DOMAIN-AWARE REPRESENTATION OF SMALL MOLECULES FOR EXPLAINABLE PROPERTY PREDICTION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

The advances in deep learning algorithms have impacted the drug discovery pipeline in many ways. Specifically, generative artificial intelligence (AI) algorithms can now explore the large chemical space and design novel and diverse molecules. While there has been significant progress in generative AI models, it is equally important to develop predictive models for various properties, which can help to characterize novel drug-like molecules. Further, the predictive model acts as a critic to design multi-property optimized molecules, which can potentially reduce the late stage attrition of drug candidates. Nevertheless, understanding the reason behind model predictions can guide the medicinal chemist to modify substructures that can make the molecules undesirable during the lead optimization stage of drug discovery. However, current explainable approaches are mostly atom-based where, often only a fraction of a fragment is shown to be significant. To address the above challenges, we have developed a novel domain-aware fragment-based graph input representation based on a molecular fragmentation approach termed pBRICS, which can fragment small molecules into their functional groups. Both single and multi-task models were developed to predict various properties including ADMET properties. The fragment level explainability were obtained using the Grad-CAM approach. The method was further validated with the available Matched Molecular Pairs (MMP) for blood brain barrier permeability (BBBP) and Ames Mutagenicity.

1 INTRODUCTION

The advent of AI-based algorithms has revolutionized the complete drug design process. There has been a significant progress in the field of generative AI to design novel drug-like molecules by effectively exploring the chemical space (Segler et al. (2018)Olivecrona et al. (2017); Pope et al. (2018); Bung et al. (2021); Born et al. (2020); Krishnan et al. (2021a); Krishnan et al. (2021b)). Some studies have shown the potential of these models to truly accelerate the early-stage drug discovery process and design drug-like molecules in few weeks (Zhavoronkov et al. (2019); Stokes et al. (2020)). Although generation of novel drug-like small molecules specific to a target protein of interest is highly essential to explore the potentially vast chemical space, it is also necessary to optimize various physicochemical and late stage properties including absorption, distribution, metabolism, excretion and toxicity (ADMET) to reduce the attrition rate of novel molecules during the clinical trial process (Daoud et al. (2021); Bung et al. (2022)). The ability to accurately predict various ADMET properties can provide insights right from the administration of a molecule to its elimination, along with its probable pharmacological action in the human body (Cumming et al. (2013); Van De Waterbeemd & Gifford (2003); Selick et al. (2002)). Studies have shown that, around 50% of potential drug candidates fail while being evaluated for their pharmacokinetic properties, leading to a marked increase in the time taken to produce admissible drug molecules (Kola & Landis (2004); Waring et al. (2015)). However, the recent incorporation of AI-based models to predict drug potency and various other properties in parallel to the drug design phase, can eliminate undesirable drug candidates in the early stages of the drug discovery pipeline (Siramshetty et al. (2021); Blay et al. (2022)). One of the challenges for such predictive models is to differentiate between highly similar molecules that can be categorized into different classes depending on their property profile.

Even the best performing models have been shown to mispredict such cases and lead to a higher number of false positives and false negatives. Consequently, there has been considerable interest, to validate various machine and deep learning models by assessing their performance on classifying similar molecules with very different potency, also known as activity cliffs (van Tilborg et al. (2022); Tamura et al. (2023)).

Apart from measuring the performance on similar molecules, the interpretability of model predictions can accelerate the lead optimization stage of drug discovery (He et al. (2021); Vangala et al. (2022); Baptista et al. (2022); Tian et al. (2022)). The ability to quantify the contribution of each substructure or molecular fragment towards the property being predicted, can guide the medicinal chemists to optimize the lead molecules. The choice of input representation of molecules can influence what the model learns and consequently the explanations obtained for model predictions. Popular input representations for development of predictive models include, the Simplified Molecular Input Line Entry System (SMILES), atomic-level molecular graphs, fingerprints based on functional group frequency, and descriptor-based feature vectors such as PaDEL (Yap (2011)) and Mordred (Moriwaki et al. (2018)). For models developed using each of the above input representations, corresponding model explanations have also been extracted using gradient-based explainability methods such as Gradient-based Class Activation Mapping (Grad-CAM) and saliency maps, and non-gradient explainability methods such as SHAPley scores (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). Although explanations from molecular graph representations, typically involving coloring schemes to highlight the magnitude of gradient change, are in good agreement with the experimental knowledgebase of medicinal chemists, these methods tend to produce partial explanations in molecules involving ring systems (Jiménez-Luna et al. (2021)). To remedy this, functional-group level molecular graph representations could be a solution by appropriate fragmentation of molecules, and reconstitution of the molecular graph with functional groups as nodes and fragmented bonds as edges (Zhang et al. (2021b); Zhang et al. (2021a); Hajiabolhassan et al. (2022)). However, most of the existing fragmentation methods such as BRICS, RECAP and SynDIR (Firth et al. (2015)) lead to singleton atoms upon fragmentation, cannot handle very small molecules, and fail to fragment all substituents of a core scaffold in a molecule. Specifically, some substituents including halogen atoms, methyl and hydroxyls groups can be considered valid fragments of interest to medicinal chemists, due to the possiblilty of extensive change in the physicochemical and ADMET properties of a molecule upon their substitution (Gentry et al. (1999); Cramer et al. (2019)). The above limitations necessitate the development of a novel fragmentation method that could address the above challenges, and could be used as input to train deep learning models by capturing a chemist's perception.

In this study, we have addressed two challenges through the development of a novel molecular representation: 1) training property prediction models on standardized datasets, and 2) using gradient-based explainability methods to quantify fragment-level contributions to the property prediction. We have developed a novel molecular fragmentation method termed pBRICS for finegrained fragmentation of any novel small molecule. With the fragments of pBRICS, a molecular graph representation is proposed, which was used to train graph convolution network (GCN) models for prediction of over 23 different ADMET properties. The models developed were compared with an existing atom-level graph model from literature trained using a similar architecture. Further, the explanations obtained from the predictive models were validated based on their relevance to the existing knowledgebase of substructure contributions using matched molecular pairs (MMP) analysis.

2 Materials and Methods

2.1 Post-processing BRICS (pBRICS) fragmentation method

In this work, we have introduced a novel fragmentation method named pBRICS. The pBRICS method performs a post-processing of fragmentation results from the commonly used BRICS (Breaking of Retrosynthetically Interesting Chemical Substructures) method (Degen et al. (2008)). BRICS method considers 16 chemical environments and their corresponding fragment prototypes for fragmentation of every bond in a molecule. Due to the expanded set of cleavage rules possible for a molecule, the BRICS method has been shown previously (Degen et al. (2008)) to fragment 10% more molecules than the RECAP (REtrosynthetic Combinatorial Analysis Procedure) fragmentation method (Lewell et al. (1998)). pBRICS attempts to iteratively fragment the molecule such that, the

smallest possible substituents of a scaffold are also enumerated, which can be of interest in building a knowledgebase for molecular optimization problems.

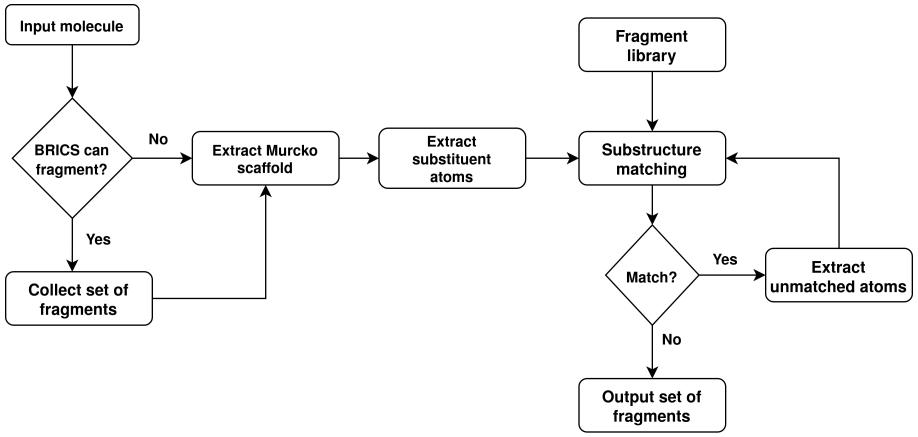

Figure 1: Flowchart of the pBRICS fragmentation method.

pBRICS method uses a comprehensive library of fragments (ring systems and non-ring substituents) manually curated from the Open Chemistry project database and literature. If a molecule can be fragmented by the BRICS method, the resultant fragments are primarily classified into scaffolds and substituents. The scaffold of a molecule or a fragment is defined based on the Bemis-Murcko frameworks (Bemis & Murcko (1996)) as implemented in RDKit. The atoms corresponding to substituents are further matched against the library of fragments to obtain the largest possible fragment match, and the process is iterated until no further fragment matches are possible. If a molecule cannot be fragmented by the BRICS method, the atoms corresponding to scaffold and substituents of the molecule are enumerated and the iterative fragmentation procedure is applied on the substituent atoms, instead of fragments from the BRICS method (Fig. 1). The pBRICS method was implemented in Python using utilities from the RDKit library.

2.2 Graph construction from pBRICS fragmentation of a molecule

The fragments obtained from pBRICS were used to construct a domain-aware fragment-level graph representation. For generating the graph, fragments were considered as nodes and an edge was defined between two nodes, if the two fragments were connected in the molecule by at least one bond (see Fig. 2). Once the graph construction is complete, the next step is to define the feature

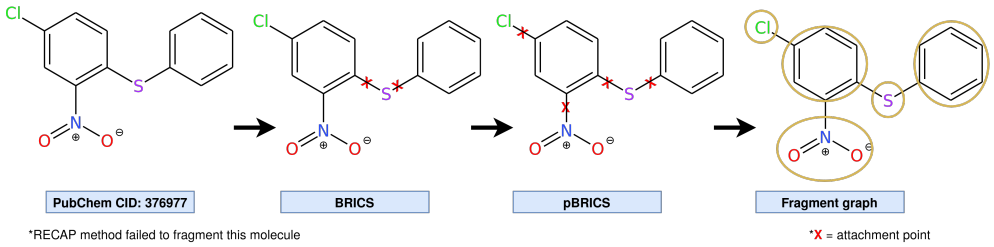

Figure 2: Graph construction process from pBRICS fragments and subsequent fragment graph formation. The post processing of BRICS allowed the pBRICS method to further divide the molecules into smaller and chemically interesting fragments.

vector for each node in the graph. The MACCS (Molecular ACCess System) 167 keys (Durant et al. (2002)) and ECFP2 (Extended-connectivity fingerprints) 256 bit fingerprint (Rogers & Hahn (2010)) representation of fragment (node) are considered as the feature vector for each node in the graph. The type of bond connecting the fragments is encoded as a value, to be used as edge weight while aggregating the information from neighbours.

2.3 DATASET CURATION

To validate the method, the datasets from the ADMETLab2.0 study (Xiong et al. (2021)) were curated. The curated dataset consists of data associated with 23 pharmacokinetic properties. For each property, the data was split into 80% for the training set, 10% for the validation set and remaining 10% for the test set for training the models. The individual datasets for these 23 properties were further combined to form a multi-task dataset to be used for training a multi-task property prediction model.

2.4 GRAPH CONVOLUTIONAL NETWORK (GCN)

The domain-aware graph generated using the pBRICS fragmentation method was used as input to the deep learning models involving Graph Convolutional Network (GCN) (Kipf & Welling (2016)) layers as implemented in the Deep Graph Library (Wang (2019)). The GCN-based architecture has been proven to yield promising results for predicting a wide range of molecular characteristics (Wieder et al. (2020)). Typically, a graph is defined as $G = (V, E)$, wherein fragments are nodes (V) and bonds connecting them are edges (E). GCN layer modifies each node's embedding in the following way: 1) Using the node feature matrix $X \in F^{n \times d}$ and an adjacency matrix $A \in \{0, 1\}^{n \times n}$, GCN aggregates the information from neighbouring nodes (Buffelli & Vandin (2020)). 2) Applies a non-linear activation function to the aggregated feature vector embedding of nodes (Buffelli & Vandin (2020)) to get the prediction.

2.5 RELATIONAL GRAPH CONVOLUTIONAL NETWORK (RGCN)

The Relational Graph Convolutional Network (RGCN) (Schlichtkrull et al. (2018)) was introduced to capture the relations among different nodes of the complex graph unlike in GCN network only one relation that exists between nodes (whether the node is a neighbour or not). An RGCN layer transforms the embedding of each node almost similar to a GCN layer with one major change in the way the aggregation happens. RGCN layer aggregates the information from neighbouring nodes (or atoms), via an adjacency matrix $A \in \{0, 1\}^{n \times n \times r}$ where, r corresponds to each relation in the graph, and a node feature matrix $X \in F^{n \times d}$. The RGCN layer was implemented using the Deep Graph Library (DGL) (Wang (2019)).

2.6 CONSTRUCTING THE MULTI-TASK GRAPH NEURAL NETWORK FRAMEWORK

The multi-task graph neural network model (Fig. 3) was prepared so as to benefit from the related tasks and improve the prediction for the task with low data. Nevertheless, single task models were also built to quantify the difference in performance compared to the multi-task model. The multi-task model can be used for predicting both classification and regression properties simultaneously, unlike the single task predictive models where individual models have to be built independently for each task of interest. Recent studies have shown that multi-task models can outperform single task models, as the hidden layers are shared among all the tasks thereby, helping the model to learn more about the related tasks, which will improve the performance for the task in hand. In the multi-task model, a batch of fragment graphs are first passed into a 2-layered RGCN/GCN each of size 64. The aggregated feature vector from the final RGCN/GCN layer for each graph is multiplied by their respective attention weight, and the resultant vector is fed into an independent stack of fully connected (FC) layers (one FC stack per property). Here, all FC stacks share an aggregated feature vector from the final RGCN/GCN layer, enabling the model to learn common features between all combinations of properties involved. The architecture of multi-task model can be converted to single task models by considering only one FC stack in the final layer instead of 23, as in case of the multi-task model.

2.7 EVALUATION METRICS

The binary cross-entropy-loss (BCE) between the actual values and model predictions is calculated for each of the 23 classification tasks mentioned in Table 2. The loss is back-propagated to update the weights of the neurons in the fully connected layers and GCN layers. Area Under Receiver Operating System Curve (auROC) was used for measuring the performance of the classification

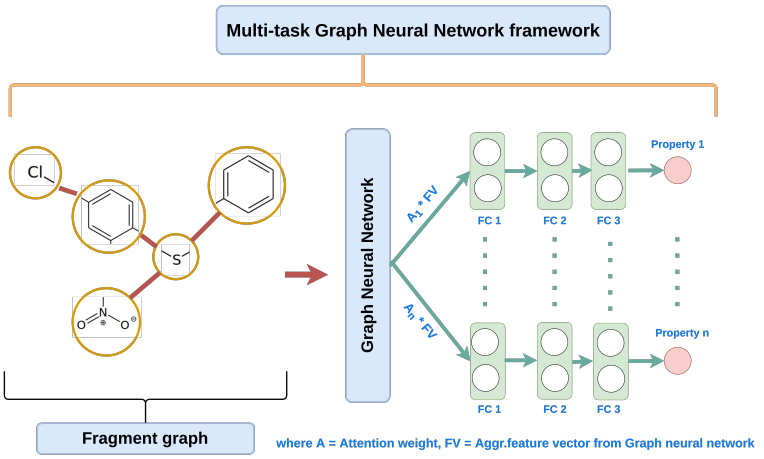

Figure 3: Multi-task graph neural network framework proposed in this study for simultaneous prediction of 23 pharmacokinetic properties.

models in this study.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)) \tag{1}$$

$N : number\_of\_samples, \hat{y} : predicted\_value, y : true\_value$

2.8 Explainability of model predictions using Gradient Class Activation Map (Grad-CAM)

To interpret the predictions obtained from the deep learning model, the Gradient Class Activation Map (Grad-CAM) method was used. The Grad-CAM method was initially developed for analysing convolutional neural network (CNN) models, and was later adapted to work with graph convolutional networks (GCN) (Pope et al. (2018)). Here, consider c to be the class for which explanations are generated, L to be the final GCN layer of the model, $\alpha_k^{L,c}$ to be the Grad-CAM weight for the k^{th} feature of class c at layer L, N to be the number of nodes, and $F_{k,n}^L$ be the feature map of k^{th} feature map at layer L of node n:

$$\alpha_k^{L,c} = \frac{1}{N}\sum_{n=1}^{N}\frac{\partial y^c}{\partial F_{k,n}^L} \tag{2}$$

Node level importance scores for a graph obtained using the Grad-CAM method are given by the following equation 2. Based on these scores, one can identify the fragments (nodes) contributing positively to the class c.

$$L_{Grad-CAM}^c[L, n] = \sum_k \alpha_k^{L,c} F_{k,n}^L(X, A) \tag{3}$$

2.9 Validation of the pBRICS method with Matched Molecular Pairs (MMP) analysis

Matched Molecular Pairs (MMP) (Kenny & Sadowski (2005)) are experimentally validated pair of molecules that differ by a single substitution of functional group/atom, which leads to conflicting chemical properties. Since one of the major challenges for these predictive models is to be able to correctly predict similar molecules that belong to conflicting classes in a classification problem, MMPs provide an opportunity to validate such models by testing their ability to differentiate similar molecules belonging to opposite classes. Further, by extracting the corresponding reasoning from explainable AI methods, model predictions can also be justified.

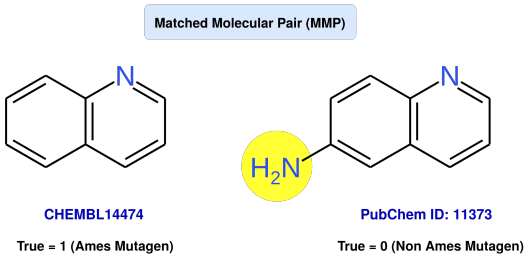

Figure 4: An example of a matched molecular pair (MMP) from the ChEMBL database. where,The the substitution (here, Hydrogen to amino) make the molecule pair active to inactive or vice-versa. which acts as the MMP transformation for this pair is highlighted in yellow circle.

Table 1: Comparison of the performance metrics between different fragmentation methods on the ChEMBL database.

Metric	pBRICS	BRICS	RECAP	SynDiR
No.of uncut molecules	5,908	55,677	152,177	97,823
Total no.of fragments generated (with duplicates)	11,913,61	8,328,962	4,964,387	5,672,031
No of unique fragments	120,552	242,208	242,208	342, 263
Average heavy atom count for the unique fragments	12,201	15,593	15,911	17,693

3 Results and Discussion

3.1 Benchmarking pBRICS against other fragmentation methods

The pBRICS fragmentation approach proposed in this study was compared with the existing fragmentation methods namely, BRICS, RECAP and SynDiR, in terms of four metrics defined earlier (Firth et al. (2015)) with the ChEMBL dataset as the benchmark dataset (Table 1). Based on the comparison results it is notable that, pBRICS generates the least number of unique fragments with smallest heavy atom count, resulting in a more generic fragment library due to the fine-grained fragmentation procedure developed. Further, pBRICS can fragment 16-times, 5-times and 10-times more molecules than RECAP, BRICS and SynDiR methods, respectively. Several examples were identified wherein, only pBRICS could fragment the molecule of interest (see SI Fig. S1).

3.2 Performance of the multi-task property prediction model

As mentioned in the methods section, both multi-task and single task-models were trained for predicting 23 properties that are crucial for small molecules to be considered as drug-like molecule. The multi-task and single-task predictive models trained using the domain-aware molecular graph representation based on the pBRICS method and GCN are referred to as MT-FraGCN and ST-FraGCN models, respectively. And MT-FraRGCN in Table 2 referred to pBRICS methods with RGCN. The performance of multi-task models on the 23 properties is summarized in Table 2.

The performance of the domain-aware fragment-based multi-task model was compared with the atom-level graph-based multi-task model from ADMETLab2.0. The ADMETLab2.0 model was retrained on the multi-task property prediction dataset, using an atom-level graph-based input representation. MT-FraRGCN model after hyperparameter tuning achieved a mean auROC of 84.47%, which is 1.12% higher than that of the ADMETLab2.0 model (83.35%) and 1.53% higher than that of the MT-FraGCN model (82.94%). The increase in performance of the MT-FraRGCN model can be attributed to the different relations it considers among fragments as shown in ADMETLab2.0 such as, bond type, type of atoms forming the bond, and stereochemistry. For each property provided in Table 2, a separate single task GCN (ST-FraGCN) model was also trained and compared with the multi-task model. For the 23 properties considered in the study, it was observed that there is no significant difference in the performance of single-task and multi-task models. To validate the

Table 2: Comparison of the performance of different models developed in this study with that of the multi-task (MT) model developed with the ADMETlab2.0 source code (Xiong et al. (2021)).

Property	MT-FraGCN	MT-FraRGCN	ST-FraGCN	ADMETLab2.0
Ames Mutagenicity	87.75	89.42	88.18	86.57
Carcinogenicity	63.97	74.98	67.65	71.62
CYP1A2	92.31	95.44	93.41	92.65
CYP2C19	79.91	81.11	80.82	80.67
CYP2C9	83.65	84.24	84.01	84.91
CYP2D6	84.91	85.41	81.36	87.03
CYP3A4	90.13	88.45	89.24	89.57
Eye Corrosion	98.53	99.04	97.73	97.89
Eye Irritation	96.14	95.93	95.44	96.13
Respiratory Toxicity	83.28	89.41	82.19	81.07
BBBP	88.81	89.75	88.21	88.57
NR-AR	75.71	74.31	76.07	77.72
NR-AR-LBD	79.51	84.58	79.36	76.46
NR-AhR	85.04	85.51	85.36	85.72
NR-Aromatase	83.38	81.85	76.93	77.71
NR-ER	66.35	68.21	67.71	67.38
NR-ER-LBD	81.78	84.58	79.64	83.84
NR-PPAR-gamma	73.86	79.85	82.27	79.40
SR-ARE	79.32	78.59	76.65	80.47
SR-ATAD5	86.46	81.61	82.78	82.72
SR-HSE	74.43	78.47	71.23	75.73
SR-MMP	88.37	87.19	85.53	86.17
SR-p53	84.11	84.83	77.19	87.37
Mean auROC (%)	**82.94**	**84.47**	**82.12**	**83.35**

pBRICS method, Matched Molecular Pairs (MMP) dataset was curated for each property. However, except for BBB permeability and Ames mutagenicity, there were not many MMP pairs for rest of the properties.

3.3 Matched Molecular Pairs (MMP) analysis for BBB Permeability (BBBP)

The challenge for the predictive model is to accurately predict small molecules that belong to the MMP dataset and are not part of the same class. The MMP dataset for BBBP property consists of 102 unique pairs of molecules along with their corresponding ground-truth labels. To check the model performance on the MMP dataset, the molecules from all pairs of MMP entries were combined to form a unique set of 110 molecules. The MMPs identified were supplied to the two trained models namely, MT-FraGCN and ADMETLab2.0, and one external model (SwissADME (Daina et al. (2017))) to get the predictions. The corresponding results are summarized in SI Table S1. Gradient Class Activation Map (Grad-CAM) algorithm was applied on all the molecules in the MMP dataset to get the gradient weights corresponding to each node (fragment) in the graph. Correct predictions are defined as cases where the positively labelled molecules obtain a positive gradient or weight for the corresponding MMP transformation identified and vice versa. Fig. 5 depicts two examples of correctly predicted MMP pairs from the dataset, whose Grad-CAM values were found to match with their labels. In MMP A (Fig. 5a), adding a hydroxyl group to the molecule transformed the molecule from BBB permeable to BBB impermeable, and in another case (Fig. 5b), replacing the halogen (Fluorine) with hydrogen atom transformed the molecules from BBB permeable to BBB impermeable. It is notable that these transformations are also experimentally proven to disrupt BBBP property of a molecule (Gentry et al. (1999)). To understand the false positives and false negatives from the Grad-CAM analysis, incorrectly predicted MMPs were analyzed to justify the rationale behind the model prediction. This analysis showed that the prediction was mostly in-

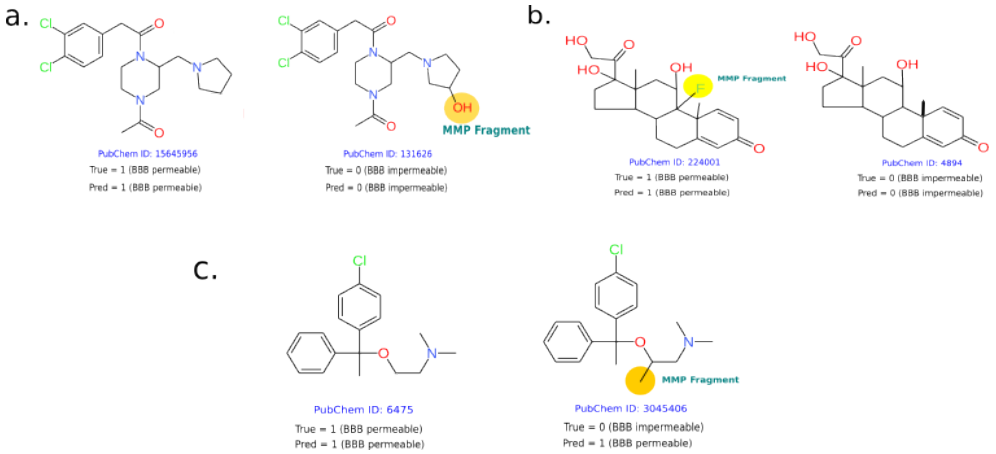

Figure 5: Examples of correctly predicted MMPs (a & b) and incorrectly predicted MMPs (c) using the MT-FraGCN model for BBB property. The MMP transformation is highlighted in yellow.

Table 3: Top MMP transformations identified for the BBB dataset.

ID	MMP Transformation	Active to Inactive (I)	Inactive to Active (A)	Count (C)	Mean Change (ratio = (A – I) / C)
1	*—CH_3 ⟶ *—Cl	0	8	8	Inactive to Active (1.00)
2	*—OH ⟶ *—H	0	7	7	Inactive to Active (1.00)
3	*—F ⟶ *—H	5	1	6	Active to Inactive (-0.66)

fluenced by the scaffold of the molecule even though the corresponding MMP fragment got a correct weight for the label.

In Fig. 5c, although the second molecule is BBB impermeable, the model predicted it as a BBB permeable molecule. Grad-CAM analysis showed that scaffold of the corresponding molecule got a positive Grad-CAM score, while the MMP transformation got a negative score. Since the frequency of the corresponding scaffold was found to be higher in the set of BBB permeable molecules compared to the fragment involved in the MMP transformation, the model could have interpreted it as a BBB permeable molecule.

For each MMP transformation present in the BBBP dataset, the number of MMP pairs with matching model predictions were identified and filtered to extract entries with at least 5 matches (Table 3), which are considered to be significant. Here, the transformation of C[*] to [*]Cl, was observed to change the label from impermeable to permeable. If the mean label of a transformation was positive, then the corresponding transformation was considered to have a significant effect in transforming the impermeable molecules to permeable molecules and vice versa.

3.4 Matched Molecular Pairs (MMP) analysis for Ames Mutagenicity

A similar analysis of MMP dataset was performed for Ames mutagenicity property. The Ames mutagenecity consists of 600 unique pairs of molecules along with their corresponding ground-truth labels. To check the model performance on the MMP dataset, the molecules from all pairs of MMP entries were combined to form a unique set of 712 molecules. The MMPs identified were supplied to both trained models and one external model similar to the BBBP property model, to obtain the predictions and corresponding results (SI table S2). It was observed that FraGCN could correctly predict maximum MMP pairs (SI table S2). Based on Grad-CAM analysis, replacing an amino group with methyl group in the molecule transformed the molecule from mutagen to non-mutagen (Fig. 6a), and replacing an amino with hydrogen atom transformed the molecules from mutagen to non-mutagen (Fig. 6b). Similar to the BBB model, we have also analysed the Grad-CAM scores for few of the incorrect predictions. The analysis again confirmed that the graph model's predictions are influenced by the scaffold of the corresponding molecule. In Fig. 6c, the molecule on the right

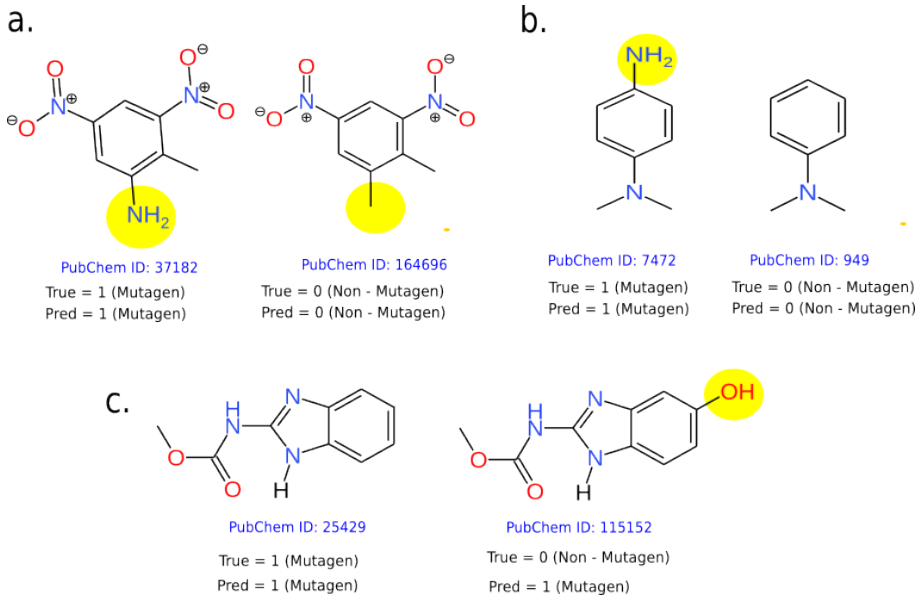

Figure 6: Examples of correctly predicted MMPs (a & b) and incorrectly predicted MMPs (c) using the MT-FraGCN model. The MMP transformation is highlighted in yellow.

Table 4: Top MMP transformations for Ames Mutagenicity.

ID	MMP Transformation	Active to Inactive (I)	Inactive to Active (A)	Count (C)	Mean Change (ratio = (A – I) / C)
1	*—NH_2 ⟶ *—H	72	8	80	Active to Inactive (- 0.80)
2	*—Cl ⟶ *—CH_3	11	4	15	Active to Inactive (-0.46)
3	*—Cl ⟶ *—OH	14	1	15	Active to Inactive (-0.88)
4	*—NH_2 ⟶ *—CH_3	9	4	13	Active to Inactive (-0.38)
5	*—CH_3 ⟶ *—OH	12	0	12	Active to Inactive (-1.00)
6	*—NH_2 ⟶ *—OH	10	1	11	Active to Inactive (-0.82)
7	*—CH_3 ⟶ *—NH_2	2	8	10	Inactive to Active (0.60)

is a non-mutagen, but was predicted to be an Ames Mutagen. However, a large part of the molecule (i.e., scaffold) was found to have a positive Grad-CAM score, since it was found to be most frequent in the mutagen set of molecules.

For each MMP transformation in the Ames Mutagenicity dataset, the number of MMPs with matching model predictions were identified and filtered to extract the entries with at least 10 matching labels (Table 4). Here the transformation of Chlorine to hydroxyl, methyl or hydrogen, was frequently found to change the label from mutagen to non-mutagen.

4 Conclusion

In summary, we have proposed a new fragmentation method named pBRICS for representing small molecules such that it captures a chemist's intution. The performance of the model trained using pBRICS fragments is higher when compared to atom-based representation. Also, the multi-task property prediction model showed marginal improvement over the corresponding single task models. Further, the explanations obtained from the trained model are easily interpretable by a medicinal chemist. While there was some success with the prediction of MMP for BBB permeability and Ames mutagenicity, the predictive performance is limited due to the amount of data that is available to train AI-based models. The proposed pBRICS method could guide the optimization of lead molecules, where the explanation obtained from deep learning models can be used to improve molecular properties.

ACKNOWLEDGMENTS

The authors thank their colleagues G Bulusu, D Das and B Chakrabarty for valuable suggestions.

REFERENCES

Delora Baptista, João Correia, Bruno Pereira, and Miguel Rocha. Evaluating molecular representations in machine learning models for drug response prediction and interpretability. *Journal of Integrative Bioinformatics*, 19(3), 2022.

Guy W Bemis and Mark A Murcko. The properties of known drugs. 1. molecular frameworks. *Journal of medicinal chemistry*, 39(15):2887–2893, 1996.

Vincent Blay, Xiaoyu Li, Jacob Gerlach, Fabio Urbina, and Sean Ekins. Combining dels and machine learning for toxicology prediction. *Drug Discovery Today*, pp. 103351, 2022.

Jannis Born, Matteo Manica, Ali Oskooei, Joris Cadow, and María Rodríguez Martínez. Paccmannrl: Designing anticancer drugs from transcriptomic data via reinforcement learning. In *Research in Computational Molecular Biology: 24th Annual International Conference, RECOMB 2020, Padua, Italy, May 10–13, 2020, Proceedings*, pp. 231–233. Springer, 2020.

Davide Buffelli and Fabio Vandin. A meta-learning approach for graph representation learning in multi-task settings. *arXiv preprint arXiv:2012.06755*, 2020.

Navneet Bung, Sowmya R Krishnan, Gopalakrishnan Bulusu, and Arijit Roy. De novo design of new chemical entities for sars-cov-2 using artificial intelligence. *Future medicinal chemistry*, 13 (06):575–585, 2021.

Navneet Bung, Sowmya Ramaswamy Krishnan, and Arijit Roy. An in silico explainable multiparameter optimization approach for de novo drug design against proteins from the central nervous system. *Journal of Chemical Information and Modeling*, 62(11):2685–2695, 2022.

Jonathan Cramer, Christoph P Sager, and Beat Ernst. Hydroxyl groups in synthetic and natural-product-derived therapeutics: a perspective on a common functional group. *Journal of medicinal chemistry*, 62(20):8915–8930, 2019.

John G Cumming, Andrew M Davis, Sorel Muresan, Markus Haeberlein, and Hongming Chen. Chemical predictive modelling to improve compound quality. *Nature reviews Drug discovery*, 12 (12):948–962, 2013.

Antoine Daina, Olivier Michielin, and Vincent Zoete. Swissadme: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, 7(1):42717, 2017.

Nour El-Huda Daoud, Pobitra Borah, Pran K Deb, Katharigatta N Venugopala, Wafa Hourani, Muhammed Alzweiri, Sanaa K Bardaweel, and Vinod Tiwari. Admet profiling in drug discovery and development: perspectives of in silico, in vitro and integrated approaches. *Current Drug Metabolism*, 22(7):503–522, 2021.

Jörg Degen, Christof Wegscheid-Gerlach, Andrea Zaliani, and Matthias Rarey. On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem: Chemistry Enabling Drug Discovery*, 3(10):1503–1507, 2008.

Joseph L Durant, Burton A Leland, Douglas R Henry, and James G Nourse. Reoptimization of mdl keys for use in drug discovery. *Journal of chemical information and computer sciences*, 42(6): 1273–1280, 2002.

Nicholas C Firth, Butrus Atrash, Nathan Brown, and Julian Blagg. Moarf, an integrated workflow for multiobjective optimization: implementation, synthesis, and biological evaluation. *Journal of chemical information and modeling*, 55(6):1169–1180, 2015.

CL Gentry, RD Egleton, T Gillespie, TJ Abbruscato, HB Bechowski, VJ Hruby, and TP Davis. The effect of halogenation on blood–brain barrier permeability of a novel peptide drug. *Peptides*, 20 (10):1229–1238, 1999.

Hossein Hajiabolhassan, Zahra Taheri, Ali Hojatnia, and Yavar Taheri Yeganeh. Funqg: Molecular representation learning via quotient graphs. *arXiv preprint arXiv:2207.08597*, 2022.

Jiazhen He, Huifang You, Emil Sandström, Eva Nittinger, Esben Jannik Bjerrum, Christian Tyrchan, Werngard Czechtizky, and Ola Engkvist. Molecular optimization by capturing chemist's intuition using deep neural networks. *Journal of cheminformatics*, 13(1):1–17, 2021.

José Jiménez-Luna, Miha Skalic, Nils Weskamp, and Gisbert Schneider. Coloring molecules with explainable artificial intelligence for preclinical relevance assessment. *Journal of Chemical Information and Modeling*, 61(3):1083–1094, 2021.

Peter W Kenny and Jens Sadowski. Structure modification in chemical databases. *Chemoinformatics in drug discovery*, pp. 271–285, 2005.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Ismail Kola and John Landis. Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery*, 3(8):711–716, 2004.

Sowmya Ramaswamy Krishnan, Navneet Bung, Gopalakrishnan Bulusu, and Arijit Roy. Accelerating de novo drug design against novel proteins using deep learning. *Journal of Chemical Information and Modeling*, 61(2):621–630, 2021a.

Sowmya Ramaswamy Krishnan, Navneet Bung, Sarveswara Rao Vangala, Rajgopal Srinivasan, Gopalakrishnan Bulusu, and Arijit Roy. De novo structure-based drug design using deep learning. *Journal of Chemical Information and Modeling*, 62(21):5100–5109, 2021b.

Xiao Qing Lewell, Duncan B Judd, Stephen P Watson, and Michael M Hann. Recap retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *Journal of chemical information and computer sciences*, 38(3):511–522, 1998.

Hirotomo Moriwaki, Yu-Shi Tian, Norihito Kawashita, and Tatsuya Takagi. Mordred: a molecular descriptor calculator. *Journal of cheminformatics*, 10(1):1–14, 2018.

Marcus Olivecrona, Thomas Blaschke, Ola Engkvist, and Hongming Chen. Molecular de-novo design through deep reinforcement learning. *Journal of cheminformatics*, 9(1):1–14, 2017.

Phillip Pope, Soheil Kolouri, Mohammad Rostrami, Charles Martin, and Heiko Hoffmann. Discovering molecular functional groups using graph convolutional neural networks. *arXiv preprint arXiv:1812.00265*, 2018.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pp. 593–607. Springer, 2018.

Marwin HS Segler, Thierry Kogej, Christian Tyrchan, and Mark P Waller. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS central science*, 4(1): 120–131, 2018.

Harold E Selick, Alan P Beresford, and Michael H Tarbit. The emerging importance of predictive adme simulation in drug discovery. *Drug Discovery Today*, 7(2):109–116, 2002.

Vishal Siramshetty, Jordan Williams, Dac-Trung Nguyen, Jorge Neyra, Noel Southall, Ewy Mathe, Xin Xu, and Pranav Shah. Validating adme qsar models using marketed drugs. *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, 26(10):1326–1336, 2021.

Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M Donghia, Craig R MacNair, Shawn French, Lindsey A Carfrae, Zohar Bloom-Ackermann, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.

Shunsuke Tamura, Tomoyuki Miyao, and Jürgen Bajorath. Large-scale prediction of activity cliffs using machine and deep learning methods of increasing complexity. *Journal of Cheminformatics*, 15(1):1–11, 2023.

Hao Tian, Rajas Ketkar, and Peng Tao. Admetboost: a web server for accurate admet prediction. *Journal of Molecular Modeling*, 28(12):1–6, 2022.

Han Van De Waterbeemd and Eric Gifford. Admet in silico modelling: towards prediction paradise? *Nature reviews Drug discovery*, 2(3):192–204, 2003.

Derek van Tilborg, Alisa Alenicheva, and Francesca Grisoni. Exposing the limitations of molecular machine learning with activity cliffs. *Journal of Chemical Information and Modeling*, 62(23): 5938–5951, 2022.

Sarveswara Rao Vangala, Navneet Bung, Sowmya Ramaswamy Krishnan, and Arijit Roy. An interpretable machine learning model for selectivity of small-molecules against homologous protein family. *Future Medicinal Chemistry*, 14(20):1441–1453, 2022.

Minjie Yu Wang. Deep graph library: Towards efficient and scalable deep learning on graphs. In *ICLR workshop on representation learning on graphs and manifolds*, 2019.

Michael J Waring, John Arrowsmith, Andrew R Leach, Paul D Leeson, Sam Mandrell, Robert M Owen, Garry Pairaudeau, William D Pennie, Stephen D Pickett, Jibo Wang, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature reviews Drug discovery*, 14(7):475–486, 2015.

Oliver Wieder, Stefan Kohlbacher, Mélaine Kuenemann, Arthur Garon, Pierre Ducrot, Thomas Seidel, and Thierry Langer. A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies*, 37:1–12, 2020.

Guoli Xiong, Zhenxing Wu, Jiacai Yi, Li Fu, Zhijiang Yang, Changyu Hsieh, Mingzhu Yin, Xiangxiang Zeng, Chengkun Wu, Aiping Lu, et al. Admetlab 2.0: an integrated online platform for accurate and comprehensive predictions of admet properties. *Nucleic Acids Research*, 49(W1): W5–W14, 2021.

Chun Wei Yap. Padel-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, 32(7):1466–1474, 2011.

Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34:15870–15882, 2021a.

Ziqiao Zhang, Jihong Guan, and Shuigeng Zhou. Fragat: a fragment-oriented multi-scale graph attention model for molecular property prediction. *Bioinformatics*, 37(18):2981–2987, 2021b.

Alex Zhavoronkov, Yan A Ivanenkov, Alex Aliper, Mark S Veselov, Vladimir A Aladinskiy, Anastasiya V Aladinskaya, Victor A Terentiev, Daniil A Polykovskiy, Maksim D Kuznetsov, Arip Asadulaev, et al. Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Nature biotechnology*, 37(9):1038–1040, 2019.

SUPPLEMETARY INFORMATION

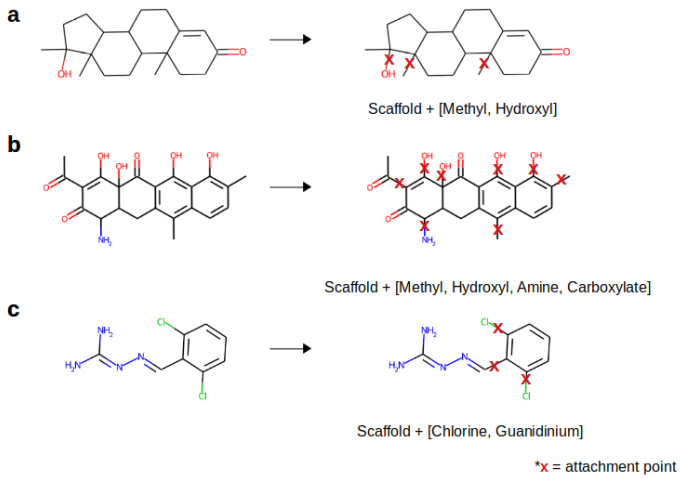

Figure S1: Examples of small molecules where only pBRICS can fragment the scaffold and substituents in comparison with BRICS, RECAP and SynDiR methods.

Table S1: MMP Analysis for BBB Permeability.

Task	Model	Performance
Combined SMILES list of all pairs of MMPs	MT-FraGCN	68 / 110
	ADMETLab2.0	80 / 110
	swissADME	60 / 110
Matching both labels of MMPs	MT-FraGCN	22 / 102
	ADMETLab2.0	31 / 102
	swissADME	33 / 102

Table S2: MMP Analysis for Ames Mutagenicity.

Task	Model	Performance
Combined SMILES list of all pairs of MMPs	MT-FraGCN	433 / 712
	ADMETLab2.0	436 / 712
	swissADME	405 / 712
Matching both labels of MMPs	MT-FraGCN	135 / 600
	ADMETLab2.0	98 / 600
	swissADME	33 / 102