10-2002

A comparison of the Low Mode and Monte Carlo conformational search methods

Carol A. Parish
University of Richmond, cparish@richmond.edu

Rosina Lombardi

Kent Sinclair

Emelyn Smith

Alla Goldberg

See next page for additional authors

Recommended Citation

Parish, Carol A., Rosina Lombardi, Kent Sinclair, Emelyn Smith, Alla Goldberg, Melissa Rappleye, and Myrianne Dure. "A Comparison of the Low Mode and Monte Carlo Conformational Search Methods." *Journal of Molecular Graphics and Modelling* 21, no. 2 (October 2002): 129-50. doi:10.1016/S1093-3263(02)00144-4.

Authors
Carol A. Parish, Rosina Lombardi, Kent Sinclair, Emelyn Smith, Alla Goldberg, Melissa Rappleye, and Myrianne Dure

ELSEVIER

Journal of Molecular Graphics and Modelling 21 (2002) 129–150

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

A comparison of the Low Mode and Monte Carlo conformational search methods

Carol Parish*, Rosina Lombardi, Kent Sinclair, Emelyn Smith, Alla Goldberg, Melissa Rappleye, Myrianne Dure

Department of Chemistry, Hobart and William Smith Colleges, Geneva, NY 14456, USA

Received 17 April 2002; accepted 2 July 2002

Abstract

The Low Mode (LM) and Monte Carlo (MC) conformational search methods were compared on three diverse molecular systems; (4*R*, 5*S*, 6*S*, 7*R*)-hexahydro-5,6-dihydroxy-1,3,4,7-tetrakis(phenylmethyl)-2H-1,3-diazapin-2-one (**1**), 2-methoxy-2-phenyl-2-triflouromethyl-*N*-α-methyl benzyl propanamide (**2**) and a trimeric 39-membered polyazamacrolide (**3**). We find that either method, or a combination of the methods, is equally efficient at searching the conformational space of the smaller molecular systems while a 50:50 hybrid of Low Mode and Monte Carlo is most efficient at searching the space of the larger molecular system.

Keywords: Low Mode; Monte Carlo; Conformational analysis; Molecular flexibility; Cyclic urea; Mosher amide; Polyazamacrolide

1. Introduction

Conformational search methods are used to explore molecular potential energy surfaces (PESs) and to identify low energy structures on the surface. The identification of low energy, accessible states yields information about molecular flexibility and conformational behavior, two concepts that are important in understanding a variety of phenomena at the molecular level. Biological, physical, and chemical properties depend on the nature of the accessible conformations [1]. Our ability to understand molecular flexibility will lead to advances in the design of synthetic materials, the synthesis and efficacy of drugs, the understanding of surface catalysis, and the development of environmental and biological sensors.

With recent advances in software algorithms and computer hardware, it is now possible to study the conformational behavior of much larger, more flexible systems than ever before. However, for the results of a conformational search to be representative of the actual behavior of a molecular system, three criteria must be satisfied [2,3]. First, the molecular potential energy surface upon which the search takes place must be well defined and include environmental effects such as solvent. Many conformational search techniques are performed on surfaces generated using molecular mechanics force fields. In these cases, it is necessary to evaluate the accuracy of the force field used to represent the molecular system under study. Second, conformational sampling methods must be highly efficient and able to locate all low energy structures on the PES in a reasonable amount of computing time. Conformational search results may not be representative of the molecular behavior and flexibility if the method is only able to sample part of the multidimensional potential energy surface. Third, reasonable convergence criteria should be used to determine if a search is exhaustive. Searches started from different points on the PES must identify the same global minimum energy structure and the same number of unique local minima. Results must be time invariant—i.e. searches must be run until global minimum energies no longer change and no new structures are found.

Many of the early conformational search methods were based on a systematic search of a small number of rotatable bonds; however, applications quickly outgrew this method as the number of conformations that need to be evaluated energetically grow exponentially as the number of such bonds in the molecule increases [4]. Today, a variety of efficient conformational searching methods are available using simulated annealing [5–7], distance geometry [8–12], Monte Carlo (MC) [13–18], and eigenvector-following [19–21] algorithms. Comprehensive reviews of conformational searching algorithms have been published elsewhere [1,22–24]. This study compares the Low Mode (LM) [19,20] and Monte Carlo [15] conformational search methods as

* Corresponding author.
E-mail address: parish@hws.edu (C. Parish).

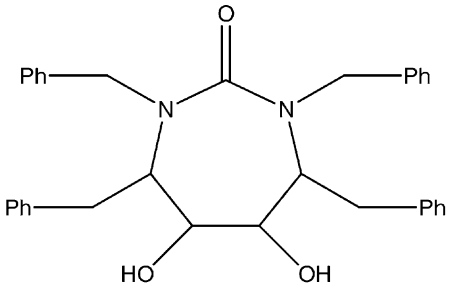
1093-3263/02/$ – see front matter
PII: S1093-3263(02)00144-4

they are implemented in the MacroModel/Maestro molecular modeling software program [25].

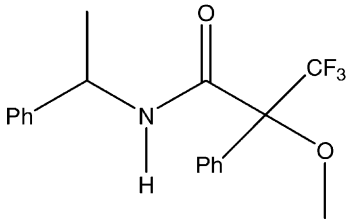
Conformational searching methods generate starting conformations that are locally energy minimized to locate low energy minima on the PES. The MC conformational search method generates starting conformations by random variation of molecular torsions and proceeds as follows. The starting conformation is randomly chosen from the potential energy surface and then minimized into its local minimum energy well. The torsional bonds of the newly minimized structure are then rotated by random increments to move to a different point on the potential energy surface. The search is biased toward regions of low energy by choosing starting conformations for each step from among the previously identified low energy structures. Surface coverage is obtained by selecting the least used structures as starting geometries for the next step and by randomly varying a random number of torsional rotations in each conformational search step. When the same conformations have been visited many times over and no new conformations are found, the search may be considered converged.

The LM conformational search method is based upon an eigenvector-following technique [26] whereby the network of interconnected minima on a PES are randomly searched. The starting conformation for an LM search is a local minimum structure that is subjected to a normal mode analysis to identify the low energy or low mode directions of motion. The starting structure is then perturbed by taking a fixed step along the direction of one of the eigenvectors chosen at random from among the low energy eigenvectors. Subsequent energy minimization identifies a second local minimum structure that is associated with the starting conformation's saddle point. This process is then repeated to find additional minima. The LM method is quite efficient due to the fact that the search is focused on the minimum energy paths along which one degree of freedom (d.f.) is at its maximum and the remaining d.f. are at their minima.

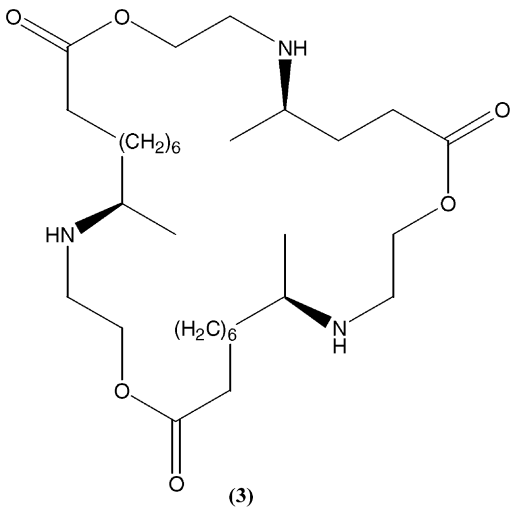
The present report compares the LM and MC conformational search methods for determining the conformational flexibility of three diverse molecular systems; a potent, cyclic urea HIV-1 protease inhibitor [27] (4*R*, 5*S*, 6*S*, 7*R*)-hexahydro-5,6-dihydroxy-1,3,4,7-tetrakis(phenylmethyl)-2H-1,3-diazapin-2-one (**1**); 2-methoxy-2-phenyl-2-triflouromethyl-*N*-α-methyl benzyl propanamide (**2**), which is a diastereomer formed from α-methylbenzyl amine and a Mosher's acid chloride [28] and a trimeric 39-membered



(1)



(2)



(3)

Fig. 1. Molecular systems used for comparing the Low Mode and Monte Carlo conformational search methods.

polyazamacrolide (**3**) [29]. The molecular structures are shown in Fig. 1. Molecule **2** contains 6 rotatable bonds and should be well managed by any modern conformational search method while **1**, which is a cyclic system with 14 rotatable bonds, is a perfect candidate for exploring the limits of a conformational search method. Compound **3**, which is a macrocycle containing 34 rotatable bonds, represents a major computational challenge to any method. We are interested in the conformational behavior of each of these systems and believe that a comprehensive study comparing LM and MC techniques will establish the criteria for the selection of the most appropriate methods to use.

2. Methods

2.1. Generating potential energy surfaces

The conformational ensembles that are generated in this study were calculated using the MacroModel V7.0 (version 7.0) [25] suite of software programs running on 800 MHz Athlon PC's under the RedHat LINUX 6.2 operating system. Seven different force fields are available in the MacroModel program. It is important to choose a force field that is well parameterized for the molecular system under study. Accurate torsional parameters are particularly important in flexible molecular systems since they control conformational interconversions. A comparison of the parameter quality for each of the three molecular systems studied is summarized in Table 1. The AMBER94 [30] force field as implemented in MacroModel V7.0 did not contain the parameters necessary to calculate the energies of each of these systems and was, therefore, not included in the comparison.

AMBER* [31] and OPLS-AA [32,33] as implemented in MacroModel V7.0 were the best parameterized force fields as they contained the fewest low quality torsion parameters for molecules **1**–**3**. AMBER* was chosen to describe the potential energy surfaces due to its ability to represent the molecule with (all-atom representation) and without (united-atom representation) hydrogen atoms. United-atom calculations are computationally less demanding than all-atom calculations since there are fewer d.f. We were interested in comparing the results from the united-atom AMBER* surface with those from the all-atom AMBER* surface, so we chose to perform our calculations with AMBER*. Solvent effects were included using the generalized born/surface area (GB/SA) continuum model [34–39] for water and chloroform. GB/SA has been shown to reproduce accurately the hydration free energies for various molecular systems [36]. Non-bonded interactions within 8 Å for van der Waals' and 20 Å for electrostatic interactions were included in all calculations employing the GB/SA model.

Table 1
Force field parameter analysis for (4*R*, 5*S*, 6*S*, 7*R*)-hexahydro-5,6-dihydroxy-1,3,4,7-tetrakis(phenylmethyl)-2H-1,3-diazapin-2-one (**1**), a 38-membered cyclic polyazamacrolide (**3**), and 2-methoxy-2-phenyl-2-triflouromethyl-*N*-α-methyl benzyl propanamide (**2**)

Force field	Interaction type	CU			AMBA			PAML		
		H	M	L	H	M	L	H	M	L
AMBER* (all-atom)	Stretch	49	3	0	28	0	3	96	18	0
	Bend	108	23	0	53	20	0	198	18	3
	Torsion	63	44	0	33	18	1	243	69	0
AMBER* (united-atom)	Stretch	17	3	0	11	0	3	30	18	0
	Bend	36	23	0	18	20	0	36	18	3
	Torsion	26	22	0	20	15	1	27	33	0
MMFF	Stretch	76	0	0	43	0	0	114	0	0
	Bend	129	0	2	71	0	2	219	0	0
	Torsion	179	0	24	80	0	20	309	0	3
MM2*	Stretch	52	4	0	29	4	0	111	12	0
	Bend	88	20	1	50	12	0	192	51	0
	Torsion	39	48	8	14	31	7	225	129	6
MM3*	Stretch	21	3	4	10	1	8	96	6	12
	Bend	28	18	3	16	8	9	207	3	9
	Torsion	41	20	22	11	10	19	276	15	21
OPLS-A*	Stretch	49	3	0	28	0	3	108	6	0
	Bend	128	1	2	62	0	11	189	21	9
	Torsion	69	18	20	33	13	6	243	69	0
OPLS-AA	Stretch	76	0	0	43	0	0	114	0	0
	Bend	131	0	0	73	0	0	219	0	0
	Torsion	199	4	0	96	4	0	312	0	0

Parameter qualities: H = high, M = medium, L = low.

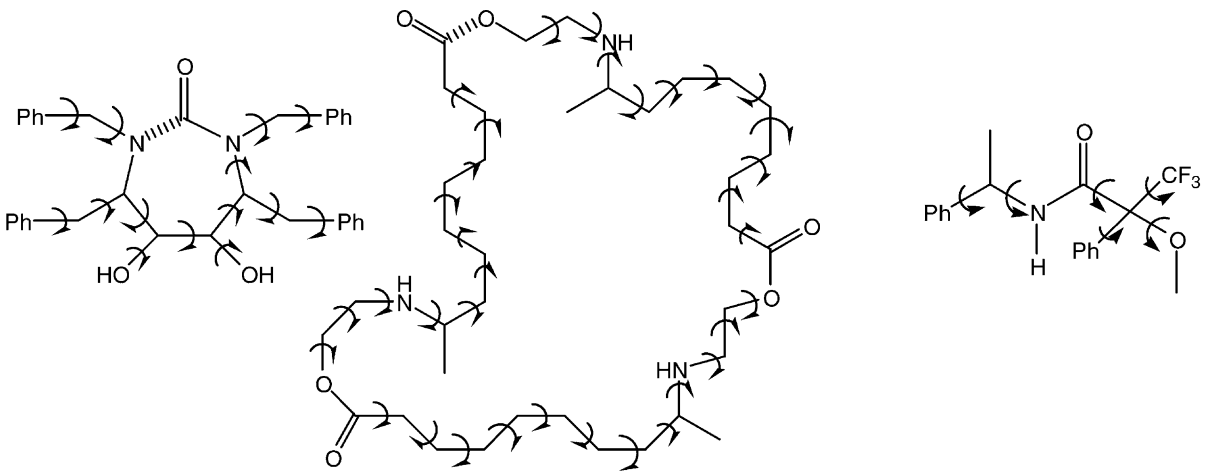


Fig. 2. The d.f. varied during the conformational searches. Arrows represent torsions that were allowed to vary and hashed lines indicate ring-opening bonds.

2.2. Conformational searches

2.2.1. MC conformational searches

The MC conformational searches in this study explored torsional space. This method has been shown to provide efficient conformational interconversion and sampling of complex potential energy surfaces [15,40–42]. Interconversion of ring structures was enabled using the ring-opening method of Still and Galynker [43]. In this method, the ring is opened by breaking a bond between two ring atoms while allowing other ring torsions to vary. After generating a new set of ring torsions the bond is closed subject to closure constraints. Variable torsions and ring-opening bonds are illustrated in Fig. 2 for each of the systems in this study. The minimum and maximum allowable closure distances for **1** were 0.5 and 2.0 Å, respectively. The minimum and maximum allowable closure distances for the larger ring system **3** were 0.1 and 30.0 Å, respectively. Closure distances outside of these ranges were discarded before minimization. Each MC conformational search step varied a random number of torsional d.f. between a minimum of two and a maximum of *N*, where *N* is the total number of variable torsion angles. There were 14 variable d.f. in **1**, 6 in **2** and 34 in **3**.

2.2.2. LM conformational searches

The potential energy surfaces of the three molecular systems were also explored using the LM conformational search method [19]. The major advantages of this method are not having to define rotatable torsions or needing to open and close the ring to allow conformational ring interconversions. The ability to sample a potential energy surface without specifying d.f. is particularly advantageous when performing conformational searches on large structures or multiconformer structure files. With LM, conformational interconversions occur by movement along the potential energy surface that follows the low frequency vibrational modes as the only d.f. In this report, LM frequencies corresponding to the 10 lowest eigenvectors were explored. The total traveling distance for each step was selected randomly between 3 and 6 Å.

2.2.3. Combined LM:MC conformational searches

The LM search method was also used in combination with the MC algorithm. Various ratios of LM to MC were tested. In this hybrid method, explicit torsional rotations are combined with movement along the minimum energy paths that connect low energy structures. It is expected that the hybrid method will take advantage of the strengths of each individual method, i.e. the LM method will provide effective local sampling while the MC method will make global movements on the potential energy surface, locating previously non-sampled regions.

2.3. General search strategy common in all three types of conformational searches

Starting structure chirality was preserved throughout the conformational searching. Sampling of amide bonds was constrained to the *trans* orientation (i.e. $\theta = 180 \pm 90^{\circ}$). Where possible, searches were run in blocks of 1000 or 5000 steps until no new unique conformations were found within 25 kJ/mol of the global minimum and the energy of the global minimum was converged. Unique conformations were determined by superimposition of all heavy (non-hydrogen) atoms as well as reflection and/or rotation of the atom-numbering scheme. Structures were considered to be duplicates and were rejected if the maximum interatomic distance was 0.25 Å or less following optimal RMS superposition. Structures that were found in previous searches were used to seed subsequent searches. Searches utilized the usage-directed structure selection method [15] that selects the least used structure from among all known conformations as the starting structure for each new search. This insures that a variety of different starting structures

from different regions of the potential energy surface are used to initiate each search. During the conformational search, all structures were subjected to 2000 (**1** and **2**) or 6500 (**3**) steps of the truncated Newton conjugate gradient (TNCG) [44] minimization method to within a derivative convergence criterion of 0.01 kJ/(Åmol). The ability of a method to search exhaustively a potential energy surface was judged by monitoring (1) the energy of the lowest energy structure, (2) the number of conformations found, and (3) the frequency with which the simulation visited the lowest energy structure.

2.3.1. Clustering ensembles

Ensembles generated with each conformational search method were grouped into geometrically similar families using the XCluster [45] program. XCluster calculates the pairwise distance between each structure, in either torsional or Cartesian space. It partitions the conformations into geometrically similar subsets in an agglomerative, hierarchical fashion. The process begins with every structure as the only member of its own cluster. Individual structures are then grouped into clusters using the shortest distance between points as the threshold distance. At each clustering level, the next shortest distance is used to form new, agglomerative clusters, with later clusters formed from groupings of earlier clusters. This process continues until all structures are a member of the same final cluster. The goal is to find the clustering level at which the distance between members of clusters is much smaller than the distance between clusters; i.e. the minimum separation ratio. Separation ratios greater than 2 that occur at high cluster levels indicate significant clustering [45]. For a given clustering level, the full distance matrix was used to visualize the clustering of molecular structures in the ensembles that were generated by each method. The clustering mosaics were used to illustrate how the clusters agglomerated as the clustering proceeded from the first level to the $N - 1$ level.

3. Results and discussion

The three different molecular systems shown in Fig. 1 were used to compare the efficiency of the LM, MC, and combined LM:MC conformational search methods.

*3.1. (4R, 5S, 6S, 7R)-Hexahydro-5,6-dihydroxy-1,3,4, 7-tetrakis(phenylmethyl)-2H-1,3-diazapin-2-one (**1**)*

The cyclic urea **1** contains four chiral centers and exists in 10 different stereoisomeric forms. The RSSR stereoisomer was focused on in this study since it shows the highest binding affinity for the HIV-1 protease [27]. The conformational flexibility of the RSSR stereoisomer was investigated using the AMBER* all-atom force field, the GB/SA model for water and five different conformational search strategies: pure LM, pure MC and three combinations of the LM and MC algorithms—LM:MC 75:25, LM:MC 50:50, and LM:MC 25:75. The searching efficiency of each method in a fixed number of steps for **1** was compared. Table 2 illustrates the number of unique structures of **1** that were found in the first 5000 steps for each algorithm. All searches were initiated from the same starting conformation that happened to be 0.74 kJ/mol above the lowest energy structure found overall. All methods find the same lowest energy structure of **1** in 5000 steps. This structure will be shown by lengthier calculations, which are described later, to be the overall lowest energy structure on the AMBER* potential energy surface. Pure MC is much slower than the other methods and, while it finds the maximum number of structures in a 5000 step search, the number of conformations found per unit time is significantly less than with the other methods. At this stage, the LM and 25:75 LM:MC hybrid methods are the most efficient since they find the same lowest energy structure as the other methods and the largest number of structures per hour.

Another way of evaluating conformational search results is to ask how many conformational search steps are necessary to obtain exhaustive coverage of the PES. To judge convergence, calculations on **1** were run in blocks of 5000 steps for a total of 55,000 steps while monitoring the energy of the lowest minimum and recording the number of unique structures found. Each 5000 step search block began with a new starting geometry. The energy of the lowest structure, the number of conformations found and the frequency with which the simulation visits the lowest structure are shown in Table 3 for each of the five methods. All methods found the same lowest energy structure (same energy with superimposition RMSD = 0.0 Å) and each ensemble resulting from 55,000 steps contain a similar number of structures (∼700). The energy of the lowest structure remains constant very early in each calculation and this structure was visited ∼100

Table 2
Comparing LM, MC, and LM:MC 50:50 search methods on **1** after 5000 steps

Method	Number of conformations found	Number of global minimum visits	Minimum energy (kJ/mol)	CPU time (h)	Number of conformations found per hour
LM	485	13	−81.09	5.43	89.32
LM:MC 75:25	461	7	−81.08	6.10	75.57
LM:MC 50:50	556	5	−81.08	6.42	86.60
LM:MC 25:75	601	11	−81.08	6.63	90.65
MC	648	14	−81.09	9.18	70.59

Table 3
Comparing conformational search results on **1** using the LM, MC, and LM:MC methods for a 55,000 step search run in blocks of 5000 steps

Number of steps	Number of conformations found	Number of global minimum visits	Global minimum energy (kJ/mol)	CPU time (h)
Method: LM				
5000	485	13	−81.09	5.43
5000	563	24	−81.09	7.25
5000	589	39	−81.09	7.38
5000	611	51	−81.09	7.90
5000	637	67	−81.09	7.92
5000	626	77	−81.10	7.80
5000	642	87	−81.09	8.00
5000	642	94	−81.09	8.40
5000	640	106	−81.09	7.98
5000	659	117	−81.09	8.17
5000	663	128	−81.10	8.15
Total				84.38
Method: LM:MC 75:25				
5000	461	7	−81.08	6.10
5000	594	12	−81.08	7.37
5000	638	15	−81.08	8.02
5000	661	19	−81.08	8.20
5000	663	27	−81.10	8.23
5000	680	33	−81.08	8.65
5000	689	39	−81.09	8.40
5000	702	45	−81.09	8.67
5000	685	53	−81.09	8.55
5000	697	60	−81.09	8.87
5000	690	72	−81.09	8.50
Total				89.56
Method: LM:MC 50:50				
5000	556	5	−81.08	6.42
5000	621	12	−81.09	8.03
5000	652	24	−81.09	8.93
5000	663	33	−81.09	9.08
5000	675	41	−81.09	9.73
5000	686	50	−81.09	9.30
5000	686	60	−81.09	9.23
5000	675	68	−81.09	8.88
5000	685	74	−81.09	10.00
5000	690	84	−81.10	9.27
5000	695	92	−81.08	8.85
Total				97.72
Method: LM:MC 25:75				
5000	601	11	−81.08	6.63
5000	670	19	−81.09	8.55
5000	694	26	−81.09	9.22
5000	684	43	−81.09	8.95
5000	687	48	−81.09	8.83
5000	703	53	−81.09	9.00
5000	699	63	−81.09	9.20
5000	705	75	−81.09	9.47
5000	698	85	−81.09	9.17
5000	695	95	−81.09	8.95
5000	702	108	−81.09	9.47
Total				97.44
Method: MC				
5000	648	14	−81.09	9.18
5000	680	21	−81.09	11.10
5000	686	30	−81.09	11.33
5000	681	37	−81.09	11.28
5000	694	42	−81.09	11.48

Table 3 (*Continued*)

Number of steps	Number of conformations found	Number of global minimum visits	Global minimum energy (kJ/mol)	CPU time (h)
5000	691	49	−81.09	11.33
5000	690	64	−81.09	11.45
5000	693	70	−81.09	11.52
5000	695	82	−81.09	11.57
5000	696	87	−81.08	11.58
5000	701	91	−81.09	11.50
Total				123.32

times per 5000 steps in the later stages of the study. In addition, the rate of finding new structures decreased substantially after the first few blocks of 5000 search steps. These results suggest that the search was no longer finding new regions of conformational space and was instead revisiting already sampled regions. Any new structures found during the later stages of the search were simply not fully converged duplicates of existing ensemble structures.

The random barrier "hopping" nature of MC and the minimum energy path mode following nature of LM are clearly evidenced for **1** in the data (Table 3). Thus, MC found a large number of structures very early in the search whereas LM seemed to find new structures more slowly. This result is seen by comparing the number of structures found between the first and second block of 5000 steps with each method. For example, MC finds 648 structures for **1** in the first block and only 32 more (a 4.9% increase) structures for **1** in the second block. On the other hand, LM finds 485 structures for **1** in the first 5000 steps and 78 more (a 16% increase) in the second block.

Each structure in each ensemble, which was generated after 55,000 search steps, was subjected to 14,000 additional steps of TNCG minimization to ensure a well-minimized ensemble of unique structures. Table 4 shows the resulting number of conformations along with a summary of the total time necessary for 55,000 conformational search steps and the efficiency of each method as determined by the number of conformations found per CPU time (h). Since the total number of fully minimized, unique structures is less than the number of structures found after each block of 5000 search steps in the later stages of the conformational search, it is likely that the search results in Table 3 were converged in the early stages of the search. The LM and LM:MC 75:25 methods have generated a converged ensemble after approximately 15,000 steps while the LM:MC 50:50 and LM:MC 25:75 methods appear converged after 10,000 steps and the MC after 5000 steps. This information suggests that the lowest energy structure found is, indeed, the global minimum energy structure and that the generated ensembles are representative of the low energy structures of **1** on the AMBER* surface. Of course, only quantum mechanical calculations can guarantee a correct energetic ordering of the minimum energy structures.

Not surprisingly, the results generated with the LM:MC hybrid methods lie between those of pure LM and pure MC methods with respect to the number of conformers found and the time consumed. The efficiency of each method can be determined by comparing the number of conformations found per hour (Table 4). It is clear that pure LM is the most efficient method for searching the conformational space of **1** since it generated 6.94 conformations per hour. Pure MC, on the other hand, appears to be the least efficient search method for **1** since it finds only 5.04 low energy structures per hour. Interestingly, the LM:MC 25:75 results are more efficient than would be expected from a simple analysis of the efficiency versus LM:MC ratio in each method. This may signal a greater influence of the pure LM method than is indicated by its percentage weighting.

Efficiency considerations must include an analysis of the structures in the resulting ensembles. For instance, the LM method may be the most efficient for searching the conformational search of **1** as judged by the rate of conformations found per hour. However, the LM ensemble contains fewer structures than do the ensembles generated by the other methods. This may indicate that the LM method is not as efficient as the MC method at finding new minima distant

Table 4
Number of conformations of **1** after 14,000 additional minimization steps

Method	Number of conformations found	Number of global minimum visits	CPU time (h)	Number of conformations found per hour
LM	586	128	84.38	6.94
LM:MC 75:25	611	72	89.56	6.82
LM:MC 50:50	611	92	97.72	6.25
LM:MC 25:75	622	108	97.44	6.38
MC	621	91	123.32	5.04

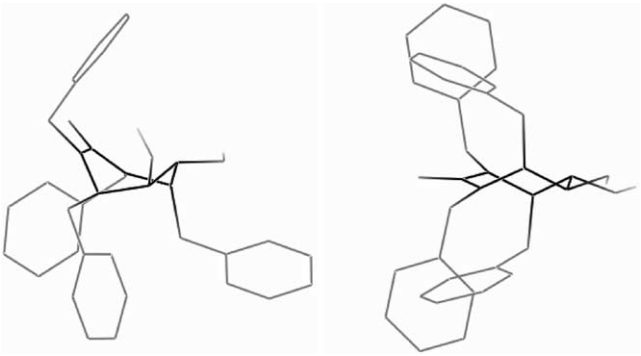


Fig. 3. AMBER*/GBSA(water) low energy structures of (**1**). The global minimum structure is shown on the left in the boat configuration and the second lowest energy structure is shown on the right in the chair conformation.

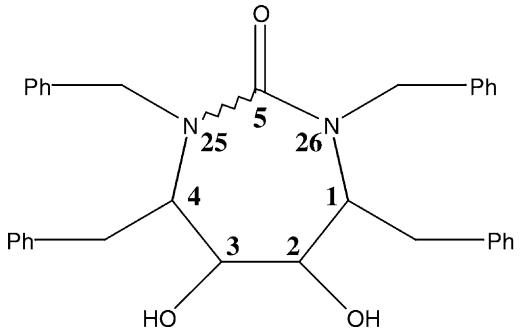


Fig. 4. Numbering system used for clustering ensembles of (**1**).

from existing minima on the PES. To investigate this matter further, the five ensembles of **1** were subjected to a detailed structural analysis to determine the differences and similarities in their structures.

*3.1.1. Comparison of the conformational ensembles of **1***

The lowest energy structure in each of the five ensembles is the same (RMSD = 0.0 Å) and exists in the boat form (Fig. 3). The second structure in each ensemble is also the same from ensemble to ensemble (RMSD = 0.0 Å) and is in the chair conformation. This conformation is on average 0.383 kJ/mol higher in energy than the lowest energy structure. The dimensionality of this system is too high to allow a one-to-one visual comparison of each structure in each ensemble. However, a comparison of the energetic ordering of the structures in each ensemble is shown in Table 5 and indicates that (a) the ensembles contain similar numbers of structures in fixed energetic windows above the lowest energy structure, (b) the number of low lying structures and the complexity of the PES increases as the energy increases, and (c) the ensembles are quite similar, at least by this measure.

The XCluster program was used to determine if the ensembles naturally form structurally related groupings and if the groupings are the same or different for each ensemble. Clustering by atomic RMS after rigid body superposition of all heavy atoms did not lead to strong clustering in any of the ensembles as evidenced by distance maps, mosaics, and separation ratios. Clustering the ensembles with respect to ring conformation, using a superposition of ring atoms 1–2–3–4–5–25–26 (Fig. 4), resulted in very strong clustering. The generic ordered distance maps for each ensemble are shown in Fig. 5. Generic ordered maps contain structures that have been reordered so that all conformations belonging to the same cluster are grouped together. These maps indicate that each ensemble clearly clusters into major groups as evidenced by the regions of darkest block diagonal patterns in the distance map (see arrows). The clustering behavior of the LM and LM:MC 75:25 ensembles is very similar. Each ensemble clearly contains a large cluster containing boat conformations and a smaller cluster with chair conformations. The LM:MC 75:25 chair grouping shows further clustering as evidenced by the small cluster between the large boat and chair clusters. This cluster contains 30 structures in the chair conformation that are distinct from the larger chair cluster. The conformers in this cluster contain 1- and 4-equatorial benzyl groups and 2- and 3-axial hydroxyl groups whereas the larger chair cluster contains 1- and 4- axial benzyl groups and 2- and 3- equatorial hydroxyl groups.

The LM:MC 50:50 ensemble is very similar to the LM and LM:MC 75:25 ensemble except that a small boat cluster appears as a fourth grouping between the larger and smaller chair clusters. This small boat cluster contains 26 structures that are distinct from the larger boat cluster in the orientation of the 1-, 4-, 25-, 26-benzyl "arms" of the molecule as shown in Fig. 6. These structures also appear in the LM and LM:MC 75:25 ensembles but appear as sub-clustering of the major boat cluster rather than as a separate boat sub-cluster. In the LM:MC 25:75 ensemble, the first boat cluster has decreased in size, which results in a corresponding increase in the size of the second boat cluster. Upon further examination of the individual structures it is clear that the distinction between these two clusters is simply a difference in the orientation of the benzyl groups.

The MC ensemble contains two boat and two chair clusters; however, the order of the clusters has changed relative to the order in the other ensembles. It should be noted that the XCluster ordering is not unique [45] and, therefore,

Table 5
Comparison of the number of low energy structures in each ensemble of **1** generated with the LM, MC, and LM:MC methods found within a 1–3, 5 and 10 kcal/mol energy window of the global minimum structure

Method	Number of conformations found with X kcal/mol of the global minimum				
	$X = 1$	$X = 2$	$X = 3$	$X = 5$	$X = 10$
LM	33	95	222	528	663
LM:MC 75:25	31	95	226	547	690
LM:MC 50:50	34	96	226	549	695
LM:MC 25:75	33	95	226	550	702
MC	31	93	220	549	701

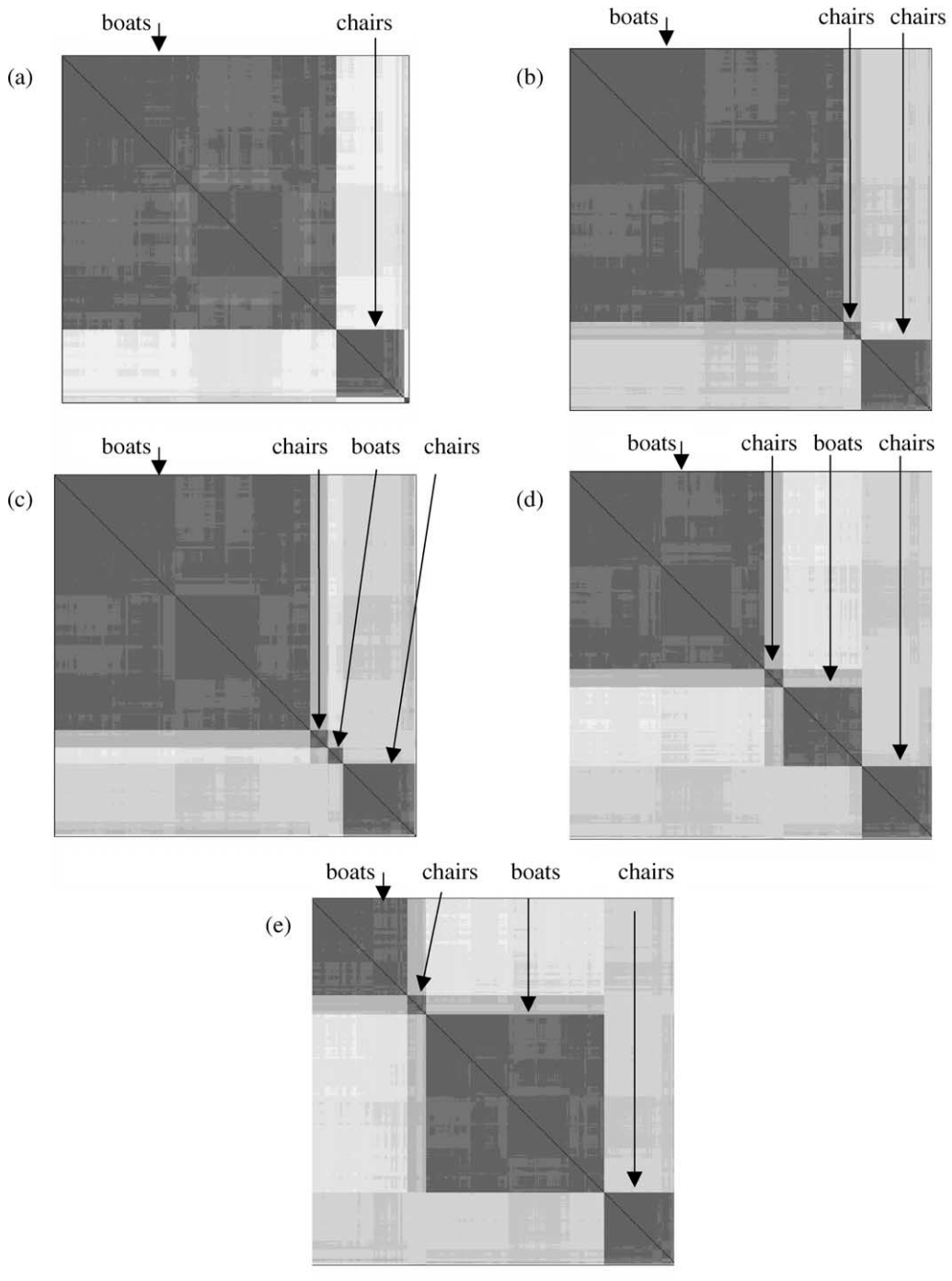


Fig. 5. Generically ordered distance maps from the conformational searches of **1**: (a) LM ensemble, (b) 75:25 LM:MC ensemble, (c) 50:50 LM:MC ensemble, (d) 25:75 LM:MC ensemble, and (e) MC ensemble.

the clustering behavior of the MC ensemble is very similar to the clustering behavior of the other ensembles. Table 6 quantifies the clustering behavior and demonstrates that the clusters are similarly sized and the leading member of most of the clusters has the same conformation from ensemble to ensemble. The average percentage of chair conformations in all the ensembles is 23.6 with ±1.5% standard deviation between ensembles. The clustering mosaics (Fig. 7) show that the ensemble clusters formed in relatively similar fashion by agglomeration of other similarly sized lower ordered clusters. This information, together with the distance maps, provide further evidence that the ensembles generated with each method contain a similar distribution of structures, however, the LM and LM:MC 75:25 methods appear to have missed a conformational class identified with the other methods. This reflects the local exploration strengths

Table 6
Clustering statistics for **1** based upon atomic RMS superposition of atoms 5–26–1–2–3–4–25

Method	Leading member of cluster	Cluster size (no. of structures)	Conformation	Leading member of cluster	Cluster size (no. of structures)	Conformation	Leading member of cluster	Cluster size (no. of structures)	Conformation	Leading member of cluster	Cluster size (no. of structures)	Conformation
LM	1	463	Boat	2	123	Chair						
LM:MC 75:25	1	463	Boat	2	118	Chair	16	30	Chair			
LM:MC 50:50	1	432	Boat	2	123	Chair	16	30	Chair	17	29	Boat
LM:MC 25:75	1	334	Boat	2	120	Chair	16	32	Chair	8	136	Boat
MC	1	165	Boat	2	120	Chair	16	32	Chair	8	304	Boat

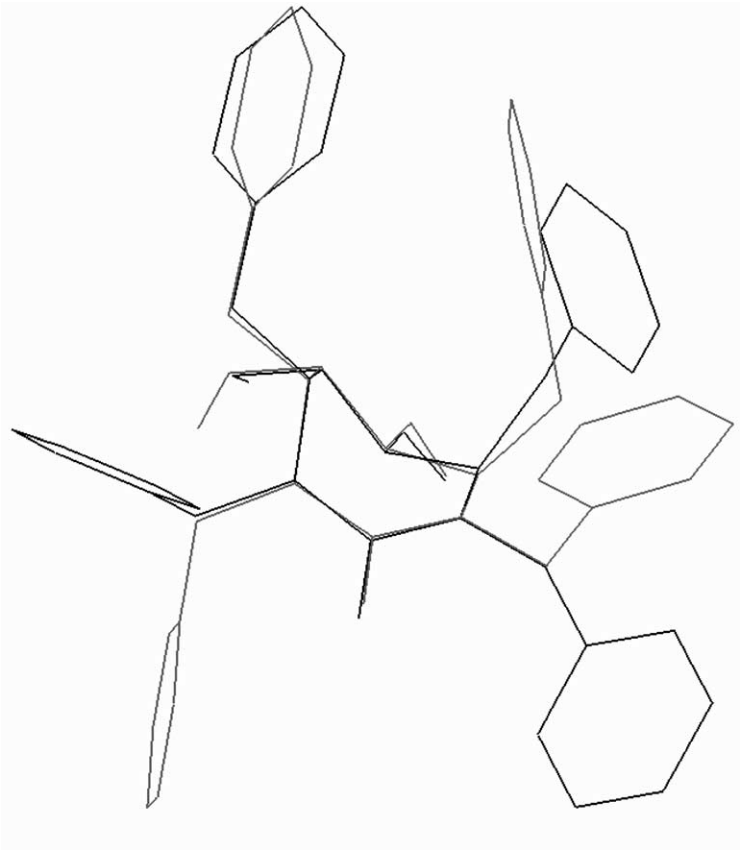


Fig. 6. Representative structures of the two boat clusters of **1** from the LM:MC 50:50 ensemble. The structure in grey is a member of the larger boat cluster (Fig. 5(c)) whereas the structure in black is a member of the smaller boat cluster.

of the LM method and the global search strengths of the MC method. In this study, LM exploration occurred by taking fixed steps of 3–6 Å along the low energy eigenvectors. It is possible that a larger range of motion would enable more global searching however this was not investigated.

3.1.2. Comparison of RSSR and SRRS ensembles

One of the convergence criteria outlined in the introduction is that a conformational search method must generate the same ensemble of structures for two molecules that are symmetrically equivalent. Consequently, an AMBER*/GBSA(water) LM:MC 50:50 conformational search was carried out on the SRRS enantiomer of **1** and compared with the RSSR ensemble for the same molecule generated with the same method. The energy of the lowest structure, the number of conformations found and the frequency with which the simulation visits the lowest structure are shown in Table 7 for a 55,000-step search on the

Table 7
Conformational search results for the SRRS enantiomer of **1** for a 55,000 step search run in blocks of 5000 steps

Number of steps	Number of conformations	Number of global minimum visits	Minimum energy (kJ/mol)	CPU time (h)
5000	535	12	−81.09	6.18
5000	639	22	−81.09	8.22
5000	662	33	−81.10	8.42
5000	681	39	−81.10	8.47
5000	669	46	−81.10	8.33
5000	685	54	−81.10	8.65
5000	680	66	−81.10	8.60
5000	698	75	−81.10	8.93
5000	691	80	−81.10	8.95
5000	703	88	−81.10	8.98
5000	697	92	−81.10	8.83

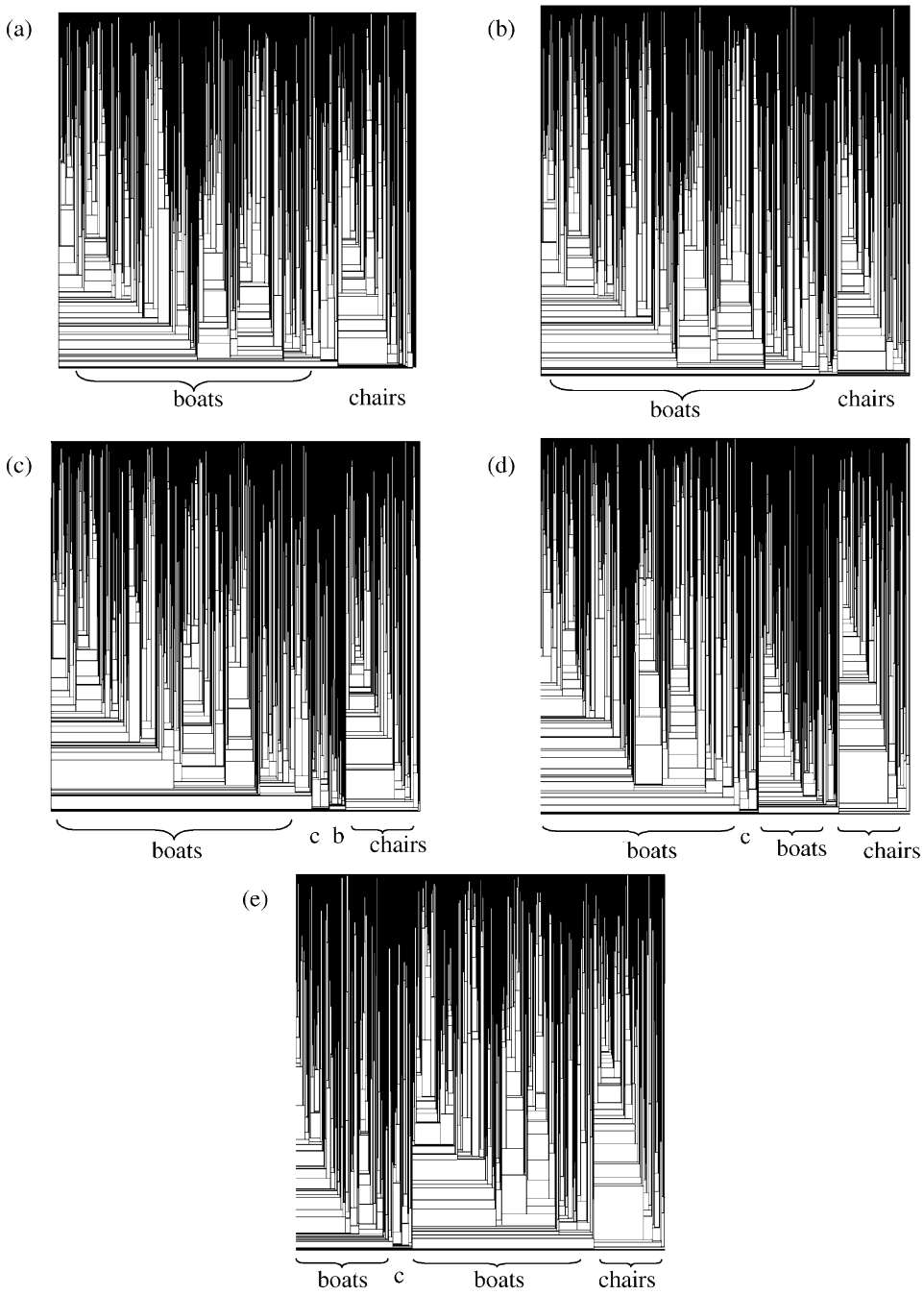


Fig. 7. Clustering mosaics from the conformational searches of **1**: (a) LM ensemble, (b) 75:25 LM:MC ensemble, (c) 50:50 LM:MC ensemble, (d) 25:75 LM:MC ensemble, and (e) MC ensemble.

SRRS enantiomer of **1**. The energy of the lowest structure is converged very early in the search and this structure is visited ~80–90 times per 5000 steps in the later stages of the study. This behavior is very similar to the LM:MC 50:50 convergence results for the RSSR enantiomer (Table 3). The SRRS LM:MC 50:50 ensemble contains 611 structures after 14,000 additional minimizations. The RSSR and SRRS global minimum energy structures are energetically equivalent, non-superimposable mirror images.

3.2. α-Methyl benzylamide (2)

The conformational flexibility of α-methyl benzylamide (**2**) was investigated using the AMBER* all-atom force field, the GB/SA model for chloroform, and three different conformational search strategies: pure LM, pure MC, and a 50:50 combination of the LM and MC algorithms. This molecule was chosen because it represents a structure whose internal d.f. are tractable enough to allow exhaustive coverage

Table 8
Comparison of LM, MC, and LM:MC search methods on **2** after 1000 search steps

Method	Number of conformations found	Number of global minimum visits	Global minimum energy (kJ/mol)	CPU time (h)	Number of conformations found per hour
LM	17	51	145.21	0.27	62.96
LM:MC	16	62	145.21	0.30	53.33
MC	18	70	145.19	0.37	48.65

of the resulting multidimensional potential energy surface. Molecule **2** is smaller than **1** (42 versus 72 atoms), contains only six torsional d.f. and does not contain any ring structures whose conformational space needs to be sampled. The analysis began by examining the results of 1000 search steps using the LM, MC, and LM:MC 50:50 methods. All searches were initiated from the same starting conformation that was 9.26 kJ/mol above the lowest energy structure found overall. Each method completes 1000 search steps in approximately the same amount of CPU time and finds the same lowest energy structure (Table 8). MC revisits the global minimum most often. LM is the most efficient method as it finds 62.96 structures per hour as opposed to 53.33 and 48.65 structures per hour for LM:MC and MC, respectively.

The ability to perform an exhaustive search of this space was probed by running each search in blocks of 1000 steps for a total of 5000 steps. Table 9 indicates that each method finds the same lowest energy structure and requires approximately the same amount of CPU time. As the search progresses, the global minimum energy structure is found increasingly more often for each method; however, MC revisits this structure most often.

To ensure convergence, each method was used to perform 25,000 additional search steps in blocks of 5000. No new structures were found during this time; and the global minimum energy structure remained the same leading us to believe that this is a converged and exhaustive conformational search of **2** (Table 10). It is difficult to distinguish between these methods since all three are very efficient owing to the small number of conformational d.f. Further minimization of the ensembles generated with each search method resulted in 13 unique structures within a 25 kJ/mol window of the global minimum. An RMSD comparison between structures in each ensemble revealed identical structures with the same energetic ordering.

We are interested in the conformational behavior of **2** because it is a simple example of a Mosher amide; i.e. a diastereomer formed by derivatizing a chiral amine with Mosher's acid chloride [28,46,47]. The differing intramolecular orientation of the resulting pair of diastereomers allows for the assignment of absolute chirality in the starting amine. The Mosher argument is based on the tendency of the CF_3–C–C=O dihedral angle to adopt a predominately *cis* conformation and the amide to adopt a *trans* conformation leading to a distinct intramolecular orientation for each diastereomer and a non-equivalence in the corresponding H^1-NMR spectra. For example, in the (*S*)–amide (**2a**) (Fig. 8), protons in the L^3 portion of the molecule are shifted upfield while those in the L^2 portion of the molecule appear further downfield. In the (*R*)-amide (**2b**), however, the

Table 9
Comparison of conformational search results for **2** using the LM, MC, and LM:MC methods for a 5000 step search run in 1000 step blocks

Number of steps	Number of conformations found	Number of global minimum visits	Global minimum energy (kJ/mol)	CPU time (h)
Method: LM				
1000	17	51	145.21	0.27
1000	20	120	145.20	0.28
1000	17	167	145.23	0.30
1000	21	216	145.20	0.28
1000	20	272	145.22	0.35
Method: LM:MC				
1000	16	62	145.21	0.30
1000	16	124	145.21	0.28
1000	19	199	145.22	0.27
1000	19	254	145.20	0.30
1000	21	325	145.23	0.35
Method: MC				
1000	18	70	145.18	0.37
1000	18	145	145.21	0.30
1000	17	212	145.23	0.28
1000	19	275	145.21	0.32
1000	18	337	145.23	0.33

Table 10
Comparison of conformational search results for **2** using the LM, MC, and LM:MC methods for a 25,000 step search run in 5000 step blocks

Number of steps	Number of conformations	Number of global minimum visits	Global minimum energy (kJ/mol)	CPU time (h)
Method: LM				
5000	17	279	145.19	1.65
5000	21	534	145.20	1.72
5000	21	872	145.24	1.90
5000	22	1349	145.23	1.75
5000	20	1756	145.25	1.63
Method: LM:MC				
5000	21	369	145.19	1.48
5000	23	692	145.22	1.83
5000	23	1053	145.19	1.68
5000	21	1459	145.21	1.75
5000	24	1821	145.25	1.78
Method: MC				
5000	21	400	145.20	1.55
5000	23	790	145.20	1.83
5000	22	1168	145.24	1.67
5000	21	1538	145.23	1.72
5000	21	1978	145.25	1.87

opposite behavior is observed; e.g. the protons in the L^2 region are shifted upfield while those in the L^3 region appear downfield.

Mosher's model works quite well [48–54] and we wanted to investigate how the flexibility and dynamical nature of the diastereomers affects this model. The RS diastereomer of **2** included in this study corresponds to structure **2a** with a phenyl and a methyl group in the L^3 and L^2 positions, respectively. The corresponding RR diastereomer corresponds to structure **2b** with the phenyl and methyl groups also in the L^3 and L^2 positions, respectively. Conformational flexibility will affect the distance between the phenyl ring on the carboxy end of the molecule and the methyl protons on the amino end of the molecule as well as the "stiffness" of the CF_3–C–C=O dihedral angle. In this report, the molecular flexibility of the RS diastereomer of **2** only is reported. A more detailed study of the conformational flexibility of Mosher amides is in progress.

A detailed structural analysis of the ensembles of **2** generated after 25,000 search steps is shown in Table 11. LM, MC, and LM:MC all generate the same structures with minor

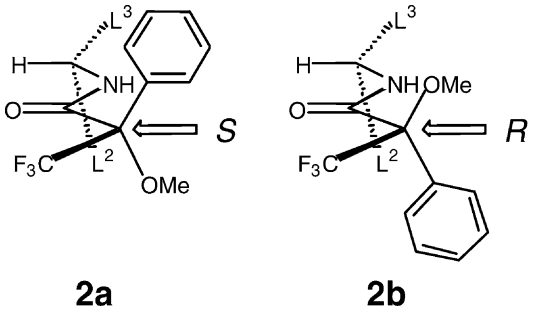


Fig. 8. Mosher amide diastereomers.

structural variances that provide further evidence of an exhaustive search. The global minimum energy structure is in agreement with the Mosher "static" model. Note that the phenyl rings are on the same side of the backbone formed from the amide and CF_3–C–C=O dihedral angles. However, it is clear from an examination of the other structures in the ensemble that the C(=O)–N–C–C(Ph) rotational d.f. is being well sampled and generates structures with the phenyl rings on the same and opposite sides of the molecular backbone. (The AMBER*/GBSA ($CHCl_3$) energetic barrier to rotation for this bond is 21 kJ/mol.) The ensemble structures of **2** can be grouped into two families with respect to phenyl orientation; an orientation with both phenyl rings on the same side of the plane as in **2a** and the other with the phenyl rings on opposite sides of the plane as in **2b**. Fig. 9 shows the first and fifth structure in the energetically ordered ensemble as representatives of the two possible orientations. Of the 13 low energy structures found in this conformational search, 9 structures contained phenyl groups on the same side of the molecule in agreement with Mosher's model and four contained phenyls on opposite sides of the molecule. If the entropic contribution of each of these states is be assumed to be similar, than the larger number of enthalpic states of **2** with the phenyls on the same side of the molecule provides eutropic evidence in support of Mosher's model.

The CF_3–C–C=O angle is between +30 and −30 for 4 of the 14 structures and three of these structures are among the four lowest energy conformations. Interestingly, an examination of the ensemble structures indicates that the CF_3–C–C=O angle is well sampled. Furthermore, it is not necessary for this dihedral angle to be close to zero in order to orient the phenyl groups on the opposite or the same side of the molecular plane; i.e. the amide *trans* conformation is enough to provide a molecular plane. (The AMBER*/GBSA

Table 11
A structural comparison of the ensembles of **2** generated with the LM, MC, and LM:MC conformational search methods

Structure number	CF_3–C–C=O dihedral angle			Angle between the phenyl groups[a]			Number of times the structure was visited			Global minimum energy[b] (kJ/mol)	Phenyl orientation[c]
	LM	LM:MC	MC	LM	LM:MC	MC	LM	LM:MC	MC		
1	24.7	24.7	24.7	−31.7	−31.7	−31.7	272	325	337	145.25	S
2	71.8	72.4	72.4	18.2	16.8	16.8	833	736	843	149.73	S
3	28.9	28.9	28.9	47.9	47.9	47.8	106	152	212	150.72	S
4	28.0	28.0	28.0	−168.3	−168.3	−168.3	22	140	248	154.41	O
5	70.6	69.2	70.6	15.2	12.2	15.3	201	244	296	155.32	S
6	71.8	71.8	71.8	−126.7	−126.7	−126.7	97	233	345	157.18	O
7	69.7	69.7	69.7	85.1	85.1	85.1	178	154	140	158.17	S
8	109.7	109.7	109.7	59.0	59.0	59.0	72	103	117	159.84	S
9	109.1	109.1	109.1	129.4	129.3	129.3	52	93	150	162.81	S
10	48.9	49.6	49.7	−2.4	−1.3	0.8	274	228	173	163.15	S
11	69.7	69.7	69.7	−128.8	−128.8	−128.8	112	134	198	163.83	O
12	−14.9	−14.8	−14.9	−58.6	−58.6	−58.6	17	102	136	164.81	S
13	107.3	107.3	107.4	−87.5	−87.5	−87.5	68	113	152	167.92	O

[a] The angle is defined in Fig. 10.
[b] The LM, MC, and LM:MC energies for each structure are equivalent.
[c] Conformations containing phenyl rings oriented on the same side of the molecular backbone are indicated with an S and oriented on opposite side of the backbone with an O.

($CHCl_3$) energetic barrier to rotation about CF_3–C–C=O is 48 kJ/mol.) This is a purely enthalpic analysis, and current work is underway to investigate entropic effects in these molecules.

While all the search methods generate the same ensemble for **2**, an analysis of the number of times that each structure is sampled (Table 11 and Fig. 10) indicates that each method follows quite different search paths. This is a natural consequence of the random nature of the methods. For instance, structure number 4 in the ensembles was visited quite often by MC, but infrequently by LM. On the other hand, structure number 10 was visited relatively often by LM, but not as often by MC. All the methods visited structure number 2 most often. Structure number 2 is 4.48 kJ/mol higher in energy than the global minimum.

3.3. A 38-membered cyclic polyazamacrolide (*3*)

The conformational behavior of **3** is interesting because this macrocyclic compound has been recently reported as a constituent component in a remarkably complex defense secretion isolated from the ladybird beetle *Epilachna borealis* [55]. The secretion contains a combinatorial library of differently sized polyazamacrolides built up from three different

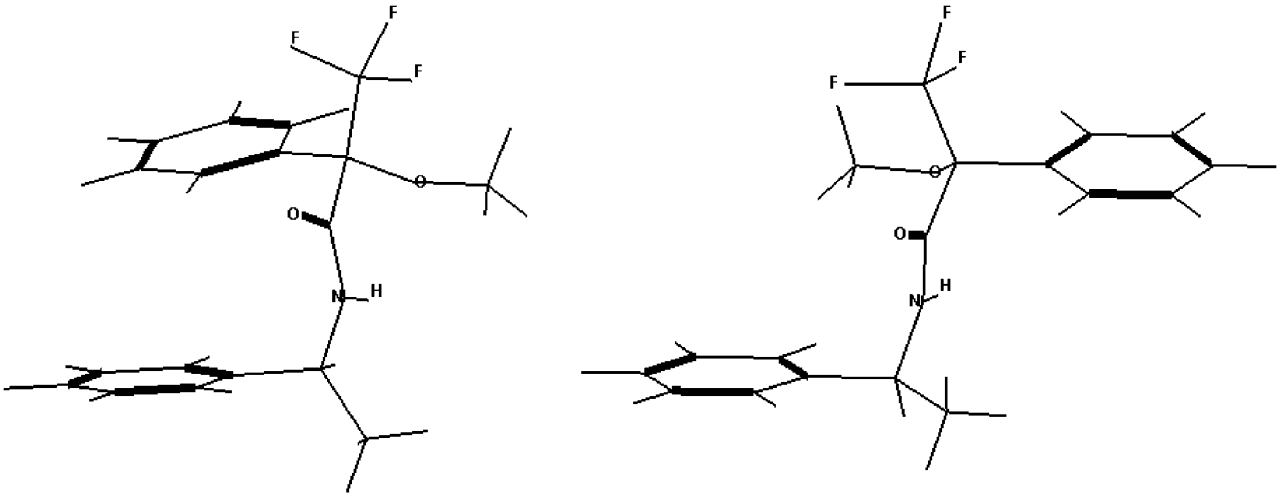


Fig. 9. Two conformations of **2**. The structure on the left is the global minimum energy structure ($E = 145.25$ kJ/mol). Note that the methoxy protons are oriented away from the shielding effect of the phenyl ring on the amino end of the molecule. The structure on the right is the fifth structure in the ensemble ($E = 154.41$ kJ/mol). The methoxy protons in this structure are oriented 3.5 Å away from the phenyl ring, likely near enough to experience a through-space shielding effect.

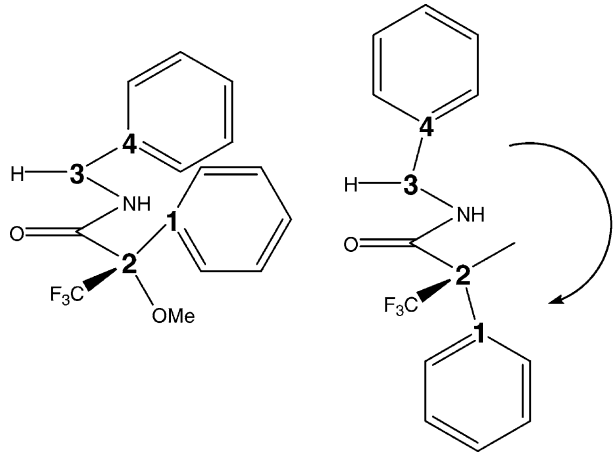


Fig. 10. Relative orientation of phenyl rings of **2** as defined by the torsional angle between carbon atoms 1–2–3–4.

saturated (ω-1)-hydroxyethyl-amino acids. The structural diversity of the arthropod exudate is even further enhanced because the polyazamacrolides have been shown to undergo multiple intramolecular O–N acyl migrations [56–58]. How the conformational flexibility and dynamic nature of this system affects the ability to initiate an O–N acyl migration via a five-membered ring intermediate was investigated.

Molecule **3** is much larger than **1** and **2** and contains a very large macrocyclic ring that likely interconverts among low energy conformational states through concerted motion of multiple ring torsions. Because of its size, the calculations were computationally more demanding than those for the smaller systems. Hence, only pure LM, pure MC, and a 50:50 combination of the two methods were compared. Molecule **3** was included in our comparison since it represents a challenge to any modern conformational search method due to the large number of d.f. and the cyclic nature of the molecule; i.e. **3** contains 114 atoms and 34 torsional d.f.

The conformational flexibility of **3** was investigated using the AMBER* all-atom force field and the GB/SA model for chloroform. The analysis began by examining the results of 5000 search steps using LM, MC, and LM:MC 50:50. All searches were initiated from the same conformation that was chosen at random and happened to be 154.88 kJ/mol above the lowest energy structure found overall. The MC and LM:MC 50:50 methods completed the 5000 search steps in approximately the same amount of CPU time (Table 12); however, the LM method takes approximately 60% longer to finish the same number of search steps. The LM method finds significantly more low energy conformations and is the most efficient since it finds 29 conformations per hour relative to 20 and 2 conformations per hour for the LM:MC 50:50 and MC methods, respectively. The LM:MC 50:50 method finds the lowest energy structure.

In an attempt to search exhaustively, the space of **3**, each search was continued for 45,000 steps in blocks of 5000 steps. Table 13 summarizes the results of each method and indicates significant differences. The MC method, for example, ran much faster than either LM or LM:MC 50:50; however, it found significantly fewer structures and failed to find many low energy structures that were identified by the other methods. The LM method found many more structures than the MC method but failed to find some of the lowest energy structures found by the LM:MC 50:50 method. Table 14 lists the number of low energy conformations that each method found within particular energetic windows above the lowest energy structure. A graphical comparison of these data (Fig. 11) illustrates that MC missed more than 5000 structures that were found within 16.736 kJ/mol (4 kcal/mol) of the lowest energy structure with the LM:MC 50:50 method. As indicated in Table 13, it is very unlikely that any of these methods are exhaustively searching the potential energy surface of **3** although the LM and LM:MC 50:50 methods obtain better surface coverage than the MC method.

In an effort to present a complete analysis, the MC search was continued in blocks of 5000 steps for the same amount of CPU time as was necessary to complete 45,000 steps using the LM:MC 50:50 method; e.g. approximately 635 CPU time (h). Table 15 shows the results from 15,000 additional MC steps. The lowest energy structure is still 11.02 kJ/mol higher than the lowest energy structure found with the LM:MC 50:50 method and only 79 new structures were found.

The lowest energy structure found with each method is shown in Fig. 12. This is the first molecular system in this study in which the different conformational search methods fail to find the same lowest energy structure. All three structures contain one intramolecular hydrogen bond between an ester oxygen on one side and an amine hydrogen on the opposite side of the ring. Each lowest energy structure contains ester groups in the *s-trans* orientation and all methyl groups are oriented external to the ring. The LM structure can be described as L-shaped whereas the LM:MC 50:50 structure is spiral or figure-eight-shaped and the MC structure is C-shaped. While these conformations are clearly different, it

Table 12
Comparison of the LM, MC, and LM:MC search results on 3 after 5000 search steps

Method	Number of conformations found	Number of global minimum visits	Minimum energy (kJ/mol)	CPU time (h)	Number of conformations found per hour
LM	1787	2	−182.54	61.37	29.12
LM:MC 50:50	804	1	−185.59	38.73	20.76
MC	103	1	−174.08	36.98	2.79

Table 13
Comparison of the conformational search results of **3** using the LM, MC, and LM:MC methods for a 45000 step search run in blocks of 5000 steps

Number of steps	Number of conformations found	Number of global minimum visits	Global minimum energy (kJ/mol)	CPU time (h)
LM				
5000	1787	2	−182.54	61.37
5000	3395	2	−182.54	77.70
5000	4808	4	−182.57	94.88
5000	5864	2	−183.93	107.10
5000	7056	2	−183.92	113.82
5000	8250	2	−183.91	121.68
5000	8045	1	−185.90	132.08
5000	9266	1	−185.90	136.80
5000	9998	1	−185.90	153.20
LM:MC 50:50				
5000	804	1	−185.59	38.73
5000	1606	1	−187.28	39.25
5000	2527	1	−187.27	48.00
5000	3314	1	−187.91	46.32
5000	3840	1	−189.13	50.27
5000	4592	1	−189.15	54.65
5000	5364	1	−189.77	58.77
5000	6113	1	−189.78	64.85
5000	6888	1	−189.78	66.35
MC				
5000	103	1	−174.08	36.98
5000	262	1	−174.39	36.15
5000	408	1	−173.61	47.01
5000	550	1	−174.35	37.23
5000	596	1	−175.75	35.50
5000	750	1	−175.69	37.77
5000	888	1	−175.77	36.13
5000	1043	1	−175.68	37.12
5000	1038	1	−176.83	39.12

appears that the driving force for each structure is the formation of an intramolecular transannular hydrogen bond and a preference for the macrolides to adopt an orientation that allows a close (less than 2.5 Å) interaction between each ester oxygen and the hydrogen atom on a neighboring amine.

A comparison of the ensembles generated with each method indicates that, while the LM method takes almost 50% more CPU time (h), a lower energy structure and many more unique conformations are found relative to the MC method. MC completes 5000 steps much faster than either LM or LM:MC 50:50 but a closer examination of the results indicates that the MC method generates many random, high energy structures that are discarded before minimization. The pure MC search is just as likely to move to a region of high energy, which may be discarded by energetic criteria, as it is to move to a lower energy region that will lead to a new structure. As the minimization step is the most time consuming the MC method runs quickly because it does not find many low energy structures. This leads to fast timings for MC but very few structures in the final ensemble. On the other hand, LM generates a significant number of structures and finds a lower energy than MC. This is in agreement with the results of Kolossvary and Guida [19,20] that the LM method is particularly well suited for larger cyclic systems where interconversion of conformers occurs via concerted, ring torsional rotations. Also in a large system such as this, the LM method is searching a smaller space than the MC method; i.e. LM is searching the conformational sub-space of low mode directions whereas the MC method is randomly searching the full 34-dimensional torsion space. As expected, LM performs better than MC, but not quite as well as the hybrid method, LM:MC 50:50. LM finds many more unique structures than MC, but does not find the lowest energy structures identified with the

Table 14
Comparison of the low energy structures of **3** found with the LM, LM:MC 50:50, and MC conformational search methods

Energy window	Number of structures found		
	LM	LM:MC 50:50	MC
−189.80 to −185.62	2	44	0
−185.62 to −181.44	30	240	0
−181.44 to −177.25	276	760	0
−177.25 to −173.07	1376	3982	7

This table lists the number of structures found within specific energetic windows above the lowest energy structure ($E = -189.80$ kJ/mol) found by the LM:MC 50:50 method.

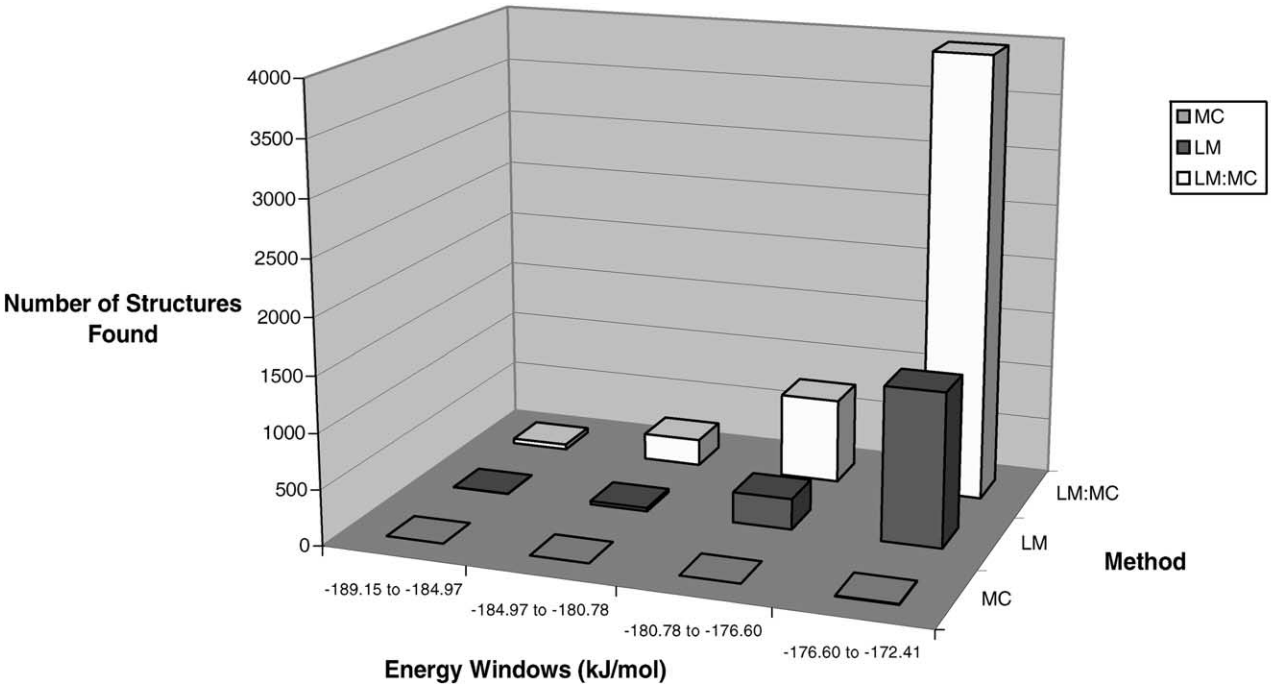


Fig. 11. A comparison of the number of structures of **3** found within 16.736 kJ/mol of the lowest energy structure for the LM, LM:MC 50:50, and MC conformational search methods.

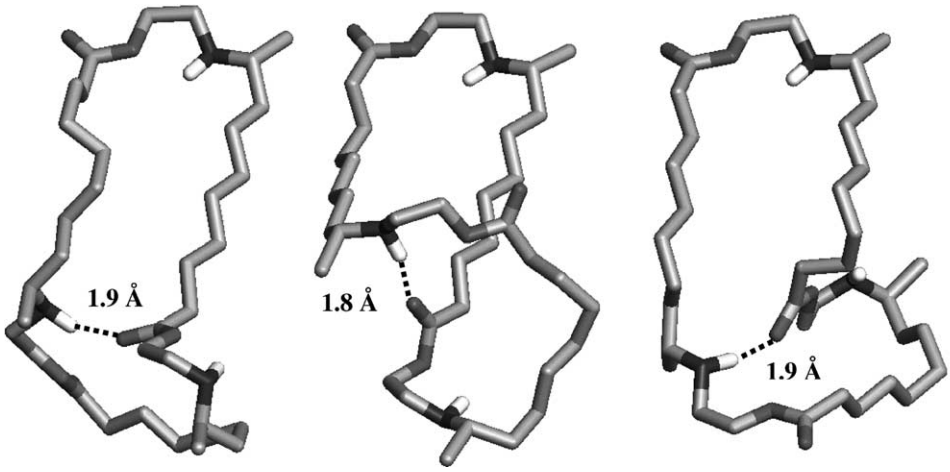


Fig. 12. The lowest energy structures of **3** from the LM (a), LM:MC 50:50 (b), and MC (c) methods.

Table 15
The results of 15,000 additional MC conformational search steps on **3**

Number of steps	Number of conformations found	Number of new conformations found	Number of global minimum visits	Global minimum energy (kJ/mol)	CPU time (h)	Number of conformations found per hour
5000	909	0	1	−178.77	41.33	21.99
5000	1011	0	1	−178.75	39.05	25.89
5000	1117	79	1	−178.76	37.53	29.76

Table 16
A structural comparison of the ensembles of **3** generated with the LM, LM:MC, and MC conformational search methods

Filter criteria	Structures matching filter criteria (%)		
	LM	LM:MC 50:50	MC
1,5-Amine H–ester O interaction[a]	98.6	98.6	95.4
Transannular H-bond[b]	7.3	69.4	17.9
Amine–carbonyl "near attack"[c]	98.2	96.4	95.4

[a] An interaction was defined when the distance between the ester oxygen and the amine hydrogen was less than or equal to 3 Å.
[b] An intramolecular transannular hydrogen bond was defined as one or more carbonyl oxygens within 3 Å of one or more amine hydrogens.
[c] An amine was defined as being oriented for attack on a carbonyl if the distance was less than or equal to 4 Å.

LM:MC 50:50 method. LM:MC 50:50 consumed 6% less CPU time than LM and 38% more than MC.

The ensembles of **3** that are generated with each method are significantly different as evidenced by the clustered distance maps. The clustering patterns are very different; the LM ensemble shows very little clustering in the generically ordered map with small clusters forming in the middle of the map. The LM:MC 50:50 ensemble shows very weak clustering into three rather large on-diagonal and two off-diagonal groupings. The MC ensemble shows the least amount of clustering with very weak evidence of a few very small on-diagonal clusters.

A detailed analysis of the conformational ensembles using the filtering mechanism available in MacroModel was also used (Table 16). While the distance maps clearly show that the ensembles are different, the filtering results indicate that there are some structural similarities. For instance, more than 95% of all the low energy structures found in any ensemble contains a 1,5-intramolecular amine hydrogen-to-ester oxygen interaction distance of less than 3 Å (Fig. 13). This strongly electrostatic 1,5 orientation positions the amine nitrogen within 4 Å of a neighboring ester carbonyl group in more than 98% of the structures that were found. It is concluded that this electrostatically driven interaction is responsible for the preponderance of "near-attack" N–C (O) conformations [59,60] seen in the ensembles of **3** generated in this study and by others [57]. More than 60% of the low energy structures are in a folded state where the intramolecular cavity has collapsed on itself. In many instances the folding seems to be caused by a transannular hydrogen bond; however, folded structures that do not contain such a bond also exist. Approximately 69% of the structures in the LM:MC ensemble contains at least one transannular hydrogen bond while the LM and MC ensembles contain only 7 and 17% transannular hydrogen-bonded structures, respectively. A closer examination of the hydrogen-bonding behavior provides conclusive evidence that none of the ensembles generated represent converged results. Table 17 provides detailed information regarding transannular hydrogen

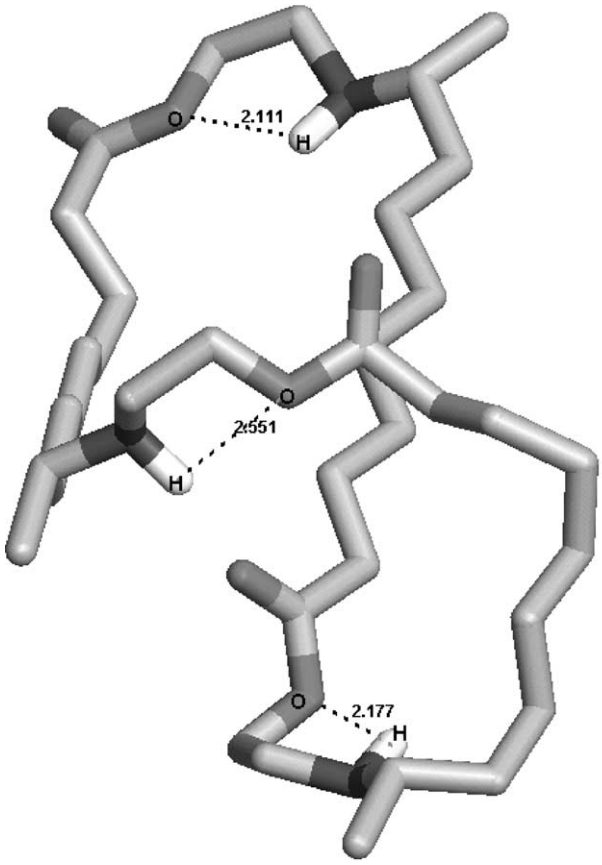


Fig. 13. Lowest energy conformation of **3** showing 1,5 amine hydrogen–ester oxygen interaction. More than 95% of all structures found contains a 1,5-amine hydrogen–ester oxygen interaction of less than 3 Å.

bonding. Since **3** has C_3 symmetry, symmetry arguments dictate that a converged ensemble should contain the same percentage of hydrogen bonding interactions between atoms O13–H43 as between O45–H27 or O29–H11 (Fig. 14). This is not the case for any of the ensembles generated. The MC percentages are the most evenly distributed and this likely reflects the barrier crossing ability of the method. However, the LM:MC 50:50 method generates an ensemble that appears to be "stuck" in the vicinity of the PES that corresponds to a transannular hydrogen bond between atoms O45

Table 17
A transannular hydrogen bonding comparison of the ensembles of **3** generated with the LM, LM:MC, and MC conformational search methods: the percentage of structures that contain a C=O···H–N bond less than 3 Å

O-atom number	H-atom number	Structures matching H-bond criteria (%)		
		LM	LM:MC 50:50	MC
13	43	0.8	0.0	3.2
13	11	0.0	0.3	2.0
29	11	5.4	0.4	0.7
29	27	0.0	0.0	1.4
45	27	0.6	68.7	9.7
45	43	0.5	0.0	0.9

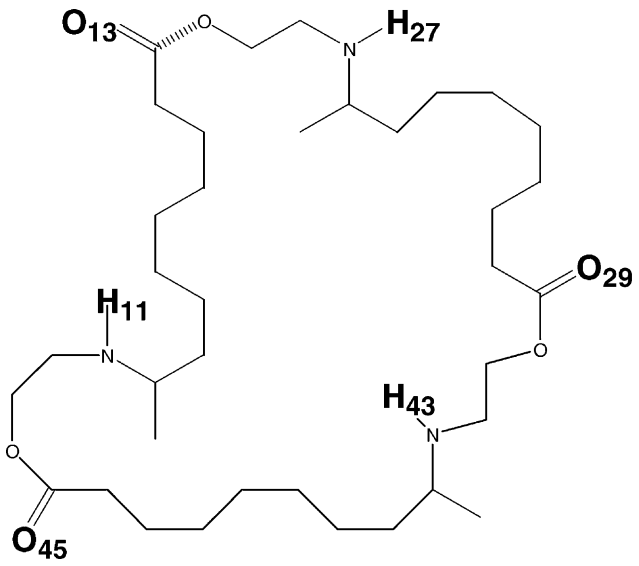


Fig. 14. Atom numbering system for **3** used to investigate hydrogen-bonding behavior.

and H27, and rarely samples the other symmetrically equivalent conformations. The size and complexity of the conformational space of **3** necessitates additional conformational searching. These calculations are underway and will be reported in a separate article focusing on the flexibility of this system and its derivatives.

4. Conclusions

The conformational flexibility of three differently sized, rather diverse molecular systems has been investigated using the pure LM, pure MC, and hybrid LM:MC conformational search methods. The smallest system, **2**, contains six rotatable bonds. All three methods are able to search exhaustively and efficiently the conformational space of **2**, generating the same ensemble containing 13 unique structures. The medium-sized molecule, **1**, is a cyclic system containing 14 variable torsions. All three methods find the same global minimum structure and generate very similar ensembles, which suggests that each method is quite capable of exhaustively searching the conformational space of **1**. Of the three methods, LM is the most efficient for searching the space of **1** as it generates the largest number of conformations per unit time; however, MC generates a converged ensemble more rapidly. This is likely due to the random nature of pure MC, which enables good surface coverage in a reasonable amount of CPU time. The largest system, **3**, is a macrocycle containing 34 torsional d.f. This molecule represents a challenge to any modern conformational search method and is the only system that is not exhaustively searched by any of the methods. MC was least efficient for searching the space of **3** due to the random nature of the method. The LM and LM:MC 50:50 methods for **3** obtain much better surface coverage of the space than MC. LM finds the largest number of unique structures; however, it fails to find some of the low energy structures identified with LM:MC 50:50. The optimum search method for **3** appears to be LM:MC 50:50 which appears to dovetail the local exploration strengths of LM with the random surface "hopping" capability of MC.

Acknowledgements

Acknowledgement is made to the Donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research. This work was also supported, in part, by a grant from the Merck Foundation and the American Association for the Advancement of Science. Thanks are also due to Prof. Wayne Guida (Eckerd College), Dr. Peter Shenkin (Schrodinger Inc.) and Marc Zimmer (Connecticut College) for valuable comments on this manuscript. We are also grateful for support from Hobart and William Smith Colleges.

References

[1] A.R. Leach, Molecular Modeling Principles and Applications, 2nd Edition, Prentice Hall, Pearson Education, Harlow, England, 2001.
[2] H. Senderowitz, C. Parish, W.C. Still, Carbohydrates: united-atom AMBER* parameterization of pyranoses and simulations yielding anomeric free energies, J. Am. Chem. Soc. 118 (1996) 2078–2086.
[3] M. Saunders, K.N. Houk, Y.-D. Wu, W.C. Still, M. Lipton, G. Chang, W.C. Guida, Conformations of cycloheptadecane: a comparison of methods for conformational searching, J. Am. Chem. Soc. 112 (1990) 1419–1427.
[4] G. Keseru, I. Kolossvary, Molecular Mechanics and Conformational Analysis in Drug Design, Blackwell Scientific, Oxford, 1999.
[5] S.R. Wilson, W. Cui, J.W. Moskowitz, K.E. Schmidt, Conformational analysis of flexible molecules: location of the global minimum energy conformation by the simulated annealing method, Tetrahedron Lett. 29 (1988) 4373.
[6] S.R. Wilson, W. Cui, Applications of simulated annealing to peptides, Biopolymers 29 (1990) 225.
[7] S.R. Wilson, W. Cui, J.W. Moskowitz, Conformational analysis of flexible molecules: location of the global minimum energy conformation by the simulated annealing method, J. Comput. Chem. 12 (1991) 342.
[8] P.K. Weiner, S. Profeta, G. Wipff, T. Havel, I.D. Kuntz, R. Langridge, P.A. Kollman, A distance geometry study of ring systems: application to cyclooctane, 18-crown-6, cyclododecane and androstanedione, Tetrahedron 38 (1983) 1113.
[9] G.M. Crippen, T.F. Havel, Distance Geometry and Molecular Conformation, Wiley, New York, 1988.
[10] T.F. Havel, An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance, Prog. Biophys. Mol. Biol. 56 (1991) 43.
[11] C.E. Peishoff, J.S. Dixon, Improvements to the distance geometry algorithm for conformational sampling of cyclic structures, J. Comput. Chem. 13 (1992) 565.
[12] G.M. Crippen, Exploring the conformation space of cycloalkanes by linearized embedding, J. Comput. Chem. 13 (1992) 351.
[13] R.M. Fine, H. Wang, P.S. Shenkin, D.L. Yarmush, C. Levinthal, Predicting antibody hypervariable loop conformations: minimization and molecular dynamics studies of mcpc603 from many randomly

generated loop conformations, Proteins: Struct., Funct., Genet. 1 (1986) 342.
[14] P.S. Shenkin, D.L. Yarmush, R.M. Fine, H. Wang, C. Levinthal, Predicting antibody hypervariable loop conformation: ensembles of random conformations for ring-like structures, Biopolymers 26 (1987) 2053.
[15] G. Chang, W.C. Guida, W.C. Still, An internal coordinate Monte Carlo method for searching conformational space, J. Am. Chem. Soc. 111 (1989) 4379–4386.
[16] N. Weinberg, S. Wolfe, A comprehensive approach to the conformational analysis of cyclic compounds, J. Am. Chem. Soc. 116 (1994) 9860–9868.
[17] Z. Li, H.A. Scheraga, Monte Carlo-minimization approach to the multiple-minima problem in protein folding*, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 6611–6615.
[18] B.V. Freyberg, W. Braun, Efficient search for all low energy conformations of met-enkephalin by Monte Carlo methods, J. Comput. Chem. 12 (1991) 1065.
[19] I. Kolossvary, W.C. Guida, Low mode search—an efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides, J. Am. Chem. Soc. 118 (1996) 5011–5019.
[20] I. Kolossvary, W.C. Guida, Low mode confomational search elucidated: application to c39h80 and flexible docking of 9-deazaguanine inhibitors into pnp, J. Comput. Chem. 20 (1999) 1671–1684.
[21] I. Kolossvary, G. Keseru, Hessian-free low mode conformational search for large-scale protein loop optimization: application to c-jun n-terminal kinase jnk3, J. Comput. Chem. 22 (2001) 21–30.
[22] A.R. Leach, A survey of methods for searching the conformational space of small and medium-size molecules, in: K.B.L.A.D.B. Boyd (Ed.), Reviews of Computational Chemistry, Vol. 2, Wiley, New York, 1991, pp. 1–55.
[23] O. Becker, Computational Biochemistry and Biophysics, Marcel Deeker, New York, 2001.
[24] J.M. Goodman, Chemical Applications of Molecular Modeling, Royal Society of Chemistry, Cambridge, 1998.
[25] F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W.C. Still, MacroModel: an integrated software system for modeling organic and bioorganic molecules using molecular mechanics, J. Comput. Chem. 11 (1990) 440–467.
[26] P. Culot, G. Dive, V.H. Nguyen, J.M. Ghuysen, Vibrational analysis in Cartesian coordinates, Theor. Chim. Acta 82 (1992) 189.
[27] R.F. Kaltenbach, D.A. Nugiel, P.Y.S. Lam, R.M. Klabe, S.P. Seitz, Stereoisomers of cyclic urea HIV-1 protease inhibitors: synthesis and binding affinities, J. Med. Chem. 41 (1998) 5113–5117.
[28] J.A. Dale, H.S. Mosher, Nuclear magnetic resonance enantiomer regents. Configurational correlations via nuclear magnetic resonance chemical shifts of diastereomeric mandelate, *o*-mandelate, and α-methoxy-α-trifluoromethylphenylacetate (MTPA) esters, J. Am. Chem. Soc. 95 (1973) 512–519.
[29] F.C. Schroeder, A combinatorial library of macrocyclic polyamines produced by a ladybird beetle, J. Am. Chem. Soc. 122 (2000) 3628–3634.
[30] W.D. Cornell, R. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179.
[31] S.J. Weiner, P.A. Kollman, D. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P.K. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, J. Am. Chem. Soc. 106 (1984) 765.
[32] G. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides, J. Phys. Chem. 105 (2001) 6474–6487.
[33] W.L. Jorgensen, J. Tirado-Rives, The opls potential functions for peptides: energy minimizations for crystals of cyclic peptides and crambin, J. Am. Chem. Soc. 110 (1988) 1657.
[34] W. Hasel, T.F. Hendrickson, W.C. Still, A rapid approximation to the solvent accessible surface areas of atoms, Tetrahedron Comput. Methodol. 1 (1988) 103–116.
[35] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics, J. Am. Chem. Soc. 112 (1990) 6127–6129.
[36] D. Qiu, P.S. Shenkin, F.P. Hollinger, W.C. Still, The GB/SA continuum model for solvation: a fast analytical method for the calculation of approximate born radii, J. Phys. Chem. A 101 (1997) 3005–3014.
[37] J. Weiser, A.A. Weiser, P.S. Shenkin, W.C. Still, Neighbor-list reduction: optimization for computation of molecular van der Waals' and solvent-accessible surface areas, J. Comput. Chem. 19 (1998) 797–808.
[38] J. Weiser, P.S. Shenkin, W.C. Still, Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO), J. Comput. Chem. 20 (1999) 217–230.
[39] J. Weiser, P.S. Shenkin, W.C. Still, Fast, approximate algorithm for detection of solvent-inaccessible atoms, J. Comput. Chem. 20 (1999) 586–596.
[40] J.M. Goodman, W.C. Still, An unbounded systematic search of conformational space, J. Comput. Chem. 12 (1991) 1110–1117.
[41] W.C. Guida, R.S. Bohacek, M.D. Erion, Probing the conformational space available to inhibitors in the thermolysin active site using monte carlo energy minimization techniques, J. Comput. Chem. 13 (1992) 214.
[42] I. Kolossvary, W.C. Guida, 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. 115 (1993) 2107–2119.
[43] W.C. Still, I. Galynker, Chemical consequences of conformation in macrocyclic compounds, Tetrahedron 37 (1981) 3981–3996.
[44] J.W. Ponder, F.M. Richards, An efficient Newton-like method for molecular mechanics energy minimization of large molecules, J. Comput. Chem. 8 (1987) 1016.
[45] P.S. Shenkin, D.Q. Mcdonald, Cluster analysis of molecular conformations, J. Comput. Chem. 15 (1994) 899–916.
[46] J.A. Dale, D.L. Dull, H.S. Mosher, α-Methoxy-α-trifluoromethylphenylacetic acid, a versatile reagent for the determination of enantiomeric composition of alcohols and amines, J. Org. Chem. 34 (1969) 2543.
[47] G.R. Sullivan, J.A. Dale, H.S. Mosher, Correlation of configuration and 19f chemical shifts of α-methoxy-a-trifluoromethylphenylacetate derivatives, J. Org. Chem. 38 (1973) 2143.
[48] T.R. Hoye, M.K. Renner, MTPA (Mosher) amides of cyclic secondary amines: conformational aspects and a useful method for assignment of amine configuration, J. Org. Chem. 61 (1996) 2056–2064.
[49] I. Ohtani, T. Kusumi, Y. Kashman, H. Kakisawa, High-field FT-NMR application of Mosher's method: the absolute configurations of marine terpenoids, J. Am. Chem. Soc. 113 (1991) 4092–4096.
[50] S.K. Latypov, J.M. Seco, E. Quinoa, R. Riguera, Conformational structure and dynamics of arylmethoxyacetates: D-NMR spectroscopy and aromatic shielding effect, J. Org. Chem. 60 (1995) 504–515.
[51] S.K. Latypov, J.M. Seco, E. Quinoa, R. Riguera, Determination of the absolute stereochemistry of chiral amines by H^1-NMR of arylmethoxyacetic acid amides: the conformational model, J. Org. Chem. 60 (1995) 1538–1545.
[52] B.M. Trost, R.C. Bunt, S.R. Pulley, On the use of *o*-methylmandelic acid for the establishment of absolute configuration of a-chiral primary amines, J. Org. Chem. 59 (1994) 4202–4205.

[53] T.R. Hoye, M.K. Renner, Applications of MTPA (Mosher) amides of secondary amines: assignment of absolute configuration in chiral cyclic amines, J. Org. Chem. 61 (1996) 8489–8495.
[54] S.K. Latypov, J.M. Seco, E. Quinoa, R. Riguera, MTPA vs. mpa in the determination of the absolute configuration of chiral alcohols by H^1-NMR, J. Org. Chem. 61 (1996) 8569–8577.
[55] F.C. Schroder, J.J. Farmer, A.B. Attygalle, S.R. Smedley, T. Eisner, J. Meinwald, Combinatorial chemistry in insects: a library of defensive macrocyclic polyamines, Science 281 (1998) 428–431.
[56] F.C. Schroder, J.J. Farmer, S.R. Smedley, A.B. Attygalle, T. Eisner, J. Meinwald, A combinatorial library of macrocyclic polyamines produced by a ladybird beetle, J. Am. Chem. Soc. 122 (2000) 3628–3634.
[57] S. Garcia-Rubio, J. Meinwald, Synthesis and ring contraction reactions of polyazamacrolides, J. Org. Chem. 66 (2001) 1082–1096.
[58] M.R. Gronquist, J. Meinwald, Synthesis of unsaturated polyazamacrolides from the ladybird beetle subcoccinella vigintiquatuorpunctata, J. Org. Chem. 66 (2001) 1075–1081.
[59] F.C. Lightstone, T.C. Bruice, Ground state conformations and entropic and enthalpic factors in the efficiency of intramolecular and enzymatic reactions 1. Cyclic anhydride formation by substituted glutarates, succinate, J. Am. Chem. Soc. 188 (1996) 2595.
[60] T.C. Bruice, F.C. Lightstone, Ground state and transition state contributions to intramolecular and enzymatic reactions, Acc. Chem. Res. 32 (1999) 127.