Graph Neural Networks for Identifying Protein-Reactive Compounds

Victor Hugo Cano Gil and Christopher N. Rowley*

Department of Chemistry, Carleton University, Ottawa, ON

E-mail: christopherrowley@cunet.carleton.ca

Phone: +(613) 520-2600 x 1647

Abstract

The identification of protein-reactive electrophilic compounds is critical to the design of new covalent modifier drugs, screening for toxic compounds, and the exclusion of reactive compounds from high throughput screening. In this work, we employ traditional and graph machine learning algorithms to classify molecules being reactive towards proteins or nonreactive. For training data, we built a new dataset, ProteinReactiveDB, comprised primarily of covalent and noncovalent inhibitors from DrugBank, BindingDB, and CovalentInDB databases. To assess the transferability of the trained models, we created a custom set of covalent and noncovalent inhibitors, which was constructed from recent literature. Baseline models were developed using Morgan fingerprints as training inputs, but they performed poorly when applied to compounds outside the training set. We then trained various Graph Neural Networks (GNNs), with the best GNN model achieving an Area Under the Receiver Operator Characteristic (AUROC) curve of 0.84, precision of 0.92, and recall of 0.73. We also explore the interpretability of these GNNs using Gradient Activation Mapping (GradCAM), which shows regions of the molecules GNNs deem most relevant when making a prediction.

1

These maps indicated that our trained models can identify electrophilic functional groups in a molecule and classify molecules as protein-reactive based on their presence.

Introduction

Proteins can undergo a range of chemical reactions with endogenous and exogenous molecules.[1–3] The amino acids cysteine, serine, lysine, threonine, and tyrosine can act as nucleophiles in reactions with electrophilic compounds. The covalent linkage formed through these reactions provides a more durable connection to the ligand than intermolecular interactions alone, so these reactions are often used to inhibit or label proteins.[4,5] These reactions typically occur between the amino acid side chain and a reactive moiety of the molecule, referred to as the covalent "warhead." Michael acceptors like acrylamides and α-haloacetamides commonly modify cysteine residues, while epoxides and lactones often target serine residues. In recent years, many warheads have been identified, including alkynes, cyclopropanes, chloropyridines, and benzaldehydes. The reactive warhead of a variety of covalent inhibitors is highlighted in Figure 1. The number of types of covalent warheads is large and growing; a database of covalent inhibitors (CovalentInDB) is organized into 63 different warhead categories.[6] Additional protein-reactive electrophilic functional groups are still being identified.[7]
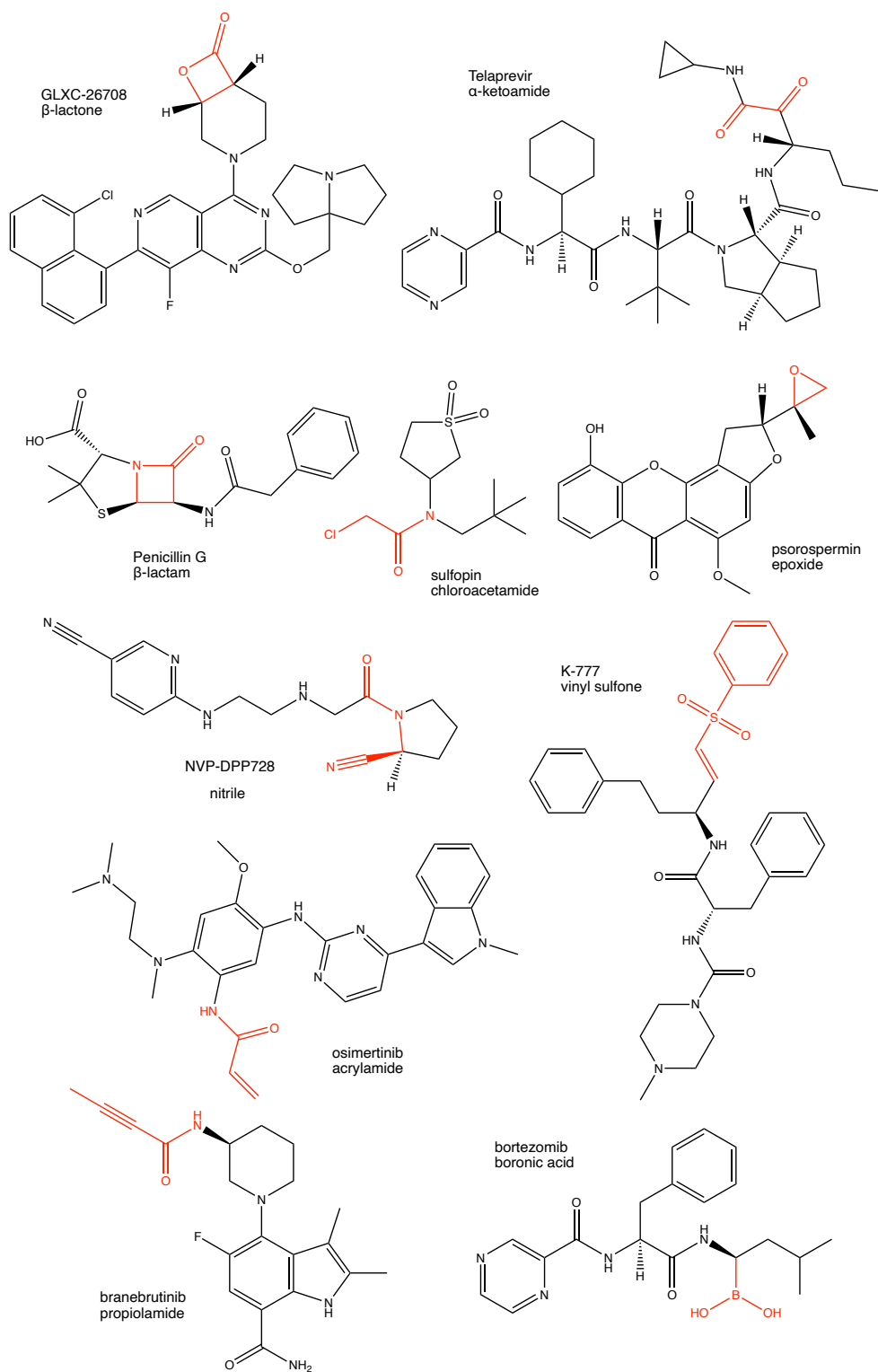
Figure 1: Examples of protein-reactive inhibitors. The substructure that reacts with the protein side chain (a.k.a., the warhead) is indicated in red.

3

While covalent inhibitors have significant therapeutic uses, there are other instances where it is important to detect protein reactivity because it is a liability in a specific application. Protein-reactive compounds can have off-target activity due to promiscuous reactions with other cellular components[8] and can be metabolized at faster rates due to higher electrophilicity.[9] Likewise, the development of non-covalent inhibitors now routinely uses high-throughput screening of the compounds in large chemical datasets to a protein target.[10,11] Alternatively, generative AI methods are now being used to design new compounds optimized to bind to a target.[12,13] Both of these cases, protein-reactive compounds should generally be excluded from the searches for non-covalent inhibitors. These applications would benefit from an efficient, automatic approach for identifying protein-reactive compounds.

There have been several efforts to predict the reactivity of compounds towards proteins using quantum chemistry.[14,15] Some model the reaction of a specific covalent inhibitor with its target,[16,17] while others attempt to predict the intrinsic reactivity of a warhead to model thiols.[18–21] These limitations require 3D structures of the warheads to be constructed and for quantum chemical calculations to be performed. Thio-Michael additions are a uniquely challenging chemical reaction for conventional DFT models,[22,23] and these methods have been limited to narrow classes of warheads and are not amenable to automated high-throughput screening.

One approach to identifying protein-reactive compounds would be to search for warhead substructures in a molecule. The **P**an-**A**ssay **IN**terference compound**S** (PAINS) criteria includes some electrophile motifs because compounds that promiscuously modify proteins can be false positives in high-throughput screening campaigns. Methods have been developed to automatically check if a compound matches the criteria set for PAINS compounds, such as the PAINSfilter[24] set of SMARTS search strings. Although these strings are efficient and well-defined, these filters are not entirely effective for detecting protein reactive compounds; only 7% of the CovalentInDB are identified as PAINS compounds, so many modes of protein reactivity are missed by these searches. The diversity of warheads means it is less

4

practical to define search patterns for each of them individually. Further, searching for a specific substructure ignores the reality that neighboring atoms in the molecule can amplify or attenuate the reactivity of any of these groups; for example, certain acrylamides that are normally non-reactive become potent covalent inhibitors of S6 kinase RSK2 if they are β-substituted with cyano groups.[25]

A machine learning classifier could provide a more general approach for predicting protein reactivity without requiring a researcher to define specific warhead substructures individually. These methods can leverage large quantities of data to define algorithms to classify molecules or predict their properties. These methods require a method to encode the molecular structure into a representation that is amenable to machine learning methods. Chemical fingerprints are a popular input to ML algorithms.[26–28] These fingerprints are vectors that contain ordered elements encoding for physical, chemical, and structural properties. A widely used class of chemical fingerprints is Extended Connectivity Fingerprints (ECFP),[29] which is based on the Morgan algorithm.[30] This produces binary sequences of fixed length, where a positive bit at a given position indicates the presence of a chemical substructure inside the molecule. Morgan fingerprints have been successfully used as a molecular representation in numerous chemical machine learning applications.[27,31–33] Graph representations of molecules are an alternative to these fingerprint methods.[34] In these models, atoms are represented as nodes of a graph and the bonds between them are represented as edges. Atomic and bond properties can be added as features to the graph nodes and edges, respectively. This allows extensive chemical data to be encoded in the graph. These graphs can be used as the inputs to Graph Neural Networks,[35] which can be trained for both classification and regression tasks.[34–36]

GNNs have been used to predict some modes of protein-molecule reactions. For example, Xenosite is a machine learning method that can predict if a compound can undergo a bioorganic transformation like epoxidation, glutathione conjugation, or alkylation.[37] A drawback of this model is that it was trained using data from the Accelrys Metabolite Database, which

cannot be distributed openly, so neither the model nor the training set are widely available. Generally, an open and extensible model for protein reactivity will require the use of publicly available datasets that can be extended as new compounds are synthesized and their modes of inhibition are reported. In this paper, we use machine learning techniques to develop a classifier to designate a molecule as being reactive towards proteins or non-reactive. We construct a training dataset using publicly accessible databases and an external test set from the recent literature.

Methods

The general workflow of the project is described in Figure 2. The methods developed in this paper take the chemical structure of a molecule as its input and output a classification of the molecule as being protein-reactive (positive class) or non-protein-reactive (negative class). These methods are trained using machine learning methods from datasets of molecules that are labeled as protein-reactive or non-reactive. A separate test set was curated to assess the transferability of these models. The construction of the training and test sets are described in the following sections. The training data and source code for all our models are deposited on GitHub.[38]

https://doi.org/10.26434/chemrxiv-2023-d0dqp **ORCID:** https://orcid.org/0000-0002-0205-952X Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0
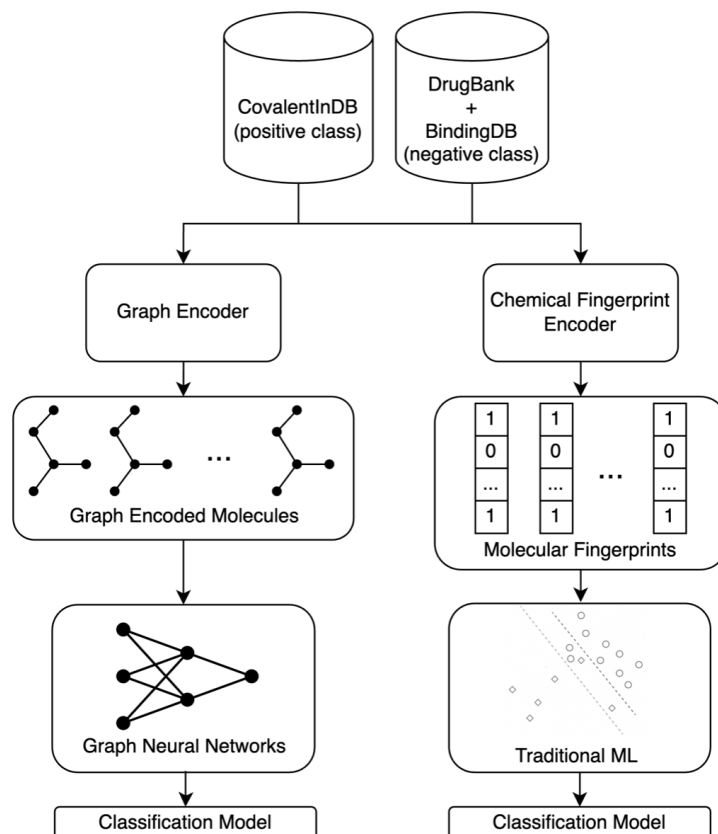
Figure 2: Schematics for the workflow of the ML protein reactivity classifiers. The models are trained exclusively from sets of inhibitors in public databases of covalent and noncovalent inhibitors. This labeled data is used to train models to classify molecules as being covalent or non-covalent using GNN and fingerprint-based ML models.

Data - Training

For training, we have built a new dataset, ProteinReactiveDB. This dataset was constructed from the data in three publicly available datasets: DrugBank,[39] BindingDB,[40] and CovalentInDB.[6] The DrugBank is predominantly composed of drug molecules. The BindingDB contains a broader set of molecules reported in the chemical literature. These two datasets served as the bulk of the negative (non-protein-reactive) set of molecules in the training set. The CovalentInDB is a database of inhibitors that have been determined to inhibit their targets by covalently modifying them. This dataset served as the bulk of the positive (protein-reactive) set of molecules in the training set.

7

These molecules from these datasets were curated and combined into a dataset appropriate for the available representations. All compounds containing inorganic components were excluded (i.e., containing only the elements H, B, C, N, S, O, F, Cl, Br). The RDKit (version 2023.03.2) toolkit[41] was used to convert the database entry into a molecular representation. If this library failed to generate a structure for a compound, it was not included in the data set.

An immediate challenge was that both the DrugBank and BindingDB contain some compounds that are covalent modifiers, so an extensive manual effort was made to identify these compounds and move them from the non-protein-reactive class training set to the protein-reactive class. This included 88 compounds in the DrugBank database that were annotated as DNA alkylating agents, insecticides, or broad-spectrum antibacterial compounds. Additional compounds that were believed to be misannotated or were not suitable for the representations used in these models (e.g., metal-containing compounds, antibodies, medical adhesives, etc.) were removed from the training set entirely (n=291). Compounds annotated as prodrugs were also excluded (n=64).

Compounds were removed from the protein-reactive set if there was experimental evidence in the literature that they act through a covalent mechanism. All compounds that were present in both the DrugBank and the CovalentInDB were categorized as protein-reactive. We performed an additional search of the compounds in our non-protein-reactive set that our first models classified as positive to determine if they act through a covalent mechanism. For these compounds, we searched the Protein-Databank for crystallographic structures of protein-ligand complexes and searched the macromolecular Crystallographic Information File (mmCIF) file for a covalent linkage between the compounds and the proteins. Lastly, a literature search was performed to identify any published studies where the enzyme kinetics were analyzed to determine if the mode of inhibitor was reversible or irreversible. 162 compounds in the DrugBank and 285 compounds from the BindingDB dataset were added to the covalent set through this process.

8

In total, the training set used in this study were composed of 45,244 noncovalent inhibitors and 6,259 covalent inhibitors. The dataset and lists of compounds included from the source databases are included in our GitHub repository.[38]

Data - Testing

The models presented in this paper are evaluated using two test sets. The first test set is generated by randomly extracting 10% of the compounds in ProteinReactiveDB using stratified sampling. We will refer to this set as the Internal Test Set. A second test set was constructed to test the transferability of these models to the types of compounds that might be evaluated in a modern medicinal chemistry campaign. This set will be referred to as the External Test Set, which is composed of covalent and non-covalent inhibitors that were not present in the training set (Table 1). These compounds were manually curated from the recent chemical literature, and are split into three groups: covalent inhibitors, first disclosures, and nonreactive decoys.

Covalent Inhibitors (positive class)

This test is composed of compounds reported to be covalent inhibitors, mostly collected from the recent literature highlighted on the weblog Covalent Modifiers.[42] This set is divided into subcategories of covalent warheads of inhibitors with a variety of covalent warheads, including aldehyde, alkyne, aziridine, boronic acid, chloropyridine, epoxide, furan, alkenes, lactones, nitrile, epoxides, furans, haloacetamides, isothiocyanate, lactones, nitriles, quinones, and sulfonyl. Compounds that do not fall into any of those groups are combined into a group called atypical covalent inhibitors.

First Disclosures (negative class)

The noncovalent component of the test set was collected from experimental drugs first disclosed 2021–2023, sourced from journal articles and DrugHunter.com.[43] These compounds

9

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

were selected because were not present in the versions of the DrugBank and BindingDB used in the training set but have the chemical features of modern drug candidates. None of these compounds were reported to act through a covalent mechanism in their disclosures and were manually inspected to ensure they did not contain a potential covalent warhead, so the classifier should assign these as being not protein-reactive.
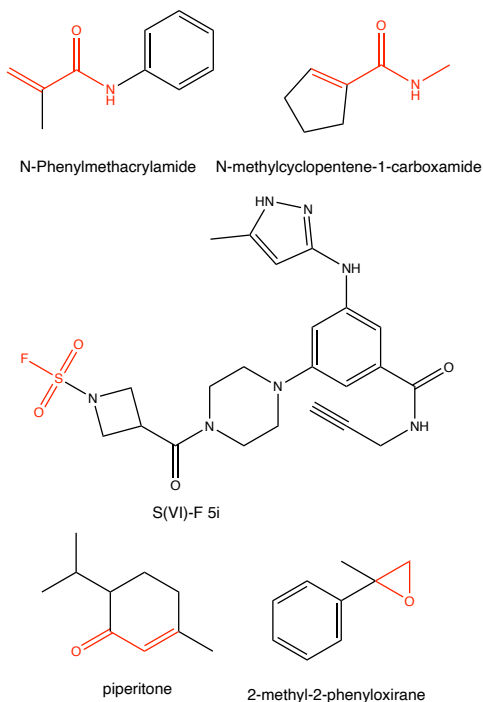
10

Nonreactive Decoys (negative class)

Figure 3: Examples of compounds from the non-reactive "decoy" set that contain a deactivated warhead (red). These compounds have been determined experimentally determined to react with nucleophiles at a slow rate.

One challenge for classifiers of covalent inhibitors is that functional groups that are protein-reactive in some molecules can be deactivated by their chemical environment to the degree that they will not be significantly reactive towards protein nucleophiles. For example, endocyclic cyclohexadienones like piperitone,[44] α-substituted acrylamides,,[45,46] deactivated sufonyl fluorides,[47] and substituted aliphatic epoxides[48,49] have been found to have limited reactivity with protein nucleophiles. These present an additional challenge for classification, because simple recognition of motifs like an epoxide or Michael acceptor would misclassify these compounds as reactive. To test if our ML classifiers can discern when an electrophile is deactivated, we constructed another test set of compounds that contain an electrophilic moiety (e.g., epoxide or α-β-unsaturated ketone), but have been determined experimentally to have slow or negligible rates of reaction with nucleophiles. This set of 49 compounds is

11

evaluated separately from the external test set because these compounds served as a distinct and more challenging test of the negative classification of these inhibitors.

Metrics

To measure the performance of our models, we employ common classification metrics such as precision, recall, and area under receiver operating characteristic (AUROC). For the external test set, we also review the accuracy of each prediction for each group. The classification metrics are defined as:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Where TP, FP, FN, TN are true positive, false positive, false negative, and true negative respectively. A model that classifies all molecules in the test set correctly will have both a precision and recall of 1. A lower precision indicates that the model tends to falsely classify molecules as being protein-reactive when they are not, while a lower recall indicates that the model tends to classify molecules as being non-protein-reactive when they are. The ROC curve is the plot of recall against false positive rate (FPR), defined as:

$$FPR = \frac{FP}{FP + TN} \tag{3}$$

The integral of this curve provides the AUROC. This metric indicates the success of the model for both positive and negative classification. A model with an AUROC of 1 indicates that it classified all molecules in the test set correctly.

12

Models and Features

Morgan Fingerprint Models

To establish a baseline of how effective conventional chemoinformatic methods are for this classification task, we have trained models using Morgan fingerprints. These fingerprints will have bits that indicate the absence or presence of a chemical fragment within a molecule. As a result, they should in principle be capable of representing the presence of an electrophilic group in a compound. We performed a grid hyperparameter search of logistic regression (LR), support vector classifier (SVC),[50] random forest (RF) classifier, histogram gradient boosting (HGB),[51] and multilayer perceptron (MLP)[52] models in the Scikit-Learn package (version 1.3).[53] The input features of the molecules in this model were the Morgan fingerprint generated using RDKit. Models were evaluated with various bit lengths and radii. A balanced loss function was used to train the LR, SVC, RF, and HGB classifiers. The full details of the hyperparameter search are included in the Supporting Information.

Graph Neural Networks

The second type of classifier we investigated was Graph Neural Networks (GNNs), where each molecule is represented as a graph where the nodes correspond to atoms and the edges correspond to bonds connecting the atoms. In particular, graph convolutional layers were employed. Following the definition of Kipf and Welling,[54] a vanilla graph convolutional layer can be defined as:

$$F^l(X,\ A)\ =\ \sigma\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\ \widetilde{D}^{-\frac{1}{2}}F^{(l-1)}(X,\ A)\,W^l\right) \tag{4}$$

Where A is the adjacency matrix, X is the node attributes of a graph with N nodes and adjacency matrix A. The degree of matrix A is $D_{ii} = \sum_j A_{ij}$, F^l is the convolutional activations at the layer l, $F_0 = X, \widetilde{A}\ =\ A\ +\ I_N$ is the adjacency matrix with added self-connections where I_N is the identity matrix, W^l are the trainable convolutional weights,

13

$\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$, and σ is the nonlinear activation function.

In this work, the GNNs were implemented using the Molgraph library[55] (version 0.5.8), which also provides a wrapper to RDKit descriptors that were used to generate atomic and bond features. The atomic and bond features included common chemical descriptors such as chemical symbol, total number of hydrogens, being a part of aromatic system, etc. A full list of the atomic and bond features is presented in the Supporting Information. Additionally, CDFT derived Fukui functions and electrophilicity indices were calculated and used as part of the atomic features in some models. Molecules were converted into 3D structures using RDKit and then the Fukui functions were calculated using AIMNET.[56]

Gradient Activation Mapping (GradCAM)

A drawback of neural networks is that it can be difficult to interpret how a classification decision is reached. This can make it difficult to determine if the model is making a classification based on relevant, generalizable properties of the input molecules or on a spurious correlation. As such, there has been an effort to understand their predictions better. In particular, the field of computer vision has seen several developments to better understand neural network predictions, with one of the more prominent techniques being gradient activation mapping. Pope *et al.*, have shown that GradCAM can be adapted to the graph neural networks;[57] first, we can calculate the class-specific weights for class c at layer l and for feature k using the following expression:

$$\alpha_k^{l,c} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial y^c}{\partial F_{k,n}^l} \tag{5}$$

Then, using Eqn. 4 , we can define $L_{Grad-CAM}^c$ as the heatmap from layer l:

$$L_{GradCAM}^c[l,n] = ReLU(\sum_k \alpha_k^{l,c} F_{k,n}^l(X,A)) \tag{6}$$

These values can be presented visually as a heatmap where the nodes are colored ac-

14

cording to the magnitude of L for a node. In this work, the heatmaps were produced using $L^c_{Grad-CAM}[l, \ n]$ Avg, defined by

$$L^c_{GradCAM}Avg[n] = \frac{1}{L}\sum_{l=1}^{L}L^c_{GradCAM}[l,n] \tag{7}$$

Results and Discussion

Morgan Fingerprint Models

The performance of the models trained using the Morgan fingerprints of the inhibitors as features is summarized in Table 2. The optimal models for all five classifiers performed reasonably well on the internal test set, with AUCROCs ranging from 0.82 to 0.96; however, the transferability of these models to the external test set was modest, with the AUCROCs between 0.55 and 0.69. These models have very high precisions on the external test set, ranging from 0.89 to 1, but have recalls that range from 0.11 to 0.51. This indicates that these models are likely skewed towards the noncovalent majority class, and are prone to classifying a candidate molecule as negative.

The HGB model had the best recall of any of these methods (0.51), although this still indicates a high false negative rate. Further, it falsely identified 42% of the decoy set as being protein-reactive, indicating this method has modest performance for negatively classifying molecules that contain motifs that are common in protein-reactive molecules but are deactivated due to the broader structure of the molecule. The RF model is the opposite extreme - it has a precision of 1.0 and a 0% false positive rate on the decoy set but had a particularly poor recall on the external test set (0.11).

All these models showed significantly poorer performance on the external test set than on the internal test set. In general, these models show that the approach of using Morgan fingerprints has limited transferability to compounds outside the training set. A chemical substructure indicated by a specific Morgan fingerprint bit can be connected to protein

15

reactivity, but these models fail to generalize in cases where the specific substructure is lost but protein-reactive activity is still present. This led us to explore more advanced methods using molecular graphs.

Table 2: Metrics for optimal ML models for predicting protein reactivity using Morgan fingerprint features for each classifier.

Model	Internal Test AUROC	External Test AUROC	External Test Precision	External Test Recall	Nonreactive Decoy FPR
SVC	0.93	0.64	0.89	0.42	0.21
HGB	0.95	0.68	0.89	0.51	0.42
LR	0.96	0.69	0.93	0.45	0.15
RF	0.82	0.55	1.0	0.11	0
MLP	0.96	0.64	0.92	0.34	0.15

Graph Neural Network Models

Several variants of GNN were evaluated in this work. For each model, optimal hyperparameters were found using a random search with 10-fold cross-validation. The full details of the hyperparameter search and best hyperparameters for each type of model are described in the Supporting Information. The performance of GNN models is summarized in Table 3. All these GNNs performed better than the fingerprint models on classifications of compounds in the external test set. Most significantly, these methods consistently had much higher recall rates, which ranged from 0.68 to 0.82. These results suggest that GNNs are significantly better at classifying protein-reactive compounds outside their training. These models all had significant false positive rates on the decoy set (false positive rates that ranged from 0.28 to 0.61).

All graph models perform similarly for the internal test AUROC, although the Graph Transformer (GT) model was the lowest. All models performed comparably on external test set AUCROC, although there were some differences in the external test set recall rates, with values ranging from 0.7 to 0.82. Models with higher recalls also had higher nonreactive decoy FPRs, indicating these models are biased towards positive classification. For our

16

immediate use, false positives are a greater concern than false negatives, so we have chosen the Graph Convolutional via Initial residue and Identity mapping[58] (GCNII) model based on its lower FPR, while its internal testset AUROC is superior to the GT. The GCNII model was developed to address issues with oversmoothing,[58] which is an advantage in these systems where a covalent substructure can span 3–4 bonds (edges) and deactivation of these substructures involves even more distant atoms.

Table 3: 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	54
GCNII	**0.95**	**0.84**	**0.92**	**0.73**	**0.37**	58
GraphSage	0.95	0.83	0.88	0.80	0.61	59
GAT	0.94	0.84	0.88	0.82	0.57	60
GatedGCN	0.96	0.83	0.90	0.70	0.39	61
GIN	0.97	0.81	0.90	0.72	0.49	62
GT	0.90	0.85	0.92	0.70	0.28	63
GMM	0.97	0.83	0.92	0.68	0.39	64
GATv2	0.93	0.82	0.88	0.76	0.59	65

Conceptual Density Functional Theory Features

Conceptual Density Functional Theory (CDFT) is often used to rationalize chemical reactivity.[66–68] The Fukui function is one of the most significant CDFT concepts. The electrophilic Fukui function (f^+) describes the rate at which the electron density at a point in space will change when an electron is added to the molecule. Electrophiles transfer electron density to the molecule, so the points where this function has a high magnitude have a high propensity for an electrophilic attack. The nucleophilic Fukui function (f^-) is defined as the rate that electron density at a point in space changes as an electron is removed from the molecule. Nucleophiles transfer electron density from the molecule, so the points where this function

has a high magnitude have a high propensity for nucleophilic attack. These functions can be condensed onto individual atoms to define atomic Fukui functions by calculating the partial atomic charges of the neutral, anionic, and cation states of a molecule and estimating the Fukui functions by finite difference. These condensed Fukui functions can be multiplied by the CDFT molecular electrophilicity to provide the positive (ω^+) and negative (ω^-) electrophilicity indices, which have been noted as useful descriptors for the prediction of warhead reactivity.[69,70]

CDFT features like the Fukui functions could be useful as atomic node features in GNNs for predicting chemical reactivity, but traditionally, calculating these terms would require a quantum chemical calculation. Isayev and coworkers have implemented CDFT predictions into AIMNET, a message-passing neural network approach that approximates ωB97x/def2-TZVPP minimal basis iterative stockholder charge analysis, without a quantum chemical calculation.[56] Using AIMNET, the nucleophilic and electrophilic Fukui functions can be calculated from these data with a very small computational cost, making it practical to include these charges as features in high-throughput GNN models.

To test whether GNNs with CDFT features perform better for predicting protein-reactivity, we trained a second GNN classifier with AIMNET-calculated atomic charges, positive Fukui function and negative condensed Fukui functions, and positive and negative condensed electrophilicity indices functions included as atomic features. Calculation of the AIMNET CDFT features requires the generation of a 3D structure, search for an optimal conformation, optimization of the structure, and calculation of CDFT properties using a message passing NN. The AIMNET NN currently only allows CDFT properties to be calculated for neutral molecules and other failures in this workflow reduced the training set to 5875 covalent inhibitors and 43373 non-covalent inhibitors. For comparison, a second GNN classifier was trained using this dataset but without the CDFT features. The metrics for both models are presented in Table 4. Both GNNs used the same architecture as the GCNII from Table 3.

18

Table 4: External test data performance of GCNII architecture with and without CDFT features,

Architecture	Internal Test AUROC	External Test AUROC	External Test Precision	External Test Recall	Nonreactive Decoy FPR
Without CDFT	0.95	0.85	0.92	0.71	0.49
With CDFT	0.96	0.84	0.94	0.74	0.78

The CDFT model performed similarly to the non-CDFT model across classification metrics; however, it performed significantly worse on the decoy set (false positive rate of 0.78 compared to 0.49). This is surprising because CDFT properties like the Fukui function are standard quantum chemical methods for quantifying the electrophilicity of an atom in a molecule. Hughes *et al.* also investigated the utility of CDFT features in their Xenosite GNN classifier for mechanisms of biomolecular reaction and metabolism and found that they did not result in a large improvement.[71] We suspect that the existing atomic features defined based on the bonding connectivity are sufficient for the GNN to make predictions of protein reactivity that are already near the limit of these graph architectures given the limited training data, so CDFT features do not provide data that can improve upon this.
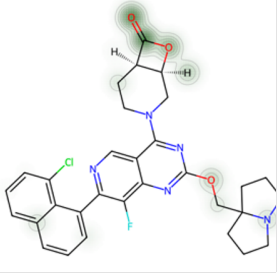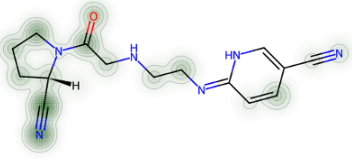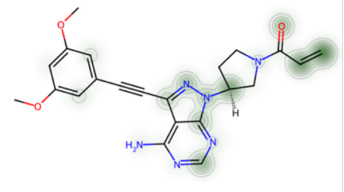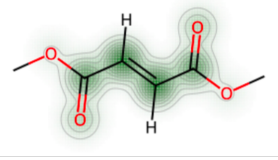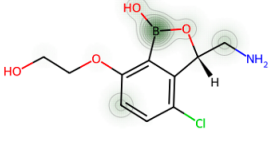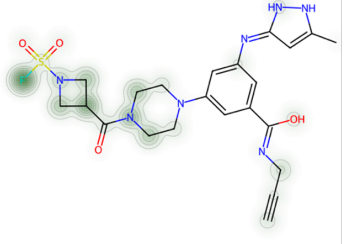
There are several drawbacks associated with including CDFT features vs our main GNN classifier. Calculation of the AIMNET CDFT features requires the generation of a 3D structure, generation of an optimal conformation, optimization of the structure, and calculation of CDFT properties. In contrast, all the features in our previous model can be calculated from the 2D structure alone. Calculating a 3D structure is computationally intensive and occasionally fails in an automated workflow, so adding these features significantly complicates the workflow.

Gradient Activation Maps

Like with any neural network architecture, GNNs are not directly interpretable. As we have constructed the positive and negative classes of our datasets from different sources, there is

some risk that the trained network would make classifications based on characteristics that are not generalizable. When adapted to graph inputs, GradCAM is capable of producing graph heatmaps (see Eqn. 6). To assess whether the models developed here classify based on generalizable criteria, we calculated the GradCAM heatmaps for a variety of molecules (Table 5), allowing us to visualize which atoms in a molecule are contributing most to its classification as a protein-reactive or non-protein-reactive molecule.

Table 5: Selected examples of known covalent drugs, their classifier confidence score of being of covalent as classified by GCNII model, and their gradient activation heatmaps

Name	Heatmap	Classifier Confidence,%	Ref.
G12Si-5		87.5	72
NVP-DPP-728		78	73
Futibatinib		99.5	74
Dimethylfumarate		82	75
Ganfeborole		99.7	76
S(VI)–F 5i		99.9	47

G12Si-5 features a β-lactam warhead, which forms a covalent bond with the mutant Serine-12 residue in KRAS G12S.[77] The GNN correctly classifies it as being a covalent in-

hibitor with a classifier confidence score of 87.5%. The heatmap highlights the β-lactam warhead, indicating that classification is correctly based on the presence of this electrophile in the molecule. Likewise, the covalent inhibitors NVP-DPP-728, dimethyl fumarate, futibatinib, and ganfeborole are all correctly classified as being protein-reactive. The heatmaps highlight their nitrile, acrylate, acrylamide, and cyclic borate warheads, respectively, indicating that their positive classification was correctly based on the presence of these motifs.

Futibatinib is a notable example because it contains two nominally electrophilic groups: an acrylamide and an alkyne. The heatmap indicates that the acrylamide group was the most significant class for the positive classification. This is in keeping with the mode of action of this inhibitor, which inhibits FGFR1–4 through the chemical modification a of P-loop cysteine and the acrylamide, while the alkyne is unmodified.[78] This demonstrates that the GNN model can recognize that the acrylamide is activated while the reactivity of the alkyne is muted by conjugation with two aromatic rings.

The sulfamoyl fluoride compound S(VI)–F 5i is another instance of a false-positive classification where the classifier categorizes the compound as protein-reactive with very high confidence (99.9%). Although the compound contains a sulfonyl fluoride warhead, Gilbert *et al.* found the sulfamoyl ring attenuates the reactivity to the point that its rate of modification of the CDK2 protein was negligible.[47] The failure of the model in this case likely reflects the large number of sulfonyl fluoride covalent modifiers and an insufficient number of inert sulfonyl fluorides in the training set. This example indicates that these methods have a significant risk of falsely classifying a compound as protein-reactive if it contains a deactivated electrophile. Additional training data or atomic features may address this issue.

Limitations

Although the metrics of the GNN classifier are good for a chemical application of this type, there are some areas where the performance is weaker. This is evident when the true and false classifications of the external test set are grouped by type for the GCNII model (Figure 4)

The compounds in the first disclosure set are non-covalent inhibitors that have been reported in the literature recently, so they are not present in the training sets. The GCNII classifier is generally effective in classifying them as non-covalent inhibitors but incorrectly classifies 7 of the 97 molecules as being protein-reactive. The decoy compounds are a set of molecules that contain electrophilic functional groups, but their reactivity has been determined to be very slow or insignificant by experimental measurements. The classifier has mixed performance here, indicating that it has some ability to exclude non-reactive compounds but frequently misclassifies them.

The model shows good performance on compounds with a warhead featuring an unsaturated bond, which can be explained by them being well represented in training data – CovalentInDB contains a large number of this type of Michael Acceptor warheads. It performed poorer on "atypical" warhead portion of the test set, which includes novel functional groups that have only recently been identified as protein reactive (e.g., isoxazoline-based electrophiles[79]); less than half were classified correctly. This is likely due to very limited training data and a lack of transferability. For several other groups, the model underperforms either because those groups are not well represented in the training data. There are only 10 chloropyridines and 1 aziridine in the positive training set, so the network is likely undertrained in recognizing when these structures will be protein-reactive. Further, both epoxides and aziridines have triangular elements, which are not amenable to graph convolutional methods.[80] Hughes *et al.* introduced a special expoxide atomic feature and additional training data to train their GCN Xenosite model to predict the reactivity of expoxides correctly, which may also be needed to improve the performance of this model on epoxides.[71]
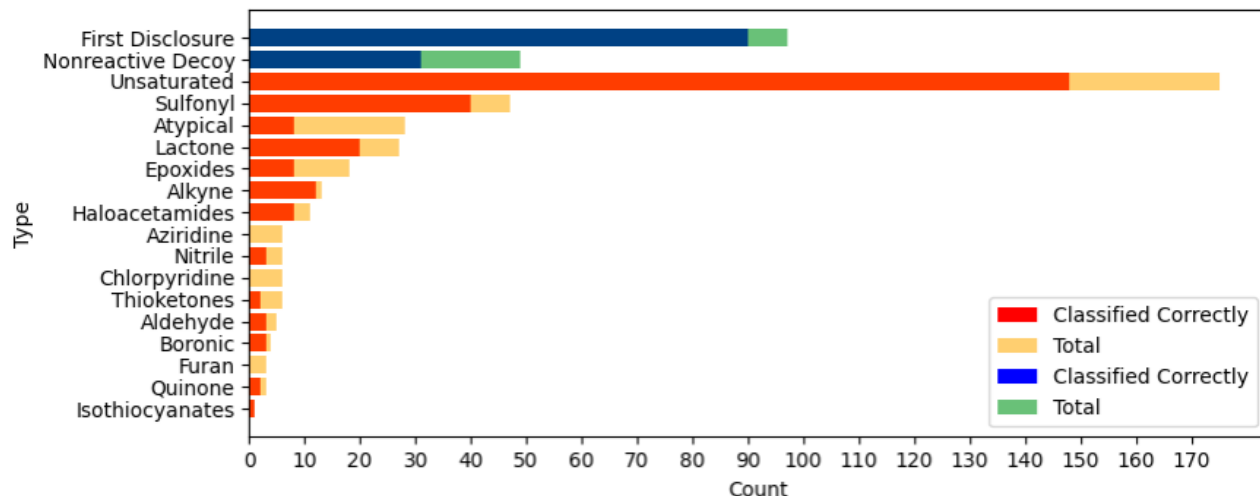
23

Figure 4: Performance of the GCNII model on the external test set by type of inhibitor. Covalent inhibitors are subdivided by the warhead type

The reaction between a protein side chain and a molecule is often mediated by the environment inside the protein, such as neighboring charged residues and hydrogen bonding networks inside the active site of an enzyme.[81] While some covalent inhibitors are promiscuous,[82,83] Kuljanin *et al.*, found that there was a high level of selectivity for a particular covalent inhibitor in whole-cell assays.[84] Our categorization of inhibitors as either covalent or noncovalent ignores these distinctions, so this classifier cannot predict if a compound will covalently modify a specific protein, but rather it predicts whether the molecule could covalently modify a protein provided there is a protein with a binding site that can accommodate the ligand in an orientation that will put its warhead in contact with a reactive side chain.

Generally, the composition of our training set imposes significant limitations on our methods. All three source datasets are based on compounds where inhibition experiments have been performed. Many highly electrophilic compounds would not be present in the training set because they are too unstable to perform inhibition studies of. Further, novel covalent warheads that employ unprecedented chemical motifs are still being identified. These structures are inherently absent from the training sets and these models have only limited abilities to predict reactivity in compounds dissimilar to those in their training set. More generally,
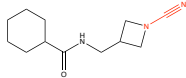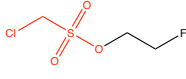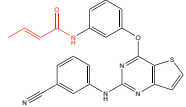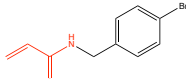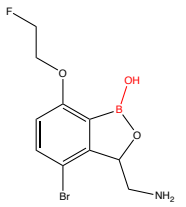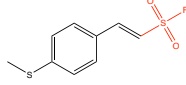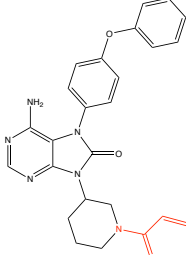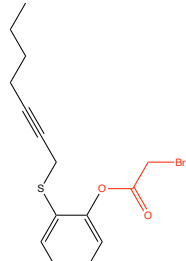
the labeling of the data is "noisy" because currently covalent inhibitors must be manually separated from the non-covalent training sets and is not always apparent when an inhibitor acts through a covalent mechanism[85–87] The expansion of the training set and developments to make these models more transferable to new chemical substrates may help address this issue.

Protein-Reactive Molecules in the ChEMBL Database

Libraries of chemical structures are often used in high-throughput screening campaigns. Generally, these screenings are intended to identify non-covalent inhibitors, so molecules likely to react with proteins would create risks of off-target inhibition or toxicity. Generally, protein- reactive molecules should be excluded from these searches. The ChEMBL database is a widely used library of drug-like compounds collected from medicinal chemistry journals and patents,[88–90] but it is not currently separated into covalent and non-covalent inhibitors.

This led us to apply the GNN classifier developed in this work to the ChEMBL 33 database to identify how many potentially protein-reactive molecules are in this set. The GNN classifier developed in the previous section was used with a threshold of 0.9 to minimize false positives. 5.1% of the ChEMBL database was flagged as potentially being protein-reactive by these criteria. Eight examples are presented in Table 6. These compounds were confirmed to be covalent-modifier inhibitors through a literature search. Researchers using the ChEMBL database to search for non-covalent inhibitors may consider testing if the compounds are protein-reactive using this classifier to exclude these molecules from the search because there is a risk that they will react with a protein other than the target.

25

Table 6: Examples of compounds in the ChEMBL database that were correctly identified as being protein-reactive. The warhead is highlighted in red. Compounds were selected using a classifier score threshold of 0.9. The classifier score and the literature reference providing evidence of a covalent mode of inhibition are provided.

ChEMBL ID	Chemical Structure	Classifier Score, %	Ref.
ChEMBL8796		99.7	91
ChEMBL17428		90.0	92
ChEMBL4751575		98	93
ChEMBL2086469		99.5	94
ChEMBL4116142		99.5	95
ChEMBL4435627		99.9	96
ChEMBL4303189		99.9	97
ChEMBL144547		96.7	98

Currently, protein-reactive compounds are often identified by using PAINS criteria. These

have been implemented by defining SMARTS patterns[99] that will match specific chemical substructures within molecules. Many of these PAINS patterns correspond to electrophilic groups that would be chemically unstable or react promiscuously with proteins. We find that the compounds identified by our classifier are generally distinct from those identified as PAINS compounds, with an overlap coefficient of only 0.065 for compounds in the ChEMBL dataset. This suggests that the PAINS patterns currently in use would not detect a large fraction of protein-reactive compounds.

Conclusions

Machine learning methods for predicting if a molecule is protein-reactive were developed. A new dataset, ProteinReactiveDB was constructed from public datasets of molecular inhibitors. These data were used to train classifiers to designate a molecule as being protein-reactive or not protein reactive. 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 non-reactive 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.

Conventional ML methods using Morgan fingerprints as features had limited transferability and performed poorly in identifying protein-reactive molecules in the external test set. The HGB classifier was the best-performing model of this type, with an AUCROC of 0.95 on the internal test set but degraded to 0.68 on the external test set. The primary limitation of these models is a high false negative rate; the HGB model had a precision 0.89 but a recall of only 0.51 for compounds in the external test set.

The GNNs showed improved performance over the models based on Morgan fingerprints, with GCNII model performing the best across most metrics. This model had an AUCROC of 0.95 for the internal test set and 0.84 for the external test set. Notably, the recall of these models was much improved. Analysis of the GNN using the gradient activation map indicates

that these models successfully identify the relevant reactive regions of these inhibitors and can distinguish electrophilic groups that are made less electrophilic by their environment.

These models were effective using only basic atomic and bond properties as features and adding more sophisticated CDFT properties did not provide a model that was systematically improved. Analysis of the GradCAM heatmaps showed that these models can successfully identify the electrophilic warhead of the compound, indicating the classification being made based on chemically sensible criteria.

The GNN models have a small but significant false-positive rate, so when these are applied to large databases, there will be a significant number of compounds incorrectly classified as protein-reactive. These models also struggle with the "decoy" test set of compounds that contain similar functional groups as covalent inhibitors but are not sufficiently reactive to be considered a practical covalent inhibitor. This is generally more challenging because it requires that the degree of electrophilicity is estimated rather the just the presence or absence of a reactive motif. Likewise, these models had limited success in identifying protein-reactive compounds with newly developed warheads that are not well-represented in the training set.

Despite these limitations, 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 non-covalent 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.

Data 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.

Acknowledgement

Supporting Information Available

Details of the hyperparameter optimization, list of GNN atomic and bond features, list of GNN features

References

(1) Coles, B. Effects of Modifying Structure on Electrophilic Reactions with Biological Nucleophiles. *Drug Metabolism Reviews* **1984**, *15*, 1307–1334.

(2) Rudolph, T. K.; Freeman, B. A. Transduction of Redox Signaling by Electrophile-Protein Reactions. *Science Signaling* **2009**, *2*, re7–re7.

(3) Enoch, S. J.; Ellison, C. M.; Schultz, T. W.; Cronin, M. T. D. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Critical Reviews in Toxicology* **2011**, *41*, 783–802.

(4) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The resurgence of covalent drugs. *Nature Reviews Drug Discovery* **2011**, *10*, 307–317.

(5) Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angewandte Chemie International Edition* **2016**, *55*, 13408–13421.

(6) Du, H.; Gao, J.; Weng, G.; Ding, J.; Chai, X.; Pang, J.; Kang, Y.; Li, D.; Cao, D.; Hou, T. CovalentInDB: a comprehensive database facilitating the discovery of covalent inhibitors. *Nucleic Acids Research* **2021**, *49*, D1122–D1129.

(7) Péczka, N.; Orgován, Z.; Ábrányi Balogh, P.; Keserű, G. M. Electrophilic warheads in covalent drug discovery: an overview. *Expert Opinion on Drug Discovery* **2022**, *17*, 413–422.

(8) Dahal, U. P.; Obach, R. S.; Gilbert, A. M. Benchmarking in Vitro Covalent Binding Burden As a Tool To Assess Potential Toxicity Caused by Nonspecific Covalent Binding of Covalent Drugs. *Chemical Research in Toxicology* **2013**, *26*, 1739–1745.

(9) Shibata, Y.; Chiba, M. The Role of Extrahepatic Metabolism in the Pharmacokinetics of the Targeted Covalent Inhibitors Afatinib, Ibrutinib, and Neratinib. *Drug Metabolism and Disposition* **2015**, *43*, 375–384.

30

(10) Clyde, A. et al. High-Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Noncovalent Inhibitor. *Journal of Chemical Information and Modeling* **2022**, *62*, 116–128.

(11) Garland, O.; Ton, A.-T.; Moradi, S.; Smith, J. R.; Kovacic, S.; Ng, K.; Pandey, M.; Ban, F.; Lee, J.; Vuckovic, M.; Worrall, L. J.; Young, R. N.; Pantophlet, R.; Strynadka, N. C. J.; Cherkasov, A. Large-Scale Virtual Screening for the Discovery of SARS-CoV-2 Papain-like Protease (PLpro) Non-covalent Inhibitors. *Journal of Chemical Information and Modeling* **2023**, *63*, 2158–2169.

(12) Arnold, C. Inside the nascent industry of AI-designed drugs. *Nature Medicine* **2023**, *29*, 1292–1295.

(13) Martinelli, D. D. Generative machine learning for de novo drug discovery: A systematic review. *Computers in Biology and Medicine* **2022**, *145*, 105403.

(14) Awoonor-Williams, E.; Walsh, A. G.; Rowley, C. N. Modeling covalent-modifier drugs. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2017**, *1865*, 1664–1675.

(15) Awoonor-Williams, E.; Rowley, C. N. How Reactive are Druggable Cysteines in Protein Kinases? *Journal of Chemical Information and Modeling* **2018**, *58*, 1935–1946.

(16) Awoonor-Williams, E.; Rowley, C. N. Modeling the Binding and Conformational Energetics of a Targeted Covalent Inhibitor to Bruton's Tyrosine Kinase. *Journal of Chemical Information and Modeling* **2021**, *61*, 5234–5242.

(17) Martí, S.; Arafet, K.; Lodola, A.; Mulholland, A. J.; Świderek, K.; Moliner, V. Impact of Warhead Modulations on the Covalent Inhibition of SARS-CoV-2 Mpro Explored by QM/MM Simulations. *ACS Catalysis* **2022**, *12*, 698–708.

(18) Schwöbel, J. A. H.; Wondrousch, D.; Koleva, Y. K.; Madden, J. C.; Cronin, M. T. D.; Schüürmann, G. Prediction of Michael-Type Acceptor Reactivity toward Glutathione. *Chemical Research in Toxicology* **2010**, *23*, 1576–1585.

(19) Awoonor-Williams, E.; Kennedy, J.; Rowley, C. N. In *The Design of Covalent-Based Inhibitors*; Ward, R. A., Grimster, N. P., Eds.; Annual Reports in Medicinal Chemistry; Academic Press, 2021; Vol. 56; pp 203–227.

(20) Lonsdale, R.; Burgess, J.; Colclough, N.; Davies, N. L.; Lenz, E. M.; Orton, A. L.; Ward, R. A. Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity. *Journal of Chemical Information and Modeling* **2017**, *57*, 3124–3137.

(21) Smith, J. M.; Rowley, C. N. Automated computational screening of the thiol reactivity of substituted alkenes. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 725–735.

(22) Smith, J. M.; Jami Alahmadi, Y.; Rowley, C. N. Range-separated DFT functionals are necessary to model Thio-Michael additions. *J. Chem. Theory Comput.* **2013**, *9*, 4860–4865.

(23) Awoonor-Williams, E.; Isley III, W. C.; Dale, S. G.; Johnson, E. R.; Yu, H.; Becke, A. D.; Roux, B.; Rowley, C. N. Quantum Chemical Methods for Modeling Covalent Modification of Biological Thiols. *Journal of Computational Chemistry* **2020**, *41*, 427–438.

(24) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *Journal of Medicinal Chemistry* **2010**, *53*, 2719–2740.

(25) Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.; McFarland, J. M.; Miller, R. M.; Frödin, M.; Taunton, J. Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat. Chem. Biol.* **2012**, *8*, 471–476.

(26) Baptista, D.; Correia, J.; Pereira, B.; Rocha, M. Evaluating molecular representations in machine learning models for drug response prediction and interpretability. *Journal of Integrative Bioinformatics* **2022**, *19*.

(27) Qiao, Z.; Li, L.; Li, S.; Liang, H.; Zhou, J.; Snurr, R. Q. Molecular fingerprint and machine learning to accelerate design of high-performance homochiral metal–organic frameworks. *AIChE Journal* **2021**, *67*, e17352.

(28) Yang, M.; Tao, B.; Chen, C.; Jia, W.; Sun, S.; Zhang, T.; Wang, X. Machine Learning Models Based on Molecular Fingerprints and an Extreme Gradient Boosting Method Lead to the Discovery of JAK2 Inhibitors. *Journal of Chemical Information and Modeling* **2019**, *59*, 5002–5012.

(29) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **2010**, *50*, 742–754.

(30) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation* **1965**, *5*, 107–113.

(31) Riniker, S.; Fechner, N.; Landrum, G. A. Heterogeneous Classifier Fusion for Ligand-Based Virtual Screening: Or, How Decision Making by Committee Can Be a Good Thing. *Journal of Chemical Information and Modeling* **2013**, *53*, 2829–2836.

(32) Banerjee, P.; Preissner, R. BitterSweetForest: A Random Forest Based Binary Classifier to Predict Bitterness and Sweetness of Chemical Compounds. *Frontiers in Chemistry* **2018**, *6*.

(33) Zhu, X.; Polyakov, V. R.; Bajjuri, K.; Hu, H.; Maderna, A.; Tovee, C. A.; Ward, S. C. Building Machine Learning Small Molecule Melting Points and Solubility Models Using CCDC Melting Points Dataset. *Journal of Chemical Information and Modeling* **2023**, *63*, 2948–2959.

(34) Wieder, O.; Kohlbacher, S.; Kuenemann, M.; Garon, A.; Ducrot, P.; Seidel, T.; Langer, T. A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies* **2020**, *37*, 1–12.

(35) Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P. S. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* **2021**, *32*, 4–24.

(36) Dwivedi, V. P.; Joshi, C. K.; Luu, A. T.; Laurent, T.; Bengio, Y.; Bresson, X. Benchmarking Graph Neural Networks. `http://arxiv.org/abs/2003.00982`.

(37) Hughes, T. B.; Miller, G. P.; Swamidass, S. J. Modeling Epoxidation of Drug-like Molecules with a Deep Machine Learning Network. *ACS Central Science* **2015**, *1*, 168–180.

(38) GitHub. 2023; `https://github.com/RowleyGroup/covalent-classifier`.

(39) Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* **2018**, *46*, D1074–D1082.

(40) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research* **2007**, *35*, D198–201.

(41) RDKit: Open-Source Cheminformatics Software. **2023**,

(42) Covalent Modifiers. 2023; `https://covalentmodifiers.blogspot.com/`.

(43) Drug Hunter. 2023; `https://drughunter.com/`.

(44) Avonto, C.; Taglialatela-Scafati, O.; Pollastro, F.; Minassi, A.; Di Marzo, V.; De Petrocellis, L.; Appendino, G. An NMR Spectroscopic Method to Identify and Classify Thiol-Trapping Agents: Revival of Michael Acceptors for Drug Discovery? *Angewandte Chemie International Edition* **2011**, *50*, 467–471.

(45) Böhme, A.; Thaens, D.; Paschke, A.; Schüürmann, G. Kinetic Glutathione Chemoassay To Quantify Thiol Reactivity of Organic Electrophiles—Application to α,β-Unsaturated Ketones, Acrylates, and Propiolates. *Chemical Research in Toxicology* **2009**, *22*, 742–750.

(46) Birkholz, A.; Kopecky, D. J.; Volak, L. P.; Bartberger, M. D.; Chen, Y.; Tegley, C. M.; Arvedson, T.; McCarter, J. D.; Fotsch, C.; Cee, V. J. Systematic Study of the Glutathione Reactivity of N-Phenylacrylamides: 2. Effects of Acrylamide Substitution. *Journal of Medicinal Chemistry* **2020**, *63*, 11602–11614.

(47) Gilbert, K. E.; Vuorinen, A.; Aatkar, A.; Pogány, P.; Pettinger, J.; Grant, E. K.; Kirkpatrick, J. M.; Rittinger, K.; House, D.; Burley, G. A.; Bush, J. T. Profiling Sulfur(VI) Fluorides as Reactive Functionalities for Chemical Biology Tools and Expansion of the Ligandable Proteome. *ACS Chemical Biology* **2023**, *18*, 285–295.

(48) Wade, D.; Airy, S.; Sinsheimer, J. Mutagenicity of aliphatic epoxides. *Mutation Research/Genetic Toxicology* **1978**, *58*, 217–223.

(49) Blaschke, U.; Paschke, A.; Rensch, I.; Schüürmann, G. Acute and Chronic Toxicity toward the Bacteria Vibrio fischeri of Organic Narcotics and Epoxides: Structural Alerts for Epoxide Excess Toxicity. *Chemical Research in Toxicology* **2010**, *23*, 1936–1946.

(50) Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995**, *20*, 273–297.

(51) Friedman, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* **2001**, *29*, 1189 – 1232.

(52) Pal, S.; Mitra, S. Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on Neural Networks* **1992**, *3*, 683–697.

(53) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(54) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. International Conference on Learning Representations. 2017.

(55) Kensert, A.; Desmet, G.; Cabooter, D. MolGraph: a Python package for the implementation of molecular graphs and graph neural networks with TensorFlow and Keras. 2022; `https://arxiv.org/abs/2208.09944v4`.

(56) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Science Advances* **2019**, *5*, eaav6490.

(57) Pope, P. E.; Kolouri, S.; Rostami, M.; Martin, C. E.; Hoffmann, H. Explainability Methods for Graph Convolutional Neural Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019; pp 10764–10773.

(58) Chen, M.; Wei, Z.; Huang, Z.; Ding, B.; Li, Y. Simple and Deep Graph Convolutional Networks. 2020; `http://arxiv.org/abs/2007.02133`.

(59) Hamilton, W. L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. 2018; `http://arxiv.org/abs/1706.02216`.

(60) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. 2018; `http://arxiv.org/abs/1710.10903`.

(61) Bresson, X.; Laurent, T. Residual Gated Graph ConvNets. 2018; `http://arxiv.org/abs/1711.07553`.

(62) Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How Powerful are Graph Neural Networks? 2019; `http://arxiv.org/abs/1810.00826`.

(63) Müller, L.; Galkin, M.; Morris, C.; Rampášek, L. Attending to Graph Transformers. 2023; `http://arxiv.org/abs/2302.04181`.

(64) Monti, F.; Boscaini, D.; Masci, J.; Rodolà, E.; Svoboda, J.; Bronstein, M. M. Geometric deep learning on graphs and manifolds using mixture model CNNs. 2016; `http://arxiv.org/abs/1611.08402`.

(65) Brody, S.; Alon, U.; Yahav, E. How Attentive are Graph Attention Networks? 2022; `http://arxiv.org/abs/2105.14491`.

(66) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chemical Reviews 103*, 1793–1874.

(67) Geerlings, P.; Chamorro, E.; Chattaraj, P. K.; De Proft, F.; Gázquez, J. L.; Liu, S.; Morell, C.; Toro-Labbé, A.; Vela, A.; Ayers, P. Conceptual density functional theory: status, prospects, issues. *Theoretical Chemistry Accounts* **2020**, *139*, 36.

(68) Flores-Holguín, N.; Frau, J.; Glossman-Mitnik, D. Conceptual DFT as a chemoinformatics tool for the study of the Taltobulin anticancer peptide. *BMC Research Notes* **2019**, *12*, 442.

(69) Palazzesi, F.; Grundl, M. A.; Pautsch, A.; Weber, A.; Tautermann, C. S. A Fast Ab Initio Predictor Tool for Covalent Reactivity Estimation of Acrylamides. *Journal of Chemical Information and Modeling* **2019**, *59*, 3565–3571.

(70) Hermann, M. R.; Pautsch, A.; Grundl, M. A.; Weber, A.; Tautermann, C. S. Covalent inhibitor reactivity prediction by the electrophilicity index-in and out of scope. *J. Comput. Aided Mol. Des.* **2021**, *35*, 531–539.

(71) Hughes, T. B.; Dang, N. L.; Miller, G. P.; Swamidass, S. J. Modeling Reactivity to Biological Macromolecules with a Deep Multitask Network. *ACS Central Science* **2016**, *2*, 529–537.

(72) Zhang, Z.; Guiley, K. Z.; Shokat, K. M. Chemical acylation of an acquired serine suppresses oncogenic signaling of K-Ras(G12S). *Nature Chemical Biology* **2022**, *18*, 1177–1183.

(73) Rummey, C.; Metz, G. Homology models of dipeptidyl peptidases 8 and 9 with a focus on loop predictions near the active site. *Proteins: Structure, Function, and Bioinformatics* **2007**, *66*, 160–171.

(74) Sootome, H. et al. Futibatinib Is a Novel Irreversible FGFR 1–4 Inhibitor That Shows Selective Antitumor Activity against FGFR-Deregulated Tumors. *Cancer Research* **2020**, *80*, 4986–4997.

(75) Andersen, J. L.; Gesser, B.; Funder, E. D.; Nielsen, C. J. F.; Gotfred-Rasmussen, H.; Rasmussen, M. K.; Toth, R.; Gothelf, K. V.; Arthur, J. S. C.; Iversen, L.; Nissen, P. Dimethyl fumarate is an allosteric covalent inhibitor of the p90 ribosomal S6 kinases. *Nature Communications* **2018**, *9*, 4344.

(76) Li, X. et al. Discovery of a Potent and Specific M. tuberculosis Leucyl-tRNA Synthetase Inhibitor: (S)-3-(Aminomethyl)-4-chloro-7-(2-hydroxyethoxy)benzo[c][1,2]oxaborol-1(3H)-ol (GSK656). *Journal of Medicinal Chemistry* **2017**, *60*, 8011–8026.

(77) Zhang, Z.; Guiley, K. Z.; Shokat, K. M. Chemical acylation of an acquired serine suppresses oncogenic signaling of K-Ras(G12S). *Nature Chemical Biology* **2022**, *18*, 1177–1183.

(78) Meric-Bernstam, F.; Bahleda, R.; Hierro, C.; Sanson, M.; Bridgewater, J.; Arkenau, H.-T.; Tran, B.; Kelley, R. K.; Park, J. O.; Javle, M.; He, Y.; Benhadji, K. A.; Goyal, L. Futibatinib, an Irreversible FGFR1-4 Inhibitor, in Patients with Advanced Solid Tumors Harboring FGF/FGFR Aberrations: A Phase I Dose-Expansion Study. *Cancer Discovery* **2022**, *12*, 402–415.

(79) Bruno, S.; Pinto, A.; Paredi, G.; Tamborini, L.; De Micheli, C.; La Pietra, V.; Marinelli, L.; Novellino, E.; Conti, P.; Mozzarelli, A. Discovery of Covalent Inhibitors of Glyceraldehyde-3-phosphate Dehydrogenase, A Target for the Treatment of Malaria. *Journal of Medicinal Chemistry* **2014**, *57*, 7465–7471.

(80) Tolmachev, A.; Sakai, A.; Todoriki, M.; Maruhashi, K. Bermuda Triangles: GNNs Fail to Detect Simple Topological Structures. 2021; `http://arxiv.org/abs/2105.00134`.

(81) Awoonor-Williams, E.; Rowley, C. N. Evaluation of Methods for the Calculation of the pKa of Cysteine Residues in Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 4662–4673.

(82) Jöst, C.; Nitsche, C.; Scholz, T.; Roux, L.; Klein, C. D. Promiscuity and Selectivity in Covalent Enzyme Inhibition: A Systematic Study of Electrophilic Fragments. *J. Med. Chem.* **2014**, *57*, 7590–7599.

(83) Rao, S. et al. Leveraging Compound Promiscuity to Identify Targetable Cysteines within the Kinome. *Cell Chemical Biology* **2019**, *26*, 818–829.e9.

(84) Kuljanin, M.; Mitchell, D. C.; Schweppe, D. K.; Gikandi, A. S.; Nusinow, D. P.; Bulloch, N. J.; Vinogradova, E. V.; Wilson, D. L.; Kool, E. T.; Mancias, J. D.; Cravatt, B. F.; Gygi, S. P. Reimagining high-throughput profiling of reactive cysteines for cell-based screening of large electrophile libraries. *Nature Biotechnology* **2021**, *39*, 630–641.

(85) Roth, G. J.; Stanford, N.; Majerus, P. W. Acetylation of prostaglandin synthase by aspirin. *Proceedings of the National Academy of Sciences* **1975**, *72*, 3073–3076.

(86) Ortlund, E.; Lacount, M. W.; Lewinski, K.; Lebioda, L. Reactions of Pseudomonas 7A Glutaminase-Asparaginase with Diazo Analogues of Glutamine and Asparagine Result in Unexpected Covalent Inhibitions and Suggests an Unusual Catalytic Triad Thr-Tyr-Glu,. *Biochemistry* **2000**, *39*, 1199–1204.

(87) Su, H. et al. Identification of pyrogallol as a warhead in design of covalent inhibitors for the SARS-CoV-2 3CL protease. *Nat. Commun.* **2021**, *12*, 3623.

(88) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* **2011**, *40*, D1100–D1107.

(89) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL bioactivity database: an update. *Nucleic Acids Research* **2013**, *42*, D1083–D1090.

(90) Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J. P. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Research* **2015**, *43*, W612–W620.

(91) Falgueyret, J.-P.; Oballa, R. M.; Okamoto, O.; Wesolowski, G.; Aubin, Y.; Rydzewski, R. M.; Prasit, P.; Riendeau, D.; Rodan, S. B.; Percival, M. D. Novel, Nonpeptidic Cyanamides as Potent and Reversible Inhibitors of Human Cathepsins K and L. *J. Med. Chem.* **2001**, *44*, 94–104.

(92) Shealy, Y. F.; Krauth, C. A.; Struck, R. F.; Montgomery, J. A. 2-Haloethylating agents for cancer chemotherapy. 2-Haloethyl sulfonates. *Journal of Medicinal Chemistry* **1983**, *26*, 1168–1173.

(93) Xiao, Z.; Zhou, Z.; Chu, C.; Zhang, Q.; Zhou, L.; Yang, Z.; Li, X.; Yu, L.; Zheng, P.; Xu, S.; Zhu, W. Design, synthesis and antitumor activity of novel thiophene-pyrimidine derivatives as EGFR inhibitors overcoming T790M and L858R/T790M mutations. *European Journal of Medicinal Chemistry* **2020**, *203*, 112511.

(94) Prime, M. E. et al. Discovery and Structure–Activity Relationship of Potent and Selective Covalent Inhibitors of Transglutaminase 2 for Huntington's Disease. *Journal of Medicinal Chemistry* **2012**, *55*, 1021–1046.

(95) Li, X. et al. Discovery of a Potent and Specific M. tuberculosis Leucyl-tRNA Synthetase Inhibitor: (S)-3-(Aminomethyl)-4-chloro-7-(2-hydroxyethoxy)benzo[c][1,2]oxaborol-1(3H)-ol (GSK656). *Journal of Medicinal Chemistry* **2017**, *60*, 8011–8026.

(96) Zha, G.-F.; Wang, S.-M.; Rakesh, K.; Bukhari, S.; Manukumar, H.; Vivek, H.; Mallesha, N.; Qin, H.-L. Discovery of novel arylethenesulfonyl fluorides as potential candidates against methicillin-resistant of Staphylococcus aureus (MRSA) for overcoming multidrug resistance of bacterial infections. *European Journal of Medicinal Chemistry* **2019**, *162*, 364–377.

(97) Kozaki, R.; Yoshizawa, T.; Tohda, S.; Yasuhiro, T.; Hotta, S.; Ariza, Y.; Ueda, Y.; Narita, M.; Kawabata, K. Development of a Bruton's Tyrosine Kinase (Btk) Inhibitor, ONO-WG-307: Efficacy in ABC-DLBCL Xenograft Model – Potential Treatment for B-Cell Malignancies,. *Blood* **2011**, *118*, 3731.

(98) Kalgutkar, A. S.; Kozak, K. R.; Crews, B. C.; Hochgesang, G. P.; Marnett, L. J. Covalent Modification of Cyclooxygenase-2 (COX-2) by 2-Acetoxyphenyl Alkyl Sulfides, a New Class of Selective COX-2 Inactivators. *Journal of Medicinal Chemistry* **1998**, *41*, 4800–4818.

(99) Daylight Chemical Information Systems, Inc.,. `https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html`, Accessed October 26, 2023.

TOC Graphic