# An Efficient Algorithm for Searching Low-energy Conformers of Cyclic and Acyclic Molecules

**Hitoshi Gotō[a] and Eiji Ōsawa*[,b]**
[a] *Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060, Japan*
[b] *Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi 441, Japan*

A set of strategies for exhaustively finding low-energy conformers of cyclic, acyclic or alicyclic molecules is presented. Starting from any conformation, local perturbation is systematically applied to all the flexible portions of the molecule in question to produce candidates of new conformation. The perturbations consist of flapping and/or flipping for endocyclic bonds and stepwise rotation for acyclic bonds. The conformations they produced are believed to lie close to the initial geometry in the conformational space. The global energy minimum (GEM) structure of the starting domain of conformational space can be quickly reached by always choosing the most stable of the conformers produced in the last perturbation cycle as the next initial structure. Once GEM of the domain is reached, the local perturbations direct the search gradually to higher and higher energy regions while exhaustively finding all the low-energy conformers therein. The variable search-limit strategy allows one to use unstable conformers as the initial structure for perturbation to ensure the exhaustiveness of the search in the low-energy region. By further increasing the search-limit, new domains of conformational space may be found. A program CONFLEX3 containing several additional strategies for improved performance has been tested for *n*-alkanes up to decane and cycloalkanes up to cyclododecane.

Among the fundamental problems in molecular modelling, 'the most difficult to overcome is the global minimum problem. Even if one could exhaustively search conformational space, one still needs to correctly evaluate and rank the relative free energies of all the conformations. This is currently impossible even for systems of 100 atoms.'[1] There is indeed a great demand for predicting the conformer distributions of large and flexible molecules, especially those of biologically interesting systems and of polymeric materials, within reasonable computer time. Whereas the prospect is bright as to the development of ultrafast computers,[2] really powerful software is lacking. Recent intensive research activity on the algorithms of conformational space search[3–5] and their applications[6,7] reflects the situation.

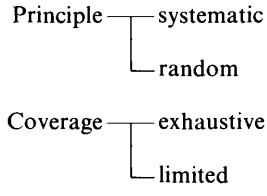
In practice, conformational space search is more difficult in acyclic than in cyclic molecules, the open-chain structures being generally much more flexible than the rings.[8] For this reason, the algorithm research in this area has primarily been directed towards cyclic structures,[4,9] whereas past attempts to cover acyclic conformations are sporadic at best and much less successful than the cyclic conformation search.[5]

This paper is a full account of our systematic study aimed at efficient coverage of low-energy regions of conformational energy space, which works equally well for cyclic and acyclic structures.[10]

## Algorithm Developments

The known algorithms for conformational space search may be categorized into either systematic (deterministic) or random (stochastic) depending on the basic search principle, and each of them can be further divided into exhaustive or limited with regard to the extent of coverage.

The obvious choice, namely the exhaustively systematic method,[5a,11] is not useful, except for very small systems, because the required computer time for this method increases exponentially as the target molecule becomes larger and as the search grid is made finer. Fine grids assure a thorough search but cause

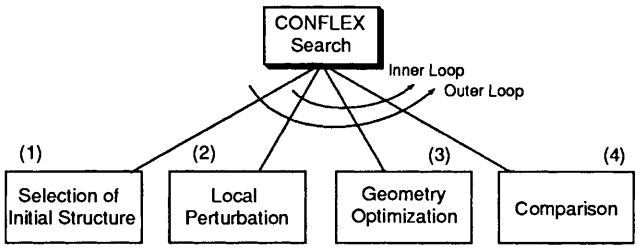



inefficiency owing to the fact that many grid points merge into the same energy minimum (redundancy problem). In order to reduce computing time, attempts have been made to remove trial structures having too short van der Waals contacts.[12] However, too early pruning can be dangerous, because stable conformers sometimes emerge from unstable and seemingly insignificant conformations (*vide infra*).

Random strategies[4c–e,4l,9,13] tend to avoid the redundancy problem, at least in the early stages of the search, by performing wide-range sampling of target points from the whole conformational space. Nevertheless, as Still states,[4e] the later stages of the random search suffer from the rapidly decreasing probability of finding new conformers and also from difficulties in deciding when the search is to be terminated. The most serious problem inherent in the random search is the dense population in the high energy region. Even for small acyclic systems it has recently been noticed that the high-energy space is more crowded than previously thought.[7e]

Covering the conformational space of a flexible molecule is clearly a difficult task. From the above considerations, it must be concluded that a realistic approach can be neither totally systematic nor completely random. We have no other choice but to cover a limited space. An obvious choice is to search only the chemically meaningful, low-energy regions of conformational space, quickly and exhaustively. This strategy may be categorized as efficient.

Our objectives are (1) to avoid going into high-energy regions, (2) to search the low-energy region exhaustively, and (3) to complete the conformation search in reasonable computer time.




| Process | I/O Structural data in | out |
|---|---|---|
| (1) | Input/saved | Initial |
| (2) | Initial | Trial |
| (3) | Trial | Optimized |
| (4) | Optimized | Stored/(none) |

**Fig. 1** Illustration of the four elementary steps in the general conformational space search algorithm

**Table 1** Thirty-seven rotational isomeric states of $C_6$-molecule **1** as expressed by rotation indices at α, β and γ bonds

| | | | | | |
|---|---|---|---|---|---|
| No gauche bond | | | | | |
| AAA | | | | | |
| One gauche bond | | | | | |
| GAA | G′AA | AGA | AG′A | AAG | AAG′ |
| Two gauche bonds | | | | | |
| GGA | G′G′A | AGG | AG′G′ | | |
| **GG′**A | **G′**A | AGG′ | A**G′**G | | |
| GG′A | G′GA | AG**G′** | AG′**G** | | |
| GAG | GAG′ | G′AG | G′AG′ | | |
| Three gauche bonds | | | | | |
| GGG | G′G′G′ | | | | |
| **GGG′** | **G′**GG | GG′**G** | GG′G′ | **G′**GG′ | G′**G′G** |
| GG**G′** | G′GG | **GG′**G | GG′G′ | **G′**GG′ | G′G′**G** |

[a] A = anti, G = gauche with positive sign G′ = gauche with negative sign. Those containing the GG′ sequence are grouped in the boxes. The ±95° dihedral angle can be taken by either G or G′ (bold), giving rise to the vertical isomer pair. Hexane in GG′G (and enantiomeric G′GG′) conformation (underlined) is not an energy-minimum, ref. 7d.

One clear advantage in the conformation search of acyclic molecules, in contrast to cyclic ones, is that it is generally a straightforward task to estimate the global energy minimum (GEM) conformation (*e.g.* the all-anti form for *n*-alkanes), hence we can almost always start the search from the bottom of the energy surface. We consider in this paper only small *n*-alkanes as the target acyclic molecule; more complicated cases of larger, branched or substituted structures could have many domains in the conformational space, each having deep well(s), and they will be the subject of our future study.

*General Scheme.*—An overview of the search algorithm is presented in Fig. 1. The four-process scheme of conformation search is actually common to most of the known conformational space search methods.[4,5] It may be noted that structures are designated differently in these processes. Whole scheme contains the following nested loops: the outer loop for selecting an initial structure (process 1) and the inner loop for generating a new conformer therefrom. The latter consists of the following three steps: perturbation of initial structure to produce a candidate for a new conformation (process 2), geometry optimization of the trial structure (process 3), and comparison of the optimized structure with those already found in order to remove duplicates (process 4).

For every process, new and unique algorithms have been implemented, except for the last one in the table below which is taken from Saunders:[4d]

| process | algorithm |
|---|---|
| 1 | down stream, reservoir filling, variable search-limit |
| 2 | corner flap, edge flip, stepwise rotation, stepping-stone |
| 3 | pre-check |
| 4 | conformational distance |

Features of these algorithms are explained below.

Throughout this paper, we tentatively classify the bond rotation by its dihedral angle $\phi$ (deg) as follows:

| | |
|---|---|
| $\phi = 0$ | syn (S) |
| $0 > \phi > 120$ | gauche (G) |
| $-120 > \phi > 0$ | −gauche (G′) |
| $120 \leq \lvert\phi\rvert \leq 180$ | anti (A) |

*Local Perturbation.*—In our algorithm, local perturbation (process 2, Fig. 1) is the only step in which endocyclic and open-chain bonds are treated separately.

(a) *Stepwise rotation for acyclic bonds.* All internal bonds (*n*) in acyclic parts of the molecule under examination are successively rotated according to the desired resolution, for example ±120° for a $C_{sp^3}$–$C_{sp^3}$ bond of alkane. A maximum of 2*n* new conformers may be discovered in one inner loop. If we start the conformation search of **1** using the all-anti form (AAA) as the input structure, the first inner loop will produce six new conformations (Table 1): GAA, G′AA (these two are obtained by the rotation of α bond), AGA, AG′A (β bond), AAG and AAG′ (γ bond), all of them containing one gauche bond.

$$CH_3\text{–}CH_2\text{–}CH(CH_3)\text{–}CH_2\text{–}CH_2\text{–}CH_3$$

α β γ

**1**

In the second outer loop, the next initial structure is decided according to the down stream algorithm (*vide infra*). Here, let us assume that GAA is chosen. From this, six structures, AAA, G′AA, GGA, GG′A, GAG and GAG′, will be produced by the stepwise rotation. The fourth structure, GG′A, merits special attention. This conformation is actually either **GG′**A or G**G′**A where the bold-faced **G** (or **G′**) indicates the abnormally expanded gauche bond (±95°) due to severe steric congestion.[7e] Because of the asymmetric deformation, **GG′**A and G**G′**A are different energy minima (Table 1). This fact [7e] is referred to later in this paper.

The first two structures, AAA and G′AA, in the second loop are the products of reverse paths: geometry-optimization of these structures only leads to already known conformers, and they are quickly pruned by the 'pre-check' mechanism (*vide infra*). Hence, among the six structures only four are new. In principle a one-gauche initial structure like GAA produces two-gauche conformers. Similarly, the two-gauche initial structures generate three-gauche rotamers and so on. Thus, each inner loop creates conformers having one more gauche bonds than the initial structure.

The purpose of this algorithm is to step up the search space slowly, namely by *ca.* 0.7 kcal* (the gauche increment) [7d] in one step. Note that our strategy differs from the brute-force systematic search algorithm, wherein $3^n$-conformations are generated at once and indiscriminately.[5a] As will be mentioned below,

* 1 cal = 4.184 J.

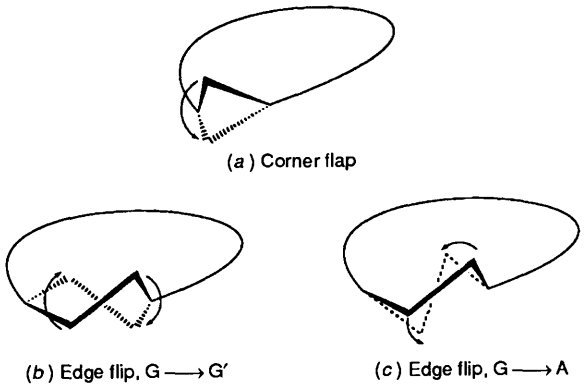



**Fig. 2** Illustration of local perturbation modes for ring bonds

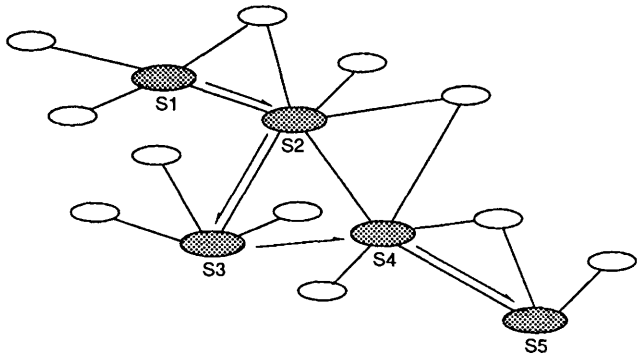
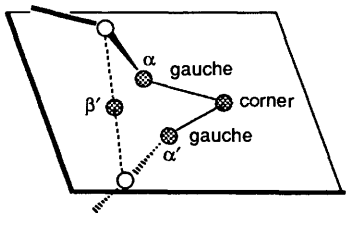



**Fig. 3** Illustration of down stream strategy. Ellipse: conformer; $S_n$: initial structure chosen for the $n$-th loop; ____: conformational interconversion path.

most of the $3^n$ conformations are high energy, hence not interesting to chemists.

(b) *Edge flip and corner flap for endocyclic bonds.* Edge flip is a modification of corner flap [Fig. 2(*a*)].[4f] Soon after our preliminary work[4f] was published, we realized the following problematic cases. (1) Corner flapping did not work when the corner atom was flanked by two gauche bonds of the same sign.




Because the plane defined by three points, α, α′ and β′ (the mid point between the two β atoms), contains the corner atom, the corner cannot be flapped. (2) Corner flap does not work when either of the bonds flanking the corner atom is anti. (3) Sometimes the flapping operation induces unduly strained deformation when the corner is flanked by an anti bond and the flap motion pushes the corner atom into the inside of ring. These defects arose since we did not take into account the environment of the corner being flipped.

We have hence implemented a new local perturbation mode called edge flip, which flaps a pair of adjacent corners simultaneously in opposite directions so that the bond between the two corners changes its rotation. There are six possibilities for such a change: a gauche bond changes its sign or becomes anti, while an anti bond transforms itself into the two gauche bonds of opposite sign:


The G-to-G′ [and *vice versa*, Fig. 2(*b*)] as well as A-to-G (and A-to-G′) transformations are essentially the double corner flap, similar to that of chair cyclohexane to twist-boat form, whereas the G-to-A [and G′-to-A, Fig. 2(*c*)] transformation of an endocyclic bond can be achieved by very small inward displacements of the two corner atoms. In practise, the movements of corner atoms in a ring structure are adjusted depending on the rotation patterns of the two neighbouring bonds in order to prevent too highly strained structures from forming.[14]

*Down Stream and Reservoir Filling Strategies.*—The 'down stream' strategy is a small but useful device for driving the conformational search towards lower and lower energy regions, in the outer loop. This surprisingly powerful device simply consists of selecting the lowest-energy conformer for the initial structure of the next local perturbation loop from among those stored conformers which have not yet been used for the initial structure.

Suppose we start the search using conformer S1 of Fig. 3 as the input structure. Our first goal is to find a low-energy region in the conformational space where many good conformers might be found. Perturbations (*vide infra*) of the input structure, followed by geometry optimization, generally produce several new conformers. When the perturbation is small and localized, the resulting conformers are similar to S1 and can be considered to lie in the vicinity of it in the conformational space. If conformers may be compared to stones scattered in space, those generated during one local perturbation loop may be likened to stepping-stones,[10c] all one step away from the initial structure. As mentioned above, the lowest-energy stepping-stone S2 is then selected as the initial structure for the next loop.

In this way, the search moves slowly but steadily down the space towards lower and lower energy space. If, as illustrated, the third loop initiated by S3 generates only stepping-stones which are of higher energy than the initial structure, the lowest energy of the unprocessed conformers, namely S4, is chosen as the next initial structure. After S5, which is the GEM in this illustration, is reached, the search continues by covering all the stepping stones about the next most stable and unprocessed conformer in turn, thus moving, on the whole, slowly upwards. The search ends when no more unprocessed conformers remain in the stored structures. The number of structures to be stored is controlled by setting beforehand the highest energy value for the conformers to be saved (*vide infra*).

The whole scheme resembles the process of pouring water into an empty reservoir or dam: the water stream quickly goes down along the rugged surface while filling and overflowing all the local dips on the way, and, once the real bottom of the reservoir (which corresponds to the GEM) is reached, then the water level keeps going up by filling all the available space until the top of the dam (which corresponds to the pre-fixed limit of search) is reached. Note that there can be more than one domain containing stable and meaningful conformations in the whole conformational space. This point will be referred to later.

*Variable Search Limit Strategy.*—Let us define the purpose of a conformational space search as follows: to find, identify and rank all possible conformers whose energies lie between the GEM and a certain 'chemically significant limit' (CSL, Fig. 4). The limit can be conveniently expressed in term of either the energy relative to the GEM or the lowest possible population.

**Table 2** Test of 'pre-check' option

| Molecule | Perturbation method | Number of Conformers found[a] | Elapsed time/ min (no pre-check) | Elapsed time/ min (pre-check) | Reduction rate (%) |
|---|---|---|---|---|---|
| Cyclodecane | Corner flap | 48 | 19 | 12 | 63 |
| | Edge flip | 57 | 47 | 28 | 59 |
| | Flap and flip | 65 | 68 | 37 | 54 |
| Heptane | Rotation | 30 | 18 | 6 | 31 |

[a] Only the low-energy conformers are sought.

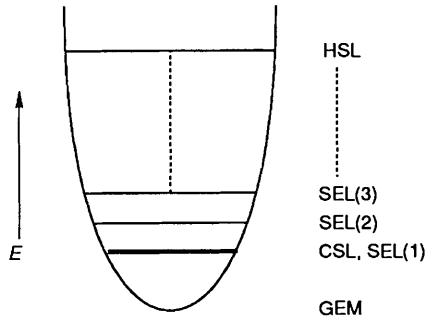



**Fig. 4** Schematic illustration of the conformational space divided by energy levels. The general purpose of the conformation search is to obtain all the conformers between the global energy minimum (GEM) and chemically significant limit (CSL), *e.g.*, those having Boltzmann populations greater than 0.1%. Increasing the search limit from SEL(1) to SEL(2), SEL(3), ... helps to ensure exhaustive coverage of the space between GEM and CSL. HSL = highest storage limit.

The content of conformers in the GEM/CSL range keeps changing during the search as new low-energy conformers are found. The program updates the Boltzmann distribution whenever a new conformer is found. For practical purposes, 3 kcal $mol^{-1}$ in energy or 0.1% in population is recommended as the CSL.

Subjecting all the conformers stored in the region between the GEM and the CSL to the perturbation–optimization process is not a sufficient condition to find all the conformers within this range, because new conformers often emerge by perturbation of conformers above the CSL. In such a case, the conformation search often goes into a new domain in the conformational space separate from the previous one by way of some barrier above the CSL. For this reason, after finishing the GEM/CSL region, the search limit (SEL) is pushed up from CSL [ = SEL (1)] to SEL (2), to subject conformers between CSL and SEL (2) to the outer loop. If a new conformer having an energy level below the present search-limit SEL ($n$) appears, it is immediately used as the next initial structure. SEL is pushed up gradually until no new conformer appears below CSL after several successive increases in the search limit, when the search can be terminated.

All the high-energy conformers should be stored as they are found, since the final SEL is unknown. In practice, however, owing to the finite storage memory size, a highest storage limit (HSL) is set and those conformers having energies higher than HSL are abandoned. We use a default HSL equal to $N/2$ kcal $mol^{-1}$ where $N$ is the total number of atoms in the molecule.

*Pre-check during Geometry Optimization.*—Our search algorithm involves the following redundancy. Suppose the local perturbation of the first bond in an initial structure A produces a new conformer B after geometry-optimization. Later, when the conformer B is selected as the initial structure and its first bond is perturbed, it is most likely that the conformer A will emerge after geometry optimization. Basically, for each conformational transformation found, the reverse path should exist. The knowledge of the reverse path may be desirable when interconversion among conformers is studied. However, we temporarily ignore it because this aspect of the energy hypersurface is not our present concern (but will be a future project). In the 'pre-check' option that we implemented here as a time-saving device to remove redundancy, the structure under geometry optimization is compared, after 10, 25, 50 and 100 iterations and after every 100 more iterations thereafter, with all the stored structures using the dihedral angle criteria (*vide infra*). Whenever the optimizing structure is judged to be identical with one of the stored structures, this structure is immediately abandoned and the next bond is perturbed to produce the next trial structure.

Table 2 compares elapsed time with and without 'pre-check' for cyclododecane and heptane. It may be reasonably concluded that this strategy reduced computer time by 30–60%.

*Comparison of Conformers.*—Quick removal of redundant structures is important for reducing computer time. According to our working criteria, a pair of acyclic structures are considered identical if all the endocyclic rotation angles of one structure agree with the corresponding angles of the other within ±8°. For cyclic structures, conformation distance[4d] is calculated between a pair of structures and they are considered identical when the distance is less than 10°.[15]

For symmetrical molecules like $n$-alkanes, the conformation being compared must be rotated by 180° and mirror-reflected in order to avoid redundancy and identify unique conformations.[16] Namely, dihedral angles are compared in one direction and then in the reverse direction (twice), and the same comparison is repeated after changing the signs of all dihedral angles (twice).

Additional checks regarding the search limit and the preservation of chirality (configurations of asymmetric centres) are performed before finally storing a unique conformer in the conformer storage. A flow chart of this sequence is given in Fig. 5.

## Computational Methods

The new algorithm mentioned above has been implemented into a program CONFLEX3.2,[17] which was originally written for cyclic conformation search.[4f] The present version uses MM2(77)[18] as the minimizer. Geometry optimization is first carried out with the block diagonal Newton–Raphson procedure implemented in MM2. After a conformation space search job is finished, all of the conformers obtained are subjected to re-optimization with the full-matrix Newton–Raphson method implemented in BIGSTRN3[19] using MM2 in order to check if any of them will merge into another conformer and to confirm that they are true energy minima containing six zero or nearly zero eigenvalues in their force constant matrices. These recalculations also allow us to estimate relative free energies of conformers taking into account vibrational and

**Table 3** Changes in the distribution of conformers for lower alkenes with increasing search limit (CONFLEX3/MM2)[a]

| *n*-Alkane | Rel. steric $E^b$/ kcal mol$^{-1}$ | Search limit (%)[c] 1 | 0.2 | 0.1 | 0.02 | 0.01 | 0.002 | 0.001 | 0.0002 | 0.0001 |
|---|---|---|---|---|---|---|---|---|---|---|
| Heptane | 0–1 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | — | — |
| | 1–2 | **18** | 18 | 18 | 18 | 18 | 18 | 18 | — | — |
| | 2–3 | **12** | **14** | **14** | 14 | 14 | 14 | 14 | — | — |
| | 3–4 | 24 | 24 | **24** | **24** | 24 | 24 | 24 | — | — |
| | 4–5 | 8 | 20 | 27 | **28** | **28** | 28 | 28 | — | — |
| | 5–6 | 0 | 0 | 4 | 6 | 6 | **6** | **6** | — | — |
| | 6–7 | 0 | 0 | 1 | 3 | 6 | 8 | **10** | — | — |
| | Total | 71 | 85 | 97 | 102 | 105 | 107 | 109 | — | — |
| | Trial structures | 184 | 312 | 424 | 584 | 680 | 800 | 848 | — | — |
| | Efficiency (%)[d] | 38.6 | 27.2 | 22.9 | 17.5 | 15.4 | 13.4 | 12.9 | — | — |
| | CPU time/min[e] | 3.8 | 5.9 | 7.9 | 11.6 | 15.0 | 17.5 | 19.9 | — | — |
| | Ave. opt. time/s[f] | 1.2 | 1.1 | 1.1 | 1.2 | 1.3 | 1.3 | 1.4 | — | — |
| Octane | 0–1 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| | 1–2 | **32** | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| | 2–3 | 34 | **48** | **48** | 48 | 48 | 48 | 48 | 48 | 48 |
| | 3–4 | 30 | 46 | **48** | **48** | 48 | 48 | 48 | 48 | 48 |
| | 4–5 | 0 | 68 | 78 | 96 | **100** | 100 | 100 | 100 | 100 |
| | 5–6 | 0 | 0 | 4 | 28 | 36 | **36** | **36** | 36 | 36 |
| | 6–7 | 0 | 1 | 3 | 16 | 33 | 48 | **48** | **48** | 48 |
| | 7–8 | 0 | 0 | 0 | 0 | 0 | 14 | 16 | **16** | 16 |
| | 8–9 | 0 | 0 | 0 | 0 | 0 | 4 | 4 | 4 | **4** |
| | 9– | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 4 |
| | Total | 107 | 206 | 224 | 279 | 308 | 341 | 343 | 347 | 347 |
| | Trial structures | 210 | 710 | 910 | 1 470 | 1 910 | 2 650 | 2 770 | 3 290 | 3 310 |
| | Efficiency (%)[d] | 51.0 | 29.0 | 24.6 | 19.0 | 16.1 | 12.9 | 12.4 | 10.5 | 10.5 |
| | CPU time/min[e] | 5.7 | 19.3 | 24.4 | 42.0 | 58.5 | 83.6 | 89.6 | 112.3 | 112.8 |
| | Ave. opt. time/s[f] | 1.6 | 1.6 | 1.6 | 1.7 | 1.8 | 1.9 | 1.9 | 2.0 | 2.0 |
| Nonane | 0–1 | **13** | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 |
| | 1–2 | **50** | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| | 2–3 | 38 | **108** | **108** | 108 | 108 | 108 | 108 | 108 | 108 |
| | 3–4 | 31 | 98 | **122** | **134** | **134** | 134 | 134 | 134 | 134 |
| | 4–5 | 0 | 99 | 193 | 228 | **242** | **242** | 242 | 242 | 242 |
| | 5–6 | 0 | 3 | 17 | 104 | 162 | **192** | **192** | 192 | 192 |
| | 6–7 | 0 | 0 | 2 | 12 | 75 | 142 | 149 | **150** | 151 |
| | 7–8 | 0 | 0 | 0 | 0 | 1 | 78 | 113 | 132 | **146** |
| | 8–9 | 0 | 0 | 0 | 0 | 0 | 5 | 13 | 48 | 57 |
| | 9– | 0 | 0 | 0 | 0 | 0 | 1 | 5 | 6 | 8 |
| | Total | 132 | 371 | 505 | 649 | 785 | 965 | 1 019 | 1 075 | 1 101 |
| | Trial structures | 240 | 1 092 | 1 812 | 31 132 | 4 428 | 7 188 | 8 292 | 10 200 | 10 980 |
| | Efficiency (%)[d] | 55.0 | 34.0 | 27.9 | 20.7 | 17.7 | 13.4 | 12.3 | 10.5 | 10.0 |
| | CPU time/min[e] | 8.7 | 39.2 | 66.8 | 117.4 | 179.2 | 318.5 | 371.4 | 480.5 | 536.1 |
| | Ave. opt. time/s[f] | 2.1 | 2.1 | 2.2 | 2.2 | 2.4 | 2.6 | 2.6 | 2.8 | 2.9 |
| Decane | 0–1 | **15** | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| | 1–2 | **72** | **72** | 72 | 72 | 72 | 72 | 72 | 72 | 72 |
| | 2–3 | 0 | **188** | **198** | 198 | 198 | 198 | 198 | 198 | 198 |
| | 3–4 | 20 | 156 | 268 | **354** | **354** | 354 | 354 | 354 | 354 |
| | 4–5 | 0 | 91 | 385 | 516 | **546** | **554** | 554 | 554 | 554 |
| | 5–6 | 0 | 4 | 6 | 348 | 480 | **685** | **688** | 688 | 688 |
| | 6–7 | 0 | 0 | 1 | 16 | 113 | 380 | 458 | **493** | **493** |
| | 7–8 | 0 | 0 | 0 | 3 | 3 | 141 | 357 | 568 | **587** |
| | 8–9 | 0 | 0 | 0 | 1 | 1 | 6 | 23 | 200 | 254 |
| | 9– | 0 | 0 | 0 | 0 | 0 | 1 | 7 | 33 | 48 |
| | Total | 107 | 526 | 945 | 1 523 | 1 782 | 2 406 | 2 726 | 3 175 | 3 263 |
| | Trial Structures | 210 | 1 414 | 3 010 | 6 538 | 8 750 | 15 974 | 20 594 | 29 932 | 33 936 |
| | Efficiency (%)[d] | 51.0 | 37.2 | 31.4 | 23.3 | 20.4 | 15.1 | 13.2 | 10.6 | 9.6 |
| | CPU time/min[e] | 10.2 | 67.0 | 160.9 | 362.2 | 511.0 | 1002.5 | 1 384.1 | 2 189.4 | 2 593.3 |
| | Ave. opt. time/s[f] | 2.9 | 2.8 | 3.2 | 3.2 | 3.4 | 3.6 | 3.8 | 4.1 | 4.3 |

[a] Bold face numbers are the approximate highest energy values corresponding to the limit in terms of the lowest acceptable Boltzmann population. [b] Relative to the global energy minimum. [c] The lowest limit of Boltzmann distribution. [d] Total number of unique conformers/number of trial structures. [e] Measured on an HP Apollo DN10000. [f] Average time needed to complete geometry-optimization per structure.

symmetry contributions to the partition functions. Nevertheless, since MM2 parameters are not calibrated for vibration, the accuracy of the free energy calculation cannot be assessed. All the calculations have been carried out on an HP Apollo DN10000 workstation.

## Results

*Tests with Small* n-*Alkanes*.—The algorithms mentioned above evolved in the course of studying a variety of structures over a few years. The results for heptane to decane presented below illustrate the performance and characteristics of our

**Table 4** Performances of CONFLEX2 and CONFLEX3 in generating low-energy conformers of cycloalkanes $C_nH_{2n}$

| n | No. of conformers found: CONFLEX2[a] | CONFLEX3 | SEL[b,c]/ kcal mol$^{-1}$ | HSL[b,d]/ kcal mol$^{-1}$ | Comp. time[b]/ min | Total no. of conformers[e] |
|---|---|---|---|---|---|---|
| 9 | 5 | 6 | 5.0 | 13.5 | 2 | 8 |
| 10 | 15 | 17 | 5.0 | 15.0 | 11 | 18 |
| 11 | 27 | 32 | 10.0 | 16.5 | 50 | 40 |
| 12 | 87 | 96 | 9.0 | 18.0 | 141 | 111 |

[a] Ref. 4*f*. [b] Pertains to edge flip runs. [c] The final search limit. [d] Highest storage limit ($= N/2$ kcal mol$^{-1}$, $N$ is the total number of atoms). [e] Ref. 4*d*.

**Table 5** Conformers of cycloundecane[a]

| No. | Rel. $E$[b] kcal mol$^{-1}$ | Boltz. dist.[c] (%) | Stochastic search[d] | Corner flap[e] | Edge flip[f] |
|---|---|---|---|---|---|
| 1 | 0.00 | 48.11 | ○ | ○ | ○ |
| 2 | 0.26 | 31.01 | ○ | ○ | ○ |
| 3 | 0.92 | 10.17 | ○ | ○ | ○ |
| 4 | 1.02 | 4.29 | ○ | ○ | ○ |
| 5 | 1.69 | 1.38 | ○ | ○ | ○ |
| 6 | 1.86 | 1.04 | ○ | ○ | ○ |
| 7 | 2.12 | 0.67 | ○ | ○ | ○ |
| 8 | 2.28 | 1.02 | ○ | ○ | ○ |
| 9 | 2.37 | 0.88 | ○ | ○ | ○ |
| 10 | 2.46 | 0.75 | ○ | ○ | ○ |
| 11 | 2.75 | 0.46 | ○ | ○ | ○ |
| 12 | 3.12 | 0.12 | ○ | ○ | ○ |
| 13 | 4.35 | 0.03 | ○ | ○ | ○ |
| 14 | 4.47 | 0.03 | ○ | ○ | ○ |
| 15 | 4.78 | 0.01 | ○ | ○ | ○ |
| 16 | 5.08 | 0.01 | ○ | ○ | ○ |
| 17 | 5.12 | 0.00 | ○ | ○ | ○ |
| 18 | 6.74 | 0.00 | ○ | ○ | ○ |
| 19 | 7.33 | 0.00 | ○ | ○ | ○ |
| 20 | 8.09 | 0.00 | ○ | ○ | ○ |
| 21 | 8.24 | 0.00 | ○ | — | ○ |
| 22 | 8.34 | 0.00 | ○ | ○ | ○ |
| 23 | 8.42 | 0.00 | ○ | ○ | ○ |
| 24 | 8.87 | 0.00 | ○ | ○ | ○ |
| 25 | 9.46 | 0.00 | ○ | — | ○ |
| 26 | 9.61 | 0.00 | ○ | ○ | ○ |
| 27 | 9.67 | 0.00 | ○ | ○ | ○ |
| 28 | 9.92 | 0.00 | ○ | — | ○ |
| 29 | 10.04 | 0.00 | ○ | ○ | ○ |
| 30 | 10.51 | 0.00 | ○ | — | — |
| 31 | 11.63 | 0.00 | ○ | — | ○ |
| 32 | 12.35 | 0.00 | ○ | ○ | ○ |
| 33 | 13.50 | 0.00 | ○ | — | ○ |
| 34 | 18.21 | 0.00 | ○ | × | × |
| 35 | 18.48 | 0.00 | ○ | × | × |
| 36 | 18.88 | 0.00 | ○ | × | × |
| 37 | 21.08 | 0.00 | ○ | × | × |
| 38 | 25.47 | 0.00 | ○ | × | × |
| 39 | 26.80 | 0.00 | ○ | × | × |
| 40 | 33.98 | 0.00 | ○ | × | × |

[a] '○' means that the conformer is generated. '—' conformer is missed. '×' conformers are out of the significance limit (CSL). [b] Relative MM2-steric energy. [c] Based on relative steric energy. [d] Ref. 4*d*. [e] Ref. 4*f*. [f] Present work.

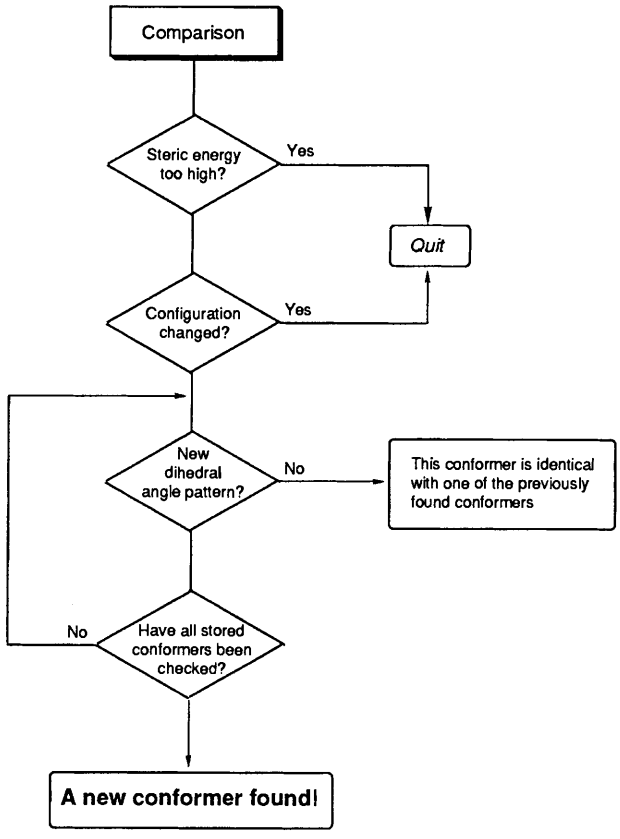



**Fig. 5** Flow diagram for comparison of a freshly optimized conformer with stored structures to remove redundancy

method and disclose inherent difficulties in exploring the conformational space of chain molecules.

The courses of conformational space search for these small alkanes have been monitored in detail during the test calculations (Table 3). The search limit, which was expressed in terms of the lowest permissible population, was increased from 1% to the somewhat too high range of 0.0001% in order to drive the search to completion in these tests. Hence, the extent of search at the end of each SEL, can be computed later by dividing the number of conformers found so far by the final number of conformers. No provision was made to remove redundancy due to symmetry, *i.e.* the results are presented for the general asymmetric $C_6$ to $C_{10}$ chain.

$C_7$ *chain.* We will first check when and how conformers within 4 kcal mol$^{-1}$ from GEM have been found. For such a small system, the conformer search is highly efficient: with a SEL of 0.2%, all of the 65 conformers in the 4 kcal mol$^{-1}$ range have been found after trying 312 structures (enclosed in a box in Table 3). At this point, the search had covered 78% of the whole space.

The efficiency, defined as the probability of finding a new conformer per every trial structure is high, 20–30%, and retains a 12% level even at the end of the search. The efficiency decreases with time as the probability of obtaining duplicated trial structures increases. Duplication is an inherent defect in our algorithm but can be removed as quickly as possible by the 'pre-check' option. Average time of optimizing a trial structure initially tends to remain unchanged but increases with the progress of the search due to the appearance of high-energy, difficult structures, in addition to the increased duplication.

As shown in Fig. 6(*a*), there are three peaks in the distribution of $C_7$ conformers. The first peak is due to the conformers containing two gauche bonds (contiguous in the same sign or separate), the second to a pair of contiguous gauche bonds with

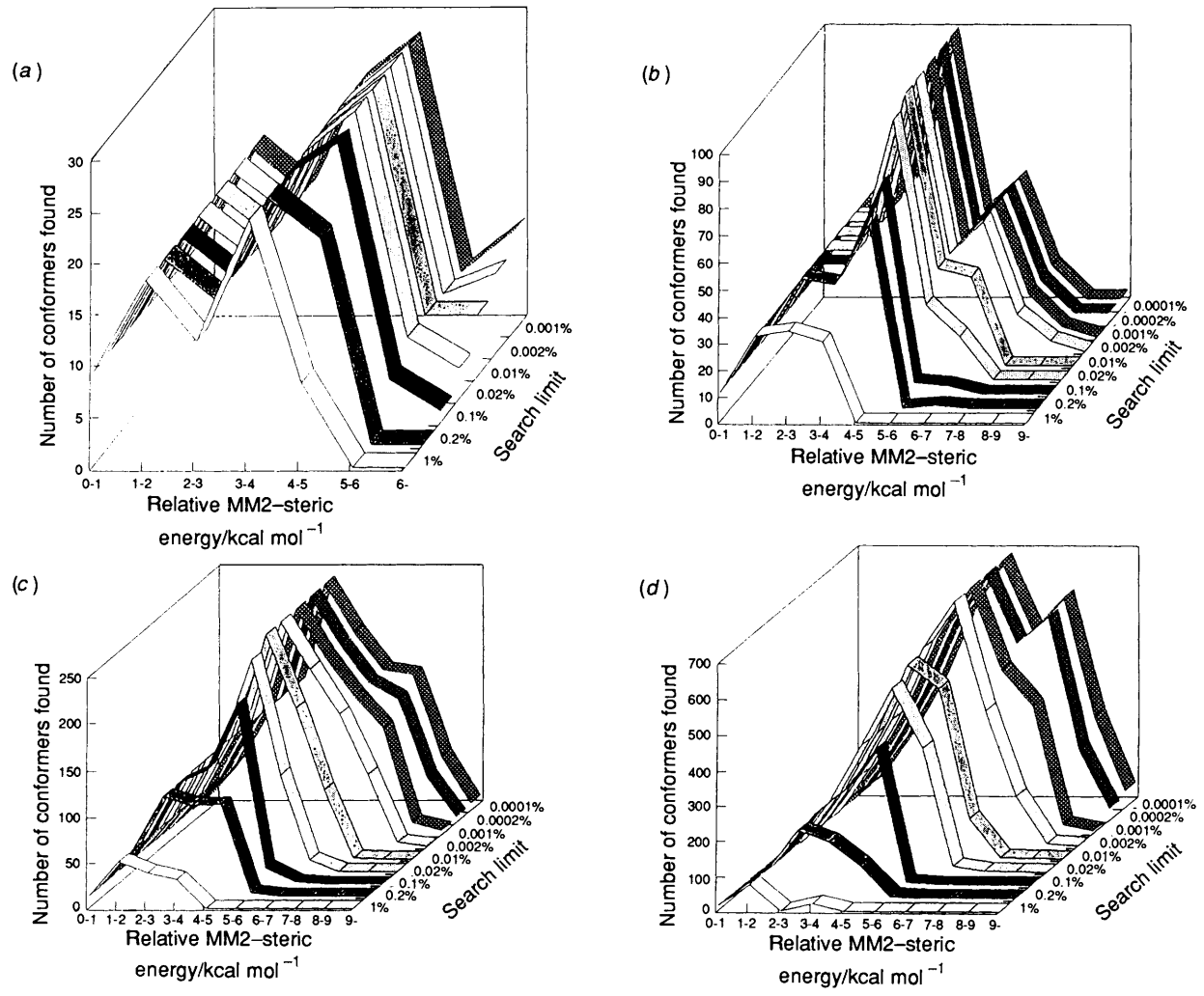



**Fig. 6** Graphic representation of the change in conformer distribution of *n*-alkane (treated as an asymmetric chain) with increasing search limit (SEL), which is given by the lowest population of the high-energy conformer: (*a*) heptane, (*b*) octane, (*c*) nonane, (*d*) decane

opposite sign (GG′) and the last small peak to a pair of GG′ sequences which are independent of each other.[7e]

*$C_8$, $C_9$ and $C_{10}$ chains.* For octane, the 4 kcal mol$^{-1}$ search was likewise highly efficient, having found all of the 139 conformers (boxed) after trying 910 structures and covering 65% of the whole space. The distribution of $C_8$-conformers [Fig. 6(b)] in terms of energy blocks differs from the $C_7$ chain in that the first peak for conformers with two independent gauche bonds is almost unrecognizable, as the result of the increase in the second peak. For nonane, the 4 kcal mol$^{-1}$ search ended when all of the 305 conformers had been found and 59% of the whole space covered, retaining an efficiency of 21% at a SEL value of 0.02%.

The shift of conformer distribution towards higher energy in going to longer chains becomes more evident in nonane [Fig. 6(*c*)] and in decane [Fig. 6(*d*)], wherein the highest peak position moved to the 5–6 kcal mol$^{-1}$ range. Fig. 6(*d*) is especially suggestive of the history of the conformational space search, filling the lowest-energy domain first and gradually covering higher and higher energy space with increasing SEL. The 4 kcal mol$^{-1}$ search of decane was complete when 639 conformers were found at a SEL of 0.02%. At this point, the extent of search was 47% with the efficiency still at the level of 23%.

It is clear from these analyses that the CONFLEX3.2 program indeed effectively finds low-energy conformers. Extrapolation of the above results leads to the prediction that, for longer chain molecules, we may need to cover only about one third of the whole conformational space to find all conformers existing within a few kcal mol$^{-1}$ of the GEM, while maintaining the success rate of about 20%.

It should be added here that the total numbers of conformers after exhaustive search for heptane to decane are considerably larger than $3^m$ where $m$ is the number of internal bonds: $C_7$-chain ($m$ = 4, $3^m$ = 81), $C_8$ (5, 243), $C_9$ (6, 729), $C_{10}$ (7, 2187). The reason is that, when the highly strained GG′ segment appears in the chain, this part deforms unsymmetrically to produce a pair of energy minima having dihedral angles (95°, −63°) or (63°, −95°) for this sequence. The difference between the actual number of conformers and $3^m$ increases with $m$.[7e] The unexpectedly high density of high-energy conformations in the conformational energy space of chain molecules renders the random search method almost unpracticable for long chains.

*Tests with Medium Cycloalkanes.*—Low-energy regions of the conformation space of cycloalkanes $C_9$ to $C_{12}$ have been searched and the over-all results are briefly summarized in Table 4. The new ring perturbation method, edge flip, indeed found more conformers than corner flap did.[4f] The edge flip gives two types of local perturbations for each endocyclic bond, producing twice as many starting structures as the corner flap. The right-end column of Table 4 refers to the results of 'exhaustive' stochastic searches[4d] and these numbers can be regarded as the 'total number' for each cycloalkane tested here. Conformers found by the edge flip search agreed completely up to 6th, 17th, 29th and 48th conformers of cyclononane to cyclododecane, respectively, while those of the corner flap

| No. | Conformation | Steric E. | Successive loop 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | [1331'4'] | 25.57 | x | | | | | | | | | | | | | | | | |
| 16 | [13'44'] | 25.44 | | | | | | o | | | | | | | | | | | x |
| 15 | [143'4'] | 25.43 | | o | | | | | | | | | | | | | | x | |
| 14 | [343*4*] | 25.39 | | o | | | | | | | | | | | | | x | | |
| 13 | [23'3'4] | 25.32 | | o | | | | | | | | | | | | x | | | |
| 12 | [2424] | 25.23 | | o | | | | | | | | | x | | | | | | |
| 11 | [23'4'3] | 25.13 | | | | | | o | | | | x | | | | | | | |
| 10 | [333'3'] | 24.48 | | | | | | | | | | | o | x | | | | | |
| 9 | [1'344'] | 24.11 | o | | | | | | | | x | | | | | | | | |
| 8 | [133'1'4'] | 24.05 | o | | | | | | | x | | | | | | | | | |
| 7 | [13'31'4'] | 23.91 | | | | | | | | | | | | o | x | | | | |
| 6 | [2343] | 23.48 | o | | | | | | x | | | | | | | | | | |
| 5 | [333*4'] | 23.22 | | | o | | | x | | | | | | | | | | | |
| 4 | [2343] | 23.08 | | o | | | x | | | | | | | | | | | | |
| 3 | [1'434'] | 22.35 | | o | | x | | | | | | | | | | | | | |
| 2 | [2334] | 21.68 | o | x | | | | | | | | | | | | | | | |
| 1 | [3333] | 20.61 | | o | x | | | | | | | | | | | | | | |

**Fig. 7** History of conformational space search of cyclododecane starting from the [1331'4'] conformer as the input (the first initial) structure. Columns 1 to 17 correspond to the seventeen outer loops. The initial structure in each loop is denoted by '×', and the newly found conformers by 'O'. Steric energies are given in kcal mol$^{-1}$. For nomenclature of ring conformations, see ref. 15.

**Table 6** Low-energy conformers of cyclononane

| No. | Nomenclature: Dale | Nomenclature: This work | Steric $E$/ kcal mol$^{-1}$ | Rel. $E$/ kcal mol$^{-1}$ | Point group | Boltz. pop. (%) |
|---|---|---|---|---|---|---|
| 1 | [333][a] | [333] | 23.39 | 0.00 | $D_3$ | 22.57 |
| 2 | [225][a] | [225][c] | 24.14 | 0.75 | $C_2$ | 21.10 |
| 3 | [144][b] | [33$\bar{3}$'] | 24.16 | 0.77 | $C_2$ | 46.67 |
| 4 | — | [36'][c] | 25.61 | 2.22 | $C_1$ | 8.62 |
| 5 | [234][a] | [1323'] | 26.55 | 3.16 | $C_1$ | 0.99 |
| 6 | — | [1'44'] | 29.06 | 5.67 | $C_2$ | 0.05 |

[a] Ref. 20. [b] J. Dale, *Top. Stereochem.*, 1977, **9**, 199. [c] This conformer has torsional angles close to 120° on the longest side which cannot be recognized by our nomenclature. See refs. 15 and 20.

search up to 5th, 14th, 20th, and 37th. The 49th cyclododecane conformer generated by our edge flip search is not recorded in the results of the stochastic search.[4d]

Table 5 gives more detailed results for cycloundecane. In our calculations, HSL was set to 16.5 kcal mol$^{-1}$, hence conformers above No. 34 were ignored even if they were discovered. Corner flap found all the conformers down to No. 20, but edge flip did better, finding down to No. 29. While several higher conformers were missed in our calculations, the population becomes virtually zero beyond conformer No. 16 at room temperature.

Hence both of our local perturbation methods can be considered adequate for practical purposes. At the moment, we have not completely abandoned corner flap. As mentioned above, highly strained conformations that are produced from time to time during the corner flap cycle often help the search to jump into a new domain of the conformational space. Combined use of edge flip and corner flap gives the most reliable results at the expense of increased computer time. In CONFLEX3, options are available to perform edge flip or corner flap alone, or a combination of these.

Fig. 7 illustrates how the 'down stream' and 'reservoir filling' strategies worked for a medium-ring molecule. A known cyclododecane conformer,[4f] [1331'4'],[15] having a steric energy 5 kcal mol$^{-1}$ above the GEM was used as the input structure. When the first (inner) loop was completed, four new conformers (No. 9, 8, 6 and 2) were found. The current lowest energy conformer No. 2 was then selected as the next initial structure. The second loop found the GEM in addition to six other new conformers. Since the third loop did not locate any lower point, the most stable of the unprocessed, stored structures, No. 3, became the next initial structure. As the search proceeded, higher and higher points were processed until the eleventh loop, wherein a lower point (No. 10) was found. The next loop initiated from this conformer discovered a still lower point, No. 7, the bottom of a local well which can be considered as belonging to a separate domain. After seventeen loops, all the sixteen conformers of cyclododecane known to exist below the input conformer were found. Following the marks of initial structures (× in Fig. 7), we noticed that the reservoir filling algorithm indeed adjusts itself to the local topography of the conformational space.

This is actually a somewhat oversimplified statement, however. The observation of 'overflow' that occurred in the eleventh loop, suggests possibilities of overflowing over barriers higher than No. 17 leading to conformers more stable than that. In order to be sure that all the interesting conformers, *e.g.* within 5 kcal mol$^{-1}$ of the GEM, are found, it is generally necessary to expand the search beyond this CSL. The 'variable search limit' strategy is used for this purpose. In this case, the search was repeated by setting SEL at 7, then 8, and finally 9 kcal mol$^{-1}$ to confirm that no new conformers appeared below No. 17 (CSL = 5 kcal mol$^{-1}$). In general, the GEM is unknown, hence the actual energy values of SEL and CSL must be updated after every loop based on the steric energy of the current GEM. When large systems like cycloheptadecane[4a] are treated by a long job, it is necessary to check the conformers below CSL frequently and adjust SEL during a job.

Tables 6–8 present all the low-energy conformers of cyclononane to cycloundecane found below the specified HSEL values (Table 4), while Table 9 shows those of cyclododecane lying within 7 kcal mol$^{-1}$ of its GEM. Boltzmann distributions are based on free energy at room temperature calculated by the thermodynamic treatment implemented in BIGSTRN3. Perusal

**Table 7** Low-energy conformers of cyclodecane

| No. | Nomenclature | | Steric *E*/ kcal mol$^{-1}$ | Rel. *E*/ kcal mol$^{-1}$ | Point group | Boltz. pop. (%) |
|---|---|---|---|---|---|---|
| | Dale[a] | This work | | | | |
| 1 | [2323] | [2323] | 24.50 | 0.00 | $C_{2h}$ | 14.10 |
| 2 | [1333] | [1′333′] | 24.92 | 0.42 | $C_2$ | 45.70 |
| 3 | [2233] | [2233] | 25.62 | 1.12 | $C_2$ | 6.66 |
| 4 | [1414] | [1′4′1′4′] | 25.63 | 1.13 | $C_{2h}$ | 9.03 |
| 5 | [1324] | [1′324′] | 26.03 | 1.53 | $C_1$ | 11.25 |
| 6 | — | [13′33′] | 26.76 | 2.26 | $C_1$ | 6.78 |
| 7 | — | [334′] | 27.49 | 2.99 | $C_1$ | 4.15 |
| 8 | — | [232′3′] | 27.68 | 3.18 | $C_2$ | 1.06 |
| 9 | — | [23′23′] | 28.26 | 3.76 | $C_2$ | 0.42 |
| 10 | — | [2224] | 28.58 | 4.08 | $C_s$ | 0.08 |
| 11 | — | [33*33*] | 28.72 | 4.22 | $D_2$ | 0.21 |
| 12 | — | [223′3′] | 28.76 | 4.26 | $C_1$ | 0.47 |
| 13 | — | [1′3′3′3′] | 29.25 | 4.75 | $C_2$ | 0.15 |
| 14 | — | [13′3′3] | 29.42 | 4.92 | $C_2$ | 0.03 |
| 15 | — | [123′13′] | 33.27 | 8.77 | $C_1$ | 0.00 |
| 16 | — | [55] | 33.30 | 8.80 | $D_2$ | 0.00 |
| 17 | — | [12$\underline{5}$2] | 34.81 | 10.31 | $C_2$ | 0.00 |

[a] Ref. 20*a*.

**Table 8** Low-energy conformers of cycloundecane

| No. | Nomenclature | | Steric *E*/ kcal mol$^{-1}$ | Rel. *E*/ kcal mol$^{-1}$ | Point group | Boltz. pop. (%) |
|---|---|---|---|---|---|---|
| | Dale[a] | This work | | | | |
| 1 | [12323] | [233′3] | 24.22 | 0.00 | $C_1$ | 34.49 |
| 2 | [335] | [1334′] | 24.48 | 0.26 | $C_1$ | 29.58 |
| 3 | [12314][b] | [1′33′4′] | 25.14 | 0.92 | $C_1$ | 19.39 |
| 4 | [13223] | [1′3223′] | 25.24 | 1.02 | $C_2$ | 3.08 |
| 5 | [344] | [131′33′] | 25.91 | 1.69 | $C_2$ | 1.34 |
| 6 | — | [1′3′1′33′] | 26.08 | 1.86 | $C_2$ | 3.35 |
| 7 | — | [3333*] | 26.34 | 2.12 | $C_2$ | 2.97 |
| 8 | — | [13′3′4′] | 26.50 | 2.28 | $C_1$ | 1.82 |
| 9 | — | [344] | 26.59 | 2.37 | $C_1$ | 1.57 |
| 10 | — | [1′4*34′] | 26.68 | 2.46 | $C_1$ | 1.76 |
| 11 | — | [23′24] | 26.97 | 2.75 | $C_1$ | 0.30 |
| 12 | — | [22223] | 27.34 | 3.12 | $C_2$ | 0.05 |
| 13 | — | [223′4] | 28.57 | 4.35 | $C_1$ | 0.09 |
| 14 | — | [1′3′1′3′3] | 28.69 | 4.47 | $C_1$ | 0.19 |
| 15 | — | [13′223′] | 29.00 | 4.78 | $C_1$ | 0.03 |
| 16 | — | [134′3] | 29.30 | 5.08 | $C_1$ | 0.01 |
| 17 | — | [11′4′1′4′] | 29.34 | 5.12 | $C_2$ | 0.01 |
| 18 | — | [232′$\underline{4}$′] | 30.96 | 6.74 | $C_1$ | 0.00 |
| 19 | — | [13′13′3] | 31.55 | 7.33 | $C_1$ | 0.00 |
| 20 | — | [12233] | 32.31 | 8.09 | $C_1$ | 0.00 |
| 21 | — | [13223] | 32.46 | 8.24 | $C_2$ | 0.00 |
| 22 | — | [2222′3′] | 32.56 | 8.34 | $C_2$ | 0.00 |
| 23 | — | [13′22′3′] | 32.64 | 8.42 | $C_2$ | 0.00 |
| 24 | — | [122′3′3] | 33.09 | 8.87 | $C_2$ | 0.00 |
| 25 | — | [13133] | 33.68 | 9.46 | $C_2$ | 0.00 |
| 26 | — | [12′4′13′] | 33.83 | 9.61 | $C_1$ | 0.00 |
| 27 | — | [13*31′4′] | 33.89 | 9.67 | $C_1$ | 0.00 |
| 28 | — | [23*4′3] | 34.14 | 9.92 | $C_1$ | 0.00 |
| 29 | — | [124′1′3] | 34.26 | 10.04 | $C_1$ | 0.00 |
| 30 | — | [12332] | 34.73 | 10.51 | $C_2$ | 0.00 |
| 31 | — | [123′1′4′][c] | 35.85 | 11.63 | $C_1$ | 0.00 |
| 32 | — | [122123′] | 36.57 | 12.35 | $C_1$ | 0.00 |
| 33 | — | [12332] | 37.72 | 13.50 | $C_2$ | 0.00 |

[a] Ref. 20 unless otherwise marked. [b] F. A. L. Anet and T. N. Rawdah, *J. Am. Chem. Soc.*, 1978, **100**, 7810. [c] Edge flip search missed this conformer. Taken from ref. 4*d*.

of these Tables reveals an interesting fact; there is a large energy gap of 2.51 kcal mol$^{-1}$ between No. 5 and No. 6 for cyclononane, 3.85 kcal mol$^{-1}$ between No. 14 and 15 for cyclodecane, and 1.62 kcal mol$^{-1}$ between No. 17 and No. 18 for cycloundecane, all with populations of *ca.* 0.01%. These gaps seem to give a working definition of the chemically significant energy limit in these conformational spaces.

On the other hand, energies of cyclododecane conformers (Table 9) are closely spaced, except for a medium-sized gap of 1.07 kcal mol$^{-1}$ between the GEM [3333] and the second lowest energy minimum [2334]. A similar gap (1.08 kcal mol$^{-1}$) is known to exist between the GEM and the second lowest energy minimum of cyclotetradecane. The absence of pronounced energy gaps is commonly observed for large even-membered cycloalkanes.[4p]

It is interesting to note that the GEM's and second lowest

**Table 9** Low-energy conformers of cyclododecane

| No. | Nomenclature | | Steric $E$/ kcal mol$^{-1}$ | Rel. $E$/ kcal mol$^{-1}$ | Point group | Boltz. pop.(%) |
|---|---|---|---|---|---|---|
| | Dale[a] | This work | | | | |
| 1 | [3333] | [3333] | 20.61 | 0.00 | $D_2$ | 78.96 |
| 2 | [2334] | [2334] | 21.68 | 1.07 | $C_1$ | 12.96 |
| 3 | — | [1′434′] | 22.35 | 1.74 | $C_2$ | 4.18 |
| 4 | [2343] | [2343] | 23.08 | 2.47 | $C_s$ | 1.20 |
| 5 | — | [333′4*] | 23.22 | 2.61 | $C_1$ | 0.96 |
| 6 | — | [2343] | 23.48 | 2.87 | $C_1$ | 0.62 |
| 7 | — | [13′31′4′] | 23.91 | 3.30 | $C_s$ | 0.30 |
| 8 | — | [133′1′4′] | 24.05 | 3.44 | $C_1$ | 0.24 |
| 9 | — | [1′344′] | 24.11 | 3.50 | $C_1$ | 0.21 |
| 10 | — | [333′3′] | 24.48 | 3.87 | $C_2$ | 0.11 |
| 11 | — | [23′4′3] | 25.13 | 4.52 | $C_s$ | 0.04 |
| 12 | — | [2424] | 25.23 | 4.62 | $C_2$ | 0.03 |
| 13 | — | [23′3′4] | 25.32 | 4.71 | $C_1$ | 0.03 |
| 14 | — | [343*4*] | 25.39 | 4.78 | $C_s$ | 0.02 |
| 15 | — | [143′4′] | 25.43 | 4.82 | $C_1$ | 0.02 |
| 16 | — | [13′44] | 25.44 | 4.83 | $C_1$ | 0.02 |
| 17 | — | [1331′4′] | 25.57 | 4.96 | $C_1$ | 0.02 |
| 18 | — | [333′3′] | 25.63 | 5.02 | $C_2$ | 0.02 |
| 19 | — | [1353] | 26.07 | 5.46 | $C_2$ | 0.01 |
| 20 | — | [33′33′] | 26.12 | 5.51 | $C_i$ | 0.01 |
| 21 | — | [1′425′] | 26.17 | 5.56 | $C_1$ | 0.01 |
| 22 | — | [1323′3] | 26.47 | 5.86 | $C_1$ | 0.00 |
| 23 | — | [2325] | 26.67 | 6.06 | $C_2$ | 0.00 |
| 24 | — | [2232′3] | 26.69 | 6.08 | $C_1$ | 0.00 |
| 25 | — | [1′323*4′] | 26.71 | 6.10 | $C_1$ | 0.00 |
| 26 | — | [22323′] | 26.73 | 6.12 | $C_1$ | 0.00 |
| 27 | — | [1′313′4′] | 26.95 | 6.34 | $C_1$ | 0.00 |
| 28 | — | [13′233] | 26.99 | 6.38 | $C_1$ | 0.00 |
| 29 | — | [1′4′3′4′] | 27.00 | 6.39 | $C_2$ | 0.00 |
| 30 | — | [4′4′4′] | 27.09 | 6.48 | $C_1$ | 0.00 |
| 31 | — | [2244] | 27.11 | 6.50 | $C_1$ | 0.00 |
| 32 | — | [44′4′] | 27.12 | 6.51 | $C_2$ | 0.00 |
| 33 | — | [2323*3] | 27.19 | 6.58 | $C_2$ | 0.00 |
| 34 | — | [1′5′1′5′] | 27.30 | 6.69 | $D_2$ | 0.00 |
| 35 | — | [23′34′] | 27.31 | 6.70 | $C_1$ | 0.00 |
| 36 | — | [2′33′4] | 27.35 | 6.74 | $C_1$ | 0.00 |
| 37 | — | [2′3′34] | 27.41 | 6.80 | $C_1$ | 0.00 |
| 38 | — | [1′33*33′] | 27.57 | 6.96 | $C_2$ | 0.00 |
| 39 | — | [1′32′3′3] | 27.58 | 6.97 | $C_1$ | 0.00 |

[a] Ref. 20.

energy minima of even-membered cycloalkanes are either square or rectangular conformations, and are characterized by four genuine corners which we have recently found to decrease steric energy significantly owing to attractive van der Waals forces acting within the GG segment.[7d]

## Discussion

Our initial purpose, to develop an algorithm that can efficiently cover the low energy domains of the conformational space of chain molecules, seems to have been fulfilled to a considerable extent. What we report here is probably one of the most powerful strategies for effectively covering chemically significant domains across the vast conformational space of large and flexible molecules.

Some of our basic ideas are close to that of Still.[4e,o] The concepts underlying Still's usage-direct algorithm and our down stream/reservoir filling algorithm are similar, in that the initial structure which undergoes perturbation is selected from the low energy region. However, whereas Still selects at random from the low-energy window, we do it in a simpler and more determined way: to choose the most stable one from the hitherto unperturbed structures. Another point of difference is that we generate as many trial structures as possible from an initial structure, whereas Still's method perturbs one structure only once. His method will eventually achieve similar effect as ours by choosing the same structure many times in the course of random selection.

Our flip/flap perturbation changes three to five contiguous dihedral angles. To our surprise, Still also found that the best results were obtained when three to five dihedral angles were chosen and changed.[4e,o] His observation indirectly supports our assumption that 'a good conformation is surrounded by equally good conformations differing only in local structure in the conformational space'.[10c]

In summary, both of us drive the search to low-energy regions, in a random (Still) or systematic (us) way. It may be noted that, whereas Still's algorithm retains the defects of a random method, we achieved the goal by means of an improved systematic approach. Namely, we maintained the advantages (exhaustive search with clearly defined end-point) while avoiding the disadvantages (explosive increase in the number of trial structures) of systematic searches. We may call it a variable segmented systematic search. In his latest report, Still improved his approach by introducing a variable resolution method.[4o]

One could ask why we did not perform multiple local perturbations to accelerate the search. For example, instead of flapping only one corner at a time, one can flap two or more randomly selected corners simultaneously. Such an operation

may appear to cover the search space more quickly than the succession of single flap/flip process. We tested this possibility but there was essentially no difference between the single and multiple flap/flip processes, as far as the number of discovered conformations was concerned. In some examples, the multiple process missed several conformers that the single process found. This means that accumulation of the single processes may be more systematic than the multiple process.

Like any other methods of conformational space search, our method cannot absolutely guarantee completeness of search. Nevertheless, by comparison with the results of highly exhaustive stochastic search [4d] and the cycloheptadecane case,[4a] we are sure that our program is sufficiently effective for exhaustively searching the chemically significant regions of conformational space.

There are, however, two major problems that remain to be solved. One is redundancy, namely the high rate of repeated appearance of the previously examined structures during the stepwise rotation process. Although the pre-check option works well for small alkanes, we anticipate that the time for pre-check will increase significantly in larger structures (see Table 3). To the author's knowledge, none of the known algorithms of conformation space search has ever avoided the time-wasting redundancy problem.

The second point is that the rotation of one bond in an open chain structure at a time probably does not exactly mimic the thermal conformational interconversion processes actually taking place. The actual process is likely to involve simultaneous or concerted rotation of several contiguous bonds, as in the 'hula twist' movements in the conformational process of rhodopsin,[21] propagation of kinks in monolayers,[22] and migration of vibrational energy along a chain molecule.[23] Replacement of the stepwise rotation with some segmental rotation scheme is expected to increase the efficiency of the conformational space search.

It is to be noted that the three ways of locally perturbing conformation, namely edge flip, corner flap (for rings) and stepwise bond rotation (for chain), convert a conformation not by walking over the edge of a conformational border but by tunneling underneath. It is hoped to improve the local perturbation method in the future to pass over the conformational transition state. Such a device will make it possible to draw conformational interconversion maps of large and flexible molecules, which has so far eluded the molecular modelling capabilities.

Final words should refer to the minimizer. Except for the 'pre-check' option, CONFLEX is completely independent of the minimizer program. In view of the large number of trial structures that must be geometry optimized, at least partly, the choice of practicable minimizer is limited to molecular mechanics types. Still, at least for small systems, semiempirical molecular orbital programs might seem usable. However, a recent comparison of MM2, AM1 and 3-21G methods for the conformers of $C_5$–$C_8$ cycloalkanes reveals some serious defects in the semiempirical MO method. On the other hand, structures and relative energies obtained by MM2 and 3-21G calculations agree well.[24] For this reason, our choice of MM2 as the minimizer can be considered appropriate.

## Acknowledgements

We thank Professor M. Saunders for reading the manuscript at various stages and Dr. C. Jaime for suggesting the 'pre-check' option. This work was partly supported by the Ministry of Education, Science and Culture through Grants-in-Aid for Scientific Research (Nos. 03214105 and 01303013), and by the Taisho Pharmaceutical Company.

## References

1 P. A. Kollman and K. M. Merz, Jr., *Acc. Chem. Res.*, 1990, **23**, 246.
2 E. Clementi, S. Chin, G. Corongiu, J. H. Dietrich, M. Dupuis, D. Folsom, G. C. Lie, D. Logan and V. Sonnad, *Int. J. Quant. Chem.*, 1989, **35**, 319. E. Clementi, *Phil. Trans. Roy. Soc. London, Ser. A*, 1988, **326**, 445.
3 (*a*) A. E. Haward and P. A. Kollman, *J. Med. Chem.*, 1988, **31**, 1669; (*b*) F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson and W. C. Still, *J. Comput. Chem.*, 1990, **11**, 440.
4 Cyclics: (*a*) H. Gotō and E. Ōsawa, *Tetrahedron Lett.*, 1992, **33**, 1343 and references cited therein; (*b*) M. Saunders, K. N. Houk, Y.-D. Wu, W. C. Still, M. Lipton, G. Chang and W. C. Guida, *J. Am. Chem. Soc.*, 1990, **112**, 1419; (*c*) D. M. Ferguson, W. A. Glauser and D. J. Raber, *J. Comput. Chem.*, 1989, **10**, 903; (*d*) M. Saunders, *J. Comput. Chem.*, 1991, **12**, 645; (*e*) J. M. Goodman and W. C. Still, *J. Comput. Chem.*, 1991, **12**, 1110; (*f*) H. Gotō and E. Ōsawa, *J. Am. Chem. Soc.*, 1989, **111**, 8950; (*g*) Y. Sun and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 33; (*h*) D. A. Wertz, C.-X. Shi and C. A. Venanzi, *J. Comput. Chem.*, 1992, **13**, 41; (*i*) T. D. J. Perkins and D. J. Barlow, *J. Mol. Graph.*, 1990, **8**, 156; (*j*) Y.-d. Chen, *J. Chem. Phys.*, 1981, **74**, 2034; (*k*) R. Eminger and S. Fallab, *Chimia*, 1991, **45**, 349; (*l*) G. M. Crippen, *J. Comput. Chem.*, 1992, **13**, 351; (*m*) S. D. Morley, D. E. Jackson, M. R. Saunders and J. G. Vinter, *J. Comput. Chem.*, 1992, **13**, 693; (*n*) C. E. Peishoff and J. S. Dixon, *J. Comput. Chem.*, 1992, **13**, 565; (*o*) G. Chang, W. C. Guida and W. C. Still, *J. Am. Chem. Soc.*, 1989, **111**, 4379; (*p*) E. Osawa, unpublished result.
5 Acyclics: (*a*) C. Jaime, E. Ōsawa, Y. Takeuchi and P. Camps, *J. Org. Chem.*, 1983, **48**, 4514; (*b*) R. W. W. Hooft, J. A. Kanters and J. Kroon, *J. Comput. Chem.*, 1991, **12**, 943; (*c*) S. H. Nilar, *J. Comput. Chem.*, 1991, **12**, 1008; (*d*) R. Czerminski and R. Elber, *J. Chem. Phys.*, 1990, **92**, 5580; (*e*) P. J. Flory, *Statistical Mechanics of Chain Molecules*, Wiley, New York, 1969; (*f*) B. von Freyberg and W. Braun, *J. Comput. Chem.*, 1990, **12**, 1065; (*g*) W. C. Guida, R. S. Bohacek and M. D. Erion, *J. Comput. Chem.*, 1992, **13**, 214; (*h*) I. L. Shamovsky and I. Y. Yarovskaya, *J. Mol. Struct. (Theochem)*, 1991, **236**, 333; *J. Chim. Phys.*, 1991, **88**, 2675; (*i*) I. L. Shamovsky, I. Y. Yarovskaya, N. G. Khrapova and E. B. Burlakova, *J. Mol. Struct. (Theochem)*, 1992, **253**, 149; (*j*) K. A. Palmer and H. A. Scheraga, *J. Comput. Chem.*, 1992, **13**, 329; (*k*) M. E. Snow, *J. Comput. Chem.*, 1992, **13**, 579.
6 (*a*) A. B. Smith, III, J. L. Wood, W. Wong, A. E. Gould, C. J. Rizzo, S. Funayama and S. Omura, *J. Am. Chem. Soc.*, 1990, **122**, 7425; (*b*) Y. Murakami, J. Kikuchi, T. Ohno, T. Hirayama, Y. Hisaeda, H. Nishimura, J. P. Snyder and K. Steliou, *J. Am. Chem. Soc.*, 1991, **133**, 8229.
7 (*a*) E. J. Corey, S. Iguchi, J. O. Albright and B. De, *Tetrahedron Lett.*, 1983, **24**, 37; (*b*) E. J. Corey and J. W. Ponder, *Tetrahedron Lett.*, 1984, **25**, 4325; (*c*) E. Ōsawa, K. Imai, T. Fujiyoshi-Yoneda, C. Jaime, P. Ma and S. Masamune, *Tetrahedron*, 1991, **47**, 4579; (*d*) S. Tsuzuki, L. Schäfer, H. Gotō, E. D. Jemmis, H. Hosoya, K. Siam, K. Tanabe and E. Ōsawa, *J. Am. Chem. Soc.*, 1991, **113**, 4665; (*e*) H. Gotō, E. Ōsawa and M. Yamato, *Tetrahedron*, in the press. (*f*) N. Nakajima, K. Uoto, T. Matsushima, O. Yonemitsu, H. Gotō and E. Ōsawa, *J. Org. Chem.*, 1990, **55**, 1129; (*g*) N. Nakajima, T. Matsushita, O. Yonemitsu, H. Gotō and E. Ōsawa, *Chem. Pharm. Bull.*, 1991, **39**, 2819; (*h*) K. Shimazaki, K. Mori, K. Okada, T. Chuman, H. Gotō, E. Ōsawa, K. Sakakibara and M. Hirota, *J. Chem. Soc., Perkin Trans. 2*, 1992, 811; (*i*) T. Chuman, K. Shimazaki, M. Mori, K. Okada, H. Gotō, E. Ōsawa, K. Sakakibara and M. Hirota, *J. Chem. Ecol.*, 1990, **16**, 2877; (*j*) H. Gotō and E. Ōsawa, *Pure and Appl. Chem.*, 1989, **61**, 597.
8 J. A. Semlyen, *Chem. Brit.*, 1982, **18**, 704; *Pure Appl. Chem.*, 1981, **53**, 1797.
9 G. M. Crippen and T. F. Havel, *Distance Geometry and Molecular Conformation*, Research Studies Press, Somerset, 1988.
10 Previous brief communications: (*a*) ref. 4*a*; (*b*) ref. 4*f*; (*c*) E. Ōsawa and H. Gotō, in *Computer Aided Innovation of New Materials*, eds. M. Doyama, T. Suzuki, J. Kihara and R. Yamamoto, Elsevier Sci. Publ. B.V. (North-Holland), Amsterdam, 1991, p. 411–416.
11 RINGMAKER: W. C. Still, in *Current Trends in Organic Synthesis*, ed. H. Nozaki, Pergamon Press, Oxford, 1983, p. 233–246.
12 J. W. H. M. Uiterwijk, S. Harkema, B. W. van der Waals, F. Gobel and H. T. M. Nibbeling, *J. Chem. Soc., Perkin Trans. 2*, 1983, 1843.
13 MULTIC: M. Lipton and W. C. Still, *J. Comput. Chem.*, 1988, **9**, 343.
14 H. Gotō, M.A. Thesis, Hokkaido University, 1991.
15 H. Gotō, *Tetrahedron*, 1992, **48**, 7131.
16 K. Hassler, *Monatsh. Chem.*, 1979, **110**, 419.

17 CONFLEX3.2: H. Gotō and E. Ōsawa, *Japan Chemistry Program Exchange* P040. See also, JCPE P021 and QCPE #592.
18 U. Burkert and N. L. Allinger, *Molecular Mechanics*, American Chemical Society, Washington, D.C., 1982.
19 R. B. Nachbar, Jr. and K. Mislow, QCPE #514.
20 (*a*) J. Dale, *Acta Chem. Scand.*, 1973, **27**, 1115; (*b*) G. Borgen, J. Dale and G. Teien, *Acta Chem. Scand., Ser. B*, 1979, **33**, 15.
21 R. S. H. Liu and D. T. Browne, *Acc. Chem. Res.*, 1986, **19**, 42.
22 S. H. Northrup, *J. Phys. Chem.*, 1984, **88**, 3441.
23 A. Lami and G. Villani, *J. Phys. Chem.*, 1990, **94**, 6959.
24 D. M. Ferguson, I. R. Gould, W. A. Glauser, S. Schroeder and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 525.

*Paper* 2/04655E
*Received* 1*st September* 1992
*Accepted* 26*th October* 1992