Tetrahedron Vol. 27, pp. 789 to 797. Pergamon Press 1971. Printed in Great Britain

GEMO, A COMPUTER PROGRAM FOR THE CALCULATION OF THE PREFERRED CONFORMATIONS OF ORGANIC MOLECULES

N. C. COHEN

Centre de Recherches Roussel-Uclaf, B.P. 9 Romainville (93) France

(*Received in the UK* 17 *April* 1970; *Accepted for publication* 5 *June* 1970)

Abstract—A computer program is presented which permits the calculation of the geometries and corresponding energies of any molecule which can be submitted to treatment by the strain energy minimization method.

The strain and the molecular geometries of the preferred conformations of two ketosteroids compounds have been computed and related to the stereochemistry of the reduction products by metal hydrides.

OUR INTEREST in the conformational analysis of steroids has led us to write an elaborate computer program for the calculation of the equilibrium conformation and the strain energy which is applicable to a wide range of molecules.

This program, called GEMO, is written in FORTRAN IV for use on an IBM 360/75, Univac 1108 and CDC 6600. In this form it is able to study large molecules (more than 150 atoms); the compound may be linear, cyclic or polycyclic with substituents, and may contain single or multiple bonds, heteroatoms and polar groups.

We use the method of strain-energy minimization which has been widely used in recent years for cycloalkanes,[1] cycloalkenes[2] and small molecules, based on the concepts of Westheimer.[3] In particular, we have drawn our inspiration from the recent work on cycloalkanes of Lifson and Bixon.[1c]

The data for the program are the following:

(a) Definition of the molecule.
(b) Definition of the molecular energy.
(c) Approximate conformation of the molecule.

1. DEFINITION OF THE MOLECULE

A tree structure with nodes (atoms), and links (bonds) is used to define the topology of the molecule (see Fig 1). There are four types of nodes (C, H, X, Y),* and two types of links (single or multiple bonds).

The spatial location of the atoms is defined if we add to this graph, information dealing with geometrical parameters corresponding to internal molecular coordinates; these geometrical parameters are:

The distances between connected nodes (bonds)
The angles between connected links (valency angles)
The dihedral angles defined by three successive connected links (torsional angles).

The graph is topologically equivalent to the chemical formula if the initial tree structure is completed by connecting nodes that are not connected in the tree but

---

* X and Y are heteroatoms.

which correspond to bonded atoms in the molecule (in the case of Fig 1 : 1–6, 5–10, 9–14, 13–17).

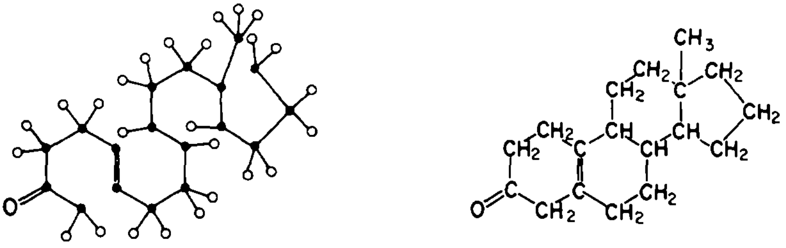


FIG 1. Initial tree structure of 5(10)oestrene 3-one and the numbering of carbon atoms by GEMO

2. THE MOLECULAR ENERGY

The molecular energy is defined by semi-empirical functions and may be written as

$$E = E(1) + E(\theta) + E(\Phi) + E(nb) + E(es)$$

where $E(1)$, $E(\theta)$ and $E(\Phi)$ are respectively the energies associated with bond stretching, bond-angle bending and torsional strain, $E(nb)$ is the total non-bonded interactions, and $E(es)$ is the electrostatic energy. The analytical forms of the energy functions and the numerical values of the different parameters used are given in the Appendix.

3. GENERAL SCHEME OF COMPUTATION

The scheme of the computation is as follows:

1—Spatial location of all the atoms of the initial conformation from the independent internal coordinates* by the classical procedure of the vector method.[4]

2—Calculation of all internal coordinates and the corresponding energies.

3—Calculation of non-bonded distances and energies related to each kind of interaction.

4—If dipoles are involved in interaction, calculation of the electrostatic energy.

5—The total molecular energy ("strain energy") is then computed from the different independent contributions for the given conformation.

The conformation of minimum energy is given by trying out deformations of the initial conformation and repeating the scheme so as to minimize its energy.

Hydrogens

In order to reduce the computing time, we assume that hydrogen atoms are not free. Their positions are constructed from the positions of carbon atoms situated in their vicinity.

---

* The independent coordinates for a ramification of n nodes are the $n - 1$ bond lengths, the $n - 2$ valency angles and the $n - 3$ torsional angles.

Methylenic hydrogens are located as in ref. 1a, symmetrically on both sides of the C—C—C planes, with an angle H—C—H (ω) calculated from the relation

$$\omega - \omega_0 = 0{\cdot}294\,(\theta - \theta_0)\left\{\begin{array}{l}\omega_0 = 106°\text{, and relevant values of } \theta_0\\ \text{are listed in the Appendix.}\end{array}\right.$$

The methine hydrogens are constructed by the same process from the C—C—C angle of the initial tree structure. Ethylenic hydrogens are placed as methylenics but with a value $\omega = 0$.

In all these constructions the C—H bond is considered as constant with a value fixed at 1·108 Å. Methyl hydrogens were allowed to move in their three degrees of freedom: the C—H bond, the C—C—H angle and the C—C—C—H torsional angle, with the assumed simplification that the three hydrogens conform to three-fold symmetry about the C—C direction.

4. MINIMIZATION

Our minimization process, which does not need the calculation of derivatives, is rapidly convergent[5] and may study functions of 200 independent variables. To each independent internal coordinate x_i is associated a small increment Δx_i (at the beginning of the minimization the increments are generally $\Delta x = 10^{-3}$ Å for all bond lengths and $\Delta x = 0{\cdot}5°$ for all valency and torsional angles). Indicators ε_i, related to each coordinate to be optimized, are used in the iterative process for the evaluation of adequate new increments.

A value $\varepsilon_i = -1$ indicates that if the coordinate x_i is changed by the amount Δx_i, it results in an energy-lowering; on the contrary, a value $\varepsilon_i = +1$ indicates that such an operation leads to an increasing of the energy; the value $\varepsilon_i = 0$ indicates that no variation occurs.*

The new direction of search depends on the related values of ε_i and is calculated as part of the iteration cycles. The minimization continues until the energy improvement for a given number N of successive entire cycles (e.g. $N = 4$) is less than a prespecified amount (e.g. 0·01 kcal).

Although the description of the molecule and consequently the motions of carbon atoms is symmetry-independent, symmetrical minima result for molecules having symmetry elements.

The method does not avoid the problem common to all search procedures of the false minima in multidimensional space. The more one tries out different initial conformations, the more the assurance of having the true minimum is great.

Experience shows that the total time of computation for a given case is not so drastically dependent on the number of independent parameters as in methods using the gradient vector.

The total time required to reach the minimum for a molecule such as one of the steroids described below is about five minutes on an IBM 360/75.

The output of the program consists of a complete description of the preferred molecular geometry with tables giving the final atomic coordinates, the bond lengths,

---

* This may arise at the end of the minimization where increments have diminished almost to zero, but this also happens if one or more particular increments have been fixed to zero for all the computations. Such an operation is useful for the study of the capabilities of a molecule to adopt conformations different to that of its ground state.

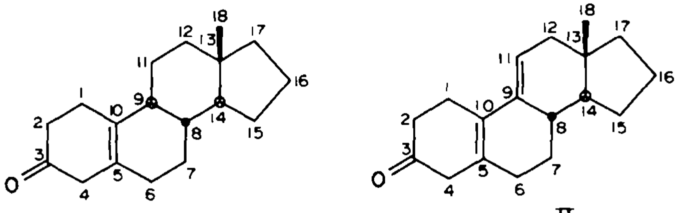
valency and torsional angles, non-bonded distances and the energies associated with each kind of strain.

A perspective view of the molecule may be obtained using another computer program[6] which draws a coloured (one colour for each kind of atom) perspective view; hidden lines are eliminated by this program.

5. RESULTS AND DISCUSSION

The examples that have been chosen to illustrate the capabilities of our computer program are related to the problem of the conformational preference of ring A in 5–10 unsaturated keto steroids such as

5(10) oestrene 3-one *I*
5(10)-9(11) oestradiene 3-one *II*

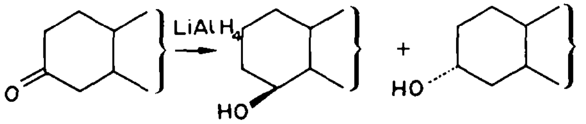


I

II

Fig 2. 5(10) oestrene 3-one *I* and 5(10)-9(11) oestradiene 3-one *II* and the numbering of atoms according to the steroid convention

The conformations of these molecules have been investigated to ascertain those features which can be of significance in the steric course of the reduction of these two ketones by hydrides. This reaction yields different mixtures of the two possible epimeric alcohols *III* and *IV*.



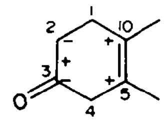
III

IV

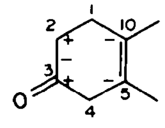
Fig 3. The two epimeric alcohols formed in the reduction of the ketones

Experiment shows that the proportion of the alcohols *III* and *IV* is in the ratio 1:4 in the reduction of *I* (with 17β O—CO—C_2H_5)[7] and about 1:1 in the reduction of *II* (with 17β—OCO—C_6H_5).[8] The understanding of the enhanced stereospecificity in the reduction of *I* relative to its Δ-9,11 homologue *II* may be elucidated by the study of the transition states.

One may attempt the quantitative calculation of such transition states, but in order to avoid the too numerous hypotheses that have to be assumed, we have chosen to consider not the transition states but the compounds involved in the



V



VI

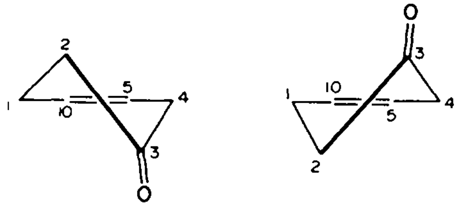
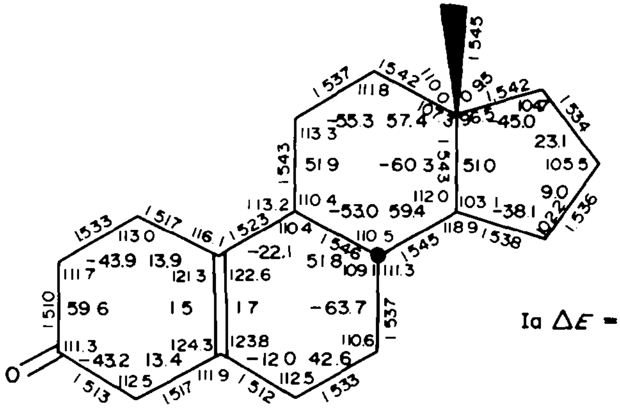
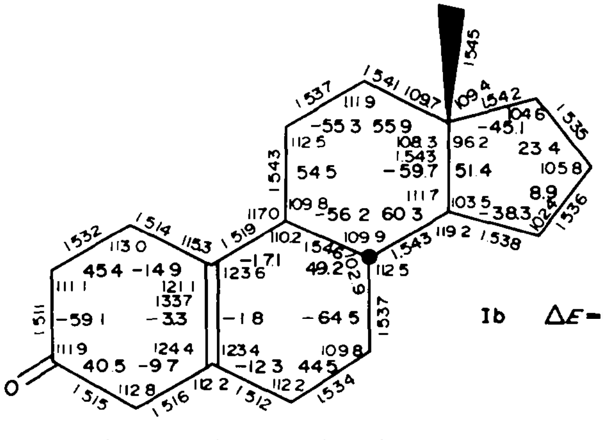


Fig 4. The two alternative half-chair conformation of ring A and signs of dihedral angles (convention of Klyne[10])



Ia ΔE = 0·9 kcal/mole

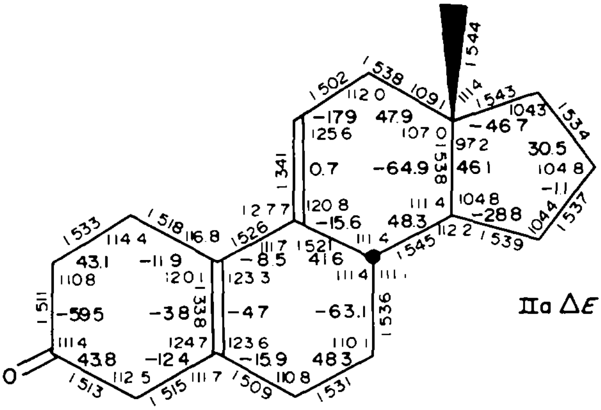


Ib ΔE = 0

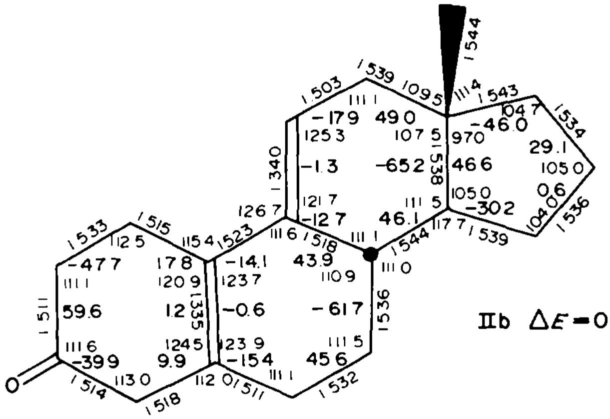
Fig 5. Computed conformations of 5(10) oestrene 3-one *I*.

reaction. More simply, it is possible to consider only the reactants as actually proposed in the literature:[9] in the case of the reduction of unhindered ketones by hydrides, the reactant-like transition state is the most probable.

The calculations were carried out for both enones *I* and *II* using in each case the two possible enantiomeric half-chair conformations *V* and *VI* of ring A. The computed optimal conformations of the molecules can be conveniently represented by indicating in the chemical formula the values of the final internal rotation angles together with bond lengths and bond angles (Figs 5 and 6).



IIa ΔE = 0·3 kcal/mole



IIb ΔE = 0

FIG 6. Computed conformations of 5(10):9(11) oestradiene 3-one *II*

The conformational transmission due to the modification of ring C from *I* to *II* may be seen from these figures. In particular the modifications of ring D geometries have to be compared with X-ray structure determinations by the elegant quantitative description of Altona Geise and Romers.[13] Phase angles of pseudorotation in ring D of *Ia*, *Ib*, *IIa* and *IIb* are respectively 15·2°, 15·3°, 38·5°, and 34·3°. The shapes of ring A in the preferred conformers of *I* and *II* are different (opposite signs for the dihedral angles of ring A in *Ib* and *IIb*), and in both cases the conformation is a half-chair type.

Our calculations show that the preferred conformation of olefin *I* is *Ib* which is more stable than *Ia* by 0·9 kcal/mole, whereas the difference between *IIb* and *IIa* is only 0·3 kcal/mole in favour of *IIb*.

The origin of these two different stabilities is due solely to unequal non-bonded interactions between the two possible conformers a and b.* The energy of interaction between non-bonded atoms is distributed among about 150 pairs of atoms for each conformer, but in the final resulting conformers of *I* and *II*, the most important ones were 1,11 and 4,6 non-bonded interactions with relative contributions in the ratio about 100:36 in both compounds. However, it is interesting to note that in *I* both the 1,11 and 4,6 act the same way to destabilise the *I*a conformer, whereas, in *II*, they are opposed. This is the principal reason for the smaller difference of energy between *II*a and *II*b relative to that between *I*a and *I*b, which suggests that ring A in *II* has the greater conformational mobility.

The observed different mixtures of the two epimeric alcohols *IV* and *V* may therefore be explained, with the hypothesis of a reactant-like transition state, in the light of our results. Thus, by the axial attack of ion hydride on the preferred conformer *I*b, the reduction of *I* will give preferential formation of alcohol *IV*, but, since the stabilities of the two conformers *II*a and *II*b are almost the same, there will be no stereospecificity in the course of the reduction of *II*.

The general shape of the four conformers Ia, Ib, IIa and IIb is represented in Figs 7–10 as a set of stereodiagrams which have been drawn with a computer program.[6] The relevant data are the position of the view point, the location and the name of each atom and the colour to be used for the drawing of each kind of atom. The time needed for one molecule is about 20 seconds of computing time in a CDC 6600 and seven minutes of drawing on a Calcomp 763 plotter.

The final Cartesian coordinates of the different conformers can be obtained from the author.

7. CONCLUSION

As shown in our paper, the optimization treatment of complex organic molecules is now available, and should make possible new developments of the concepts pioneered by Westheimer which will certainly be useful in an empirical approach to the study of reaction paths in chemistry.

Acknowledgement—The author wishes to record his warm appreciation to Dr. R. Bucourt for stimulating discussions and to thank the referee for his helpful suggestions.

APPENDIX I

The following energy functions are used in the computer program.

Stretching $E(l) = K_1(l - l_0)^2$

Bending $E(\theta) = K_\theta(\theta - \theta_0)^2 + \psi(\theta - \theta_0)$ (12)

---

* This was demonstrated by a calculation which was performed in the same manner, but without taking into account the Van der Waals forces between non-bonded atoms. In these conditions, there is no difference of energy between the resulting conformers a and b.

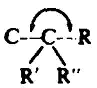
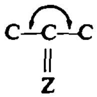
$$\textit{Torsional}\begin{cases}\text{single bond } E(\Phi) = q[1 + \cos n(\Phi + s)]\\ \text{double bond } E(\Phi) = a(|\Phi| - t)^2\end{cases}$$

$$\textit{Non-bonded} \qquad E(nb) = Ae^{-\mu r} - \frac{B}{r^6}$$

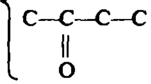
$$\textit{Electrostatic} \qquad E(es) = \frac{1}{D|R^3|}\left[\vec{\mu}_1 \cdot \vec{\mu}_2 - 3\frac{(\vec{\mu}_1 . \vec{R})(\vec{\mu}_2 . \vec{R})}{|R|^2}\right] \quad (11)$$

The numerical values of the different parameters used in the calculations are:*

Stretching	l_0(A)	K_l (Kcal mole^{-1} Å^{-2})
C(sp^3)—C(sp^3)	1·533	300
C(sp^2)—C(sp^3)	1·505	300
C(sp^2)═C(sp^2)	1·331	640

Bending				θ_0(deg)	K_θ(Kcal mole^{-1} deg^{-2})
	R = C	R′ = H	R″ = H	112·4	0·01950*
	R = C	R′ = H	R″ ≠ H	111·0	0·01751
	R = C	R′ ≠ H	R″ ≠ H	109·5	0·01751
	R = H	R′ = H	R″ = H	110·1	0·03828*
	R = H	R′ ≠ H	R″ ≠ H	107·1	0·01206
	Z = C			112·0	0·01887
	Z = O			116·0	0·02830*

* global constant for the bending of all the valency angles around the central atom.

Torsional		q(Kcal mole^{-1})
σ bonds	C—C—C—C	1·40
	C—C—C═C	0·99
	C═C—C═Z	3·50
	 $C{-}C({=}O){-}C{-}C$	0·40
π bonds	C—C═C—C	a = 26·2 Kcal mole^{-1} rad^{-2}

Non-bonded interactions (1d)

	A	B	μ
H...H	2300	49·2	3·6
C...H	4012	125	3·4
C...C	7000	325	3·2

APPENDIX II

After this paper was submitted for publication, a recent work of Levine and Eudy was published[15] which shows that the reduction of 17β hydroxy estra 5(10), 9(11) diene-3-one acetate by lithium tri-t-butoxy aluminium hydride occurs with only a low degree of stereospecificity. Using Dreiding models and Hendrickson's H—H interaction potential (1d), the authors have estimated a difference of energy between *II*a and *II*b of 0·26 kcal/mole in favour of *II*b which is very close to the result of our computations. But the authors have considered this result as erroneous because this would predict rather more of alcohol 3β than of alcohol 3α whereas experimentally, the alcohols 3β and 3α are produced in ca 35:65 ratio. We consider that the small preference for conformer *II*b is real and suggest the following explanation for the

---

* The computed molecular structure of androsterone obtained with these values agrees well with that of its X-ray determination. The activation energy of the inversion process of cyclohexene together with the different conformers involved in this process, have been computed by this force field and also by molecular orbital calculations using the CNDO/2 approximation. GEMO's computations are in good agreement with the available experimental data, and the results obtained by these two different methods will be discussed elsewhere.[14]

apparent discrepancy between the predictions and the experimental results: in the reduction of unhindered ketones, when the conformational preference between the conformers is small, the relative stability between the transition states does not necessarily correspond to that of the ketones. Thus, if the conformational analysis of the reactants reveals small differences of energies between them, one may only predict that the reaction will occur with a low degree of stereospecificity. The prediction of what compound will be present in a larger amount is possible by the knowledge of the differences of energies between the related transition states themselves.

REFERENCES

[1] [a] J. B. Hendrickson, *J. Amer. Chem. Soc.* **83**, 4537 (1961)
[b] K. B. Wiberg, *J. Amer. Chem. Soc.* **87**, 1070 (1965)
[c] M. Bixon and S. Lifson, *Tetrahedron* **23**, 769 (1967)
[d] J. B. Hendrickson, *J. Amer. Chem. Soc.* **89**, 7036 (1967)
[e] R. Bucourt and D. Hainaut, *Bull. Soc. Chim. France* 4562 (1967)
[2] [a] R. Bucourt and D. Hainaut, *C.R. Acad. Sci.* **258**, 3305 (1964)
[b] N. L. Allinger, J. A. Hirsch, M. A. Miller and I. J. Tyminski, *J. Amer. Chem. Soc.* **90**, 5773 (1968)
[c] G. Favini, G. Buemi and M. Raimondi, *J. Mol. Struct.* **2**, 137 (1968)
[d] G. Favini, F. Zuccarello and G. Buemi, *J. Mol. Struct.* **3**, 385 (1969)
[3] A recent review of quantitative conformational analysis is given by J. E. Williams, P. J. Stang and P. von R. Schleyer, *Ann. Rev. Phys. Chem.* **19**, 531 (1968)
[4] H. B. Thompson, *J. Chem. Phys.* **47**, 3407 (1967)
[5] Subroutine SCOOP supplied by the IBM Company
[6] N. C. Cohen, unpublished work.
[7] S. G. Levine, N. H. Eudy and C. F. Leffler, *J. Org. Chem.* **31**, 3995 (1966)
[8] L. Nedelec, *Bull. Soc. Chim. France*, in press
[9] M. Chérest and H. Felkin, *Tetrahedron Letters* 2205 (1968)
[10] W. Klyne and V. Prelog, *Experientia* **16**, 521 (1960)
[11] J. M. Lehn and G. Ourisson, *Bull. Soc. Chim. France*, 1119 (1963)
[12] N. L. Allinger, J. A. Hirsch, M. A. Miller, I. J. Tyminski and F. A. Van-Catledge, *J. Amer. Chem. Soc.* **90**, 1199 (1968)
[13] C. Altona, H. J. Geise and C. Romers, *Tetrahedron* **24**, 13 (1968)
[14] R. Bucourt and N. C. Cohen, unpublished results.
[15] S. G. Levine and N. H. Eudy, *J. Org. Chem.* **35**, 549 (1970)