Pure & Appl.Chem., Vol.54, No.12, pp.2425-2442, 1982.
Printed in Great Britain.
0033-4545/82/122425-18$03.00/0
Pergamon Press Ltd.

THE DENDRAL PROJECT: COMPUTATIONAL AIDS TO NATURAL PRODUCTS STRUCTURE ELUCIDATION*

Carl Djerassi, Dennis H. Smith,** Christopher W. Crandell,
Neil A.B. Gray,*** James G. Nourse and Matthew R. Lindley

Department of Chemistry, Stanford University, Stanford, CA, USA 94305

Abstract - We discuss the use of several computer programs designed to aid chemists in structure elucidation. These programs are highly interactive in that a chemist works together with a program to derive structural candidates for an unknown, to evaluate the candidates using spectroscopic data, and to explore relationships between structure and biological activity. An example is carried through the discussion to illustrate our methods.

INTRODUCTION

For more than a decade, the DENDRAL Project has been conducting research on computational techniques applied to structural analysis of organic compounds. From the very beginning, we have explored ways of developing computer programs that are specifically designed to aid structural chemists in their investigations of molecular structure problems. We seek in these efforts to augment the problem-solving capabilities of chemists, not to replace chemists. Our feeling has always been that a powerful synergism can be created between the chemist and computer, given properly written computer programs designed to interact with a chemist during investigation of a new problem.

To achieve these goals, we have developed programs that handle those parts of structural analysis that are most difficult to perform manually: exhaustive generation of structural candidates in a structure elucidation problem, prediction of spectral properties, and perception and manipulation of molecular symmetry and stereochemistry. We have left to the chemist the computationally more difficult tasks of data interpretation, application of ancillary chemical knowledge unknown to the computer, and exercise of judgement and chemical intuition. Using carefully built interfaces between the chemist and the computer, we have provided mechanisms that allow the chemist to express such insights into a problem.

Our efforts have led to a number of interactive computer programs that have been applied to many different problems of structural analysis, primarily in the area of structure elucidation. A recent review summarizes many of these programs, using several different examples to illustrate how the programs work.[2] In this paper we take a somewhat different approach to demonstrate the utility of our programs. We have chosen a single example to provide a unifying theme to our presentation, and carry this example from structure elucidation through exploration of relationships between structure and biological activity.

The example we have chosen is the compound warburganal, a potent insect antifeedant originally isolated by Kubo and co-workers from the East African plant Warburgia ugandensis.[3] The identity of warburganal was established by conventional analysis of its chemical and spectral properties. Several syntheses have subsequently been devised[4] and the biological activities of warburganal and related compounds have been investigated.[5] In the following sections, we take a retrospective look at this structural problem and show how our programs can be used for both structure elucidation and for establishing structure/activity relationships. Obviously, no single example can bring out all of the different ways our programs can be used, but this example serves as an adequate introduction to our computational approaches, and should stimulate thoughts about similar applications to related problems.

---

*Part XLII of the series "Applications of Artificial Intelligence for Chemical Inference". For part XLI see ref. 1.

**Current address: Building 105, Lederle Laboratories, Pearl River, New York 10965
***Current address: Department of Computer Science, University of Wollongong, NSW 2500, Australia

COMPUTER-ASSISTED STRUCTURE ELUCIDATION OF WARBURGANAL

There are several computer programs in existence that can aid chemists in determining the structure of an unknown compound, using data obtained from chemical and spectroscopic studies. For this example we focus primarily on one of those programs, GENOA[6] an outgrowth of our earlier CONGEN[7] program. Typically, the process of structure elucidation by the chemist involves three distinct phases:

1. data interpretation. The presence and absence of particular substructural fragments are inferred through analysis of spectral and chemical data that characterize the unknown;

2. structure generation. The substructures thus derived are pieced together to yield one or more complete structures that represent possible candidates for the unknown;

3. structure evaluation. The candidate structures are evaluated against existing data to eliminate some possibilities and new experiments are planned to discriminate among remaining candidates.

Much of phases 1-3, particularly the spectral interpretation in phase 1 and structure evaluation in phase 3, is intuitive. A chemist exploits analogies to other, previously identified compounds, intelligent guesswork, and general chemical experience. In contrast, the process of assembling the fragments into valid structures, phase 2, is essentially combinatorial. Computer programs can assist in all three phases of the structure elucidation process. The area in which programs currently excel is the process of combinatorial structure assembly, or structure generation, phase 2. This area is least intuitive and most difficult for the chemist because there are no simple techniques or procedures to follow to guarantee that all possible candidates have been written down. Structure generating programs such as GENOA,[6] and its predecessor CONGEN,[7] provide this guarantee. Such programs are more correctly called isomer generating programs because they generate, in the computer, structural isomers of a given molecular formula. Structure generators are often interfaced to other programs that perform automated spectral interpretation and structure evaluation. Such combined programs will be discussed in subsequent sections. For the purposes of this example, however, we will limit ourselves to illustrating computer programs to assist in phase 2, structure generation, and phase 3, structure evaluation.

STRUCTURE GENERATION: OBTAINING STRUCTURAL CANDIDATES

In this section we use GENOA as a prototypical structure generator, a tool with the aid of which a chemist can establish the scope of the structural problem and can explore the implications of various structural hypotheses. Programs like GENOA are typically interactive. A chemist, working with the program, defines first the molecular formula of an unknown compound and then, as they are identified, the substructural constraints that can be inferred from available data. We illustrate the processes of defining and using substructural constraints, and obtaining a complete set of structural candidates that obey all constraints, for the structure of warburganal. For your amusement, and to put you into the mode of thinking systematically about the structural inferences, as GENOA must do, you should get a pencil and paper to see if you can verify the results of the program at each step of the analysis.

Warburganal was one of four components isolated, using several chromatographic techniques, from the antifeedant fraction obtained by extraction of the ground bark of Warburgia ugandensis.[3] The molecular formula, determined by mass spectrometry, is $C_{15}H_{22}O_3$. Key infrared (i.r.) spectral data, and substructural constraints derived therefrom[3] are summarized in Table 1.

TABLE 1. I.R. spectral data and related substructural constraints for warburganal

Absorption (cm^{-1})	Substructural Constraint
3460	OH (intramolecular hydrogen bonded)
2850, 1722	CHO
1687, 1650	-C=C-CHO | |

The ultraviolet (u.v.) spectrum (λ_{max}224nm,ε6300 in methanol) and circular dichroism (c.d.) spectrum ($\Delta\varepsilon$ -2.06 and $\Delta\varepsilon$ -0.78 in methanol) were stated by the authors[3] as indicative of the presence of two aldehyde groups, one of which is α,β-unsaturated. These assignments, notably the c.d. interpretation, were based in part on examination of additional data. Indeed,

these interpretations are confirmed by the analysis of ^{1}H (Fig. 1) and ^{13}C (Fig. 2) n.m.r., and mass (Fig. 3) spectra.****

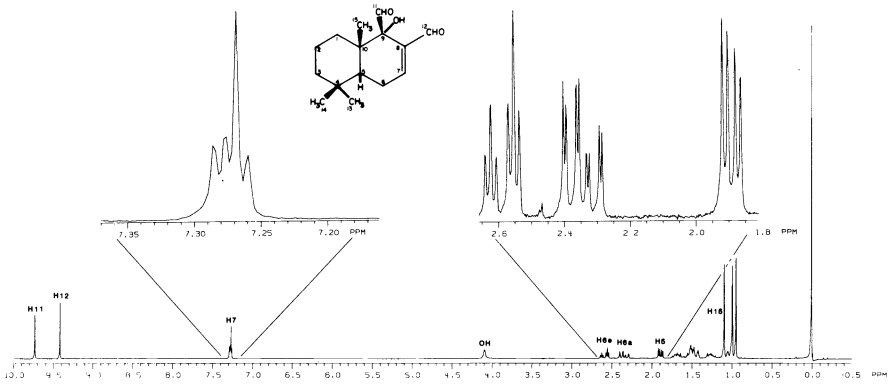


Fig. 1. 300 MHz ^{1}H n.m.r. spectrum of warburganal using TMS as internal standard.

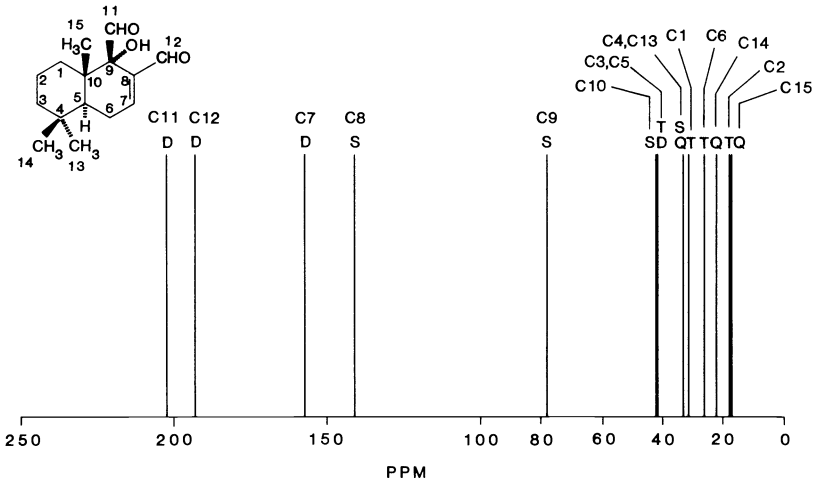


Fig. 2. ^{13}C n.m.r. spectrum of warburganal using TMS as internal standard. The letters S,D,T,Q refer to the multiplicity (singlet, doublet,...) of the resonance lines in the SFORD spectrum.

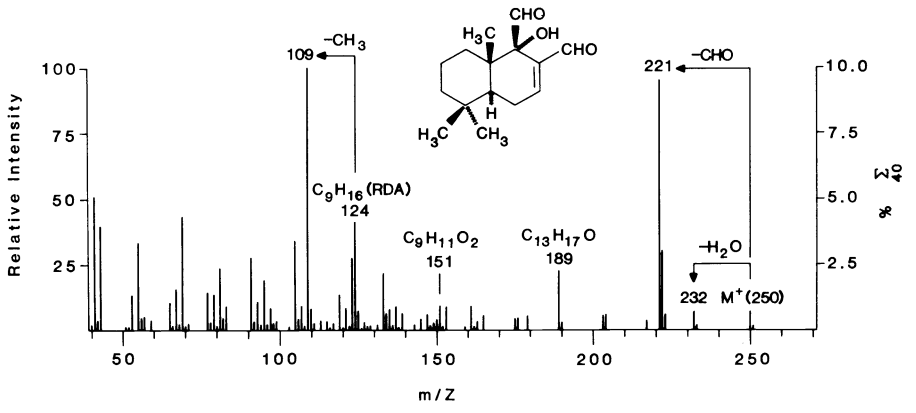


Fig. 3. Low resolution mass spectrum of warburganal.

****We express our thanks to Professor Koji Nakanishi for making available to us a sample of warburganal which was used to obtain these complete spectra which have so far not been reported in the literature.

In order to arrive efficiently at a set of candidate structures, the GENOA program must be supplied with all known information about a structure. In principle, any datum that can be expressed as a substructural constraint can be used within the program, as a substructure. The first and most important constraint is the molecular formula, and this is required before GENOA will begin constructing structures. Other constraints consist of substructural features together with some statement about the number of occurrences of the substructures in the unknown molecule. The program possesses a library of common substructural fragments that can be utilized directly without requiring their continued redefinition. However, most structural problems involve specialized substructures that could not reasonably be provided in a general purpose library, for example, a specific chain of methylene and methine carbons derived through a proton decoupling experiment. Therefore, the GENOA program provides a simple mechanism for substructure definition which allows the chemist to build up any desired substructure through an appropriate sequence of commands for creating rings and chains of atoms of specified types.

In the subsequent discussion we define and employ several substructural constraints using GENOA. The order in which these constraints are used is at the discretion of the chemist who interacts with the program. Some ways may be computationally more efficient than others, but generally the differences are of little consequence. After each new constraint is applied, the results are immediately available to the chemist for examination. In this way she or he can monitor the gradual growth of complete structures and, in the process, discover additional constraints which can be brought to bear to limit the problem further. Examples of this step by step examination of results, stimulation of new ideas, and use of new constraints are given below.

We illustrate the initial steps of interaction with GENOA in Fig. 4, which is an annotated transcript of actual use of the program for this example. Commands supplied by the chemist are underlined. At [1] in Fig. 4, the molecular formula of warburganal, $C_{15}H_{22}O_3$,[3] verified by our measurements of the mass spectrum (Fig. 3)***** at high resolving power, is defined.

```
[1]     #DEFINE MOLFORM
        MOLECULAR FORMULA:C 15 H 22 O 3
        MOLECULAR FORMULA DEFINED

[2]     #DEFINE SUBSTRUCTURE C-OH
        (NEW SUBSTRUCTURE)
  [2a]  >CHAIN 2
  [2b]  >ATNAME 2 O
  [2c]  >HRANGE
        ATOM:1
        MINIMUM NUMBER OF H'S:0
        MAXIMUM NUMBER OF H'S:0
  [2d]  >HRANGE 2 1 1
  [2e]  >HYBRIDIZATION
        ATOM: 1
        TYPE:SP3
  [2f]  >DONE
        C-OH DEFINED

[3]     #CONSTRAINT
        SUBSTRUCTURE NAME:C-OH
        RANGE OF OCCURRENCES:AT LEAST 1
        .
        1 CASE WAS OBTAINED
```

Fig. 4. Initial interaction with the GENOA program to define the molecular formula of warburganal and to define and use as a constraint a substructure, C-OH, representing a tertiary hydroxyl group.

The presence of some form of hydroxyl group was obvious from the i.r. spectrum (Table 1). Analysis of the ^1H n.m.r. (Fig. 1) and ^{13}C n.m.r. (Fig. 2) spectra suggested that the compound must be a tertiary alcohol. Therefore, a named substructure representing a carbon bearing a hydroxyl group and no hydrogens must be defined for GENOA. Although this simple substructure is in fact in a substructural library, we define it in this example to provide a brief illustration of how all substructures, of arbitrary complexity, can be defined. The name assigned to the substructure is only important as a "tag" which can be used to refer to the substructure at other points in the program, for example, in the CONSTRAINT command

***** The low resolution spectrum was derived by plotting a high resolution mass spectrum, obtained in our laboratory, at nominal mass resolution.

(Fig. 4). Subsequently in this example we will assume that all other substructures are named and defined for GENOA in like manner.

The substructure, named arbitrarily C-OH, is created by specifying first a chain of two atoms, ([2a],Fig. 4), which in the absence of further specifications are assumed by the program to be carbon atoms. Next we specify the name of one of the atoms, in this case atom 2, to be an oxygen ([2b], Fig. 4). Persons experienced with GENOA and related interactive programs soon learn how to abbreviate and combine commands. Persons who only occasionally use GENOA are assisted by the program, which prompts the person for missing information. This is illustrated at [2c] and [2d],(Fig. 4), where at [2c] we step through the HRANGE command to fix the number of hydrogens on the carbon as zero, while at [2d] we merely type all such information about the single hydrogen atom on the oxygen atom on one line. Additional information on the properties of the carbon atom is required because we know that this carbon is sp^3 hybridized. The HYBRIDIZATION command, at [2e], specifies this information, thereby avoiding the potential ambiguity of the requirement for a tertiary alcohol being accidentally satisfied by the presence of an enol system. Finally, the substructure definition is completed by issuing the DONE command at [2f], (Fig. 4).

The sequence of commands in step [2], (Fig. 4) merely defines to GENOA the substructural details of something called C-OH; we have not yet specified how many such substructures are required for this structural problem. The CONSTRAINT command is used to tell GENOA the number of occurrences of any defined substructure; a substructure may occur no times, i.e. it is forbidden, or it may occur a number or range of times, under the complete control of the chemist.[6] Each constraint causes GENOA to construct, in the computer, the given number of occurrences of the specified substructure. The starting point in each instance is what was obtained from the previous constraint. The construction procedure has been described previously.[6] For the first constraint, (Fig. 4) the starting point is the molecular formula and the constructive procedure is trivial in that there is only one unique way to construct the C-OH group from the atoms in the molecular formula.

The results of incorporating a constraint are cases, and the number obtained at each step is given by GENOA (Fig. 4). Cases are partial structures in which all previous constraints are incorporated. Complete structural isomers for a problem are never constructed until later in a problem when considerably more substructural constraints have been specified. There are excellent strategic reasons for this approach to structure generation, because it keeps the size of a problem, in terms of numbers of cases or structures, as small as possible for the given substructural constraints. Essentially the program does no more work than the minimum required to incorporate the next item of structural information. In examples given below we illustrate cases obtained for warburganal after additional constraints have been supplied.

Before inferring and using new constraints we summarize the results obtained and refer back to these results in subsequent discussion. In Table 2 we show the number of cases which result from incorporation of each named substructural constraint. The substructures themselves are given in Fig. 5.

TABLE 2. Number of cases obtained by GENOA on incorporation of each substructural constraint, and final number of structures (see Fig. 5 for the substructures corresponding to each name)

Substructure Name	Range of Occurrences	Evidence from	Resulting Number of Cases
C-OH	At least 1	IR, ^1H, ^{13}C	1
C-ALDEHYDE	At least 1	^1H	2
ENAL	At least 1	IR, ^1H, ^{13}C, CD, UV	2
C-SP2	Exactly 4	^{13}C	2
C-CH3	At least 3	^1H	17
T-BUTYL	None	^1H, ^{13}C	15
CH3	Exactly 3	^{13}C, ^1H	14
CH3-C-OH	None	Lanthanide shift study	6
DECOUPLE	At least 1	^1H decoupling	9
CH2	Exactly 4	^{13}C	9
CHAIN3	At least 1	^1H	9
CYCP	None	^1H	9

Final Number of Structures = 42

The spectral data characterizing warburganal imply the presence of two aldehyde groups, one forming part of an enal system (Table 1). These aldehyde groups are evidenced by the resonances at 9.72 (aldehyde) and 9.41 ppm (enal) in the ^1H n.m.r. spectrum (Fig. 1), and at

about 201 and 192 ppm in the ^{13}C n.m.r. spectrum (Fig. 2). The aldehyde can be assumed to be attached to a quaternary alkyl carbon because its only coupling in the proton spectrum is a long range 1 Hz coupling to the hydroxyl proton. The first aldehyde functionality is expressed by defining C-ALDEHYDE (Fig. 5). Use of this substructure as a constraint yields two new cases (Table 2) because there are two ways to incorporate this new fragment. The quaternary alkyl carbon attached to the aldehyde may or may not be the same as the carbon to which the hydroxyl group in substructure C-OH is bonded, resulting in partial structures A and B.

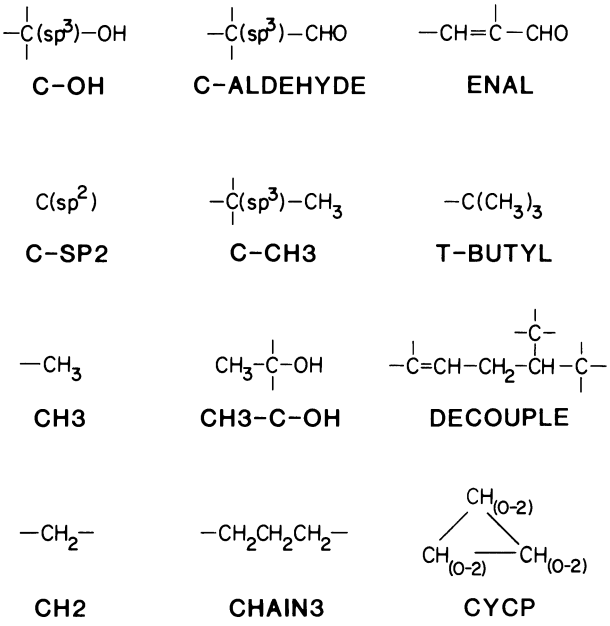
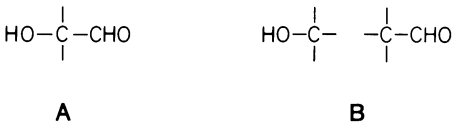


Fig. 5. Substructural constraints used in the structure elucidation of warburganal (see Table 2 for range of occurrence of each substructure).

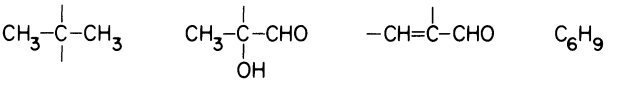


The next few substructures defined and used as constraints included (see Table 2 and Fig. 5):

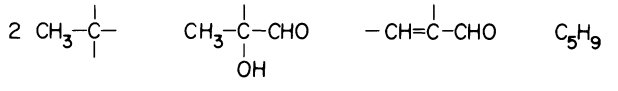
1. the required enal system, ENAL;
2. a limit of exactly four sp^2 hybridized carbons, C-SP2, as established from the number of resonances in the 100-240 ppm region of the ^{13}C n.m.r. spectrum (Fig. 2);
3. a requirement for at least three methyl groups bonded to quaternary alkyl carbons, C-CH3. These methyls correspond to the 3H-singlets at 1.10, 1.00 and 0.96 ppm in the ^{1}H n.m.r. spectrum (Fig. 1).

These constraints result in seventeen possible cases (the fifth entry in Table 2). Although each case is still comprised only of several small, disconnected substructural fragments, it is instructive to examine several of them to illustrate the point mentioned earlier that such examination often reveals that additional constraints can be implemented. Four of the seventeen cases are shown as C-F.

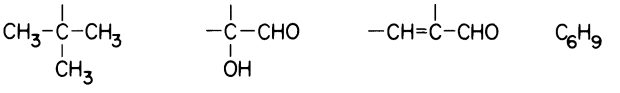
In this example, some of the partial structures proved to contain t-butyl groups, e.g. E. Such substructures do not seem compatible with the proton resonance data. The constraint T-BUTYL with range of occurrence "none" removes two cases containing a t-butyl group. Case F possess six methyl groups. In F, three new quaternary centers were created to satisfy the



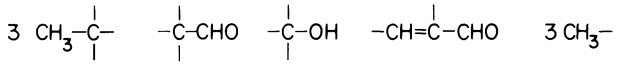
C



D



E



F

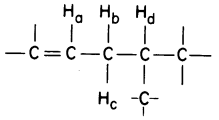
requirement for three C-CH3 substructures. GENOA has perceived that for F, given the constraints used so far (see Table 2), there are so many remaining hydrogens that three additional methyl groups would have to be formed. Constraining the problem so that exactly three methyls, CH3, are allowed eliminates F, the only case that possesses more than three-leaving fourteen cases to this point (see Table 2).

A partially completed problem can be saved at any stage in the analysis in order to wait for additional data. Meanwhile, the analysis can proceed on the basis of plausible but unproven, structural inferences. In this example, the methyl resonance shifts (1.10, 1.00, and 0.96 ppm) in the proton spectrum (Fig. 1) suggest that it is unlikely that any methyl be alpha to the hydroxy group. Methyl groups in environments like CH_3-C-OH typically have shifts of 1.3 ppm or greater. Also, addition of lanthanide shift reagent was not reported to result in shift of methyl resonances.[3] A constraint expressing these observations, CH3-C-OH, was defined and used with a range of occurrence "none", leaving only six cases (Table 2).

The structure of warburganal was proposed[3] based on the few observations mentioned above, together with comparison of spectral properties with analogs whose structures were determined previously.[3] For the purpose of our example, we proceed with further examination of the spectral data for warburganal to illustrate the more common circumstances where analogs with proven structures are not available. Decoupling experiments in the ^1H n.m.r. spectrum provide some of the most definitive substructural information about a molecule. Unlike constraints based upon correlations of substructure and chemical shift/absorption frequency, constraints based upon positive results from proton decouplings are unquestionable. Such experiments frequently allow substantial fragments of a molecule to be mapped out clearly. We have performed a series of decoupling experiments (summarized in Table 3) on warburganal in our laboratory, which resulted in the specification of the following additional substructure (named DECOUPLE, see Table 3) that could be used as a further constraint.

The vinylic proton H_a at 7.26 ppm is coupled (J-2.57, 5.01 Hz) to the two protons H_b and H_c comprising a non-equivalent methylene resonating at 2.58 and 2.33 ppm. These methylene protons are further coupled to each other and to a methine at 1.88 ppm, H_d; the methine exhibits no further coupling. This interpretation is confirmed when the decoupling information is taken into consideration. Irradiating at the frequency of the vinylic proton H_a (decoupling experiment 1 in Table 3) causes the methylene proton signals at 2.53 and 2.33 ppm to collapse to doublets of doublets and has no effect on the methine signal at 1.88 ppm, indicating that the vinylic proton is coupled to each of the protons corresponding to the first two signals, H_b and H_c, but not to the third. Decoupling at the frequency corresponding to the methylene protons (decoupling experiments 2 and 3), demonstrates that each of

TABLE 3. Chemical shifts, coupling constants, and decoupling data for warburganal. Shifts, multiplicity, and coupling constants for the undecoupled proton spectrum are shown in columns two and three. The last four columns present the results of four decoupling experiments, and the effects on the corresponding protons. In each column the asterisk indicates where the decoupler was set and the resulting effects on the multiplicities.



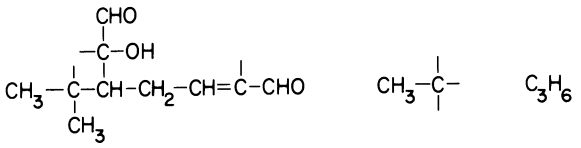
Decouple

Partial ^{1}H N.M.R. Spectrum

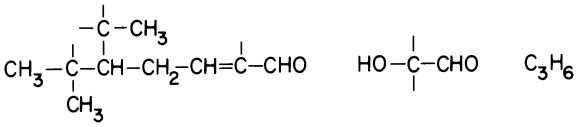
	Undecoupled		Decoupled			
Proton	Shift	mult.(J in Hz.)	1	2	3	4
H_a	7.26	dd(2.57,5.01)	*	d	d	dd
H_b	2.58	ddd(4.97,5.01,20.97)	dd	*	[a]	dd
H_c	2.33	ddd(2.57,11.80,20.97)	dd	[a]	*	dd
H_d	1.88	dd(4.97,11.80)	dd	d	d	*

[a]Results unclear due to incomplete decoupling and complexity of signals.

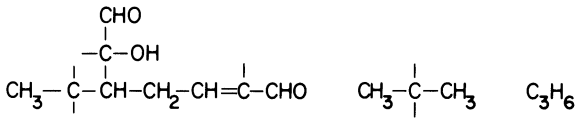
these two protons are coupled to each other and to the other two protons (at 7.28 and 1.88 ppm). Finally, irradiating the methine proton H_d resonating at 1.88 ppm affects only the signals at 2.58 and 2.33 ppm (decoupling experiment 4), indicating that it is coupled only to H_b and H_c. These data support the presence of substructure DECOUPLE. The assumption that the methine is in fact bonded to two quaternary carbons is reasonable (based on the absence of further coupling) but not conclusively proven. DECOUPLE was used as a constraint, bearing in mind that the data may have been over-interpreted, leading to nine new cases (Table 2). Four of the nine are shown as G-J. With the addition of the new constraint the structures are beginning to take definite shape.



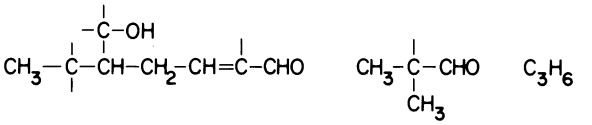
G



H



I



J

The ^{13}C spectrum indicates the presence of exactly four methylenes (Fig. 2). Requiring the presence of exactly four of substructure CH2 merely fixes precisely the degree of the three remaining carbons (C_3H_6 in G-J) to be two. The proton spectrum reveals no sharp methylene 2-H singlets indicating the absence of isolated methylene groups (Fig. 1). Given the structural inferences already made, the requirement for no isolated methylenes implies that three must form a chain, substructure CHAIN3. This constraint results in connecting the three methylenes together without changing the number of cases (Table 2).

The constraints applied so far represent all that one can obtain easily from the available spectral data. At this point it is reasonable to generate final structures from the existing cases, thereby obtaining the complete set of candidate structures. Prior to structure generation, it is useful to specify additional constraints to prevent the construction of undesirable structural features, thereby saving the time required to find and remove them from the set after all are built. In this example there is no evidence (absence of high-field resonances characteristic of cyclopropyl hydrogens in the ^1H n.m.r. spectrum) for a cyclopropyl group, so the constraint CYCP is defined and applied with range of occurrence "none". This constraint has no effect on the existing cases, none of which contain a cyclopropyl group, but it will discard automatically any structure which possesses such a functionality during structure generation. A total of 42 final structures are obtained, all of which (#1-#42) are shown in Fig. 6 to indicate the variety of structural types compatible with the constraints. The "#" signs in Fig. 6 and throughout the rest of the paper are included merely as a reminder that these are the computer-generated structural candidates and do not necessarily represent isolated, characterized compounds. Without the constraint of "no cyclopropyl groups", the total number of structural candidates would have been 57. The assigned structure of warburganal[3] is #16.

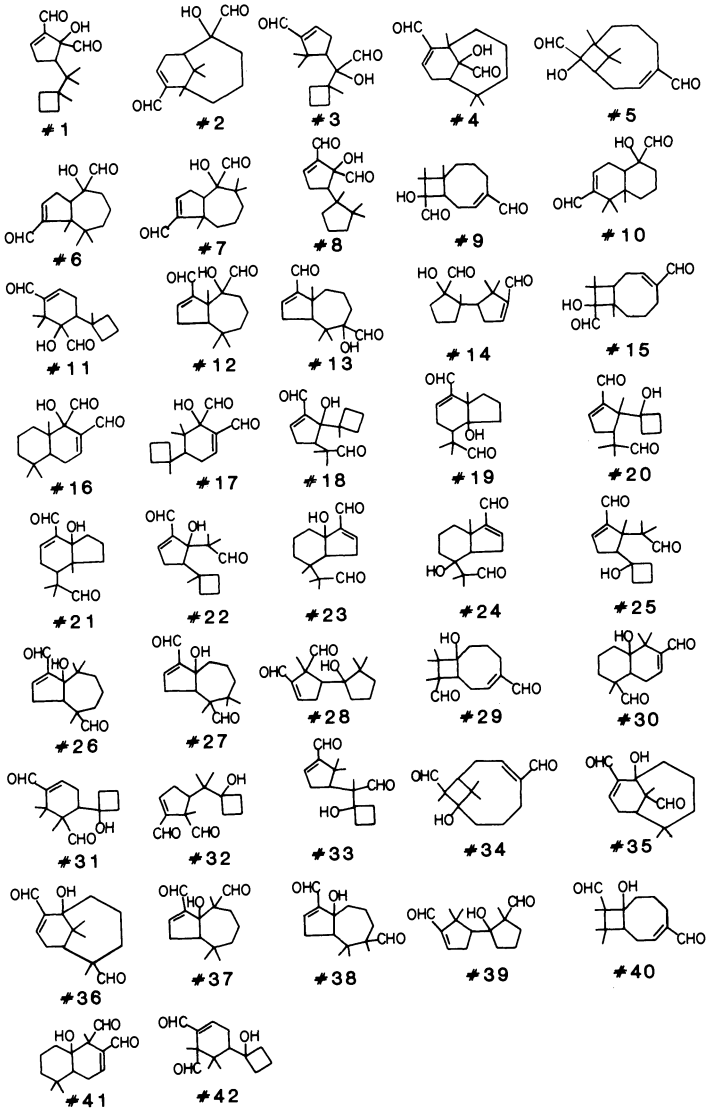


Fig. 6. The 42 structural candidates obtained for warburganal.

As with many other structural problems, the number of possible solutions consistent with numerous and detailed structural constraints is often surprising. In the previous instance it is obvious that no chemist, in the absence of a computer program such as GENOA, would have generated all of the possible structures. Even if such a remarkable feat were accomplished, the chemist would not have been sure that all conceivable structures had been generated.

A distinct advantage of carrying out structural studies on the computer is that no sample and little time is consumed by exploring alternative interpretations of data. Thus, we can examine the results, in terms of numbers of structural possibilities, using different assumptions. For example, if we assume that the ^{1}H n.m.r. data do not exclude CH_3-C-OH groups, then the only limit placed on methylene groups is to exclude -CH_2- groups between two quaternary carbons. Further, if no assumptions are made about the bonding of the alkyl methine then some two thousand structures are compatible with these unambiguous spectral inferences. However, if the origin of warburganal is assumed to imply that it incorporates a standard decalin system or perhydroazulene skeleton (as are found in many sesquiterpenes) then only eighty of the two thousand structures are compatible with these skeleta and unambiguous spectral interpretations.

STRUCTURE EVALUATION: FOCUSING ON THE CORRECT STRUCTURE

Once structure generation is complete, transfer is made to the STRCHK (STRucture CHecKing) program which provides the chemist with assistance in carrying out a number of procedures to evaluate the structural candidates,[2,8] including program modules to:

1. survey the generated structures to identify any containing standard skeletal systems, or any other specified combination of substructural features;
2. evaluate the candidates through automated spectral analysis;
3. explore stereochemical aspects of the structural problem.

None of the tests that can be performed using these program modules can guarantee selection of the correct structure from among a set of candidates. What we are looking for is an accumulation of evidence pointing to a small subset of the structures, or a single structure, as being more likely correct than the remaining structures. Verification of the correct structure must then be done by other, unambiguous methods such as total synthesis.

Surveying the structural candidates

The first step in the evaluation of the 42 candidates found for warburganal would most appropriately be surveying them for the presence of standard sesquiterpene skeleta using the SURVEY module of STRCHK. This is a simple method for testing structures under the hypothesis that there is a high probability that a new bicyclic sesquiterpene possesses a previously reported skeleton. The library of standard bicyclic sesquiterpene skeletons used in SURVEY[9] was taken from an earlier compilation,[10] which is now somewhat out of date. This compilation includes complete skeletons of reported sesquiterpenes, (e.g. the chamigrane, valerane, tutin, and trichothecane skeletons, shown in Fig. 7), and some skeletal fragments (identified by code-names like S005 and S076, also shown in Fig. 7).

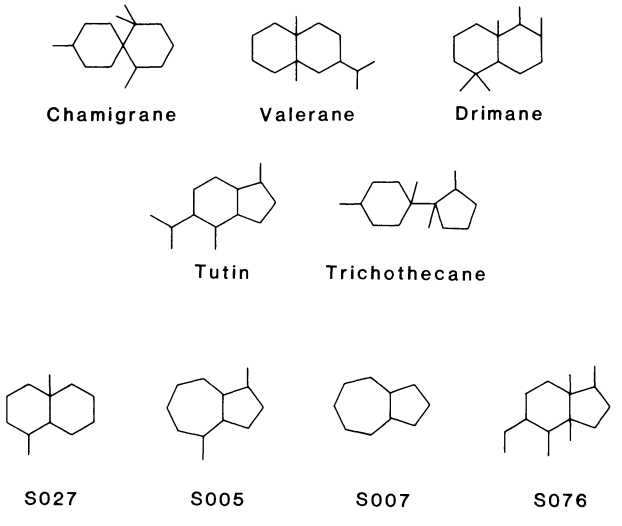


Fig. 7. Examples of reported[10] sesquiterpene skeletons and skeletal fragments.

SURVEY works by automatically matching selected items (substructures or structures) against each of the candidate structures. A record is maintained within the computer of what items were found to be present in which candidates. Subsequently, the chemist receives a report of this record and can selectively display structures that possess any given logical combination of items.

The computer interaction with the STRCHK program to perform this analysis is shown in Fig. 8 as an annotated typescript. Commands typed by the chemist are underlined in the figure.

```
Step

[1]   #SURVEY
[2]   Do you want to use a library of substructures?Y
      Which library file? BSTERP
      Which substructures  :ALL
[3]   READING ENTRIES FROM LIBRARY FILE.
      ..................................................

      SCANNING THROUGH STRUCTURES.
      ...................................

      THE FOLLOWING LIBRARY FEATURES WERE NOT FOUND IN ANY STRUCTURE.
      CHAMIGRANE        VALERANE       TUTIN          CARYOPHYLLANE
      CUPARANE          ACORANE        CAROTANE       PSEUDO-GUIAIANE
          .                 .              .                .
          .                 .              .                .
      TRICHOTHECANE     S023           S016           S076

      STRUCTURES WITH DISCRIMINATING FEATURES:
                 1     DRIMANE
                 2     S027
                 8     S005
                 8     S007
[4]   Do you want to select structures with combinations of features?Y
      ->SELECT
      Desired features >S005
      8 structures.
      ->INDEX
      8 structures currently selected.
      Index numbers are:
          6   7   12   13   26   27   37   38
      ->RESET
      42 structures.
[5]   ->SELECT DRIMANE
      1 structures
      ->INDEX
      1 structures currently selected.
      Index numbers are:
          16
      ->DONE
```

Fig. 8. Results of SURVEY's comparison of the 42 candidate structures for warburganal against a library of bicyclic sesquiterpene skeletons.

With reference to Fig. 8, first the SURVEY module is run (step 1). In step 2, a request is made for all substructures in the library file, BSTERP, which contains computer representations of the bicyclic sesquiterpene skeletons. These are the substructures that will be matched against the candidate structures.

In step 3, SURVEY informs the chemist of its progress by printing out one dot for each skeleton read from the file (55 total skeletons). Each candidate structure is analyzed in turn, by comparing each of the 55 skeletons to the structure, with one dot printed on the terminal on completion of the comparison, 42 dots total. This comparison takes only a few seconds of computer time. SURVEY then prints out its report, first summarizing those skeletons not found in any structure (this list has been abbreviated in the figure). If there were any skeletons common to all 42 candidates, they would be printed out next. There are none for this problem, so this output is skipped. Finally, a report of discriminating features is made. These are skeletons that were found in less than 42 of the structures and thus represent differences among the candidates. This report indicates that only one structure possesses a complete skeleton, the drimane skeleton (see Fig. 7). The skeletal fragments S027, S005, S007 (see Fig. 7) are found in 2, 8, and 8 structures respectively.

In step 4, Fig. 8, the chemist chooses to determine precisely which structures possess these discriminating features by specifying which feature, in this case S005, a perhydroazulene skeleton. The command INDEX prints out the structure numbers of those possessing this skeleton; the result can be verified by checking the list of structure numbers against the structures in Fig. 6. The structures can be drawn on the chemist's terminal at this point if desired.

In step 5, the single structure possessing the drimane skeleton is selected. It is index #16 in the set of 42 candidates, and corresponds to the assigned structure of warburganal.

Prediction of spectral properties

Although many spectral data are exploited in the initial phases of data interpretation and structure generation, it is often possible to obtain considerably more structural information from the data once candidate structures have been produced. During spectral interpretation, an analyst must rely on spectral-feature/substructure correlations which generally yield only small fragments of a much larger structure.

However, once candidate structures are available more extensive investigations can be made exploiting spectral correlations that depend on larger substructures, for example mass spectral fragmentations and ^{13}C chemical shifts, and on complete structures, for example configurational stereochemistry. Given complete structures, we can predict the spectral properties of each candidate structure on the basis of a simple algorithmic model and rank the candidates on the basis of comparison between predicted and observed spectra.

Mass Spectral Prediction and Ranking. There are a number of advantages in using mass spectral prediction and structure ranking procedures as an aid to structure evaluation. Although configurational stereochemical differences can sometimes induce subtle changes in the mass spectra of stereoisomers, most of the features in a compound's mass spectrum can be adequately described in terms of cleavages of the bonds in a constitutional (topological) representation of that structure. In contrast, prediction and ranking using ^{13}C n.m.r. spectra, discussed below, require consideration of configurational stereochemistry. A representation which included molecular conformation would be required for other techniques such as ^{1}H n.m.r. or circular dichroism.

A further advantage of mass spectral analysis is that we can exploit some fairly general models describing how molecules fragment within a mass spectrometer. Empirical methods for predicting other spectral properties usually depend on data bases of specific rules derived from standard reference compounds.

The 42 candidate structures for warburganal were ranked according to their compatibility with the recorded high resolution mass spectrum. The "half-order" model was used for mass spectral predictions,[11] using the program module MSANALYZE. Given a set of constraints upon the complexity of fragmentation processes allowed, the "half-order" model finds the most plausible fragmentation process that, applied to a given structure, could lead to each specific observed ion. A score is derived that describes how readily the observed mass spectrum can be rationalized in terms of simple fragmentations of a structure. These scores are used to rank the various candidates.[11] Our goal in this analysis is not to identify the correct structure unambiguously, but to obtain a subset of candidates each of which yields a significantly better explanation of the mass spectrum than any candidate in the remainder of the set.

In this example, MSANALYZE was restricted to simple single-step fragmentations involving, at most, a ring cleavage and the transfer of one hydrogen atom into or out from the ion. More complex fragmentation processes can be allowed if desired, together with specification of substructural templates that express enhanced plausibility of fragmentations such as α-cleavage.[11]

In Table 4 we summarize the results of prediction and ranking. MSANALYZE provides as output information on the distribution of scores, a tabular presentation of the distribution of structures with the highest scores and a list of the top-ranked structures by index number and score (abbreviated here to save space). The scores range from a high of 24 to a low of 9 (the maximum score of 100 would indicate that every observed ion in the spectrum was predicted with plausibility 1[11]). Two candidates, #16 and #17 (Fig. 6), rationalize the observed mass spectrum better than any of the other forty candidates. Structure #16 is the drimane-type sesquiterpene identified previously by SURVEY in the search against the file of bicyclic sesquiterpenes, and corresponds to the assigned structure of warburganal.

TABLE 4. Results of mass spectral prediction and ranking for warburganal

Distribution of scores.
Maximum : 24
Mean : 13
Median : 12
Mode : 12
Minimum : 9

Score	24	23	22	21	20	19	18	17	16	15	14	13	12
Frequency	2	0	0	0	0	0	0	4	3	2	7	2	9

Top ranked structures and their scores :
#16 24
#17 24
8 17
#26 17
#27 17
.
.
.

Generation of Configurational Stereoisomers. There are several other techniques which we might apply to evaluate further a set of structural candidates. For example, we might wish to perform spectrum prediction and ranking using n.m.r. data or construct geometric representations of the structures for viewing on a computer graphics system. These techniques require stereochemical representations of the structures. Thus, a logical next step is to obtain the set of configurational stereoisomers for the candidates. This, too, can now be done with the aid of a computer program. This program called STEREO, is itself a constrained structure generator, but devoted strictly to obtaining a set of candidate stereoisomers under stereochemical constraints.[12-14]

In our example of warburganal, there were few data given which allow formulation of a large number of stereochemical constraints. The molecule is clearly chiral from the observed circular dichroism spectrum.[3] The results of experiments using lanthanide shift reagents revealed no shifts of the methyl singlets indicating they are remote from the hydroxyl or disposed on opposite sides of a ring if on adjacent carbons. We can presume that the double bond, if in a ring, is in the cis configuration.

These constraints can be applied to the list of possible stereoisomers of the 42 structures. The results are summarized in Table 5. There are 576 possible stereoisomers for these structures and all are chiral so this constraint does not help. Application of the constraint which restricts double bonds to be cis in rings under size 8 reduces this total to 336 stereoisomers for 42 constitutional structures. Application of the constraint which forbids a cis relationship between the methyl and hydroxyl in a ring of size under 8 reduces the total to constitutional structures with 152 stereoisomers. Notice that all of the stereoisomers of some of the structures were eliminated by this constraint. These are primarily structures with the hydroxyl adjacent to a gem-dimethyl in a ring and some polycyclic structures in which the cis relationship is forced for one or more of the rings.

TABLE 5. Summary of the cumulative effects of applying stereochemical constraints to the 576 stereoisomers of the 42 structures

Constraint	Number of Structures	Total Number of Stereoisomers
none	42	576
chiral	42	576
cis C=C	42	336
no cis CH_3,OH	28	152

For this particular problem, the absence of more detailed stereochemical constraints does not allow one to reduce the number of structural candidates further. In other problems, however, detailed constraints can focus quickly on the correct structure.[14]

Prediction and ranking using ^{13}C n.m.r. data. So far we have used the ^{13}C n.m.r. spectrum only to derive very crude constraints such as C-SP2 and CH2 (Table 2). We now illustrate how the ^{13}C n.m.r. spectrum can be used to suggest which of the 42 candidates are best choices for the structure of warburganal in a manner analogous to use of mass spectral data described in a previous section. We have developed a method for prediction of the ^{13}C n.m.r. spectrum

of a structure. Prediction of the spectra of all candidate structures followed by comparison of the predicted and observed spectra allows us to rank-order the candidates on the basis of agreement between predicted and observed spectra. Our method has been described in detail in two papers;[15,16] a brief summary follows herewith.

We have assembled a library, or data base, consisting of descriptions of carbon atoms and their substructural environments together with associated chemical shifts. The library currently contains approximately twenty thousand unique substructures (out to a four bond radius, hereafter referred to as a four shell environment) gleaned from about 1500 compounds (predominantly steroids, terpenoids, and alkaloids).

The data base is used for predicting a spectrum by determining a chemical shift range expected for each of the carbon atoms in a structure. First, the carbon atom and its substructural environment are characterized by a special code. Second, the data base is searched for all occurrences of this code. Associated resonances are retrieved; the predicted resonance for the carbon atom is the mean value of the range of retrieved shifts. If the shell 4 environment is not found in the data base, the successively smaller environments are scanned until a match is found. Thus, along with the mean resonance, the shell level at which a match was found is also reported (other statistics on the distribution of retrieved resonances are available, but are not used in prediction).

For the example of warburganal, this process results in a predicted spectrum for each of the 42 candidate structures. Each of these predicted spectra is then matched to the observed spectrum, and a score reflecting the dissimilarity of the spectra is calculated.[17]

Table 6 summarizes the results of this procedure for the 20 highest ranked structures. In the first four columns of the table are found, respectively, the rank of a particular compound, its score (based on a scoring function which we have developed which takes into account the form of the substructural codes, and possible inadequacies of the data base[17]), the mean shell level on which the predictions are based and the index number of the structure in the original list.

Structure #16 is top-ranked by this procedure, but the distribution of scores is such that, based on these data alone, at least the top ten or fifteen structures must still be considered as good possibilites. It is interesting to note that when the top ranked structures from this analysis are compared with the top ranked structures from the mass spectral analysis (Table 4), only structures #16, #8, and #27 are found in common.

TABLE 6. Results of the ranking of the predicted ^{13}C n.m.r. spectra of the 42 candidate structures (only the top 20 ranked candidates are included) with the observed ^{13}C n.m.r. spectrum of warburganal.

Rank	Score	Shell	Structure	Rank	Score	Shell	Structure
1	46.2	2.3	16	11	66.3	1.4	7
2	53.5	1.2	9	11	66.3	1.5	35
3	53.6	1.2	29	13	67.2	1.6	4
4	60.8	1.3	34	14	69.7	1.8	41
5	62.2	1.2	8	15	69.9	1.1	21
6	62.9	1.2	23	16	71.8	1.1	19
7	63.0	1.2	27	17	73.1	1.1	39
8	64.0	1.2	24	18	73.5	1.2	38
8	64.0	1.2	28	19	74.1	1.1	33
8	64.0	1.4	40	20	75.6	1.4	37

CONCLUSIONS ON STRUCTURE DETERMINATION

The constitution of warburganal is indeed represented by structure #16 of the 42 (Fig. 6). The results of our more detailed evaluation of the spectral data add significant weight to the structural assignment because structure #16 is found to possess a common sesquiterpene skeleton and to offer among the best explanations of both the mass and ^{13}C n.m.r. spectra. However, it is not possible on these data alone to assign the structure unambiguously. One method of proof is unambiguous synthesis, and as mentioned in the introduction, several syntheses have been carried out,[4] thereby establishing the structure of warburganal as #16.

Our presentation describes one important use of our computer programs, as an aid to verification of a structural assignment. It is simple to determine what other candidates remain as possibilities based on existing structural information. Because all candidates are available for examination, structural assignment reduces to the problem of eliminating all but one of the structures. A second important use of these programs is from the very beginning of a problem to help guide collection of new experimental data based on examination of intermediate results, the cases described earlier. In this way, large numbers of structures can be eliminated from further consideration by recognition and removal of those cases

possessing implausible combinations of substructures.

STRUCTURE/ACTIVITY CORRELATIONS IN WARBURGANAL AND RELATED COMPOUNDS

Warburganal and a number of related compounds possess potent anti-feedant activity against African army worms. In addition, these same molecules are powerful helicocides (snail-killers). It is conceivable that such compounds might be exploited for controlling pest insects in the field or in crop storage. The helicocidal properties are also of considerable interest because a number of diseases, such as schistosomiasis, are spread by parasitic nematodes transmitted by snails. It is the bioactivity of these compounds, in addition to the challenge presented by their unusual structures with enal and hydroxy-aldehyde units in the same ring, that has helped inspire the numerous synthetic studies which have yielded a number of analogs. Several of these compounds have been tested for their biological activity and correlations between specific structural features and activity have been developed.[5]

We will illustrate the use of computer techniques in examining the relationship between molecular structure and biological activity, using warburganal and several naturally occurring, active analogs, all of which are illustrated in Fig. 9. Given a set of structures displaying similar biological activity, our methods are designed to determine common features of the structures that might give rise to the activity. Here we are interested in common geometric features, assuming that the activities are due to the actual structures in their three-dimensional representations.

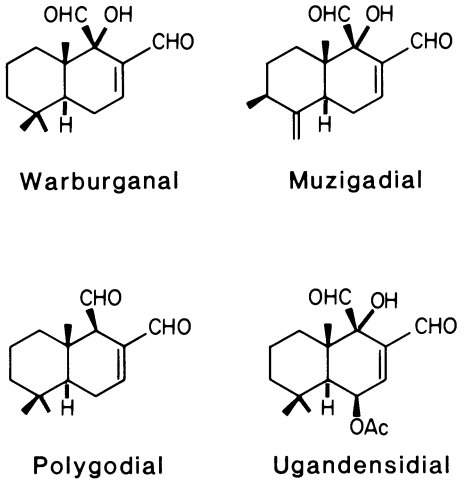


Fig. 9. Warburganal and three other naturally occurring helicocides.

Deriving three-dimensional representations of warburganal and related compounds

In the absence of X-ray crystallographic data on compounds, as is the case for warburganal and its analogs, our BUILD3D[18] program can be used to help determine three-dimensional, X,Y,Z, coordinate-based representations for structures. The connection table augmented with stereochemical information for each compound of interest, i.e., the atom, bond and stereochemical designations given in Fig. 9, is used in the determination of cartesian coordinates for all atoms, based upon the method of Crippen, et al.[9] These coordinates allow the structure to be displayed on a graphics terminal using an interactive graphics package which is part of the BUILD3D program.

The limitation of this method is that the first guess of the atom positions is random. Thus the method yields only a random sampling of the conformational space and is ill-suited to a systematic exploration of possible conformations. To overcome this, a method of optimizing geometries has been coupled with BUILD3D, specifically Allinger's empirical strain energy minimization program.[20] Allinger's program works by taking a set of defined coordinates and calculating an initial energy for that structure. A new set of coordinates is then derived and a new energy is calculated. The program continues to iterate the energy minimization process until the difference in the change of the total energy has reached a minimum value. The final set of coordinates which is returned represents the energy minimized conformation of the compound.

For each of the 4 compounds in Fig. 9, BUILD3D was used to generate an initial set of coordinates which were then subjected to the MMII energy minimization process to obtain a final, optimal set of coordinates. Figure 10 illustrates the calculated, energy minimized conformation of warburganal.

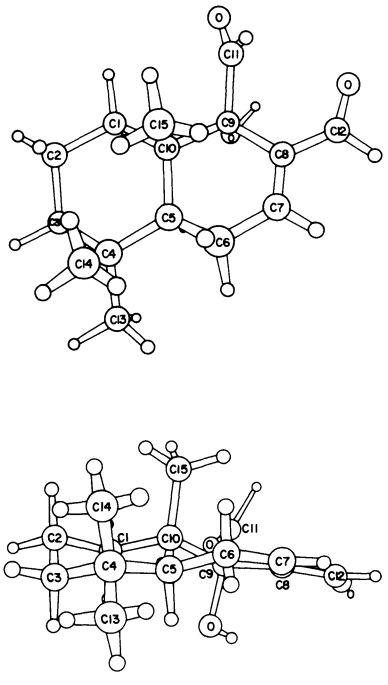


Fig. 10. Two views of the energy minimized conformation obtained for warburganal. Hydrogen atoms are unlabeled and unnumbered.

Search for common three-dimensional substructures

Starting from the point of coordinate-based representations, previous approaches to determining common features, or substructures, have involved use of computer graphics to superimpose two or more structures to view their common (overlapping) features, or to compute the degree of overlap of a set of structures.[21,22] These approaches have been limited by the total number of structures which can be compared and by initial assumptions as to where each structure is placed in the coordinate space. We have developed a somewhat more general approach to this problem of identifying three-dimensional common structures.

Briefly, our search for common substructures (which represents work currently in progress) begins with a set of biologically active compounds, their three-dimensional structures, and a set of definitions of what atom properties are important (e.g. atom type, hybridization, degree, and aromaticity are presently implemented as such constraints) as supplied by the analyst. A preprocessing step transforms interatomic distances into distance "types", atom properties into a condensed descriptor, and then puts this information into a distance matrix that is a record of all the interatomic distances in a structure between pairs of atoms, bonded or nonbonded. This distance matrix is an efficient way of storing information in a form which is easily accessible, and which is of the desired format for use by subsequent processing steps. From this point on, it is simply a matter of determining common two atom substructures, then using the common substructures from this first generation as parents from which to grow the common substructures for the succeeding generation and iterating until no further common substructures are found. Thus, each new generation is one atom larger in size than the previous generation. Preliminary details of this procedure have been described in an earlier paper.[23]

A summary of the numbers of common substructures for each generation is given in Table 7. The table provides a summary of the total number of different types of substructures, along with the total number of representative node-sets for each generation to provide a sense of the magnitude of the problem we are solving. Each node-set is a unique occurrence of a substructure in any of the structures. For smaller common substructures, there may be several occurrences of a given substructure in each structure. For large substructures there tends to be only a single occurrence in each structure. Thus, for generation 10, Table 7, there are five common substructures of size 10 and a total of 20 node-sets representing a single occurrence of each substructure in each of the four compounds (Fig. 9).

TABLE 7. Summary of the numbers of common substructures of different sizes. The first column gives the size of the substructure, the second column gives the total number of common substructures, and the third column gives the total number of representative node-sets making up the substructures.

Generation	Substructures	Node-sets
2	31	486
3	161	1567
4	435	2164
5	534	2244
6	448	1811
7	269	1076
8	114	456
9	33	132
10	5	20

Figure 11 illustrates the node-sets, using the structure of warburganal as an example, making up the five common substructures of size 10 for warburganal, muzigadial, polygodial and ugandensidial (Fig. 9). These common substructures include portions of both the A and B rings of the compounds and thus represent extensions of the commonality of the ring B dial functionalities noted in the previous study.[5] In addition, our program has verified that these substructures are common geometrically, adding powerful support to the previous hypotheses relating structures to activities.

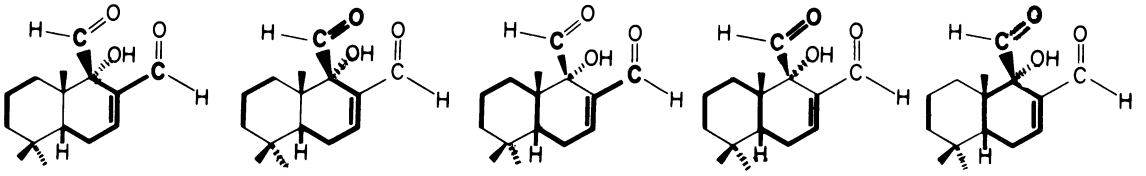


Fig 11. The five common substructures of size 10 illustrated for warburganal. Bonds connecting common atoms have been marked for clarity, but are not formally considered part of the substructure because connectivity is not a consideration in the search for three-dimensional common substructures.

A close examination of the five substructures illustrated in Fig. 11 leads one to believe that a number of substructures of larger size, incorporating additional ring A and B atoms could have been obtained. Further investigation reveals that this is indeed the case if stereochemistry is ignored. Our method is specially designed to include stereochemistry, specifically, distinction between enantiomeric substructures,[2,3] because it is well-known that enantiomers often display markedly different biological activities. Our current method for stereochemical differentiation is proving to be too sensitive in situations where four-atom substructures (the smallest substructure at which stereochemical differentiation is possible) have all four atoms nearly in a plane. In such instances, slight differences in the geometries of the structures result in effective inversion of configuration for some of the substructures, which are then no longer viewed as being in common. These kinds of problems are characteristic of new methods. For the present, we can examine the potential commonality of larger substructures in our example by taking the (stereochemically) common substructures of a smaller size and continuing the generation process ignoring stereochemistry.

Beginning with common substructures of size 8, we obtain by this method two common substructures of size 11, illustrated in Fig. 12. The second must be regarded as an artifact of the rigid models obtained from the molecular modeling program, because it is unlikely that the oxygen of the C(11) aldehyde functionality would indeed be so conformationally locked in place.

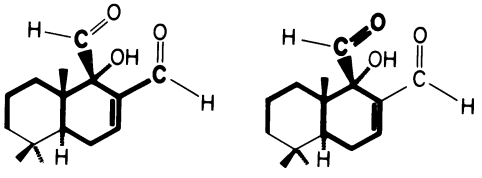


Fig. 12. The two common substructures of size 11, again using warburganal to illustrate the substructure. As before, bonds connecting substructural atoms are marked for easier perception, but are only implicitly part of the computed substructures.

We suspect that our results have over-determined the pharmacophoric pattern required for activity. All four structures are based on the same basic ring system, so it is not surprising that elements of the ring system itself are found to be in common. The answer to this question must await further studies on synthetic analogs based on different ring systems. The important point is that now computer programs are available to help chemists in the detailed comparison of geometric representations of structures for common patterns without the difficulties inherent in manual methods and the biases inherent in previous computer-based approaches to the problem.

Acknowledgement - We wish to thank the National Institutes of Health for their generous financial support (RR 00612) and for their support of the SUMEX computer resource on which the programs were developed (RR 0785). We would also like to thank the National Science Foundation for its support (CHE81-08064) of our spectroscopic facilities, and Craig Barnes for his assistance in obtaining and interpreting some of the spectra presented herein. The programs described in this paper are available through Molecular Design, Ltd., 1122 B St., Hayward, CA 94541.

REFERENCES

1. H. Egli, D. H. Smith, and C. Djerassi, Helv. Chim. Acta (1982), submitted for publication.
2. D. H. Smith, N. A. B. Gray, J. G. Nourse, and C. W. Crandell, Anal. Chim. Acta, Computer Techniques and Optimization 133, 471-479 (1981).
3. I. Kubo, Y-W Lee, M. Pettei, F. Pilkiewicz, and K. Nakanishi, J. Chem. Soc., Chem. Commun., 1013-1014 (1976).
4. A. Ohsuka and A. Matsukawa, Chem. Lett. 635-636 (1979); A. J. G. M. Peterse, J. H. Roskam, and A. de Groot, Rec. Trav. Chim. Pays-Bas 97, 249-278 (1978); S. P. Tanis and K. Nakanishi, J. Am.Chem. Soc. 101, 4398-4400 (1979); T. Nakata, H. Akita, T. Naito, and T. Oishi, Chem. Pharm. Bull. 28, 2172-2177 (1980); A. S. Kende and T. J. Blacklock, Tetrahedron Lett. 21, 3119-3122 (1980); D. J. Goldsmith and H. S. Kezar,III, ibid. 21, 3543-3546 (1980); S. V. Ley and M. Mahon, ibid. 22, 3909-3912 (1981); P. A. Wender and S. L. Eck, ibid, 23, 1871-1874 (1982).
5. K. Nakanishi and I. Kubo, Israel J. Chem. 16, 28-31 (1977).
6. R. E. Carhart, D. H. Smith, N. A. B. Gray, J. G. Nourse, and C. Djerassi, J. Org. Chem. 46, 1708-1718 (1981).
7. R. E. Carhart, D. H. Smith, H. Brown, and C. Djerassi, J. Am. Chem. Soc. 97, 5755-5762 (1975).
8. C. Djerassi, D. H. Smith, and T. H. Varkony, Naturwiss. 66, 9-21 (1979).
9. D. H. Smith and R. E. Carhart, Tetrahedron 32, 2513-2519 (1976).
10. T. K. Devon and A. I. Scott, Handbook on Naturally Occurring Compounds, Vol. II, Terpenes, Academic Press, New York, 1972.
11. N. A. B. Gray, R. E. Carhart, A. Lavanchy, D. H. Smith, T. Varkony, B. G. Buchanan, W. C. White, and L. Creary, Anal. Chem. 52, 1095-1102 (1980).
12. J. G. Nourse, J. Am. Chem. Soc., 101, 1210-1216 (1979).
13. J. G. Nourse, R. E. Carhart, D. H. Smith, and C. Djerassi, J. Am. Chem. Soc. 101, 1216-1223 (1979)
14. J. G. Nourse, D. H. Smith, and C. Djerassi, J. Am. Chem. Soc. 102, 6289-6295 (1980).
15. N. A. B. Gray, C. W. Crandell, J. G. Nourse, D. H. Smith, M. L. Dageforde, and C. Djerassi, J. Org. Chem. 46, 703-715 (1981)
16. N. A. B. Gray, J. G. Nourse, C. W. Crandell, D. H. Smith and C. Djerassi, Org. Magn. Reson. 15, 375-389 (1981).
17. C. W. Crandell, N. A. B. Gray, and D. H. Smith, J. Chem. Inf. Comp. Sci. 22, 48-57 (1982).
18. J. C. Wenger and D. H. Smith, J. Chem. Inf. Comp. Sci. 22,29-34(1982).
19. G. M. Crippen, J. Comput. Phys. 24, 29-34 (1977).
20. N. L. Allinger, J. Am. Chem. Soc. 99, 8127-8134 (1977).
21. D. C. Rohrer and H. Perry, "FITMOL" in Public Procedures: A Program Exchange for PROPHET Users, J. J. Wood, ed., Bolt, Beranek and Newman, Inc., Cambridge, Mass., (1978).
22. N. C. Cohen, Computer Assisted Drug Design, E. C. Olson and R. E. Christoffersen, eds., Am. Chem. Soc., Washington, D. C., 377-381 (1979)
23. D. H. Smith, J. G. Nourse, and C. W. Crandell, Structure Activity Correlation as a Predictive Tool in Toxicology, L. Goldberg, ed., Hemisphere Publishing Corp., New York, 1982 (in press).