MOL3D—A modular and interactive program for molecular modeling and conformational analysis: II. Extended modules

D. Pattou,* L.H. Bach,† J.P. Kocher‡ and B. Maigret

Laboratoire de Résonance Magnétique et de Modélisation Moléculaire Université L. Pasteur, Strasbourg, France

The new release of MOL3D, a molecular modeling program written in FORTRAN, contains not only enhanced graphic capabilities, but also an improved module for intermolecular calculations that allows rigid and flexible docking. Various interfaces have been added to some well-known and widely diffused programs, such as MM2, AMBER and MOPAC, and to the Cambridge Crystallographic Database. Finally a graph manager and a samples database have been added, which allow efficient searches with various requirements concerning structural templates, pharmacophoric three-dimensional (3D) constraints, and the field of biological activity, if any.

Keywords: molecular modeling, docking, MM2, AMBER, MOPAC, Cambridge Crystallographic Database, structure–activity relationships

INTRODUCTION

Knowledge of the preferred geometry of a molecule and its corresponding properties is of fundamental interest in drug design. From this information, one gains insight into the relationship between structure and activity of compounds, which can be compared and discussed. Unfortunately, the construction by computers of reliable molecular models and the associated conformational searches are not simple tasks. In the case of flexible molecules, the available conformational space is so large that it is necessary to reduce the conformational search to a tractable size, eliminating certain regions by using the experimental available data (especially coming from NMR experiments) or by introducing simplified assumptions (for example, the united atom representation). The potential functions used to calculate the conformational energy must be as general as possible, while giving accurate results. The multiple-minima problem and the significance of the discovered conformations must be answered in some way. Thus, although impressive achievements have been recorded during recent years in the field of computerized molecular modeling, several problems remain to be solved, and there is still room for extensive molecular modeling algorithms and developments.

Several years ago, we decided to build our own molecular modeling package, starting with the basic modules necessary to solve conformational problems associated with most of the commonly found molecules in organic chemistry. Following our general basic ideas, the MOL3D package[1] makes no distinction between small and large biological molecules, in order to remain as general as possible. To give as many choices and flexibility as possible to the user, the system has been extended now to molecular mechanics and dynamics, intermolecular interactions, quantum chemistry, and crystallographic database searches and retrieval, as well as molecular storage, analysis, and comparison. This paper describes the extensions now available in the MOL3D package.

MODIFICATIONS OF THE GRAPHIC FEATURES

The general flowchart of the MOL3D modeling package is illustrated in Figure 1 (modules discussed here are in bold characters). As previously, all of the new modules are written in ANSI FORTRAN 77 without extension. The previous release of the package ran its own graphical package,[2] and was designed for Tektronix terminals of the 4*xxx* series. In the new version, the graphics routines are compatible with the PHIGGS graphics library. Extended graphic features, such as real-time 3D rotations and manipulations, CPK with shading, and depth-cueing are available for molecular display.

*Groupement de Recherches de Lacq Groupe Elf-Aquitaine, France

†Ciba-Geigy AG, Pharmaceutical Division, Basel, Switzerland

‡Laboratoire de Chimie Générale, Faculté des Sciences, Université Libre de Bruxelles, Belgium

Address reprint requests to Dr. Pattou at Groupement de Recherches de Lacq, Groupe Elf-Aquitaine, BP 34, F-64170 Lacq, France.

Received 12 November 1991; accepted 10 December 1991

© 1992 Butterworth–Heinemann

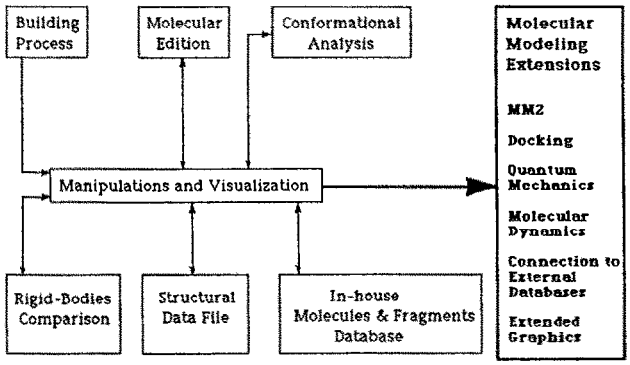


Figure 1. General flowchart of the MOL3D modeling package.

INTERMOLECULAR CALCULATIONS

Rigid and flexible docking

Once the desired molecules are built, or loaded from a given database, one can study the possible interactions between them: a docking module is available, allowing both rigid and flexible conformational intermolecular interactions. This module, which is menu driven, simultaneously can handle docking problems for up to five molecules. As in the conformational analysis module, Claverie's potential functions[3–7] are used here. As long as only a rigid fit is performed, six variables (three Euler angles and three translation variables) are used to move one molecule with respect to the other. When flexibility is taken into account, all the dihedral angles are added to the list of the possible variables. More, if cyclic molecules are considered during flexible optimization, the algorithm of Haarhof and Buys[8] is used to take into account the ring-closure constraints. Similarly, several local symmetry conditions can be introduced to keep a molecule in the correct conformation during the docking procedure. Particular constraints also can be added that consider particular purposes: for instance, the particular shape of the ligand molecule can be maintained during the docking procedure.

Solvent effects

In order to study solvent effects, several possibilities are available to fill solvent boxes, or several shells of solvent molecules around a solute. Optimization of the positions of solvent molecules with respect to the solute can be achieved using Metropolis sampling[9] or energy minimization. In the latter case, the approach followed here is similar to that described by Claverie et al.[10] Another possibility consists of using the AMBER package[11] for energy minimization, through the AMBER interface described below.

INTERFACES TO EXTERNAL SOFTWARES

Molecular mechanics connection to MM2

This interface has been written in order to broaden the capabilities of MOL3D regarding molecular geometry optimization procedures. To date, the package has handled exclusively dihedral angles to minimize the potential energy through Simplex-type algorithms or Monte Carlo methods.[12] A supplementary module was necessary, in order to allow MOL3D to deal simultaneously with dihedral angles, bond lengths, and valence angles during energy optimizations, and therefore to obtain a satisfactory molecular geometry, even when the building procedure misguides the construction of models.

MM2, a molecular mechanics program written by N.L. Allinger[13] was chosen because of its wide distribution among the scientific community, the frequent updates of its force field parameters, and its inclusion in the Quantum Chemistry Program Exchange library (QCPE). Moreover, the MM2 force field, initially designed for small molecules and recently extended to peptidic molecules,[14] makes this program an attractive tool for the range of molecules studied in MOL3D. The MOL3D–MM2 interface (Figure 2) consists of three parts:

(1) Conversion of the type of each atom into an MM2 internal atom type
(2) Generation of the molecular connectivity file in MM2 format
(3) Check of the presence of all the force field parameters needed. This operation is performed in order to prevent MM2 runs from aborting due to the lack of information. Missing values can be introduced interactively, as shown for thiophene in Figure 3. These values thus will be automatically stored and added to the parameter database.

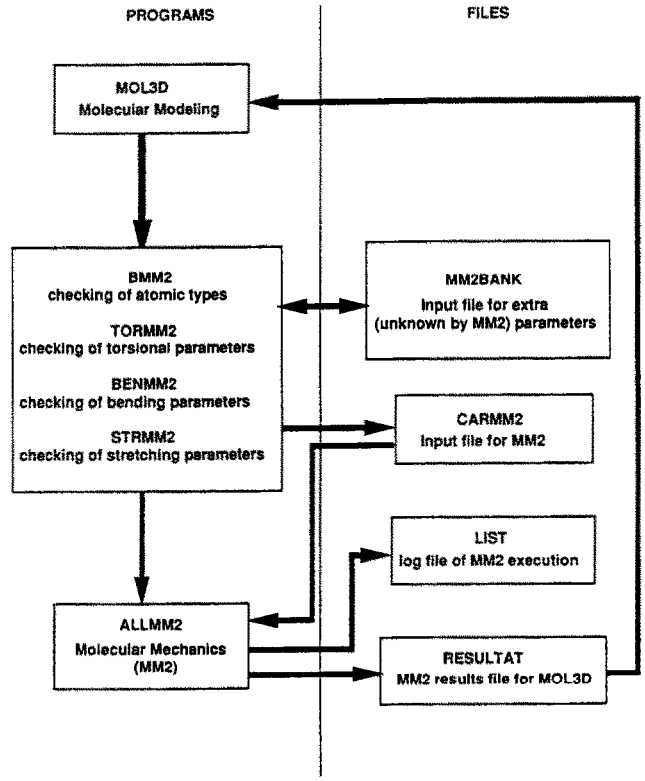
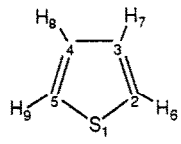


Figure 2. Flowchart of the MOL3D-MM2 interface.




```
CHECKING TORSIONAL PARAMETERS...
    Torsional parameters are missing for atoms
N°  3  2  1  5  -- TYPE  2  2  42  2
    Input V1   V2   V3 :   0  2.3  0

    Torsional parameters are missing for atoms
N°  6  2  1  5  -- TYPE  5  2  42  2
    Input V1   V2   V3 :   0  2.3  0

    Torsional parameters are missing for atoms
N°  4  3  2  1  -- TYPE  2  2  2  42
    Input V1   V2   V3 :   0  15  0

    Torsional parameters are missing for atoms
N°  7  3  2  1  -- TYPE  5  2  2  42
    Input V1   V2   V3 :   0  15  0

CHECKING STRETCHING PARAMETERS...
    Stetching parameters are missing for atoms
N°  1  2  -- TYPE  2  42
    Input KS, L1,L2 and dipole moment : 9  43  1.467  0  0.4

CHECKING BENDING PARAMETERS...
    Bending parameters are missing for atoms
N°  2  1  5  -- TYPE  2  42  2
    Input force constant and angle :   0.68   98.5

    Bending parameters are missing for atoms
N°  3  2  1  -- TYPE  2  42  2
    Input force constant and angle :   0.38   119.0

    Bending parameters are missing for atoms
N°  6  2  1  -- TYPE  5  2  42
    Input force constant and angle :   0.4   120.0
```

Figure 3. Example of an interactive input of missing force-field parameters (thiophene molecule).

Molecular Dynamics Module

Molecular dynamics is used widely to perform molecular simulations and to calculate energies in relation to observed experimental properties. This method completes stochastic (Metropolis sampling) or systematic (grid scanning) searches already available in MOL3D. Several well-known computer programs are available: AMBER,[11] CHARMm,[15] BIOGROMOS,[16] etc. We have chosen to use MOL3D with AMBER, the related interface integrating the software at three different levels (Figure 4a):

(1) Preparation of all the necessary data to make the MOL3D and AMBER data structures compatible. This was achieved in the following way:
   (a) For peptides or nucleic acids, only sequences of building blocks are needed. Then, as both programs use amino acid or nucleic acid dictionaries that present similar geometries and charges, there are no peculiar problems in the data preparation.
   (b) As concerns organic molecules, a program has been written to translate MOL3D molecular topology, atom types, etc. into AMBER parameters. A given molecule is therefore "seen" by AMBER as a single entity at the PREP level. The EDIT program was modified to allow the molecule to present all the necessary features needed in MOL3D.

(2) Submission of batch jobs performing the AMBER minimization, normal modes, and molecular dynamics calculations.

(3) Retrieval and analysis of the batch job results.

Analysis modules have been added to obtain graphical representations of several quantities versus time (e.g., distances, angles, surfaces, and volumes), to perform a dynamics animation from the coordinates files obtained in this way, etc. (Figure 4b). Such graphical tools are useful for viewing conformational changes and to check possible anomalous behaviors.

Quantum mechanics calculations

Molecular mechanics is not accurate enough to handle several important problems, such as electronic delocalization, charge transfer effects, or reactivity. Therefore, as the determination of reasonable atomic charges is the basis of any molecular modeling study, an interface to quantum chemistry packages is a necessary tool in any modeling software. AMPAC/MOPAC (see 17 and references therein) is easily available (QCPE 455/539) and contains the most widely used semiempirical molecular orbitals methods—MNDO, AM1, MINDO/3 and PM1—with several interesting options. The MOPAC menu of MOL3D (Figure 5) depicts all the possible options of this software, prepares all the necessary data, submits the jobs, gets the results and updates the atomic positions and charges of the current molecule for future use. The DENSITY menu integrates the drawing of atomic densities, using the corresponding QCPE program (QCPE 492/540). If more accurate calculations are needed, MOL3D also proposes a simple interface to the GAUSSIAN 8*x* series (see for instance 18 and references therein) of *ab initio* programs (QCPE 446), transferring only coordinates in that case.

Connection to the Cambridge Crystallographic Data Base

X-ray crystallography is one of the most powerful sources of experimental information for molecular modelers. Rapid advances in the structural determination of small organic molecules make about 70,000 crystal structures available in the Cambridge Crystallographic Data Base,[19] the so-called CCDB. This database is an established research tool in sorting and analyzing known molecular geometries. It is commonly used in conjuction with both bibliographic and connectivity files, which are queried to decide which dataset should be retrieved for a particular problem.

To make such possibilities easily available to the MOL3D user, an integrated interface is proposed in the package, that allows the investigation and retrieval of molecules and fragments by alphanumeric queries, or by substructure searches. The recovered structure thus can be stored in the internal MOL3D fragment database for future use, for example to

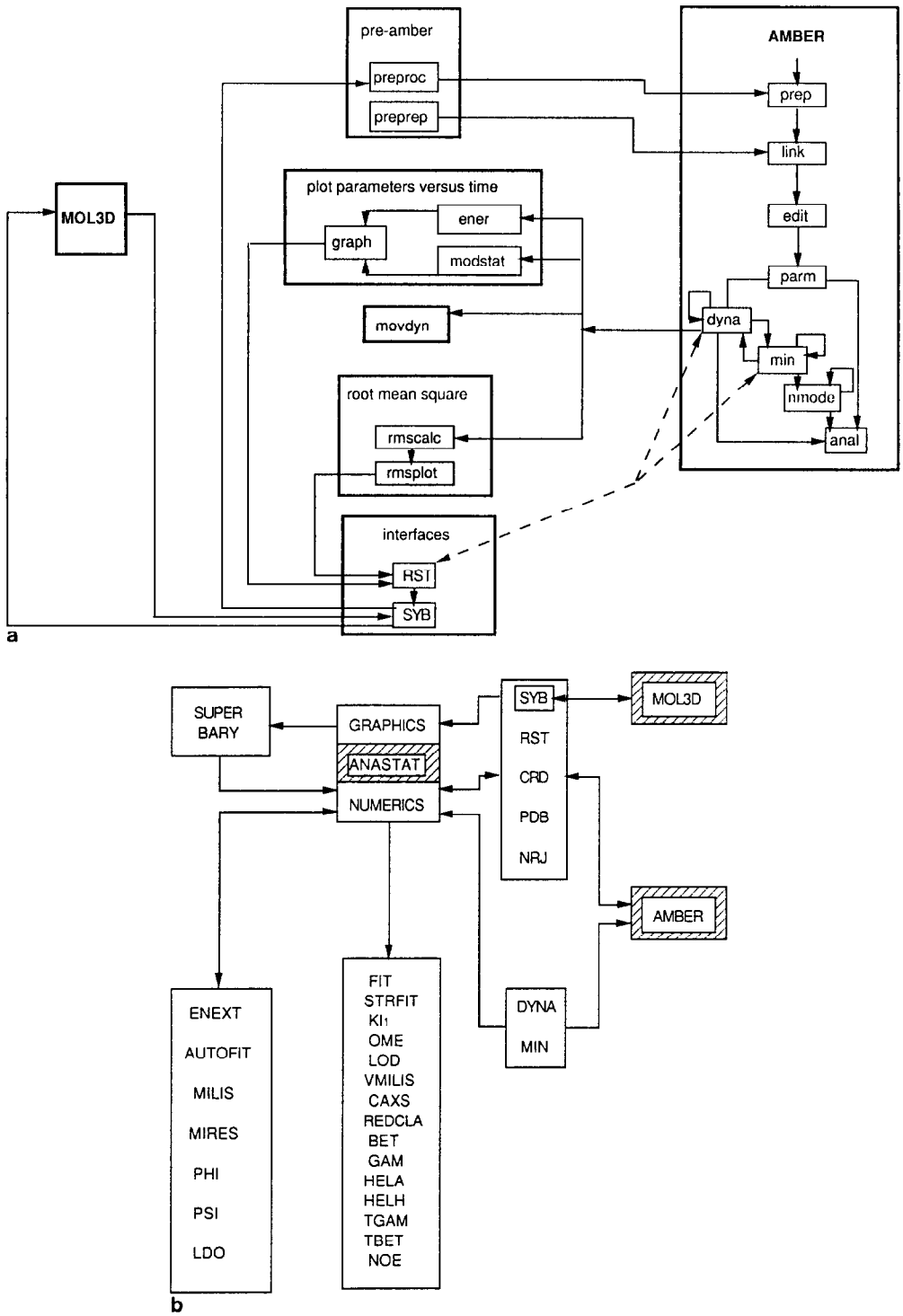


Figure 4. **a,** *Flowchart of the interactions between MOL3D and AMBER;* **b,** *Flowchart of the analysis package.*

build more complex molecules or to compare calculated and crystallographic conformations.

The CCDB translator reads the data stored in a sequential file (written in CCDB format), which encloses information about the topological connectivity and the spatial coordinates of the atoms—with respect to an orthonormal reference. This sequential input file results from inquiries made in CCDB.

DATABASE SEARCHING AND RETRIEVAL

The Graph Manager

Concerning the problem of building an internal MOL3D database, which should contain chemicals and molecular samples, as well as a managing system, it seems logical to start the job with a two-dimensional (2D) molecular graphs base manager, called GAGE. However, in order to further

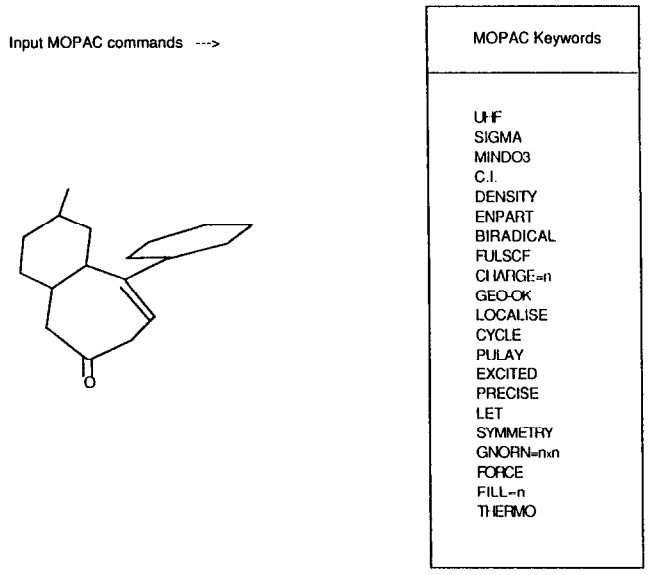


Figure 5. AMPAC/MOPAC menu of MOL3D.

install a SAR-oriented database management module, some other essential features have been added to this basic level:

(1) The possibility of storing both 2D or 3D molecular fragments
(2) A user-friendly interface (graphical display and manipulations), with especially a more detailed menu for graphical works on 3D molecules
(3) A sample manager, and an extra module involving the molecular structures.

As each sample needs specific information (e.g., biological activity, chemical and geometrical features), the molecules and the sample databases are strictly independent one of the other. Thus, the selection of a given sample of compounds is at least submitted to the constraint of the 3D information accuracy. The notion of 3D distances between assumed active sites represents, indeed, one of the underlying concepts to be taken into account during a compound-receptor site interaction. Such a constraint has lead us to prefer using directly 3D molecular entries for the selection steps. The graph manager therefore uses as an input the interface to CCDB described above.

Samples management gives access to a secondary menu through which the molecular samples database can be updated (creation/removal/renaming of a sample, listing of a specific sample content, or deletion of the whole samples database), while the molecules database is not modified. Proceeding to the formation of a sample deserves some explanation. To us, a sample contains only 3D molecular structures. It is generated whenever the user transmits to the system the following types of selection criteria:

- Biological activity (a qualitative coding, such as ANXIO for anxiolytics) of the selected compounds of the future sample. This criterion is compulsory.
- The presence (or absence) of specified atom types, or templates, in the chosen structures.
- Specified 3D distances or distance ranges between two given atoms in the compounds examined.

The last two criteria remain facultative, since they will permit a skilled or perspicacious user to confirm (or infirm) his or her intuition concerning the structural origin of the studied biological activity.

Example: beta-carbolines

Figures 6 and 7 depict an example of a study made using our graph manager. In a first step, all the structures related to the keyword ANALGE (for "analgesics") were extracted from CCDB and stored in memory. A cleaning up was then done, which consisted in throwing away poorly defined or incomplete structues and eleminating solvent molecules counter-ions, etc., only 71 different molecules were kept in memory. From this samples database, we extracted the subset of molecules containing the beta-carboline template (see Figure 6). Only 5 molecules were then selected (most of the others are benzodiazepines) that one can see in Figure 7. Their names in CCDB are, respectively, *bimsar, biprib, citrig01, citrom10,* and *etcrbl.* These molecules can be fitted to each other and thus discriminated according to any physical property or 3D requirement, by using the relational database. (In Figure 8 the first two requirements of the fit concern both nitrogens of the beta-carboline templates, and the third requirement concerns either the exocyclic or amide nitrogen, or the ester oxygen atom).

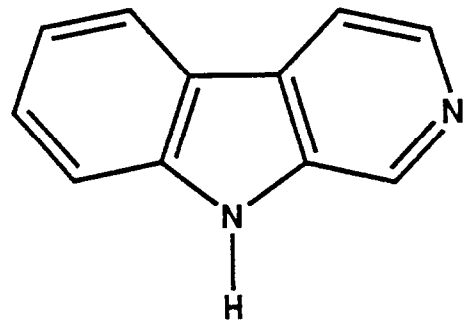


Figure 6. Structural template of beta-carbolines.

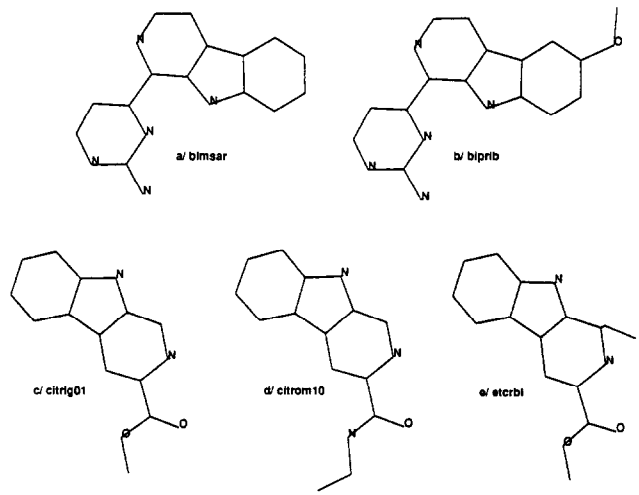


Figure 7. The molecules selected from the 71 initial ones.

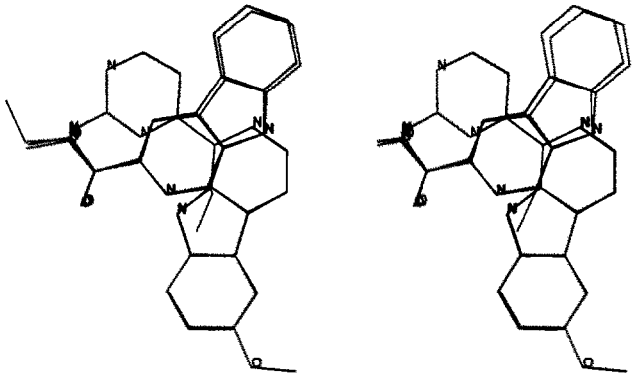


Figure 8. Fit of the 5 beta-carbolines obtained.

CONCLUSION

With this new release, MOL3D becomes a very attractive molecular modeling package, allowing various molecular mechanics and quantum chemistry calculations. Its rigid and flexible docking module, as well as its graphics capabilities, are adapted for drug design, while its interface to CCDB, and its relational samples database with different levels of interrogation make it an efficient tool for SAR-oriented studies.

REFERENCES

1 Pattou, D. and Maigret, B. MOL3D, a modular and interactive program for molecular modeling and conformational analysis: I. basic modules. *J. Mol. Graphics* 1988, **6**, 112–121

2 Cromer, J.M. *Conception d'un logiciel graphique avec visualisation de molécules 3D pour la CAO en chimie organique*. Thesis, IRMA-University Louis Pasteur, Strasbourg, 1985

3 Huron, M.J., and Claverie, P. Practical improvements for the calculation of intermolecular energies. *Chem. Phys. Lett.* 1969, **4**, 429–434

4 Huron, M.J., and Claverie, P. Calculation of the interaction energy of one molecule with its whole surrounding. I. Method and application to pure nonpolar compounds. *J. Phys. Chem.* 1972, **76**, 2123–2133

5 Caillet, J. and Claverie, P. Theoretical evaluation of the intermolecular interaction energy of a crystal: application to the analysis of crystal geometry. *Acta Crystallogr. Sect. A* 1975, **31**, 448–460

6 Caillet, J., Claverie, P. and Pullman, B. On the conformational varieties of adrenaline; the free molecule and the molecule in the crystal. *Acta Crystallogr. Sect. B* 1976, **32**, 2740–2745

7 Gresh, N., Claverie, P. and Pullman, A. Theoretical studies of molecular conformation. Derivation of an additive procedure for the computation of intramolecular interaction energies. Comparison with *ab initio* SCF computations. *Theoret. Chim. Acta (Berl.)* 1984, **66**, 1–20

8 Haarhof, P.C. and Buys, J.D. A new method for the optimization of a nonlinear function subject to nonlinear constraints. *Comput. J.* 1970, **13**, 178–184

9 Metropolis, N.A., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 1953, **21**, 1087–1092

10 Claverie, P., Daudey, J.P., Langlet, J., Pullman, B., Piazzola, D. and Huron, M.J. Studies of solvent effects: 1. discrete, continuum, and discrete-continuum models and their comparison for some simple cases: ammonium(1+) ion, methanol, and substituted ammonium(1 +) ion. *J. Phys. Chem.* 1978, **82**, 405–418

11 Weiner, P.K. and Kollman, P.A. AMBER: assisted model building with energy refinement. A general program for model molecules and their interactions. *J. Comput. Chem.* 1981, **2**, 287–299

12 Pattou, D., Maigret, B., Fourné-Zaluski, M.C., and Roques, B.P. Computational analysis of conformational behavior of cholecystokinin fragments: I. CCK4, CCK5, CCK6 and CCK7 molecules. *Int. J. Peptide Protein Res.* 1991, **37**, 440–450

13 Burkert, U. and Allinger, N.L. *Molecular Mechanics* American Chemical Society, Washington, D.C., 1982

14 Lii, J.H., Gallion, S., Bender, C., Wikstroem, H., Allinger, N.L., Flurchick, K.M., and Teeter, M.M. Molecular Mechanics (MM2) calculations on peptides and on the protein Crambin using the Cyber 205. *J. Comput. Chem.* 1989, **10**, 503–513

15 Brooks, B.R., Bruccoleri, R.E. Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculation. *J. Comput. Chem.* 1983, **4**, 187–217

16 Van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J., Postma, J.P.M. Computer simulation of the dynamics of hydrated protein crystals and its comparison with x-ray data. *Proc. Natl. Acad. Sci. USA* 1983, **80**, 4315–4319

17 Stewart, J.J.P. MOPAC, a semi-empirical molecular orbital program. *J. Comput.-Aided Mol. Design* 1990, **4**, 1–103

18 Clark, T. *A Handbook of Computational Chemistry*. Wiley, New York, 1985, pp. 233–317

19 Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.R.S. and Watson, D.G. The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. *Acta Crystallogr. sect. B* 1979, **35**, 2331–2339