MOL3D, a modular and interactive program for molecular modeling and conformational analysis: I — basic modules

D. Pattou and B. Maigret*

Laboratoire de Modèles Informatiques Appliqués à la Synthèse (U.A. CNRS 490), Institut LeBel — Université L. Pasteur, 4 rue Blaise Pascal, 67000 Strasbourg, France

MOL3D is a generalized machine-independent computer program that lets the user interactively build 3D structures with different display options, such as wire, ball-and-stick and CPK representations. The program, which uses its own graphics package and driver, is designed to be very user friendly through the use of commands and menus. It has powerful transformation capabilities, such as software rotations, superpositions and zooming, and it is equipped with a fragment database that allows the user to build complex structures. The algorithm presented here is designed to perform computations in all the conformational space and therefore can be used to predict experimentally available quantities, such as NMR coupling constants. The program is efficient in the sense that it handles only dihedral angles in the first steps; as a result, it allows a rapid sampling of a great number of points through the entire conformational space. The user can choose between grid and Monte-Carlo searches of energy minimization, using a reasonable amount of computer time.

Keywords: computer modeling, conformational analysis

Received 14 October 1987
Accepted 24 November 1987

INTRODUCTION

Three-dimensional (3D) models of molecules are indispensable tools in many areas of organic chemistry, molecular pharmacology and biochemistry. For many years, scientists have been routinely using mechanical CPK, Dreiding, and other models to obtain crude estimates of spatial relationships in molecules where 3D information is essential.

However, if you want to determine which requirements a molecule should satisfy to be active, you must know its steric properties and compare them to those shown by other molecules acting the same way. Because no detailed information on the receptor structure is generally available, you have to infer the target shape and properties by receptor mapping techniques. This assumes some knowledge of the steric features of a series of molecules that, on an experimental basis, are expected to behave similarly. To understand such molecules, it is important to build and manipulate many spatial models and to compare them in their approximate stable conformations obtained from crystallographic experiments, when known, or by theoretical computations. These goals require an enormous number of attempts, even if the problem can be reduced by taking into account different data from physico-chemical experiments (NMR seems to be the most suitable technique for that purpose).

The remarkable advances of computer technology in the last few years have made considerable impact on molecular modeling. Since the pioneering work of Levinthal,[1] many computer-assisted molecular packages have become available, both from academic and industrial company sources, to help building molecular structures as complex as biological ones. Moreover, the latest developments in computer technology have led to the availability of powerful systems, so that individual research groups — or individuals themselves — can have their own modeling systems. But the price of source

*To whom all correspondence should be sent at: Groupe de RMN et Modélisation Moléculaire, Laboratoire de Chimie Organique Physique, U.A. CNRS, No. 422, Institut LeBel — Université L. Pasteur, 4 rue Blaise Pascal, 67000 Strasbourg, France

code and the continual changes in hardware impede the portability of programs. The portability requirement, in fact, may conflict with the use of specialized high-speed hardware, for which machine-dependent codes are necessary to reach the required performances.

In fact, an efficient multipurpose molecular modeling package must handle the simplest as well as the most sophisticated tasks to tackle a large variety of problems. To overcome conflicting interests between all the possible users, such a system needs to be highly modular, so that it can be adapted easily to specific requirements. Moreover, users should be able to run parts of the program immediately on personal computers as well as on mainframes. Such a molecular modeling package must also consider the availability of existing distributed algorithms, such as QCPE, and molecular databases, such as Cambridge Crystallographic Data Base.

These various reasons pushed us to develop our own modeling package. The technical choices that we retained take into account the possibility of accessing the system from various devices, depending on the user's interests and the problem's complexity. The system can thus be used in a network, from low-cost raster terminals linked to a mainframe computer to more sophisticated 3D workstations.

GENERAL DESCRIPTION OF MOL3D PACKAGE

Today, many different molecular modeling systems can be found all over the world.[2] Some of them need high-cost graphics terminals, including high-resolution display, real-time rotations and depth-cuing. Many preliminary modeling applications, however, do not need such sophisticated workstations, so packages intended for personal computers and limited to basic applications are also available. In order to extend the choice to applications in the medium price range, we have developed a modular molecular package using low-cost 2D terminals (Tektronix raster terminals of the 41xx series) connected to an IBM mainframe or to a micro-VAX network. MOL3D software is written in ANSI Fortran 77 without extensions, using its own graphic driver.[3] The system's core is presented in Figure 1.

MOLECULAR EDITOR

The user draws the 2D chemical structure directly on the screen, using the terminal's joystick and the keyboard for defining atom and bond types (the default is a carbon atom and a single bond). This method seems natural for the user and makes the program easy to use for beginners (see Figure 3). Of course, the calculation of the preliminary 3D coordinates does not depend on the quality of the drawing. The user can calculate and display the 3D molecule at any time by picking the appropriate option in the editor's menu.

For noncyclic moieties in a molecule, 3D Cartesian coordinates are obtained using the Thompson's procedure.[4] The corresponding internal coordinates are taken from tables (bond lengths, bond angles) or are chosen to zero (dihedral angles).

The 3D Cartesian coordinates of individual rings are also obtained from internal coordinates. In this particular case, dihedral angles are calculated from the pseudoangle approximation for ring size smaller than 8, or from Go and Scheraga's or Madison's works[5,6] for bigger rings. For five-membered rings, we used the works of Pitzer,[7,8] Geise,[9,10] Dunitz[11] and other authors.[12-15] Six-membered rings are built similarly, according to the results obtained by Hendrickson,[16] Dunitz,[11,17] Cremer and Pople,[18] Buys and Geise.[19] The precise determination of seven-membered ring conformations requires many parameters (Esteban[20] or

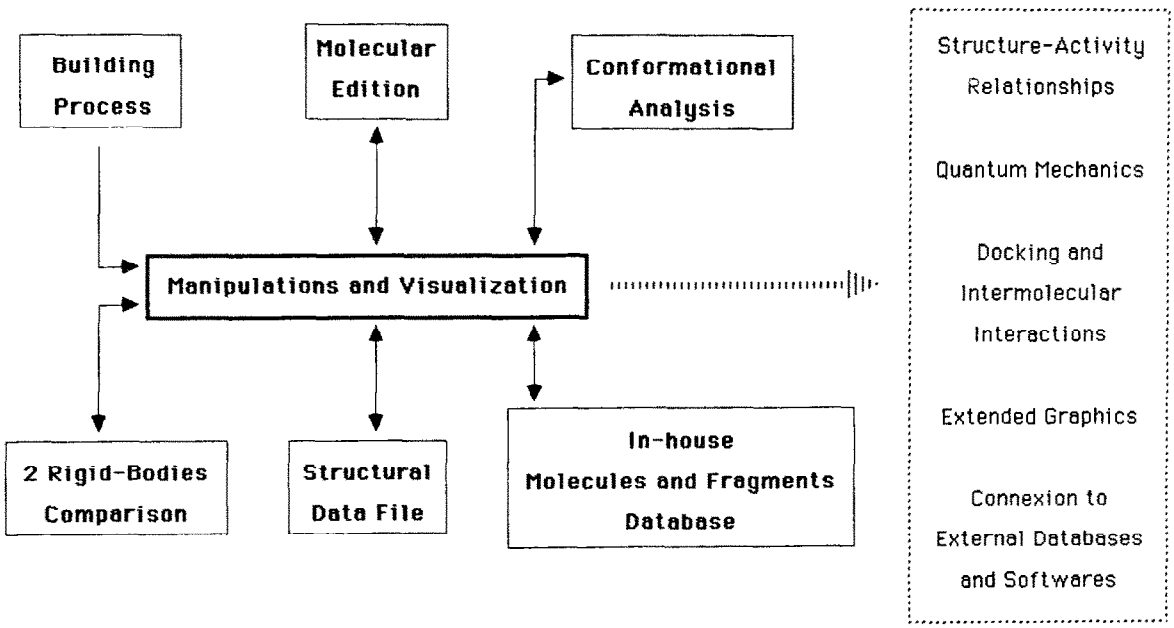


Figure 1. General organization of the MOL3D package

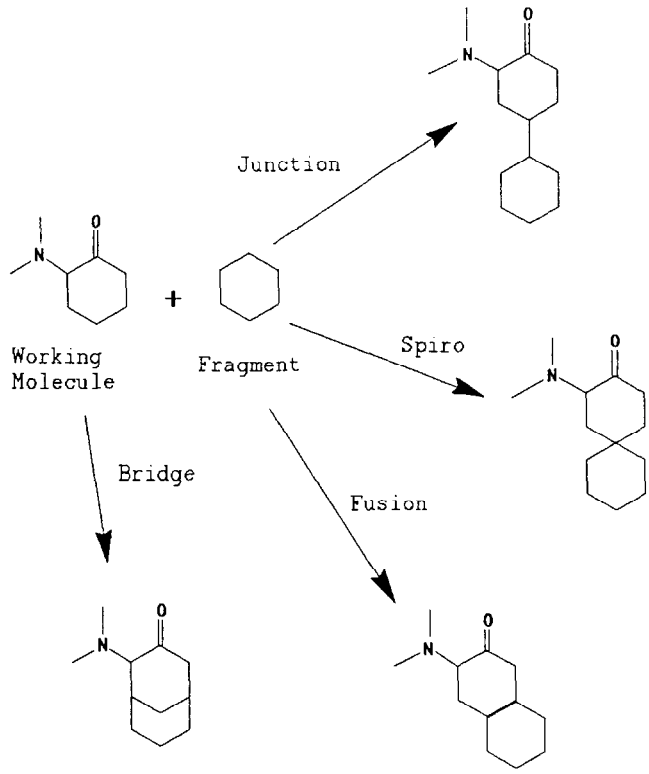


Figure 2. Example of the MOL3D-allowed gluing between the working molecule and a molecular fragment

Flappers and Romers[21]), so the program proposes directly only the more common structures (chair, boat, twisted-chair, twisted-boat). To obtain the other ones, the user enters the corresponding values of the parameters.

At this level, the data concerning a molecular structure are given as follows:

- Atom identification: atom number, atom type and connectivity
- Bond identification: bond number, bond type (single, double, triple or aromatic) and numbers of the atoms it links
- Atom coordinates: Cartesian and internal ones
- Atomic charge, if any

When building a molecule, the user can also merge a 3D fragment to the working structure. These fragments can be small common molecules (cyclohexan, benzene, cyclopropan, etc.) whose coordinates have been experimentally obtained by crystallography, but they can also be user-specific molecules, for which 3D coordinates have been obtained experimentally or by previous theoretical computations. These fragments are contained in a personal standard box. A fragment can be merged to the working structure, using four different methods (see Figure 2):

- Junction: a bond is created between an atom of the molecule and an atom of the fragment

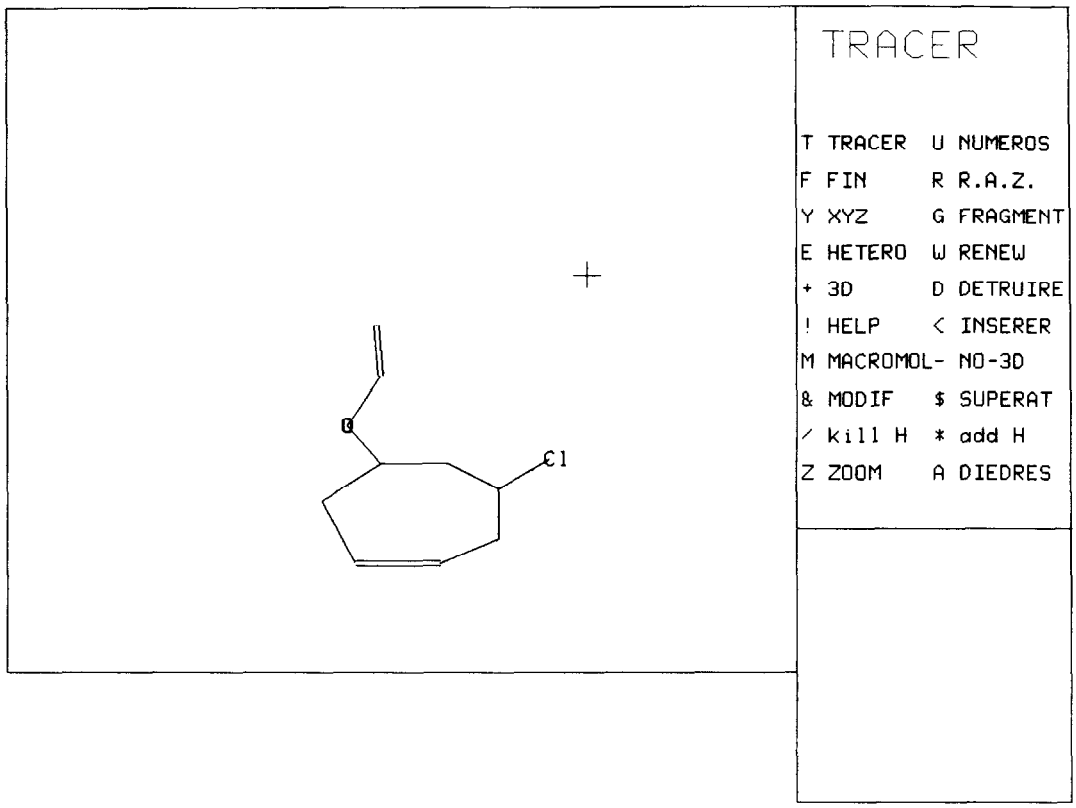


Figure 3. Drawing of the molecular graph of the working molecule using the mouse and the keyboard

- Spiro: the (cyclic) fragment is linked to a ring of the working structure in a spiro way
- Fusion: the fragment is fused to the molecule
- Bridge: an m-n-1 bicyclo molecule is made from the fragment and the main structure. So far, the bridge has been limited to one atom.

During this operation, both the fragment and the molecule are displayed on the screen (Figure 4) so the user can give all the necessary information for a correct addition. When it is finished, the whole molecule is displayed in 3D (Figure 5).

Finally, a 3D molecular model of a molecule built from repeated units (peptide, nucleic acid) can be obtained easily, according to the method given by Hermans and Ferro.[22] For such polymeric molecules, this method offers the great advantage of obtaining large structures using a very limited database, the "dictionary," containing the 3D information for each building unit (amino acid, nucleic base, etc.). The user enters the molecule's sequence from the keyboard, and the program requests the desired conformational type (helix, etc.), as Figure 6 shows.

CONFORMATIONAL ANALYSIS MODULE

Empirical energy function

Empirical formulas used in intramolecular energy calculations are often derived from intermolecular ones.[23] We did likewise by using the potential functions first proposed by Claverie[24-27] and further adapted to intramolecular calculations.[28] The atom-atom contributions include:

- a monopole–monopole electrostatic term (atomic charges are here calculated according to Delre[29,30] for sigma terms and Huckel[31] for pi terms)
- a Van Der Waals term, of Buckingham's type, derived from Kitaygorodskii's works.[32] This Van Der Waals contribution is corrected to take into account hydrogen bonding[26] and very short-range interactions,[27] not usefully described with a classic Buckingham potential

A polarization electrostatic term is also included to take into account possible induced electrostatic effects.

This representation of the potential energy of a molecule does not include any stretching or bending terms, as in many force fields. Such a choice has been made because of the conformational analysis strategy we used. At this level, the program used only the dihedral angles as possible parameters of the conformational searches. Thus, because of the separation between "soft" (bond lengths and angles) and "hard" (dihedral angles) variables in conformational energy minimization processes, the molecular geometry can be kept during the conformational search and refined next using one of the commonly used force fields, such as MM2 and AMBER.

As a test for Claverie's potential use, we have studied

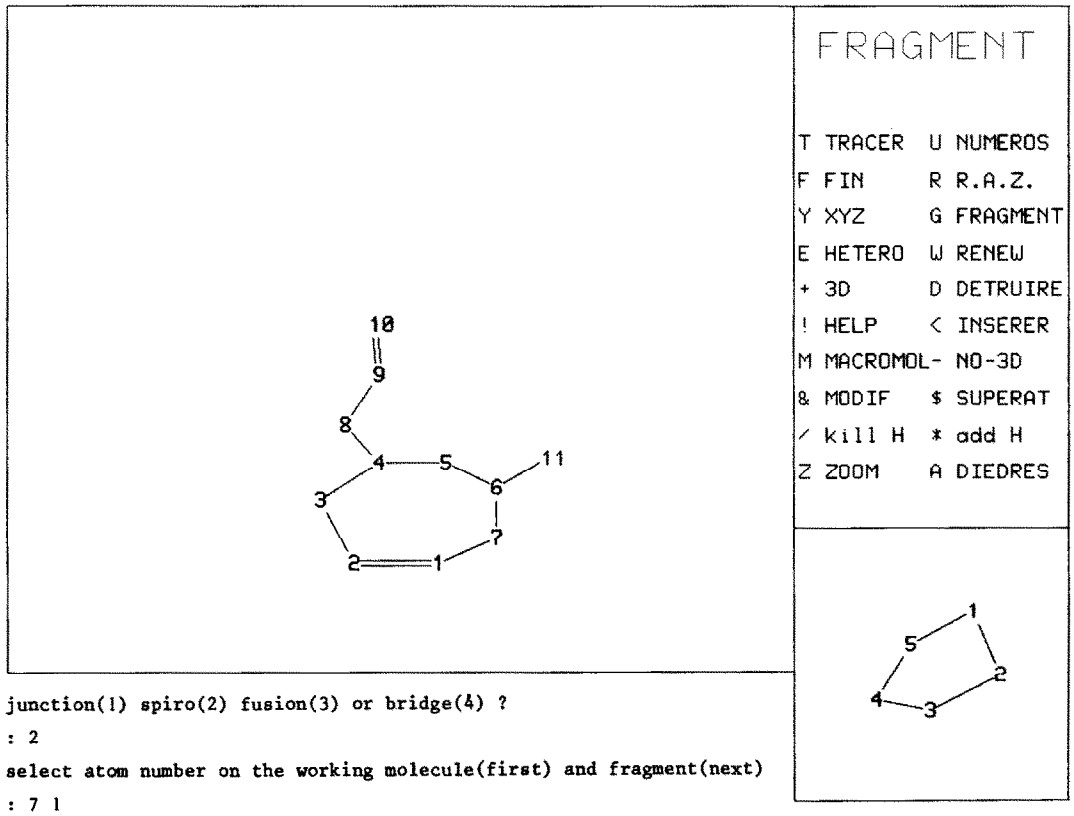


Figure 4. Drawing of the working molecule and of the called fragment

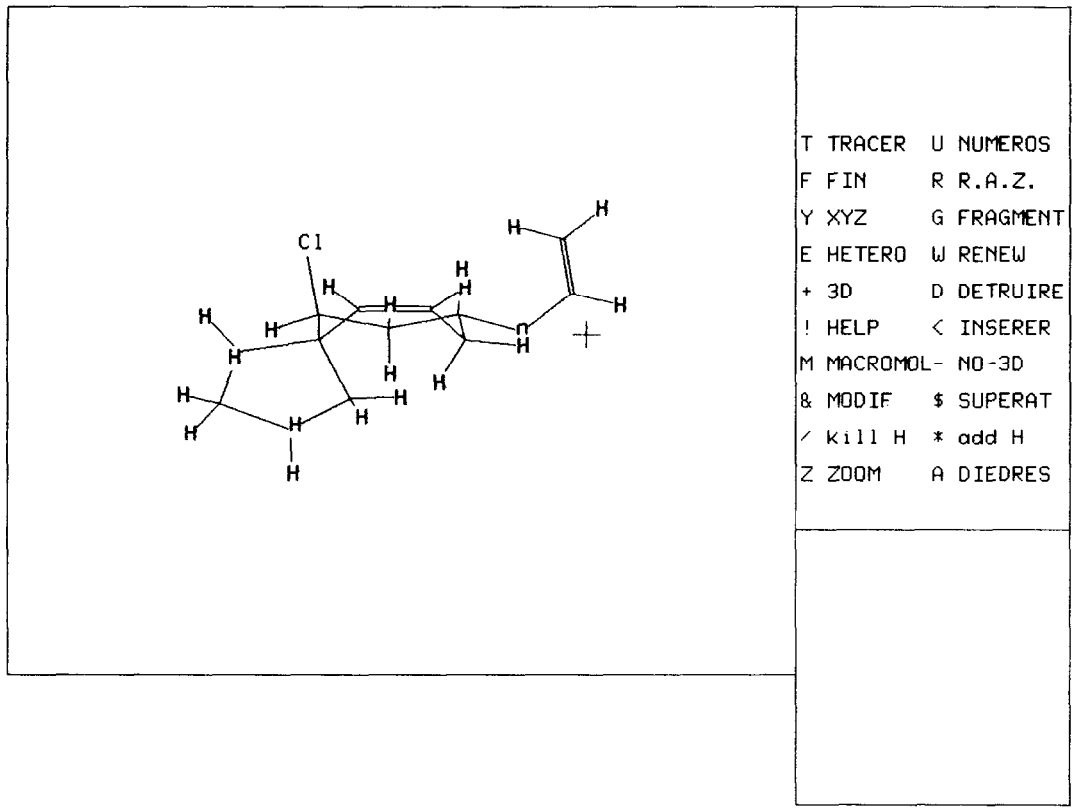


Figure 5. 2D to 3D transformation of the resulting structure

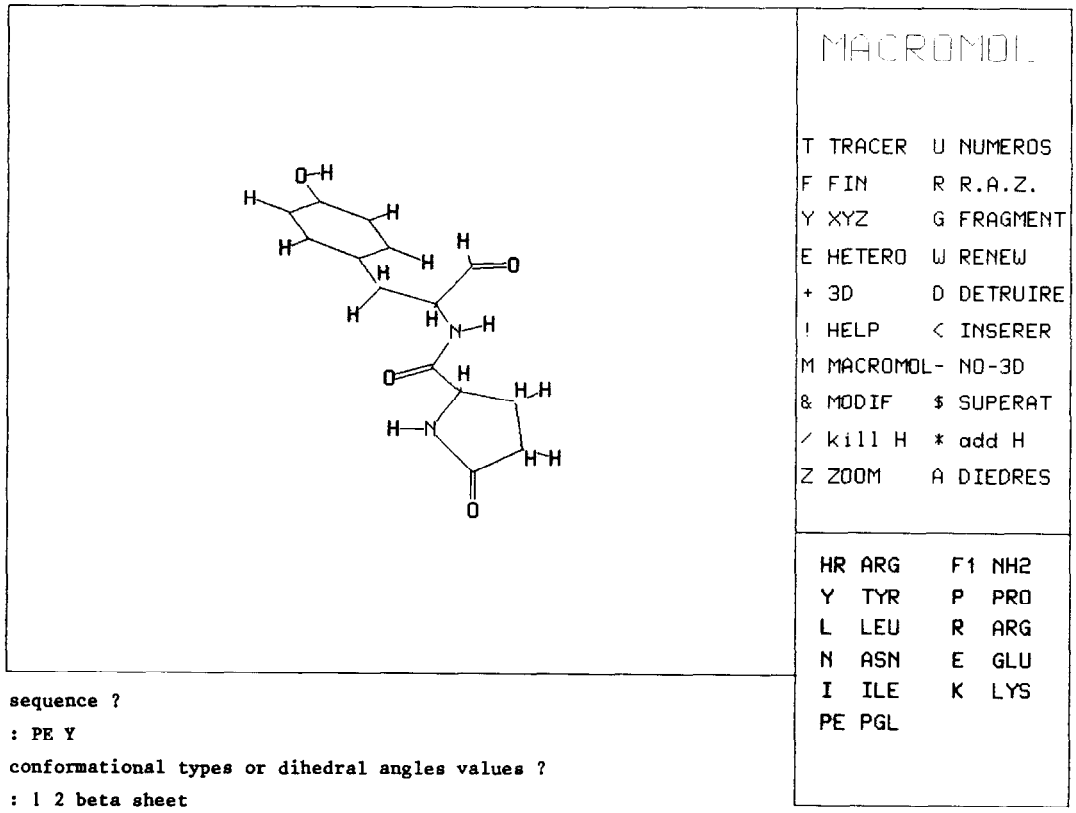


Figure 6. Example of a small peptide building using the user's amino acids dictionary

the conformational energy map of an alanyl dipeptide (N-acetyl-N'-methyl alanylamide). The literature includes many theoretical energy calculations[33-35] for this dipeptide, because of its conformational simplicity (only two free rotation bonds), its relative chemical complexity (existence of a lateral chain, peptidic bond, etc.) and the abundance of available experimental data.[36-38] As you can see in Figure 7, the results we obtained can be successfully compared to experimental and theoretical ones.

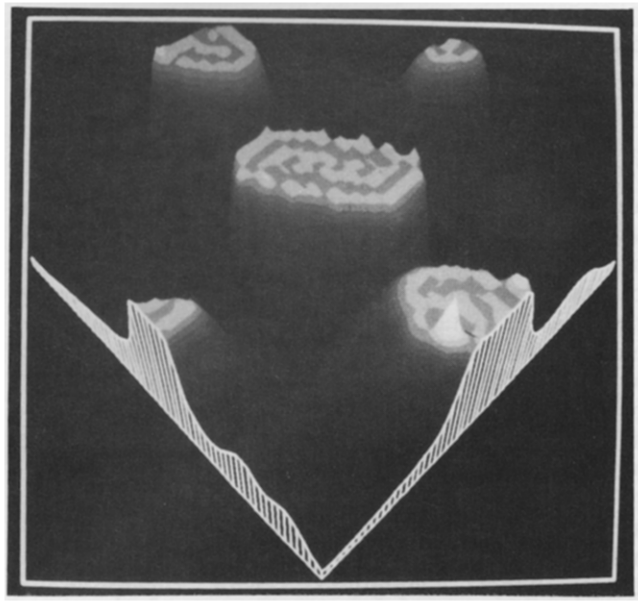


Figure 7. 3D representation of the usual Ramachandran energy map of the N-Acetyl-N'-methyl-Alanyl-dipeptide

Conformational search strategy

The problem now is determining the stable (i.e., low-energy) conformations of the investigated molecules. Because many minimization strategies exist, we decided to give the user as many choices as possible; the best method depends on the user's specific problems.

Systematic explorations of the molecule's whole conformational space can be obtained by a grid method. This is the simplest and the most efficient method, but it can be used only with a very restricted number of variables (Figure 8 includes an example of a one-parameter search, and Figure 7 shows the result of a two-variable exploration).

For larger conformation spaces, this systematic exploration can be efficiently replaced by a statistical Monte-Carlo search. The method proposed here is the Metropolis[39] one, modified by Premilat and Maigret[40,41] for peptidic chains and extended to common organic molecules. In this method, a preliminary set of N_1 independent conformations is randomly selected in the conformational hypersurface. Then, for each one of these N_1 conformations, N_2 perturbations are made, by choosing randomly one of the dihedral angles and giving it a random value. At each perturbation step, the energy E_{i+1} of the new conformation S_{i+1} is calculated and compared to the one E_i of the precedent state S_i. If $E_{i+1} < E_i$, the new conformation is accepted and is the basis of a new perturbation. But if $E_{i+1} > E_i$, the new conformation is accepted with the probability $P_{i+1} = e^{-\Delta E/RT}$, where $\Delta E = E_{i+1} - E_i$. A random number $X(0 < X < 1)$ is then compared to P_{i+1}: conformation S_{i+1} is accepted if $X < P_{i+1}$. Each of the N_1 states of the initial sample grows therefore progressively to high-probability states, and the method lets

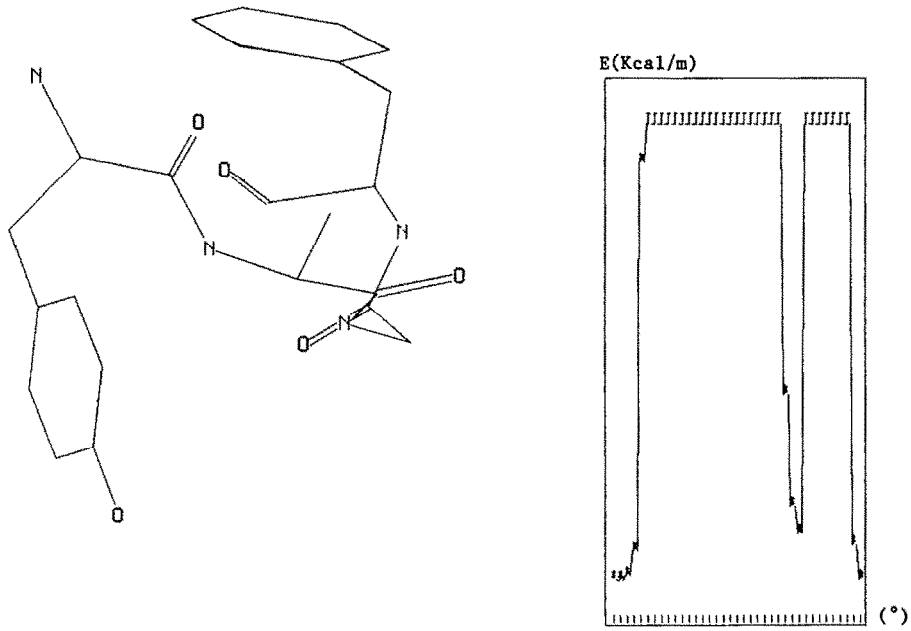


Figure 8. Representation of the energy plot corresponding to a grid rotation around one single bond of the working molecule

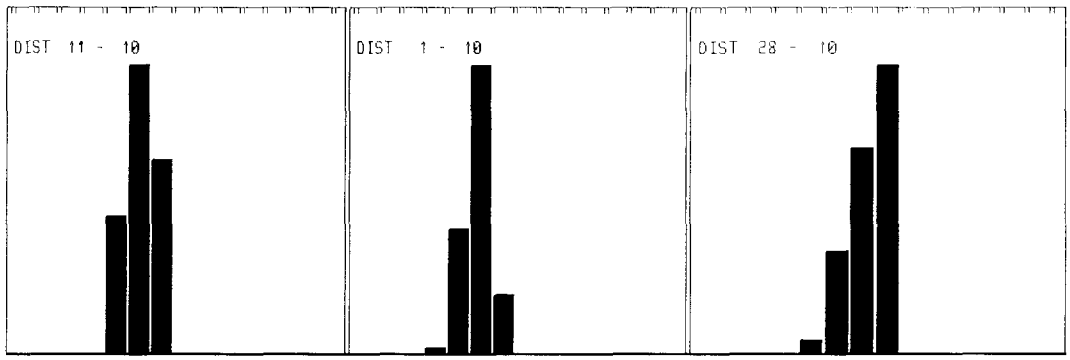


Figure 9. Histograms showing the selected interatomic distances distributions in the sample of conformers obtained through the Metropolis procedure

users obtain several stable conformations. Moreover, this process lets users calculate mean values of coupling constants (to be related to dihedral angles through Karplus curves) or of distances between selected atoms or atom groups (see Figure 9). Such calculated averages can then be compared to experimental results.

If the grid of Metropolis methods allow the determination of conformational-stability areas on the energy hypersurface, none of them permit a precise localization of the minima. Such a refinement (i.e., looking for a minimum energy conformation around a stable initial one) is available in the MOL3D conformational analysis module, using two iterative direct methods, derived from the Simplex one. The first possible algorithm is deduced from that proposed by Nelder and Mead.[42] It allows fast, but not always precise, minimum determinations. The other one is derived from Chandler's Stepit program[43] and is more precise, but much slower.

To include possible experimental data, or to tackle specific goals in the minimization search, we added a constrained-energy optimization procedure. The associated algorithm is a modified version of the CONMIN program proposed by Haarhoff *et al.*[44] Several choices are available for defining the constraint, which can be either of equality or inequality types. Such choices include dihedral angle selections, interatomic distance ones or adaptations to a given pharmacophore (see Figure 10).

MOLECULAR PICTURES

In the MOL3D program, molecules are displayed on the screen as wire models. However, users are often interested in viewing the volume occupied by the studied structure in 3D space; the program thus offers the possibility of CPK and ball-and-stick representations. This part of the program has been taken from J. M. Cromer's work in our lab.[3] The molecules' volumic representation is obtained easily on raster terminals from software or hardware solutions of the *Z*-buffer algorithm.[45,46] The image obtained can be shadowed[47,48] if needed, using different light sources (see Figures 11 and 12). Lastly, the program driver is forecast for a cheap screen's hard copy — on an HP 7475, for instance.

SUPERPOSITION OF MOLECULES

Another basic module of the MOL3D software lets users perform simple steric comparison between two selected molecular models. For example, a molecule, assumed to contain a given pharmacophore, can be called from the structure database and superposed onto the one under study. Thus, it is possible to compare the spatial arrangement of some atoms and then propose hypothesis for biological activity. This can be done either by entering on the keyboard 2 or 3 atoms to be superposed or by using a least-squares method.[49] In both cases, the comparison is made between two rigid sets of atoms, one for the target molecule and one for the studied molecule (Figure 13). The problem of the dynamic conformational adaptation[50] (with superpositions optimized by rotations around the bonds) can be solved by using the constrained conformational energy menu.

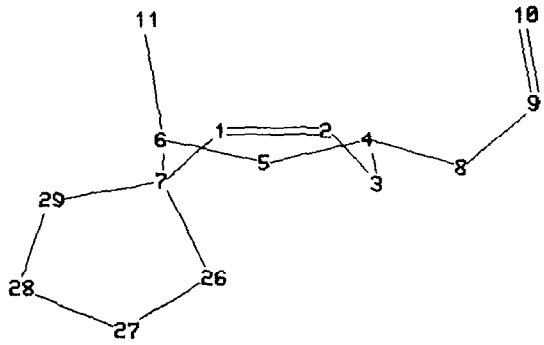
VIEWING AND MANIPULATING MOLECULES

Because it is impossible to obtain, on low-cost 2D terminals, a real-time rotation of the molecule or of the molecule-target representation, our last basic module lets users visualize, manipulate, and calculate the steric properties of molecular models.

The visualization module displays simultaneously on the screen four different user-defined views of the working molecule. Users can then rotate the molecule and choose the best orientation of the structure. Because one of these views, which occupies only one-quarter of the screen, is not always very clear, it can be zoomed onto the whole screen. Rotations, of course, remain possible. On the other hand, molecules can be displayed either with or without hydrogen atoms at any moment of the process.

Molecular distortions are possible, provided that the user indicates which bond must be distorted and by what angular value. The new conformation is then displayed, as well as the energy increase (or decrease) it causes. The user can then verify if the conformation obtained is chemically possible.

All the distances between atoms, atom groups, or ring



DEFINITION OF

CONSTRAINTS

Distances

Bond Angles

Diedral angles

Ring closure

Target

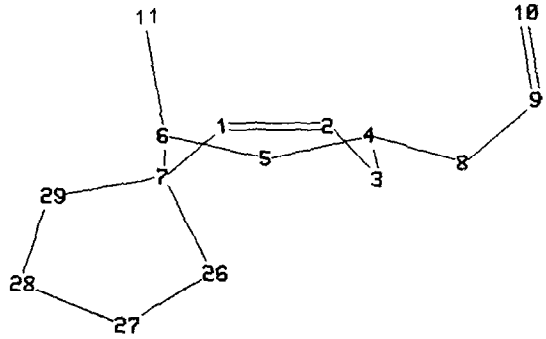
Surfaces

Electronics

N.M.R

exit

a



DEFINITION OF

CONSTRAINTS

INTERATOMIC
DISTANCES

enter -->

Atome1,atome2,

value ,range

b

```
Constraint number -->          1
?
10 11 5.2 1.0
```

Figure 10. (a) General menu of the constrained energy minimization module; (b) example of interatomic distance restraint

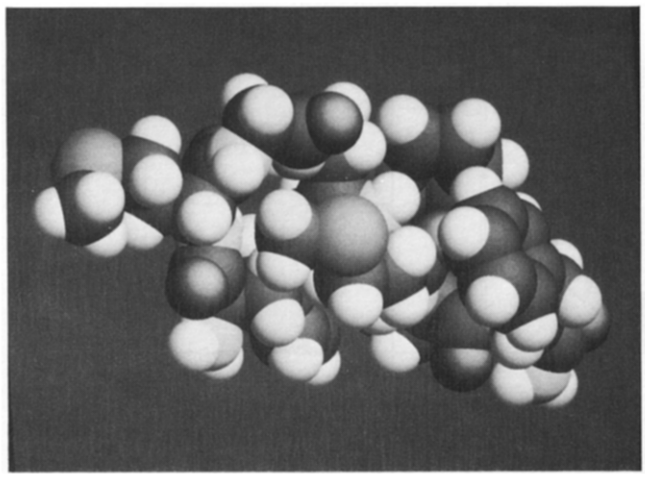


Figure 11. CPK picture obtained from the molecular drawing module

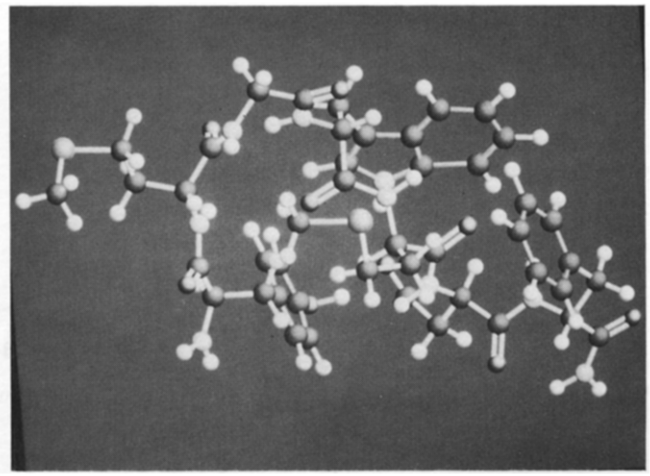


Figure 12. Ball-and-stick picture obtained from the molecular drawing module

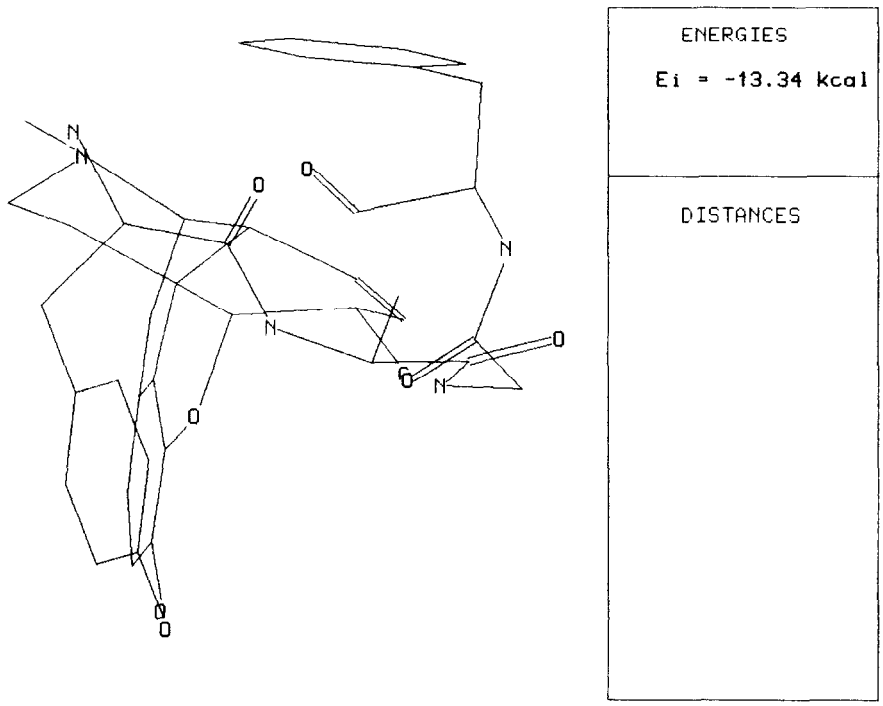


Figure 13. 2-bodies superposition example (black: working molecule, here the Tyr-D-Ala-Gly-Phe peptide; pink: the target, here morphine; the shown energy is that of the peptide molecule in the used conformation; this superposition is obtained from a least-square fit procedure)

centers entered by the user are displayed on the screen. This number must be smaller than 5 (10 crossed distances) simultaneously, but if the user wants to know the complete distance matrix of the molecule, it can be sent on the printer connected to the terminal. Similarly, the user can display dihedral angle values by picking the corresponding atoms.

CONCLUSION

MOL3D, a molecular graphic modeling package integrated in our in-house computational chemistry framework, is a versatile modular and interactive program. It was designed primarily for chemists to build and display 3D molecular structures of up to 256 atoms for theoretical calculations. We started to develop this system of programs about four years ago, after we realized no such program was available for working on low-cost raster graphics terminals.

The MOL3D system is split into two functional subsystems. The first one, which we have presented here, consists of basic operations needed in model building

and conformational analysis. It has successfully been used for theoretical conformational computations.[51] The second subsystem develops and consists of several extensions of the core system, which we will describe in our next paper.[52] We will soon present a more ambitious view of our system, which will permit, on one hand, an interface with Evans & Sutherland PS 300 series graphics workstations and, on the other hand, the availability of top-level molecular modeling tools (utilization of MM2 force field, connection to external databases, etc.).

REFERENCES

1 Levinthal, C. *Scientific American,* 1966, **214,** 42–53
2 Barry, C. D., and McAlister, J. P. High performance molecular graphics, a hardware review. *Computational Crystallography,* D. Sayre, Ed. Oxford University Press, 1982, 274
3 Cromer, J. M. Conception d'un logiciel graphique avec visualisation de molécules 3D pour la CAO en chimie organique. Thesis, IRMA, Strasbourg, 1985
4 Thompson, H. B. *J. Chem. Phys.,* 1967, **47**(9), 3407–3410
5 Madison, V. *Biopolymers,* 1973, **12,** 1837–1852
6 Go, N., and Scheraga, H. A. *Macromolecules,* 1970, **3,** 178–187
7 Kilpatrick, J. E., Pitzer, K. S., and Pitzer, R. *J. Amer. Chem. Soc.,* 1947, **69,** 2483–2488
8 Pitzer, K. S., and Donath, W. E. *J. Amer. Chem. Soc.,* 1959, **81,** 3213–3218
9 Geise, H. J., Altona, C., and Romers, C. *Tetrahedron Lett.,* **15,** 1383–1386
10 Altona, C., Geise, H. J., and Romers, C. *Tetrahedron,* 1968, **24,** 13–32
11 Dunitz, J. D. *Tetrahedron,* 1972, **28,** 5459–5467
12 Cremer, D., and Pople, J. A. *J. Amer. Chem. Soc.,* 1975, **97**(6), 1358–1367
13 Lifson, C., and Wharshel, A. *J. Chem. Phys.,* 1968, **49,** 5116–5129
14 Abillon, E. *Biophys. Struc. Mech.,* 1982, **8,** 257–270
15 Abillon, E. *Biophys. Struct. Mech.,* 1982, **9,** 11–27
16 Hendrickson, J. B. *J. Amer. Chem. Soc.,* 1967, **89**(26), 7036–7061
17 Dunitz, J. D., and Waser, J. *J. Amer. Chem. Soc.,* 1972, **94**(16), 5645–5650
18 Cremer, D., and Pople, J. A. *J. Amer. Chem. Soc.,* 1975, **97**(6), 1354–1358
19 Buys, H. R., and Geise, H. J. *Tetrahedron Lett.,* 1968, **54,** 5619–5624
20 Esteban, A. L. *et al. J. Chem. Soc.,* Perkin Trans. II, 1982, 657–662
21 Flappers, M. J. M., and Romers, C. *Tetrahedron,* 1975, **31,** 1705–1713
22 Hermans, J. R., and Ferro, D. *Biopolymers,* 1971, **10,** 1121–1138
23 Williams, J. E., Stang, P. J., and Von Schleier, P. *Ann. Rev. Phys. Chem.,* 1968, 531–558
24 Huron, M. J., and Claverie, P. *Chem. Phys. Lett.,* 1969, **4**(7), 429–434
25 Huron, M. J., Claverie, P. *J. Phys. Chem.,* 1972, **76,** 2123–2133
26 Caillet, J., and Claverie, P. *Acta Crystallogr.,* 1975, sect. A, **31,** 448–460
27 Caillet, J., Claverie, P., and Pullman, B. *Acta Crystallogr.,* 1976, sect. B, **32,** 2740–2745
28 Gresh, N., Claverie, P., and Pullman, A. *Theor. Chim. Acta,* 1984, **66,** 1–20
29 Delre, G. *J. Chem. Soc.,* 1958, 4031–4040
30 Delre, G. *Biochim. Biophys. Acta,* 1963, **75**(2), 153–182
31 Julg, A. Chimie quantique. Dunod, Paris, 1971
32 Kitaygorodskii, A. I. *Tetrahedron,* 1961, **14,** 230–236
33 Brant, D. A., Miller, W. G., and Flory, P. J. *J. Mol. Biol.,* 1967, **23,** 47–65
34 Chuman, H., Momany, F. A., and Schafer, L. *Int. J. Pept. Protein Res.,* 1984, **24,** 233–248
35 Weiner, S. J., *et al. J. Amer. Chem. Soc.,* 1984, **106**(3), 765–784
36 Mizushima, S., *et al. J. Amer. Chem. Soc.,* 1957, **79,** 5357–5361
37 Bystrov, V. F., *et al. Tetrahedron,* 1969, **25,** 493–515
38 Avignon, M., *et al. Biopolymers,* 1969, **8,** 68–69
39 Metropolis, N. A., *et al. J. Chem. Phys.,* 1953, **21,** 1087–1092
40 Premilat, S., and Maigret, B. *C. R. Acad. Sci.,* 1976, Paris, Ser. C, **282,** 225–228
41 Premilat, S., and Maigret, B. *J. Chem. Phys.,* 1977, **66,** 3418–3425
42 Nelder, J. A., and Mead, R. *Comput. J.,* 1965, **7,** 308–313
43 Chandler, J. P. Quantum Chemistry Program Exchange, QCPE program no. 67, Chemistry Dept., Indiana University, Bloomington, IN, USA
44 Haarhoff, P. C., and Buys, J. D. *Comput. J.,* 1970, **13**(2), 178–184
45 Eyries, F. Synthèse d'images composées de sphères. *Rapport de Recherche de l'Inria,* 1982, No. 163
46 Porter, T. K. *Comm. A. C. M.,* 1979, **4**
47 Brickmann, J., and Raab, H. U. *J. Mol. Graph.,* 1983, **1,** 78–82
48 Max, N. L. Computer representation of molecular surfaces. Lawrence Livermore Laboratory, Livermore, CA, 1981, UCRL 85942
49 McLachlan, A. D. *Acta Crystallogr.,* 1972, sect. A, **28,** 656–657
50 Barino, L. *Comput. and Chem.,* 1981, **5,** 85–90
51 Pattou, D., *et al.* Theoretical analysis of the conformational behaviour of cholecystokinin fragments: I- CCK4, CCK5, CCK6 and CCK7. *J. Mol. Pharm.* (submitted)
52 Pattou, D., *et al.* MOL3D, a modular and interactive program for molecular modelling and conformational analysis: II — Extended Modules (in preparation)