Analytical Methods

DOI: 10.1002/anie.200605128

Boosting the Throughput of Separation Techniques by "Multiplexing"**

*Oliver Trapp**

The search for highly efficient catalysts, potential drugs and lead structures, and disease markers in medical diagnostics is of considerable scientific and economic interest.[1,2] Parallelized high-throughput assays in combination with sophisticated analytical techniques are currently used to identify and quantify reaction products and to determine conversions.[3–5] Depending on the sample complexity from such high-throughput (ht) assays, chromatography or electrophoresis in combination with spectroscopy or mass spectrometry is required to analyze the composition.[6] As these separation steps are time-consuming and limit the overall performance, efforts have been made to improve speed by optimizing continuous injections and using fast and parallelized separation techniques.[7,8] Miniaturization is another approach to enhance analytical performance, reduce analysis time, and increase sample throughput,[9–11] but it often requires time-consuming optimizations for varying analyte concentrations. Nevertheless, continuous real-time sampling in parallelized high-throughput assays is desirable to perform kinetic studies of catalysts or to detect activation and deactivation processes. However, this is often restricted to single-batch systems or must be performed sequentially, for example, the GC analysis of the gas-phase composition of a 7×7 parallel reactor with a typical analysis time of 20 min per sample (Figure 1a) requires more than 16 h for a single cycle. As in continuous-wave spectroscopy, the overall duty cycle of chromatographic systems is low, and typically most of the acquisition time is spent recording detector noise. Despite these limitations, kinetic studies on large catalyst libraries are important for conclusive insights into reaction mechanisms for the future development of advanced materials and catalysts. The major challenge is to increase the duty cycle of the separation system in order to maximize information and minimize analysis time.

"Multiplexing" techniques, for example, Fourier and Hadamard transformations (FT and HT),[12] are commonly used in spectroscopy[13] and mass spectrometry[14,15] to increase the duty cycle and improve the signal-to-noise ratio (SNR, Fellgett advantage). Previous attempts to apply multiplexing to separation techniques focused only on the SNR improvement for a single sample.[16–19] Here, a novel technique based on multiplexing is described to inherently increase the sample throughput of a chromatographic separation system (cf. Figure 1). In contrast to a conventional chromatographic separation the obtained chromatogram consists of $2^n/2$ overlapped chromatograms, for example, 1023 in Figure 1b, representing the analyte concentrations of 100 different samples (Figure 1c).

The experimental setup for analyzing samples emerging from an N-fold chemical parallel reactor is depicted schematically in Figure 2. The samples are transferred continuously to a six-port multiplexing injector, which is attached to a GC injector (Figure 2a). The samples are rapidly injected by the multiplexing injector onto the separation column according to an n-bit binary pseudo-random sequence (bar code, cf. Figure 2 b, black bar code). In turn the binary pseudo-random sequence is divided into subsequences (bar codelets, cf. Figure 2b, colored bar code assigned to individual reactor channels) encoding individual samples with similar analyte composition but different concentrations. These bar codelets are used to control the time-shifted repetitive injections of each sample by a computer. Structured repetitive injections are necessary to unambiguously identify the individual samples. The analytes of each injection are separated in the separation column yielding time-shifted chromatograms (Figure 2b). However, the measured chromatogram is a convolution of these overlapping time-shifted chromatograms (Figure 2c). The information for individual analytes is conserved in the complete modulation sequence (bar code), and the individual sample information is stored within the subsequences (bar codelets) of the structured sequence. To increase the sample throughput the following were used: 1) long binary pseudo-random sequences to modulate a high number of sample injections, 2) short injection intervals, 3) stable and reproducible sample injections onto the separation column, and finally 4) division of the binary pseudo-random sequences into subsequences with equal and sufficient information content to encode the individual samples. Long n-bit binary pseudo-random sequences with N elements ($N = 2^n - 1$), which are derived from Hadamard matrices, consist of 50% of the elements "0" (no sample injection) and of 50% of the elements "1" (sample injection), and therefore the overall duty cycle of the separation system is increased to 50%. These sequences are unique, and therefore encoded information can be later unambiguously identified by application of the HT. Furthermore, with increasing sequence length N the SNR is improved (maximum gain $= \sqrt{N}/2$).

[*] Dr. O. Trapp
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1, 45470 Mülheim/Ruhr (Germany)
Fax: (+49) 208-306-2995
E-mail: trapp@mpi-muelheim.mpg.de
Homepage: http://www.mpi-muelheim.mpg.de/kofo/institut/arbeitsbereiche/trapp/trapp_e.html

[**] Generous financial support by the MPI für Kohlenforschung and the DFG (Emmy Noether program TR 542/3-1) is gratefully acknowledged. I thank B. List, M. T. Reetz, F. Schüth, and R. N. Zare for helpful discussions and suggestions.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

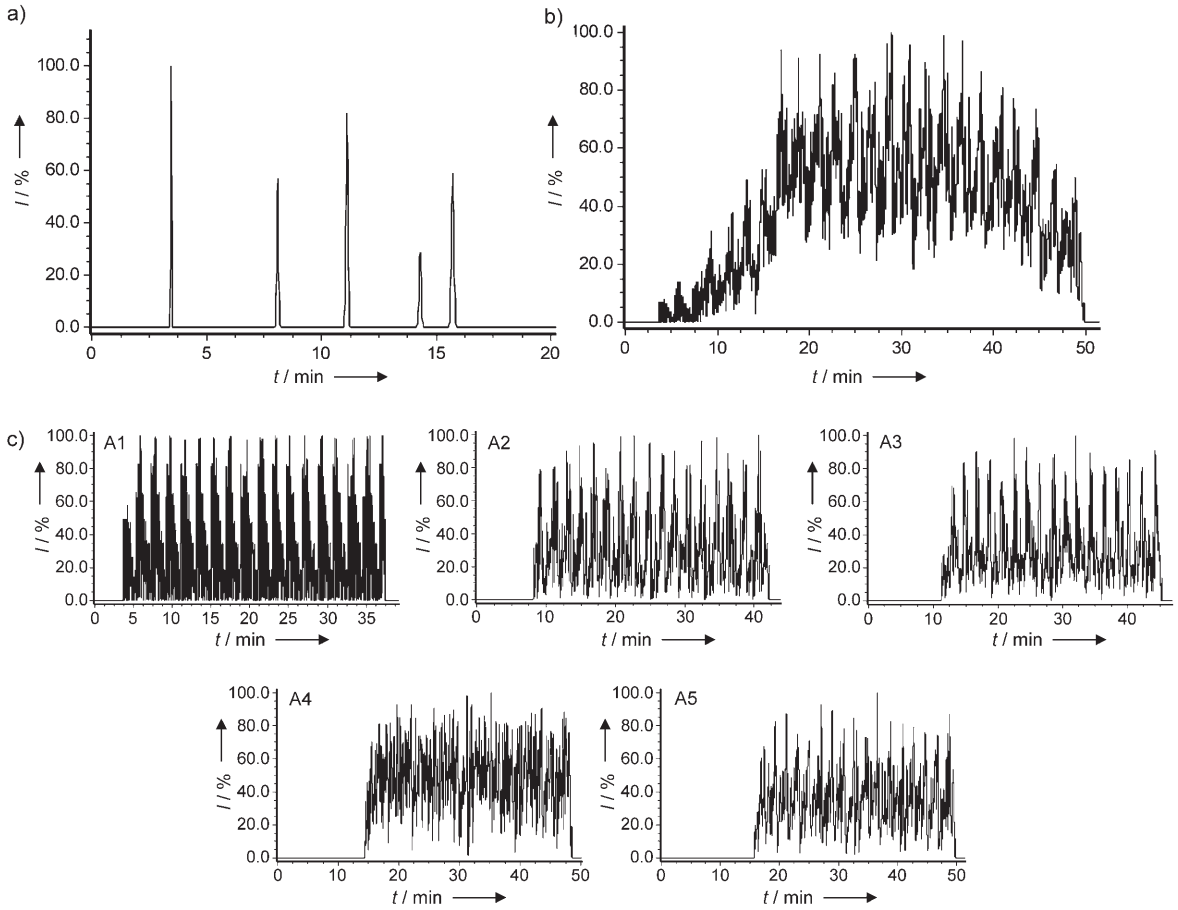

Figure 1. a) Conventional gas chromatogram of a single sample containing five analytes. b) Convoluted chromatogram of 1000 sample injections obtained by high-throughput multiplexing GC (htMPGC) of 100 samples containing five analytes at varying concentrations. c) Time-dependent analyte concentrations (A1 to A5) obtained by deconvolution of (b).

To achieve high sample throughput and to minimize the overall analysis time, short and precise injection intervals ("time bin intervals" Δt) in the range of 500 to 2000 ms and short and highly reproducible injection pulses were applied. As it is necessary to inject different samples consecutively without any delay, a six-channel continuous-flow split/splitless multiplexing injector was designed, built, and mounted with an injection needle onto a conventional split injector of a GC (Figure 2a). The samples are continuously evaporated in the sample ports of the heated multiplexing injector and injected by short pressure pulses of 400 kPa with a solenoid valve as fast as 1 ms, producing highly reproducible sample amounts (see Figures 1b, 2, and 3 in the Supporting Information). Compared to a conventional injection system, for example, a liquid autosampler, these short injection pulses lead to very narrow sample volumes. In average the peak width w_h decreases by a factor of 3 ($\pm 20\,\%$), which in turn improves the separation efficiency ($\approx (1/w_h)^2$) and the multiplexed chromatogram.

Whereas typically the time bin interval Δt also defines the injection pulse duration and leads to rectangular concentration profiles on the separation column, here the injection pulses were chosen considerably shorter (in the range of ms) than the time interval Δt to improve the resolution of the obtained chromatogram and therefore to further increase the information density. The split injector and the electronically controlled gas flow of the GC reduce pressure and gas-flow fluctuations, and thus minimize deviations in the retention times t_R of the injected analytes. Furthermore, through the injector design, sample mixing could be prevented by purging the injection needle and the sample ports with the carrier gas flow from the split/splitless injector of the GC. With this injection system, excellent signal stability could be achieved for repetitive sample injections of analytes with different concentrations (see Figure 1b in the Supporting Information). Peaks of same height can be assigned to the same analyte; deviations are contributed to peak overlapping of different samples. With the multiplexing injector up to 3000 injections h^{-1} ($\Delta t = 0.6$ s, 50 % duty cycle) from different sample ports were achieved.

To unambiguously self-encode individual sample injections the n-bit binary pseudo-random sequence (bar code) is divided into sections of equal information content (bar codelets). In the experiments each sample was assigned to a number of repetitive injections ranging from 50 down to 5, depending on the number of analytes per sample. With increasing sample complexity more repetitive injections are necessary.

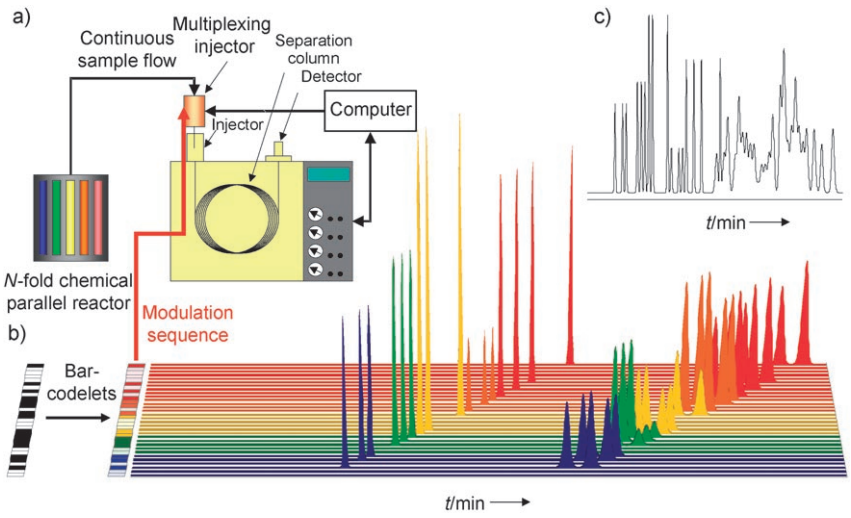

Figure 2. High-throughput multiplexing gas chromatography (htMPGC). a) Schematic experimental setup for analyzing the sample composition from an *N*-fold parallel reactor. The samples are sequentially injected by short pressure pulses (1–5 ms) onto the separation column by the multiplexing injector according to an *n*-bit binary pseudo-random sequence ($n=5$) consisting of 2^n-1 time bins with time bin intervals Δt on the order of seconds. This sequence is divided into bar codelets each encoding an individual channel of the *N*-fold parallel reactor for unambiguous identification of the samples. b) Temporally shifted chromatograms obtained by repetitive sample injections according to the bar codelets of the *n*-bit binary pseudo-random sequence. c) Convoluted chromatogram, which represents the sum of the chromatograms depicted in (b).

The minimum number of necessary repetitive injections r_{min} can be obtained from the number of elements "1" in the modulation sequence and the maximum number of samples i_{max} that can be analyzed [Eq. (1)]. For the unambiguous

$$r_{min} = \frac{N}{2\,i_{max}} \tag{1}$$

determination of a single analyte j of sample i, the product $(j_{max} \times i_{max})$ of the maximum numbers of analytes j_{max} and maximum number of samples i_{max} must not exceed the maximum number of data points acquired by high-throughput multiplexing GC (htMPGC). The number of data points can be calculated from the sum of the sequence length N and the maximum retention time t_R^{max} divided by the time bin interval Δt, and multiplied by the data acquisition frequency f [Hz]. Therefore, the maximum number of samples i_{max} that can be unambiguously determined is given by Equation (2).

$$i_{max} = f\frac{N + (t_R^{max}/\Delta t)}{j_{max}} \tag{2}$$

The throughput power P of an htMPGC analysis is expressed by the ratio of the maximum number of samples i_{max} and the total duration t_{max} [Eq. (3)]. Here the capability of this

$$P = \frac{i_{max}}{t_{max}} \tag{3}$$

technique is demonstrated by the analysis of 100 samples containing various amounts of five analytes (methanol, *n*-butanol, *tert*-butanol, *n*-heptane, and toluene; cf. Figure 3).

The samples were injected consecutively with the six-channel continuous-flow split/splitless multiplexing injector onto a separation column according to a predefined *n*-bit binary pseudo-random sequence, time bin interval Δt, and number of repetitive injections. The results of three experiments, performed with an 11-bit sequence (2047 elements) with decreasing time bin intervals Δt (from 2 s to 1 s) and decreasing number of repetitive injections per sample (from 25 to 5) are depicted in Figure 3 a–c (throughput of 32, 150, and 299 samples per hour, respectively). By reducing the time bin interval Δt to 600 ms (3000 sample injections per hour, 5 repetitive injections) the throughput was pushed to 453 samples per hour, corresponding to an increase by a factor of 38 compared to the respective conventional separation (see Figure 7 in the Supporting Information). This represents an extraordinary improvement over a state-of-the-art separation, where only 12 samples per hour

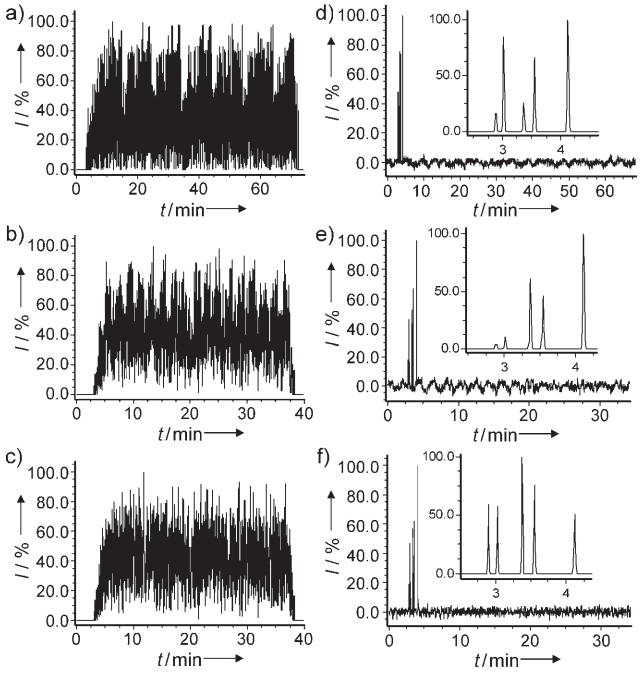

Figure 3. Experimental htMPGC data. Samples with five analytes of varying concentrations were injected by the six-channel continuous-flow split/splitless multiplexing injector according to a 11-bit (2047 time bins) binary pseudo-random sequence with injection pulses of 1 ms. a) Time bin interval $\Delta t = 2$ s (900 injections per hour), 25 injections per sample (32 samples per hour). b) $\Delta t = 1$ s (1800 injections per hour), 10 injections per sample (150 samples per hour). c) $\Delta t = 1$ s (1800 injections per hour), 5 injections per sample (299 samples per hour). d)–f) Hadamard transformed chromatograms of raw data from (a–c), respectively. The inset chromatograms show the composition of selected samples after data deconvolution.

can be analyzed.

Despite the rapid modulation (900 injections per hour for Figure 3a, 1800 injections per hour for Figure 3c,e) a baseline signal can be observed in these chromatograms indicating the high separation efficiency. Because of the different sample compositions the envelope has a pronounced sawtooth structure for longer time bin intervals Δt and more repetitive injections. Figure 3d–f are the corresponding HT overview chromatograms, which are characterized by an apparent noise level arising from the convolution of the different analyte concentrations of the samples. The insets are selected chromatograms of single samples obtained after data deconvolution.

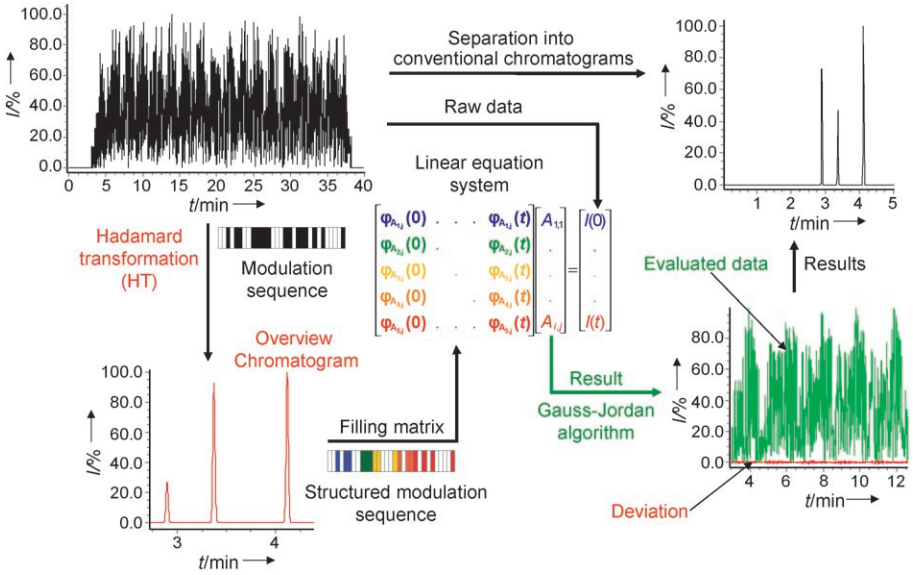

Figure 4. Flowchart of the data deconvolution process of chromatograms obtained by htMPGC to determine the relative analyte concentrations of the individual samples. The HT of the raw chromatogram (top left) gives an overview chromatogram with the peaks of the analytes. The peak information of the analytes (retention times and peak shapes) is used to fill matrix $\varphi_{A_{i,j}}(t)$ with normalized peak profiles according to the structured modulation sequence. Each row represents the chromatogram of the repetitive injection of a single analyte of a single sample. Multiplication of this matrix with the analytes' concentration vector of all samples equals the raw chromatogram represented as vector $I(t)$. This linear equation system is solved by a Gauss–Jordan algorithm to obtain the analytes' concentrations to calculate conventional chromatograms (top, right, chromatogram of sample 7) of the individual samples.

To obtain the relative concentrations of the analytes in each sample, a deconvolution algorithm for the raw data was developed and applied to these data sets (Figure 4). In the first step the raw chromatogram is Hadamard transformed to obtain an overview chromatogram containing the information about the number of analytes and their retention times t_R and peak width w_h. This data is used to fill a matrix $\varphi_{A_{i,j}}(t)$ according to the structured modulation sequence. Therefore in the first step elution profiles considering the chromatographic data and the bar codelets are calculated for the single analytes, representing a single row of this matrix. The complete matrix $\varphi_{A_{i,j}}(t)$ represents the peak shapes of the individual sample injections and their analytes. The product of the matrix with the unknown concentration vector (A_i) is the convoluted raw data (cf. Figure 4). To solve this linear equation system a Gauss–Jordan algorithm is applied. It was found that the computational time could be reduced considerably when the chromatogram and linear equation system is fractioned into shorter sections and these sections are solved separately by a Gauss–Jordan algorithm (see Figures 13 and 14 in the Supporting Information). The final result is the concentration vector (A_i) which can be multiplied with the normalized analyte peak shape matrix $\varphi_{A_{i,j}}(t)$ to calculate a conventional chromatogram of a single sample. The relative peak areas were determined by integration of these subchromatograms. Deviations in the peak area integration caused by the deconvolution process are determined by subtraction of the deconvoluted chromatograms from the raw chromatogram (Figure 4). In Table 1 statistical data of the evaluation of the experiment depicted in Figure 3b,e with 102 samples is summarized. The maximum mean standard deviation between experimental data and the evaluated data is 1.47% and the average standard deviation is 1.02% of the peak area ratios in relation to the first eluted peak of each sample. This data was also compared to the peak area ratio of the conventionally measured samples. There, the maximum mean deviation is 0.63% and the average is 0.36%, which demonstrates a high agreement between conventional GC and htMPGC.

Table 1: Statistics of the results obtained by the deconvolution of a htMPGC analysis of 102 samples each containing five analytes.[a]

Injection channel	Mean standard deviation of peak ratios [%]	Mean deviation from conventionally determined peak ratios [%]
1	0.98	0.10
2	0.91	0.35
3	1.47	0.41
4	1.19	0.63
5	0.65	0.37
6	0.89	0.28
Average	1.02	0.36

[a] Binary 11-bit (2047 time bins) pseudo-random sequence, injection pulses of 1 ms, time bin interval $\Delta t = 1$ s, and 10 injections per sample.

The technique presented here increases the information content and at the same time decreases the overall time for the chromatographic analysis. Time-resolved quantification of analytes, even in more complex mixtures, could be useful for real-time analysis in kinetic and mechanistic studies conducted in parallelized chemical reactors. Coupling with other analytical techniques, for example, mass spectrometry or a second chromatographic dimension, could further increase the information density and allow identification of products at the same time. The application of this approach to

other chromatographic and electrophoretic techniques and in particular to ultrafast separation techniques can be envisaged to maximize the sample throughput in future applications.

Received: December 19, 2006
Revised: March 23, 2007
Published online: June 21, 2007

Keywords: analytical methods · chromatography · Hadamard transformation · high-throughput screening · multiplexing

[1] M. T. Reetz, *Angew. Chem.* **2001**, *113*, 292 – 320; *Angew. Chem. Int. Ed.* **2001**, *40*, 284 – 310.
[2] P. S. Dittrich, A. Manz, *Nat. Rev. Drug Discovery* **2006**, *5*, 210 – 218.
[3] K. Jähnisch, V. Hessel, H. Löwe, M. Baerns, *Angew. Chem.* **2004**, *116*, 410 – 451; *Angew. Chem. Int. Ed.* **2004**, *43*, 406 – 446.
[4] C. Jäkel, R. Paciello, *Chem. Rev.* **2006**, *106*, 2912 – 2942.
[5] F. Schüth, L. Baumes, F. Clerc, D. Demuth, D. Farrusseng, J. Llamas-Galilea, C. Klanner, J. Klein, A. Martinez-Joaristi, J. Procelewska, M. Saupe, S. Schunk, M. Schwickardi, W. Strehlau, T. Zech, *Catal. Today* **2006**, *117*, 284 – 290.
[6] R. Aebersold, M. Mann, *Nature* **2003**, *422*, 198 – 207.
[7] S. Yao, D. S. Anex, W. B. Caldwell, D. W. Arnold, K. B. Smith, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5372 – 5377.
[8] B. M. Paegel, C. A. Emrich, G. J. Wedemayer, J. R. Scherer, R. A. Mathies, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 574 – 579.
[9] D. J. Harrison, K. Fluri, K. Seiler, Z. Fan, C. S. Effenhauser, A. Manz, *Science* **1993**, *261*, 895 – 897.
[10] G. M. Whitesides, *Nature* **2006**, *442*, 368 – 373.
[11] D. Janasek, J. Franzke, A. Manz, *Nature* **2006**, *442*, 374 – 380.
[12] R. R. Ernst, W. A. Anderson, *Rev. Sci. Instrum.* **1966**, *37*, 93 – 102.
[13] M. B. Comisarow, A. G. Marshall, *Chem. Phys. Lett.* **1974**, *25*, 282 – 283.
[14] O. Trapp, J. R. Kimmel, O. K. Yoon, I. A. Zuleta, F. M. Fernandez, R. N. Zare, *Angew. Chem.* **2004**, *116*, 6703 – 6707; *Angew. Chem. Int. Ed.* **2004**, *43*, 6541 – 6544.
[15] A. G. Marshall, *Fourier, Hadamard, and Hilbert Transforms in Chemistry*, Plenum, New York, **1982**.
[16] R. Annino, M.-F. Gonnord, G. Guiochon, *Anal. Chem.* **1979**, *51*, 379 – 382.
[17] D. C. Villalanti, M. F. Burke, J. B. Phillips, *Anal. Chem.* **1979**, *51*, 2222 – 2225.
[18] T. Kaneta, Y. Yamaguchi, T. Imasaka, *Anal. Chem.* **1999**, *71*, 5444 – 5446.
[19] K. L. Braun, S. Hapuarachchi, F. M. Fernandez, C. A. Aspinwall, *Anal. Chem.* **2006**, *78*, 1628 – 1635.

