MolStruc: A force field calculation program allowing interactive modifications of the force field parameters

D. Siri and G. Pèpe

Centre de Recherche sur les Mécanismes de la Croissance Cristalline (Laboratoire propre du CNRS associé aux Universités d'Aix-Marseille II et III), Campus de Luminy, Marseille, France

J.M. Bernassau

SANOFI Recherche, Montpellier, France

To analyze the influence of parameters and functions on the energy and geometry obtained through different force field calculations, we have developed program MolStruc. This software allows the user to choose between two sets of functions and parameters, MM2 and AMBER.

The MM2 option of the program was developed to compute the coulombic energy in a dipole or monopole approximation. To establish comparisons between the energy values, the coulombic contribution is computed in the same way in the Amber and MM2 options of the program.

The force field parameters can be handled interactively (through addition or modification).

The program was used to study molecules of a representative sample displaying most of the problems encountered in molecular mechanics (MM).

Keywords: molecular mechanics, MM2, AMBER

INTRODUCTION

MolStruc is an original program for studying molecular mechanics. It was developed to analyze the influence of force field parameters and functions on the final geometry of a given molecule.

Address reprint requests to Dr. G. Pèpe at the Centre de Recherche sur les Mécanismes de la Croissance Cristalline, Campus de Luminy, Case 913, 13288 Marseille Cedex 9, France.

Received 30 April 1990; accepted 10 July 1990

For the electrostatic energy calculation, the well known MM2 program[1] uses a dipole approximation. This will be compared to the simpler monopole approximation.

MolStruc offers the user a choice between two sets of functions and parameters: those from AMBER[2,3] and those from MM2.[1] The main feature of the program is the option to manipulate the force field parameters interactively. A companion program allows one to display any parameter to modify it. If any of the parameters needed to define the input molecule are missing, they can be introduced.

MolStruc will load in memory only the parameters necessary for the calculation; this feature permits one to reduce table sizes and calculation times.

To ensure that the optimization algorithm has not stopped on an energy plateau, we have implemented two optimizers: conjugate gradient[4] and Newton.[5] The user can switch from one optimizer to the other at the end of the calculation.

MolStruc is about 10 times faster than MM2, in real situations its speed being related to the fact that the molecular geometry is standardized in an analytical way and an original algorithm automatically defines the atom pairs to be taken into account for the nonbonded interactions.

FORCE FIELD FUNCTIONS

The classical function for computing the steric energy in MM is:

$$E_{\text{steric}} = \Sigma E_S + \Sigma E_B + \Sigma E_T + \Sigma E_v + \Sigma E_C + \Sigma E_H$$

where E_S, E_B, E_T, E_v, E_C and E_H are the stretching, bending, torsion, van der Waals, electrostatic (coulombic) and hydrogen bond energies, respectively.

The terms of this function derived from MM2, and used in MolStruc are:

$$E_S = 143.88 \frac{K_S}{2} \Delta l^2 (1 - 2\Delta l)$$

$$E_B = 0.043828 \frac{K_B}{2} \Delta\theta^2 (1 - 0.006\Delta\theta)$$

When an sp^2 atom is the central atom of a bending angle, a modification is made in computing the bending energy to account for deformation of this atom out of the plane.

$$E_T = \frac{V_1}{2}(1 + \cos\omega) + \frac{V_2}{2}(1 - \cos 2\omega) + \frac{V_3}{2}(1 + \cos 3\omega)$$

$$E_v = \epsilon_{ij}^* \left(8.28 \times 10^5 \exp\left(\frac{1}{0.0736} P_{ij}\right) - 2.25 P_{ij}^6\right)$$

Here

$$P_{ij} = \frac{r^*}{R_{ij}}$$

r^* is the sum of the van der Waals radii of the atom pair and R_{ij} is the distance between atoms i and j. If $P_{ij} > 4.039$, then

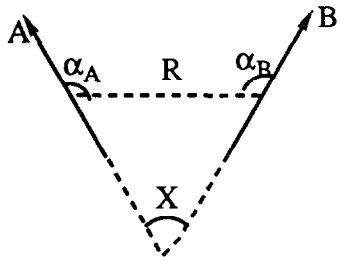
$$E_v = \epsilon^* \, 1157.276 P_{ij}^2$$

$\epsilon_{ij}^* = \sqrt{\epsilon_i \epsilon_j}$ (ϵ_i and ϵ_j are the hardness of the atoms)

$\epsilon_{ij}^* = 0.0508$ kcal/mole when hydrogen atoms are involved

In this representation, hydrogen atom bonds are shortened by a factor of 0.925.

$$E_C = \frac{14.39418 \mu_A \mu_B (\cos X - 3 \cos\alpha_A \cos\alpha_B)}{R^3}$$



The electrostatic energy can also be computed in a monopole approximation, as in the AMBER option. (See below.)

This force field does not consider the hydrogen bonds. (In MolStruc, the stretch-bend and torsion-bend contributions to the steric energy are neglected. Calculations performed on different molecule families indicate that these contributions are seldom necessary.)

For AMBER the functions are:

$$E_S = K_S \Delta l^2$$

$$E_B = K_B \Delta\theta^2$$

$$E_T = \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$

where $n = 2$ or 3, depending on the hybridization of the bond atoms, and $\gamma = 180°$ for sp^2 atoms and 0° for sp^3 atoms.

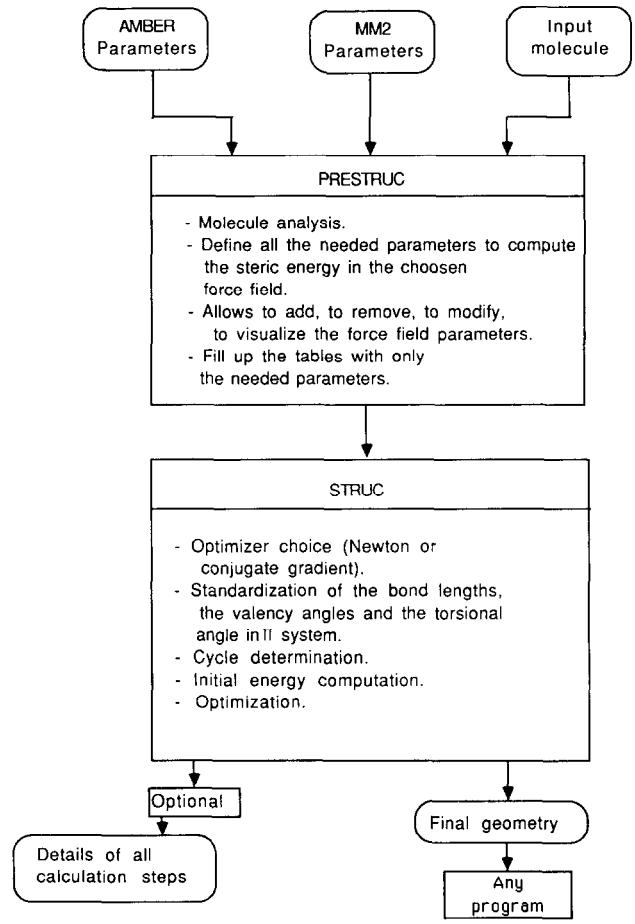
To account for the out-of-plane deformation of sp^2 atoms, AMBER uses the improper torsion represented by the same formula with different values for the parameters V_n, n and γ.

$$E_v = \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6}$$

$$E_C = \frac{1}{\epsilon} \frac{q_i q_j}{R_{ij}}$$

Here q_i and q_j are the net atomic charges of the atoms i and j. These charges are automatically computed with the Del Ré method[6] for the σ part with a new set of atomic parameters.[7] For the Π part, charges are corrected in an empirical way to obtain the Pariser and Parr[8] values. In the original AMBER program, these charges are tabulated and come from quantum mechanics calculations. We chose $\epsilon = R_{ij}$.

$$E_H = \frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}$$



DESCRIPTION OF THE PROGRAM

The technical features are:

- Language: FORTRAN 77
- Size: Source, 11000 lines; Memory, 8MBytes
- Computer: Any 32-bit computer

The memory size corresponds to a version of the program able to study a molecular size of up to 500 atoms.

The input data consist of:

- Parameter libraries
- Atomic Cartesian coordinates and atom codes

The input files are sequential for portability.

Output data consist of

- The molecule file with atom names, Cartesian coordinates, connection table and net atomic charges
- Optional: An output file with all the calculation details:
  - Steric energy and average displacements of the atoms at each calculation step
  - Atomic coordinates, derivative values of the strain energy towards the coordinates and root mean square values (RMS) of the derivatives to verify the optimizer convergence, at the beginning and end of the computation

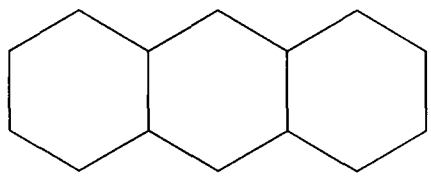
Program specifications include

- Choice of the force field
- Choice of the optimizer
- Geometry standardization
- Cycle determination (allows the definition of the correct bonded parameters)

An original algorithm was written on the following principles:

> A molecule is described by a connection table; this information will be used to define the rings. We move forward on the bonds. When an atom already used in a connection path is found, we are in a ring if this atom is the starting point. If we are not at the starting point we backtrack and take the next connected atom.

This algorithm finds all of the cycles present, even if they are juxtaposed. Redundancies are suppressed to minimize the set of the smallest rings. As an example, consider anthracen.



Six rings are found:

Three 6-atom rings
Two 10-atom rings
One 14-atom ring

The three 6-atom rings are kept.

Conjugated systems

When two sp^2 atoms are bonded, the problem is to distinguish whether the bond is a true double bond, a conjugation between two Π systems or part of an aromatic ring. MolStruc automatically defines the nature of each bond and attributes the correct stretching parameters.

RESULTS AND DISCUSSION

The problems relative to MM can be summarized in a sample of representative molecules. As the AMBER option of the program can only refine peptide or oligonucleotide molecules, we chose to compute two peptides in both options of MolStruc to compare the influence of the force field functions and parameters on the final geometries.

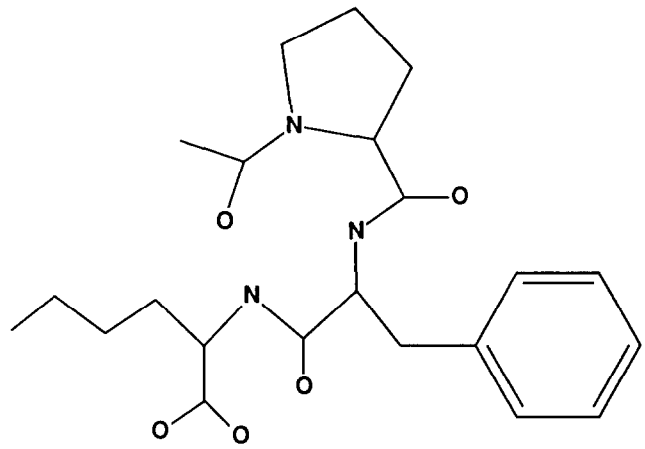
The other compounds are calculated only with the MM2 version of the program. In all cases, the geometries are compared with the known X-ray structures. For each molecule, the steric energy and the RMS difference between the calculated structure and the X-ray structure are given. (The best molecular fit is performed on all the atoms.)

To analyze the effect of the coulombic energy on the steric energy and the geometry, we display the results obtained with MolStruc in the dipole option and in the charge option.

Peptide molecules

Two peptides for which X-ray structures are available were used to test the program.

-N-Acetyl-L-Prolyl-L-Phenylalanyl-L-Leucine (LEU).[9]



-L-Phenylalanyl-L-Proline (PRO)[10].

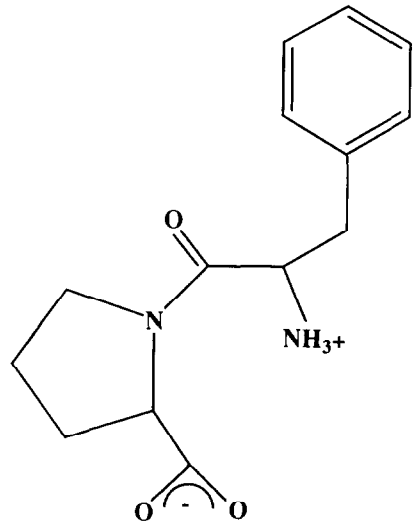


Table 1 gives the energy values obtained with the peptide X-ray structures; the hydrogen atoms have been placed in theoretical positions. The high values of the stretching and bending energies result essentially from the hydrogen atom positions. For the coulombic energy, the same potential and dielectric constant ($\epsilon = R_{ij}$) were used. The steric energy components are of the same magnitude.

In the dipole approximation of the MM2 option, the coulombic energy values are very different: −1.5 kcal/mole instead of 19.4 kcal/mole for PRO and −25.8 kcal/mole instead of 34 kcal/mole for LEU.

The energy and RMS values corresponding to the different options of the program are reported in Table 2. For initial geometries we notice important differences at the coulombic interaction level, depending on the approximation chosen. The best RMS values, obtained with MM2 in the charge option, are not very different from the AMBER option results. The corresponding geometries are superimposed on the X-ray structures in Figure 1 for PRO and in Figure 2 for LEU.

In both cases, the molecules crystallize with a water molecule inducing strong hydrogen bonds. The differences between the observed and the calculated molecules might be related to the limit of the force field calculations on such molecules, and not to this interaction, because the RMS value between the calculated geometries in the different options of the program are of the same order of magnitude.

Table 1. Details of the energy values (kcal/mole) of the PRO and LEU peptides corresponding to the X-ray geometries computed in the *MM2* and *AMBER* options of MolStruc. Contributions are designated as follows: E_v = van der Waals, E_C = coulombic, E_H = hydrogen bonds, E_T = torsion, E_S = stretching, E_B = bending, E_{op} = out of plane and E_{ti} = improper torsion

Molecule	PRO			LEU		
Force field / Energies	MM2 charge	AMBER	Diff	MM2 charge	AMBER	Diff
E_s	16.8	18.0	1.8	38.6	40.2	1.6
E_b	15.1	14.1	−1.	16.4	19.4	3.0
E_{ti}	—	0.9	—	—	0.2	—
E_{ofp}	0	—	—	0	—	—
E_t	3.2	1.4	−1.8	7.1	2.7	−4.4
E_v	11.8	13.4	1.6	21.3	20.1	−1.2
E_c	−19.4	−19.4	0	3.4	3.4	0
E_h	—	−0.4	—	—	−0.7	—
E_{steric}	66.2	67.1	0.9	86.8	85.4	−1.4

Table 2. Final energy (kcal/mole) and RMS (Å) values obtained on the two peptides PRO and LEU in the different options of MolStruc, showing huge differences between the coulombic energies computed in dipole and monopole approximations. The best molecular fit is performed with the X-ray structures

Molecule	PRO			LEU		
Force field / Energies	MM2 dipole	MM2 charge	AMBER	MM2 dipole	MM2 charge	AMBER
E_s	1.	1.1	1.1	1.5	1.8	1.4
E_b	7.3	7.5	3.5	7.8	8.	3.7
E_{ti}	—	—	0.2	—	—	0.
E_t	0.3	0.1	0.5	−0.3	−0.6	0.4
E_v	5.2	4.8	4.6	7.6	7.7	4.
E_c	−1.8	17.3	18.8	−26.8	1.9	2.8
E_h	—	—	−0.1	—	—	−0.7
E_{steric}	12.	30.8	28.6	9.1	18.8	11.7
RMS with X-ray structures	0.21	0.12	0.17	0.3	0.25	0.28

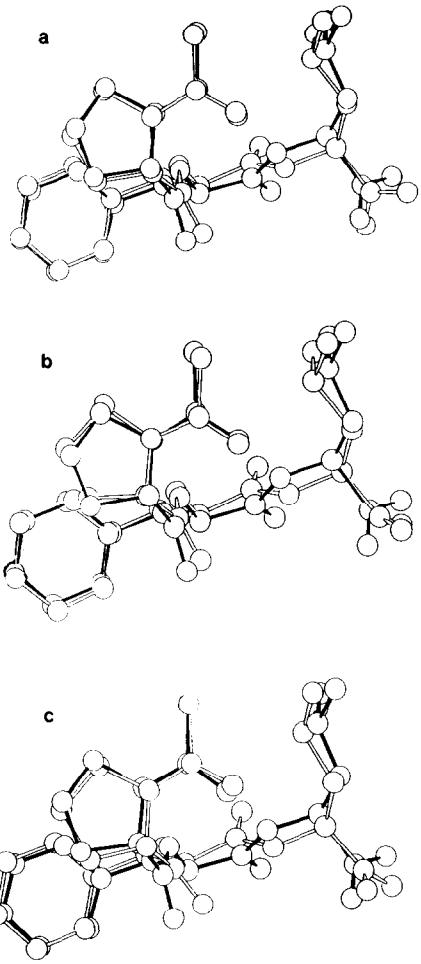


Figure 1. Molecule of the peptide LEU computed in the three options of MolStruc—a, MM2 dipole; *b, MM2* charge; *and c, AMBER—superimposed on the X-ray molecule. (The filled line corresponds to the X-ray structure.) The energies and corresponding RMS values are reported in the Table 2*

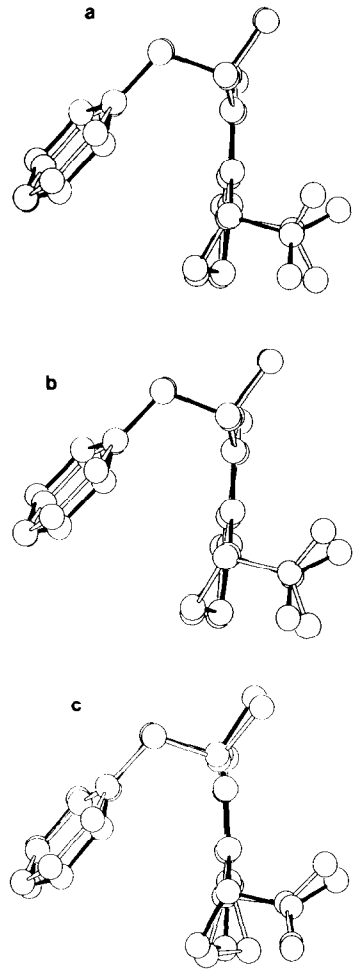
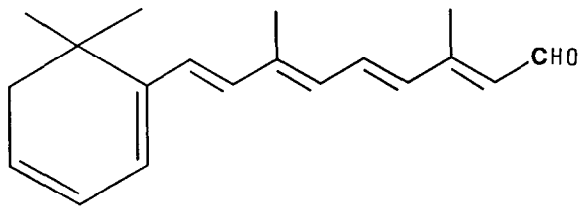


Figure 2. Molecule of the peptide PRO computed by MolStruc as indicated in Figure 1

Molecule with an extended conjugated Π system

All-Trans-3,4-Dihydroretinal (REA).[11]



This molecule presents an extended Π system having alternating double and single bonds.

In the original MM2 program, the Π systems are taken into account by a strengthening of the torsion constant, without distinguishing the bond nature: partially Π from truly double bonds. The consequence is a flattening of the Π system: All bonds between sp^2 atoms are considered as double bonds having the same length, while in the MM2 option these bonds are distinguished. The steric energy values and the RMS with the experimental structures are equivalent, as indicated in Table 3. The corresponding superimposed geometries are shown in Figure 3.

As the net atomic charges are very weak in such molecules, the electrostatic energies issued from the dipole or the monopole approximation are not very different (0 kcal/mole and −3.1 kcal/mole).

In the experimental structure the torsion angle between the chain and the ring is 55.6°. With the original MM2 program this angle becomes 17° while MolStruc in the MM2 option gives 60°.

A soft molecule completely σ

Dodecyldimethylpropylammonium (DOD).[13]

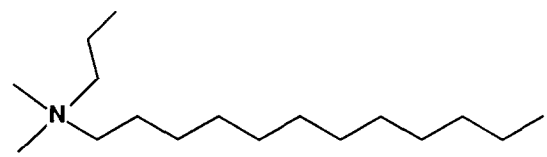


Table 3. Computational results (energy in kcal/mole and RMS in Å) obtained for three molecules (REA and DOD) representative of MM problems. The best molecular fit always corresponds to the coulombic energy calculated with the net atomic charges

Molecule	REA		DOD			
			Conformation 1		Conformation 2	
Force field			MM2	MM2		MM2
Energies	MM2 dipole	MM2 charge	dipole	charge	MM2 dipole	charge
E_s	1.8	1.8	2.7	2.7	2.8	2.8
E_b	4.4	4.5	5.3	5.4	5.6	5.8
E_t	−8.	−8.	1.5	1.5	1.9	1.9
E_v	15.4	15.4	13.0	12.9	13.0	12.8
E_c	0.	2.2	0.	6.8	0.	6.8
E_{steric}	13.6	11.5	22.5	29.3	23.3	30.1
RMS with x-ray structures	0.12	0.12	0.07	0.07	1.73	1.74

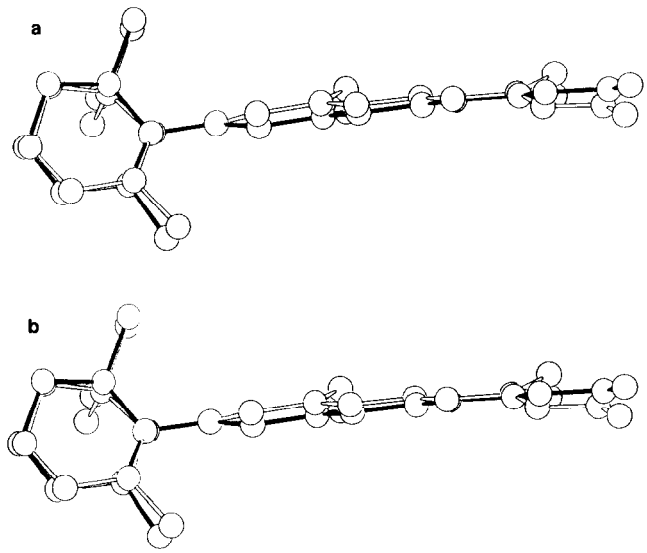


Figure 3. REA, a molecule with an extended conjugated Π *system, computed in the MM2 options—a,* dipole; *and b,* charge—*superimposed with the X-ray structure (Table 3). The force field parameters flatten the geometry*

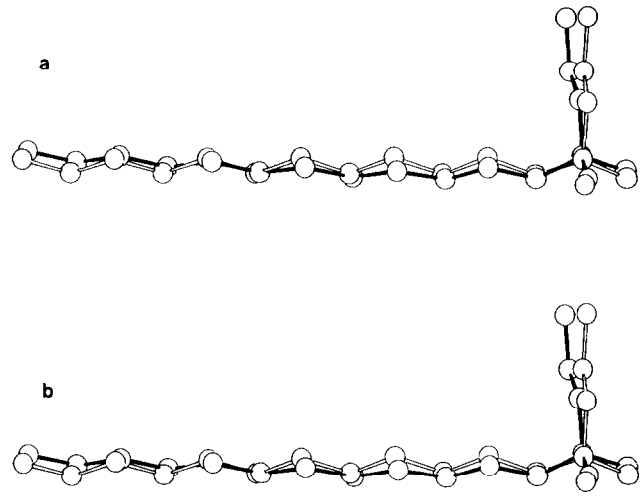


Figure 4. DOD, a complete σ *molecule. The superposition with the X-ray structure is very good in both options of MolStruc: a, dipole; and b, charge. The RMS value is 0.07 Å in both cases (Table 3)*

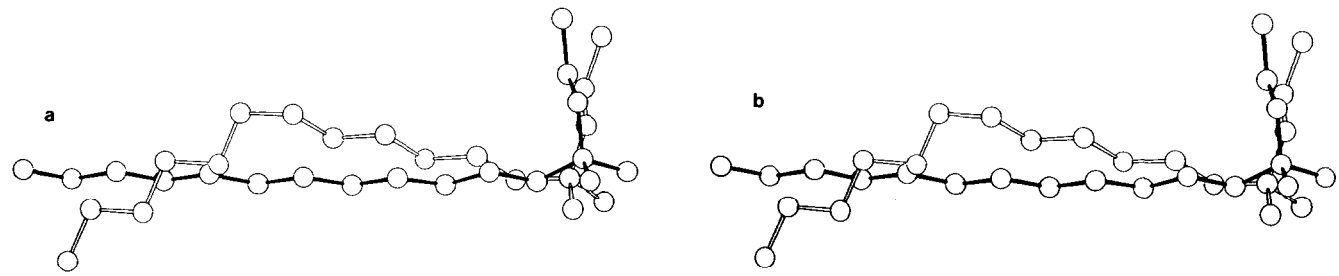


Figure 5. The molecule DOD in another conformation expressing the problem of the local minima of the MM. An optimizer algorithm is unable to find the global minimum (Table 3)

This molecule belongs to the σ type. It is an ideal model for molecular mechanics. This method was initially developed to treat saturated hydrocarbons. In this case, no matter what approximation is used we obtain a very good superposition between the calculated and the observed geometries. (Residual atomic charges are very weak, except on the nitrogen atom.) The RMS is 0.07Å (Table 3, Figure 4), although the coulombic energy is also underestimated in the dipole approximation: 0 kcal/mole instead of 6.8 kcal/mole.

To show that it is not possible to get out of a local minimum in the steric energy using a minimizer algorithm, we chose another conformation of the molecule, very close energetically to the preferred one (30.1 kcal/mole instead of 29.3 kcal/mole). The optimizer remains at the same conformation, as indicated by the RMS value: 1.75 Å (displayed on Figure 5).

CONCLUSION

The calculations performed on peptide molecules in the MM2 or AMBER option of the program indicate that the difference between the obtained geometries are within the limits of the molecular mechanics approximation.

MolStruc with MM2 parameters and functions gives good results for pure σ molecules, while significant differences are observed for molecules with Π systems.

In all cases, the best results are obtained with the net atomic charges instead of diatomic dipoles.

The final geometry of a soft molecule will correspond to the starting conformation, which means that one must be very careful when extrapolating MM results.

All the drawings were performed by the ORTEP program.[14]

REFERENCES

1 Allinger, N.L. and Yuh, Y.H. in *Programme MM2 (1980) Quant. Chem. Prog. Exc. No. 395,* Bloomington, Indiana

2 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. Jr. and Weiner, P. *J. Am. Chem. Soc.* 1984, **106**, 765

3 Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A. *J. Comp. Chem.* 1986, **7**, 2, 230

4 Polak, E. *Computational Methods in Optimization* Academic P., New York (1971) Sect. 2.3

5 Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. in *Numerical Recipes* Cambridge Univ. P., Cambridge (1986)

6 Del Ré, G. *J. Chem. Soc.* 1958, **40**, 31

7 Pèpe, G., Serres, B., Laporte, D., Del Ré, G. and Minichino, C. *J. Theor. Biol.* 1985, **115**, 571

8 Pariser, R. and Parr, R.G. *J. Chem. Phys.* 1953, **21**, 446, 767

9 Precigoux, G., Joffre, S. and Ouvrard, E. *Acta Crystallogr., Sect. C* 1986, **42**, 721

10 Panneerselvam, K. and Chacko, K.K. *Acta Crystallogr., Sect. C* 1989, **45**, 106

11 Simmons, C.J., Asato, A.E. and Liu, R.S.H. *Acta Crystallogr., Sect. C* 1986, **42**, 711

12 Tinant, B., Declercq, J.P. and van Meerssche, M. *Acta Crystallogr., Sect. C* 1986, **42**, 579

13 Taga, T., Machida, K., Kimoura, N., Hayashi, S., Umemura, J. and Takenaka, T. *Acta Crystallogr., Sect. C* 1986, **42**, 608

14 Johnson, C.K. *ORTEP II.* Oak Ridge Natl. Lab. report ORNL-5138 (1976)