Graph Neural Networks for Identifying Protein Reactive Compounds

Victor Hugo Cano Gil
Carleton University, Canada
victorcanogil@cmail.carleton.ca

Christopher Rowley
Carleton University, Canada
christopherrowley@cunet.carleton.ca

Abstract

In chemistry, electrophilic and nucleophilic reactions are utilized in the design of new protein reactive drugs, identification of toxic compounds, and the exclusion of reactive compounds from high throughput screening. In particular, covalent drugs comprise a class of protein reactive compounds that have seen a lot of interest due to their potential advantages such as better selectivity, longer effective dose, and overcoming drug resistances. Despite that, there are currently no reliable screening tools that go beyond basic substructure matching. In this work, we demonstrate that graph neural networks models are capable of predicting covalent reactivity and capturing chemical motifs by looking at gradient activation heatmaps and how they correlate with chemical theory. We also propose a new dataset, ProteinReactiveDB, which was used to train graph-based models in this work.

1 Introduction

Most FDA-approved drugs react via reversible intermolecular interactions [1]. Covalent drugs form covalent linkages that provide more durable connection to the target protein, so they are commonly used to inhibit or label proteins. Covalent inhibition typically occurs between the amino acid sidechain and reactive substructure of the molecule, typically referred to as the *covalent warhead*. The variety of covalent warheads is large and growing - additional protein reactive substructures are still being identified [2, 3]. The warhead diversity means that it is not practical to define search patterns for each of them individually. Moreover, neighboring atoms in a molecule can amplify or reduce both reactivity and/or efficiency of any of these groups [4].

While covalent inhibitors have significant therapeutic uses, there are other instances where protein reactivity must be avoided. Protein reactive compounds can have off-target activity due to promiscuous reactions with other cellular components and can be metabolized at faster rates due to higher electrophilicity. High-throughput screening methods evaluate the noncovalent binding of the compounds in large chemical datasets to a protein target. Likewise, generative AI method are now being used to design new compounds optimized to bind to a target [5, 6]. In each case, protein reactive compounds should be generally excluded from the searchers. Conversely, covalent modification is a common mode of toxicity, so predicting if a compound is protein reactive is also important in toxicology.

Currently, the most common way to screen for protein reactive compounds is based on the Pan-Assay INterference compoundS (PAINS) criteria which include motifs to identify compounds that contain electrophilic groups that are unstable or react promiscuously with proteins. Methods have been developed to automatically check if compound matches the criteria set for PAINS compounds, such as the PAINSfilter set of chemical search strings [7]. Although these strings are well-defined, these filters are not entirely effective as they rely entirely on string matching.

Deep learning and Graph Neural Networks (GNNs) in particular have been widely applied to chemical problems ranging from property prediction and structure generation to lab experiment

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

automation[8, 9, 10]. In this work, we train a variety of graph-based models to predict covalent activity. We also explore the interpretability of their predictions by looking at heatmaps generated using gradient activation mapping (GradCAM), and how they correlate to known chemical theory.

2 Methods

2.1 Datasets

ProteinReactiveDB was built on top of three datasets: CovalentInDB [11], DrugBank [12], and BindingDB [13]; it acted as training data for all models. CovalentInDB is a dataset of covalent inhibitors, and makes up the majority of ProteinReactiveDB's positive class. DrugBank is a dataset composed of predominantly drug molecules; BindingDB is a broader dataset of organic compounds.

A significant effort was made to refine the data, including removing irrelevant or misannotated molecules (such as metallorganics) as well as finding false positive and false negatives and correcting their label. In total, ProteinReactiveDB is composed of 51503 noncovalent inhibitors and 5875 covalent inhibitors.

External Test Data was built to test the transferability of the models on the types of compounds that might be evaluated in a modern medicinal chemistry campaign. It is comprised of three major types of molecules:

Table 1: Breakdown of the external test by class and by type of compounds

Class	Type	Count
Noncovalent	First Disclosures	105
	Nonreactive Decoys	49
Covalent	Aldehyde	5
	Alkyne	13
	Aziridine	6
	Chloropyridine	6
	Epoxides	18
	Furan	3
	Haloacetamides	11
	Isothiocyanates	1
	Lactone	27
	Nitrile	6
	Atypical	28
	Quinone	3
	Sulfonyl	47
	Thioketones	6
	Boronic	4
	Alkenes	175

- **Covalent Inhibitors (positive class)** - collected from recent literature; composed of molecules reported to be covalent inhibitors that are not present in the ProteinReactiveDB.
- **Nonreactive Decoys (negative class)** - constructed from drug-like molecules that contain substructures typical to covalent inhibitors, but were experimentally determined not to be such.
- **First Disclosures (negative class)** - collected from experimental drugs first disclosed 2021-2023, sourced from literature. These molecules contain features of modern drug candidates; none of them were reported to act through a covalent mechanism. As such, they were assigned to the negative class.

Table 1 breaks down the external test data by both class and type.

2.2 Models

Graph Convolutional Neural Networks proved to be the most successful type of models in this work. Chemical structures can be encoded into molecular graphs, where atoms are represented as nodes and chemical bonds are represented as edges. For each molecular graph, we let A be the adjacency matrix, X be the node attributes of the a graph N with n nodes. Let also the degree of matrix be $D_{ii} = \sum_j A_{ij}$. From the definition posited by Kipf and Weiling [14], a vanilla graph convolutional layer can then be defined as:

$$F^l(X, A) = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}F^{l-1}(X, A)W^l)$$

where F^l is the convolutional activation at the layer l, $F^0 = X$, $\tilde{A} = A + I_N$ is the adjacency matrix with added self-connections where I_N is the identify matrix, W^l are the trainable convolutional weights, $\tilde{D_{ii}} = \sum_j \tilde{A}_{ij}$, and σ is the nonlinear activation function.

Gradient Activation Mapping is an interpretability technique that first originated to explain how convolutional neural networks make predictions when classifying images by generating a heatmap [15]. Pope et al. [16] have shown that it can be also adopted to graph convolutions to generate graph heatmaps for class c at layer l and for feature k:

$$L^c_{GradCAM}[l,n] = ReLU(\sum_k \alpha^{l,c}_k F^l_{k,n}(X,A))$$

where $\alpha^{l,c}_k = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial y^c}{\partial F^l_{k,n}}$ are the class specific weights for class c at layer l and for feature k.

In this work, graph heatmaps were produced using $L^c_{GradCAM}Avg[l,n]$, defined by:

$$L^c_{GradCAM}Avg[n] = \frac{1}{L}\sum_{l=1}^{L} L^c_{GradCAM}[l,n]$$

3 Results and Discussion

The performance of different graph architectures are summarized in Table 2. Best performing model for each architecture was found using random hyperparameter search in conjunction with 10-fold validation. Most GNNs performed reasonably well, as summarized by their Precision, Recall, and Area Under Receiver Operating Characteristic curve (AUROC). However, all the GNN models struggled with the decoy set, as indicated by their false positive rates. We have found the most optimal model to have the following hyperparameters: two graph convolutional via initial residue and identity mapping (GCNII) layers (SeLU activation function), followed by a mean readout layer, a dense layer (ReLU activation function), and the output layer with sigmoid activation function. The optimizer was set to Adam (learning rate = 5×10^{-5}), the batch size was set to 64, each GCNII layer had a dropout layer after it (rate = 0.15). The model was trained against binary cross entropy as the loss function.

Table 2: Performance of various graph architectures, as measured by the internal and external AUROC, and external precision and recall. Also displayed is the FPR on the nonreactive decoy part of the external test set. The GCNII model discussed in the rest of the paper is highlighted. The full details of each model are described in the Supporting Information.

Graph Architecture	Internal Test AUROC	External Test AUROC	External Test Precision	External Test Recall	Nonreactive Decoy FPR	Ref.
GCN	0.96	0.84	0.90	0.71	0.49	14
GCNII	**0.95**	**0.84**	**0.92**	**0.73**	**0.37**	17
GraphSage	0.95	0.83	0.88	0.80	0.61	18
GAT	0.94	0.84	0.88	0.82	0.57	19
GatedGCN	0.96	0.83	0.90	0.70	0.39	20
GIN	0.97	0.81	0.90	0.72	0.49	21
GT	0.90	0.85	0.92	0.70	0.28	22
GMM	0.97	0.83	0.92	0.68	0.39	23
GATv2	0.93	0.82	0.88	0.76	0.59	24

We then looked at the heatmaps produced by GCNII when applied to known covalent drugs. Figure 1 shows the class activation maps of several known covalent drugs and how they correlate to actual covalent warheads. We can see that the GCNII model is recognizing covalent warheads as important to the classification, which agrees with chemical theory and resembles how a medicinal chemist would assess chemical structures.

4 Limitations

The composition of our training set imposes significant limitations on our methods. All datasets are based on compounds where inhibition experiments have been performed. Several recent novel covalent warheads employ unprecedented chemical motifs that are inherently absent from the training sets. The GCNII model identified correctly less than half of "atypical" warhead portion of the test set (Figure 2), which is composed of novel warheads that are not present in the training set. Other warheads (such as epoxides and aziridines) have triangular elements, which have been shown to be not amenable to graph convolutional methods [25]. These methods show limited abilities to identify covalent inhibitors with unusual warheads or cases there maybe external factors that affect covalent reactivity (e.g., nonreactive decoys). In short, it has all the drawbacks of data-driven methods.

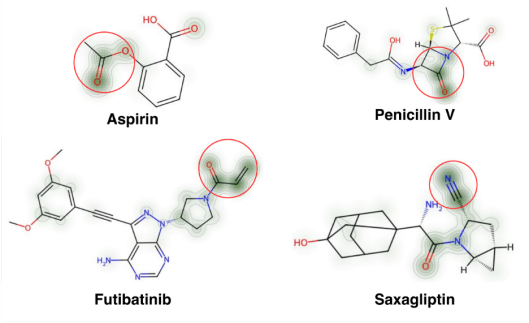

Figure 1: Selected examples of class activation maps of known covalent inhibitors. Circled in red are the actual covalent warheads and highlighted in green are the class activation maps.

It is also important to note that GradCAM explains the predictions of one model at a time, which itself depends on its architecture and the data it was trained on. From our own observations, GradCAM performed best (i.e. the class activation map corresponded to the actual covalent warhead) when classifying a covalent molecule correctly and with high confidence. Moreover, while GradCAM provides insight into interpretability of the model, it does not provide information of how model attributes importance.

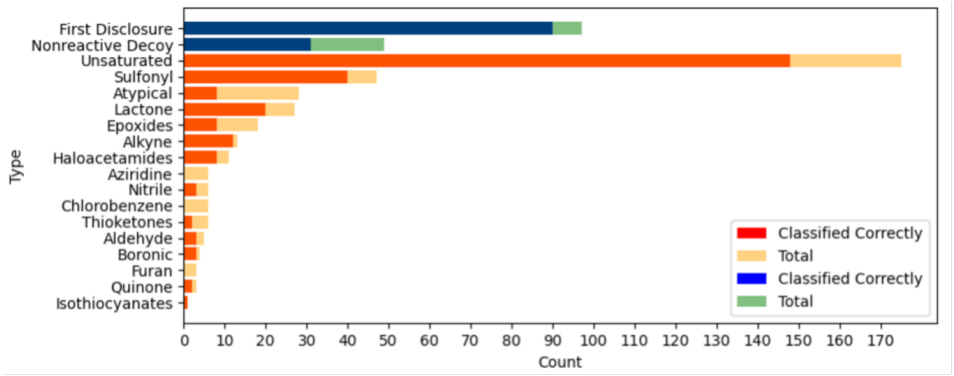

Figure 2: Performance of the GCNII model on the external test set by category.

5 Conclusion

Machine learning methods for predicting if a molecule is protein reactive were developed. A new dataset, ProteinReactiveDB, was established and as training data for protein reactivity classification models. To test the transferability of these models, an external test set was constructed from compounds that are not present in these sets, as well as a nonreactive decoy test set of compounds that contain functional groups that can be protein reactive but are not reactive in the chemical context of that molecule. Despite the limitations of the data and techniques employed, this study demonstrates the remarkable ability of GNNs to learn to recognize reactive chemical substructures based exclusively on the classification of compounds as covalent and noncovalent inhibitors. This suggests that the substantial libraries of covalent and noncovalent inhibitors are an effective training set for machine perception of electrophilicity. Currently, there are only a modest number of experimental chemical datasets that have the quality and extent that is suitable for machine learning, so the success of these models using these data opens new possibilities in chemical reaction prediction.

6 Data and Code Availability

The ProteinReactiveDB, our test set, and our complete code for both the fingerprint and graph models are all distributed on our GitHub repository: https://github.com/RowleyGroup/covalent-classifier.

Broader Impact

Drug discovery is considered to be one of the most important and complex applications of chemistry. The time it can take a potential drug candidate to reach the market can be over decade; even then, there are known cases of certain drugs being recalled due to unforeseen long-term side effects. The first phase of drug development - the research phase, heavily relies on careful screening and lab experimentation of potential drug candidates. The latter is especially difficult due to to the effectivily infinite chemical space. Current screening filters heavily rely on simple substructure matching, and are often not effective enough. This work is a step towards better screening tools that would allow for better screening of promising drug candidates, which would allow to better streamline the drug discovery process. Finally, protein reactivity is not exclusive to covalent modifier drugs, and the ability to detect if a compound is protein-reactive it will be beneficial to other areas, such as toxicology.

References

[1] Juswinder Singh, Russell C. Petter, Thomas A. Baillie, and Adrian Whitty. The resurgence of covalent drugs. *Nature Reviews Drug Discovery*, 10(4):307–317, 2011.

[2] Lydia Boike, Nathaniel J. Henning, and Daniel K. Nomura. Advances in covalent drug discovery. *Nature Reviews Drug Discovery*, 21(12):881–898.

[3] Matthias Gehringer and Stefan A. Laufer. Emerging and re-emerging warheads for targeted covalent inhibitors: Applications in medicinal chemistry and chemical biology. *Journal of Medicinal Chemistry*, 62(12):5673–5724.

[4] Iana M. Serafimova, Miles A. Pufall, Shyam Krishnan, Katarzyna Duda, Michael S. Cohen, Rebecca L. Maglathlin, Jesse M. McFarland, Rand M. Miller, Morten Frödin, and Jack Taunton. Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nature Chemical Biology*, 8(5):471–476.

[5] Liam Wilbraham, S. Hessam M. Mehr, and Leroy Cronin. Digitizing chemistry using the chemical processing unit: From synthesis to discovery. *Accounts of Chemical Research*, 54(2):253–262.

[6] Rizwan Qureshi, Muhammad Irfan, Taimoor Muzaffar Gondal, Sheheryar Khan, Jia Wu, Muhammad Usman Hadi, John Heymach, Xiuning Le, Hong Yan, and Tanvir Alam. AI in drug discovery and its clinical relevance. *Heliyon*, 9(7):e17575.

[7] Jonathan B. Baell and Georgina A. Holloway. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry*, 53(7):2719–2740.

[8] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry.

[9] Laurianne David, Amol Thakkar, Rocío Mercado, and Ola Engkvist. Molecular representations in AI-driven drug discovery: a review and practical guide. *Journal of Cheminformatics*, 12(1):56.

[10] Jianyuan Deng, Zhibo Yang, Iwao Ojima, Dimitris Samaras, and Fusheng Wang. Artificial intelligence in drug discovery: applications and techniques. *Briefings in Bioinformatics*, 23(1):bbab430.

[11] Hongyan Du, Junbo Gao, Gaoqi Weng, Junjie Ding, Xin Chai, Jinping Pang, Yu Kang, Dan Li, Dongsheng Cao, and Tingjun Hou. CovalentInDB: a comprehensive database facilitating the discovery of covalent inhibitors. *Nucleic Acids Research*, 49:D1122–D1129, 2021.

[12] David S. Wishart, Yannick D. Feunang, An C. Guo, Elvis J. Lo, Ana Marcu, Jason R. Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam Maciejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, Diana Le, Allison Pon, Craig Knox, and Michael Wilson. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46:D1074–D1082, 2018.

[13] Tiqing Liu, Yuhmei Lin, Xin Wen, Robert N. Jorissen, and Michael K. Gilson. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research*, 35:D198–201, 2007.

[14] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

[15] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2):336–359.

[16] Phillip E. Pope, Soheil Kolouri, Mohammad Rostami, Charles E. Martin, and Heiko Hoffmann. Explainability methods for graph convolutional neural networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10764–10773, 2019.

[17] Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks, 2020.

[18] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2018.

[19] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018.

[20] Xavier Bresson and Thomas Laurent. Residual gated graph ConvNets. (arXiv:1711.07553), 2018.

[21] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?, 2019.

[22] Luis Müller, Mikhail Galkin, Christopher Morris, and Ladislav Rampášek. Attending to graph transformers, 2023.

[23] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Jan Svoboda, and Michael M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model CNNs, 2016.

[24] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks?, 2022.

[25] Zhengdao Chen, Lei Chen, Soledad Villar, and Joan Bruna. Can graph neural networks count substructures?, 2020.