

# Prediction of $^{13}\text{C}$ Nuclear Magnetic Resonance Chemical Shifts<sup>1a</sup>

Dennis H. Smith\*<sup>1b</sup> and Peter C. Jurs<sup>1c</sup>

Contribution from the Departments of Chemistry and Genetics, Stanford University, Stanford, California 94305, and the Department of Chemistry, The Pennsylvania State University, State College, Pennsylvania 16801.† Received September 29, 1977

**Abstract:** We describe a computer-based method for prediction of  $^{13}\text{C}$  NMR chemical shifts. The method involves computation of molecular descriptors from topological and three-dimensional representations of molecules, the latter afforded by a molecular mechanics program. Subsequently, the method of stepwise, multiple linear regression calculates coefficients relating the descriptors to observed chemical shifts. We predict the  $^{13}\text{C}$  chemical shifts of 287 primary, secondary, and tertiary carbon centers in our data set of methyl-substituted cycloalkanes with a standard error of about 1.1 ppm.

Computer-based methods for analysis of analytical data are receiving widespread interest as the cost of computer hardware decreases and sophisticated computer programs become more widely available. One area of special interest is computer-assisted structure elucidation.<sup>2</sup> The two major uses for analytical data in structure elucidation are *interpretation* and *prediction*. *Interpretation* of original data in terms of structural features forms the basis for proposing structural candidates. Subsequently, each of the candidates can be tested in detail against the same or other data by *prediction* of the physical or chemical behavior of each candidate. These predictions can be used to rank the candidates. Significant work has been done in automated interpretation of mass spectral, infrared, and  $^1\text{H}$  NMR data,<sup>2</sup> and, to a lesser extent,  $^{13}\text{C}$  NMR data.<sup>3-7</sup> More recently, reports have appeared discussing the use of computers to predict spectroscopic and chemical behavior of sets of molecules.<sup>2,8,9</sup>

In this report we discuss an approach to prediction of  $^{13}\text{C}$  NMR chemical shifts. Recent studies correlating chemical shifts with structural features of molecules have resulted in highly class specific rules<sup>10-19</sup> with limited capability for application outside the structural context in which they were formed. It would be useful to devise methods for prediction of  $^{13}\text{C}$  NMR chemical shifts which are less strongly coupled to the concept of chemical class, thus allowing prediction of shifts for a variety of types of compounds. We propose an approach which is, in principle, capable of such generality. We provide an example of an application where  $^{13}\text{C}$  NMR shifts are accurately predicted for related, but different, structural types.

Our approach relates a variety of structural features of a set of molecules with observed chemical shifts. Structural features are derived automatically by a computer program from *both* topological and three-dimensional ("geometrical") representations of the molecules. The calculated topological and geometrical features, or descriptors, are regressed (multiple linear regression) against observed shifts. In this respect our approach is closely related to the work of Grant and co-workers,<sup>13-16</sup> who used regression analysis based on numbers of various features which could be derived by manual examination of molecular models. The key difference in our approach is that we attempt to obtain more widely applicable rules relating chemical shifts to structural features by extracting certain features from computed, three-dimensional models of each compound. From the set of atomic coordinates obtained by molecular modeling the relationships and interactions among atoms can be easily determined and a variety of structural descriptors easily computed. We obtain, for a set of

saturated cycloalkanes, a set of regression coefficients with high correlation. These coefficients can be used to predict  $^{13}\text{C}$  NMR spectra of candidate structures for an unknown.

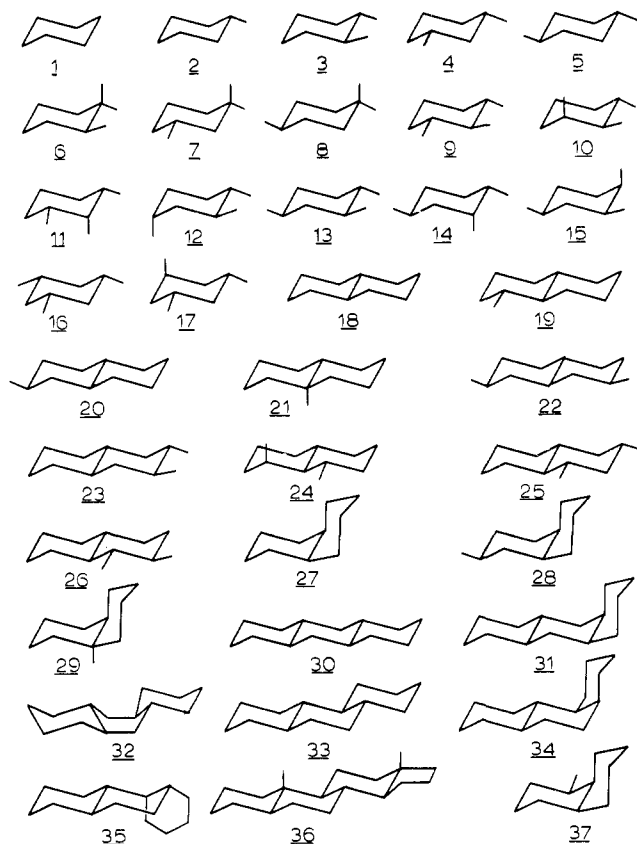
## The Data Set

Initial correlations of structure with  $^{13}\text{C}$  NMR shift related topological features of structures to spectral peaks.<sup>10,11,20</sup> Such correlations proved adequate for spectral predictions and interpretations for alkanes and aliphatic amines.<sup>4,7,18,20</sup> However, topological features alone proved inadequate for treatment of cyclic systems. For example, the topologically equivalent tertiary carbons of *cis*- and *trans*-decalin exhibit resonances differing by 7.3 ppm. Consideration of descriptors based on three-dimensional properties is required to "explain" observed spectra.<sup>13-17,19,21</sup> Selection of descriptors has been based on manual examination of models and chemical intuition. Counts of the numbers of each descriptor are used with multiplicative constants to modify predictions based on parent compounds.<sup>13-17,19,21</sup>

We decided to attempt to parametrize geometrical influences on  $^{13}\text{C}$  NMR chemical shifts more systematically. To do so, we required a set of molecules (with assigned spectra) with well-characterized conformations. Assigned spectra for a number of cyclic, saturated hydrocarbons (described in more detail below) have been presented.<sup>13-16</sup> Although related, the molecules have been treated as separate "classes", with parameter sets generated separately for each class.<sup>13-16</sup> Thus, this set of compounds meets our criteria of being sufficiently different to merit treatment, while allowing isolation of effects of topology and geometry from other influences of heteroatoms and multiple bonds.

The set of carbon centers used in this work is shown in Table I. The first subset, **1-17**, consists of cyclohexane plus 16 methyl-substituted cyclohexanes.<sup>13-14</sup> Four other methyl-substituted cyclohexanes reported in those papers were not used because the molecules are too conformationally mobile. The second subset, **18-29**, consists of *trans*- and *cis*-decalin plus ten methyl-substituted decalins.<sup>15</sup> *cis-syn*-1-Methyl-decalin (**37**) is placed at the bottom of Table I because it required special attention. It was not used in the derivations of any of the parameters from multiple linear regression. Shifts for **37** were predicted using the parameters derived from all the other compounds listed in Table I. Six perhydroanthracenes and perhydrophenanthrenes, **30-35**, were also used.<sup>16</sup> A single four-ring compound, 5- $\alpha$ -androsterane (**36**), was available.<sup>17</sup> Thus, the entire data set consists of 36 compounds with a total of 51 primary, 157 secondary, and 79 tertiary carbon centers. There were an insufficient number of quaternary centers to allow them to be included in this study. The reported shifts (vs.  $\text{Me}_4\text{Si}$ ) for these compounds range from a low of 5.3 ppm to a high of 34.0 ppm for the primary centers, from 20.5 to 49.7

† Work performed while on sabbatical leave at Stanford University, 1976-1977.



ppm for the secondary centers, and from 26.3 to 55.1 ppm for the tertiary centers. Following earlier studies, primary, secondary, and tertiary centers are treated separately.<sup>10</sup>

### The Descriptor Set

The objective in the development of the descriptor set was to obtain a set of descriptors for each carbon center that would allow a multiple linear regression fit to predict shifts accurately. These descriptors should be calculable from a knowledge of the topology and geometry of the molecule. Since we expected to use a method of molecular mechanics<sup>22,23</sup> to derive a model for each compound, and since the development of such models can only be done by computer, we decided from the start to allow all the descriptors to be computer generated. The topological descriptors are developed from the connection table of the molecule. They can be calculated directly from a knowledge of the atom and bond identities and connectivities given in the connection table. The geometrical descriptors are based on the proximity of atoms of various types to the carbon center in question. These descriptors are calculated from the atomic coordinates of each atom subsequent to model building. The complete set of descriptors generated for each carbon center is shown in Table 11, and described in detail below.

**Topological Descriptors.** The topological descriptors (Table 11) generated for each carbon center included a variety of features which we assumed might be important based on previous relationships of topology to chemical shift.<sup>4,10,11</sup> Note that there is no term for the number of nearest neighbors ( $\alpha$ -substituents) because the carbon centers are already partitioned by the degree of the carbon atom. The connectivity index (CI) (originally developed by Randić<sup>24</sup>) descriptors were chosen because this descriptor has been shown to be correlated with a large number of molecular properties.<sup>25</sup> It has usually been calculated for entire molecular structures as a summation of contributions over all the bonds in the structure. The term in the summation associated with each bond is related to the connectivity of the atoms bound together by it. Equation 1 shows how the connectivity index (CI) is calculated for an

entire structure:

$$CI = \sum_k C(k) = \sum_k \frac{1}{[L(i)L(j)]^{1/2}} \quad (1)$$

where  $k$  is an index over all the bonds in the molecule, and  $L(i)$  and  $L(j)$  are the connectivities of the two atoms which are joined by the  $k$ th bond. Because we wish to characterize individual carbon centers rather than entire molecules, we determined the CI descriptors for each center by summing over all the second nearest-neighbor bonds (C12), all the third nearest-neighbor bonds (C13), etc. The CI descriptors are measures of the degree of substitution of atoms on  $n$ th nearest-neighbor bonds. Thus, the descriptors are highly correlated to the number of  $\beta$ ,  $\gamma$  and more distant substituents.

**Geometrical Descriptors.** From the beginning of this study, it was recognized that there must be geometrical information available to the prediction routine, since  $^{13}\text{C}$  NMR shifts are sensitive to changes in orientation and conformation. As mentioned previously, we minimized effects of conformational changes by including in the data set only compounds whose conformations were locked by substituent patterns. The geometries obtained from molecular mechanics routines are known to be quite close to those obtained from x-ray data and have been used for a wide variety of research in organic chemistry.<sup>22,23,26,27</sup> In our study we used the molecular mechanics modeling program of Allinger et al.<sup>22</sup> to obtain three-dimensional models (see Experimental Section). The resulting atomic coordinates for all carbon and hydrogen atoms were provided to another routine that calculated the geometrical descriptors.

There seems to be little agreement in the literature regarding details of relationships among structural features and chemical shifts. For example, several theories have been advanced to rationalize " $\beta$  effects"<sup>15,16,28</sup> and " $\gamma$  effects"<sup>16,29,30</sup> including considerations of both steric and electronic effects. We did not attempt to parametrize electronic features of molecules in our data set. Indeed, recent work found no simple relationship between computed charge densities and  $^{13}\text{C}$  NMR shielding constants for butanes and butenes.<sup>29</sup> Structural descriptors which are measures of interatomic distances ought, however, to capture steric effects.

We focussed on three types of geometrical descriptors. The first two are different measures of the proximity of other atoms in the molecule. The first is the count of the number of atoms of each type within spherical shells whose radii are given in Table 11. These radii were determined by noting that, in plots of interatomic distances to other atoms for each of the carbon centers in a given structure, natural clusters form near the center of each spherical shell. This is due, of course, to the regular geometries (the six-membered rings and their conformations) among members of our data set. The radial descriptors are simply, for each carbon center, summations of one over the radial distance (to the  $n$ th power) to all other carbons or hydrogens in the molecule.

The third descriptor is a measure of the van der Waals strain energy associated with pairwise interactions between hydrogen atoms on a given carbon center with other hydrogens in the vicinity. The reasons for selecting this parameter, which we have used only for tertiary carbon atoms, are discussed below (see Results and Discussion).

We are aware that many other geometrical descriptors might have been selected. We are also aware that our particular choices are not necessarily extendable to other structural classes and that reasonable, but accidental, correlations can be obtained by our method. Nevertheless, our descriptors reflect, to a first approximation, the environment of a given center in terms of various measures of distance to neighboring atoms. The important point is that the availability of a three-

Table 1. Compounds Comprising the Data Set

Compd	Number of		
	Primary centers	Secondary centers	Tertiary centers
1 Cyclohexane		1	
2 Methylcyclohexane	1	3	1
3 <i>trans</i> -1,2-Dimethylcyclohexane	1	2	1
4 <i>cis</i> -1,3-Dimethylcyclohexane	1	3	1
5 <i>trans</i> -1,4-Dimethylcyclohexane	1	1	1
6 1,1,2-Trimethylcyclohexane	3	4	1
7 1,1,3-Trimethylcyclohexane	3	4	1
8 1,1,4-Trimethylcyclohexane	3	2	1
9 1- <i>trans</i> -2- <i>cis</i> -3-Trimethylcyclohexane	2	2	2
10 1- <i>trans</i> -2- <i>trans</i> -3-Trimethylcyclohexane	3	3	3
11 1- <i>cis</i> -2- <i>cis</i> -3-Trimethylcyclohexane	2	2	2
12 1- <i>trans</i> -2- <i>cis</i> -4-Trimethylcyclohexane	3	3	3
13 1- <i>trans</i> -2- <i>trans</i> -4-Trimethylcyclohexane	3	3	3
14 1- <i>cis</i> -2- <i>trans</i> -4-Trimethylcyclohexane	3	3	3
15 1- <i>cis</i> -2- <i>cis</i> -4-Trimethylcyclohexane	3	3	3
16 1- <i>cis</i> -3- <i>cis</i> -5-Trimethylcyclohexane	1	1	1
17 1- <i>cis</i> -3- <i>trans</i> -5-Trimethylcyclohexane	3	2	2
18 <i>trans</i> -Decalin		2	1
19 <i>trans-anti</i> -1-Methyldecalin	1	7	3
20 <i>trans-syn</i> -2-Methyldecalin	1	7	3
21 <i>trans</i> -9-Methyldecalin	1	4	1
22 <i>trans-syn</i> -2- <i>syn</i> -7-Dimethyldecalin	1	3	3
23 <i>trans-syn</i> -2- <i>syn</i> -3-Dimethyldecalin	1	3	2
24 <i>trans-anti</i> -1- <i>syn</i> -8-Dimethyldecalin	2	6	4
25 <i>trans-anti</i> -1- <i>syn</i> -3-Dimethyldecalin	2	6	4
26 <i>trans-anti</i> -1- <i>syn</i> -2-Dimethyldecalin	2	6	4
27 <i>cis</i> -Decalin		4	1
28 <i>cis-syn</i> -2-Methyldecalin	1	7	3
29 <i>cis</i> -9-Methyldecalin	1	8	1
30 <i>trans-syn-trans</i> -Perhydroanthracene		3	1
31 <i>cis-trans</i> -Perhydroanthracene		10	4
32 <i>trans-anti-trans</i> -Perhydroanthracene		3	1
33 <i>trans-anti-trans</i> -Perhydrophenanthrene		5	2
34 <i>trans-anti-cis</i> -Perhydrophenanthrene		10	4
35 <i>trans-syn-cis</i> -Perhydrophenanthrene		10	4
36 5- $\alpha$ -Androstane	2	11	4
37 <i>cis-syn</i> -1-Methyldecalin	1	7	3

dimensional model which can be manipulated in a computer provides an easy route to consideration of many different descriptors. Our set is potentially general and, with the exception of the van der Waals parameter, is not overly sensitive to the details of the molecular model.

### Experimental Section

Model building and calculation of geometrical descriptors were performed on the SUMEX-AIM Digital Equipment Corp. PDP-10 at Stanford University. A DEC-10 version of the molecular mechanics program of Allinger et al., "MM1",<sup>22</sup> was used to construct models based on starting coordinates obtained in one of two ways. An x-ray structure of androstane (**36**) provided coordinates used as a starting point for (1) androstane itself; and, with appropriate structural modifications, (2) the substituted cyclohexanes, based on ring A of androstane; and (3) the substituted *trans*-decalins based on rings A and B. Coordinates for the carbon atoms of the *cis*-decalins and the perhydroanthracenes and phenanthrenes were approximated using the program of Wipke et al.<sup>31</sup> These coordinates were then passed to MM1 for final minimization, now also including all attached hydrogen atoms. Computation of the geometrical descriptors (Table II) was straightforward based on these coordinates. Radial distances are measured from atom center to atom center.

Determination of topological descriptors and the regression analyses were performed on the Stanford Center for Information Processing (SCIP) IBM 370/168. The topological descriptors were obtained using an algorithm which, given the connection table of the structure, determined the numbers of *n*th nearest neighbors. Connectivity indexes were calculated using eq 1. The stepwise multiple linear re-

gression programs used were BMDO2R and the multiple linear regression program included in the SPSS package.

### Results and Discussion

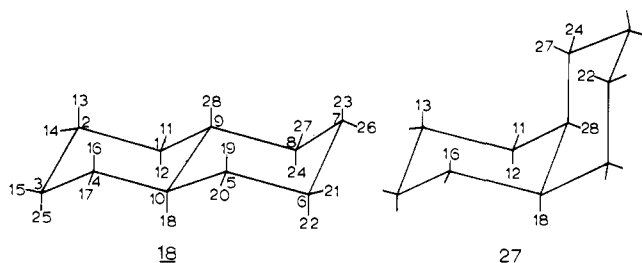
Our initial experiments treated discrete subsets of the data, for example, the cyclohexanes, the *trans*-decalins, and so forth, to ensure that we could achieve results comparable to those of Grant et al.<sup>13-16</sup> with our set of descriptors (Table 11). These experiments were successful until we included tertiary centers which were at *cis* ring junctures, at which point the fit of predicted and observed shifts was substantially degraded. The reason for this quickly became obvious. The sets of topological and geometrical descriptors (Table 11) are completely equivalent for some tertiary centers, while the observed chemical shifts are quite different. For example, ring juncture atoms, C-9,10, of *cis*- (**27**) and *trans*-decalin (**18**) have identical descriptor sets, while their chemical shifts differ by 7.3 ppm (shifts are 44.2 and 36.9 ppm, relative to Me<sub>4</sub>Si, for **18** and **27**, respectively). Although the geometrical descriptors appear to describe the different spatial arrangements of primary and secondary centers in compounds with *cis* vs. *trans* ring junctures, the descriptors are inadequate for the tertiary centers.

As a result, we made a careful comparison of the models calculated for pairs of compounds differing by *cis* vs. *trans* ring junctures. The final steric energy output from MM1 is, as expected, higher for compounds possessing *cis* ring junctures. For example, the final steric energy for *cis*-decalin (**27**) is 15.44

**Table II.** Descriptor Set Generated for Each Carbon Center

Topological Descriptors	
CI2	Connectivity index for all second nearest-neighbor (NN) bonds
CI3	Connectivity index for all third NN bonds
CISUM	CI1 + CI2 + CI3
L2	Number of second NN bonds
L3	Number of third NN bonds
(CI1)	Average connectivity index for first NN bonds
(CI2)	Average connectivity index for second NN bonds
(CI3)	Average connectivity index for third NN bonds
M1	Number of primary carbons bonded to the carbon of interest
M2	Number of secondary carbons bonded to the carbon of interest
M3	Number of tertiary carbons bonded to the carbon of interest
Geometrical Descriptors	
NCS3	Number of carbon atoms found between 2.7 and 3.4 Å
NCS4	Number of carbon atoms found between 3.4 and 4.0 Å
NCS5	Number of carbon atoms found between 4.0 and 5.0 Å
NHS1	Number of hydrogen atoms found between 0.0 and 1.5 Å
NHS2	Number of hydrogen atoms found between 1.5 and 2.4 Å
NHS3	Number of hydrogen atoms found between 2.4 and 3.2 Å
NHS4	Number of hydrogen atoms found between 3.2 and 3.6 Å
NHS5	Number of hydrogen atoms found between 3.6 and 5.0 Å
RCS1	Sum of $1/d(\text{C})$ over all carbon atoms in the molecule
RCS2	Sum of $1/d(\text{C})^2$ over all carbon atoms in the molecule
RCS3	Sum of $1/d(\text{C})^3$ over all carbon atoms in the molecule
RHS1	Sum of $1/d(\text{H})$ over all hydrogen atoms in the molecule
RHS2	Sum of $1/d(\text{H})^2$ over all hydrogen atoms in the molecule
RHS3	Sum of $1/d(\text{H})^3$ over all hydrogen atoms in the molecule
VDW	van der Waals strain energy associated with the single attached hydrogen of a tertiary center

kcal, for *trans*-decalin (**18**), 12.94 kcal. This additional strain is reflected largely in the terms for bending and 1,4 (or higher) van der Waals interactions, both of which are higher for *cis*-



decalin (**27**). Examination of individual energy terms revealed no obvious correlation of chemical shift differences with reported bond length and angle deformations, nor with bending and stretch-bending terms in these relatively strain-free systems. In addition, there were no obvious correlations with shifts of primary or secondary carbon centers with 1,4 van der Waals interactions of the centers themselves or their attached hydrogen atoms. However, the 1,4 (or higher) van der Waals terms are consistently higher for hydrogen atoms bonded to the tertiary centers which are *cis* ring junctures. For example, considering interactions which are greater than 0.1 kcal, H-18 and H-28 of *cis*-decalin (**27**) each contribute about 2.5 kcal

**Table III.** Results for Secondary Centers for Selected Subsets of the Data Set

	Cyclohexanes	Decalins	Trans ring junctures <sup>a</sup>
<i>n</i>	42	63	60
<i>r</i> <sup>2</sup>	0.997	0.992	0.992
Std error	0.47	0.61	0.61
No. of residuals > 1 ppm	0	4	4

<sup>a</sup> All *trans*-decalins, the secondary centers from the all *trans*-perhydroanthracenes and phenanthrenes, plus androstane.

**Table IV.** Results for Tertiary Centers for Selected Subsets of the Data Set

	Cyclohexanes	Trans ring junctures <sup>a</sup>
<i>n</i>	29	39
<i>r</i> <sup>2</sup>	0.995	0.975
Std error	0.41	1.16
No. of residuals > 1 ppm	0	8

<sup>a</sup> All *trans* compounds' ring junctures and points of methyl substitution.

**Table V.** Summary of Results of Multiple Linear Regression for All Carbon Centers in the Complete Data Set

	Primary centers	Secondary centers	Tertiary centers
<i>n</i>	51	157	79
<i>r</i> <sup>2</sup>	0.961	0.973	0.965
<i>F</i>	100	524	217
Std error	1.11	1.09	1.18
No. of residuals > 2.0 ppm	3	6	5
> 1.5 ppm	3	19	12
> 1.0 ppm	11	43	22
> 0.5 ppm	23	89	42
Largest residual	3.69	-6.34	3.32

to the steric strain, while the contribution is only about 0.9 kcal each for H-18 and H-28 of *trans*-decalin (**18**). The additional interactions are, for example, between H-28 and H-11, -12, -24 and -18, interactions which do not exist for *trans*-decalin (**18**). Interestingly, these additional interactions are greater in magnitude than the  $\gamma$  H-H interactions between H-27 and H-13 and H-16, and H-22 and H-16 of **27**, interactions usually presumed responsible for greater strain in the *cis* system.

The above pattern is repeated for other compounds bearing *cis* ring junctures and for tertiary centers formed by substitution of an axial vs. equatorial methyl group. The correlation even by manual observation is obvious. In all cases the tertiary carbon center whose attached hydrogen is involved in greater steric interactions displays a shift upfield with respect to  $\text{Me}_4\text{Si}$ . We have no explanation for the actual cause of this shift, although it is not clearly attributable to bond length or angle deformations. We note that this correlation would also be captured by extending the geometrical descriptors to include, for each carbon center, the radial distances between its attached hydrogen atoms and other hydrogen atoms in the structure. We also note that H-H interactions were used to parametrize shifts for the cyclohexanes and decalins,<sup>14,15</sup> although the particular interactions discussed above were not used in those studies.

Representative results using the complete descriptor set are

**Table VI.** Coefficients Obtained in Fits of All Centers<sup>a</sup>

Variable	Primary centers coefficients	Secondary centers coefficients	Tertiary centers coefficients
C12		10.757 (1)	33.280 (1)
L2	25.018 (2)		
M1			-5.778 (5)
M2			-4.881 (8)
(C12)	-66.541 (9)		
(C13)	-13.991 (10)		
NCS3	1.170 (1)	-4.137 (3)	-1.294 (3)
NCS4		0.752 (4)	1.857 (6)
NCS5	1.115 (8)		
NHS2		-8.140 (7)	-0.225 (4)
NHS3	-0.262 (7)	-2.222 (10)	-5.205 (9)
NHS4	-1.366 (5)	-1.316 (2)	-1.750 (10)
NHS5		-0.484 (8)	
RCS1	6.059 (3)		
RCS3	-261.74 (6)		15.346 (7)
RHS1		-1.854 (9)	
RHS2		-10.003 (6)	
RHS3	24.067 (4)	124.55 (5)	
VDW			-3.562 (2)
Constant	46.624	-154.94	16.564

<sup>a</sup> Based on chemical shifts relative to Me<sub>4</sub>Si. Numbers in parentheses indicate the order of choice during regression.

presented in Tables III and IV for subsets of secondary and tertiary centers, respectively. These results are certainly comparable to those obtained by Grant and co-workers, but the more important question is the ability of the descriptor set to predict shifts across the entire data set.

Table V shows the results obtained for primary, secondary, and tertiary centers for the entire data set. The standard errors are each approximately 1.1 ppm. This standard error is very reasonable considering the diversity of compound types and carbon centers. It indicates that the descriptor set captures important structure/chemical shift relationships even though direct, causal relationships cannot be inferred from these data alone. The standard errors are low enough to lend a high degree of confidence in the predictive power of the parameter set. Also shown in Table V are several measures of the quality of the fits in addition to the standard error. The counts of numbers of residuals larger than a certain threshold are reported cumulatively to show that there are relatively few predictions in large error.

Table VI shows which descriptors were chosen by the stepwise inclusion, multiple linear regression program used for each of the fits. The values of the coefficients are given, and the numbers in parentheses show the sequence in which the descriptors were chosen. No attempt was made to scale the magnitude and range of each descriptor to a common value, nor to include aliphatic carbon centers of known conformation (e.g., methane). Thus, the constant terms bear no relationship to any "basis" shift. Although completely suitable for computer analysis, manual application of the descriptors and the coefficients to predict a spectrum is difficult without a three-dimensional model.

It is dangerous to attribute a great deal of physical significance to particular descriptors and whether or not they were chosen. The descriptors are not completely independent. Selection of one descriptor during the stepwise procedure may account for the variance of another, related descriptor, which would be chosen in the absence of the former. However, certain descriptors in our set which were found to be important are interesting in light of past correlations. The C12 and L2 topological descriptors are indirect measures of the numbers of  $\beta$  and  $\gamma$  carbon atoms, structural features used in almost all previous correlations. The term NCS3 is a direct measure of

**Table VII.** Observed and Predicted Chemical Shifts Relative to Me<sub>4</sub>Si for Compounds, 22, 28, 33, and 36

Carbon	Obsd shift <sup>a</sup>	Pred shift	Residual	Std error
<b>22</b>				
1,8	43.05	43.24	-0.19	
2,7	32.89	32.90	-0.01	
3,6	35.60	34.76	0.84	
4,5	34.05	34.09	-0.04	
9	42.84	43.03	-0.19	
10	42.60	42.64	-0.04	
11,12	23.0	23.51	0.51	0.39
<b>28</b>				
1	35.63	36.44	-0.81	
2	34.22	34.48	-0.26	
3	30.33	30.79	-0.46	
4	33.22	33.13	0.09	
5	26.37	25.28	1.09	
6	27.63	27.61	0.02	
7	21.43	22.34	-0.91	
8	33.22	33.72	-0.50	
9	36.98	37.05	-0.07	
10	36.47	36.60	-0.13	
11	23.16	22.75	0.41	0.56
<b>33</b>				
1,8	34.59	34.51	0.08	
2,7	26.84	26.66	0.18	
3,6	27.14	26.26	0.88	
4,5	30.13	30.47	-0.34	
9,10	35.02	34.79	0.23	
11,14	43.61	42.82	0.79	
12,13	48.21	48.36	-0.15	0.48
<b>5-<math>\alpha</math>-Androstane (36)</b>				
1	38.8	37.15	1.65	
2	22.3	21.73	0.57	
3	26.9	26.50	0.40	
4	29.2	29.89	-0.69	
5	47.1	46.90	0.20	
6	29.2	27.19	2.01	
7	32.6	32.02	0.58	
8	36.0	35.75	0.25	
9	55.1	51.78	3.32	
10	36.4			
11	20.9	22.43	-1.53	
12	39.0	40.74	-1.74	
13	40.8			
14	54.7	55.30	-0.60	
15	25.5	26.08	-0.58	
16	20.5	21.88	-1.38	
17	40.5	40.60	-0.10	
18	17.6	17.25	0.35	
19	12.3	12.51	-0.21	1.27

<sup>a</sup> Observed shifts taken from ref 9 and 10.

the number of " $\gamma$  gauche"<sup>17</sup> carbon atoms in this data set, due to the regular geometries of the fused, six-membered rings systems. Similarly, NCS4 represents the number of other, more remote  $\gamma$  carbon atoms. Although selected frequently, the NHS terms are most important for the secondary carbon centers, NHS4 in particular. This term includes a variety of hydrogen atoms including hydrogens on  $\gamma$  carbons in spatial relationships which capture additional  $\gamma$  H-H interactions in, for example, *cis*- vs. *trans*-decalin. The RCS and RHS terms can be regarded as inverse measures of the "nearness" of atoms in the rest of the molecule. The reasons for the greater importance of these descriptors for the primary centers are un-

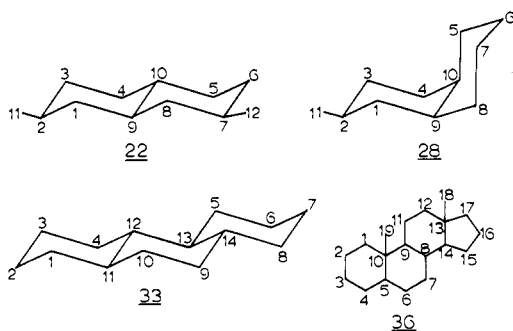
**Table VIII.** Prediction of Shifts for *cis-syn*-1-Methyldecalin (**37**)<sup>a</sup>

Obsd shifts	Pred shifts <sup>10</sup>	Carbon	Pred shifts <sup>b</sup>
85.82	87.03	9	85.80
90.12	90.12	10	91.68
91.64	91.61	1	94.58
95.27	95.14	5	95.00
99.28	98.11	2	97.84
101.40	102.04	3	100.93
101.40	102.04	7	100.90
103.05	102.06	4	103.33
106.93	106.60	6	106.32
108.78	106.62	8	107.96
109.04	109.54	11	108.72

<sup>a</sup> With respect to benzene as the internal standard. <sup>b</sup> This work.

clear. These terms seem less useful than the CI, NCS, and NHS terms.

Table VII shows the quality of fit obtained from the multiple linear regression program for three compounds (**22**, **28**, **33**) randomly chosen from the data set, and androstane (**36**). These



data are representative of the quality of the predictions. For many compounds the predictions allow unambiguous assignment of the spectrum. Many remaining ambiguities can be resolved by considering data from partially decoupled spectra where the degree of each carbon atom can be ascertained.

The coefficients shown in Table VI were used to predict the shifts of *cis-syn*-1-methyldecalin (**37**). Initial experiments including this compound showed dramatic degradation of fits, indicating potential misassignment of chemical shifts. Subsequently, it was found that the published structure of this compound was drawn incorrectly (although named correctly) and we had repeated the error.<sup>32</sup> Our predicted spectrum is shown in Table VIII together with observed and predicted shifts<sup>15</sup> for this compound.

## Conclusions

$^{13}\text{C}$  NMR shifts can be accurately predicted based on topological and geometrical descriptions of structural features obtained from three-dimensional models of known compounds. Both model building and derivation of a wide variety of de-

scriptors can be done utilizing computer techniques, opening the possibility for more systematic studies of correlations of molecular structure with  $^{13}\text{C}$  NMR chemical shifts. Although we have demonstrated the utility of this method for a variety of saturated, carbocyclic systems, further studies are required to test the generality of the descriptor set and to extend the set to systems possessing unsaturation or heteroatoms.

A limitation of the technique is the difficulty of prediction of spectra of new compounds. Although computations are possible to do manually, they are far more efficiently done by computer. Another limitation, one common to all current methods of  $^{13}\text{C}$  NMR structure/spectrum correlations, is the absence of strong theoretical models of sufficient predictive power on which to base derivation of descriptors. We hope that our empirical correlations and our observations on the influence of steric strain will assist further theoretical work.

## References and Notes

- (1) (a) One of us (D.H.S.) wishes to thank the National Institutes of Health (NIH RR 00612) for partial financial support of this work. We thank Raymond Carhart and Gretchen Schwenger for helpful discussions and suggestions. (b) Stanford University. (c) The Pennsylvania State University.
- (2) Dennis H. Smith, Ed., "Computer-Assisted Structure Elucidation", American Chemical Society, Washington, D.C., 1977.
- (3) W. Bremser, M. Klier, and E. Meyer, *Org. Magn. Reson.*, **7**, 97 (1975).
- (4) R. E. Carhart and C. Djerassi, *J. Chem. Soc., Perkin Trans. 2*, 1753 (1973).
- (5) B. A. Jezl and D. L. Dalrymple, *Anal. Chem.*, **47**, 203 (1975).
- (6) R. S. Schwarzenbach, J. Meili, H. Konitzer, and J. T. Cierc, *Org. Magn. Reson.*, **8**, 11 (1976).
- (7) A. L. Burlingame, R. V. McPherron, and D. M. Wilson, *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 3419 (1973).
- (8) C. A. Shelley and M. E. Munk, Abstracts, 173rd National Meeting of the American Chemical Society, New Orleans, La., March 20-25, 1977.
- (9) C. A. Shelley, H. B. Woodruff, C. R. Snelling, and M. E. Munk in ref 2, p 92.
- (10) L. P. Lindeman and J. Q. Adams, *Anal. Chem.*, **43**, 1245 (1971).
- (11) H. Eggert and C. Djerassi, *J. Am. Chem. Soc.*, **95**, 3710 (1973).
- (12) J. E. Sarneski, H. L. Surprenant, F. K. Molen, and C. N. Reilley, *Anal. Chem.*, **47**, 2116 (1975).
- (13) D. K. Dalling and D. M. Grant, *J. Am. Chem. Soc.*, **89**, 6612 (1967).
- (14) D. K. Dalling and D. M. Grant, *J. Am. Chem. Soc.*, **94**, 5318 (1972).
- (15) D. K. Dalling, D. M. Grant, and E. G. Paul, *J. Am. Chem. Soc.*, **95**, 3718 (1973).
- (16) D. K. Dalling and D. M. Grant, *J. Am. Chem. Soc.*, **96**, 1827 (1974).
- (17) H. Eggert, C. L. Van Antwerp, N. S. Bhacca, and C. Djerassi, *J. Org. Chem.*, **41**, 71 (1976).
- (18) T. M. Mitchell and G. M. Schwenger, *Anal. Chem.*, submitted for publication.
- (19) H. Eggert and C. Djerassi, *J. Org. Chem.*, **38**, 3788 (1973).
- (20) G. M. Schwenger and T. M. Mitchell in ref 2, p 58.
- (21) S. H. Grover and J. B. Stothers, *Can. J. Chem.*, **52**, 870 (1974).
- (22) N. L. Allinger, M. T. Tribble, M. A. Miller, and D. H. Wertz, *J. Am. Chem. Soc.*, **93**, 1637 (1971).
- (23) E. M. Engler, J. D. Andose, and P. v. R. Schleyer, *J. Am. Chem. Soc.*, **95**, 8005 (1973).
- (24) M. Randic, *J. Am. Chem. Soc.*, **97**, 6609 (1975).
- (25) L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York, N.Y., 1976.
- (26) C. Altona and D. H. Faber, *Top. Curr. Chem.*, **45**, 1 (1974).
- (27) S. A. Godleski, P. v. R. Schleyer, E. Osawa, Y. Inamoto, and Y. Fujikura, *J. Org. Chem.*, **41**, 2596 (1976).
- (28) H. Beierbeck and J. K. Saunders, *Can. J. Chem.*, **53**, 1307 (1975).
- (29) D. M. Grant and B. V. Cheney, *J. Am. Chem. Soc.*, **89**, 5315 (1967).
- (30) K. Seidman and G. E. Maciel, *J. Am. Chem. Soc.*, **99**, 659 (1977).
- (31) W. T. Wipke and P. Gund in "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. T. Feldman, and E. Hyde, Ed., Wiley-Interscience, New York, N.Y., 1974, p 147.
- (32) D. K. Dalling, personal communication.