Factor Analysis of Long Range Substituent Effects on C-13 NMR Chemical Shifts and Target Factor Testing of Different Types of Substituent Parameters

Ghazwan F. Fadhil

Chemistry Department, Science College, Basrah University, Basrah, Iraq

Z. Naturforsch. **47 a**, 775–780 (1992); received December 10, 1991

A factor analysis technique has been used to analyse 186 long range C-13 substituent chemical shift (scs) of 90 compounds of the general formula X–G–Y; X represents the set of substituents, G is a hydrocarbon skeleton, in some cases unsaturated (phenylene) and in other cases saturated. Y consists of a variety of groups which contain the probe site.

The data were taken from the literature. Although they were measured in different laboratories, the solvent was $CDCl_3$ in every case. From the value of the factor indicator function (IND) it was concluded that two factors were significant. The validity of several empirical and theoretical substituent parameters has been investigated by Target testing. Models were constructed for the substituent effect. Results from Target factor testing, model designing and linear regression analysis have shown that Reynolds' σ_R^0 and σ_F are the best substituent parameters in reproducing the C-13 substituent chemical shift.

Key words: C-13 nmr, Substituent effect, Factor analysis, Target factor testing.

Introduction

The effect of substituents on the long range C-13 nmr substituent chemical shift (scs) of para substituted benzene has been explained by mathematical models. Those models have divided the workers in this field into two schools of thought, those who attribute the effect of the substituent on the C-13 nmr scs to a single substituent parameter [1], i.e. the Hammett relationship

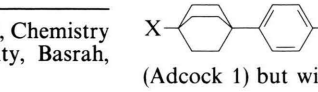
$$\Delta\delta = \varrho\,\sigma_p \qquad (1)$$

($\Delta\delta$ = C-13 nmr scs, σ_p = a substituent parameter, ϱ = a weighting factor for that substituent parameter), and those who attribute the effect of substituent on C-13 nmr scs for para substituted benzene to two substituent parameters [2–4]

$$\Delta\delta = \varrho_F\sigma_F + \varrho_R\,\sigma_R \qquad (2)$$

($\Delta\delta$ as explained above, σ_F and σ_R are substituent parameters which measure the effect of field and resonance, respectively. ϱ_F and ϱ_R are weighting factors for the field and resonance effect, respectively). Another problem has risen during the past two decades when many substituent parameters were compiled, viz.: R (Swain) [2], σ_R^n (Wepster) [5], M (Dewar) [6, 7], σ_R^0 (Brownlee et al.) [4], σ_R^0 (Bromilow et al.) [8], σ_R^0 (Katritzky) [9], S (Godfrey) [10], σ_R^0 (Reynolds) [11], PA [10], σ_R^0 (theor.) [12], $\Sigma\, g_\pi(C_2H_3X/CT)$ [12], $\Sigma\, q_\pi(C_6H_5/CT)$ [12], F (Swain) [2], σ_I (Taft) [8], σ_I (Grob) [13], σ_I (Charton) [14], F (Dewar) [6, 7], σ_I (Adcock 1)* and σ_I (Adcock 2) [15], F (Godfrey) [10], σ_F (Reynolds) [11]. Different workers advocate the use of different substituent parameters in interpreting C-13 nmr scs models. To resolve the controversies concerning the minimum number of factors responsible for C-13 nmr scs and the validity of those published substituent parameter, we have made use of new techniques, Factor analysis and Target factor testing [16]. They have been used successfully in interpreting data matrices of solvent-induced chemical shift [17] and C-13 nmr scs [18].

In this work we shall analyse a data matrix of long range C-13 nmr scs of p-disubstituted benzene and 1-X-4-phenylbicyclo[2.2.2]octanes by means of a factor analysis technique. And then we shall test the validity of the above substituent parameters by using a target factor technique. We first give a brief description of these techniques.

σ_I (Adcock 1) is based on the F-19 nmr chemical shift for  in CCl_4. σ_I (Adcock 2) the same as σ_I (Adcock 1) but with CH_3OH used as solvent.

Reprint requests to Prof. Dr. Ghazwan F. Fadhil, Chemistry Department, Science College, Basrah University, Basrah, Iraq.

0932-0784 / 92 / 0600-0775 $ 01.30/0. – Please order a reprint rather than making your own copy.

The Number of Factors

Malinowski and Howery [16] have defined two functions to detect the correct number of factors in a data matrix, namely the factor indicator function (IND) and the imbedded error function (IE):

$$\mathrm{IND} = \frac{\mathrm{RE}}{(c-n)^2}, \tag{3}$$

$$\mathrm{IE} = \left[\frac{n \sum_{j=n+1}^{c} \lambda_j}{r\,c\,(c-n)}\right]^{1/2}, \tag{4}$$

$$\mathrm{RE} = \left[\frac{\sum_{j=n+1}^{c} \lambda_j}{c\,(c-n)}\right]^{1/2}. \tag{5}$$

RE is the real error, λ_j the secondary eigenvalue (eigenvalue due to error), and r, c, and n are the number of rows, columns and factors in the data matrix, respectively.

These two functions are functions of the secondary eigenvalues, the number of rows and columns in the data matrix and the correct number of factors. Hence the behaviour of these two functions varies with the number of factors. We increase the number of factors gradually and calculate each time the IND and IE functions. As the number of factors increases, the IND and IE functions decrease in value and reache a minimum when the correct number of factors is employed.

Target Factor Testing

To test the validity of a factor, Malinowski and Howery [16] have designed a target factor test to examine the validity of a test factor in reproducing the data matrix. They have defined a function called Spoil as

$$\mathrm{Spoil} = \frac{\mathrm{RET}}{\mathrm{EDM}} \simeq \frac{\mathrm{RET}}{\mathrm{REP}}. \tag{6}$$

RET is the real error in the target, REP the real error in the predicted target, and EDM the error from the data matrix. According to Malinowski and Howery a Spoil value between 0 and 3 indicates an acceptable factor. A factor which gives a Spoil value between 3 and 6 is moderately acceptable while a factor with a Spoil value greater than 6 is not acceptable.

Results and Discussion

The system under investigation had the general formula X–G–Y. X represents the substituent. Six substituents were used in the analysis, namely OCH_3, CH_3, F, Cl, Br, and NO_2. In fact they form a basis set of substituents. G is a hydrocarbon skeleton. Two types of skeletons were involved: unsaturated, viz. phenyl group, and saturated, viz. bicyclooctyl group and methylene group attached to phenyl group. Y forms a variety of groups where the probe lies in. For full description of Y see Table 1.

The C-13 nmr scs data matrix was constructed as follows: Each substituent constituted a row, and the long range C-13 nmr scs for the probe carbon atom in each type of G–Y constituted a column. Table 2 shows the variety of C-13 nmr scs probes.

The Number of Factors

Table 3 gives the results of IND and IE functions. The IND function reaches a minimum at two abstract factors and starts increasing at three abstract factors, which indicates that the third factor is carrying nothing but error.

Hence there are two significant factors in our C-13 nmr scs data matrix. This finding is in accordance with the dual substituent parameters concept. The IE function value drops sharply from abstract factors one to two and then continues to decrease slowly. This misbehaviour of the IE function has been found by other workers [16] and has been attributed to the existence of a non uniform error in the data matrix.

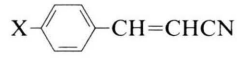
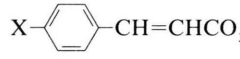
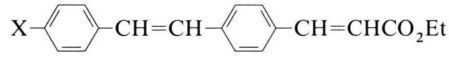
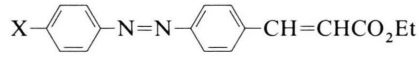
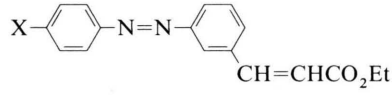
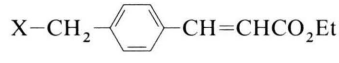
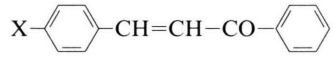
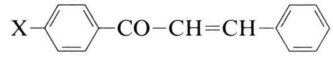
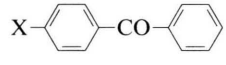
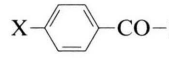
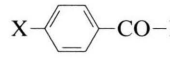
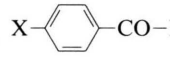
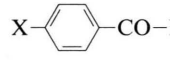
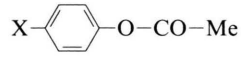
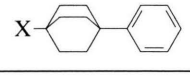
Our data matrix is a collection from different laboratories and the experimental error will not be uniform.

Nevertheless the IND function has given a better indication for the (true) significant number of factors and has been used successfully in situations in which the IE function exhibits no minimum [16].

The Validity of the Substituent Parameters

The validity of the substituent parameters mentioned in the introduction has been tested by using the target factor technique. Table 4 gives Spoil values for the case of two factors since the results from factor analysis give conclusive indication that two factors are sufficient to reproduce the C-13 nmr scs data matrix.

Table 1. Types of molecules used in the study.

No	X–G–Y
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

Reynolds' σ_R^0 and σ_F give the lowest Spoil value. Reynolds et al. [11] have derived these substituent parameters from iterative multiple linear regression analysis of substituent C-13 nmr chemical shifts in 3–4-substituted styrene derivatives.

R (Swain) and M (Dewar) showed approximately similar Spoil values. This is because both parameters depend in their dependence on σ_p.

Table 2. Factor analysed C-13 scs probes of the molecules indicated in Table 1.

X–G–Y	C-13 probe	References
1	β-carbon atom [a]	[19]
2	α-carbon atom [b]	[20]
2	β-carbon atom	[20]
3	α-carbon atom [c]	[20]
3	β-carbon atom [c]	[20]
4	α-carbon atom	[20]
4	β-carbon atom	[20]
5	α-carbon atom	[20]
5	β-carbon atom	[20]
6	α-carbon atom	[21]
6	β-carbon atom	[21]
7	α-carbon atom	[22]
7	β-carbon atom	[22]
7	α-carbonyl carbon atom	[22]
8	α-carbon atom	[22]
8	β-carbon atom	[22]
8	carbonyl carbon atom	[22]
9	carbonyl carbon atom	[23]
9	carbon atom no. 1 [d]	[23]
9	carbon atom no. 2 [d]	[23]
9	carbon atom no. 3 [d]	[23]
9	carbon atom no. 4 [d]	[23]
10	carbonyl carbon atom	[24]
11	carbonyl carbon atom	[24]
12	carbonyl carbon atom	[24]
13	carbonyl carbon atom	[24]
14	carbonyl carbon atom	[24]
15	carbon atom no. 1 [d]	[25]
15	carbon atom no. 2 [d]	[25]
15	carbon atom no. 3 [d]	[25]
15	carbon atom no. 4 [d]	[25]

[a] In this table β always refers to the remote carbon atom of the vinyl group.
[b] For the other carbon atom in the same vinyl group.
[c] For the remote vinyl group.
[d] As in the corresponding reference.

Table 3. Behaviour of IND and IE functions as the number of factors increased.

Number of abstract factors	IND function	IE function
1	2.39×10^{-2}	0.245
2	7.27×10^{-3}	7.2×10^{-2}
3	9.75×10^{-3}	6.21×10^{-2}
4	1.48×10^{-3}	4.83×10^{-2}
5	4.26×10^{-3}	3.89×10^{-2}

σ_R^0 (Brownlee et al.) and σ_R^0 (Bromilow et al.), which are derived from C-13 nmr data of the para carbon atom of substituted benzene in CCl_4, showed a rather high Spoil value when compared with σ_R^0 (Reynolds). We attribute this to the method by which σ_R^0 (Rey-

Table 4. Spoil values of some common substituent parameters at two abstract factors.

No.	Tested factor	Spoil
1	σ_R^0 (Reynolds)	1.89
2	R (Swain)	2.32
3	σ_R^0 (Bromilow et al.)	2.47
4	M (Dewar)	2.63
5	S (Godfrey)	3.04
6	PA	3.32
7	σ_R^0 (Brownlee et al.)	3.71
8	σ_R^0 (Katritzky)	5.41
9	σ_R^n (Wepster)	7.57
10	σ_R^0 (theor.)	7.69
11	Σq_π (C_2H_3X/CT)	8.63
12	Σq_π (C_6H_5X/CT)	9.49
13	σ_F (Reynolds)	1.99
14	σ_I (Grob)	2.51
15	σ_I (Adcock 1)	2.66
16	σ_I (Charton)	2.78
17	F (Swain)	2.86
18	F (Dewar)	2.9
19	σ_I (Adcock 2)	3.01
20	σ_I (Taft)	3.41
21	U	10.7
22	F (Godfrey)	12.7

nolds) was derived as well as to the use of a large set of substituted styrene derivatives in this method.

PA and S (Godfrey) substituent parameters showed acceptable Spoil values. This confirms a previous finding [10] that the S (Godfrey) parameter is equivalent to the σ_R^0 parameter.

σ_R^n (Wepster) did not give an acceptable Spoil value. This could be attributed to the inefficiency of the original substituent parameter which σ_R^n was derived from, which depends on thermodynamical data of phenylacetic acid in different solvents.

The validity of semiempirical resonance substituent parameters, viz. R (Swain), M (Dewar), σ_R^n (Wepster), σ_R^0 (Brownlee et al.) is not only dependent upon the method used to design these parameters but also depends upon the σ_I parameter since the latter parameter is used in their derivation. Hence the quality of σ_I will definitely affect the quality of these semiempirical resonance parameters.

$\Sigma q_\pi(C_2H_3X)$, $\Sigma q_\pi(C_6H_5X)$ σ_R^0 (theor.) and σ_R^0 (Katritzky), did not give acceptable Spoil values when each was tested separately. The high Spoil value for these parameters indicates that the resonance component of the long range C-13 nmr scs is not simply dependent upon the amount of π-charge transfer between the substituent and the phenyl group attached to it.

σ_I (Grob), σ_I (Adcock 1), σ_I (Adcock 2), σ_I (Charton), F (Swain), F (Dewar) and σ_I (Taft) all gave similar Spoil values in the range 2.5–3.4, whereas σ_F (Reynolds) gave the lowest value 1.99.

Reynolds et al. [11] have shown that their σ_F parameter gave a better correlation coefficient than σ_I (Charton) and σ_I (Grob) when applied to chemical models where only a field effect is operative using aprotic solvents.

Our findings lends supports to the utility of the Reynolds σ_F parameter, since the size of our data matrix is larger and covers a wide range of C-13 nmr scs probes.

The Unity (U) substituent parameter did not give an acceptable Spoil value. This justifies the dissection of the C-13 nmr scs effect into field and resonance effects only.

The F (Godfrey) substituent parameter did not give an acceptable Spoil value. This theoretical parameter has shown to be similar to a group electronegativity parameter [10]. Thus we may conclude that the group electronegativity parameter is not a factor in determining the C-13 nmr scs.

Model Designing

To cast more light on the validity of substituent parameters, we have constructed several substituent models from those substituent parameters which have acceptable and moderately acceptable Spoil values.

The chemical resonance behind the combination of these substituent parameters was also considered. The efficiency of the model was given by the Root Mean Squares (RMS) of the data matrix reproduction. The combined substituent parameters which give the lowest RMS will be the best substituent parameters to reproduce the data matrix. Table 5 gives the details of the constructed models with their RMS values. The model with the lowest RMS is the one which combines σ_R^0 (Reynolds) and σ_F (Reynolds).

Our data matrix contained one π-acceptor group, namely the nitro group, because of limited number of data available in the literature. In order to test the validity of Reynolds' substituent parameter for data sets which involve more than one π-acceptor group, we have performed linear regression analysis between Reynolds' substituent parameters (σ_R^0 and σ_F) and the

Table 5. Designed models using two substituent parameters (factors).

No.	Combined factors	RMS
1	σ_R^0 (Reynolds) + σ_F (Reynolds)	0.109
2	R (Swain) + F (Swain)	0.119
3	M (Dewar) + F (Dewar)	0.122
4	σ_R^0 (Brownlee et al.) + σ_I (Taft)	0.126
5	σ_R^0 (Bromilow et al.) + σ_I (Adcock 1)	0.130
6	σ_R^0 (Katritzky) + σ_I (Adcock 1)	0.164
7	σ_R^0 (theor.) + σ_I (Charton)	0.216
8	Σq_π (C_2H_3X/CT) + σ_I (Adcock 1)	0.235
9	σ_R^n (Wepster) + σ_I (Grob)	0.235
10	PA + U	0.252
11	S (Godfrey) + F (Godfrey)	0.841

C-13 substituent chemical shift for different X–G–Y systems. Table 6 presents the molecular systems used in the study as well as the correlation coefficient (r) and the (f) value which is sd/rms for the regression analysis of C-13 substituent chemical shift against Reynolds' substituent parameters and against other substituent parameters together with their references. The (r) and (f) values for the correlation of scs with Reynolds' substituent parameters are either as good as of those obtained by other workers using other types of σ_R^0 (cf. Table 6) or better than them. In fact using Reynolds' σ_R^0 and σ_F in correlating scs of different X–G–Y systems gives a better understanding for the correlation than using different types of σ_R^0 for different molecular systems.

Table 6. Statistical results for correlating C-13 scs for different probes with Reynolds' σ_F and σ_R^0.

Molecular system	Correlated property	Statistical results using DSP model of Reynolds		Substituent parameter(s) used by other workers BA	Statistical results		Ref.
		r	f		r	f	
1 X–C$_6$H$_4$–	C-13 (C1)	0.988	0.029	σ_R, σ_I BA	0.996	0.09	[26]
2 X–C$_6$H$_4$–NH–CO–NH$_2$	C-13 (C1)	0.971	0.198	σ_R, σ_I	0.977	0.23	[27]
3 X–C$_6$H$_4$–OH	C-13 (C1)	0.999	0.029	σ_R, σ_I BA	0.998	0.08	[27]
4 X–C$_6$H$_4$–O–CO–C_2H_5	C-13 (C1)	0.991	0.127	σ_R, σ_I BA	0.994	0.11	[27]
5 X–C$_6$H$_4$–NH–CO–CH_3	C-13 (C1)	0.991	0.115	σ_R, σ_I BA	0.997	0.08	[26]
	C-13 (CH_3)	0.992	0.11	σ_R, σ_I BA	0.997	0.15	[26]
	C-13 (CO)	0.989	0.009	σ_R, σ_I	0.991	0.14	[26]
6 X–C$_6$H$_4$–CH–O–C$_6$H$_4$–NO_2	C-13 (C1)	0.992	0.128	σ_R, σ_I	0.993	0.12	[26]
7 X–C$_6$H$_4$–CH_2^{α}–N=CH–C$_6$H$_5$	C-13 (Cα)	0.997	0.079	σ^0	0.995	–	[1 b]
	C-13 (Ci)	0.997	0.026	σ	0.595	–	[1 b]
8 X–C$_6$H$_4$–CH=N–CH_2–C$_6$H$_4$ (1, 4)	C-13 (Ci)	0.995	0.063	σ	0.881	–	[1 b]
	C-13 ($C\bar{1}$)	0.994	0.107	σ	0.994	–	[1 b]
	C-13 ($C\bar{4}$)	0.994	0.113	σ^0	0.986	–	[1 b]
9 X–C$_6$H$_4$–CH=N–C$_6$H$_5$	C-13 (Ci)	0.979	0.130	σ_p	0.926	–	[1 c]

Substituent set for **1** NH_2, OMe, Me, F, Cl, Br, CN, NO_2; **2** NMe_2, OMe, Me, F, Cl, Br, CN, NO_2; **3** NH_2, Me, F, Cl, COMe, NO_2; **4** NH_2, Me, F, Cl, COMe, NO_2; **5** NH_2, OMe, Me, F, Cl, Br, CF_3, CN, NO_2; **6** OMe, Me, F, Cl, CN, NO_2; **7** and **8** NMe_2, OMe, Me, F, Cl, CN; **9** NMe_2, OMe_3, Me, Cl, Br, NO_2.

Calculations

The calculations were carried out on an NEC 800 computer, using FACTANAL computer programme.

Conclusion

The factor analysis has manifested the duality of long range C-13 nmr scs. Target factor testing and model designing have indicated the superiority of Reynolds' σ_R^0 and σ_F parameters over any other substituent parameters. Results of linear regression analysis of C-13 scs for molecular systems not included in the C-13 data matrix also indicated the superiority of Reynolds' substituent parameters.

[1] (a) G. L. Nelson, G. C. Levy, and J. D. Crigioli, J. Amer. Chem. Soc. **94,** 3089 (1972). – (b) J. E. Arrowsmith, M. J. Cook, and D. J. Hardstone, Org. Magn. Reson. **11,** 160 (1978). – (c) A. Echevarria, J. Miller, and M. G. Nascimento, Magn. Reson. Chem. **23,** 809 (1985).
[2] C. G. Swain and E. C. Lupton, J. Amer. Chem. Soc. **90,** 4328 (1968).
[3] P. R. Wells, S. Ehrenson, and R. W. Taft, Prog. Phys. Org. Chem. **6,** 147 (1968).
[4] S. Ehrenson, R. T. C. Brownlee, and R. W. Taft, Prog. Phys. Org. Chem. **10,** 1 (1973).
[5] A. J. Hoefnagel and B. M. Wepster, J. Amer. Chem. Soc. **90,** 7007 (1970).
[6] M. J. S. Dewar and P. J. Grisdale, J. Amer. Chem. Soc. **84,** 3539 (1962).
[7] M. J. S. Dewar, and P. J. Grisdale, J. Amer. Chem. Soc. **84,** 3548 (1962).
[8] J. Bromilow, R. T. C. Brownlee, V. O. Lopez, and T. W. Taft, J. Org. Chem. **44,** 4766 (1979).
[9] A. R. Katritzky and R. D. Topsom, Chem. Rev. **77,** 639 (1977).
[10] G. F. Fadhil and M. Godfrey, J. Chem. Soc. Perkin Trans. II, **1982,** 933.
[11] W. F. Reynolds, A. Gomes, A. Maron, D. W. Macintyre, A. Tanin, G. K. Hamer, and I. R. Peat, Can. J. Chem. **61,** 2376 (1983).
[12] S. Marriott and R. D. Topsom, J. Chem. Soc. Perkin Trans. II, **1985,** 1045.
[13] C. A. Grob, B. Schaub, and M. G. Schlageter, Helv. Chim. Acta **63,** 57 (1980).
[14] M. Charton, Prog. Phys. Org. Chem. **13,** 199 (1981).
[15] W. Adcock and A. N. Abeywickrema, J. Org. Chem. **47,** 2957 (1982).
[16] E. R. Malinowski and D. G. Howery, Factor Analysis in Chemistry, Wiley Intersci., New York 1980.
[17] G. F. Fadhil, Ph.D. Thesis, University of Southampton, 1983.
[18] W. F. Reynolds, A. Gomes, A. Maron, D. W. Macintyre, R. G. Maunder, A. Tanir, H. E. Wong, G. K. Hamer, and I. R. Peat, Can. J. Chem. **61,** 2367 (1983).
[19] G. Butt and R. D. Topsom, Spectrochem. Acta **36A,** 811 (1980).
[20] D. Christoforou and D. A. R. Happer, Aust. J. Chem. **35,** 729 (1982).
[21] D. A. R. Happer and B. E. Steenson, Aust. J. Chem. **38,** 337 (1985).
[22] E. Solaniova, S. Toma, and S. Gronowitz, Org. Mag. Res. **8,** 439 (1976).
[23] M. J. Shapiro, Tetrahedron **33,** 1091 (1977).
[24] J. Bromilow, R. T. C. Brownlee, D. J. Craik, P. R. Fiska, J. E. Rowe, and M. Sadek, J. Chem. Soc. Perkin Trans. II, **1981,** 753.
[25] W. Adcock and Thong-Chak, Khor, J. Amer. Soc. **100,** 7799 (1978).
[26] C. J. O'Connor, D. J. McLennan, D. J. Calvert, T. D. Lomax, A. J. Porter, and A. Rogers, Aust. J. Chem. **37,** 497 (1984).
[27] C., J. O'Connor, D. J. McLennan, D. J. Calvert, and A. S. H. Mitha, Aust. J. Chem. **40,** 677 (1987).