Application of a New Expert System for the Structure Elucidation of Natural Products from Their 1D and 2D NMR Data

Mikhail E. Elyashberg,[†] Kirill A. Blinov,[†] Antony J. Williams,*[,†] Eduard R. Martirosian,[†] and Sergey G. Molodtsov[‡]

Advanced Chemistry Development Inc., 90 Adelaide Street West, Suite 702, Toronto, Ontario, M5H 3V9 Canada, and Novosibirsk Institute of Organic Chemistry, Siberian Branch of Russian Academy of Science, Lavrentiev Avenue 9, Novosibirsk 630090, Russia

Received June 29, 2001

Described herein are applications of the latest version of the StrucEluc expert software system, enhanced to use 2D NMR data, to the structure elucidation of 60 recently isolated natural products. In this study, selected molecules containing between 15 and 65 skeletal atoms and having molecular masses ranging from 200 to 900 amu have been investigated. The correct structure was determined unambiguously for 58 of these molecules. The structures for 75% of the data sets were determined in less than one minute, while 90% of the analyses required no more than 30 minutes. The strategy of structure elucidation by this expert system is described, and several examples are discussed. These illustrate that StrucEluc is a powerful and versatile analytical tool for the structure elucidation of natural products.

A large number of reports have been devoted to the structure determination of natural products using 2D NMR spectroscopy. Not surprisingly, a number of computer-assisted methods, particularly expert systems, have been devised to build on the structure-solving techniques developed in these investigations. By increasing the speed and reliability of structure elucidation, these expert system methods are poised to become significant contributors in the development of new pharmaceuticals. The first-generation expert systems (such as EXPERT,[1] CHEMICS,[2] SESAMI,[3] SpecSolv,[4] and the first version of StrucEluc[5]) used only 1D NMR data in conjunction with other types of analytical information and did not provide a general solution to the structure elucidation issue. Reported data suggest that these applications will usually fail to determine the structure of molecules containing more than about 25 skeletal atoms. This limitation is largely due to the severe underdetermination of the problem using only 1D NMR data in conjunction with other analytical methods. In the early 1990s, the first publications describing second-generation expert systems that used 2D NMR data for structure determination of organic molecules began to appear. New versions of SESAMI,[6,9] CHEMICS,[10] and StrucEluc[11,12] are now available that can use 2D NMR data. In addition, several new programs that can use 2D NMR information have been released. These include CISOCSES,[13−16] LSD,[17] LUCY,[18] and COCON.[19] A review of a number of the 2D NMR expert systems has been recently published by Jaspars.[20]

Studies of the structure determination of specific natural products using some of these systems have been published.[6−9,14−18,21−22] Unfortunately, only some of these reports present detailed information on the structure elucidation process.[7,14−16] In others, this information is limited or missing, making meaningful comparisons between these programs almost impossible. The authors believe a thorough and systematic analysis of the features, molecular size capacity, computational metrics, and limitations of these expert systems using a large number of natural product test cases has not been published. Few summaries of actual structure elucidation case studies exist that discuss the difficulties encountered in the elucidation process and their possible resolutions. In particular, the issue of contradictions generated from 2D NMR data has received relatively little attention and calls for detailed study.

The aim of the present investigation was twofold: first, to address the above gaps in the literature of expert system structure elucidators; and, second, to report on tests of the efficiency of a 2D NMR-based expert system as a routine tool for structure determination of newly isolated natural products. To meet these goals, the StrucEluc system[11,12] has been applied to the structure elucidation of natural compounds using 1D and 2D NMR, IR, and mass spectra. To make these results more easily reproducible, molecules from recently published natural product structure determinations were chosen that contained listings of 2D NMR data.

The 1D version of StrucEluc, comprising two interacting processes, was described previously.[5] This system relied on a knowledge base consisting of 140 000 assigned ^{13}C NMR spectra, a fragments library containing about 500 000 fragments with assigned ^{13}C NMR subspectra, and spectrum−structure correlations for NMR and IR spectra. The program could use all of the 1D NMR, IR, and mass spectra available. Two structure generators utilized different principles, one library-based, the other molecular formula-based, to provide the system with operational flexibility. However, the 1D StrucEluc shared the limitation common to all similar systems constrained to using only 1D NMR data: it failed to analyze molecules containing more than 20−25 skeletal atoms due to the severely underdetermined nature of the structural search.

Recently, a new process[11,12] has been developed that can be used either independently or in conjunction with the two original processes. It is intended primarily for structure elucidation of large organic molecules and is based on the use of 2D NMR data. The use of these data, in combination with a knowledge base, spectrum−structure correlations, candidate structure spectral prediction, and numerous other software aids, has significantly extended the system's

* To whom correspondence should be addressed. Tel: 919-570-0217. Fax: 425-790-3749. E-mail: Tony@acdlabs.com.
[†] Advanced Chemistry Development Inc.
[‡] Novosibirsk Institute of Organic Chemistry.

10.1021/np0103315 CCC: $22.00
Published on Web 04/25/2002

ability to handle molecules containing more than 25 skeletal atoms.

This work describes the results of application of the StrucEluc for structure determination of 60 new compounds from natural sources. 1D and 2D NMR spectra of most of these products were published in the *Journal of Natural Products* during the year 2000. These studies demonstrate that StrucEluc can be successfully applied to the structure elucidation of new and novel natural product molecules containing more than 60 skeletal atoms.

Results and Discussion

The 2D NMR module is based on a number of programs developed for deducing the molecular structure from a combination of 2D NMR spectra. The most typical combination providing an appropriate basis for the structure determination includes $^1H-^{13}C$ HMQC and HSQC (see ref 23; more recently the ADSQC experiment has been applied due to superior line shape and resolution relative to HMQC[24]), $^1H-^1H$ COSY, and HMBC. The StrucEluc system presently operates with the following 2D NMR methods: HMQC, HETCOR, HSQC, HMBC, COLOC, INADEQUATE, COSY, TOCSY, ROESY, and NOESY. The program can utilize data from both $^1H-^{13}C$ and $^1H-^{15}N$ heteronuclear correlation experiments. In addition to the spectral data, the program also needs at least one possible molecular formula.

The following section illustrates the use of the 2D module for the structure elucidation of polycarpol ($C_{30}H_{48}O_2$; compound **1**, Figure 1). The initial data were taken from a literature description of the structure elucidation of polycarpol using the LUCY program.[18]

To begin, the molecular formula, the 1D ^{13}C NMR spectrum, and the coordinates of all cross-peaks (COSY, HMQC, and HMBC) were entered into the program. The 2D data[18] shown in Table 1 were entered into StrucEluc manually. However, it should be noted that, when available, raw data can be processed, analyzed, and easily transferred to StrucEluc using a companion software program.[25] This program allows the processing of both 1D and 2D NMR data as well as reading in preprocessed spectral data files from the primary spectrometer vendors. The program allows automated peak-picking through specific algorithms, but ultimately it is the responsibility of the user to ensure that the appropriate input data set is fed into the StrucEluc program. This can include identification and removal of appropriate artifacts or spectral responses including one-bond responses from HMBC experiments. An advantage of using these companion programs is that the cross-peak intensities are automatically transferred to StrucEluc. When available, StrucEluc uses cross-peak intensity data to estimate the number of skeletal bonds separating correlated nuclei. This is generally more appropriate for COSY experiments rather than long-range heteronuclear experiments. In those situations where short-range couplings may be weak, for example, two-bond HMBC peaks in olefinic or aromatic fragments, the default values will account for the range of possible values. In those cases where the actual detail of the couplings has been proven to be 2-,3- or even higher bond range, the actual bond order of the coupling can be explicitly defined in the input table even though these data are rarely published. Because the cross-peak intensity data were unavailable for the test cases reported here, the program default values were relied on. These assign the ranges of 2−3 and 1−3 intervening bonds between correlated nuclei in COSY and HMBC experiments, respectively. In those situations where a two-bond peak is so weak that it cannot be observed, this may lead to an incomplete representation of the skeletal framework since certain connectivities will not be available. In these cases the structure generation mode of the problem will attempt to fill in the appropriate gaps left by these absences.

Table 1. NMR Data Used for Elucidating the Structure of Polycarpol (**1**)

DEPT	δ_C	C,H-COSY cross-signal with δ_H	HMBC cross-signal with δ_H	H,H-COSY cross-signal with δ_H
C	146.0		6.00, 2.00/2.26, 1.01	
C	142.0		5.30, 1.04	
C	131.0		1.66, 160	
CH	125.5	5.20	1.66, 1.60	
CH	122.0	6.00	2.05/2.14	2.05/2.14
CH	116.3	5.30	2.00/2.26	2.00/2.26
CH	79.0	3.15		1.68
CH	74.8	4.30	1.04	1.83/1.93
C	52.5		1.04, 0.64	
CH	49.7	1.14	1.00, 1.01, 0.91	
CH	49.3	1.65	0.87, 0.64	1.83/1.93
C	44.4		5.30, 1.65, 1.04, 0.64	
CH2	39.8	1.83/1.93		4.30, 1.65
C	39.0		1.14, 1.00, 1.68, 0.91	
CH2	38.8	2.00/2.26	0.64	5.30
C	37.7		1.14, 1.68, 1.01	
CH2	36.8	1.07/1.45	0.87	
CH	36.4	1.32	0.87	0.87
CH2	36.3	1.38/1.92	1.01	1.68
CH3	28.3	1.00	0.94	
CH2	27.9	1.68		3.15, 1.38/1.92
CH3	25.9	1.66	1.60	
CH2	25.3	2.00		
CH2	23.3	2.05/2.14		6.00
CH3	23.0	1.01	1.14	
CH3	18.7	0.87		1.32
CH3	17.8	1.60	1.66	
CH3	16.1	0.64	1.65, 2.00/2.26	
CH3	16.0	0.91	1.14, 1.00	

Once the data were entered, the NMR cross-peak data tables were transferred into tables of carbon atom connectivities. The program is supplied with convenient aids for detecting and removing user mistakes that can arise in the connectivity data during the data entry process. Since the information in these tables provides the foundation for the structure generation process, they must be examined for consistency (i.e., the absence of contradictions). The main cause of contradictions arising from 2D data is during the interpretation of cross-peak information: if the maximum number of intervening bonds assigned to a cross-peak is less than the true number of bonds, a contradiction will arise. Because contradictions can produce wrong connectivities, particularly for heteronuclear long-range correlations across five or more bonds (see primary review),[26] methods have been devised to resolve them.

Contradictions arising from 2D NMR data sets have been discussed in the literature, and an approach that allowed one to perform structure generation in the presence of definite contradictions and ambiguous connectivities was offered.[13−16] Unfortunately, the robustness of this approach was not proven by a significant number of examples, so a method was developed for automatically detecting and resolving contradictions prior to the actual structure generation process. In so doing, a series of specific criteria for contradiction detection was created.[11] In the polycarpol (**1**) example, the contradiction detection routine utilized found that the data taken from the article were fully self-consistent.

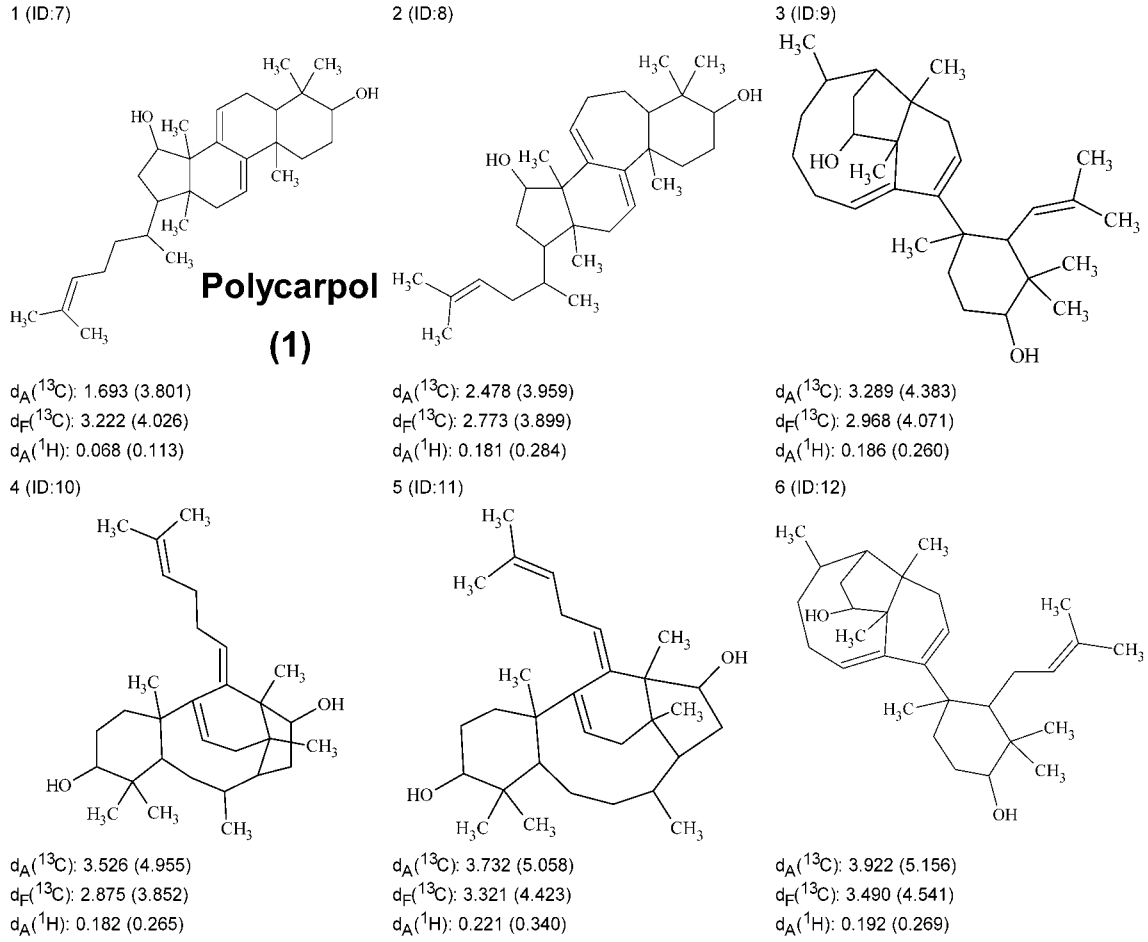

Figure 1. Polycarpol (**1**) structure elucidation. Output structural file ranked in order of increasing d_A values.

Once the data were determined to be free of contradictions, structure generation was conducted without altering any other program default constraints. The generator produced six structures in an elapsed time of 6 s (Celeron operating at 500 MHz, Window98, RAM 128 Mb). Then StrucEluc was used to predict the "accurate" ^{13}C NMR spectra for the six generated structures, ranking the structures in order of increasing average deviation between these accurate estimates and the experimental data, d_A. The correct structure of polycarpol (**1**) had the lowest d_A; the ranked structures are shown in Figure 1. The deviation of the polycarpol molecule was 1.71 ppm; the d_A value of the next-best candidate structure was 2.49 ppm. It should be noted that the LUCY program produced the same six structures, but took 2 h (Pentium PC operating at 100 MHz) and left the selection of the correct structure to the chemist (the presence of spectral prediction aids was not mentioned in ref 18). In StrucEluc, stereochemistry is utilized in the prediction of both ^{1}H and ^{13}C chemical shifts. Certainly *E*/*Z* isomers can be clearly distinguished and extracted automatically. At present, 2D frameworks are provided for the molecules assembled through the elucidation process without specific spatial stereochemistry defined other than for *E*/*Z* isomers. In those cases where the input data provide a direct hit in the assigned structure database the stereochemistry may be distinguished directly.

Features of StrucEluc. StrucEluc provides a graphical user interface (GUI) that allows easy and intuitive visual analysis and editing of connectivity data. Carbon resonance multiplicity, heteroatoms, primary structural blocks (PSBs), and "unattached" hydrogen atoms (usually belonging to a heteroatom) are displayed. The default display properties color-code the different atoms and groups, as well as the different connectivities; the user can easily change the color assignments. This scheme permits "at-a-glance" knowledge of whether two atoms are linked by exactly one bond, exactly two bonds, or a specific range of bonds. Information from different experiments can be viewed or suppressed by clicking appropriate buttons on the toolbar. Values of the associated ^{1}H and/or ^{13}C NMR chemical shifts can also be displayed, if desired, as shown in Figure 2. This figure represents the StrucEluc interface. The upper left corner represents the molecular formula of the molecule represented as distinct nuclear centers. The numbers of quaternary, methane, methylene, and methyl fragments have already been deduced using a combination of the spectral data available. The display options used include the display of the ^{1}H NMR chemical shift below each nuclear center, and the lines drawn between fragments represent the connectivities deduced using the COSY data. The chemist is also able to draw bonds of any multiplicity between appropriate atoms to set user fragments (C=O, O−C=O, O−H, etc.) This provides a quick and intuitive mechanism for entering structural information evident from ^{1}H NMR and/or IR spectra without the need for typing exhaustive tables of numerical data. The rest of the figure represents the chemical shift projections of both the ^{1}H and ^{13}C data extracted from the spectral data inputs. The upper right-hand corner displays a reconstruction of the 2D HMBC data set.

The program analyzes ^{13}C and ^{1}H NMR chemical shifts of CH_n groups (n = 0−3) and automatically sets atom properties for each carbon atom. These properties are the

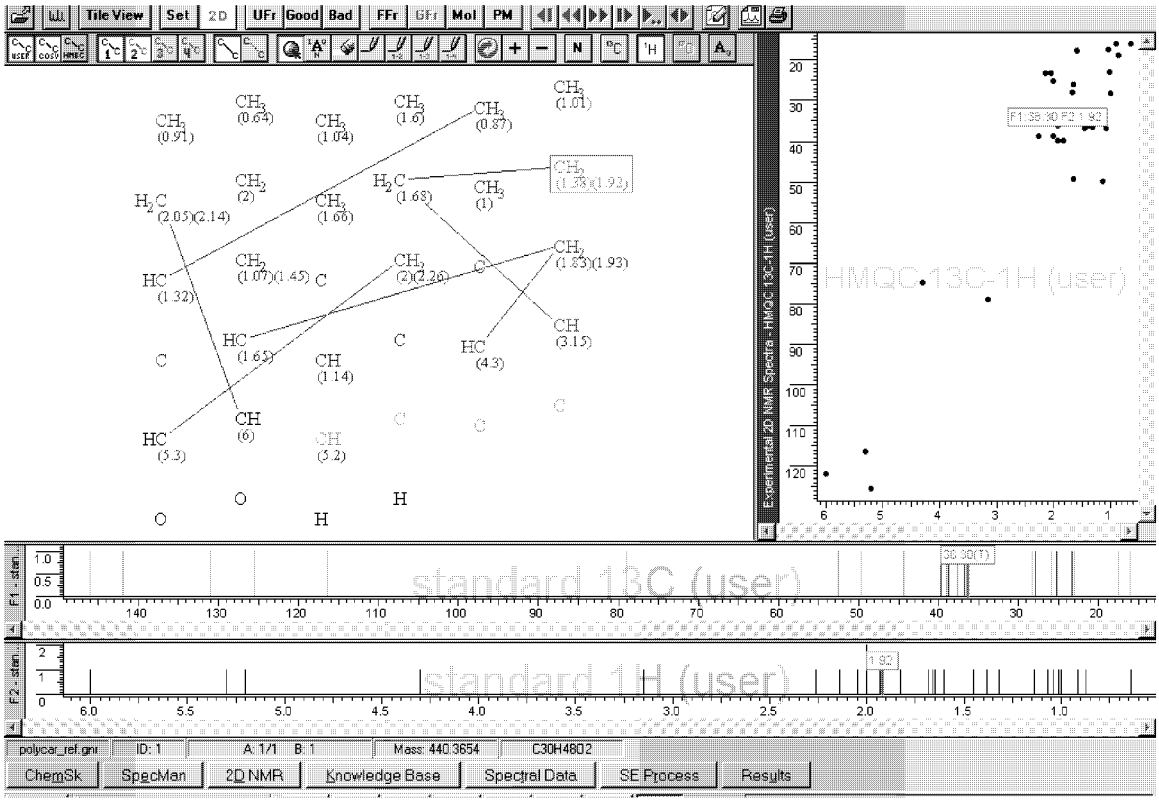

Figure 2. Polycarpol (**1**) ^{1}H−^{1}H COSY NMR connectivities.

allowed hybridizations (sp^3, sp^2, sp, not defined) and the possibilities of heteroatom connectivity (obligatory, forbidden, not defined). To set these atom properties, StrucEluc generates special correlation tables from appropriate fragments and structures stored in the system database. It is very important that both ^{13}C and ^{1}H NMR chemical shifts are taken into account when setting the atom parameters, so this requirement has been built into the program. The chemist can edit these parameters to introduce a priori background information via the connectivity GUI. The data generated at this stage are used as input for the specialized 2D structure generator, which assembles atoms and fragments into larger fragments and structures using the constraints imposed by the topological distances between correlated atoms. While NOESY and ROESY data can be included in the input data, they provide only supporting connectivity information to the COSY and TOCSY homonuclear correlation experiments. At present, these experiments cannot be used for determining stereochemistry directly, but this capability is being examined.

Structures are generated from the structural blocks, PSBs, and user-defined fragments (if entered) under constraints imposed by the tables of carbon atom connectivities. This includes the map of connectivities and any additional constraints imposed by the chemist (maximum bond multiplicity, allowed ring sizes, etc.).

Generated structures can be verified by means of the aids common for all processes of the StrucEluc system (filtering with libraries of spectrum−structure correlations, GOODLIST, BADLIST, etc.). The GOODLIST and BADLIST may be formed automatically, or manually as a result of a fragment search using the approach described in the literature.[5] In addition, application of spectral filtering is very effective and allows the program to reject an enormous number of invalid structures.

As the generation is carried out from structural blocks containing assigned carbon and hydrogen atoms, all the structures generated intrinsically possess assigned spectra. Even within the constraints imposed by long-range connectivity limits, the structure generator may produce identical structures with *slightly* different carbon atom assignments. The reason for this is due to the fact that a series of structure−spectral pairs can be created using the spectral data available to the program. These may result from the combination of the data in subtly different ways, often with the interchange of only two resonances in close chemical shift proximity, to give rise to the same structure but with different nuclear assignments. For this reason, the structural file was checked for structure identity, and all but one structure in each set of duplicated structures was eliminated. The structure retained in each group of duplicates was the one whose intrinsic ^{13}C NMR spectrum best matched the ^{13}C NMR spectrum calculated "from scratch". This process may be performed automatically upon user request through a menu selection.

One of the most important and distinctive attributes of the StrucEluc system is its array of powerful predictive tools. These include several types of spectral estimation and routines for the prediction of many physicochemical properties. StrucEluc is capable of estimating the spectra of candidate structures using the following methods: (1) a fast ^{13}C NMR chemical shift prediction based on increment rules; (2) an accurate ^{13}C NMR chemical shift prediction; (3) ^{1}H NMR chemical shift and, optionally, full spectral prediction including coupling patterns correctly modeling dihedral angle dependencies and second-order effects; (4) calculation of candidate structure correspondence to the experimental mass spectrum with fragmentation (if available).

In the authors' experience, the fast ^{13}C NMR chemical shift prediction provides the best compromise between speed and accuracy for the first level of candidate structure filtering. The fast ^{13}C NMR spectrum prediction is performed for all structures included into the output file, and

the average deviation, d_F, between each of these estimated spectra and the experimental 1D spectrum (or the chemical shifts derived from suitable projections of the 2D data) are calculated. The structures are then ranked by increasing d_F. The speed of the fast spectral prediction (1−2 s per structure, even for large molecules) makes it practical for use as an initial screen of all structure candidates, even for lists containing thousands of structures. This ability is very important, as it frees the user from imposing otherwise unjustifiable constraints (which may exclude the correct structure) in order to make the program finish in a reasonable amount of time.

The correct structure usually falls near the top of the d_F-ranked list. To better discriminate between the most likely candidates, a more accurate ^{13}C NMR spectral calculation is performed for the first 10−20 structures of the d_F-ranked list. A second statistic, d_A, is calculated for each structure from the average deviation between the more accurately predicted and experimental spectra. As discussed below, the correct structure is almost always the one possessing the smallest d_A statistic. Once computed, the user can have StrucEluc sort the structures in ascending d_A order to bring the most likely candidate to the top of the list.

Three more criteria can be used to evaluate structures possessing similar d_A values or to provide independent confirmation of the most likely structure indicated by this parameter. First, the ^1H NMR chemical shifts can be predicted and ranked by deviation from the corresponding shifts obtained from suitable projections of the 2D data. Second, if an experimental 1D ^1H NMR spectrum is available, full ^1H NMR spectra, including not only chemical shift information but expected coupling patterns (derived from database values and adjusted for dihedral-angle dependencies and second-order perturbations) can be predicted for each candidate structure of interest. Visual comparison between these predicted proton spectra and the experimental one provides another criterion for accepting or rejecting candidate structures. During this work it was observed that, when present, the near first-order multiplets associated with methyl groups in a calculated ^1H NMR spectrum are diagnostically valuable and frequently provide convincing fingerprint-like confirmation of the most likely structure. Third, if an experimental mass spectrum is available, the program can analyze each structure for assignable fragments. The resulting percentage of assigned experimental MS peaks provides another criterion for ranking proposed structures. Finally, StrucEluc also incorporates routines for the prediction of several physicochemical parameters (log P, pK_a, boiling point, solubility, surface tension, etc.) When experimental physicochemical data are available, prediction of these properties provides several more independent criteria for potential structure evaluation and aids in further characterizing the compound once its structure is determined.

All but the fast ^{13}C NMR spectral calculations can be improved by prior 3D optimization of the structure underlying the predictions. The 3D optimization algorithm allows the planar (2D) structure to be rapidly translated into a realistic three-dimensional structure. It is based on modified molecular mechanics which take into account bond stretching, angle bending, internal rotation, and van der Waals nonbonded interactions. Modifications include minor simplification of potential functions and enforcement of the minimization scheme by additional heuristic algorithms for dealing with poor starting conformations. The proprietary 3D optimization algorithm uses a force field initially based on classical CHARMM parametrization. The modifications involve some simplification and were intended to increase the stability and speed of computation. It should be emphasized that the 3D optimization, as well as all of the spectral predictions, is performed automatically. They may be applied to the entire list of possible structures generated by the program (or to any user-defined subset of this list) by using a menu to choose the spectra and options required, then clicking a button.

StrucEluc can also be used in a "reverse" mode to check the validity of 2D NMR assignments of a given structure from multiple experiments. If any contradictions are found, the program displays a conflict notification message explaining the cause(s) of the conflict(s).

One of the unique advantages of the StrucEluc is its ability to allow the user to utilize the system knowledge base effectively in solving underdetermined structures with 2D NMR data. In some cases, the total number of 2D NMR correlations is insufficient for structure elucidation. This is usually revealed by a seemingly endless structure generation stage. If such a situation arises, the user has an opportunity to perform fragment searches in the knowledge base, which can be combined with user-generated auxiliary databases (described in the next paragraph), if these were not provided to the program at the start of the data analysis. In such cases, one may use a menu command to initiate a search for fragments whose spectra match parts of the experimental spectra. Any fragments found can be integrated into the 2D connectivities map using another menu command. This can significantly accelerate the structure generation process.

StrucEluc also provides for the automatic creation of a user database containing fragments obtained from a set of assigned structures similar to the molecule under investigation (e.g., derived from the same or related biological sources). These fragments can be employed in two ways: they can be directly drawn into the 2D connectivities map, and they may also be used in automatic fragment set generation. These fragment sets can be used for structure generation as described previously.[5] This approach of creating and using a user fragments database is analogous to the "common sense" method frequently utilized by chemists: to elucidate the structure of a new compound, a chemist often makes use of comparisons and contrasts between data from a new compound and that from similar compounds (or compounds from similar sources). The evident advantage of the system application is that all procedures are performed either automatically or in an interactive mode.

Selection of Test Data. The performance of StrucEluc was initially tested by elucidating the structures of 10 natural products, using 2D NMR data taken from refs 27 and 28 and a series of articles published by authors of previous expert systems.[6−9,14−18] Articles were carefully chosen where the 2D NMR spectral data were presented clearly and without evident mistakes. The goal of these initial studies was not only to test StrucEluc but also to compare the results described here with those obtained by other expert systems. Utilizing StrucEluc, all 10 problems were successfully solved in reasonably short time periods; the results are summarized in Figure 3. The correct structure was identified unambiguously by d_A-ranking in all but one case: structure 3.8 was ranked second.

Comparison of the speed of StrucEluc with other systems was difficult because documentation of processing times was available only for CISOC-SES[13] and LUCY[18] systems. The results described in this work showed that the

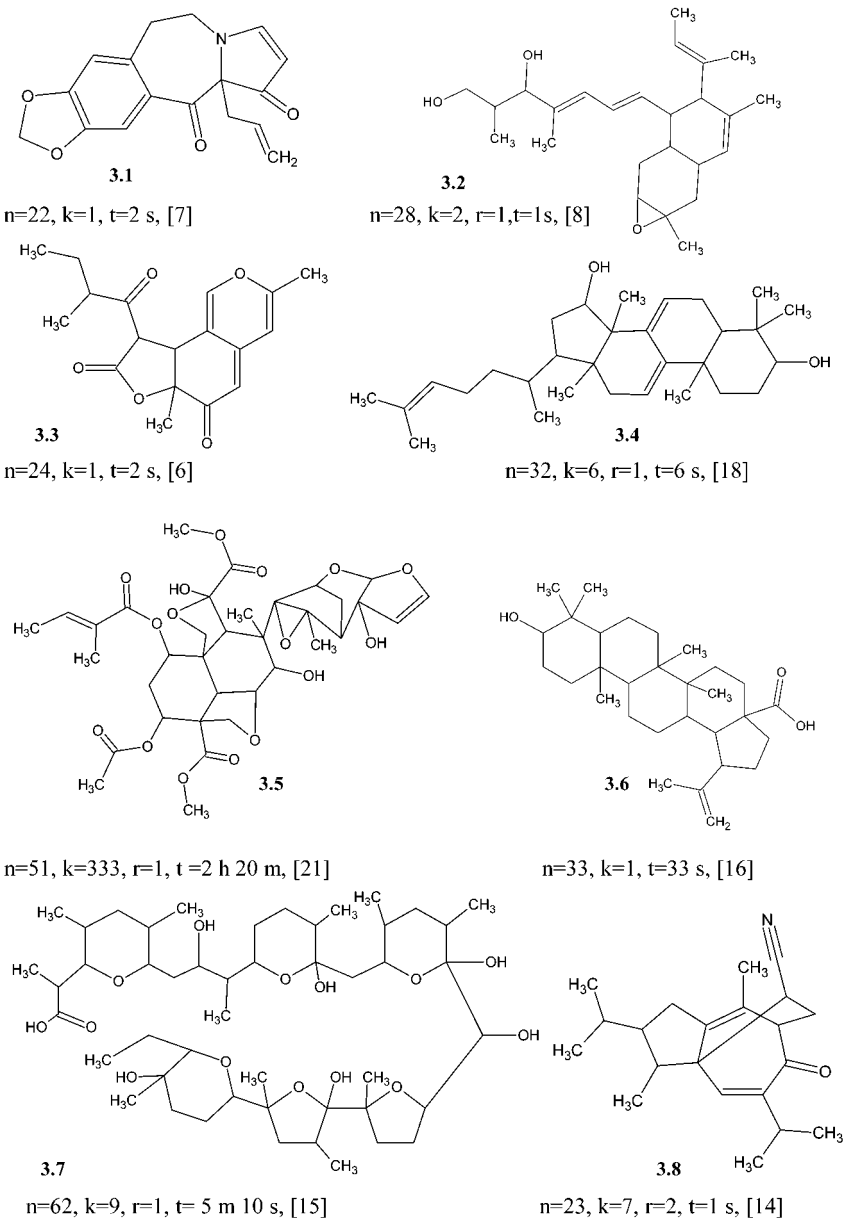

Figure 3. Results of the preliminary system testing. Designations: n = number of skeletal atoms, k = number of structures in an output file, r = position of the right structure in an output file, t = elapsed time of structure generation. Figures in square brackets show the references from which the 2D data were taken.

structure generation speed of CISOC-SES was comparable to that of StrucEluc, while the LUCY system was several orders of magnitude slower than StrucEluc.

To further characterize the performance of StrucEluc in the structure determination of natural products, 2D NMR data from 50 natural products studies[29−62] published in the *Journal of Natural Products* mainly during 2000 were used. Structural formulas of these compounds, together with the StrucEluc results, are shown in Figures 4−6. The 60 molecules totally discussed in this article covered a broad range of masses and skeletal sizes, as illustrated by the histograms presented in Figures 1 and 2 (Supporting Information). In summary, the molecules were reasonably large and fairly complex, ranging in mass from 200 to 900 amu and having between 15 and 65 skeletal atoms.

In these studies, StrucEluc was configured to attempt automatic contradiction resolution whenever the contradiction check found an inconsistency. To resolve these problems, the automated routine increments (by one bond) the upper limit on the number of linking bonds. Ideally, one could preclude contradictions by unambiguously associating each cross-peak with the number of intervening bonds. When available, data from INADEQUATE-type experiments are invaluable in this regard. Unfortunately, the concentration and instrument time requirements of this experiment render such information inaccessible in the majority of cases. When additional experiments are not feasible, attempts to resolve contradictions by educated trial-and-error can be quite successful. Methods for recognizing likely sources of contradiction are discussed below, and strategies for alleviating the most common of these problems are discussed in the next section.

The 2D NMR data in the source articles were presented either in tables of chemical shifts or as structures with graphical correlation schema. Since cross-peak intensities are usually not published, it was not possible to use them as criteria to determine the most probable bond separation between correlated atoms. Therefore, each solution was started using the StrucEluc default values mapping cross-peaks to bond separation; these default values depended

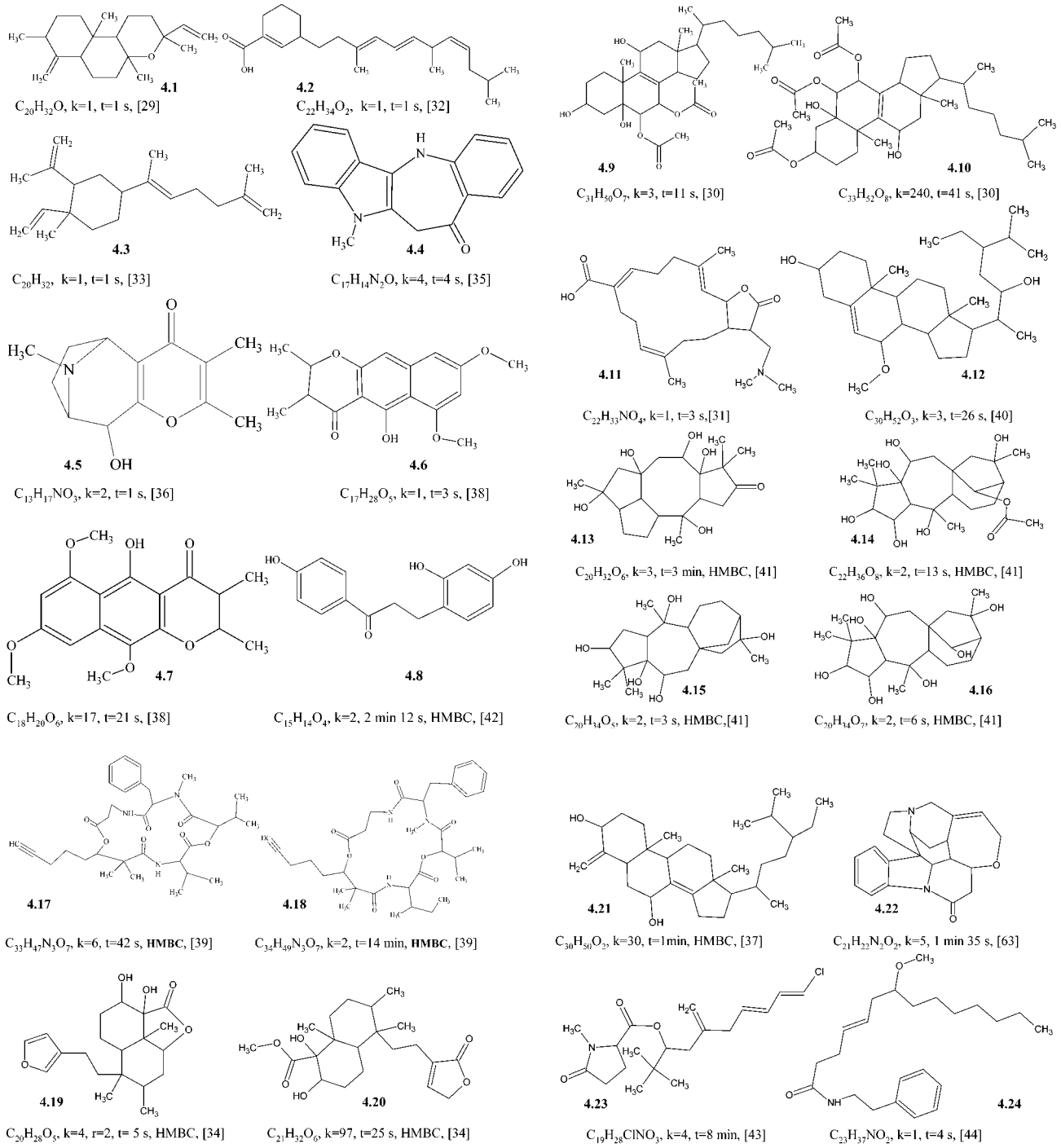

Figure 4. Structures recognized when 2D NMR data either contained no contradictions or the contradictions were removed by the program. Designations: k = number of structures in the output file, r = position of the right structure in an output file, t = elapsed time of structure generation, HMBC = only HMBC long-range correlations were available. Figures in square brackets show the references from which the 2D data were taken.

only on the type of 2D data (COSY, HMBC, etc.). These defaults were appropriate for about 75% of the cases tested. This ratio indicated that the defaults set by the program represent a reasonable compromise between thoroughness and required computational time. Investigations of the source of the problems in the remaining cases indicated that they all had exhaustive lists of long-range connectivities, one or more of which (per structure) fell outside the default limits used by StrucEluc.

An in-depth investigation of those cases that generated contradictions was performed. Most often, the contradictory data arose from COSY correlations (up to five per structure in some cases) or HMBC correlations (again, up to five per structure) across more than three bonds.

Some of these exceptions were very easy to recognize. For instance, it is readily apparent that a problem exists for a CH_2 group assigned three one-bond connectivities to non-hydrogen atoms, or for a CH_3 group assigned more than one one-bond connection to non-hydrogen atoms, based simply on the tetravalent nature of carbon. These contradictions can be resolved manually by replacing all connectivities to a "defective" atom (e.g., one where the tetravalency rule was in violation) with fuzzy (or fuzzier) connectivities. [A fuzzy connectivity is one where a range

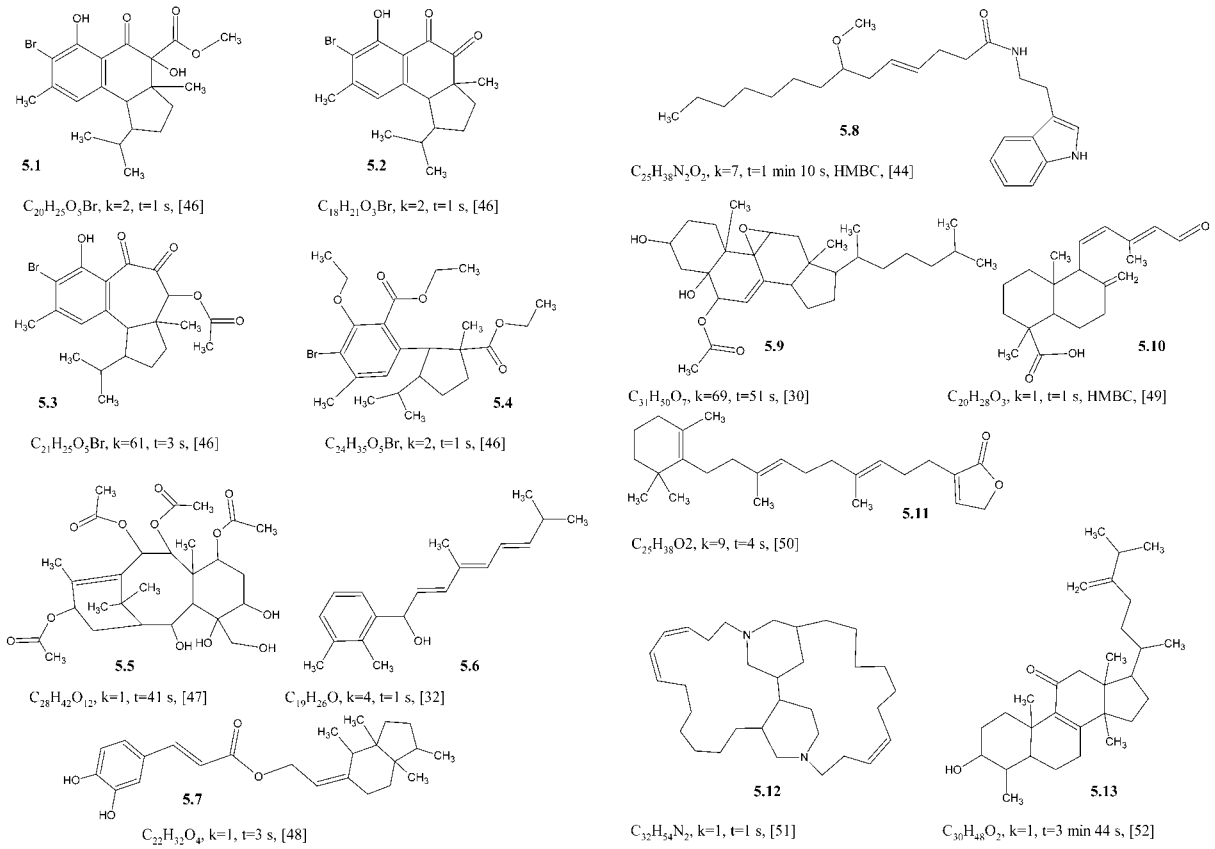

Figure 5. Structures recognized when contradictions were manually removed from the 2D NMR data. Designations: k = number of structures in the output file, t = elapsed time of structure generation, HMBC = only HMBC long-range correlations were available. Figures in square brackets show the references from which the 2D data were taken.

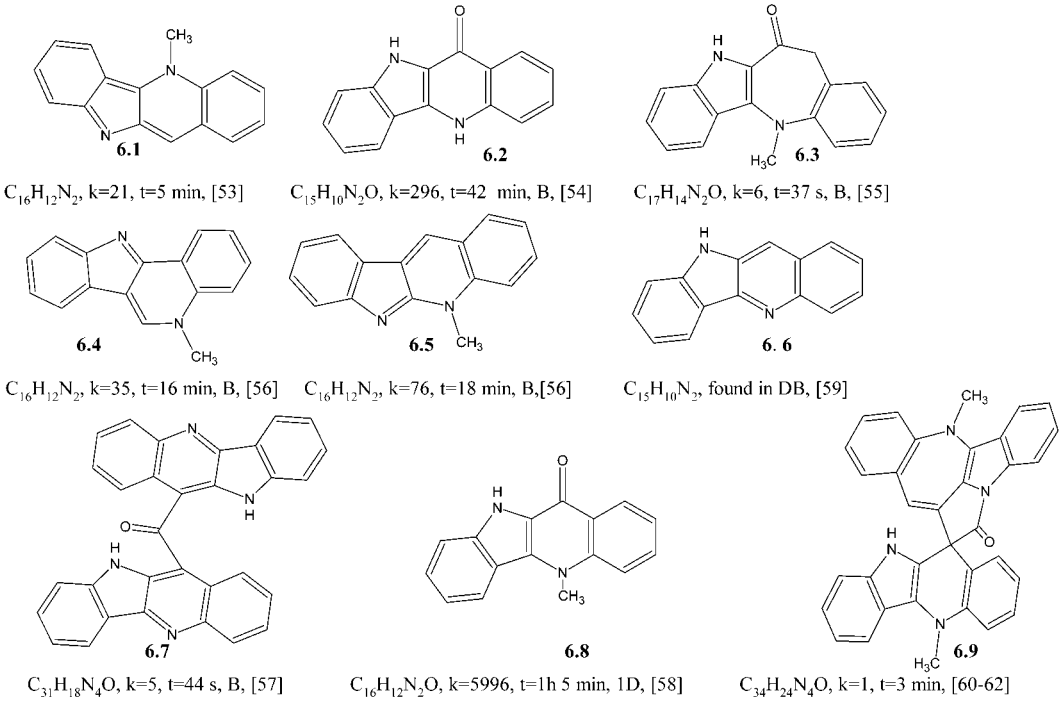

Figure 6. Elucidation of the structure of *Cryptolepis* alkaloids. Designations: k = number of structures in the output file, t = elapsed time of structure generation, B = only HMBC long-range correlations were available. Figures in square brackets show the references from which the 2D data were taken.

of values is specified, as opposed to a definite connectivity such as one expects in HMQC data, where only one-bond (direct) H−C connectivities are expected.]

By increasing the maximum allowed number of intervening bonds by one for all the correlations to the offending atom, contradictions in the data set may be resolved. This action effectively tells the program that one or more of the correlations may represent a $^4J_{HH}$ or a $^4J_{CH}$ rather than a $^3J_{HH}$ or a $^3J_{CH}$ cross-peak. However, the loosened restraints come at the expense of increased computational time and

a larger output list of possible structures. The number of possibilities to be checked increases nonlinearly as constraints become fuzzier, and only one (or a few, allowing for duplication) of the structures found to fit the data can actually be correct. This is the basis of the earlier statement about the appropriateness of the StrucEluc default values for these parameters.

Other sources of contradictory data could be found by analyzing the connectivities map for several simple, well-known situations where coupling can be observed across more bonds than usual. One or more of the well-known mechanisms that propagate *J*-couplings beyond the "normal" number of bonds could usually explain these exceptional cases: allylic, homoallylic, or conjugated systems, and zigzag (*W*-geometry) arrangement of the intervening bonds.

In general, a preliminary contradiction check was run on each test case before launching the fully automated elucidation program, and any simple contradictions were resolved manually. All contradictions must be resolved before the program will start to generate structures. Because manual correction of obvious contradictions such as those presented above is faster than relying on the more rigorous, fully automated resolution protocol, this manual pretreatment of the connectivity parameters helped to decrease the time required to reach a solution.

Several levels of rigor are available for the contradiction tests; normally the "not checked" mode was employed. This meant that there were no constraints (beyond valence considerations) in the bonding to and between carbon atoms. The defaults for the other types of allowed bonding were kept (no heteroatom−heteroatom bonds at all, no increase in connectivity link when merging connectivities). In addition, the "Use NMR Shifts Correlation Table" option was used for all spectra except those with very unusual chemical shifts.

In most cases, if the program did not detect contradictions, the solution set included the correct structure. However, in some cases, the program failed to detect contradictions present in the 2D data, and in those cases the program either refused to generate any structures or produced a solution set that did not contain the correct structure. This occurred because the search for contradictions is based on heuristic principles that do not exhaustively check for all possible contradictions. For instance, connectivities across more than five bonds that are in conflict with other connectivities may not be detected. This represents a design tradeoff: exhaustive contradiction analysis yields diminishing returns and increases processor time.

Contradiction Resolution Strategies. The number of correlations obtained from 2D NMR experiments can number in the hundreds. Therefore, attempts to resolve contradictions by arbitrarily increasing the limit on the number of intervening bonds for *all* correlations will be grossly inefficient. Since there can be only one correct structure, nearly all structures generated by this action will be incorrect. Furthermore, the calculation time requirements would become impractical. Fortunately, experience has shown that the majority of contradictions can be resolved by judicious choice of the structure blocks for which the upper bound on the number of intervening bonds is increased.

This work demonstrated that, in most cases, long-range correlations propagating through more than the "usual" number of intervening bonds involve methyl groups. The intensity of a long-range cross-peak to the three equivalent protons comprising a methyl group is usually much more intense than the corresponding cross-peaks to methylene or methine protons. This is not only because of the increased number of protons in the methyl group but also because this intensity is concentrated in a narrower spectral region, as the number of nuclei coupled to a methyl group with large (>3 Hz) couplings is usually much smaller than for the other types of proton environments. As a result, $^4J_{HH}$ and $^4J_{CH}$ involving the protons of a methyl group are frequently visible in 2D NMR data sets with a cross-peak intensity comparable to those of $^3J_{HH}$ and $^3J_{CH}$ cross-peaks involving only methine and/or methylene groups. This knowledge was applied to those test cases where StrucEluc was unable to automatically resolve contradictions. By incrementing the upper limit on number of linking bonds to CH_3 groups only, nearly 50% of the contradictions were resolved. The resulting increases in the number of possible structures incurred a small time penalty for structure generation and resulted in a substantially larger output file.

In addition to the intensity phenomenon associated with methyl groups, the presence of allylic or conjugated systems and the presence of a *W*-geometry can all extend the transmission range of scalar coupling interactions. Treatment of the structural blocks $=CH_2$ and $=CH$ (usually identifiable on the 2D NMR connectivities map) as described above for methyl groups was also helpful. However, these strategic approaches are most effective when the number of structural blocks requiring adjustment is small, although they may be the only option available if additional NMR experiments are infeasible.

Identification of the Best Candidate Structure. In the experience of the authors, an unusually large d_A value (>5 ppm) for the first structure in a d_A-ranked file often indicates an invalid solution. This conclusion is supported by the histogram presented in Figure 3 (Supporting Information), which shows the distribution of correct answer structures as a function of d_A. It is clear from this figure that, for 85% of the test cases, the deviation did not exceed 4 ppm. Only in a small number of cases did a d_A statistic greater than 5 ppm correspond to a correct structure; one example is cryptomisrine (structure 6.8), which had a deviation of 5.5 ppm. These results indicate that a $d_A > 5$ ppm can be used as a criterion indicating that the solution obtained must be thoroughly checked. Such checks may include a priori information, known to the chemist but not provided to the program, that explains differences between the observed and predicted ^{13}C NMR spectra (e.g., unusually strong solvent effects). They may also include one or more additional NMR experiments, or acquisition of a mass spectrum if one was not already provided to StrucEluc. The use of the percentage assignability of an experimental mass spectrum by StrucEluc as a structure discriminator was discussed above.

Utilization of a User Fragments Database. This investigation showed the high efficiency of the combined use of all three (library, MF, and 2D NMR) processes in the structure generation system described in an earlier section of this article. In particular, compound 6.9[60−62] proved especially challenging; it was not identifiable using 2D H−H, C−H, and N−H correlations due to a large number of missing and/or overlapped cross-peaks. Attempts to generate structures using only these available data proved to be unsuccessful: no structure was generated during several tens of hours. To augment the basic StrucEluc program, it was decided to create a user fragments library from eight biosynthetic precursors of compound 6.9.

The 1D and 2D NMR data for these compounds[53−59] were entered in StrucEluc, and the elucidation process was performed on each one in turn. StrucEluc determined the structures of compounds 6.1−6.5 and 6.7 by 2D NMR correlations; molecule 6.6 was recognized from its 1D ^{13}C NMR spectrum, which was already present in the StrucEluc internal database. Molecules 6.1−6.7, with their ^{13}C NMR assignments, were put into a separate reference file from which StrucEluc automatically created a user fragment library. The eighth compound was more challenging than the first seven. StrucEluc required the user library just described to reach a solution. During the structure generation process for compound 6.8, 20 fragments were selected from the combined internal and user fragments library, and these fragments were used to generate 5996 structures. The correct structure was identified unambiguously by performing an accurate prediction of the ^{13}C NMR spectra identified as the top candidates by d_F rank. The structure with the lowest d_A statistic, 0.27 ppm, was the correct structure.

The assigned cryptolepinone molecule (compound 6.8) was added to the reference file, allowing StrucEluc to enhance the user's fragments database. The search through this enhanced user database using the 1D ^{13}C NMR spectrum of cryptospirolepine (structure 6.9) detected several fragments whose addition to the 2D connectivities map provided much of the information missing from the experimental 2D data. As a result, the structure generation process took only three minutes and produced only one structure—the correct one, cryptospirolepine. This approach is of significant importance, and its details will be published separately.[64]

The example of cryptolepinone illustrates that StrucEluc can successfully detect the correct structure in an output file of nearly 6000 structures, underscoring the high efficiency of the StrucEluc strategy for correct structure identification. Interestingly, it has been shown that large output files were more often the exception than the rule. The high selectivity of the system can be realized by inspection of the histogram in Figure 4 (Supporting Information). The histogram shows that, in 75% of the cases, the output list contains 10 or fewer structures.

The robust nature of the StrucEluc structure evaluation strategy is further supported by the fact that in only 2 of 60 test cases (structure 3.8, Figure 3, and structure 4.19, Figure 4) was the correct structure ranked second by the d_A statistic; in all other cases the d_A-ranking placed the correct structure at the top of the list. This research has shown that even preliminary ranking of the output structures by the d_F statistic placed the correct structure at the top of the list in 80% of the test cases. Given the speed and economy of modern personal computers, the issue of processing time required for structure determination of a new natural product is inconsequential. When the extraction and purification of a new natural product takes months and sometimes even years, a matter of minutes, hours, or even a few days to obtain a structural solution on a PC cannot be considered excessive. Nevertheless, the histogram presented in Figure 5 (Supporting Information) shows that, for 75% of the test cases, the solution was found in under one minute; for 90% of the cases the solution time did not exceed 30 minutes. An analysis of the computation time with respect to the available data showed that almost all tasks requiring more than 10 minutes for solution were solved on the basis of HMBC correlations alone. Clearly, for the structure determination of complex natural products, utilization of H−H COSY data in conjunction with HMBC data is highly desirable.

Recent advances in 2D NMR spectroscopy are catalyzing the widespread use of $^{1}H-^{15}N$ 2D NMR correlations for the structure elucidation of natural compounds (see a recent comprehensive review).[65] Further, new pulse sequences are being developed that differentiate between two- and three-bond long-range heteronuclear correlations.[66] Such precise knowledge of the correspondence between cross-peaks and the number of intervening bonds will streamline the structure generation process and reduce the number of incorrect structures proposed. Continuing advances in spectrometer hardware indicate that 2D NMR spectra from samples as small as about 10 μg will be generally available in the near future as recent reports indicate.[67−69] The integration of these technological advances with expert systems such as StrucEluc will have a tremendous impact on natural products structure determination. The speed at which new structures can be elucidated, combined with the ability of chemists to devote their time to less tedious tasks than structure assignments, will help produce a synergistic increase in the rate at which new pharmaceutical leads become available.

Experimental Section

General Experimental Procedures. The software used for the results reported here was the ACD/Structure Elucidator, StrucEluc, version 5.08, a Windows-based software program composed of a number of separate modules.[70] The entire software suite was composed of 1D and 2D NMR data processing, MS data processing, ^{1}H and ^{13}C NMR chemical shift prediction and assigned structure databasing tools, and an integrated chemical structure drawing program. During the elucidation process the program displayed a number of possible structures, with comparison of on-screen experimental and fragment spectra or, in the case of failure, a set of structural fragments corresponding to portions of the spectrum that were used to assemble the structure of the unknown compound. StrucEluc included filters for ^{1}H NMR, IR peaks, mass spectral (MW) data, and elemental composition and a self-training system. If the accuracy of spectral calculations for a new class of compounds was poor, user databases with experimental chemical shifts could be constructed and utilized during the elucidation process. Both ^{1}H and ^{13}C databases of assigned structures with NMR chemical shifts and coupling constants were available for searching by chemical structure and substructure. The number of entries in each database was >110 000 for ^{1}H NMR and >111 000 for ^{13}C NMR.

Acknowledgment. The authors thank Dr. Gary E. Martin of Pharmacia Corporation for the provision of many challenging data sets to challenge the StrucEluc system discussed in this publication. We especially appreciate the many stimulating conversations regarding the application of H-N heteronuclear correlation experiments to the structure elucidation of natural products. We also thank Dr. Dean Carlson of the Advanced Chemistry Development Technical Support Group, Toronto, for applying his exacting and stringent standards of proofreading to the manuscript. His thorough evaluation of the science reported here significantly enhanced the quality of the submission.

Supporting Information Available: Problem distribution histograms (Figures 1−5) providing a graphical summary of the success rates of StrucEluc for solving the structures for the 60 molecules examined in this work. This information is available free of charge via the Internet at http://pubs.acs.org

References and Notes

(1) (a) Elyashberg, M. E.; Martirosian, E. R.; Karasev, Yu. Z.; Thiele, H.; Somberg, H. *Anal. Chim. Acta* **1997**, *337*, 265−286. (b) Elyashberg, M. E.; Martirosian, E. R.; Karasev, Yu. Z.; Thiele, H.; Somberg, H. *Anal. Chim. Acta* **1997**, *348*, 443−463.

(2) Miyabayashi, N.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18−28.
(3) Munk, M. E.; Christie, B. D. *Anal. Chim. Acta* **1989**, *216*, 57−68.
(4) Will, M.; Fachinger, W.; Richert, J. R. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 221−227.
(5) Elyashberg, M. E.; Blinov, K. A.; Martirosian, E. R. *Lab. Automat. Inf. Manage.* **1999**, *34*, 15−30.
(6) **)** Christie, B. D.; Munk, M. E. *J. Am. Chem. Soc.* **1991**, *113*, 3750−3757.
(7) Munk, M. E.; Velu, V. K.; Madison, M. S.; Robb, E. W.; Baderstscher, M.; Christie, B. D.; Razinger M. In *Recent Advances in Chemical Information II;* Colier, H., Ed.; Royal Society of Chemistry: Cambridge, U.K., 1993; pp 247−263.
(8) Munk, M. E.; Madison, M. S.; Schulz, K.-P.; Korytko A. 13th Workshop, Nov 13−15, 1998, Bad Durkheim, Germany, 1998.
(9) Munk, M. E. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 997−1009.
(10) Funatsu, K.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 190−204.
(11) Blinov, K. A.; Elyashberg, M. E.; Molodtsov, S. G.; Williams A. J.; Martirosian, E. R. *Fresenius J. Anal. Chem.* **2001**, *369*, 709−714.
(12) Elyashberg, M. E.; Blinov, K. A.; Martirosian, E. R. European Conference on Analytical Chemistry (EUROANALYSIS XI), September 3−9, 2000, Lisbon; Book of Abstracts, OC-34.
(13) Peng, C.; Yuan, S.; Zheng, C.; Hui. Y. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 805−813.
(14) Peng, C.; Yuan, S.; Zheng, C.; Hui, Y.; Wu, H.; Ma, K.; Han, X. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 814−819.
(15) Peng, C.; Yuan, S.; Zheng, C.; Shi, Z.; Wu, H. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 539−546.
(16) Peng, C.; Bodenhausen, G.; Qiu, S.; Fong, H. H. S.; Farnsworth, N. R.; Yuan, S.; Zheng, C. *Magn. Reson. Chem.* **1998**, *36*, 267−278.
(17) Nuzillard, J.-M.; Massiot, G. *Tetrahedron* **1991**, *47*, 3655−3664.
(18) Steinbeck, C. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1984−1986.
(19) Lindel, T.; Junker, J.; Koeck, M. *Eur. J. Org. Chem.* **1999**, *3*, 573−577.
(20) Jaspars, M. *Nat. Prod. Rep.* **1999**, *16*, 241−248.
(21) Ley, S. V.; Doherty, K.; Massiot, G.; Nuzillard, J.-M. *Tetrahedron* **1994**, *50*, 12267−12280.
(22) Nuzillard, J.-M.; Connolly, J. D.; Delaude, C. D.; Richard, B.; Zeches-Hanrot, M.; Le Men-Olivier, L. *Tetrahedron* **1999**, *55*, 11511−11518.
(23) Reynolds, W. F.; McLean, S.; Tay, L.-L.; Yu, M.; Enriquez, R. G.; Estwick, D. M.; Pascoe, K. O. *Magn. Reson. Chem.* **1997**, *35*, 455−462.
(24) Hadden, C. E.; Angwin, D. T. *Magn. Reson. Chem.* **2001**, *39*, 1−8.
(25) ACD SpecManager Program. This program is capable of reading data presented in practically all formats common in NMR spectroscopy. http://www.acdlabs.com/products/spec_lab/exp_spectra/
(26) Araya-Maturana, R.; Delgad-Castro, T.; Cardona, W.; Weiss-Lopez, B. E. *Curr. Org. Chem.* **2001**, *5*, 253−263.
(27) Koehn, F. E.; Gunasekera, S. P.; Niel, D. N.; Cross, S. S. *Tetrahedron Lett.* **1991**, *32*, 169−172.
(28) Jimeno, M. L.; Rumbero, A.; Vazquez, P. *Magn. Reson. Chem.* **1995**, *33*, 408−411.
(29) Kinouchi, Y.; Ohtsu, H.; Tokuda, H.; Nishino, H.; Matsunaga, S.; Tanaka, R. *J. Nat. Prod.* **2000**, *63*, 817−820.
(30) Leone, P. A.; Redburn, J.; Hooper, J. N. A.; Quinn, R. J. *J. Nat. Prod.* **2000**, *63*, 694−697.
(31) Rashid, M. A.; Gustafson, K. R.; Boyd, M. R. *J. Nat. Prod.* **2000**, *63*, 531−533.
(32) Capon, R. J.; Miller, M.; Rooney, F. *J. Nat. Prod.* **2000**, *63*, 821−824.
(33) Chai, M.-C.; Wang, S.-K.; Dai, C.-F.; Duh, C.-Y. *J. Nat. Prod.* **2000**, *63*, 843−844.
(34) Ragasa, C. Y.; Cruz, M. C.; Cula, R.; Rideout, J. *J. Nat. Prod.* **2000**, *63*, 509−511.
(35) Zhang, P.; Bierer, D. E. *J. Nat. Prod.* **2000**, *63*, 643−645.
(36) Butler, M. S.; Katavic, P. L.; Davis, R. A.; Forster, P. I.; Guymer, G. P.; Quinn, R. J. *J. Nat. Prod.* **2000**, *63*, 688−689.
(37) Qureshi, A.; Faulkner, D. J. *J. Nat. Prod.* **2000**, **63**, 841−842.
(38) Macias, M.; Ulloa, M.; Camboa, A.; Mata, R. *J. Nat. Prod.* **2000**, *63*, 757−761.
(39) Sitachitta, N.; Williamson, R. T.; Gerwick, W. H. *J. Nat. Prod.* **2000**, 63, 197−200.
(40) Pettit, G. R.; Numata, A.; Cragg, G. M.; Herald, D. L.; Takada, T.; Iwamoto, C.; Riesen, R.; Schmidt, J. M.; Doubek, D. L.; Goswami, A. *J. Nat. Prod.* **2000**, *63*, 72−78.
(41) Li, C.-J.; Wang, L.-Q.; Chen, S.-N.; Qin, G.-W. *J. Nat. Prod.* **2000**, *63*, 1214−1217.
(42) Gonzales, A. G.; Leon, F.; Sanchez-Pinto, L.; Pardon, J. I.; Bermejo, J. *J. Nat. Prod.* **2000**, *63*, 1297−1297.
(43) Gallimore, W. A.; Galario, D. L.; Lacy, C.; Zhu, Y.; Scheuer, P. J. *J. Nat. Prod.* **2000**, *63*, 1022−1026.
(44) Tan, L. T.; Okino, T.; Gerwick, W. H. *J. Nat. Prod.* **2000**, *63*, 952−995.
(45) Nagle, D. C.; Zhou, Y.-D.; Park, P. U.; Paul, V. J.; Rajbhandary, I. *J. Nat. Prod.* **2000**, *63*, 1431−1433.
(46) Wellimgton, K. D.; Cambie, R. C.; Rutledge, P. S.; Bergquist, P. R. *J. Nat. Prod.* **2000**, *63*, 79−85.
(47) Shen, Y. C.; Lo, K.-L.; Chen, C.-Y.; Kuo, Y.-H.; Hung, M.-C. *J. Nat. Prod.* **2000**, *63*, 720−722.
(48) Burgess, E. J.; Larsen, L.; Perry, N. B. *J. Nat. Prod.* **2000**, *63*, 537−539.
(49) **(**49) Iwamoto, M.; Ohtsu, H.; Matsunaga, S.; Tanaka, R. *J. Nat. Prod.* **2000**, *63*, 1381−1383.
(50) Kirsch, G.; Kong, G. M.; Wright, A. D.; Kaminsky, R. *J. Nat. Prod.* **2000**, *63*, 825−829.
(51) Torres, Y. R.; Berlinck, R. G. S.; Megalhaes, A.; Schefer, A. B.; Ferreira, A. G.; Hajdu, E.; Muricy, G. *J. Nat. Prod.* **2000**, *63*, 1098−1105.
(52) Tanaka, R.; Kasubuchi, K.; Kita, S.; Tokuda, H.; Nishino, H.; Matsunaga, S. *J. Nat. Prod.* **2000**, *63*, 99−103.
(53) Ablordeppey, S. Y.; Hufford, C. D.; Borne R. F.; Dwumu-Badu, D. *Planta* **1990**, *56*, 416−417.
(54) Crouch, R. C.; Davies, A.; Spitzer, T. D.; Martin, G. E.; Phoebe, C. H., Jr.; Sharaf, M. H. M.; Schiff, P. L., Jr.; Tackie, A. N. *J. Heterocycl. Chem.* **1995**, *32*, 1077−1080.
(55) Sharaf, M. H. M.; Schiff, P. L., Jr.; Crouch, R. C.; Davies, A.; Andrews, C. W.; Martin, G. E.; Phoebe, C. H., Jr.; Tackie, A. N. *J. Heterocycl. Chem.* **1995**, *32*, 1631−1636.
(56) **(**56) Sharaf, M. H. M.; Schiff, P. L., Jr.; Martin, G. E.; Phoebe, C. H., Jr.; Tackie, A. N. *J. Heterocycl. Chem.* **1996**, *33*, 239−243.
(57) Sharaf, M. H. M.; Schiff, P. L., Jr.; Minick, D.; Johnson, R. L.; Crouch, R. C.; Andrews, C. W.; Martin, G. E.; Phoebe, C. H., Jr.; Tackie, A. N. *J. Heterocycl. Chem.* **1996**, *33*, 789−797.
(58) Hadden, C. E.; Duholkc, W. K.; Guido, J. E.; Robins, R. H.; Martin, G. E.; Sharaf, M. H. M.; Schiff, P. L., Jr. *J. Heterocycl. Chem.* **1999**, *36*, 525−531.
(59) Spitzer, T. D.; Crouch, R. C.; Martin, G. E.; Sharaf, M. H. M.; Schiff, P. L., Jr.; Tackie, A. N.; Boye, G. L. *J. Heterocycl. Chem.* **1991**, *28*, 2065−2070.
(60) Tackie, A. N.; Boye, G. L.; Sharaf, M. H. M.; Schiff, P. L.; Crouch, R. C.; Spitzer, T. D.; Johnson, R. L.; Dunn, J.; Minick, D.; Martin, G. E. *J. Nat. Prod.* **1993**, *54*, 653−670.
(61) Hadden, C. E.; Martin, G. E.; Tackie, A. N.; Schiff, P. L., Jr. *J. Heterocycl. Chem.* **1999**, *36*, 1115−1117.
(62) Martin, G. E.; Hadden, C. E.; Tackie, A. N.; Sharaf, M. H. M.; Schiff, Jr., P. L. *Magn. Reson. Chem.* **1999**, *37*, 529−537.
(63) 2D NMR spectra of strychnine were kindly provided to us by Dr. Gary E. Martin, Pharmacia, Kalamazoo, MI.
(64) Williams, A. J.; Elyashberg, M. E.; Blinov, K. A.; Molodtsov, S. G.; Martirosian, E. R.; Martin, G. E.; Hadden, C. E. *Magn. Reson. Chem.* (manuscript in preparation).
(65) Martin, G. E.; Hadden, C. E. *J. Nat Prod.* **2000**, *63*, 543−585.
(66) Krishnamurthy, V. V.; Russell, D. J.; Hadden, C. E.; Martin, G. E. *J. Magn. Reson.* **2000**, *146*, 232−239.
(67) Martin, G. E.; Guido, J. E.; Robins, R. H.; Sharaf, M. H. M.; Schiff, P. L., Jr.; Tackie, A. N. *J. Nat. Prod.* **1998**, *61*, 555−559.
(68) Russell, D. J.; Hadden, C. E.; Martin, G. E.; Gibson, A. A.; Zens, A. P.; Carolan, J. L. *J. Nat. Prod.* **2000**, *63*, 1047−1049.
(69) Reynolds, W. F.; Yu, M.; Enriquez, R. G. *Magn. Reson. Chem.* **1997**, *35*, 614−618.
(70) A detailed explanation of the modules comprising the StrucEluc program can be found at http://www.acdlabs.com/products/spec_lab/complex_tasks/str_elucidator/.

NP0103315