analytical chemistry

pubs.acs.org/ac Article

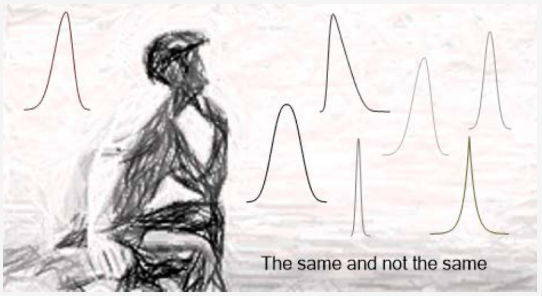
Shape-Based Peak Identity Confirmation in Liquid Chromatography

Akinde F. Kadjo,* Purnendu K. Dasgupta, and Kannan Srinivasan

Cite This: *Anal. Chem.* 2021, 93, 3848−3856 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Chromatography is used to separate analytes in a mixture. In the absence of a detector providing analyte-specific information, the retention time is used to identify the analyte, typically by comparing the retention time of the unknown with that of a standard. This is the simplest of identification routines, but it remains the most widely used. The challenge is that analyte retention times can and do shift; sole reliance on retention time for identification is not prudent. Retention time is influenced by the total sample concentration, elution of components near the peak of interest, aging of a column, and so on, complicating time window-based identification. We propose a sameness threshold (ST) that is based on numerical comparison of shapes of chromatographic peaks for identification confirmation. We demonstrate shape-based identification confirmation or lack thereof of an unconfirmed peak by comparison with a shape library of standards. In numerous cases we look at, only one analyte in the shape library fits the putative identity. ST can be based on r^2, the commonly used coefficient of determination used in regression analysis, or on an "index of width mismatch" that compares the widths of the peaks throughout the available height range.



INTRODUCTION

Chromatography is a process where analytes travel through a column propelled by a mobile phase while interacting with a stationary phase. Separation is achieved due to the differing affinities of the analytes for the stationary phase *versus* the mobile phase.[1] A common practice is to first separate a standard mixture of the desired analytes under optimum conditions. Then, the sample containing unknown analytes is injected under the same conditions. Analyte identification is based on "sameness" comparison of retention times of sample peaks *versus* those of the known injected standards. An unknown having the "same" retention time as that of a particular standard is then identified as that particular standard. The tolerance for "sameness" is arbitrary; it is largely based on user experience or default settings in the chromatographic software used. This method of identification is not infallible; retention time shifts can occur due to a variety of reasons.[2−5] For the same analyte, the retention time can change as a function of injected concentration. A large amount of a single analyte can affect the retention times of all other analytes present. Analytes eluting close to each other may change the apparent peak location(s), leading to misidentification.

Detectors providing solute-specific information can facilitate identification. While spectral libraries can be useful in some applications,[6] few analytes exhibit UV−visible spectral absorption that is sufficiently unique. Even when the spectral signature is unique, additional steps such as momentarily pausing chromatographic movement may be required to obtain a high-fidelity absorption spectrum. In the case of ion chromatography (IC), UV−visible absorbance detection has never been mainstream; many common ions of interest have no useful UV absorption. Vacuum UV absorbance detectors, recently introduced for gas chromatography, provide far more solute-specific information and are useful for confirming analyte identities.[7] For unequivocal peak identification, the use of mass spectrometry (MS) is the gold standard. The first pairing of a chromatograph with a mass spectrometer is now 70 years old; the first GC−MS instrument by McLafferty and Gohlke is now a National Historic Chemical Landmark.[8]

Shape-Based Peak Description. Chromatography is one of the more mature areas of analytical sciences. Still, there have been few efforts to use the shape of a peak as an identification tool.[9] An early effort to describe the shape of a chromatographic peak was a curve defined by eight separate parameters.[10] Subsequently, Mott and Grushka[11] described the shape of a chromatographic peak of a particular analyte with such parameters and used these to identify compounds, at least to assign them to different structural classes.

Keeping such multiparametric descriptions aside, chromatographic peaks are theoretically expected to be purely Gaussian.[12−15] Gross departure from the model Gaussian behavior is often assigned to system issues and is to be avoided.[16] Indeed, fronting and tailing behaviors always

Received: October 20, 2020
Accepted: February 10, 2021
Published: February 18, 2021

negatively impact the separation and quantitation of analytes. Regardless, chromatographic peaks are rarely perfectly Gaussian,[17−19] as would be indicated by the large number of extant chromatographic peak shape models.[20−23] Molecular interactions, multiple and nonlinear partition mechanisms including overloading effects, and dispersion in various system components are known to influence the shape of chromatographic peaks.[24−27] The story of the journey of an analyte from the injector to the detector is written in the observed peak profile. An appropriate beholding should reveal the uniqueness of the peak shape, which can then be exploited for confirming identification in conjunction with the retention time.

Why Seek Alternative Means for Confirming Peak Identities? Tandem MS unquestionably provides more information about analyte identities than any other technique. Single quadrupole MS is far more affordable but has limited capabilities. For most IC users, however, the economics is still formidable. Taking Thermo Fisher as a representative vendor, although the base price of an IC instrument ranges from 10 to 50 K, adding the lowest-end MS more than doubles the system cost even for the highest-end IC, not even including costs of essential high-purity N_2 supply. Moreover, in suppressed conductometric ion chromatography (SCIC), none of the MS detectors are effective for any analyte below m/z 50 at meaningful concentrations, in part because of background leaching of suppressor membranes.[28] This is especially relevant in a common longstanding and vexing problem in IC related to the close elution (often, coelution) of poorly retained hydrophilic ions, notably, fluoride (m/z 19) and a number of organic acids, leading to potentially false identification and/or erroneous quantitation.[29] In this common case alone, it would be of considerable value if an inexpensive *in silico* means can be provided to establish if what is being quantitated as fluoride is purely (or if at all) fluoride.

Strategy. We use the amplitude-normalized shape of a peak. We compare this for an unconfirmed peak with that of a known analyte of identical peak area to determine if a preassigned sameness threshold (ST) has been reached. We propose that a specific analyte peak shape, arising from a particular concentration and volume, injected on a particular column, undergoing a given elution condition, is adequately unique to confirm the identity when observed with sufficient temporal and amplitude resolution. While it may be applicable to chromatography in general, herein, we mainly examine examples from SCIC and a single example from reverse-phase liquid chromatography with fixed wavelength UV detection. Admittedly, IC represents a favorable test case. At least two interaction mechanisms, hydrophobic and electrostatic, are always operative in ion exchange-based separations. In addition, in SCIC, the analyte concentration profile exiting the suppressor is transformed to the conductance signal both in a linear fashion by the analyte mobility value and in a nonlinear fashion by the analyte pK_a.

We transform multilevel analyte calibration data, routinely generated for analyte quantitation, to a "shape calibration matrix" (SCM) for each analyte. The SCM essentially enables the (re)construction of that analyte peak for any desired peak area within the data span. A quantitative comparison of the shapes, of the unconfirmed peak and of a known analyte of the same peak area generated from its SCM, can judge if a preassigned ST is reached. Even if two analytes have the same retention time, correct identification as to which is eluting is possible without the use of any analyte-specific detector information.

PRINCIPLES

Shape Calibration Matrix. Analyte peaks vary not only in their lateral shapes but also in their amplitudes. To remove this from numeric considerations, all peaks are first normalized to unity height, that is, all ordinate values are scaled such that the normalized peak height H_N at any point is calculated as H/H_{max}, where H is the analyte response at any time point and H_{max} is the peak amplitude (H_N has been previously referred[9,17] to as $1/\bar{h}$; the present notation is less cumbersome). To compare peaks regardless of their actual retention times, the apex is centered at an abscissa value of zero. This means that half the peak will occupy negative abscissa values. To avoid ambiguities, the abscissa is referred to as θ, rather than time.

When peaks are thus plotted in the H_N−θ space, for some analytes, the peak shape is seen to be concentration-independent over a large range of concentrations. This is the case for chloride, as seen in Figure 1A. More commonly, however, the exact shape depends on the injected amount. Even for an analyte such as chloride that exhibits concentration-independent shape at low loading, the shape may change at higher amounts because of column overloading, nonlinear isotherms, the sample acting as its own eluent (self-elution), and/or other factors. In other cases, the shape is influenced by the concentration because the monitored property is not linear with concentration. In SCIC in particular, an anion is detected conductometrically as the corresponding acid. For weak acids, the detector signal is not linear with concentration. Broadening of the peak with increasing concentration, more so toward the base of the peak than the apex, is to be expected as the extent of ionization decreases with increasing concentration and increasing pK_a of the analyte acid. A change in the peak shape, as registered by the detector, will be expected even if the normalized concentration profile does not change. Nitrite illustrates such a behavior (Figure 1B). However, just being a strong acid anion does not assure isomorphous peaks as a function of concentration. First, for all analytes, starting at actual detected ionic concentrations of ∼1 mM, the log (conductance)—log (concentration) slope becomes less than 1 and the same peak broadening effect as observed for weak acids sets in, albeit to a lesser degree. For nitrate (Figure 1C), a combination of a nonlinear adsorption isotherm, self-elution, and possibly other factors may lead to the observed increasing asymmetry with concentration.

Because of concentration dependence of shape, it is necessary for the comparison peak generated from the SCM to have the same concentration as that of unconfirmed peak. The concentration of the unknown analyte is also obviously unknown *a priori*, but one can *create* a benchmark that has the same peak height or area. Herein, below, we use the actual original peak area for this purpose (this is not the normalized peak area). An infinite number of comparison standards to match every unknown peak area encountered is obviously impractical. An SCM is constructed for each analyte that is intended to be used as a comparison standard. The SCM is nothing more than a numerical matrix derived from a finite number of quantitation standards that allows the reconstruction of a standard comparison peak of any desired area (within the data span) using third-order polynomial interpolation. The

https://dx.doi.org/10.1021/acs.analchem.0c04432
Anal. Chem. 2021, 93, 3848−3856

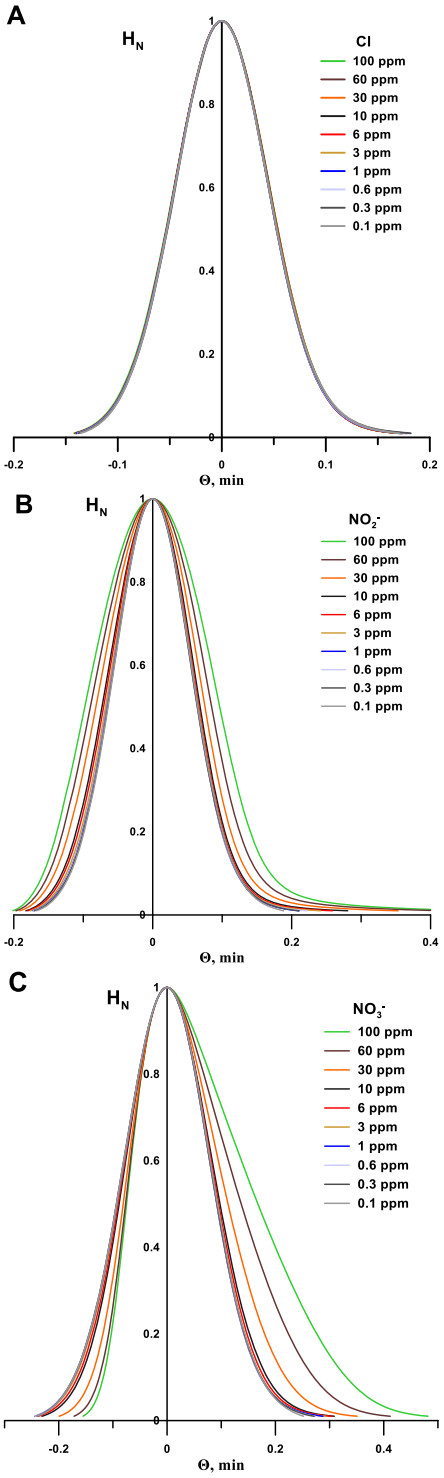


Figure 1. Changes in the normalized peak shape (θ–H_N plots) as a function of concentration. (A) Peak shape for chloride remains isomorphous over a 4 order of magnitude injected concentration range. (B) Normalized peak widths systematically increase with concentration for weak acid anions such as nitrite. (C) Nitrate peak asymmetry increases with increasing concentration. Column set B, see the Experimental Section for eluent and injection volume details.

only input necessary is multilevel calibration data for the analyte that are routinely generated as a prelude to quantitation. Much as a calibration curve for quantitation must include both the highest and lowest concentrations to be quantitated, the SCM must encompass the entire concentration span anticipated for an unconfirmed analyte.

The first step is to convert the raw quantitation data spanning n different analyte concentrations for a given analyte to the normalized form. Then, we have n $\theta - H_N$ arrays, each for a different peak area A. Now, we express the normalized ordinate value H_{Ni} as a polynomial function of the peak area A for each θ value (θ_i). While we found a second-order polynomial to provide insufficient agreement with the observed values, a third-order polynomial was sufficient

$$H_{Ni} = H_{Ni0} + a_i A + b_i A^2 + c_i A^3 \quad (1)$$

where A is the peak area, H_{Ni0} is the area-independent term, generally the largest contributor to H_{Ni}, and a, b, and c are the first-, second-, and third-degree empirical coefficients. The n available data points are used to obtain the best fit values of H_{Ni0} and a_i, b_i, and c_i. The process is then repeated for the entire range of θ_i values of interest (which will depend on the analyte and span a greater range for broader peaks). An array thus generated comprises best-fit values for all of the four abovementioned parameters for every θ_i value of interest. This array thus constitutes the SCM for the analyte, specific of course to the chromatographic conditions. To generate a normalized peak for any desired peak area (within the span of the peak area range in the library), one merely calculates H_{Ni} for every θ_i value in the SCM using eq 1, inputting the desired value of A and the H_{Ni} and a_i, b_i, and c_i values from the SCM.

Figure 2 illustrates an example for 10 normalized θ–H_N calibration data sets for fluoride from which the SCM is constructed; an abbreviated illustrative SCM derived from these data is shown in Table 1. Table S1 in the Supporting Information provides the unabridged version of the SCM. Figure S1 in the Supporting Information illustrates the process.

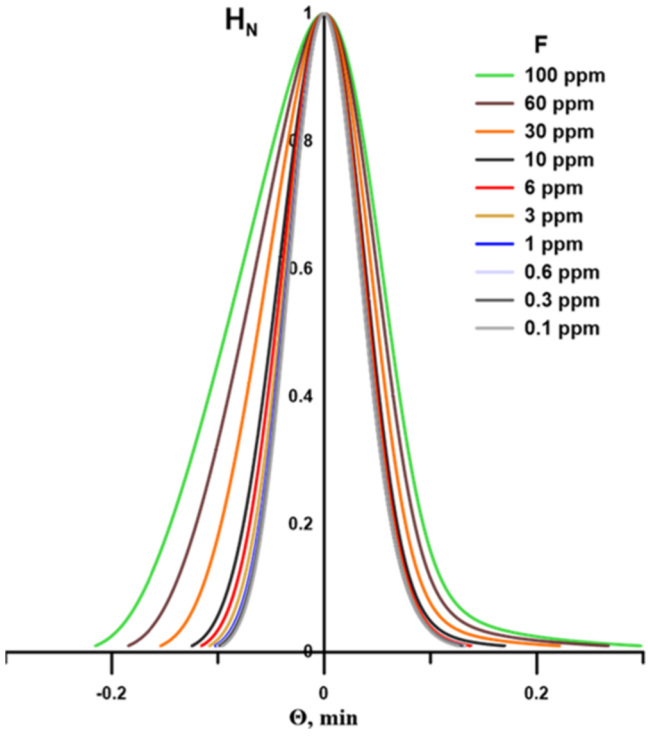


Figure 2. Normalized calibration data for fluoride peaks as shape library input. Column set B.

https://dx.doi.org/10.1021/acs.analchem.0c04432
Anal. Chem. 2021, 93, 3848−3856

Table 1. Illustrative Eq 1 Calibration Matrix Generated from Data in Figure 2[a]

θ_i	H_{Ni0}	a_i	b_i	c_i
−0.2000	-1.40×10^{-4}	2.95×10^{-4}	-6.49×10^{-4}	3.36×10^{-6}
−0.1500	7.42×10^{-4}	-1.62×10^{-3}	3.38×10^{-4}	-1.23×10^{-6}
−0.1000	5.73×10^{-3}	1.32×10^{-2}	6.45×10^{-4}	1.91×10^{-5}
−0.0500	3.16×10^{-1}	5.88×10^{-2}	-3.03×10^{-3}	5.56×10^{-5}
0.0000	1.00×10^{0}	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
0.0500	3.51×10^{-1}	2.27×10^{-2}	-6.56×10^{-4}	8.60×10^{-6}
0.1000	3.29×10^{-2}	3.74×10^{-3}	9.78×10^{-5}	-2.33×10^{-6}
0.1500	4.91×10^{-3}	1.99×10^{-3}	-2.60×10^{-5}	2.71×10^{-7}
0.2000	3.65×10^{-4}	1.55×10^{-3}	-3.88×10^{-5}	5.43×10^{-7}
0.2500	-4.80×10^{-4}	1.11×10^{-3}	-3.16×10^{-5}	4.49×10^{-7}
0.3000	-3.30×10^{-4}	8.17×10^{-4}	-3.03×10^{-5}	5.19×10^{-7}
0.3003	-3.29×10^{-4}	8.15×10^{-4}	-3.02×10^{-5}	5.19×10^{-7}
0.3007	-3.28×10^{-4}	8.13×10^{-4}	-3.02×10^{-5}	5.18×10^{-7}
0.3010	-3.27×10^{-4}	8.11×10^{-4}	-3.01×10^{-5}	5.17×10^{-7}

[a]For the complete matrix, see Table S1.xlsx in the Supporting Information.

Figure S2 in the Supporting Information illustrates how well the third-degree polynomial represents the fluoride peak shape.

Beyond this, how well the overall peak shape is described by eq 1 is obviously related to the concentration resolution (value of n) and the value of i (how many θ_i points are used to describe the peak). As a computational alternative, we also explored the use of a generalized Gaussian model that uses two parameters to model each side of a peak and thus four parameters in all.[17] Again, the dependence of these parameters on the peak area was described with a third-order polynomial. As seen in see Table S2 in the Supporting Information, this approach did not provide results as good as those using the present approach based on eq 1. The fundamental difference is that in the present construct, the peak shape is a completely empirical entity; there is no model.

Peak Identity Confirmation: Indices for Sameness. Identity confirmation involves comparison of the unconfirmed peak profile in the θ−H_N space with the peak profile for the same peak area of a standard analyte computed from its SCM. Figure S3A,B in the Supporting Information indicate visually obvious examples of what most would regard as *same* and *not the same*, respectively. However, a quantitative matching index is obviously needed. Various data comparison methods may be used; here, we suggest two simple indices.

Coefficient of Determination. The first proposed index is a single global fit parameter, namely, the coefficient of determination, r^2, r being the Pearson correlation coefficient. As is well known, an r^2 value of unity would indicate a perfect match. What would constitute the acceptable threshold for sameness remains up to the user. This judgment in turn is governed by how different peak shapes tend to be within that milieu. Consider that presently, there is no preconceived mathematical construct for a peak model. Regardless, chromatographic peaks, different in shape as they may be, look more like each other than, for example, a rectangular pulse. While an r^2 value of 0.9500 is not a terrible match in the purely numerical world, Figure S4 illustrates a pair of peaks with an r^2 of 0.9644 that are obviously *not the same*. Much as in linear regression, there are no absolute standards on what r^2 value constitutes acceptable linearity or what tolerance can be placed around the nominal retention time of a peak for an unconfirmed peak to be identified as that analyte based on retention time, here also, this is a choice made by the user. In reverse-phase HPLC, on well-constructed columns, the peak shapes tend to be highly similar to each other. Based on our experience with such data, we set the sameness threshold bar at an r^2 value of ≥0.9995, whereas for the somewhat more diverse IC peaks, we use an ST bar of ≥0.9990 to confirm the identity.

Index of Width Mismatch (IWM). This is a distributed metric that reveals the degree of (mis)match of the peak width at different values of H_N; the latter is specified in the subscript. Specifically, it is the scalar percent difference between the actual width of a peak ($W_{act,h}$) and the width of the comparison peak derived from the SCM ($W_{SCM,h}$) as a fraction of the actual peak width, at a given H_N value

$$\mathrm{IWM}_h = \left| \frac{W_{act,h} - W_{SCM,h}}{W_{act,h}} \right| \times 100\% \quad (2)$$

Note that this index may be used across the full width of a peak or the left or the right halfwidth; such separate consideration is useful when an impurity affects one side of a peak. See Figure S5 in the Supporting Information for an illustration.

To use the IWM_h as a confirmatory index, ST bars can be set by the user within stated bounds of H_N, for example, in the HPLC example described in this paper, ST was based on an H_N range of 0.01−1, where ≥99% of computed IWM_h values must be ≤1%. A different ST could be that *all* computed IWM_h values in this range are ≤5%. A greater insight can be obtained also from a graphical plot of IWM_h (as the x -coordinate error bar, centered on $x = 0$) with H_N as the ordinate (see Figure S6 in the Supporting Information).

While a similar index based on the height mismatch at various values of θ is possible, this parameter is not preferred. It is intrinsically biased toward producing a greater relative mismatch at greater magnitudes of θ (*i.e.*, at lower H_N values).

DATA CONSIDERATIONS

Data Density. In any shape-matching algorithm, the number of data points considered per peak is critical. Much as in an image, the number of pixels provides for image quality, increased amplitude (voltage) resolution and higher sampling frequencies provide a more highly resolved image of the peak and increase the discriminating ability of any matching algorithm. As to the first, our data were acquired with 24-bit resolution, which is a fairly standard current practice. Regarding sampling frequency, our experience indicates that

https://dx.doi.org/10.1021/acs.analchem.0c04432
Anal. Chem. 2021, 93, 3848−3856

1000 points per peak (considered across H_N bounds of 0.01−0.01) provide sufficient data density. For a Gaussian peak of 8 s halfwidth, a sampling frequency of ∼50 Hz meets this goal. A correspondingly higher sampling frequency will be needed for narrower peaks. A greater number of data points increase the accuracy in locating the peak apex. If sufficient data points are not available, the data density can be increased by interpolation. Table 2 compares the r^2 values of fluoride analytes for fluoride identity at an initial collected frequency of 5 Hz *versus* data that were "upsampled" to 50 Hz using cubic spline interpolation and data directly collected at 50 Hz. In all cases, the match is better at 50 Hz, both for the upsampled data and the directly collected data; indeed, a datum in the 5 Hz set fell below our ST bar (shaded in rose).

Table 2. Identification of Fluoride Analytes at Different Sampling Frequencies[a]

Sample	5 Hz	50 Hz	5 Hz upsampled to 50 Hz
0.100 ppm F^-	0.9992	0.9991	0.9996
0.300 ppm F^-	1.0000	0.9996	1.0000
0.600 ppm F^-	0.9989	1.0000	1.0000
1.00 ppm F^-	0.9999	0.9999	0.9999
3.00 ppm F^-	0.9998	0.9999	0.9999
6.00 ppm F^-	0.9995	0.9999	1.0000
10.0 ppm F^-	0.9996	0.9997	1.0000
30.0 ppm F^-	1.0000	1.0000	1.0000
60.0 ppm F^-	1.0000	1.0000	1.0000
100 ppm F^-	1.0000	1.0000	1.0000

[a]Column set B. Rose shading indicates values below the r^2 acceptance threshold of 0.9990.

Spacing of Calibrant Concentrations. How many separate concentrations are needed per decade for accurate quantitation of an analyte depends on the linearity of the response behavior. More calibrants are needed if the response is nonlinear. The same is true when those data are to be used to construct an SCM. Our experience indicates that one minimally needs two (reasonably equispaced) concentrations (*e.g.*, 1, 3, and 10) per decade although three per decade (*e.g.*, 1, 2, 5, and 10 or 1, 3, 6, and 10) are preferred, especially when the response is nonlinear. For fluoride, which exhibits a nonlinear response behavior in the indicated concentration span, the r^2 values increase with a greater number of calibrants per decade (Table 3).

Impact of Noise. If it is sufficiently large, noise will predictably affect the reliability of peak identification. To quantitatively test this, we added increasing amounts of simulated random noise to a simulated peak, to attain peak amplitude-to-noise amplitude (signal to noise, S/N) ratios of 10,000, 1000, 100, 50, 30, 20, and 10. The r^2 values decreased as 1.0000, 1.0000, 0.9999, 0.9997, 0.9992, 0.9981, and 0.9817, thus failing the ST bar at S/N ≤ 30. To reliably perform shape-based peak confirmation, the test peaks and the input data to the SCM all should minimally have an S/N ≥ 30. For comparison, the S/N for 0.02 ppm F^- and 0.1 ppm Cl^-, NO_2^-, SO_4^{2-}, Br^-, and NO_3^-, lowest concentrations presently used, ranged from 120 to 160 under typical IC conditions.

Table 3. Identification of Fluoride Samples Using Different Concentration Sets for Calibration[a]

Sample	r^2 for Fluoride identity		
	3 calibrants /decade	2 calibrants /decade	1 calibrant /decade
0.100 ppm F^-	0.9998	0.9999	1.0000
0.300 ppm F^-	1.0000	1.0000	1.0000
0.600 ppm F^-	1.0000	1.0000	1.0000
1.00 ppm F^-	1.0000	0.9999	1.0000
3.00 ppm F^-	0.9999	0.9999	0.9995
6.00 ppm F^-	0.9998	1.0000	0.9992
10.0 ppm F^-	1.0000	1.0000	1.0000
30.0 ppm F^-	0.9998	0.9992	0.8347
60.0 ppm F^-	1.0000	1.0000	0.8252
100 ppm F^-	1.0000	1.0000	1.0000

[a]Column set A.

EXPERIMENTAL SECTION

Ion chromatography (IC) data were generated with a Thermo Scientific Dionex ICS-6000, comprising a pump, an eluent generator, a loop injector (2.5 μL except as stated), a temperature-controlled oven (30 °C), a conductivity detector, and an autosampler. Other specific experimental conditions are listed below for individual column sets: (A) AG15 2 × 50 mm + AS15 2 × 250 mm, 38 mM KOH @ 0.30 mL/min, 5 μL injected; (B) AS19 2 × 250 mm 4 μm column, 25.5 mM KOH @ 0.25 mL/min; (C) AS18 2 × 150 mm 4 μm, 23 mM KOH @ 0.25 mL/min; (D) AS19 2 × 250 mm 4 μm, 20 mM KOH @ 0.25 mL/min; (E) AS19 4 × 250 mm 4 μm, 20 mM KOH @ 1.0 mL/min, 10 μL injected; (F) AS22 Fast 4 × 150 mm 4 μm, 4.5 mM Na_2CO_3/1.4 mM $NaHCO_3$ (manually prepared) @1.20 mL/min; and (G) CS16 2 × 150 mm 4 μm, 30 mM methanesulfonic acid (MSA) @ 0.16 mL/min.

HPLC data was generated with a Vanquish Horizon UHPLC system. The experimental conditions followed the EPA method 532 guidelines for the separation of phenylurea pesticides. Column: Thermo Scientific prototype C18, 2.1 × 100 mm 1.5 μm, solvent A water, Solvent B acetonitrile, mobile phase temporal gradient (t_{min},% B): 0,40; 1,40; 3,80; flow 0.6 mL/min, temperature: 30 °C, injection volume: 1 μL, UV detection @ 245 nm, 100 Hz data.

RESULTS AND DISCUSSION

Analytes Separated on a 2 × 250 mm AS15 Column. The two chromatographic peaks from a sample containing 2.85 ppm fluoride (peak A, area 1.13 μS·cm^{-1}·min) and 1.00 ppm chloride (peak B, area 0.22 μS·cm^{-1}·min) were individually examined to see if the shape of either one can be confirmed as that of fluoride. Using A = 1.13 in eq 1 and using a_i, b_i, and c_i values previously stored, H_{Ni} values are computed for all θ_i values. This predicted peak profile (red circles, a limited number shown) is plotted along with the experimental peak A (Figure 3A, blue line). Within bounds of H_N 0.01−0.01 (this is the default range used throughout this paper), the two data sets exhibit an r^2 value of 1.0000. Thus, attribution of peak A as fluoride is confirmed. Similar analysis is now done for peak B. The predicted profile for fluoride for this peak area is shown in Figure 3B (red circles), while the actual experimental data are shown as the blue line. Correlation of the available data (from H_N 0.01−0.01) showed an r^2 value of 0.8454, commensurate

https://dx.doi.org/10.1021/acs.analchem.0c04432
Anal. Chem. 2021, 93, 3848−3856

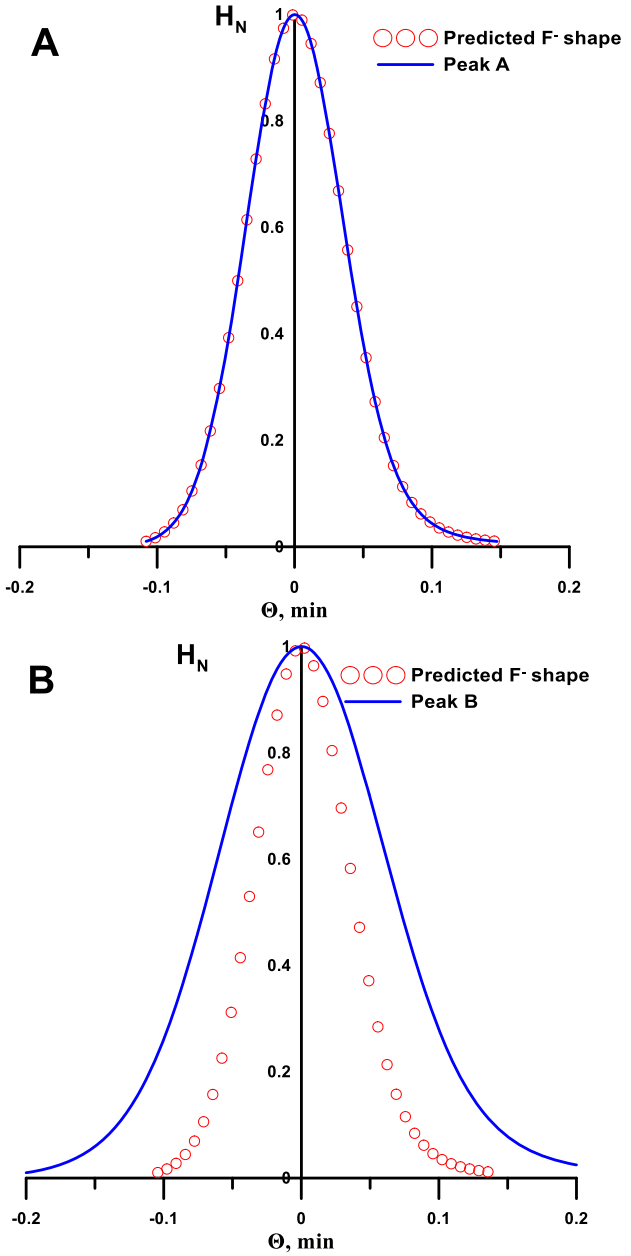


Figure 3. (A) Peak A matched with the predicted fluoride profile for the same peak area, r^2 = 1.0000. (B) Peak B matched with the predicted fluoride profile of the same peak area, r^2 = 0.8454. Column set A.

with the observed poor visual match. Peak B will thus be designated as *not the same as fluoride*.

The shape calibration matrix was constructed similarly for all six analytes. The entire set of peaks (six analytes at 10 concentrations) was then tested for an identity match in comparison with a peak generated for the same area from the respective calibration matrices. It will be observed that the procedure provides correct identification in all cases, without either false positives or negatives. Clearly, shape-based peak identification as outlined is feasible even with detectors that provide no analyte-specific information.

Note that shape uniqueness may originate both chromatographically as well as nonlinear and unique detector response behavior that happens in SCIC for weak acids as a function of concentration.

Analyte Identification in a Drinking Water Sample. The chromatogram of a water sample separated on a 4 × 250 mm 4 μm AS19 column is shown in Figure S8. The correlation matrix (r^2 values) for the same peaks generated from corresponding standards is shown in Table 5. With the ST

Table 4. Identification Summary of Test Analytes Separated on an AS15 Column[a,b]

Sample / Concentration mg/L	r^2 Values					
	F-	Cl-	NO_2^-	SO_4^{2-}	Br-	NO_3^-
F^- 0.02 - 20.0	1.0000 - 1.0000	0.8759 - 0.9111	0.7991 - 0.8142	0.7209 - 0.7672	0.6703 - 0.7376	0.6572 - 0.7310
Cl^- 0.1 - 100.0	0.5095 - 0.9096	0.9999 - 1.0000	0.4837 - 0.9724	0.7754 - 0.9025	0.1035 - 0.8172	0.0436 - 0.7892
NO_2^- 0.1 - 100.0	0.6107 - 0.7347	0.9036 - 0.9715	0.9999 - 1.0000	0.9607 - 0.9900	0.7866 - 0.8962	0.6711 - 0.8607
SO_4^{2-} 0.1 - 100.0	0.5073 - 0.6238	0.8401 - 0.8706	0.9540 - 0.9935	0.9998 - 1.0000	0.4875 - 0.9552	0.5062 - 0.9164
Br^- 0.1 - 100.0	0.3852 - 0.4637	0.6319 - 0.6884	0.7907 - 0.8216	0.8700 - 0.9380	0.9990 - 1.0000	0.9780 - 0.9900
NO_3^- 0.1 - 100.0	0.3249 - 0.3679	0.4986 - 0.5982	0.6663 - 0.7281	0.7334 - 0.8703	0.9682 - 0.9883	0.9993 - 1.0000

[a]A complete version of this table, where the results for each concentration appear separately, is presented in the Supporting Information as Table S3. See Figure S7 for the chromatogram. [b]Each analyte was tested at 10 individual concentration levels; see Table S3.

Table 5. Drinking Water Chromatogram[a] Peak Identification

Peak #1	F^- Identity	Cl^- Identity	NO_2^- Identity	Br^- Identity	SO_4^{2-} Identity	NO_3^- Identity
1	0.9999	0.9438	0.8750	0.7942	0.7430	0.6700
2	0.9521	0.9999	0.9821	0.9387	0.9008	0.8378
3	0.8644	0.9805	0.9998	0.9853	0.9611	0.9058
4	0.7148	0.8768	0.9472	0.9918	1.0000	0.9814
5	0.4414	0.5831	0.6792	0.7802	0.8447	0.9128
6	0.6193	0.7742	0.8681	0.9427	0.9781	0.9999

[a]See Figure S8, Column set E.

bar set at r^2 = 0.9990, identification is possible without any regard to retention times in this case and without the need for detectors providing analyte-specific information. Note that bromide was found to be not present and peak 5 did not match the shape of any of the standards—all these conclusions were possible without any retention information at all.

Identification of Peaks Having Very Close Retention Times. As Table 5 indicates, the present approach can prevent peak misidentification. A common situation is that despite optimized conditions, two analytes have nearly the same retention time and cannot be resolved. However, both are not necessarily present in the sample. Can one identify an analyte eluting in that retention window either as one or the other? As an example, when using 25.5 mM KOH as an eluent on a 2 × 250 mm AS19 column, nitrate and sulfate have essentially the same retention time (Figure S9). Comparison of the peak at ~9.5 min shown in the black trace, when compared with respective peaks generated for the same area from nitrate and sulfate SCMs, respectively, produces r^2 values of 0.9968 and

https://dx.doi.org/10.1021/acs.analchem.0c04432
Anal. Chem. 2021, 93, 3848−3856

1.0000. The peak at ~9.5 min in the red trace in the same comparison exhibits r^2 values of 1.0000 and 0.9965. With a ST bar of r^2 = 0.9990, the black peak would be identified the *same as* sulfate and the red peak *same as* nitrate. Importantly, the black peak is found to be *not the same as* nitrate and the red peak is found to be *not the same as* sulfate.

A similar experiment was carried out in the reversed-phase HPLC mode, with the acetonitrile content of the eluent adjusted until acetophenone and dimethylphthalate, both analytes used for illustrative exemplary separations, eluted at the same retention time (Figure S10). The r^2 value was 0.9982, well below the ST bar of 0.9995.

Impact of Close-Eluting/Neighboring Peaks. The presence of another analyte of significantly higher concentration can affect both the retention time and the shape of neighboring analyte peaks. This is quantitatively examined on column set E by increasing the amounts of nitrate on proximally and later eluting bromide (R_s = 3.2) and earlier and more distally eluting fluoride (R_s = 31.8). Neither is easily affected. However, predictably, bromide identity confirmation is compromised by the time nitrate amounts reach a ratio ≥500:1, while fluoride identity confirmation is unaffected (Table 6).

Table 6. Impact on Identification as a Function of a Spectator Analyte Concentration

	F^- identity		Br^- identity
Samples	r^2 values	Samples	r^2 values
10 μM F^-	1.0000	10 μM Br^-	1.0000
10 μM F^- + 10 μM NO_3^-	1.0000	10 μM Br^- + 10 μM NO_3^-	0.9998
10 μM F^- + 100 μM NO_3^-	0.9999	10 μM Br^- + 100 μM NO_3^-	0.9997
10 μM F^- + 500 μM NO_3^-	0.9999	10 μM Br^- + 500 μM NO_3^-	0.9996
10 μM F^- + 1 mM NO_3^-	1.0000	10 μM Br^- + 1 mM NO_3^-	0.9992
10 μM F^- + 5 mM NO_3^-	0.9999	10 μM Br^- + 5 mM NO_3^-	0.9919
10 μM F^- + 10 mM NO_3^-	0.9998	10 μM Br^- + 10 mM NO_3^-	0.9800

Applicability Across Analytes and Across Different Stationary Phases. Tables S4−S6 in the Supporting Information show the correlation matrices for six anions (up to 10 different concentrations each) in almost the same manner as Table 4, respectively, for separations on AS18, AS19, and AS 22 column sets. A similar correlation matrix is shown in Table S7 for six cations separated on a CS16 column.

The applicability of shape-based identity confirmation was explored in reversed-phase HPLC for the separation of 11 different phenylurea pesticides using UV detection at 245 nm (Figure S11). Two different concentrations, differing by 20×, were examined. All peaks exhibited remarkable symmetry (asymmetry factors ranged from 1.00 to 1.05) on this state-of-the art 1.5 μm column. Most would regard these as benchmark ideal HPLC peaks. The highest absorbances involved were well below 0.25 AU, so there was no detector nonlinearity. The relevant r^2-based correlation matrix is shown in Table S8 in the Supporting Information, once again attesting to the applicability of the approach. The peak shapes were more comparable to each other than in the IC cases; between-analyte r^2 values were uniformly higher. In Table S8, the ST bar was raised to $r^2 \geq 0.9995$. It was of interest to explore the performance of the other proposed index, IWM_h. This analysis is shown below in Table 7: we indicate here the percentage of data points that (in the H_N 0.01 to 1 domain) had a width index mismatch of ≤1%. It will be apparent that all of the peaks would have still been correctly identified even if we set the ST bar at 90% of the considered data having <1% difference. The IWM_h analysis may be a more critical discriminator for peaks more similar in shape.

Limitations. The proposed approach is not a substitute for MS or multiwavelength UV detection (for cases where distinctive UV absorption is present). It also provides no information as to what an analyte peak might be. It can only conclude whether the peak can be unambiguously identified as one already in the shape library. Shape-based identity confirmation entails a very detailed comparison with the library-generated shape, and as indicated, at least for simulated noise addition to a simulated peak, a minimum S/N of 30 is needed for reliable identification.

The ST bars are set based on the user experience. A numerical probability specification of the two peaks in comparison may perhaps be a better acceptance criterion, but it will not alter the responsibility for the user to set the precise acceptance criterion.

The comparison system is only as good as the ongoing applicability of the calibration data under precisely the same column and elution conditions it is based on. It may seem that more frequent calibration checks will be needed, but this will also assure continued quantitation accuracy. Fouling of a column over a period of use and the resulting effects are difficult to simulate. However, most commonly, this causes a (gradual) loss of capacity. We can simulate this using a slightly higher eluent concentration that causes retention to decrease. On a 2 × 150 mm AS18 column using 33 mM KOH @0.25 mL/min as the base case, we examined two analytes: chloride, which remains remarkably isomorphous with concentration, and nitrite, which does not. A 2.5 μL sample containing each analyte at a concentration of 10 mg/L was injected. Chloride and nitrite eluted at 2.656 and 3.109 min, respectively, with corresponding peak areas of 1.752 and 1.141 μS·min. Fouling of the column was now simulated using 35 mM KOH as an eluent. The retention time for chloride and nitrite decreased by 2.00 and 2.50%, respectively, to 2.603 and 3.031 min, with the corresponding areas decreasing 1.03 and 0.35% to 1.734 and 1.137 μS·min. For the demanding user, the shift in retention is likely sufficient to initiate recalibration. The sameness analysis on the normalized peaks, however, indicated that the correlations for chloride as chloride and nitrite as nitrite remained at r^2 = 0.9999, far above the ST bar.

CONCLUSIONS

This paper demonstrates the applicability of shape-based peak identification in suppressed conductometric ion chromatography. We have looked at six anions on four columns and six cations on one column, at 7−10 different concentrations for each analyte. Concentrations can also affect the shape, so this represents a total of 282 potentially different shapes. In all of the comparisons carried out, on a sameness threshold based on an r^2 value of 0.9990, we did not see a single false positive or false negative case of identity confirmation. For a more limited experiment in reverse-phase liquid chromatography of 11 phenylurea pesticides separated on a single column by the EPA method 532 at two different concentrations, on a somewhat higher sameness threshold based on a r^2 value of 0.9995, a single false negative case of identity confirmation resulted. For the same data set using a different test method (IWM_h,

https://dx.doi.org/10.1021/acs.analchem.0c04432
Anal. Chem. 2021, 93, 3848−3856

Table 7. IWM_h Results[a,b,c]

	% of data points with $IWM_h \leq 1\%$										
concentration	tebuthiuron	thidiazuron	monuron	fluometuron	diuron	propanil	siduron A	siduron B	liunuron	carbazole	diflubenzuron
tebuthiuron[b]	100	0	0	0	0	0	0	0	0	0	0
tebuthiuron[b]	100	0	0	0	0	0	0	0	0	0	0
thidiazuron[b]	0	100	0	0	0	0	0	0	0	8	2
thidiazuron[c]	0	100	0	0	0	0	0	0	0	5	11
monuron[b]	0	0	100	36	0	0	0	1	5	1	19
monuron[c]	0	0	100	13	0	0	0	1	5	10	13
fluometuron[b]	0	0	36	100	0	0	0	1	6	0	3
fluometuron[c]	0	1	1	100	0	0	0	0	2	11	6
diuron[b]	0	0	0	0	100	1	32	12	0	0	0
diuron[c]	0	0	0	0	100	0	11	25	0	0	0
propanil[b]	0	0	0	0	1	100	0	0	0	0	0
propanil[c]	0	0	0	0	0	100	0	0	0	0	0
siduron A[b]	0	0	0	0	32	0	99	1	1	0	0
siduron A[c]	0	0	1	0	20	0	100	1	0	0	0
siduron B[b]	0	0	1	1	13	0	1	100	2	0	1
siduron B[c]	0	0	0	0	42	0	1	100	0	0	0
liunuron[b]	0	0	5	6	0	0	1	2	100	0	0
liunuron[c]	0	0	1	5	0	0	0	0	100	0	1
carbazole[b]	0	8	1	0	0	0	0	0	0	100	0
carbazole[c]	0	1	1	13	0	0	0	1	0	100	1
diflubenzuron[b]	0	2	19	3	0	0	0	1	0	0	100
diflubenzuron[c]	0	8	13	1	0	0	0	0	0	0	100

[a]Phenylurea pesticides, EPA method 532. [b]Concentration 1 mg/L. [c]Concentration 20 mg/L.

sameness threshold being ≥99% of the compared data must have <1% difference), no false positives or negatives were seen. It is remarkable that this was possible in reverse-phase HPLC on a state-of-the-art column where all the peaks had excellent symmetry and all the analytes belonged to the same functional class.

In the absence of a detector that can provide analyte-specific information, analyte attribution is solely based on retention time. Presently, we transform the analyte quantitation calibration data set offline to compose a shape correlation matrix or "shape library". To facilitate utility and wide use, this process will need to be incorporated into software such that the necessary library is automatically generated. When integrated into resident software, such a system will automatically provide identity confirmation based on comparison with analytes in its shape library, if an analyte elutes within corresponding time windows. In IC, certain analytes such as chloride and sulfate are present in nearly all samples: when the identity check for such essentially known constituents begins to fail, chances are that quantitation accuracies are also compromised. Whether or not this is accompanied by a shift in retention time(s), this approach can alert the operator of the need for recalibration.

Finally, that the shape of a chromatographic peak is not a perfect Gaussian is perhaps not to be bemoaned. It is good that the shape of an analyte peak under a set of given conditions is unique; it provides a ready means to distinguish between *the same and not the same*.[30]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c04432.

Calibration mapping for fluoride, ability of a third-degree polynomial to predict the fluoride peak shape, prediction accuracy: present strategy *versus* general Gaussian model, illustrative good *versus* poor match, illustration of a poor match and corresponding regression plot, illustration of index of width mismatch, and so on (PDF)

Fluoride shape calibration matix (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Akinde F. Kadjo − *Thermo Fisher Scientific, Sunnyvale, California 94085, United States;* orcid.org/0000-0002-2718-6023; Email: akindeflorence.kadjo@mavs.uta.edu

Authors

Purnendu K. Dasgupta − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019-0065, United States;* orcid.org/0000-0002-8831-7920

Kannan Srinivasan − *Thermo Fisher Scientific, Sunnyvale, California 94085, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c04432

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank M. Sengupta, S. Bhardwaj, R. Lin, X. Sun, and J. Madden for providing some of the presented data. We also thank C. Phillip Shelor for valuable suggestions. P.K.D. acknowledges the support of National Science Foundation Grant CHE-2003324 and the Hamish Small

Chair Endowment at the University of Texas at Arlington for enabling his participation in this project. For the TOC, we acknowledge the inspiration from Roald Hoffmann for his book of the same name.

REFERENCES

(1) Small, H. *Ion Chromatography*; Plenum Press: New York, 1990.
(2) Åsberg, D.; Chutkowski, M.; Leśko, M.; Samuelsson, J.; Kaczmarski, K.; Fornstedt, T. *J. Chromatogr. A* **2017**, *1479*, 107−120.
(3) Sousa, P. F. M.; de Waard, A.; Åberg, K. M. *Anal. Bioanal. Chem.* **2018**, *410*, 5229−5235.
(4) Dolan, J. W. *LC-GC N. Am.* **2014**, *32*, 546−551.
(5) Bechtold, M.; Felinger, A.; Held, M.; Panke, S. *J. Chromatogr. A* **2007**, *1154*, 277−286.
(6) Herzler, M.; Herre, S.; Pragst, F. *J. Anal. Toxicol.* **2003**, *27*, 233−242.
(7) Diekmann, J. A., III; Cochran, J.; Hodgson, J. A.; Smuts, D. J. *J. Chromatogr. A* **2020**, *1611*, 460607.
(8) American Chemical Society. Gas Chromatography-Mass Spectrometry. A National Historic Chemical Landmark. https://www.acs.org/content/acs/en/education/whatischemistry/landmarks/gas-chromatography-mass-spectrometry.html (accessed June 5, 2020).
(9) Kadjo, A. F.; Liao, H.; Dasgupta, P. K.; Kraiczek, K. G. *Anal. Chem.* **2017**, *89*, 3893−3900.
(10) Chesler, S. N.; Cram, S. P. *Anal. Chem.* **1973**, *45*, 1354−1359.
(11) Mott, S. D.; Grushka, E. *J. Chromatogr.* **1976**, *126*, 191−204.
(12) Kadjo, A.; Dasgupta, P. K. *Anal. Chim. Acta* **2013**, *773*, 1−8.
(13) Craig, L. C. *J. Biol. Chem.* **1943**, *150*, 33−45.
(14) Craig, L. C. *J. Biol. Chem.* **1944**, *155*, 519−534.
(15) Fritz, J. S.; Scott, D. M. *J. Chromatogr. A* **1983**, *271*, 193−212.
(16) Wahab, F. M.; Patel, D. C.; Armstrong, D. W. *LC-GC N. Am.* **2017**, *35*, 846−853.
(17) Kadjo, A. F.; Dasgupta, P. K.; Su, J.; Liu, S.; Kraiczek, K. G. *Anal. Chem.* **2017**, *89*, 3884−3892.
(18) Baeza-Baeza, J. J.; Ortiz-Bolsico, C.; García-Álvarez-Coque, M. C. *Anal. Chim. Acta* **2013**, *758*, 36−44.
(19) Dolan, J. W. *LC-GC N. Am.* **2008**, *26*, 610−616.
(20) Low, G. K. C.; Haddad, P. R.; Duffield, A. M. *J. Chromatogr., Biomed. Appl.* **1984**, *336*, 15−24.
(21) Di Marco, V. B.; Bombi, G. G. *J. Chromatogr. A* **2001**, *931*, 1−30.
(22) Nikitas, P.; Pappa-Louisi, A.; Papageorgiou, A. *J. Chromatogr. A* **2001**, *912*, 13−29.
(23) Caballero, R. D.; García-Alvarez-Coque, M. C.; Baeza-Baeza, J. J. *J. Chromatogr. A* **2002**, *954*, 59−76.
(24) Cavazzini, A.; Dondi, F.; Jaulmes, A.; Vidal-Madjar, C.; Felinger, A. *Anal. Chem.* **2002**, *74*, 6269−6278.
(25) Baeza-Baeza, J. J.; García-Alvarez-Coque, M. C. *J. Chromatogr. A* **2004**, *1022*, 17−24.
(26) Moretti, P.; Vezzani, S.; Garrone, E.; Castello, G. *J. Chromatogr. A* **2004**, *1038*, 171−181.
(27) Baeza-Baeza, J. J.; Ruiz-Ángel, M. J.; García-Álvarez-Coque, M. C. *J. Chromatogr. A* **2007**, *1163*, 119−127.
(28) Liao, H.; Dasgupta, P. K. *Anal. Chem.* **2016**, *88*, 2198−2204.
(29) Dasgupta, P. K. *Environ. Sci. Technol.* **1998**, *32*, 427.
(30) Hoffman, R. *The Same and Not the Same*; Columbia University Press: New York, 1995.

https://dx.doi.org/10.1021/acs.analchem.0c04432
Anal. Chem. 2021, 93, 3848−3856