ACKNOWLEDGMENT

We thank the National Institutes of Health (Grant No. RR-00612-12) for their generous financial support. Computer resources were provided by the SUMEX facility at Stanford University under National Institutes of Health Grant RR-0785.

REFERENCES AND NOTES

(1) Part 44 of the series "Applications of Artificial Intelligence for Chemical Inference". For part 43 see: Lindley, M. R.; Shoolery, J. N.; Smith, D. H.; Djerassi, C., *Org. Magn. Reson.*, in press.
(2) Cone, M. M.; Venkataraghavan, R.; McLafferty, F. W. "Molecular Structure Comparison Program for the Identification of Maximal Common Substructures". *J. Am. Chem. Soc.* **1977**, *99*, 7668–7671.
(3) Varkony, T. H.; Shiloach, Y.; Smith, D. H. "Computer-Assisted Examination of Chemical Compounds for Structural Similarities". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 104–111.
(4) Gund, P.; Wipke, W. T.; Langridge, R. "Computer Searching of a Molecular Structure File for Pharmacophoric Patterns". In "Proceedings International Conference on Computers in Chemical Research and Education"; Elsevier: Amsterdam, 1973; pp 5–33.
(5) Lesk, A. M. "Detection of Three-Dimensional Patterns in Chemical Structures". *Commun. ACM* **1979**, *22*, 219–224.
(6) Rohrer, D. C.; Perry, H. "FITMOL". In "Public Procedures: A Program Exchange for PROPHET Users"; Wood, J. J., Ed.; Bolt, Beranek, and Newman, Inc.: Cambridge, MA, 1978.
(7) Cohen, N. C. "Beyond the 2-D Chemical Structure"; In "Computer Assisted Drug Design"; Olson, E. C., Christoffersen, R. E., Eds.; American Chemical Society: Washington, DC, 1979; pp 377–381.
(8) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. "The Conformational Parameter in Drug Design: The Active Analog Approach"; In "Computer-Assisted Drug Design"; Olson, E. C., Christoffersen, R. E., Eds.; American Chemical Society: Washington, DC, 1979; Chapter 9, pp 205–226.
(9) Avidon, V. V.; Pomerantsev, I. A.; Golender, V. E.; Rozenblit, A. B. "Structure-Activity Relationship Oriented Languages for Chemical Structure Representation". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 207–214.
(10) Nakanishi, K.; Kubo, I. "Studies on Warburganal, Muzigadial, and Related Compounds". *Isr. J. Chem.* **1977**, *16*, 28–31.
(11) Wenger, J. C.; Smith, D. H. "Deriving Three-Dimensional Representations of Molecular Structure from Connection Tables Augmented with Configuration Designations Using Distance Geometry". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 29–34.
(12) Allinger, N. L. "Conformational Analysis. 130. MM2. A Hydrocarbon Force Field Utilizing V_1 and V_2 Torsional Terms". *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.
(13) Harary, F. "Graph Theory"; Addison Wesley: Reading, MA, 1971.
(14) Smith, D. H.; Carhart, R. E.; Crandell, C. W.; Venkataraghavan, R. "Constructive Perception of Shared Three-Dimensional Substructures". "Abstracts of Papers"; 186th National Meeting of the American Chemical Society, Washington, DC, Aug 28–Sept 2, 1983; American Chemical Society: Washington, DC, 1983; CINF No. 3.
(15) Richards, M.; Whitby-Stevens, C. "BCPL – the Language and its Compiler"; Cambridge University Press: Cambridge, 1979.

Carbon-13 Nuclear Magnetic Resonance Spectral Interpretation by a Computerized Substituent Chemical Shift Method†

H. N. CHENG* and S. J. ELLINGSEN

Hercules Incorporated, Research Center, Wilmington, Delaware 19899

Received December 30, 1982

A FORTRAN computer program (called CSHIFT) is developed for the rapid estimation of the ^{13}C NMR chemical shifts of aliphatic organic compounds. The method is based on additive ^{13}C shift relationships, using empirical substituent chemical shift parameters. Examples are given that illustrate its use.

INTRODUCTION

It is generally recognized that ^{13}C NMR spectroscopy is a very powerful tool for organic structure determination. A major task in the interpretation of ^{13}C NMR spectra is to estimate the chemical shifts of compounds known or suspected to be present. Two approaches are generally used: (1) look up the chemical shifts in spectral libraries of either the compound in question or, if not available, compounds with similar structures; (2) calculate the ^{13}C shifts by using empirical substituent chemical shift rules.

For the first approach the spectral collections of Sadtler,[1] Bremser,[2] Breitmaier,[3] and Stothers,[4] among others, are very useful. In the last few years, many computer-assisted structure-determination methods have been developed.[5] Some of the earliest, the CNMR program[6] of Chemical Information Systems and its variants, have been generally available for several years. Recently the Stanford group has developed an array of sophisticated methods.[7] Several other groups are also very active in advancing this important area.[8-15]

In the second approach, there exist empirical rules such as those formulated by Grant and Paul,[15] Lindeman and Adams,[16] and Carman, et al.[17] for hydrocarbons, by Eggert and Djerassi[18] and Sarneski et al.[19] for amines, by Roberts[20] and Ejchart[21] for alcohols, and by Hagen and Roberts for carboxylic acids[22], along with numerous others observed for other functional groups.[23,24] Clerc and Pretsch have devised general additive rules for 28 functional groups.[25] Dubois has used a topological parameter to model the alkyl environment.[26] Levy and Nelson[27] and Ejchart[28,29] have proposed substitution methods whereby the ^{13}C shifts are first estimated for the hydrocarbons, and heteroatoms are substituted later. Although these rules have varying accuracy, they serve as good starting points for spectral interpretation, especially when simple analogues cannot be located in the spectral libraries. A drawback to this approach is that it is labor-intensive and occasionally prone to arithmetic error.

One way to facilitate the application of substituent chemical shift rules is to computerize them. One such effort was made by Clerc and Sommerauer.[30] In this work we have modified and extended the Clerc–Pretsch rules and computerized them using a different approach. Our program (called CSHIFT) was written in a high-level language (FORTRAN IV) and has many special features. It is applicable to aliphatic carbons carrying 30 functional groups including the 28 listed by Clerc and Pretsch.[25] It can also take care of alicyclic compounds, although its accuracy tends to be lower.

METHOD

In the Grant–Paul scheme,[15] the ^{13}C shifts are thought of as arising from empirical additive parameters that are char-

† Hercules Research Center Contribution No. 1762.

0095-2338/83/1623-0197$01.50/0

acteristic of the neighboring atoms as in eq 1, where k_0 is a

$$\delta_{obsd} = k_0 + n_\alpha k_\alpha + n_\beta k_\beta + n_\gamma k_\gamma + n_\delta k_\delta + S \quad (1)$$

base value, n_i is the number of carbons in the ith position, k_i is an empirical parameter, and S is a steric correction term. Clerc and Pretsch[25] generalized these rules by providing substituent chemical shift values for 28 common organic functional groups. The steric correction parameters were also generalized. We have revised their values for the halogens since it appears that the α effect depends strongly on whether the halide is primary, secondary, or tertiary. From the haloalkanes reported in the literature the following α-substituent values are obtained:

	Primary	Secondary	Tertiary
F	70.1	69.0	66.0
Cl	31.0	35.0	43.0
Br	18.9	27.9	36.9
I	−7.2	3.8	20.8

It is of interest that the α effect stays fairly constant from primary to secondary to tertiary for fluorine but increases for the rest of the halogens. The rules are strictly valid only for alkanes with monosubstituted halogens. When multiple halogens are added to the alkane structure in close proximity, the incremental chemical shifts cease to be additive, and the rules must take on more complex forms.[31]

Two additional functional groups are added to the list. The first is alkanesulfonic acids (RSO_3H). The additive parameters have been given by Freeman and Angeletakis.[32] The second one is the peroxy (and hydroperoxy) functionalites (ROOR′). The ^{13}C shifts have been found[33] for a long-chain secondary hydroperoxide to occur at 86.2, 33.2, and 25.9 ppm, giving 56.3, 3.3, and −4.0 as the α, β, and γ parameters. More recent studies[34,35] provided shifts for additional compounds; the numbers used for this work are 55.0, 2.7, and −4.0.

For amides, alkyl substituents can occur on both sides of the functional group (i.e., RCONHR′). Clerc and Pretsch[25] originally only supplied the rules for alkyl carbons on the C=O side (i.e., R group). A new set of values is devised here for the carbons on the N side (i.e., R′). For this group, $\alpha = 28.0$, $\beta = 6.8$, $\gamma = -5.1$, and $\delta = 0$. In addition, the values for thiocyanate and thiol are revised. Minor adjustments are also applied to several other functional groups. The complete list of substituent parameters is given in Table I, and the steric corrections in Table II.

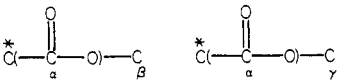
In the use of steric correction term (S) in eq 1, Clerc and Pretsch[25] recommended that one only counts the number of nonhydrogen substituents on the *most branched* α substituent. However, in Grant and Paul's original formulation,[15] the steric correction should be used on *all* α substituents. In our extensive computations, we noticed that Grant and Paul's approach to steric correction produces somewhat better results. We have therefore followed Grant and Paul's approach in this work.

The operation of the rules is best shown in an example. The base value k_0 is taken to be −2.3 ppm, and the ^{13}C shifts for the carbons in choline chloride will be computed:

$$\underset{1}{(CH_3)_3}—\overset{+}{N}—\underset{2}{CH_2}—\underset{3}{CH_2}OH\ Cl^-$$

C_1		C_2		C_3	
base	−2.3	base	−2.3	base	−2.3
α-N^+	30.7	α-N^+	30.7	α-C	9.1
3β-C	27.3	α-C	9.1	α-OH	49.0
α-C	−2.5	3β-C	27.3	β-N^+	5.4
δ-OH	0	β-OH	10.1	3γ-C	−7.5
$S(p,4)$	−3.4	$S(s,4)$	−7.5	$S(s,2)$	0
calcd	50.8	calcd	67.4	calcd	53.7

In the application of these rules, caution is needed in dealing with some functionalities containing more than one atom. For example, in methyl acetate, there are two possible ways to count the neighboring substituents (see below). The same



possibility 1 possibility 2

problem exists for C=C, C≡C, and C(O)N. For the substituent values given in Table I, possibility 2 is actually the correct one to use. The second atom in the group (O in this case) counts as a β atom, although its contribution is not figured in the calculation. (This is because this contribution

Table I. Substituent Chemical Shift Parameters (in ppm) for ^{13}C NMR To Be Used with Eq 1 with $k_0 = -2.3$

substituent	code	α[b]	β	γ	δ
-C-[a]	C	9.1	9.4	−2.5	0.3
-O-[a]	O	49.0	10.1	−6.0	0.3
-N<[a]	N	28.3	11.3	−5.1	0.3
-N-[a]	N+	30.7	5.4	−7.2	−1.4
S[a]	S	11.0	12.0	−3.0	−0.5
C_6H_5	PH	22.1	9.3	−2.6	0.3
F	F	70.1 (1,2) 69.0 (3) 66.0 (4)	7.8	−6.8	0.0
Cl	CL	31.1 (1,2) 35.0 (3) 43.0 (4)	10.0	−5.1	−0.5
Br	BR	18.9 (1,2) 27.9 (3) 36.9 (4)	11.0	−3.8	−0.7
I	I	−7.2 (1,2) 3.8 (3) 20.8 (4)	10.9	−1.5	−0.9
NH_3^+	NH3+	26.0	7.5	−4.6	−0.1
CN	CN	3.1	2.4	−3.3	−0.5
NO_2	NO2	62.0	4.4	−4.0	0
-OO-	OO	55.0	2.7	−4.0	0
C=NOH syn	CNOS	11.7	0.6	−1.8	0.0
C=NOH anti	CNOA	16.1	4.3	−1.5	0.0
SCN	SCN	21.0	7.2	−4.0	0.3
S(O)-	SO	31.1	9.0	−3.5	0.0
SO_3H	SO3H	38.9	0.5	−3.7	0.2
CHO	CHO	29.9	−0.6	−2.7	0.0
C(O)-	CO	22.5	3.0	−3.0	0.0
COOH	COOH	20.1	2.0	−2.8	0.0
COO^-	COO-	24.5	3.5	−2.5	0.0
COCl	COCL	33.1	2.3	−3.6	0.0
C(O)O	COO	22.6	2.0	−2.8	0.0
OC(O)	OCO	54.5 (1,2,3) 62.5 (4)	6.5	−6.0	0.0
CONH	CON	22.0	2.6	−3.2	−0.4
NHCO	NCO	28.0	6.8	−5.1	0.0
C=C[a]	C=C	21.5	6.9	−2.1	0.4
C≡C	C3C	4.4	5.6	−3.4	−0.6

[a] Steric correction parameters (Table II) apply to these substituents. [b] Number(s) in parentheses denotes the number of non-H substituents on the carbon in question.

Table II. Steric Correction Parameters[a]

i	$j=1$	$j=2$	$j=3$	$j=4$
primary	0.0	0.0	−1.1	−3.4
secondary	0.0	0.0	−2.5	−7.5
tertiary	0.0	−3.7	−9.5	−15
quaternary	−1.5	−8.4	−15	

[a] Designation = $S(i, j)$, where i = the carbon in question, and j = number of nonhydrogen substituents directly attached to the α substituent (applicable only to α-substituents marked with footnote *a* in Table I).

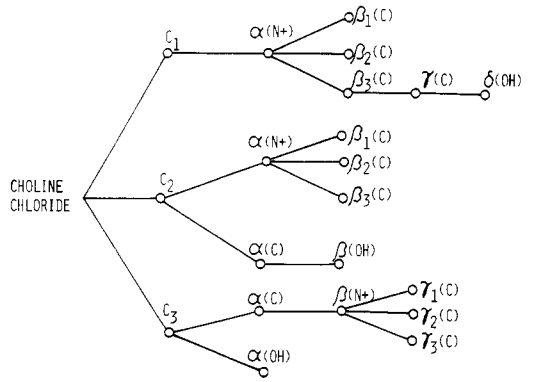


Figure 1. Logic used in the program CSHIFT for choline chloride. The calculation proceeds as follows: C_1–α(N+)–β_1(C)–β_2(C)–β_3(C)–γ-(C)–δ(OH)–C_2–α(N+) etc.

has already been incorporated in the COO group contribution.)

PROGRAM CSHIFT

Structuring of the Program. In writing a FORTRAN computer program for the modified Clerc–Pretsch rules, we sought to incorporate all the necessary calculations in the program so that the user needs only to feed in the molecular structure to obtain the chemical shifts with no further work. After some preliminary investigation, we decided to use an algorithm with a *tree*like logic. An example is provided by the choline chloride example we used earlier (Figure 1).

Thus, we start with the given molecular structure (the trunk of the tree) and examine each carbon in turn (the branches). For each carbon, whenever neighboring atoms are found, the subroutines corresponding to α, β, γ, and δ atoms are called, and appropriate values are added to the chemical shift.

A simplified block diagram of the algorithm is given in Figure 2. The *organization* of the program consists of the following routines.

(1) The main program reads in the molecular structure and initiates the computation one carbon at a time.

(2) Subroutine ALPHA deciphers how many and what kind of α atoms are present and adds the α-substituent values and the steric correction terms to the calculated chemical shifts. This routine also contains the "dictionary" with all structural symbols and parameter values.

(3) Subroutine BETA deciphers how many and what kind of β atoms are present.

(4) Subroutine GAMMA deciphers how many and what kind of γ atoms are present.

(5) Subroutine DELTA deciphers how many and what kind of δ atoms are present.

(6) Subroutine SWITCH contains features to handle some special cases (vide infra).

Two versions of program CSHIFT are available. The first was designed for batch operation. The structure needed is coded on cards, and the computed chemical shifts are obtained on the computer printout. The computer used for testing is a Perkin-Elmer Model 7/32. The second version was written for a Nicolet 1280 interactive computer with a raster CRT accessory. The structure can be entered directly on the keyboard with provisions for corrections or amendments. The computed chemical shifts are either displayed on the raster screen or printed on the keyboard. Needless to say, the second version is more convenient to use and is recommended for routine applications.

Operational Details. One advantage of the Nicolet version of CSHIFT is the conversational nature of the program and the simple input procedures. Basically one enters the molecular structure into the terminal exactly as one draws on paper. The only restriction is that the format for each atom or bond is

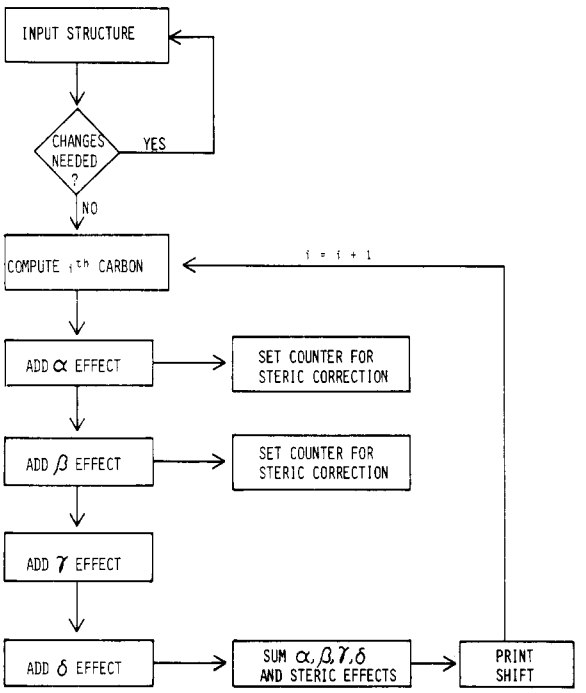


Figure 2. Schematic diagram for the program CSHIFT.

A4, such that an entry of C (for carbon) or an asterisk (for single bond) must be followed by three spaces, e.g.

```
C * C * O        NH3+ * C * COOH
    *                    *
    C                    C
```

2-propanol alanine

For simplicity, all hydrogens are omitted, and the structures are made up with the codes given in the second column of Table I.

Since we attempt to depict three-dimensional structures in two dimensions, crowding of atoms and bonds is sometimes unavoidable. This may lead to ambiguity in deciphering the intended structure. To obviate this difficulty, one can use any number of asterisks as a single bond; for example

```
C * C * C * * C * COOH
    *   *
    C   *
        *
        C
```

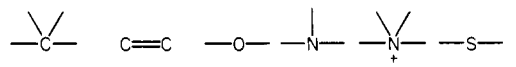
The only qualification is that for each bond the asterisks must be either all horizontal or all vertical. An additional advantage of this feature is that *cyclic* structures with odd number of carbons can be easily depicted, e.g.

```
C * C * C          C * C * C
*     *            *     *
C * C * C          C * * * C
```

cyclohexane cyclopentane

Alicyclic carbons usually give less accuracy than linear molecules in using additive shift rules. Nevertheless, the computer program calculates them just like linear molecules. Of course the rules do not work well for ring sizes less than or equal to 4.

As noted in Table II, the *steric correction term* applies only to the structures shown below. The subroutine ALPHA can



specifically trap for these groups. The correction term is not used for the other functional groups. Furthermore, provision has been made in the case of COO, CON, C=C, and C≡C

```
C PROGRAM CSHIFT (NICOLET 1280 VERSION)
C WRITTEN BY H. N. CHENG AND S. ELLINGSEN
C            HERCULES RESEARCH CENTER, WILMINGTON, DELAWARE 19899
      CHARACTER*4 BOND,CARB,IEND,IADD,TITLE(18),M(18,40),IM
      CHARACTER IL,IX,IN,IE,IR,IC,IS
      COMMON M,B,G,B,LA,LB,ICHEK
      DATA IL/'L'/,IN/'N'/,IE/'E'/,IR/'R'/,IC/'C'/,IS/'S'/
      BOND='*   '
      CARB='C   '
      IEND='$   '
      IADD='+   '
    1 DO 10 I=1,18
      DO 10 J=1,40
   10 M(I,K)='    '
      ICHEK=0
C ENTER THE STRUCTURE
C     TYPE GROUP CODE DIRECTLY UNDER EACH NUMBER (FORMAT =I4)
C     TYPE '$   ' AT POSITION 2 TO TERMINATE STRUCTURE INPUT
C     OR '+   ' TO TERMINATE AND TO OBTAIN SUBSTITUENT VALUES.
C     TO USE MULTIPLE * FOR SINGLE BOND, ** MUST BE ALL HORIZ OR VERT.
  100 WRITE(2,101)
  101 FORMAT(' ENTER TITLE (72 CHARACTERS MAX)'/)
  105 READ(1,110) (TITLE(I),I=1,18)
  110 FORMAT(18A4)
  120 WRITE(2,121)
  121 FORMAT(' ENTER STRUCTURE'/)
      WRITE(2,130)(I,I=2,8),(I,I=10,17)
  130 FORMAT('     ',7(I1,3X),'9',2X,8(I2,2X),/)
  140 DO 190 K=2,39
      WRITE(2,135) K
  135 FORMAT(' ',I2,2X)
      READ(1,141)(M(I,K),I=2,17)
  141 FORMAT(17A4)
      IM=M(2,K)
      IF(IM.EQ.IEND)GO TO 180
      IF(IM.EQ.IADD)GO TO 170
  190 CONTINUE
  170 ICHEK=1
  180 KM=K-1
      M(2,K)='
  215 FORMAT(A1)
C STRUCTURE MODIFICATION (OPTIONAL)
  300 WRITE(2,310)
  310 FORMAT('0ENTER OPTION:    ECHO STRUCTURE(E),',
     *T41,'RE-ENTER STRUCTURE(R),')
      WRITE(2,311)
  311 FORMAT('  CHANGE LINE(L), CHANGE SINGLE ENTRY(S),',
     * T41,'OR INITIATE CALCULATION(C)...')
      READ(1,215) IX
      IF(IX.EQ.IE) GO TO 320
      IF(IX.EQ.IR) GO TO 1
      IF(IX.EQ.IL) GO TO 350
      IF(IX.EQ.IS) GO TO 380
      IF(IX.EQ.IC) GO TO 400
      GO TO 300
  320 WRITE(2,321)(I,I=2,17)
  321 FORMAT('0 ',17(2X,I2)/)
      DO 322 K=2,KM
  322 WRITE(2,325)K,(M(I,K),I=2,17)
  325 FORMAT('0',I2,2X,17A4)
      GO TO 300
  350 WRITE(2,351)
  351 FORMAT(' CHANGE LINE # ? ')
  355 READ(1,356)KCHG
  356 FORMAT(I2)
      IF(KCHG-KM) 360,360,358
  358 KM=KCHG
  360 WRITE(2,361)
  361 FORMAT(' ENTER NEW LINE')
      WRITE(2,130) (I,I=2,8),(I,I=10,17)
      WRITE(2,135) KCHG
      READ(1,141)(M(I,KCHG),I=2,17)
      GO TO 300
  380 WRITE(2,381)
  381 FORMAT(' CHANGE ENTRY',T16,'VERTICAL POSITION # ?  ')
  382 READ(1,356)ICH
  383 WRITE(2,384)
  384 FORMAT(T15,' HORIZONTAL POSITION # ? ')
  385 READ(1,356) KCH
  390 WRITE(2,391)
  391 FORMAT(' PROVIDE NEW ENTRY (FORMAT A4): ')
  392 READ(1,393) M(ICH,KCH)
  393 FORMAT(A4)
      IF(KCH-KM) 300,300,395
  395 KM=KCH
      GO TO 300
C INITIATE CALCULATION
  400 WRITE(2,401)(TITLE(I),I=1,18)
  401 FORMAT('0',18A4,/)
      DO 450 K=2,KM
      DO 450 I=2,17
      IM=M(I,K)
      IF (IM.EQ.CARB) CALL ALPHA(I,K)
  450 CONTINUE
  460 WRITE(2,490)
  490 FORMAT('0ENTER OPTION:  EXIT(E),  ENTER NEW STRUCTURE(N),',/,
     *'                OR MAKE CHANGES ON OLD STRUCTURE(C)...')
      READ(1,215) IX
      IF(IX.EQ.IE) GO TO 999
      IF(IX.EQ.IN) GO TO 1
      IF(IX.EQ.IC) GO TO 300
      GO TO 400
  999 STOP
      END
      SUBROUTINE ALPHA(I,J)
      CHARACTER*4 SUB(31),BOND,M(18,40),IM,IS
      COMMON M,B,G,B,LA,LB,ICHEK
      DIMENSION ADD(31,4),KSTORE(4),LSTORE(4)
      REAL LCT(4,4),LT
      BOND='*   '
      SUB(1)='C   '
      SUB(2)='O   '
      SUB(3)='N   '
      SUB(4)='N+  '
      SUB(5)='S   '
      SUB(6)='*   '
      SUB(7)='PH  '
      SUB(8)='F   '
      SUB(9)='CL  '
      SUB(10)='BR  '
      SUB(11)='I   '
      SUB(12)='NH3+'
      SUB(13)='CN  '
      SUB(14)='NO2 '
      SUB(15)='OO  '
      SUB(16)='CHOS'
      SUB(17)='CHOA'
      SUB(18)='SCN '
      SUB(19)='SO  '
      SUB(20)='SO3H'
      SUB(21)='CHO '
      SUB(22)='CO  '
      SUB(23)='COOH'
      SUB(24)='COO-'
      SUB(25)='COCL'
      SUB(26)='COO '
      SUB(27)='OCO '
      SUB(28)='CON '
      SUB(29)='NCO '
      SUB(30)='C=C '
      SUB(31)='CJC '
      DATA ADD/9.1,49.0,28.3,30.7,11.6,0.0,22.1,78.1,31.1,18.9,
     * -7.2,26.0,3.1,62.0,55.0,11.7,16.1,21.0,31.1,38.9,29.9,
     * 22.5,20.1,24.5,33.1,22.6,54.5,22.0,28.0,21.5,4.4,
     * 9.4,18.1,11.3,5.4,12.0,0.0,9.3,7.8,16.0,11.0,
     * 18.9,7.5,2.4,4.4,2.7,6.6,4.3,7.2,9.0,0.5,-0.6,
     * 3.0,2.0,3.5,2.3,2.0,6.5,2.6,6.0,6.9,5.6,
     * -2.5,-6.0,-5.1,-7.2,-3.0,0.0,-2.6,-6.0,-5.1,-3.0,
     * -1.5,-4.6,-3.3,-4.6,-4.0,-1.0,-1.5,-4.0,-3.5,-3.7,-2.7,
     * -3.0,-2.0,-2.5,-3.6,-2.0,-6.0,-3.2,-5.1,-2.1,-3.4,
     * 0.3,0.3,0.3,-1.4,0.5,0.0,0.3,0.0,-0.5,-0.7,
     * -0.9,-0.1,-0.5,0.0,0.0,0.0,0.0,0.3,0.0,0.2,0.0,
     * 0.0,0.0,0.0,0.0,0.0,0.0,0.0,-0.4,0.0,0.0,0.4,-0.6/
      DATA LCT/0.0,0.0,0.0,-1.5,0.0,0.0,-3.7,-8.4,-1.1,-2.5,-9.5,-15.0,
     * -3.4,-7.5,-15.0,0.0/
C SINGLE BOND (*) IS CALLED SUB(6)
C NAMES LXXX ARE ALL USED TO DETERMINE CORRECTION TERMS
      LA=0
      LT=0.
      A=0.
      B=0.
      G=0.
      D=0.
C START CHECKING THE NEIGHBORING POSITIONS
      DO 100 NI=1,2
      DO 100 NJ=1,2
      LB=0
      I2=NI+NJ+I-3
      J2=NI-NJ+J
      IM=M(I2,J2)
      IF(IM.EQ.BOND)GO TO 175
      GO TO 100
  175 I3=2*I2-I
      J3=2*J2-J
   19 DO 110 KS=1,31
      IS=SUB(KS)
      IM=M(I3,J3)
      IF(IM.EQ.IS)GO TO 150
  110 CONTINUE
      WRITE(2,125) I3,J3,M(I3,J3)
  125 FORMAT('0ILLEGAL INPUT AT POSITION (',I2,',',I2,') ',A4)
      RETURN
  150 IF(KS-6) 180,170,180
  170 IX4=2*I3-I2
      JX4=2*J3-J2
      I2=I3
      J2=J3
      I3=IX4
      J3=JX4
      GO TO 19
  180 IF (KS-26) 420,410,400
  400 IF (KS-29) 410,410,420
  410 CALL SWITCH(KS,I2,I3,J2,J3)
  420 LA=LA+1
      A=A+ADD(KS,1)
      KSTORE(LA)=KS
      IF (KS-26) 200,210,210
  210 CALL GAMMA(I2,J2,I3,J3,SUB,ADD)
      GO TO 100
  200 CALL BETA(I2,J2,I3,J3,SUB,ADD)
      IF (KS-30) 500,520,500
  500 IF(KS-5)520,520,510
  510 LB=-1
  520 LSTORE(LA)=LB+1
  100  CONTINUE
C THE CHEMICAL SHIFT IS CALCULATED AND PRINTED OUT
      DO 300 K=1,LA
      LB=LSTORE(K)
      IF (LB) 320,320,321
  321 LT=LT+LCT(LA,LB)
  320 KS=KSTORE(K)
      GO TO (300,300,343,344),LA
  343 IF(KS.EQ.6) A=A-0.9
      IF(KS.EQ.9) A=A+4.0
      IF(KS.EQ.10)A=A+9.0
      IF(KS.EQ.11)A=A+11.0
      GO TO 300
```

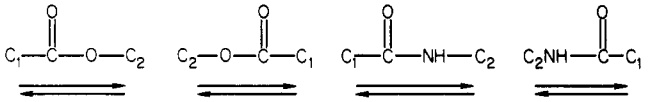
```
  344  IF(KS.EQ.27)A=A+8.0
       IF(KS.EQ.8) A=A-3.9
       IF(KS.EQ.9) A=A+12.0
       IF(KS.EQ.10)A=A+18.1
       IF(KS.EQ.11)A=A+28.0
  300  CONTINUE
       SHIFT=-2.3+A+B+G+D+LT
       WRITE(2,250)I,J,SHIFT
  250 FORMAT(' POSITION ',I2,',',I2,T20,F10.3,' PPM')
       IF(ICHEK)240,240,241
  241  WRITE(2,245) A,B,G,D,LT
  245  FORMAT('0  (A=',F6.2,'  B=',F6.2,'  G=',F6.2,'  D=',F6.2,
     *  '  S=',F6.2,' )',/)
  240  RETURN
       END



       SUBROUTINE SWITCH(KS,II,III,JJ,JJJ)
       IF((III+JJJ)-(II+JJ))200,200,100
   200 IF(KS-2*(KS/2))21,20,21
C IF KS=26 OR 28, CHANGE TO 27 OR 29, AND VICE VERSA.
    20 KS=KS+1
       RETURN
    21 KS=KS-1
   100 RETURN
       END



       SUBROUTINE BETA(I2,J2,I3,J3,SUB,ADD)
       CHARACTER*4 BOND,SUB(31),M(18,40),IK,IS
       COMMON M,B,G,D,LA,LB,ICHEK
       DIMENSION ADD(31,4)
       BOND='*   '
       DO 100 NI=1,2
       DO 100 NJ=1,2
       I4=NI+NJ+I3-3
       J4=NI-NJ+J3
   300 IF(I4-I2)320,310,320
   310 IF(I4-I3)100,320,100
   320 IF(J4-J2)350,330,350
   330 IF(J4-J3)100,350,100
   350 IK=M(I4,J4)
       IF(IK.EQ.BOND)GO TO 110
       GO TO 100
   110 I5=2*I4-I3
       J5=2*J4-J3
    82 DO 78 KS=1,31
       IS=SUB(KS)
       IK=M(I5,J5)
       IF(IK.EQ.IS)GO TO 79
    78 CONTINUE
    79 IF(KS-6)80,81,80
    81 IX6=2*I5-I4
       JX6=2*J5-J4
       I4=I5
       J4=J5
       I5=IX6
       J5=JX6
       GO TO 82
    80 IF(KS-26)420,410,400
   400 IF(KS-29)410,410,420
   410 CALL SWITCH (KS,I4,I5,J4,J5)
   420 LB=LB+1
       B=B+ADD(KS,2)
       IF(KS-26)200,210,210
   210 CALL DELTA(I4,J4,I5,J5,SUB,ADD)
       GO TO 100
   200 CALL GAMMA(I4,J4,I5,J5,SUB,ADD)
   100 CONTINUE
       RETURN
       END
```

```
       SUBROUTINE GAMMA(I4,J4,I5,J5,SUB,ADD)
       CHARACTER*4 SUB(31),BOND,M(18,40),IK,IS
       COMMON M,B,G,D,LA,LB,ICHEK
       DIMENSION ADD(31,4)
       BOND='*   '
       DO 100 NI=1,2
       DO 100 NJ=1,2
       I6=NI+NJ+I5-3
       J6=NI-NJ+J5
   300 IF(I6-I4)320,310,320
   310 IF(I6-I5)100,320,100
   320 IF(J6-J4)350,330,350
   330 IF(J6-J5)100,350,100
   350 IK=M(I6,J6)
       IF(IK.EQ.BOND)GO TO 110
       GO TO 100
   110 J7=2*J6-J5
       I7=2*I6-I5
    82 DO 78 KS=1,31
       IS=SUB(KS)
       IK=M(I7,J7)
       IF(IK.EQ.IS)GO TO 79
    78 CONTINUE
    79 IF(KS-6)80,81,80
    81 IX8=2*I7-I6
       JX8=2*J7-J6
       I6=I7
       J6=J7
       I7=IX8
       J7=JX8
       GO TO 82
    80 IF(KS-26)420,410,400
   400 IF(KS-28)410,410,420
   410 CALL SWITCH(KS,I6,I7,J6,J7)
   420 G=G+ADD(KS,3)
       IF(KS-26)200,100,100
   200  CALL DELTA(I6,J6,I7,J7,SUB,ADD)
   100  CONTINUE
        RETURN
        END



        SUBROUTINE DELTA(I4,J6,I7,J7,SUB,ADD)
        CHARACTER*4 SUB(31),BOND,M(18,40),IK,IS
        COMMON M,B,G,D,LA,LB,ICHEK
        DIMENSION ADD(31,4)
        BOND='*   '
        DO 100 NI=1,2
        DO 100 NJ=1,2
       I8=NI+NJ+I7-3
        J8=NI-NJ+J7
   300 IF(I8-I6)320,310,320
   310 IF(I8-I7)100,320,100
   320 IF(J8-J6)350,330,350
   330 IF(J8-J7)100,350,100
   350 IK=M(I8,J8)
       IF(IK.EQ.BOND)GO TO 110
       GO TO 100
   110 I9=2*I8-I7
       J9=2*J8-J7
    82 DO 78 KS=1,31
       IS=SUB(KS)
       IK=M(I9,J9)
       IF(IK.EQ.IS)GO TO 79
    78 CONTINUE
    79 IF(KS-6)80,81,80
    81 IX10=2*I9-I8
       JX10=2*J9-J8
       I8=I9
       J8=J9
       I9=IX10
       J9=JX10
       GO TO 82
    80 IF(KS-26)420,410,400
   400 IF(KS-29)410,410,420
   410 CALL SWITCH(KS,I8,I9,J8,J9)
   420 D=D+ADD(KS,4)
   100  CONTINUE
        RETURN
        END
```

Figure 3. Program listing for the Nicolet version of CSHIFT.

to count the second atom in the group in the computation. (This is actually done by leapfrogging to the second subroutine in the hierarchy, e.g., BETA calling DELTA, skipping GAMMA.)

A problem in the computer approach is the asymmetric groups, e.g., esters and amides.



If one inputs the ester structure as COO, the calculation is fine for C_1. However, for C_2, a different set of rules would apply because we must calculate from right to left and the substituent rules for OCO must be used. This problem is taken care of in the subroutine SWITCH, where the direction of the computation is specifically taken into account.

Program Listing. The program listing for the Nicolet version is given in Figure 3. Readers interested in the batch version may write to the authors for the program listing and input instructions.

RESULTS AND DISCUSSION

The computer printout for choline chloride is given in Figure 4. An intentional error is introduced in the input structure, and provisions for corrections are illustrated. The program as presently written can handle molecular structures with 16 × 40 structural elements. Since structural elements include atoms and bonds, a total of 8 × 20 atoms can be handled. For molecules larger than this, they can be broken up into smaller pieces and calculated piecewise. Aternatively, the dimension statements in the program may be changed.

A comparison of calculated vs. observed ^{13}C shifts for a variety of compounds is provided in Table III. Four classes of compounds are noted: acyclic hydrocarbons (class I), hydrocarbons with one substituent (class II), hydrocarbons with two substituents (class III), and cyclic compounds (class IV). The observed shift values are mostly taken from Johnson and Jankowski.[36]

For the vast majority of organic compounds, the modified Clerc and Pretsch rules as given in the CSHIFT program work

Table III. Comparison of Calculated vs. Observed (in Parentheses) ^{13}C Shifts

Class I. Acyclic Hydrocarbons

compd	shift, ppm 1	2	3	4	5	6
$CH_3CH_2CH_2CH_2CH_3$	14.0 (13.5)	22.8 (22.2)	34.7 (34.1)			
$CH_3CH(CH_3)CH_2CH_2CH_3$	22.3 (22.7)	28.2 (27.9)	41.6 (41.9)	20.3 (20.8)	14.3 (14.3)	
$CH_3CH(CH_3)CH(CH_3)CH_2CH_3$	19.8 (20.0)	31.8 (31.9)	40.0 (40.6)	27.2 (26.8)	11.8 (11.6)	17.0 (14.5)

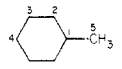
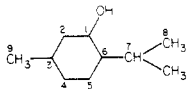
Class II. Hydrocarbons with 1 Substituent

compd	shift, ppm 1	2	3	4	5	6	7	8
$CH_3C(O)CH_2CH_2CH_3$	27.4 (29.25)	48.1 (45.10)	16.4 (17.60)	13.5 (13.4)				
$CH_3CH_2CH_2CH_2Cl$	44.8 (44.6)	35.3 (35.2)	20.2 (20.4)	13.2 (13.4)				
$CH_3CH_2CH_2CH_2CH(CH_2CH_3)CH_2OH$	67.4 (65.1)	40.6 (42.1)	30.6 (30.3)	30.3 (29.3)	23.4 (23.2)	14.0 (14.1)	24.0 (23.5)	11.8 (11.1)
$CH_3CH{=}CHCH_2CH_2CH_2CH_2CH_3$ (trans)	17.0 (17.8)	33.0 (32.7)	30.0 (29.9)	32.6 (31.7)	23.2 (22.8)	14.0 (14.1)		

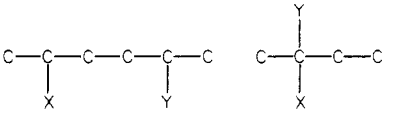
Class III. Hydrocarbons with 2 Substituents

compd	shift, ppm 1	2	3	4	5
$SHCH_2CH_2COOH$	38.9 (38.2)	19.8 (19.3)			
$CH_3CH(OH)CH_2OH$	72.8 (68.2)	71.3 (67.7)	19.2 (18.7)		
$CH_3NHCH_2CH_2OH$	64.6 (60.3)	54.6 (54.3)	33.2 (36.0)		
$CH_3CH_2OCH_2CH_2C{\equiv}N$	17.8 (18.9)	65.1 (65.1)	62.2 (66.5)	14.7 (14.9)	
$NH_2CH_2CH_2CH_2CH_2CH_2NH_2$	42.3 (42.1)	34.4 (33.8)	24.5 (24.0)		

Class IV. Cyclic Compounds

compd	shift, ppm 1	2	3	4	5	6	7	8	9
	70.2 (68.6)	31.0 (27.2)	23.3 (24.2)						
	77.1 (74.7)	29.2 (34.0)	34.4 (34.7)	68.0 (67.6)	16.6 (17.9)				
	32.0 (33.0)	37.2 (35.6)	27.8 (26.6)	30.9 (33.0)	20.1 (22.9)				
	68.0 (71.3)	42.9 (45.1)	26.6 (31.7)	35.9 (34.6)	23.7 (23.2)	46.2 (50.1)	23.9 (25.8)	20.4 (21.0)	20.4 (22.2)

reasonably well. From our experience the rules work best for hydrocarbons and acyclic monofunctional compounds. Cyclic structures tend to give somewhat less accuracy. Multiply substituted compounds (e.g., structures below) also give slightly



less accuracy; in general, the closer the two groups (X and Y), the larger the discrepancy. The reason is that the substituent shifts in these cases are probably not additive. This discrepancy is especially severe for halogens.

Although we have aimed to make the program general, there are obviously cases where the rules would not apply. The following is a list of these exclusions: (A) Molecules with functional groups not in the list in Table I cannot be computed. (B) Molecules with fixed conformations need corrections. Clerc and Pretsch[25] noted that, for these molecules, the γ effect should be additionally incremented as follows: cis, −4.0; trans, +2.5. This is not incorporated in the program because one has to designate in the molecular structure which features are rigid, and this would unduly complicate data input. The prospective user may add these values himself to the computer-generated shifts. (C) For olefins and amides, at present only doubly substituted functional groups can be computed.

possible	not possible
$RCH{=}CH_2$	$R_1CH{=}CR_2R_3$
$RCH{=}CHR'$	$R_1R_2C{=}CR_3R_4$
$RCONH_2$	$RCONR_1R_2$
$RCONHR'$	

Furthermore, the rules only give the ^{13}C shifts of the trans isomers. The cis isomer is usually 2–6 ppm (amides) and 5–10 ppm (olefins) upfield.

```
*RUN CSHIFT

ENTER TITLE (72 CHARACTERS MAX)
CHOLINE CHLORIDE

ENTER STRUCTURE

    2  3  4  5  6  7  8  9  10 11 12 13 14 15 16 17

 2          C

 3          *

 4  C   *   N+  *   C   *   C   *   0

 5          *

 6          C

 7  #


ENTER OPTION:    ECHO STRUCTURE(E),      RE-ENTER STRUCTURE(R),
 CHANGE LINE(L), CHANGE SINGLE ENTRY(S), OR INITIATE CALCULATION(C)...S

CHANGE ENTRY    VERTICAL POSITION # ?  10

                HORIZONTAL POSITION # ? 4

PROVIDE NEW ENTRY (FORMAT A4): O


ENTER OPTION:    ECHO STRUCTURE(E),      RE-ENTER STRUCTURE(R),
 CHANGE LINE(L), CHANGE SINGLE ENTRY(S), OR INITIATE CALCULATION(C)...C


CHOLINE CHLORIDE
POSITION  4, 2          50.699 PPM
POSITION  2, 4          50.699 PPM
POSITION  6, 4          68.299 PPM
POSITION  8, 4          53.700 PPM
POSITION  4, 6          50.699 PPM

ENTER OPTION:  EXIT(E),  ENTER NEW STRUCTURE(N),
               OR MAKE CHANGES ON OLD STRUCTURE(C)...E

STOP 0
```

Figure 4. Computer output for the choline chloride example.

ACKNOWLEDGMENT

We thank W. J. Freeman for encouragement and comments, O. W. Marks and H. G. Lee for assistance in using the Perkin-Elmer (Interdata) 7/32 computer, and N. P. Carlson for providing feedback on the use of the program.

REFERENCES AND NOTES

(1) Sadtler Research Laboratories, Inc., Philadelphia, PA 19104.
(2) Bremser, W.; et al. "Carbon-13 NMR Spectral Data"; Verlag Chemie: Weinheim/Bergstr., Germany, 1981.
(3) Breitmaier, E.; Haas, G.; Voelter, W. "Atlas of Carbon-13 NMR data"; Heyden: London, 1979.
(4) Stothers, J. B. "Carbon-13 NMR Spectroscopy"; Academic Press: New York, 1972.
(5) For general reviews and monographs, see: (a) Gribov, L. A.; Elyashberg, M. E. *CRC Crit. Rev. Anal. Chem.* **1979**, *8*, 110. (b) Hippe, Z.; Hippe, R. *Appl. Spectrosc. Rev.* **1980**, *16*, 135. (c) Smith, D. H., Ed. "Computer Structure Elucidation". *ACS Symp. Ser.* **1977**, *No. 54*. (d) Varmuza, K. "Pattern Recognition in Chemistry"; Springer Verlag: West Berlin, 1980.
(6) Also known as the NIH-EPA-NIC data base. See for example: (a) Jezl, B. A.; Dalrymple, D. L. *Anal. Chem.* **1975**, *47*, 203. (b) Dalrymple, D. L.; Wilkins, C. L.; Milne, G. W. A.; Heller, S. R. *Org. Magn. Reson.* **1978**, *11*, 535. (c) Zupan, J.; Heller, S. R.; Milne, G. W. A.; Miller, J. A. *Anal. Chim. Acta* **1978**, *103*, 141. (d) Milne, G. W. A., Zupan J.; Heller, S. R.; Miller, J. A. *Org. Magn. Reson.* **1979**, *12*, 289.
(7) More recent papers from the Stanford group include: (a) Lindley, M. R.; Gray, N. A. B.; Smith, D. H.; Djerassi, C. *J. Org. Chem.* **1982**, *47*, 1027. (b) Crandall, C. W.; Gray, N. A. B.; Smith, D. H. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 48. (c) Gray, N. A. B.; et al. *J. Org. Chem.* **1981**, *46*, 703. (d) Gray, N. A. B.; et al. *Org. Magn. Reson.* **1981**, *15*, 375. (e) Nourse, J. G.; Smith, D. H.; Djerassi, C. *J. Am. Chem. Soc.* **1980**, *102*, 6289.
(8) For references on the CHEMICS system, see: (a) Sasaki, S.; et al. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 211. (b) Abe, H.; Sasaki, S.; Tanaka, K.; Osada, H. *CODATA Bull.* **1981**, *40*, 31.
(9) Bremser, W.; Wagner, H.; Franke, B. *Org. Magn. Reson.* **1981**, *15*, 178 and references therein.
(10) Schwarzenbach, R.; Meili, J.; Konitzer, H.; Clerc, J. T. *Org. Magn. Reson.* **1976**, *8*, 11.
(11) Kwiatkowski, J.; Riepe, W. *Anal. Chim. Acta* **1982**, *135*, 285, 293, 381.
(12) Szalontai, G.; Simon, Z.; Czapo, Z.; Farkas, M.; Pfeifer, G. *Anal. Chim. Acta* **1981**, *133*, 31.
(13) Shelley, C. A.; Munk, M. E. *Anal. Chem.* **1982**, *54*, 516.
(14) Mlynarik, V.; Vida, M.; Kello, V. *Anal. Chim. Acta* **1980**, *122*, 47.
(15) Paul, E. G.; Grant, D. M. *J. Am. Chem. Soc.* **1964**, *86*, 2984.
(16) Lindeman, L. P.; Adams, J. Q. *Anal. Chem.* **1971**, *43*, 1245.
(17) Carman, C. J.; Tarpley, A. R.; Goldstein, J. H. *Macromolecules* **1973**, *6*, 719.
(18) Eggert, H.; Djerassi, C. *J.* Am. *Chem. Soc.* **1973**, *95*, 3710.
(19) Sarneski, J. E.; Suprenant, H. L.; Molen, F. K.; Reilley, C. N. *Anal. Chem.* **1975**, *47*, 2116.
(20) Roberts, J. D.; Weigert, F. J.; Kroschwitz, J. I.; Reich, H. J. *J. Am. Chem. Soc.* **1970**, *92*, 1338.
(21) Ejchart, A. *Org. Magn. Reson.* **1977**, *9*, 351.
(22) Hagen, R.; Roberts, J. D. *J. Am. Chem. Soc.* **1969**, *91*, 4504.
(23) Wehrli, F. W.; Wirthlin, T. "Interpretation of Carbon-13 NMR Spectra"; Heyden: London, 1978.
(24) Some recent studies include hydrocarbons,[24a] amines,[24a,d] carboxylic acids and esters,[24a-e] amino acids,[24a,d] amides,[24f] and nitroalkanes.[24g] See: (a) MacDonald, J. C. *J. Magn. Reson.* **1979**, *34*, 207. (b) Marr, D. H. *Org. Magn. Reson.* **1980**, *13*, 28. (c) Couperus, P. A.; Clague, A. D. H.; van Dougen, J. P. C. M. *Ibid.* **1978**, *11*, 590. (d) D. L. Rabenstein and T. L. Sayer, *J. Magn. Reson.* **1976**, *24*, 27. (e) Dostovalova, V. I.; et al. *Org. Magn. Reson.* **1983**, *21*, 11. (f) Newmark, R. A.; Hill, J. R. *J. Magn. Reson.* **1976**, *21*, 1. (g) Ejchart, A. *Org. Magn. Reson.* **1977**, *10*, 263. A lot of interest was shown for *haloalkanes*. See for example: (h) Bailey, W. F.; Ciofii, E. A.; Wiberg, K. B. *J. Org. Chem.* **1981**, *46*, 4219. (i) Dostovalova, V. I.; et al. *Org. Magn. Reson.* **1981**, *16*, 251. (j) Nouguier, R.; Surzar, J. M.; Virgili, A. *Ibid.* **1981**, *15*, 155 and references cited in ref 24h-j.
(25) Clerc, J. T.; Pretsch, E.; Sternhell, S. "^{13}C-Kernresonanzspektroskopie"; Akademische Verlagsgesellschaft: Frankfurt am Main, 1973.
(26) Dubois, J.-E.; Carabedian, M. *Org. Magn. Reson.* **1980**, *14*, 265 and references therein.
(27) Levy, G. C.; Nelson, G. L. "Carbon-13 NMR for Organic Chemists"; Wiley-Interscience: New York, 1972.
(28) Ejchart, A. *Org. Magn. Reson.* **1980**, *13*, 368.
(29) Ejchart, A. *Org. Magn. Reson.* **1981**, *15*, 22.
(30) Clerc, J. T.; Sommerauer, H. *Anal. Chim. Acta* **1977**, *95*, 33. The program was written in machine language for the Bruker BNC minicomputer.
(31) For example, see ref 24i.
(32) Freeman, F.; Angeletakis, C. N. *Org. Magn. Reson.* **1983**, *21*, 86.
(33) Cheng, H. N.; Schilling, F. C.; Bovey, F. A. *Macromolecules* **1976**, *9*, 363.
(34) Budinger, P. A.; et al. *Anal. Chem.* **1981**, *53*, 884.
(35) Aganov, A. V.; Antonovskii, V. L. *Izv. Akad. Nauk SSSR, Ser. Khim.* **1982**, 271.
(36) Johnson, L. F.; Jankowski, W. C. "Carbon-13 NMR Spectra"; R. E. Krieger Publishing Co.: Huntington, NY 1978.