

# Single-Resonance Subspectra/Substructure Investigations of the $^{13}\text{C}$ DARC Databank. Representation of Local and Global Topological Knowledge

JACQUES-EMILE DUBOIS\* and MICHEL CARABÉDIAN

Institut de Topologie et de Dynamique des Systèmes de l'Université Paris 7, associé au CNRS, URA 34, 1 rue Guy de la Brosse, 75005 Paris, France

Received March 27, 1991

The knowledge content of a  $^{13}\text{C}$  NMR databank, which contains a large amount of associated structural (St) and spectral (Sp) information (a number of special terms used in this paper are defined in Table IV), has been approached through a combination of two complementary models: single-resonance subspectra/substructure (SRS) and DARC structural organization. Various representations of (St,Sp) pairs, both topological and statistical, are defined in order to capture local and global structural environmental factors. The databank has been organized and its content investigated with the SRS model. Substructures (St) are described using the DARC concept of an Environment that is Limited, Concentric, and Ordered (ELCO). The general profile of the databank has been defined in terms of the distribution of all the  $\delta^{13}\text{C}$  occurrences, and different knowledge levels have been explored via partial neglect of active features in the ELCO fragments. It is shown that the ELCO<sub>0</sub>, which identifies the first neighbors of a central carbon atom with their connectivities and unspecified second neighbors, is a good compromise for the definition and use of generic knowledge in various applications of the SRS model. The growth in global St knowledge caused by the updating of the databank was evaluated and found to level off rapidly for the more generic ELCOs (A and b). This emphasizes the basic differences between data, information storage, and knowledge acquisition. This SRS model was found to have limitations which stem from an imbalance between the strong structural data and the weak spectral data.

## INTRODUCTION

More than ten million chemical compounds have been registered to date by Chemical Abstracts Service, and each year a further 500 000 new compounds are added to this file. Faced with the extreme diversity of chemistry, spectroscopic databanks, which may contain tens of thousands of structures, together with the corresponding spectra, cannot hope to satisfy all chemists' potential queries by means of simple documentation procedures. Strategies for assignment of structures to new compounds from spectra are an absolute necessity. The definition of generic knowledge with its broad field of application and of appropriate methods for its representation seems increasingly to be the key to the development of systems capable of supplying the assistance required to identify a new structure or to interpret its spectrum—the different operations depicted in Figure 1.

In NMR spectroscopy, since the discovery of the relationship between the chemical shift of the signal of a nucleus and its environment, advances in experimental techniques have not been accompanied by the development of the theoretical models required for physical interpretation of the information produced. To mitigate this absence and to exploit the information, chemists very quickly formalized the empirical knowledge needed to solve the different problems shown in Figure 1, which are centered on the structure/spectrum relationship. The requirements in the structural (St) and the spectral (Sp) components of this St.Sp knowledge for solving all these problems are not necessarily the same.

Two parallel approaches have been adopted to formalize this complex knowledge. The first, inspired by the pragmatic model developed by Grant and Paul,<sup>1</sup> concerns the definition of additive parameters. In the Grant and Paul model, the contributions of the environment sites of a  $^{13}\text{C}$  nucleus to its chemical shift were described for a particular family of compounds. Today, more sophisticated systems based on the use of topological, geometrical, and electronic descriptors simulate precisely the influence of the structural environment on the  $^{13}\text{C}$  chemical shifts.<sup>2</sup> These parametric functions are very sensitive to the homogeneity of their reference population to

which their field of application is closely tied. In such models, one deals with small and coherent sets of structures, and researchers seem more motivated by interpretation of mechanisms than by prediction through correlation.

The second approach, developed more pragmatically for greater generality, aims at prediction and interpretation of the relationships linking structures to their spectra. Pairs of corresponding structural and spectral information were identified and selected, very often for their success in some problems of interpretation or structure elucidation. This type of correspondence throughout the various sets of (St,Sp) knowledge pairs constitutes the core of the model called Single-Resonance Subspectra/Substructure (SRS).<sup>3</sup> It has been developed, more often on a case-by-case basis, together with computer systems which can extract St.Sp knowledge from spectroscopic databanks. It supports numerous applications in spectral simulation<sup>4,5</sup> and most structure elucidation systems, including DENDRAL,<sup>6</sup> CASE,<sup>7</sup> CHEMICS,<sup>8</sup> and ACCESS.<sup>9</sup> This model is a direct transcription of the chemical shift concept. It expresses the uncertainty or range of chemical shifts for given types of carbon perturbed by a variable environment. It was proposed as early as 1953<sup>10</sup> in connection with information supplied by  $^1\text{H}$  NMR.

The expression of parametric functions, as well as that of the SRS rules, relies upon an accurate description of the environmental effects experienced by each nucleus in the molecule. The search for the precision leads to an increased specificity of these environments or of their descriptors. The improved precision in this type of St.Sp relationship is then accompanied by a restriction of their field of application.

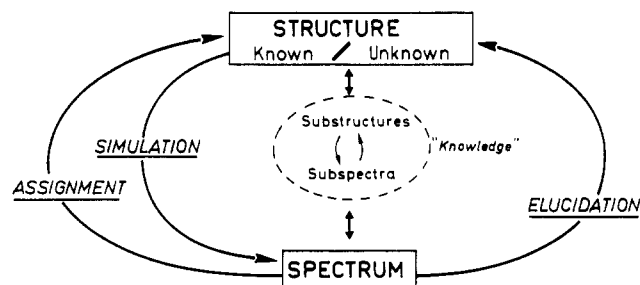
The diversity of chemical compounds and the complexity of the environmental effects, expressed implicitly in the  $^{13}\text{C}$  NMR spectra, cannot be globally modeled by the juxtaposition of such knowledge linking precision and specificity. The limitations inherent in this approach, which are often seen in hindsight,<sup>2,4,5,11</sup> are here analyzed concretely by means of a global evaluation of the SRS model.

The efficiency of this SRS model lies in the linkage of each single  $\delta^{13}\text{C}$  with its environment. The nature and definition

**Table I.** Evolution of the  $H(\delta, E)$  Distribution According to the Number of Reference Structures Considered<sup>a</sup>

no. of St	no. of $\delta$	$\delta^*$	Occ <sup>*</sup>	% <sup>*</sup>	$\delta^*$	Occ <sup>*</sup>	% <sup>*</sup>	$\delta^*$	Occ <sup>*</sup>	% <sup>*</sup>
1000	18659	22	425	2.7	71	125	0.7	128	300	1.7
1100	20042	22	453	2.7	68	126	0.6	128	326	1.7
1200	21493	28	498	2.3	71	136	0.6	128	354	1.7
1300	23349	29	561	2.4	71	139	0.6	128	473	2.2
1400	24852	29	624	2.5	68	148	0.6	128	528	2.5
1500	26038	29	662	2.5	68	148	0.6	128	616	2.5
1600	27268	29	695	2.5	68	148	0.5	128	618	2.4
1700	28847	29	728	2.5	72	157	0.5	128	657	2.4
1800	30274	29	740	2.4	72	173	0.6	128	665	2.3
1900	32552	29	752	2.3	71	217	0.7	128	668	2.2
2000	34393	29	786	2.3	71	243	0.7	128	679	2.1
15867	95043	29	3810	1.9	71	1131	0.6	128	5042	2.6

<sup>a</sup>No. of  $\delta$  = number of  $\delta^{13}\text{C}$  values.  $\delta^*$  = maximum occurrence (Occ<sup>\*</sup>) value for  $\delta^{13}\text{C} \leq 60$  ppm.  $\delta^*$  and Occ<sup>\*</sup> for  $60 < \delta^{13}\text{C} \leq 90$  ppm.  $\delta^*$  and Occ<sup>\*</sup> for  $90 < \delta^{13}\text{C}$ . % = Occ/no. of  $\delta$ .

**Figure 1.** Structure spectrum correlations and various applications using St.Sp knowledge at substructure/subspectra levels.

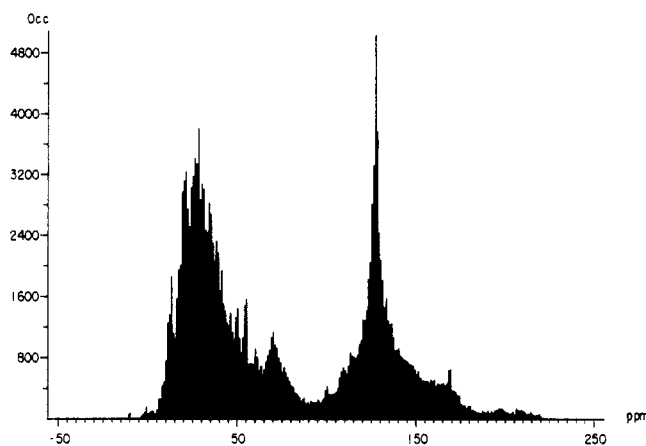
of the environment will be considered here through statistics and an evaluation of the effect of partial neglect of the information retained in the substructural part of the (St.Sp) pair. This evaluation concerns the population of the  $^{13}\text{C}$  NMR DARC-PLURIDATA.III ( $^{13}\text{CDP}_{\text{III}}$ ) databank and the various ELCOs present in the knowledge base of our EPIOS system.<sup>12,13</sup> The comparison between the original structural and spectral information and the knowledge obtained from its SRS model shows the difficulty involved in reconciliation of the precision of this type of knowledge with the extension of its area of application.

#### HOLISTIC REPRESENTATION OF BASIC "STRUCTURE/ $\delta^{13}\text{C}$ INFORMATION" OF THE $^{13}\text{CDP}_{\text{III}}$ DATABANK

The structure/spectrum reference population used to define the knowledge in a system largely determines the limits of the system's competence. This population forms, after modeling, the heart of the potential field of application engendered. It also constitutes the first test area permitting evaluation of the formalized knowledge. In spectral simulation systems, this evaluation is carried out by calculation of the spectra of available structures, followed by comparison of the spectral data with their experimental values. Conversely, for elucidation, the spectra are interpreted to generate the corresponding reference structures. Such operations are more difficult in an application field of indeterminate boundaries.

Here we propose a global treatment to evaluate, a priori, the capacity of SRS-type knowledge to describe the diversity of structure/ $\delta^{13}\text{C}$  relationships. This treatment involves a comparison of the set of reference structural and spectral information gathered in the databank with the resulting set of knowledge extracted and organized according to the SRS model. With this end in mind, we propose a **holistic representation** of our  $^{13}\text{CDP}_{\text{III}}$  databank, which currently contains 15867 structures and 21499 spectra (195043  $\delta^{13}\text{C}$  values). The distribution of chemical shifts in this databank is shown in Figure 2.

The occurrence frequency associated with each value between -30 and +250 ppm, relative to TMS, is the sum of the

**Figure 2.** Distribution  $H(\delta, E)$  of the 195043  $\delta^{13}\text{C}$  shifts in the  $^{13}\text{CDP}_{\text{III}}$  databank. The different  $E$  environments of  $^{13}\text{C}$  carbons in the 15867 reference structures are implicitly described by this  $\delta^{13}\text{C}$  distribution.

occurrences of peaks at that frequency in the 21499 spectra of the databank. A structure may be associated with several spectra measured under different conditions, and so several chemical shifts may be associated with the same carbon. Each occurrence of a chemical shift is recorded in Figure 2, which represents the number of carbon atoms in the whole databank assigned to a given chemical shift. Similarly, a given  $\delta^{13}\text{C}$  can correspond to different carbons, chemically equivalent in a single structure, to carbons with an identical local environment in different structures, or to carbons with different environments in different structures. The number and diversity of these structural situations, as well as the effects of experimental conditions, are described implicitly by the single spectral information, the  $\delta^{13}\text{C}$ .

We retain the profile of the distribution obtained,  $H(\delta, E)$ , as a characteristic and significant image of the influence of the structural environments  $E$  on the  $\delta^{13}\text{C}$  chemical shifts. These environments correspond to those having an effect on the carbons of the 15867 reference structures.

In order to verify the stability of this physical image,  $H(\delta, E)$ , of the databank and, more importantly, to show its general character, its development was followed as the number of chemical shifts and structures considered was increased. The results are shown in Table I. Here, the distribution  $H(\delta, E)$  is characterized schematically by the appearance of three maxima centered at  $\delta^{13}\text{C} = 29, 71$ , and  $128$  ppm and by their relative heights. The final values of these descriptors over the whole databank are approximated well by the values obtained with a subpopulation of 2000 spectra or 34393 chemical shifts, as can be seen from the last two lines of Table I.

This result underscores the nuance that exists between knowledge and information. According to the defined knowledge, the position and height of the peaks in the dis-

tribution, the accumulation of information does not necessarily lead linearly to the enrichment of knowledge—a point that will be discussed later. However, it is reasonable to suppose that the representation obtained with this databank is characteristic of experimental structure/ $\delta^{13}\text{C}$  relationships and corresponds overall to that which would be obtained from larger populations.

### MORPHOLOGY OF THE $\delta^{13}\text{C}/E$ DISTRIBUTION

In order to situate the terms of the comparisons proposed below between the  $\delta^{13}\text{C}/E$  information and its image derived from SRS knowledge of increasing specificity, a partial interpretation of the distribution in Figure 2 is necessary. This interpretation consists only of some very general identification of the structural characteristics associated with the main peaks in the distribution.

The two main bands in the  $H(\delta, E)$  distribution shown in Figure 2 and Table I represent the proportions of  $\text{C}_{\text{sp}^3}$  and  $\text{C}_{\text{sp}^2}$  carbons in the population. These wide bands, called  $\sigma$  and  $\pi$ , respectively, have Gaussian shapes, and this suggests that the broader diversity of  $E$  environments affects the  $\text{C}_{\text{sp}^3}$  carbons in the  $\sigma$  band ( $\delta^{13}\text{C} > 90$  ppm). The exclusive presence of  $\sigma$ -bonds does indeed allow the most highly varied substitutions, qualitatively and quantitatively, in the immediate environment of these carbons. The wide  $\pi$ -band of  $\text{C}_{\text{sp}^2}$  carbons centered at 128 ppm, on the other hand, reveals both the homogeneity of their behavior when linked to another carbon and the highest occurrence frequency (Occ = 5.042). The third peak in Figure 2, with a maximum at  $\delta^{13}\text{C} = 71$  ppm, has fewer occurrences and corresponds primarily to a  $\text{C}_{\text{sp}^3}$  carbon carrying a heteroatom (typically O or N). Such carbons are defined as "x" in Table I.

This distribution curve, which provides a global characterization of the variability of the  $\delta^{13}\text{C}$  chemical shift with varying environments  $E$ , is the reference representation  $H(\delta, E)$  that must be accounted for by the SRS model. To evaluate the potentials of different ELCO, such as  $\text{ELCO}_A$ ,  $\text{ELCO}_b$ , or  $\text{ELCO}_B$ , different  $H(\Phi, \text{ELCO})$  distributions have been constructed and compared to this reference distribution in order to illustrate the distortions associated with these structural descriptors and the SRS model.

### ELCO MODELING AND SRS RESOLUTION

The SRS model supports most of the applications in spectral simulation or structure elucidation. The choice of the depth and/or the genericity of the fragments used are not justified explicitly. They are generally selected pragmatically to improve the efficiency of the resulting systems. This choice is usually class- or family-oriented.<sup>1-5</sup>

In this paper, on the contrary, the choice of substructure is based on a statistical analysis of an extended databank, organized topologically rather than by family or category. The topology of the structural environment of a carbon, as used in the additive model of Grant and Paul,<sup>1</sup> constituted the first element which allowed rationalization of the  $\delta^{13}\text{C}$  shifts in alkanes. It expresses, in fact, the intuitive perception of environmental effects in simple terms of the connectivity between atoms and the distances from the signaling nucleus. Described according to the concepts ELCO defined in the DARC system,<sup>14</sup> it provides appropriate support for the modeling of structure/ $\delta^{13}\text{C}$  relationships,<sup>15,16</sup> their organization in databanks,<sup>17,18</sup> and their exploiting in the knowledge bases of structure elucidation systems.<sup>7,9,13</sup>

The influence of different types of topological primitives on SRS models of structure/ $\delta^{13}\text{C}$  relationships is evaluated with the help of the ELCO used in the EPIOS knowledge base.<sup>13,19</sup> These ELCO, which increase in specificity from  $\text{ELCO}_A$  through  $\text{ELCO}_b$  to  $\text{ELCO}_B$  according to the number of atom

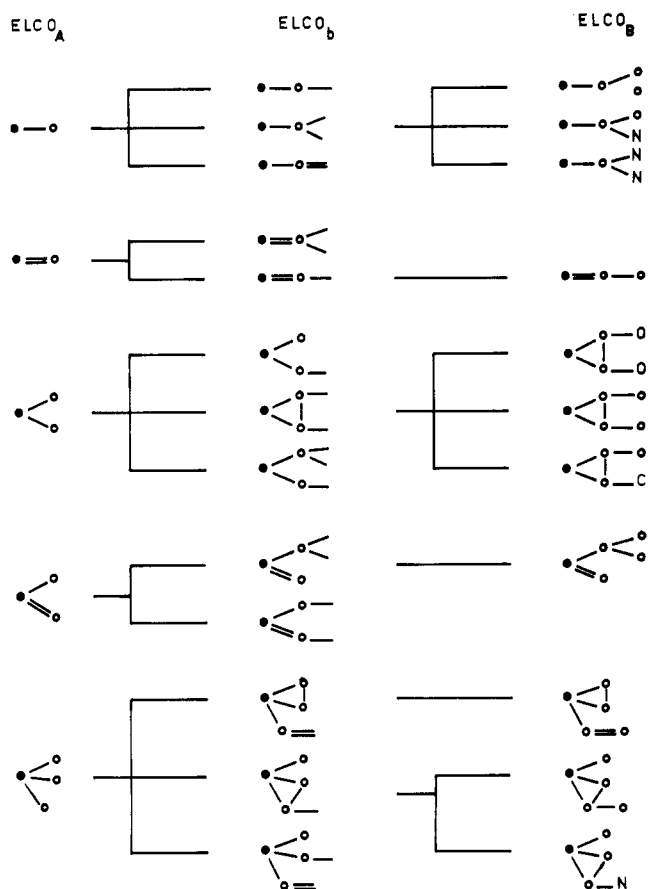


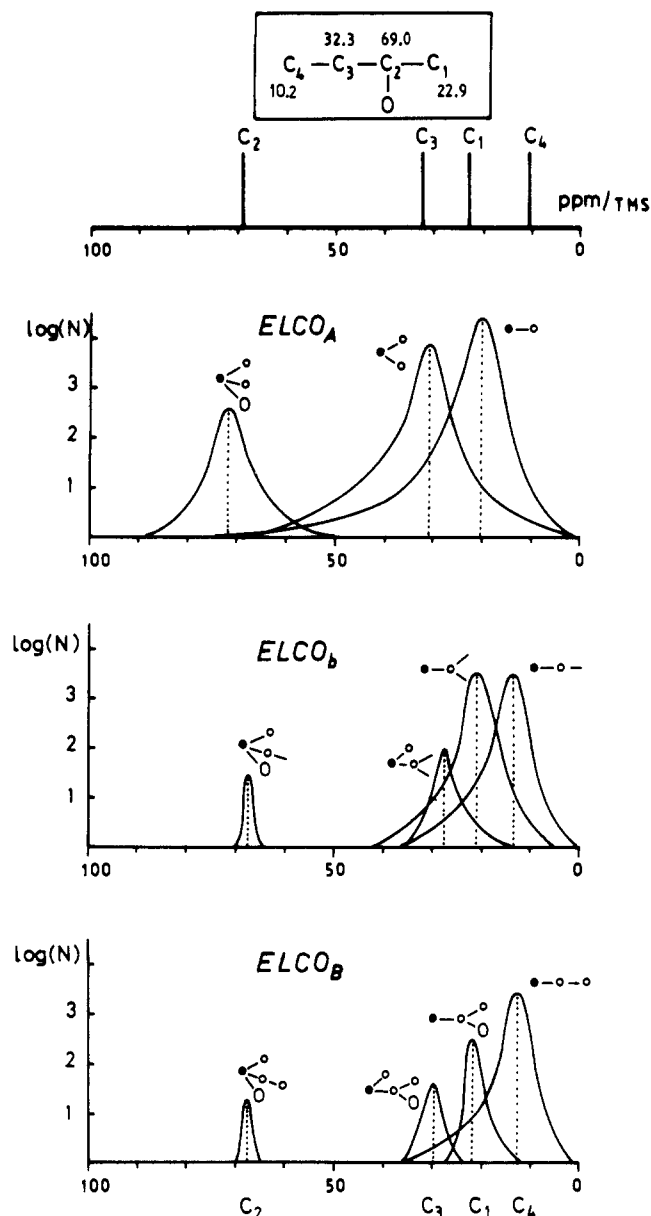
Figure 3. Hierarchical filiation of Environments that are Limited, Concentric, and Ordered (ELCO) in the EPIOS system knowledge base. ● = focus atom,  $\text{C}_{F_0}$ , ○ =  $\text{C}_{A_i}$  and  $\text{C}_{b_j}$ .

shells around their focus, are hierarchically affiliated in this knowledge base, as shown in Figure 3. Atoms  $A$  and  $B$  correspond respectively to the first and second shells around the focus atom;  $a$ ,  $b$ , etc. are the labels associated with the bond to the next atom, which may be defined specifically or generically.

In the preceding  $H(\delta, E)$  representation, the occurrence of each  $\delta^{13}\text{C}$  value corresponds to the occurrences of one or more  $E$  environments around a  $^{13}\text{C}$  atom, all producing the same behavior for this carbon. These various  $E$  environments, whose active part typically extends to the position  $\delta$  to the focal atom, may be different or may correspond to a similar local environment for this carbon. The definition of ELCO aims at elucidating, in the latter case, a common part of these local environments. Depending upon the specificity of the chosen ELCO, more or less of the environment is topologically encoded.

Ideally, a set of  $\text{ELCO}_B$ , describing the environment up to the  $\delta$ -position, relative to the central atom, would make it possible to define the influence of all the environments acting on the carbons of the reference structure. Practically, as is shown later, such a set is difficult to exploit in simulation or elucidation systems where the generic nature of the knowledge used is a necessary condition. Experience led us to limit the specificity of our ELCO primitives to that of the  $\text{ELCO}_B$ , including therefore only two rows of neighbors.

Each ELCO ( $A$ ,  $b$ ,  $B$ ) appears with varying frequency in the reference  $E$  environments, and each of their occurrences corresponds to one of the chemical shifts determined for their central carbon by one of these values of  $E$ . A set of chemical shifts is thus associated with an ELCO to characterize the behavior of its central carbon. Only the limits of the range of variation of these values,  $\delta^{13}\text{C}_{F_0}$ , and its average value are considered by the SRS rules.

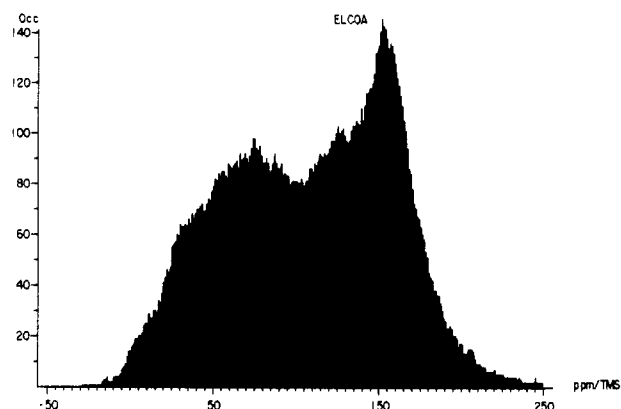


**Figure 4.** Increases in the precision of the spectrum simulation with the specificity of the ELCO considered. Each ELCO is represented schematically by its  $\delta^{13}\text{C}$  distribution.

As ELCOs increase in specificity, the SRS rules provide greater precision, as can be seen from Figure 4. The simulated spectrum of 2-butanol is progressively sharpened by going from  $\text{ELCO}_A$  to the specific  $\text{ELCO}_B$  that constitute it. This determining role of specificity on the precision of the SRS model motivated the introduction of explicit terms to characterize the shell level of substructures used in spectral simulation systems.<sup>4,5</sup> Such increased specificity, which is not always sufficient to achieve the precision of the reference spectra, is generally desirable. Calibration of these spectra, for compounds that are particularly sensitive to changes in experimental conditions, should be carried out beforehand to ensure compatibility of SRS knowledge with the potential field of application.<sup>20</sup>

#### INTERVAL SCALING DISTRIBUTIONS AND ELCO DISPERSION

Modeling of specific information,  $\delta^{13}\text{C}/E$ , by generic SRS knowledge,  $\Phi\delta^{13}\text{C}/\text{ELCO}$ , leads to a significant compression of this information. A single ELCO is proposed for different  $E$  environments and a  $\Phi\delta^{13}\text{C}$  range is substituted for a  $\delta^{13}\text{C}$  distribution. In order to appreciate the prospective advantage



**Figure 5.** Distribution  $H(\Phi, \text{ELCO}_A)$ . Each  $\delta^{13}\text{C}$  value is associated with the number of  $\text{ELCO}_A$  for which behavior of the focus  $\Phi\delta^{13}\text{C}_{F_0}$  is compatible with this value ( $\delta^{13}\text{C} \in \Phi\delta^{13}\text{C}_{F_0}$ ).

**Table II<sup>a</sup>**

	$\text{ELCO}_A^\sigma$	$\text{ELCO}_B^\sigma$	$\text{ELCO}_C^\sigma$	$\text{ELCO}_A^\pi$	$\text{ELCO}_B^\pi$	$\text{ELCO}_C^\pi$
$N$	314	3983	7937	320	2431	6135
$\Phi$	21.1	7.6	4.8	26.4	14.4	6.3
$\Sigma\delta$		115 794			72 076	

<sup>a</sup>  $N$  = number of ELCO;  $\Phi$  = average width of  $\delta^{13}\text{C}_{F_0}$  ranges;  $\Sigma\delta$  = number of  $\delta^{13}\text{C}_{F_0}$  associated with  $\text{ELCO}^\sigma$  and  $\text{ELCO}^\pi$ .

of this knowledge organization, its ability to describe the reality of the  $\delta^{13}\text{C}/E$  reference relations shown in Figure 2 must be studied. To this end, the distributions of knowledge obtained with the different ELCO primitives were compared. The results are shown in Figures 5, 6, and 7. Each ELCO is related to a subpopulation of reference structures which determine a set of  $\delta^{13}\text{C}$  values covering a  $\Phi\delta^{13}\text{C}$  interval. This  $\Phi\delta^{13}\text{C}$  interval translates the effects of environments outside the ELCO throughout the whole databank onto the  $\delta$  axis. These distributions  $H(\Phi, \text{ELCO})$  no longer express the effective occurrences of each  $\delta^{13}\text{C}$  value as a function of the environments  $E$  of the  $^{13}\text{C}$  nucleus, but they are now **interval scaling distributions** representing the number of ELCO associated with each  $\delta^{13}\text{C}$ .

Such structural and spectral approximations in the SRS model, ELCO, and  $\Phi\delta^{13}\text{C}$  cause the distortions in the initial reference distribution  $H(\delta, E)$ . The distortions correspond to a partial neglect of the structural information, linked to the depth or the nature of the ELCO. With more information, the distortions are lessened and the profiles of the global knowledge curves,  $H(\delta, E)$  and  $H(\Phi, \text{ELCO})$  should become increasingly similar. In the following section, interpretation of these distributions  $H(\Phi, \text{ELCO})$  differentiates between  $\text{ELCO}^\sigma$ , centered on a  $\text{C}_{\text{sp}^3}$  carbon, and  $\text{ELCO}^\pi$ , which is centered upon a  $\text{C}_{\text{sp}^2}$  carbon.

**Generic Knowledge through  $\text{ELCO}_A$ .** Modification of the initial  $H(\delta, E)$  distribution shown in Figure 2, using the model based upon the  $\text{ELCO}_A$ , results in a higher occurrence of  $\text{ELCO}_A^\sigma$  (0–90 ppm) than of  $\text{ELCO}_A^\pi$  (90–250 ppm) as shown in Figure 5. Although the databank contains similar numbers of  $\text{ELCO}_A^\sigma$  (314) and  $\text{ELCO}_A^\pi$  (320), the  $\text{ELCO}_A^\sigma$  appear almost twice as frequently in the reference structures, where their occurrences are respectively 115 794 and 72 076, as shown in Table II. This difference in representation explains the change in the relative heights of the two wide bands of the initial distribution (Figure 2). On average, each  $\text{ELCO}_A^\sigma$  corresponds to a compression of 369  $\delta^{13}\text{C}/E$ , while for each  $\text{ELCO}_A^\pi$ , 225  $\delta^{13}\text{C}/E$  values are grouped together. The first wide band of the initial distribution, centered at 29 ppm, corresponds essentially to the 5420  $\delta^{13}\text{C}$  signals associated with the four  $\text{ELCO}_A^\sigma$  in the 25–35-ppm interval (the  $\text{ELCO}_A^\sigma$  entries in Table III).

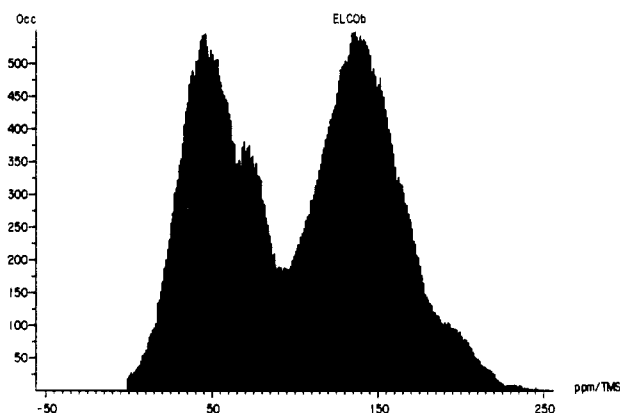
Table III<sup>a</sup>

ELCO <sub>A</sub> <sup>σ</sup>				
$n\delta^{13}\text{C}$	25 115	38 099	14 858	6025
$\Phi\delta^{13}\text{C}$	0-40	0-60	0-80	5-90
max_oc	1800	2400	800	420
$\delta_{\text{max}}$	25	30	30	35

ELCO <sub>A</sub> <sup>π</sup>				
$n\delta^{13}\text{C}$	1038	8347	4523	26 243
$\Phi\delta^{13}\text{C}$	70-155	80-160	75-175	95-145
max_oc	90	400	170	4000
$\delta_{\text{max}}$	115	130	130	128

<sup>a</sup>  $n\delta^{13}\text{C}$  = number of  $\delta^{13}\text{C}_{F_o}$ ,  $\Phi\delta^{13}\text{C}$  = variation range of  $\delta^{13}\text{C}_{F_o}$ ,  
 Max\_oc = maximum occurrence of a  $\delta^{13}\text{C}_{F_o}$ ,  $\delta_{\text{max}}$  =  $\delta^{13}\text{C}_{F_o}$  for max\_oc.

Figure 6. Distribution  $H(\Phi, \text{ELCO}_b)$ : (see Figure 5).

The second peak, that of the  $\text{ELCO}_A^\pi$ , centered at 128 ppm, corresponds primarily to the five motifs shown in the lower half of Table III. This peak thus contains the 4900  $\delta^{13}\text{C}$  shifts in the 120–135-ppm range. There is an asymmetry to the wide  $\sigma$  band, whose maximum is now at 75 ppm (cf. 29 ppm in the initial distribution). This indicates the relative importance assumed by the subpopulation of the  $\text{ELCO}_A^\sigma$  in which a heteroatom, usually oxygen or nitrogen, is found in the  $\alpha$ -position relative to the focus atom, causing a shift in the absorption of the focus atom to beyond 50 ppm. This subpopulation is less generic than its carbon homologue and groups proportionately less information on more  $\text{ELCO}_A^\sigma$ . Some 100  $\text{ELCO}_A^\sigma$  can be associated with the 71-ppm peak and some 60 with the 29-ppm value. The occurrences of these same shifts in the reference distribution,  $H(\delta, E)$ , are 1131 and 3810, respectively (see Table I).

Given the average width of the  $\Phi\delta^{13}\text{C}$  ranges of the  $\text{ELCO}_A$  (Table II) and of their overlap, each  $\delta^{13}\text{C}$  is associated with 60  $\text{ELCO}_A$  on average or 9.5% of the whole set of 634  $\sigma$  and  $\pi$   $\text{ELCO}_A$ . This figure can be taken as a measure of the selectivity of the  $\Phi\delta^{13}\text{C}/\text{ELCO}_A$  rules that can be compared to that of the more specific  $\text{ELCO}$ . The approximations associated with the  $\text{ELCO}_A$  and the sizable distortions they cause in the effective  $\delta^{13}\text{C}/E$  relations are not conducive to their direct use in elucidation systems. They can nonetheless make it possible to diminish the risk of silence introduced into strategies based upon more specific  $\text{ELCO}$  primitives. Put another way, search recall will be higher, but precision will be lower.

**Resolution Potential of  $\text{ELCO}_b$ .** A more satisfactory description of the distribution  $H(\delta, E)$  values is obtained with the  $\text{ELCO}_b$ , as can be seen from Figure 6. The  $\text{ELCO}_b$  allow

Table IV. Glossary of Terms Used in This Paper

$\text{ELCO}_i$	Environment that is Limited, Concentric, and Ordered. A primitive structural fragment defined around a focus atom ( $F_o$ ) describing its environment out to the $i$ th neighbor ( $i = A, B$ , or $C$ ) or as far as the bonds reaching this layer ( $i = a, b$ , or $c$ ). $\text{ELCO}_b$ include "a" bonds, $A_i$ atoms, and "b" bonds, but "B" atoms are not specified and are generic atoms.
$\text{ELCO}^\sigma$	$\text{ELCO}$ whose focus is an $\text{sp}^3$ carbon.
$\text{ELCO}^\pi$	$\text{ELCO}$ whose focus is an $\text{sp}^2$ carbon.
$E$	Structural Environment of an atom or of an $\text{ELCO}$ , defined by a DARC order.
$\Phi\delta^{13}\text{C}$	$\delta^{13}\text{C}$ shift variation range of a carbon.
data	Represents the results of an action on objects. For example, $\delta^{13}\text{C}$ is a datum from the effect of a magnetic field on a particular carbon nucleus.
information	Data transformed into information required for an objective. This production of information is controlled by knowledge associated with the subject.
knowledge	Interaction of symbols derived from information, often handled as symbolic expressions. Knowledge can be structured in successive layers and has organizing power.
(St.Sp)K	St is organized around a focus and its $\text{ELCO}_b$ environment plus the structural features associated with the various $F_o-A_i$ pairs and the $\text{ELCO}_i$ environments of each $A_i$ . The structural overlap leads to a circular description that goes as far as the atoms of the C shell, which are encoded generically. Sp in the MRS description deals with the $\delta^{13}\text{C}$ shifts of the focus and the $A_i$ atoms.
SRS	Single-Resonance Subspectra/Substructure model in which a substructure is characterized by the $\delta^{13}\text{C}_{F_o}$ behavior of its focus. The substructure is encoded as a DARC $\text{ELCO}_i$ .
MRS	Multi-Resonance Subspectra/Substructure model in which an $\text{ELCO}_i$ is characterized by the $\delta^{13}\text{C}_{F_o}$ behavior of its focus and by the $\delta^{13}\text{C}_{A_i}$ of its neighbors. The MRS model is controlled by the $F_o-A_i$ pairs of the $\text{ELCO}_i$ .
holograph	A graph of distribution representing all the holistic information pertaining to a given set. Designated as $H(\delta, E)$ to describe $E$ environments effects on $\delta^{13}\text{C}_{F_o}$ with the SRS model and $H(\Phi_{F_o}, \Phi_{A_i}, \text{ELCO})$ to describe the behavior of $F_o-A_i$ carbon pairs of $\text{ELCO}$ s with the MRS model.
correlation plane	Plane defined by the two chemical shift axes assigned respectively to the $\delta^{13}\text{C}_{F_o}$ and $\delta^{13}\text{C}_{A_i}$ behavior of the $\text{ELCO}$ s.
spectral map	Area of correlation plane limited by the $\delta^{13}\text{C}_{F_o}$ and $\delta^{13}\text{C}_{A_i}$ behavior of the $\text{ELCO}$ s.
assignment table	On a correlation plane, this determines the zones pertaining to different types of component carbons of $\text{ELCO}_i$ .

the introduction of structural primitives which are more important both for their size and their connectivities.

The distribution shown in Figure 5,  $H(\Phi, \text{ELCO}_b)$ , restores the shape of the first wide band (0–90 ppm) formed by the  $\text{ELCO}_b^\sigma$ . This improvement is due mainly to a greater increase of their number (3983  $\text{ELCO}_b^\sigma$ ) as compared to 2431  $\text{ELCO}_b^\pi$ . The  $\text{ELCO}_b^\sigma$  acquires a representation identical to that of the  $\text{ELCO}_b^\pi$ ; an  $\text{ELCO}_b$  ( $\sigma$  or  $\pi$ ) corresponds on average to 29  $\delta^{13}\text{C}/E$  values.

A considerable gain in selectivity is also obtained, resulting from the reduced range widths of  $\Phi\delta^{13}\text{C}$  (Table II). Some 261  $\text{ELCO}_b$ , or 4.1% of the 6414  $\text{ELCO}_b$  as opposed to 9% of the  $\text{ELCO}_A$ , are associated with each  $\delta^{13}\text{C}$ . This range reduction, combined with the multiplication of the  $\text{ELCO}_b$ , helps to sharpen the description. Thus in the wide band of the  $\text{ELCO}_b^\sigma$ , we again encounter relative heights (i.e., numbers of  $\text{ELCO}_b^\sigma$ ) for the 29- and 71-ppm values in a relationship closer to that

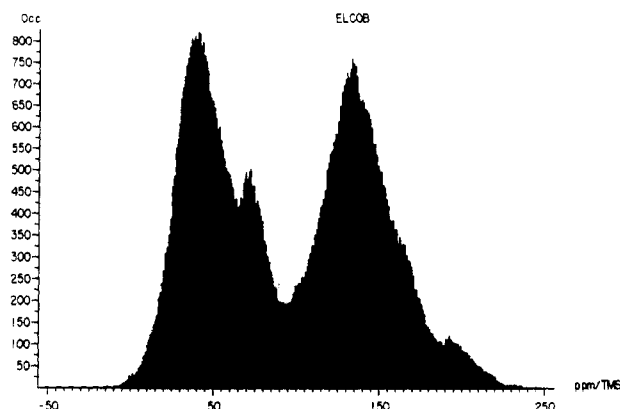


Figure 7. Distribution  $H(\Phi, ELCO_B)$ : (see Figure 5).

of their occurrences, as shown in Figure 2. The chemical shifts corresponding to the distribution peaks at 45 and 135 ppm are also closer to those associated with the greatest occurrences (29 and 128 ppm).

The advantage of the  $ELCO_B$  module, which has been adopted as the standard module of the EPIOS system,<sup>12</sup> is also shown by the results of the spectral simulation systems using substructures of variable depth.<sup>4-6</sup> In these systems, including substructures with four<sup>4,6</sup> or even five<sup>5</sup> shells of atoms the mean shell level of the substructures that are actually used for prediction falls between 1 and 2, which describes only the  $\alpha$  and  $\beta$  effects on the central carbon. In structure elucidation systems, modules of the  $ELCO_B$  type have been adopted by Munk<sup>7</sup> as a basis of the new COCOA system.<sup>21</sup>

**Advantages and Disadvantages of the  $ELCO_B$ .** According to the defined selectivity criteria, the most specific but also most numerous  $ELCO_B$  offer the best precision, as can be seen from Figure 7. Their narrow  $\Phi\delta^{13}C$  range, which can be seen from Table II, explains this improved precision.

However, their specificity complicates their use by an elucidation system both for retrieval procedures and for structure generation. An average of 307  $ELCO_B$  are associated with each  $\delta^{13}C$ , or 2.2% of the 14072  $ELCO_B$ . As a result, structure generation from an  $ELCO_B$  selected for a query spectrum involves a number of combinatoric problems, which can however sometimes be simplified by strategic decisions. Thus in ACCESS,<sup>9,22</sup> where  $ELCO_B$  are used, arbitrarily only the first eight of each list of candidates associated with input chemical shifts are retained for structure generation.

The major disadvantage of the  $ELCO_B$  in structure elucidation is related to the risk that they introduce of silence in spectral interpretation. Such risk of nonselection of a fragment which is in fact in the target structure results first of all from the probable absence of this fragment from the knowledge base. It is also linked to the narrow ranges characterizing the available  $ELCO_B$ . These ranges, defined from a reduced number of  $\delta^{13}C$  reference values, only partially reflect the behavior of focus atoms. The  $\delta^{13}C$ /substructure relations are simplified and subject to fluctuations due to the variability in the conditions under which they were measured. Aside from their own structural specificity then,  $ELCO_B$  are also penalized by the excessively restrictive spectral characterization.<sup>13</sup>

#### SUBSTRUCTURAL LEARNING AND KNOWLEDGE

The significance of silence with respect to the use of  $ELCO_B$  appears more clearly in the curves shown in Figure 8. These illustrate progress in learning in the EPIOS system in terms of the new ELCO substructures entered into its knowledge base. The curves represent the numbers of ELCO of each type, A, b, and B, that are obtained from an increasing number of structures in the databank. Figure 8 provides a graphic display of knowledge growth, formalized by the ELCOs, as a function

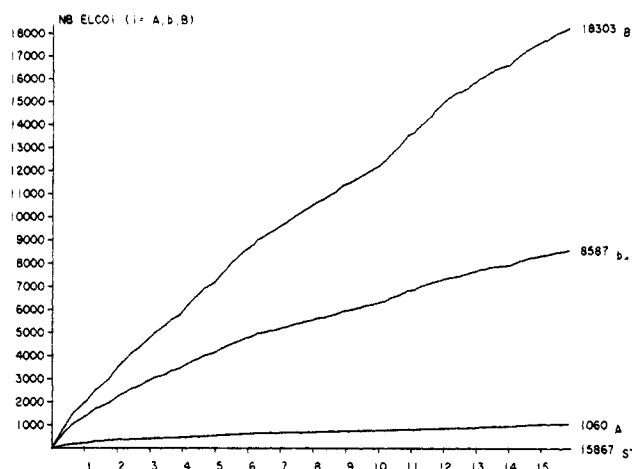


Figure 8. Knowledge growth, in terms of ELCO primitives, according to the reference structural information considered. The acquisition rate of new ELCOs starting from new structures increases with the ELCO specificity.

of the reference information provided—the structures.

The number of  $ELCO_A$  in the knowledge base has already stabilized at 1060 as can be seen from Figure 8. This conveys the ability of  $ELCO_A$  to describe the vast majority of normal chemical compounds. The  $ELCO_b$  curve, on the other hand, shows an increase as more structures are added. The decreasing slope of this curve shows that a situation equivalent to that of the  $ELCO_A$  will soon be reached as the process continues. Currently,  $ELCO_b$  that are new to the databank represent relatively rare structural fragments, and their absence has little effect upon the operation of the EPIOS system.

In contrast, the number of indexed  $ELCO_B$  continues to increase approximately linearly with the number of reference structures; on average, every new structure added to the reference population supplies at least one new  $ELCO_B$ . Such linear growth of knowledge increasingly resembles a *duplication* of the structural information rather than an *enrichment* of the knowledge associated with it. The specificity of the  $ELCO_B$ , insofar as it reduces the probability of their presence in an unknown structure, makes it difficult to use them in a system designed to cover a wide range of application.

#### DISCUSSION

More knowledge is needed to grasp the structural environment concept, with its static and dynamic aspects, in all its complexity. The theoretical approach to ordinary molecules is still too superficial to account for short-range and long-range interactions, linked or not. The various aspects of complex molecular situations can only be fully understood by means of lengthy and painstaking pragmatic research.  $^{13}C$  NMR spectroscopy is just one methodology that is helpful in furthering our structural knowledge.

In this area, the transformation process by which we proceed from experimental data to interpreted data, or information stored in a databank, and thus to modeled information, or knowledge that can be exploited by a system, is essentially empirical. Each step in the process poses choices concerning the nature of the data or the information to be considered. Depending upon the choices that are made, distortions or losses can occur and can limit or severely handicap the applications that are planned as well as the strategies adopted for carrying them out. Accumulation of an increasing quantity of structural and spectral information in a databank does not necessarily lead to a relevant knowledge base which is adapted to solving the various problems posed of it by chemists.

In this paper, we have attempted to deal with these ideas concerning information and knowledge, which are often per-

ceived intuitively and often confused. We have approached the problem of structural and spectral knowledge and its representation by formalization in terms of a classical model, the Single-Resonance Subspectra/Substructure (SRS) representation. The criterion determining this model's precision is substructure specificity. This determines the ability of the captured knowledge to define structure/ $\delta^{13}\text{C}$  relationships. Its effect upon the modeling of this knowledge is analyzed with respect to the different representations obtained from the same reference information derived from the  $^{13}\text{CDP}_{\text{III}}$  databank. An overall physical image of this test population is defined by the profile of a distribution  $H(\delta, E)$  which translates the diversity of the environment  $E$  effect on the  $\delta^{13}\text{C}$  shifts. This image of experimental reality is taken as a reference and compared successively to those produced by different SRS models of the initial information. The distributions  $H(\Phi, \text{ELCO})$  obtained for the increasingly specific topological ELCO modules represent in graphic form the knowledge provided by the SRS model. This comparison of the real structure/ $\delta^{13}\text{C}$  relation with that created from the SRS knowledge reveals the distortions introduced by ELCO primitives and by modeling their  $\delta^{13}\text{C}$  behavior using the interval  $\Phi\delta^{13}\text{C}$  that characterizes it. The study of these distortions and of their reduction by successive models involving different degrees of neglect of information within the ELCO context permit certain compromises in the choice of an optimum reference ELCO. This final choice is in fact often linked to the potential application. The work reported here shows that the ELCO<sub>b</sub> module is a good compromise for the formulation of structural knowledge necessary for computer-aided structure elucidation. However, the associated SRS spectral knowledge remains somewhat too limited to reach the discriminatory power necessary for the selective interpretation of  $^{13}\text{C}$  NMR spectra.

It would be too optimistic to believe that the same substructure could satisfy constraints derived from diverse applications. In compound identification by databank query<sup>23</sup> where the responses are contained in the population consulted, the specificity of the substructures used determines both the precision and the recall of the search. In structural elucidation, where the structure of the unknown compound must be generated from structural fragments, introduction into the substructures of too high a degree of specificity leads to a significant risk of *silence*, i.e., nonselection of one of the component fragments. The basic disadvantage of the SRS model stems directly from this substructural specificity. This leads not only to silence but also to combinatoric problems associated with the large numbers of ELCO. In the singly focused framework of the SRS model, the active physical environment of the central carbon coincides, more or less, with that encoded in the substructure. This one-to-one correlation between the structural environment and its perceived influence on the spectrum allows a very exclusive view of the carbon by isolating it from the molecule of which it is a part. In this SRS model, spectral knowledge is too weak in comparison with the structural knowledge it characterizes, and this imbalance is the principal cause of the model's limitations.

To overcome this difficulty, which is a serious problem in some applications of the SRS model, the basic ELCO<sub>b</sub> module was retained and enhanced with further spectral knowledge.<sup>12,13,24</sup>

The ELCO description used in DARC is particularly advantageous in that it offers representation on different levels of structural knowledge and even provides levels of abstraction for the  $B$  atoms. The need for an extension to the basic structural description seems clear, and our new Multi-Resonance Subspectra/Substructure (MRS) model<sup>24</sup> provides a richer perception of both the St and the Sp components of the active environment of  $^{13}\text{C}$  nuclei. More spectral information

must be introduced to redress the imbalance referred to above. The spectral information must be spread over all the carbon atoms, especially the focus and its immediate neighbors, of the ELCO<sub>b</sub> module. This should harmonize the structural and spectral components of the defined knowledge.

The concepts associated with the SRS model are nonetheless essential for data structuring in NMR databanks and useful for complex strategies which use a combination of models. Moreover, the ELCO<sub>b</sub> fragment that is seen here to be a good compromise for the structural portion of the (St.Sp) pairs needed to drive a computer-aided structure elucidation system is an excellent candidate for further developments such as those used in the MRS model. In the following paper,<sup>24</sup> an investigation of the capability of this new MRS representation in handling more complex knowledge and perceiving environmental effects with increased sensitivity will be described.

## REFERENCES AND NOTES

- (1) Grant, D. M.; Paul, E. G. Carbon-13 Magnetic Resonance. 2. Chemical Shift Data for the Alkanes. *J. Am. Chem. Soc.* **1964**, *86*, 2984-2989.
- (2) Jurs, P. C.; Sutton, G. P.; Ranc, M. L. Carbon-13 NMR Spectral Simulation. *Anal. Chem.* **1989**, *61*, 1115A-1122A.
- (3) Gray, N. A. B. Computer-Assisted Analysis of Carbon-13 NMR Spectral Data. *Prog. Nucl. Magn. Reson. Spectrosc.* **1982**, *15*, 201-248.
- (4) Crandell, C. W.; Gray, N. A. B.; Smith, D. H. Structure Evaluation Using Predicted  $^{13}\text{C}$  Spectra. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 48-57.
- (5) Kalchauer, H.; Robien, W. CSEARCH: A Computer Program for Identification of Organic Compounds and Fully Automated Assignment of Carbon-13 Nuclear Magnetic Resonance Spectra. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 103-108.
- (6) Gray, N. A. B.; Crandell, C. W.; Nourse, J. G.; Smith, D. H.; Dageforde, M. L.; Djerassi, C. Computer-Assisted Structural Interpretation of Carbon-13 Spectral Data. *J. Org. Chem.* **1981**, *46*, 703-715.
- (7) Munk, M. E.; Lind, R. J.; Clay, M. E. Computer-Mediated Reduction of Spectral Properties to Molecular Structures. *Anal. Chim. Acta* **1986**, *184*, 1-19.
- (8) Fujiwara, I.; Okuyama, T.; Yamasaki, T.; Abe, H.; Sasaki, S. Computer-Aided Structure Elucidation of Organic Compounds with the CHEMICS System. *Anal. Chim. Acta* **1981**, *133*, 527-533.
- (9) Bremser, W.; Fachinger, W. Multidimensional Spectroscopy. *Magn. Reson. Chem.* **1985**, *23*, 1056-1071.
- (10) Meyer, L. H.; Saika, A.; Gutowsky, H. S. Electron Distribution in Molecules. 3. The Proton Magnetic Spectra of Simple Organic Groups. *J. Am. Chem. Soc.* **1953**, *75*, 4567-4573.
- (11) Smith, D. H.; Jurs, P. C. Prediction of  $^{13}\text{C}$  Nuclear Magnetic Resonance Chemical Shifts. *J. Am. Chem. Soc.* **1978**, *100*, 3316-3321.
- (12) Dubois, J.-E.; Carabédian, M.; Ancian, B. Élaboration Structurale Automatique par RMN du Carbone-13: Méthode DARC-EPIOS. Description de l'élucidation Progressive par Intersection Ordonnée de Sous-Structures. *C. R. Acad. Sci., Ser. C* **1980**, *290*, 369-372, 383-386.
- (13) Dubois, J.-E.; Carabédian, M.; Dagane, I. Computer-Aided Elucidation of Structures by Carbon-13 Nuclear Magnetic Resonance. The DARC-EPIOS Method: Characterization of Ordered Substructures by Correlating the Chemical Shifts of Their Bonded Carbon Atoms. *Anal. Chim. Acta* **1984**, *158*, 217-233.
- (14) (a) Dubois, J.-E. Principles of the DARC Topological System: Applications Pointing to Structural Influence on Oxidation of Hydrocarbons. Plenary Lecture at the 12th International Conference on Combustion, Poitiers, July, 1968; *Entropie* **1969**, *27*, 1-9. (b) Dubois, J.-E. Système Darc: La Chimie à l'Heure des Ordinateurs. *Entropie* **1968**, *21*, 5-13. (c) Dubois, J.-E. Rethinking Chemistry for Computers. The DARC System: A New Approach to Interactive Computer-Assisted Description and Conception. *CNRS Res.* **1975**, *2*, 32-43. (d) Dubois, J.-E. DARC System in Chemistry. In *Computer Representation and Manipulation of Chemical Information*; John Wiley: New York, 1973; pp 239-262.
- (15) Dubois, J.-E.; Doucet, J. P. Carbon-13 NMR of Aliphatic Alkynes: Topological Analysis of Alkyl Substituent Effects on sp Carbon Chemical Shifts by the DARC PELCO Method. *Org. Magn. Reson.* **1978**, *11*, 87-97.
- (16) (a) Dubois, J.-E.; Carabédian, M. Modeling of the Alkyl Environment Effects on the Carbon-13 Chemical Shift. *Org. Magn. Reson.* **1980**, *14*, 264-271. (b) Doucet, J. P.; Panaye, A.; Dubois, J.-E. Topological Correlations of Carbon-13 Chemical Shifts by Perturbation on a Focus: DARC-PULFO Method. Attenuation and Inversion of  $\alpha$ -Methyl Substituent Effects. *J. Org. Chem.* **1983**, *48*, 3174-3182.
- (17) (a) Dubois, J.-E.; Bonnet, J. C. The DARC Pluridata System: The  $^{13}\text{C}$ -N.M.R. Data Bank. *Anal. Chim. Acta* **1979**, *112*, 245-252. (b) Dubois, J.-E. Actualité et Perspective de l'Information Chimique (Editorial). *Entropie* **1977**, *78*, 2-3. (c) Nouvelles de Pluridata:



- Banques de Données du Système DARC-C13/NMR-Crystallography-Mass Spectra-Graphic Input and Output. *Entropie* 1977, 78, 53-55.
- (18) Bremser, W. Expectation Ranges of  $^{13}\text{C}$  NMR Chemical Shifts. *Magn. Reson. Chem.* 1985, 23, 271-275.
- (19) Carabédian, M.; Dagane, I.; Dubois, J.-E. Elucidation by Progressive Intersection of Ordered Substructures from Carbon-13 Nuclear Magnetic Resonance. *Anal. Chem.* 1988, 60, 2186-2192.
- (20) Small, G. W.; McIntyre, M. K. Structure Elucidation Methodology for Disaccharides Based on Carbon-13 Nuclear Magnetic Resonance Spectrum Simulation. *Anal. Chem.* 1989, 61, 666-674.
- (21) Christie, B. D.; Munk, M. E. Structure Generation by Reduction: A New Strategy for Computer-Assisted Structure Elucidation. *J. Chem. Inf. Comput. Sci.* 1988, 28, 87-93.
- (22) Bremser, W. Structure Elucidation and Artificial Intelligence. *Angew. Chem., Int. Ed. Engl.* 1988, 27, 247-260.
- (23) Attias, R.; Dubois, J.-E. Substructure Systems: Concepts and Classifications. *J. Chem. Inf. Comput. Sci.* 1990, 30, 2-7.
- (24) Carabédian, M.; Dubois, J.-E. A Combined Model of Multi-Resonance Subspectra/Substructure and DARC Topological Structure Representation. *J. Chem. Inf. Comput. Sci.* 1991, 31, 564-574.

## A Combined Model of Multi-Resonance Subspectra/Substructure and DARC Topological Structure Representation. Local and Global Knowledge in the $^{13}\text{C}$ NMR DARC Database

MICHEL CARABÉDIAN and JACQUES-ÉMILE DUBOIS\*

Institut de Topologie et de Dynamique des Systèmes de l'Université Paris 7, associé au CNRS, URA 34,  
1 rue Guy de la Brosse, 75005 Paris, France

Received March 27, 1991

The structural and spectral information in a  $^{13}\text{C}$  NMR database can be represented by means of a model which relates substructural fragments to subspectral features for multiple resonances. The substructural part of this model contains a concise DARC description of the structural part with a partially generic  $\text{ELCO}_b$  which is associated with all the spectral information pertaining to the focal atom ( $F_0$ ) and its neighboring carbons ( $A_i$ ). In the spectral information, the concentric environmental view is shifted from the focal atom to the neighbor positions. This leads to overlap in the views and redundancy in the information and a dissymmetrical physical perception which formally, is broader than the substructural view. New substructural/subspectral local and global knowledge functions of this model are managed with holographic techniques. Formalized local and global knowledge is described statistically by juxtaposition of the  $\delta^{13}\text{C}_{F_0} \times \delta^{13}\text{C}_{A_i}$  correlation plane supporting the 3D occurrence distributions. Use of the inferential ability of these planes is facilitated by a table which correlates the repartitioning of the  $\sigma$ - and  $\pi$ -bonds in  $F_0$ - $A_i$  atom pairs.

### INTRODUCTION

In most structure elucidation systems,  $^{13}\text{C}$  NMR data are assigned a classical one-to-one relationship with specific carbon atoms in the structure. With this single resonance subspectra/substructure model, the central carbon of each substructural fragment is considered exclusively, and its  $\delta^{13}\text{C}$  signal, usually described as falling within a band in the overall scale of  $^{13}\text{C}$  chemical shifts, is the only datum that is used to identify the substructure in question.

A more penetrating analysis of the relationships between structural environments and spectral properties of such atoms is possible.<sup>1</sup> This approach uses substructural fragments defined by DARC, in which a structure is treated as a focal atom ( $F_0$ ) and its environment. The environment is organized so as to be limited, concentric, and ordered, hence the acronym ELCO. Use of these substructural fragments has shown that a single resonance-single substructure model cannot account for the complex relationships between structure and spectra.<sup>2</sup>

The multi-resonance spectra/substructure model that is described here has advantages in the analysis of these relationships. A number of different ELCOs have been studied<sup>2</sup> and of these, the specific ELCO that seems to be most appropriate here is the  $\text{ELCO}_b$ . This is the classical atom-centered fragment which consists of a central carbon (the focus atom,  $C_F$ ), its immediate neighbors, and the bonds attaching it to its immediate neighbors. The next neighbors, atoms  $\beta$  to the focus atom, are not explicitly defined, and this imprecision affords the  $\text{ELCO}_b$  with a generic quality which permits the grouping together of numerous diverse environments for

the focus atom. The data which define these environments are all derived from the  $^{13}\text{C}$  NMR DARC PLURIDATA III ( $^{13}\text{CDP}_{III}$ ) Database.<sup>3</sup> This database contains 15 867 structures and allows the generation of environments having statistically significant populations.

**Extended Perception of Environment.** The shielding of a  $^{13}\text{C}$  atom is determined in a complex manner by different factors. Localized site contributions are considered to be approximately additive<sup>4</sup> and any atoms in positions  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  relative to the central or focus atom may affect the chemical shift. The  $\delta$  position corresponds to the D shell in the concentric description used by DARC, but the most significant contributions usually come from the  $\gamma$  positions—the DARC C shell. There are few correlation functions which permit the interpretation of experimental spectra in terms of these interactions, and so purely empirical models must be used.

In the classical single-resonance model, the chemical shift of the focal carbon is used in isolation but in the multi-resonance model, the focal carbon chemical shift is considered together with the chemical shifts of as many as four carbons directly bonded to it. If these neighboring carbons are in turn regarded as focus atoms, then a much deeper perspective on the actual external environment of the  $\text{ELCO}_b$  is revealed. The local spectral information is enriched by the data redundancy in the  $\text{ELCO}_b$  and by the broader perception of its external environment. This multi-resonance model leads to more complex statistics than those proposed for use with the single-resonance model. The  $H(E, \delta)$  holographic curves used in