Organic Mass Spectrometry, 1974, Vol. 9, pp. 536 to 547. © Heyden & Son Limited. Printed in Northern Ireland

PATTERN RECOGNITION

AN APPLICATION TO THE IDENTIFICATION OF SOME STEREOISOMERIC *N*-ACETYLHEXOSAMINES BY MASS SPECTROMETRY

J. VINK and W. HEERMA
Laboratory for Analytical Chemistry, University of Utrecht, The Netherlands

J. P. KAMERLING and J. F. G. VLIEGENTHART
Laboratory for Organic Chemistry, University of Utrecht, The Netherlands

(*Received* 27 *September* 1973; *accepted* 23 *January* 1974)

Abstract—A method is reported for the identification of trimethylsilylated stereoisomeric *N*-acetylhexosamines by application of a pattern recognition procedure to mass spectral data. The selection of characteristic mass spectral data and the pattern recognition procedure are discussed.

INTRODUCTION

IN THE framework of our investigations concerning the structural analysis of biopolymers by means of modern instrumental techniques, we are interested in the possibilities of mass spectrometry for the structural determinations of carbohydrates. The study of this type of compound is hampered by the great number of parameters which define the structure of a saccharide. Previously, we reported for a number of classes of trimethylsilyl (TMS) di- and higher saccharides the applicability of electron-impact mass spectrometry to the determination of the glycosidic bond and/or the determination of the monomer sequence in case of units with different molecular weight.[1–4] Discrimination within a certain class of oligosaccharides could be performed by utilising the absence or presence of specific peaks. Whenever this criterion could not be applied, selected peak intensity ratios were used to complete the discrimination.

The study of stereoisomeric mono- and disaccharides by electron-impact mass spectrometry is still more complicated as these spectra show a great similarity. For this reason some authors concluded that mass spectrometry is insensitive to stereochemical differences, i.e. comparison of peak intensity data is an inaccurate method to distinguish stereoisomeric saccharides.[5,6] Indeed, the values of the intensity measurements can fluctuate because the production of ions in a mass spectrometer also depends upon small inadjustable ion source conditions and is therefore not reproducible to a high extent. This feature can cause misinterpretations in comparing spectra of unknown and reference compounds. However, in a study on the identification of some stereoisomeric TMS-disaccharides[7] we have shown that the use of specific peak intensity ratios is a more reliable criterion for distinguishing stereoisomeric compounds, because to a large extent the influence of the above mentioned secondary effects is eliminated. Furthermore, we demonstrated that the peak intensity ratio criterion can be applied for the discrimination between a number of TMS-aldohexoses.[8,9] Recently, the use of peak intensity ratios for the identification of stereoisomeric compounds was also reported by some other authors.[10,11] In our articles[8,9] we introduced briefly the use of computer programs based on pattern recognition techniques[12–15] for the comparison of a large number of mass spectral data of reference

monosaccharides with the unknown. We now describe in more detail the use of pattern recognition techniques for the computerised identification of trimethylsilyl-aldohexosamines.

The mass spectra of 2-acetamido-2-deoxy-tetrakis-O-(trimethylsilyl)-D-aldohexopyranoses

The mass spectra of the TMS derivatives of six 2-acetamido-2-deoxy-aldohexopyranoses, viz. α- and β-D-*N*-acetylglucosamine, α- and β-D-*N*-acetylgalactosamine, and α- and β-D-*N*-acetylmannosamine were recorded. Figure 1 shows the mean spectrum of each monosaccharide obtained by averaging seventy-five individual scans. From the different spectra in Fig. 1 it is obvious that the galactose derivatives stand definitely apart from the glucose and mannose compounds. The mass spectra of the two mannosamine anomers are almost indistinguishable.

In Table 1 the main fragment ions in the mass spectra of TMS-*N*-acetylhexosamines are reported. The origin of the fragment ions, indicated by an asterisk in Table 1, has been determined with the Barber-Elliott defocusing technique. A precursor ion of the base peak (*m/e* 173) could not be found. If the formation of this ion proceeds in a very fast reaction, no metastable transitions can be expected. We therefore assume that the *m/e* 173 ion is exclusively produced in the ion source and originates directly from the molecular ion which absence indicates a very poor stability. Furthermore, the defocusing method indicated that the *m/e* 450 ion, which is formed by elimination of an NH_2COCH_3 molecule from the molecular ion via a McLafferty rearrangement, is the precursor of the ions at *m/e* 305, 233 and 218. For the fragment ion at *m/e* 305 no specific ion structure is proposed. Its production requires rearrangement of a TMS group to an O atom or an OTMS group to a C atom before elimination of the neutral. If no rearrangements occur, the ions at *m/e* 233 and 218 can be explained by cleavage reactions of the *m/e* 450 ion resulting in C-4—C-5—O—C-1 and C-6—C-5—C-4 skeleton fragments, respectively.

The *N*-acetylgalactosamines yield more abundant fragment ions at *m/e* 233 and *m/e* 218 compared with the *N*-acetylglucosamines and mannosamines. The *m/e* 233 and *m/e* 218 ions originate from C-3—C-4 bond fissions which indicate that the splitting of the C-3—C-4 bond in the galactose compounds is favoured by the different position of the OTMS group attached to C-4 in the galactose compounds compared with the glucoses and mannoses. To relate the differences in abundances of fragment ions to differences in the stereochemical structure remains speculative as long as the origin of the atoms in fragment ions and the conformation of the ions produced under mass spectral conditions are not established.

Preprocessing of the peak intensity data

The differences in the stereochemical structure are not reflected in the abundance of each ion present in the mass spectra of stereoisomeric compounds. For identification purposes significant peak intensity data has to be selected from the available mass spectral information. The preprocessing of the data is very important in computer based pattern recognition methods.[13] The number of features, in this case peak intensity ratios, has to be reduced not only to limit computing time, but also to ensure that the number of spectra measured for each reference compound should be at least threefold the number of selected features.[13,15,16] The selection procedure was

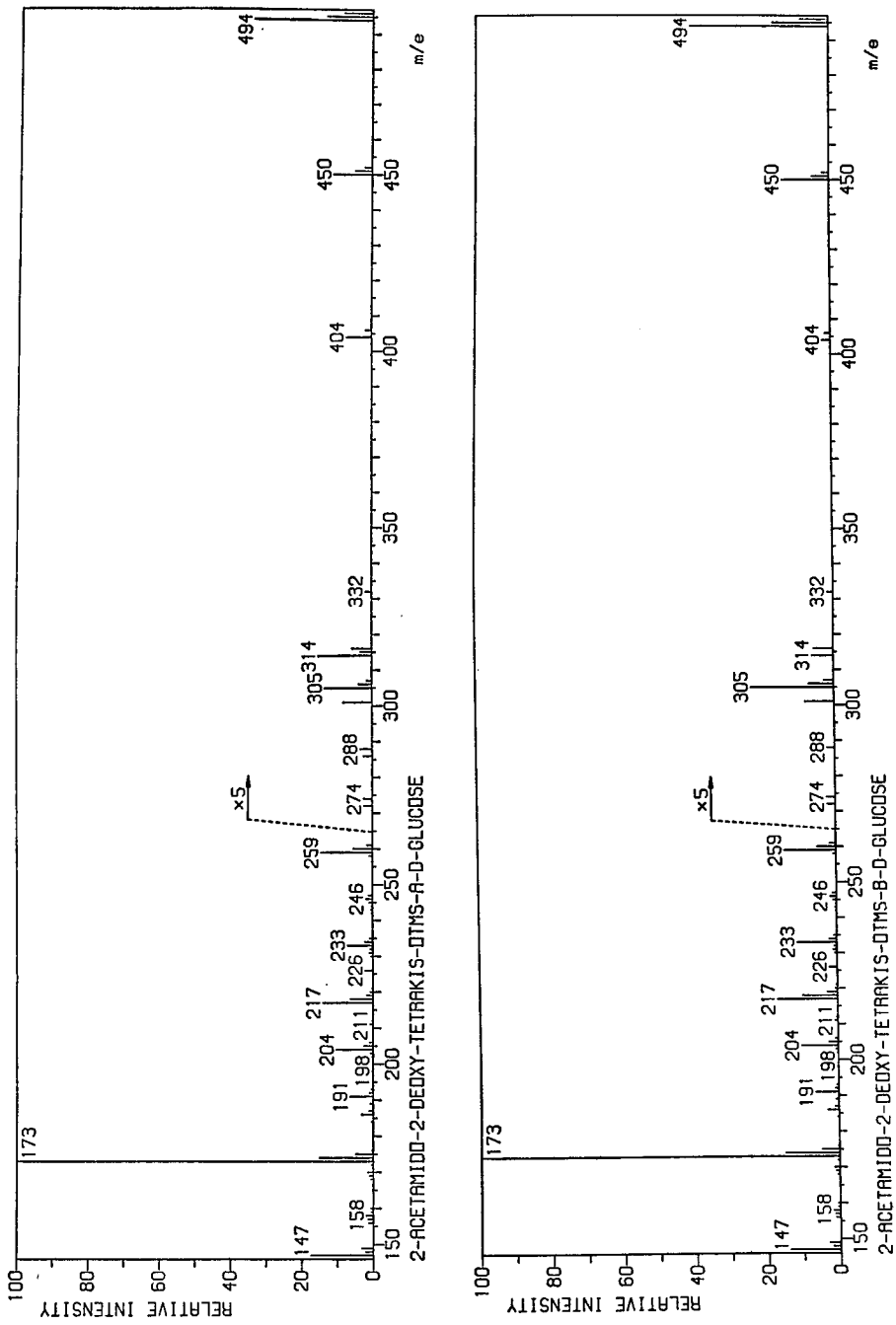

FIG. 1. The 70 eV mass spectra of six trimethylsilylated *N*-acetyl-D-hexosamines obtained by multi scan averaging of seventy-five scans. Ions beneath the *m/e* 147 ion or with abundances less than 0·3% relative to the base peak are omitted.

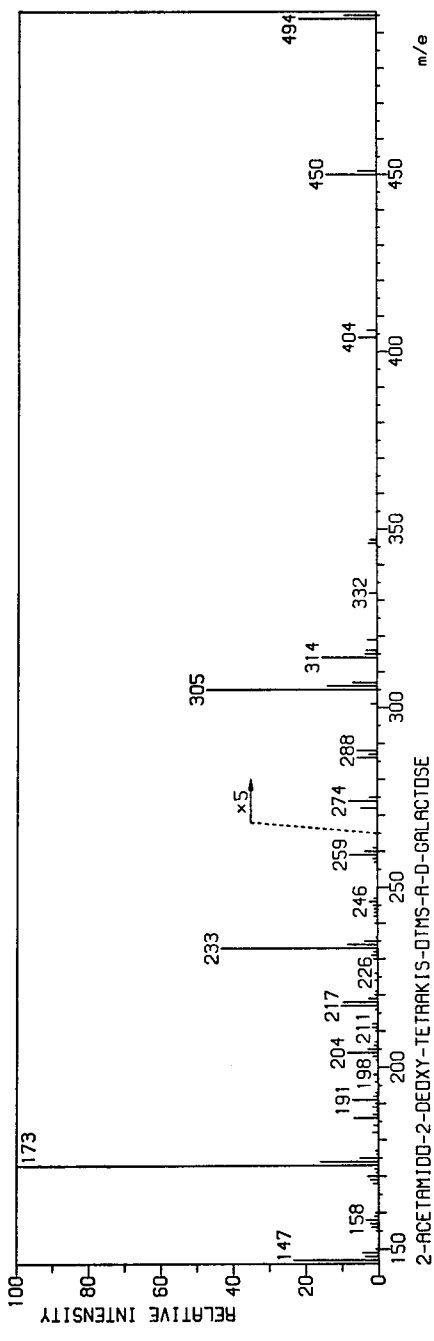
RELATIVE INTENSITY
m/e
2-ACETAMIDO-2-DEOXY-TETRAKIS-OTMS-A-D-GALACTOSE
×5
RELATIVE INTENSITY
m/e
2-ACETAMIDO-2-DEOXY-TETRAKIS-OTMS-B-D-GALACTOSE
×5

FIG. 1(b).

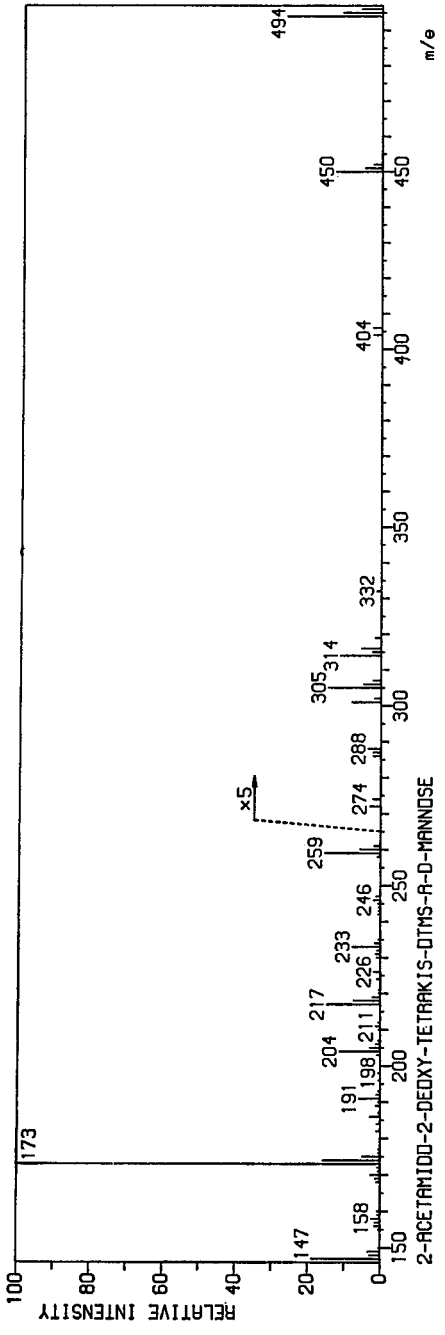
RELATIVE INTENSITY
m/e
2-ACETAMIDO-2-DEOXY-TETRAKIS-OTMS-A-D-MANNOSE
2-ACETAMIDO-2-DEOXY-TETRAKIS-OTMS-B-D-MANNOSE
×5
173
147
158
191
198
204
211
217
226
233
246
259
274
288
305
314
332
404
450
494

FIG. 1 (c).

TABLE 1. THE MAIN FRAGMENT IONS PRESENT IN THE MASS SPECTRA OF TMS-*N*-ACETYLHEXOSAMINES

m/e	Elemental composition	Fragment
494	$C_{19}H_{44}NO_6Si_4$	$[M - CH_3]^+$
450	$C_{18}H_{42}O_5Si_4$	$[M - NH_2COCH_3]^{+\cdot}$
406	not determined	$[M - CH_2OTMS]^+$
404*	$C_{16}H_{34}NO_5Si_3$	$[494 - TMSOH]^+$
362	$C_{14}H_{32}NO_4Si_3$	$[M - CH_3 - TMSOH - CH_2{=}C{=}O]^+$
360	$C_{15}H_{32}O_4Si_3$	$[M - NH_2COCH_3 - TMSOH]^{+\cdot}$
347	$C_{14}H_{31}O_4Si_3$	$[M - CH_2OTMS - NH_2COCH_3]^+$
332	$C_{14}H_{32}O_3Si_3$	$[M - NH_2COCH_3 - O{=}CH{-}OTMS]^{+\cdot}$
316*	$C_{13}H_{26}NO_4Si_2$	$[406 - TMSOH]^+$
314*	$C_{13}H_{24}NO_4Si_2$	$[404 - TMSOH]^+$
305*	$C_{12}H_{29}O_3Si_3$	originates from $[450]^{+\cdot}$.[a]
301	$C_{13}H_{27}NO_3Si_2$	$[M - O{=}CH{-}OTMS - TMSOH]^{+\cdot}$
288	$C_{12}H_{26}NO_3Si_2$	$[M - CH_2OTMS - O{=}CH{-}OTMS]^+$
286	$C_{12}H_{24}NO_3Si_2$	$[M - CH_3 - TMSOH - O{=}CH{-}OTMS]^+$
274	$C_{11}H_{24}NO_3Si_2$	$[M - CH_2OTMS - TMSOH - CH_2{=}C{=}O]^+$
272	$C_{11}H_{22}NO_3Si_2$	$[M - CH_3 - TMSOH - TMSOH - CH_2{=}C{=}O]^+$
259*	$C_{11}H_{25}NO_3Si_2$	$[301 - CH_2{=}C{=}O]^{+\cdot}$
246*	$C_{10}H_{24}NO_2Si_2$	$[288 - CH_2{=}C{=}O]^+$
233*	$C_9H_{21}O_3Si_2$	$TMSO{-}CH{=}CH{-}O^+{=}CH{-}OTMS$
226*	$C_{10}H_{16}NO_3Si$	$[316 - TMSOH]^+$
224	$C_{10}H_{14}NO_3Si$	$[M - CH_3 - TMSOH - TMSOH - TMSOH]^+$
218*	$C_9H_{22}O_2Si_2$	$TMSO{-}CH_2{-}\dot{C}H{-}CH{=}O^+TMS$
217	$C_9H_{21}O_2Si_2$	$TMSO{-}CH{=}CH{-}CH{=}O^+TMS$
211	$C_{10}H_{17}NO_2Si$	$[301 - TMSOH]^{+\cdot}$
204	$C_8H_{20}O_2Si_2$	$TMSO{-}\dot{C}H{-}CH{=}O^+TMS$
198	$C_9H_{16}NO_2Si$	$[M - OTMS - TMSOH - TMSOH - CH_2{=}C{=}O]^+$
191	$C_7H_{19}O_2Si_2$	$TMSO{-}CH{=}O^+TMS$
189	$C_7H_{17}O_2Si_2$	$(CH_3)_2Si{=}O^+{-}CH{=}CH{-}OTMS$
186	$C_6H_{16}NO_2Si$	$TMSO{-}CH{=}CH{-}CH{=}\overset{+}{N}H{-}CO{-}CH_3$
173*	$C_7H_{15}NO_2Si$	$TMSO{-}\dot{C}H{-}CH{=}\overset{+}{N}H{-}CO{-}CH_3$
158	$C_6H_{12}NO_2Si$	$(CH_3)_2Si{=}O^+{-}CH{=}CH{-}NH{-}CO{-}CH_3$
147	$C_5H_{15}OSi_2$	$TMSO^+{=}Si(CH_3)_2$

[a] In our previous investigation of 2-acetamido-2-deoxyglycose containing disaccharides we suggested that the fragment ion at *m/e* 305 in the TMS-*N*-acetylhexosamines originates from the molecular ion by elimination of a TMSO—CH=CH—OTMS molecule. However, the data reported in the present study are not in agreement with this proposal.

built up as follows: (i) Peaks with a smaller mass value than *m/e* 147 are not taken into account to avoid interference of 'background' present in the low mass range. (ii) Only peaks which are obviously not isotope peaks and occur in each spectrum with an intensity of at least 0·5% with respect to the base peak (*m/e* 173) are used. (iii) Only peak intensity ratios are used of which the corresponding peaks are recorded within a period of less than two seconds.

These three limitations result in 154 possible peak intensity ratios. Each ratio is computed for each single scan, the ratios are averaged and corresponding standard deviations are calculated. Those ratios are then selected which can distinguish the

compounds by successive dichotomies. The first selection step was the discrimination between the *N*-acetylgalactosamines and the other four stereoisomers. The next steps were made to distinguish the two anomers of galactosamine from each other, the glucosamines from the mannosamines, the two anomers of glucosamine from each other and finally the two mannosamine anomers from each other. The Fisher distance[13] was used as a measure of quality for a feature to separate two classes. For each feature the Fisher distance between two classes is given by $(m_1 - m_2)^2/(\sigma_1^2 + \sigma_2^2)$, where m_1, m_2 are the means and σ_1^2, σ_2^2 the variances of the feature for the two classes.

In Table 2 the resulting seven most characteristic peak intensity ratios, their mean

TABLE 2. THE MOST CHARACTERISTIC PEAK INTENSITY RATIOS, THEIR MEAN VALUES AND CORRESPONDING STANDARD DEVIATIONS (S.D.)[a] OF THE STEREOISOMERIC *N*-ACETYLHEXOAMINES INVESTIGATED

	Trimethylsilylated *N*-acetyl derivatives of											
	glucosamine				galactosamine				mannosamine			
	α		β		α		β		α		β	
Ratio	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
I494/I450	3·23	0·68	2·94	0·28	1·66	0·36	3·77	0·67	2·13	0·45	3·18	0·89
I404/I316	1·22	0·10	0·42	0·10	1·84	0·86	2·37	0·86	0·44	0·11	0·81	0·19
I314/I305	1·19	0·23	0·26	0·04	0·33	0·08	0·12	0·04	0·78	0·17	0·79	0·19
I259/I233	2·16	0·47	1·30	0·08	0·18	0·02	0·13	0·02	2·03	0·36	2·18	0·33
I246/I186	0·70	0·05	0·56	0·08	0·33	0·05	0·35	0·03	0·67	0·07	0·58	0·07
I226/I211	2·26	0·22	2·59	0·34	0·49	0·15	0·72	0·13	1·43	0·18	1·63	0·26
I218/I217	0·48	0·07	0·58	0·02	0·96	0·12	1·22	0·19	0·50	0·08	0·48	0·09

[a] The number of determinations for the calculation of the mean values and standard deviations amounted to 75.

values and corresponding standard deviations are shown. In this investigation the number of spectra of each reference compound measured is seventy-five and the number of features amounts to seven. Consequently, the ratio between the number of measurements and the number of features is >10, which is tolerable.[16]

Pattern recognition

Pattern recognition[12,13] can be described as the comparison of a set of data (the pattern) of the compound to be identified with equivalent patterns of reference compounds. The unknown is given the identity of that reference compound whose pattern fits best with the pattern of the unknown.

The basis of the pattern recognition procedure can, to great extent, be explained in geometric terms. A set of a number of n data $(x_1, x_2, \ldots, x_n)$ can be represented as a point in an n-dimensional space. If, for example, two features x_1 and x_2 are measured several times for two different reference compounds, each pattern (x_1, x_2) becomes a point in a 2-dimensional space. Each point can be labelled by the digit 1 or 2 depending on to which reference compound (the class) it belongs. In Fig. 2 a series of measurements are mapped onto the x_1, x_2-plane. To classify an unknown sample, a decision rule has to be derived from the set of samples (the measurements of the reference compounds) by pattern classification algorithms. This can be done by constructing a hyperplane which divides the space into regions of different classes. In Fig. 2 such a plane is indicated by g. The plane g intersects the x_1, x_2-plane in the

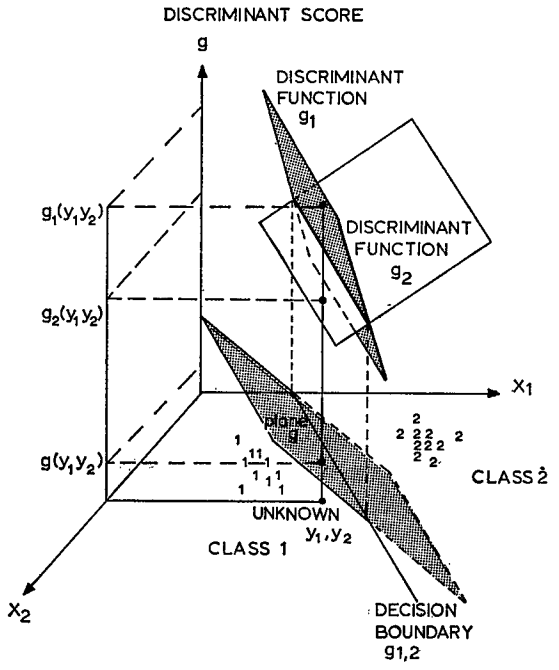

FIG. 2. Pattern recognition. The set-up of a decision rule for two classes explained in geometric terms.

line $g_{1,2}$, the decision boundary, which separates class 1 from class 2. A point (y_1, y_2) of an unknown sample in this 2-dimensional space belongs to class 1 if the point (y_1, y_2) lies at the side of class 1 of the decision boundary $g_{1,2}$ (mathematically expressed as $g(y_1, y_2) > 0$), or belongs to class 2 if the point (y_1, y_2) is at the other side of the line $g_{1,2}$ $(g(y_1, y_2) < 0)$. A point on the decision boundary $(g(y_1, y_2) = 0)$ must be classified arbitrarily. For instance, the unknown pattern (y_1, y_2), indicated by a dot in Fig. 2, will be classified as a member of class 1.

In the case of two classes one decision boundary can be designed, N classes require $N(N-1)/2$ decision boundaries. Working with a large number of classes gives rise to a vast amount of decision steps. This can be avoided by defining a number of N discriminant functions, one for each class: $g_1, g_2, \ldots, g_N$ (N is the total number of classes). In Fig. 2 the discriminant functions, represented by the surfaces g_1 and g_2, are shown for class 1 and 2, respectively. The hyperplane g, and the decision surfaces g_1 and g_2, are related by $g = g_1 - g_2$ (1) in the 2-dimensional case. (In Fig. 2 the same arbitrary constant was added to each of the discriminant functions without altering the following conclusions.) The decision boundary $g_{1,2}$ is given by the projection in the x_1, x_2-plane of the intersection of the discriminant functions g_1 and g_2. The decision rule that an unknown point (y_1, y_2) belongs to class 1 because it is at the class 1 side of the decision boundary $g_{1,2}$ $(g(y_1, y_2) > 0)$, can be transformed in the expression that the discriminant function g_1 has the largest value $(g_1(y_1, y_2) > g_2(y_1, y_2))$. Thus, the unknown pattern (y_1, y_2) in Fig. 2 is still classified as belonging to class 1 because $g_1(y_1, y_2) > g_2(y_1, y_2)$. In general, an unknown sample point $(x_1, x_2, \ldots, x_n)$ is considered to be a member of class i if $g_i\ (x_1, x_2, \ldots, x_n) > g_j\ (x_1, x_2, \ldots, x_n)$

for $i, j = 1, 2, \ldots, N$ and $i \neq j$. Instead of deciding for each of the $N(N-1)/2$ boundaries for N classes at which side of the decision boundaries an unknown sample point has been placed, the largest value from a number of N discriminant functions has to be selected.

The basic principles as outlined above still hold in spaces of more than three dimensions. For the determination of the decision boundaries or discriminant functions from a large number of measurements of different features and many classes, the use of computer programs is essential to perform efficient calculations.

The identification procedure for an unknown compound can be summarised as follows: based on measurements of reference compounds a discriminant function for each reference compound is estimated by a computer program; the values of equivalent measurements of the unknown are substituted in the discriminant functions and the measurements of the unknown are classified as belonging to measurements of that reference compound for which the discriminant function yields the largest value.

Discriminant functions

The discriminant functions used in this study are linear functions, computed according to a parametric pattern recognition method, which is an optimum classifier for normal distributions of the feature values and assumes equal covariance matrices for each distribution.[12] For each reference compound i a linear discriminant function g_i is derived

$$g_i(X) = X^t \Sigma^{-1} M_i + \log p_i - \tfrac{1}{2} M_i^t \Sigma^{-1} M_i \qquad (2)$$

where X and M_i are the column vector notations for the pattern of the unknown and the mean M for class i, respectively, Σ^{-1} is the inverse of the covariance matrix and p_i is the *a priori* probability of class i. The character t (transpose) is attached to some parameters in Eqn. (2) to indicate that the vector is represented as a row, and not as a column vector. Eqn. (2) can be simplified as follows

$$g_i(x_1, x_2, \ldots, x_n) = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + w_{n+1} \qquad (3)$$

with $w_1, w_2, \ldots, w_{n+1}$ as weight factors, and $x_1, x_2, \ldots, x_n$ the components of the pattern of the unknown, viz. the peak intensity ratios.

In this investigation eight weight factors $w_1, w_2, \ldots, w_8$ are calculated according to Eqn. (2) by use of the values of the most characteristic peak intensity ratios (see Table 2) for each single scan of each hexosamine derivative. To compute the weight factor w_8 an equal *a priori* probability of each sugar is supposed (p_i in Eqn. (2) is given the value $\frac{1}{6}$). The peak intensity data of seventy-five mass spectra per reference compound are used to compute the weight factors of the linear discriminant functions. In Table 3 the weight factors of the linear discriminant functions are shown for each trimethylsilylated *N*-acetylhexosamine investigated.

RESULTS AND DISCUSSION

To determine the reliability of the classification method, the data of 450 mass spectra which are used as a learning set to design the decision rules are also used as test set. For all stereoisomers except the *N*-acetylmannosamines only correct classifications are found. Eight spectra of TMS-*N*-acetyl-α-D-mannosamine are classified as the β-form and six spectra of the β-D-mannosamine derivative are classified as the

TABLE 3. THE WEIGHT FACTORS $w_1, w_2, \ldots, w_8$ IN THE LINEAR DISCRIMINANT FUNCTIONS COMPUTED FOR THE SIX STEREOISOMERIC AMINOSUGARS INVESTIGATED

Compound	Weight factors in the linear discriminant functions							
	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
α-D-glucosamine	11·557	10·544	117·049	73·226	190·663	61·417	125·520	−339·843
β-D-glucosamine	16·812	5·481	33·064	45·745	157·186	61·713	114·006	−217·153
α-D-galactosamine	10·734	4·987	36·663	23·878	91·519	19·509	123·496	−100·791
β-D-galactosamine	23·563	6·793	7·395	26·967	102·843	25·528	158·155	−178·576
α-D-mannosamine	9·632	4·692	85·905	61·280	186·703	40·543	112·726	−226·665
β-D-mannosamine	14·486	7·079	81·852	65·607	163·268	45·756	116·302	−242·030

α-form. The *N*-acetylmannosamines are confused with each other because of the nearly similar mass spectra from which significant features could hardly be derived to distinguish the α- from the β-compounds.

For a further check of the classification method the mass spectra of a newly prepared series of derivatives are recorded by another operator with the same instrument and under the same experimental conditions. This test set was recorded in a period of three to six months after the recording of the learning set. The results of the classification of the spectra in the test set with the decision rules derived by the learning set are presented in Table 4. In the column 'no classification possible' the number of times is

TABLE 4. THE RESULTS OF THE IDENTIFICATION OF A NUMBER OF MASS SPECTRA NOT INCLUDED IN THE LEARNING SET

TMS-*N*-acetyl derivative	Number of spectra recorded	Number of misclassifications	Misclassified as	No classification possible
α-D-glucosamine	73	0	—	9
β-D-glucosamine	59	4	α-D-mannosamine	12
α-D-galactosamine	53	4	β-D-galactosamine	0
β-D-galactosamine	28	1	α-D-galactosamine	0
α-D-mannosamine	46	2	β-D-mannosamine	8
β-D-mannosamine	57	25	α-D-mannosamine	23

indicated that the two largest discriminant scores differ by less than 1% of the largest score. This means that the decision boundary $g_{1,2}$ between two classes in Fig. 2 is extended with a region wherein no decision is taken. The 1% limit used in this investigation is of practical use for the linear discriminant functions derived. Theoretical consideration to determine the exact 'no classification possible' region would go beyond the scope of this report. The results shown in Table 4 indicate that for the *N*-acetylhexosamines except for the β-mannosamine derivative the probability of misclassification for one single scan is less than 8%. TMS-*N*-acetyl-β-D-mannosamine is incorrectly identified as the α-D-mannosamine derivative in nearly one of two cases. On the basis of the experience with the learning set used as test set, confusion of the two *N*-acetylmannosamine anomers could be expected. The rather good score in the identification of TMS-α-D-mannosamine points to a difference between the test and learning set. A TMS-*N*-acetylhexosamine can be classified with high probability as a *N*-acetylmannosamine if the two largest discriminant scores of

nearly equal magnitude are obtained for the two mannosamine anomers, and a difference of more than 1 % of the largest score is present between the *N*-acetylmannosamines and the other stereoisomers. Determination of the position of the anomeric OTMS group for the mannosamines is doubtful by this mass spectral identification method.

The misclassifications mentioned in this study are found with our learning and test set, recorded with our instrument under our experimental conditions. In order to ensure representative learning and test sets, it is desirable that all measurements are performed under identical conditions. This indicates that every investigator has to build up his own learning set. In previous studies[8,9] we found that for the TMS-aldohexoses correct identifications could also be obtained with spectra recorded with mass spectrometers of different design. However, because the identification method is based on comparison, compounds which are not included in the set of reference compounds can never be identified correctly.

The simplicity of the method demands complete automation. An example of an automated *N*-acetylhexosamine analysis is presented in Fig. 3. This identification has been performed on-line according to the decision rules within less than a second. In Fig. 3 the α-D-glucosamine derivative has obtained the highest discriminant score

ANAL CHEM LAB 03-08-73 TEST HEXOSAMINE SYSTEM

IDENTIFICATION OF MONOSACCHARIDES

HEXOSAMINE DERIV.	DISCRIMINANT SCORE
A-D-GLU	360.591
B-D-MAN	335.862
A-D-MAN	335.676
B-D-GLU	298.520
A-D-GAL	203.354
B-D-GAL	169.694

FIG. 3. An example of the automated identification of 1 μg TMS-*N*-acetyl-α-D-glucosamine.

and is most likely to match with the α-D-glucosamine data included in the learning set.

The identification method as described in this investigation, is very useful for the identification of strongly related compounds if normally no identification can be performed due to nearly identical mass spectra. Moreover, the principles developed in this investigation can be applied to each class of related compounds.

EXPERIMENTAL

The mass spectra were recorded with an AEI MS-902 mass spectrometer connected on-line with a Ferranti Argus 500 computer. The mass spectrometer-computer system, the peripherals and the features of the system are described elsewhere.[17] The samples were introduced into the mass spectrometer by the direct insertion probe. The mass spectral conditions were: ion source temperature 100 to 120 °C, electron energy 70 eV, trap current 500 μA, accelerating voltage 8 kV. Recording of the mass spectra was performed under low resolving power conditions at a resolving power of 1000 (10% valley), a scan speed of 8 s/dec. and a bandwidth of 1000 Hz. The data acquisition took place with a sample interval of 100 μs and 131072 samples. The rejection voltage was 100 mV. The peak intensities are normalised to the total ion current monitor signal.[17]

To determine the elemental compositions shown in Table 1, mass spectra under high resolving power conditions were recorded under the following conditions: resolving power 12500 (10% valley), scan speed 16 s/dec., bandwidth 2000 Hz, sample interval 50 μs, 524288 samples and rejection voltage 50 mV.

The metastable transitions have been measured using the Barber-Elliott defocusing technique. The defocusing device used did not allow the establishment of a metastable reaction $m_1^+ \rightarrow m_2^+ + m_3$, in which $m_1/m_2 > 4$.

N-Acetyl-D-glucosamine was purchased from Serva; *N*-acetyl-D-galactosamine and *N*-acetyl-D-mannosamine were obtained from Pierce Chemical Company. After anomerisation in H_2O for 24 h at room temperature and lyophilisation, the aminosugars were silylated as described previously.[2] By preparative gas-liquid chromatography the α- and β-pyranose ring forms were separated from each other and from furanose ring forms sometimes present (Pye 105 gas chromatograph, 10% OV-17 on Chromosorb W AW-DMCS, 30 to 60 mesh, column temperature 230 °C, column length 175 cm). All isolated TMS derivatives were crystalline compounds. The existence of these compounds in the pyranose ring form ($^4C_1$ chair conformation) was established by ^{1}H n.m.r. spectroscopy.[18] This technique was also used for differentiation between the α- and β-forms of glucose and galactose, respectively, while the α- and β-forms of mannose were distinguished on the basis of optical rotation experiments.

To get a learning set which is a representative sample of the infinite population, each derivative was prepared at least three times and the mass spectra were recorded on three different days.

Acknowledgement—The authors are indebted to Mrs L. R. Smaling-Hakkert and Dr D. G. Streefkerk for sample preparations and purifications, and to Dr. W. Dekker, Dr H. A. van 't Klooster, Mr A. J. de Meijer, Dr P. C. Sander and Mrs J. S. Vaarkamp-Lijnse for computer programming. The authors also acknowledge the work of Mr F. van Geel and Mr G. van de Werken performed as a part of the study for their master's degree. Ir R. P. W. Duin is thanked for the valuable discussions in the field of pattern recognition. This work was supported in part by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for the Advancement of Pure Research (ZWO).

REFERENCES

1. J. P. Kamerling, J. F. G. Vliegenthart, J. Vink and J. J. de Ridder, *Tetrahedron Letters* 2367 (1971).
2. J. P. Kamerling, J. F. G. Vliegenthart, J. Vink and J. J. de Ridder, *Tetrahedron* **27**, 4275 (1971).
3. J. P. Kamerling, J. F. G. Vliegenthart, J. Vink and J. J. de Ridder, *Tetrahedron* **27**, 4749 (1971).
4. J. P. Kamerling, J. F. G. Vliegenthart, J. Vink and J. J. de Ridder, *Tetrahedron* **28**, 4375 (1972).
5. D. C. DeJongh, T. Radford, J. D. Hribar, S. Hanessian, M. Bieber, G. Dawson and C. C. Sweeley, *J. Amer. Chem. Soc.* **91**, 1728 (1969).
6. N. K. Kochetkov, O. S. Chizhov and N. V. Molodtsov, *Tetrahedron* **24**, 5587 (1968).
7. J. Vink, J. J. de Ridder, J. P. Kamerling and J. F. G. Vliegenthart, *Biochem. Biophys. Res. Commun.* **42**, 1050 (1971).
8. J. Vink, J. H. W. Bruins Slot, J. J. de Ridder, J. P. Kamerling and J. F. G. Vliegenthart, *J. Amer. Chem. Soc* **94**, 2542 (1972)
9. J. Vink, P. H. Melis and W. Heerma, in A. Frigerio (Ed.), *Proceedings of the International Symposium on Gas Chromatography-Mass Spectrometry*, Tamburini, Milano, 1972, p. 289.
10. V. Kováčik, C. Peciar, Š. Bauer and H. F. Grützmacher, *Carbohydrate Res.* **19**, 169 (1971).
11. R. Bentley, in E. E. Snell (Ed.), *Annual Review of Biochemistry*, Volume 41, Annual Reviews Inc, Palo Alto, 1972, p. 953.
12. N. J. Nilsson, *Learning Machines*, McGraw-Hill, New York, 1965.
13. W. S. Meisel, *Computer-Oriented Approaches to Pattern Recognition*, Academic Press, New York, 1972.
14. T. L. Isenhour and P. C. Jurs, *Anal. Chem.* **43**, 20A (1971) and references cited therein.
15. B. R. Kowalski and C. F. Bender, *J. Amer. Chem. Soc.* **94**, 5632 (1972).
16. J. W. Sammon, D. Foley and A. Proctor, *IEEE Symposium on Adaptive Processes*, Austin, Texas, December 1970, p. IX.2.1.
17. H. A. van 't Klooster, J. S. Vaarkamp-Lijnse and G. Dijkstra, *Org. Mass Spectrom.* **8**, 303 (1974).
18. D. G. Streefkerk, M. J. A. de Bie and J. F. G. Vliegenthart, *Carbohydrate Res.* in press.