Tetrahedron Vol. 36, pp 3273 to 3280
0040-4020/80/1115-3273/$02.00/0.

^{13}C NMR-SPECTROSCOPY OF SUBSTITUTED XANTHONES–III

COMPUTER SUPPORTED STRUCTURE ELUCIDATION OF POLYHYDROXY XANTHONES

H. HAMBLOCH and A. W. FRAHM*
Institut für Pharmazeutische Chemie der Universität, Kreuzbergweg 26, D-5300 Bonn 1, West Germany

(*Received in U.K.* 15 *January* 1980)

Abstract—A procedure for developing ^{13}C NMR chemical shift additivity rules for all 136 polyhydroxy xanthones, which is based on the method of multiple linear regression analysis with dummy parameters, is reported. A computer search program is presented, which rapidly identifies unkown polyhydroxy xanthones with the help of these additivity rules. The working mode is described and its selectivity demonstrated by means of examples from literature.

Recently several papers on ^{13}C NMR-spectroscopy of polyhydroxy xanthones[1-4] have been published. These investigations also attempted to develop substituent shift increments for hydroxy groups.[1-3] Apart from the use of increments generated from substituted benzenes[1] or simple formation of differences between the chemical shifts of the four monosubstituted xanthones and those of the unsubstituted xanthone,[2] there are reports[3] about establishing increments depending on substituent position as well as on the position of the influenced carbon with regard to interactions caused by multiple substitution. Based on this principle a method for developing ^{13}C NMR-chemical shift additivity rules is described, which deals with the multiple linear regression analysis with dummy parameters. Since the successful interpretation of ^{13}C NMR-spectra of multisubstituted basic structures still remains a tedious operation, a computer program nominated SEOX 1 (Structure Elucidation of Xanthones) is introduced, which rapidly identifies unknown polyhydroxy xanthones with the help of the above additivity rules and thus represents a remarkable facilitation in structure elucidation.

I. MULTIPLE LINEAR REGRESSION ANALYSIS WITH DUMMY PARAMETERS

It is the aim of the multiple linear regression analysis, to establish an optimum linear forecast equation for mathematical magnitudes, which are dependent on several parameters simultaneously.

Therewith it is anticipated, that the parameter influences behave in an additive manner. In general this requirement is always fulfilled with the ^{13}C NMR chemical shifts, if an appropriate parameter selecting has been done. In these cases excellently fitting additivity rules are obtained.

With the hydroxylated xanthones the ^{13}C NMR chemical shifts of the single C-atoms were selected as dependent magnitudes and the hydroxy groups in the various ring positions as the parameters. These can only take the values "0" or "1" in a linear regression analysis with so called dummy parameters.

Figure "1" stands for a hydroxy group in a certain position; figure "0" indicates its absence. It is advisable for such an analysis to develop a balanced distribution design for the hydroxy groups of the polyhydroxy xanthones under investigation, i.e. each of the eight C-atoms, which are ready for substitution, should be substituted at equal frequency. Table 1, column 2 shows 10 oxygenated xanthones as have been chosen for the analysis. For reduction of the number of parameters it was assumed that hydroxy groups in one ring influence the chemical shifts of C-atoms of the other ring only to a very minor extent.‡ The validity of this supposition is shown in several publications.[2,3,5] On these grounds only six of the twelve carbon atoms have been introduced into the analysis i.e. only those carbon atoms which are situated in the substituted ring itself. For those xanthones which are hydroxylated in both rings, only the C-atoms in ring A are considered (Fig 1) but are treated as those of C, i.e. C-8 = C-1; C-7 = C-2, etc.

The ten selected xanthones, simplified in this manner are listed in Table 1, column 3; the distribution patterns are found in columns 4 to 7.

As shown in the last row of Table 1, the principle of an equalised distribution of the hydroxy groups to the various ring positions is well maintained. The multiple linear regression analyses for the chemical shifts have been performed just for the C-atoms 1, 2, 3, 4, 4a and 8b of ring C for the reasons given above. Out of the six systems of equations, that for the chemical shift of C-atom-1 has been chosen as an example in Table 2.

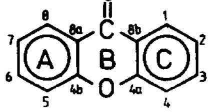


Fig. 1. Xanthone skeleton

---

‡The same assumption was made in the cases of compound **9** and **10** (Table 1), which contain a methoxy group in position 3 (see lit. 5).

Table 1. Polyoxygenated xanthones and their simplified structures with the appropriate distribution pattern

	Compound	Simplified Structure	Distribution Pattern			
			C-1	C-2	C-3	C-4
1	(X)	(X)	0	0	0	0
2	(1-HX)	(1-HX)	1	0	0	0
3	(2-HX)	(2-HX)	0	1	0	0
4	(3-HX)	(3-HX)	0	0	1	0
5	(4-HX)	(4-HX)	0	0	0	1
6	(1,3-DHX)	(1,3-DHX)	1	0	1	0
7	(1,3,5,6-TeHX)	(3,4-DHX)	0	0	1	1
8	(1,3,6,7-TeHX)	(2,3-DHX)	0	1	1	0
9	(1,7,8-TH-3-MX)	(1,2-DHX)	1	1	0	0
10	(1,5,8-TH-3-MX)	(1,4-DHX)	1	0	0	1
		Sum	4	3	4	3

Table 2. Multiple linear regression analysis starting equations for the ^{13}C NMR chemical shift of C-atom-1

Structure	Simplified Structure	Shift^{+} of C-atom-1 (ppm)		C-1	C-2	C-3	C-4
1	(X)	125,92	=	(ax0) +	(bx0) +	(cx0) +	(dx0) + e
2	(1-HX)	160,98	=	(ax1) +	(bx0) +	(cx0) +	(dx0) + e
3	(2-HX)	108,58	=	(ax0) +	(bx1) +	(cx0) +	(dx0) + e
4	(3-HX)	127,96	=	(ax0) +	(bx0) +	(cx1) +	(dx0) + e
5	(4-HX)	115,25	=	(ax0) +	(bx0) +	(cx0) +	(dx1) + e
6	(1,3-DHX)	162,91	=	(ax1) +	(bx0) +	(cx1) +	(dx0) + e
7	(3,4-DHX)	115,93	=	(ax0) +	(bx0) +	(cx1) +	(dx1) + e
8	(2,3-DHX)	108,16	=	(ax0) +	(bx1) +	(cx1) +	(dx0) + e
9	(1,2-DHX)	147,11	=	(ax1) +	(bx1) +	(cx0) +	(dx0) + e
10	(1,4-DHX)	151,83	=	(ax1) +	(bx0) +	(cx0) +	(dx1) + e

$^{+}$values from lit.[3]

The coefficients a to e, derived therefrom, constitute the increments of C-atom-1, the equation of the chemical shift of C-atom-1 thus is as follows:

$$\delta\,\text{C-1} = 125.81 + 36.61\,\text{C-1} - 16.95\,\text{C-2} + 0.66\,\text{C-3} - 10.56\,\text{C-4}.$$

The variables C 1 to C 4, contained therein, take the values "0" or "1", depending on the substitution pattern. The first value (125.81 = basic increment e) represents the corrected chemical shift of C-atom-1 of the unsubstituted xanthone. The statistical data of this regression equation are quite satisfactory. So, with 0.9989 the correlation coefficient r is very high, the standard deviation SD lies at the low level of 1.34 ppm, the test value F figures at 574.5, from which the extremely high significance S > 99.9999 is calculated. The basic increments and the increments of the respective *ipso*, *ortho*, *meta* and *para* positioned substitutents for all the C-atoms 1, 2, 3, 4, 4a and 8b, which have been evaluated by the above procedure, are listed in Table 3.

The appropriate statistical data are summarized in Table 4 and show, that the postulated additivity of hydroxy group increments with respect to the ^{13}C NMR chemical shifts is fulfilled to a high degree.

Table 3. Basic and hydroxy shift increments for C-atom-1, -2, -3, -4, -4a and 8b

C-atom	Basic Shift Increments (ppm)	Hydroxy Shift Increments (ppm)			
1	125,81	IP11[+] = 36,61	OR12[++] = -16,95	ME13[+++] = 0,66	PA14[++++] = -10,56
2	124,44	OR21 = -14,57	IP22 = 30,15	OR23 = -10,96	ME24 = - 0,41
3	135,11	ME31 = 1,49	OR32 = -11,27	IP33 = 29,69	OR34 = -13,55
4	117,93	PA41 = -10,35	ME42 = - 0,21	OR43 = -14,74	IP44 = 29,25
4a	156,06	ME4a1 = - 0,76	PA4a2 = - 7,00	ME4a3 = 1,77	OR4a4 = -11,52
8b	121,93	OR8b1 = -13,17	ME8b2 = - 1,26	PA8b3 = - 7,82	ME8b4 = - 0,69

[+] IP11: Influence of Hydroxysubstitution at C-atom-1 on C-atom-1 (ipso)

[++] OR12: Influence of Hydroxysubstitution at C-atom-2 on C-atom-1 (ortho)

[+++] ME13: Influence of Hydroxysubstitution at C-atom-3 on C-atom-1 (meta)

[++++] PA14: Influence of Hydroxysubstitution at C-atom-4 on C-atom-1 (para)

Table 4. Statistical data of the multiple linear regression equations for C-atom-1, -2, -3, -4, -4a and 8b

C-atom	r	SD	F	S
1	0,9989	1,34	574,5	>99,9999
2	0,9997	0,60	1950,4	>99,9999
3	0,9989	1,12	551,4	>99,9999
4	0,9985	1,26	423,9	99,9998
4a	0,9971	0,55	211,1	99,9991
8b	0,9926	1,13	83,5	99,9910

II. COMPUTER PROGRAM SEOX 1

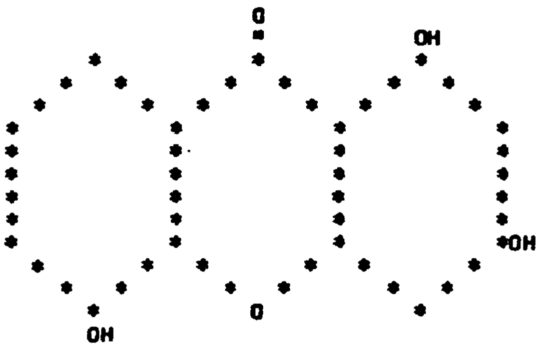
The above calculated increments served to establish a computer program comparing the chemical shifts of the twelve carbons of an unknown polyhydroxy xanthone with the sets of shifts of all 136 theoretically feasible polyhydroxy xanthones or a distinct selection of them, and selecting those sets which fit best with that of the unknown xanthone. The chemical shift of the carbonyl C-atom is not taken into consideration. An example of an output is shown in Fig. 2.

Basically the program works as follows:

A matrix of the size 136 × 13 is developed. In each row in column 1 to 8 a code is positioned for each of the 8 xanthone-C-atoms, which may be substituted. Figure "2" is put in, if the position bears a hydroxy group, and the figure "1", if the position is free. Columns 9–12 are reserved for the C-atoms 4a, 4b, 8a and 8b and kept open at the beginning. The code sum of each row is placed in column 13. It expresses the grade of substitution, as for instance an "8" for the unsubstituted xanthone and an "11" for all trihydroxy xanthones etc. An example is given in Fig. 3. It shows the codes for the 1,3,5-trihydroxy xanthone.

In order to establish the code sum of an unknown xanthone, the number of hydroxy groups must be known. It is derived from ^{1}H-NMR-spectrum, recorded in DMSO-d_6, by integration of the hydroxy proton area. On adding the number of hydroxy groups to the value 8, the code sum is obtained, which then will be entered into the program. The code sum serves for reducing the number of comparison calculations. The set of chemical shifts of an unknown trihydroxy xanthone needs not be compared with all those of the 136 polyhydroxy xanthones, but only with all those of the 28 feasible trihydroxy xanthones (code sum = 11). The same is valid for other substitution grades. Thereafter the calculations of every twelve chemical shifts of those xanthones are performed, the code sum of which is equal to the value of the above established code sum of the unknown xanthone. The calculated chemical shifts substitute the codes in the respective row of the matrix. This operation is demonstrated for C-atom 1 of 1,3,5-trihydroxy xanthone. Since C-atom 1 is coded with a figure "2" (see Fig. 3) the *ipso* increment IP11 is added to the basis increment of C-atom-1 (Table 3) (provisional result). The *ortho* and *para* increments OR12 and PA14 are not taken into

```
                    1.3.5.-TRIHYDROXYXANTHONE
                    -------------------------------------
```



```
CORRELATION COEFFICIENT: 0.99961    STANDARD DEVIATION:  0.80    LIMIT:  3.00  OPTION:  1.
------------------------------------------------------------------------------------------


CARBON:      C-1       C-2      C-3       C-4      C-4A     C-8B      C-5      C-6      C-7      C-8      C-4B     C-8A

            163.08    98.91   166.29     92.84   157.07   100.94    147.18   121.56   124.03   115.25   144.54   121.24
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Fig. 2. Computer search program SEOX 1 output for the given chemical shifts of the "unknown" 1,3,5-Trihydroxy xanthone.

Matrixrow of 1,3,5-Trihydroxy Xanthone

ColumnNr.:	1	2	3	4	5	6	7	8	9	10	11	12	13
C-Atom :	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-4a	C-4b	C-8a	C-8b	Sum
Code :	2	1	2	1	2	1	1	1	-	-	-	-	11

Fig. 3. Matrixrow of 1,3,5-Trihydroxy xanthone

account because of the code Figure "1" of both C-atom-2 and C-atom-4. For C-atom 3, which is coded with a "2" the *meta* increment ME13 is added to the provisional result. In this way the calculated chemical shift of C-atom 1 is obtained, which is placed in column 1 of the respective matrix row. The same procedure is applied for the remaining eleven C-atoms. This whole process is to be repeated with full details for every row with the code sum "11" (27 times).

It is followed by entering the twelve chemical shifts of the unknown polyhydroxy xanthone into the program. These shifts as well as those, which have been calculated in each of the 28 rows, are then automatically put in descending order. Thereafter 28 simple linear regression analyses of the twelve chemical shifts of the unknown xanthone (X-values) are performed with every 12 calculated shifts of the 28 trihydroxy xanthones (Y-values). Correlation coefficients and standard deviations are calculated. The standard deviations are compared with a preselected, freely eligible limit. All xanthones, the standard deviations of which lie under this value, are printed out analogously Fig. 2.

By comparison of the measured chemical shifts of the 1,3,5-trihydroxy xanthone (in our example the "unknown xanthone") with those of the 28 existing trihydroxy xanthones besides the 1,3,5- no further trihydroxy xanthone is found in the output. For further checking of the selectivity of the program the measured chemical shifts of the hydroxy xanthones 1–8 (Table 1), which have been used for the increment calculations, were entered into the computer and were subjected to search with a chosen limit of 3.0 ppm for the standard deviation. The results are summarized in Table 5.

It proves, that in each of the eight cases only the xanthone subjected to search is printed out. Accordingly the selectivity of the program must be very high. For security's sake the limit for the standard deviation should not fall below 3.0 ppm, even though the result turns out to be identical with a limit of even 1.5 ppm. The findings are identical, if the specification of the code sum is dropped. Then the measured chemical shifts of the xanthones 1–8 are compared with the calculated shift sets of all 136 theoretically feasible polyhydroxy xanthones. If, however, more than one xanthone should appear in the output, in all probability that one with the lowest standard deviation will correspond to the unknown xanthone.

Besides this mathematical criterion one can use spectroscopic data for discrimination, like the chemical shifts of the hydroxy proton signals in the ^{1}H NMR spectrum (in DMSO-d_6)[6] or the splitting pattern of distinct C-atom signals and the size of their C-H coupling§ constants derived from the coupled ^{13}C NMR spectrum.

The high selectivity of the program could be corroborated with the 1,2,3-trihydroxy xanthone, the ^{13}C NMR-data of which have been borrowed from literature.[2] This xanthone has not been taken into account for the calculation of the increments (Section I) and represents an especially suitable test compound

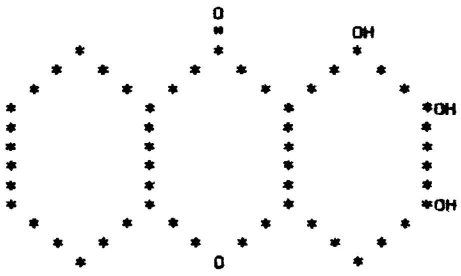
Table 5. Computer searching of compounds **1–8** with correlation coefficients and standard deviations (calculated vs observed chemical shifts)

Compound given	Compound found	SD	r
1	(X)	0,47	0,9995
2	(1-HX)	0,99	0,9989
3	(2-HX)	1,16	0,9977
4	(3-HX)	0,74	0,9994
5	(4-HX)	0,55	0,9992
6	(1,3-DHX)	0,99	0,9994
7	(1,3,5,6-TeHX)	0,68	0,9997
8	(1,3,6,7-TeHX)	0,92	0,9996

---

§ In preparation.

```
                 1.2.3.-TRIHYDROXYXANTHONE
              -------------------------------
```



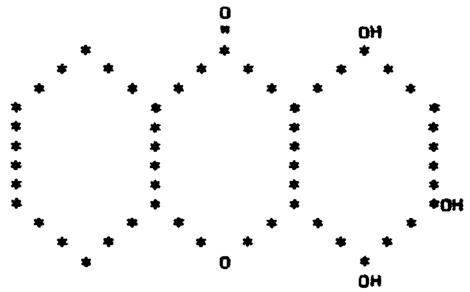
```
 CORRELATION COEFFICIENT: 0.99755    STANDARD DEVIATION:  1.60    LIMIT:  3.00  OPTION:  2.
 ---------------------------------------------------------------------------------------------

  CARBON:   C-1     C-2     C-3     C-4     C-4A    C-8B    C-5     C-6     C-7     C-8     C-4B    C-8A

           146.13  129.06  155.02   92.63  150.07   99.68  117.93  135.11  124.44  125.81  156.06  121.93
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

CALCULATED: 156.06  155.02  150.07  146.13  135.11  129.06  125.81  124.44  121.93  117.93   99.68   92.63
  OBSERVED: 155.80  155.80  150.30  148.90  135.40  128.90  125.40  124.20  119.80  117.90  102.80   94.20
```

```
1.3.4.-TRIHYDROXYXANTHONE
```



```
CORRELATION COEFFICIENT: 0.99286    STANDARD DEVIATION:  2.65    LIMIT:  3.00  OPTION:  2.

  CARBON:    C-1     C-2     C-3     C-4     C-4A    C-8B    C-5     C-6     C-7     C-8     C-4B    C-8A
            152.52   98.50  152.74  122.09  145.55  100.25  117.93  135.11  124.44  125.81  156.06  121.93
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

CALCULATED: 156.06  152.74  152.52  145.55  135.11  125.81  124.44  122.09  121.93  117.93  100.25   98.50
  OBSERVED: 155.80  155.80  150.30  148.90  135.40  128.90  125.40  124.20  119.80  117.90  102.80   94.20
```

Fig. 4a + 4b. Computer search program SEOX 1 output for the given chemical shifts of 1,2,3-Trihydroxy xanthone.

because of the vicinal hydroxy groups. Figure 4a + b show the output of the search for this xanthone.

The expected 1,2,3-trihydroxy xanthone (Fig. 4a) is seen together with the 1,3,4-trihydroxy xanthone (Fig. 4b), irrespective of using the code sum or comparing with all 136 xanthones. As a support for a decision in favour of the 1,2,3-trihydroxy xanthone, its lower standard deviation can be used for the chemical shifts as well as the splitting pattern in the coupled ^{13}C-NMR-spectrum. A 2-bond-coupling of the C-atom-4a is observed with the hydrogen at C-atom-4. The 1,3,4-trihydroxy xanthone would not show this coupling. The detection of the splitted signals of the respective C-atoms in the coupled spectrum is facilitated by the program, which associates the C-atom under consideration with a distinct shift (in the uncoupled spectrum).

SEOX 1 has been developed for ease of structure elucidation of new naturally occuring hydroxy xanthones. Its general capability and rather high selectivity may be checked by means of a card batch of SEOX 1, which is available together with a detailed user instruction from the authors. A program for polymethoxy xanthones as well as for polyacetoxy xanthones is under preparation.

EXPERIMENTAL

Multiple Linear Regression Analyses were executed with the Program Package SPSS (Statistical Package for the Social Sciences).

SEOX 1 is written in FORTRAN IV (leaning against IBM Language Reference Manual 1974) and run on an IBM/370-168 System. Number of Statements: 290; Object Code: c. 12 Kbytes; Array Area: c.9 Kbytes; Compile Time: <0.5 seconds; Execution Time: <5 seconds.

REFERENCES

[1] M. Sumb. Hj. Idris, A. Jefferson and F. Scheinmann, *J Chem. Soc., Perkin I* 2158 (1977).

[2] P. W. Westerman, S. P. Gunasekera, M. Uvais, S. Sultanbawa, R. Kazlauskas, *Org. Mag. Res.* **9**, 631 (1977).

[3] A. W. Frahm and R. K. Chaudhuri, *Tetrahedron* **35**, 2035 (1979).

[4] I. Miura, K. Hostettmann and K. Nakanishi, *Nouveau J. Chem.* **2**, 653 (1978).

[5] R. K. Chaudhuri, F. Zymalkowski and A. W. Frahm *Tetrahedron* **34**, 1837 (1978).

[6] P. Arends and P. Helboe, *Acta Chem. Scand.* **26**, 4180 (1972).