Quick Method for Anti-Bredt Structure Detection

J. M. Nuzillard

Laboratorie de Pharmacognosie, Faculté de Pharmacie, Unité Associée au CNRS No. 492, 51 Rue Cognacq-Jay, F-51096 Reims Cedex, France

Received June 29, 1993®

A simple algorithm for the search of anti-Bredt molecular structures is described. It relies on a topological criterion and requires neither the search of the smallest set of smallest rings of the molecular graph nor adjacency matrix manipulation. This algorithm, first implemented in Prolog and then in C language, is used in the LSD program (logic for structure determination) to discard unrealistic structures. The present article is the first report of an attempt to sort computer-generated organic structures according to this type of steric constraint.

INTRODUCTION

The automated structural analysis of organic compounds is a problem of great theoretical and practical importance. The programs related to the well-known Dendral[1] project provide an example of how computers can be used to solve chemical problems. Similar programs described to date deal with small molecules. They are intended to give the practicing chemist planar molecular formulas consistent with MS, UV, IR, and NMR (mainly ^{1}H and ^{13}C NMR) spectroscopic data.[2] Resulting structures are first represented as ordinary molecular drawings, in which bond lengths and bond angles are not significant. The next step for the chemist is the determination of the stereochemistry from coupling constants and nuclear Overhauser effects (NOE). This kind of structural data is generally used by programs handling macromolecules.[3] However, efforts have been made to rationalize ^{13}C chemical shifts as a source of stereochemical information.[4] The aim of this paper is to show how a particular class of highly strained molecules can be easily discarded on the basis of connectivities and without resorting to energy calculations. To our knowledge such considerations have never been taken into account by authors of structure-solving programs.

The LSD program[5] uses mainly direct and remote carbon–proton chemical shifts correlation maps to generate molecular structures. The analysis protocol has been tested on many compounds, and it appeared in the course of these trials that some generated planar structures could be rejected on the ground of steric constraints. Analyzing a sesquiterpenic lactone,[6] the correct structure A (Figure 1) has been found, but also the structure B, which is clearly anti-Bredt. The Bredt rule[7] has first been proposed to account for the regioselectivity of bicyclic alcohol dehydration (Figure 2). By extension, an anti-Bredt structure contains a double bond located at a bridgehead. Bredt's rule does not hold for large bicyclic systems.

PRINCIPLE

The chemical instability of anti-Bredt compounds is due to the prohibitive geometrical strain that may be induced by the coplanarity of a double bond and its connecting bonds. The evaluation of the steric energy relying on 3D structure modeling[8] may be used as a discriminating criterion for such structures. However, this way of handling the molecular stability problems is hardly possible at the topological level,

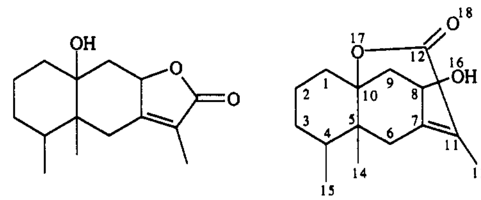

Figure 1. (A) Structure of a sesquiterpene derived by the LSD program from 1D and 2D NMR data. (B) Alternative anti-Bredt structure. Atom numbers are written according to ref 6.

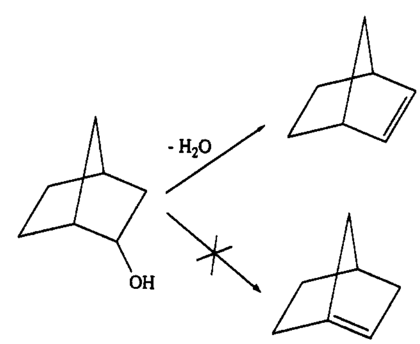

Figure 2. Origin of the Bredt rule.

as far as stereochemical information is generally not yet available. Moreover, the speed of the anti-Bredt screening process must be high in order to treat the hundreds of solutions LSD can generate when data are weakly restrictive. A well-accepted stability criterion relies on ring sizes: a seven-membered ring or a smaller one cannot contain a trans double bond.[9] This rule is the one implemented in the LSD program.

The final stage of the structure elaboration in LSD contains a filter that removes duplicated structures and checks substructural information. The position of double bonds, connecting pairs of sp^2 hydridized atoms, are also determined at this stage. Immediately after, tri- and tetra-substituted double bonds are now checked according to Bredt's rule, because they are the only ones possibly located at bridgeheads. A trisubstituted bond (Figure 3) will be first considered.

It is meaningless to decide on the sole basis of connectivities that atoms *a* and d_1 or *a* and d_2 are in a cis or trans arrangement. However, it is sure that if *a* and d_1 are in the cis position, then *a* and d_2 must be in trans position. In other words, the a–b–c–d_1 and a–b–c–d_2 paths must not both be embedded in rings of size less than or equal to 7. This means that neither d_1 nor d_2 belongs to the set V_4 of the atoms x, verifying $d(x,a) \leq 4$,

® Abstract published in *Advance ACS Abstracts*, March 15, 1994.

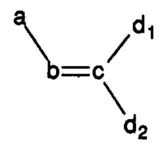

Figure 3. Atom labeling of a trisubstituted double bond.

so that the path from a to x does not contain either b or c. The distance d is the number of bonds between a and x. The problem is now reduced to the construction of the set V_4. Questions of this kind are well-documented and could involve the manipulation of the molecular graph's adjacency matrix.[10] The current and previous versions of LSD do not use this structure representation. In LSD, each atom a is associated to the set $N(a)$ of its neighbors, and hence V_4 is determined from the redundant connectivity table of the molecule. Two series of sets V_i and W_i ($1 \leq i \leq 4$) are built. They are recursively defined by

$$W_1 = \{x/x \in N(a), x \notin \{b,c\}\}$$

$$V_n = \bigcup_{i=1}^{n} W_i$$

$$W_i = \{x/x \in \bigcup_{j \in W_{i-1}} N(j), x \notin V_{i-1} \cup \{a,b,c\}\} \qquad i > 1$$

The process is stopped when $n = 4$ or $W_i = \varnothing$. Practically an algorithm has been written that returns V_4 as a function of a, b, and c.

The complete analysis of a molecule is performed by a second algorithm:

for each sp[2] atom b bonded to two or three atoms
 find the sp[2] atom c bound to b
 if c is bonded to three atoms then
  find $d_1 \neq b$ and $d_2 \neq b$, $d_2 > d_1$ bonded to c
  for each atom $a \neq c$ bonded to b
   determine $V_4(a, b, c)$
   if $d_1 \in V_4$ and $d_2 \in V_4$ then return FALSE

return TRUE.

The returned value provides the validity of the proposed molecular structure according to the topological stability criterion cited hereabove. A tetrasubstituted double bond is treated as four trisubstituted double bonds: thus all cis–trans relationships between substituents are explored.

A more global approach would start with the search of the smallest set of smallest rings of the molecule. The determination of the double bond position, relative to bridgeheads, also yields the desired topological information. The corresponding algorithm was found to be unnecessarily complex to be written in PROLOG, the language in which the LSD program was first implemented. The LSD program, including the Bredt rule checker, has recently been rewritten in C language, without noticeable changes in the algorithms.

EXAMPLES

Structure B of Figure 1 does not survive the Bredt rule test described here. With $b = 11$, $c = 7$, $a = 12$, we obtained successively $W_1 = \{17,18\} = V_1$, $W_2 = \{10\}$, $V_2 = \{10,17,18\}$, $W_3 = \{1,5,9\}$, $V_3 = \{1,5,9,10,17,18\}$, $W_4 = \{2,4,6,8,14\}$, $V_4 = \{1,2,4,5,6,8,9,10,14,17,18\}$. Both atoms $d_1 = 6$ and $d_2 = 8$ belong to $V_4(12,11,7)$. This proves that the double bond between atoms 7 and 11 is embedded in the two rings 7–11–12–17–10–9–8–7 and 7–11–12–17–10–5–6–7, of size less than or equal (equal here) to 7. Either atom 6 or atom 8 is a substituent in the trans position relative to atom 12. Solution structure B must therefore be discarded.

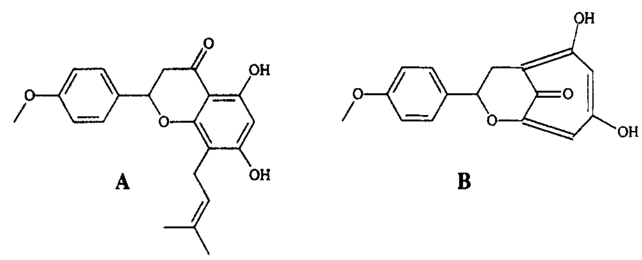

Figure 4. (A) Structure of a flavonoid derived by the LSD program from 1D and 2D NMR data. (B) Strained alternative structure.

The energetic approach to an automated stability analysis would require the modeling of all the possible diastereoisomers of identical constitution formula and must therefore be much more time consuming. Moreover, the strain caused by double bond nonplanarity may arise in situations where it may be difficult to detect by the use of simple rules. For example, structure A of Figure 4 is the correct answer to a structural problem investigated by LSD. Structure B is not in contradiction with the Bredt rule, although it is unrealistic. The carbonyl group is subjected to a strong steric interaction with the surrounding double bonds. Even though this group could be replaced by another one, as small as possible, the bicyclic system would then contain two trans double bonds in a nine-membered ring, resulting in an unacceptably strained compound. These more complex cases are not treated by the present algorithm.

CONCLUSION

This work was undertaken in order to easily and quickly detect highly strained molecular structures rather than to provide a quantitative strain analysis similar to the one given by molecular modeling calculations. The price of the simplicity is the acceptance of molecules in which the strain is not caused by the inappropriate location of a single double bond.

REFERENCES AND NOTES

(1) Smith, D. H.; Gray, N. A. B.; Nourse, I. G.; Crandell, C. W. The Dendral project: Recent advances in computer-assisted structure elucidation. *Anal. Chim. Acta* **1981**, *133*, 471–497.
(2) (a) Dubois, J. E.; Carabedian, M.; Dagane, I. Computer-aided elucidation of structures by carbon-13 nuclear magnetic resonance. The DARC-EPIOS method: characterisation of ordered substructures by correlating the chemical shifts of their bonded carbon atoms. *Anal. Chim. Acta* **1984**, *158*, 217–233. (b) Christie, B. D.; Munk, M. E. The role of two-dimensional nuclear magnetic resonance spectroscopy in computer-enhanced structure elucidation. *J. Am. Chem. Soc.* **1991**, *113*, 3750–3757. (c) Funatsu, K.; Susuta, Y.; Sasaki, S. Introduction of two-dimensional NMR spectral information to an automated structure elucidation system, CHEMICS. Utilization of 2D INADEQUATE information. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 6–11.
(3) Duncan, B.; Buchanan, B. G.; Hayes-Roth, B.; Lichtarge, O.; Altman, R.; Brinkley, J.; Hewett, M.; Cornelius, C.; Jardetzky, O. PROTEAN: a new method for deriving solution structures of proteins. *Bull. Magn. Reson.* **1986**, *8*, 111–119.
(4) Gastmans, J. P.; Zurita, J. C.; Sahao, J. R.; de Emerenciano, V. Prévision des spectres de résonance magnétique nucléaire de [13]C par intelligence artificielle: le probléme de la codification. *Anal. Chim. Acta* **1989**, *217*, 85–100.
(5) Nuzillard, J. M.; Massiot, G. Logic for structure determination. *Tetrahedron* **1991**, *47*, 3655–3664.
(6) Massiot, G.; Nuzillard, J. M.; Le Men-Olivier, L.; Aclinou, P.; Benkouider, A.; Khelifa, A. Eremophilenolides from *Hertia cherifolia*. *Phytochemistry* **1990**, *29*, 2207–2210.
(7) Bredt, J.; Thouet, H.; Schmitz, J. Steric hindrance in the bridge ring (Bredt's rule) and the *meso-trans*-position in condensed ring systems of hexamethylenes. *Liebigs Ann. Chem.* **1924**, *437*, 1–13.
(8) Mayer, W. F.; von Ragué Sclayer, P. Evaluation and prediction of the stability of bridgehead olefins. *J. Am. Chem. Soc.* **1981**, *103*, 1891–1900.
(9) Wiseman, J. R.; Pletcher, W. A. Bredt's rule. 3. The synthesis and chemistry of bibyclo[3.3.1]non-1-ene. *J. Am. Chem. Soc.* **1970**, *92*, 956–962.
(10) Gondran, M.; Minoux, M. *Graphes et Algorithmes*; Eyrolles: Paris, 1979; p 74.