Topological Approach of ^{13}C NMR Spectral Simulation: Application to Fuzzy Substructures

Annick Panaye, Jean-Pierre Doucet,* and Bo Tao Fan

Institut de Topologie et de Dynamique des Systèmes, associé au CNRS, URA-34, Université Paris 7, 75005 Paris, France

Received August 12, 1992

2D experiments (HETCOR and COSY) provide carboneous substructures bearing hanging bonds (free sites). As a part of the SERC system (Shift Evaluation for Resonating Carbons), for sp^3 carbons, a topological model for chemical shift simulation in fuzzy substructures is used to predict the nature of the missing α substituents.

INTRODUCTION

Spectral simulation, i.e., reconstruction of the spectra associated to a given structural formula initially intervened only as a filter step in structural elucidation systems. Thanks to a wider structural scope and more precise predictions, it now becomes an independent area, and even experienced spectroscopists can use spectral simulation as a help for structural analysis. The largely widespread use of ^{13}C NMR, during the last years, for conformity assessment of chemical samples (verifying a structural formula) leads to increased recourse to spectral simulation processes.

Furthermore, spectral simulation still remains an important part of structural elucidation systems. Indeed, these systems generally process through recognition of relevant structural fragments from their spectral characteristics and generation of candidates. The choice of the solution is then carried out by simulation of their spectra, comparison with the experimental spectrum, and determination of the best match.[1] In ^{13}C NMR-based systems, elementary substructures identified are generally constituted by concentric fragments centered on the resonating carbon, the shift of which is the only parameter examined for identification. Recognition of n-tuples,[2] taking simultaneously into account the shift of the central carbon and those of the neighboring carbons as in the EPIOS system, not only allows for an easier recognition of relevant spectral features but also guides and largely speeds up the structural expansion step. 2D NMR spectra now afford powerful tools to identify larger structural fragments by defining the carbon chain and to give valuable clues about their assembling in the molecular framework.[3-6]

In this paper, we stress another possible application of spectral simulation in the framework of structural elucidation. Beside the usual application for ranking candidates, spectral simulation operating on fuzzy substructures can be used upstream and provides a tool for defining the nature of the substructures constituting the "molecular puzzle". First, we will present the basis of our method of calculation of the ^{13}C shifts from a topological description of their environment. We will then develop the application of these models to the simulation of fuzzy substructures, a step we propose to include in structural elucidation.

Spectral simulation is interfaced to the conflicting requirements of a greatest precision in spectral information and a wider structural scope to be able to treat very varied situations. In the search for coupling molecular substructures and chemical shifts, two approaches are traditional: one directly exploits databases through substructures/subspectra associations,[2,7,8] the other one uses relationships quantifying the influence of the various atoms of the substructure.[9-13] Since the early times of ^{13}C NMR, various additive-increment models were successfully developed within structurally homogeneous chemical families.[14-16] However, two avenues still deserve special attention: a refined description of environment effects on chemical shifts[17,18] with incorporation of topographical features[19] and the treatment of polyfunctionalized compounds.[10b,11b]

Within the DARC system, the concept of FREL (Fragment Reduced to an Environment that is Limited) offers an efficient and flexible framework to generate, for multispectroscopy simulation approaches, structural descriptors centered on atoms (NMR) or bonds (IR) and provided with adaptable precision (variable depth and chromatic fuzziness).[20] When associating the relevant spectral information to the structural fragments constituting the input molecule, this highly flexible description can be used for either recognition of characteristic substructures from substructures/subspectra databases or application of correlation models.

Within the framework of the SERC system currently developed in our laboratory and following varied studies on structure–chemical shift relationships,[17,18,21] we propose here an exploratory module for ^{13}C NMR simulation relying on topological relationships.

PRINCIPLES OF THE SERC SIMULATION

As in most chemical shift calculations, we consider concentric fragments centered on the resonating carbon. The model we have chosen obeys the double constraints of accurate chemical shift prediction with understandable parameters and possible future extensions to families not yet considered. According to our previous studies, two topological approaches have been investigated. First, in an homogeneous family, the DARC-PELCO method identifies the contribution of every site of the ordered environment around the resonating carbon.[21] The environment is organized into concentric layers where sites are labeled with a linear order. Organizing the structural family into topological filiations allows for a statistical determination of the site contributions displayed in a valuated graph. Second, the influence of an alkyl environment on the ^{13}C shift has been rationalized in linear relationships ($\omega\Sigma\lambda$ model).[17,18] These methods will be briefly summarized below.

In the approach developed here for the chemical shift calculation, these two models are used. A structure is conventionally assimilated to an alkane skeleton perturbed with functional groups, this virtual alkane corresponding to the isotopologous framework. In a first step, the influence of the alkyl environment so generated is evaluated on the resonating carbon according to the PELCO model proposed for the alkane family. Then, the influence of the function is taken into account, using the $\omega\Sigma\lambda$ formalism. Replacing the function by an isotopologous carbon moiety rather than a

0095-2338/93/1633-0258$04.00/0

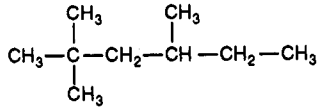
hydrogen atom to evaluate the participation of the framework aims at maintaining the connectivity and possible steric constraints and generally allows for a more accurate evaluation of the chemical shifts.[18a]

Alkane Framework. The influence of alkyl branching depends on the connectivity of the resonating carbon. For this reason, the topological analysis has been carried out on four separate subsets of the alkane population published by Lindeman and Adams,[14] according to the ^{13}C connectivity (shifts are expressed in ppm, with TMS as the reference).

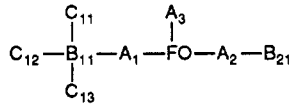
In the DARC-PELCO method, the environment of the resonating carbon is described in concentric layers A, B, C, and D provided with a linear order. The sites in the environment are labeled by indices which specify the order induced in this concentric organization. For example, on the second layer, sites B_{ij}, the first index refers to the position of the preceding layer (A_i) to which they are bonded. The second index j corresponds to the order for sites linked to a same seed site (B_{12} is the second B site linked to the first A site). Lower indices correspond to the more substituted positions. The environment of each ^{13}C is, therefore, considered as progressively built starting from the focus by successive addition of ordered sites: (increasing indices i then j on each layer). In this (formal) series of generation, the evolution of the property investigated (here the chemical shift) going from one environment to the following one is considered as a perturbation conventionally attributed to the new site introduced. These perturbation terms are evaluated by least-squares procedures and gathered in a valuated graph. For a given environment, the property sought is calculated by summation of the contributions associated to the focus and the occupied sites in the environment, and possibly interaction terms to cope with nonadditivity of perturbations when specific sites are simultaneously occupied. (For more details see, for instance, refs 20 and 21.)

The four topological correlations obtained are good (Figure 1), as indicated by the statistical tests gathered in Table I. Although these populations comprise some relatively hindered compounds, a simple summation of site contributions is sufficient to evaluate the chemical shift with an accuracy better than 2 ppm for the subsets of primary, secondary, and quaternary ^{13}C (this last case being not surprising since effects are largely dampened). Two interaction terms only are needed for the tertiary ^{13}C subset to achieve the same prediction reliability. This analysis leads to four different valuated graphs gathered in Figure 1. Although the sampling of the environments given by the population investigated is rather large, some sites are obviously lacking. We have estimated them by analogy with the existing sites to approximate new environments not included in the initial population.

For example, for the calculation of the CH chemical shift in 2,2,4-trimethylhexane



the graph is organized around the resonating carbon as



According to the valuated graph of the tertiary carbon

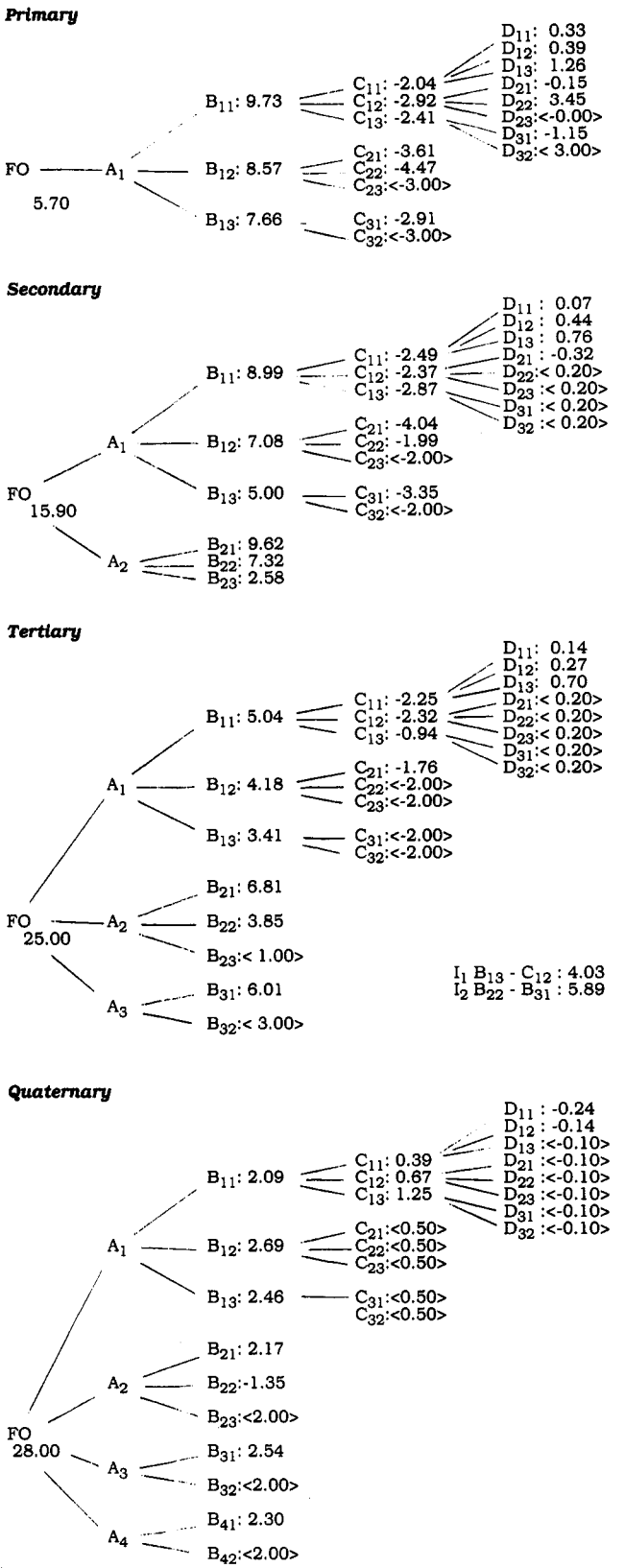


Figure 1. Alkane valuated graphs (site contributions in ppm Ref TMS). Within each layer of carbons around the resonating ^{13}C (FOcus), sites are ordered according to connectivity. Each substructure is described as a topological vector and least-squares treatment yields site values. Terms within brackets are estimated from analogous sites.

(Figure 1):

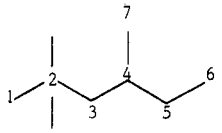
$$\underset{25.0}{(FO + A_1 + A_2 + A_3)} + \underset{5.0}{B_{11}} + \underset{6.8}{B_{21}} + \underset{-2.2}{C_{11}} + \underset{-2.3}{C_{12}} + \underset{-0.9}{C_{13}}$$

where $\delta_{cal} = 31.4$ and $\delta_{exp} = 31.9$ ppm. The whole set of the

Table I. Statistical Tests for DARC-PELCO Topological Correlations of Alkane ^{13}C Shifts[a]

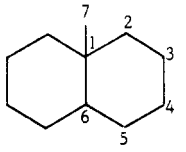
	n	n'	R	ψ	SD
primary	130	4	0.997	0.08	0.50
secondary	121	4	0.999	0.04	0.36
tertiary	54	4	0.997	0.08	0.50
quaternary	24	6	0.997	0.10	0.22

[a] n = total number of environments; n' = number of environments which exhibit a site present only once and are excluded of the correlation; R = correlation coefficient; SD = standard deviation; ψ = Exner's test.



n	δ_{exp}	δ_{cal}
1	29.9	30.4
2	31.0	31.0
3	51.0	51.4
4	31.9	31.4
5	31.0	30.8
6	11.2	10.8
7	21.9	20.4

Figure 2. ^{13}C SERC simulation of 2,2,4-trimethylhexane.



n	δ_{exp}		δ_{cal}
1	32.9	34.8	35.8
2	36.4	42.4	36.8
3	22.8	22.2	23.4
4	24.5	27.4	26.7
5	28.1	29.4	28.9
6	41.8	46.2	43.5
7	28.2	15.7	23.4

Figure 3. ^{13}C SERC simulation of 1-methyldecalines.

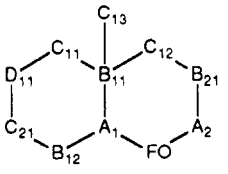
Table II. Corrective Cyclic Parameters

	size			
atom	3	4	5	6
CH_2	−20	−3.5	−4	−3.5
CH	−18	1	−0.5	−2.5
C	−16	5	4	−2.5
α-substituent	−6	0.5	3	2.5

chemical shifts of this compound is gathered in Figure 2.

For cyclic environments, an accurate evaluation of the chemical shifts could be only achieved by taking into account stereochemical features. Very often, calculations are carried out in using specific patterns. However, for a first evaluation, an acceptable approximation is attainable with the preceding topological models provided that corrective terms are introduced in the calculation (Table II). This rough model leads to an averaged accuracy of about 4 ppm for simple (mono- and bi-) cyclic systems.

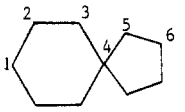
For example, for the C_5 of 1-methyldecaline system (Figure 3):



the analogous alkane is calculated as (Figure 1, graph of secondary carbon):

$$(FO + A_1 + A_2) + B_{11} + B_{12} + B_{21} + C_{11} + C_{12} + C_{13} + C_{21} + D_{11}$$
$$15.9 \quad 9.0 \quad 7.1 \quad 9.6 \quad -2.5 \quad -2.4 \quad -2.9 \quad -4.0 \quad 0.1$$

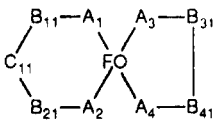
Two corrective terms are added: −3.5 for a CH_2 in a six-membered cycle and 2.5 since this CH_2 is α-substituent to a six-membered cycle; δ_{cal} = 28.9 and δ_{exp} = 28.1 (cis) or 29.4 (trans) ppm.



n	δ_{exp}	δ_{cal}
1	26.8	28.0
2	24.5	19.9
3	38.7	35.8
4	42.6	41.5
5	38.7	39.2
6	24.1	25.8

Figure 4. ^{13}C SERC simulation of cyclohexane spirocyclopentane.

The same parameters can be used for spiro systems (Figure 4):



The analogous alkane is calculated as (Figure 1, quaternary carbon)

$$(FO + A_1 + A_2 + A_3 + A_4) + B_{11} + B_{21} + B_{31} + B_{41} + C_{11}$$
$$28.0 \quad 2.1 \quad 2.2 \quad 2.5 \quad 2.3 \quad 0.4$$

The corrective term 4 ppm expresses that this quaternary carbon belongs to a five-membered cycle (only the less-membered cycle is taken into account); δ_{cal} = 41.5 and δ_{exp} = 42.6 ppm.

Function Influence. The shifts of sp^3 carbons bearing a chemical function in their near environment are usually calculated by adding to the analogous alkane an increment depending on the nature of the function, its distance to the resonating carbon, and the linear (or not) character of the carbon framework. In fact, the influence of a functional group is by far more complex since it depends on its environment and that of the resonating carbon.[18a]

For any given substituent group G, α-substituent shifts evaluated by reference to hydrogen (H-^{13}C → G-^{13}C) largely vary with the ^{13}C environment (for example, the substituent shift induced by an iodine atom varies from −7.5 ppm in I-$^{13}CH_2Me$ to 28.9 ppm in I-$^{13}CMe_2tBu$). On the contrary, these variations are significantly dampened when evaluated by reference to a methyl group (Me-^{13}C → G-^{13}C). This probably results from the fact that the connectivity of the ^{13}C is maintained during the substitution.

Our previous studies have established that it is possible to sharply evaluate the influence of a function in adding two terms: $\delta = \delta_0 + \omega\Sigma\lambda$. The basic parameter δ_0, depending on the ^{13}C connectivity, corresponds to the reference compound in the family (alkyl environment reduced to methyl groups). In the corrective term $\omega\Sigma\lambda$, the parameter $\Sigma\lambda$ describes the branching of the alkyl environment, and the proportionality coefficient ω is related to the connectivity of the resonating carbon and represents the sensitivity of the ^{13}C to structural perturbations. So, δ_0 and ω are specific of the chemical family, and $\Sigma\lambda$ expresses the alkyl environment.[17,18] The topological parameters λ have been evaluated from the topological correlation of aliphatic alkynes chosen as reference family. Extensive studies showed their transferability to a wide range of chemical families.[17] This is the approach we are using when data are available (Table III). Otherwise, we select an increment in agreement with the usual values.[5,16]

Table III. α Functional Effect[a]

F	F-$^{13}CH_3$ δ_0	F-$^{13}CH_2R$ δ_0	F-$^{13}CH_2R$ $\Delta\omega$	F-$^{13}CHR_2$ δ_0	F-$^{13}CHR_2$ $\Delta\omega$	F-$^{13}CR_3$ δ_0	F-$^{13}CR_3$ $\Delta\omega$
F	66	63.9	−0.63	63.9	−0.29	66.1	−0.09
Cl	20	23.1	−0.22	30.2	−0.08	39.3	0.29
Br	4	10.6	0.04	21.5	0.37	34.9	0.60
I	−30	−18.8	0.57	−2.3	0.83	16.5	1.16
OH	43.5	40.6	−0.31	40.0	−0.16	40.4	−0.03
NH_2	22.5	20.8	−0.20	19.5	−0.11	19.5	−0.11
COMe	6.8	3.6	−0.06	7.7	0.05	12.0	0.16
COtBu	8.1	5.1	−0.10	7.4	0.04	16.4	0.18
CO_2H	−3.2	−4.8	0.02	−0.6	0.12	6.9	0.18
$CC{=}CH_2$	−1.2	−1.3	0.08	2.6	0.24	3.4	0.26
C≡CH		−13.1	−3.2	−9.2	−1.0	−3.2	
C≡CC	−10.1	−11.9	0.05	−7.0	0.17	−3.6	0.30

[a] δ_0 represents the function-induced shift on the reference terms (alkyl environment reduced to methyl), and $\Delta\omega$ is the sensitivity of the functional effect to changes in the alkyl environment where the function is introduced.

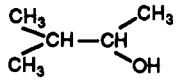
Table IV. β and γ Functional Effects[a]

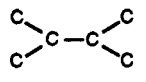
F	β	γ
F	−0.5	−3.5
Cl	1.5	−0.5
Br	2.5	0
I	3.5	1.0
OH	0.5	−1.0
OR	−1.5	−1.0
OCOR	−1.3	−1.0
NH_2	1.5	0.
NR_2	−3.0	1.5
NO_2	−3.0	−1.0
HCO	−5.0	0.
COR	−1.5	1.2
$HC{=}CH_2$	−1.5	0.5
C≡CR	0.5	0.

[a] These values must be weighted by the following coefficients according to the ^{13}C connectivity: $\Delta\delta_p = 1.94$, $\Delta\delta_t = 1.13$, $\Delta\delta_s = 1.56$, and $\Delta\delta_q = 0.38$.

For β- and γ-substitutions, the environment influence is weak, and (but for halogens) the contribution can be considered as only slightly dependent on the ^{13}C connectivity[18c] (Table IV).

So for the 3-methylbutanol-2, the analogous alkane is built





For the carbon bearing the hydroxyl group, the shift calculated in the corresponding alkane is

$$(FO + A_1 + A_2 + A_3) + B_{11} + B_{12}: \ 25.0 + 5.0 + 4.2 = 34.2$$

The hydroxyl α-effect is evaluated as

$$\delta = \delta_0 + \omega\Sigma\lambda = 40.0 - (0.16 \times 10.7) = 38.3$$

(λ values taken from ref 16). So

$$\delta_{cat} = 34.2 + 38.3 = 72.5 \qquad \delta_{exp} = 72.5 \text{ ppm}$$

For the CH situated in β, the analogous alkane is identical to the preceding one, and the β-effect is added:

$$\delta_{cal} = 34.2 + 1.13 \times 0.5 = 34.8 \qquad \delta_{exp} = 35.4 \text{ ppm}$$

Application of the above models to polyfunctionalized compounds points out the problem of interactions between functions (nonadditivity of functional effects) illustrated by the well-known example of halogen-substituted methanes

Table V. Chemical Shift Ranges of α-Substituent Effects for O, N, and C

α	^{13}C CH_3	^{13}C CH_2	^{13}C CH	^{13}C C
O	42/55	38/53	38/51	38/49
N	20/41	15/40	14/31	13/28
$-NO_2$	55	55	55	54
C	10/25	8/20	6/18	−2/13
−C≡	−5/−3	−6/−4	−5/−3	−3/−6

Table VI. Chemical Shift Ranges of β- and γ-Substituent Effect [for O, N, and C] according to the Connectivity of Resonating Carbon

atom	position β	position γ
CH_3	−10/+3	−3/+3
CH_2	−7.5/+2.5	−2/+2
CH	−5.5/1.5	−1.5/+1.5
C	−2/0.5	−0.5/+0.5

$CH_{4-n}X_n$).[16] Some corrective terms have been proposed and can improve calculated results.[10,11]

SIMULATION FROM FUZZY SUBSTRUCTURES

Frequently 2D NMR spectra provide carbon fragments with hanging bonds (as will be illustrated below). Is it possible to define what are (or may be) the atoms born by these hanging bonds? Indeed it is well-known that substituent effects heavily depend on the connectivity of the resonating carbon and the nature of its environment. So, clearly the direct examination of the chemical shifts is not sufficient. However, a more refined scrutiny of substituent shifts can give some insights. In this preliminary study, we limit the problem to sp^3 carbons with hanging bonds bearing either carbon, oxygen, or nitrogen, which represent the more frequent atoms encountered in structural elucidation.

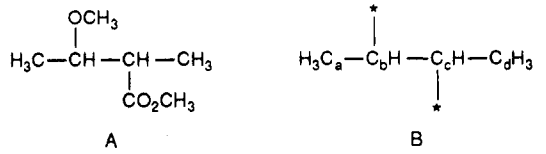
By examination of a series of simple systems, ranges of α-substituent shifts can be defined for the three types of atoms considered: C, N, O. For example, on a $^{13}CH_3$ carbon, substituent shifts associated to O—H, O—CH_3, O—$COCH_3$, and O—$CH{=}CH_2$ are respectively 43.5, 54.7, 45.0, and 46.0 ppm (ref methyl). So, an α-oxygen introduced on $^{13}CH_3$ can be associated to shifts above 42 ppm. Similar ranges can be derived for $^{13}CH_2$ and ^{13}CH. These ranges are gathered in Table V.

β-Substituent shifts do not vary much with the nature of G (about 10 ppm for Me and other groups, except for a few groups such as CHO or NO_2), and in any event they remain within a range of 10 ppm (Table VI). Similarly γ-effects can be taken as identical to that of methyl with a tolerance of ±2 ppm (except specific stereochemical situations).

In a carbon fragment, for a carbon bearing a hanging bond, we will consider that its shift is determined by the contribution of the α-unknown atom (that to determine) and that of the known part of its environment. Let us note that at this point the unknown part is considered as reduced to a single atom (that α- to ^{13}C), ignoring more remote atoms. As a consequence, only shift ranges and not precise values can be associated to the nature (C,N,O) of the first atom in the unknown part. Once the known moiety is evaluated, the difference to the experimental shift will provide the substituent shift due to the unknown α-group (within an acceptable tolerance). Comparison with ranges previously determined in simple systems will give some indication about the nature of this α-atom. In some cases, all the sites of the known

environment are not fully determined (other hanging bonds in β or γ). These free sites are modeled by methyls and given possible supplementary shifts on account of a preceding remark (see Table VI).

As an example, let us consider molecule (A) leading to fragment (B) where free sites are indicated by asterisks:



A B

For carbon C_b ($\delta = 78.3$), the carbon framework of the known environment is analogous to $^{13}CH(Me)_2i\text{-}Pr$, that corresponds to the value:

$$(FO + A_1 + A_2 + A_3) + B_{11} + B_{12}: \qquad 25.0 + 5.0 + 4.2 = 34.2$$

The free site on C_c corresponds to a β contribution in the range (−5.5 to +1.5). So

$$\delta_{exp} = 34.2 + \alpha + \beta$$

$$78.3 = 34.2 + \alpha + (-5.5/+1.5) \qquad \alpha = (42.6/49.6)$$

This value clearly indicates that the α-substituent on C_b begins by oxygen (see Table V).

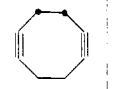
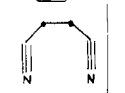
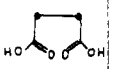
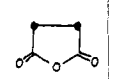
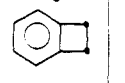
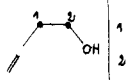
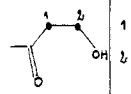
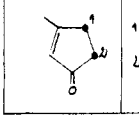
Similarly, for C_c ($\delta_{exp} = 45.6$):

$$45.6 = 34.2 + \alpha + (-5.5/+1.5) \qquad \alpha = (9.9/16.9)$$

The α-substituent begins probably with carbon; nitrogen being also possible with lower likelihood (see Table V).

From the examples treated up to now, for sp^3 carbon bearing an only free site, it appears that the tolerance range accepted also encompasses interactions between groups situated in the 1 and 2 positions, without explicit requirements to corrective terms to take them into account. Some of these border examples are gathered in Table VII.

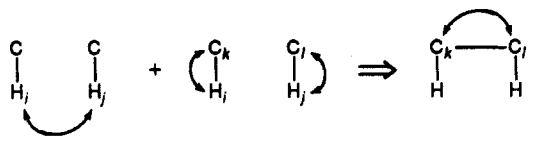
Table VII. Some Examples of Expanding $*CH_2{-}CH_2*$ Structures

Structure	δ_{exp}	α-subs.	Structure	δ_{exp}	α-subs.
	20.2	C		1 54.3 2 60.3	N N (O)
	14.6	C		58.7	N (O)
	28.9	C		1 45.9 2 65.0	N (C) O (N)
	28.2	C		64.5	O (N)
	29.5	C		54.7	N
	1 36.9 2 61.3	C (N) N (O)		1 55.9 2 66.8	N O (N)
	37.0	C (N)		67.6	O (N)
	1 46.1 2 57.5	N(C) N (O)	— — —	— —	— — —
				40.6	N (C)
	1 33.1 2 35.8	C (N) C (N)			

GENERATING FUZZY SUBSTRUCTURES FROM 2D NMR DATA

New techniques of 2D NMR now offer efficient tools to identify connected fragments of the carbon framework. According to experimental requirements, various facilities are provided. 2D INADEQUATE sequences would give the more complete information, allowing it to traverse the all carbon chain, going from every carbon to those bonded to it, and so on (stopping only when one heteroatom is encountered). Indeed, it was indicated that "if all the contours in a 2D-INADEQUATE spectrum are visible, its interpretation requires little time and can be carried out without any particular spectroscopic or chemical experience".[6] However, this rather insensitive technique requires large amounts of sample and cannot be retained as the only basis when building tools for an elucidation system. More structurally limited, but more widely accessible, diverse other sequences give some useful information about the carbon framework and allow the retrieval of information similar (although less precise) to that given by INADEQUATE.

In this paper we will use simultaneously the more conventional ^1H-^1H COSY and ^1H-^{13}C HETCOR correlations to fix what carbons are bonded to each other. Assembling information given by COSY peaks (H_i correlated to H_j by ^3J coupling) and HETCOR (H_i bonded to C_k and H_j bonded to C_l) indicates that carbons k and l are bonded. The information is quite comparable to that given by INADEQUATE, except that now the derived carbon chain also stops at quaternary carbons (bearing no protons).



In an elucidation process, once connected-carbon fragments have been determined, thanks to 2D data, the next step is to assemble them to each other to expand the structure. It is well-known, in the present state, that it is difficult to associate even a limited number of substructures to a given chemical shift. However, for most common compounds, those bearing only carbon, hydrogen, nitrogen, and oxygen and without small cycles (three- or four-membered), we use the simulation of fuzzy substructures for defining to what atoms (no-hydrogenated carbon or heteroelement) the hanging links of the recognized connected-carbon fragments are bonded.

This elucidation-directed process comprises three main steps:

determining connected-carbon fragments and the number of hydrogen atoms born by the carbons (COSY and HETCOR)

defining the hybridization of the carbons and the multiplicity of the C–C bonds to determine the hanging links

deriving, from simulation of fuzzy substructures, some insight about the possible atoms bonded to these hanging links.

This approach will be illustrated as a formal example on the structure of the quinolizidine alkaloid anagyrine recently

Table VIII. ^{13}C and Connectivity Data for Anagyrine (from Ref 22)

atom numbers	δ	no. of H[b]	no. of C[c]	hybrid	free sites
1	9.0	2	2	sp^3	
2	20.5	2	2	sp^3	
3	22.5	2	2	sp^3	
4	25.3	2	2	sp^3	
5	32.4	1	3	sp^3	
6	35.2	1	2	sp^3	1
7	51.3	2	1	sp^3	1
8	52.8	2	1	sp^3	1
9	54.2	2	1	sp^3	1
10	62.9	1	2	sp^3	1
11	104.5	1	1	sp^2	1
12	116.3	1	1	sp^2	1
13	138.5	1	2	sp^2	1
14	151.7	0		sp^2	2
15	163.4	0		sp^2	2

[a] The numbering is arbitrary (and is not identical to that of Rycroft et al.). [b] Number of neighboring hydrogens. [c] Number of neighboring carbons.

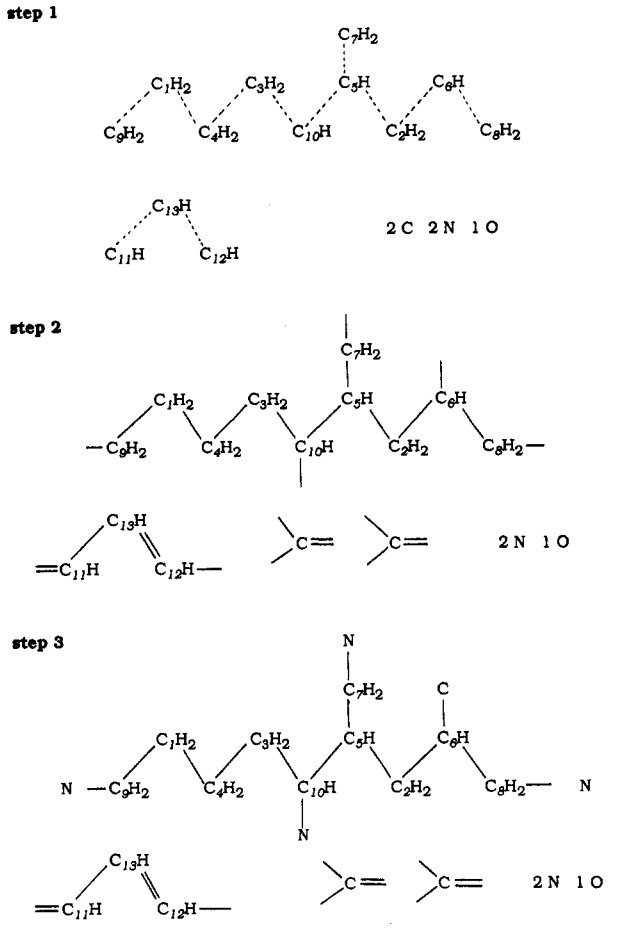


Figure 5. Expanding connected-carbon fragments for anagyrine. Step 1: determining the connected carbon fragments. Step 2: determining the hybridization of the carbons. Step 3: deriving the possible linked atom.

studied.[21] The available 2D data are gathered in Table VIII. The successive steps of the elucidation process are illustrated in Figure 5.

Determining the Connected Carbon Fragments. As previously indicated, this is easily carried out by examination of COSY and HETCOR patterns. Input data consist of a list of the COSY couples of correlated protons (i,j) and of bonded hydrogen/carbon HETCOR couples (i,k), from an arbitrary numbering of protons and carbons peaks (for instance, sequentially from low to high shifts...). Then, the program

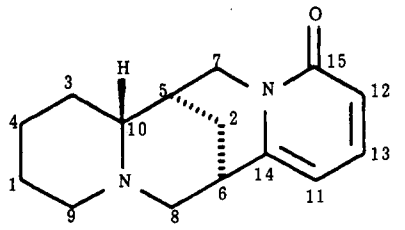


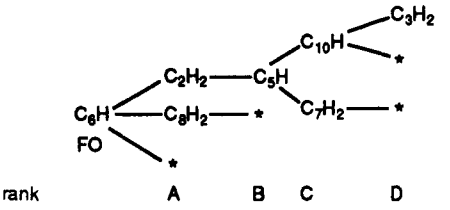
Figure 6. Structure of anagyrine (from ref 22).

determines for each carbon (k) the carbons $(l, ...)$ to which it is bonded.

Defining the Hybridization of the Carbons. This is performed by examination of their chemical shifts. From usual shift-range tables, ranges of chemical shifts values for sp^3, sp^2 (alkenes, aromatics, carbonyl, ...), and sp hybridization can be determined with few overlaps, provided that CH_3, CH_2, and CH carbons are treated separately (Figure 7). Joined to the number of carbons and hydrogens bonded, this new information allows for determining the number and the nature of the hanging links.

On the example of anagyrine, these first two steps lead to two connected-carbon fragments plus five isolated atoms. On the 15 carbons, five atoms have already four neighbors (hydrogens plus carbons) and, consequently, are sp^3 hybridized. Among the others, five atoms are determined as sp^3- with one free site, the last ones being sp^2-hybridized with, respectively, two carbons with 1 and 2 free sites, and one carbon with 0 free sites. So, seven carbons bear only a hanging bond. *What can we say about the first atom of this unknown substituent?*

Deriving the Possible Linked Atom. This is carried out via simulation from the fuzzy substructures as illustrated. For example



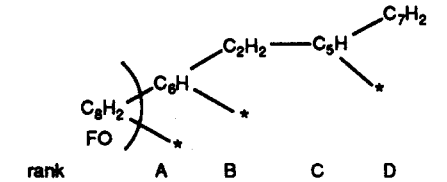
C_6 is tertiary; applying the corresponding valuated graph leads to:

$$(FO + A_1 + A_2 + A_3) + B_{11} + B_{21} +$$
$$25.0 \qquad 5.0 \quad 6.8$$
$$C_{11} + C_{12} + 3D_{ij} = 32.9$$
$$-2.2 \quad -2.3 \quad 0.2 \times 3$$

$$\delta_{exp}: \quad 35.2 = 32.9 + \alpha + \beta$$

$$\alpha + \beta = 2.3$$

According to the values gathered in Table VI for the substituents under study, the range of β effects is $-5.5 \rightarrow +1.5$. So, the α effect can be estimated as lying in the range of $+0.8 \rightarrow +7.8$. This corresponds to a substituent linked to the resonating carbon by a non-hydrogenated carbon (see Table V).



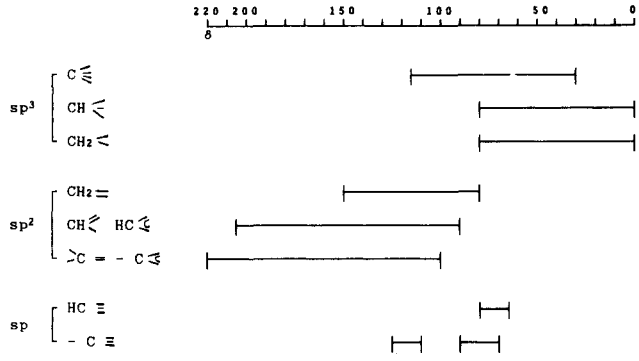


Figure 7. Chemical shift ranges according to the hybridization of the resonating carbon (for C, O and N substituents studied in this paper).

C_8 is secondary; from the corresponding graph

$$(FO + A_1 + A_2) + B_{11} + B_{12} + C_{11} + 2D_{ij} = 29.9$$
$$15.9 \quad 9.0 \quad 7.1 \quad -2.5 \quad 0.2 \times 2$$

$$\delta_{exp}: \quad 52.8 = 29.9 + \alpha + \beta$$

$$\alpha + \beta = 22.9$$

β effect on $^{13}CH_2$ is between −7.5 and +2.5; the α effect estimated between 20.4 and 30.4 corresponds to a substituent beginning by nitrogen.

In the same manner, we determine that the hanging bond for the carbons C_7, C_9, and C_{10} correspond to nitrogen.

The three preceding steps are automated and checked (the program is written in C++ and currently implemented on PC). After that, some additional information can be manually input. So, if for non-sp^3 carbons, the study is just beginning, in some favorable cases, substructures can be already defined. In the current problem, the moiety constituted by three sp^2-carbons is symmetric, the two ending atoms must each bear one hanging bond: one being simple, the other being double:

$$=C-CH=C-$$

According to the usual range of values, the atom linked by the double hanging bond is likely to be a carbon (—N=^{13}CH—CH= 145–160 ppm, O=^{13}CH—CH=, 180–200 ppm).

All these data are consistent with the published structure of anagyrine[22] (see Figure 6).

CONCLUSION

The models which are submitted have proved their efficiency for chemical shift prediction in compounds of moderate complexity, and in this preliminary work, we show the capability of simulation to provide structural information before the molecular puzzle step. After finding sp^3 carbons located in a saturated immediate environment, we will keep on investigating the possible information given by fuzzy substructures for more complex situations. This requires more refined models taking into account the nature of the cyclic fragments, stereochemistry, and polyfunctional interactions.

As final remarks, we would stress that we retained here only two of the most common sequences of 2D NMR (COSY and HETCOR) in their general use, although more sophisticated sequences[23,24] could provide helpful supplementary data (such as heteronuclear multiple-bond correlation experiments, HMBC, which correlate carbon and hydrogen separated by two or three bonds). On the same point of view, incorporating more specific information would of course improve the elucidation: Cosy peaks involving NH protons (proton nonlinked to a carbon) or aldehydic (O=C)H protons (with their characteristic shift), for instance, are examples of such useful data. In further steps, efficiency would be largely increased by integrating the SERC module in a multispectroscopy artificial intelligence approach able to cope with information issued from varied techniques.

REFERENCES AND NOTES

(1) Gray, N. A. B *Computer-Assisted Structure Elucidation*; Wiley-Interscience: New York, 1986.

(2) (a) Dubois, J.-E.; Carabédian, M.; Dagane, I. Computer-Aided Elucidation of Structures by ^{13}C-NMR. The DARC-EPIOS Methods: Characterizing Ordered Structures by Correlating the Chemical Shift of their Bonded Carbon Atoms. *Anal. Chim. Acta* **1984**, *158*, 217–233. (b) Carabédian, M.; Dubois, J.-E. A Combined Model of Multi-Resonance Subspectra/Substructure and DARC Topological Structure Representation. Local and Global Knowledge in the ^{13}C NMR DARC Database. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 564–574.

(3) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principle of NMR in One and Two Dimensions*; Oxford Science Publications, Clarendon Press: Oxford, 1987.

(4) Sanders, J. K. M.; Hunter, B. K. *Modern NMR Spectroscopy: A Guide for Chemists*; Oxford University Press: Oxford, 1987.

(5) Silverstein, R. M.; Bassler, G. C.; Morrill, T. C. *Spectrometric Identification of Organic Compounds*, 5th ed.; John Wiley and Sons: New York, 1991.

(6) Buddrus, J.; Bauer, H. Direct Identification of the Carbon Skeleton of Organic Compounds Using Double Quantum Coherence ^{13}C-NMR Spectroscopy. The INADEQUATE Pulse Sequence. *Angew. Chem., Int. Ed. Engl.* **1987**, *26*, 625–642.

(7) Bremser, W. HOSE—A Novel Substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355–365.

(8) Gray, N. A. B.; Nourse, J. G.; Crandell, C. W.; Djerassi, C. Stereochemical Substructures Codes for ^{13}C Spectral Analysis. *Org. Magn. Reson.* **1981**, *15*, 375–389.

(9) Shoulders, B.; Welch, S. C. A Very Brief, Rapid, Simple, and Unified Method for Estimating Carbon-13 NMR Chemical Shifts. *J. Chem. Educ.* **1987**, *64*, 915–918.

(10) (a) Fürst, A.; Pretsch, E. A Computer Program for the Prediction of ^{13}C-NMR Chemical Shifts of Organic Compounds. *Anal. Chim. Acta* **1990**, *229*, 17–25. (b) Fürst, A.; Pretsch, E.; Robien, W. Comprehensive Parameter Set for the Prediction of the ^{13}C-NMR Chemical Shifts of sp^3-hybridized Carbon Atoms in Organic Compounds. *Anal. Chim. Acta* **1990**, *233*, 213–222. (c) Fürst, A.; Pretsch, E.; Robien, W. Parameter Set for the Prediction of the ^{13}C-NMR Chemical Shifts of sp^2- and sp-hybridized Carbon Atoms in Organic Compounds. *Anal. Chim. Acta* **1991**, *248*, 415–428. (d) Pretsch, E.; Clerc, T.; Seibl, J.; Simon, W. *Tables of Spectral Data for Structure Determination of Organic Compounds*; Springer: Berlin, 1983 and 1989.

(11) (a) Cheng, H. N.; Ellingsen, S. J. Carbon-13 Nuclear Magnetic Resonance Spectral Interpretation by a Computerized Substituent Chemical Shift Method. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 197–203. (b) Cheng, H. N.; Bennet, M. A. Trends in Shift Rules in Carbon-13 Nuclear Magnetic Resonance Spectroscopy and Computer-Aided Shift Prediction. *Anal. Chim. Acta* **1991**, *242*, 43–56.

(12) (a) Lah, L.; Tusar, M.; Zupan, J. Simulation of ^{13}C NMR spectra. *Tetrahedron Comput. Methodol.* **1989**, *2*, 5–15. (b) Tusar, M.; Tusar, L.; Bohanec, S.; Zupan, J. ^{1}H and ^{13}C NMR Spectral Simulation. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 299–303.

(13) Ranc, M. L.; Jurs, P. C. Simulation of Carbon-13 Nuclear Magnetic Resonance Spectra of Quinolines and Isoquinolines. *Anal. Chim. Acta* **1991**, *248*, 183–193.

(14) Lindeman, L. P.; Adams, J. Q. Carbon 13 Nuclear Magnetic Resonance Spectrometry. *Anal. Chem.* **1971**, *43*, 1245–1252.

(15) Martin, G. J.; Martin, M.; Odiot, S. Theoretical and Empirical Calculations of the Carbon Chemical Shift in Terms of the Electronic Distribution in Molecules. *Org. Magn. Reson.* **1975**, *7*, 2–17.

(16) Kalinowski, H.-O.; Berger, S.; Braun, S. *Carbon-13 NMR Spectroscopy*; John Wiley and Sons: New York, 1988; pp 113–467 and references cited therein.

(17) Dubois, J.-E.; Carabédian, M. Modeling of Alkyl Environment Effects on the ^{13}C Chemical Shift. *Org. Magn. Reson.* **1980**, *14*, 264–271.

(18) (a) Doucet, J.-P.; Yuan, S. G.; Dubois, J.-E. Evolvement of Alpha Substituent Shifts in ^{13}C NMR: DARC PULFO Topological Correlation. *J. Chim. Phys.* **1984**, *81*, 219–224. (b) Doucet, J.-P.; Panaye, A.; Yuan, S. G.; Dubois, J.-E. Evolvement of Alpha Functional Shifts in ^{13}C NMR: Application of the DARC PULFO Topological Model for Acyclic Derivatives. *J. Chim. Phys.* **1985**, *82*, 607–611. (c) Yuan, S. G. Thesis. University Paris, 1984.

(19) Beierbeck, H.; Saunders, J. K. Analysis of ^{13}C Nuclear Magnetic Resonance Chemical Shifts of Acyclic Hydrocarbons. *Can. J. Chem.* **1980**, *58*, 1258–1265.

(20) Dubois, J.-E.; Panaye, A.; Attias, R. DARC SYSTEM: Notions of Defined and Generic Substructures. Filiations and Coding of FREL Substructures (SS) Classes. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 74–82.

(21) (a) Dubois, J.-E.; Doucet, J.-P.; Tiffon, B. Carbon 13 NMR: Alkyl Substituent Effects on the Chemical Shift of the Carbonyl Carbon in Aliphatic Ketones. *J. Chim. Phys.* **1973**, 805–806. (b) Dubois, J.-E.; Doucet, J.-P. ^{13}C NMR of Aliphatic Alkynes: Topological Analysis of Alkyl Substituent Effects on the Chemical Shift of sp Carbons by the DARC PELCO Method. *Org. Magn. Reson.* **1978**, *11*, 87–96. (c) Doucet, J.-P.; Dubois, J.-E. Alkyl Substituent Shifts in ^{13}C-NMR Spectra of Alkynes and Alkynols. Part 1. Topological Model for the Evaluation of the sp Carbon Shifts. *J. Chem. Res.* **1980**, S82–83 and M1101–1129. (d) Doucet, J.-P.; Dubois, J.-E.; Alkyl Substituent Shifts in ^{13}C-NMR Spectra of Alkynes and Alkynols. Part 2. Perturbations Induced by CC or OH groups on the sp^3 Carbon Shifts. *J. Chem. Res.* **1980**, S84–85 and M1130–1149.

(22) Rycroft, D. S.; Robins, D. J.; Sadler, I. H. Assignment of the ^1H and ^{13}C NMR Spectra of the Quinolizidine Alkaloid Anagyrine and Determination of its Conformation. *Magn. Reson. Chem.* **1991**, *29*, 936–940.

(23) Nuzillard, J. M.; Massiot, G. Computer-Aided Spectral Assignment in Nuclear Magnetic Resonance Spectroscopy. *Anal. Chim. Acta* **1991**, *242*, 37–41.

(24) Eggenberger, U.; Bodenhausen, G. Analysis of Two Dimensional Nuclear Magnetic Resonance Spectra with Relayed Proton-Proton-Carbon Magnetization Transfer: A Step toward Automated Structure Elucidation. *Anal. Chem.* **1989**, *61*, 2298–2306.