Inherent Limitations of Linear Additivity Models for the Estimation of ¹³C-NMR Chemical Shift Values of Polysubstituted Benzenes: Some New Findings Using the SCSD Algorithm

Lingran Chen† and Wolfgang Robien*

Department of Organic Chemistry, University of Vienna, Währinger Strasse 38, A-1090 Vienna, Austria

Received December 23, 1992

A large number of substituent-induced chemical shift difference (SCSD) values have been automatically derived from both mono- and polysubstituted benzenes for 15 different substituents by a new algorithm based on systematic investigation of all the possible structure pairs available within a large database. The average SCSD values were found to be in good agreement with the corresponding SCSD values derived from monosubstituted benzenes. However, very large SCSD ranges were found for almost all functionalities under investigation. The scattered distribution of SCSD values and the approximate feature of additivity equations are two inherent factors which jointly restrict the accuracy of the additivity model method.

INTRODUCTION

Substituent-induced chemical shift differences (SCSD)¹ play an important role in both theroretical studies of the nature of substituent effects on chemical shift values and the practical application to the simulation of NMR spectra of a known molecular structure. The SCSD value is simply defined as the difference of the chemical shift values between two corresponding compounds differing by a substituent under investigation. Extensive effort has been directed toward the analysis of substituted benzenes, and a large number of SCSD values for monosubstituted benzene derivatives (SCSD_{mo}) have been collected for a wide range of substituents. 2 These $SCSD_{mo}$ values have widely been used to estimate chemical shifts of polysubstituted benzene derivatives.^{3,4} Some computer programs based on that SCSD_{mo} table have also been written in order to perform automatic estimation of ¹³C-NMR spectra. ^{5,6} However, although it has been noted that the main problem of additivity models is the limited accuracy as compared to other methods, the real reasons of the problem have not been completely clarified so far.⁷

The application of our novel SCSD algorithm¹ allows automatic extraction and analysis of the SCSD values from all of the substituted benzenes available in a large data collection for different kinds of substituents under investigation. In this paper we report some new findings by using SCSD method and investigate the reasons for errors inherent to linear additivity models.

EXPERIMENTAL SECTION

The databases accessed consist of some 80 000 ¹³C-NMR spectra, including the libraries of the University of Vienna, SADTLER Research Laboratories, and the German Cancer Research Center at Heidelberg. The SCSD algorithm was implemented in FORTRAN 77 under the UNIX operating system on a Silicon Graphics workstation and an IBM-R6000 workstation. The program consists of about 5000 lines of source code. First, all those compounds which contain exactly one six-membered ring system were retrieved from the available databases, and then a partial structure search was performed using benzene as query structure. Finally, the SCSD algorithm was applied to this hit list for each substituent under investigation.

RESULTS AND DISCUSSION

Relationship between SCSD_{mo} and SCSD_{av}. The accuracy of chemical shift estimation using additivity models depends heavily upon the deviations between SCSD values considered and their alternative SCSD_{mo} values really used during calculation. In order to improve the estimation accuracy, a new set of increments derived from monosubstituted benzenes has recently been published.⁴ In this section, we investigate the relationship between SCSD_{mo} values and their corresponding average SCSD values (SCSD_{av}) calculated from all the possible SCSD values derived from both mono- and polysubstituted benzene derivatives. Table I gives the statistic information of the obtained SCSD values for 15 different substituents, including SCSD_{av} values, the highest and the lowest SCSD values, and their corresponding structure pair inforamtion allowing easy access to the reference data sets (see Figure 1). The SCSD_{mo} values from ref 4 are also listed in the right column of Table I for ease of comparison.

By inspection of Table I, we can easily find the interesting fact that almost all the $SCSD_{av}$ values are surprisingly consistent with their corresponding $SCSD_{mo}$ values. For fluorine substituent, for instance, the absolute differences between $SCSD_{av}$ and $SCSD_{mo}$ values are less than 0.5 ppm for all carbon atoms in the benzene ring. In only a few cases, the differences between $SCSD_{av}$ and $SCSD_{mo}$ are larger than 1 ppm. The largest difference (2.7 ppm) has been found for a thiomethyl substituent at the ipso-position.

This fact seems to tell us that using $SCSD_{mo}$ values for estimation of $^{13}C\text{-NMR}$ spectra of polysubstituted benzenes is suitable. If this is true, we can further deduce, from the viewpoint of statistics, that in those cases where $SCSD_{av}$ and $SCSD_{mo}$ values have larger differences, using $SCSD_{av}$ instead of $SCSD_{mo}$ values to estimate chemical shifts should, in most situations, lead to better results because more SCSD values are located within the range with the $SCSD_{av}$ value as the center than those within the identically sized range with the $SCSD_{mo}$ value as the center.

However, when checking Table I again, we can find another fact that there are very large SCSD ranges for most of the substituents. The SCSD value of ipso-carbon atom for iodine substituent has the largest range of over 55 ppm, from +11.6 ppm for structure pair 20445A-5665A to -43.6 ppm for structure pair 9512B-8812B (see Figure 1). From Table I it can be seen that SCSD ranges depend upon the number of SCSD values. Usually more occurrences of SCSD values

[†] On leave from the University of Science and Technology of China, Hefei, Anhui 230026, The People's Republic of China.

Table I. Statistic Data of the Average, Highest, and Lowest SCSD Values and Their Corresponding Structure Pair Information for 15 Substituents Obtained by SCSD Algorithm^a

| | | | differences of | chemical shifts | in each position | n | | | |
|-------------------------|------------------|----------------------------|------------------------|-----------------|------------------|---------------------------------------|---------------------------------|--|--|
| positions considered | highest value | entries used 1-2 | average values | no. of hits | lowest values | entries used 1-2 | SCSD _{mo} ⁴ | | |
| | | <u> </u> | 1 | | | | 3 COD mo | | |
| 1 | 43.5 | 1096F-7257C | 33.7 | 2744 | 7.2 | 20446A-5705A | 33.6 | | |
| 2 | 0.8 | 25988B-14441A | -12.9 | 2745 | -30.5 | 11852A-14337A | -13.0 | | |
| 2 3 | 6.1 | 4810B-22571A | 1.4 | 2746 | -7.8 | 6972B-954C | 1.6 | | |
| 4 | 18.0 | 16771B-6517B | -4.0 | 2746 | -30.1 | 25988B-14441A | -4.4 | | |
| 5 | 6.3 | 1988A-2119A | 1.5 | 2744 | -7.7 | 17920A-970B | | | |
| 5 6 | 0.8 | 25988B-14441A | -13.1 | 2745 | -21.9 | 19345B-1988A | | | |
| | | | 2 | Cl | | | | | |
| 1 | 18.2 | 1097F-7256C | 5.1 | 4763 | -5.0 | 12142B-2095A | 5.3 | | |
| 2 | 16.1 | 4675B-28569B | -0.9 | 4753 | -14.0 | 241B-6127A | 0.4 | | |
| 3 4 | 11.8 | 3920A-3397B | 1.1 | 4750 | -10.2 | 16692 B -8816B | 1.4 | | |
| 4 | 9.8 | 12203B-8521B | -1.7 | 4771 | -13.4 | 594F-2631B | -1.9 | | |
| 5 6 | 9.3 | 3920A-3397B | 0.9 | 4746 | -14.1 | 3008B-7951C | | | |
| 6 | 5.6 | 11543B-2361A | 0.1 | 4757 | -12.6 | 594F-2631B | | | |
| | | | 3 | | | . | | | |
| 1 | 8.4 | 3336B-678A | -6.0 | 2822 | -20.2 | 21438B-3520A | -5.4 | | |
| 2 3 | 16.2 | 10347C-10338C | 1.8 | 2843 | -11.0 | 8601B-2208B | 3.3 | | |
| 3 | 17.3 | 8601B-4090A | 1.3 | 2841 | -14.2 | 10347C-10338C | 2.2 | | |
| 4 | 2.7 | 16303B-7401B | -1.3 | 2856 | -5.2 | 23757B-6009A | -1.0 | | |
| 5 | 6.1 | 1100F-7255C | 1.4 | 2839 | 5.3 | 8601B-2208B | | | |
| 6 | 7.8 | 3445C-808B | 3.0 | 2840 | -6.5 | 18079A-18078A | | | |
| é | 11.7 | 204454 56654 | 4. – | | 12.6 | 05100 00100 | 21.0 | | |
| 1 | 11.6 | 20445A-5665A | -33.5 | 1934 | -43.6 | 9512B-8812B | -31.2 | | |
| 2 3 | 17.5 | 15108A-5729B | 7.2 | 1930 | -3.9 | 3718B-20927A | 8.9 | | |
| 3 | 11.4 | 6988B-9515B | 1.2 | 1933 | -7.3 | 15146A-677A | 1.6 | | |
| 4 | 8.4 | 21423A-9296C | -0.6 | 1932 | -4 .8 | 15395A-10303B | -1.1 | | |
| 5 | 12.6 | 9512B-8812B | 1.4 | 1934 | -9.8 | 6988B-9515B | | | |
| 6 | 15.7 | 3443C-808B | 9.2 | 1929 | -11.3 | 17625B-1601H | | | |
| | | | 5. –CH | | | | | | |
| 1 | 13.3 | 3435 B -13578A | 9.3 | 609 | 3.9 | 1513F-1512F | 9.3 | | |
| 2 | 3.8 | 20583B-1084A | 0.0 | 609 | -3.5 | 15314B6031A | 0.3 | | |
| 3 | 6.9 | 4226B-8413B | 0.4 | 609 | -4.5 | 3435B-1544F | 0.2 | | |
| 4 | 3.9 | 16653 B -366A | 0.3 | 609 | -4.8 | 4226B-8413B | 0.0 | | |
| 5 6 | 2.8 | 8413B-2345A | 0.3 | 608 | -2.2 | 16653B-6009A | | | |
| 6 | 4.3 | 20583B-454K | 0.3 | 609 | -5.7 | 4226B-8413B | | | |
| | | #011D 105E01 | 6. –CH | | 4.0 | 244427 (264 | | | |
| 1 2 | 13.6 | 7011B-13578A | 9.6 | 587 | 4.3 | 25443B-6060A | 9.5 | | |
| 2 | 3.6 | 24271B-1815H | 0.5 | 587 | -3.4 | 7011B-1544F | 0.7 | | |
| 3 | 3.6 | 7063C-2119A | 0.5 | 587 | -4 .2 | 7011B-1544F | 0.3 | | |
| 4 | 3.0 | 23259A-22908A | 0.1 | 587 | -3.4 | 7367C-7360C | 0.2 | | |
| 5 6 | 2.8 | 3531B-2119A | 0.5 | 586 | -2.8 | 7400C-7360C | | | |
| 6 | 4.0 | 7367C-7361C | 0.7 | 587 | -2.2 | 24111B-6927A | | | |
| | 46.0 | 44155D 00/4D | 7. –0 | | | 07040D 4744D | ••• | | |
| 1 | 46.0 | 22177B-8963B | 28.5 | 5426 | 6.1 | 27252B-1731B | 28.8 | | |
| 2 3 | 2.2 | 16771B-18633B | -13.2 | 5414 | -30.9 | 15349B-19820B | -12.8 | | |
| 3 | 17.9 | 8001B-5729B | 1.4 | 5424 | -12.8 | 10382B-4687C | 1.4 | | |
| 4 | 15.6 | 24691B-1253F | -8.3 | 5404 | -32.3 | 22177B-8963B | -7.4 | | |
| 5 6 | 15.3 1.4 | 837A-900B 864A-9119B | 1.5 -12.7 | 5429 5410 | -14.0 -23.9 | 21481B-4570B 8001B-571A | | | |
| 0 | 1.4 | 004A-7117D | | | -23.9 | 0001B-3/1A | | | |
| 1 | 45.4 | 22019B-2367A | 8. –O – 31.3 | CH₃ 4650 | 17.3 | 13158 B -11750 B | 33.5 | | |
| 1 2 | 16.5 | 2348A-7259C | -12.7 | 4622 | -23.5 | 6434B-18072A | -14.4 | | |
| 3 | 22.1 | 8963B-2367A | 1.0 | 4629 | -23.3 -12.3 | 17403B-17395B | 1.0 | | |
| 4 | 16.7 | 24998B-24839B | _7.7 | 4642 | -12.5 -27.5 | 7644B-2348A | -7.7 | | |
| 5 | 16.7 | 13158B-11750B | 1.0 | 4621 | -12.3 | 17403B-17395B | -7.7 | | |
| 6 | 8.5 | 8963B-6798B | -15.4 | 4649 | -31.7 | 21400B-18063A | | | |
| _ | | | 9O-CI | | | | | | |
| 1 | 42.1 | 3839B-800B | 30.6 | 769 | 23.0 | 4715B-6060A | 30.3 | | |
| 2 | -4.9 | 2295B-2852B | -13.0 | 768 | -17.8 | 2852B-695C | -14.3 | | |
| 3 | 12.8 | 3839B-800B | 1.1 | 768 | -10.2 | 7183B-577A | 0.6 | | |
| 3 4 5 | 2.1 | 21420A-9296C | -7.6 | 767 | -11.1 | 273B-6060A | -8.4 | | |
| | 8.7 | 3839B-800B | 1.0 | 768 | -3.0 | 6395B-3525B | | | |
| 6 | -5.0 | 7183 B –11538A | -14.6 | 768 | -23.1 | 7854 B -684A | | | |
| _ | 10SH | | | | | | | | |
| 1 2 3 | 11.3 | 26546B-366A | 2.9 | 445 | -4.3 | 13037A-2275A | 4.0 | | |
| 2 | 19.8 | 27918A-5589A | -0.2 | 441 | -20.1 | 6466B-8560A | 0.7 | | |
| | 3.9 | 26546B-366A | 0.5 | 445 | -4.1 5.6 | 15865A-1088F | 0.3 | | |
| 4 | 4.8 | 27918A-5589A | -2.5 0.4 | 445 | -5.6 | 24824B-6153A | -3.2 | | |
| 5 6 | 3.0 18.6 | 13037A-22569A 6466B-50K | 0.4 1.7 | 441 445 | −3.0 −7.4 | 18221 B -8568A 27918A-5589A | | | |
| O | 0.61 | 0400D-30K | 1.7 | 443 | - / .4 | 21710A-JJ07A | | | |

Table I (Continued)

| positions considered | differences of chemical shifts in each position | | | | | | | | |
|-------------------------|---|--------------------------------|-------------------------|----------------|------------------|-----------------------|--------------------|--|--|
| | highest value | entries used 1-2 | average values | no. of hits | lowest values | entries used 1-2 | SCSD _{mo} | | |
| | | | 11. -S - | CH, | | | | | |
| 1 | 18.4 | 16143B-450K | 9.8 | 1179 | -16.8 | 1253F-2280A | 7.1 | | |
| 2 | 7.6 | 1252F-5300C | -1.2 | 1200 | -7.7 | 1545F-1312H | -1.9 | | |
| 3 | 4.4 | 16143B-450K | 0.5 | 1199 | -5.0 | 1270F-1088F | 0.2 | | |
| 4 | 9.4 | 16968B-6714B | -2.5 | 1171 | -6.8 | 7389C-7360C | -3.6 | | |
| 5 | 6.7 | 20914B-2068A | 0.6 | 1198 | -3.0 | 7423C-7360C | 5.0 | | |
| 6 | 7.6 | 1252F-5300C | -0.8 | 1200 | -7. 4 | 15755A-19987B | | | |
| U | 7.0 | 12321-3300C | | | -/. - | 13733A-19907 D | | | |
| | 22.6 | 252004 225504 | 12. –CH= | | 2.4 | 07334 017 | | | |
| 1 | 22.6 | 25280A-22569A | 8.4 | 1206 | 3.4 | 9733A-91B | 6.4 | | |
| 2 | 1.9 | 12245B-22560A | -2.2 | 1206 | -6.0 | 12433B-6058A | -2.3 | | |
| 3 | 4.1 | 25309A-2119A | -0.3 | 1207 | -3.1 | 25298A-8573A | -0.1 | | |
| 4 | 3.0 | 12075B-22559A | -1.5 | 1206 | -4.5 | 9744A-6068A | -0.8 | | |
| 5 | 2.9 | 12245B-364A | -0.3 | 1206 | -3.0 | 2357A-6068A | | | |
| 6 | 1.9 . | 9735A-1389H | -2.2 | 1206 | -4.8 | 25310A6009A | | | |
| | | | 13. –C≡ | ≡CH | | | | | |
| 1 | -1.3 | 4426C-5325C | -5.8 | 346 | -8.3 | 4429C-1312H | -6.2 | | |
| 2 | 6.9 | 3404C-1815H | 3.4 | 346 | -0.3 | 54K-8560A | 3.6 | | |
| 3 | 2.3 | 1991A-5293C | -0.2 | 346 | -3.5 | 4426C-1088F | -0.4 | | |
| 4 | 2.1 | 4427C-5325C | 0.3 | 346 | -3.1 | 27755 B -7360C | -0.3 | | |
| 5 | 2.3 | 1991A-5293C | -0.1 | 346 | -2.8 | 3404C-1815H | | | |
| 6 | 6.9 | 3404C-1815H | 4.0 | 346 | 1.6 | 17769B-9445B | | | |
| | | | 14. –C0 | ЮН | | | | | |
| 1 | 16.2 | 9785B-13827A | 1.7 | 2392 | -6.9 | 19349B-6495C | 2.1 | | |
| 2 | 19.8 | 8403B-7794B | 1.4 | 2392 | -7.0 | 19071B-9785B | 1.6 | | |
| 3 | 12.0 | 8325B-7078B | 0.1 | 2389 | -8.3 | 4480B-6503C | -0.1 | | |
| 4 | 12.4 | 19348B-1675H | 3.9 | 2394 | -5.4 | 494K-1319H | 5.2 | | |
| 5 | 8.7 | 24109B-8011B | -0.3 | 2392 | -15.5 | 15349B-1225B | J. L | | |
| 6 | 19.8 | 8403B-7794B | 1.4 | 2389 | -9.8 | 4560B-21674B | | | |
| Ü | 17.0 | 01000 11710 | | | 7.0 | 10002 2107 12 | | | |
| 1 | 33.3 | 8924B-6135B | 15. –N 19.4 | 4635 | -1.3 | 1075H-1074H | 19.9 | | |
| 1 | 33.3 30.9 | 26557B-364K | -4.5 | 4674 | -1.3 -17.7 | 8566B-4630B | -4.9 | | |
| 2 3 | | 19129B-9522B | -4 .5 1.2 | 4674 4670 | -17.7 -21.2 | 17625B-15148A | -4.9 0.9 | | |
| 3 | 18.6 | | | | | | | | |
| 4 | 15.5 | 8454B-6495C | 5.5 | 4676 | -16.0 | 26557B-364K | 6.1 | | |
| 5 | 15.2 | 15338A-15908A | 0.5 | 4664 | -20.8 | 11852A-15329A | | | |
| 6 | 17.9 | 21481 B -11363 B | -4.4 | 4677 | -18.1 | 6616B-14826A | | | |

The data in the right column are cited from ref 4. The numbering of the carbon atoms in the benzene ring is shown on structure B in Figure 3.

Table II. Chemical Shift Deviations Resulting Mainly^a from Substituent Interactions

| | chemical of carbo the hal | | |
|-----------------|---------------------------------|---------------------------|-----------|
| compound | calculated ^a | experimental ^a | deviation |
| o-fluoroanisole | 148.67 a | 152.90 a | -4.2 |
| m-fluoroanisole | 163.27 c | 160.10 b | 3.2 |
| p-fluoroanisole | 155.27 a | 157.40 a | -2.1 |
| o-chloroanisole | 119.97 a | 121.20 a | -1.2 |
| m-chloroanisole | 135.37 a | 134.90 a | 0.5 |
| p-chloroanisole | 126.57 a | 125.50 a | 1.1 |
| o-bromoanisole | 108.32 a | 111.70 a | -3.4 |
| m-bromoanisole | 123.50 d | 122.80 d | 0.7 |
| p-bromoanisole | 114.92 a | 112.58 a | 2.3 |
| o-iodoanisole | 80.12 a | 86.00 a | -5.9 |
| m-iodoanisole | 95.30 d | 94.30 d | 1.0 |
| p-iodoanisole | 86.84 c | 82.85 c | 4.0 |

^a The letter indicates the solvent: $a = \text{chloroform-}d_1$; b = DMSO; $c = DMSO \cdot d_6$; $d = CDCl_3$.

enlarge the corresponding SCSD range. These facts demonstrate that using $SCSD_{mo}$ values to predict chemical shifts for polysubstituted benzene derivatives is a very dangerous procedure.

In view of the above mentioned findings, it would be helpful to establish a criterion for checking when SCSD_{mo} values can be properly used in spectrum simulation. Unfortunately, it has been pointed out⁵ that such a criterion does not exist. Thus, we would like to know how many spectrum estimation

problems can be solved by using $SCSD_{mo}$ values. This problem will be discussed in detail in the following paragraphs.

Distribution of SCSD Values. Parts of the individual distribution diagrams of SCSD values for 15 different substituents are given in Figure 2.

The visual inspection of the various distribution diagrams shows that in most cases SCSD values have the distributions of approximate Gaussian-curve shape. The peak maximum is usually located at or near the corresponding $SCSD_{av}$ or $SCSD_{mo}$ values.

The scattered distributions of SCSD values are mainly caused by various substituent interactions. Suppose all of the spectra used in calculation were measured under identical experimental conditions and there would be no interaction between the substituents within the benzene ring; then the SCSD values corresponding to the same carbon atoms at the benzene ring should be exactly identical for one substituent under investigation, independent of the calculation of the SCSD values from mono- or polysubstituted benzenes. However, there exist various interactions between substituents located at the ring system. It is not appropriate to discuss these effects in detail here; this topic has been reviewed elsewhere.² Some illustrative examples shown in Table II can help us better understand the interactions of substituents. For disubstituted benzenes, the deviations between calculated and experimental chemical shift values directly reflect the extent of interaction between two substituents at the benzene ring.

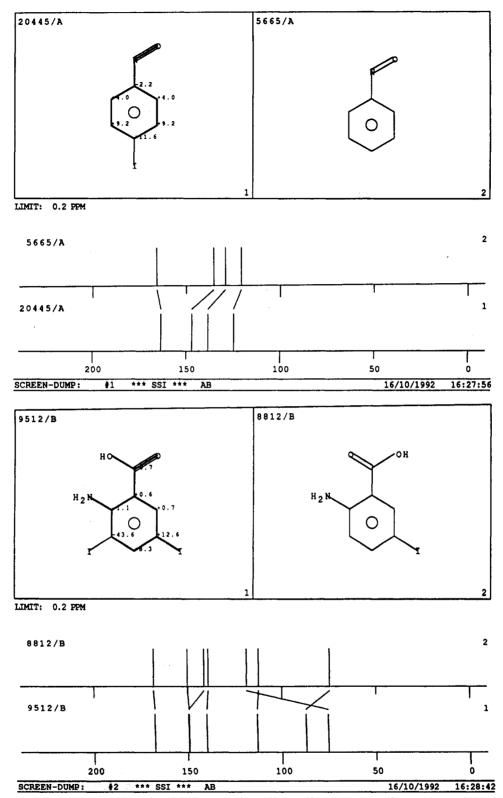


Figure 1. Two structure pairs which give the largest SCSD range of over 55 ppm, from +11.6 ppm for structure pair 20445A-5665A to -43.6 ppm for structure pair 9512B-8812B for the ipso-carbon in the benzene ring. The maximal common substructural fragment (MCSS) is marked by bold lines.

From Table II, it can be seen that, for all three fluoroanisoles, the interactions between fluorine and the methoxy group are quite different, depending upon the substitution pattern; the other halogen-anisoles show a similar behavior. On the other hand, for the same substitution pattern, different halogen anisoles also show different interactions between halogen atom and methoxy group. Besides, the differences of experimental conditions used in the measurement of spectra also contribute the scattering of SCSD values to some minor extent.

From distribution diagrams we can roughly estimate that about 50% of all SCSD values are within the range of SCSD $_{mo}$

 ± 1 ppm. This fact seems to indicate that SCSD_{mo} values can be properly applied to some 50% simulation problems for substituted benzenes. But the real situation is more complicated. In order to clarify this point we should first investigate the inherent problem of additivity models.

Limitation of Linear Additivity Models. As has been most recently pointed out by Pretsch et al., 7 the principal drawback of the additivity models is their limited accuracy as compared to other approaches. In this section the inherent problem of this method will be discussed.

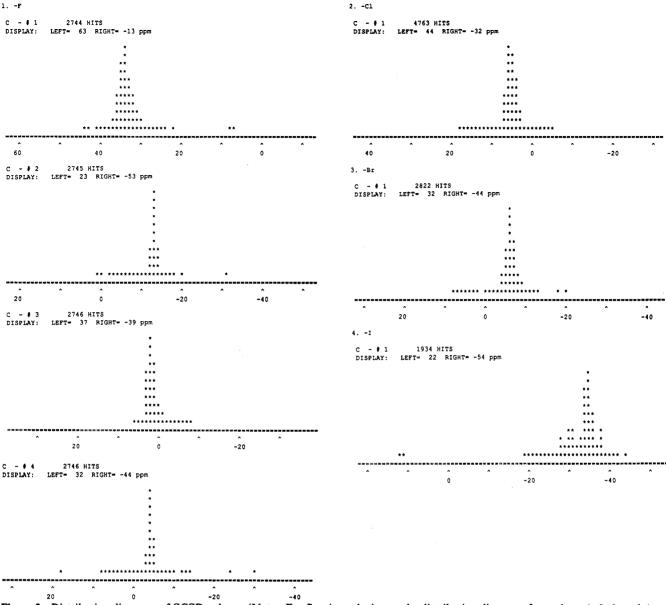


Figure 2. Distribution diagrams of SCSD values. (Note: For fluorine substituent the distribution diagrams for carbons 1, 2, 3, and 4 are given; for chlorine, bromine, and iodine substituents only, one distribution diagram corresponding to the ipso-carbon atom at benzene ring is shown.)

First, we consider disubstituted benzenes. In order to estimate the chemical shift value of carbon 1 of structure A shown in Figure 3, we can deduce a mathematically accurate equation. The mathematical description of the SCSD definition is given as follows:

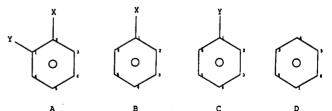
$$\Delta \delta^{h,k}_{i,j} = \delta^h_{i} - \delta^k_{j} \tag{1}$$

where, $\Delta \delta^{h,k}_{i,j}$ is the SCSD value of carbon h of structure i and its corresponding carbon k of structure j, δ^{h}_{i} and δ^{k}_{i} represent the chemical shift values of carbons h and k of structures i and j, respectively. Thus, we can easily write the following expressions:

$$\Delta \delta^{1,6}_{A,B} = \delta^1_A - \delta^6_B \tag{2}$$

$$\Delta \delta^{6,3}_{B,D} = \delta^6_{B} - \delta^3_{D} \tag{3}$$

$$\Delta \delta^{1,4}_{C,D} = \delta^1_C - \delta^4_D \tag{4}$$



Four structures used in deducing eqs 1-6. X, Y: Figure 3. Substituents.

From definitions 2 and 3, we obtain

$$\delta^{1}_{A} = \delta^{3}_{D} + \Delta \delta^{6,3}_{BD} + \Delta \delta^{1,6}_{AB}$$
 (5)

Equation 5 is accurate because no hypothesis was introduced during the above deduction process. However, when we try to use this equation, the problem immediately occurs. Because of the lack of a δ^{1} _A value (which is just what we want to calculate), the $\Delta \delta^{1,6}_{A,B}$ value is in fact unknown. Considering that both $\Delta \delta^{1,6}{}_{A,B}$ and $\Delta \delta^{1,4}{}_{C,D}$ are SCSD values of the ipsocarbon atom for the substituent Y, we may use the latter to

estimate $\Delta \delta^{1.6}_{A,B}$; thus we get

$$\delta_{A}^{1} = \delta_{D}^{3} + \Delta \delta_{B,D}^{6,3} + \Delta \delta_{C,D}^{1,4}$$
 (6)

It can be easily seen that eq 6 is exactly the same as the additivity model presented in ref 4 for solving the problem except that 128.5 ppm has been selected as the value of δ^3_D in the literature. Therefore, from the above deduction process we know that the additivity model for carbon 1 of structure A is in fact an approximate equation which contains one approximate term ($=\Delta\delta^{1,4}_{C,D}$). This conclusion is also tenable for any other carbon in the benzene ring of disubstituted benzenes

Now it is clear that, for disubstituted benzenes, the deviation of calculated shift values directly depends upon the difference between the SCSD value derived from disubstituted benzene and its alternative SCSD mo value really used in the calculation. For the above example, for instance, the deviation of calculated shift value via eq 6 is solely determined by $\Delta \delta^{1,4}{}_{C,D} - \Delta \delta^{1,6}{}_{A,B},$ provided the reference shift data are free from a variety of experimental conditions. Therefore, taking ± 1 ppm as the threshold values of the largest acceptable deviations in shift simulation, we can reasonably conclude that $SCSD_{mo}$ values can be properly applied to some 50% simulation problems for disubstituted benzenes.

For polysubstituted benzenes, however, the situation is more complicated. It can be proved in a similar manner that for an *n*-substituted benzene $(n \ge 3)$, there are n-1 approximate terms in the corresponding additivity equation. Thus, the total deviation between the calculated shift value and the experimental one is equal to the sum of deviation contributions of n-1 approximate terms. In the worst case, all the individual deviations have the same sign. Therefore, even in those cases where all the individual absolute deviation values are less than 1 ppm, the total absolute deviation value may be greater than 1 ppm. On the other hand, in some cases, although all the individual absolute deviation values are larger than 1 ppm, the total absolute deviation value may be less than 1 ppm because of the counteractions of the individual deviations with different signs. For most of the polysubstituted benzenes, the situation is between the above two extreme cases. Therefore, it is difficult to estimate the application scope of SCSD_{mo} values for polysubstituted benzenes. However, using several approximate terms in an additivity equation without any correction for substituent interaction is by no means a reliable method.

CONCLUSION

A large number of SCSD values derived from both monoand polysubstituted benzenes were obtained for 15 different substituents by means of the SCSD approach. For most carbon atoms, their SCSD values have a distribution similar to a Gaussian-curve shape. The SCSD_{av} values show good consistencies with the corresponding SCSD_{mo} values. For most substituents, there exist considerably large ranges of SCSD values.

Linear additivity models are approximate equations containing n-1 approximate terms for n-substituted benzenes. This fact, together with the scattered distribution of SCSD values, determines that the method based on the linear additivity equations described in the chemical literature has the inherent disadvantage of limited accuracy.

Better results can be achieved by using a new approach which has the ability to utilize automatically SCSD values derived from polysubstituted benzenes.⁸⁻¹⁰

ACKNOWLEDGMENT

L.C. thanks the Austrian Academic Exchange Service for the research fellowship. This project was supported by the European Academic Supercomputing Initiative (EASI).

REFERENCES AND NOTES

- Chen, L.; Robien, W. Sophisticated Algorithm for Automatic Extraction and Analysis of Substituent-Induced Chemical Shift Differences of ¹³C-NMR Spectra. J. Chem. Inf. Comput. Sci., preceding paper in this issue
- (2) Ewing, D. F. ¹³C Substituent Effects in Monosubstituted Benzenes. Org. Magn. Reson. 1979, 12, 499-524.
- (3) Pretsch, E.; Clerc, J. T.; Seibl, J.; Simon, W. Tables of Spectral Data for Structure Elucidation of Organic Compounds, 2nd ed.; Springer-Verlag: 1989.
- (4) Pretsch, E.; Fürst, A.; Robien, W. Parameter Set for the Prediction of the ¹³C-NMR Chemical Shifts of sp²- and sp-Hybridized Carbon Atoms in Organic Compounds. *Anal. Chim. Acta* 1991, 248, 415-428.
- (5) Fürst, A.; Pretsch, E. A Computer Program for the Prediction of ¹³C-NMR Chemical Shifts of Organic Compounds. *Anal. Chim. Acta* 1990, 229, 17-25.
- (6) Hearmon, R. A.; Liu, H.-M.; Laverick, S.; Tayler, P. Microcomputer Prediction and Assessment of Substituted Benzene ¹³C NMR Chemical Shifts. Magn. Reson. Chem. 1992, 30, 240-248.
- (7) Pretsch, E.; Fürst, A.; Badertscher, M.; Bürgin, R.; Munk, M. E. C13-SHIFT: A Computer Program for the Prediction of ¹³C NMR Spectra Based on an Open Set of Additivity Rules. *J. Chem. Inf. Comput. Sci.* 1992, 32, 291-295.
- (8) Chen, L.; Robien, W. Optimized Prediction of ¹³C-NMR Spectra Using Increments. Anal. Chim. Acta 1993, 272, 301-308.
- (9) Chen, L.; Robien, W. Optimized Prediction of ¹³C-NMR Spectra Using Increments. Comparison with Other Methods. Fresenius Z. Anal. Chem. 1992, 344, 214-216.
- (10) Chen, L.; Robien, W. Submitted for publication.