

# Report III : Torsional Geometric Generation of Molecular 3D Conformer Ensembles (GEOMOL)

Irem Begüm Gündüz - 7026821

## Contents

GEOMOL Approach . . . . .	1
Discussion . . . . .	2
References . . . . .	2

## GEOMOL Approach

Recent breakthroughs in deep learning allow us to predict protein 3-Dimensional structure from the protein linear sequence using specifically designed tools, such as *AlphaFold2* with great accuracy. However, the prediction of 3-D molecule structures with great accuracy remains still as an open problem. Existing techniques for determining molecular conformations are expensive thus, predictive models have been developed over the past few decades. These approaches, typically divided into two categories: stochastic models, which rely on random fluctuations, and systematic models, which are based on rules. Stochastic methods are traditionally based on molecular dynamics (MD) or Markov chain Monte Carlo (MCMC) techniques, possibly combined with genetic algorithms (GA). They allow extensive exploration of the energy landscape and accurately sample equilibrium structures, but they are extremely slow for large molecules. For a single drug-like molecule, they require several CPU minutes. The approaches to resolving protein folding challenges are not transferable to general molecular data, because molecules may contains cycles, different bonds that may yield highly branched graph structures which is hard to deal with machine learning frameworks. Thus, a need for an approach to generate distributions of low-energy molecular 3D conformers remained. *GEOMOL* is an end-to-end, non-autoregressive and SE(3)-invariant machine learning approach that is specifically designed to generate distributions of low-energy molecular 3D conformers. Ganea et al. (2021)

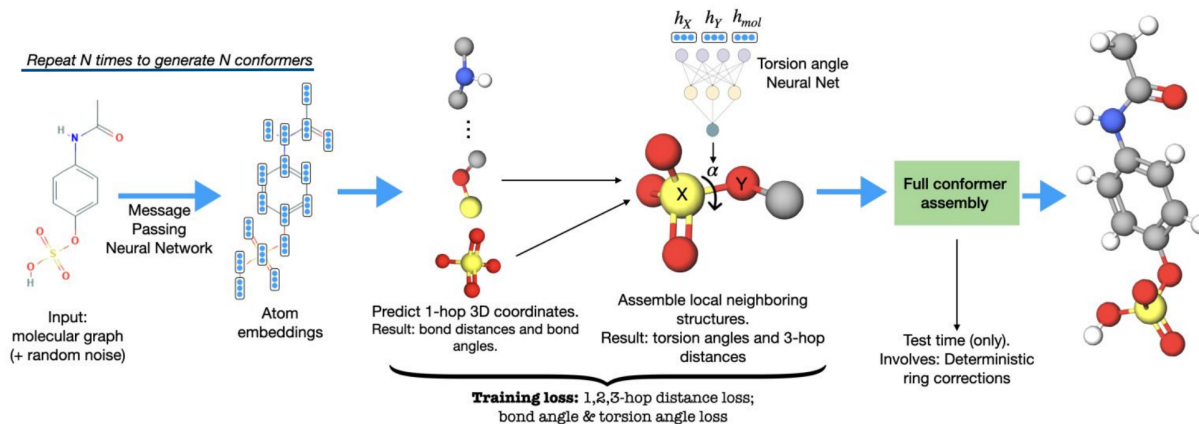


Figure 1: Overview of GEOMOL methodology.

*GEOMOL* tries to answer this question: “Can we design fast ML generative models of high quality, representative, diverse, and generalizable low-energy 3D conformational ensembles from molecular graphs?” as stated by the authors. The model is trainable end-to-end, non-autoregressive, and does not require distance geometry techniques to work correctly. The model outputs a minimal set of geometric quantities. Therefore, we can fully reconstruct the 3D conformation of the molecule by using the correct angles and distances. By design, the model adapts to the shapes of the molecules in a SE(3)-invariant way. The desired inductive bias has either been completely or partially captured in the past utilizing multi-step distance geometry approaches.

*Step 1:* Each molecular graph in addition to random noise inputted to message passing neural network. Message passing framework of *GEOMOL* computes molecular embedding (ie. atom embedding of an input molecule).

*Step 2:* *GEOMOL* predict the 3-dimensional position of of all its 1-hop neighbouring nodes for each atom in a permutation equivariant way.

## Discussion

The fundamental components of molecular geometry, such as torsion angles and local 3D structures, are explicitly represented in this model (bond distances and angles adjacent to each atom). These are combined with the input molecular network to compute the k-hop distance during training and the entire deterministic conformation assembly during testing. It is crucial to note that we compute a single twist angle for each non-terminal bond, regardless of the quantity and configuration of surrounding atoms at each of the associated bond’s endpoints, in order to avoid over-parameterizing these predictions.

## References

Ganea, Octavian Eugen, Lagnajit Pattanaik, Connor W. Coley, Regina Barzilay, Klavs F. Jensen, William H. Green, and Tommi S. Jaakkola. 2021. “GeoMol: Torsional Geometric Generation of Molecular 3D Conformer Ensembles.” *Advances in Neural Information Processing Systems* 17 (June): 13757–69. <https://doi.org/10.48550/arxiv.2106.07802>.