Pseudosystematic Conformational Search. Application to Cycloheptadecane

J. Thomas Ngo*,[†,‡] and Martin Karplus*

Contribution from the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received April 5, 1996. Revised Manuscript Received March 17, 1997[⊗]

Abstract: We describe a deterministic approach to conformational searching. The algorithm is a best-first, depth-first traversal of the *conformer graph*, *i.e.*, a graph generated by a small set of deterministic deformation operators. No deformation operator is ever applied to the same starting structure more than once. Thus, this algorithm avoids the source of diminishing returns, common in Monte Carlo searches, that is caused by the rising probability that a given starting structure is subjected to the same or similar perturbations more than once. We apply this technique to the task of finding all locally optimal conformations of cycloheptadecane whose MM2 energies lie within 3 kcal/mol of its global optimum. This task, which is considered to be challenging for contemporary algorithms and computer hardware, was used by Saunders *et al.* (*J. Am. Chem. Soc.* **1990**, *112*, 1419−1427) as a benchmark for comparing existing techniques in terms of their efficiency and thoroughness. The new algorithm compares favorably, by both criteria, with all of the methods tested by Saunders *et al.* Moreover, the conformer graph can be used for an analysis of the potential-energy surface that is not directly made possible by other search methods. The particular set of deformation operators used is specialized for medium-sized ring molecules. Ways to design deformation operators for larger molecules such as proteins are suggested.

1. Introduction

The observed physical properties of a highly flexible molecule are produced not by a single, globally optimal conformational state, but by a thermal distribution of states that may represent diverse regions of its conformational space. This realization has deeply influenced current understanding of the behavior of biological macromolecules, especially that of proteins.[2] For this reason, enumerating the low-energy conformers of a given molecule is an important task in computational chemistry.

While dynamical simulations may be used to explore the conformational spaces of macromolecules, such explorations can rarely be complete. However, in simpler cases an exhaustive enumeration of low-energy conformers may be possible. Saunders *et al.*[1] have tested a suite of conformational search techniques against one such problem: that of enumerating the conformational states of cycloheptadecane that are within 3 kcal/mol of its global optimum on its MM2[3] energy surface. They identified this problem as “just at the limit of what can be accomplished in a reasonable length of time, [and therefore] a good test of the effectiveness of new methods as they are developed”.

We present a new algorithm for conformational searches and tests of the algorithm against the Saunders *et al.* benchmark. The main criterion for success is thoroughness. Efficiency is also a consideration, but it is secondary; for example, an algorithm that runs 25% faster than the alternatives but misses 10% more of the low-energy conformers may be of little interest. The algorithm that we describe is more thorough and more efficient than the ones tested by Saunders *et al.*[1]

The algorithm is very similar to that used in the CONFLEX3 algorithm of Gotō and Ōsawa,[4] which was published after this work was completed[5] and studied no cycloalkanes larger than cyclododecane. Also closely related are papers by Kolossvàry and Guida[6] and by Koča,[7] which focused on characterizing low-energy interconversion paths (including transition states) in cycloalkanes no larger than cyclododecane and cyclohexane, respectively.

Approaches to conformational enumeration often involve repeating a three-step cycle.[1] First, a crude starting geometry is produced, possibly by using a restricted set of degrees of freedom. Second, the structure is refined by energy minimization with all degrees of freedom permitted to vary. Third, if the resulting energy is acceptable, the structure is compared with previously found conformers to test for possible duplication. Because energy minimization is usually the computationally intensive step, the total execution time is roughly a product of two factors: the average time required to minimize and the total number of structures submitted to minimization. Saunders *et al.* varied both factors simultaneously. We focus solely on the number of crude starting geometries minimized as a measure of efficiency; the time required for subsequent treatment of each crude structure can be optimized independently.

[†] Graduate Biophysics Program of the Committee on Higher Degrees in Biophysics.

[‡] Present address: Interval Research Corp., 1801-C Page Mill Rd., Palo Alto, CA 94304-1216. E-mail: ngo@acm.org.

[⊗] Abstract published in *Advance ACS Abstracts,* May 1, 1997.

[1] Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. Comformations of cycloheptadecane: A comparison of methods for conformational searching. *J. Am. Chem. Soc.* **1990**, *112*, 1419−1427.

[2] Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Proteins. *Advances in Chemical Physics*; Wiley: New York, 1988; Vol. LXXI.

[3] Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V_1 and V_2 torsional terms. *J. Am. Chem. Soc.* **1977**, *99* (25), 8127−8132.

[4] Gotō, H.; Ōsawa, E. An efficient algorithm for searching low-energy conformers of cyclic and acyclic molecules. *J. Chem. Soc., Perkin Trans. 2* **1993**, pages 187−198.

[5] Ngo, J. T. Global Optimization for Articulated Figures: Molecular Structure Prediction and Motion Synthesis for Animation. Ph.D. Thesis, Harvard University, Cambridge, MA, June 1993.

[6] Kolossvàry, I.; Guida, W. C. Comprehensive conformational analysis of the four- to twelve-membered ring cycloalkanes: Identification of the complete set of interconversion pathways on the MM2 potential energy hypersurface. *J. Am. Chem. Soc.* **1993**, *113*, 2107−2119.

[7] Koča, J. Computer program CICADA−travelling along conformational potential energy hypersurface. *J. Mol. Struct: THEOCHEM* **1994**, *308*, 13−24.

S0002-7863(96)01132-8 CCC: $14.00

Saunders *et al.* tested a variety of conformational search techniques. Two of these techniques, distance geometry and molecular dynamics, lagged far behind the other algorithms in terms of both efficiency and thoroughness, when tested on cycloheptadecane. The most effective techniques were stochastic (Monte Carlo) searches, conducted in either Cartesian[1,8] or torsional[9] coordinates, and modified so that a given search step could begin from any previously stored structure, not just the one most recently found. At an intermediate level of effectiveness were torsional tree searches.[10] These seem attractive because in principle they can provide a thorough, uniform coverage of conformational space, have obvious termination criteria, and automatically restrict the search to relevant degrees of freedom. In practice they are prohibitively expensive unless the search tree is pruned, and in general pruning requires *a priori* assumptions that may eliminate low-energy structures that should be included. An objective of the present work is to develop an algorithm that has the strengths of a deterministic algorithm but requires no such *a priori* pruning criteria.

The pseudosystematic algorithm that we describe is a best-first, breadth-first traversal of the conformer graph: a graph whose nodes are local potential-energy minima, and whose arcs represent the action of a small set of deterministic deformation operators. The deformation operators are tailored specifically for the molecule in question and are designed to move the conformational state of the molecule from one local potential-energy minimum to a different nearby neighbor.

A cyclic molecule such as cycloheptadecane presents special challenges for a conformational search algorithm. If bond lengths and angles are taken to be fixed, then the molecule has 17 degrees of freedom, all of which are torsional. The cyclic alkane ring-closure condition imposes 6 constraints, leaving 11 soft degrees of freedom.[11] Thus, all of the low-energy conformational minima of cycloheptadecane can be expected to lie in or near an 11-dimensional submanifold (surface) within its full conformational space of 147 ($3N - 6$) dimensions.

A systematic search is hindered by the difficulty of parametrizing this submanifold. The best preexisting search techniques side step this difficulty by searching in the higher-dimensional search space in which the submanifold is embedded. Deviations from the submanifold are corrected by a combination of pruning based on *a priori* assumptions about acceptable geometries (as in a torsional tree search[10,12]) and energy minimization (as in a torsional tree search or a Monte Carlo search, whether based on torsional[9] or Cartesian[1,8] coordinates).

In the present implementation of the pseudosystematic search algorithm for cycloheptadecane, a deformation operator that we refer to as the "kinematic twist" confines the search directly to the submanifold of closed configurations, entirely without resort to pruning or energy minimization. Instead it employs inverse kinematics, a body of techniques that have found pervasive application in robotics, where an end effector must typically be brought to a given position with a particular orientation.[13] Here, the "end effector" is an imaginary 18th methylene group on the end of *n*-heptadecane, and the ring-closure condition is met if and only if that 18th methylene group is perfectly superimposed on the first, assuming that all bond lengths and angles are constrained to be in their respective local minima.

2. Related Work

Conformational search problems have been approached by a number of methods. While in-depth reviews of these techniques are available elsewhere,[14] it is instructive to situate the pseudosystematic algorithm in relation to some of the promising methods currently available.

Unlike torsional tree searches,[10,12] it begins with a crude starting geometry for the entire molecule and explores conformational space by means of small modifications to the geometry. Like Monte Carlo Minimization,[15,16] it executes each modification by a rapid geometric change followed by energy minimization, and therefore "sees" a discrete space consisting only of local energy minima[17] as opposed to the full continuous energy surface. However, unlike the Monte Carlo method[18] and its variants as applied to molecular structure,[8,9,15,19−22] it uses a small, discrete set of deterministic deformation operators, each tailored specifically for the molecule in question and designed to move the conformational state of the molecule from one local energy minimum to a different nearby neighbor.

In the version of the pseudosystematic algorithm that we have applied to cycloheptadecane, the deformation operators are kinematic twists, which are guaranteed to treat all bond lengths and angles as holonomic constraints. The idea of giving special treatment to stiff degrees of freedom in a molecule is not new. It has been employed in a number of techniques[23−37] that

[8] Saunders, M. J. Stochastic exploration of molecular mechanics energy surfaces. Hunting for the global minimum. *J. Am. Chem. Soc.* **1987**, *109*, 3150.

[9] Chang, G.; Guida, W. C.; Still, W. C. An internal coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *11*, 4379.

[10] Lipton, M.; Still, W. C. The multiple minimum problem in molecular modeling. Tree searching internal coordinate conformational space. *J. Comput. Chem.* **1988**, *9* (4), 343−355.

[11] Gō, N.; Scheraga, H. A. Ring closure and local conformational deformations of chain molecules. *Macromolecules* **1970**, *3* (2), 178−187.

[12] Bruccoleri, R. E.; Karplus, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* **1987**, *26*, 137−168.

[13] Korein, J. U. A Geometric Investigation of Reach. *ACM Distinguished Dissertations*; MIT Press: Cambridge, MA, 1985.

[14] Howard, A. E.; Kollman, P. A. An analysis of current methodologies for conformational searching of complex molecules. *J. Med. Chem.* **1988**, *31* (9), 1669−1675.

[15] Li, Z.; Scheraga, H. A. Monte Carlo-Minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611−6615.

[16] Caflisch, A.; Niederer, P.; Anliker, M. Monte Carlo Minimization with thermalization for global optimization of polypeptide conformation conformations in Cartesian coordinate space. *Proteins: Struct., Funct. Genet.* **1992**, *14*, 102−109.

[17] Nayeem, A.; Vila, J.; Scheraga, H. A. A comparative study of the simulated-annealing and Monte Carlo-with-minimization approaches to the minimum-energy structures of polypeptides: [met]-enkephalin. *J. Comput. Chem.* **1991**, *12* (5), 594−605.

[18] Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671−680.

[19] Saunders, M. J. Stochastic search for the conformations of bicyclic hydrocarbons. *J. Comput. Chem.* **1989**, *10* (2), 203−208.

[20] Ferguson, D. M.; Raber, D. J. A new approach to probing conformational space with molecular mechanics: Random incremental pulse search. *J. Am. Chem. Soc.* **1989**, *111* (12), 4371−4378.

[21] Berg, B. A.; Neuhaus, T. Multicanonical algorithms for first-order phase transitions. *Phys. Lett.* **1991**, *B267*(2), 249−253.

[22] Hansmann, U. H. E.; Okamoto, Y. Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple-minima problem. *J. Comput. Chem.* **1993**, *14* (11), 1333−1338.

[23] van Gunsteren, W. F.; Berendsen, H. J. C. Algorithms for macromolecular dynamics of constraint dynamics. *Mol. Phy.* **1977**, *34* (5), 1311−1327.

[24] Swindoll, R. D.; Haile, J. M. A multiple time-step method for molecular dynamics simulations of fluids of chain molecules. *J. Comput. Phys.* **1984**, *53* (2), 289−298.

[25] Teleman, O.; Jönsson, B. Vectorizing a general purpose molecular dynamics simulation program. *J. Comput. Chem.* **1986**, *7* (1), 58−66.

[26] Tuckerman, M. E.; Martyna, G. J.; Berne, B. J. Molecular dynamics algorithm for condensed systems with multiple time scales. *J. Chem. Phys.* **1990 1990**, *93* (2), 1287−1291.

[27] Tuckerman, M. E.; Berne, B. J.; Rossi, A. Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *J. Chem. Phys.* **1990**, *94* (2), 1465−1469.

Table 1. A Brief Glossary of Terms

conformer	a configuration of cycloheptadecane that lies in a local minimum of the MM2 energy function; Synonyms: structure, state, node
encountering a conformer	generating a conformer by energy minimization
visiting a conformer	applying all 34 kinematic twist operators to a conformer
instant global minimum	the conformer of lowest energy that has been encountered prior to any given time during the search
purported global minimum	the 19.09 kcal/mol conformer of cycloheptadecane identified originally by Saunders *et al.*; the conformer of lowest MM2 energy that has been found by any conformational search of cycloheptadecane to date
low-energy conformer	a conformer whose energy is no more than 3 kcal/mol greater than that of the purported global minimum (a high-energy conformer is any conformer that is not a low-energy conformer)
seed	the first low-energy conformer encountered during the pseudosystematic search

improve the computational speed of dynamical simulation or energy minimization without changing its intended behavior.

In contrast with such techniques, the kinematic twist operator maintains the bond-length and bond-angle constraints (and therefore closure) entirely without computation of energies. It is comparable in spirit to the dynamic Monte Carlo (DMC) technique of Morley *et al.*, which (when applied to ring systems) employs "corner-flapping" motions driven by "short bursts of molecular dynamics".[38] It is also related to the "local moves" protein-structure algorithm of Elofsson *et al.*, in which dihedral angles within a window of up to five residues are changed at random, and residues just outside the window are restored to their original positions by least-squares minimization.[39] It is similar both in spirit and in approach to work done by Dudek and Scheraga,[40] in which the objective is to locate the global (ECEPP) energy minimum of a polypeptide. An essential step in that procedure is to generate a collection of backbone structures by deforming a seven-residue subsegment of a given starting structure in a manner that keeps the end points of the subsegments approximately fixed. To generate these closure-preserving deformations, twelve torsional degrees of freedom are permitted to vary: six using grid-based enumeration, and the remaining six using the analytic solution of Gō and Scheraga, which is specialized for 6 degrees of freedom.[11] (We note that the torsional tree-search program CONGEN[12] also employs the Gō−Scheraga analytic technique to complete partial structures.) Unlike the Dudek−Scheraga procedure, which generates a number of structural variants that is exponential in the number of degrees of freedom and concentrates strain in the analytically determined torsional degrees of freedom, the kinematic twist operator generates a linear number of evenly strained structural variants.

Finally, in a technique related to the kinematic portion of the algorithm described here, Crippen[41] described the cycloalkane ring constraints in terms of the metric matrix, each of whose N^2 elements is a dot product between the direction vectors of two C−C bonds. These coordinates permitted the writing of analytic expressions for the submanifolds of closed conformations of cyclohexane and cycloheptane. In contrast, the kinematic twist operator used here is expressed in terms of torsional angles, which are fewer, though less amenable to analytic treatment.

3. Pseudosystematic Search

The pseudosystematic search begins with its only random procedure, the generation of an initial structure, called the seed. The value of each torsion angle is chosen from a uniform, continuous distribution in the range $-180° < \omega \leq 180°$. The structure is closed kinematically (see later section), minimized, placed in the list of known structures, and marked as "unvisited". (A glossary of terms appears in Table 1.) Thereafter, the search proceeds deterministically.

At each iteration of the search (Figure 1), the unvisited structure of lowest energy is marked as "visited". To *visit* the structure, each of the 34 possible kinematic twists (see later section) is applied to it, generating 34 variants. Each variant is minimized, and if it is not a duplicate, it is placed in the list of known structures and marked as unvisited.

In principle, this cycle may be repeated indefinitely. However, the probability of identifying new low-energy conformers decreases with time, particularly after the difference between the highest and lowest energies represented among the visited structures exceeds the desired threshold, in this case 3 kcal/mol. The run described here was terminated soon after that condition was met.

4. Theoretical Framework: The Conformer Graph

We have called this search algorithm "pseudosystematic" for two reasons. First, unlike truly systematic procedures it begins at a randomly chosen point in the search space. Second, the term "systematic" has come to be associated with torsional

[28] Tuckerman, M. E.; Berne, B. J. Molecular dynamics in systems with multiple time scales: Systems with stiff and soft degrees of freedom and with short- and long-range forces. *J. Chem. Phys.* **1991**, *95* (11), 8362−8364.

[29] Tuckerman, M. E.; Martyna, G. J.; Berne, B. J. Molecular dynamics algorithm for multiple time scales: Systems with long-range forces. *J. Chem. Phys.* **1991**, *94* (10), 6811−6815.

[30] Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990−2001.

[31] Watanabe, M.; Karplus, M. Dynamics of molecules with internal degrees of freedom by multiple time-step methods. *J. Chem. Phys.* **1993**, *99*, 8063.

[32] Némethy, G.; Pottle, M. S.; Scheraga, H. A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J. Phys. Chem.* **1983**, *87*, 1883.

[33] Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* **1975**, *79* (22), 2361−2381.

[34] Durup, J. Protein molecular dynamics constrained to slow modes. Theoretical approach based on a hierarchy of local modes with a set of holonomic constraints: The method and its tests on citrate synthase. *J. Phys. Chem.* **1991**, *95*, 1817−1829.

[35] Mazur, A. K.; Dorofeev, V. E.; Abagyan, R. A. Derivation and testing of explicit equations of motion for polymers described by internal coordinates. *J. Comput. Phys.* **1991**, *92*, 261−272.

[36] Head-Gordon, T.; Brooks, C. L., III. Virtual rigid body dynamics. *Biopolymers* **1991**, *31*, 77−100.

[37] Surles, M. C. An algorithm with linear complexity for interactive, physically-based modeling of large proteins. *Comput. Graphics* **1992**, *26* (2), 221−230.

[38] Morley, S. D.; Jackson, D. E.; Saunders, M. R.; Vinter, J. G. DMC: A multifunctional hybrid dynamics/Monte Carlo simulation algorithm for the evaluation of conformational space. *J. Comput. Chem.* **1992**, *13* (6), 693−703.

[39] Elofsson, A.; Le Grand, S. M.; Eisenberg, D. Local moves: An efficient algorithm for simulation of protein folding. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 73−82.

[40] Dudek, M. J.; Scheraga, H. A. Protein structure prediction using a combination of sequence homology and global energy minimization I. Global energy minimization of surface loops. *J. Comput. Chem.* **1990**, *11* (1), 121−151.

[41] Crippen, G. M. Exploring the conformational space of cycloalkanes by linearized embedding. *J. Comput. Chem.* **1992**, *13* (3), 351−361.

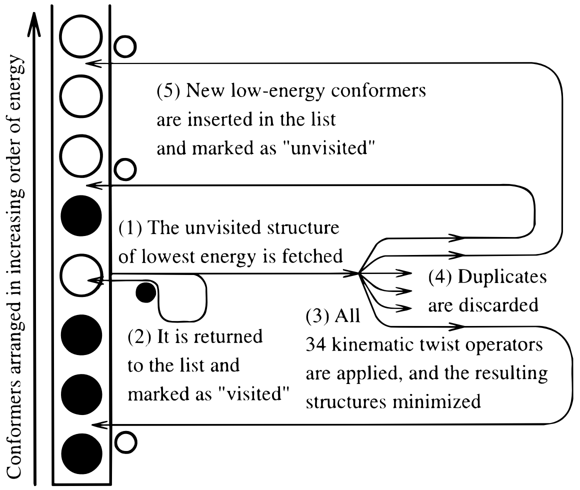
Conformers arranged in increasing order of energy
(5) New low-energy conformers are inserted in the list and marked as "unvisited"
(1) The unvisited structure of lowest energy is fetched
(4) Duplicates are discarded
(3) All 34 kinematic twist operators are applied, and the resulting structures minimized
(2) It is returned to the list and marked as "visited"

Figure 1. Schematic illustration of one iteration of the pseudosystematic search. The column of larger circles represents the growing set of known conformers, in ascending order of energy. Unfilled circles are unvisited; filled circles are visited.

tree searches, and as we have pointed out, the pseudosystematic search is much closer in approach to Monte Carlo with minimization. Specifically, the underlying structure being searched is a directed graph, not a tree. We call this graph the conformer graph. Although the concept of the conformer graph is not critical to understanding how the algorithm works, it is helpful in understanding how thorough the search can be expected to be; the search is not guaranteed to be exhaustive. We develop the idea informally in a manner that is independent of the choice of deformation operators, which could, in principle, be based on an existing technique such as simple torsional change[9] or the "Cartesian kick".[8] A somewhat more formal treatment is given in the Supporting Information.

4.1. Set of Local Minima. Let U be the empirical potential-energy function, in this case the MM2 energy function. Let S be the set of all unique local minima of U. We intend the common-sense meaning of the term "unique": two minima are considered to be identical if the corresponding structures are related to each other by rigid translation or rotation, reflection through a spatial plane, or reversal or cyclic permutation of the carbon-numbering system.

We name the local minima s_1, s_2, ..., in nondecreasing order of their energies. Thus, s_1 is the global optimum, s_2 is the next best, and so on. For a given energy threshold ΔU we define the set of low-energy minima to be the subset $S_{\Delta U}$ of S whose elements have energy not exceeding $U(s_1) + \Delta U$. The conformational search task in the present case is to enumerate the elements of $S_{\Delta U}$, where ΔU is 3 kcal/mol.

4.2. Set of Deformation Operators. Let W be the set of deformation operators to be used by the search algorithm, and let $\{w_1, w_2, ..., w_{|W|}\}$ be its elements. (The set might be very large if the deformations are generated randomly, because in principle each member of W corresponds to one possible assignment of the random parameters that control the deformation. For example, the operator w_{89} might be one that adds 12.1° to the fifth torsion angle. However, the set of deformations used here is small: $|W| = 34$.)

If a given deformation operator w_k is applied to a given local minimum s_i and the structure is subsequently minimized, the result will be a local minimum (either the original one or a different one). We will use the notation $s_i \rightarrow s_j$ to signify that s_j can be generated by applying some operator $w_k \in W$ to s_i and subsequently minimizing its energy. (From this point forward, all definitions are dependent on the particular choice of energy-minimization algorithm, since two such algorithms might in practice arrive at different minima when applied to the same unminimized structure.)

4.3. Full Conformer Graph. The properties of the solution space "seen" by the search algorithm are completely described by the full conformer graph G. This graph is not a fundamental property of the molecule; it depends on the choice of energy function U, the energy-minimization algorithm, and the deformation operators W. It is a directed graph whose nodes are the local minima (the elements of S), and in which an arc from s_i to s_j exists if and only if $s_i \rightarrow s_j$.

Note that the conformer graph contains no information about *kinetics* because the process of applying a deformation operator and subsequently minimizing is independent of energy-barrier heights. The conformer graph determines what moves are permissible *computationally*. Stated simply, if the conformer graph is compared to a road map, the algorithm must respect one-way streets.

4.4. Low-Energy Conformer Graph. The run described here was terminated soon after it had visited all low-energy conformers (those in $S_{\Delta U}$) that it had encountered, *i.e.*, before visiting any conformers outside the ΔU energy bound. To analyze the effects of terminating the search in this manner, we define the low-energy conformer graph, which is the graph that is left when all high-energy nodes and the edges attached to them are deleted.

Because of this edge deletion, some node in $S_{\Delta U}$ may not be found by the search if every path to it in the full conformer graph from the starting node contains a high-energy conformer. The kinematic twist operators described below successfully generated low-energy conformers of cycloheptadecane when applied to low-energy conformers (see the Results). Applying the pseudosysematic search technique to long-chain molecules is likely to be complicated by the possibility of backbone self-intersection; in the Discussion we explore what is involved in designing deformation operators for such molecules.

5. Kinematic Twist Operator

We now describe the *kinematic twist* operator used in the work presented here. There are 34 possible kinematic twist operators (two for each methylene group). To help understand the design of the operator, let us suppose temporarily that the two carbon−carbon bonds emerging from a methylene group are collinear. To exploit the 3-fold periodicity of the torsional potential, each kinematic twist operator treats a methylene group i as if it were a three-state thumbwheel with the states equally spaced at 120° intervals; it attempts to move the thumbwheel to one of the two other states[42] without moving the two neighboring methylenes $i \pm 1$ by rotating the two torsions ω_{i-1} and ω_i in opposite senses by 120°. Since the carbon−carbon bonds are in fact not collinear, the kinematic twist operator must simultaneously adjust the rest of the ring in order to maintain closure. The kinematic technique described below adjusts the torsions in the rest of the ring minimally in a least-squares sense.

5.1. Kinematic Chain. Let ω_i be the torsional angle about the C−C bond that connects methylene group i to methylene group $i - 1$. Let $\vec{C}, \vec{H}^+, \vec{H}^- \in \mathrm{R}^3$ be the positions of the carbon and two hydrogens of a methylene group in some arbitrary reference position.

We relate the actual positions of the atoms in the ith methylene group to these reference positions a rigid-body transformation operator $\hat{U}$:

[42] This is similar to the corner-flapping motion sought by Morley *et al.* in their Dynamic Monte Carlo technique.[38]

$$\vec{C}_i = \hat{U}_i\vec{C}$$

$$\vec{H}_i^+ = \hat{U}_i\vec{H}^+$$

$$\vec{H}_i^- = \hat{U}_i\vec{H}^-$$

Without loss of generality, we fix methylene group 1 to the reference position by defining $\hat{U}_1$ to be the identity transformation. Then, as usual,[10] the remaining $\hat{U}$ quantities are computed recursively:

$$\hat{U}_i = \hat{U}_{i-1}\hat{\Omega}_{i-1}$$

Each Ω is a function of the corresponding torsional angle only:

$$\hat{\Omega}_i = \hat{\Omega}_i(\omega_i)$$

5.2. Closure Condition. The kinematic chain is closed if and only if an imaginary 18th methylene coincides with the first group, so that:

$$\hat{U}_{18} = I$$

It is convenient to express this condition in terms of a ring-closure vector

$$\vec{R} = \Lambda(\hat{U}_{18})$$

where Λ is an operator that converts its argument, a rigid-body transformation, into a vector of three rotational and three translational parameters[43] $\Lambda(\hat{I}) \equiv \vec{0}$.

5.3. Jacobian Matrix and Its Pseudoinverse. Let **J** be the Jacobian matrix that, to first order, relates changes in the ring-closure vector $\vec{R}$ and the torsional angles $\vec{\omega}$:

$$J_{ij} = \frac{\partial R_i}{\partial \omega_j}$$

Let $\mathbf{J}^+$ be its Moore−Penrose pseudoinverse,[44] which may be computed by singular-value decomposition (SVD).[45]

5.4. Closing an Unclosed Ring. Given a ring that is not completely closed, we can bring about closure by repeatedly adding to the torsional angles a quantity $\Delta\vec{\omega}$, where

$$\Delta\vec{\omega} = \mathbf{J}^+\cdot(-\mu\vec{R})$$

The scalar quantity μ is either 1 or a smaller positive number chosen so that no component of $\Delta\vec{\omega}$ exceeds 0.1 rad. (When many ring-closure conditions in a polycyclic molecule must be satisfied simultaneously (see the Discussion), the dimensionality of $\vec{R}$ will be a multiple of 6.) This operation is repeated until $\vec{R} = \vec{0}$ to six or seven digits of precision.[46]

5.5. Deforming a Closed Ring. Given a ring that is already closed, and a closure-violating set of proposed changes $\Delta\vec{\omega}$ to the torsional angles, we can "filter" the changes so that, to first order, they leave the ring closed. This is done by projecting $\Delta\vec{\omega}$ onto the null space of **J**, i.e., replacing it with the quantity

$$(I - \mathbf{J}^+\mathbf{J})\cdot\Delta\vec{\omega}$$

5.6. Kinematic Twist. The kinematic twist operator begins by setting $\Delta\vec{\omega}$ to one of the 17 possible cyclic permutations of (10°, −10°, 0°, 0°, ...) or (−10°, 10°, 0°, 0°, ...), a 10° step of the idealized "thumbwheel" rotation, in which two adjacent torsions are rotated in opposite senses. The vector $\vec{\omega}$ of torsion angles is replaced by

$$\vec{\omega} + (I - \mathbf{J}^+\mathbf{J})\cdot\Delta\vec{\omega}$$

which executes the closest possible approximation to the idealized rotation $\Delta\vec{\omega}$ that keeps the ring closed to first order. Thus, 15 of the torsions will have changed by the minimum amount necessary to maintain ring closure, which is desirable because the original structure is known to contain energetically reasonable torsion angles. The ring is then reclosed kinematically, as described above, to remove any second-order errors.

As a result of this two-step procedure, the two torsions associated with the 10° components of $\Delta\vec{\omega}$ will have moved by about 10°, and other torsions will have moved slightly. The two-step procedure is repeated as many times as necessary (approximately 12 times) to bring about at least a 120° net change in one of the torsion angles. This termination condition ensures that at least one torsion in the ring is moved to a new well in its own potential curve prior to reminimization. Interestingly, the two lowest-energy conformers of cycloheptadecane, as found below, are related by torsional changes that are close to an "idealized" kinematic twist; see Figure 2.

6. Results

6.1. Search Efficiency. In assessing the efficiency of various promising conformational search techniques, Saunders *et al.* used MicroVax II cpu time as the basic measure of computational cost. Thus, their comparisons do not factor out differences in implementation, such as the use of coarse energy functions in the initial stages of refinement, and the rejection of highly strained structures after partial minimization. In the present work we consider the inherent efficiency of the search algorithms in a manner that is independent of such essentially orthogonal factors.[47] We therefore use the number of structures minimized as the basic measure of cpu cost.

[43] The operator Λ converts a rigid-body transformation to a representation containing six coordinates known as the *canonical coordinates* of the transformation.[67] Formally, the canonical coordinates are defined in terms of the logarithm of the rigid-body transformation—an infinite power series of matrices. In practice, no matrix logarithm is required. Consider some rigid-body translation $\hat{T}$ that is equivalent to some rigid-body rotation A followed by some translation $\vec{b}$, *i.e.*, $\hat{T}\vec{r} = A\vec{r} + \vec{b}$ for every vector $\vec{r}$. Further, let A represent rotation by θ rad about an axis $\vec{n}$. Then the six components of $\Lambda(\hat{T})$ are θn_x, θn_y, θn_z, λb_x, λb_y, and λb_z. The parameter λ, which controls the relative weighting of the translational and rotational ring-closure constraints, was chosen for cycloheptadecane to be 0.01 rad Å^{-1}, or about 0.57 deg Å^{-1}. This causes the maximum values of the rotational and translational components of $\Lambda(\hat{T})$ to have similar magnitude.

[44] Strang, G. *Linear Algebra and Its Applications*, 2nd ed.; Academic Press: New York, 1976.

[45] Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C. The Art of Scientific Computing*; Cambridge University Press: Cambridge, U.K., 1988.

[46] The procedure has never failed to converge during the deterministic portion of the pseudosystematic search. However, it can become unstable if the random starting conformation happens to be almost fully extended, so that the rotational portion of $\hat{U}_{18}$ is nearly of magnitude 180°. In such a case, the axis of rotation becomes indeterminate. This singular condition, a well-known complication in the control of robot arms,[13] can arise only at the beginning of the pseudosystematic search, when the first random structure is generated. It is easily overcome by bending the extended structure in any direction, or by choosing a new random starting structure.

[47] Our implicit assumption is that the average time required to minimize each structure would be no higher, given the same hardware, software, and energy minimization parameters, than the time required when a different search technique is employed. This assumption is reasonable because, unlike the trial structures produced by the Cartesian and torsional Monte Carlo searches, the trial structures produced by the pseudosystematic search contain no unfavorable bond lengths and angles and, because of the use of operations that respect the kinematics of the molecule, are unlikely to contain unfavorable torsional values.

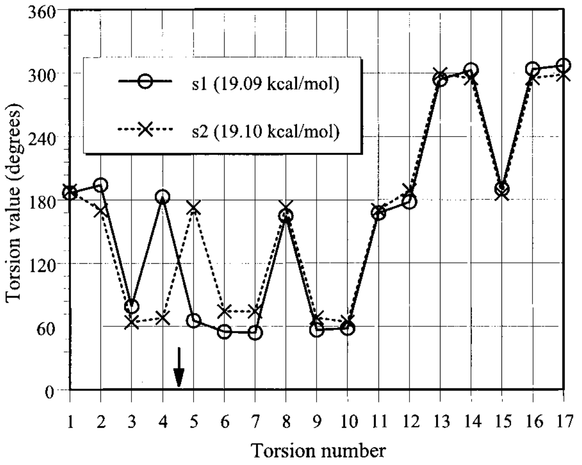

Figure 2. Generation of s_1 (19.09 kcal/mol) by means of a kinematic twist applied to s_2 (19.10 kcal/mol). All torsional values are similar, except ω_4 and ω_5, which undergo +gauche → trans (+60° → 180°) and trans → +gauche transitions, respectively. An arrow indicates the position of the methylene group rotated during the kinematic twist. The approximate C_2 symmetry observed in s_2 by Saunders *et al.* is about an axis drawn from the middle of ω_{15} to the methylene group between torsions ω_6 and ω_7.

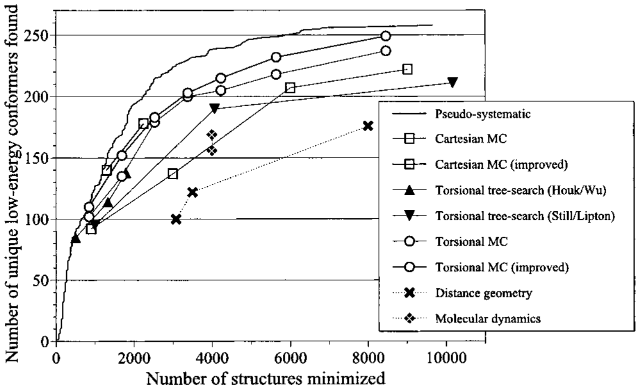

Figure 3. Comparison of search techniques in terms of their efficiency and thoroughness. The vertical axis represents the number of conformers found within 3 kcal/mol of the best structure that has been found by any search to date, and is therefore presumed to be the global minimum. The horizontal axis represents the number of structures minimized. An arrow indicates when the pseudosystematic search had visited all of the 257 low-energy conformers that it had encountered until that point; before the run was terminated, one more conformer (s_{145}) was found. For all algorithms except the pseudosystematic search, the horizontal location of each point was inferred by dividing run times in Table 1 of Saunders *et al.* by average minimization times quoted in the text. All of the inferred values have been confirmed to be within 10% of their true values (C. Still, personal communication). Not shown is a three-day torsional Monte Carlo run on a Convex C210, which found 260 of the 262 low-energy conformers known at that time. The three-day run is estimated to have required 30 000 structure minimizations.

Efficiency figures for the Saunders *et al.* tests are plotted in Figure 3. Those runs identified a total of 262 low-energy conformers. A single three-day run of the torsional Monte Carlo algorithm on a Convex 210, which required approximately 30 000 energy minimizations and found 260 of the 262 conformers, is omitted from the plot. Of the remaining Saunders *et al.* runs, the most thorough one (an "improved" torsional MC run) identified 249 of the low-energy conformers after approximately 10 000 energy minimizations. (Their set of 262 conformers was found by combining the results from different searches.)

In a similar[48] number of minimizations (see Figure 3), the pseudosystematic search found 258 conformers: 257 of the 262 previously known conformers, plus one new one. The new conformer was confirmed by vibrational analysis to be a true minimum, and was determined to be different from the other 262 conformers.[49] The pseudosystematic search found more unique low-energy conformers than any of the other algorithms tested, except for the 30 000-step run of the torsional Monte Carlo algorithm. Although efficiency is not the primary objective, we note that the pseudosystematic search was also the most efficient algorithm; its progress is also plotted in Figure 3.

Combining the results from all of the algorithms tested by Saunders *et al.* with those reported here, 263 conformers are now known.[50] In the subsequent analysis, the set of low-energy minima $S_{\Delta U}$ is taken to be this set of 263 conformers.

6.2. Density of States. The density of states agrees well with a Gaussian fit (see Supporting Information). If the Gaussian fit is assumed valid, the mean energy is 22.6 kcal/mol and the total number of states is 900. Although these values are obviously approximate, they are of interest, particularly because the numbers are such that they could be checked in a reasonable amount of time. Further, the apparent Gaussian density of states suggests that, even for this small, relatively constrained system, a random-energy model is applicable. This is of interest in relation to the random-energy models that have been applied to polymers and to protein models.[51]

6.3. Order of Discovery. Conformers are found by the pseudosystematic search nearly in ascending order of energy. A plot of these energies as the search progresses is helpful in visualizing how the algorithm operates (see Figure 4). The lower envelope of the plot follows a characteristic two-stroke pattern, even though the algorithm is not explicitly divided into two phases. During the downward stroke (bottom panel), each visit happens to produce a new instant global minimum. The new instant global minimum is, of course, the "unvisited structure of lowest energy" called for by the pseudosystematic search, and is therefore the next to be visited. This greedy behavior continues to be successful in producing new instant global minima until the seventh visit. For the remainder of the run, no new instant global minima are produced, so the other minima are visited essentially in order of ascending energy; this produces the rising stroke of the plot. The energies of nearly all of the unique conformers discovered at a given iteration of the search lie above a well-behaved "weak envelope"; *i.e.*, there are only a few outliers, which represent about 10% of the unique conformers. Parallel to the weak envelope, approximately 0.8 kcal/mol below it, is a "strong envelope" that incorporates all of the outliers. Each outlier corresponds to a node s_i that is not accessible from the structure from which the search started, except via paths that involve some intermediate node of energy exceeding $U(s_i)$. This "energy hump" is generally less than 0.8

[48] Although the pseudosystematic search can, in principle, be run until all encountered conformers have been visited, we terminated the run soon after visiting a conformer with energy 22.17 kcal/mol, *i.e.*, 3.08 kcal/mol above $U(s_1)$.

[49] Duplicates were eliminated by the same criterion employed by Houk and Wu:[1] two conformers are deemed identical if all of their torsions match within 10° after 1 of the 68 possible symmetry operations.

[50] A change in precision between versions 2.1 and 3.5 of the BATCHMIN program (C. Still, personal communication) caused energies to shift slightly. In particular, $U(s_1)$ changed from 19.06 to 19.09 kcal/mol, and $U(s_{263}) - U(s_1)$ increased to just below 3.002 kcal/mol. If the coordinates obtained by Saunders *et al.* are reminimized, all energies of corresponding conformers match those obtained here to at least 0.0001 kcal/mol precision. The results reported here employ the 3.002 kcal/mol threshold.

[51] Shakhnovich, E. I.; Gutin, A. M. Implications of thermodynamics of protein folding for evolution of primary sequences. *Nature* **1990**, *346*, 773−775.

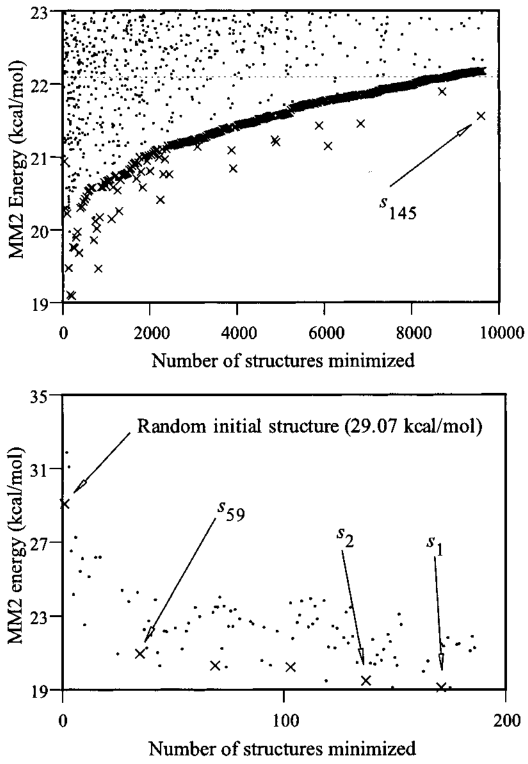

Figure 4. Order in which conformers are first encountered by the pseudosystematic search. Each × represents the visit of a structure, *i.e.*, the application of 34 kinematic twist operators to it. Subsequent minimization of each structure and elimination of duplicates produce up to 34 new conformers, indicated by dots. A broken line on the main plot marks the energy 3 kcal/mol above the purported global minimum; note that a last low-energy minimum (s_{145}, 21.55 kcal/mol) was found while visiting a high-energy conformer slightly above that threshold (22.17 kcal/mol). The second plot differs from the first in the scale and extent of its axes.

Table 2. Nodes Mentioned Frequently in the Text

node	energy (kcal/mol)	remark
s_1	19.0917	purported global minimum
s_{55}	20.9045	found only by pseudosystematic search
s_{108}	21.3214	found only by combined Saunders *et al.* searches
s_{155}	21.5883	
s_{159}	21.6253	
s_{195}	21.8127	
s_{227}	21.9442	
s_{145}	21.5537	found via high-energy node (22.17 kcal/mol)
s_{246}	22.0293	no outgoing edges to low-energy nodes

kcal/mol, which is about the difference in energy between a gauche torsion and a trans torsion in the MM2 potential.

This two-envelope pattern in the upward stroke provides a fairly reliable empirical method for determining when all conformers below a given energy bound are likely to have been found. Specifically, assuming that the observed behavior would remain consistent were the search to be continued indefinitely, it is reasonable to infer from the plot in Figure 4 that the first 9000 iterations of the search are likely to have identified nearly all of the low-energy conformers, and that all of the missed conformers are likely to have energy at least 2.2 kcal/mol above the purported global minimum. This expectation is confirmed (see Table 2). A corollary inference is that the purported global minimum is almost certainly the true global minimum. To be reasonably certain to enumerate all of the low-energy conformers, *i.e.*, those less than 3.002 kcal/mol above the purported global minimum, the run would have to be continued until the weak envelope were to reach approximately 3.8 kcal/mol.

6.4. Diminishing Returns. From the shape of the upward stroke one can obtain a semiquantitative estimate for the diminishing returns of continuing the search. The number of unique conformers below a given energy bound identified per iteration is expected to decrease gradually as the run proceeds, drop almost to zero when the weak envelope intersects the given bound, and drop exactly to zero when the strong envelope intersects the bound.

A conventional Monte Carlo minimization search[15,16] also produces diminishing returns, for two reasons. First, the probability that a newly generated conformer is a duplicate of a previously generated conformer increases as the search proceeds, simply because the number of previously generated conformers increases. Second, the probability that an existing starting structure is subjected to the same or similar perturbations twice increases as the search proceeds. The latter source of diminishing returns is absent in a pseudosystematic search: a given conformer is subjected to each deformation operator at most once. *In other words, the pseudosystematic search does not avoid encountering the same node twice, but it does avoid traversing the same edge twice.* This is the principal advantage of the pseudosystematic search over conventional Monte Carlo methods. The effect of this advantage increases as the number of structures tested increases, because of the rising probability that a conformer in a conventional Monte Carlo run is subjected to the same (or similar) perturbations twice.

6.5. Properties of the Conformer Graph. Because the pseudosystematic search is a straightforward traversal of the conformer graph from a random seed, analysis of the conformer graph for cycloheptadecane yields answers to a number of practical questions about the robustness of the algorithm and how it would behave if modified. This analysis extends and validates many of the observations made by Gotō and Ōsawa.[4] (Except where explicit reference is made to a figure in this paper, corroborating data are provided in the Supporting Information.)

First, two initially attractive shortcuts to avoid duplicated effort cannot be used without possibly compromising the thoroughness of the search: (1) From the success of this run in producing a new instant global minimum after each of the first six visits, and the fact that the sixth such minimum is s_1, the purported global minimum, one might wonder whether the global energy minimum of cycloheptadecane can always be found by traversing edges only "downhill" in energy. Unfortunately, nearly all paths to the global energy minimum contain increases in energy of up to 0.8 kcal/mol (see Figure 5). This 0.8 kcal/mol figure matches the distance between the weak and strong envelopes in Figure 4. (2) The existence of an edge (s_i, s_j) does not necessarily imply the existence of an edge (s_j, s_i). Thus, the algorithm cannot be made more efficient by avoiding backward traversal of previously traversed arcs.

Second, we find that the kinematic twist operator is efficient at escaping from local minima and avoids producing isolated groups of local minima with similar energy: (1) For many of the low-energy conformers, at least 1 of the 34 possible kinematic twist operators produces the original conformer after minimization. However, the number of times this happens for each conformer in most cases does not exceed two. (2) The energies of neighboring low-energy nodes on the graph are uncorrelated. (However, conformers very close in energy to the global optimum have more incoming edges than the others. The best two conformers, s_1 and s_2, have 25 and 22 incoming edges, respectively.)

Third, the search procedure always finds the purported global minimum quickly: (1) The purported global optimum, s_1, is first encountered within 550 minimizations after encountering

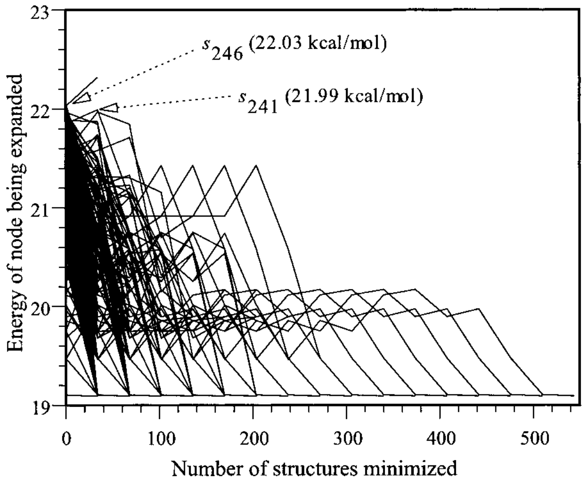

Figure 5. Energies along the path from each seed s_i to s_1. Each trace corresponds to a separate run of the pseudosystematic algorithm in which a different low-energy conformer is used as the seed, and shows the energies of the nodes visited during the search. Traces are terminated at s_1, since subsequent behavior is essentially identical in each case.

one of the other low-energy conformers when one starts from nearly any one of the possible seeds. (2) The single exception to this rule is when the seed is s_{246} (22.03 kcal/mol). (Selected nodes are listed in Table 2.) The best neighbor of s_{246} is never visited because its energy (22.32 kcal/mol) is more than 3 kcal/mol above $U(s_1)$ (see Figure 5). However, recall that the energy bound on visited nodes is measured relative to the instant global minimum. A search beginning with s_{246} would have found one of the other low-energy conformers before terminating, and therefore would have gone on to find s_1.

Fourth, the vivid metaphor of the search as "the process of pouring water into an empty reservoir or dam",[4] which evokes a broad, rugged basin, is found to be slightly misleading (at least in the case of cycloheptadecane). The energies of the low-energy conformers are very weakly correlated with their distance in the graph from the purported global minimum.

Finally, of prime interest are the known low-energy conformers[1] that were not found by the run presented here.[52] For clarity we denote this set of five conformers by S^{missed}:

$$S^{\text{missed}} \equiv \{s_{108}, s_{155}, s_{159}, s_{195}, s_{227}\}$$

None of the five nodes has any incoming edge from any node in $S - S^{\text{missed}}$; otherwise, it would have been found by the pseudosystematic search. However, each has an outgoing edge to the remainder of $G_{\Delta U}$, indicating that none is completely isolated. This situation closely resembles that of s_{145}, the node found just after the 3 kcal/mol threshold was reached by the search. For example, when s_{145} or any of the nodes in S^{missed} is the seed, the number of low-energy conformers found by the search exceeds 257 (see Supporting Information). (When either s_{108} or s_{227} is the seed, the total number of conformers found is 260. This is because the two nodes are connected bidirectionally, and via an outgoing edge to s_{155}.) The nodes in S^{missed}, like s_{145}, are likely to be found by running the pseudosystematic algorithm until the strong envelope described above reaches the 3 kcal/mol energy bound.

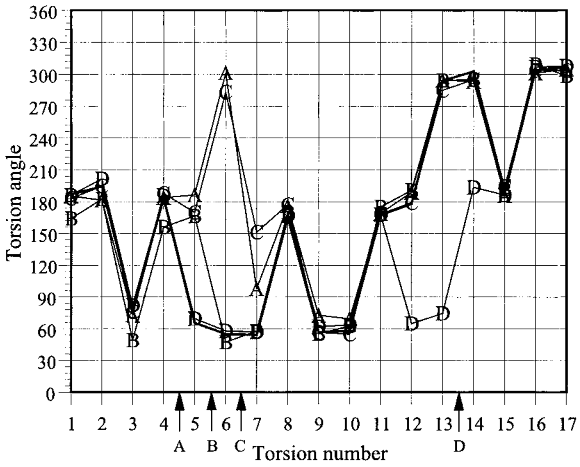

Figure 6. Selected nearest neighbors of s_1 in $G_{\Delta U}$ with energy within 1.5 kcal/mol. In this and the next figure, neighbors connected to s_1 via both incoming and outgoing edges are considered. Torsions are renumbered to reflect any symmetry operations performed during the search to identify the result of a deformation as a duplicate conformer. An arrow indicates the position of the methylene group rotated during each kinematic twist. The following notation summarizes the effects of the kinematic twist and subsequent minimization. The minimized structures are shown. Asterisks mark the torsions involved in each kinematic twist. (A) $S_1 \rightarrow S_7$ (ω_5*, ω_6*, ω_7): (+g, +g, +g) (+a, −g, +g) (+a, −g, +95°) (a = anti) (B) $s_{22} \rightarrow s_1$ (ω_4*, ω_5*): (a, a) (−g, +g) (a, +g). (C) $s_1 \rightarrow s_{24}$ (ω_5, ω_6*, ω_7*): (+g, +g, +g) (+g, −g, a) (a, −g, a). (D) $s_{26} \rightarrow s_1$ (ω_{12}, ω_{13}*, ω_{14}*): (+g, +g, a) (+g, −g, −g) (a, −g, −g).

6.6. Steric and Closure Effects. The effect of the kinematic twist operator and subsequent reminimization may be studied by comparing pairs of neighbors in the symmetrized conformer graph. In some cases, only the two torsions directly controlled by the kinematic twist operator undergo appreciable change, as in the $s_1 \rightarrow s_2$ transition depicted in Figure 2. In other cases, additional effects are produced by kinematic maintenance of ring closure during the deformation, and by subsequent relief of steric strain during energy minimization.

Figure 6 shows transitions between s_1 and selected very low-energy conformers. (Transitions both to and from s_1 are considered.) Most departures from the simple two-torsion behavior of the kinematic operator are accounted for by the steric overlap that would be present in a true +g/−g (g = gauche) pentane fragment.[1,53,54] In the transition $s_{26} \rightarrow s_1$ (D), an unfavorable +g/−g pentane array (ω_{12}, ω_{13}) is avoided by the movement of ω_{12} to the trans position. A similar event occurs at ω_5 in the transition $s_1 \rightarrow s_{24}$ (C). Occasionally one of the torsions under kinematic control returns to its original value, as in $s_{22} \rightarrow s_1$ (B), where a +g/−g array (ω_4, ω_5) is avoided by the return of ω_4 to 180°. In some cases a +g/−g configuration is nearly adopted, but with +95°/−65° instead of +65°/−65° geometry, as observed in *ab initio* (MP3/6-31G*) calculations with pentane.[53] An example of this is in $s_1 \rightarrow s_7$ (A), where $\omega_7 \doteq 98°$. The presence of eight trans torsions (compared to seven in s_1 and s_2) gives s_7 a very low energy (19.76 kcal/mol) despite the relatively unfavorable arrangement at ω_7.

Transitions between s_1 and some of its neighbors close to the 3 kcal/mol threshold are shown in Figure 7. Each of the three neighbors contains the characteristic +95°/−g fragment,

[52] Because the pseudosystematic search found only the subset of low-energy conformers $S^{\text{found}} \equiv S_{\Delta U}$- S^{missed}, its output can be used to reconstruct only $G \cap (S^{\text{found}} \times S^{\text{found}})$. As required by the definition that $G_{\Delta U} \cap (S_{\Delta U} \times S_{\Delta U})$, *i.e.*, that the low-energy conformer graph include all 263 known low-energy conformers, $G_{\Delta U}$ was constructed by augmenting the graph $G \cap (S^{\text{found}} \times S^{\text{found}})$ with the outgoing edges from each node in S^{missed}.

[53] Wiberg, K. B.; Murcko, M. A. Rotational barriers. 2. Energies of alkane rotamers. An examination of gauche interactions. *J. Am. Chem. Soc.* **1988**, *110*, 8029−8038.

[54] Dunbrack Jr., R. L.; Karplus, M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.* **1994**, *1*, 334−340.

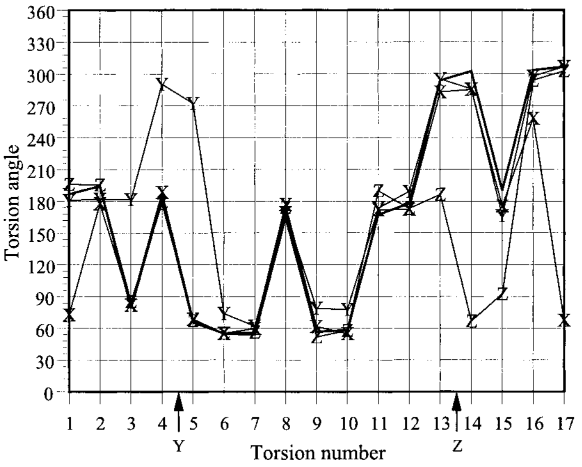

Figure 7. Selected nearest neighbors of s_1 in $G_{\Delta U}$ with energy close to the 3 kcal/mol threshold. An arrow indicates the position of the methylene group rotated during each kinematic twist, except in the case of the $s_{192} \rightarrow s_1$ transition (X), in which torsions ω_1 and ω_{17} were rotated. (X) $s_{194} \rightarrow s_1$ (ω_{16}, ω_{17}*, ω_1*): (−95°, +g, +g) (−95°, −g, a) (−g, −g, a). (Y) $s_{194} \rightarrow s_1$ (ω_1, ω_2, ω_3, ω_4*, ω_5*, ω_6): (a, a, a, −g, −95°, +75°) (a, a, a, a, +g, +75°) (a, a, +80°, a, +g, +g). (Z) $s_{242} \rightarrow s_1$ (ω_{13}*, ω_{14}*, ω_{15}, ω_{16}): (a, +g, +90°, −g) (−g, −g, +90°, −g) (−g, −g, a, −g).

and in each case the 95° torsion disappears as a result of the kinematic twist that relates the neighbor to s_1.

In $s_{192} \rightarrow s_1$ (X), the −95° value at ω_{16} relaxes to $-g$ because the adjacent +g becomes −g. In $s_{194} \rightarrow s_1$ (Y), the kinematic twist attempts to create a sequence of four contiguous 180° torsions (ω_1 through ω_4), whose presence would hinder ring closure; this is resolved by the change of ω_3 from trans to the locally less favorable +gauche position.[55] In $s_{242} \rightarrow s_1$ (Z), a highly unfavorable −g/+95°/−g triplet centered on ω_{15} is avoided by the rotation of ω_{15} to 180°. Plots showing the correlation between adjacent torsions in the ring and the effects of the "1,5 pentane" interaction associated with the +g/−g configuration are provided in the Supporting Information.

7. Discussion

7.1. Comparison with Other Algorithms. Saunders *et al.* carefully compared a range of conformational search techniques, using exhaustive enumeration of the low-energy conformers of cycloheptadecane as a benchmark. A possible conclusion of their study, which they refrained from stating categorically, is that stochastic methods provide a number of advantages over systematic methods. Chief among these is that a stochastic conformational search may be run to any desired degree of convergence, without the need for *a priori* decisions. In contrast, a torsional tree search requires an inordinate amount of time unless pruned, and pruning requires predetermined assumptions that can cause a significant number of low-energy conformers to be missed. An additional disadvantage of torsional tree searches is that they are no faster for symmetric molecules than for asymmetric ones.[1]

Systematic methods are expected to provide certain advantages if they can be made to run with reasonable efficiency. In particular, the behavior of deterministic algorithms can be easier to study and predict, and when they fail, the reasons can sometimes be stated more precisely than would be possible with a stochastic algorithm. Moreover, systematic algorithms can often avoid duplicated effort when stochastic methods cannot. An objective of this work was to discover whether a deterministic conformational search algorithm with the described properties could be designed.

Finding new conformers is inherently more difficult at the end of an enumerative conformational search than at the beginning. Two reasons for the diminishing returns in such a search exist. First, as the number of known conformers increases, the odds of rediscovering a conformer by a pathway in the conformational space not previously traversed by the search increases. Second, the probability of applying the *same* deformation operator to the *same* conformation twice increases unless some bookkeeping is done to avoid such redundant operations.

The first source of redundant effort, which might be possible to reduce in certain special cases, is in general difficult to eliminate. However, the second source is completely avoided by the algorithm we have described; it accomplishes the necessary bookkeeping by performing a systematic traversal of the conformer graph. Two of the methods within the Saunders *et al.* study recognized this effect but were able to avoid it only partially: their "usage criterion" [1] ensures that each known low-energy conformer is used as a starting structure an equal number of times regardless of whether it was found early or late in the search, but cannot control which perturbations are applied.

Hence, the approach presented here is a synthesis of two other approaches usually thought to be dichotomous. Like a stochastic search, it exploits information about known low-energy conformers by generating new trial structures via perturbation, but like a conventional systematic search, it executes its search in a controlled fashion that prevents some types of duplicated effort.

7.2. Scaling Properties of the Pseudosystematic Search. The pseudosystematic algorithm may be used with a variety of different deformation operators. The particular deformation operator employed here, the kinematic twist, exploits the ring-closure condition and lack of branching in a rather direct manner. Furthermore, cycloheptadecane is special in a number of other ways. It is small enough that all of its low-energy conformers are essentially circular, so that the backbone does not cross itself in moving from one torsional well to another. It is highly symmetric: its chemical structure is invariant with respect to reversal or cyclic permutation of the carbon-numbering system, and any given structure has an enantiomer of identical energy that may be constructed by reflection through any spatial plane.

Thus, it is reasonable to ask the following question: would the pseudosystematic algorithm continue to function well when used for a more general class of molecules? In this subsection, we address the scaling properties of the pseudosystematic algorithm. In the next, we lay out issues associated with the design of new deformation operators for larger molecules.

The running time of the algorithm is expected to scale in a reasonable manner for larger molecules. Unlike a torsional tree search, its time complexity is not inherently exponential. Specifically, the total time required to complete the search is slightly more[56] than $T_{\min}\cdot|S_{\Delta U}|\cdot|W|$, where $T_{\min}$ is the average time required to minimize a structure, $|S_{\Delta U}|$ is the number of low-energy conformers, and $|W|$ is the number of deformation operators. $|W|$ is 34 in the case of cycloheptadecane, or more generally,

[55] Saunders *et al.*[1] report that sequences of four contiguous anti torsions appear twice among their conformers within 2 kcal/mol of the purported global minimum and 18 times among their conformers within the 3 kcal/mol bound.

[56] In practice, $|S_{\Delta U}|$ must be replaced by a slightly larger integer to account for the time spent finding the first low-energy conformer. For the run described here, the appropriate figure is $|S_{\Delta U}| + 6$, since six conformers were visited before s_1 was found.

$$|W| = \sum_k (n_k - 1)$$

for the kinematic twist operator, where n_k is the number of preferred values associated with the kth degree of freedom in the molecule. Bearing in mind that $T_{\min}\cdot|S_{\Delta U}|$ is the amount of time required to accomplish the task given an infallible generator of coarse structures (one that generates a unique low-energy conformer every time it is invoked), this time complexity is quite acceptable.[57] From the above, the running time of the algorithm is approximately proportional to the number of low-energy conformers (*i.e.*, those with energy below the ΔU threshold), not the size of the search space. This holds true only to a point, of course; for example, setting $\Delta U = 0$ does not cause the globally optimal structure to be generated instantly! From inspection of the traces in Figure 5, it is likely in the present case that the algorithm would fail to find the globally optimal structure of cycloheptadecane if ΔU were significantly smaller than 1 kcal/mol.

A recent theoretical result stated that an algorithm for global minimization of a molecular potential-energy function cannot be simultaneously (1) always guaranteed to locate the globally optimal solution in less than exponential time and (2) sufficiently general to accommodate a set of hypothetical molecules described in the proof construction,[58] unless "P = NP", a proposition whose most immediate practical implication would be that all NP-complete problems (*e.g.*, traveling salesman) could be solved "efficiently", *i.e.*, in polynomial time.[59] The pseudo-systematic search algorithm does satisfy the second condition, and therefore cannot meet the first unless P = NP. In keeping with that result, the globally optimal solution is not guaranteed to be among the low-energy conformers identified, and the total running time of the algorithm is not guaranteed to be bounded by any function that is polynomial in the size of the problem (because $|S_{\Delta U}|$ is not). However, we find that, for cycloheptadecane, the global solution identified by Saunders *et al.* is found quickly and reliably.[60] The characteristic order in which conformers are found (Figure 4) provides strong evidence that this purported global minimum is the global minimum.

The loss of the high symmetry present in cycloheptadecane would greatly increase the time required by the pseudosystematic algorithm to enumerate the low-energy conformers because the total number of such conformers would increase by a factor of 68. This observation is not specific to the pseudosystematic search; all algorithms for exhaustive enumeration of low-energy conformations must have running time asymptotically linear or superlinear in the number of low-energy conformers.

Thus, the small size and high symmetry of cycloheptadecane make it unrepresentative of macromolecules for which structure prediction would be of biological interest. However, because of its cyclic constraint, it is unlikely to be so simple that its structure can be predicted in polynomial time: end point constraints and restraints are, by one interpretation of the theoretical result cited above,[58] a key source of intractability in molecular-structure prediction.

7.3. Extension of the Kinematic Twist. The particular form of the kinematic twist operator presented here is based on conjectures about geometric relationships between the low-energy conformers of cycloheptadecane. First, it is assumed that no local minimum in the 17-dimensional *torsional* potential contains more than one local minimum in the full potential. This assumption will break down for a long-chain molecule, in which a single well in the torsional potential could correspond to many topologically distinct arrangements of the backbone. Similar bifurcations in individual torsional wells could be caused by mutual obstruction among large substituents. The consequence of the 1,5 pentane interaction for the structure of the +gauche/−gauche potential well is an example (see Supporting Information). Second, the kinematic twist operator is designed so as to produce *minimal change* in the path taken through space by the backbone, as torsions are altered. With a long-chain molecule, large rearrangements of the backbone structure would clearly be essential to a thorough conformational search.

An obvious question is whether the pseudosystematic algorithm could be extended for use with proteins and other long-chain molecules. A new level of difficulty is introduced because the low-energy conformer graph generated by local deformation operators (*e.g.*, the kinematic twist operator) is likely to be disconnected by sterically unfavorable conformations. Paths between disconnected portions of the low-energy conformer graph might be provided by augmenting the set of deformation operators with "rearrangement operators": geometric manipulations that produce concerted changes on the basis of prior knowledge or guesses about secondary and tertiary structure. Rearrangement operators might, for example, impose predetermined secondary structure on small portions of the chain, or "redock" a pair of α helices so as to change their relative alignment by one groove. These manipulations could be carried out using a coarse-level representation of the type employed by Friesner and co-workers.[61,62]

Minor extensions to the algorithm might permit its application to protein loop structures, which are intermediate in complexity between full proteins and the cycloalkane treated here. First, a dihedral degree of freedom that is not involved in a ring-closure condition may be handled by replacing the kinematic twist by a simple twist, *i.e.*, rotation of a single dihedral. Second, a closure condition in which a kinematic chain must meet fixed end points in space, as opposed to form a ring with itself, is easily accommodated by a simple generalization of the ring-closure condition.[63] Third, a dihedral degree of freedom that is involved in more than one closure condition (produced, for example, by a disulfide bridge) may be dealt with by extending the dimensionality of the ring-closure vector to whatever multiple of 6 is required. Finally, long-range interactions present in larger peptide loops might necessitate the use of rearrangement operators.

[57] The algorithm is also easily parallelized, especially for MIMD parallel machines and workstation clusters. If the generation and energy minimization of one trial structure are assigned to each processor, then the total number of processors that can be utilized effectively at any given stage in the search is 34 times the number of known, unvisited solutions. At the beginning of the search, this number can be as low as 34, but soon afterward it rises into the thousands.

[58] Ngo, J. T.; Marks, J. Computational complexity of a problem in molecular-structure prediction. *Protein Eng.* **1992**, *5* (4), 313−321.

[59] Garey, M. R.; Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*; W. H. Freeman and Co.: San Francisco, CA, 1979. P is the set of problems that can be solved exactly by a deterministic machine in polynomial time. NP is the set of problems that can be solved exactly by a non-deterministic machine in polynomial time.

[60] This empirical observation does not contradict the theoretical result; any NP-hard problem can have special cases that are solved relatively easily. What the result prohibits (unless P = NP) is the existence of a conformational search algorithm that can be guaranteed to complete in polynomial time for *all* molecules.

[61] Monge, A.; Friesner, R. A.; Honig, B. An algorithm to generate low-resolution protein tertiary structures from knowledge of secondary structure. *Proc. Natl. Acad. Sci.* **1994**, *91*, 5027−5029.

[62] Gunn, J. R.; Monge, A.; Friesner, R. A.; Marshall, C. H. Hierarchical algorithm for computer modeling of protein tertiary structure: Folding of myoglobin to 6.2å resolution. *J. Phys. Chem.* **1994**, *98*, 702−711.

[63] The condition $\hat{U}_{18} = \hat{I}$ would be replaced by $\hat{U}_{\mathrm{end}} + \hat{U}_{\mathrm{target}}$, where $\hat{U}_{\mathrm{end}}$ represents the position and orientation of the mobile end point of the chain and $\hat{U}_{\mathrm{target}}$ is a constant transformation determined by the required position and orientation.

Algorithm design for molecular-structure prediction begins, implicitly or explicitly, with assumptions about what order exists in the energy landscape. Conventional stochastic methods depend upon statistical correlation between the energies of neighboring states in the conformational space. However, such correlations may be weak or nonexistent, for example, as expected from the random-energy model (REM) of Derrida,[64] which gives a Gaussian energy spectrum that apparently describes the present system. The pseudosystematic algorithm does not attempt to exploit such correlation. Instead, it relies on geometric order: it is assumed that given a low-energy conformer, every other low-energy conformer in the vicinity can be reached by applying a small, discrete set of deformation operators. It remains to be seen whether proteins have enough of this type of geometric order to be amenable to structure prediction by a pseudosystematic algorithm.

Acknowledgment. We thank the reviewers for numerous helpful suggestions, including citations of closely related work.[4,6,7] We thank Oren Becker for ideas in formalizing the conformer graph and John Zao for helping us to strike a compromise between rigor and clarity in presenting the kinematic twist operator. We thank Clark Still for answering questions and providing data. Amedeo Caflisch, Roland Dunbrack, and Leo Caves provided useful assistance and advice. The searches were performed using a Convex C220 and SGI Personal Iris, Indigo, and 4D/230 workstations. This work was supported in part by a National Science Foundation grant. J.T.N. is grateful for a Graduate Fellowship from the Fannie and John Hertz Foundation. An earlier version of this paper appeared as a chapter in the dissertation of J.T.N.[5]

Supporting Information Available: Description of the density of states, properties of the conformer graph, and steric and closure effects (10 pages). See any current masthead page for ordering and Internet access instructions.

JA961132O

[64] Derrida, B. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B* **1981**, *24* (5), 2613−2624.

[65] Rechenberg, I. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*; Frommann-Holzboog Verlag: Stuttgart, 1973.

[66] Bäck, T.; Hoffmeister, F.; Schwefel, H.-P. A survey of evolution strategies. In *Proceedings of the 4th International Conference on Genetic Algorithms*, San Mateo, CA, 1991; Belew, R. K., Booker, L. B., Eds.; Morgan Kaufmann: San Mateo, CA, 1991; pp 2−9.

[67] Choquet-Bruhat, Y.; Dewitt-Morette, C.; Dillard-Bleick, M. *Analysis, Manifolds, and Physics: Part I*; North-Holland: Amsterdam, 1982.