



CALCULATIONS ON INTRAMOLECULAR HYDROGEN BONDS BY THE MOLECULAR FORCE FIELD CNDO/2 METHOD

The molecular force field method of molecular geometry optimization is presented in a version which uses cartesian forces, together with the CNDO/2 method for the calculation of wavefunctions. Some preliminar results are presented for ethylene and for the enol tautomer of malonaldehyde.

1 — INTRODUCTION

The hydrogen bond, considered as an interaction between an hydrogen atom of low electronic density and a region of high electronic density, has for a long time raised the interest of the theoretical chemists, both because of its bonding characteristics (in principle, an hydrogen atom cannot participate in more than a single covalent bond) and because of the fundamental role played by the hydrogen bonds in the chemistry of many biological molecules.

However, many molecular systems of practical interest are too complex to be treated by conventional *ab initio* techniques; therefore, many of the calculations of hydrogen bonded systems already reported in the literature were made for very simple systems, or else, they were based on very drastic approximations.

The recent development of more powerful semi-empirical methods of calculation (CNDO[1], INDO[2], MINDO[3], etc.) has made possible the study of molecular systems complex enough to be of real interest [4-8].

Another problem frequently raised is the determination of equilibrium molecular geometries. Equilibrium geometries will occur at minima of the potential energy hypersurface, and this is the basis of all the theoretical methods used for the calculation of molecular geometries.

The choice of the method of geometry optimization is closely linked to the characteristics of the method used to evaluate the molecular wavefunctions, and to the degree of complexity of the molecular systems to be studied.

As the aim of the authors is to study systems of intermediate complexity, such as malonaldehyde and some β -diketones, the CNDO/2 method of Pople, Santry and Segal was chosen for the calculation of the wavefunctions and a conveniently adapted version of the molecular force field method, first reported by PULAY [9-12], was chosen for the geometry optimization; this method allows the analytical localization of the minima of the potential energy hypersurface and this becomes rather easy if the ZDO approximation is used, as it is the case for the CNDO/2 method.

In this paper, the basic formulation of the molecular force field method adapted to cartesian forces is presented, and some preliminary results obtained

for ethylene and for the enol tautomer of malonaldehyde are reported.

2 — THE METHOD OF THE MOLECULAR FORCE FIELD

The *ab initio* methods of Molecular Quantum Mechanics have been used quite successfully for the calculation of equilibrium molecular geometries, but their usefulness is limited to rather small molecules because of their complexity.

For larger molecules the semi-empirical methods including all the valence electrons are, for the time being, the only accessible means of calculating the wavefunctions. However, even with these approximated methods, many of the current techniques of molecular geometry optimization are not efficient enough to deal with complex molecules.

Most of these techniques are based on a numerical search of the potential energy hypersurface and this requires an SCF calculation at each of the configuration space points needed for a satisfactory definition of the hypersurface. The number of points will increase near the potential energy minimum, and even more so, if the calculation of molecular properties (molecular geometries, force constants, etc.) is intended. It is obvious that for rather complex systems (i.e., systems with many internal degrees of freedom) the number of SCF calculations becomes forbidding and the numerical methods are impracticable.

It is also convenient to examine how accurate these approximations are. All the methods for the optimization of the molecular geometries are based on the knowledge of the potential energy first derivatives relative to the nuclear coordinates. If a numerical technique is used, then considerable errors can be introduced when the first derivatives are calculated and, even more considerable errors will be made if the obtained equilibrium geometry is used for the calculation of other molecular properties. For instance, the numerical calculation of force constants may be impracticable because of the errors entailed by the evaluation of the second derivatives. The analytical methods of optimizing the molecular geometries avoid these problems because the first derivatives are evaluated analytically, only the force constants being calculated by numerical methods. While in the numerical methods the potential energy is directly minimized, the analytical methods use an

indirect method where a configuration is searched for such that all the partial derivatives of the potential energy will vanish. At least in principle, the analytical methods are more exact than the numerical methods, but they have the obvious disadvantage of requiring the evaluation of a fairly large number of additional two electron integrals. Therefore, the analytical methods are not used in *ab initio* calculations, because it takes about four times longer to compute the potential energy gradient than to make an SCF calculation. However, these methods prove to be very useful at the level of the semi-empirical ZDO methods, as most of the additional integrals are neglected in the ZDO approximation.

The molecular force field method is based on the analytical calculation of the first partial derivatives (negative forces) of the potential energy relative to the molecular geometry parameters. The method is more difficult to program than a numerical method but this is more than compensated by the lower number of necessary SCF calculations and because the errors in the results are smaller. The method was initially developed by PULAY [9] for SCF calculations and, later on, was used at the level of the semi-empirical methods with all the valence electrons [10-13].

In this paper a modification of the original molecular force field is developed, where all the calculations are made in terms of nuclear cartesian coordinates, far easier to manipulate than the internal coordinates or the symmetry coordinates.

3 — FORMULATION OF THE METHOD

At the level of a ROOTHAAAN-HARTREE-FOCK (RHF) [14] approximation the molecular potential energy presents a double functional dependence upon the molecular parameters: an explicit dependence upon the nuclear coordinates and an implicit dependence through the density matrix, D :

$$\langle \Psi | \hat{H} | \Psi \rangle = \epsilon = \epsilon(\mathbf{R}, D(\mathbf{R})) \quad (1)$$

where $\mathbf{R} = [\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N]$ is a matrix with elements \mathbf{R}_j which are three-component column vectors — the three cartesian coordinates of nucleus j . Then, the total derivative of the potential energy relative to the nuclear coordinates has the form:

$$\frac{d\varepsilon}{d\mathbf{R}} = \frac{\partial\varepsilon}{\partial\mathbf{R}} + \frac{\partial\varepsilon}{\partial\mathbf{D}} \cdot \frac{\partial\mathbf{D}}{\partial\mathbf{R}} \quad (2)$$

Within an RHF approximation the second term of (2) is zero, because the variational method makes the energy, ε , stationary relative to the coefficients, \mathbf{C} , of the atomic orbitals (AO) on the molecular orbitals (MO) and the same is true for the density matrix:

$$\mathbf{D} = \mathbf{C} \cdot \mathbf{C}^+$$

The forces acting on the atoms within a configuration \mathbf{R} , are evaluated as:

$$\mathbf{F} = - \frac{\partial\varepsilon}{\partial\mathbf{R}} = - \langle \Psi | \frac{\partial\hat{H}}{\partial\mathbf{R}} | \Psi \rangle - 2 \langle \frac{\partial\Psi}{\partial\mathbf{R}} | \hat{H} | \Psi \rangle \quad (3)$$

where \mathbf{F} (with a structure identical to matrix \mathbf{R}) is a matrix whose element F_j is the cartesian force acting on atom j of the molecular system.

The first term on the right hand side of (3) is the Hellmann-Feynman force. This force equals exactly $-\frac{\partial\varepsilon}{\partial\mathbf{R}}$ (exact force) when Ψ is an exact wave function or when Ψ is within the Hartree-Fock limit (Hellmann-Feynman theorem). If Ψ is an approximated wave function, the Hellmann-Feynman theorem is no longer valid, as the second term on the right hand side of (3) ceases to vanish and the Hellmann-Feynman force will differ from the exact force.

This means that when the molecular force field is used with a semi-empirical method all the partial derivatives in (3) must be evaluated, thus losing some of the simplicity of the Hellmann-Feynman method, because the Hellmann-Feynman forces, being the expectation value of a monoelectronic operator, are rather easy to compute.

However, it can be shown that even for wave functions close to the Hartree-Fock limit, the Hellmann-Feynman force is a poorer approximation to the exact force than the negative derivative of the potential energy. Indeed, for a wave function with an error of order ε [15] it can be shown that the Hellmann-Feynman force is also affected by an error of order ε ; but the energy associated with this wave function is only affected by an error of order

ε^2 , and the same applies to the first derivative of the potential energy. For these reasons the molecular force method does not use the Hellmann-Feynman forces.

The basic principles behind the molecular force field method are rather trivial: starting with an assumed geometry (necessarily not very different from the optimum geometry), and for each SCF calculation, the method produces a sequence of geometries, each one closer to the optimum geometry. It can be said that in the molecular force field method each SCF calculation has the same value as 3N-6 SCF calculations (N is the number of atoms in the molecule) in a numerical method of molecular geometry optimization.

The basic idea of the method is that for the equilibrium molecular configuration all the molecular forces will vanish; thus, instead of minimizing the energy, the method searches the configuration where the molecular force field will vanish.

Let \mathbf{R}_{00} be the initially assumed geometry. An SCF calculation performed with this geometry will yield a density matrix, $\mathbf{D}(\mathbf{R}_{00})$, and a potential energy, ε_0 , and the cartesian forces acting on the atoms will then be evaluated for this configuration \mathbf{R}_{00} :

$$\mathbf{F}(\mathbf{R}_{00}) = - \left. \frac{\partial\varepsilon(\mathbf{R}, \mathbf{D}(\mathbf{R}_{00}))}{\partial\mathbf{R}} \right|_{\mathbf{R}=\mathbf{R}_{00}} \quad (4)$$

A new configuration $\mathbf{R}_{0(1)}$ is then searched for, such that

$$\mathbf{F}(\mathbf{R}_{0(1)}) \simeq - \left. \frac{\partial\varepsilon(\mathbf{R}, \mathbf{D}(\mathbf{R}_{00}))}{\partial\mathbf{R}} \right|_{\mathbf{R}=\mathbf{R}_{0(1)}} = 0 \quad (5)$$

and this means that all the atoms are then located at positions where all the cartesian forces will vanish independently.

The search for the positions where all the cartesian forces will vanish is made independently for each atom and for each coordinate, while all the other coordinates remain at the values they have in \mathbf{R}_{00} . This means that when searching for the x coordinate of atom j where the x component of the cartesian force at this atom will vanish, all the derivatives in (5) but that on the x coordinate of atom j will be evaluated for the values that the coordinates have in

\mathbf{R}_{00} . In this way, for each of the 3N-6 unidimensional searches, $\mathbf{R}_{0(i)}$ equals the matrix \mathbf{R}_{00} with the unique change of the coordinate where the search is performed. In the present program all the unidimensional searches are made by means of the *regula falsi* method.

In order to simplify the notation, $\mathbf{R}_{0(i)}$ will represent from now on the matrix whose elements are already the results of those 3N-6 independent searches. From the new configuration $\mathbf{R}_{0(i)}$ the displacements produced on the atoms by the forces are evaluated:

$$\mathbf{X}_{00} = \mathbf{R}_{0(i)} - \mathbf{R}_{00}$$

as well as a new geometry:

$$\mathbf{R}_{01} = \mathbf{R}_{00} + \lambda \mathbf{X}_{00}$$

where λ ($0 < \lambda < 1$) is a damping factor. New forces (4) are then evaluated by substituting \mathbf{R}_{01} for \mathbf{R}_{00} and a new sequence of geometry optimization is started based once more upon the density matrix $\mathbf{D}(\mathbf{R}_{00})$ calculated for the initial molecular geometry \mathbf{R}_{00} . These geometry optimization sequences are repeated up to the k th sequence where

$$|(\mathbf{R}_{0k} - \mathbf{R}_{0k-1})_{IJ}| \ll d \quad (I = 1, 2, \dots, N; J = 1, 2, 3)$$

d being a previously set numerical value (in this program $d = 0.01 \text{ \AA}$). Once this convergence has been attained a new SCF cycle is performed but starting from an initial configuration $\mathbf{R}_{10} \equiv \mathbf{R}_{0k}$ and this leads to a new density matrix $\mathbf{D}(\mathbf{R}_{10})$ and to a new potential energy ϵ_1 ($\epsilon_1 < \epsilon_0$). This procedure of geometry optimization sequences followed by an SCF cycle is repeated up to the SCF cycle, $l+1$, where the potential energy has converged:

$$\epsilon_l - \epsilon_{l+1} < \delta$$

δ being a previously set small positive value (in this program $\delta = 0.001 \text{ eV}$). After this double convergence has been attained, the molecular geometry \mathbf{R}_{lk} obtained on the last sequence of the last but one SCF cycle is considered as the best molecular geometry within the CNDO approximation and within the molecular force field method.

The effort of setting up a program at the level of the CNDO/2 method is not too discouraging because

most of the necessary additional integrals are neglected in the ZDO approximation. Thus, it is only necessary to evaluate the derivatives of the overlap integrals and of the electronic repulsion integrals (which are simple functions of the internuclear distances) and the derivatives of the internuclear distances.

When using the program it was found that the complete optimization of a molecular geometry still takes a rather high number of SCF cycles, but it is hoped that this number of cycles will be reduced by using more efficient techniques of interpolation and better «damping» factors.

4 — RESULTS OBTAINED

In this work the molecular force field method was used together with the CNDO/2 method of SCF calculations; the choice of the CNDO/2 method was made for two reasons: firstly because CNDO/2 calculations have already produced good results for hydrogen bond problems [4-8] and secondly because this method produces good results for molecular geometries.

In order to test the method of calculation, as well as the program, some preliminary calculations were done for simple systems with well known molecular geometries; the results of the calculations for ethylene are reported. Later on, a study of the malonaldehyde was done.

ETHYLENE (D_{2h})

The initial geometry (fig. 1, cycle I) was deliberately chosen to be very different from the experimental geometry ($R_{CC} = 1.337 \text{ \AA}$, $R_{CH} = 1.086 \text{ \AA}$, $\angle \text{H-C-H} = 117^\circ$).

The results obtained (figs. 1 to 4) lead to the conclusion that the geometry obtained after each cycle will converge quickly to a geometry close to the experimental one as it could be hoped from a method with the characteristics of CNDO/2. It should be noted that although the C-H distance suffers a steep increase on the first cycles, its difference to the experimental values will tend to vanish on the following cycles. On the other hand, an analysis of the variations of the potential energy (fig. 2a)) and of the various force components (figs. 3,4) (the notation is that of fig. 2(b)) through some

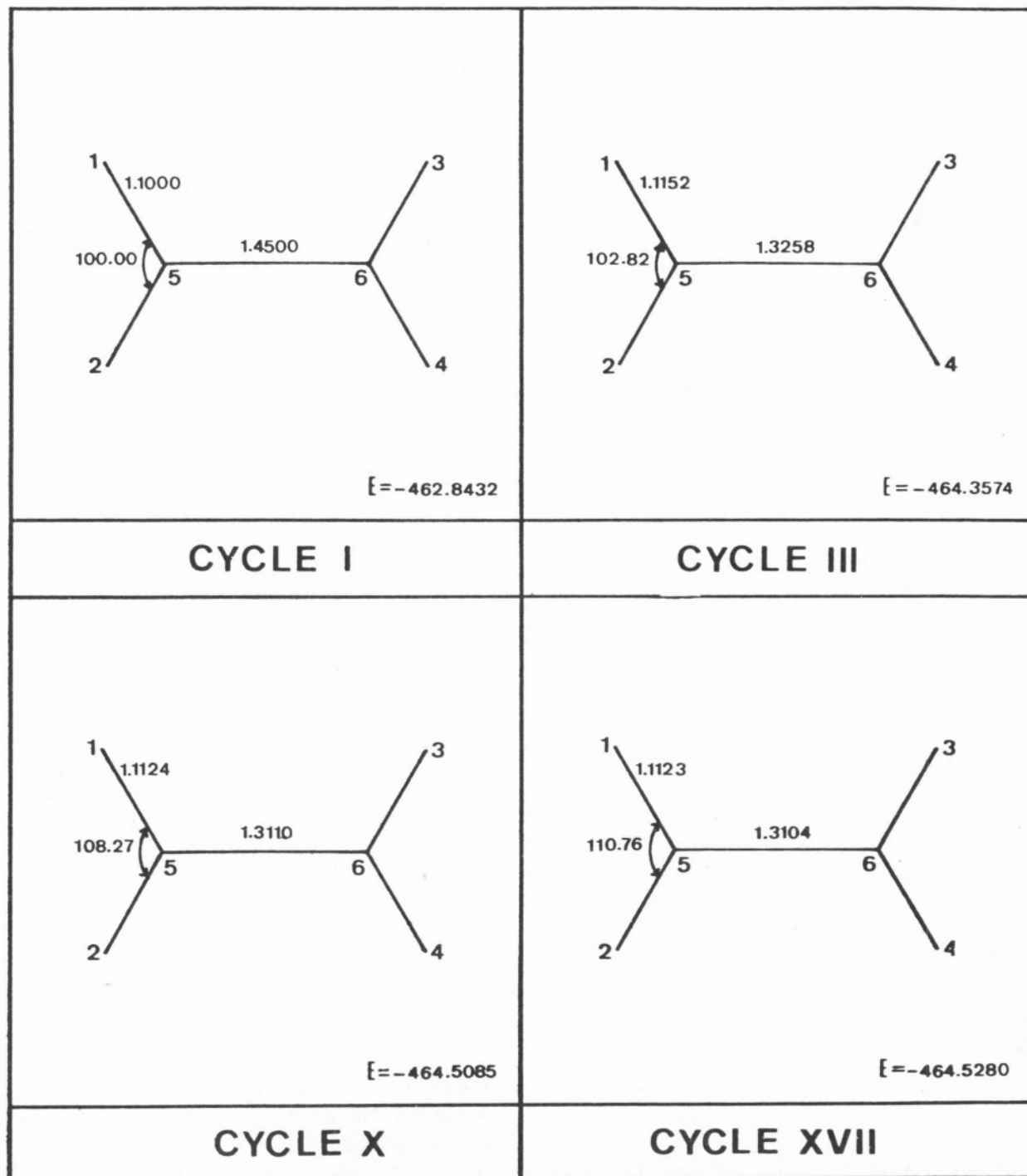


Fig. 1

MALONALDEHYDE

geometry optimization cycles suggests that any attempt to improve the iterative process will not produce much better results and the iteration was stopped. Initially f_3 was the cartesian force with the biggest modulus ($|f_3| = 16.45 \text{ eV/\AA}$) and in the end it was f_2 ($|f_2| = 0.071 \text{ eV/\AA}$); the energy stabilization was about 38.85 kcal/mol.

In order to try to clarify the problems [4,16-22] raised by the molecular geometry of the enol form of malonaldehyde and by the stereochemistry of its intramolecular hydrogen bond, calculations were made for three structural models of that enolic

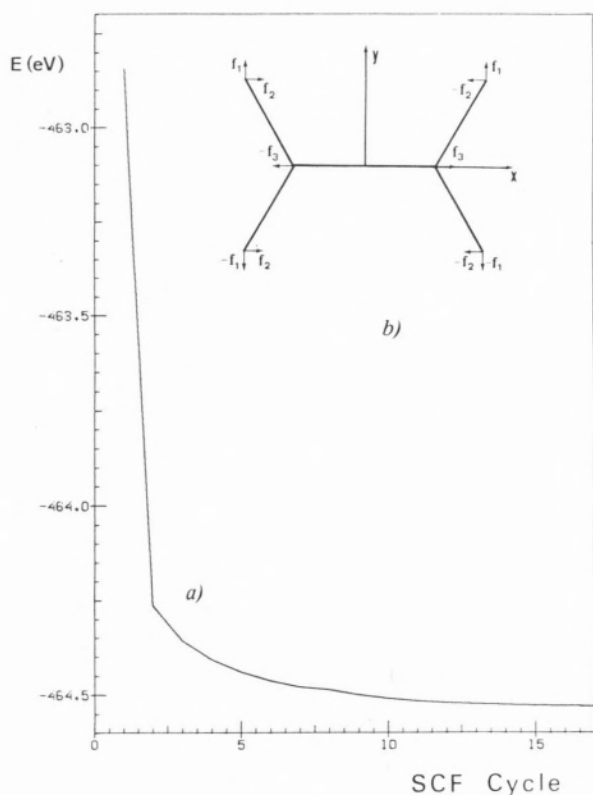


Fig. 2

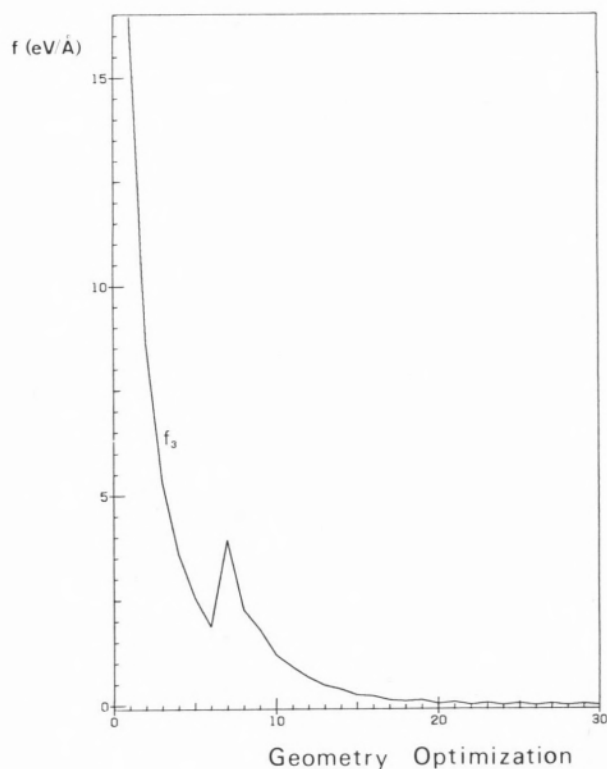


Fig. 4

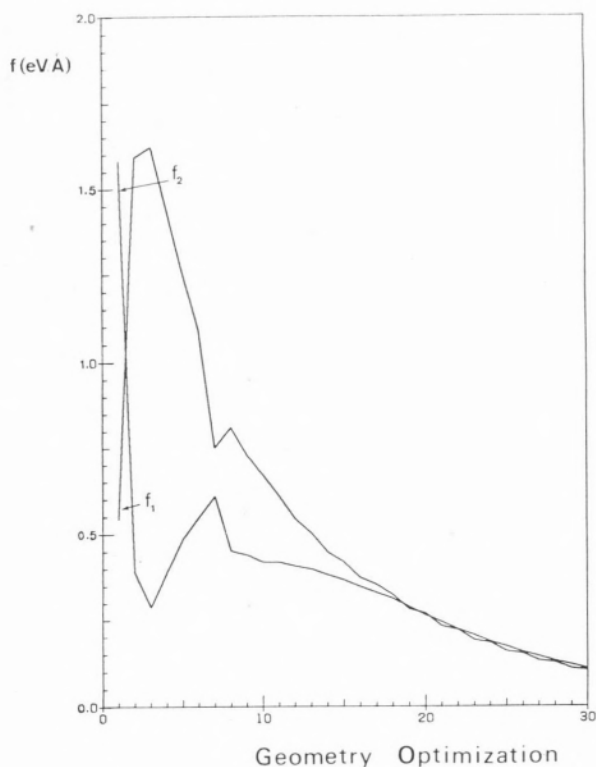


Fig. 3

tautomer: one of the models with C_{2v} symmetry and two other models with symmetry C_s and with the following characteristics:

- (1) Bond lengths C-C and C=C equal
Bond lengths C-O and C=O equal
- (2) Bond lengths C-C and C=C different
Bond lengths C-O and C=O different

The geometries initially assumed for models with symmetries C_{2v} and C_s (2) were those proposed by SCHUSTER [4], while for model C_s (1) the initial geometry was the optimum geometry calculated for model C_{2v} (except, of course, for the hydrogen bonded proton). The results obtained for C_{2v} (figs. 5, 6, 7) will show once more how efficient the molecular force field method is.

For the C_{2v} model the cartesian force with maximum absolute value was initially f_5 ($|f_5| = 5.81 \text{ eV/Å}$) while at the end of the calculation the maximum force was f_7 ($|f_7| = 0.38 \text{ eV/Å}$), the stabilization energy being about 19.71 kcal/mol. For the other models similar convergence characteristics were obtained.

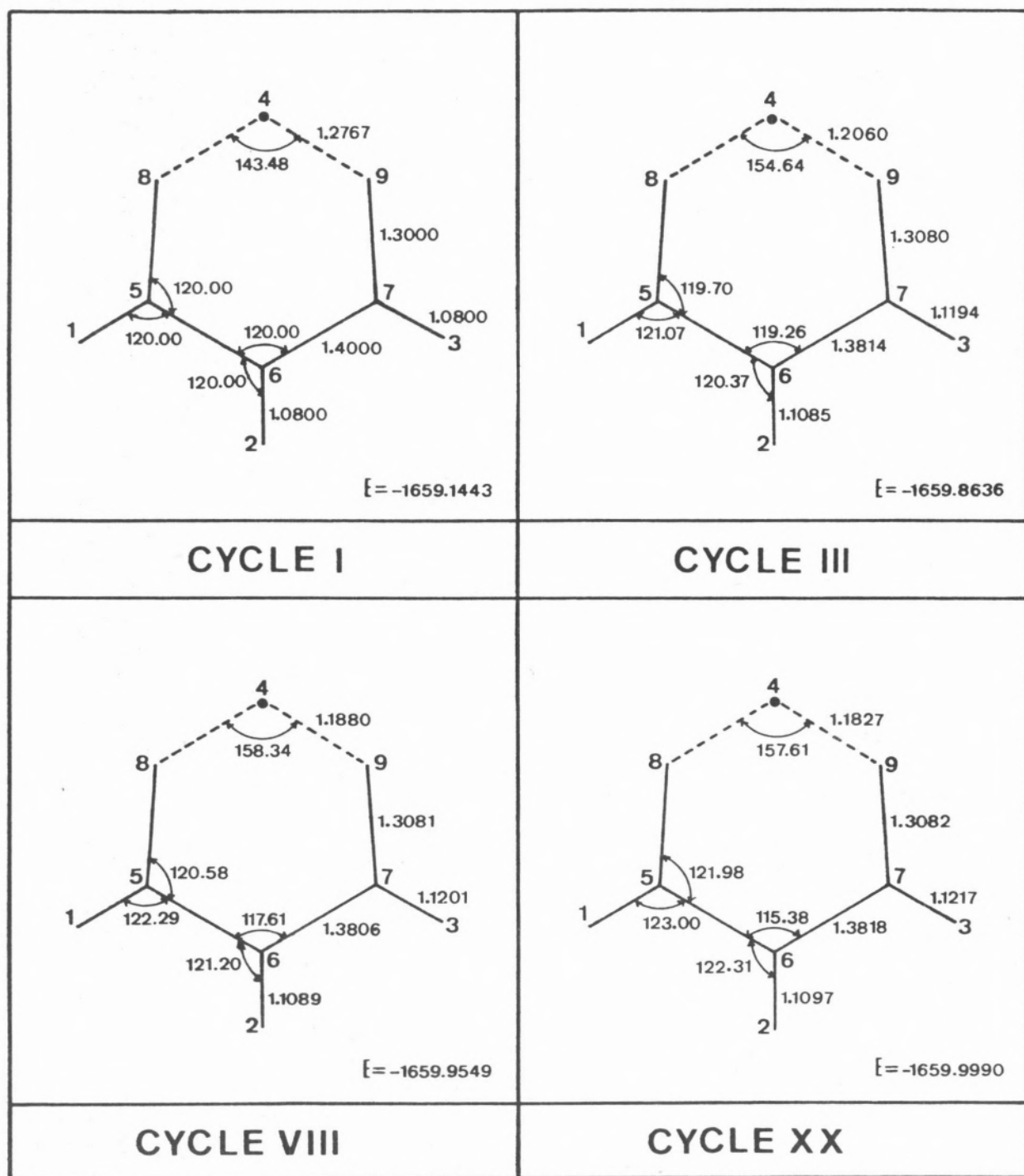


Fig. 5

With models C_{2v} and $C_s(1)$ it was found that the most favoured position for the enolic hydrogen was the symmetric one (just on the C_{2v} symmetry axis) while the most stable structure was also symmetric. However, when more flexibility (more degrees of freedom) was given to the model, by allowing

«single» and «double» bonds a new structure was obtained, more stable than any of the previous ones. The existence of such a structure suggests that the potential energy surface for the hydrogen bonding proton has a double symmetric potential well and this agrees with the conclusions of some

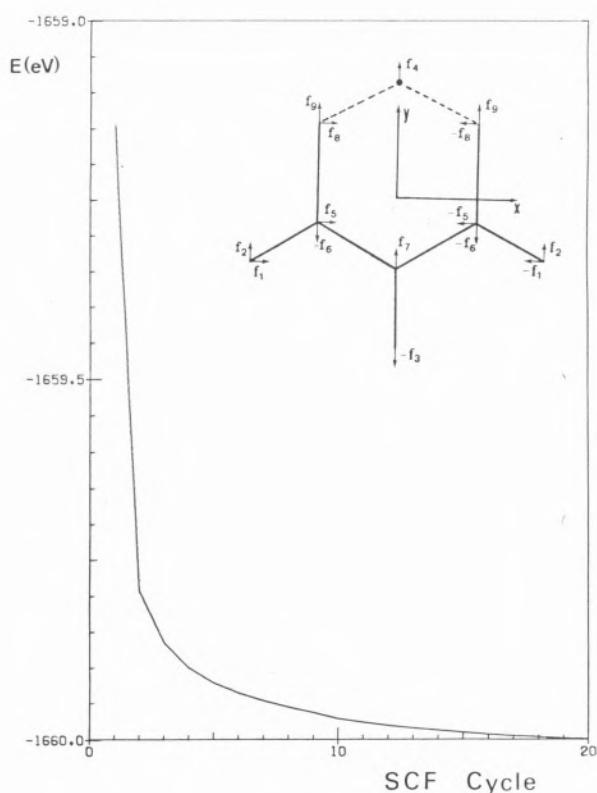


Fig. 6

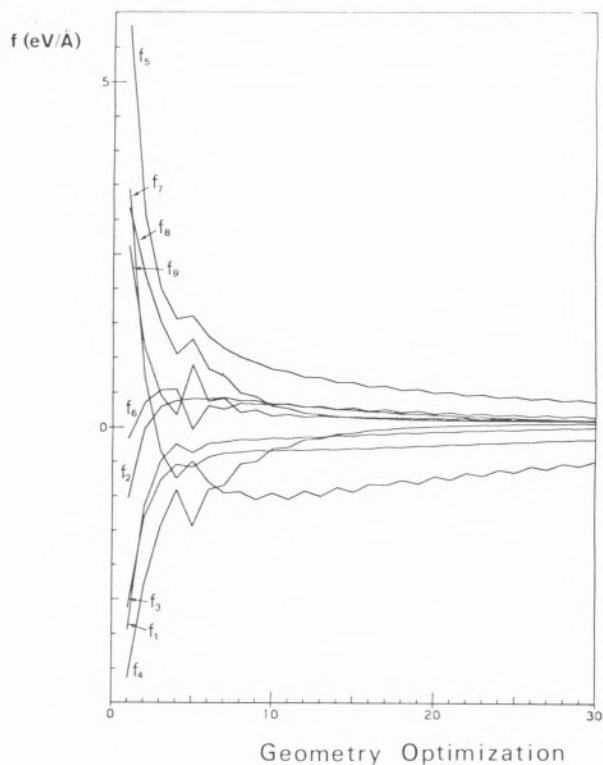


Fig. 7

other researchers. The hydrogen atom is bonded to the oxygen at a distance of 1.08\AA and the H-O-H angle equals 153.73° . The optimum distance between the oxygen atoms is found to be 2.41\AA .

The potential barrier between the two minima of the potential well has a calculated value of 0.57 kcal/mol (in good agreement with the value given by SCHUSTER [4]) and this is low enough for a fast tunneling effect to occur; therefore the transition of the proton can not be detected by the more conventional experimental techniques.

Received 7.March.1980

REFERENCES

- [1] J. A. POPLE, D. P. SANTRY, G. A. SEGAL, *J. Chem. Phys.*, **43**, S129 (1965).
J. A. POPLE, G. A. SEGAL, *J. Chem. Phys.*, **43**, S136 (1965).
J. A. POPLE, G. A. SEGAL, *J. Chem. Phys.*, **44**, 3289 (1966).
D. P. SANTRY, G. A. SEGAL, *J. Chem. Phys.*, **47**, 158 (1967).
- [2] J. A. POPLE, D. L. BEVERIDGE, P. A. DOBOSH, *J. Chem. Phys.*, **47**, 2026 (1967).
- [3] N. C. BAIRD, M. J. S. DEWAR, *J. Chem. Phys.*, **50**, 1262 (1969).
N. C. BAIRD, M. J. S. DEWAR, R. SUSTMANN, *J. Chem. Phys.*, **50**, 1275 (1969).
M. J. S. DEWAR, E. HASELBACH, *J. Am. Chem. Soc.*, **92**, 590 (1970).
N. BODOR, M. J. S. DEWAR, A. HARGET, E. HASELBACH, *J. Am. Chem. Soc.*, **92**, 3854 (1970).
M. J. S. DEWAR, D. H. LO, *J. Am. Chem. Soc.*, **94**, 5296 (1972).
R. C. BINGHAM, M. J. S. DEWAR, D. H. LO, *J. Am. Chem. Soc.*, **97**, 1285 (1975).
- [4] P. SCHUSTER, *Int. J. Quantum Chem.*, **3**, 851 (1969); *Chem. Phys. Lett.*, **3**, 433 (1969); *Mh. Chem.*, **100**, 2084 (1969); *Theoret. Chim. Acta*, **19**, 212 (1969).
- [5] A. OCVRK, A. AZMAN, D. HADZI, *Theoret. Chim. Acta*, **10**, 187 (1968).
- [6] H. MORITA, S. NAGAKURA, *Theoret. Chim. Acta*, **27**, 325 (1972).
- [7] A. PULLMAN, H. BERTHOD, *Theoret. Chim. Acta*, **10**, 461 (1968).
- [8] J. R. HOYLAND, L. B. KIER, *Theoret. Chim. Acta*, **15**, 1 (1969).
- [9] P. PULAY, *Mol. Phys.*, **17**, 197 (1969); *Mol. Phys.*, **18**, 473 (1970); *Mol. Phys.*, **21**, 329 (1971).
- [10] R. F. NALEWAJSKI, *J. Mol. Struct.*, **40**, 247 (1977).
- [11] P. PULAY, F. TOROK, *Mol. Phys.*, **25**, 1153 (1973).
- [12] J. PANCIR, *Theoret. Chim. Acta*, **29**, 21 (1973).
- [13] F. TOROK, P. PULAY, *J. Mol. Struct.*, **46**, 43 (1978).
- [14] C. C. J. ROOTHAAN, *Rev. Mod. Phys.*, **23**, 69 (1951).

- [15] L. SALEM, E. B. WILSON, *J. Chem. Phys.*, **36**, 3421 (1962).
- [16] W. F. ROWE, JR., R. W. DWERST, E. B. WILSON, *J. Am. Chem. Soc.*, **98**, 4021 (1976).
- [17] A. D. ISAACSON, K. MOROKUMA, *J. Am. Chem. Soc.*, **97**, 4453 (1975).
- [18] G. KARLSTROM, B. JONSSON, H. WENNERSTROM, S. FORSEN, B. ROOS, J. ALMLÖF, *J. Am. Chem. Soc.*, **97**, 4153 (1975).
- [19] G. KARLSTROM, H. WENNERSTROM, B. JONSSON, S. FORSEN, J. ALMLÖF, B. ROOS, *J. Am. Chem. Soc.*, **97**, 4188 (1975).
- [20] G. KARLSTROM, B. JONSSON, B. ROOS, H. WENNERSTROM, *J. Am. Chem. Soc.*, **98**, 6851 (1976).
- [21] M. S. GORDON, R. D. KOOB, *J. Am. Chem. Soc.*, **95**, 5863 (1973).
- [22] R. S. BROWN, *J. Am. Chem. Soc.*, **99**, 5497 (1977).

RESUMO

Cálculo de ligações de hidrogénio intramolecular pelo método do campo de força molecular/CNDO/2.

Apresenta-se uma versão do método do campo de forças moleculares para a optimização de geometrias moleculares, com utilização de forças cartesianas e usando o método CNDO/2 para o cálculo das funções de onda. Apresentam-se resultados prévios para o etileno e para a forma enólica do tautómero do aldeído malónico.