Tetrahedron. Vol 32, pp 1807 to 1810. Pergamon Press 1976 Printed in Great Britain

APPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR CHEMICAL INFERENCE—XXI[1]

CHEMICAL STUDIES OF MARINE INVERTEBRATES—XVII.[2] THE COMPUTER-ASSISTED IDENTIFICATION OF [+]-PALUSTROL IN THE MARINE ORGANISM *CESPITULARIA* sp., aff. *SUBVIRIDIS*

C. J. CHEER,[3] DENNIS H. SMITH and CARL DJERASSI*
Department of Chemistry, Stanford University, Stanford, CA 94305, U.S.A.

and

B. TURSCH, J. C. BRAEKMAN[3] and D. DALOZE
Collectif de Bio-écologie, Unité de Chimie Bio-organique, Université Libre de Bruxelles, Belgique

(*Received in UK* 21 *January* 1976; *accepted for publication* 16 *February* 1976)

Abstract—In order to demonstrate the power and utility of the interactive computer program CONGEN,[5] an example describing its use in the identification of the tricyclic sesquiterpene alcohol, [+]-palustrol, in extracts of a marine Xeniid (*Cespitularia* sp., aff *subviridis*) is provided. The rotation of [+]-palustrol further supports the remarkable antipodal relationship between sesquiterpenes from marine Coelenterates and their corresponding terrestrial forms.[6]

INTRODUCTION

The "art" of chemical structure elucidation normally involves the process of piecing together structural fragments inferred from a variety of sources of information. Chemical experience and intuition frequently play a significant role in constraining and directing the construction of plausible structural candidates during this process. CONGEN[5] (for CONstrained structure GENeration) is an interactive computer program which allows the chemist to utilize his intuition and experience in conjunction with a systematic and thorough procedure for constructing molecular structures.

Structure generation by CONGEN involves assembly of atoms and molecular fragments (called "superatoms"[5]) into intermediate structures followed by a technique known as "imbedding".[5] The intermediate structures (e.g. see **8–12** below) are constructed using only the names of superatoms and thus each may represent a whole class of final structures. The ability to apply (interactively) constraints at all stages of structure generation and imbedding allows the user to guide the process of structure generation and thus prevent construction of unmanageably large numbers of undesired structures. Examination of intermediate structures frequently suggests additional constraints which were previously overlooked.

Although the scenario presented in this example leads to the identification of the enantiomorph of a previously reported substance, it nevertheless provides a cogent demonstration of the practical utility of the program in structure elucidation problems.

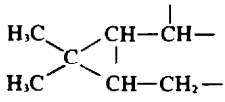
Structural information

Mass spectroscopic analysis established the molecular formula to be $C_{15}H_{26}O$. The presence of an OH group was evident from the IR spectrum. The 60 MHz ^{1}H NMR spectrum exhibited signals suggestive of cyclopropyl methine hydrogens, between 0.4 and 0.8 ppm and Me groups at 0.9 to 1.1 ppm. There were no signals below 2.4 ppm. The ^{13}C NMR spectrum exhibited signals due to all carbons and the following assignments were based on the undecoupled spectrum: –C–OH (s, 85.48 ppm downfield from internal TMS); –C– (s, 19.79); 5x –CH (d's, 47.11, 46.09, 34.91, 27.06 and 21.97); 4x $-CH_2-$ (t's 38.83, 32.73, 32.22 and 24.45); and 4x $-CH_3$ (q's, 28.81, 18.55, 17.64 and 15.52). The 360 MHz ^{1}H NMR spectrum indicated the presence of two Me doublets at 1.10 and 1.18 ppm, and two Me singlets at 1.23 and 1.27 ppm. Examination of changes of the ^{1}H Me resonances produced upon addition of Eu(fod)$_3$ to deuteriochloroform solutions of the alcohol indicated a profound downfield shift of one of the Me doublets. The two Me singlets and remaining doublet were only weakly affected. This behavior strongly suggests that there are no Me groups directly attached to the carbinol carbon. The 360 MHz ^{1}H spectrum clearly exhibits only two cyclopropyl hydrogens at 0.625 ppm (d × d) and 0.80 ppm (d × d × d). Both quaternary methyls must therefore be attached to the lone quaternary carbon, and the presence of only two cyclopropyl hydrogens fixes this carbon within the 3-membered ring. The multiplicities of the cyclopropyl hydrogen signals further indicate the attachment of methine and methylene groups.

DISCUSSION

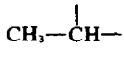
From the above spectral data and corresponding structural inferences we may define the superatom **1**, (arbitrarily named **A**) where bonds with an unspecified terminus are called "free valences".[5,7] These free valences represent the connection points by which superatoms are attached to other atoms or superatoms in the generation of structures. Because the degree of each carbon atom in **1** which bears free valences is known, the additional constraint that these free valences be bonded to nonhydrogen atoms is further imposed Similarly, from the ^{13}C and ^{1}H NMR spectra it is evident that the hydroxyl function is tertiary and no H atoms are permitted to be

attached to the carbinol carbon in the superatom **2** (named **B**).



1(A)

2(B)



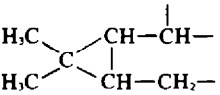
3(ETH)

The unknown alcohol contains two remaining Me groups, both attached to methine carbons. However, at this point in the problem, available evidence does not exclude the attachment of one of these methyls to the methine carbon present in superatom **A(1)** and therefore only one methyl-methine attachment may be explicitly defined. This is done in the third superatom structure **3** (named **ETH**, above).

In addition to superatoms **A**, **B** and **ETH**, the unknown structure contains one additional Me, three methylenes, and one methine group. These are defined as superatoms CH3, CH2 and CH respectively. These superatoms comprise a COMPOSITION† list shown in Table 1.

While it would be possible to direct CONGEN to generate all possible structures by systematically connecting the superatoms of the COMPOSITION list it is possible and desirable to "constrain"[7] this process using additional information. From the ^{13}C NMR spectrum it is evident that there are no multiple bonds present in the unknown structure. In addition, there is only one cyclopropane ring, and the absence of ethyl and isopropyl groups is evident from the ^{1}H spectra. We may then impose the constraints summarized in Table 2 on the initial generation of intermediate structures. The absence of double bonds is implemented by the entry 2 as a BADRINGS constraint. The GOODRINGS entry 3 (exactly 1‡) insures that all structures generated contain one and only one cyclopropyl ring. **MET** represents the class of substructures (**4**) in which superatom CH3 is bonded to any of the parenthetically enclosed superatoms CH2, **B** and **ETH**. The BADLIST entre **MET** specifically

Table 1. The superatoms used to generate initial structures

Structure[a]	Superatom name	Number
	**A**	1
	**B**	1
CH_3—	CH3	1
—CH_2—	CH2	3
	CH	1
	**ETH**	1

[a] In this example all free valences are bonded to non-hydrogen atoms. The program also provides the option of permitting selected atoms to be associated with a specific range of hydrogen atoms.

†The COMPOSITION list is the list of atoms and/or superatoms which comprise the empirical formula.

‡Since this ring is already defined in superatom A, this constraint prevents assembly of structures with additional 3-membered rings.

Table 2. CONSTRAINTS LIST[a] used for generation of intermediate structures

Constraint type	Entry[a]
BADRINGS	2
GOODRINGS	3 (exactly 1)
BADLIST	**MET**[b] (structure 4)

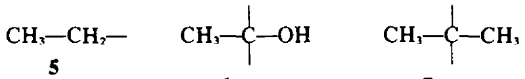
[a] Entry refers to responses given by the user to requests from CONGEN.

[b] See text for explanation.

prevents generation of all structures containing the molecular fragments **5–7**.

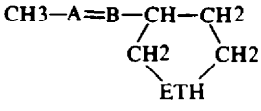
CH3–(CH2, **B**, **ETH**)

4 (MET)

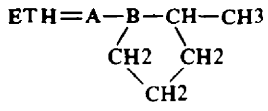


5 **6** **7**

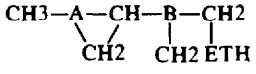
A total of 177 intermediate structures were generated at this stage, of which **8–12** are representative examples. Recall that each of these intermediate structures may



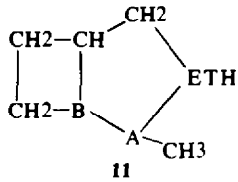
8



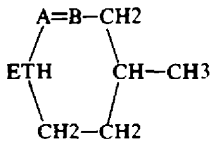
9



10



11



12

represent a whole class of final structures since there may be more than one way of imbedding unsymmetrical polyvalent superatoms within intermediate structures. (Monovalent superatoms can be imbedded in only one way). This concept is best illustrated by the sequential imbedding of the superatoms, summarized in Table 3.

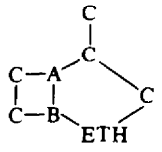
Table 3. Effect of the imbedding process on the number of structures

Superatom imbedded	Number of resulting structures
CH3	177
CH2 (3x)	177
CH	177
ETH	124[a]
B	124
A	272[b]

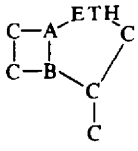
[a] Some previously unrecognized duplications are now discarded.

[b] Since no superatoms remain, these are final structures.

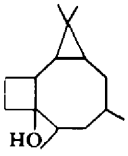
Imbedding the Me (CH3), methylene (CH2) and methine (CH) superatoms occurs without affecting the number of intermediate structures. Upon imbedding the superatom **ETH** the number of intermediate structures is reduced since half of all equivalent structural pairs, such as **13** and **14**, previously unrecognized as identical are now eliminated. Imbedding superatom **A** yields 272 final structures. This represents a more than doubling of the number of structures after imbedding superatom **B** (Table 3) since each intermediate structure may lead to up to three final structures. For example, intermediate structure **13** (or **14**) leads to final structures **15–17**. Further constraints were obviously needed at this stage to reduce the problem to a manageable size.



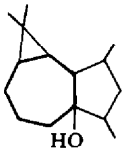
13



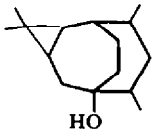
14



15

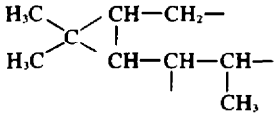


16



17

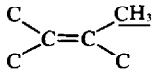
Homonuclear decoupling of the 360 MHz ^{1}H spectrum (Table 4) revealed the relationship of one Me group to the cyclopropane ring. This relationship is defined in the substructure **18(HEP)** shown below. When the list of 272



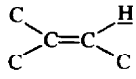
18(HEP)

structures was "pruned"[5] with the GOODLIST entry of at least one **HEP**, only 88 structures survived. While this additional constraint significantly reduces the scope of the problem, further reductions are highly desirable.

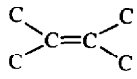
Dehydration of the unknown alcohol (phosphoryl chloride/pyridine, steambath, 30 minutes) produced a mixture of three new compounds (in the ratio of 1:6:14), isolated by preparative TLC on silver nitrate impregnated silica gel. The ^{1}H NMR spectrum of the major component exhibited a new methyl resonance at 1.59 ppm indicating direct attachment to a double bond (**19**). The second component did now show any Me resonances below 1.1 ppm but did display a single vinyl hydrogen at 5.1 ppm (**20**). The third and minor component exhibited neither vinyl methyl nor vinyl hydrogen signals (**21**). The partial structures of the three dehydration products (**19–21**) permit the inference of the substructure **22(PEN)** describing the immediate environment of the OH group. At this point we must bear in mind the inherent weakness of the



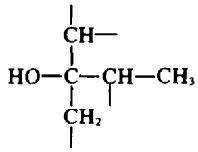
19



20

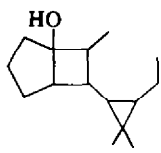


21

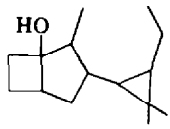


22(PEN)

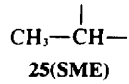
assumption that dehydration occurs without concomitant skeletal rearrangement, a notoriously common occurrence in terpene chemistry. Aware of this caveat, it is still informative to prune the current list of 88 structures with the GOODLIST constraint of at least one **PEN** substructure present. Twenty-two of the 88 structures survive this pruning. Examination of the 22 remaining structures revealed that we had failed to test for the presence of Et groups during the imbedding of **A**. Two structures (**23** and **24**) were pruned from the list of 22 with the GOODLIST constraint requiring exactly 2 **SME(25)** substructures to be present.



23



24



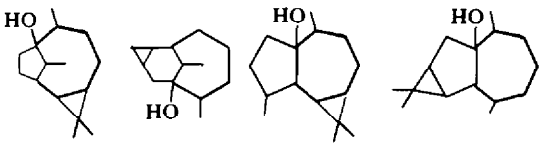
25(SME)

At this point the problem is no longer unmanageable and a variety of avenues exist for further reduction. It is instructive to briefly digress to consider some of these

Table 4. Summary of 360 MHz NMR spectrum of (+)-palustrol

Signal	Intensity	δ ppm ($CDCl_3$)	Multiplicity	J_{Hz}	Coupling
A	1	0.52	d×d	9.0, 11.2	
B	1	0.65	d×d×d	6.7, 9.0, 10.5	
C	3	0.92	d	6.5	
D	3	0.99	d	7.2	
E	3	1.0	s		
F	3	1.06	s		
G	6	1.21–1.54	m		
H	1	1.60	bs		
I	1	1.60	d×d	11.2, 11.2	
J	3	1.72–2.21	m		
K	1	2.15	d×q	7.2, 11.2	

alternatives before describing the actual method employed. For example, examination of the list of 20 remaining structures reveals that in 7 of these the OH group is bonded to the bridgehead of a bridged bicyclic system (e.g. see **26** and **27**) and may be excluded on the basis of Bredt's rule. Alternatively, the list of structures may be surveyed generally for the presence of isoprenoids, or specifically compared to a library file of known terpene skeletons.[6,8] Implementation of this latter test was in a developmental phase† at this point of the problem and the list of compounds was therefore pruned to exclude non-isoprenoids. Initially the constraint was strictly applied, and only those isoprenoids with two head-to-tail linkages were allowed. Application of this constraint resulted in 4 surviving structures (**26–29**). Of these, **26** and **27** may be excluded on the basis of Bredt's rule. Structure **28** contains the known aromadendrane skeleton and is in



26 **27** **28** **29**

fact identical to the structure reported for (−)-palustrol.[9] The published IR spectrum of (−)-palustrol was identical in all respects to that of the unknown alcohol, thus confirming its structure. Palustrol isolated from *Ledum palustre*[10] is levoratatory ($[\alpha]_D^{20} = -17.8°$) while the compound isolated from *C. sp. aff. subviridis* is dextrorotatory ($[\alpha]_D^{20} = +14.3°$), providing another example of the curious enantiomeric relationship between sesquiterpenes from marine Coelenterates and their corresponding terrestrial forms.[6]

The approximate total time elapsed from the initiation of this problem to its conclusion was less than 3 weeks. We feel that reduction of the problem to no more than 20 structures in this period of time represents a demonstration of the significant utility of CONGEN as a practical aid to problems in structure elucidation.

EXPERIMENTAL

Optical rotations were made on a Perkin-Elmer model 141 polarimeter, in chloroform solution. NMR were obtained on a Varian T60 and XL100 NMR spectrometers and on a Bruker HXS 10 360 MHz NMR spectrometer. Low resolution mass spectra were taken on AEI MS-9, Varian MAT 711, or Atlas CH-4 mass spectrometers, while element-mapping high resolution spectra utilized the Varian MAT 711 instrument. GLC was carried out at 145° on a Hewlett Packard high efficiency instrument using a 6′ × 4 mm glass column packed with 3% OV25 on Gaschrom Q (100–120 mesh) obtained from Applied Science Laboratories. TLC was performed using 20 × 20 cm, 750 μ silica gel HF_{254} impregnated with 12% $AgNO_3$. Visualization was effected by spraying with a 2% ceric sulfate soln in 1M H_2SO_4 followed by heating.

Isolation and purification of (+)-*palustrol.* Sun-dried specimens (117 g) of *Cespitularia* sp aff. subviridis (Coelenterata, Octocorallia, Alcyonacea) collected at Albatros Rocks (Seychelles Islands) were extracted with hexane to give 3.42 g (yield 2.9%) of a viscous residue which was chromatographed on 340 g silica gel (eluents: hexane, hexane:acetone 9:1, hexane:acetone 8:2, hexane:acetone 5:5 and acetone). The fraction eluted with hexane:acetone 9:1 (1.59 g) contains mainly (+)-palustrol contaminated by a slightly more polar compound as observed by TLC. (+)-Palustrol and its companion were further purified by silica gel column chromatography using hexane:dichloromethane 8:2 as eluent. (+)-Palustrol: oil, $[\alpha]_D^{20} = +14.3°$ ($c = 0.30$; $CHCl_3$), M^+ 222.19617, *m/e* $C_{15}H_{26}O$ requires: M 222.198355.

Dehydration of (+)-*palustrol.* To an ice-cold soln of 100 mg (0.45 mmole) of (+)-palustrol in 0.5 ml dry pyridine, was added 0.1 ml $POCl_3$. The mixture was allowed to warm to room temp and then heated on the steam bath for 30 min. The resulting dark mixture was cooled, poured over ice, and extracted with hexane. The combined hexane extracts were washed successively with water, 1N HCl, saturated salt, water, dried (Na_2SO_4) and concentrated to leave 82 mg of an orange oil. GLC analysis indicated the presence of a small amount of starting material and three new compounds, A, B and C, in the ratio of 1:6:14 in order of increasing retention time. Separation was effected by preparative TLC on $AgNO_3$ impregnated silica gel. Three developments with hexane provided complete separation. The order of elution of the three components on TLC is reversed from their order of appearance on GLC, with the major component (C) exhibiting the largest R_f value. All three components were estimated to be greater than 98% pure by GLC analysis.

The individual components exhibited the following pertinent ^1H NMR signals: Component A; δ 0.88 (s, 3H, CH_3), 0.99 (d, 6 Hz, 3H, CH_3), 1.02 (d, 6 Hz, 3H, CH_3) and 1.07 (s, 3H, CH_3). The signals due to cyclopropyl hydrogens were no longer visible upfield from the Me signals suggesting conjugation of the cyclopropyl ring with the double bond.[11] There were no signals below 2.5 ppm. Component B 0.42 (d × d, 10 Hz, 1H, cyclopropyl-H), 0.53 (m, 1H, cyclopropyl-H), 1.02 (s, 3H, CH_3), 1.03 (s, 3H, CH_3), 1.04 (d, 7 Hz, 3H, CH_3), 1.10 (d, 7 Hz, 3H, CH_3), 5.20 (m, 1H, vinyl-H). Component C; 0.55 (m, 1H, cyclopropyl-H), 0.65 (d × d, 9 Hz, 9 Hz, 1H, cyclopropyl-H), 0.94 (d, 7 Hz, 3H, CH_3), 0.98 (s, 3H, CH_3), 1.05 (s, 3H, CH_3) and 1.56 (bs, 3H, vinyl-CH_3).

Acknowledgements—This work was supported by the National Institutes of Health (Grants No. GM 06840, RR 00621-05A1, and RR 785-01A1), the latter in support of the Stanford University Medical Experimental Computer Facility, SUMEX, and by N.A.T.O. (Research Grant No. E.S. 003). We thank Dr. J. Verseveldt for the identification of the animal. The 360 MHz NMR spectra were obtained at the Stanford Nuclear Magnetic Resonance Laboratory Supported by the National Institutes of Health (Grant No. RR-00711) and the National Science Foundation (Grant No. GP 23633).

REFERENCES

[1]For Part XIX in this series see D. H. Smith, J. Konopelski and C. Djerassi, *Org. Mass Spectrom* in press; for Part XX see D H. Smith and R. E. Carhart, *Computers in Chemistry*, submitted for publication.

[2]For Part XVI in this series see B. Tursch, C. Hootele, M. Kaisin, D. Losman and R. Karlsson, *Steroids* in press.

[3]Postdoctoral research associate, 1975–76, while on leave from the University of Rhode Island.

[4]Chercheur qualifié du Fonds National de la Recherche Scientifique Belge.

[5]R. E. Carhart, D. H. Smith, H. Brown and C. Djerassi, *J. Am. Chem. Soc.* **97**, 5755 (1975).

[6]L. S. Ciereszko and T. K. B Karns, *Comparative Biochemistry of Coral Reef Coelenterates*, in *Biology and Geology of Coral Reefs*, Vol. II(1), pp. 183, Academic Press, New York (1973).

[7]L. M. Masinter, N. S. Sridharan, J. Lederberg and D. H. Smith, *J. Am. Chem. Soc.* **96**, 7702 (1974).

[8]T. K. Devon and A. I. Scott, *Handbook of Naturally Occurring Compounds*, Vol. II. Academic Press, New York (1972).

[9]L Dolejs, V. Herout and F. Sorm, *Coll. Czech. Chem. Comm.* **26**, 811 (1961).

[10]N. P. Kryialov, *Z. Obsch. Chim.* **24**, 1271 (1954).

[11]A. D. Ketley and J. L. McClanahan, *J. Org. Chem.* **30**, 941 (1965).

†A subsequent computer comparison of the list of 20 compounds with all known tricyclic sesquiterpene skeletons indicated that only structure **28** contained a known skeleton.