ELSEVIER

Journal of Chromatography A, 962 (2002) 161–173

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Chemometric detection of thermally degraded samples in the analysis of drugs of abuse with gas chromatography–Fourier-transform infrared spectroscopy

M. Praisler[b], J. Van Bocxlaer[c], A. De Leenheer[a,*], D.L. Massart[d]

[a]*Laboratory of Toxicology, Gent University, Harelbekestraat 72, B-9000 Gent, Belgium*
[b]*Department of Physics, University of Galati, Domneasca St. 47, 6200 Galati, Romania*
[c]*Laboratory for Medical Biochemistry and Clinical Analysis, Gent University, Harelbekestraat 72, B-9000 Gent, Belgium*
[d]*Laboratory of Pharmaceutical and Biomedical Analysis, Free University of Brussels, Laarbeeklaan 103, B-1090 Brussels, Belgium*

Received 28 May 2001; received in revised form 17 April 2002; accepted 25 April 2002

Abstract

We present a chemometric procedure for the identification of the reference standard chromatographic peak in cases where the GC–FTIR analysis of commercial standards results in the appearance of more than one peak in the GC chromatogram. The procedure has been designed for phenethylamines, which represent the class with the largest number of individual molecules on the illicit drug market, and which are abused for their stimulant and/or hallucinogenic effects. The similarity between their vapor-phase FTIR spectra was modeled using principal component analysis (PCA), and class identity was assigned on the basis of soft independent modeling of class analogy (SIMCA). Additional peaks could be assigned to impurities in the standards, but most often they were artifacts formed during the GC–FTIR analysis of thermolabile or chemically unstable compounds. The latter case is illustrated by the identification of the reference standard chromatographic peak and FTIR spectrum of the potent psychotropic amphetamine derivative *N*-methyl-1-(3,4-methylenedioxyphenyl)-2-butanamine (MBDB), and by the elucidation of the chemical changes that occur in the molecule of MBDB due to thermal degradation.

Keywords: Thermal degradation; Principal component analysis; Fourier-transform infrared spectroscopy; Soft independent modeling of class analogy; Drugs of abuse; *N*-methyl-1-(3,4-methylenedioxyphenyl)-2-butanamine

1. Introduction

As many drugs of abuse are reasonably volatile substances, gas chromatography–Fourier-transform infrared spectroscopy (GC–FTIR) has become a powerful technique for their identification in tablets or powders of judicial origin. IR spectra are well suited to validate structural candidates. The IR spectrum not only depends on the particular functional groups, it also reflects the arrangement of these functional groups within the molecule. The IR spectrum is, thus, in contrast to NMR or MS spectra, predominantly a property of the whole molecule and not just the sum of the properties of its constituents.

Since GC–FTIR spectra became available in digitized form, much effort has been done to find a

*Corresponding author. Tel.: +32-9-264-8121; fax: +32-9-264-8197.
E-mail address: andre.deleenheer@rug.ac.be (A. De Leenheer).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)00536-8

faster way of solving chemical structure problems by the wide use of databases containing information on the structures and properties of compounds [1–6]. Computer methods of extracting the structural information from the spectral data were intensively developed hand in hand with the improvement of the spectral equipment. If the structure of a new compound has to be established or the spectrum of the unknown compound are absent from the database, the analyst unavoidably faces a problem of restoring the molecular structure from the spectral information given by fragments of its molecule. The solution to this problem becomes easier when an expert system is used. Expert systems are known to simulate a chemist's reasoning during molecular structure elucidation on the basis of spectrum–structure correlations and (in the more recent developments) of quantitative spectrum–activity relationships (QSAR). A number of expert systems for the spectral analysis are described in the literature, many of which are dedicated to the processing of IR data [7–9] and MS data [10–13].

In our previous papers, we have presented an expert system representing an automated tool screening for novel illicit amphetamines (Fig. 1) found in tablets and powders [14]. The amphetamines represent the class with the largest number of individual molecules on the illicit drug market, and which are abused for their stimulant and/or hallucinogenic effects [15]. The main objective was to obtain a computerized procedure performing a general unknown analysis and predict with high accuracy the class identity of the compound when the reference spectrum is not present in the spectral library. First, we have examined the possibility of using an in-house-made infrared spectral database for the recognition of the amphetamines according to their biological activity (stimulant and/or hallucinogenic). During the exploratory analysis, based on principal component analysis (PCA) [16], emphasis was placed on the selection of the most adequate sample preparation and data pre-processing techniques yielding the best differentiation and recognition of the structural and of the substitution patterns (determining the associated activity) of amphetamines.

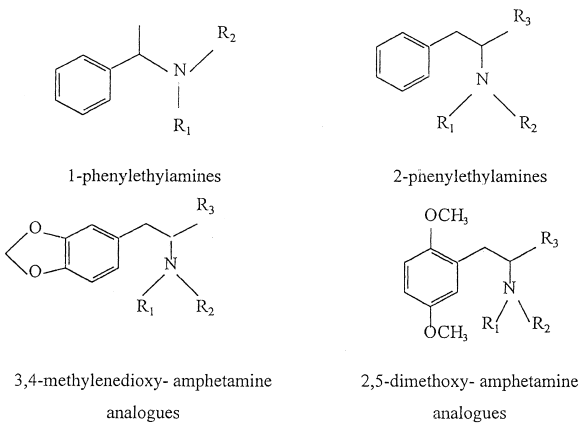

Fig. 1. Molecular structures of the modeled amphetamine analogues.

Once the most adequate form of the knowledge base was determined, we focused on the development of the expert system itself [17]. A classification procedure assigning the class identity was implemented, using soft independent modeling of class analogy (SIMCA) [16] and properties of analytical merit were further assessed [18]. The appropriate sensitivity and selectivity were found according to whether the requirements for the investigation of illicit drugs was for epidemiological, clinical, administrative and forensic purposes [19,20].

Practice has shown that the key issue in obtaining good results with computer-assisted analytical tools, such as expert systems, is to have a knowledge base formed with good quality data. For data obtained with hyphenated techniques, the optimization of the separation process is as important as recording the spectral data with accuracy. In order to optimize the entire analytical process, chemometrics has been applied in many successful attempts to improve chromatogram quality [21,22], using a variety of criteria related to the performance goals of the chromatographic analysis (e.g. resolution between adjacent peaks, peak asymmetry, retention time, duration of the analysis) [23–25]. One of the recent trends is to adapt computer-assisted programs, initially designed for compound identification (library search based on the chemometrical quantification of structural similarities), to the assessment of chromatographic peak quality [26].

During the process of building the GC–FTIR database representing the knowledge base of the expert system, an analytical problem that was encountered was that some samples degrade during the

chromatographic analysis (e.g. in the injection liner at high temperatures). Generally, peaks additional to the reference standard chromatographic peak could be assigned to impurities in the standards, but most often they were artifacts formed during the GC–FTIR analysis of thermolabile or chemically unstable compounds. The aim of the work presented in this paper is to further develop the expert system into an integrated analytical tool that would not only perform the identification of the class identity of the compounds of interest but also the identification of the reference standard chromatographic peak in cases where the GC–FTIR analysis of commercial standards results in the appearance of more than one peak in the Gram–Schmidt (GC) chromatogram. The efficiency of the system is illustrated by the identification of the reference standard chromatographic peak and FTIR spectrum of the potent psychotropic amphetamine derivative *N*-methyl-1-(3,4-methylenedioxyphenyl)-2-butanamine (MBDB, $R_1=-CH_3$, $R_3=-CH_2-CH_3$), and by the elucidation of the chemical changes that occur in the molecule of MBDB due to thermal degradation. To provide a more complete picture of the discriminating power and of the benefits of using the expert system for the detection of the correct chromatographic peak through the screening of the corresponding FTIR spectrum, a critical evaluation comparing its results with the techniques (i.e. automated spectral library search and structure elucidation) ''traditionally'' performed by the commercially available programs is presented.

2. Experimental

A Perkin-Elmer (Buckinghamshire, UK) Autosystem GC was interfaced with a light pipe GC–IR System 2000 and connected to a FTIR System 2000 with a mid-infrared source and a medium band liquid nitrogen-cooled mercury cadmium telluride (MCT) detector. Temperature-programmed separations were carried out on a Hewlett-Packard (Palo Alto, CA, USA) Ultra-1 methylsilicone capillary column (25 m×0.32 mm I.D., 0.52-μm film thickness). The carrier gas was helium at a flow-rate of 1.8 ml/min. The analytical column outlet stretched into the light pipe inlet. Helium was added as make-up gas at the connection between the capillary column and the light pipe. The gold-coated light pipe (12 cm×1 mm I.D.) was heated at a constant temperature of 270 °C. Real time spectra were obtained by addition of two scans, with a spectral resolution of 8 cm^{-1} and 32 background scans. The scan range was from 4000 to 580 cm^{-1}. Chromatograms were calculated by the Gram–Schmidt vector orthogonalization method.

Methanolic stock solutions (1 mg/ml) of the reference standards were injected into the GC–FTIR system. Gram–Schmidt reconstruction was performed using 10 basic vectors throughout the run. Baseline correction was performed on the reconstructed Gram–Schmidt chromatogram (GS) and low-noise vapor-phase FTIR spectra were generated after co-addition. The obtained reference vapor-phase FTIR spectra were stored in a computer-based library after normalization (i.e. the peak absorbance of the most intense band was set to unity). The spectral data were stored in the database at 5-cm^{-1} intervals.

In order to enhance FTIR sensitivity, correct the chromatographic peak shape, and improve, sometimes, compound thermostability, a derivatization procedure was included in the sample preparation phase. The derivatization agents were heptafluorobutyric anhydride (HFBA) and to a lesser extent pentafluoropropionyl anhydride (PFPA). The 159 compounds included in the FTIR database are forensic substances such as drugs of abuse (mainly central stimulants, hallucinogens, sympathomimetic amines, narcotics and other potent analgesics), precursors and derivatized counterparts. The samples represent reference standards, as well as compounds synthesized at the laboratory. This in-house-made database is constantly up-dated with spectra of new designer drugs, in order to increase the variety of illegal substances that can be positively identified in judicial exhibits. The same GC–FTIR based analytical procedure was used for the analysis of suspect powders and tablets from judicial origin.

To form its knowledge base of the expert system, all spectra in the library were reduced in size by eliminating the spectral windows where the compounds in the database have no FTIR absorption. Hence, data effectively ranged from 3750 to 2550 cm^{-1}, and from 2000 to 600 cm^{-1}. A PCA model was calculated for hallucinogenic amphetamines

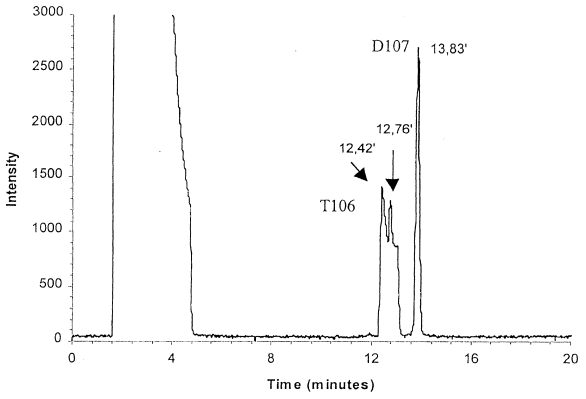

Fig. 2. The Gram–Schmidt reconstructed chromatogram obtained for the reference standard of MBDB.

(class code T), stimulant amphetamines (class code M), their HFB derivatives (class code DT and DM, respectively), and for counter-examples (class code N). Data were mean-centered and scaled on the basis of the variables' standard deviation. The validation method was full cross-validation. A total number of three principal components (PCs) were used to model each class. Then SIMCA classification was run, with a 5% confidence level, for all spectra in the database. The quality of the classification was evaluated on the basis of the classification matrix, and the total correct classification rate. The system was optimized by an iterative process enhancing the model true positive rates and maximizing the total correct classification rate. PCA and SIMCA were performed using the software package The Unscrambler® (Camo, Trondheim, Norway).

The reference standard of MBDB was provided by RBI (Natick, MA, USA). In the Gram–Schmidt reconstructed chromatogram more than one peak was obtained (Fig. 2). The FTIR spectrum was recorded for each peak (retention time 12.42, 12.76, and 13.83 min). The same FTIR spectrum (code T106) was obtained for the 12.42- and 12.76-min chromatographic features, showing that in fact they are generated by the same compound and belong to a single chromatographic peak. The spectrum (code D107) recorded for the highest peak in the chromatogram (retention time 13.83 min) was different (Fig. 3) from the spectrum of T106. In order to identify the reference standard chromatographic peak, the molecular structure of the compounds yielding the two chromatographic peaks was analyzed by classifying their FTIR spectra with the help of the expert system.

3. Results and discussion

3.1. The expert system

PCA modeling enables the detection of sample patterns in infrared spectra, like any particular grouping, and discriminates the useful information from noise or meaningless variation contained in the data.

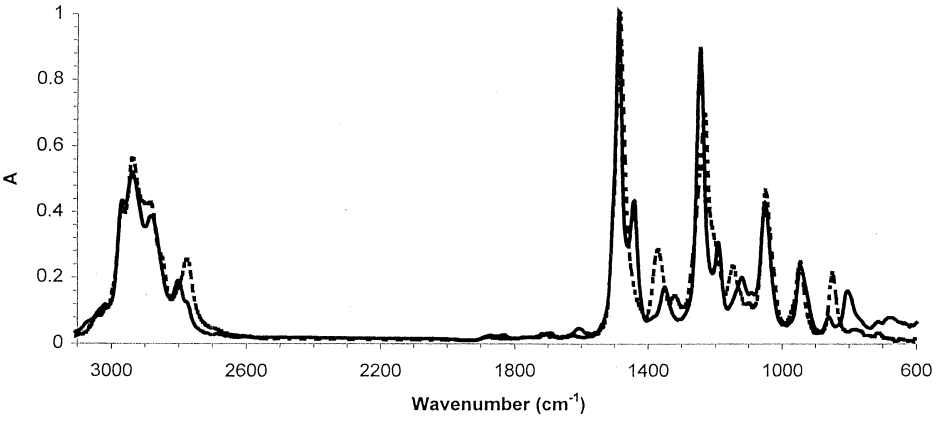

Fig. 3. The FTIR spectra obtained for the first (———, spectrum code T106) and for the second chromatographic peak (···, spectrum code D107).

In matrix representation, a model **K** with a given number of PCs has the following equation:

$$\mathbf{X_K} = \bar{\mathbf{X}}_\mathbf{K} + \mathbf{P_K}(n,r)\,\mathbf{V_K^T}(r,p) + \mathbf{E_K}(n,p) \quad (1)$$

where $\mathbf{X_K}$ is the data matrix containing the training set of model **K**, $\bar{\mathbf{X}}_\mathbf{K}$ the mean centered data matrix, $\mathbf{P_K}$ the scores matrix obtained for n samples and r selected PCs, $\mathbf{V_K^T}$ the transpose of the loading matrix obtained for r selected PCs and p variables, and $\mathbf{E_K}$ the error (residual) matrix. The linear combination is a better descriptor for the pattern in the data than the original measurements because it discriminates the useful information from noise or irrelevant absorption variation contained by the original data.

Once PCA models are built for several classes (training sets), they allow the initiation of classification-specialized algorithms, such as SIMCA, which performs disjoint class modeling by supervised pattern recognition. In comparison with other multivariate techniques, SIMCA has the advantage that it can decide whether unknown samples belong to more than two modeled classes and, thus, may best fit the functions of a screening test. Each class is modeled separately based on the similarity of the samples within the class. The class boundaries (confidence limits) are constructed around each model by taking into account the distribution of the Euclidean distances between the sample and the origin in the space of the residual PCs (i.e. based on the error matrix $\mathbf{E_K}$ of the class **K**). A critical sample-to-model distance S_{crit} is computed with the help of an F-test, at a selected level of confidence, on the basis of the mean distance S_0 between the samples belonging to the class and the class model:

$$S_0 = \sqrt{\sum_{j=1}^{n} \cdot \sum_{i=1}^{p} e_{ji}^2 \Big/ [(p-r)(n-r-1)]} \quad (2)$$

$$S_{crit} = \sqrt{F_{crit} \cdot S_0^2} \quad (3)$$

where e_{ji}^2 is the squared residual of the jth sample for the ith (latent) variable.

For the identification of a new sample, its spectrum is projected into the PC space defined by the PCA model and its distance $S_{new,\,\mathbf{K}}$ towards the class model **K** is compared to S_{crit}:

$$\tilde{\mathbf{x}}_{new}(1,p) = \bar{\mathbf{x}}_\mathbf{K} + (\mathbf{x}_{new} - \bar{\mathbf{x}}_\mathbf{K})\,\mathbf{V_K}\mathbf{V_K^T} \quad (4)$$

$$\mathbf{e}_{new} = \mathbf{x}_{new} - \tilde{\mathbf{x}}_{new} \quad (5)$$

$$S_{new,\,\mathbf{K}} = \sqrt{\sum_{i=1}^{p} e_{new,i}^2/(p-r)} \quad (6)$$

If $S_{new} < S_{crit}$, the new sample is classified as a member of class **K**, as the sample is well described by the class model. The influence of a sample on the model is evaluated by the leverage $L_{new,\,\mathbf{K}}$ (distance from the center of the model to the projection of the sample point on the model).

The expert system was optimized using SIMCA as a trail and error test applied to the whole FTIR database. The feedback criteria were the correct classification rates and the nature of the false positives and false negatives. An optimum was reached for **M** and **T** models built with the spectra of the compounds listed in Tables 1 and 2. The training set of the corresponding **N** model contained the spectra of bemegride, β-butyrolactone, cadaverine and its HFB derivative, codeine and its PFP derivative, caffeine, γ-butyrolactone, the trimethylsilyl (TMS) derivative of γ-hydroxy butyric acid, the TMS derivative of γ-hydroxy valeric acid, γ-valerolactone, nicotamide, piracetam, putrescine, dextromoramide, nicotine, prolintane, and safrole. The training set of **DM** contained the vibrational spectra of the HFB derivatives of amphetamine, methamphetamine, *N*-ethylamphetamine, *N*-*n*-propylamphetamine, α-phenethylamine, β-phenethylamine, and *N*-methyl-α-phenethylamine. The training set of the **DT** model included the spectra of the HFB derivatives of the following hallucinogenic amphetamines: 3,4-methylenedioxyamphetamine, 3,4-methylenedioxy-*N*-ethylamphetamine, 3,4-methylenedioxymethamphetamine, 1-(3,4-methylenedioxyphenyl)-2-butanamine, and *N*-methyl-1-(3,4-methylenedioxyphenyl)-2-butanamine. The optimized system classified 81.13% of the 159 tested compounds with a confidence level of 5%, with a total correct classification rate of 93.93%. A 96.30% true positive rate was obtained for **M** and **T** amphetamines, and 85.71% **DM** and **DT** true positive HFB derivatives.

The phenyl disubstituted amphetamines classified as such show that the structural prototype for the **T** model is a structural skeleton containing a disubstituted phenyl ring linked by an aliphatic side chain with two carbon atoms to an amino group (Table 2).

Table 1
Class **M** identity assignment for (phenyl non-substituted) stimulant amphetamines with their spectral library entry code (ID), acronym, status as controlled compound, substituents (R_1, R_2 and R_3), and classification scores

ID	Name of the drug	Acronym	CSA schedule	Control in Belgium	Found in illicit samples	R_1	R_2	R_3	Sample-to-**M** model distance	Leverage $L_{i,\,\mathbf{M}}$
2-Phenethylamines										
M96	*N*-Ethylamphetamine	EAMP	I	Yes	Yes	$-CH_2-CH_3$	H	$-CH_3$	0.596	0.169
M7	Amphetamine	AMP	II	Yes	Yes	H	H	$-CH_3$	0.365	0.197
M32	Dextroamphetamine		II	Yes	Yes	H	H	$-CH_3$	0.837	0.329
M74	Methamphetamine	MAMP	II	Yes	Yes	$-CH_3$	H	$-CH_3$	0.666	0.182
M12	Benzphetamine		III	No	No	CH_3	$-CH_2-C_6H_5$	$-CH_3$	0.573	0.336
M45	Fenproporex		IV	Yes	No	$-(CH_2)_2-CN$	H	$-CH_3$	0.830	0.184
M46	Phentermine		IV	No	Yes	H	H	$-(CH_3)_2$	0.481	0.263
M159	Clobenzorex		No	Yes	No	$-CH_2-(C_6H_4Cl)$	H	$-CH_3$	0.640	0.127
M71	Mephentermine		No	No	No	$-CH_3$	H	$-(CH_3)_2$	0.403	0.279
M114	*N*-*n*-Propylamphetamine	PAMP	No	No	No	$-(CH_2)_2-CH_3$	H	$-CH_3$	0.495	0.225
M47	1-Phenyl-2-butanamine		No	No	No	H	H	$-CH_2-CH_3$	0.606	0.145
M17	β-Phenylethylamine	BPEA	No	No	No	H	H	H	0.655	0.313
M112	*N*,*N*-Dimethyl-1-phenyl-2-Ethanamine	MMBPEA	No	No	No	$-CH_3$	$-CH_3$	H	0.790	0.253
1-Phenethylamines										
M4	α-Phenylethylamine	APEA	No	No	Yes	H	H		0.517	0.329
M102	*N*-Methyl-α-phenylethylamine	MAPEA	No	No	Yes	$-CH_3$	H		0.490	0.087

The allowed structural variations are the nature and the position of the substituents of the phenyl ring (i.e. 3,4-methylenedioxy analogues, or 2,5-dimethoxy analogues), and the presence of small aliphatic substituents on the second carbon atom of the side chain, or on the amino group, if any. While all tested phenyl disubstituted amphetamines have been classified as such (no false negative), the deaminated precursor safrole (4-allyl-1,2-methylenedioxybenzene, code D134) has been identified as a false **T** positive. In our trials, hallucinogens having one or three substituents on the phenyl ring were all

Table 2
Class **T** identity assignment for (phenyl disubstituted) hallucinogenic amphetamines with their spectral library entry code (ID), acronym, status as controlled compound, substituents (R_1, R_2 and R_3) and classification scores

ID	Name of the drug	Acronym	CSA schedule	Control in Belgium	Found in illicit samples	R_1	R_2	R_3	Sample-to-**T** model distance	Leverage $L_{i,\,\mathbf{T}}$
3,4-Methylenedioxy analogues										
T79	3,4-Methylenedioxyamphetamine	MDA	I	Yes	Yes	H	H	$-CH_3$	0.551	0.298
T87	3,4-Methylenedioxymethamphetamine	MDMA	I	Yes	Yes	$-CH_3$	H	$-CH_3$	0.445	0.243
T82	3,4-Methylenedioxy-*N*-ethylamphetamine	MDEA	I	Yes	Yes	$-CH_2-CH_3$	H	$-CH_3$	0.312	0.479
T106	*N*-Methyl-1-(3,4-methylenedioxyphenyl)-2-butanamine	MBDB	No	No	Yes	$-CH_3$	H	$-CH_2-CH_3$	0.571	0.357
T113	*N*,*N*-Dimethyl-3,4-methylenedioxyamphetamine	MDMMA	No	No	Yes	$-CH_3$	$-CH_3$	$-CH_3$	0.611	0.260
T85	1-(3,4-Methylenedioxyphenyl)-2-butanamine	BDB	No	No	No	H	H	$-CH_2-CH_3$	0.537	0.267
2,5-Dimethoxy analogues										
T145	2,5-Dimethoxyamphetamine	DMA	I	Yes	Yes	H	H	$-CH_3$	0.295	0.504
T37	2,5-Dimethoxyphenylethylamine		No	No	No	H	H	H	0.304	0.475

classified as negatives. A very important aspect for their recognition was that they were the *only* negatives having the **T** model as second closest model.

3.2. Analysis of MBDB

The system classified both T106 and D107 as phenyl disubstituted amphetamine analogues (Fig. 4). The critical value of the sample-to-model distance $S_{\text{crit, T}}$ for the **T** model was 1.036, corresponding to a leverage $L_{\text{crit, T}}$ of 0.800. Interestingly, although the corresponding chromatographic peak did not have an adequately shaped form, the sample-to-**T** model distance and the leverage of the compound T106 ($S_{\text{T106, T}} = 0.571$, $L_{\text{T106, T}} = 0.357$) was very similar to those of 3,4-methylenedioxyamphetamine (T79) or 3,4-methylenedioxymethamphetamine (T87) (Table 2), indicating that T106 is a phenyl disubstituted amphetamine analogue of the 3,4-methylenedioxy-type. The sample-to-distance value obtained for the compound D107 was similar to those of the 3,4-methylenedioxy analogues ($S_{\text{D107, T}} = 0.532$) as well. On the other hand, the value of its leverage was nearly twice larger ($L_{\text{D107, T}} = 0.614$) than the leverage of T106 or of the other 3,4-methylenedioxy analogues. The SIMCA scores of D107 were also much larger than those of the 2,5-dimethoxy analogues. Although the identity of compounds with a high leverage is more difficult to predict accurately than are the more ''central'' samples, the SIMCA scores obtained for D107 suggest that this compound is a phenyl disubstituted amphetamine analogue that has substituents of a *different* nature than the 3,4-methylenedioxy or 2,5-dimethoxy analogues. Thus, the expert system indicates that the reference standard chromatographic peak of MBDB is the first peak in the chromatogram.

The results of this class identity assignment were confirmed by visual inspection of the FTIR spectra of the two compounds. Indeed, the spectrum of T106 matches perfectly the mean spectrum of the modeled 3,4-methylenedioxy analogues (Fig. 5a), while the spectrum of D107 displays several absorption bands that are not specific to the spectra of the 3,4-methylenedioxy analogues. The comparison made in the spectral region associated with the C–H stretching vibrations (Fig. 5b) revealed that the absorptions of D107 in this region are stronger than those of T106, and than the mean value encountered for the rest of the 3,4-methylenedioxy analogues. This suggests that the number of alkyl groups encountered in the molecule of D107 might be larger than in the case of the 3,4-methylenedioxy analogues. Moreover, D106 has an atypical absorption at 2775 cm^{-1},

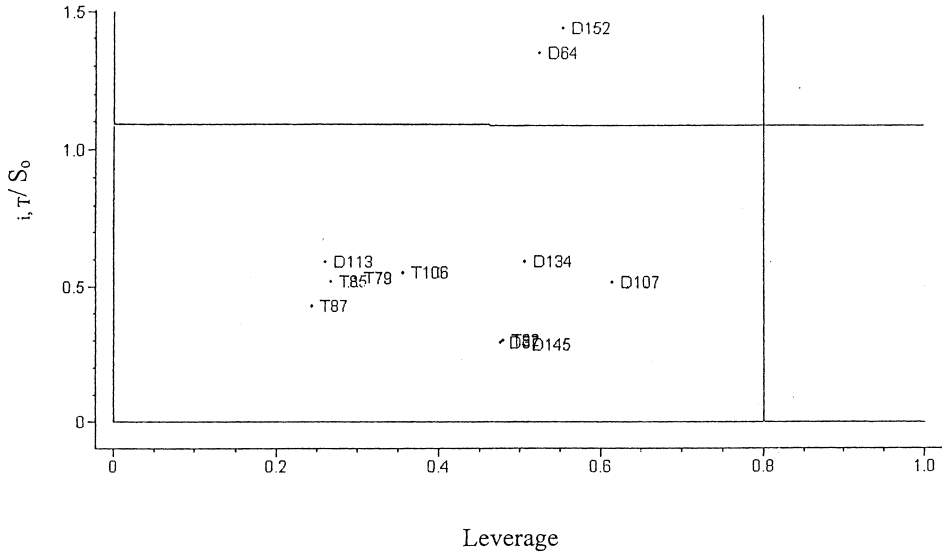

Fig. 4. Assignment of the **T** class identity for both T106 and D107 with the help of the expert system.

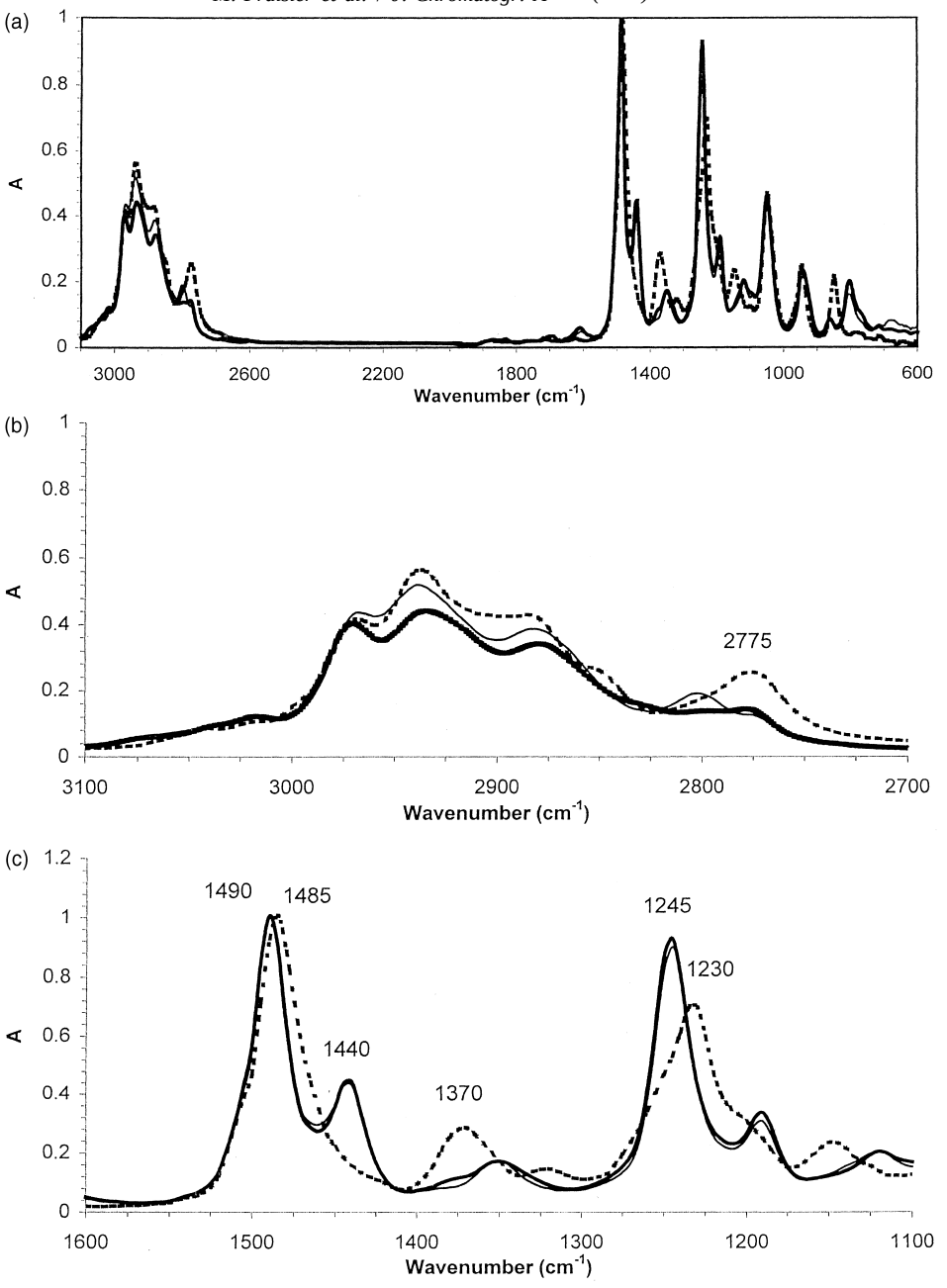

Fig. 5. (a) Comparison of the FTIR spectra of T106 (———), D107 (···) and of the mean spectrum of the modeled 3,4-methylenedioxy analogues (▬▬▬). (b) Comparison of the FTIR spectra of T106 (———), D107 (···) and of the mean spectrum of the modeled 3,4-methylenedioxy analogues (▬▬▬), in the spectral region of the absorptions associated with the C–H stretching vibrations. (c) Comparison of the FTIR spectra of T106 (———), D107 (···) and of the mean spectrum of the modeled 3,4-methylenedioxy analogues (▬▬▬), in the spectral region of the absorptions associated with the breathing vibrations of the phenyl ring (1490 cm^{-1}) and with the C–O stretching vibrations (1245 cm^{-1}). (d) The band-parameter stability of the original FTIR spectra of the 3,4-methylenedioxy analogues forming the training set of the **T** model, in the spectral region of the breathing vibrations of the phenyl ring and of the C–O stretching vibrations. (e) Comparison of the FTIR spectra of T106 (———), D107 (···) and of the mean spectrum of the modeled 2,5-dimethoxy analogues (▬▬▬), in the spectral region of the absorptions associated with the breathing vibrations of the phenyl ring and with the C–O stretching vibrations.

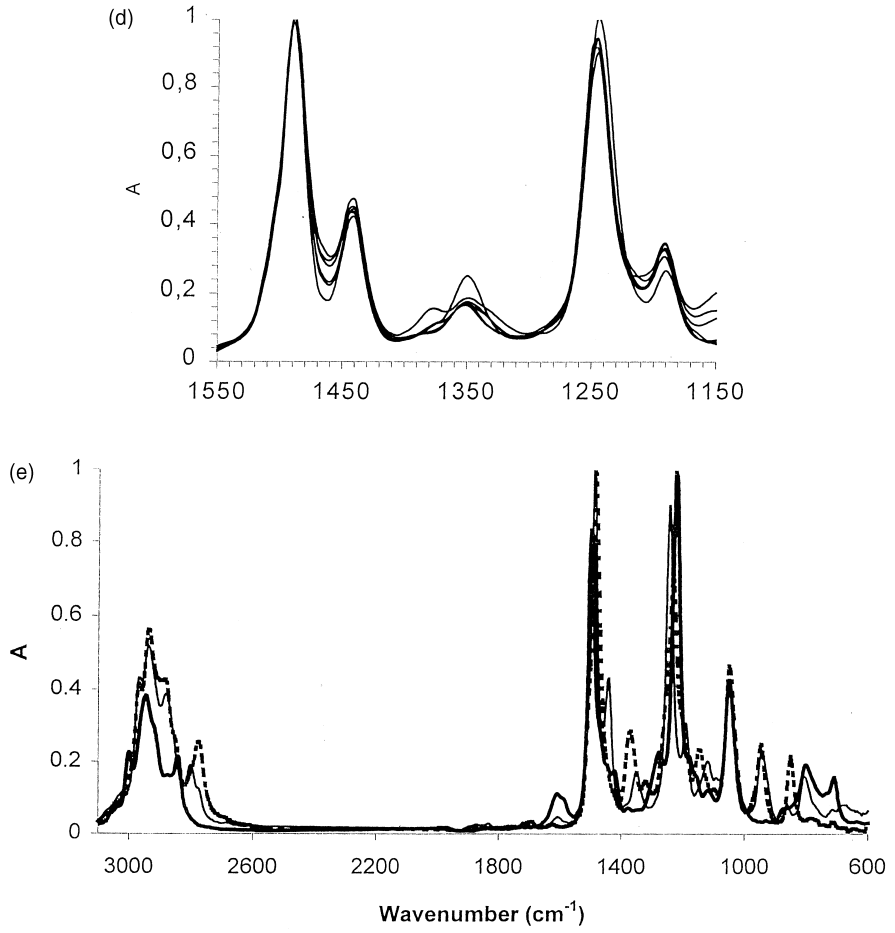

Fig. 5. (*continued*)

suggesting that D107 contains C–H bonds belonging to a chemical group that is not encountered in the molecular structure of the 3,4-methylenedioxy analogues. This absorption may be the cause of the high leverage of D107, as the variation of the number of the alkyl groups present in the side-chain of the true positive 3,4-methylenedioxy analogues is reflected in a leverage variation (around 0.3) of less than 0.1 (Fig. 4 and Table 2).

The absorptions are associated with the breathing vibrations of the phenyl ring (~1490 cm^{-1}) and with the stretching vibrations of the C–O bond (~1245 cm^{-1}) brought the most reliable confirmation that T106 is the spectrum of MBDB, as a 3,4-methylenedioxy analogue (Fig. 5c). The 1490 cm^{-1} absorption band is the strongest in the spectra of the hallucinogenic analogues and all the bands in the spectral region between 1550 and 1150 cm^{-1} are characterized by a remarkable band-parameter stability (Fig. 5d). In this region, the spectrum of T106 is practically identical with the mean spectrum of the modeled 3,4-methylenedioxy analogues. On the other hand, the band associated with the breathing vibrations of the phenyl ring is shifted towards lower wave numbers (1485 cm^{-1}) in the spectrum of D107, consistent with a phenyl ring affected by a stronger electron delocalization than in the case of 3,4-methylenedioxy analogues. Oxygen substitution

has a major impact on the intensity and position of the latter absorption, associated with the ring C=C bonds, because of the high force constant of C–O single bonds. Delocalization of the π-electrons is expected on the attachment of any oxygen atom, resulting in a band shift towards lower wave numbers.

Another spectral feature distinguishing the 3,4-methylenedioxy analogues is the fact that its strongest absorption occurs at 1490 cm^{-1}, as opposed to the stimulants, or to the hallucinogens having free (non-cyclic) phenyl substituents (Fig. 5e). In the latter case, the strongest absorption is that around 1245 cm^{-1}, associated with the stretching vibrations of the C–O link. Whereas the band around 1490 cm^{-1} ensures discrimination between stimulants and hallucinogens, the C–O band is the most sensitive in discriminating among hallucinogens due to the conjugation effect in rings and the bond-angle effect. The change in C–O bond character (variations in electron distribution) is reflected in an altered force constant and produces a shift of the C–O band towards higher wave numbers, thus in the opposite direction than the shift of the bands around 1490 cm^{-1}. The C–O band has only medium intensity in the case of both T106 and D107, consistent with the presence of an exocycle attached to the phenyl ring. The position of the C–O band of T106 confirms the 3,4-methylenedioxy substitution, while the same band appears in the spectrum of D107 shifted towards lower wave number, between the position of the C–O band of 3,4-methylenedioxy analogues and 2,5-methylenedioxy analogues. Thus, D107 seems to have an exocycle attached to the phenyl ring, of a nature that produces an electron delocalization stronger than the 3,4-methylenedioxy substituent.

Derivatization yields, besides increased volatility, enhanced thermostability. Indeed, GC–FTIR analysis of the HFB derivative of MBDB (code DT108) yielded much clearer results. In this case, the chromatogram contains only one strong peak (Fig. 6); the second peak having an intensity characteristic to impurities. The FTIR spectrum (Fig. 7) obtained for the main chromatographic peak was again analyzed with the help of the expert system, which assigned the **DT** class identity to the compound. The identity of DT108 was also confirmed by the visual inspection of its spectrum in comparison with the spectra of the other HFB derivatives of 3,4-methylenedioxy analogues. The region between 1550 and 1150 cm^{-1}, where the best band-parameter stability is encountered, provided the most reliable confirmation that DT108 is an HFB derivative of a 3,4-methylenedioxy analogue. The absorption band associated with the breathing vibrations of the phenyl ring (1490 cm^{-1}) appears in this region together

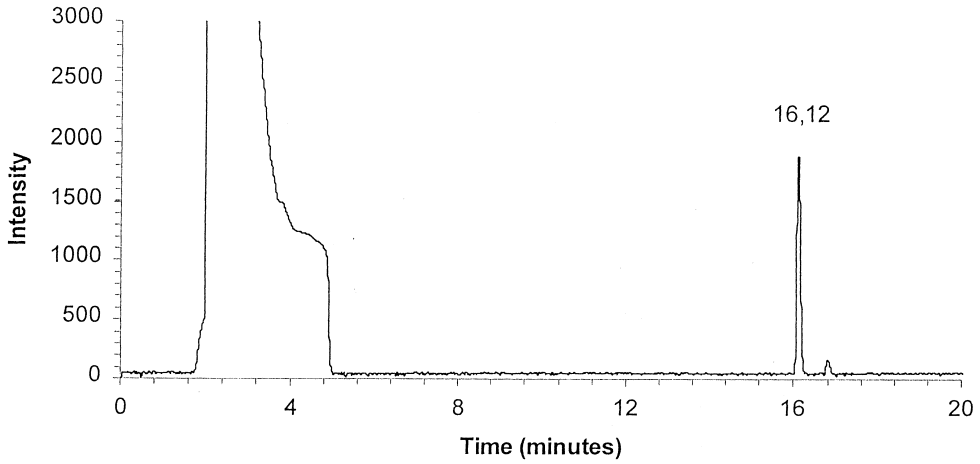

Fig. 6. The GC reconstructed chromatogram obtained for the HFB derivative of the reference standard of MBDB (MBDB-HFB, DT108).

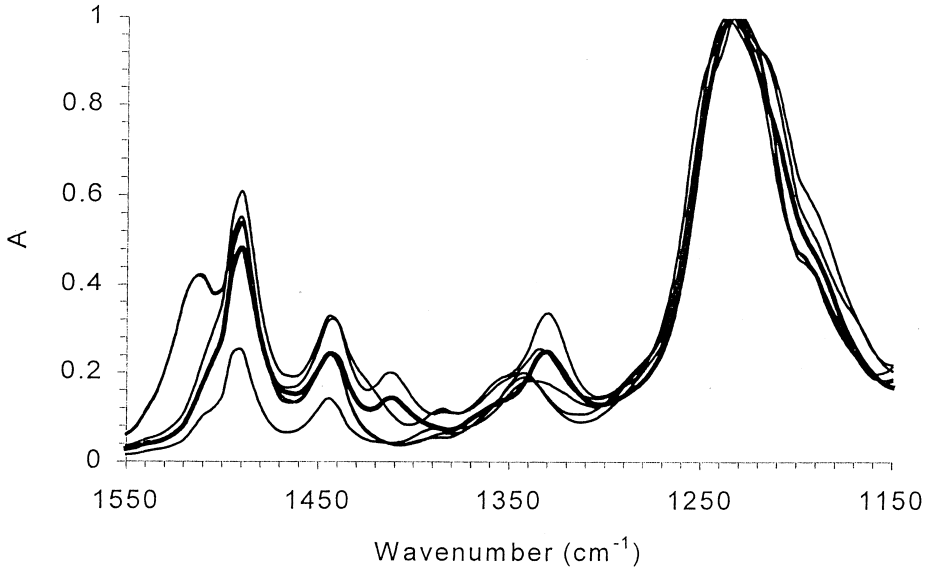

Fig. 7. The band-parameter stability of the original FTIR spectra of the HFB derivatives of 3,4-methylenedioxy analogues [(——) DT108, and the (———) HFB derivatives of the rest of the compounds forming the training set of the **DT** model] in the spectral region of the breathing vibrations of the phenyl ring and of the C–F stretching vibrations (1235 cm^{-1}).

with the stretching vibrations of the C–F bond (1235 cm^{-1}), the latter being the strongest band in the spectra of the HFB-derivatized compounds.

3.3. Comparison with methods "traditionally" applied for computer-assisted identification of molecular structure

To highlight the benefits of using the approach of pattern recognition for the detection of the correct chromatographic peak as described in this paper, we critically evaluated the system by comparing its results with the "traditional" ones performed by the commercially available programs. The methods applied for computer-assisted identification of compounds using IR spectra can be classified into three groups: structure elucidation systems in which spectroscopic and chemical information expertise is encoded to assist spectra interpretation, the routine approach of searching in the spectral library, and pattern recognition techniques that have the ability to recognize structural properties by classifying spectral data.

In the case of the structure elucidation method, the approach is based on IR peak recognition, which provides information about the presence of the substructures in organic compounds using features directly derived from peak tables. Fragment codes are widely used in computer-assisted structure elucidation [27,28] and particularly to interpret and predict IR spectra [29,30]. When applied to MBDB, the information obtained with this method did not provide the means to distinguish between the non-degraded and degraded samples and thus was irrelevant for the selection of the correct chromatographic peak. The instrument's routine program indicated, for both T106 and D107, the presence of an alkyl group, while for D107 it also indicated the presence of an aromatic ring, possibly 1–4 substituted. The latter information could not serve as an indication of a degradation consisting of a change of the substitution pattern of the aromatic ring in the case of D107, as the same structural pattern is suggested in the case of some 3,4-methylenedioxyamphetamines (i.e. T79, T82, T85) but not for others (i.e. T87, T113).

The model used by the commercial library search

programs [31–37] is basically a nearest neighbor search in a structured library. The compounds of the reference library are ranked in order of decreasing similarity with the input query (target spectrum and structure). The similarity between the target and the spectra present in the library is assessed using measures such as the Euclidean distance, or Tanimoto coefficient. The rationale for such a searching mechanism is the reasonable assumption that similar structural features usually will give rise to similar IR spectra. Thus, the top ranking molecules are expected to have the highest probability of exhibiting a similar spectrum. The structures and the spectra are considered as a whole to preserve the spectral behavior uncharacterized in structure elucidation tables (and programs), such as skeletal vibrations, the higher harmonics and combination bands, as well as effects of molecular symmetry that are usually not listed in structural fragment–subspectra correlation tables.

The library search process, especially the nearest neighbor approach, has the distinct advantage (as compared to the pattern recognition techniques) that it does not require predefined classes. However, this second approach is also of little help when the spectrum of the unknown sample *is not* present in the spectral library, as it happened in the analysis of MBDB. For example, the library search process has indicated that the Euclidian distance between the (normalized) spectrum of T106 and the spectra of the 3,4-methylenedioxyamphetamines present in the digital library may be as large as 2.477145 (the case of 3,4-methylenedioxy-*N*-hydroxyamphetamine), while the distance between T106 and safrole (4-allyl-1,2-methylenedioxybenzene, a precursor *without* hallucinogenic activity due to the lack of the amino group) is only 2.123325. Similar distances are obtained while quantifying the similarity of D107 with T113 (1.885174), T82 (1.9671284), T87 (2.060229) or T85 (2.768716). This result is not surprising. Small molecules, such as the phenethylamines, display significant spectral differences when small structural changes occur, which result in relatively large differences between the similarity coefficients computed for very similar structures.

In conclusion, in cases where a compound shows several chromatographic peaks, GC–FTIR users may often find it difficult so select the correct chromatographic peak using the well-established techniques of spectral searching, such as structure elucidation or library searching programs. Unless more elaborate techniques—such as pattern recognition—are used, the only way to select the correct peak is to ignore automatized procedures and perform the time-consuming classical interpretation of the spectra by a human expert.

4. Conclusion

A knowledge-based system, simultaneously screening for the main stimulant and hallucinogenic amphetamines with GC–FTIR, has been built by computing PCA models that initiated the SIMCA classification. An optimization process allowed a 93.92% total correct classification rate, and the classification with a 5% confidence level of 81.13% of the tested compounds. The discrimination was enabled by consideration of chemical factors influencing the vibrational spectra of the modeled compounds. The admitted structural variations among the modeled hallucinogens (nature of the substituents and ring conjugation) yield specific band intensity alterations and band shifts due to electron delocalization, distinguishing their spectral behavior from that of the non-modeled compounds.

Applied as a screening test, the expert system has proven to be highly specific and sensitive. When more than one strong peak is obtained in the chromatogram, the additional peaks are very often artifacts formed during the GC–FTIR analysis of thermolabile or chemically unstable compounds, and the reference standard chromatographic peak has to be determined. The results presented in this paper show that the expert system may also be a useful analytical tool for the identification of such peaks. The SIMCA scores (sample-to-model distance and leverage) were useful quantitative measures of the molecular similarity of a compound with modeled structural prototypes.

As it has been shown for the case of the psychotropic amphetamine derivative *N*-methyl-1-(3,4-methylenedioxyphenyl)-2-butanamine (MBDB), the shape of the chromatographic peak may be misleading. The identity assignment performed with the help of the expert system indicated that the compound

generating a well-defined chromatographic peak is a phenyl disubstituted amphetamine analogue with a substitution pattern of the phenyl ring different than that of the 3,4-methylenedioxy analogues. Vibrational analysis confirmed that the chemical changes that occur in the molecule of MBDB may be due to thermal degradation yielding a compound that seems to have an exocycle (attached to the phenyl ring) producing an electron delocalization different than the 3,4-methylenedioxy substituent, and possibly containing a larger number of carbon atoms. Including HFB derivatization in the sample preparation may be a solution to avoid thermal degradation and the resulting analytical ambiguity.

Acknowledgements

This study was conducted within projects (BIL 97/75 and BIL 00/39) included in and financed by the Bilateral Scientific and Technological Cooperation Agreement between Flanders and Romania.

References

[1] C. Cleva, C. Cachet, D. Cabrol-Bass, T.P. Forrest, Anal. Chim. Acta 348 (1997) 255.
[2] K. Baumann, J.T. Clerc, Anal. Chim. Acta 348 (1997) 327.
[3] H.J. Luinge, M.H. Langoor, J.H. van der Maas, Anal. Chim. Acta 348 (1997) 471.
[4] P.N. Penchev, G.N. Andreev, K. Varmuza, Anal. Chim. Acta 388 (1999) 145.
[5] F. Ehrentreich, Anal. Chim. Acta 393 (1999) 193.
[6] V.N. Piottukh-Peletsky, B.G. Derendyaev, Anal. Chim. Acta 396 (1999) 99.
[7] M.E. Elyashberg, E.R. Martirosian, Yu.Z. Karasev, H. Thiele, H. Somberg, Anal. Chim. Acta 337 (1997) 265.
[8] M.E. Elyashberg, Yu.Z. Karasev, E.R. Martirosian, H. Thiele, H. Somberg, Anal. Chim. Acta 348 (1997) 443.
[9] G. Gini, V. Testaguzza, E. Benfenati, R. Todeschini, Anal. Chim. Acta 43 (1998) 135.
[10] B.K. Alsberg, R. Goodacre, J.J. Rowland, D.B. Kell, Anal. Chim. Acta 348 (1997) 389.
[11] A. Eghbaldar, T.P. Forrest, D. Cabrol-Bass, Anal. Chim. Acta 359 (1998) 283.
[12] C.G. Georgakopoulos, M.C. Statheropoulos, G. Montaudo, Anal. Chim. Acta 359 (1998) 213.
[13] C.S. Tong, K.C. Cheng, Chemometr. Intell. Lab. Syst. 49 (1999) 135.
[14] M. Praisler, I. Dirinck, J.F. Van Bocxlaer, A.P. De Leenheer, D.L. Massart, Anal. Chim. Acta 404 (2000) 303.
[15] S.B. Karch, Drug Abuse Handbook, CRC Press, New York, 1998.
[16] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Data Handling in Science and Technology, Handbook of Chemometrics and Qualimetrics. Part B, Elsevier, Amsterdam, 1998.
[17] M. Praisler, I. Dirinck, J.F. Van Bocxlaer, A.P. De Leenheer, D.L. Massart, Talanta 53 (2000) 155.
[18] M. Praisler, I. Dirinck, J.F. Van Bocxlaer, A.P. De Leenheer, D.L. Massart, Talanta 53 (2000) 177.
[19] M. Praisler, I. Dirinck, J.F. Van Bocxlaer, A.P. De Leenheer, D.L. Massart, J. Anal. Toxicol. 25 (2001) 45.
[20] M. Praisler, J.F. Van Bocxlaer, A.P. De Leenheer, D.L. Massart, Turk. J. Chem. 26 (2002) 45.
[21] A.V. Pirogov, M.M. Platonov, I.V. Pletnev, O.N. Obrezkov, O.A. Shpigun, Anal. Chim. Acta 369 (1998) 47.
[22] A. Nijhuis, S. De Jong, B.G.M. Vandeginste, Chemometr. Intell. Lab. Syst. 38 (1997) 51.
[23] M. Pompe, M. Razinger, M. Noviè, M. Veber, Anal. Chim. Acta 348 (1997) 215.
[24] E. Forgács, T. Cserháti, Anal. Chim. Acta 348 (1997) 481.
[25] B. Divjak, M. Moder, J. Zupan, Anal. Chim. Acta 358 (1998) 305.
[26] C.E. Anderson, R.G. Nieves, J.H. Kalivas, Chemometr. Intell. Lab. Syst. 41 (1998) 115.
[27] M.E. Munk, B.D. Christie, Anal. Chim. Acta 216 (1989) 57.
[28] W.A. Warr, Anal. Chem. 65 (1993) 1087A.
[29] E.W. Robb, M.E. Munk, Mikrochim. Acta I (1990) 131.
[30] M.E. Munk, M.S. Madison, E.W. Robb, Mikrochim. Acta II (1991) 505.
[31] P. Willet, Similarity and Clustering in Chemical Information Systems, Chemometrics Series 12, Research Studies Press, London, 1987.
[32] P. Willet, V. Winterman, Quant. Struct.-Act. Relatsh. 5 (1986) 18.
[33] H.J. Luinge, E.D. Leussink, T. Visser, Anal. Chim. Acta 345 (1997) 173.
[34] P.A. Bath, C.A. Morris, P. Willet, J. Chemom. 7 (1993) 549.
[35] D.L. Venton, C.P. Woodbury, Chemometr. Intell. Lab. Syst. 48 (1999) 131.
[36] A.K. Leung, F. Chau, J. Gao, T. Shih, Chemometr. Intell. Lab. Syst. 43 (1998) 69.
[37] D.H. Drewry, S.S. Young, Chemometr. Intell. Lab. Syst. 48 (1999) 1.