

# The Multiple Minimum Problem in Molecular Modeling. Tree Searching Internal Coordinate Conformational Space

---

Mark Lipton and W. Clark Still

*Department of Chemistry, Columbia University, New York, New York 10027*

*Received 2 June 1987; accepted 8 September 1987*

A general algorithm is described which exhaustively searches conformational space using an internal coordinate tree search. Using only geometrical operations and a set of criteria for eliminating chemically unreasonable structures, the algorithm generates starting geometries for optimization by molecular mechanics. An implementation of this algorithm is exceedingly fast and finds all known minima, as well as several new ones, for four test alkanes. This method makes feasible global conformational searches for molecules with up to  $10^9$  conformational possibilities.

Since the earliest days of molecular modeling, it has been clear that construction of geometries approximating all of the low-energy conformations of flexible molecules is problematic for all but the simplest of structures. Even using the common simplification of fixed bond lengths and angles manual methods of conformer generation become impractical when more than three or four internal degrees of freedom (e.g., dihedral angles) are allowed to vary. The basic problem is that the required sampling of points in conformational space having  $N$  dimensions (e.g.,  $N$  dihedral angles) and  $M$  points in each dimension (e.g.,  $30^\circ$  dihedral angle resolution =  $360/30 = 12$  points) yields  $M^N$  possible structures. While the actual number of structures resulting from a given conformational search will always be less than this theoretical maximum due to the connectivity requirements of rings and the avoidance of high-energy non-bonded interactions, all  $M^N$  structures must be considered to be sure that no valid structure is overlooked. Thus highly flexible structure ( $N$  large) or high resolution ( $M$  large) conformational searches are problematic simply because the  $M^N$  number of possible structures is so large. Whereas manual searches become impractical beyond  $M^N = 10^2$ , computational methods for conformer generation can extend the accessible range of  $M^N$  by more than four orders of magnitude.

Apart from manual methods, the most common schemes for conformer generation are molecular dynamics<sup>1</sup> and distance geometry.<sup>2</sup>

The former method begins with an initial low-energy structure and integrates Newton's second law of motion over time to obtain classical atomic trajectories which define conformational changes and from which low-energy structures may be periodically sampled. The time periods over which many conformational changes occur, however, are rather long and as a result, molecular dynamics is most useful as a method for local conformational searching. The distance geometry method approaches the problem differently by defining molecular geometry by a set of interatomic distances. This definition of geometry makes it convenient to define conformers by specific 1, 2- and 1, 3-distances corresponding to standard bond lengths and angles, respectively, and sets of 1, 4-distances representing the values of various torsion angles. Since this set of distances is equivalent to a set of simultaneous equations in distance, cartesian coordinates for the distance geometry-defined conformers may be readily computed. Regardless of method used for the conformer generation, the resulting set of starting structures is geometrically refined by subsequent energy minimization to yield a unique subset of structures representing the low-energy conformers of the molecule of interest.

In this article, we describe an internal coordinate conformer generator<sup>3</sup> which has a number of advantages over the methods described above. First, it is quite fast. No energy gradient needs to be evaluated and no matrix

of distance equations needs to be solved with a result that the speed increase for conformer generation ranges from 10- to 100-fold over molecular dynamics or distance geometry methods. Second, operation of the method in an internal coordinate framework allows assured sampling over all accessible regions of conformational space. Third, the method is readily applied as a tree search, which makes feasible analyses of acyclics with  $M^N \leq 10^6$  and cyclic molecules with  $M^N \leq 10^9$ .

## GENERAL METHOD

The approach we employ to computationally generate molecular conformers is analogous to the way in which a chemist might go about the same task using mechanical molecular models. We start with a geometry-optimized structure and perform torsional rotations about all rotatable bonds, exploring every possible combination of rotamers at some convenient torsion angle resolution and retaining only those structures which pass geometrical tests designed to exclude high-energy molecular geometries.

This process amounts to a grid search sampling of conformational space—a  $3N-6$ -dimensional hyperspace whose axes are the molecule's  $3N-6$  degrees of freedom expressed in internal coordinates (bond lengths, bond angles, torsion angles). While degrees of freedom corresponding to bond lengths and angles could be explored along with torsion angles, the single-minimum, parabolic nature of the potential functions for bond stretching and bending allows the use of only single points for these variables. By thus restricting sampling to torsional space, the scope of the problem is reduced to at most, an  $N-3$ -dimensional one. Each axis  $i$  is partitioned into  $n_i$  segments, dividing the hyperspace into  $n_1 \cdot n_2 \cdot n_3 \dots n_{N-3}$  volumes. Conformational space is rigorously sampled by choosing one point from within each volume.

The problem of performing a point-by-point, methodical sampling of this sort becomes clear when we look at the simple case of  $n$ -decane. In order to completely explore its C-C-C-C torsional degrees of freedom with a sampling density of 6 points per torsional rotation ( $60^\circ$  dihedral angle resolution) and without allowances for symmetry, we would have to examine  $6^{(10-3)} = 280,000$  conformations.

To speed the conformational search, we perform the angular search in a sequential unidirectional fashion. In an acyclic system we start at one end of the structure and alter succeeding dihedral angles as we travel down the chain. In a cyclic system, rings are temporarily opened to form a pseudoacyclic molecule which is then processed as described for the acyclic case but with additional constraints enforcing ring closure. For example, the acyclic path chosen for the decalin ring system could be that illustrated in Figure 1, with dashed lines denoting the ring-closure bonds which are temporarily broken for the operation of the conformer-generating algorithm described here. The torsional angles in this example would be varied in the sequence  $\Phi_1, \Phi_2 \dots \Phi_5$ .

This torsion angle ordering permits the complete conformer generation process to be graphed as a tree structure. At the lowest level of the tree, the leaves stand for each of the possible products (conformations) of the generation process and each edge stands for a torsional rotation. Intermediate nodes in the tree represent partially completed structures. The number of nodes at level  $k$  is given by the product  $r_1 \cdot r_2 \cdot r_3 \dots r_k$  and each node at that level gives rise to  $r_k$  offspring nodes.

At any point in the tree, certain atoms in the molecule become spatially defined since they will not move in any torsional rotation farther down the tree. By applying constraint tests (e.g., tests for interatomic contacts) to these newly fixed portions of the molecule, we can eliminate in a single operation all structures which would arise from a partially defined structure having some undesirable

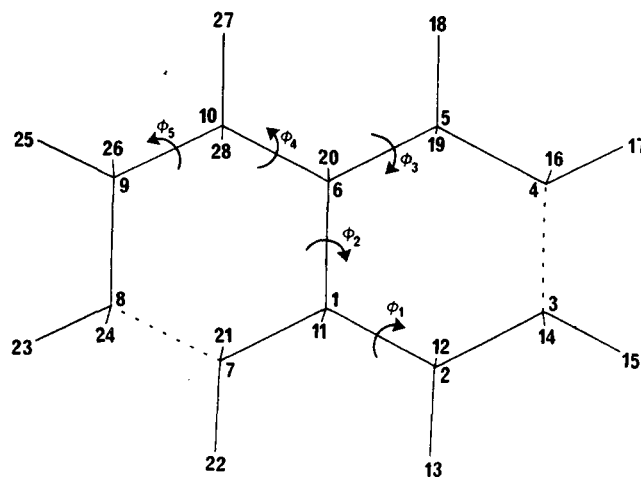


Figure 1. Decalin ring system as varied by MULTIC.

component. This elimination corresponds to pruning an entire branch from the structure generation tree.

A second consequence of ordering the rotations is the continual elaboration of a fixed portion of the molecule as one proceeds down the tree and through the sequence of torsional rotations. This sequential fixing of atomic positions permits additional streamlining of the operation by requiring calculation of an atom's coordinates only when it becomes fixed. To illustrate, let us again look at the decalin system in Figure 1. If the current rotation modifies torsion angle  $\Phi_{i2}$ , then atoms 1–3, 7, 11–15, 21, and 22 are fixed and atoms 5, 6, 10, and 20 will not move during the subsequent rotations about  $\Phi_{i3}$  through  $\Phi_{i5}$ . Alternatively, the rest of the atoms—though also moved by the current  $\Phi_{i2}$  rotation—are not yet fixed and will move again in later rotations. Therefore the coordinates of atoms 5, 6, 10, and 20 are the only ones immediately affected by the current rotation and are the only ones evaluated at this point.

Taken together, early constraint testing and minimal atomic coordinate computation provide an efficient scheme by which low-energy conformers may be selected from a large number of potential structures. Structures are rejected primarily by interatomic distance constraint tests which eliminate chemically unreasonable geometries or other structures which do not pass user-defined constraint tests. These tests are described in the following paragraphs and are generally lenient to avoid the loss of structures which may become low in energy and conformationally unique upon subsequent geometry refinement by energy minimization.

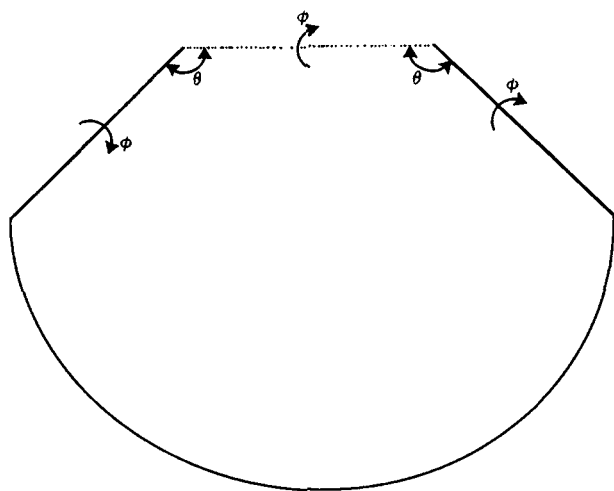
The first and most general constraint checks for minimum allowable separations between certain nonbonded atoms. It is a simple distance constraint test which eliminates conformations which would have high steric energies due to severe nonbonded contacts. It is useful to distinguish contacts between nonbonded atoms by the number of bonds separating the atoms. Since the severity of steric interactions between atoms separated by three bonds is effectively limited by realistic values of intervening bond distances and angles, 1,3- and 1,4-nonbonded interactions are not subject to nonbonded constraint testing. A 1,5-interaction represents

the most closely connected nonbonded arrangement we test for contact. Here, we often apply a rather stringent 3.0 Å cutoff between nonhydrogen atoms to eliminate conformations of the +gauche/–gauche or gauche/eclipsed pentane type. More remote interactions are tested with less restrictive distance cutoffs (generally 1.5 Å for all-atom structures and 2.0 Å for united-atom structures) since remote contacts may be effectively relieved by relatively small changes in the intervening bond distances, angles, and dihedrals. Nonbonded distance tests provide one of the most efficient ways to eliminate chemically unreasonable structures from those generated because the tests can be applied early in the structure-generation process and thus allow highly effective structure tree pruning.

A second distance test is employed with cyclic molecules. As noted above, one bond in each ring is defined as a ring-closure bond and is temporarily broken to produce a pseudoacyclic molecule for the purpose of conformer generation. Since we wish to retain the cyclic nature of the final structure, several additional constraints must be introduced. The most important of these, the closure distance constraint, is simply a range of acceptable distances between the two atoms forming the ring-closure bond. Two bond angles and three dihedral angles comprise the remaining closure-dependent internal coordinates and may also be constrained to reasonable values for final structure selection (see Fig. 2).

One problem can arise from breaking a ring bond to create a pseudoacyclic structure. If either atom defining the ring-closure bond is a stereocenter, its chirality could be lost when the ring is temporarily cleaved. To prevent such an occurrence, two dummy atoms are attached to each of the two ring closure atoms along the axis of the original closure bond. After a valid ring conformation has been generated, the proper stereochemistry can be reestablished by torsional rotations to realign the dummy atoms with the closure bond. Such rotations are described in this paper as "implicit" rotations.

Because closure constraints can only be applied to completed rings—necessarily late in the structure-generation process—rejecting structures via closure constraints is inher-



**Figure 2.** Closure parameters: 1 bond, two internal angles, and three torsional angles.

ently less efficient than rejection by nonbonded contacts. A partial remedy is provided by a long-range closure test described below.

At any point after a ring has been half completed, it can be tested to see if the remaining, geometrically undefined ring atoms can span the distance necessary to close the ring. By using the fixed internal coordinates in the starting structure (i.e., the unvaried bond distances, bond angles, and dihedral angles) as well as the optimal choice from the list of possible values for those internal coordinates which do change, the maximum distance which can be spanned by each section of the ring may be calculated by the method of Bixon and Lifson.<sup>4</sup> We define the sum of this maximum spanning distance and the maximum ring closure distance as the long-range closure cutoff. This calculation makes possible an early comparison of the distance between the geometrically defined ends of a partially formed ring with the appropriate long-range closure cutoff to determine whether or not the ring being generated has any chance of closing.

Although the scheme described here provides a rapid and well-defined method for uniformly searching internal coordinate space, one aspect of using simple distance constraint tests to detect high-energy conformations is less than ideal. Consider a 60° resolution torsional search with a 2.0 Å nonbonded rejection constraint. The search might generate, *inter alia*, a structure having a 60° torsion angle and a single nonbonded contact of 1.99 Å. Such a structure would be rejected by the 2.0 Å nonbonded test even if a slight ad-

justment of said 60° torsion angle to 61° would remove the offending nonbonded contact. Although the rigor of the search would be improved by allowing previously fixed angles to vary within limits related to the angular resolution in use, such a solution would be quite time-consuming and subject to multiple minimum problems of its own. As an expedient alternative, we use lenient constraints to limit the rejection of structures to only the most problematic of geometries.

## COMPUTATIONAL DETAILS

The routines which perform the above-described conformer generation were implemented as the MULTIC submode of the MacroModel interactive molecular modeling package and were written in FORTRAN-77 on a DEC VAX-11/780 running the VMS operating system. To produce an algorithm suitable for large molecular systems, the routines were optimized for both speed and memory allocation. This optimization was effected in large part through design of the data structure. The limitations we imposed were:

1. The data base (excluding the starting coordinates and connection table) would occupy no more than 2.5 Mb.
2. Operations would accommodate a maximum of 1000 atoms, 100 angles varied, 50 rings in a molecule, and 10 rings opened.
3. The number of possible conformations stored would depend on the amount of space remaining in the 2.5 Mb allocation.

The three features of the data structure which bear further discussion are the interaction matrix, which stores the nonbonded interactions; the arrays needed to perform rotations; and the array which stores valid conformations.

As much as possible of the arithmetic and data manipulation is performed in advance since the generation of conformers is a highly iterative procedure. Thus whether or not two atoms are to be tested for nonbonded contact is determined at an initial stage and stored for fast retrieval.

For this particular case the fastest possible data structure—a 1000 × 1000 matrix of

LOGICAL variables—is unreasonably large, even when using a nonstandard LOGICAL\*1 array. To reduce the memory cost it was necessary to reference individual bits. This was done by means of a 32-element INTEGER\*4 single-bit mask array, which allows the program to selectively test or set any bit in a 32-bit word. Further savings were obtained by eliminating the redundancy of a symmetrical square matrix and reducing it to a lower-left triangular matrix without the diagonal elements (which encode the trivial case of an atom interacting with itself). These modifications reduce the size of the “matrix” to 63.5 kb, well within our limit. The heavy atom 1,5 interactions, in addition to being stored in INTMAT, are also identically stored in a separate matrix (I15MAT).

FORTRAN does not provide for triangular matrices, so the interaction matrices are actually 15610 element INTEGER\*4 arrays (INTMAT and I15MAT). The proper bit for the interaction of atoms  $i$  and  $j$  ( $i > j$ ) is calculated using the formula  $(i - 3)i/2 + j$ , but a pointer array containing the value  $(i - 3)i/2$  in element  $i$  was used to further increase the speed. In this way the need to check an interaction is determined quickly by testing the proper bit in INTMAT and I15MAT using the mask array and the pointer array.

Another notable feature of the data structure is the way in which rotation information is stored. This problem may be broken down into how the atoms which move are stored and how rotation information is transmitted to lower levels of the tree.

The fastest way to encode atom movement consists of an INTEGER array dimensioned 1000 by 100, each column containing a variable number of nonzero elements corresponding to the atoms which must be moved during each rotation. Because this scheme is too wasteful of space (the columns would consist mostly of zeroes), we use an alternative which relies on the fact that each atom will move only once. The program creates a temporary variable dimensioned to 32 (1024 bits) by 100 words. For rotation number  $i$ , every atom moved by rotation has its corresponding bit set to 1 in column  $i$ . From this array, a permanent 1000-element INTEGER\*4 array is established which contains a list of the atoms arranged in the order in which they stop mov-

ing. Two pointer arrays indexed by rotation number provide quick access to atoms which move during each rotation. Using the decalin in Figure 1 as an example, we would produce the following sequence of atoms in the list: 1 2 3 12 13 6 7 11 5 10 20 4 18 19 9 27 28 8 25 26 14 15 16 17 21 22 23 24. Note that hydrogens attached to closure atoms stop moving last due to the implicit rotations at the end of the sequence.

Rotations are stored as  $4 \times 4$  rotation-translation matrices to account for translation of the origin during rotation and translation back to the initial frame of reference afterward. The actual matrix is the product of a translation matrix, a rotation matrix coding for a rotation about an arbitrary axis and the inverse of the first translation matrix:

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & -x_0 \\ 0 & 1 & 0 & -y_0 \\ 0 & 0 & 1 & -z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos \theta \cos \varphi & \cos \theta \sin \varphi & -\sin \theta & 0 \\ \sin \varphi & -\cos \varphi & 0 & 0 \\ -\sin \theta \cos \varphi & -\sin \theta \sin \varphi & -\cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Transmission of this matrix to a later rotation is done by matrix multiplication. Although one could simply initialize all matrices to the identity matrix, for speed we employ a LOGICAL array to indicate when no multiplication is necessary; under those conditions the current matrix is simply used to overwrite the matrix for the later rotation. Since lower levels of the tree may be reached infrequently, the actual matrices applied may be the product of hundreds or thousands of multiplications. Using 32-bit arithmetic, this scheme does not lead to significant numerical error (see below).

Other arrays are created from the temporary array which enable MULTIC to relate information from one rotation to another further down the tree. Because this information is transmitted down the tree and never in the other direction, a linked list suffices for

straight paths. Branched paths, however, need a second linked list to provide simultaneous transmission to two or more different rotations at a branch point. The links needed to describe the decalin's rotations are: rotation  $\text{Phi}_1$  affects  $\text{Phi}_2$ , which in turn affects  $\text{Phi}_3$ . Because of branching, however,  $\text{Phi}_2$  also affects  $\text{Phi}_4$ —which is shown by an "equivalence" link from  $\text{Phi}_3$  to  $\text{Phi}_4$ . Rotation  $\text{Phi}_4$  affects  $\text{Phi}_5$ ;  $\text{Phi}_5$  points to nothing, thereby ending the sequence.

The data structure also stores the valid conformations. Although one could simply write the structures to a permanent storage medium (e.g., disk) as they are generated, this approach was not adopted for two reasons. First, storing them in memory allows further operations such as the elimination of near-duplicate conformations, ordering by energy or the application of further constraints prior to output. Second, without any input/output, the routine can generate structures much faster than it could if it also had to write them to disk. As a result, the user can quickly generate a set of conformations, examine the number produced, redefine his constraints if need be, and perform any of the subsequent functions.

Valid conformations are stored as the values of the varied angles. For all but implicit rotations, the angle can be related back to an element of the array which holds the specified values for each angle. Hence, only the integer subscript needed to find this value is stored. The `INTEGER*4` subscript is converted to an 8-bit signed integer and stored in a non-standard `BYTE` array to conserve memory.

For implicit rotations there is no list of possible angles, so the current angle is stored as an angle relative to the initial value (an arbitrary reference point). This difference is placed in the interval  $-180/180^\circ$ , divided by two and converted from `REAL*4` to an 8-bit signed integer for storage in the `BYTE` array.

The actual task of generating conformations is done by a recursive subroutine since it amounts to an exploration of the tree of conformations. While `VAX/VMS FORTRAN-77` does not allow recursion, it could be simulated using a dummy calling routine and placing variables which would normally go onto the stack instead into `COMMON` as arrays dimensioned to a maximum level of 100.

MULTIC was implemented as a standard menu-driven submode of the MacroModel mo-

lecular modeling system since the structure generations described are rapid enough for use in an interactive environment. The user thus starts with an energy minimized starting geometry of the molecule of interest, selects bonds and allowable values for torsion angle rotation, adds constraints for the conformational search and then starts the generation process. The amount of time necessary for the analysis depends on the size of the structure, the number of bonds being rotated, and the angular resolution of the search but is typically several minutes (see below), far less than the time necessary to energy minimize all the structures produced.

## RESULTS AND DISCUSSION

To assess the capabilities of MULTIC and to gain an understanding of the relationship between nonbonded cutoff, closure parameters and torsional resolution, an investigation of *n*-hexane, cyclooctane, cyclononane, and cyclodecane was undertaken. The objective of the study was to define a set of parameters which could be used with reasonable confidence for global conformational searching without producing unnecessarily large numbers of starting geometries.

*n*-Hexane was used to determine the optimal nonbonded distance cutoff as well as the number of torsional rotations per bond ("resolution") needed to produce all known conformations. Cyclooctane was used to explore and optimize the ring closure parameters. Finally, cyclononane and cyclodecane were used to test the optimal values for nonbonded cutoffs, torsion angle resolution and ring-closure parameters.

Structures from the conformation generator were subjected to block diagonal Newton-Raphson energy minimization (final RMS gradient  $<0.01$  kJ/Å) using the MacroModel implementation of the MM2 molecular mechanics force field.<sup>5</sup> Each minimum energy conformer was superimposed by the least-squares method of Kabsch<sup>6</sup> with all other minimum energy structures and was saved as unique only if no fit better than 0.1 Å RMS was obtained. Each unique conformer was then subjected to full matrix Newton-Raphson energy minimization to reduce the gradient to less than 0.001 kJ/Å and finally tested for imaginary vibrational frequencies by calculating the eigenvalues of the

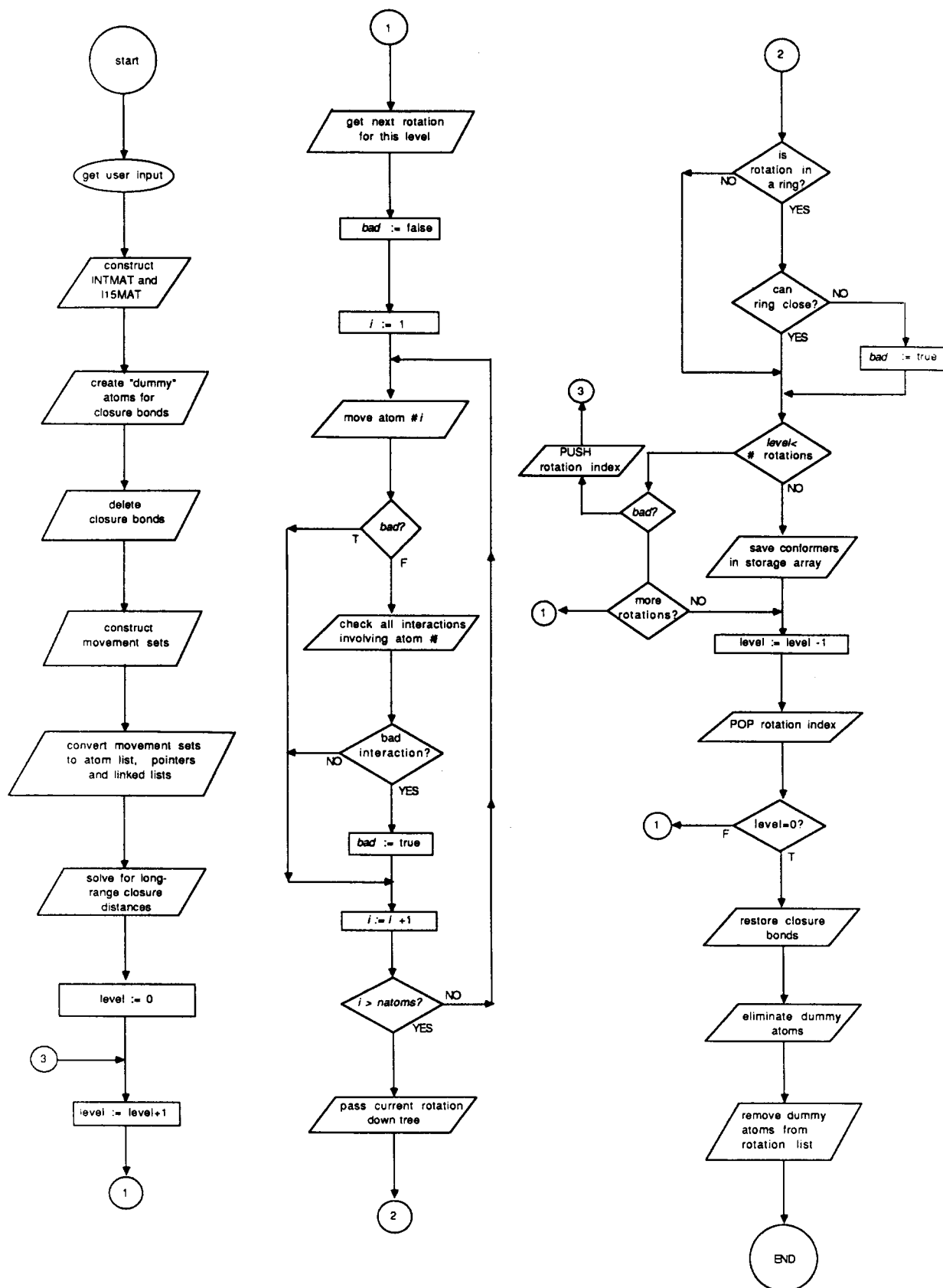


Figure 3. Flow chart of the operation of MULTIC.

second derivative matrix. All of these operations were performed by the MacroModel batch mode energy minimization routine BATCHMIN. To determine what set of structures represented all possible conformers, the structure generation was run at double the apparently adequate resolution (i.e., at 30°) without rejecting structures having close nonbonded contacts on the assumption that at so high a resolution no valid conformation of these small molecules would escape. Finally, the complete set of conformations produced were shown to include all previously reported conformations of the hydrocarbons under study unless it could be demonstrated that the literature conformations were not minima. It must be remembered however that this scheme, like all others yet proposed, is probabilistic and there is no guarantee that low energy minima have not been missed. The results of the tests with *n*-hexane are summarized in Table I.

Several conclusions may be drawn from these results. First, in order to find all minimum energy conformations it is necessary to

use a dihedral angle resolution higher than 120 degrees. Since we wish to obtain all the staggered positions for the  $sp^3 - sp^3$  linkages, 60° resolution is the next choice. As shown above, the combination of 60° resolution with a 1.5 Å nonbonded cutoff distance produced the minimum number of starting geometries needed to obtain all low-energy conformations upon energy minimization. In this study and in the others below, the special case 1,5-heavy atom cutoff was omitted. It should be noted that two of the conformations found in this study of *n*-hexane have not been previously reported.<sup>7</sup> Indeed, global conformation searching commonly finds new low-energy minima and the new conformers found during this study are indicated in Table V.

These conclusions were tested using cyclooctane (Table II) in which the torsional resolution and nonbonded cutoff were varied along with a closure distance and two closure angles. In this and other cyclic situations, we must also consider the closure parameters consisting of one closure distance, two closure angles and three closure torsions. Since, however, torsion is energetically the softest of the internal coordinates, we have investigated only the results of modifying the closure distance and angles. At 120° resolution, important low-energy conformations of cyclooctane were missed unless the closure and nonbonded distance constraints were severely relaxed; however, as with *n*-hexane, 60° resolution proved adequate so long as a closure distance window of at least 1 Å was used. When the 2 Å window was used, the closure bond angles could be constrained

Table I. *n*-Hexane.

Torsional resolution	Nonbonded cutoff (Å)	Number of conformations	
		Generated	Minimized
120°	1.50	17	7
120°	0.00	27	10
60°	1.75	69	11
60°	1.50	99	12
60°	0.00	144	12
30°	2.00	292	12
30°	1.50	708	12
30°	0.00	1008	12

Table II. Cyclooctane.

Torsional resolution	Nonbonded cutoff	Closure		Refinement (Y,N)	Number of conformations	
		Dist	Angle		Generated	Minimized
120°	0.00	1.0-3.0	65-155	Y	2	1
120°	0.00	1.0-3.0	0-180	Y	5	3
120°	0.50	1.0-5.0	65-155	Y	13	4
120°	0.50	1.0-5.0	0-180	Y	26	4
60°	1.50	1.4-1.6	90-130	N	2	2
60°	1.50	1.4-1.6	0-180	N	3	3
60°	1.50	1.0-2.0	90-130	N	7	2
60°	1.50	1.0-2.0	0-180	N	16	4
60°	1.50	1.0-3.0	90-130	N	22	4
60°	1.50	1.0-3.0	0-180	N	72	4
60°	2.00	1.0-3.0	0-180	N	5	4
30°	1.50	1.4-1.6	90-130	N	20	4
30°	1.50	1.4-1.6	0-180	N	68	4
30°	1.50	1.0-2.0	0-180	N	371	4
30°	1.50	0.5-2.5	90-130	N	928	4



to chemically reasonable values without a loss of minimum energy conformations. In Table II, the Refinement column notes whether or not the structure has been crudely minimized in internal coordinates prior to storage. This refinement uses a steepest descent algorithm using numerical derivatives of selected internal coordinates with respect to strain energy and limits structural changes to small variations (maximum 10°) in the variable dihedral angles. Such refinements are used to delocalize the strain of any abnormally long ring-closure bonds over the entire molecule prior to minimization in cartesian space.

Results with cyclononane and cyclodecane are summarized in Tables III and IV. The results shown are for conformer generations having the first dihedral angle set to +60° on the assumption that at least one gauche-like torsion must appear somewhere in each conformer and also to avoid generation of enantiomeric conformers. Again, the combination of 60° torsional resolution, 1.5 Å nonbonded cutoffs, a 1–3 Å closure window and 65–155° closure angles yielded all available conformations with the smallest number of starting geometries. Several other interesting observations were made. First, narrowing the closure angle window with cyclononane to 40° with 60° dihedral resolution reduced the

number of starting geometries dramatically but resulted in the loss of the global minimum energy structure. Alternatively, the 90° bond angle window was useful for reducing the number of starting geometries without loss of low energy conformers. Second, for the systems studied, a 2 Å closure window was adequate although a 1 Å window could be used with cyclooctane provided no angular constraints were used. Among 2 Å closure windows, a closure distance range of 0.5–2.5 Å was less effective than the range of 1.0–3.0 Å. Third, for a given number of generated structures, internal coordinate refinement (described above) prior to cartesian coordinate minimization generally yields more minima presumably due to improved maintenance of the uniform coverage of conformational space during the minimization process. Thus our conformation searching protocol for medium rings emerges as one with 60° dihedral angle resolution, 1.5 Å nonbonded cutoffs, a 65–155° closure angle window and a closure distance window starting at 1.0 Å and extending to a distance numerically equal to 1.0 plus one fourth of the ringsize. With the several dozen medium ring compounds we have studied in recent years, these parameters have been adequate to yield all of the final conformers we have been able to discover with higher resolution searches.

Table III. Cyclononane.

Torsional resolution	Nonbonded cutoff	Closure		Refinement (Y,N)	Number of conformations	
		Dist	Angle		Generated	Minimized
60°	1.5	1.4–1.6	0–180	N	2	2
60°	1.5	1.3–1.7	0–180	N	4	3
60°	1.5	1.0–3.0	0–180	N	57	6
60°	1.5	1.0–3.0	65–155	N	40	6
60°	1.5	1.0–3.0	90–130	N	7	5

Table IV. Cyclodecane.

Torsional resolution	Nonbonded cutoff	Closure		Refinement (Y,N)	Number of conformations	
		Dist	Angle		Generated	Minimized
120°	1.5	1.0–4.0	0–180	N	14	8
120°	1.5	1.0–4.0	0–180	Y	14	8
60°	1.5	1.0–3.0	0–180	Y	317	17
60°	1.5	1.0–3.0	65–155	Y	153	17
60°	1.5	0.5–2.5	65–155	Y	104	15
60°	1.5	1.0–2.0	0–180	N	25	12
60°	1.5	1.0–2.0	0–180	Y	25	15
60°	1.5	1.3–1.7	0–180	N	23	9
30°	1.5	1.3–1.7	90–130	N	240	17

For larger molecules having more torsional degrees of freedom, 60° resolution conformer generation may produce too many structures to minimize in a realistic amount of computer time. In such cases, there are several alternatives to the protocol above which can simplify the problem though not without increasing the uncertainty of searching conformational space adequately. First of all, 120° resolution can be used for  $-C(sp^3)-C(sp^3)$ -torsional arrays with 60° resolution being retained for the typically lower barrier  $-C(sp^3)-C(sp^2)$ - or  $-C(sp^3)$ -heteroatom- substructures. Whenever 120° resolution is used, it is important to relax ring-closure constraints. We typically use a closure distance window equal to half the size of the ring being closed with no closure angle constraints whatsoever when such low-resolution searches are being made. We have also found it important to deemphasize or remove nonbonded contact tests when conducting searches on large structures since small angular changes can make large differences in remote nonbonded distances. We must stress, however, that when it becomes necessary to use 120° dihedral angle resolution then it is probable that conformational space is not adequately covered by the search and that low energy conformers, possibly even the global minimum, will be missed. Even with 60° resolution, there is no guarantee that all low-energy minima will be found.

Other schemes for limiting the number of starting geometries rely on the avoidance of well known high-energy substructures or on the incorporation of *ad hoc* constraints to fit available experimental data (e.g., dihedral angles from NMR coupling constants or distances from NOE observations). Conjugated functionality with its typically high rotational barriers (e.g., esters, enones, enol ethers) provides a common and relatively safe opportunity to reduce the associated dihedral angle resolution to 180°. More specialized constraints could include, for example, searching only the low-energy regions of the appropriate Ramachandran map for peptides. It is interesting that one of the cornerstones of hydrocarbon conformational analysis, the  $+/-$  gauche pentane substructure, is not necessarily as energetically damaging as is frequently assumed. Thus while such a substructure in a diaxial *cis* 1,3-dimethylcyclohexane carries an energetic penalty of >3 kcal/mol due to ring-buttressed non-

bonded interactions, the unbuttressed acyclic variant ( $+/-$  gauche pentane itself) is strained relative to  $+/+$  gauche pentane by only ca. 1 kcal/mol according to MM2. Thus it is not safe to routinely delete all starting structures having such components. In MULTIC, we use a special 1,5-nonbonded test to eliminate such substructures only if they occur multiple (typically defined as 2–4) times.

Dihedral angles and relative steric energies (MM2) for the unique conformations of *n*-hexane, cyclooctane, cyclononane, and cyclodecane found during this work are listed in Table V. Previously undescribed conformers,<sup>7</sup> a number of which are symmetrical, are designated in the table.

Timing tests were used to establish the dependence of the operation time on the number of atoms in the molecule and number of output conformations produced. To simplify the tests, we removed the nonbonded cutoff constraint and tested acyclic hydrocarbons with and without attached hydrogens. The CPU times in seconds observed for pentane, hexane, and heptane are shown in Tables VI and VII.

Tests were made which isolate each component of the structure-generating operation. Thus, the same molecule was searched at various dihedral angle resolutions to gauge the effect of increasing the number of rotational increments. We also changed the substituents from none to hydrogens to methyls, thereby increasing the number of atoms without affecting the number of structures produced. A final test involved increasing the chain length of the alkane which changes both the number of atoms and number of rotations simultaneously. The data from Tables V and VI support a relationship between operation time (*t*) and the two variables mentioned above with the empirical form:

$$t = n^2(cm + t_0) \quad (1)$$

where *t* is the conformer generation time, *n* is the number of atoms, *m* is the number of conformations produced and *t*<sub>0</sub> and *c* are constants. These constants were determined by linear regression analysis of the time divided by *n*<sup>2</sup> versus *m*. The line produced by *c* = 7.97 10<sup>-4</sup> and *t*<sub>0</sub> = 0.022 fit the data with a correlation coefficient *r* = 0.999.

The expression above shows that the time of operation is proportional to the number of

**Table V.** Unique minima of hexane, cyclooctane, cyclononane, and cyclodecane.

Compound	Energy kcal/mol	Torsion angles									
		Phi <sub>1</sub>	Phi <sub>2</sub>	Phi <sub>3</sub>	Phi <sub>4</sub>	Phi <sub>5</sub>	Phi <sub>6</sub>	Phi <sub>7</sub>	Phi <sub>8</sub>	Phi <sub>9</sub>	Phi <sub>10</sub>
Hexane	0.0	180.0	180.0	180.0							
	0.9	179.6	175.8	65.3							
	0.9	175.4	65.8	175.5							
	1.3	175.3	59.9	58.9							
	1.6	64.5	171.6	64.5							
	1.8	66.4	179.8	-66.8							
	2.2	59.7	56.4	59.7							
	<sup>a</sup> 3.1	62.2	-95.4	178.7							
	3.2	173.0	62.4	-95.0							
	3.8	93.3	-64.4	-61.5							
	<sup>a</sup> 4.1	62.6	-94.7	-71.5							
	5.4	94.8	-65.8	94.6							
Cyclo- octane	0.0	-101.7	43.6	64.8	-64.8	-43.6	101.7	-67.5	67.5		
	1.0	-84.9	111.7	-84.9	62.5	-84.9	111.7	-84.9	62.5		
	1.9	-88.9	91.9	-48.8	-47.2	117.5	-47.2	-48.8	92.0		
	2.9	-36.6	64.8	36.6	-64.8	-36.6	64.8	36.6	-64.8		
Cyclo- nonane	0.0	-56.1	125.4	-56.1	-56.1	125.4	-56.1	-56.1	125.4	-56.1	
	0.7	-51.0	102.9	-85.9	102.9	-51.0	-70.5	66.6	66.6	-70.5	
	0.8	-121.9	85.6	-73.1	117.7	-64.8	-64.8	117.7	-73.1	85.6	
	<sup>a</sup> 2.2	-102.7	117.6	-88.1	55.3	-90.2	148.2	-59.8	-47.2	96.4	
	3.1	-129.0	103.7	-38.6	-51.7	134.0	-51.0	-68.9	57.1	53.3	
	<sup>a</sup> 3.3	-130.5	122.0	-39.9	-39.9	122.0	-130.5	84.2	-68.1	84.2	
Cyclo- decane	0.0	-151.1	55.2	65.7	-65.7	-55.2	151.1	-55.2	-65.7	65.7	55.2
	0.6	53.2	-96.0	151.4	-64.4	-57.7	130.0	-57.7	-64.4	151.4	-96.0
	1.2	68.3	-76.6	-61.3	138.4	-53.1	-53.1	138.4	-61.3	-76.6	68.3
	1.3	-145.5	145.5	-83.6	67.2	-83.6	145.5	-145.5	83.6	-67.2	83.6
	1.6	-142.8	135.3	-84.3	61.5	-89.3	156.2	-61.2	-71.7	57.8	51.9
	<sup>a</sup> 2.4	-120.3	61.3	70.9	-140.9	109.8	-45.1	-59.8	152.6	-92.8	71.5
	<sup>a</sup> 3.1	-138.3	85.5	-83.0	121.1	-68.4	-72.7	149.6	-48.0	-45.4	128.9
	<sup>a</sup> 3.3	-108.8	131.1	-108.8	94.4	-53.9	-57.5	168.8	-57.5	-53.9	94.4
	<sup>a</sup> 3.8	-151.3	112.9	-92.9	50.0	58.6	-151.3	112.9	-92.9	50.0	58.5
	4.1	-125.2	54.5	64.8	-72.9	-72.3	72.3	72.9	-64.8	-54.6	125.2
	4.4	-68.2	120.7	-89.0	120.7	-68.2	-68.2	120.7	-89.0	120.7	-68.2
	<sup>a</sup> 4.4	-127.7	71.0	-91.3	127.3	-100.3	108.7	-65.6	-67.3	87.3	64.8
	<sup>a</sup> 4.9	-124.0	88.5	-82.7	148.1	-98.3	38.5	-98.3	148.1	-82.7	88.5
	<sup>a</sup> 4.9	-145.9	78.6	-80.4	118.3	-80.4	78.6	-145.9	54.2	25.5	54.2
	<sup>a</sup> 8.6	-133.1	60.3	44.4	-97.4	140.2	-58.1	-72.0	36.9	67.0	-2.1
	8.8	-147.6	121.0	-43.0	-43.0	121.0	-147.6	121.0	-43.0	-43.0	121.0
	<sup>a</sup> 10.3	110.0	-48.2	-75.5	20.8	78.1	20.8	-75.5	-48.2	110.0	-77.7

<sup>a</sup>Conformations not previously described.**Table VI.** Timing tests for united atom molecules.

Molecule	Resolution	Number of confs	Total time	S/structure
<i>n</i> -Pentane (5 atoms)	120°	9	0.12	0.0130
	60°	36	0.23	0.0083
	30°	144	0.73	0.0065
<i>n</i> -Hexane (6 atoms)	120°	27	0.65	0.0240
	60°	216	1.09	0.0050
	30°	1728	6.48	0.0038
<i>n</i> -Heptane (7 atoms)	120°	81	0.98	0.0120
	60°	1296	6.52	0.0050
	30°	20736	95.98	0.0046

conformations found and the square of the number of atoms. These relations are supported by the fact that nonbonded checks increase as  $n^2$  (MULTIC still performed the tests although the result was always to keep the conformer) and that the number of edges

in a tree increases more slowly than the number of leaves.

It should be noted that the timing tests above represent a worst-case analysis. When tree pruning occurs due to nonbonded interactions, the operation time will be signifi-

**Table VII.** Timing tests for all atom molecules.

Molecule	Resolution	Number of confs	Total time	S/structure
<i>n</i> -Pentane	120°	9	0.39	0.043
(17 atoms)	60°	36	0.96	0.027
<i>n</i> -Hexane	120°	27	1.32	0.049
(20 atoms)	60°	216	6.68	0.031
<i>n</i> -Heptane	120°	81	3.24	0.040
(23 atoms)	60°	1296	47.53	0.037

cantly faster than the tables above suggest. To test this assumption, several of the examples from Tables VI and VII were repeated using the standard nonbonded cutoff. The resulting times were 50–100% faster than the predictions based on Eq. (1).

We also needed to gauge the extent to which this algorithm was dependent on the starting structure. While the precise number of structures produced by MULTIC will vary with starting structure, the number of minima found by energy minimization should be invariant provided that adequate resolution is used in the conformation search. A number of different starting structures were used to test this point. In all cases, no change in the number of minima was observed when the optimal parameters were used. An example is shown in the results for cyclononane (Table III). Although conformer generations usually began with the so-called "twist chair-boat" conformation, conformer generation starting with the highest energy conformer (the "boat-chair") led to the same six minima.

A significant problem with any iterative routine is the loss of numerical accuracy due to roundoff errors. To assess the impact of roundoff error on the output conformer coordinates, the following test was conducted. Eleven torsional angles in two rings of a 74-atom macrobicyclic molecule were varied at 60° resolution. MULTIC produced 203 structures in approximately five CPU minutes. The rotations had been arranged in such a way that the final structure produced should have been identical to the starting conformation. When superimposed using the least-squares algorithm of Kabsch,<sup>6</sup> the coordinates of all atoms in the two structures matched to within an RMS deviation of 0.0006 Å. This result demonstrates that the numerical inaccuracies produced by single precision arithmetic in this context are insignificant.

## CONCLUSION

Internal coordinate tree-searching provides a rapid method for generating molecular geometries approximating the various conformers of small to medium-sized organic molecules. In the molecules examined, 60° dihedral angle resolution was found to be adequate for location of all minima as judged by the fact that increasing the dihedral angle resolution to 30° provided no new conformers. To the extent that 60° resolution is adequate for all structures, this method makes feasible complete conformational analyses of acyclic molecules having up to six independent torsional angles and cyclic structures having up to nine independent torsional angles. Thus molecules such as nonane, 2-decene, cyclododecane, and cyclotetradecadiene represent practical upper limits for this approach to the multiple minimum problem with a VAX-class computer. In our heads, application of this approach to the conformational analysis of macrocyclic organic molecules commonly leads to the discovery of previously undescribed, low-energy conformers and is thus clearly superior to any manual method of conformer discovery yet described. Nevertheless, this method like all others is a probabilistic one and, regardless of the search resolution used or the number of starting geometries produced, there can be no guarantee that all low-energy minima have been found.

Even at this point, practical limitations reflect the much more time-consuming energy minimization stage of the analysis rather than the structure-generation stage described above. Thus the use of full conformational space searching for larger molecules will be feasible only when substantially faster energy minimizations become possible and/or when those starting structures which are destined to become conformational duplicates can be eliminated early in the analysis. Even with such improvements and others yet to be

devised, many interesting molecules will remain beyond the scope of the full space conformational analysis which appears necessary to assure detection of all distinct minima or even the global minimum. A simple protein composed of 50 amino acids is so far beyond present capabilities that it must be considered noncomputable from a global search standpoint: even the peptide backbone analysis alone at 120° torsion angle ( $\phi, \psi$ ) resolution would keep a supercomputer which is able to energy minimize one structure/second fully occupied for more than  $10^{40}$  years.

This work was supported by a grant from the National Science Foundation.

### Notes and References

1. J. A. McCammon and M. Karplus, *Acc. Chem. Res.*, **16**, 187 (1983).
2. G. Crippen, *Distance Geometry and Conformational Calculations*, Research Studies Press, Wiley, New York, 1981.
3. The predecessor (RINGMAKER) of the conformer-generating program described here was outlined previously: W. C. Still and I. Galynker, *Tetrahedron*, **37**, 3981 (1981). A tree-searching algorithm similar to ours has been described recently for finding peptide conformations: R. Bruccoleri and M. Karplus, *Biopolymers*, **26**, 137 (1987). See also P. J. De Clercq, *Tetrahedron*, **40**, 3717, 3729 (1984); Y. Fukazawa, S. Usui, Y. Uchio, Y. Shiobara, and M. Kodama, *Tetrahedron Lett.*, **27**, 1825 (1986); D. P. Dolata and R. E. Carter, *J. Chem. Inf. Comput. Sci.*, **27**, 36 (1987).
4. M. Bixson and S. Lifson, *Tetrahedron*, **23**, 769 (1967).
5. N. L. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977).
6. W. Kabsch, *Acta Cryst.*, **A32**, 922 (1976).
7. For a description of *n*-hexane conformations see, for example: L. Schaefer, K. Siam, J. D. Ewbank and E. Osawa, *J. Mol. Struct. (THEOCHEM)*, **32**, 125 (1986). For those of cyclooctane, cyclononane, and cyclodecane see: J. B. Hendrickson, *J. Am. Chem. Soc.*, **86**, 4584 (1964); *Ibid.*, **89**, 7036, 7043, 7047 (1967); K. B. Wiberg, *J. Am. Chem. Soc.*, **87**, 1070 (1965); N. L. Allinger, J. A. Hirsch, M. A. Miller, I. J. Tyminski, F. A. Van Catledge, *J. Am. Chem. Soc.*, **90**, 1199 (1968); N. L. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977); F. A. L. Anet and J. Krane, *Isr. J. Chem.*, **20**, 72 (1980); P. A. Kollman, *et al.*, *Tetrahedron*, **39**, 1113 (1983).