

Exploring Insecticidal Molecules with Random Forest: Toward High Insecticidal Activity and Low Bee Toxicity

Wei Guo, Xiangmin Song, Yongchao Gao, Shuai Yang, Jiahong Tang, Chen Zhao, Haojing Wang, Jiajun Ren,* Lingda Zeng,* and Hanhong Xu*



Cite This: *J. Agric. Food Chem.* 2025, 73, 5573–5584



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information

ABSTRACT: Insecticidal molecules with high activity are crucial for global pesticide reduction and food security. However, their usage is limited by their concomitant high toxicity to bees. Balancing insecticidal activity and bee toxicity remains a critical challenge in the exploitation of new insecticidal molecules. In this study, we propose a novel strategy for exploiting molecules that are both highly effective against pests and minimally harmful to bees. A series of molecules were synthesized and tested to train a machine learning (ML) model for predicting insecticidal activity against pests. Meanwhile, another ML model was trained by using publicly available data to predict bee toxicity. The models demonstrated good performance, with mean AUC values of 0.88 ± 0.05 for insecticidal activity and 0.91 ± 0.01 for bee toxicity. By integrating these two models, we successfully predicted and experimentally validated a molecule that exhibited a high insecticidal activity and low bee toxicity. This dual-ML-model approach offers a promising pathway for the development of insecticidal molecules that are both effective and environmentally safe, thereby contributing to sustainable agricultures.

KEYWORDS: insecticidal molecule exploitation, machine learning, insecticidal activity, bee toxicity

1. INTRODUCTION

Insecticides play an important role in global food security.^{1–3} According to the food and agriculture organization of the United Nations, the global consumption of insecticides reached 757,002.90 tons in 2021.^{4,5} These insecticides have greatly protected crops, enhanced farmers' produce, and increased crop yields.^{6,7} However, with rising concerns over food safety and policy shifts toward reducing pesticide usage, there is a growing emphasis on discovering highly effective insecticidal molecules.^{8,9} A major challenge with these molecules is their toxicity to beneficial insects such as bees.¹⁰ Molecules with high insecticidal activity often exhibit high toxicity to bees, limiting their practical application. For instance, any pesticide with bee toxicity exceeding 2 $\mu\text{g}/\text{bee}$ is prohibited from being registered in China. The traditional "trial and error" approach to discovering new insecticides is time-consuming and expensive. Thus, it is crucial to develop new strategies based on rational design for the rapid and efficient discovery of highly effective insecticidal molecules with low bee toxicity.

In recent years, machine learning (ML) has emerged as a powerful tool for uncovering complex patterns and relationships within data, facilitating accurate predictions and informed decisions. ML has been extensively applied in fields such as MRI image synthesis, drug discovery, disease diagnosis, biological image analysis, and materials science.^{11–25} Despite their great potential in the discovery of insecticidal molecules, ML applications in this area are still nascent.

In this study, we proposed a dual-ML-model strategy, where one model predicts the insecticidal activity of molecules and the other predicts their bee toxicity. For the insecticidal activity, given the absence of a standardized public database and the

challenges in constructing a large, diverse data set from scratch, we developed a specialized database of 181 molecules. This database was created by synthesizing and testing structural analogues of fluralaner, an isoxazoline molecule known for its high insecticidal activity but also high bee toxicity.^{26–30} A specialized database allows for the training of accurate ML models even with a small data set, and the predictions are likely to identify molecules with high insecticidal activity similar to fluralaner.^{20,31} For bee toxicity, since there are standard international guidelines for determining bee toxicity, there are many publicly available high-quality databases.^{32–34} Therefore, data from these databases were used to train an ML model for predicting molecular toxicity.

We employed random forest algorithms for training the two ML models due to their robustness, interpretability, scalability, and ability to mitigate overfitting.^{35,36} These models were used to predict the insecticidal activity and bee toxicity of 10 newly designed molecules, which were then validated experimentally. The accuracy of the ML model for the 10 molecules was 70%, demonstrating the effectiveness of our strategy. Among the 10 molecules, one molecule that exhibited high insecticidal activity and low bee toxicity was further investigated in detail as a potential candidate for pest control. Our work presents a rapid and efficient strategy for discovering highly effective insecticidal

Received: September 12, 2024

Revised: February 6, 2025

Accepted: February 7, 2025

Published: February 20, 2025



molecules with minimal bee toxicity, contributing to the advancement of sustainable agriculture.

2. MATERIALS AND METHODS

2.1. Data Preparation. A total of 181 structural analogues of fluralaner were synthesized (Supplementary Data 1) and tested for their insecticidal activity against *Plutella xylostella* (*P. xylostella*), forming the *P. xylostella* activity data set. A 100% insecticide rate at a concentration of 25 mg/L was used as the criterion for the dichotomous classification. Molecules causing less than 100% mortality are classified as inactive and otherwise as active.

In the realm of predicting molecular activity through ML models, a pivotal stage involves translating the structural attributes of molecules into numerical representations, commonly referred to as molecular descriptors.^{37,38} In this study, a total of 208 molecular descriptors were calculated in RDKit v2023.3.2 for each molecule.³⁹ To mitigate the risks of overfitting, careful preprocessing of these molecular descriptors was necessary.^{40,41} We adopted a concurrent approach of feature selection and model training using the recursive feature elimination (RFE) feature selection algorithm. Specifically, the RFE feature selection algorithm from sklearn was utilized to systematically remove less significant descriptors while training the model.⁴² This strategic fusion of RFE contributed to the optimization of the model's predictive capability.

Ultimately, 14 key molecular descriptors with the highest correlation were identified to train the *P. xylostella* activity prediction model (PAP model), which was key to establishing a mapping relationship between structure and activity. Their importance to the PAP model was calculated using the `feature_importances_` function of the random forest algorithm (Figure S1).

Bee toxicity data were collected from different public databases: the Pesticide Properties DataBase (PPDB) created by the Agriculture & Environment Research Unit (AERU) at the University of Hertfordshire, the Terrestrial US-EPA ECOTOX database present in the OECD QSAR Toolbox v4.6 (www.qsartoolbox.org), and the Pesticide Action Network Pesticide Database (PANPD) by Pesticide Action Network North America (<http://www.pesticideinfo.org>). A total of 502 insecticide molecules were collected to form the final bee toxicity data set (Supplementary Data 2). According to the Chinese bee toxicity classification criteria, insecticides with an $LD_{50} \leq 2 \mu\text{g}/\text{bee}$ are classified as highly toxic. We used three times this criterion ($LD_{50} \leq 6 \mu\text{g}/\text{bee}$) as the classification standard for the honeybee toxicity data set. Fourteen key molecular descriptors were identified to train the bee toxicity prediction (BTP model), and their importance to the BTP model was calculated (Figure S2).

2.2. ML Model Development and Validation. After the preprocessing of descriptors was completed, the focus shifted to model training. The PAP model and BTP model were constructed using the random forest algorithm. Internal validation plays a crucial role in the development of random forest models. Statistical parameters, including the area under the curve (AUC), serve as yardsticks for evaluating performance. Internal validation was conducted through 5-fold cross-validation to assess the internal predictive capability of the developed model.

The *P. xylostella* activity data set and bee toxicity data set were randomly divided into training set and validation set using 5-fold cross-validation in a 4:1 ratio, where four subsets were used to train the model and the remaining subset was used for the cross-validation process. Traversing the number of descriptors, number of estimators, maximum depth, minimum sample split, and maximum features through a grid searching method obtained a group of parameters that achieved the best AUC values for the PAP model and the BTP model. The optimized parameters for the random forest algorithm were 110 estimators, a maximum depth of 20, a minimum sample leaf of 1, and a minimum sample split of 2 for the PAP model, and 110 estimators, a maximum depth of 8, a minimum sample leaf of 3, and a minimum sample split of 2 for the BTP model.

Beyond internal validation, external validation is crucial for ascertaining a model's generalization capacity and authentic predictive

potential. An external test set of 37 molecules from the *P. xylostella* activity data set and 101 molecules from the bee toxicity data set was used to evaluate the generalization ability of the PAP model and BTP model.

2.3. Prediction and Synthesis of New Molecules in the Dual-ML Model. Ten new molecules were designed, and their activity and bee toxicity were first predicted using the dual-ML model we constructed. Molecules classified as positive by the PAP model and negative by the BTP model were synthesized and validated by bioassays.

In this study, all 10 molecules were synthesized. Details of synthesis procedures, reagents, and instrumentation are presented in the Supporting Information.

2.4. Bioassay of New Molecules on *P. xylostella*. For the purpose of conducting the activity assays, all *P. xylostella* was kindly supplied by National Key Laboratory of Green Pesticide/Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education (South China Agricultural University, Guangzhou, Guangdong, China). The *P. xylostella* was cultivated on an artificial diet under controlled environmental conditions. This encompassed maintaining a temperature of $25 \pm 1^\circ\text{C}$, a relative humidity of $65 \pm 5\%$, and a photoperiod cycle of 16:8 h (light:dark).

All experiments were conducted in small plastic Petri dishes (3 cm diameter, 5 cm height) containing round filter paper and a strip of paper towel. With 10 third-instar larvae of *P. xylostella* per plate, the plates were placed in an incubator at $25 \pm 1^\circ\text{C}$ at 65% relative humidity, with an L:D photoperiod of 16:8 h.

The assessment of molecular activity was executed through meticulous procedures. SAF-A, SAF-E, and SAF-G from the test set were solubilized in DMSO and subsequently diluted with distilled water containing Tween-80 (0.1% v/v) to yield a solution of 25 mg/L and 5 mg/L. In the initial stages of each experiment, leaf discs (1.4 cm in diameter) were excised from greenhouse-grown fresh cabbages. These discs were immersed in the test solution for a duration of 30 s, after which they were left to air-dry naturally.

The experimental process involved supplying three leaf discs treated with the respective molecule to third-instar larvae. After 24 h, fresh leaves were provided as required. The negative control comprised treatment with a solution containing 0.1% (v/v) Tween-80 and 0.01% (v/v) DMSO dissolved in distilled water. Each experiment was meticulously replicated three times to ensure reliability. The mortalities (%) were conducted 48 h post-treatment, constituting the core outcome measure for the activity assessment.

Based on the above process, the insecticidal activities of the molecules were determined and classified as positive by the PAP model and negative by the BTP model and then their bee toxicity was verified in turn.

2.5. Bioassay of New Molecules on Bees (*Apis cerana*).

Bioassays were conducted following the method outlined in the *Test guidelines on environmental safety assessment for chemical pesticides-Part 10: Honeybee acute toxicity test*, with certain modifications. To elaborate, the experimental procedure is presented as an example of the highest ranked SAF-A. Ten bees were placed in a wooden beehive with dimensions of 20 cm on each side and starved for 2 h before the start of the test. Subsequently, the bees were granted free access to a sucrose solution (50% w/v) containing SAF-A for a duration of 4 h. The bees were provided with a sucrose solution devoid of SAF-A for the remaining duration of the experiment. For the control group, bees were exclusively administered the sucrose solution. A concentration of 6 $\mu\text{g}/\text{bee}$ was used to verify bee toxicity, and the experiment was performed in three replicates. Beehives were maintained at $25 \pm 1^\circ\text{C}$ and $75 \pm 5\%$ RH in the dark except during observations. The mortalities of honeybees were recorded at 48 h after treatment.

2.6. Careful Study of the Properties of SAF-A. In order to comprehensively study the efficacy of SAF-A in controlling the binge-feeding stage of the cabbage moth, feeding bioassays were conducted on third-instar larvae and fourth-instar larvae specimens. Each developmental stage was subjected to four replicates, where one replicate comprised either 10 third-instar larvae or 10 fourth-instar larvae individuals within a single Petri dish.

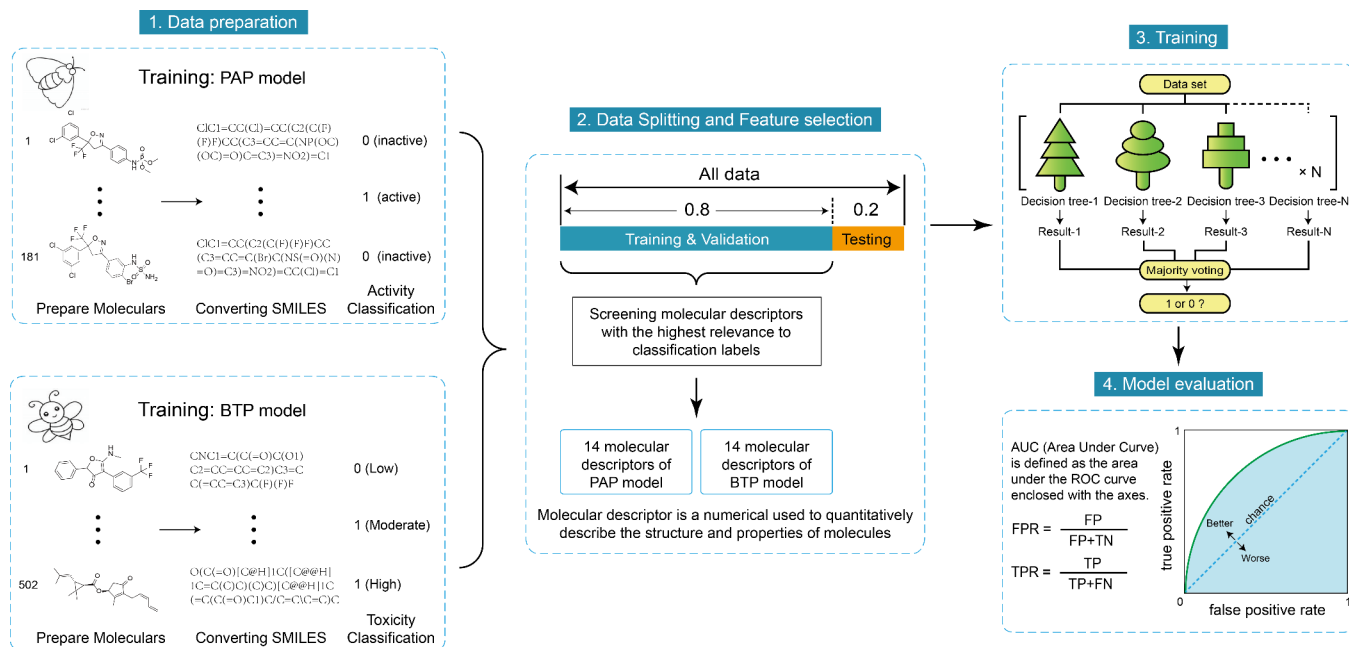


Figure 1. Schematic of construction process of the PAP model and the BTP model. The molecules were converted to SMILES strings and classification by *P. xylostella* activities or bee toxicities. The molecules were then randomly split, and the key molecular descriptors were identified using the RDKit Python package. The models were built with the random forest-based machine learning method. AUC values were evaluated to obtain the best model by optimizing the parameters.

For the larval bioassays involving **SAF-A** treatments, four distinct concentrations were employed: 0.5, 2.5, 5, and 25 mg/L. The activity of **SAF-A** against *P. xylostella* was determined using the leaf dipping method. At the initiation of each experiment, the third-instar larvae were supplied with three leaf discs, while the fourth-instar larvae were offered four. Leaf discs were replaced every 24 h, and only the first 24 h leaf discs were loaded with the drug. As a reference, a control group was treated with distilled water containing 0.1% (v/v) Tween-80 and 0.01% (v/v) DMSO.

To investigate the effect of **SAF-A** on the feeding inhibition of *P. xylostella*, leaf discs were subjected to treatment with solutions of either 0.5 or 2.5 mg/L of **SAF-A**, following the procedure outlined earlier. One treated leaf disc, within a single Petri dish and per replicate, was exposed to being eaten by one third-instar larva. Periodic observations were made, and fresh leaf discs were introduced 24 h after the start of the experiment. Photographic documentation was carried out at 12, 24, 36, and 48 h. It is worth mentioning that 10 replications were conducted for each treatment group.

A specific focal point was the investigation of the effect of **SAF-A** on the development of third-instar larvae. This assessment was conducted at a concentration of 0.5 and 2.5 mg/L. Two control groups were set up: one control group was fed normally, and the other was a fasting group. Photographic documentation of the developmental stages of surviving larvae for each treatment was undertaken 48 h after the commencement of feeding. Notably, each treatment consisted of three replicates of 10 third-instar larvae each.

The adult bioassay experiments necessitated a modified approach. A small aperture, approximately 5 mm in diameter, was introduced into the lid of the Petri dish. This opening was plugged with a cotton material. **SAF-A** was dissolved in DMSO and subsequently diluted using a 10% (v/v) honey solution, resulting in concentrations of 5, 50, 100, 200, and 400 mg/L. Five nymphs were placed in each Petri dish, and each concentration was repeated 18 times. The commencement of the experiment coincided after the emergence of adult eclosion. The number of fledged nymphs in each Petri dish was counted, and the unfledged nymphs were removed from the Petri dish. Eventually, ck (without **SAF-A**), 5, 50, 100, 200, and 400 mg/L had 73, 54, 46, 48, 50, and 53 adults for the bioassay, respectively. A pipet gun was employed to administer 300 μ L of the prepared honey solution containing the

SAF-A onto the cotton, ensuring the cotton remained moist without excess liquid dripping off. Once again, the control group was treated with a honey solution containing 0.1% (v/v) Tween-80 and 0.01% (v/v) DMSO. Mortality was observed and counted at 48 h.

2.7. Insecticidal Activity—Bee Toxicity Window of SAF-A. Using the methods described above, the activities of fluralaner and **SAF-A** were determined against *P. xylostella* and their toxicity assay to bees, respectively. For activity bioassays to *P. xylostella*, eight concentrations of **SAF-A**, 0.25, 0.5, 1, 2, 3, 4, 5, and 25 mg/L, and five concentrations of fluralaner, 0.005, 0.01, 0.02, 0.03, and 0.04 mg/L, were used for the treatments, respectively. For bioassays of bee toxicity, four concentrations of **SAF-A**, 50, 100, 150, and 300 mg/L, and five concentrations of fluralaner, 0.00625, 0.0125, 0.025, 0.05, and 0.1 mg/L, were used for the treatments, respectively. Three replicates per treatment and 10 experimental subjects per replicate were used.

2.8. Homology Modeling and Molecular Docking Studies.

The 3D structure of the γ -aminobutyric acid receptor (GABAR) of *P. xylostella* was constructed by homology modeling using an online protein structure prediction server (SWISS-MODEL). Molecular docking was finished by YASARA version 16.7.22, and the details are described in the [Supporting Information](#).

3. RESULTS

3.1. Dual-ML Model. The workflow of our dual-ML-model strategy is illustrated in Figure 1. We synthesized 181 structural analogues of fluralaner (Supplementary Data 1) and assessed their insecticidal activities against a harmful pest, *P. xylostella*.^{43,44} The dichotomous classification criteria of these molecules were set as a 100% insecticide rate at a concentration of 25 mg/L to identify those with high activity. Among the total 181 molecules, 144 were randomly selected to form a training set to develop the PAP model. Using the RFE feature selection algorithm, we identified 14 key molecular descriptors most correlated with insecticidal activities (Figure S1).⁴⁵ The PAP model was carefully fine-tuned to ensure a balance between model complexity and prediction accuracy, achieving a mean AUC value of 0.88 ± 0.05 (Figure 2a). The high AUC value

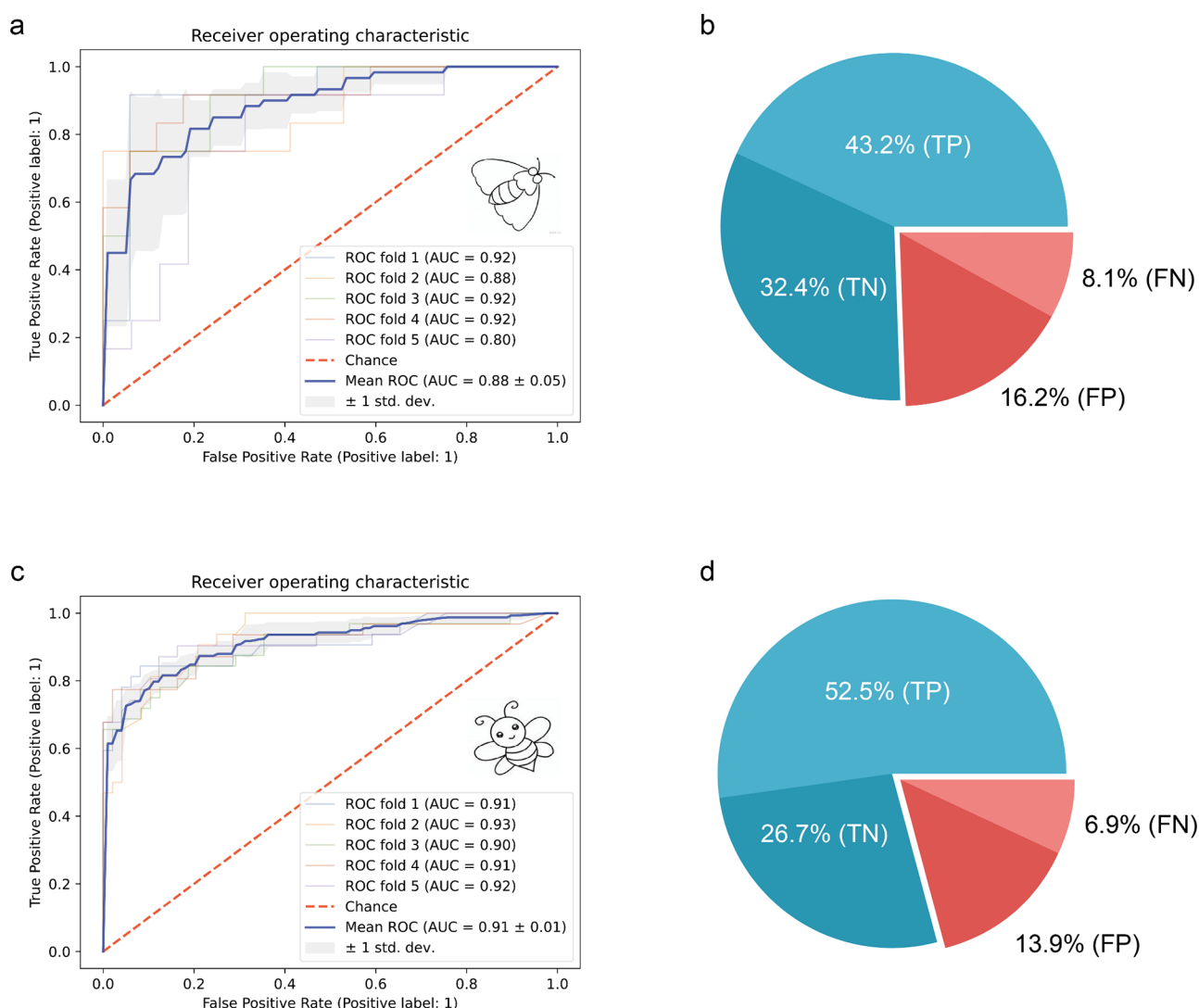


Figure 2. Evaluation and validation of the PAP model and the BTP model. **a** The ROC curves of the PAP model (5-fold cross-validation). **b** External data set validation of the PAP model (another 37 molecules in the 181 molecules). TP (true positive), TN (true negative), FP (false positives), and FN (false negative). **c** The ROC curves of the BTP model (5-fold cross-validation). **d** External data set validation of the BTP model (another 101 molecules in the 502 molecules).

suggests the ability to effectively distinguish between positive (active) and negative (inactive) molecules, while the small error suggests robustness and consistency of the PAP model.

An external test set of 37 molecules was used to evaluate the generalization ability of the PAP model. As shown in Figure 2b, the model correctly predicted 16 molecules as active (true positive, TP) and 12 as inactive (true negative, TN), with an overall accuracy of 75.7%, indicating that the PAP model possesses a high generalization ability. Although only 144 molecules were used to train the PAP model, the data's high quality enabled the creation of a robust model. Another reason for obtaining such high-quality predictive models is attributed to the homogeneity of the data set, consisting of a series of structural analogues. In contrast to data sets containing diverse molecular skeletons, data sets consisting of structural analogues reduce the data size needed for model construction. This result indicates that building a data set from a highly active structure can expedite the development of effective ML models, facilitating the discovery of highly active insecticidal molecules.

Similarly, the BTP model was constructed using a data set of 401 molecules with experimental toxicity data from public

databases (Supplementary Data 2). According to the Chinese guidelines for bee toxicity, insecticides with $LD_{50} \leq 2 \mu\text{g}/\text{bee}$ are prohibited. Here, we used three times this concentration as our classification criterion (that is, insecticides with $LD_{50} \leq 6 \mu\text{g}/\text{bee}$ are considered to be highly toxic and unacceptable) to identify molecules with low bee toxicity. The RFE algorithm identified 14 key molecular descriptors for training the BTP model (Figure S2). The BTP model achieved a mean AUC value of 0.91 ± 0.01 (Figure 2c), indicating strong predictive performance. The model's generalization ability was further validated using an external test set of 101 molecules, achieving an accuracy of 79.2% with 27 correctly predicted as nontoxic (TN) and 53 as toxic (TP) (Figure 2d).

The high predictive accuracy on the external test set and the high AUC values of both models illustrate the capability to predict highly insecticidally active molecules with low bee toxicity.

3.2. Prediction of Potential Insecticidal Molecules. To demonstrate the effectiveness of the dual-ML model, we designed 10 new structural analogues of fluoroaners (SAFs) shown in Table 1 and predicted their activity and bee toxicity.

Table 1. Structural Formula and Prediction Results of SAFs

Molecule	Structural formula	PAP model prediction ^a	BTP model prediction ^b
SAF-A		1	0
SAF-B		0	0
SAF-C		0	1
SAF-D		0	1
SAF-E		1	0
SAF-F		0	1
SAF-G		1	0
SAF-H		1	1
SAF-I		0	0
SAF-J		0	1

^aThe data set negative (inactive) is 0 and positive (active) is 1. ^bThe data set negative (nontoxic) is 0 and positive (toxic) is 1.

Molecules classified as positive by the PAP model and negative by the BTP model, namely, molecule **SAF-A**, **SAF-E**, and **SAF-G**, were screened out. These three molecules were synthesized according to the synthetic route shown in Figure S3. Their structures were confirmed by ¹H NMR, ¹³C NMR, and MS as shown in Figures S4–S6. Furthermore, the *P. xylostella* activity of **SAF-A**, **SAF-E**, and **SAF-G** was tested experimentally. As shown in Figure 3a, at 25 mg/L, both **SAF-A** and **SAF-G** could lead to 100% mortality of *P. xylostella*. At a lower concentration of 5 mg/L, **SAF-A** showed superior lethality, achieving a 100% mortality rate (Figure 3b). **SAF-A** possesses the highest insecticidal activity among **SAF-A**, **SAF-E**, and **SAF-G**. Finally, the bee toxicity of **SAF-A** was assessed at 6 μg/bee. As shown in Figure 3c, the mortality of bees was 40.8 ± 2.9% < 50%, indicating that the LD₅₀ value of **SAF-A** against bees was more than 6 μg/bee. Therefore, **SAF-A** was confirmed as a molecule with a high insecticidal activity and low bee toxicity.

The overall molecular discovery process is depicted in Figure 3d. The dual-ML-model approach allows for efficient virtual screening, reducing the number of molecules that need to be experimentally tested for insecticidal activity and bee toxicity. For the 10 structural analogues of fluralaner, only three

molecules required insecticidal activity testing and one required bee toxicity testing to identify the candidate molecules. In fact, all 10 molecules were synthesized (Figures S7–S14) and their insecticidal activities were tested at 25 mg/L. As shown in Table S1, the accuracy of the PAP model on these 10 molecules was 70%, further confirming the feasibility of this strategy.

3.3. Insecticidal Activity and Bee Toxicity of SAF-A. As **SAF-A** was one molecule with high insecticidal activity and low bee toxicity, the property of **SAF-A** was carefully studied to provide a candidate for the *P. xylostella* control. As the third and fourth-instar larvae are mainly responsible for widespread plant damage caused by *P. xylostella*, the insecticidal activity of **SAF-A** against third-instar larvae and fourth-instar larvae was investigated. As shown in Figure 4a, in the case of third-instar larvae, 0.5 mg/L **SAF-A** led to a significantly lower survival rate than the control treatment after 36 h, while for fourth-instar larvae, the time that 0.5 mg/L **SAF-A** resulted in significantly lower survival rate than the control treatment was delayed to 60 h (Figure 4b). The situation was similar at the other concentrations, suggesting that **SAF-A** was more effective against third-instar larvae than against fourth-instar larvae.

Analyses of the dose–mortality relationship at the same time and the mortality–time relationship at the same virulence further illustrate this view. As shown in Figure 4c, the insecticidal activity of **SAF-A** against third-instar larvae of *P. xylostella* was significantly higher than that against fourth-instar larvae of *P. xylostella*, at both 48 and 72 h. The LC₅₀ of **SAF-A** against third-instar larvae of *P. xylostella* was 0.99 mg/L (48 h) and 0.368 mg/L (72 h). Besides, **SAF-A** took effect more rapidly on third-instar larvae than on fourth-instar larvae when both doses used resulted in 80% mortality at 72 h, as shown in Figure 4d. Therefore, **SAF-A** is recommended for use in the third instar of *P. xylostella*.

The bee toxicity of **SAF-A** was further scrutinized. As shown in Figure 4e, the LC₅₀ of **SAF-A** against bees was 222 mg/L (8.90 μg/bee). In order to evaluate the selectivity of insecticidal molecules in a better way, we proposed the concept of the insecticidal activity–bee toxicity window, that is, log(LC_{50, bee}) – log(LC_{99.9, P. xylostella}), referring to the therapeutic window in medicine. For fluralaner, although its *P. xylostella* activity was higher than **SAF-A**, its insecticidal activity–bee toxicity window was 0, lacking selectivity, while for **SAF-A**, its insecticidal activity–bee toxicity window was 1.6, with good selectivity. Considering its high insecticidal activity and low bee toxicity, the molecule **SAF-A** we screened demonstrates great potential for *P. xylostella* control.

3.4. Effects of SAF-A on the Third-Instar Larvae of P. xylostella. We further explored how **SAF-A** affects the survival of third-instar larvae of *P. xylostella*. As shown in Figure 5a,b, the third-instar larvae treated with 2.5 mg/L **SAF-A** consumed less leaf material compared to both the 0.5 mg/L treatment group and the control group after the initial 12 h of feeding. At 24 h, the leaf consumption of both treatment groups was significantly lower than that of the control group. The feeding inhibition in the treatment groups continued after replacement of new leaf discs without **SAF-A**, as shown in Figure 5c,d. The third-instar larvae in the 2.5 mg/L treatment group even showed no consumption of leaves at 36 and 48 h, despite that the third-instar larvae were not completely dead at 36 and 48 h. These results indicate that **SAF-A** can inhibit the feeding behavior of third-instar larvae of *P. xylostella*, which facilitates the reduction of crop losses due to feeding by *P. xylostella*.

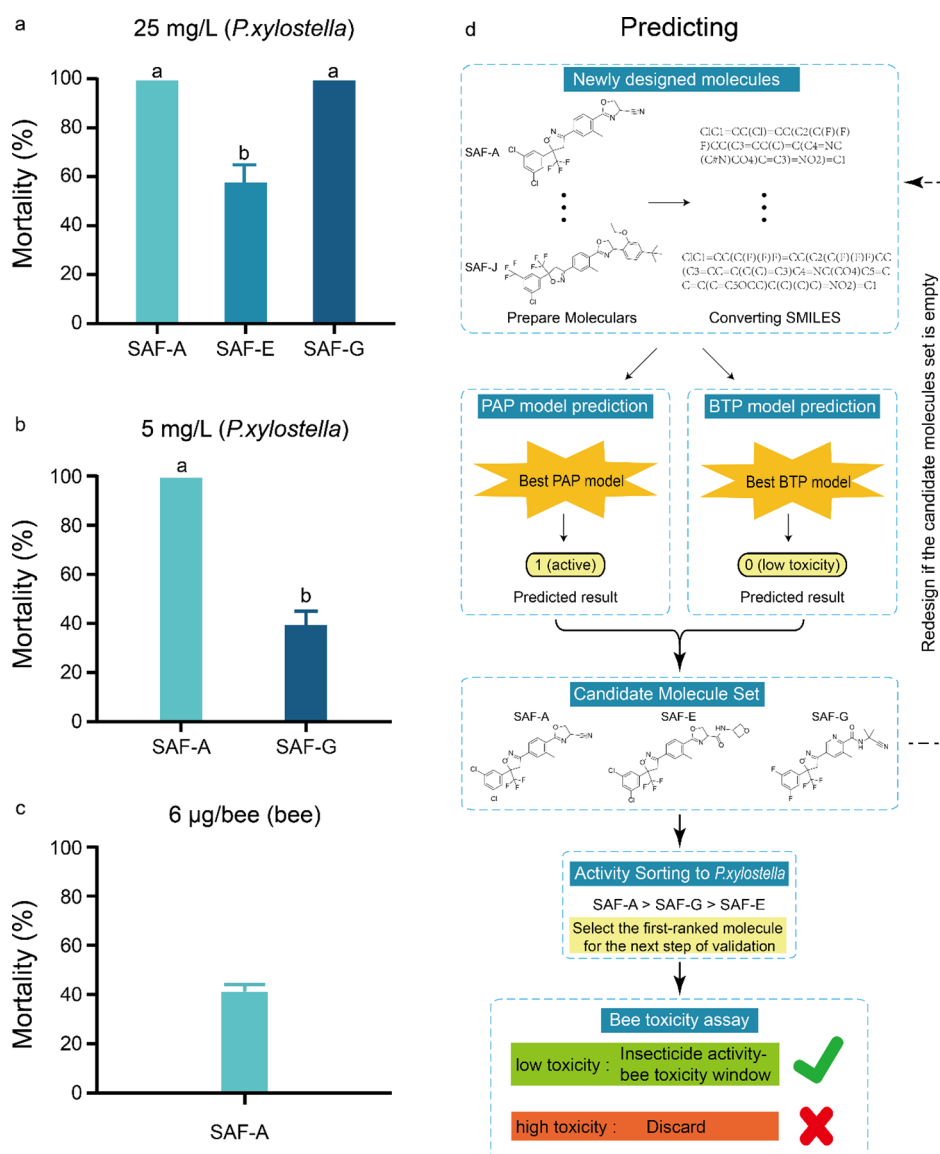


Figure 3. Discovery process of SAFs with high insecticidal activity and low bee toxicity. The mortality of *P. xylostella* after being fed with leaves soaked by **a** 25 mg/L and **b** 5 mg/L SAFs. Mean $\pm$ SE, $n = 3$, analysis by one-way ANOVA followed by Tukey's pairwise test, $P < 0.05$. **c** The mortality of bees after being fed with a sucrose solution (50% w/v) containing SAF-A. Mean $\pm$ SE, $n = 3$. **d** Schematic of the insecticidal molecular discovery process with the dual-ML model. The newly designed molecules were predicted by the dual-ML model, and if there were molecules classified as positive by the PAP model and negative by the BTP model, the next step of validation and ranking of activity was performed; if the set of candidate molecules is empty, then a new batch of molecules was redesigned and the previous step was repeated.

For the growth of larvae, as shown in Figure 5e,f, the few individuals that survived the SAF-A treatment were significantly inhibited in their development, which were still in the early stages of third-instar larvae, whereas larvae in the control group had grown to fourth-instar larvae. Besides, although the body lengths of the third-instar larvae in the 2.5 mg/L treatment group were close to that in the fasting group, the fasted third-instar larvae could grow normally when feeding was resumed, whereas the third-instar larvae treated with 2.5 mg/L SAF-A eventually died altogether. In addition, significant epidermal darkening was observed in the third-instar larvae treated with SAF-A. Considering that SAF-A is a structural analogue of fluralaner, whose target is insect γ -aminobutyric acid receptor,^{46–48} and that both feeding inhibition and epidermal darkening are consistent with the manifestation of impaired

normal function of GABAR,^{49–54} it is hypothesized that the target of SAF-A may also be GABAR.

3.5. Molecular Docking of SAF-A to GABAR of *P. xylostella* and Bees. Molecular docking was used to verify whether GABAR is a target of SAF-A. The GABAR models of *P. xylostella* were first built with human $\beta 3$ GABAR as the homologous templates (Figure S15).⁵⁵ The Z-score value of GABAR models of *P. xylostella* was -4.04 , which is within the range of typical scores for similarly sized native proteins.⁵⁶ A Ramachandran plot showed that residues in most favored regions accounted for 93.2%.^{57,58} These results indicated that the GABAR models were of high quality and could be further used. The stable conformation of SAF-A and GABAR of *P. xylostella* is shown in Figure 6a, with binding energy of 6.40 kcal/mol, which indicates that GABAR is a target of SAF-A.

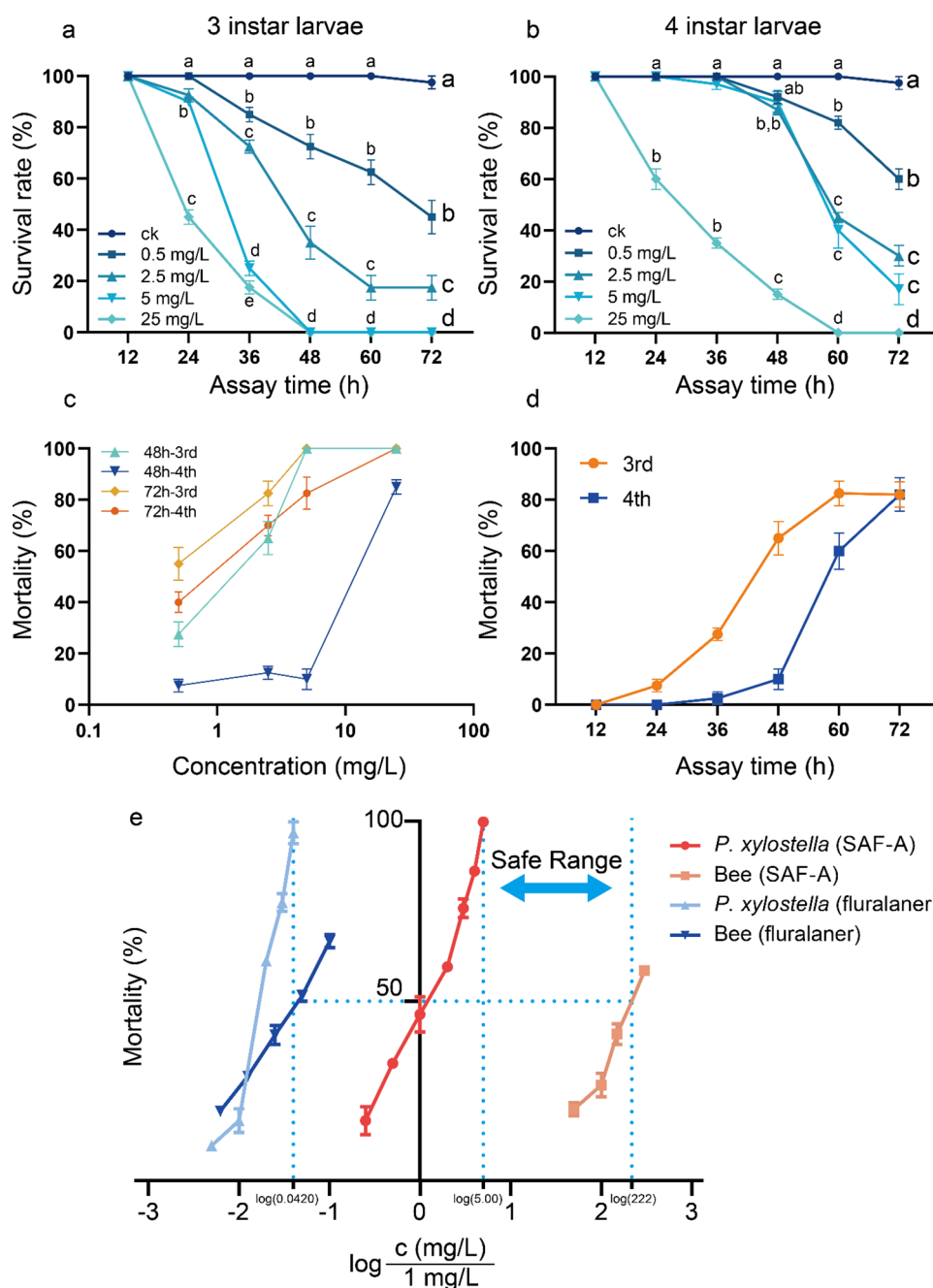


Figure 4. Detailed study of the insecticidal properties of SAF-A. **a** Effect of feeding SAF-A on the survival rate (%) of *P. xylostella* with assay time (h) of third-instar larvae (mean $\pm$ SE, $n = 4$ biologically independent experiments). **b** Effect of feeding SAF-A on the survival rate (%) of *P. xylostella* with assay time (h) of fourth-instar larvae (mean $\pm$ SE, $n = 4$ biologically independent experiments). **c** Dose–mortality relationship of third- and fourth-instar larvae at 48 and 72 h, respectively. Graphing used the logarithm of concentrations. **d** Mortality–time relationship between third- and fourth-instar larvae of the same virulence at 80%, respectively. **e** Study of the insecticidal activity–bee toxicity window of SAF-A and fluralaner, i.e., $\log(\text{LC}_{50, \text{bee}}) - \log(\text{LC}_{99.9, P. xylostella})$ (mean $\pm$ SE, $n = 3$ biologically independent experiments). Concentrations were standardized and plotted using a logarithm. The activity of SAF-A against *P. xylostella* was determined using the leaf dipping method, where leaf discs were replaced every 24 h, and only the first 24 h leaf discs were loaded with the drug. Data represent mean $\pm$ SE ($n = 4$ biologically independent experiments). Curves with different letters indicate significant differences ($P < 0.05$). Source data are provided as a Source Data file.

We wondered whether the interaction of SAF-A and GABAR was one reason for the differing activity of SAF-A against *P. xylostella* and bees. To this end, the effect of the insect instar on activity was first excluded. As shown in Figure S16, the LC_{50} values of SAF-A against adults of *P. xylostella* were 81 mg/L, higher than the LC_{50} of SAF-A against third-instar larvae of *P. xylostella* but lower than the LC_{50} of SAF-A against adult bees, suggesting that insect instar is not the only reason that affects

selectivity. Molecular docking was further performed to study the interaction of SAF-A and GABAR of bees. As shown in Figure S17 and Figure 6b, the binding energy of SAF-A to the GABAR of the bee was only 4.69 kcal/mol, which was much lower than the binding energy of SAF-A to the GABAR of *P. xylostella*. Therefore, the difference in binding energy, together with the fact that SAF-A acts on third-instar larvae of *P. xylostella*

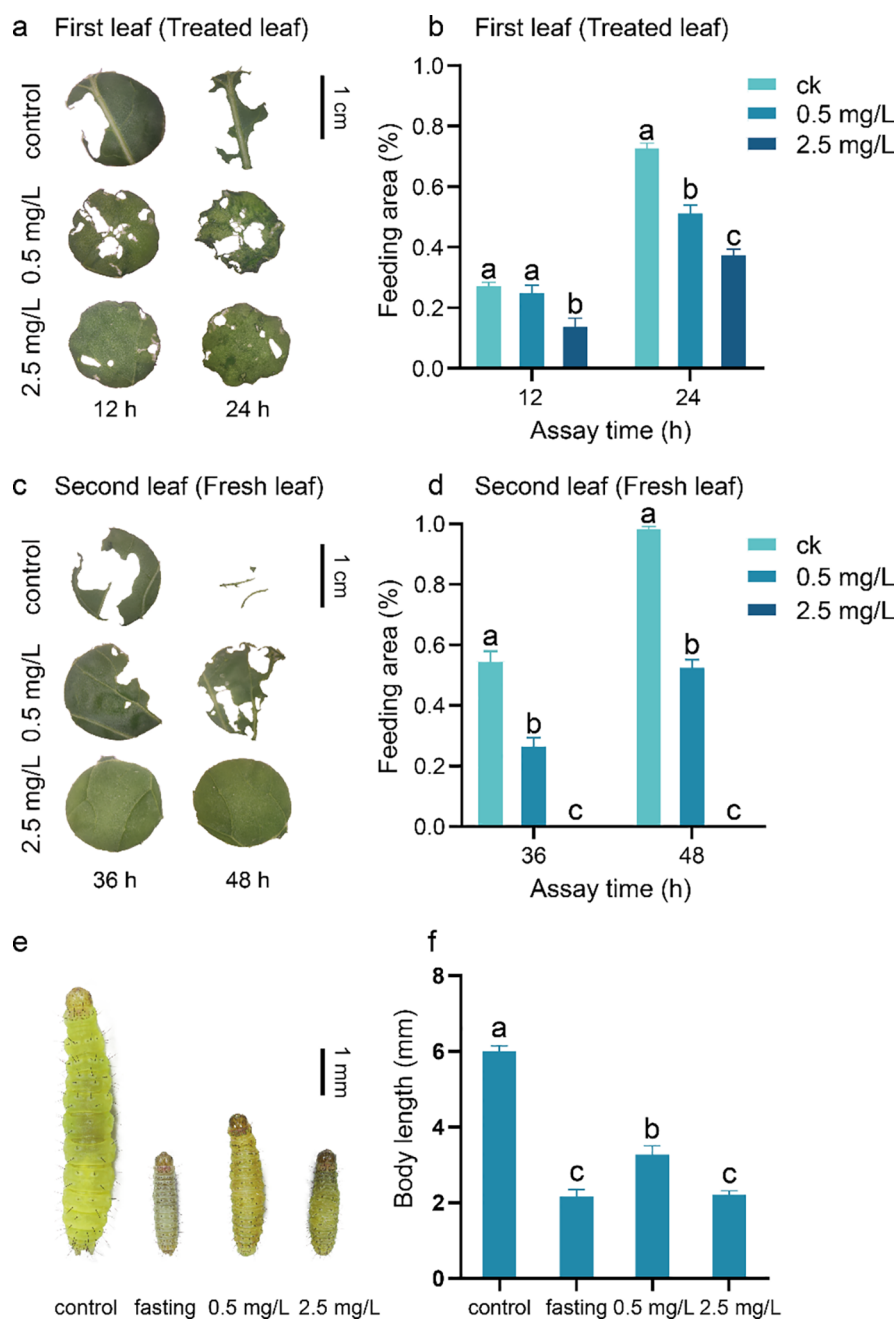


Figure 5. Effects of SAF-A on the feeding and development of third-instar larvae of *P. xylostella*. **a** The area of leaves consumed by a third-instar larva at 12 and 24 h after feeding on leaf discs soaked in SAF-A at concentrations of 0.5 and 2.5 ppm and control solutions, respectively. **b** Percentage of leaves consumed by third-instar larvae at 12 and 24 h (mean $\pm$ SE, $n = 3$ biologically independent experiments). **c** A fresh leaf disc was replaced at 24 h, and the area of the leaf disc consumed at 36 and 48 h was recorded after treatment. **d** Percentage of leaves consumed by third-instar larvae at 36 and 48 h (mean $\pm$ SE, $n = 3$ biologically independent experiments). **e** Photograph of the larval developmental stages at 48 h following the start of the experiment after feeding on two concentrations of SAF-A, 0.5 and 2.5 mg/L. One control group was fed normally for 48 h, and the other control group was starved for 48 h. **f** The body length of *P. xylostella* in the experiment of **e** (mean $\pm$ SE, $n = 3$ biologically independent experiments). The activity of SAF-A against *P. xylostella* was determined using the leaf dipping method, where leaf discs were replaced every 24 h, and only the first 24 h leaf discs were loaded with the drug. Data represent mean $\pm$ SE ($n = 3$ biologically independent experiments). Columns with different letters indicate significant differences ($P < 0.05$). Source data are provided as a Source Data file.

but adults of bees, leads to the high selectivity between *P. xylostella* and bees of SAF-A.

4. DISCUSSION

Designing insecticidal molecules with both high activity and low bee toxicity is of vital importance for green agriculture. However, achieving this balance is challenging, as molecules toxic to pests

often also harm bees. In this study, we introduced a dual-ML-model strategy, consisting of an insecticidal activity prediction model and a bee toxicity prediction model, to identify molecules with both of the desired properties.

In building these ML models, the scale and quality of the molecule data set play crucial roles. A larger data set generally provides more information for the models to learn from, which

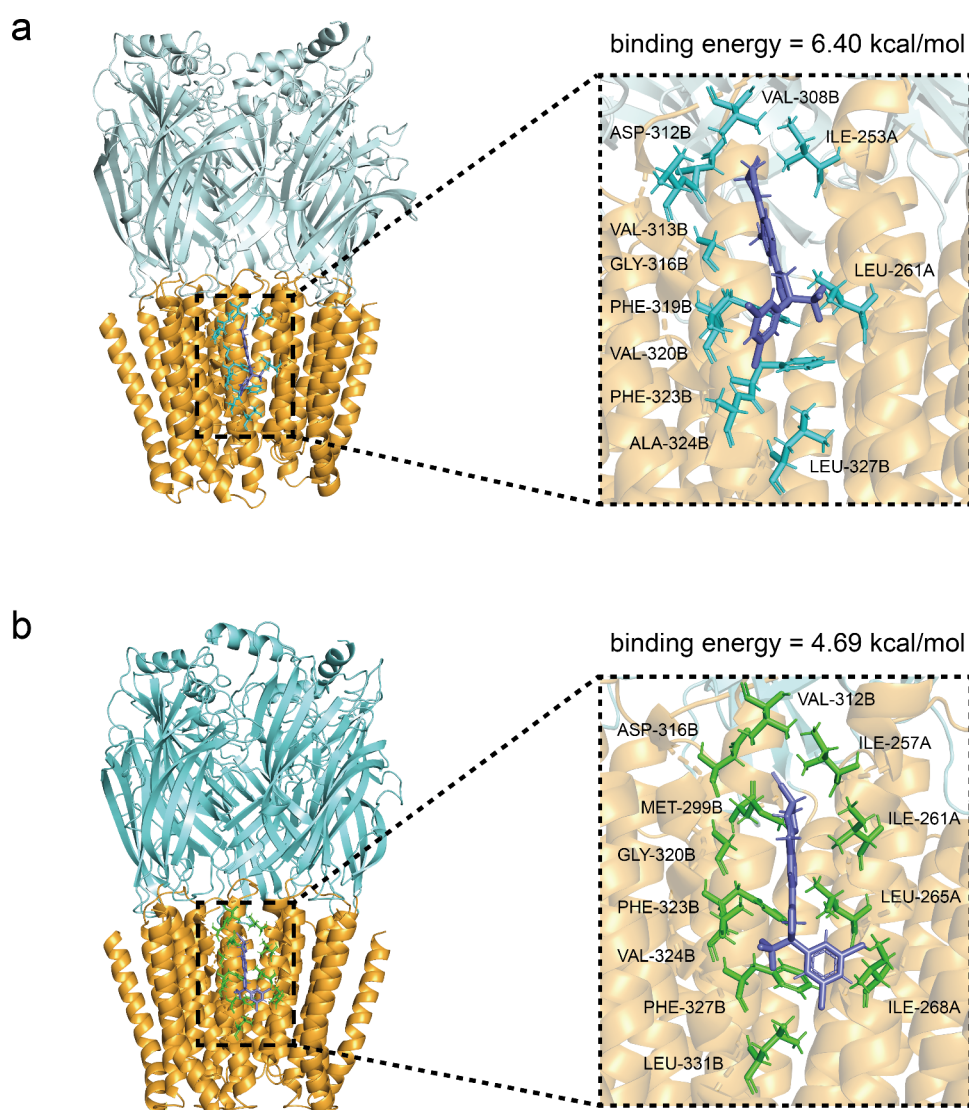


Figure 6. Molecular docking studies. **a** The stable conformation of **SAF-A** exhibits a binding energy of 6.40 kcal/mol with GABAR of *P. xylostella*. Binding energy is contributed by ILE-253A, LEU-261A, VAL-308B, ASP-312B, VAL-313B, GLY-316B, PHE-319B, VAL-320B, PHE-323B, ALA-324B, and LEU-327B together. **b** The stable conformation of **SAF-A** exhibits a binding energy of 4.69 kcal/mol with the GABAR of bees. Binding energy is contributed by ILE-257A, ILE-261A, LEU-265A, ILE-268A, MET-299B, VAL-312B, ASP-316B, GLY-320B, PHE-323B, VAL-324B, PHE-327B, and LEU-331B together.

helps in improving the generalization ability of the models. High-quality data consisting of accurate bioassay results are essential for the reliability of the models. However, for discovering molecules with high insecticidal activity and low bee toxicity through ML, high-quality data remain scarce, especially for insecticidal activity. Given this current situation, we developed predictive models using structural analogues derived from highly active molecules. This approach requires only a small number of experiments to generate sufficient training data.^{59,60} For bee toxicity, we leveraged existing extensive databases of commercial chemical molecules, which provided ample data for training the toxicity prediction model. The BTP model thus constructed exhibits a high generalizability. Although the dual-ML model we constructed is only suitable for predicting structural analogues of fluralaner, it can be extended to other types of molecules by inputting new data sets.

Using this dual-ML model, we predicted **SAF-A** as a molecule with high insecticidal activity and low bee toxicity. Experimental validation confirmed **SAF-A**'s efficacy, with an LC_{50} of 0.99 mg/

L for *P. xylostella* and an LC_{50} of 222 mg/L for bees, indicating its potential as an effective and safe insecticide.

In summary, the dual-ML-model strategy significantly enhances the efficiency of insecticidal molecule discovery, providing agriculture with effective and safe insecticidal molecules. This approach is anticipated to support the development of sustainable agriculture.

■ ASSOCIATED CONTENT

Data Availability Statement

The Python code used to generate all results is publicly available only at <https://github.com/DrugAP/drug-activity-prediction>.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jafc.4c08587>.

Instruments and chemicals used in the study, key molecular descriptors of the ML models, ^1H NMR and ^{13}C NMR spectra of molecules **SAFs**, bioassay of **SAFs**

against *P. xylostella*, molecular docking of SAF-A to GABAR of *P. xylostella* and bees, and bioassay of SAF-A on *P. xylostella* adults at 48 h (PDF)

Supplementary Data 1: Structural analogues of fluralaner (XLSX)

Supplementary Data 2: Insecticide molecules forming the final bee toxicity data set (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Jiajun Ren — Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, 100875 Beijing, People's Republic of China; orcid.org/0000-0002-1508-4943; Email: jjren@bnu.edu.cn

Lingda Zeng — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China; orcid.org/0000-0002-6191-4370; Email: zldvictor@163.com

Hanhong Xu — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China; orcid.org/0000-0001-7841-2396; Phone: +86-20-85285127; Email: hxxu@scau.edu.cn

Authors

Wei Guo — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China

Xiangmin Song — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China

Yongchao Gao — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China

Shuai Yang — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China

Jiahong Tang — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China

Chen Zhao — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China; orcid.org/0000-0002-8042-7025

Haojing Wang — State Key Laboratory of Green Pesticide, Key Laboratory of Natural Pesticide and Chemical Biology,

Ministry of Education, College of Plant Protection, South China Agricultural University, Guangzhou, Guangdong 510642, People's Republic of China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jafc.4c08587>

Author Contributions

Wei Guo and Xiangmin Song contributed equally. Lingda Zeng and Jiajun Ren conceived the project. Hanhong Xu and Lingda Zeng designed and led the project; Wei Guo developed the dual-ML-model approach and wrote the manuscript; Lingda Zeng and Jiajun Ren improved the manuscript; Xiangmin Song, Yongchao Gao, Shuai Yang, and Jiahong Tang synthesized the structural analogues of fluralaner; Chen Zhao directed the synthesis of molecules; Wei Guo, Xiangmin Song, and Haojing Wang performed the experimental work and analyzed the data.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by the Guangdong Province Key Research and Development Plan (Grant No. 2023B0202080001); the National Natural Science Foundation of China (Grant No. 22273005); and Guangdong Provincial Department of Agriculture Rural Revitalisation Strategy Special (Grant No. 2024-440000-90040000-0193).

REFERENCES

- (1) Colin, T.; Monchanin, C.; Lihoreau, M.; Barron, A. B. Pesticide dosing must be guided by ecological principles. *Nat. Ecol. Evol.* **2020**, *4*, 1575–1577.
- (2) Matson, P. A.; Parton, W. J.; Power, A. G.; Swift, M. J. Agricultural intensification and ecosystem properties. *Science* **1997**, *277*, 504–509.
- (3) Barathi, S.; Sabapathi, N.; Kandasamy, S.; Lee, J. Present status of insecticide impacts and eco-friendly approaches for remediation—a review. *Environ. Res.* **2024**, *240*, 117432.
- (4) Ritchie, H.; Roser, M.; Rosado, P. Pesticides: Published online at OurWorldInData.org. 2022, <https://ourworldindata.org/pesticides>.
- (5) Deutsch, C. A.; Tewksbury, J. J.; Tigchelaar, M.; Battisti, D. S.; Merrill, S. C.; Huey, R. B.; Naylor, R. L. Increase in crop losses to insect pests in a warming climate. *Science* **2018**, *361*, 916–919.
- (6) Martinez, D. A.; Loening, U. E.; Graham, M. C.; Gathorne-Hardy, A. When the medicine feeds the problem; Do nitrogen fertilisers and pesticides enhance the nutritional quality of crops for their pests and pathogens? *Front. Sustain. Food Syst.* **2021**, *5*, 701310.
- (7) Knapp, J. L.; Bates, A.; Jonsson, O.; Klatt, B.; Krausl, T.; Sahlin, U.; Svensson, G. P.; Rundlöf, M. Pollinators, pests and yield—Multiple trade-offs from insecticide use in a mass-flowering crop. *J. Appl. Ecol.* **2022**, *59*, 2419–2429.
- (8) Baer-Nawrocka, A.; Sadowski, A. Food security and food self-sufficiency around the world: A typology of countries. *PLoS One* **2019**, *14*, No. e0213448.
- (9) Carvalho, F. P. Agriculture, pesticides, food security and food safety. *Environ. Sci. Policy* **2006**, *9*, 685–692.
- (10) Goulson, D.; Nicholls, E.; Botías, C.; Rotheray, E. L. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* **2015**, *347*, 1255957.
- (11) Catacutan, D. B.; Alexander, J.; Arnold, A.; Stokes, J. M. Machine learning in preclinical drug discovery. *Nat. Chem. Biol.* **2024**, *20*, 960–973.
- (12) Lo, Y. C.; Rensi, S. E.; Torng, W.; Altman, R. B. Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today* **2018**, *23*, 1538–1546.

- (13) Dara, S.; Dhamecherla, S.; Jadav, S. S.; Babu, C. M.; Ahsan, M. J. Machine learning in drug discovery: A Review. *Artif. Intell. Rev.* **2022**, *55*, 1947–1999.
- (14) Diaoy, Y. Y.; Liu, D. D.; Ge, H.; Zhang, R. R.; Jiang, K. X.; Bao, R. H.; Zhu, X. Q.; Bi, H. J.; Liao, W. J.; Chen, Z. Q.; Zhang, K.; Wang, R.; Zhu, L.; Zhao, Z. J.; Hu, Q. Y.; Li, H. L. Macrocyclization of linear molecules by deep learning to facilitate macrocyclic drug candidates discovery. *Nat. Commun.* **2023**, *14*, 4552.
- (15) Müller-Franzes, G.; Huck, L.; Arasteh, S. T.; Khader, F.; Han, T. Y.; Schulz, V.; Dethlefsen, E.; Kather, J. N.; Nebelung, S.; Nolte, T.; Kuhl, C.; Truhn, D. Using machine learning to reduce the need for contrast agents in breast MRI through synthetic images. *Radiology* **2023**, *307*, 3.
- (16) Ouyang, Y.; Huang, J. J.; Wang, Y. L.; Zhong, H.; Song, B. A.; Hao, G. F. In Silico resources of drug-likeness as a mirror: what are we lacking in pesticide-likeness? *J. Agric. Food Chem.* **2021**, *69*, 10761–10773.
- (17) Ding, Y.; Chen, S. Z.; Liu, H.; Liu, T.; Yang, Q. Discovery of multitarget inhibitors against insect chitinolytic enzymes via machine learning-based virtual screening. *J. Agric. Food Chem.* **2023**, *71*, 8769–8777.
- (18) Hwang, J.; Lee, Y.; Yoo, S.; Kim, J. Image-based deep learning model using DNA methylation data predicts the origin of cancer of unknown primary. *Neoplasia* **2024**, *55*, 101021.
- (19) Tong, X. C.; Qu, N.; Kong, X. T.; Ni, S. K.; Zhou, J. Y.; Wang, K.; Zhang, L. H.; Wen, Y. M.; Shi, J. S.; Zhang, S. L.; Li, X. T.; Zheng, M. Y. Deep representation learning of chemical-induced transcriptional profile for phenotype-based drug discovery. *Nat. Commun.* **2024**, *15*, 5378.
- (20) Eckmann, P.; Anderson, J.; Yu, R.; Gilson, M. K. Ligand-based compound activity prediction via few-shot learning. *J. Chem. Inf. Model.* **2024**, *64*, 5492–5499.
- (21) Berg, S.; Kutra, D.; Kroeger, T.; Straehle, C. N.; Kausler, B. X.; Haubold, C.; Schiegg, M.; Ales, J.; Beier, T.; Rudy, M.; Eren, K.; Cervantes, J. I.; Xu, B.; Beuttenmueller, F.; Wolny, A.; Zhang, C.; Koethe, U.; Hamprecht, F. A.; Kreshuk, A. Ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **2019**, *16*, 1226–1232.
- (22) Lilienfeld, O. A. V.; Burke, K. Retrospective on a decade of machine learning for chemical discovery. *Nat. Commun.* **2020**, *11*, 4895.
- (23) Li, J. N.; Mu, D.; Zheng, S. C.; Tian, W.; Wu, Z. Y.; Meng, J.; Wang, R. F.; Zheng, T. L.; Zhang, Y. L.; Windsor, J.; Lu, G. T.; Wu, D. Machine learning improves prediction of severity and outcomes of acute pancreatitis: a prospective multi-center cohort study. *Sci. China Life Sci.* **2023**, *66*, 1934–1937.
- (24) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.
- (25) Li, G. L.; Chen, C. J.; Chen, P. K.; Meyers, B. C.; Xia, R. sRNAmminer: A multifunctional toolkit for next-generation sequencing small RNA data mining in plants. *Sci. Bull.* **2024**, *69*, 784–791.
- (26) Sheng, C. W.; Huang, Q. T.; Liu, G. Y.; Ren, X. X.; Jiang, J.; Jia, Z. Q.; Han, Z. J.; Zhao, C. Q. Neurotoxicity and mode of action of fluralaner on honeybee *Apis mellifera* L. *Pest Manag. Sci.* **2019**, *75*, 2901–2909.
- (27) Yang, S.; Tang, J. H.; Li, B. J.; Yao, G. K.; Peng, H. X.; Pu, C. M.; Zhao, C.; Xu, H. H. Rational design of insecticidal isoxazolines containing sulfonamide or sulfonamide structure as antagonists of GABA receptors with reduced toxicities to honeybee and zebrafish. *J. Agric. Food Chem.* **2023**, *71*, 14211–14220.
- (28) Gao, Y. C.; Song, X. M.; Jia, T. H.; Zhao, C.; Yao, G. K.; Xu, H. H. Discovery of new *N*-Phenylamide Isoxazoline derivatives with high insecticidal activity and reduced honeybee toxicity. *Pestic. Biochem. Physiol.* **2024**, *200*, 105843.
- (29) Hu, X. P.; Ma, X. J.; Cui, J. L.; Liu, H. S.; Zhu, B.; Xie, J.; Liang, P.; Zhang, L. Identification of 1-phenyl-4-cyano-5-aminopyrazoles as novel ecdysone receptor ligands by virtual screening, structural optimization, and biological evaluations. *Chem. Biol. Drug Des.* **2021**, *97*, 184–195.
- (30) Zhang, C. L.; Li, X. L.; Song, D. L.; Ling, Y.; Zhou, Y. L.; Yang, X. L. Synthesis, aphicidal activity and conformation of novel insect kinin analogues as potential eco-friendly insecticides. *Pest Manag. Sci.* **2020**, *76*, 3432–3439.
- (31) Li, K. M.; Persaud, D.; Choudhary, K.; DeCost, B.; Greenwood, M.; Hattrick-Simpers, J. Exploiting redundancy in large materials datasets for efficient machine learning with less data. *Nat. Commun.* **2023**, *14*, 7283.
- (32) Wang, F.; Yang, J. F.; Wang, M. Y.; Jia, C. Y.; Shi, X. X.; Hao, G. F.; Yang, G. F. Graph attention convolutional neural network model for chemical poisoning of honey bees' prediction. *Sci. Bull.* **2020**, *65*, 1184–1191.
- (33) Bernardes, R. C.; Botina, L. L.; Silva, F. P. D.; Fernandes, K. M.; Lima, M. A. P.; Martins, G. F. Toxicological assessment of agrochemicals on bees using machine learning tools. *J. Hazard. Mater.* **2022**, *424*, 127344.
- (34) Como, F.; Carnesecchi, E.; Volani, S.; Dorne, J. L.; Richardson, J.; Bassan, A.; Pavan, M.; Benfenati, E. Predicting acute contact toxicity of pesticides in honeybees (*Apis mellifera*) through a k-nearest neighbor model. *Chemosphere* **2017**, *166*, 438–444.
- (35) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR modeling: where have you been? where are you going to? *J. Med. Chem.* **2014**, *57*, 4977–5010.
- (36) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
- (37) Redžepović, I.; Furtula, B. Predictive potential of eigenvalue-based topological molecular descriptors. *J. Comput. Aided Mol. Des.* **2020**, *34*, 975–982.
- (38) Mauri, A.; Consonni, V.; Todeschini, R. Molecular descriptors. *Handbook of computational chemistry* **2017**, 2065–2093.
- (39) Landrum, G. RDKit: Open-source cheminformatics. 2006, <http://www.rdkit.org>.
- (40) Hamadache, M.; Benkortbi, O.; Hanini, S.; Amrane, A. QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (*Apis mellifera* L.). *Environ. Sci. Pollut. Res.* **2018**, *25*, 896–907.
- (41) Hamadache, M.; Benkortbi, O.; Hanini, S.; Amrane, A.; Khaouane, L.; Moussa, C. S. A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction. *J. Hazard. Mater.* **2016**, *303*, 28–40.
- (42) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Muller, A.; Nothman, J.; Louppe, G.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (43) Xie, M.; Ren, N. N.; You, Y. C.; Chen, W. J.; Song, Q. S.; You, M. S. Molecular characterisation of two α -esterase genes involving chlorpyrifos detoxification in the diamondback moth, *Plutella xylostella*. *Pest Manag. Sci.* **2017**, *73*, 1204–1212.
- (44) Zhang, L. L.; Jing, X. D.; Chen, W.; Wang, Y.; Lin, J. H.; Zheng, L.; Dong, Y. H.; Zhou, L.; Li, F. F.; Yang, F. Y.; Peng, L.; Vasseur, L.; He, W. Y.; You, M. S. Host plant-derived miRNAs potentially modulate the development of a cosmopolitan insect pest. *Plutella xylostella*. *Biomolecules* **2019**, *9*, 602.
- (45) Su, R.; Liu, X. Y.; Wei, L. Y. MinE-RFE: determine the optimal subset from RFE by minimizing the subset-accuracy-defined energy. *Briefings Bioinf.* **2020**, *21*, 687–698.
- (46) Asahi, M.; Yamato, K.; Ozoe, F.; Ozoe, Y. External amino acid residues of insect GABA receptor channels dictate the action of the isoxazoline ectoparasiticide fluralaner. *Pest Manag. Sci.* **2023**, *79*, 4078–4082.
- (47) Yamato, K.; Nakata, Y.; Takashima, M.; Ozoe, F.; Asahi, M.; Kobayashi, M.; Ozoe, Y. Effects of intersubunit amino acid substitutions on GABA receptor sensitivity to the ectoparasiticide fluralaner. *Pestic. Biochem. Physiol.* **2020**, *163*, 123–129.
- (48) Zhang, Y. C.; Huang, Q. T.; Sheng, C. W.; Liu, G. Y.; Zhang, K. X.; Jia, Z. Q.; Tang, T.; Mao, X.; Jones, A. K.; Han, Z. J.; Zhao, C. Q. G3^M_{TMD3} in the insect GABA receptor subunit, RDL, confers

resistance to broflanilide and fluralaner. *PLoS Genet.* **2023**, *19*, No. e1010814.

(49) Liu, S. S.; Wang, M.; Li, X. C. Overexpression of Tyrosine hydroxylase and Dopa decarboxylase associated with pupal melanization in *Spodoptera exigua*. *Sci. Rep.* **2015**, *5*, 11273.

(50) Inagaki, H. K.; De-Leon, S. B. T.; Wong, A. M.; Jagadish, S.; Ishimoto, H.; Barnea, G.; Kitamoto, T.; Axel, R.; Anderson, D. J. Visualizing neuromodulation in vivo: TANGO-mapping of dopamine signaling reveals appetite control of sugar sensing. *Cell* **2012**, *148*, 583–595.

(51) Roberts, B. M.; Lopes, E. F.; Cragg, S. J. Axonal modulation of striatal dopamine release by local γ -aminobutyric acid (GABA) signalling. *Cells* **2021**, *10*, 709.

(52) Patel, J. C.; Sherpa, A. D.; Melani, R.; Witkovsky, P.; Wiseman, M. R.; O'Neill, B.; Aoki, C.; Tritsch, N. X.; Rice, M. E. GABA co-released from striatal dopamine axons dampens phasic dopamine release through autoregulatory GABA_A receptors. *Cell Rep.* **2024**, *43*, 113834.

(53) Zhao, C. Q.; Casida, J. E. Insect gamma-aminobutyric acid receptors and isoxazoline insecticides: toxicological profiles relative to the binding sites of [³H]fluralaner, [³H]-4'-ethynyl-4-n-propylbicycloorthobenzoate, and [³H]ivermectin. *J. Agric. Food Chem.* **2014**, *62*, 1019–1024.

(54) Dong, L. F.; Wang, W. G.; Zhou, L. Q.; Yang, W. L.; Xu, Z. P.; Cheng, J. G.; Shao, X. S.; Xu, X. Y.; Li, Z. Design, synthesis, and bioactivity of trifluoroethylthio-substituted phenylpyrazole derivatives. *J. Agric. Food Chem.* **2024**, *72*, 11949–11957.

(55) Miller, P. S.; Aricescu, A. R. Crystal structure of a human GABA(A) receptor. *Nature* **2014**, *512*, 270.

(56) Wiederstein, M.; Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, 407–410.

(57) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

(58) Laskowski, R. A.; Rullmann, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR.* **1996**, *8*, 477–486.

(59) Song, J.; Ma, Z. Y.; Zhang, Y. C.; Li, T. T.; Yu, G. Rim: A reusable iterative model for big data. *Knowl. Based Syst.* **2018**, *153*, 105–116.

(60) Yang, Y. M.; Wu, Q. M. J.; Feng, X. X.; Akilan, T. Recomputation of dense layers for the performance improvement of DCNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2912–2925.



The graphic features a collage of scientific images and text. At the top, it says 'CAS INSIGHTS™' in yellow. Below that, in large white and blue letters, is 'EXPLORE THE INNOVATIONS SHAPING TOMORROW'. A smaller white text block reads: 'Discover the latest scientific research and trends with CAS Insights. Subscribe for email updates on new articles, reports, and webinars at the intersection of science and innovation.' A yellow button with the text 'Subscribe today' is positioned below this. At the bottom right is the CAS logo, which includes the text 'CAS' and 'A division of the American Chemical Society' next to a stylized molecular structure icon.