Generative AI for Co-Crystal Design with Property Control

Nina Gubina,[1] **Andrei Dmitrenko,** [1] **Lyubov Yamshchikova,** [1] **Ivan Lebedev,** [2] **Grigory Kirgizov,** [1]
Nikita Serov, [1] **Nikolay Nikitin,** [1] **Vladimir Vinogradov** [1]

[1] ITMO University, St. Petersburg, Russia
[2] Ivanovo State University of Chemistry and Technology, Ivanovo, Russia

Abstract

Co-crystallization is an accessible way to control physico-chemical characteristics of organic crystals. However, design of co-crystals with predefined properties is an extremely non-trivial task, since it requires a large number of experiments to select a suitable coformer. In practice, experimental screening is often focused on testing a limited number of candidate compounds to keep it feasible within a single study. Artificial intelligence has revolutionized the scientific method and proposed alternatives to tedious experimental work in many fields of science. In this work, for the first time we present a pipeline for automated co-crystal screening *in silico*, which stands on three pillars: generation of candidates for co-crystallization, evaluation of co-crystal properties and estimation of probability of formation. For that, we first trained a GAN on the ChEMBL dataset of 1.75M molecular structures and then fine-tuned it on the state-of-the-art coformer dataset. We then developed a novel approach to predict the tabletability of co-crystals by training machine learning models to predict key properties of co-crystals reflecting powder plasticity and integrating an existing solution to account for the probability of co-crystallization. Finally, we added evolutionary optimization of coformer molecules to reinforce desired mechanical properties of the predicted co-crystals. The resulting pipeline enables fast screening of co-crystals that could reduce the lead optimization time in drug design from several years to a few months including experimental testing. We validated the pipeline by controlling the validity and similarity of the generated coformers and by standard evaluation protocol for machine learning problems. The proposed pipeline is a promising tool for reducing costs in co-crystal design by minimizing the number of necessary experiments.

Introduction

The use of multi-component crystals, specifically co-crystals, have become increasingly popular in various industries including energy (Li et al. 2022), electronics (Wang and Zhang 2020), optoelectronics (Zhu et al. 2015; Sun et al. 2018), food (Dias, Lanza, and Ferreira 2021), and especially in pharmaceuticals (Guo et al. 2021; Duggirala et al. 2019; Yousef and Vangala 2019). Pharmaceutical co-crystals are defined as solids that are crystalline singlephase materials composed of a drug molecule and an additional pharmaceutically acceptable molecule (coformer) (Bolla, Sarma, and Nangia 2022). Co-crystals have a different crystal structure from the original components, leading to unique physico-chemical properties. They are appealing because the resulting solid can exhibit better physicochemical properties compared to either of the pure molecules (Karimi-Jafari et al. 2018). The formation of co-crystals has been shown to enhance characteristics such as bioavailability (Emami et al. 2018), solubility (Good and Rodríguez-Hornedo 2009; He et al. 2017), stability (Babu, Sanphui, and Nangia 2012; Li et al. 2019), pharmacokinetics (Shan et al. 2014), and mechanical properties (He et al. 2017; Karki et al. 2009). Plasticity is a mechanical property that is particularly important for the pharmaceutical industry. It is known that highly plastic materials tend to produce stronger tablets compared to those exhibiting elastic behavior (Bryant, Maloney, and Sykes 2018). Therefore, it is essential to control for plasticity while designing a tablet form of a therapeutic agent.

Despite all the robustness and versatility of co-crystals, determining the combination of a coformer and parent component with the desired property modification is an extremely non-trivial task, usually addressed by experimental high-throughput screening (Childs et al. 2008; Surov et al. 2016). Due to the large amounts of time and effort required, such studies remain targeted, focusing on rather narrow classes of candidate compounds.

Artificial intelligence (AI) methods have recently found their way into the field of chemistry (Cerchia and Lavecchia 2023; Baum et al. 2021; Gormley and Webb 2021; Chan et al. 2019; De Almeida, Moreira, and Rodrigues 2019). Since then, the accumulated experimental data has become the basis for predictive models transforming the traditional way science works. With big data and machine learning (ML), it is now possible to consider a much larger set of candidate molecules for a given problem, rather than being satisfied with a limited number of experiments. Among the pioneering works in the co-crystal domain are the studies aimed at determining the probability of co-crystallization of a particular molecular pair (Vriza et al. 2022; Jiang et al. 2021). However, the sole fact of co-crystallization with no information about the properties of the resulting co-crystals is not enough to inform decision making for a specific use case. Accordingly, another direction of research investigated co-crystal properties with AI methods (Gamidi and Rasmuson 2020; Guo et al. 2023). Still, prediction of most properties has been possible only in the case of already known co-
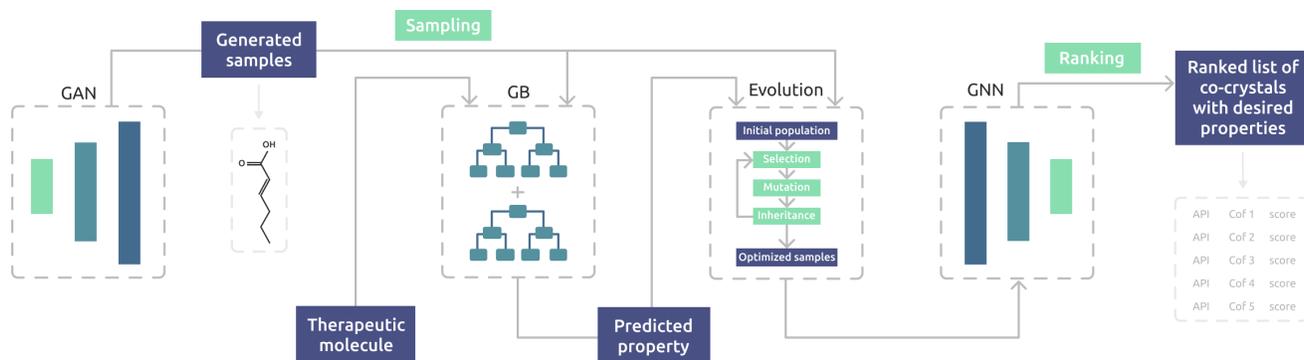
Figure 1: Pipeline for generative co-crystal design consisting of a GAN generating coformer candidates, a gradient boosting (GB) classification model predicting the mechanical properties of co-crystals based on the generated coformers, evolutionary algorithm, and a graph neural network (GNN) ranking co-crystals according to the probability of formation.

crystallising molecular pairs. *De novo* design of co-crystals with predefined properties leveraging big data to cover a large chemical space remains an actual task of great application value.

Therefore, here for the first time we develop a pipeline that generates coformer candidates based on the structure of a drug molecule to form a co-crystal with predefined mechanical properties. We trained a generative adversarial network (GAN) on a dataset of 1.75M chemical structures and then fine-tuned it on the state-of-the-art dataset of co-formers. We then trained a Gradient Boosting (GB) ML model to predict plasticity parameters of the generated coformer candidates. We employed evolutionary optimization leveraging the trained GB models to improve the tabletability profiles of the generated coformers. Finally, we applied a pretrained graph neural network (GNN) to rank the molecular pairs according to the probability of successful co-crystal formation. The output of the proposed pipeline is a set of coformers forming a co-crystal with improved tabletability properties for a selected drug compound. Thus, the pipeline can serve as a tool for selecting the best molecular combination of an active pharmaceutical agent and a coformer delivering the desired properties of the co-crystal.

To the best of our knowledge, our work makes two novel contributions:

- We are the first to develop machine learning models predicting mechanical properties of co-crystals.
- Also, for the first time in the field, we present a generative pipeline for *de novo* co-crystal design.

Related Work

Molecule generation

Traditionally, the process of discovering new molecules or selecting chemical structures for a particular task relies on existing experimental evidence and subjective research experience, both limiting the number and variety of possible compounds to consider. Generative models allow efficient exploration of the molecular space, which has al-

ready caused a rapid growth of molecular generative design. Cutting-edge generative models use string (Blaschke et al. 2020; Hu et al. 2020), 2D (Li, Zhang, and Liu 2018; Jin, Barzilay, and Jaakkola 2018) and 3D (Skalic et al. 2019; Li, Pei, and Lai 2021) molecular graphs as molecular representations. The most common way is the SMILES (Simplified molecular-input line-entry system) notation, as the other approaches have not yet shaped the field to such an extent (Martinelli 2022). Recurrent neural networks (Grisoni et al. 2020; Li et al. 2020), variational autoencoders (Kadurin et al. 2017; Blaschke et al. 2018; Gómez-Bombarelli et al. 2018), generative adversarial networks (GANs) (Guimaraes et al. 2017; Prykhodko et al. 2019; Bian et al. 2019), evolutionary algorithms (Yoshikawa et al. 2018; Leguy et al. 2020a) and hybrid models using reinforcement learning techniques (Putin et al. 2018; Zhavoronkov et al. 2019) have been successfully applied for various problems in chemistry. However, to our knowledge, generative approaches have not yet been used to generate coformer structures. Our work addressed this problem in the context of co-crystal design by training and fine-tuning a GAN model.

Co-crystal property prediction

Research in co-crystal property prediction is targeted at determining various parameters, such as the lattice energy, density, melting temperature, crystal density, enthalpy and entropy of melting, as well as ideal mole fraction solubility of co-crystals (Gamidi and Rasmuson 2017; Rama Krishna et al. 2018; Fathollahi and Sajady 2018; Yue, Wang, and Lu 2023). However, a limited number of samples is typically used in the training phase. For example, Gamidi and Rasmuson trained an artificial neural network on the data of 30 co-crystal systems for 8 different drugs (Gamidi and Rasmuson 2020). Such models are likely to have very limited generalization power beyond the training data. The most recent model predicting the co-crystal density (Guo et al. 2023) used a large training set of 4144 molecular pairs covering a much wider chemical space of possible co-crystals. In this work, we predict several mechanical properties of co-
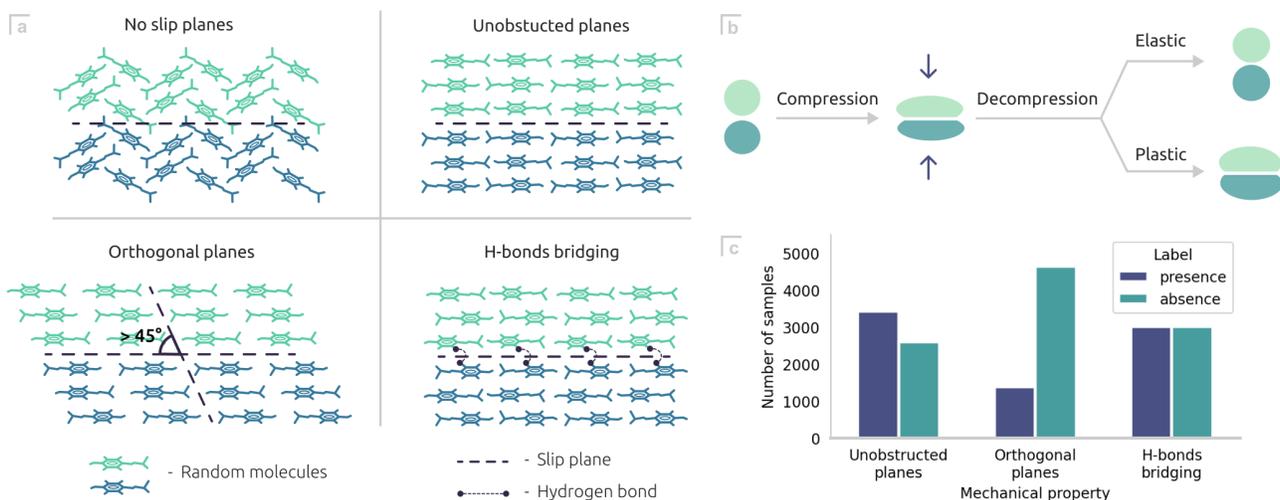
Figure 2: (a) Schematic representation of the mechanical properties of co-crystals. No slip plane is associated with low tabletability. The other three properties positively correlate with tabletability and are predicted in this work. (b) Schematic representation of the particle deformation during powder compression. (c) Number of coformer samples of each category per mechanical property.

crystals for the first time and we do that based on an even larger amount of data (6029 samples), which makes our approach more versatile and better generalizable for different pharmaceutical applications.

Co-crystallization possibility prediction

Determining the possibility of co-crystallization by molecular pairing is an important step in the co-crystal design. For this reason, many works attempted to solve this problem with AI (Yang et al. 2023; Vriza et al. 2022; Wicker et al. 2017). Most works that are closely related to our problem do not provide code to reproduce or reuse their results (Mswahili et al. 2021; Wang et al. 2020; Devogelaer et al. 2020). In our pipeline, we employed CCGNet, which was trained on a set of molecular data with 6819 positive and 1052 negative examples (Jiang et al. 2021). Unlike many of the previous works, CCGNet achieves state-of-the-art performance predicting co-crystal formation while being 100% open-source and easily reproducible.

Results

We present the pipeline for generative co-crystal design with improved tabletability properties. The pipeline consists of four key components, as depicted on Figure 1.

First, a trained and fine-tuned GAN model generates SMILES representations of coformer-like chemical structures. The generated molecules are then fed into the trained ML models along with the therapeutic compounds, where the mechanical properties of co-crystals are predicted. In addition, an evolutionary algorithm is used in combination with the ML models to further improve the tabletability of the generated coformers. Finally, co-crystals with the desired properties are selected for the next step,

where CCGNet (GNN) scores and ranks molecular pairs of drugs and coformers according to the probability of co-crystallization. Thus, the pipeline outputs a list of potential coformers with the desired mechanical properties of the co-crystal, ranked according to the probability of successful co-crystallization. In the following sections, we describe the individual components of the pipeline in more detail.

Data

Datasets of molecules and coformers. In order to train a generative model capable of suggesting reasonable chemical structures, a dataset of molecules from the ChEMBL database was collected. From the large variety of molecular structures available in the database, 1.75M samples were selected according to the following criteria:

- Structural type: molecule.
- Class: small molecules.
- Molecular weight of each component <600 Da.
- Number of hydrogen bond donors (HBD) less than 3 and hydrogen bond acceptors (HBA) less than 8.
- Number of rotatable bonds up to 9.
- Polar surface area up to 138 nm.
- Number of heavy atoms in molecular structure up to 39.

The aforementioned criteria are based on the distributions of relevant parameters in the known coformers and ensure that the GAN is trained on molecules capable of forming co-crystals. However, chemical structures in the ChEMBL database are still substantially different from the structures composing co-crystals. Coformers most often have more basic chemical structures and a smaller variety of functional groups. Therefore, we used an open dataset of 6819 two-component co-crystals (Jiang et al. 2021), which contains

4227 unique chemical structures of the co-formers, for fine-tuning.

For the mechanical properties of co-crystals, we used the Cambridge Structural Database (CSD) and a recently proposed protocol for geometric analysis of co-crystalline materials available with a CSD Python API (Bryant, Maloney, and Sykes 2018). For each of the 6819 available co-crystals, we used the API to query additional experimental data from the CSD and calculate the following binary parameters of plasticity: presence of non-overlapping Miller planes (Unobstructed Planes), presence of orthogonal planes (Orthogonal Planes), and presence of hydrogen bonds between the planes (H-bond bridging). Since some of the co-crystals were missing in CSD, this process yielded a total of 6029 records. This data was then used for training ML models to predict each of the three plasticity parameters.

The obtained mechanical properties of the co-crystals determine their viscoelastic nature. The presence of unobstructed planes and additional slip planes orthogonal to the stacked layers lead to the improved plasticity (Sun 2009). There is also evidence that the lack of hydrogen bonding between the layers has a positive effect on the plasticity of the crystal (Krishna et al. 2015; Reddy et al. 2006). Therefore, an "ideal" co-crystal in terms of plasticity should have a non-overlapping slip planes, the presence of additional orthogonal planes and the absence of hydrogen bonding between the planes (Figure 2a). The compaction properties of many pharmaceutical powders depend on their viscoelastic nature. The closer the material to a perfectly plastic body, the larger the bonding area after compaction of the powder and the denser (less porous) the compressed tablet (Figure 2b). Therefore, accurate prediction of the plasticity parameters is essential for data-driven co-crystal design.

We analyzed the number of samples for each plasticity parameter in the collected dataset (Figure 2c). In the case of orthogonal planes, we observed a dramatic difference between the two groups. When training the corresponding ML model, we accounted for this disproportion by adjusting a threshold probability for predicting a positive class.

Representation of molecules. Traditionally, molecules are represented as structural diagrams with bonds and atoms, but such representations are not well suited for efficient computation. Alternatively, molecules can be represented with SMILES and molecular fingerprints, which have been extensively used for various applications, including the generative models (David et al. 2020). SMILES notation is often used to describe the composition and structure of a chemical molecule by means of short strings (Figure 3a). Whereas molecular fingerprints is a way of representing molecules in the vectorized form (Figure 3b). Therefore, molecular fingerprints enable comparing different structures by calculating similarity measures.

Feature selection. We used RDKit to generate 43 molecular descriptors for each coformer with its SMILES representation (Figure 3c). Since co-crystals consist of two coformer components, each one was described by 86 numerical features in total. Before training ML models for the prediction of mechanical properties, we applied a set of preprocess-
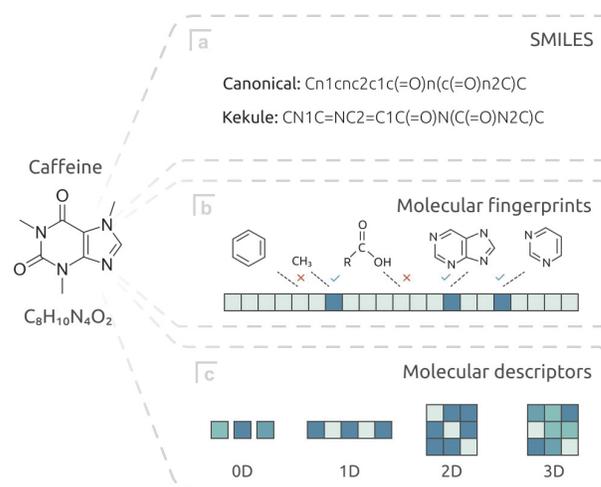

Figure 3: Molecular representation using the chemical structure of caffeine as an example in the form of SMILES, molecular fingerprints, and molecular descriptors.

ing steps. We engineered new features by aggregating the molecular features of the co-formers of the same co-crystal with summation and averaging. To reduce redundancy in the feature space, we investigated the feature importances using embedded methods and the degree of linear association with target variables through correlation coefficients. After feature engineering and filtering, the datasets for the prediction of non-overlapping planes, orthogonal planes, and hydrogen bonding contained 29, 24, and 30 features, respectively.

Generation of coformers

Methods. We used the generative adversarial network (GAN) architecture as one of the most advanced generative approaches. GANs typically consist of two neural networks, a generator and a discriminator, playing an adversarial game against each other while learning the data distribution $p^*(x)$. The generator network receives a random input signal and generates data distribution $p_\theta(x)$, while the discriminator network $D_\phi(x)$ evaluates the generated data and tries to distinguish it from the real training examples (Killoran et al. 2017). In the original formulation, both networks are improved by competing with each other following a min-max optimisation procedure:

$$\min_\theta \max_\phi E_{p^*(x)}[logD_\phi(x)] + E_{p_\theta(x)}[log(1 - D_\phi(x))]$$

Goodfellow et al. proposed alternate generator losses providing better gradients for the generator (Goodfellow et al. 2014):

$$E_{p_\theta(x)}[-log(D_\phi(x))]$$

Since 2014, GANs have been successfully used for numerous applications, including modeling of astronomical phenomena (Schawinski et al. 2017), experiments in particle and high-energy physics (De Oliveira, Paganini, and

Nachman 2017), medical imaging (Wang et al. 2021), and molecule generation (Guimaraes et al. 2017; Prykhodko et al. 2019). Here, we use an open-source GAN implementation[2] inspired by the work of (d'Autume et al. 2020). The GAN takes SMILES representations of molecular structures as input. In the training process, the generator network creates molecular representations from the Gaussian noise and the discriminator network tries to differentiate those from the tokenized SMILES of the real chemical compounds. As a result, the generator learns to output new molecular structures similar to those in the training set.

Implementation details. Performance of GANs is known to strongly depend on hyperparameters and random restarts (Lucic et al. 2018). Therefore, we implemented a grid search to select such hyperparameters as batch sizes and learning rates and performed several repetitive trainings for the best combination. In total, we evaluated 35 different hyperparameter sets by training the GAN for 10,000 steps and calculating validity of the generated molecular structures. Validity refers to the ratio of predicted molecules deemed chemically plausible, it is estimated by RDKit taking into account the valence of atoms in the molecule and the consistency of bonds in aromatic rings. The model showing the steepest increase in the number of valid molecular structures was considered the best and retrained for 30,000 steps.

Ultimately, the generative model was targeted at generation of coformer-like chemical structures. Therefore, after pretraining on the ChEMBL dataset, the GAN was fine-tuned on the dataset of coformers, described in the data section. At this stage, we tested 125 different sets of hyperparameters including learning rates, batch sizes, and the number of additional training steps. We evaluated the validity and percent of unique chemical structures to select the final fine-tuned GAN model.

Finally, we used the Tanimoto index, also known as Jaccard index, to quantify structural similarity of the generated molecules (Bajusz, Rácz, and Héberger 2015). For that, we obtained fingerprint representations of the generated molecules with RDKit and calculated the Tanimoto index as follows:

$$S_{A,B} = \frac{c}{a + b - c} \, ,$$

where S denotes similarities, a is the number of *on* bits in molecule A, b is the number of *on* bits in molecule B, while c is the number of bits that are *on* in both molecules. We assessed the distribution of this metric for the predicted coformers to control for GAN overfitting and ensure the rich variety of the generated molecules.

Results. The GAN trained on the ChEMBL dataset with batch size of 512 and learning rate of 0.001 consistently produced molecules with validity >75 % after 25,000 training steps. After 30,000 steps, this model was fine-tuned on the coformer dataset with a smaller batch size of 256 for additional 1,000 steps (Figure 4a).

To illustrate the importance of fine-tuning in this case, we employed the t-distributed stochastic neighbor embedding
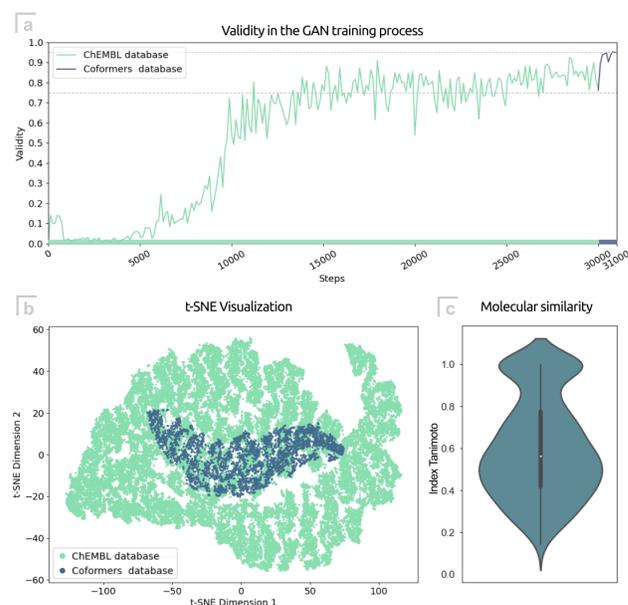
Figure 4: GAN training results on ChEMBL datasets and coformers: (a) plot of the growth of the percentage of valid chemical structures in a batch, (b) t-SNE visualization of molecules from the ChEMBL dataset and coformers, (c) Distribution of the Tanimoto index between the trained GAN-generated molecules and the structures from the training dataset.

(t-SNE) technique and plotted samples of the corresponding datasets in 2D (Figure 4b). The t-SNE analysis reveals that the molecular space of coformers is considerably more constrained compared to that of ChEMBL. Therefore, fine-tuning was critical to shift chemical compound generation towards the molecular space of coformers. The final model was able to produce >95 % of valid and >86 % of unique chemical structures molecules in the test generation of 1000 molecules at 5 times repetition.

To verify the diversity of the predicted molecules, we plotted the distribution of the maximum Tanimoto similarity measure calculated for each predicted molecule against the entire training set (Figure 4c). We observed the median value of 0.56 indicating that the generated molecules were only moderately similar to those in the training data. This suggests that the final GAN does not suffer from overfitting and is perfectly capable of predicting unique and valid coformer structures.

Prediction of mechanical properties of co-crystals

Methods. Since the number of training examples available for prediction of mechanical properties was only 6,029, we resorted to the classical machine learning algorithms. We formulated a binary classification problem for each of the mechanical properties and implemented a number of ML models as a first screen, including logistic regression, k-nearest neighbors classifier, support vector machines, decision trees, multilayer perceptron, as well as ensemble mod-
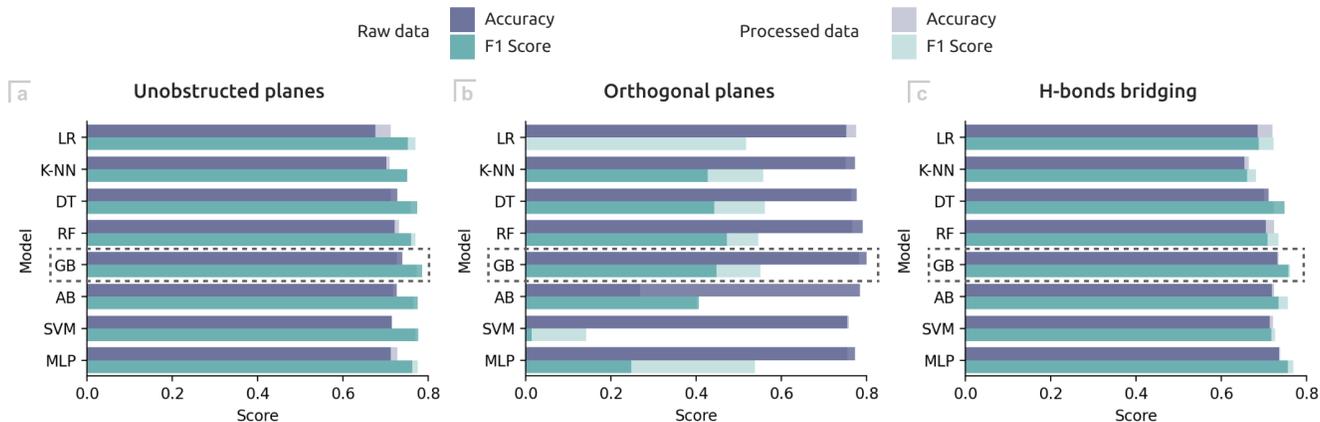
Figure 5: Accuracy and F1 score metrics for the ML models predicting three mechanical properties of co-crystals. (a) Unobstructed planes. (b) Orthogonal planes. (c) H-bonds bridging. The performance of each model is shown before ("Raw data") and after ("Processed data") the feature engineering and feature selection steps.

Task	Property	Data points	Best metric	Generative design	Ref
Regression	Crystal density	26	$R^2 = 0.993$	No	(Fathollahi and Sajady 2018)
Regression	Melting temperature		$R^2 = 0.992$	No	
Regression	Melting enthalpy		$R^2 = 0.999$	No	
Regression	Melting entropy	30	$R^2 = 0.997$	No	(Gamidi and Rasmuson 2020)
Regression	Ideal solubility		$R^2 = 0.953$	No	
Regression	Melting temperature	61	RSD = 2.89%	No	(Gamidi and Rasmuson 2017)
Regression	Lattice energy		RSD = 2.40%	No	
Regression	Crystal density	61	RSD = 1.77%	No	(Rama Krishna et al. 2018)
Regression	Melting temperature	84	$R^2 = 0.998$	No	(Yue, Wang, and Lu 2023)
Regression	Crystal density	4144	$R^2 = 0.985$	No	(Guo et al. 2023)
Classification	Unobstructed planes		Accuracy = 0.731		
Classification	Orthogonal planes	6029	Accuracy = 0.785	Yes	Our work
Classification	H-bonds bridging		Accuracy = 0.734		

Table 1: Comparative table with model metrics on prediction of various co-crystals properties

els, such as random forest (RF) and gradient boosting (GB).

Implementation details. The preprocessed dataset was randomly split into train and test sets in proportion 4:1. The train set was used to optimize hyperparameters of the models with a grid search using the 10-fold cross-validation (CV). The random grid size was 500 and concerned the following parameters: learning rate, number of estimators, subsample, maximum depth of the individual estimators. The test set was used only once, to evaluate and report the performance of the optimized models. We calculated accuracy and F1 score during the CV to select the best hyperparameter set. The use of the two metrics was important given the imbalanced nature of the "Orthogonal planes" and "Unobstructed planes" target variables (Figure 2c).

To account for the disproportion, we also adjusted the threshold for the probability of the positive class by calculating precision and recall metrics. When plotted against the threshold values, the intersection of the two curves represents the optimal point for balancing the number of false positives and false negatives.

Finally, we employed SHapley Additive exPlanations (SHAP) to interpret model predictions. SHAP is a model-independent method based on sensitivity analysis investigating the effect of systematic changes in feature values on the model output (Lundberg and Lee 2017). Such analysis serves as an additional validation of the trained models by inspecting the set of the most important features in the decision-making process and putting it into the context of the original domain. But also, it enables domain-specific hypothesis generation while contributing to the explainability of the predictive model, which is a huge benefit for potential applications.

Results. Overall, the GB model showed the best accuracy and F1 score compared to the other models across all tasks (Figure 5). Despite the high accuracy for the orthogonal planes parameter, we obtained a moderate F1 score suggesting that the final model is more likely to predict the absence of the orthogonal planes. This is attributed to the dispro-

portion in the training examples discussed earlier. Although we demonstrated a significant improvement in metrics by introducing the probability threshold evaluating the model trained on the processed data, it was not enough to entirely resolve this issue.

We optimized the hyperparameters of the Gradient Boosting (GB) model, which resulted in the performance metrics outlined in Table 1. Furthermore, we conducted a thorough review of the existing research on the prediction of co-crystal properties to compare with our results. Notably, we are the first to develop predictive models for the plasticity parameters, so our metrics set the state of the art. In addition, our work clearly stands out by the number of data points used for training.

With SHAP analysis, we learned that the number of atoms among the molecular pairs forming a co-crystal is a decisive factor in the prediction of non-overlapping and orthogonal planes. In both cases, the decrease in the number of atoms in the coformer molecules significantly contributed to the presence of non-overlapping and orthogonal planes. The descriptors associated with the number of hydrogen bond donors (HBD) also had a high degree of importance. As expected, an increase in the number of HBD resulted in the hydrogen bonds forming between planes of the co-crystal.

Evolutionary optimization of coformers

Methods. To increase the quality of coformer generation, we apply a graph-based evolutionary algorithm to structures produced by the GAN. The software implementation is obtained from self-developed *GOLEM*[1] library. The fitness function is designed to reinforce the mechanical characteristics of the molecules being optimized based on predictions of the classification models described above:

$$f(x) = (1 - p_{up}(x), 1 - p_{op}(x), p_{hbb}(x))^T,$$

where x is an evaluated molecule of coformer, $p_{up}(x)$ is the probability of the positive class for unobstructed planes, $p_{op}(x)$ is the same probability for orthogonal planes, and $p_{hbb}(x)$ - for H-bond bridging. Therefore, minimization of the fitness function f leads to generation of coformer molecules having an improved tabletability profile.

Implementation details. The multi-objective optimization algorithm used in this work considers molecules as undirected graphs and follows the generational evolutionary scheme SPEA2 (Zitzler, Laumanns, and Thiele 2001). First, a population of individuals is evaluated with the fitness function. Then, SPEA2-based selection is applied to pick individuals from the population to undergo mutation. After the variation by mutation is done, the inheritance operator is used to form the new population of individuals to proceed to the next iteration.

The initial population of coformer structures (obtained with GAN) were varied by the set of mutation operators (inspired by (Leguy et al. 2020b)). The set of mutations includes simple operations (add, delete, or replace an atom, delete or replace a bond) and more complicated, multi-step
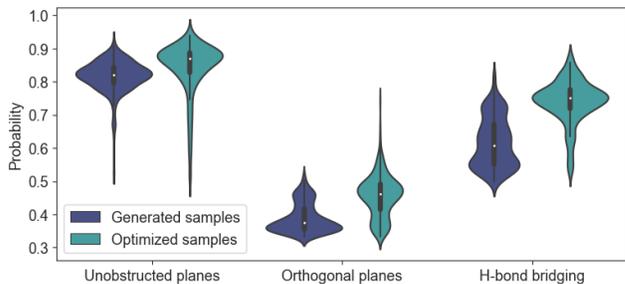
[1]https://github.com/aimclub/GOLEM

Figure 6: Comparison of probability distributions for the key mechanical properties of coformers generated by the GAN and further optimized with the evolutionary algorithm.

Metric type	Median probability		p-value
	Generated	Optimized	
Unobstructed planes	0.82	0.87 (+6.1%)	2.415e-31
Orthogonal planes	0.37	0.46 (+24.3%)	1.144e-51
H-bond bridging	0.61	0.75 (+23.0%)	2.705e-84

Table 2: Results and statistical significance of the evolutionary optimization.

actions (delete or move a functional group, insert carbon, remove an atom if it has only two neighbors).

More than 128,000 new molecules with suboptimal characteristics were generated after 20 independent runs of optimization. For each run of evolution, we used 1000 random structures generated by the GAN as the initial population. Analyzing the dataset of known coformers (Jiang et al. 2021), we estimated the maximal number of heavy atoms to be 50 and the available elements to be C, N, O, F, P, S, Cl, Br, and I. Population sizes were set to 200, number of iterations to 200 and timeout to 60 minutes.

Evolutionary algorithms tend to produce redundantly complicated structures due to overfitting (Gonçalves and Silva 2011). To avoid unrealistic molecules, synthetic accessibility score (SA) (Ertl and Schuffenhauer 2009) was calculated for all obtained molecules. Only coformers with $SA \leq 3$ were selected for further consideration.

Results. To evaluate results of the evolutionary search, we extracted the best molecules from the Pareto frontiers of the independent runs and compared them to the best molecules produced by the GAN. The results are presented in Figure 6 and described in Table 2. We observed a significant increase in probabilities of generated co-crystals to have the desired mechanical properties after the evolutionary optimization, assessed by the non-parametric one-sided Mann-Whitney test. Notably, the median probability for the desired properties of H-bond bridging and orthogonal planes increased by 23.0% and 24.3%, respectively.

Estimation of probability of co-crystal formation

To account for the probability of co-crystallization, we applied an existing GNN-based deep learning framework, called CCGNet (Jiang et al. 2021). With an average balanced accuracy of 98.6%, CCGNet efficiently scores and

ranks coformer candidates according to the probability of co-crystal formation. Since CCGNet was originally trained on the same database of coformers, we did not perform any fine-tuning and simply integrated the model from the open GitHub repository[3] into the pipeline.

The pipeline for *de novo* co-crystal design

To demonstrate the functionality of the pipeline, we generated a set of coformers for Theophylline, a well-known drug against respiratory diseases that has recently been proposed as an adjuvant in the treatment of COVID-19 patients (Montaño et al. 2022). The pipeline produced 760 candidate molecules with good tabletability properties, of which 225 coformers received high scores for the co-crystallization potential with Theophylline (162 produced by the GAN and 63 by the evolutionary optimization). Figure 7 presents eleven high-score molecules picked at random.

Based on our observations, the GAN produces molecules with a higher likelihood of co-crystallization compared to the evolutionary algorithm. On the other hand, molecules refined through the evolutionary algorithm exhibit the desired mechanical properties with higher probability. Thus, coformer generation with the proposed pipeline necessitates a trade-off between the desired properties and the potential for co-crystallization, given the critical importance of both of these parameters.

One of the common functional groups in the generated chemical structures is carboxyl. Notably, this functional group often forms the hydrogen bonding synthons in the Theophylline co-crystals (Kakkar et al. 2018). Moreover, 3,4-dichlorobenzoic acid, which differs by only one halogen group from the generated compound (molecule with Score=20.5 in Figure 7) is now a confirmed coformer of Theophylline, while its tabletability properties were not discussed in the original study of Kakkar et al.

The evidence presented above looks very promising for the practical applications of our pipeline. However, a comprehensive experimental validation is required to confirm its utility. It involves organic synthesis of coformers and co-crystal formation followed by a tablet compression experiment. This work is currently in progress.

Based on our empirical results, we anticipate the following limitations of the proposed pipeline:

- The molecular space of the generated coformers may be too narrow for some of the applications. This is the consequence of the small sample size of the coformer dataset used for fine-tuning the GAN.
- As discussed earlier, the GB model is biased towards predicting the absence of the orthogonal planes, which inevitably results in an increased number of false negatives in the predicted set of coformers. In this case, we recommend considering an alternative set of coformers selected by the other two mechanical properties but not all three of them.
- Low-scale screening may still produce coformers failing to form co-crystals, especially those produced by the evo-

---

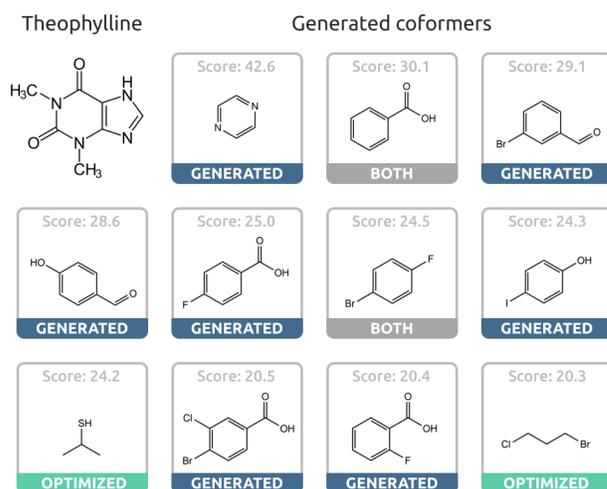[3]https://github.com/Saoge123/ccgnet

Figure 7: Predicted coformers for Theophylline with good tabletability properties, scored and ranked according to the probability of co-crystallization. "Generated" refers to a set of coformers produced by the GAN. "Optimized" is a set of generated coformers that were further improved with the evolutionary algorithm. "Both" represents the intersection of the two sets.

lutionary optimization. Although the generated coformers are ranked according to the probability of formation, there is currently no method to precisely estimate this probability. Therefore, the more coformers are screened for a particular therapeutic agent, the higher the chance to obtain molecular pairs forming co-crystals.

Most of the limitations of the proposed pipeline can be solved with more data available for training. Data availability remains the major challenge for AI applications in co-crystallization.

Conclusion

In this work, for the first time we developed a generative pipeline for *de novo* co-crystal design. The pipeline consists of a GAN generating coformer molecules for a given therapeutic agent, a gradient boosting model predicting mechanical properties associated with tabletability and a GNN-based model ranking coformer candidates according to the probability of forming co-crystals. In addition, an evolutionary optimization step allows to explore an even wider space of molecules while reinforcing the desired mechanical properties. We validated the pipeline by predicting coformers for Theophylline, a drug against respiratory diseases. Despite limitations associated with data availability, the pipeline enables fast generation of unique and valid chemical structures of coformers with predefined tabletability properties and high probability of co-crystallization. This research paves the way for more efficient co-crystal design benefiting pharmaceutical industry and beyond.

Acknowledgments

References

Babu, N. J.; Sanphui, P.; and Nangia, A. 2012. Crystal Engineering of Stable Temozolomide Cocrystals. *Chemistry - An Asian Journal*, 7(10): 2274–2285.

Bajusz, D.; Rácz, A.; and Héberger, K. 2015. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics*, 7(1): 20.

Baum, Z. J.; Yu, X.; Ayala, P. Y.; Zhao, Y.; Watkins, S. P.; and Zhou, Q. 2021. Artificial Intelligence in Chemistry: Current Trends and Future Directions. *Journal of Chemical Information and Modeling*, 61(7): 3197–3212.

Bian, Y.; Wang, J.; Jun, J. J.; and Xie, X.-Q. 2019. Deep Convolutional Generative Adversarial Network (dcGAN) Models for Screening and Design of Small Molecules Targeting Cannabinoid Receptors. *Molecular Pharmaceutics*, 16(11): 4451–4460.

Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; and Patronov, A. 2020. REINVENT 2.0: an AI tool for de novo drug design. *Journal of chemical information and modeling*, 60(12): 5918–5922.

Blaschke, T.; Olivecrona, M.; Engkvist, O.; Bajorath, J.; and Chen, H. 2018. Application of generative autoencoder in de novo molecular design. Mol Inform 37: 1700123.

Bolla, G.; Sarma, B.; and Nangia, A. K. 2022. Crystal Engineering of Pharmaceutical Cocrystals in the Discovery and Development of Improved Drugs. *Chemical Reviews*, 122(13): 11514–11603.

Bryant, M. J.; Maloney, A. G. P.; and Sykes, R. A. 2018. Predicting mechanical properties of crystalline materials through topological analysis. *CrystEngComm*, 20(19): 2698–2704.

Cerchia, C.; and Lavecchia, A. 2023. New avenues in artificial-intelligence-assisted drug discovery. *Drug Discovery Today*, 28(4): 103516.

Chan, H. S.; Shan, H.; Dahoun, T.; Vogel, H.; and Yuan, S. 2019. Advancing Drug Discovery via Artificial Intelligence. *Trends in Pharmacological Sciences*, 40(8): 592–604.

Childs, S. L.; Rodríguez-Hornedo, N.; Reddy, L. S.; Jayasankar, A.; Maheshwari, C.; McCausland, L.; Shipplett, R.; and Stahly, B. C. 2008. Screening strategies based on solubility and solution composition generate pharmaceutically acceptable cocrystals of carbamazepine. *CrystEngComm*, 10(7): 856.

d'Autume, C. d. M.; Rosca, M.; Rae, J.; and Mohamed, S. 2020. Training language GANs from Scratch. ArXiv:1905.09922 [cs, stat].

David, L.; Thakkar, A.; Mercado, R.; and Engkvist, O. 2020. Molecular representations in AI-driven drug discovery: a review and practical guide. *Journal of Cheminformatics*, 12(1): 56.

De Almeida, A. F.; Moreira, R.; and Rodrigues, T. 2019. Synthetic organic chemistry driven by artificial intelligence. *Nature Reviews Chemistry*, 3(10): 589–604.

De Oliveira, L.; Paganini, M.; and Nachman, B. 2017. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis. *Computing and Software for Big Science*, 1(1): 4.

Devogelaer, J.; Meekes, H.; Tinnemans, P.; Vlieg, E.; and Gelder, R. 2020. Co-crystal Prediction by Artificial Neural Networks**. *Angewandte Chemie International Edition*, 59(48): 21711–21718.

Dias, J. L.; Lanza, M.; and Ferreira, S. R. 2021. Cocrystallization: A tool to modulate physicochemical and biological properties of food-relevant polyphenols. *Trends in Food Science & Technology*, 110: 13–27.

Duggirala, N. K.; LaCasse, S. M.; Zaworotko, M. J.; Krzyzaniak, J. F.; and Arora, K. K. 2019. Pharmaceutical cocrystals: formulation approaches to develop robust drug products. *Crystal Growth & Design*, 20(2): 617–626.

Emami, S.; Siahi-Shadbad, M.; Adibkia, K.; and Barzegar-Jalali, M. 2018. Recent advances in improving oral drug bioavailability by cocrystals. *BioImpacts*, 8(4): 305–320.

Ertl, P.; and Schuffenhauer, A. 2009. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1: 1–11.

Fathollahi, M.; and Sajady, H. 2018. Prediction of density of energetic cocrystals based on QSPR modeling using artificial neural network. *Structural Chemistry*, 29(4): 1119–1128.

Gamidi, R. K.; and Rasmuson, A. C. 2017. Estimation of Melting Temperature of Molecular Cocrystals Using Artificial Neural Network Model. *Crystal Growth & Design*, 17(1): 175–182.

Gamidi, R. K.; and Rasmuson, C. 2020. Analysis and Artificial Neural Network Prediction of Melting Properties and Ideal Mole fraction Solubility of Cocrystals. *Crystal Growth & Design*, 20(9): 5745–5759.

Gonçalves, I.; and Silva, S. 2011. Experiments on controlling overfitting in genetic programming. In *15th Portuguese conference on artificial intelligence (EPIA 2011)*, 10–13.

Good, D. J.; and Rodríguez-Hornedo, N. 2009. Solubility Advantage of Pharmaceutical Cocrystals. *Crystal Growth & Design*, 9(5): 2252–2264.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Networks. Publisher: arXiv Version Number: 1.

Gormley, A. J.; and Webb, M. A. 2021. Machine learning in combinatorial polymer chemistry. *Nature Reviews Materials*, 6(8): 642–644.

Grisoni, F.; Moret, M.; Lingwood, R.; and Schneider, G. 2020. Bidirectional Molecule Generation with Recurrent Neural Networks. *Journal of Chemical Information and Modeling*, 60(3): 1175–1183.

Guimaraes, G. L.; Sanchez-Lengeling, B.; Outeiral, C.; Farias, P. L. C.; and Aspuru-Guzik, A. 2017. Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models. Publisher: arXiv Version Number: 3.

Guo, J.; Sun, M.; Zhao, X.; Shi, C.; Su, H.; Guo, Y.; and Pu, X. 2023. General Graph Neural Network-Based Model To Accurately Predict Cocrystal Density and Insight from Data Quality and Feature Representation. *Journal of Chemical Information and Modeling*, 63(4): 1143–1156.

Guo, M.; Sun, X.; Chen, J.; and Cai, T. 2021. Pharmaceutical cocrystals: A review of preparations, physicochemical properties and applications. *Acta Pharmaceutica Sinica B*, 11(8): 2537–2564.

Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; and Aspuru-Guzik, A. 2018. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science*, 4(2): 268–276.

He, H.; Zhang, Q.; Li, M.; Wang, J.-R.; and Mei, X. 2017. Modulating the Dissolution and Mechanical Properties of Resveratrol by Cocrystallization. *Crystal Growth & Design*, 17(7): 3989–3996.

Hu, F.; Wang, D.; Hu, Y.; Jiang, J.; and Yin, P. 2020. Generating novel compounds targeting SARS-CoV-2 main protease based on imbalanced dataset. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 432–436. IEEE.

Jiang, Y.; Yang, Z.; Guo, J.; Li, H.; Liu, Y.; Guo, Y.; Li, M.; and Pu, X. 2021. Coupling complementary strategy to flexible graph neural network for quick discovery of coformer in diverse co-crystal materials. *Nature Communications*, 12(1): 5950.

Jin, W.; Barzilay, R.; and Jaakkola, T. 2018. Junction Tree Variational Autoencoder for Molecular Graph Generation. In Dy, J.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 2323–2332. PMLR.

Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; and Zhavoronkov, A. 2017. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. *Molecular Pharmaceutics*, 14(9): 3098–3104.

Kakkar, S.; Bhattacharya, B.; Reddy, C. M.; and Ghosh, S. 2018. Tuning mechanical behaviour by controlling the structure of a series of theophylline co-crystals. *CrystEngComm*, 20(8): 1101–1109.

Karimi-Jafari, M.; Padrela, L.; Walker, G. M.; and Croker, D. M. 2018. Creating cocrystals: A review of pharmaceutical cocrystal preparation routes and applications. *Crystal Growth & Design*, 18(10): 6370–6387.

Karki, S.; FrisÌˇcÌˇicÌˇ, T.; FÃ¡biÃ¡n, L.; Laity, P. R.; Day, G. M.; and Jones, W. 2009. Improving Mechanical Properties of Crystalline Solids by Cocrystal Formation: New Compressible Forms of Paracetamol. *Advanced Materials*, 21(38–39): 3905–3909.

Killoran, N.; Lee, L. J.; Delong, A.; Duvenaud, D.; and Frey, B. J. 2017. Generating and designing DNA with deep generative models. Publisher: arXiv Version Number: 1.

Krishna, G. R.; Shi, L.; Bag, P. P.; Sun, C. C.; and Reddy, C. M. 2015. Correlation Among Crystal Structure, Mechanical Behavior, and Tabletability in the Co-Crystals of Vanillin Isomers. *Crystal Growth & Design*, 15(4): 1827–1832.

Leguy, J.; Cauchy, T.; Glavatskikh, M.; Duval, B.; and Da Mota, B. 2020a. EvoMol: a flexible and interpretable evolutionary algorithm for unbiased de novo molecular generation. *Journal of cheminformatics*, 12(1): 1–19.

Leguy, J.; Cauchy, T.; Glavatskikh, M.; Duval, B.; and Mota, B. D. 2020b. EvoMol: a flexible and interpretable evolutionary algorithm for unbiased de novo molecular generation. *Journal of Cheminformatics*, 12.

Li, D.; Li, J.; Deng, Z.; and Zhang, H. 2019. Piroxicam–clonixin drug–drug cocrystal solvates with enhanced hydration stability. *CrystEngComm*, 21(28): 4145–4149.

Li, H.-j.; Liu, J.-c.; Yang, L.; Yan, Z.-z.; Lu, Y.-w.; Han, J.-m.; Ren, X.-t.; and Li, W. 2022. Theoretical predict structure and property of the novel CL-20/2,4-DNI cocrystal by systematic search approach. *Defence Technology*, 18(6): 907–917.

Li, X.; Xu, Y.; Yao, H.; and Lin, K. 2020. Chemical space exploration based on recurrent neural networks: applications in discovering kinase inhibitors. *Journal of cheminformatics*, 12(1): 1–13.

Li, Y.; Pei, J.; and Lai, L. 2021. Structure-based de novo drug design using 3D deep generative models. *Chemical science*, 12(41): 13664–13675.

Li, Y.; Zhang, L.; and Liu, Z. 2018. Multi-objective de novo drug design with conditional graph generative model. *Journal of cheminformatics*, 10: 1–24.

Lucic, M.; Kurach, K.; Michalski, M.; Gelly, S.; and Bousquet, O. 2018. Are GANs Created Equal? A Large-Scale Study. ArXiv:1711.10337 [cs, stat].

Lundberg, S.; and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. Publisher: arXiv Version Number: 2.

Martinelli, D. D. 2022. Generative machine learning for de novo drug discovery: A systematic review. *Computers in Biology and Medicine*, 145: 105403.

Montaño, L.; Sommer, B.; Gomez-Verjan, J.; Morales-Paoli, G.; Ramírez-Salinas, G.; Solís-Chagoyán, H.; Sanchez-Florentino, Z.; Calixto, E.; Pérez-Figueroa, G.; Carter, R.; Jaimez-Melgoza, R.; Romero-Martínez, B.; and Flores-Soto, E. 2022. Theophylline: Old Drug in a New Light, Application in COVID-19 through Computational Studies. *International Journal of Molecular Sciences*, 23(8): 4167.

Mswahili, M. E.; Lee, M.-J.; Martin, G. L.; Kim, J.; Kim, P.; Choi, G. J.; and Jeong, Y.-S. 2021. Cocrystal Prediction Using Machine Learning Models and Descriptors. *Applied Sciences*, 11(3): 1323.

Prykhodko, O.; Johansson, S. V.; Kotsias, P.-C.; Arús-Pous, J.; Bjerrum, E. J.; Engkvist, O.; and Chen, H. 2019. A de novo molecular generation method using latent vector based generative adversarial network. *Journal of Cheminformatics*, 11(1): 74.

Putin, E.; Asadulaev, A.; Ivanenkov, Y.; Aladinskiy, V.; Sanchez-Lengeling, B.; Aspuru-Guzik, A.; and Zhavoronkov, A. 2018. Reinforced Adversarial Neural Computer for *de Novo* Molecular Design. *Journal of Chemical Information and Modeling*, 58(6): 1194–1204.

Rama Krishna, G.; Ukrainczyk, M.; Zeglinski, J.; and Rasmuson, C. 2018. Prediction of Solid State Properties of Cocrystals Using Artificial Neural Network Modeling. *Crystal Growth & Design*, 18(1): 133–144.

Reddy, C. M.; Kirchner, M. T.; Gundakaram, R. C.; Padmanabhan, K. A.; and Desiraju, G. R. 2006. Isostructurality, Polymorphism and Mechanical Properties of Some Hexahalogenated Benzenes: The Nature of HalogenHalogen Interactions. *Chemistry - A European Journal*, 12(8): 2222–2234.

Schawinski, K.; Zhang, C.; Zhang, H.; Fowler, L.; and Santhanam, G. K. 2017. Generative Adversarial Networks recover features in astrophysical images of galaxies beyond the deconvolution limit. *Monthly Notices of the Royal Astronomical Society: Letters*, slx008.

Shan, N.; Perry, M. L.; Weyna, D. R.; and Zaworotko, M. J. 2014. Impact of pharmaceutical cocrystals: the effects on drug pharmacokinetics. *Expert Opinion on Drug Metabolism & Toxicology*, 10(9): 1255–1271.

Skalic, M.; Sabbadin, D.; Sattarov, B.; Sciabola, S.; and De Fabritiis, G. 2019. From target to drug: generative modeling for the multimodal structure-based ligand design. *Molecular pharmaceutics*, 16(10): 4282–4291.

Sun, C. C. 2009. Materials Science Tetrahedron—A Useful Tool for Pharmaceutical Research and Development. *Journal of Pharmaceutical Sciences*, 98(5): 1671–1687.

Sun, L.; Zhu, W.; Yang, F.; Li, B.; Ren, X.; Zhang, X.; and Hu, W. 2018. Molecular cocrystals: design, charge-transfer and optoelectronic functionality. *Physical Chemistry Chemical Physics*, 20(9): 6009–6023.

Surov, A. O.; Solanko, K. A.; Bond, A. D.; Bauer-Brandl, A.; and Perlovich, G. L. 2016. Cocrystals of the antiandrogenic drug bicalutamide: screening, crystal structures, formation thermodynamics and lattice energies. *CrystEngComm*, 18(25): 4818–4829.

Vriza, A.; Sovago, I.; Widdowson, D.; Kurlin, V.; Wood, P. A.; and Dyer, M. S. 2022. Molecular set transformer: attending to the co-crystals in the Cambridge structural database. *Digital Discovery*, 1(6): 834–850.

Wang, C.; Yang, G.; Papanastasiou, G.; Tsaftaris, S. A.; Newby, D. E.; Gray, C.; Macnaught, G.; and MacGillivray, T. J. 2021. DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis. *Information Fusion*, 67: 147–160.

Wang, D.; Yang, Z.; Zhu, B.; Mei, X.; and Luo, X. 2020. Machine-Learning-Guided Cocrystal Prediction Based on

Large Data Base. *Crystal Growth & Design*, 20(10): 6610–6621.

Wang, Z.; and Zhang, Q. 2020. Organic Donor-Acceptor Cocrystals for Multiferroic Applications. *Asian Journal of Organic Chemistry*, 9(9): 1252–1261.

Wicker, J. G. P.; Crowley, L. M.; Robshaw, O.; Little, E. J.; Stokes, S. P.; Cooper, R. I.; and Lawrence, S. E. 2017. Will they co-crystallize? *CrystEngComm*, 19(36): 5336–5340.

Yang, D.; Wang, L.; Yuan, P.; An, Q.; Su, B.; Yu, M.; Chen, T.; Hu, K.; Zhang, L.; Lu, Y.; and Du, G. 2023. Cocrystal virtual screening based on the XGBoost machine learning model. *Chinese Chemical Letters*, 34(8): 107964.

Yoshikawa, N.; Terayama, K.; Sumita, M.; Homma, T.; Oono, K.; and Tsuda, K. 2018. Population-based de novo molecule generation, using grammatical evolution. *Chemistry Letters*, 47(11): 1431–1434.

Yousef, M. A.; and Vangala, V. R. 2019. Pharmaceutical cocrystals: molecules, crystals, formulations, medicines. *Crystal Growth & Design*, 19(12): 7420–7438.

Yue, H.; Wang, J.; and Lu, M. 2023. Neural Network Prediction Model of Cocrystal Melting Temperature Based on Molecular Descriptors and Graphs. *Crystal Growth & Design*, 23(4): 2540–2549.

Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V.; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; Volkov, Y.; Zholus, A.; Shayakhmetov, R. R.; Zhebrak, A.; Minaeva, L. I.; Zagribelnyy, B. A.; Lee, L. H.; Soll, R.; Madge, D.; Xing, L.; Guo, T.; and Aspuru-Guzik, A. 2019. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9): 1038–1040.

Zhu, W.; Zheng, R.; Fu, X.; Fu, H.; Shi, Q.; Zhen, Y.; Dong, H.; and Hu, W. 2015. Revealing the Charge-Transfer Interactions in Self-Assembled Organic Cocrystals: Two-Dimensional Photonic Applications. *Angewandte Chemie*, 127(23): 6889–6893.

Zitzler, E.; Laumanns, M.; and Thiele, L. 2001. SPEA2: Improving the strength Pareto evolutionary algorithm. *TIK report*, 103.