Wolfgang H. Kohn,[1] *Ph.D. and Alfons N. Jeger*[2]

Identification of Drugs by Their Near Infrared Spectra

REFERENCE: Kohn, W. H. and Jeger, A. N., **"Identification of Drugs by Their Near Infrared Spectra,"** *Journal of Forensic Sciences*, JFSCA, Vol. 37, No. 1, Jan. 1992, pp. 35–41.

ABSTRACT: Fourier transform spectrometers in the near infrared-region, equipped with optical fiber bundles and driven by fast computers, can measure and identify drugs in a few seconds. A library with 37 spectra of drugs and other powders for stretching drugs, such as saccharoses, was generated. Assessment and optimization of a reference library by mathematical means is discussed. The method is simple, fast, and reliable and can be validated.

KEYWORDS: toxicology, drug identification, spectroscopic analysis, near infrared region, Fourier transform spectrometer

Reliable identification of drugs and other chemical substances is an important task of the forensic science laboratory. Two methods are predominantly used. Thin-layer chromatography is commonly employed because of its simplicity. Comparison of mid-infrared spectra is used because of its greater selectivity, but the samples have to be mixed with potassium bromide (KBr) and pressed into pellets or a nujol mull has to be made before they can be measured. Both preparation methods are time-consuming, and often a given situation demands an answer in a short time. We therefore investigated the near infrared (NIR) region with respect to its selectivity. The mid-infrared region records vibrations of the chemical bonds; the adjacent NIR region shows the overtones and combination vibrations, especially the CH, NH, and OH stretching vibrations. Both spectral regions share a common information content. The advantage of the NIR region is its lower absorptivity, which makes sample preparation not necessary. The spectra of powders are obtained in the diffuse reflectance mode directly from the surface of the sample.

Equipment

NIR spectra were obtained within 5 s from 4800 to 10 000 cm^{-1} with a resolution of 25 cm^{-1} using a Bran + Luebbe InfraProver, a Fourier transform (FT) spectrometer. This instrument uses polarized light from a halogen lamp and the birefringent properties of a moving quartz wedge to generate an interferogram. The detector is a lead sulfide (PbS) detector. Remote sensing is made possible by a quartz optical fiber bundle with a probe head.

Received for publication 9 Oct. 1990; revised manuscript received 25 March 1991; accepted for publication 28 May 1991.

[1]Scientist, Analytical Promotion, Ciba-Geigy Ltd., Basle, Switzerland.

[2]Chemist and deputy leader, Forensic Laboratory, Police and Military Department of Basle City, Basle, Switzerland.

Materials

The substances, which are listed with their Chemical Abstracts Service (CAS) registry numbers in Table 1, were chosen to reflect everyday identification problems. The majority of the 37 substances are highly potent drugs of interest to a forensic science laboratory. The remainder are powders used potentially for stretching drugs, such as the carbohydrates lactose, mannitol, sucrose, glucose, lactose, maltose, sorbitol, and ascorbic acid.

Measurements

The substances were measured directly as available, without any pretreatment and in their original containers, using a bidirectional fiber bundle probe. The probe head was pressed firmly against the surface of the sample, illuminating a circular area with a diameter of 5 mm. The recorded spectra were an average of five scans. Measurements were obtained of five different subsamples by shaking the bottle contents between measurements.

TABLE 1—*List of the investigated substances with their CAS registry numbers.*

Substance No.	CAS Registry No.	Substance Name
1	63-42-3	lactose crystals
2	63-42-3	lactose, H_2O-free
3	69-65-8	mannitol PH
4	57-50-1	sucrose
5	50-99-7	glucose, H_2O-free
6	6363-53-7	maltose, monohydrate Fluka
7	50-70-4	sorbitol
8	10 504-35-5	ascorbic acid
9	58-08-2	caffeine, pure
10	2870-71-5	methylatropine bromide
11	51-55-8	atropine
12	51-05-8	atoxicocaine
13	299-39-8	sparteine sulfate
14	1622-62-4	flunitrazepam
15	439-14-5	diazepam
16	50-36-2	cocaine
17	76-57-3	codeine
18	561-27-3	heroin, 70%
19	357-57-3	brucine
20	57-27-27	morphine
21	76-58-4	ethylmorphine
22	469-79-4	ketobemidone
23	58-74-2	papaverine
24	57-24-9	strychnine
25	8001-45-4	hashish (cannabis)
26	13 093-77-1	amphethamine tartrate
27	5965-13-9	dihydrocodeine bitartrate
28	71-68-1	hydromorphone hydrochloride
29	51-57-0	methamphetamine hydrochloride
30	316-42-7	emetine hydrochloride
31	1095-90-5	methadone hydrochloride
32	4901-03-05	narceine hydrochloride
33	912-60-7	noscapine hydrochloride
34	50-13-5	pethidine hydrochloride
35	956-90-1	phencyclidine hydrochloride
36	850-57-7	thebaine hydrochloride
37	1152-76-7	mescaline sulfate

Spectra

Figure 1 shows four spectra of different drugs. They are typical spectra for the NIR region, showing strongly overlapped bands, relatively high absorbances at lower frequencies, and lower absorbances toward the visible part of the spectrum. There are only a few distinct peaks and areas of fine structures, for example, between 5000 and 6000 cm^{-1}. Nevertheless, there are considerable differences between the spectra.

Data Reduction

Identification was made by comparison of a sample spectrum with either a reference spectrum or a library of reference spectra. The second approach is more appropriate for the tasks of a forensic science laboratory since it does not require any previous knowledge of the samples.

With NIR spectra, there are several mathematical ways of estimating "similarity." Linear discriminant analysis, in conjunction with Mahalanobis distances, is one example [*1*]. Typically three to six wavelengths are determined for which the absorbance values allow the best discrimination between spectra of different substances in a library. The spectra are then represented in a three- to six-dimensional space as points with coordinates determined by the corresponding absorbance values. For identification purposes, the Mahalanobis distances from the test spectrum to the reference spectra are assessed.

A second approach evaluates the residual spectra after "multiplicative scatter correction" between the test spectrum and the reference spectra [*2*]. The calculation is

$$y_j = u_i + v_i x_{ij} + r_{ij}$$

where y_j is the absorption of the sample at wavelength j; x_{ij} and r_{ij} are the absorptions of the reference spectrum and residual spectrum i at wavelength j; and u_i and v_i are estimated by a least squares calculation. The similarity of the sample spectrum y and reference spectrum x_i is measured by the residual sum of squares RSS_i, which is defined as

$$RSS_i = \Sigma_j r_{ij}^2$$

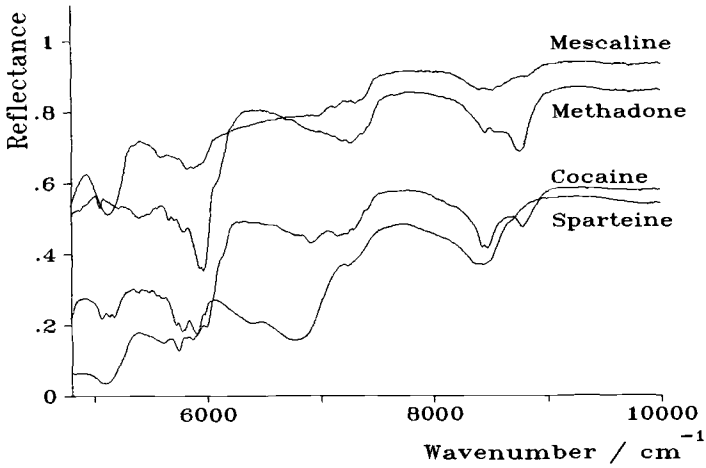

FIG. 1—*Spectra of four investigated drugs: mescaline, methadone, cocaine, and sparteine. The spectra are measured in diffuse reflectance in the region of 4800 to 10 000 cm^{-1}.*

Low or high values of RSS_i correspond to similar or dissimilar spectra, respectively.

A third approach is given by factor analysis [3,4]. It determines the principal components or factors of a library of reference spectra. A reference spectrum is the average of several spectra measured from different samples of the same chemical substance. Each reference is given by a sum of the factors, weighted with their corresponding loadings, and a residual spectrum. It can then be represented by a point in factor space, spanned by the principal components. In this case, the factor loadings are the coordinates of the reference point. The spectrum of an unknown sample is decomposed into the given factors, and the distances from its point in factor space to all the reference points are computed in a fraction of a second. Normally, the next neighbor is the correct reference.

Linear discriminant analysis uses only a very restricted part of the available information. Modifications in the sample spectra in regions not covered by the chosen wavelengths may easily be overlooked. The second approach is calculation-intensive after the measurement of an unknown sample and less selective.

These disadvantages are overcome by the third approach. The decomposition of reference spectra into their principal components is combined with striking data reduction. The first 15 principal components of the library and the corresponding loadings for the reference spectra were therefore calculated.

Afterward, the model has to be validated and perhaps optimized, as will be described in the following paragraphs. Then the system can identify an unknown sample in a few seconds.

Validation of the Library

To estimate the risk of not being able to distinguish between two substances, it is necessary to evaluate the distances of all corresponding reference points in factor space. The distances D to the next neighbor are given in Table 2. The distances alone are difficult to judge. The scatter effect due to sample presentation or grain size distribution has to be taken into account. This is why Table 2 also contains the distance to that subsample which is furthest away from the reference point. This is d_1 for the reference and d_2 for the next neighbor. To facilitate the interpretation, a relative figure, interference $I = (d_1 + d_2)/D$, has been calculated and listed. It is proportional to the risk of not distinguishing between next neighbors. $I = 1$ would mean that the spheres with radii d_1 and d_2 around the reference and next neighbor would touch; $I = 0.1$ would mean that 90% of space between the spheres is empty. The maximum interference in Table 2 is 0.137 for the pair methadone hydrochloride (No. 31) and pethidine hydrochloride (No. 34), which is largely due to the scatter of methadone subsamples. Most interferences are below 0.1.

Optimization of an Identification Model

If the scatter effect is mainly due to grain size distribution, it can be reduced by differentiating the spectra. This eliminates the horizontal displacements between spectra. In this case, differentiation causes the spheres to shrink and thus improves the selectivity of an identification model. The corresponding figures are also shown in Table 2. Note that the I values are now smaller, the maximum is now 0.028, and that the next neighbor of the differentiated spectra may change. The first derivative is not always the means of choice to reduce the interferences, because information on particle size is reduced. This is of no importance in this application.

If the interference value for a reference spectrum and its next neighbor, even after such an optimization, is not below one, then these two substances are indistinguishable with this kind of identification method and other techniques have to be applied.

TABLE 2—*Validation for the library of drugs with spectra without pretreatment and with differentiation.*[a]

No.	Substance[b]	Without Pretreatment					With Differentiation				
		d_1	NN	D	d_2	I,%	d_1	NN	D	d_2	I,%
1	lactose crystals	0.000 33	7	0.258 35	0.004 08	1.708	0.000 19	2	0.220 69	0.000 16	0.158
2	lactose, H_2O-free	0.000 71	6	0.450 04	0.000 57	0.283	0.000 16	7	0.033 15	0.000 75	2.776
3	mannitol	0.001 07	6	0.316 46	0.000 57	0.518	0.000 63	5	0.199 46	0.001 28	0.960
4	sucrose	0.003 42	7	0.589 81	0.004 08	1.271	0.002 53	7	0.550 66	0.000 75	0.595
5	glucose, H_2O-free	0.003 24	1	0.308 28	0.000 33	1.157	0.001 28	7	0.138 08	0.000 75	1.471
6	maltose	0.000 57	26	0.255 46	0.007 29	3.083	0.000 14	5	0.204 56	0.001 28	0.690
7	sorbitol	0.004 08	1	0.258 35	0.000 33	1.708	0.000 75	2	0.033 15	0.000 16	2.776
8	ascorbic acid	0.000 81	7	0.607 32	0.004 08	0.805	0.000 17	25	0.199 99	0.000 19	0.178
9	caffeine, pure	0.000 86	22	0.241 37	0.001 60	1.022	0.000 35	32	0.345 39	0.000 23	0.169
10	methylatropine	0.001 20	11	0.248 56	0.000 72	0.771	0.000 46	11	0.127 90	0.000 32	0.608
11	atropine	0.000 72	37	0.167 89	0.001 62	1.395	0.000 32	10	0.127 90	0.000 46	0.608
12	atoxicocaine	0.001 15	32	0.906 53	0.000 53	0.185	0.000 46	10	1.023 97	0.000 46	0.090
13	sparteine sulfate	0.004 34	11	0.746 43	0.000 72	0.679	0.000 88	7	0.214 95	0.000 75	0.757
14	flunitrazepam	0.000 36	15	0.070 90	0.000 73	1.538	0.000 35	22	0.532 28	0.000 48	0.157
15	diazepam	0.000 73	14	0.070 90	0.000 36	1.538	0.000 50	10	0.542 45	0.000 46	0.177
16	cocaine	0.000 55	24	0.327 03	0.000 99	0.472	0.000 29	33	0.355 55	0.000 30	0.165
17	codeine	0.000 88	35	0.957 45	0.000 54	0.148	0.000 20	8	0.988 89	0.000 17	0.037
18	heroin, 70%	0.014 82	36	0.268 46	0.005 60	7.620	0.001 86	22	0.630 12	0.000 48	0.371
19	brucine	0.000 97	35	0.161 93	0.000 54	0.934	0.000 48	35	0.766 13	0.000 33	0.105
20	morphine	0.000 61	6	0.380 86	0.000 57	0.309	0.000 25	6	0.448 08	0.000 14	0.086
21	ethylmorphine	0.002 77	1	0.560 06	0.000 33	0.553	0.002 17	1	0.431 84	0.000 19	0.545
22	ketobemidone	0.001 60	32	0.195 22	0.000 53	1.095	0.000 48	26	0.094 70	0.000 97	1.529
23	papaverine	0.001 06	32	0.314 84	0.000 53	0.506	0.001 04	32	0.628 73	0.000 23	0.203
24	strychnine	0.000 99	33	0.287 92	0.000 86	0.642	0.000 45	11	0.748 91	0.000 32	0.103
25	hashish	0.001 28	26	0.547 91	0.007 29	1.565	0.000 19	32	0.104 33	0.000 23	0.405
26	amphethamine	0.007 29	6	0.255 46	0.000 57	3.083	0.000 97	22	0.094 70	0.000 48	1.529
27	dihydrocodeine	0.006 79	30	0.369 66	0.001 62	2.272	0.001 71	30	0.223 41	0.000 67	1.065
28	hydromorphone	0.000 53	19	0.475 27	0.000 97	0.316	0.000 37	35	0.139 29	0.000 33	0.504
29	methamphetamine	0.008 90	34	0.213 47	0.000 38	4.358	0.004 99	26	0.215 19	0.000 97	2.773
30	emetine	0.001 62	37	0.194 40	0.001 62	1.670	0.000 67	37	0.125 48	0.000 76	1.138
31	methadone	0.013 49	34	0.100 60	0.000 38	13.739	0.001 31	34	0.230 51	0.000 33	0.709
32	narceine	0.000 53	37	0.168 26	0.001 62	1.282	0.000 23	25	0.104 33	0.000 19	0.405
33	noscapine	0.000 86	32	0.210 47	0.000 53	0.661	0.000 30	32	0.140 19	0.000 23	0.379
34	pethidine	0.000 38	31	0.100 60	0.013 49	13.739	0.000 33	31	0.230 51	0.001 31	0.709
35	phencyclidine	0.000 54	19	0.161 93	0.000 97	0.934	0.000 33	28	0.139 29	0.000 37	0.504
36	thebaine	0.005 60	18	0.268 46	0.014 82	7.620	0.002 56	33	0.662 54	0.000 30	0.431
37	mescaline sulfate	0.001 62	11	0.167 89	0.000 72	1.395	0.000 76	30	0.125 48	0.000 67	1.138

[a]The table lists the maximum distance d_1 between a subsample and the reference spectrum, the next neighbor NN, the distance D between the reference and the next neighbor, the maximum distance d_2 between the next neighbor and its subsamples, and the percent interference value $I = 100^* (d_1 + d_2)/D$ for all investigated substances. Differentiation of spectra reduces the sometimes dominant physical effects, such as different grain sizes or packing density of powders, and can therefore reduce the interference values. The wavelength range reaches from 4800 to 10 000 wave numbers, and the spectra are deconvoluted by 15 factors.

[b]Substances No. 28 through 36 are in hydrochloride form.

After the validation and optimization procedure have been applied on the model, the user can start to identify unknown samples.

Identification

Using NIR spectra makes the identification of chemical substances, especially of powders, an easy task; the diffuse reflectance spectrum is obtained directly by measuring the sample with the fiber optic probe in its container, without sample preparation. The spectrum is then pretreated in the same manner as the reference spectra in the library, that is, by differentiation, if useful, followed by factorial decomposition. The computation of distances between the sample-spectrum and all the references is fast and results in a list sorted by the smallest distance.

An example is given in Table 3 for sucrose. A sample of sucrose is measured, and the distances d_s to the next neighbors in the drug library are computed. The reference point of sucrose has the smallest distance, $d_s(1) = 0.001\ 27$, and is smaller than the largest scatter distance d_1 of all sucrose subsamples listed in Table 2 ($d_1 = 0.003\ 42$). The sample is therefore correctly identified. For comparison, the distances d_s to the next five neighbors are also listed in Table 2. To evaluate the reliability of an identification, the smallest distance may be compared with the second smallest distance $d_s(2)$ or the scatter distance d_1 or both. Altogether, this takes only a few seconds.

Discussion

The identification of chemical substances, a routine procedure in a forensic science laboratory, can be achieved reliably by near infrared spectroscopy. The decomposition of a reference library into principal components has two advantages, it uses the full spectral information and achieves a considerable data reduction. The reliability of the method may be improved using the first derivative of spectra, rather than the spectra themselves. As has been demonstrated, identification can even be performed with structurally related substances, as is shown here for the different carbohydrates. Remote measurement by a quartz optical fiber probe allows the identification to be made in any container within a few seconds and without sample preparation. A drawback of the equipment used presently is the need for several grams of material for an analysis.

TABLE 3—*Identification of a sucrose sample using the drug library.*[a]

Substance No.	Next Neighbor	d_s(No.)	d_s(No.)/d_s(2)	d_s(No.)/d_1
1	sucrose	0.001 27	0.002 14	0.371
2	sorbitol	0.593 41	1.000 00	173.511
3	lactose, H_2O-free	0.694 94	1.171 09	203.199
4	noscapine hydrochloride	0.719 70	1.212 82	210.439
5	caffeine, pure	0.721 24	1.215 42	210.889
6	narceine hydrochloride	0.745 17	1.255 74	217.886

[a]The sample is correctly identified. The table lists the next nearest neighbors and the distances d_s to the sample. For evaluation purposes, the distances d_s are divided by the distance to the second next neighbor $d_s(2)$ and by the maximum distance d_1 between the reference spectrum and its subsamples.

Acknowledgment

The authors gratefully acknowledge J. Griffin and H. Bartels of Ciba-Geigy Analytical Promotion for helpful discussions.

References

[1] Mark, H. L. and Tunnel, D., "Qualitative Near-Infrared Reflectance Analysis Using Mahalanobis Distances," *Analytical Chemistry*, Vol. 57, No. 7, 1985, pp. 1449–1456.

[2] Salamin, P. A., Cornélis, Y., and Bartels, H., "Identification of Chemical Substances by Their Near-Infrared Spectra," *Chemometrics and Intelligent Laboratory Systems*, Vol. 3, 1988, pp. 329–333.

[3] Gemperline, P. J., Webber, L. D., and Cox, F. O., "Raw Materials Testing Using Soft Independent Modeling of Class Analogy Analysis of Near-Infrared Reflectance Spectra," *Analytical Chemistry*, Vol. 61, No. 2, 1989, pp. 138–144.

[4] Shah, N. K. and Gemperline, P. J., "Combination of the Mahalanobis Distance and Residual Variance Pattern Recognition Techniques for Classification of Near-Infrared Reflectance Spectra," *Analytical Chemistry*, Vol. 62, No. 5, 1990, pp. 465–470.

Address requests for reprints or additional information to
Dr. Wolfgang H. Kohn
Analytical Promotion
Ciba-Geigy Ltd.
Klybeckstrasse
K-127.2.30
4002 Basle, Switzerland, CH