AMCG: A Graph Dual Atomic-Molecular Conditional Molecular Generator

Carlo Abate[1,2], Sergio Decherchi[3], Andrea Cavalli[2,4]

[1]Department of Pharmacy and Biotechnology (FaBiT), Alma Mater Studiorum, University of Bologna, Italy
[2]Computational & Chemical Biology, Fondazione Istituto Italiano di Tecnologia, Genoa, Italy
[3]Data Science and Computation Facility, Fondazione Istituto Italiano di Tecnologia, Genoa, Italy
[4]Centre Européen de Calcul Atomique et Moléculaire (CECAM), Ecole Polytechnique Fédérale de Lausanne, Switzerland

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ISTITUTO ITALIANO
DI TECNOLOGIA

INTRODUCTION

De novo drug design:

- Crucial for pharmaceutical innovation
- Challenging and expensive endeavor
- High impact on drug discovery process

We present AMCG, a GNN-based VAE-like framework for conditional *de novo* drug design.

METHODS

Key Components:

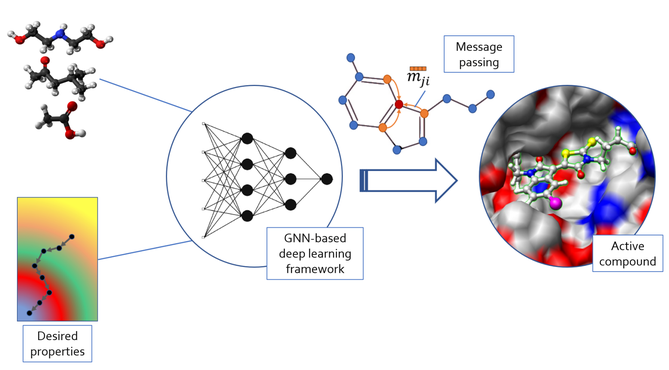
- Encoder: Processes input graphs
- Combiner: Merges atomic and molecular representations
- Molecular Decoder: Generates surrogate atom-wise representation
- Shared Decoder: Reconstructs molecular structure

Flexible Latent Space Sampling:

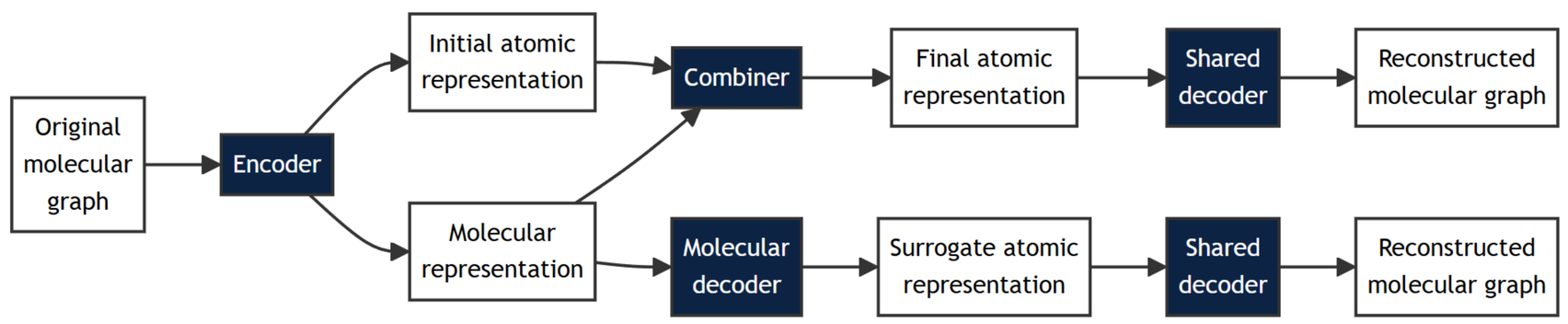
- Gaussian Mixture Model (GMM) priors for targeted exploration
- Enables conservative and explorative molecular generation

Conditional generation:

- Property optimization via gradient ascent in latent space
- Explicit control over the atom types histogram



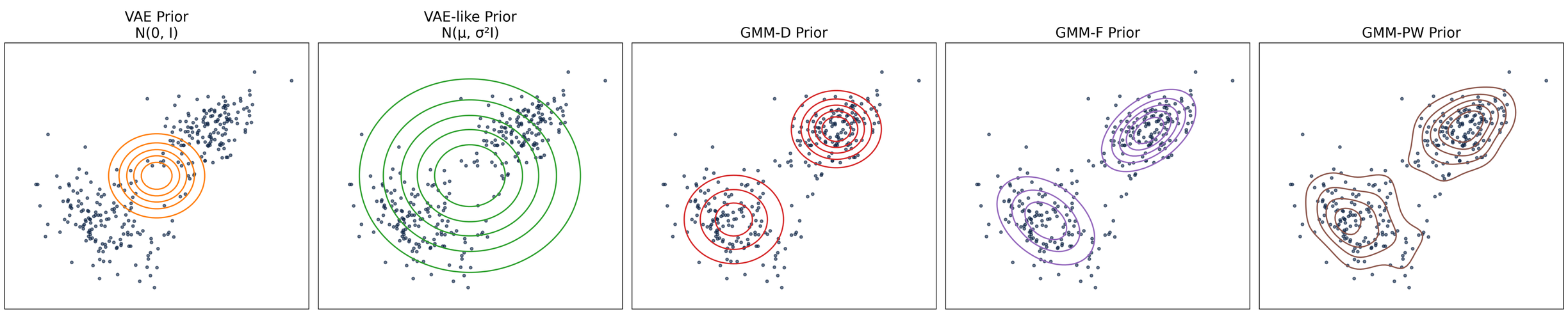
Conditional generation of molecular graphs via GNNs



Schematic representation of the AMCG architecture

GMM PRIORS OVERVIEW

AMCG employs various Gaussian Mixture Model priors for flexible latent space sampling.



Comparison of different priors in AMCG

VUN ASSESSMENT

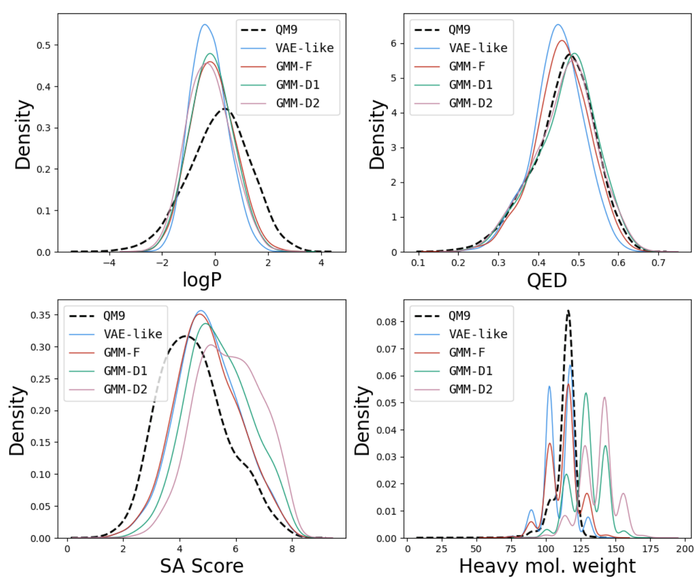
Model	Validity	Validity w/o check	Uniqueness	Novelty	VUN
MPG-VAE	-	0.9100	0.6800	0.540	0.3340
GraphNVP	-	0.8310	0.9920	0.582	0.4797
GRF	-	0.8450	0.6600	0.586	0.3268
GraphAF	1.000	0.6700	0.9451	0.8883	0.8395
MoFlow	1.000	0.8896	0.9853	0.9604	0.9462
GraphDF	1.000	0.8267	0.9762	0.9810	0.9576
Ours - VAE	1.000	0.4006	0.1293	0.8987	0.1162
Ours - VAE-like	1.000	0.5803	0.7756	0.8829	0.6848
Ours - GMM-F	1.000	0.4075	0.9428	0.8001	0.7543
Ours - GMM-D1	1.000	0.1653	0.9693	0.9640	0.9344
Ours - GMM-D2	1.000	0.0555	0.9982	0.9964	0.9946

Comparison of AMCG models with competing methodologies on QM9 dataset

Model	Validity	Validity w/o check	Uniqueness	Novelty	VUN
GraphNVP	-	0.426	0.948	1.000	0.4038
GRF	-	0.734	0.537	1.000	0.3942
GraphAF	1.000	0.68	0.991	1.000	0.9910
MoFlow	1.000	0.5030	0.9999	1.000	0.9999
GraphDF	1.000	0.8903	0.9916	1.000	0.9916
Ours - VAE	1.000	0.2323	0.0437	0.8902	0.0389
Ours - VAE-like	1.000	0.0262	0.7054	1.000	0.7054
Ours - GMM-D	1.000	0.0144	0.9900	1.000	0.9900
Ours - GMM-PW	1.000	0.2630	0.9190	0.7636	0.7017

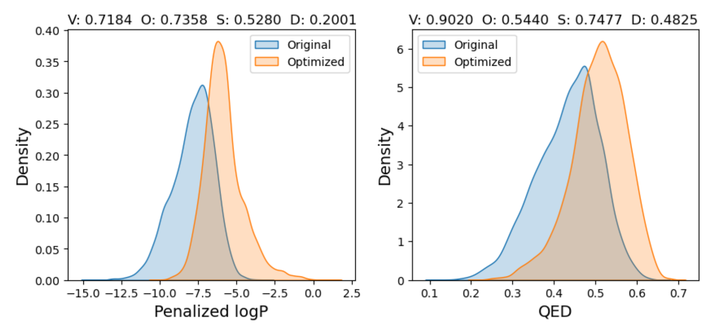
Comparison of AMCG models with competing methodologies on ZINC dataset

LEARNT PROPERTY DISTRIBUTIONS

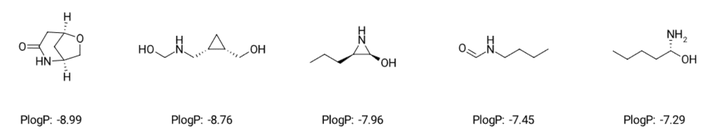


Molecular property distributions on QM9 dataset

PROPERTY OPTIMIZATION

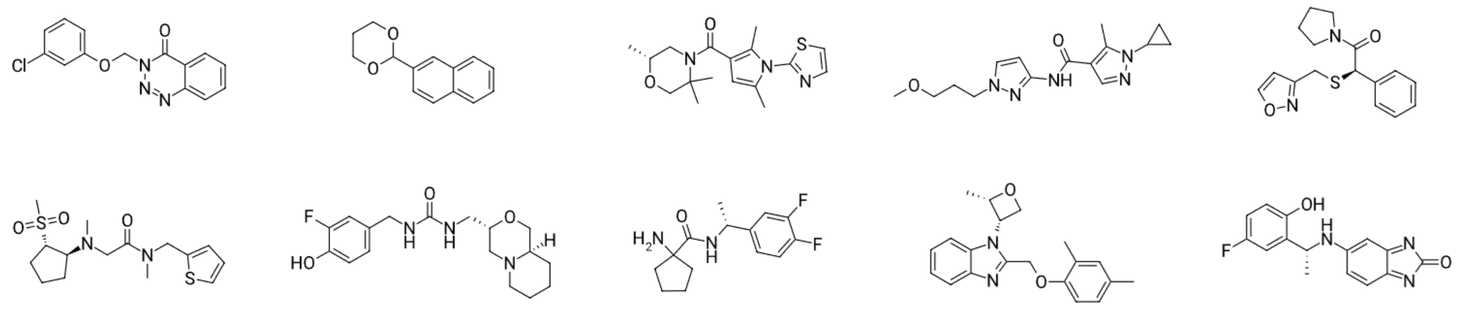


Property distribution shift



Optimization path

GENERATED MOLECULES



Examples of molecules generated from ZINC dataset

CONCLUSIONS AND FUTURE RESEARCH

We introduced AMCG, a conditional molecular graph generator performing competitively or better than state-of-the-art latent variable models on QM9 and ZINC datasets.

Future works:

- Improving the technical components of AMCG framework
- Integrating additional molecular properties

Expected outcomes:

- Better stability in model training
- Expanded capabilities in molecular design and optimization

CONTACT INFORMATION

- GitHub: `carloabate`
- Email: `carlo.abate@iit.it`

FULL PAPER AVAILABLE HERE

Abate, C., Decherchi, S., & Cavalli, A. (2024). AMCG: a graph dual atomic-molecular conditional molecular generator. *Machine Learning: Science and Technology*, **5**(3), 035004.