

Computer-Assisted Structure Generation from a Gross Formula. 6. Reducing the Structural Redundancy by the Employment of 2D NMR Spectral Information¹

Ivan P. Bangov^{*,†} and Svetlana Simova

Bulgarian Academy of Sciences, Institute of Organic Chemistry, Building 9, Sofia 1113, Bulgaria

Daniel Cabrol-Bass and Isabelle Laude

LARTIC, Université de Nice, Sophia-Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

Received May 3, 1993[®]

The use of ¹³C chemical shift/structure relationships and 2D NMR information from different techniques as a constraint for the structure generation is discussed, and the development of a hierarchical multilevel generation procedure is made, the input information being molecular empirical formula and ¹³C and ¹H spectra with their cross peak assignments from the 2D maps. These form the basis for a heuristic search procedure over the generation tree. This search implies that at each level of the structure formation all possible extensions are investigated and only those with the best fit to the available spectral information are taken onto the next level of structure construction. Since the spectral information is uncertain in its nature, some elements of fuzzy logic are used in the determination of the hybridization states of the carbon atoms arbitrarily assigned with ¹³C NMR signals, and in the breadth-first-check procedure for selection of the proper extensions.

INTRODUCTION

The simulation of a spectrum can be expressed mathematically by the following equation:²

$$\text{spectrum} = f(\text{structure}) \quad (1)$$

The structure elucidation problem is the reverse of eq 1 and can be written as

$$\text{structure} = f^{-1}(\text{spectrum}) \quad (2)$$

The explicit solution of eq 2 is the ultimate goal of everyone initiating the build up of a computer-assisted structure elucidation system. However, the information content of any real spectrum is usually insufficient to this end. Accordingly, a procedure usually applied is to derive a set of chemical groups and fragments $\{f_i\}$ according to the following equation:

$$\{f_i\} = F^{-1}(\{p_i\}) \quad (3)$$

where $\{p_i\}$ is a set of spectral parameters. The operation F^{-1} is not a simple function; it is a procedure represented by a specific algorithm for transforming the spectral information into structural data. Both, the fragments f_i and the spectral parameters p_i are either derived from spectral databases or computed by a theoretical or empirical model of the structure/spectrum relationship.

Further, from the molecular empirical formula MF, the set of fragments $\{f_i\}$, and a set of constraints $\{c_i\}$ (they could be either chemical or spectral), a set $\{S_i\}$ of complete structures is generated:

$$\{S_i\} = \Gamma(\text{MF}, \{f_i\}, \{c_i\}) \quad (4)$$

Here Γ is a structure generation procedure. The set $\{S_i\}$ must be exhaustive and nonredundant; i.e. all the structures subject to the set of constraints $\{c_i\}$ without duplications have to be generated.

For each structure a spectrum may be simulated according to eq 1 and compared with the measured one. The structure

providing the best fit is then considered to be the correct one. However, several severe problems have been encountered with the practical application of this scheme.

First is the recognition problem. On the one hand, as mentioned above, most of the spectra provide insufficient structural information. On the other hand, all the spectral parameters p_i (chemical shifts, IR frequencies, etc.) are derived from complete structures but applied to the recognition of structural fragments. The real values of such fragment parameters are unknown. In addition to conventional spectroscopic methods, very rich sources of structural information are provided by the 2D NMR techniques. The *INAD-EQUATE* method is the most powerful, directly giving the carbon-carbon connectivities. But, this technique is extremely instrumentally time and substance extensive, and it has not yet become common routine practice. More interestingly for practical utilization are the *DEPT*, *¹H-¹H correlated 2D (H,H,COSY)*, and *¹³C-¹H correlated 2D (C,H,COSY or HETCOR)* spectra. The first one, in most cases, gives unambiguously the ¹³C multiplicities due to the adjacent ¹³C-¹H couplings. The second, gives the ¹H-¹H coupling map, and the third gives the ¹³C-¹H coupling map. However, yet again, a great deal of uncertainty is observed.

The second problem is the combinatorial problem. Owing to the uncertainty, discussed above, rather small fragments can be derived by applying any procedure generalized by eq 3, except in the *INADEQUATE* case. This leads to a combinatorial explosion in the generation procedure (4). To alleviate this problem fragments of greater size should be derived. However, the derivation of large fragments is subjected to the same difficulties as the determination of the whole complete structures.

Hereafter we shall discuss a novel approach to the treatment of the aforementioned problems. It is based on the following principles: First, *we cannot create information where it is absent. The less information we have means that more candidate structures will be generated.* However, a complete generation of all candidate-structures increases the *informational entropy*, rather than direct the decision-taking process to the correct solution. Second, *the spectral information is*

[†] Presently, a Senior Guest Scientist on a NATO grant at the University of Nice Sophia-Antipolis, Nice, France.

[®] Abstract published in *Advance ACS Abstracts*, February 15, 1994.

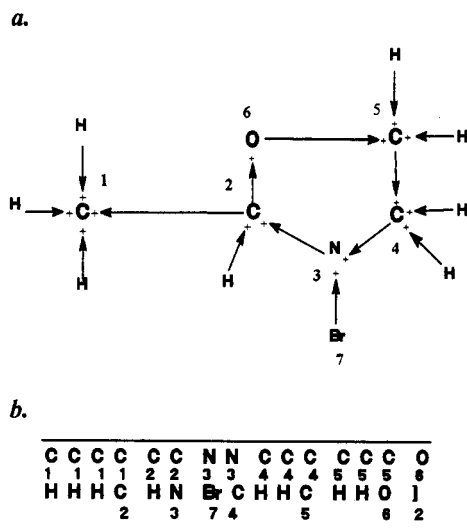


Figure 1. Directed graph (a) and two-row matrix representation (b) of chemical structure.

more or less fuzzy in its nature.³ Hence, the decision-making logic must reflect this fact. In order to obtain a positive solution a multicriteria inference procedure has to be developed. Accordingly, the development of an approach toward the effective utilization of available information content from different ¹³C NMR spectra, and toward combining the automated information treatment with the human reasoning, is made. The heart of this approach is a flexible generation procedure which guides the structure generation process either to one or to several possible structures, depending on the structural information available.

STRUCTURE GENERATION

The main features of our structure generation method, developed by one of us (I.P.B.) has been discussed in previous papers.^{1,4-7} Here, the basic ideas are outlined which help to give a greater understanding of the current developments.

We consider the mathematical representation of chemical structure as a directed graph. An example is provided in Figure 1. Following the graph-theory terms,⁸ a graph is defined as

$$G = [V, E] \quad (5)$$

where V is a vertex set and E is a pair relation applied on V . The vertices in constitutional chemistry are the atoms forming the structure. The relation E gives the mode of bond formation within the structure. Directed graphs consist of ordered pairs of vertices (v_i, v_j), which form arcs (in contrast to nondirected graphs which form edges), corresponding to the bonds in the chemical structure. Each free atom, depending on its valence state, contains n free valences, which correspond to the bonding sites (BSs) of the vertices. An arc is formed by linking two bonding sites formally considered to be of different type. They are denoted in Figure 1a by arrows and by crosses and have been named in previous papers^{1,6,7} as saturating valences (SVs) and saturation sites (SSs), respectively.

As a consequence, the bonding sites are partitioned into two sets: the set of saturating valences, $SVs \in S_V$ and the set of saturation sites $SSs \in S_s$. Practically the former are the BSs from where an arc originates, and the latter are the BSs where the arc ends. Our two-row matrix representation of the structure is given in Figure 1b. The first row elements are the SSs given by their identifiers and the numbering of the corresponding atoms. The second row elements are the SVs which form the connectivity within the structure by

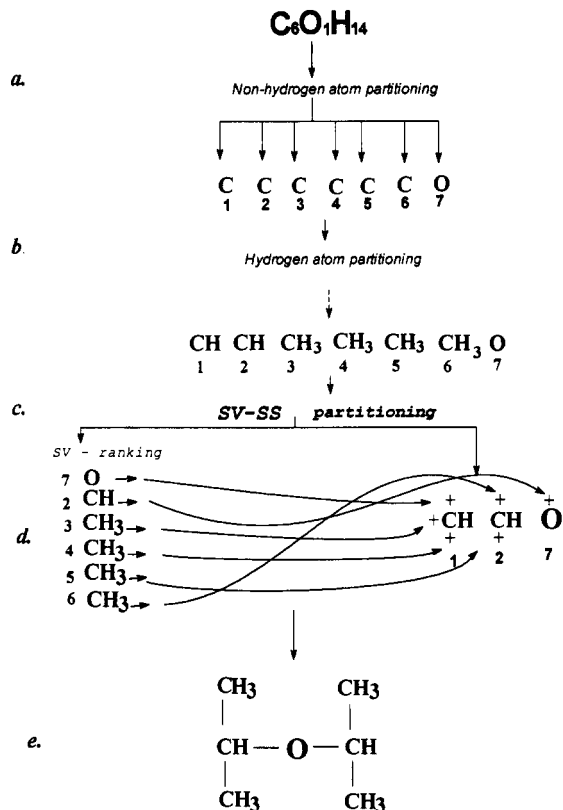


Figure 2. Partitioning of the non-hydrogen (a) and hydrogen (b) gross formula atoms, SV-SS partitioning (c), SV ranking (d), and SV-SS saturation (d), resulting in chemical structure (e).

juxtaposing to the first row SSs. Each SV-SS pair of matrix elements represents a bond. In cases of fragments the unsaturated SSs are denoted as "+" in the second row (not depicted in Figure 1). One can see that the first atom contains n SSs and no SV, while the other atoms have $(n - 1)$ SSs and 1SV. For atoms where a cycle is closed one more bonding site is transformed into SV. It is denoted by "J" (see atom 2 and the formation of the O_6-J_2 closure bond in Figure 1). The structure generation procedure consists of the production of all different transpositions of the SVs with respect to the SSs, so that each separate SS is saturated by a SV (Figure 2d).

The SV/SS partitioning of the bonding sites may look too artificial, but one can see from Figure 1a that it represents the father/son relationship⁹ of the nodes in rooted trees. The arrow of the SV points to the father node. The root of the tree is always the first atom. Hence, it has no father node, respectively no SV. Each one of the other nodes is connected with 1 father node and $(n - 1)$ son nodes, thus having 1 SV and $(n - 1)$ SSs.

This process is illustrated in Figure 2. The gross formula $C_6O_7H_{14}$ is first partitioned into non-hydrogen atoms (Figure 2a); then the hydrogen atoms are partitioned among the former (Figure 2b). Here, information is required as to the way of this partitioning. This may be either the number of single C atoms, CH, CH₂, and CH₃ groups, or the C-H direct multiplicities from either off-resonance or DEPT ¹³C NMR spectra (the spectral assignment will be discussed in the following sections).

The partitioning of the bonding sites into SVs and SSs and the subsequent generation of a structure is presented in Figure 2c-e. Even though the bonding sites in single atoms are equivalent, the selection of one out of all BSs of a given atom as SV is arbitrary, it is usually the first BS which is taken. However, this is not the case with the fragment atoms.

The fragments are treated in a way similar to single atoms: 1 BS from an atom of each fragment, except that of the first,

Table I. Valences, Atom Codes, and Initial Values R_0 for Different Types and Hybridization States

type and hybridization	valence	code	R_0
C			
sp ³	4	C	4
sp	2	#C	7
sp ² (olefinic)	3	=C	11
sp ² (aromatic)	3	Ph	13
N			
sp ³	3	N	15
sp	1	#N	18
sp ² (non-aromatic)	2	=N	20
sp ² (aromatic)	2	N:	20
O			
sp ³	2	O	23
sp ² (non-aromatic)	1	+O	25
sp ² (aromatic)	1	O:	25
F	1	F	32
Cl	1	Cl	33
Br	1	Br	34
I	1	I	35
S			
sp ³	2	S	28
sp ² (non-aromatic)	1	=S	29
sp ² (aromatic)	1	S:	29

is transformed into a SV; all the other ($n - 1$) BSs, into SSs. However, because the separate atoms in a fragment are not equivalent, all the nonequivalent atoms in turn provide one BS to be transformed into a SV, the others being SSs.

The atoms and their SVs are split into equivalent classes, the main classes being as follows: h, SVs of heteroatoms; m, SVs of multiple bond atoms; c, SVs of single carbon atoms or carbon atoms forming chemical groups with valences equal to or greater than 2; 1, SVs of carbon atoms from univalent fragments and chemical groups.

Each of these classes are additionally partitioned into subclasses according to a local index:^{1,6,7,10,11}

$$R = R_0 + N_H - q_C \quad (6)$$

where R_0 is a constant chosen for each type of atom and hybridization state in such a way that its values for two atoms or two hybridisation states of the same atom do not overlap. The R_0 values are provided in Table I. N_H is the number of directly attached hydrogen atoms, and q_C is the charge density, calculated by a fast procedure (the method of Gasteiger et al.¹² was used to this end). R has the same values for equivalent atoms and distinct values for nonequivalent atoms.

Finally, the atoms (their SVs, respectively), being of the same equivalence class, are ranked according to their numbering; the smaller the number, the lower the level it occupies. Thus, a hierarchical system of levels is constructed, each SV occupying one level (Figure 2d). The structure generation process is a depth-first procedure with backtracking from the first to the highest 1 level. It consists of the following steps:

SS Selection. At each level, starting from level 0 (before the first), a selection of the appropriate SSs for the next (higher) level SV-SS saturation is executed. It traces all the SSs and selects those which are eligible to form SV-SS atom extensions. Because the SV-SS saturation produces a linkage between two atoms, this procedure can be subjected to various rules which guide the structure formation to one or another result. The most trivial are those related to the chemical nature of the generated structure: (i) *Any selection of SSs from the atom providing the higher level SV is forbidden (atoms cannot connect with themselves, i.e. loops are prohibited).* Thus, no SS is selected from the oxygen atom 7 for the first level O(7) SV (see the curved arrows in Figure 3a and the dashed arrows in Figure 3b). The same can be observed for the higher levels presented in Figure 3.

(ii) *Only one SS from each atom is selected for a given SV (each atom in the molecule forms one, and only one bond with each other atom).* According to rule ii multigraphs are not treated in this approach. The double, triple, and aromatic bonds formed of atoms in different hybridization states are considered different atoms, such as =C, #C, Ph for carbon, =N, #N, N: for nitrogen, or =O, O: for oxygen, corresponding to sp² olefinic and sp and sp² aromatic states, respectively. The SS-selection procedure also allows some rules concerning the avoidance of the isomorphism redundancy to be applied.^{1,6,7} (iii) *If the higher and the current level SV atoms are in the same equivalence class, the selection of SSs starts at the SS saturated at the current level otherwise it starts from the first SS.* This rule avoids permutations between equivalent SVatoms yet allows them between non-equivalent atoms; e.g., in the case of the 4th level in Figure 3a, no SS is supposed to be selected from the first carbon atom CH(1). However, the selection in Figure 3, as well as some other selections provided here, violates both this rule and the next rules (iv and v) for reasons which will be discussed later. (iv) *Only one (from the atom bearing the lowest number) out of all SSs belonging to a given equivalence class is selected for the next higher level SV-SS saturation step; i.e. the selected SSs are of atoms belonging to different equivalence classes.* One can see from Figure 3a that a successive selection of SSs from carbon atoms CH(1) and CH(2) of the first substructure at the third level would lead to isomorphic extensions. (v) *If the SVs and the SSs are of the same equivalence class, only SSs from atoms bearing numbering lower than the corresponding SV atom are selected.* Rules (iii, iv and v) order the atoms belonging to a given equivalence class in a unique canonical form.

At each level a new partitioning of the atoms into equivalence classes is carried out (eq 6 and the procedure described above), taking into account the new extension. The problems concerning isomorphism in the case of the employment of 2D NMR spectra are more specific; hence rules iii-v are modified in this development. The particularities emerging from the analysis of the spectral information are discussed below.

The SS selection procedure gives a list of the possible SSs, i.e. of the possible extensions, which form alternative substructures. Any set of rules (knowledge base) developed to validate some or/and discard other extensions, SSs, respectively, could be applied at this stage. For instance, if the INADEQUATE spectrum is available, then only SSs producing C-C bonds with the higher order SV carbon atom corresponding to the 2D map patterns may be selected, the others being discarded. The set of ranked SVs with their lists of SSs can be viewed as a set of potential alternative graphs (structures). If more information is available producing more constraints, the list of SSs will be shorter (to one SS in the ideal case). Since the selection of the SSs is dynamic, i.e. the selection of each level is dependent on the lower level, there is an element of *heuristics* in this generation scheme. Furthermore, if we assign a *fuzziness number* to each selected SS, i.e. define a distribution of *possibilities*,¹³ we can consider this graph system as a *fuzzy graph set*.

SV-SS saturation. This is the stage where a new extension of the structure under construction is produced. All the selected SSs are in sequence saturated through a depth-first and backtracking procedure. The complete structures are generated at the highest level. Here, again, some checks concerning the full structure may be applied. The multilevel generation of an acyclic structure is exemplified in Figure 3a and the corresponding matrix representation of the SS-selection and SV-SS saturation steps in Figure 3b.



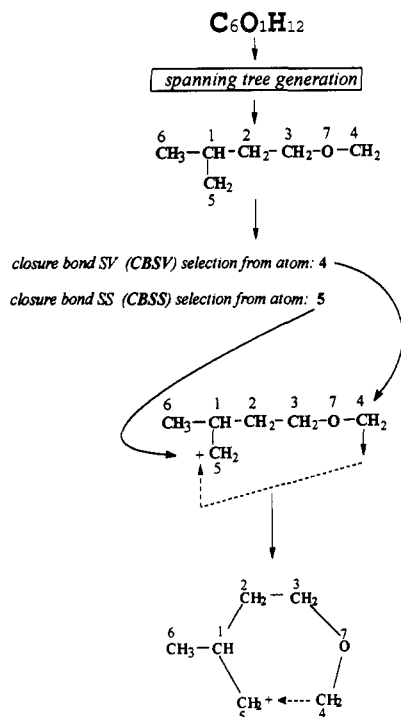


Figure 4. Selection of closure bond SV and SSs and cyclic structure generation.

Cyclic Structure Generation. In this case each generated acyclic structure is considered a spanning tree of the molecular graph, and a set of SVs, called closure bond SVs (CBSVs), from atoms having nonsaturated bonding sites is derived. The selection is subjected to the following rules: (vi) *all but the last atom BSs are traced and a combinatorial procedure selects D_C out of all CBSVs at a time.* Here D_C is the number of rings that the gross formula implies, and it is calculated from the following expression:

$$D_C = D_u - N_{db} - 2N_{tb} \quad (7)$$

where D_u is the degree of unsaturation and N_{db} and N_{tb} are the double and triple bond numbers, respectively. These CBSVs are ranked according to the numbering of their atoms in a system of D_C ring-closure levels. The last atom BS is not considered because there is no partner CBSS for this atom according to the rules given below. (vii) *Only one (having the lowest number) out of all BSs belonging to a given equivalence class is selected.* For each CBSV a set of partner SSs is selected, called closure bond SSs (CBSSs), according to the following rules: (viii) *Each CBSS has a number greater than the partner SV.* (ix) *No more than one CBSS from two or more equivalent and equidistant atoms (being at the same distance from the CBSV atom) is selected.* The same constraints as the aforementioned may be applied to the CBSS selection. Additionally, some constraints concerning the ring system, e.g. the size of rings, etc., can be imposed. The ring-closure process is also a recursive, hierarchical multilevel procedure of the stages *CBSV selection*, *CBSS selection* (for each higher level CBSV) and *CBSV-CBSS saturation*. The generation of a cyclic structure is illustrated in Figure 4. If the isomorphism-avoiding rules iii–v are not applied in the way outlined above, some rings may close before the CBSV–CBSS saturation step. Subsequently, the latter will produce the final (although acyclic) linking of the structure.

REDUCING THE STRUCTURAL REDUNDANCY

The process of structure generation within this hierarchical scheme can be illustrated by the generation tree shown in

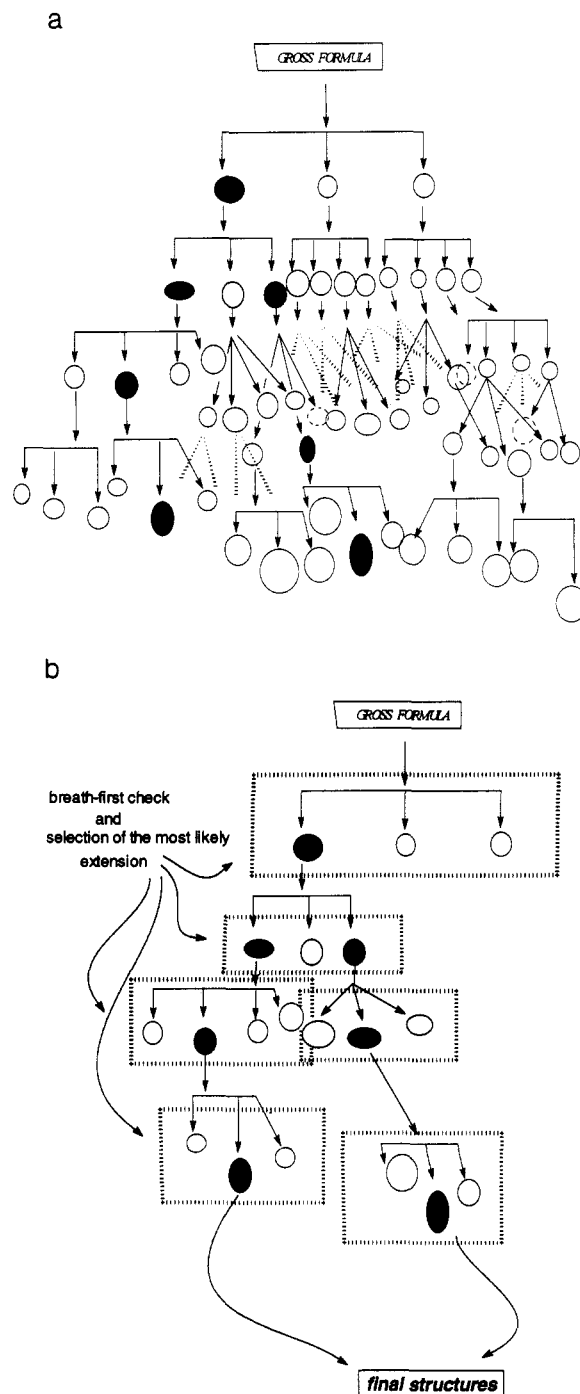


Figure 5. (a) Depth-first structure generation tree of combinatorial complexity. (b) Heuristic search over the generation tree.

Figure 5a. Each node of this tree, being an *extension* of the molecular graph, is either one or several substructures (when disconnected) generated at the lower levels and a complete structure at the highest levels. Evidently, if we try to generate all the structures, starting from a gross formula and using the *depth-first* procedure, discussed above, the problems of *combinatorial complexity* arise, and for most molecules, except the smallest, it leads to so called *combinatorial explosion*. In order to constrain this process the following approach is suggested:

(j) At each level all the SV–SS saturations emanating from a lower level tree node are traced (dash line boxes in Figure 5b). The corresponding extensions are generated and only one (in the most favored case) or a few extensions (SSs) are selected (the black circles in Figure 5a,b) for the next higher level. Thus, the generation process goes along one (or a few) branches of the generation tree straightforward to one (or a

few) solution, as shown in Figure 5b. The other branches of the generation tree are discarded (pruned); hence the generation process is greatly alleviated. Such a preliminary search over all nodes on a given level resembles the *breadth-first search* over the solution tree. Hereafter we call it *breadth-first-check* procedure.

(jj) The breadth-first-check approach requires the development of tools for the selection or discrimination of one or another substructure generated from a SV-SS extension at a given level. These might be rules or knowledge bases used to analyze the input spectral information. Hence, the structure is examined in the course of its formation (breadth-first-check in Figure 5b) instead of at the end, and the generation process is directed to one or a few most likely solutions. This type of search is known as a *heuristic search*¹⁴ (compare a with b in Figure 5).

Procedure j resembles the logic of CONGEN.¹⁵ However, in the latter case the function PRUNE is executed on a set of intermediate dummy structures generated from *super-atoms*—uniform moieties without any chemical details. Their chemical identification comes afterward, with the application of the function EMBED. In contrast, our program generates real substructures at each step, and this allows the construction of different sets of selection rules based on real chemical and/or spectroscopic knowledge according to jj.

Hereafter we discuss the utilisation of both single resonance and 2D NMR spectral information for the development of such discrimination tools.

Signal Assignment. The usual intuitive way of thinking of each spectroscopist is *first generate a structure and then assign the spectrum to it*. However, such an approach leads to additional combinatorial and signal recognition problems. Our assumption is the following: *Each signal of a proton-decoupled ¹³C NMR spectrum is associated with a carbon atom*. The occurrence of overlapping signals is easily detected and their number calculated by comparing the number of signals and the number of carbon atoms in the gross formula. Each existing overlapping signals should be user-identified and entered as separate entities but have the same chemical shifts and multiplicities. In the next step, the signals are split into different groups, depending upon their multiplicity and hybridization state. The multiplicities of the signals can be derived either from off-resonance or DEPT spectra. So long as the carbon atoms in the gross formula are nondistinguishable, the program assigns arbitrarily a signal to each carbon. Thus, each carbon atom obtains two attributes: chemical shift and multiplicity. The former transforms it into one of the hybridization types on the basis of the chemical shift values.

Intensive investigations of large ¹³C spectral databases have shown¹⁶ that single-resonance subspectra/substructure relationships cannot lead to accurate determination of chemical shift ranges associated with single carbon atoms. The use of classical chemical shift ranges tables¹⁷ with a house collection of 1300 spectra, corresponding to 8899 carbon atoms, led us to a rate of ambiguous or unclassified cases unacceptable for our purpose. Thus it was decided to use the fuzzy set theory¹³ in order to take into account the variability of chemical shifts for each type of carbon atom. Instead of crisp chemical shift ranges, we define fuzzy ranges for each multiplicity ($M = [1, 2, 3]$) and hybridization state (sp^3, sp^2, sp), leading to eight different cases, namely, $sp^3 C$, $sp^2 C$ (olefinic and aromatic), $sp C$, $sp^3 CH$, $sp^2 CH$, $sp CH$, $sp^3 CH_2$, and $sp^2 CH_2$. For each case k , a membership function $m_k(\delta_i)$ has been set up. For a given chemical shift assigned to the carbon atom number i , the value of $m_k(\delta_i)$ is an estimate of the degree of support that this atom is in fact a member of the fuzzy set k . This

membership function takes the value 1 when it is certain that the atom is a member of the set k ; thus it is in the associated hybridization state. It takes the value 0 when it is certain that this is not the case. The membership functions are built up by comparing distribution curves of observed chemical shifts for each set in our collection of spectra. When these distribution curves do not overlap, the membership function takes the value 0 for regions outside of the observed ranges and 1 for regions inside. In overlapping regions, an estimate of relative confidence is made by the ratio of the number of observed cases in each set. Transitions between regions (outside, overlapping, inside) are modeled by the use of a sigmoid function

$$m(\delta_i) = P + \frac{H}{1 + e^{-a(\delta_i - b)}} \quad (8)$$

The parameters P, H, a and b are determined empirically to fit the ratio in chemical shifts distribution. Depending upon the composition of the gross formula, three different sets of parameters have been determined: one for pure hydrocarbons, one for compounds containing only oxygen as heteroatoms, and one for compounds containing other heteroatoms. For each input signal the program calculates the membership functions, taking into account the gross formula, the multiplicity, and the chemical shift, resulting in a list of possible hybridization states sorted by decreasing order of support. The final decision of choosing the possible state is left to be made by the user.

Evidently, in the case of input fragments such an arbitrary assignment is not applicable. Hence, the fragment signals must be assigned by the user. Assuming that he has recognized the fragment from the spectral information, he could easily accomplish the assignment. So, as the employment of fragments within this approach is not so crucial for the constraint of the generation process, their input is justified only in cases where the user has a firm knowledge about their constitution and signal assignment. All these atoms, chemical groups, and/or fragments are the starting elements for the structure generation scheme described above, and they form a *hypothesis* for the subsequent elucidation procedure following rules j and jj. Accordingly, the structure elucidation problem may be redefined as follows: *find a set of connectivities between the atoms from the gross formula which gives the best fit to this arbitrary assignment*.

¹³C Chemical Shift/Chemical Structure Relationships. As discussed in the literature,¹⁶ the information content of the single-resonance ¹³C NMR spectra is insufficient to provide a full description of chemical structure. However, the chemical shift/carbon atom environment relationship might be extremely useful as an additional support of a given structural inference. Unfortunately the theoretical chemistry was unable to provide fast and reliable methods for the spectrum simulation according to expression 1. Consequently, two empirical approaches formalizing the chemical knowledge on this relation have been adopted. The first one is based on the development of inference knowledge bases from large sets of compiled structures and spectra. It requires the development of large databases.^{18–20} The second approach initiating from the Grant and Paul model²¹ consists of the development of empirically derived additivity rules.^{22–25} The latter was adopted because it was more appropriate to the development of our system. However, we must stress again that both the former and especially the latter are developed using spectral information derived from complete structures, and their application to substructure elucidation implies some fuzziness of the results. Hence, the additivity rule method was implemented here as an additional tool within a multicriteria scheme for the

prediction of the proper extension during the generation process. For each substructure (or set of substructures) resulting from a new extension, the Hamilton agreement factor (HAF)²⁶

$$\text{HAF} = \left(\frac{\sum_i (\delta_i^{\text{obs}} - \delta_i^{\text{calc}})^2}{\sum_i (\delta_i^{\text{obs}})^2} \right)^{1/2} \quad (9)$$

is calculated. δ_i^{obs} are the experimental shifts, i.e. they are fixed within the whole process of structure generation, and δ_i^{calc} are calculated by the additivity parameter method²⁷ shifts for each newly generated substructure(s). This factor shows how close to the structure in query, the substructure(s) generated at the current level is. Additionally, an alternative factor estimating the role of the chemical shift environment has been included in this system. It is based on the ¹³C chemical shift/charge density linear relationship:^{28,29}

$$\delta_{\text{C13}} = Aq_{\text{C}} + B \quad (10)$$

Here δ_{C13} are again the experimental chemical shift values, and q_{C} are the charge densities of the carbon atoms of the substructure(s) generated at the current level, calculated by a fast computational scheme. As discussed above, the charges had already been generated in order to be used in the local index R estimation according to expression 6. As in the case of the δ_i^{calc} the q_{C} are representing the substructure(s) generated at the current level, and they alter in the process of structure generation. The coefficients A and B are computed from the linear regression method, and the standard approximation error (SAE):

$$\text{SAE} = \left(\frac{\sum_i (\delta_i^{\text{obs}} - (Aq_i + B))^2}{N - 3} \right)^{1/2} \quad (11)$$

is consequently estimated. Both HAF and SAE make up the structure/chemical shift relationship. The lower their values, the closer the (sub)structure constitution is to that of the query structure. However, owing to their different nature they produce different results. We are not going to discuss their reliability here because it is a subject of a more comprehensive study. It should be mentioned though that for the case of substructure(s) relationship 9, respectively (10), is more flexible, perhaps because it was not derived from any complete structure spectral information. However, since the coefficients A and B have been calculated by a regression analysis, this factor is not applicable at the lowest levels, where the number of carbon atoms N is too small. Additionally, no reliable and fast π -electron charge density computational scheme has been available to us, up to now (such a method is under development). Hence, this factor is not in use in the cases of conjugated unsaturated compounds. It is clear that the reliability, and hence the selectivity, of the both factors will be lower in the beginning of the structure formation and will increase with its augmentation along with the increase in the environment of the carbon atom nuclei. The HAF value is divided by 10, rendering comparable SAE and HAF values.

An additional contradiction comes from the fact that the generation and breadth-first-check selection processes are carried out in one direction from the lowest to the highest level. This makes some chemical shifts follow the order of the the corresponding SVs to participate in early stages of the structure formation process than the others. However, as discussed above the chemical shifts within the same multiplicity are assigned arbitrarily. Hence, some extensions which might be not compatible at an early level owing to the SV carbon

atom chemical shift contribution to the HAF or SAE factors may appear compatible if those factors are calculated with a chemical shift assigned to a higher level SV carbon atom. Accordingly, the following procedure was developed: for each new extension produced at a given level by saturation of a new carbon atom SV all the higher level chemical shifts within the same multiplicity are assigned in turn to this atom, and the chemical shift producing the minimal HAF or/and SAE is taken as the correct assignment. Our practice shows that this leads to a more correct estimation of the plausible extensions at the lower levels which are more important for the alleviation of the combinatorial problem.

Isomorphism Problem. It is now understood that this problem emerges from the symmetry properties of the graph vertices and their partitioning into equivalence classes.^{1,6,7,30} To represent chemical structure in an algebraic form we are forced to label all the vertices with different numbers. Evidently, this labeling does not estimate the chemical nature and the equivalence status of the different nodes. A procedure for avoiding the isomorphism redundancy by avoiding permutations between BSs belonging to the same equivalence class was discussed above (rules iii–v). However, the notion of equivalence has to be re-defined in the lime light of the signal assignment. All the gross formula carbon atoms are initially equivalent, but after being attributed with ¹³C chemical shifts and multiplicities they become distinct. So long as all the connections between carbon atoms (hence between different signals) should be considered, the equivalence class partitioning is subjected to the following definition: *Two carbon atoms (graph vertices) are considered equivalent in the spectroscopic sense if they are topologically equivalent (e.g. having the same R values) and they have the same ¹³C NMR chemical shifts and multiplicities.* This means that the magnetic rather than the topological equivalence has been accepted here. Rules iii and iv are applied to our *heuristic search* generation procedure (rules j, jj), only within the framework of this definition. Thus, although some extensions result in (sub)structures which are isomorphic in topological sense, they are not considered to be duplicated in spectroscopic sense, because they emerge from different chemical shift combinations and produce different HAF and SAE values. Our measure for the topological isomorphism is the global charge related topological index (CTI):¹¹

$$\text{CTI} = \sum_{i \neq j} \frac{R_i R_j}{D_{ij}} \quad (12)$$

where R_i are the local indices given with expression 6 and D_{ij} are the distance matrix elements. Only extensions producing (sub)structures of equal CTI, HAF, and/or SAE values are pruned automatically; the others are left to be considered by the user within the breadth-first-check process.

2D Spectral Information. As stated above, the 2D techniques provide rich structural information. There are several structure elucidation systems^{31–33} utilizing these techniques as powerful structure inference tools. The first work in this field was reported by Lindley et al.³¹ The structural fragments were user-deduced from the 2D INADEQUATE spectrum and included as input to GENOA. The 2D INADEQUATE automated analysis was optionally included in the CHEMICS system.³² However, the data from other 2D techniques were not used in this system. This might be attributed to the fact that in CHEMICS the NMR signals are not assigned to the carbon and hydrogen atoms and this makes difficult the computerized analysis of the different 2D NMR COSY techniques. Christie and Munk³³ have recently reported their latest development: SESAMI. The work of this system is

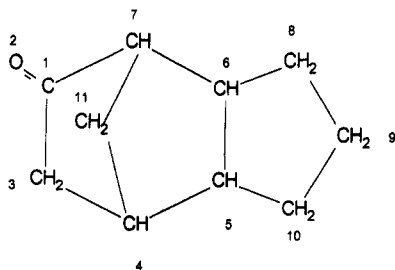


Figure 6.

divided into two tasks: spectrum interpretation (carried out by the procedure INTERPRET) and structure generation (use of the COCOA generator). INTERPRET is a two-track procedure: the first track PRUNE produces a shortlist of one-concentrically-layered atom-centered fragments (ACFs); the other track INFER produces substructural inferences serving as constraints on the structure generation process. Finally the COCOA generator generates all the alternative structures. However, the design of such a system requires the development of a comparatively large base of ACFs and its installation on more powerful computer facilities (SESAMI is installed on a VAX station). Evidently, our approach differs from these systems. Here we shall outline the use of the 2D spectral information within it.

2D INADEQUATE Spectrum. The information from this technique could be directly employed in the SS selection step. However, here again some ambiguities may be observed. Thus for the compound given in Figure 6 the coupling of the carbonyl carbon with the other carbons is not provided in the 2D map.³⁴ Furthermore, the 2D pattern indicates a non-existing carbon 3-carbon 7 bond. This pushed us to program the INADEQUATE module as a breadth-first-check rather than as a fully automated SS-selection procedure.

The ¹³C NMR spectrum is fed into the computer by entering the chemical shifts and multiplicities. Each signal is assigned with the corresponding cross peaks from the 2D pattern. Then the generation process starts. At each level a breadth-first-check to all possible extensions is carried out by checking each new bond (extension) whether it corresponds or not to the 2D signal pattern. Thus, if a correspondence is observed a "Y" (Yes) message is associated with this extension. In the case of a carbon-heteroatom extension a "?" message is provided, and in the case of receiving no signal for a carbon-carbon linkage a "N" message is associated with the generated extension. The user can scroll around the list of generated extensions and is able to call a graphical representation for each one of them. There are two options of graphical representation, the former giving the carbon atoms labeled as C, CH, CH₂, and CH₃ groups and the latter giving them with their assigned signal numbering. Accordingly, the final decision for whether an extension is to be pruned from the list or to continue is left up to the user. But his judgement is supported by several inference factors such as the HAF, SAE values, the (Y, ?, N) message (hereafter called approval message (AM)), and the graphical representation and by any previous information which the user has for the structure in question. Thus, for the structure presented in Figure 6 the user may consider the IR band at 1740 cm⁻¹ suggesting a five-membered-ring ketone, as discussed in ref 34, and select those extensions only, leading to the formation of such ring. Hence, although not fully automatic this approach allows any ambiguity which exists in the spectral information to be treated by multiple factors.

2D H,H,COSY and HETCOR Spectra. These spectral techniques are more accessible. The logic behind their usage

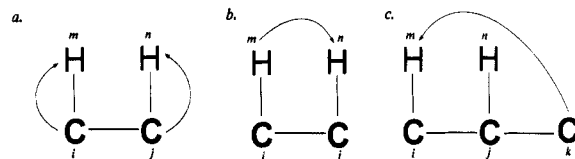


Figure 7. (a) C-H direct coupling in the HETCOR spectrum; (b) H-H vicinal coupling in the H,H,COSY spectrum; (c) C-H long-range coupling in the COLOC spectrum.

is presented in Figure 7a,b. The HETCOR spectrum provides the ¹³C-¹H direct couplings. As we consider the ¹³C chemical shifts as attributes of the carbon atoms, they bear their numbering. The ¹H NMR signals are entered as chemical shifts having a numbering which follows their successive feed into the computer. By observing the 2D HETCOR patterns, the user assigns each carbon atom ¹³C chemical shift with the corresponding ¹H cross peaks, e.g. chemical shift *i* (carbon atom *i*) is coupled to either to proton *m* or to protons *m*, *m'* (in the case where they are non-equivalent). In this way a list of the coupled protons is associated with each carbon atom.

The H,H,COSY spectrum provides the ¹H-¹H coupling patterns. The geminal couplings are automatically eliminated as they are linked to the same carbon atoms. Most interestingly are the vicinal couplings (Figure 7b). The logic is the following: *if both carbon atom i and carbon atom j are coupled to protons m and n, respectively, in the HETCOR spectrum and protons m and n are coupled in the H,H,COSY spectrum, then carbon atoms i and j form a bond (i,j).* This simple logic allows, in the most ideal case, where all couplings in the H,H,COSY spectrum are vicinal, an automatic construction of large pieces of the query structure to be accomplished. However, in many cases some longer range couplings appear. The latter are difficult to distinguished from the former in the COSY spectrum. Even a vicinal coupling in some rare cases may be absent, but longer-range couplings may be present. Moreover, the quaternary carbon atom and heteroatom connectivities cannot be treated in this way. This all makes the problem more uncertain, and some fuzziness in the logic should be inserted. Hence, a procedure similar to that previously described for the INADEQUATE case is developed. All substructures generated at a given level are scanned through a breadth-first-check procedure. For each substructure(s) the HAF, SAE factors and (N, Y, ?) messages are provided as described above. However, in this case the ? message is attributed to both connectivity with heteroatoms and connectivity with a quaternary carbon atom. Additionally, a possibility index (PI) was developed indicating the *possibility distribution*¹³ between the alternative solutions at the current level. It is based on the following formal rules:

- | | |
|----------|---|
| PI = 0.5 | in case the current SV-SS extension is a carbon-heteroatom bond (either SV or the SS atoms are heteroatoms) |
| PI = 0.5 | if either the current SV or SS atoms are quaternary carbon atoms, i.e. having ¹³ C- ¹ H multiplicity of 1 |
| PI = 1.0 | if the 2D spectra analyzed by using the logic described above indicate the presence of connectivity between the current SV and SS carbon atoms and such a connectivity is observed in the substructure(s) |

PI = 0.1 if the 2D spectra indicate the presence of connectivity between the current SV and SS carbon atoms but it is not found in the substructure(s), while a longer range C-C connectivity (through 2 or 3 atoms) is observed (thus, a PI \neq 0 is assigned, indicating a small possibility that this signal is a longer-range signal)

PI = 0.03 if 2D analysis rejects the presence of connectivity between the current SV and the SS carbon atoms but such connectivity is observed in the substructure(s), a small possibility is left for the rare cases where a signal does not appear

The cardinality of all possibilities is formed by summing the PIs of all extensions at the current level, and a possibility distribution is formed by assigning each structure with a *possibility weight* (PW) given as a percentage. Although rather formal this PW gives extremely valuable information about the 2D spectrum/substructure relationship.

In this way, at each level of the structure construction, the user has several indices indicating different aspects of the spectrum/structure relationship: HAF, SAE, the (Y, N, ?) message and the PW index plus the graphical representation of the substructure(s). All of them contribute to the formation of a multicriteria decision toward the elimination of the extensions not complying with the input spectral information and toward the selection of one or several extensions for the highest level generation process.

C, H, COLOC Spectrum. This option was developed in addition to the H,H,COSY-HETCOR procedure. As one can see from Figure 7c the 2D COLOC pattern gives the long range C-H couplings, usually through three bonds.³⁴ Hence, it may give information about the connectivity of quaternary carbon atoms. However, in many cases C-H couplings between two bonds and even direct couplings appear in it. The COLOC spectrum assignment is carried out by the user in the same way as in the HETCOR case. The two assignments are compared by the program, and the cross peaks originating from the direct couplings are automatically eliminated. For each substructure(s) at a given extension the distance matrix is calculated and the program checks whether a long range connectivity (through two or three bonds) between the current level SV carbon atom and the other carbon atoms with adjacent H atoms corresponding to the C-H COLOC coupling is present. If such a connectivity is encountered a Y message is produced, otherwise a N message appears. In the case where the SV atom is a heteroatom a ? message is produced.

PROGRAM IMPLEMENTATION AND APPLICATIONS

This approach has been implemented in a new development of the ASEC13 (automated structure elucidation from C-13 NMR data) system with the STRGEN structure generation module written in BORLAND Pascal 7 for DOS for the PC compatible computers. The initial input data are the molecular (gross) formula and the ¹³C NMR and/or ¹H NMR spectra. Several options are coded in this version: (1) use of the chemical shifts from broad band proton decoupled ¹³C NMR spectra; (2) use of chemical shifts and multiplicities from off-resonance or DEPT ¹³C NMR spectra; (3) use of 2D INADEQUATE; (4) use of H,H,COSY and HETCOR

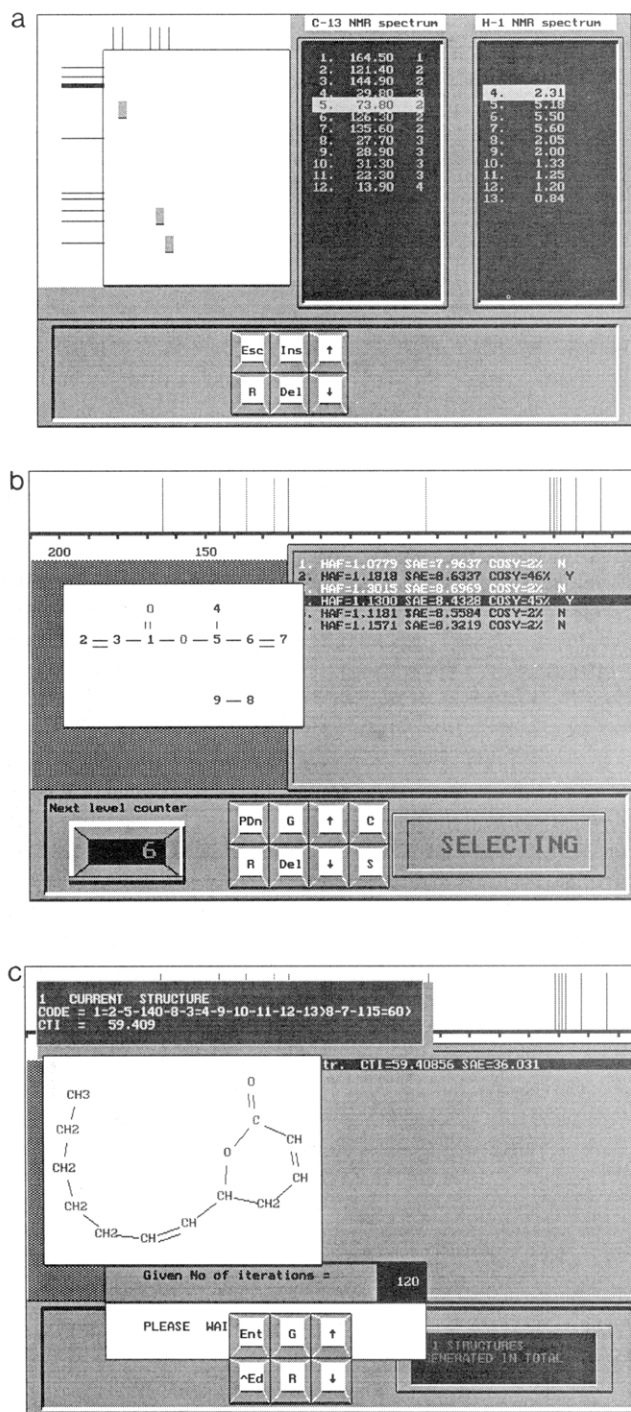


Figure 8. Screens from the program execution: (a) user-driven 2D spectrum cross peak assignment; (b) breadth-first-check selection of the most likely extensions; (c) final structure.

spectra; (5) use of the C,H,COLOC spectrum. Only in the last two cases is the input of the ¹H NMR spectrum required. There is also an option for a joint use of two or all 2D options, provided the spectral information is available.

First, the empirical formula is entered. Then the ¹³C NMR signals (chemical shifts and multiplicities) are fed into computer. For each ¹³C signal a list of the different hybridization states is provided with the corresponding possibility distributions given in percentage, as explained in the Signal Assignment section. The user is prompted to select the most likely hybridization state for each input signal.

For the cases in which either H,H,COSY or COLOC spectra are used, the proton spectrum is entered. The cross peaks of the 2D spectra are user-assigned (Figure 8a).

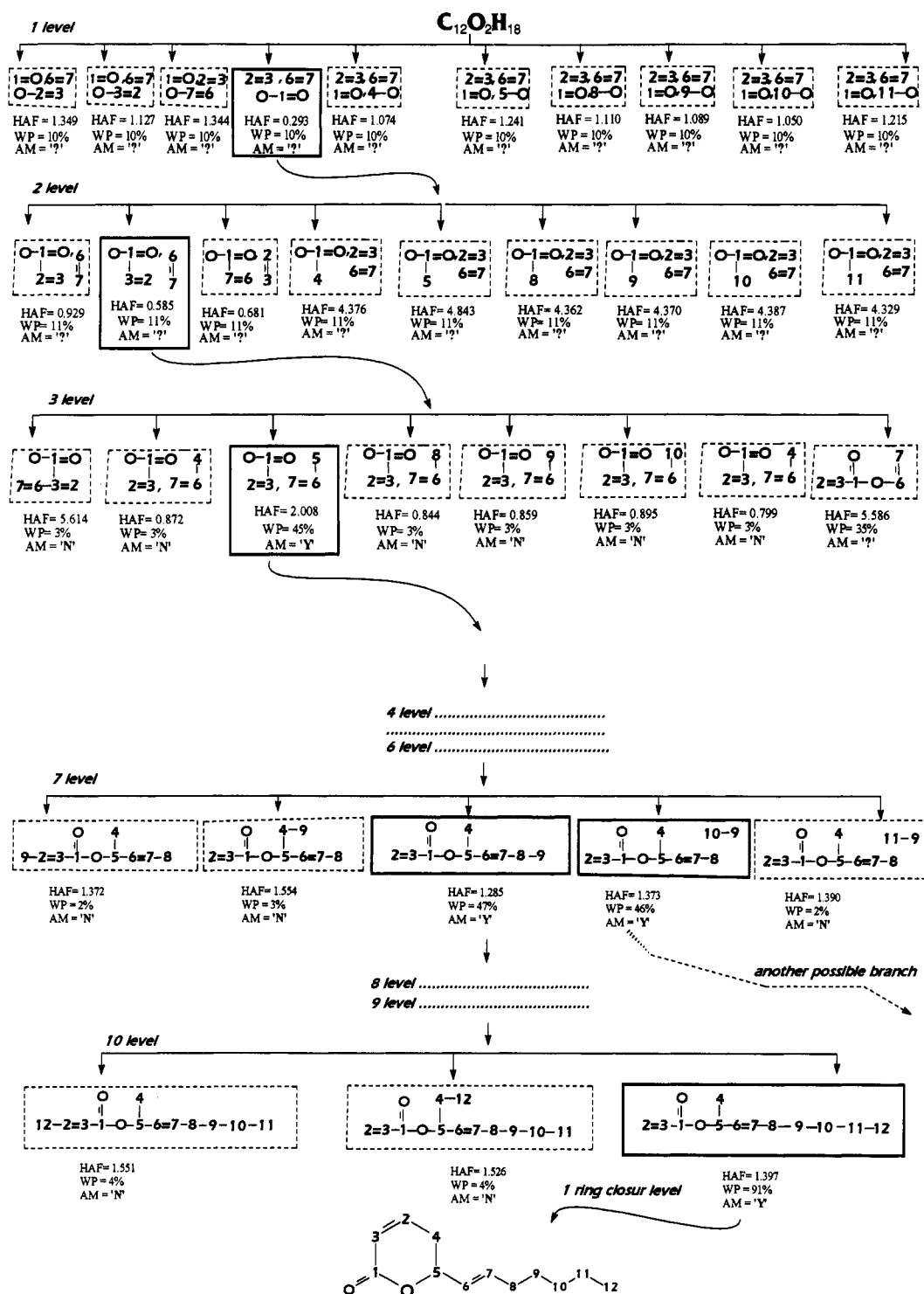


Figure 9. Heuristic search and multicriteria decision making based on 2D NMR spectral information.

In the next step, the program asks for the number of previously known fragments. In the case where the user has precise information about the constitution and spectral assignment of such structural pieces, they are fed into the computer by the graphical structure editor *FRAGED*. Their ^{13}C NMR signals must be user-assigned to the carbon atoms. However, in the most common cases such precise information is not available, in which case their number is specified as 0.

Having the above information, the program starts the level-by-level construction of the possible chemical structures. At each level a list of the extensions is displayed with the corresponding structure/spectrum compatibility factors: HAF and SAE for all cases 1-4, (Y, N, ?) message in addition to the 2D cases 3 and 4, the PW factor for the COSY-HETCOR case 4, and an additional (Y, N, ?) message for the case of

the COLOC spectrum. The user is prompted to delete from the list the substructures which do not comply to the best values of these factors and/or to his preliminary information about the query structure constitution (Figure 8b). In this way the breadth-first-check is carried out and the heuristic generation continues from a few extensions or, in the most favorable case, from one extension to the higher level structure augmentation.

At the end, several (or one in the most favorable case) final structures are listed and ordered according to their HAF factors (Figure 8c). It is assumed that the structure having the lowest HAF is the correct structure.

The process of structure generation and elimination is depicted in Figure 9. The 2D H,H,COSY and HETCOR spectra with the assignment of the cross peaks are taken from

Table II. ^{13}C and ^1H Chemical Shifts with the Corresponding ^{13}C - ^1H Multiplicities and Their C,H, HETCOR and H,H, COSY 2D NMR Cross Peak Assignments

signal nl.	^{13}C NMR		^1H NMR chemical shift	cross peak assignment
	chemical shift	multiplicity		
1.	164.5	1		
2.	121.4	2	5.99	H-3
3.	144.9	2	6.84	H-2, H-4
4.	29.8	3	2.34	H-3, H-5
			2.31	
5.	73.8	2	5.18	H-4, H-6
6.	126.3	2	5.50	H-5, H-7
7.	135.6	2	5.60	H-6, H-8
8.	27.7	3	2.05	H-7, H-9
			2.00	
9.	28.9	3	1.33	H-8, H-10
10.	31.3	3	1.25	H-9, H-11
11.	22.3	3	1.20	H-10, H-12
12.	13.9	4	0.84	H-11

ref 34 (Exercise 12). The ^{13}C and the corresponding ^1H chemical shifts and multiplicities (assigned according to the C,H,HETCOR) and their cross-peak H,H,COSY assignments are provided in Table II. The two olefin and the only carbonyl groups are entered as fragments and assigned as discussed above.

Ten extensions are possible at the first level and only four of them are chemically non-equivalent (non-isomorphic). However, the isomorphic extensions have different HAFs; i.e., they are magnetically non-equivalent, so they must be considered separately. All the extensions at this and the highest second levels have the same WPs and AMs because they originate from the linkage of an oxygen atom (first level) and of a carbonyl carbon atom (second level) to the remaining carbon atoms. Obviously, the C,H,HETCOR and H,H,COSY 2D spectra cannot provide information for their discrimination (AM = ? for all extensions). Consequently, the lowest HAF value criterion may be used to this end. Hence, the fourth extension (in the thick line box) for the first level and the second extension for the second level are selected as continuations for the highest level structure construction. All other extensions given in dashed line boxes, are pruned.

Further, the third extension at level 3 is undoubtedly selected as being the only extension with certain C-C linkage supported by the 2D spectral information (WP = 45%, and AM = Y). However, it produces a comparatively high HAF value. This might be explained with the fact that atom 5, having a chemical shift corresponding to a C-O simple bond, is still not bonded at this level to the oxygen atom. The same procedure is applied at levels 4-6. In the case of level 7 the two extensions 3 and 4 have similar values of the HAF, WP, and AM factors. Here both extensions lead to correct results since they give the different directions of the side chain formation, but in other cases they might produce different results. Hence, both branches must be considered.

The selection at the next levels 8-10 is unambiguous as shown at level 10. A six-membered cycle is formed by the only possible bonding of atoms 2 and 4 at the highest (first ring-closure) level. In this way the final structure is selected at the end.

This example is fortuitously simple. It was taken here for a better depiction of our approach. It should be noted that most of the real cases are much more ambiguous. They require both a more substantial involvement of the user's reasoning and more than one extension to be selected at some levels, which result in several final structures. Nevertheless, this procedure highly alleviates the combinatorial problem by directing the elucidation process to a restricted set of plausible structures. Furthermore, any combined utilization of data

from several 2D NMR techniques additionally increases the predicting power of this method.

CONCLUSIONS

The method and its program realization discussed in this paper are a new approach toward the automation of the structure elucidation process. It allows flexible and effective utilization of the whole structural content of the NMR information which is derived from both chemical shift environment/structure relationships and from the 2D patterns of the homo and heteronucleus spin-spin couplings. An important feature of this system is that it is developed for the cheapest range of computers, the PC-compatible, which you may find in all laboratories. It works perfectly on the 386 and 486 machines, but it can be also used on the older 286/287 processor computers. It is written in a compact form requiring only 223K of RAM (323K disk space).

ACKNOWLEDGMENT

This work has been partially supported by the National Scientific Fund at the Ministry of Culture, Science, and Education, Sofia, Bulgaria (Contract X96), and by a NATO grant for I.P.B. Their support is gratefully acknowledged.

REFERENCES AND NOTES

- (1) Paper 5 of this series: Bangov, I. P. Toward the Solution of the Isomorphism Problem in Generation of Chemical Graphs: Generation of Benzenoid Structures. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 167-173.
- (2) Munk, M. E. Symposium on Spectral Simulation. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 263. Relation 1 is given in this introduction to the symposium as properties = f(structure).
- (3) Otto M. Fuzzy Sets. Applications to Analytical Chemistry. *J. Anal. Chem.*, **1990**, 62 (14), 797A-802A.
- (4) Bangov, I. P. Computer-Assisted Generation of Molecular Structures from a Gross Formula. I. Acyclic Saturated Compounds. *Commun. Math. Chem. (MATCH)* **1983**, 14, 235-246.
- (5) Bangov, I. P.; Kanev, K. D. Computer-Assisted Structure Generation from a Gross Formula: 2. Multiple Bond Unsaturated and Cyclic Compounds. Employment of the Fragments *J. Math. Chem.* **1988**, 2, 31-48.
- (6) Bangov, I. P. Computer-Assisted Structure Generation from a Gross Formula. 3. Alleviation of the Combinatorial Problem. *J. Chem. Inform. Comput. Sci.* **1990**, 30, 277-289.
- (7) Bangov, I. P. Computer-Assisted Structure Generation from a Gross Formula. 4. Fighting against Isomorphism Decease. *Commun. Math. Chem. (MATCH)* **1992**, 27, 3-30.
- (8) Polansky, O. E. Elements of Graph Theory for Chemists. In *Mathematical Chemistry*; Bonchev, D., Rouvray, D. H., Eds.; Gordon and Breach: Chichester, U. K., 1991; Vol. 1, p 43.
- (9) Trinajstić, N.; Nikolic, S.; Knop, J. V.; Muller, W. R.; Szimanski, K. In *Computational Chemical Graph Theory*; Mellor, J. M., Ed.; Ellis Horwood: New York, London, 1991; Chapter 3, p 79. Contreras, M. L.; Valdivia, R.; Rozas, R. Exhaustive Generation of Organic Isomers. I. Acyclic Structures. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 223-230.
- (10) Bangov, I. P. Use of ^{13}C NMR Chemically Equivalent Signals and Topological Symmetry for Removal of Redundant Structures. *Commun. Dep. Chem. (Bulg. Acad. Sci.)* **1988**, 21, 194-198.
- (11) Demirev, P. A.; Dyulgerov, A.; Bangov, I. P. CTI: A Novel Charge-Related Topological Index with Low degeneracy. *J. Math. Chem.* **1991**, 8, 367-382.
- (12) Gasteiger, J.; Marsili, M. Iterative Partial Equalisation of Orbital Electronegativities-A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, 36, 3219-3228.
- (13) Zadeh, L. A. Fuzzy Sets as a Basis for a Theory of Possibility. *J. Fuzzy Sets Syst.* **1978**, 1, 3-28.
- (14) Nilsson, N. L. *Problem-Solving Methods in Artificial Intelligence*, McGraw-Hill: New York, 1971; Chapter 3, p 53.
- (15) Masinter, L. M.; Sridharan, N. S.; Lederberg, J.; Smith, D. H. Applications of artificial intelligence for chemical inference. XII. Exhaustive generation of cyclic and acyclic isomers. *J. Am. Chem. Soc.* **1974**, 96, 7702-7723.
- (16) Dubois, J. E.; Carabedian, M. Single-Resonance Subspectra/Substructure Investigations of the C13 DARC Databank. Representation of Local and Global Topological Knowledge. *J. Chem. Inf. Comput. Sci.* **1991**, 31, 557-564.
- (17) Wehrli, F. W.; Marchand, A. P.; Wehrli, S. *Interpretation of Carbon-13 NMR Spectra*. Wiley: Chichester, New York, 1988.
- (18) Bremser, W.; Klier, M.; Meyer, E. Mutual Assignment of Subspectra and Substructures-A Way to Structure Elucidation by C13 NMR Spectroscopy. *Org. Magn. Reson.* **1975**, 7, 97-105.

- (19) Dubois, J.-E.; Carabedian, M.; Dagane, I. Computer-aided elucidation of structures by carbon-13 nuclear magnetic resonance. The DARC-EPIOS method: Characterisation of ordered substructures by correlating the chemical shifts of their bonded carbons. *Anal. Chim. Acta* **1984**, *158*, 217-233.
- (20) Robien, W. Computer-Assisted Structure Elucidation of Organic Compounds. III: Automatic Fragment Generation from ^{13}C -NMR Spectra. *Mikrochim. Acta* **1986**, *II*, 271-279.
- (21) Grant, D.M.; Paul, E.G. Carbon-13 Magnetic resonance. II. Chemical Shift Data for the Alkanes. *J. Am. Chem. Soc.* **1964**, *86*, 2984-2990.
- (22) Pretsch, E.; Clerc, T.; Seibl, J.; Simon, W. *Tables of Spectral Data for Structure Determination of Organic Compounds*; Springer: Berlin, 1989.
- (23) Brown, D. W. A Short Set of ^{13}C NMR Correlation Tables. *J. Chem. Educ.* **1985**, *65*, 209-212.
- (24) Cheng, H. N.; Bennett, M. A. Trends in shift rules in carbon 13 nuclear magnetic resonance spectroscopy and computer aided shift prediction. *Anal. Chim. Acta* **1991**, *242*, 43-56.
- (25) Pretsch, E.; Furst, A.; Badertscher, M.; Burgin, M. E. C13 Shift: A Computer Program for the Prediction of C13 NMR Spectra Based on an Open Set of Additivity Rules. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 291-295.
- (26) Hamilton, W. C. Significance Test on the Crystallographic R Factor. *Acta Crystallogr.* **1965**, *18*, 502-510.
- (27) The parameters coded in our program are taken from refs 22 and 23.
- (28) Bangov, I. P. Use of the ^{13}C NMR Chemical Shift/Charge Density Linear Relationship for Recognition and Ranking of Chemical Structures. *Anal. Chim. Acta* **1988**, *209*, 29-43.
- (29) Bangov, I. P. Computer-Assisted Structure Elucidation Based on ^{13}C NMR Spectral Information. *XXVI CSI Select. Pap.* **1990**, *8*, 267-283.
- (30) Faulon, J.-L. On Using Graph Equivalent Classes for the Structure Elucidation of Large Molecules. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 338-348.
- (31) Lindley, M.; Schoolery, J.; Smith, D.; Djerassi, C. Application of the Computer Program GENOA and Two-Dimensional NMR Spectroscopy to Structure Elucidation. *Org. Magn. Reson.* **1983**, *21*, 405-411.
- (32) Funatsu, K.; Susuta, Y.; Sasaki, S. Introduction of Two-Dimensional NMR Spectral Information to an Automatic Structure Elucidation System, CHEMICS. Utilization of 2D-INADEQUATE Information. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 6-11.
- (33) Christie, B. D.; Munk, M. The Role of the Two-Dimensional Nuclear Magnetic Resonance in Computer-Enhanced Structure Elucidation. *J. Am. Chem. Soc.* **1991**, *113*, 3750-3757.
- (34) Duddeck, H.; Dietrich, W. *Structure Elucidation by Modern NMR. A Workbook*; Steinkopff Verlag, Springer-Verlag: Darmstadt, New York, 1992.