Macrocycle Conformational Sampling with MacroModel

K. Shawn Watts,[†] Pranav Dalal,[‡] Andrew J. Tebben,[§] Daniel L. Cheney,[||] and John C. Shelley*,[†]

[†]Schrödinger, Inc., 101 SW Main Street, Suite 1300, Portland, Oregon 97204, United States

[‡]D. E. Shaw India Software, Private Limited, Sanali Infopark, 8-2-120/113, Road No. 2, Banjara Hills, Hyderabad 500 034, Andhra Pradesh, India

[§]Bristol-Myers Squibb, 3551 Lawrenceville Road, Princeton, Lawrence Township, New Jersey 08648, United States

[||]Bristol-Myers Squibb, 311 Pennington−Rocky Hill Road, Pennington, New Jersey 08543, United States

S *Supporting Information*

ABSTRACT: Sampling low energy conformations of macrocycles is challenging due to the large size of many of these molecules and the constraints imposed by the macrocycle. We present a new conformational search method (implemented in MacroModel) that uses brief MD simulations followed by minimization and normal-mode search steps. The method was parametrized using a set of 100 macrocycles from the PDB and CSD. It was then tested on a publicly available data set for which there are published results using alternative methods; we found that when the same force field is used (in this case MMFFs in vacuum), our method tended to identify conformations with lower energies than what the other methods identified. The performance on a new set of 50 macrocycles from the PDB and CSD was also quite good; the mean and median RMSD values for just the ring atoms were 0.60 and 0.33 Å, respectively. However, the RMSD values for macrocycles with more than 30 ring-atoms were quite a bit larger compared to the smaller macrocycles. Possible origins for this and ideas for improving the performance on very large macrocycles are discussed.

Crystal
Elongated

Generated conformations
More symmetric
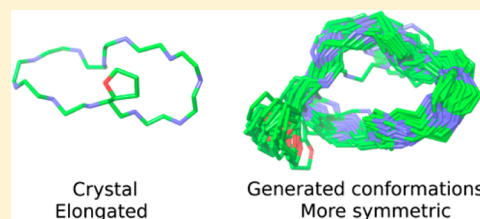
INTRODUCTION

Macrocycle-containing molecules are diverse and have the potential for high affinity, selectivity, and improved pharmacokinetic properties such as permeability.[1] They are present in more than one hundred drugs currently in the market,[2−4] and there are significant efforts underway to develop new macrocyclic drugs.[2−16] Macrocycle-containing molecules present challenges for molecular modeling because the molecules as a whole are often atypically large, containing many rotatable bonds inside of and, in many cases, outside of the macrocycle.

Two recent studies have focused on the challenge of generating three-dimensional structures for both peptidic and nonpeptidic macrocycles from a methodological standpoint. In the first, Bonnet and co-workers[17] evaluated how well a number of conformational search methods, including some that they developed, performed on 19 macrocycles. This well-thought-out study has helped focus the field and provided a publicly available set of molecules to use in benchmarking conformational search methods going forward. We will refer to this set of molecules as the Bonnet macrocycle test set or BMTS. Since the goal was to measure the inherent capability of various search methods, independent of other factors such as force field quality, comparison with experiment was not sought. The calculations were performed in vacuum and all structures were energy minimized after the search using the same software and force field. Performance was judged by the lowest energy conformer identified for each molecule, the number of distinct conformers found and the number of unique pharmacophore triplets within the conformers. In that study some of the authors' own distance geometry methods performed the best, with MacroModel[18] generally out-performing the remaining methods. However, the authors ran MacroModel in a nonstandard manner, effectively without energy minimizations or energy filtering during the search; something that unintentionally degraded its performance relative to using the default setting. Subsequently, Labute developed the LowModeMD method[19] (LMMD), which is a new type of quenched molecular dynamics; a class of conformational searches carried out by repeatedly alternating a short molecular dynamics simulation and an energy minimization and accumulating the conformations visited. A distinguishing feature of this approach is that at the start of each MD simulation the initial randomized atom velocities are scaled atom by atom such that components in directions of high curvature (stiffness) in the potential energy surface are reduced. The velocities are rescaled collectively to recover the target temperature. The net effect of this procedure is to concentrate the thermal energy into motions along directions of low curvature in the potential energy surface. He obtained results superior to those provided by Bonnet and co-workers' best performing method, SPE(SOS2), for the BMTS and since the SPE(SOS2) method was reported as out-performing many others,[17] including the low-mode search method[20,21] in MacroModel, this implied that LMMD was superior to those methods as well.

MacroModel has a wide range of search methods that to some extent can be used in combination, making it possible to take advantage of their various characteristic strengths. Indeed, it has been demonstrated that using multiple search methods, namely a Monte Carlo torsional search method (MCMM)[22−24] and a low mode based conformational search method (LMOD),[20,21] in MacroModel improves the results for a macrocyclic compound over using these methods separately.[25] In addition, the defaults for the search methods in MacroModel have been selected to perform well on either typical small ligands or protein−ligand complexes and may not be appropriate for macrocycles. As a result it should be possible to improve MacroModel's performance on macrocycles, particularly those relevant for drugs, using the appropriate combination of methods and parameters. In this paper we describe a MacroModel-based conformational searching method for macrocycle-containing molecules, characterize its sensitivity to parameter settings, and its performance on a diverse and challenging set of macrocyclic molecules. This new method has been available in MacroModel in one form or another for several releases.

Recently Chen and Foloppe[26] in a thorough and informative study examined the relative advantages of parameter variations on the stochastic and LMMD methods available from MOE[27] and MacroModel's LMOD,[19] LLMOD[28,29] and mixed mode search (which is a combination of MCMM[22,23] and LMOD)[30] methods for drug-like, very flexible and macrocyclic molecules. They also applied MacroModel's macrocycle search method (described herein) to their 30 molecule macrocycle set. Focusing on the results for reproducing experimental macro-cycle geometries as reported in Tables 4 and 5 of that paper, MacroModel's macrocycle search strategy performed better than MOE's stochastic and LMMD methods and Macro-Model's mixed mode and LMOD methods when only default parameters are used for all methods. MacroModel's macrocycle search strategy with default parameters also out-performed all parameter variations explored for MOE's stochastic and LMMD methods. Some parameter variations in MacroModel's mixed mode and LMOD methods performed as well as or slightly better than MacroModel's macrocycle search method with default parameters. However, some parameter variations for the macrocycle method slightly outperformed those methods as well.

Jacobson's group has adapted a method for generating conformations for protein loops[31] to generating conformations on a broad range of molecules[1,11−13] including macrocycles.[1] This method first rapidly generates many candidate ring structures which are then clustered. Representative ring conformation candidates are selected from the clusters for further processing which includes the generation of con-formations for side chains and energy minimization. While some approaches such as Jacobson's may eventually search general macrocycles relatively quickly, macrocycle conforma-tional sampling is challenging and the technology in Macro-Model inherently requires significant CPU time. As such we focus on establishing a high-quality and reliable macrocycle search procedure and explore the speed/quality trade-offs of adjusting various parameters.

The main goal for this work is to reproduce, as closely as possible, the bioactive conformations of macrocycles as represented by crystal structures, as this can be useful for drug design including both ligand-based methods, such as pharmacophore modeling, and structure based methods, such

as docking. Since many crystal structures of macrocyclic compounds are available, quantitative evaluation of the performance of the search method is possible. The idea of performing conformational searches on isolated molecules, even when an implicit solvent model is used, and then comparing the macrocycle conformations with crystal struc-tures has a potential limitation because of the specific, and in some cases, collective interactions between the macrocycle and the protein, for protein-macrocycle complexes, or with other macrocyclic molecules in pure crystals. Intermolecular inter-actions involving formally charged groups at close range may well strongly influence which conformations are favored in these continuum environments limiting how well our searches should be able to match such structures. Thus, bioactive conformations may not coincide with local minima for the isolated or solvated ligand[32−34] and the ensemble of conformations generated may not overlap with those found in crystal structures. Despite these limitations we believe that a key measure of the success of a conformational search calculation that encompasses not only search strategies but also the force fields and solvent models employed, is its ability to reproduce experimentally determined crystal structures well enough to be useful.

In the next section we will briefly describe the BMTS. We then describe our large, diverse set of macrocyclic molecules, as well as how it was selected and split into training and test sets. Following that the baseline macrocycle conformational searching technology within MacroModel and the methodology for evaluating its performance are outlined. We then compare the results obtained using a vacuum adapted version of this method with those published by Bonnet and Labute using the BMTS. Subsequently, we explore the sensitivity of the baseline search method to parameter variations using the training set. Performance is then quantified using our test set of molecules. Finally, we summarize the results and discuss potential directions for further advances.

■ METHODS

Test Set Selection. The BMTS contains 19 structures falling into three classes: polyglycines, cyclodextrins, and more complex peptidic macrocycles. Some of the molecules in the last category have extensive, flexible side-chains attached to the macrocycle. Each of these classes presents distinct challenges for a search technique and thus this is a good general test set. However, only the six nonpolyglycinic peptides closely resemble the predominant types macrocycles found in drugs. Since we want to explore parameter sensitivity and characterize a MacroModel-based[30] search procedure for macrocycles a larger and broader set of molecules consistent with those considered as drug candidates is required.

Our set of macrocycle containing molecules (MCM) has been selected from the PDB[35] (as of March 2011) and the CSD[36] (November 2010 updates). The selection criteria were different for each source. For the PDB, ligands with between 30 and 400 atoms containing rings with more than 7 atoms were considered. Exploratory calculations did not demonstrate a correlation between resolution and ligand strain-energy so we decided against a general ranking and elimination of ligands based upon experimental resolution. The suitability of ligands based upon crystal structure quality measures is examined retrospectively in the discussion of the results. 245 unique ring systems remained after duplicate removal using Canonical SMILES generated by Canvas.[37] Some structures with very
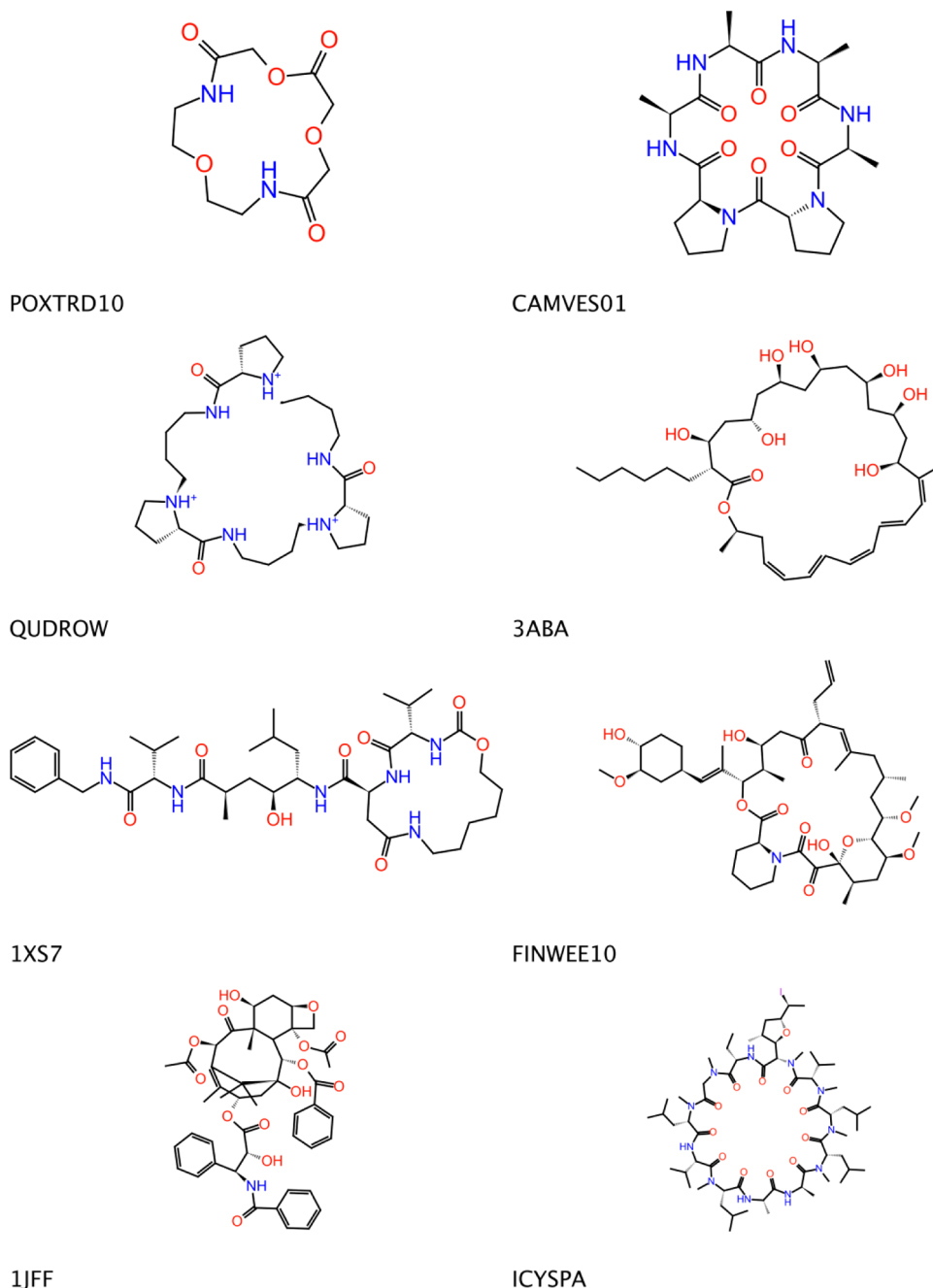
Figure 1. Qualitatively representative molecules from the MCM set. The four letter labels are the codes for molecules from the PDB while the longer labels are the entry identifiers for molecules from the CSD.

poor resolution or inconsistent information between the crystal structure and the description in the corresponding publication were removed. The CSD was searched using Conquest for organic molecules with 3D coordinates using the query NM-C-NM-NM where NM means any nonmetal. This gave 120 105 hits. This collection was filtered by retaining only those molecules in which the smallest set of smallest rings[38] (SSSR) contained at least one ring with more than 8 atoms. As with the PDB duplicate structures were identified using Canonical SMILES generated by Canvas.[37] CSD structures containing cocrystallized species, including counterions were eliminated from consideration. The PDB and CSD structures were then combined into a single set. For cases where crystal structures containing the same or very similar macrocycles were available

in the PDB and the CSD, the former were retained on the presumption that the PDB structures are more likely to be biologically relevant. Structures with little or no flexibility in the rings (e.g., porphyrins) were removed, as well as structures that contain functionalities that are usually ruled out in drug development efforts. Additionally, candidates with heavily represented ring systems, or rigid ring systems, were culled by visual inspection. All culling was done before attempting macrocycle conformational searches on this set of molecules. The resulting MCM set consists of 150 structures, 67 and 83 of which originated in the PDB and CSD, respectively. Selected qualitatively, representative molecules from this set are depicted in Figure 1. Additional information, including the identifying PDB or CSD codes is provided in Table S1. For PDB structures

Table 1. Average, Minimum, and Maximum Values for General Characteristics of the MCM

	training set			test set		
property	average	minimum	maximum	average	minimum	maximum
number of nonring rotors	8.16	0	28	8.04	0	36
number of in-ring rotors	23.9	9	55	23.5	9	48
largest ring size	20.2	8	40	20.6	9	38
number of heavy atoms	49.5	18	98	49.5	20	114
molecular weight	699.1	260.2	1371.9	705.0	284.3	1665.9
sum of magnitudes of formal charges	0.55	0	8	0.37	0	5

this table also includes quality measures for the crystal structure. Characterization of MCM in terms of overall characteristics of the molecules and chemical composition is available in Tables S2 and S3, respectively.

We clustered the MCM with Canvas[37] using *k*-means clustering seeded with 50 clusters and Molprint2D, 64 bit fingerprint as the descriptors. Forty-eight centroid members and two other molecules from structures with PDB codes 1E9W and 2VYP that were too big to generate fingerprints were placed in the test set which was held in reserve to test the performance of the search method. The remaining 100 compounds constituted the training set. Table S1 lists which set contained each molecule. When selecting molecules to include in the test set, care was taken to ensure that the overall characteristics and representation of chemical space of the two sets are similar. Table 1 provides some of the overall characteristics of the training and test sets. In general we will use the largest ring in the SSSR to gauge the size of the macrocycle and simply refer to it as the largest ring. Table S4 also includes information on the hydrogen bond donors and acceptors. Figure 2 depicts the distributions of the largest ring

Figure 2. Distributions of the largest ring size in each molecule in the training and test sets and in the overall MCM set (all) and broken down by number of nonring rotatable bonds. Count refers to the number of molecules matching the qualifications from the bottom and right labels. Each horizontal row of plots corresponds to a range of nonring rotatable bonds, e.g., ≤5 NR indicates that that row of plots is for molecules containing 5 or fewer nonring rotatable bonds.

and nonring rotatable bonds in the MCM as a whole and in the two subsets. Here, 27% and 18% of the molecules in the training and test sets, respectively, have at least one atom with a formal charge with 11% and 8% of the molecules having two or more atoms with formal charges.

To avoid seeding searches with the experimental conformation all structures were converted to SMILES strings and then processed back into 3D structures using LigPrep[39] and Epik.[40,41] This conversion used the command: $SCHRODINGER/ligprep −ismi file.smi −omae file.mae −W e,−ms,1 −epik. This conversion process is challenging for macrocycles, in part because of their size and the large the number of stereo centers in many of these molecules, and manual intervention was required. For instance, an unusually large number of molecules, 9 (6%), were dropped in this process and had to be drawn in by hand in a conformation distinct from the crystal structure. Also, 32% and 12% of the prepared structures required the inversion of one or more of the chiral centers or reversal of the nonamide bond *E/Z* geometries to match the experimental structures, respectively. Additionally, 3% of the molecules had a tautomeric state different from that inferred from the crystal structure and needed to be adjusted. The need for tautomer state adjustments may be due to the −W e,−ms, 1 which was used for convenience to require that Epik produce only one output structure, the one deemed most appropriate for bulk solution. Amide bond geometries were not retained in the SMILES strings so LigPrep would not necessarily reproduce the cis/trans crystal structure conformation for these bonds. Many of these molecules contain amide bonds that may or may not need to be sampled in practical applications depending on what is known about the molecules beforehand. We chose to manually and arbitrarily adjust the amide bond geometries to produce three sets of input structures: the ideal set with amide bond geometries as found in the crystal structure, the all trans set with all amide bonds trans and the cis set with all amide bonds cis. To remove excessive strain present after manual adjustment, energy minimization was performed. The median RMSD values for the largest rings in these seed structures as compared to the crystal structures were 1.09, 1.19, and 1.05 Å for the ideal, trans, and cis sets, respectively.

Search Protocol. Our initial applications of MacroModel's mixed mode conformational search technique, using default parameters, to the BMTS performed similarly well to the SPE(SOS2) method.[17] However, this mixed mode search technique clearly struggled with the larger polyglycines and cyclodextrins in that data set. Given that this method is optimized for sampling more typical drug-like molecules in which the rings are smaller, adjusting the parameters and considering additional search methods seemed like promising ways to improve the results for macrocycles. Extensive trials with variations on methods available in MacroModel resulted in
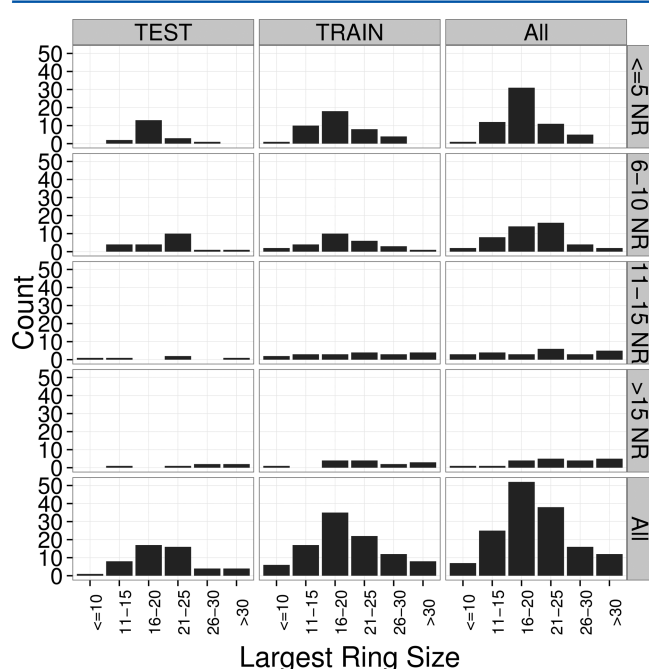
2683

dx.doi.org/10.1021/ci5001696 | *J. Chem. Inf. Model.* 2014, 54, 2680−2696

Table 2. Energy of the Lowest Energy Conformer and Number of Conformers for Molecules in the BMTS Produced by SPE(SOS2), LMMD, and MMBSV[a]

macrocycle	SPE (SOS2)	LMMD	MMBSV20	MMBSV10	method producing the lowest energy conformer	lowest energy relative to MMBSV 20	10
D6	522.64/*20*	522.33/232	521.76/**1626**	521.77/**359**	MMBSV20	0.57	0.56
D8	687.32/*19*	686.70/*26*	684.29/**1116**	684.1/**239**	MMBSV20	**2.41**	**2.6**
D10	863.37/*59*	871.94/121	859.52/**2013**	857.73/**298**	MMBSV20	**3.85**	**5.64**
D12	1045.60/*14*	1045.94/132	1025.94/**1677**	1023.25/**287**	MMBSV20	**19.66**	**22.35**
D14	1212.23/*1*	1207.18/*24*	1179.56/**1151**	1178.81/**200**	MMBSV20	**27.62**	**28.37**
G6	28.74/23	*28.75/*53	28.72/**418**	28.72/*51*	MMBSV20	0.02	0.02
G8	26.2/132	*26.21/*275	26.19/**542**	26.19/*73*	MMBSV20	0.01	0.02
G10	35.12/1335	35.08/**4484**	35.05/1418	35.05/479	MMBSV20	0.03	0.03
G12	39.09/1618	35.13/**5106**	35.72/1226	35.72/203	LMMD	−0.59	−0.59
G14	40.07/1753	40.69/**5594**	38.21/1204	38.99/463	MMBSV20	**1.86**	**1.08**
G16	36.94/300	41.73/**3429**	36.92/442	41.69/176	MMBSV20	0.02	-4.75
G18	49.57/395	44.34/**2314**	34.33/338	34.33/*10*	MMBSV20	**10.01**	**10.01**
G20	51.33/328	42.14/474	46.21/**1059**	43.69/168	LMMD	-4.07	-1.55
P1	26.17/*77*	*7.94/*217	−4.45/**910**	−6.21/102	MMBSV20	**12.39**	**14.15**
P2	157.00/*52*	*145.15/*199	141.90/**1484**	142.81/**357**	MMBSV20	**3.25**	**2.34**
P3	119.83/1276	118.8/**6168**	120.24/2082	118.70/595	LMMD	-1.44	0.10
P4	−146.44/214	−158.05/**1294**	−162.94/1041	−162.70/270	MMBSV20	**4.89**	**4.65**
P5	399.45/*93*	383.10/**980**	380.04/585	380.04/*94*	MMBSV20	**3.06**	**3.06**
P6	32.35/*22*	*13.73/*325	9.93/**2214**	9.93/**557**	MMBSV20	**3.80**	**3.80**

[a]D, G, and P in the first column indicate that the molecule is a cyclodextrin, polyglycine, or a peptide, respectively. For D and G the number of monomers is indicated by the value in the name (e.g. G10 means that this macrocycle is composed of 10 glycines). The results are presented as energies [(kcal/mol)/number of conformations] in each column. The results for the SPE(SOS2) and LMMD methods were obtained from the literature.[17,19] The two right most columns give the energy difference between MMBSV20 and MMBSV10 and the lower of the SPE(SOS2) or LMMD results. Positive values indicate that MMBSV produced a lower energy structure and bold text for energies emphasizes differences larger than 1 kcal/mol. Bold text for the number of conformers indicates which of SPE(SOS2), LMMD, and MMBSV20 found more conformers. Bold text for the number of conformers in the MMBSV10 column indicates that it found more conformers than SPE(SOS2) and LMMD. Searches finding fewer than 100 conformers are italicized. The LMMD method, as published, uses a mechanism to stop searches before 10 000 steps if the search is does not find more conformers in 200 consecutive steps. Values for searches involving less than 10 000 steps are indicated with an asterisk.

better performance. However, they all gave poorer results than we felt could be obtained. Of these, MacroModel's large-scale low-mode search implementation, LLMOD,[28,29] with eigenvector updates during the search with a larger range of atom displacements was one of the top performers. Since LMMD performed reasonably well on the BMTS, we explored variations on quenched MD. We found that a cycle, which we refer to as a quench cycle or QC, composed of 0.5 ps stochastic dynamics at 1000 K followed by another 0.5 ps at a target temperature of 300 K, in turn followed by an energy minimization, performed well. Since using a mixture of conformational searching techniques often makes a search perform well on a broader range of systems[25] and conditions, such as different solvents, we elected to use a compound search composed of 5000 QCs and up to 5000 steps of LLMOD search depending on the number of rotatable bonds. All energy minimizations are carried out more thoroughly than is typical for MacroModel calculations to a RMS gradient of 0.01 kJ/(mol Å) for a maximum of 50 000 iterations. Table S5 gives a more complete listing of parameters for this protocol which we call MacroModel's macrocycle baseline search (MMBS). The development of the MMBS has taken a number of years and employed the current training set, but not the test set, resulting in modifications to the protocol (e.g., switching from 4rDDD to GB/SA).

Modified Search Protocol for Comparison with Other Methods. The Bonnet[17] and Labute[19] studies were performed in vacuum, with infinite cutoffs (i.e., without truncating nonbonded interactions), and used MMFFs[42−48] to provide a standard set of conditions for evaluating the strengths of different search methods. A version of the MMBS employing these conditions, MMBSV, will be used for comparison with their results. All studies reject conformers that have an energy more than an energy window, E_w, above the lowest energy conformer found by that method. Bonnet and Labute used a wider energy window of 20 kcal/mol rather than the 10 kcal/mol window used by the MMBS so we also ran using this window (MMBSV20). We will primarily compare their studies to MMBSV20; however, we will include our results for the 10 kcal/mol window (MMBSV10) for completeness. Bonnet provided the lowest energy conformations in the Supporting Information for all molecules in their study, and Labute provided the corresponding conformations for the P1−P6 molecules. As a check we did energy evaluations on these conformations using a vacuum, effectively infinite cut-offs, and MMFFs. Our energies agreed within the precision of the values reported in their papers, 0.01 kcal/mol, except for P2 where our estimate was 0.02 kcal/mol higher than reported by Labute. So we believe that we are working on essentially the same potential energy surface for the molecules in the BMTS.

Methodology for MMBS Characterization. While the BMTS contains six complex oligopeptides similar to those present in macrocyclic drugs, the other molecules do not resemble molecules typically considered as candidates for macrocyclic drugs. This set is not large enough to optimize MMBS parameters. Instead, we previously used the macrocycle training set to establish useful default values for the search

parameters. Here we use this training set to examine the effect of changing key search parameters in the MMBS, namely:

- the search length
- the relative amount of QC to LLMOD search steps
- the RMSD threshold ($RMSD_T$) values used for distinguishing conformers
- the width of the energy window (E_w) above the lowest energy conformation for retained conformers
- the solvation treatment
- when eigenvectors are updated in LLMOD calculations
- the size of LLMOD search steps
- the use of alternate sampling protocols
- the force field

On average it takes about 2 h on one 2.2 GHz core from a 6 core AMD Opteron(tm) processor with 512 kB cache and 2 GB memory per core to search a molecule in the test or training sets using the MMBS parameters. The calculations for different molecules are independent and thus can be run simultaneously and automatically using the macrocycle sampling script in MacroModel. The MMBS will be subsequently evaluated using the macrocycle test set which has not been used to for MMBS parametrization.

How to evaluate and refine a procedure depends on what is being measured. Our primary goal is to reproduce the conformation of macrocyclic molecules from crystal structures. Since sampling macrocyclic conformations is our primary focus, unless otherwise indicated, reported RMSD values are the smallest for any conformer generated with respect to the crystal structure for the atoms in the largest ring in the SSSR. In some cases we also report the smallest RMSD values for any conformer for all heavy atoms (RMSDRH) with the largest ring superimposed and in others with all heavy atoms optimally superimposed (RMSDH). In all cases distinguishing conformers and thus conformer counts are based upon the RMSDH. From a pragmatic standpoint a search should produce sufficient conformations to reproduce the crystal structures well without also generating a large number of irrelevant conformations in the output of the conformational search. Excessive conformations require computer resources in terms of disk space, computer memory and downstream processing time (e.g., docking each conformation generated or quantum optimizations). In addition, including irrelevant conformations can make it harder to focus on the appropriate ones. While other studies have used measures such as the number of unique three point pharmacophores[17,19,26,49−51] or the ability to reliably find the lowest energy conformer for a given method,[17,19] we feel that the ability to reproduce the crystal structure geometry and the number conformers generated are concise and sufficiently effective measures for our purposes.

■ RESULTS

Table 2 lists the energies of the lowest energy conformer and number of conformers obtained from the search for each member of the BMTS using the SPE(SOS2), LMMD, and MMBSV methods. The results for the SPE(SOS2) and LMMD methods are reproduced from the original papers.[17,19] All methods used a total of 10 000 search cycles except that the LMMD protocol determined that the searches for some molecules were converged enough to terminate the search early.

The number of conformers found differs significantly for all molecules depending on the search method used. Since many of the test molecules contain significant symmetry, convergence criteria for the minimizations differ, and the RMSD threshold may not be the same in the Bonnet set, it is unclear whether conformers are identified and redundant conformers eliminated in a similar manner by the different methods. Another complication with these counts is that they are for all conformations within 20 kcal/mol, or 10 kcal/mol for MMBSV10, of the lowest energy conformation found in each search which means that different absolute energy windows are applied in each study for many of these molecules. This matters because there is a significant increase in the number of conformations at higher energies (see below) so methods that find lower energy conformations for a given molecule could be expected to have a tendency to find fewer conformations, all other factors being equal. With these caveats in mind, MMBSV20 and LMMD find the greatest number of conformations for 11 and 8 molecules, respectively. With a narrower energy window MMBSV10 finds more conformations for 7 of the molecules while LMMD ($E_w = 20$ kcal/mol) does so for 12. In contrast to expectations, across all methods there is a tendency to find fewer conformers for the largest members of the polyglycine and cyclodextrin series than for slightly smaller molecules in the same series. This behavior suggests that all of these methods are producing inadequate conformation subsets for these larger molecules. For some applications, uniform but sparse coverage of conformational space may be adequate (e.g., pharmacophore modeling). For instance, for typical drug-like ligands a fairly small number of conformations is often adequate to reproduce crystal structure conformations reasonably well.[52] Macrocycles, being larger, could be expected to require more conformations. With this in mind we have highlighted the entries in Table 2 for which fewer than an arbitrary threshold of 100 conformations were found. For the smallest polyglycine, G6, all methods, except MMBSV20, produce fewer than 100 conformers, perhaps reflecting an inherently small number of low-energy conformations available for this molecule. The number of molecules for which each method finds fewer than 100 conformations is 0, 3, 3, and 10, for MMBSV20, MMBSV10, LMMD, and SPE(SOS2).

Table S6 summarizes the results for each of the BMTS molecules from 20 separate MMBSV20 calculations each seeded with different initial conformations. For 9 of the molecules in the BMTS the lowest energy conformer was found only in 1 simulation and only 6 molecules had more than 10 searches which found the lowest energy conformation. Given the variation among the energies for the three methods and among different MMBS runs, it seems unlikely that the larger polyglycines, cyclodextrins and the natural products have found the global minimum energy. Focusing on only those instances in which the lowest energy result differs by more than 1 kcal/mol from the next lowest, the counts for the method generating the lowest energy conformer are 9, 2, and 0 for MMBSV20, LMMD, and SPE(SOS2), respectively. The corresponding counts for MMBSV10, LMMD, and SPE(SOS2) are the same. Both the MMBSV20 and MMBSV10 give the lower energies for all four instances where the energy differences are larger than 5 kcal/mol. Table S6 shows that the standard deviation for the lowest MMBS energy is at least 2.75 times smaller than these energy differences for these 4 molecules. The protocol used to terminate the LMMD conformational search was
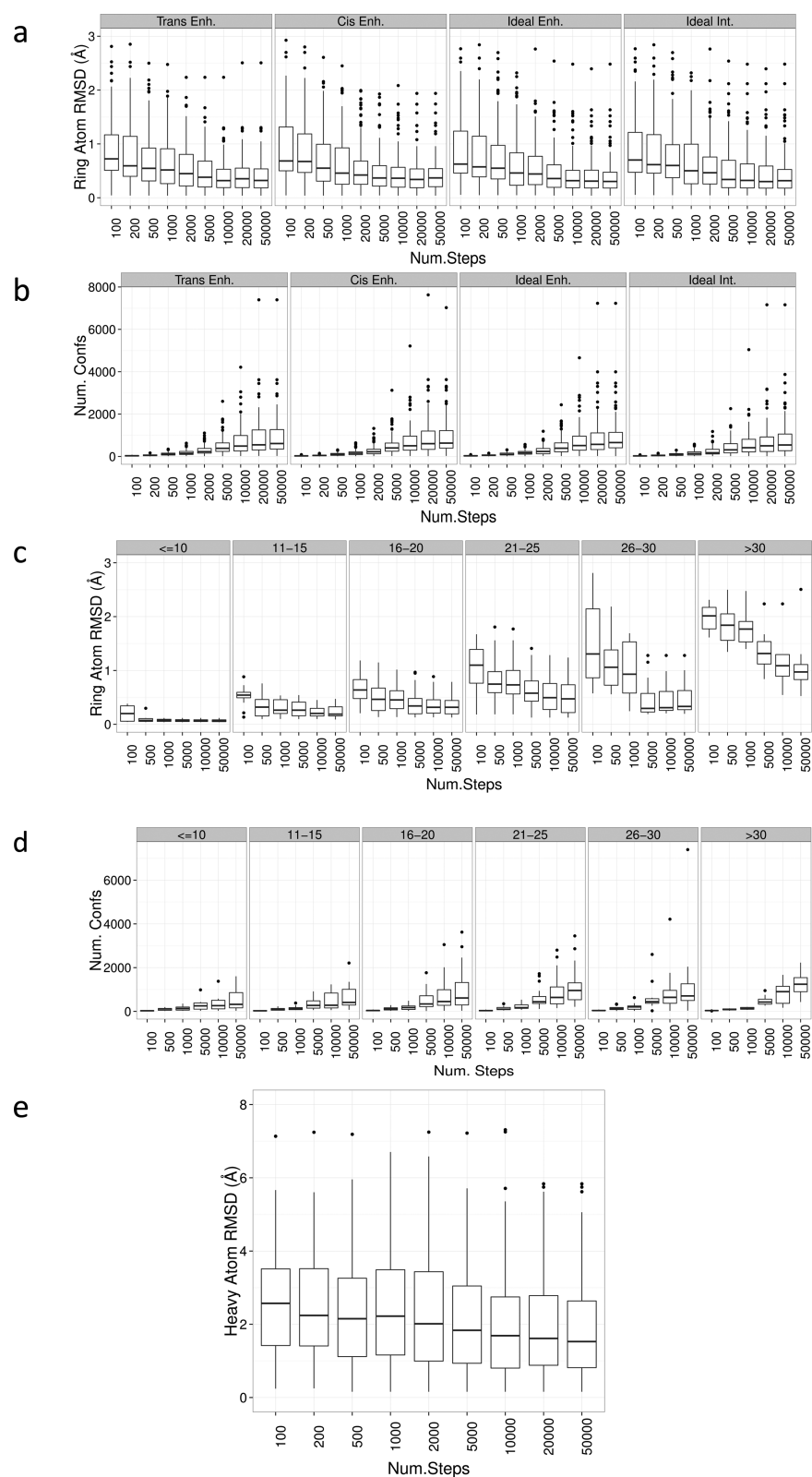
Figure 3. Conformational search results as a function of the number of QC and LLMOD search steps. The number of search steps used for each of QC and LLMOD is given on the horizontal axis. The RMSD and number of distinct conformers are plotted in a and b as a function of starting geometry and bonds considered rotatable. Enh. stands for enhanced searches which sample amide bond conformations while Int. stands for intermediate searches which do not. (c and d) RMSD and number of distinct conformers broken down by largest ring size. RMSDRH, the heavy atom RMSD, with the atoms in the largest ring superimposed is presented in e. The box plot representation of the data is explained in the text. The points for which the number of steps begins with 2 have been omitted in c and d to reduce crowding.

triggered for three of the oligopeptides, namely P1, P2, and P6, all of which had energies significantly higher than MMBSV20 suggesting that this mechanism may prematurely terminate the search for this type of molecule.

Since the MMBS method involves very thoroughly minimizing conformations for up to 50 000 steps or until the RMS gradient drops below 0.01 kJ/(mol Å), it is not clear from the data in Table 2 whether that alone is responsible for its ability to find lower energy conformers. To address this issue we performed two additional calculations for the molecules in the Bonnet test set in which the default minimization parameters were changed to 500 iterations with a threshold RMS gradient of 0.05 kJ/(mol Å) which are MacroModel's default values for most calculations or 3000 iterations with a threshold gradient of 0.05 kJ/(mol Å) which are values used by Chen and Foloppe.[26] This data is presented in Table S7. While the results vary with molecule, the average energy across all molecules drops by 0.81 and 0.37 kcal/mol when using the MMBS minimization parameters as compared to MacroModel's general defaults and Chen and Foloppe's parameters, respectively. These nontrivial shifts in energy demonstrate the MMBS method benefits from more thorough minimizations. However, the average drop in energy between the lower of the two energies from SPE(SOS) or LMMD and the MMBSV20 energy (see Table S7) is much larger at 4.6 kcal/mol demonstrating that only a relatively small portion of this energy difference between these methods can be attributed to the thoroughness of the MMBS minimizations.

Since Chen and Foloppe explored the application of LLMOD to macrocycles[26] we conducted pure LLMOD searches (i.e., use the MMBSV20 method with 10 000 LLMOD and 0 QC steps) to determine whether the QC cycles improve the performance of the MMBSV20 for the BMTS. The results in Table S8 indicate that pure LLMOD calculations are somewhat less effective since they find the lowest energy conformer less often and the average energy of the lowest energy conformers is 1.58 kcal/mol higher.

Based upon these results MMBSV20 and MMBSV10 out-perform the SPE(SOS2) and LMMD methods for the BMTS in terms of finding lower energy conformers. MMBSV20 more reliably produces at least 100 conformers than the other methods. MMBSV20 and LMMD produce the most con-formers for roughly the same number of molecules and between them account for all of the largest conformer counts. The MMBSV10 method performs essentially as well as the MMBSV20 method based upon its ability to find low-energy conformers however it does find significantly fewer conformers.

MMBS Characterization. RMSD values and the number of conformers as a function of the number of search steps for each of the QC and LLMOD search methods keeping a 1:1 ratio of QC to LLMOD steps are plotted in Figure 3 (e.g., in these plots a value of 5000 corresponds to 5000 steps of QC and 5000 steps of LLMOD for a total of 10 000 steps). The results are depicted in this, and subsequent figures, using whisker box plots[53] which are used to represent the spread of the data and identify outliers. Figure S1 contains an annotated box plot. Briefly, for each data point the central horizontal line indicates the median value, the bars show the first quartile on either side of that (essentially the median of the values above or below the overall median). The interquartile range (IQR) is defined as the difference between the values for the third and first quartiles. The highest point within 1.5 IQR above the third quartile and the lowest point within 1.5 IQR below the first quartile are used

to define the limits of the "whisker" lines extending out from the boxes while data values outside this range are considered outliers and appear as distinct points.

Macrocycle searches are normally conducted with "extended" selection of bonds to search which includes sampling amide bonds. We also include results for a search with an "intermediate" setting which does not vary amide bond geometries on the ideal set as a reference. The RMSD (Figure 3a) and number of conformers generated (Figure 3b) are insensitive to the initial amide bond geometry and sampling method. Indeed, the smallest paired t-test[53] p-value for the RMSD between the trans with "enhanced" sampling and the other three choices is 0.79 indicating that the differences in the averages are not statically significant. For the remainder of this paper only the results with the starting structures set to trans with "enhanced" sampling (sample amide bond geometries) will be discussed since this mimics a common usage scenario in which users do not know the correct amide bond geometries, so they are arbitrary and the search must explore these degrees of freedom.

Starting from the shortest searches the RMSD steadily decreases up until about 10 000 search steps total and remains relatively unaffected for additional search steps. The largest paired t test p-value for the RMSD between the 10 000 step total search and the other search durations is 0.0082 indicating the differences among these search levels are statistically significant. The number of conformers found continues to increase with additional search steps. RMSDRH (Figure 3e) also improves up to about 10 000 steps total where the median value is 1.84 Å. This value is much larger than the median RMSD found for just the largest ring 0.38 Å. Part of this is due to only using the ring atoms for superimposing conformers, however, the median RMSDH value, 0.91 Å is still much larger than that for just the largest ring. Reasons for this large difference include: the set of atoms in the ring is smaller and inherently generally more compact, and the large flexible side chains present in many of the molecules are more influenced by the environment than the rings are. Our searches of isolated, solvated molecules typically will not mimic the effect of the crystal packing environment suggesting that macrocycle searching and side-chains searching might be separated with the latter determined in in situ calculations such in a docking for protein−ligand complexes.

Figure 3c breaks down the RMSD results by largest ring size while Figure 3d does the same for number of conformers found. For rings containing up to 20 atoms the results are essentially converged by 10 000 steps total. For larger rings, particularly those containing more than 30 atoms, using more steps helps somewhat.

We also varied the percent of the search cycles that are QC while keeping the total number of search steps fixed at 10 000. The overall results for the RMSD and number of conformations are given in Figure S2 while Figure S3 provides a breakdown of the RMSD as a function of ring size and the % QC. While the median RMSD is not sensitive to this parameter, it does degrade somewhat above 70% QC. As well, the number of outliers noticeably increases below 50% QC and the performance of the search method for rings larger than 25 atoms systematically degrades. Recall that pure LLMOD MMBSV20 searches also performed poorer than the corre-sponding searches with 50% QC cycles.

Decreasing the threshold RMSD value ($RMSD_T$) for distinguishing conformations from 1.25 to 0.25 Å generally
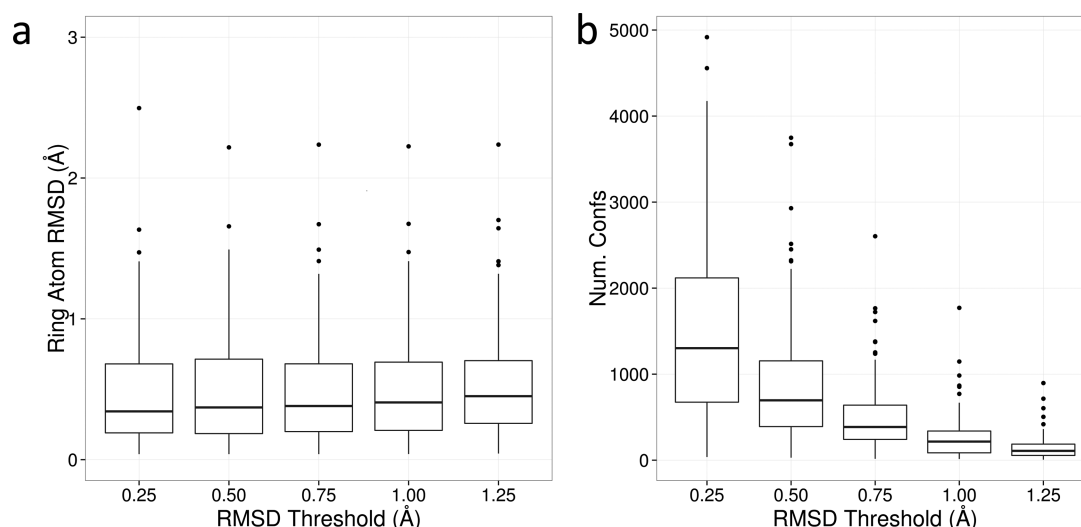
Figure 4. Influence of the RMSD threshold used to distinguish redundant conformers on the ring atom RMSD for the conformer closest to the crystal structure (a) and the number of distinct conformers generated (b).
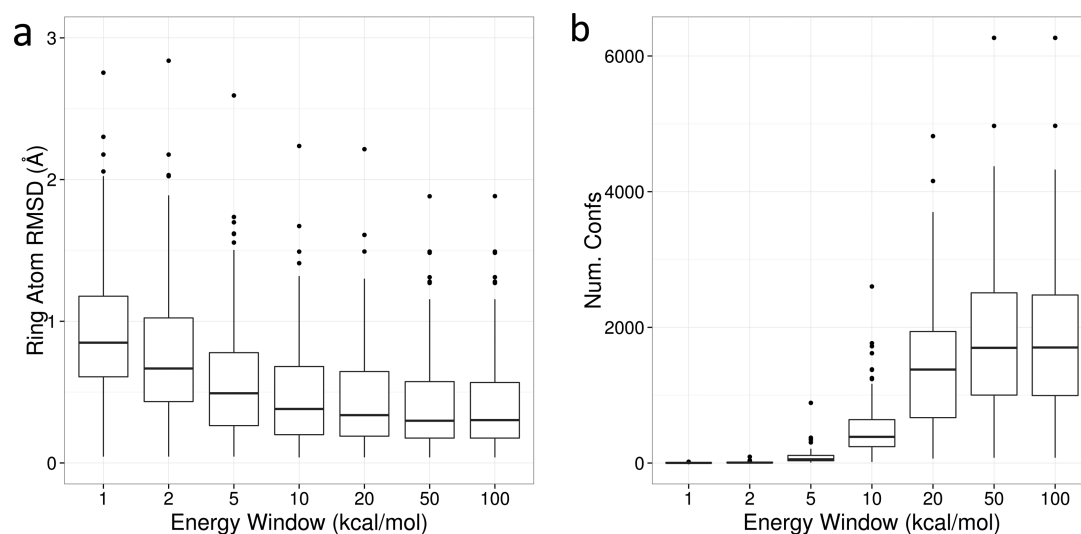

Figure 5. Influence of the size of the energy window, E_w, on the RMSD (a) and number of distinct conformers (b).

decreases the ring atom RMSD (see Figure 4a), however, the trend is fairly weak. The paired t test p-values for the RMSD for an $RMSD_T$ value of 0.75 Å versus 1.25, 1.0, 0.5, and 0.25 Å are 0.000 88, 0.089, 0.24, and 0.48, respectively, indicating that the 0.75 Å results are statistically significantly different than those for 1.25 Å, marginally different than 1.0 Å and not significantly different than those for 0.5 or 0.25 Å. On the other hand, as shown in Figure 4b, the number of distinct conformations produced (N_C) increases exponentially. Figure S4 and S5 plot $\ln(\langle N_C \rangle)$ versus $RMSD_T$ broken down by largest ring size and overall, respectively. The term $\ln(\langle N_C \rangle)$ is a nearly linear function of $RMSD_T$ for the data set as a whole:

$$\ln(\langle N_C \rangle) = -2.42 RMSD_T + 7.61, \quad R^2 = 0.999$$

While the specific coefficient values may vary with macrocycle set, this equation can provide a useful guide as to roughly how many conformations to expect and thus the disk space and downstream processing requirements at a given resolution. Borodina et al.[54] provided a function giving the logarithm of the number of conformers for a given molecule at a specific resolution to the logarithm of the number of conformers at a

different resolution taking into account measures of flexibility such as the number of rotatable bonds. However, their relation included a term involving the logarithm of the $RMSD_T$. The reference MMSB parameters with a threshold of 0.75 Å gives an average RMSD value of 0.50 Å (median 0.38 Å) while for a threshold of 0.25 Å the average RMSD value is 0.49 Å (median 0.34 Å) despite the corresponding approximately 3-fold increase in the number of conformations from 503.7 to 1513.5. These results illustrate that increasing the number of conformers can have distinctly diminishing returns when the primary goal is to reproduce experimental conformations. The fact that the these two very different resolutions give similar results suggests that this method has effectively reached the limit in its ability to obtain agreement between experimental and generated conformations given the limits of the force field, the use of a solvated environment to mimic crystals, and the uncertainty in the atomic positions as determined by experiment. This limit is quite low.

Increasing the energy window, E_w, from 1 to 10 kcal/mol steadily reduces the RMSD (Figure 5a). Further increases do not strongly affect the median RMSD; however, we observe
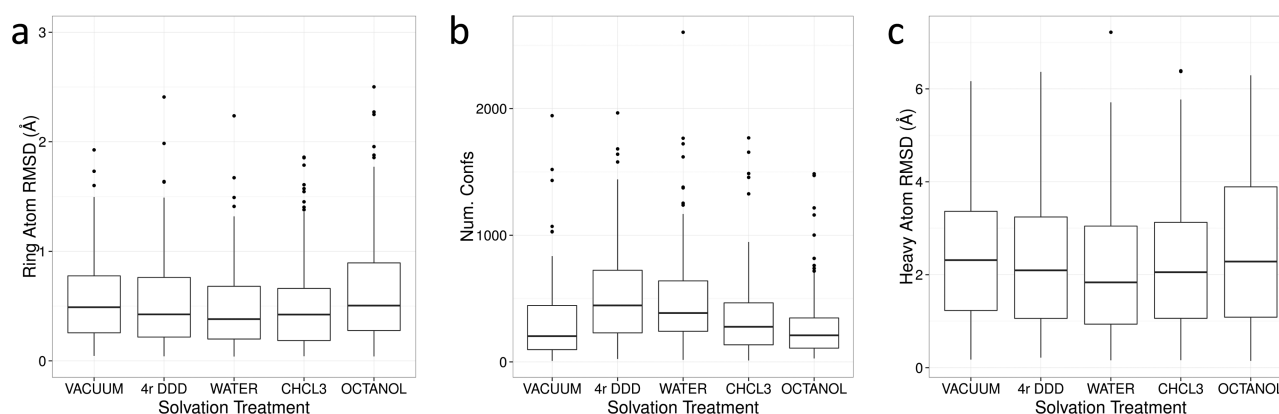
Figure 6. Influence of solvation treatment. 4rDDD stands for distance dependent dielectric given by $\varepsilon_r = 4r$ where ε_r is the relative permittivity and r is the distance while GBSA stands for the GB/SA solvation model for water as implemented in MacroModel. Other solvation treatments supported in MacroModel, vacuum, CHCl$_3$, and octanol were also evaluated. (a) RMSD. (b) Corresponding number of distinct conformations. RMSDRH, the heavy atom RMSD, with the atoms in the largest ring superimposed is presented in c.
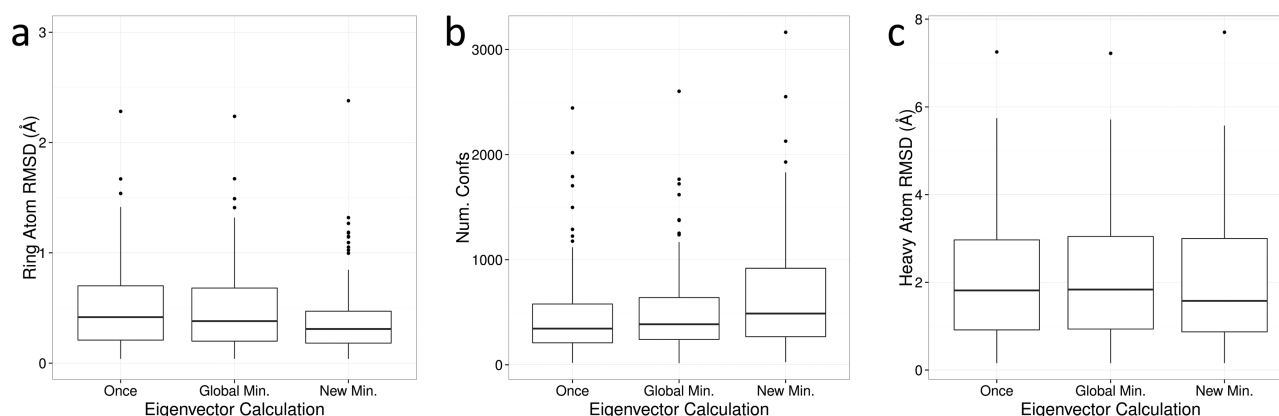
Figure 7. Influence of eigenvector recalculation frequency on search results. "Once" means that the eigenvectors are calculated for the input conformer only and reused in all subsequent LLMOD search steps unchanged. "Global Min." means that the eigenvectors are updated whenever a new conformer with an energy lower than all conformers previously found is generated. "New Min." corresponds to regenerating the eigenvectors in each search step using the conformer that will be perturbed to generate a new conformation. (a) Ring atom RMSD. (b) Number of conformations generated. (c) RMSDRH.

somewhat improved results for outliers. The largest paired t test p-value for the RMSD between the 10 kcal/mol E_w setting and the other settings is 0.014 indicating that the 10 kcal/mol results are statistically significantly different than those for the other settings. The number of conformations increases dramatically as E_w increases (Figure 5b) in a roughly sigmoidal pattern that is similar for all ring sizes except that it saturates somewhat earlier for smaller rings (not shown). In general the ability of conformational searches to match experimental crystal structures improves with larger energy windows.[26,49,55] While this observation makes sense, including high energy conformers should be justified since they may prove problematic in downstream analysis. For instance, for molecules that have a low energy conformer that matches the crystal structure well, including a high energy conformer that matches it better adds little value and may be detrimental because such high energy conformers likely have local distortions that are unacceptable. The value of including higher energy conformers should be noticeable in how much they improve the results for molecules for which a poor match to the crystal structure would otherwise be obtained. For molecules with an RMSD greater than 1 Å with $E_w = 5$ kcal/mol, the RMSD improves by 0.28 Å on average with 25% of these molecules dropping below the 1 Å

threshold with $E_w = 10$ kcal/mol. When E_w is increased from 10 to 20 kcal/mol, corresponding improvement for the molecules with RMSD values greater than 1 Å was 0.09 Å and the RMSD for 16% of these molecules drop below 1 Å. For molecules that have RMSD values less than 1 Å the average energy for the lowest energy conformer was 3.34, 6.66, and 8.87 kcal/mol for E_w values of 5, 10, and 20 kcal/mol, respectively. For the $E_w = 20$ kcal/mol window 12.5% (11 out of 88) of the structures with RMSD values less than 1 Å have energies 15 kcal/mol or more above the lowest energy conformer and all but one of those had an RMSD less than 1 Å with a 10 kcal/mol window. While there is no definitive choice for E_w, 10 kcal/mol window seems like a good compromise. It reproduces crystal structures nearly as well as larger windows without including as many high energy conformers for molecules that have a low energy conformer with a small RMSD value. Energy $E_w = 10$ kcal/mol produces significantly smaller RMSD values than for smaller thresholds and RMSD values nearly as good as those for larger windows with significantly fewer conformers. This window is larger than that commonly used for searches of typically drug-like molecules using MacroModel (5 kcal/mol) which is consistent with the idea that strain energy increases with the number of rotatable bonds.[26,32,33,55,56]

GB/SA implicit solvation calculations[57−63] parametrized for water are more computationally expensive, taking about 4 times longer (7.8 h on average per molecule), as compared to using a distance dependent dielectric in which the relative permittivity, ε_r, is given by $\varepsilon_r = 4r$ where r is the distance between the interacting atoms (4rDDD). The RMSD values for these two solvation treatments are statistically significantly different (paired t test p-value = 0.024). Figure 6 shows that RMSD and RMSDRH improve from 0.42 to 0.38 Å and 2.09 to 1.84 Å when GB/SA water is used rather than 4rDDD while the number of distinct conformations generated decreases by about 16%. Similar improvements were noted by Chen and Foloppe.[26] This suggests that GB/SA water may be a better choice in many cases especially when the computer time for processing fewer conformations in subsequent calculations, such as docking, is taken into account. Vacuum and GB/SA octanol environments perform the worst with poorer RMSD values and a relatively small number of conformers (paired t test p-values for the RMSD versus GB/SA water < 0.02). The RMSD and RMSDRH values for GB/SA CHCl$_3$ are similar to those for 4rDDD with significantly fewer conformers. The RMSD values for GB/SA CHCl$_3$ as compared to GB/SA water are not statistically significant different (paired t test p-value = 0.33) while the former produces fewer conformers.

LLMOD search moves have several parameters that can affect results, including how often the computationally demanding updates to the eigenvectors are done and the size range for the changes in atom coordinates. We tested three choices for updating the eigenvectors: only using the eigenvectors from the input structure, updating the eigenvectors each time a new global minimum (a conformer with an energy lower than all previous conformations) is found and updating the conformers for each search step. The differences in the RMSD values between these three settings are statistically significant (largest paired t test p-value < 0.0004). Figure 7 shows that updating the eigenvectors for each search step gives the best RMSD (median 0.31 Å) and RMSDRH (median 1.58 Å), and produces more conformers (median 488). Updating the eigenvectors for each new global minimum, the default in MSB, reduces the median RMSD slightly and gives more conformers as compared using the original eigenvectors. Updating for each new global minimum is only 6% slower than using the original vectors and is more than 3 times faster on average than updating every step. Chen and Foloppe also found that using the original eigenvectors produced poorer results than updating more frequently.[26] In an LLMOD search step the atomic displacements for the selected eigenvector are scaled such that the atom that moves the most is displaced by an amount randomly chosen between specified upper and lower bounds. The RMSD and the number of conformations identified are not very sensitive to these limits as long as the upper bond is increased to between 18 and 100 Å which is significantly larger than MacroModel's default of 6 Å (see Figure S6).

We performed a number of other explorations, which for brevity we will not describe in detail. These include the following:

1. The MMFF94s, OPLS2005, and OPLS2.1[64] force fields (results not shown) give similar quality results for these macrocycles despite work showing that the use of OPLS2.0 gives better results for solvation free energies.[65]

2. Reversing the order of the application of the methods, so that LLMOD is applied before QC sampling, did not significantly affect the results.

3. Eliminating the 0.5 ps MD simulation at 300 K (but not the 1000K simulation) in the QC only marginally decreased the overall CPU time. This change did not significantly affect the RMSD values for macrocycles containing fewer than 20 atoms in the largest ring. For macrocycles containing larger rings, performance degraded. For instance, for rings containing 26−30 atoms, the median RMSD rose from a value of 0.3 Å to 1.25 Å, and for rings containing more than 30 atoms, it increased from 1.4 to 1.6 Å. These results suggest that the 300 K window allowed the system to, at least sometimes, relax into a different, more appropriate, local minimum than sampled at 1000 K for the larger rings.

Table 3 summarizes the results for parameter variations and indicates what variations may help either speed up the searches or provide better results. For instance, increasing the number of search steps to 20 000 and regenerating the LLMOD eigenvectors for each step will slow the calculation down by about a factor of 6 while generally performing better for larger macrocycles.

Application to the MMBS Test Set. The results from searches of the training and test sets using the MMBS parameters are summarized in Table 4 (median and averages of various RMSD measures) and Table 5 (bioactive conformation recovery rates based upon RMSD measures). The results from these tables and from plots of the performance broken down by largest ring size in Figure 8, indicate that the search results are generally of similar quality for the training and test sets with values for the test set being somewhat poorer, especially for larger rings. The largest systematic differences between the sets are in the averages of the various RMSD values. Much of this difference can be attributed to the inclusion of two large molecules, which could not be clustered, in the test set; specifically 1E9W (2 large macrocycles containing 27 and 26 atoms with 20-nonring rotors) and 2VYP (38 atoms in the largest ring with 36 nonring rotors). If the results for those two molecules are excluded the average RMSD values for the test set drop close to those for the training set. Our recovery rates for bioactive conformations (Table 5) are somewhat lower than those obtained by Chen and Foloppe[26] because our set includes some larger molecules. Even so the recovery rates are high with over 80% of the largest rings having RMSD values less than 1 Å.

■ DISCUSSION

In Figure 8a and c, the RMSD values for the ring atoms and the heavy atoms are quite large for rings containing 30 or more atoms. Figure 8b shows that the number of distinct conformers generated for the molecules containing the largest rings (25 or more atoms) does not significantly increase relative to that for smaller rings. A similar pattern of larger molecules giving similar or fewer conformers is also apparent for the BMTS set in Table 2 across the various programs. It seems doubtful that the number of conformers would not generally increase with molecular size, particularly within a series of related molecules such as the cyclodextrins and polyglycines in the BMTS. Two possible causes for this behavior are the following:

1. The search methods tend to revisit conformations similar to those visited before as the extent of conformation space increases.

Table 3. Summary of the Sensitivity of Search Results to Parameter Variation

search feature	summary	action/recommendation
number of search steps	10000 steps works well for most molecules; 20000 steps can help for more difficult cases	continue to use 10000 steps as the default; recommend using 20000 steps for difficult cases
%QC	results are relatively insensitive to %QC however 30–50% works best	continue to use 50% as the default
RMSD threshold for distinguishing conformers	noted a relatively weak dependence of RMSD for atoms in the largest ring while the number of conformations increases exponentially as the RMSD threshold is reduced	continue to use 0.75 Å as the default
energy window, E_w	values less than 10 kcal/mol significantly degrade RMSD values while values larger than this do not improve the results much; the number of conformers increases significantly as this value is increased	continue to use 10 kcal/mol as the default
solvation treatment	GB/SA water solvation reproduces ring conformations more accurately with fewer conformers than a distance dependent dielectric (4rDDD); the next best solvation treatment; GB/SA CHCl3 produces results statistically similar to GB/SA water with fewer conformers	use GB/SA water or CHCl3 except in cases where faster processing is important, in which case use 4rDDD; GB/SA CHCl3 may be a useful alternative to GB/SA water
LLMOD eigenvector recalculation	recalculating the eigenvectors when a new global energy minimum is found to produce better results than not updating the eigenvectors; updating the eigenvectors each step works best, but requires triple the CPU time	update the eigenvectors for each new global minimum; CPU resources permitting update in each search step
limits on the size of LLMOD search steps	using 3 Å for the lower bound and 18 Å for the upper bound works well	continue to use 3 Å for the lower bound and 18 Å for the upper bound

Table 4. Median and Average Values for Key Search Measures for the Training and Test Sets[a]

property	training		test	
	median	average	median	average
RMSD (Å)	0.38	0.50	0.33 (0.33)	0.60 (0.49)
RMSDRH (Å)	1.84	2.15	1.74 (1.67)	2.36 (2.09)
RMSDH (Å)	0.91	1.38	1.12 (1.08)	1.64 (1.41)
number of conformers	386	503.7	357 (348.5)	478.1(479.3)

[a]Test set values in brackets exclude the two molecules deliberately included in the test set because they could not be clustered, namely, 1E9W and 2VYP.

Table 5. Reproduction of Crystal Structure Geometries[a]

set/measure	0.5 Å	1.0 Å	1.5 Å	2.0 Å
test/RMSD	62%	86%	92%	94%
train/RMSD	67%	88%	98%	99%
test/RMSDH	18%	44%	64%	78%
train/RMSDH	19%	54%	68%	79%
Chen and Foloppe[26]/RMSDH	37%	77%	93%	97%

[a]The values given are the percent of molecules with at least one conformer having an RMSD less than a specified threshold.

2. The potential energy surfaces explored (vacuum or implicit solvent) systematically biases conformational space away from the crystal structure conformation as the molecules grow larger.

Methods such as SPE(SOS2) and LLMOD are designed to generate very different conformations for large molecules so the first explanation seems unlikely. We examined the poorest performing molecules visually anticipating that either strong interactions among highly charged and polar groups or alternately excessive clumping of large hydrophobic groups would be apparent. No consistent pattern stood out. Examination of distributions of energy components was similarly uninformative. To further explore the idea that the potential energy surface might be selecting conformations different that than those present in crystal structures we calculated a number of features including radius of gyration, moments of inertia, eccentricity and solvent accessible surface area among others to characterize molecular extent and shape. Of these eccentricity, e, using the definition:

$$e = 1 - (I_{min}/I_{ave})$$

where I_{min} and I_{ave} are the smallest and average of the principal moments of inertia seemed informative. The value of e is 0 for a spherical top molecule, 1/3 for a ring, and 1 for a linear molecule. Plots comparing the eccentricity of the conformations generated versus that for the reference crystal structures for the molecule as a whole, Figure 9a, and the largest ring within a molecule, Figure 9b, show that in all cases the median eccentricity for the generated conformations is smaller than that for the experimental structures with the exception of the eccentricities of the least challenging macrocycles; those containing 15 or fewer atoms. The median eccentricity of the molecules is fairly independent of macrocycle size at around 0.4 except for the molecules containing smaller macrocycles where it is 0.6. In contrast the eccentricity of the largest ring in each molecule steadily increases with the size of the ring, indicating the rings are becoming more elongated as they grow larger. The trend toward increasing eccentricity with increasing ring size is
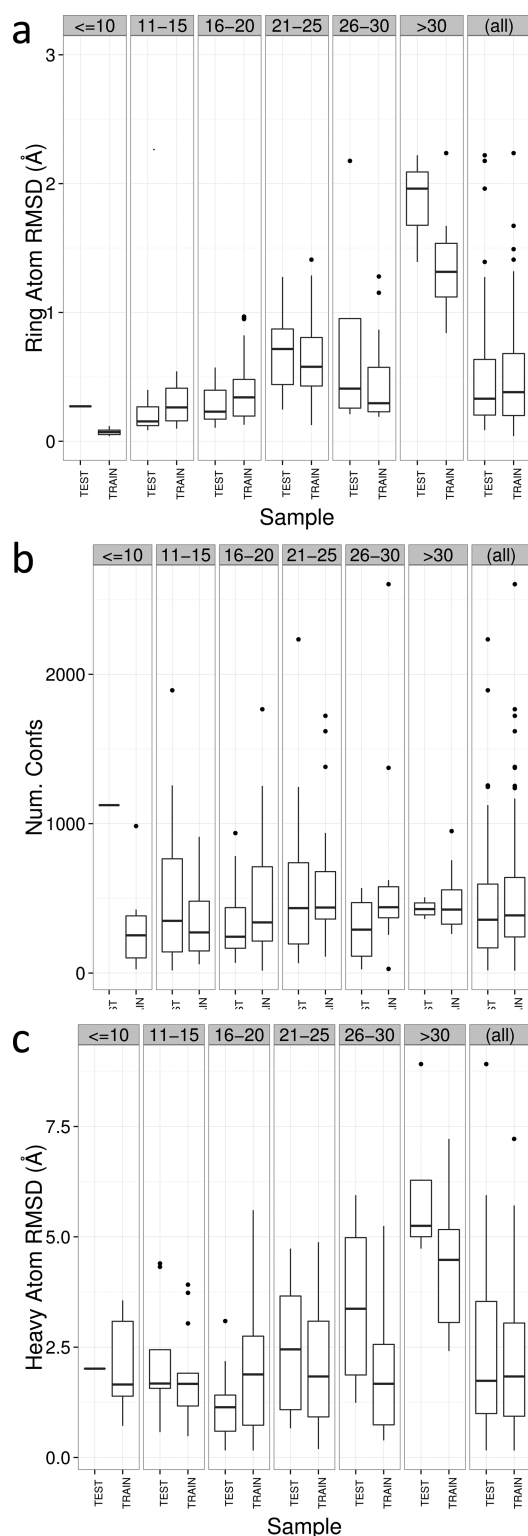
Figure 8. Comparison of results for the training and test sets using the MMBS parameters overall and as a function of the largest ring size. (a) Ring atom RMSD. (b) Number of conformations generated. (c) RMSDRH.

conformations, respectively. As well, the Welch Sample t test indicates that it is highly unlikely (less than 1 chance in 10 000) that the crystal structure and ensemble of conformations have the same mean value for rings containing 16 or more atoms. For rings containing more than 25 atoms 84% of the generated conformations have smaller values for e than its median value for the crystal structures. This suggests that the poorer results for larger rings may be due to the generated ensemble containing conformations that predominantly have the wrong shape for the macrocycle.

Table 6 provides information on the five molecules for which the RMSD is the largest. All of these molecules have rings containing at least 27 atoms. This table includes a comparison of eccentricities of the largest ring in the crystal structure with that for the lowest energy conformer. In 4 of the 5 cases the eccentricity of the largest ring is significantly larger for the crystal structure reflecting the fact the ring conformations in the lowest energy conformers tend to be more compact. The side chains in the generated conformers tend to lie closer to the macrocycle as compared to the more extended crystal structures. Overall these structures are consistent with the potential energy function biasing the results away from crystal structures in a manner that is quite noticeable for larger rings and larger molecules.

In selecting molecules from the PDB for use in this study we did not utilize measures of the quality of the crystal structure except where the resolution was very poor. Recently more effort has gone into validating crystal structures[67−70] and providing useful quality measures to guide the use of such structures. Much of this is directed toward characterizing protein features as opposed to ligands, salts, water, etc. present in the crystal structure. The RSR (Real Space R) and RSCC (Real-Space density Correlation Coefficient) measures can be used to characterize the quality of portions of the structure and have been suggested as being useful for characterizing the ligands[67,69,70] with tentative thresholds of RSR < 0.2 and the RSCC > 0.9. for acceptable structures for the ligand quality.[70] Table S9 lists the 10 ligands originating from PDB structures that violate these conditions along with the RMSD and RMSDH values for the nearest conformation generated. The average largest ring size is nearly the same as that for the whole MCM set (20.3). Nine out of the 10 molecules have RMSD values less than 1 Å and the average RMSD value for these structures (0.57 Å) is only marginally larger than that found for the whole MCM set (0.50 Å). The average RMSDH value for these structures (1.57 Å) is also only marginally larger than that for the whole set (1.47 Å). None of these 10 molecules are among the 3 molecules from the PDB molecules listed in Table 6 for which the MMBS performs the poorest. See the Supporting Information for more information. Retrospectively, the evidence does not support the use of these RSR and RSCC thresholds for selecting macrocycle containing structures from the PDB. As well, for the 20 structures (Table S10) that were selected from the PDB that did not have electron density information and thus RSR and RSCC values were unavailable, the average RMSD value for the closest conformation generated by the MMBS is 0.37 Å while the corresponding value for all of the MCM is 0.50 Å. This result suggests that even though these structures lacked key information used to evaluate the quality of the crystal structure information this was not problematic in the current study.

For cases when macrocycles were present in both the CSD and PDB we selected structures for macrocycles from the PDB,
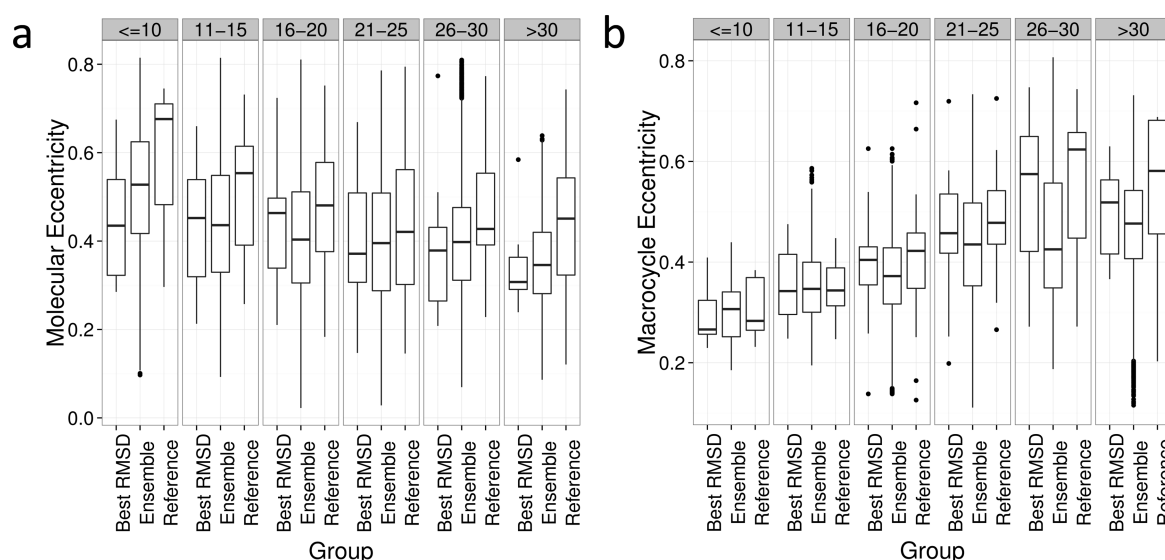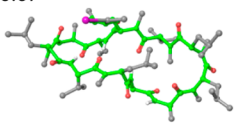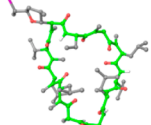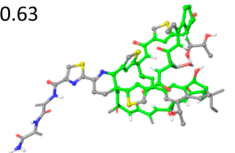
weaker for the generated conformations. The Welch Sample t test[66] indicates that it is extremely unlikely, 5 parts in 10^{10} and 2 parts in 10^{16} that the average eccentricities are the same for rings containing 20 or few atoms as compared to rings with more than 20 atoms, for the crystal structure and generated

2692

dx.doi.org/10.1021/ci5001696 | J. Chem. Inf. Model. 2014, 54, 2680−2696
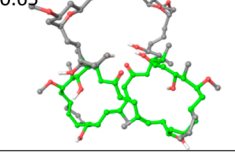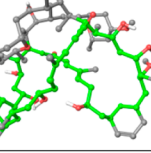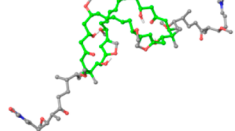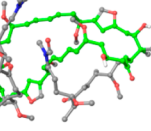
Figure 9. Box plots of the eccentricity of the generated conformations versus the experimental structures. The eccentricities from the crystal structure (Reference), the macrocycle conformation closest to the crystal structure (Best RMSD), and for the distribution of conformations (Ensemble) are plotted by the largest ring size for the molecules as a whole (a) and just for the largest ring in each macrocycle (b).

Table 6. Crystal Structure As Compared to the Lowest Energy Macrocycle Conformation for Five Molecules with the Largest RMSD Values between Any Generated Structure and the Crystal Structure

Code/ Set	Best RMSD	# of atoms in largest ring	Crystal structure with eccentricity of largest ring	Lowest energy conformer with eccentricity of largest ring
ICYSPA (CSD) / Test	1.96	33	0.67	0.38
1E9W (PDB) / Test	2.18	27	0.63	0.43
HAXMOI10 (CSD) / Test	2.22	32	0.69	0.47
1YXQ (PDB) / Training	2.24	40	0.65	0.30
2VYP (PDB) / Test	3.33	38	0.37	0.62

despite the expectation that the CSD structures might be expected to be more accurate, because reproducing conformations for large rings from protein−ligand complexes is of more interest to MacroModel users. Figure S7 plots the RMSD values for conformation generated using MMBS for the MCM

broken down by origin (CSD or PDB) and ring size. For rings containing between 11 and 20 atoms molecules from the CSD typically have smaller RMSD values however for rings containing between 21 and 30 atoms the PDB ligands usually have smaller RMSD values. There are only 2 molecules from

the PDB with rings larger than 30 atoms preventing meaningful comparisons for rings this size. The Welch Two Sample *t* test in which the null hypothesis is that the mean RMSD values for molecules from the PDB and CSD are the same, yields a *p*-value of 0.457. Thus, retrospectively one should not presume selecting structures from either source would permit the MMBS to reproduce the conformation more accurately.

CONCLUSIONS

A specialized MacroModel-based methodology, MMBS, has been developed for conformational sampling of macrocycles. In comparison with published results using the same molecules (BMTS) and, as much as possible, search conditions, we obtain results that are collectively better than the methods published by Bonnet et al.[17] and by Labute[19] at producing low energy conformers. We have characterized the method's sensitivity and identified parameter variations that might be useful for performing faster or more thorough conformational searches. The MMBS performs well on our test set, giving a median RMSD of 0.33 Å for the atoms in the largest ring, 1.74 Å for all heavy atoms when the largest rings are superimposed, and 1.12 Å for all heavy atoms when superimposing using all of the heavy atoms while producing a manageable number of conformers (median 357) for downstream processing. At least 85% of the largest rings are reproduced within 1 Å RMSD. Overall, our RMSDH values are similar to those recently published by Chen and Foloppe[26] although our values are somewhat larger because our test set of molecules contains some larger and more flexible molecules. Most macrocyclic ring conformations are well reproduced indicating that there typically are local minima close to the crystal structure conformation. The MMBS method is available in MacroModel from both the command-line and from a panel within Maestro.[71]

While the MMBS method performs well on the whole, the conformations for rings containing more than 30 atoms typically have RMSD values larger than 1.0 Å, indicating that improvement for these systems is desirable. The plateauing or even in some cases decreases in the number of conformers produced across search methods (MMBS, LowModeMD, SPE(SOS2)) and data sets as the size of the macrocycle grows for the largest macrocycles is puzzling. One interesting observation is that the eccentricities of the largest ring from the crystal structure tends to increase with ring size indicating that the rings are becoming more elongated they become larger. The eccentricities of the rings in the generated conformations also generally increase with ring size, but more slowly than in the crystal structures. For rings containing more than 25 atoms, 84% of the conformers generated have eccentricities smaller than the median eccentricity of the crystal structures. This divergence between the eccentricities of the generated versus crystal structure conformations of the larger rings suggests that these rings in continuum solvent models and perhaps in bulk solution, may have qualitatively different conformations than those found in crystals. We plan to see if improvements to the search method including alterations to the potential energy surface can better reproduce ring conformations for larger macrocycles from crystal structures including their eccentricities. Another avenue for improvement that we plan to explore is to sample side-chain conformations using a systematic approach, such as that used by ConfGen[52,72] which may more readily provide more diverse conformations of the nonring

portions of the molecule without consuming significantly more computer time.

ASSOCIATED CONTENT

Ⓢ Supporting Information
This includes the following:

- Table S1 which contains the CSD or PDB IDs for all molecules from the MCM including their membership in the training and test sets along with crystallographic measures of the quality of the structures for those from the PDB.
- Tables S2−S4 which characterize the molecules in the MCM.
- Table S5 which lists the search parameters for the MMBS.
- Tables S6−8 which provide additional information regarding the application of the vacuum variants of the MMBS to the BMTS.
- Figure S1 which describes in detail the information encoded in a box plot.
- Figures S2−7 which provide additional information on the application of the MMBS to the MCM.
- Table S9, Figure S7, and Table S10 which compare information on crystal structure quality versus the ability of the MMBS to generate similar conformations.
- Table S11 MMBS results for the PDB members of the MCM.
- Reference PDB crystal structure conformations for the macrocyclic compounds in two forms:
  - Original coordinates with minimal processing.
  - Processed structures (translated/rotated,with hydrogen atoms added)

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: John.Shelley@schrodinger.com.

Notes
The authors declare the following competing financial interest(s): MacroModel is a product sold by Schrödinger, LLC., in which both K.S.W. and J.C.S. own stock.

ABBREVIATIONS

4rDDD, distance dependent dielectric with $\varepsilon_r = 4r$ where r is the distance; BDCO, bond-dipole cutoffs; BMTS, Bonnet Macrocycle Test Set; CSD, Cambridge Structure Database; E_w, conformers with energies within this window relative to the lowest energy conformer are retained; GB/SA, a continuum solvent model that includes both generalized Born and surface area terms; IQR, interquartile range, the value for the third quartile minus the first quartile; LLMOD, a conformational search method using low frequency normal modes available in MacroModel that is suitable for large systems; LMMD, LowModeMD, a conformational search method available in MOE; LMOD, a conformational search method using low frequency normal modes available in MacroModel; MCM, the macrocycle containing set of molecules; MMBS, MacroModel's Baseline Search, a conformational search protocol developed for macrocycles; MMBSV, a version of MMBS used for comparisons with previously published results for the BMTS, MMBSV10 and MMBSV20 use a 10 and 20 kcal/mol energy

windows, respectively; MD, molecular dynamics; MMFFs, Merck Molecular Force-Field with modifications to enforce planarity for some functional groups; MOE, Molecular Operating Environment; OPLS, Optimized Potentials for Liquid Simulations; PDB, Protein Databank; QC, quench cycle; RMSD, root-mean-square displacement for just the ring atoms; RMSDH, RMSD for all heavy atoms with all heavy atoms superimposed; RMSDRH, RMSD for all non-hydrogen atoms with the ring atoms superimposed; $RMSD_T$, the threshold RMSD for distinguishing conformers; SMILES, simplified molecular-input line-entry system; SPE(SOS2), Stochastic Proximity Embedding, Self-Organizing Superposition with boosting

■ REFERENCES

(1) Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J. Am. Chem. Soc.* **2006**, *128*, 2510−2511.

(2) 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 Discovery* **2008**, *7*, 608−624.

(3) Oyelere, K. A. Hot topic: Macrocycles in Medicinal Chemistry and Drug Discovery. *Curr. Top. Med. Chem.* **2010**, *10*, 1359−1360.

(4) Drahl, C. Big Hopes Ride On Big Rings. *Chem. Eng. News* **2009**, *87*, 54−57.

(5) Brandt, W.; Haupt, V. J.; Wessjohann, L. A. Chemoinformatic analysis of biologically active macrocycles. *Curr. Top. Med. Chem.* **2010**, *10*, 1361−1379.

(6) Johnson, V. A.; Singh, E. K.; Nazarova, L. A.; Alexander, L. D.; McAlpine, S. R. Macrocyclic inhibitors of Hsp90. *Curr. Top. Med. Chem.* **2010**, *10*, 1380−1402.

(7) Avolio, S.; Summa, V. Advances in the development of macrocyclic inhibitors of hepatitis C virus NS3−4A protease. *Curr. Top. Med. Chem.* **2010**, *10*, 1403−1422.

(8) Kritzer, J. A. Stapled peptides: Magic bullets in nature's arsenal. *Nat. Chem. Biol.* **2010**, *6*, 566−567.

(9) Guo, Z.; Mohanty, U.; Noehre, J.; Sawyer, T. K.; Sherman, W.; Krilov, G. Probing the α-helical structural stability of stapled p53 peptides: Molecular dynamics simulations and analysis: Research article. *Chem. Biol. Drug Des.* **2010**, *75*, 348−359.

(10) Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular hydrogen bonding in medicinal chemistry. *J. Med. Chem.* **2010**, *53*, 2601−2611.

(11) Rezai, T.; Bock, J. E.; Zhou, M. V.; Kalyanaraman, C.; Lokey, R. S.; Jacobson, M. P. Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: successful in silico prediction of the relative permeabilities of cyclic peptides. *J. Am. Chem. Soc.* **2006**, *128*, 14073−14080.

(12) Leung, S. S. F.; Mijalkovic, J.; Borrelli, K.; Jacobson, M. P. Testing physical models of passive membrane permeation. *J. Chem. Inf. Model.* **2012**, *52*, 1621−1636.

(13) Rafi, S. B.; Hearn, B. R.; Vedantham, P.; Jacobson, M. P.; Renslo, A. R. Predicting and improving the membrane permeability of peptidic small molecules. *J. Med. Chem.* **2012**, *55*, 3163−3169.

(14) Ledford, H. Complex synthesis yields breast-cancer therapy. *Nature* **2010**, *468*, 608−609.

(15) Mayer, A. M.; Glaser, K. B.; Cuevas, C.; Jacobs, R. S.; Kem, W.; Little, R. D.; McIntosh, J. M.; Newman, D. J.; Potts, B. C.; Shuster, D. E. The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends Pharmacol. Sci.* **2010**, *31*, 255−265.

(16) Foloppe, N.; Matassova, N.; Aboul-Ela, F. Towards the discovery of drug-like RNA ligands? *Drug Discovery Today* **2006**, *11*, 1019−1027.

(17) Bonnet, P.; Agrafiotis, D. K.; Zhu, F.; Martin, E. Conformational analysis of macrocycles: finding what common search methods miss. *J. Chem. Inf. Model.* **2009**, *49*, 2242−2259.

(18) *MacroModel*, version v9.1; Schrödinger, Inc.: New York, NY, 2008.

(19) Labute, P. LowModeMD–implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. *J. Chem. Inf. Model.* **2010**, *50*, 792−800.

(20) 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.

(21) Kolossvary, I.; Guida, W. C. Low-mode conformational search elucidated; Application to C39H80 and flexible docking of 9-deazaguanine inhibitors into PNP. *J. Comput. Chem.* **1999**, *20*, 1671−1684.

(22) 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.

(23) Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. Conformations of Cycloheptadecane: A Comparison of Methods for Conformational Searching. *J. Am. Chem. Soc.* **1990**, *112*, 1419−1427.

(24) *MacroModel 10.0 Reference manual*; Schrödinger, Inc.: New York, NY, 2013.

(25) Parish, C.; Lombardi, R.; Sinclair, K.; Smith, E.; Goldberg, A.; Rappleye, M.; Dure, M. A comparison of the Low Mode and Monte Carlo conformational search methods. *J. Mol. Graphics Modell.* **2002**, *21*, 129−150.

(26) Chen, I. J.; Foloppe, N. Tackling the conformational sampling of larger flexible compounds and macrocycles in pharmacology and drug discovery. *Bioorg. Med. Chem.* **2013**, *21*, 7878−7920.

(27) *MOE*, Chemical Computing Group, 2013.

(28) Keseru, G. M.; Kolossváry, I. Fully flexible low-mode docking: application to induced fit in HIV integrase. *J. Am. Chem. Soc.* **2001**, *123*, 12708−12709.

(29) Kolossváry, I.; Keseru, G. M. Hessian-Free Low-Mode Conformational Search for Large-Scale Protein Loop Optimization: Application to c-jun N-Terminal Kinase JNK3. *J. Comput. Chem.* **2001**, *22*, 21−30.

(30) *MacroModel*, version 10.0; Schrödinger, Inc.: New York, NY, 2013.

(31) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351−367.

(32) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. Conformational changes of small molecules binding to proteins. *Bioorg. Med. Chem.* **1995**, *3*, 411−428.

(33) 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.

(34) Sitzmann, M.; Weidlich, I. E.; Filippov, I. V.; Liao, C.; Peach, M. L.; Ihlenfeldt, W. D.; Karki, R. G.; Borodina, Y. V.; Cachau, R. E.; Nicklaus, M. C. PDB ligand conformational energies calculated quantum-mechanically. *J. Chem. Inf. Model.* **2012**, *52*, 739−756.

(35) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(36) Allen, F. H. *Acta Crystallogr.* **2002**, *B58*, 380−388.

(37) *Canvas*, version 1.4; Schrödinger, Inc.: New York, NY, 2011.

(38) Figueras, J. Ring perception using breadth-first search. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 986−991.

(39) *LigPrep*, version 2.5; Schrödinger, Inc.: New York, NY, 2011.

(40) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **2007**, *21*, 681−691.

(41) *Epik*, version 2.3; Schrödinger, Inc.,: New York, NY, 2011.

(42) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(43) 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.

(44) Halgren, T. A. Merck Molecular Force Field. III. Molecular Geometrics and Vibrational Frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553−586.

(45) Halgren, T. A.; Nachbar, R. B. Merck Molecular Force Field. IV. Conformational Energies and Geometries. *J. Comput. Chem.* **1996**, *17*, 587−615.

(46) 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.

(47) Halgren, T. A. Merck Molecular Force Field. VI. Option for Energy Minimization Studies. *J. Comput. Chem.* **1999**, *20*, 720−729.

(48) Halgren, T. A. Merck Molecular Force Field. VII. Characterization of MMFF94, MMFF94s and Other Widely Available Force Fields for Conformational Energies and for Intermolecular Interaction Energies and Geometries. *J. Comput. Chem.* **1999**, *20*, 730−748.

(49) Chen, I. J.; Foloppe, N. Conformational sampling of druglike molecules with MOE and catalyst: implications for pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* **2008**, *48*, 1773−1791.

(50) Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E. Conformational sampling of bioactive molecules: a comparative study. *J. Chem. Inf. Model.* **2007**, *47*, 1067−1086.

(51) Chen, I. J.; Foloppe, N. Drug-like bioactive structures and conformational coverage with the LigPrep/ConfGen suite: comparison to programs MOE and catalyst. *J. Chem. Inf. Model.* **2010**, *50*, 822−839.

(52) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: a conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534−546.

(53) The R Manuals. http://cran.r-project.org/manuals.html (accessed July 1, 2014).

(54) Borodina, Y. V.; Bolton, E.; Fontaine, F.; Bryant, S. H. Assessment of conformational ensemble sizes necessary for specific resolutions of coverage of conformational space. *J. Chem. Inf. Model.* **2007**, *47*, 1428−1437.

(55) 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.

(56) Hao, M. H.; Haq, O.; Muegge, I. Torsion angle preference and energetics of small-molecule ligands bound to proteins. *J. Chem. Inf. Model.* **2007**, *47*, 2242−2252.

(57) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(58) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.

(59) Reddy, M. R.; Erion, M. D.; Agarwal, A.; Viswanadhan, V. N.; McDonald, D. Q.; Still, W. C. Solvation free energies calculated using the GB/SA model: Sensitivity of results on charge sets, protocols, and force fields. *J. Comput. Chem.* **1998**, *19*, 769−780.

(60) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate solvent-accessible surface areas from tetrahedrally directed neighbor densities. *Biopolymers* **1999**, *50*, 373−380.

(61) Weiser, J.; Shenkin, P. S.; Still, W. C. Optimization of Gaussian surface calculations and extension to solvent-accessible surface areas. *J. Comput. Chem.* **1999**, *20*, 688−703.

(62) Weiser, J.; Shenkin, P. S.; Still, W. C. Fast, approximate algorithm for detection of solvent-inaccessible atoms. *J. Comput. Chem.* **1999**, *20*, 586−596.

(63) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217−230.

(64) *OPLS2.1*, version 2.0; Schrödinger, Inc.: New York, NY, 2012.

(65) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. Improving the prediction of absolute solvation free energies using the next generation opls force field. *J. Chem. Theory Comput.* **2012**, *8*, 2553−2558.

(66) See for instance: en.wikipedia.org/wiki/Welch's_t_test (accessed Sept. 20, 2014).

(67) Read, R. J.; Adams, P. D.; Arendall, W. B., 3rd; Brunger, A. T.; Emsley, P.; Joosten, R. P.; Kleywegt, G. J.; Krissinel, E. B.; Lutteke, T.; Otwinowski, Z.; Perrakis, A.; Richardson, J. S.; Sheffler, W. H.; Smith, J. L.; Tickle, I. J.; Vriend, G.; Zwart, P. H. A new generation of crystallographic validation tools for the protein data bank. *Structure* **2011**, *19*, 1395−1412.

(68) Gore, S.; Velankar, S.; Kleywegt, G. J. Implementing an X-ray validation pipeline for the Protein Data Bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 478−483.

(69) Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* **2013**, *280*, 5705−5736.

(70) Warren, G. L.; Do, T. D.; Kelley, B. P.; Nicholls, A.; Warren, S. D. Essential considerations for using protein-ligand structures in drug discovery. *Drug Discovery Today* **2012**, *17*, 1270−1281.

(71) *Maestro*, version 9.4; Schrödinger, Inc.: New York, NY, 2013.

(72) *ConfGen*, version 2.3; Schrödinger, Inc.: New York, NY, 2012.