

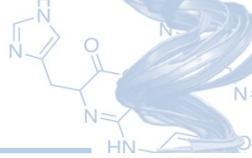


# Improving Conformer Quality in ETKDG

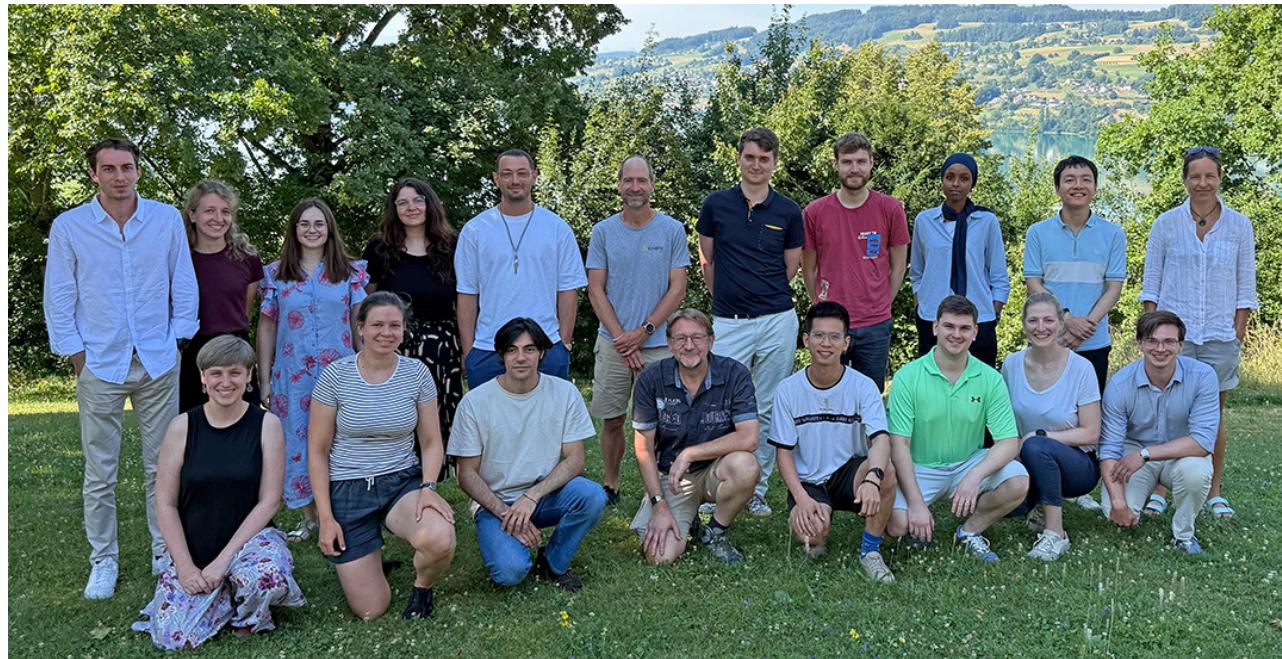
RDKit UGM

10.09.2025

Niels Maeder, CCG @ ETHZ

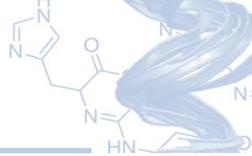


# Acknowledgments



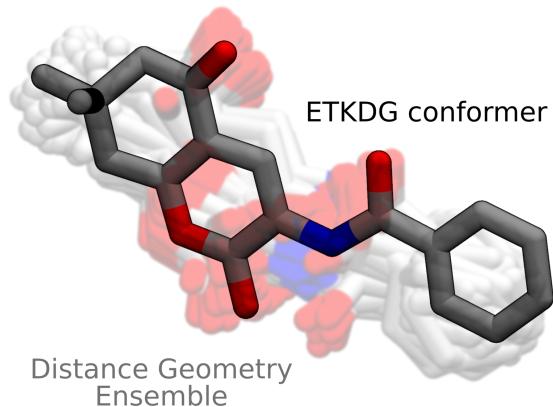
Prof. Sereina Riniker  
Dr. Gregory Landrum

Jessica Braun



# Goals for a Conformer Generator

## Ensemble Quality



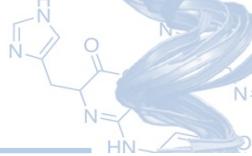
Is there something in my ensemble that's close to a reference?

## Runtime Performance

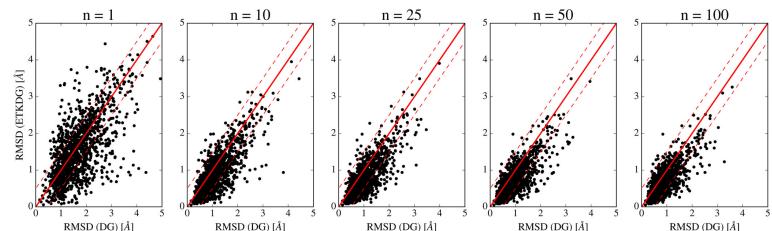
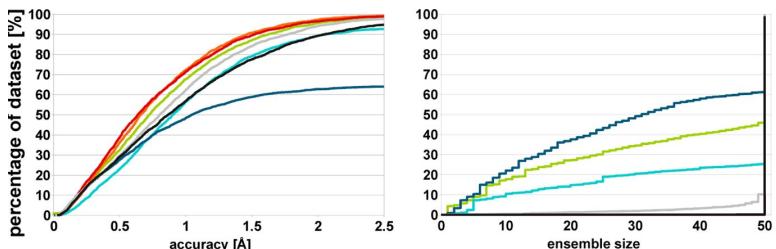


How much does it cost me?

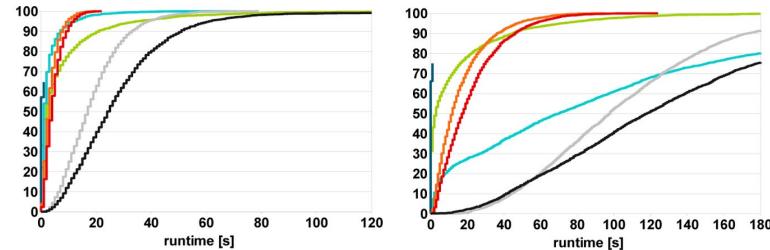
# Validation



## Ensemble Quality



## Runtime Performance



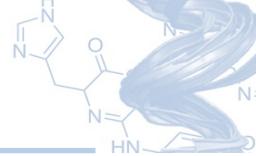
Accuracy

Speed

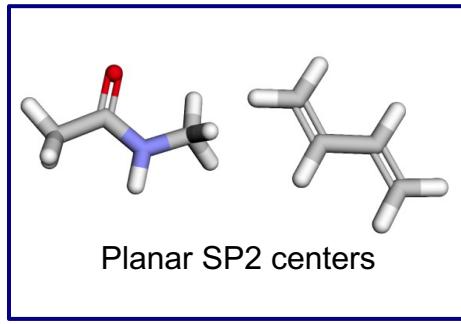
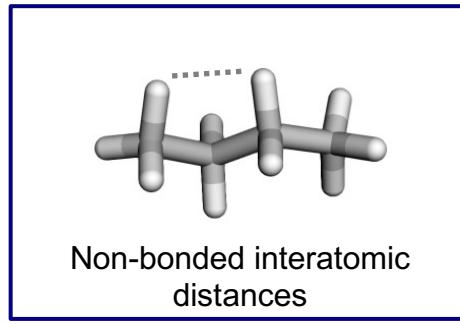
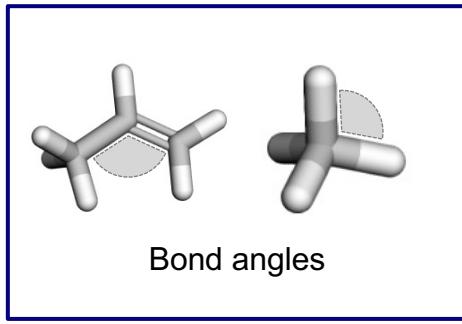
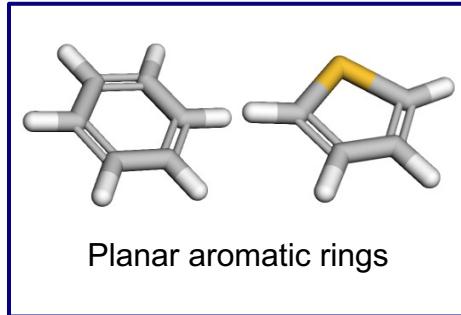
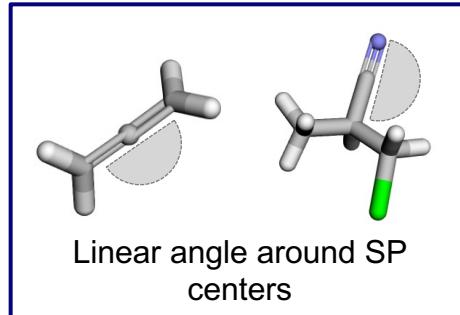
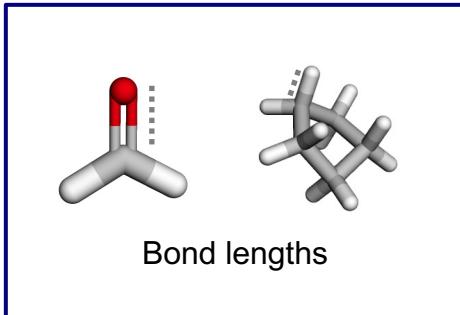
Tradeoff between speed and accuracy

Aim for the fastest possible implementation for a given accuracy

# Missing Third Pillar

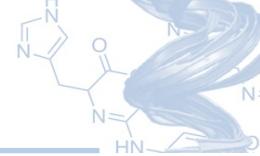


## Conformer Quality

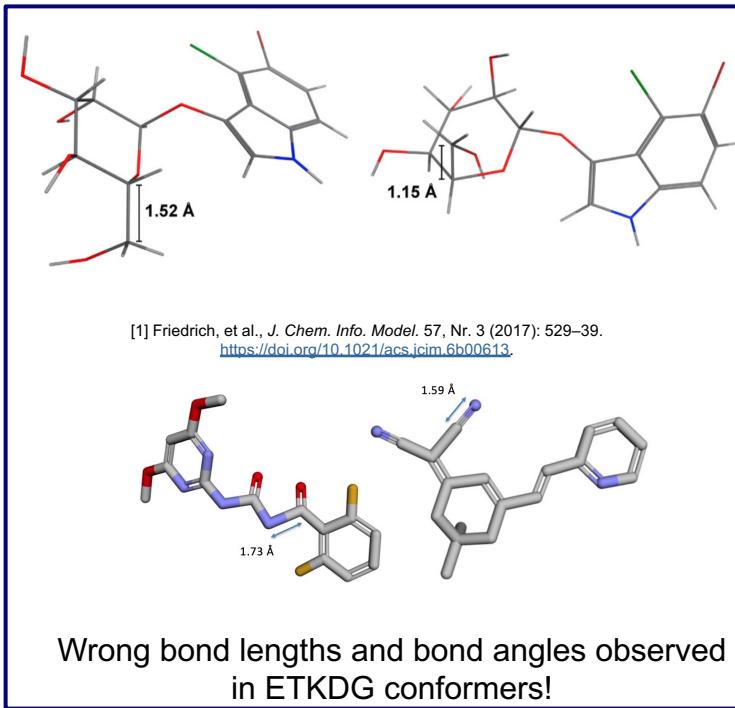


Every conformer generated should be physically reasonable; however, this is often overlooked in benchmarks.

# How Is This Tested?



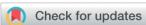
Basic conformer quality is often assumed and verified by manually reviewing examples.



Buttenschoen *et al.* recently published a test suite that checks docked poses for physical reasonableness.

## EDGE ARTICLE

[View Article Online](#)  
[View Journal](#) | [View Issue](#)



Cite this: *Chem. Sci.*, 2024, 15, 3130

All publication charges for this article have been paid for by the Royal Society of Chemistry

**PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences†**

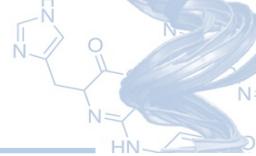
Martin Buttenschoen, Garrett M. Morris and Charlotte M. Deane \*



# PoseBusters

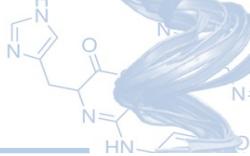
Buttenschoen, et al., *Chem. Sci.* 15, no. 9 (2024): 3130–39  
<https://doi.org/10.1039/D3SC04185A>.

<https://github.com/maabuu/posebusters?tab=readme-ov-file>

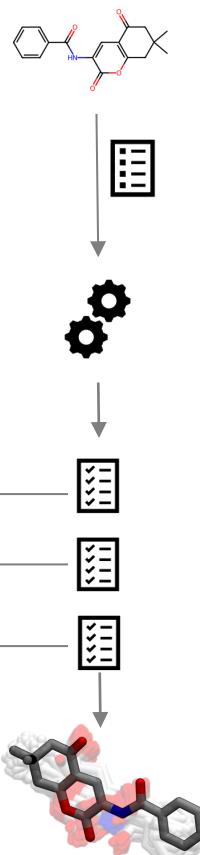
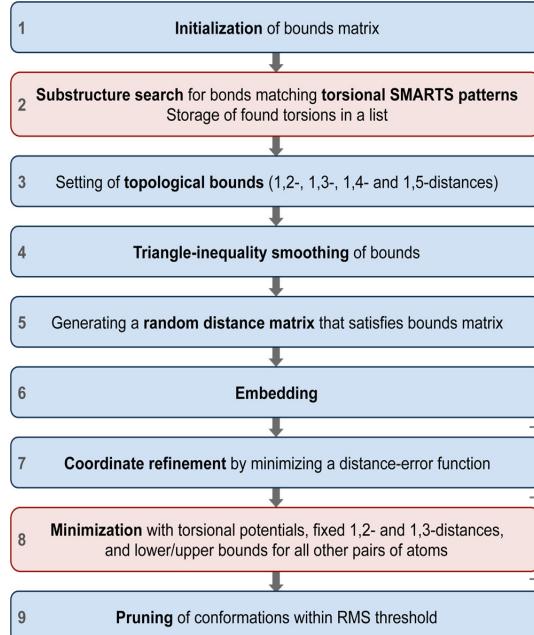


Where can we improve?

What do we aim for?



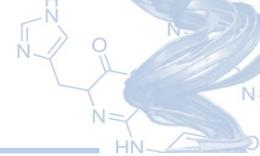
# Possible Areas of Improvement



- Force Field used for bounds matrix generation
- SMARTS patterns / torsion fits
- Heuristics in bounds matrix
- Improve initial embedding
- Add / remove K-terms
- Change coordinate refinement procedure
- Add / adapt / remove checks

C  
O  
S  
t

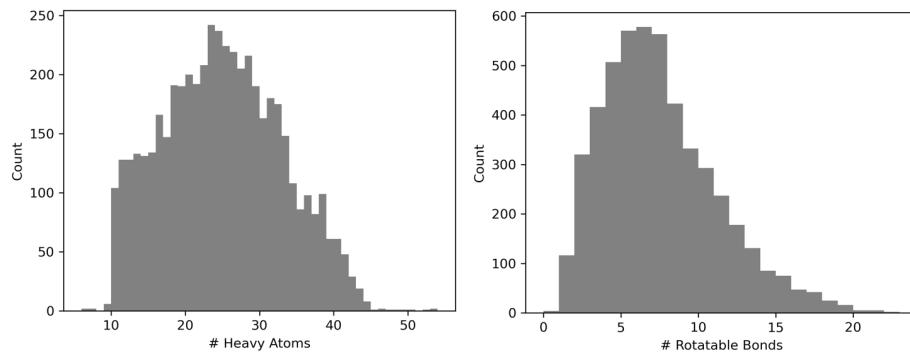
# Expectation Management



Platinum dataset [1] (~4500 molecules) including coverage of small ring and macrocycle torsion patterns

Selection of COD structures (~425 mols) that cover as many torsions patterns as possible

Discard unphysical crystal structures



Devised our own benchmarking tool, checking for things we ask the algorithm for:

Steric clashes

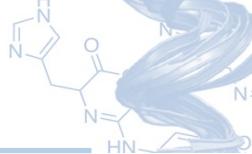
Violation of bounds matrix

Dihedral angles in aromatic rings

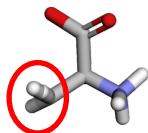
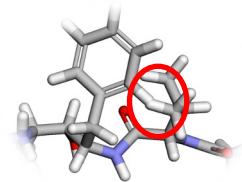
Planarity of SP2 centers

Bond angle around SP2 centers

Bond angle around SP center



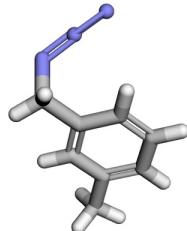
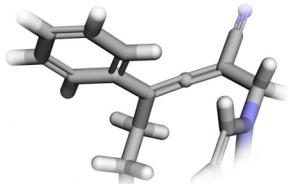
# Logical Fixes



No distance term added to topologically distant atom pairs

Remaining 4D component so distance is satisfied in 4D but not 3D

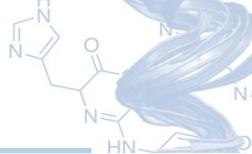
Easily fixable by adding a new check after embedding!



No K term for alene / cumulene like structures

Adding the term fixes this easily!

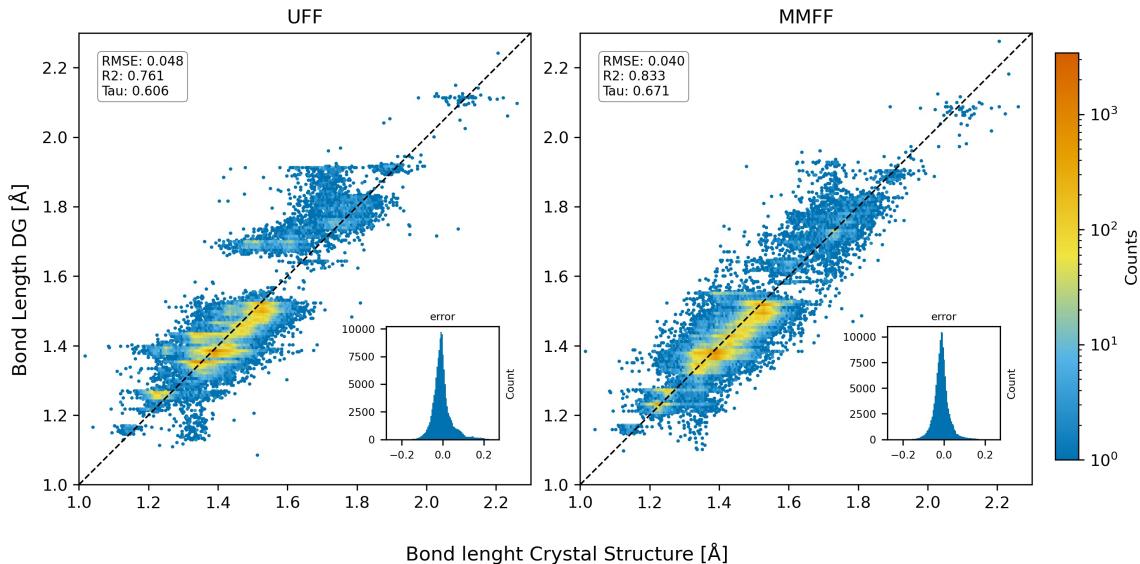
-  Force Field used for bounds matrix generation
-  SMARTS patterns / torsion fits
-  Heuristics in bounds matrix
-  Improve initial embedding
-  Add / remove K-terms
-  Change coordinate refinement procedure
-  Add / adapt / remove checks



# Bond Lengths and Bond Angles

Currently, the UFF [2] is used for estimating the bounds matrix 1-2 and 1-3 distances

Can bond lengths and angles be improved by switching to the MMFF? [3]

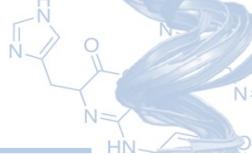


- Force Field used for bounds matrix generation
- SMARTS patterns / torsion fits
- ⚙️ Heuristics in bounds matrix
- ⚙️ Improve initial embedding
- ⚙️ Add / remove K-terms
- ⚙️ Change coordinate refinement procedure
- ⚙️ Add / adapt / remove checks

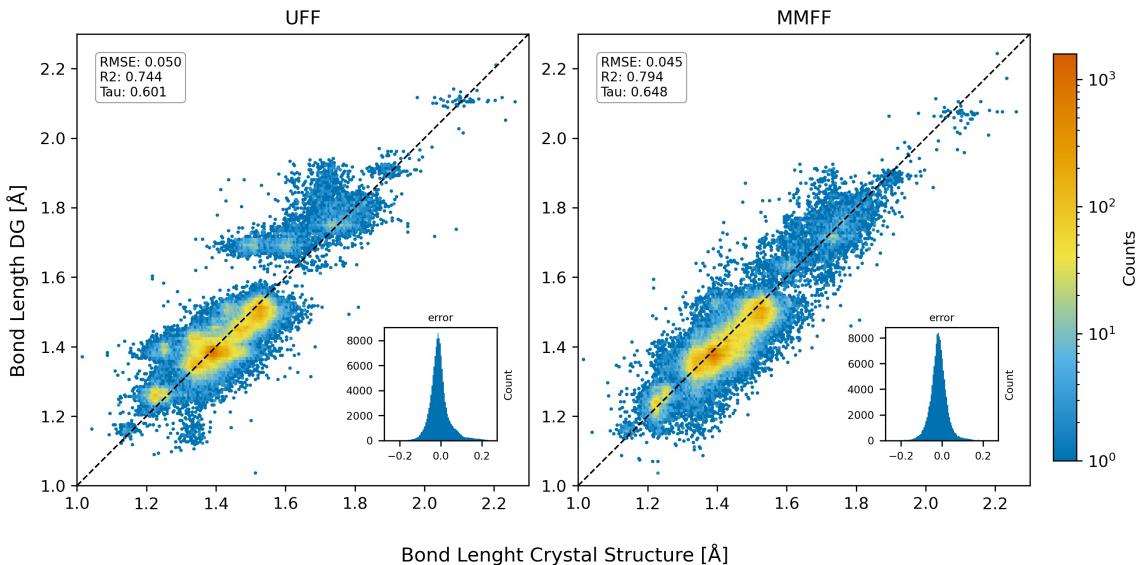
[2] Rappe, et al., *J. Am. Chem. Soc.* 114, Nr. 25 (1992): 10024–35. <https://doi.org/10.1021/ja00051a040>.

[3] Halgren, Thomas A. *J. Comp. Chem.* 17, Nr. 5–6 (1996): 490–519. [https://doi.org/10.1002/\(SICI\)1096-987X\(199604\)17:5/6%253C490::AID-JCC1%253E3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6%253C490::AID-JCC1%253E3.0.CO;2-P).

# Bond Lengths and Bond Angles

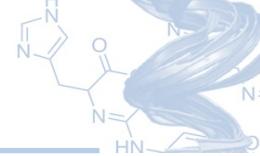


Applying the ET / K terms decreases the improvement significantly!



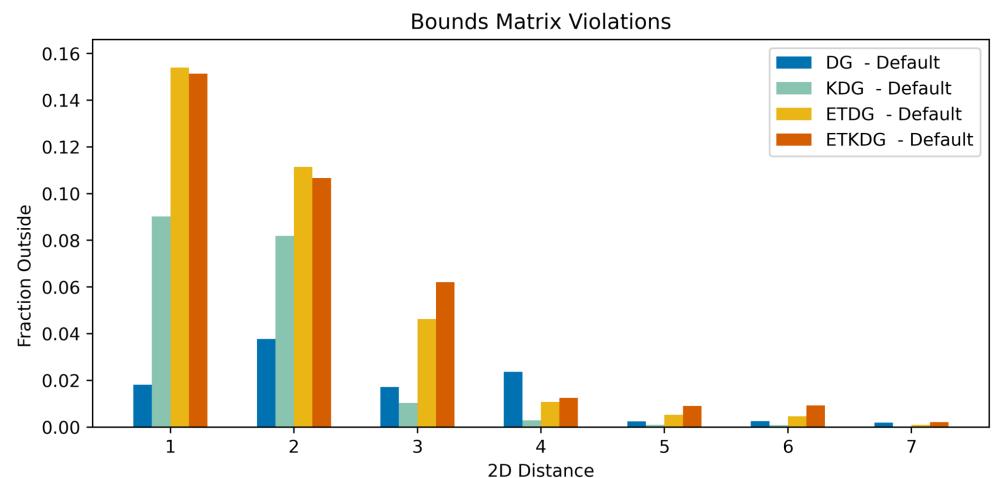
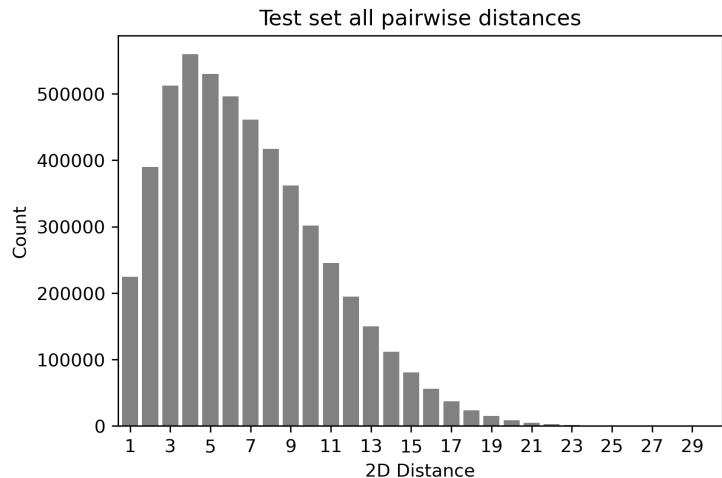
- Force Field used for bounds matrix generation
- SMARTS patterns / torsion fits
- ⚙️ Heuristics in bounds matrix
- ⚙️ Improve initial embedding
- ⚙️ Add / remove K-terms
- ⚙️ Change coordinate refinement procedure
- 📝 Add / adapt / remove checks

# Bounds Matrix Violations

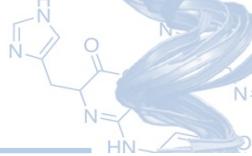


How many of our distances are inside the bounds matrix (possible conformer space)?

Extra minimization with ET and K terms push our distances massively out of bounds!

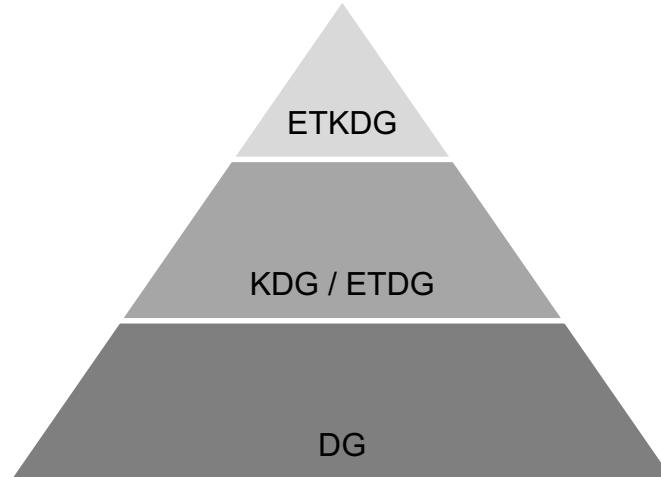


We first need to make sure the algorithm acknowledges the input before further fine-tuning the input itself!



# Optimizing the Optimization

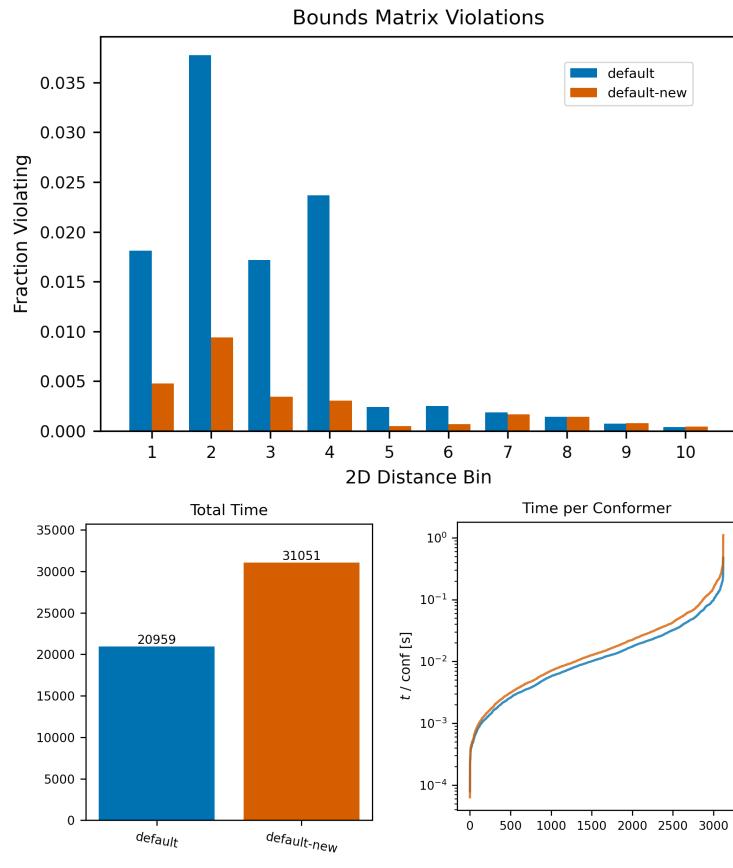
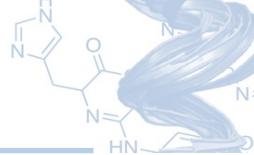
-  Force Field used for Bounds Matrix Generation
-  SMARTS Patterns / Torsion Fits
-  Heuristics in bounds matrix
-  Improve initial embedding
-  Add / remove K Terms
-  Change coordinate refinement procedure
-  Add / adapt / remove Checks



Fine-tune the relative weights of the force constants in the three minimizations to improve coherence with the bounds matrix

Start with plain DG and build on a good foundation

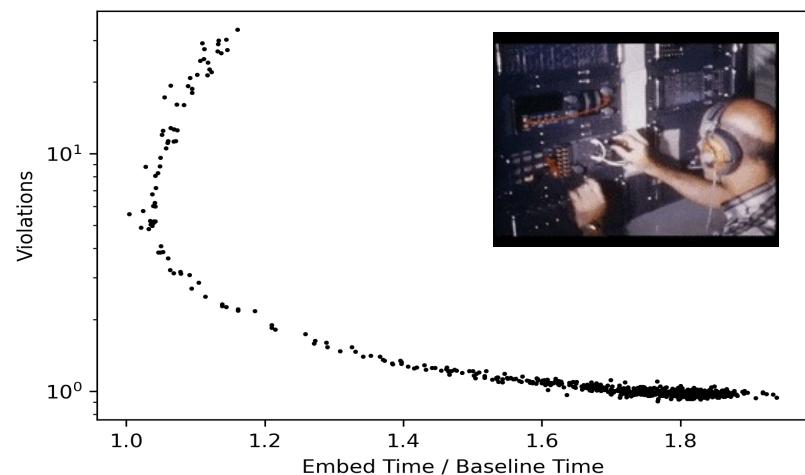
# First Try



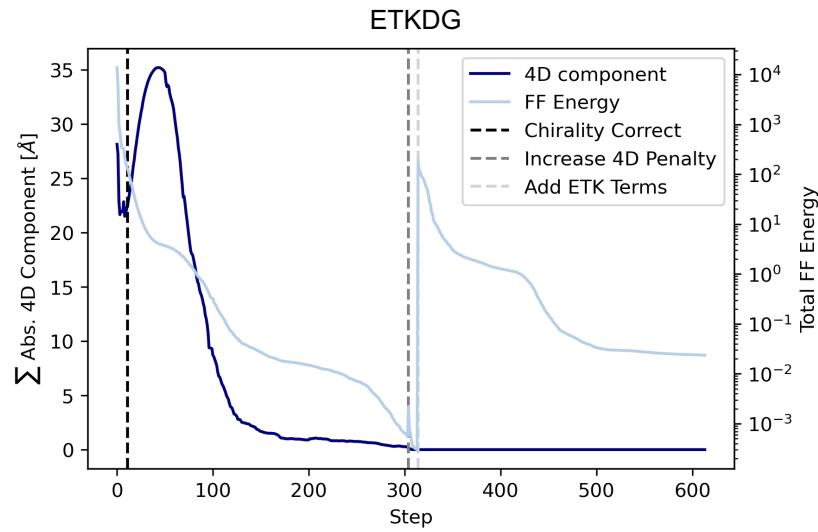
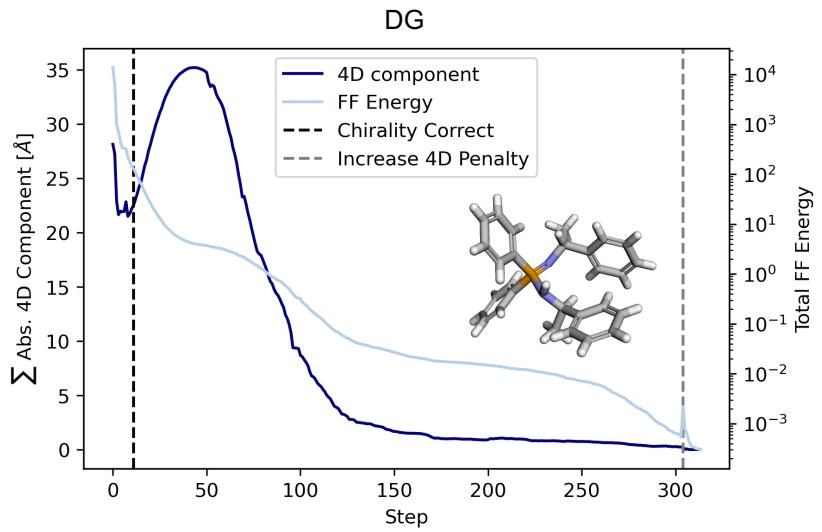
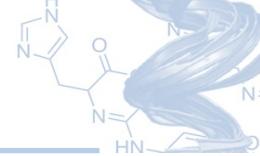
Tens of parameters to tune just for the DG part itself

Reduction of bounds matrix violations achievable by increasing force constants

Better bounds coherence needs longer minimization time!



# Time Spent in Different Minimizations



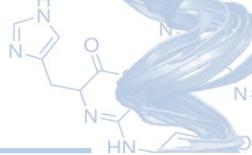
$$V_1 = V_{4D}^{Dist} + V_{3D}^{Chiral}$$

$$V_2 = V_{4D}^{Dist} + V_{4D}^{Penalty}$$

$$V_3 = V_{3D}^{Dist} + V_{3D}^{ET} + V_{3D}^K$$

Overhead from sequential setup of the coordinate refinement

Force field creation (and deletion) up to 3 times, 4D minimization often for only 10 minimization steps



# All In One - Idea

our purposes. Calculating analytical first and second derivatives with respect to some coordinate  $x$  of one of the four atoms is actually not too hard because

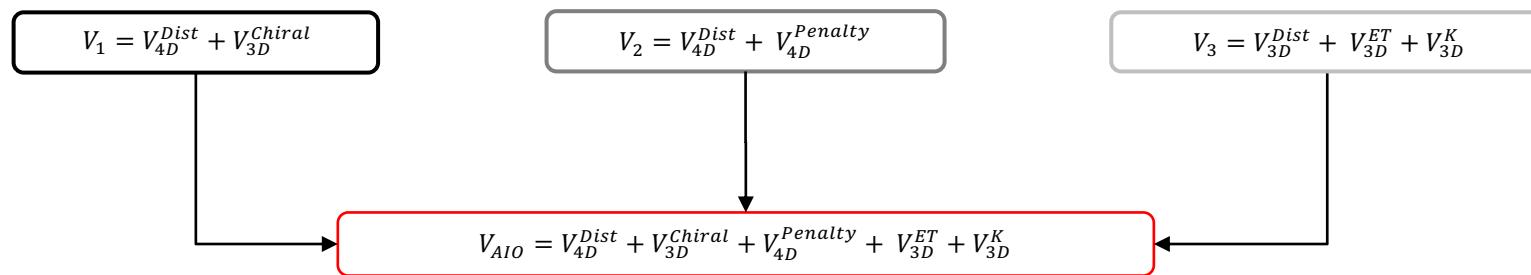
$$\mathbf{a} \cdot \mathbf{b} = -\vec{ij} \cdot \vec{kl} + \frac{(\vec{ij} \cdot \vec{kj})(\vec{kl} \cdot \vec{kj})}{\vec{kj} \cdot \vec{kj}} \quad (12)$$

and similarly for  $\mathbf{a} \cdot \mathbf{a}$  and  $\mathbf{b} \cdot \mathbf{b}$ . Then

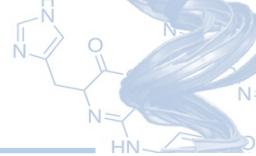
$$\frac{\partial \cos \phi}{\partial x} = \frac{2 \frac{\partial \mathbf{a} \cdot \mathbf{b}}{\partial x} (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) - (\mathbf{a} \cdot \mathbf{b})(\mathbf{b} \cdot \mathbf{b}) \frac{\partial \mathbf{a} \cdot \mathbf{a}}{\partial x} - (\mathbf{a} \cdot \mathbf{b})(\mathbf{a} \cdot \mathbf{a}) \frac{\partial \mathbf{b} \cdot \mathbf{b}}{\partial x}}{2(\mathbf{a} \cdot \mathbf{a})^{3/2}(\mathbf{b} \cdot \mathbf{b})^{3/2}} \quad (13)$$

Second-derivative expressions are not actually all that much worse.

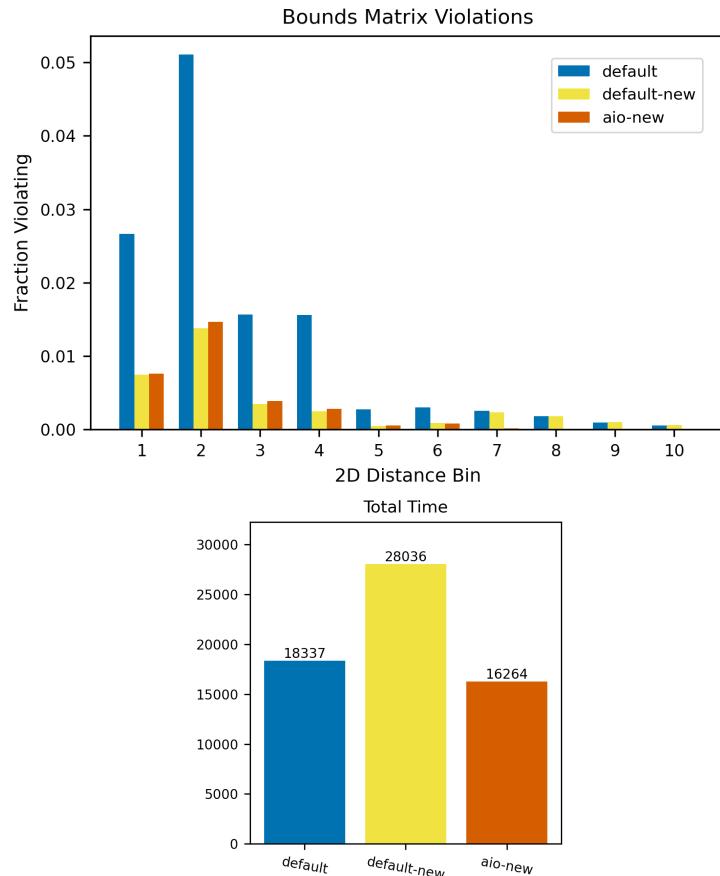
chirality of the group of atoms considered. Since a set of four points (e.g. a tetrahedron) has no well-defined chirality in four dimensions, there is no way to distinguish between the two forms and hence we cannot obtain the sign of  $\xi$ . We have therefore chosen to evaluate the improper dihedral only in  $xyz$ -space, i.e. the 3D-projection of the molecule should have the correct chirality. The improper dihedral interaction can now be evaluated in the



Crippen, G. M., *J. Chem. Phys.* 91, no. 25 (1987): 6341–43. <https://doi.org/10.1021/j100309a007>.  
van Schaik, et. al., *J. Mol. Biol.* 234, no. 3 (1993): 751–62. <https://doi.org/10.1006/jmbi.1993.1624>.



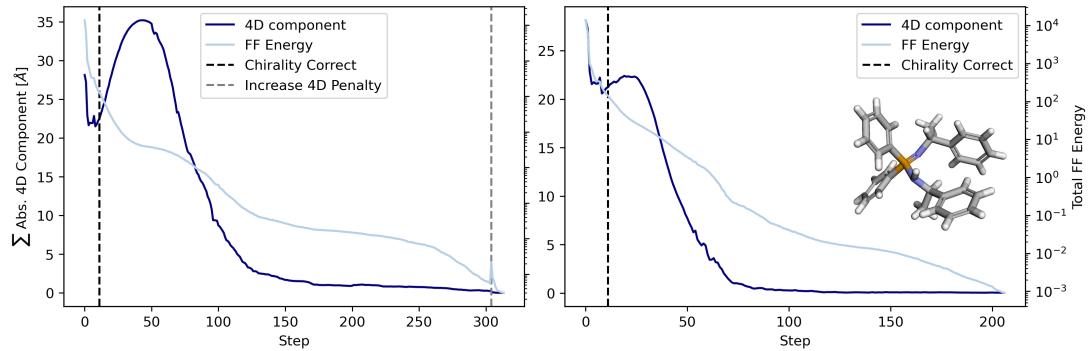
# All In One - DG



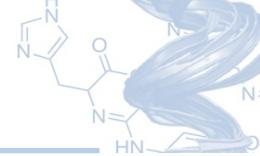
For compounds with no chiral sets, there is no difference in performance between default and all in one, since the force field is practically the same

For chiral compounds, drastic decrease in run time performance observed

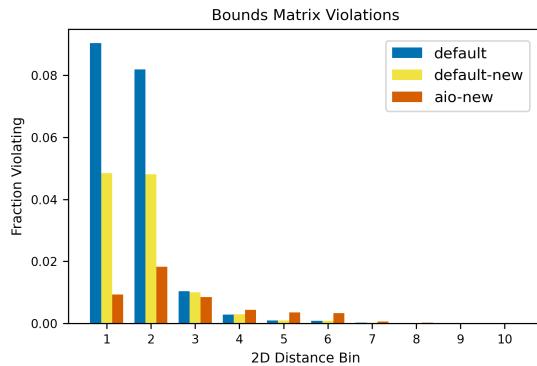
This comes both, from having less overhead creating and deleting objects as well as spending less time in the minimizer itself!



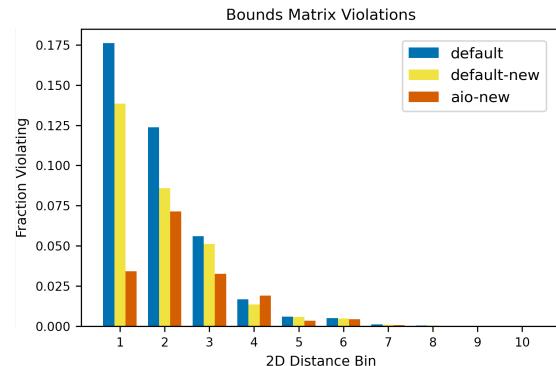
# All In One – K- / ET- / ETKDG



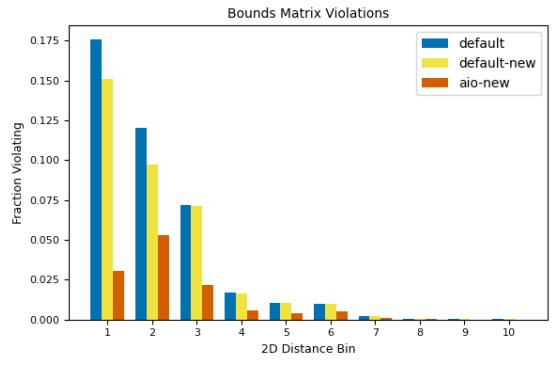
KDG



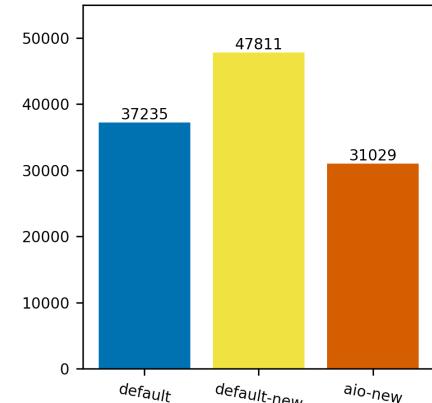
ETDG



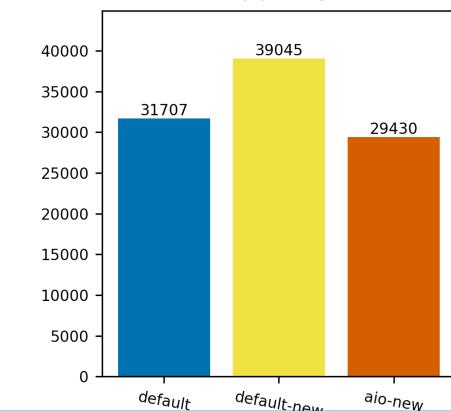
ETKDG



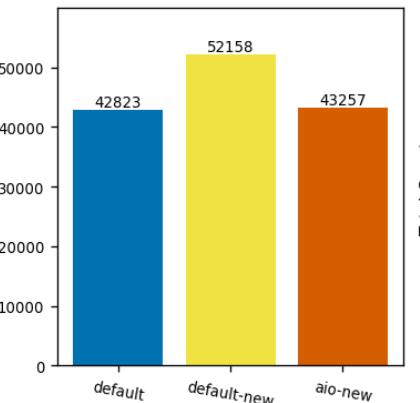
Total Time



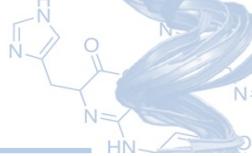
Total Time



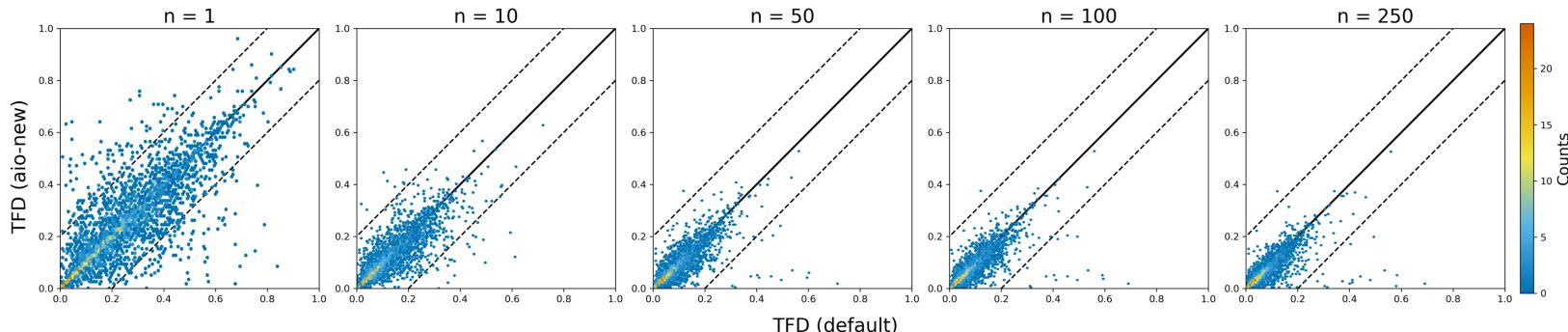
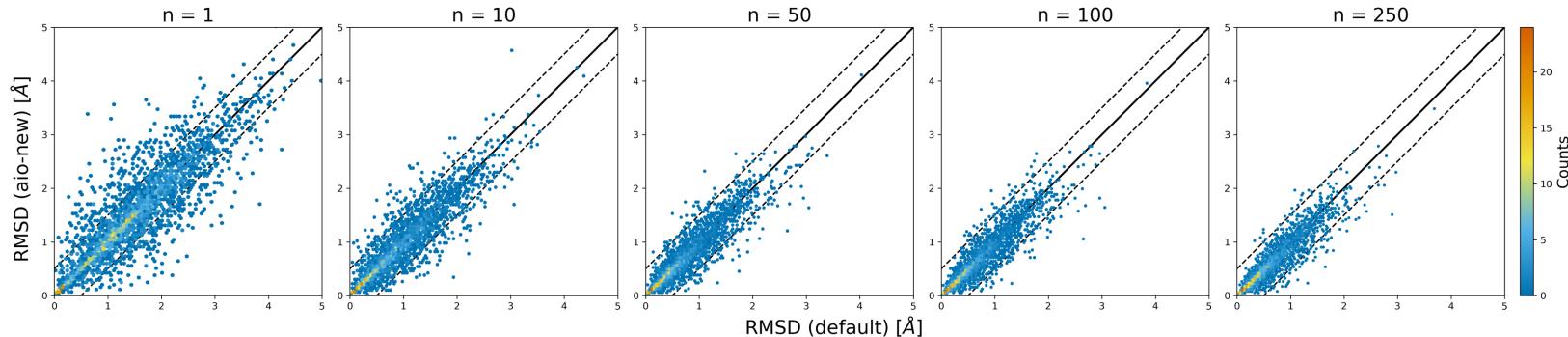
Total Time



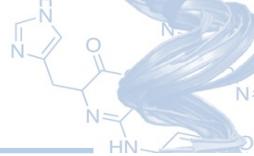
# Ensemble Quality



For ETKDG, both TFD and RMSDs are mostly unchanged

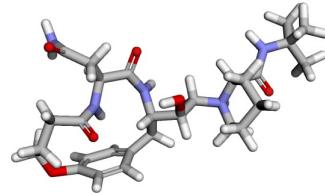
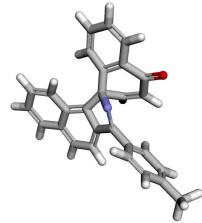


# Problems

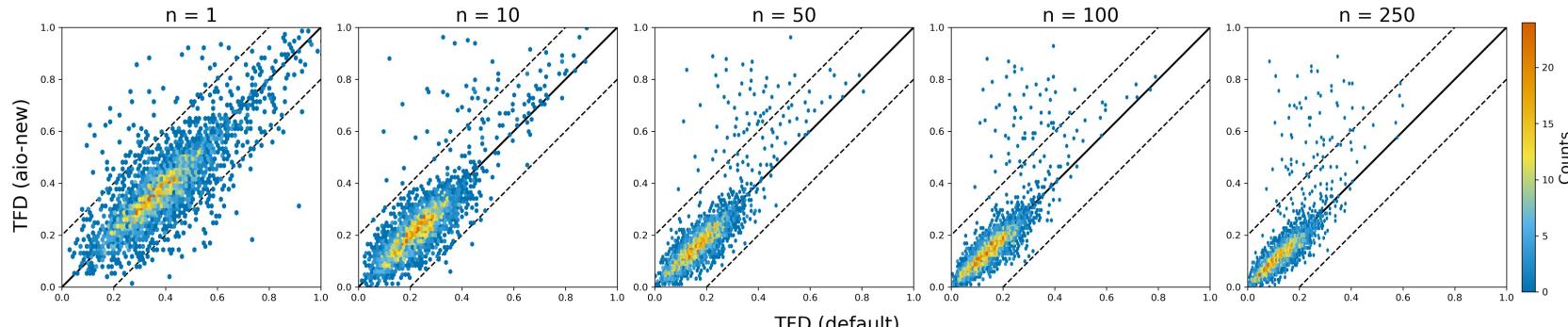


Conflicting ET and K terms lead to increase in failures, and prolonged minimizations.

99.5 % of conformers generated with K terms are perfectly fine, we need to investigate the last 0.5%!

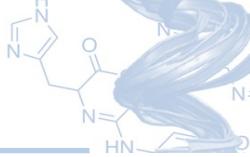


Increase in TFD values for observed for 'better quality' molecules, improvement of 1-4 heuristics in bounds matrix necessary?

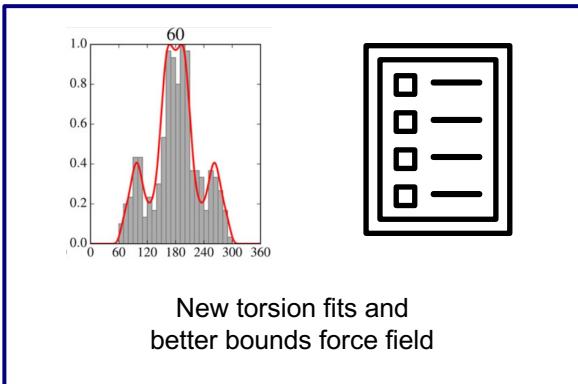
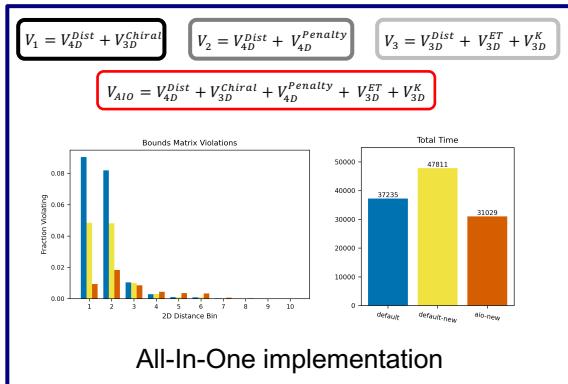
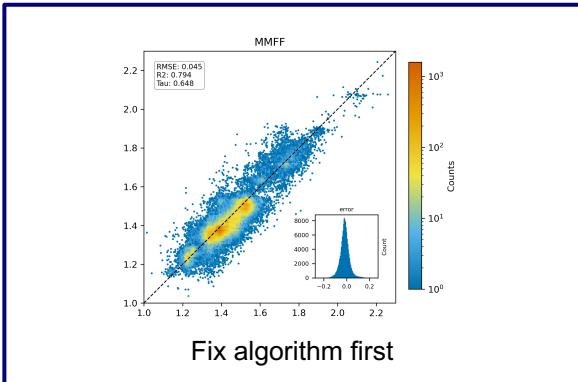
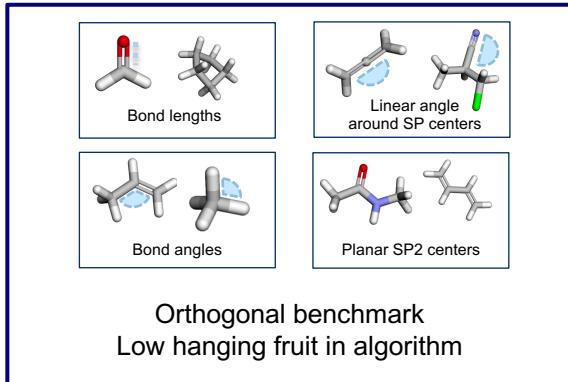


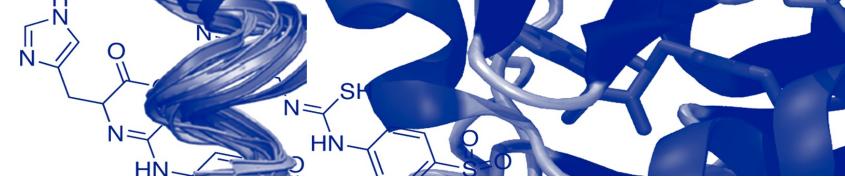


# When Will This Be Out?



# Conclusions & Outlook





# Thank You!