



Tensor Train Optimization for Conformational Sampling of Organic Molecules

Life Sciences

Published by Journal of Chemical Theory and Computation
(American Chemical Society)

2025

terraquantum.swiss

Supporting Information: Tensor Train Optimization for Conformational Sampling of Organic Molecules

Christopher Zurek,[†] Ruslan A. Mallaev,[‡] Alexander C. Paul,[‡] Nils van
Staalduinen,[†] Philipp Pracht,[¶] Roman Ellerbrock,^{*,‡} and Christoph Bannwarth^{*,†}

[†]*Institute of Physical Chemistry, RWTH Aachen University, Aachen, Germany*

[‡]*Terra Quantum AG, St. Gallen, Switzerland*

[¶]*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge,
Massachusetts, USA*

E-mail: romanellerbrock@gmail.com; bannwarth@pc.rwth-aachen.de

1 Setup of Internal Coordinates

Fig. S1 outlines the algorithm to determine BAT coordinates from the graph representation, which ensures a consistent and reproducible definition of the coordinates.

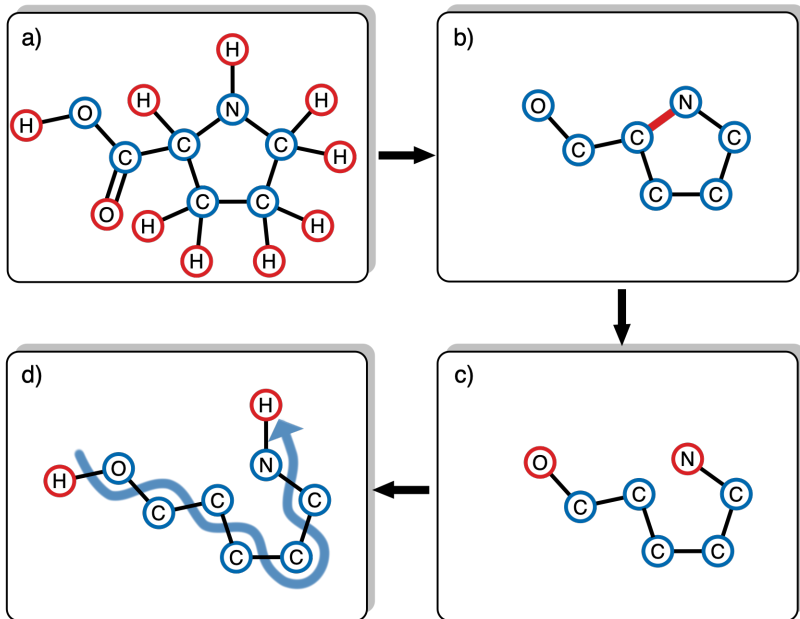


Figure S1: Algorithm for automated generation of BAT coordinates. a) Remove all terminal nodes, b) remove one edge of every ring, c) add one node to the terminal atoms, d) determine origin as one of the two most distant nodes and add atoms to the BAT coordinate matrix in order of their shortest-path distance from the origin.

First of all, terminal nodes are removed and subsequently one edge of every ring is removed from the molecular graph, leading to a graph containing only heavy atoms. For every terminal node in this modified graph, one of the previously removed neighbors is added, to provide reference atoms for the definition of proper dihedral angles. As origin for the construction of the BAT coordinate matrix one of the two most distant nodes is chosen, with respect to their shortest-path distance. In the following, all atoms contained in the modified graph are added in order of their distance from the origin, using exactly one proper dihedral angle per bond. Finally, every atom, which is not part of the modified graph is added referenced via improper dihedral angles.

2 Structural Formulas

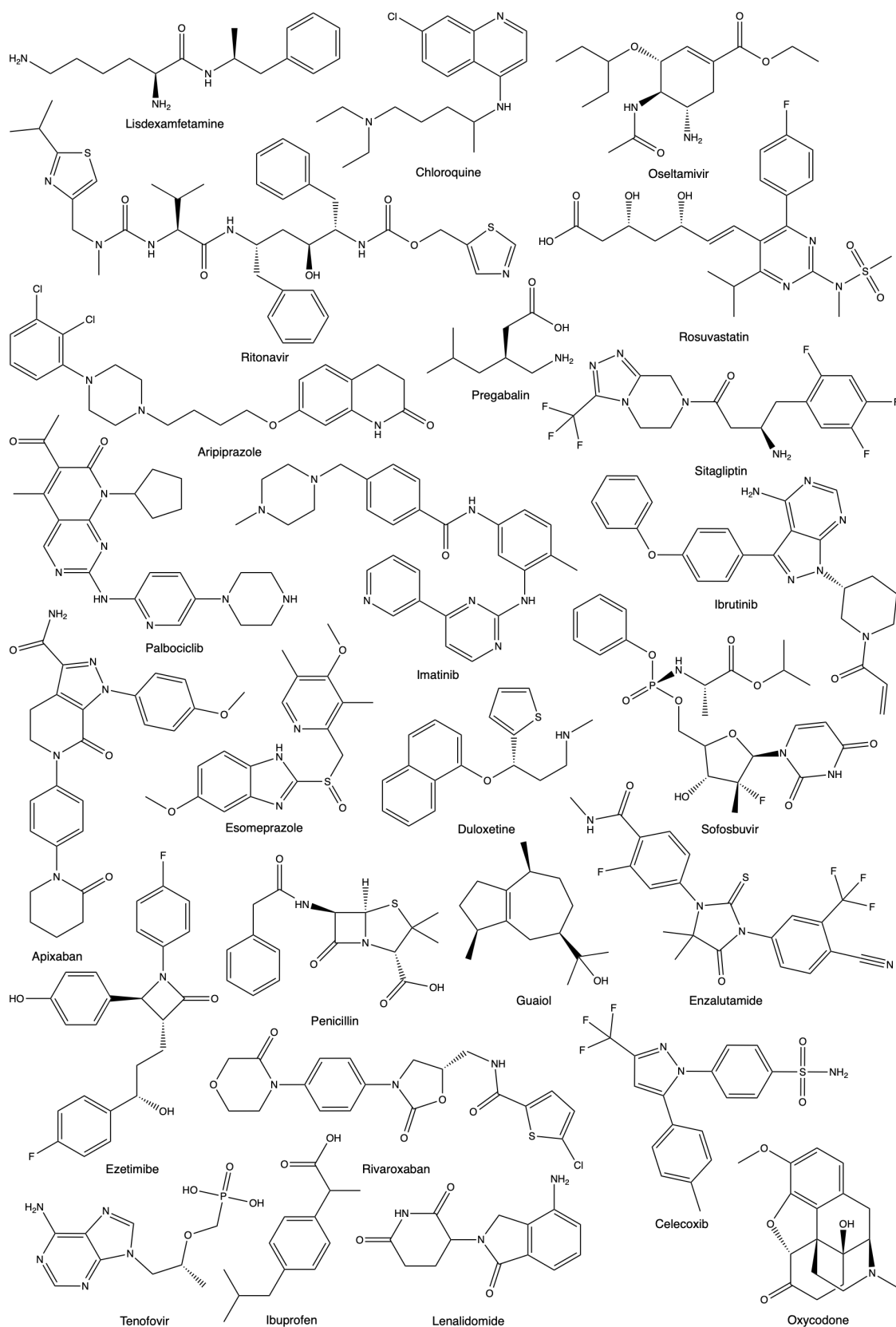


Figure S2: Structural formulas of all molecules included in the CD25 benchmark set.¹

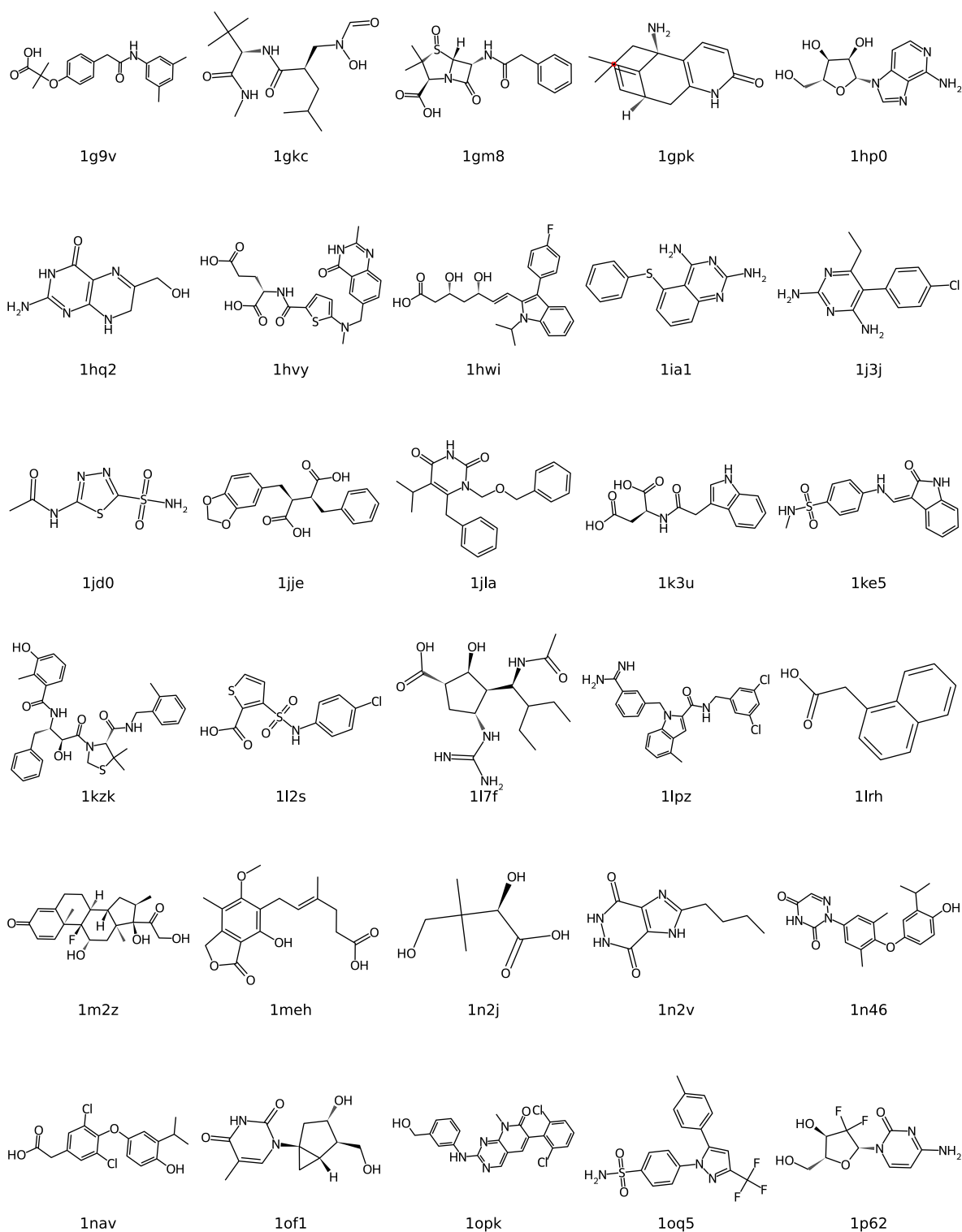
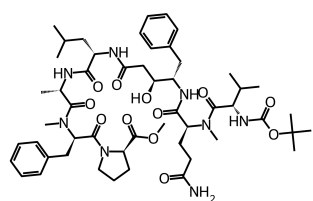
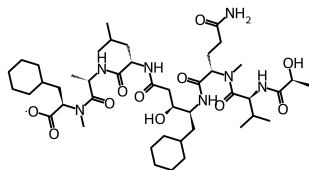


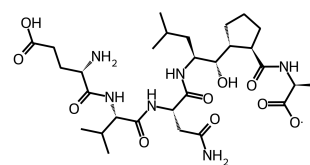
Figure S3: Structural formulas of a subset of molecules from the Astex dataset.²



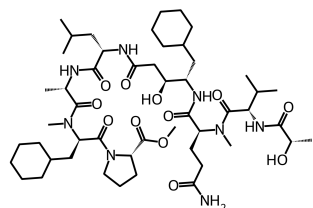
MOL0003



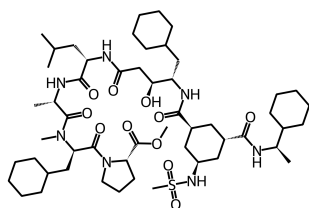
MOL0006



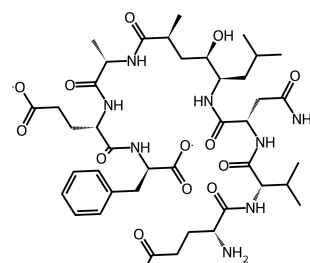
MOL0007



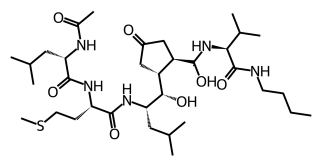
MOL0014



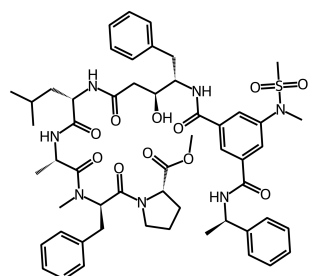
MOL0024



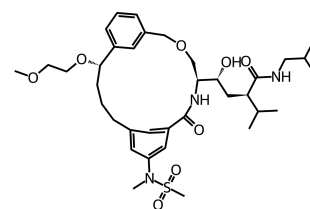
MOL0029



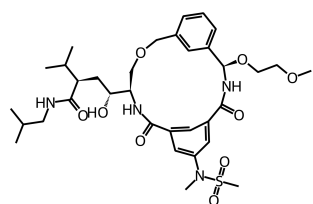
MOL0030



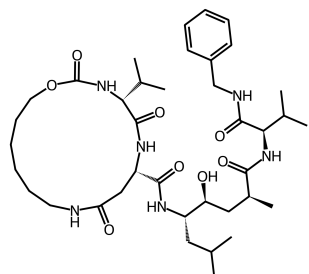
MOL0032



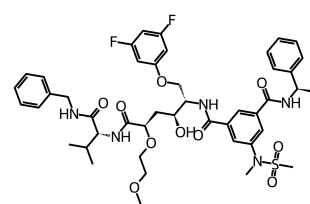
MOL0072



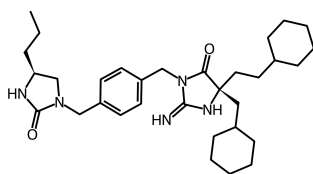
MOL0085



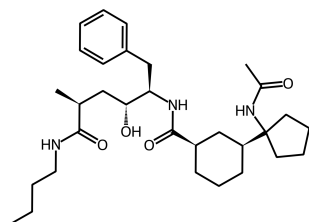
MOL0115



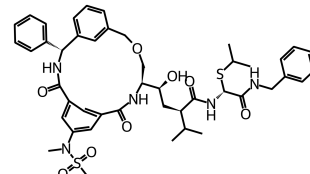
MOL0265



MOL0902



MOL1025



MOL1361

Figure S4: Structural formulas of a subset of large molecules from the BACE dataset.³

3 Additional Benchmark

3.1 Astex set

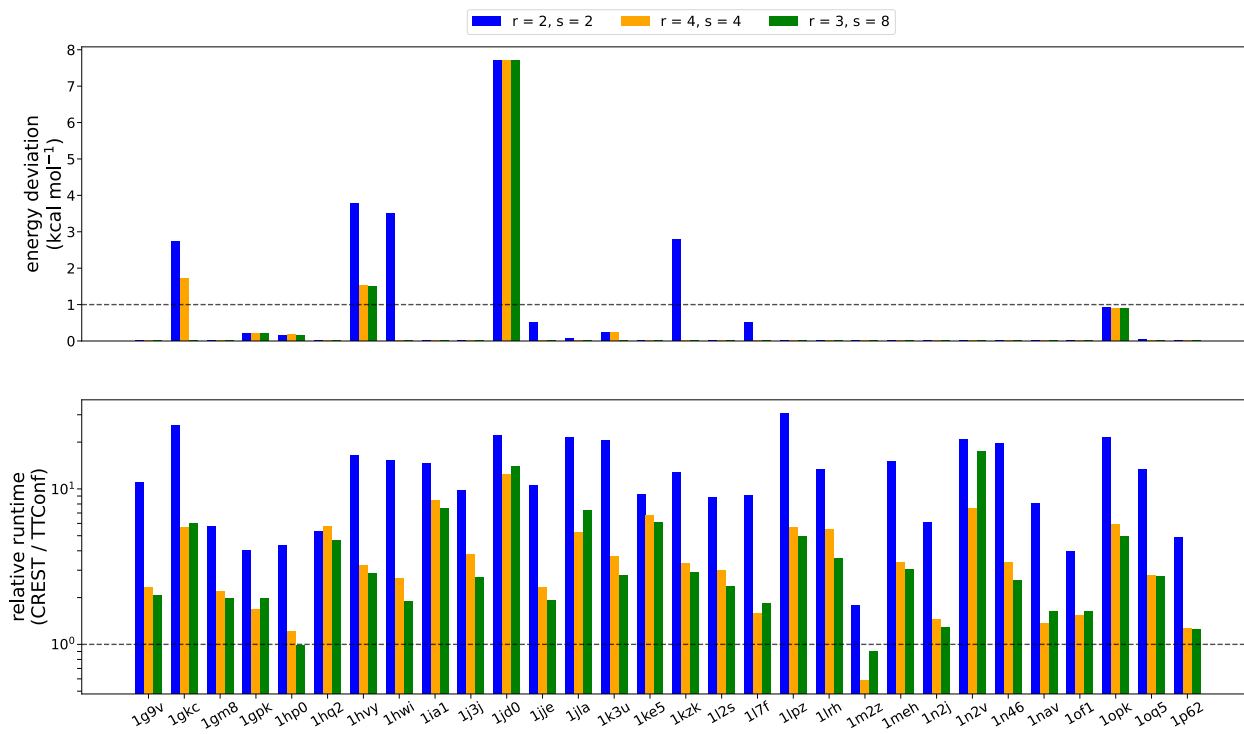


Figure S5: Individual energy deviations and relative runtimes of a subset of the Astex dataset (see Fig. S3) for three different TTConf settings.²

3.2 Optimal Settings

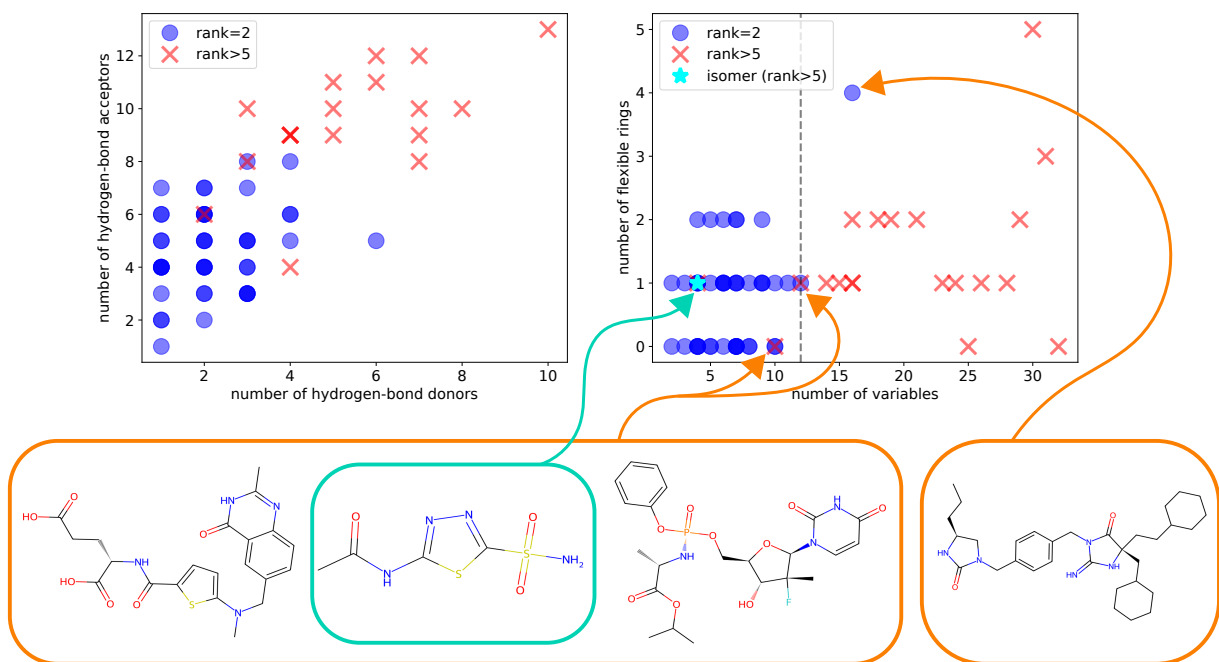


Figure S6: Assessment of the required rank to achieve a maximum ΔE of 1 kcal mol^{-1} . On the left side, the molecules are categorized by their number of hydrogen bond donors and acceptors, while on the right side they are categorized by the number of variables and flexible rings. The set of molecules consists of the CD25 dataset (Fig. S2), a subset of the Astex dataset (Fig. S3), and a subset of large molecules from the BACE dataset (Fig. S4).

3.3 Alanine oligomers

3.3.1 Runtime estimation

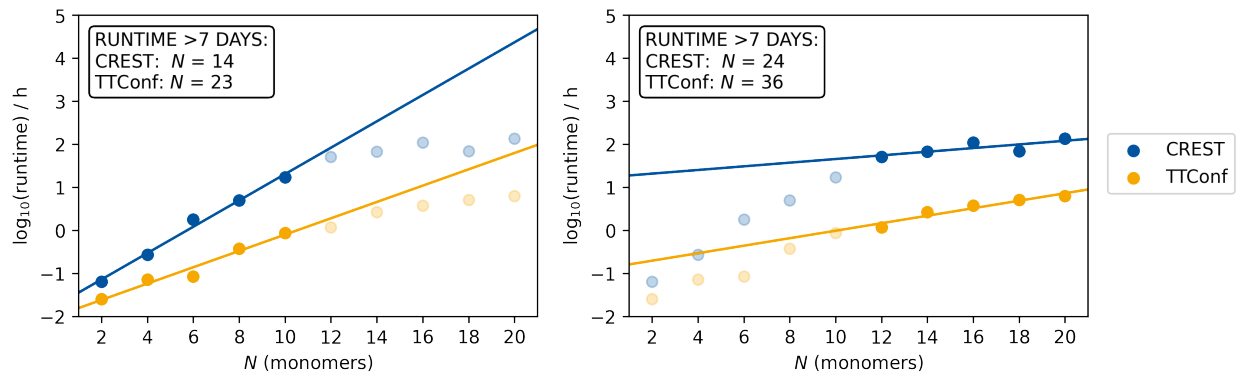


Figure S7: Linear extrapolation of the respective CREST and TTConf runtimes for alanine oligomers. As CREST uses a limited MD/MTD runtime for Ala₁₂ to Ala₂₀ the regression is split into 2-10 (left) and 12-20 monomer units (right).

3.3.2 Combining TTConf and MD

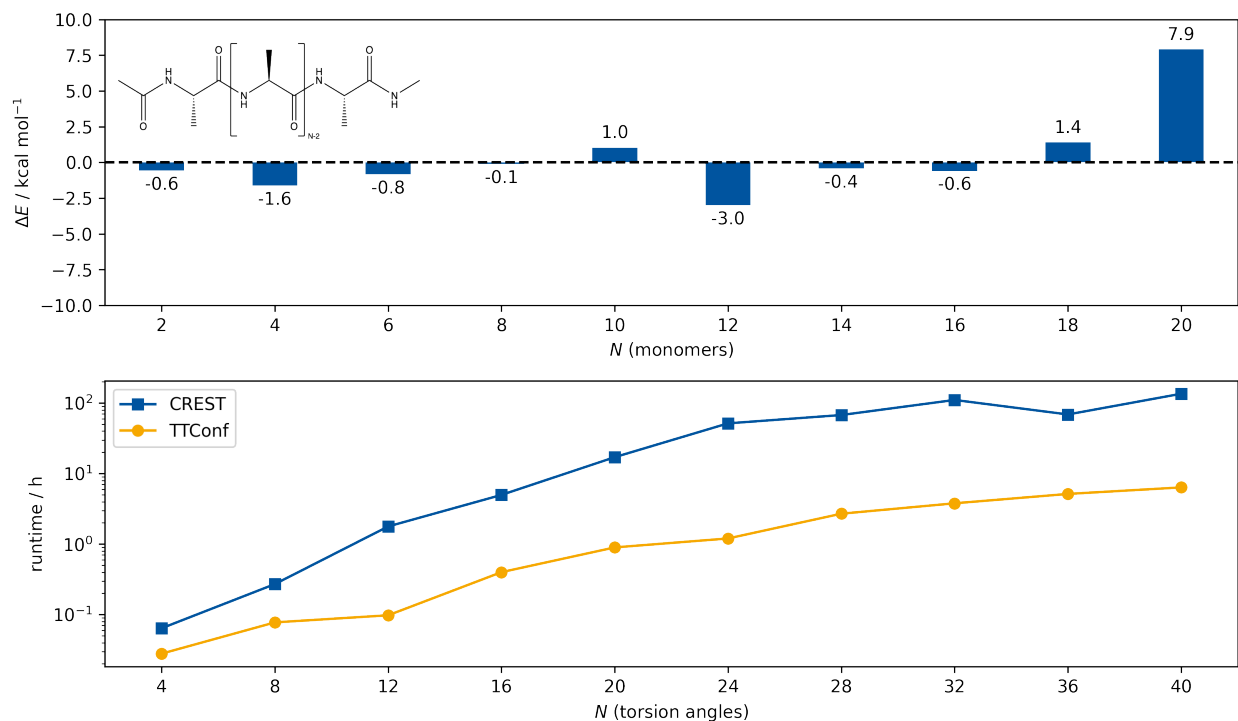


Figure S8: Relative accuracy of TTConf ($r = 3, s = 8$) in terms of energy deviation of the lowest energy conformers in comparison to CREST (top), in addition to the respective runtimes (bottom) for alanine oligomers of 2 to 20 monomer units corresponding to 4 to 40 flexible torsion angles. Here, the regular TTConf runs were extended by short MD simulations of 10 ps at 400 K on the lowest energy conformer in analogy to CREST, which lead to significant improvements for Ala₁₄ and Ala₁₆.

4 Conformational ensemble

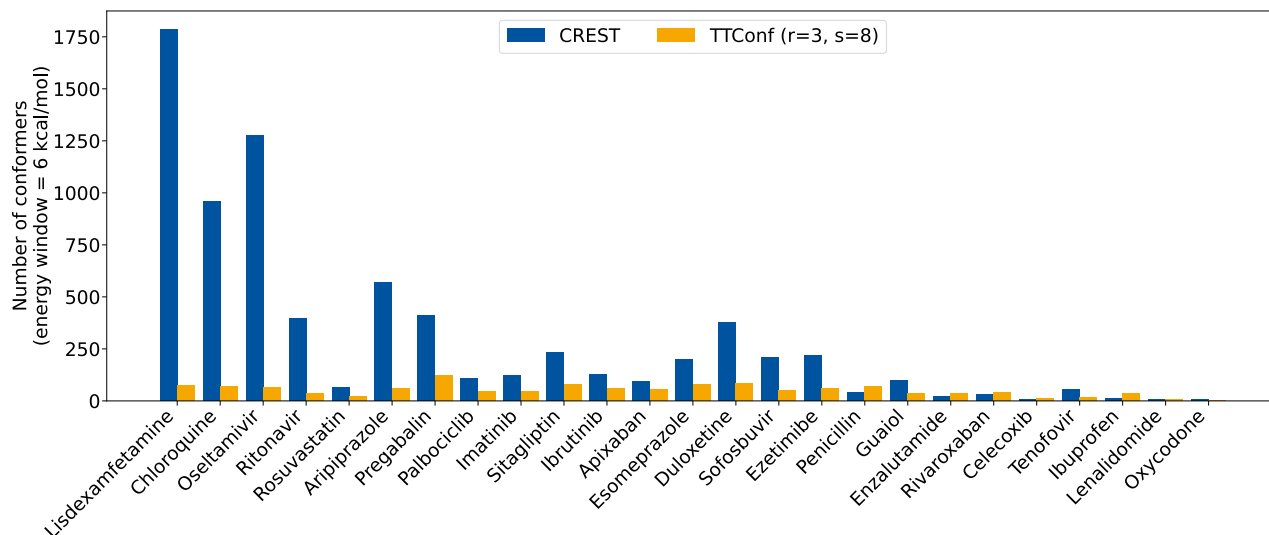


Figure S9: Comparison of the number of conformers found in the conformational ensembles obtained with TtConf ($r = 3, s = 8$) and CREST for the molecules of the CD25 dataset within an energy window of 6 kcal mol^{-1} with respect to the lowest energy conformer found among both methods. The conformational ensembles are filtered using the default parameters of conformer filtering implemented in CREST: RMSD threshold = $0.125 \text{ \AA}$, energy threshold between conformer pairs = 0.05 kcal/mol , rotational constant threshold = 1% .

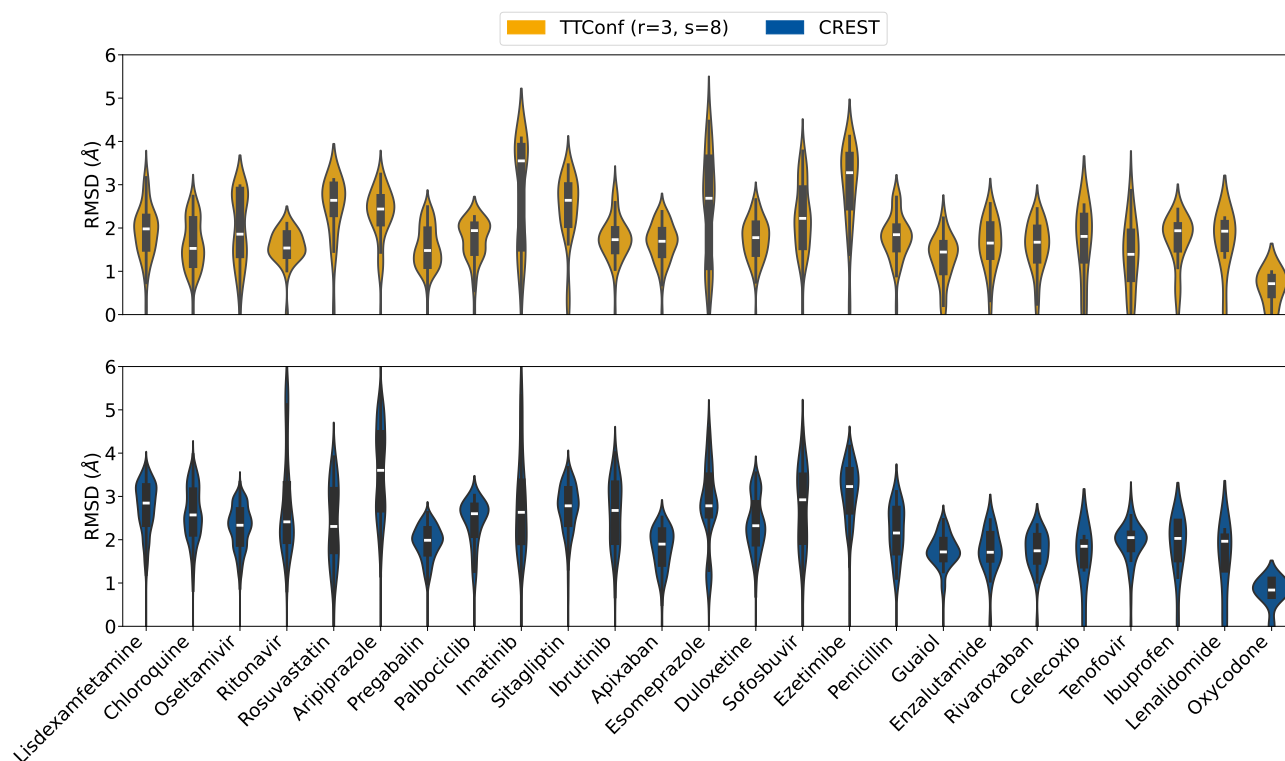


Figure S10: Distribution of conformers in a conformational ensemble by RMSD compared with the minimum energy conformer. Results obtained in TTConf ($r = 3, s = 8$) calculations (top) are compared with calculations obtained in CREST (bottom) for the CD25 dataset at an energy window of 6 kcal mol^{-1} from the lowest energy conformer across both methods, filtered with default CREST parameters ($\text{rthr} = 0.125 \text{ \AA}$, $\text{ethr} = 0.05 \text{ kcal/mol}$, $\text{bthr} = 0.01$). The white line on the violin plot indicates the median.

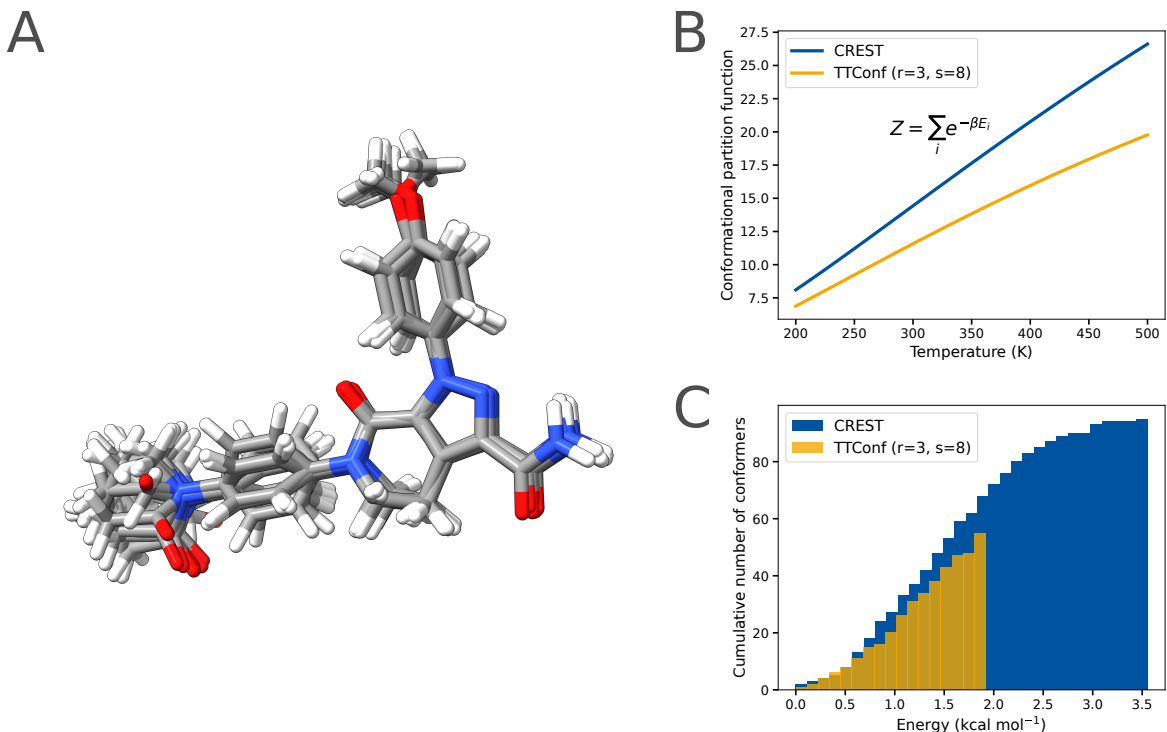
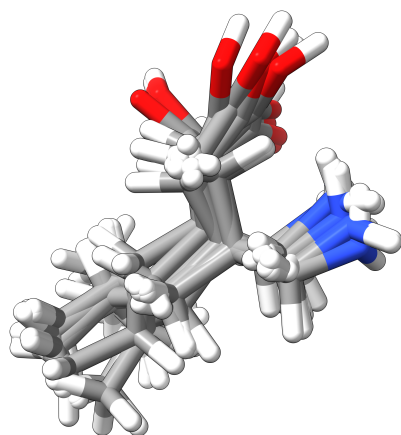
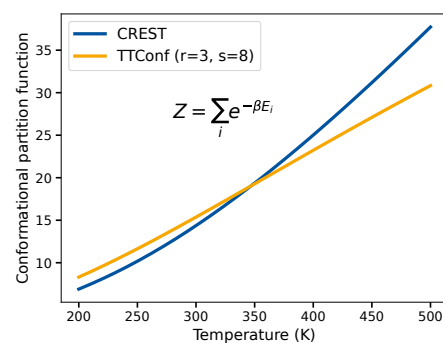


Figure S11: Conformational ensemble of Apixaban (see Fig. S2). Panel A shows the conformational ensemble of the Apixaban molecule with an energy window of 1 kcal/mol and RMSD threshold of 5 Å, obtained with TTConf ($r = 3, s = 8$). Panels B and C compare the conformational ensembles of Apixaban in an energy window of 6 kcal/mol, RMSD threshold of 0.125 Å and rotational constant threshold of 1% (CREST default settings), obtained with CREST and TTConf ($r = 3, s = 8$). Panel B shows conformational partition function as a function of temperature. Panel C shows the cumulative number of conformers in the ensemble as a function of the energy window.

A



B



C

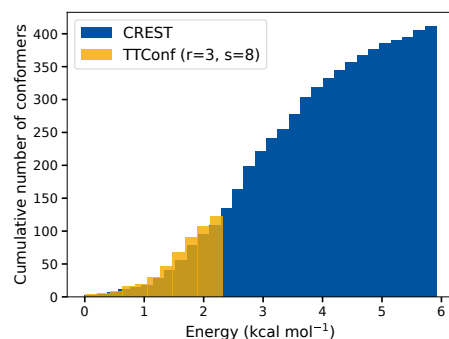


Figure S12: Conformational ensemble of Pregabalin (see Fig. S2). Panel A shows the conformational ensemble the Pregabalin molecule with an energy window of 1 kcal/mol and RMSD threshold of 5 Å, obtained with TTConf ($r = 3, s = 8$). Panels B and C compare the conformational ensembles of Pregabalin in an energy window of 6 kcal/mol, RMSD threshold of 0.125 Å and rotational constant threshold of 1% (CREST default settings), obtained with CREST and TTConf ($r = 3, s = 8$). Panel B shows the conformational partition function as a function of temperature. Panel C shows the cumulative number of conformers in the ensemble as a function of the energy window.

References

- (1) Pracht, P.; Grimme, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chem. Sci.* **2021**, *12*, 6551–6568.
- (2) 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.
- (3) Subramanian, G.; Ramsundar, B.; Pande, V.; Denny, R. A. Computational Modeling of β -Secretase 1 (BACE-1) Inhibitors Using Ligand Based Approaches. *J. Chem. Inf. Model.* **2016**, *56*, 1936–1949.