ACS OMEGA

http://pubs.acs.org/journal/acsodf Article

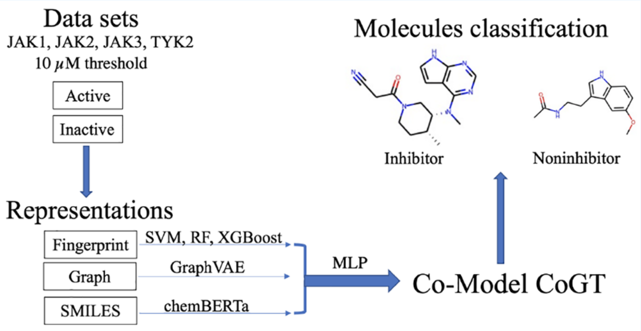
CoGT: Ensemble Machine Learning Method and Its Application on JAK Inhibitor Discovery

Yingzi Bu, Ruoxi Gao, Bohan Zhang, Luchen Zhang, and Duxin Sun*

Cite This: *ACS Omega* 2023, 8, 13232−13242 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The discovery of new drug candidates to inhibit an intended target is a complex and resource-consuming process. A machine learning (ML) method for predicting drug−target interactions (DTI) is a potential solution to improve the efficiency. However, traditional ML approaches have limitations in accuracy. In this study, we developed a novel ensemble model CoGT for DTI prediction using multilayer perceptron (MLP), which integrated graph-based models to extract non-Euclidean molecular structures and large pretrained models, specifically chemBERTa, to process simplified molecular input line entry systems (SMILES). The performance of CoGT was evaluated using compounds inhibiting four Janus kinases (JAKs). Results showed that the large pretrained model, chemBERTa, was better than other conventional ML models in predicting DTI across multiple evaluation metrics, while the graph neural network (GNN) was effective for prediction on imbalanced data sets. To take full advantage of the strengths of these different models, we developed an ensemble model, CoGT, which outperformed other individual ML models in predicting compounds' inhibition on different isoforms of JAKs. Our data suggest that the ensemble model CoGT has the potential to accelerate the process of drug discovery.



■ INTRODUCTION

Janus kinases (JAKs) are a family of enzymes that play a crucial role in the intracellular signaling of cytokine receptors,[1] which are involved in many biological processes such as cell proliferation, apoptosis, and immune regulation.[2,3] Dysregulation of JAKs and JAK-related pathways leads to malignancies and autoimmune disorders such as myelofibrosis, rheumatoid arthritis, inflammatory bowel diseases, multiple sclerosis, and psoriasis.[4] Accordingly, several JAK inhibitors have been approved for the treatment of these diseases.

However, all approved JAK inhibitors have commonly observed side effects, which may be due to their pan-inhibition of different JAK isoforms.[3] The JAK families, JAK1, JAK2, JAK3, and TYK2, have seven homology domains (JH), where JH1 serves as the kinase domain that phosphorylates downstream signaling proteins.[3,4] Most JAK inhibitors are designed to compete with adenosine triphosphate (ATP) for the binding site in the JH1 kinase domain. However, the JH1 is a highly evolutionarily conserved domain,[4] which makes it difficult to develop isoform-selective inhibitors. Thus, an *a priori* tool to predict JAK selectivity of designed molecules will be of valuable help to develop more isoform-specific inhibitors to reduce their side effects. Machine learning can significantly improve the efficiency and accuracy of these processes.

To develop an isoform-specific inhibitor, high-throughput screening and lead compound optimization are usually performed, which are time-consuming and not economically efficient. On the other hand, machine learning methods, such as random forest (RF),[5] support vector machine (SVM),[6] K nearest neighbors (KNN),[7] and extreme gradient boosting (XGBoost),[8] could be applied to accelerate these processes. For instance, XGBoost has shown promising prediction on JAK2 inhibitors, using fingerprint as drug molecule representation.[9] In our study, we also explored the abilities of graph neural network (GNN) models, which directly used a molecule graph as input. In addition, we attempted to experiment with a transformer-based model on JAK inhibition prediction using simplified molecular input line entry systems (SMILES) as input. By integrating different aspects of the ML methods, we developed an ensemble model CoGT (conventional ML models + graph-based models + transformer-based models), aiming to leverage the predicting ability for drug−target interactions (DTI). This novel method could be further applied and validated on drug development of other molecular targets.

Received: January 26, 2023
Accepted: March 16, 2023
Published: March 27, 2023

■ MATERIALS AND METHODS

Data Preparation. This data set was extracted from ChEMBL,[10,11] BindingDB,[12] PubChem,[13,14] and Liu et al.[15] We removed duplicated drugs or drugs with controversial labels (e.g., one drug with both active and inactive labels) in the data set based on compound ID (CID) or compounds' SMILES strings. More than 2,130,000 compounds were extracted from ChEMBL without a label and used to pretrain neural network structured models mentioned later. For four types of JAKs and the number of compounds collected are summarized in Table 1.

Table 1. Number of Molecules Collected in Each JAK Category

molecule number	JAK1	JAK2	JAK3	TYK2
total	7373	10161	7722	2424
active (label 1)	5606	6846	5250	1627
inactive (label 0)	1767	3315	2472	797

The threshold of active drugs is those with IC_{50}, inhibition, EC_{50}, and K_i to a certain JAK below 10 μM. For model training, the data sets were randomly split into training, validation, and test sets in 8:1:1 ratio. To make sure the training, validation, and test set are the same when training all categories of models, we set the random state seed at 42. Therefore, during all of our training processes including later comodel training, compound information from the test set did not leak.

Molecular Fingerprints Calculation. A fingerprint of a molecule is a list of binary bits, which contains information on drug substructure. For instance, each bit of the fingerprint list could be a Boolean determination of certain element presence, ring structure, or atom pairing.[16] In our work, all molecules were represented by MACCS fingerprints (166 bits). Those fingerprints were calculated based on a compound's SMILES using the RDKit package.

Model Building. *SVM.* This defines a margin or decision plane to separate data from different classes. Here, we used MACCS fingerprints as features. We tried different SVM methods: linear, poly, rbf, and sigmoid. Model evaluations for SVM are summarized in Table S1, and area under the curve (AUC) results are shown in Figure S1. We found that SVM poly performed best overall on 4 JAKs, while the SVM sigmoid showed the worst performance (nearly random guessing). Thus, we chose SVM poly for model comparison and later comodel building.

Random Forest. RF consists of individual decision trees, and each tree is trained on a subversion of the data set. We used fingerprints as features, and the number of trees was optimized for each JAK category. The n_estimator for JAK1, JAK2, JAK3, and TYK2 is 53, 91, 48, and 13, respectively.

Extreme Gradient Boosting. XGBoost[17] is a scalable machine learning system for tree boosting. Compared with RF, XGBoost has a range of adjustable parameters to optimize for each JAK category. We did a grid search on parameters listed in Table S2 for each JAK and built the final XGBoost model using the determined optimal parameters.

Graph Model. To leverage the natural structure of chemicals, we attempted to solve the problem by using graph neural networks[18−20] because of their performance and interpretability.[21] For each chemical molecule, we built one graph by taking atoms as nodes and chemical bonds as edges. For each node, we used 6 attributes of the atom: (1) atomic number, (2) atom degree, (3) formal charge, (4) hybridization, (5) aromatic, (6) chiral tag. We removed all hydrogen atoms so the related nodes, edges, and atom degrees would not be included in the graph construction. The architecture figure of our graph model is shown in Figure 1A.

Apart from basic connection information between atoms, bond types (single, double, triple, and aromatic) are employed as edge relations between two nodes to supply more edge information for the constructed graphs, and a relational graph convolutional network (RGCN) is applied to adapt this data structure. The node embedding is initialized by an embedding layer, and the RGCN convoluton layers update the node embedding using the neighbors and relation information:

$$\mathbf{h}_{\mathbf{i}}^{(k+1)} = \sigma\left(\sum_{r\in R}\sum_{j\in\mathcal{N}_r(i)}\frac{1}{|\mathcal{N}_r(i)|}\mathbf{W}_{\mathbf{r}}^{(k)}\mathbf{h}_{\mathbf{j}}^{(k)} + \mathbf{W}_{\mathbf{0}}\mathbf{h}_{\mathbf{i}}^{(k)}\right) \tag{1}$$

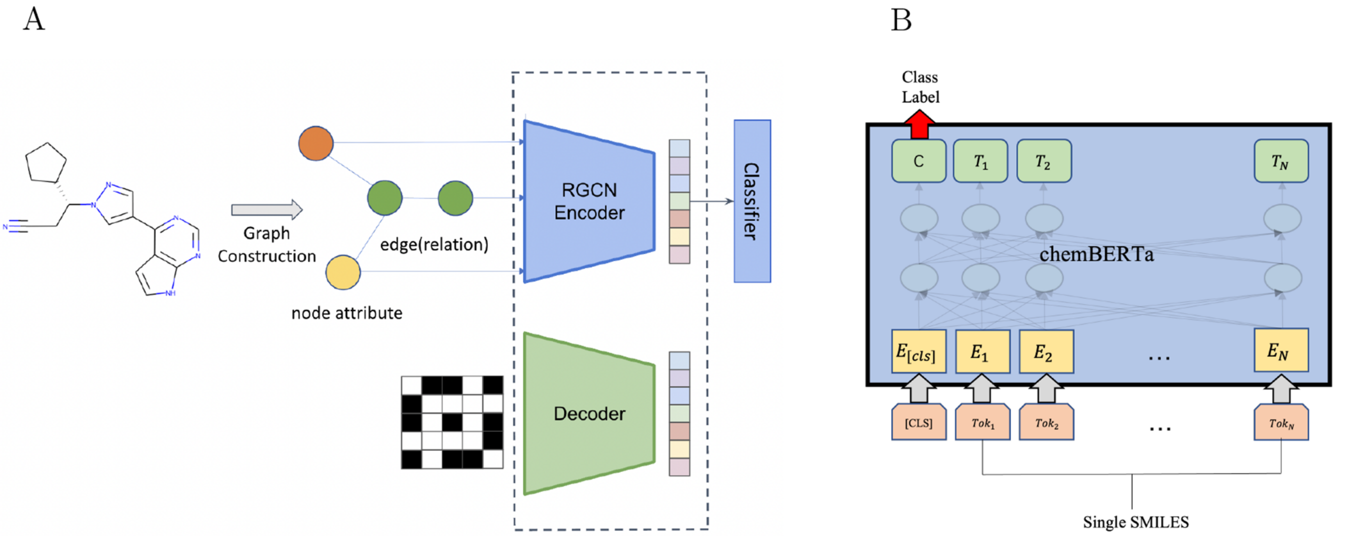


Figure 1. (A) Graph model, where the encoder is a RGCN that maps the drug to latent space Z, and the decoder is implemented by $\sigma(ZZ^T)$. (B) ChemBERTa model.

where $\mathbf{h}_i^{(k)}$ is the node embedding of the ith node after the kth layer, R is the relation set, and $\mathcal{N}_r(i)$ denotes the neighbor set that has r relation of the ith node.

A two-layer RGCN[22] is used to experiment with embedding dimension 4 (embedding layer), hidden dimension 64 (the first RGCN convolutional layer), and output dimension 128 (the second RGCN convolutional layer).

Variational autoencoder (VAE)[23] is considered as a pretraining tool to eliminate the effect of unbalanced data and train a more robust model. We modified variational graph autoencoders (VGAE)[24] as relation-employed graph autoencoders (GraphVAE). The RGCN is the simple inference model, i.e., encoder. The generative model is given by an inner product between latent variables to learn the adjacency matrix. We optimize the variational bound on negative log likelihood:

$$\mathcal{L} = \mathbb{E}_x[-\mathbb{E}_{\mathbf{z}\sim q_\theta(\mathbf{z}|\mathbf{x})}\log(A|\mathbf{z}) + D_{KL}(q_\theta(\mathbf{z}|\mathbf{x}), \mathcal{N}(\mathbf{0}, \mathbf{I}))] \tag{2}$$

where q_θ is the encoder distribution, $\mathbf{z}$ represents the drug representation on latent space, and A denotes the adjacency matrix of a drug graph.

After the GraphVAE training process, the pretrained encoder followed by a global attention pool layer and a linear layer is fine-tuned as a JAK classifier.

chemBERTa. Large pretrained neural networks, especially transformer-based, have made breakthroughs in many domains like language,[25] vision,[26] as well as protein prediction.[27] However, the progress of chemical property prediction using large transformers is not significant compared to these domains. Previous work did not fully take advantage of the capacity of large transformer models as they were either pretrained on smaller language models like recurrent neural networks or tuned on smaller data sets and narrow applications like reaction predictions,[28] which may cause the overfitting of models and be unable to generalize to other tasks. In recent years, a new chemical transformer called chemBERTa[29] makes one of the first attempts to systematically evaluate large transformers on molecular property prediction tasks. As shown in Figure 1B, the chemBERTa is originally pretrained on 77 million unique SMILES of chemicals from PubChem on RoBERTa[30] with a SMILE-based tokenizer[31] to predict corresponding Morgan fingerprints and is then applied to several downstream properties' prediction tasks. The SMILE-based tokenizer, first designed for another pretrained transformer model,[31] tokenizes SMILES strings more reasonably than regular hard tokenization, which turns SMILES into single letters but may lose information when two or more consecutive letters should stay in the integrity. The backbone model, RoBERTa, shares a similar architecture with BERT[25] but shows more robust performances specifically on classification tasks under different training strategies from BERT. The natural of RoBERTa can be a good fit for the chemical property predictions which are usually classifications.

Here, we fine-tuned a chemBERTa with additional two million SMILES, as mentioned in Data Preparation above, where the inputs to the chemBERTa were SMILES of chemicals and the targets were MACCS fingerprints. Even though the original chemBERTa was pretrained to predict Morgan fingerprints, we believe the transformation from a MACCS fingerprint to a Morgan fingerprint can be handled easily by deep neural net models. Also, to be consistent and able to easily ensemble with other methods we explored in this paper, we need to use MACCS fingerprints in the downstream JAK classification task, so we decided to also use MACCS fingerprints in the pretraining stages. In this stage, the model was pretrained in 30 epochs as the loss starts to converge. The optimizer was AdamW[32] with $1e^{-4}$ learning rate and $1e^{-2}$ weight decay. We added a linear layer on top of the pretrained model to fine-tune and cross-validate the JAK classification data set. In the stage of tuning for the JAK prediction task, the model was trained in 20 epochs and the learning rate is $1e^{-5}$, while other settings remain the same as pretraining. For both stages, the batch size was 16 and a 0.5 dropout was applied before the final linear layer.

Convolutional Neural Network. To compare with the large pretrained model, we implemented a convolutional neural network (CNN) as a neural baseline model. The CNN model was also pretrained on the same set of data as the chemBERTa and then fine-tuned on the JAK data set. In the CNN architecture, we had 3 convolution kernels with kernel sizes of 1, 2, and 3 as unigram, bigram, and trigram filters, respectively, which is analogous to common settings in language tasks. The output of 3 kernels after max-pooling was concatenated together to be fed into a final linear prediction layer. The embedding dimension of each character is 256. The output size of all 3 convolution layers is 128. The activation function is LeakyReLu, and the dropout rate is 0.25. In pretraining, the model was trained in 20 epochs, and the optimizer was a stochastic gradient descent with 0.9 learning rate and $1e^{-2}$ weight decay. In the stage of tuning for JAK prediction task, the learning rate was 0.1 while other settings remain the same as pretraining. For both stages, the batch size was 1024.

Baseline Model. In this study, we chose K nearest neighbor (KNN) to evaluate data set as a simple base model. By utilizing a simple model, one could examine the data set and validation with rapid feedback. We would use the base model's performance to contrast with other models.[33] We used the MACCS fingerprints as inputs for fingerprint-based non-neural classification models.

CoGT. To fully utilize the advantages of conventional ML models, graph-based models and transformer-based models, we built a comodel CoGT using simplified multilayer perceptron (MLP). In detail, predicted probabilities of compounds calculated by SVM, RF, XGBoost, GraphVAE, and chemBERTa were taken as input, and probability calculated by sigmoid function was the output by using the SGD optimizer to minimize the weighted BCE loss.

Model Evaluation. All models listed above were evaluated on test sets. Model performance was evaluated based on accuracy, active recall (or sensitivity, SE), negative recall (or specificity, SP), weighted accuracy (average of SE and SP), Matthew's correlation coefficient (MCC), F1 score, AUC, and average precision (AP). The equations to calculate each metric are listed below, in which TP is true positive, TN is true negative, FP is false positive, and FN is false negative. In the AP formula, R_n and P_n denote the precision and recall at the nth threshold, respectively.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{TP} + \text{FP} + \text{FN}} \tag{3}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

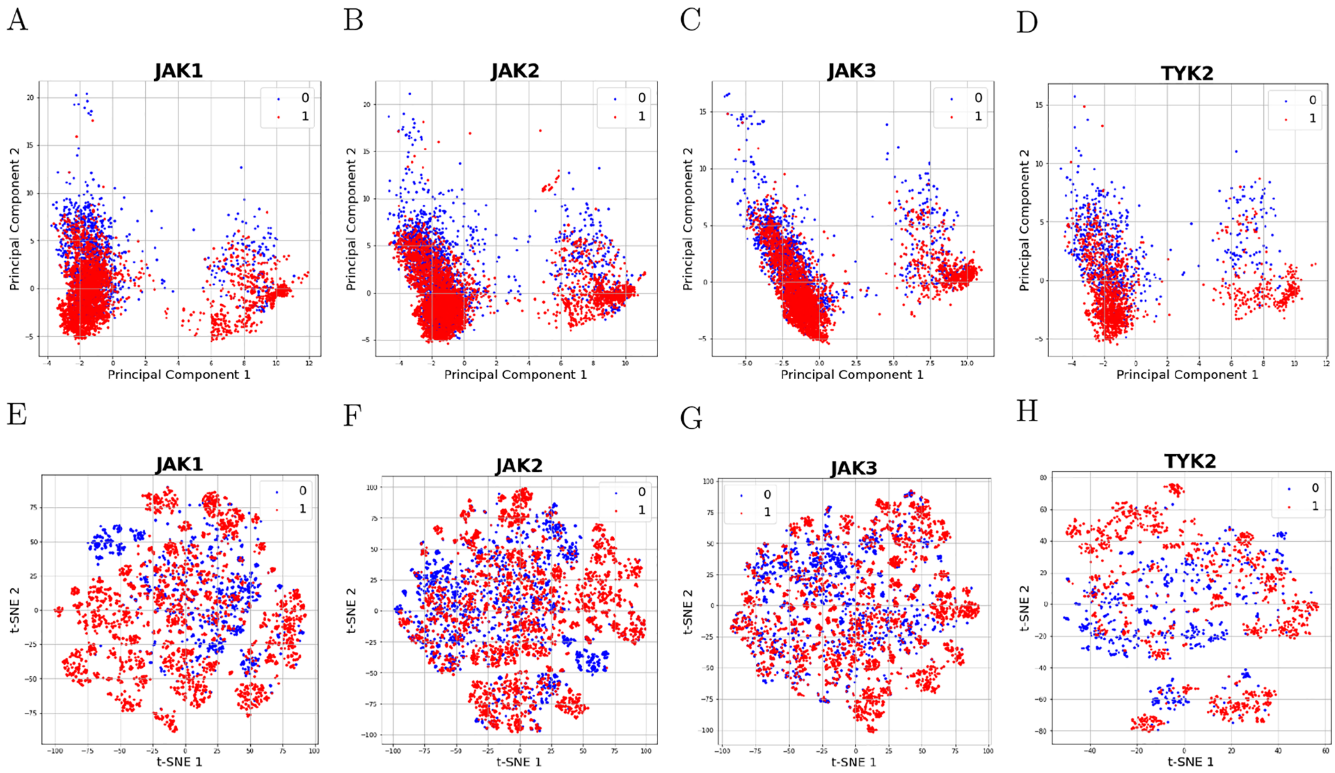


Figure 2. Data visualization based on MACCS fingerprint with PCA and t-SNE. PCA for (A) JAK1, (B) JAK2, (C) JAK3, and (D) TYK2; t-SNE for (E) JAK1, (F) JAK2, (G) JAK3, and (H) TYK2. Blue and red dots represent noninhibitors and inhibitors, respectively.

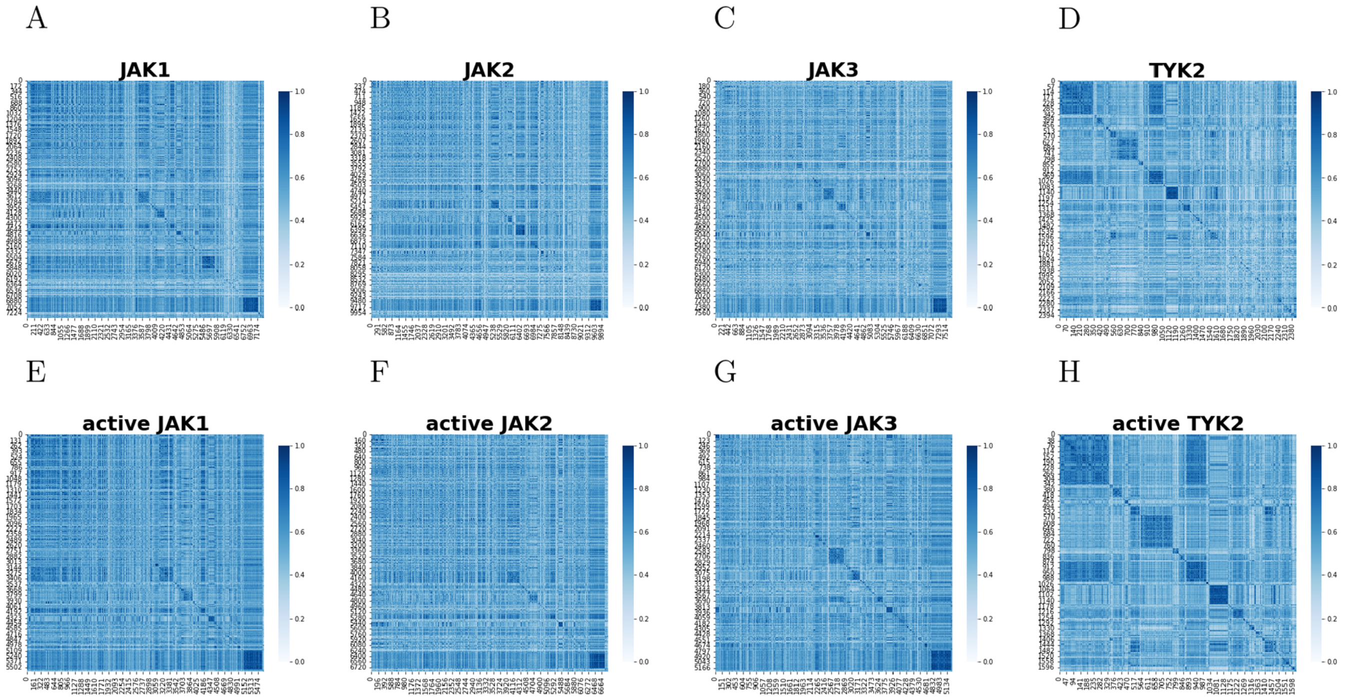


Figure 3. Data visualization based on Tanimoto similarity. Tanimoto similarity for all compounds in (A) JAK1, (B) JAK2, (C) JAK3, and (D) TYK2 data sets; Tanimoto similarity for (E) JAK1, (F) JAK2, (G) JAK3, and (H) TYK2 inhibitors.

$$\text{recall (SE)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{6}$$

Table 2. Results of Test Sets in JAK1, JAK2, JAK3, and TYK2 (Best Performances of Each Metric Are Shown in Bold)

target	model	acc	weighted acc	precision	recall	SP	F1	AUC	MCC	AP
JAK1	KNN	0.942	0.920	0.965	0.960	0.881	0.962	0.920	0.835	0.957
	SVM	0.958	0.939	0.972	0.974	0.905	0.973	0.974	0.880	0.990
	RF	0.954	0.932	0.969	0.972	0.893	0.970	0.986	0.868	0.996
	XGBoost	0.955	0.937	0.972	0.970	0.905	0.971	0.989	0.873	0.997
	CNN	0.744	0.720	0.887	0.765	0.674	0.821	0.765	0.392	0.888
	GraphVAE	0.902	0.924	0.988	0.884	0.964	0.933	0.948	0.770	0.986
	chemBERTa	0.957	0.938	0.972	0.972	0.905	0.972	0.989	0.877	0.997
	CoGT	**0.989**	**0.985**	**0.993**	**0.993**	**0.978**	**0.993**	**0.999**	**0.970**	**1.000**
JAK2	KNN	0.908	0.877	0.947	0.933	0.821	0.940	0.877	0.743	0.935
	SVM	0.923	0.893	0.952	0.947	0.839	0.950	0.943	0.782	0.981
	RF	0.924	0.896	0.954	0.947	0.845	0.951	0.948	0.786	0.979
	XGBoost	0.905	0.878	0.907	0.956	0.800	0.931	0.953	0.781	0.973
	CNN	0.668	0.623	0.747	0.760	0.486	0.753	0.646	0.248	0.748
	GraphVAE	0.901	0.900	0.948	0.902	0.899	0.924	0.965	0.783	0.981
	chemBERTa	0.896	0.887	0.930	0.913	0.860	0.922	0.950	0.766	0.973
	CoGT	**0.975**	**0.974**	**0.986**	**0.977**	**0.971**	**0.981**	**0.996**	**0.943**	**0.998**
JAK3	KNN	0.882	0.836	0.926	0.921	0.750	0.923	0.836	0.667	0.914
	SVM	0.879	0.836	0.927	0.916	0.756	0.921	0.912	0.662	0.972
	RF	0.878	0.824	0.920	0.923	0.726	0.921	0.926	0.652	0.978
	XGBoost	0.867	0.830	0.895	0.919	0.740	0.907	0.926	0.673	0.965
	CNN	0.696	0.500	0.696	1.000	0.000	0.821	0.503	N/A	0.699
	GraphVAE	0.894	0.889	0.946	0.901	0.877	0.923	0.956	0.755	0.972
	chemBERTa	0.875	0.849	0.912	0.910	0.789	0.911	0.943	0.698	0.976
	CoGT	**0.970**	**0.969**	**0.986**	**0.970**	**0.969**	**0.978**	**0.993**	**0.930**	**0.997**
TYK2	KNN	0.855	0.772	0.892	0.925	0.619	0.908	0.880	0.571	0.941
	SVM	0.866	0.833	0.931	0.893	0.774	0.911	0.893	0.638	0.957
	RF	0.882	0.808	0.907	0.944	0.673	0.925	0.923	0.651	0.967
	XGBoost	0.942	0.931	0.959	0.959	0.903	0.959	0.975	0.862	0.987
	CNN	0.718	0.593	0.716	0.960	0.225	0.820	0.733	0.289	0.812
	GraphVAE	0.951	0.945	0.970	0.959	0.931	0.965	0.977	0.883	0.991
	chemBERTa	0.926	0.891	0.923	0.977	0.806	0.949	0.981	0.819	0.993
	CoGT	**0.988**	**0.985**	**0.987**	**0.994**	**0.977**	**0.990**	**0.999**	**0.973**	**0.999**

$$\text{weighted accuracy} = \frac{\text{SE} + \text{SP}}{2} \tag{7}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FN} \times \text{FP}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FN})(\text{TN} + \text{FP})}} \tag{8}$$

$$F1 = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \tag{9}$$

$$\text{AP} = \sum_{n} (R_n - R_{n-1})P_n \tag{10}$$

RESULTS AND DISCUSSION

Chemical Diversity Analysis. To visualize the chemical diversity of our data sets, principal component analysis (PCA) was performed on all molecules collected with the MACCS fingerprint as input. If the features' relationship is nonlinear, PCA may not perform well to cluster data. Therefore, we also performed t-distributed stochastic neighbor embedding (t-SNE) to avoid PCA under fitting. Figure 2 shows that the distribution of active and inactive compounds for all 4 JAKs are still overlapped for both PCA and t-SNE, which suggests that in the chemical space active and inactive compounds could not be separated easily since some of them share similar structure and properties.

In addition, we did similarity quantification for all JAKs using Tanimoto similarity.[34] MACCS fingerprints of each drug in the data set were used to calculate Tanimoto similarity index. As shown in Figure 3A−D, the similarity for all molecules is relatively low in 4 JAKs, suggesting that molecules in our data set have a wide distribution with rather diverse structures. Furthermore, we examined active molecules for 4 JAKs, and the similarity is also not high overall, as shown in Figure 3E−H. These suggest that our data set is representative and our models have strong flexibility to identify active molecules from inactive ones.

Performance Evaluation and Comparison of Models. We summarized our models' evaluation in Table 2. Due to the imbalance of our data, we focused more on weighted accuracy than accuracy when evaluating the performance of different models. For all 4 types of JAKs, neural models significantly outperformed other conventional models in more than half of all metrics and performed comparable to the best performances on other metrics. This implies the potential ability of transformer-based and graph-based models to better grasp chemical structure information and be applied to downstream chemical tasks. The pretrained CNN model performed poorly compared with

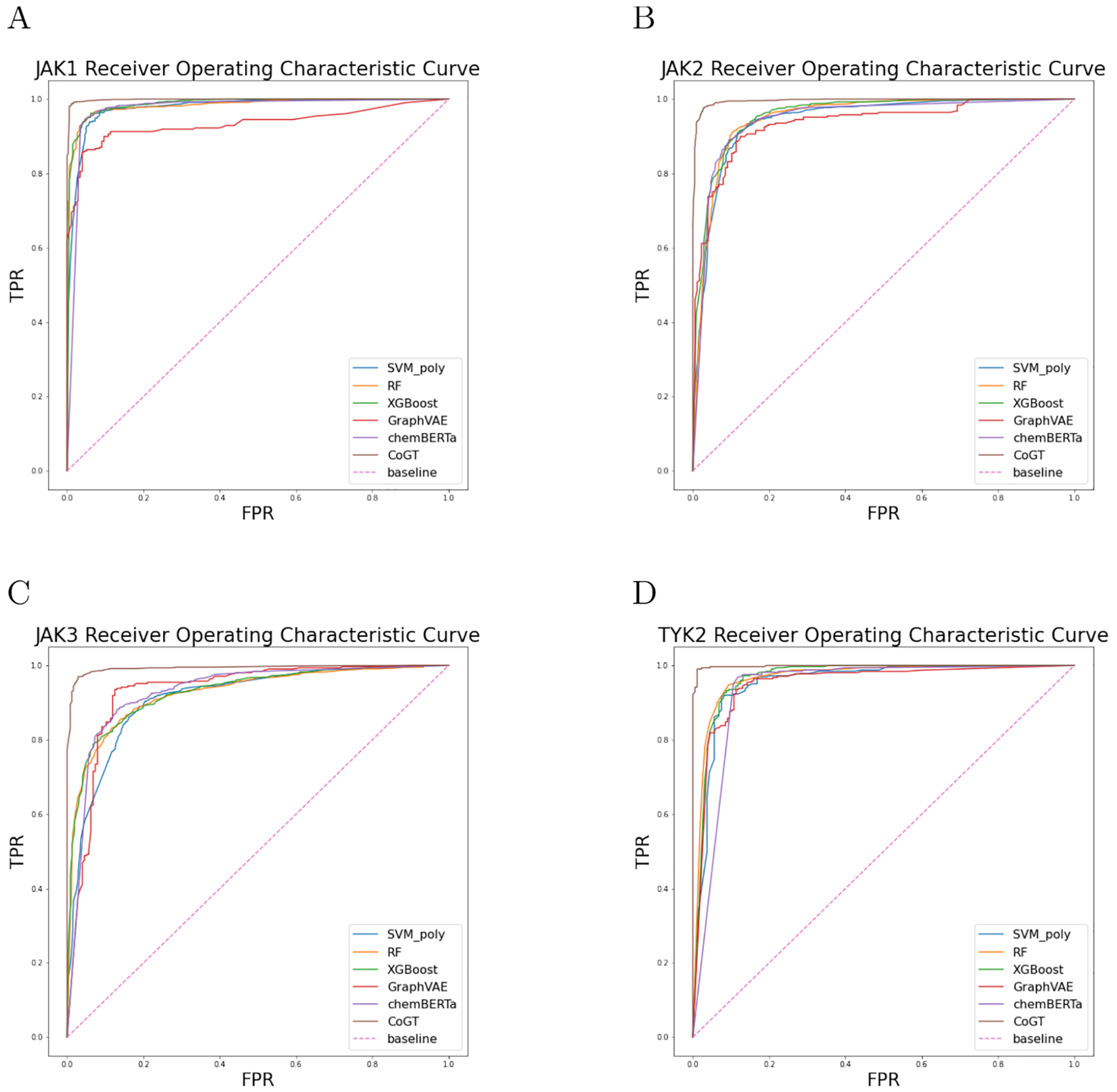


Figure 4. AUC−ROC curves of 5 selected models on (A) JAK1, (B) JAK2, (C) JAK3, and (D) TYK2 test sets.

chemBERTa and RGCN, which were also pretrained. This suggests that CNN is small to take advantage of a large volume of pretrained data and the lack of ability to extract helpful structural information from SMILE inputs.

Other than the graph-based model and the transformer-based model, traditional ML method SVM and tree-based models (RF and XGBoost) also performed well on JAK inhibition prediction. For instance, SVM performed well on JAK3 data sets and XGBoost showed impressive performance on TYK2. Compared with the base model, SVM, RF and XGBoost achieved high weighted accuracy on all 4 JAK prediction tasks and those 3 algorithms were chosen for later comodel training.

To fully utilize the advantages of all different models, the ensemble models CoGT were built on all 4 JAKs with MLP as second-level model via a stacking technique. The stacking technique builds a two-level model: the first level contains SVM, RF, XGBoost, GraphVAE, and chemBERTa to estimate a probability of a drug being a JAK inhibitor as an intermediate prediction, and the second level is a MLP which takes the prediction of three models in the first level as input to achieve a final prediction.[35] We visualized the normalized weight of the comodel in Figure 5. Each model weight and bias for each JAK can be found in Table S3. Our results show that CoGT performs impressively among all 4 JAK inhibition prediction tasks, scoring the highest for all metrics listed in Table 2. For AUC−ROC (receiver operating characteristic) curves on test sets shown in Figure 4, the comodel CoGT outperforms among all other models, with an AUC score nearly equal to 1 for all 4 JAKs.

As simple machine learning methods may already achieve similar performance compared with state-of-the-art ML

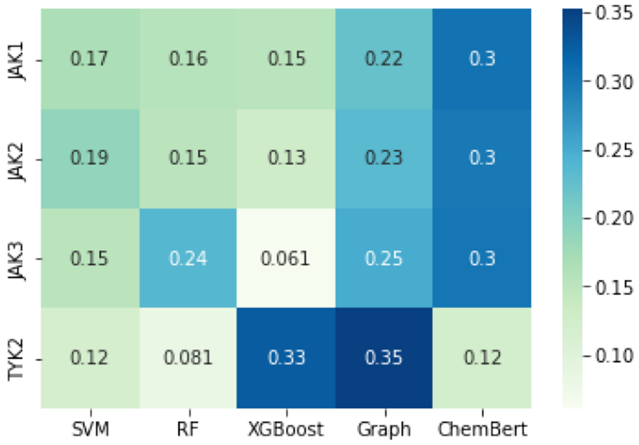


Figure 5. Normalized weights visualization for comodel CoGT.

methods,[33] we examined the performance of CoCM (comodel using conventional models only, i.e., SVM, RF, and XGBoost). Results shown in Table 3 indicate that our comodel CoGT exceeds the accuracy of conventional ML comodels. This demonstrates that incorporating models which utilize different ways of molecule representation could extract more information than simply using fingerprint-based conventional ML models.

We also examined the structure similarity between compounds that our model gave a wrong prediction. Tanimoto similarity did not reveal a common substructure between wrongly predicted molecules, as shown in Figure S2.

Comparison with Previous Models for JAK-Related ML Methods. There are several works on JAK inhibitor prediction including deep learning models and traditional learning models, as summarized in Table 4. Previous related work used XGBoost to predict JAK2 inhibition activity, and RF models were also utilized to predict JAK inhibition.[9,36] We also used XGBoost and RF in our model building, and our data showed that both models performed well compared with the base model, yet our comodel CoGT showed better performance on all metrics.

Graph-based model methods have recently emerged in the field of JAK inhibitor discovery.[15,37] These studies demonstrate the promising predicting power of graph-based models on JAK inhibitor discovery and design. As a consequence, we also included graph-based model graphVAE in our comodel building.

Table 4. Previous Work and Comparison with Ours

method	molecular representation	task category	target category
MTATFP[37]	graph	regression	JAK1, JAK2, JAK3, TYK2
MolGNN[15]	graph	classification	JAK1, JAK2, JAK3
RF[36]	fingerprint	classification	multiple kinases
XGBoost[9]	fingerprint	classification, regression	JAK2
CoGT (ours)	fingerprint, graph, SMILES string	classification	JAK1, JAK2, JAK3, TYK2

Several recent studies[38−40] have investigated using machine learning techniques to address JAK-related problems. However, these endeavors either involve a combination of the aforementioned methods or do not directly predict JAK types but rather do have an effect on the exploration of JAK inhibitors. Consequently, we did not incorporate them in our experiment comparisons.

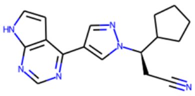
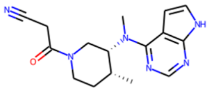
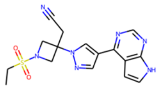
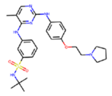
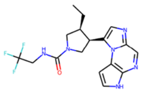
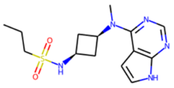
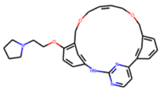
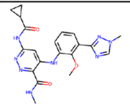
Overall, our comodel CoGT is the only model which tries to incorporate different representation information from a compound. Previous work mainly focused on using fingerprint or graph-based representation, neglecting the possibility that more information could be extracted through different representation. Here, we not only include fingerprint and graph representation but also utilize large pretrained transformer-based models to extract information directly from SMILES strings.

Comodel CoGT Prediction on Approved Drugs. To further validate our ensemble model CoGT, we used approved drugs and drugs in clinical trials as input and predicted their JAK inhibition. Results are summarized in Table 5. Our model prediction aligns well with real-world JAK inhibition for most drugs. For the four out of eight FDA-approved JAK inhibitors (i.e., Ruxolitinib, Tofacitinib, Baricitinib, Upadacitinib), our comodel gives accurate prediction on their inhibition profiles, which cannot be achieved by one single model. Besides, to further explore the phenomenon of the active cliff, which is the phenomenon that compounds with similar structures may have significant efficacy difference,[41] we examined our data set to see whether there are similar structures with FDA-approved drugs, and compounds highly similar to the approved drugs are shown in Figure 6. Results show that especially for JAK1−JAK3, there exists an active cliff in our data set, i.e., compounds highly similar

Table 3. Results of Test Sets in JAK1, JAK2, JAK3, and TYK2 on CoCM (Comodel Using Conventional Models, i.e., SVM, RF, XGBoost) and CoGT (Comodel Using Conventional, Graph, and Transformer-Based Models, i.e., SVM, RF, XGBoost, GraphVAE, chemBERTa)

target	model	acc	weighted acc	precision	recall	SP	F1	AUC	MCC	AP
JAK1	CoCM	0.982	0.975	0.988	0.989	0.961	0.988	0.998	0.952	0.999
	CoGT	**0.989**	**0.985**	**0.993**	**0.993**	**0.978**	**0.993**	**0.999**	**0.970**	**1.000**
JAK2	CoCM	0.966	0.958	0.971	0.979	0.937	0.975	0.993	0.921	0.996
	CoGT	**0.975**	**0.974**	**0.986**	**0.977**	**0.971**	**0.981**	**0.996**	**0.943**	**0.998**
JAK3	CoCM	0.949	0.949	0.977	0.948	0.951	0.962	0.987	0.884	0.994
	CoGT	**0.970**	**0.969**	**0.986**	**0.970**	**0.969**	**0.978**	**0.993**	**0.930**	**0.997**
TYK2	CoCM	0.977	0.972	0.975	0.990	0.955	0.982	0.998	0.951	0.999
	CoGT	**0.988**	**0.985**	**0.987**	**0.994**	**0.977**	**0.990**	**0.999**	**0.973**	**0.999**

Table 5. Probability as Inhibitors Based on CoGT Prediction and Their IC_{50} on FDA-Approved Drugs for 4 JAKs (Drugs Existing in the Training Sets Are Marked with *)[43−48]

Drug Name	Structure	$IC_{50}(\mu M)$ / Probability of Inhibitor			
		JAK1	JAK2	JAK3	TYK2
RUXOLITINIB[43]		6.4×10^{-3}	8.8×10^{-3}	0.487	0.0301
		0.931	0.896	0.885	0.834
TOFACITINIB[43]		0.0151*	0.0774	0.055	0.489
		0.905	0.689	0.899	0.848
BARICITINIB[43]		4×10^{-3}*	6.6×10^{-3}	0.787*	0.061
		0.941	0.643	0.895	0.83
FEDRATINIB[44]		0.105*	3×10^{-3}	1*	0.405
		0.951	0.911	0.103	0.841
UPADACITINIB[45]		0.047*	0.12*	2.304	4.69
		0.952	0.930	0.826	0.811
ABROCITINIB[46]		0.029*	0.803	>15	1.25
		0.958	0.094	0.108	0.840
PACRITINIB[47]		1.28	0.023*	0.52	0.05
		0.233	0.942	0.704	0.850
DEUCRAVACITINIB[48]		>10	>10	>10	$2\ \times10^{-4}$
		0.924	0.464	0.762	0.834

to approved drugs yet may have different levels of potency, and our model provides quite accurate prediction.

Among the remaining four approved inhibitors, the activity of two inhibitors with similar core structures (Fedratinib and Abrocitinib) is incorrectly predicted on JAK3 and JAK2, respectively. Our model also predicts that Pacritinib is a JAK1 noninhibitor based on the threshold of 10 μM. We can observe that the wrong predictions all happen at the values of ∼1 μM. Such discrepancy may be partially explained by the fact that most kinase-targeting small molecules are ATP-competitive inhibitors, and thus their measured IC_{50} can be greatly affected by the concurrent ATP concentration. For example, in one previous work elucidating the JAK2 binding sites of Fedratinib, the authors showed that the IC_{50} values of Fedratinib were measured to be 4.9 and 90 nM at the corresponding ATP concentrations of 10 and 100 μM, respectively.[42] Therefore, inhibitors with weaker activity (i.e., IC_{50} values closer to the set threshold) may exhibit opposing categorizations depending on the testing conditions. Given that the conditions applied in inhibition assays (i.e., ATP concentrations) can be slightly different across different research groups, while the collected IC_{50} values in the data sets do not necessarily include such information, a more consistent reporting format of IC_{50} values will be of valuable help in eliminating such uncertainty.

The remaining inhibitor Deucravacitinib shows least satisfactory prediction, where our model indicates it to be inhibitor on JAK1, JAK3, and TYK2, while it only inhibits TYK2. This discordance is most possible due to the unique incorporation of deuterium into the compound. Such a tiny replacement of three hydrogen atoms into isotope deuterium may not be universally incorporated in the available data sets, and thus the prediction accuracy suffers.[49]

To further analyze the 5 wrong predictions among all 32 predictions for approved drugs, we searched for a similar structure among wrongly predicted compounds in a separate data set, and most similar compounds are shown in Figure 7. Results show that there are compounds wrongly predicted whose structures are similar to FDA-approved drugs. Especially for the JAK1 data set, there is a compound highly similar to Pacritinib with Tanimoto similarity as high as 0.963. Thus, our model has difficulty giving accurate predictions for those moieties with high structural similarities, and more labeled compounds should be collected during the training process.

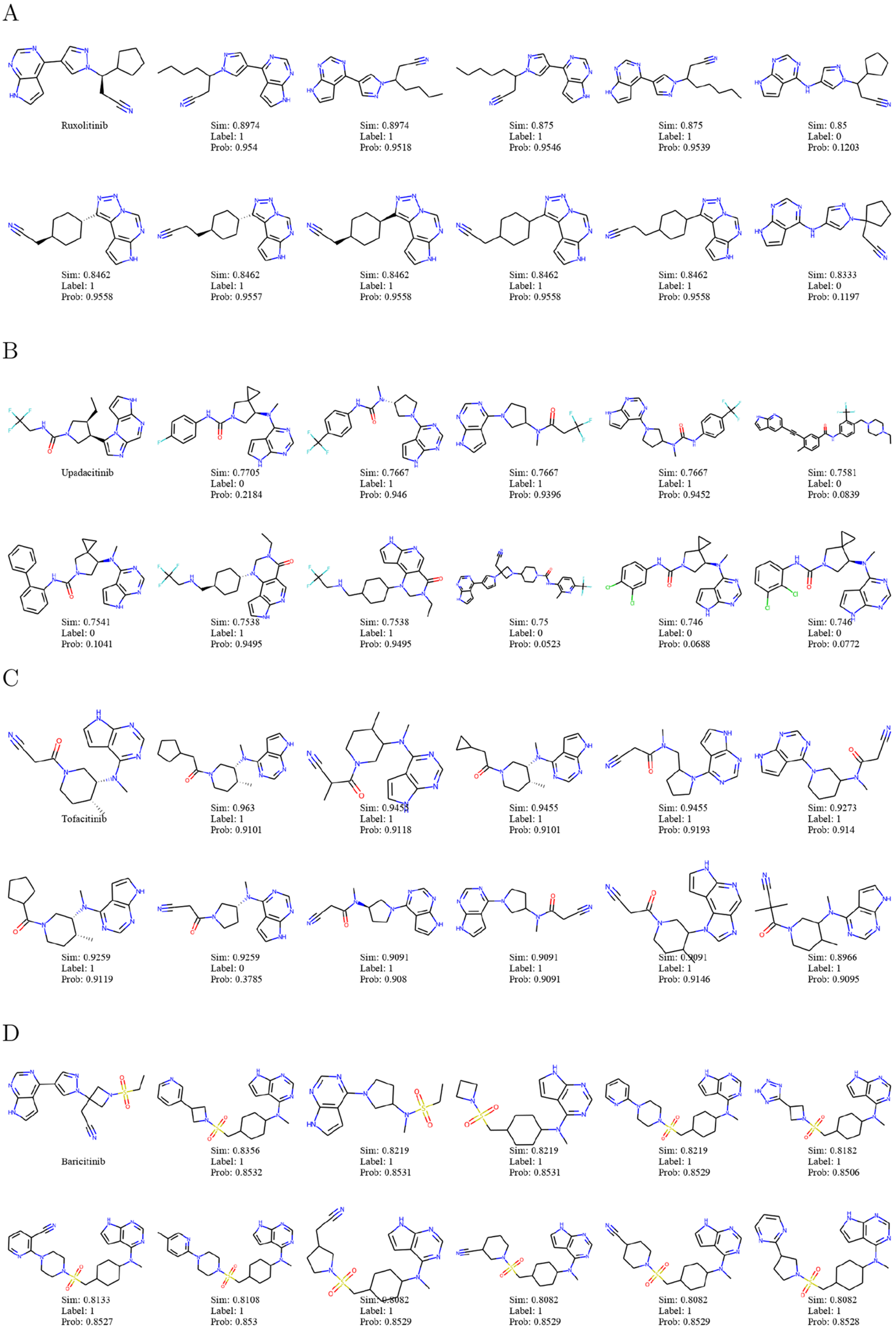

A
Ruxolitinib
Sim: 0.8974 Label: 1 Prob: 0.954
Sim: 0.8974 Label: 1 Prob: 0.9518
Sim: 0.875 Label: 1 Prob: 0.9546
Sim: 0.875 Label: 1 Prob: 0.9539
Sim: 0.85 Label: 0 Prob: 0.1203
Sim: 0.8462 Label: 1 Prob: 0.9558
Sim: 0.8462 Label: 1 Prob: 0.9557
Sim: 0.8462 Label: 1 Prob: 0.9558
Sim: 0.8462 Label: 1 Prob: 0.9558
Sim: 0.8462 Label: 1 Prob: 0.9558
Sim: 0.8333 Label: 0 Prob: 0.1197
B
Upadacitinib
Sim: 0.7705 Label: 0 Prob: 0.2184
Sim: 0.7667 Label: 1 Prob: 0.946
Sim: 0.7667 Label: 1 Prob: 0.9396
Sim: 0.7667 Label: 1 Prob: 0.9452
Sim: 0.7581 Label: 0 Prob: 0.0839
Sim: 0.7541 Label: 0 Prob: 0.1041
Sim: 0.7538 Label: 1 Prob: 0.9495
Sim: 0.7538 Label: 1 Prob: 0.9495
Sim: 0.75 Label: 0 Prob: 0.0523
Sim: 0.746 Label: 0 Prob: 0.0688
Sim: 0.746 Label: 0 Prob: 0.0772
C
Tofacitinib
Sim: 0.963 Label: 1 Prob: 0.9101
Sim: 0.9455 Label: 1 Prob: 0.9118
Sim: 0.9455 Label: 1 Prob: 0.9101
Sim: 0.9455 Label: 1 Prob: 0.9193
Sim: 0.9273 Label: 1 Prob: 0.914
Sim: 0.9259 Label: 1 Prob: 0.9119
Sim: 0.9259 Label: 0 Prob: 0.3785
Sim: 0.9091 Label: 1 Prob: 0.908
Sim: 0.9091 Label: 1 Prob: 0.9091
Sim: 0.9091 Label: 1 Prob: 0.9146
Sim: 0.8966 Label: 1 Prob: 0.9095
D
Baricitinib
Sim: 0.8356 Label: 1 Prob: 0.8532
Sim: 0.8219 Label: 1 Prob: 0.8531
Sim: 0.8219 Label: 1 Prob: 0.8531
Sim: 0.8219 Label: 1 Prob: 0.8529
Sim: 0.8182 Label: 1 Prob: 0.8506
Sim: 0.8133 Label: 1 Prob: 0.8527
Sim: 0.8108 Label: 1 Prob: 0.853
Sim: 0.8082 Label: 1 Prob: 0.8529
Sim: 0.8082 Label: 1 Prob: 0.8529
Sim: 0.8082 Label: 1 Prob: 0.8529
Sim: 0.8082 Label: 1 Prob: 0.8528

Figure 6. Compounds with high structure similarities compared to FDA-approved drugs. Panels (A–D) each represents the grouped compounds with high structure similarity to Ruxolitinib, Upadacitinib, Tofacitinib, and Baricitinib, respectively. Sim, Tanimoto similarity; Label, true label; Prob, model predicting probability.

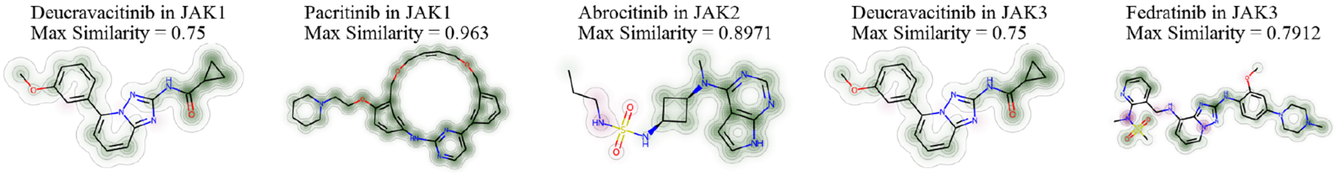


Figure 7. Similarity visualization between wrongly predicted molecules in Table 5 and the most similar drug compound in data set.

CONCLUSIONS

In this research, we developed an ensemble model, called CoGT, which combined multiple machine learning models to achieve better accuracy than any individual model in predicting DTI for four JAK isoforms. We first compiled a comprehensive data set for JAK inhibitors. Using this data set, we compared different ML methods in predicting JAK inhibition, which included a graph-based model (RGCN applied GraphVAE), a pretrained RoBERTa model (chemBERTa), and traditional machine learning models. Our experiments revealed that the graph model was superior to conventional ML methods to effectively extract structural information for all JAK inhibitors. In addition, large pretrained transformer-based model chemBERTa could also be effective for chemical predictions of these JAK inhibitory structures. Traditional models such as SVM, RF, and XGBoost performed well, despite their relatively low computational costs. By fully leveraging the strengths of various models, our ensemble mode CoGT performed best with prediction accuracy of DTI of JAK inhibitors. Further improvement can be achieved by optimizing parameters in GraphVAE model for a better description of the bond and atom types, as well as by utilizing the distribution of chemical fragments in the data set for the model training.

ASSOCIATED CONTENT

Data Availability Statement

Data sets, code, python package version are available at https://github.com/yingzibu/JAK_ML. Other data are available from the corresponding authors with reasonable request.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c00160.

> Details of SVM results and XGBoost parameters, the stacking parameters, the evaluation on the validation set, the Tanimoto similarity visualization of wrongly predicted molecules by CoGT (PDF)

AUTHOR INFORMATION

Corresponding Author

Duxin Sun − *Department of Pharmaceutical Sciences, College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109, United States;* orcid.org/0000-0002-6406-2126; Email: duxins@umich.edu

Authors

Yingzi Bu − *Department of Pharmaceutical Sciences, College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109, United States;* orcid.org/0000-0002-2600-8946

Ruoxi Gao − *Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109, United States*

Bohan Zhang − *School of Information, University of Michigan, Ann Arbor, Michigan 48109, United States*

Luchen Zhang − *Department of Pharmaceutical Sciences, College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109, United States;* orcid.org/0000-0001-9163-9990

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c00160

Author Contributions

Y.B., R.G., and B.Z. contributed equally to this work. D.S. directed this study. Y.B., R.G., and B.Z. contributed equally to design the experiment and run the models. L.Z. performed data analysis. All authors have contributed on writing and reviewing the manuscript, and have given approval to the final version.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported in part through computational resources and services provided by Advanced Research Computing (ARC), a division of Information and Technology Services (ITS) at the University of Michigan, Ann Arbor.

REFERENCES

(1) O'Shea, J. J.; Schwartz, D. M.; Villarino, A. V.; Gadina, M.; McInnes, I. B.; Laurence, A. The JAK-STAT pathway: impact on human disease and therapeutic intervention. *Annual review of medicine* **2015**, *66*, 311−328.

(2) Salas, A.; Hernandez-Rocha, C.; Duijvestein, M.; Faubion, W.; McGovern, D.; Vermeire, S.; Vetrano, S.; Vande Casteele, N. JAK−STAT pathway targeting for the treatment of inflammatory bowel disease. *Nature Reviews Gastroenterology & Hepatology* **2020**, *17*, 323−337.

(3) Villarino, A. V.; Kanno, Y.; O'Shea, J. J. Mechanisms and consequences of Jak−STAT signaling in the immune system. *Nature immunology* **2017**, *18*, 374−384.

(4) Hu, X.; Li, J.; Fu, M.; Zhao, X.; Wang, W. The JAK/STAT signaling pathway: From bench to clinic. *Signal Transduction and Targeted Therapy* **2021**, *6*, 402.

(5) Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5−32.

(6) Mavroforakis, M.; Theodoridis, S. A geometric approach to Support Vector Machine (SVM) classification. *IEEE Transactions on Neural Networks* **2006**, *17*, 671−682.

(7) Peterson, L. E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883.

(8) Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T.; et al. *Xgboost: extreme gradient boosting*. R package version 0.4-2, 2015; Vol. *1*, pp 1−4.

(9) Yang, M.; Tao, B.; Chen, C.; Jia, W.; Sun, S.; Zhang, T.; Wang, X. Machine Learning Models Based on Molecular Fingerprints and an Extreme Gradient Boosting Method Lead to the Discovery of JAK2 Inhibitors. *J. Chem. Inf. Model.* **2019**, *59*, 5002.

(10) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research* **2012**, *40*, D1100−D1107.

(11) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte,

E.; et al. The ChEMBL database in 2017. *Nucleic acids research* **2017**, *45*, D945−D954.

(12) Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic acids research* **2016**, *44*, D1045−D1053.

(13) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Bryant, S. H. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic acids research* **2009**, *37*, W623−W633.

(14) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2019 update: improved access to chemical data. *Nucleic acids research* **2019**, *47*, D1102−D1109.

(15) Liu, Y.; Wu, Y.; Shen, X.; Xie, L. COVID-19 multi-targeted drug repurposing using few-shot learning. *Frontiers in Bioinformatics* **2021**, *1*, DOI: 10.3389/fbinf.2021.693177.

(16) Kim, S.; Thiessen, P.; Bolton, E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. PubChem Substance and Compound databases. *Nucleic acids research* **2016**, *44*, D1202−D1213.

(17) Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016; pp 785−794.

(18) Kipf, T. N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, https://arxiv.org/abs/1609.02907.

(19) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, https://arxiv.org/abs/1710.10903.

(20) Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How powerful are graph neural networks? *arXiv* **2018**, https://arxiv.org/abs/1810.00826.

(21) Zhang, Z.; Chen, L.; Zhong, F.; Wang, D.; Jiang, J.; Zhang, S.; Jiang, H.; Zheng, M.; Li, X. Graph neural network approaches for drug-target interactions. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102327.

(22) Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Berg, R. v. d.; Titov, I.; Welling, M. Modeling relational data with graph convolutional networks. *European semantic web conference*, 2018; pp 593−607.

(23) Kingma, D. P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, https://arxiv.org/abs/1312.6114.

(24) Kipf, T. N.; Welling, M. Variational graph auto-encoders. *arXiv* **2016**, https://arxiv.org/abs/1611.07308.

(25) Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, https://arxiv.org/abs/1810.04805.

(26) Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, https://arxiv.org/abs/1804.02767.

(27) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.

(28) Sun, C.; Ohodnicki, P. R.; Stewart, E. M. Chemical sensing strategies for real-time monitoring of transformer oil: A review. *IEEE Sensors Journal* **2017**, *17*, 5786−5806.

(29) Chithrananda, S.; Grand, G.; Ramsundar, B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. *arXiv* **2020**, https://arxiv.org/abs/2010.09885.

(30) Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, https://arxiv.org/abs/1907.11692.

(31) Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science* **2019**, *5*, 1572−1583.

(32) Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, https://arxiv.org/abs/1711.05101.

(33) Janela, T.; Bajorath, J. Simple nearest-neighbour analysis meets the accuracy of compound potency predictions using complex machine learning models. *Nature Machine Intelligence* **2022**, *4*, 1246−1255.

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

(35) Wolpert, D. H. Stacked generalization. *Neural networks* **1992**, *5*, 241−259.

(36) Cooper, K.; Baddeley, C.; French, B.; Gibson, K.; Golden, J.; Lee, T.; Pierre, S.; Weiss, B.; Yang, J. Novel development of predictive feature fingerprints to identify chemistry-based features for the effective drug design of sars-cov-2 target antagonists and inhibitors using machine learning. *ACS omega* **2021**, *6*, 4857−4877.

(37) Wang, Y.; Gu, Y.; Lou, C.; Gong, Y.; Wu, Z.; Li, W.; Tang, Y.; Liu, G. A multitask GNN-based interpretable model for discovery of selective JAK inhibitors. *Journal of cheminformatics* **2022**, *14*, 16.

(38) Yang, Z.; Tian, Y.; Kong, Y.; Zhu, Y.; Yan, A. Classification of JAK1 Inhibitors and SAR Research by Machine Learning Methods. *Artificial Intelligence in the Life Sciences* **2022**, *2*, 100039.

(39) Rodriguez, S.; Hug, C.; Todorov, P.; Moret, N.; Boswell, S. A.; Evans, K.; Zhou, G.; Johnson, N. T.; Hyman, B. T.; Sorger, P. K.; et al. Machine learning identifies candidates for drug repurposing in Alzheimer's disease. *Nat. Commun.* **2021**, *12*, 1033.

(40) Faquetti, M. L.; Grisoni, F.; Schneider, P.; Schneider, G.; Burden, A. M. Identification of novel off targets of baricitinib and tofacitinib by machine learning with a focus on thrombosis and viral infection. *Sci. Rep.* **2022**, *12*, 7843.

(41) Stumpfe, D.; Hu, Y.; Dimova, D.; Bajorath, J. Recent progress in understanding activity cliffs and their utility in medicinal chemistry: miniperspective. *Journal of medicinal chemistry* **2014**, *57*, 18−28.

(42) Kesarwani, M.; Huber, E.; Kincaid, Z.; Evelyn, C. R.; Biesiada, J.; Rance, M.; Thapa, M. B.; Shah, N. P.; Meller, J.; Zheng, Y.; et al. Targeting substrate-site in Jak2 kinase prevents emergence of genetic resistance. *Sci. Rep.* **2015**, *5*, 14538.

(43) Clark, J. D.; Flanagan, M. E.; Telliez, J.-B. Discovery and development of Janus Kinase (JAK) inhibitors for inflammatory diseases: Miniperspective. *Journal of medicinal chemistry* **2014**, *57*, 5023−5038.

(44) Wernig, G.; Kharas, M. G.; Okabe, R.; Moore, S. A.; Leeman, D. S.; Cullen, D. E.; Gozo, M.; McDowell, E. P.; Levine, R. L.; Doukas, J.; et al. Efficacy of TG101348, a selective JAK2 inhibitor, in treatment of a murine model of JAK2V617F-induced polycythemia vera. *Cancer cell* **2008**, *13*, 311−320.

(45) Parmentier, J. M.; Voss, J.; Graff, C.; Schwartz, A.; Argiriadi, M.; Friedman, M.; Camp, H. S.; Padley, R. J.; George, J. S.; Hyland, D.; et al. In vitro and in vivo characterization of the JAK1 selectivity of upadacitinib (ABT-494). *BMC rheumatology* **2018**, *2*, 23.

(46) Xu, H.; Jesson, M. I.; Seneviratne, U. I.; Lin, T. H.; Sharif, M. N.; Xue, L.; Nguyen, C.; Everley, R. A.; Trujillo, J. I.; Johnson, D. S.; et al. PF-06651600, a dual JAK3/TEC family kinase inhibitor. *ACS Chem. Biol.* **2019**, *14*, 1235−1242.

(47) William, A. D.; Lee, A. C.-H.; Blanchard, S.; Poulsen, A.; Teo, E. L.; Nagaraj, H.; Tan, E.; Chen, D.; Williams, M.; Sun, E. T.; et al. Discovery of the Macrocycle 11-(2-Pyrrolidin-1-yl-ethoxy)-14, 19-dioxa-5, 7, 26-triaza-tetracyclo [19.3. 1.1 (2, 6). 1 (8, 12)] heptacosa-1 (25), 2 (26), 3, 5, 8, 10, 12 (27), 16, 21, 23-decaene (SB1518), a Potent Janus Kinase 2/Fms-Like Tyrosine Kinase-3 (JAK2/FLT3) Inhibitor for the Treatment of Myelofibrosis and Lymphoma. *Journal of medicinal chemistry* **2011**, *54*, 4638−4658.

(48) Wrobleski, S. T.; Moslin, R.; Lin, S.; Zhang, Y.; Spergel, S.; Kempson, J.; Tokarski, J. S.; Strnad, J.; Zupa-Fernandez, A.; Cheng, L.; et al. Highly selective inhibition of tyrosine kinase 2 (TYK2) for the treatment of autoimmune diseases: discovery of the allosteric inhibitor BMS-986165. *J. Med. Chem.* **2019**, *62*, 8973−8995.

(49) Mullard, A. First de novo deuterated drug poised for approval. *Nature reviews. Drug Discovery* **2022**, *21*, 623.