Applying Computer-Assisted Structure Elucidation Algorithms for the Purpose of Structure Validation: Revisiting the NMR Assignments of Hexacyclinol

A. J. Williams,*[,†] M. E. Elyashberg,[‡] K. A. Blinov,[‡] D. C. Lankin,[§] G. E. Martin,[⊥] W. F. Reynolds,[∥] J. A. Porco, Jr.,[▽] C. A. Singleton,[▽,⬡] and S. Su[▽,⬢]

ChemZoo, 904 Tamaras Circle, Wake Forest, North Carolina 27587, Advanced Chemistry Development, Moscow Department, 6 Akademik Bakulev Street, Moscow, 117513, Russian Federation, Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois 60612-7231, Schering-Plough Research Institute, Rapid Structure Characterization Laboratory, Pharmaceutical Sciences, Summit, New Jersey 07901, Department of Chemistry, University of Toronto, 80 St George St., Toronto, Ontario M5S 3H6, Canada, and Department of Chemistry and Center for Chemical Methodology and Library Development, Boston University, 24 Cummington Street, Boston, Massachusetts 02215

Received October 3, 2007

Computer-assisted structure elucidation (CASE) using a combination of 1D and 2D NMR data has been available for a number of years. These algorithms can be considered as "logic machines" capable of deriving all plausible structures from a set of structural constraints or "axioms", defined by the spectroscopic data and associated chemical information or prior knowledge. CASE programs allow the spectroscopist not only to determine structures from spectroscopic data but also to study the dependence of the proposed structure on changes to the set of axioms. In this article, we describe the application of the ACD/Structure Elucidator expert system to help resolve the conflict between two different hypothetical hexacyclinol structures derived by different researchers from the NMR spectra of this complex natural product. It has been shown that the combination of algorithms for both structure elucidation and structure validation delivered by the expert system enables the identification of the most probable structure as well as the associated chemical shift assignments.

In recent years, computer-assisted structure elucidation (CASE) has offered an additional option to scientists challenged by difficult chemical structure elucidation problems. The reasons for considering CASE methods should be obvious: there may be significant time savings when applying algorithmic approaches; the subjective bias of a scientist can be reduced relative to an algorithm; the thought process associated with the analysis of a large and diverse data collection made up of spectrum-structural information used to elucidate a chemical structure offers a significant logical and combinatorial challenge. The nature of this process was already revealed in the pioneering works published in the late 1960s and 1970s.[1–4] When scientists attempt to solve structural problems, they logically draw definitive structural conclusions from a set of spectroscopic data and *a priori* knowledge assembled into a set of initial axioms. It has been shown[5] that the definition of a spectrum-structural problem is equivalent to formulating a specific *axiomatic theory*.

The solution of a structural problem using 2D NMR spectroscopy data can be divided into four main stages:

1. Prepare the experimental data to create spectrum-structure axioms that serve as the basis of the structure elucidation process. This stage encompasses both raw data processing and peak picking. 1D and 2D peak tables are produced as an output from this stage.

2. Create axioms and hypotheses on the basis of NMR peak tables. The information contained in the peak tables is considered as *true* and *consistent*. The following axioms are the most common: (a) If the hydrogen atom H(i) shows an HSQC (HMQC) correlation with the carbon atom C(i), then the atom H(i) is attached to the atom C(i); (b) if the hydrogen atom H(i) attached to the carbon atom C(i) shows a COSY correlation with the hydrogen atom H(j) attached to the carbon atom C(j), then carbons C(i) and C(j) make up a chemical bond in a molecule; (c) if the hydrogen atom H(i) shows an HMBC correlation to a carbon atom C(j), then the distance between those atoms in a molecule is 2 or 3 bonds; (d) if the hydrogen atom H(i) shows a NOESY (ROESY) correlation with the hydrogen atom H(j), then the distance through space between atoms H(i) and H(j) is less than 5 Å. Additional structural constraints are produced on the basis of *chemical knowledge* including information regarding sample origin and associated structural knowledge as well as Bredt's rule, etc.

3. The logical inference of *all* direct structural conclusions, if one is possible, comes from following the set of axioms outlined above. These conclusions form a chemical structure file where each structure within the file has to satisfy *all* the axioms applied (i.e., be entirely consistent with the experimental data) and has to contain the assigned experimental chemical shifts.

4. The verification of all structural hypotheses proceeds in order to choose the most probable solutions consistent with the data. For this purpose the prediction of NMR spectra associated with the candidate structures in combination with specific chemical considerations is applied and hypothetical structures are rank ordered based on the agreement between the experimental data and the algorithmically generated structures.

If all initial axioms used to deduce the structural formula of an unknown contain no contradictions, then the structural file produced will be valid[6] and will contain the actual structure. It is evident that if at least one of the axioms in the set is false or conflicts with the other axiom(s), then the correct structure will not be determined. A contradictory system of axioms simply produces no structure. Research has shown that the information obtained from 2D NMR spectra can frequently be fuzzy, contradictory, and incomplete.[7–10] These problems primarily arise due to the presence of long-range 2D NMR correlations that span more than the typical number of bonds allowed by the axioms outlined above (so-called nonstandard correlations, NSC[8]) or severe overlap in the NMR spectra. Overlap

* To whom correspondence should be addressed. Tel: +919.341 8375. Fax: +919 300-5321. E-mail: antony.williams@chemspider.com.
† ChemZoo.
‡ Advanced Chemistry Development, Moscow Department.
§ University of Illinois at Chicago.
⊥ Schering-Plough Research Institute.
∥ University of Toronto.
▽ Boston University.
⬡ Current address: Waters Corporation, 34 Maple St., Milford, MA, 01757.
⬢ Current address: The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037.

10.1021/np070557t CCC: $40.75
Published on Web 02/08/2008

leads to equivocal signal assignments in the 2D spectra and consequently to ambiguities in the spectroscopic interpretations. These possibilities mean that the initial axioms may fail (for instance, the hydrogen atom H(*i*) in the real molecule is attached not to the carbon atom C(*i*) but to another carbon atom) and some axioms can be violated (for instance, the topological distance between two HMBC correlated atoms H(*i*) and C(*j*) is longer than three bonds ($^3J_{CH}$)).

To overcome these difficulties, a human expert employs all of their knowledge, experience, and intuition. One common approach is the suggestion of a structure for the unknown using a series of similar molecules as a basis that have assigned chemical shifts in both ^{1}H and ^{13}C NMR spectra. The comparison of the molecules with the proposed structure and its NMR spectra allows chemical shift assignment that appears to be consistent with the suggested structure. The correctness of the assignments is validated by checking the consistency of the chemical shifts with the topological distances between the intervening atoms: an assignment is considered acceptable if all distances are in agreement with the postulated axioms. The problem is reduced to either confirming or refuting the proposed structure. Commonly, some fragments of the unknown compound can be inferred by the chemist through an understanding of the origin of the sample (reactants, plant genus, etc.), and these fragments can also be utilized in the elucidation process.

Data analysis can vary from simple to very complex depending on the nature of the problem. For complex problems chemists can arrive at a structure for the unknown compound that is incorrect. Our experience[7–10] shows that CASE methods offer the possibility of dramatically accelerating the process of structure elucidation and validation, as well as increasing the reliability of the derived structure.

Expert systems now exist that are capable of modeling all stages of the structure elucidation process via mathematical algorithms implemented in the corresponding software programs. To date the most advanced CASE system is ACD/Structure Elucidator (StrucEluc), which has been described in a large number of publications.[7,10,11] The StrucEluc system mimics the sequence of steps performed by spectroscopists during the process of structure elucidation. It should be noted that inside the program the structure elucidation process is combined with structure verification in such a way that the structure verification procedure can be carried out in an independent manner. StrucEluc is supplied with an extensive and branched knowledge base, sophisticated algorithms for inferring plausible structures from an initial set of axioms, and multiple algorithms for the accurate and rapid prediction of ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P nuclei chemical shifts. The efficiency of this system has been amply demonstrated by solving hundreds of structural problems.[9] Moreover, the application of StrucEluc allowed authors[12] to recognize the structure of a complex natural product that had previously remained unsolvable by highly qualified spectroscopists.

In this article we will demonstrate how the StrucEluc system can be applied to the validation of different hypothetical structures derived by different researchers from the same initial experimental data. For this aim, two important capabilities of StrucEluc were utilized: (a) the system makes it possible to follow the impact of how changes in the initial axioms influence the inferred structures; (b) the application of spectrum prediction for different proposed structures allows comparison of different structural hypotheses and allows the choice of the most probable structure(s).

This study was initiated by a discussion in the literature and on Web site blogs regarding the complex chemical structure of a newly identified natural product, hexacyclinol.[13–16] The essence of the problem will be explained in the next section.

The History of Hexacyclinol and Its Various Structural Forms

Hexacyclinol was first described by Gräfe and co-workers in 2002.[13] The compound was isolated in Siberia from basidiospores collected from *Panus rudis* strain HKI 0254. The proposed structure (Figure 1a) contained a reactive epoxyketone and a highly strained endoperoxide moiety.

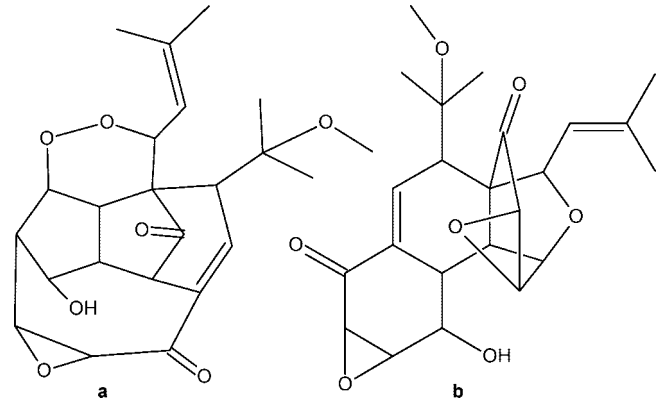


Figure 1. Two different hexacyclinol structures proposed by Gräfe et al.[13] (a) and by Rychnovsky[15] (b).

The total synthesis of the structure proposed by Gräfe[13] was first reported by La Clair in 2006.[14] Shortly after this work, the structure of hexacyclinol was revisited by Rychnovsky based on calculated ^{13}C NMR chemical shift correlations.[15] To determine the structure of hexacyclinol, highly oxygenated and unsaturated molecules were compared with predictive NMR calculations. When evaluating the chemical shift difference between calculated and experimental hexacyclinol, it was found that the Gräfe[13] structure of hexacyclinol had an unusually high deviation, with an average of 6.8 ppm difference, and five carbon chemical shifts with more than a 10 ppm difference from the calculated structure. Using GIAO NMR predictions, the structure of hexacyclinol was revised from the one proposed by Gräfe[13] to the structure shown in Figure 1b.[15]

Following this, the revised hexacyclinol structure was indeed synthesized by Porco and co-workers.[16] The analytical data were identical to that reported for natural hexacyclinol for ^{1}H and ^{13}C NMR and optical rotation studies. In addition, an X-ray crystal structure was obtained, providing unequivocal structural confirmation. These data allow for a thorough assessment of the power of CASE systems in elucidating heretofore unknown, complex natural products.

Structure Comparison Using ACD/Labs Methods

The initial question identified for investigation was assuming that the NMR data obtained by Gräfe et al.[13] are appropriately recorded and accurately reported. The NMR data presented in the original publication are shown in Table 1. These data can be considered as the initial set of spectroscopic-structural axioms defining the given problem.

The molecular formula, $C_{23}H_{28}O_7$, and the spectroscopic information from the ^{1}H, ^{13}C, COSY, HMQC, and HMBC data sets were manually input into the StrucEluc program. A molecular connectivity diagram (MCD)[7–11] was created using the criteria that chemical bonds between oxygen atoms would be allowed, but triple bonds are forbidden (see Supporting Information Figure A). Using the most reliable multiplicities of the ^{1}H NMR signals presented in Table 1 (where the measured coupling constants coincide for intervening nuclei), the following numbers of attached hydrogen atoms were set for specific carbon nuclei: 18.6 (0), 26.1(0), 120.7 (1), 75.8 (1), 47.8 (2), 139.6 (1), 40.9(1), 26.6(0), 24.7(0), 49.1(0). This additional structural information significantly reduces the number of plausible structures. Note that the atom properties and the number of attached hydrogen atoms displayed in the MCD complement the system of axioms given in Table 1.

According to the methodology of the StrucEluc application the MCD was checked for the presence of contradictions using the usual

Table 1. NMR Data Utilized by Gräfe et al.[13] for Structure Elucidation

position	δC	δH	COSY	HMBC
1	18.6 q	1.77 s		142.2, 120.7
2	142.2 s			
3	26.1 q	1,72 s		142.2, 120.7
4	120.7 d	4.82 d, 10.1	H-5	
5	75.8 d	5.46 d, 10.1	H-4	60.5, 202.9
6	60.5 s			
7	202.9 s			
8	53.1 d	3.23 d, br, 3.5	H-9, **H-10**	202.9, 54.5
9	54.5 d	3.64 m	H-8, H-10, H-13	47.8
10	47.8 d	2.74, dd, 5.2, 7.8	H-9, H-11	54.5, 60.5
11	71.5 d	4.99 dd,5.2 br	H-10, H-12	
12	40.4 d	3.55 m	H-11, H-13	61.0, 71.5, 72.7
13	72.7 d	3.8 dd, 9.5, 1.5, 2.54 br (OH)	H-12, H-9	40.4, 54.5, 47.8, 53.1
14	61.0 d	3.51 dd, 2.9, 0.5	H-12, H-15	40.4
15	53.2 d	3.29 d, 3.2	H-14	132.5, 192.8
16	192.8 s			
17	132.5 s			
18	139.6 d	6.73 dd 5.3, 2.4 (allyl)	H-19	192.8, 40.9
19	40.9 d	3.59 d, 5.3	H-18	139.6
20	77.3 s			
21	26.6 q	1.26 s		77.3, 40.9
22	24.7 q	1.15 s		77.3, 40.9
23	49.1 q	3.02 s		77.3

assumptions,[8] and the program reported that the analyzed data were consistent. As has been reported elsewhere,[10] the most appropriate structure generation fuzzy mode uses $m = 0$–15, $a = 16$, with generation stopped at $m = m_g$. m is the number of connectivities to be lengthened or removed and a is the number of bonds to be added to the connectivity length ($a = 16$ represents the removal of connectivities), m_g is the m value at which the resultant structural file is not empty. It is evident that if $m = 0$ (when the data contain no NSCs), then strict structure generation will be carried out. The fuzzy structure generation procedure was initiated utilizing structural filtering via spectroscopic libraries, the permanent BADLIST (containing fragments that are unlikely in organic chemistry), and the checking of chemical structures by Bredt's rule. Four-membered rings were forbidden since these are relatively rare in natural products. These constraints accelerate the process of structure generation.

As a result of structure generation, 5425 molecules were generated at $m_g = 1$ (denoted as $k = 5425$). The structure generation process took $t_g = 56$ s. ^{13}C NMR chemical shift prediction was carried out for all generated structures by both the incremental and the neural net approaches using ACD\CNMR Predictor.[17] The calculations consumed 66 s. The resulting file was reduced to 4961 structures after removal of duplicates ($k = 5425 \rightarrow 4961$). Structures containing the O-OH group were filtered out since this fragment was deemed to be unlikely in natural products. Excluding hydroperoxide moieties reduced the k value to 3097. According to the methodology described earlier,[7,11] structures were ranked in order of the increasing deviation between the experimental and calculated chemical shifts of the predicted structures. Four structures at the top of the file ranked by the deviation of spectra calculated via increments d_I(C) are shown in Figure 2.

Figure 2 shows the most probable structure when ranking is performed by d_I(C). Figure 3 displays the COSY and HMBC connectivities of this structure. The next three structures are very similar to Tanimoto structure similarity coefficients of 0.99, 0.99, and 0.97, and two of them are distinguished as preferable using other calculation approaches. Structure 3 in Figure 2 is distinguished as the best structure overall by both the ^{13}C HOSE and ^{1}H neural net calculations as well as by the complex match factor. Structure 4 was ranked as the best structure by the ^{13}C neural net approach (the upper right-hand corner of the structure boxes in the quadrants displayed in Figure 2 are labeled appropriately). The deviations calculated for Gräfe's structure are relatively large, but when the complexity of the structure is taken into account, they are certainly acceptable.

All of these structures contain only one NSC ($a = 1$) between the carbon resonances at 47.8 ppm and 53.1 ppm. The presence of at least one NSC is frequently observed for organic molecules. Since the structures are both very similar and very complex, the applied methods of chemical shift prediction cannot reliably prove or refute any of the structures. The results obtained show that the structure suggested by Gräfe et al.[13] from the data displayed in Table 1 was also inferred by the CASE algorithms as one of the most probable. The location of the structure proposed by Gräfe[13] in the CASE output file should be considered as a consequence of the total set of initial axioms and assumptions used by Gräfe, including the allowance of the presence of one NSC and O−O bond in the target structure.

The number of NSCs in a structure could be very large, as has been reported in our earlier work.[10] It is interesting to identify what type or types of structure(s) are generated if we allow the presence of up to four NSCs in the analyzed structures as well as forbidding the peroxide group, as was suggested for Rychnovsky's structure.[15] Setting a larger number of possible NSCs is too time-consuming for this structural problem. Fuzzy structure generation was performed in the secure mode {$m = 4$, step by step, $a = 16$}, implying that all m values between 0 and 4 will be taken into account.[10] To prevent the production of too large an output file, incremental ^{13}C NMR spectrum prediction was performed during structure generation and only structures that had $d_I(^{13}C)$ value less than 4 ppm were saved to disk. The results obtained were as follows: 290 141 molecules were generated and 503 molecules were stored, removal of duplicates gave $k = 503 \rightarrow 386$, and the process was performed in a generation time of $t_g = 16$ min 42 s. *All* of the structures in the output file contained two epoxy-containing cycles as represented in Rychnovsky's structure. The "best" structure (**1**) distinguished by the minimum sum of deviations $D(\text{sum}) = d_A(\text{C}) + d_I(\text{C}) + d_{Net}(\text{C})$ when the structures were ranked by the complex match factor $d_A(\Sigma) = d_A(\text{C}) + d_A(\text{H})$ is shown below (structure **1**).

Four nonstandard COSY correlations are shown by the two-headed arrows, and one of them between the proton bearing carbon atoms resonating at 40.40 and 61.00 ppm was unusually long, a $^{6}J_{HH}$ coupling. There is no convincing explanation of why this coupling would exist. Comparison between the deviations determined for Gräfe's structure (Figure 3) and the structure **1** are given in Table 2.

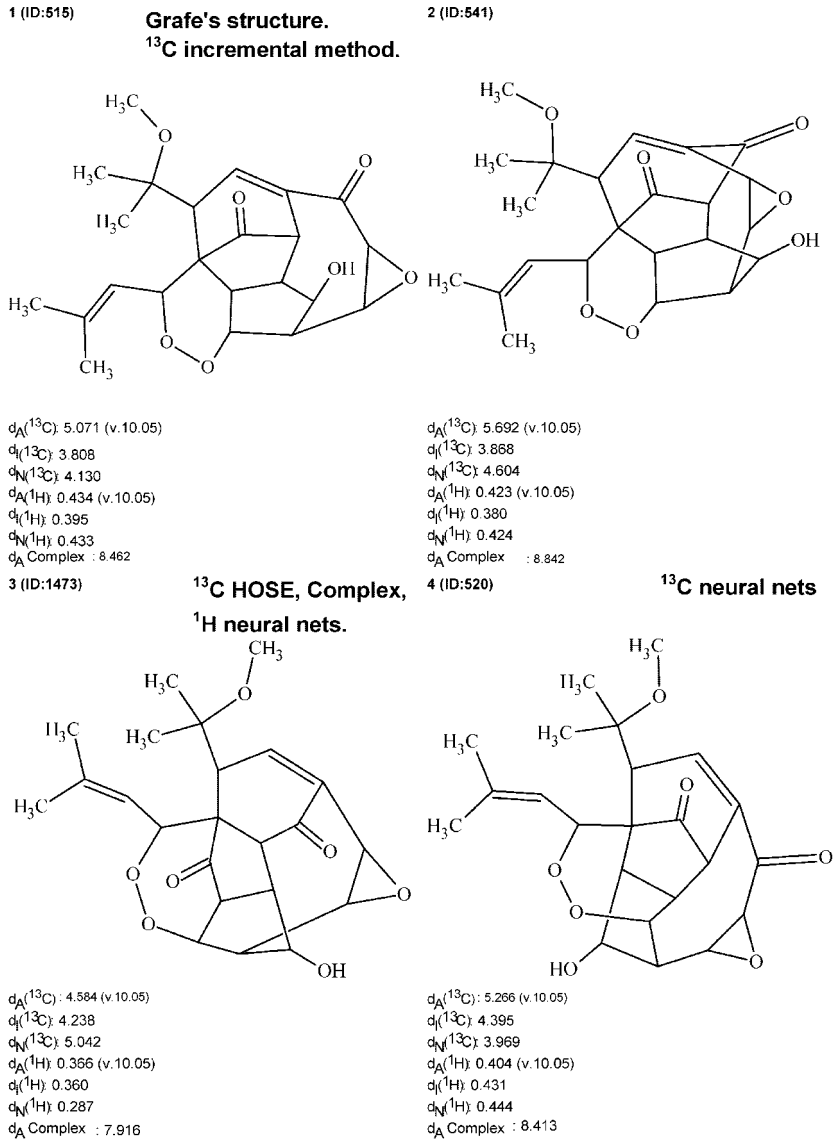
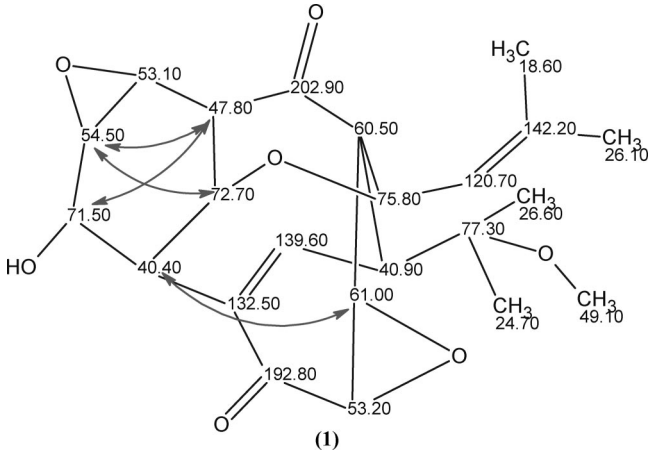


Figure 2. The four structures at the beginning of the CASE structure elucidation output file are ranked by the deviation of incrementally calculated spectra d_I(C). The inscriptions accompanying the structures list the type of spectrum prediction method for which a given structure was ranked in first position.



(1)

The table shows that, on one hand, the deviations found for structure **1** are markedly smaller than Gräfe's structure, but on the other hand four NSCs including one corresponding to a $^6J_{HH}$ coupling appear in the structure. It is to be expected that further refining of the possible structure by increasing the *m* value during fuzzy structure generation would result in further violation of the initial axioms declaring that the lengths of the COSY and HMBC correlations should be standard. The results obtained suggest that a revision of the initial data displayed in Table 1 may be necessary.

As mentioned earlier, Rychnovsky[15] suggested the hexacyclinol structure differing from that suggested by Gräfe's and offered chemical shift assignments for this newly suggested structure using the spectroscopic data published by Gräfe et al.[13] Deviations based on NMR predictions were calculated for both structures suggested by Rychnovsky and Gräfe. The results are listed in Table 3.

All deviations calculated for the structure suggested by Rychnovsky are slightly smaller than those for the structure suggested by Gräfe. However, checking Rychnovsky's structure relative to the 2D NMR data shows the presence of 12 NSCs with unusually long topological distances between the intervening nuclei. Using a designation introduced previously,[10] the set of NSCs can be symbolized as follows:

$$m = \text{COSY}\{1a(5) + 2a(4) + 2a(3) + 2a(2)\} + \text{HMBC}\{2a(5) + 2a(4) + 1a(3)\} = 12$$

where $na(p)$ denotes that there are n connectivities that exceed by p bonds the standard length assumed for a given 2D NMR technique (see Table 4). Rychnovsky's structure with the corresponding NSCs

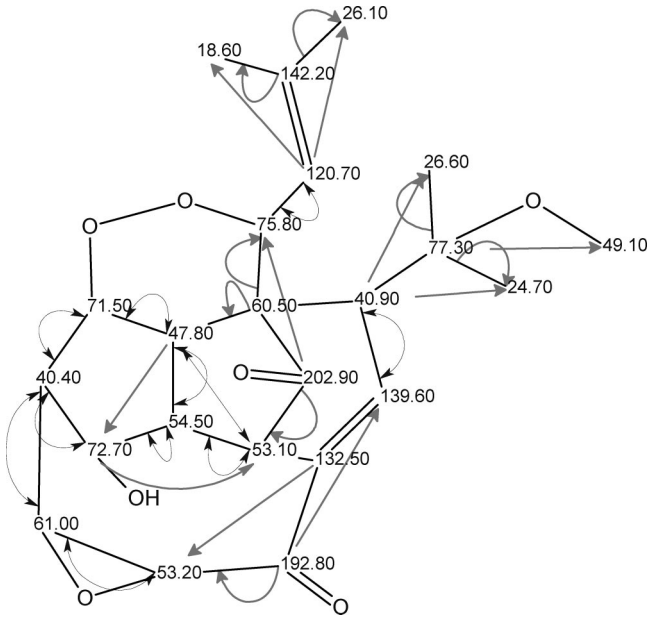


Figure 3. Structure of hexacyclinol proposed by Gräfe and the ^{13}C NMR chemical shift assignment suggested in his work.[13] The COSY (two-headed arrows) and HMBC (unidirectional arrows) connectivities are shown.

Table 2. Comparison of Deviations Calculated for Structure **1** and for Gräfe's Structure (Figure 3)

structure **1**	Gräfe's structure[13]
$d_A(^{13}C)$ 3.808	$d_A(^{13}C)$ 5.071
$d_I(^{13}C)$ 2.873	$d_I(^{13}C)$ 3.808
$d_N(^{13}C)$ 2.564	$d_N(^{13}C)$ 4.130
$d_A(^{1}H)$ 0.412	$d_A(^{1}H)$ 0.434
$d_I(^{1}H)$ 0.420	$d_I(^{1}H)$ 0.395
$d_N(^{1}H)$ 0.356	$d_N(^{1}H)$ 0.433
$d_A(\Sigma)$ 6.126	$d_A(\Sigma)$ 8.462

Table 3. Comparison of Deviations for Structures Suggested by Gräfe and Rychnovsky Using the Data Listed in Table 1

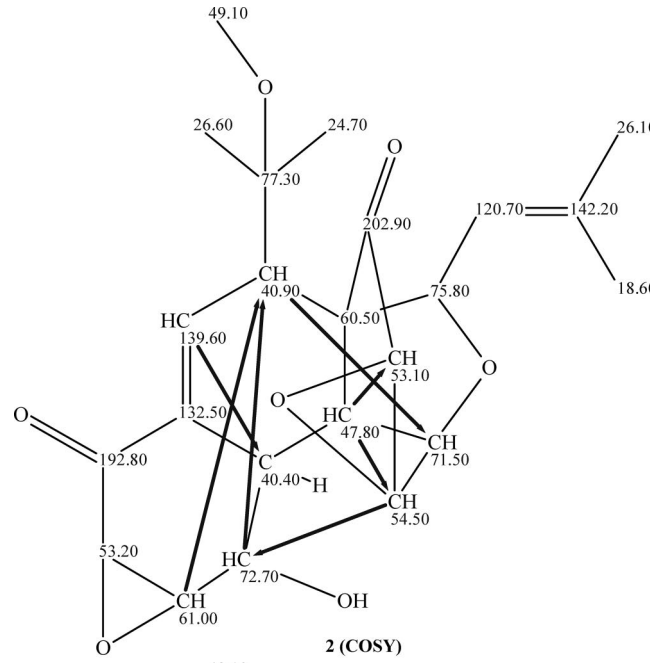
Gräfe's structure[13]	Rychnovsky's structure[15]
$d_A(^{13}C)$ 5.071	$d_A(^{13}C)$ 5.027
$d_I(^{13}C)$ 3.808	$d_I(^{13}C)$ 3.699
$d_N(^{13}C)$ 4.130	$d_N(^{13}C)$ 4.018
$d_A(^{1}H)$ 0.434	$d_A(^{1}H)$0.343
$d_I(^{1}H)$ 0.395	$d_I(^{1}H)$ 0.297
$d_N(^{1}H)$ 0.433	$d_N(^{1}H)$ 0.289
$d_A(\Sigma)$ 8.462	$d_A(\Sigma)$ 8.457

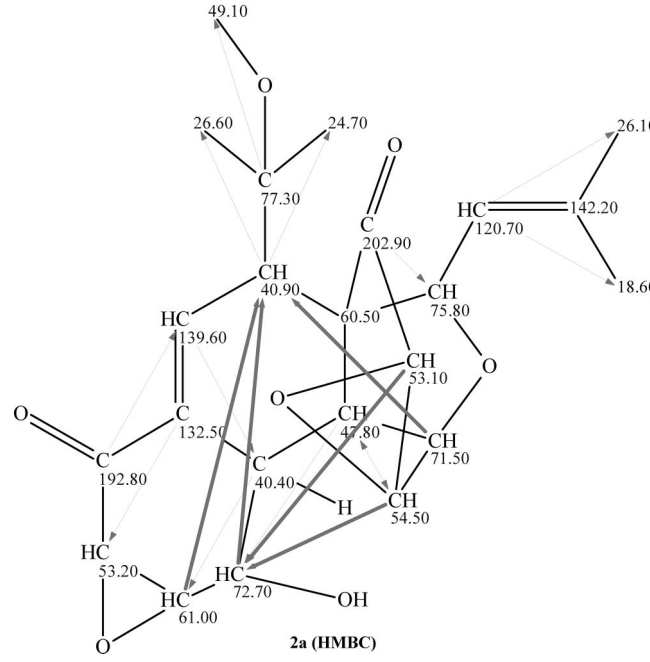
Table 4. Nonstandard Connectivities Present in the Structure Suggested by Rychnovsky (marked in bold)

COSY	HMBC
40.9→61.00 $^{7}J_{HH}$	40.9→61.0 $^{6}J_{CH}$
40.9→72.7 $^{6}J_{HH}$	40.9→72.7 $^{5}J_{CH}$
40.9→71.5 $^{5}J_{HH}$	40.9→71.5 $^{4}J_{CH}$
40.4→139.6 $^{4}J_{HH}$	54.5→72.7 $^{5}J_{CH}$
72.7→54.5 $^{6}J_{HH}$	53.1→72.7 $^{6}J_{CH}$
54.5→47.8 $^{4}J_{HH}$	
53.1→47.8 $^{5}J_{HH}$	

is shown below with the nonstandard connectivities marked by bold lines (structure **2** (COSY) and **2a** (HMBC)).

Analysis of the spectroscopic data led Reynolds[18] to conclude that the structure and NMR data could be made more consistent if the ^{1}H NMR chemical shifts at 3.55 and 3.59 ppm were exchanged. This exchange only slightly reduced the deviations, while the number of NSCs and their lengths still remained fairly large (a = 2–4). Up to this point in the present study attempts were made to





resolve the problem assuming that only the 2D NMR data of Gräfe[13] were available and that the structure suggested by Rychnovsky[15] has chemical shift assignments based on these data.

The investigation led to the following conclusions: all possibilities to apply the 2D NMR data published by Gräfe for clarifying the conflict between the two suggested structures were exhausted, and it was impossible to either prove or refute either of them. It became clear that either the structure suggested by Rychnovsky was incorrect or that the chemical shift assignments suggested in his work should be reexamined using new 2D NMR data. Rychnovsky's structure for hexacyclinol has been unambiguously confirmed by X-ray analysis,[16] so the challenge remains in aligning the chemical shift assignments of the structure in agreement with all of the NMR data.

NMR Confirmation of the Structure of Hexacyclinol

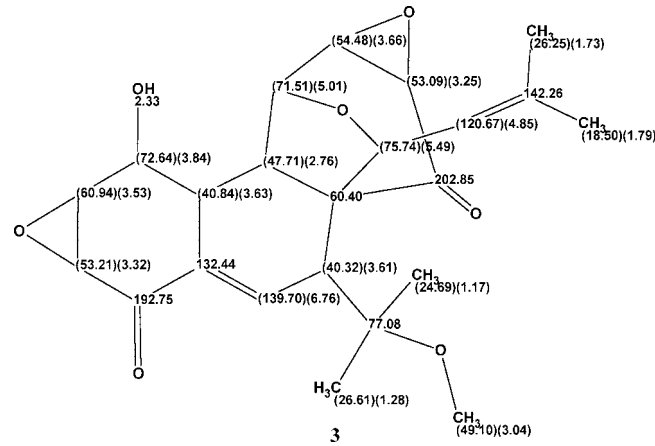
In order to determine appropriate chemical shift assignments for hexacyclinol, 1D and 2D NMR spectra of the compound synthesized by Porco et al.[16] were acquired under conditions described in the

Experimental Section. Positions and numbers of resonances localized in the strongly overlapped areas of ^{1}H NMR were refined using the ^{1}H NMR spectrum registered at the frequency of 900 MHz. Using ACD/SpecManager[19] we automatically determined peak positions, multiplicities, and coupling constants (when it was possible) in the ^{1}H NMR spectrum. With these data, assignments of series of peaks observed in 2D NMR spectra were revised. The problem was reformulated in the following manner: once the structure is determined unambiguously, determine those chemical shift assignments that remove the extremely long-range correlations. Generally speaking, the criterion based on the minimization of the number of NSCs and their length is of a heuristic nature, so this criterion should be added to the system of axioms used for confirmation of the structure of hexacyclinol.

To assist the process of understanding, we will denote *new* experimental chemical shifts in the ^{1}H and ^{13}C NMR spectra using an *italic* font. The consideration of nonstandard connectivities presented in structure **2**(**2a**) shows that exchange of pairs of carbon atom chemical shifts, specifically exchanging the pair 40.90(*40.84)* and 40.40(*40.32*) ppm, immediately removes the nonstandard lengths of three of the unusually long HMBC connectivities: 40.90–71.50 (*40.84–71.51*), 40.90–72.70 (*40.84–72.64*), and 40.90–61.00 (*40.84–60.94*) ppm. This exchange of assignments also adjusts the proton–proton connectivities in the COSY spectrum between the protons attached to the following pairs of carbons: 40.90–71.50 (*40.84–71.51*), 40.90–72.70 (*40.84–72.64*), 40.90–61.00 (*40.84–60.94*), and 40.4–139.6 (*40.32–139.70*) ppm to acceptable lengths ($^{3-4}J_{HH}$). To edit the COSY ($^6J_{HH}$) and HMBC ($^5J_{CH}$) connectivities 72.70–54.50 (*72.64–54.48*), we took into account that four ^{1}H NMR resonances, *3.53, 3.61, 3.63*, and *3.66* ppm, are observed in a very narrow spectroscopic interval (see Supporting Information, Figure B, where an expansion of the ^{1}H NMR spectrum obtained at 900 MHz is displayed) related to several equivalent possibilities for HSQC peak assignment may be expected. The ^{1}H NMR chemical shifts determined from this spectrum are very close to those found from the HSQC spectrum. Assuming that *72.64*(*3.84*) and *40.84*(*3.63*) ppm are newly assigned HSQC assignments, then both the noted COSY and HMBC connectivities will assume standard lengths. The $^4J_{HH}$ COSY connectivity 54.50–47.80 (*54.48–47.71*) ppm is adjusted to a standard length when the COSY correlation *3.66–2.76* ppm is replaced by the correlation *3.63–2.76* ppm. The COSY peak 3.23–2.74 (53.10–47.8) ppm shown in Table 1 might be assigned to a cross-peak *3.20–2.76* ppm in the experimental spectrum, but careful analysis of the spectrum shows that this peak is absent from the 2D NMR data and should be considered as an error. According to Table 1 the HMBC cross-peak at 54.50–72.70 (*54.48–72.64*) ppm is related to two peaks: 3.64–72.70 (*3.66–72.64*) ppm and 3.80–54.50 (*3.84–54.48*) ppm. Exchanging the assignments between *3.66* ppm and the nearby shift of *3.63* ppm leads to a $^2J_{CH}$ correlation length. The HMBC peak *3.84–54.48* ppm is not observed in the experimental spectrum. The remaining long ($^6J_{CH}$) HMBC connectivity 53.10–72.70 (*53.09–72.64*) ppm corresponds to the cross-peak at 3.80–53.1 (*3.84–53.09*) ppm and is transformed into a standard correlation when the ^{13}C NMR chemical shift *53.09* ppm is exchanged with the peak at *53.21* ppm.

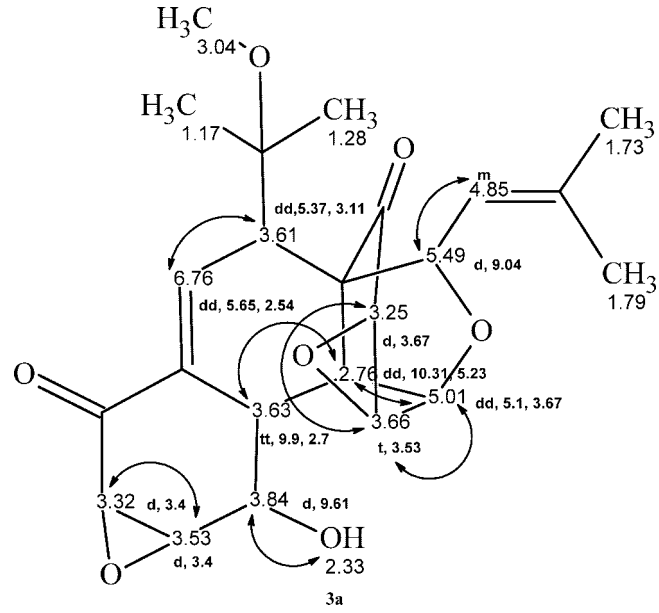
The final chemical shift assignment of the Rychnovsky[15] structure is given in Table 5 (see ref 21), where NMR data obtained in this work are shown in comparison with Gräfe's data.[13] The ^{13}C and ^{1}H NMR chemical shifts determined in this work differ only very slightly from those determined in the original work.[13] The near coincidence of the chemical shifts provides us with the basis to conclude that the compound identified by Gräfe et al.[13] had a structure consistent with that suggested by Rychnovsky.[15] This conclusion is further confirmed by the coincidence of multiplicities and coupling constants in the ^{1}H NMR spectra obtained by Gräfe et al.[13] and in this report.

The ^{13}C and ^{1}H NMR chemical shift assignments are displayed in structure **3** below:



3

The COSY correlations accompanied by the peak multiplicities and the measured coupling constants are shown in structure **3a**:



3a

When the new NMR data presented in Table 5 were entered into the StrucEluc program, only three structures were produced in 0.095 s. After selection of the best structure, the program placed structure **3** at the first position and with the same ^{1}H and ^{13}C NMR assignments shown. Table 6 lists the deviations calculated for hexacyclinol with both the old (see Table 3) and the new assignments.

It is evident that the deviations calculated by all methods are smaller for the structure for which chemical shifts were assigned in this work. Note that all 2D NMR correlations in the structure **3** are now of standard length.

Experimental Section

The 900 MHz NMR data were obtained on a Bruker AVANCE-900 NMR spectrometer, which was equipped with a Bruker 5 mm TCI cryoprobe. A synthetic sample of (+)-hexacyclinol (~1.2 mg) was dissolved in ~120 uL of $CDCl_3$, contained in a 3 mm o.d. NMR sample tube (~3 cm sample height) and centered in the receiver coil of the cryoprobe. ^{1}H NMR spectra were acquired with 64K data points and zero-filled to 256K data points prior to Fourier transformation. All 2D experiments (gCOSY, gHSQC, gHMBC) employed gradient-enhanced pulse sequence versions and were acquired with proton detection in f2. The gCOSY data was acquired

Table 5. Comparison of 1D and 2D NMR Data Presented in Ref 13 with the Data Obtained in the Current Work (with those chemical shifts marked in italics)

#	δC	δH	COSY	HMBC
1	18.60 q	1.77 s		142.20, 120.70
	18.50	*1.79*	*4.85*	*142.26; 120.67; 26.25*
2	142.20 s			
	142.26			
3	26.10 q	1,72 s		142.20, 120.70
	26.25	*1.73*	*4.85*	*142.26; 120.67; 18.50*
4	120.70 d	4.82 d, 10.1	5.46	
	120.67	*4.85 m*	*5.49; 1.7, 1.79*	*75.74; 26.25; 18.50; 60.40*
5	75.80 d	5.46 d, 10.1	4.82	60.50, 202.90
	75.74	*5.49 d, 9.04*	*4.85*	*60.40; 202.85; 142.26; 40.32; 120.67; 71.51*
6	60.50 s			
	60.40			
7	202.90 s			
	202.85			
8	53.10 d	3.23 d, br, 3.5	3.64, 2.74	202.90, 54.50
	53.09	*3.25 d,3.67*	*3.66*	*202.85; 60.40*
9	54.50 d	3.64 m	3.23, 2.74, 3.80	47.8
	54.48	*3.66 t, 3.53*	*3.25, 5.01*	
10	47.80 d	2.74, dd, 5.2, 7.8	3.64, 4.99	54.50, 60.50
	47.71	*2.76 dd, 10.31, 5.23*	*5.01;* **3.63**	*54.48; 60.40; 40.32; 202.85; 71.51; 75.74*
11	71.50 d	4.99 dd,5.2 br	2.74, 3.55	
	71.51	*5.01 dd 5.10, 3.67*	*2.76, 3.66*	*53.09; 75.74; 60.40*
12	40.40 d	3.55 m	4.99, 3.80	61.00, 71.50, 72.70
	40.32	*3.61 dd, 5.37, 3,11*	*6.76*	*60.40; 77.08; 139.70; 47.71; 202.85; 132.44; 24.69; 26.61*
13	72.70 d	3.80 dd, 9.5, 1.5, 2.54 br (OH)	3.55, 3.64	40.40, 54.50, 47.80, 53.10
	72.64	*3.84 d, 9.61*	*2.33*	*47.71; 53.21; 60.94*
14	61.00 d	3.51 dd, 2.9, 0.5	3.55, 3.29	40.40
	60.94	*3.53 d 3.4*	*3.32*	*40.84; 53.21; 72.65*
15	53.20 d	3.29 d, 3.2	3.51	132.50, 192.80
	53.21	*3.32 d, 3.4*	*3.53*	*132.44; 192.75; 60.94*
16	192.80 s			
	192.75			
17	132.50 s			
	132.44			
18	139.60 d	6.73 dd 5.3, 2.4 (allyl)	3.59	192.80, 40.90
	139.70	*6.76 dd, 5.65, 2.54*	*3.61*	*192.75; 40.32; 60.40*
19	40.90 d	3.59 d, 5.3	6.73	139.60
	40.84	*3.63 tt, 9.9, 2.7*	*2.76*	*139.70; 72.64; 132.44*
20	77.30 s			
	77.08			
21	26.60 q	1.26 s		77.30, 40.90
	26.61	*1.28*		*77.08; 40.32; 24.69*
22	24.70 q	1.15 s		77.30, 40.90
	24.69	*1.17*		*77.08; 40.32; 26.61*
23	49.10 q	3.02 s		77.30
	49.10	*3.04*		*77.08*
	OH		*3.84*	*72.64; 40.84*

Table 6. Comparison of the Deviations Calculated for the Rychnovsky Structure with both Old and New Chemical Shift Assignments (all entries are in ppm)

old assignment	new assignment
$d_A(^{13}C)$ 5.03	$d_A(^{13}C)$ 4.46
$d_I(^{13}C)$ 3.70	$d_I(^{13}C)$ 3.07
$d_N(^{13}C)$ 4.02	$d_N(^{13}C)$ 2.60
$d_A(^{1}H)$ 0.34	$d_A(^{1}H)$ 0.33
$d_I(^{1}H)$ 0.30	$d_I(^{1}H)$ 0.29
$d_N(^{1}H)$ 0.29	$d_N(^{1}H)$ 0.28
$d_A(\Sigma)$ 8.46	$d_A(\Sigma)$ 5.40

with 2K points in f2 with 256 increments in f1. The f1 increments were linear predicted to 2K, and both f1 and f2 were zero-filled to 4K points to yield a 4K × 4K data set after Fourier transformation. The gHSQC and gHMBC experiments were acquired with 2K points in f2 and 128 increments in f1 (gHSQC) and 256 increments in f1 (gHMBC) and subsequently linear predicted to 1K increments and 2K increments in f1, zero-filled to 4K × 2K (gHSQC) and 4k × 4k (gHMBC), respectively, prior to Fourier transformation. The frequency range of the carbon-13 sweep widths for the gHSQC and gHMBC experiments was 180 and 220 MHz, respectively.

Conclusions

It is rather an uncommon situation when different research groups publish different chemical structures for newly separated or synthesized organic molecules, especially when 2D NMR data are used for the structure elucidation. This is more likely to happen if there is severe overlap in the NMR spectra since this eventuality leads to ambiguous interpretation of the data. With severe overlap in the proton spectrum, incorrect chemical shift assignments can carry through the homonuclear and heteronuclear HSQC (HMQC) spectra. If two researchers employ slightly different initial data, then it is possible that they can arrive at different structures. In this article, we have demonstrated that resolving contradiction(s) between two or more proposed structures the capabilities within the expert system StrucEluc[7–10] can be extremely advantageous since the system is capable of both generating and validating different structural solutions derived from the data.

The system was applied to the selection of the most appropriate structure from two suggested variants of the natural product hexacyclinol. This compound was first separated and structurally characterized by Gräfe et al.[13] Subsequently the structure of hexacyclinol was revised[15,16] and an alternative structure was confirmed via total synthesis. Combining the procedures of both

structure generation and spectrum prediction incorporated into the expert system we checked both structural hypotheses and independently confirmed the revised structure.

In order to help in a fresh analysis of the NMR data, we reacquired 1D and 2D NMR spectra at 900 MHz for a synthetic sample as described elsewhere.[16] With these data we have been able to reassign both the ^{1}H and ^{13}C chemical shifts. The reassignment was governed by the heuristic requirement of eliminating unusually long COSY and HMBC correlations. As a result, a number of the chemical shift pairs with very similar chemical shifts were permuted and all 2D NMR connectivities were converted to standard lengths. Simultaneously, the chemical shift deviations calculated by all available prediction methods were smaller than those suggested by Rychnovsky, thereby endorsing the new assignment. The number of exchanged chemical shifts is actually quite large, and the criterion of minimizing the number of nonstandard correlations is heuristic in nature and leads to the question, is the suggested assignment correct? Turning to the works of Sigmund Freud[20] we quote: "An attribute of scientific thinking is the possibility to be content with an approximation to the truth and to continue creative work in spite of the absence of final confirmation". We have no further data available[21] to us to assist in checking the suggested assignment, so we accept the conclusions as being accurate and appropriate.

Acknowledgment. The purchase of the 900 MHz NMR spectrometer and construction of the UIC Center for Structural Biology were funded by NIH Grant GM068944.

Supporting Information Available: A graphic indicating the nature of a molecular connectivity diagram generated by the Structure Elucidator software program during the computer-assisted structure elucidation process. An expansion of the 900 MHz ^{1}H NMR spectrum of hexacyclinol is shown as evidence of the nature of peak overlap in the spectrum as discussed in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Lederberg, J.; Sutherland, G. L.; Buchanan, B. G.; Feigenbaum, E. A.; Robertson, A. V.; Duffield, A. M.; Djerassi, C. *J. Am. Chem. Soc.* **1968**, *91*, 2973–2976.

(2) Elyashberg, M. E.; Gribov, L. A. *Zh. Prikl. Spectrosc.* **1968**, *8*, 296–300.

(3) Sasaki, S. I.; Abe, H.; Ouki, T.; Sakamoto, M.; Ochiai, S. *Anal. Chem.* **1968**, *40*, 2220–2223.

(4) Nelson, D. B.; Munk, M. E.; Gasli, K. B.; Horald, D.L. *J. Org. Chem.* **1969**, *34*, 3800–3805.

(5) Elyashberg, M. E.; Gribov, L. A.; Serov, V. V. *Molecular Spectral Analysis and Computer* (in Russian); Nauka: Moscow, 1980.

(6) Elyashberg, M. E.; Martirosian, E. R.; Karasev; Yu., Z.; Thiele, H.; Somberg, H. *Anal. Chim. Acta* **1997**, *348*, 443–463.

(7) Elyashberg, M. E.; Blinov, K. A.; Molodtsov, S. G; Williams, A. J.; Martin, G. E. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 771–792.

(8) Molodtsov, S. G.; Elyashberg, M. E.; Blinov, K. A.; Williams, A. J.; Martin, G. E.; Lefebvre, B. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1737–1751.

(9) Elyashberg, M. E.; Blinov, K. A.; Molodtsov, S.; G; Williams, A. J.; Martin, G. E. *J. Chem. Inf. Model.* **2006**, *46*, 1643–1656.

(10) Elyashberg, M. E.; Blinov, K. A.; Molodtsov, S.; G; Williams, A. J.; Martin, G. E. *J. Chem. Inf. Model.* **2007**, *47*, 1053–1066.

(11) Blinov, K. A.; Carlson, D.; Elyashberg, M. E.; Martin, G. E.; Martirosian, E. R.; Molodtsov, S. G.; Williams, A. J. *Magn. Reson. Chem.* **2003**, *41*, 359–372.

(12) Blinov, K. A.; Elyashberg, M. E.; Martirosian, E. R.; Molodtsov, S. G.; Williams, A. J.; Sharaf, M. M. H.; Schiff, P. L. Jr.; Crouch, R. C.; Martin, G. E.; Hadden, C. E.; Guido, J. E.; Mills, K. A. *Magn. Reson. Chem.* **2003**, *41*, 577–584.

(13) Schlegel, B.; Härtl, A.; Dahse, H.-M.; Gollmick, F. A.; Gräfe, U.; Dörfelt, H.; Kappes, B. *J. Antibiot.* **2002**, *55*, 814–817.

(14) La Clair, J. J. *Angew. Chem., Int. Ed.* **2006**, *45*, 2769–2773.

(15) Rychnovsky, S. D. *Org. Lett.* **2006**, *8*, 2895–2898.

(16) Porco, J. A., Jr; Su, S.; Lei, X.; Bardhan, S.; Rychnovsky, S. D. *Angew. Chem., Int. Ed.* **2006**, *45*, 1–4.

(17) *ACD/CNMR Predictor* v.10.0; Advanced Chemistry Development Inc.: 110 Yonge St., Toronto, Ontario, Canada M5C 1T4.

(18) Reynolds, W. F. The Hexacyclinol Controversy; What Went Wrong With The Original Structure Determination?, Poster #34, SMASH NMR Meeting., Burlington, , 2006.

(19) *ACD/SpecManager* v.10.0; Advanced Chemistry Development Inc.: 110 Yonge St., Toronto, Ontario, Canada M5C 1T4.

(20) Freud, S. *New Introductory Lectures on Psycho-analysis*; London, 1933.

(21) The current assignments given in Table 5 are consistent with the reassignments reported in the hexacyclinol total synthesis report. See the Supporting Information associated with ref 16 for further details.

NP070557T