Fuzzy Structure Generation: A New Efficient Tool for Computer-Aided Structure Elucidation (CASE)

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

Advanced Chemistry Development, Moscow Department, 6 Akademik Bakulev Street, Moscow 117513, Russian Federation, Advanced Chemistry Development, Inc., 110 Yonge Street, 14th Floor, Toronto, Ontario, Canada M5C 1T4, Novosibirsk Institute of Organic Chemistry, Siberian Branch of Russian Academy of Science, Lavrentiev Avenue 9, Novosibirsk 630090, Russian Federation, and Schering-Plough Corporation, Pharmaceutical Sciences, Rapid Structure Characterization Laboratory, Summit, New Jersey 07901

Received November 21, 2006

Contemporary Computer-Aided Structure Elucidation (CASE) systems are heavily based on the utilization of 2D NMR spectra. The utilization of HMBC/GHMBC and COSY/GCOSY correlations generally assumes that these correlations result from $^{2-3}J_{CH}$ and $^{2-3}J_{HH}$ spin−spin couplings, respectively, and consequently these values are used as the default setting in these systems. Our previous studies[1,2] have shown that about half of the problems studied actually contain some correlations of 4−6 bonds, so-called "nonstandard" correlations. In such cases the initial 2D NMR data are contradictory, and the correct solution is therefore not directly attainable. Unfortunately nonstandard correlations and the number of intervening bonds usually cannot be identified experimentally. In this work we suggest a new approach that we term Fuzzy Structure Generation. This allows the solution of structural problems whose 2D NMR data contain an unknown number of nonstandard correlations having different and unknown lengths. Suggested methods for the application of Fuzzy Structure Generation are described, and their application is illustrated by a series of real-world examples. We conclude that Fuzzy Structure Generation is efficient, and there is no real alternative at present in terms of a universal practical method for the structure elucidation of organic molecules from 2D NMR data.

1. INTRODUCTION

In our previous reports[1,2] we have already discussed the problems that arise when an expert system is used for molecular structure elucidation from 2D NMR data. As a rule, a combination of GHMQC/GHSQC, GHMBC, and COSY spectra make up an experimental data set with the GHMBC and COSY correlations assumed to correspond to $^{2-3}J_{CH}$ and $^{2-3}J_{HH}$ coupling constants, respectively. In our previous work[1,3,4] we defined such correlations as "standard" correlations. For $^{n>3}J_{HH/CH}$ couplings we label these correlations *nonstandard*. The origin and nature of nonstandard correlations (NSCs) are discussed elsewhere in the literature.[5] Unfortunately, there are no reliable and routine experimental methods so far that unambiguously distinguish between standard and nonstandard correlations though there have been significant efforts to address the differentiation of $^{2}J_{CH}$ from $^{3}J_{CH}$ long-range heteronuclear couplings (note that although the couplings are listed here as $^{n}J_{CH}$ correlations, this statement can also, albeit it less frequently, apply to $^{n}J_{NH}$ correlations as well).[5]

It is generally believed by others in this field of study that correlations of nonstandard length are observed fairly rarely.[6−8] Results obtained in our work[2] contradict this opinion. Previously we investigated the solutions of more than 250 problems whereby the expert system Structure Elucidator (StrucEluc)[3,4] was applied to the structural identification of complex natural products from 2D NMR and MS spectra. The studies indicated that almost half of the problems (45%) contained nonstandard correlations in the 2D NMR data. Nonstandard long-range heteronuclear correlations are even more common when some of the accordion-optimized long-range heteronuclear shift correlation experiments are employed.[3] These additional HMBC connectivities can give additional constraints and simplify the process of solving the problem. However, at the same time new NSCs can appear so a method to address these NSCs is necessary. This is the benefit of the fuzzy generation approach discussed in this paper. If the spectroscopist can indicate the connectivities associated with the NSCs, then this information will of course be useful. Meanwhile, expert systems are usually optimized to structure elucidation assuming a set of correlations of common (standard) length. The presence of nonstandard correlations within one or more 2D NMR data sets generally produces a result that is inconsistent with the real structure. For instance, the observation of a COSY correlation between hydrogen atoms attached to two carbon atoms C-1 and C-2 suggests that these atoms are connected by a carbon−carbon bond. In the case when a given correlation is of nonstandard length, the distance between the C-1 and C-2 atoms is actually of two or more bonds in the real structure. Prior to actually establishing the

* Corresponding author e-mail: tony@acdlabs.com.
† Advanced Chemistry Development, Inc.
‡ Siberian Branch of Russian Academy of Science.
§ Advanced Chemistry Development, Moscow Department.
|| Schering-Plough Corporation, Pharmaceutical Sciences.

10.1021/ci600528g CCC: $37.00
Published on Web 03/27/2007

structure of the unknown molecule, the presence or absence of nonstandard correlations as well their number and real lengths remains unknown. Therefore, the problem adds up to molecular structure elucidation from spectrum-structural information that is not only fuzzy by nature ($^{2-3}J_{CH}$ in HMBC) but can also be both contradictory and uncertain (i.e., the number of nonstandard connectivities and their lengths are unknown).

In a previous report[1] we suggested approaches for solving problems in the presence of correlations of nonstandard lengths. The program first attempts to find whether nonstandard correlations are present in the spectral data. The method is based on the logical analysis of a full set of connectivities including HMBC and COSY data, and most frequently the algorithm is capable of detecting skeletal atoms that are involved in connectivities of a nonstandard length. If nonstandard connectivities are associated with particular atoms, then the program automatically lengthens all connectivities emanating from such atoms by one bond and attempts to generate structures. It is important to emphasize that not all connectivities are lengthened, only the ones identified by the algorithm as being nonstandard. As a result the program is capable of generating structural solutions consistent with the data within a reasonable period of time even if there are a large number of nonstandard correlations.

In those cases when correlations are present in the 2D NMR data with ^nJ where $n > 4$, this method, unfortunately, does not work. The augmentation of the path between two intervening nuclei by one bond obviously cannot lead to the generation of a correct structure. Moreover, even in those cases when $n \leq 4$ the algorithm gives no guarantee that all nonstandard correlations will be found and corrected due to a lack of constraints that are to be logically analyzed. For example, the greater the number of carbon atoms with accurately defined properties (the type of hybridization and different heteroatom neighborhoods) and/or the higher the total number of available 2D NMR connectivities, the higher the probability of successfully performing logical analysis to arrive at the correct structure. In contrast, severely proton-deficient molecules can be among the most challenging. Obviously, the problem becomes more computationally complicated as the complexity of a molecule increases. Our experience has shown that the number of nonstandard correlations contained within the 2D NMR data associated with a molecule, *m*, can be rather large—up to about 20 correlations. At the same time the augmentation of standard correlation lengths, *a*, could be 1−3. As an example of such situations several structures taken from literature[9−12] are used to demonstrate those examples with a large number of nonstandard correlations including ^5J and ^6J coupling constants (see Figure 1).

To overcome the described shortcomings a computational approach is suggested that we have defined as *Fuzzy Structure Generation (FSG)*.[1] During the process of fuzzy generation the number of nonstandard connectivities is restricted to a parameter *m*, and their lengths can be augmented by a number of bonds equaling *a*. A strategy of determining the values of the parameters *m* and *a* has not previously been elaborated. The goal of our work was therefore to develop a methodology that would identify both the number and length of the nonstandard connectivities to facilitate the structure elucidation of unknown organic

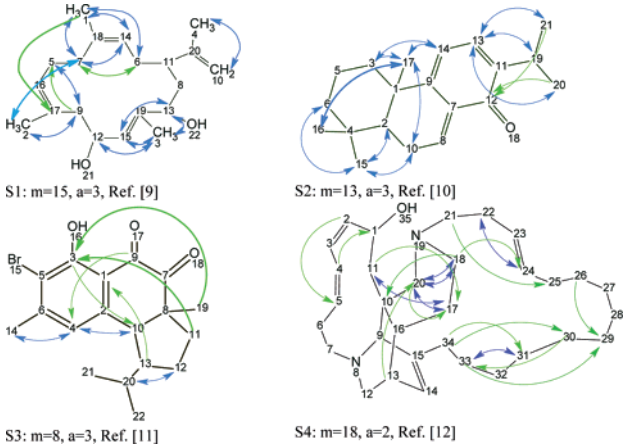


Figure 1. An illustration of a number of structures containing multiple nonstandard correlations. The nonstandard COSY correlations are shown as blue arrows, and the GHMBC correlations are shown by green arrows. In the legends for the structures *m* is the total number of nonstandard correlations, and *a* is the value of correlation lengthening allowed during the process of Fuzzy Generation (see below).

molecules from 2D NMR data. This goal is achieved by fuzzy structure generation on the basis of *m* and *a* parameters whose actual values become known during the process of problem solving. As a result of our studies a solution to the problem posed was determined, and the efficiency of fuzzy structure generation was examined using a series of real-world examples. We have shown that the procedure of determining the correct parameters associated with fuzzy structure generation can be conducted as a result of series of iterations controlled by the user, and, in principle, it is amenable to automation.

2. SOLVING PROBLEMS IN THE PRESENCE OF NONSTANDARD CORRELATIONS

2.1. Modes of Fuzzy Structure Generation. Prior to describing the strategy of fuzzy structure generation some specific concepts will be described. As stated previously[13] a solution to the problem is *valid* if the resulting structural file contains the correct structure. Otherwise, the solution is considered *invalid*. The number of structures included in the output file is denoted as *k*. According to the methodology common for StrucEluc, the solution quality is evaluated on the basis of NMR spectral prediction in order to allow identification of the most probable structure. As will be shown, a comparison of structural output files obtained with the aid of fuzzy structure generation at different stages of the structure elucidation plays an important role in identifying the validity of the final answer. Therefore we will briefly explain the main features of the approach used since details and numerous examples have been reported previously in our articles.[1,3,4,14]

During the first step ^{13}C NMR spectra are predicted for all generated structures using an incremental method, our so-called "Fast" method,[3,4,15] and d_F values, the average deviation of an experimental ^{13}C NMR spectrum versus predicted chemical shifts, are calculated. Recently the incremental prediction algorithms were revamped (an article is presently in preparation reporting the associated advances). The most recent iteration of the algorithm predicts 5000−7000 shifts per second with an average chemical shift

deviation of 1.8 ppm. In parallel a fast algorithm for ^{13}C NMR prediction using artificial neural networks was also implemented in StrucEluc. Since both algorithms have comparable speeds and accuracy, both were employed during one specific run for initial spectrum prediction. After duplicate removal (among the duplicates, the structures with the smallest deviations are retained[4]), the structures are ranked by the d_F value and sorted in ascending order. The smallest d_F value indicates the best match between the experimental and calculated spectra, and this structure will be the first in the output list. Usually the fast calculation of ^{13}C NMR spectra and their subsequent ranking places the correct structure (if it exists in the output file) as the first or the second in the list. Only in very rare instances is the correct structure listed below fifth place. During the next stage so-called "accurate" ^{13}C NMR spectra are calculated for the first 20−50 structures of the ranked file. These predictions are performed using the database of 175 000 structures with the corresponding assigned ^{13}C and ^{1}H NMR spectra.[15] The description of each nuclear environment is defined using the HOSE code approach[16] (Hierarchical Ordering of Spherical Environments). The average deviation values between the experimental and calculated values (d_A) are found, and the structures are again rank-ordered. For additional control over the correct choice of the output structure, the accurate proton chemical shifts can be predicted and displayed together with the corresponding deviation value d_H (the mentioned[15] database is used). The same can also be done for ^{15}N chemical shift predictions when the investigator has access to direct or long-range ^{1}H-^{15}N heteronuclear chemical shift correlation data. For proton NMR prediction, the predicted proton−proton couplings can be enhanced by three-dimensional optimization of the structure. An additional complex match factor $d_\Sigma = d_A + 10 \cdot d_H$ is also calculated. The complex match factor reflects how well the structure matches to both the ^{13}C and ^{1}H NMR spectra. Recalling that the ^{13}C NMR shift range is about 0−200 ppm, while for ^{1}H shifts the range is about 0−15 ppm we ensured an approximate equal weight to the two deviations, $d(^{13}C)$ and $d(^{1}H)$, by multiplying d_H by 10 and defining the complex match factor as $d_\Sigma = d_A + 10 \cdot d_H$**.** The position of the correct structure in the file determines its rank depending on the type of ranking parameter, i.e., d_A, d_F, d_H, or d_Σ correspondingly. The rates of the correct structure in the ranked file are denoted as r_A, r_F, r_H, and r_Σ. If the correct structure assumes primary position in the list ranked by d_A values, then $r_A = 1$, and the corresponding deviation is denoted as $d_A(1)$. As a rule, the final structural ranking is carried out according to the increasing values of the most significant d_A or d_Σ parameters, while the magnitudes of the d_F and d_H parameters serve as additional aids for estimating the reliability of the correct structure selection.

The first ranked structure is considered as the most probable one. If the deviation $d_i(1)$ calculated for the *first* ranked structure is less than a threshold D_i which depends on the precision of spectrum prediction, then the solution is classified as *acceptable*. Otherwise the solution is deemed to be *unacceptable*. Experience has shown that acceptable solutions are most frequently valid though exceptions can occur.

Numerous computational experiments have allowed us to conclude that if the program detects the presence of nonstandard correlations but fails to resolve contradictions in the 2D NMR data using algorithms,[1] then fuzzy structure generation should be used to solve the problem. Moreover, it is quite probable that structure elucidation from 2D NMR data on the basis of fuzzy structure generation can be considered as a general CASE strategy because it is almost independent of the presence or absence of nonstandard correlations in the 2D NMR data.

Fuzzy structure generation can easily be controlled by parameters that make up a set of options. The two main parameters are *m*, the number of nonstandard connectivities, and *a*, the number of bonds by which some connectivity lengths should be augmented. Unfortunately, 2D NMR spectral data cannot deliver definitive information regarding the values of these variables and, as a matter of fact, both of them can be determined only during the process of structure elucidation. We have concluded that in many cases the risk of choosing an erroneous value for *a* can be avoided, and the solution of a problem can be considerably simplified if the lengthening of the *m* connectivities is replaced by their *deleting*. When set in the options the program can ignore by deleting connectivity responses that have to be augmented (by convention, the parameter *a* is set to a value of 16 in these cases). Such an approach can be successful in those cases when the number of 2D NMR connectivities is in some sense optimal. In this sense we mean that the total number of connectivities (structural constraints), *N*, must be large enough to facilitate a description of the chemical structure. In many instances, there are sufficient numbers of correlations in the ensemble of 2D NMR data acquired to essentially overdetermine the structure—in other words there is redundancy in some of the connectivity information. It can then be expected that deletion of *m* of the connectivities will not dramatically influence either the generation time or the size of the output file. On the other hand, the number of combinations of *N* connectivities taken *m* at a time can be very large. This can dramatically impede problem solving to a point that it is not feasible to solve the problem. Indeed, some workers have commented that some of the accordion-optimized long-range heteronuclear shift correlation experiments actually provide too many long-range correlations of the type $^{n}J_{XH}$ where $n \geq 4$.

If the number of connectivities, *N*, is small, then further decreasing *N* by *m* in a connectivity combination can lead to an excessive decrease in the number of structural constraints required for solving the problem. In such a case the problem may be difficult to solve because the 2D NMR data structural constraints will only reduce the total number of possible isomers very slightly.

Independent of the use of augmentation or removal of connectivities, the crucial point in application of fuzzy structure generation is the number of connectivity combinations that should be checked during structure generation. For instance, if $N = 60$ and $m = 5$, then the number of connectivity combinations, $n_{math} = C_N^m$, is equal to ~5.5 million. Any attempt at structure generation has to be performed using each of these combinations. It is necessary to perform generation of structures from each of the C_N^m data sets and obtain the output file as a unification of all of the intermediate results. Even though the StrucEluc structure generator is fast, the productivity is certainly insufficient in

terms of coping with a combinatorial problem as outlined here.

To overcome this difficulty the system is delivered with an algorithm capable of reducing the number of combinations without the risk of losing the correct solution. The first step is to reduce the total number of connectivities N down to N_0, where N_0 is the number of connectivities used to form the connectivity combinations. The data are preprocessed according to the following rules: (1) ambiguous connectivities are excluded from consideration. Ambiguous connectivities are those that appear due to accidental degeneracy of chemical shifts associated with two or more nonequivalent atoms; (2) if two connectivities C-1 to C-2 and C-2 to C-1 are present, then only one of them is included in a data set. One of the two equivalent correlations is redundant and corresponds to overdetermination of the data needed for solution of the structure.

The second and most important step is based on the results of logical analysis of the initial 2D NMR data. If connectivity sets containing nonstandard connectivities are identified (see details in ref 1), then groups of these connectivities are utilized to produce connectivity combinations. As a consequence connectivities that are suspected to be nonstandard are included in all resulting combinations and the initial number of combinations reduces (as will be shown later this number can be reduced by many factors). In addition, the algorithm is capable of immediately detecting combinations of connectivities from which structure generation is impossible—a connectivity combination of this kind still contains at least one nonstandard connectivity. These combinations are skipped during the structure generation process. As a result fuzzy structure generation can be performed in a reasonable time even in those cases when n_{math} is very large. If the MCD checking process fails to detect nonstandard correlations in the 2D NMR data (according to our studies the probability of failure is about 10%), then the program is forced to try all C_N^m connectivity combinations. This can drastically increase the time to solve the problem, and the described approach is inefficient. In these cases User Fragments and Found Fragments[3−4] can frequently be helpful. The ability of the program to calculate and display the real number of connectivity combinations to be validated during fuzzy structure generation allows the user to approximately evaluate the complexity of a given task even at the first stage of the structure elucidation process.

When option parameters are combined in a different way, it is possible to initiate the following modes of fuzzy structure generation:

Mode 1. Structures are generated such that the number of correlations that are extended is specified ($m=m_0$) and connectivity augmentation is also assigned ($a=a_0$). In this case for a GHMBC correlation having a length of 1−2 skeletal bonds both the lower and upper length limits are updated and the connectivity length is extended to 3 bonds.

Mode 2. Structure generation is performed using the following options: it is assumed that the number of extendable (or ignored) connectivities cannot exceed m_{max}, ($m=1,2,...,m_{\text{max}}$), while a is equal to a_0. The m_{max} value is defined as the *maximum* allowed number of nonstandard correlations in the 2D NMR data. Typically the m_{max} value is set equal to 20 thereby covering a wide range of nonstandard connectivities (see Figure 1). The program initially performs structure generation with a value of $m = 1$. If the attempt is unsuccessful, then the m value is *automatically* incremented by 1, and a new run is made with $m = 2$ and so on. An iteration is declared unsuccessful if either no structure is stored after structure generation and spectral filtration or if an *unacceptable* solution was found. When m reaches the m_g value, then the program considers the 2D NMR data to be consistent, and then fuzzy structure generation is initiated with $m = m_g$. The program stops after completing structure generation with $m = m_g$ if the output structural file is not empty and if an *acceptable* solution is provided.

Mode 3. The number of connectivities m is allowed to vary between m_{min} and m_{max} values ($m_{\text{min}} \le m \le m_{\text{max}}$), while the fixed number of bonds a_0 is set. The minimum number m_{min} is usually derived as a result of checking the 2D NMR data for consistency. The program stops when similar conditions as described for mode 2 are achieved.

Mode 4. This mode is a generalization of mode 3 where the interval for m value variation is defined by the condition $m_{\text{min}} \le m \le m_{\text{max}}$ at $m_{\text{min}} = 0$. The peculiarity of this mode is that it is a "generalized" mode of structure generation and can be initiated with $m = 0$. In this mode, the program starts by checking the hypothesis that nonstandard correlations are absent in a given 2D NMR data set. If the data set does not contain nonstandard connectivities, then the program completes the process of structure generation, and the further solution of the problem is carried out as described previously.[3−4,14] If an attempt with $m = 0$ proves to be unsuccessful, then the program automatically performs fuzzy structure generation starting with $m = 1$, $a = 16$ and continues problem solving in the manner described earlier for mode 3. The merit of such an approach is that no assumption regarding the a value is necessary.

Mode 5. This mode is initiated if it is necessary to perform fuzzy structure generation iteratively covering all values of m starting from m_{mim} to m_{max} without exclusion. For example, if structure generation is successful at $m = m_g$, then the program automatically switches to $m = m_g + 1$ and so on until it reaches $m = m_{\text{max}}$. The structures generated at each step are added to those generated during the previous step. This mode is useful to check the solution for stability to make sure that the best structures found at steps $m = m_g$ and $m = m_g + 1$ or higher are equivalent.

Mode 6. This mode resembles mode 5, but the function of this mode is to generate all structures for which the number of nonstandard connectivities is *less or equal to m* at the given value of a. The corresponding options are denoted as $\{m \le m_0,\ a=a_0\}$. The number of connectivity combinations from which fuzzy structure generation is performed depends only on the N_0 and m values. In contrast to the "step-by-step" modes some combinations of the connectivities are united by this approach, and this in principle can speed up the calculations. When this procedure is performed, only the maximal lengths of GHMBC connectivities (i.e., two skeletal bond lengths) are enlarged. For example, consider a GHMBC connectivity between C-1 and C-2 atoms whose "standard" length is varied from 1 to 2 skeletal bonds. In this mode the updated connectivity length varies from 1 to 3 skeletal bonds. It is important to note that the number of nonisomorphic structures generated in this mode is equal to the total number

Table 1. Main Features of Fuzzy Structure Generation in Different Modes

generation mode	*m*, number of extendable or ignored correlations	*a*, connectivity length augmentation	peculiarities of the fuzzy structure generation process
mode 1	$M = m_0$	$a = a_0$	Generation is stopped after checking all possible combinations of extendable correlations.
mode 2	$m = 1 \rightarrow m_{max}$, $m = 1, 2,...$ m_{max}, m_{max} is defined by the user	$a = a_0$	Stop after completing structure generation with $m = m_g$ if the output structural file is not empty and if an *acceptable* solution is provided.
mode 3	$m = m_{min} \rightarrow m_{max}$, $m_{min} \leq m \leq m_{max,}$ m_{min} and m_{max} are defined by the user	$a = a_0$	the same conditions as for mode 2
mode 4	$m = m_{min} \rightarrow m_{max}$, $m_{min} = 0$, m_{max} is defined by the user	Start with strict generation at $m_0 = 0$, $a = 0$. If unsuccessfully, continue with $m_{min} = 1$, $a = 16$	The same conditions as for mode 3. No assumption about *a* value is necessary.
mode 5	$m = m_{min} m_{min} \rightarrow m_{max}$, $m_{min} \leq m \leq m_{max,}$ m_{min} and m_{max} are defined by the user	$a = a_0$ or $a = 16$ at user discretion	Perform fuzzy structure generation iteratively covering *all* values of *m* starting from m_{min} to m_{max} without exclusion
mode 6	Generate all structures for which the number of NSCs is *less or equal to m*, $\{m \leq m_0, a = a_0\}$	$a = a_0$ or $a = 16$ at user discretion	Allows a problem to be solved in which the 2D NMR data contain an *unknown* number of nonstandard connectivities of an *unknown* length.

of nonisomorphic structures generated during all steps of mode 5. However, the total time necessary for completion of fuzzy structure generation can be significantly different between these modes.

Mode 6 with the parameter *a* set equal to 16 can be considered as the most comprehensive mode since in principle it will solve a problem in which the 2D NMR data contain an *unknown* number of nonstandard connectivities of an *unknown* length. Experience has shown that depending on the complexity of the problem the *m* value is typically equal to 5, 10, or 15. If the problem is successfully solved with a given set of options, then the real *m* and *a* values are simply determined visually from the resulting structure which is displayed along with all COSY and GHMBC connectivities. In other cases these parameters can be estimated only by the trial method.

All modes described above are summarized in Table 1. In addition to the approaches mentioned for controlling fuzzy structure generation there is also a possibility to exclude the COSY data from the process of fuzzy structure generation as a user option. In some cases, especially those when the COSY data contain many nonstandard correlations requiring $a > 1$ while the GHMBC data are rich enough, the exclusion of the COSY data both simplifies and accelerates the solution of the problem.

2.2. The Strategy of Fuzzy Structure Generation Application. The possibility of employing several different modes of fuzzy structure generation provides a very flexible analytical tool. However, the diversity of modes available is also a source of complexity since the user has to choose the optimal mode when solving a specific problem. Before starting the calculations it is unclear which mode will lead to a solution in a reasonable time. An attempt was made to answer the question of whether there is a general strategy of structure elucidation using fuzzy structure generation that works best. A set of more than 100 problems was selected where either the GHMBC or COSY spectra or both contained a total of 1−18 nonstandard connectivities corresponding to a range of coupling constants ${}^nJ_{HH,CH}$ where $n = 4-6$. The structures under investigation were all natural products, and the number of skeletal atoms in the molecules varied between 15 and 75 skeletal atoms. The experimental data were obtained from articles published mainly in the *Journal of Natural Products* or from collaborations with various laboratories.

For each problem the NMR spectral data were entered into the program and graphically represented as MCDs (Molecular Connectivity Diagrams, see refs 3 and 4). The procedure for checking the 2D NMR data for contradictions[1] was then applied to every problem. If the presence of nonstandard connectivities was revealed, then the program displayed the minimum number of nonstandard connectivities and made an attempt to automatically resolve the contradictions as described in our work.[1] In successful cases the updated MCDs were displayed with modified connectivities marked by specific color.

As a result of these studies all problems were classified into three sets as follows: (1) 53 problems were identified where NSCs were detected and the initial MCDs were updated; (2) 34 problems were identified where the program revealed presence of NSCs but failed to update the MCDs; and (3) 13 problems were identified where the program failed to detect NSCs.

This classification describes all conceivable results of checking the MCDs. Depending on the results of checking the MCD, various modes or combinations of modes can lead to solution of the problem. Attempts to solve each problem were made using different fuzzy structure generation modes to investigate possible approaches. The problems for which valid solutions could not be found during the first attempt were eventually solved after utilizing different fuzzy generation options. Logical data preprocessing frequently allowed significant reduction of the number of connectivity combinations to be tested during the fuzzy structure generation. Figure 2 shows the ratio ρ describing the number n_{real} of tested connectivity combinations to the theoretically calculated number of combinations, $n_{math} = C_{N_0}^m$ for the entire problem set. Figure 3 examines in greater detail these combinations.

The figures demonstrate that the theoretical number of combinations can be hundreds of billions, but the real

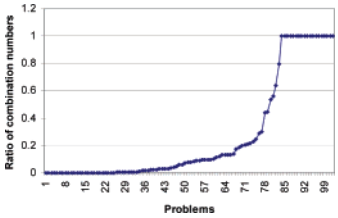


Figure 2. The ratio of numbers of real connectivity combinations to the numbers of theoretically possible combinations for the problems solved using fuzzy structure generation. The program failed to reduce the number of combinations mainly in those cases when nonstandard connectivities were not detected during checking of the MCD.

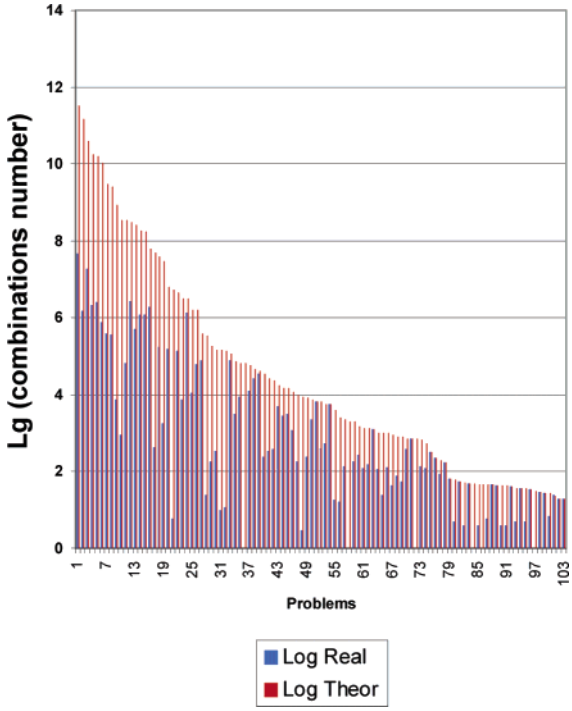


Figure 3. A plot of the logarithms of the theoretical (red) and real (blue) numbers of connectivity combinations.

numbers reduce down to manageable dimensions. For instance, in 20 problems the theoretical number dropped by 10^4-10^6 times, but the real numbers of combinations still remained rather large. Nevertheless, the speed of the structure generator algorithm was fast enough to solve almost all problems.

Fuzzy structure generation did however fail for the elucidation of structure S4 ($C_{32}H_{50}NO_2$) in Figure 1. The 2D NMR data contain 18 nonstandard connectivities (12 GHMBC and 6 COSY nonstandard connectivities; 5 connectivities are of the type ^5J). The theoretical number n_{math} of connectivity combinations is equal to $\sim 43 \times 10^{12}$ for this case. The difficulty could be circumvented by using the Fragment Mode, but no large appropriate fragment was found in the database during the ^{13}C NMR search. The application of a large User Fragment led to an extremely large set of MCDs with each containing the User Fragment with different distributions of the carbon chemical shifts[2]. As a result these two combinatorial "explosions" hampered problem solving. The solution of such computationally difficult problems will hopefully be eased by further development of the algorithm providing fragment "implementation" in MCDs. Work in this direction is presently in progress.

As a result of our studies, general traits were identified that could help to find appropriate ways to solve a problem. These strategies, as applied to the three problem subsets mentioned above, are described in the following sections.

NSCs Were Identified and the MCD Was Updated. Assuming that the MCD updating process was performed correctly (with the lengths of all NSCs increased), then *strict* structure generation is performed. If an acceptable solution is obtained, then it should be checked for suitability. Fuzzy structure generation with the options $\{m=m_{min}\div 20$; stop at $m=m_g$, $a=16\}$ is started from the initial MCD, not the updated MCD. The previously found solution will be confirmed if the first ranked structures for both strict and fuzzy solutions coincide. When an inequality $d^{st}_A(1) > d^{fuz}_A(1, m_g)$ is observed ($d^{st}_A(1)$—the deviation calculated for the first ranked structure of the solution found by strict structure generation, $d^{fuz}_A(1, m_g)$—the same found by fuzzy structure generation at $m=m_g$), then it is concluded that not all NSCs were lengthened and fuzzy structure generation should be repeated with $m_g + 1$ and so on until the minimum value of $d^{fuz}_A(1, m_g+\upsilon)$ and a valid solution is achieved at $m = m_g + \upsilon$. The corresponding structure is then considered as the most probable.

An unacceptable solution can be obtained as a result of strict structure generation from the updated MCD, i.e., a solution will be found for either $d^{st}_A(1) > D_A$, where D_A is a threshold value or an empty structural file is obtained ($k=0$). In both cases the program is automatically switched to the mode where $\{m=m_{min}\div 20$, stop at $m=m_g$, $a=16\}$. Depending on the m_g values and the complexity of the problem (the size of m_{real} and the calculation time) evaluated during the first stages of solving the problem the user can initiate fuzzy structure generation with the options $\{m \le m_0$, $a=16\}$, $m_0 =$ 5, 10, or 15 to obtain the most reliable solution.

NSCs Were Identified but the MCD Failed To Be Updated. If the program identified NSCs but failed to update the MCD, then fuzzy structure generation is one manner by which to solve such a problem. Since the software application only displays the minimum number of NSCs while their associated lengths remain unknown, the solution should be used in Common Mode[4] of FSG with the options $\{m=m_{min}-20$, stop at $m=m_g$, $a=16\}$. The real numbers of the connectivity combinations, n_{real}, are displayed as well as the number of combinations for a given $m = m_g$, and the predicted time for structure generation allow the user to easily evaluate the complexity of the problem and suggested time for execution. If mode 6 can be applied as on the time estimates then it should be used.

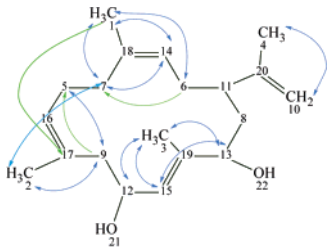
NSCs Were Not Detected. If nonstandard connectivities were not revealed by checking the MCDs, then there are two ways to interpret this result: either the 2D NMR data are free of nonstandard connectivities or the NSCs are present, but the program failed to detect them. Both of these situations are covered by fuzzy structure generation with the options $\{m=0\div 20$, stop at $m=m_g$, $a=16\}$. If NSCs are indeed absent from the 2D NMNR data, then structure generation is performed with $m = 0$ with a nonzero output file and the values of deviation allow the user to determine whether the solution determined is acceptable. Obtaining deviation values that exceed the threshold D_A or deriving an empty output file after spectral filtering both serve as hints to the presence of latent nonstandard connectivities.

When NSCs are not detected by the logical data analysis, then the number of connectivity combinations that must be tested during fuzzy structure generation cannot be reduced, and it is equal to $C_{N_0}^m$, $m = 1, 2, 3, \ldots$ at each mth step of the fuzzy structure generation process. This situation can cause significant difficulties due to an unmanageable number of connectivity combinations needing to be processed; as discussed previously, both Found and User Fragments can assist in this situation.

It is difficult to describe the myriad of nuances associated with fuzzy structure generation since these depend on each 2D NMR data set associated with a given problem. A series of examples illustrating the strategies leading to valid solutions with the minimum number of user assumptions will be presented. Example problems were chosen where automatic updating of the MCD to resolve contradictions was inefficient. The structures are shown in Figure 1 where there are a large number of NSCs used as examples.

2.3. Problem Solution in the Common Mode. *Example 1*. In the analysis of *cleospinol A*[9] with molecular formula $C_{20}H_{32}O_2$ (**1**), the 2D NMR data are comprised of 21 COSY and 55 HMBC correlations. These data were used to evaluate the possibility of solving a problem in those cases when a large number of nonstandard correlations were present. In this case the 2D NMR data contained the following combination of NSCS: 3 HMBC[2*a*(1), 1*a*(3)] + 12COSY-[8*a*(1), 3*a*(2), 1*a*(3)] = 15. This nomenclature describes the fact that there are 3 HMBC nonstandard correlations, two of which must be lengthened by 1 bond and one by 3 bonds; the information about the 12 COSY correlations is interpreted analogously. The total number of NSCs is 15. In this article such expressions are used to provide short and unambiguous designations of the numbers and lengths of the NSCs contained within the 2D NMR data.

The COSY connectivities are represented below on the structure by blue double-headed arrows, while the HMBC correlations are defined by green unidirectional arrows from the proton to the carbon to which it is long-range coupled.



1

The COSY, HMQC, and HMBC spectral data associated with the compound were fed to the program, and the MCD was generated. A check of the MCD was accompanied by the automatic removal of contradictions. The software program displayed a message declaring that the contradictions had been detected and resolved, while the minimum number of NSCs was estimated to be equal to 7. Unfortunately strict structure generation from the automatically edited MCD resulted in an empty output file. This result was interpreted as evidence of the presence of either undetected additional nonstandard correlations or those whose lengths must be augmented by more than one bond.

There are two possible trajectories from this point to solve the problem. Since in general there is no information about the number of NSCs and their lengths, these values can be determined using a trial and error method. If it turns out that $a = 1$, then there is a chance to find a solution in a short time. The second approach is more general and allows the user to ignore the problem of determining the maximal a value. The cost, however, may be longer structure generation times and a consequently larger output file. Both approaches are described in detail below as applied to this specific example.

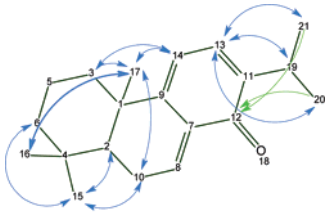
With the first approach fuzzy structure generation was initiated from the initial (not updated) MCD using mode 3 with the options {m=7−20, stop at m=m_g, a=1}. The program started the generation process automatically with $m_g = 14$ (m=7−13 were immediately rejected), and the process was aborted by the operator at $m = 16$ with a zero result. An attempt to repeat fuzzy generation with $a = 2$ again gave an empty result file.

The possibility that one or more of the nonstandard connectivities needed to be augmented by 3 bonds was assumed. When the options {m_{max}=20, stop at m=m_g, a=3} were set, then the program automatically started with $m_g = 10$ and completed the fuzzy generation process in 9 m 15 s (Pentium IV, 2.8 MHz) with $m = 14$. Three molecules were generated and two were stored in the results file after the removal of duplicate structures. The highest ranked structure coincided with the actual structure, **1**. The second structure gave a $d_A(2)$ value with $\Delta_{2\text{-}1} = d_A(2) - d_A(1) = 1.5$ ppm. A solution was found at $m = 14$ and not at $m = 15$ since the COSY and HMBC connectivities between C-5 and C-9 carbons are coincidental. The program displays the final structure where all connectivities and their associated lengths can be visualized.

The second more universal and systematic approach was applied, and fuzzy structure generation was initiated assuming only that the number of nonstandard connectivities is not more than 15 (mode *6*), i.e. options {$m \leq 15$, a=16} were set. In this case 18 281 379 connectivity combinations from 40 225 345 056 theoretically possible combinations were used for structure generation. The following result was obtained: 769 structures were generated, 430 structures were stored after spectral filtering, 245 structures remained after removing duplicates (this is denoted as k=769→430→245), a generation time of $t_g = 29$ min 9 s, and the correct structure was ranked first by all methods of spectrum prediction, $r_{all} = 1$.

The program therefore identified the correct solution even when 15 nonstandard connectivities existed in the 2D NMR data and especially in the presence of HMBC and COSY connectivities representing both $^6J_{CH}$ and $^6J_{HH}$ correlations. Note that only $\sim 10^{-4}$ of the theoretically possible connectivity combinations were processed. The real number of processed connectivity combinations n_{real} is more than 18 million. Nevertheless, the high-speed structure generator present in the Structure Elucidator program completed the process in a reasonable time.

Example 2. The 2D NMR data associated with a natural product with molecular formula $C_{20}H_{28}O$ isolated and identified by Mensah et al.[10] contains 14 NSCS: 2HMBC-[2*a*(1)]+12COSY[4*a*(1), 6*a*(2), 1*a*(3)] = 13. The presence of a minimum of 6 NSCs was detected, and the program

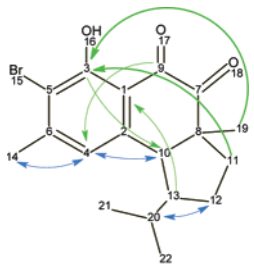


2

displayed a message indicating that the MCD was successfully updated. Strict structure generation of the updated MCD led to an empty output file indicating that fuzzy structure generation must be used. Experience shows that the maximum value of a usually does not exceed 3. A solution to this problem can be obtained with the options $\{m=6-15$, stop at $m=m_g$, $a=3\}$. The program initiated generation at a value of $m_g = 11$ and stopped when 4 structures were generated. The large deviations calculated for the best structure suggested that the solution was invalid (see Table 2). The results of subsequent structure generations with $m_g = 11$, 12, and 13 are presented in the Table 2. All deviations calculated for the best structure achieved minimum values for the solution obtained at a value of $m_g = 13$: the best structure coincided with structure **2**. Carbon atom assignments were identical to those suggested by the authors.[10]

The practical approach was applied with the options $\{m \leq 15$, $a=16\}$. The following results were obtained: $k = 66\,538 \rightarrow 38\,407 \rightarrow 12\,070$, $t_g = 2$ h 41 m, $n_{math} \approx 63 \cdot 10^9$, $n_{real} = 48\,525\,735$, $\rho = 7.6 \cdot 10^{-4}$, $r_A = r_\Sigma = 1$. It is worth noting that the ^{13}C NMR spectra of 38407 structures were calculated in 2 min 23 s for preliminary ranking using the incremental method.

Example 3. Natural product **3**

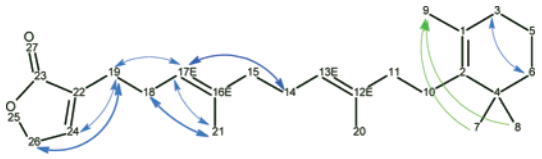


3

was isolated and identified by Wellington et al.[11] using a combination of HMBC and COSY spectra. There are 8 NSCs in the 2D NMR data: 5HMBC[2a(3), 3a(1)] +3COSY-[3a(1)] = 8. As a result of checking the MCD it was updated, and a further analysis showed that the minimum number of NSCs m_{min} was 4. An attempt to perform strict structure generation was unsuccessful, and the program produced an empty output file. The problem could be solved using "step-by-step" methods analogous to those described in the previous example, but we present here only the result of the practical approach.

For modeling a situation where nothing is known about the number and lengths of NSCs except that $m \geq 4$ the following fuzzy structure generation options to ensure minimization of the risk of losing the correct solution were set: $\{m \leq 10$, $a=16\}$. The result is defined by $k = 114\,638 \rightarrow 68\,668 \rightarrow 23\,213$; $t_g = 1$ h 52 m; $n_{real} = 52\,427\,715$, $n_{math} \approx 0.6 \cdot 10^9$; $\rho = 0.08$; $r_{all} = 1$. Using mode 6 the program performed an exhaustive search of all possibilities to generate structures (see section 2.1), and the output file is large. The ^{13}C NMR spectrum prediction for about 70 000 structures using the incremental method took about 5 min. The correct structure was distinguished by examining the values of the deviations calculated using all methods within the StrucEluc system.

Example 4. Kirsch and coauthors[17] reported the isolation and structure elucidation of a natural product with a molecular formula of $C_{25}H_{38}O_2$ (**4**) using 2D NMR data containing 23 COSY and 41 HMBC correlations. As illustrated by the arrows the following set of NSCs are observed in structure **4**: 2HMBC[2a(2)]+7COSY[4a(1), 3a(2)] = 9.



4

The following steps allowed identification of a correct solution to this problem:

(1) When the MCD was checked, nonstandard correlations were detected, but the program declared that the contradictions in the 2D NMR data could not be resolved. The minimum number of NSCs had a value of 6.

(2) Fuzzy structure generation was performed with the options $\{m=6 \div 10$, stop at $m=m_g$, $a=1\}$ but resulted in an empty output file.

(3) Fuzzy generation with the options $\{m=6-10$, stop at $m=m_g$, $a=2\}$ provided a valid solution in less than 7 min. The solution was obtained with $m = 9$ since $m = 1-8$ resulted in empty output files.

When the more general approach was applied with options $\{m=6 \div 20$, stop at $m=m_g$, $a=16\}$, a single and correct structure was found in 2 min. With this 370 950 connectivity combinations of $\approx 2.5 \cdot 10^9$ theoretically possible ones combinations ($\rho \approx 10^{-4}$) were processed in this time.

The validity of the solution was checked by a time-consuming process using fuzzy structure generation with the options $\{m \leq 10$, $a=16\}$ to give the following result: $k = 18 \rightarrow 17 \rightarrow 15$, $t_g = 28$ min 35 s, $n_{real} = 4\,830\,600$, $n_{math} \approx 10 \cdot 10^9$, $\rho = 4.7 \cdot 10^{-4}$, $r_{all} = 1$. The first ranked structure and associated atom assignment coincided with that determined by the authors.

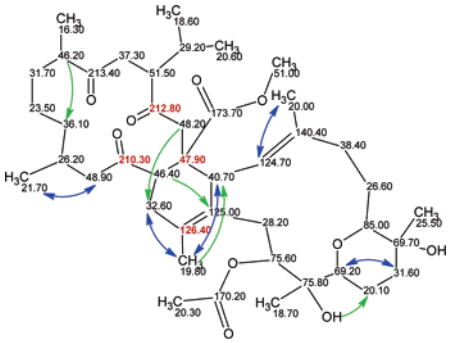
Example 5. Computational difficulties associated with structure generation rise with an increase in the complexity of the molecule even in those cases when the 2D NMR data contain correlations of standard lengths only. The difficulties become especially serious if the nonstandard correlations exist in COSY and HMBC spectra. The possibility of solving problems using fuzzy structure generation for a large molecule with a large number of NSCs in 2D NMR data is illustrated in the following example.

Feller et al.[18] reported the isolation and structure determination of a new terpenoid **5** with a molecular formula of $C_{43}H_{66}O_{10}$ and therefore containing 53 skeletal atoms:

The 2D NMR data were composed of 53 COSY and 94 HMBC correlations including 10 nonstandard connectivities

Table 2. Results of Four Subsequent Steps of Fuzzy Structure Generation

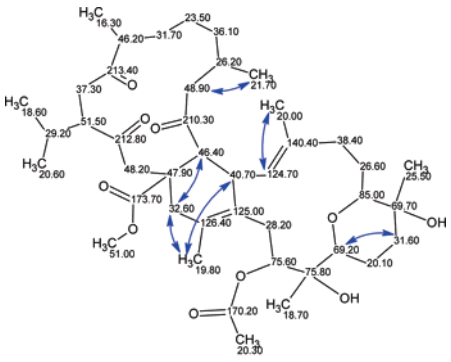
m	N_0	n_{math} (approximately)	n_{real}	ρ (n_{real}/n_{math})	results
11	41	$3.1 \cdot 10^9$	17 629	$5.6 \cdot 10^{-6}$	$k = 4 \rightarrow 4$, $t_g = 1$ m 35 s, $d_A(1) = 6.22$,
12	41	$7.9 \cdot 10^9$	240 001	$3 \cdot 10^{-5}$	$k = 25 \rightarrow 19$, $t_g = 2$ m 40 s, $d_A(1) = 5.15$,
13	41	$17.6 \cdot 10^9$	1 601 574	$1.2 \cdot 10^{-4}$	$k = 218 \rightarrow 176 \rightarrow 110$, $t_g = 12$ m, $d_A(1) = 2.13$



5

as shown on the representation of structure **5** and enumerated as follows: 5HMBC[5*a*(1)] + 5COSY[4*a*(1), 1*a*(2)] = 10. When the MCD was created, four carbon atoms, whose chemical shifts are marked in red, were not involved in any correlations. The presence of such "free" atoms introduces an additional obstacle to solving a problem and generally leads to an increase in structure generation time.

MCD checking accompanied by automated resolution of contradictions in the 2D NMR data produced a program message declaring that the minimum number of nonstandard connectivities was 7. The MCD was then updated by the program to resolve the contradictions. Structure generation from the updated MCD gave $k = 56 \rightarrow 28$ and $t_g = 21$ s. Structure **5′** was distinguished as the best one.



5′

The deviation value of $d_A(1) = 2.74$ ppm is typical for a correctly recognized structure[2] so there was no reason to reject the first ranked structure. The visualization of the nonstandard correlations in structure **5′** displays the presence of 6 nonstandard COSY connectivities, including the connectivity 19.8 to 40.7 corresponding to a $^5J_{HH}$ coupling constant. This connectivity was lengthened by two bonds following analysis of 2D data. Five of the six NSCs exist in structure 5, but lengthening the connectivity 46.4−32.6 was a mistake and leads to an incorrect structure similar to the genuine one. To obtain a reliable solution fuzzy structure generation was performed as outlined below.

Fuzzy structure generation was performed using the options {m=7−15, stop at m=m_g, a=16}. A nonempty structural file was generated with $m_g = 9$ ($n_{math} \approx 328 \cdot 10^9$, $n_{real} \approx 28 \cdot 10^6$, $\rho = 8 \cdot 10^{-5}$) due to the accidental degeneracy of two COSY and HMBC nonstandard correlations 40.7−19.8 (otherwise only m=10 would be successful). The following results were obtained: $k = 36 \rightarrow 21 \rightarrow 11$, t_g = 7 h 50 min, r(all) = 1, $d_A(1) = 2.62$ ppm, and the best structure coincided with that deduced by the authors.[18] The difference between the d_A deviations for the correct (**5**) and incorrect (**5′**) structures is only 0.12 ppm. The calculated similarity coefficient was equal to 0.98 for these structures. The structures differ only by the permutation of the carbons at 47.9 and 46.4 ppm which resulted in the transformation of part of the correct structure into its mirror reflection.

The solution was fairly time-consuming due to a large number of connectivity combinations ($28 \cdot 10^6$) as well as the presence of four carbon atoms with no connectivities demonstrated to other atoms. The example provides evidence that the approach is efficient even in a situation when the analyzed molecule is large and the number of NSCs is big enough including correlations corresponding to $a > 1$.

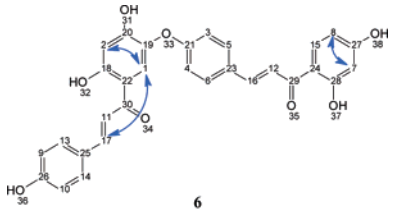
The examples presented in this section demonstrate the high efficiency of the procedure suggested for logical analysis of 2D NMR data. The application of this procedure reduces the total number of connectivity combinations by about 10^4−10^6 times, and this allows the program to complete the fuzzy structure generation within a reasonable time. The examples given lead to the conclusion that the capability of fuzzy structure generation in the StrucEluc system enables complex tasks to be successfully resolved even in those cases when the number of nonstandard correlations is large (10−15), and when correlation lengths exceed the default values of 2−3 bonds by 2 or even 3 additional bonds.

2.4. Problem Solution in the Fragment Mode. As shown in our previous publications,[3−4,14] the use of fragments stored in a content database (the so-called Found Fragments selected as a result of a database search using ^{13}C NMR spectra as inputs) as well as User Defined Fragments allows problems to be solved even when Common Mode structure generation is time-consuming and cannot be completed in a reasonable time. In this work peculiarities of the fragment approach have been investigated as a tool for solving problems in the presence of nonstandard connectivities.

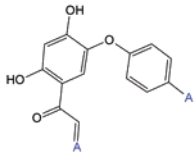
As shown previously[3] the need to apply fragments to solve a problem arises when there is a deficit of hydrogen atoms in the molecular formula or when the number of connectivities in the 2D NMR data is simply not ample enough to produce a set of efficient structural constraints. One might assume that the larger the fragment involved in forming the MCD then the quicker the time to arrive to a solution. In reality for a large fragment there is a huge number of permutations for different assignments of the experimental chemical ^{13}C shifts to the fragment carbon atoms, and this can lead to an extremely large number of MCDs for processing during structure generation. When fuzzy structure generation is applied to the set of MCDs containing fragments, the m_g value can be different for each MCD, and the

m_{min} value estimated during checking of the MCD generated in the Common Mode has low predictability. The specificity of fuzzy structure generation from a set of MCDs containing fragments will be demonstrated by examining a series of examples.

Example 1. Mdee and co-workers[19] isolated and identified using 2D NMR data (9 COSY and 44 HMBC correlations) a new bichalcone of molecular formula $C_{30}H_{22}O_8$.



6



7

The COSY spectrum contained a set of nonstandard correlations 3COSY[1*a*(1), 1*a*(2), 1*a*(3)] as illustrated by the blue arrows shown in structure **6**. Checking the MCD and using automatic contradiction removal detected a minimum number of 2 NSCs and resulted in updating of the initial MCD. No structures could be generated from the resulting MCD however. An attempt to perform fuzzy structure generation from the initial MCD with the options {m=2÷10, stop at m=m_g, a=16} was also unsuccessful: even though only 213 connectivity combinations were processed with m_g = 2 and a = 16 the fuzzy generation process turned out to be very time-consuming and tens of hours were predicted for the value t_g. A search of the Fragment Database using the ^{13}C NMR spectrum as an input was performed, and 1138 fragments were selected. The first ranked fragment, **7,** showed good coincidence of its ^{13}C subspectrum with the chemical shifts of the unknown compound (spectrum comparison was easily performed due to a visual display capability for spectrum representation[4]). The program created four MCDs from this fragment.

Fuzzy structure generation was initiated {m=2−15, stop at m=m_g, a=16} and was completed with the following result: k = 46, t_g = 6 min 30 s, r_{all} = 1, i.e., the correct structure was unambiguously identified. Analysis of the conditions applied for fuzzy structure generation was performed from each MCD which delivered a result and gave

MCD #1: $m_g = 10$, $n_{real} = 6060$, $n_{math} \approx 20 \cdot 10^6$, $\rho = 3 \cdot 10^{-4}$

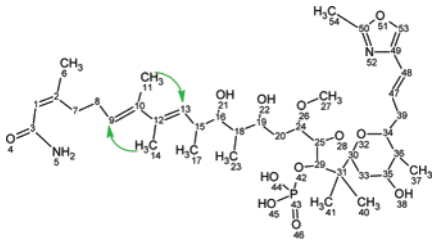
MCD #2: $m_g = 6$, $n_{real} = 96$, $n_{math} \approx 0.3 \cdot 10^6$, $\rho = 3.2 \cdot 10^{-4}$

MCD #3: $m_g = 6$, $n_{real} = 3319$, $n_{math} \approx 0.5 \cdot 10^6$, $\rho = 7 \cdot 10^{-3}$

MCD #4: $m_g = 2$, $n_{real} = 27$, $n_{math} = 351$, $\rho = 8 \cdot 10^{-2}$

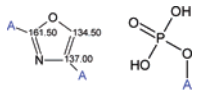
In spite of the different conditions 156 structures were generated from each MCD and resulted in 46 nonisomorphic structures. It is evident that the fuzzy structure generation process was not performed in the most rational way when applied to each MCD, and further work is necessary to optimize this procedure.

Example 2. Kehraus et al.[20] reported the separation and identification of a new natural product geometricin A, $C_{39}H_{63}N_2O_{12}P$ (**8**).



8

The reported 2D NMR data included 11 COSY and 99 HMBC correlations. Two NSCs characterized with a = 1 are present in HMBC data set and are represented by green arrows. The program detected the presence of NSCs and determined m_{min} to be equal to 2 but failed to resolve the contradictions. An attempt to apply fuzzy structure generation to the MCD showed that the problem would be too time-consuming without the introduction of some key fragments. The following user fragments were introduced to assist in solution of the problem.



9 **10**

Kehraus et al.[20] reported that the ^{13}C NMR chemical shifts for carbons C-49, C-50, and C-53 together with a singlet proton resonance at δ 7.76 characterized an oxazole ring. With the chemical shifts calculated for fragment **9** the program produced 18 MCDs containing both user fragments. With m_{min} = 2 fuzzy structure generation was performed with {$m \leq 3$, a=16} to produce the following result: k = 6180 → 4900 → 2450; t_g= 3 min 30 s, n_{math} = 57 154, n_{real} = 1056, ρ = 0.02, r_{all} = 1. In reality a result was obtained only from one of the 18 MCDs since the other MCDs contained fragment **9** in a format whereby the carbon atom assignment did not correspond to that specific to the target structure. ^{13}C NMR spectral prediction for the 4900 structures via the incremental method took only 27 s. As a result of prediction the correct structure was ranked first in the output file.

2.6. Is There an Alternative to Fuzzy Structure Generation? To the best of our knowledge the question regarding to what extent the lengthening of *all* 2D NMR correlations can act as a method for contradiction resolution in 2D NMR data has never been investigated. In this study an attempt to identify the quantitative characteristics describing how structure generation time increases and the amount of structural information obtained decreases if all correlations

belonging in the ^{2-4}J range have been made. In previous work[13] it was suggested that in principle the amount of structural information obtained as a result of the application of an expert system can be measured in the manner described below.

Assume that the entire number of possible isomers N corresponding to the molecular formula of an unknown compound is fixed and nothing is known about the structure to be analyzed. Let p_i be the probability that the ith ($1 \leq i \leq N$) isomer is the genuine structure. Before the task can be solved all isomers must be equally probable and $p_i = 1/N$. According to Shannon[21] the initial entropy E_0 characterizing the solution can then be calculated from the equation $E_0 = \log_2 N$.

If the task results in an output file containing k structures and the solution contains the genuine structure, then the entropy of the correct solution, E_c, can be calculated as $E_c = \log_2 k$. The amount of structural information obtained as a result of solving the problem can be expressed as follows:

$$I_0 = E_0 - E_c = \log_2 N - \log_2 k = \log_2(N/k)$$

Obviously, if $k = 1$ (the structure of the unknown compound is unambiguously and correctly elucidated), then the general amount of *structural* information I_0 obtained as a result of solving the problem will be equal to $I_0 = \log_2 N$. The value μ(%), the portion of the full structural information obtained at any given stage of the task, can be expressed as follows:

$$\mu = (I_c/I_0)\cdot 100 = (1 - \log_2 k/\log_2 N)\cdot 100$$

Our computational experiments have demonstrated that it is possible to generate a full set of isomers corresponding to a given molecular formula for small molecules containing no more than 17 skeletal atoms[2] using the GENM[22-24] generator present in StrucEluc. It takes tens or even hundreds of hours when a PC Pentium IV, 2.8 GHz is used. For our work examining structural problems via 2D NMR data only problems with less than 20 skeletal atoms were selected. The experimental data were borrowed from publications.

In Table 3 the structures, their molecular formulas, and the total number of structural isomers are given. The table shows that all selected structures can be considered as small structures. For each problem the two sets of solution results obtained with two different options enabled are listed. In the first case all correlations were assumed to be ^{2-3}J by default. In the second case the option settings allowed the connectivity length to vary between 2 and 4. Also listed is the number of structures k contained in the output file and the generation times $t_g^{(2-3)}$ and $t_g^{(2-4)}$ and their ratio $\tau = t_g^{(2-4)}/t_g^{(2-3)}$. The τ value demonstrates a slowing down of the generation process if generation is performed using default correlation lengths corresponding to ^{2-4}J. The μ coefficient estimates the loss of structural information that takes place in this case.

Analysis of the table shows that even in the case of small molecules the output file size increases considerably when the ^{2-4}J couplings are set as default. Under those conditions the portion of extracted structural information drops from 95−100% to 60−70%. At the same time the generation time increases by many times to hundreds or even tens of thousands times greater. Problem 6 is the most distinctive example where the size of the output file increased from 2 to 3036 structures, while the generation time increased by 6.5 million times!

For problems 2 and 5 in Table 3, the 2D NMR data contained nonstandard correlations, and both tasks were solved automatically using fuzzy generation with the output file containing only one structure at the conclusion of the run. The solutions to these problems found by simply lengthening all correlations to ^{2-4}J resulted in an increase in the number of structures up to 25 (problem 2) and up to 1211 (problem 5) structures. The computing time grew by 25 times and 11 900 times, respectively.

Table 3 lists some examples that allow the examination of the dependence of the results on the default option settings. However, the most typical problems of the same structure size are presented in Table 4. In example 1[34] 8 out of 28 HMBC correlations were of nonstandard length. The task was solved using the fuzzy generation mode and resulted in a single correct structure with a generation time of 41 min 35 s. Strict generation with the coupling constant values ^{2-4}J set by default gave $k = 4$ and $t_g = 24$ min. This example shows that if a molecule is small and the number of nonstandard correlations ^{2-4}J is large, then while the longer correlations (^5J, ^6J) are absent in 2D NMR data the application of fuzzy generation and the lengthening of all correlations both give comparable results in a case when the number of skeletal atoms is less than 20.

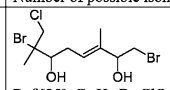
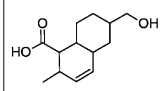
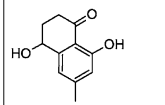
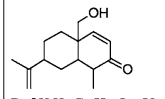
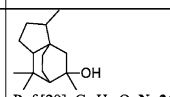
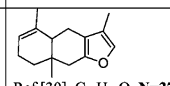
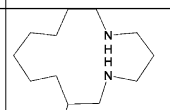
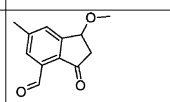
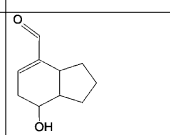
The problem solution time and the number of generated structures increase dramatically when the size of the molecule increases, and it is evident that the default setting of ^{2-4}J for the correlation length can only assist when the number of skeletal atoms is around 20.

The solutions for problems 2 and 3[35,36] with ^{2-3}J coupling constants set as the default were identified in several seconds. However, when the coupling constants were set to ^{2-4}J as default, the program was aborted by the operator in about 15 h. At the same time the number of generated structures was around 10^5, and it was impossible to predict the time left in order for structure generation to be completed.

Investigation of the results of problem 4[37] also provided interesting results. The correct solution was determined by two approaches: using the method of automatic correction of the MCD (k=1, t_g =0.009 s) and using fuzzy structure generation with the options {m_{max}=10, stop at m=m_g, a =2} and resulted with $k = 1$, $t_g = 0.031$ s. Structure generation with all correlations set to ^{2-4}J by default gave 61 structures in about 6 min. Rank ordering of the structural file according to the deviation values gave a structure with the deviations $d_A(1) = d_F(1) = 8.6$ ppm and $d_H(1) = 0.66$ ppm placed at the top of the ranked file. According to previously defined criteria $d_A(1)$ should be less than 5.5 ppm for the correct structure, and under this constraint the solution is likely incorrect. This implies that the generation process should be repeated with the coupling constants set to ^{2-5}J as the default.

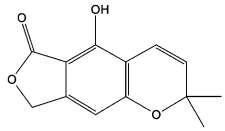
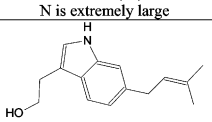
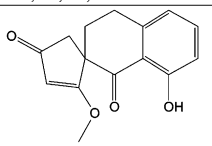
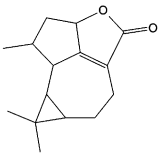
It could be expected that the number of structures would be large so the structure generation process was executed in a mode whereby the structures were not written to disk. The program was later aborted by the operator after 6 min since the number of resultant structures had already reached half a million.

Table 3. Dependence of the Amount of Structural Information Extracted from 2D NMR Data Based on the Nature of the Coupling Constant ^nJ Value Set as the Default during the Structure Generation Process[a]

No	Structure. Molecular formula. 2D NMR data available. Number of possible isomers.	Comparison of problem solutions obtained with $^{2\text{-}3}J$ and $^{2\text{-}4}J$ correlations.	Structural information, μ and τ values
1	 Ref.[25], $C_{10}H_{17}Br_2ClO_2$, **N=50,502,293** HMBC (24)	$^{2\text{-}3}J$ k=1 t_g=0.008 s	I_0=25.59 μ=100%
		$^{2\text{-}4}J$ k=103 (E=6.69) t_g=2.36 s	E_0=25.59 I=18.90 μ= 74%, τ=295
2	 Ref.[26],$C_{13}H_{20}O_3$, **N=14,431,269,166** COSY(33), HMBC (48)	$^{2\text{-}3}J$ 3 NSCs, COSY, a=1 2 NSCs HMBC, a=1 Fuzzy generation: k=1 t_g=0.014 s	I_0=33.75 μ=100%
		$^{2\text{-}4}J$ k=25 (E=4.64) t_g =0.342 s	E_0=33.75 I=29.11 μ= 86%, τ=25
3	 Ref.[27],$C_{11}H_{12}O_3$, **N=4,703,963,545** HMBC(20)	$^{2\text{-}3}J$ k=2 (E=1) t_g=0.011 s	I_0=32.13 μ=97%
		$^{2\text{-}4}J$ k=933 (E=9.86) t_g=1 m 46 s	E_0=32.13 I=22.27 μ=69%, τ=9630
4	 Ref.[28], $C_{15}H_{22}O_2$, **N=138,136,211,624** HMBC(48)	$^{2\text{-}3}J$ k=1 t_g=0.008 s	I_0=37 μ=100%
		$^{2\text{-}4}J$ k=823 (E=9.69) t_g=1 m 28 s	E_0=37 I=27.31 μ=74%, τ=11,100
5	 Ref.[29], $C_{15}H_{26}O$, **N=261,045,917** COSY(12), HMBC(56)	$^{2\text{-}3}J$ k=1 t_g=0.009 s 2 NSC COSY, a=1 Fuzzy generation	I_0=27.95 μ=100%
		$^{2\text{-}4}J$ k=1211 (E=10.23) t_g=1 m 47 s	E_0=27.95 I=17.72 μ=63%, τ=11,900
6	 Ref.[30], $C_{15}H_{20}O$, **N=37,568,150,635** HMBC (32)	$^{2\text{-}3}J$ k=2 (E=1) t_g=0.011 s	E_0=35.14 I=34.14 μ=97 %
		$^{2\text{-}4}J$ k=3036 (E=11.57) t_g=2 h	E_0=35.14 I=23.57 μ=67 %, τ=$6.5\cdot10^6$
7	 Ref.[31], $C_{13}H_{28}N_2$, **N=6,332,846** HMBC (35)	$^{2\text{-}3}J$ k=2 (E=1) t_g=0.012 s	E_0=22.60 I=21.60 μ=95 %
		$^{2\text{-}4}J$ k=2 (E=1) t_g=6 s	E_0=22.60 I=21.60 μ=95 %, τ=500
8	 Ref.[32], $C_{12}H_{12}O_3$, **N=68,930,547,646** HMBC (26)	$^{2\text{-}3}J$ k=2 (E=1) t_g=0.050 s	E_0=36 I=35 μ=97%
		$^{2\text{-}4}J$ k=22 (E=4.46) t_g=11 s	E_0=36 I=31.54 μ=88%, τ=220
9	 Ref.[33],$C_{10}H_{14}O_2$, **N=16,422,284** HMBC (39)	$^{2\text{-}3}J$ k=1 t_g=0.008 s	I_0=23.95 μ=100%
		$^{2\text{-}4}J$ k=132 (E=6.77) t_g=10 s	E_0=23.95 I=17.18 μ=72% τ=1250

[a] Calculations were performed with a PC Pentium IV, 2.8 GHZ with 1 Gbyte of RAM. The abbreviation NSC represents Non-Standard Connectivity.

Table 4. Dependence of Structural Information Extracted Based on the J Value Range Utilized for a Series of Fairly "Large-Sized" Molecules

1	 Ref.[34], $C_{13}H_{12}O_4$ HMBC (28) N is extremely large	$^{2\text{-}3}J$ 8 nonstandard HMBC connectivities (a=1). Fuzzy generation was performed using options {m_{max}=10, stop at m=m_g, a=1} Results: k=1 ; t_g=41 m 35 s $^{2\text{-}4}J$ k=4; t_g=24 m
2	 Ref.[35], $C_{15}H_{19}NO$ HMBC (30) 600,000,000,000 < N → $10^{13\text{-}14}$	$^{2\text{-}3}J$ k=4 ; t_g=2.6 s $^{2\text{-}4}J$ Structure generation was aborted after 14 h with k~116,000 structures produced. An indicator predicted that at least double this number of structures would be produced. The problem could not be solved.
3	 Ref. [36], $C_{15}H_{14}O_4$ HMBC (31) N is extremely large	$^{2\text{-}3}J$ One nonstandard HMBC correlation (a=1). Fuzzy structure generation with options: {m_{max}=10, stop at m=m_g, a=1} Results: k=5; t_g=2 s $^{2\text{-}4}J$ Structure generation was aborted after 15 h with k~10^5 .
4	 Ref. [37], $C_{15}H_{20}O_2$ COSY(22), HMBC(62) N is extremely large	$^{2\text{-}3}J$ 2 nonstandard COSY correlations (a=2) 2 nonstandard HMBC correlations (a=1) *Automatic MCD correction*: k=1, t_g=0.009 s *Fuzzy generation* {m_{max}=10, stop at m=m_g, a=2}: k=1, t_g=0.031 s $^{2\text{-}4}J$ k=61; t_g=5 m 54 s Chemical shift deviations of the best structure: $d_A(1)$=$d_F(1)$=8.6 ppm; $d_H(1)$=0.66 ppm Wrong solution $^{2\text{-}5}J$ Run in the mode of isomer counting. The program was aborted in 6 min when k reached ~500,000 structures

The last example shows that even in the case of a molecule being fairly small, in this case with n < 20, and even when the both COSY and HMBC spectra contain a large number of standard correlations, the process of increasing the intervals allowed for the correlation lengths offered no solution. In such a situation, the most effective way to solve the problem appears to be fuzzy structure generation.

We assume that the results obtained in this work strongly support the guideline that the lengthening of *all* correlations should be rejected as a general method of solving problems arising from the presence of nonstandard correlations in 2D NMR data. This conclusion is even more obvious if we take into account the distribution of problems solved using StrucEluc.[2] The distribution shows that molecules with less than 20 skeletal atoms are rarely found among natural products.

4. CONCLUSIONS

This study has shown that fuzzy structure generation should be considered as the most general method for structure generation in expert systems based on 2D NMR spectra. It allows the process of structure elucidation to be initiated under conditions whereby the user has no idea about the presence or absence of NSCs or details regarding their real lengths. This is attained by setting appropriate generator options that provide a varying number of NSCs in the range of 0 up to m_{max}. Experience has shown that the value of m_{max}

is usually set equal to 10, 15, or 20 depending on the minimum number of NSCs detected by the program during the logical analysis of 2D NMR data. The real number of NSCs and their lengths can be determined by two approaches.

The first approach allows identification of these variables along with the structure of the unknown using a trial method. In so doing, both variables, *m* (true number of NSCs) and *a* (augmentations of connectivity lengths), are varied.

The second approach assumes only that the *m* variable does not exceed some m_{max} value, while no information about *a* value is necessary. This approach is advisable since it is fully automated and more universal. Its shortcoming is that fuzzy structure generation can in some cases be more time-consuming and in addition the output file can be larger. This can be of little impact since calculation time can be relatively inexpensive in the present era of computational cost and structure generation can be performed in background mode anyways. At the very least computations can be left overnight as is common for 2D NMR data acquisition. We suppose that $t_g = 5-15$ h is still acceptable because the separation and identification of a natural product by traditional methods usually takes weeks and even months. To elucidate the structure of a new compound without any danger associated with the presence of an unknown number of NSCs of unknown lengths is attractive enough to afford time-consuming structure generation.

As a result of dramatically increasing the speed of ^{13}C NMR spectrum prediction (5000−7000 shifts/s) a large output file no longer hampers fast candidate structure spectral prediction that is necessary for the optimal elimination of isomorphic structures and file ranking in order to select the most probable structure. For instance, the ^{13}C spectrum prediction and average deviation calculation for 4900 isomers generated from the molecular formula $C_{39}H_{63}N_2O_{12}P$ (see structure **8**) took only 27 s. Fuzzy structure generation can be concluded to be an appropriate analytical tool for application to the structure elucidation of organic molecules using 2D NMR spectra.

In each individual case a strategy for problem solving is chosen using an estimation of the problem complexity. This is possible on the basis of computing the number of connectivity combinations which can be processed during fuzzy structure generation. Preliminary logical analysis of the 2D NMR data allows the reduction of the calculated number of combinations by 10^3-10^5 times. In spite of a large enough number of remaining combinations to be processed (thousands and millions) the speed of the structure generator in StrucEluc is such that it copes with a task in a reasonable time.

The strategy of fuzzy generation has been illustrated in this work using a series of real-world examples including molecules whose 2D NMR spectra contained up to 15 NSCs with lengths varying between four and six bonds. When a data set lacked 2D NMR correlations but was accompanied by the presence of NSCs, then the problem can be solved using the fragments found in a Fragment Databases using a ^{13}C NMR search and employing user proposed substructures. However, the solution of such problems can be difficult since a large number of MCDs can be created from the fragments. To circumvent this difficulty in the future an algorithm for fragment "implementation" in a MCD as well as fuzzy structure generation from fragment-containing MCDs need to be improved. This work is underway.

In the literature examples are described where lengthening all correlations by one bond circumvented the problem of nonstandard correlations in 2D NMR data. This possibility has been investigated here and has shown that even the structures of small molecules with less than 20 carbon atoms in general cannot be elucidated using this approach. The lengthening of connectivities by one bond is evidently in vain when the NSCs are characterized by ^{5-6}J coupling constants present in the 2D NMR data.

REFERENCES AND NOTES

(1) Molodtsov, S. G.; Elyashberg, M. E.; Blinov, K. A., Williams, A. J.; Martin, G. M.; Lefebvre, B. Structure Elucidation from 2D NMR Spectra Using the *StrucEluc* Expert System: Detection and Removal of Contradictions in the Data. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1737−1751.

(2) Elyashberg, M. E.; Blinov, K. A.; Williams, A. J.; Molodtsov, S. G.; Martin G. M. Are deterministic expert systems for computer assisted structure elucidation obsolete? *J. Chem. Inf. Model.* **2006**, *46*, 1643−1656.

(3) 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 Data: Application of the StrucEluc System. *J. Magn Reson. Chem.* **2003**, *41*, 359−372.

(4) Elyashberg, M. E.; Blinov, K. A.; Molodtsov, S. G; Williams, A. J.; Martin, G. E. *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.

(5) Gunther, H. *NMR Spectroscopy: Basic Principles, Concepts, and Applications in Chemistry*, 2nd ed.; Wiley: 1995.

(6) 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.

(7) Han, Y.; Steinbeck, C. Evolutionary-Algorithm-Based Strategy for Computer-Assisted Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 489−498.

(8) Steinbeck, C. Recent Developments In Automated Structure Elucidation Of Natural Products. *Nat. Prod. Rep*. **2004**, *21*, 512−518.

(9) Collins, D. O.; Reynolds, W. F.; Reese, P. B. New Cembranes from *Cleome spinosa*. *J. Nat. Prod*. **2004**, *67*, 179−183.

(10) Mensah, A. Y.; Houghton, P. J.; Bloomfield, S.; Vlietinck, A.; Berghe, D. V. Known and Novel Terpenes From *Buddleja Globosa* Displaying Selective Antifungal Activity Against Dermatophytes. *J. Nat. Prod.* **2000**, *63*, 1210−1213.

(11) Wellington, K. D.; Cambie, R. C.; Rutledge, P. S.; Bergquist, P. R. Chemistry of Sponges. 19. Novel Bioactive Metabolites from *Hamigera tarangaensis*. *J. Nat. Prod.* **2000**, *63*, 79−85.

(12) Oliveira, J. H. H. L.; Grube, A.; Köck, M.; Berlinck, R. G. S.; Macedo, M. L.;Ferreira, A. G.; Hajdu, E. Ingenamine G and Cyclostellettamines G-I, K, and L from the New Brazilian Species of Marine Sponge *Pachychalina* sp. *J. Nat. Prod*. **2004**, *67*, 1685−89.

(13) Elyashberg, E. M.; Martirosian, E. R.; Karasev, Y. Z.; Thiele, H.; Somberg. H. Expert Systems as a Tool for the Molecular Structure Elucidation by Spectral Methods. Strategies of Solution to the Problems. *Anal. Chim. Acta* **1997**, *348*, 443−463.

(14) Elyashberg, M. E.; Blinov, K. A.; Martirosian, E. R.; Molodtsov, S. G.; Williams, A. J.; Martin, G. E. Automated Natural Product Structure Elucidation − the Benefits of a Symbiotic Relationship between the Spectroscopist and the Expert System. *J. Heterocycl. Chem*. **2003**, *40*, 1017−1029.

(15) ACD/NMR predictors-Advanced Chemistry Development, 110 Yonge Street, 14th floor, Toronto, ON, M5H 3V9, Canada, http://www.acd-labs.com: Prediction suite includes ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P NMR prediction.

(16) Bremser, W. HOSE − A Novel Substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355−365.

(17) Kirsch, G.; Kong, G. M.; Wright, A. D.; Kaminsky, R. A New Bioactive Sesterterpene and Antiplasmodial Alkaloids from the Marine Sponge *Hyrtios* cf. *erecta*. *J. Nat. Prod*. **2000**, *63*, 825−829.

(18) Feller, M.; Rudi, A.; Berer, N.; Goldberg, I.; Stein, Z.; Benayahu, Y.; Schleyer, M.; Kashman, Y. Isoprenoids of the Soft Coral *Sarcophyton glaucum*: Nyalolide, a New Biscembranoid, and Other Terpenoids. *J. Nat. Prod.* **2004**, *67*, 1303−1308.

(19) Mdee, L. K.; Yeboah, S. O.; Abegaz B. M. Rhuschalcones II-VI, Five New Bichalcones from the Root Bark of *Rhus pyroides. J. Nat. Prod.* **2003**, *66*, 599−604.
(20) Kehraus, S.; Köning, G. M.; Wright, A. D. A New Cytotoxic Calyculinamide Derivative, Geometricin A, from the Australian Sponge *Luffariella geometrica. J. Nat. Prod.* **2002**, *65*, 1056−1058.
(21) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379−423.
(22) Molodtsov, S. G. Computer-Aided Generation Of Molecular Graphs. *Commun. Math. Chem. (MATCH*) **1994**, *30*, 213−224.
(23) Molodtsov, S. G. The Generation Of Molecular Graphs With A Given Set Of Nonoverlapping Fragments. *Commun. Math. Chem. (MATCH*) **1994**, *30*, 203−212.
(24) Molodtsov, S. G. The Generation Of Molecular Graphs With Obligatory, Forbidden And Desirable Fragments. *Commun. Math. Chem. (MATCH)* **1998**, *37*, 157−162.
(25) Diaz-Marrero, A. R.; Rovirosa, J.; Darias, J.; San-Martin, A.; Cueto, M. Plocamenols A-C, Novel Linear Polyhalohydroxylated Monoterpenes from *Plocamium cartilagineum. J. Nat. Prod.* **2002**, *65*, 585−588.
(26) Klemke, C.; Kehraus, S.; Wright, A. D.; König, G. M. New Secondary Metabolites from the Marine Endophytic Fungus *Apiospora montagnei. J. Nat. Prod.* **2004**, *67*, 1058−1063.
(27) Gu, J. Q.; Graf, T. N.; Lee, D.; Chai, H. B.; Mi, Q.; Kardono, L. B. S.; Setyowati, F. M.; Ismail, R.; Riswan, S.; Farnsworth, N. R.; Cordell, G. A.;. Pezzuto, J. M; Swanson, S. M.; Kroll, D. J.; Falkinham, J. O., III; Wall, M. E.; Wani, M. C.; Kinghorn, A. D.; Oberlies, N. H. Cytotoxic and Antimicrobial Constituents of the Bark of *Diospyros-maritime* Collected in Two Geographical Locations in Indonesia. *J. Nat. Prod.* **2004**, *67*, 1156−1161.
(28) Iken, K. B.; Baker, B. J. Ainigmaptilones, Sesquiterpenes from the Antarctic Gorgonian Coral *Ainigmaptilon antarcticus. J. Nat. Prod.* **2003**, *66*, 888−890.
(29) Fukushi, Y.; Yajima, C.; Mizutani, J.; Tahara, S. Tricyclic Sequiterpenes from *Rudbeckia Laciniata*. *Phytochemistry* **1998**, *49*, 593−600.
(30) Gavagnin, M.; Mollo, E.; Castelluccio, F.; Crispino, A.; Cimino, G. Sesquiterpene Metabolites of the Antarctic Gorgonian *Dasystenella acan. J. Nat. Prod.* **2003**, *66*, 1517−1519.
(31) Heinrich, M. R.; Kashman, Y.; Spiteller, P; Steglich, W. Revision of The Structure of Haliclorensin to (S)-7-Methyl-1, 5-Diazacyclotetradecane and Confirmation of the New Structure by Synthesis. *Tetrahedron* **2001**, *57*, 9973−9978.
(32) Nagle, D. G.; Zhou, Y.-D.; Park, P. U.; Paul, V. J.; Rajbhandary, I. A New Indanone from the Marine Cyanobacterium *Lyngbya majuscula* that Inhibits Hypoxia-Induced Activation of the VEGF Promoter in Hep3B Cells. *J. Nat. Prod*. **2000**, *63*, 1431−1433.
(33) Moon, S.-S.; Lee, J.-Y.; Cho, S.-C. Isotsaokoin, an Antifungal Agent from *Amomum tsao-ko*. *J. Nat. Prod.* **2004**, *67*, 889−891.
(34) Joshi, B. S.; Singh, K. L.; Roy, R. Structure of a new isobenzofuranone derivative from *Nigella sativa* Linn. *Magn. Reson. Chem*. **2001**, *39*, 771−772.
(35) López, J. M. S.; Insua, M. M.; Baz, J. P.; Puentes, J. L. F.; Hernańdez, L. M. C. New Cytotoxic Indolic Metabolites from a Marine *Streptomyces. J. Nat. Prod.* **2003**, *66*, 863−864.
(36) Hernández-Romero, Y.; Rojas, J.-I.; Castillo, R.; Rojas, A.; Mata, R. Spasmolytic Effects, Mode of Action, and Structure-Activity Relationships of Stilbenoids from *Nidema boothii. J. Nat. Prod.* **2004**, *67*, 160−167.
(37) Stærk, D.; Skole, B.; Jørgensen, F. S.; Budnik, B. A.; Ekpe, P.; Jaroszewski, J. W. Isolation of a Library of Aromadendranes from *Landolphia dulcis* and Its Characterization Using the VolSurf Approach. *J. Nat. Prod.* **2004**, *67*, 799−805.

CI600528G