Hybrid Generative AI for De Novo Design of Co-Crystals with Enhanced Tabletability

Nina Gubina,[1] Andrei Dmitrenko,[1,2] Gleb Solovev,[1] Lyubov Yamshchikova,[1] Oleg Petrov,[1] Ivan Lebedev,[3] Grigory Kirgizov,[1] Nikita Serov,[1] Nikolay Nikitin,[1] Vladimir Vinogradov [1]

[1] ITMO University, St. Petersburg, Russia
[3] D ONE AG, Zurich, Switzerland
[3] Ivanovo State University of Chemistry and Technology, Ivanovo, Russia

Introduction

There exists no open platform for fast *in silico* screening of co-crystals with target tabletability profiles

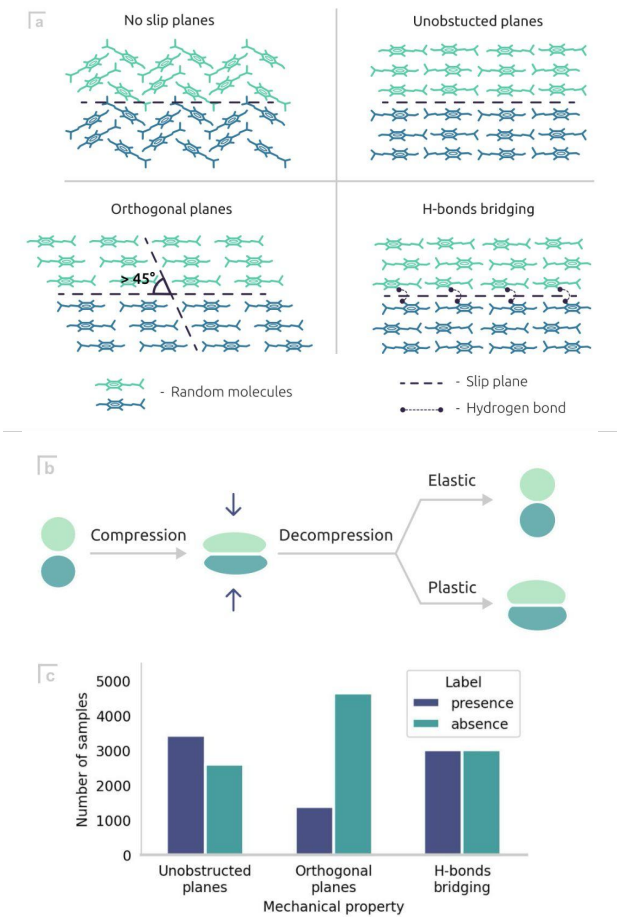
Co-crystals play an important role in many industries, such as energy, electronics, optoelectronics, food, and **pharma**, especially

Tabletability of therapeutic agents can be achieved by co-crystallization

Tabletability is defined by a set of **mechanical properties**, such as plasticity

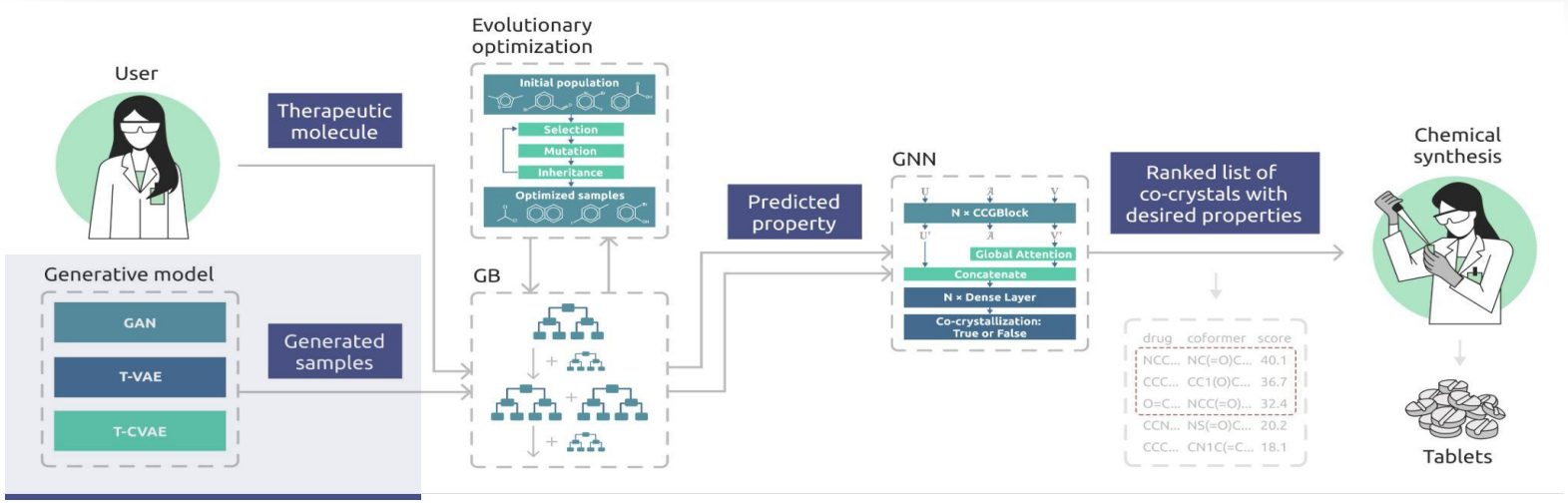
Introduction

Target properties:
"Orthogonal planes",
“H-bonds bridging” and
"Unobstructed planes."



Contribution

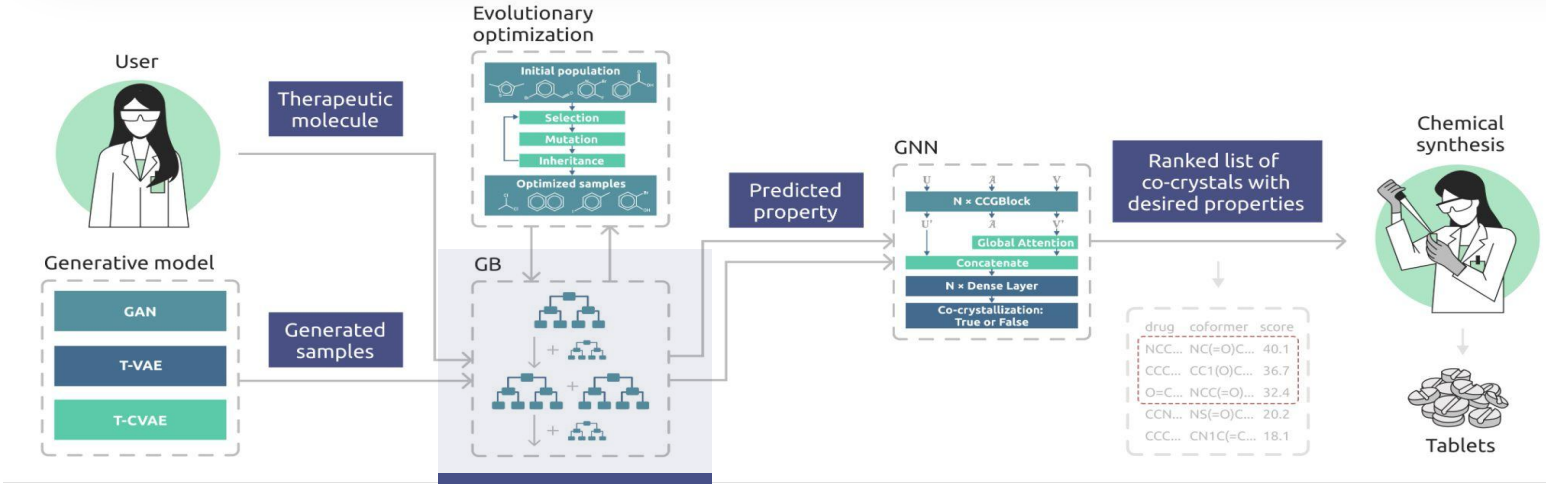
We enable *de novo* co-crystal design with mechanical property control



Coformer generation with generative models ▶ Prediction of mechanical properties with classical ensemble learning ▶ Coformer optimization with graph-based evolutionary algorithm ▶ Estimation of co-crystallization probability with a pretrained graph neural network

Contribution

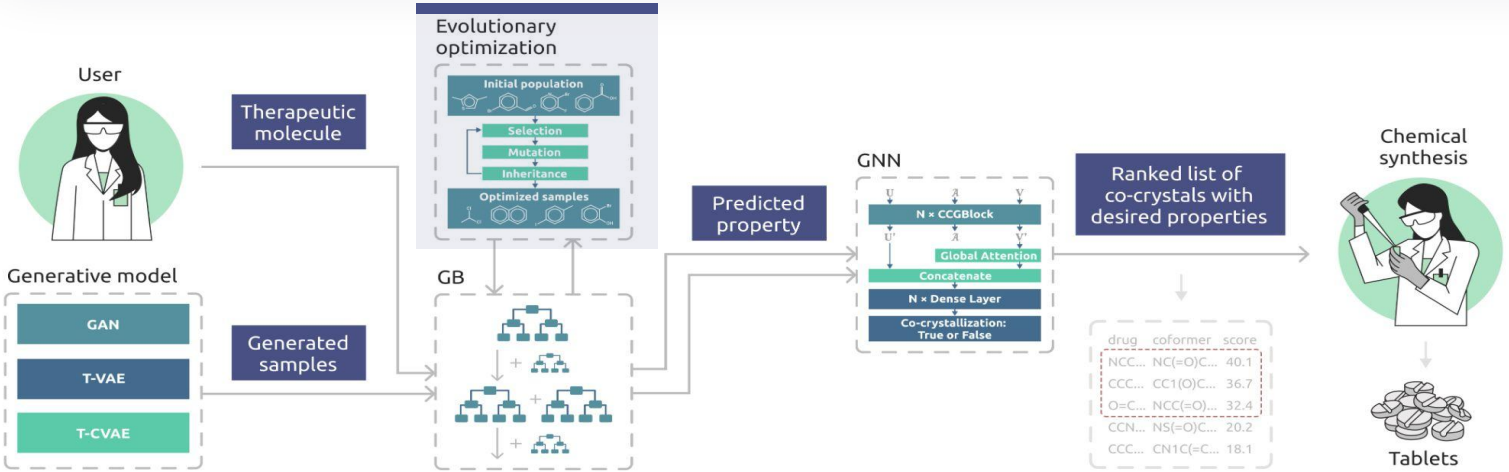
We enable *de novo* co-crystal design with mechanical property control



Coformer generation with LSTM-based GAN ▶ **Prediction of mechanical properties with classical ensemble learning** ▶ Coformer optimization with graph-based evolutionary algorithm ▶ Estimation of co-crystallization probability with a pretrained graph neural network

Contribution

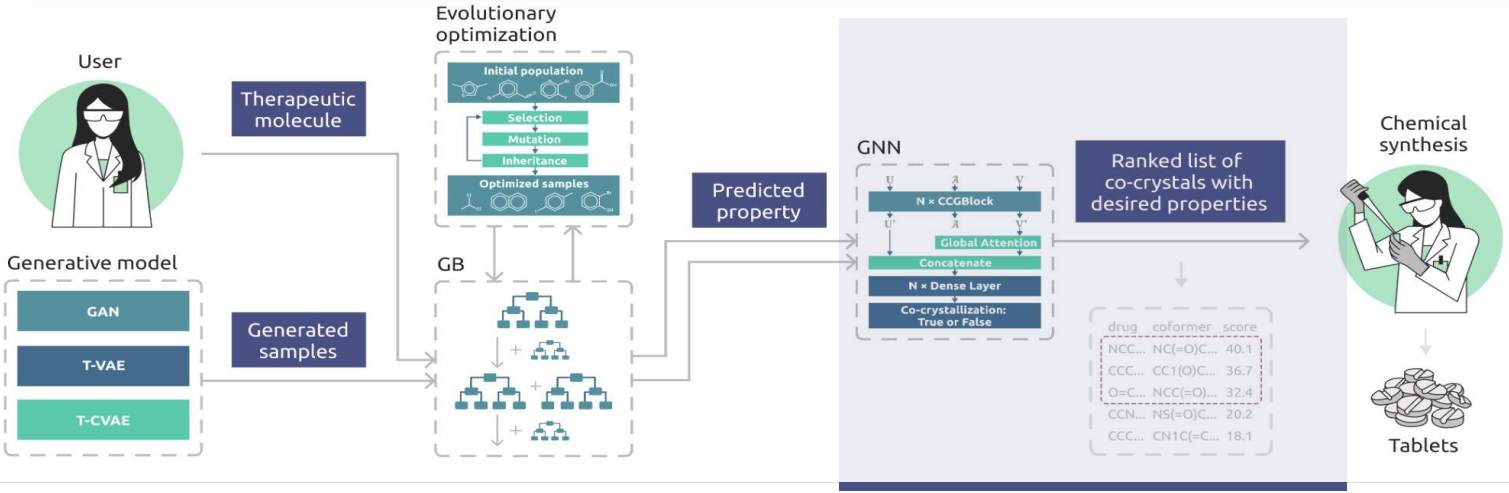
We enable *de novo* co-crystal design with mechanical property control



Coformer generation with LSTM-based GAN ▶ Prediction of mechanical properties with classical ensemble learning ▶ **Coformer optimization with graph-based evolutionary algorithm** ▶ Estimation of co-crystallization probability with a pretrained graph neural network

Contribution

We enable *de novo* co-crystal design with mechanical property control



Coformer generation with LSTM-based GAN ▶ Prediction of mechanical properties with classical ensemble learning ▶ Coformer optimization with graph-based evolutionary algorithm ▶ **Estimation of co-crystallization probability with a pretrained graph neural network**

Data

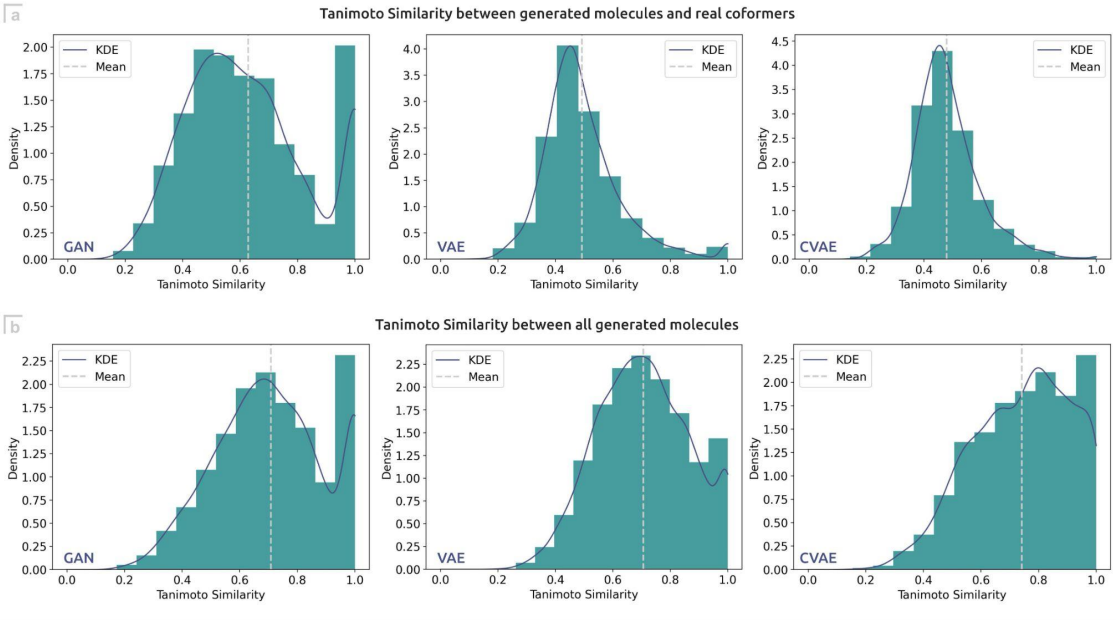
We selected 1.75M samples from the **ChEMBL database** based on the relevant parameter distributions of the known coformers

We retrieved mechanical properties for 6k coformers from the **Cambridge Structural Database (CSD)**

We used SMILES representations to extract molecular features with RDKit

We performed feature engineering and filtering as preprocessing steps

Experimental results

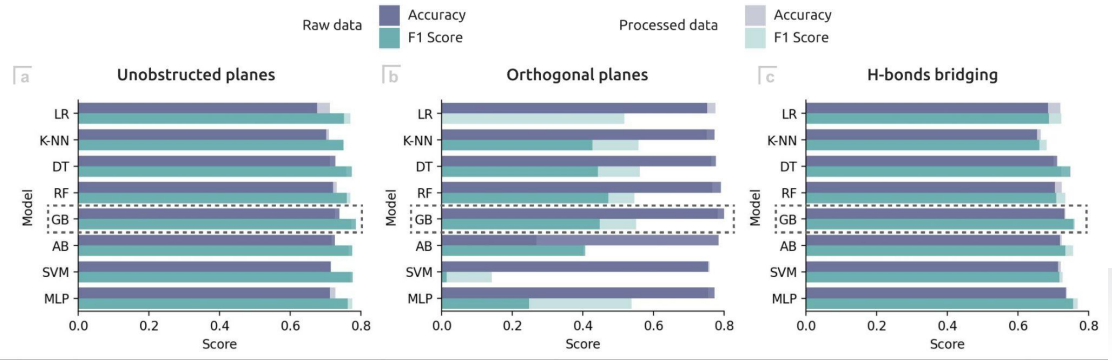


We trained generative models on 1.75M molecules from ChEMBL and fine-tuned on the curated 6k coformers from CSD

We trained ML models and selected the best one predicting mechanical properties of coformers

We employed **evolutionary optimization** to significantly improve the tabletability profiles of the generated coformers

Experimental results

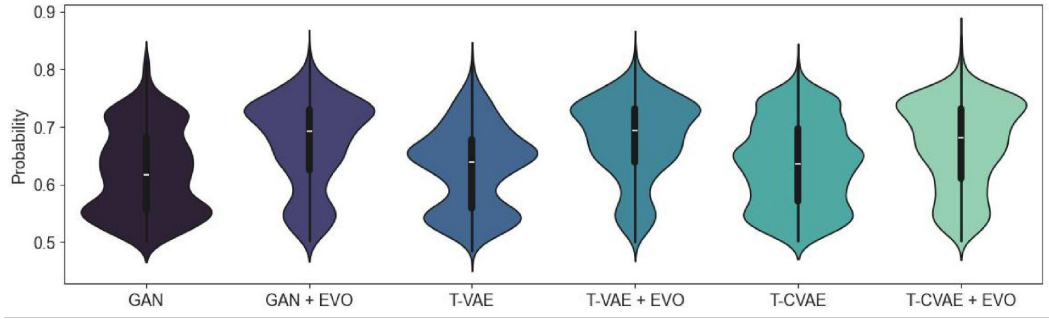


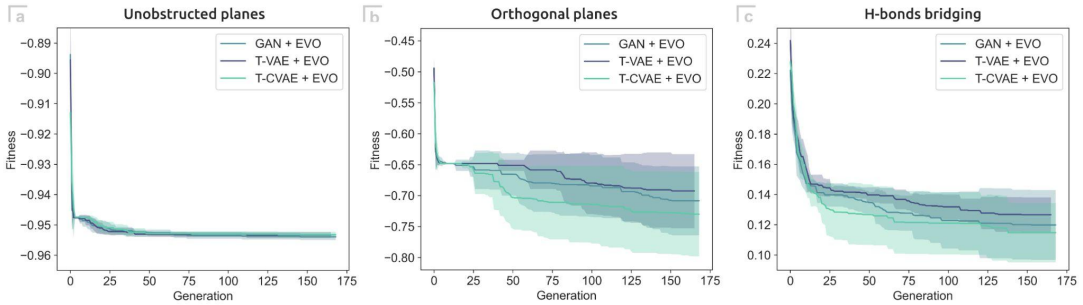
We trained generative models on 1.75M molecules from ChEMBL and fine-tuned on the curated 6k coformers from CSD

We trained **ML models** and selected the best one predicting mechanical properties of coformers

We employed **evolutionary optimization** to significantly improve the tabletability profiles of the generated coformers

Experimental results





We trained generative models on 1.75M molecules from ChEMBL and fine-tuned on the curated 6k coformers from CSD

We trained ML models and selected the best one predicting mechanical properties of coformers

We employed **evolutionary optimization** to significantly improve the tabletability profiles of the generated coformers

Conclusion

We presented a generative pipeline for *de novo* co-crystal design "**GEMCODE**" with target property control

We systematically investigated performance of its individual components to achieve the best results

We demonstrated utility of the pipeline in the **Theophylline case** study and discussed its current limitations

Experimentally validated coformers improving drug tabletability generated by GEMCODE.

Drug	Generated SMILES	CSD Refcode	Model
Nicorandil	O=C(O)C=CC(=O)O	WAHGEV	GAN / T-VAE / T-CVAE
Rivaroxaban	O=C([O-])CC(=O)[O-]	YORVEJ	T-VAE
Paracetamol	C1=CC=C2C=CC=CC2=C1 C[N+](C)(C)CC(=O)[O-]	LUJSIT CUQKAC	GAN / T-VAE / T-CVAE T-CVAE

Nina Gubina
gubina@
scamt-itmo.ru

Andrei Dmitrenko
dmitrenko@
scamt-itmo.ru

Gleb Solovev
glebsolo46@
gmail.com

Nikolay Nikitin
nicl.nno@
gmail.com

Prof. Vladimir
Vinogradov

Contacts

Center for
AI in Chemistry

NSS Lab

Natural Systems
Simulation lab