In Silico Target Prediction for Elucidating the Mode of Action of Herbicides Including Prospective Validation

Rucha K. Chiddarwar [1], Sebastian G. Rohrer [2], Antje Wolf [3], Stefan Tresch [2], Sabrina Wollenhaupt [3], Andreas Bender [1,]*

[1] Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[2] Global Research Crop Protection, BASF SE, Speyerer Strasse 2, 67177 Limburgerhof, Germany

[3] Computational Chemistry and Biology, BASF SE, Carl-Bosch-Strasse 38, 67056 Ludwigshafen, Germany

Keywords: Herbicide targets; Mode of Action; *in silico* target prediction; MOLPRINT 2D

Abstract

The rapid emergence of pesticide resistance has given rise to a demand for herbicides with new mode of action (MoA). In the agrochemical sector, with the availability of experimental high throughput screening (HTS) data, it is now possible to utilize *in silico* target prediction methods in the early discovery phase to suggest the MoA of a compound *via* data mining of bioactivity data. While having been established in the pharmaceutical context, in the agrochemical area this approach poses rather different challenges, as we have found in this work, partially due to different chemistry, but even more so due to different (usually smaller) amounts of data, and different ways of conducting HTS. With the aim to apply computational methods for facilitating herbicide target identification, 48,000 bioactivity data against 16 herbicide targets were processed to train Laplacian modified Naïve Bayesian (NB) classification models. The herbicide target prediction model ("HerbiMod") is an ensemble of 16 binary classification models which are evaluated by internal, external and prospective validation sets. In addition to the experimental inactives, 10,000 random agrochemical inactives were included in the training process, which showed to improve the overall balanced accuracy of our models up to 40%. For all the models, performance in terms of balanced accuracy of ≥ 80% was achieved in five-fold cross validation. Ranking target predictions was addressed by means of z-scores which improved predictivity over using raw scores alone. An external testset of 247 compounds from ChEMBL and a prospective testset of 394 compounds from BASF SE tested against five well studied herbicide targets (ACC, ALS, HPPD, PDS and PROTOX) were used for further validation. Only 4% of the compounds in the external testset lied in the applicability domain and extrapolation (and correct prediction) was hence impossible, which on one hand was surprising, and on the other hand illustrated the utilization of using applicability domains in the first place. However, performance better than 60% in balanced accuracy was achieved on the prospective testset,

where all the compounds fell within the applicability domain, and which hence underlines the possibility of using target prediction also in the area of agrochemicals.

Introduction

In agriculture, crop protection is carried out by application of chemical pestices of which herbicides are the extensively used class that account for nearly 60% of pesticide applications.[1] But application of herbicides have been found to have environmental effects concerning off-target organisms and resistance issues. Herbicide resistance can be defined as the condition of a plant's ability to withstand the standard dose of a herbicide as a consequence of genetic response to frequent exposure to herbicides with an analogous mode of action (MoA).[2] More than 450 herbicide-resistant weeds have been catalogued across the world, and this number is expected to grow even further.[3] To combat the resistnace issue, an often-used approach is one where herbicide-resistant weed is managed by rotating with an alternative herbicide from a different MoA class in practice.[4]

As a result of such resistance developments, the search for herbicides with novel modes of action remains a high priority[5] - "There is an urgent need for new herbicide options or a new weed management paradigm", as has been commented by Tranel *et al*.[4] In addition to loosing biological efficacy, regulatory issues also led to some of the older herbicides with unique MoAs losing their marketplace in a few countries, for example the banning of paraquat in some European countries, making the development of ingredients with novel (and in some way superior) properties important at the current stage.[6] Over the last 40 years, 270 herbicide active ingredients possessing 17 identifiable and distinct modes of action have been found, based on the empirical screening of chemicals on the whole target organisms. For new herbicide chemistries, the mode of action has always been discovered in retrospect even though subsequent attempts have been made to utilize this knowledge to aid in discovery and optimization of further instances.[7]

In the last couple of decades, however, the initial stages of herbicide discovery have adopted the application of *in vitro* assays against specific molecular targets as well as high throughput

screening (HTS), in addition to the previously existing direct testing of compounds on whole plants (i.e., conventional phenotypic screens).[8,9]

This wealth of target-based bioactivity data can now be utilized for *in silico* target prediction, which has been established in other areas for a while now - in drug discovery, the concept of employing bioactivity data for ligand-target prediction (for a review on target prediction, please refer the work of Jenkins et al.[10] and Koutsoukas et al.[11]) has received considerable attention in recent times for identification of mode of action. [10,12,13,14,15] The purpose of *in silico* approaches is that by exploiting prior knowledge of ligand-target interactions, available in various databases either public or commercial, make knowledge based predictions for novel molecules or to suggest new probable target interactions for previously marketed compounds. This further aids in the hypothesis of mode of action for new molecules. The most commonly used *in silico* target prediction methods are the multiclass Naïve Bayes[16], Similarity Ensemble Approach (SEA)[17,18], Support Vector Machines (SVM)[19], the PASS method (Prediction of the Activity Spectra of Substances)[20] , Random Forests (RFs)[21,22], Parzen-Rosenblatt Window[23], and the Winnow algorithm[24]. Although a number of methodologically different *in silico* target prediction approaches exist, in the following we will limit our analysis to ligand-based target prediction algorithms that employs fingerprints combined with Bayes-based approaches.

In one of the earlier studies, Multiple-Category Bayesian models were employed by Nidhi et al.,[16] for 964 target classes where data from the WOMBAT database[25] were used for model training which predicted the three most likely protein targets for all MDDR database compounds, leading to on an average of 77% correct predictions. In the studies by Nigsch et al.,[24] Naïve Bayesian and Winnow algorithms were applied on 20 drug targets from WOMBAT database to compare and evaluate their performances. Both classifiers were observed to perform similar overall, but differed significantly between target classes and

among individual structures. Similarly, Koutsoukas et al.[15] compared the Naïve Bayes and Parzen-Rosenblatt Window by training the classifiers on a benchmark dataset on 894 human protein targets with more than 155,000 compound-protein pairs; and achieved a recall of 63.3% and 66.6% on an external dataset. The study also suggested that the model performance significantly depends on the class size and underlying diversity of the chemical compounds; where with low structural similarity and small class size the performance is badly affected.

Several target prediction studies – such as the ones above - have been based on the Naïve Bayes classifier (NB) in the field of drug discovery, and generally good prediction performance has been obtained at suitable speed of training models, which was likely helped by the ability of the models to handle noisy data reasonably well.[16]

Computational methodologies have been previously explored and developed for the rational design of agrochemicals.[26] In the last couple of decades, approaches based on graph-theoretical descriptors have been explored for the design of bioactive agents.[27,28] There also have been computational approaches for identification[29] and classification of fungicides based on toxicity[30], and QSAR models designed towards the design of fungicides with a defined resistance risk using sub-structural descriptors.[31] These methodologies, especially in fungicides, were intended for the discovery and identification of novel leads that are potential candidates with a wide spectrum of action against various fungicide species and possibly act by means of different modes of action with low resistance issues. The objective of such methods was to execute substantial screenings of available databases of heterogeneous series of compounds and to extract possible structural information at different levels of molecular diversity and complexity. These methods, supported by computer-aided drug design techniques, have been developed rapidly in recent years.[31]

In this work, taking into consideration this previous experience, we now attempted to extend

the work of *in silico* target prediction to the area of agrochemicals, namely to herbicide target prediction and elucidation of modes of action (visualized in Figure 1). Figure 1 also summarizes the data we had at hand for model training, comprising 16 herbicide target classes with nearly 48,000 tested compounds (known actives and inactives) across different species and including external and prospective test sets for further validation. Hence the herbicide target prediction model ("HerbiMod") is an ensemble of 16 binary classification models which are evaluated by internal, external and prospective validation sets (see following sections for details).

Materials and Methods

Data preparation and standardization

In the current work, BASF proprietary compound screening database was queried and a subset of all herbicide protein targets and tested compounds of the whole bioactivity data were made available. Herbicide targets possessing either less than 50 tested compounds (including both actives and inactives) or very few (<10) active compounds were excluded from further analysis since it was apparent from previous studies that models generated on very small datasets would not be reliable.[32] After filtering out such targets, the dataset consisted of 16 targets associated with around 48,000 compounds that were tested in enzyme assays, with bioactivities measured in either dose-response curves (i.e. providing IC_{50} values) or percentage inhibition values (Table 1; Note that the number of datapoints listed here does not necessarily represent all data available within BASF, but rather the subset used for training this particular type of model).

This herbicide dataset consisted of compounds tested on some of the well-explored herbicide targets that were organism specific (plant species), namely ACC (Acetyl-CoA carboxylase, Uniprot-Q94FR5)[6,33,34,35,36,37], ALS (acetolactate synthase, Uniprot-Q94FR5)[6,33,7,38,39], HPPD (4-hydroxyphenylpyruvate dioxygenase, Uniprot-P93836)[20,7,28,41,42,1], PROTOX (protoporphyrinogen IX oxidase, Uniprot-Q9FYV8)[33,7,1,43,44], and PDS (phytoene desaturase, Uniprot-P26294)[33,7,42,45,46] as well as a number of proprietary targets which cannot be mentioned in this study. The targets are from here on referred to as H01-H16.

Bioactivity threshold

Defining a bioactivity threshold for considering compounds as actives or inactives varied considerably between target classes – and this was a major task in the current work, compared to (relatively) more homogenous bioactivity data available in the pharmaceutical

area. In the pharmaceutical context, e.g., in HTS campaigns, often compounds with a bioactivity of better than 10μM ($pIC_{50} > 5$; where $pIC_{50} = -\log_{10}IC_{50}$) are considered as actives.[26, 27] However, for the data at hand in the current study, this would in some cases have led to a very small number of actives (or even no actives), and in other cases to too large a number of active compounds – it seems that agrochemical screening data, and even less so than pharmaceutical screening data, rarely follows a neat normal distribution of bioactivities amenable to such consistent bioactivity cut-offs across assays. As the compounds for this study come from historical data (and from sometimes multiple assays which have been performed for a given target), the bioactivity threshold to be chosen *a priori* in the majority of assays was unknown. Hence, two options were investigated to proceed in practice. One option was the utilization of fixed activity thresholds, since for some target assays the activity threshold was known (as suggested by biological experts) *i.e.*, for actives $pIC_{50} \geq 6$ and inactives $pIC_{50} \leq 5.3$. However for the majority of target assays this information was not available, and hence after manually examining the activity distributions in different bioactivity classes a fixed threshold of $pIC_{50} \geq 5$ (10 μM) was used for actives, and $pIC_{50} \leq 4.3$ (50 μM) was used for inactives. A gap of 0.7 log unit was defined to increase the likelihood of 'active' compounds being indeed active and 'inactive' compounds being indeed inactive, given experimental uncertainty. (This gap is similar to the one suggested in recent studies discussing experimental uncertainty, which mentions a mean error of 0.44 pK_i units[47] for bioactivity data from public sources, which similar to the current study also span different types of experimental assay setups.) The other option was using compounds with all IC_{50} values as actives (since in cases where dose-response curves have been determined this led us to conclude that the compounds have been considered as 'actives' in this particular screen), and all compounds with percentage inhibition values below 50% as inactives. Both the options were investigated on all target classes for model building and based on model

performance class-specific decision was taken.

Tautomeric compounds and duplicates were manually analysed and annotated as active or inactive or discarded. Subsequently, a unique target identifier from UniProt was assigned to all targets, enabling easy access to protein information on all the target classes.

Increasing the size of the inactive dataset

In order to increase the chemical space covered by the inactive dataset a set of 10,000 inactive compounds was randomly extracted from the BASF in-house database comprising all the pesticide indications (herbicides, fungicides and insecticides) which were then included in all the 16 inactive datasets. This step has been found to be necessary to generate sets of inactive compounds of sufficient size, given that some targets only contained an insufficient number of experimentally annotated inactive compounds. In machine learning experiments, it has previously been observed that the inclusion of random inactive training instances influences the performance of the classifier, and at least in some cases also shows improvement in precision.[48] The influence of adding random inactives to the known inactives on model performance has been previously analysed[48,49] and was also further evaluated in our work on this particular pesticide dataset, where two separate models, one with, and the other without inclusion of random inactives were trained and analysed. The average predictive measures were calculated from five-fold cross validation and compared for both the models on individual targets.

External validation dataset

With the pesticide-like (agrochemical) data recently becoming available in ChEMBL database, ChEMBL_20[50] was used to generate an external test set for the five targets mentioned in the current study that are tested only on the plant species. As not all the

compounds were annotated as actives or inactives we applied the fixed threshold for this dataset where compounds with IC_{50}, K_i and ED_{90} values below 10 µM were considered as actives, and those with values equal to or above 10 µM, as well as compounds with less than 50% inhibition values, were considered as inactives. Duplicates were removed and compounds with conflicting bioactivities were discarded. A total of 250 compounds thus formed the external test set, comprising 165 active and 85 inactive compounds.

Prospective validation dataset

With the aim to apply HerbiMod to novel unseen compounds and to further estimate the predictive ability on such new molecules, 394 compounds were prospectively tested in the assays against ACC, ALS, HPPD, PDS and PROTOX by BASF. Standard inhibitors were included in the dataset that was checked for consistency in order to allow comparison of data. This testset consisted of very few active compounds in each class; 7 in ACC; 11 in ALS; 2 in HPPD; 6 in PDS and 65 in PROTOX. None of the 394 compounds were included in any of the previous datasets used for model building or model evaluation and hence constitute a true prospective validation dataset.

Molecular fingerprints

In the current work, Atom Environments[39] also known as MOLPRINT 2D descriptors, were used as molecular representation. These atom environments are similar to Scitegic ECFP fingerprints, Signature Molecular Descriptors[35] and also Augmented Atoms.[36] In MOLPRINT 2D [51], an individual atom fingerprint is calculated for each of the heavy atoms in a molecule which have been assigned Sybyl mol2 atom types, and stored in a count vector, where the vector elements are counts of atom types at a certain distance from the central atom. In this work, calculations were performed by employing distances from 0 up to 2

bonds.

Feature selection

Feature selection[40] is originally employed to select the top features for the nodes of a decision tree; however the underlying concept of information content is in general applicable.[51] The information gain measure of Quinlan[41,42] was employed for computation of information content of individual atom environments. Bender *et al.*[51] in their work have shown how a number of selected features can influence the mean percentage of active compounds found in the Top 5% of the ranked library and so 250 features were selected in the current study for each model generation.

Classification using Laplacian-modified Naïve Bayesian classifier

The Naïve Bayes (NB) classifier[51] was used as the classification method which follows the Bayes rule of conditional probability. The classifier was trained on the feature vector (*F*) containing features f_i with their associated target classes (*TC*) for a given dataset. For a new feature vector the classifier then predicts the class it belongs to as the class with maximum probability of *P(TC|F)* which is calculated as follows:

$$P(TC_n|F) = \frac{P(TC_n)\,P(F|TC_n)}{P(F)}$$

where $P(TC_n)$ is the probability of class *n*, *P(F)* is the probability of feature vector and $P(F|TC_n)$ is the probability of feature *F* for target class *n*. This classifier is said to be 'naïve' because of its assumption that the features are independent. Based on this assumption, for two datasets (actives and inactives for a given target class), the binary NB classifier used for classification is given as

$$\frac{P(TC_1|F)}{P(TC_2|F)} = \frac{P(TC_1)}{P(TC_2)} \prod_i \frac{P(f_i|TC_1)}{P(f_i|TC_2)}$$

where molecules for the given dataset are characterised by their feature vectors *F*. Laplacian correction was employed in order to keep the calculations running when the feature for a new molecule is not observed in the training datasets. We consider highest probability ratios to most likely belong to class 1 (active class in our case) and lowest values to class 2 (inactive class) for a given target class.

Model assessment and validation

Studies[43] have shown that model accuracy can be determined by building a model using a portion of the training dataset, and computing the prediction accuracy of the model using the rest of the dataset.[44]

To assess our models, *v*-fold cross validation[52] technique of model assessment was performed. Five-fold cross-validation method was performed where the datasets were split into five equal folds. Each time one fold was used as the test set while the rest was used for training the classifier. The trained classifier was then tested with the test set. This process was repeated five times over each fold, thus ensuring that every compound was used in the prediction once. The prediction results over all the folds were then averaged to compute the average predictive measures.

Predictive measures

The predictive measures, viz., true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rate were computed for a range of score cut-offs. For all the classification models, a mean of all the five-folds was calculated followed by sensitivity (SN) and specificity (SP). Sensitivity and specificity given by recall and precision, respectively, are complementary measures commonly used for information retrieval and are used to measure the effectiveness of a classifier.[46]

The sensitivity is calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN}$$

and specificity is calculated as follows:

$$Specificity = \frac{TN}{TN + FP}$$

Balanced accuracy measures the performance of a model by valuating the two classes equally regardless of the class size and is considered to be more effective than using accuracy alone. It is computed using the equation given below,

$$Balanced\ accuracy = \frac{Sensitivity + Specificity}{2}$$

In machine learning, to measure the quality of binary (two class) classifications, the Matthews correlation coefficient (MCC)[47] is one of the frequently used predictive measures and was also calculate in our work as follows:

$$MCC = (TP * TN) - (FP * FN) / \sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}$$

However, it can fail to provide a fair evaluation, for example, MCC will be high in cases where the predictive models give very few or no false positives but at the same time very few true positives.[53]

Optimization of classification cut-offs

Probabilistic modelling techniques generally result in a predictive probability which is then transformed into a "yes or no" (active or inactive in our study) classification. Conventionally, the default is set to use a threshold of 0.5 but this does not certainly result in the highest prediction accuracy and hence represents parameters that need to be optimized.[54] The thresholds, alternatively, can also be chosen to optimize classification accuracy which is to

empirically determine the prior as is described Bayesian approaches.[55] In this work, for each model a range of thresholds were computed based on the raw scores. At each threshold all the predictive measures were computed and a cut-off was chosen based on the highest balanced accuracy.

Transforming raw scores from Naïve Bayesian to standard score (z-score)

The probable targets for the test compounds were ranked based on z-score statistics (also referred to as a 'standard score'). The z-score approach[56] allows to calculate the probability of a score occurring within a distribution (generally, normal distribution) and aids to compare two scores that are from two different target distributions.[57] The z-score is calculated from raw score of each data point in a sample or population relative to the sample's mean and standard deviation as follows:

$$Z = \frac{x - \mu}{\sigma}$$

where x is the raw score, and μ and σ are the mean and standard deviation of a population (active training set for each model, in our study), respectively. In the current work, each of the active dataset per model was input into its own model to calculate the mean and standard deviation obtained for active compounds from their raw scores. Z-scores were then calculated for both actives and inactives using the above equation.

Based on z-scores for each target model, recall analysis was then performed to evaluate every model's ability to predict the true target in the top 3 positions. For this analysis, active compound set of each target class were tested on all the final 16 models. The Naïve Bayes raw score for each tested (active) compound was transformed into a z-score for all the predicted targets. The z-scores were sorted in descending order, with the hypothesis that moving from raw scores to z-scores would allow for an improved comparison between models.

Applicability domain (AD)

In order to define AD for our herbicide and fungicide models, the k-nearest neighbours method[58] (k-NN, k=5 in our work) was employed. For all the target classes, the dataset was divided into five sections as done in five-fold cross validation. For each fold – as a test set, the final model was applied to predict activity class (active/inactive). Each prediction was then analysed for prediction error, where for each false prediction the prediction error was counted. Further, for each test compound, distance was calculated by averaging similarities (in terms of Tanimoto coefficient) of five nearest training neighbours (NN) and computing 1-NN. Theoretically, lower distance values correspond to a higher similarity, whereas the increasing distances indicate higher levels of structural difference. These calculations were performed for all the five folds and merged for each target class. The distances were binned into thresholds (nbins = 40) and frequencies of number of compounds along with number of prediction errors at the distance were estimated. Plotting prediction error against the distance assisted in defining the maximum acceptable prediction error and also aided in deriving the maximum allowed distance for the new test compounds.

Results and Discussions

Diversity analysis

For understanding the chemical diversity of our herbicide dataset, pair-wise Tanimoto similarity was computed for each active compound with every other active compound in its activity class[59], and distributions were plotted as shown in Figure 2. It is observed that the majority of the target classes consisted of rather diverse active compounds, where the mean similarity was less than 0.3 in all the target classes. Taking into consideration the active datasets for four of the five well studied targets, it can be seen that target classes ACC, HPPD, PDS and PROTOX on average had less than 0.1 Tanimoto similarity, while for ALS the similarity was fairly broadly distributed with a mean of 0.27. Hence, this analysis shows that the current dataset is composed of a diverse set of bioactive compounds which is used for model building and that might capture a relatively wider chemical space for our predictive models. Employing a diverse (or heterogeneous) data however represents a heavy limit for most of the *in silico* methods given that the models are extensively dependent on the data used.

Model validation and evaluation

Table 2 summarizes the predictive measures MCC, sensitivity, specificity and balanced accuracy, obtained at the NB score threshold (cut-off) where highest balanced accuracy was observed. The model performance was observed to be more than 80% for all the models with a mean performance of 92.42% and standard deviation of 6.6%. It was observed that for herbicide models H03, H04, H06, H08 and H16 had classification accuracy of 0.9 to 1.0 in terms of balanced accuracy at a target-specific threshold. In Figure 2, it can be seen that for these target classes the compounds were observed to be relatively less diverse with average similarity between 0.2 – 0.3 in terms of Tc, thus indicating high predictive ability of these

models. Also, the z-score distributions between actives and inactives of these classes showed a clear separation (refer Figure S1), further suggesting such high performance. While for the other target models that included target like ACC, ALS, HPPD, PDS, PROTOX, slightly lower model performance was observed which may be because of the high diversity (similarity less than 0.2 as seen in Figure 2) and also due to the overlapping region in the z-score distributions (seen in Figure S2) of these target class.

The results of the recall analysis in top 1-3 positions based on raw scores and z-scores, were plotted as seen in Figure 3. It was observed that for 10 of the 16 modes, raw score gave higher predictions in top 3 positions than z-score. For all the models a recall of more than 80% was achieved in the top 3 positions together based on z-scores. Except for target H01-ACC, for all the other target models more than 90% of the true targets of compounds were predicted in the first position. Based on these results, raw scores were used for activity classification (active/inactive), and since z-scores provide confidence to the predictions given the fact that they are drawn from the target-specific distributions, the ranking of active predicted targets was determined by z-score.

Influence of adding random inactives on model performance

On evaluating the balanced accuracies for all the targets, it was found that inclusion of random inactives increased the model performance (balanced accuracy) by an average of 13% (see Figure 4 and Table S1). Hence overall it was clearly seen that the addition of random inactives played a positive role in the successful classification of compounds which was also observed by Kurczab et al.[48] and Heikamp et al.,[60] in their studies.

Defining applicability domain

By employing k-nearest neighbours approach (k-NN, k=5 in our work), AD was defined by computing the distance (1-Tc) for each compound of the test set to the nearest neighbour in the training set and was achieved for all the folds as divided in a five-fold cross validation method. The distances were binned into thresholds (0.0 to 1.0) and frequencies of number of compounds along with number of prediction errors at the distance were estimated. Plotting the distance to prediction error assisted in defining the maximum acceptable prediction error and also aided in deriving the maximum allowed distance for the new test compounds, which in case of our models was set to 0.09 to 0.42. The maximum allowed distance (in terms of Tc) for each target model is listed in Table 3. Further, five nearest training neighbours to the test set were computed based on Tanimoto similarity coefficient to assist in extrapolation of the applicability domain.

External validation

Next, the compounds extracted from ChEMBL database were investigated for model validation, with the first step being the estimation whether they lie in the applicability domain of our models. Surprisingly, only 15 compounds (listed in Table S2) were found to have Tc similarity > 0 to the training sets, meaning that the remainder of the 235 compounds had no similarity, or in fact shared features, at all, to the compounds used in the training set and hence lied outside the applicability domain of our models. Accordingly, results from HerbiMod showed that for the compounds with no similarity to the compounds in the training set, no targets were predicted above the cut off, and for the 15 compounds with similarity greater than zero (maximum Tc of 0.034) the predictive measures were calculated, TP = 6, TN = 2, FP = 4 and FN = 3. Also, the dataset contained three well-known marketed HPPD

inhibitors like sulcotrione, mesotrione and nitisinone (Figure 5) that were correctly predicted by HerbiMod as inhibitors of HPPD. The average Tc similarity of the five nearest neighbours in the active set of HPPD was 0.034, 0.033 and 0.035 for nitisinone, mesotrione and sulcotrione, respectively, and even though the structural similarity of these compounds was very low compared to the training dataset, the model could predict the target HPPD correctly for these compounds.

However, overall due to small sample size not many conclusions can be drawn from this result, except that chemistry even against the same target, when derived from different sources, can be very different, and that it is hence important to determine if a model is applicable to those new parts of chemical space in the first place. On a positive note, the current models can be updated with the data becoming available from either in-house or public databases, and trained in a semi-automated process, and then be employed for prediction on a wider chemical space.

Prospective validation

Next, a prospective test set consisting of nearly 400 compounds were experimentally tested in-house against ACC, ALS, HPPD, PDS and PROTOX. The Tc similarity of these compounds to the nearest neighbour in the training set was analysed and found to be in the range of 0.21 to 0.6 (mean = 0.4 and stdv =0.2), with none of the compounds having a similarity of zero, therefore suggesting that this dataset lies in the applicability domain of our models. The results of predictive measures (sensitivity, specificity and accuracy) for all 5 target models on the prospective test set are summarized in Figure 6. All the five target models achieved an accuracy of more than 60%. In case of targets like ACC and PROTOX, the specificity was 78% to 90%, but the sensitivity was only around 43% and 46%,

respectively. The targets ALS, HPPD and PDS fared better, where the sensitivity was 73%, 100% and 67%, respectively, and specificity amounted to 84%, 96% and 68%. It is also important to note that very few compounds that were actual actives were present in this dataset, for example, in case of HPPD only two actives were tested. Although these results are lower for some target classes than the results of the five-fold cross validation, they are likely to be closer to real world applications, where new compounds may come from a previously un- or underexplored area of the chemical space.

Conclusion

In the current work, with the aim to utilize historical agrochemical bioactivity data from a large agrochemical company, an *in silico* herbicide target prediction model was developed by employing a Naïve Bayesian classifier trained on molecules tested against 16 herbicide targets.

Assigning activity thresholds turned out to be less trivial than one might have anticipated, and on investigating the approaches for differentiating actives and inactives; for five target classes a fixed threshold was set, while for the other eleven targets the type of reported activity units (compounds with IC_{50} values as actives and inhibition values as inactives) was used for differentiation.

All the herbicide models achieved an averaged balanced accuracy of more than or equal to 80% in the five-fold cross validation. Classification was achieved by utilising target-specific raw NB score cut-off and ranking of the probable targets was established by z-scores, since z-score represented the underlying distribution of compounds (actives/inactives) for each target class. When generating datasets, 10,000 agrochemical random inactives were included in the inactive training set, a step which was found to have a minimum of 3% and maximum of 40% positive influence on all of the model's performance in terms of balanced accuracy, which was also observed to be in agreement with similar previous studies.

The external validation set used for assessing our models represented completely new chemistry and lied outside the applicability domain of the models, and hence the model was unable to make predictions on such a dataset. This underlines the difference in chemotypes present in bioactivity data from different sources, even against the same protein targets, and the necessity to evaluate applicability domains before applying a model. On the contrary, the prospective validation set represented compounds that lied in the applicability domain of our models and hence extrapolation in this case was possible with 60% overall accuracy on five

targets. The predictive ability of our models on the external test set and prospective set provided a realistic assessment of *in silico* modelling approach and hence presents a real world application. Five nearest training neighbours to the test set were computed based on Tanimoto similarity coefficient to assist in extrapolation of the applicability domain.

With these results obtained from our approach, we conclude that exploiting in silico target prediction indeed presents a way to elucidate mode of actions for new agrochemical compounds. This however is highly dependent on the applicability domain of the models, as was apparent from the application of ChEMBL external dataset. Still, on the positive side any new activity data for molecules as well as new targets can easily be added to the model, thereby taking advantage of all data available at any point in time.

Author Information

Corresponding Author

* Phone: +44 1223 762 983. E-mail: ab454@cam.ac.uk. Address: Centre for Molecular Informatics, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, UK. (A. Bender)

Acknowledgments

R.K.C. acknowledges research and funding support by BASF SE. AB thanks Unilever for funding and the European Research Council for a Starting Grant (ERC-2013-StG-336159 MIXTURE).

References

(1) Shaner, D. L. Herbicide Safety Relative to Common Targets in Plants and Mammals. *Pest Manag. Sci.* **2004**, *60* (1), 17–24.

(2) Warwick, S. I. Herbicide Resistance in Weedy Plants: Physiology and Population Biology. *Annu. Rev. Ecol. Syst.*, 1991, *22*, 95–114.

(3) Heap, I. International Survey of Herbicide Resistance Weeds http://weedscience.org/ (accessed Jun 27, 2015).

(4) Tranel, P. J.; Riggins, C. W.; Bell, M. S.; Hager, A. G. Herbicide Resistances in Amaranthus Tuberculatus: A Call for New Options. *J. Agr. Food Chem.* **2011**, *59* (11), 5808–5812.

(5) Müller, U. Chemical Crop Protection Research. Methods and Challenges *. *Pure. Appl. Chem.* **2002**, *74* (12), 2241–2246.

(6) Duke, S. O. Why Have No New Herbicide Modes of Action Appeared in Recent Years? *Pest Manag. Sci.* **2012**, *68* (4), 505–512.

(7) Cole, D.; Pallett, K.; Rodgers, M. Discovering New Modes of Action for Herbicides and the Impact of Genomics. *Pestic. Outlook.* **2000**, *11* (6), 223–229.

(8) Harrison, W. Agrochemical Research and Development in the 21st Century: Bloom or Bust. *J. Biomol. Screen.* **1999**, *4* (2), 61–65.

(9) Lein, W.; Bornke, F.; Reindl, A.; Ehrhardt, T.; Stitt, M.; Sonnewald, U.; Börnke, F.; Reindl, A.; Ehrhardt, T.; Stitt, M.; et al. Target-Based Discovery of Novel Herbicides. *Curr. Opin. Plant Biol.* **2004**, *7*, 219–225.

(10) Jenkins, J. L. Large-Scale QSAR in Target Prediction and Phenotypic HTS Assessment. *Mol. Inf.* **2012**, *31*, 508–514.

(11) Koutsoukas, A.; Simms, B.; Kirchmair, J.; Bond, P. J.; Whitmore, A. V; Zimmer, S.; Young, M. P.; Jenkins, J. L.; Glick, M.; Glen, R. C.; et al. From in Silico Target Prediction to Multi-Target Drug Design : Current Databases , Methods and Applications. *J. Proteomics* **2011**, *74* (12), 2554–2574.

(12) Jenkins, J. L.; Bender, A.; Davies, J. W. In Silico Target Fishing: Predicting Biological Targets from Chemical Structure. *Drug Discov. Today Technol.* **2006**, *3* (4), 413–421.

(13) Plouffe, D.; Brinker, A.; McNamara, C.; Henson, K.; Kato, N.; Kuhen, K.; Nagle, A.; Adrián, F.; Matzen, J. T.; Anderson, P.; et al. In Silico Activity Profiling Reveals the Mechanism of Action of Antimalarials Discovered in a High-Throughput Screen. *PNAS* **2008**, *105* (26), 9059–9064.

(14) Martínez-Jiménez, F.; Papadatos, G.; Yang, L.; Wallace, I. M.; Kumar, V.; Pieper, U.; Sali, A.; Brown, J. R.; Overington, J. P.; Marti-Renom, M. a. Target Prediction for an Open Access Set of Compounds Active against Mycobacterium Tuberculosis. *PLoS Comput. Biol.* **2013**, *9* (10), e1003253.

(15) Koutsoukas, A.; Lowe, R.; Kalantarmotamedi, Y.; Mussa, H. Y.; Klaffke, W.; Mitchell, J. B. O.; Glen, R. C.; Bender, A. In Silico Target Predictions: Defining a Benchmarking Data Set and Comparison of Performance of the Multiclass Naive Bayes and Parzen-Rosenblatt Window. *J. Chem. Inf. Model.* **2013**, *53* (8), 1957–1966.

(16) Nidhi; Glick, M.; Davies, J. W.; Jenkins, J. L. Prediction of Biological Targets for Compounds Using Multiple-Category Bayesian Models Trained on Chemogenomics Databases. *J. Chem. Inf. Model.* **2006**, *46* (3), 1124–1133.

(17) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating Protein Pharmacology by Ligand Chemistry. *Nat. Biotechnol.* **2007**, *25* (2), 197–206.

(18) Chen, B.; McConnell, K. J.; Wale, N.; Wild, D. J.; Gifford, E. M. Comparing Bioassay

Response and Similarity Ensemble Approaches to Probing Protein Pharmacology. *Bioinformatics* **2011**, *27* (21), 3044–3049.

(19) Zernov, V. V.; Balakin, K. V.; Ivaschenko, A. a.; Savchuk, N. P.; Pletnev, I. V. Drug Discovery Using Support Vector Machines. The Case Studies of Drug-Likeness, Agrochemical-Likeness, and Enzyme Inhibition Predictions. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 2048–2056.

(20) Poroikov, V. V; Filimonov, D. a; Borodina, Y. V; Lagunin, a a; Kos, A. Robustness of Biological Activity Spectra Predicting by Computer Program PASS for Noncongeneric Sets of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (6), 1349–1355.

(21) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32.

(22) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947–1958.

(23) Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33* (3), 1065–1076.

(24) Nigsch, F.; Bender, A.; Jenkins, J. L.; Mitchell, J. B. O. Ligand-Target Prediction Using Winnow and Naive Bayesian Algorithms and the Implications of Overall Performance Statistics. *J. Chem. Inf. Model.* **2008**, *48* (12), 2313–2325.

(25) Kenny, P.; Sadowski, J. Chemoinformatics in Drug Discovery. In *Chemoinformatics in Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA, 2005; pp 271–285.

(26) Speck-Planche, A.; Guilarte-Montero, L.; Yera-Bueno, R.; Rojas-Vargas, J. A.; García-López, A.; Uriarte, E.; Molina-Pérez, E. Rational Design of New Agrochemical Fungicides Using Substructural Descriptors. *Pest Manag. Sci.* **2011**, *67* (4), 438–445.

(27) Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds. *J. Med. Chem.* **2000**, *43* (10), 1975–1985.

(28) Estrada, E.; Peña, A.; García-Domenech, R. Designing Sedative/hypnotic Compounds from a Novel Substructural Graph-Theoretical Approach. *J. Comput. Aided. Mol. Des.* **1998**, *12* (6), 583–595.

(29) González-Díaz, H.; Prado-Prado, F. J.; Santana, L.; Uriarte, E. Unify QSAR Approach to Antimicrobials. Part 1: Predicting Antifungal Activity against Different Species. *Bioorg. Med. Chem.* **2006**, *14* (17), 5973–5980.

(30) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Predicting Multiple Ecotoxicological Profiles in Agrochemical Fungicides: A Multi-Species Chemoinformatic Approach. *Ecotoxicol. Environ. Saf.* **2012**, *80*, 308–313.

(31) Speck-Planche, A.; Kleandrova, V. V.; Rojas-Vargas, J. a. QSAR Model toward the Rational Design of New Agrochemical Fungicides with a Defined Resistance Risk Using Substructural Descriptors. *Mol.Divers.* **2011**, *15* (4), 901–909.

(32) Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **2010**, *29* (6-7), 476–488.

(33) Wakabayashi, K.; Böger, P. Target Sites for Herbicides: Entering the 21st Century. In *Pest Manag. Sci.*; 2002; Vol. 58, pp 1149–1154.

(34) Kukorelli, G.; Reisinger, P.; Pinke, G. ACCase Inhibitor Herbicides – Selectivity, Weed Resistance and Fitness Cost: A Review. *Int. J. Pest Manag.* **2013**, *59* (3), 165–173.

(35) Casida, J. E. Pest Toxicology: The Primary Mechanisms of Pesticide Action. *Chem. Res. Toxicol.* **2010**, *22* (4), 609–619.

(36) Lichtenthaler, H. K. Mode of Action of Herbicides Affecting Acetyl-CoA Carboxylase and Fatty Acid Biosynthesis. *Zeitschrift für Naturforsch. C* **1990**, *45* (5), 521–528.

(37) Howard, J. L.; Ridley, S. M. Acetyl-CoA Carboxylase: A Rapid Novel Assay Procedure Used in Conjunction with the Preparation of Enzyme from Maize Leaves. *Febs Lett.* **1990**, *261* (2), 261–264.
(38) Brown, H. M. Mode of Action, Crop Selectivity, and Soil Relations of the Sulfonylurea Herbicides. *Pestic. Sci.* **1990**, *29* (3), 263–281.
(39) Duke, S. O. Overview of Herbicide Mechanisms of Action. *Environ. Heal. Persp.* **1990**, *87*, 263–271.
(40) Camper, D. L.; Parker, M.H. A Structure-Based Design Approach to Plant Selective 4-Hydroxyphenylpyruvate Dioxygenase Inhibitors. In *Synthesis and Chemistry of Agrochemicals VIIA*; 2007; Vol. 948, pp 105–117.
(41) López-Ramos, M.; Perruccio, F. HPPD: Ligand- and Target-Based Virtual Screening on a Herbicide Target. *J. Chem. Inf. Model.* **2010**, *50*, 801–814.
(42) Pallett, K. E.; Little, J. P.; Sheekey, M.; Veerasekaran, P. The Mode of Action of Isoxaflutole: Physiological Effects, Metabolism, and Selectivity. *Pest. Biochem. Physiol.* **1998**, *62*, 113–124.
(43) Theodoridis, G.; Bahr, J. T.; Hotzman, F. W.; Sehgel, S.; Suarez, D. P. New Generation of Protox-Inhibiting Herbicides. In *Crop Prot.*; 2000; Vol. 19, pp 533–535.
(44) Matringe, M.; Camadro, J.; Labbet, P.; Scalla, R. Protoporphyrinogen Oxidase as a Molecular Target for Diphenyl Ether Herbicides. *Biochem. J.* **1989**, *260*, 231–235.
(45) Ohki, S.; Miller-Sulger, R.; Wakabayashi, K.; Pfleiderer, W.; Böger, P. Phytoene Desaturase Inhibition by O-(2-Phenoxy)ethyl-N-Aralkylcarbamates. *J. Agric. Food Chem.* **2003**, *51* (10), 3049–3055.
(46) Michel, A.; Arias, R. S.; Scheffler, B. E.; Duke, S. O.; Netherland, M.; Dayan, F. E. Somatic Mutation-Mediated Evolution of Herbicide Resistance in the Nonindigenous Invasive Plant Hydrilla (Hydrilla Verticillata). *Mol. Ecol.* **2004**, *13*, 3229–3237.
(47) Kramer, C.; Kalliokoski, T.; Gedeck, P.; Vulpetti, A. The Experimental Uncertainty of Heterogeneous Public K I Data. *J. Med. Chem.* **2012**, *55* (11), 5165–5173.
(48) Kurczab, R.; Smusz, S.; Bojarski, A. J. The Influence of Negative Training Set Size on Machine Learning-Based Virtual Screening. *J. Cheminf.* **2014**, *6* (1), 1–9.
(49) Smusz, S.; Kurczab, R.; Bojarski, A. J. The Influence of the Inactives Subset Generation on the Performance of Machine Learning Methods. *J. Cheminf.* **2013**, *5* (1), 17.
(50) Gaulton, A.; Kale, N.; van Westen, G. J. P.; Bellis, L. J.; Bento, A. P.; Davies, M.; Hersey, A.; Papadatos, G.; Forster, M.; Wege, P.; et al. A Large-Scale Crop Protection Bioassay Data Set. *Sci. data* **2015**, *2*, 150032.
(51) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1708–1718.
(52) Baumann, K. Cross-Validation as the Objective Function for Variable-Selection Techniques. *Trends Anal. Chem.* **2003**, *22* (6), 395–406.
(53) Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. Assessing the Accuracy of Prediction Algorithms for Classification: An Overview. *Bioinformatics* **2000**, *16* (5), 412–424.
(54) Freeman, E. a.; Moisen, G. G. A Comparison of the Performance of Threshold Criteria for Binary Classification in Terms of Predicted Prevalence and Kappa. *Ecol. Model.* **2008**, *217* (1-2), 48–58.
(55) Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*, Second.; Springer Series in Statistics; Springer New York: New York, NY, 1985.
(56) Baldi, Pierre and Nasr, R. When Is Chemical Similarity Significant? The Statistical

Distribution of Chemical Similarity Scores and Its Extreme Values. *J. Chem. Inf. Model.* **2011**, *50* (7), 1205–1222.

(57) Standard Score https://statistics.laerd.com/statistical-guides/standard-score.php (accessed Jun 20, 2015).

(58) Sahigara, F.; Ballabio, D.; Todeschini, R.; Consonni, V. Defining a Novel K-Nearest Neighbours Approach to Assess the Applicability Domain of a QSAR Model for Reliable Predictions. *J.Cheminf.* **2013**, *5* (27), 1–9.

(59) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of Fingerprint-Based Methods for Virtual Screening Using Multiple Bioactive Reference Structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 1177–1185.

(60) Heikamp, Kathrin & Bajorath, J.; Heikamp, K.; Bajorath, J. Comparison of Confirmed Inactive and Randomly Selected Compounds as Negative Training Examples in Support Vector Machine-Based Virtual Screening. *J. Chem. Inf. Model.* **2013**, *53* (7), 1595–1601.

Figures

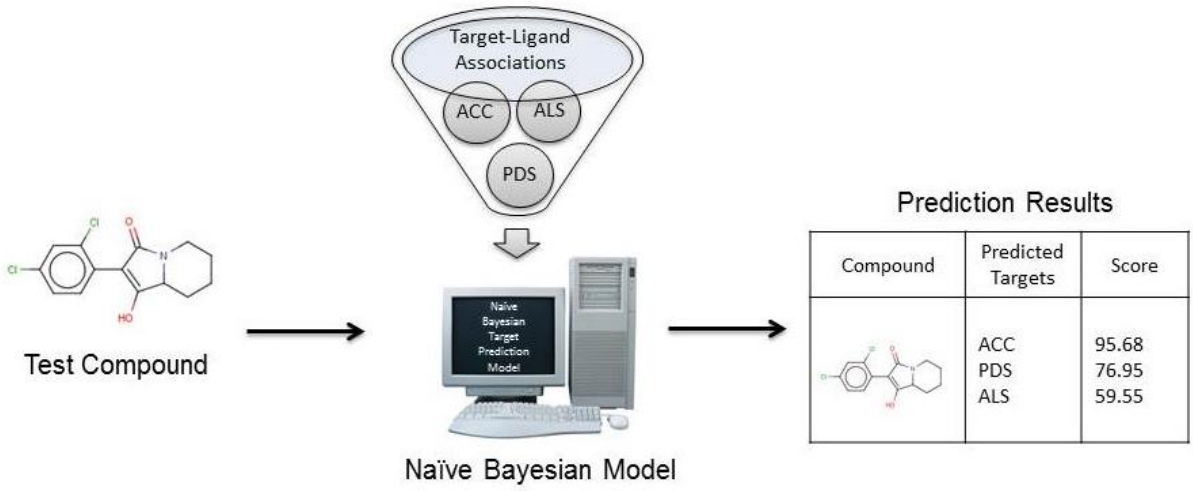


Figure 1. Schematic representation of an *in silico* target prediction workflow. For a computational model to predict protein targets of small molecules, firstly target class models for up to 16 proteins targets with 48,000 compounds are constructed with application of the Naïve Bayesian Classifier. A chemical structure input (shown on the left) in the model (shown in the center) can then be annotated with its likely targets based on the probability classification score (shown on the right) generated by the model. These predictions can then be ranked based on the z-scores.

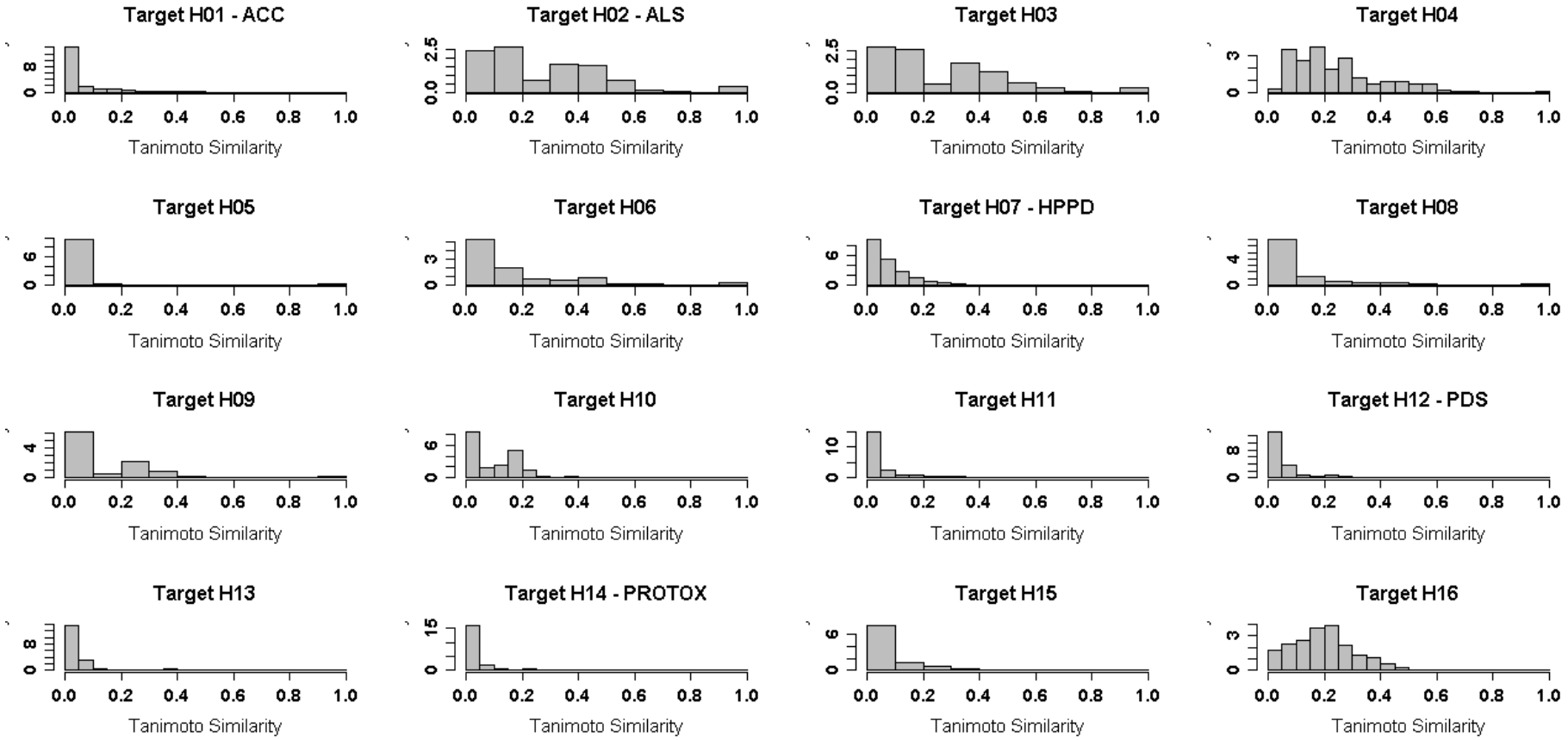


Figure 2. Distribution plots of pairwise compound similarity among the actives per target class for all the herbicide target classes. On the x-axis is the similarity measure (Tanimoto coefficient) and on the y-axis is the density of the occurrences in the respective datasets. The compounds in majority of the classes are seen to be diverse with a mean Tanimoto coefficient of less than 0.3. The diversity of datasets contributes in covering a wider chemical space but at the same time can pose challenges in the predictive ability of the prediction models.

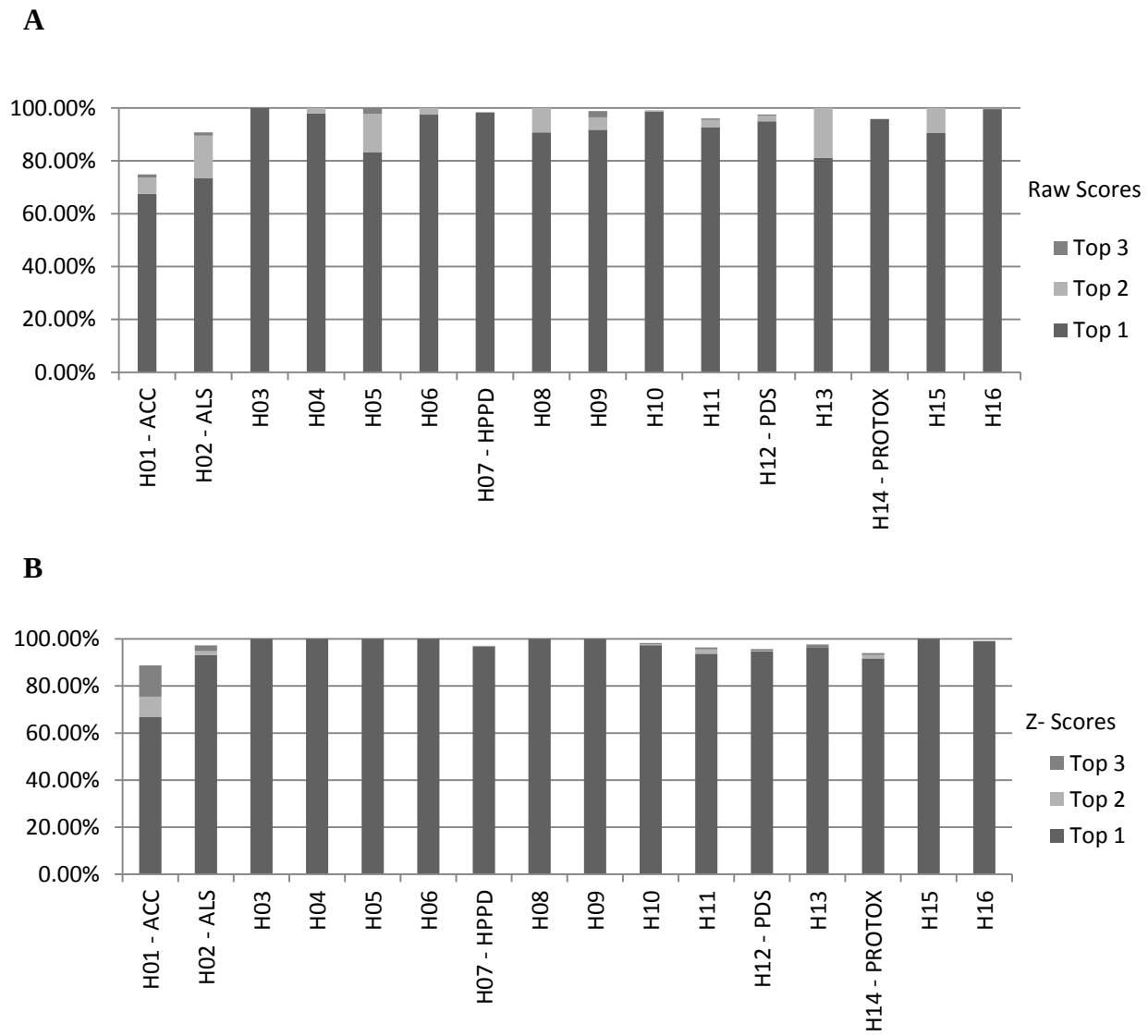


Figure 3. Recall analysis assessing predictive ability of the herbicide models based on raw scores (A) and z-score (B) in top 1, top 2 and top 3 positions. From the plot it is seen that recall based on Z- scores generated higher prediction results than raw scores and consistently yielded more than 80% correct predictions for all the target models in top 1-3 positions.

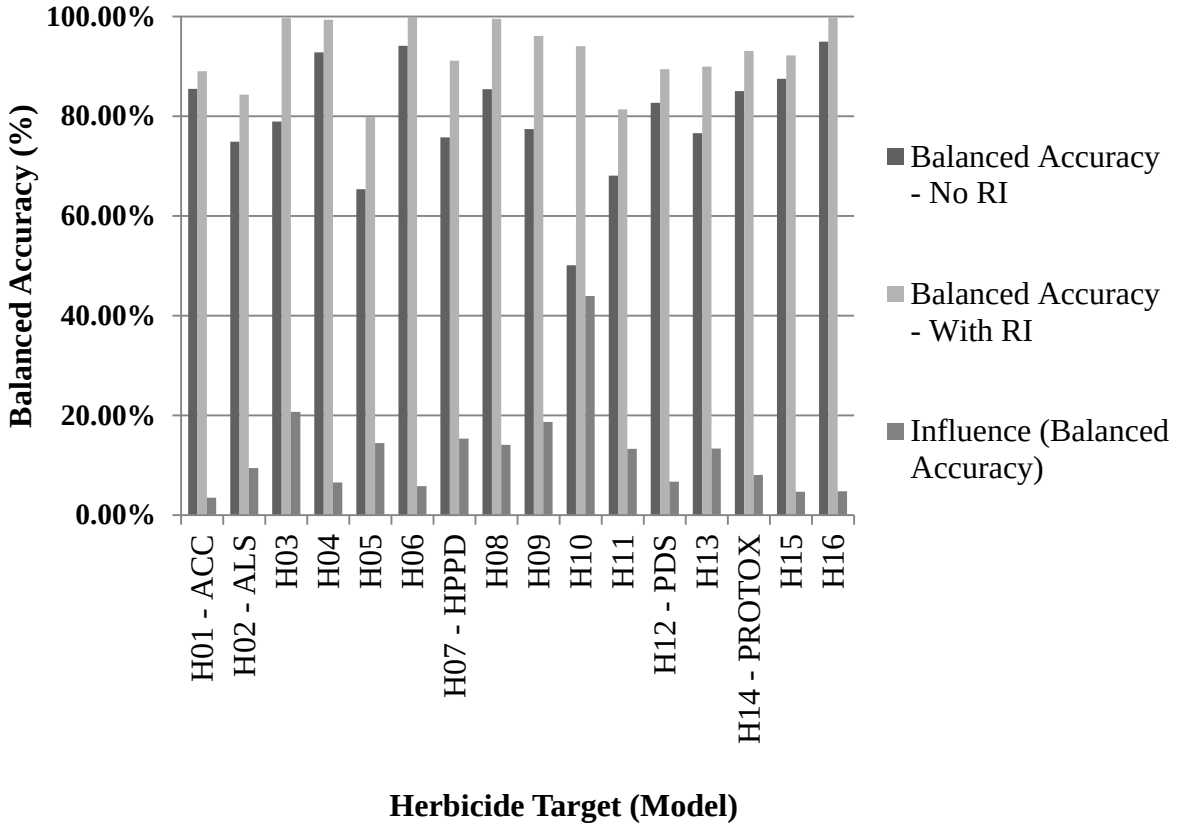
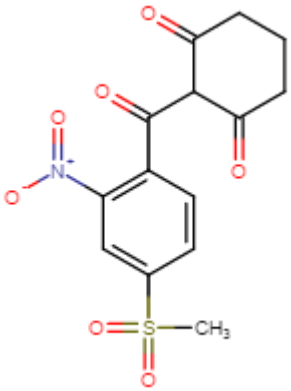
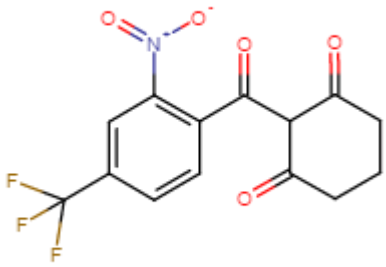


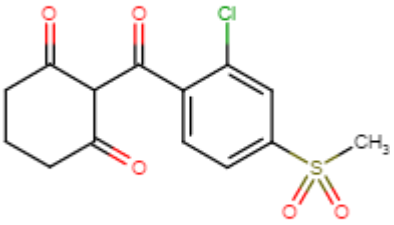
Figure 4. Balanced accuracy achieved by performing 5-fold cross validation for all the herbicide models built with and without random inactives (RI) and the percent influence of adding RI. It is observed that the models with inclusion of RI performed nearly 3% - 40% better than the model with no RI, which is in agreement with previous studies.



Mesotrione



Nitisinone



Sulcotrione

Figure 5. Examples of a class of triketone compounds in the external testset, which are marketed herbicides and are known to specifically inhibit HPPD. The average Tc similarity of the 5 nearest neighbours (5-NN) in the active set were 0.034, 0.033 and 0.035 for nisinone, mesotrione and sulcotrione respectively, and despite those low similarities they were correctly identified as inhibitors of HPPD by HerbiMod.

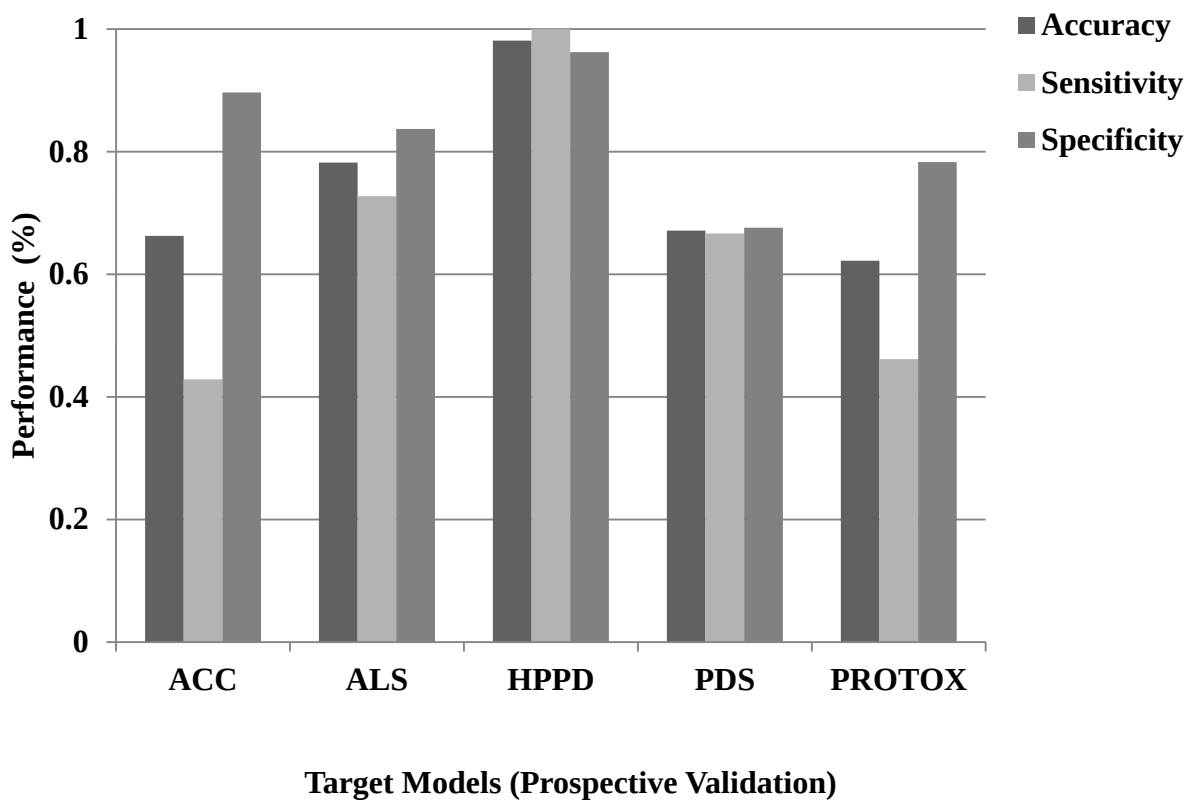


Figure 6: Results of the prospective validation test set for 394 compounds tested on the five well studied herbicide targets. Overall, balanced accuracy of more than 60% is seen to be achieved for all the target classes. These results illustrate the predictive ability of HerbiMod for elucidating the mode of action of new compounds which, importantly, fall into the applicability domain of the model.

Table 1: The number of active and inactive compounds tested against 16 herbicide target classes and the total number of compounds including the random inactives used for training the NB models. (Note that the number of datapoints listed here does not necessarily represent all data available within BASF, but rather the subset used for training this particular type of model.) 'H' in the model-column represents 'herbicide' target. Five well studied herbicide targets, ACC, ALS, HPPD, PDS and PROTOX are named in the table. It can be seen that target classes are highly imbalance where some classes contain more actives than inactives and vice a versa.

Target ID	Actives	Inactives	Actual tested compounds	Total training compounds (including 10,000 random inactives)
H01 – ACC	310	3,054	3,364	13,364
H02 – ALS	1,454	4,129	5,583	15,583
H03	34	477	511	10,511
H04	277	636	913	10,913
H05	94	857	951	10,951
H06	41	87	128	10,128
H07 - HPPD	4,161	1,884	6,045	16,045
H08	40	638	678	10,678
H09	87	780	867	10,867
H10	517	432	949	10,949
H11	1,192	3,125	4,317	14,317
H12 – PDS	737	1,339	2,076	12,076
H13	88	601	689	10,689
H14 - PROTOX	4,741	2,139	6,880	16,880

H15	76	849	925	10,925
H16	922	1,489	2,411	12,411

Table 2. Summary of model performance statistics for 16 herbicide target models at specific NB score cut-off. The cut-offs are chosen based on highest balanced accuracy for each model.

Model	Cut-off	MCC	Sensitivity	Specificity	Balanced accuracy
H01 – ACC	0	0.77	0.89	0.89	0.89
H02 - ALS	10	0.69	0.89	0.80	0.84
H03	10	0.99	1.00	0.99	1.00
H04	50	0.99	1.00	0.99	0.99
H05	0	0.60	0.94	0.65	0.80
H06	30	1.00	1.00	1.00	1.00
H07- HPPD	0	0.82	0.97	0.86	0.91
H08	10	0.99	1.00	0.99	1.00
H09	10	0.92	1.00	0.92	0.96
H10	0	0.88	0.99	0.89	0.94
H11	20	0.63	0.87	0.76	0.81
H12 – PDS	0	0.79	0.98	0.81	0.89
H13	10	0.80	0.82	0.98	0.90
H14 – PROTOX	220	0.86	0.93	0.94	0.93
H15	10	0.84	0.87	0.98	0.92
H16	360	1.00	1.00	1.00	1.00

Table 3. For each target model, a maximum allowed distance was set at the defined classification error of 0.2. The maximum allowed distance was observed to be 0.4 (in case of H07 - HPPD).

Herbicide Model	Maximum allowed distance (Tc)
H01 – ACC	0.27
H02 – ALS	0.24
H03	0.18
H04	0.19
H05	0.21
H06	0.22
H07 – HPPD	0.41
H08	0.1
H09	0.12
H10	0.23
H11	0.09
H12 – PDS	0.1
H13	0.1
H14 – PROTOX	0.42
H15	0.09
H16	0.39

Supplementary Material

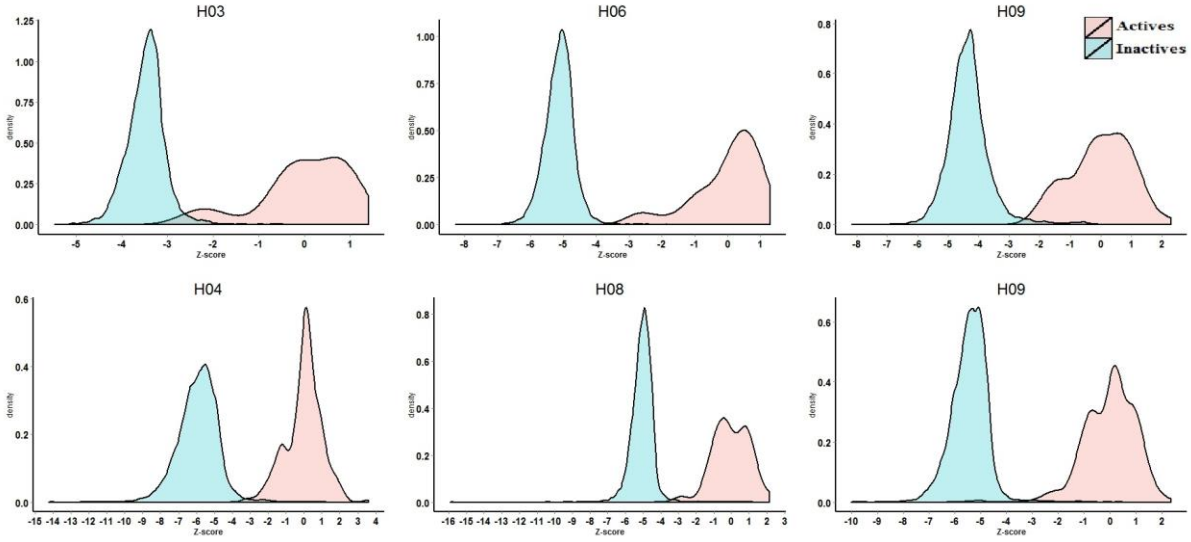


Figure S1. Z-score distribution of active and inactive compounds in the training sets of herbicide target classes H03, H04, H06, H08 and H16 showing relatively clear separation among the actives and inactives. It is seen that target H03 serves a case overlapping region which indicates false positives and/or false negatives.

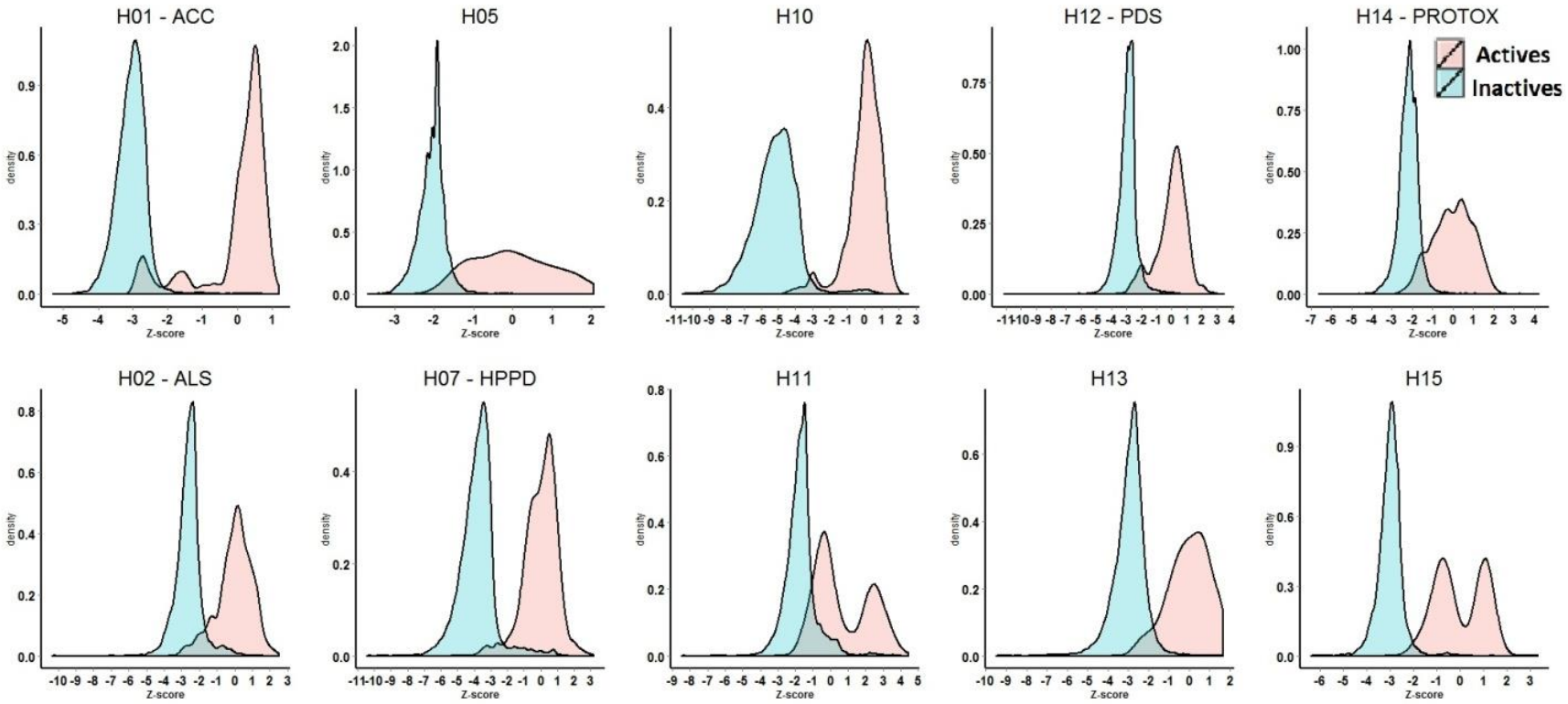
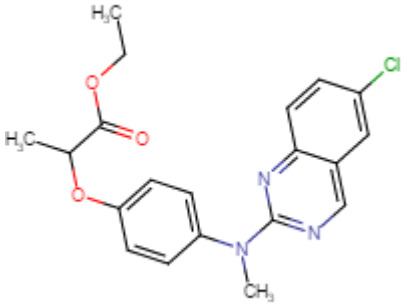
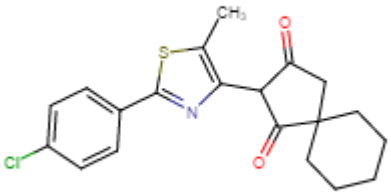
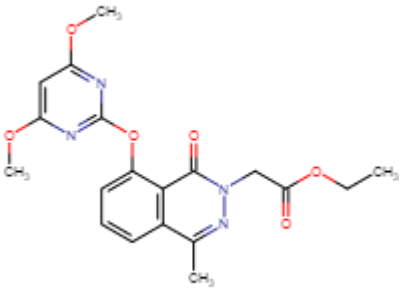


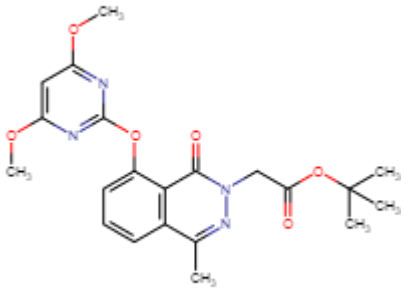
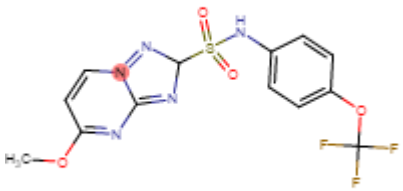
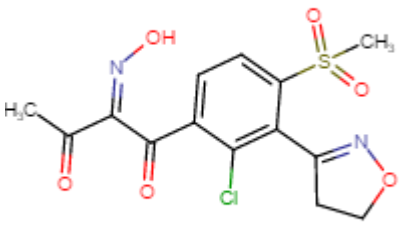
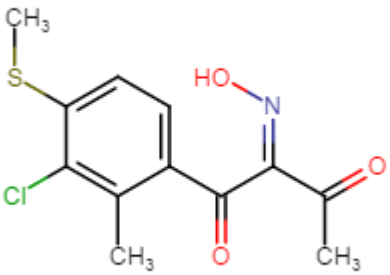
Figure S2. Z-score distribution plots of active and inactive compounds in the training sets of 10 herbicide target classes that show overlapping distributions among the actives and inactives suggesting that area to constitute false positive and false negative predictions.

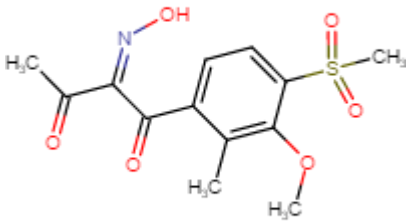
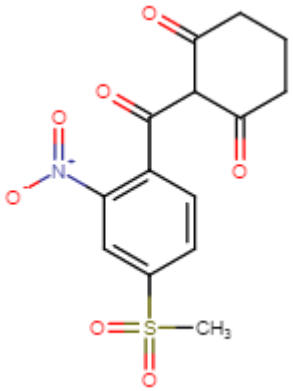
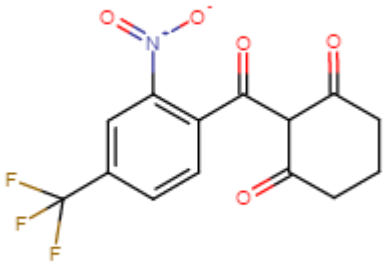
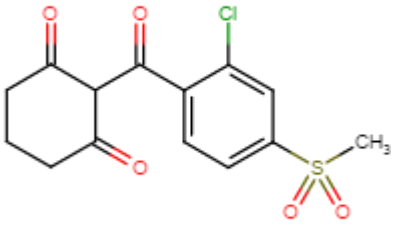
Table S1. Influence of adding random inactives on the model performance of each of the individual herbicide target class.

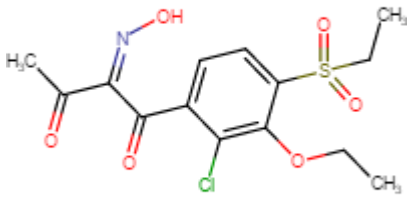
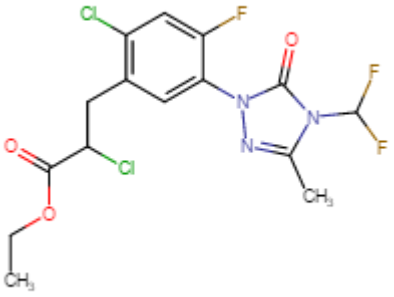
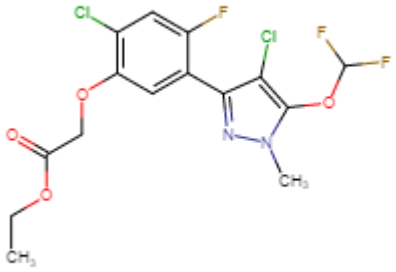
Target ID	Balanced Accuracy - No Random Inactives	Balanced Accuracy - With Random Inactives	Influence (Balanced Accuracy)
H01 - ACC	0.85492	0.89008	3.52%
H02 - ALS	0.748986	0.843457	9.45%
H03	0.789474	0.996659	20.72%
H04	0.927917	0.99365	6.57%
H05	0.653509	0.798109	14.46%
H06	0.941176	0.999256	5.81%
H07 - HPPD	0.757697	0.911414	15.37%
H08	0.854331	0.995299	14.10%
H09	0.774194	0.961253	18.71%
H10	0.501214	0.940683	43.95%
H11	0.680908	0.813871	13.30%
H12 - PDS	0.826913	0.894295	6.74%
H13	0.765931	0.899495	13.36%
H14 - PROTOX	0.850276	0.930826	8.05%
H15	0.874951	0.922038	4.71%
H16	0.949495	0.997606	4.81%

Table S2. List of 12 compounds from the external dataset extracted from ChEMBL database associated to the five herbicide targets with their structures and associated targets. These compounds lie in the applicability domain of our models and for which predictions were generated.

Herbicide target	ChEMBL ID	Chemical structure
ACC	CHEMBL38166	
	CHEMBL2271423	
ALS	CHEMBL2253256	

	CHEMBL2253235	
	CHEMBL2251940	
HPPD	CHEMBL2251435	
	CHEMBL2251434	

	CHEMBL2251433	
	CHEMBL1873440	
	CHEMBL1337	
	CHEMBL2252422	

	CHEMBL2251432	
PROTOX	CHEMBL2145069	
	CHEMBL1905010	
PDS	CHEMBL1863061