Pergamon

Tetrahedron Vol. 51, No. 24, pp. 6789-6796, 1995

0040-4020/95 $9.50+0.00

0040-4020(95)00312-6

Searching Conformational Space with Molecular Mechanics Method. Dihedral Angle Random Increment Search

Ljiljana Došen-Mićović

Faculty of Chemistry, University of Belgrade, Studentski trg 16, P.O.Box 550, 11000 Belgrade, Yugoslavia, and
Institute of Chemistry, Technology and Metallurgy (IHTM) - Centre of Chemistry, University of Belgrade

Abstract. A computational method for finding different molecular conformations of organic molecules is described. The method is based on Allinger's MM2 (MMP2) force-field program where DRIVER subroutine was supplemented by Monte Carlo type routine for generation of new conformations by random variation of the randomly selected torsional angles. The method works in two directions: conformational search (using torsional or energetic criteria) providing conformational geometries, energies and molar fractions, and search for global minimum using simulated annealing procedure. The use of "temperature shaking" procedure facilitates escape from deep local minima. The method works for open chain as well as for cyclic molecules of any size acceptable by MM2 (MMP2) program. It has been tested by searching conformational space for several acyclic and cyclic (medium- and large- ring) molecules.

Introduction

The use of computational methods to simulate properties of the molecules is now widely accepted in most areas of chemistry[1-4]. The method of molecular mechanics has proved to be a powerful tool in studying molecular structure and related properties[3-6]. However the major difficulty in the molecular mechanics study of flexible molecules of any size, and particularly in the study of large and flexible molecules, is the multiple-minimum problem.

The molecular mechanic method is based on the minimization of the empirically defined energy function depending on relative positions of the atoms in the molecule. With increasing size of a molecule the number of local minima in the energy-configuration space increases drastically and the problem of finding the global minimum, and the other low-energy conformations, gains more and more importance. Any standard molecular mechanics program, after entering a trial structure, leads to a nearest energy minimum with a single energy-minimized structure. That structure may not be the global minimum structure and there may be other minima of comparable energy on the conformational surface.

In order to deal with multiple-minimum problem the entire conformational space of the molecule has to be searched. Available approaches to conformational searching include systematic searching by varying internal coordinates[7] and interatomic distances[8]. These methods work so that numbers of trial geometries are generated by systematically varying internal coordinates or interatomic distances. These geometries are starting points for molecular mechanics optimization to nearest minimum. For large and flexible molecules, however, the available conformational space is large and too many structures are generated to be optimized in a reasonable length of time. Alternatives to the full-space systematic search are methods based on random variation of coordinates such as molecular dynamics[8c] or Monte Carlo (MC) searching in Cartesian space[9] and MC searching in dihedral space[10].

Although the MC method of random searching in Cartesian coordinates proved to be efficient in searching conformational space of a series of cycloalkanes[9] , there are strong arguments in favor of conducting conformational searches in torsional angles. The conformational stereoisomers differ primarily in their dihedral (torsional) angles which, compared to the searches in Cartesian space, reduce the size of global search problem and enable selective sampling in the region of conformational space where stable conformations are found. It is therefore to be expected, and on several examples it has been demonstrated[10a], that dihedral-angle conformational searching would be more efficient than searching in Cartesian coordinates.

The torsion-based random search method has previously been used in the study of thermal conformational fluctuations of proteins[11] with force-field reduced to torsional angles, and in the determination of conformer statistical weights and mean values of structure-related NMR parameters of dermenkephalin[12]. The torsion-based method, with MM2 force-field, has been used in global minimum search[10b,13]. The torsion-based method of Chang, Guida and Still[10a] provides not only the global minimum but also all the other stable conformations consistent with predefined energetic and geometrical constrains. The rings were temporarily fragmented, in this method, and ring-closure distance/angles were used as constraints to avoid geometries with unreasonable ring-closure geometries. This method[10a] was claimed to be applicable to the study of organic molecules having up to 14 symmetric cyclic or 10 acyclic or unsymmetrical cyclic torsion angles.

In this paper we describe a dihedral-angle random-variation method which may be used for searching the conformational space of organic molecules of any size acceptable by the MM2 (or MMP2) program. Depending on the way the temperature was treated, this method leads to a global minimum, or to all the populated minima, or to a selected set of low energy minima in the conformational space.

Methods

The method of MC searching in dihedral angle space, applied in this work, proceeds as follows: a starting structure is chosen, the selected variable (torsional angle, or two angles) is randomly altered, geometry is optimized,and the resultant energy and geometry are compared to those found in previous search steps. The resulting structure is stored if new, or rejected if duplicate, and the cycle is repeated using a stored structure as a starting structure. The starting structure is selected with the help of the PCMODEL program[14] which has been used to build the structure from fragments and to roughly optimize its geometry. The PCMODEL optimized structure was subjected to a short MC run to afford several MM2 (or MMP2) optimized geometries of the target molecule. One of these was randomly selected to be a starting structure of a MC run. The starting structure of a new step in an MC run was the last stored structure of the previous steps. The PCMODEL program, as well as the DTMM program[15] were used for visualization of the final structures. Following questions are common to all the random-search methods: how to recognize a new structure, how to be sure that the true global minimum was found, and how to be sure that all conformations were located.

In response to the first question the two criteria described earlier[9b] were in use. One criterion is the energy. The structure is considered to be new if the absolute value of its energy difference, with respect to the energy of any of the previously stored geometries, is equal or larger than some user-specified value. The other criterion consists in dihedral angles. The list of dihedral angles is generated by MM2 (MMP2). The structure is considered to be new, even if the energy criterion is not satisfied, when at least one dihedral angle differs from the value of the corresponding angle, in any of the previously stored geometries, by more than the value allowed by the user.

The second and the third question are more delicate since there is no explicit test to determine whether a random method has located all minima or has found a global minimum. However, several search variables may help in establishing criteria of a good MC run. The first is the number[10a] of MC steps in one MC run. If several different MC runs of different length (i.e. different number of MC steps in a MC run) yield the same set of final conformers, this is a solid evidence of search convergence. Duplication rate is another criterion[10a],the large number of times one conformation is repeated in a MC run represents an indication of good convergence, too. The temperature is another search variable that may be used to establish convergence. Searches at different temperatures, used in this work, which yield the same global minimum, or the same set of final conformations in a low energy region, may be considered convergent with high probability.

Program

The MC-MM2 (Monte Carlo MM2) program presented here uses the subroutines from MM2(MMP2) program[16]. It can be used for global minimum search as well as for the search of conformational space. The only difference is in the manner the temperature is treated in the program.

The program starts with MM2(MMP2) optimized coordinates supplied by the user. Using dihedral driver routine, the DRIVER of MM2 (MMP2) program randomly selects a dihedral angle (or two dihedral angles for cyclic structures), and then the fragment is rotated a random number of degrees, ranging from -120 to 120 degrees for open chain molecules, and -90 to 90 degrees for cyclic molecules. An option is available which allows user specified dihedral angles to be exempt from being perturbed. In the next step open chain molecules and cyclic molecules were treated differently. After the dihedral angle perturbation (crude rotation) an acyclic molecule has been subjected to energy minimization (by corresponding MM2 (MMP2) subroutines) to the nearest energy minimum, the minimum energy was compared to the energy of the previous step. If $\Delta E \leq 0$, the step was accepted, and the new structure was used as a starting point for the next step, provided it passed the energy and dihedral angle tests. In the case $\Delta E > 0$ the probability that the structure is accepted is $P=\exp(-\Delta E/RT)$. A random number (FFACT) is generated between 0 and 1 and compared to P. If $P \geq \mathrm{FFACT}$, i.e. if ΔE is sufficiently small, the structure is accepted. Otherwise the previous structure is restored.

For cyclic molecules the gradual rotation for energy minimization scheme of DRIVER subroutine in MM2 (MMP2) was adapted and used to generate a new structure which was then subjected to energy minimization. The ΔE was calculated and compared to P as described above. This scheme may be used for open chain molecules as well, and for molecules of any size acceptable by the MM2 (MMP2) program.

When looking for a global minimum conformation the temperature may be treated as a variable during one MC run. In that case T is reduced by a user-specified value after each user-specified number of steps. This way the probability acceptance of high energy conformations is continually reduced until a certain conformation is frozen out.

When in conformational space option, the use of low temperature enables conformational search in low energy regions . If however the temperature is kept high enough to make the probability P equal (or nearly equal) to 1, every conformation is accepted and the conformational space is searched completely. In order to prevent the program to become trapped in some local minimum a "temperature shaking" procedure may be used. Two options are available which consist in raising the temperature from room temperature value to a very high value, (1000000 K for example) and retaining that value for the next 20 steps. This way a new, high energy local minimum was usually found which served as the starting point of a search in a new conformational region. The temperature was raised either every 500 steps, or alternatively only when the program was trapped in a local minimum i.e. when it kept returning to the same minimum more than 50 times successively.

Occasionally the program ends up in local minima of high energy due to abnormal bond lengths or angles. In order to avoid these unreasonable geometries, molecular energy and stretching energy are compared to some user specified values and if higher, the structure is discarded. Care was taken, as well, to maintain chirality and olefin configuration during the course of the search. As mentioned earlier in the text in order to accept some structure as new, bothe the energy criterion and the dihedral angle criterion must be fulfilled and the search may be done employing energy criterion or both.

In order to check for complete optimization, and to test whether the optimized geometries are true minima or saddle points, refinement of a minimum energy structure in each step was done, by a small perturbation of the atomic coordinates followed by energy minimization. The refined structure was accepted if its energy was lower than the energy of the unrefined structure. All calculations described below were done with refinement.

In a search for a global minimum the program may be made more efficient by running several short runs using minimum energy structure of one run as a starting structure for a new MC run. The MM2 program allows one or two dihedral angles to be varied simultaneously. While this small number of varied torsions can lock a search in a vicinity of some local minimum, searching by varying a large number of dihedral angles often increases the number of undesired high-energy conformations found[10a]. Instead of increasing a number of dihedral angles varied in each MC step, we used, in case of large compounds, several MC runs with starting geometries found in mutually remote regions of conformational space. Alternatively we used "temperature shaking" as described above.

Results and Discussion

In order to demonstrate the MC-MM2 method we report here the results of conformational and global minimum searches done on several compounds of different size. Calculations were performed on an IBM 3090 computer bound through the control station 3070 to an IBM PC 386. Energy refinements were carried for each structure with maximum coordinate increment of 0.05 A, and the other user-defined parameters have been varied.

Ten independent runs with 100 steps have been done for butane at high temperature (9000 K). Three conformations were found by using the dihedral angle criterion (with dihedral angle tolerance equal 6°), the anti and both gauche forms. Ordinarily, 28 steps were needed in order to locate all three conformations of n-butane, and 53 steps were needed in a longest run.

At most 161 steps (73 steps on an average) in twelve independent runs were needed to locate eight cyclohexane conformations (two chairs and six twist-boat conformations). The dihedral angle criterion was 6° ,and the temperature used was 9000 K. Searching by the energy criterion yielded single chair and single twist boat forms.

Several independent runs (less than 500 steps), with dihedral angle criterion ranging from 12° to 22°, and the energy tolerance of 0.02 kcal/mole were used in the conformational search of cycloheptane. They afforded all the 14 twist-chair conformations (energy equal 14.31 kcal/mole) and different number of boat/twist-boat forms in each run. The number of boat/twist-boat conformations ranged between 25 (dihedral angle criterion equal to 12°) and 12 conformations (dihedral angle criterion equal to 22°). This is in accord with the very small energy difference between the boat and the twist-boat forms (17.46-17.47 kcal/mole) and with their free interconversion by pseudorotation.

Due to the great flexibility and a large number of conformations only the energy criterion (energy tolerance equal to 0.01 kcal/mole) was used in the conformational search of cyclooctane. Fourteen independent runs were done. The same five conformations were obtained in each case, one more than reported[9b] in Cartesian coordinate search. These five conformations are: boat chair, twist chair chair, twist boat chair, twist boat and chair chair with energies : 19.41, 20.38, 21.07,22.53 and 20.56 kcal/mole, respectively. About 90 steps were needed on the average (149 steps at the most) to locate these five conformations. Much longer runs, about 400 steps and longer, afforded occasionally boat-boat conformation with energy equal 22.95 kcal/mole, and crown conformation with energy equal 20.57 kcal/mole, indicating that these two conformations may be saddle points, as suggested earlier[9b]. The high energy forms (boat, chair and twist chair) were not found.

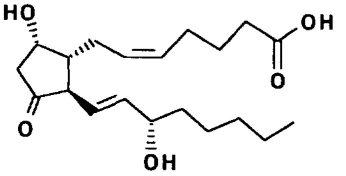
Four independent runs, 1500 steps each, were done for n-octane at each of the two temperatures (9000 K and 298 K) to illustrate the role of temperature in conformational searching. Only the energy criterion was applied with energy tolerance of 0.01 kcal/mole. The same four conformations were used as starting structures at both temperatures. At 9000 K the total number of conformations located in the four runs was: 79,80,80, and 77 respectively, and the number of conformations with the energy below 2.5 kcal/mole, relative to the global minimum conformations, was 19,18,19, and 19 respectively; 18 conformations were found by systematic search method[10a], in this energy region. The complete set of low energy conformations (19) was located after 1051 steps, on the average. The global minimum was found after 146 steps, ordinarily. Runs at even higher temperature (1000000 K) did not afford any new low energy conformation. Much longer runs (3000 steps), at the temperature of 9000 K, did not locate any new low energy conformation either.

At 298 K the total number of conformations found in four independent runs was 26,43,23, and 35 respectively, and the number of conformations found within 2.5 kcal/mole above the global minimum was 16,17,15 and 18 respectively. The last one in the set of low energy conformations found was located ordinarily after 1120 steps, and the global minimum was found after 66 steps, on the average.

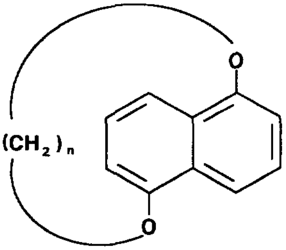
In each run, at both temperatures, the same global minimum was found (MM2 energy equal 4.75 kcal/mole) corresponding to the completely extended structure.

The lower total number of conformations found at the temperature of 298 K is a consequence of a temperature constraint of the search to low energy regions of the conformational space. The time was not spent on less important high energy regions. However a search at low temperature may be locked in global (or some low local) minimum more easily than the high temperature search. Therefore additional searches at higher temperatures, or more extended runs, should be used to ensure search convergence in a low energy region. The search for global minimum was almost twice more efficient at low temperature, ordinarily 66 steps were needed to locate the global minimum at the temperature of 298 K, compared to 146 steps at the high temperature.

In order to illustrate how the MC-MM2 method works with large molecules the global minimum search calculations on prostaglandin D2 molecule (1), and one on dioxa-(1,5)-naphthalenophanes, (n=14) (2), are reported below.



1



2

The starting geometry of (1) was the one with extended side chains and the MM2 calculated energy equal 31.52 kcal/mole, Fig.1.

Some modifications have been introduced to the calculational method in order to speed the search for global minimum. Instead of doing one long run, several short runs (about 200 steps each) were done. Each new run used the lowest energy conformation of a previous run as the starting conformation. Each new run had the record of all the conformations located in previous runs in order to avoid repeated searches in the same conformational regions. After 788 steps, at 298 K, and 55 conformations found, the search was trapped in a local minimum (Fig.2) with the MM2 energy equal

8.32 kcal/mole, and double hydrogen bond between side chains.

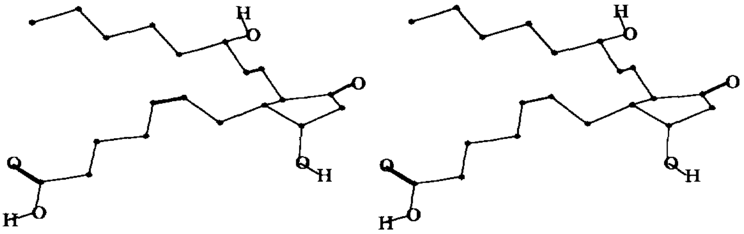


Fig. 1. Stereoview of the starting conformation of (1)

In the next 507 steps at room temperature (298 K), the search was not capable of escaping the local minimum . The temperature was therefore raised, first to 9999 K (422 steps), and then to 1000000 K (196 steps). The highest energy conformation (MM2 calculated energy equal to 20.18 kcal/mol) with partially broken hydrogen bond, was used as the starting conformation in the next low temperature (298 K) run (577 steps) which located the global minimum with energy equal to 6.36 kcal/mole, Fig. 3. The same global minimum conformation was found in another independent run (760 steps) starting with a different high energy conformation as the starting structure.

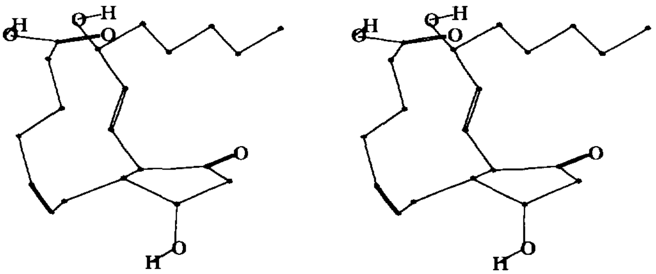


Fig. 2. Stereoview of the local minimum conformation of (1)

The third run, starting from the third arbitrarily chosen high energy conformation, afforded at first the local minimum of the energy equal to 9.35 kcal/mole, characterized by two H-bonds between the COOH group and the ring OH. After two runs of about 200 steps, at 9000 K, and 298 K respectively, the same global minimum was found as above.

The 2600 steps run,at 298 K, starting with molecular geometry of (1) corresponding to 31.52 kcal/mole of conformational energy, and using the "temperature shaking" procedure, afforded the global minimum conformation (energy equal 6.36 kcal/mole) after 750 steps.

Compound (2) was previously studied[17] by MM2 force field using systematic search approach. The [8363] conformation (by Dale's nomenclature[18]) was reported to be the global minimum,

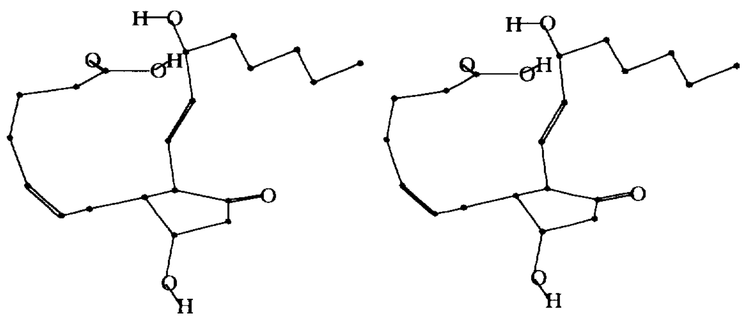


Fig. 3. Stereoview of the global minimum conformation of (1)

followed by the [7463] conformation, 1.4 kcal/mole above the global minimum, then two non-Dale structures, and the [9353] and [6464] conformations. The force-field used in the work described in ref.17 and ref.19 is not identical to the one used here (MM2 with parameters updated in 1989) so the results are not directly comparable. Yet, the same global minimum was found in both studies, corresponding to the [8363] conformation, Fig.4.

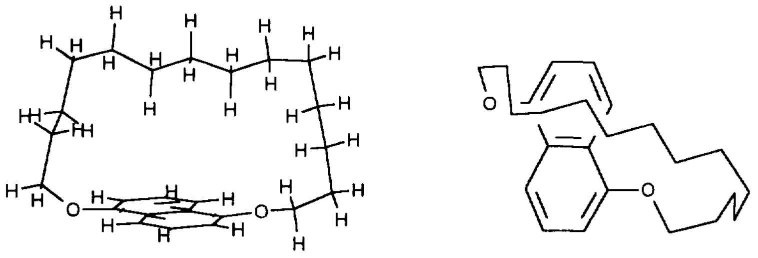


Fig. 4. Two views of the ground-state conformation of (2)

The other low energy conformations (within 2 kcal/mole above the global minimum) of (2) found in our study are [8372], [7463], [8462], [6464], in the order of increasing energy, and the seven non-Dale conformations.

Conclusion

We believe that the MC-MM2 method is a reliable technique for probing conformational space locating global minimum as well as local minima in a desired energy region.

The random-search methods in torsional coordinates are recognized as more efficient than the corresponding methods in Cartesian coordinates. However the primary advantage of Cartesian-based search is that it allows conformational searches of molecules of any size while internal coordinate search methods, which locate local minima besides global minimum, were limited to the molecules of up to 10 unsymmetrical variable torsional angles[10a]. The MC-MM2 method enables torsion coordinate search for molecules of any size acceptable by MM2 program, leading to the global minimum, as well as to the other, local minima in the conformational space.

The use of temperature adds additional flexibility to the program. Besides being used in a search for global minimum by simulated annealing procedure the temperature serves in MC-MM2 program: for additional testing of search convergence (runs at different temperatures should provide the same set of low energy conformations), and for escaping from deep local minima (by "temperature shaking"). Perhaps the most interesting use of temperature is in limiting the search to the desirably low energy regions of conformational space, by adjusting the temperature high while preserving, through the probability factor, the possibility of occasional jumps to high energy regions.

REFERENCES

1. Allison,E.H.; Kollman,P.A. *J.Med.Chem.* 1988, *31*, 1669.
2. Kollman, P.A.; Merz, K.M.Jr. *Acc. Chem. Res.* 1990, 23,246.
3. Papers in *Chem.Rev.* 1993, 93, N°7.
4. Clark, T. *Handbook of Computational Chemistry,* John Wiley and Sons: New York, 1985.
5. Burkert, U.; Allinger, N.L. *Molecular Mechanics,* American Chemical Society: Washington,DC, 1982.
6. Kollman, P. *Ann.Rev.Phys.Chem.* 1987, 38, 303.
7. (a) Wiberg, K.B., Boyd, R.H. *J.Am.Chem.Soc.* 1972, 94, 8426.(b) Lipton, M.; Still, W.C. *J.Comput.Chem.*, 1988, 9, 343.(c) Bruccoleri, R.; Karplus, M. *Biopolymers* 1987,26, 137.
8. (a) Crippen, G. *Distance Geometry and Conformational Calculations,* Research Studies Press, Wiley: New York, 1981.(b) Weiner, P.K.; Profeta, S.; Wipff, G.; Havel,T.; Kuntz, I.D.;Langridge, R.; Kollman, P.A. *Tetrahedron* 1983, 39, 1113.c) McCammon, J.A.; Harvey, S.C. *Dynamics of Proteins and Nucleic Acids,* Cambridge University Press, New York, 1987.
9. (a) Saunders, M. *J.Am.Chem.Soc.* 1987, 109, 3150. (b) Ferguson,D.M.; Raber, D.J. *J.Am.Chem.Soc.* 1989, 111, 4371.
10. (a) Chang, G.; Guida, W.C.; Still, W.C . *J.Am.Chem.Soc.* 1989, 111, 4379. (b) Wilson, S.R.; Moskowitz, J.W.; Schmidt, K.E. *Tetrahedron Lett.* 1988, 4373.
11. Noguti, T.; Go, N. *Biopolymers,* 1985, 24, 527.
12. Nikiforovich, G.V.; Prakash, O.; Gehrig, C.A., Hruby, V.J. *J.Am.Chem.Soc.* 1993, 115, 3399.
13. Li, Z.; Scheraga, H.A. *Proc.Natl.Acad.Sci. USA* 1987, 84, 6611.
14. Serena Software, Box 3076, Bloomington, IN 47402-3076.
15. Crabe, M.J.; Appleyard, J.R. *Desktop Molecular Modeller,* Oxford,University Press, 1990.
16. Alllinger, N.L. *J.Am.Chem.Soc.* 1977, 99,8127.
17. Masek, B.B.; Santarsiero, B.D.; Dougherty, D.A. *J.Am.Chem.Soc.* 1987, 109, 4373.
18. Dale, J. *Top. Stereochem.* 1976, 9, 199.
19. Chang, M.H.; Masek, B.B.; Dougherty, D.A. *J.Am.Chem.Soc.* 1985,107, 1124.

(*Received in UK* 13 *February* 1995; *revised* 12 *April* 1995; *accepted* 21 *April* 1995)