A Comprehensive Approach to the Conformational Analysis of Cyclic Compounds

Noham Weinberg* and Saul Wolfe*

Contribution from the Department of Chemistry, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada

Received June 7, 1994®

Abstract: Currently available strategies for the conformational analysis of cyclic compounds include the molecular mechanics energy minimization of randomly distorted previously optimized cyclic structures (Ōsawa, Raber, Saunders, Wilson) or minimization of the cyclic structures contained within a set of acyclic structures (Fromowitz, Hruby, Karplus, Pavitt, Scheraga, Still, Veber). Reported herein is a novel constrained stochastic search strategy which affords large numbers of *only* cyclic initial structures. The algorithm is based on a random variation of the dihedral angles of a chain *within the range of angles that are compatible with ring closure.* With efficiency defined as the number of cyclic probe structures lacking internal close contacts that can be produced per second, for 10-, 15-, 20-, 30-, and 50-membered rings our strategy produces these structures 40, 80, 160, 400, and 4500 times more efficiently than other methods tested. The structures generated in this way appear to be well distributed over the conformational space: all previously reported MM2 and MM3 minima for 9-, 10-, 11-, 12-, and 17-membered cyclic hydrocarbons have been found by minimization of 50, 100, 1000, 2500, and 60000 cyclic probe structures, respectively. Numerous previously unreported (usually high energy) minima have also been found. These include structures in which one or more hydrogens or methylene groups are directed toward the interior of the ring, one trefoil knot, and artifacts that contain pyramidal carbon. The latter are found only with MM2, because of the different expressions for bending employed in the MM2 and MM3 force fields. Using a version of the algorithm that avoids high energy structures, a total of 12513 MM2 minima has been found for cycloheptadecane.

Introduction

Polyatomic molecules have so many degrees of freedom that the discovery of all of their stable conformations is a formidable challenge. A brute force search of the conformational space ("tree search" [1]) is possible in principle, but since this approach is practical only for molecules having fewer than about 50 atoms,[1g,h,2r] many workers now apply random search procedures to the problem.[2,3] These are of two types, which differ conceptually in how the random structures are generated.

The more widely used of these includes all of those methods in which each successive structure is found by energy minimization of a randomly distorted (kicked,[2p,q] pulsed,[2n,o] flapped,[2l,m] etc.) previously optimized structure (Figure 1). In this "inductive" approach, information is gathered first within a

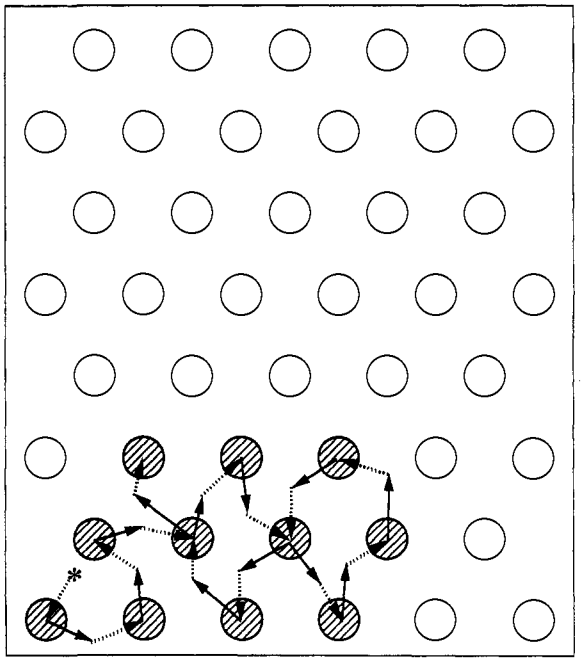


Figure 1. A random walk over multiminimum conformational space. The search begins with an arbitrarily selected initial structure (*), whose energy is minimized (broken arrow). The new structure (hashed circle) is subjected to a random distortion (solid arrow), and the process is reiterated. Open circles refer to minima that have not yet been identified.

® Abstract published in *Advance ACS Abstracts,* September 15, 1994.

(1) (a) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 188. (b) Dygert, M.; Gō, N.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 750. (c) Still, W. C.; Galynker, I. *Tetrahedron* **1981**, *37*, 3981. (d) Still, W. C. In *Current Trends in Organic Synthesis*; Nozaki, H., Ed.; Pergamon Press: Oxford, 1983; p 233. (e) Lipton, M.; Still, W. C. *J. Comput. Chem.* **1988**, *9*, 343. (f) Smith, G. M.; Veber, D. F. *Biochem. Biophys. Res. Commun.* **1986**, *134*, 907. (g) Fromowitz, M.; Hruby, V. *Int. J. Pept. Protein Res.* **1989**, *34*, 88. (h) Fromowitz, M. *Biopolymers* **1990**, *30*, 1011. (i) Hall, D.; Pavitt, N. *Biopolymers* **1984**, *23*, 1441. (j) Hall, D.; Pavitt, N. *Biopolymers* **1985**, *24*, 935. (k) Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1987**, *26*, 137. (l) Howard, A. E.; Kollman, P. A. *J. Med. Chem.* **1988**, *31*, 1669.

(2) (a) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611. (b) Nayeem, A.; Vila, J.; Scheraga, H. A. *J. Comput. Chem.* **1991**, *12*, 594. (c) Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, H.; Levinthal, C. *Biopolymers* **1987**, *26*, 2053. (d) Chang, G; Guida, W.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379. (e) Gulda, W. C.; Bohacek, R. S.; Erion, M. D. *J. Comput. Chem.* **1992**, *13*, 214. (f) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; John Wiley: New York, 1988. (g) Havel, T. F. *Progr. Biophys. Mol. Biol.* **1991**, *56*, 43. (h) Peishoff, C. E.; Dixon, J. S. *J. Comput. Chem.* **1992**, *13*, 565. (i) Wilson, S. R.; Cui, W. *Tetrahedron Lett.* **1988**, *29*, 4373. (j) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. *J. Comput. Chem.* **1991**, 12, 342. (k) Guarnieri, F.; Wilson, S. R. *Tetrahedron* **1992**, *48*, 4271. (l) Gotō, H.; Ōsawa, E. *J. Am. Chem. Soc.* **1989**, *111*, 8950. (m) Gotō, H.; Ōsawa, E. *Tetrahedron Lett.* **1992**, *33*, 1343. (n) Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371. (o) Ferguson, D. M.; Glauser, W. A.; Raber, D. J. *J. Comput. Chem.* **1989**, *10*, 903. (p) Saunders, M. *J. Am. Chem. Soc.* **1987**, *109*, 3150. (q) Saunders, M. *J. Comput. Chem.* **1991**, *12*, 645. (r) Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419. (s) Gotō, H.; Ōsawa, E. *J. Chem. Soc., Perkin Trans. 2* **1993**, 187. (t) Kolossváry, I.; Guida, W. C. *J. Am. Chem. Soc.* **1993**, *115*, 2107.

(3) (a) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1391. (b) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1986**, *25*, 1547. (c) Wolfe, S.; Bruder, S.; Weaver, D. F.; Yang, K. *Can. J. Chem.* **1988**, *66*, 2703. (d) Wolfe, S.; Yang, K.; Bowers, R. J.; Shin, H.-S.; Sohn, C.-K. *Heterocycles* **1989**, *28*, 639.

0002-7863/94/1516-9860$04.50/0

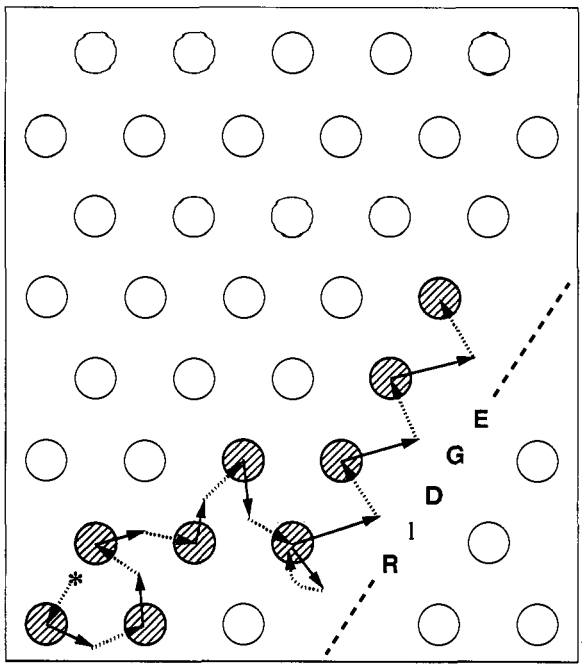


Figure 2. A random walk over a multiminimum energy surface with two domains separated by a wide energy ridge. This ridge cannot be surmounted by a small distortion, and one of the domains remains unexplored.

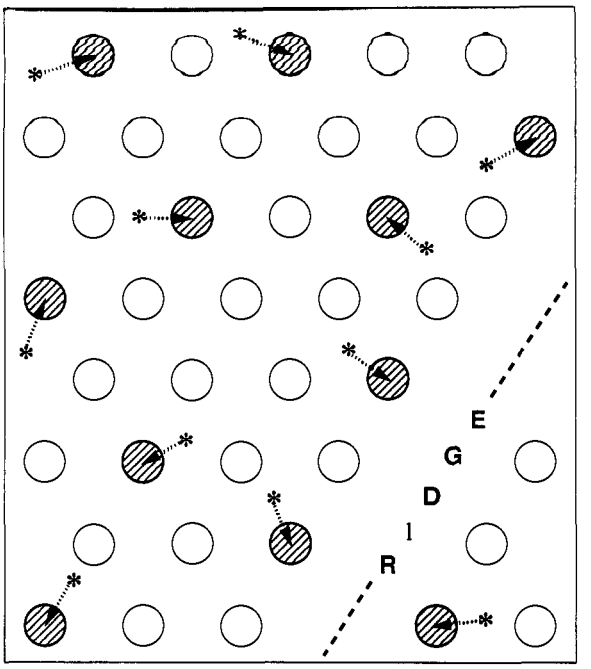


Figure 3. Minimization from a set of initial structures (*) randomly distributed over the entire conformational space allows each domain to be explored.

small region of the space; then, as the duration of the search increases, the size of this region also increases.

This strategy has produced excellent results,[2m,q,r] but there are some potential pitfalls. Because algorithms of this type explore only an exponentially small fraction of conformational space,[4] even a moderately sized molecule would require an unacceptably long computing time. If a search has to be truncated, the starting region selected initially may not yet have expanded to sample all of the important conformations. It is also possible that a conformational space consists of several disjoint regions separated by wide barriers, e.g., Figure 2, which cannot be traversed by a *small* distortion. Only the starting region will then be explored.

Random search procedures of the second type[3] are free of this problem. They are based on a "deductive" strategy, in which *every* probe conformation is selected independently (Figure 3).

Although the deductive strategy, or some combination of the two, would seem to offer certain advantages, only inductive methods have so far been used to treat cyclic compounds. An obvious reason for this choice is that a small distortion of a cyclic structure will produce another cyclic structure, but the assignment of a randomly selected set of dihedral angles to a chain of atoms does not necessarily lead to a ring. It follows that any attempt to incorporate the deductive approach into an overall strategy for such compounds should begin with an algorithm for the *de novo* production of the cyclic probe structures.

There are, at present, three such methods.[1c-k,5] In the first,[1c,d] cyclic structures are selected from a set of randomly generated acyclic structures whose termini are separated by a distance that is close to the length of a bond (Figure 4). The limitation of this procedure is that the average distance between the termini of a random chain of N atoms is proportional to N.[6] The number of cyclic structures generated with this strategy will therefore decrease as N increases.

The data of Table 1 refer to N-membered cyclic hydrocarbons

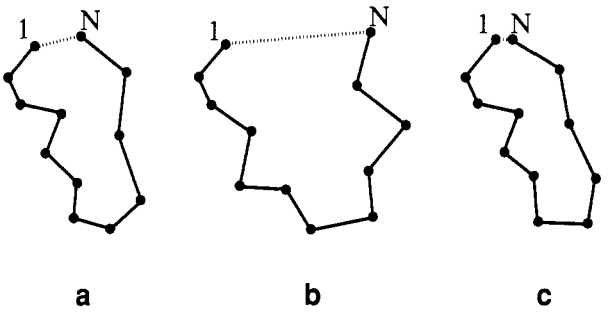


Figure 4. Selection of ring structures from randomly generated open chains: (a) the structure is acceptable; (b) the termini are too far apart; (c) the termini are too close.

Table 1. The Number of Cyclic Probe Structures (ν) and the Time To Produce One Such Structure (τ), Using Different Ring-Making Procedures[a]

ring size (N)	simple selection		early selection		present method	
	ν	τ (s)	ν	τ (s)	ν	τ (s)
10	510	0.08	430	0.07	36 500	0.003
15	280	0.24	250	0.21	39 100	0.004
20	210	0.45	180	0.43	39 700	0.006
30	110	1.4	90	1.4	38 600	0.01
50	50	5	50	5	35 000	0.02
100	20	31	10	40	28 700	0.06

[a] Average of three runs of 100 000 trial conformations, with an average 1–N distance threshold of ±0.3 Å and (N – 1)–N–1 and N–1–2 angle thresholds of ±40°.

in which the criteria for ring closure are a distance of 1.54 ± 0.3 Å between atoms 1 and N and (N – 1)–N–1 and N–1–2 bond angles of 114 ± 40°.[7] These criteria have been applied to 100 000

(4) Slade, G. *The Mathematical Intelligencer*. **1994**, *16*, 29. Madras, N.; Sokal, A. D. *J. Stat. Phys.* **1987**, *47*, 573.

(5) (a) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178. (b) Palmer, K. A.; Scheraga, H. A. *J. Comput. Chem.* **1991**, *12*, 505. (c) Bruccoleri, R. E.; Karplus, M. *Macromolecules* **1985**, *18*, 2767.

(6) Flory, P. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953.

(7) The C–C–C bond angles in MM2- and MM3-minimized cyclic hydrocarbons range from 111 to 117°. A deformation of 40° in C–C–C bond angles is energetically equivalent to a deformation of 0.3 Å in C–C bond lengths.[8,9]

(8) MMP2(85): Available from Quantum Chemistry Program Exchange, Indiana University, Bloomgington, IN, 1985. See: Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.

(9) MM3: Operating Instructions for MM3 Program. Technical Utilization Corporation, 235 Glen Village Court, Powell, OH 43065, 1989. See: Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

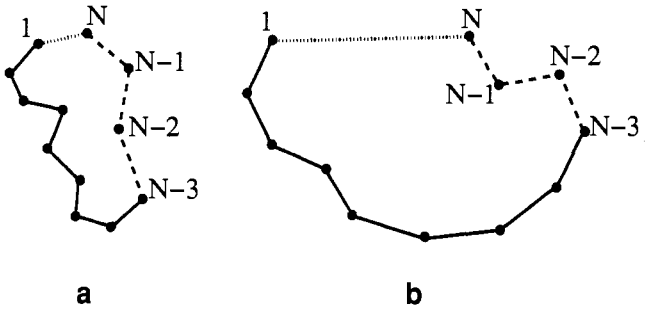


Figure 5. Linking of atoms 1 and $N-3$ by an analytically defined[5a] three-atom bridge: (a) termini 1 and $N-3$ are close enough to permit at least one analytical solution; (b) termini 1 and $N-3$ are too far apart.

conformations generated in different ways by random variation of the dihedral angles in 30° increments. For each method reported in this table the results shown are the average of three independent searches and are expressed in two ways: ν, the number of cyclic structures found in the search; and τ, the time required to find one cyclic structure, using one processor of a Silicon Graphics R-4D380S system. No attempt is made at this stage to determine how many of the structures are unique.

The "simple selection" data refer to the cyclic structures found in a randomly generated set of acyclic structures. As expected,[6] ν decreases and τ increases with ring size. Thus, the probabilities that rings will be found in the random generation of 10-membered and 30-membered chains are 5×10^{-3} and 1×10^{-3}, respectively. In the case of a molecule with more than one ring, the probability that these rings will be found from the random generation of chain structures becomes the product of the individual probabilities. For example, crambin, a 46-residue peptide which contains 8-, 29-, and 31-membered disulfide rings,[10] will exhibit a probability of *ca* 5×10^{-9}, i.e., *only 5 of 1 000 000 000 randomly generated structures will have acceptable initial geometries.*

One might expect that the situation would be improved by the imposition of an analytical ring-closure condition.[1i-k,5] Once the first $N-6$ torsional angles have been specified, the final torsional angles (Figure 5a) might be determined by solution of an equation[5] which expresses the ring-closure condition. However, since the existence of a solution requires that the distance between atoms 1 and $N-3$ be sufficiently small (*cf.* Figure 5b), the efficiency of this procedure for an N-membered ring will not differ greatly from that obtained by the application of "simple selection" to an $(N-3)$-membered ring, and this efficiency will decrease as the size of the ring increases.

An "early rejection" strategy[1e-h] attempts to establish at an early stage whether a growing chain can eventually have termini within bonding distance. After n atoms of an N-membered chain have been connected, a comparison of the distance between atoms 1 and n with d_{max}, the maximum length available for the fully unfolded remainder of the chain (Figure 6), determines whether an N-membered ring can be created. If this distance is greater than d_{max} (Figure 6a), the remainder of the chain is too short to allow eventual ring closure, and this structure is rejected. If the distance is less than d_{max}, the position of atom $n+1$ is specified, and the rejection condition is again imposed. Although the productivity, ν, cannot be increased by this procedure, one might expect a significant decrease in τ.

However, as can be seen in Table 1 under "early selection", there is only a marginal reduction in τ. The reason for this is that rejected structures comprise most of the structures that are produced, and most of the time is used to produce these structures.

In the present work, the problems associated with the *de novo* production of cyclic structures have been overcome by a strategy in which the range of torsions compatible with ring closure is first established and is then *followed by a random search within this range.* As can be seen under "present method" in Table 1, the algorithm described below leads to a substantial increase in ν and a concomitant substantial decrease in τ.

(10) Whitlow, M.; Teeter, M. M. *J. Am. Chem. Soc.* **1986**, *108*, 7163.

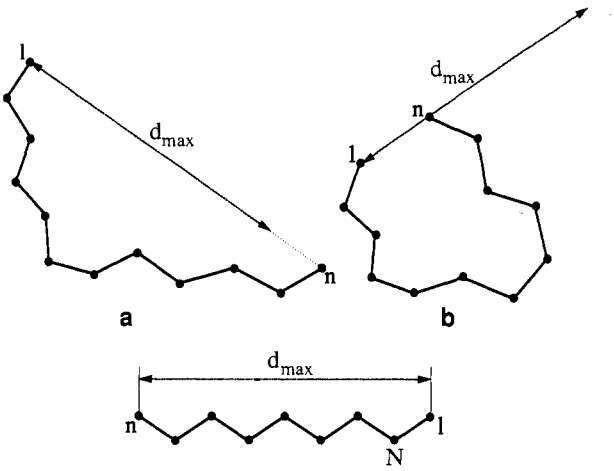


Figure 6. Early rejection of an acyclic structure; d_{max} is the length of the unfolded portion of the chain containing atoms n to N. In (a) the distance between 1 and n is greater than d_{max}; in (b) the distance between 1 and n is less than d_{max}.

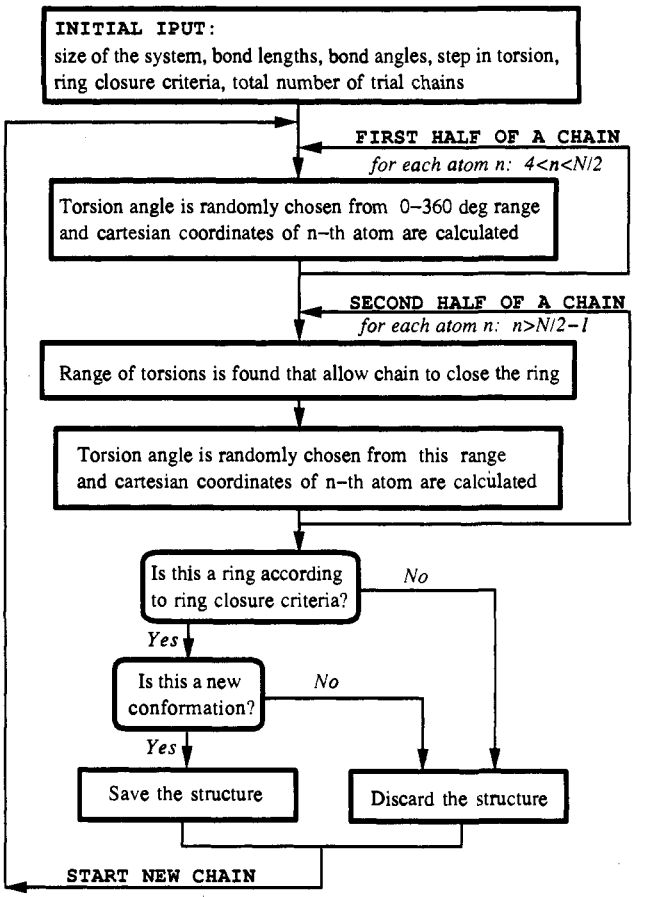


Figure 7. Flow chart of an algorithm for the production of cyclic probe structures.

Algorithm

Generation of Cyclic Structures. Figure 7 is a flow chart of our ring-making procedure. We begin with the recognition that the dihedral angles which determine the positions of atoms 1 to $N/2$ in an N-membered ring require no constraint; the problem is to determine the appropriate constraints upon the dihedral angles of the second half of the chain (atoms $N/2$ to N).

Given the coordinates of the first $n-1$ atoms of the chain, a proper range of torsion angles for atom n ($n > N/2$) is found as

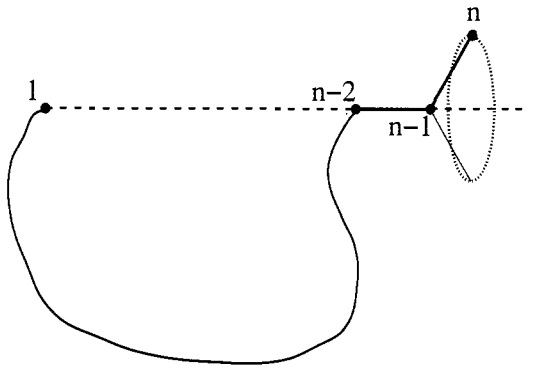


Figure 8. An arrangement that places no constraints upon the range of torsions available to atom *n*.

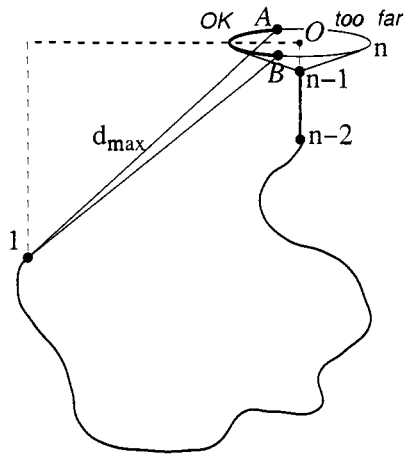


Figure 9. Long-distance constraint: torsion angles outside arc *AB* do not permit ring closure and are avoided. The distance constraint d_{max} is the same as in Figure 6.

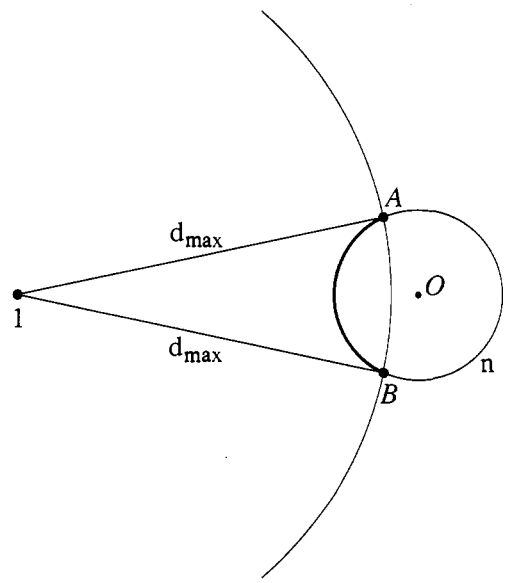


Figure 10. The range of torsions available to atom *n*: the special case in which atom 1 lies in the plane traced by the rotations of atom *n*.

follows. If atoms 1, $(n-2)$, and $(n-1)$ are collinear (Figure 8), all possible positions of atom *n* will be equidistant from atom 1. No constraint will be necessary in this case, and torsion about the bond between atoms $(n-2)$ and $(n-1)$ may take any value between 0 and 360°. In the general case (Figure 9), it is required that the distance between atoms 1 and *n* be less than d_{max} (Figure 6). If atom 1 lies in the plane of the circular path traced by atom *n* (Figure 10), the limits *A* and *B* of the allowed range are the intersection points of this path with the circle of radius d_{max}

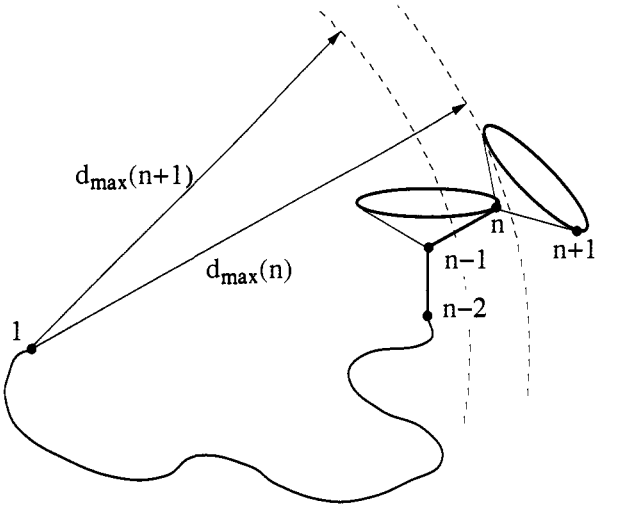


Figure 11. The case in which d_{max} alone fails to select the proper range of torsions.

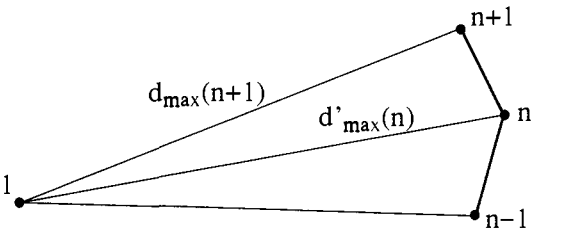


Figure 12. Calculation of the secondary long-distance constraint $d'_{max}(n)$ for atom *n* from the primary long-distance constraint $d_{max}(n+1)$ for atom $(n+1)$.

centered at atom 1. In the more general case, arc *AB* is defined by the intersection of the path with the sphere of radius d_{max} centered at atom 1. Once the allowed range (arc *AB*) of torsions has been decided, the torsional angle that determines the position of atom *n* can be selected randomly from within this range; the process just described is then repeated for atom $(n+1)$.

However, since the magnitude of d_{max} necessarily decreases as *n* increases, it is possible, as depicted in Figure 11, that the path traced by atom *n* lies entirely within the sphere of radius $d_{max}(n)$, but the path traced by atom $(n+1)$ lies entirely outside the sphere of radius $d_{max}(n+1)$. In this event, no torsional angle for atom $(n+1)$ will lead to ring closure.

To avoid this situation an additional constraint $d'_{max}(n)$ upon atom *n* is necessary. This is defined as the maximum distance between atoms 1 and *n* that is compatible with the constraint $d_{max}(n+1)$ for atom $(n+1)$. Given a position of atom $(n-1)$ and a distance $d_{max}(n+1)$ between atoms 1 and $(n+1)$, the maximum separation $d'_{max}(n)$ between atoms 1 and *n* can be achieved if 1, $(n-1)$, *n*, and $(n+1)$ are coplanar (Figure 12). Subsequently, the more restrictive of the constraints, $d_{max}(n)$ and $d'_{max}(n)$, is selected when the allowed range of $(n-2)-(n-1)$ dihedral angles is specified.

Although the long-distance constraints d_{max} and d'_{max} overcome the problem of Figure 4b, their imposition can create the new problem depicted in Figure 4c. Therefore, in the final stages of the procedure, when atoms $N-1$ and N are reached, torsions are restricted to positions *A* and *B* only (Figure 9).[11]

Removal of Internal Close Contacts. The numbers shown in Table 1 refer to cyclic structures generated as described above and do not take into account the transannular close contacts between pairs of atoms that may develop in the interior of a growing chain. Following Smith and Veber,[1f] we define a "close contact" as an interatomic distance less than 70% of the sum of

(11) Alternatively, the analytical procedures described in ref 5 might be incorporated at this stage.

Table 2. The Abilities of Different Ring-Making Procedures To Produce Cyclic Structures That Have No Transannular Close Contacts[a]

ring size	simple selection[b]		early selection[b]		present method[c]	
	ν	τ (s)	ν	τ (s)	ν	τ (s)
10	100	0.4	75	0.4	9500	0.01
15	30	2.5	20	2.5	5300	0.03
20	12	8.1	7	11	3800	0.06
30	2	75	1.5	85	1800	0.2
50	0.03	1×10^4	0.03	1×10^4	330	2.2
100	0		0		5	400

[a] Average of three runs, with distance and angle thresholds as in footnote *a* of Table 1. [b] Rescaled by 0.1 from the results for 1 000 000 trial structures. [c] Calculated for 100 000 trial structures.

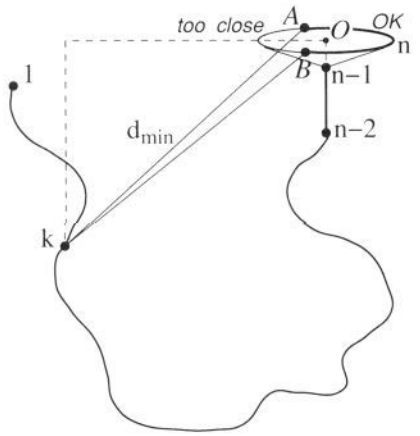


Figure 13. van der Waals constraint: torsion angles outside arc *AB* bring atom *n* too close to atom *k*; d_{min} is 70% of the sum of the van der Waals radii of atoms *n* and *k*.

Table 3. Efficiency of Production of Cyclic Structures That Contain No Transannular Interactions[a]

ring size	without constraints		with constraints	
	ν	τ (s)	ν	τ (s)
10	960	0.011	830	0.024
15	530	0.035	1500	0.027
20	390	0.065	1450	0.047
30	180	0.22	1340	0.11
50	33	2.2	1120	0.34
100	1	200	770	1.8

[a] Average of three runs of 10 000 trial structures.

the van der Waals radii of two atoms of the chain. Table 2 lists the new values of ν and τ that result when the structures of Table 1 are examined, and all those that contain one or more close contacts are rejected.

A comparison of Tables 1 and 2 illustrates a point made earlier: a procedure in which structures are first generated and then rejected is not efficient. The identification of close contacts was, therefore, introduced at an earlier stage. In the new algorithm summarized in Figure 7a,[12] the range of dihedral angles available to atom *n* of a growing chain is restricted to values that avoid close contacts of atom *n* with atoms $k < n$, as explained in Figure 13 (*cf.* Figure 9). Typical results are shown in Table 3.

Although, in principle, *all* transannular contacts can be eliminated from the cyclic probe structures, it is too restrictive to proceed in this manner. For example, Table 4 shows the effect of elimination of only the first close contact. The effect on the *distribution* of the close contacts is particularly striking, as can be seen for 50-membered rings in Figure 14. The corresponding data for 10-, 15-, 20-, 30-, and 100-membered rings are shown in Figures 14a–e.[12]

(12) This figure can be found in the supplementary material.

Table 4. Average Number of Close Contacts in the *N*-Membered Rings Produced without van der Waals Constraints and When the First Possible Close Contact Is Avoided[a]

ring size	av no. of close contacts	
	without constraint	*with constraint*
10	1.8	1.7
15	6.3	2.6
20	10	3.4
30	19	4.6
50	35	6.1
100	78	8.4

[a] Average of three trials of 10 000 structures.

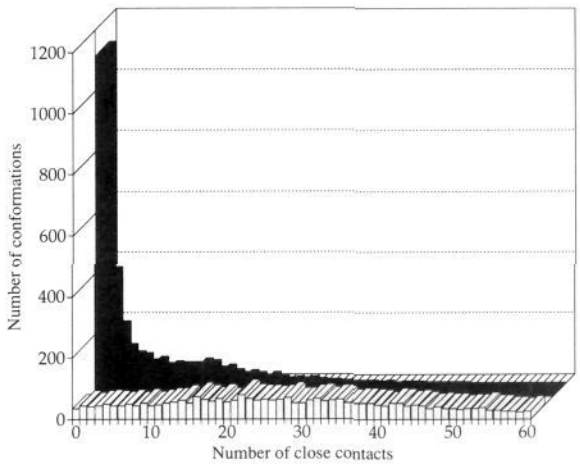


Figure 14. The distribution of close contacts in 50-membered rings generated with the algorithm of Figure 7 (light) and with the first close contact avoided (dark). The corresponding results for 10-, 15-, 20-, 30-, and 100-membered rings are shown in Figures 14a–e.[12]

Three versions of the algorithm were, therefore, developed, which differ in the degree to which close contacts are taken into account in the production of the probe structures: **Version 1** is the ring-making procedure already described. No van der Waals constraints are applied, and all cyclic probe structures are carried forward to the next stage of analysis. **Version 2** avoids the first possible close contact according to the scheme shown in Figure 13. Subsequently, all cyclic structures are retained to the next stage, regardless of the number of additional close contacts. **Version 3** excludes all probe structures that have close contacts.

The probe structures produced in the manner described above are characterized by their internal coordinates. The conversion of these to Cartesian coordinates, the specification of atom types for the MM2[8] and/or MM3[9] force fields, and the preparation of input data files for MM3 are then straightforward. The nature of the compound to be examined is specified at the beginning of the procedure by the specification of the bond lengths and bond angles of the probe structures and then by the specification of the atom types and attached atoms for the conversion to MM2/MM3. All minima have been characterized by vibrational analysis. Many transition states have also been characterized; these will be discussed elsewhere.

Table 5 shows the results of calculations on cyclononane (C9), cyclodecane (C10), cycloundecane (C11), and cyclododecane (C12), using each of Versions 1–3 to produce the initial structures. As expected from the design of the program, the average number of close contacts in the initial structures is Version 1 > Version 2 > Version 3. The higher initial energies from Versions 1 and 2 reflect the greater number of close contacts in the starting structures produced by these procedures. For the data of Table 5, a higher range of initial energies leads to a higher range of final energies but, despite the differences in the range of energies of the structures produced by the different versions, in each case the global minimum is included among these structures. In the general

Table 5. Average Number of Close Contacts in the Initial Structures and Average Initial and Final MM2 Energies of the Cyclic Hydrocarbons Generated Using Different Degrees of van der Waals Constraint.[a,b]

	av no. of close contacts per structure			av initial MM2 energy			av final MM2 energy			
compd	V1[c]	V2[d]	V3[e]	V1	V2	V3	V1	V2	V3	energy of the global min
C9	1.1	1.1	0	860	850	175	32	30	25	23.4
C10	1.8	1.6	0	1020	650	180	69	51	27	24.5
C11	2.8	1.5	0	1530	880	200	64	47	27	24.2
C12	3.7	1.4	0	1950	700	215	52	36	27	20.6

[a] Data obtained from one run using 100 000 trial structures. [b] Energies are in kcal/mol. [c] van der Waals interactions are ignored (Version 1). [d] The first possible close contact is avoided (Version 2). [e] All transannular close contacts are excluded (Version 3).

Table 6. Selection of Unique Minima[a]

	version 1			version 2			version 3		
ring size	crude[b]	weak[c]	strong[d]	crude	weak	strong	crude	weak	strong
10	3691	3072	2891	4493	786	776	834	334	324
15	3943	3943	3943	3844	3616	3616	1502	1495	1495
20	3972	3972	3972	3618	3616	3616	1452	1452	1452
30	3890	3890	3890	3479	3479	3479	1344	1344	1344
50	3503	3503	3503	3160	3160	3160	1115	1115	1115
100	2904	2904	2904	2775	2775	2775	769	769	769

[a] Average of three runs for 10 000 trial conformations. [b] Number of cyclic structures before selection. [c] Number of cyclic structures after weak selection for repeats. [d] Number of cyclic structures after strong selection for repeats.

case, we suggest that the low-energy structures produced by Version 3 are more likely to include the global minimum, but the structures produced by Version 1 will encompass more of the conformational space.

Selection of Unique Structures. To identify and eliminate duplications, each newly generated structure was compared to all previously identified unique structures, using four tests: (i) direct comparison; (ii) comparison to the mirror image; (iii) comparison with the reverse numbering; and (iv) comparison to the mirror image with the reverse numbering. If all torsion angles of two structures coincided to within ±15° in any of these comparisons, the structures were considered to be equivalent, and the new structure was rejected. This procedure is termed "weak selection".

A more rigorous comparison was also applied. Within each of the four subtests, all cyclic permutations of the sets of torsional angles, representing a total of $4N$ comparisons for each pair of N-membered ring structures, were examined. This procedure is termed "strong selection". The data of Table 6 indicate that the numbers of structures selected for minimization by the "weak" and "strong" criteria do not differ greatly, so that, in the initial stages, the simpler and faster "weak selection" can be recommended as a reliable test of the uniqueness of a probe structure for minimization. In the final stages, however, because C_n symmetry is more likely to be encountered in the minimized structures, "strong selection" is more appropriate.

Performance of the Method

The foregoing strategy produces large numbers of cyclic probe structures. To determine how these are distributed over the conformational space, we optimized the probe structures to find the energy minima of the C9–C12 saturated hydrocarbons[1e,2l,o,q,t,u] on their MM2 and MM3 energy surfaces (Tables 7 and 8). The dynamics of the accumulation of these minima are shown in Tables 9 and 10. We also examined cycloheptadecane (C17), using MM2 (Table 11). In the latter case, the calculations were terminated after 140 000 probe structures had been minimized. At this point the 262 0–3 kcal/mol minima[2r] had been found, but the number of higher energy structures was still increasing.

The data of Tables 7 and 8 show that the present method provides the most complete coverage of the conformational space. As seen in Table 9, only 50, 100, 1000, and 2500 probe structures are required to find *all* previously determined C9, C10, C11, and C12 MM2 minima. Referring to Table 10, all C9, C10, C11, and C12 MM3 minima previously reported by Saunders[2q] are found using 25, 50, 1000, and 2500 probe structures, respectively.

Table 7. The Number of MM2 Minima Found for Cycloalkanes by Different Methods

ring size	ref 21	ref 2s	ref 1e	ref 2o	ref 2q	ref 2t	present work
9	5	6	6	7	8	8	8[a]
10	15	17	17		18	18	18[b]
11	27	32			40	38	41[c]
12	87	96			111	117	121[d]

[a] Two high-energy minima containing pyramidal carbon were also found (see text). [b] Fourteen high-energy minima containing pyramidal carbon were also found. [c] Sixty-nine high-energy minima containing pyramidal carbon were also found. [d] More than 170 high-energy minima containing pyramidal carbon were also found.

Table 8. The Number of MM3 Minima Found for Cycloalkanes by Different Methods

ring size	ref 2q	present work
9	8	9
10	16	18
11	29	32
12	89[a]	97[a,b]

[a] Minimum 44, reported by Saunders to have an imaginary frequency of 6.4i cm^{-1}, was also found in the present work and is considered to be a transition state. [b] One of the new structures is a trefoil knot (see text).

Methylene-Inside Conformations. Saunders found that both MM2 and MM3 produce a C9 conformation in which one methylene group is turned toward the inside of the ring.[2q] Numerous new structures of this type, containing one, two, or three methylene-inside groups were found in the present work, using both MM2 and MM3. These are collected in Figures 15–20.

Pyramidal Carbon, an Artifact of MM2. As seen in the footnotes to Table 7, there are many more C9–C12 minima on the MM2 energy surface than on the MM3 energy surface. Most of these additional structures contain a pyramidal carbon atom. This was also seen in the work of Saunders,[2q] who found several such structures using both MM2 and MM3. A typical example from the present work is shown in Figure 21. This MM2 structure is obtained by minimization of the probe structure shown in Figure 22. Since the same initial structure is minimized by MM3 to the methylene-inside conformation of Figure 15, pyramidal MM2 "minima" appear to be an artifact of the MM2 force field.

The reason for this can be understood by using methane as an example. As expected, in both force fields the tetrahedral

Table 9. The Number of MM2 Minima of the C9–C12 Cycloalkanes as a Function of the Number of Cyclic Probe Structures Minimized[a]

no. of probe structures minimized	no. of different minima[b]			
	C9	C10	C11	C12
10	7	6	7	8
25		13	17	19
50	8	16	22	33
100	9	18	24	55
250			31	82
500			39	98
1000	10	19	43	114
2500		21	49	131
5000		27	56	147
10000		32	72	172
15000			85	201
20000			97	221
25000			104	245
30000			110	265

[a] 30 000 probe structures were generated. [b] The numbers shown include minima containing pyramidal carbon.

Table 10. The Number of MM3 Minima of the C9–C12 Cycloalkanes as a Function of the Number of Cyclic Probe Structures Minimized

no. of probe structures minimized	no. of different minima			
	C9	C10	C11	C12
10	7	6	8	9
25	8	12	16	21
50		18	18	30
100			21	44
250			26	68
500			28	78
1000	9		30	86
2500				90
5000				95
10000			31	
20000			32	96
25000				97

Table 11. The Number of Low-Energy MM2 Conformations of Cycloheptadecane That Are Found Using Different Methods

method	energy range (kcal/mol)		
	0–1	0–2	0–3
Cartesian coordinate stochastic search[2r]	11	67	222
torsional tree-search[2r]	9	63	211
torsional Monte Carlo, random walk[2r]	11	67	237
torsional Monte Carlo, usage-directed[2r]	11	69	260
flap search[2m]	11	68	249
flap/flip search[2m]	11	69	262
distance geometry[2r]	7	44	176
distance geometry[2h]			223
molecular dynamics[2r]	9	57	169
present method	11	69	262

structure is the global minimum, with zero strain energy. However, pyramidal structures result in both cases when minimization is initiated from a pyramidal conformation. In the case of MM3, the structure is a transition state,[13] but with MM2 the structure is a minimum.

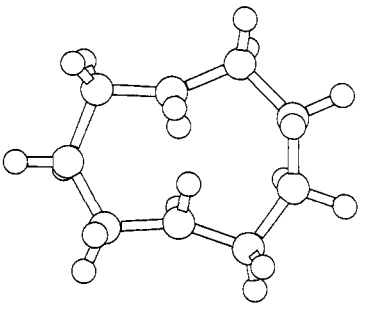
The difference between MM2 and MM3 is caused by their use of different equations to express the bending energy.[8,9] In MM2, the bending energy (eq 1) is an even function of $(\theta - \theta_0)$, and it

$$E_\theta^{MM2} = {}^1/_2 k_\theta(\theta - \theta_0)^2[1 + k_6(\theta - \theta_0)^4]$$

$$k_6 = 7 \times 10^{-8}\ \text{deg}^{-4} \qquad (1)$$

is almost harmonic for $|(\theta - \theta_0)| < 30$–40 deg, whereas in MM3

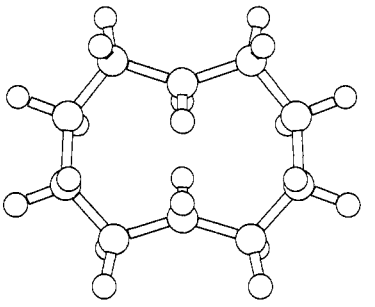
(13) For an ab initio treatment of this point, see: Gordon, M. S.; Schmidt, M. W. *J. Am. Chem. Soc.* **1993**, *115*, 7486.



min 9: 105.22

159.3, 25.3, -42.8, 170.8, -161.3, 56.6, -83.7, 54.3, 179.7

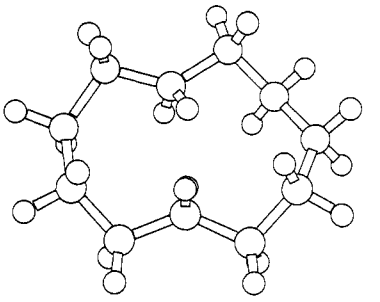
Figure 15. A new MM3 minimum of cyclononane.



min 17: 74.26

61.3, 179.2, -179.2, -61.3, 87.9, -61.3, -179.2, 179.2, 61.3, -87.9

Figure 16. A new MM3 minimum of cyclodecane. A second new minimum is shown in Figure 16a.[12]



min 41: 62.65

-80.9, 55.8, 55.8, -80.9, -170.9, 169.0, 65.2, -87.2, 65.2, 169.0, -170.9

Figure 17. A new MM2 minimum of cycloundecane.

the bending energy (eq 2) is characterized by substantial cubic

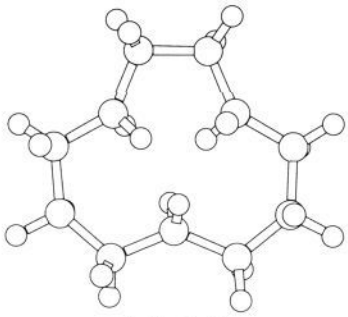
$$E_\theta^{MM3} = {}^1/_2 k_\theta(\theta - \theta_0)^2[1 - k_3(\theta - \theta_0) + \ldots]$$

$$k_3 = 1.4 \times 10^{-2}\ \text{deg}^{-1} \qquad (2)$$

anharmonicity. This makes the pyramidal structure unstable with respect to its transformation to the tetrahedral structure and explains why none of our MM3 minima contain pyramidal carbon(s).

Cycloheptadecane. A Challenge from Martin Saunders. Saunders has commented that "looking for 'all' of the possible conformers is a more difficult test for a search method (than finding the lowest energy conformers) and therefore should serve better for comparison of different procedures."[2q,14] The 262 0–3 kcal/mol MM2 minima reported previously for C17[2m,r] (Table 11) were accmulated after 60 000 probe structures containing no close contacts (Version 3) had been minimized. At this point there were also 1342 minima in the 3–5 kcal/mol range, 6879

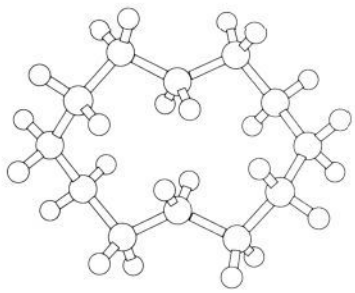
(14) Saunders, M. Personal communication to the authors.



min 32: 81.51

168.3, -26.5, 168.3, -178.8, 53.6, -48.2,
-158.2, -158.2, -48.2, 53.6, -178.8

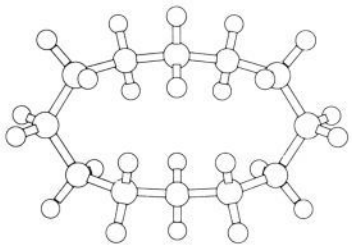
Figure 18. A new MM3 minimum of cycloundecane. Two additional new minima are shown in Figures 18a and 18b.[12]



min 121: 50.14

-71.9, 61.0, 61.0, -71.9, -162.3, -175.4,
-79.5, 60.6, 60.6, -79.5, -175.4, -162.3

Figure 19. A high-energy MM2 minimum of cyclododecane. Nine additional new minima are shown in Figures 19a–i.[12]



min 75: 39.32

123.9, -44.8, -44.8, 123.9, -160.0, 160.0,
-123.9, 44.8, 44.8, -123.9, 160.0, -160.0

Figure 20. A new MM3 minimum of cyclododecane. Six additional new minima are shown in Figures 20a–f.[12]

minima in the 5–10 kcal/mol range, and 1956 higher energy minima.

The minimization of an additional 80 000 probe structures, also generated with Version 3, did not increase the number of 0–3 kcal/mol minima; however, an additional 26 3–5 kcal/mol, 1286 5–10 kcal/mol, and 762 higher energy minima were found. The distribution of the complete set of C17 structures obtained from the minimization of the 140 000 probe structures is summarized in Figure 23 and comprises our current response to the challenge issued by Professor Saunders.[15]

The 140 000 probe structures were not independently subjected to MM3 minimization, because the present calculations were

(15) As mentioned, only Version 3 was applied to C17, because the principal objective here was to find all previously reported 0–3 kcal/mol minima. We expect that this version of the close contacts constraint causes higher energy minima to be underrepresented and that relaxation of this constraint will alter the distribution of minima by an increase in the number of higher energy structures.

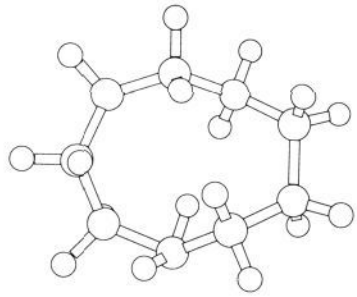


Figure 21. An artifact of the MM2 minimization of cyclononane which contains a pyramidal carbon atom.

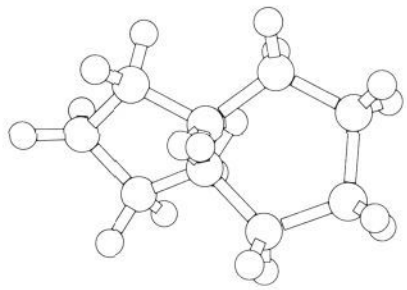


Figure 22. A cyclic probe structure of cyclononane that minimizes to the artificial MM2 structure of Figure 21.

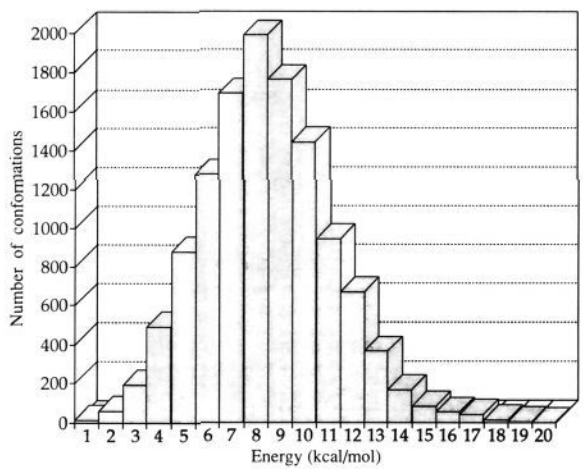


Figure 23. Distribution of the MM2 minimum energy conformations of cycloheptadecane.

Table 12. Distribution of the 262 0–3 kcal/mol Structures of Table 11 after Reminimization Using MM3

MM2 distribution		MM3 distribution				
energy range	no. of structures	0–1	1–2	2–3	3–4	4–5
0–1	11	3	8			
1–2	58		16	41	1	
2–3	193			64	128	1
0–3	262	3	24	105	129	1

directed principally toward comparisons to the earlier benchmarks discussed in Tables 9–11. The 262 low-energy MM2 structures were, however, reminimized with MM3. The new distribution of the 262 structures is shown in Table 12.

A Cyclododecane Knot. The highest energy structure identified on the MM3 energy surface of C12 was found to lie 1000 kcal/mol above the global minimum but was, nevertheless, identified as a minimum by frequency analysis. The reason for its stability toward the minimization procedure is that the structure is a trefoil knot (Figure 24). Such a structure has no physical significance in this case, because the C–C bonds are too long (1.7–2.0 Å) and

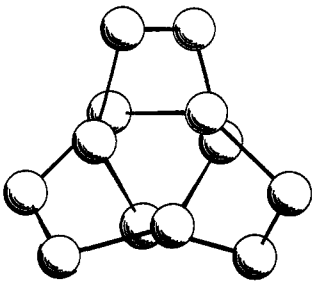


Figure 24. Trefoil knot conformation of cyclododecane (hydrogens have been deleted for clarity). An MM3 input file for the optimized structure is included in the supplementary material.

many of the distances between nonbonded atoms are too small. The structure is stable only within a molecular mechanics force field having a predetermined and fixed topology of bonding. Nevertheless, the discovery of this structure is an important result, because of its relevance to the situation noted in Figure 2. It exemplifies our ability to reach remote, unexpected regions of conformational space. Although, in the case of C12, the high energy of the knot makes this structure attainable but physically meaningless, with larger rings ($N > 50$),[16] the numbers, varieties, and relative stabilities of the knots discovered by our procedure increase substantially,[17] and their physical significance becomes comparable to that of unknotted[16] structures.

The trefoil knot is obtained only with MM3. Figure 25 shows the initial C12 conformation which minimizes to the knot. With MM2, optimization of the same initial structure produces one of the low-energy minima (the fifth in the sequence, with an energy of 23.22 kcal/mol), also shown in Figure 25.

Summary and Conclusions

The strategy developed here for the conformational analysis of cyclic structures seems to have numerous advantages. The programming is simple, fast, and efficient, and the process is data-independent. The method is therefore suitable for parallel processing and vectorization, in contrast to the data-dependent Markov-type procedures.[2] The method produces starting points for minimization that randomly cover all of the conformational space. Because it can be modified to produce a non-uniform sampling of this space, the method will be applicable to importance-sampling Monte Carlo calculations of cyclic molecules (e.g., ref 3a). Finally, in addition to the ring-closure and van der Waals constraints employed here, other kinds of distance constraints, e.g., NMR constraints, can also be taken into account.

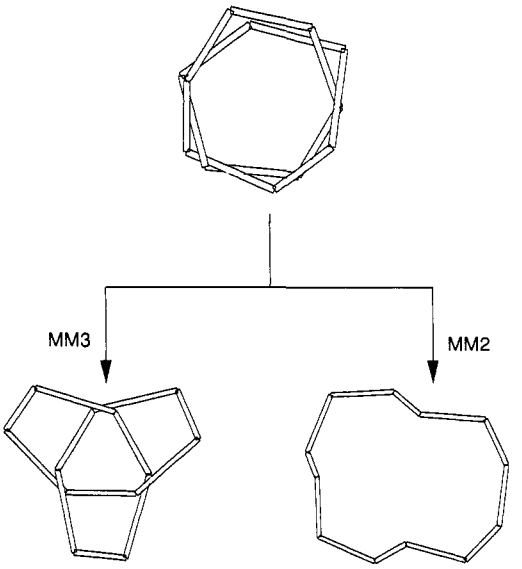


Figure 25. The initial C12 structure (top), which minimizes to a trefoil knot (lower left) using MM3 and to minimum number five (lower right) using MM2.

Acknowledgment. This work was supported by the Operating and Equipment Grants Programmes of the Natural Sciences and Engineering Research Council of Canada. The authors thank Professors K. N. Houk and W. C. Still for details of the 262 C17 minima and Professors Kurt Mislow and Martin Saunders for interesting discussions.

Supplementary Material Available: A listing of the MM2 and MM3 energies of all C9–C12 minima found in this work, Figures 7a, 15a–e, 17a, 18a–b, 19a–i, and 20a–f, and an MM3 input for the optimized C12 knot (41 pages). This material is contained in libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.

(16) Frisch, H. L.; Wasserman, E. *J. Am. Chem. Soc.* **1961**, *83*, 3789. Wasserman, E. *Sci. Am.* **1962**, *207* (5), 94.

(17) Weinberg, N.; Wolfe, S.; Mislow, K. To be published. See: Liang, C.; Mislow, K. *J. Am. Chem. Soc.* **1994**, *116*, 3588.