# Regular article Prediction of <sup>13</sup>C NMR chemical shifts in substituted naphthalenes

C.A.L. Mahaffy<sup>1</sup>, J.R. Nanney<sup>2</sup>, R.E. Jetton<sup>3</sup>

<sup>1</sup> Department of Chemistry, Auburn University at Montgomery, Montgomery, AL 36117, USA

<sup>2</sup> Department of Mathematics, Auburn University at Montgomery, Montgomery, AL 36117, USA

<sup>3</sup> Computer Sciences Corporation, 201 Technacenter Drive, Montgomery, AL 36117, USA

Received: 19 June 1998 / Accepted: 20 October 1998 / Published online: 16 March 1999

Abstract. The prediction of the  $^{13}$ C NMR signals for derivatives of naphthalene has been investigated using mathematical modeling techniques. Two empirical multiple regression models which utilize the field, resonance, and Charton's steric parameters together with molar refractivity were developed, one for  $\alpha$ - and the other for  $\beta$ -substituted naphthalene derivatives. In the  $\alpha$  case the model had a correlation coefficient of observed versus predicted line positions of  $r = 0.973$  with a standard deviation of 2.2 ppm while in the  $\beta$  case  $r = 0.979$  with the standard deviation being 2.3 ppm. The database consisted of 3152 signals from 394 naphthalene derivatives. We also report the use of the Taft steric parameter in place of the Charton steric parameter in the abovementioned prediction equations.

Key words:  $^{13}$ C NMR of naphthalenes – mathematical modeling

## 1 Introduction

We have investigated two different techniques that can be used to predict the NMR line positions using statistical methods. The first method is to use the statistical substituent chemical shift (SSCS) values. An SSCS value is a measure of the shift in an NMR line position induced by substituting a particular group on a structure which acts as a standard. We have reported SSCS values for a number of systems: the  ${}^{11}B$  NMR spectra of trigonal boranes [1], the  $^{13}$ C NMR spectra of arenes  $[2-3]$ , and the <sup>19</sup>F NMR signals for fluoro derivatives of arenes [4], six-membered aromatic nitrogen heterocycles [5], arenetricarbonylchromium(0) complexes [6], and ethylenes [7].

The second technique is to establish an equation using known parameter values for the electronic, steric, or polarization properties of the groups attached to the base structure in order to provide an estimate of the NMR line position. We have had success using the values for the field, resonance, and Charton's steric parameters  $(v)$  together with molar refractivity  $(MR)$  in the prediction of the  $^{11}$ B NMR of trigonal boranes [8], the  $13^{\circ}$ C NMR of arenes [9-10], the nitrogen NMR spectra for derivatives of ammonia [11-12], and the  $^{19}F$  NMR signals for fluoroarenes  $[13-14]$  and fluoroarenetricarbonyl-chromium $(0)$ complexes [15–16].

The studies reporting the  $^{13}$ C NMR spectra [10] and the  $^{19}$ F NMR spectra [14] of arenes were done utilizing symmetrical coefficients in the prediction equations, a technique which does not require any arbitrary orientation of the molecule before the prediction equation can be employed. In the above-mentioned older studies some arbitrary orientation of the molecule was necessary before the prediction equations could be employed which led to asymmetrical prediction equations.

In the present paper we report the application of the parameter method to the <sup>13</sup>C NMR spectra of substituted naphthalenes.

## 2 Experimental

Statistical computations were done on a Sun SPARC-10 using Statistical Analysis System (SAS) and on a Pentium Pro 200 computer running the Linux operating system.

## 3 Results and discussion

#### 3.1 The database

We surveyed the literature from the period from 1970 until 1985. From this we extracted our database in which

- 1. The compound was a naphthalene derivative containing one or more groups. The groups are listed in Table 1.
- 2. All signals were converted so as to use tetramethylsilane as a reference. The database consists of 3152 signals from 394 naphthalene derivatives.

This article is supplemented by an internet archive which can be obtained electronically from the Springer-Verlag server located at http://link.springer.de/journals/tca

Table 1. Group numbers, codes, groups, field  $(F)$ , resonance  $(R)$ , steric parameters (v), and molar refractivity  $(MR)^{2}[17-19]^{a}$ 

No.	Code	Group	F	R	v	MR
1		Н	0.00	0.00	0.00	0.00
2	А	Br	0.44	$-0.17$	0.65	8.88
3	B	Cl	0.41	$-0.15$	0.55	6.03
4	C	NH <sub>2</sub>	0.02	$-0.68$	0.35	5.42
5	D	CH <sub>2</sub> CH <sub>3</sub>	$-0.05$	$-0.10$	0.56	10.30
6	E	F	0.43	$-0.34$	0.27	0.92
8	F	OН	0.29	$-0.64$	0.32	2.85
11	G	OCH <sub>3</sub>	0.26	$-0.51$	0.36	7.87
12	I	CH <sub>3</sub>	$-0.04$	$-0.13$	0.52	5.65
13	J	NO <sub>2</sub>	0.67	0.16	0.59	7.36
14	K	SCH <sub>3</sub>	0.20	$-0.18$	0.64	13.82
15	L	COCH <sub>3</sub>	0.32	0.20	0.50	11.18
16	М	$CH(CH_3)$	$-0.05$	$-0.10$	0.76	14.96
17	N	$C(CH_3)_3$	$-0.07$	$-0.13$	1.24	19.62
18	P	$N(CH_3)$	0.10	$-0.92$	0.43	15.55
21	Q	CH <sub>2</sub> OH	0.00	0.00	0.53	7.19
22	R	$CH=CH2$	0.07	$-0.08$	1.31	10.99
156	S	<b>COOEt</b>	0.33	0.15	0.69	17.47

<sup>a</sup>We have used the same group numbers as in our previous papers  $[1-16]$  for consistency. The group codes are used for brevity in Table 4

#### 3.2 Statistical analysis

Because of the geometry of the naphthalene molecule, there are three categories of  $^{13}$ C NMR signal, the  $\alpha$ signals from carbons in positions 1, 4, 5, and 8 and the  $\beta$  signals from the carbons in positions 2, 3, 6, and 7. In addition there are also signals from the carbons in positions 9 and 10. This necessitates dividing the signals into three populations before models for prediction of the  $^{13}$ C NMR signals are computed. Thus we derive one model for predicting  $\alpha$  signals and another model for predicting  $\beta$  signals. The signals from the carbons in positions 9 and 10 are nearly constant at around 130 ppm, and we were not successful in computing a model to accurately predict these signals.



For the  $\alpha$  population each atom has four encodings in the file used to compute the model, one for each of the four  $\alpha$  signals. The encoding is determined by orienting the atom so that the carbon with which the signal is associated is in position 1. The subscripts in the model refer to the numbered positions after this orientation has been made. For example, compound number 25, 1-acetyl-2 fluoronaphthalene has the following four encodings in the database used for computing the model for  $\alpha$  signals.



The integers in the encoding refer to the substituent numbers, 6 being fluorine, 15 being acetyl, and 1 being hydrogen. The numbers for the substituents can be found in Table 1. The signals at the end of the line are from the carbons in position 1. The 394 compounds used in the study thus resulted in a database of 1576 alpha signals from which the model was computed.

There seems to be little interaction between adjacent groups. This may result partly from the fact that the compounds in this database are sparcely substituted. The small interactions between adjacent substituents means that the signals can probably be predicted well from SCS values, though very few SCS values can be determined directly from this database.

The variables used in the model are the field parameter,  $F$ , the resonance parameter,  $R$ , Charton's steric parameter,  $v$ , and MR. A scarcity of well determined values of  $\nu$  limits somewhat the compounds which can be predicted by this method. In the model,  $F_i$ ,  $R_i$ ,  $v_i$ , and  $MR_i$  denote the field, resonance, steric and molar refractivity respectively of the substituent in the ith position,  $i = 1...8$ . The model does not allow for interactions between different substituents, though it does seem to require self-interaction of parameters for the same substituent for substituents that are close to the signal being predicted. For  $\alpha$  signals self-interaction variables are used for substituents in positions 1, 2, and 8. The coefficients of the variables were determined by stepwise multiple regression, and all variables statistically significant at the  $5\%$  level of significance were retained in the model resulting in 40 predictor variables including the self-interaction variables in positions 1, 2, and 8. The predicted variable  $\delta$  is of course the <sup>13</sup>C NMR shift of the carbon in position 1.



Note that none of the variables in positions 5 and 6 were statistically significant which means that the signal for position 1 is scarcely influenced at all by the substituents in positions 5 and 6. The correlation of observed versus predicted signals, shown in Fig. 1, has a correlation



Fig. 1. Correlation between 1576 pairs of observed and predicted  $\alpha$ signals.  $r = 0.973$ 

coefficient of  $r = 0.973$  with a standard deviation of 2.2 ppm. The average absolute error of prediction is therefore 1.5 ppm. Standard errors of estimate of the coefficients as well as  $p$  values are given in Table 2. The  $F$ value for the model was 682.5 which yields a  $p$  value under 0.0001. Most of the predictions with large residuals are from carbons with an attached amino group, though most carbons with an amino group attached are predicted quite accurately. We are unable to explain this anomoly.

To assess the internal validity of the model, we ran the leave-one-out cross-validation procedure. The sum of squares of errors (SSE) for the generated pseudovalues (PRESS) was 8438, only slightly higher than 7769, the (SSE) for the model. This indicates that the model ought to predict new data of similar type about as well as it predicts the data from which the model was generated.

In the database used for computing the model to predict  $\beta$  signals, the encodings are the same as for the  $\alpha$ database except that the  ${}^{13}$ C NMR signal listed at the end of a line is from the carbon in position 2. This results in the following four lines in this database for 1-acetyl-2-fluoronaphthalene, one line for each  $\beta$  signal.



As with  $\alpha$  signals this complete database from which the prediction model was computed, has 1576 lines. Also  $F_i$ ,  $R_i$ ,  $v_i$ , and MR<sub>i</sub>,  $i = 1...8$  are defined as in the  $\alpha$ model. For  $\beta$  signals we used self-interaction among the parameters for positions 1, 2, and 3. The model, determined by a stepwise regression procedure, has 39 variables significant at the 5% level.



 $\delta = 126.54 - 17.415 F_1 + 161.04 R_1 + 46.617 F_1 F_1$  $+ 12.289 F_1R_1 + 40.675 R_1R_1 - 387.55 R_1v_1$  $+3.0365 R_1MR_1 - 23.103 v_1v_1 + 26.377 F_2$  $-338.06 R_2 + 20.681 v_2 - 119.55 F_2F_2$  $-51.866 F<sub>2</sub>R<sub>2</sub> + 93.790 F<sub>2</sub>v<sub>2</sub> - 6.3687 F<sub>2</sub>MR<sub>2</sub>$  $-81.079 R_2R_2 + 971.80 R_2v_2 - 10.142 R_2MR_2$  $+56.138$   $v_2v_2 - 42.203 F_3 + 17.351 v_3$  $-0.46576 \text{ MR}_3 - 55.468 F_3F_3 - 37.592 F_3R_3$  $+132.59 F_3v_3 - 2.9597 F_3MR_3 + 72.538 R_3v_3$  $-1.3890 R_3 MR_3 - 1.7734 F_4 - 1.2537 R_4$  $+2.8366 F_5 + 1.0181 R_5 - 0.075999 M R_5$  $+2.1966 F_6 + 5.1689 R_6 + 3.3826 F_8 + 2.5651 R_8$  $+2.7794 v_8 - 0.13394 MR_8$ 

 $(2)$ 



Fig. 2. Correlation between 1576 pairs of observed and predicted  $\beta$ signals.  $r = 0.979$ 

The correlation of observed versus predicted signals, shown in Fig. 2, has a correlation coefficient of  $r = 0.979$ with a standard deviation of  $s = 2.3$  ppm. The average absolute error of prediction is 1.5 ppm. Standard errors of estimate for the coefficients as well as  $p$  values are given in Table 3. The F value for the model was 1000.0 which yields a p value under 0.0001. For  $\beta$  signals the variables in position 7 are not statistically significant, and thus the substituent in position 7 has a negligible effect on the signal from the carbon in position 2.

The take-one-out cross-validation procedure yields  $SSE = 7810$  and PRESS = 8479. This represents only a small increase in SSE for the pseudovalues.

The 394 compounds used in this study are given in Table 4. This table lists the substituents in positions 1–8 as well as the observed and predicted NMR signals from these positions. Because Table 4 contains a large amount of data, for brevity we omitted signals 9 and 10, which we were unable to predict and encoded the compounds with a letter code rather than with the group numbers. The group numbers are also retained in this paper for consistency with our earlier papers in which we used these numbers.

## 3.3 Comparison of Taft and Charton steric factors in the above models

The advantage of the parameter method outlined above is that the number of variables is kept to a minimum; however there are many groups for which good values of the Charton steric factor is not known  $[17-23]$ . The Taft

Table 3. Variables, coefficient estimates, standard errors of estimates, and  $p$  values of the estimates for the  $\beta$ -signal prediction model

Variable	Coefficient	Standard error	<i>p</i> value
Intercept	126.537797	0.11	0.0001
$F_1$	$-17.414998$	2.66	0.0001
$R_1$	161.035505	5.62	0.0001
$F_1F_1$	46.617338	5.02	0.0001
$F_1R_1$	12.289252	4.66	0.0085
$R_1R_1$	40.675068	3.55	0.0001
$R_1v_1$	$-387.548948$	13.36	0.0001
$R_1MR_1$	3.036502	0.19	0.0001
$v_1v_1$	$-23.102638$	1.05	0.0001
F <sub>2</sub>	26.377376	7.27	0.0003
$R_2$	$-338.059053$	20.81	0.0001
$v_2$	20.680575	1.41	0.0001
$F_2F_2$	$-119.548130$	18.09	0.0001
$F_2R_2$	$-51.865701$	7.02	0.0001
$F_2v_2$	93.790184	34.16	0.0061
$F_2MR_2$	$-6.368708$	1.03	0.0001
$R_2R_2$	$-81.078698$	5.66	0.0001
$R_2v_2$	971.799037	62.87	0.0001
$F_2MR_2$	$-10.142096$	0.65	0.0001
$v_2v_2$	56.137944	4.35	0.0001
$F_3$	$-42.203092$	4.62	0.0001
v <sub>3</sub>	17.350969	2.13	0.0001
MR <sub>3</sub>	$-0.465757$	0.14	0.0013
$F_3F_3$	$-55.467863$	7.95	0.0001
$F_3R_3$	$-37.591915$	4.90	0.0001
$F_3v_3$	132.587760	8.21	0.0001
$F_3MR_3$	$-2.959668$	0.43	0.0001
$R_3v_3$	72.537843	4.80	0.0001
$R_3MR_3$	$-1.388971$	0.21	0.0001
$F_4$	$-1.773441$	0.39	0.0001
$R_4$	$-1.253745$	0.39	0.0016
$F_5$	2.836571	0.50	0.0001
$R_5$	1.018114	0.42	0.0157
$MR_5$	$-0.075999$	0.02	0.0040
$F_6$	2.196643	0.47	0.0001
R <sub>6</sub>	5.168925	0.38	0.0001
$F_8$	3.382608	0.57	0.0001
$R_8$	2.565087	0.42	0.0001
$v_8$	2.779425	0.92	0.0026
$MR_8$	$-0.133938$	0.05	0.0177

parameter,  $E_s$ , is known for a different set of groups [24], and it has been reported that a linear relationship exists between the two parameters [25]. This relationship is given by

$$
E_s = -2.041480 \, v - 0.288578 \quad (r = 0.98)
$$
 (3)

We examined the two prediction equations given above with values of  $E_s$  used in place of v, and we found that although the coefficients changed in the equations, the predictive ability of the models was unaltered from those given above. As the  $\nu$  values are known for some groups and  $E_s$  for others, this refinement should lead to the use of larger databases in future studies of this type.

Acknowledgements. We thank the Auburn University at Montgomery Grant-in-Aid program, together with the Chemistry and Mathematics Departments for support.



#### References

- 1. Jetton RE, Nanney JR, Mahaffy CAL (1996) Electron J Theor Chem 2:24
- 2. Jetton RE, Nanney JR, Mahaffy CAL (1994) In: Leszczyuski J Abstracts from the 3rd Conference on Current Trends in Computational Chemistry. p 97
- 3. Jetton RE, Nanney JR, Mahaffy CAL, Price CC (1995) Struct Chem 6:301
- 4. Mahaffy CAL, Nanney JR (1994) J Fluorine Chem 67:67
- 5. Nanney JR, Mahaffy CAL (1994) J Fluorine Chem 68:181
- 6. Mahaffy CAL, Nanney JR (1994) J Fluorine Chem 68:165
- 7. Jetton RE, Nanney JR, Mahaffy CAL (1995) J Fluorine Chem 72:121
- 8. Nanney JR, Jetton RE, Mahaffy CAL (1996) J Mol Struct (Theochem) 388:19
- 9. Jetton RE, Nanney JR, Mahaffy CAL (1995) Abstracts from the 4th Conference on Current Trends in Computational Chemistry. p 97
- 10. Jetton RE, Nanney JR, Mahaffy CAL (1996) J Mol Struct (Theochem) 388:139
- 11. Jetton RE, Nanney JR, Mahaffy CAL (1997) Abstracts from the 6th Conference on Current Trends in Computational Chemistry
- 12. Jetton RE, Nanney JR, Mahaffy CAL (1998) J Mol Struct (Theochem) 429:29
- 13. Nanney JR, Traylor AK, Mahaffy CAL (1993) J Fluorine Chem 64:217
- 14. Nanney JR, Jetton RE, Mahaffy CAL (1996) J Fluorine Chem 80:105
- 15. Armstrong B, Grier A, Hamilton JB, Khuu H, Mahaffy CAL, Rawlings J, Nanney JR (1992) Inorg Chim Acta 191:189
- 16. Nanney JR, Mahaffy CAL (1992) Inorg Chim Acta 201:55 17. Hansch C, Leo A, Unger SH, Kim KH, Nikaitani D, Lien EJ
- (1973) J Med Chem 16:1207
- 18. Hansch C, Leo A (1979) Substituent constants for correlation analysis in chemistry and biology. Wiley, New York
- 19. Charton M (1975) J Am Chem Soc 97:1552
- 20. (a) Charton M (1975) J Am Chem Soc 97:3691; (b) Charton M (1975) J Am Chem Soc 97:3694
- 21. Charton M (1976) J Org Chem 41:2217
- 22. Charton M (1977) In: Roche EB (ed) Design of biopharmaceutical properties through prodrugs and analogues. American Pharmaceutical Association, Academy of Pharmaceutical Sciences, Washington, DC pp 228-280
- 23. Hansch C, Leo A, Taft RW (1991) Chem Rev 91:165
- 24. Unger SH, Hansch C (1976) Prog Phys Org Chem 12:91
- 25. Gallo RJ (1983) Prog Phys Org Chem 14:115
- 26. Dalling DK, Ladner KH, Grant DM, Woolfenden WR (1977) J Am Chem Soc 99:7142
- 27. Bulbarela A, Tlahuext H, Morales HR, Cuéllar L, Uribe G, Contreras R (1986) Magn Reson Chem 24:1093
- 28. Doddrell D, Barfield M, Adcock W, Aurangzeb M, Jordan D (1976) J Chem Soc Perkin II 402
- 29 Breitmaier E, Haas G, Voelter W (1979) Atlas of carbon-13 NMR data, Heyden, London
- 30. Adcock W, Kitching W, Alberts V, Wickham G, Barron P, Doddrell D (1977) Org Magn Reson 10:47
- 31. Ernst L (1977) Org Magn Reson 9:35
- 32. Bullpitt M, Kitching W, Adcock W, Doddrell D (1976) J Organomet Chem 116:161
- 33. Bullpitt M, Kitching W, Adcock W, Doddrell D (1976) J Organomet Chem 116:187
- 34. (a) Ernst L (1976) J Magn Reson 22:279; (b) Ernst L (1976) J Magn Reson 22:402
- 35. Kitching W, Bullpitt M, Gartshore D, Adcock W, Kohr TC, Doddrell D, Rae ID (1977) J Org Chem 42:2411
- 36. Bullpitt M, Kitching W, Doddrell D, Adcock W (1976) J Org Chem 41:760
- 37. Hansen PE (1979) Org Magn Reson 12:109
- 38. Seita J, Sandstrom J, Drakenberg T (1978) Org Magn Reson 11:239
- 39. Ernst L (1975) J Magn Reson 20:544
- 40. Ernst L (1975) Z Naturforsch B 30:788
- 41. Ernst L (1975) Z Naturforsch B 30:794
- 42. Ernst L (1975) Chem Ber 108:2030
- 43. Wilson NK, Zehr RD (1978) J Org Chem 43:1768
- 44. Takai H, Odani A, Sasaki Y (1978) Chem Pharm Bull Jpn 26:1966
- 45. van de Weijer P, Mohan C, van den Ham DMW (1977) Org Magn Reson 10:165
- 46. Ewers U, Günther H, Jaenicke L (1975) Angew Chem 87:356
- 47. Kitching W, Bullpit M, Doddrell D, Adcock W (1974) Org Magn Reson 6:289
- 48. Wells PR, Arnold DP, Doddrell D (1974) J Chem Soc Perkin 2:1745
- 49. Wilson NK, Strothers JB (1974) J Magn Reson 15:31
- 50. Hunter DH, Strothers JB (1973) Can J Chem 51:2884
- 51. Mechin B, Richer JC, Odiot S (1980) Org Magn Reson 14:79
- 52. Johns SR, Willing RI (1976) Aust J Chem 29:1617
- 53. Adcock W, Cox DP, Kitching W (1977) J Organomet Chem 133:393
- 54. Claret PA, Osborne AG (1976) Org Magn Reson 8:147
- 55. Ernst L (1976) Org Magn Reson 8:161