# RT-Tranformer: Retention Time Prediction for Metabolite Annotation to Assist in Metabolite Identification

Jun Xue,[†] Weihua Li,[*,†] and Bingyi Wang[‡]

†*School of Information Science and Engineering, Yunnan University, Kunming, China*

‡*Faculty of Drug Control, Yunnan Police College, Kunming, China*

E-mail: liweihua@ynu.edu.cn

**Abstract**

Liquid chromatography retention times (RTs) prediction can assist in metabolite identification, which is a critical task and challenge in non-targeted metabolomics. However, different chromatographic methods (CM) may result in different RTs for the same metabolite. Current RT prediction methods lack sufficient scalability to transfer from one specific chromatographic method to another. Therefore, we present RT-Transformer, a novel deep neural network model coupled with 1D-Transformer and graph attention network (GAT) that can predict RTs under any chromatographic methods. First, we obtain a pre-trained model by training RT-Transformer on the large small molecule retention time (SMRT) dataset containing 80038 molecules, and then project the resulting model onto different chromatographic methods based on transfer learning. When tested on the METLIN dataset, as other authors did, the average absolute error reached 27.3 after removing samples with retention times fewer than five minutes. Still, it reached 33.5 when no samples were removed. The pre-trained

RT-Transformer was further transferred to 5 datasets corresponding to different chromatographic conditions and fine-tuned. According to the experimental results, RT-Transformer achieves competitive performance compared to state-of-the-art methods. In addition, RT-Transformer was applied to 30 external molecular RT datasets. Extensive evaluations indicate that RT-Transformer has excellent scalability in predicting RTs for liquid chromatography and improves the accuracy of metabolite identification.

# Introduction

Metabolomics systematically identifies and quantifies all metabolites in a given organism or biological sample.[1] Metabolite annotation and identification is the main bottleneck in untargeted metabolomics.[2–5] Liquid chromatography-mass spectrometry (LC-MS) has become the most widely used method for metabolite identification because of its enhanced resolution and excellent sensitivity.[6,7] Typically, the original liquid chromatography (LC) data consists of hundreds of original mass spectra. Therefore, various approaches[8–16] have been developed to identify metabolites by searching against structural databases, such as PubChem[17] and ChemSpider.[18] Unfortunately, all of these approaches return multiple candidates with similar structures. To reduce the cost of the experiment, it's necessary to filter out as many false candidates as possible. Previous studies have shown that retention times(RTs) obtained by chromatographic separation would enable filtering candidates with similar spectra but different RTs, further facilitating the identification of metabolites.

Experimental methods for obtaining RTs are costly, so CM-specific datasets only contain a tiny portion of known compounds. To predict compounds that lack experimental RTs, numerous researchers have developed various RT prediction methods.[19–26] Traditional machine learning approaches, such as multiple linear regression, random forest, support vector machine, and gradient boosting, are frequently employed for RT prediction.[27–32] For example, Bouwmeester et al.[33] compared different ML methods for RT prediction and found an ensemble of multiple ML-based models performed optimally. Recently, METLIN small molecule

retention time (SMRT) dataset[34] of 80038 molecules were released to the public, which stimulating deep learning-based RT prediction methods, such as DLM,[34] DNNpwa,[35]and 1D-CNN.[36] More deep learning-based methods, such as GNN-RT,[37] CPORT,[38] MPNN,[39] and Blender,[40] apply transfer learning[41] to predict the retention times in specific chromatographic separation systems. These methods alleviate the limitation of small training data by pre-training the neural networks on SMRT and further reusing some parameters in the pre-trained networks. More specifically, GNN-RT[37] and MPNN[39] exploit graph neural networks (GNN) to learn effective molecular representations from the structures of small molecules, improving the transferability of the models.

These methods have been developed and have made significant progress in RT prediction. However, RTs result from the combination of the metabolite with the chromatographic method(CM) used, so the retention time of the same molecule can vary from CM to CM. Thus, the models developed using the SMRT dataset still need more scalability to predict the RT of a molecule in other CMs. Besides, most current methods heavily rely on molecular fingerprints or molecular descriptors, while neglecting node and edge attributes in molecular structures, resulting in the inability to learn effective molecular representations. It is still being determined how to exploit and combine molecular fingerprints and structures to predict RTs better. Third, recent methods, such as GNN-RT, attempt to use GNNs to embed a molecule structure as a fixed feature vector for RTs prediction. Nonetheless, GNNs tend to treat atoms and chemical bonds equally, making models fail to capture inter-atomic correlations effectively. Graph attention networks(GAT),[42] one of the most popular GNN architectures, employ the attention mechanism[43] to update the attributes of every node and are insensitive to the selection order of neighbors. It has been shown that GAT can better exploit the graph structure, node information, and edge information, and obtain their representations in low-dimensional space.

Guided by these, we propose a new deep neural network model, RT-Transformer, coupled with 1D-Transformer is a model we developed based on the Transformer Encoder[44] for

processing one-dimensional data and GAT, to learn the effective molecular representations from molecular graphs and molecular fingerprints for RT prediction. We train the model using the SMRT dataset, freeze the feature extraction layer of the resulting model, and further fine-tune the model on other datasets for prediction. The results show that our model significantly outperforms the previous methods. In addition, the pre-trained model on SMRT dataset is evaluated with 30 external RT datasets obtained from PredRet.[20] Extensive evaluations demonstrate that the RT-transformer has excellent and robust scalability.

# Methods

## Preparation of the SMRT Dataset

The SMRT dataset from the METLIN library was released by Domingo et al.[34] It provides RT data for 80038 small molecules, including metabolites, natural products, and drug-like small compounds, all obtained using RP chromatography and HPLC-MS. The majority of molecules can be categorized into seven groups, including organic heterocyclic compounds (63.9%), benzene compounds (24.7%), organic acids and derivatives (6.6%), organic nitrogen compounds (1.6%), organic oxygen compounds (1.18%), organic sulfur compounds (0.66%), and other compounds (1-25%). Furthermore, other compounds consist of lipids, lignans, nucleosides, nucleotides, phenylpropane, and polyketides. SMRT provides RTs in seconds, PubChem molecule numbers, molecular structure data in SDF format, as well as several chemical descriptors and extended connectivity fingerprints ECFP. The experimental RTs for 80038 small molecules range from 0.3 to 1471.7 seconds. In previous studies, new molecules were omitted, and samples with retention periods longer than 300 seconds were utilized for training. These retained molecules have a relative atomic mass of 113.08-741.46 Da and mainly comprise metabolites, natural products, and tiny compounds resembling drugs. The relative atomic mass in all available data sets varies from 104.18 to 741.46 Da. We obtained two pre-trained models on the data sets of retained molecules and SMRT, respectively.

The two-pertained models were then transferred to several datasets, and their effects were compared.

## Molecular Graph Data

We transform molecules into graphs, with the nodes and edges representing the atoms and chemical bonds in the molecule, respectively. The node features consist of chirality, relative atomic mass, degree, formal charge, orbital hybrid mode, valence, radical electrons, whether or not it is on a ring, etc. The edge attributes include the type of the bond, whether it is a ring, aromatic, or conjugated. All the type attributes mentioned above are transformed into one-hot vectors concatenated with the value attribute. Besides, for each chemical bond in the molecular graphs, we convert it into a bi-directional edge. This paper uses Python package RDKit(www.rdkit.org)[45] to generate molecular graphs with 34 features per node and 5 features per bond.

## Overview of RT-Transformers

RT-Transformers take the resulting molecular graphs and Morgan fingerprints from InChl (International Chemical Identifier) as input, and extract the features using a multi-head GAT and a stacked 1D-transformer, respectively. Then, the obtained features are fused and fed into a linear layer to produce a vector representation for RT prediction. An overview of RT-transformer is illustrated in Figure 1.

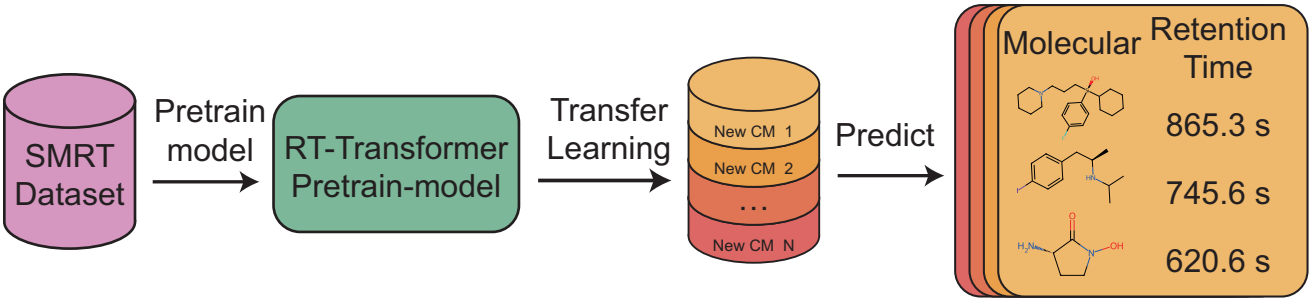


Figure 1: Procedure of RT prediction by RT-Transformer

## ResGAT

Graph Attention Networks(GAT) were proposed by Veličković et al.[42] to learn graph-structured data based on attention mechanism.[43] To learn the embedding vectors of the molecular graph, we devise a ResGAT block based on a three-head GAT with a residual connection. And, the ResGAT block uses a linear layer as an aggregation function rather than an addition or concatenation function to accelerate training. This enables any node to aggregate information from all other nodes. Then, the node features are updated by skip connections of the non-updated features. At Last, we apply layer normalization[46] to all node features.

$$y = \frac{x - \mathrm{E}[x]}{\sqrt{\mathrm{Var}[x] + \epsilon}} * \gamma + \beta$$

The mean and standard deviation are calculated over the last $D$ dimensions, where $D$ is the dimension of the inputs. $\gamma$ and $\beta$ are learnable affine transform parameters.

The input of this block is node features $h = \left\{\overrightarrow{h_1}, \overrightarrow{h_2}, \ldots, \overrightarrow{h_N}\right\}$ and edge features $d = \left\{\overrightarrow{d_1}, \overrightarrow{d_2}, \ldots, \overrightarrow{d_M}\right\}$, where $\overrightarrow{h_i} \in R^F$, $\overrightarrow{d_i} \in R^G$, $N$ and $M$ are the numbers of nodes and edges, respectively; $F$ and $G$ are the dimensions of a node feature and an edge feature, respectively. For any two nodes, their features $\vec{h}_i$ and $\vec{h}_j$ and the features of the bond between them $\vec{d}_{ij}$ are transformed into a vector $e_{ij}$ and then the attention coefficient $a_{ij}$ between these two nodes is calculated as follows :

$$e_{ij} = \mathit{Attention}\left(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j, \mathbf{W}\vec{d}_{ij}\right)$$

$$a_{ij} = \mathrm{Softmax}_j\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{k \in N_i} \exp\left(e_{ik}\right)}$$

where $\mathbf{W}$ is a weight matrix; *Attention* is a function of attention mechanism. The specific

implementation of this function, *Attention*, is as follows:

$$a_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\overrightarrow{\alpha}^{\top}\left[\mathbf{W}\vec{h}_i \middle\| \mathbf{W}\vec{h}_j \middle\| \mathbf{W}\vec{d}_{ij}\right]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(\overrightarrow{\alpha}^{\top}\left[\mathbf{W}\vec{h}_i \middle\| \mathbf{W}\vec{h}_k \middle\| \mathbf{W}\vec{d}_{ik}\right]\right)\right)}$$

where $\alpha$ is a $2N + M$ dimensional vector; $\|$ is the concatenation operation; LeakyReLU is an activation function as follows:

$$\text{LeakyReLU(x)} = \begin{cases} \text{x}, \text{x} > 0 \\ \lambda\text{x}, \text{x} \leq 0 \end{cases}$$

where $\lambda$ is 0.0001. After getting the attention coefficient, we could calculate the final output features of every node (after potentially applying a nonlinearity).

The updated eigenvectors of node $i$ are as follows:

$$\vec{h}'_i = \sigma\left(\sum_{j\in\mathcal{N}_i} a_{ij}\mathbf{W}\vec{h}_j\right)$$

Because this block uses multi-head attention, we concentrate all vectors generated by all heads as follows:

$$\vec{h}'_i = \text{concat}\left(\sigma\left(\sum_{j\in N_i}\alpha_{ij}^{k}W^{k}h_j\right)\right)$$

The detail of ResGAT is shown in Figure 2.

## 1D-Transformer

The 1D-Transformer is based on transformer architecture.[44] Transformer, an encoder–decoder architecture, is proposed for machine translation tasks and has achieved state-of-art performance in many deep learning areas such as computer vision, and natural language processing. The 1D-Transformer is a combination of attention layers and feed-forward layers. The attention layer exploits the scaled-dot attention mechanism to capture the features most relevant

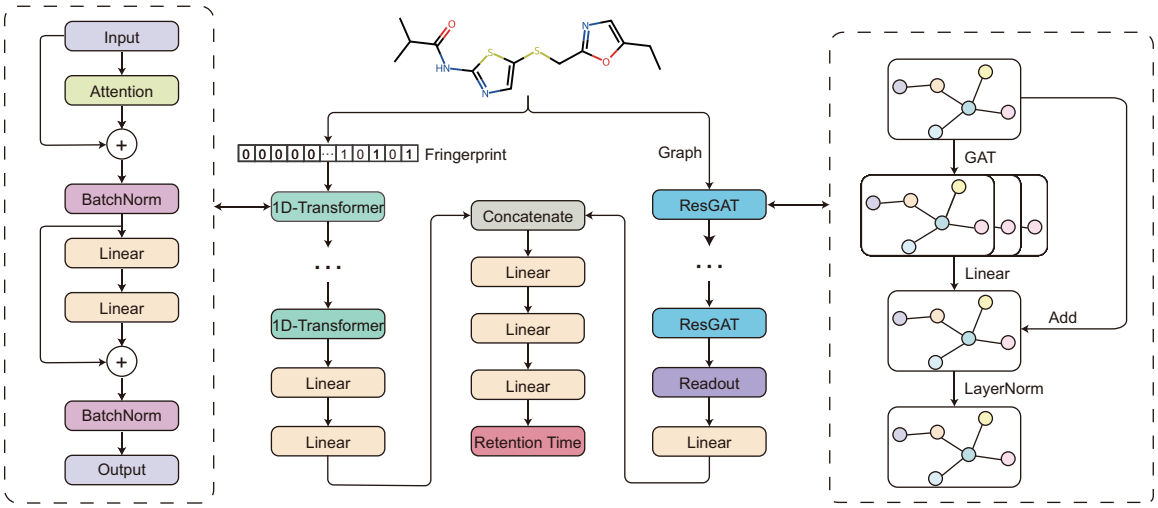


Figure 2: Structure of RT-Transformer

to RTs, and takes three inputs, i.e., the keys $\mathbf{K}$, the queries $\mathbf{Q}$, and the values $\mathbf{V}$. To make the attention layer more robust, we add a trainable matrix $\mathbf{W}$ and compute the attention as follows.

$$\text{Attention}\left(\mathbf{Q}, \mathbf{K}, \mathbf{V}\right) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d}}\right)\mathbf{V}\mathbf{W}$$

The dot product of $\mathbf{Q}$ and $\mathbf{K}$ computes how closely the keys are aligned with the queries. If the query and key are aligned, their dot product will be big, and vice versa. Each key has a value vector multiplied by the softmax output, normalizing the dot products and emphasizing the greatest components. $d$ is a scaling factor that changes based on the layer size. We use it as self-attention mechanism, so $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ are all the input vectors $\vec{f} \in R^d$, where $d$ is the number of the input features. The feed-forward layers are composed of 2 linear layers. The detail of 1D-Transformer is shown in Figure 2.

## Output-Block

Output-Block receives molecular graphs output by ResGATs and fingerprints proceed by 1D-Transformers. We used an addition function to readout the molecular graph, ignoring the features of bonds on the graph. The readout layer adopts a linear layer with a 512-dimensional input channel and a 512-dimensional output channel so that the features produced by Res-

GAT are suitable for feature fusion. Then, two linear layers are used to reduce the dimension of the fused features. The features extracted by ResGAT and 1D-Transformer have been resized to 512 dimensions by several linear layers. A 1024-dimensional vector concatenating these features goes through 3 linear layers and produces a vector for RTs prediction. All the linear layers are activated by Rectified Linear Unit function(ReLU) as follows:

$$\mathrm{ReLU(x)} = \begin{cases} \mathrm{x}, \mathrm{x} > 0 \\ 0, \mathrm{x} \leq 0 \end{cases}$$

## RT-transformer

RT-Transformers could be divided into three modules: ResGAT, 1D-Transformer, and Output-Block. A linear layer embeds the node features into 512 dimensions to get the higher dimension relationship. Graph-Transformers receive molecular graphs as input. By stacking 9 ResGAT Blocks, every atom could get information from other atoms and chemical bonds. At the end of Graph-Transformers, we use an addtion function to readout all the atom features into the molecular graph features. 1D-Transformer receives a molecular fingerprint, a 2048-dimensional vector, as its input, and produces a 2048-dimensional vector as its output. We stack 12 1D-Transformer blocks to process the fingerprint feature. After processing these features, we send them to Output-Block to get the final prediction.

## Evaluation Metrics

We evaluate the model's performance with MAE, MRE, MedAE, MedRE, and $R^2$. The calculation formulas are as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i}^{N}|y_i - \hat{y}_i|$$

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i|}$$

$$\text{MedAE} = \text{median}(|y_i - \hat{y}_i|)$$

$$\text{MedRE} = \text{median}(\frac{|y_i - \hat{y}_i|}{|y_i|})$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

### Experimental Setups

The SMRT dataset employs a random partitioning strategy whereby molecules are allocated to non-overlapping training, validation, and test sets in the proportions of 80%, 10%, and 10%, respectively. During training, the model seeks to minimize L1 loss in the validation set for a maximum of 300 epochs. The RT-transformer is trained through backpropagation and optimized using AdamW with an initial learning rate of 0.0001 that decays by a factor of 0.1 every 50 epochs. The batch size is set to 64. In the transfer learning phase, the model adopts an initial learning rate of 0.001 that decreases by a factor of 0.1 per 30 epochs, while the batch size is set to 8. The model trains for 130 epochs using the AdamW optimizer.

### Code and pre-trained models available

The source code for the model is available at `https://github.com/01dadada/RT-Transformer`. The pre-trained models and all data is available for download at the following link: `https://drive.google.com/file/d/1TM-w1Y9xr6iHw0wVW7SSv_B5ivDDIJIa/view?usp=sharing`

## Results and discussion

### Evaluation of the RT-Transformer Model

The SMRT dataset is a publicly accessible collection of data which can be used for the assessment of models designed for the prediction of retention times (RTs). Previous investigations have employed SMRT molecules with RTs less than 600 seconds as training data for their

models. However, our research indicates that in certain circumstances, models trained with the complete SMRT dataset exhibit superior performance when transferred to other chromatographic methods (CMs). This improved performance may be attributed to the model's ability to acquire additional features from the unretained molecules. Consequently, we have trained our model using both the complete SMRT dataset and a subset of the SMRT dataset that exclusively contains molecules with retention times of 300 seconds or greater. Specifically, we have utilized only those SMRT molecules whose RT is greater than 300 seconds to train our RT-Transformer. On the test set, our model achieved a mean absolute error (MAE) of 27.2 seconds and a median absolute error (MedAE) of 16.8 seconds. By comparison, the MAE and MedAE errors for the model trained with all SMRT molecules are 33.2 and 17.6 seconds, respectively. A comparison of the accuracy of our RT-Transformer model with previous works is presented in Table 1. In general, RT-Transformer is superior to the other models in all the performance metrics, including MAE, MRE, MedAE, MedRE, and $R^2$.

## RT-Transformer Transfer Learning on other Chromatographic Systems

Currently, the development of specific chromatographic methods is often hampered by a paucity of training data. Transfer learning offers a viable solution to address the issue of overfitting of models on small datasets by leveraging the Shared Modification Retention Time (SMRT) dataset.

To evaluate the efficacy of the transfer learning approach, we utilized 41 datasets obtained from PredRet, which were generated by diverse chromatographic methods (CMs) and contributed by researchers from independent laboratories. We pre-trained the RT-transformer model using SMRT, then froze the parameters of ResGATs and 1d-transformer modules in the pre-trained model, and fine-tuned the parameters of the Output-Block using target data to obtain RT prediction models for various CMs. The results of the transfer of the models to

the 41 CMs, including the means and standard deviations of MAE, MedAE, MRE, MedRE, and R2 determined via 10-fold cross-validation, are presented in Tables 5, 6, 3, and 4. As these datasets have not been tested in previous studies, we compared the performance of our model with other established models using previously tested datasets, which serve as a benchmark for evaluating the performance of machine learning models in various tasks. By using these well-established datasets, we could compare our model's performance with other state-of-the-art models. As shown in Table 2, our model outperformed the state-of-the-art models on most evaluated metrics and datasets.

## Application and Evaluation of RT Prediction in Compound Annotation

Currently, metabolite identification methods based on MS2 spectra always propose different candidate motifs, resulting in a large number of false positive compounds. Especially when the molecules are structurally similar, it may be difficult to identify them. RT contains information that is orthogonal to the mass spectra. It helps filter candidates, even if they have similar structures.

From the histogram in Figure 3, prediction errors of 7790 molecules in the test set can be regarded as a normal distribution. There were 7047 molecules (90.36%) with an absolute error of less than 60s and 7503 molecules (96.22) with an absolute error of less than 120s in the RT prediction. Using ±2 SD (standard deviation of RT prediction errors in the test set is 59.89s) as the filter threshold,[37] the false negative rate decrease to 3.78%. The threshold can be adjusted to filter more false positive molecules or minimize false negative molecules.

We randomly selected 100 molecules from the test set of SMRT. And the visualization in Figure 3 presents that these molecules were distributed uniformly in the SMRT dataset. Then we searched PubChem with the exact molecular mass to get their candidates. All the candidate molecules were filtered by RT-Transformer filter. The number of selected candidates and the number of candidates filtered by RT-Transformer filter are shown in Table

7. It can be seen that the means of the filter rate in 100 test molecules were 75.71%. This result showed the ability of the RT-Transformer filter to filter out false positive molecules.

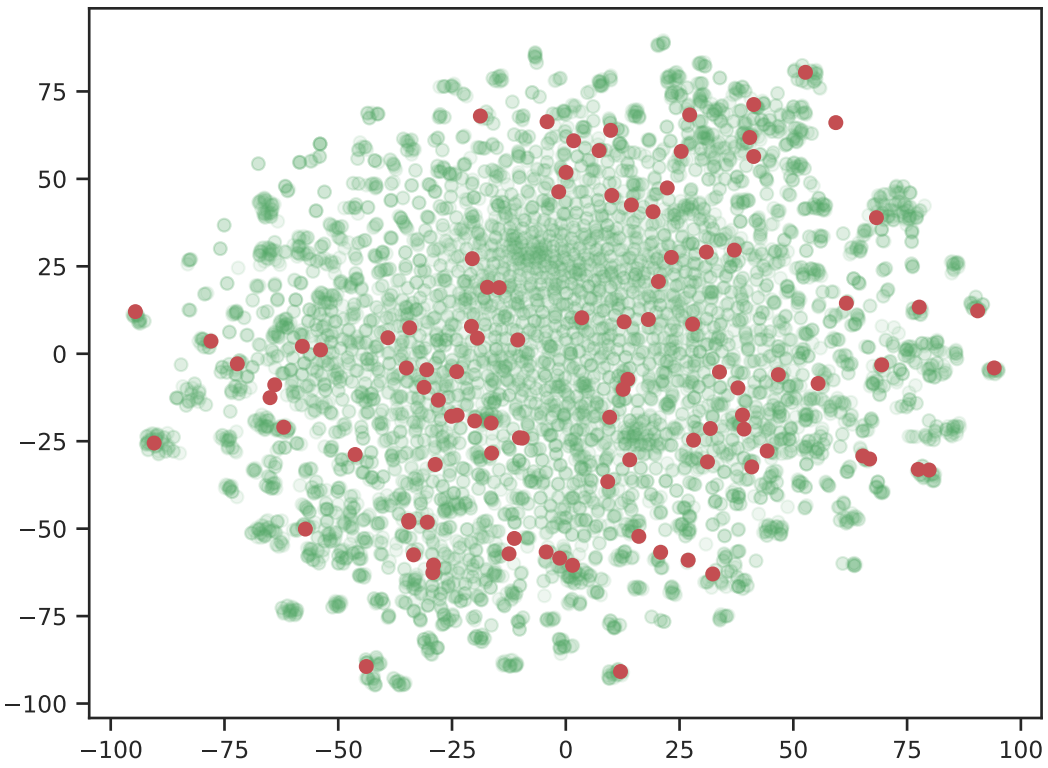


Figure 3: Distributions of 100 molecules of the test set in SMRT

We built this model to provide a filter for filtering out more isomers in non-targeted LC-MS workflow. We search isomers of all molecules from PubChem. To access the capacity of RT-Transformers in the compound annotation. We generated receiver operating characteristic (ROC) curves based on SMRT(test set) and SMRT_Retained(retained molecules test set),as shown in Figure 4 and Figure 5 respectively. The resulting curves were then utilized to determine the optimal threshold and filter out candidate compounds based on these values. The best threshold for SMRT and SMRT_Retained is 5.7% and 4.3%, and eliminating 80% and 80% false identity. We generated the boxplot with the x-axis representing the filter's effectiveness and the y-axis representing the number of molecules and the

number of molecules filtered out. The boxplot showed a significant increase in the number of molecules filtered out, indicating a notable improvement in the filter's performance. The boxplot is illustrated in Figure 5 and Figure 4.

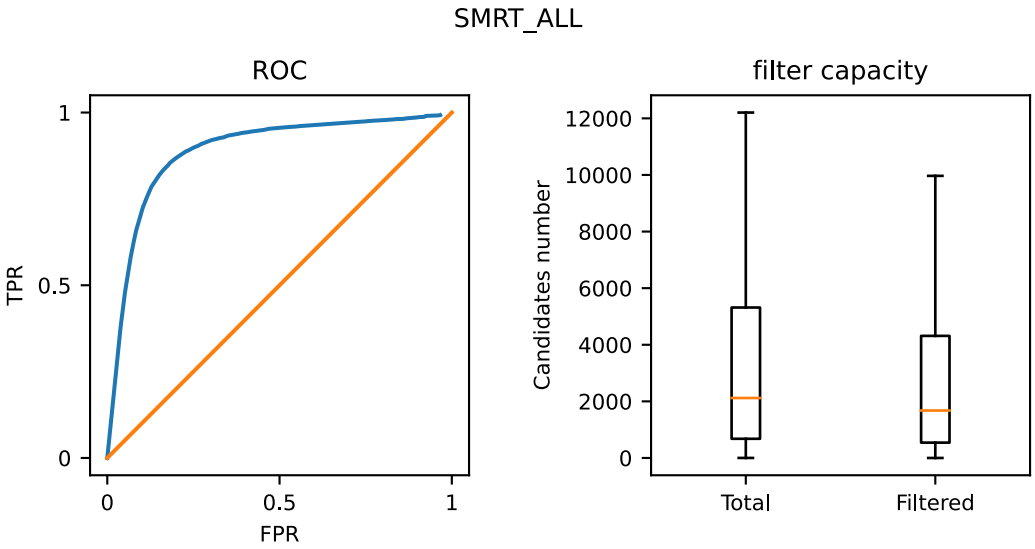


Figure 4: ROC curves and eliminated false identities of SMRT

Overall, these results indicate that RT-transformer filters can effectively filter false positive molecules with RT prediction.

## Conclusion

The retention time prediction in liquid chromatography is increasingly essential for identifying small molecules, as it provides valuable orthogonal information to tandem mass spectra. In this study, we present a robust RT prediction model, RT-Transformer, designed to aid in identifying small molecules. The model exhibits excellent scalability across different chromatographic methods, and its performance was validated on both the SMRT dataset and 41 datasets obtained using various chromatographic methods. Our results indicate that RT-Transformer outperforms state-of-the-art models when trained on the SMRT dataset. By leveraging transfer learning, the model can accurately predict RT values in any chro-

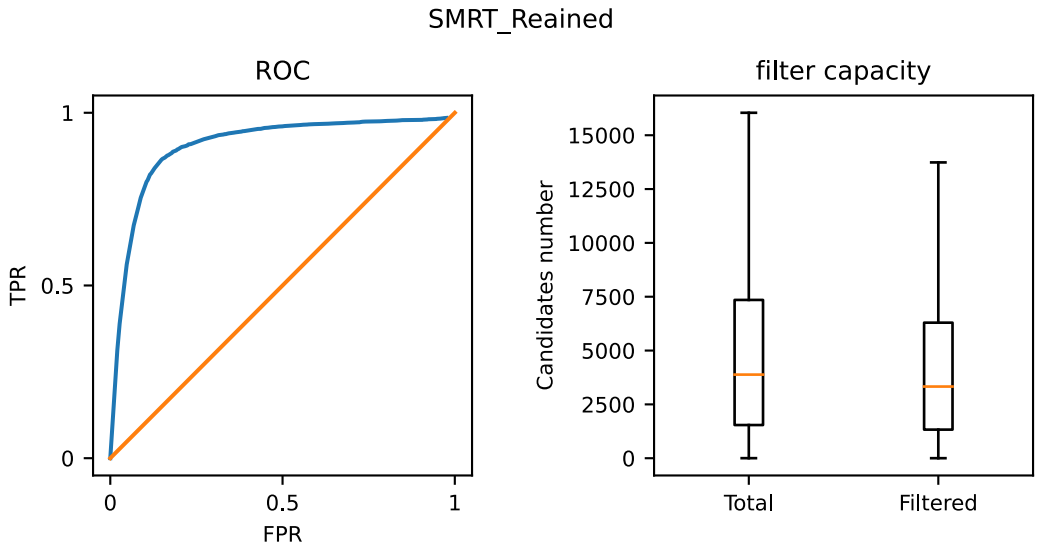


Figure 5: ROC curves and eliminated false identities of SMRT_Retained

matographic method and demonstrate superior performance to other RT prediction models. Our findings demonstrate that RT-Transformer can filter isomeric candidates based on their predicted RT values, thereby facilitating molecular identification. Furthermore, we have made the source code and pretrained-model of RT-Transformer publicly available, enabling researchers to apply this model to their datasets via transfer learning and improve the accuracy and efficiency of their chemical analyses.

## Acknowledgement

This work is supported by the National Natural Science Foundation of China (32060151), the Yunnan Provincial Foundation for Leaders of Disciplines in Science and Technology, China (202305AC160014), and the Innovation Research Foundation for Graduate Students of Yunnan University (KC-22221489)

# References


(1) Idle, J. R.; Gonzalez, F. J. Metabolomics. *Cell Metab* **2007**, *6*, 348–51.

(2) van Der Hooft, J. J. J.; Wandy, J.; Barrett, M. P.; Burgess, K. E.; Rogers, S. Topic modeling for untargeted substructure exploration in metabolomics. *Proceedings of the National Academy of Sciences* **2016**, *113*, 13738–13743.

(3) Neumann, S.; Böcker, S. Computational mass spectrometry for metabolomics: identification of metabolites and small molecules. *Analytical and bioanalytical chemistry* **2010**, *398*, 2779–2788.

(4) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N., et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic acids research* **2018**, *46*, D608–D617.

(5) Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic acids research* **2018**, *46*, W486–W494.

(6) Bell, A. W.; Deutsch, E. W.; Au, C. E.; Kearney, R. E.; Beavis, R.; Sechi, S.; Nilsson, T.; Bergeron, J. J. A HUPO test sample study reveals common problems in mass spectrometry–based proteomics. *Nature methods* **2009**, *6*, 423–430.

(7) Gika, H. G.; Theodoridis, G. A.; Plumb, R. S.; Wilson, I. D. Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *Journal of pharmaceutical and biomedical analysis* **2014**, *87*, 12–25.

(8) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *Journal of cheminformatics* **2016**, *8*, 1–16.

(9) Wang, Y.; Kora, G.; Bowen, B. P.; Pan, C. MIDAS: a database-searching algorithm for metabolite identification in metabolomics. *Analytical chemistry* **2014**, *86*, 9496–9503.

(10) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic acids research* **2014**, *42*, W94–W99.

(11) Wang, Y.; Wang, X.; Zeng, X. MIDAS-G: a computational platform for investigating fragmentation rules of tandem mass spectrometry in metabolomics. *Metabolomics* **2017**, *13*, 1–9.

(12) Ridder, L.; van der Hooft, J. J.; Verhoeven, S. Automatic compound annotation from mass spectrometry data using MAGMa. *Mass Spectrometry* **2014**, *3*, S0033–S0033.

(13) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI: FingerID. *Proceedings of the National Academy of Sciences* **2015**, *112*, 12580–12585.

(14) Djoumbou-Feunang, Y.; Pon, A.; Karu, N.; Zheng, J.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.; Wishart, D. S. CFM-ID 3.0: significantly improved ESI-MS/MS prediction and compound identification. *Metabolites* **2019**, *9*, 72.

(15) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature methods* **2019**, *16*, 299–302.

(16) Ruttkies, C.; Neumann, S.; Posch, S. Improving MetFrag with statistical learning of fragment annotations. *BMC bioinformatics* **2019**, *20*, 1–14.

(17) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A., et al. PubChem substance and compound databases. *Nucleic acids research* **2016**, *44*, D1202–D1213.

(18) Hettne, K. M.; Williams, A. J.; van Mulligen, E. M.; Kleinjans, J.; Tkachenko, V.; Kors, J. A. Automatic vs. manual curation of a multi-source chemical dictionary: the impact on text mining. *Journal of cheminformatics* **2010**, *2*, 1–7.

(19) Eugster, P. J.; Boccard, J.; Debrus, B.; Bréant, L.; Wolfender, J.-L.; Martel, S.; Carrupt, P.-A. Retention time prediction for dereplication of natural products (CxHyOz) in LC–MS metabolite profiling. *Phytochemistry* **2014**, *108*, 196–207.

(20) Stanstrup, J.; Neumann, S.; Vrhovsek, U. PredRet: prediction of retention time by direct mapping between multiple chromatographic systems. *Analytical chemistry* **2015**, *87*, 9421–9428.

(21) Cao, M.; Fraser, K.; Huege, J.; Featonby, T.; Rasmussen, S.; Jones, C. Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics. *Metabolomics* **2015**, *11*, 696–706.

(22) Falchi, F.; Bertozzi, S. M.; Ottonello, G.; Ruda, G. F.; Colombano, G.; Fiorelli, C.; Martucci, C.; Bertorelli, R.; Scarpelli, R.; Cavalli, A., et al. Kernel-based, partial least squares quantitative structure-retention relationship model for UPLC retention time prediction: a useful tool for metabolite identification. *Analytical chemistry* **2016**, *88*, 9510–9517.

(23) Bruderer, T.; Varesio, E.; Hopfgartner, G. The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition. *Journal of Chromatography B* **2017**, *1071*, 3–10.

(24) Amos, R. I.; Haddad, P. R.; Szucs, R.; Dolan, J. W.; Pohl, C. A. Molecular modeling and prediction accuracy in Quantitative Structure-Retention Relationship calculations for chromatography. *TrAC Trends in Analytical Chemistry* **2018**, *105*, 352–359.

(25) Aalizadeh, R.; Nika, M.-C.; Thomaidis, N. S. Development and application of retention

time prediction models in the suspect and non-target screening of emerging contaminants. *Journal of Hazardous materials* **2019**, *363*, 277–285.

(26) Pasin, D.; Mollerup, C. B.; Rasmussen, B. S.; Linnet, K.; Dalsgaard, P. W. Development of a single retention time prediction model integrating multiple liquid chromatography systems: Application to new psychoactive substances. *Analytica Chimica Acta* **2021**, *1184*, 339035.

(27) Aicheler, F.; Li, J.; Hoene, M.; Lehmann, R.; Xu, G.; Kohlbacher, O. Retention time prediction improves identification in nontargeted lipidomics approaches. *Analytical chemistry* **2015**, *87*, 7698–7704.

(28) Wolfer, A. M.; Lozano, S.; Umbdenstock, T.; Croixmarie, V.; Arrault, A.; Vayer, P. UPLC–MS retention time prediction: A machine learning approach to metabolite identification in untargeted profiling. *Metabolomics* **2016**, *12*, 8.

(29) Bach, E.; Szedmak, S.; Brouard, C.; Böcker, S.; Rousu, J. Liquid-chromatography retention order prediction for metabolite identification. *Bioinformatics* **2018**, *34*, i875–i883.

(30) Bonini, P.; Kind, T.; Tsugawa, H.; Barupal, D. K.; Fiehn, O. Retip: retention time prediction for compound annotation in untargeted metabolomics. *Analytical chemistry* **2020**, *92*, 7515–7522.

(31) Feng, C.; Xu, Q.; Qiu, X.; Ji, J.; Lin, Y.; Le, S.; She, J.; Lu, D.; Wang, G., et al. Evaluation and application of machine learning-based retention time prediction for suspect screening of pesticides and pesticide transformation products in LC-HRMS. *Chemosphere* **2021**, *271*, 129447.

(32) Liapikos, T.; Zisi, C.; Kodra, D.; Kademoglou, K.; Diamantidou, D.; Begou, O.; Pappa-Louisi, A.; Theodoridis, G. Quantitative structure retention relationship (QSRR) modelling for Analytes' retention prediction in LC-HRMS by applying different Machine

Learning algorithms and evaluating their performance. *Journal of Chromatography B* **2022**, *1191*, 123132.

(33) Bouwmeester, R.; Martens, L.; Degroeve, S. Comprehensive and empirical evaluation of machine learning algorithms for small molecule LC retention time prediction. *Analytical chemistry* **2019**, *91*, 3694–3703.

(34) Domingo-Almenara, X.; Guijas, C.; Billings, E.; Montenegro-Burke, J. R.; Uritboonthai, W.; Aisporna, A. E.; Chen, E.; Benton, H. P.; Siuzdak, G. The METLIN small molecule dataset for machine learning-based retention time prediction. *Nature communications* **2019**, *10*, 5811.

(35) Ju, R.; Liu, X.; Zheng, F.; Lu, X.; Xu, G.; Lin, X. Deep neural network pretrained by weighted autoencoders and transfer learning for retention time prediction of small molecules. *Analytical Chemistry* **2021**, *93*, 15651–15658.

(36) Fedorova, E. S.; Matyushin, D. D.; Plyushchenko, I. V.; Stavrianidi, A. N.; Buryak, A. K. Deep learning for retention time prediction in reversed-phase liquid chromatography. *Journal of Chromatography A* **2022**, *1664*, 462792.

(37) Yang, Q.; Ji, H.; Lu, H.; Zhang, Z. Prediction of liquid chromatographic retention time with graph neural networks to assist in small molecule identification. *Analytical Chemistry* **2021**, *93*, 2200–2206.

(38) Zaretckii, M.; Bashkirova, I.; Osipenko, S.; Kostyukevich, Y.; Nikolaev, E.; Popov, P. 3D chemical structures allow robust deep learning models for retention time prediction. *Digital Discovery* **2022**, *1*, 711–718.

(39) Osipenko, S.; Nikolaev, E.; Kostyukevich, Y. Retention Time Prediction with Message-Passing Neural Networks. *Separations* **2022**, *9*, 291.

(40) García, C. A.; Gil-de-la Fuente, A.; Barbas, C.; Otero, A. Probabilistic metabolite annotation using retention time prediction and meta-learned projections. *Journal of Cheminformatics* **2022**, *14*, 1–23.

(41) Weiss, K.; Khoshgoftaar, T. M.; Wang, D. A survey of transfer learning. *Journal of Big data* **2016**, *3*, 1–40.

(42) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903* **2017**,

(43) Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* **2014**,

(44) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*.

(45) Landrum, G., et al. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum* **2013**, *8*.

(46) Ba, J. L.; Kiros, J. R.; Hinton, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450* **2016**,

Table 1: RT Prediction Accuracy of Different Models

| Methods | MAE(s) | MRE(%) | MedAE(s) | MedRE(%) | $R^2$ |
|---|---|---|---|---|---|
| GNN-RT | 39 | 5 | 24 | — | 0.85 |
| 1D-CNN | 34.7 | 4.3 | 18.7 | 2.4 | — |
| Blender | 34 | — | 17.2 | — | — |
| CPORT | 44 | 5.5 | 26 | 3.4 | — |
| MPNN | 31.5 | 4 | 16 | — | 0.879 |
| RT-Transformer | 27.3 | 3.2 | 11.88 | 1.58 | 0.87 |

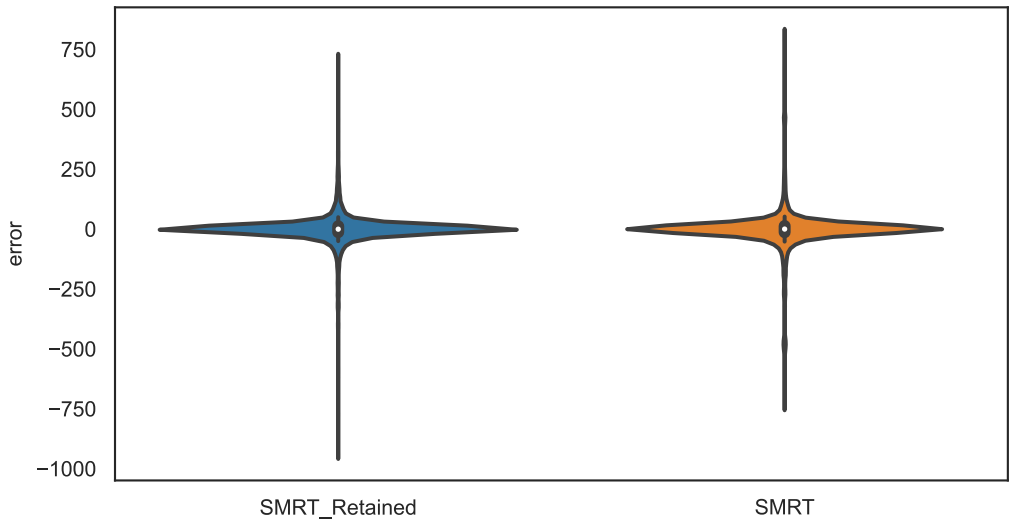


Figure 6: distribution of errors

Table 2: MAEs of different models on multiple datasets

| Model | RIKEN Retip | FEM_long | Eawag_XBridgeC18 | LIFE_new |
|---|---|---|---|---|
| 1D-CNN | 32.4 | — | — | 23.6 |
| MPNN | 38.2 | 204.6 | 80.9 | 22.1 |
| GNN-RT | — | 235.01 | 112.78 | 29.38 |
| RT-Transformers(SMRT_Retained) | 37.16 | 184.04 | 73.34 | 20.36 |
| RT-Transformers(SMRT) | — | 176.53 | 69.8 | 22.12 |

Table 3: Means of MAE, MedAE, MRE, MedRE, and R2 of 10-fold Cross-Validation of 41 CMs transfer by the model trained with retained molecules

| dataset | mae(s) | mre(%) | medAE(s) | medRE(%) | R2 |
|---|---|---|---|---|---|
| Acquity HSST3-RP-60mn | 328.73 | 49.10 | 238.73 | 19.12 | 0.62 |
| AjsTestF | 77.57 | 41.21 | 50.11 | 25.60 | 0.58 |
| AjsUoB | 61.25 | 51.01 | 32.91 | 26.60 | 0.68 |
| BDD_C18 | 52.07 | 46.67 | 18.88 | 13.58 | 0.78 |
| BfG_NTS_RP1 | 78.84 | 20.52 | 53.09 | 10.54 | 0.72 |
| CBM_Test_A | 150.65 | 41.93 | 131.50 | 34.07 | -0.05 |
| CBM_Test_A_ | 57.63 | 20.20 | 41.33 | 13.30 | -0.15 |
| CBM_Test_B | 137.66 | 46.14 | 112.38 | 37.94 | -0.04 |
| CBM_Test_E | 75.47 | 26.30 | 58.15 | 19.72 | -0.01 |
| CBM_Test_F | 23.26 | 3.39 | 17.88 | 2.58 | -0.82 |
| CBM_Test_G | 133.39 | 45.91 | 103.12 | 23.17 | 0.35 |
| CS1 | 32.95 | 10.31 | 21.86 | 3.65 | 0.88 |
| Cao_HILIC | 101.74 | 64.94 | 52.28 | 36.21 | 0.48 |
| Eawag_XBridgeC18 | 73.34 | 27.43 | 58.40 | 13.41 | 0.82 |
| FEM_long | 184.04 | 36.85 | 122.45 | 11.80 | 0.93 |
| FEM_orbitrap_plasma | 50.29 | 18.63 | 35.87 | 8.77 | 0.94 |
| HILIC_BDD_2 | 148.64 | 47.83 | 91.77 | 17.93 | 0.54 |
| HILIC_tip | 80.51 | 42.63 | 50.79 | 26.49 | 0.51 |
| IJM_TEST | 92.66 | 51.26 | 36.84 | 11.30 | 0.27 |
| IPB_Halle | 48.75 | 38.98 | 29.04 | 20.61 | 0.64 |
| KI_GIAR_zic_HILIC_pH2_7 | 97.36 | 43.87 | 80.22 | 23.03 | 0.51 |
| LIFE_new | 20.36 | 26.61 | 11.82 | 12.53 | 0.87 |
| LIFE_old | 11.57 | 16.04 | 8.27 | 9.31 | 0.91 |
| MTBLS87 | 72.54 | 12.30 | 51.62 | 7.88 | 0.65 |
| MTBLS_36 | 61.16 | 38.45 | 29.26 | 18.79 | -0.01 |
| Mceachran HPLC | 50.54 | 26.08 | 42.32 | 16.19 | 0.70 |
| Meister zic-pHILIC pH9.3 | 71.23 | 35.49 | 54.24 | 24.10 | 0.51 |
| RIKEN | 45.29 | 81.57 | 17.25 | 30.79 | 0.62 |
| RPFDAMM | 23.11 | 8.20 | 14.37 | 4.59 | 0.84 |
| RPLC_zorbax150_JH | 32.94 | 30.41 | 20.43 | 20.06 | 0.60 |
| RPMMFDA | 33.48 | 18.14 | 22.28 | 7.01 | 0.79 |
| SNU_RIKEN_POS | 37.39 | 15.12 | 28.60 | 10.83 | 0.80 |
| SNU_RP_indole_annotation | 34.34 | 16.31 | 18.94 | 7.53 | 0.94 |
| SNU_RP_indole_order | 36.89 | 17.47 | 24.88 | 8.65 | 0.83 |
| SNU_organoid | 95.81 | 43.91 | 60.87 | 11.99 | 0.87 |
| UFZ_Phenomenex | 128.79 | 21.63 | 81.19 | 6.18 | 0.64 |
| UniToyama_Atlantis | 44.80 | 5.12 | 27.50 | 3.05 | 0.90 |
| Waters ACQUITY UPLC with Synapt G1 Q-TOF | 135.62 | 48.76 | 97.53 | 31.20 | 0.29 |
| Waters STA Forensic | 132.33 | 54.71 | 110.49 | 32.78 | -0.03 |
| cecum_JS | 36.89 | 17.47 | 24.88 | 8.65 | 0.83 |

Table 4: Standard deviation of MAE, MedAE, MRE, MedRE, and R2 of 10-fold Cross-Validation of 41 CMs transfer by the model trained with all molecules

| dataset | mae | mre | medAE | medRE | R2 |
|---|---|---|---|---|---|
| Acquity HSST3-RP-60mn | 87.15 | 39.14 | 90.71 | 8.17 | 0.18 |
| AjsTestF | 8.60 | 6.97 | 8.06 | 3.69 | 0.11 |
| AjsUoB | 6.64 | 6.53 | 7.39 | 5.04 | 0.07 |
| BDD_C18 | 9.82 | 14.37 | 8.01 | 3.07 | 0.09 |
| BfG_NTS_RP1 | 8.60 | 4.04 | 7.32 | 1.43 | 0.07 |
| CBM_Test_A | 10.39 | 4.09 | 18.57 | 3.27 | 0.05 |
| CBM_Test_A_ | 23.79 | 8.72 | 26.84 | 8.05 | 0.30 |
| CBM_Test_B | 13.94 | 5.94 | 19.29 | 6.56 | 0.07 |
| CBM_Test_E | 15.30 | 5.64 | 15.91 | 5.56 | 0.13 |
| CBM_Test_F | 8.38 | 1.29 | 9.77 | 1.40 | 1.16 |
| CBM_Test_G | 43.49 | 26.09 | 45.35 | 15.43 | 0.39 |
| CS1 | 11.44 | 12.14 | 7.96 | 1.22 | 0.07 |
| Cao_HILIC | 13.24 | 13.31 | 14.95 | 6.86 | 0.09 |
| Eawag_XBridgeC18 | 8.19 | 8.26 | 11.57 | 2.81 | 0.05 |
| FEM_long | 40.01 | 17.49 | 37.76 | 3.88 | 0.04 |
| FEM_orbitrap_plasma | 20.21 | 13.51 | 19.84 | 4.83 | 0.05 |
| HILIC_BDD_2 | 22.73 | 12.87 | 26.37 | 5.36 | 0.13 |
| HILIC_tip | 8.35 | 5.83 | 7.96 | 3.62 | 0.10 |
| IJM_TEST | 26.07 | 18.45 | 11.90 | 3.70 | 0.31 |
| IPB_Halle | 14.07 | 13.26 | 14.93 | 9.27 | 0.28 |
| KI_GIAR_zic_HILIC_pH2_7 | 10.14 | 7.28 | 13.15 | 4.96 | 0.11 |
| LIFE_new | 6.39 | 11.79 | 5.77 | 5.84 | 0.08 |
| LIFE_old | 2.61 | 4.70 | 2.51 | 3.23 | 0.08 |
| MTBLS87 | 27.52 | 5.90 | 28.99 | 3.74 | 0.29 |
| MTBLS_36 | 35.53 | 22.21 | 25.23 | 8.55 | 1.16 |
| Mceachran HPLC | 12.39 | 7.49 | 11.64 | 5.31 | 0.21 |
| Meister zic-pHILIC pH9.3 | 14.83 | 8.32 | 16.65 | 6.50 | 0.18 |
| RIKEN | 16.32 | 25.34 | 5.39 | 9.64 | 0.28 |
| RPFDAMM | 9.54 | 4.65 | 8.75 | 2.95 | 0.14 |
| RPLC_zorbax150_JH | 7.24 | 7.78 | 6.26 | 6.07 | 0.25 |
| RPMMFDA | 3.24 | 3.79 | 2.15 | 0.78 | 0.06 |
| SNU_RIKEN_POS | 3.42 | 1.71 | 3.31 | 1.37 | 0.04 |
| SNU_RP_indole_annotation | 14.15 | 9.37 | 11.09 | 3.51 | 0.06 |
| SNU_RP_indole_order | 18.78 | 13.29 | 16.19 | 5.50 | 0.37 |
| SNU_organoid | 47.68 | 38.69 | 37.89 | 4.47 | 0.13 |
| UFZ_Phenomenex | 30.13 | 12.70 | 18.94 | 1.46 | 0.11 |
| UniToyama_Atlantis | 16.10 | 1.99 | 15.09 | 1.69 | 0.08 |
| Waters ACQUITY UPLC with Synapt G1 Q-TOF | 17.44 | 8.49 | 15.37 | 4.68 | 0.22 |
| Waters STA Forensic | 19.06 | 15.50 | 25.62 | 8.26 | 0.07 |
| cecum_JS | 18.78 | 13.29 | 16.19 | 5.50 | 0.37 |

Table 5: Means of MAE, MedAE, MRE, MedRE, and R2 of 10-fold Cross-Validation of 41 CMs transfer by the model trained with all molecules

| dataset | mae | mre | medAE | medRE | R2 |
|---|---|---|---|---|---|
| Acquity HSST3-RP-60mn | 311.33 | 32.39 | 225.11 | 17.42 | 0.65 |
| AjsTestF | 73.95 | 38.07 | 45.75 | 23.49 | 0.60 |
| AjsUoB | 51.92 | 45.91 | 28.19 | 21.17 | 0.77 |
| BDD_C18 | 47.95 | 49.22 | 20.36 | 12.53 | 0.78 |
| BfG_NTS_RP1 | 76.16 | 18.76 | 49.77 | 10.13 | 0.72 |
| CBM_Test_A | 158.05 | 43.76 | 139.76 | 35.29 | -0.07 |
| CBM_Test_A_ | 76.62 | 25.92 | 59.44 | 19.15 | -0.10 |
| CBM_Test_B | 141.89 | 47.65 | 115.75 | 37.55 | -0.07 |
| CBM_Test_E | 85.32 | 28.26 | 67.59 | 21.78 | 0.03 |
| CBM_Test_F | 53.58 | 11.38 | 21.13 | 3.10 | -0.07 |
| CBM_Test_G | 174.07 | 61.69 | 125.25 | 29.07 | -0.02 |
| CS1 | 36.63 | 20.73 | 16.90 | 2.89 | 0.85 |
| Cao_HILIC | 105.88 | 60.98 | 52.18 | 34.46 | 0.50 |
| Eawag_XBridgeC18 | 69.80 | 22.65 | 48.04 | 12.52 | 0.81 |
| FEM_long | 176.53 | 47.24 | 106.03 | 10.46 | 0.91 |
| FEM_orbitrap_plasma | 68.28 | 22.54 | 39.34 | 9.14 | 0.88 |
| HILIC_BDD_2 | 116.60 | 38.25 | 57.87 | 11.87 | 0.66 |
| HILIC_tip | 73.69 | 38.10 | 44.90 | 24.07 | 0.58 |
| IJM_TEST | 104.98 | 44.60 | 31.15 | 9.80 | 0.26 |
| IPB_Halle | 27.97 | 24.21 | 15.53 | 10.91 | 0.86 |
| KI_GIAR_zic_HILIC_pH2_7 | 81.02 | 34.77 | 59.41 | 16.98 | 0.64 |
| LIFE_new | 22.12 | 30.29 | 9.87 | 9.93 | 0.81 |
| LIFE_old | 13.09 | 17.31 | 7.65 | 8.94 | 0.86 |
| MTBLS87 | 138.18 | 22.27 | 86.90 | 13.56 | 0.33 |
| MTBLS_36 | 93.06 | 42.57 | 38.43 | 20.72 | 0.18 |
| Mceachran HPLC | 84.21 | 33.01 | 47.58 | 18.30 | 0.32 |
| Meister zic-pHILIC pH9.3 | 80.18 | 34.06 | 50.63 | 23.13 | 0.42 |
| RIKEN | 50.70 | 61.01 | 11.51 | 27.15 | 0.52 |
| RPFDAMM | 33.74 | 26.27 | 18.51 | 6.04 | 0.76 |
| RPLC_zorbax150_JH | 47.36 | 38.89 | 24.50 | 19.77 | 0.35 |
| RPMMFDA | 30.90 | 14.69 | 20.94 | 6.28 | 0.82 |
| SNU_RIKEN_POS | 35.85 | 13.86 | 25.56 | 9.64 | 0.80 |
| SNU_RP_indole_annotation | 42.59 | 17.97 | 17.68 | 7.41 | 0.88 |
| SNU_RP_indole_order | 46.09 | 19.16 | 21.99 | 8.47 | 0.87 |
| SNU_organoid | 151.22 | 51.93 | 62.88 | 16.57 | 0.59 |
| UFZ_Phenomenex | 105.13 | 13.78 | 67.53 | 5.30 | 0.79 |
| UniToyama_Atlantis | 95.30 | 13.32 | 55.00 | 5.99 | 0.67 |
| Waters ACQUITY UPLC with Synapt G1 Q-TOF | 123.82 | 48.71 | 87.91 | 26.85 | 0.43 |
| Waters STA Forensic | 149.06 | 56.21 | 125.59 | 36.79 | -0.08 |
| cecum_JS | 46.14 | 19.22 | 22.04 | 8.51 | 0.87 |

Table 6: Standard deviation of MAE, MedAE, MRE, MedRE, and R2 of 10-fold Cross-Validation of 41 CMs transfer by the model trained with all molecules

| dataset | mae | mre | medAE | medRE | R2 |
|---|---|---|---|---|---|
| Acquity HSST3-RP-60mn | 62.22 | 9.86 | 83.40 | 6.50 | 0.17 |
| AjsTestF | 7.84 | 6.12 | 9.13 | 3.65 | 0.09 |
| AjsUoB | 5.77 | 8.59 | 4.59 | 5.31 | 0.06 |
| BDD_C18 | 11.99 | 18.69 | 9.87 | 3.70 | 0.15 |
| BfG_NTS_RP1 | 8.77 | 3.69 | 5.68 | 1.29 | 0.11 |
| CBM_Test_A | 11.73 | 4.84 | 20.14 | 3.63 | 0.06 |
| CBM_Test_A_ | 20.96 | 8.47 | 33.88 | 10.37 | 0.29 |
| CBM_Test_B | 15.08 | 7.42 | 19.69 | 6.50 | 0.10 |
| CBM_Test_E | 14.46 | 5.57 | 19.47 | 5.30 | 0.13 |
| CBM_Test_F | 31.98 | 8.79 | 13.37 | 2.05 | 0.63 |
| CBM_Test_G | 34.52 | 20.47 | 51.31 | 17.34 | 0.36 |
| CS1 | 15.26 | 21.12 | 8.44 | 1.36 | 0.12 |
| Cao_HILIC | 14.40 | 13.02 | 15.08 | 6.22 | 0.08 |
| Eawag_XBridgeC18 | 13.61 | 6.51 | 10.23 | 3.17 | 0.11 |
| FEM_long | 42.36 | 31.66 | 30.43 | 3.03 | 0.09 |
| FEM_orbitrap_plasma | 33.48 | 15.63 | 26.15 | 4.29 | 0.15 |
| HILIC_BDD_2 | 18.95 | 8.51 | 17.75 | 3.12 | 0.11 |
| HILIC_tip | 7.13 | 4.79 | 7.26 | 3.97 | 0.09 |
| IJM_TEST | 27.62 | 19.84 | 11.04 | 3.68 | 0.27 |
| IPB_Halle | 10.84 | 13.17 | 10.88 | 6.89 | 0.18 |
| KI_GIAR_zic_HILIC_pH2_7 | 9.94 | 5.99 | 12.09 | 2.60 | 0.12 |
| LIFE_new | 6.04 | 11.86 | 4.32 | 4.26 | 0.12 |
| LIFE_old | 4.03 | 6.31 | 3.00 | 2.86 | 0.15 |
| MTBLS87 | 51.15 | 7.63 | 46.76 | 7.23 | 0.38 |
| MTBLS_36 | 39.46 | 18.71 | 38.86 | 12.47 | 0.73 |
| Mceachran HPLC | 34.19 | 11.64 | 19.06 | 6.47 | 0.33 |
| Meister zic-pHILIC pH9.3 | 20.68 | 8.58 | 17.66 | 6.38 | 0.23 |
| RIKEN | 15.08 | 22.46 | 7.18 | 15.06 | 0.28 |
| RPFDAMM | 19.40 | 42.98 | 9.00 | 2.61 | 0.31 |
| RPLC_zorbax150_JH | 20.27 | 13.37 | 11.29 | 6.21 | 0.48 |
| RPMMFDA | 3.01 | 2.64 | 2.36 | 0.65 | 0.07 |
| SNU_RIKEN_POS | 3.95 | 1.51 | 3.84 | 1.60 | 0.04 |
| SNU_RP_indole_annotation | 18.66 | 12.09 | 11.40 | 3.75 | 0.11 |
| SNU_RP_indole_order | 24.32 | 17.37 | 18.13 | 4.35 | 0.15 |
| SNU_organoid | 87.31 | 29.18 | 44.94 | 13.66 | 0.46 |
| UFZ_Phenomenex | 15.33 | 5.01 | 15.04 | 1.23 | 0.09 |
| UniToyama_Atlantis | 45.24 | 10.81 | 41.01 | 4.24 | 0.24 |
| Waters ACQUITY UPLC with Synapt G1 Q-TOF | 14.87 | 9.00 | 17.92 | 5.89 | 0.15 |
| Waters STA Forensic | 22.29 | 15.54 | 30.47 | 7.96 | 0.10 |
| cecum_JS | 24.30 | 17.34 | 18.13 | 4.32 | 0.15 |

Table 7: The results of RT-Transformer trained with retained molecules of SMRT in ranking capability

| formula | experimental RT | Num of candidates searched from PubChem | Num of candidates filtered by RT-TFM | Rate of candidates filtered by RT-TFM | candidate is filtered out |
|---|---|---|---|---|---|
| C16H16N4O | 796.3 | 8896 | 7505 | 84.36 | TRUE |
| C27H33N3OS | 855.8 | 307 | 279 | 90.88 | FALSE |
| C15H21N3O | 518.3 | 27937 | 18393 | 65.84 | TRUE |
| C23H28N6O2S | 1108.2 | 992 | 983 | 99.09 | FALSE |
| C19H23N3O3 | 888.9 | 21452 | 19283 | 89.89 | TRUE |
| C21H23BrN4O3 | 621.4 | 375 | 257 | 68.53 | FALSE |
| C22H22N2O5S | 802.5 | 3753 | 2817 | 75.06 | TRUE |
| C19H20N2O3 | 869.6 | 15864 | 11871 | 74.83 | TRUE |
| C16H20N4O3 | 709.0 | 10154 | 7148 | 70.40 | TRUE |
| C21H19N5O | 673.8 | 3683 | 2354 | 63.92 | TRUE |
| C17H14ClN3O2S | 903.0 | 1691 | 1152 | 68.13 | TRUE |
| C23H21FN6O4S | 782.7 | 101 | 61 | 60.40 | TRUE |
| C20H19N5O2S | 737.8 | 3192 | 2029 | 63.57 | TRUE |
| C23H17N3O6 | 1198.2 | 541 | 461 | 85.21 | TRUE |
| C23H32N2O4 | 764.5 | 3137 | 2423 | 77.24 | TRUE |
| C19H19ClN4O2S | 891.5 | 1966 | 1457 | 74.11 | TRUE |
| C18H26N4OS | 914.5 | 3414 | 3205 | 93.88 | TRUE |
| C24H27N5O2 | 858.6 | 5646 | 5038 | 89.23 | TRUE |
| C18H15F2N3O | 647.6 | 715 | 527 | 73.71 | TRUE |
| C22H21N5O2 | 1083.7 | 5390 | 5204 | 96.55 | TRUE |
| C14H19NO2 | 821.0 | 24904 | 19803 | 79.52 | FALSE |

| | | | | | |
|---|---|---|---|---|---|
| C24H28N4O2 | 692.7 | 7952 | 4971 | 62.51 | TRUE |
| C19H15BrN4O | 1146.6 | 305 | 276 | 90.49 | TRUE |
| C18H13N3O4 | 1147.9 | 1163 | 1125 | 96.73 | TRUE |
| C17H27NO3 | 685.7 | 13286 | 7888 | 59.37 | FALSE |
| C20H24N4O4S | 791.7 | 3384 | 2637 | 77.93 | TRUE |
| C22H27N3O6S2 | 747.3 | 569 | 381 | 66.96 | TRUE |
| C10H13N5 | 726.4 | 3899 | 3711 | 95.18 | FALSE |
| C21H26N4O4S | 617.4 | 3155 | 1785 | 56.58 | TRUE |
| C17H16N2O2S | 800.2 | 4575 | 3151 | 68.87 | TRUE |
| C13H22N2O3 | 561.2 | 14740 | 6257 | 42.45 | TRUE |
| C15H15N3O3S | 728.0 | 3777 | 2284 | 60.47 | TRUE |
| C24H36N4O3 | 627.0 | 2151 | 1029 | 47.84 | TRUE |
| C16H19N3O2S3 | 972.6 | 148 | 131 | 88.51 | TRUE |
| C22H31N3O4 | 1062.6 | 4355 | 4191 | 96.23 | TRUE |
| C18H26N2O2 | 745.5 | 14671 | 10605 | 72.29 | TRUE |
| C17H22ClFN2O3S | 885.2 | 69 | 55 | 79.71 | TRUE |
| C25H29N5O2S | 1058.8 | 2331 | 2138 | 91.72 | TRUE |
| C17H18N4O2S | 838.0 | 6287 | 5198 | 82.68 | TRUE |
| C16H10ClN5O | 796.7 | 135 | 96 | 71.11 | TRUE |
| C19H22N4O2 | 604.3 | 14431 | 6598 | 45.72 | TRUE |
| C20H21N3O4S | 918.0 | 7945 | 6329 | 79.66 | TRUE |
| C23H22N4O | 673.4 | 3755 | 2508 | 66.79 | TRUE |
| C22H21N5O4S | 761.0 | 1358 | 902 | 66.42 | TRUE |
| C15H22N2OS | 933.6 | 9324 | 8437 | 90.49 | TRUE |
| C20H16FN3O4S | 924.0 | 373 | 271 | 72.65 | TRUE |
| C19H18ClN3O3 | 816.1 | 4212 | 2976 | 70.66 | TRUE |
| C22H23ClFN5O | 723.1 | 235 | 159 | 67.66 | FALSE |

| | | | | | |
|---|---|---|---|---|---|
| C26H33N3O4S | 1017.5 | 2420 | 1883 | 77.81 | TRUE |
| C24H28N4O3 | 852.0 | 8984 | 7571 | 84.27 | TRUE |
| C20H23N5O2 | 697.1 | 9960 | 6992 | 70.20 | TRUE |
| C21H21FN2O4S | 872.0 | 1128 | 726 | 64.36 | TRUE |
| C18H23NO3 | 621.6 | 9160 | 7151 | 78.07 | TRUE |
| C23H24ClN3O4S | 795.7 | 1046 | 838 | 80.11 | TRUE |
| C19H20ClN3O | 1058.0 | 2089 | 1881 | 90.04 | TRUE |
| C19H25N3O4S | 928.6 | 5333 | 4572 | 85.73 | FALSE |
| C18H24N4O3S2 | 1104.5 | 1400 | 1350 | 96.43 | TRUE |
| C22H29NO5 | 1246.6 | 2295 | 2134 | 92.98 | TRUE |
| C17H21ClFN3O2 | 705.4 | 298 | 214 | 71.81 | TRUE |
| C16H18FN3O3S | 720.7 | 1096 | 555 | 50.64 | TRUE |
| C15H12FN3O2S | 976.1 | 627 | 565 | 90.11 | TRUE |
| C20H16N4O4S | 681.2 | 1242 | 928 | 74.72 | TRUE |
| C15H16N2O4 | 975.6 | 11207 | 10744 | 95.87 | FALSE |
| C20H22N2O4 | 603.4 | 15634 | 13447 | 86.01 | FALSE |
| C20H27N3O3S | 776.9 | 6293 | 4913 | 78.07 | TRUE |
| C23H24N6O3 | 565.6 | 2525 | 1964 | 77.78 | TRUE |
| C21H16FN3O2S | 1285.3 | 751 | 732 | 97.47 | TRUE |
| C22H32FN3O3 | 724.2 | 363 | 260 | 71.63 | FALSE |
| C19H18N4O2 | 799.1 | 8289 | 6300 | 76.00 | TRUE |
| C20H24Br2N2O | 906.2 | 42 | 35 | 83.33 | TRUE |
| C21H28ClN5O | 585.0 | 636 | 398 | 62.58 | FALSE |
| C21H22ClN3O2 | 811.9 | 3962 | 3070 | 77.49 | TRUE |
| C24H30FN3O | 642.2 | 828 | 449 | 54.23 | TRUE |
| C29H41N3O3 | 924.4 | 933 | 797 | 85.42 | TRUE |
| C21H28N2O3 | 653.7 | 9554 | 6099 | 63.84 | TRUE |

| | | | | | |
|---|---|---|---|---|---|
| C21H17N3O2S | 745.8 | 2527 | 2104 | 83.26 | TRUE |
| C10H13N3O3S2 | 738.1 | 645 | 559 | 86.67 | TRUE |
| C18H15FN4O4 | 907.3 | 434 | 332 | 76.50 | TRUE |
| C23H31N5O2 | 646.4 | 4223 | 1990 | 47.12 | TRUE |
| C31H35NO10 | 721.2 | 107 | 85 | 79.44 | TRUE |
| C18H23N3O4 | 750.0 | 10416 | 7474 | 71.75 | TRUE |
| C19H25N3O3 | 1021.0 | 18650 | 18102 | 97.06 | TRUE |
| C20H19N5O3 | 649.6 | 4171 | 2208 | 52.94 | TRUE |
| C20H26N4O3 | 783.0 | 11688 | 9684 | 82.85 | TRUE |
| C24H33N5O3 | 873.9 | 2678 | 2479 | 92.57 | TRUE |
| C20H19ClN2O6 | 1143.4 | 536 | 487 | 90.86 | FALSE |
| C19H20N2O | 879.0 | 5684 | 4173 | 73.42 | TRUE |
| C16H18FN3O3S | 685.3 | 1096 | 513 | 46.81 | TRUE |
| C15H19FN4O2S | 736.0 | 625 | 375 | 60.00 | TRUE |
| C14H13BrN4O | 684.8 | 989 | 644 | 65.12 | TRUE |
| C9H10N2O2 | 655.4 | 5495 | 3163 | 57.56 | FALSE |
| C24H28N4O3 | 656.6 | 8984 | 5066 | 56.39 | TRUE |
| C16H16N4O | 1025.2 | 8896 | 8730 | 98.13 | TRUE |
| C19H20N4O3 | 642.0 | 9785 | 5317 | 54.34 | FALSE |
| C23H24FN5OS | 1108.7 | 377 | 335 | 88.86 | FALSE |
| C22H24FN5O4S | 922.2 | 314 | 271 | 86.31 | TRUE |
| C27H31N3O3S | 674.1 | 2334 | 1744 | 74.72 | TRUE |
| C17H16N2O6 | 777.5 | 1900 | 1220 | 64.21 | TRUE |
| C26H27N3O3S | 830.9 | 2862 | 2365 | 82.63 | TRUE |
| C18H16FN3OS | 783.6 | 1136 | 822 | 72.36 | TRUE |