

University of Groningen

Benchmark of Generic Shapes for Macrocycles

Reyes Romero, Atilio; Ruiz-Moreno, Angel Jonathan; Groves, Matthew R; Velasco-Velázquez, Marco; Dömling, Alexander

Published in:
Journal of chemical information and modeling

DOI:
[10.1021/acs.jcim.0c01038](https://doi.org/10.1021/acs.jcim.0c01038)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Reyes Romero, A., Ruiz-Moreno, A. J., Groves, M. R., Velasco-Velázquez, M., & Dömling, A. (2020). Benchmark of Generic Shapes for Macrocycles. *Journal of chemical information and modeling*, 60(12), 6298-6313. <https://doi.org/10.1021/acs.jcim.0c01038>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Benchmark of Generic Shapes for Macrocycles

Atilio Reyes Romero, Angel Jonathan Ruiz-Moreno, Matthew R. Groves, Marco Velasco-Velázquez, and Alexander Dömling*

Cite This: *J. Chem. Inf. Model.* 2020, 60, 6298–6313

Read Online

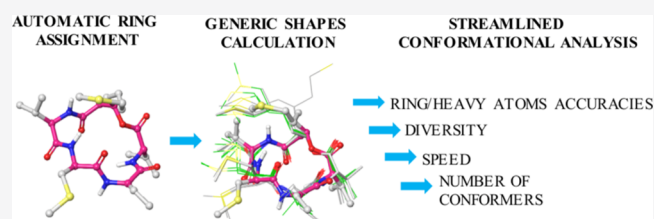
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Macrocycles target proteins that are otherwise considered undruggable because of a lack of hydrophobic cavities and the presence of extended featureless surfaces. Increasing efforts by computational chemists have developed effective software to overcome the restrictions of torsional and conformational freedom that arise as a consequence of macrocyclization. Moloc is an efficient algorithm, with an emphasis on high interactivity, and has been constantly updated since 1986 by drug designers and crystallographers of the Roche biostructural community. In this work, we have benchmarked the shape-guided algorithm using a dataset of 208 macrocycles, carefully selected on the basis of structural complexity. We have quantified the accuracy, diversity, speed, exhaustiveness, and sampling efficiency in an automated fashion and we compared them with four commercial (Prime, MacroModel, molecular operating environment, and molecular dynamics) and four open-access (experimental-torsion distance geometry with additional “basic knowledge” alone and with Merck molecular force field minimization or universal force field minimization, Cambridge Crystallographic Data Centre conformer generator, and conformator) packages. With three-quarters of the database processed below the threshold of high ring accuracy, Moloc was identified as having the highest sampling efficiency and exhaustiveness without producing thousands of conformations, random ring splitting into two half-loops, and possibility to interactively produce globular or flat conformations with diversity similar to Prime, MacroModel, and molecular dynamics. The algorithm and the Python scripts for full automatization of these parameters are freely available for academic use.



INTRODUCTION

Macrocycles comprise a (hetero)cyclic core of at least 12 atoms, with molecular weight typically between 500 and 2000 Da. Ring sizes of 8–11 atoms and 3–7 atoms are classified as medium and small cycles. Although some naturally occurring rings contain up to 50 atoms, 14-, 16-, and 18-membered rings occur at a higher frequency.¹ Generally, they encompass a large variety of chemical structures that originate from macrocyclization of simple building blocks, for example, cyclopeptide,² cycloalkanes, and cyclodextrins,³ or as a result of *de novo* total synthesis or semisynthetic routes.⁴ Among their clinical applications as drugs, macrocycles are used in oncology (temsirolimus and^{5,6} epothilone B derivatives^{7,8}), as antibiotics (vancomycin, macrolides, and rifampicin), immunology (sirolimus and zotarolimus), and in dermatology (pimecrolimus).⁹ Other applications of macrocycles are in supramolecular chemistry (crown ethers,¹⁰ cryptands, catenanes, rotaxanes,¹¹ and calixarenes). Recently, macrocycles have received growing attention in medicinal chemistry^{12–15} because of their unique ability to disrupt protein–protein interactions,¹⁶ improve metabolic stability,¹⁷ and improve cellular permeability by conformational restriction^{18–21}—resulting in a higher oral bioavailability compared to noncyclic congeners. Although macrocycles are outside of Lipinski’s rule of five, these molecules are able to bind proteins that are otherwise considered challenging because of their lack of

hydrophobic cavities where functional groups can be anchored.^{22,23} It has been estimated that nearly 25% of the ring atoms can contribute to the contact area with the protein surface through nonpolar contacts. Nevertheless, both ring atoms and peripherals/substituents show the same probability to match a hotspot, suggesting that ligand-based drug design of macrocycles should take into account these two components in order to identify potent binders.²⁴ We have recently described multiple scaffolds of artificial macrocycles which are readily synthesizable using multicomponent reaction chemistry (MCR)^{25–30} and investigated the structural basis of macrocycles targeting PD1–PDL1, p53–MDM2, and IL17A receptor interactions.^{30–33} Thus, we are highly interested in computational tools to rapidly screen conformational space of large virtual macrocycle libraries as a filter to synthesize bioactive compounds. To date, several benchmarks demonstrated the feasibility of algorithms with the aim of producing macrocycle conformations with enough accuracy and uniqueness for common computer-aided drug design (CADD)

Received: September 2, 2020

Published: December 3, 2020



Table 1. Free (Green) and Commercial (Salmon) Software for the Conformation Generation of Macrocycles and Their Working Principles

methodology	description	usage
Moloc	macrocycle shapes are characterized by a selection of harmonics which occur in an approximate Fourier representation of the atomic coordinates of the rings. ⁴²	free
Conformator	incremental construction of conformers with torsional angle assignment and a new deterministic cluster algorithm. ⁴⁶	free
CCDC	ring template libraries to describe ring geometries using based on the wealth of experimental data in CSD.	free
ETKDG	stochastic search method that utilizes DG together with knowledge derived from experimental crystal structures. ^{47,48}	free
MOE	perturbation of an existing conformation along a MD' trajectory using initial atomic velocities with kinetic energy focused on the low-frequency vibrational modes and energy minimization. ³⁸	commercial
Prime	ring splitting to create to two half rings that are sampled independently and recombined. ⁴⁹	commercial
MD	Desmond from Schrödinger suite 2014-4 chosen as a baseline method (MaestroDesmond Interoperability Tools; Schrödinger: New York, NY, 2014).	commercial
MM	brief MD simulations followed by minimization and normal-mode search steps. ³⁹	commercial

strategies, such as docking and pharmacophore screening.³⁴ Some of these algorithms are based on distance geometry (DG),³⁵ inverse kinematics,³⁶ genetic algorithms,³⁷ molecular dynamics (MD) simulations implementing either low-frequency modes³⁸ or normal-mode search steps plus energy minimization,³⁹ and, most recently, Monte Carlo multiple minimum/mixed torsional/low mode.⁴⁰

Generally, these software programs are distinguished on the basis of the strategy adopted to generate conformations, systematic or stochastic. For example, molecular operating environment (MOE), MacroModel (MM), Cambridge Crystallographic Data Centre (CCDC) conformer generator, and experimental-torsion DG with additional “basic knowledge” (ETKDG) belong to the stochastic search category. Nevertheless, a major issue with these techniques is the generation of large numbers of representative conformers. On the other hand, a problem related to systematic search methods is the constrained flexibility of the ring, which is often insufficiently sampled by rotating a single bond at a time. In contrast to noncyclic molecules, the change in a single bond rotation impacts all bonds in macrocycles. Developing methods for sampling macrocycle conformations or improving upon the currently existing methods without generating a large number of conformers is a key step in the exploration of macrocycles in drug discovery.

The computational basis of finite Fourier transform of ring structures was developed in 1985⁴¹ and its first embedding within a specialized conformer generator for macrocycle conformational sampling was shown in the publication of Gerber and co-workers in 1988.⁴² Fourier representation of the atomic position for macrocycle sampling has the advantage of generating a number of conformations that depend solely on the number of atoms in the ring, with few other user defined parameters. In the original publication, the author assessed the extensive conformational space covered by the Moloc software by taking (*E*)-cyclodecene and *s-cis/s-trans*-caprolactam as two study cases, investigating the potential of their method in combination with NMR spectroscopy of a macrocyclic tetrapeptide as a third example. This resulted in an exhaustive set of low-energy conformations of macrocyclic systems generated automatically, reproducing the experimented observed conformations, including *s-cis/s-trans*-isomers and, finally, showing the potential application in modeling surface loops of proteins.

Herein, we benchmark the Fourier-based algorithm using a database of 208 macrocycle crystal structures and compare the performances of Moloc with the commercial software Prime, MOE, MD, MM, and four open-access packages—exper-

imental-torsion DG with additional “basic knowledge” and with the minimization steps employing the Merck molecular force field (MMFF94s⁴³) or the universal force field (UFF⁴⁴), CCDC, and conformator. We systematically assess the accuracy, structural diversity, and speed. Moreover, concepts of exhaustiveness and sampling efficiency (SE) are introduced. The aim of our work is to identify software capable of producing diverse and accurate conformations for daily virtual screening (*i.e.*, docking). Moreover, because significant conformational changes in total shape and volume guide the bioavailability of certain macrocycles,⁴⁵ we believe that the application of this approach could efficiently identify generic shapes of membrane-permeating conformations.

A summary of the different software and the theoretical principles behind their functionality are presented in Table 1.

■ MATERIALS AND METHODS

Dataset. For a direct comparison of Moloc with the commercial and free software, we used the dataset of 208 macrocycles of Sindhikara and co-workers,⁴⁹ consisting of 130 crystal structures from the Cambridge crystallographic dataset,⁵⁰ a subset of 60 structures from the Protein Data Bank (PDB⁵¹) selected by Watts and co-workers³⁹ accounting for diverse and challenging macrocyclic topologies (disulfide bridges, cross-linking amide bonds, and polycyclic rings, including cyclodextrins, polyglycines, cycloalkanes, and peptidic macrocycles) and 18 crystals from the Biologically Interesting Molecule Reference Dictionary (BIRD) dataset chosen on the basis of quality (low-temperature factors and/or resolution < 2.1 Å) and structural diversity. Further details about the full dataset composition can be found in the Supporting Information from Sindhikara and co-workers.⁴⁹

Preparation of the Input Structures. Nonbiased starting conformations were prepared by removing the initial crystallographic coordinates, the partial charges, and the explicit hydrogens. Processed structures were converted to isomeric SMILES, preserving the stereochemistry flags. The resulting SMILES codes were employed as input for conformational sampling by conformator, CCDC conformer generator, and ETKDG alone or in combination with the minimization steps employing the MMFF94s or UFF while for Moloc, a set of random three-dimensional (3D) structures were generated using Mol3d.

Software Tested and Parametrization. MOE, Prime, MM, and MD. Macrocycle sampling description and initial condition for Prime, MOE, MM, and MD can be found in the Methods section of Sindhikara and co-workers while the

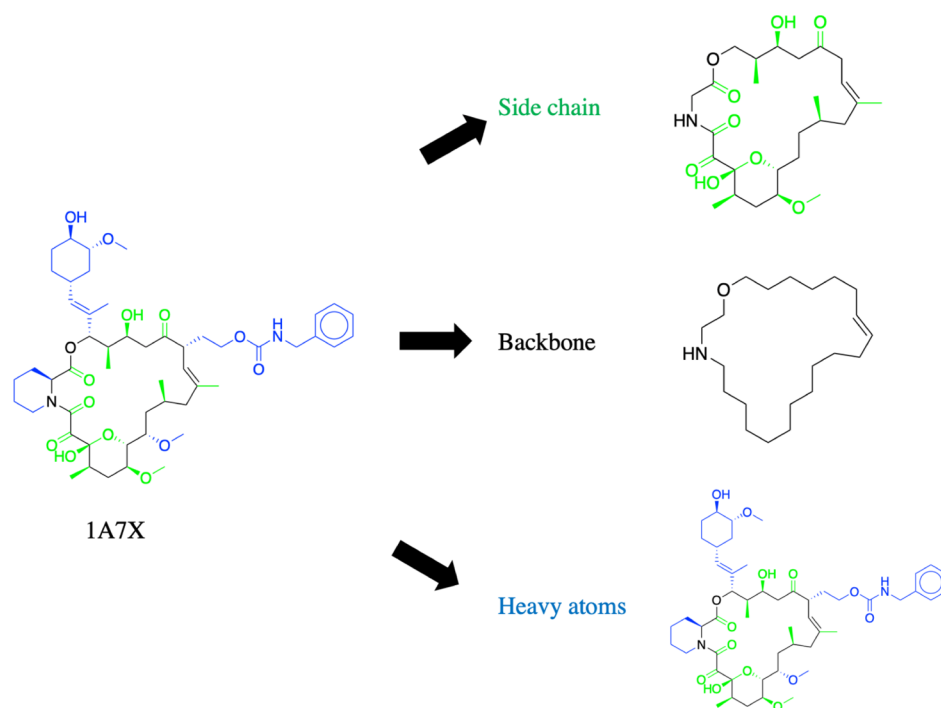


Figure 1. Example of separation of a 21-membered macrocycle into three atomic categories for the calculation of the $\text{RMSD}_{\text{backbone}}$ and $\text{RMSD}_{\text{heavy atoms}}$. Side chains, backbone, and heavy atoms are colored green, black, and blue, respectively.

results of accuracy, diversity, and speed can be found in the [Supporting Information](#).⁴⁹

Moloc. Moloc is one of the first molecular modeling packages and has since been updated regularly in close collaboration with drug designers and crystallographers of the Roche biostructural community, encompassing numerous functions, such as conformational sampling, generation of 3D pharmacophores,⁵² similarity analysis, peptide and protein modeling, modules for X-ray data handling, and ligand-based drug design. The generic Fourier description of the shape of the ring atoms is based on the generation of a series of harmonics.⁴² Radial and axial deviations are then applied until a generic shape is found. Once it is identified, the algorithm starts to build a number of conformations that is proportional to the ring size. Geometric deviations, such as bond length and angles, are fixed by minimizing against the MAB force field.⁵³ In order to launch a sampling job, the “Mcnf” module was run in batch with the parameters “w0” and “c3” to initiate randomization of input atomic 3D coordinates and preserve the stereochemistry of both E/Z bonds and sp^3 carbon, respectively. The selection of unique conformations is based on energetic (0.1 kcal/mol) and structural [0.1 Å root mean square deviation (RMSD) for cross-rigid body superimposition] thresholds. The conformations were kept within an energetic threshold of 10 kcal/mol. A conformational job can be launched using either two-dimensional (2D) or 3D atomic coordinates that are generated using Mol3d. During the conformational sampling, inner symmetries and permutations are enumerated. The number of generic shapes used as a start guide for the generation of the conformers grows as the square of $N(\ln N)$ where N represents the number of ring atoms. Finally, for assessment, the flexibility of the software, energetic threshold, and hydrogen bond term were activated for the conformational job.

Conformer. Conformer is a conformer generator focused on the enhancement of molecular torsion based on the assessment of torsion angles from the rotatable bonds. Conformer consists of a torsion driver enhanced by an elaborate algorithm for the assignment of torsion angles to rotatable bonds and a new clustering component that efficiently compiles ensembles by taking advantage of lists of partially presorted conformers. The clustering algorithm minimizes the number of comparisons between pairs of conformers that are required to effectively derive individual RMSD thresholds for molecules and to compile the ensemble. For this purpose, conformer features two conformer generation modes, “fast” and “best”, where “best” and “fast” focuses on the accuracy or speed of conformer search to generate conformers with the lowest RMSD values against a reference, respectively. Both modes attempt to ensure chemically correct bond angles and lengths as well as the planarity of aromatic rings and conjugated systems. After conformer generation, conformer performs a local optimization employing the macrocyclic optimization score which includes several well-known components from common force fields and some components specific to the optimization of macrocycles.⁴⁶ For optimal comparison of the software, we selected the “best” feature for macrocycle conformational sampling using the isomeric SMILES codes described above and requesting one thousand conformers per entry.

CCDC Conformer Generator. Conformer generator from CCDC is a knowledge-based method that uses data derived from CSD libraries and heuristic rules. For instance, conformer generator uses rotamer libraries to characterize preferred rotatable bond geometries and ring template libraries to describe ring geometries. Conformations are sampled based on CSD-derived rotamer distributions and ring templates. A final diverse set of conformers, clustered according to conformer similarity, are returned. Each conformer is locally optimized in

torsion space.^{48,54} For this work, the input structures described previously were loaded into the CCDC conformer generator through the CSD Python application programming interface (API). Conformer generator runs a minimization using the Tripos force field prior to conformational sampling, for which one thousand conformers were requested for each entry.

ETKDG Alone and with Minimization. RDKit is an open-source toolkit for cheminformatics, comprising a wide variety of analysis and synthesis tools including similarity search, fingerprint calculations, 2D and 3D descriptor calculation, and conformer generation (<https://www.rdkit.org/>). Currently, RDKit is able to generate conformers using DG and an improved new method called ETKDG. The ETKDG algorithm is based on DG including experimental torsion angle termed experimental-torsion DG (ETDG) and “basic knowledge” (ETKDG) of molecular terms, including linear triple bonds and planar aromatic rings. The ETKDG method has been demonstrated to be more accurate in reproducing crystal structure conformations than DG alone. In addition, this algorithm has been recently optimized by the implementation of knowledge-based terms, preference for the *trans*-amide configuration, and the control of eccentricity from 2D elliptical geometry.⁴⁸ Thereby, we decided to explore the ETKDG approach for macrocycle sampling. Because ETKDG conformational sampling lacks any step of minimization, we ran minimization steps after the ETKDG conformational job using MMFF94s or UFF over 400 iterations per conformer in order to explore the minimization effect on macrocycle conformational sampling. We used the Python API of RDKit to generate one thousand conformers per entry from the input structures.

Comparison Parameters. *Exhaustiveness.* Not all the software compared exhaustively sampled conformational space but stopped before because some of them were not able to generate conformations for some of the input structures. For instance, no sampling was performed in the case conformator if the assignment of torsion angles to rotatable bonds failed for a specific structure because this is the flexibility determination method employed using such a software. Thus, we defined the term exhaustiveness as follows

$$\text{Exhaustiveness} = \frac{\text{num. entries sampled}}{\text{total entries}}$$

Accordingly, exhaustiveness values equal to 1 indicate full sampling of all entries in the dataset. Correspondingly, decreased exhaustiveness values indicate fewer entries sampled.

Accuracy. Based on previous benchmarks of conformational sampling,^{38,39,46,49,55,56} we have used RMSD to quantify the accuracy of the conformers in reproducing the reported bioactive crystallographic coordinates.

The lowest RMSD values between each conformational ensemble to the reference structure were calculated. Notably, we have quantified the ring atom accuracy ($\text{RMSD}_{\text{backbone}}$) in a separate manner from heavy atom accuracy ($\text{RMSD}_{\text{heavy atoms}}$), as indicated in Figure 1. This is based on the recently described classification of contacts between the macrocycle and its target: side chain, peripheral functional groups, and backbone atoms to the receptor.²⁴ Typically, a relative RMSD cutoff below 2.0 Å is considered an acceptable accuracy.⁵⁷ However, because macrocycles are more complex and larger than small molecules, we considered an $\text{RMSD}_{\text{heavy atoms}}$ value up to 2.5 Å as reasonably accurate and $\text{RMSD}_{\text{heavy atoms}}$ values below 1.0 Å were treated as highly accurate. Finally, we used the

cumulative function distribution (CDF) to evaluate the performance of the algorithm in sampling a specific percentage of the dataset below two $\text{RMSD}_{\text{backbone}}$ threshold values 0.5 Å (highly accurate) and 1.0 Å (accurate).

Diversity and SE. In order to systematically assess the structural diversity of each conformational ensemble, we used torsional fingerprints (TFs) in a similar manner to Sindhikara and co-workers.⁴⁹ The unique conformers were identified using a torsional scan on multiple conformations of a truncated version of the molecule comprising only the macrocycle backbone. Correspondence between related molecules was assessed by atom mapping from a maximum common substructure analysis. Then, a comparison of the fingerprints between the conformers was calculated using the torsional fingerprint deviation (TFD).⁵⁸ Conformers with unique fingerprints were identified and kept if TFD was nonzero. As a further descriptor for assessment of shape diversity, we used the span in the radius of gyration (RoG), which is defined as the difference between the highest and the lowest RoG conformers.⁵⁹ Aiming to establish a relation among the exhaustiveness and the capability of the software to generate unique conformers, we introduced the SE as

$$\text{SE} = \text{exhaustiveness} \left(\frac{\text{unique torsional fingerprints}}{\text{num. conformers}} \right)$$

SE values equal to 1 mean that each conformer represents a unique conformation within taking in account the number of entries sampled, while values close to 0 indicate high redundancy among conformers and/or lower exhaustiveness.

Speed. Time efficiency for each software was quantified by calculating the difference between the start and end time for conformer generation per entry. Batch scripts were generated for calculation of the time consumption for Moloc and conformator. Because of the usage of Python API for RDKit and CCDC conformer generator, a tailored Python script was implemented in order to calculate the time consumption for CCDC conformer generator, ETKDG, and its further minimizations steps (UFF or MMFF94s). Moloc, conformator, and ETKDG alone or with minimization and CCDC conformer generator were run in a machine utilizing a 4-core Intel Xeon 3500 CPU-processor, 12 GB RAM, and 25 GB of data storage in a 1 TB HDD. The speed of MOE, MM, Prime, and MD was retrieved from the Supporting Information of the Prime benchmark publication.⁴⁹

Statistical Analysis. Data representation was carried out using the Python library matplotlib 3.1.1.⁴⁸ Statistical comparison of data was computed using a nonparametric Kruskal–Wallis *H*-test among study groups using the stats module of SciPy.⁶⁰ All the *p*-values of the pairwise comparisons among the software can be found in the Supporting Information.

RESULTS

Exhaustiveness. According to our observations from conformational sampling of macrocycles employing different software, some methods were incapable of sampling all entries into the database. Conformer resulted in the least exhaustive sampling (190 out of 208 entries). Although the ETKDG algorithm was able to generate conformers for all input structures, the subsequent minimization step using UFF or MMFF94s force fields resulted in less exhaustiveness than the ETKDG algorithm alone (197 out of 208). All the remaining

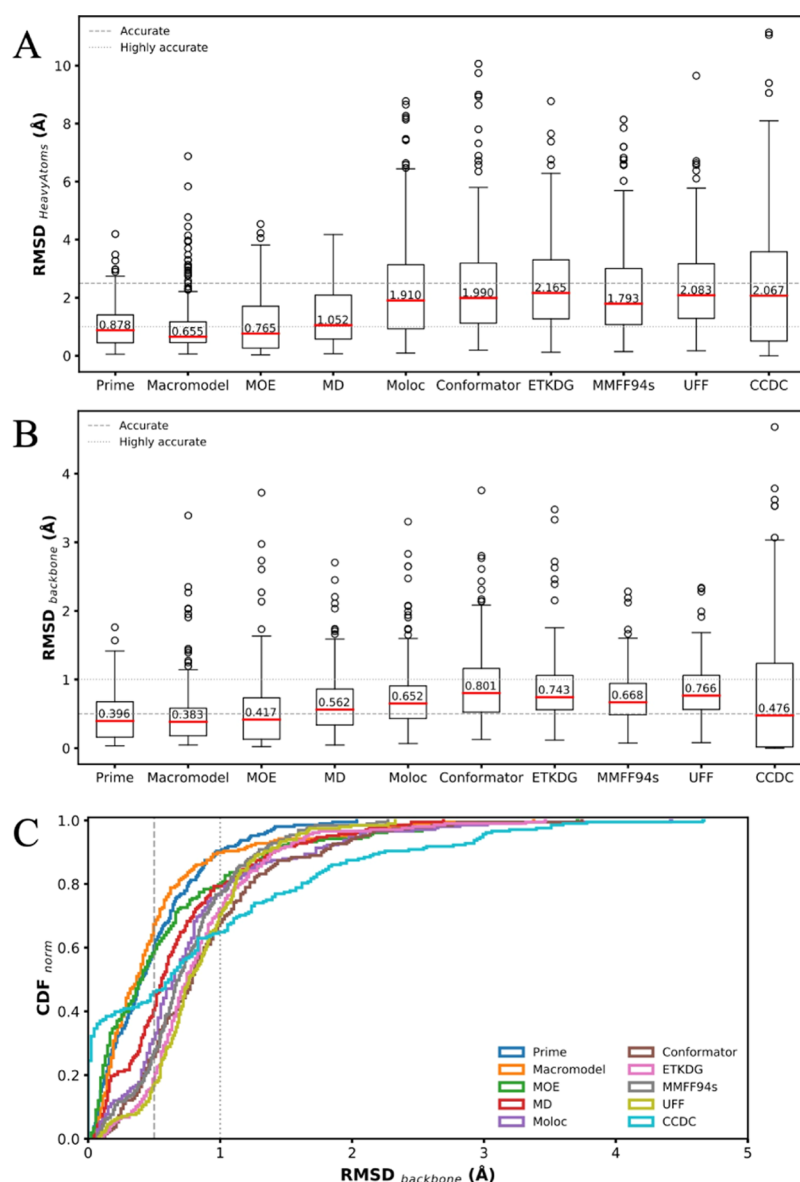


Figure 2. Crystal structure accuracies for each method displayed as (A) $\text{RMSD}_{\text{heavy atoms}}$ and (B) $\text{RMSD}_{\text{backbone}}$, respectively. (C) Normalized cumulative distribution function (CDF_{norm}). The accuracy threshold values, median, and outliers are presented as gray dots, red lines, and black-contoured circles, respectively.

software tested (Moloc, CCDC conformer generator, and ETKDG) or previously reported (Prime, MOE, MM, and MD) was able to generate conformers for all input structures (Table 3).

Accuracy. Figure 2 indicates that all the software can generate conformers with reasonable accuracy ($\text{RMSD}_{\text{heavy atoms}} < 2.5 \text{ \AA}$) and MM, MOE, and Prime generated conformers with median $\text{RMSD}_{\text{heavy atoms}}$ values below a threshold of 1.0 \AA with no statistical difference among the methods (Table S1). Among the six other software tested in this work, ETKDG algorithm plus MMFF94s minimization and Moloc were able to generate conformers with the lowest median $\text{RMSD}_{\text{heavy atoms}}$ value. However, in contrast to ETKDG plus MMFF94s minimization (0.9471), Moloc retained superior exhaustiveness (1), indicating that it is able to generate reasonably accurate conformers across a complex and diverse dataset of macrocycle molecules. No statistical difference was found among all open-source methods, including CCDC conformer

generator. Finally, MD showed a median $\text{RMSD}_{\text{heavy atoms}}$ value slightly higher for the highly accurate threshold, and statistical difference versus all the remaining private and open-access methods. In $\text{RMSD}_{\text{backbone}}$ and CDF analysis, Figure 2A shows that Prime, MM, MOE, and CCDC conformer generator produced the highest accurate conformers ($\text{RMSD}_{\text{backbone}} < 0.5 \text{ \AA}$) with no statistical difference among these four methods (Table S2), returning a fraction of entries sampled for each method of 0.63 , 0.67 , 0.58 , and 0.46 , respectively (Figure 2B and Table 2). In addition, our data indicate that all the remaining methods generated conformers below 1.0 \AA . No statistical difference was observed among MD, Moloc, and ETKDG with MMFF94s, whose fraction of sampled entries was, respectively, 0.79 for the first two and 0.78 .

Such results indicate similar accuracy among these methods to reproduce the reference macrocycle backbone structure. Similarly, no statistical difference was found between Moloc and MMFF94s and both produced a similar fraction of entries

Table 2. Fraction of Entries Sampled below the Two RMSD Backbone Thresholds Chosen as Highly Accurate (<0.5 Å) and Accurate (<1.0 Å)

method	<0.5 Å	<1.0 Å
Prime	0.63	0.90
MM	0.67	0.90
MOE	0.58	0.80
MD	0.40	0.79
Moloc	0.31	0.79
conformator	0.26	0.68
CCDC	0.46	0.65
ETKDG	0.19	0.72
MMFF94s	0.27	0.78
UFF	0.17	0.70

sampled above the threshold (Moloc: 0.77, MMFF94s: 0.79). Finally, comparison between conformator, ETKDG, and ETKDG plus UFF minimization did not show any statistical differences. A statistical difference was found when comparing conformator, ETKDG, and ETKDG plus UFF minimization versus Moloc or ETKDG plus MMFF94s minimization with a fraction of entries sampled being 0.68 for conformator, 0.72 for ETKDG, and 0.70 for ETKDG plus UFF minimization steps. However, among these last groups of methods, ETKDG is the most exhaustive followed by ETKDG plus UFF minimization and conformator.

Diversity and SE. Although all software was challenged with a one thousand conformers per entry request, not all of them succeeded in accomplishing the task, either retrieving fewer conformers per entry or unable to sample some, resulting in poor exhaustiveness. Among the methods studied, only MD and ETKDG succeeded in generating all conformers requested. Nevertheless, we compared the TFs of the conformers for each method in order to assess the number of unique conformers generated and, furthermore, we employed the exhaustiveness value to calculate the SE of each software. We identified Moloc and ETKDG followed by ETKDG plus minimization with either MMFF94s or UFF as the most efficient methods to perform conformational search of macrocycles (Table 3). On the contrary, although MD showed an exhaustiveness value of 1, it is also a highly redundant method generating only a median of 59 unique conformers across 1000 conformers retrieved, obtaining the lowest SE value (0.059) among all reported methods. In a similar fashion to MD, MM showed a low SE. Despite being a highly exhaustive methodology, the

Table 3. Summary Table of the Exhaustiveness and SE, Number of Conformers, and TFs

method	exhaustiveness	unique TF (median)	number of conformers (median)	SE
Prime	208/208 = 1	707	932	0.7586
MM	208/208 = 1	100	300	0.3333
MOE	208/208 = 1	48	76	0.6316
MD	208/208 = 1	59	1000	0.0590
Moloc	208/208 = 1	67	67	1
conformator	190/208 = 0.91	246	338	0.6648
ETKDG	208/208 = 1	1000	1000	1
MMFF94s	197/208 = 0.95	998	998	0.9471
UFF	197/208 = 0.95	535	535	0.9471
CCDC	208/208 = 1	6	8	0.7500

relation between the number of conformers generated and their uniqueness results in an SE of 0.333. Thus, Moloc and ETKDG are three times more efficient in macrocycle conformation sampling than MD. However, Prime (exhaustiveness: 1) was able to produce a median of 707 unique conformers for a median of 932 conformers, resulting in an SE of 0.7586. A similar behavior was observed for MOE, which obtained exhaustiveness equal to 1 and an SE of 0.6316. CCDC conformer generator showed an SE of 0.7500 with the lowest number of unique conformers generated (Figure 3A) across all the software studied.

Figure 4A compares the results obtained from the span of RoG as a parameter to study the 3D conformational diversity of the conformers moving from a globular to a flat-shaped conformation (Figure 4B). Our data indicate that ETKDG algorithm plus MMFF94s minimization (1.13 Å) achieved the highest span in RoG with no statistical difference with Prime (1.02 Å) and ETKDG with UFF minimization (1.08 Å) (Table S4). On the other hand, the conformations produced by Moloc (0.86 Å) were proven to be statistically similar to MM (0.93 Å), MOE (0.74 Å), MD (0.85 Å), conformator (0.87 Å), and ETKDG alone without minimization (0.82 Å). Finally, with a span in RoG of 0.15 Å, the conformers produced by CCDC conformer generator were identified as having the lowest diversity among all the software tested.

Speed. Surprisingly, the speed of macrocyclic conformation generation differed dramatically between the software ranging from seconds to more than a day. This will have consequences for usage in virtual screening of large macrocycle libraries. Because sampling is carried out under similar conditions, comparisons allow analysis of the time required to accomplish the conformational task. The overall results of the computational speed are shown in Figure 5. With 2.6 s per entry, CCDC conformer generator outperformed the other software in time needed to finish a conformational job. On the other hand, MD was the slowest followed by conformator, which required 17.9 h. Prime, Moloc, and MOE produced conformations with a similar speed within 1 h with nonsignificant differences between MOE and Moloc (Table S5). More interestingly, we observed a statistical difference between ETKDG alone and UFF/MMFF94s resulting in a median of 35.1 s, 1.3 min, and 17.6 per entry.

Study Cases. In addition to the benchmark results described above, we report cases of effective accuracy in predicting the crystallographic coordinates of macrocycles using Moloc both in terms of lowest RMSD_{backbone}/RMSD_{heavy atoms} and in relation with the ring size. For convenience, we kept the same categories as previously reported,⁴⁹ binning the database in three groups containing 10–19, 20–29, and over 30 ring atoms, respectively. We referred to Prime as a comparative example among other commercial software.

10–19-Ring-Sized Macrocycles. 10–19-ring-sized macrocycles represent a challenge in the context of organic synthesis because of the high energetic strain. Similarly, medium-sized rings suffer from increased ring strain over their 5- and 6-membered or macrocyclic congeners.^{62,63} This can be quantitatively captured in deviations from ideal antiperiplanar conformations, transannular strain, and Pitzer strain components. Out of the total 208, 117 macrocycles belong to this class, including 30 from PDB, 79 from CSD, and 8 from BIRD datasets. According to our findings, Moloc predicted the coordinates of ACOPUF (Figure 6A), a 12-ring-sized macro-

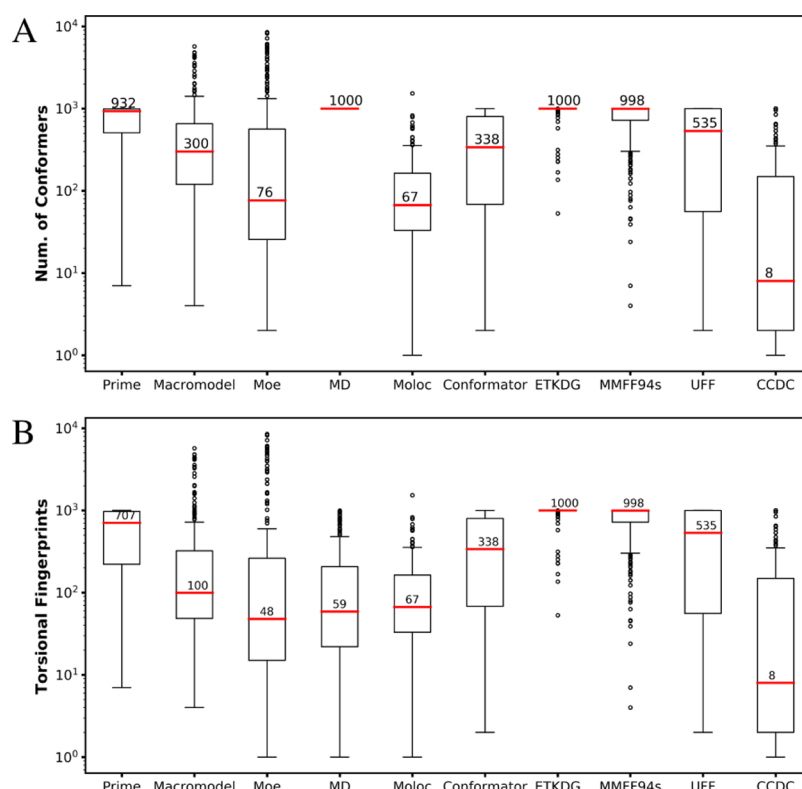


Figure 3. Panel showing (A) box plot of number of the conformers and (B) TFs for each method. Graphical description of median and outliers is the same as in Figure 2.

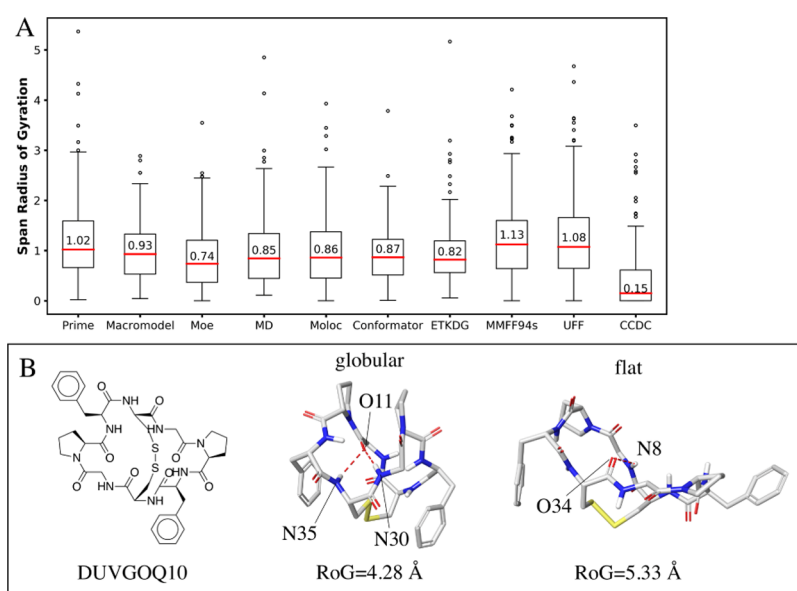


Figure 4. (A) Box plot of span RoG for each method and (B) example of a cyclic octapeptide⁶¹ in its globular (lowest RoG) and flat-like conformations (highest RoG) with intramolecular hydrogen bonds predicted with Moloc (red dotted lines).

cycle from the CSD database, with an $\text{RMSD}_{\text{backbone}}$ of 0.07 Å—slightly better than Prime (0.12 Å)—and with less conformations (requiring only 93 for the former against 871 for the latter). In a similar fashion, Moloc predicted the bioactive conformation of cytochalasin D (Figure 6C), an 11-membered ring macrocycle from the PDB database, with a high accuracy (0.12 Å) employing only 9 conformers, whereas Prime (0.15 Å) employed 185. BANROX (Figure 6B) and DOZWUL (Figure 6D) were two CSD macrocycles of 13- and

14-atom backbone, respectively, with an $\text{RMSD}_{\text{heavy atoms}}$ of 0.09 and 0.10 Å. These data indicate that this software is highly accurate for medium-sized rings. In contrast to Prime, Moloc also proved to be superior in terms of the number of conformations, producing only 33 and 93 conformers rather than 95 for BANROX and 388 for DOZWUL, and accuracy with $\text{RMSD}_{\text{heavy atoms}}$ values of 0.44 and 0.41 Å for Prime.

20–29-Ring-Sized Macrocycles. This category includes 67 X-ray structures, 27 from PDB, 34 from CSD, and 6 from

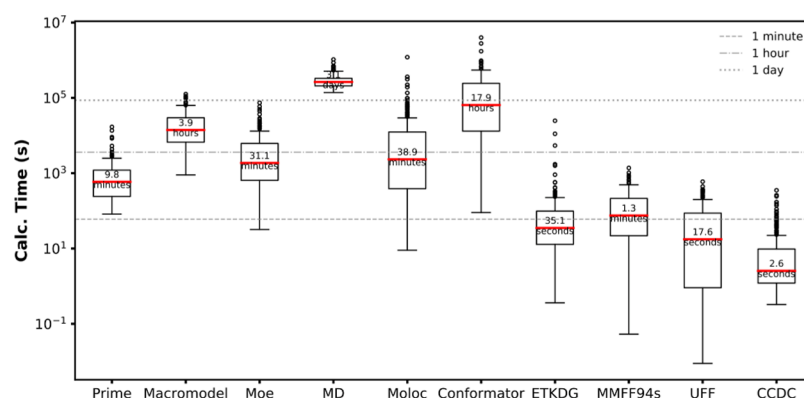


Figure 5. Box plot showing the distribution of the speed ranges for each entry. The reader is referred to Figure 2 for the legend. Three significant threshold values were added to visualize the differences in the performance level in completing a conformation work, *i.e.*, 1 min, 1 h, and 1 d.

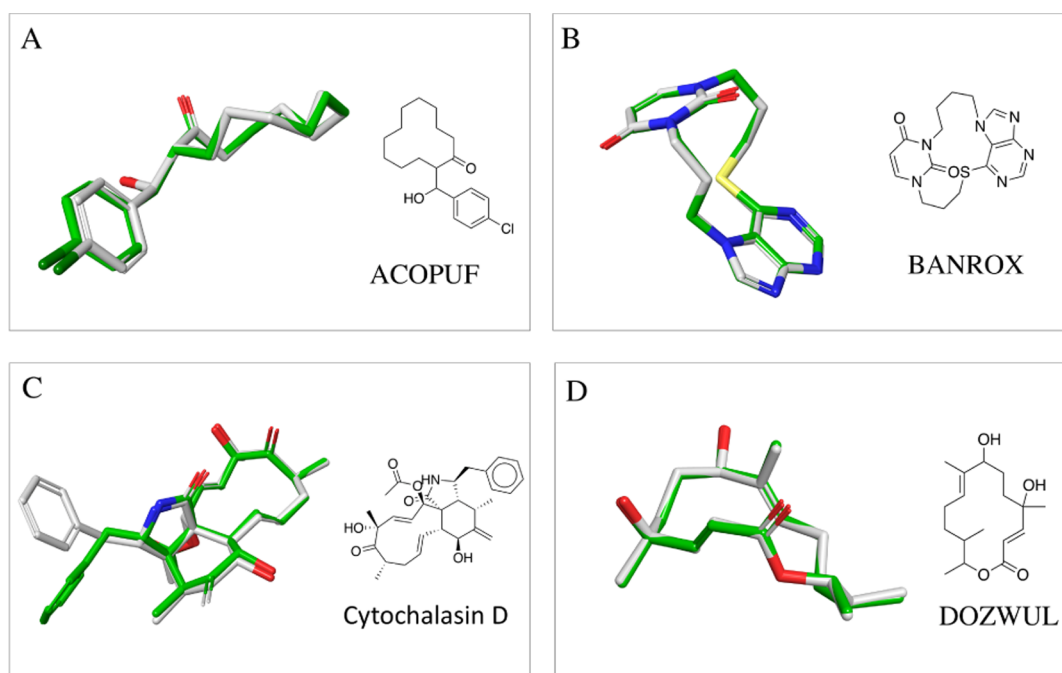


Figure 6. Examples of macrocycles having a flexibility of 10–19-atom backbone and indication by their dataset identifier (A–D). The atoms of the crystallographic structure to which the lower RMSD conformer has been aligned are colored in gray, whereas those of the conformer predicted using Moloc are in green.

BIRD database. On the one hand, Moloc reproduced 7 entries with high accuracy (<0.5 Å) and 38 with accuracy <1.0 Å, with the best being DEMJAG10 (Figure 7A) and kabiramide C (Figure 7B), two macrocycles of 22 and 25 ring size from the CSD and PDB dataset, whose closest coordinates to the bioactive molecule were 0.13 and 0.17 Å $\text{RMSD}_{\text{backbone}}$, respectively. Despite producing 789 and 172 conformations, Moloc remained superior to Prime, for which the closest coordinates for the two referred macrocycles were 0.82 and 0.35 Å, respectively (1000 conformations per entry). On the other hand, it is also interesting to assess the robustness of Moloc in generating accurate conformations of the heavy atoms. In that respect, only 11 crystal structures resulted in an interval of $\text{RMSD}_{\text{heavy atoms}} < 1.0$ Å—mostly belonging to the CSD (10) with only one from the PDB dataset (Figure 7C). Among these macrocycles, it is noteworthy to mention WURVEL (Figure 7D), a 27-membered ring entry from the CSD database, whose closest atomic coordinates (1.0 Å) indeed were not dissimilar from those predicted using Prime

(1.06 Å); nevertheless, Moloc produced 163 conformations while Prime produced 983.

>30-Ring-Sized Macrocyces. Highly flexible macrocycles represent a challenge for every conformational algorithm, given the large number of rotatable bonds and possible values of torsional angles around the ring. Another problem is the number of replacements that attach to the ring and their degree of branching. In this subset, a total of 24 crystalline structures can be found and, specifically, 5 are cross-linked and another 5 are cyclopeptides that were originally included by the Prime developers in order to make the benchmark more challenging. Five macrocycles, all belonging to the CSD database, appeared in the list predicted with $\text{RMSD}_{\text{backbone}} < 1.0$ Å. Among them, Moloc predicted the crystallographic coordinates of OCERET (Figure 8A), a 35-atom backbone macrocycle, with an $\text{RMSD}_{\text{backbone}}$ of 1.04 Å with 168 conformations. On comparison, Prime performed slightly better with 0.83 Å but produced 957 conformations. Only SUMMOC (Figure 8B) and LENPEA (Figure 8C) were

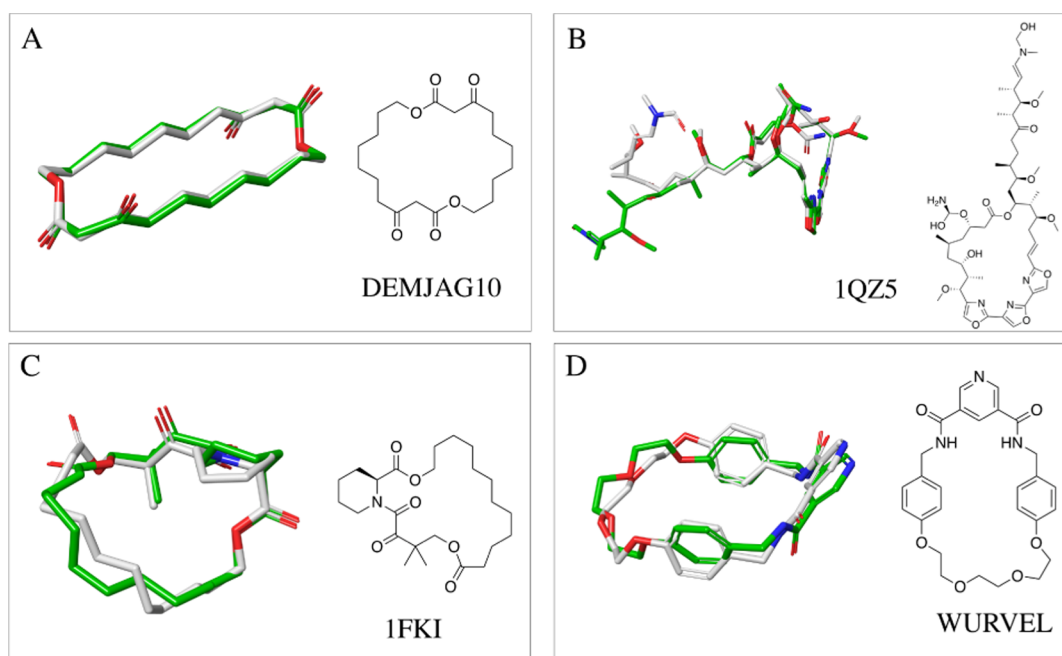


Figure 7. Examples of macrocycles having a flexibility of 20–29-atom backbone and their dataset identifier (A–D). The atoms of the crystallographic structure to which the lower RMSD conformer has been aligned are colored in gray, whereas those of the conformer predicted using Moloc are in green.

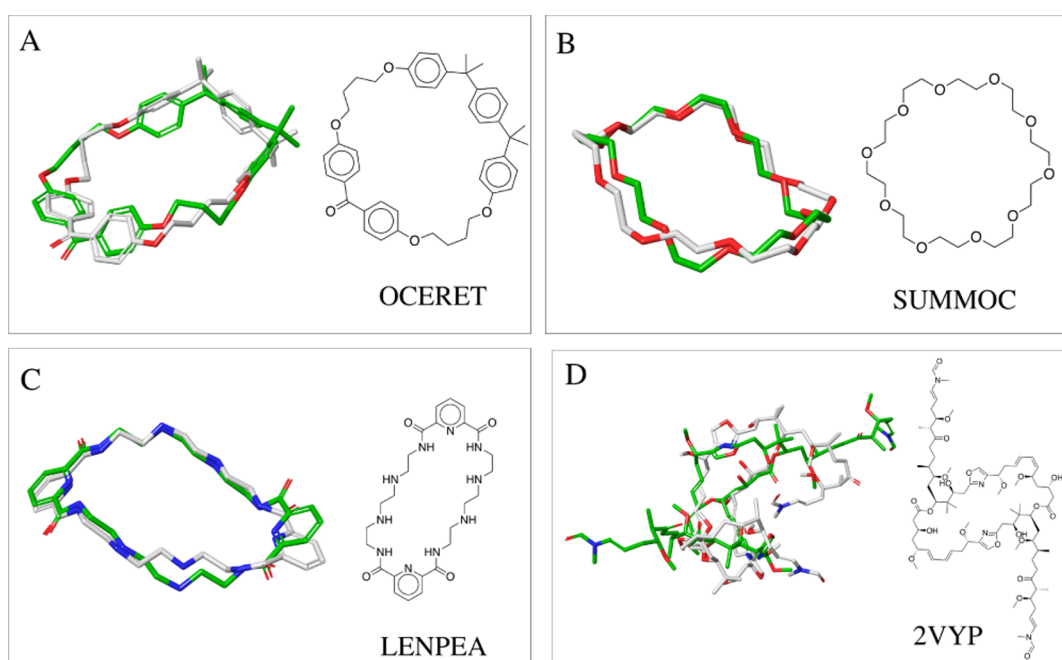


Figure 8. Examples of macrocycles indicated by their dataset identifier (A–D). The atoms of the crystallographic structure to which the lower RMSD conformer has been aligned are colored in gray, whereas those of the conformer predicted using Moloc are in green.

predicted below the threshold of 1.0 Å with values of $\text{RMSD}_{\text{heavy atoms}}$ of 0.74 and 0.92 Å, respectively. In addition to the advantage of Moloc being able to handle large-sized macrocycles, we noticed a limitation of Moloc in the complexity of the functional groups—expressed in terms of degree of branching. An example of this limit is shown in Figure 8D. The measured $\text{RMSD}_{\text{heavy atoms}}$ of (–)-rhizopodin (PDB: 2VYP), a potent actin-binding anticancer molecule,⁶⁴ decreases from 6.444 to 1.49 Å upon pruning the lateral substituents. This evidence can be explained by the ability of

Prime to randomly cleave the macrocycle and reconnect the two generated semiloops.

Intramolecular Interactions. The ideal software is required to predict intramolecular interactions as it is generally appreciated that they play a pivotal role in defining both overall shape of a molecule⁶⁵ and the stabilization of the functional groups by masking or exposing them to the external environment.⁶⁶ This change regulates the passive membrane permeability of macrocycles which adopt a globular shape while passing through the lipidic environment of the

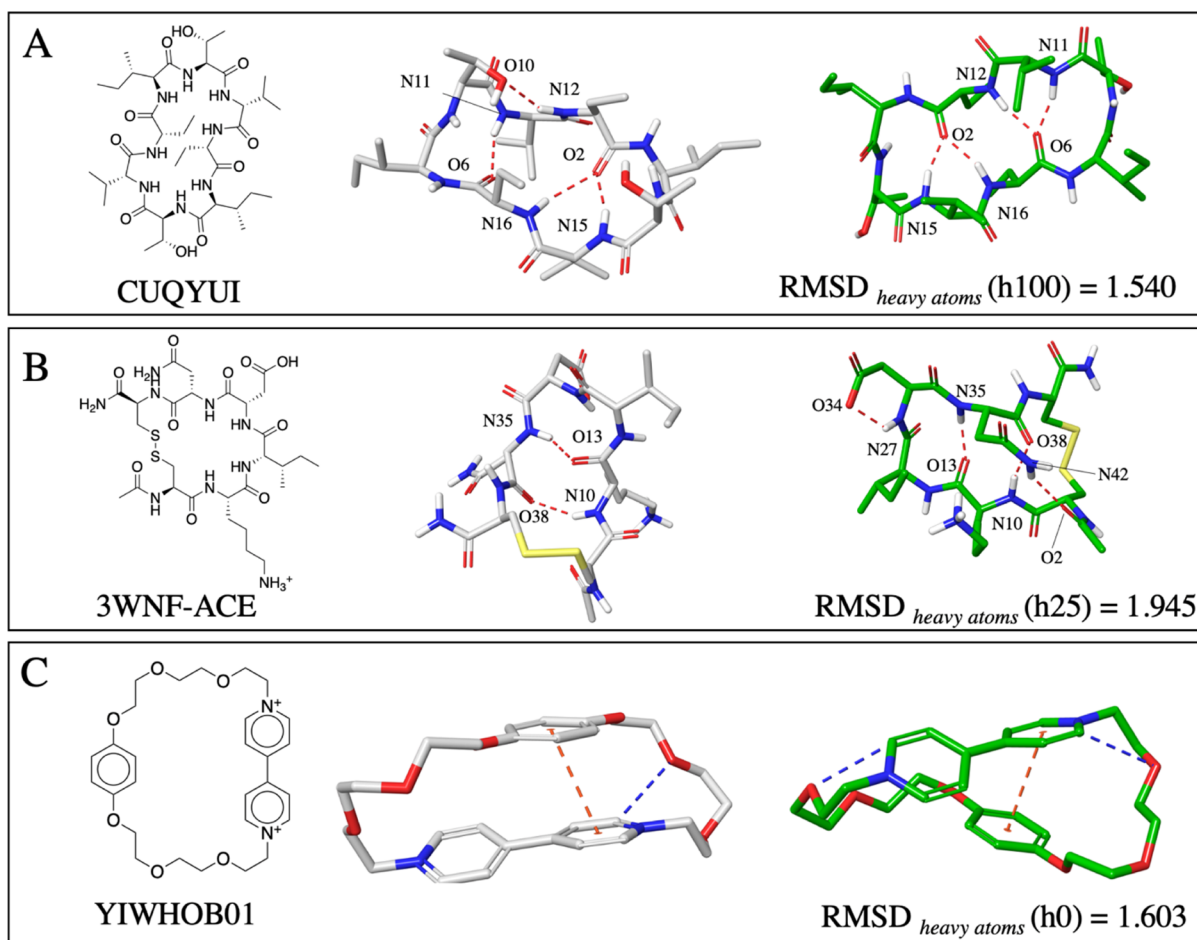


Figure 9. Panel showing the intramolecular interactions predicted using Moloc (green sticks) for (A) CUQYUI, (B) 3WNF-ACE, and (C) YIWHOB0 alongside with the RMSD_{heavy atoms} calculated for the hydrogen bond weight applied in the MAB force field. Hydrogen bonds, π -stacking, and aromatic hydrogen bonds are, respectively, colored as red, blue, and orange dotted lines while the crystal structure atoms are represented as gray sticks.

membrane and adopt a stretched conformation in the cytosol/extracellular environment.⁴⁵ Knowledge of the chameleonic properties of macrocycles has recently expanded far beyond the historical case of cyclosporin A.^{67,68}

As exemplified by the crystal structures of cyclosporin A in chloroform (CSD ID P2₁2₁2₁) and in the protein bound form (PDB ID: 2X2C⁶⁹), the conformational change is followed by the formation of new intramolecular hydrogen bonds, underlying their role in the dynamics of binding. As can be seen in Figure 9A, the crystal structure of CUQYUI, the 24-atoms backbone of the non-cross-linked cyclopeptide has 4 internal hydrogen bonds (between N15 and O2, N16 and O2, and O6 and N11) as well as one transannular interaction between N12 and O10).

Moloc successfully predicted three of these internal hydrogen bonds with an RMSD_{heavy atoms} of 1.365 Å and, most notably, matched the lowest global minimum among the 38 local minima, with a potential energy of 5.33 kcal/mol. 3WNF-ACE (Figure 9B) is a 20-atom backbone hexacyclic peptide whose binding affinity for HIV-1 integrase was measured in the low millimolar range by surface plasmon resonance and HSQC-NMR while the binding mode with the target was confirmed by X-ray crystallography.⁷⁰ Visual inspection of the cocrystal structure revealed the presence of two internal hydrogen bonds between N35 and O13, and N10

and O38 and two transannular interactions, between O34 and N27, and O2 and N10. Moloc was able to predict three of these four interactions with reasonable accuracy (RMSD_{heavy atom} = 1.945 Å) and a local minimum with a potential energy of 11.13 kcal/mol. YIWHOB01 (Figure 9C) is a 30-atom backbone non-cross-linked artificial macrocycle used as a charge transfer system in the field of supramolecular chemistry.⁷¹ Visual inspection of the CSD structure revealed the presence of a π -stacking interaction between the pyridine and phenyl rings. Again, Moloc predicted the conformation with the bipyridinium units being parallel to the phenyl ring with an RMSD_{heavy atom} of 1.642 Å and a potential energy of 9.846 kcal/mol, despite minor deviations at the dioxaryl moiety.

User-Defined Energy Threshold for Improved Accuracy and Diversity. In a standard Moloc conformational job, the structures are only kept if their energy is less than 10 kcal/mol above the lowest-energy conformation. Such an energetic cutoff is typical for many other conformational software. However, Prime sets the cutoff to 100 kcal/mol. Thus, we have quantified the diversity and the accuracy at 100 kcal/mol and chose 4MNW and 4KEL, two cyclopeptides, cross-linked macrocycles with 42-atom backbone. Based on our data (Table S6), no improvement over the diversity was observed independently from the chosen threshold because the number

of unique fingerprints for 4MNW (192) and 4KEL (290) remained unchanged. However, when the energy threshold was increased to 100 kcal/mol, Moloc produced new conformers with expanded globularity because the span RoG increased from 1.179 to 1.660 Å for 4KEL and from 1.041 to 1.704 Å for 4MNW. Additionally, we observed a marginal improvement in both the ring and the heavy atom structure accuracies: -0.42 Å/ -0.23 Å (4MNW) and -0.22 Å/ -0.08 Å (4KEL) at 20 kcal/mol and -0.83 Å/ -0.76 Å (4MNW) and -0.25 Å/ -0.39 Å (4KEL) at 100 kcal/mol (Figure S2A). As the number of conformations for both cases exponentially increased (Figure S2B), the global minimum energy of the most accurate conformer of 4MNW displays an increase in the potential energy by 6 and 15 kcal/mol, whereas for 4KEL, the equivalent values were 8 and 5 kcal/mol (Figure S2C,D).

DISCUSSION

Computational screening of large virtual macrocycle libraries is an effective way to prioritize compounds for expensive and time-consuming synthesis in the laboratory. We have recently described convergent and short syntheses of macrocycles using MCR. One synthesis consisted of a short two-step assembly of macrocycles from cyclic anhydrides, diamines, oxo components (aldehydes and ketones), and isocyanides. Based on commercial availability of the building blocks, a very large chemical space is spanned: 20 (cyclic anhydrides) \times 20 (diamines) \times 1000 oxo components \times 1000 isocyanides = 400 million macrocycles. Computational generation of conformers for such a large chemical space requires fast and optimized software. Therefore, in this manuscript, we have benchmarked Moloc versus available commercial and freeware for their performance as defined by accuracy, speed, exhaustiveness, diversity, and SE.

Our results confirmed that Prime, MM, and MOE possess higher accuracy in reproducing both the heavy atoms and ring coordinates of the crystallographic macrocycle references. According to our results, conformational sampling with ETKDG algorithm could be improved by subsequent minimizations steps with MMFF94s but not UFF. This finding could be related to the existence of out-of-plane bending and dihedral torsion parameters to planarize certain types of delocalized trigonal N atoms applied by the MMFF94s force field, thus providing a better match to the reference crystal structures. However, UFF contains basic parameters for all types of atoms on hybridization and connectivity and thereby is able to parameterize the restricted patterns of dihedral angles and rotatable bonds, both present in macrocycles.⁴⁴ Nevertheless, these data lead us to suggest that the implementation of minimization steps employing specific force fields after conformational sampling of macrocycles would lead to improvements of sampling. For instance, the OPLS-2005 in Prime or MAB force field in Moloc represent the most accurate commercial and open software, respectively. Such an evidence could allow further analysis to study the effect of different force fields to improve macrocycle sampling. On the other hand, we show that the use of DG methods as ETKDG could be improved to generate conformers closely related to the crystal structures. In this sense, a modification to the ETKDG algorithm for macrocycle sampling has been recently published by the developer team of RDKit and will be available in the upcoming RDKit release 2020.03.⁴⁷ Along with a restriction in search space for macrocycles, the new implementations in ETKDG will include additional torsional-angle potentials to

describe small aliphatic rings and adapt the previously developed potentials for acyclic bonds to facilitate the sampling of macrocycles. Nevertheless, because of the novelty of this algorithm, more testing is needed to evaluate its capability in diverse and challenging macrocycle datasets, such as those presented in this work.

MD was performed only under solvated conditions⁴⁹ with no major improvement in generating high-quality conformers according to the SE value. However, other reported MD-based approaches using different simulation conditions have reported the importance of solvation for the generation of bioactive conformations of macrocycles.⁷² An enhanced sampling method has been reported using MD simulations that resulted in a reliable method to reproduce the experimentally determined structure of three macrocycles.⁷³ Nevertheless, the major drawback for MD-based methods relies on its low scalability of large and diverse macrocycle datasets. As a result, such methods can be an option when working with a limited number of macrocyclic structures but not for virtual screening approaches such as Prime, MM, Moloc, ETKDG, or other software reported here.

Although CCDC conformer generator was one of the most efficient software for conformer generation in terms of speed and exhaustiveness, it suffers a low rate of conformational sampling exploration as only one single conformer was generated for 37 structures. The most noticeable exception relies on 76 cases where the RMSD_{backbone} values were unrealistically lower than 0.1 Å and hence equal to the crystallographic reference. This behavior could be explained by a bias in the sampling of entries from CSD: the CCDC conformer generator assigns the crystallography coordinates prior to conformation sampling. The CCDC conformer generator uses bond lengths and valence angles taken from CCDC Mogul and one of its best strengths consists in the use of dynamic rotamer libraries that are automatically updated with new data inside of CCDC.^{74,75} However, although CCDC conformer generator has implemented strategies to deal with conformer generation of rings as set preclustered templates for isolated, fused, spiro-linked, and bridged ring systems,⁷⁵ there is no specific method regarding macrocyclic conformers yet described. For instance, in rings for which no template is obtainable from Mogul data, the templates are generated on the fly using rotamer distributions for cyclic bonds.^{74,75} If ring generation fails and no template structure can be generated, the ring conformation from the 3D input structure is used. According to our results, the conformational sampling with CCDC conformer generator for the CSD entries, bond lengths, and valence angles were taken from CCDC Mogul retrieving conformers with conformations close to the crystal structures. Thus, for the macrocycles not present in CSD database, the conformers were generated either from an on-the-fly template assignment or using the input coordinates. This could explain the lowest number of conformers generated per entry and the reduced number of unique TFs. Furthermore, the span in RoG values from CCDC conformer generator suggests a tendency to retain conformations with higher compaction in comparison with any other methods for macrocycle conformational sampling described here, thus omitting possible extended states. Taking these results together, the restricted usage of CCDC conformer generator within the macrocycle conformational sampling could lead to poor results in terms of conformational space exploration or even a lack of conformers, suggesting that this tool is useful only to generate conformers

for small molecules or for the assignment of crystallographic coordinates to macrocycle structures.

Overall, our analysis indicated conformator as the lowest efficiency conformational sampling software tested in this work. This tool showed one of the lowest exhaustiveness values among the studied methods, just below that of MD. The accuracy of conformator reproducing the macrocycle backbone is also the lowest and is also one of the slowest conformational sampling methods—generating structures with the lowest span in RoG of all methods tested. Nevertheless, the authors of conformator have tested this algorithm employing 49 different macrocyclic structures.⁴⁶ These evidences suggest that the use of conformator could be restricted to small-to-medium macrocycles. Further analysis and testing are needed to assess the feasibility of conformator in generating conformers for a dataset containing large and complex structures. Furthermore, this software produces conformations that differ from each other by rotation of one single bond at a time which may limit its use to macrocycle with few rotatable bonds.

As for Moloc, we are indeed aware that reproducing the accuracy of all heavy atoms, as our $\text{RMSD}_{\text{heavy atoms}}$ data demonstrate, represents its main limitation. However, we would like to emphasize that one of the main challenges in the conformational analysis of macrocycles is the accuracy of ring atoms. Based on our $\text{RMSD}_{\text{backbone}}$ data, Moloc has a similar accuracy to the negative control (MD) and MD, Moloc, and ETKDG alone or in combination with MMFF94s, implying that it can be used as a valid alternative to these two methodologies to produce conformations with a similar accuracy. Most importantly, Moloc retains good exhaustiveness, SE, and economy in terms of least numbers of conformers to generate high quality conformers without requiring 1000 or more conformers for the exhaustive exploration of the chemical space, saving computational resources and avoiding redundancy in the conformers generated, suggesting this software as an acceptable alternative to Prime, MM, and MD for sampling. One major drawback of Moloc is that it relies on the number of symmetry elements within the macrocycle structure needed for the sampling. This is particularly evident in the case of POGLIH, a macrocycle from the CSD, for which 5 days were necessary to complete the conformational sampling. Indeed, the enumeration of topological symmetries is intended to avoid the counting of identical conformations that vary only by altered atom-numbering (e.g., 180° rotation of a phenyl ring in the structure). Such enumeration takes an (exponentially) increasing time in accordance with the number of symmetry elements. For POGLIH, all 8 phenyl rings can be rotated, and methyl groups can be exchanged, as well as oxygen in the sulfates. In addition, the whole structure has a twofold symmetry. All in all, there are over 32,000 symmetry elements present, meaning that the same conformation may occur 32,000 times—indicating that a threshold or restricted search of symmetries and their calculation could improve the speed of sampling. Another limitation of Moloc consists in sampling macrocycles with complex side chains: this has been seen in rhizopodin (PDB: 2VYP), a potent actin-binding anticancer agent.⁶⁴ Aiming to understand the relation between the accuracy and the side-chain complexity, we first trimmed the two 15-atom-branched symmetrical side chains of rhizopodin and subsequently sampled again the macrocycle (Figure S1). As a result, we observed an improvement of heavy atom

accuracy (from 6.27 to 2.17 Å) and an increased number of conformers (increasing from 62 to 205).

Nevertheless, several parameters allow the user a full control of the output ensembles, making Moloc a flexible piece of software for the molecular modeling of macrocycles. Our data indicate that the number of ensembles can be interactively controlled by applying either by energy thresholds (parameter “e”) or hydrogen bond weight (parameter “h”) term in the batch mode, allowing the enumeration of globular or flat conformations, the identification of intramolecular hydrogen bonds, and potentially predicting the most accurate ones in nonpolar environments. Taken altogether, these applications of Moloc indeed represent a “nice-to-have” tool in the molecular modeling toolkit of permeable macrocycles. Not lastly, the user can decide whether to apply a final energy minimization after conformational sampling followed by the addition of hydrogens to heteroatoms by invoking the parameter “q1”. As a result, Moloc returns all the energetic components calculated by MAB per conformer produced, bonds, valence angles, torsions, pyramidalities, 1–4 repulsion, van der Waals interactions, hydrogen bonds, and polar repulsion. To our knowledge, recent algorithms were published with already built-in protocols including the maximum ensemble size, RMSD or energy thresholds, and further constrains such as NMR data, enforcement of the chirality, geometry check before sampling, and application of a filter to retain the conformers according to a certain *R* value of the crystal structures.^{38,46,49,76} MM presents indeed the advantage of tuning several parameters such as electrostatic treatment and possibility to choose two different force fields (OPLS-2005 or MMFF94s).³⁹ In the case of open-access software, such as ETKDG, recently, new improvements were released in order to favor certain interactions or orientation angles.⁴⁸ Additionally, we would like to point out that CCDC conformer generator as well as ETKDG and conformator are knowledge-based systems with pre-existing rotational libraries of small-medium rings. This implies that if a test set entry is derived from the CSD, it will have prior information and make use of these coordinates. Nevertheless, CSD entries were retained in knowledge-based systems.

Finally, a possible strategy to improve the accuracy of complex macrocycles could be the implementation of further shape constrains accounting for the crystallographic packing forces—because most of the macrocyclic crystal structures are flattened in a high-energy level conformation.

Additional improvement of Moloc should also consider the flexibility of the complex side chains because the current version of the algorithm starts the identification of the first generic shape from a polar coordinate of a circle with an acceptable degree of accuracy and time.

CONCLUSIONS

In this work, we have benchmarked the shape-guided algorithm using a dataset of 208 macrocycles from Prime publication, carefully selected on the basis of structural complexity (e.g., ring size, cyclopeptide/aliphatic, cross-linkings) and we have quantified accuracy, diversity, speed, exhaustiveness, and SE with four conformational commercial (Prime, MM, MOE, and MD) and five open-access (ETKDG, MMFF94s, UFF, CCDC, and conformator) software packages. A Python script to streamline the whole data collection of these parameters has been written *ad hoc*. The results of our benchmark are summarized in Table 4. Although Prime, MM,

Table 4. Summary Table of the Benchmark^a

methodology	Prime	MM	MOE	MD	Moloc	conformator	ETKDG	MMFF94s	UFF	CCDC
RMSD _{heavy atoms} (Å)	0.878	0.655	0.765	1.052	1.910	1.990	2.165	1.793	2.083	2.067
RMSD _{backbone} (Å)	0.396	0.383	0.417	0.562	0.652	0.801	0.743	0.668	0.766	0.476
number of conformations	972	300	76	1000	67	338	1000	998	535	8
TF	707	100	48	59	67	338	1000	998	535	8
span RoG (Å)	1.02	0.93	0.74	0.85	0.86	0.87	0.82	1.13	1.08	0.15
exhaustiveness	1.00	1.00	1.00	1.00	1.00	0.91	1.00	0.95	0.95	1.00
SE	0.76	0.33	0.63	0.06	1.00	0.66	1.00	0.95	0.95	0.75
speed	9.8 min	3.9 h	31.1 min	3.1 d	38.9 min	17.9 h	35.1 s	1.3 min	17.6 s	2.6 s

^aData are medians.

MOE, and MD remained the most accurate software tested in this paper in reproducing macrocycle heavy atoms, Moloc retained the same exhaustiveness. However, Moloc stood out for the highest SE in producing an acceptable number of conformations per entry and three-quarters of the database were processed with high accuracy (RMSD_{backbone} < 1.0 Å). Interactive control of the hydrogen bond terms allows the enumeration of globular and flat conformers and prediction of intramolecular interaction in a nonpolar solvent. However, the structural accuracy of Moloc is hampered by long-branched side chains. In that respect, side chain pruning in the batch mode with “Mdfy”, a built-in module within Moloc, and subsequent reattachment to the ring could be an option for future improvement. Surprisingly, minimization with UFF and MMFF94s managed to produce macrocycles with the most diverse shapes in terms of RoG, suggesting these types of software as a valid free alternative for the prediction of the most likely shape that the macrocycles can adopt in their bulk environment, for example, the cellular membrane or water. Follow-up studies could include modifications to ETKDG algorithm or the use of force field minimization in order to predict the X-ray structure. For instance, the evaluation of ETKDG conformational sampling was combined with OPLS-2005 and/or MAB as minimization methods.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.0c01038>.

Full datasets of results from Moloc, experimental-torsion DG with additional “basic knowledge” alone and in combination with Merck molecular force field/UFF, conformator, and CCDC software can be found in Data sets.zip file; speed data can be retrieved from ConforTime.zip file; pairwise Krustal–Wallis *H*-test calculated for the median RMSD_{heavy atoms}; RMSD_{backbone}; TFs, span RoG, speed, and crystal structure of rhizopodin; summary table of complex macrocycles predicted using Moloc at 100 kcal/mol as energy cutoff and related bar plots of the number of conformations, accuracy, and energy calculated for 10/20 kcal/mol thresholdscan be found in the Supporting Information.docx; detailed installation of Moloc can be found at <http://www.moloc.ch/installation.html>; and Python script for calculation of accuracy, diversity, TFs, the number of conformers can be found at https://github.com/AngelRuizMoreno/Macro_analyzer (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Alexander Dömling – Drug Design, Department of Pharmacy, University of Groningen, 9713 AV Groningen, The Netherlands; orcid.org/0000-0002-9923-8873; Phone: +31 50 36 33307; Email: a.s.s.domling@rug.nl

Authors

Atilio Reyes Romero – Drug Design, Department of Pharmacy, University of Groningen, 9713 AV Groningen, The Netherlands

Angel Jonathan Ruiz-Moreno – Drug Design, Department of Pharmacy, University of Groningen, 9713 AV Groningen, The Netherlands; Departamento de Farmacología y Unidad Periférica de Investigación en Biomedicina Trasnacional, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM), 04510 Ciudad de México, Mexico; Programa de Doctorado en Ciencias Biomédicas, UNAM, 04510 Ciudad de México, Mexico

Matthew R. Groves – Drug Design, Department of Pharmacy, University of Groningen, 9713 AV Groningen, The Netherlands; orcid.org/0000-0001-9859-5177

Marco Velasco-Velázquez – Departamento de Farmacología y Unidad Periférica de Investigación en Biomedicina Trasnacional, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM), 04510 Ciudad de México, Mexico

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.0c01038>

Author Contributions

A.R.R. and A.J.R.-M. contributed equally. All authors have given approval to the final version of the manuscript.

Funding

A.J.R.-M. would like to acknowledge scholarship CONACYT grant number 584534. This research has been supported by the National Institute of Health (NIH) (2R01GM097082-05), the European Lead Factory (IMI) (grant agreement number 115489), the Qatar National Research Foundation (NPRP6-065-3-012), COFUNDs ALERT (grant agreement no. 665250), Prominent (grant agreement no. 754425), KWF Kankerbestrijding grant (grant agreement no. 10504), and PAPIIT UNAM IN219719. This project is funded from the European Union’s Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie grant agreement no. 675555, accelerated early-stage drug discovery (AEGIS).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to Paul Gerber for constructive discussion, implementing the generic shape algorithm in the current version of Mcfnf, and correction of minor bugs.

■ ABBREVIATIONS

CCR2, CC chemokine receptor 2; CCL2, CC chemokine ligand 2; CCR5, CC chemokine receptor 5; TLC, thin layer chromatography

■ REFERENCES

- (1) Frank, A. T.; Farina, N. S.; Sawwan, N.; Wauchope, O. R.; Qi, M.; Brzostowska, E. M.; Chan, W.; Grasso, F. W.; Haberfield, P.; Greer, A. Natural Macrocyclic Molecules Have a Possible Limited Structural Diversity. *Mol. Diversity* **2007**, *11*, 115–118.
- (2) Hill, T. A.; Shepherd, N. E.; Diness, F.; Fairlie, D. P. Constraining Cyclic Peptides To Mimic Protein Structure Motifs. *Angew. Chem., Int. Ed.* **2014**, *53*, 13020–13041.
- (3) D'Souza, V. T.; Lipkowitz, K. B. Cyclodextrins: Introduction. *Chem. Rev.* **1998**, *98*, 1741–1742.
- (4) Palei, S.; Mootz, H. D. Preparation of Semisynthetic Peptides Macrocycles Using Split Inteins. *Methods Mol. Biol.* **2017**, *1495*, 77–92.
- (5) Kwitkowski, V. E.; Prowell, T. M.; Ibrahim, A.; Farrell, A. T.; Justice, R.; Mitchell, S. S.; Sridhara, R.; Pazdur, R. FDA Approval Summary: Temsirolimus as Treatment for Advanced Renal Cell Carcinoma. *Oncologist* **2010**, *15*, 428–435.
- (6) Raymond, E.; Alexandre, J.; Faivre, S.; Vera, K.; Materman, E.; Boni, J.; Leister, C.; Korth-Bradley, J.; Hanauske, A.; Armand, J.-P. Safety and Pharmacokinetics of Escalated Doses of Weekly Intravenous Infusion of CCI-779, a Novel MTOR Inhibitor, in Patients With Cancer. *J. Clin. Oncol.* **2004**, *22*, 2336–2347.
- (7) Goodin, S. Novel Cytotoxic Agents: Etoposides. *Am. J. Health-Syst. Pharm.* **2008**, *65*, S10–S15.
- (8) Goodin, S. Ixabepilone: A Novel Microtubule-Stabilizing Agent for the Treatment of Metastatic Breast Cancer. *Am. J. Health-Syst. Pharm.* **2008**, *65*, 2017–2026.
- (9) Stotani, S.; Giordanetto, F. Overview of Macrocycles in Clinical Development and Clinically Used. In *Practical Medicinal Chemistry with Macrocycles*; John Wiley & Sons, Ltd., 2017; pp 411–499.
- (10) Pedersen, C. J. The Discovery of Crown Ethers. *Science* **1988**, *241*, 536–540.
- (11) Batten, S. R.; Robson, R. Catenanes and Rotaxanes Motifs in Interpenetrating and Self-Penetrating Coordination Polymers. In *Molecular Catenanes, Rotaxanes and Knots*; John Wiley & Sons, Ltd., 2007; pp 77–106.
- (12) Yudin, A. K. Macrocycles: Lessons from the Distant Past, Recent Developments, and Future Directions. *Chem. Sci.* **2014**, *6*, 30–49.
- (13) Marsault, E.; Peterson, M. L. Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. *J. Med. Chem.* **2011**, *54*, 1961–2004.
- (14) 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.
- (15) Mallinson, J.; Collins, I. Macrocycles in New Drug Discovery. *Future Med. Chem.* **2012**, *4*, 1409–1438.
- (16) Dougherty, P. G.; Qian, Z.; Pei, D. Macrocycles as Protein-Protein Interaction Inhibitors. *Biochem. J.* **2017**, *474*, 1109–1125.
- (17) Bell, I. M.; Gallicchio, S. N.; Abrams, M.; Beese, L. S.; Beshore, D. C.; Bhimnathwala, H.; Bogusky, M. J.; Buser, C. A.; Culberson, J. C.; Davide, J.; Ellis-Hutchings, M.; Fernandes, C.; Gibbs, J. B.; Graham, S. L.; Hamilton, K. A.; Hartman, G. D.; Heimbrook, D. C.; Homnick, C. F.; Huber, H. E.; Huff, J. R.; Kassahun, K.; Koblan, K. S.; Kohl, N. E.; Lobell, R. B.; Lynch, J. J.; Robinson, R.; Rodrigues, A. D.; Taylor, J. S.; Walsh, E. S.; Williams, T. M.; Zartman, C. B. 3-Aminopyrrolidinone Farnesyltransferase Inhibitors: Design of Macro-

cyclic Compounds with Improved Pharmacokinetics and Excellent Cell Potency. *J. Med. Chem.* **2002**, *45*, 2388–2409.

- (18) Leung, S. S. F.; Sindhikara, D.; Jacobson, M. P. Simple Predictive Models of Passive Membrane Permeability Incorporating Size-Dependent Membrane-Water Partition. *J. Chem. Inf. Model.* **2016**, *56*, 924–929.

- (19) 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.

- (20) 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.

- (21) Giordanetto, F.; Kihlberg, J. Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties? *J. Med. Chem.* **2014**, *57*, 278–295.

- (22) Dömling, A. Small Molecular Weight Protein-Protein Interaction Antagonists: An Insurmountable Challenge? *Curr. Opin. Chem. Biol.* **2008**, *12*, 281–291.

- (23) Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J. Oral Druggable Space beyond the Rule of 5: Insights from Drugs and Clinical Candidates. *Chem. Biol.* **2014**, *21*, 1115–1142.

- (24) Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; Porco, J. A.; Kozakov, D.; Vajda, S.; Whitty, A. How Proteins Bind Macrocycles. *Nat. Chem. Biol.* **2014**, *10*, 723–731.

- (25) Beck, B.; Larbig, G.; Mejat, B.; Magnin-Lachaux, M.; Picard, A.; Herdtweck, E.; Dömling, A. Short and Diverse Route Toward Complex Natural Product-Like Macrocycles. *Org. Lett.* **2003**, *5*, 1047–1050.

- (26) Liao, G. P.; Abdelraheem, E. M. M.; Neochoritis, C. G.; Kurpiewska, K.; Kalinowska-Thuścik, J.; McGowan, D. C.; Dömling, A. Versatile Multicomponent Reaction Macrocyclic Synthesis Using α -Isocyno- ω -Carboxylic Acids. *Org. Lett.* **2015**, *17*, 4980–4983.

- (27) Madhavachary, R.; Abdelraheem, E. M. M.; Rossetti, A.; Twarda-Clapa, A.; Musielak, B.; Kurpiewska, K.; Kalinowska-Thuścik, J.; Holak, T. A.; Dömling, A. Two-Step Synthesis of Complex Artificial Macrocyclic Compounds. *Angew. Chem., Int. Ed. Engl.* **2017**, *56*, 10725–10729.

- (28) Vishwanatha, T. M.; Bergamaschi, E.; Dömling, A. Sulfur-Switch Ugi Reaction for Macrocyclic Disulfide-Bridged Peptidomimetics. *Org. Lett.* **2017**, *19*, 3195–3198.

- (29) Abdelraheem, E.; Shaabani, S.; Dömling, A. Artificial Macrocycles. *Synlett* **2018**, *29*, 1136–1151.

- (30) Wang, W.; Groves, M. R.; Dömling, A. Artificial Macrocycles as IL-17A/IL-17RA Antagonists. *Medchemcomm* **2018**, *9*, 22–26.

- (31) Magiera-Mularz, K.; Skalniak, L.; Zak, K. M.; Musielak, B.; Rudzinska-Szostak, E.; Berlicki, L.; Kocik, J.; Grudnik, P.; Sala, D.; Zarganes-Tzitzikas, T.; Shaabani, S.; Dömling, A.; Dubin, G.; Holak, T. A. Bioactive Macrocyclic Inhibitors of the PD-1/PD-L1 Immune Checkpoint. *Angew. Chem., Int. Ed.* **2017**, *56*, 13732–13735.

- (32) Neochoritis, C. G.; Kazemi Miraki, M.; Abdelraheem, E. M. M.; Surmiak, E.; Zarganes-Tzitzikas, T.; Łabuzek, B.; Holak, T. A.; Dömling, A. Design of Indole- and MCR-Based Macrocycles as P53-MDM2 Antagonists. *Beilstein J. Org. Chem.* **2019**, *15*, 513–520.

- (33) Estrada-Ortiz, N.; Neochoritis, C. G.; Twarda-Clapa, A.; Musielak, B.; Holak, T. A.; Dömling, A. Artificial Macrocycles as Potent P53-MDM2 Inhibitors. *ACS Med. Chem. Lett.* **2017**, *8*, 1025–1030.

- (34) Kaserer, T.; Beck, K.; Akram, M.; Odermatt, A.; Schuster, D. Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules* **2015**, *20*, 22799–22832.

- (35) Spellmeyer, D. C.; Wong, A. K.; Bower, M. J.; Blaney, J. M. Conformational Analysis Using Distance Geometry Methods. *J. Mol. Graphics Modell.* **1997**, *15*, 18–36.

- (36) Coutias, E. A.; Lexa, K. W.; Wester, M. J.; Pollock, S. N.; Jacobson, M. P. Exhaustive Conformational Sampling of Complex

Fused Ring Macrocyces Using Inverse Kinematics. *J. Chem. Theory Comput.* **2016**, *12*, 4674–4687.

(37) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462–2474.

(38) Labute, P. LowModeMD—Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocyces and Protein Loops. *J. Chem. Inf. Model.* **2010**, *50*, 792–800.

(39) Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. Macrocycle Conformational Sampling with MacroModel. *J. Chem. Inf. Model.* **2014**, *54*, 2680–2696.

(40) Olanders, G.; Alogheli, H.; Brandt, P.; Karlén, A. Conformational Analysis of Macrocyces: Comparing General and Specialized Methods. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 231–252.

(41) Vulis, M. Ring Structures and the Discrete Fourier Transform. *Adv. Appl. Math.* **1985**, *6*, 350–372.

(42) Gerber, P. R.; Gubernator, K.; Müller, K. Generic Shapes for the Conformation Analysis of Macrocylic Structures. *Helv. Chim. Acta* **1988**, *71*, 1429–1441.

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

(44) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(45) Whitty, A.; Zhong, M.; Viarengo, L.; Beglov, D.; Hall, D. R.; Vajda, S. Quantifying the Chameleonic Properties of Macrocyces and Other High-Molecular-Weight Drugs. *Drug Discovery Today* **2016**, *21*, 712–717.

(46) Friedrich, N.-O.; Flachsenberg, F.; Meyder, A.; Sommer, K.; Kirchmair, J.; Rarey, M. Conformer: A Novel Method for the Generation of Conformer Ensembles. *J. Chem. Inf. Model.* **2019**, *59*, 731–742.

(47) Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S. Improving Conformer Generation for Small Rings and Macrocyces Based on Distance Geometry and Experimental Torsional-Angle Preferences. *J. Chem. Inf. Model.* **2020**, *60*, 2044–2058.

(48) Taylor, R.; Cole, J.; Korb, O.; McCabe, P. Knowledge-Based Libraries for Predicting the Geometric Preferences of Druglike Molecules. *J. Chem. Inf. Model.* **2014**, *54*, 2500–2514.

(49) Sindhikara, D.; Spronk, S. A.; Day, T.; Borrelli, K.; Cheney, D. L.; Posy, S. L. Improving Accuracy, Diversity, and Speed with Prime Macrocycle Conformational Sampling. *J. Chem. Inf. Model.* **2017**, *57*, 1881–1894.

(50) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171–179.

(51) 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.

(52) Gerber, P. R. Topological Pharmacophore Description of Chemical Structures Using MAB-Force-Field-Derived Data and Corresponding Similarity Measures. In *Fundamentals of Molecular Similarity*; Carbó-Dorca, R., Gironés, X., Mezey, P. G., Eds.; Mathematical and Computational Chemistry; Springer US: Boston, MA, 2001; pp 67–81.

(53) Gerber, P. R.; Müller, K. MAB, a Generally Applicable Molecular Force Field for Structure Modelling in Medicinal Chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 251–268.

(54) Cole, J. C.; Korb, O.; McCabe, P.; Read, M. G.; Taylor, R. Knowledge-Based Conformer Generation Using the Cambridge Structural Database. *J. Chem. Inf. Model.* **2018**, *58*, 615–629.

(55) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the Performance of 3D Virtual Screening Protocols: RMSD Comparisons, Enrichment Assessments, and Decoy Selection—What Can We Learn from Earlier Mistakes? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 213–228.

(56) Friedrich, N.-O.; de Bruyn Kops, C.; Flachsenberg, F.; Sommer, K.; Rarey, M.; Kirchmair, J. Benchmarking Commercial Conformer Ensemble Generators. *J. Chem. Inf. Model.* **2017**, *57*, 2719–2728.

(57) Bai, F.; Liu, X.; Li, J.; Zhang, H.; Jiang, H.; Wang, X.; Li, H. Bioactive Conformational Generation of Small Molecules: A Comparative Analysis between Force-Field and Multiple Empirical Criteria Based Methods. *BMC Bioinf.* **2010**, *11*, S45.

(58) Schulz-Gasch, T.; Schärfer, C.; Guba, W.; Rarey, M. TFD: Torsion Fingerprints as a New Measure to Compare Small Molecule Conformations. *J. Chem. Inf. Model.* **2012**, *52*, 1499–1512.

(59) Todeschini, R. Molecular Descriptors. *Recent Advances in QSAR Studies*; Springer, 2010.

(60) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.

(61) Cusack, R. M.; Grøndahl, L.; Abbenante, G.; Fairlie, D. P.; Gahan, L. R.; Hanson, G. R.; Hambley, T. W. Conformations of Cyclic Octapeptides and the Influence of Heterocyclic Ring Constraints upon Calcium Binding. *J. Chem. Soc., Perkin Trans. 2* **2000**, 323–331.

(62) Hendrickson, J. B. Molecular Geometry. IV. The Medium Rings. *J. Am. Chem. Soc.* **1964**, *86*, 4854–4866.

(63) Engler, E. M.; Andose, J. D.; Schleyer, P. V. R. Critical Evaluation of Molecular Mechanics. *J. Am. Chem. Soc.* **1973**, *95*, 8005–8025.

(64) Hagelueken, G.; Albrecht, S. C.; Steinmetz, H.; Jansen, R.; Heinz, D. W.; Kalesse, M.; Schubert, W.-D. The Absolute Configuration of Rhizopodin and Its Inhibition of Actin Polymerization by Dimerization. *Angew. Chem., Int. Ed.* **2009**, *48*, S95–S98.

(65) Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular Hydrogen Bonding in Medicinal Chemistry. *J. Med. Chem.* **2010**, *53*, 2601–2611.

(66) Alex, A.; Millan, D. S.; Perez, M.; Wakenhut, F.; Whitlock, G. A. Intramolecular Hydrogen Bonding to Improve Membrane Permeability and Absorption in beyond Rule of Five Chemical Space. *MedChemComm* **2011**, *2*, 669–674.

(67) Danelius, E.; Poongavanam, V.; Peintner, S.; Wieske, L. H. E.; Erdélyi, M.; Kihlberg, J. Solution Conformations Explain the Chameleonic Behaviour of Macrocylic Drugs. *Chem.—Eur. J.* **2020**, *26*, 5231–5244.

(68) Sebastiano, M. R.; Doak, B. C.; Backlund, M.; Poongavanam, V.; Over, B.; Ermondi, G.; Caron, G.; Matsson, P.; Kihlberg, J. Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5. *J. Med. Chem.* **2018**, *61*, 4189–4202.

(69) Lammers, M.; Neumann, H.; Chin, J. W.; James, L. C. Acetylation Regulates Cyclophilin A Catalysis, Immunosuppression and HIV Isomerization. *Nat. Chem. Biol.* **2010**, *6*, 331–337.

(70) Northfield, S. E.; Wielens, J.; Headey, S. J.; Williams-Noonan, B. J.; Mulcair, M.; Scanlon, M. J.; Parker, M. W.; Thompson, P. E.; Chalmers, D. K. Cyclic Hexapeptide Mimics of the LEDGF Integrase Recognition Loop in Complex with HIV-1 Integrase. *ChemMedChem* **2018**, *13*, 1555–1565.

(71) Pia, E.; Toba, R.; Chas, M.; Peinador, C.; Quintela, J. M. Synthesis of New Viologen Macrocyces with Intramolecular Charge Transfer. *Tetrahedron Lett.* **2006**, *47*, 1953–1956.

(72) Kamenik, A. S.; Kraml, J.; Hofer, F.; Waibl, F.; Quoika, P. K.; Kahler, U.; Schauerl, M.; Liedl, K. R. Macrocycle Cell Permeability Measured by Solvation Free Energies in Polar and Apolar Environments. *J. Chem. Inf. Model.* **2020**, *60*, 3508–3517.

(73) Kamenik, A. S.; Lessel, U.; Fuchs, J. E.; Fox, T.; Liedl, K. R. Peptidic Macrocyces - Conformational Sampling and Thermodynamic Characterization. *J. Chem. Inf. Model.* **2018**, *58*, 982–992.

(74) Cottrell, S. J.; Olsson, T. S. G.; Taylor, R.; Cole, J. C.; Liebeschuetz, J. W. Validating and Understanding Ring Conformations Using Small Molecule Crystallographic Data. *J. Chem. Inf. Model.* **2012**, *52*, 956–962.

(75) Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G. Retrieval of Crystallographically-Derived Molecular Geometry Information. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2133–2144.

(76) Cleves, A. E.; Jain, A. N. ForceGen 3D Structure and Conformer Generation: From Small Lead-like Molecules to Macrocyclic Drugs. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 419–439.