ELSEVIER

Journal of Chromatography A, 974 (2002) 169–184

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Unravelling the composition of very complex samples by comprehensive gas chromatography coupled to time-of-flight mass spectrometry
Cigarette smoke

Jens Dallüge[a,*], Leo L.P. van Stee[a], Xiaobin Xu[b], Jonathan Williams[b], Jan Beens[a], René J.J. Vreuls[a], Udo A.Th. Brinkman[a]

[a]*Department of Analytical Chemistry and Applied Spectroscopy, Vrije Universiteit, de Boelelaan 1083, 1081 HV Amsterdam, Netherlands*

[b]*Air Chemistry Department, Max Planck Institute for Chemistry, J.J. Becher Weg 27, 55020 Mainz, Germany*

Abstract

The potential and current limitations of comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GC×GC–TOF-MS) for the analysis of very complex samples were studied with the separation of cigarette smoke as an example. Because of the large number of peaks in such a GC×GC chromatogram it was not possible to perform manual data processing. Instead, the GC–TOF-MS software was used to perform peak finding, deconvolution and library search in an automated fashion; this resulted in a peak table containing some 30 000 peaks. Mass spectral match factors were used to evaluate the library search results. The additional use of retention indices and information from second-dimension retention times can substantially improve the identification. The combined separation power of the GC×GC–TOF-MS system and the deconvolution algorithm provide a system with a most impressive separation power.

Keywords: Gas chromatography, comprehensive two-dimensional; Cigarette smoke; Data processing

1. Introduction

In the past few years, comprehensive two-dimensional gas chromatography (GC×GC) has been shown to be an extremely powerful technique for the analysis of complex samples [1–4]. From among the many application areas that are now being studied, two may be cited here—the analysis of mixtures of a petrochemical nature, and mixtures of polyhalogenated compounds such as polychlorobiphenyls, -dibenzodioxins and -dibenzofurans. Impressive results have been obtained in terms of separation efficiency and, also, compound classification based on the presence of ordered structures in the GC×GC chromatogram of (classes of) structurally related compounds. This has considerably helped in unravelling the composition of petrochemical samples (limited

*Corresponding author. Tel.: +31-20-444-7525; fax: +31-20-444-7543.

E-mail address: acas@chem.vu.nl (J. Dallüge).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)01384-5

number of classes) and halogenated compounds (large number of individual congeners available).

The aforementioned analyses represent the state of the art, but with time-of-flight mass spectrometers being available today to provide compound-specific information, one can try to go one step further. One route, the route we intend to follow here, is to analyse samples containing a huge number of analytes, partly identified and partly unknown, which belong to many different classes. The power of the ordered-structure principle now will be, at best, limited. In our opinion, cigarette smoke is a good example of that sample type.

Cigarette smoke is a very complex mixture and, over the years, many attempts have been made to separate and identify its constituents [5]. More than 4700 compounds have been identified in mainstream cigarette smoke [6] and, in some reports, the number of unidentified compounds has been estimated to be as high as 100 000 [7]. Clearly, conventional GC cannot provide sufficient resolution to separate even a fair proportion of these compounds. As a consequence, most methods used for the analysis of cigarette smoke focus on a relatively small number of target analytes. Appropriate sample preparation techniques are usually applied to isolate the target analytes and selective detection is frequently used to further increase the performance. Several publications report the use of coupled chromatographic (heart-cutting) techniques, e.g. LC–GC or GC–GC [8,9]. However, even with these techniques, insufficient separation is obtained. Moreover, as is well known, the inherent drawback of heart-cutting techniques is that detailed information is obtained only for the (few) fraction(s) subjected to the two-dimensional operation. If the operation is extended to include the complete mixture, the total time of analysis (which may well be several days) becomes the main stumbling block. With GC×GC, where the second-dimension separations are completed within the run time of the first-dimension separation, the latter problem no longer exists. This makes it highly interesting to study this alternative to the more classical approaches. However, it should immediately be made clear that the present study has an exploratory nature and merely intends to study the general potential of the new technique. A final solution to cigarette-smoke separation should certainly not be expected here.

2. Instrumentation

2.1. Sampling system

The sampling system consisted of a thermal desorber (Unity, Markes International, Llantrisant, UK), controlled by the GC computer. The thermal desorber allows the transfer of collected volatile compounds from a sample tube (Silcosteel, 89 mm× 5 mm I.D.) into a cold trap (quartz, 12 mm long× 2 mm I.D.) for focusing and subsequent injection. The sample tube contained three beds of sorbents in a sequence of increasing adsorption strength, i.e. Tenax TA, Carbopack B, and Carboxen 1000. The cold trap contained two beds of sorbents, i.e. Tenax TA and Carbograph; it can be cooled to −10 °C using a two-stage Peltier element, which enables quantitative retention of compounds as volatile as ethane.

Mainstream cigarette smoke of a (filterless) roll-your-own cigarette was first collected in an Erlenmeyer flask. The smoke was then drawn through the sample tube for 20 s using a vacuum pump. The sampling tube was placed in the thermal desorber and purged to waste for 4 min at ambient temperature using helium at a flow-rate of 50 ml/min to remove water vapour and oxygen. Subsequently, the sample tube was switched in-line with the cold trap and was heated to 280 °C at 20 °C/s to desorb the organic compounds and transfer them to the cold trap which was held at −10 °C. Desorption was performed for 5 min. Next, the sample was injected into the GC system by heating the cold trap to 250 °C in less than 5 s. During injection, the direction of the flow through the cold trap was reversed. The flow from the cold trap was split prior to entering the GC column, with 2.5 ml/min being directed to the GC and 5 ml/min to waste.

2.2. GC×GC system

The GC×GC–time-of-flight mass spectrometry (TOF-MS) system consisted of a HP 6890 (Agilent Technologies, Palo Alto, CA, USA) gas chromatograph and a Pegasus II time-of-flight mass spectrometer (LECO, St. Joseph, MI, USA). A 30 m× 0.25 mm I.D., 1 μm DB-5 column (Agilent Technologies) was used as the first-dimension column, and a 1 m×0.1 mm I.D., 0.1 μm Carbowax column

(Quadrex, Woodbridge, CT, USA) as the second-dimension column. The columns were connected by means of a press-fit connector. The second-dimension column was installed in a separate oven, which was housed in the main GC oven. The separate oven provided a more flexible system since it allows fine-tuning of the retention in the second column by using a somewhat higher or lower temperature relative to the first-dimension column [10]. Cryogenic modulation was performed using a jet-type modulator (KT 2001; Zoex, Lincoln, NE, USA) which was installed at the top of the second-dimension column. This modulator consists of two cold and two hot jets, with the nozzles providing the cold jets mounted orthogonal to the hot jets. Nitrogen gas is cooled by heat exchange through copper tubing immersed in liquid nitrogen outside the GC system and delivered through vacuum-insulated tubing to the cold jets, which provide two continuous jets of approx. 10 l/min of cold nitrogen gas. Remobilization of the analytes is achieved by alternatively applying short pulses of hot air (at 70 l/min). The modulation time was 6 s; during the first 0.3 s of each modulation cycle, the ''upstream'' hot jet (i.e. the one close to the GC injector) was activated. After a delay of 0.4 s, the second hot jet (''downstream'') was activated for 0.3 s. When the hot downstream pulse is fired the analytes are effectively injected into the second-dimension column. A detailed description of the setup and its operation is given elsewhere [11].

The carrier gas was helium (99.999%, Praxair, Oevel, Belgium) at a constant pressure of 276 kPa. The temperature of the first-dimension column was programmed from 50 (no hold) to 200 °C at 2.5 °C/min; the second-dimension column was programmed from 30 to 180 °C at 2.5 °C/min. The time-of-flight mass spectrometer was operated at a spectrum storage rate of 100 Hz, using a mass range of m/z 35–300 and a multi-channel plate voltage of −1800 V.

For data transformation into a two-dimensional array and visualization as a contour plot, two additional programmes were used, since the original Pegasus software (version 1.40) did not provide these options. One programme converted the raw data into a two-dimensional array (laboratory written software) and the second generated contour plots from this array (''Transform'', part of Noesys software package; Research Systems International, Crowthorne, UK).

3. Results and discussion

3.1. GC×GC separation

The GC×GC analyses were performed on a system with a non-polar thick-film column in the first dimension and a second-dimension column containing a thin-film Carbowax stationary phase. This column combination was used previously for the analysis of air samples and proved to be a good choice for volatile analytes [12]. However, the thick-film column causes the elution temperatures in the first dimension to be high. Since the Carbowax phase can be used only up to a temperature of 250–260 °C, the present GC×GC analyses were restricted to the volatile compounds in cigarette smoke. The GC×GC separation was performed in the elution range of *n*-butane to *n*-pentadecane.

Fig. 1 shows a major part of the GC×GC total ion chromatogram (TIC) contour plot in the elution range of *n*-heptane to *n*-dodecane. The data were baseline-corrected by the Pegasus GC–MS software before generating the contour plot to improve the graphical presentation. It should be noted that, since the software was not able to handle the very different peak shapes in this complex chromatogram, excessive broad peaks and long tails were not recognized in some cases and do not show up in the corrected chromatogram. The final part of the chromatogram (retention times of 2500–3500 s) is not included since it contains some large peaks which cover the entire contour plot in that area and make it impossible to recognize smaller peaks. This is demonstrated by the GC×GC plot shown in the insert of Fig. 1, in this case without baseline correction. The major peaks appear as intense vertical bands due to tailing in the second dimension; of these, the nicotine peak is indicated by an arrow.

The contour plot of Fig. 1 reveals the presence of a very large number of peaks which essentially fill the entire GC×GC plane. Most strikingly, in the lower part of the plot, viz. at second-dimension retention times of 0.25–1 s, a virtual continuum of non-resolved peaks can be seen. This band mainly contains non-polar saturated and unsaturated hydro-

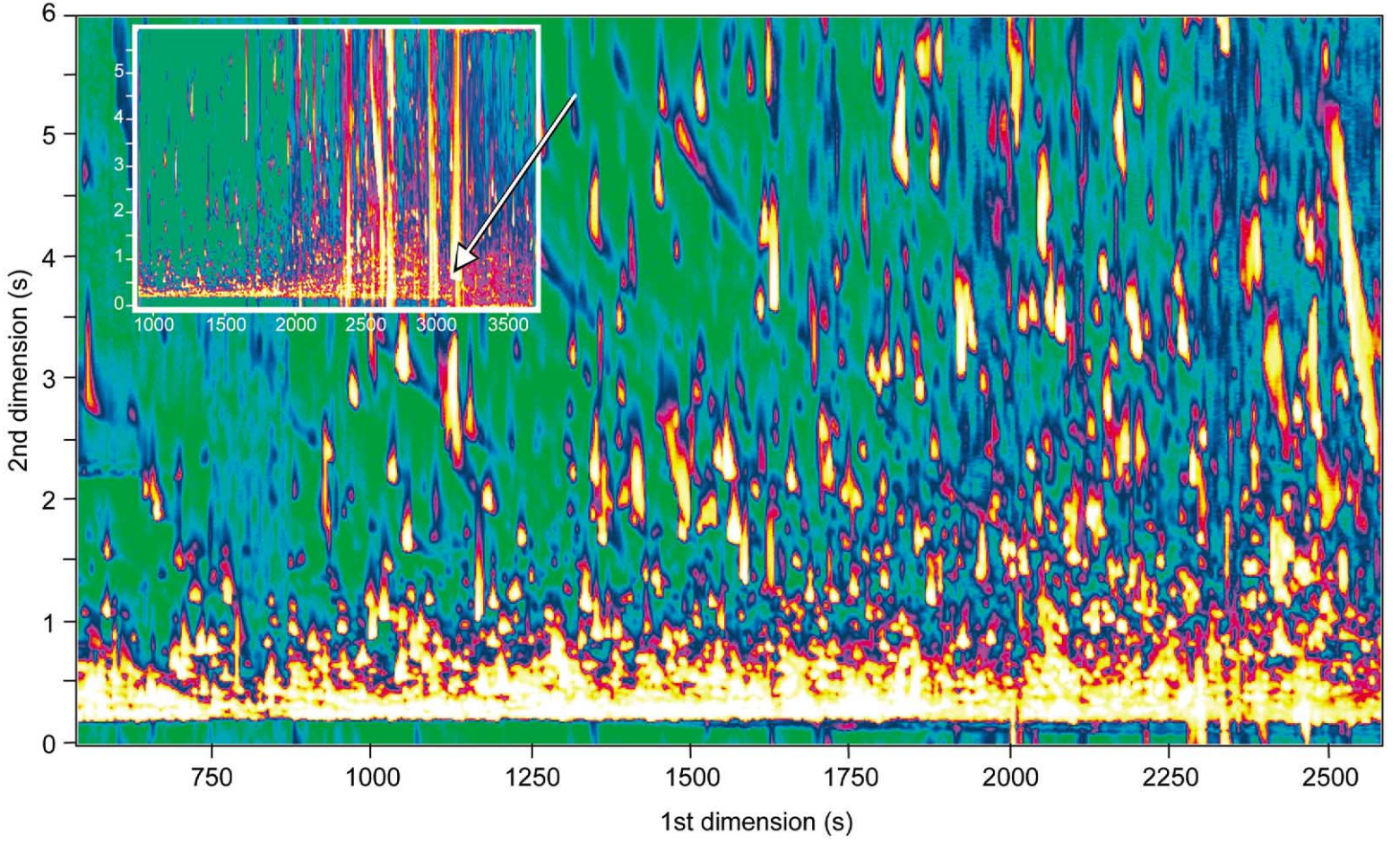

Fig. 1. GC×GC–TOF-MS contour plot (total ion chromatogram, TIC) of cigarette smoke showing the first-dimension range of 500–2600 s.

carbons, which have very low activity coefficients for the second-dimension stationary phase and, consequently, very short second-dimension retention times. Compounds with higher second-dimension retention times are seen to be much better separated.

In GC×GC, analytes should elute within their own modulation cycle; that is, there should be no ''wrap around'' in order to prevent their co-elution with other analytes eluting during subsequent modulation cycle(s). On the other hand, long modulation times will adversely affect the separation in the first dimension [13]. In this study, a modulation time of 6 s was used as a compromise, since most analytes were found to have second-dimension retention times of less than 6 s. The small number of exceptions could be tolerated. These analytes were basically very polar compounds such as formic and ethanoic acid, which could be recognized by their peak widths: since the second-dimension chromatograms are essentially isothermal chromatograms, peak width increases with retention time.

3.2. Data processing

3.2.1. Contour plot

Contour plots were mainly used to evaluate the general quality of the separation and for manual peak identification. First, total or extracted ion chromatograms were exported from the GC–TOF-MS software. Next, these data were converted into a two-dimensional matrix by means of laboratory written software, with the number of columns equalling the number of data points of a single second-dimension separation and the number of rows the number of modulations. Contour plots were generated from this matrix using the ''Transform'' software.

3.2.2. Manual identification of single peaks

In order to identify individual peaks in the contour plot, first their total retention time was calculated by adding the first- and second-dimension retention times, to generate the retention time in the raw data file. These peaks were then identified in this file by the same procedure as is used in conventional one-dimensional GC. As an example, in Fig. 2 a blow-up of a part of the GC×GC chromatogram of Fig. 1 is shown, with the assignment of some 20 peaks being indicated (see Table 1). It should be noted that the peak identification of Fig. 2 is based on library mass spectra only [i.e. no retention indices (*I* values) were used, and no pure standards were analyzed].

The above example shows that manual peak finding in the contour plot and subsequent identification is a possibility; however, there are some serious limitations. First, different peaks may have markedly different peak heights; it is, therefore, often necessary to use different scales for the *Z* (response) axis to detect both large and small peaks in the contour plot. Secondly, and more importantly, co-elution is a very general phenomenon in such highly complex chromatograms, even with GC×GC separation. This occurs even in the less crowded parts of a chromatogram, as is illustrated in the insert of Fig. 2. A rather small peak, that was identified as 3-methyl-1-butanol, was found to elute very close to the front of the much larger peak of pyrazine, although not the slightest indication of such co-elution could be gleaned from peak 9 in the contour plot. In such situations, the appropriate mass spectra had to be deconvoluted prior to identification. A third argument against the use of manual peak identification is, of course, the large number of peaks appearing in each chromatogram, which will make the procedure very laborious and time-consuming. Consequently, automated data processing, including peak finding, deconvolution and library searching, is highly desirable. In this study, only software tools from conventional one-dimensional gas chromatography were available for the (only partially automated) processing of the raw GC×GC chromatogram.

3.2.3. Automated data processing—generation of the peak table

As the first step of automated data processing, the GC–TOF-MS software of the LECO Pegasus was used to find all peaks in the raw GC×GC chromatogram with a signal-to-noise ratio larger than 30 in at least one of the extracted ion traces. The threshold value of 30 is, admittedly, somewhat arbitrary. However, in our experience, it offers a fair compromise between being swamped by an avalanche of peaks and losing too many relevant compounds. Next, for all peaks found, mass spectra were calculated by means of deconvolution. The maximum number of peaks that can be listed in the peak table is limited to 9999 by the GC–TOF-MS software.

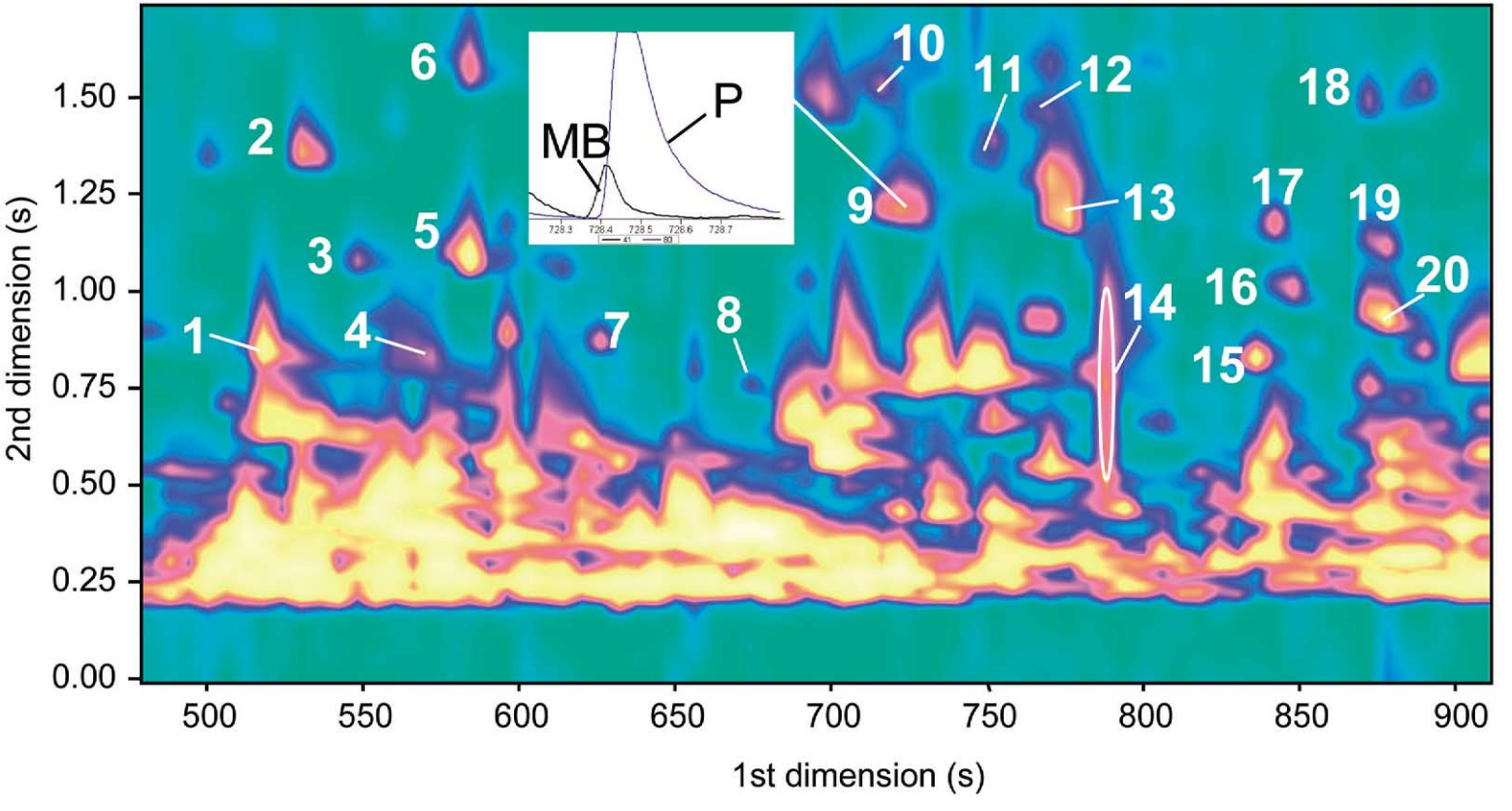

Fig. 2. Blow-up of part of the GC×GC–TOF-MS contour plot of Fig. 1; for peak assignment, see Table 1. The insert shows part of the second-dimension chromatogram recorded at ${}^{1}t_{R}$ 722–723 s; 3-methyl-1-butanol (MB), which co-elutes with pyrazine (P), is not visible in the contour plot.

Table 1
Peak assignment for Fig. 2

No.	Analyte	No.	Analyte
1	2-Butenal	11	Isothiocyanatomethane
2	3-Butenenitrile	12	Dimethylaminoacetonitrile
3	1-Butanol	13	Pyridine
4	Butanenitrile	14	Pyrrole
5	2-Butenenitrile	15	3-Methylpyridazine
6	1-Chloro-2-propanone	16	Pentanenitrile
7	2-Methylenebutanenitrile	17	Methallyl cyanide
8	3-Chloro-2-butanone	18	3-Pentenenitrile
9	Pyrazine	19	2-Pentenenitrile
10	Thiazole	20	3-Methyl-2-butenal

Because the raw chromatogram contained many more peaks, separate data processing was performed for each (adjacent) 500-s window. A library search was now carried out for all peaks using the NIST/EPA/NIH 98 mass spectral library and the results were combined in a single peak table. The three steps—peak finding, deconvolution and library search—were performed by the GC–TOF-MS software fully automated without any user interaction. It should be pointed out that this software feature, which was not only essential for the processing of such complex chromatograms as discussed here, is—to the best of our knowledge—a unique characteristic of this GC–TOF-MS software. Therefore, the data processing procedure and its results are explained in some detail in this and the following sections.

The data processing for the chromatogram displayed in Fig. 1 took approximately 7 h, which is, admittedly, a very long time but, at the same time, a marked improvement over what a manual search would require. Moreover, such processing can be performed overnight and/or by using additional computers. In addition, the on-going rapid development of new computer hardware will no doubt help to accelerate data processing in the future.

The final outcome was a large table containing the library search results such as compound name, mass spectral match factors and CAS number for all peaks. The NIST library search provided three numbers for each analyte—''similarity'', ''reverse'' and ''probability'', which will be explained in detail in Section 3.2.4. The peak table was exported to a spreadsheet programme for further processing.

As briefly illustrated in Section 3.2.2, the GC×GC chromatogram contains a large number of co-eluting peaks. To quote an example, non-polar analytes, such as alkanes and olefins, eluted in a band at the bottom of the GC×GC contour plot. Although the peaks eluting in this band were not baseline-separated, the deconvolution algorithm of the GC–TOF-MS software could calculate ''pure'' mass spectra for many of these analytes and identify them. Fig. 3A shows part of the contour plot of Fig. 1. The vertical line in Fig. 3A indicates a single second-dimension chromatogram. Since colours or shades of grey have some limitations when used to indicate signal intensities, the second-dimension chromatogram is shown separately in Fig. 3B. The GC–MS software could deconvolute and identify a number of compounds in this band; in this 2-s part of a single second-dimension chromatogram alone, 18 peaks were found by the GC–TOF-MS software, which are indicated by horizontal lines in Fig. 3B. Half of them could be actually identified by their mass spectra; the identified compounds are shown on the right-hand side of the figure. As an illustration of the quality of the mass spectra obtained, Fig. 3C and D show the deconvoluted mass spectrum of the peak found at the second-dimension time of 0.24 s and the corresponding library spectrum, respectively.

3.2.4. Processing of the peak table

The peak table generated as described in Section 3.2.3 contained more than 30 000 peaks for the entire 60-min GC×GC chromatogram. However, since most compounds were modulated into two or three second-dimension peaks, their names appeared two to three times in the peak table. In addition, large

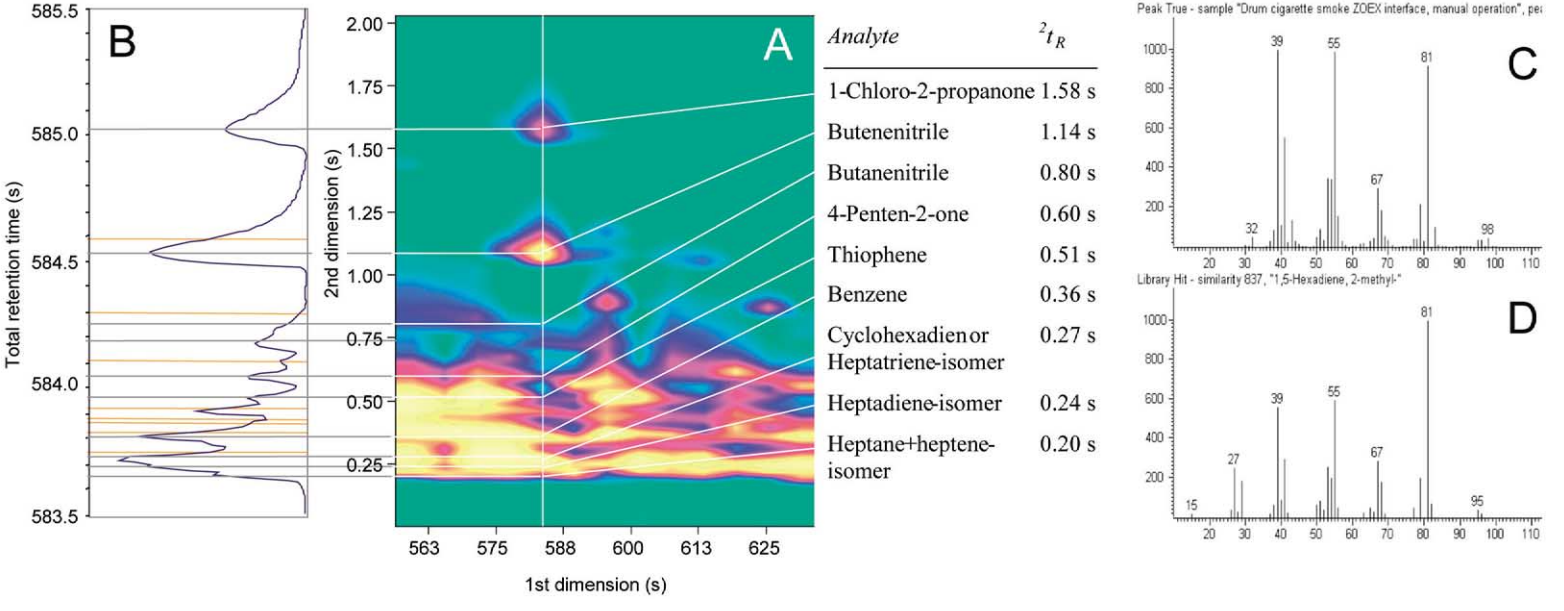

Fig. 3. (A) Detail of the GC×GC contour plot of Fig. 1. The vertical line at 583 s indicates the second-dimension chromatogram, which is shown separately in (B). In (B) the horizontal lines indicate the positions where peaks were found by the deconvolution algorithm of the GC–TOF-MS software. Provisional identifications are summarized on the right-hand side of the figure. (C) The deconvoluted mass spectra of the peak at 0.24 s. (D) The corresponding library spectrum.

peaks that showed fronting and/or tailing in the first dimension showed up more often in the table because they were modulated into more than three second-dimension peaks. Finally, peaks having an elongated shape in the second dimension due to long second-dimension retention times, fronting or tailing, were often identified as several individual peaks in a single second-dimension chromatogram—a problem that is caused by the software, as was explained in Ref. [13]. As a consequence, the peak table contained some 7500 rather than 30 000 different compound names. It should, however, be noted that isomers with similar mass spectra were sometimes given the same compound name (see Section 3.2.4.2).

The aim of further data processing was to obtain an overview of the sample constituents, i.e. to compile a list of provisionally identified compounds. The data processing was performed similar to the strategy used in the ''screening for unknowns'' described in detail in Fig. 7 of Ref. [13]. Because it was not possible to perform a manual verification for all 30 000 peaks (cf. above), mass spectral match factors had to be used to decide whether a peak was correctly identified or not. First, criteria had to be defined to correlate the match factors and the reliability of the identification by the library search.

3.2.4.1. Similarity/reverse

Similarity and *reverse* are mass spectral match factors in the range 0–999, with a higher value meaning a better fit. ''Similarity'' describes how well the library hit matches the peak when using all masses, and ''reverse'' how well the library hit matches the peak when using only the masses present in the library spectrum. The peak table was found to contain many peaks which were obviously not correctly identified, such as compounds as divergent as, for example, l-gala-l-ido-octose and rifamycin that are not at all GC-amenable. Such compounds typically had similarity and reverse factors below 750 and 800, respectively. Therefore, as a first step, peaks displaying such low values were removed from the peak table—in total, some 20 000 peaks. The low match factors were caused by, for example, low peak intensities, which resulted in ''noisy'' mass spectra, non-successful deconvolution, and/or mass spectra that had no counterpart in the NIST library. It should be noted that some of these 20 000 peaks could still be identified correctly. However, this had to be done on the basis of a manual peak identification, but was impossible by means of the procedure based on the match factors.

Earlier experience had shown that similarity and reverse numbers above 800 and 900, respectively, indicate that an acquired mass spectrum usually shows a good match with the library spectrum. About 1500 peaks (520 different compound names) met that requirement. Table 2 shows a selection of compound classes identified by the library search. The remaining 8000–9000 peaks, with match factors in what can be called the ''grey zone'', generally showed satisfactory library matches. However, for a number of these peaks, identification was not correct. Such peaks require more careful attention, for example manual inspection of the mass spectra and/or the use of additional data, such as *I* values. In this study, only those target compounds listed in Table 5 and with ''grey zone'' parameters were manually inspected.

3.2.4.2. Probability factor and apex plots

The similarity and reverse factors indicate only how well a mass spectrum matches the library spectrum. In many cases, however, different compounds, for example isomers, have very similar mass spectra. Therefore, additional criteria are required to

Table 2
Groups of analytes found in cigarette smoke with similarity >800 and reverse >900

Compound group	Number of peaks	Number of different compounds
(Alkyl-)benzenes	115	31
(Alkyl-)naphthalenes	17	7
(Alkyl-)phenols	92	19
(Alkyl-)pyridines	53	11
(Alkyl-)pyrroles	39	12
Alkyl-nitriles	51	14
Alkenes	25	17
(Alkyl-)indoles	18	6
Alkane-dienes	120	33
(Alkyl-)quinolines	8	4
Chlorinated compounds	10	7
Sulfur-containing compounds (except SO_2)	25	14

reveal whether peaks with the same compound name belong to one compound or to several compounds with similar spectra. This information could be obtained (i) from the probability factor provided by the library search and (ii) from so-called apex plots. *Probability* (range 0–9999) describes how unique a spectrum is compared with all other spectra in the library [14]. If the mass spectrum of a peak matches only one spectrum in the library, a high probability value is assigned to it during the NIST library search. *Apex plots* are calculated GC×GC two-dimensional plots in which only the apexes of peaks are indicated. They were generated in a spreadsheet programme using the first- and second-dimension retention times, which were calculated from the total retention time using the modulation time and modulation start time as parameters.

Table 3 shows a selection of peaks which occurred several times in the peak table and were identified with good match factors, but with either distinctly low or, alternatively, very high probability values. In the apex plot of Fig. 4 the positions of the peaks/compounds of Table 3 are examined. As an example, three peaks were identified with good match factors as 1-(2-pyridinyl)-ethanone, i.e. 2-vinylpyridine. The probability values (9654–9758) indicate that the mass spectrum is highly unique. This suggests that the three peaks are different modulations of a single compound. This could be confirmed by the apex plot: the three apexes elute in a very narrow first- and second-dimension retention time range and obviously belong to the same compound. In general, probability values above 9000 indicate that the mass spectra are highly unique and that identification based on mass spectra is justified. Of the 1500 peaks

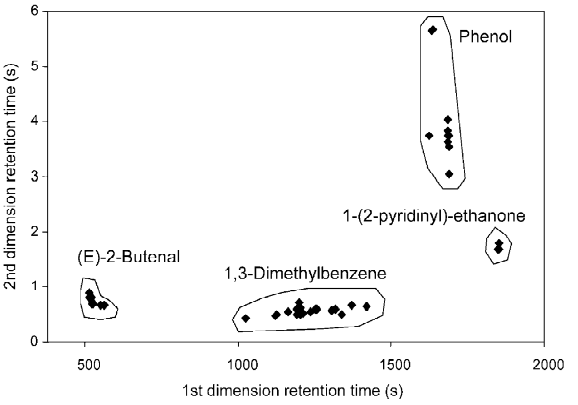

Fig. 4. Apex plot of the peaks described in Table 3.

in the peak table with high similarity and reverse values, 148 had such high probability factors.

As another example, 19 peaks were identified as 1,3-dimethylbenzene with very high similarity and reverse, but low probability, values (1526–3536). This indicates that several other library spectra are highly similar to the acquired mass spectrum—in this example, 1,2- and 1,4-dimethylbenzene, ethylbenzene and dimethylfulvene. In other words, a group-type identification is now all that can be provided. The apex plot revealed that the 19 peaks—which span a first-dimension elution time range of no less than 400 s—indeed belong to several different compounds.

The eight peaks identified as (*E*)-2-butenal have probability values lower than 1-(2-pyridinyl)-ethanone, but higher than the dimethylbenzenes, which suggests that they may belong to several compounds. However, the apex plot shows that all

Table 3
Selection of peaks with good match factors[a]

Compound	Number of peaks found	Similarity range	Reverse range	Probability range	^1t_R range (s)	No. of compounds
(*E*)-2-Butenal	8	868–936	910–941	3554–6540	518–566	1
1-(2-Pyridinyl)-ethanone	3	848–909	916–945	9654–9758	1849–1855	1
Phenol	10	874–938	912–944	3845–5457	1630–1695	1
1,3-Dimethyl-benzene	19	810–923	901–954	1526–3536	1027–1423	≥3

[a] Peaks with the same name are combined (see text).

peaks belong to a single GC×GC peak; due to tailing in both dimensions the peak was identified eight times. The probability values obtained for phenol are in the same range as for (*E*)-2-butenal. However, the apex plot has a completely different shape. This can be explained by fronting in both the first and second dimension; phenol is one of the major sample constituents and the high concentration causes overloading of the stationary phase: all apexes in Fig. 4 belong to the same compound. The appearance and handling of such intense peaks is explained in detail in Section 3.2.5.

For further improvement of the identification potential, the use of *I* values and second-dimension retention times is required, as will be discussed in Sections 3.2.6 and 3.2.7.

3.2.5. Manual processing of large peaks

The cigarette smoke sample contained a number of very intense peaks. Some of these show up vividly as intense vertical bands in the 2500–3200 s region of the insert of Fig. 1. Such peaks typically display peak deformation—i.e. fronting and/or tailing—in the second dimension and, in some cases, also in the first dimension. This is caused by overloading of the stationary phase in the first and/or second column, and is often observed in complex samples containing analytes with widely different concentrations. The peak finding algorithm, which only operates in the second dimension and assumes equal peak widths for all peaks, cannot satisfactorily handle such a situation (cf. Ref. [13]). Consequently, these intense peaks are identified several, or even many, times in a single second-dimension chromatogram—as already mentioned in Section 3.2.4.2. As an extreme example, nicotine was found 107 times, and 1,2-benzenediol no less than 149 times, in the peak table. The latter compound was modulated into 11 (adjacent) second-dimension chromatograms and was identified in each of these chromatograms some 15 times. It is obvious that the correct retention time of the real apexes of such large peaks cannot be read from the peak table. Instead, the apexes were determined manually, either in the contour plot or in the raw chromatogram.

Not unexpectedly, the intensity of the major peaks exceeded the dynamic range (more accurately, the range of the detector electronics) of the time-of-flight mass spectrometer. As a result, most of the higher signal intensities (in the extracted ion traces) were cut off—which is a common phenomenon with most GC detectors. As a consequence, the ratios of the masses in the (deconvoluted) mass spectra were in error—most masses in the mass spectrum were at 100% intensity. Therefore, the mass spectra of these compounds had to be obtained manually; good-quality spectra were then obtained on the lower slopes of the chromatographic peaks with their lower intensities.

The contour plot (cf. Fig. 1) was used to locate the major peaks and obtain their mass spectra as described above. Table 4 summarizes information on a selected number of prominent compounds. Identification was based on a mass spectral library search, and also on *I* values (see Section 3.2.6). Although all peaks were identified with good or acceptable similarity and reverse values, only two of the major peaks have high probability values (above 9000). The lower values for the majority of the peaks in Table 4 reflect the fact that most of the major constituents do not have ''unique'' mass spectra.

3.2.6. Linear retention indices

Because the combination of *I* values and GC–MS data can improve analyte identification considerably compared to GC–MS only [15], linear *I* values of the first-dimension separation were calculated for all compounds using the first-dimension retention times of the *n*-alkanes from *n*-pentane to *n*-pentadecane, which were present in the cigarette smoke. All peaks for which at least satisfactory similarity and reverse match factors had been found were also searched for via the Sadtler *I* index [16]. From among the approximately 2500 compounds searched, 238 were present in the Sadtler index. For 152 of these, the experimentally determined *I* values showed good agreement with the literature values; that is, the differences were less than 15 (this is a slightly greater difference than is sometimes used, but this seems justified in view of the overloading of the stationary phase). For a selected number of analytes from the group of 152, relevant data are shown in Table 5. Most of these compounds are known to be constituents of cigarette smoke [6,7,17]. One may add that some of the experimentally determined *I* values for the major compounds listed in Table 4 are

Table 4
Selection of major peaks in cigarette smoke[a]

Compound[b]	CAS	^1t_R (s)	^2t_R (s)	Similarity	Reverse	Probability	*I*	
							Measured	Lit.[c]
2-Furanmethanol	98-00-0	1128.6	3.09	895	895	4959	851	853
2-Cyclopentene-1,4-dione	930-60-9	1242.5	4.46	850	850	8170	880	
Phenol	108-95-2	1631.9	3.82	883	887	4060	985	984
3-Methyl-1,2-cyclo-pentane-dione	765-70-8	1829.6	5.17	914	914	5945	1038	
o-Cresol	95-48-7	1919.1	3.09	950	952	7918	1058	1055
p-Cresol	106-44-5	1973.9	2.21	919	920	5742	1072	1075
m-Cresol	108-39-4	2012.8	5.10	911	925	6182	1083	1078
2,4-Dimethylphenol	105-67-9	2332.9	5.90	936	938	3902	1160	
1-Methylindene	767-59-9	2350.8	0.74	926	928	1741	1163	
3-Ethylphenol	620-17-7	2416.8	1.46	867	868	3212	1180	1170
2,4-Dimethylphenol	105-67-9	2524.6	1.40	863	911	1880	1208	
Benzoic acid	65-85-0	2530.6	3.92	838	904	4749	1210	1172
1,2-Benzenediol (catechol)	120-80-9	2644.4	5.62	870	892	8833	1243	
2-Ethyl-4-methylphenol	3855-26-3	2752.3	5.57	854	859	4052	1275	
Indole	120-72-9	2902.0	1.59	873	873	5878	1321	
4-Hydroxy-2-methyl-acetophenone	875-59-2	2950.0	3.07	830	833	1149	1337	
2,6-Dimethoxyphenol	91-10-1	3081.8	3.10	922	922	9633	1376	
Nicotine	54-11-5	3117.7	1.33	840	848	9528	1380	1340
3-Methyl-1*H*-indole	83-34-1	3207.6	5.21	924	928	3035	1408	
Myosmine	532-12-7	3357.4	1.60	834	836	8687	1454	

[a] ^1t_R and ^2t_R, retention times in the first and second dimension, respectively.
[b] As identified by the GC–TOF-MS software; names according to NIST mass spectral library.
[c] Measured on a 25 m×0.31 mm I.D., 0.52 μm SE-54 column at 2 °C/min [16].

distinctively higher than the literature values; examples are benzoic acid, nicotine and isoeugenol. This can be easily explained by the peak distortion (fronting) in the first-dimension separation, which causes the peak apex to shift to a higher retention time.

In some cases, the elution order—and, consequently, the identity—of isomers which could not be derived from their (identical) mass spectra, was determined using their *I* values. To give an example, all cresols included in Table 4 were identified as 2-methylphenol (*o*-cresol) by the library search. By subsequently using their *I* values, the elution order could be determined.

Finally, it should be added that no *I* values are as yet available to describe the second-dimension separation. Actually, there is not even a generally accepted model on which the required calculation can be based (see also Conclusions). Even so, however, the second-dimension retention times provided valuable information regarding functional groups and substitution patterns of many classes of compounds.

3.2.7. Ordered/structured chromatogram

As was briefly indicated in the Introduction, the column combination used in this study provided two almost independent separations. On the first (non-polar) column, analytes were separated according to their vapour pressure, and on the second-dimension column, according to their activity coefficient for the Carbowax stationary phase. The Carbowax column shows high retention for compounds with high hydrogen-bond acidities. Consequently, compounds with similar vapour pressures will have similar retention times in the first dimension, and analytes with similar hydrogen-bond acidities will display similar second-dimension retention times. As a result, structurally related compounds will exhibit ‘‘elution order’’, i.e. they will elute within a cluster

Table 5
Selection from 152 provisionally identified analytes[a]

Name[b]	CAS	^1t_R (s)	^2t_R (s)	Similarity	Reverse	Probability	*I*	
							Measured	Lit.
1*H*-Pyrrole, 1-methyl-	96-54-8	745.2	0.83	865	867	7693	750	735
2(3*H*)-Furanone, dihydro-5-methyl-	108-29-2	1512.0	2.97	914	922	7024	950	948
2,5-Hexanedione	110-13-4	1398.2	2.20	884	886	9110	920	926
2-Furancarboxaldehyde, 5-methyl-	620-02-0	1554.0	2.23	889	900	9582	961	960
2*H*-1-Benzopyran-2-one	91-64-5	3279.4	3.22	775	846	5972	1432	1431
2-Propanol, 1-chloro-	127-00-4	619.4	2.74	911	946	9560	703	694
2-Propanone, 1-chloro-	78-95-5	577.4	1.63	926	937	9743	683	678
2-Pyrrolidinone	616-45-5	1985.3	4.88	926	926	9826	1076	1080
2-Vinylpyridine	100-69-6	1449.3	1.07	921	945	8405	932	929
3-Acetylpyridine	350-03-8	2147.3	2.72	891	902	6954	1117	1102
3-Butenenitrile	109-75-1	529.5	1.39	956	957	4211	658	653
5-Hepten-2-one, 6-methyl-	110-93-0	1643.8	0.67	897	900	8178	984	986
Benzaldehyde	100-52-7	1560.0	1.84	932	948	8065	962	955
Benzene	71-43-2	511.5	0.50	750	909	6271	648	654
Benzoic acid, methyl ester	93-58-3	2099.2	1.25	852	933	6456	1105	1092
Benzonitrile	100-47-0	1649.8	2.28	917	954	6441	986	980
Benzyl chloride	100-44-7	1787.6	1.18	893	941	5370	1023	1010
Benzyl nitrile	140-29-4	2266.9	3.30	925	933	4191	1148	1138
Butanal, 3-methyl-	590-86-3	529.5	0.40	806	815	7855	658	646
Butyrolactone	96-48-0	1350.3	4.13	955	966	6510	908	911
Crotonaldehyde	123-73-9	523.5	0.66	908	937	6045	657	642
Cyclohexanone	108-94-1	1296.4	0.84	861	943	4447	894	889
Ethanol, 2-(2-butoxyethoxy)-	112-34-5	2446.6	1.47	880	915	5588	1192	1191
Ethanol, 2-(2-ethoxyethoxy)-	111-90-0	1721.7	2.18	920	925	9550	1006	1008
Ethanol, 2-chloro-, acetate	542-58-5	1044.7	1.37	787	873	9274	832	834
Ethanone, 1-(2-furanyl)-	1192-62-7	1350.3	2.33	922	927	8759	908	909
Hexanenitrile	628-73-9	1206.5	1.01	899	911	9155	871	876
Isoquinoline	119-65-3	2692.3	2.04	858	930	6067	1255	1247
Methoxybenzene	100-66-3	1380.2	0.97	933	954	9355	915	914
m-Toluidine	108-44-1	2015.3	3.16	874	887	3570	1082	1074
N,*N*-Dimethylacetamide	127-19-5	1188.5	1.98	926	970	9587	867	863
Naphthalene	91-20-3	2374.7	1.26	720	911	3408	1174	1176
Octanenitrile	124-12-9	2021.3	0.87	866	885	5753	1084	1082
Pentanenitrile	110-59-8	841.1	1.02	812	823	8350	777	773
Pentanoic acid, 4-oxo-, methyl ester	624-45-3	1631.9	1.95	884	950	9667	981	989
Phenylethyne	536-74-3	1230.5	1.25	937	953	8918	877	871
Propanenitrile, 2-methyl-	78-82-0	475.6	0.58	911	912	9578	626	620
Pyrazine, methyl-	109-08-0	1014.8	1.12	919	921	8523	825	817

[a] In the case of multiple identification, only the peak with the best mass spectral match factors and *I* value match is shown.
[b] Name according to NIST library.

in the GC×GC plane or along ''strings'' (cf. Refs. [1,2,4,18]).

In one-dimensional GC–MS, extracted ion chromatograms showing only one or a few selected m/z values are occasionally used to reveal the presence of structurally related compounds. With the present complex sample, however, it was generally speaking not possible to find sufficiently selective masses to generate meaningful extracted ion chromatograms in the two-dimensional GC plane. Instead, the peak table was searched for selected compound classes (using compound names and formulae) and the

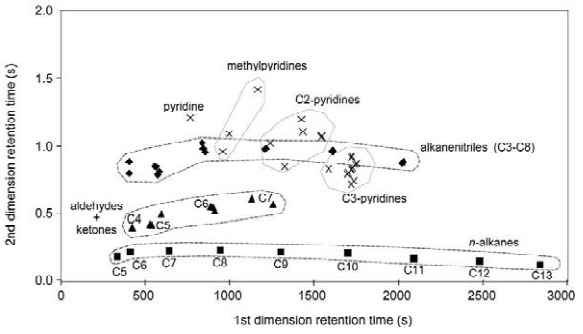

Fig. 5. GC×GC apex plot of some selected analyte classes. Bands or clusters formed by structurally related compounds are indicated by polygons.

retention times of identified peaks were used to generate apex plots. As an example, Fig. 5 shows that structurally related compounds such as *n*-alkanes, aldehydes/ketones and alkanenitriles elute in bands or clusters. For some classes, such as the alkyl-substituted pyridines, a further sub-division into methyl-, C_2- and C_3-substituted compounds was observed (where the term ''C_2-substitution'' includes dimethyl- and ethyl-substituted compounds). The value of such information to enhance the reliability of identification will be clear.

Another example regarding the practical usefulness of ordered chromatograms is shown in Fig. 6. The retention of alkyl-substituted pyrroles on the second-dimension column strongly depends on the substitution pattern. Pyrroles with a hydrogen in the 1 position (1*H*-alkyl-pyrroles, which are secondary amines) can form hydrogen bonds with the Carbowax stationary phase and, consequently, have high second-dimension retention times, whereas pyrroles with an alkyl substituent in the 1 position (1-alkyl-pyrroles, which are tertiary amines) have much shorter retention times. Since both types of substituted pyrroles have very similar mass spectra (as shown in Fig. 6B and C), the second-dimension retention times provide much desired additional information for identification. The peak of 1*H*-pyrrole had a second-dimension retention time of approximately 7 s and, consequently, did not elute during its own modulation cycle, i.e. showed wrap around. However, for a better visualization of the structure of the apex plot, the peak is displayed at its true second-dimension retention time in Fig. 6. The decreasing second-dimension retention times with increasing degree of alkyl substitution can be explained by the positive inductive effect of the alkyl groups, which reduces the acidity of the 1*H*-alkyl-pyrroles.

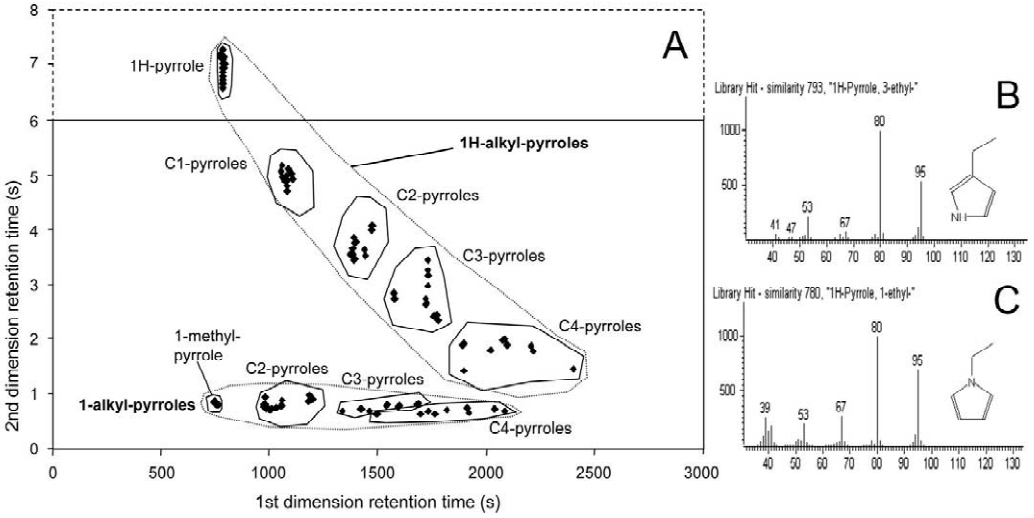

Fig. 6. (A) GC×GC apex plot of identified alkyl-substituted pyrroles. (B) and (C) show the mass spectra of 3-ethylpyrrole and 1-ethylpyrrole, respectively.

4. Conclusions

The high separation power of GC×GC–TOF-MS, combined with sophisticated peak finding and deconvolution algorithms, provides a system with an unparalleled separation and identification power. However, the huge amount of information obtained from such a separation requires new strategies for data processing and data reduction. With very complex chromatograms, such as obtained, for example, when analyzing cigarette smoke, it is not possible to perform the data processing manually—it has to be performed in an automated fashion.

In this paper, the potential of such automated processing of GC×GC–TOF-MS data, and its current limitations, are shown. The automated peak finding and deconvolution algorithm of the GC–TOF-MS software is the essential part of the data processing. The large number of peaks found in the present sample required an approach which is different from that used in conventional one-dimensional GC–MS. Since manual verification of the library search results cannot be carried out for all peaks in the table, it was necessary to rely largely on mass spectral match factors. Using this approach, more than 1500 peaks (some 520 compounds) were found in the sample with good library matches (cf. Table 6). However, for a large number of peaks (some 9000), the majority of which showed acceptable library matches, it was not possible to decide whether they were correctly identified or not based solely on these factors, and additional information was required. Linear *I* values for the first-dimension separation, group-type information derived from the second-dimension retention times and interpretation of the peak shapes in the GC×GC contour plot provided additional clues for peak identification. To quote an example, by using literature *I* values, some 660 peaks (152 compounds) from the group of 10 500 peaks with good and acceptable library matches could be identified.

While it is, in our opinion, true that the present results—which were obtained without injecting any reference compounds—clearly demonstrate the huge potential of GC×GC–TOF-MS in analyzing very complex samples, the same results also demonstrate that distinct improvements are needed to (i) significantly reduce the data processing time, and (ii) considerably increase the number of identified peaks. Four aspects may be mentioned here.

(1) Although it has been stated above that the software used in this study has several unique features, one would nevertheless like to stimulate the manufacturers to further improve the quality of the deconvolution algorithm, and of the library search strategies.

(2) As mentioned in Section 3.2.3, the generation of the peak table took about 7 h. There is little doubt that this process will become much more rapid in the near future when faster computer hardware becomes available. However, one should consider that the post-processing took much longer—and has to be measured in days rather than hours. This predominantly manual post-processing therefore constitutes a much more dramatic challenge. A significant time reduction will require an at least partial inclusion of this step in the automated processing—for example, the use of the first-dimension and, also, the second-dimension *I* values (see next item).

(3) Information complementary to that provided by the first-dimension *I* values can be obtained from second-dimension *I* values. As was briefly men-

Table 6
Classification of library search results for cigarette smoke

Description	Criteria			No. of peaks	No. of different compounds (names)
	Similarity	Reverse	Probability		
Good match, unique spectrum	≥800	≥900	9000–9999	148	74
Good match	≥800	≥900	0–9999	1500	520
Acceptable match, additional information/ inspection required	750–800	800–900	0–9999	9000	2400
Poor match	<750	<800	0–9999	20 000	6200

tioned in Section 3.2.6, collecting such information will, first of all, require designing an acceptable model. Subsequently, setting up a reference database will require the evaluation and, next, optimization of several experimental approaches. Currently, such work is in progress in several laboratories.

(4) The manual processing required to inspect peaks with less than unambiguous provisional identification should be improved by making integrated software available that combines all data processing strategies described in this paper.

Finally, even though it is true that solutions to problems such as those mentioned above are urgently required, one should at the same time realize that—already—GC×GC–TOF-MS provides a separation and identification power which is unequalled in the field of chromatography. One should also consider that, as yet, non-target analyses as discussed in this paper are never performed on a routine basis and, moreover, that non-target analyses of much less complex samples by means of conventional GC–MS are likewise time-consuming. In our opinion, it will certainly not be possible in the near future to perform such complex analyses in a fully automated fashion. Even when the additional tools mentioned above become available, some manual processing and the knowledge and expertise of an experienced analyst will still be required.

References

[1] P. Korytár, P.E.G. Leonards, J. de Boer, U.A.Th. Brinkman, J. Chromatogr. A 958 (2002) 203.
[2] J. Beens, J. Blomberg, P.J. Schoenmakers, J. High Resolut. Chromatogr. 23 (2000) 182.
[3] Ph.J. Marriott, R. Shellie, J. Fergeus, R. Ong, P. Morrison, Flav. Fragr. J. 15 (2000) 225.
[4] P.J. Schoenmakers, J.L.M.M. Oomen, J. Blomberg, W. Genuit, G. van Velzen, J. Chromatogr. A 892 (2000) 29.
[5] R.A. Jenkins, M.R. Guerin, B.A. Tomkins (Eds.), The Chemistry of Environmental Tobacco Smoke: Composition and Measurement, CRC Press, New York, 2000.
[6] C.R. Green, A. Rodgman, Recent Adv. Tobacco Sci. 22 (1996) 131.
[7] H. Wakeham, in: I. Schmeltz (Ed.), Proceedings of the Symposium on the Chemical Composition of Tobacco and Tobacco Smoke, 162nd National Meeting, American Chemical Society, Washington, DC, September 1971, Plenum Press, New York, 1972, p. 1.
[8] B.M. Gordon, M.S. Uhrig, M.F. Borgerding, H.L. Chung, W.M. Coleman, J.F. Elder, J.A. Giles, D.S. Moore, C.E. Rix, E.L. White, J. Chromatogr. Sci. 26 (1988) 174.
[9] K. Rustermeier, R. Stabbert, H.-J. Haussmann, E. Roemer, E.L. Carmines, Food Chem. Technol. 40 (2002) 93.
[10] H.-J. de Geus, A. Schelvis, J. de Boer, U.A.Th. Brinkman, J. High Resolut. Chromatogr. 23 (2000) 189.
[11] E.B. Ledford, J.B. Philips. US Pat. 6 007 602 (1999).
[12] X. Xu, J. Williams, L.L.P. van Stee, J. Beens, M. Adahchour, R.J.J. Vreuls, J. Lelieveld, Atmos. Chem. Phys. (manuscript in preparation).
[13] J. Dallüge, R.J.J. Vreuls, J. Beens, U.A.Th. Brinkman, J. Sep. Sci. 25 (2002) 201.
[14] S.E. Stein, J. Am. Soc. Mass Spectrom. 5 (1994) 316.
[15] W.P. Eckel, Am. Lab. March (2000) 17.
[16] The Sadtler Standard Gas Chromatography Retention Index Library, Philadelphia, Bio-Rad Laboratories, Sadtler Division, Philadelphia, PA, 1986.
[17] J. Dong, J.N. Glass, S.C. Moldoveanu, J. Microcol. Sep. 12 (2000) 142.
[18] H.-J. de Geus, I. Aidos, J. de Boer, J.B. Luten, U.A.Th. Brinkman, J. Chromatogr. A 910 (2001) 95.