

Conformational Analysis of Macrocycles: Finding What Common Search Methods Miss

Pascal Bonnet,^{*,†} Dimitris K. Agrafiotis,^{*,‡} Fangqiang Zhu,^{‡,¶} and Eric Martin[§]

Johnson & Johnson Pharmaceutical Research and Development, Division of Janssen Pharmaceutica N.V., Turnhoutsweg 30, 2340 Beerse, Belgium, Johnson & Johnson Pharmaceutical Research and Development, L.L.C., 665 Stockton Drive, Exton, Pennsylvania 19341, and Novartis, 4560 Horton Street, Emeryville, California 94608

Received July 6, 2009

As computational drug design becomes increasingly reliant on virtual screening and on high-throughput 3D modeling, the need for fast, robust, and reliable methods for sampling molecular conformations has become greater than ever. Furthermore, chemical novelty is at a premium, forcing medicinal chemists to explore more complex structural motifs and unusual topologies. This necessitates the use of conformational sampling techniques that work well in all cases. Here, we compare the performance of several popular conformational search algorithms on three broad classes of macrocyclic molecules. These methods include Catalyst, CAESAR, MacroModel, MOE, Omega, Rubicon and two newer self-organizing algorithms known as stochastic proximity embedding (SPE) and self-organizing superimposition (SOS) that have been developed at Johnson & Johnson. Our results show a compelling advantage for the three distance geometry methods (SOS, SPE, and Rubicon) followed to a lesser extent by MacroModel. The remaining techniques, particularly those based on systematic search, often failed to identify any of the lowest energy conformations and are unsuitable for this class of structures. Taken together with our previous study on drug-like molecules (Agrafiotis, D. K.; Gibbs, A.; Zhu, F.; Izrailev, S.; Martin, E. Conformational Sampling of Bioactive Molecules: A Comparative Study. *J. Chem. Inf. Model.*, **2007**, 47, 1067–1086), these results suggest that SPE and SOS are two of the most robust and universally applicable conformational search methods, with the latter being preferred because of its superior speed.

INTRODUCTION

Macrocycles are large, cyclic molecules widely used in medicinal, supramolecular, carbohydrate, and analytical chemistry. The discovery and the optimization of macrocycles capable of binding to biological receptors are of fundamental interest to the pharmaceutical industry.¹ Most macrocycles are polymeric in nature and are categorized based on the chemical nature of the repeating unit. The best known examples include cycloalkanes (cyclic methylene chains), cyclopeptides² (cyclic amino acid chains), cyclodextrins³ (cyclic 1→4 linked D-glucopyranoside chains, also known as cycloamyloses), crown ethers⁴ (cyclic polyethylene oxides), cryptands (cyclic polyazo-polyethers), calixarenes (cyclic chains of phenol derivatives), and the intriguingly interlocked catenanes.⁵ Most of these cyclic oligomers are involved in coordination chemistry where they chelate metal ions or in host–guest chemistry where they complex small molecules. Charles Pedersen, Donald Cram, and Jean-Marie Lehn,^{6,7} who were awarded the 1987 Nobel Prize in chemistry⁸ for their work in the field of organic synthesis of

molecules that mimic biological processes, largely studied macrocyclic polyethers and their selective complexation to alkali metal cations. There exists a large number of natural macrocycles such as heme, a porphyrin-containing molecule present in red blood cells, chlorophyll, a chlorine-containing green pigment involved in photosynthesis, and valinomycin, a selective neutral ionophore agent. Numerous biologically active cyclopeptides have been isolated from natural sources and are used along with their derivatives as antibiotic, antifungal, and antitumor clinical drugs. These include antibiotics, such as erythromycin A and rifamycin S, a potent inhibitor of DNA-dependent RNA polymerase (RNAP), immunosuppressants, such as rapamycin, which specifically inhibits mTOR signaling to its downstream targets,⁹ antifungals, such as amphotericin B,¹⁰ antiparasitics, such as avermectin, and anticancer agents, such as geldanamycin, which bind and alter the function of heat shock protein 90 (HSP90),¹¹ and maytansine, which shows some antileukemic activity.¹² Another structurally and biologically intriguing class of macrocyclic compounds is found in marine natural products, like tetramic acid-containing macrolactams. These macrocycles are used as anticancer, antibiotic, antiprotozoal, and antioxidant agents. Examples include cylindramide A,¹³ discodermide,¹⁴ and alteramide A.¹⁵

As with all biologically active molecules, the three-dimensional structure of macrocycles is critically important for molecular recognition and target binding. During the last two decades, there have been numerous conformational studies of macrocycles by molecular mechanics and molec-

* Corresponding authors. Telephone: +32-14-605961 (P.B.) and (610) 458-6045 (D.K.A.). E-mail: pbonnet@its.jnj.com (P.B.) and dagrafio@its.jnj.com (D.K.A.).

[†] Johnson & Johnson, Pharmaceutical Research & Development, L.L.C., Division of Janssen Pharmaceutica N.V., Beerse, Belgium.

[‡] Johnson & Johnson, Pharmaceutical Research & Development, L.L.C., Exton, PA.

[¶] Current address: National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health, Building 5, Room 137A, Bethesda, Maryland 20892.

[§] Novartis, Emeryville, CA.

ular dynamics. The recent availability of crystal structures of large cyclodextrins containing up to 26 glucose units¹⁶ revealed interesting physicochemical properties, molecular recognition modes, and structural features, like the double band-flip motif,¹⁷ and sparked additional studies by computational methods.¹⁸ A cyclized three-turn single helical structure comprised of 21 units was found in the crystal structure of cyclomaltooligosaccharide.¹⁹ Due to their biological interest, cyclopeptides and their derivatives have been the subject of many conformational studies^{20–22} by NMR spectroscopy,^{23,24} circular dichroism,²⁵ and molecular mechanics and dynamics calculations.^{26–28}

By their nature, macrocycles are relatively rigid molecules with higher London dispersion forces and with lower entropies compared to acyclic molecules, but larger macrocycles can still be considerably flexible.^{17,29} Despite several decades of active research, accurate prediction of free energy of binding remains a challenging problem and one of the holy grails of computer-aided drug design. Attention is increasingly shifted to empirical methods which can recover the bioactive conformations observed in cocrystal structures and can offer the best correlation of docking scores to known biological activities. Adequate sampling of conformational space is important for identifying the bioactive conformation among the myriad of available configurations as well as for obtaining reliable estimates of ΔG , which is necessary for the meaningful ranking of putative ligands. In the interest of efficiency, some scoring functions approximate the entropy term using a weighted sum of the ligand's rotatable bonds.³⁰ Obviously, such scoring functions have an inherent bias when applied to macrocyclic molecules because these systems are subject to additional conformational constraints which are not present in their acyclic analogs. Moreover, the correlation that has been observed between acceptable strain energy (defined as the energy difference between the bound and the minimized conformation of the free ligand) and the number of rotatable bonds for linear bound ligands³¹ may not hold true for macrocycles. Effective sampling of conformational space is a prerequisite for obtaining sensible results in a wide range of modeling applications, including protein docking, pharmacophore searching and flexible alignment, and this is certainly no less the case for macrocycles. Indeed, it has been shown that adequate sampling is extremely important in docking calculations, and that small changes in the search parameters can have a profound effect on performance.^{32,33}

Given its pivotal nature, it is not surprising that conformational search has been the center of attention since the inception of the computational chemistry field. A plethora of methods and software packages have been developed ranging widely in scope, speed, and sophistication,^{34–44} some of which specifically target ring systems.^{45,46} Many of these methods have been compared against each other in a number of studies, using an equally wide range of criteria to assess relative performance.^{47–55} These criteria typically include the speed, the number and diversity of unique conformations generated, and the ability to identify the lowest energy minimum and/or some other experimentally determined conformation, such as the crystal structure of the free ligand or the bound conformation to a known protein target. Since most conformational search methods usually involve a large number of tunable parameters, the vast majority of these

validation studies use best practices or default settings recommended by the original developers of these methods.


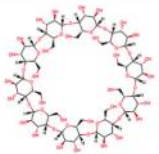





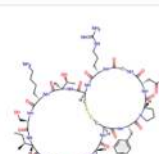

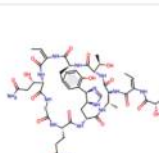
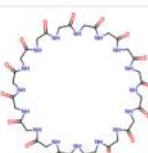
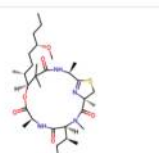
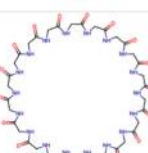
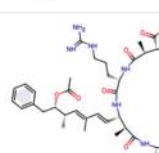
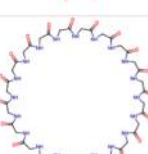
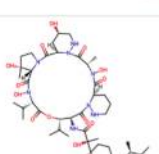
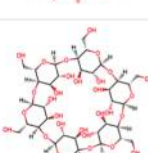

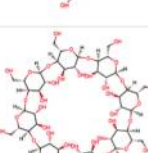
While diversity is sometimes a goal in its own right, as in many approaches to library design, thoroughness of conformational sampling is usually not an end in itself nor is it the sovereign virtue for a conformational search method. Some methods are extremely fast, and the extent of their sampling may be sufficient for some applications. However, thorough sampling is an important means to many further ends, and knowing which methods are capable of sampling the full ensemble of accessible conformations is of great value to the molecular modeling community.

One recent study aimed at comparing the stochastic proximity embedding (SPE)⁵⁶ algorithm developed by our group against several widely used conformational search algorithms implemented in commercially available molecular modeling packages.⁵² We found that SPE used in conjunction with conformational boosting, a heuristic for biasing conformational search toward more extended or more compact geometries, along with Catalyst, were significantly more effective in sampling the full range of conformational space of drug-like molecules compared to the other methods. Here we extend this type of analysis to a diverse class of macrocyclic molecules, paying particular attention to the effects of ring size. Two additional methods that became available to us since our original paper have been included, namely CAESAR⁵⁷ (conformation algorithm based on energy screening and recursive buildup) and a fast variant of SPE known as self-organizing superimposition (SOS).⁵⁸ As seen in the sequel, our results show a dramatic advantage for the three distance geometry methods (SOS, SPE, and Rubicon) followed to a lesser extent by the stochastic search algorithm implemented in MacroModel and Catalyst. The remaining techniques, particularly those based on systematic search like Omega and CAESAR, often failed to identify any low-energy conformations and are unsuitable for macrocyclic molecules (and, we believe, constrained topologies in general).

METHODS

Data Sets. The 12 conformational search algorithms were compared using three different groups of molecules: cyclopeptides, cyclodextrins, and a set of diverse naturally occurring macrocycles exhibiting biological activity (Table 1). Cyclopeptides and cyclodextrins were chosen in order to systematically study the effect of ring size on the performance of the various search methods. The cyclopeptide group consists of eight polyglycines of increasing ring size, assembled by linking 6, 8, 10, 12, 14, 16, 18, and 20 residues of glycine, respectively, into a cyclic structure (labeled G6 to G20 hereafter). Cyclodextrins, also called cycloamyloses, belong to the family of cyclic oligosaccharides. The cyclodextrin group is composed of 6, 8, 10, 12, and 14 $\alpha(1\rightarrow4)$ -linked D-glucopyranose units per ring (labeled D6 to D14 hereafter). The third group selected for this study contains a set of six naturally occurring macrocycles of different ring sizes and amino acid composition. These compounds range from monocyclic to polycyclic glycopeptides and exhibit significant biological properties. SFTI-1 (sunflower trypsin inhibitor 1) is a cyclic peptide with 14 amino acid residues and one disulfide bond.⁵⁹ Aciculitin A is an antifungal bicyclic glycopeptidolipid obtained from a marine source that

Table 1. Macrocycles Used in the Present Study

Molecule	Name	Structure	Molecule	Name	Structure
G6	Cyclo(Gly) ₆		D10	Cyclodextrin-10	
G8	Cyclo(Gly) ₈		D12	Cyclodextrin-12	
G10	Cyclo(Gly) ₁₀		D14	Cyclodextrin-14	
G12	Cyclo (Gly) ₁₂		P1	SFTI-1	
G14	Cyclo(Gly) ₁₄		P2	Aciculitin A	
G16	Cyclo(Gly) ₁₆		P3	Halipeptin A	
G18	Cyclo(Gly) ₁₈		P4	MicrocystinLR	
G20	Cyclo(Gly) ₂₀		P5	Polyoxypeptin	
D6	Cyclodextrin-6		P6	TolybyssidinA	
D8	Cyclodextrin-8				

contains an unusual histidino-tyrosine bridge.⁶⁰ Halipeptin A is a 16-membered cyclodepsipeptide with anti-inflammatory activity.⁶¹ Microcystin-LR is a cyclic heptapeptide toxin produced and released by cyanobacteria, that is made up of two natural (leucine and arginine) and five unnatural amino acids. Polyoxypeptin A, a potent inducer of apoptosis,⁶² is a 19-membered cyclic hexadepsipeptide containing an acyl side chain and a unique amino acid, (2*S*,3*R*)-3-hydroxy-3-methylproline. Finally, Tolybyssidin A is a cyclic tridecapeptide containing the unnatural amino acid, dehydro-homoalanine, that exhibits some antifungal activity.⁶³

Computational Details. Twelve representative conformational sampling techniques implemented in six different software packages were evaluated. These include CAESAR, four variants of Catalyst (best, best with poling, fast, and fast with poling), MacroModel, Omega, MOE stochastic search, Rubicon, two variants of SOS (with and without boosting), and SPE (with boosting). It is important to realize that the package names are used as shorthand for a particular conformational sampling protocol employed within that package.

As stated in our previous study, the results presented here cannot be taken as a general endorsement or an indictment of a particular package for several reasons. First, several of these packages include other conformational sampling algorithms besides the ones evaluated here, which would likely produce different results. For example, MacroModel here refers to their serial torsional/low-mode conformational sampling method as detailed below. MacroModel also supplies eight other search algorithms: torsional sampling, serial torsional sampling, systematic torsional sampling, mixed torsional/low-mode sampling, low-mode sampling, serial low-mode sampling, large-scale low-mode sampling and mixed torsional/large-scale low-mode sampling. Furthermore, most methods have adjustable parameters that can affect the results. Every effort was made to use sensible parameters for the specific problem at hand, and all nondefault values are listed in the method-specific sections below.

Conformations (10 000 total) were requested from each of these methods except Omega, CAESAR, and poling Catalyst, which use a different strategy for selecting conformations (*vide infra*). For consistency, the raw conformations generated by each method were minimized using the MMFF94s force field^{64–68} and the BFGS variable metric minimization algorithm, as implemented in the DirectedDiversity software suite.^{69,70} This software has been tested thoroughly and has passed the entire MMFF94s validation suite.⁷¹ In some cases (Omega and MOE stochastic search), it was not easy to decouple energy minimization from the actual search, so a 500 kcal/mol energy cutoff was applied during the search in order to retain all remotely plausible conformations before secondary minimization, in an effort to maximize the diversity of the final conformations.

The geometric diversity of the conformations produced by each search algorithm was quantified using the radius of gyration and the number of unique pharmacophore-like triplets in the resulting ensembles, both computed using DirectedDiversity.^{69,70} For each conformation, all unique 3-point pharmacophores were calculated using the atom type definitions described below and using 21 distance bins spanning the intervals [0, 1], (1, 2], (2, 3], ..., (19, 20] and (21, +∞). Each triplet was canonicalized, using a simple

vertex ordering algorithm, and assigned a unique identifier, or a bit location in a binary fingerprint. The fingerprints of all the conformations generated by a given method for a given molecule were combined into a method-specific, multiconformation molecular fingerprint using the binary union operator (inclusive OR). The number of '1' bits in these multiconformation fingerprints was used to quantify the pharmacophore space covered by that specific method and molecule. The precise binning strategy and atom-type definitions were chosen to provide sufficient resolution and discriminatory power and were not intended to be exhaustive. Because the three structural classes differ substantially in their functional composition and because our primary interest was to explore geometric rather than true pharmacophore diversity, we used a different set of atom types for each class. For cyclodextrins, we used oxygen in a 6-membered ring as the only atom type, for the cycloglycines we used oxygen and nitrogen, and for the bioactive peptides we used oxygen, nitrogen, and carbon in a terminal methyl group to capture in part the orientation of side chains.

Unless noted otherwise, all calculations were carried out on a series of IBM Intellistations running Windows XP Professional and Red Hat Enterprise Linux and equipped with two 3.2 GHz Xeon processors and 2048 Mb of RAM. The 12 conformational search methods are described in greater detail below.

CAESAR. CAESAR (conformer algorithm based on energy screening and recursive buildup) is a new, divide-and-conquer conformation generation algorithm developed by Accelrys.⁵⁷ The method involves three basic steps. First, the molecule is recursively partitioned into a set of "smallest" fragments, which include rings and rigid substructures. The second step identifies viable conformations for these smallest fragments including rings, which are handled using the same ring conformation generation module that is employed in the "fast" variant of Catalyst (for a problem with this approach, see Results and Discussion). Finally, the method assembles the molecular conformation in a recursive manner starting from the smallest fragments. This step takes into consideration local rotational symmetry in order to eliminate duplicate conformations arising from topological symmetry and uses energy criteria to prune the branches of the search tree that would lead to high-energy conformations. In this study, 10 000 conformers were requested for each molecule using an energy threshold of 30 kcal/mol during the search. Duplicate checking was disabled (confGen.removeTopSym-Duplicate parameter set to false), and all other parameters were kept at their default values, as defined in Discovery Studio, volume 1.7. As with the other methods, the final conformations were all minimized using DirectedDiversity.^{69,70}

Catalyst. Four sets of conformations were generated using the command-line tool "catConf" supplied with Catalyst version 4.10.⁷² Two of these sets were obtained using the "best" conformation generation routine, which produces conformations of better quality and diversity, and the second with the "fast" method, which achieves better speed. Two additional variants were explored for each method, one with and one without poling. Poling is a technique for repelling similar conformations during the sampling through an artificial potential, thereby promoting conformational variation.⁷³ 10 000 conformations were requested in each case,

and the program indeed produced as many for most molecules, although for some small molecules it generated only as few as 115. Duplicate checking was disabled (confGen.removeTopSymDuplicate parameter set to false), and all other parameters were kept at their default values.

MacroModel. A total of 10 000 conformations were generated for each molecule using the serial torsional/low-mode conformational sampling method in MacroModel version 9.1111.⁷⁴ This is a hybrid technique that combines broad Monte Carlo sampling of torsional space⁷⁵ with local low-frequency eigenvector sampling in the vicinity of the current conformation.⁷⁶ Since minimization could not be decoupled from the actual search and since we wanted to apply a consistent minimization technique across all methods, a single cycle of truncated Newton conjugate gradient⁷⁷ minimization was performed on each conformation using the MMFF94s force field, and no energy cutoff was applied to discard unreasonable conformations. Default values were used for all other parameters.

MOE. The MOE conformations were generated using the stochastic search algorithm available in MOE version 2006.8.⁷⁸ This method is similar to Ferguson's random incremental pulse search method⁷⁹ in that new conformations are generated via random perturbation of the parent conformation but differs in that the perturbation is of bonds rather than Cartesian coordinates. For each molecule, 10 000 conformations were requested, and the resulting geometries were minimized in MOE using MMFF94s with the distance-dependent dielectric disabled. A cutoff of 90 successive failed attempts at generating new conformations prior to termination was applied in order to maximize the number of conformations produced within the specified energy window. A 500 kcal/mol energy cutoff was used in order to retain all sensible conformations for subsequent minimization with DirectedDiversity,^{69,70} and all other parameters were set to their default values.

Omega. Omega version 1.8.1⁸⁰ was employed to generate conformations using a systematic rule-based approach. This algorithm divides each molecule into component fragments, which may contain up to five contiguous rotatable bonds. A library of predefined angles is used to generate conformations for each fragment, which are then assembled together to construct the conformations of the whole molecule using a depth-first, divide-and-conquer approach, driven by the fragment energies. For each molecule, 10 000 conformations were requested. Gas-phase conditions were mimicked by eliminating the coulombic and attractive van der Waals terms during the final refinement. Default values were used for all the remaining parameters.

Rubicon. Rubicon⁸¹ is a distance geometry method that attempts to generate conformations that satisfy a set of geometric constraints. These constraints are derived from the molecular connectivity table and fall into two categories: distance constraints that require the distance between two atoms i and j , d_{ij} , to fall within certain bounds $l_{ij} \leq d_{ij} \leq u_{ij}$ and volume constraints that require the signed volume, V_{ijkl} , formed by four atoms i, j, k, l to fall within certain bounds $V_{ijkl}^l \leq V_{ijkl} \leq V_{ijkl}^u$. Volume constraints are used to enforce the planarity of conjugated systems and the correct chirality of stereocenters. Collectively, these constraints capture all possible three-dimensional (3-D) geometries attainable by a given molecule.

Rubicon starts by defining the upper and lower bounds for all interatomic distances and chiral volumes present in the molecule, using a set of chemical rules encoded in the form of SMARTS patterns and derived from standard covalent geometry. It then selects a random set of distances within these bounds and uses the "metric matrix" algorithm⁸² to generate approximate 3-D coordinates, followed by conjugate gradient minimization to refine these coordinates so that they satisfy the original bounds.

In this study, a total of 10 000 conformations were generated for each molecule, using one trial per conformation, 0.5 Å maximum distance violation, 0.5 Å³ maximum volume violation (and volume constraints enabled), 1–4 bump checking enabled, redundancy checking disabled, and default values for all the remaining parameters. Since subsequent minimization was to be performed on all conformations, hydrogens were ignored even though their inclusion is known to generate better geometries.

SPE. Just like Rubicon, SPE generates molecular conformations that satisfy a set of distance and volume constraints. SPE starts from a random initial configuration and uses a self-organizing scheme to rapidly refine the atomic positions so as to satisfy all the input constraints. This scheme randomly selects and minimizes one constraint at a time by adjusting the coordinates of the atoms involved in that constraint.

In the present study, we used the boosting variant of SPE described in ref 38. Boosting is a simple heuristic that can be used in conjunction with SPE to generate increasingly extended (or compact) conformations through iterations. In the first iteration, a normal SPE embedding is performed, as described above, generating a chemically sensible conformation c_1 . The lower bounds of all atom pairs $\{l_{ij}\}$ are then replaced by the actual interatomic distances $\{d_{ij}\}$ in conformation c_1 and used along with the unchanged upper bounds $\{u_{ij}\}$ and volume constraints $\{V_{ijkl}^l\}$ and $\{V_{ijkl}^u\}$ to perform a second embedding to generate another conformation, c_2 . This process is repeated for a prescribed number of iterations. The lower bounds are then restored to their original default values, and a new sequence of embeddings is performed using a different random number seed. An analogous procedure can be used to generate increasingly compact conformations (by replacing the upper as opposed to the lower bounds). In our previous study, we found that boosting significantly improved the range of geometries sampled during the search and compensated for SPE's general tendency to produce relatively compact conformations.⁵²

A total of 10 000 conformations (9 999 in some cases) were generated for each hydrogen-depleted molecule using the default sampling and learning rate parameters. Each conformation was derived by running the algorithm with a different random number seed. For the smaller and more constrained macrocycles (G6, G8, C6, C8, and B1–B6), only one level of boosting was used, that is, two sets of 5 000 conformations were generated, one using the original distance bounds and the other using the boosted bounds from the first set. For the intermediate macrocycles (G10, G12, G14, G16, C10, C12, and C14), two levels of boosting were used to produce three sets of 3 333 conformations or a total of 9 999 conformations per molecule. Finally, for the larger rings (G18 and G20), a third level of boosting was added to produce

four sets of 2 500 conformations or a total of 10 000 conformations per molecule.

SOS. Self-organizing superimposition (or SOS) is an improved variant of SPE that accelerates convergence by decomposing the molecule into rigid fragments and by using precomputed conformations for those fragments in order to enforce the desired geometry.⁵⁸ Starting from completely random initial coordinates, the SOS algorithm repeatedly superimposes the templates to adjust the positions of the atoms, thereby gradually refining the conformation of the molecule. Coupled with pairwise atomic adjustments to resolve steric clashes (a step that is identical to the core self-organizing step in SPE), the method is able to generate conformations that satisfy all geometric constraints at a fraction of the time required by SPE. The approach is conceptually simple, mathematically straightforward, and numerically robust and allows additional constraints to be readily incorporated. Since rigid fragments are precomputed, planarity and chirality constraints are automatically satisfied after the template superimposition process, and local geometry is perfectly restored. Furthermore, because each embedding starts from completely random initial atomic coordinates, each new conformation is independent of those generated in the previous runs, resulting in greater diversity and more effective sampling (also true for Rubicon and SPE). As the algorithm only involves pairwise distance adjustments and superimposition of relatively small fragments, it is very efficient.

Two types of SOS runs were performed: one with boosting (SOS2) and one without (SOS). For the boosted runs, we used the same protocol that we employed for SPE, namely we generated $2 \times 5\,000 = 10\,000$ conformations for G6, G8, C6, C8, B1–B6, $3 \times 3\,333 = 9\,999$ conformations for G10, G12, G14, G16, C10, C12, and C14, and $4 \times 2\,500 = 10\,000$ conformations for G18 and G20.

RESULTS AND DISCUSSION

The effectiveness of the 12 conformational search algorithms was assessed using five different metrics: the ability to identify the global minimum, the number of unique conformations within 20 kcal/mol from the lowest energy structure identified by each method as well as from the global minimum, the distribution of energies and radii of gyration of these conformations, and the number of unique pharmacophore triplets encoded by each conformational ensemble. The results are summarized in Tables 2–3 and Figures 1–11.

Figures 1, 3, and 5 contain a series of panels for each class of compounds showing the relative energies of the minimized conformations obtained for each method and molecule (cycloglycines in Figure 1, cyclodextrins in Figure 3, and bioactive peptides in Figure 5). Only conformations within 20 kcal/mol of the lowest energy structure identified by each method were included. All the energies are plotted relative to the global minimum, which, for the purposes of this study, is defined as the lowest energy structure identified by any method for a given molecule. Figures 2, 4, and 6 contain a similar series of panels showing the distribution of the radii of gyration of all the conformations obtained by each method after minimization. For ease of comparison, all the heatmaps are identically scaled to cover the full range of energies/radii encountered across all conformations,

molecules, and methods. These plots were generated using Third Dimension Explorer, which is part of the ABCD informatics platform.⁸³

The lowest energies identified for each method and molecule are summarized in Figure 7. Since the absolute energies range widely across the 19 molecules, the data bars in Figure 7 are scaled independently for each one, and they range from the lowest to the highest energy in each row. This graphical representation reveals some clear trends. The four distance geometry (DG) methods (Rubicon, SOS, SOS2, and SPE) outperform all other search algorithms by a significant margin, and they are the only ones that show consistently good performance across all three classes of macrocycles (cycloglycines, cyclodextrins, and bioactive cyclopeptides). SOS2 and SPE identified the global minimum in 6 out of the 19 molecules (G6, G8, G16, D6, D8, and D10 for SOS2; G6, G8, G18, G20, P3, and P6 for SPE), and Rubicon and SOS identified it in 5 molecules (G6, G8, G10, P1, and P5 for Rubicon; G6, G8, G10, G12, and G14 for SOS). For the remaining cases, all of these DG methods identified alternative conformations very close to the global minima, as indicated by the consistently short bars across the corresponding columns. SOS2 shows the best and the most consistent performance, followed by SOS, SPE, and Rubicon, with SPE being a little less effective with the cyclodextrins and Rubicon with the larger cycloglycines and cyclodextrins. SOS2 is the only method that identified at least one conformation within 20 kcal/mol of the global minimum for every single molecule in our study.

From the remaining methods, MacroModel was the most effective, followed by CatalystB, CatalystBP, MOE and Omega. MacroModel identified the global minimum for 4 out of the 19 molecules (G6, G8, D12, and D14) and did much better for the cycloglycines and the cyclodextrins compared to the bioactive cyclopeptides. In fact, for the first two classes of macrocycles, MacroModel's performance was comparable to that of the four DG methods, but it had significant difficulty with the structurally more intricate bioactive peptides. CatalystB and CatalystBP did reasonably well with the smaller cycloglycines, but their performance declined as the number of residues increased. However, both of these methods did rather poorly with the bioactive peptides and did very poorly with the cyclodextrins. The diminishing performance with increasing ring size was also true for MOE and Omega, but the former did significantly better with bioactive peptides. Omega's performance was comparable to MOE's for the cycloglycines and cyclodextrins, but it was significantly worse for the bioactive peptides with the exception of Aciculitin A (P2), a molecule with a long and a relatively linear side chain that lends itself to the systematic search algorithm embedded in Omega. CatalystB, CatalystBP, MOE, and Omega identified the global minimum for 2 out of the 19 molecules, and CAESAR, CatalystF and CatalystFP identified it for none of the molecules.

The average performance of each method for each structural class as well as the entire collection is summarized in Figure 11. To account for the large differences in absolute energies, the plot shows the difference between the lowest energy conformation identified by each method relative to the global minimum for that particular molecule expressed as a fraction of the global minimum averaged over all molecules (i.e., the height of the bar y_j is given by

Table 3. Median Energies and CPU times Required for the Generation and Minimization of 100 Conformations of the Bioactive Cyclopeptide P4 (MicrocystinLR) using Different Methods^a

method	CPU ^b	Confo ^c	T_{gen}^d	T_{min}^e	T_{min}^f	T_{min}^g	T_{min}^h	T_{min}^i	$T_{\text{gen}} + T_{\text{min}}^j$	median energy (kcal/mol)				
										E_0^k	E_1^l	E_{100}^m	E_{100}^n	E_{100}^o
CAESAR	Intel Core2 Duo T7800 2.60 GHz	100	5.00	2.89	4.81	19.67	140.46	826.59	831.59	2665.38	1445.41	259.58	-40.34	-114.83
CatalystB	Intel Core2 Duo T7800 2.60 GHz	100	305.00	2.95	4.34	16.53	132.17	706.96	1011.96	-11.96	-22.43	-44.58	-76.82	-103.81
CatalystBP	Intel Core2 Duo T7800 2.60 GHz	100	339.00	2.98	4.32	16.54	131.87	726.40	1065.40	-11.57	-22.66	-45.68	-76.46	-104.59
CatalystF	Intel Core2 Duo T7800 2.60 GHz	98	7.00	2.81	4.32	17.34	132.65	740.57	747.57	67.53	53.56	-7.69	-73.69	-102.97
CatalystF ^{sp}	Intel Core2 Duo T7800 2.60 GHz	98	7.00	2.87	4.41	17.69	135.36	755.68	762.68	67.53	53.56	-7.69	-73.69	-102.97
CatalystFP	Intel Core2 Duo T7800 2.60 GHz	100	7.00	2.92	4.31	17.40	134.37	747.46	754.46	60.81	46.85	-17.86	-75.78	-105.42
MacroModel	Intel Xeon 3.6 GHz	58	12.00	2.53	5.87	19.26	121.20	686.28	698.28	3071.78	2493.96	752.24	-29.01	-115.13
MacroModel ^{sp}	Intel Xeon 3.6 GHz	58	12.00	4.36	10.12	33.21	208.97	1183.24	1195.24	3071.78	2493.96	752.24	-29.01	-115.13
MOE	Intel Xeon 3.6 GHz	5	27.00	1.12	0.32	0.96	7.06	42.76	69.76	426.51	396.99	278.44	33.31	-92.66
MOE ^{sp}	Intel Xeon 3.6 GHz	5	27.00	22.40	6.40	19.20	141.20	855.20	882.20	426.51	396.99	278.44	33.31	-92.66
Omega	Intel Xeon 3.6 GHz	99	141.00	2.92	4.29	16.85	130.45	679.75	820.75	1.00	-12.43	-58.63	-90.92	-112.84
Omega ^{sp}	Intel Xeon 3.6 GHz	99	141.00	2.89	4.25	16.68	129.15	672.95	813.95	1.00	-12.43	-58.63	-90.92	-112.84
Rubicon	AMD Opteron 850 2.4 GHz	100	39.39	9.62	11.09	24.42	143.32	821.93	861.32	322.37	239.83	100.37	-37.00	-96.77
SOS	Intel Xeon 3.8 GHz	100	1.71	9.95	11.87	27.03	149.15	835.50	837.21	2178.39	1350.24	361.66	-30.11	-112.85
SOS2	Intel Xeon 3.8 GHz	100	2.25	9.89	11.70	26.21	147.32	857.64	859.89	1395.85	822.99	243.02	-39.03	-115.03
SPE	Intel Xeon 3.8 GHz	100	40.75	9.78	11.39	24.79	143.50	792.12	832.87	198.91	148.98	27.57	-69.14	-115.61

^a Different CPUs were used only for the generation of the conformations. All minimizations were carried out on the same Intel Xeon 3.80 GHz processor. Five CPU times and energies are reported: those obtained after 0, 1, 10, and 100, and up to 1000 minimization steps (sometimes the minimization converged in fewer steps). All programs were requested to generate 100 conformations, but in the case of CatalystF, MacroModel, MOE, and OMGEA fewer conformations were produced and minimized. In those cases, two minimization times are reported: the time T required to minimize only the resulting K conformations (outlined in gray), and the time it would have required if K were equal to 100 ($T^* = T \times 100/K$). ^b Processor used to generate the conformation. ^c Number of conformations generated (100 requested). ^d CPU time required to generate the conformations. ^e Time required to compute the energy of the raw conformations (single point). ^f Time required for 1 minimization step. ^g Time required for 10 minimization steps. ^h Time required for 100 minimization steps. ⁱ Time required for up to 1000 minimizations steps (minimization may have converged in fewer steps). ^j Total time required to generate and to minimize the conformations (note that for CAESAR, Catalyst, MacroModel, MOE, Omega, and Rubicon generation and minimization were done on different CPU's). ^k Median energy of raw conformations. ^l Median energy of conformations after 1 minimization step. ^m Median energy of conformations after 10 minimization steps. ⁿ Median energy of conformations after 100 minimization steps. ^o Median energy of conformations after up to 1000 minimization steps (minimization may have converged in fewer steps). ^p Minimization times scaled to 100 conformations (i.e., if a method produced only $K < 100$ conformations, the time reported is $T^* = T \times 100/K$).

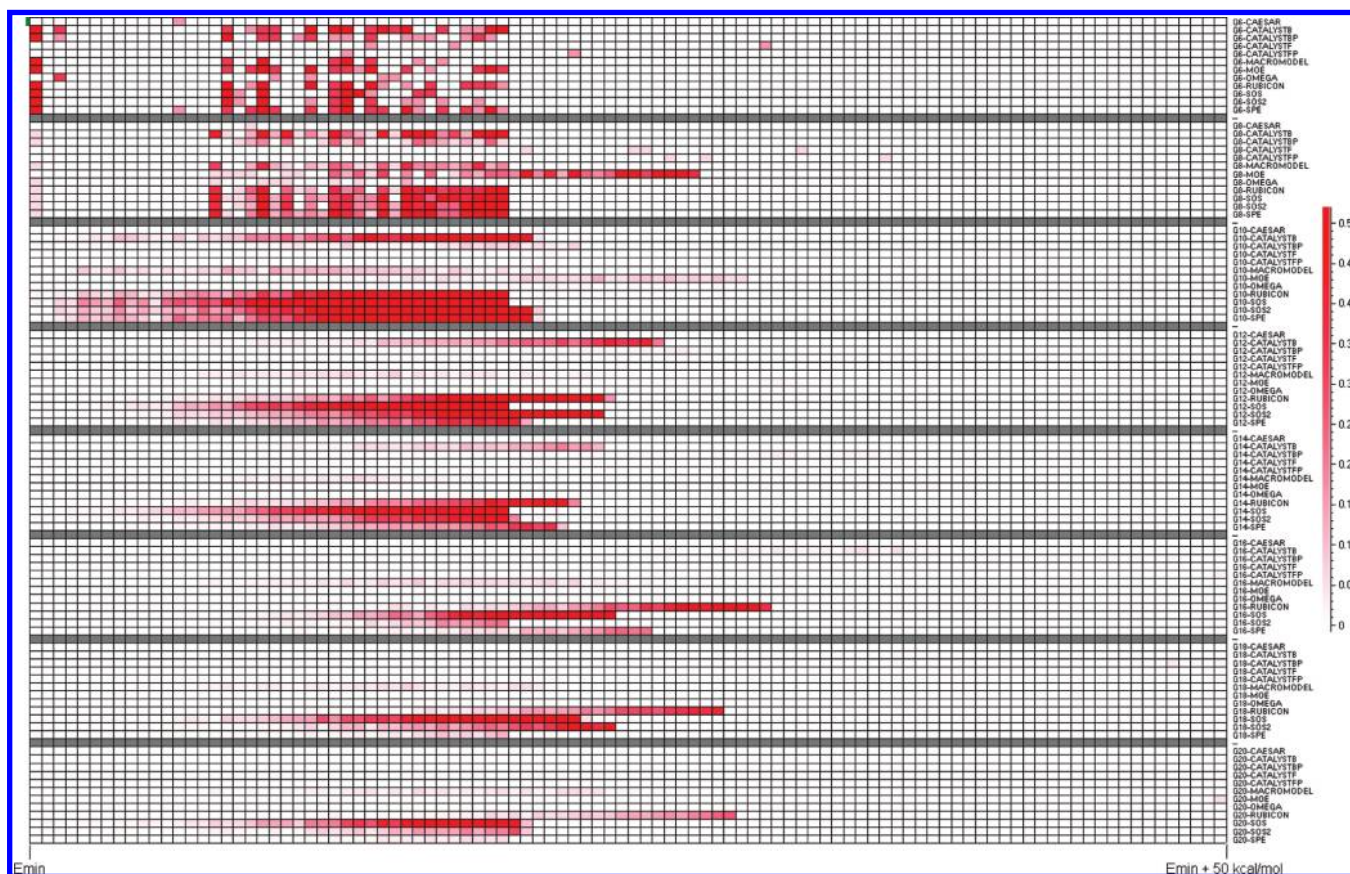


Figure 1. Heatmap showing the distribution of relative energies (in kcal/mol) of the minimized conformations of cycloglycines obtained by each method. Energies are relative to the global minimum identified by any method for a given molecule. Each row ranges from 0 to 50 kcal/mol and is divided into 100 evenly sized bins of 0.5 kcal/mol width each. The bins are color coded according to their relative densities (number of conformations that fall within them), which are normalized separately for each molecule, so that the most populated bin across all methods for each particular molecule is assigned a density of 1. To reveal finer differences in the low-density regions of the map, we scaled the colors from 0 to 0.5 (see color scale on the right), so that all cells with density >0.5 appear in the same bright-red color.

$y_j = (1)/(N) \sum_{i=1}^N (E_{ij}^{\min} - \min_j E_{ij}^{\min}) / \min_j E_{ij}^{\min}$, where E_{ij}^{\min} is the lowest energy of the i -th molecule identified by the j -th method, and N is the number of molecules in the class). Four averages were computed and highlighted in different colors: the average over all cycloglycines is shown in gray, the average over all cyclodextrins is red, the average over all bioactive cyclopeptides is green, and average over all molecules is blue. This plot reveals that the four DG methods (Rubicon, SOS, SOS2, and SPE) have a very significant advantage over the other techniques, producing conformations of consistently lower energy for all three structural series. The fast variants of Catalyst (CatalystF and CatalystFP) did extremely poorly as did CAESAR, producing minima 8 to 16 times higher in relative energy compared to those of SOS, SOS2, SPE, and Rubicon.

The four DG methods also identified the greatest number of unique conformations within 20 kcal/mol of the lowest energy structure identified by each method individually (Figure 8a) as well as all methods collectively (Figure 8b). In many cases, the advantage is overwhelming, particularly for SOS, which identifies almost twice as many conformations as any other method for some of the more flexible molecules. Figures 7 and 8 show a high correlation between the number of unique conformations and the ability to identify the global minimum. Both are linked to the sampling power of the underlying search algorithm. The greater the sampling capacity of the search method, the more conforma-

tions it will uncover and the greater the likelihood that a lower energy structure will emerge among them (and, therefore, the greater the probability of finding the global minimum).

Just as with the energy minima, MacroModel is the most effective of all the remaining non-DG methods, particularly for the larger cycloglycines and cyclodextrins. The limitations of the remaining methods are striking. As shown in Figure 8b, Omega, the fast variants of Catalyst (CatalystF and CatalystFP), and CAESAR and to a lesser extent MOE and the best variants of Catalyst (CatalystB and CatalystBP) cannot find even a single conformation within 20 kcal/mol of the global minimum for most of the molecules under study. CatalystB is more successful with the smaller cycloglycines, but its relative performance drops rapidly with ring size. MOE shows a similar pattern but, unlike the other methods, it does reasonably well for the smaller bioactive peptides P3 and P5.

CAESAR's performance with cycloglycines was a little more perplexing. CAESAR and the fast variant of Catalyst share the same ring conformation generation module, and one would expect that the method would have identified more than one conformation for at least some of the cycloglycines. After repeating the experiments to confirm our findings, we contacted the developers of CAESAR who traced the problem to a bug that caused the program to return only one conformation if the molecule was comprised of a single ring

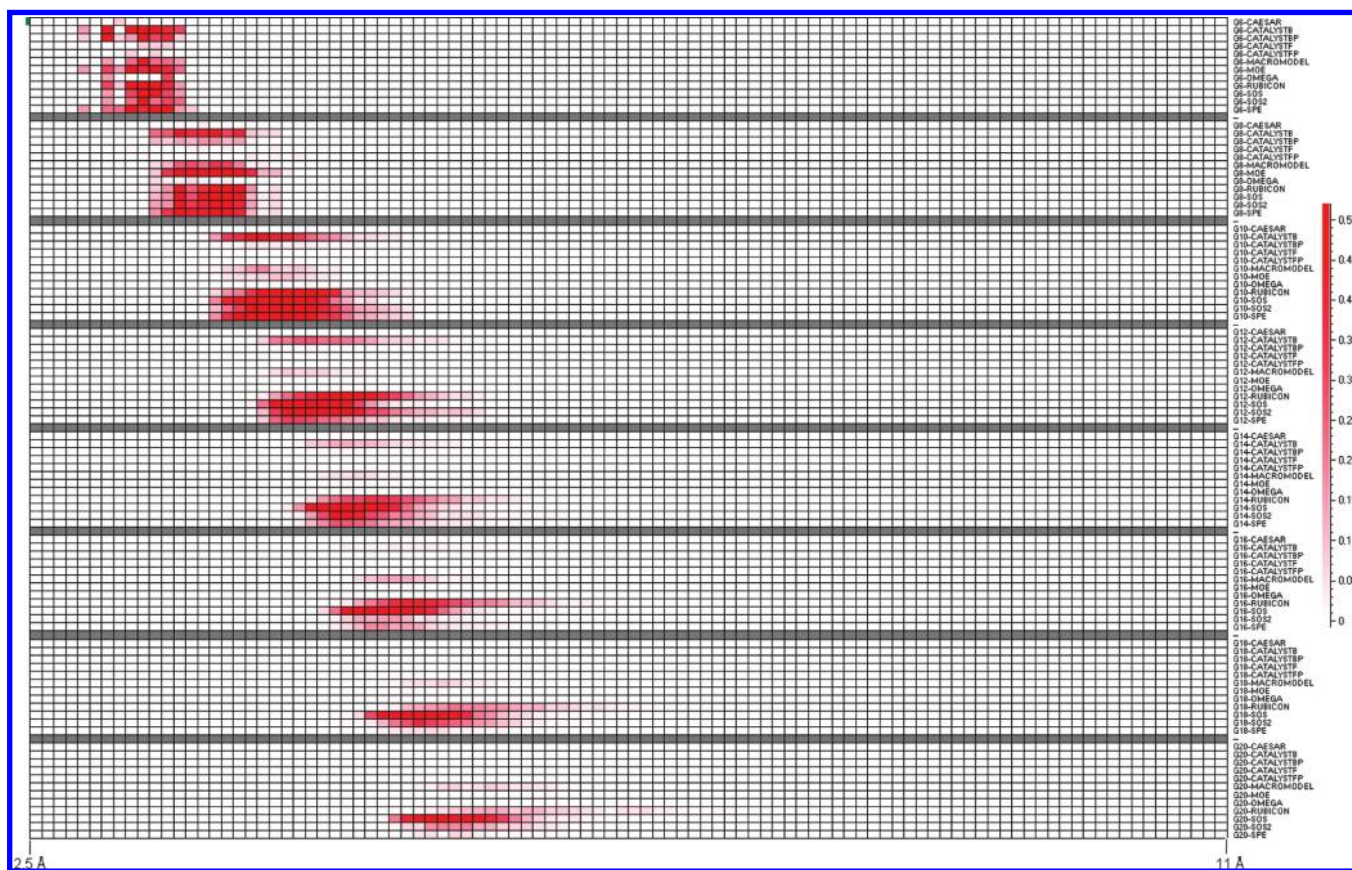


Figure 2. Heatmap showing the distribution of radius of gyration of the minimized conformations of cycloglycines obtained by each method. All rows range from 2.5 to 11 Å, which covers the full range of radii encountered across all conformations, molecules, and methods. Similarly to Figure 1, the bins are color coded according to their relative densities (number of conformations that fall within them), which are normalized separately for each molecule, so that the most populated bin across all methods for each particular molecule is assigned a density of 1. To reveal finer differences in the low-density regions of the map, we scaled the colors from 0 to 0.5 (see color scale on the right) so that all cells with density >0.5 appear in the same bright-red color.

and no side chains. We were told that the problem has been corrected, and a fix will be available in the upcoming release of Discovery Studio (version 2.5).⁸⁴

While the number of unique conformations is usually indicative of the sampling capacity of a search algorithm, it leaves open the question as to whether these conformations cover an entire conformational space or a relatively narrow region of it. To address this question, we used the distribution of the radii of gyration and the number of unique pharmacophore triplets encoded by the conformational ensembles to further quantify conformational diversity. The results are summarized in Table 2 and Figures 9 and 10. Once again, only conformations within 20 kcal/mol of the global minimum were considered. Table 2 lists the number of unique conformations (N), the minimum energy (E_{\min}), the minimum and maximum radius of gyration (R_{\min} , R_{\max}), and the total number of unique pharmacophore triplets (N_p) identified for each method and molecule. This table and the plots in Figures 9 and 10 reveal similar trends to those described above and show a compelling advantage for SOS2, SOS, SPE, and Rubicon over the remaining methods, with MacroModel a distant fifth and with CatalystB with and without poling a distant sixth and seventh, respectively. The differences become increasingly pronounced as the flexibility of the molecule increases. In general, SOS identifies the most compact conformations as manifested by the minimum radius of gyration R_{\min} , while SOS2 and SPE identify the most extended conformations as manifested by the maximum

radius of gyration R_{\max} . This is fully consistent with our previous findings that both SPE and SOS have a general tendency to generate compact conformations and that boosting is very effective at probabilistically biasing the search towards more extended geometries. The net result is ensembles that contain many more pharmacophore-like triplets (Figure 10) and span a much wider range of molecular radii (Figure 9).

It is pretty clear that, for the more flexible molecules, 10 000 trials are insufficient to sample the entire conformational space and that the global minimum may not be present in the generated ensembles. It is also interesting that the number of unique conformations is not a simple function of ring size. For example, SOS2 identified 3 087 and 3 673 conformations for G12 and G14 but only 747, 1 347, and 1 469 for G16, G18, and G20. The corresponding global minima are shown in Figure 12 in the Supporting Information along with the actual 3-D structures in SD format.

In all cases, the global minima are characterized by an extensive network of intramolecular hydrogen bonds. For the larger macrocycles, this network can be quite intricate, leading to conformations of considerably lower energy than random ones. The global minima of the eight cycloglycines contain 6, 8, 8, 10, 13, 16, 16, and 16 hydrogen bonds and belong to the symmetry groups S_6 , C_2 , C_i , C_i , C_1 , C_2 , C_1 , and C_1 , respectively. The global minima for G14, G18, and G20 show no apparent symmetry, but it is possible that a lower energy structure with higher symmetry may exist for

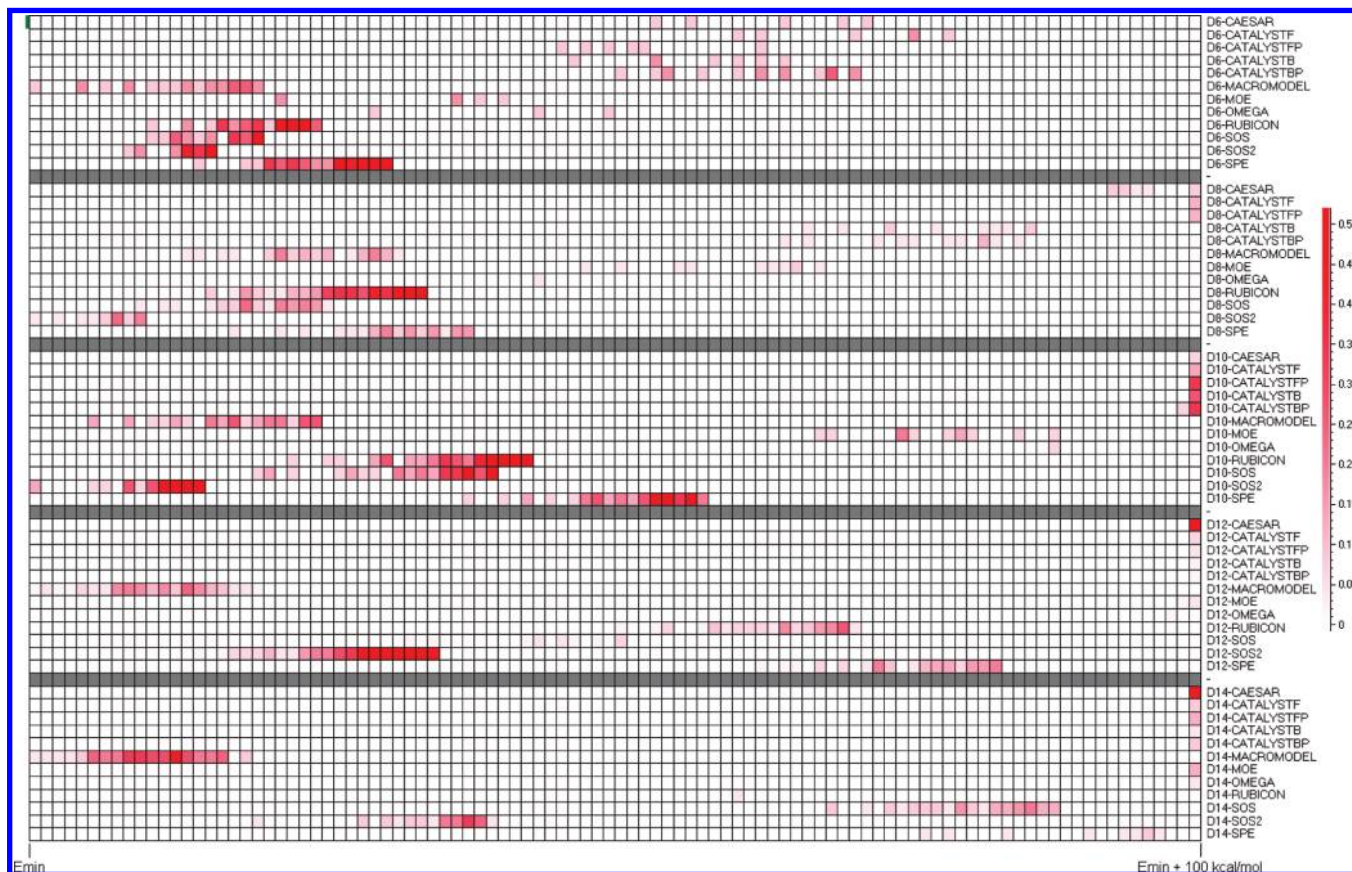


Figure 3. Heatmap showing the distribution of relative energies (in kcal/mol) of the minimized conformations of cyclodextrins obtained by each method. The map is similar to that of Figure 1, except that the energy window ranges from 0 to 100 kcal/mol.

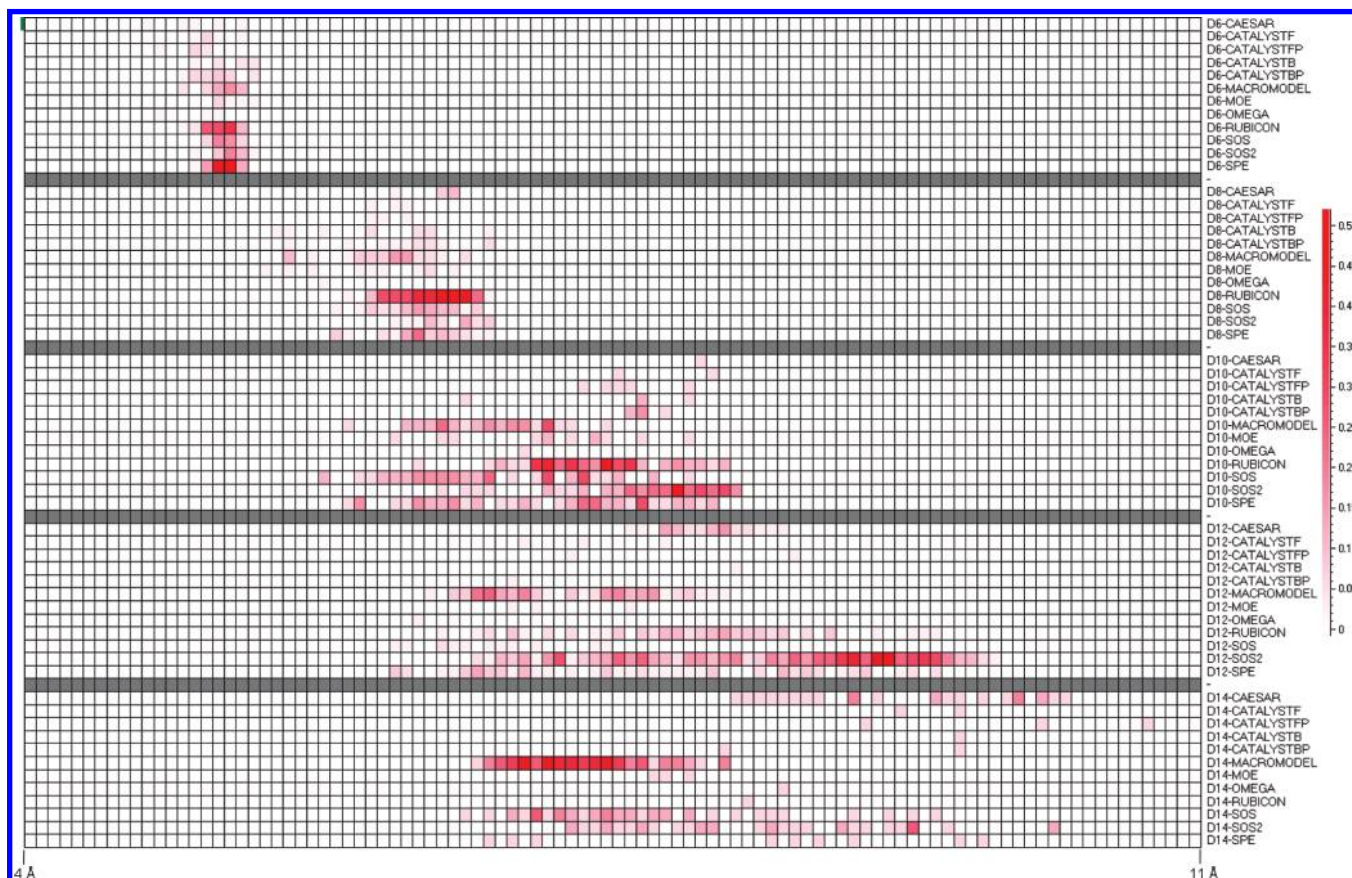


Figure 4. Heatmap showing the distribution of radius of gyration of the minimized conformations of cyclodextrins obtained by each method. The map is similar to that of Figure 2, except that the window of radii ranges from 4 to 11 Å.

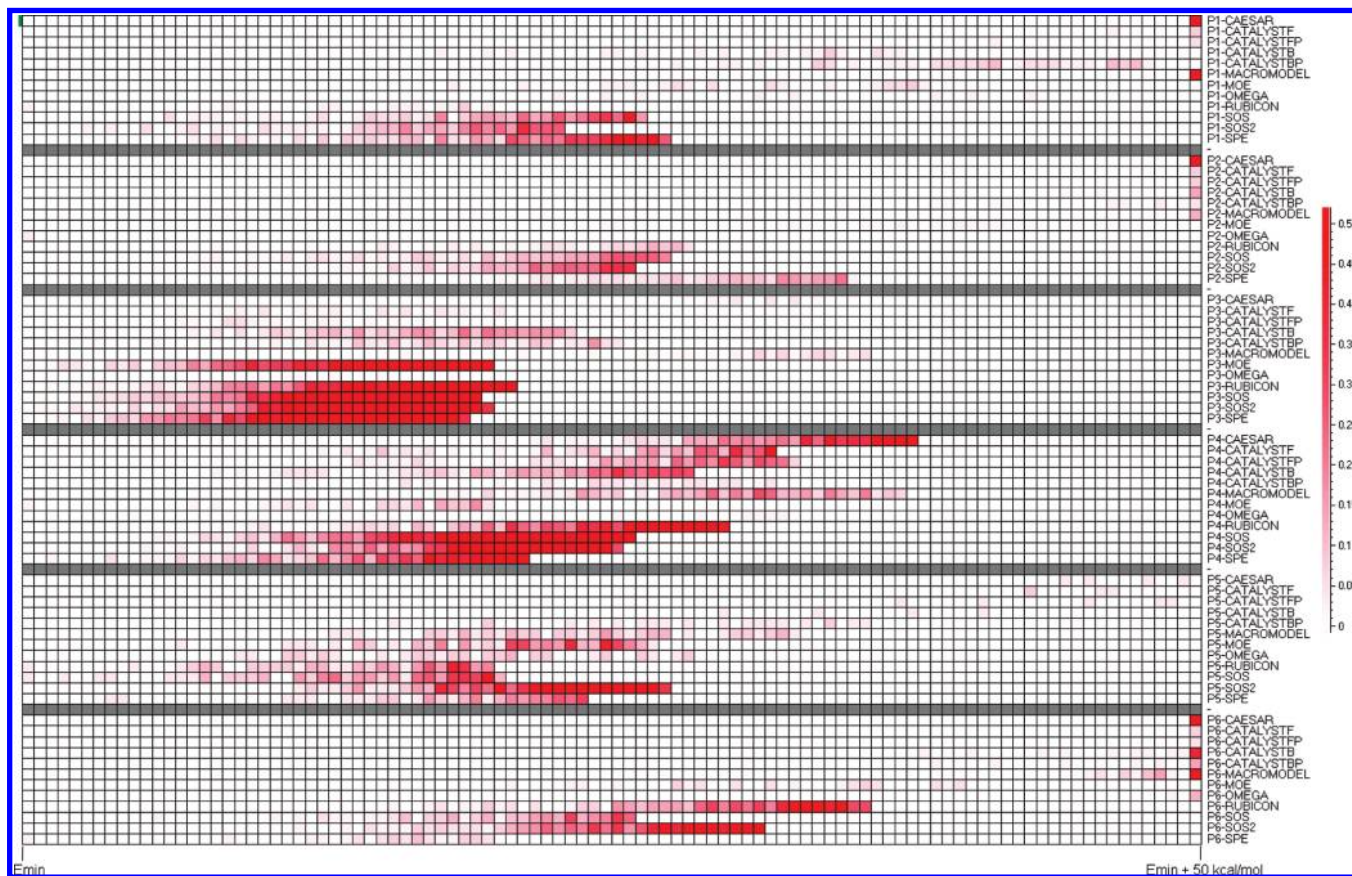


Figure 5. Heatmap showing the distribution of relative energies (in kcal/mol) of the minimized conformations of bioactive cyclopeptides obtained by each method. The map is similar to that of Figure 1, with the energy window ranging from 0 to 50 kcal/mol.

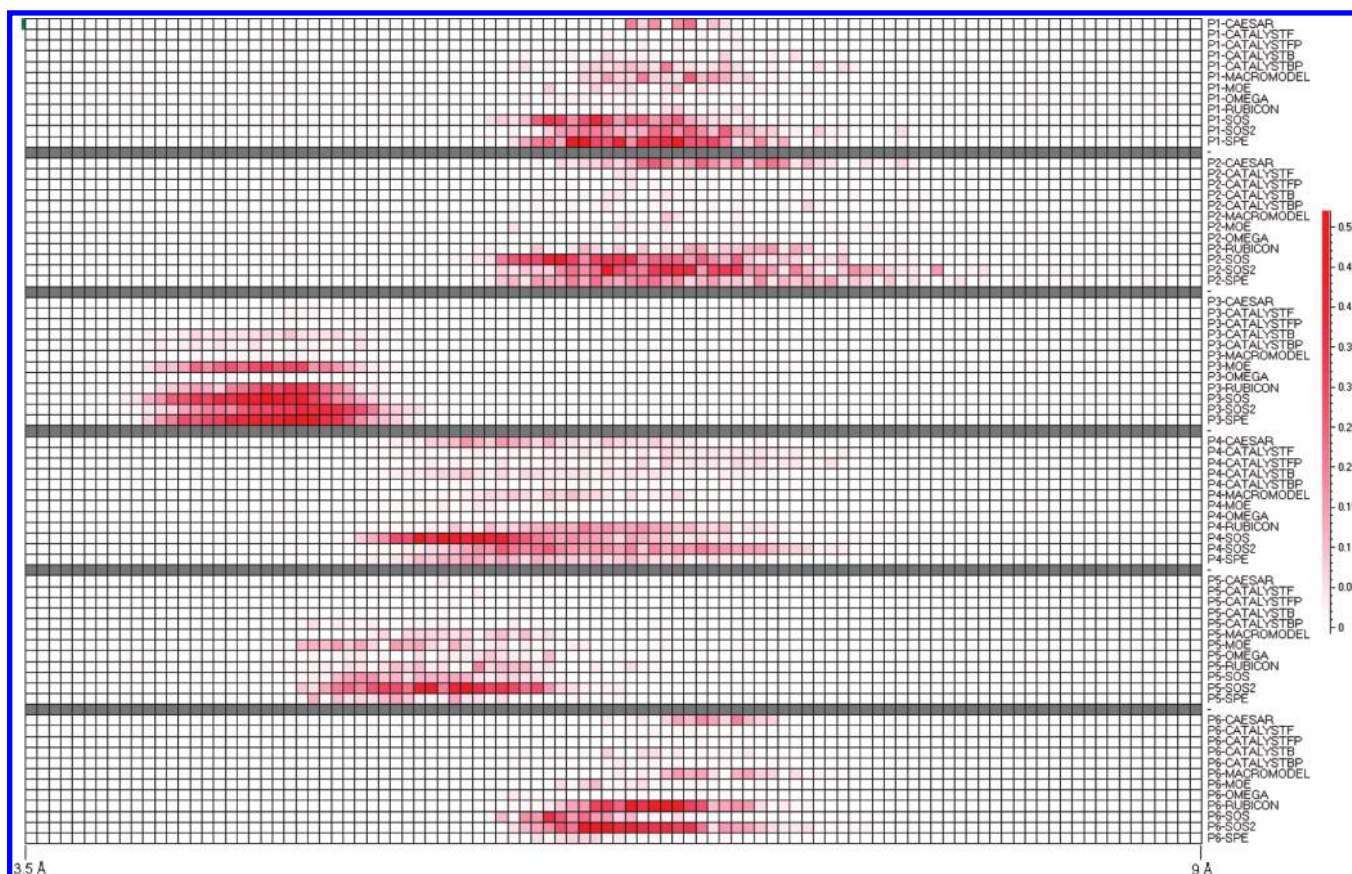


Figure 6. Heatmap showing the distribution of radius of gyration of the minimized conformations of bioactive cyclopeptides obtained by each method. The map is similar to that of Figure 2, except that the window of radii ranges from 3.5 to 9 Å.

MOLECULE	CAESAR	CATALYSTB	CATALYSTBP	CATALYSTF	CATALYSTFP	MACROMODEL	MOE	OMEGA	RUBICON	SOS	SOS2	SPE
G6	34.89	28.74	28.74	43.08	41.92	28.74	28.74	30.05	28.74	28.74	28.74	28.74
G8	35.20	26.20	26.20	47.09	53.11	26.20	34.10	26.20	26.20	26.20	26.20	26.20
G10	43.58	36.12	37.90	57.01	63.54	36.95	44.82	45.43	35.07	35.07	36.12	36.12
G12	52.86	41.34	43.37	67.57	73.08	39.64	47.48	53.75	39.26	35.13	39.09	35.75
G14	54.52	43.65	56.21	63.51	65.75	40.39	59.40	57.03	42.57	39.89	40.07	42.06
G16	61.04	54.72	65.02	78.15	92.21	43.99	60.84	59.85	47.73	41.41	36.94	42.79
G18	66.29	64.78	75.66	83.70	93.33	46.84	75.58	80.11	53.95	48.09	49.57	45.07
G20	70.91	68.91	76.22	81.62	86.08	58.08	94.53	76.06	60.03	51.10	51.33	50.55
D6	580.07	572.78	577.21	586.92	572.70	526.71	547.84	556.35	531.18	526.76	522.64	537.44
D8	782.62	763.78	762.60	799.40	796.86	710.89	745.41	741.58	711.58	702.89	687.32	715.25
D10	980.00	971.41	967.05	989.43	989.66	873.35	936.29	955.68	890.94	887.83	863.37	906.11
D12	1174.08	1186.65	1150.50	1197.85	1183.23	1031.29	1129.99	1129.16	1081.44	1063.32	1045.60	1094.29
D14	1427.64	1345.15	1345.07	1425.51	1430.16	1193.96	1309.79	1322.02	1254.61	1261.97	1213.23	1270.46
P1	127.03	41.07	51.27	69.47	64.92	87.02	41.92	46.40	23.43	29.57	26.17	30.60
P2	204.68	195.31	185.86	200.33	204.21	207.55	175.65	150.64	159.31	158.40	157.00	166.08
P3	146.57	123.61	125.71	126.64	126.38	136.48	120.17	131.78	121.06	119.63	119.83	118.78
P4	-133.64	-143.10	-139.64	-139.54	-138.92	-134.00	-151.53	-126.35	-141.79	-145.72	-146.44	-150.18
P5	436.57	422.19	408.21	435.14	429.63	405.66	398.93	401.52	392.32	392.40	399.45	396.31
P6	71.24	59.20	59.41	140.39	142.28	64.88	41.47	69.50	36.68	26.73	32.35	20.77

Figure 7. Lowest energy identified by each method. The data bars are scaled independently for each molecule, and the range is from the lowest to the highest energy in each row.

MOLECULE	CAESAR	CATALYSTB	CATALYSTBP	CATALYSTF	CATALYSTFP	MACROMODEL	MOE	OMEGA	RUBICON	SOS	SOS2	SPE
G6	1	49	26	3	2	19	37	6	35	24	23	44
G8	1	122	38	4	3	65	149	9	121	140	132	145
G10	1	1009	92	5	3	227	158	18	1264	1862	1449	1553
G12	1	1182	75	4	3	260	32	15	2561	3087	2763	1156
G14	1	585	87	2	1	186	51	9	1878	3675	1804	1171
G16	1	84	23	3	4	241	10	6	1260	1989	300	387
G18	1	13	21	3	3	163	8	6	792	2303	944	142
G20	1	5	1	4	1	161	11	3	407	1637	385	87
D6	5	7	16	6	6	24	6	3	49	23	20	61
D8	8	10	12	3	3	34	10	1	134	36	19	35
D10	1	4	6	2	5	36	12	1	70	51	59	54
D12	32	1	1	3	2	87	3	1	67	12	294	69
D14	25	1	2	2	3	88	3	1	1	41	36	9
P1	41	20	50	4	3	66	24	8	13	155	137	243
P2	153	14	17	7	7	11	15	3	90	281	310	200
P3	13	249	126	36	20	44	1016	4	948	1603	1381	1757
P4	302	193	35	153	162	165	49	6	432	1016	764	364
P5	4	4	20	4	5	57	102	41	84	111	367	72
P6	71	26	11	3	2	70	28	6	279	119	348	30

(a)

MOLECULE	CAESAR	CATALYSTB	CATALYSTBP	CATALYSTF	CATALYSTFP	MACROMODEL	MOE	OMEGA	RUBICON	SOS	SOS2	SPE
G6	1	49	26	2	1	19	37	6	35	24	23	44
G8	1	122	38			65	42	9	121	140	132	145
G10	1	896	65			220	39	9	1263	1861	1335	1399
G12	1	384	18			224	5	1	1276	3087	1618	1046
G14	1	271	10			186	1	1	1154	3673	1753	784
G16		1				170			78	747	300	78
G18		1				141			83	1347	395	142
G20		1				101			48	1469	328	87
D6						15			16	13	20	7
D8										2	19	
D10						11					59	
D12						87					14	
D14						88					1	
P1		2					1		13	46	77	56
P2								3	16	60	52	4
P3		132	59	31	20	1	894	4	735	1439	1276	1757
P4	1	38	4	18	14	6	49		49	342	214	275
P5			3			14	28	24	84	108	93	29
P6									9	35	22	30

(b)

Figure 8. (a) Number of unique conformations within 20 kcal/mol of the lowest energy structure identified by each individual method. (b) Number of unique conformations within 20 kcal/mol of the global minimum (lowest energy structure identified by any method). The data bars are scaled independently for each molecule, and the range is from zero to the largest number of conformations in each row.

MOLECULE	CAESAR	CATALYSTB	CATALYSTBP	CATALYSTF	CATALYSTFP	MACROMODEL	MOE	OMEGA	RUBICON	SOS	SOS2	SPE
G6	0.00	0.70	0.69	0.11	0.00	0.50	0.71	0.40	0.50	0.52	0.50	0.80
G8	0.00	0.84	0.64			0.75	0.85	0.52	0.84	0.84	0.84	0.84
G10	0.00	1.60	1.36			1.16	1.01	0.97	1.56	1.63	1.75	1.82
G12	0.00	1.53	0.85			1.32	0.56	0.00	1.74	1.51	1.69	1.69
G14	0.00	2.27	0.76			1.13	0.00	0.00	2.69	1.67	2.74	2.75
G16		0.00				0.97			1.08	1.46	1.39	1.17
G18		0.00				1.52			1.88	1.59	2.93	1.56
G20		0.00				1.61			1.68	1.89	2.22	1.85
D6						0.19			0.28	0.37	0.33	0.12
D8										0.13	0.57	
D10						1.23					1.75	
D12						1.70					2.82	
D14						1.50					0.00	
P1		0.90					0.00		0.66	1.17	1.79	1.83
P2								0.39	1.70	1.84	1.78	0.56
P3		1.16	1.15	0.71	0.65	0.00	1.34	0.23	1.20	1.50	1.47	1.56
P4	0.00	2.03	0.82	1.53	1.58	0.81	1.55		1.84	2.05	2.26	2.30
P5			0.64			0.86	1.78	1.17	1.82	1.20	1.26	1.30
P6									1.00	0.86	1.02	1.21

Figure 9. Difference between the largest and smallest radius of gyration of the conformations that were produced by each method and were within 20 kcal/mol of the global minimum (lowest energy structure identified by any method). The data bars are scaled independently for each molecule, and they range from zero to the largest range of radii in each row.

MOLECULE	CAESAR	CATALYSTB	CATALYSTBP	CATALYSTF	CATALYSTFP	MACROMODEL	MOE	OMEGA	RUBICON	SOS	SOS2	SPE
G6	133	497	443	233	124	391	477	336	478	434	416	476
G8	257	1081	899			975	938	634	1101	1067	1085	1090
G10	620	2562	2053			1987	1694	1452	2672	2600	2781	2772
G12	1100	3265	1966			3021	1601	1070	3760	3598	3779	3606
G14	1074	4455	2350			3578	1212	1318	5538	4634	5859	5450
G16		1193				3823			3886	4674	4770	3980
G18		1666				5469			6114	6240	7276	5427
G20		2571				6444			6671	7406	7541	6244
D6						90			85	62	82	56
D8										78	194	
D10						370					616	
D12						978					702	
D14						1080					219	
P1		7587					5083		14254	15658	20799	19725
P2								5743	13282	18378	19174	7954
P3		7412	6273	5437	4594	993	8752	2217	7956	8488	9010	8907
P4	2645	15878	7393	10884	10287	8156	14314		15228	19236	22157	20720
P5			4438			8324	9474	9900	12971	12331	12695	9670
P6									16638	17437	18197	19700

Figure 10. Number of unique pharmacophore triplets encoded in the conformations that were produced by each method and were within 20 kcal/mol of the global minimum (lowest energy structure identified by any method). The data bars are scaled independently for each molecule, and they range from zero to the largest number of pharmacophore triplets in each row.

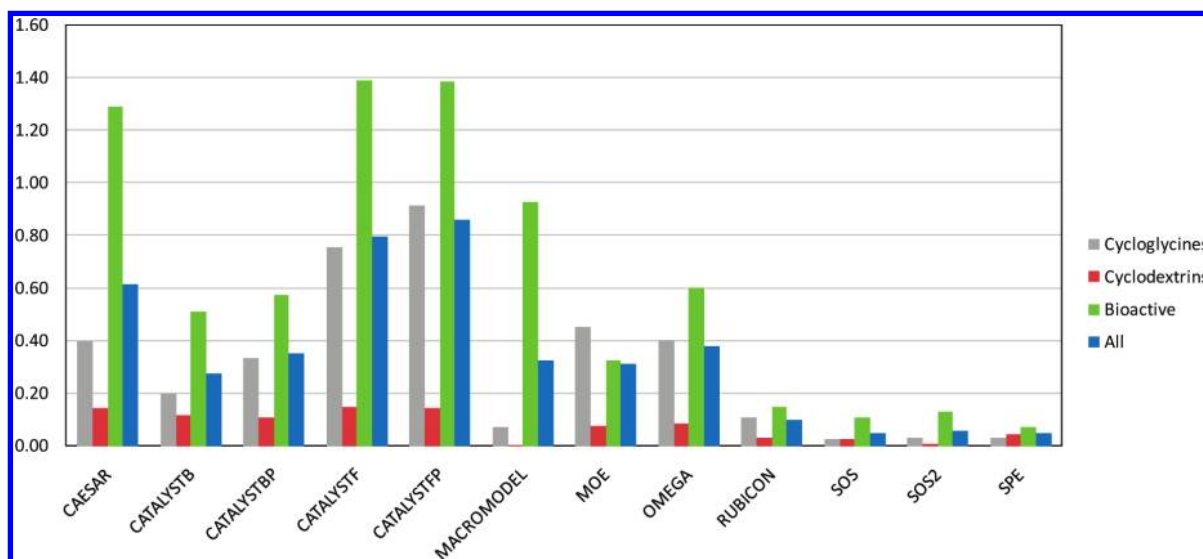


Figure 11. Difference between the lowest energy conformation identified by each method relative to the global minimum for that particular molecule expressed as a fraction of the global minimum and averaged over all molecules (i.e., $y_j = (1/(N) \sum_{i=1}^N (E_{ij}^{\min} - \min_j E_{ij}^{\min}) / \min_j E_{ij}^{\min})$, where E_{ij}^{\min} is the lowest energy of the i -th molecule identified by the j -th method, and N is the number of molecules). Gray: average over all cycloglycines; red: average over all cyclodextrins; green: average over all bioactive cyclopeptides; and blue: average over all molecules.

these molecules. Indeed, more than 30% of the generated conformations are unique for G14, and more than 10% are unique for G18 and G20, which suggests that alternative conformations of lower energy may have gone undetected due to insufficient sampling.

A more interesting situation is observed for the cyclodextrins. Here, the number of unique conformations is relatively small compared to the number of conformations sampled, which offers more confidence that the global minima have been identified. The global minima of the five cyclodextrins contain 6, 8, 11, 21, and 17 hydrogen bonds and belong to the symmetry groups C_3 , C_4 , C_1 , C_1 , and C_1 , respectively. The α - and γ -cyclodextrins (D6 and D8) have a three- and four-fold rotational symmetry, respectively, and show the typical truncated cone structure that these molecules are known to adopt, with the primary hydroxyl groups crowning the narrow rim and the two secondary hydroxyl groups crowning the wide rim. This structure creates a hydrophobic cavity inside a hydrophilic exterior, which makes it possible for these molecules to form inclusion complexes with a wide range of hydrophobic compounds. Starting with D10, the conformations begin to bend and eventually twist into double-helical structures (D12 and D14). It is interesting to

note that, unlike the reported crystal structure, the global minimum that we identified for D10 maintains the general all-cis geometry of the smaller cyclodextrins, though it is slightly bent and asymmetric. D12 exhibits a single band-flip involving a trans bond between two adjacent glucose units, and D14 assumes a double-helical fold that differs substantially from the reported boat-like crystal structure.⁸⁵ In fact, the crystal structure was not among the conformations within 20 kcal/mol of the global minimum, which all exhibited a similar folding pattern. (A detailed comparison of the generated conformations vis-à-vis the crystal structures is beyond the scope of this paper and will be discussed in a subsequent publication.)

The global minima of the bioactive peptides have no symmetry, but they also show extensive intramolecular hydrogen bonding with the exception of P3, which is the simplest of all. The six bioactive peptides (P1–P6) contain 11, 12, 1, 7, 7, and 15 hydrogen bonds. P4 features two occurrences of an interesting pattern of triple bifurcation, where three functional groups are linked to each other by three hydrogen bonds in a triangular fashion (with each group being hydrogen bonded to the other two). While it is conceivable that important conformations may have been

missed, the generally low number of unique conformations (except P3) suggests that most of the important conformations have been identified.

Before we conclude, we would like to address two important remaining issues. The first concerned the relative CPU requirements of these methods. Our preceding analysis focused exclusively on conformational space coverage and did not take into account the computational cost associated with each of these methods, which ultimately determines their practical utility, particularly for large collections of compounds. To provide the reader with a relative estimate of this cost, we asked each method to generate 100 conformations of the bioactive peptide P4 (MicrocystinLR) using identical parameter settings as those described above and recorded the CPU times required for the generation and the minimization of these conformations. The results are summarized in Table 3. For CAESAR, this second set of calculations was carried out using version 2.5 because our original license had expired and our version 1.5 was no longer available.

Because of licensing limitations and operating system dependencies, it was impossible to run all of the conformation generation methods on the same machine, so the times reported are not directly comparable unless they were carried out on the same or similar CPU. In contrast, all minimizations were carried out on the same Intel Xeon 3.8 GHz processor, so these times are directly comparable.

Five CPU times are reported for each method: the time required to generate the conformations, T_{gen} ; the time required to compute the single-point energy of the raw conformation, T_{min0} ; the time required for 1 minimization step, T_{min1} ; the time required for 10 minimization steps, T_{min10} ; the time required for 100 minimization steps, T_{min100} ; and the time required for up to 1 000 minimization steps, T_{min1000} (though some minimizations converged in fewer steps). In addition to the CPU times, we also provide the median energies of the respective conformations.

Table 3 reveals several interesting trends. SOS is the fastest conformation generation method followed by CAESAR, FAST variants of Catalyst and MacroModel, Rubicon and SPE, MOE, Omega, and finally the BEST variants of Catalyst. However, MOE produced only 5 valid geometries, and MacroModel produced only 58, suggesting that the time it would take them to generate 100 chemically sensible conformations could be much greater (note that the remaining conformations were eliminated because of bad contacts and not because of redundancy, as duplicate checking was turned off for every program).

The conformation generation algorithms which required the longest time also produced the best starting geometries, as manifested by the median single-point energies, E_0 . The "worst" starting geometries were produced by MacroModel, CAESAR, and SOS, which had strain energies greater than 3 000 kcal/mol, but these starting geometries eventually minimized to better minima (see E_{1000}). It is important to note that no method produced 'perfect' geometries, i.e., conformations that were very close to the corresponding local minima. The lowest energy structures were produced by the BEST variants of Catalyst and Omega, which partially minimize the structures during the course of the search using the very same force field that was used for the subsequent refinement (MMFF94s). Despite that fact, the starting

geometries were nearly 100 kcal/mol higher in energy than their nearest local minima, suggesting that parts of the molecule may not have been fully minimized. Indeed, visual inspection of the conformations produced by Omega (which also used MMFF94s minimization during the search and produced relatively low-energy geometries) revealed that the large ring was not sampled at all, only the side chains.

It is also important to note that 100 minimization steps were sufficient to bring the SPE, SOS, Rubicon, and MacroModel geometries lower in energy than the starting geometries of any method evaluated, including CatalystB and Omega. In other words, no method produces perfect starting geometries, and SPE, SOS, Rubicon, and MacroModel arrive at the same level of 'perfection' (or 'imperfection') faster than any other method through partial minimization. In conclusion, these methods are not only substantially more effective in sampling the full range of conformational space but also they appear to be faster as well (though to prove this point conclusively, a more systematic analysis with a broader range of diverse compounds would be necessary).

The second point concerns the potential dependence of our results on the particular force field used for the minimization. Indeed, our choice of MMFF94s may have favored methods which use this force field internally during the course of the search. We believe that this is not the case but have no direct evidence to substantiate this claim. The only argument we can bring forward at this time is that the use of a very large energy window to filter out nonplausible conformations should have eliminated such dependencies. Besides, the most effective methods (those based on distance geometry) did not employ energetics at all. The point, however, is a valid one, and a more thorough comparison against multiple force fields is currently under way and will be presented in due course.

CONCLUSIONS

As the quest for structural novelty intensifies, the need for conformational search methods that work well for many different types of molecules is greater than ever. Macrocycles are of interest for several reasons: they have shown promise as drugs, they represent a viable strategy for exploring novel chemical space unencumbered by intellectual property constraints, and they test the limits of many conformational search algorithms that have seen widespread utility because of their speed, cost, availability, and integration with larger molecular modeling packages. In this paper, we report a rigorous comparison of twelve different conformational search algorithms on three distinct classes of macrocycles and demonstrate an overwhelming advantage for methods rooted in distance geometry. These methods excel in many important respects: they produce the greatest number of unique low-energy conformations, they identify the lowest-energy structure (or some other structure close to it), and they yield the greatest geometric diversity. SOS emerges as the most competitive method among them, particularly when it is combined with conformational boosting. These results, coupled with our previous validation studies on nonmacro-cyclic drug-like molecules, lead us to believe that SPE and SOS are among the best and the most universally applicable conformational sampling methods invented to date, with the latter being preferred because of its superior speed.

Supporting Information Available: Global minima of cycloglycines, cyclodextrins, and bioactive peptides. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- (1) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The exploration of macrocycles for drug discovery - an underexploited structural class. *Nat. Rev. Drug Discov.* **2008**, *7*, 608–624.
- (2) Schröder, E.; Lübke, K. *The Peptides*, Academic Press: New York, 1965–1966; Volumes 1 and 2.
- (3) Cyclodextrins *Chem. Rev.* **1998**, *98* (5), all pages.
- (4) Pedersen, C. J. The discovery of crown ethers. *Science* **1988**, *241*, 536–540.
- (5) Sauvage, J. P.; Dietrich-Buchecker, C. *Molecular catenanes, rotaxanes and knots*. Wiley-VCH: Weinheim, Germany, 1999.
- (6) Lehn, J. M. Supramolecular chemistry - Scope and perspectives. Molecules, supermolecules, and molecular devices. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 89–112.
- (7) Lehn, J. M. Perspectives in supramolecular chemistry. From molecular recognition towards molecular information processing and self-organization. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 1304–1319.
- (8) *Nobel Lectures, Chemistry 1981–1990*; Malmstrom, B. G., Ed.; World Scientific: London, UK, 1992; pp. 411–511.
- (9) Wulschleger, S.; Loewith, R.; Hall, M. TOR signaling in growth and metabolism. *Cell* **2006**, *124*, 471–484.
- (10) (a) Giordani, R.; Regli, P.; Kaloustian, J.; Mikail, C.; Abou, L.; Portugal, H. Antifungal effect of various essential oils against *Candida albicans*. Potentiation of antifungal action of amphotericin B by essential oil from *Thymus vulgaris*. *Phytother. Res.* **2004**, *18*, 990–995. (b) Kauffman, C. A. Clinical efficacy of new antifungal agents. *Curr. Opin. Microbiol.* **2006**, *9*, 483–488.
- (11) Bedin, M.; Gaben, A. M.; Saucier, C.; Mester, J. Geldanamycin, an inhibitor of the chaperone activity of HSP90, induces MAPK-independent cell cycle arrest. *Int. J. Cancer* **2004**, *109*, 643–652.
- (12) Remillard, S.; Rebhun, L. I.; Howie, G. A.; Kupchan, S. M. Antimitotic activity of the potent tumor inhibitor maytansine. *Science* **1975**, *189*, 1002–1005.
- (13) Kanazawa, S.; Fusetani, N.; Matsunaga, N. Bioactive marine metabolites. 45. Cylindramide: cytotoxic tetramic acid lactam from the marine sponge *Halichondria cylindrata* Tanita & Hoshino. *Tetrahedron Lett.* **1993**, *34*, 1065–1068.
- (14) Gunasekera, S. P.; Gunasekera, M.; McCarthy, P. Discodermide: a new bioactive macrocyclic lactam from the marine sponge *Discodermia dissoluta*. *J. Org. Chem.* **1991**, *56*, 4830–4833.
- (15) Shigemori, H.; Bae, M. A.; Yazawa, K.; Sasaki, T.; Kobayashi, J. Alteramide A, a new tetracyclic alkaloid from a bacterium *Alteromonas* sp. associated with the marine sponge *Halichondria okadaei*. *J. Org. Chem.* **1992**, *57*, 4317–4320.
- (16) (a) Gessler, K.; Usón, I.; Takaha, T.; Krauss, N.; Smith, S. M.; Okada, S.; Sheldrick, G. M.; Saenger, W. V-Amylose at atomic resolution: X-ray structure of a cyclodextrin with 26 glucose residues (cyclomaltohexaosaose). *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4246–4251. (b) Nimz, O.; Gessler, K.; Usón, I.; Saenger, W. An orthorhombic crystal form of cyclomaltohexaosaose, CA26·32.59 H₂O: comparison with the triclinic form. *Carbohydr. Res.* **2001**, *336*, 141–153.
- (17) Larsen, K. L. Large Cyclodextrins. *J. Inclusion Phenom. Macrocycl. Chem.* **2002**, *43*, 1–13.
- (18) Ivanov, P. M.; Jaime, C. Insights into the structure of Large-Ring Cyclodextrins through Molecular Dynamics Simulations in Solution. *J. Phys. Chem. B* **2004**, *108*, 6261–6274.
- (19) Kitamura, S.; Isuda, H.; Shimada, J.; Takada, T.; Takaha, T.; Okada, S.; Mimura, M.; Kajiwara, K. Conformation of cyclomaltooligosaccharide (“cyclodextrin”) of dp21 in aqueous solution. *Carbohydr. Res.* **1997**, *304*, 303–314.
- (20) Ovchinnikov, Y. A.; Ivanov, V. T. Conformational states and biological activity of cyclic peptides. *Tetrahedron* **1975**, *31*, 2177–2209.
- (21) (a) Thevenard, J.; Floquet, N.; Ramont, L.; Prost, E.; Nuzillard, J. M.; Dauchez, M.; Yezid, H.; Alix, A. J. P.; Maquart, F. X.; Monboisse, J. C.; Brassart-Pasco, S. Structural and antitumor properties of the YSNSG cyclopeptide derived from tumstatin. *Chem. Biol.* **2006**, *13*, 1307–1315. (b) Terracciano, S.; Bruno, I.; Bifulco, G.; Avallone, E.; Smith, C. D.; Gomez-Palomaa, L.; Riccio, R. Synthesis, solution structure, and bioactivity of six new simplified analogues of the natural cyclodepsipeptide jaspamide. *Bioorg. Med. Chem.* **2005**, *13*, 5225–5239.
- (22) Bach, A. C.; Eyerma, C. J.; Gross, J. D.; Bower, M. J.; Harlow, R. L.; Weber, P. C.; DeCradol, W. F. Structural studies of a family of high affinity ligands for GP I I b/III a. *J. Am. Chem. Soc.* **1994**, *116*, 3201–3219.
- (23) Meyer, B.; Peters, T. NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem., Int. Ed.* **2003**, *42*, 864–890.
- (24) Haubner, R.; Schmitt, W.; Hollzemann, G.; Goodman, S. L.; Jonczyk, A.; Kessler, H. Cyclic RGD peptides containing b-turn mimetics. *J. Am. Chem. Soc.* **1996**, *118*, 7881–7891.
- (25) Freeman, D. J.; Pattenden, G.; Drake, A. F.; Siligardi, G. Marine metabolites and metal ion chelation. Circular dichroism studies of metal binding to *Lissoclinum cyclopeptides*. *J. Chem. Soc., Perkin Trans.* **1998**, *2*, 129–135.
- (26) Kolossváry, I.; Guida, W. C. Low mode search. An efficient, automated computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. *J. Am. Chem. Soc.* **1996**, *118*, 5011–5019.
- (27) Keasar, C.; Rosenfeld, R. Empirical modifications to the Amber/OPLS potential for predicting the solution conformations of cyclic peptides by vacuum calculations. *Fold Des.* **1998**, *3*, 379–388.
- (28) Kahn, K.; Bruce, T. C. Parameterization of OPLS-AA force field for the conformational analysis of macrocyclic polyketides. *J. Comput. Chem.* **2002**, *23*, 977–996.
- (29) Raithby, P. R.; Shields, G. P.; Allen, F. H. Conformational analysis of macrocyclic ether ligand. II. 1,4,7,10,13-pentaoxacyclopentadecane and 1,4,7,10,13-pentathiacyclopentadecane. *Acta Crystallogr.* **1997**, *B53*, 476–489.
- (30) Böhm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243–256.
- (31) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499–2510.
- (32) Good, A. C.; Cheney, D. L.; Sitkoff, D. F.; Tokarski, J. S.; Stouch, T. R.; Bassolino, D. A.; Krystek, S. R.; Li, Y.; Mason, J. S.; Perkins, T. D. J. Analysis and optimization of structure-based virtual screening protocols 2. Examination of docked ligand orientation sampling methodology: mapping a pharmacophore for success. *J. Mol. Graph. Modell.* **2003**, *22*, 31–40.
- (33) Joseph-McCarthy, D.; Thomas, B. E. IV; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins: Struct., Funct., Bioinf.* **2003**, *51*, 172–188.
- (34) Bruccoleri, R. E.; Karplus, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* **1987**, *26*, 137–168.
- (35) Gippert, G. P.; Wright, P. E.; Case, D. A. Distributed torsion angle grid search in high dimensions: a systematic approach to NMR structure determination. *J. Biomol. NMR* **1998**, *11*, 241–263.
- (36) Sadowski, J.; Boström, J. MIMUMBA revisited: torsion angle rules for conformer generation derived from X-ray structures. *J. Chem. Inf. Model.* **2006**, *46*, 2305–2309.
- (37) Chandrasekhar, J.; Sanders, M.; Jorgensen, W. Efficient exploration of conformational space using the stochastic search method: application to α -peptide oligomers. *J. Comput. Chem.* **2001**, *22*, 1646–1654.
- (38) Izrailev, S.; Zhu, F.; Agrafiotis, D. K. A distance geometry heuristic for expanding the range of geometries sampled during conformational search. *J. Comput. Chem.* **2006**, *27*, 1962–1969.
- (39) Agrafiotis, D. K.; Bandyopadhyay, D.; Carta, G.; Knox, A. J. S.; Lloyd, D. G. On the effects of permuted input on conformational sampling of druglike molecules: an evaluation of stochastic proximity embedding (SPE). *Chem. Biol. Drug Des.* **2007**, *70*, 123–133.
- (40) Leach, A.; Smellie, A. combined model-building and distance-geometry approach to automated conformational analysis and search. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 379–385.
- (41) Crippen, G. M. Exploring the conformation space of cycloalkanes by linearized embedding. *J. Comput. Chem.* **1992**, *13*, 351–361.
- (42) Peishoff, C. E.; Dixon, J. S. Improvements to the distance geometry algorithm for conformational sampling of cyclic structures. *J. Comput. Chem.* **1992**, *13*, 565–569.
- (43) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M.; Peterson, M. L. Conformational searching methods for small molecules. II. Genetic algorithm approach. *J. Comput. Chem.* **1993**, *14*, 1407–1414.
- (44) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.
- (45) Wang, C.-S. Efficient algorithm for conformational search of macrocyclic molecules. *J. Comput. Chem.* **1997**, *18*, 277–289.
- (46) Saunders, M. Stochastic search for the conformations of bicyclic hydrocarbon. *J. Comput. Chem.* **1989**, *10*, 203–208.
- (47) Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T. Comparative analysis of protein-bound ligand conformations with respect to Catalyst's conformational space subsampling algorithms. *J. Chem. Inf. Model.* **2005**, *45*, 422–430.

- (48) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499–2510.
- (49) Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. Comparative performance assessment of the conformational model generators Omega and Catalyst: A large-scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.* **2006**, *46*, 1848–1862.
- (50) Boström, J. Reproducing the conformations of protein-bound ligands: a critical evaluation of several popular conformational searching tools. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1137–1152.
- (51) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of automatic three-dimensional model builders using 639 x-ray structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000–1008.
- (52) Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F. Q.; Izrailev, S.; Martin, E. Conformational sampling of bioactive molecules: a comparative study. *J. Chem. Inf. Model.* **2007**, *47*, 1067–1086.
- (53) Böhm, H. J.; Klebe, G.; Lorenz, T.; Mietzner, T.; Siggel, L. Different approaches to conformational analysis: A comparison of completeness, efficiency, and reliability based on the study of a nine-membered lactam. *J. Comput. Chem.* **1990**, *11*, 1021–1028.
- (54) Christensen, I. T.; Jørgensen, F. S. Conformational analysis of six- and twelve-membered ring compounds by molecular dynamics. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 385–394.
- (55) Kawai, T.; Tomioka, N.; Ichinose, T.; Takeda, M.; Itai, A. High-temperature simulation of the dynamics of cyclohexane. *Chem. Pharm. Bull.* **1994**, *42*, 1315–1321.
- (56) Xu, H.; Izrailev, S.; Agrafiotis, D. K. Conformational sampling by self-organization. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1186–1191.
- (57) Li, J.; Ehlers, T.; Sutter, J.; Varma-O'Brien, S.; Kirchmair, J. *J. Chem. Inf. Model.* **2007**, *47*, 1923–1932.
- (58) Zhu, F.; Agrafiotis, D. K. Self-organizing superimposition algorithm for conformational sampling. *J. Comput. Chem.* **2007**, *28*, 1234–1239.
- (59) Luckett, S.; Garcia, R. S.; Barker, J. J.; Konarev, A. V.; Shewry, P. R.; Clarke, A. R.; Brady, R. L. High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *J. Mol. Biol.* **1999**, *290*, 525–533.
- (60) Bewley, A. C.; He, H.; Williams, D. H.; Faulkner, D. J. Aciculitins A-C: Cytotoxic and antifungal cyclic peptides from the lithistid sponge *aciculites orientalis*. *J. Am. Chem. Soc.* **1996**, *118*, 4314–4321.
- (61) Randazzo, A.; Bifulco, G.; Giannini, C.; Bucci, M.; Debitus, C.; Cirino, G.; Gomez-Paloma, L. Halipeptins A and B: Two novel potent anti-inflammatory cyclic depsipeptides from the Vanuatu marine sponge *haliclona* species. *J. Am. Chem. Soc.* **2001**, *123*, 10870–10876.
- (62) Umezawa, K.; Nakazawa, K.; Uemura, T.; Ikeda, Y.; Kondo, S.; Naganawa, H.; Kinoshita, N.; Hashizume, H.; Hamada, M.; Takeuchi, T.; Ohba, S. Polyoxypeptin isolated from streptomycetes: A bioactive cyclic depsipeptide containing the novel amino acid 3-hydroxy-3-methylproline. *Tetrahedron Lett.* **1998**, *39*, 1389–1392.
- (63) Jaki, B.; Zerbe, O.; Heilmann, J.; Sticher, O. Two novel cyclic peptides with antifungal activity from the cyanobacterium *tolypothrix byssoidea* (EAWAG 195). *J. Nat. Prod.* **2001**, *64*, 154–158.
- (64) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17*, 616–641.
- (65) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
- (66) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520–552.
- (67) Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553–586.
- (68) Halgren, T. A.; Nachbar, R. B. Merck molecular force field. IV. conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587–615.
- (69) Agrafiotis, D. K.; Lobanov, V. S.; Salemme, F. R. Combinatorial informatics in the post-genomics era. *Nat. Rev. Drug Discov.* **2002**, *1*, 337–346.
- (70) Agrafiotis, D. K.; Bone, R. F.; Salemme, F. R.; Soll, R. M. System and method for automatically generating chemical compounds with desired properties. United States patents 5,463,564; 5,574,656; 5,684,711; 5,901,069; 6,421,612; 6,434,490.
- (71) Labanowski, J. K. CCL HOME. <http://www.ccl.net/ccs/data/MMFF94s/> (accessed January 1, 2009).
- (72) *Catalyst*, version 4.10; Accelrys, Inc.: San Diego, CA; <http://www.accelrys.com>.
- (73) Smellie, A.; Teig, S. L.; Towbin, P. Poling: Promoting conformational variation. *J. Comput. Chem.* **1995**, *16*, 171–187.
- (74) *MacroModel*, version 9.1111; Schrödinger, Inc.: Portland, OR; <http://www.schrodinger.com>.
- (75) Chang, G.; Guida, W. C.; Still, W. C. An internal-coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.
- (76) Kolossvary, I.; Guida, W. C. Low mode search. An efficient, automated computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. *J. Am. Chem. Soc.* **1996**, *118*, 5011–5019.
- (77) Ponder, J. W.; Richards, F. M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **1987**, *8*, 1016–1024.
- (78) *MOE*, version 2006-08; Chemical Computing Group: Montreal, Quebec; <http://www.chemcomp.com> (accessed September 2009).
- (79) Ferguson, D. M.; Raber, D. J. A new approach to probing conformational space with molecular mechanics: random incremental pulse search. *J. Am. Chem. Soc.* **1989**, *111*, 4371–4378.
- (80) *Omega*, version 1.8.1; OpenEye Scientific Software: Santa Fe, NM; <http://www.eyesopen.com>. (accessed September, 2009).
- (81) *Rubicon*, version 1.0; Daylight Chemical Information Systems, Inc.: Aliso Viejo, CA; <http://www.daylight.com> (accessed September, 2009).
- (82) Crippen, G. M.; Havel, T. F. Stable calculation of coordinates from distance information. *Acta Crystallogr.* **1978**, *A34*, 282–284.
- (83) Agrafiotis, D. K.; Alex, S.; Dai, H.; Derkinderen, A.; Farnum, M.; Gates, P.; Izrailev, S.; Jaeger, E. P.; Konstant, P.; Leung, A.; Lobanov, V. S.; Marichal, P.; Martin, D.; Rassokhin, D. N.; Shemanarev, M.; Skalkin, A.; Stong, J.; Tabruyn, T.; Vermeiren, M.; Wan, J.; Xu, X. Y.; Yao, X. Advanced Biological and Chemical Discovery (ABCD): Centralizing discovery knowledge in an inherently decentralized world. *J. Chem. Inf. Model.* **2007**, *47*, 1999–2014.
- (84) Accelrys, Inc., personal communication.
- (85) Jacob, J.; Gebler, K.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T.; Saenger, W. Strain-induced “band flips” in cyclodecaamylose and higher homologues. *Angew. Chem., Int. Ed.* **1998**, *37*, 605–609.

CI900238A