Multivariate Derivation of Descriptive Scales for Monosaccharides

Lennart Eriksson, Jörgen Jonsson, Michael Sjöström*, Conny Wikström and Syante Wold

Research Group for Chemometrics, Department of Organic Chemistry, Umeå University, S-901 87 Umeå, Sweden

> Eriksson, L., Jonsson, J., Sjöström, M., Wikström, C. and Wold, S., 1988. Multivariate Derivation of Descriptive Scales for Monosaccharides. - Acta Chem. Scand., Ser. B 42: 504-514.

> Thirteen TLC systems were used to characterize sixteen monosaccharides. Principal components analysis (PCA) of the 16×13 data matrix gave four principal components that together explained 97.8% of the variance. The first PC described 90 % of the variance and was tentatively interpreted as a hydrophobicity scale. The data for the eight aldohexoses were further analyzed by performing a separate PCA. This resulted in a two-component model explaining 95.5 % of the variance. This model relates to how the three hydroxy groups on each aldohexose are located (either equatorially or axially).

General considerations. Carbohydrates such as oligo- and polysaccharides are of current interest both in mechanistic studies and in the pharmaceutical industry. Recent advances in the biochemistry of these compounds have led to the creation of a number of carbohydrate-based drugs. The development of such drugs requires an optimization of the carbohydrate structure with regard to the corresponding biological activity. To perform and facilitate this optimization it is essential that quantitative structure-activity relationships (QSAR's) are established, which mathematically relate structural descriptors to biological response data.

The parametrization of oligo- and polysaccharide structures is often difficult because of the complexity and great variety of such compounds. One way to accomplish the structural description in a QSAR is to parametrize each single carbohydrate unit, i.e. the monosaccharides. The characteristics of each monosaccharide are then used to parametrize higher saccharides, by arranging the derived parameters according to the monosaccharide sequence.

modelling their biological activities.

An analogous problem has recently been dealt

with in our laboratory. With some success, it has

been demonstrated that a multivariate parametrization of amino acids can be used to con-

struct QSAR's for certain peptide families. 1-3 The variations in amino acid sequence within the se-

ries of peptide analogues were described by three

variables per amino acid position. The characterizing variables (or principal properties) were ex-

tracted from multi-property matrices with several

measurements for the amino acids. With the con-

structed QSAR's it was also possible to predict

descriptive scales (principal properties) for

monosaccharides, which hopefully can be used in future QSAR's for oligo- and polysaccharides

This work is the first step towards analogous

biological activities for new peptide analogues.

Selection of data. An experimental method used

to characterize small biomolecules such as amino acids or monosaccharides should preferably be rapid and give a range of stable, reproducible data. The choice of a chromatographic technique seems appropriate in this case, since this methodology is well established and gives rapid results. Chromatographic procedures have been used

^{*}To whom correspondence should be addressed.

Table 1. A two-level factorial design in three variables for the composition of TLC eluents. Acetone (A), 1-butanol (B) and water (C) were mixed in different proportions according to the design. For acetone, high level (+) was 8 volume parts and low level (-) 2 volume parts. The corresponding values for the other solvents were 4 and 1, respectively. The randomized experimental order was 7, 3, 4, 1, 6, 5, 2 and 8.

Eluent No.	Α	В	С	Composition
1	+	+	+	8:4:4
2	_	+	+	2:4:4
2 3 4 ^a 5	+	_	+	8:1:4
4 ^a	_	-	+	2:1:4
5	+	+	_	8:4:1
6	_	+	_	2:4:1
7	+	-	_	8:1:1
8	_	_	-	2:1:1

^aEluent 4 was unusable because of too high a water content. See also text.

earlier to identify and characterize carbohydrates in a variety of ways. Clark⁴ used paper chromatography to identify simple sugars in the body fluids of marine worms. Churms⁵ reported many applications of GC, HPLC and TLC to this subject. The use of pyrolysis gas chromatography (PGC) and cluster analysis to characterize carbohydrate structures was reported by Morgan and Jacques.⁶

Thin layer chromatography has been used to

parametrize amino acids.^{3,7} Thus, it appears reasonable that TLC also could be used to give a good parametrization for monosaccharides. The method produces reproducible multivariate data that are obtainable under stable experimental conditions. Moreover, it is easy to simultaneously change both stationary phases and eluents to further increase the range of the TLC data.

For this study we measured R_f values for sixteen sugars in thirteen TLC systems. The monosaccharides were eight aldohexoses, four substituted aldohexoses and four aldopentoses.

Materials and methods

Monosaccharides. The highest purity grades of sixteen monosaccharides commercially available from Sigma were used without further purification. The sugars were dissolved in distilled water to give standard solutions of 20 mM. These solutions were stored in a refrigerator, but were used at room temperature.

Thin layer plates. The pre-coated 200×200 mm plates were all obtained from Merck. The stationary phases were silica gel 60 F-254 (with 0.5 mm layer thickness) and cellulose (0.1 mm).

Eluents. With the aim of constructing a basis for the selection of appropriate eluents, we performed a principal components analysis^{8,9} on literature data. ^{5,10,11} Guided by the resulting loading plots we selected mobile phases as different as

Table 2. Variable description. All proportions for the mobile phases are given as volume parts. The stationary phases were silica gel (variables 1–4 and 7–13) and cellulose (5 and 6).

Var. No.	Composi	tion		Proportions		
1	1-Propan	iol/wa	ter	85:15		
2	Butanone	e/ace	ic acid/methanol	60:20:20		
3	Acetone/	chlore	oform/water/ethanol/2-Propanol	50:6:25:25:10		
4			one/formic acid/water	60:20:10:20		
5	Ethyl ace	etate/i	pyridine/water	40:30:40		
6	1-Propan			85:15		
7	"design"			see Table 1		
8	"	"	3	11		
9	"	"	1	**		
10	"	"	6	99		
11	"	"	5	11		
12	"	"	2	"		
13	,,	17	8	17		

Table 3. TLC migration values for sixteen monosaccharides.

No.	Sugar	Variable No.ª												
		1	2	3	4	5	6	7	8	9	10	11	12	13
1	Xylose	66	59	86	50	65	28	59	96	84	60	57	60	82
2	Lyxose	64	58	86	50	66	33	60	92	83	61	60	62	81
3	Arabinose	52	51	77	42	63	26	46	86	75	48	42	57	76
4	Ribose	58	51	81	48	68	35	56	88	80	57	51	61	79
5	Fucose	61	58	84	50	66	36	55	91	82	60	52	61	79
6	Rhamnose	73	69	89	62	70	45	71	94	89	71	73	68	83
7	N-acetylgalactosamine	55	45	82	40	65	33	47	90	80	52	44	55	75
8	N-acetylglucosamine	61	50	85	45	66	31	51	92	83	58	51	57	78
9	Allose	48	42	72	39	60	21	35	80	70	42	32	52	70
10	Altrose	59	50	83	46	64	30	51	89	80	55	51	56	75
11	Galactose	42	41	68	37	56	17	26	79	67	36	25	48	69
12	Glucose	50	45	76	40	59	20	33	84	74	43	32	53	72
13	Gulose	52	44	78	41	61	24	39	84	75	46	36	52	72
14	Mannose	55	45	80	43	61	26	37	86	77	49	37	53	74
15	Talose	49	43	76	42	64	30	40	84	73	47	38	52	71
16	Idose	61	51	85	48	66	35	54	91	82	59	55	56	76

^a For explanation of the variables, see Table 2.

possible, since we are interested in optimal characterization rather than maximum separation or resolution.

We also used other eluents, but they had compositions determined by a two-level factorial design, ¹² (see Table 1). According to the design, acetone, 1-butanol and water were mixed in different proportions, which resulted in eight additional eluents. One of these mobile phases was, however, unusable as a result of too high a water content (too much water prevents good migration through the stationary phase). The experimental order of the remaining "design"-eluents was randomized to diminish possible systematic experimental errors (Table 1).

In all, a total of thirteen eluents were used. Six of them originated from the PCA of the literature data and seven were mixed according to the statistical design. For a complete listing of the mobile phases and their compositions, see Table 2.

Almost all the solvents used were commercially available from Merck. Formic acid and pyridine were obtained from Reidel-de-Haen, and chloroform from May & Baker.

Chromatographic tank. The chromatographic tank used was a sandwich chamber. This is a very thin chamber which can accommodate a single

200×200 mm plate and a 10 ml solvent volume for ascending chromatography. In comparison with rectangular tanks, the S-chamber has some

Table 4. PC scores for sixteen monosaccharides. The first PC (t_1) explains 89.9 % of the variance. The other components (t_2-t_4) explain 2.7, 3.7 and 1.5 % of the variance, respectively.

4 Ribose 1.76 0.52 0.80 0.7 5 Fucose 2.60 0.36 0.03 0.0 6 Rhamnose 7.21 1.26 0.12 -0.5 7 N-acetyl- galactosamine -0.18 -1.07 0.76 0.3 8 N-acetyl- glucosamine 1.60 -1.60 0.02 0.1 9 Allose -4.32 0.64 0.12 0.0 10 Altrose 0.62 -0.37 0.06 -0.3 11 Galactose -6.37 0.72 -0.50 -0.1 12 Glucose -3.43 -0.02 -0.72 -0.1 13 Gulose -2.60 -0.23 -0.02 -0.2	No.	Sugar	<i>t</i> ₁	<i>t</i> ₂	<i>t</i> ₃	<i>t</i> ₄
2 Lyxose 3.32 0.11 -0.49 0.2 3 Arabinose -1.14 0.53 -0.25 0.7 4 Ribose 1.76 0.52 0.80 0.7 5 Fucose 2.60 0.36 0.03 0.0 6 Rhamnose 7.21 1.26 0.12 -0.5 7 N-acetyl- galactosamine -0.18 -1.07 0.76 0.3 8 N-acetyl- glucosamine 1.60 -1.60 0.02 0.1 9 Allose -4.32 0.64 0.12 0.0 10 Altrose 0.62 -0.37 0.06 -0.3 11 Galactose -6.37 0.72 -0.50 -0.1 12 Glucose -3.43 -0.02 -0.72 -0.1 13 Gulose -2.60 -0.23 -0.02 -0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	_	Vidaa	2.07	0.52	1 42	0.26
3 Arabinose						
4 Ribose 1.76 0.52 0.80 0.7 5 Fucose 2.60 0.36 0.03 0.0 6 Rhamnose 7.21 1.26 0.12 -0.5 7 N-acetyl- galactosamine -0.18 -1.07 0.76 0.3 8 N-acetyl- glucosamine 1.60 -1.60 0.02 0.1 9 Allose -4.32 0.64 0.12 0.0 10 Altrose 0.62 -0.37 0.06 -0.3 11 Galactose -6.37 0.72 -0.50 -0.1 12 Glucose -3.43 -0.02 -0.72 -0.1 13 Gulose -2.60 -0.23 -0.02 -0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0		•		• • • • •	-	
5 Fucose 2.60 0.36 0.03 0.0 6 Rhamnose 7.21 1.26 0.12 -0.5 7 N-acetyl- galactosamine -0.18 -1.07 0.76 0.3 8 N-acetyl- glucosamine 1.60 -1.60 0.02 0.1 9 Allose -4.32 0.64 0.12 0.0 10 Altrose 0.62 -0.37 0.06 -0.3 11 Galactose -6.37 0.72 -0.50 -0.1 12 Glucose -3.43 -0.02 -0.72 -0.1 13 Gulose -2.60 -0.23 -0.02 -0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	3	Arabinose	-1.14			0.70
6 Rhamnose 7.21 1.26 0.12 -0.5 7 N-acetyl- galactosamine -0.18 -1.07 0.76 0.3 8 N-acetyl- glucosamine 1.60 -1.60 0.02 0.1 9 Allose -4.32 0.64 0.12 0.0 10 Altrose 0.62 -0.37 0.06 -0.3 11 Galactose -6.37 0.72 -0.50 -0.1 12 Glucose -3.43 -0.02 -0.72 -0.1 13 Gulose -2.60 -0.23 -0.02 -0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	4	Ribose	1.76	0.52	0.80	0.72
7 N-acetyl- galactosamine	5	Fucose	2.60	0.36	0.03	0.04
galactosamine -0.18 -1.07 0.76 0.3 8 N-acetyl-glucosamine 1.60 -1.60 0.02 0.1 9 Allose -4.32 0.64 0.12 0.0 10 Altrose 0.62 -0.37 0.06 -0.3 0.1 11 Galactose -6.37 0.72 -0.50 -0.1 0.1 12 Glucose -3.43 -0.02 -0.72 -0.1 0.7 13 Gulose -2.60 -0.23 -0.02 -0.2 0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	6	Rhamnose	7.21	1.26	0.12	-0.57
8 N-acetyl-glucosamine 1.60 -1.60 0.02 0.1 9 Allose -4.32 0.64 0.12 0.0 10 Altrose 0.62 -0.37 0.06 -0.3 11 Galactose -6.37 0.72 -0.50 -0.1 12 Glucose -3.43 -0.02 -0.72 -0.1 13 Gulose -2.60 -0.23 -0.02 -0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	7	N-acetyl-				
glucosamine 1.60 -1.		galactosamine	-0.18	-1.07	0.76	0.35
9 Allose	8	N-acetyl-				
10 Altrose 0.62 -0.37		glucosamine	1.60	-1.60	0.02	0.14
11 Galactose -6.37 0.72 -0.50 -0.1 12 Glucose -3.43 -0.02 -0.72 -0.1 13 Gulose -2.60 -0.23 -0.02 -0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	9	Allose	-4.32	0.64	0.12	0.01
12 Glucose -3.43 -0.02 -0.72 -0.1 13 Gulose -2.60 -0.23 -0.02 -0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	10	Altrose	0.62	-0.37	0.06	-0.37
13 Gulose -2.60 -0.23 -0.02 -0.2 14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	11	Galactose	-6.37	0.72	-0.50	-0.14
14 Mannose -1.75 -0.46 -0.24 -0.4 15 Talose -2.43 0.14 1.18 -0.0	12	Glucose	-3.43	-0.02	-0.72	-0.11
15 Talose -2.43 0.14 1.18 -0.0	13	Gulose	-2.60	-0.23	-0.02	-0.28
	14	Mannose	-1.75	~0.46	-0.24	-0.43
16 Idose 1.82 -0.60 0.54 -0.4	15	Talose	-2.43	0.14	1.18	-0.07
** *****	16	Idose	1.82	-0.60	0.54	-0.49

Var. No.	Weights	Means	<i>p</i> ₁	p ₂	p ₃	<i>p</i> ₄
1	0.128	7.26	0.284	-0.102	-0.183	-0.280
2	0.131	6.55	0.270	0.403	-0.357	-0.060
3	0.176	14.20	0.276	-0.437	-0.029	-0.205
4	0.162	7.33	0.273	0.408	-0.065	-0.480
5	0.279	17.77	0.269	-0.040	0.529	0.360
6	0.142	4.17	0.262	0.089	0.628	-0.148
7	0.084	4.01	0.286	0.071	0.064	0.207
8	0.205	18.04	0.270	-0.474	-0.219	0.146
9	0.174	13.66	0.283	-0.291	-0.043	-0.138
10	0.122	5.89	0.289	-0.053	0.073	-0.136
11	0.080	3.71	0.287	0.042	0.019	-0.112
12	0.200	11.27	0.276	0.376	-0.074	0.350
13	0.232	17.56	0.277	0.020	-0.309	0.508

Table 5. Weights, means and loadings for the four-dimensional PC model obtained for sixteen monosaccharides.

distinct advantages. It gives better reproducibility and faster gas-phase equilibration.¹³

TLC procedure. The 20 mM monosaccharide samples were spotted in 2 μ l aliquots with Drummond micropipets. These volumes were applied in small increments on top of each other. To minimize the size of the spots, we immediately removed the application solvent (evaporation) after each application. The chromatography was ended when the solvent front had migrated 100 mm (measured from the application line). The eluent was then evaporated by heating the plate

to $110\,^{\circ}\text{C}$ for five minutes. Upon spraying with anilinium phthalate (Merck) and another heating to $110\,^{\circ}\text{C}$ for five minutes, the sugar spots were visualized. The R_{f} -values (multiplied by 100) are given in Table 3.

Data analysis methods

Principal components analysis, PCA. To extract the systematic information in Table 3, the data of this table were subjected to PCA. 8,9 This multivariate data analysis method contracts the original 13 variables to a few descriptive dimensions, so-

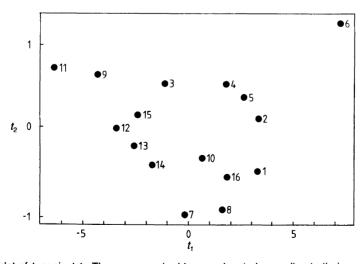


Fig. 1. Score plot of t2 against t1. The monosaccharides are denoted according to their numbers in Table 3.

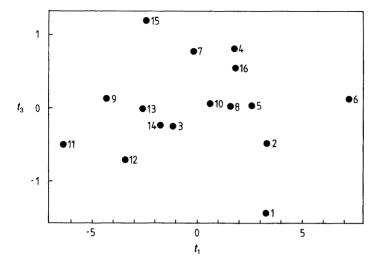


Fig. 2. Score plot of t_3 against t_1 . Notation as in Fig. 1.

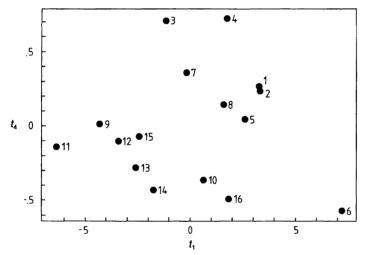


Fig. 3. Score plot of t_4 against t_1 . Notation as in Fig. 1.

called "principal components" (PC's), aimed to illuminate the "principal properties" of the monosaccharides. Since all variables in Table 3 are assumed to be equally important, they are initially scaled to unit variance.

It is quite customary to transform chromatographic data into logarithmic form. However, in the present case the variation in data is fairly small and the logarithmic transformation not needed.

In PCA, the data matrix X (Table 3) is decomposed to means (x_k) , scores (t_{ia}) , loadings (p_{ak}) and residuals (e_{ik}) . In equation form this is given by:

$$x_{ik} = x_k + \sum_{a=1}^A t_{ia} p_{ak} + e_{ik}$$

where the x_{ik} elements consist of the measurements made in the multivariate characterization phase. Index i is used for the monosaccharides and index k for the TLC variables. In the equation above, A denotes the number of significant model dimensions. The numerical value of A, i.e. the number of appropriate PC's, is determined by cross-validation.¹⁴

In this paper PCA is used to calculate PC's for the sixteen monosaccharides. We also establish a

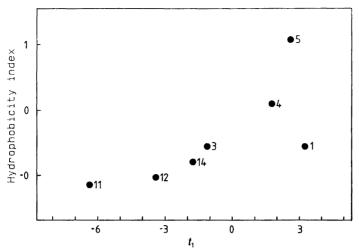


Fig. 4. Correlation between t_1 and the hydrophobicity index presented in Ref. 17. The correlation coefficient is 0.73.

separate PC model for the eight unsubstituted aldohexoses.

Partial least-squares projections to latent structures, PLS. The PLS method relates the information in an X-matrix to the information in a single dependent y-variable or a Y-matrix. Here, PLS is used to check whether the thirteen descriptor variables (matrix X) can forecast how many equatorial and/or axial hydroxy groups the eight unsubstituted hexoses have. Since the OH groups

on the C-2 to C-4 carbon atoms are located either equatorially or axially, it is of interest to see to what extent these patterns are explained by the chromatographic data. The y-variable that marks the number of equatorial OH groups is an external variable, whose value ranges from zero to three.

The PLS algorithm will not be outlined in detail here. For further reading, we refer to more comprehensive discussions in the literature. ^{15,16}

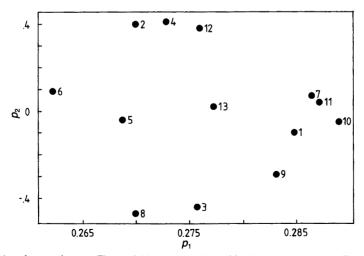


Fig. 5. Loading plot of p_2 against p_1 . The variables are numbered in the same way as in Table 2.

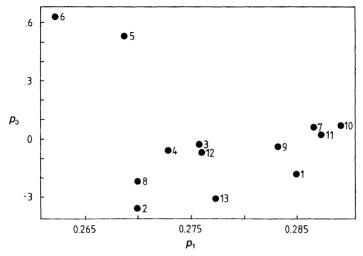


Fig. 6. Loading plot of p_3 against p_1 . Notation as in Fig. 4.

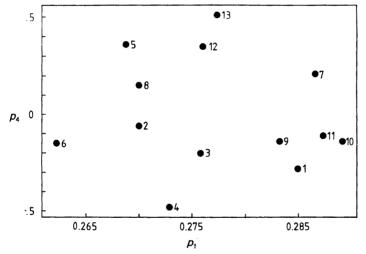


Fig. 7. Loading plot of p_4 against p_1 . For notation, see Fig. 4.

Results and discussion

PCA of sixteen monosaccharides. The PCA for the sixteen monosaccharides resulted in a four-component (t_1-t_4) model. In all, these components explained 97.8% of the variance in the data. The first PC dominated and accounted for 89.9% of the variance. In comparison with the first PC, the remaining components were small and of moderate importance. They described 2.7, 3.7 and 1.5% of the variance, respectively. The PC scores are given in Table 4 and loadings in Table 5.

The derived PC's are plotted in Figs. 1-3. Fig.3 (where t_4 is plotted against t_1) shows an interesting pattern. The distribution of the monosaccharides in this figure shows that they can be divided into three different classes, namely aldopentoses (sugars 1-4), substituted aldohexoses (5-8) and unsubstituted aldohexoses (9-16). Here, rhamnose (No. 6) is a moderate outlier, but the omission of this sugar did not alter the class separation.

It is not evident how to interpret the derived components. An evaluation of the score plots (Figs. 1-3) suggests that PC 1 could correspond

Table 6. PC scores from the separate modelling of the eight aldohexoses. These scales are denoted with a "prime" to avoid confusion with other scorevectors.

No.	Sugar	<i>t</i> ′ ₁	ť ₂
9	Allose	-2.67	0.38
10	Altrose	3.97	-0.28
11	Galactose	-5.37	-0.12
12	Glucose	-1.38	-0.98
13	Gulose	-0.46	-0.06
14	Mannose	0.74	-0.62
15	Talose	-0.32	1.49
16	Idose	5.50	0.19

to a hydrophobicity scale, since it distinguishes between hexoses and substituted hexoses. This assignment is also partly verified by the hydrophobicity index for eight monosaccharides reported by Janado and Yano. These authors determined partition coefficients (K_{av}) for the partitioning of sugars between a polystyrene gel and aqueous solvents. Based on K_{av} , they then derived an index that described the hydrophobicity of the examined monosaccharides. The correlation between this index and t_1 is seen in Fig. 4. Evidently, the existing relationship is not linear but rather slightly curved. The outcome of Fig. 4

strengthens our interpretation of t_1 as a hydrophobicity scale for sugars.

The other components (t_2-t_4) are very small and their meaning unclear. What can be stated concerning t_4 is that it supplements t_1 in the separation of the monosaccharides into sub-classes (Fig. 3).

Loading plots corresponding to Figs. 1–3 are those given in Figs. 5–7. We can see that there is no division of the eluents into different groups. This means that the "design" eluents (7–13) and the others (1–6) contribute with the same type of information.

PC-modelling of eight unsubstituted aldohexoses. From Fig. 3 we see that the eight aldohexoses form a separate class. To have a closer look at their features we performed a new PCA concerning only these eight sugars. This PCA gave two significant components that together described 95.5% of the total variance. The first PC alone explained 91,3% of the variance. To avoid confusion, these two scales are denoted t'_1 and t'_2 (see Table 6).

The PC's are plotted in Fig. 8 as t'_2 against t'_1 . Inspection of the monosaccharide distribution reveals that these PC's express fairly well the content of equatorial hydroxy groups (or equivalently the number of axial OH groups) in each aldohexose (cf. Fig. 9). For example, going in the

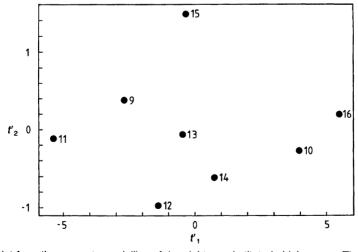


Fig. 8. Score plot from the separate modelling of the eight unsubstituted aldohexoses. The monosaccharides have the same numbers as in Table 3.

34. 511

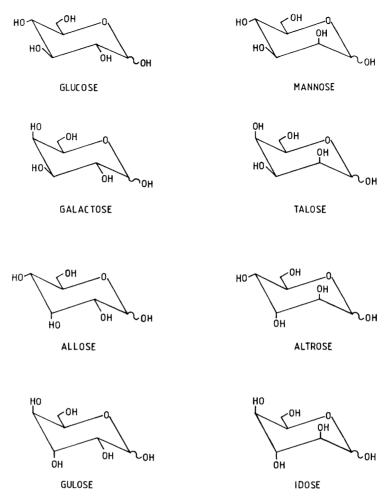


Fig. 9. The OH groups positioned on carbons 2–4 can be located either equatorially or axially. The figure shows that the number of equatorial hydroxy groups goes from one (idose) to three (glucose). There are three sugars (altrose, gulose and talose) with one eq. OH group, and three sugars (allose, galactose and mannose) with two eq. OH groups.

right direction along the line formed by sugars 12, 14, 10 and 16 in the lower right part of Fig. 8, glucose (No. 12) with three equatorial OH groups is followed by mannose (No. 14, with two eq. OH groups) and altrose (No. 10, with one eq. OH group). To the right, this line ends with the three-axial monosaccharide idose (No. 16).

The remaining aldohexoses are also positioned according to their hydroxy substitution pattern. Thus, galactose (No. 11) and allose (No. 9) with two equatorial hydroxy groups are situated near mannose. In the same way, talose (No. 15) and

gulose (No. 13) with one equatorial OH group lie in the same region as altrose.

PLS analysis. A question that was raised during this work was whether the chromatographic data obtained could be used to model the number of equatorial OH groups in each aldohexose. To examine this, we carried out a multivariate PLS analysis with a single y-variable indicating this number. Thus, this external variable ranged from zero (idose) to three (glucose).

The PLS computation gave three significant

Table 7. Observed and calculated number of equatorial hydroxy groups on the *un*substituted aldohexoses.

No.	Sugar	y _{obs} ^a	Y calc
9	Allose	2	2.17
10	Altrose	1	0.87
11	Galactose	2	1.84
12	Glucose	3	2.89
13	Gulose	1	1.46
14	Mannose	2	1.89
15	Talose	1	0.83
16	Idose	0	0.03

^aThe number of observed equatorial OH groups can be seen in Fig. 9.

dimensions that described 95.9% of the variance in the X-block and 89.9% of the variance in the y-variable. In the next step, a calculation of the number of equatorial hydroxy groups was made (Table 7) according to the derived PLS model. The correlations between observed and calculated y-values are shown in Fig. 10. The correlation coefficient is 0.97 and, as can be seen in Fig. 10, this agreement is excellent.

Conclusions

We report here the use of TLC as a rapid method for producing multivariate data that can be used to characterize monosaccharides. PCA of the resulting 16×13 data matrix gives four principal components (t_1-t_4) . The first PC (t_1) explains almost 90% of the variance, while the others are small but statistically significant. When related to a hydrophobicity index (Ref. 17), t_1 correlates in such a manner that it is possible to label it as a scale reflecting hydrophobic properties of the investigated sugars. This fact is also confirmed by observing that t_1 distinguishes between hexoses and substituted hexoses (Figs. 1–3).

An apparent criticism, raised by a referee, against the presently derived parameters for use in QSAR of polysaccharides is that we have investigated the properties of monosaccharides existing in equilibrium between all ring and open forms as well as their α/β forms. However, we may assume that the same intrinsic factors govern both the partial biological activity in polysaccharides and the distribution among the different forms in equilibrium. Hence, the derived PC's may, at least partly, be applicable to the QSAR of polysaccharides where the monomer units exist in a single form, although they are derived for properties of mixtures in equilibrium.

The present parameters should also be regarded as a first effort in the outlined direction

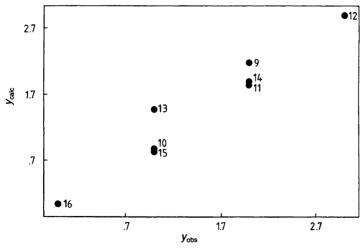


Fig. 10. Correlation between the observed and the calculated number of equatorial hydroxy groups. The correlation coefficient is 0.97.

ERIKSSON ET AL.

and they will be further developed, perhaps by including data for monosaccharide derivatives locked in one single anomer form. It is also quite possible that with other experimental techniques, new and complementary scales describing other sugar properties can be established. We are planning to examine the monosaccharides by FT-IR. Such studies will hopefully lead to the illumination of properties that hitherto have not been described by numerical scales.

Together with the established t_1 these "possible" scales might be used in future QSAR's concerning oligo- and polysaccharides. Then, if these scales are found to be appropriate in such QSAR's, the next step will be to expand them to cover additional monosaccharides.

Acknowledgements. Grants from the Swedish Natural Science Research Council (NFR) and the National Swedish Board for Technical Development (STU) are gratefully acknowledged.

References

- Hellberg, S., Sjöström, M. and Wold, S. Acta Chem. Scand., Ser. B 40 (1986) 135.
- Hellberg, S., Sjöström, M., Skagerberg, B. and Wold, S. J. Med. Chem. 30 (1987) 1126.
- Wold, S., Eriksson, L., Hellberg, S., Jonsson, J., Sjöström, M., Skagerberg, B. and Wikström, C. Can. J. Chem. 65 (1987) 1814.
- 4. Clark, M. E. Biol. Bull. 127 (1964) 63.
- Churms, S. C. Handbook of Chromatography: Carbohydrates, CRC Press, Boca Raton, Florida 1982, Vol. 1.

- Morgan, S. and Jacques, C. Anal. Chem. 54 (1982) 741.
- Eriksson, L., Jonsson, J., Sjöström, M. and Wold, S. Quant. Struct.-Act. Relat. In press.
- Wold, S., Albano, C., Dunn, W.J., III, Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W. and Sjöström, M. In: Kowalski, B. R., Ed., Chemometrics: Mathematics and Statistics in Chemistry, NATO ISI Series C 138, D. Reidel Publ. Co., Dordrecht, Holland 1984, pp. 4-96.
- 9. Wold, S., Esbensen, K. and Geladi, P. Chemomet. Intell. Lab. Syst. 2 (1987) 37.
- 10. Ghebregzabher, M., Rufini, S., Monaldi, B. and Lato, M. J. Chromatogr. 127 (1976) 133.
- Stahl, E. Thin Layer Chromatography, 2nd ed., Springer Verlag, Berlin 1969.
- Box, G. E. P., Hunter, W. G. and Hunter, J. S. Statistics for Experimenters, Wiley, New York 1978
- Fried, B. and Sherma, J. Thin-Layer Chromatography: Techniques and Applications, 2nd ed., Marcel Dekker, New York 1987, pp. 100-110.
- 14. Wold, S. Technometrics 20 (1978) 379.
- Wold, S., Albano, C., Dunn, W.J., III, Esbensen, K., Hellberg, S., Johansson, E., Sjöström, M. In: Martens, H. and Russwurm, H., Jr., Eds., Food Research and Data Analysis. Proc. of the IUFOST Symp. on Data Analysis in Food Research, Oslo 1982, Applied Science Publishers Ltd., London 1983, pp. 147-188.
- Sjöström, M., Wold, S., Lindberg, W., Persson, J. and Martens, H. Anal. Chim. Acta 150 (1983) 61.
- 17. Janado, M. and Yano, Y. J. Solution Chem. 14 (1985) 891.

Received January 20, 1988.