Beyond Validity and Novelty: Scaffold-based Metrics for Evaluating Chemical Structure Generators

Valeriia Fil [1], Svozil Daniel [1,2]

[1] *Department of Informatics and Chemistry & CZ-OPENSCREEN: National Infrastructure for Chemical Biology, Faculty of Chemical Technology, University of Chemistry and Technology, Technická 5, 16628, Prague, Czech Republic.*
[2] *CZ-OPENSCREEN: National Infrastructure for Chemical Biology, Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic.*

Introduction

The exploration of chemical space is crucial for creating new virtual compounds, which are vital for the next generation of drug-like molecules. To achieve this goal, researchers have developed various molecular generation tools. However, evaluating their effectiveness presents a complex challenge. Current evaluation methods primarily focus on technical aspects, verifying that a generated molecule is valid and unique. However, they often neglect the potential biological activity, the ultimate goal in drug discovery. **This work proposes a novel approach to benchmarking chemical structure generators. We focus on scaffolds, the core structure of a molecule, to identify biologically active patterns that the generator can discover, even if they weren't part of the training data. This allows us to assess a generator's ability to generate new active chemotypes.**

Metrics

1. True positive recall all (TUPOR) - measures recall rate of unique scaffolds.

$$TUPOR = \frac{\mathrm{NAS}}{\mathrm{UAS}}$$

2. Set Scaffold yield (SESY) - assesses diversity of output setset

$$SESY = \frac{\mathrm{NS}}{\mathrm{SS}}$$

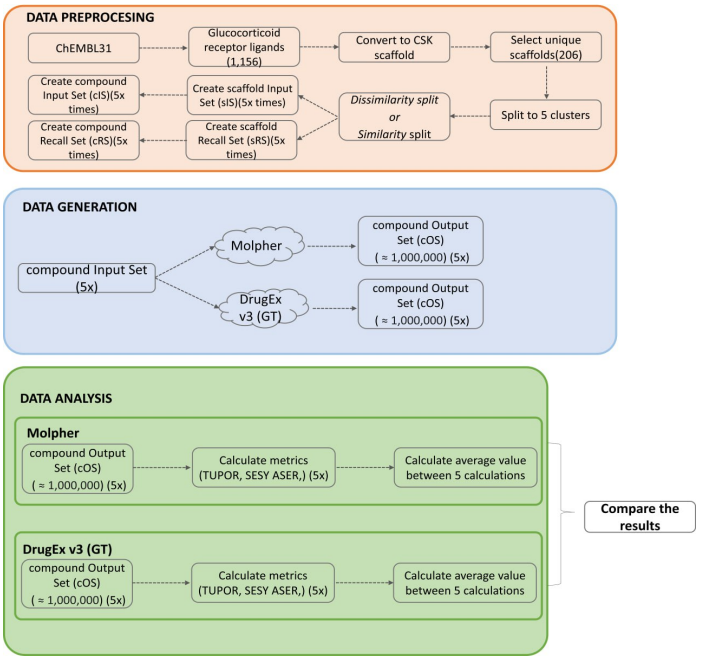
3. Absolute set scaffold recall (ASER) - evaluates generator's effectiveness

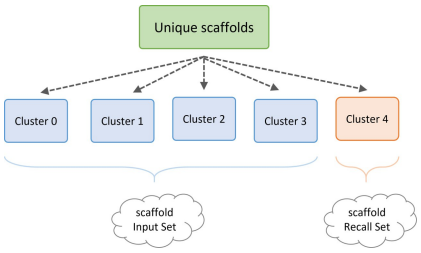
$$ASER = \frac{\mathrm{tRS}}{\mathrm{SS}}$$

Split approach

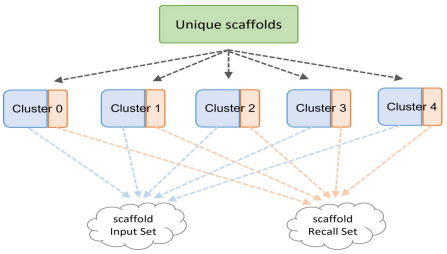
Using two different approaches for split data for Input Sets(IS) and Recall Sets(RS) to test the bottom and top limits of our metrics. This helps evaluate how generators perform with varying distances between the IS and RS.
Dissimilarity Split: sRS was assembled from Cyclic Skeleton Scaffold (CSK) scaffolds that differed from the CSK scaffolds in the sIS.
Similarity Split: sRS was assembled from scaffolds similar to those in sIS.

Project





Dissimilarity split



Similarity split

Results

Name	Splitting_type	Scaffold	Set_size	TUPOR	SESY	ASER
Molpher_mean	dis	csk	1 035 435	0,5136	0,1242	0,0046
DrugEx_mean	dis	csk	999 912	0,5555	0,3221	0,0127

Our metrics indicate that DrugEx outperforms Molpher in several key areas:
- **TUPOR (True Positive Recall All):** DrugEx has a higher TUPOR value, indicating it generates more unique active scaffolds compared to Molpher.
- **SESY (Set Scaffold Yield):** DrugEx also shows a better SESY value, suggesting it produces a more diverse output set.
- **ASER (Absolute Set Scaffold Recall):** DrugEx has a superior ASER value, meaning it generates more compounds containing active scaffolds. This indicates DrugEx is more effective in exploring the desired chemical space, rich in active chemotypes.
Based on these results, we recommend using DrugEx over Molpher for generating active compounds targeting the Glucocorticoid Receptor.

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

CZ OPENSCREEN