THE PREDICTION OF R_F VALUES OF ALDOHEXOSE AND ALDOMETHYLOSE SUGARS

HERMAN AMATO
Research Laboratories, Eczacibaşi Ilaçlari Limited Sirketi, Levent-Istanbul (Turkey)
(Received October 9th, 1967)

SUMMARY

Mathematical equations are derived in order to predict, with the knowledge of the R_F values of six aldohexose (or aldomethylose) sugars, the R_F values of the remaining two. The aldohexose and aldomethylose sugars are treated separately.

ΔR_M values and vicinal correction factors were calculated for the hydroxyl groups at C-2, C-3 and C-4, using the R_F values given by ISHERWOOD AND JERMYN[1]. A self-consistent set of constants was obtained, which gives good agreement between the experimental and calculated R_M values for all the cases in which data were available in the literature.

Several studies have been devoted to the correlation of the structure of sugars and their chromatographic behaviour[1-8]. In the case of aldohexose and aldomethylose sugars, a quantitative correlation between the steric positions of the hydroxyl groups and the R_F values of the sugars has not been established as far as we know.

Such a correlation may be helpful in the identification of some of the sugars which are not easily obtainable and may be interesting from a theoretical point of view.

THEORETICAL CONSIDERATIONS

The present state of knowledge related to MARTIN's additivity principle can be summarized as follows according to LEDERER AND PIRELLI[9]: "The additivity of ΔR_M values should not be entirely rejected as unreliable, but it should be born in mind that certain solvents obey this rule while others do not".

BUSH[10] has given an extensive review of the R_M theory and its applications. Early and recent examples of the application of the theory are also given in refs. 11–13.

These authors have pointed out that there are two main sources of difficulty in applying the theory, and BUSH[10] has emphasized that these must be distinguished carefully. In the first place many substances contain substituents which interact with one another and do not fulfil the premises of MARTIN's[11] original theory. Secondly, many sources of experimental error have not been adequately controlled by authors claiming to have invalidated the theory[9,10,13].

It seems to the present author that the additivity principle can be still obeyed in some cases in which it previously appeared to fail, if this principle is used in a more refined form. In the case of the isomeric aldohexose (or aldomethylose) sugars which are treated in the present paper, the requirements of MARTIN's theory do not seem to be fulfilled, and a direct application of the additivity principle leads to unsatisfactory results. The fact that paper is not an ideal support for water, the manner of separation of sugars, is still subject to doubt (surface adsorption or liquid-liquid partition) and the separation of isomers seems, at first glance, to be contradictory to MARTIN's statement[11]: "Thus all isomers containing the same groups (note that the degree of ionisation, etc., must not be changed) would be expected to have the same partition coefficient" suggests, in fact, the absence of the requirements of MARTIN's theory. Since the aldohexose (or aldomethylose) sugars are steric isomers one can expect that it would be impossible to separate them from each other according to MARTIN's statement. However, this is not the case.

These facts suggest that a refinement of MARTIN's original simple theory is desirable. Thus, BUSH[10] has pointed out that interacting substituents, while not having simple additive effects on R_M values, should give constant ΔR_M values if treated as complex (multi-substituent) structures. Alternatively the effect of a neighbouring substituent on the ΔR_M value of another should be constant in different substances.

We have tried to discover such invariant groups within sugars by an empirical examination of their R_M values. For example, consider the aldohexose (or aldomethylose) sugars whose configurations differ predominantly from each other by the steric positions of their hydroxyl groups at C-2, C-3 and C-4. We have tried to correlate the change in R_M values with the steric deviations of hydroxyl groups at these positions, taking their orientations in galactose (galactomethylose in the case of aldomethyloses) as the reference structure. The steric deviations of aldohexoses or aldomethyloses from their reference compound can then be divided into two classes:

(a) Sugars that differ from the galacto-configuration by the steric position of one hydroxyl group (gluco- (C-4), gulo- (C-3), talo- (C-2).

(b) Sugars that differ from the galacto-configuration by the steric positions of more than one hydroxyl group.

In case (a) the effect of each deviation on the R_M value can be estimated by simply subtracting the R_M value of galactose (or galactomethylose) from the R_M value of the sugar having the deviation under consideration. As an example, the effect of the deviation at C-4 in the aldohexose sugars can be evaluated as follows:

$$\Delta R_M^{(4)} = R_M^{(4)}\,(\text{glucose}) - R_M^{(0)}\,(\text{galactose}) \qquad (1)$$

Similar expressions can be written for the aldomethylose sugars, provided galactose is replaced by galactomethylose. The superscript (4) refers to the deviation at C-4. Similar notations will be used throughout to indicate that the deviations occur at C-2 and/or C-3, and/or C-4. In case (b), the individual effects of the deviations can not simply be summed up, because the remaining part of the sugar under consideration is not identical to that of galactose if any one of the deviations is removed. Although glucose is reduced to galactose if the deviation at C-4 is removed, the removal of the deviation at C-4 in mannose, for instance, reduces it to talose, which differs from galactose by the deviation at C-2. This fact suggests that the R_M value of any sugar belonging to class (b) cannot be obtained by adding the R_M value of

galactose and the ΔR_M values referring to the individual effects of the deviations defined in the manner given in eqn. (1).

$$R_M^{(2,4)}\ (\text{mannose}) \neq R_M^{(0)}\ (\text{galactose}) + \Delta R_M^{(2)} + \Delta R_M^{(4)}$$

These considerations require that some correction factors must be introduced in order to express the R_M values of the sugars belonging to case (b) in terms of the R_M value of galactose and the individual effects of the deviations. For this purpose a correction factor is associated with each deviation, say d_2, d_3, and d_4 for the deviations at C-2, C-3, and C-4, respectively. It is assumed that the total correction for any sugar is calculated as the arithmetic mean of the individual corrections associated with the deviations appearing in that sugar; and that the individual correction for any deviating group is the sum of the correction factors associated with the remaining deviations. This last statement is justified by the fact that the removal of the remaining deviations reduces the sugar to the one which has only one group deviating from galactose, thus completing the correction procedure, since the resultant "compound" falls in class (a).

In the following, a sugar will be said to be of the first, second, or third class if it has one, two or three groups deviating from galactose. Under the above assumptions, it is easily seen that we can write, for instance, for allose:

$$R_M^{(3,4)} = R_M^{(0)} + \Delta R_M^{(3)} + \Delta R_M^{(4)} + (d_4 + d_3)/2 \qquad (2)$$

and that the double arithmetic mean of the molecular corrections of the sugars of the second class gives the correction of the sugar of the third class, that is:

$$2 \times \frac{(d_3 + d_4)/2 + (d_2 + d_4)/2 + (d_2 + d_3)/2}{3} = \frac{(d_3 + d_4) + (d_2 + d_4) + (d_2 + d_3)}{3} \qquad (3)$$

In eqn. (2), the correction factors d_4 and d_3 are used to correct the deviations at C-3 and C-4, respectively. In eqn. (3), the three terms in the numerator on the left-hand side are the molecular corrections of allose, mannose, and idose, respectively, and on the right-hand the molecular correction of altrose is given. Similar formulas can easily be written for the aldomethylose sugars.

Denoting the arithmetic mean of the molecular corrections of the sugars of the second class by D, eqn. (3) can be simply written as:

$$2D = D^{(2,3,4)} \qquad (4)$$

where the right-hand side is the molecular correction of altrose. Assuming that the mean correction factor D of the second class sugars does not differ significantly from their individual molecular corrections, we can write the following approximate equation which is valid for all the sugars considered:

$$R_M^{(i)} = R_M^{(0)} + (n - 1)D + \sum_{k=1}^{n} \Delta R_M^{(k)} \qquad (5)$$

where n denotes the class number of the sugar, and k must take the group numbers of the deviations in the sugar considered. For example, in mannose k must take the

values 2 and 4, because the deviating groups from galactose are at C-2 and C-4. In cases of the first class sugars we have $n = 1$, and the term in eqn. (5) containing D disappears. For the sugars of the second and third classes the coefficients of D are 1 and 2, respectively. The evaluation of the R_M value of altrose by eqn. (5) using the remaining sugars justifies eqn. (4). Eqn. (5) cannot be used for aldomethylose sugars, since the R_F value of glucomethylose for which $n = 1$ is not given in the work of ISHERWOOD AND JERMYN[1] on which the present paper is based.

In order to overcome this difficulty, a further assumption is introduced suggested by the correction factors of the aldohexose sugars which consists of taking:

$$D^{(3,4)} \cong D^{(2,3)} \tag{6}$$

where $D^{(3,4)}$ is the molecular correction for the allo-nucleus and $D^{(2,3)}$ is that for the ido-nucleus. This new assumption enables us to evaluate the R_M value of any two sugars using the remaining six sugars. Similar calculations are made for the aldomethylose sugars under the assumption:

$$5/4\, D^{*(3,4)} = D^{*(2,3)} \tag{6a}$$

where the coefficient 5/4 is used empirically as giving a better fit with existing data.

CALCULATIONS

As an example, the R_F values of glucose and allose are calculated with the help of the R_F values of galactose, gulose, talose, mannose, idose and altrose. According to eqns. (4), (5), and (6) we can write:

$$\begin{aligned}
R_M^{(3)} &= R_M^{(0)} + \Delta R_M^{(3)} \text{ for gulose,}\\
R_M^{(2)} &= R_M^{(0)} + \Delta R_M^{(2)} \text{ for talose,}\\
R_M^{(2,4)} &= R_M^{(0)} + D^{(2,4)} + \Delta R_M^{(2)} + \Delta R_M^{(4)} \text{ for mannose,}\\
R_M^{(2,3)} &= R_M^{(0)} + D^{(2,3)} + \Delta R_M^{(2)} + \Delta R_M^{(3)} \text{ for idose,}\\
R_M^{(2,3,4)} &= R_M^{(0)} + D^{(2,3,4)} + \Delta R_M^{(2)} + \Delta R_M^{(3)} + \Delta R_M^{(4)} \text{ for altrose,}\\
D^{(2,3)} &= D^{(3,4)}\\
2\,\frac{D^{(2,3)} + D^{(3,4)} + D^{(2,4)}}{3} &= D^{(2,3,4)}
\end{aligned} \tag{7}$$

The set of eqns. (7), where the R_M values are replaced by their experimental values obtained from the R_F values of ISHERWOOD AND JERMYN[1] with the help of a conversion table[14], are solved with respect to the ΔR_M values and the correction factors. By adding $\Delta R_M^{(4)}$ to the R_M value of galactose, $R_M^{(0)}$, we obtained $R_M^{(4)}$, the R_M value of glucose. Similarly, the R_M value of allose, $R_M^{(3,4)}$, is obtained by adding $\Delta R_M^{(3)}$, $\Delta R_M^{(4)}$, and $D^{(3,4)}$ to $R_M^{(0)}$. The R_M values obtained in this way are transformed to R_F values by means of the conversion table.

TABLE I

STERIC DEVIATIONS OF ALDOHEXOSE AND ALDOMETHYLOSE SUGARS FROM THE GALACTO-NUCLEUS AND THE PREDICTIONS OF R_F VALUES MADE ON THE BASIS OF THESE DEVIATIONS

Steric deviations of —OH groups (galacto-nucleus as reference)			*Class (n)*	*Aldohexose sugars*			*Aldomethylose sugars*		
C-2	*C-3*	*C-4*			R_F[1]	R_F *predicted*		R_F[1]	R_F *predicted*
0	0	0		Galactose	0.175	—	Galactomethylose	0.265	—
0	0	1	I	Glucose	0.195	0.19–0.205	Glucomethylose	—	0.315–0.33
0	1	0		Gulose	0.23	0.24	Gulomethylose	0.36	0.35
1	0	0		Talose	0.285	0.29	Talomethylose	0.48	0.475
0	1	1	II	Allose	0.22	0.21	Allomethylose	0.36	0.37
1	0	1		Mannose	0.24	0.23	Mannomethylose	0.38	0.38
1	1	0		Idose	0.31	0.31	Idomethylose	0.51	0.51
1	1	1	III	Altrose	0.27	0.265	Altromethylose	0.45	0.45

Numerical example of the calculation

The R_M values corresponding to the experimental R_F values shown in Table I, are as follows:

$R_M^{(3)}$ (gulose) = 0.525 $R_M^{(2)}$ (talose) = 0.399
$R_M^{(2,4)}$ (mannose) = 0.501 $R_M^{(2,3)}$ (idose) = 0.347
$R_M^{(2,3,4)}$ (altrose) = 0.432 $R_M^{(0)}$ (galactose) = 0.673

Replacing these figures in the set of eqns. (7), we obtain:

$$0.525 = 0.673 + \Delta R_M^{(3)} \text{ for gulose}$$
$$0.399 = 0.673 + \Delta R_M^{(2)} \text{ for talose}$$
$$0.501 = 0.673 + D^{(2,4)} + \Delta R_M^{(2)} + \Delta R_M^{(4)} \text{ for mannose}$$
$$0.347 = 0.673 + D^{(2,3)} + \Delta R_M^{(2)} + \Delta R_M^{(3)} \text{ for idose}$$
$$0.432 = 0.673 + D^{(2,3,4)} + \Delta R_M^{(2)} + \Delta R_M^{(3)} + \Delta R_M^{(4)} \text{ for altrose}$$
$$D^{(2,3)} = D^{(3,4)}$$
$$2\frac{D^{(2,3)} + D^{(3,4)} + D^{(2,4)}}{3} = D^{(2,3,4)}$$

Solving these equations we obtain:

$\Delta R_M^{(3)} = -0.148$ $\Delta R_M^{(2)} = -0.174$
$\Delta R_M^{(4)} = -0.045$ $D^{(2,3)} = D^{(3,4)} = 0.096$

Calculation of the R_F value of glucose gives:

$$R_M^{(4)} = R_M^{(0)} + \Delta R_M^{(4)}$$
$$R_M^{(4)} = 0.673 + (-0.045) = 0.628$$

In the conversion table the R_M value 0.627 corresponds to an R_F value of 0.191.

Calculation of the R_F value of allose gives:

$$R_M^{(3,4)} = R_M^{(0)} + D^{(3,4)} + \Delta R_M^{(3)} + \Delta R_M^{(4)}$$
$$R_M^{(3,4)} = 0.673 + 0.096 + (-0.148) + (-0.045) = 0.576$$

In the conversion table R_M 0.575 corresponds to an R_F value of 0.21.

In the above calculation the R_M values of glucose and allose were omitted. In the next calculation another sugar besides glucose was omitted in order to obtain a new set of equations similar to eqns. (7), and this gave a second "predicted R_M value" for glucose and one R_M value for the sugar omitted. Repeating the procedure of changing the sugar omitted, six predicted R_M values for glucose, and one R_M value for each omitted sugar, other than glucose, were obtained. While the R_M values for glucose and glucomethylose were obtained by six calculations, the other R_M values were obtained by one calculation only. Because there was no experimental R_F value for glucomethylose, it was impossible to do six calculations for each sugar in the

aldomethylose group. For the sake of symmetry as well as better comparisons, the R_M value of glucose, is not introduced in the calculations of the R_F values of sugars belonging to the aldohexose group. The range of the R_F values of glucose shown in Table I may give an idea of the scatter due to different calculations which has a magnitude of about $\pm$ 0.01 R_F unit oscillating around the mean which almost corresponds to the experimental R_F value.

RESULTS

The R_F value of glucomethylose which is not given experimentally is evaluated by a similar set of formulas in six different ways by combining galactomethylose with five of the remaining six sugars each time. At the end of each evaluation, the R_F value of the sugar which is not used in this evaluation is also obtained as a secondary result. These values and the range of R_F values of glucomethylose are indicated in Table I. For comparison, the aldohexose sugars are treated similarly and the values so obtained are also indicated in Table I.

DISCUSSION

The above calculations seem interesting from the following points of view:

(1) They may provide a means of predicting the R_F values of various nuclei, with the help of other nuclei. The information about the predicted nuclei can give further information related to other groups of a substance having the same nuclei as shown by LEVY[3] (information related to aldohexoses may provide information related to aldomethyloses and hexuronic acids, and perhaps about glycosides having the same nuclei and *vice versa*).

Before being too optimistic, the theory has to be tested in more than the one solvent mixture available at present (ethyl acetate–pyridine–water).

(2) The characteristic pattern of D values may provide a means of measuring the degree of complex formation, in solvent mixtures tending to form complexes (*e.g.* phenol) and to explain the inversion of some of the orders of R_F values, for instance, those of galactose and glucose. If the complex is formed by galactose, the predicted R_F values are expected to deviate from their experimental values, because the R_F value of galactose is used as a basis in these predictions. In the case of a complex formed by glucose, the predicted value of glucose must be expected to deviate from the experimental value of glucose, while the predicted values of the remaining sugars would be expected to be in good agreement with experimental values.

(3) The values of the parameters of any consistent set of equations will depend upon the R_M value of one hexose which is taken as a starting point for the derivation of the equations. For various reasons it seemed simplest and most logical to use galactose as the parent compound for developing the set of equations given in this paper. In a previous study[7] it was assumed that the time spent by a sugar in the stationary phase increases according to its similarity to the cellulose structure. This fact, criticized by REICHSTEIN[8], may be explained by the structural similarity of cellulose and galactose.

The following excludes accidental results:

(a) The aldohexose and aldomethylose sugars were treated separately. Each predicted value is found on the basis of the relationship between the correction factors

and the use of experimental R_F values of the remaining six sugars by applying a self-consistent set of equations. No additional constant is introduced. And in the prediction of each value, the R_F of the predicted sugar is not introduced in the calculation. This fact shows that the equations are self-consistent.

(b) Supposing that the results in aldohexose sugars were accidental, it is difficult to explain similar results by using aldomethylose sugars.

ACKNOWLEDGEMENTS

I would like to thank Prof. Dr. T. REICHSTEIN, Dr. I. SEDAT IMRE, Docent Dr. HALIL YÜKSEL and Prof. Dr. I. E. BUSH for their help and criticism.

REFERENCES

1 F. A. ISHERWOOD AND M. A. JERMYN, *Biochem. J.*, 48 (1951) 515.
2 D. FRENCH AND G. M. WILD, *J. Am. Chem. Soc.*, 75 (1953) 2612.
3 M. F. LEVY, *Anal. Chem.*, 26 (1954) 1849.
4 H. JAEGER, A. RAMEL AND O. SCHINDLER, *Helv. Chim. Acta*, 40 (1957) 1310.
5 J. A. THOMA AND D. FRENCH, *Anal. Chem.*, 29 (1957) 1645 (cf. 12).
6 M. T. KRAUSS, H. JAEGER, O. SCHINDLER AND T. REICHSTEIN, *J. Chromatog.*, 3 (1960) 63.
7 H. AMATO, *New Istanbul Contr. Clin. Sci.*, 7 (1964) 32.
8 T. REICHSTEIN, *New Istanbul Contr. Clin. Sci.*, 7 (1964) 326.
9 M. LEDERER AND A. PIRELLI, *Sci. Rept. Ist. Super. Sanità*, 1 (1961) 582 (cf. ref. 10).
10 I. E. BUSH, in D. GLICK (Editor), *Methods of Biochemical Analysis*, Vol 13, Interscience, New York, 1965, p. 357.
11 A. J. P. MARTIN, *Biochem. Soc. Symp. (Cambridge, Engl.)*, 3 (1950) 4 (*cf.* H. G. CASSIDY, *Fundamentals of Chromatography*, Interscience, New York, 1957).
12 M. LEDERER, *Technical Bulletin C2*, Reeve Angel and Co. Ltd., London.
13 S. MARCINKIEWICZ, J. GREEN AND D. MCHALE, *J. Chromatog.*, 10 (1963) 42.
14 E. STAHL, *Thin-Layer Chromatography*, Springer-Verlag, Berlin, 1965, p. 509.