Structure Elucidation from 2D NMR Spectra Using the *StrucEluc* Expert System: Detection and Removal of Contradictions in the Data

Sergey G. Molodtsov,[†] Mikhail E. Elyashberg,[‡] Kirill A. Blinov,[‡] Antony J. Williams,[*,§]
Eduard E. Martirosian,[‡] Gary E. Martin,[||] and Brent Lefebvre[§]

Novosibirsk Institute of Organic Chemistry, Siberian Branch of Russian Academy of Science,
Lavrentiev Avenue 9, Novosibirsk 630090, Russia, Advanced Chemistry Development, Moscow Department,
6 Akademik Bakulev Street, Moscow 117513, Russian Federation, Advanced Chemistry Development Inc.,
90 Adelaide Street West, Suite 600, Toronto, Ontario M5H 3V9, Canada, and Rapid Structure
Characterization Group, Global Research and Development, Pfizer, Kalamazoo, Michigan 49001-0199

Received January 26, 2004

The elucidation of chemical structures from 2D NMR data commonly utilizes a combination of COSY, HMQC/HSQC, and HMBC data. Generally COSY connectivities are assumed to mostly describe the separation of protons that are separated by 1 skeletal bond ($^3J_{HH}$), while HMBC connectivities represent protons separated from carbon atoms by 1 to 2 skeletal bonds ($^2J_{CH}$ and $^3J_{CH}$). Obviously COSY and HMBC connectivities of lengths greater than those described have been detected. Though experimental techniques have recently been described to aid in the identification of the nature of the couplings the detection of whether a coupling is 2-bond or greater still remains a challenge in most laboratories. In the *StrucEluc* software system the common lengths of the connectivities, 1-bond for COSY and 1- or 2-bond for HMBC, derived from 2D NMR data are set as the default. Therefore, in the presence of any extended connectivities contradictions can appear in the 2D NMR data. In this article, algorithmic methods for the detection and removal of contradictions in 2D NMR data that have been developed in support of *StrucEluc* are described. The methods are based on the analysis of molecular connectivity diagrams, MCDs. These methods have been implemented in the *StrucEluc* system and tested by solving 50 structural problems with 2D NMR spectral data containing contradictions. The presence of contradictions was detected by the algorithm in 90% of the cases, and the contradictions were automatically removed in ∼50% of the problems. A method of "fuzzy" structure generation in the presence of contradictions has been suggested and successfully tested in this work. This work will demonstrate examples of the application of developed methods to a number of structural problems.

1. INTRODUCTION

2D NMR spectra are the primary form of spectral data used today for the elucidation, rather than the confirmation, of molecular structures. During the past decade, a series of expert systems based on 2D NMR spectral data have been reported, for instance.[1−5] Previously the *ACD/Structure Elucidator* (*StrucEluc*) expert system was described, and the results of structure elucidation for ca. 150 natural products were discussed.[6−11] The conclusion from these reports is that the system is able to solve very complicated problems if the 2D NMR data are free of contradictions. Since the system is adjusted by default to account for the coupling constants $^{2,3}J_{HH}$ and $^{2,3}J_{CH}$ which are common for COSY and HMBC correlations correspondingly (referred to as "standard" correlations), contradictions will appear when at least one correlation of >3 bonds results in a response in the 2D NMR data. Despite recent developments to aid in the identification of the correlation lengths[12,13] there is no presently available NMR technique capable of distinguishing couplings of different lengths in a reliable fashion. Therefore, development of theoretical methods for 2D NMR data analysis that identify the presence of "nonstandard" correlations is of considerable importance.

During this work algorithms and programs have been developed that are able to detect the presence of nonstandard correlations in 2D NMR data in the majority of cases. Moreover, algorithms that help to remove contradictions by lengthening certain connectivities have been delivered. An algorithm for "fuzzy" structure generation was elaborated, which allows structure generation to be performed using the condition that the lengths of several connectivities can be varied by means of searching a set of connectivities. The resulting programs were tested on 50 challenging elucidation problems where the 2D NMR data were known to contain nonstandard connectivities and the results are detailed in this report.

2. DETERMINATION AND REMOVAL OF CONTRADICTORY CONNECTIVITIES

2.1. Basic Terms and Definitions. In the *StrucEluc* expert system, a molecular connectivity diagram (MCD) is used as

[*] Corresponding author phone: (919)570-0217; fax: (425)790-3749; e-mail: tony@acdlabs.com.
[†] Siberian Branch of Russian Academy of Science.
[‡] Advanced Chemistry Development, Moscow Department.
[§] Advanced Chemistry Development Inc.
[||] Pfizer.

10.1021/ci049956+ CCC: $27.50
Published on Web 08/04/2004

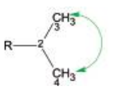
a graphical representation of connectivities that are obtained from processed 2D NMR data (see examples reported in refs 6−11). Connectivities are designated as $C_{kl}(\upsilon_i\text{-}\upsilon_j)$, to indicate that vertices υ_i and υ_j are at a distance of k to l bonds from each other in the chemical structure, while it is assumed that υ_i and υ_j, in the general case, are skeletal atoms. For instance, the vicinal coupling constant $^3J_{HH}$ observed in COSY implies a connectivity $C_{11}(\upsilon_i\text{-}\upsilon_j)$ of length *one* bond and $^{2,3}J_{CH}$ HMBC coupling constant is associated with a connectivity $C_{12}(\upsilon_i\text{-}\upsilon_j)$ whose length varies from one to two bonds. In other words, $C_{12}(\upsilon_i\text{-}\upsilon_j)$ (or *1,2- connectivities)* are connectivities whose lengthes may be 1−1, 1−2, or 2−2 bonds.

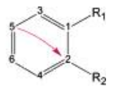
A molecular connectivity diagram can contain connectivities of two types: *fixed* and *formal*. *Fixed* connectivities are those of a fixed length that are either specified by the user on the basis of some prior information or are input on the basis of certain 2D NMR methods (for example ref 12). In particular, *chemical* bonds in a structure can be defined as fixed connectivities. Examples of fixed connectivities having length 1, 2, and 3 bonds correspondingly are $C_{11}(\upsilon_i\text{-}\upsilon_j)$, $C_{22}(\upsilon_i\text{-}\upsilon_j)$, and $C_{33}(\upsilon_i\text{-}\upsilon_j)$.

Formal connectivities are those drawn by the MCD program itself on the basis of 2D NMR data analysis and with default correlation lengths assumed for each type of two-dimensional spectra. In particular, default connectivities $C_{11}(\upsilon_i\text{-}\upsilon_j)$ are produced from COSY data and $C_{12}(\upsilon_i\text{-}\upsilon_j)$ from HMBC data. Formal connectivities can be divided into two categories—*standard* and *nonstandard* ones. *Formal* connectivities that correspond to the structure of the analyzed molecule will be called *standard* connectivities, and they allow the program to generate a correct structure.

If a 2D NMR spectrum contains correlations of a *nondefault* length (exceeding the default upper limits) and their lengths were not identified explicitly by the chemist, the corresponding *formal* connectivities will be called *nonstandard*. Note that in the case when there is no prior information about the presence of nondefault correlations all formal connectivities are considered as standard. Obviously *formal nonstandard* connectivities contradict the structure of the molecule under study.

Examples of nonstandard COSY (two C−C bonds, green two-sided arrow) and HMBC (three C−C bonds, red one-sided arrow) connectivities are shown below:





The nonstandard COSY connectivity C_{22}(C3−C4) and HMBC connectivity C_{33}(C5−C2) will be accepted by the program as C_{11}(C3−C4) and C_{12}(C5−C2) connectivities correspondingly, if their true lengths were not specified by the chemist on the basis of preliminary 2D NMR experiments.

Since the structure generator is adjusted to utilize connectivities of default (standard) lengths, nonstandard connectivities, if present in a diagram, yield either a zero result since no structure can be generated, or an invalid result where the results file does not contain a correct structure. This happens as the algorithm considers the topological distances between corresponding skeletal atoms (they are specified by inherent chemical shifts) of a nonstandard connectivity equal to standard length, while in reality this distance is longer. Therefore, nonstandard connectivities appear “nonstandard” to the software algorithms. Nonstandard connectivities may also appear due to heavy overlap of the NMR resonances or accidental degeneracy which can lead to permuted pairs of assignments.

We assume the fixed connectivities, unlike the formal ones, are standard (correct) in all cases. A set of vertices in the structure separated by a distance of k bonds from the υ vertex will be denoted as the *kth layer* of the υ vertex environment. The set of vertices in the structure separated by less than or equal to k bonds from the υ vertex shall be the *kth sphere of the υ vertex environment*. Generally, an MCD is considered to be a graphical representation whose vertices are skeletal atoms of the molecule under study and with the hydrogen atoms attached.

Assume that an MCD diagram is created on the basis of a series 2D NMR spectra for an unknown. In other words, there is a set of fixed and formal connectivities. The task is to detect the nonstandard connectivities, or, at least, locate these by detecting small groups of connectivities that contain nonstandard responses. A correct solution and structure determination of the compound under study can be achieved only after the correction or elimination, of *all* nonstandard connectivities.

Experience indicates that all nonstandard connectivities can be partitioned into at least four different types which are dependent on their position in the MCD and the extent to which they contradict the rest of the connectivities:

1. Explicit nonstandard connectivities that contradict the environment of a skeletal atom (vertex) and some fixed connectivities that refer to this vertex.

2. A group of connectivities containing *at least one* nonstandard connectivity. Vertices and vertex pairs with a set of connectivities that contain nonstandard ones are detected.

3. A group of connectivities referring to a vertex that contradict other connectivities that refer to other vertices.

4. Implicit nonstandard connectivities that are not obviously contradictory but nevertheless prevent the elucidation of a correct structure.

It is necessary to consider methods that allow the detection and removal of each type of nonstandard connectivity.

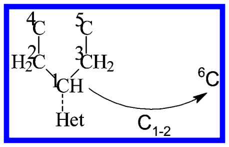
2.2. Explicitly Nonstandard Connectivities. An explicitly nonstandard connectivity can be detected if a vertex exists whose *fixed* connectivities fully determine *all vertices and all bonds between them* in the *rth* sphere of this vertex environment. If there are formal connectivities $C_{kl}(\upsilon_i\text{-}\upsilon_j)$, where $l \le r$, defined for this υ vertex, it is possible to determine the appropriateness of each of the connectivities.

Furthermore, the validity of a formal connectivity $C_{kl}(\upsilon_i\text{-}\upsilon_j)$ may sometimes be determined even if *not all the vertices and bonds* in the first sphere of the υ_1 vertex are defined. It is only possible if a set of the allowed environment spheres is defined for each of the C-vertices. These spheres are defined using the Atom Property Correlation Table (APCT)[11] and is based on chemical shifts of the appropriate carbon atoms.

It should be noted that there is a significant difference between the ambiguity associated with the HMBC or COSY correlations and that associated with the APCT. For instance, the HMBC correlations (in particular, the standard ones of

2 or 3 bond lengths) are defined by nature and cannot be overcome without the development of new experimental techniques for distinguishing correlations of different lengths. This is true to a greater extent as related to the nonstandard correlations. At present *StrucEluc* uses structure generators to search all possible variants for the different distance lengths of standard bond length selective NMR correlations within a reasonable time frame since these bond length selective NMR experiments do not presently exist. The ambiguity involved in the elucidation process using APCT data has a different character and is a consequence of the fuzziness of knowledge regarding the interrelation between units used to construct the molecular structure and their associated spectral features. The diversity of the different kinds of atom environments means that it is impossible to enumerate all conceivable atom combinations and their related spectral features. It is therefore necessary to determine some typical atom combinations and associate assigned ranges of spectral features which characterize them. The boundaries of these spectral ranges are fuzzy since such a combination could have a spectral feature (for instance, a ^{13}C chemical shift) that is observed slightly out of the limits of a specific range. While using the Atom Property Correlation Table can be deemed to be a risky procedure, it is difficult to avoid since it helps overcome the possibility of a combinatorial explosion and the enormous time associated with structure generation. Such a situation is inevitable in the case of big molecules when there are a large number of atoms and no assumptions can be made about the nearest neighbors of a given carbon atom having specific chemical shifts.

Example 1. Assume that the figure below represents a fragment with the fixed bonds indicated for the v_l vertex environment (in the figure the vertices are denoted by their numbers, and the bond with a heteroatom provided for by the APCT is represented by a dashed line):



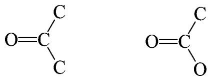
Assume there is a formal HMBC connectivity $C_{12}(v_1\text{-}v_6)$ and the APCT indicates that the v_6 *vertex has no bonds with heteroatoms.* In this case, the v_6 vertex cannot obviously be located both in the first and second layers of the v_1 *vertex,* and the $C_{12}(v_1\text{-}v_6)$ connectivity is, therefore, nonstandard.

Explicitly nonstandard connectivities can be removed either by simply deleting them or by increasing the maximum possible distance between the vertices. Replacing the $C_{12}(v_1\text{-}v_6)$ connectivity with $C_{13}(v_1\text{-}v_6)$ in the above example ensures that it is no longer nonstandard. Explicitly nonstandard connectivities are the primary targets for detection and removal during the elucidation process.

2.3. Sets Containing Nonstandard Connectivities. Obviously it is possible to determine the maximum possible number of vertices in each layer and therefore in each sphere of a vertex. The CH_3 vertex, for example, always has one vertex in the first layer, and no more than 3 vertices in the second layer (e.g., $CH_3{-}C(C)_3$). The second sphere of a CH_3 vertex therefore has no more than 4 vertices. Thus, if a CH_3 vertex has more than 4 different 1,2-connectivities, at least one of these must be nonstandard. The data regarding the maximum possible number of vertices in the second sphere are used as a criterion for detecting nonstandard connectivities in the vertices and vertex pairs with a relatively large number of formal 1,2-connectivities.

2.3.1. Vertices with Nonstandard Connectivities. Consider the molecular connectivity diagram of a compound defined on a set of N vertices, where N is the number of skeletal atoms in the molecule. The task is to determine if the set of 1,2-connectivities of a given vertex v contains at least one nonstandard connectivity.

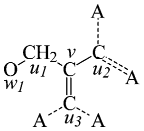
As defined previously each C-vertex has a set of allowed *environment spheres* defined on the basis of the corresponding carbon atom chemical shift value and the total set of vertices. A quaternary carbon atom with a chemical shift of 180 ppm (at the boundary of the chemical shift range characteristic for esters and ketones) in a compound containing only C, H, and O atoms, for example, has the following allowed spheres:



Having analyzed these sets, we can calculate the maximum possible number of adjacent C-vertices and the maximum possible number of adjacent vertices for each C-vertex. Define the maximum possible number of the adjacent C-vertices for the C-vertex v as $D^C(v)$ and the maximum possible number of the adjacent vertices as $D(v)$. Obviously, $D^C(v) \leq \mathrm{val}(v)$ and $D(v) \leq \mathrm{val}(v)$, where $\mathrm{val}(v)$ − is the valence of the v-vertex. The following equations $D^C(u) = \mathrm{val}(u)$ and $D(u) = \mathrm{val}(u)$ are presumed to be true for the non C-vertex u.

It is assumed that there is a set of 1,2-connectivities for the v vertex. The term "*the admissible bond distribution of the v vertex*" is introduced here and described as follows. Take into account that bonds between separate vertices in a structure can be provided as fixed connectivities and a bond distribution of the v vertex to other vertices can be constructed. For this purpose, bonds are drawn from the v vertex to other vertices so that the sum of bond multiplicities is equal to the valence of the v vertex and the resulting sphere of the v vertex is *allowable.*

Let $U = \{u_j\}$ be the set of vertices adjacent to the v vertex and $W = \{w_k\}$ be the set of vertices adjacent to at least one vertex of U. The drawing below shows an example fragment whose real bonds are shown as solid lines and hypothetical bonds as dotted lines. Undefined vertices that can correspond to either carbon atoms or to heteroatoms are marked as A. The figure indicates that $D^C(u_1) = 1$, $D(u_1) = 2$, $D^C(u_2) = 3$, $D(u_2) = 3$, $D^C(u_3) = 3$, and $D(u_3) = 3$.



Consider whether the 1,2-connectivities of the v vertex correspond to the obtained fragment. Obviously, the correspondence for each formal $C_{11}(v{-}v_i)$ connectivity of the v vertex can be determined. The $C_{11}(v{-}v_i)$ connectivity of the v vertex corresponds to the fragment if $v_i \in U$. If at least one C_{11} connectivity of the v vertex does not correspond to the fragment, the distribution of bonds for the v vertex is *inadmissible*. Assume that all C_{11} connectivities of the v vertex correspond to the fragment. Now check to determine if any of C_{12} and C_{22} connectivities of the v vertex do not correspond to the fragment. The $C_{12}(v-v_i)$ connectivity of the v vertex *does not correspond to the* fragment if the vertex $v_i \in U \cup W$ and cannot be connected to any of the vertices $u_j \in U$ regarding the allowable *environment* spheres of the v_i and u_j vertices. If at least one C_{12} or C_{22} connectivity of the v vertex does not correspond to the fragment, then this bond distribution of the v vertex is *inadmissible*.

It is noteworthy that both the C_{12} and C_{22} connectivities of the v vertex can be partitioned into three groups:

1. connectivities corresponding to the fragment;
2. connectivities not corresponding to the fragment;
3. connectivities whose correspondence to the fragment are undefined.

Connectivities corresponding to the fragment contain information regarding the distances between fragment-compatible vertices. These distances will still be valid regardless of the bonds between the vertices $u_j \in U$ (i.e. the vertices from the first layer of the v vertex) and other vertices. However, in addition to these, there may be connectivities that are feasible or infeasible depending on the vertices that the u_j vertex is bound to. The correspondence of the $C_{12}(v-v_i)$ (or $C_{22}(v-v_i)$) connectivity to the vertex is regarded as *undefined* if the vertex $v_i \notin U \cup \mathrm{W}$ and there exists such a vertex $u_j \in U$ so that the bond between the v_i and u_j vertices is possible.

Assume that there are no $C_{12}(v-v_i)$ and $C_{22}(v-v_i)$ connectivities of group 2 (connectivities not corresponding to the fragment). For the time being disregard the C_{12} and C_{22} connectivities of the v vertex whose correspondences to the fragment are confirmed. Later the rest of the connectivities (group 3) whose correspondence to the fragment remains undefined will be analyzed. Let $K_2(v)$ be the number of the remaining C_{12} and C_{22} connectivities of the v vertex. In other words, $K_2(v)$ is the number of vertices that are to be additionally included in the second layer of the v vertex under the bond distribution of the v vertex under consideration.

Let $d^C(u_j)$ and $d(u_j)$ be the number of the adjacent C-vertices and the number of the adjacent vertices of the u_j vertex in the examined fragment, respectively. Denote as $D_2^C(v) = \sum_j[D^C(u_j) - d^C(u_j)]$ and $D_2(v) = \sum_j[D(u_j) - d(u_j)]$ the maximum possible numbers of the additional C-vertices and vertices in the second layer of the v vertex, respectively. Thus, the following is true $d^C(u_1) = 1$, $d(u_1) = 2$, $d^C(u_2) = 1$, $d(u_2) = 1$, $d^C(u_3) = 1$, $d(u_3) = 1$ and $D_2^C(v) = 4$, $D_2(v) = 4$ for the fragment in the drawing.

Depending on the type of vertices, there may be two cases: (1) all vertices v_i in the remaining $C_{12}(v-v_i)$ and $C_{22}(v-v_i)$ connectivities are C-vertices and (2) there is a non-C v_i vertex. Consider the first case. Let all vertices v_i in the remaining $C_{12}(v-v_i)$ and $C_{22}(v-v_i)$ connectivities be C-vertices. Then, if $K_2(v) > D_2^C(v)$, it is impossible to locate all vertices v_i from the remaining $C_{12}(v-v_i)$ and $C_{22}(v-v_i)$ connectivities in the second layer of the v vertex; therefore, the examined bond distribution of the v vertex is *inadmissible*.

If, however, there is a non-C v_i vertex (second case), the examined bond distribution of the v vertex is *inadmissible*, when $K_2(v) > D_2(v)$. The examined bond distribution of the v vertex is termed *admissible*, if the fact that it is *inadmissible* is not proven.

Let $\{v_i\}, i = 1,...N$ be the set of vertices in the MCD. To determine whether the given set of 1,2-connectivities of the v_i vertex contains at least one nonstandard connectivity, all possible bond distributions for the v_i vertex are consecutively generated. Only if all bond distributions of the v_i vertex appear to be *inadmissible*, can it be concluded that the set of its 1,2-connectivities contains at least one nonstandard connectivity. The algorithm below describes the process for determining the vertices with nonstandard 1,2-connectivities:

0. $i = 0$

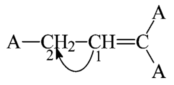
1. $i = i + 1$

2. Testing feasibility of recurrent bond distribution of the v_i vertex. If the distribution is not allowable, it is concluded that *the set of 1,2-connectivities of the* v_i vertex contains at least one nonstandard connectivity, go to 1.

3. The next distribution of bonds made with other vertices in the v_i vertex is generated.

4. Verification of the bond distribution of the v_i vertex for being admissible. If the bond distribution is inadmissible, transfer to 2, otherwise, go to 1.

Example 2



Consider five 1,2-connectivities $C_{11}(v_1\text{-}v_2)$, $C_{12}(v_1\text{-}v_3)$, $C_{12}(v_1\text{-}v_4)$, $C_{12}(v_1\text{-}v_5)$, and $C_{12}(v_1\text{-}v_6)$ to be specified for the v_1(CH) vertex in the fragment above, where the hypothetical heavy atoms attached to the free bonds of the fragment are denoted by A. It is also known that all the given vertices are C-vertices, and the chemical shift values of the corresponding C atoms indicate that the v_1 atom is in the sp^2-hybridization state and that the v_2-v_6 atoms are in the sp^3-hybridization state. Presume that it is known that the v_2 vertex corresponds to the $-CH_2-$ group. Now, it will be proven that the given set of 1,2-connectivities of the v_1 vertex contains at least one nonstandard connectivity.

The input data indicates that the v_1 vertex can be bound to two vertices only. Presume that all of the connectivities associated with the v_1 vertex are standard. Then, the v_1 vertex is to be connected to the v_2 vertex by an ordinary bond, and the second bond should have a multiplicity equal to two. The vertices v_3-v_6 corresponding to the C-atoms in the sp^3-hybridization state cannot have incident multiple bonds and, therefore, cannot be connected to the v_1 vertex. As a result, these 4 vertices are to be located in the second environment layer of the v_1 vertex regardless of the distribution of bonds of the v_1, i.e., $K_2(v_1) = 4$. On the other hand, the maximum possible number of vertices in the second layer of the v_1 vertex is 3 (see the figure above). Therefore, all vertices represented by $v_3\text{-}v_6$ cannot be simultaneously located in the second layer of the v_1 vertex regardless of its bond distribution. Hence, the given set of 1,2-connectivities of the v_1 vertex contains at least one nonstandard connectivity, as suggested.

2.3.2. Selection of Vertex Pairs Having Nonstandard Connectivities. If nonstandard connectivities are *not revealed* as a result of an analysis of individual vertices, it does not necessarily indicate that they are really absent. It may happen that among the vertices for which the algorithm fails to detect the presence of nonstandard connectivities it is possible to find groups of vertices that possess the following property: the union of connectivities included in the groups must contain a nonstandard connectivity. At the same time, the vertices included in the group may not have common connectivities (see example below).

Currently consideration is given to groups containing only two vertices and having 1,2-connectivities. This is for the following reasons: (a) the number of possible groups formed from more vertices would be rather large, and (b) in those cases where a group containing many vertices with nonstandard connectivities was selected, the union of connectivities belonging to these vertices would, in turn, contain a great number of connectivities. This reduces the significance of selecting a group of vertices. The smaller the set of connectivities among which nonstandard connectivities exist, the greater is the probability of removing any contradictions in the initial connectivity set.

If a MCD contains N vertices, then the number of different groups consisting of two vertices is obviously equal to $N\cdot(N - 1)/2$. In reality, only groups where both vertices contain 1,2-connectivities can be considered, while the presence of nonstandard connectivities is not revealed for these particular vertices. As an example, assume that two vertices A and B are examined and at least one of them has an associated nonstandard connectivity. The algorithm can fail to detect the presence of the nonstandard connectivity. However, when these vertices are unified in a group of two vertices, the algorithm could become capable of detecting the presence of a nonstandard connectivity in the connectivity set formed from the connectivities belonging to *both* vertices.

Let $\{v_1,v_2\}$ be a group of two vertices where each contains 1,2-connectivities. Consider all bond distributions at the vertex ν_1 one at a time. For each bond distribution of vertex ν_1 consider all bond distributions at ν_2 vertex. This will help to determine the admissibility of the bond distributions of both the ν_1 and ν_2 vertices. If it is shown that there are no admissible bond distributions for ν_1 and ν_2 vertices, it will prove that the 1,2-connectivity union of both ν_1 and ν_2 vertices contains at least one nonstandard connectivity. Consider the algorithm describing the connectivity analysis of the vertex pair.

1. Check whether the next bond distribution of the vertex ν_1 is possible. If the distribution is not possible, then end the analysis at this point. It is concluded that the *set of 1,2-connectivities of* ν_1 *and* ν_2 *vertices contains at least one nonstandard connectivity*.

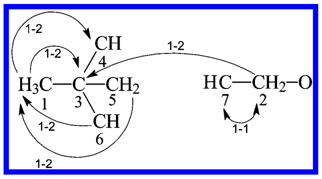
2. The next bond distribution of the ν_1 vertex is then generated for the rest of the vertices.

3. Check whether the next bond distribution of the vertex ν_2 is possible. If the distribution is not possible then go to 1.

4. The next bond distribution of the ν_2 vertices is then generated for other vertices.

5. Check whether the bond distribution of the ν_1 and ν_2 vertices is admissible. If the bond distribution of least one of these vertices is not admissible, then go to 3. Otherwise, end the analysis at this point. The conclusion is that *the set of 1,2-connectivities of* ν_1 *and* ν_2 *vertices is admissible*.

Example 3. Let $v_1(CH_3)$, $v_2(CH_2)$, $v_3(C)$, $v_4(CH)$, $v_5(CH_2)$, $v_6(CH)$, and $v_7(CH)$ be vertices, and the NMR chemical shift of the corresponding C atom indicates that the ν_2 vertex is bonded to an oxygen atom (designated as vertex O). In addition, the following formal connectivities are given for the ν_1 vertex: $C_{12}(v_1\text{-}v_3)$, $C_{12}(v_1\text{-}v_4)$, $C_{12}(v_1\text{-}v_5)$, $C_{12}(v_1\text{-}v_6)$ and for the ν_2 vertex: $C_{11}(v_2\text{-}v_7)$, $C_{12}(v_2\text{-}v_3)$. It can be shown that the given 1,2-connectivity set related to the ν_1 and ν_2 vertices contains at least one nonstandard connectivity.



Assume all connectivities specified at vertices ν_1 and ν_2 are standard. It is evident from the connectivities of vertex ν_1 that a unique admissible bond distribution at vertex ν_1 exists: it must be bonded to the ν_3 vertex. In reality the CH_3 vertex always contains only one vertex in the first layer of the environment, while no more than 3 vertices can exist in the second layer (see the figure above). Therefore, the first and second layers of the CH_3 vertex environment may not contain more than a total of 4 vertices. In this case, exactly four vertices will be present in the environment only if the CH_3 vertex is connected with a tetravalent C-vertex having no multiple bonds. Since there are four C_{12} connectivities at the ν_1 vertex, only vertices v_3, v_4, v_5, and v_6 located at opposite ends of the connecting bonds can be present in both the first and second layers of the ν_1 vertex environment, and the ν_1 vertex has to be connected to the tetravalent C-vertex. As is shown in the figure among all the vertices v_3, v_4, v_5, and v_6 only the ν_3 vertex is a quaternary C-vertex. Consequently, the ν_1 vertex must be connected to the ν_3 vertex, while the vertices v_4, v_5, and v_6 are located in the second layer of the ν_1 vertex environment as shown in the figure.

Furthermore, taking into account the fact that the ν_2 vertex is connected with the O-vertex and considering the connectivity $C_{11}(v_2\text{-}v_7)$, it is obvious that the ν_2 vertex has a unique bond distribution: the ν_2 vertex must be connected with both the ν_7 and O vertices. The figure illustrates that the ν_7 and O vertices adjacent to the ν_2 vertex cannot be connected with the ν_3 vertex, that is, the unique bond distribution of the ν_2 vertex is inadmissible. It follows that the union of 1,2-connectivities at the ν_1 and ν_2 vertices must contain at least one nonstandard connectivity. It is impossible to prove the presence of a nonstandard connectivity at the ν_1 and ν_2 vertices if at least one of the mentioned connectivities is removed. This explains why the algorithm is not always capable of indicating the possibility of a nonstandard connectivity present in an MCD.

2.3.3. Nonstandard Connectivity Removal When Vertices and Groups of Vertices with Nonstandard Connectivities Are Revealed. To remove problems from a set of connectivities containing nonstandard connectivities two methods are suggested:

1. Lengthen the shortest connectivities of the set by one bond.

2. Lengthen *all* connectivities of the set by one bond.

The first method is used when it can be supposed that the presence of nonstandard connectivities C_{11} is more probable compared with the probability of the presence of C_{12} nonstandard connectivities in a given problem. At the same time all connectivities of equal length are assumed to have the same probability of being nonstandard. The choice of method for the removal of a nonstandard connectivity is the choice of the user. The second method is generally employed more frequently however. In each specific case, the best manner to remove contradictions is determined by trial and error.

The general strategy for contradiction removal in 2D NMR data consists of two steps: first, all vertices and groups of vertices having, according to the algorithm, nonstandard connectivities are determined, and then *all* connectivities included in these "suspicious" connectivity sets are lengthened by one bond. If the contradictions are removed as soon as vertices with nonstandard connectivities are detected it is possible that the presence of nonstandard connectivities at another vertex will become undetectable after the error is removed from the connectivity set of a given vertex.

Assume ν_1 and ν_2 are two vertices whose connectivity sets contain nonstandard connectivities. Allow a connectivity $C_{11}(v_1\text{-}v_2)$. Lengthen the connectivities at the ν_1 vertex only. In particular the connectivity $C_{11}(v_1\text{-}v_2)$ will be replaced by $C_{12}(v_1\text{-}v_2)$. Now two cases are possible:

- Repeated analysis shows that the set of connectivities at the ν_2 vertex does not contain nonstandard connectivities. Consequently, lengthening only one connectivity at the ν_2 vertex removes the "inaccuracy" in the set of connectivities. It is assumed that all short connectivities are *nonstandard with equal probability*.
- As before, repeated analysis shows that the connectivity set at the ν_2 vertex contains nonstandard connectivities. Then, lengthening *all* connectivities at the ν_2 vertex will lead to the replacement of the $C_{12}(v_1\text{-}v_2)$ connectivity by $C_{13}(v_1\text{-}v_2)$, i.e. the length of the initial $C_{11}(v_1\text{-}v_2)$ connectivity will be extended more than those of any of the other connectivities. This is undesirable since it will make structural constraints buried in the 2D NMR data more fuzzy and will, as a result, increase in more structures being generated. Additional structures, where ν_1 and ν_2 vertices are separated by three bonds, will be generated in the presence of the $C_{13}(v_1\text{-}v_2)$ connectivity, which would be impossible in the presence of the $C_{12}(v_1\text{-}v_2)$ connectivity.

As a result of lengthening the connectivity we obtain a *renewed* initial connectivity set. At this stage the search for nonstandard connectivities begins over in the renewed connectivity set.

2.4. Nonstandard Connectivities that Prevent Structure Generation. Assume that explicit nonstandard connectivities as well as vertices and groups of vertices having nonstandard connectivities are not found in the initial or renewed set of connectivities. Nevertheless, when the presence of contradictory data prevents the generation of chemical structures corresponding to all connectivities, it may be possible to detect the presence of nonstandard connectivities in the initial connectivity set. Moreover, it is possible to predict a vertex containing nonstandard connectivities.

A bond distribution for a vertex ν is termed *accurate* if the bond distributions of all filled vertices, including the ν vertex, are *admissible.* If *all* accurate bond distributions at the ν vertex are analyzed, then it may be possible to determine the *obligatory* and *forbidden* bonds between the ν vertex and other vertices. If *all* accurate bond *distributions*of the ν vertex contain any bonds, then all of these bonds are defined as *obligatory* ones. Bonds that are absent in all the accurate bond distributions are *forbidden*. Detecting obligatory and forbidden bonds in a chemical structure is based on the assumption that all the given connectivities are *standard*.

It is necessary to sequentially conduct a bond analysis of all vertices. If, when analyzing bonds at the ν vertex, it is found that there is no accurate bond distribution of the given vertex, it means that the connectivities of the ν vertex are at variance with the connectivities of the vertices analyzed earlier. It follows that the assumption regarding the validity of all connectivities is incorrect, and as a result the MCD will therefore contain some nonstandard connectivities. The bond refinement of any vertex directly influences the bond analysis of all the other vertices. Hence, after the bond analysis of the last vertex, it is necessary to start analyzing the bonds of the first vertex. Bond analysis is only complete when new obligatory and forbidden bonds are no longer identified as a result of successive bond analysis of *all* vertices.

Let *i* be the number of the current vertex under examination, *j*, be the number of the vertex for which the analysis of connectivities is completed, and let *N* be the total number of vertices in a structure. The algorithm for checking the MCD for the presence of nonstandard connectivities can be written in a series of steps as follows:

1. $i = 1$, $j = 1$.

2. Perform the bond analysis of the ν_i vertex. If the accurate bond distributions of the ν_i vertex are not found by bond analysis of this vertex, then END the procedure. It can be concluded that the connectivities belonging to the ν_i vertex are in contradiction with all remaining connectivities, and the initial data set therefore contains nonstandard examples.

3. Determine new obligatory and forbidden bonds of the ν_i vertex with other vertices. If refinement is possible, then $j = i$.

4. $i = i + 1$. If $i = N + 1$, then $i = 1$.

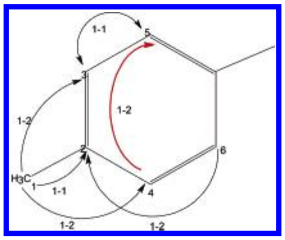
5. If $i = j$, then END the procedure. The conclusion will be that nonstandard connectivities cannot be identified.

6. Return to step 2.

It should be noted that in practice the vertex for which the contradiction was identified during bond analysis often contains nonstandard connectivities. This follows from the fact that the number of vertices having nonstandard connectivities is usually small. Most frequently, bond analysis is carried out initially for vertices with standard connectivities. As a result the correct obligatory and forbidden bonds are determined in the structure. When a contradiction is identified during the bond analysis of a recurrent vertex, it is very probable that the nonstandard connectivities of a given vertex are in conflict with those belonging to vertices which have already been analyzed. In any case, the connectivities of the found vertex are then lengthened to attempt removal of the contradictions, after which the check for the presence of nonstandard connectivities is restarted.

Example 4. Assume that the formal connectivities $C_{11}(v_1\text{-}v_2)$, $C_{12}(v_1\text{-}v_3)$, $C_{12}(v_1\text{-}v_4)$, $C_{12}(v_2\text{-}v_6)$, $C_{11}(v_3\text{-}v_5)$, and

$C_{12}(v_4\text{-}v_5)$ are set as the fragment vertices for $v_1(CH_3)$, $v_2({=}C{<})$,$v_3({=}CH\text{-})$, $v_4({=}CH\text{-})$, $v_5({=}CH\text{-})$, and $v_6({=}CH\text{-})$ as illustrated in the drawing below:

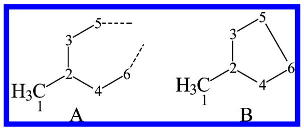


In a situation where the fragment structure is unknown it can be shown that the connectivity set given above contains at least one nonstandard connectivity.

Assume that all the suggested connectivities are standard. Bond analysis of the v_1-v_4 vertices provides the following conclusions:

- The ν_1 vertex with connectivities *$C_{11}(v_1\text{-}v_2)$*, *$C_{12}(v_1\text{-}v_3)$*, and *$C_{12}(v_1\text{-}v_4)$* can be connected only with the ν_2 vertex.
- The ν_2 vertex can be connected with the ν_1, ν_3, and ν_4 vertices only, since only the bond distribution of the ν_1 vertex will be admissible in this case. All the connectivities of the ν_1 vertex will be standard.
- The ν_3 vertex with the *$C_{11}(v_3\text{-}v_5)$* connectivity can only be connected with the ν_2 and ν_5 vertices.
- To allow the possibility of the *$C_{12}(v_2\text{-}v_6)$* connectivity at the ν_2 vertex (see fragment A in the drawing below), the ν_4 vertex can be connected with the ν_2 and ν_6 vertices only.

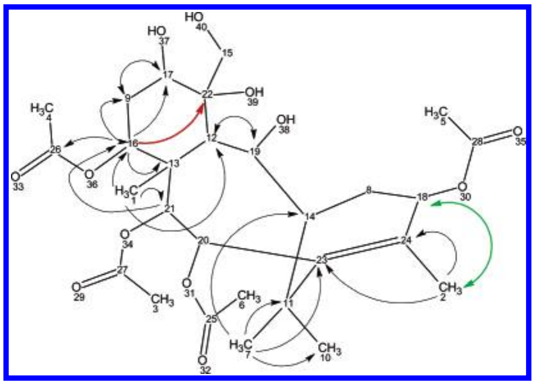
Analyze the bonds at the ν_5 vertex. To provide an accurate description of the bond distribution at the ν_5 vertex it must be connected to the ν_3 and ν_6 vertices since the *$C_{12}(v_4\text{-}v_5)$* connectivity will not correspond to the obtained fragment for all the other bond distributions of the ν_5 vertex. As a result of vertex bond analysis fragment B is obtained as illustrated below:



All the vertices of the given fragment already have the necessary number of adjacent vertices. The fragment vertices cannot therefore be connected with the remaining vertices of the structure. Consequently, a contradiction is identified during the bond analysis of the ν_5 vertex, i.e., the given set of connectivities contains nonstandard members.

Consider the operating sequence for the determination and removal of nonstandard connectivities using an example for which the 2D NMR data were borrowed from the literature.[14] The solution for this problem will be given below.

Example 5. Comparison of the experimental data with the true molecular structure supplied with assigned ^{13}C NMR chemical shifts showed that there were two nonstandard connectivities. These are the COSY formal connectivity *$C_{11}(2-18)$* marked with the bold green two-sided arrow and the HMBC formal connectivity *$C_{12}(16-22)$* marked by a bold red one-sided arrow. From hereon only the vertex numbers are shown to describe the connectivities.



The following description outlines how the algorithm reveals the connectivity sets containing nonstandard connectivities. During the early stage of the analysis a search for explicitly nonstandard connectivities is carried out. In this case it showed that nonstandard connectivities were absent. In the second stage, the check for the presence of vertices and vertex pairs having nonstandard connectivities is performed. For this data set, the algorithm revealed that only the $\nu_2(CH_3)$ vertex had nonstandard connectivities. The program arrived at this observation in the following way.

The starting point is that the following three connectivities are given at the ν_2 vertex: *$C_{11}(2-18)$*, *$C_{12}(2-23)$*, and *$C_{12}(2-24)$*. According to the experimental data the $\nu_{18}(CH)$ vertex has $\delta_C = 69.8$ ppm and $\delta_H = 5.65$ ppm. Comparison of these values with the data of the Atom Property Correlation Table shows that the ν_{18} vertex cannot be connected with the three neighboring C-vertices, i.e., $D^C(\nu_{18}) = 2$. Assuming that the *$C_{11}(2-18)$* connectivity is standard, the ν_2 vertex must be connected with the ν_{18} vertex and, consequently, $d^C(\nu_{18}) = 1$ and $D_2^C(\nu_2) = D^C(\nu_{18}) - d^C(\nu_{18}) = 1$. On the other hand, $K_2(\nu_2) = 2$, as two connectivities *$C_{12}(2-23)$* and *$C_{12}(2-24)$* remain at the ν_2 vertex. As a result $K_2(\nu_2) > D_2^C(\nu_2)$ and consequently, there exists at least one nonstandard connectivity in the connectivity set at the ν_2 vertex.

If the user defines that short connectivities, those having only one bond *$C_{11}(\nu_i\ \text{-}\nu_j)$*, should first be lengthened, then the connectivity *$C_{11}(2-18)$* will be replaced by *$C_{12}(2-18)$*. This corresponds to the real relative positions of the ν_2 and ν_{18} vertices in the structure. As a result of connectivity lengthening a renewed set of connectivities is obtained. The search for nonstandard connectivities now continues in the renewed set of connectivities. However, in this example repeated analysis does not identify nonstandard connectivities in the first and second stages, so the presence of the nonstandard connectivity *$C_{12}(16-22)$* remains undetected.

In the third stage, the presence of nonstandard connectivities that prevent the building of structures is checked. Vertices whose connectivities are in contradiction with the connectivities of other vertices are now searched. The analysis starts with the ν_1 vertex, while none of the bonds implied by the formal C_{11} connectivities are set.

As a result of bond analysis of the $\nu_1(CH_3)$ vertex, with connectivities *$C_{12}(1-12)$*, *$C_{12}(1-16)$*, and *$C_{12}(1-21)$*, it is shown that the ν_1 vertex can be connected only with the ν_{11}, ν_{12}, or ν_{13} vertices. From the indicated connectivities belonging to the ν_1 vertex it follows that three C-vertices ν_{12}, ν_{16}, and ν_{21} are present in first and second layers of the ν_1 vertex environment.

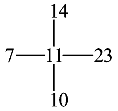
Initially consider the possibility of the ν_1 vertex being connected to the ν_{12}, ν_{16}, and ν_{21} C-vertices. To provide for feasibility of all the connectivities at the ν_1 vertex, the ν_1 C-vertex must be connected with one of the mentioned vertices that, in turn, must be connected with the two remaining C-vertices. Therefore the ν_1 vertex can only be connected to a vertex with bonds to three C-vertices. Only the ν_{12} vertex possesses this property among the trivalent vertices ν_{12}, ν_{16}, and ν_{21}, since the ν_{16} and ν_{21} vertices must be connected with the O-vertex, as follows from the ^{13}C and ^{1}H chemical shifts of the corresponding atoms.

Now suppose that the ν_1 vertex is *not* connected with any of the C-vertices ν_{12}, ν_{16}, and ν_{21}. From this it follows that the ν_{12}, ν_{16}, and ν_{21} vertices have to be situated 2 bonds from the ν_1 vertex, all of them being C-vertices. Thus the ν_1 vertex can be connected only with vertices that, in turn, can be connected with four C-vertices. As with all structural vertices only the tetravalent C-centers may have 4 neighbor vertices. Therefore the ν_1 vertex can be connected only with tetravalent C-vertices that, in turn, are connected with four C-vertices.

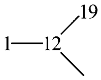
Among all vertices describing the structure, only two C-vertices, ν_{11} and ν_{13}, can be connected with four C-vertices. For instance, the tetravalent C-vertex ν_{22} cannot be connected with four C-vertices because its ^{13}C chemical shift $\delta(C_{22}) = 78.2$ ppm indicates that an oxygen atom is a neighbor of this carbon atom. Consequently, the ν_{22} vertex cannot be connected with four C-vertices. Thus, the ν_1 vertex can be connected only with one of the ν_{11}, ν_{12}, and ν_{13} vertices.

To determine which of the possible vertices that v_1 is bound to, we must consider other correlations associated with each of the possibilities. In doing so, we begin with an analysis of the bonds emanating from the ν_7 (CH_3) vertex with connectivities *C*$_{12}$*(7−10)*, *C*$_{12}$*(7−11)*, *C*$_{12}$*(7−14)*, and *C*$_{12}$*(7−23)* which shows that the ν_7 vertex can be connected with only one of the vertices ν_{10}, ν_{11}, ν_{14}, and ν_{23}; in addition, this vertex must be tetravalent. Among the mentioned vertices, only the ν_{11} vertex possesses this property indicating the ν_7 vertex must be connected with the ν_{11} vertex.

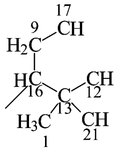
Next, the ν_9 (CH_2) vertex has connectivities *C*$_{11}$*(9−16)* and *C*$_{11}$*(9−17)* and is obviously connected with the ν_{16} and ν_{17} vertices. For the next vertex, ν_{11}, which is already connected with the ν_7 vertex, it must be connected to the ν_{10}, ν_{14}, and ν_{23} vertices to conclude all connectivities belonging to the ν_7 vertex. This means that the ν_{11} vertex cannot be connected with the ν_1 vertex. As a result, the following fragment is constructed:



The ν_{12} (CH) vertex has the connectivity *C*$_{11}$*(12−19)*. This vertex cannot be connected with the ν_1 vertex as this would require the following fragment

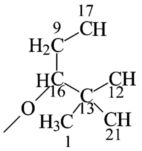


and consequently, both connectivities *C*$_{12}$*(1−16)* and *C*$_{12}$*(1−21)* cannot be standard. It can be concluded that the ν_1 vertex can be connected only with the ν_{13} vertex. Bond analysis of the ν_{13} vertex leads to the conclusion that this vertex has to be connected with the ν_1, ν_{12}, ν_{16}, and ν_{21} vertices to provide a set of connectivities at the ν_1 vertex. Thus, the following fragment is established:



Continuing, the bonds of the ν_{16}(CH) vertex are analyzed. This vertex has the following 7 connectivities: *C*$_{11}$*(16−9)*, *C*$_{12}$*(16−1)*, *C*$_{12}$*(16−13)*, *C*$_{12}$*(16−17)*, *C*$_{12}$*(16−21)*, *C*$_{12}$*(16−22)*, and *C*$_{12}$*(16−26)*.

The drawing shows that the first five connectivities correspond to the assembled fragment. The chemical shifts measured for the ν_{16} vertex $\delta_C = 68.9$ ppm and $\delta_H = 5.59$ ppm indicate that, according to the APCT, this vertex cannot be connected with the three neighboring C-vertices, and consequently the ν_{16} vertex has to be connected with O-vertex, as shown in the representation below:



Counting up the maximum number of vertices that can also be present in the second layer of the ν_{16} vertex environment only the O-vertex has one free valence bond in the first layer of the ν_{16} vertices shown in the drawing above. As a result the second layer of the ν_{16} vertex environment can therefore contain one additional vertex only, i.e. $D_2^C(v_{16}) = 1$. At the same time, from the connectivities *C*$_{12}$*(16−22)* and *C*$_{12}$*(16−26)* it follows that two vertices, ν_{22} and ν_{26}, must also be present in the second layer of the n_{16} vertex environment, i.e., $K_2(v_{16}) = 2$. As a result $K_2(v_2) > D_2^C(v_2)$ and consequently the unique bond distribution of the ν_{16} vertices is *inadmissible*. The contradiction was identified when the bonds of the ν_{16} vertices were analyzed. To remove the contradiction, all connectivities belonging to the ν_{16} vertex, specifically the nonstandard connectivity *C*$_{12}$*(16−22)*, are lengthened by one bond.

Note that if the vertex numbering started from the vertex that currently has the number 16, the analysis would also be started here, and the contradiction would be found during the analysis of other vertices depending on the vertex numbering.

2.5. Implicit Nonstandard Connectivities. Thus far, we have considered only the detection of nonstandard connectivities that prevents final structure generation. However, there are situations (see the examples in section 3) when building the structures corresponding to all the defined connectivities, including the nonstandard ones, is possible. In these cases it is impossible to identify the nonstandard connectivities and this leads to *invalid* solutions. It is possible, even after successfully checking the data for contradictions that nonstandard connectivities remain unnoticed by the algorithm. Their presence can only be identified in indirect ways, including the following: (a) large value(s) for the ^{13}C NMR spectral deviations[11] calculated for the most probable structure(s); (b) inconsistencies between the most probable structure and additional experimental data (for instance, NOESY, ROESY, etc.); (c) comparison of the chemical shifts and multiplicities of the experimental and calculated data of ^{1}H NMR spectra; and (d) by checking the structures using infrared correlation tables (if an IR spectrum is available) and/or interpretation of a mass spectrum, if available. The nonstandard connectivities that do not prevent the structure from being built are termed *implicit* nonstandard connectivities.

As the identification of implicit nonstandard connectivities cannot be guaranteed since any of the given connectivities may be nonstandard, a method of removing nonstandard connectivities was developed and tested that in some cases allows the identification of a valid solution even in this situation.

To remove implicit nonstandard connectivities, the following approach is suggested. Some connectivities are declared as a series to be *suspicious* and are lengthened or eliminated. Each time the structure generation is initiated with a renewed connectivity set. This process is termed *Fuzzy Structure Generation*.

Let n be the total number of connectivities and m be the number of connectivities that are suggested to be nonstandard. In this case it is necessary to consider $\binom{n}{m} = (n!/m!(n - m)!)$ different sets from m connectivities that will be declared as suspicious. The calculation time increases dramatically as the number of tasks resulting in structure generation sharply increases with the rise of the m value. At present we consider those cases when $m \leq 5$. The suspicious connectivity concept has a number of applications. Declaring members of the connectivity sets to be suspicious is sometimes useful for search vertices and pairs of vertices with nonstandard connectivities as well as the direct determination of the presence of nonstandard connectivities. In these cases, the program usually lengthens all connectivities belonging to the vertices selected during data analysis. In so doing both nonstandard and standard connectivities are deliberately lengthened, which correspondingly leads to an increase in the number of structures generated. If only the definite connectivities (related to vertices at which the presence of nonstandard connectivities are revealed, termed *found* vertices) are lengthened, then the number of generated structures will be considerably lower. In some cases the total time of structure generation will decrease despite the repeated initiation of the generation process. This is explained by the fact that most connectivity sets that do not contain connectivities emanating from the found vertices are excluded from consideration. Furthermore, during the vertex bond analysis process and the lengthening of definite connectivities, a greater number of obligatory and forbidden bonds can be established relative to the case when all the connectivities at the *found* vertices are lengthened. In this case the structure generation time with the new connectivity sets can sharply decrease.

It is worth noting that in the presence of vertices with nonstandard connectivities, sets of connectivities declared as suspicious must contain connectivities belonging to all *found* vertices. This indicates that the considered number of suspicious connectivity sets will decrease.

The problem considered in Example 5 with the number of nonstandard connectivities equal to two (m=2) can be solved using the *Fuzzy Structure Generation*. As discussed in section 3.1 (Example 1), a single correct structure was generated in 13 s. For this article all calculations were performed on a PC Celeron operating at 500 MHz, Windows98, RAM 128 Mb.

3. MOLECULAR STRUCTURE ELUCIDATION IN THE PRESENCE OF CONTRADICTIONS IN 2D NMR DATA

Over 150 separate elucidations using 2D NMR literature data as well as raw 2D NMR spectra have been performed in this work. In accordance with the methodology described in earlier reports[6,7,11] spectral data taken from the literature or other sources were entered into the program, and the structures were assumed to be unknown. Later, an attempt was made to elucidate the structures on the assumption that no information existed regarding the presence of contradictions in the 2D NMR data. For one-third of all problems analyzed, about 50 tasks, this subset was used to perform experiments to determine and remove the contradictions. This was feasible since the 2D NMR data related to these problems certainly contained COSY and/or HMBC connectivities of nonstandard length. For convenience this term “connectivities of nonstandard length” is used to denote spin−spin couplings that are present over distances longer than those set as defaults in the system options.

In this section the results of this work will be analyzed. It was shown that the program was capable of determining the *presence* of connectivities of nonstandard length in 90% of all cases using the MCD checking procedure described in section 2. These results are very encouraging since experimental methods guaranteeing the precise determination of COSY and HMBC connectivity lengths are not available. Knowledge of the presence of contradictions in 2D NMR data can give the investigator valuable information that can determine the strategy of structure elucidation with these data. An invalid suggestion regarding the presence of nonstandard connectivities may appear only if there are contradictions between the atom properties and the Atom Property Correlation Table (APCT) used during the data analysis. The validity of the statement can be checked by repeated data verification with the APCT disabled. Such cases actually aid in the development and correction of the correlation tables. A false statement can also appear in rare cases when an unknown under study contains a pair of bonded hetetroatoms. The absence of such atomic pairs is set in the program options as the default. In these cases a message regarding the “existence” of contradictions can help the chemist to reveal the presence of bonded heteroatoms. For instance, it allowed the detection of the O−O group from only the published shift data, during the elucidation of *mycaperoxide H*.[15] With this in mind there were no cases of incorrect detection of nonstandard connectivities when solving ca. 100 tasks where contradictions in 2D NMR data were not present.

For about 50% of cases studied, the program not only identified the contradictions in the data correctly but also was able to successfully remove them automatically to allow determination of the correct structure. Examples were encountered where the program resolved contradictions caused by the presence of a large number of nonstandard connectivities (up to 8).

In six of 50 cases the program was unable to determine the presence of connectivities of nonstandard length. In five of those six cases the 2D NMR data contained only one nonstandard HMBC connectivity. This occurrence can be explained by the fact that if there are only one or two HMBC nonstandard connectivities in the data, the atoms in the structure may be arranged so that their arrangement complies with the standard length of all connectivities. If the number of nonstandard connectivities is large, such an arrangement of atoms is unlikely. The presence of implicit nonstandard connectivities can become apparent as a result of structure generation and their subsequent filtration with the use of spectral libraries: if all the generated structures obviously contradict the spectral data, the program generates an empty results file. Indirect evidence of the possibility that contradictions were not detected may be an empty results file or large values, more than 5.5 ppm, of the chemical shift deviation, d_A, calculated for the first ranked structure.[7,11] Investigations have shown that nonstandard connectivities were detected by both direct and indirect methods for 95% of the analyzed tasks containing contradictory data. If there are reasons to assume that the program did not detect contradictions in the initial data it would be highly likely that the problem could be solved with the use of "fuzzy" structure generation as described in section 2.

Since it is possible that 2D NMR data could contain implicit nonstandard connectivities, the most probable structure generally requires additional verification by independent methods. In particular, if the structure is generated after automatic contradiction removal, it is still desirable to check for the presence of nonstandard connectivities. For this purpose a preview mode in the Structure View window exists to view the connectivities. If such connectivities are found, they can be verified with appropriate experimental parameter optimization to probe the values of the spin couplings.[16,17] It has been shown[18] that incorrect structures are often rejected on the basis of predicted chemical shifts and multiplets in the ^{1}H NMR spectrum (see the examples below).

Experience gained in this work shows it is not always possible to find nonstandard connectivities and to automatically resolve the contradictions in 2D NMR data sets. These investigations show that in practice the following difficult situations may typically arise:

1. The program detects the presence of nonstandard connectivities and makes an attempt to remove the contradictions in the data but then reports that contradictions cannot be removed. Sometimes fuzzy structure generation can help to solve the problem. Generally additional experiments are required in an effort to detect nonstandard connectivities.

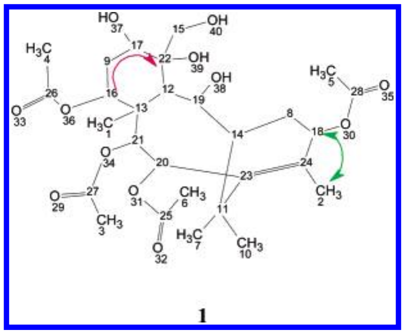
2. The program fails to detect nonstandard connectivities and displays a message informing the researcher about the absence of contradictions. In this case, structure generation is initiated. The following outcomes are possible: (a) no structure is generated and (b) the wrong structure(s) is generated.

3. The program detects the presence of nonstandard connectivities, makes an attempt to remove the contradictions in the data, and displays a message that the contradictions were removed, though, in fact, contradictions remain. This is due to the fact that not all nonstandard connectivities are lengthened. Possible consequences are similar to those listed for point 2 above.

It should be obvious that the most dangerous situations, when incorrect solutions are offered, can occur for points 2 and 3 and will be considered later. Even in the case when the program is not able to remove detected contradictions, case 1, the fact that contradictions are identified is still extremely important.

In this section attention will be focused on examples that demonstrate how a correct result can be obtained using the *StrucEluc* system in those cases where the program does not remove all contradictions. First we will cite several examples of the successful automatic removal of contradictions from 2D NMR data.

3.1. Examples of Successful Contradictions Removal. *Example 1.* In the work of Shen et al.[14] a natural product (**1**), related to the class of *taxoids*, was identified. For this compound the authors presented ^{1}H and ^{13}C NMR data as well as COSY and HMBC correlations in the form of peak tables. These data were used as inputs to the *StrucEluc* system. The given molecular formula was $C_{28}H_{42}O_{12}$ with MW = 570 and the number of skeletal atoms being n_{sk} = 40:



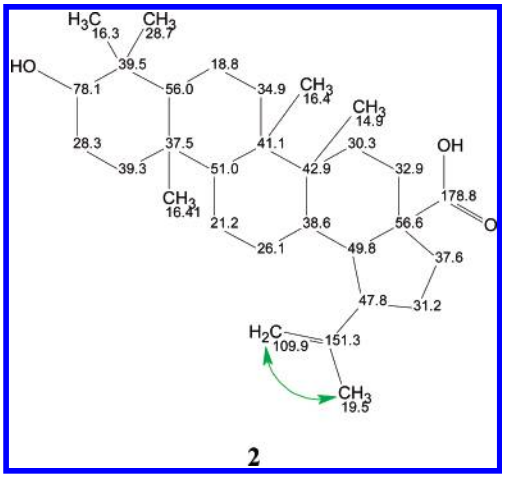
1

The molecular formula and the cross-peaks from the COSY, HMQC, and HMBC spectra were fed as inputs into the program. These formed tables containing 22 HMQC direct correlations, 21 COSY cross-peaks, and 47 HMBC correlations. As mentioned above, the default settings of the program are such that the COSY and HMBC spectra distances between intervening spins are within 2−3 bonds. This means that the default connectivity lengths are one *skeletal* bond for COSY correlations and from one to two− for HMBC experiments (see section 2.1). Comparison of the cross-peaks with the molecular structure, a process that can only be performed after the structure of an unknown compound is determined, showed that one long-range correlation $^4J_{HH}$ (H2−H18) exists in the COSY spectrum and one exists in the HMBC $^4J_{CH}$ (H16−C22). These long-range COSY and HMBC connectivities are shown on the structure **1** with the corresponding two-sided green and one-sided red arrows. These correlations contradict the default options, and consequently the correct structure will not be generated. If additional 2D experiments are not performed, it is impossible to determine in advance which correlations may be nonstandard. Information regarding experiments utilized to check the correlation lengths was not available in ref 14.

The MCD was automatically created by the program using the peak tables as a basis. When the *Check MCD for Contradictions* command was applied, the program displayed the appropriate message that some correlations contradicted the standard values. The algorithms attempted to resolve the contradictions in an iterative mode. The procedure required two iterations and 28 s to successfully resolve the contradictions. The process for detection and removal of contradictions was described earlier in section 2.4 (Example 5). An important feature of an advanced expert system is the ability to create an audit trail of the actions and decision-making process. This is particularly useful when a scientist needs to review, repeat, or "tweak" the elucidation. To support this function a "contradiction resolution protocol" is generated by the *StrucEluc* system. In addition, connectivities that were edited by the program appear in the MCD marked by colors corresponding to the new connectivity lengths. For this example, one C_{11} connectivity and all 1,2-connectivities belonging to the carbon 16 as well as one C_{11} connectivity at the carbon 2 were lengthened by 1 bond. The protocol indicates the results of each iterative step of the program and produces a complete list of the new lengths of connectivities introduced.

Based on the MCD with the corrected connectivity lengths, the generation and filtration of structures using spectral libraries and a structural BADLIST (Bredt's rule, unstable fragments) were performed. No changes were made to the default atom properties. The program generated 1 molecule identical in structure to structure **1** in about 3 s (This is denoted as $k = 1$, $t_g = 3$ s.). The deviation values ($d_A = 1.34$ ppm, $d_F = 3.09$ ppm, $d_H = 0.21$ ppm) are approppropriate for a correct structure determination. Here the subscripts A and F are used for denoting different types of ^{13}C NMR spectral prediction, correspondingly known as the accurate and fast methods, while H defines ^{1}H NMR spectral prediction.[11] This task was also successfully solved using the *Fuzzy Structure Generation* mode. Assuming that the total number of nonstandard connectivities in both types of spectra, COSY, and HMBC did not exceed 2, the correct structure was generated in 13 s.

Example 2. The program was applied to the structure generation of *betulinic acid* (**2**)[19] $C_{30}H_{48}O_3$, whose COSY spectrum contained only one nonstandard connectivity $C_{22}(109.9-19.5)$. Note that here the COSY connectivity is represented via the chemical shifts of carbon atoms to which the intervening hydrogen atoms are attached.

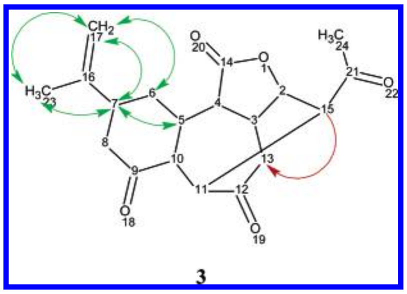


2

Contradictions in the 2D NMR data were searched and the one contradiction identified was removed by the program, and, as a result, only one correct structure was generated in 1 s. A similar result was obtained in the *Fuzzy Structure Generation* mode with m equaling 1 and 2.

In the examples described above, the program successfully overcame contradictions in the presence of one and two nonstandard connectivities. Note that the number of nonstandard connectivities may be much larger, up to 10 or more.

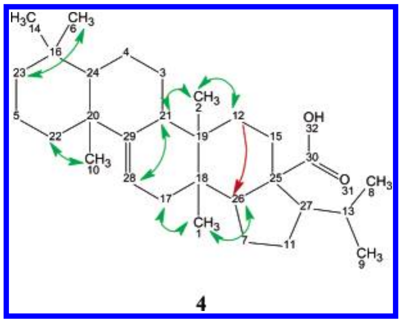
Example 3



3

In case of the analysis of *horiolide* (**3**),[20] whose 2D NMR spectra contain five nonstandard connectivities $C_{22}(17-23)$, $C_{22}(7-23)$, $C_{22}(7-17)$, $C_{22}(7-5)$, $C_{33}(6-17)$ in the COSY data and 1 nonstandard $C_{33}(15-13)$ connectivity in the HMBC spectrum (remember that they are accepted by the program as having lengths common for COSY and HMBC) as indicated with arrows on the structure, the program found and automatically removed all contradictions. As a result, the only structure to be generated was the correct structure, and $t_{\mathrm{g}} = 1$ s ($d_{\mathrm{A}} = 3$ ppm).

Example 4. The 2D NMR spectra of *triterpene E*, $C_{30}H_{48}O_2$ (**4**), which was isolated and characterized by Reynolds et al.,[21] contained a total of 8 nonstandard connectivities: 7 in the COSY spectrum and 1 in the HMBC.



4

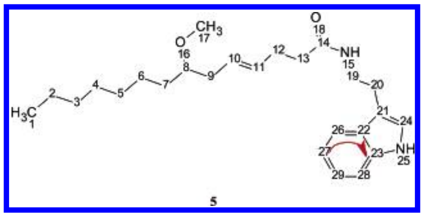
All contradictions were resolved automatically in 3 iterations, and one, correct structure, was generated in 7 s.

3.2. Resolving Contradictions in Complicated Cases. *Example 1.* It is possible that the program will not detect the presence of connectivities of nonstandard length. In these cases when the program fails, for a data set containing HMBC and COSY connectivities the COSY data are often excluded from consideration, and an attempt is made to solve the problem from the HMBC data only. In so doing, the HMBC data are analyzed again, as these data may contain no contradictions or the contradictions may be resolvable and, consequently, a solution can be found. In the previous example if the COSY data of *triterpene E* are excluded from consideration, only one nonstandard HMBC connectivity $C_{33}(12-26)$ is left. The new MCD was checked, but the program did not detect the one remaining contradiction.

An attempt to generate structures, both using and ignoring the APCT, resulted in empty output files, characteristic of the presence of contradictions in data. If the contradictions are not found at the checking stage, but structure generation indicates they are present as noted above then this may be indirect evidence that there are a small number (1 or 2) of nonstandard connectivities.

Fuzzy Structure Generation was attempted without imposing any constraints. The *m* parameter was set to 1. The result was $k = 120$, $t_g = 4$ min, $d_A(1) = 1.81$ ppm, $\Delta_{(2-1)} = d_A(2)$-$d_A(1) = 0.22$ ppm, $r_A = r_H = 1$, $r_F = 2$. Here $\Delta_{(2-1)}$ is the difference in deviation values between the second and first structures in the structural output file ranked in ascending order of $d_A(i)$ values (i=1÷k); r_A, r_F, and r_H are the ranks assigned to the *correct* structure by the different methods of spectral prediction.[11] The generation was repeated after manual correction of the nonstandard connectivity (12−26), which resulted in $k = 81$, $t_g = 25$ s, $r_A = 1$. In this case, the fuzzy mode using $m = 1$ was 9 times slower, and the number of generated structure was 1.5 times greater.

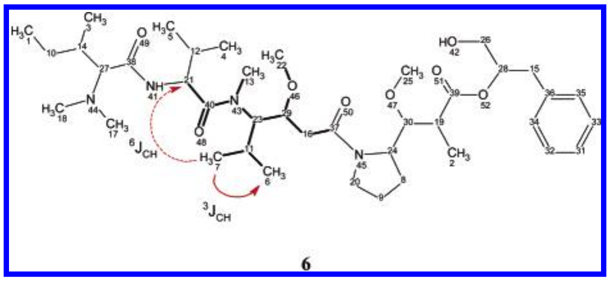
Example 2. When the program is not able to detect connectivities of nonstandard length, the structure generation results in either an empty or invalid result file. In the latter case the large $d_A(1)$ deviations usually indicate that the results are incorrect. When the structure of *hermitamide* (**5**) was elucidated from the published[22] HMBC data, the program did not reveal the presence of the *C*$_{33}$*(27−23)* connectivity, and as a result, two structures were generated in 12 s.



5

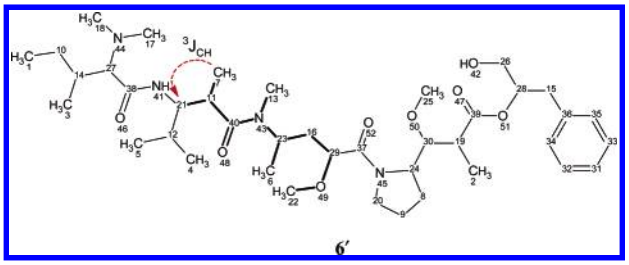
The deviations $d_A(1) = 5.75$ ppm, $d_F(1) = 5.67$ ppm are large enough that the obtained solution is suspicious. For this reason, the generation was repeated in the fuzzy mode with $m = 1$. The result was $k = 394 \rightarrow 42$, $t_g = 6$ min 30 s, $d_A(1) = 0.90$ ppm, $d_F(1) = 1.60$ ppm, and $d_H = 0.09$ ppm (the arrow indicates how the size of the output file is reduced as a result of removing identical structures). The resulting preferred structure and the chemical shifts assignment of the ^{13}C NMR spectrum coincided with those given by the authors.[22]

Example 3. This example demonstrates the ability of the program to verify the validity of the 2D NMR data and the results of structure elucidation reported in the literature. Horgen et al.[23] reported the structure of a novel natural product (**6**), isolated and identified as *maleυamide-D* on the basis of spectroscopic data:



6

The high-resolution ESI mass spectrum of (**6**) gave a molecular ion *m/z* 733.5114, suggesting the molecular formula $C_{40}H_{68}N_4O_8$. The ^{1}H, ^{13}C, COSY, HMQC, and HMBC NMR spectra were recorded and presented in a peak table without any indication of the presence of correlations of nonstandard lengths. The information captured in this table was used to determine the structure using the *StrucEluc* program. Checking for contradictions did not reveal the presence of any connectivities of nonstandard length, and structure generation was completed in the automated mode without any limitations imposed. The results were $k = 40 \rightarrow 10$, $t_g = 4$ s. Ranking the obtained structures suggested the following structure as the most likely:



6′

This solution does not formally allow us to question the validity of the generated structure, since the deviations of the three different types of calculated spectra from the experimental ones are minimal and the absolute values of the deviations are well within the statistical limits: $d_A(1) = 1.452$, $d_F(1) = 2.725$, and $d_H(1) = 0.240$ ppm. Comparison of the resultant structure with that obtained by the authors shows clearly that they differ in the functional groups highlighted in the central part of the molecule. The differing fragments are marked with bold lines in structures (**6** and **6′**). It can be seen that the differences are fairly subtle. The calculation of the Tanimoto similarity match factor between the structures gave a value of 0.96. The validity of the generated structure is doubtful when the difference between the experimental and calculated ^{1}H NMR spectra is considered. Using the ACD/HNMR prediction component of the program, comparison of the spectra indicated differences in the region 0.5−1.5 ppm. The predicted doublet of the methyl group (C6) at $\delta_H = 1.4$ ppm is absent in the experimental spectrum. There are two suggestions that will explain this observation: either the authors elucidated the wrong structure, or the 2D NMR data contains nonstandard connectivities that were not revealed by the program.

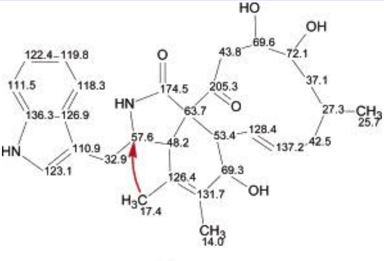
Comparison of the HMBC correlations with structure **6** identified the presence of the $^6J_{CH}$ correlation *C*$_{55}$*(7−21)* as indicated with the dotted arrow in structure **6**. However, the search for contradictions in the 2D NMR data produced no results. This indicates that the program identified an arrangement(s) of atoms under which the corresponding topological distances between the carbon nuclei, represented by their chemical shifts, are the same as the default values. In other words, the program found that it was possible to generate a chemical structure that meets both the constraints and the atom properties assigned by the program with the help of the APCT. As can be seen in structure **6′** atoms 7 and 21 are separated by two C−C bonds that corresponds to the standard HMBC connectivity.

This observation was communicated to the authors[23] with the result of a misprint in the published tables being identified. The experimental data actually displayed a HMBC connectivity *C*$_{12}$*(6−7)* of standard length (see structure **6**). After the initial data were corrected, the generation was repeated with the following results: $k = 156 \rightarrow 44$, $t_g = 4$ s, $d_A(1) = 0.721$ ppm, $d_F(1) = 2.148$ ppm, $d_H(1) = 0.170$ ppm,

$r_A = r_F = r_H = 1$. The most highly ranked structure that was generated, coincided with that reported by the authors, structure **6**. The deviation values $d_A(1)$, $d_F(1)$, and $d_H(1)$ were also less than those calculated initially for the wrong structure. In fact, the $d_A(1)$ value was decreased by a factor of 2. Note that the HMBC correlation between CH_3-6 and CH_3-7 would generally not be possible in the structure **6′**, or, if a correlation were observed, it would be weak.

The conclusion of this study is that if the program is not able to reveal nonstandard connectivities, it is possible that the program can generate a structure similar to the correct one. As noted above it is always desirable to verify the solutions suggested by the software for their correctness and stability. It can be very effective to check the structural validity by comparing the experimental and predicted ^{1}H NMR spectral patterns visually. The stability of the result can be verified by lengthening some connectivities to which the weakest peaks of the 2D NMR spectra correspond, by loosening the limitations on the atom properties, or by using the fragments from the knowledge base. It may also be of value to use the *ACD/Structure Elucidator* system in the mode which uses the 1D ^{13}C NMR spectrum only to perform structure elucidation.[11,18] If the first structures of the ranked result files from several generation processes are different, the most probable structure is the one with the minimal $d_A(1)$ value.

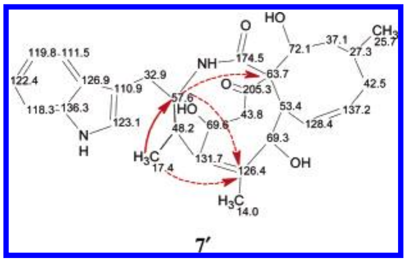
Example 4. The next example indicates how the combined usage of different system tools can help in solving challenging problems. In this case,[24] the HMQC, COSY, and CIGAR−HMBC spectra were used to elucidate *cytochalasin* (**7**), $C_{29}H_{36}N_2O_5$. Comparison of the table of 2D NMR peaks cited in the original publication with the structure suggested by the authors revealed the presence of one nonstandard connectivity *C_{33}(17.4−57.6)* in the HMBC data.



7

Solution of this problem proceeded using an iterative method with the use of various system tools.

Iteration 1. Checking the MCD revealed the presence of contradictions. Repetitive checking with the automated correction of the connectivities gave a message informing that the contradictions had been removed. Visual analysis of the corrected MCD showed that the connectivity (17.4−57.6) had not been lengthened, but other correct connectivities (126.4−14.0), (126.4−17.4), (126.4−57.6), (63.7−57.6), and (63.7−NH) had been lengthened. In reality the program had not managed to remove the contradictions in the data, and the user may be oblivious to this fact. The program made this conclusion because it eventually became possible to generate structures from the "corrected" MCD. So, with the assumption that the program correctly adjusted the MCD, structures were generated with only one limitation, no four-membered cycles should be present, with the following results: $k = 3486$, $t_g = 32$ min. To identify the most probable structure without calculating the ^{13}C NMR spectra of ∼3500 structures, the system knowledge base was searched for fragments using the method described previously.[8,11] The program selected 4427 fragments. Searching for functional groups in the selected fragments showed that around 500 fragments contained a benzene ring. Experience has shown that in this case, it would be safe to assume that there is a benzene ring in the molecule under analysis. Filtering the resulting file using a GOODLIST containing a benzene ring reduced the file to 48 structures. After rank ordering following NMR spectral prediction the most preferable one was structure **7′** characterized by a deviation of $d_A(1) = 3.40$ ppm. The structure with the assigned experimental chemical shifts is shown below:



7′

The atom chemical shifts, topological peculiarities of the structure, and $d_A(1)$ values are acceptable, and there was no basis to reject the structure. The length of the connectivity 17.4−57.6 in this structure, marked with the solid arrow in the drawing, is also "standard". Investigation of the structure in the *Connectivity Viewing* mode showed that it had three HMBC $^4J_{CH}$ connectivities as indicated with the dashed arrows. If the experimental spectra were available to the authors a check for found nonstandard connectivities using the potential relationship with the intensities of the corresponding peaks of HMBC spectrum would be made or alternatively additional experimental data would be generated.

In the absence of such opportunities the ^{1}H NMR spectrum was calculated and compared with the experimental one. A considerable difference in the region 1−2 ppm was observed: instead of a singlet $\delta_H = 1.52$ ppm ($\delta_C = 17.4$ ppm) for the methyl group observed in the experimental spectrum, in the calculated spectrum a doublet at $\delta_H = 1$ ppm was assigned to this methyl group. A conclusion was made that the generated structure was incorrect and *Fuzzy Structure Generation* may be an appropriate path forward from the initial, not corrected, MCD.

Iteration 2. *Fuzzy Structure Generation* was performed using $m = 1$ and a restriction on the presence of four-membered ring cycles. The result was $k = 4552$, $t_g = 36$ min. As with the previous iteration, to reduce the size of the resulting file, the structures were filtered using a GOODLIST containing a single benzene ring. The result gave $k = 4552 \rightarrow 112$, $d_A(1) = 1.94$ ppm, $d_F(1) = 3.09$ ppm, $d_H(1) = 0.30$ ppm, and $r_A = r_F = 1$ and the most preferable generated structure corresponded to that reported in the publication.[24]

Iteration 3. Comparison of the ^{13}C NMR subspectra of the first fragments from the beginning of the list indicated that they were quite consistent with the spectrum of the analyzed molecule. A solution was sought using these found fragments. The following calculation parameters (see definitions[11]) were selected: $l = 100$, $E = 2.5$ ppm, $q = 0$, $n_f = 1$ which resulted in $n_{(MCD)} = 11$. Fuzzy structure generation using $m = 1$ gave $k = 16 \rightarrow 12$, $t_g = 8$ s, $r_A = r_F = 1$. *Fuzzy Structure Generation* using fragments elucidated the structure 270 times faster.

4. CONCLUSION

This work has analyzed the various situations which may appear when there are COSY and HMBC connectivities whose length is greater than the values set as default in the *ACD/Structure Elucidator* system (1 skeletal bond for COSY and from 1 to 2 skeletal bonds for HMBC). From this a series of classifications has been defined. "Nonstandard" connectivities are classified as those connectivities that resulted in an incorrect solution or gave no solution resulting in an empty output file. A series of nonstandard connectivity classifications has been identified; sets of connectivities containing nonstandard connectivities at one or several vertices; sets of connectivities of one vertex contradicting the connectivities of other vertices; and implicit nonstandard connectivities that caused no contradictions but caused the wrong structure generation. Each case has been examined using practical examples with an analysis of the correspondence between subgraphs built from the vertices (skeletal atoms) of the considered molecule and the connectivities connected to the vertices of graphs.

The algorithms developed iteratively as a result of this work have been implemented into the *StrucEluc* system and tested on 50 tasks whose 2D NMR data contained correlations with lengths exceeding the default values in the system. The presence of nonstandard connectivities was automatically revealed by the program in 90% of the cases studied. In almost 50% of cases the program removed the contradictions by lengthening the connectivities at corresponding vertices or vertex pairs, which led to successful problem solution. It has been shown that the program is capable of resolving contradictions and solving problems in cases where the number of nonstandard connectivities in the data set can be large, up to 8 connectivities to date.

A series of typically challenging situations has been identified that may appear during the elucidation of a molecular structure from 2D NMR data containing contradictions. Complications appear when the presence of contradictions in the initial data is not recognized by the program. These are the so-called implicit nonstandard connectivities. Also complications arise when a false message is generated that suggests that the contradictions have been removed. Nonstandard connectivities could remain undetected mainly in those cases where the data contain one or two nonstandard HMBC connectivities. The probability of detecting the contradictions increases with a concomitant increase in the number of carbon atoms for which the properties are strictly set, for example the presence of C−X bonds where X is explicitly specified as a carbon or a heteroatom eases the search through the MCD for the presence of contradictions. Methods of structure elucidation depending on the various situations that can be encountered have been outlined in this article. In those cases where the number of nonstandard connectivities did not exceed five, a solution could be found with the help of *Fuzzy Structure Generation* in a reasonable time.

The authors are continuing to work on improving the efficiency of the data analysis processes and specifically on the removal of contradictions. Experience suggests that this can be achieved by further analysis of the atomic environments in the MCDs (Molecular Connectivity Diagram). At present a publication is in preparation which describes the results of further development of the Fuzzy Structure Generation algorithm. This allows the solution of structural problems even in the presence of 15−20 nonstandard connectivities (^4J or ^5J). These results were obtained during the review period for this manuscript.

REFERENCES AND NOTES

(1) Munk, M. E. Computer-based structure determination: then and now. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 997−1009

(2) Nuzillard, J.-M.; Massiot, G. Logic for structure determination. *Tetrahedron* **1991**, *47*, 3655−3664.

(3) Peng, C.; Yuan, S.; Zheng, C.; Hui, Y. Efficient application of 2D NMR correlation information in computer-assisted structure-elucidation of complex natural products. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 805−813.

(4) Lindel, T.; Junker, J.; Koeck, M. 2D-NMR-guided constitutional analysis of organic compounds employing the computer program COCON. *Eur. J. Org. Chem.* **1999**, *573*, 3−577.

(5) Steinbeck, C. SENECA: a platform-independent, distributed, and parallel system for computer-assisted structure elucidation in organic chemistry. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1500−1507.

(6) Blinov, K. A.; Elyashberg, M. E.; Molodtsov, S. G.; Williams, A. J.; Martirosian, E. R. An expert system for automated structure elucidation utilizing ^{1}H-^{1}H, ^{13}C-^{1}H and ^{15}N-^{1}H 2D NMR correlations. *Fresenius J. Anal. Chem.* **2001**, *369*, 709−714.

(7) Elyashberg, M. E.; Blinov, K. A.; Williams, A. J.; Molodtsov, S. G.; Martirosian, E. R. Application of a new expert system for the structure elucidation of natural products from their 1D and 2D NMR data. *J. Nat. Prod.* **2002**, 65, 693−703.

(8) Blinov, K. A.; Carlson, D.; Elyashberg, M. E.; Martin G. E.; Martirosian, E. R.; Molodtsov, S. G.; Williams, A. J. Computer assisted structure elucidation of natural products with limited 2D NMR data: application of the *StrucEluc* system. *Magn. Reson. Chem.* **2003**, *41*, 359−372.

(9) Martin, G. E.; Hadden, C. E.; Kaluzny, B. D.; Russell, D. J.; Stiemsma, B. A.; Thamann, T. J.; Crouch, R. C.; Blinov, K. A.; Elyashberg, M. E.; Williams, A. J.; Martirosian, E. R.; Schiff, P. L., Jr. Identification of degradants of a complex alkaloid using NMR cryoprobe technology and ACD/Structure Elucidator. *J. Het. Chem.* **2002**, *39*, 1241−1250.

(10) Blinov, K. A.; Elyashberg, M. E.; Martirosian, E. R.; Molodtsov, S. G.; Williams, A. J.; Sharaf, M. M. H.; Schiff, P. L., Jr.; Crouch, R. C.; Martin, G. E.; Hadden, C. E.; Guido J. E. Quindolinocryptotackieine: the elucidation of a novel indoloquinoline alkaloid structure through the use of computer-assisted structure elucidation and 2D-NMR. *Magn. Reson. Chem.* **2003**, *41*, 577−584.

(11) Elyashberg, M. E.; Blinov, K. A.; Williams, A. J.; Molodtsov, S. G.; Martin, G. E.; Martirosian E. R. *Structure Elucidator*: a versatile expert system for molecular structure elucidation from 1D and 2D NMR data and molecular fragments. *J Chem Inf. Comput. Sci*. **2004**, *44*, 771−792.

(12) Krishnamurthy, V. V.; Russell, D. J.; Hadden, C. E.; Martin, G. E. ^{2}J,^{3}J-HMBC: a new long-range heteronuclear shift correlation technique capable of differentiating $^2J_{CH}$ from $^3J_{CH}$ correlations to protonated carbons. *J. Magn. Reson.* **2000**, *146*, 232−239.

(13) Sprang, T.; Bigler, P. A new technique for differentiating between ^{2}J(C,H) and $^{3/4}$J(C,H) connectivities. *Magn. Reson. Chem*. **2003**, *41*, 177−182.

(14) Shen, Y.-C.; Lo, K.-L.; Chen, C.-Y.; Kuo, Y.-H.; Hung, M.-C. New taxanes with an opened oxetane ring from the roots of *Taxus mairei*. *J. Nat. Prod*. **2000**, *63*, 720−722.

(15) Phuwapraisirisan, P.; Matsunaga, S.; Fusetani, N.; Chaitanawisuti, N. Mycaperoxide H, A new cytotoxic norsesterterpene peroxide from a Thai marine sponge *Mycale. J. Nat. Prod*. **2003**, *66*, 289−291.

(16) Marquez, B. L.; Gerwick, W. H.; Williamson, R. T. Survey of NMR experiments for the determination of ^{n}J(C, H) heteronuclear coupling constants in small molecules. *Magn. Reson. Chem*. **2001**, *39*, 499−530.

(17) Martin, G. E. *Exploitation of Long-range Heteronuclear Coupling Constants −New Techniques for Long-Range Heteronuclear Shift Correlation Spectroscopy*; *Ann. Rep. NMR Spectrsosc*., G. A. Webb, Ed., Academic Press: New York, 2002; Vol. 46, pp 37−100.

(18) Elyashberg, M. E.; Blinov, K. A.; Martirosian, E. R. A new approach to computer-aided molecular structure elucidation: the expert system *Structure Elucidator*. *Autom. Inf. Manage.* **1999**, *34*, 15−30.

(19) Peng, C.; Bodenhausen, G.; Qiu, S.; Fong, H. H. S.; Farnsworth, N. R.; Yuan, S.; Zheng, C. Computer-assisted structure elucidation: application of CISOC−SES to the resonance assignment and structure generation of betulinic acid. *Magn. Reson. Chem*. **1998**, *36*, 267−278.

(20) Radhika, P.; Rao, P. V. S.; Anjaneyulu, V.; Asolkar, R. N.; Laatsch. Horiolide, a novel norditerpenoid from Indian Ocean soft coral of the genus *Sinularia*. *J. Nat. Prod*. **2002**, *65*, 737−739.

(21) Reynolds, W. F.; McLean, S.; Burke, S. J.; Jacobs, H. Identification and complete 1H and 13C spectral assignments for the triterpene fern-9(11)-en-28-oic acid. *Magn. Reson. Chem.* **2001**, *39*, 757−758.

(22) Tan, L. T.; Okino, T.; Gerwick, W. H. Hermitamides A and B, toxic malyngamide-type natural products from the marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod*. **2000**, *63*, 952−955.

(23) Horgen, F. D.; Kazmierski, E. B.; Westenburg, H. E.; Yoshida, W. Y.; Scheuer, P. J. Malevamide D: isolation and structure determination of an isodolastatin H analogue from the marine cyanobacterium *Symploca hydnoides*. *J. Nat. Prod*. **2002**, *65*, 487−491.

(24) Feng, Y.; Blunt, J. W.; Cole, A. L. J.; Munro, M. H. G. Three novel cytochalasins X, Y, and Z from *Pseudeurotium zonatum*. *J. Nat. Prod*. **2002**, *65*, 1274−1277.

CI049956+