Large Virtual Enhancement of a ^{13}C NMR Database. A Structural Crowning Extrapolation Method Enabling Spectral Data Transfer

Michel Carabedian and Jacques-Emile 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 April 10, 1997[⊗]

A method for creating virtual spectral information by ordered combination of generic DARC primitives is proposed. Structural crowns are generated and define ∼300 million potential environments around the 8587 available primitives. The associated spectral data are induced by controlled transfer from spectral maps characterizing the primitives forming the crown. This virtual enhancement of knowledge fills the gaps inherent to an experimental data bank by defining the potential influence of environments absent from this bank on their core primitive.

INTRODUCTION

It is becoming increasingly obvious that knowledge extracted from data bases is the true key to the performance of systems that use them. This idea constitutes an important step in the work carried out in the field of structure elucidation by ^{13}C NMR. Knowledge for a long time occupied a subsidiary role in strategies that put the emphasis on the exhaustive enumeration of candidate structures created by means of isomer generators. Its use, restricted to filtering and validating candidate stuctures, proved finally to be inadequate for overcoming the multiple problems intrinsic to the exploration of overextended research spaces.[1−3]

In contrast to these systems conceived around structure generators, EPIOS (elucidation by progressive intersection of ordered substructures) was developed on the basis of an original modeling of structure/δ^{13}C relationships.[4] The priority given right from the beginning to knowledge was expressed by the initial definition of the multi-resonance/substructure model (MRS).[5] This model characterizes the generic primitives, $ELCO_b$ (environments that are limited, concentric, and ordered), consisting of a focus Fo and its neighbors A*i*, by the behavior of all their carbons. The primitives and their associated chemical shifts are extracted from our ^{13}C NMR DARC-PLURIDATA data bank (16 000 structures; 195 000 δ^{13}C). In the course of an elucidation, these primitives are deduced from the interpretation of a spectrum, then progressively assembled to generate the candidate structures. The joint use of the two components, structural (St) and spectral (Sp), of the knowledge completely controls the search for valid solutions to the problem.

However, because of the diversity of chemical structures, knowledge extracted from experimental data can prove to be inadequate.[6] This inadequacy leads in some cases to "silence" of the system (i.e., failure to provide an answer) when certain primitives of a target structure are unknown to the system. The generic character of the $ELCO_b$ primitives minimizes this risk by assuring that they occur in very varied structures, which potentially are targets of the system. The complete set of these structures, formed exclusively of the available $ELCO_b$, determines the scope of the system.

The second cause of silence is the more common. Not infrequently, the behavior of the $ELCO_b$ primitives based on the δ^{13}C values assigned to the reference structures is too restricted. Certain δ^{13}C values associated with structural contexts absent from the data bank, and likely to be associated with potential target structures, may be overlooked by this characterization. These gaps in the spectral component of knowledge consequently reduce the scope of the system, to what is provided by the available experimental data.

In this article we propose a method for increasing the amount of spectral information and the knowledge extracted from our ^{13}C NMR bank without modifying the initial content. This method is based on the St and Sp properties of the MRS model. The first step consists of generating around the available $ELCO_b$ the complete set of their virtual environments in the potential structures. Because there are too many of these potential structures for complete enumeration, we represent them generically by the $ELCO_c$ environments (Figure 1a).

Each of these environments defines the nature of a particular set of atoms B located β from the focus of the $ELCO_b$ and their bonds. The set of the possible $ELCO_c$ represents in this way the set of the possible structural contexts of the central $ELCO_b$ in an external generic crown. To generate the $ELCO_c$ that correspond to an elementary expansion of the $ELCO_b$, we use the EPIOS principle of progressive generation by overlap (Figure 1b). According to this principle, an $ELCO_c$ results from the partial overlap of a central $ELCO_b$ consisting of N pairs of atoms (Fo, A*i*) by a combination of N neighbor $ELCO_b$ having one of these atom pairs in common. In this representation, the focus itself becomes a common neighbor (denoted $A^{n'}$ in Figure 1b) of these N neighbor $ELCO_b$.

The second step of the method consists of simulating the spectral perturbations of the $ELCO_b$ carbons induced by the structural variations of the external crown. This step corresponds to the application at the spectral level of the structural overlap of the $ELCO_b$. This overlap is the

[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1997.

S0095-2338(97)00024-3 CCC: $15.00
Published on Web 01/09/1998

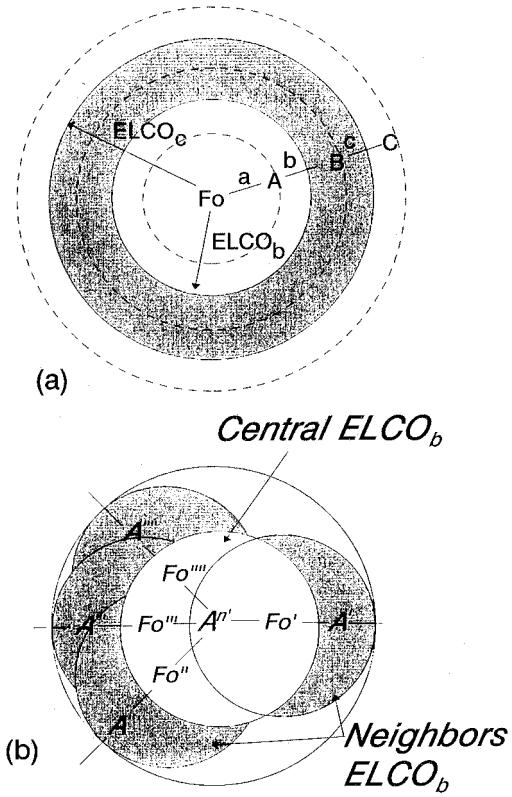


Figure 1. (a) Definition of the $ELCO_b$ primitives and the $ELCO_c$ environments. The external crown, shown in grey, includes the atoms and the bonds of the B layer of the focus Fo. (b) The external crown is defined by overlap of a central $ELCO_b$ with *N* of its neighbor $ELCO_b$. Fo becomes the common neighbor $A^{n'}$ of these neighbor $ELCO_b$.

necessary condition for transferring to the central $ELCO_b$ the $\delta^{13}C$ values assigned to the neighbor $ELCO_b$. The redundancy of the focus in the preceding definition of the $ELCO_c$ is used to ensure the reliability of this transfer by filtering the relevant $\delta^{13}C$ values supplied separately by each of the neighbor $ELCO_b$.

We shall describe here the principles and the practical application of this pragmatic method for enhancing spectral knowledge.

DESCRIPTION OF THE VIRTUAL CROWN METHOD OF EXTRAPOLATION

Definition of Spectral Maps by the Multiresonance/ Substructure Model. The interest of the MRS model is that it expresses the partial overlap of the primitives in the structures from which they are extracted. An overall correspondence is established between the description of the $ELCO_b$ primitives and the representation of their $\delta^{13}C$ behavior. All the carbons of these primitives, the focus Fo and its neighbors A*i*, are characterized by spectral maps (δFo, δA*i*). These maps are used to interpret the spectrum of an unknown compound. In their experimental version they describe only the effect of the environments really associated with the $ELCO_b$ primitives in the reference structures. To extend their scope beyond these experimental limits, we enhance them with new data expressing the behavior induced by all the virtual $ELCO_c$ environments of these primitives.

This production of complementary information, structural and spectral, is carried out in three phases consisting of (Figure 2): (i) enlarging the structural space St by constructing all the extended virtual $ELCO_c$ environments centered on the $ELCO_b$ primitives; (ii) completing the spectral maps by transferring to the central $ELCO_b$ the $\delta^{13}C$ values of the neighbor $ELCO_b$ forming these $ELCO_c$; and (iii) validating the values transferred, by determining the limits of the maps corresponding to the intersection of the spectral maps associated with the neighbor $ELCO_b$.

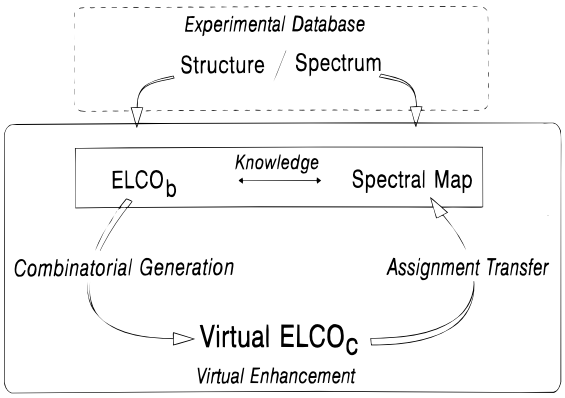


Figure 2. Principle of the method.

In what follows, these differents steps are presented formally and then with the help of practical examples.

Generation of $ELCO_c$ Environments by $ELCO_b$ Overlap. The object of the generation of $ELCO_c$ environments is to create around the available $ELCO_b$ primitives new structural contexts, more numerous and more varied than those present in the reference structure population. These $ELCO_c$ environments consist of two concentric layers of atoms and their bonds, and they cover the most significant part of the active environment of a ^{13}C nucleus. It is possible to generate them formally by combining their atoms and bonds. Such a treatment was considered by Munk[7] for the atom-centered fragment (ACF) primitives of the $ELCO_b$ type. However, in the absence of a general theoretical model of the structure/$\delta^{13}C$ relationships, it is not possible to associate the generated ACF with an exhaustive characterization of their $\delta^{13}C$ behavior. This has to be defined in classical fashion from the ^{13}C NMR bank of the CIS (7300 structures), where only 70% of them have an occurrence sufficient to characterize them. Besides the absence of chemical shift for 30% of the primitives, the characterization obtained remains purely experimental and inherits therefore the limitiations of the reference population.

To avoid this distortion between the structural and spectral components, the $ELCO_c$ and the perturbations that they induce are defined from the available $ELCO_b$ primitives and their behavior. This method restores to the $ELCO_b$ primitives the role they play in the course of progressive structure generation by EPIOS. In this process, an $ELCO_c$ is an intermediate environment on the way to a possible target structure and results from the elementary expansion of its central $ELCO_b$ taken as the root of this structure. Figure 3 illustrates this construction of an $ELCO_c$ by partial overlap of a central $ELCO_b$ with two of its neighbor $ELCO_b$. The overlap of the two $ELCO_b$ is determined by the presence of a same pair of bonded atoms in both of them. The pairs, (Fo, Al) and (Fo, A2), forming the central $ELCO_b$ in this example appear in their transposed forms, (A′, Fo′) and (A″, Fo″), respectively, in the neighbor $ELCO_b$ (Figure 3). It should be noted that the idea of "transposed (Fo, Ai) pair"

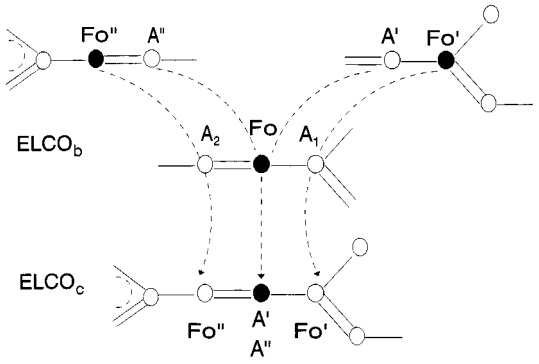


Figure 3. Example of the construction of an ELCO$_c$ by overlapping a central ELCO$_b$ with two of its neighbor ELCO$_b$.

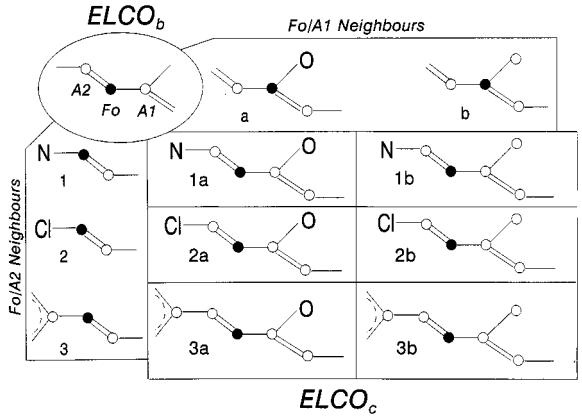


Figure 4. Different ELCO$_c$ obtained by combining the neighbors associated with the two pairs, (Fo,Al) and (Fo,A2), of the central ELCO$_b$.

makes it possible to tackle simply that of the partial overlap of two ELCO$_b$ resulting from the shift of their focus relative to their common bond.

The set of possible ELCO$_c$ around a central ELCO$_b$ is generated by combining the neighbors associated with its constituent pairs. We show in Figure 4 six ELCO$_c$ constructed around the ELCO$_b$ from two neighbors associated with its (Fo, Al) pair and three neighbors associated with its (Fo, A2) pair. Two types of ELCO$_c$ are obtained by this combination of neighbors. The first are the experimental ELCO$_c$ featuring in the reference structures and whose effect is described by the available δ^{13}C values. The second type corresponds to virtual ELCO$_c$ that exist in no reference structure. These virtual ELCO$_c$ can result from the combination around a central ELCO$_b$: (i) of a virtual neighbor ELCO$_b$ with which a connection is formally possible but of which no example features in the reference structures; (ii) of an experimental neighbor ELCO$_b$ coexisting with the central ELCO$_b$ but in different reference structures; and (iii) of a virtual neighbor ELCO$_b$ and an experimental neighbor ELCO$_b$. In the method proposed, these different types of experimental and virtual ELCO$_c$ are not differentiated and are treated in the same way.

Moreover, it should be noted that the number of these ELCO$_c$ that can be generated around an ELCO$_b$ depends on the number of different pairs, on the number of identical pairs, and on the number of neighbor ELCO$_b$ associated with these pairs. The different situations encountered are shown in Figure 5. The ELCO$_b$ considered in this example are formed of two different pairs, (Fo, Al) and (Fo, A2), so the

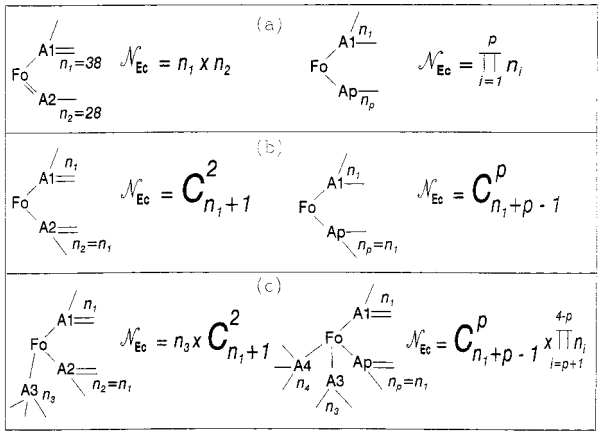


Figure 5. Calculation of the number *Nec* of ELCO$_c$ depending on the number n_i of neighbor ELCO$_b$ associated with the *N* (Fo,A*i)* pairs of a central ELCO$_b$: (a) al1 pairs different; (b) all pairs identical; (c) *p* of the *N* pairs are identical.

number of possible ELCO$_c$ is equal to the product of the numbers of neighbor ELCO$_b$ associated web these pairs (here $38 \times 68 = 1064$). Of the 1064 possible ELCO$_c$, only 71 exist in our reference population.

For an ELCO$_b$ with *p* identical pairs, the *n* neighbors associated with each of these pairs are also identical. In this case the number *Nc* of possible ELCO$_c$ corresponds to the number of combinations of *p* elements chosen from the *n* neighbors: a same neighbor can be taken several times, but no permutations are allowed among the *p* neighbors (Figure 5). For the ELCO$_b$ with both identical and non-identical pairs, the number of possible ELCO$_c$ is the product of the two previous numbers.

With this method very many new ELCO$_c$ environments are generated. Their number expresses the contribution of potential overlaps established between the ELCO$_b$. For the 8587 ELCO$_b$ primitives, 740 753 overlap possibilities have been defined between these primitives, whereas <10% (61 449) are really observed in the 16 000 reference structures.[8] These 10% of observed overlaps correspond to 49 193 experimental ELCO$_c$ environments (i.e., appearing the reference structures). With the complete set of overlaps, virtuals and experimental, 295 674 848 different ELCO$_c$ environments can be constructed. These virtual environments do not all correspond to chemically stable entities. Nevertheless, for the system, which ignores the idea of chemical stability, they represent environments that can be considered during an elucidation.

The characterization of the effect of the new ELCO$_c$ environments generated is the necessary condition for the selection of their central ELCO$_b$ in the interpretation of the spectrum of a target structure possibly containing one of these ELCO$_c$. It is obtained by transferring the pairs of δ^{13}C values assigned to neighbor ELCO$_b$ according to the method described later.

Principle of Controlled Transfer of δ^{13}C Assignments. The extension of the spectral maps aims at completing them by characterizing the new perturbations induced by the set of ELCO$_c$ generated, and extending their field of application to the set of possible target structures containing these ELCO$_c$. To describe the influence of these virtual ELCO$_c$, which are not part of the reference population, the spectral information used corresponds to pairs of δ^{13}C values assigned

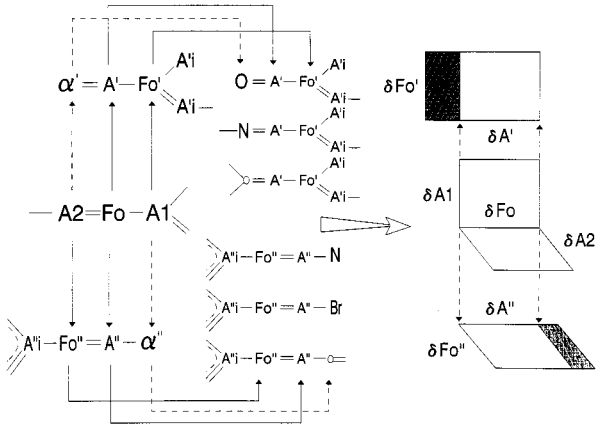


Figure 6. Transfer of dc δ^{13}C values. The maps, (δFo,δAl) and (δFo,δA2), of the central $ELCO_b$ are defined from the maps, (δFo′, δA′) and (δFo″,δA″), associated with its neighbors. The grey areas correspond to values δA′ and δA″ that cannot be attributed to the focus of the central $ELCO_b$.

in a transposed form to the neighbor $ELCO_b$. The behavior of a central $ELCO_b$ is thus extrapolated from that of its atom pairs observed in the neighbor $ELCO_b$. It is characteristic of the perturbations induced by the possible connection of these neighbors. Figure 6 shows schematically the transfer of the chemical shifts of the neighbor $ELCO_b$ to the central $ELCO_b$ considered in the previous example (Figure 2). Each of the atom pairs (Fo, A*i*) of the central $ELCO_b$ features the following in its transposed form in the neighbor $ELCO_b$: (i) the focus corresponds to one of the A*i* positions (here A′ and A″) of the neighbor $ELCO_b$; (ii) each atom A*i* corresponds to the focus (here Fo′ and Fo″) of one of the neighbor $ELCO_b$; and (iii) the other atoms $A_{j \neq i}$ correspond to fuzzy positions (here α′ and α″) of these neighbor $ELCO_b$.

Transposition of the roles played by the atoms Fo and A leads to a modification of their explicit environment. For the atoms A*i*, which become focuses (Fo′ and Fo″) of a neighbor $ELCO_b$, their fuzzy α positions in the central $ELCO_b$ are specified by new atoms (A′*i*, A″*i*) and their bonds (Figure 6). The behavior induced by these new atoms on the Fo′ and Fo″ positions is *a priori* compatible with what could be accepted by the A*i* positions of the central $ELCO_b$. The situation is different for the focus Fo. When it occupies the positions A′ and A″ in the neighbor $ELCO_b$, $N - 1$ of its neighbors $A_{j \neq i}$ in the $ELCO_b$ are replaced by $N - 1$ fuzzy α positions: its *N*th neighbor A*i* occupies the position of the focus in the neighbor $ELCO_b$. In Figure 6 a fuzzy α′ position (or α″) replaces atom Al (or A2) in each of the neighbor $ELCO_b$. This shift of the fuzzy positions around a pair (Fo, A*i*) of the central $ELCO_b$ implies very varied new environments for its focus Fo (A′ and A″). Consequently, the transfer of values (δA′ and δA″) observed in these environments towards the focus of the central $ELCO_b$ is not unrestricted. It assumes that the perturbations induced on the positions A′ and A″ are compatible with those that could really be exerted on Fo in the central $ELCO_b$. Because there is no one-to-one relationship between the environment and the behavior that it causes, it is not possible to impose *a priori* the restrictions necessary for this transfer. A given environment can be associated with different behaviors and *vice versa.* It is evident that this ambiguity in the structure/δ^{13}C relationships also affects the characterization of the $ELCO_b$ by the δ^{13}C values of their carbon pairs. It is

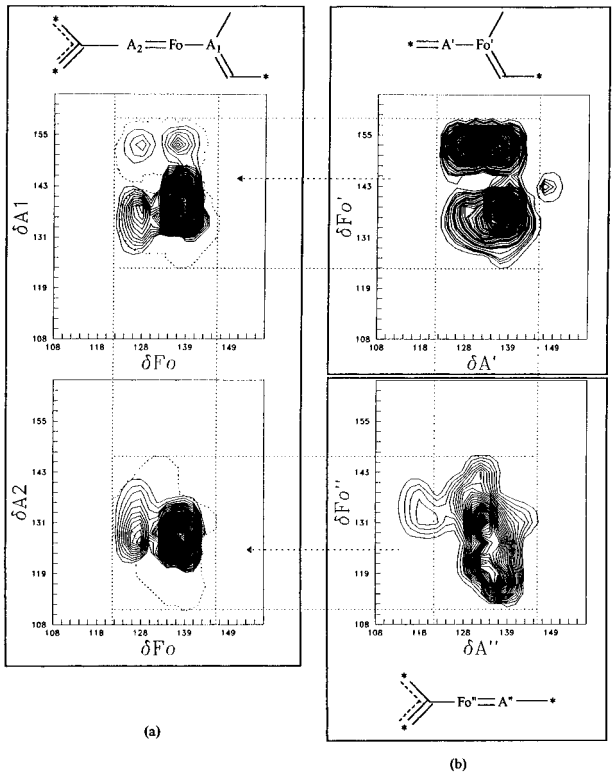


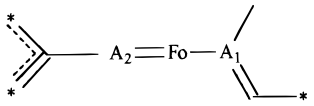
Figure 7. Effect of a virtual $ELCO_c$. (a) The maps (δFo,δAl) and (δFo,δA2) include the experimental values available for the $ELCO_c$. Their contour (dotted) corresponds to that of the maps (δFo′,δA′) and (δFo″, δA″) established for the neighbor $ELCO_b$. (b) These maps are based on a population from which the previous experimental values, (δFo,δAl) and (δFo, δA2), are excluded.

nevertheless reduced by the MRS model, which by definition is based on the multidimensional description of their behavior. Any δ^{13}C value assigned to the focus of an $ELCO_b$ is part of *N* pairs (δFo, δA*i*) characteristic of its *N* carbon pairs. This fundamental property of the MRS model is applied here to transposed pairs of δ^{13}C values provided by its neighbor $ELCO_b$. Any value δA′ coming from a transposed pair and to be attributed to the focus Fo of the $ELCO_b$ must also appear among the values (δA″) provided by the other transposed pairs. In other words, the coherent values of δFo are those belonging to the intersection of the *N* sets of values (δA′, δA″) assigned to its *N* transposed pairs in its neighbor $ELCO_b$.

In the case of an experimental $ELCO_c$, the central $ELCO_b$ is a neighbor common to its neighbors. Its own δFo values are therefore found in the set of those assigned (δA′ and δA″) to its neighbors. On the other hand, for a virtual $ELCO_c$, by definition absent from the reference population, the environments of the neighbor $ELCO_b$ can be incompatible with those which can exist around the central $ELCO_b$. This is the case in Figure 6 where the O and N atoms occupying the α′ and α″ positions are incompatible with the Al and A2 carbons of the central $ELCO_b$. This incompatibility of the environments is revealed by the incoherence of the behavior (δA′ and δA″) to which must correspond at the same time the focus Fo of the central $ELCO_b$. The practical example given in Figure 7 shows that the intersection of the multiple behaviors (here δA′ and δA″) transferred to the focus Fo proves to be sufficiently selective to eliminate the incompatible environments.

VIRTUAL EXTENSION OF SPECTRAL MAPS

Perturbation of an $ELCO_b$ by a Virtual $ELCO_c$. To illustrate the ability of the method to simulate the effect of a virtual environment, the experimental and virtual spectral maps of the following $ELCO_c$ are set up and then compared.



In the first step, the spectral map describing the effect of this "real" $ELCO_c$ on its central $ELCO_b$ is constructed from the available experimental data (Figure 7a). In the second step, the structures containing this $ELCO_c$ (and their spectra) are eliminated from our reference population. Because it is no longer part of this population, this $ELCO_c$ becomes a virtual environment. To characterize its influence on its central $ELCO_b$, we consider the neighbor $ELCO_b$ and the data provided by their spectral maps.

The spectral maps characteristic of the pairs (A′, Fo′) and (A″, Fo″) of the neighbor $ELCO_b$ are defined by 63 pairs (δA′, δFo′) and 39 pairs (δA″, δFo″), respectively, arising from structures not containing the $ELCO_c$ considered (Figure 7). Each of the pairs, (Fo, A1) and (Fo, A2), of the central $ELCO_b$ features individually in the structures containing these neighbor $ELCO_b$, where they are affected by the various environments responsible for the behaviors (δA′, δFo′) and (δA″, δFo″), respectively. The behavior of the central $ELCO_b$, formed by these two pairs, is reconstructed from their behavior in the neighbor $ELCO_b$. By this transfer of the values (δA′, δFo′) and (δA″, δFo″), the two sets of values δA′ and δA″ are attributed to the focus Fo. Its neighbors, A1 and A2, correspond to the values δFo′ and δFo″, respectively. The double definition of the behavior, δFo, of the focus is used to validate the overall behavior attributed at the central $ELCO_b$. According to the MRS model, all the δFo values attributed to Fo are simultaneously involved in the N pairs (δFo, δAi) characteristic of a $ELCO_b$. The set of values obeying this condition corresponds to the intersection of the sets of δA′ and δA″ values. Conversely, it is impossible to find attributions for A1 and A2 that are mutually compatible for all δFo values (δA′, δA″) taken outside this intersection (Figure 7).

This transfer of the δ^{13}C values of the neighbor $ELCO_b$ towards the central $ELCO_b$ defines the behavior of the latter on the basis of those observed for its pairs considered separately in more generic contexts. The set of values selected by the intersection of the behaviors of the pairs corresponds, therefore, to an extended approximation of the specific behavior of the central $ELCO_b$. The only condition for validating these values is that they belong to the intersection, which outlines an enlarged area necessarily containing the specific behavior of the $ELCO_b$ as it would be described by actual experimental data. This restriction on the behavior of a $ELCO_b$ does not only correspond to the logic of the MRS model. We show in the following analysis that it results also from its ability to describe the reality of the effects of environments on the δ^{13}C behaviors.

Shown in Figure 8 are the structures eliminated to define the behavior δFo of the focus of the $ELCO_b$ considered in this example. These structures correspond to the limiting

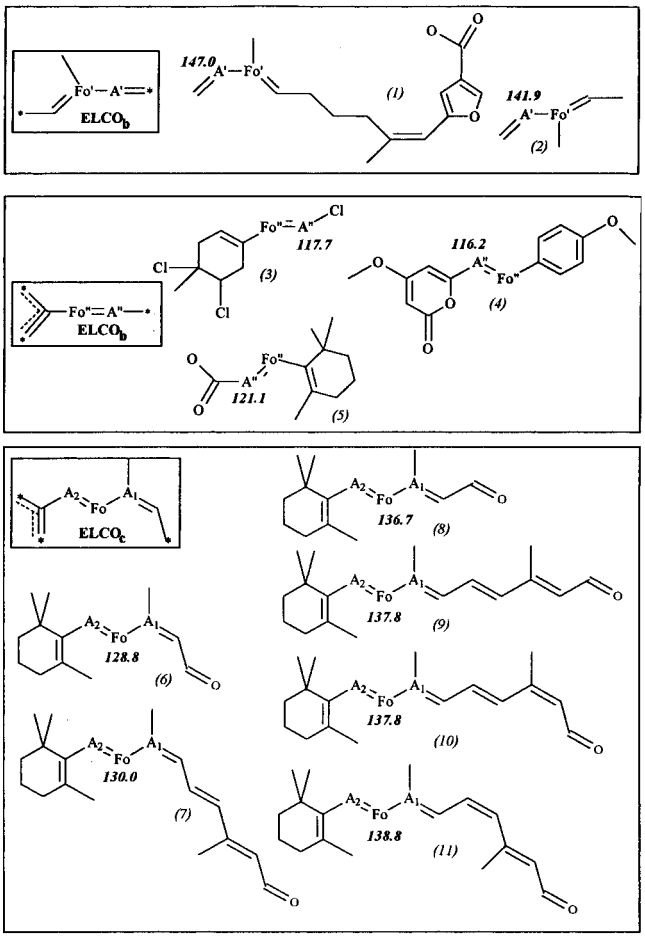


Figure 8. Structures *(1)-(5)* correspond to values δA′ and δA″ eliminated by intersection of the maps of the neighbor $ELCO_b$ in Figure 7. Structures *(6)* and *(7)* and *(8)*−*(11)* correspond to *cis* and *trans* isomers responsible for the peaks on the (δFo, δA2) map of Figure 7.

behaviors δA′ ≥ 141.7 and δA″ ≤ 125.8 ppm observed for each of the neighbor $ELCO_b$. The values δA′ = 141.7 and 147 ppm are introduced by structures *(1)* and *(2)*, which lie outside the range of variation of δA″ and cannot be transferred to δFo because they correspond to no δFo″ value (i.e., δA2). In structures *(1)* and *(2)*, the A′ carbon (i.e., Fo) is bonded to an ethylenic carbon, which is incompatible with the A2 carbon of the central $ELCO_b$. In the same way, in structure *(3)*, where δA″ ≤ 125.8 ppm, A″ (i.e., Fo) is bonded to a Cl atom that is incompatible with the Al carbon of the central $ELCO_b$. In structures *(4)* and *(5)*, A″ is bonded to environments that introduce β oxygen atoms, which are absent from the $ELCO_c$ considered. The influence of this virtual $ELCO_c$ on its central $ELCO_b$ is finally described by the maps of the neighbor $ELCO_b$, (δA′, δFo′) and (δA″, δFo″), in their common range, δFo = δA′ ∩ δA″.

Comparison of Virtual and Experimental Behavior. It should be remembered that the maps of the neighbor $ELCO_b$ represented in Figure 7 are based on a reference population from which we have excluded the structures (and their spectra) containing the $ELCO_c$ considered in this example. On the contrary, the experimental maps (δFo, δA1) and (δFo, δA2), describe the behavior observed for this $ELCO_c$ in the structures previously eliminated. Generally, the maps (δA′, δFo′) and (δA″, δFo″) cover the experimental maps (δFo, δA1) and (δFo, δA2) (Figure 7). The behavior described

by a small number of experimental values on the (δFo, δA1) map is confirmed by the (δA′, δFo′) behaviors, as is the case for the area limited by δA1 ≥ 143 ppm for which only three values are available. On the map (δA′, δFo′), this zone is described by 34 values grouped homogeneously around the experimental behaviors. The profile suggested by the experimental map is thus confirmed by the distribution of the virtual values that also respect the position of its summits.

The virtual map (δA″, δFo″) corresponds more approximately to the experimental map (δFo, δA2). The first largely covers the second but the distribution of the behaviors described is different. Note that on the experimental map the distribution of available (δFo, δA2) values is concentrated around a single summit at δFo ≅ 139 ppm. This distribution is not the case for the virtual map (δA″, δFo″). The latter shows a more complex distribution organized around three summits in the δA″ = 132−137 ppm range. This delocalization of the experimental summit can be explained by the composition of the reference populations used.

In the population containing the $ELCO_c$, most of the structures [*(8)*−*(11)*, Figure 8] are conjugated polyenes where the focus of the $ELCO_c$ is β to a *trans* bond. The values for its behavior (136 ≤ δFo ≤ 139 ppm) determine the peak on the experimental map (δFo, δA2). On the other hand, the less numerous *cis* isomers [*(6)* and *(7)*, Figure 8] correspond to upfield values: δFo ≤ 130 ppm.[9,10] This dichotomy of the reference population (*cis/trans*) does not exist in that used to characterize the neighbor $ELCO_b$ because the previous structures were excluded. One sees, however, that the pair (A″, Fo″) appears in both more numerous and varied environments, inducing less specific δA″ behaviors. The treatment described previously for a particular $ELCO_b$ is now extended to all the possible $ELCO_c$ around a given central $ELCO_b$.

Perturbation of a Central $ELCO_b$ by the Set of its Virtual $ELCO_c$. Figure 9 shows the enhancement of the experimental spectral maps obtained for the $ELCO_b$ considered. This $ELCO_b$ feautures in 264 structures of the reference population. From their spectra, 350 pairs of values, (δFo, δA1) and (δFo, δA2), are formed to set up the experimental spectral maps shown in Figure 9. The behavior described by these maps corresponds to the effect exerted by 71 experimental $ELCO_c$ environments associated with this $ELCO_b$ in the 264 reference structures. Potentially, 1064 $ELCO_c$ environments can be generated by combining the neighbor $ELCO_b$ containing each of the (Fo, Al) and (Fo, A2) pairs. A total of 502 (δFo′, δA′) pairs and 1782 (δFo″, δA″) pairs attributed to these neighbor $ELCO_b$ are used to set up the virtual maps (Figure 9). In this way, one now describes not only the experimental $ELCO_c$ environments but also those of the set of potential $ELCO_c$. The values δA′ and δA″ corresponding to δFo cover the ranges 80−185 and 90−168 ppm, respectively. The final range selected for δFo by intersection of these ranges is, therefore, 90−168 ppm.

The structures responsible for the limiting behaviors excluded from this range (δA′ ≥ 168 ppm and δA′ ≤ 90 ppm) are shown in Figure 10. In structures *(24)* and *(25)*, the atom A′ (i.e., Fo) is bonded to an oxygen, which is incompatible with the ethylenic carbon A2 of the central $ELCO_b$. The behavior of A′ is that of a carbonyl (δA′ =

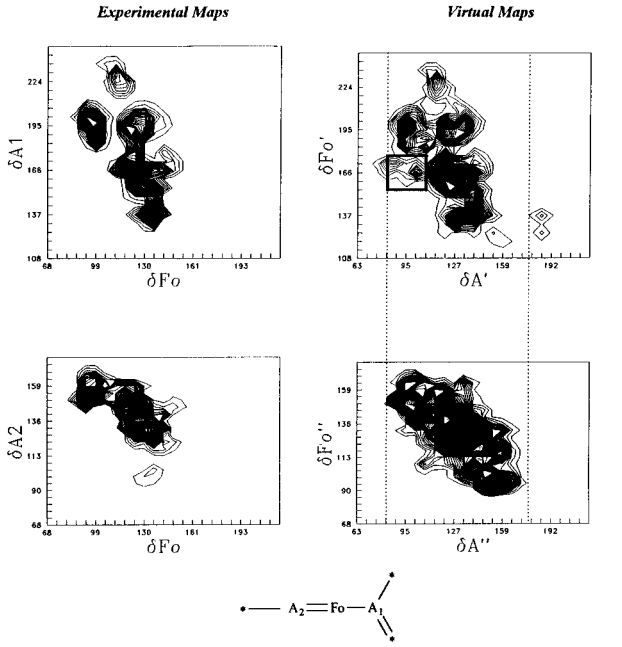


Figure 9. Effect of the $ELCO_c$ set on a $ELCO_b$. The experimental maps include 350 pairs of (δFo, δA1) and (δFo, δA2) values associated with 264 reference structures. For the virtual maps, 502 (δFo, δA1) pairs and 1782 (δFo, δA2) pairs are associated with the set of 1064 possible $ELCO_c$.

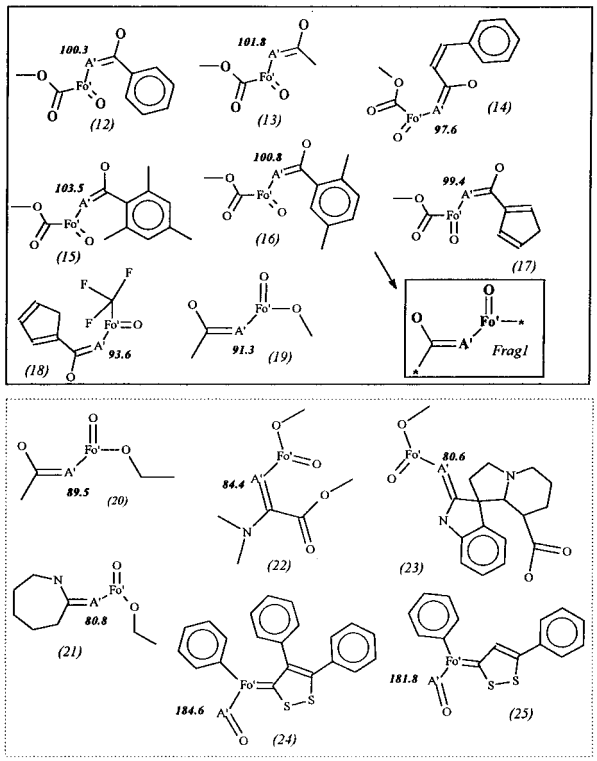


Figure 10. Structures associated with the (δFo, δA1) values responsible for the new area outlined in Figure 9. Structures *(20)*−*(25)* correspond to δA′, δA″ values eliminated by intersection of the virtual maps in Figure 9.

181.8 and 184.6 ppm) and cannot be transferred to the focus Fo of the $ELCO_b$. In structures *(20)*−*(23)*, atom A′ is bonded to a disubstituted sp^2 carbon that transmits to it across the double bond the β^{Π} effects of a carbon and an oxygen *(20)* or a nitrogen *(21)*−*(23)*. The elimination of these structures corresponds to the suppression of the effects of this double substitution, which is incompatible with the monosubstituted

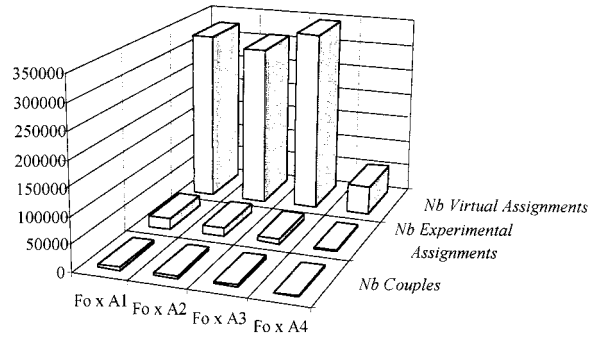


Figure 11. Comparison of the numbers of experimental and virtual attributions for the 14 663 carbon (Fo,A*i*) pairs making up the 8587 available $ELCO_b$.

carbon A2 of the central $ELCO_b$. The magnitude of the downfield shift (+10 ppm) of the new minimum value, δFo = 90 ppm instead of 80 ppm, expresses this suppression of the effect of this extra β^{Π} substituent. In structures *(12)*−*(19)*, atom A′ is also bonded to a disubstituted sp^2 carbon. However, the $\delta A' = 91.3-103.5$ ppm range assigned to these structures is included in the δA″ range. Therefore, these structures are considered as being consistent, and structures *(12)*−*(19)* are kept.

Perturbation of the Set of $ELCO_b$ by the Set of their Virtual $ELCO_c$. After describing the effect of a particular $ELCO_c$ on a central $ELCO_b$, then the effect of the set of $ELCO_c$ on this $ELCO_b$, we consider now the set of 8587 $ELCO_b$ taken from our data bank. The set of 300 million virtual environments that can be created around the $ELCO_b$ primitives represents a considerable contribution to the knowledge base of the EPIOS system. In other words, the experimental maps completed by the new virtual data extend its field of application at the same time as they reduce the risk of silence. The magnitude of the enhancement obtained is illustrated by comparing the number of (δFo, δA*i*) pairs brought together to define the experimental and virtual maps (Figure 11). For the 14 663 carbon pairs (Fo, A*i*) forming the 8587 available $ELCO_b$, 1 030 833 pairs (δFo, δA*i*) are formed from the 195 000 reference δ^{13}C to establish the virtual spectral maps.

The approximation introduced by the generic character of the (Fo, A*i*) atom pairs, used to describe the effect of the very varied environments considered, is good enough for the final spectral maps to be used to guide the exploration of an extended field of application with adequate discriminating power. The representation of these maps, defined as experimental maps by means of the MRS model, is identical to that of these experimental maps. They can therefore be used directly in the EPIOS system and require no special treatment, neither for recording nor for consultation.

The interest of the method presented here is also that it guaranties the coherence of the enhancement of knowledge of the EPIOS system. The defined virtual knowledge conserves a strict relationship with the experimental reality it describes. This proximity of the knowledge and its extended field of application is a necessary condition for the pertinence of an elucidation system. The method of virtual crowns is inspired in its principle by the generation law of the EPIOS system and the enhanced knowledge is directly useable by its elucidation strategy.

We shall now illustrate the interest of the relationships established between the reality of the experimental data and

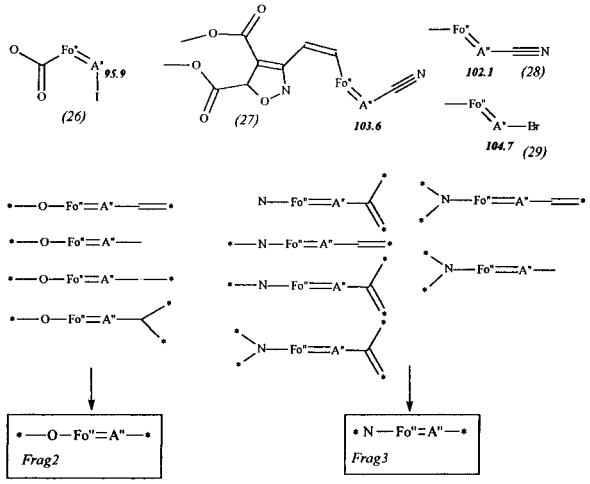


Figure 12. Structures associated with the (δFo″, δA″) pairs corresponding to the new area on the (δFo′, δA′) map of Figure 9.

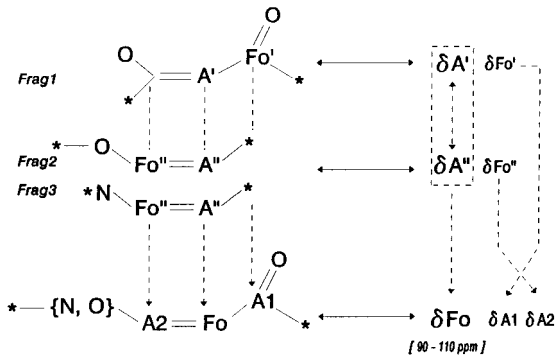


Figure 13. Structural interpretation of the new area of behavior of Figure 9 from the structural fragments *Frag1* − *Frag3*, defined in Figures 10 and 12.

the extrapolated virtual knowledge by analyzing a specific δ^{13}C behavior, revealed by this knowledge.

STRUCTURAL INTERPRETATION OF VIRTUAL BEHAVIOR

Generally, the variety of the virtual environments considered leads to an overall modification of the contours of the experimental maps. This modification is accompanied in the example considered in Figure 9 by the appearance of a new zone of behavior for the (Fo, Al) pair in the range where δFo = [90−110 ppm and δA1 = 160−175 ppm]. The δA′ behaviors responsible for this zone are associated with structures *(12)*−*(19)* discussed previously. They express the presence of the generic environment represented by *Frag1* in Figure 10 around the pair (Fo′, A′).

In this range of behavior, where δFo = 90−110 ppm, the second pair (Fo″, A″) appears in 75 structures for which 126 pairs of (δFo″, δA″) values are available. The atom Fo″ is attached to a carbon in only four of these structures [*(26)*−*(29)*, Figure 12], and in the 71 others, it is bonded to an oxygen or nitrogen atom. These latter are represented by the generic fragments *Frag2* and *Frag3* corresponding to environments common to the (Fo″, A″) pair. From these fragments and *Frag1* defined for the (Fo′, A′) pair one can attribute the behavior of the pair (Fo, Al) in the new region to environments of the β keto-vinyl ether or β keto-enarnine type shown in Figure 13. These interpretations are those that the EPIOS system would possibly propose in an

elucidation if the pairs of $\delta^{13}C$ values of the problem spectrum fell in this range and if oxygen or nitrogen atoms were present in the target structure.

CONCLUSION

The method for extrapolating spectral data described in this article has made it possible to considerably increase the spectral information extracted from our ^{13}C NMR data bank. It uses the structural and spectral relationships defined by the MRS model for the available experimental population. The rational generation of virtual crown environments leads to a generic representation of the set of potential environments of the experimental $ELCO_b$ primitives of the EPIOS elucidation system. This enhancement by virtual knowledge resulting from a combinatory generation represents a convenient tool for increasing the intelligence of a specific data bank by generic meta-data. The transfer of the $\delta^{13}C$ values associated with the component primitives of the crown environments towards their central $ELCO_b$ is filtered by the spectral maps characteristic of these primitives. It must be noted, however, that the exclusion of $\delta^{13}C$ values by filtering depends only on the currently available data, which are kept in the bank, and the contribution of new experimental data can eventually lead to their inclusion in future versions of the spectral maps.

The treatment cycle of our proposed extrapolation method is based on the $ELCO_b$ primitives, the generation of $ELCO_c$ environments, the spectral attribution by transfer of specific $\delta^{13}C$ data and the control of the virtual information deduced by establishing the range in which they are valid. The tools developed in this process in themselves represent an enhancement of the data bank. The extension of maps associated with the primitives increases their field of application and significantly improves the EPIOS decision power for elucidation. The extension leads at the same time to the selection of a greater number of these primitives during the interpretation of a spectrum. We shall show in a future article that handling the noise caused by these more numerous alternatives represents only a minor disadvantage compared with the extension of the scope of the EPIOS system.

ACKNOWLEDGMENT

The authors are greatly indebted to Dr. J. S. Lomas for help with the translation of this paper.

REFERENCES AND NOTES

(1) 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.

(2) Shelley, C. A.; Munk, M. E. CASE, A Computer Model of the Structure Elucidation Process. *Anal. Chim. Acta* **1981**, *133*, 507−516.

(3) Sasaki, S.; Kudo, Y. Structure Elucidation System Using Structural Information from Multisources: Chemics. (A Review about the CHEMICS Family) *J. Chem. Inf. Comput. Sci.* **1985**, *25, 252−257*.

(4) 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.

(5) Carabédian, M.; Dubois, J.-E. A Combined Model of Multi-Resonance Subspectra/Substructure and DARC Topological Structure Representation. Local and Global Knowledge in ^{13}C-NMR DARC Database. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 564−574.

(6) Warr, W. A. Computer-Assited Structure Elucidation. Part l. Library Search and Spectral Data Collections. *Anal. Chem.* **1993**, *65*, 1045A−1050A.

(7) Munk, M. E.; Lind, R. J.; Clay, M. E. Computer-Mediated Reduction of Spectral Properties to Molecular Structures. General Design and Structural Building Blocks. *Anal. Chim. Acta* **1986**, *184*, 1−19.

(8) Carabédian, M.; Dubois, J.-E. Inferring Extended Virtual Knowledge from an EPIOS Conversion Graph of Overlapping Subtructures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 701−706.

(9) Rowan, R., III; Sykes, B. D. A Carbon-13 Nuclear Magnetic Resonance Study of the Visual Chromophores and Model Compounds. *J. Am. Chem. Soc.* **1974**, *96*, 7000−7008.

(10) Becker, R. S.; Berger, S.; Dalling, D. K.; Grant, D. M.; Pugmire, R. J. Carbon-13 Magnetic Resonance Investigation of Retinal Isomers and Related Compounds. *J. Am. Chem. Soc.* **1974**, *96*, 7008−7014.

CI970024Y