# Interpretation of Compound Fragments via Attentive Recursive Tree

Nural Ozel[1 [0000−0002−7197−9071]*], Berk Atıl[2 [0000−0002−5979−2394]*], Asu Büşra Temizer[3 [0000−0003−1811−093X]], Taha Khoulani[3 [0000−0001−7159−6006]], Elif Ozkirimli[4 [0000−0002−3206−8427]], Nilgün Karalı[3 [0000−0002−6916−122X]], Kutlu O. Ulgen[1,5 [0000−0003−3668−3467]], and Arzucan Özgür[2 [0000−0001−8376−1056]]

[1] Department of Computational Science and Engineering, Boğaziçi University, İstanbul, Turkey
[2] Department of Computer Engineering, Boğaziçi University, İstanbul, Turkey
[3] Department of Pharmaceutical Chemistry, İstanbul University, İstanbul, Turkey
[4] Data and Analytics Chapter, Pharma International Informatics, F. Hoffmann-La Roche AG, Switzerland
[5] Department of Chemical Engineering, Boğaziçi University, İstanbul, Turkey

**Abstract.** The discovery of new drug-like chemicals with desired properties is a challenging and costly process in the pharmaceutical industry. To facilitate this process in the preclinical phase, many different neural network models have been proposed for different tasks (e.g., drug-target affinity prediction, molecular property prediction, target-specific molecule generation). Despite producing successful results, they usually lack interpretability. Because it is impossible to determine how and what the neural network model has learned, they behave like a black box. To comprehend the significance of each fragment in the relevant compounds, we employed the Attention Recursive Tree (AR-Tree) model. Thanks to its task-specific attention mechanism for parsing compounds, AR-Tree highlights the significant fragments of compounds by positioning them closer to the root of the tree structure. In this way, the identified significant fragments can be used to design new compounds with desired properties in future research. Additionally, to overcome the problem of exploring the vast chemical space, pre-training strategy is applied by using the PubChem database. On top of AR-Tree representations, we experiment with three different classifiers (i.e., XGBoost, Random Forest, and Fully Connected Neural Network Layers) on four different classification tasks of the MoleculeNet datasets as downstream tasks. The results of the experiments show that the proposed architecture succeeded in finding chemically meaningful fragments for the corresponding tasks.

**Keywords:** interpretability · compound · fragment.

* These authors contributed equally to the work.

## 1 Introduction

Determining the physicochemical and functional properties of novel compounds (e.g., blood-brain barrier penetration, toxicity, and lipophilicity) is an important step in the discovery of new drugs. However, this is time-consuming and costly as it requires complicated experiments in the laboratory. Therefore, computational methods (e.g., deep learning (DL)-based) have been developed to reduce the time and cost involved in these processes [20]. One of the most important parts of DL algorithms is learning molecular representations that should contain the most distinctive features in a compact form [16, 6]. However, the interpretability of the models is weak despite their impressive performance on known compounds, because it is very difficult to understand how the models make predictions and what types of patterns are captured by the model because the models work like a "black box."

In the literature, some approaches to calculate the feature importances have been proposed with the intent of interpretability. [34, 35, 45]. For Convolutional Neural Networks (CNN), to create saliency maps of the images that highlight the discriminative regions of the image, using gradient of the output with respect to the input has been proposed [35]. An earlier work proposes a probabilistic approach [45] to see the effects of features for each type of deep neural network (DNN). Here, the difference between the class scores of known and unknown feature sets is analyzed based on the conditional probabilities. As several forward passes of the data can be necessary, this kind of technique may not be efficient. In particular, to reduce the limitations of the gradient-based approaches when the gradients are close to zero. [34] explains the difference between the outputs of the neurons and some predetermined reference values in terms of the difference between the neurons' inputs and the predetermined reference inputs. One pass of backpropagation is sufficient to calculate the importance scores, but this approach requires domain-specific knowledge to determine the reference values.

In this work, we use an interpretation strategy with pre-trained Attentive Recursive Tree (AR-Tree) [33] to learn general compound representations. We then fine-tune the model to tailor these representations to downstream tasks. As a Tree-structured Long Short-Term Memory (Tree-LSTM)-based sentence embedding model, AR-Tree was developed to learn task-specific sentence embeddings considering the importance of fragments to the task at hand. It emphasizes the important fragments by placing them closer to the root of the tree, thanks to its task-specific attention mechanism for parsing sentences. In addition, owing to construction of a latent tree based on the importance of the fragments, it is also possible to interpret the embeddings created by the model. In other words, for each task, it is easy to identify which fragments are necessary for each activity by looking at their placements in the tree structure. This is an important aspect since understanding the crucial components of compounds can help experts make better predictions of physicochemical and functional features.

## 2 Materials and Methods

### 2.1 Tokenization

We model compounds as documents derived from a chemical language by using the representations of the Simplified Molecular Input Line Entry System (SMILES) [39] . To identify the words of this language, Byte-Pair Encoding (BPE) algorithm is employed [10], whose effectiveness for biomolecules has recently been shown [1, 19, 24]. The BPE strategy suggests that frequent subsequences in a large corpus can be meaningful language elements. Therefore, every character is first taken to represent a word, and BPE then examines the 2-character subsequences to increase its vocabulary by incorporating the most frequent subsequences. Then, this process is repeated, considering each element of the vocabulary as a single character, until the desired vocabulary size ($V$) is reached. When the BPE algorithm is finished, we obtain a vocabulary of size $V$ consisting of the most frequent units. We experiment with thirty-two thousands as the vocabulary size, since this size performs well in drug-target affinity prediction task [28].

### 2.2 Pre-training

DL models require large and well-representative data from the chemical space to learn generalizable representations. However, due to time and cost constraints, it is not feasible to obtain a large number of labeled data. Therefore, supervised DL models may not be able to learn task-specific representations due to data scarcity. To address this problem, pre-training approaches for DL models have been proposed [43, 44, 15, 4, 8, 42, 14, 38]. Although there are different pre-training strategies, we can briefly divide them into 2: self-supervised and supervised strategies.

For pre-training, we experiment with two different approaches. The first is a supervised method, molecular descriptor prediction (MD), where the model predicts surface descriptors, and the other is a self-supervised method, contrastive learning (CL), where the model tries to discriminate compounds. We used mean squared error (MSE) loss for the MD approach, and a variant of noise-contrastive estimation (NCE) loss is used for the CL approach, namely InfoNCE [27] with the implementation of [7].

### 2.3 Fine-tuning

We experimented with four different datasets from MoleculeNet [40] as downstream tasks. According to [40]; BBBP, ClinTox and Tox21 datasets are categorized as physiology and BACE dataset is categorized as biophysics.

Blood-Brain Barrier Penetration (BBBP) dataset contains compounds and their classification labels based on their ability to penetrate the blood-brain barrier (BBB). The dataset consists of 2039 compounds.

$\beta$-secretase (BACE) dataset contains compounds that inhibitors of human $\beta$-secretase 1(BACE-1). The dataset contains regression (the half maximal inhibitory concentration (IC50)) and classification binding labels of the compounds. The dataset consists of 1513 compounds.

Clinical Trial Toxicity (ClinTox) dataset contains compounds that have not been approved and approved by the Food and Drug Administration (FDA) due to toxicity. The dataset contains two classification tasks. In the experiments, the CT_TOX [40] task was used. The dataset consists of 1478 compounds.

Toxicology in the 21st Century (Tox21) dataset contains information on the toxicity of compounds according to different toxicity criteria. In the experiments, the SR_p53 [40] criterion was used to determine the toxicity labels of the compounds. The dataset consists of 7831 compounds.

### 2.4 Experimental Setup

For the molecular descriptor prediction, the RDKit library [22] is used for the calculations of surface descriptor subset which consists of 48 molecular descriptors. In the CL approach, we use the same data and create positive examples via SMILES augmentation. For the training, validation and test splits of the downstream tasks, the same splits with [38] are preferred in the sense of performance comparison.

In addition to the pre-trained models, the model is also tested without pre-training with different hyperparameters. In the search for hyperparameters, 32 and 64 for batch size, 300 and 500 for the number of hidden dimensions of the tree structure, and 0.3 and 0.5 for the dropout ratio were tested in all experiments. Among the hyperparameter combinations, 32 for batch size, 500 for the number of hidden dimensions of the tree structure, and 0.3 for the dropout ratio were found to be the best hyperparameters based on validation performance. Bidirectional LSTM [13], ST Gumbel [18] and Adam (learning rate is 0.003) [21] are preferred as RNN type, model type and optimizer. The remaining hyperparameters of the model are the default parameters. The PyTorch library [29] is used to build the model. The training was continued with 25 epochs.

In the experiments, 3 different classifiers are used on top of tree representations which is basically a 300 dimensional vector. After training AR-Tree, XGBoost classifier [3] and Random Forest classifiers are trained from scratch by using AR-Tree representations of the compounds. No second training was required for the final classifier, which consists of fully connected neural network layers, since the layers can be directly connected to AR-Tree.

For XGBoost classifier, the learning rate and the number of estimators were set to 0.03 and 600, respectively, and the XGBoost library [3] was used. For Random Forest classifier, the number of estimators was set to 600 and the Scikit-learn

library [31] was used. For fully connected neural network layers, 3 dense layers and 1 output layer are connected in series. The number of hidden dimensions of the layers was set to 1024 and the PyTorch library [29] was used. The remaining hyperparameters of the classifiers are the default parameters.

## 3 Results

The cross-entropy loss function is employed as the training objective because all experiments were performed with classification tasks. To evaluate the performance of the model, we used the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) as the metric. The ROC-AUC scores of our model and the state-of-the-art models [38] are shown in **Table 1**. The results show that our model outperformed the state-of-the-art models in ClinTox task and achieved moderate scores in the other downstream tasks.

**Table 1.** ROC-AUC scores of the down-stream tasks. CL, MD, and - indicate contrastive learning, molecular descriptor prediction, and without pre-training, respectively. Models in the first 9 rows are the models that we suggest in this paper. The results of the other models are retrieved from [38].

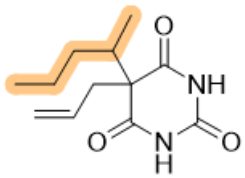
| Pretrain Task | Model | %BBBP | %BACE | %ClinTox | %Tox21 |
|---|---|---|---|---|---|
| - | AR-Tree-FC | 67.5 | 66.6 | 99.2 | 59.9 |
| CL | AR-Tree-FC | 66.2 | 70.4 | 94.3 | 66.2 |
| MD | AR-Tree-FC | 68.6 | 71.3 | 98.9 | 67.4 |
| - | AR-Tree-XGB | 67.7 | 66.8 | 98.9 | 62.1 |
| CL | AR-Tree-XGB | 67.4 | 67.8 | 94.6 | 62.4 |
| MD | AR-Tree-XGB | 67.7 | 67.2 | 99.3 | 64.0 |
| - | AR-Tree-RF | 68.0 | 67.3 | 99.3 | 61.9 |
| CL | AR-Tree-RF | 68.6 | 71.6 | 94.5 | 62.7 |
| MD | AR-Tree-RF | 68.5 | 68.8 | 93.9 | 66.5 |
| - | RF | 71.4 | 86.7 | 71.3 | 76.9 |
| - | SVM | 72.9 | 86.2 | 66.9 | 81.8 |
| - | GCN | 71.8 | 71.6 | 62.5 | 70.9 |
| - | GIN | 65.8 | 70.1 | 58.0 | 74.0 |
| - | SchNet | 84.8 | 76.6 | 71.5 | 77.2 |
| - | MGCN | 85.0 | 73.4 | 63.4 | 70.7 |
| - | D-MPNN | 71.2 | 85.3 | 90.5 | 68.9 |
| MD | Hu et al. [15] | 70.8 | 85.9 | 78.9 | 78.7 |
| MD | N-Gram | 91.2 | 87.6 | 85.5 | 76.9 |
| MD | MolCLR(GCN) | 73.8 | 78.8 | 86.7 | 74.7 |
| MD | MolCLR(GIN) | 73.6 | 89.0 | 93.2 | 79.8 |

## 4 Interpretability

The fragments should be rated using a scoring process in order to evaluate the model's interpretability. In the scoring procedure, each leaf is first assigned 1 point and each node is assigned a point increasing by 1 point node-by-node from the leaves to the root. In this way, the root receives the maximum point. This maximal point is then divided by each node's points so that all of the outcomes fall between 0 and 1. Among the compounds in the corresponding dataset, the points of each fragment are added up. After summation, the total point of each fragment is divided by the number of the compounds that the fragment is in it. Thus, once more, all of the outcomes fall inside the range of 0 and 1. Finally, all fragments are sorted in descending order and the first 50 fragments are examined by using Structure-Activity Relationship (SAR) analysis.

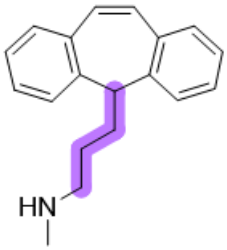
The fragments, the names of the corresponding compounds, and the results of the SAR analyses for different tasks are presented in the following subsections. According to the model, the colored fragments were determined to be the most meaningful fragments for the corresponding tasks. Unfortunately, the model could not find any meaningful fragment for BACE task, as revealed by the SAR analysis.

### 4.1 BBBP SAR Analysis



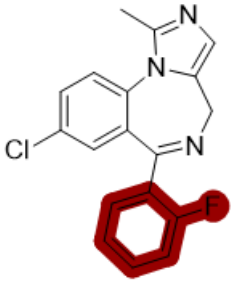

Secobarbital

Hydrogen atoms at the position 5- of barbituric acid cause the molecule to become acidic, thereby preventing a sedative-hypnotic effect. For maximum activity, at least one of the hydrogen atoms located at the position 5- must be replaced with different substituents, including an alkyl chain. Thus, the 2-pentyl substituent at the position 5- increases the lipophilicity of the molecule and helps it pass the BBB to affect the central nervous system and exhibit the necessary sedative-hypnotic effect [2, 23, 17].



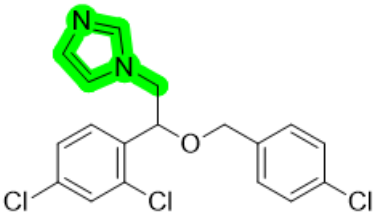

Protriptyline

In tricyclic anti-depressants, the alkyl chain connecting the dibenzocycloheptene ring to the secondary amine group should consist of 3 carbon atoms. If the alkyl chain contains 4 carbons and branches, the activity decreases. Although the dibenzocycloheptene structure is the most important structure for the antidepressant effect, the 'CCCC1' fragment selected by the algorithm is an important substituent in terms of lipophilicity and selectivity to the receptor [32, 25].



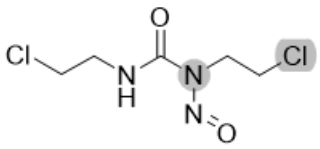

Midazolam

The substitution of the benzodiazepine ring with an ortho-halogenated phenyl group at the position 5- is known to increase lipophilicity, efficacy, and sedative potential, especially in long-acting compounds [12].



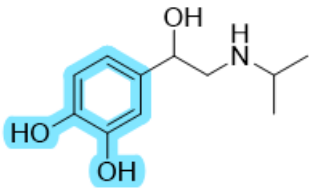

Econazole

In econazole, the highlighted part selected by the algorithm denotes the essential imidazole ring system. Antifungal azoles carry triazole and imidazole rings [41].



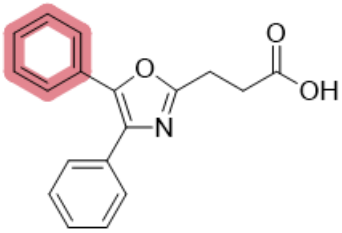

Carmustine

Antineoplastic compounds carrying a nitrosourea structure cross the BBB due to their high lipophilicity and are used in the treatment of brain tumors. The nitrosourea group converts to isocyanates and binds to the DNA as a carbamoylating group. 2-Chloroethyl groups are convert first to ethyleneimmonium (aziridinium) derivatives and then to carbonium ions and, bind to DNA as alkylating groups [26].




Isoprenaline

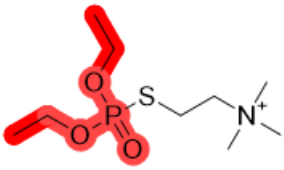
Isoprenaline carries a catechol ring. It has been determined in the literature, that the catechol ring binds to beta receptors with hydrogen bond, thanks to its hydroxy groups [30].




Oxaprozine

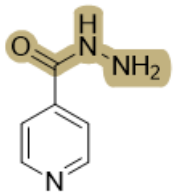
The key pharmacophore group of nonsteroidal anti-inflammatory drugs (NSAIDs) is the propionic acid. The benzene ring chosen by the algorithm increases the lipophilicity of the molecule, and binds to the flat area in the cyclooxygenase (COX) enzyme with Van der Waals bonds [30].

### 4.2 ClinTox SAR Analysis



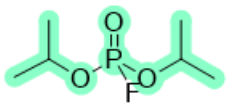

Ecothiophate

In organophosphorus compounds, the diethylphosphoryl group of the cholinesterase inhibitor, ecothiophate, is the part responsible for its irreversible binding to the hydroxy group of serine [9].




Izoniazide

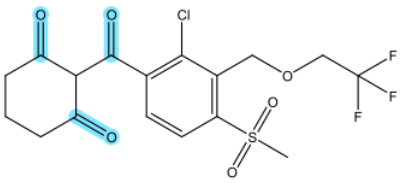
The hydrazide group of the isoniazid selected by the algorithm is converted into strong acylating radicals by the enzyme catalase-peroxidase in mycobacteria. Mycolic acid synthesis and thus cell wall synthesis are inhibited. The major metabolite of the isoniazid is N-acetyl isoniazid. Acetyl radicals formed after the advanced metabolization stages are hepatotoxic [37].




Fluostigmine

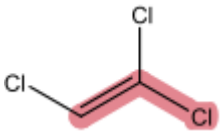
Due to its selected moiety, diisopropyl phosphoryl fluoride is used as a pesticide and is an acetylcholinesterase inhibitor [9].

### 4.3 Tox21 SAR Analysis



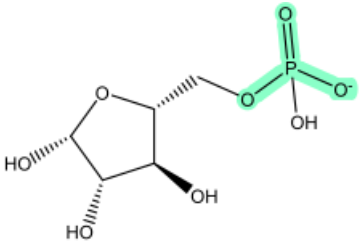

Tembotrione

The carbonyl groups found in the $\beta$-triketone herbicide, tembotrione, contribute to the formation of more toxic chlorination by-products through a haloform-like mechanism since the $\alpha$-carbon of the three carbonyl groups is highly acidic, which leads to the production of chloroform and other toxic chlorophenolic compounds during the purification process of drinking water with chlorine [36].




Trichloroethylene

The chlorovinyl group of the industrial solvent, trichloroethylene, which is known to be metabolized by oxidative and conjugative pathways, bears a significant role in the transformation of the molecule to genotoxic chloral hydrate. Mutagenic S-(1,2-dichlorovinyl)glutathione metabolite in a mechanism which involves the substitution of sulfur atom in glutathione with the leaving chlorine atom of the chloroethene moiety [5].




$\beta$-L-arabinofuranose

The phosphonate moiety selected at $\beta$-L-arabinofuranose can be found at the active form of many anticancer drugs. Antimetabolite effective chemotherapeutics bearing this structure convert to triphosphate forms and cause cytotoxic effects by inhibiting enzymes like DNA polymerase [11].

## 5 Conclusion

We here offer an interpretation strategy for compounds and their fragments using the AR-Tree model. Thanks to the task-specific attention mechanism of the model, it can highlight the important fragments for the corresponding tasks. The results of the experiments show that our approach outperforms the state-of-the-art models in ClinTox task. In terms of interpretability, the model succeeded in finding chemically meaningful fragments in the compounds for BBBP, ClinTox and Tox21 tasks.

Determining the most significant fragments for the given tasks is an important feature because when predicting physicochemical and functional properties, recognizing the important parts of compounds can give insights to experts for experimental studies. Our approach allows to construct novel compounds with desired features more precisely and hopefully, by this way, individuals suffering from various incurable diseases may find the necessary compounds.

## Acknowledgments

This work is supported by Scientific and Technological Research Council of Turkey (TÜBİTAK) [Grant Numbers 119E133].

**Conflict of Interest:** E.O. is an employee of Roche AG.

## References

1. Asgari, E., McHardy, A.C., Mofrad, M.R.: Probabilistic variable-length segmentation of protein sequences for discriminative motif discovery (dimotif) and sequence embedding (protvecx). Scientific reports **9**(1), 1–16 (2019)
2. Bialer, M.: How did phenobarbital's chemical structure affect the development of subsequent antiepileptic drugs (aeds)? Epilepsia **53**, 3–11 (2012)
3. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016)
4. Chithrananda, S., Grand, G., Ramsundar, B.: Chemberta: Large-scale self-supervised pretraining for molecular property prediction. arXiv preprint arXiv:2010.09885 (2020)
5. Cichocki, J.A., Guyton, K.Z., Guha, N., Chiu, W.A., Rusyn, I., Lash, L.H.: Target organ metabolism, toxicity, and mechanisms of trichloroethylene and perchloroethylene: key similarities, differences, and data gaps. Journal of Pharmacology and Experimental Therapeutics **359**(1), 110–123 (2016)
6. David, L., Thakkar, A., Mercado, R., Engkvist, O.: Molecular representations in ai-driven drug discovery: a review and practical guide. Journal of Cheminformatics **12**(1), 1–22 (2020)
7. Elbers, R.: info-nce-pytorch. https://github.com/RElbers/info-nce-pytorch (2021)

8. Fabian, B., Edlich, T., Gaspar, H., Segler, M., Meyers, J., Fiscato, M., Ahmed, M.: Molecular representation learning with language models and domain-relevant auxiliary tasks. arXiv preprint arXiv:2011.13230 (2020)
9. Fukuto, T.R.: Mechanism of action of organophosphorus and carbamate insecticides. Environmental health perspectives **87**, 245–254 (1990)
10. Gage, P.: A new algorithm for data compression. The C Users Journal **12**(2), 23–38 (1994)
11. Gandhi, V., Plunkett, W.: Cellular and clinical pharmacology of fludarabine. Clinical pharmacokinetics **41**(2), 93–103 (2002)
12. Ghose, A.K., Crippen, G.M.: Modeling the benzodiazepine receptor binding site by the general three-dimensional structure-directed quantitative structure-activity relationship method remotedisc. Molecular pharmacology **37**(5), 725–734 (1990)
13. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional lstm networks. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005. vol. 4, pp. 2047–2052. IEEE (2005)
14. Guo, Z., Sharma, P., Martinez, A., Du, L., Abraham, R.: Multilingual molecular representation learning via contrastive pre-training. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 3441–3453 (2022)
15. Hu, W., Liu, B., Gomes, J., Zitnik, M., Liang, P., Pande, V., Leskovec, J.: Strategies for pre-training graph neural networks. arXiv preprint arXiv:1905.12265 (2019)
16. Huang, B., Von Lilienfeld, O.A.: Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. The Journal of Chemical Physics **145**(16), 161102 (2016)
17. J Ernst, B., F Clark, G., Grundmann, O.: The physicochemical and pharmacokinetic relationships of barbiturates–from the past to the future. Current Pharmaceutical Design **21**(25), 3681–3691 (2015)
18. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax (2017)
19. Kawano, K., Koide, S., Imamura, C.: Seq2seq fingerprint with byte-pair encoding for predicting changes in protein stability upon single point mutation. IEEE/ACM transactions on computational biology and bioinformatics **17**(5), 1762–1772 (2019)
20. Kim, J., Park, S., Min, D., Kim, W.: Comprehensive survey of recent drug discovery using deep learning. International Journal of Molecular Sciences **22**(18), 9983 (2021)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
22. Landrum, G.: Rdkit documentation. Release **1**(1-79), 4 (2013)
23. Langer, M.K., Rahman, A., Dey, H., Anderssen, T., Zilioli, F., Haug, T., Blencke, H.M., Stensvåg, K., Strøm, M.B., Bayer, A.: A concise sar-analysis of antimicrobial cationic amphipathic barbiturates for an improved activity-toxicity profile. European Journal of Medicinal Chemistry **241**, 114632 (2022)
24. Li, X., Fourches, D.: Smiles pair encoding: A data-driven substructure tokenization algorithm for deep learning. Journal of Chemical Information and Modeling **61**(4), 1560–1569 (2021)
25. Maxwell, R.A., White, H.L.: Tricyclic and monoamine oxidase inhibitor antidepressants: Structure-activity relationships. In: Handbook of psychopharmacology, pp. 83–155. Springer (1987)
26. Montgomery, J.A.: Chemistry and structure-activity studies of the nitrosoureas. Cancer Treat Rep **60**(6), 651–664 (1976)

27. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
28. Özçelik, R., Öztürk, H., Özgür, A., Ozkirimli, E.: Chemboost: A chemical language based approach for protein – ligand binding affinity prediction. Molecular Informatics **40**(5), 2000212 (2021). https://doi.org/https://doi.org/10.1002/minf.202000212, https://onlinelibrary.wiley.com/doi/abs/10.1002/minf.202000212
29. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
30. Patrick, G.L.: An introduction to medicinal chemistry. Oxford university press (2013)
31. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
32. van Schaik, N., Graf, U.: Structure-activity relationships of tricyclic antidepressants and related compounds in the wing somatic mutation and recombination test of drosophila melanogaster. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **286**(2), 155–163 (1993)
33. Shi, J., Hou, L., Li, J., Liu, Z., Zhang, H.: Learning to embed sentences using attentive recursive trees (2018)
34. Shrikumar, A., Greenside, P., Kundaje, A.: Learning important features through propagating activation differences (2019)
35. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps (2014)
36. Tawk, A., Deborde, M., Labanowski, J., Gallard, H.: Chlorination of the $\beta$-triketone herbicides tembotrione and sulcotrione: kinetic and mechanistic study, transformation products identification and toxicity. Water Research **76**, 132–142 (2015)
37. Wang, P., Pradhan, K., Zhong, X.b., Ma, X.: Isoniazid metabolism and hepatotoxicity. Acta pharmaceutica sinica B **6**(5), 384–392 (2016)
38. Wang, Y., Wang, J., Cao, Z., Barati Farimani, A.: Molecular contrastive learning of representations via graph neural networks. Nature Machine Intelligence **4**(3), 279–287 (2022)
39. Weininger, D.: Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. Journal of chemical information and computer sciences **28**(1), 31–36 (1988)
40. Wu, Z., Ramsundar, B., Feinberg, E.N., Gomes, J., Geniesse, C., Pappu, A.S., Leswing, K., Pande, V.: Moleculenet: A benchmark for molecular machine learning (2018)
41. Xu, X., Xu, B., Wang, X., Li, J.: Quantitative structure-activity relationships study of a series of niacinamide analogues as androgen receptor antagonists. Journal of Molecular Structure **1201**, 127128 (2020)
42. Xu, Z., Wang, S., Zhu, F., Huang, J.: Seq2seq fingerprint: An unsupervised deep molecular embedding for drug discovery. In: Proceedings of the 8th ACM international conference on bioinformatics, computational biology, and health informatics. pp. 285–294 (2017)

43. Xue, D., Zhang, H., Xiao, D., Gong, Y., Chuai, G., Sun, Y., Tian, H., Wu, H., Li, Y., Liu, Q.: X-mol: large-scale pre-training for molecular understanding and diverse molecular analysis. bioRxiv pp. 2020–12 (2021)
44. Zhu, J., Xia, Y., Qin, T., Zhou, W., Li, H., Liu, T.Y.: Dual-view molecule pre-training. arXiv preprint arXiv:2106.10234 (2021)
45. Zintgraf, L.M., Cohen, T.S., Adel, T., Welling, M.: Visualizing deep neural network decisions: Prediction difference analysis (2017)