Journal of Chromatography, 419 (1987) 135–154
Biomedical Applications
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMBIO. 3680

IDENTIFICATION OF DRUGS IN HUMAN SERUM BY HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY WITH PHOTODIODE ARRAY DETECTION AND A SEARCH ALGORITHM FOR ULTRAVIOLET SPECTRA

ELISABETH I. MINDER* and RENÉ SCHAUBHUT

Central Laboratory of Clinical Chemistry, University Hospital, CH-8091 Zurich (Switzerland)

CHRISTOPH E. MINDER

Institute for Social and Preventive Medicine, University of Bern, CH-3010 Bern (Switzerland)

and

DIETER J. VONDERSCHMITT

Central Laboratory of Clinical Chemistry, University Hospital, CH-8091 Zurich (Switzerland)

(First received September 8th, 1986; revised manuscript received February 23rd, 1987)

SUMMARY

A system is described that identifies a number of unknown drugs (benzodiazepines, antidepressants and neuroleptics) in blood specimens for clinical-toxicological purposes. Reversed ion-pair high-performance liquid chromatography with a photodiode array detector saves the ultraviolet spectrum of every chromatographically significant peak. Post-run data-processing, provided by a microcomputer, retrieves candidate substances from a library of ultraviolet spectra. Selected standard ultraviolet spectra are compared with the unknown by five different similarity tests. The discriminatory efficiency of these algorithms has been determined. Multicomponent analysis, a built-in program of the spectrophotometer, provided the most reliable results.

INTRODUCTION

A challenging task in clinical toxicology is the identification of the causative agent(s) during an acute poisoning episode [1]. The actual condition of the patient determines its necessity, as the clinical evaluation alone may be insufficient for adequate treatment of the severely poisoned. The patient's history may be missing, unreliable or incomplete [1], mixed poisoning may change, mask or potentiate the clinical symptomatology and, further, modern therapeutic tools

0378-4347/87/$03.50

increasingly require fast and precise diagnosis. The chemical evaluation of the patient's fluids therefore has to prove the existence of a poison, if present, as well as determine its nature and eventually to quantify it.

In contrast to forensic toxicology, clinical toxicology has to deliver its result within a short time span, beyond which the analysis will be of only academic interest [2]. Therefore, a number of fast qualitative tests have been developed [1,3,4], although most of them allow only group identification. As the relative potencies of chemically related drugs can differ greatly, more precise identification methods may deliver useful and eventually decisive information to the clinician.

Different drug identification systems have been described: combinations of different thin-layer systems and gas chromatography [5–7]; infrared (IR) spectrometry; gas chromatography combined with mass spectroscopy [8,9]; and, more recently, high-performance liquid chromatography (HPLC) [5,10]. Ultraviolet (UV) spectroscopy is an alternative tool of satisfactory reliability and short analysis time [11]. Drug identification by UV spectroscopy is accomplished by comparing the unknown spectrum with tables and spectra of reference books [7,11] or spectra of reference compounds [12]. If HPLC is combined with an on-line photodiode array spectrophotometer as detector, its supplementary use can be made of UV spectroscopy. As a data-processing device usually is connected to a photodiode array spectrophotometer, the on-line recorded UV spectra can easily be compared with standard spectra.

The application of this method to clinical toxicology has been proposed [13], but no detailed system has yet been described. In order to automate the comparison procedure, we have implemented a computerized search algorithm that retrieves similar UV spectra from a library, compares them with the spectrum of the unknown and calculates the similarity. Whereas for IR and mass spectra a number of search algorithms exist, to our knowledge this is the first search algorithm for UV spectra that has been described and tested in some detail.

In order to render such an analytical system practical for the intended purpose, we have developed an isocratic HPLC system suitable for the analysis of the three drug classes, benzodiazepines, antidepressants and neuroleptics. They were chosen with regard to the epidemiology of acute intoxications in Switzerland [14].

The aim of the present work is mainly to test whether UV spectra suffice for reliable drug identification by means of a computerized search algorithm. The algorithm combines five different comparison methods. The discriminatory efficiencies for the whole algorithm and for each of these five comparison methods have been determined and will be discussed. Although a rather small spectra library of only 48 different compounds (all belonging to one of the three mentioned pharmacological classes) was used for reliability testing, the results can be considered to reflect the potential of computerized drug identification by UV spectroscopy, as each of the standards had close chemical similarities with other compounds in the library. Thus, the identical counterpart of an unknown had to be distinguished from a number of close contenders for unique identification. Obviously, UV spectroscopy will not be able to discriminate compounds of identical chromophoric structures.

The question arises as to what happens if the unknown has no identical counterpart in the library. Therefore, tests were carried out to see if the calculated similarity score value gives any hint whether or not the unknown substance is contained in the archive. Further, the grade of chemical similarity between unknown and selected candidates was determined, in cases where no identical counterpart was contained in the library.

In order to improve the practicality of the system further expansion of the library will be necessary. However, an increase in library size will render more difficult an optimal selection of compounds that may be compared in detail with the unknown. Marker points, such as maxima and minima, which can serve as selection criteria for archive retrieval, are fewer in UV spectra than in IR or mass spectra. The addition of a further selection criterion, independent of UV spectroscopy, will cope with this problem. The retention time of the unknown, as determined by three-dimensional chromatography, is suitable as a further identification criterion and will greatly increase the discriminatory power of the method. This will be visualized by using measured variations of retention times and of wavelengths of absorbance maxima.

EXPERIMENTAL

Instrumentation

The HPLC system consists of an HPLC pump (Model 414, Kontron, Zurich, Switzerland), a U6K injector (Waters Assoc., Millford, MA, U.S.A.), a Supelcosil LC 18DB column (250×4.6 mm I.D., 5 μm) (Supelco, Bellefonte, PA, U.S.A.), a photodiode array spectrophotometer (Model 8450, Hewlett-Packard, Palo Alto, CA, U.S.A.) equipped with a flow-cell No. 178.32 (Hellma, Müllheim, Baden, F.R.G.). A Hewlett-Packard 85B microcomputer supplemented with a 10-Mbyte Winchester disk station (HP 9153) and a plotter (Model HP 7440) served to control the spectrophotometer and for data-processing. The software to support communication between the HP 85 and the spectrophotometer and to enable post-run data-processing was obtained from Hewlett-Packard (Switzerland). It was modified and supplemented to meet our needs by one of the authors (E.I.M.). Its structure (Table I) will be explained in detail below. Modified parts are marked with one asterix, newly developed ones by two. Reference drug substances were obtained from the respective pharmaceutical companies (see Acknowledgements). Solvents and other chemicals were reagent or chromatographic grade and were purchased from Merck (Darmstadt, F.R.G.). Water was quartz-distilled.

HPLC conditions

The mobile phase was 500 ml of 0.1 *M* phosphate buffer (pH 2.7), 440 ml of acetonitrile, 60 ml of methanol and 200 μl of triethylamine. The flow-rate was 1 ml/min, and temperature ambient. The recorded wavelengths were 205, 215, 230, 240, 254, 280 and 300 nm, and the reference wavelength range was 780–800 nm. The interval time was 1.5 s. The recorded spectra range was 200–400 nm.

TABLE I

STRUCTURE OF THE PROGRAM LINKING THE HP 85 AND THE SPECTROPHOTOMETER

SET-UP-OF DATA-RECORDING CONDITIONS★
Number of simultaneously recorded wavelengths: 1–7
Wavelength values for recording: any from 200 to 400 nm and any even from 402 to 800 nm
Reference wavelength range (optional): e.g. 780–800 nm (must be outside the range where drugs absorb)
Time interval for measurements: e.g. 1.5 s
Wavelength range for UV spectra: e.g. 200–400 nm
DATA RECORDING★
Real-time chromatogram★
Retention time and peak height★★
Storing data of up to seven simultaneous wavelength chromatograms (evt. permanently)
Recording of UV–visible spectra★
Automated after each significant peak★★
additional by keyboard entry
POST-RUN DATA-PROCESSING★
Computation of peak areas and heights
Pseudo-three-dimensional plotting★★
IDENTIFICATION PROCESS★★
Smoothing of UV spectra★★
Automated search for absorbance maxima★★
(optional: additional absorbance maxima by keyboard)
Establishment and update of library★★
archive of names, absorbance maxima and retention times and UV spectra digitized in intervals of 1 nm★★
Identification search algorithm★★

★Modified parts.
★★Newly developed parts.

Real-time chromatogram and data acquisition

Before each chromatographic run a so-called "balance operation" of the photometer is performed. This procedure sets the baseline to zero over the whole wavelength range and corrects for dark current of the photodiodes as well as for absorbance changes between different batches of mobile phase. During the chromatographic run, absorbances from up to seven wavelengths are recorded simultaneously. The performance of the present photometer allows a minimum interval of 1.5 s between two measurements. The real-time chromatogram represents the highest absorbance measured per unit of time in any one of the synchronously monitored wavelengths. Chromatographic peak maxima are followed by storage of a UV spectrum in the memory of the photometer. Additional spectra can be saved by manual keyboard entry.

Optional post-run data processing

The wavelength chromatograms may be plotted either singly or as a pseudo-three-dimensional graph. These chromatograms can be integrated for quantifi-

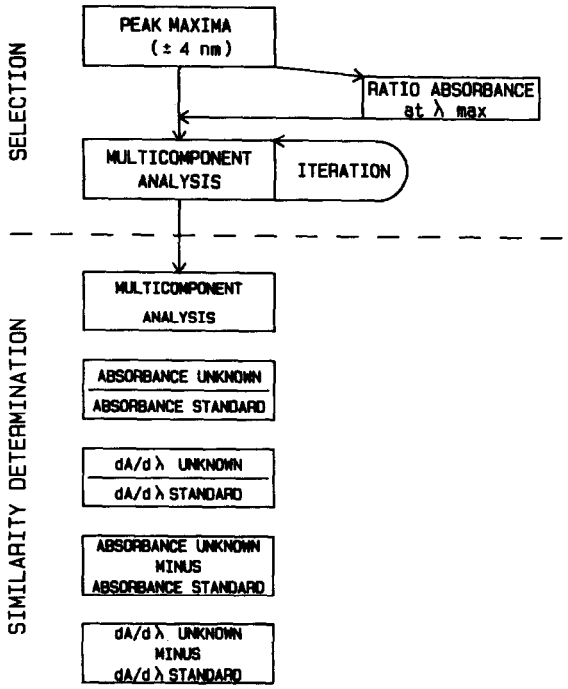

Fig. 1. Flow chart of the search algorithm.

cation. Because data acquisition is interrupted during spectra storage, peak heights rather than peak areas have to be used. Calculation and visualization of absorbance ratios at different wavelengths versus suitable time period can be performed as check for peak purity [15].

The spectra library

The archive of standards consists of three parts. Part 1 contains the names of standards plus their identification numbers. Part 2 consists of identification numbers, wavelengths and absorbances of absorbance maxima and (optionally) retention times. This information is ordered according to the wavelengths of absorbance maxima, ranging from 200 to 400 nm at intervals of 1 nm. Part 3 stores the digitized standard UV spectra. They have been recorded under HPLC conditions identical with those used to analyse the unknowns. The contents of the archive can be updated.

The search algorithm (Fig. 1)

If a significant peak appears on the real-time chromatogram of an unknown sample and its UV spectrum has been recorded, this peak may be subjected to the search process. The first step is to localize the absorbance maxima by means of a built-in algorithm of the spectrophotometer. These maxima serve as keys for the archive retrieval. The wavelength window may be chosen from ± 1 to ± 4 nm, but for the present work it was fixed at ± 4 nm [11] resulting in a total window of 9 nm. If the unknown UV spectrum exhibits more than one absorbance maximum,

the search process retrieves corresponding absorbance maxima of standards and compares their relative absorbances. If the absorbance ratio of standard peak maxima is at least in the limit of ±50% of the unknown, this substance will be positively selected.

The UV spectra of the selected candidates are then transferred into the memory of the spectrophotometer. The number of candidates is restrained by means of multicomponent analysis, an algorithm built in to the spectrophotometer [16], whereby the unknown UV spectrum is considered as the summation spectrum of the selected standards. Using eqn. 1, the sum of the squared differences of the absorbances of the unknown minus the summarized absorbances of the standards over the specified wavelength range is calculated.

$$\mathrm{SUM} = \sum_{i=1}^{n} \left[A_i - \sum_{k=1}^{p} C_k \, (A_{aik}/C_{ak}) \right]^2 \tag{1}$$

where A_i = absorbance of unknown at wavelength i; n = number of wavelengths; C_k = concentration of the kth standard; p = number of standards; A_{aik}/C_{ak} = absorptivity of standard k at the ith wavelength of its reference spectrum "a".

This sum is minimized by estimation of the factor C_k for each standard. According to the Beer–Lambert law these factors (C_k) can be considered as concentrations. The best fit between the unknown and the summation spectrum of standards may be achieved if one or several of these factors have negative values. Thus apparently negative concentration values may result from the calculations for one or several standards. As negative concentrations are meaningless in chemical terms, these substances are excluded and the multicomponent analysis is again calculated for the remaining standards, thus reducing the number of candidates progressively. Iterations are repeated until there are fewer than four candidates left, or no further substance can be excluded for its apparent negative concentration value.

The remaining candidates are then subjected to five different comparison algorithms, the first being the last iteration of multicomponent analysis. Beside the concentrations, the algorithm estimates two statistical terms. The relative fit error quantifies the difference between the spectrum of the unknown and the calculated summation spectrum of standards. The size of the relative fit error indicates whether the unknown compound is present among the standards tested or not. A value less than 1 suggests a present counterpart, but a value greater than 1 probably means that no counterpart is among the retrieved standards. The relative standard deviation (R.S.D.) of each component is an indicator of the precision of the estimation of concentration that can be expected for repeated measurement of the same sample. Experience shows that the R.S.D. is generally less than 0.1 when the unknown compound is identical with the standard considered. The statistical methods used are a first estimate by least-squares and finer estimates by maximum likelihood. The exact procedures are not available from Hewlett-Packard.

The UV spectra in the region 200–350 nm are usually used for the residual

comparison methods, as the drugs analysed exhibit only low absorptivity above 350 nm in most instances. Absorptivity is low even below 350 nm for some substances (e.g. nortriptyline). Their spectra are truncated and only the significant part is retained. In practice, this covers at least the range 200–280 nm.

The remaining similarity tests are the following.

(1) The absorbance ratio. The absorbance ratio of unknown divided by the library standard is calculated at intervals of 1 nm over the significant part of the UV spectrum. The mean, S.D. and R.S.D. of the resulting absorbance ratios are computed.

(2) The ratio of first derivative spectra. Values of $dA/d\lambda$ of the unknown divided by $dA/d\lambda$ of the standard are recorded at intervals of 1 nm. The mean, S.D. and R.S.D. of the resulting ratio values are calculated.

(3) Difference of normalized absorbance spectra. The absorbance spectra of both the unknown and the standard are normalized to a mean absorbance of one over the significant part of the spectrum. Care must be taken that both spectra are truncated to the same extent. For each interval of 1 nm the difference of the two absorbances is calculated and the mean, S.D. and standard error of mean (S.E.M.) were calculated.

(4) Difference of normalized first derivative spectra. The procedure was analogous to that described under (3), except that instead of absorbance data first derivative data were used.

Lastly, the cumulative score was calculated. This is the product of the score components mentioned in the following steps 1–6. (The Appendix shows an example of the procedure as used with a serum spiked with flunitrazepam as "unknown".)

(1) If only one absorbance maximum exists, the contribution to the score is the square of wavelength difference of the absorbance maxima of the unknown and standard, multiplied by 10. If there are several corresponding absorbance maxima for the unknown and standard, the standard is selected several times, once for every maximum located in the defined search window. Then tests for similarity of their absorbance ratios R [i.e. $(E_{P1}/E_{P2})_{unknown}$:$(E_{P1}/E_{P2})_{standard}$, where E_{P1} = absorbance maximum 1 and E_{P2} = absorbance maximum 2] are performed. If R is not within the limits (>0.5 and <1.5), the standard is not selected. For n absorbance maxima the formula for the score contribution is:

$$S = 2^{(n-1)} \sqrt{\prod_{i=1}^{n} R_i^* \, (\lambda_{ui} - \lambda_{si})^2}$$

$$R_i^* = \begin{pmatrix} R \text{ when } R > 1 \\ 2 - R \text{ when } R < 1 \end{pmatrix}$$

λ_{ui} = wavelength at absorbance maximum i of unknown
λ_{si} = wavelength at absorbance maximum i of standard

(2) The R.S.D. is calculated by multicomponent analysis. If repetitive estimates are made, these factors are multiplied.

(3) The R.S.D. of ratio absorbance.

(4) The R.S.D. ratio of first derivative spectra.

(5) The S.E.M. difference of normalized absorbance spectra.

(6) The S.E.M. difference of first derivative spectra.

This scoring system is designed to imitate the judgement of an expert making a decision on the similarity of UV spectra. The best similarity score is represented by the lowest value. If the UV spectrum of a library standard is used as unknown, a cumulative score value of zero results.

The first aim of the current work was to determine the overall reliability of the algorithm in identifying unknown samples. Identification was only accepted as correct if the identical library counterpart of the test substance analysed as unknown received the lowest cumulative score value. Even if a closely related substance exhibited the best similarity score, identification was defined as incorrect.

Secondly, the relative discriminatory potency of the five similarity tests mentioned above was checked. Any interference in the analytical procedure, such as instrumental noise, baseline shift and coeluting substances, may affect each of the methods differently [13,17].

We injected ca. 0.2–1 μg of different drug substances either as standard solutions or as plasma samples that had been spiked and extracted as unknowns. Twenty-eight different runs were subjected to the search algorithm. The sequential analysis described by Bross [18,19] was applied at a significance level of $\alpha = 0.05$ to monitor the efficiency of each test, including the similarity score (cumulative weight). The results are shown in Fig. 2. Each small field represents one search. Correct identifications of the particular similarity test resulted in progression of one field on the abcissa, every incorrect one in one field on the ordinate.

RESULTS AND DISCUSSION

Mobile phase

When the photodiode array spectrophotometer is used as a detector, a mobile phase with low absorbance over the whole wavelength range is required. Further, it must elute all the different substances with satisfactory peak shape and in a short time. Buffers that were compared include glycine, phthalate, citrate and PIC B7 (heptanesulphonic acid buffered to pH 3.5 by a compound not mentioned by the manufacturer (Waters Assoc.). Phosphate buffer (pH 2.7) supplemented with 0.147 μM triethylamine and with the organic solvent acetonitrile, and only a small proportion of methanol, served our intentions best. In particular, peak tailing observed in some compounds was considerable reduced by adding triethylamine [20], whereas the ion-pairing reagent heptanesulphonic acid was less efficient in improving the peak shapes. However, one of the substances, flupentixol, was eluted as a double peak. The retention times of all but one drug are less than 12 min: penfluridol was recovered after 20 min. The k' values (relative capacity factors) were uniformely distributed [21].

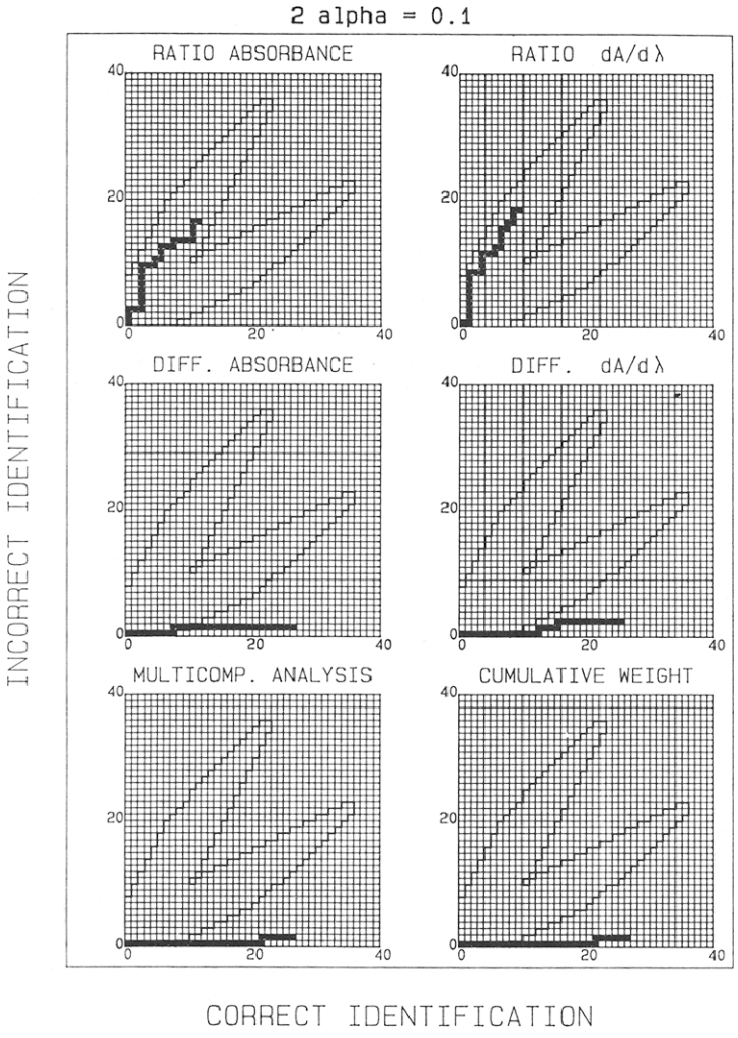

Fig. 2. Discriminatory efficiency of six different comparison metrics (specified in the text). Sequential analysis diagrams showing one field progression for each correct identification on the abcissa, for each incorrect on the ordinate. The bold line marks the significance limit $\alpha = 0.05$; in the lower right parts, the correct identification is significant more often than the incorrect; in the upper left part the incorrect identification occurs more often than the correct; in the middle part, the correct and incorrect identifications are equally frequent.

Processing the unknown sample

The amount of unknown analysed ranged between 0.2 and 1 μg. If 1 ml of serum is assayed, this corresponds to the lower toxicity level in most of these drugs [22,23]. But some of them, such as alprazolam, triazolam and flunitrazepam, exhibit considerably higher pharmacological potency, and therefore lower toxicity levels are to be expected. Peak absorbance of 1 μg of substance was usually ca. 0.1 absorbance units. Scattering, predominantly in low UV region, was rarely found to interfere with the peakfind algorithm of the HP 8450 spectrophotome-

ter. Nonetheless, to deal with these rare events a supplementary subroutine was introduced that allows manual localization of the UV peak maxima on the screen of the spectrophotometer and then keyboard entry into the microcomputer.

The search algorithm: test compound contained in library

An example of a search with explanatory comments is given in the Appendix. Avoidance of any distortion of baseline was found to be crucial, as otherwise interference with the identification process occurs [13]. Therefore, the instrumental baseline was set to zero after steady-state HPLC conditions were established, in particular the mobile phase composition, flow-rate and warming-up of the detector. The baseline was again reset to zero just before each chromatographic run. Thus, baseline subtraction was implicitly performed. As we used an isocratic system, baseline shift during chromatography was not a major problem. Extracts or spiked sera were prepared to be as clean as possible, to prevent coelution of endogenous material with drugs eventually present.

If different compounds coelute in a single chromatographic run, their identification is impaired or impossible, as in other chromatographic identification procedures. Additional UV spectra saved from the flanks of such a peak may represent relatively pure component spectra. Their incongruity reveals peak overlapping and they are suitable for identification, if their signal-to-noise ratio is adequate. Another way to check for peak purity is to plot the absorbance ratio at two different wavelengths versus time [15]. A horizontal line should appear in a pure peak. However, this test proved to be not very reliable in our experience.

The first two steps of the search process (selection according to peak maxima and eventually according to their absorbance ratios; see Fig. 1) retrieved possible candidates from the library. At this stage their number was median 8, the range 2–11 ($n=20$). The iterative process of multicomponent analysis reduced the number to median 3, ranging from 2 to 5.

The evaluation of the components of the search algorithm by means of sequential analysis (Fig. 2) demonstrates that three single tests and cumulative scores (cumulative weight) identify the correct component significantly more often than any incorrect. Moreover, the graph reflects the performance of each of the tests by the trend of the symbols. Thus, multicomponent analysis proved best from all single tests. Difference of normalized absorbance spectra and difference of normalized first derivative spectra were almost as reliable as multicomponent analysis. But absorbance and first derivative ratios did not discriminate well between closely related drugs, such as different benzodiazepines or several tricyclic antidepressants.

The final results of the algorithm represented by the cumulative scores achieved correct identification in all but one case. This only exception was the wrong identification of chlorprothixene [α-2-chloro-9-(3′-dimethylaminopropyliden)thioxanthene] instead of zuclopenthixol [*cis*(Z)-2-chloro-9-(3-(N′-β-hydroxyethylpiperazino)propyliden)thioxanthene]. Both compounds are neuroleptics, have exactly the same chromophore, and differ only in their saturated side-chain.

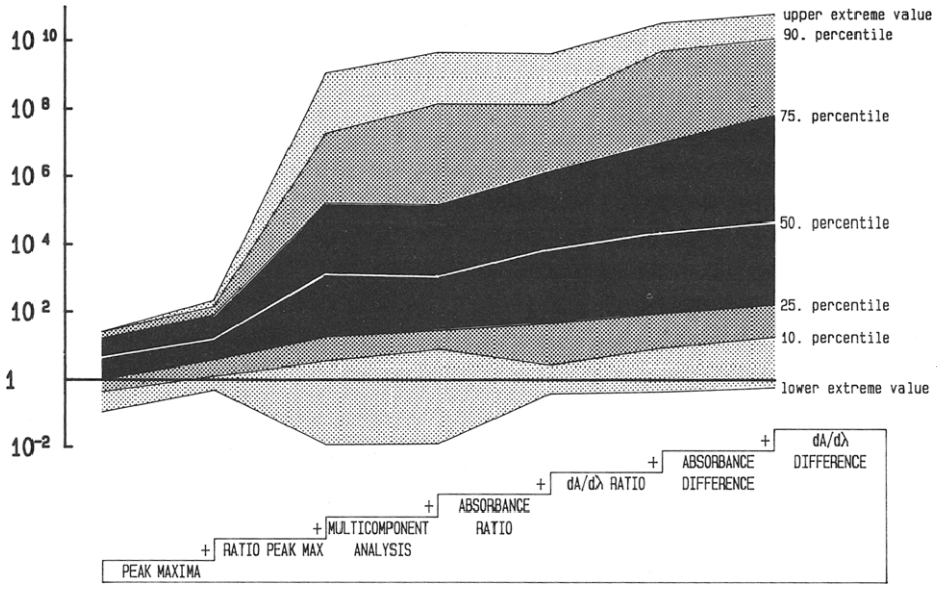

Fig. 3. Development of cumulative scores during the different steps of the search process (abcissa). Weight values of all retrieved, but non-identical, standards in relation to the identical (scaled on ordinate) are depicted as percentiles of their distribution. The values of seventeen search processes (at least 39 values per step) have been used.

The progressive development of the similarity score (cumulative weight) along the different steps of the search algorithm has been followed in seventeen identification runs (Fig. 3). Thereby the values of all candidates, not only of the best ones as in the sequential analysis, have been considered. The scores scaled on the ordinate are given in relation to the identical counterpart of the unknown. Thus, candidates with values less than one have a better score than the identical, those greater than one are worse. The distribution of scores is given in percentiles. In addition, both upper and lower extreme values are shown. If wavelengths of absorbance maxima alone are considered, one quarter of the score values are less than 1, three quarters greater than 1. Obviously, this relation can be influenced by the search window. If relative absorbance at the maxima can be added, more than 90% of values are greater than 1. Multicomponent analysis again seems to have great impact on the results, especially if the drastic reduction of the number of candidates during this step is taken into account. In addition, the four tests, absorbance and first derivative spectra ratios and difference of absorbance and first derivative spectra, positively influence the discriminatory power of the cumulative score test. It must be borne in mind that the sequence of the tests has a great influence on their apparent contribution to the cumulative score. This sequence is largely arbitrary, but determined by the ease of application to a large number of candidates. Easily applied tests are performed first to reduce the number of candidates to be analysed in the more complex tests.

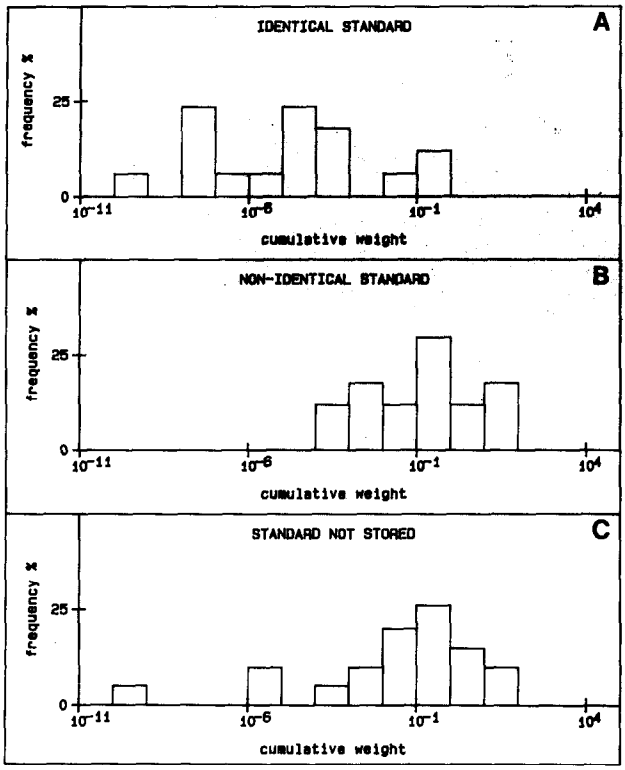

Fig. 4. Frequency histogram of absolute cumulative weight values (similarity scores). Graph A is the distribution for standards identical with the unknown ($n=17$). Graph B shows the distribution for best (lowest) values of standards not identical with the unknown ($n=17$). Graph C is the distribution of best (lowest) values when the unknown was not in the library ($n=20$).

Search algorithm: test compound not contained in the library

As yet only compounds contained in the library have been discussed. However, a substance to be identified may not yet appear in the library. In mass spectrometry a search is expected to deliver a closely related substance, if the compound itself is not stored [24]. However, molecular structures are less clearly represented by their UV than by their mass spectrum [25]. Direct elucidation of molecular structures by UV spectroscopy is scarcely ever successful and requires additional manipulation, e.g. addition of shift reagents [26]. Our intention was to elucidate what happens if a compound is not in the library. For this purpose, twenty drug compounds that were not yet part of the library were subject to the identification algorithm. These compounds were β-adrenoceptor blocking agents and antihistamines. The first are chemically unrelated to the substances already in the library, but some antihistamines are phenothiazines as some neuroleptics. These and other substances were subsequently added to the library, thus increasing it to 97 compounds during the study.

TABLE II

CUMULATIVE SCORE AND STRUCTURAL SIMILARITY BETWEEN TEST DRUG (UNKNOWN) AND RETRIEVED STANDARD (NO IDENTICAL COUNTERPART IN LIBRARY)

Cumulative score	Chemical similarity★			
	Class 3	Class 2	Class 1	Class 0
Less than 10^{-3}	1	2	0	1
10^{-3}-10^{-1}	0	4	3	3
10^{-1}-10	0	4	8	18
Greater than 10	0	1	0	12

★See text for explanation of classes.

In this experiment we tried to evaluate whether the absolute values of the cumulative score reflect the presence or absence of an identical counterpart in the library. In Fig. 4 the cumulative score (weight) values are depicted as frequency histograms in dependence on their absolute values for three different groups. Graph C shows cumulative scores of unknowns without an identical counterpart in the library ($n=20$). Graphs A and B illustrate the results of seventeen searches with the counterpart of the unknown contained in the library. Graph A shows the values for the identical counterpart. Only the best (lowest) values of the non-matching compounds of each search are depicted in graph B. For this kind of representation, absolute values are visualized rather than paired differences between the identical and non-identical standards. Therefore, a decrease in discriminatory power has to be expected, as shown by the overlap of groups A and B.

Nevertheless, groups A and B or A and C, respectively, show significantly different mean values ($\alpha < 0.005$, Wilcoxon range test). The lowest value in group C was $9.039 \cdot 10^{-10}$ and was achieved by analysis of the spectrum of promethazine [10-(2-dimethylaminopropyl)phenothiazine]. The algorithm retrieved promazine [10-(3-dimethylaminopropyl)-phenothiazine], a structural isomer with exactly the same chromophore. The two next lowest were doxylamine analysed and lofepramine retrieved, with a cumulative weighing factor of $4.554 \cdot 10^{-6}$. Further, terfenadine was injected, but diphenhydramine retrieved with a value of $8.328 \cdot 10^{-6}$. If the best cumulative score in a search is less than 10^{-4}, this is a hint for indentity. If it is greater than 10^{-3} then the compound tested is probably not in the library.

In addition, we looked to see if the substances retrieved by the algorithm are related to the test substance. During the 20 analyses of test compounds not in the library, 57 compounds were retrieved and their similarity scores calculated. As mentioned above, the 20 test substances were mostly chemically unrelated to the standards, but during the course of the study these test substances were added to the archive so that the number of related archive standards increased for the compounds later analysed.

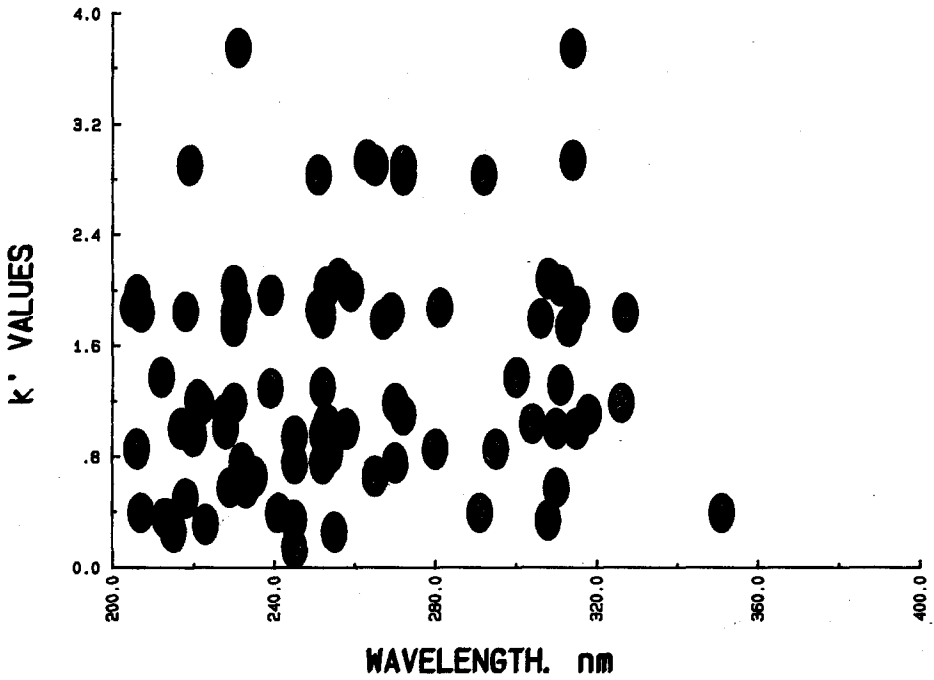

Fig. 5. Enhancement of the selection process by adding the k' values as an identification criterion. Absorbance maxima of 48 different compounds are located by their wavelengths (abcissa) and their k' values (ordinate). All maxima that a compound exhibits in the mobile phase have been shown. The shaded areas correspond to a wavelength interval of ± 3 nm and $\pm 0.14\,k'$ factor (equivalent to a retention time of 16). Horizontal shading indicates benzodiazepines ($n=15$); shading from top left to bottom right indicates antidepressant ($n=16$); shading from bottom left to top right indicates neuroleptics ($n=17$).

The 57 compounds were judged according their chemical relation to the test compound. Class 0 were substances without relation (it must be mentioned here that most of the compounds contain one or several benzene rings, but this feature was not classified as a relation). Class 1 were substances with differences within the conjugated double bond systems. Class 2 were substances with the same chromophoric ring system(s) but different auxochromes. Class 3 were compounds with identical chromophores (including at least three atoms of every side-chain). Table II shows the distribution of the substances with respect to their attributed cumulative weight values. This shows that the likelihood of a chemical relation increases as the weight factor decreases. However, the numbers of both the library standards and the test drugs are small, so that these results must be considered as preliminary.

The current study was designed to use only UV spectroscopy as a selection and comparison criterion. This allowed us to show that the developed search algorithm for UV spectra may be a valuable tool for drug identification. One may ask whether such a system will perform well once its standards library contains 200–500 compounds. In these circumstances, it is feasible to include retention time as additional discriminatory variable. The chromatographic system described

supplies the retention time of each compound as a measured characteristic. This feature, which is independent of UV spectroscopy, may serve to limit the number of candidates. Thus, as illustrated by Fig. 5, the retention time adds a further dimension to the selection process. On the abcissa the absorbance maxima of the 48 compounds of the three pharmacological classes, benzodiazepines, antidepressants and neuroleptics, are plotted versus the relative capacity factors. The area shaded is equal to a variation of ± 3 nm in the x-axis and ± 16 s in retention time (0.14 relative capacity factor, respectively). In our experience, using the described HPLC system, the absorbance maximum wavelength of a specific compound never varied more than 3 nm. For retention times we found a median variation coefficient of 1.15% (range 0.11–3.79%) in a total of 173 measurements using ten different substances and extended over several weeks. This corresponds to a median variation of 5.5 s. Using these values as search windows, very little overlapping occurs, as shown by the shaded areas on Fig. 5.

Thus, it may be speculated that an increase in the size of the library may be well tolerated by the search algorithm, if retention time is introduced as additional criterion. Such a system is now under evaluation, to show its practicality and to compare its performance with established clinical–toxicological methods.

ACKNOWLEDGEMENTS

We thank Hewlett-Packard Switzerland for providing part of the software package and Allan Olgivie for linguistic corrections of the manuscript. The following pharmaceutical companies are thanked for gifts of their pure drug substances: Angelini SA, Astra Pharmaceutica AG, Bayer AG Schweiz, Beecham AG, Boehringer Mannheim AG, Bristol-Myers SA, Ciba-Geigy AG, Cilag AG, Clin-Midy (Sanofi Pharma SA), Degussa AG, Duphar Schweiz, E. Merck AG Schweiz, Essex Chemie AG, Hoechst Pharma AG, Hoffman-LaRoche & Co. AG, ICI-Pharma AG, Inpharzam AG, Janssen Pharmaceutica AG, Kali-Duphar Pharma AG, Kramer-Synthélabo SA, Laboratoire Délagrange SA, Laboratoire Délalande SA, Laboratories Sauter SA, Laboratories Théraplix, Lundbeck & Co A/S, Mack Pharma AG, Medichemie AG, Merck Sharp & Dohme AG, Merrel Dow Pharmaceutica AG, Organon International AG, Pfizer AG, Pharmuka SA, Roussel AG, Sandoz AG, Searle SA (Suisse), Serum SA, Siegfried AG, Specia SA (Rhône-Poulenc Pharma Suisse), Squibb AG, Upjohn SA, Wander AG, Wellcome AG, Wyeth Laboratories (Doetsch, Grether & Cie AG), Zyma SA.

APPENDIX

APPENDIX: example of the search algorithm:
(flunitrazepam was used as "unknown")

example:	commentary:
*** IDENTIFICATION OF DRUGS ***	head: mentioning search conditions.
==========	
RUN NUMBER: #12/7/110786	
UNKNOWN NUMBER 12	
SEARCH WINDOWS:	
TIME WINDOW OPEN	
WAVELENGTH: 9nm	
==========	
SELECTION OF λMAX	selection from archivated standards.
MAX OF UNKNOWN 218nm	first absorbance maximum of
1 SUBSTANCE # 30	unknown at 218 nm, search
RATIO UNKNOWN/STANDARD 1.055	window +/- 4nm.
2 SUBSTANCE # 24	extended format of one retrieved substance:
RATIO OF UNKNOWN (max 218, 252)	abs.max. of unknown at
abs.:.0404, .0251	218nm:.0404, 252nm:.0251
1.6096	.0404/.0251=1.6096
RATIO OF STANDARD (max 218, 252)	abs.max. of std # 24 at
abs:.356, .228	218nm:.356, 252nm:.228
1.5614	.356/.228=1.5614
RATIO UNKNOWN/STANDARD 1.031	1.6096/1.5614=1.031
3 SUBSTANCE # 24	2nd retrieval of # 24
RATIO UNKNOWN/STANDARD 1.050	s.below
4 SUBSTANCE # 29	substances with only one
5 SUBSTANCE # 18	abs.max. within the search
6 SUBSTANCE # 33	window.
MAX OF UNKNOWN 252nm	2nd abs.max. of unknown
7 SUBSTANCE # 17	
8 SUBSTANCE # 5	
9 SUBSTANCE # 8	
10 SUBSTANCE # 16	
11 SUBSTANCE # 24	retrieves # 24 several times
RATIO UNKNOWN/STANDARD .970	once for every corresponding
12 SUBSTANCE # 24	pair of max. at each wave-
RATIO UNKNOWN/STANDARD 1.019	length. At each abs. max. 2
13 SUBSTANCE # 4	corresponding ones are found
14 SUBSTANCE # 28	results in 6 retrievals.

```
MAX OF UNKNOWN 314nm
 15 SUBSTANCE # 30
  RATIO UNKNOWN/STANDARD .948
 16 SUBSTANCE # 21
 17 SUBSTANCE # 24
  RATIO UNKNOWN/STANDARD .952
 18 SUBSTANCE # 24
  RATIO UNKNOWN/STANDARD .982
```

3rd abs.max. of unknown.

```
12 SUBSTANCES SELECTED
NAMES OF SELECTED SUBSTANCES:
FLUNITRAZEPAM # 24
 WEIGTH 1.0972
MIDAZOLAM # 29
 WEIGHT 10
FLUVOXAMIN # 8
 WEIGHT 10
NITRAZEPAM # 30
 WEIGHT 10.5469
IMIPRAMIN # 16
 WEIGHT 10
DESIPRAMIN # 5
 WEIGHT 40
CLOMIPRAMIN # 4
 WEIGHT 40
TRIMIPRAMIN # 17
 WEIGHT 40
MEDAZEPAM # 28
 WEIGHT 90
ALPRAZOLAM # 18
 WEIGHT 160
TRIAZOLAM # 33
 WEIGHT 160
CLONAZEPAM # 21
 WEIGHT 160
```

names of selected standards are given together with their accumulated weights.

```
multicomponent analysis
relative fit error 2.731E-01
independence of standards 3.452E+04
```

up to 11 stds plus a dummy spectrum can be processed in one batch.

STD	CONC	rel std dev
0 dummy	9.82E-02	8.99E-01
50 # 24	1.14E-01	6.07E-02
51 # 30	-3.69E-03	-1.10E+00*
52 # 29	-6.09E-04	-7.04E+00*
53 # 18	-1.51E-03	-1.87E+00*
54 # 33	7.31E-03	1.04E+00
55 # 17	-2.95E-02	-2.53E+00*
56 # 5	3.36E-02	1.16E+00
57 # 8	7.24E-04	2.22E+00
58 # 16	-2.34E-02	-3.21E+00*
59 # 4	-1.66E-03	-3.80E+00*
60 # 28	1.49E-04	9.17E+00

s, page 140 for explanation of the terms Conc and rel std dev.
Stds with neg. conc values marked with asterix (*) are are excluded from further processing. Note that nitrazepam (#30), with a closely matching spectrum is excluded.

#24 (flunitrazepam) exhibits the lowest rel std dev.

```
multicomponent analysis
relative fit error 2.064E+00
independence of standards 4.390E+04
STD        CONC         rel std dev
0 dummy    2.61E-01     6.55E-01
60 # 28    6.92E-03     7.33E-02
61 # 21    8.38E-02     1.20E-02
```

the 11th std is also processed. The relative fit error is high, as no standard matches the unknown.

```
multicomponent analysis
relative fit error 3.270E-01
independence of standard 1.273E+04
STD        CONC         rel std dev
0 dummy     1.25E-01     5.52E-01
50 # 24     9.64E-02     4.68E-02
54 # 33     2.91E-03     2.33E-01
56 # 5     -1.31E-03    -8.99E-01*
57 # 8     -1.63E-04    -5.99E+00*
60 # 28     1.49E-03     4.25E-01
61 # 21     9.27E-03     4.08E-01
```

first iteration:
6 stds are left for multicomponent analysis.

two further substances are excluded for their negativity.
Again # 24 exhibits the lowest rel std dev.

```
multicomponent analysis
relative fit error 3.241E-01
independence of standards 1.825E+04
STD        CONC         rel std dev
0 dummy    1.12E-01     6.07E-01
50 # 24    1.00E-01     2.69E-02
54 # 33    2.46E-03     2.09E-01
60 # 28    7.801E-04    3.47E-01
61 # 21    6.17E-03     3.56E-01
```

2nd iteration: only 4 stds left (+ dummy).

No negative concentration values any more.

```
multicomponent analysis
.....

1. selection:
#          rel std dev  weight
24         2.691E-02    2.169E-05
33         2.089E-01    5.074E+00
28         3.472E-01    9.295E+00
21         3.564E-01    2.995E-01
```

3rd iteration:(not shown) same values as before: iterations interrupted.
1st similarity estimate: the last rel std dev (2nd row) and the accumulated weight (similarity score) are given.

```
absorbance ratio:
 unknown/std # 24:
  mean (200-350 nm): 1.099E-01
  rel std dev: 3.764E-02

 unknown/std # 33:
  mean (200-350 nm): 6.036E+00
  rel std dev: 4.812E+00

 unknown/std # 28:
  mean (200-350 nm):-1.521E+00
  rel std dev:-1.452E+01

 unknown/std # 21:
  mean (200-350 nm): 9.527E-02
  rel std dev: 1.22E-01
```

Calculation of abs ratio:
the range of truncated spectrum is shown: 200-350nm
the mean of the 150 abs ratios and its rel std dev are calculated.

small baseline shift in a wavelength region where abs is low, may lead to negative values.

the rel std dev is used as similarity criterion.

ratio of first derivative spectra:
 unknown/std # 24:
  mean (200-350 nm): 2.013E-01
  rel std dev: 3.797E+00

 unknown/std # 33:
  mean (200-350 nm): 1.298E+00
  rel std dev: 8.901E+00

 unknown/std # 28:
  mean (200-350 nm): -1.118E-01
  rel std dev: -1.742E+01

unknown/std # 21:
  mean (200-350 nm): 4.661E-02
  rel std dev.: 5.972E+00

ratio of first derivatives essentially analogous as abs ratios.

difference of normalized absorbance spectra:
 unknown - std # 24:
  mean: 6.867E-05
  std error of mean: 2.617E-03

 unknown - std # 33:
  mean: -1.753E-03
  std error of mean: 7.394E-02

 unknown - std # 28:
  mean: 3.833E-04
  std error of mean: 4.877E-02

 unknown - std # 21
  mean: 4.253E-05
  std error of mean: 1.452E-02

The absorbance spectra are normalized to mean absorbance of 1. Then the differences of the 2 spectra are calculated. The same 150 wavelength points as above are used.
The standard error of mean was used as similarity criterion.

difference of normalized first derivative spectra:
 unknown - std # 24:
  mean: 6.867E-05
  std error of mean: 6.951E-02

 unknown - std # 33:
  mean: 1.753E-03
  std error of mean: 1.516E-01

 unknown - std # 28:
  mean: 3.833E-04
  std error of mean: 2.390E-01

 unknown - std # 21:
  mean: 4.253E-05
  std error of mean: 1.032E-01

analogous as for normalized absorbance spectra.

```
final selection (cumulative weight):  final listing ordered
FLUNITRAZEPAM          24             according to cumulative
 WEIGHT: 5.638E-10                    weight.
                                      Lowest weight is equivalent
CLONAZEPAM             21             to best similarity score.
 WEIGHT: 3.273E-04

TRIAZOLAM              33
 WEIGHT: 2.437E+00

MEDAZEPAM              28
 WEIGHT: 2.741E+01

Abbreviations: abs: absorbance; conc: concentration; dummy;
baseline spectrum; max: maximum; neg: negative rel std dev:
relative standard deviation; s: see; std: standard; λ: wavelength
#: number;
```

REFERENCES

1 B.R. Hepler, C.A. Sutheimer and I. Sunshine, Med. Toxicol., 1 (1986) 61-75.
2 H. Brandenberger, Fresenius' Z. Anal. Chem., 324 (1986) 230.
3 I. Sunshine, Handbook of Analytical Toxicology, CRC Press, Boca Raton, FL, 1969.
4 H. Schütz, Mitteilung IV der Senatskommission, VCH Verlagsgesellschaft, Weinheim, 1986.
5 T. Daldrup, F. Susanto, P. Michalke, Fresenius' Z. Anal. Chem., 308 (1981) 413-427.
6 R. Gill, B. Law, C. Brown and A. Moffat, Analyst, 110 (1985) 1059-1065.
7 A.C. Moffat, in A.C. Moffat, J.V. Jackson, M.S. Moss and B. Widdop (Editors), Clarke's Isolation and Identification of Drugs, The Pharmaceutical Press, London, 2nd ed., 1986, pp. 160-175.
8 E.H. Foerster, J. Dempsey and J.C. Garriott, J. Anal. Toxicol., 3 (1979) 87-91.
9 H. Leach and J.D. Ramsay, in A.C. Moffat, J.V. Jackson, M.S. Moss and B. Widdop (Editors), Clarke's Isolation and Identification of Drugs, The Pharmaceutical Press, London, 2nd ed., 1986, pp. 178-200.
10 R. Gill, in A.C. Moffat, J.V. Jackson, M.S. Moss and B. Widdop (Editors), Clarke's Isolation and Identification of Drugs, The Pharmaceutical Press, London, 2nd ed., 1986, pp. 201-220.
11 I. Sunshine, in Handbook of Spectrophotometric Data of Drugs, CRC Press, Boca Raton, FL, 1981.
12 M. Martens, F. Martens, P. Maenhout and A. Heyndrickx, Anal. Chem., 47 (1975) 458-461.
13 A.F. Fell, B.J. Clark and H.P. Scott, J. Chromatogr., 316 (1984) 423-440.
14 Jahresbericht des Schweizerischen Toxickologischen Informationszentrums, Zurich, 1986.
15 A.F. Fell, H.P. Scott, R. Gill and A.C. Moffat, J. Chromatogr., 282 (1983) 123-140.
16 G.E. James, UV/VIS Application Note AN 295-4, and B.G. Willis, UV/VIS Application Note AN 295-3, Hewlett-Packard.
17 H.J. Hediger, Quantitative Spektroskopie, Dr. Alfred Hüthig Verlag, Heidelberg, 1985.
18 I. Bross, Biometrics, 8 (1952) 188.
19 Documenta Geigy. Wissenschaftliche Tabellen, J.R. Geigy, Basel, 7th ed., 1968, p. 197.
20 R. Gill, S.P. Alexander and A.C. Moffat, J. Chromatogr., 247 (1982) 39-45.
21 E.I. Minder, R. Schaubhut and D.J. Vonderschmitt, Human Vet. Toxicol., in press.
22 T.A. Gossel and J.D. Bricker, Principles of Clinical Toxicology, Raven Press, New York, 1984, pp. 335-336.
23 A.H. Stead and A.C. Moffat, Human Toxicol., 3 (1983) 437-464.
24 G.T. Rasmussen and T.L. Isenhour, J. Chem. Inf. Comput. Sci., 19 (1979) 179-186.
25 A.E. Gillan, Ultraviolet and Visible Spectroscopy, Butterworths, London, 1975.
26 K. Hofstettmann, IV Schweizerisches HPLC Symposium, Bürgenstock, Sept. 24-25, 1985.