Tetrahedron, Vol. 37, pp. 1711 to 1721, 1981
Printed in Great Britain. All rights reserved

SCRIPT: INTERACTIVE MOLECULAR GEOMETRICAL TREATMENTS ON THE BASIS OF COMPUTER-DRAWN CHEMICAL FORMULA

N. C. COHEN,* P. COLIN and G. LEMOINE
Centre de Recherches Roussel-UCLAF, 102 route de Noisy, 93230 Romainville, France

(*Received in France* 9 *July* 1980)

Abstract—SCRIPT is a conversational computer program which reveals the conformers and their relative energies for a given organic compound. The exploration of the conformational potential surface is based on a molecular mechanics model using a strain energy minimization criterion. The description of the investigated molecule consists of the drawing of its chemical formula on a cathode ray tube. The perception of the molecule through transforming the chemical formula into all the necessary molecular descriptors (topologic, geometric and energetic) is entirely automatic. The logical basis for the generation of the characteristic starting points representing all the possible conformers of the molecule is presented, and some specific algorithms of general utility in the conformational analysis of cyclic moieties are developed. Some features of the program are shown in a study of 17 substituted Δ^{14}-unsaturated steroids in conjunction with Toromanoff's theory of the steric course of the *cis* addition to cyclo-olefins. Our calculations show that bulky substituents at 17α lead to a more stable *quasi-cis* C/D conformation which cannot be constructed with standard Dreiding models and, to our knowledge, would be the first example of such an unusual geometry of the C/D moiety of Δ^{14}-steroids.

In the course of our theoretical studies on the conformations and the molecular geometries of organic molecules, it appeared necessary to reduce the required input data for the computer to the minimum, in order to avoid the laborious preparations for encoding the molecular descriptors and checking the corresponding data. The mere drawing of the chemical formula on a cathode ray tube connected to a computer appeared to be a possible basis aiming at this goal and permitting one to conduct subsequent molecular geometrical treatments such as that involving a molecular mechanical model.

In particular we have solved the problem of the conformational treatment of complicated molecules having, for example, fused, polycyclic, spiro and bridged ring systems, where the generation of all the representative points on the conformational potential surface is entirely controlled by the program without any external supervision.[1] The logical basis of the method lies on conformational assembly rules presented here as a function of the hybridizations of the atoms and the stereochemistries of the atoms of the molecules concerned, and not on empirical time-consuming and inaccurate trials of geometrical complementarity.[2]

In the present approach which is described here, conformational moieties are vehiculated with the aid of the "conformational diagram" symbolism,[3] which consists of a representation of the compound's chemical formula, with a description of the sign of torsion angles for each fragment of the molecule (elementary rings and chain fragments).

For example, Fig. 1 shows two possible conformational diagrams of one tricyclic molecule.

The advantage of this formalism is that it provides a logical basis for analysing the conformational possibilities of the studied compound, which also appears practical for the computer manipulation of conformational entities.

The 3-dimensional structures which are representative points on the conformational potential surface are then directly constructed on the basis of the generated conformational diagrams without any trial and error.

GENERAL FEATURES OF THE SCRIPT PROGRAM: MOLECULAR FRAGMENTS AND CONFORMATIONAL DIAGRAMS

The molecule is considered in the Script program as the assembly of chain and ring fragments, possibly possessing geometrical constraints.

Chain fragments

A chain fragment is defined as a four atom sequence with possible rotation around the central bond; the geometrical constraints in this case are the relationships between the torsion angles around the central bond with respect to the specific hybridization and stereochemistry of its two atoms; for example in Fig. 2 each of the nine torsion angles around the 2–3-central bond can be expressed as a function of one of them (i.e. 1–2–3–4) and a constant of 0, $(2\Pi/3)$ or $(4\Pi/3)$ value. For example we have $\Phi\ (6,2,3,8) = \Phi(1,2,3,4) + (4\Pi/3)$.

Torsional constraints for the constitution of the conformational diagrams for the chains

In the case of the rotation around bonds the conformational diagrams are represented by the description of the normal equilibrium position around the central

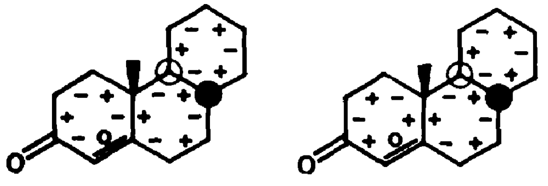

A/B *quasi-trans* A/B *quasi-cis*

Fig. 1. Conformational diagrams: two possible conformers of one tricyclic molecule (Klyne's convention[4] for the signs of dihedral angles).

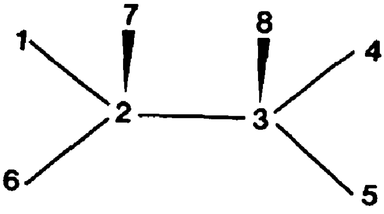


Fig. 2. Chain fragment.

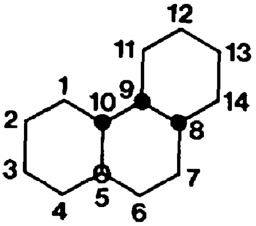


Fig. 3. Ring fragment.

bond. According to the type of fragment concerned, the number of equilibrium positions varies from one (e.g. sp^2=sp^2 double bond) to six (e.g. sp^3-aromatic bond) with the more frequent value being three (as for example in sp^3–sp^3 σ bonds).

Ring fragments

A ring fragment is defined as a minimum set of independent fused rings, and the geometrical constraints are those involving the torsion angles around the junctions between the rings: for example in Fig. 3 the main geometrical constraints are: for the *trans*-fused rings, $\Phi(1,10,5,4) = -\Phi(9,10,5,6)$; for the *cis*-fused rings, $\Phi(10,9,8,7) = \Phi(11,9,8,14)$.

GENERAL TORSIONAL CONSTRAINS FOR THE CONSTITUTION OF CONFORMATIONAL DIAGRAMS FOR THE RINGS

Ring conformers

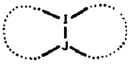
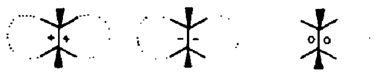
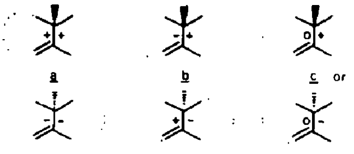
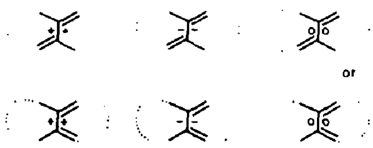
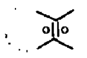
All the conformers of rings from three to eight have been introduced in special routines consisting of the sequence of the torsion angles of the specific forms chosen for these rings. The different forms are described in Appendix I, each ring being divided into two conformer's groups: level 1 (the more stable) and level 2 (higher energy). Level 2 is only activated on demand, otherwise Script constructs the conformers merely on the basis of level 1 forms.

For a ring of more than eight atoms (maximum 10-membered rings) the forms have to be introduced manually onto the screen.

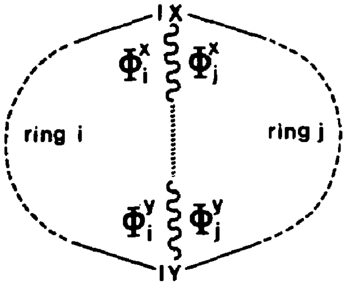
(a) *Fused rings.* When two fused rings have two atoms in common, the general constraints rules for the torsion angles which have been considered, are dependent on the hybridizations and the stereochemistries of the common atoms. Table 1 illustrates the conformational assembly rules of a carbon–carbon junction. Similar constrains have been formulated for junction with one or two nitrogen atoms with the possible cases: sp^3, sp^2 or enamine nitrogen.

(b) *Bridged ring systems.* In the case of bridged ring systems the fusion's conditions consist of torsional con-

Table 1. Constraints rules for fused rings (C–C junction)* *Note*: in the particular case of atom J being aromatic, only the *quasi-trans* diagram (IIb) is considered

	HYBRIDIZATIONS			Number of forms	CONSTRAINTS RULES FOR FUSED RINGS (C-C Junction) 
	Atom I	Atom J	Bond I-J		DIAGRAM OF THE CONSTRAINTS
I	sp^3	sp^3	σ	(cis) 3	
				(trans) 1	
II*	sp^3	sp^2	σ	3	
III	sp^2	sp^2	σ	3	
IV	sp?	sp^2	π	1	

ditions combined with stereochemical conditions. With regard to the former, four torsion angles involving the bridgehead atoms are considered (see Fig. 4), and the corresponding conditions express that the signs of these four angles must be altered as given in the following expressions:



$$\Phi_i^x \cdot \Phi_i^y < 0$$

$$\Phi_j^x \cdot \Phi_j^y < 0$$

$$\Phi_i^x \cdot \Phi_j^x < 0$$

$$\Phi_i^y \cdot \Phi_j^y < 0$$

Fig. 4. Torsional conditions: the torsion angles involved in a bridged ring fragment.

The stereochemical conditions have to be used for all sp^3 bridgehead atoms. These conditions relate the sign of the torsion angle Φ (IA, IB, IC, ID) (Klyne convention) to the sign of the rotation angle ω (trigonometric convention) if one follows the sequence IA–IB–IC as described in Fig. 5 and Table 2.

These stereochemical conditions enable one to reduce the set of elementary ring conformers which are required, while the torsion conditions permit one to obtain the only possible valid ring assembly.

The combinatory product of the conformational diagrams for rings and chains will produce all possible conformational diagrams of the molecule, therefore identifying all possible conformers.

Generation of the 3-D entities

Once the conformational diagrams have been obtained, three dimensional structures are subsequently constructed as follows. A tree structure is first generated onto the molecular frame. The function of this tree is to define all the atoms of the molecule through a set of independent internal coordinates (bond lengths, bond angles and torsion angles). The 3-dimensional construction becomes possible if the topological tree is transformed into a three dimensional tree. The latter is possible if we associate numerical values to the connections (bond lengths), to adjacent bonds (valency angles) and to four connected atom sequences (torsion angles) of this tree.

The different conformers will depend on the numerical values of the torsion angles defining it.

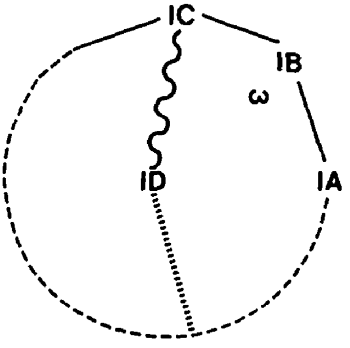


Fig. 5. Stereochemical conditions: the stereochemistry of the IC bridgehead atom is given along the IC–ID bond.

While there is no particular problem in associating numerical values to the "bonds" and "valency angles" of the tree, it is necessary to pay some particular attention to the generation of the "torsion angles" values (the conformational diagrams do not have all the necessary torsion angles; for example in the case of Fig. 3 the torsion angle Φ(8–10, 14) is not known, and all the torsion angles of the tree are not included in the torsion angles of the conformational diagram). Any torsion angle of a given conformer can be obtained on the basis of the conformational diagram with the knowledge of the stereochemistries, the orientations of the substituents (axial, equatorial) and the hybridizations of the central atoms. Specific algorithms have been conceived to this end and are of such general importance in conformational analysis that we describe these algorithms in detail in Appendix II.

Calculating the energies of the conformers

When the three dimensional conformers are constructed, it is possible to calculate the corresponding molecular energy. The energy used in the actual SCRIPT version is the sum of different terms such as stretching, bending, torsional, van der Waals, electrostatic and hydrogen bond, as generally used in molecular mechanics theoretical models.[5] A force field atomic code has been used to simplify the calculation of these energy contributions.

Minimization

Any of the rough generated structures can then be relaxed according to all its degrees of freedom in order to reach the corresponding minimum on the conformational potential surface; the precise details of the minimized

Table 2. Stereochemical conditions for bridged rings

ω(IA−IB−IC)	Stereochemistry of IC − ID	Sign of Φ(IA · IB · IC · ID)
+	α	−
+	β	+
−	α	+
−	β	−

conformer are given geometrically and energetically onto the printer.

Manipulations

It is possible to manipulate any of the three dimensional entities interactively on the screen: perspective views in a Dreiding representation, stereoviews or space filling drawings (for the molecular shape) are produced or particular answers to specific geometrical questions can be obtained. The solvent effect is calculated from the atomic exposed surface areas contributions in the interaction of the conformer with the molecules of the solvent. Our solvent model at the present time is water.[6] Some aspects of the screen are shown in Figs. 7–13.

EXAMPLE

The example we have chosen to present some of the possibilities of the program concerns 17 substituted unsaturated Δ^{14} A-nor steroids such as I and II whose conformations were studied in conjunction with Toromanoff's theory of the steric course of the suprafacial *cis*-addition to cyclopentenes.[7,8]

This theory requires, in particular, that compound II should be in a C/D *quasi-cis* conformation which is very unusual and, moreover, impossible to construct with molecular Dreiding models.

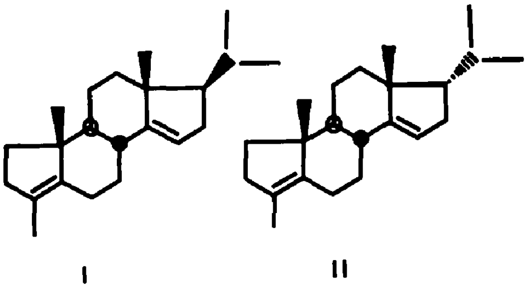

I II

Fig. 6.

With the Script program we have calculated the stable conformer of both compounds and the results are as follows. In the case of compound I, the expected C/D *quasi-trans* conformer has been found to be the more stable, whereas the C/D *quasi-cis* form has not been found as a minimum in the conformational potential surface, but as a set of points belonging to the same valley of the previously identified *quasi-trans* conformer and whose minimization always leads to this form. We have estimated the C/D *quasi-cis* forms to be approximately 5 kcal mole^{-1} over the stable conformer. In the case of compound II, the two distinct conformers have been found with the unusual result that the more stable conformer has a C/D *quasi-cis* fusion, the quasi-trans conformer being 2.4 kcal mole^{-1} over the preceding one. The calculations show that, in this case, the preference of the isopropyl substituent to be in an equatorial position has been determinant in stabilizing the C/D *quasi-cis* form. In passing, one should notice that the impossibility of constructing such a form with Dreiding models is due to the fixed standard values of the valency angles of the sp^3 and sp^2 carbons of the C/D junction. With these constraints, the fusion torsion angle of ring C would be about −70° in this conformation. Our calculations show that subtle valency angle deformation is sufficient to allow the construction of such a conformer having reasonable torsion angle values at the C/D ring junction.

To our knowledge the only X-ray study published on molecules of this type is the recent study on a Δ^{14}-17β-hydroxy steroid[9] related to structure I where the expected C/D *quasi-trans* conformation has been found.

Some aspects of the screen are shown here in the study of compound II. It shows, in particular, what the screen looks like when the drawing of the chemical

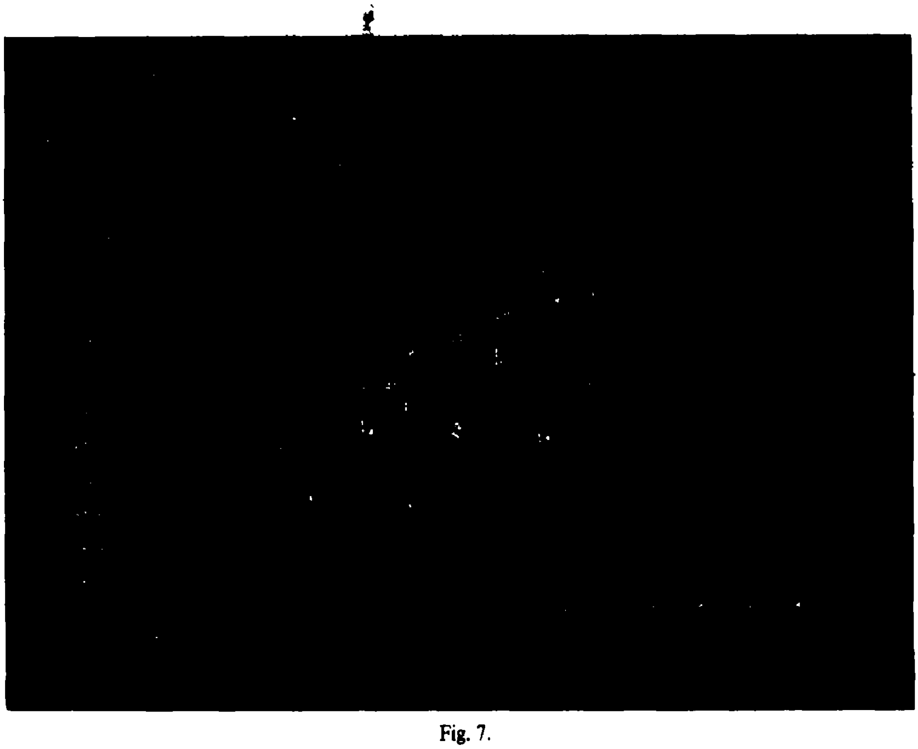


Fig. 7.

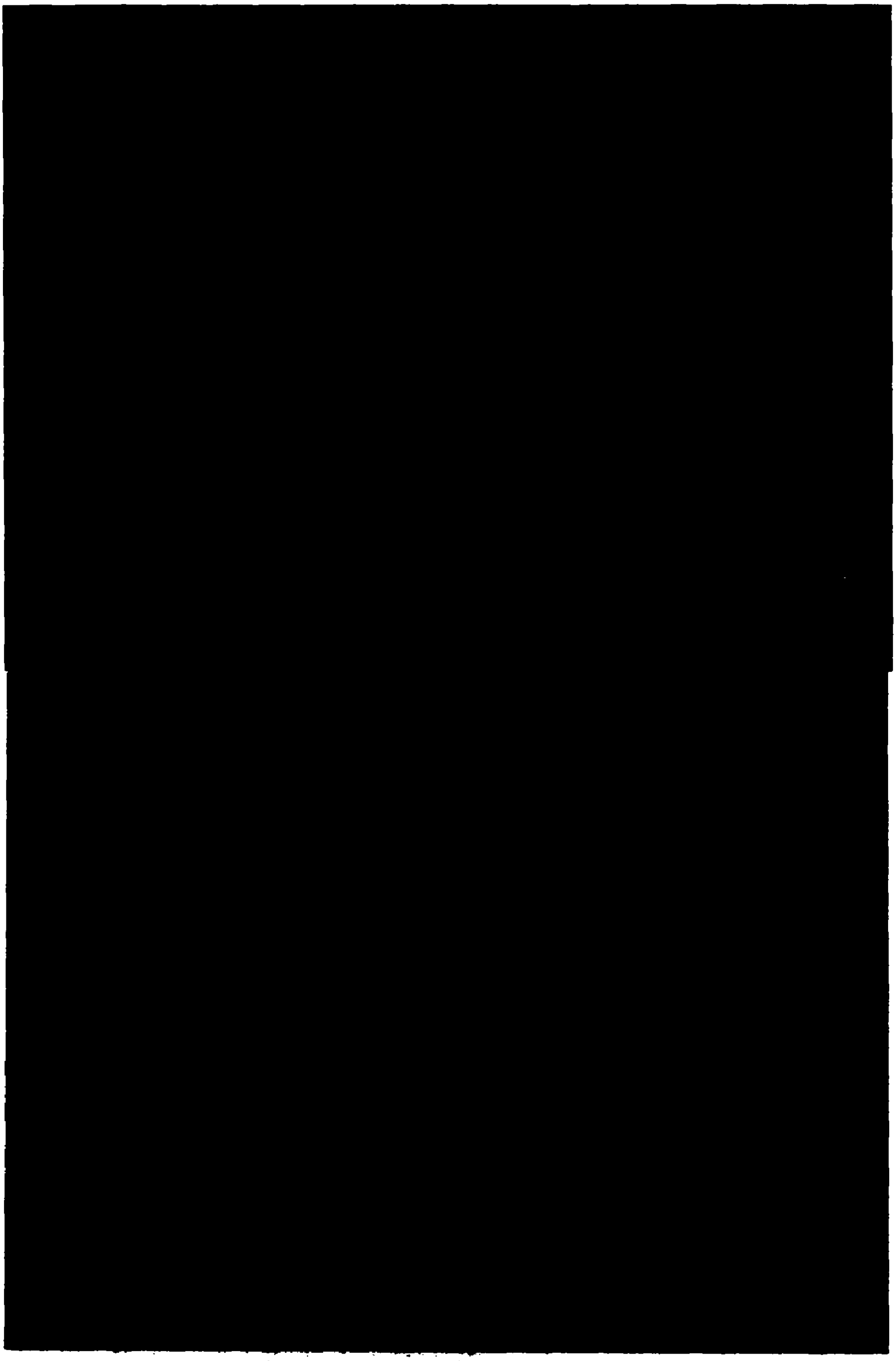


Fig. 8.

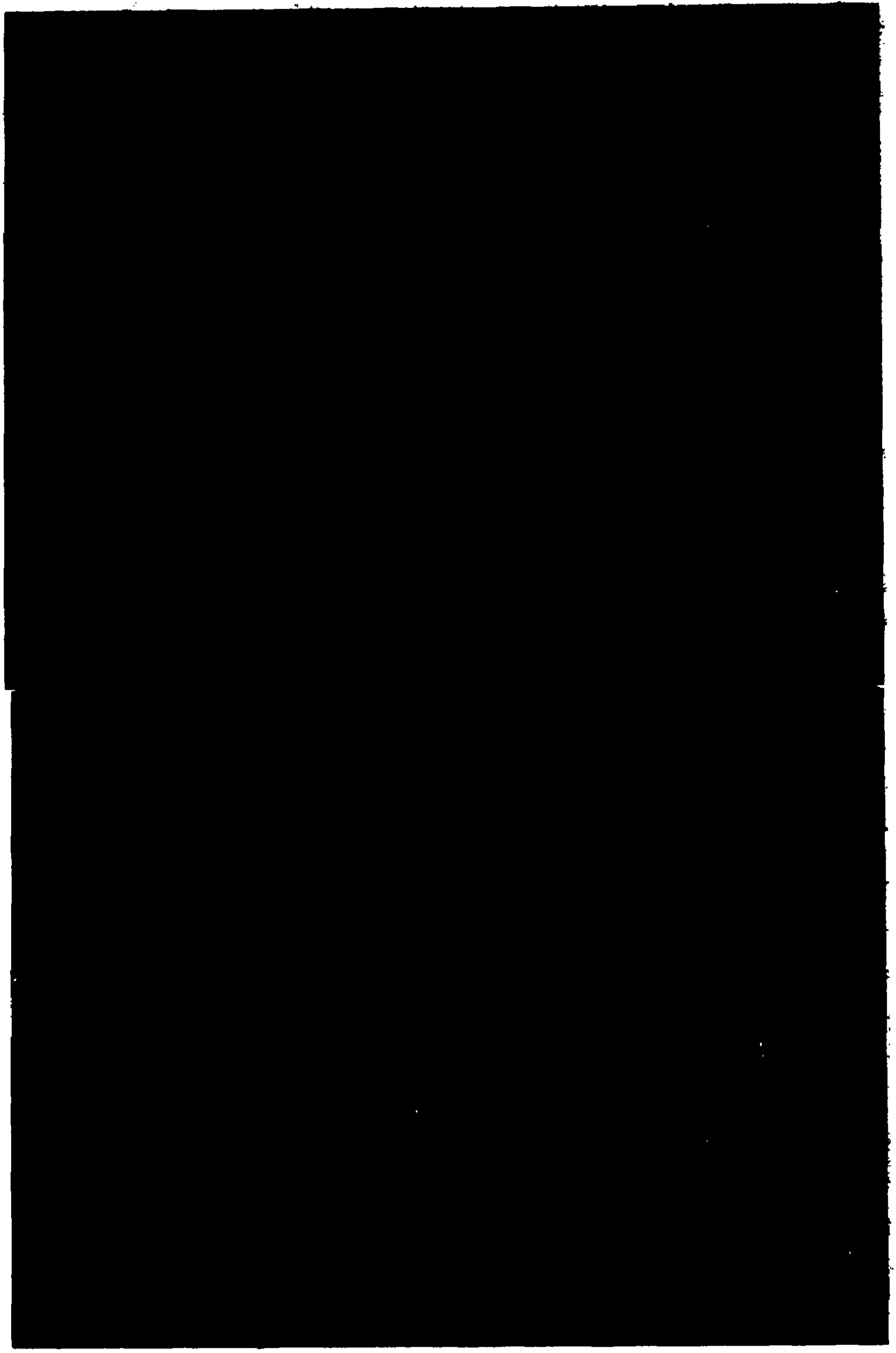


Fig. 9.

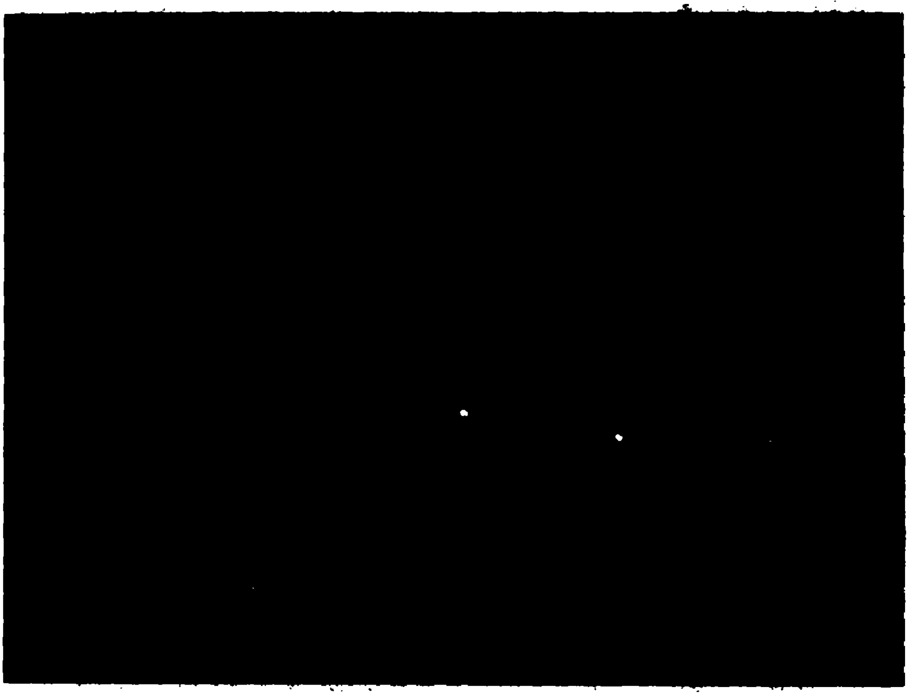


Fig. 10.

Fig. 11.

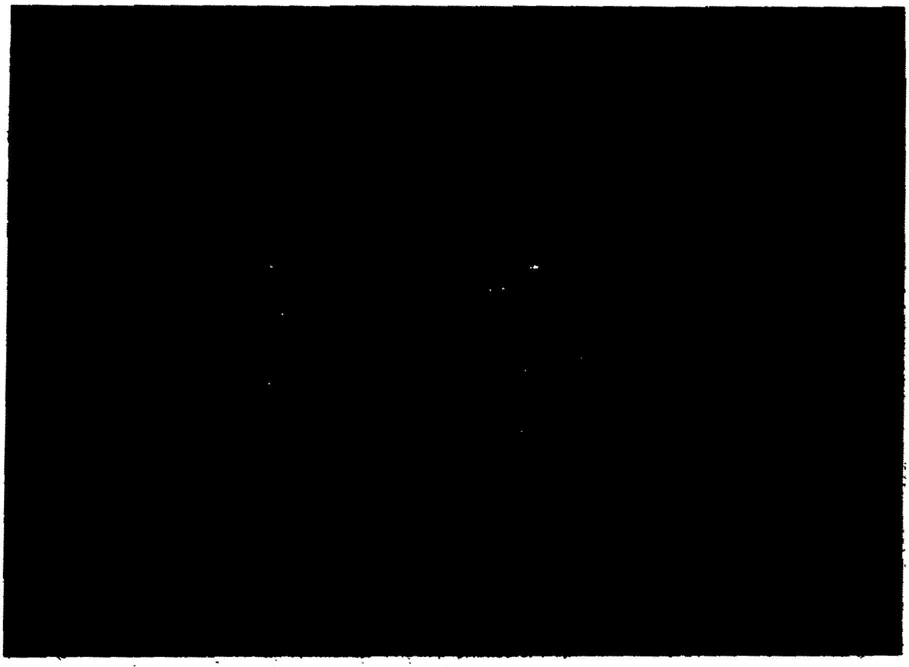


Fig. 12.

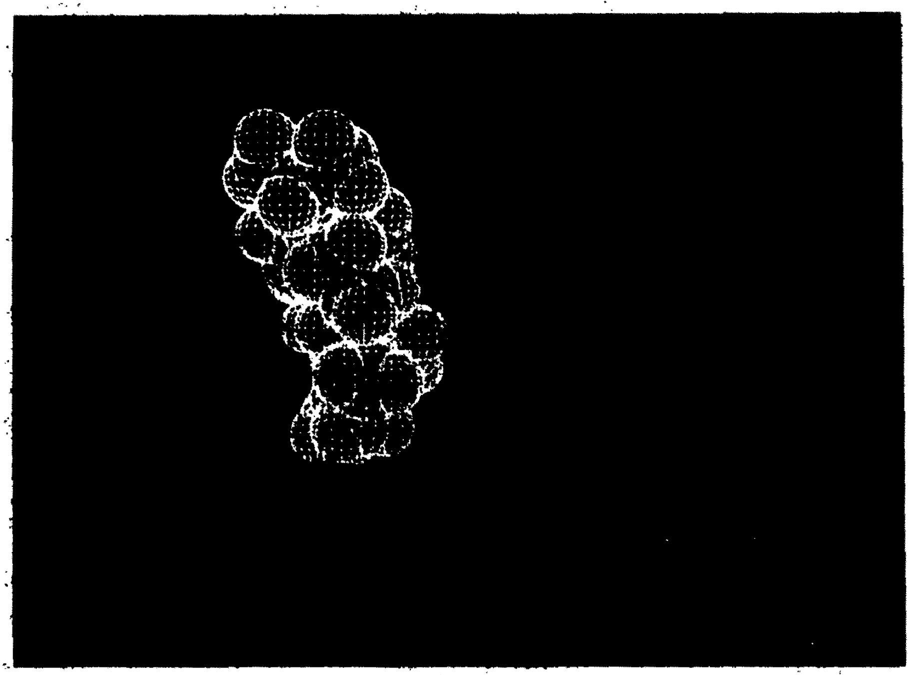


Fig. 13.

formula is complete (Fig. 7), the perspective views of the generated conformers (Fig. 8) with the corresponding conformational diagrams (Fig. 9), the minimization phase of one conformer (Fig. 10), the conformational diagram of the minimized C/D *quasi-cis* conformer (Fig. 11), stereoviews of this form (Fig. 12) together with its overall molecular shape (Fig. 13).

The program described here has been implemented on an IBM 3032 with a TSO operating system and operates in a 350k byte region; the terminal is a TEKTRONIX 4015 screen having a joy stick acting as a "light pen". SCRIPT is essentially written in FORTRAN; only one function is written in the IBM Assembly language.

Acknowledgement—Thanks are due to Mr J. E. Lemoine for programming assistance in the chemical formula input phase and the realization of the fundamental ring's search module.

REFERENCES

[1] A list of some current interactive molecular modeling systems is given in the recent paper of P. Gund, J. D. Andose, J. B. Rhodes and G. M. Smith, *Science* **208**, 1425 (1980).

[2a] W. C. Still, Emory University, Ph.D.; University Microfilms, Ann Arbor, Michigan No. 72-25, 946 (1972); [b] W. C. Still and A. J. Lewis, *Chemtech* February 1974, p. 118.

[3] R. Bucourt, *Topics in Stereochemistry* (Edited by E. L. Eliel and N. L. Allinger), Vol. 8, p. 159. Interscience, New York (1974).

[4] W. Klyne and V. Prelog, *Experientia* **16**, 521 (1960).
[5] N. C. Cohen, *Tetrahedron* **27**, 789 (1971).
[6] N. C. Cohen and G. Regnier, *Bull. Soc. Chim. Fr.* 2034 (1976).
[7] E. Toromanoff, *Tetrahedron* **35**, 893 (1979).
[8] E. J. Brunke, R. Böhm and H. Wolf, *Tetrahedron Letters* 3137 (1976).
[9] W. L. Duax, P. D. Strong and D. C. Rohrer, *Acta Cryst.* **B36**, 824 (1980).

APPENDIX I

In Table 3 the elementary conformations of the rings, which have been included in the Script program, are described.

When a general ring R is found in fusion with a planar ring (e.g. when the atoms concerned are both sp^2 like in benzene, etc.) a homologous ring R′ is constructed which differs from the R ring in that its common bond with the planar ring has been generated as a double bond.

Fig. 14.

APPENDIX II

Algorithms for establishing the values of any torsion angle of a given conformer using the knowledge of its conformational diagram.

We define a conformational diagram as representing the chemical formula with torsion angles associated to all isolated fragments of the molecule: elementary rings and chain fragments.

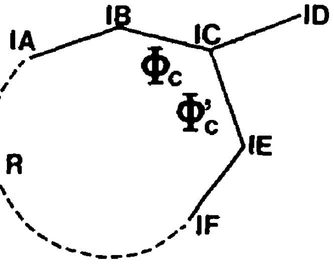


Fig. 15.

This diagram defines entirely the conformation of the compound, but there are some torsion angles, in particular those necessary for the tree, the values of which are not yet known, however, by using the appropriate algorithms, these values can be calculated as a function of the known ones. These torsion angles can exist within the following situations:

(A) *Two bonds of the three bond sequence are in the same ring R* (Fig. 15)

The problem lies in the attribution of a numerical value to the torsion angle $\Phi_0 = \Phi(\mathrm{IA, IB, IC, ID})$ with the knowledge of the value of the cyclic torsion angle $\Phi_c = \Phi(\mathrm{IA, IB, IC, IE})$.

(a) If the hybridization of atom IC is sp^2 we simply have:

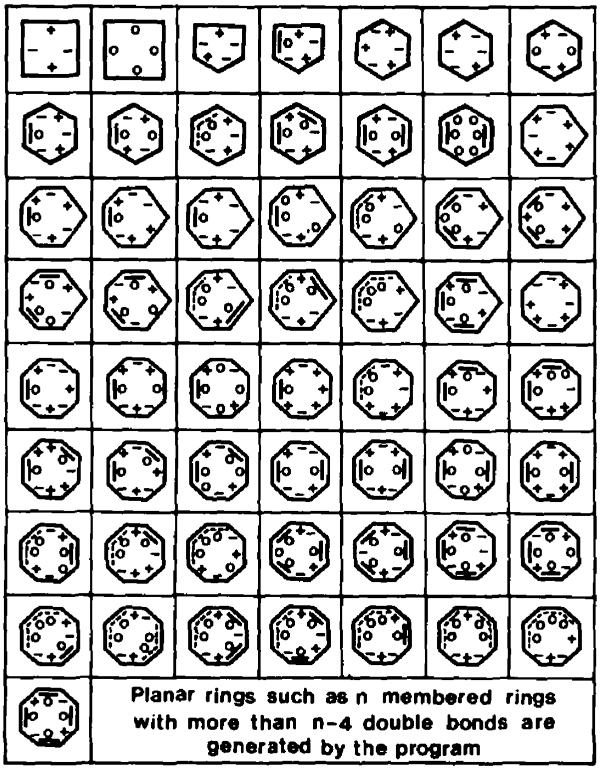
$$\Phi_0 = \Phi_c + \Pi \quad \text{(algorithm a)}$$

(b) If the hybridization of atom IC is sp^3, Φ_0 can be calculated as a function of Φ_c, Δ, ω and IOR with the following definitions:

$$\left.\begin{array}{ll} \Delta = \Phi_c - \Phi_c' & \text{if } \omega < 0 \\ \Delta = \Phi_c' - \Phi_c & \text{if } \omega > 0 \end{array}\right\} \quad \Phi_c' = \Phi(\mathrm{IB, IC, IE, IF})$$

$\omega = \overrightarrow{\Omega M_1} \cdot \overrightarrow{\Omega M_2}$ (Ω, M_1 and M_2 being the middle of segments IB–IE, IB–IC and IC–IE, respectively)

Table 3. The elementary conformations of the rings



Planar rings such as n membered rings with more than n-4 double bonds are generated by the program

Table 4. b algorithms. *Note*: F120 and F240 are constant depending on the size of the ring, the exact values are given in Table 5

Φ_c	Δ	ω	IOR	Φ(IA–IB–IC–ID)
+	+	−	1	Φ_c + F 240
+	−	−	1	Φ_c + F 120
+	+	−	0	Φ_c + F 120
+	−	−	0	Φ_c + F 240
−	+	−	1	Φ_c + F 240
−	−	−	1	Φ_c + F 120
−	+	−	0	Φ_c + F 120
−	−	−	0	Φ_c + F 240
+	+	+	1	Φ_c + F 120
+	−	+	1	Φ_c + F 240
+	+	+	0	Φ_c + F 240
+	−	+	0	Φ_c + F 120
−	+	+	1	Φ_c + F 120
−	−	+	1	Φ_c + F 240
−	+	+	0	Φ_c + F 240
−	−	+	0	Φ_c + F 120

Table 5. Angular values of the algorithmic constants according to the size of the ring

ring	3	4	≥ 5
P60	72.5	63.8	60
P120	145.0	127.6	120
F120	107.5	116.2	120
P240	215.0	232.4	240
F240	252.5	243.8	240
P300	287.5	296.2	300

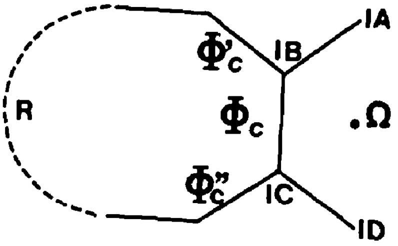


Fig. 16.

Table 6. d algorithms. *Note*: For P120 and P240 constants see Table 5

IOR_A	IOR_D	Φ_C	Δ_A	Δ_D	Φ_0
1	0	+	−	+	Φ_c
1	0	+	+	+	Φ_c + P240
1	0	+	+	−	Φ_c
1	0	+	−	−	Φ_c + P120
1	0	−	+	−	Φ_c
1	0	−	+	+	Φ_c + P240
1	0	−	−	+	Φ_c
1	0	−	−	−	Φ_c + P120
0	1	+	−	+	Φ_c
0	1	+	+	+	Φ_c + P120
0	1	+	+	−	Φ_c
0	1	+	−	−	Φ_c + P240
0	1	−	+	−	Φ_c
0	1	−	+	+	Φ_c + P120
0	1	−	−	+	Φ_c
0	1	−	−	−	Φ_c + P240
1	1	+	−	+	Φ_c + P120
1	1	+	+	+	Φ_c
1	1	+	+	−	Φ_c + P240
1	1	+	−	−	Φ_c
1	1	−	+	−	Φ_c + P240
1	1	−	+	+	Φ_c
1	1	−	−	+	Φ_c + P120
1	1	−	−	−	Φ_c
0	0	+	−	+	Φ_c + P240
0	0	+	+	+	Φ_c
0	0	+	+	−	Φ_c + P120
0	0	+	−	−	Φ_c
0	0	−	+	−	Φ_c + P120
0	0	−	+	+	Φ_c
0	0	−	−	+	Φ_c + P240
0	0	−	−	−	Φ_c

IOR A function of the two parameters: Δ and the stereochemistry. It has the value 0 if the IC–ID bond is equatorial to R and 1 if IC–ID is axial.

The corresponding relations are shown in Table 4 (algorithms b).

(B) *Only the central bond belongs to ring R* (Fig. 16)

(a) If the hybridization of the two atoms of the central bond is sp^2 the torsion angle Φ_0 is given by:

$$\Phi_0 = \Phi_c \text{ (algorithm c).}$$

(b) If the hybridization of the two atoms IB and IC is sp^3, Φ_0 can be calculated as a function of IOR_A, IOR_D, Φ_c, Δ_A, Δ_D and ω with IOR_A and IOR_D the orientations of the two bonds IB–IA and IC–ID (equatorial = 0, axial = 1)

$$\Delta_A = \Phi_c' - \Phi_c, \quad \Delta_D = \Phi_c - \Phi_c'' \quad \text{and} \quad \omega = \overrightarrow{\Omega IB} \cdot \overrightarrow{\Omega IC}.$$

In the Script program the situation is organized in order to always have a positive ω angle and the corresponding algorithms are given in Table 6 (d algorithms).

(c) If one of the two atoms of the central bond is sp^3 and the other sp^2, similar algorithms to those established in Table 4 can be used provided the two constants F120 and F240 are replaced by P300 and P60 respectively, and so long as in the case of IB sp^3 or IC sp^3, the ω rotation angle, which must be used is that involving atoms (IB, IC, ID) or (IC, IB, IA) respectively.