P37

Breakthroughs that change patients' lives

Can Generative Models Learn Privileged Substructures?

Vishnu Sresht[1], Stephen Ra[1], Liying Zhang[2]

[1]Simulation & Modeling Sciences
[2]Internal Medicine Research Unit, Medicinal Design

Pfizer

Can Generative Models Learn to Transfer SAR?

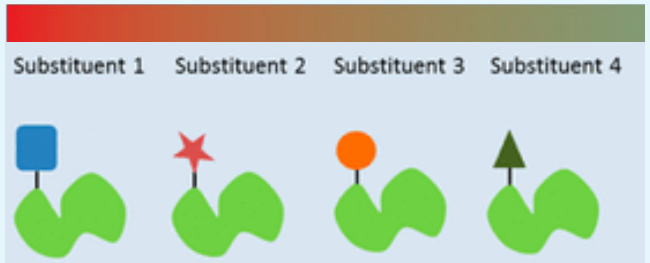
Can ‘AI’ identify the types of patterns that human medicinal chemists look for?

SAR from a previous project

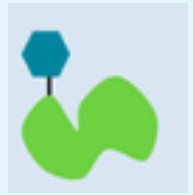
potency →

SAR of my current project

potency →



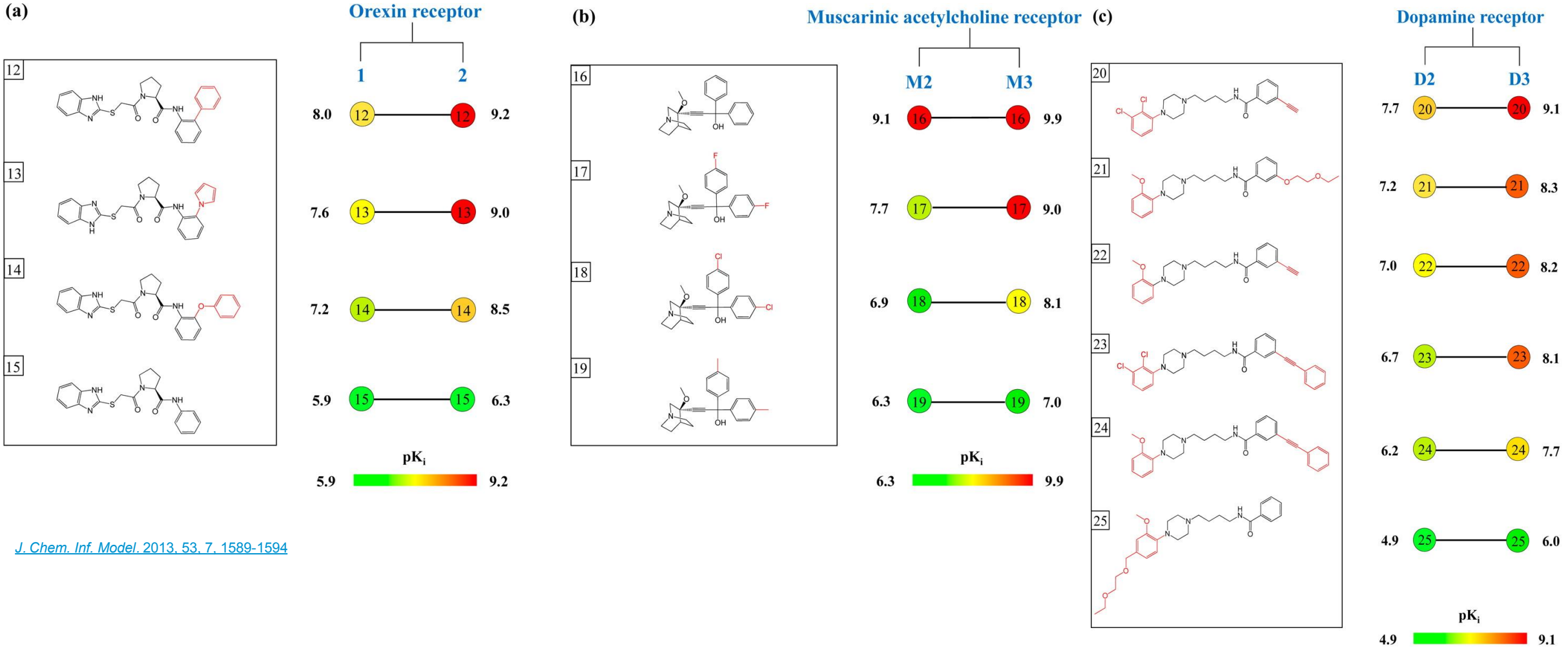
I should make this next



J. Chem. Inf. Model. 2017, 57, 5, 1187-1196

Experienced medicinal chemists develop an intuition for which substituents to try next based on the structure-activity relationships (SAR) they have seen on previous projects.

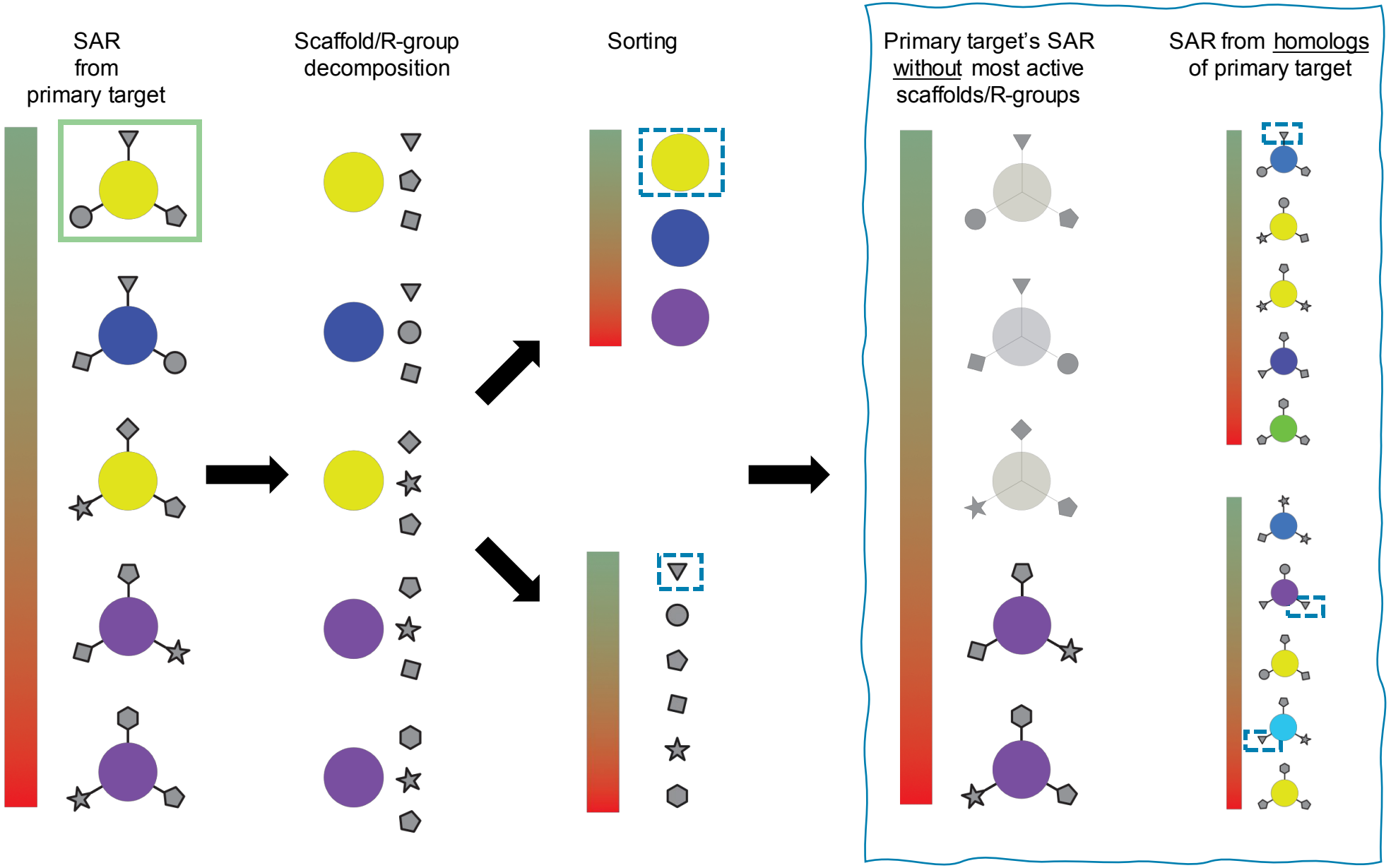
Can Generative Models Learn to Transfer SAR?



J. Chem. Inf. Model. 2013, 53, 7, 1589-1594

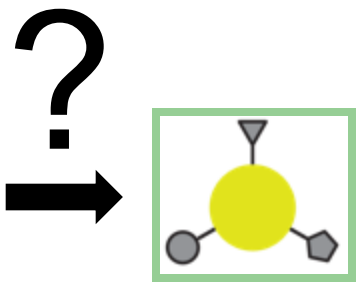
This motivates automated methods of SAR transfer such as “Matched Molecular Series”.

The Experiment



Training Data

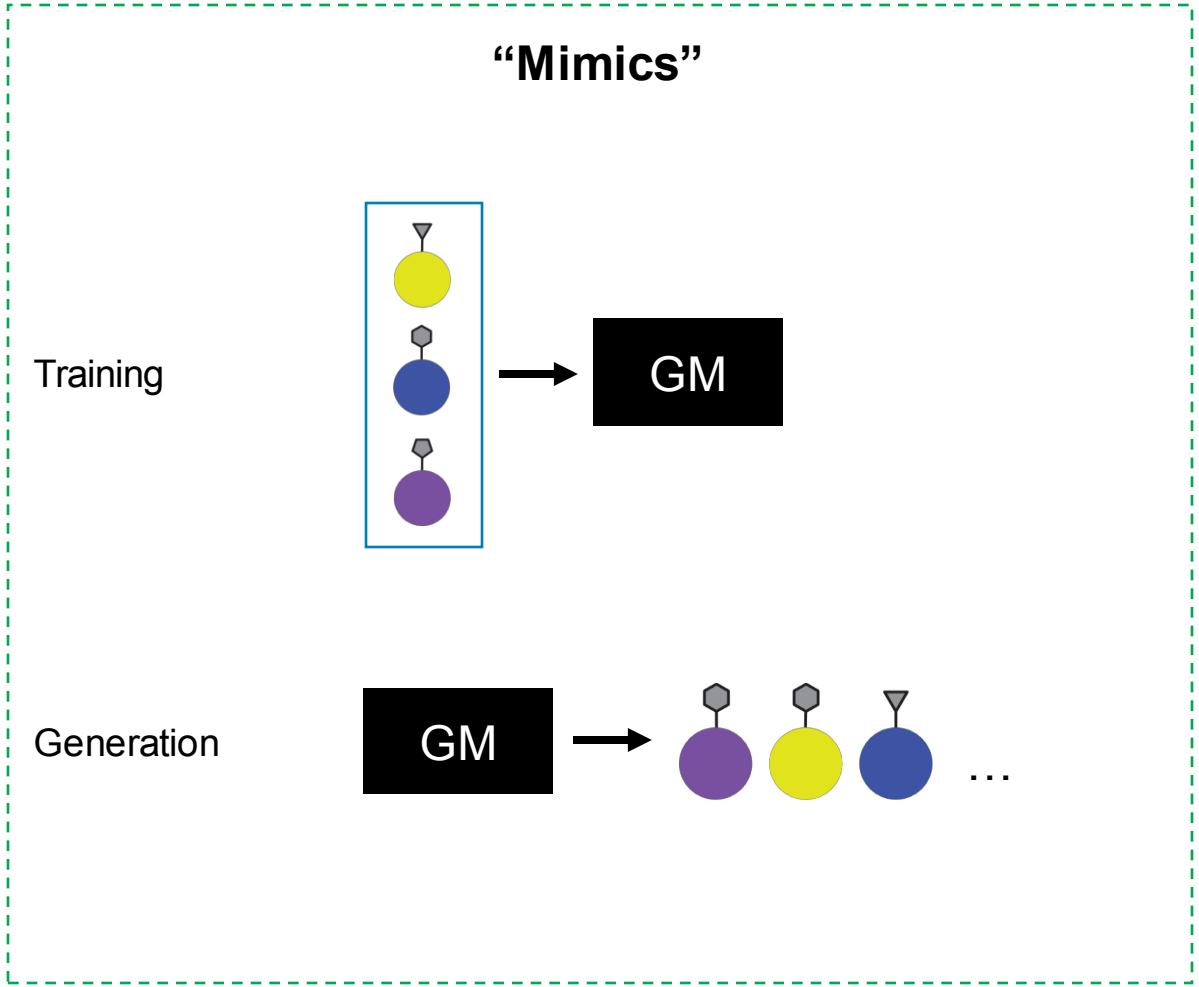
*Can a generative model trained on this **data** recover potent combinations of scaffolds and R-groups from the original SAR of the primary target?*

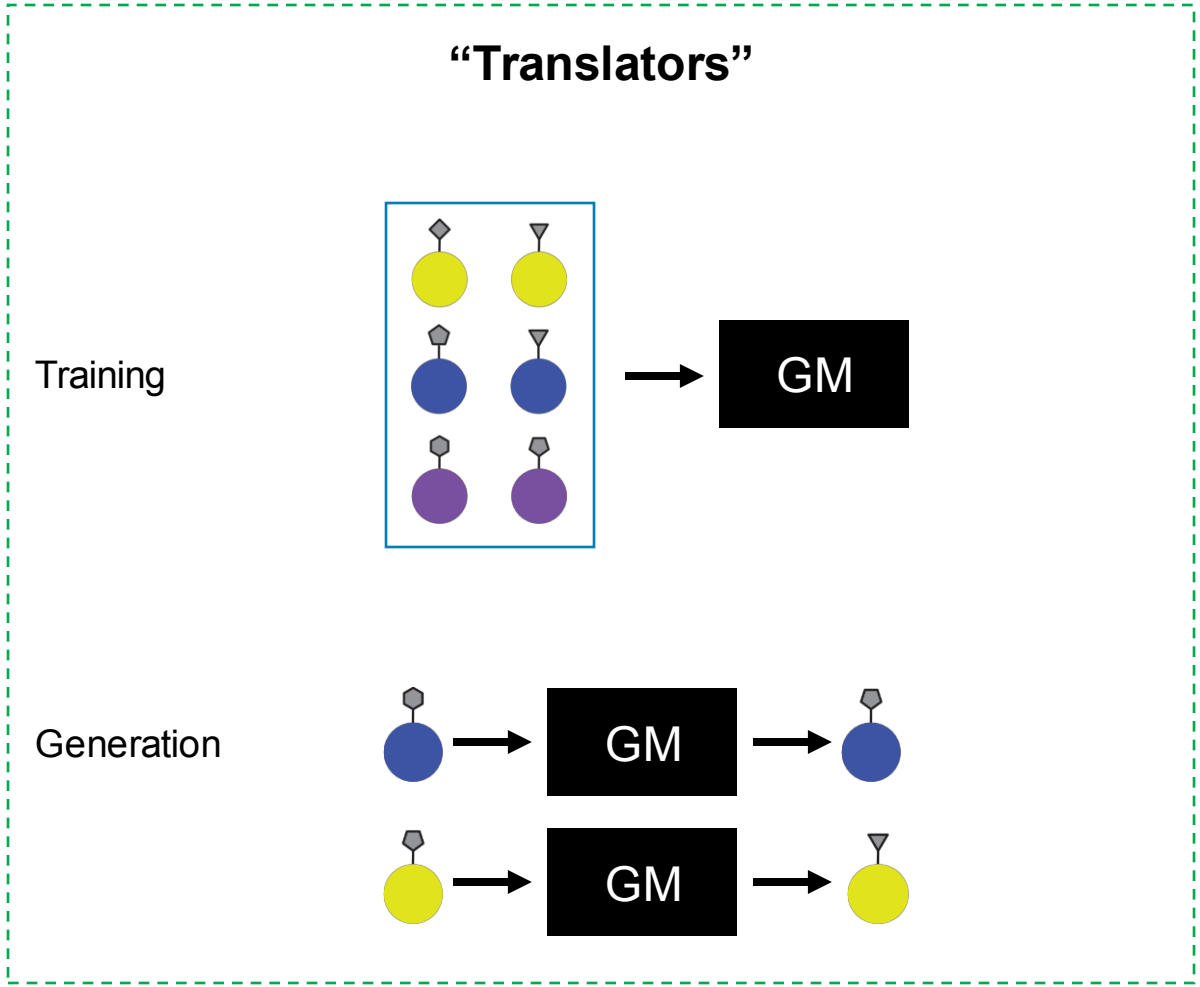


The Generative Models

We tested 6 Generative Models in this work

We looked at two generative model paradigms: “mimics” and “translators”





We tested 6 Generative Models in this work

3 models use SMILES/RNNs, 1 model uses SELFIES/transformers, 2 models use graphs

"Mimics"

B

Bidirectional Molecule Generation with Recurrent Neural Networks

Francesca Grisoni,* Michael Moret, Robin Lingwood, and Gisbert Schneider*

Cite This: *J. Chem. Inf. Model.* 2020, 60, 1175–1183 | Read Online

doi:10.1021/acs.jcim.9b00943

C

Cite This: *ACS Cent. Sci.* 2018, 4, 120–131

Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks

Marwin H. S. Segler,*[,†] Thierry Kogej,[‡] Christian Tyrchan,[§] and Mark P. Waller*[,∥]

doi:10.1021/acscentsci.7b00512

R

Research article | Open Access | Published: 21 November 2019

Randomized SMILES strings improve the quality of molecular generative models

Josep Arús-Pous, Simon Viet Johansson, Oleksii Prykhodko, Esben Jannik Bjerrum, Christian Tyrchan, Jean-Louis Reymond, Hongming Chen & Ola Engkvist

doi:10.1186/s13321-019-0393-0

Journal of Cheminformatics **11**, Article number: 71 (2019) | Cite this article

"Translators"

S2S

arXiv.org > cs > arXiv:1912.10156

Computer Science > Machine Learning

[Submitted on 21 Dec 2019]

Black Box Recursive Translations for Molecular Optimization

Farhan Damani, Vishnu Sresht, Stephen Ra

arXiv:1912.10156

G2G

arXiv.org > cs > arXiv:1812.01070

Computer Science > Machine Learning

[Submitted on 3 Dec 2018 (v1), last revised 28 Jan 2019 (this version, v3)]

Learning Multimodal Graph-to-Graph Translation for Molecular Optimization

Wengong Jin, Kevin Yang, Regina Barzilay, Tommi Jaakkola

arXiv:1812.01070

HG2G

arXiv.org > physics > arXiv:1907.11223

Physics > Chemical Physics

[Submitted on 11 Jun 2019 (v1), last revised 18 Oct 2019 (this version, v2)]

Hierarchical Graph-to-Graph Translation for Molecules

Wengong Jin, Regina Barzilay, Tommi Jaakkola

arXiv:1907.11223

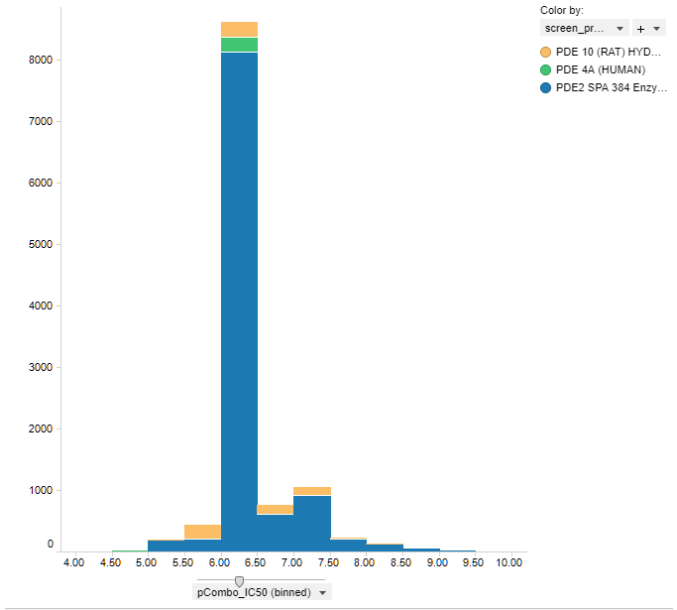
The Training Data

Targets and Selected R-groups

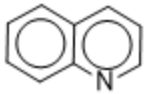
We ran this experiment on 2 datasets

- **Target 1: PDE10A**

  PDE10A: 830 cmpds
  PDE2A: 5403 cmpds
  PDE4A: 274 cmpds



Example Privileged substructure



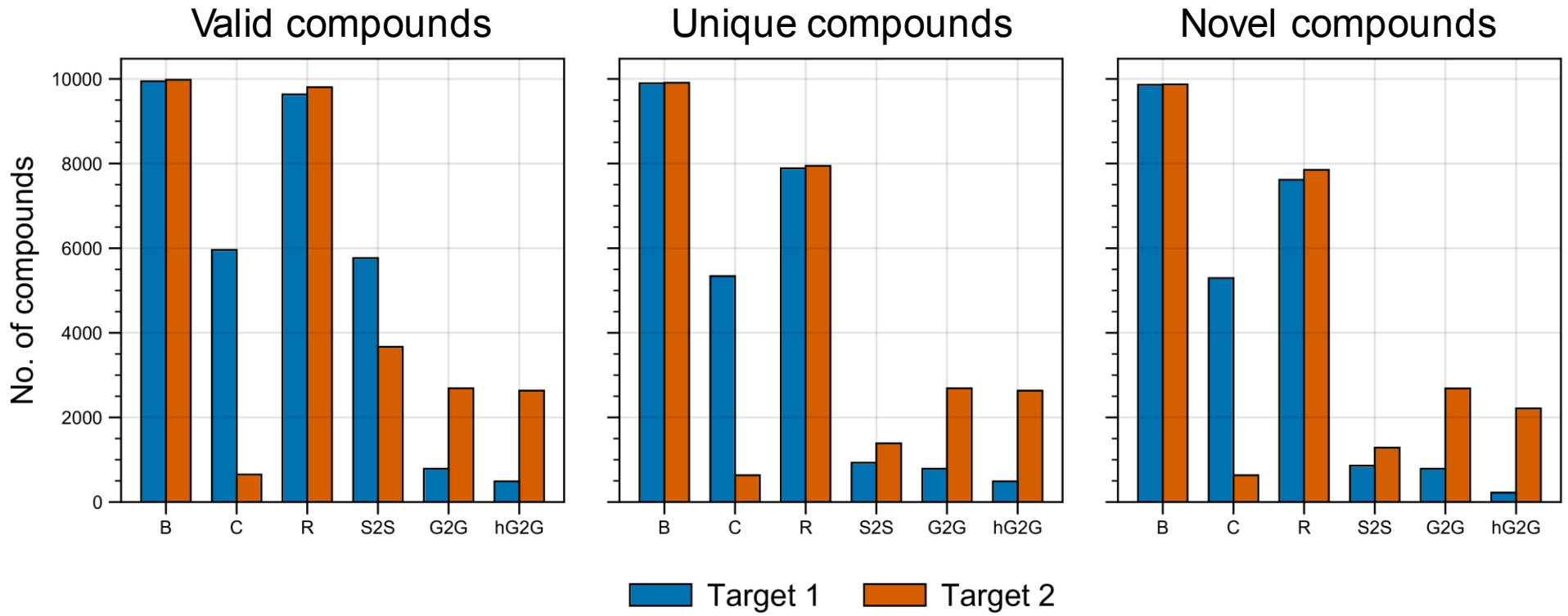
Example Substructure filter (mask)

- **Target 2: Proprietary Target**
  Primary target: 3028 cmpds
  Analog targets: 4260 cmpds

The Results

Validity, Uniqueness, and Novelty

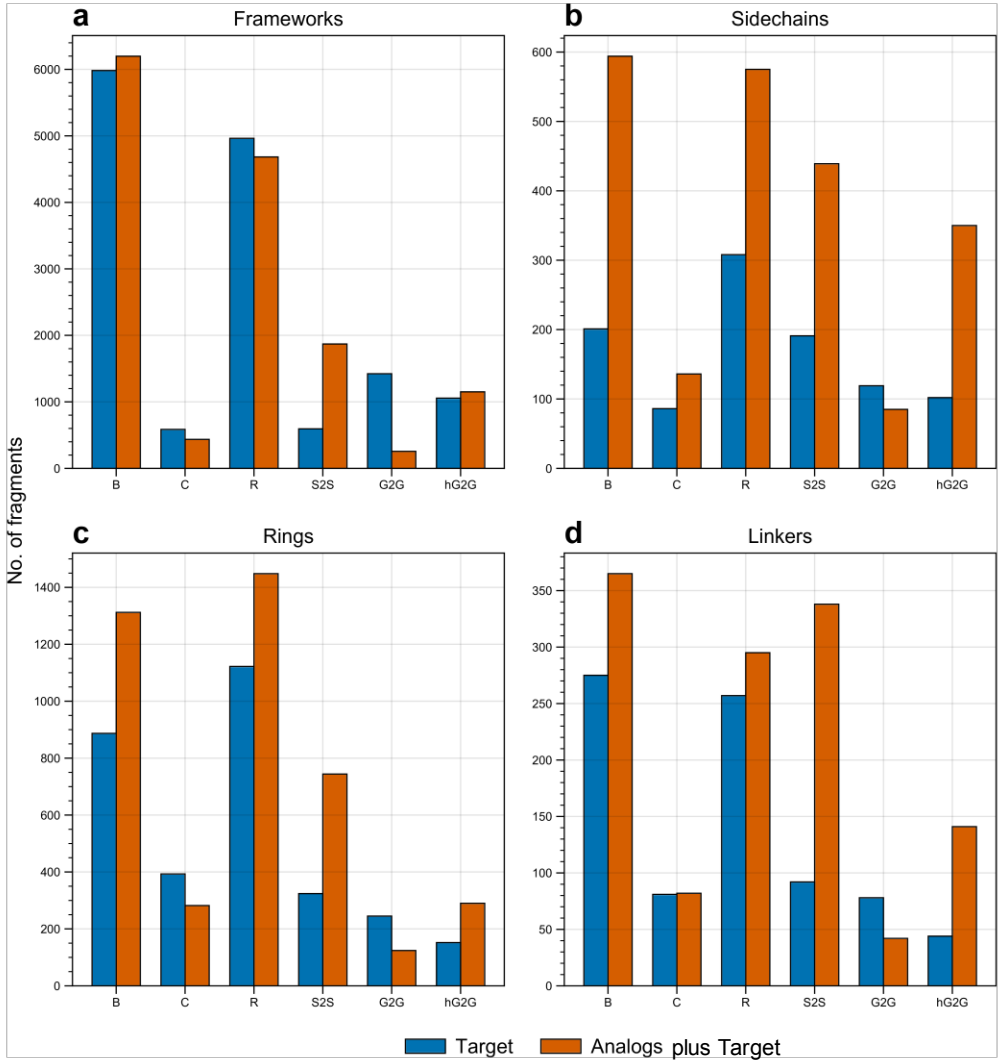
B generates the most novel compounds. All methods have high uniqueness and novelty ratios.



These numbers are gathered from 10,000 generation attempts.

Diversity within the Generated Compounds

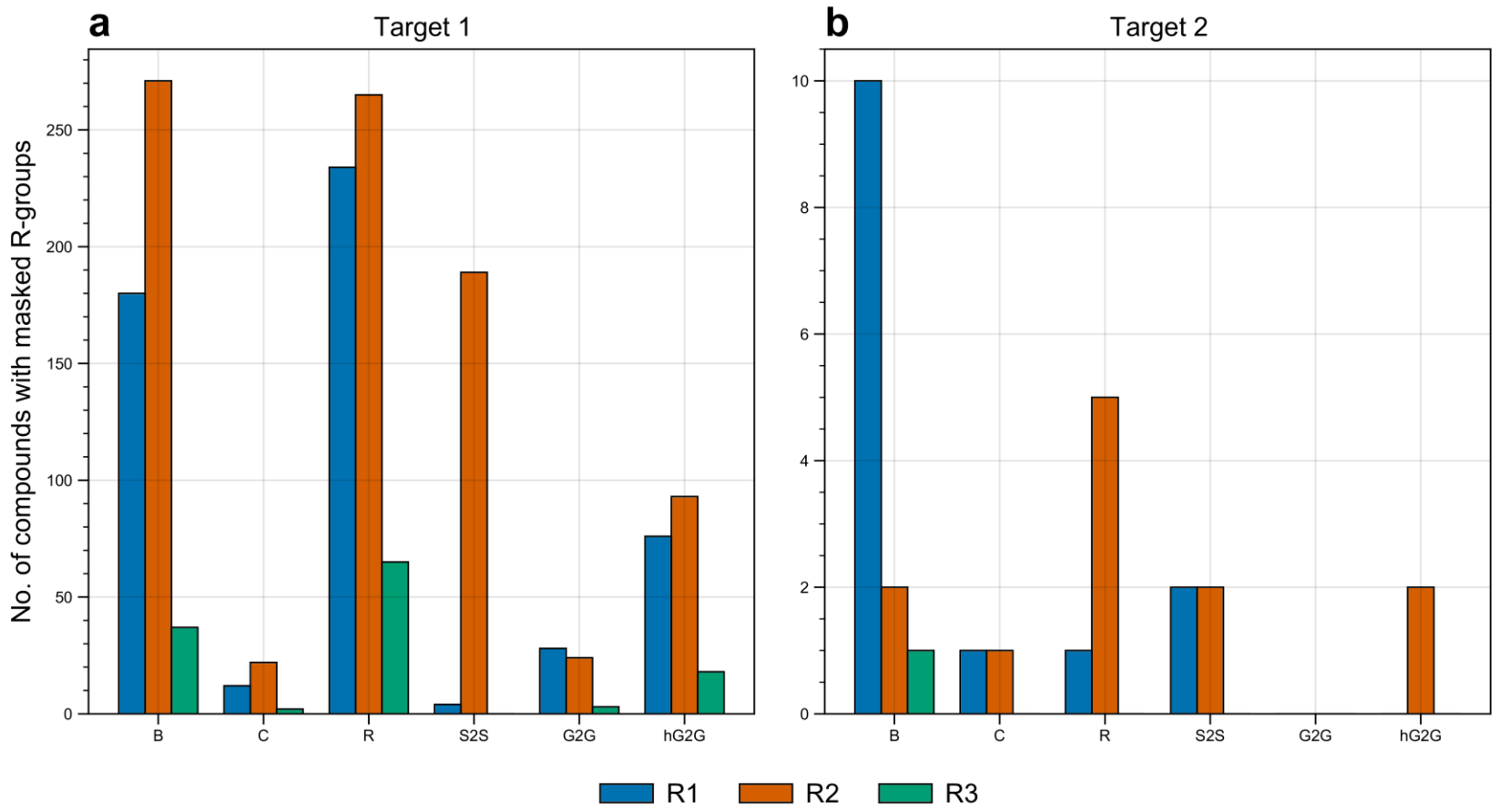
Most generative models can extract additional diversity when trained on Target + Analogs



Breakthroughs that change patients' lives

Substructure Recovery

B, R, and S2S methods do a much better job of recovery than graph-based methods



Conclusions

- We compared 6 different generative models on the task of recognizing privileged structures
  - 3 ‘Mimic’-type models used SMILES
  - 3 ‘Translator’-type models used either SELFIES or Graphs

- BIMODAL, Randomized REINVENT and SELFIES Seq2Seq do much better at recovering masked, potent substructures than graph-based methods

- When trained on SAR from Target + Analogs, these 3 methods also generate more diverse scaffolds, rings, sidechains and linkers.