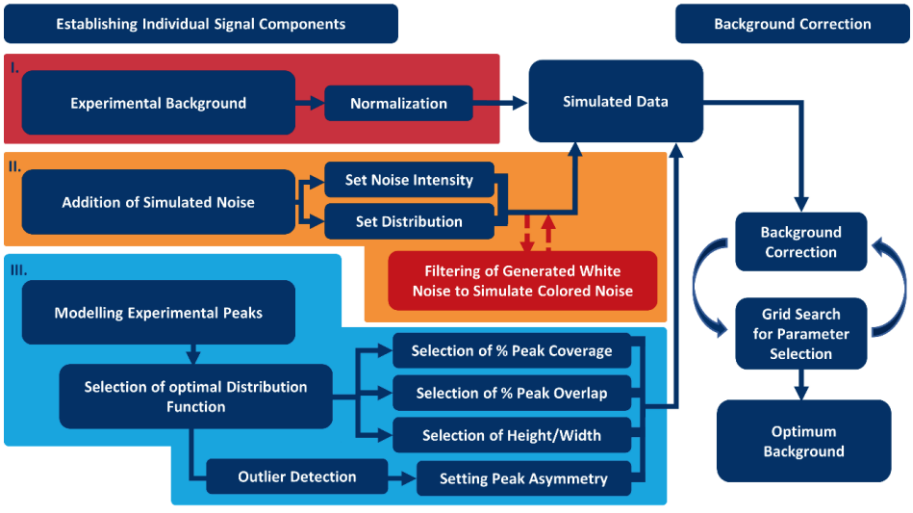

Figure 7.1: Scheme illustrating the developed approach for data simulation and subsequent background correction. Different coloured blocks (red, orange, light blue) indicate the three signal components (background, noise, and peaks)

7.4.1.1. Establishing experimental background and adding noise

The first step in the creation of the simulated data was the establishment of the low frequency drift component. As drift can be highly unpredictable and, therefore, difficult to model, an empirical approach was taken, where the background signals were obtained by compiling a library of different blanks from a variety of chromatographic experiments. This library can be further expanded with future research. Naturally, such experimental backgrounds contain an initial amount of noise in addition to the drift component. To establish an estimate of the initial noise in these backgrounds, the median absolute deviation (MAD) was used [37]. This is a robust measure of the deviation around the local average (i.e. the noise) present in the signal and is calculated using **Equation 7.12a**. However, in the presence of a baseline and peaks it has been suggested that a more representative value can be obtained by calculating the MAD from the first derivative of the background signal as given by **Equation 7.12b** [15].

$$\sigma = k * \text{median}|x_i - \text{median}(x)| \quad (i = 1, \dots, N) \tag{7.12a}$$

$$\sigma = k * \text{median}|dx_i - \text{median}(dx)| \; (i = 1, \dots, N) \tag{7.12b}$$

In which dx is the derivative of the signal and dx_i is the i^{th} point in this derivative, k is a (constant) scaling factor which for normally distributed data equals 1.4826. For an overview of the five experimental backgrounds that were used see Supplementary Material **Figure S-1**, section S-1.

In some cases, the experimental backgrounds contained one or more system peaks. These were manually removed from the signals by curve fitting and subtraction, followed by smoothing across this range. This was deemed necessary, because our approach ideally requires an experimental background that contains only low frequency drift and a small amount of initial noise. Their removal had to be performed manually and was hence tedious and time-consuming. However, when algorithms are not compared as presented in this work, the removal of such peaks is not required, as long as these are positive. Only in the presence of negative peaks will this be critical for most drift correction algorithms, as such peaks are generally treated as background drift.

The experimental backgrounds were perturbed with additional white noise, which was simulated as numbers randomly drawn from a normal distribution.

7.4.1.2. Establishing experimental peak shape

For the last step in the creation of the library (step III in **Figure 7.1**) experimental peak shapes were extracted from real chromatographic data. It is of key importance that the peaks are accurately modelled. This may be achieved by fitting empirical peak-shape models or distribution functions, such as the Gaussian, exponentially modified Gaussian (EMG), or Pearson distributions to the data. For chromatographic data deviations from the expected ideal Gaussian profile are expected, due to heterogenous mass-transfer kinetics and non-linear adsorption isotherms [38]. Such deviations usually come in the form of tailing (or sometimes fronting) peak profiles. For each peak the extent of tailing and/or fronting may be different. To describe all possible peaks mathematically, a function must be used that is flexible enough to describe any amount of tailing and fronting. Several comparisons of distribution functions have previously been performed [39–42]. From these studies a general consensus emerged that the EMG distribution described chromatographic peak shapes most accurately [42]. In the present study these common distributions were also evaluated, along with several alternatives, such as Gaussian, Bi-gaussian [43,44],

Pearson VII [45] and Modified Pearson VII [46] distributions. To successfully perform curve fitting two requirements must be met, *i.e.* (*i*) the approximate peak location must be known, and (*ii*) no background must be present [47]. In this study, the first prerequisite was met through manual selection of peak locations. While this can be performed automatically using peak-detection approaches, this would induce a risk of overlooking overlapping or small peaks. This may result in incorrect fitting for overlapping peaks. To meet the second requirement, either some form of background correction must be applied or data containing little or predictable background drift must be used. We opted to perform a linear background correction from the first to the last point in the selected peak regions. This involves the assumption that within the region of the peak the baseline does not show significant curvature. There are cases in which the approach cannot be used to describe peaks, for example when a large number of overlapping peaks is present, or when the background drift is significantly non-linear directly under the peak.

After the locations and peak regions were established, each peak was subjected to a least-squares curve-fitting procedure with 15 different distributions. In case of overlapping peaks, all peaks in the selected region were included and curve fitting was performed with two or more distributions of the same type. The possibility that overlapping peaks required different types of distribution functions was not considered in the current study, however even with a single distribution function it is still possible to describe a variety of peak shapes. The goodness-of-fit of the distributions to the experimental peaks was assessed using the Akaike information criterion (AIC) calculated using **Equation 7.13:**

$$AIC = N * \ln\frac{SSE}{N} + 2K \tag{7.13}$$

Where N is the number of data points, SSE is the sum of squared errors, in our case normalized for peak height, and K corresponds to the number of variables in the distribution function.

As an example, the results of this fitting approach applied to a selection of peaks from a single 2D-LC modulation (second-dimension chromatogram) containing significant background drift are shown in **Table 7.1** along with the types of distribution tested. The AIC values for the five best-performing distribution functions for each peak are shown in **Figure 7.2,** along with the individual fits for peaks 1-5.

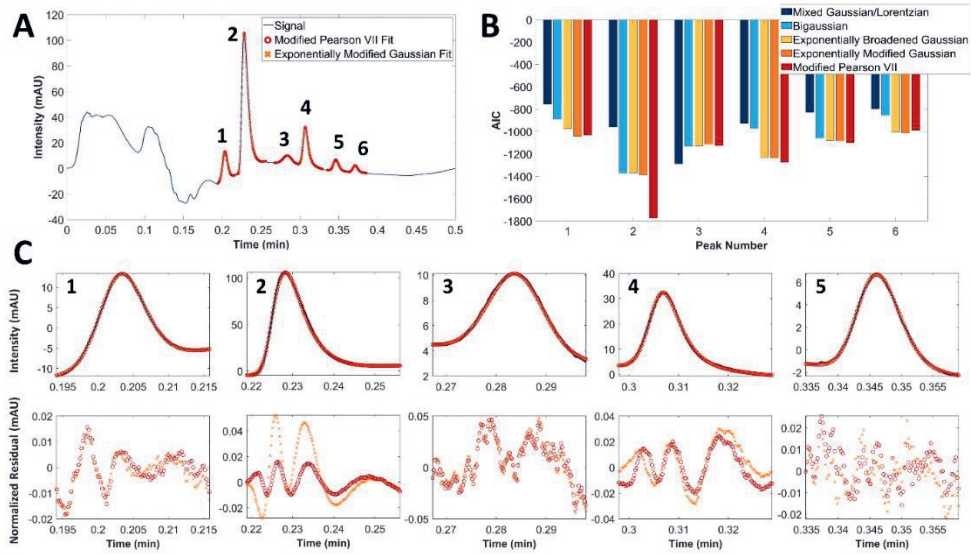

Figure 7.2: A) Fit for Modified Pearson VII and EMG distributions on experimental data, B) AIC values for the five best distribution models for each of the fitted peaks and C) zoomed-in fits and residuals for five individual peaks.

Table 7.1: AIC values obtained for a selection of distribution functions for the fitted peaks shown in Figure 7.2.

Distribution	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	$\sum$ AIC
Modified Pearson VII	**-1.03×10^{3}**	**-1.77×10^{3}**	**-1.13×10^{3}**	**-1.28×10^{3}**	**-1.10×10^{3}**	**-9.86×10^{2}**	**-7.20×10^{3}**
Exponentially Modified Gaussian	**-1.04×10^{3}**	**-1.39×10^{3}**	**-1.11×10^{3}**	**-1.24×10^{3}**	**-1.08×10^{3}**	**-1.01×10^{3}**	**-6.91×10^{3}**
Exponentially Broadened Gaussian	**-9.75×10^{2}**	**-1.37×10^{3}**	**-1.13×10^{3}**	**-1.23×10^{3}**	**-1.08×10^{3}**	**-1.00×10^{3}**	**-6.29×10^{3}**
BiGaussian	**-8.89×10^{2}**	**-1.37×10^{3}**	**-1.13×10^{3}**	**-9.69×10^{2}**	**-1.06×10^{3}**	**-8.55×10^{2}**	**-5.59×10^{3}**
Mixed Gaussian/Lorentzian	**-7.55×10^{2}**	**-9.59×10^{2}**	**-1.29×10^{3}**	**-9.26×10^{2}**	**-8.28×10^{2}**	**-7.98×10^{2}**	**-5.38×10^{3}**
Pearson	**-7.56×10^{2}**	**-9.45×10^{2}**	**-1.13×10^{3}**	**-9.12×10^{2}**	**-8.30×10^{2}**	**-7.70×10^{2}**	**-5.30×10^{3}**
Log-normal	**-7.78×10^{2}**	**-9.58×10^{2}**	**-1.12×10^{3}**	**-8.72×10^{2}**	**-8.39×10^{2}**	**-7.81×10^{2}**	**-5.30×10^{3}**
Gaussian	**-7.47×10^{2}**	**-9.44×10^{2}**	**-1.13×10^{3}**	**-8.63×10^{2}**	**-8.22×10^{2}**	**-7.70×10^{2}**	**-5.14×10^{3}**
Logistic	**-6.64×10^{2}**	**-9.35×10^{2}**	**-1.23×10^{3}**	**-8.24×10^{2}**	**-7.56×10^{2}**	**-7.07×10^{2}**	**-4.08×10^{3}**
Exponentially Broadened Lorentzian	**-5.35×10^{2}**	**-1.06×10^{3}**	**-6.86×10^{2}**	**-9.08×10^{2}**	**-6.35×10^{2}**	**-6.09×10^{2}**	**-3.82×10^{3}**
Lorentzian	**-5.04×10^{2}**	**-8.98×10^{2}**	**-6.65×10^{2}**	**-8.30×10^{2}**	**-5.86×10^{2}**	**-5.73×10^{2}**	**-3.04×10^{3}**

The results shown in **Table 7.1** and **Figure 7.2** are representative of those obtained on large numbers of treated datasets. The modified Pearson VII and EMG distributions were found capable of describing a range of different peak shapes. Other distribution functions performed well in some cases, but were less generally applicable. These conclusions are in agreement with previous studies on this subject [39,41]. Because the modified Pearson VII distribution [46] provided a slightly better fit compared to the EMG function this distribution was chosen for the creation of peaks in the simulated data. This distribution is described by **Equation 7.14**

$$f(x) = H\left(1 + \frac{(x-\mu)^2}{m(\sigma + A_S(x-\mu))^2}\right)^{-m} \quad (7.14)$$

in which μ corresponds to the mean of the distribution (the retention time), σ indicates the width of the peak, m is a parameter related to the kurtosis of the peak, covering a range between a fully Gaussian and a Lorentzian peak shape, A_s describes the asymmetry, or the extend of fronting or tailing, of the peak and H corresponds to the height of the peak..

7.4.1.3. Simulation of model chromatograms and spectra

Chromatograms or spectra are generated based on several input parameters as summarized in **Table 7.2**. The first set of parameters is used to describe peak shape and height while the second set of parameters dictates the intensity of the noise and the number of peaks in the chromatogram as well as their spacing.

Table 7.2: All parameters to be selected for signal generation.

Parameter Set	Parameter	Symbol	Range	Description
1st Set:	Kurtosis	m	3.5-51	Describes peak kurtosis
	Asymmetry	A_s	0.01-0.28	Describes peak asymmetry
	Amplitude	H	0.1-1	Multiplier for peak height
	Width	σ	0.007-0.008	Peak Width
	Position	μ	Random	Peak Retention Time
2nd Set:	Min. Peak Spacing		10%	Minimum space between two adjacent peaks (*e.g.* $\mu_1 and \mu_2$)
	Peak Coverage		10-100%	% of data points corresponding to peaks within selected region
	Noise Intensity		0-0.1	Amplitude of noise added to the chromatogram
	Background		1,2,3,4 or 5	Experimental blank
	Region Selection		Based on blank t_0	Region in which peak μ are generated.

The first set of input parameters determine the peak shape. These are drawn randomly and uniformly from the established maximum and minimum of a respective shape-defining parameter, *i.e.* m or A_s. As certain peaks in the input data may be significantly more asymmetrical than other ones, an outlier test was first performed on the m and A_s parameters obtained from the fitting procedure. Parameters were marked as outliers if they were more than three times the median absolute deviation away from the median. This test was performed to ensure that the simulated signals would represent realistic signals *i.e.* it was assumed that most peaks in a chromatogram would be of a typical shape. Outliers were subsequently

removed. To ensure that peak heights are consistent between all signals, the experimental background drift was first "min-max normalized" and peak heights were uniformly selected from a range between 0.1 and 1. However, in realistic data often some peaks are much higher or lower in intensity than the "average peak". To simulate this, 10% of the generated peaks were randomly selected to have an intensity of either 5 or 200% of the maximum intensity of the drift (*i.e.* 0.05 or 2). Peak widths were chosen based on the fitted peaks. This seems justified, since for signals of different length (*e.g.* slow one-dimensional chromatograms or very fast second-dimension chromatograms) the parameters from the first set (m and A_s) did not seem to change significantly, rather the peak width itself changed. However, it should be considered that in the present study primarily gradient-RPLC data of small, uncharged molecules were used to model the peaks, hence peak widths are expected to be relatively constant. The peak-shape parameters may change significantly in other modes of chromatography.

The second set of input parameters determine the number of peaks, their locations and the amount of overlap allowed between individual peaks. Firstly, regions where no peaks are to be generated, can be selected. Such regions were selected based on the experimental measurements. Secondly, a minimum peak spacing is selected, based on peak widths, as well as a total peak coverage. The peak overlap was calculated based on the width at 5% of the peak height and it was chosen in accordance with the selected peak coverage. For example, a minimum spacing of 0% may result in completely overlapping peaks, while a minimum spacing of 100% will result in a signal where all peaks are completely separated. The chosen total peak coverage is defined as the number of data points containing information on peaks, again measured from the width at 5% of the peak height, to the total region available for peak generation. The coverage, width and shape roughly determine the number of peaks that must be generated, but the coverage alone does not account for varying amounts of peak overlap. The signal coverage is calculated based on the minimum spacing only. This means that the actual coverage will always be lower than the chosen value. To account for this the actual peak coverage is determined once more after the full signal has been generated.

By adding the generated peaks to the experimental background perturbed with additional white noise, many realistic chromatograms can be rapidly created. To

allow other researchers to retrieve data from earlier publications (*e.g.* the present study) the tool features the setting of a seed, which is a number that serves to initialize the random number generator. This ensures that all "random" signal generation is controlled and reproducible. The same "random" signals can be generated again at any time, when required for future work.

7.4.2. Background correction

Using the developed data-simulation tool, background-correction methods could now be accurately compared. To this end different signals were created with peak coverages of 10, 21, 32, 43, 54, 66, 77, 88, 99 and 110%. Additionally, noise was added to each signal at ten different levels with "intensity" or standard deviation (σ) of 0 to 0.1 (up to ≈20% of the average peak height). Minimum peak spacing in all signals was set to 10%. This ensures the occurrence of severely overlapping peaks, while there were no peaks generated at exactly the same location.

Realistic data were generated in this manner for each of five different experimental backgrounds, resulting in 500 different signals for background comparison. A small representative fraction of the signals simulated in this way (different noise and coverage levels, three different experimental baselines) are shown in **Figure 7.3.**

To compare the background-correction and smoothing methods it is vital that the input parameters for each method are set such that the estimated background is as close as possible to the actual background. For the smoothing methods ideal input parameters were obtained by minimizing the root-mean-square error (RMSE, given by **Equation 7.15**) between the simulated noise and the noise obtained from subtracting the smoothed signal from the original signal, using a grid-search approach within manually defined constraints. For the drift-correction methods a similar approach was used where the RMSE was minimized between the known drift component and the background as determined by a drift-correction algorithm.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(b_i - z_{b,i}\right)^2}{N}} \tag{7.15}$$

Where b_i corresponds to i-th data point in the known background (either noise or drift component), $z_{b,i}$ corresponds to the i-th data point in the estimated background and N corresponds to the total number of data points in either z or b. For both smoothing and drift-correction methods this RMSE was minimized separately, starting with optimization of the smoothing, followed by optimization of the drift correction. Examples of some of the corrections obtained in this way are illustrated in **Figure 7.4.** In this case Savitsky-Golay was used for smoothing, while the result of six different drift-correction methods are shown.

Based on visual inspection of this specific signal the LMV and arPLS methods seem to perform slightly better than the other methods. The next step was to quantitatively compare all methods using the entire collection of generated data (500 chromatograms).

7.4.3. Quantitative Comparison of Correction Performance

7.4.3.1. Influence of the Smoothing Method

Using the simulated data and automatic parameter selection, a quantitative comparison was made between all methods or combinations of methods by evaluating the RMSE obtained for each signal as well as the % error in obtained peak areas. First, the influence of the smoothing method, applied before the drift correction, was investigated based on the calculated RMSE values. This resulted in a response surface where RMSE as a function both the added noise intensity, and the peak coverage could be visualized. A comparison was made by overlaying all surfaces and maintaining only the lowest RMSE values. For many of the smoothing methods the response surfaces were fairly similar (see Supplementary Material **Figure S-2**, section S-3), indicating that there were only minor differences in the performance of the smoothing methods. Background 5 (see Supplementary Material **Figure S-1**) yielded larger deviations. However, this may be explained by the fact that this is by far the shortest signal in terms of the number of data points, as a result of a very low detector frequency. Therefore, the way in which we added noise (by randomizing each data point) may not be realistic in this case and the frequency of peaks in this signal is much closer to the frequency of the noise. Overall, the degree of noise and/or the choice of the smoothing algorithm appear to only marginally affect the performance of the various background-correction algorithms.

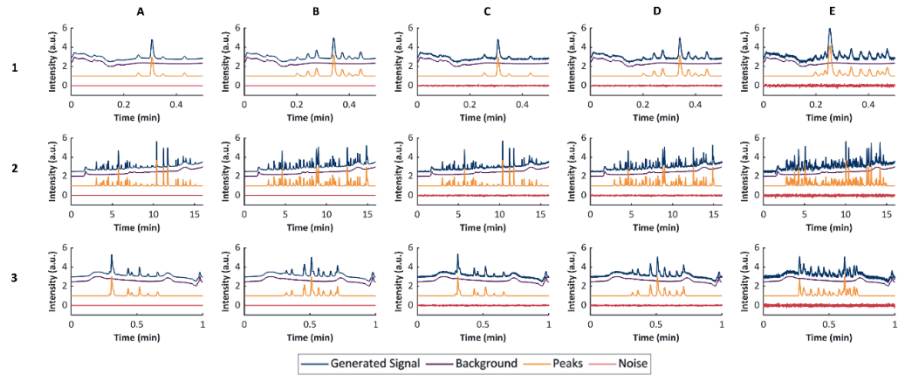

Figure 7.3: Representative examples of the generated data. A) no added noise + low peak coverage (44%), B) no added noise + medium peak coverage (76%), C) medium added noise ($\boldsymbol{\sigma}$ = 0.044) + low peak coverage (44%), D) medium added noise ($\boldsymbol{\sigma}$ = 0.044) + medium peak coverage (76%), E) high added noise ($\boldsymbol{\sigma}$ = 0.078) + high peak coverage (92%). Numbers indicate the different backgrounds.

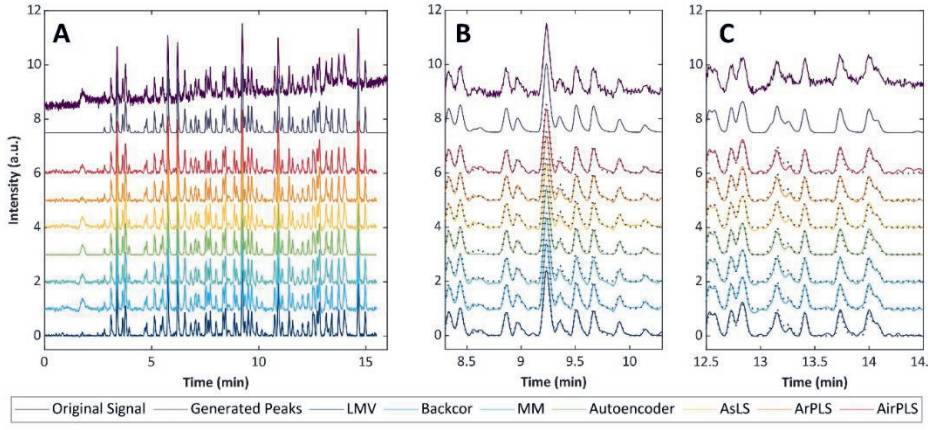

Figure 7.4: A) Correction of generated data (added noise of 0.078, coverage of 67.5%) by the combination of Savitsky-Golay (window width: 23, polynomial order: 6) smoothing followed by drift correction using the LMV (window width: 5), Backcor (s: 0.0579, polynomial order: 18), MM (number of b-splines: 73, $\boldsymbol{\lambda}$: 10^5), Autoencoder, AsLS ($\boldsymbol{\lambda}$: 10^5, p: 0.0306), arPLS ($\boldsymbol{\lambda}$: 3×10^5) and airPLS (lambda: 3×10^4) methods. B) Expansion of the region 8.3-10.3 min. C) Expansion of the region 12.5 – 14.5 min. For information on the methods see section 2. Dotted lines in (B) and (C) correspond to the generated (true) peak signals.

7.4.3.2. Influence of the Drift Correction Method

To provide a clearer overview of the influence of noise and peak coverage for the different-drift correction methods the RMSE was calculated between the known drift and the background determined by the algorithms. This was performed both with and without prior smoothing to evaluate how well the different algorithms performed in the presence of additive noise. Calculating the RMSE between the sum of known noise and drift components and the estimated background illustrates clearly that most methods cannot perform smoothing and drift removal simultaneously. In this case the Autoencoder method performed best. However, when the RMSE is calculated between the estimated background and the known drift component it is shown that most methods do not perform worse at determining the underlying signal drift in the presence of additive noise. This data is included in the Supplementary Material (**Figure S-4** and **Figure S-5**). Because the Autoencoder was capable of describing both noise and drift while many of the other algorithms could not it was decided to first perform smoothing using SASS for the comparison of the methods. In **Figure 7.5** the RMSE values obtained using background 2, initially smoothed using SASS are illustrated.

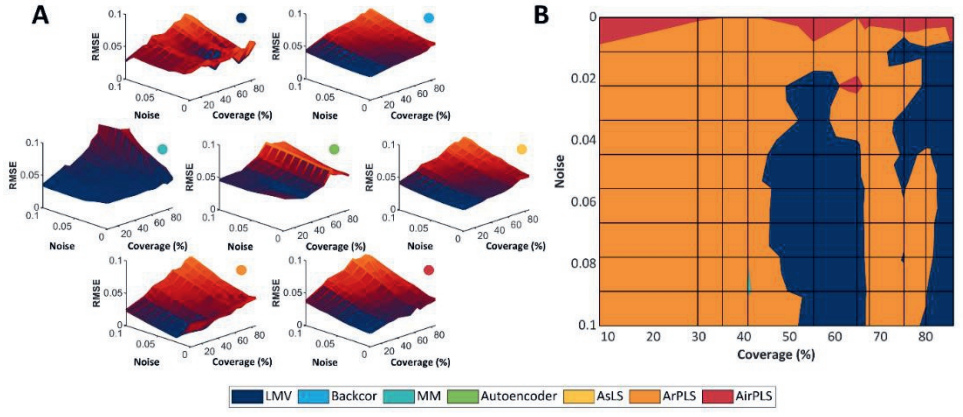

Figure 7.5: A) RMSE surfaces obtained for the various drift-correction methods in combination with the SASS smoothing algorithm and for background 2 in Figure 3. Methods are indicated by the coloured dots. B) Bottom view (lowest values) resulting from the overlaid RMSE surfaces. For an explanation of the methods see section 2.

From this comparison arPLS is seen to be the best performing method in most cases. However, this conclusion was found to depend on the background used. For other backgrounds the LMV or Backcor algorithms performed better than arPLS. However, if arPLS did not perform best, it was usually the second-best method and arPLS showed consistently good performance for all backgrounds investigated in this study (for an overview of all minimum RMSE surfaces see Supplementary Material **Figure S-3**).

7.4.3.3. Determination of Error in Peak Areas

It is interesting to know which method results in the smallest error in peak area. Therefore, the error in peak area obtained after correction has been evaluated. These errors were determined and compared to an approximate peak area obtained by trapezoidal integration from the simulated peaks. In case of overlapping peaks these were treated as one, therefore this is a "peak region" comparison rather than a comparison of individual peaks. This avoids reliance on curve fitting and the possibility of baseline or noise resulting in incorrect results. The regions were selected using the simulated peaks. The error in peak areas obtained from this approach for each method and for the signal with a peak coverage of 35.2% and a noise intensity of 0.033 from background 2 is portrayed in **Figure 7.6**. This is a representative case, in which a few peak regions contain overlapping peaks, but many peaks are isolated.

From **Figure 7.6-B** it is clear that Backcor performed worse than the other methods (indicated by the larger spread) followed by the MM and Autoencoder methods. The best-performing methods were LMV, asLS and airPLS, with mean errors of 3.8, 9.9 and 7.9% respectively. In many cases errors were still substantial, especially for low-intensity peaks, as could be expected. Peaks in regions 2, 4, 5, 13 and 17 showed the largest relative errors of up to nearly -150%, which implies that the drift is significantly over-corrected in this region relative to the peak's height, resulting in a much smaller determined peak area. However, the peaks in these regions typically also had very low signal-to-noise ratios (S/N) on the order of ~1.5-3. Relative errors of 10 or even 20% in peak area were quite common, also for peaks of average height (for more in-depth figures of % error vs peak height for the different methods see Supplementary Material **Figure S-6)**. This surprising finding highlights the importance of appropriate drift correction and noise removal when accurate

quantification is desired. The errors were similar when a different smoothing method was used prior to the drift correction. The results for the same signal and selected peak regions as in **Figure 7.6**, but with four different noise intensities (0.02, 0.06, 0.08, and 0.1), are illustrated in **Figure 7.7.**

It is seen in **Figure 7.7** that every method performs worse at higher noise levels, as is evident from the larger spread and the higher number of outliers. The deterioration of the performance is strongest for Backcor, MM and Autoencoder methods. Specifically in the case of the largest noise level it is likely that the deterioration of the Autoencoder's performance is because signals with this noise intensity (and correspondingly low S/R) were not included in the training set data. For all methods the reduced performance is not reflected in the corresponding RMSE surfaces (all of which appear relatively similar).

7.4.3.4. Outlook

Various improvements can still be made to the approach followed in this work and our evaluation method still has some limitations. The greatest challenges to the validity of the approach remain (*i*) the method used to extract the background, peaks and noise from the experimental input data, and (*ii*) whether the experimental data can be accurately recreated using simulations. We have avoided the first issue partially by using a set of experimental blanks for the simulations, which do not contain any peaks. The second point is especially critical when reconstructing peak shapes. In this work we have extracted peak shapes using an automatic curve-fitting approach. This is critically important, as manually evaluating the performance of background-correction and peak-detection algorithms will be incredibly time-consuming for large data sets. A further improvement can be parameter optimization. While a grid-search may currently be the best method for the determination of the parameters, there is still a certain risk that sub-optimal parameters are selected and the approach is very time-consuming. Furthermore, the influence of differently correlated noise on background-correction may be significant and this should be a subject of future study, as different detectors will produce different types of noise.

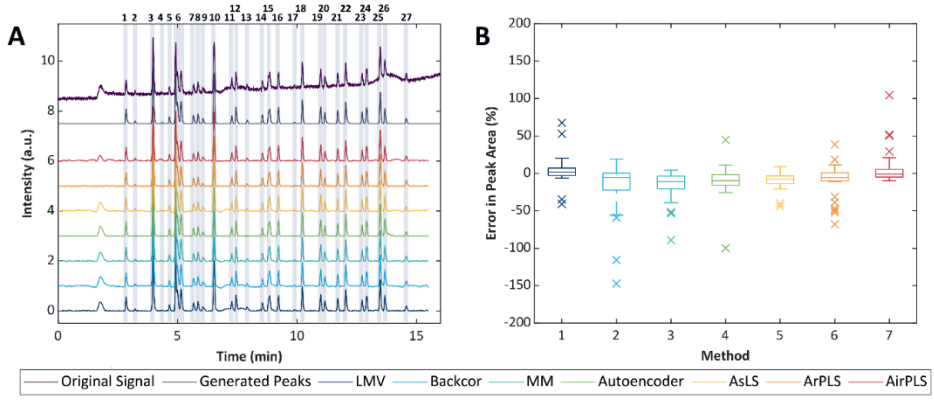

Figure 7.6: A) Signal with (in blue) markings indicating peak regions used for the % error calculation, the different traces represent (from top to bottom) the uncorrected signal and the generated peaks, followed by signals corrected using the LMV (window width: 113), Backcor (s: 0.1271, polynomial order: 20), MM (number of b-splines: 100, $\boldsymbol{\lambda}$: 10^5), Autoencoder, AsLS ($\boldsymbol{\lambda}$: 10^5, p: 0.0510), arPLS ($\boldsymbol{\lambda}$: 5.4×10^3) and airPLS (lambda: 4×10^4) methods. B) Error in peak area for each method. Regions 3, 6, 15 and 25 contain overlapping peaks; all other regions are individual peaks.

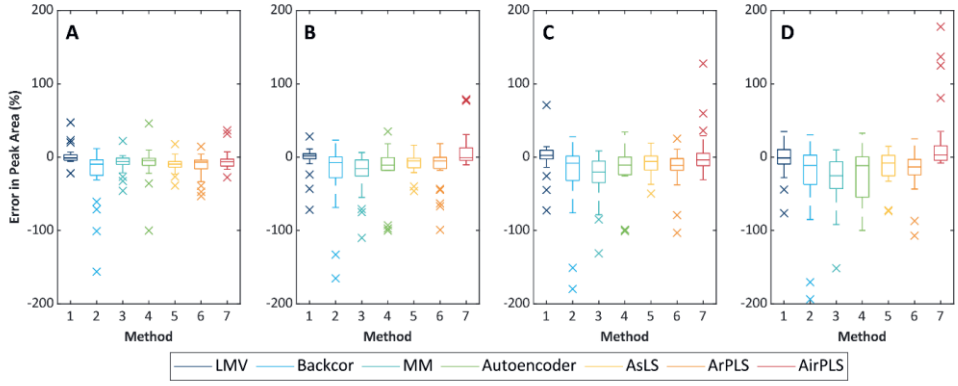

Figure 7.7: Relative errors for six different-drift correction methods for the signal shown in Figure 8-A for four levels of added noise corresponding to noise intensities of A) $\boldsymbol{\sigma}$ = 0.02, B) $\boldsymbol{\sigma}$ = 0.06, C) $\boldsymbol{\sigma}$ = 0.08 and D) $\boldsymbol{\sigma}$ = 0.1. Signals smoothed prior to this analysis using SASS.

7.5. Conclusion

A data-simulation tool has been developed which makes it possible to compare different background-correction and peak-detection methods. We have used this tool to compare a variety of data (pre-)processing methods. From the methods compared in this study, a combination of SASS and arPLS most often resulted in the lowest RMSE. Based on visual inspection this combination also showed the best looking results. However, it did not result in the smallest errors in peak area. The combination of SASS and LMV methods performed best in this respect. In terms of speed the Backcor and LMV algorithms provided the fastest drift correction, generally with evaluation times of less than half a second, while the arPLS algorithm performed slowest. However, this algorithm still generally provided results in less than a second if the number of data points remained below 10 thousand. For the smoothing algorithms nearly all algorithms performed equally, typically with evaluation times of less than 0.1s. It should also be specially mentioned that while in this case a pre-trained Autoencoder model was used, the results of this method were still relatively good. This indicates that in the future it may be possible to perform automatic background correction using similar methods as long as the training set is sufficiently large. The best combination of methods seems to depend on the nature of the background, which implies that it cannot always be a-priori predicted. However, the present study has provided valuable tools and methods to improve quantification in case the true background is unknown.

Supplementary material

References

[1] R.C. Allen, M.G. John, S.C. Rutan, M.R. Filgueira, P.W. Carr, Effect of background correction on peak detection and quantification in online comprehensive two-dimensional liquid chromatography using diode array detection, Journal of Chromatography A. 1254 (2012). https://doi.org/10.1016/j.chroma.2012.07.034.

[2] S. Samanipour, P. Dimitriou-Christidis, J. Gros, A. Grange, J.S. Arey, Analyte quantification with comprehensive two-dimensional gas chromatography: Assessment of methods for baseline correction, peak delineation, and matrix effect elimination for real samples, Journal of Chromatography A. 1375 (2015). https://doi.org/10.1016/j.chroma.2014.11.049.

[3] D.F. Thekkudan, S.C. Rutan, P.W. Carr, A study of the precision and accuracy of peak quantification in comprehensive two-dimensional liquid chromatography in time, Journal of Chromatography A. 1217 (2010). https://doi.org/10.1016/j.chroma.2010.04.039.

[4] I. Latha, S.E. Reichenbach, Q. Tao, Comparative analysis of peak-detection techniques for comprehensive two-dimensional chromatography, Journal of Chromatography A. 1218 (2011). https://doi.org/10.1016/j.chroma.2011.07.052.

[5] G. Vivó-Truyols, H.G. Janssen, Probability of failure of the watershed algorithm for peak detection in comprehensive two-dimensional chromatography, Journal of Chromatography A. 1217 (2010). https://doi.org/10.1016/j.chroma.2009.12.063.

[6] S. Peters, G. Vivó-Truyols, P.J. Marriott, P.J. Schoenmakers, Development of an algorithm for peak detection in comprehensive two-dimensional chromatography, Journal of Chromatography A. 1156 (2007). https://doi.org/10.1016/j.chroma.2006.10.066.

[7] G. Vivó-Truyols, J.R. Torres-Lapasió, A.M. van Nederkassel, Y. vander Heyden, D.L. Massart, Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals: Part I: Peak detection, Journal of Chromatography A. 1096 (2005). https://doi.org/10.1016/j.chroma.2005.03.092.

[8] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry. 36 (1964). https://doi.org/10.1021/ac60214a047.

[9] Y. Liu, X. Zhou, Y. Yu, A concise iterative method using the Bezier technique for baseline construction, Analyst. 140 (2015). https://doi.org/10.1039/c5an01184a.

[10] Z. Li, D.J. Zhan, J.J. Wang, J. Huang, Q.S. Xu, Z.M. Zhang, Y.B. Zheng, Y.Z. Liang, H. Wang, Morphological weighted penalized least squares for background correction, Analyst. 138 (2013). https://doi.org/10.1039/c3an00743j.

[11] V. Mazet, C. Carteret, D. Brie, J. Idier, B. Humbert, Background removal from spectra by designing and minimising a non-quadratic cost function, Chemometrics and Intelligent Laboratory Systems. 76 (2005). https://doi.org/10.1016/j.chemolab.2004.10.003.

[12] V. Mazet, D. Brie, J. Idier, Baseline spectrum estimation using half-quadratic minimization, in: European Signal Processing Conference, 2015.

[13] J.A. Navarro-Huerta, J.R. Torres-Lapasió, S. López-Ureña, M.C. García-Alvarez-Coque, Assisted baseline subtraction in complex chromatograms using the BEADS algorithm, Journal of Chromatography A. 1507 (2017). https://doi.org/10.1016/j.chroma.2017.05.057.

[14] X. Ning, I.W. Selesnick, L. Duval, Chromatogram baseline estimation and denoising using sparsity (BEADS), Chemometrics and Intelligent Laboratory Systems. 139 (2014). https://doi.org/10.1016/j.chemolab.2014.09.014.

[15] H.Y. Fu, H.D. Li, Y.J. Yu, B. Wang, P. Lu, H.P. Cui, P.P. Liu, Y. bin She, Simple automatic strategy for background drift correction in chromatographic data analysis, Journal of Chromatography A. 1449 (2016). https://doi.org/10.1016/j.chroma.2016.04.054.

[16] P.H.C. Eilers, H.F.M. Boelens, Baseline Correction with Asymmetric Least Squares Smoothing, Life Sciences. (2005).

[17] P.H.C. Eilers, A perfect smoother, Analytical Chemistry. 75 (2003). https://doi.org/10.1021/ac034173t.

[18] J.J. de Rooi, P.H.C. Eilers, Mixture models for baseline estimation, Chemometrics and Intelligent Laboratory Systems. 117 (2012). https://doi.org/10.1016/j.chemolab.2011.11.001.

[19] Z.M. Zhang, S. Chen, Y.Z. Liang, Baseline correction using adaptive iteratively reweighted penalized least squares, Analyst. 135 (2010). https://doi.org/10.1039/b922045c.

[20] S.E. Reichenbach, M. Ni, D. Zhang, E.B. Ledford, Image background removal in comprehensive two-dimensional gas chromatography, in: Journal of Chromatography A, 2003. https://doi.org/10.1016/S0021-9673(02)01498-X.

[21] I. Selesnick, Sparsity-assisted signal smoothing (revisited), in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2017. https://doi.org/10.1109/ICASSP.2017.7953017.

[22] F. Vogt, Data filtering in instrumental analyses with applications to optical spectroscopy and chemical imaging, Journal of Chemical Education. 88 (2011). https://doi.org/10.1021/ed100984c.

[23] D.F. Thekkudan, S.C. Rutan, Denoising and Signal-to-Noise Ratio Enhancement: Classical Filtering, in: Comprehensive Chemometrics, 2009. https://doi.org/10.1016/b978-0-444-64165-6.02002-4.

[24] S.-J. Baek, A. Park, Y.-J. Ahn, J. Choo, Baseline correction using asymmetrically reweighted penalized least squares smoothing, The Analyst. 140 (2015). https://doi.org/10.1039/c4an01061b.

[25] T.S. Bos, W.C. Knol, S.R.A. Molenaar, L.E. Niezen, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Recent applications of chemometrics in one- and two-dimensional chromatography, Journal of Separation Science. 43 (2020). https://doi.org/10.1002/jssc.202000011.

[26] C.A. Lieber, A. Mahadevan-Jansen, Automated Method for Subtraction of Fluorescence from Biological Raman Spectra, Applied Spectroscopy. 57 (2003). https://doi.org/10.1366/000370203322554518.

[27] F. Gan, G. Ruan, J. Mo, Baseline correction by improved iterative polynomial fitting with automatic threshold, Chemometrics and Intelligent Laboratory Systems. 82 (2006). https://doi.org/10.1016/j.chemolab.2005.08.009.

[28] S. Cappadona, F. Levander, M. Jansson, P. James, S. Cerutti, L. Pattini, Wavelet-based method for noise characterization and rejection in high-performance liquid chromatography coupled to mass spectrometry, Analytical Chemistry. 80 (2008). https://doi.org/10.1021/ac800166w.

[29] M.F. Wahab, T.C. O'Haver, Wavelet transforms in separation science for denoising and peak overlap detection, Journal of Separation Science. 43 (2020). https://doi.org/10.1002/jssc.202000013.

[30] A. Mani-Varnosfaderani, A. Kanginejad, K. Gilany, A. Valadkhani, Estimating complicated baselines in analytical signals using the iterative training of Bayesian regularized artificial neural networks, Analytica Chimica Acta. 940 (2016). https://doi.org/10.1016/j.aca.2016.08.046.

[31] A. Kensert, G. Collaerts, K. Efthymiadis, P. van Broeck, G. Desmet, D. Cabooter, Deep convolutional autoencoder for the simultaneous removal of baseline noise and baseline drift in chromatograms, Journal of Chromatography A. 1646 (2021). https://doi.org/10.1016/j.chroma.2021.462093.

[32] E.T. Whittaker, On a New Method of Graduation, Proceedings of the Edinburgh Mathematical Society. 41 (1922). https://doi.org/10.1017/s0013091500077853.

[33] C.M. Rader, B. Gold, Digital Filter Design Techniques in the Frequency Domain, Proceedings of the IEEE. 55 (1967). https://doi.org/10.1109/PROC.1967.5434.

[34] A. Graps, An Introduction to Wavelets, IEEE Computational Science and Engineering. 2 (1995). https://doi.org/10.1109/99.388960.

[35] J. Carlos Cobas, M.A. Bernstein, M. Martín-Pastor, P.G. Tahoces, A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data, Journal of Magnetic Resonance. 183 (2006). https://doi.org/10.1016/j.jmr.2006.07.013.

[36] Z.M. Zhang, S. Chen, Y.Z. Liang, Z.X. Liu, Q.M. Zhang, L.X. Ding, F. Ye, H. Zhou, An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy, Journal of Raman Spectroscopy. 41 (2010). https://doi.org/10.1002/jrs.2500.

[37] M. Daszykowski, K. Kaczmarek, Y. vander Heyden, B. Walczak, Robust statistics in data analysis - A review. Basic concepts, Chemometrics and Intelligent Laboratory Systems. 85 (2007). https://doi.org/10.1016/j.chemolab.2006.06.016.

[38] G. Guiochon, Attila Felinger, D. Golshan-Shirazi, A.M. Katti, Fundamentals of preparative and nonlinear chromatography, Second, Elsevier , San Diego , 2006.

[39] M.L. Phillips, R.L. White, Dependence of Chromatogram Peak Areas Obtained by Curve-Fitting on the Choice of Peak Shape Function, Journal of Chromatographic Science. 35 (1997). https://doi.org/10.1093/chromsci/35.2.75.

[40] J. Olivé, J.O. Grimalt, Gram-Charlier and Edgeworth-Cramér series in the characterization of chromatographic peaks, Analytica Chimica Acta. 249 (1991). https://doi.org/10.1016/S0003-2670(00)83005-6.

[41] J. Grimalt, H. Iturriaga, J. Olive, An experimental study of the efficiency of different statistical functions for the resolution of chromatograms with overlapping peaks, Analytica Chimica Acta. 201 (1987). https://doi.org/10.1016/S0003-2670(00)85337-4.

[42] Y. Kalambet, Y. Kozmin, K. Mikhailova, I. Nagaev, P. Tikhonov, Reconstruction of chromatographic peaks using the exponentially modified Gaussian function, Journal of Chemometrics. 25 (2011). https://doi.org/10.1002/cem.1343.

[43] Gottl.Friedr. Lipps, E.B. T., G.T. Fechner, Kollektivmasslehre., The Philosophical Review. 7 (1898). https://doi.org/10.2307/2177148.

[44] K.F. Wallis, The two-piece normal, binormal, or double gaussian distribution: Its origin and rediscoveries, Statistical Science. 29 (2014). https://doi.org/10.1214/13-STS417.

[45] IX. Mathematical contributions to the theory of evolution.—XIX. Second supplement to a memoir on skew variation, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character. 216 (1916). https://doi.org/10.1098/rsta.1916.0009.

[46] G.R. McGowan, M.A. Langhorst, Development and application of an integrated, high-speed, computerized hydrodynamic chromatograph, Journal of Colloid And Interface Science. 89 (1982). https://doi.org/10.1016/0021-9797(82)90124-2.

[47] A. Felinger, Data Analysis and Signal Processing in Chromatography , First, Amsterdam, 1998.

Chapter 8

Conclusions

Conclusions, challenges and recommendations

The focus of this thesis has been on developing new methods or improving existing methods, for the analysis of polymer distributions. Additionally, research has been performed that is more broadly applicable to any liquid chromatography (LC) separation, such as on the correction for gradient deformation and the comparison of background correction algorithms.

The first chapters of this thesis, Chapters 1 through 4, were primarily focused on novel liquid chromatographic methods to assess polymer distributions and the underlying theory of LC applied to (synthetic) polymers. Techniques such as thermal modulation in comprehensive two-dimensional liquid chromatography (combining reversed-phase liquid chromatography and size-exclusion chromatography, RPLC×SEC), recycling gradient-elution chromatography (LC↻LC) and gradient size-exclusion chromatography (gSEC), as well as SEC×gSEC, were investigated and possible applications were highlighted.

The later chapters, Chapters 5, 6 and 7, focused on chemometric strategies that aim to improve the ease-of-interpretation of chromatographic data, either by correcting for gradient deformation, or by applying algorithms for noise or drift removal to aid peak detection.

This final chapter serves to highlight the conclusions from the work in this thesis, to give recommendations for future studies based on these conclusions, to discuss preliminary work on subjects beyond the scope of this thesis, and to describe challenges that were not yet overcome or published. The challenges related to each chapter are shortly discussed.

8.1. Novel techniques to determine polymer distributions

8.1.1. Describing the retention of polymers

In Chapter 1 several models were introduced that aim to describe the retention of polymers. These were the linear-solvent strength (LSS) model, the polymer model (PM) and the Brun model. The use of the LSS and Brun models to predict the retention of polystyrene (PS) polymer standards in gradient-elution RPLC was investigated. Predictions were made on columns packed with stationary phases that featured different pore sizes. The models were fitted to the retention times or elution compositions in three gradients using **Equation 8.1** and **Equation 8.2**, which give the retention time under gradient conditions according to the LSS model [1] and the elution composition according to the Brun model [2], respectively.

$$t_R = \frac{1}{\varphi' S}\ln\left[1 + S\varphi' k_{\text{init}}\left(t_0 - \frac{(t_{\text{dwell}} + t_{\text{init}})}{k_{\text{init}}}\right)\right] + t_{\text{dwell}} + t_{\text{init}} + t_0 \quad (8.1)$$

$$\varphi_e = \varphi_{\text{crit}} + \left(\frac{\varphi'}{Q}\right)\ln[1 - e^{-Q}] \quad (8.2)$$

Where φ' is the gradient steepness or $\varphi' = \frac{d\varphi}{dt}$, k_0 is the retention factor extrapolated to 100% weak solvent ($\varphi = 0$), k_{init} is the retention factor at the initial mobile-phase composition, S is a slope parameter that determines how quickly k changes with φ; t_{dwell}, t_{init}, and t_0 are the dwell, programmed-initial-delay, and void times, respectively, and φ_e and φ_{crit} are the elution and critical mobile-phase compositions, respectively. In the Brun model the influence of molecular weight is captured in the Q parameter through the radius of gyration (R_g) through $Q = R_g^2 \frac{2}{Da}\frac{dc}{d\varphi}\varphi'$, where D, a, and c are an average pore diameter, the width of the interaction layer of the stationary phase, and an interaction parameter (for a particular repeating unit), respectively. When R_g varies with molecular weight as $R_g \sim \sqrt{M}$ (a relationship that corresponds relatively well to experimental data [3]), then all of the other parameters in Q may be lumped into a single A parameter, and so $Q = AM$. This means that for a polymer of a particular chemical composition that varies in molecular weight, φ_e depends solely on the parameters A and φ_{crit}. In the LSS model the elution time for each standard is fitted separately. This means that a much greater number of parameters (two for each standard) is available to describe the elution time than when the Brun model is used. **Equation 8.1** and **Equation 8.2** are fitted to the

measured elution or retention time/mobile-phase composition from two gradient experiments. Three gradient experiments were performed. For the LSS model the shortest and longest gradient duration were used to fit the model, while the gradient of intermediate duration was used to assess the quality of the prediction and was not used for fitting. The Brun model was fitted to φ_e of standards featuring different M. Each gradient was fitted separately. The fits for each model, presented as φ_e *vs.* M_w are shown in **Figure 8.1**. The data were obtained using a Phenomenex C_{18} column (4.6 × 150 mm) containing 100-Å superficially porous particles of 3.5 µm diameter and running gradients from 5 to 95% THF in ACN at a flowrate of 1 mL·min^{-1}. Gradient durations were varied from 10 to 15, and 20 min. Other columns featuring fully porous particles of larger pore size (300 and 500-Å) were also tested. However, for those columns the fit shown in **Figure 8.1-B** became progressively worse.

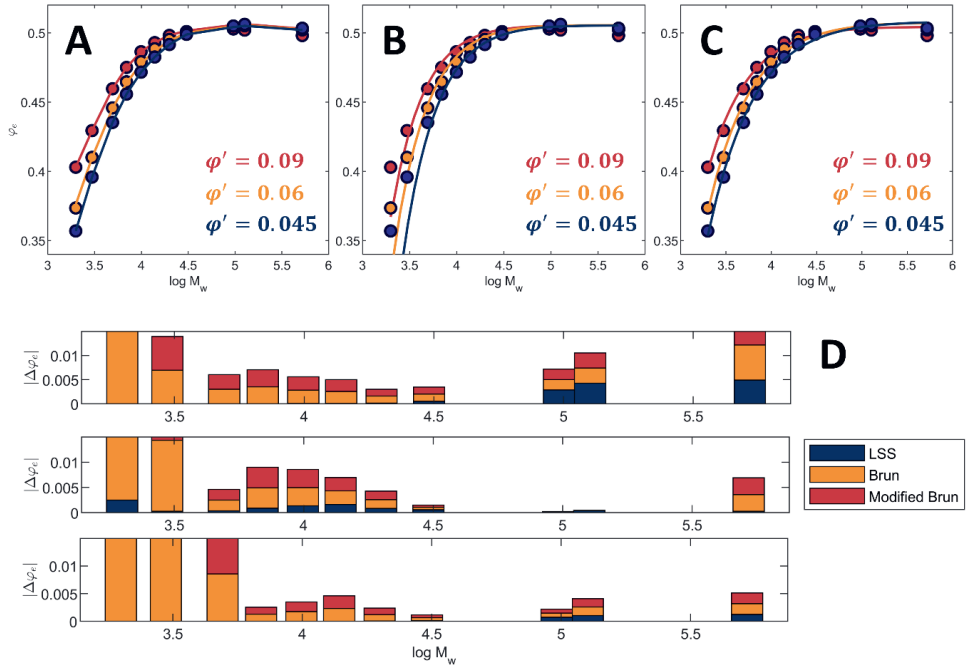

Figure 8.1: Measured elution composition (filled circles) and predicted elution composition (lines) using A) the LSS model (Equation 8.1), B) the Brun model (Equation 8.2) or C) a modified Brun Model. D) Absolute difference between predicted and measured elution composition for, from top to bottom, φ' of 0.09, 0.06 and 0.045 presented as a stacked bar plot.

As shown, in all cases the LSS model yielded a better fit. This is likely because of the much greater number of parameters. In the Brun model all of the standards of the same chemical composition, but different molecular weight, are described using only two parameters, which severely reduces the flexibility of the model. It is likely that this reduced flexibility is the cause for the worse predictions, especially for the lower-molecular-weight standards. One of the ways the model may be improved is by not assuming that the relationship $R_g \sim \sqrt{M}$ is valid, but instead fitting a relationship where R_g is allowed to vary differently with M. Instead of using $Q = AM$, a slightly different relationship, $Q = AM^b$, is used. The fits obtained using the additional parameter allowed for as much as a ten-fold improvement (for the shallowest gradient where retention occurs furthest from φ_{crit}) in the fit. These results are illustrated in **Figure 8.1-C**. However, as can be seen this model still cannot account for the earlier elution of the high-molecular-weight standards.

It is also possible to use the polymer model (PM), which typically yields better predictions [4]. However, this is a three parameter model and, as with the LSS model, the elution composition is predicted separately for each standard, rather than for a type of polymer. We conclude that, as of now, there is not a suitable and universal model that can (exactly) describe the elution composition of a polymer of a particular chemical composition as a function of M. This may be due to several factors, for example a change in the shape of the polymer with M, which also depends on its chemical composition (and sequence) and changes how the polymer interacts with the stationary phase within a pore. Additionally, differences in accessible stationary-phase volume and differences in the mobile-phase composition inside and outside of the porous structure of the stationary phase can all play a role. This will be even more complicated for copolymers or for homopolymers that feature pronounced end groups or branching. Nevertheless, the LSS predictions show that a homologues series can still be modelled relatively well, as long as each molecular-weight fraction is treated separately.

8.1.2. Thermal modulation

In Chapter 2 an increase in analyte retention with a decrease in temperature was exploited to focus polymers on small, so-called, "trap" columns, as are also commonly used in stationary-phase-assisted modulation (SPAM) [5]. Unlike SPAM,

thermal modulation does not require dilution flows or traps packed with highly retentive stationary phases. However, the use of thermal trapping faces its own challenges and issues. One of the most important of these is the bias towards the trapping of high-molecular-weight, more-hydrophobic analytes. Additionally, proper temperature control is required, which may increase the complexity of the set-up.

In the original work (Chapter 2) the cooling of the trap columns was achieved by means of a custom set-up, which was relatively bulky, as it required a large aluminium block holding the trap columns, and a separate device to pump cold mineral oil through the block. After this original publication a few alternative set-ups were investigated.

The first incorporated Peltier heating elements to control the temperature of a copper block, which had two (milled-out) spaces that held the trap columns, as well as a thermostat probe. This set-up was much smaller and could both heat and cool the trap columns electrically, so that mineral oil was not required. Furthermore, a larger range in temperature could be reached, from approximately -20°C to above 90°C. We did not attempt to reach lower or higher temperatures, although it should be relatively straightforward to expand this range by using more-powerful Peltier elements, and especially by stacking said elements on top of each other. The set-up as tested is illustrated in **Figure 8.2**

While it is likely that this set-up can also be used for thermal trapping, and reach an actual trap-column temperature closer to the programmed temperature, due to the difference in the block material (copper rather than aluminium), its operation for LC×LC was not evaluated in-depth, due to time constraints.

Another, significantly easier-to-implement, set-up was also assessed. Here an Agilent two-compartment column oven was used, where one side was heated and the other cooled. Insulating material was placed between the two sides of the oven. The oven is limited to a range of temperature from 5°C to 80°C, but does not require custom equipment. This set-up is illustrated in **Figure 8.3**, along with some of the LC×LC separations obtained with it.

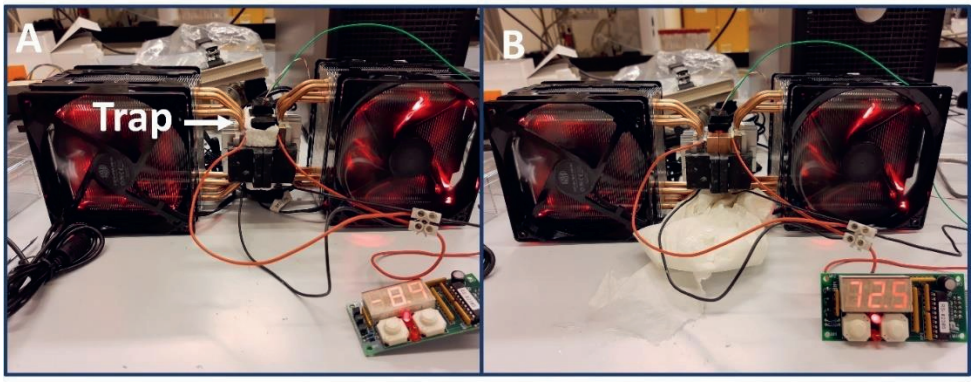

Figure 8.2: Alternative set-up for temperature control, achieved using high wattage (300 W) Peltier elements. This set-up could be used for both cooling (A) and heating (B).

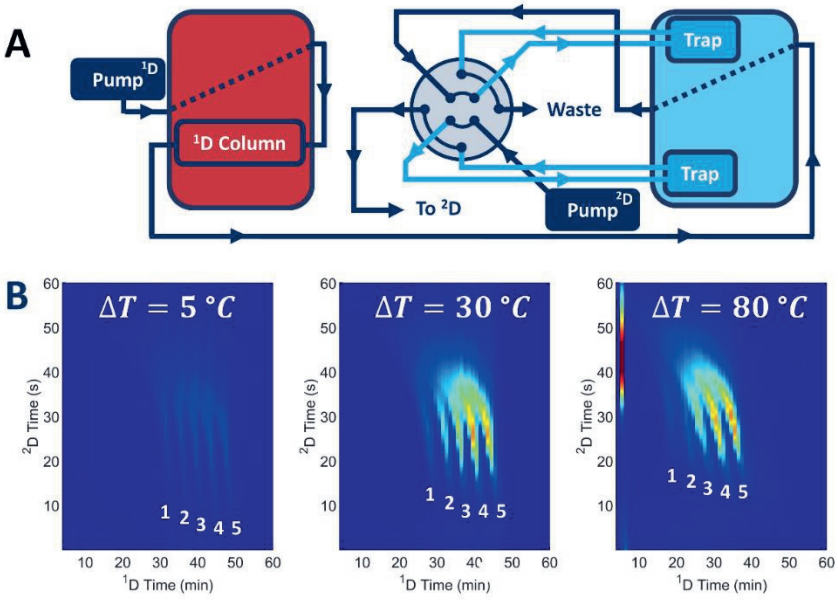

Figure 8.3: A) 2D-LC set-up for temperature modulation. Red and blue boxes indicate the heated and cooled parts of the oven, respectively; B) RPLC×SEC separation of statistical S/EA (average composition EA/S of 1: 80/20, 2: 65/35, 3: 50/50, 4: 35/65, and 5: 20/80) copolymers. ^{1}D column: XBridge BEH Shield RP18 XP (50 × 4.6 mm) column containing 2.5-µm, 130-Å pore size particles, ^{2}D column: two APC SEC (2.1 × 150mm) columns packed with 2.5-µm ethylene bridged-hybrid (BEH), 450-Å pore size particles. Trap temperature was kept constant at 5 °C, ^{1}D column temperature was varied.

Of the evaluated set-ups, this latter one is likely the most practical, because our experiments showed that most polymers do not require large differences in temperature to be successfully trapped. However, unlike the previous two set-ups, temperature control is likely worse in this set-up as the trap columns are heated and cooled primarily through the air inside the oven.

Irrespective of which set-up is used, the main flaw of thermal trapping, and any other trapping approach, is that there is a possibility that not all analytes are retained on the trap, so that quantification is jeopardized. Additionally, trap columns are challenging to consistently pack properly, and the packing stability over many 2D-LC separations, featuring high pressure that can vary quickly, will limit the repeatability of the method. Neither of these issues can be easily resolved. However, for polymer samples for which the chemical composition and molecular weight are not completely unknown, as is often the case, the thermal trapping approach may feasibly be used.

8.1.3. Recycle chromatography

In Chapter 3 recycling gradient-elution LC (LC↻LC) was described as a tool to assess the chemical-composition distribution (CCD) of a polymer. The method has several limitations, most of which may feasibly be worked around. The first potential issue is extra-column band broadening. In our set-up a photodiode array detector featuring a pressure-resistant (40 MPa) flow cell was used, which featured a relatively large volume of 8 µL. As the analyte will pass through this volume repeatedly over several cycles, such an extra-column volume may significantly degrade the performance of the separation, especially if narrow (small i.d.) columns are used. This limits the method to columns that feature relatively large internal diameters. To circumvent this issue the detector cell – and the associated connections – must be miniaturized, or other (highly-sensitive, non-destructive) detectors must be used. For non-UV-active polymers the choice for a suitable gradient-compatible detector that can be placed in-line is limited. However, for charged analytes the use of conductivity detection, such as a capacitively coupled conductivity detector (C^4D), is feasible. Even for non-charged analytes combining LC↻LC with a C^4D is interesting, as it allows one to monitor the gradient during the recycling experiment, as shown in chapter 5. Alternatively, the in-line detector may be removed completely, and only the final

cycle may be measured. In that case any gradient-compatible detector may be used (*e.g.* a mass-spectrometer or an evaporative light-scattering detector).

The second issue is that LC↻LC, as presented, necessitated the use of columns packed with low-surface-area stationary phases, which severely limits the number of columns that may be used. Stationary phases that feature larger surface areas may be used by increasing the column equilibration time, which may be achieved by reducing the gradient volume. This reduces the number of cycles that can be measured, but also reduces the number of cycles required to reach elution independent of molecular weight. Especially if the in-line detector is removed to mitigate the extra-column band broadening, a reduction in gradient volume will be less of an issue, because ultimately similar results may be obtained.

The third issue with LC↻LC is that more band broadening will occur for the lower-molecular-weight analytes than for the large analytes, which can hinder a correct assessment of the CCD. This molecular-weight-based broadening is difficult to eliminate, because small analytes also determine the required number of cycles (the duration) of the LC↻LC experiment. Furthermore, analytes cannot be refocused after the LC↻LC experiment, as this would likely reintroduce an influence of molecular weight. Therefore, analytes must be refocused during the experiment. To achieve this without compromising the results from the LC↻LC experiment, retention at φ_{init} must be high, followed by reducing the retention for all analytes. One way to achieve this is by using temperature. If the column is cooled for the first cycle, then all analytes will elute in a mobile-phase composition that is stronger than it would be had the column been heated. This does not reduce band broadening or the number of cycles required, because in a cold column φ_{crit} also shifts to a stronger mobile-phase composition, due to the increase in retention (*i.e.* it would require more cycles to reach). However, if the column is subsequently heated after the first cycle, then φ_{crit} will shift towards a weaker mobile-phase composition. Because the low-molecular-weight analytes will have eluted in a stronger mobile-phase composition in the (formerly) cold column they can now reach φ_{crit} more quickly, which reduces the time available for band broadening. Because the larger analytes will have eluted in a mobile-phase composition that correspond to SEC conditions after heating, these will catch up with the gradient and refocus towards φ_{crit}. This strategy relies on two aspects. Firstly, for all analytes the influence of temperature on retention must

be consistent, *i.e.* it must be either increasing or decreasing, and not both, within the same sample. Secondly, the largest analytes, which in a gradient experiment elute closest to φ_{crit}, must be able to refocus due to exclusion. Hence, this strategy will likely only work on a porous packing material. Alternatively, a similar effect may be obtained by placing a trap column, containing a more-retentive stationary phase than the columns used for the recycling, before the recycling set-up. However, this is a less-flexible strategy. The column temperature can be more easily adjusted to the MWD of the sample.

Irrespective of the above issues, there are certain applications of LC↻LC that have not yet been investigated. Perhaps the most interesting is the potential of the method to obtain information regarding a potential sequence distribution. According to the theory discussed in Chapter 1 of this thesis, focusing around the critical composition only occurs for statistical copolymers [2]. Therefore, the average composition over the molecular-weight distribution (MWD) within a slice taken from the CCD obtained by LC↻LC, should be more-or-less constant. If focusing is absent, as is expected for blocky copolymers, then the average composition over the MWD should not be constant within a fraction. For such copolymers earlier fractions should show more deviation in average composition in the lower-molecular-weight region, while later fractions should show deviations in the higher-molecular-weight region. For certain copolymers this experiment may be performed by taking fractions from the final cycle of the LC↻LC experiment, and subjecting these to SEC (with refractive-index or UV detection) to assess the MWD and determine the change in average composition within each fraction. Ideally a comprehensive LC↻LC ×SEC experiment could be performed to obtain this information within one experiment. The easiest way to implement such a method may be to perform the LC↻LC experiment at normal conditions, after which the flowrate can be reduced to then send fractions to the ^{2}D SEC separation using a conventional LC×LC set-up. A different way may be to select a slightly incorrect cycle timing, so that each cycle a small part of the end, or the front, of the gradient can be cut and send to the ^{2}D. An in-between focusing step may also be incorporated, *e.g.* using trap columns, although this increases the risk of re-introducing the influence of the MWD on retention.

8.1.4. Gradient size-exclusion chromatography

In Chapter 4 the determination of the CCD by size-exclusion chromatography gradients or gradient size-exclusion chromatography (gSEC) was investigated. Several conclusions could be drawn based on this work, some – but not all – of which concur with the results presented in Chapter 3. Based on the conventional SEC experiments that were performed using progressively weaker mobile-phase compositions, the influence and relative magnitude of certain effects, such as retention or a change in polymer size, could be assessed. Additionally, it was shown that polymers that are too large to enter the porous packing experience less retention, due to the smaller accessible stationary-phase surface. For high-molecular-weight analytes such effects are hardly measurable, because retention scales strongly with analyte size. This may be a potential reason for the greater deviations in retention times at "critical conditions" on the large-pore packing compared to the non-porous packing observed in Chapter 3. However, it is also possible that the observations are due to small changes in mobile-phase composition inside and outside of the pores. Further experiments are required to properly assess the reason for the larger deviations around the critical point.

In our work gSEC was shown to be primarily advantageous in cases where breakthrough in conventional gradient-elution LC was problematic. However, there are potentially other applications for gSEC. If retention on the stationary phase is sufficiently strong, one can ensure that a polymer never encounters a solvent in which it precipitates. Therefore, gSEC may be useful for specific polymers that are slow to redissolve. As this is a kinetic process, such polymers may show abnormal behaviour when separated using conventional gradient-elution-LC methods [6]. Such anomalous effects are expected to be more severe for steeper gradients.

8.2. Correction strategies applicable to chromatographic data

8.2.1. Capacitively coupled contactless conductivity detection

In Chapter 5 the use of a C^4D to measure gradient deformation was demonstrated. The C^4D, which measures the capacitance of the solution, could be used to detect changes in the bulk solvent composition. Since the C^4D was not limited by pressure and added almost no additional (dwell) volume to the system, the detector was ideal for this purpose. The C^4D may also feasibly be used for several other applications. As

mentioned previously, one of these may be LC↻LC. Since the C^4D does not add dwell volume the distortion of the gradient in successive cycles may be monitored. This approach may also feasibly be used to obtain insights in the re-equilibration of different stationary phases. This may be particularly useful for hydrophilic-interaction liquid chromatography (HILIC). Additionally, the C^4D may be used to measure the effect of every component in an LC instrument on the gradient profile or to (relatively easily) measure and compare the performance of different mixing units when combinations of aqueous and organic solvents are used.

8.2.2. Background-drift-correction and noise-removal algorithms

In Chapter 7 several background-drift-correction and noise-removal algorithms, some of which were reviewed in Chapter 6, were compared using generated data representative of complex real data. The greatest challenge to the practical usage of this approach is the extremely large number of combinations of background, peaks, and noise that may be encountered during an experiment. Hence, all possible real data can never be completely covered by generated data. Nevertheless, the approach provided useful insights regarding the performance of each algorithm and revealed which algorithms could be used to most-consistently achieve good results. Because LC×LC is becoming increasingly common and often results in greater data complexity, a study such as performed in Chapter 7, but instead focused on two-dimensional liquid or gas chromatography, may be especially interesting. While two-dimensional data adds complexity, it may also provide additional information that can actually improve the performance of the algorithms. As an example, in the case of 1D-LC it can be challenging to correct for underlying drift when the number of data points that contain information on relevant peaks is large compared to the number of points that reflect the background drift. In 2D-LC this problem will be smaller for samples of similar complexity. The larger separation space will not only provide more information on relevant signals, but also on the background, unless the sample complexity becomes so great that even 2D-LC does not provide sufficient separation space. Especially when the number of peaks is much greater in certain modulations than in other ones, the information of the background in surrounding modulations that contain fewer peaks may be used to obtain a reasonable guess for the background in those modulations where the background is difficult to correct.

8.3. The future of LC×LC and the development of chemometric strategies for the analysis of polymer distributions

As of now the use of 2D-LC in industry is still limited, although especially heart-cut 2D-LC has seen a dramatic increase in usage in recent years. The lack of acceptance is likely because of the (often repeated) notion that the technique is complex, too time-consuming to optimize, not sufficiently repeatable and/or reproducible, or simply because the data are more challenging to correctly interpret. All of these are valid points. However, during the last two decades a significant amount of work has been performed to allow for the use of 2D-LC in industry. Dedicated instrumentation and user software has become available, and chemometrics strategies have been – and are still being – developed to improve data interpretation and to reduce method-development time. All these developments make the use of 2D-LC much easier than it has ever been. Especially the introduction and further development of algorithms that can automate the entire method-development procedure [7] is expected to significantly improve the ability of a user to quickly and easily find a suitable method for a problem. Even if the resulting method may not be truly "optimal" in many cases, it likely suffices. In this way it becomes much more cost-efficient to automatically and quickly obtain a reasonable separation. Furthermore, from the iteration process of such an algorithm valuable information can be obtained that may serve as a good starting point for a "manual" fine-tuning of the separation, if needed. However, the fully automated method-development strategy faces several challenges, which are especially steep for synthetic polymers. Such polymer samples often feature an incredibly large number of very similar analytes that can also vary greatly in terms of molecular-weight. Both aspects complicate the use of mass-spectrometry, which is required to track analytes across chromatograms. Additionally, predicting the retention behaviour for polymers will be complex because even if analytes can be tracked, describing their retention consistently over multiple experiments is more difficult due to their strong dependence on very small changes in mobile-phase composition and temperature, and because of additional effects such as solubility. Finally, it is more challenging to define an optimum, because usually the type of information that is obtained is very important, which for polymers is difficult to capture in commonly used parameters such as the peak capacity or the resolution.

References

[1] P.J. Schoenmakers, F. Fitzpatrick, R. Grothey, Predicting the behaviour of polydisperse polymers in liquid chromatography under isocratic and gradient conditions, J Chromatogr A. 965 (2002) 93–107. https://doi.org/10.1016/S0021-9673(01)01322-X.

[2] Y. Brun, P. Foster, Characterization of synthetic copolymers by interaction polymer chromatography: Separation by microstructure, J Sep Sci. 33 (2010) 3501–3510. https://doi.org/10.1002/jssc.201000572.

[3] L.J. Fetters, N. Hadjichristidis, J.S. Lindner, J.W. Mays, Molecular Weight Dependence of Hydrodynamic and Thermodynamic Properties for Well-Defined Linear Polymers in Solution, J Phys Chem Ref Data. 23 (1994). https://doi.org/10.1063/1.555949.

[4] M.A. Bashir, W. Radke, Comparison of retention models for polymers. 1. Poly(ethylene glycol)s, J Chromatogr A. 1131 (2006) 130–141. https://doi.org/10.1016/j.chroma.2006.07.089.

[5] R.J. Vonk, A.F.G. Gargano, E. Davydova, H.L. Dekker, S. Eeltink, L.J. de Koning, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography with stationary-phase-assisted modulation coupled to high-resolution mass spectrometry applied to proteome analysis of saccharomyces cerevisiae, Anal Chem. 87 (2015). https://doi.org/10.1021/acs.analchem.5b00708.

[6] H.J.A. Philipsen, M. Oestreich, B. Klumperman, A.L. German, Characterization of low-molar-mass polymers by gradient polymer elution chromatography. III. Behaviour of crystalline polyesters under reversed-phase conditions, J Chromatogr A. 775 (1997). https://doi.org/10.1016/S0021-9673(97)00208-2.

[7] T.S. Bos, J. Boelrijk, S.R.A. Molenaar, B. van 't Veer, L.E. Niezen, D. van Herwerden, S. Samanipour, D.R. Stoll, P. Forré, B. Ensing, G.W. Somsen, B.W.J. Pirok, Chemometric Strategies for Fully Automated Interpretive Method Development in Liquid Chromatography, Anal Chem. 94 (2022) 16060–16068. https://doi.org/10.1021/acs.analchem.2c03160.

Chapter 9

Summary

Summary

The detailed analysis of polymeric materials is one of the necessary steps to elucidate the relationship between the chemical distributions of a polymer and the functional properties of a material. Within the UNMATCHED project (UNderstanding MATerials by CHaracterizing Essential Distributions) many techniques have been investigated – or further developed – to aid in the analysis of such materials. The primary focus of this thesis was on using liquid chromatography (LC) in innovative ways to analyze polymer chemical distributions. Additionally, chemometric strategies that help improve the interpretability of the data obtained from these methods were investigated and documented in later chapters.

In **Chapter 1** the relevant theory of liquid chromatography for small molecules and the different modes of polymer elution in isocratic and gradient-elution LC are described. For polymers the precipitation-redissolution and interaction mechanisms are briefly mentioned. Additionally, the role of the pore size of the packing and the retention behaviour of copolymers are briefly described. It is specified that the retention of a copolymer depends on the retention of the monomers that are incorporated and on how these monomers are distributed in the polymer chain (referred to as "blockiness" or "randomness"). Finally, challenges with respect to the retention modelling of such polymers are highlighted.

When multiple chemical distributions, such as a chemical-composition (CCD) and a molecular-weight distribution (MWD), are present within one sample, the use of two-dimensional liquid chromatography (2D-LC) is often beneficial. By combining separation methods, such as gradient-elution reversed-phase liquid chromatography (RPLC) and size-exclusion chromatography (SEC), information on both distributions may be obtained. In contemporary 2D-LC analytes are refocused, either at the start of the second-dimension (^{2}D) column, or within small (so-called trap) columns that replace the conventional sample loops. The latter technique is often referred to as stationary-phase-assisted modulation (SPAM), and often requires a dilution of the eluent coming from the first-dimension (^{1}D) with a weaker solvent to facilitate the trapping of the analytes. In **Chapter 2** an alternative to SPAM is presented, which does not require one to dilute the effluent. The technique termed

thermal modulation or cold-trapping was demonstrated to be a useful tool for 2D-LC separations of polymers, especially when SEC was used in the second dimension, because in that case dilution is especially undesirable. Since the effect of temperature on retention is much larger for high-molecular-weight analytes than for small molecules, it was shown that only a small difference in temperature already sufficed to trap nearly all large analytes. Additionally, a new modulation strategy was presented that reduces the pressure pulses during valve switching, which can potentially enhance the lifetime of the trap columns used.

A different approach to the analysis of copolymers that feature both a CCD and a MWD is to suppress the influence of one of these distributions on retention. In that case two one-dimensional liquid chromatography (1D-LC) separations can be sufficient to obtain the desired information. In **Chapter 3** a new technique termed recycling gradient-elution LC (LC↻LC) is described, which allows suppressing the effect of molecular weight on retention. The use of LC↻LC for the analysis of the CCD of copolymers was introduced and demonstrated. By continuously recycling the gradient, the effect of the MWD on elution could be minimized. Conventionally, very fast gradients require short durations, in combination with long columns and low flow rates, resulting in decreased peak capacities, long analysis times, and an increased risk of the system-induced gradient deformation that is described later in the thesis. Such issues could be avoided with LC↻LC and unbiased information on the CCD could be obtained

In **Chapter 4** the use of SEC gradients (or gradient-SEC, gSEC) was investigated as an alternative to RPLC for the chemical-composition analysis of polymers. The influence of the mobile-phase composition on the elution volume in SEC was explored on columns that greatly varied in terms of packing pore size. Both gSEC and RPLC were applied in 1D-LC and 2D-LC separations. It was shown that, compared to conventional gradient-elution LC, the use of SEC-gradients is primarily advantageous when breakthrough is an issue in RPLC. This is, for example, the case when comprehensive 2D-LC (LC×LC) is used with SEC is used in the first dimension, coupled to RPLC in the second dimension (SEC×RPLC). Molecular-weight-independent elution was shown to be much more challenging to achieve in gSEC, since the difference in migration velocity in gSEC is restricted to a factor of two (*i.e.*

total exclusion *vs.* total permeation). When implementing RPLC or gSEC as a ^{2}D separation technique in LC×LC, two-dimensional distributions (MWD×CCD) could be obtained. When gSEC was used breakthrough in the ^{2}D system could be avoided. However, in SEC×gSEC the residual molecular-weight dependence complicated a quantitative analysis.

In **Chapter 5** the effect of system-induced gradient deformation on retention-modelling approaches is discussed. Capacitively coupled "contactless" conductivity detection (C^{4}D) was demonstrated as a tool to measure solvent gradients. By using C^{4}D it was possible to measure gradients generated using various commonly used LC solvents. These included water and the organic solvents acetonitrile, methanol and tetrahydrofuran, to which no ionic tracers were added. In most cases, the detector response changed linearly with mobile-phase composition (*i.e.* the volume fraction of organic solvent). If needed slight non-linearities were corrected for. It was further shown that solvent- and instrument-dependent gradient profiles could be described using weighted response functions, which allowed us to predict the gradient distortion at different flowrates. Knowledge of the gradient distortion was then used to improve retention modelling, by approximating the distorted gradient as a segmented gradient consisting of 500 linear segments. Ultimately, nearly instrument-independent retention parameters could be obtained for systems that produced greatly different gradient profiles for identical gradient programs.

In **Chapter 6** an overview is given of the most-recent strategies used for the pre-processing of one-dimensional data. Different facets of pre-processing were reviewed, including smoothing, drift correction, and alignment strategies. It is emphasized that pre-processing is indispensable when the data are to be used for subsequent quantification or classification purposes. For background correction BEADS (baseline estimation and denoising using sparsity) and assisted BEADs were found to be highly promising recent developments, as these approaches seemed capable of handling many different types of background distortions and were fast. Along similar lines, the most-noteworthy strategies for peak alignment in two-dimensional chromatography were found to be those that operate not just in one-dimension, but in both simultaneously. Finally, it could be concluded that, although

there have been many interesting developments, it is often difficult to judge which methods truly perform best.

Based on the conclusions from Chapter 6, a data-simulation tool is described in **Chapter 7,** which enabled the comparison of different background-correction and peak-detection methods. The tool was used to compare a variety of data (pre-)processing methods. Drift-correction methods that were evaluated included the use of polynomial fitting using asymmetric cost functions (Backcor) and the use of asymmetrically penalized least squares (AsLS), asymmetrically reweighted penalized least squares (ArPLS), or adaptive iteratively reweighted penalized least squares (AirPLS). Additionally, the mixture model (MM) and the use of local minimum values (LMV) were evaluated. Noise-removal or smoothing methods that were evaluated included sparsity-assisted signal smoothing (SASS), Savitsky-Golay and Whittaker smoothing, and the use of low-pass or wavelet filters. The study showed that a combination of SASS and arPLS most often resulted in the lowest root-mean-square error and apparently provided the best results. However, this combination did not result in the smallest errors in peak areas. The combination of SASS and LMV methods performed best in this latter respect. In terms of speed the Backcor and LMV algorithms provided the fastest drift correction, generally with evaluation times of less than half a second, while the arPLS algorithm performed slowest. The best combination of methods seems to depend on the nature of the background, which implies that it cannot always be *a-priori* predicted. However, the present study has provided valuable tools and methods to improve quantification in case the true background is unknown.

In **Chapter 8** the conclusions from the work in this thesis were reviewed and recommendations for future studies were given. Additionally, preliminary work and challenges yet to be solved were discussed. Some feasible strategies were envisioned.

Samenvatting

De gedetailleerde analyse van polymeermaterialen is een van de benodigde stappen om de connectie tussen enerzijds de chemische verdelingen in het polymeer en anderzijds de functionele eigenschappen van het materiaal te verhelderen. In het UNMATCHED project (UNderstanding MATerials by CHaracterizing Essential Distributions) zijn veel technieken onderzocht – of verder ontwikkeld – om de analyse van dit soort materialen te realiseren. De focus van dit proefschrift lag op het gebruik van vloeistofchromatografie (LC) op nieuwe manieren om chemische verdelingen van polymeren te bepalen. In latere hoofdstukken wordt aanvullend werk beschreven, dat het gebruik van chemometrische methoden die de kwaliteit en/of de interpreteerbaarheid van de data kunnen verbeteren.

In **Hoofdstuk 1** worden zowel de relevante theorie van vloeistofchromatografie als het elutiegedrag van polymeren onder isocratische en gradiënt-elutie condities beschreven. Het verschil in precipitatie-/oplosbaarheid- en interactiemechanismes wordt verduidelijkt. Ook wordt de invloed van de kolompakking, met name de poriegrootte, op het retentiegedrag van co-polymeren kort beschreven. De retentie van een copolymeer blijkt zowel af te hangen van de verhouding van de monomeren, als van de verdeling van de monomeren in het polymeer molecuul. Tot slot worden er een aantal uitdagingen beschreven voor het nauwkeurig modelleren van de retentie van polymeren.

Als een polymeermonster wordt gekenmerkt door meerdere chemische verdelingen, zoals een chemische samenstellingsverdeling (CCD) en een moleculaire massaverdeling (MWD), kan het gebruik van tweedimensionale vloeistof chromatografie (2D-LC) voordelen bieden. Door het koppelen van scheidingsmethoden, zoals gradiënt-elutie *reversed-phase* ("omgekeerde fase") vloeistof chromatografie (RPLC) en size-exclusion chromatografie (SEC), kan er informatie over beide chemische verdelingen worden verkregen. In moderne 2D-LC scheidingen worden stoffen vaak geconcentreerd, hetzij aan het begin van de tweede dimensie (^{2}D) kolom of in kleine zogeheten "trap" kolommen, die de conventionele *sample loops* ("monsterlussen") vervangen. De laatstgenoemde methode wordt vaak aangeduid als *stationary-phase-assisted* modulatie (SPAM) en vereist vaak een verdunning van het eluens dat uit de eerstedimensie scheiding

wordt verkregen met een "zwakker" oplosmiddel, om het concentreren van analieten op de trap kolom te faciliteren. In **Hoofdstuk 2** wordt een alternatief voor SPAM beschreven dat geen extra verdunning vereist. Deze techniek wordt thermische modulatie of "*cold-trapping*" genoemd. Ze biedt voornamelijk voordelen voor 2D-LC scheidingen waarin SEC in de tweede dimensie gebruikt wordt, omdat in dat geval een verdunning van het eluens niet wenselijk is. Omdat de invloed van temperatuur op retentie vele malen groter is voor analieten met een hoog moleculair gewicht dan voor kleine moleculen kon worden aangetoond dat een klein verschil in temperatuur al voldoende is om grote analieten te concenteren op de trapkolom. Als aanvulling op dit werk werd er een nieuwe modulatiestrategie beschreven die leidde tot kleinere drukveranderingen, hetgeen belangrijk kan zijn voor de levensduur van de trapkolom.

Een alternatief voor de analyse van co-polymeren met zowel een relevante CCD als een MWD is het onderdrukken van de invloed van één van de chemische verdelingen op het scheidingsproces. In dat geval kunnen twee ééndimensionale scheidingen voldoende zijn om de benodigde informatie te verkrijgen. In **Hoofdstuk 3** wordt een nieuwe methode genaamd *recycling* gradiënt-elutie vloeistofchromatografie (LC↻LC) beschreven, die dit bewerkstelligt. Het gebruik van LC↻LC voor de analyze van de CCD van een copolymeer wordt gedemonstreerd. Door het doorlopend recyclen van de gradiënt kon de invloed van de MWD geminimaliseerd worden. Normaliter vereist dit erg snelle gradiënten van korte duur in combinatie met lange kolommen en lage stroomsnelheden. Dit leidt tot een vermindering in piek capaciteit, lange analysetijden, en een groter risico dat de gradiënt in het instrument vervormd wordt. Het laatste wordt in hoofdstuk 5 in dit proefschrift beschreven. Door middel van LC↻LC konden deze problemen vermeden worden.

In **Hoofdstuk 4** wordt het gebruik van zogeheten SEC-gradiënten (of gradiënt-SEC, gSEC) beschreven en vergeleken met RPLC voor de analyse van de CCD. Bovendien werd de invloed van de mobiele-fasesamenstelling op het elutievolume in SEC onderzocht voor een verscheidenheid van kolommen, die pakkingen bevatten met verschillende poriegroottes. Zowel gSEC als RPLC werden vervolgens toegepast voor 1D-LC en 2D-LC scheidingen. Aangetoond werd dat, in vergelijking met conventionele gradiënt RPLC methoden, het gebruik van gSEC vooral voordelen

biedt als er in RPLC doorbraak optreedt van de monsterplug. Dit is met name het geval voor *comprehensive* ("alomvattende") 2D-LC (LC×LC) scheidingen, waarbij SEC gebruikt wordt als ^{1}D scheiding en wordt gekoppeld aan RPLC als ^{2}D methode. Het bleek met gSEC moeilijker om een scheiding te realiseren, die niet beïnvloedt werd door het moleculegewicht van de analieten, omdat het verschil in migratiesnelheid maximaal een factor twee kan zijn (totale exclusie *vs.* totale permeatie). Door zowel gebruik te maken van RPLC en gSEC als ^{2}D methoden in LC×LC, kon informatie verkregen worden over tweedimensionale verdelingen (MWD×CCD). Met het gebruik van gSEC als ^{2}D methode kon monsterdoorbraak voorkomen worden, maar kwantitatieve bepalingen bleken ingewikkelder, vanwege de resterende invloed van het moleculegewicht op de scheiding.

In **Hoofdstuk 5** wordt de invloed van systeem-geïnduceerde gradiënt vervorming op het modelleren van retentie benadrukt. Het gebruik van capacitief gekoppelde "contactloze" geleidbaarheidsdetectie ($C^{4}D$) voor het meten van oplosmiddelgradiënten wordt beschreven. De $C^{4}D$ maakte het meten mogelijk van verschillende veelgebruikte oplosmiddelen, zoals water en de organische oplosmiddelen acetonitril, methanol en tetrahydrofuraan, zonder dat daaraan ionische stoffen werden toegevoegd,. In veel gevallen veranderde de respons van de detector lineair met veranderingen in de oplosmiddelsamenstelling, maar in het geval van een niet-lineaire verandering kon hiervoor gecorrigeerd worden. Ook werd aangetoond dat oplosmiddel- en instrumentafhankelijke gradiëntprofielen beschreven konden worden met gewogen responsfuncties. Dit soort functies maakte het mogelijk om de gradiëntvervorming voor verschillende stroomsnelheden te voorspellen. Informatie over de gradiëntvervorming kon vervolgens gebruikt worden voor het verbeteren van retentiemodellen. Hiervoor werd de vervormde gradiënt beschreven als een samengestelde gradiënt bestaande uit 500 lineaire segmenten. Uiteindelijk konden met deze methode min-of-meer systeemonafhankelijke retentieparameters bepaald worden voor systemen waarvan de gradiëntprofielen zeer verschillend waren voor gelijke gradiëntprogramma's.

In **Hoofdstuk 6** word een overzicht gegeven van de meest recent geïntroduceerde datavoorbewerkingstrategieën voor hoofdzakelijk eendimensionale data. Verschillende aspecten van de datavoorbewerking worden beschreven, zoals

ruisverwijderingstechnieken, correctie van zogenoemde "drift" (verlopende basislijn), en *alignment* ("uitlijn") methodes. Er kon worden geconcludeerd dat datavoorbewerkingsmethoden onmisbaar zijn als de data gebruikt worden voor verdere kwantificering of classificatie. Vooral een achtergrondcorrectiemethode genaamd *Baseline Estimation And Denoising using Sparsity* ("achtergrondbenadering en ruisverwijdering door middel van schaarsheid", BEADS) en geassisteerde BEADS methoden bleken zeer interessante recente ontwikkelingen te zijn, omdat deze technieken veel verschillende soorten achtergronden snel konden corrigeren. De meest belangrijke *alignment* technieken waren die methoden die zich richtte op tweedimensionale scheidingen. De conclusie was dat, hoewel er veel interessante recente ontwikkelingen waren, het vaak lastig was om de kwaliteit van de methoden te bepalen.

Gebaseerd op de conclusies in Hoofdstuk 6 is een data-simulatie programma ontwikkeld, waarmee verschillende achtergrondcorrectie- en piekdetectiemethoden objectief vergeleken kunnen worden. Dit is beschreven in **Hoofdstuk 7**. Het programma werd gebruikt om verschillende achtergrondcorrectie-algoritmen te vergelijken. Verschillende driftcorrectiemethoden werden vergeleken, zoals het gebruik van polynoomfitting met asymmetrische kostenfuncties ("*Backcor*"), "*asymmetrically penalized least squares* (AsLS)", "*asymmetrically reweighted penalized least squares* (ArPLS)", "*adaptive iteratively reweighted penalized least squares* (AirPLS)" en het gebruik van een mengmodel (MM) of lokale minimale waardes (LMV). Ook worden een aantal ruisverwijderingstechnieken beschreven en vergeleken, namelijk "*sparsity-assisted signal smoothing*" (SASS), "Savitsky-Golay" or "Whittaker smoothing" en het gebruik van *low-pass* of *wavelet* filters. Uit het vergelijken van de methoden bleek dat een combinatie van SASS en arPLS vaak de laagste gemiddelde kwadratische afwijking en de best ogende resultaten gaf. Deze combinatie resulteerde echter niet in de laagste afwijkingen voor het piekoppervlak. In dit opzicht werkte een combinatie van de SASS en LMV methoden beter. De *Backcor* en LMV algoritmen konden het snelst de drift corrigeren, vaak binnen een halve seconde, terwijl het arPLS algoritme het traagst was. De beste combinatie van methoden blijkt af te hangen van de feitelijke achtergrond, wat het moeilijk maakt om te voorspellen welke methoden het best zal werken. Desalniettemin zijn er in

deze studie waardevolle methoden ontwikkeld, die tot betere kwantificering kunnen leiden als de achtergrond niet tevoren bekend is.

In **Hoofdstuk 8** worden de conclusies van dit proefschrift nogmaals tegen het licht gehouden en worden aanbevelingen voor toekomstig onderzoek geformuleerd. Ook worden een aantal initiële resultaten, uitdagingen die niet opgelost werden in het kader van dit proefschrift., en strategieën die goed uitvoerbaar lijken te zijn beschreven.

Chapter 10

Sundries

10.1. List of Publications

1. **Thermal modulation to enhance two-dimensional liquid chromatography separations of polymers**

L.E. Niezen, B.B.P. Staal, C. Lang, B.W.J. Pirok, P.J. Schoenmakers

J. Chromatogr. A, 1653 (**2021**) 462429, https://doi.org/10.1016/j.chroma.2021.462429

Covered by Chapter 2

2. **Recycling gradient-elution liquid chromatography for the analysis of chemical-composition distributions of polymers**

L.E. Niezen, B.B.P. Staal, C. Lang, H.J.A. Philipsen, B.W.J. Pirok, G.W. Somsen, P.J. Schoenmakers

J. Chromatogr. A, 1679 (**2022**) 463386, https://doi.org/10.1016/j.chroma.2022.463386

Covered by Chapter 3

3. **Principles and potential of solvent gradient size-exclusion chromatography for polymer analysis**

L.E. Niezen*, J.D. Kruijswijk*, G. van Henten, B.W.J. Pirok, B.B.P. Staal, W. Radke, H.J.A. Philipsen, G.W. Somsen, and P.J. Schoenmakers

Anal Chim Acta, *accepted *Equal contribution*

Covered by Chapter 4

4. **Capacitively coupled contactless conductivity detection to account for system-induced gradient deformation in liquid chromatography**

L.E. Niezen*, T.S. Bos*, P.J. Schoenmakers G.W. Somsen and B.W.J. Pirok

Anal Chim Acta, *submitted, *Equal contribution*

Covered by Chapter 5

5. **Recent applications of chemometrics in one- and two-dimensional chromatography**

T.S. Bos*, W.C. Knol*, S.R.A. Molenaar*, L.E. Niezen*, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok

J. Sep. Sci. 43 (**2020**) 1678-1727, https://doi.org/10.1002/jssc.202000011, **Equal contribution*

Partially covered by Chapter 6

6. **Critical comparison of background correction algorithms used in chromatography**

L.E.Niezen, P.J. Schoenmakers, B.W.J. Pirok

Anal. Chim. Acta, 1201 (**2022**) 339605, https://doi.org/10.1016/j.aca.2022.339605

Covered by Chapter 7

7. **Chemometric Strategies for Fully Automated Interpretive Method Development in Liquid Chromatography**

T.S. Bos*, J. Boelrijk*, S.R.A. Molenaar*, B. van 't Veer, L.E. Niezen, D. van Herwerden, S. Samanipour, D.R. Stoll, P. Forré, B. Ensing, G.W. Somsen, B.W.J. Pirok

Anal Chem. (**2022**). https://doi.org/10.1021/acs.analchem.2c03160. **Equal contribution*

8. **Reducing the influence of geometry-induced gradient deformation in liquid chromatographic retention modelling**

T.S. Bos, L.E. Niezen, M.J. den Uijl, S.R.A. Molenaar, S. Lege, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok

J. Chromatogr. A, 1635 (**2021**) 461714, https://doi.org/10.1016/j.chroma.2020.461714

10.2. Overview of co-authors' contribution

Chapter 1: Introduction

Leon E. Niezen	Wrote the chapter
Peter J. Schoenmakers	Reviewed the chapter and provided feedback for improvements

Chapter 2: Thermal modulation to enhance two-dimensional liquid chromatography separations of polymers

J. Chromatogr. A, 1653 (**2021**) 462429, https://doi.org/10.1016/j.chroma.2021.462429

Leon E. Niezen	Co-developed the idea of this work. Designed the experimental and performed the experiments. Performed the data analysis. Wrote the manuscript and designed all the figures.
Bastiaan B.P. Staal	Co-developed the idea of this work. Reviewed the manuscript and provided feedback for improvements
Christiane Lang	Reviewed the manuscript and provided feedback for improvements
Bob W.J. Pirok	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Peter J. Schoenmakers	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.

Chapter 3: Recycling gradient-elution liquid chromatography for the analysis of chemical-composition distributions of polymers

J. Chromatogr. A, 1679 (**2022**) 463386, https://doi.org/10.1016/j.chroma.2022.463386

Leon E. Niezen	Co-developed the idea of this work. Designed the experimental and performed the experiments. Performed the data analysis. Wrote the manuscript and designed all the figures.
Bastiaan B.P. Staal	Co-developed the idea of this work. Performed part of the experiments. Reviewed the manuscript and provided feedback for improvements
Christiane Lang	Reviewed the manuscript and provided feedback for improvements
Harry J.A. Philipsen	Reviewed the manuscript and provided feedback for improvements
Bob W.J. Pirok	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Govert W. Somsen	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Peter J. Schoenmakers	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.

Chapter 4: Principles and potential of solvent gradient size-exclusion chromatography for polymer analysis

Anal Chim Acta, *accepted*

Leon E. Niezen	Co-developed the idea of this work. Designed part of the experimental and performed part of the experiments. Performed the data analysis. Wrote the manuscript and designed all the figures.
Jordy D. Kruijswijk	Co-developed the idea of this work. Designed part of the experimental and performed part of the experiments. Wrote the manuscript.
Gerben van Henten	Co-developed the idea of this work. Designed part of the experimental and performed part of the experiments. Performed part of the data analysis.
Bob W.J. Pirok	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Bastiaan B.P. Staal	Reviewed the manuscript and provided feedback for improvements.
Wolfgang Radke	Developed the gradient-SEC method. Reviewed the manuscript and provided feedback for improvements.
Harry J.A. Philipsen	Reviewed the manuscript and provided feedback for improvements.
Govert W. Somsen	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Peter J. Schoenmakers	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.

Chapter 5: Capacitively coupled contactless conductivity detection to account for system-induced gradient deformation in liquid chromatography

Anal Chim Acta, *submitted*

Leon E. Niezen	Co-developed the idea of this work. Designed part of the experimental and performed the experiments. Wrote the manuscript and designed all the figures.
Tijmen S. Bos	Co-developed the idea of this work. Designed part of the experimental. Performed the data analysis. Reviewed the manuscript and provided feedback for improvements
Peter J. Schoenmakers	Co-developed the idea of this work. Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Govert W. Somsen	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Bob W.J. Pirok	Co-developed the idea of this work. Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.

Chapter 6: Recent pre-processing strategies in one- and two-dimensional chromatography

J. Sep. Sci. 43 (**2020**) 1678-1727, https://doi.org/10.1002/jssc.202000011

Leon E. Niezen	Wrote the chapter, except for the introduction.
Peter J. Schoenmakers	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Govert W. Somsen	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Bob W.J. Pirok	Wrote the introduction, Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.

Chapter 7: Critical comparison of background correction algorithms used in chromatography

Anal. Chim. Acta, 1201 (**2022**) 339605, https://doi.org/10.1016/j.aca.2022.339605

Leon E. Niezen	Co-developed the idea of this work. Performed the data analysis and wrote the code. Wrote the manuscript and designed all the figures.
Peter J. Schoenmakers	Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.
Bob W.J. Pirok	Co-developed the idea of this work. Reviewed the manuscript and provided feedback for improvements. Overall supervisor of the project.

Chapter 8: Conclusions, challenges and recommendations

Leon E. Niezen	Wrote the chapter
Peter J. Schoenmakers	Reviewed the chapter and provided feedback for improvements

10.3. List of Abbreviations[4]

^{1}H-NMR	Proton nuclear magnetic resonance
^{1}D	First dimension
^{2}D	Second dimension
1D	One-dimensional
2D	Two-dimensional
1D-LC	One-dimensional liquid chromatography
2D-LC	Two-dimensional liquid chromatography
AC	Alternating current
ACN	Acetonitrile
ACPD	Automatic peak detection
AIC	Akaike information criterion
ALS	Alternating least-squares
AMBN	2,2'-Azodi(2-MethylButyroNitrile)
ANN	Artificial neural network
ANOVA	Analysis of Variance
ArPLS	Asymmetrically reweighted penalized least-squares
AirPLS	Adaptive iteratively reweighted penalized least squares
ASM	Active-solvent modulation
AsLS	asymmetrical least squares
ATLD	Alternating trilinear decomposition
ATSA	Automatic time-shift alignment
BDC	**B**ackground drift correction
BD-OSP	Background drift correction by orthogonal subspace projection
BEADS	Baseline estimation and denoising using sparsity
BEH	Bridged-ethylene hybrid
BLD	Block length distribution
BMA	Butyl methacrylate
BRANN	Bayesian regularized artificial neural networks
C^{4}D	Capacitively coupled “contactless” conductivity detection
CC	Corner cutting

[4] In case multiple meanings are given to an abbreviation, the additional meaning is only used in the indicated chapters.

CCD	Chemical composition distribution
CE	Capillary electrophoresis
COS	Cosine correlation
COSHIFT	Correlation-optimized shifting
COW	Correlation-optimized warping
DAD	Diode-array detector
DBD	Degree-of-branching distribution
DTW	Dynamic time warping
EA	Ethyl acrylate
ECG	Electrocardiography
ELSD	Evaporative light-scattering detector
EMG	Exponentially modified gaussian
FID	Flame-ionization detector
FIR	Finite-impulse-response
FTD	Functionality-type distribution
GC	Gas chromatography
GC×GC	Comprehensive two-dimensional gas chromatography
gSEC	Gradient size-exclusion chromatography
HILIC	Hydrophilic-interaction liquid chromatography
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass spectrometry
IEC	Ion-exchange chromatography
IPA	Isopropyl alcohol
IPC	Interaction polymer chromatography
i.d.	Internal diameter
iPF	Iterative polynomial fitting
L	Long-arm in star-block copolymer
LAC	Liquid adsorption chromatography
LC	Liquid chromatography
LCCC	Liquid chromatography at critical conditions
LC-LC	Heart-cut two-dimensional liquid chromatography
LC×LC	Comprehensive two-dimensional liquid chromatography
LC↻LC	Recycling gradient-elution liquid chromatography
LC-MS	Liquid chromatography mass spectrometry

LMV	Local minimum value
LMV-RSA	Local minimum values robust statistical analysis
LSS	Linear-solvent strength
MAD	Molecular architecture distribution
	Median absolute deviation (in Chapter 7)
MairPLS	Modified adaptive iteratively reweighted penalized least squares
MALDI	Matrix-assisted laser-desorption/ionization
MB	Copolymer of methyl methacrylate and butyl methacrylate
MCR	Multivariate curve resolution
MCR-ALS	Multivariate curve resolution alternating least-squares
MeOH	Methanol
mLC-LC	Multiple-heart-cut two-dimensional liquid chromatography
MM	Mixture model
MMA	Methyl methacrylate
MPLS	Morphologically weighted penalized least squares
MS	Mass spectrometry
MWD	Molecular-weight distribution
MWMV	Moving-window-minimum-value
NPLC	Normal-phase liquid chromatography
o.d.	Outer diameter
OSP	Orthogonal subspace projection
OSSP	Orthogonal spectral signal projection
PARAFAC	Parallel factor analysis
PB	Polybutadiene
PCA	Principal component analysis
PCC	Pearson correlation coefficient
PDI	Polydispersity index
PDR	Projected-difference-resolution
PFSM	Pump-frequency-synchronized modulation
PGC	Porous graphitic carbon
PLS	Penalized least squares
PLSDA	Partial least-squares discriminant analysis
PM	Polymer model
PS	Polystyrene

PTW	Parametric time warping
PW	Peak weighted
QM	Quadratic model
QTOF	Quadrupole time-of-flight
RF	Response function
RID	Refractive-index detection
RMSE	Root mean square error
RP	Reversed phase
RPLC	Reversed-phase liquid chromatography
S	Styrene
	Short-arm in star-block copolymer (in Chapter 2)
S/N	Signal-to-noise ratio
SASS	Sparsity-assisted signal smoothing
SEC	Size-exclusion chromatography
SG	Savitsky-Golay
sLC×LC	Selective comprehensive two-dimensional liquid chromatography
SM	Copolymer of styrene and methyl methacrylate
SPAM	Stationary-phase assisted modulation
SVD	Singular value decomposition
SWATLD	Self-weighted alternating trilinear decomposition
TASF	Temperature-assisted on-column solute focusing
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
ThFFF	Thermal field-flow-fractionation
TIC	Total ion current
TLC	Thin-layer chromatography
ToF	Time-of-flight
TPC	Total peak correlation coefficient
UHPLC	Ultra-high-performance liquid chromatography
VWD	Variable-wavelength detector
XRD	X-ray diffraction

10.4. List of Symbols[5]

$\hat{\alpha}$	Estimated P-splines coefficients or penalties
α	Selectivity Kurtosis ("tailedness") parameter in the stable distribution, standardized and centralized fourth moment (in Chapter 5) Coefficients (or penalties) of P-splines (in Chapter 7)
A	Area
A	Eddie dispersion term
A_s	Asymmetry parameter in the modified Pearson VII distribution
a	Thickness of a "monomolecular" adsorption layer
β	Column phase ratio Skewness (asymmetry) parameter in the stable distribution, standardized and centralized third moment (in Chapter 5)
$\mathbf{B}$	n × m cubic spline basis of m number of splines
B	Longitudinal diffusion term
b	Rate of change of the retention factor in a gradient per volume of mobile phase
b_i	The ith data point in the known background
C_F	Feedback capacitor
C_w	Capacitance due to capillary walls
C_S	Stray capacitance
C_L	Capacitance due to solution in the capillary
C_{18}	Octyldecylsilane
C	Mass-transfer term
C_s	Mass-transfer term in the stationary phase
C_m	Mass-transfer term in the mobile phase
C	Pure chromatographic profiles of the components in the sample
c	Interaction parameter
c_i	ith chromatographic profile representing the analyte
$c_{bk,i}$	ith chromatographic profile representing the background

[5] In case multiple meanings are given to a symbol, the additional meaning is only used in the indicated chapters.

$Đ_M$	Dispersity in molecular weight
δ_{AB}	Percentual difference in retention parameters before and after accounting for gradient deformation
δ	Mean (position) parameter in the stable distribution, first (raw) moment
$\mathbf{D}$	Derivative of the identity matrix
D_m	Diffusion coefficient in the mobile phase
D	Average pore diameter
dx_{new}	Derivative of the new initial estimate of the background
d_i	ith data point in the derivative of the new initial estimate of the background
d	Difference between fit and data (in Chapter 7)
d_t	Threshold difference
d^-	Negative part of the difference between input data and fit
d_p	Particle diameter
ε	Porosity
	Dielectric constant (in chapter 5)
ε_{AB}	Dielectric constant of solvent mixture consisting of two (A and B) solvents
ε_A	Dielectric constant of solvent A
ε_B	Dielectric constant of solvent B
E	Error matrix
F	Volumetric flowrate
f	Frequency
	Array containing background (in Chapter 6)
$\hat{f}$	Estimate of the background array
f_{coverage}	Correction factor for incomplete surface coverage in 2D-LC
$f_{\text{undersampling}}$	Correction factor for under-sampling
γ	Scale (width) parameter in the stable distribution, second centralized moment
$g(x\|\mu,\sigma)$	Normal density function
$\Delta G°$	Partial molar volume change in Gibbs free energy
ΔG^o_{mon}	Partial molar volume change in Gibbs free energy for a type of monomer unit

G	Gradient compression factor
	Conductance (in Chapter 5)
H	Plate height
	Peak height (in Chapter 7)
H	High-pass filter (in Chapter 6)
$\Delta H°$	Partial molar volume change in enthalpy
$h(x-\mu)$	Unknown density function
$\mathbf{I}$	Identity matrix (in Chapters 6 and 7)
I	Current
i	Imaginary number
	Iteration number (when used in large mathematical operators)
K	Distribution coefficient
	Number of variables (in Chapter 7)
K_{LC}	Distribution coefficient (LC)
K_{PM}	Distribution coefficient according to the polymer model
K_{SEC}	"Distribution" coefficient (SEC)
$k_{0,A}$	Retention factor extrapolated to 100% weak solvent, obtained before accounting for gradient deformation
$k_{0,B}$	Retention factor extrapolated to 100% weak solvent, obtained after accounting for gradient deformation
k_0	Retention factor extrapolated to 100% weak solvent
k	Retention factor
k_{init}	Retention factor at initial gradient conditions
k	Scaling factor (in Chapter 7)
k_{CP}	Retention factor of a copolymer calculated from the retention of two homopolymers
k_I	Retention factor of a homopolymer of monomer type I
k_{II}	Retention factor of a homopolymer of monomer type II
k_B	Boltzmann's constant
k_{crit}	Retention factor at critical conditions
k_e	Retention factor at the moment of elution
k_{mean}	Mean retention factor
λ	"Smoothing" parameter
L	Length

L	Low-pass filter (in Chapter 6)
μ	Mean
μ_d	Mean of d^-
$\mathbf{M}$	Square matrix of uniform basis functions
M_w	Weight-average molecular weight
M_n	Number-average molecular weight
M_p	Peak molecular weight
M	Molecular mass
m	Kurtosis parameter in the modified Pearson VII distribution
ν	Copolymer randomness or "blockiness"
N	Total number of data points (in Chapter 7)
N_{test}	Number of peaks in the test signal
N_{ref}	Number of peaks in the reference signal
N	Number of molecules
	Plate number
	Number of data points (in Chapter 7)
N_A	Avogadro's constant
N_I	Average chain length of monomer type I
N_{II}	Average chain length of monomer type II
n	Iteration index (in Chapter 7)
n_c	Peak capacity
$n_{c,\mathrm{eff}}$	Effective peak capacity
Ω	Number of potential microstates in the system
π	Ratio of a circle's circumference to its diameter
	Unknown mixing ratio (in Chapter 7)
Φ	Inverse of the phase ratio
φ	Volumetric fraction of the strong mobile-phase modifier
	Function cost (*e.g.* sum of squared errors, in Chapter 6 and 7)
φ_A	Volumetric fraction of solvent A
φ_B	Volumetric fraction of solvent B
φ_{crit}	Volumetric fraction of the strong mobile-phase modifier at critical conditions
$\varphi_{\mathrm{crit},CP}$	Volumetric fraction of the strong mobile-phase modifier at critical conditions for copolymer

φ_d	Actual volume fraction of strong solvent in a deformed gradient
φ_p	Programmed volume fraction of strong solvent.
φ_{init}	Fraction of strong mobile-phase modifier at the start of a solvent gradient
φ_{final}	Fraction of strong mobile-phase modifier at the end of a solvent gradient
φ'	Change in volume fraction of the strong mobile-phase modifier over the gradient duration for a linear gradient.
φ_{sol}	Fraction of strong mobile-phase modifier at which a polymer solubilizes
$\mathbf{P}$	Array containing posterior probabilities on its diagonal
P	Pressure
p_i	Posterior probability for the ith data point
p	Asymmetry parameter (in Chapter 7)
p	y-intercept of the S vs. $\ln k_0$ correlation for a homologues series
p_I	y-intercept of the S vs. $\ln k_0$ correlation for a homologues series for monomer type I
p_{II}	y-intercept of the S vs. $\ln k_0$ correlation for a homologues series for monomer type II
q	Slope of the S vs. $\ln k_0$ correlation for a homologues series
q_I	Slope of the S vs. $\ln k_0$ correlation for a homologues series for monomer type I
q_{II}	Slope of the S vs. $\ln k_0$ correlation for a homologues series for monomer type II
R	Migration rate
	Universal gas constant (in Chapter 2)
	Resistance (in Chapter 5)
	Roughness (in Chapter 6 and Chapter 7)
R_F	Feedback resistor
R_g	Radius of gyration
R_s	Resolution
	Fit of a model (In Chapter 6 and Chapter 7)
r	Reference signal for alignment
$\bar{r}$	Mean of reference signal

σ	Standard deviation
σ_d	Standard deviation of d^-
σ_t	Measure of peak width in time units
σ_V	Measure of peak width in volume units
σ_x	Measure of peak width in distance units
S	Slope parameter in the LSS model
	Entropy
	Pure spectral profiles of components in the sample (in Chapter 6)
$\Delta S°$	Partial molar volume change in entropy
$\Delta S°_{\mathrm{SEC}}$	Partial molar volume change in entropy due exclusion from the pores
$\Delta S°_{\mathrm{LC}}$	Partial molar volume change in entropy due to interaction with the stationary phase
S_1	Slope parameter in QM
S_2	Curvature parameter in QM
S_A	Slope parameter obtained before accounting for gradient deformation
S_B	Slope parameter obtained after accounting for gradient deformation
s	Noiseless input signal
	User-defined threshold value (in Chapter 7)
$\hat{s}$	Estimate of the noiseless input signal
s_i	ith spectral profile representing the analyte
$s_{bk,i}$	ith spectral profile representing the background
T	Absolute temperature
t_0	Void time
t	Time
t_{dwell}	Dwell time
t_e	Elution time
t_{mod}	Modulation time
t_R	Retention time
$t_R{}'$	Retention time adjusted with respect to the void volume
Δt_e	Average prediction errors of the retention time
u_e	Linear velocity of the analyte at the time of elution

u_i	Linear velocity of an analyte
u_m	Linear velocity of the mobile phase
V_0	Void volume
V	Volume
V_{dwell}	Dwell volume
V_e	Elution volume
V_s	Stationary-phase volume
V_s'	Accessible stationary-phase volume
V_R	Retention volume
V_m	Void volume
V_p	Pore volume
V_i	Interstitial volume
V_G	Volume of the gradient
$V_{G,\max}$	Maximum gradient volume such that all analytes reach their critical composition
V_{out}	Output voltage
V_{in}	Input voltage
$\mathbf{W}$	Weights matrix
w	Signal consisting of solely white gaussian noise
w_b	Peak width at the "base" of the peak
$w_{b,avg}$	Average peak width at peak "base"
w_i	Weight of the ith matched peak
	Weight vector (in Chapter 7)
X	Matrix of recorded data
X_I	Mass fraction of monomer type I
X_{II}	Mass fraction of monomer type II
X_{analyte}	Chemically relevant (analyte) component in matrix of recorded data
$X_{\text{background}}$	Background component in matrix of recorded data
X_C	Capacitive reactance
x_p	Array containing information of peaks
$\hat{x}_p$	Estimate of the peak array
$\bar{x}$	Mean of signal
x	One-dimensional array of signal data
x_{new}	New initial estimate of the background

$x_{\text{background}}$	Modified signal vector correspond to the background
x_i	ith data point in the signal (x)
Z	Impedance
z	Fitted data to one-dimensional array of data
$z_{b,i}$	The i-th data point in the estimated background
z_i	ith data point of fitted data (z)

Acknowledgements

And so we arrive at the final part of this thesis. Here I would like to thank everyone that has contributed in some way to the completion of this thesis, Of course the product before you is the work of many people, but since they couldn't all be on the front cover I would like to mention some of them here.

First and foremost, of course I want to thank you, **Peter**. Probably like many before me, I was not planning on doing a PhD, but somehow I guess you saw potential and pushed me to actually do it. I actually first met you around ten years ago during an ASTP interview, but I think you may have forgotten (luckily for me because that interview didn't go so smoothly for me ;) (I will blame the fact I had only done GC at this point, and all your questions were, of course, about LC). Luckily, persistence does pay off sometimes, and now we're here (although I also realize there's still a lot left to learn). Thank you for all the patience, guidance and for believing in me, both during my MSc project and my PhD. I think it's safe to say this thesis would not have existed without your help. You've definitely taught me to be critical about results, and to not take everything at face value. Luckily I have also, over these past four years, learned to love your incredibly bad and sometimes also very good jokes ;).

The next person I want to thank is you **Bastiaan**. I will admit that the first time I had to go to Germany it definitely felt a little forced, and I essentially assumed it would be a waste of my time. What I actually found was everything that I did NOT expect. Thank you for all the discussions, the never-ending fountain of crazy ideas (some of them perhaps more feasible than others) and all of the other support during these four years. I can't thank you enough.

Then I want to thank you **Bob**. I was definitely considering not putting you in here because you seemed so angry about it at one point ;) but I guess I changed my mind. I feel I have gotten to know you a lot better over these four years and I am really thankful for that. I really enjoyed our trip to the US, and also all of the time we spent playing boardgames, let's hope this will continue for a while even though I'm in Brussels now (luckily it's not the end of the world). Thank you for all the support, both in terms of scientific input and work, but mostly for being a friend.

Of course, there's no people that defined my time as a PhD more than my direct colleagues. **Stef**, **Wouter**, **Tijmen**, and **Jordy**. I am super grateful that I was in this project together with you guys. All the time spent playing boardgames, hanging out, enjoying our beers, swimming in pools on far-too-hot conference days; you definitely made these four years unforgettable. I really feel we had great chemistry and a lot of good times, but that it also didn't come to the detriment of scientific output. In my opinion (I think all our industrial partners can agree), absolutely the best kind of team you can have.

Stef, never change, je bent echt een topper. If there is one colleague that I have really gotten to know well during these four years, it's you. Thank you for all the talks and all the drinks at the Polder or elsewhere, especially around the time of covid I think I definitely needed these talks to keep going. You should forget some boardgame rules sometimes though, make it a little easier for others to cheat as well ;). **Wouter**, thank you for sharing the time in Germany and all the "coffee moments", it was nice brainstorming possible experiments in the hallway. Your outspokenness and "can-do" mentality have definitely been inspiring during these four years. **Tijmen**, if anyone knows some funky functions in MATLAB, I think it's you ;) thank you for all the help and the good times during these four years and before during our MSc projects and the trip to Edinburgh. During conferences and trips I could always rely on you to be the voice of rationality, especially after Stef and Wouter had had too many drinks again. **Jordy**, of course you joined the project a bit later but I don't feel like that created any barriers. I will never forget your laugh probably (it still haunts me to this day). I'd also like to thank **Jim** here. While you weren't a part of the same project it definitely felt like it sometimes at the end. Some work definitely wouldn't have been possible without you, so thank you for your contributions (and your great beer ;)).

Then I would like to thank a few office buddies (or ex-office buddies..), **Mimi** thank you for all the coffee talks and the trips during the final months, I think you really brightened everyone's day in the office with your incredibly enthusiastic personality (you even managed to brighten my, more-gloomy, side of the office ;)). In a similar vein I'd like to thank **Pascal** and **Rianne**. You two definitely made sure there weren't any quiet moments during work (I guess that can be a good thing, sometimes).

And then before you know it you're already at three pages, even though I told myself I'd keep this short.. apparently there's just that many people to thank. I'd also like to thank you **Govert**, even though your expertise was of course more on the bio side you definitely always ensured that I felt very welcome when I had to be at the VU.

I'd also like to thank all the industrial partners in the project. In specific I'd like to thank **Ron**, I could always count of you for your full support ;) and **Harry**, but also some of the others I met during my time at DSM (**Rene**, **Marko**, **Ivana**, **Roger**), thank you so much for all the support. Similarly I'd like to thank all the people at BASF (**Roland**, **Bianca**, **Christiane**, **Till**). Also many thanks to **Leif**, in part for hosting a great final bi-annual meeting (never thought I'd say that) in Sweden.

I would also like to thank you **Sjoerd,** you left right around the time I started my PhD, but I will never forget the interest you showed in my (and others) work right from day one. There's a few people you can tell really have a passion for chromatography and you're definitely one of them.

Tom, thanks for all the help with troubleshooting, especially at the start of my PhD I think I would have been incredibly lost in the lab without you, I'm sorry for accidentally assuming you were way older than you actually are ;). In the same vein I would also like to thank **Pascal**, and **Aleksandra** (you're legendary).

There are also a few students who have contributed to the work in this thesis who I'd like to thank, **Gerben**, **Bart**, and **Steven**. Thank you all for contributing even though the projects were maybe not always ideal. It is much appreciated.

I'd also like to thank all my other colleagues (**Noor**, **Maarten**, **Dorina**, **Suhas**, **Liana**, **Marta**, **Maria**, **Alan**, **Mirjam**, **Garry**, **Saer**, **Ruben**, **Henrik**, **Alina**, **Andrea**, **Rick**, **Nino**, **Joshka**, **Annika**, **Denice**, **Iro**, **Robert**, **Jessica**, **Ziran**, **Minghze**, **Vika**, **Masashi**). Thanks for all the good times. (Sorry, it's impossible to keep this short without a list)

Then, I would like to thank some friends, **Boris**, **Arnold**, **Ruben**, thanks for listening to all of my complaining during this PhD.

En als laatste wil ik mijn familie bedanken. **Pap**, **Esther**, **Bert**, bedankt voor alles.