

Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences

Shuzhe Wang, Jagna Witek, Gregory A. Landrum, and Sereina Riniker*



Cite This: *J. Chem. Inf. Model.* 2020, 60, 2044–2058



Read Online

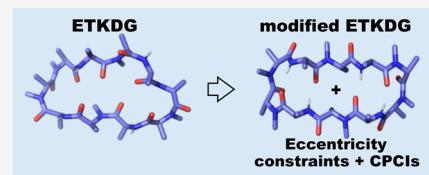
ACCESS |

Metrics & More

Article Recommendations

SI Supporting Information

ABSTRACT: The conformer generator ETKDG is a stochastic search method that utilizes distance geometry together with knowledge derived from experimental crystal structures. It has been shown to generate good conformers for acyclic, flexible molecules. This work builds on ETKDG to improve conformer generation of molecules containing small or large aliphatic (i.e., non-aromatic) rings. For one, we devise additional torsional-angle potentials to describe small aliphatic rings and adapt the previously developed potentials for acyclic bonds to facilitate the sampling of macrocycles. However, due to the larger number of degrees of freedom of macrocycles, the conformational space to sample is much broader than for small molecules, creating a challenge for conformer generators. We therefore introduce different heuristics to restrict the search space of macrocycles and bias the sampling toward more experimentally relevant structures. Specifically, we show the usage of elliptical geometry and customizable Coulombic interactions as heuristics. The performance of the improved ETKDG is demonstrated on test sets of diverse macrocycles and cyclic peptides. The code developed here will be incorporated into the 2020.03 release of the open-source cheminformatics library RDKit.



INTRODUCTION

The ability to generate 3D structures of molecules *in silico* is an integral part of drug discovery, may this be for 3D-QSAR,^{1,2} docking,³ or use as starting points for molecular dynamics-based studies,^{4,5} hence the need to generate relevant conformers. Many conformer generators have been developed over the years, including commercial products (some provide free academic licenses) such as OpenEye's Omega,^{6–8} Schrödinger's ConfGen, MacroModel and Prime,^{9–11} CSD Conformer Generator,^{12,13} MOE's LowModMD,¹⁴ BRIKARD,¹⁵ ForceGen 3D,¹⁶ TCG¹⁷ and Conformato¹⁸, and open-sourced tools such as CONFECT,¹⁹ BALLOON,²⁰ and the conformer generator implemented in RDKit.^{21,22} For a more comprehensive overview of the various conformer generation strategies, the reader is referred to refs 23 and 24 as well as different benchmarking studies (especially among the commercial toolkits).^{25–30}

The conformer generator in the RDKit utilizes stochastic search based on distance geometry (DG).³¹ DG is based on upper and lower distance bounds between all pairs of atoms in the molecule. A conformation is obtained by picking distances within the bounds for each atom pair followed by embedding. The upper and lower distance bounds are traditionally based on chemical knowledge and experimental data, although deep learning approaches begin to emerge.³² Recently, we proposed a variant of the DG method termed ETKDG that incorporates additional knowledge and experimental torsional-angle preferences for acyclic bonds obtained from the analysis of small-molecule crystal structures.²² The ability of ETKDG to sample

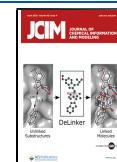
diverse and chemically-meaningful conformers and reproduce crystal conformations was improved compared to DG. Benchmarking studies have found ETKDG to be the best-performing freely available conformer generator.^{29,30}

In recent years, macrocycles have experienced a high resurgence of interest as therapeutic agents.^{33–37} The difficulty for the conformer generation of macrocycles lies in the dramatic increase of the size of the conformational space with an increasing number of atoms (curse of dimensionality).²⁴ Although all conformer generators listed above can handle macrocycles, only some of them are designed to sample macrocyclic structures well.^{8,10,11,13–16,18,26}

The experimental torsional-angle preferences in ETKDG are currently only for acyclic bonds. Therefore, we aim in this study to expand ETKDG for improved sampling of both small and large aliphatic rings. For molecules with small rings, we have developed additional torsional-angle potentials derived from the CSD³⁸ to describe the preferences of aliphatic cyclic bonds, analogous to those for acyclic bonds. For efficient sampling of macrocyclic conformers, we introduce customizable heuristics based on geometric and non-bonded interactions in order to bias the sampling toward specific regions of conformational

Received: January 9, 2020

Published: March 10, 2020



space. The use of heuristics has been found previously to improve the generation of relevant conformational ensembles.^{25,26} We demonstrate the performance of the new approach on a large and diverse set of macrocycles as well as on two specific examples of cyclic peptides. Cyclic peptides tend to have higher bioavailability than that of linear analogues, low toxicity, and restricted binding poses, making them prime drug candidates.^{39,40} For the simulation of new cyclic peptides in order to connect structure and passive membrane permeability,^{41–43} it is crucial to be able to generate relevant conformations.

The modified ETKDG is available in the docker image at github.com/rinikerlab/RDKit_mETKDG and will be incorporated into the 2020.03 release of the official RDKit.

METHODS

RDKit generates conformers using distance geometry (DG).³¹ In short, given a molecule, DG first defines a bound's matrix, which includes all the minimum and maximum atom pair distances in the molecule. Then, a stochastic process uniformly selects values between the maximum and minimum distances and yields all the atom pair distance criteria. Lastly, atomic coordinates are generated in order to satisfy these distances constraints. The experimental-torsion-knowledge distance geometry (ETKDG) variant described in ref 22 is used as the standard in our study. Its workflow is summarized in Figure 1. The improvements proposed in this study are added at different steps of the workflow, while the overall sequence is maintained.

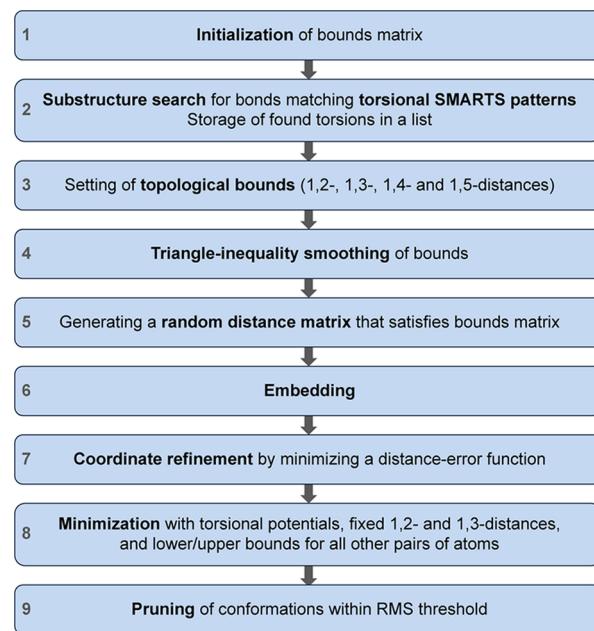


Figure 1. Workflow for the conformer generation with ETKDG²² in the RDKit. Adapted with permission from ref 22. Copyright 2015 American Chemical Society.

Van der Waals Radii. Recently, the atomic van der Waals (vdW) radii in the RDKit were updated to match those in the Blue Obelisk data repository.⁴⁴ These radii are used to calculate the lower bounds in the bounds matrix (Step 1 in Figure 1). We have assessed the effect on the conformer generator using the test sets in ref 22 (see the Supporting Information). The performance remains similar. Thus, the RDKit version with the

updated atomic vdW radii is used as the baseline in the following.

Small Rings. Torsional-Angle SMARTS Patterns. SMARTS patterns for aliphatic cyclic bonds were devised in an iterative procedure. In brief, generic patterns for all combinations of C, N, O, S, and P were used initially and subsequently refined, resulting in a total of 105 patterns (list provided in the Supporting Information). Patterns apply to bonds in rings with a maximum size of eight. Bonds in rings larger than eight have torsional preferences that are more similar to linear chains. This was verified by comparing the torsional-angle distributions using the developed patterns (Figure S2 in the Supporting Information). In addition, it was found that bridged ring systems show special torsional preferences while they are at the same time sterically well defined. Therefore, we decided to exclude them from the torsional preferences. To this end, two changes to the RDKit were introduced: (i) cyclic bonds that are part of more than one SSSR ring are excluded, and (ii) if a ring shares more than two bonds with another ring, all of its bonds are excluded.

Fitting of Torsional-Angle Potentials. Distributions of the experimental torsional angles were extracted from the CSD (status 2014, filtered for organic molecules) for each of the SMARTS patterns. The multiplicity ($m = 1–6$) and force constant were assigned manually based on the experience from the torsional-angle potentials of the acyclic bonds in ETKDG, versions 1 and 2,²² and refined by comparing the distributions obtained for the molecules in the test set with those from the CSD. Due to the ring constraints in small rings, it was found that also small peaks had to be considered explicitly, i.e., often a high multiplicity had to be chosen.

Datasets. Two test sets with small ring-containing molecules were compiled. The first set (sr-CSD-set) contains all molecules from the CSD test set in ref 22 that possess at least one aliphatic ring of a maximum size of eight and had a molecular weight below 600 g/mol. The definition of “aliphatic” in this context is very broad, i.e., at least one bond in the ring is not aromatic. Not all patterns were represented in this subset. Thus, additional molecules were selected from the CSD with the same criteria, which contained bonds that matched patterns have not yet represented. This resulted in a total of 600 molecules. Eight of the 105 patterns are still not represented in the sr-CSD-set.

The second set (sr-PDB-set) contains all molecules from the Platinum diverse set³⁰ that possess an aliphatic ring of a maximum size of eight and had a molecular weight below 600 g/mol. This resulted in a total of 1401 molecules.

The list of SMILES and database identifiers is provided in the Supporting Information.

Conformer Generation. The base conformer generator was ETKDG version 2. The following parameters were used for the conformer generator: maximum number of conformers = 100, fixed random number seed = 210,185, maximum iterations = 200 (the maximum number of failed embedding attempts allowed per conformer), and RMS threshold = 0.1 Å (a post-processing step where conformers that are too similar to already existing ones are removed). Conformers were generated either with the default embedding approach of RDKit (termed *embedding* in the following) or with random coordinates. The default embedding approach in distance geometry determines the three largest eigenvalues of the distance matrix and projects the high-dimensional metric matrix along the corresponding three eigenvectors.³¹ A consequence of this procedure is that the variance of the resulting 3D coordinates in each dimension is

maximized. The coordinates are further minimized in a distance field to ensure that all distance bounds are fulfilled. An alternative to embedding is to start from random (4D) coordinates generated within a given volume followed by minimization in the distance field. For this, the “useRandomCoords” flag of the conformer generator is set to true.

The conformer generator with the additional small-ring torsion patterns is called ETKDG+sr.

Macrocycles. The workflow of the conformer generator was modified in four different ways to improve sampling of macrocyclic conformations.

Changes to the Torsional-Angle Preferences. In the current ETKDG (version 2), experimental torsional-angle preferences are only applied for acyclic bonds. However, bonds in cycles with a size greater than eight often behave similarly to linear bonds (see above). Therefore, the existing torsional SMARTS patterns were modified such that they also apply to bonds in rings of size nine or larger. In addition, three SMARTS patterns for amides in macrocycles were added. This resulted in total 367 SMARTS patterns (without the patterns for the small aliphatic rings).

Changes to the Bound's Matrix Generation. In the generation of the bound's matrix in the RDKit, different subroutines for setting the 1,4-bounds exist depending on the number of ring bonds involved. As the torsional preferences of macrocycles are more similar to linear chains than to small rings, the code was modified such that 1,4-bounds involving macrocyclic ring bonds are treated as non-ring bonds.

Eccentricity. In 2D geometry, eccentricity e is a non-negative real number that uniquely characterizes the shape of a conic section. For ellipses, e takes values between zero and one

$$e = \sqrt{1 - \frac{b^2}{a^2}} \quad (1)$$

where a and b are the lengths of the semi-major and semi-minor axes, respectively. The ellipse perimeter p can be calculated via the Ramanujan approximation

$$p \approx \pi[3(a + b) - \sqrt{(3a + b)(a + 3b)}] \quad (2)$$

When a and b are known, a point (x, y) is on an ellipse centered at $(0,0)$ if it fulfills

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (3)$$

The circle is an ellipse with $e = 0$. Thus, eccentricity can be interpreted as a measure of how round a set of points is. Figure 2 illustrates the conformations of the cyclic peptide cyclosporine A (CsA) with different eccentricity values (roundness). For this, we assume that the coordinates of the ring atoms in a macrocycle

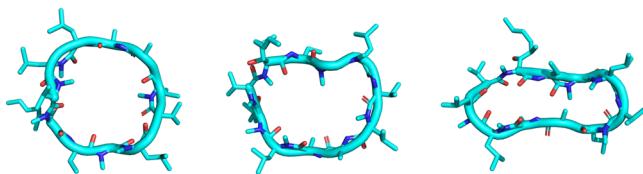


Figure 2. Conformers of the cyclic peptide cyclosporine A (CsA) with different measured eccentricity values (roundness): $e = 0.08$ (left), 0.65 (middle), and 0.92 (right). The ribbon representation provides a good approximation of the ellipse fitting that was done to obtain the 2D eccentricity values. The structures were generated *in silico*.

can be projected on a 2D plane without too much loss of information.

Furthermore, eccentricity can be used to enforce a certain “roundness” of a set of points. This takes the following steps:

1. Ring atoms of the macrocycle are identified.
2. Given an input eccentricity value e_{input} and an estimated perimeter p (by summing the lengths of the bonds connecting the ring atoms), eqs 1 and 2 form a non-linear system of equations that can be solved numerically to obtain a and b .
3. Points equal to the number of atoms identified in step 1 are placed on the ellipse perimeter. By default, the point representing the first atom location is placed at a vertex of the ellipse with the subsequent points placed in an anti-clockwise fashion around the ellipse. Alternatively, one can also specify the starting point location on the ellipse.
4. The lower and upper bounds of the pairwise distances between all ring atoms in the macrocycle are updated in the bounds matrix according to the atom locations on the ellipse.
5. The modified bounds matrix is given as input to ETKDG.

Custom Pairwise Coulombic Interactions (CPCI). In stage 8 of the ETKDG workflow (Figure 1), a minimization is performed to refine the generated conformer based on the distance bounds and the ETK terms. Note that this is not a standard force-field minimization, i.e., there are no vdW or electrostatic interactions. This means that intramolecular hydrogen bonds are not formed. In order to give the user full flexibility to favor certain interactions, we have implemented the option to add pairwise Coulombic interaction terms to the minimization step ($1/r$ decay with r being the distance between a pair of atoms). Such interaction pairs (repulsive or attractive) can be selected by the user based on chemical intuition and provided as an argument when calling the conformer generator.

Conformer Generation. The base conformer generator was ETKDG version 2. The following number of conformers was generated depending on the size of the macrocycle: 540 conformers for macrocycles with less than 15 ring atoms, 1800 conformers for macrocycles with 15–30 ring atoms, and 5400 conformers for macrocycles with more than 30 atoms. The number of generated conformers was increased with increasing ring size due to the growth of conformational space that needs to be sampled. The total number of conformers generated is always divisible by 18 to be consistent with the sampling procedure used with eccentricity constraints (see next paragraph). No filtering using an RMS threshold was applied. The default maximum iteration setting was used, which is set to 10 times the total number of atoms in the molecule. The default random number seed was used (i.e., the current time). The effect of the stochastic sampling of the distance matrices is shown in the Results and Discussion. Conformers were generated either with embedding or with random coordinates.

In this study, we used eccentricity to generate conformers as follows: the input eccentricity is fixed at 0.99. The constraints are applied in increments of 10° (from 0° to 170° due to an ellipse's twofold symmetry), each time, placing the first ring atom on the ellipse perimeter (step 3 in the workflow) at the given angle to the horizontal. Thus, 18 modified bound matrices are generated per molecule whereby the macrocycle ring is squashed at different orientations. To keep the total number of generated conformers the same as before, we only sample N/18

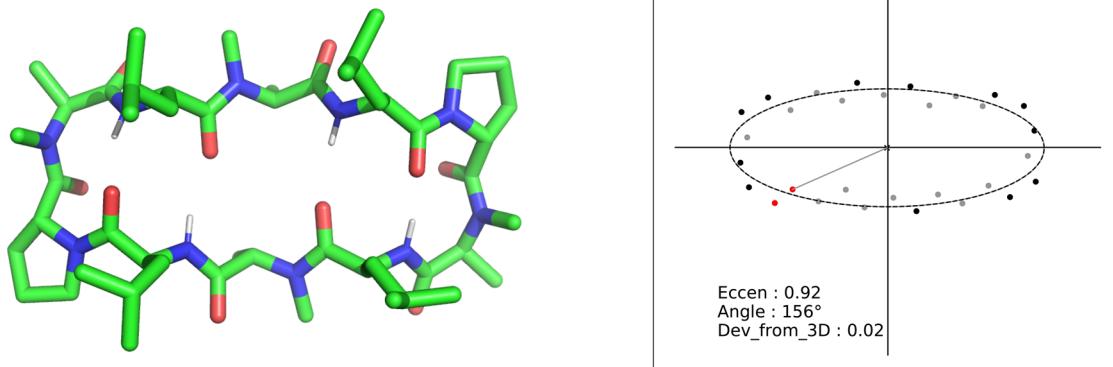


Figure 3. (Left) NMR structure of a cyclic decapeptide.^{51,52} (Right) 2D mapping of its ring atoms after PCA, including the fitted ellipse. Black markers indicate ring atoms which are also in a 2D convex hull. The remaining atoms are colored gray, except for the first two ring atoms, which are colored red. The first ring atom is also connected to the center of the fitted ellipse by a solid line. The two red dots define the direction, anticlockwise/clockwise, of the ellipse. This way the unique angle, which the directed ellipse makes with the horizontal axis, can be measured.

conformers per modified bounds matrix where N takes a value of 540, 1800, and 5400 depending on the macrocycle ring size.

Each of the 18 modified ring bounds matrices is used to amend the original full bounds matrix generated by RDKit via this update scheme: if a given upper bound entry in the modified matrix is smaller than its corresponding part in the original matrix, the new entry is adapted. At the same time, if the lower bound entry in the modified matrix is also smaller than its corresponding part in the original matrix, the new entry is adapted. All other entries in the original bounds matrix are retained. Thereafter, the bounds matrix is supplied to step 1 of Figure 1. Thus, we have monotonically reduced the range of the bounds matrix, yielding a smaller conformational search space.

The effect of CPCIs was tested on peptidic macrocycles where we introduced CPCIs between all unmethylated amine hydrogen and carbonyl oxygen pairs (HOPs) separated by more than five bonds (i.e., if H and O are not on the same or neighboring residue). We used the MMFF reference charges for H and O, applying various scaling factors to their charge product: factors of 0.1, 0.2, and 0.5 were tested.

We also experimented with including charge products for HOPs separated by more than five bonds and also adding the opposite charge products to HOPs on the same or neighboring residue (counter charges). This is an attempt to bias the amides to stay in the *trans* configuration. Again, different scaling factors were tested.

We performed tests on using the modified ETKDG combined with the various flavors of CPCIs with or without random coordinates and eccentricity constraints. A number of 5400 structures were generated as before. As a post-processing step, we subjected the top 500 rRMSD structures to energy minimization in vacuum using OpenMM⁴⁵ and OpenForceField⁴⁶ whereby the partial charges were assigned using our previously published work on machine-learning charges⁴⁷ (github.com/rinikerlab/mldec). An additional SMARTS pattern for methylated amides was added to the OpenForceField torsion parameter file with the same torsional parameters as for unmethylated amides.

Datasets. For the macrocycles, the same set of experimental structures of chemically diverse macrocycles used by OpenEye to parameterize the new OMEGA conformer generator for macrocycles^{6,7} was taken (mc-OE-set). The set originates from

four different sources: flexible macrocycles from CSD 2016,³⁸ the Prime dataset of Sindhikara et al.,¹¹ and 2016 BIRD⁴⁸ and MAC10 from PDB's LigandExpo 2016.⁴⁹ These structures were subsequently filtered to keep only those with a single large ring. This results in a total of 616 experimental structures for 480 unique single-macrocycle molecules (BIRD: 40, CSD: 262, MAC10: 53, and Prime: 261).

In addition, 19 macrocyclic structures were collected from the most recent D3R Grand Challenge 4 (druggesigndata.org/about/grand-challenge-4) (mc-D3R-set). These macrocycles were co-crystallized with their target protein β -secretase 1 (BACE) and possess a similar core (a macrocycle with either 15 or 16 atoms) with one flexible substituent.

A third set (mc-PEP-set) consists of the crystal structure of the cyclic undecapeptide cyclosporine A⁵⁰ and the NMR structure of a cyclic decapetide.^{51,52} Both adopt a “closed” conformation with a maximum number of intramolecular hydrogen bonds. Such conformations are hypothesized to be crucial for good passive membrane permeability,^{41–43} and thus, it is important to be able to generate them *in silico*.

The list of SMILES and database identifiers is provided in the Supporting Information.

Availability. A Docker container with the modified RDKit source code and examples is available at https://github.com/rinikerlab/RDKit_mETKDG. The modifications will be incorporated into the 2020.03 release of the official RDKit.

With the definition of a custom bounds matrix with eccentricity constraints, the specification of CPCIs as well as force field-based minimization is handled by a Python package we are developing termed cpeptools (<https://github.com/rinikerlab/cpeptools/>). Sample snippets on using this package in conjunction with the modified RDKit is also available in the Docker container.

Analysis. Root-Mean-Square Deviation (RMSD). The atom positional RMSD is used to measure the difference between generated conformers and experimentally derived structures

$$\text{RMSD} = \sqrt{\frac{1}{N_{\text{atoms}}} \sum_{i=1}^{N_{\text{atoms}}} (r_i - r_{i,\text{ref}})^2} \quad (4)$$

where N_{atoms} is the number of heavy atoms in the molecule, r_i is the position of atom i , and $r_{i,\text{ref}}$ is the position of atom i in the

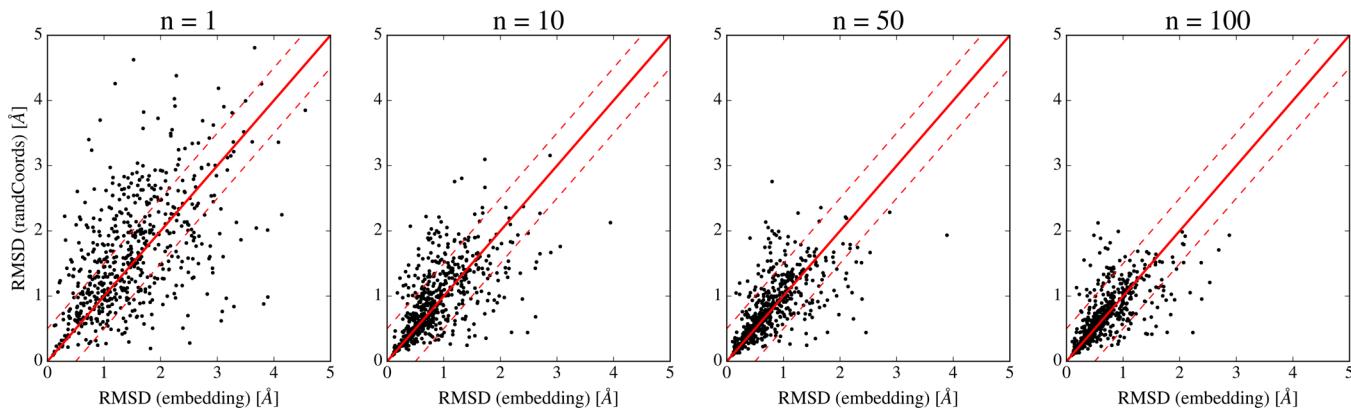


Figure 4. Best RMSD with respect to the crystal structure for ETKDG version 2 with embedding versus the same generator with random coordinates for 1, 10, 50, and 100 conformers. The sr-CSD-set (600 molecules) was used.

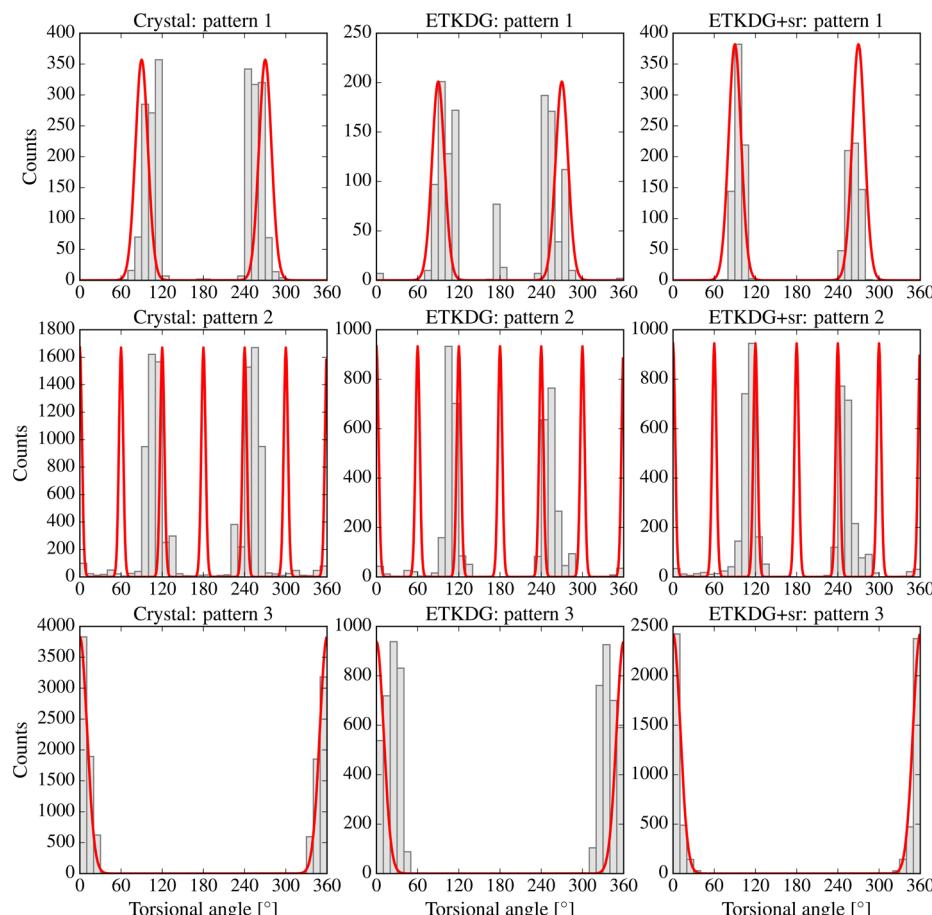


Figure 5. Torsional-angle distributions from the CSD, ETKDG (only acyclic torsion patterns), and ETKDG+sr (with small-ring torsion patterns) together with the fitted torsional-angle potentials for the first three of 105 SMARTS patterns (the list of patterns is provided in the Supporting Information). For the conformation generation, the sr-CSD-set was used and numConfs was set to 100.

experimental reference structure. The RMSD was obtained using the RDKit function AllChem.GetBestRMS() (symmetry is taken into account in the optimal rigid-body superposition).

In addition, the RMSD of the ring atoms plus the direct neighboring heavy atoms (β atoms) was calculated as

$$\text{rRMSD} = \sqrt{\frac{1}{N_{\text{ring}+\beta \text{ atoms}}} \sum_{i=1}^{N_{\text{ring}+\beta \text{ atoms}}} (r_i - r_{i,\text{ref}})^2} \quad (5)$$

The RDKit was used for determining the ring atoms and their neighbors. For small molecules, rRMSD is calculated using the RDKit AllChem.GetBestRMS() function (using the argument map). As some molecules in the test sets contain more than one ring, the rRMSD was calculated separately for each SSSR ring. For the macrocycles, the alignment and rRMSD calculation were done with MDTraj.⁵³

Torsion Deviation. An alternative to RMSD to assess the difference between two conformations is the deviation in torsional angles. For this, we employ the torsion fingerprint

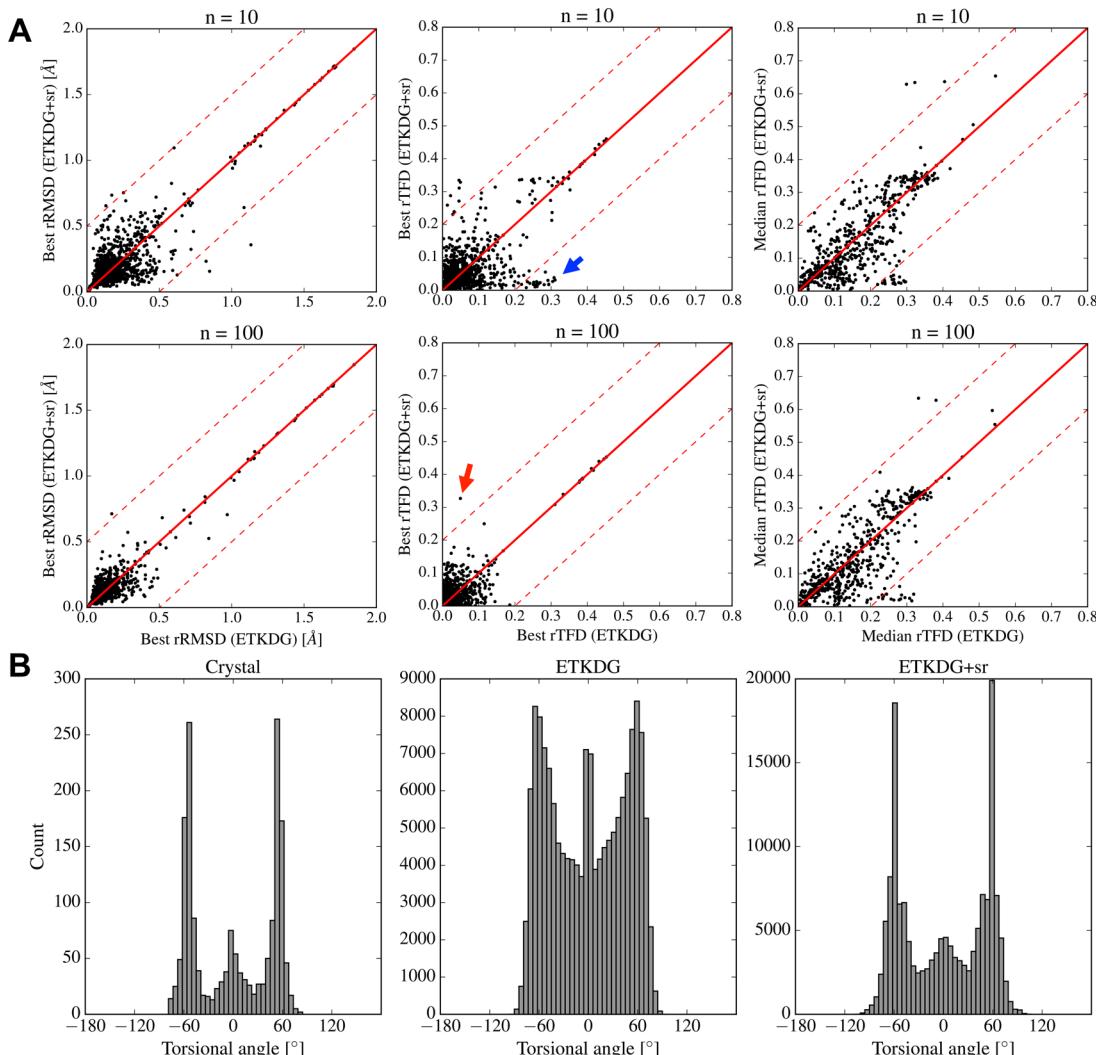


Figure 6. Performance on the sr-CSD-set. (A) Comparison of best ring RMSD (rRMSD), best ring TFD (rTFD), and median rTFD values between ETKDG and ETKDG+sr for 10 and 100 conformers. In each panel, the red solid line represents $y = x$, and the red dashed lines indicate a difference of $\Delta rRMSD = 0.5$ or $\Delta rTFD = 0.2$. (B) Distribution of the torsional angles in six-membered rings obtained from the crystal structures of the 600 molecules in the sr-CSD-set as well as from conformation generation with ETKDG and ETKDG+sr (numConfs = 100).

deviation (TFD) approach.⁵⁴ In the original implementation, the torsions of rings are averaged into a single value, which is not optimal for the analysis in this study. Thus, we have adjusted the approach for ring torsions, i.e., each ring torsion is considered separately and only ring bonds are considered. The maximum possible deviation Δt_{\max} is set to 180° , the weight is set to 1 for all ring torsions. The ring TFD (rTFD) is calculated as

$$rTFD = \frac{1}{N_{\text{tors}}} \sum_{i=1}^{N_{\text{tors}}} \frac{|\Delta t_i|}{\Delta t_{\max}} \quad (6)$$

where N_{tors} is the number of ring torsions and $|\Delta t_i|$ is the absolute deviation of torsion i . The rTFD can take values between 0 (no deviations) and 1 (maximum deviations).

Statistical Analysis. Two conformer generator variants were compared with a two-sided paired t-test using the `ttest_rel` function of the `scipy.stats` Python package.

Eccentricity. Eccentricity can also be used to evaluate the roundness of a generated conformer. Figure 3 illustrates a conformation of a cyclic peptide and its corresponding 2D fit to an ellipse. To do so requires the following steps:

- (1) Ring atoms of the macrocycle are identified.
- (2) The coordinates of the ring atoms are subjected to a principle component analysis (PCA) to obtain the coordinates projected on a 2D plane.
- (3) The optimal ellipse fitting to the points is obtained by minimizing the square error.⁵⁵
- (4) Eccentricity is calculated from the ellipse using eqs 1 and 3.

Availability. The aforementioned Docker container at https://github.com/rinikerlab/RDKit_mEtkdg also contains all the code snippets needed to replicate the analysis.

RESULTS AND DISCUSSION

Small Aliphatic Rings. Random Coordinates. For the molecules with small aliphatic rings (i.e., rings with at least one aliphatic bond), the use of random coordinates gave similar results as with embedding (confirmed with a two-sided paired t-test) (Figure 4). However, as the random coordinates reduced strong outliers and in general tend to perform better for fused ring systems, we decided to use random coordinates in the following.

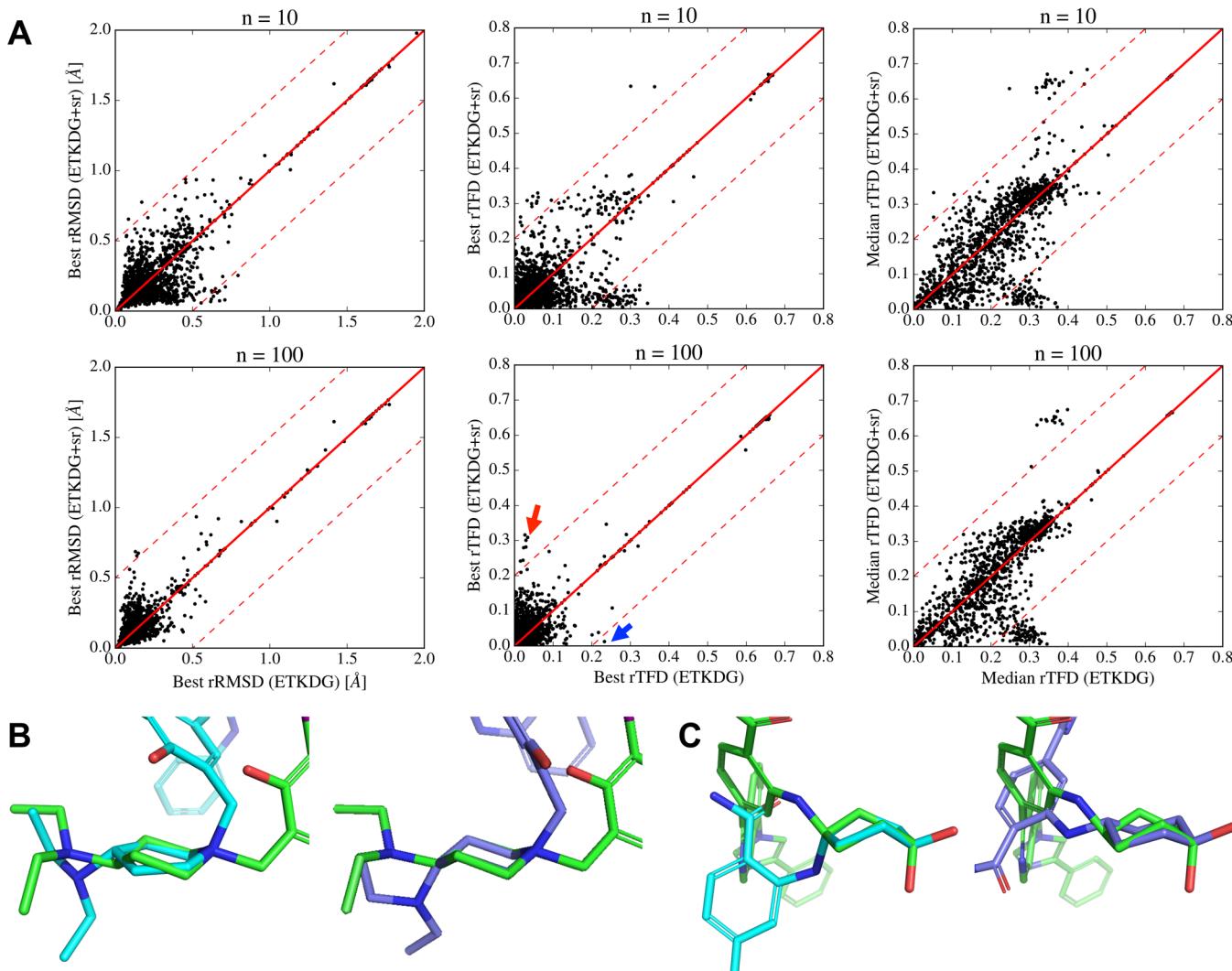


Figure 7. Performance on the sr-PDB-set. (A) Comparison of best ring RMSD (rRMSD), best ring TFD (rTFD), and median rTFD values between ETKDG and ETKDG+sr for 10 and 100 conformers. In each panel, the red solid line represents $y = x$, and the red dashed lines indicate a difference of $\Delta rRMSD = 0.5$ or $\Delta rTFD = 0.2$. (B) Example where ETKDG+sr (purple) reproduced the crystal conformation of the central ring (green) better than ETKDG (cyan) in terms of best rTFD (PDB: 4AGQ, blue arrow in (A)). (C) Example where ETKDG+sr (purple) reproduced the crystal conformation of the central ring (green) worse than ETKDG (cyan) in terms of best rTFD (PDB: 3WQ9, red arrow in (A)).

Fitting of Torsional-Angle Potentials. A total of 105 SMARTS patterns for aliphatic rings with a size of ≤ 8 were developed. For each pattern, the distribution obtained from the CSD was manually fitted with torsional-angle potentials. The patterns and corresponding parameters can be found in the RDKit. As examples, the distributions from the CSD, ETKDG (version 2), and ETKDG+sr together with the fitted torsional-angle potentials are shown in Figure 5 for the first three patterns. The sample plots for all 105 patterns are given in Figure S3 in the Supporting Information. Overall, the application of the torsional-angle potentials resulted in distributions similar to those observed in the crystal structures in the CSD. Further, it was found that, in general, higher multiplicities had to be chosen for the ring torsional-angle potentials compared to the acyclic ones due to potential ring strains. For the same reason, bonds in bridged ring systems and bonds that belong to two rings are excluded from the torsional preferences.

Reproducing Crystal Conformations. The ability of the conformer generator to reproduce crystal conformations was assessed by calculating the best rRMSD (ring atoms + β atoms)

and the best rTFD with respect to the crystal structure out of a maximum of 100 generated conformers. Note that the conformer with the best rRMSD value and the one with the best rTFD value may not be the same. The results for the sr-CSD-set are shown in Figure 6A (the same plots for the overall RMSD are given in Figure S4; again, the conformer with the best rRMSD value may not be the same as the one with the best overall RMSD value).

Overall, the best rRMSD values are similar with and without the ring torsional-angle preferences. This is surprising at first given that the individual torsional-angle distributions are better reproduced by ETKDG+sr compared to ETKDG. Similarly, when looking at the conformation of six-membered rings, a clear preference for the chair conformation (i.e., torsional angle = 60°) can be observed in the crystal structures (Figure 6B). Again, this experimental preference is much better reproduced by ETKDG+sr. When looking at the best rTFD values, it is noticeable that ETKDG+sr improves the performance for $n = 10$ conformers, but the effect vanishes for $n = 100$. Furthermore, the median rTFD over n conformers is significantly reduced with

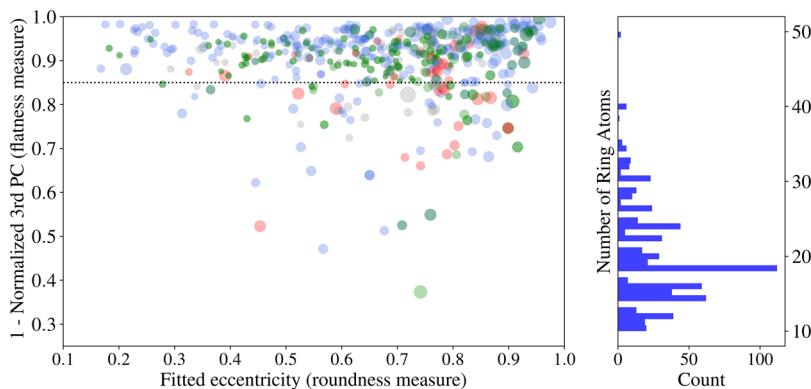


Figure 8. Left: Flatness ($1 - \text{normalized third largest principle component (PC) eigenvalue}$) vs eccentricity e (“roundness” measure) for all experimental crystal structures of the mc-OE-set. The size of markers indicates the macrocyclic ring size, whereas the color represents the source of the macrocycles: blue = CSD, green = Prime, red = BIRD, and gray = Mac10. The dotted horizontal line denotes the flatness cutoff of 0.85. Right: Distribution of the macrocycle ring size in the mc-OE-set.

ETKDG+sr for both $n = 10$ and 100 . Taken together, these observations indicate that the small-ring torsional preferences increase the probability of generating a good conformation. This effect is clearly visible in the median rTFD. In terms of best rTFD, it is most notable when generating a small number of conformers (e.g., chair conformation of six-membered rings; see blue arrow in Figure 6A); however, the effect vanishes for a larger n because a good conformation is eventually generated with ETKDG.

The negative outlier in terms of rTFD ($n = 100$) in Figure 6A (red arrow) is the crystal structure CDBMPI10 where the central six-membered ring is in a boat conformation. However, the small-ring torsions in ETKDG+sr favor chair conformations.

For the biologically active conformations in the Platinum set, similar findings were obtained (Figure 7A). The median rTFD is again clearly reduced, but the positive effect by the small-ring torsional preferences on the best rTFD values visible for $n = 10$ vanishes for a larger number of conformers ($n = 100$). Investigation of molecules with large differences between ETKDG and ETKDG+sr (examples in Figure 7B + C) shows that negative outliers in terms of best rTFD values stem again from six-membered rings, which are in the boat conformation in the crystal structure. Positive outliers, on the other hand, are molecules with six-membered rings in a chair conformation.

In conclusion, the introduction of torsional-angle preferences for small aliphatic rings results in a better reproduction of the distributions observed in crystal structures. The preference for the chair conformation for six-membered rings is also better described. This increases the likelihood to generate a good ring conformation close to the crystal conformation (as shown by the decreased median rTFD). If only a small number of conformers is generated (e.g., $n = 10$), the improvement is also visible in terms of best rTFD. However, if sufficiently many conformers are generated (e.g., $n = 100$), a good conformation is eventually obtained also without the small-ring torsion preferences. Thus, similar best rTFD or best rRMSD values are obtained with ETKDG and ETKDG+sr.

We anticipate therefore that the addition of the small-ring torsional-angle preferences will be most valuable for applications where only a small number of conformers are generated per molecule. In the RDKit, there will be an option to include the small-ring torsional patterns.

Macrocycles. Eccentricity of the mc-OE-set Reference Structures. The eccentricity e of a macrocycle (based on the

projection of the coordinates onto a 2D plane) is an indication of how round the ring is, where eccentricity = 1 represents a “squashed” (ellipsoid) macrocycle and eccentricity = 0 is obtained for a perfect ring. In the left panel of Figure 8, the eccentricity values of the experimental structures of the macrocycles are plotted against 1 minus the normalized magnitude of the third largest eigenvalue after performing principal component analysis (PCA) of each structure. The distribution of ring sizes in the macrocycles dataset is shown below on the right.

The latter is a measure of how flat a structure is and is also bound between 0 and 1, with 1 indicating a perfectly flat structure. A cutoff of 0.85 was used to classify flat and non-flat structures. Due to the 3D-to-2D projection for the calculation of the eccentricity, it is most meaningful for flat structures. As can be seen in Figure 8 (left), most of the macrocyclic crystal structures are flat and squashed, likely due to crystal packing effects.⁵⁶

Use of Random Coordinates. As for the compounds with small aliphatic rings, we tested the use of random coordinates for the macrocycles and compared the performance to that with embedding. The effect is likely higher for macrocycles than for compounds with small rings because the embedding step maximizes the variance in the three dimensions, which results in a bias toward round structures. This is in contrast to the trend observed in the crystal structures (see above), which tend to be rather squashed. Random coordinates are generated within a given volume as initial guesses. This removes the bias toward round structures but at the same time leads to a higher likelihood for non-flat structures, sometimes even folded-up.

As can be seen in the top panel in Figure 9, the generated conformers are generally less flat when using random coordinates regardless of the macrocycle source or ring size. It is also independent whether the crystal structure of a compound is flat or not (marked with yellow crosses). In general, agreement with the experimental structures in terms of the best rRMSD improved using random coordinates (bottom panel in Figure 9). This means that the reduced flatness of the conformers generated with random coordinates has little effect on the rRMSD, whereas the reduced roundness is clearly beneficial.

Torsion Patterns and 1,4-Bounds Modification. The modified ETKDG version with the torsional-angle preferences for the macrocyclic bonds (same as for acyclic bonds) and the adjusted 1,4-distances in the bounds matrix (termed mETKDG)

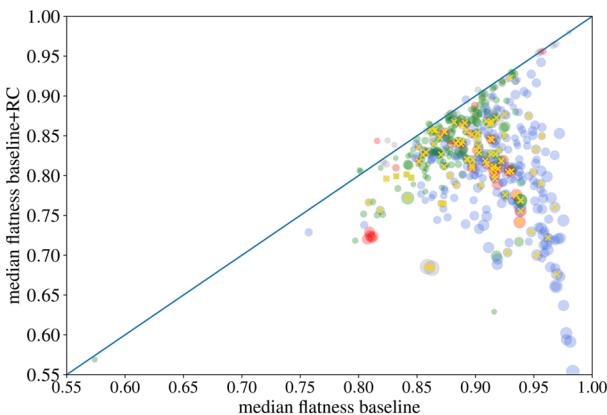


Figure 9. Comparison of ETKDG version 2 with embedding (baseline) vs the same generator with random coordinates (RC) for the mc-OE-set in terms of median flatness (top) and the best rRMSD with respect to the crystal structures (bottom). The size of markers indicates the macrocyclic ring size, whereas the color represents the source of the macrocycles: blue = CSD, green = Prime, red = BIRD, and gray = Mac10. The blue solid lines indicate $y = x$, the gray dotted lines represent $\pm 0.20 \text{ \AA}$ deviation, and the golden crosses on a marker denote compounds for which the experimental structure has a flatness value of <0.85 . The overall median rRMSD improvement of the baseline+RC over the baseline on this dataset is 0.0 \AA .

is compared against the baseline ETKDG with random coordinates. Systematic improvements in terms of the best rRMSD was observed with mETKDG, especially for flat cycles (Figure 10). Thus, we recommend using this combination as the *de facto* strategy for sampling macrocycles with the RDKit (see Figure S5 in the Supporting Information for the effect of only mETKDG without random coordinates).

Eccentricity Constraints. As the crystal conformations of many macrocycles are relatively squashed (Figure 8, left), we tested the combination of mETKDG with eccentricity constraints (mETKDG+Eccen) to bias the sampling toward such structures. Thereby, mETKDG+Eccen was used with embedding and compared to mETKDG using random coordinates. As can be seen in Figure 11, the introduction of eccentricity constraints with mETKDG results in not only a higher likelihood for squashed conformations but also increased flatness. For macrocycles whose crystal conformation is not flat, the rRMSD thus worsened (yellow crosses in Figure 11). However, for the macrocycles with flat crystal conformations (squares in Figure 11), the performance of the conformer generator improved when eccentricity constraints were applied.

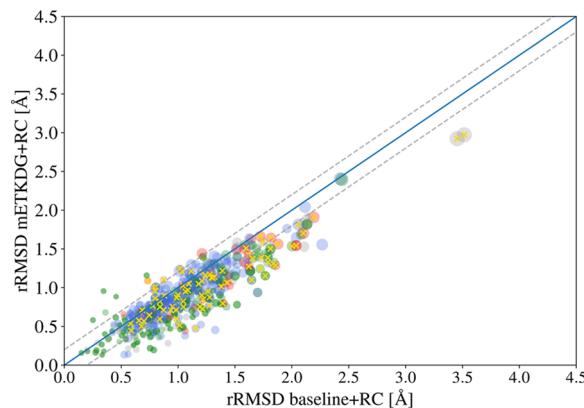


Figure 10. Comparison of ETKDG (baseline) and mETKDG for the mc-OE-set using random coordinates (RC) in terms of the best rRMSD with respect to the crystal structures. The size of markers indicates the macrocyclic ring size, whereas the color represents the source of the macrocycles: blue = CSD, green = Prime, red = BIRD, and gray = Mac10. The blue solid lines indicate $y = x$, the gray dotted lines represent $\pm 0.20 \text{ \AA}$ deviation, and the golden crosses on a marker denote compounds for which the experimental structure has a flatness value of <0.85 . The overall median rRMSD improvement of mETKDG+RC over baseline+RC on this dataset is 0.15 \AA .

Note that, although an eccentricity value of 0.99 was imposed, more round conformers are also sampled, i.e., the distribution in the generated conformational ensemble is shifted toward higher eccentricity values. This indicates that the eccentricity constraints restrict sampling perpendicular to the macrocyclic ring plane while still allowing relaxation within the ring plane.

Interestingly, we found that the use of random coordinates and the eccentricity constraints is not compatible (Figure S6 in the Supporting Information). The reason for this is that the use of random coordinates effectively impedes the modifications in the bounds matrix introduced by the eccentricity constraints. These are mainly tighter upper bounds, which are not violated by the folded-up (i.e., less flat) conformations generated using random coordinates.

Figure 12 shows conformers with the best rRMSD overlaid on the experimental structure for two macrocycles as examples: (1) mETKDG with random coordinates outperforms mETKDG+Eccen with embedding, and (2) vice versa. For clarity, only the macrocyclic ring and their direct neighbors (β atoms) are included in the depiction. For macrocycle no. 269, the crystal conformation of the ring resembles the Chinese character “匚” (with the left, right, and bottom sides parallel to the page while the top side is twisted and sticks out of the page). Due to such deviation from the (2D) elliptical approximation and strong 3D extent, imposing high eccentricity is detrimental for the reproduction of the crystal structure. On the contrary, for macrocycle no. 249, the crystal conformation of the ring is flat and shaped like a parallelogram. The introduction of eccentricity constraints biases the conformational sampling toward such structures, and a conformer with the best rRMSD = 1.56 \AA was found. The conformers generated with mETKDG using random coordinates, on the other hand, are, in general, not flat enough.

The advantages of imposing eccentricity constraints are speed (embedding is generally faster in the RDKit than the use of random coordinates) and, more importantly, the ability for the user to bias sampling toward a particular region of conformational space through the orientation angle and target eccentricity value. If, for example, the locations of the turns in a macrocycle

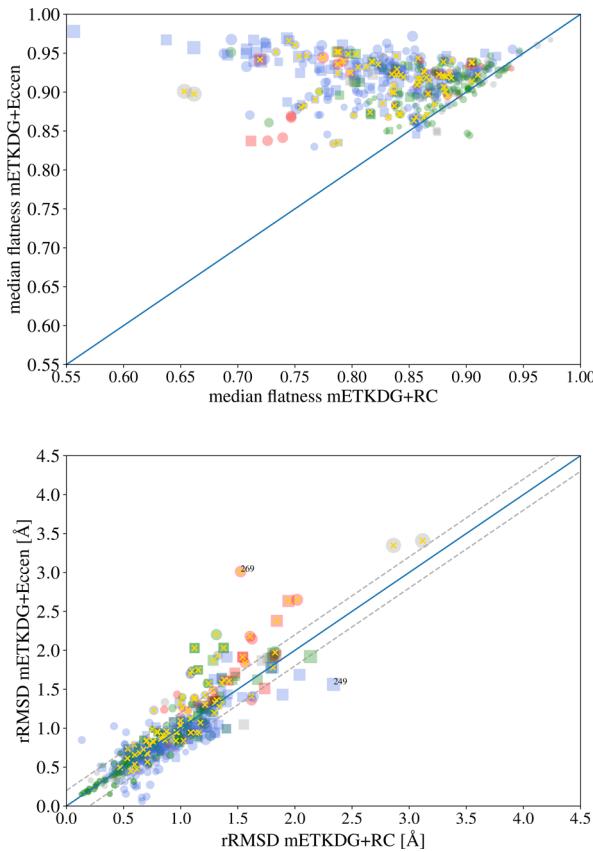


Figure 11. Comparison of mETKDG using random coordinates (RC) and mETKDG plus eccentricity (Eccen) constraints using embedding for the mc-OE-set in terms of median flatness (top) and best rRMSD with respect to the crystal structures (bottom). The size of markers indicates the macrocyclic ring size, whereas the color represents the source of the macrocycles: blue = CSD, green = Prime, red = BIRD, and gray = Mac10. The blue solid lines indicate $y = x$, the gray dotted lines represent $\pm 0.20 \text{ \AA}$ deviation, and the golden crosses on a marker denote compounds for which the experimental structure has a flatness value of < 0.85 . Squares represent macrocycles with eccentricity $e \geq 0.8$; circles represent those with $e < 0.8$. The overall median rRMSD improvement of mETKDG+Eccen over mETKDG+RC on this dataset is 0.01 \AA .

are known (or can be guessed with high certainty), only conformers with these turn locations can be generated. This is not possible using random coordinates. If the location of the turns is not known beforehand, the orientation angle can be scanned, and for each angle, a set of conformers can be generated. The latter procedure was used for the molecules in the mc-OE-set.

The rTFD metric was also applied to compare methods on the mc-OE-set. The same trend of improvement with mETKDG was observed for the best rTFD (top panel of Figure S7 in the Supporting Information) as for the best rRMSD. However, we find rTFD generally to be a less useful metric for comparing conformers of macrocycles, in contrast to small rings. This is due to the averaging over a large number of torsional angles in the calculation of the rTFD metric, which dilutes the contributions of individual deviations. This means that a conformer can already deviate greatly from a reference structure in terms of rRMSD when only a few torsional angles are very different. This effect can be seen for the example for macrocycles no. 269 and no. 249 where the effect of the eccentricity constraints is not visible using the rTFD metric (bottom of Figure S7).

Structurally, the best rTFD conformers deviate more from the reference structures than the best rRMSD conformers (shown in Figure 12).

The mc-D3R-set Macrocycles. We further tested mETKDG and mETKDG+Eccen to reproduce the co-crystal conformations of a series of 19 macrocyclic ligands in the BACE protein binding site (mc-D3R-set). These ligands were part of the pose prediction phase of the most recent D3R grand challenge 4 (drugdesigndata.org/about/grand-challenge-4). A better sampling of relevant conformers can lead to improved docking results.

The crystal conformations of these ligands are mostly flat but span a large range in terms of roundness (Figure S8 in the Supporting Information). Thus, imposing eccentricity constraints is expected to improve the sampling for these ligands. As can be seen in Figure 13, the best rRMSD obtained with mETKDG is generally lower than that with ETKDG. Biasing the sampling toward conformers with higher eccentricity further improves the results. The effect from stochastic sampling (i.e., generating conformers with the same method twice using different random number seeds) is approximately 0.15 \AA (Figure S9 in the Supporting Information). This can also be seen when comparing the full distribution of rRMSD values in the generated ensembles (Figure S10 in the Supporting Information). The rRMSD distribution obtained with mETKDG+Eccen (with embedding) is either on par or shifted to better values compared to that obtained with mETKDG using random coordinates.

Cyclic Peptides. As the ring size increases, it becomes more challenging to adequately sample the conformational space of macrocycles, and thus, different ways to restrict the search space are desired. Many macrocycles form intramolecular hydrogen bonds (H-bonds) in the crystal structure; therefore, specific conformations that allow for a maximum number of H-bonds are favored. This is especially true for cyclic peptides. Our test set mc-PEP-set contains two large cyclic peptides (Figure 14): (1) the natural product cyclosporine A (CsA),⁵⁰ an undecamer with three *trans*-annular H-bonds in the crystal conformation (a fourth H-bond is formed outside of the ring), and (2) a synthetic decapeptide^{43,51,52} with four *trans*-annular H-bonds observed in the NMR solution structure in chloroform (the same conformation was seen in the crystal structure of a closely related peptide). The structure of the decapeptide contains all *trans*-amide bonds, and the two proline residues at the turns restrict the available conformational space. The crystal structure of CsA contains a single *cis*-amide bond between residues 9 and 10 (Figure 14).

Use of Custom Pairwise Coulombic Interactions. The experimental structures of decapeptide and CsA have flatness values of 0.94 and 0.92, and both have an eccentricity (roundness) value of 0.92. Using mETKDG with random coordinates or mETKDG+Eccen with embedding performed similarly for these cyclic peptides with relatively high best rRMSD values of 1.68 \AA for the decapeptide and 1.88 \AA for CsA. The squashed and rather flat conformation of the macrocycle in the experimental structures is a consequence of the *trans*-annular H-bonds. To mimic this, we introduced custom pairwise Coulombic interactions (CPCIs) between amide atoms as described in the Methods section. Three scaling factors for the strength of the CPCIs were tested: 0.1, 0.2, and 0.5. Larger values are not recommended as the forces become too strong and atoms may overlap in the final conformation. The application of CPCIs brings atom pairs that can form

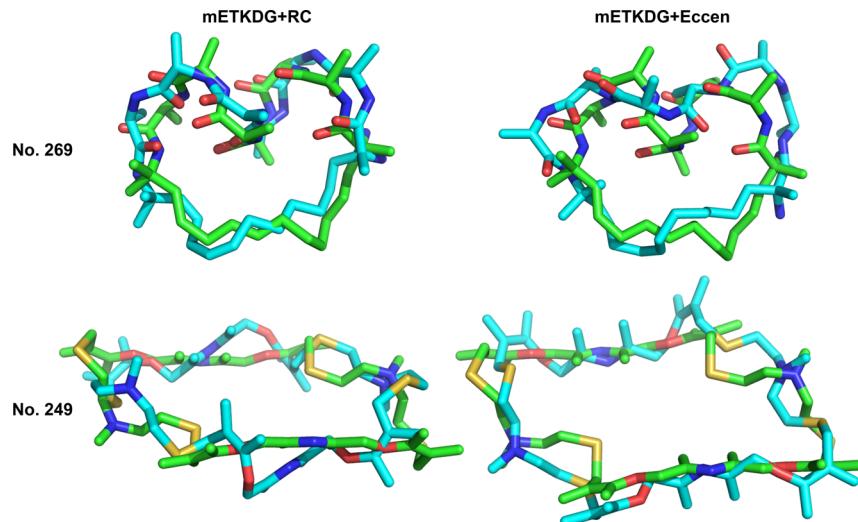


Figure 12. Overlay of the crystal structure (green) and best rRMSD conformer (cyan) for macrocycle nos. 269 (top) and 249 (bottom) generated by mETKDG using random coordinates (left) or by mETKDG+Eccen with $e = 0.99$ and embedding (right). Only the macrocyclic ring atom and their direct neighbors (β atoms) are shown (and used for alignment).

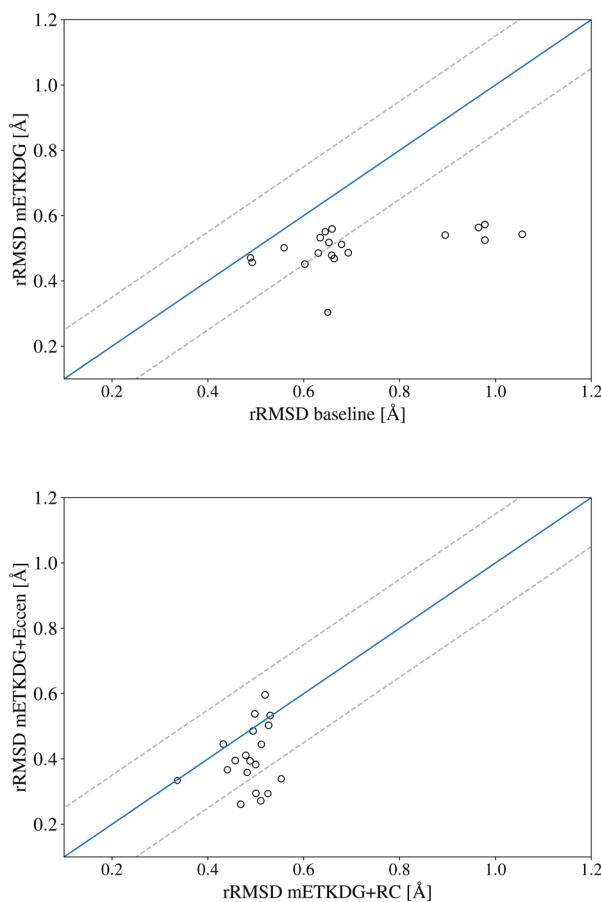


Figure 13. Comparison of conformation generators for the mc-D3R-set in terms of the best rRMSD with respect to the crystal structures. (Top) ETKDG vs mETKDG with embedding. (Bottom) mETKDG with random coordinates vs mETKDG+Eccen with embedding. Comparisons of conformation generation method pairs for the BACE ligands using rRMSD. The blue solid lines indicate $y = x$, and the gray dotted lines represent $\pm 0.15 \text{ \AA}$ deviation.

intramolecular H-bonds in close proximity. However, the optimal geometrical arrangement for an H-bond may not be

obtained. Thus, we investigated the minimization with a standard force field (OpenForceField) as a post-processing step. This was done for the 500 generated conformers with the lowest rRMSD. Subsequently, the rRMSD of the minimized conformers was recalculated with respect to the minimized experimental structure. This places the experimental and generated conformers onto the same potential-energy landscape (note that the rRMSD of the experimental structures before and after minimization can be on the order of 40 \AA).

Figure 15 shows the rRMSD distribution of the top 500 conformers for the two cyclic peptides, decapeptide (top) and CsA (bottom). Three rRMSD distributions are shown per plot, i.e., for mETKDG+Eccen without CPCIs in green, mETKDG +Eccen with CPCIs using the optimal combination of options (scaling factor = 0.5, no counter charges) in orange, and, in blue, the distribution of these 500 conformers obtained with CPCIs followed by a standard force-field minimization. Clearly, the rRMSD distribution shifts to smaller values upon using CPCIs for both cyclic peptides. In addition, the subsequent force-field minimization can produce some conformers that are even closer to the experimental structures. At the same time, other conformers become worse in terms of rRMSD. This is because the energy minima of the OpenForceField and the constraints of the conformation generator may differ. Thus, some conformers fall into energy minima, which are further away from the experimental structure.

The rRMSD distribution of all the other CPCl option combinations are shown in Figures S11 and S12 in the Supporting Information. mETKDG+Eccen together with CPCIs using a scaling factor of 0.5 consistently produces mean rRMSD scores that outperform mETKDG(+Eccen) without CPCIs for both peptides. The inclusion of counter charges or the use of random coordinates does not contribute positively.

Figure 16 illustrates the best conformers generated by mETKDG+Eccen with CPCIs (scaling factor = 0.5, no counter charges) overlaid onto the reference structures after force-field minimization. For decapeptide, all main structural features (i.e., all *trans*-amide bonds and *trans*-annular H-bonds) are well reproduced. For CsA, the conformer with the best rRMSD (bottom left) contains only one of the three *trans*-annular H-

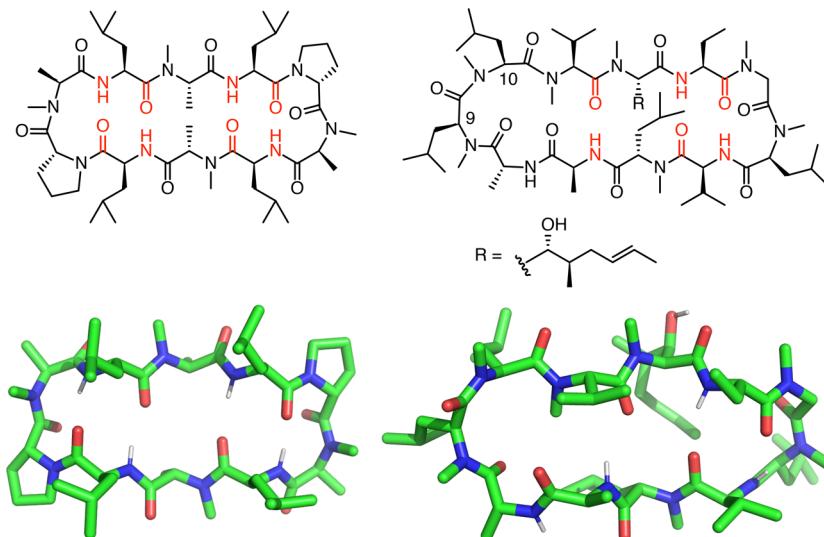


Figure 14. Schematic structure (top) and experimental structure (bottom) of the decapeptide^{51,52} (left) and cyclosporine A (CsA)50 (right). The *trans*-annular hydrogen bonds are marked in red.

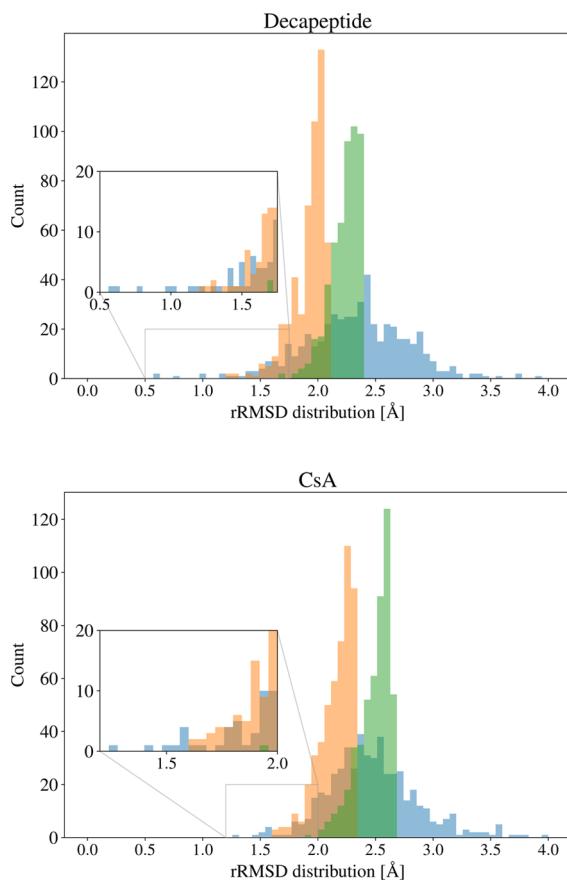


Figure 15. rRMSD distribution of the top 500 conformers for the decapeptide (top) and CsA (bottom) using three different methods: (green) mETKDG+Eccen, (orange) the best combination of options (i.e., mETKDG+Eccen with embedding plus CPCIs with a scaling factor of 1/2 and no counter charges), and (blue) the same optimal method with subsequent force-field minimization using the Open-ForceField. The best rRMSD values using mETKDG+Eccen (green) are 1.68 Å for the decapeptide and 1.94 Å for CsA. Adding CPCIs (orange) brings down the best rRMSD to 1.23 and 1.60 Å. The subsequent force-field minimization reduces the best rRMSD further to 0.60 and 1.27 Å.

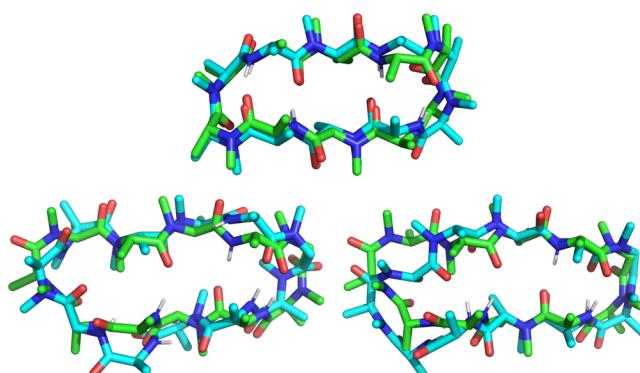


Figure 16. Overlay of experimental structure (green) to the best rRMSD conformer (cyan) for decapeptide (top) and CsA (bottom) using mETKDG+Eccen with CPCIs (scaling factor = 0.5) followed by force-field minimization. Only the macrocyclic ring atom and their direct neighbors (β atoms) are shown.

bonds in the crystal structure and the *cis*-amide bond is not reproduced (the modifications in mETKDG favor *trans*-amide bonds). As the left half of the CsA reference structure is more floppy and disordered, we focused on the reproduction of the right half that contains the three *trans*-annular H-bonds. The conformer with the best rRMSD value based on only the right half of CsA is shown on the bottom right of Figure 16. In this case, two of the three H-bonds observed in the crystal structure are formed (a third H-bond is present in the generated conformer but to a different carbonyl). In addition, the *cis*-amide bond is also reproduced.

In summary, we have demonstrated that the introduction of CPCIs can further bias the sampling of conformers toward the desired region of phase space based on chemical intuition. Just like eccentricity, the application of CPCIs is a flexible guiding principle, which can be selectively applied. The eccentricity constraints work well in conjunction with CPCIs. Finally, we showed that a standard force-field minimization as a post-processing step can be helpful, especially to establish intramolecular hydrogen bonds. All methods in this study are fully automated using our cpeptools package (github.com/rinikerlab/cpeptools/).

CONCLUSIONS

In this study, we presented modifications and additions to different stages of the ETKDG conformer generator that improve its ability to efficiently sample relevant conformations of small and large rings.

The torsional-angle potentials developed here resulted in torsional-angle distributions in the generated conformers, which better match the experimental distributions from the CSD. In particular, the propensity for the chair conformation of aliphatic six-membered rings is better reproduced. This leads to a higher likelihood to generate a good ring conformation close to the crystal structure. The effect is visible in the best rTFD for a small number of conformers, n , and in the median rTFD of the ensemble of generated conformers. However, if enough conformers are generated (e.g., $n = 100$), no significant difference was observed between ETKDG and ETKDG+sr in terms of best rTFD and best rRMSD. As the application of the ring torsional preferences can be beneficial for a small n , we will provide this option in the RDKit ETKDG conformer generator.

For macrocycles, we first demonstrated that the use of random coordinates can be preferable to embedding. Embedding tends to produce flat and round conformations, whereas random coordinates yield less flat and more squashed structures. Next, we introduced modifications to the 1,4-distances in the bounds matrix as well as torsional-angle preferences for cyclic bonds in rings with a size greater than eight (because these behave similar to acyclic bonds). These changes systematically improved ETKDG's ability to reproduce experimental structures of macrocycles. In addition, the application of eccentricity constraints on the backbone to favor squashed conformations was explored. We found that conformers obtained with these 2D elliptical constraints tend to be both squashed and flat. Thus, we suggest using this option in combination with embedding when flat conformations are desired. Finally, the introduction of custom pairwise Coulombic interactions (CPCIs) is showcased in the context of cyclic peptides, which can form intramolecular hydrogen bonds. CPCIs in combination with eccentricity constraints significantly boosted the ability of the conformer generator to reproduce the experimental structures.

The added functionalities are available at github.com/rinikerlab/RDKit_mEtkdg and will be incorporated into the 2020.03 release of the official RDKit.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.0c00025>.

Figures S1, S2, and S4–S12: effect of improved vdW radii, additional figures for the Results subsections on small aliphatic rings and macrocycles ([PDF](#))

Figure S3: torsional-angle distributions from the CSD, ETKDG (only acyclic torsion patterns), and ETKDG+sr (with small-ring torsion patterns) together with the fitted torsional-angle potentials for all 105 SMARTS patterns ([ZIP](#))

Lists of chemical identifiers of the molecules in the test sets for the small rings and macrocycles ([ZIP](#))

List of 105 small-ring SMARTS patterns ([ZIP](#))

AUTHOR INFORMATION

Corresponding Author

Sereina Riniker — Laboratory of Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland; orcid.org/0000-0003-1893-4031; Email: riniker@ethz.ch

Authors

Shuzhe Wang — Laboratory of Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland; orcid.org/0000-0001-7543-7387

Jagna Witek — Laboratory of Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland

Gregory A. Landrum — T5 Informatics GmbH, 4055 Basel, Switzerland; orcid.org/0000-0001-6279-4481

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.0c00025>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Paul Hawkins for providing the macrocycles set, Roger Sayle for the helpful discussion, and the CCDC for agreeing to the use of torsional-angle distributions obtained from the CSD. The authors gratefully acknowledge financial support by the Swiss National Science Foundation (Grant Number 200021-178762) and by ETH Zurich (ETH-34 17-2).

REFERENCES

- (1) Verma, J.; Khedkar, V.; Coutinho, E. 3D-QSAR in Drug Design – A Review. *Curr. Top. Med. Chem.* **2010**, *10*, 95–115.
- (2) Cappel, D.; Dixon, S. L.; Sherman, W.; Duan, J. Exploring Conformational Search Protocols for Ligand-based Virtual Screening and 3-D QSAR Modeling. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 165–182.
- (3) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, High-Quality Test Set for the Validation of Protein-Ligand Docking Performance. *J. Med. Chem.* **2007**, *50*, 726–741.
- (4) Shim, J.; MacKerell, A. D., Jr. Computational Ligand-based Rational Design: Role of Conformational Sampling and Force Fields in Model Development. *Med. Chem. Commun.* **2011**, *2*, 356.
- (5) de Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061.
- (6) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.
- (7) Hawkins, P. C. D.; Nicholls, A. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919–2936.
- (8) Hawkins, P. C. D.; Sader, C.; Danelius, E.; Lood, K.; Erdelyi, M.; Wlodek, S. Conformation Sampling for Molecules in the Solid-state and in Solution. *J. Chem. Inf. Model.* Submitted.
- (9) 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.
- (10) 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.
- (11) 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.

- (12) 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.
- (13) 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.
- (14) 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.
- (15) Coutsias, E. A.; Lexa, K. W.; Wester, M. J.; Pollock, S. N.; Jacobson, M. P. Exhaustive Conformational Sampling of Complex Fused Ring Macrocycles Using Inverse Kinematics. *J. Chem. Theory Comput.* **2016**, *12*, 4674–4687.
- (16) 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.
- (17) Griewel, A.; Kayser, O.; Schlosser, J.; Rarey, M. Conformational Sampling for Large-Scale Virtual Screening: Accuracy versus Ensemble Size. *J. Chem. Inf. Model.* **2009**, *49*, 2303–2311.
- (18) Friedrich, N.-O.; Flachsenberg, F.; Meyder, A.; Sommer, K.; Kirchmair, J.; Rarey, M. ConformatoR: A Novel Method for the Generation of Conformer Ensembles. *J. Chem. Inf. Model.* **2019**, *59*, 731–742.
- (19) Schärfer, C.; Schulz-Gasch, T.; Hert, J.; Heinzerling, L.; Schulz, B.; Inhester, T.; Stahl, M.; Rarey, M. CONFECT: Conformations from an Expert Collection of Torsion Patterns. *J. Chem. Inf. Model.* **2013**, *8*, 1690–1700.
- (20) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462–2474.
- (21) The RDKit: Open-Source Cheminformatics Software, version 2019.03.1. GitHub, 2019; <https://www.rdkit.org>.
- (22) Riniker, S.; Landrum, G. A Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562–2574.
- (23) Schwab, C. H. Conformations and 3D Pharmacophore Searching. *Drug Discovery Today* **2010**, *7*, e245–e253.
- (24) Hawkins, P. C. D. Conformation Generation: The State of the Art. *J. Chem. Inf. Model.* **2017**, *57*, 1747–1756.
- (25) 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.
- (26) 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.
- (27) 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*, 7898–7920.
- (28) Habgood, M. Bioactive Focus in Conformational Ensembles: A Pluralistic Approach. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 1073–1083.
- (29) Friedrich, N.-O.; Meyder, A.; de Bruyn Kops, C.; Sommer, K.; Flachsenberg, F.; Rarey, M.; Kirchmair, J. High-Quality Dataset of Protein-Bound Ligand Conformations and Its Application to Benchmarking Conformer Ensemble Generators. *J. Chem. Inf. Model.* **2017**, *57*, 529–539.
- (30) 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.
- (31) Blaney, J. M.; Dixon, J. S. Distance Geometry in Molecular Modeling. *Rev. Comput. Chem.* **1994**, 299–335.
- (32) Simm, G. N. C.; Hernández-Lobato, J. M. A Generative Model for Molecular Distance Geometry. *arXiv e-prints arXiv:1909.11459* **2019**.
- (33) van der Velden, N. S.; Kälin, N.; Helf, M. J.; Piel, J.; Freeman, M. F.; Künzler, M. Autocatalytic Backbone N-Methylation in a Family of Ribosomal Peptide Natural Products. *Nat. Chem. Biol.* **2017**, *13*, 833–835.
- (34) Poongavanam, V.; Danielius, E.; Peintner, S.; Alcaraz, L.; Caron, G.; Cummings, M. D.; Wlodek, S.; Erdelyi, M.; Hawkins, P. C. D.; Ermondi, G.; Kihlberg, J. Conformational Sampling of Macrocyclic Drugs in Different Environments: Can We Find the Relevant Conformations? *ACS Omega* **2018**, *3*, 11742–11757.
- (35) Quijano, M. R.; Zach, C.; Miller, F. S.; Lee, A. R.; Imani, A. S.; Künzler, M.; Freeman, M. F. Distinct Autocatalytic α -N-Methylating Precursors Expand the Borosin RiPP Family of Peptide Natural Products. *J. Am. Chem. Soc.* **2019**, *141*, 9637–9644.
- (36) Ono, S.; Naylor, M. R.; Townsend, C. E.; Okumura, C.; Okada, O.; Lokey, R. S. Conformation and Permeability: Cyclic Hexapeptide Diastereomers. *J. Chem. Inf. Model.* **2019**, *59*, 2952–2963.
- (37) Cummings, M. D.; Sekharan, S. Structure-Based Macrocyclic Design in Small-Molecule Drug Discovery and Simple Metrics To Identify Opportunities for Macrocyclization of Small-Molecule Ligands. *J. Med. Chem.* **2019**, *62*, 6843–6853.
- (38) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. B* **2016**, *72*, 171–179.
- (39) Doak, B. C.; Zheng, J.; Dobritzsch, D.; Kihlberg, J. How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. *J. Med. Chem.* **2016**, *59*, 2312–2327.
- (40) Qian, Z.; Dougherty, P. G.; Pei, D. Targeting Intracellular Protein–Protein Interactions with Cell-Permeable Cyclic Peptides. *Curr. Opin. Chem. Biol.* **2017**, *38*, 80–86.
- (41) Witek, J.; Keller, B. G.; Blatter, M.; Meissner, A.; Wagner, T.; Riniker, S. Kinetic Models of Cyclosporin A in Polar and Apolar Environments Reveal multiple Congruent Conformational States. *J. Chem. Inf. Model.* **2016**, *56*, 1547–1562.
- (42) Witek, J.; Mühlbauer, M.; Keller, B. G.; Blatter, M.; Meissner, A.; Wagner, T.; Riniker, S. Interconversion Rates between Conformational States as Rationale for the Membrane Permeability of Cyclosporines. *ChemPhysChem* **2017**, *18*, 3309–3314.
- (43) Witek, J.; Wang, S.; Schroeder, B.; Lingwood, R.; Dounas, A.; Roth, H.-J.; Fouché, M.; Blatter, M.; Lemke, O.; Keller, B.; Riniker, S. Rationalization of the Membrane Permeability Differences in a Series of Analogue Cyclic Decapeptides. *J. Chem. Inf. Model.* **2018**, *59*, 294–308.
- (44) O’Boyle, N. M.; Guha, R.; Willighagen, E. L.; Adams, S. E.; Alvarsson, J.; Bradley, J.-C.; Filippov, I. V.; Hanson, R. M.; Hanwell, M. D.; Hutchison, G. R.; James, C. A.; Jeliazkova, N.; Lang, A. S. I. D.; Langner, K. M.; Lonie, D. C.; Lowe, D. M.; Pansanel, J.; Pavlov, D.; Spjuth, O.; Steinbeck, C.; Tenderholt, A. L.; Theisen, K. J.; Murray-Rust, P. Open Data, Open Source and Open Standards in Chemistry: The Blue Obelisk Five Years On. *Aust. J. Chem.* **2011**, *3*, 37.
- (45) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
- (46) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076–6092.
- (47) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations. *J. Chem. Inf. Model.* **2018**, *58*, 579–590.
- (48) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **2003**, *10*, 980–980.
- (49) Feng, Z.; Chen, L.; Maddula, H.; Akcan, O.; Oughtred, R.; Berman, H. M.; Westbrook, J. Ligand Depot: A Data Warehouse for Ligands Bound to Macromolecules. *Bioinformatics* **2004**, *20*, 2153–2155.
- (50) Mikol, V.; Kallen, J.; Pfügl, G.; Walkinshaw, M. D. X-ray Structure of a Monomeric Cyclophilin A–Cyclosporin A Crystal Complex at 2·1 Å Resolution. *J. Mol. Biol.* **1993**, *234*, 1119–1130.
- (51) Fouché, M.; Schäfer, M.; Berghausen, J.; Desrayaud, S.; Blatter, M.; Piéchon, P.; Dix, I.; Martin Garcia, A.; Roth, H.-J. Design and Development of a Cyclic Decapeptide Scaffold with Suitable Properties for Bioavailability and Oral Exposure. *ChemMedChem* **2016**, *11*, 1048–1059.

(52) Fouché, M.; Schäfer, M.; Blatter, M.; Berghausen, J.; Desrayaud, S.; Roth, H.-J. Pharmacokinetic Studies around the Mono- and Difunctionalization of a Bioavailable Cyclic Decapeptide Scaffold. *ChemMedChem* **2016**, *11*, 1060–1068.

(53) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.

(54) 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.

(55) Fitzgibbon, A.; Pilu, M.; Fisher, R. Direct Least Squares Fitting of Ellipses. *Proceedings of 13th International Conference on Pattern Recognition*; IEEE, 1996; pp 476–480.

(56) Dauber, P.; Hagler, A. T. Crystal Packing, Hydrogen Bonding, and the Effect of Crystal Forces on Molecular Conformation. *Acc. Chem. Res.* **1980**, *13*, 105–112.