An Internal Coordinate Monte Carlo Method for Searching Conformational Space

George Chang,[†] Wayne C. Guida,[‡] and W. Clark Still*,[†]

Contribution from the Department of Chemistry, Columbia University, New York, New York 10027, and the Research Department, Pharmaceuticals Division, CIBA-GEIGY Corporation, Summit, New Jersey 07901. *Received October 24, 1988*

Abstract: An internal coordinate, random-search method for finding the low-energy conformations of organic molecules is described. The search is biased toward the low-energy regions of conformational space by choosing starting geometries for each step in the conformational search from among previously stored low-energy conformers. Tendencies of the search method to concentrate in certain regions of conformational space at the expense of others are reduced by uniform usage of stored structures as starting geometries and by using varying numbers or torsional rotations in each step. Tests of the procedure's ability to find all low-energy conformations of several acrylic, medium- and large-ring molecules are described. Direct comparisons of the method with Cartesian coordinate random methods and systematic internal coordinate grid searches are carried out with symmetrical and unsymmetrical molecules. The method makes it possible to carry out global conformational searches on symmetrical and unsymmetrical molecules having up to a dozen variable torsion angles.

Two major obstacles hinder the reliable prediction of molecular structures, energies, and properties by molecular modeling. One is the rapid and realistic evaluation of free energy differences in condensed phases. The other is the multiple-minimum problem. This second problem is the topic of this paper and stems from the intrinsic difficulty of locating the populated, low-energy conformers of molecules defined by a complex potential energy hypersurface which may contain an extraordinarily large number of minima. While troublesome, the multiple-minimum problem is one which is necessary to address—molecular modeling studies which omit significantly populated species cannot be expected to yield reasonable results. In this paper, we describe a search method for finding the low-energy conformations of molecules which extends the feasibility of an adequate global search to molecules having a dozen or more variable torsion angles.

A variety of schemes have been proposed and used for searching the conformational space available to small molecules.[1] In virtually all instances, a search begins with the creation of a set of three-dimensional structures taken from diverse points distributed throughout conformational space. These structures serve as starting geometries for optimization (usually by molecular mechanics energy minimization) to nearby minimum-energy conformers. Upon collecting the resulting energy-minimized structures and eliminating any duplicate conformers, a set of unique, low-energy minima is produced.

The various search methods differ primarily in the scheme used to generate the starting geometries. The methods may be classified by the coordinate system used to describe the molecular geometry and by the method used to vary the coordinates. Torsion angles (internal coordinates),[2] Cartesian coordinates (external coordinates),[3] and matrices of internuclear distances (distance geometry)[4] are the coordinate systems which have been used most frequently. To search conformational space in a chosen coordinate system, schemes have employed coordinate variations ranging from systematic to random. The foregoing procedures are commonly used for global conformational searches—i.e. searches designed to sample structures from all regions of conformational space. Starting geometries for conformation searching have also been produced by sampling structures from molecular dynamics simulations.[5] However, such simulations explore conformational space relatively slowly and in the current context are primarily valuable as local conformational-search techniques.

Each of the above approaches to the multiple-minimum problem has its own strengths and weaknesses. Distance geometry methods generate structures ≤1% of the rate of structure generations in internal or external coordinates. Such methods can become competitive with some of the other methods when large numbers of distance constraints are used. Distance geometry does not appear however to be competitive with internal coordinate tree search methods[2a] which give all structures compatible with all operative constraints in an efficient manner. A procedure known as the ellipsoid algorithm is related to distance geometry and has recently been described.[6] Structural variations in Cartesian coordinates provide the most straightforward method to produce new geometries. However, simple atomic movements in Cartesian coordinates make it difficult to favor formation of new starting geometries occupying low-energy regions of conformational space or to generate new geometries requiring large, concerted movements of multiatom substructures.

One can make a strong case for conducting conformational searches in internal coordinates and in torsion angles in particular. First of all, it is obvious that conformational stereoisomers differ primarily in their torsion angles. The bond lengths and angles of different conformers vary relatively little because they are defined by stiff, single-minimum potentials which usually constrain such degrees of freedom to nearly idealized geometries. By considering the least constrained internal degrees of freedom (i.e. torsions), a search problem formally having $3N - 6$ dimensions may be treated as one which has at most N dimensions. The reduction in available space diminishes the scope of the global-

[†] Columbia University.
[‡] CIBA-GEIGY Corp.

(1) Review: Howard, A. E.; Kollman, P. A. *J. Med. Chem.* **1988**, *31*, 1669.
(2) (a) Lipton, M.; Still, W. C. *J. Comp. Chem.* **1988**, *9*, 343. (b) Bruccoleri, R.; Karplus, M. *Biopolymers* **1987**, *26*, 137. (c) Motoc, I.; Dammkoehler, R. A.; Mayer, D.; Labanowski, J. *Quant. Struct.—Act. Relat.* **1986**, *5*, 99. (d) Fukazawa, Y.; Usui, S.; Uchio, Y.; Shiobara, Y. and Kodama, M. *Tetrahedron Lett.* **1986**, *27*, 1825. (e) White, D. N. J.; Kitson, D. H. *J. Mol. Graphics* **1986**, *4*, 112. (f) Smith, G. M.; Weber, D. F. *Biochem. Biophys. Res. Commun.* **1986**, *134*, 907. (g) Dyert, M.; Go, N.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 750. (h) Noguti, T.; Go, N. *Biopolymers* **1985**, *24*, 527. (i) Fine, R. M.; Wang, H.; Shenkin, P. S.; Yarmush, D. L.; Levinthal, C. *Proteins: Struct. Funct. Genet.* **1986**, *1*, 342. (j) Still, W. C.; Galynker, I. *Tetrahedron* **1981**, *37*, 3981.
(3) (a) Saunders, M. *J. Am. Chem. Soc.* **1987**, *109*, 3150. (b) We wish to thank Professor Douglas Raber of the University of South Florida for a preprint describing his random Cartesian search method (the preceding paper in this issue).
(4) (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.
(5) Brooks, B. R. *Proceedings from Supercomputer Research in Chemistry and Chemical Engineering*: Truhlar, D., Jensen, K., Eds.; American Chemical Society: Washington, DC, 1987. Brown, F. K.; Kollman, P. A. *J. Mol. Biol.* **1987**, *198*, 533. Fesik, S. W.; O'Donnell, T. J.; Gampe, R. T.; Olejniczak, E. T. *J. Am. Chem. Soc.* **1986**, *108*, 3165. Brunger, A. T.; Clore, G. M.; Gronenborn, A. M., Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3801.
(6) (a) Loncharich, R. J.; Seward, E.; Ferguson, S. B.; Brown, F. K.; Diederich, F.; Houk, K. N. *J. Org. Chem.* **1988**, *53*, 3479. (b) Billeter, M.; Havel, T. F.; Wuthrich, K. *J. Comput. Chem.* **1987**, *8*, 132. (c) Shor, N. Z. *Cybernetics* **1977**, *12*, 94.

0002-7863/89/1511-4379$01.50/0

search problem from one which expands as R^{3N-6} to one which expands as $\sim R^{\leq N}$, where N is the number of atoms and R is the sampling density along varied coordinates. The use of torsional internal coordinates thus facilitates selective sampling of the low-energy regions of conformational space where stable conformers are found.

Conformational searches are conducted within any coordinate system by selecting points distributed throughout the available space to produce starting geometries for subsequent refinement. We recently described an approach to exploring internal coordinate (torsion angle) space systematically.[2a] The strength of the approach is that it rapidly generates starting geometries which are sampled uniformly from all regions of torsion-angle space. Constraints may be efficiently incorporated into the search because the structures are generated using a tree-search protocol which prunes away entire branches of the search tree whenever a constraint test fails. Thus the more constraints (e.g. nonbonded distances) we supply, the more efficiently the search proceeds. The systematic tree-search approach makes it feasible to search for and find all populated, low-energy conformers of molecules having up to nine cyclic or six acyclic rotatable torsion angles. For more flexible structures, the volume of available conformational space becomes so large that searches are not feasible at a generally adequate torsion-angle resolution ($\leq 60°$). In such cases, too many structures are generated to be optimized by energy minimization in a reasonable length of time.

An alternative to the full-space, systematic search is a search based on random variation of coordinates. Such a random search of Cartesian space was recently described by Saunders.[3a] Saunder's approach operates by generating new starting geometries using random translations (ca. 0.5–3.5 Å) of the atoms of previously found conformers. The method was implemented as a continuous, random-walk process which searches on and on until examination of intermediate results indicates that the search may be terminated. The primary advantage of such a method is that it allows conformational searches of molecules of any size to be conducted conveniently to any desired degree of convergence (in principle). In practice, highly flexible molecules do not give converged results because of the large volume of the available conformational space. The method has been applied successfully to a variety of symmetrical macrocycles. A related method has been recently described by Ferguson and Raber.[3b]

Other methods seek only to find the global minimum. In the torsion-based method of Li and Scheraga,[7] a randomly selected torsion angle is rotated by a randomly selected increment and the resulting structure is optimized by energy minimization. The Metropolis criterion is then applied to determine whether to accept or to reject the new structure as the starting point for the next random torsion-angle rotation. While such global-minimum searches for a single structure may be helpful for qualitative studies, the properties of real systems usually represent averages resulting from a multitude of populated, low-energy structures. For quantitative modeling, it is insufficient to have a single structure, even if it is the global minimum. In cases where only a few conformers contribute to properties, the single, global minimum energy structure may be of value, but to find it the conformational search must be conducted with the ultimate model to be employed. In all but the simplest of molecules, it is unlikely that the single global minimum found by a gas-phase, molecular mechanics search will remain the global minimum once the effects of entropy and solvation are included. Regardless of the system being studied, the set of all populated minima (or at least a representative sampling thereof) is more useful than is a single global minimum. Sets of significantly populated minima are a requirement if quantitative modeling is the objective.

While it is clear that contemporary conformational-searching methods are far from providing a general solution to the multiple-minimum problem, the space available to many small molecules may be searched and all the minima found. High-symmetry molecules such as cycloalkanes are particularly easy conformational problems and rings as large as cyclohexadecane may be adequately searched by existing methods. Conformational searching of unsymmetrical molecules is more difficult because there are many more distinguishable conformations. In this paper we describe an internal coordinate, random-variation method which may be used for searching the conformational space of organic molecules having up to 14 symmetric cyclic or 10 acyclic or unsymmetrical cyclic variable torsion angles and/or configurational space searching in intermolecular translation/rotation coordinates. The search method has been incorporated into our BATCHMIN noninteractive modeling program as a continuous process and allows searches to be conducted on full structures of small organic molecules or on substructures of larger molecules having up to 10 000 atoms. As we will show, the approach generally leads not only to the global minimum but also to all other minima which are consistent with operative energetic and/or geometrical constraints. We believe it to provide the method of choice for searching the conformational space of flexible organic molecules.

Chart I

MCSM, conduct a Monte Carlo single global minimum search[7]
DEMX, select upper energetic bounds for saving structures
MCSS, selects between various starting structure selection options
MCNV, selects the number of coordinates changed during each MC step
MCTS, selects for certain values (e.g. staggered) of final torsional coordinates
RCA4, defines ring-closure bonds and constraints
TORS, selects torsions for variation and sets bounds for increments
MOLS, selects molecules for translation/rotation and sets bounds for movements
CRTS, selects atoms for Cartesian variation and sets bounds for movements
CHIG, saves chirality of asymmetric centers in starting structure as constraints
TORC, set torsion angle constraint
DISC, set distance constraint
COMP, atoms to be compared for deletion of duplicate conformers
ATEQ, defines atoms as equivalent (e.g., the oxygens of a carboxylate)
NSEQ, defines explicit alternative equivalent numbering systems
NSRO, rotates the numbering system for symmetrical cyclic molecules
SEED, reinitialize the random number generator

Methods

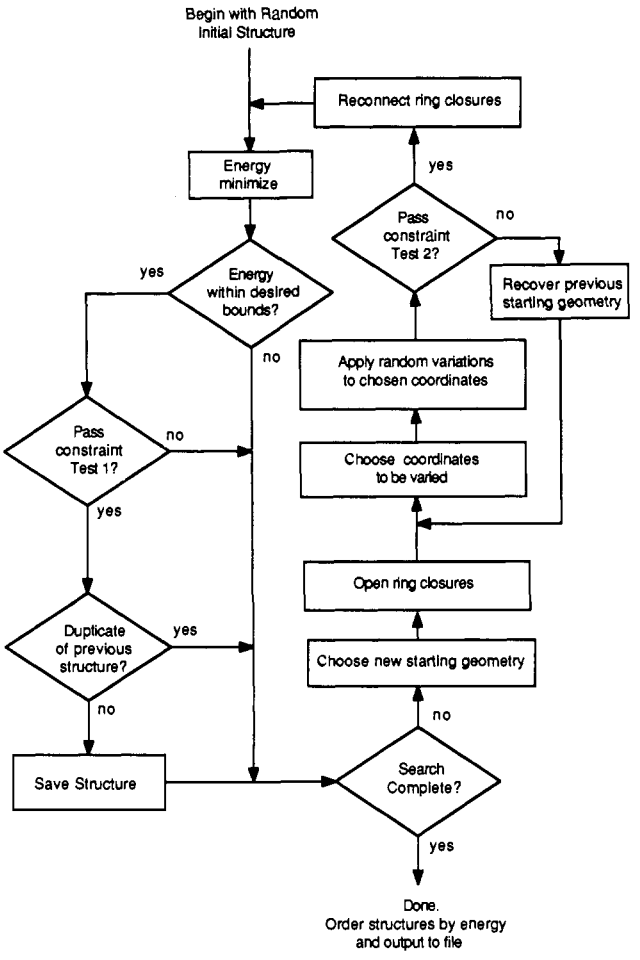
Conducting a random variational conformational search is quite simple. A starting structure is chosen, random variations to selected coordinates are applied, the structure is minimized, and the result is compared with minima found during previous conformational search steps. After the resulting structure has been stored as a new, unique conformer or rejected as a duplicate, the cycle is repeated. We describe one such cycle as a Monte Carlo (MC) step and the entire procedure as a Monte Carlo multiple-minimum (MCMM) search.

In implementing the internal coordinate MCMM search, a number of questions arose concerning the most effective search protocol: How are starting geometries to be chosen? How many coordinates should be varied in each MC step, and which coordinates should they be? How do searches in Cartesian coordinates compare with searches in torsional coordinates? To answer these questions and to evaluate the MCMM method, the search was programmed as a part of our BATCHMIN molecular mechanics program.[8] Options were provided so that the search could be conducted in internal or external coordinates, so that starting geometries could be selected in different ways and so that particular degrees of freedom could be chosen and varied as desired.

(7) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611. See also: Wilson, S. R.; Cui, W.; Moscowitz, J. W.; Schmidt, K. E. *Tetrahedron Lett.* **1988**, *29*, 4373.

(8) BATCHMIN is part of the MacroModel molecular modeling program: Still, W. C.; Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T. *MacroModel V2.0*; Department of Chemistry, Columbia University; New York.

Scheme I



BATCHMIN is controlled by a simple command language of four-character opcodes which precede integer and real-number fields. The opcodes dealing with the MC search are listed in Chart I and touch upon most of the issues with which the search must deal.

These commands provide control of the important variables in the conformational search for multiple minima (MCMM) within a given energetic upper bound relative to the global minimum (DEMX). Thus, they allow MC searches in torsion (TORS), molecular translation/rotation (MOLS), and/or Cartesian coordinates (CRTS). In the internal coordinate searches, rings are temporarily fragmented as described previously[2a] and ring-closure distance/angles are used as constraints (RCA4) to eliminate starting geometries having unreasonable ring-closure geometries. Various procedures for choosing among candidate starting structures for MC steps are selected with the MCSS opcode (see below). Selections concerning the coordinates varied and the new values used during an MC step are controlled by the TORS, CRTS, MOLS, MCNV, and MCTS opcodes. Duplicate conformers are identified by least-squares superimpositions[9] of atoms listed in COMP commands with allowances for alternative atom-numbering systems (NSEQ, ATEQ, NSRO). Finally, since energy minimization of strained starting structures occasionally changes the configuration of chiral centers or alters the geometry of olefins, opcodes (CHIG, TORC) are provided to maintain asymmetric center chirality and olefin geometry during the course of the search.

The algorithm used for the Monte Carlo multiple-minimum (MCMM) search is summarized in the flow diagram shown in Scheme I. The scheme is straightforward, but additional comments are appropriate for two of the operations, the constraint tests and the starting geometry selection. Constraint test 1 is used to eliminate energy-minimized structures (1) whose energies lie outside the selected energetic upper bound relative to the instant global minimum, (2) which have inverted chiral centers, and (3) whose interatomic distances or torsion angles do not match explicit distance or torsion constraints provided by the user. Constraint test 2 is applied to structures prior to energy minimization. This test eliminates starting geometries having poor ring-closure characteristics or high-energy nonbonded contacts (separation <one-fourth of the sum of the van der Waals radii of the atoms involved). For choosing starting geometries, two alternatives were provided: a random walk (the most recently minimized structure is chosen unless it is >100 kJ/mol above the instant global minimum) and a uniform-usage scheme (the previously saved structure which has been used the least is chosen). We also allowed energy to be used as a criterion for choosing starting geometries such that lower energy structures can be selected preferentially.

To investigate the utility of the MCMM search, we used the modified BATCHMIN program to conduct searches on *n*-octane, cyclodecane, and cyclotetradecane using the MM2[10] force field with a 7-Å soft nonbonded cutoff. These molecules were selected as representative of flexible organic acyclics, medium rings, and large rings. The ring closure distance window was defined with the RCA4 command as 0.5-3.5 Å and candidate starting geometries having ring-closure distances outside this range were rejected. Ring closure angle constraints were not used. We specified upper energetic bounds (relative to global minima) for the searches of 10, 25, and 15 kJ/mol respectively to limit the final number of conformers to approximately a dozen unique, low-energy minima for each compound. Energy minimizations on the MC starting geometries were carried out with a sufficient number of successive overrelaxation (SOR) block diagonal Newton Raphson iterations (typically ≤250) to reduce the RMS gradient to <0.05 kJ/Å mol. Structures were defined as duplicates when the residuals from a least-squares atom-by-atom superimposition[9] were <0.25 Å for all non-hydrogen atoms. Final structures were verified to be true minima and not saddle points and other metastable geometries by evaluation of normal-mode vibrational frequencies.

In the general case of a conformational search, neither the global minimum energy conformer nor the number of conformers within any specified energetic range will be known at the outset. Thus the objective of a search is not only to find low-energy minima efficiently but also to provide evidence that low-energy conformers have not been overlooked. While conclusive evidence of search convergence is elusive, empirical search convergence can be established with respect to the number of MC steps, the choice of starting geometry, and other search variables. Thus establishing that multiple, nonidentical searches always yield the same set of final conformers can be taken as evidence of search convergence. The smallest number of times any conformer has been found, i.e. the minimum duplication rate, also provides a useful index of convergence—the larger the minimum duplication rate, the more likely it is that the search has converged.

In preparation for evaluating the effect of the various search parameters, we first established the number of distinct minima within the above energetic bounds as follows. With *n*-octane and cyclodecane, a 60° resolution systematic MULTIC search[2a] found 18 and 14 conformers, respectively, within 10 and 25 kJ/mol of the global minima. No additional conformers were found by any of the MCMM searches described below. In the case of cyclotetradecane, a full-space MULTIC search was too time consuming but multiple MCMM searches found 11 conformers within 15 kJ/mol of the ground state.

We next conducted MCMM searches of the test molecules varying the search parameters discussed above. Standardized runs of 400, 500, and 1000 MC steps for cyclodecane, *n*-octane, and

(9) Kabsh, W. *Acta Crystallogr.* **1976**, *A32*, 944. Kabsh, W. *Acta Crystallogr.* **1978**, *A34*, 827.

(10) We used Allinger's MM2 force field (Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127 for all the calculations on hydrocarbons. For functionalized molecules, the MacroModel implementation of MM2 uses a partial charge treatment for electrostatics and the calculations described employed a distance-dependent ($\epsilon = R$) dielectric.

cyclotetradecane, respectively, were used. For each run, we tabulated the number of duplications of each conformer found and used it to compute the acceptance rate (A_E, the percentage of MC steps yielding structures falling within the given energetic window) and the standard deviation of the duplication rate per 100 acceptances (SD_{100}). These values are associated with efficiency of the search. A_E measures how selectively the search generates conformers within the desired energetic range. A_E = 50% indicates that half of the MC steps yielded conformers in the energetic region of interest. SD_{100} provides an indication of how uniformly the search covers the specified low-energy conformational space. The smaller is its value, the more uniform is the coverage. We also tabulated the minimum duplication rate (D_{min}) and took special note of searches in which conformers were found only once (D_{min} = 1) or were missed entirely (D_{min} = 0). Such low D_{min} values imply poor search convergence.

Results and Discussion

In the paragraphs above, we describe our implementation of a random-search method for exploring the conformational space available to flexible molecules. The method could be used as a full-space search which seeks all minima. However, the time required to complete a search generally increases with the number of structures to be found, and the structures from a full-space search would include many high-energy forms which do not contribute significantly to the properties of the compound.

We have developed a scheme which is designed to speed the conformational search by directing it toward the more populated, low-energy regions of conformational space. The working hypothesis is that low-energy conformers are generally more closely related to each other than to higher energy conformers. If this hypothesis is valid, then by using low-energy structures as starting points for MC steps and by making only limited variations in their 3-dimensional starting geometries, we should retain the geometries of low-energy substructures and favor the formation of new conformers which are low in energy. While such an approach should focus the search on low-energy regions of conformational space, it also provides an undesirable bias against finding new conformations which are very different in geometry from previously discovered ones. For this reason, we use a search protocol which includes contributions from directed searches which attempt to concentrate on selected regions of conformational space and from more random search strategies which explore conformational space less efficiently but in an unbiased way.

In the paragraphs below, we describe how choosing low-energy starting geometries in different ways and varying differing numbers of torsion angles at each MC step affect the outcome of conformational searches. We are particularly concerned with the effectiveness of such search variables in focusing the search on low-energy regions of conformational space and in minimizing tendencies to overlook low-energy conformers.

MC Structure Selections. To test the assumption that use of low-energy starting geometries for MC steps selectively favors the production of low-energy final conformers, we conducted a series of 100-step searches beginning with selected *n*-octane and cyclotetradecane conformers of varying energies. These searches could be described as local searches because each MC step of a given search began with the same starting geometry. At each step, two (*n*-octane) or three (cyclotetradecane) torsion angles were selected at random and rotated by random 0–360° increments, and the resulting structure was energy minimized. Thus in each run, a single starting geometry was used repeatedly to accumulate its progeny. A preliminary examination of the cyclotetradecane results showed that approximately 10% of the MC steps simply gave back the starting conformation. In order that this artifact of the search procedure would not bias the results, conformers identical with starting geometries were excluded in computing both the average energies of the structures produced and the acceptance rates. The results of these local searches are summarized in Table I in kJ/mol relative to the global minimum.

As expected, lower energy starting geometries tended to yield lower energy final conformations. Thus as the energy of the starting geometry increased, both the high-energy end of the range and the average of the energies of the resulting conformers increased. At the same time, the proportion of structures falling within the specified energetic bounds decreased.

Table I. Effect of Starting Structure Energy on Local Conformational Searches[a]

compound	starting-geometry energy[b]	energies[b] of conformations found: range	energies[b] of conformations found: av	A_E[c]
n-octane	0.0	0.0–13.4	5.9	88
	6.7	0.0–27.6	12.6	31
cyclotetradecane	0.0	0.0–54.2	20.1	30
	15.9	4.6–50.2	28.0	8
	26.4	13.4–55.2	28.0	5
	33.9	0.0–69.9	28.5	7
	47.3	13.4–88.3	41.4	4

[a] Results are for 100-step MC local conformational searches beginning with starting geometries having the energies shown. [b] kJ/mol. [c] Acceptance rate (see the text), in percent.

Table II. Effect of Starting Structure Selection Criteria on Global Conformational Searches

molecule	structure selection method	SD_{100}[a]	D_{min}[b]	A_E,[a] %
n-octane	1. simple RW	3.3	4, 1, 0	32
	2. RW-10 kJ	2.4	6, 4, 4	50
	3. RW-10 kJ, weighted	3.1	5, 3, 2	58
	4. UD-10 kJ	2.0	7, 7, 5	52
	5. UD-10 kJ, weighted	4.2	3, 1, 1	70
cyclotetradecane	1. simple RW	6.7	1, 0, 0	8
	2. RW-15 kJ	5.1	6, 3, 1	25
	3. RW-15 kJ, weighted	6.9	4, 3, 0	32
	4. UD-15 kJ	2.8	16, 12, 9	29
	5. UD-15 kJ, weighted	6.7	8, 7, 6	34

[a] Average of three runs using different starting geometries and random-number sequences. [b] Minimum duplication rates for each of three runs (a value of 0 indicates that one or more minima were missed by the search).

Next we tested the effectiveness of using energetic structure selection criteria for choosing starting geometries in our MCMM conformational-searching procedure. We first defined a standard run of 500 and 1000 MC steps for *n*-octane and cyclotetradecane, respectively, starting from randomly selected initial structures which were high in energy relative to the global minimum. For comparison, we conducted simple random walk (RW) searches which did not use energetic structure selections but instead always used the latest geometry for the next MC step, providing it was within 100 kJ/mol of the instant global minimum (method 1). Next we carried out various energetically directed searches as summarized in Table II. These searches differed in the nature of the tests which controlled the choice of structures as starting geometries. Energetically based structure-selection methods 2 and 4 allowed only low-energy structures to serve as starting geometries for MC steps. We define low-energy structures here to be those whose energies lie within the desired energetic upper bounds relative to the global minimum. As noted above, we used energetic ranges of 0–10 and 0–15 kJ/mol for *n*-octane and cyclotetradecane, respectively. Methods 3 and 5 used a similar energetic structure selection criterion but employed an additional weighting function to favor the lower energy structures within the 10 and 15 kJ/mol energetic bounds. The weighted energetic test operated by multiplying the energetic constraint (E_{max}, here 10 or 15 kJ/mol) by a random number (R) between 0.0 and 1.0 and then comparing the product with the energy of the structure being considered as a possible starting geometry. If the candidate structure's relative energy was less than RE_{max}, then it became the new starting geometry. Methods 4 and 5 did not use a random-walk approach but instead selected candidates for starting geometries from among the previously saved conformers based on their previous usage as starting geometries. These usage-directed (UD) structure selections operated by maintaining a tally

of the number of times a saved structure had been used as a starting geometry and always selecting the structure which had been used least. In the event that more than one structure had the same least usage, the lowest energy one of these was selected. The result of a simple UD search was that every structure found during the search was used as a starting geometry for MC steps the same number of times (±1). The UD structure selection method was invented because conformational searches tend to find some structures more frequently than others. RW searches are thus biased toward using these frequently found structures as starting geometries and therefore toward rediscovery of their progeny. The results of these studies are summarized in Table II and are averages of triplicate conformational searches which began with different starting geometries and random-number sequences.

As anticipated by the results of the local conformational searches starting from single conformers having differing energies, energy-directed search methods 2–5 are effective at focusing the search on the desired low-energy regions of conformational space. Thus in comparison to the simple RW search (method 1) in which virtually any conformer produced could serve as a starting structure, the acceptance rates (A_E) rose substantially when only low-energy starting structures were allowed as starting points for subsequent MC steps. Further increases in A_E followed from preferential selection of starting geometries from the low-energy end of the energetic selection window (methods 3 and 5); however, such weighting resulting in a significant increase in the tendency to find some structures repeatedly while neglecting others. This undesirable trend is reflected in the increased duplication standard deviations (SD_{100}) and decreased minimum duplication rates (D_{min}) of methods 3 and 5 in the table. Thus, while it is not difficult to direct a conformational search to favor the production of low-energy structures, it is less easy to do so without overdirecting the search so that the operative constraints disfavor the discovery of certain low-energy structures.

We obtained the best results with the simple uniform usage directed scheme (UD method 4) which selected starting geometries equally from among the low-energy set of previously stored conformers. As shown in Table II, this scheme gave the most uniform coverage of the conformational space within the selected energetic window. Thus it showed the largest D_{min} and the smallest SD_{100} values of all the methods examined.

Since low-energy starting geometries favor the production of low-energy conformers in a MC search, we might expect that beginning with a low-energy initial structure would improve the efficiency of the search. We therefore tested Li and Scheraga's Monte Carlo global minimum searching procedure[7] (implemented here as our MCSM search) for locating a single low-energy structure with which to initiate our standard MCMM search. The MCMM search generally locates the global minimum of small molecules like *n*-octane and cyclodecane in <100 MC steps regardless of starting geometry. We therefore studied the effectiveness of the MCSM and MCMM searches in finding low-energy conformers of the larger molecule cyclohexadecane. We began both the MCSM and MCMM searches with the same randomly selected starting geometry (+42 kJ/mol above the ground state), conducted 200 MC steps, and recorded the energy of the most stable conformer found. In all, 11 runs were conducted with each method. The MCSM procedure turned out to be marginally more effective at locating the global minimum in the limited, 200 step runs. Thus, the MCSM method found the global minimum in five out of the 11 runs whereas the MCMM search found it only three of the runs. For finding either of the two lowest energy conformers, the success/trial ratio was 7/11 for MCSM and 6/11 for MCMM. It appears therefore that, for simple molecules at least, the MCMM search finds low-energy structures very early in the search, and there is little to be gained by conducting a separate search for a low-energy initial structure.

In summary, we found that different ways of selecting starting geometries for MC steps markedly affect the efficiency of our conformational searches of *n*-octane and cyclotetradecane. While the simple random-walk search (method 1, Table II) missed at least one valid conformer in three out of the six trials, selecting starting geometries for MC steps from among the previously discovered low-energy conformers (methods 2 and 4, Table II) provided all known low-energy conformers in 12 out of 12 trials. Coverage of low-energy conformational space was also enhanced by using each of the previously found conformers an equal number of times as a starting geometry. The choice of the beginning starting geometry had little effect on the results of the search. Thus we find that our MC searches are most effective at finding all low-energy conformations when using a range of low-energy structures as starting geometries and using each such geometry uniformly. We therefore select the low-energy, uniform-usage method of starting structure selection (UD method 4) as the method of choice and use it in all of the following studies.

Table III. Dependence of Conformational Searches of Symmetrical Molecules on the Total Number of Torsions Varied

molecule	no. of varied torsions[a]	SD_{100}[b]	D_{min}[c]	A_E[d]
n-octane (0–10 kJ/mol)[e]	1	2.9	0, 3, 3	37
	2	2.7	2, 3, 4	43
	3	2.5	2, 3, 4	43
	4	2.2	3, 3, 4	38
	5	2.1	1, 1, 3	32
	6	2.6	1, 2, 3	29
	2–4	2.5	1, 2, 4	40
	2–5	2.6	1, 2, 3	40
cyclodecane (0–25 kJ/mol)[e]	1	4.9	0, 2, 2	33
	2	3.7	1, 2, 4	61
	3	3.4	5, 8, 9	82
	4	4.0	5, 7, 10	86
	5	3.7	9, 10, 14	90
	6	3.6	9, 9, 11	90
	2–4	3.4	6, 9, 10	77
cyclotetradecane (0–15 kJ/mol)[e]	1	8.6	0, 0, 0	11
	2	5.8	0, 3, 3	19
	3	5.0	0, 5, 6	19
	4	5.1	0, 2, 3	13
	5	5.2	0, 0, 2	10
	6	5.2	0, 2, 2	7
	2–4	5.3	1, 2, 5	18
	4–6	4.9	1, 2, 3	10
	2–6	5.5	0, 2, 2	12

[a] The total number of torsion angles varied in each MC step (the specific torsion angles varied were selected at random). [b] Standard deviation of duplication rate per 100 structures accepted (see the text). [c] Minimum duplication rate for the three independent runs (a value of zero indicates that a search missed one or more low-energy conformers). [d] Acceptance rate (percentage of structures generated which fell within the given energetic window). [e] Energetic range of allowable structures relative to the global minimum.

MC Coordinate Variations. In previously described Monte Carlo conformational searches, different authors have advocated varying different total numbers of coordinates at each MC step. Scheraga reported that the best results with his Monte Carlo minimization method were obtained by varying only one dihedral angle at each step.[7] The rationale behind such a selection is that localized changes in low-energy starting geometries leave large, low-energy substructures intact and are likely therefore to yield new low-energy final geometries. In contrast, the Cartesian method of Saunders operates by varying all atomic positions at each MC step.[3a]

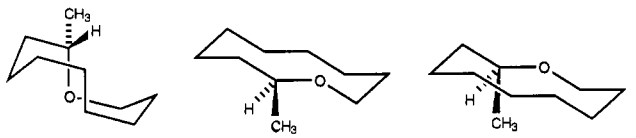
To investigate the effect of the number of torsions changed on the results of our torsional MCMM search, we varied the number of torsional rotations used in each step in a series of 500-, 400- and 1000-step searches of *n*-octane, cyclodecane, and cyclotetradecane, respectively. The searches were conducted as usage-directed searches (UD method 4 above). The results from triplicate runs using the MM2 force field are summarized in Table III. As in previous searches, starting geometries were rediscovered in a significant number of MC steps, especially when small numbers of torsion angles were being varied. Since we wish to use the duplication rate as an index of how well conformational space has been covered, we have not counted such rediscovered starting geometries in the search statistics below. Thus in the

results summarized in following tables, we include data only for conformers differing from the starting geometries used to generate them.

These studies show that the results of the search are indeed sensitive to the number of torsions varied during each MC step. First of all, in cyclotetradecane, a molecule flexible enough to have effectively independent substructures, high acceptance rates (A_E) are associated with small numbers of varied torsions. This result is in line with Scheraga's observations for acyclic peptides, although varying only a single torsion was not effective at searching the conformational space of our cyclic molecules. However, also associated with small numbers of varied torsions are large standard deviations (SD_{100}) and low minimum duplication rates (D_{min}), indicating an undesirable tendency to selectively duplicate certain conformers at the expense of finding others. Alternatively, varying a large number of torsions at each MC step had other undesirable consequences. In particular, the acceptance rate with cyclotetradecane dropped with increasing torsional variation indicating an increase in the proportion of high-energy conformers being found by the search. With cyclodecane, the reduction was not observed because the 25 kJ/mol energetic upper bound includes >90% of the cyclodecane conformers and because a 10-membered ring is too small to be approximated by a collection of independent substructures.

While the details of Table III will vary with the original starting geometry and random number sequence used, the results indicate that there is no best single value for the number of varied torsions in the general case of a search for the low-energy subset of conformational minima. A small number of varied torsions can lock a search into a subspace of the actual potential surface while large numbers of varied torsions deemphasize selection for the desired lower energy structures. With cyclotetradecane, for example, there was no single number of varied torsions which allowed all three of the 1000-step searches to discover all of the low-energy minima. One solution to the problem is to use a range of varied torsion angles. Thus, instead of always varying the same number of torsion angles at each MC step, it is better to choose randomly a number of torsions to be varied from between one and some maximum number of variable torsions.

The results in Table III imply that it is best to search the conformational space of symmetrical structures by varying considerably fewer than the total number of independent torsions. We were concerned however that energy-directed searches of unsymmetrical molecules might require a different treatment. This concern followed from certain properties of conformations of unsymmetrical medium-ring compounds like the 10-membered cyclic ether conformers shown below. Consider the question:



A B C

given conformation A, how many torsions need we vary to convert it into B in a single step? Comparison of torsion angles shows that B differs qualitatively from A by 8 of the 10 total torsions and could be generated from A by varying five to six independent torsions (depending on the location of the ring-closure bond). Thus a low-energy search starting with A in which only a maximum of four torsions were varied would miss B if no other structures were involved. On the other hand, the relatively high-energy conformer C differs from both A and B by only four torsions. Therefore with a relaxed energetic constraint which would allow C to serve as an intermediate, B could be found from A in two steps by varying as few as three to four torsion angles in each step. In an analogous situation with a fully symmetrical structure like cyclodecane, it would be unnecessary to find B explicitly since it would differ only by its numbering system from the otherwise identical conformer C. Therefore varying small numbers of torsion angles may be effective for searching structures of high symmetry but the more common case of unsymmetrical molecules requires substantially more torsional variation. Numbers of torsions varied per MC step as large as the total number of independent torsions may be appropriate when the starting geometry selection is highly constrained.

Table IV. Dependence of Conformational Searches of an Unsymmetrical 10-Membered Cyclic Ether on the Total Number of Torsions Varied

energetic constraint	no. of varied torsions[a]	SD_{100}[b]	D_{min}[c]	A_E[d]
(0–25 kJ/mol)[e]	2–4	3.2	0, 0, 1	17
(total of 28 minima)[f]	3–4	2.2	0, 2, 3	17
	4–5	2.0	2, 3, 3	14
	5–6	2.4	1, 2, 2	13
	2–6	2.2	0, 1, 1	15
(0–10 kJ/mol)[e]	2–4	36	0, 0, 0	1.1
(total of 4 minima)[f]	3–4	30	0, 0, 1	0.9
	4–5	16	1, 2, 2	0.8
	5–6	5	2, 2, 3	0.7
	2–6	30	1, 1, 2	0.8

[a] The range of the total number of torsion angles varied (the specific torsion angles varied were selected at random). [b] Standard deviation of duplication rate per 100 structures accepted (see the text). [c] Minimum duplication rate for the three independent runs (a value of zero indicates that a search missed one or more known low energy conformers). [d] Acceptance rate (percentage of structures generated which fell within the given energetic window). [e] Energetic range of allowable structures relative to the global minimum. [f] Total number of distinguishable conformers found within the indicated energetic window.

In Table IV, we summarize results of conformational searches of the unsymmetrical 10-membered cyclic ether shown in the diagram above. Each entry contains data from triplicate 2000-step MCMM searches using the AMBER united atom force field.

The results in the table show that the unsymmetrical test structure is indeed best searched with greater numbers of variable torsions than are necessary for a symmetrical system of comparable size. Whereas two to four variable torsions were adequate for searching cyclodecane, two to four torsion angle searches of the unsymmetrical cyclic ether above generally missed at least one of the known minima with 2000 MC step searches. The most effective searches required four to six torsional variations. With the narrow, 10 kJ/mol energetic range (occupied by only four conformers), the best results were obtained by varying five to six torsion angles in each MC step. Thus with these unsymmetrical cyclic systems, the maximum number of varied torsions per MC step should approach the full seven independent torsions of the 10-membered ring.

In summary, the results above sustain the expectation that small structural changes in low-energy conformers by in large favor the formation of new low-energy minima. Hence the average energy of conformers produced by MC steps increases with the number of torsions varied in each step. On the other hand, small numbers of varying torsions are also associated with poor coverage of conformational space especially with unsymmetrical structures and restrictive energetic bounds. Unsymmetrical molecules require special care for two reasons. First, they have more distinct minima than the corresponding symmetrical structures and thus require more extensive searching to give converged results. Second, interconversions between conformers of unsymmetrical molecules often require larger numbers of simultaneously varied torsions. Restrictive energetic bounds exacerbate the interconversion problem by providing fewer structures which may serve as intermediates in the pathway between torsionally remote sets of conformers. We find the best compromise for conformational searching of an unsymmetrical molecule having N independent torsions to be a random variation in the number of torsions being changed in each MC step from a minimum of two to a maximum of $N - 1$ torsions. For fully symmetrical structures such as the cycloalkanes, we favor a range of two to $N/3$ variable torsions. The wide range of varied torsions allows MC steps favoring (torsionally) nearby low-energy structures when the number of torsions changed is small and allows more random jumps to remote regions of conformational space when the number is large.

Table V. Comparison of Conformational Searches of Symmetrical and Unsymmetrical Macrocycles in Internal and External Coordinates

molecule	method	no. of minima found[d]	A_E, %
cyclotetradecane	1. UD 2–4[e] Torsions[a]	11, 11, 11	18
	2a. RW 42[f] Cartesian[a]	9, 9, 9	2
	2b. RW 14[f] Cartesian[a]	9, 9, 11	4
	2c. UD 14[f] Cartesian[a]	10, 11, 11	8
11-hydroxy-dodecanoic acid lactone	1. UD 2–7[e] Torsions[b]	359, 361, 364	58
	2c. UD 14[f] Cartesian[b]	334, 335, 335	39
	3. MULTIC[c]	357	NA

[a] A total of 1000 MC steps, 15 kJ/mol upper bound. [b] A total of 2900 MC steps, 25 kJ/mol upper bound, constrained to s-trans lactone conformations. [c] A total of 2913 starting geometries generated by 60° torsion angle resolution full space grid search. [d] Total number of unique minima found within the indicated energetic range for each of the three runs. [e] Number of torsion angles varied in each MC step. [f] Number of atoms independently translated in each MC step.

Comparison with Other Methods. To compare the MCMM torsional-search method reported above with several other established methods for exploring conformational space, we conducted searches of the symmetrical and unsymmetrical macrocycles cyclotetradecane and 11-hydroxydodecanoic acid lactone. Each Monte Carlo search was carried out in triplicate with different starting points for the random-number generator and the results are summarized in Table V. For our MCMM torsional search, we used varying numbers of rotatable torsions in the usage-directed structure selection mode as described in the paragraphs above. We defined a ring-closure bond between two sp^3 carbons in each molecule and provided a 0.5–3.5 Å ring closure distance constraint. These runs are listed as method 1 in the table and the number of torsions stated gives the range of torsional variations used in each MC step. The method 2 results are from MCMM searches in Cartesian coordinates and the number of atoms moved in each MC step is given in the table. The atomic translations were selected randomly from a range of 0.5–3.0 Å. In the run labeled 2a, all 42 atoms of cyclotetradecane were moved at each step whereas runs 2b and 2c employed the independent movement of carbons only. During the 2b and 2c runs, the hydrogens were translated along with the carbons to which they were bound. Whereas entries 2a and 2b represent random-walk runs, the 2c entries in the table were Cartesian usage-directed runs. Finally, for the unsymmetrical 12-membered lactone, we conducted a 60° full space grid search[2a] using the MULTIC submode of MacroModel.

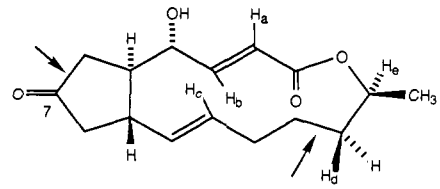
As Table V indicates, the usage-directed, torsional Monte Carlo search (method 1 in Table V) is the most efficient method for the conformational searches conducted. Compared to the corresponding Cartesian searches with the same number of steps, the torsional search gave both a higher acceptance rate, indicating enhanced coverage of the low-energy region of conformational space, and a larger number of discovered unique minima. The 2a method was designed to be as similar as possible to the stochastic search method described by Saunders.[3a] In searches using the 2a method, ~30% of the structures (after energy minimization) incorporated at least one highly distorted (inverted) tetrahedral carbon as a result of the independent translation of carbons and hydrogens. Production of such defective structures is partially responsible for the low acceptance rate and thus the relatively poor efficiency of the 2a method. The chief distinction between Saunders' stochastic method and our 2a Cartesian search seems to be the nature of the energy-minimization protocol. We minimize structures directly with the MM2 force field whereas Saunders preminimizes with his two-body force field prior to MM2 optimization. It is possible that the two-body field rectifies distorted atomic tetrahedra more effectively than does the authentic MM2 field. In any event, both acceptance rates and discovery of new minima are improved in a Cartesian search by employing the usage-directed structure selection scheme.

Overall, the tests above show that conformational searching with our method in torsional coordinates is superior to equivalent searches in Cartesian coordinates. While cyclotetradecane is handled reasonably well in either coordinate system, the MC search in torsional coordinates performed significantly better with the conformationally more complex unsymmetrical lactone. In that case, ~4100 Cartesian MC steps (method 2c) were required to give the same results as the 2900 torsional MC step (method 1) search.

The torsional MCMM search also performed well in comparison with the 60° resolution MULTIC grid search (method 3 in Table V). While the results of the 2900-step 12-membered unsymmetrical lactone test searches are similar, a comparison of the structures found indicates that neither the torsional MCMM search nor the MULTIC search found all the minima within the specified 25 kJ/mol of the ground-state conformation. By comparing the results of all the 12-membered lactone searches, we estimate the number of different minima within this energetic window to be roughly 400. Hence both searches found ~90% of the low-energy conformations. In order to find all of them, >10000 MC steps would likely be required. Such a search would require hours on a contemporary supercomputer or weeks on a machine with the computational speed of a MicroVAX II. The MULTIC method would also require a more extensive search to find all of the low-energy conformers of the 12-ring lactone above. Thus, more starting structures would have to be generated by the MULTIC method either by using a higher torsion angle resolution, by enlarging the allowable range for the ring closure distance, and/or by relaxing the nonbonded cutoff constraints. In any case, it is clear that while a 60° resolution torsional-grid search may find all minima for 7–10 membered rings, it is inadequate for the larger medium rings and large rings. For such relatively flexible structures, the continuous process implementation of the MCMM search has a distinct advantage. If we find that known structures have been missed or that minimum duplication rates are low at some point in a search, we simply allow the search to run for additional time.

The data in Table V also illustrate the conformational simplicity of the symmetrical rings relative to unsymmetrical structures of comparable size. Even though the cyclotetradecane ring is larger by two atoms than the lactone above, it is an easier problem from a conformational-search standpoint. Thus cyclotetradecane has fewer conformations by more than 1 order of magnitude than do similar structures in which the various ring atoms and substituents are distinguishable. The chiral, unsymmetrical, 12-membered ring lactone above appears to provide a conformational search problem of comparable magnitude to that of a 17-membered macrocyclic saturated hydrocarbon.

In comparison with the highly flexible 12-membered lactone above, the torsionally constrained 13-membered lactone of 7-oxobrefeldin A shown below is a relatively simple problem.



Starting with a randomly chosen high-energy conformer and defining the ring-closure bonds indicated by the arrows, three independent 2000-step usage-directed, torsional MCMM searches found the same set of 11 unique minima within 15 kJ/mol of the ground state with the MM2 force field. The structures of the lowest energy two conformers are shown as stereo pairs in Figure 1. The second structure is closely related to the x-ray structure[11] of brefeldin A found in the Cambridge crystal file. Indeed it is indistinguishable from the structure resulting from conversion of the brefeldin A crystal structure to 7-oxobrefeldin followed by energy minimization. It is also the same conformer proposed by Hutchinson and co-workers on the basis of NMR experiments.[12]

(11) Weber, H. P.; Hauser, D.; Sigg, A. P. *Helv. Chem. Acta* **1971**, *54*, 2263.

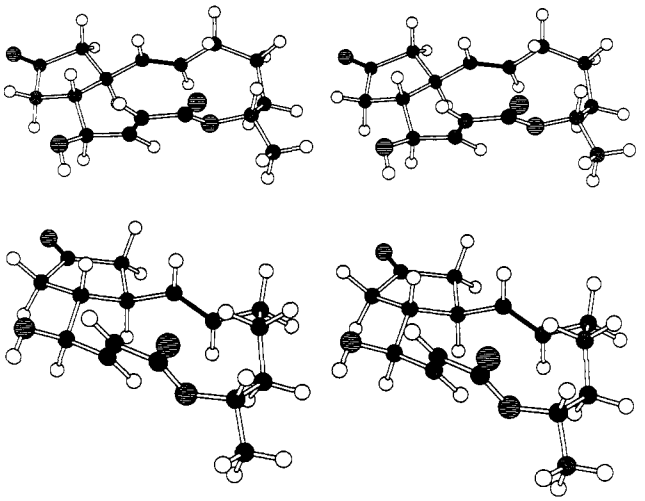


Figure 1. Stereoviews of the most stable conformer of 7-oxobrefeldin (top) and the next most stable conformer (+1.2 kJ/mol) of 7-oxobrefeldin (bottom).

Essentially the same results were found with triplicate 2000-step usage-directed Cartesian searches; however, the acceptance rate was considerably lower due largely to many instances of chiral center epimerization and olefin stereochemistry isomerization. We should also note that distance constraints originating from observables such as NOE effects are readily incorporated into the MCMM search. In the example at hand, reported[12] NOE effects were used to constrain H_a/H_e, H_b/H_c, and H_c/H_d to 2.0–5.0 Å and thus to direct the search to find nine compatible conformers. While brefeldin is small enough so that the reported NOE signals do not limit substantially the number of conformations, more flexible structures often undergo a major reduction in the number of valid conformers on applying distance constraints. We reiterate here the dangers of highly constrained searches which may hinder coverage of all regions of space when limited numbers of torsional variations are employed. When searches are highly constrained (e.g. by required internuclear proximities or energetic restrictions), it is necessary to allow torsional variations approaching the maximum number of rotatable bonds.

Conclusion

This study shows that the Monte Carlo search in torsional coordinates provides an effective method for exploring the conformational space of a variety of molecules. The way the search algorithm chooses starting geometries for MC steps plays a major role in determining the efficiency with which diverse regions of space are sampled. We find that selecting such starting geometries from among previously found low-energy structures and using all such structures an equal number of times improves the efficiency of the search at finding all low-energy conformers in comparison with a simple random walk. Another major determinant of search efficiency is the number of torsions varied in each MC step. By using a wide range of varied torsions, the search can take advantage both of energetically favorable torsional modifications and of more random jumps into remote regions of conformational space. Thus our optimal search protocol employs a combination of torsionally local search steps and more radical, global conformational changes. Taken together, these strategies provide effective conformational searches for all low-energy conformations of molecules having up to a dozen varying torsion angles.

While new structure generation in torsional coordinates is somewhat more complex than it is in Cartesian coordinates, it appears generally more efficient. Thus for a given amount of computer time, our torsional MC search finds more distinct, low-energy conformers and/or gives higher conformer duplication rates than does the corresponding Cartesian-based search. With complex structures having chiral centers and olefins of defined geometry, torsional variations maintain stereochemistry much more effectively than do Cartesian coordinate variations. On the other hand, our method of handling rings as acyclic structures with ring closure constraints increases the structure generation time relative to that required by the Cartesian search. The time in the structure generation phase however is not a limiting factor unless large numbers of geometrical constraints are employed. In the examples described above, only 0.1–1.0% of the total computer time was spent in structure generation. Energy minimization of starting geometries consumed the largest amount of the search time.

We find that implementation of the search as a continuous process provides a major practical advantage in that the extent of the search need not be specified at the outset. Instead, searches are simply continued until an examination of the accumulated results suggests that the search can be terminated. We discussed some criteria for establishing search convergence; however, providing firm evidence that all conformations have been found remains problematic. The most obvious approach to improving the confidence of finding all valid conformers is to enhance the search speed so that a more extensive search can be carried out in the same length of time. This objective could be accomplished by using faster hardware and by developing more efficient search protocols. Regardless of the techniques employed, establishment of search convergence is a significant problem and descriptions of conformational-search results should include evidence of comprehensive coverage of conformational space.

As for devising better conformational-search methods, there is considerable room for improvement. Currently, we must generate and energy minimize $\sim 10^2$ starting geometries for every unique, final geometry discovered in a converged, global conformational search for all minima. For a highly constrained search, considerably more MC steps with energy minimizations would have to be used. The conformational search problem is by current standards a tough one—it also presents one of the most significant barriers to attaining many of the contemporary objectives of molecular modeling. For simple systems, the MCMM method described here provides a useful solution. We hope our results will stimulate further work on this challenging problem.[13]

Registry No. *n*-Octane, 111-65-9; cyclodecane, 293-96-9; cyclotetradecane, 295-17-0.

(12) Hutchinson, C. R.; Kurobane, I.; Cane, D. E.; Hasler, H.; McInnes, A. G. *J. Am. Chem. Soc.* **1981**, *103*, 2477.

(13) This work was supported by grant CHE 86-05891 from the National Science Foundation.