Chemical Research in Toxicology

pubs.acs.org/crt

Article

# Cell Painting and Chemical Structure Read-Across Can Complement Each Other for Rat Acute Oral Toxicity Prediction in Chemical Early Derisking

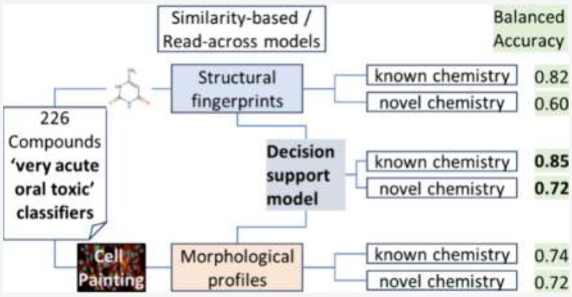
Fabrice Camilleri, Joanna M. Wenda, Claire Pecoraro-Mercier, Jean-Paul Comet, and David Rouquié*

Cite This: https://doi.org/10.1021/acs.chemrestox.4c00169

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

**ABSTRACT:** Early derisking decisions in the development of new chemical compounds enable the identification of novel chemical candidates with improved safety profiles. In vivo studies are traditionally conducted in the early assessment of acute oral toxicity of crop protection products to avoid compounds, which are considered "very acutely toxic", with an in vivo lethal dose of 50% (LD50) ≤ 60 mg/kg body weight. Those studies are lengthy and costly and raise ethical concerns, catalyzing the use of nonanimal alternatives. The objective of our analysis was to assess the predictive efficacy of read-across approaches for acute oral toxicity in rats, comparing the use of chemical structure information, in vitro biological data derived from the Cell Painting profiling assay on U2OS cells, or the combination of both. Our findings indicate that the classification of compounds as very acute oral toxic (LD50 ≤ 60 mg/kg) or not is possible using a read-across approach, with chemical structure information, morphological profiles, or a combination of both. When classifying compounds structurally similar to those in the training set, the chemical structure was more predictive (balanced accuracy of 0.82). Conversely, when the compounds to be classified were structurally different from those in the training set, the morphological profiles were more predictive (balanced accuracy of 0.72). Combining the two models allowed for the classification of compounds structurally similar to those in the training set to slightly improve the predictions (balanced accuracy of 0.85).



## ■ INTRODUCTION

Small molecule discovery involves identifying and developing novel chemical compounds with optimized safety profiles across various target and nontarget species, including laboratory animals, humans, and environmental species. This complex process requires the integration of chemical and biological exploration. Chemists design diverse compounds to interact with specific biological targets or pathways, while biological assays assess efficacy and safety, guiding further chemical optimization. Balancing potency with selectivity and minimizing off-target effects present a significant challenge. To achieve this, chemists often explore new chemical spaces to discover novel structures with the desired properties.[1,2] Additionally, the complexity of biological systems poses hurdles, as the interplay of various factors influences a molecule's properties. Success in small molecule discovery hinges on a multidisciplinary approach, where chemists, biologists, and data scientists collaborate to navigate the intricate landscape of chemical and biological interactions, ultimately advancing the development of new active substances. In agrochemical discovery, early derisking is crucial involving systematic assessment of compounds for potential safety issues to be addressed as early as possible. Addressing genotoxicity and acute oral toxicity is essential due to established cutoff criteria.[3]

Traditionally, early genotoxicity evaluations are performed with in vitro test methods (Ames and Micronucleus assays) whereas acute oral toxicity profile screening involves laboratory animal testing in rodents to obtain a first estimate of the LD50 representing the single dose level at which lethality is induced in at least in 50% of the tested animals. The traditional in vivo studies for acute oral toxicity assessment are time-consuming, expensive, and have low throughput, making it challenging to test a large number of chemicals. Animal studies also raise ethical concerns and must be reduced or if possible, eliminated. Higher throughput alternatives are desired, for the ranking of chemical candidates for their prioritization into the R&D pipeline based on their toxicological profiles. This minimizes resource waste, specifically extended R&D efforts for a chemical that is otherwise only later discovered to not meet required safety standards.

Received: April 25, 2024
Revised: September 25, 2024
Accepted: October 23, 2024

Nonanimal alternatives are already available to reduce in vivo studies, including in vitro approaches and in silico models for early estimation of the LD50. For instance, the in vitro 3T3 neutral red uptake assay (NRU)[4] can serve to categorize compounds with in vivo LD50 greater or lower than 2000 mg/kg. However, this LD50 threshold is high and may not discriminate compounds with lower LD50s. In acute toxicity testing, a low LD50 value (LD50 ≤ 60 mg/kg) means a high potential for acute toxicity, which is an unwanted safety profile.[3]

In silico models are emerging that may close this gap. Quantitative structure−activity relationship (QSAR) models rely on machine learning based on structural information on chemical compounds. One notable example is the Collaborative Acute Toxicity Modeling Suite (CATMoS), a public QSAR model built in a collaboration of several research groups. CATMoS was trained on more than 10,000 compounds and demonstrated high performance in predicting acute oral toxicity in rats.[5] As a regression model, CATMoS can predict the LD50 and classify chemicals into the five categories of the Global Harmonized System (GHS). However, it is important to note that QSAR model predictions are only reliable within their applicability domain, namely, the chemical space covered by the chemistry represented in the training set. This represents a significant limitation when using such models to predict properties of novel chemistry not represented in the training set.

In this situation, we hypothesized that using in vitro highly biologically dense profiles could be used instead of chemical structure information, especially when exploring new chemical spaces. A similar approach has been explored to expand the applicability domain of a QSAR model.[6] The Carpenter−Singh Lab at the Broad Institute has developed an in vitro high-content biologically profiling assay, Cell Painting, which captures the morphological information on cells perturbed by chemicals.[7] The primary advantage of Cell Painting lies in its untargeted nature, theoretically allowing it to capture any bioactivity that induces a change in cell morphology. Additionally, this assay is more cost-effective than other profiling assays such as transcriptomics.[8] Cell Painting has already been used successfully, in "hit" discovery and in mode of action (MoA) prediction.[9−11] In the field of toxicology, the US Environmental Protection Agency (EPA) has explored its use to screen bioactive compounds for human risk assessment.[12] More recently, Cell Painting has also been utilized for the prediction of mitochondrial toxicity[13,14] and liver toxicity.[15]

To evaluate if biological information could complement compound structure predictions for acute oral toxicity especially when exploring new chemical spaces, classifiers were employed to predict whether compounds had very high acute oral toxicity (LD50 ≤ 60 mg/kg) or not (LD50 > 60 mg/kg). Initially, a well-performing public QSAR model, CATMoS, was utilized for the prediction of a set of Bayer Cros Science compounds. Second, a simple chemical structural similarity-based classifier (utilizing a K nearest neighbor[16]) was employed for the prediction of the same set of compounds. This approach, also known as the read-across approach, is commonly used in toxicology specifically in the case of tox data-poor chemicals, such as REACH compounds with a limited number of in vivo results.[17] To simulate scenarios where predictions using the models are made on novel chemistry, a "novel chemistry" holdout strategy was created to assess the classifier. To verify the efficacy of this holdout strategy, the chemical structural similarity-based classifier was again evaluated. A Cell Painting assay was conducted on a smaller subset of compounds using the U2OS cell line to evaluate read-across morphological profiles. The results of the two similarity-based classifiers were assessed: one based on chemical structure and one based on morphological profiles derived from Cell Painting. Both classifiers were also tested in the context of the "novel chemistry" holdout strategy. Finally, a decision support model was constructed to determine which of the two similarity-based classifiers (chemical structure or morphological profile) should be recommended for the prediction of acute toxicity.

Overall, our results showed that the classification of compounds as very acute oral toxic (LD50 ≤ 60 mg/kg) is possible using a read-across approach, with chemical structure information, morphological profiles, or a combination of both. When classifying compounds structurally similar to those of the training set, the chemical structure was more predictive (balanced accuracy of 0.82). Conversely, when the compounds to be classified were structurally different from those of the training set, the morphological profiles were more predictive (balanced accuracy of 0.72). Combining both models allowed for the classification of compounds structurally similar to those used to train the classifiers to slightly enhance the predictions (balanced accuracy of 0.85).

## ■ MATERIALS AND METHODS

**Acute Oral Toxicity Compound Classes.** The compounds were divided into two classes. The class designated as "very acutely oral toxic" (abbreviated VAOT) included compounds with a lethal dose (LD50) of 60 mg/kg or less. The class designated as "not very acutely oral toxic" (abbreviated NVAOT) included compounds with a lethal dose (LD50) greater than 60 mg/kg. (Table 1).

**Table 1. Definition of the Two Acute Oral Toxicity Classes**

| acute oral toxicity classes | |
|---|---|
| very acutely oral toxic − VAOT | rat oral LD50 ≤ 60 mg/kg |
| not very acutely oral toxic - NVAOT | rat oral LD50 > 60 mg/kg |

**Compound Selection.** *Compounds with Acute Oral Toxicity Results in Rats.* To select compounds, Bayer internal databases were queried. A total of 765 compounds with in vivo rat acute oral toxicity results were found for two doses: 60 and 300 mg/kg. Of the 765 compounds, 109 compounds were identified as acute toxic at a dose of 60 mg/kg, meaning compounds belonging to the VAOT class, and 521 compounds were not toxic at 300 mg/kg, meaning compounds belonging to the NVAOT class. To create a robust contrast between the morphological profiles of the VAOT and NVAOT classes, we excluded 135 compounds that were acutely toxic at 300 mg/kg but not at 60 mg/kg. This resulted in the first data set of 630 compounds, which were used to test the acute toxicity prediction of the collaborative Acute Toxicity Modeling Suite (CATMoS)[5] and to build a chemical structure similarity-based classifier.

For the Cell Painting campaign, we checked which of the previous data set compounds were available in Bayer internal compound repository, 81 VAOT compounds were found. To complete the list of VAOT compounds, we queried the chemIDplus public database,[18] and selected 29 compounds that were available in Bayer compound logistics, making a total of 110 VAOT compounds. To have a balanced data set, 116 NVAOT compounds were selected. To have a good chemical structure diversity among them, the Butina algorithm[19] was used to cluster the 521 compounds of the previous data set, based on the Tanimoto similarity of their Morgan fingerprint and a threshold of 0.7: a maximum of clusters was selected and the number of compounds coming from the same cluster were minimized (Supporting Information Figure S1). This selection resulted in a total of 116 NVAOT compounds and 110 VAOT.

To summarize, two sets of compounds were defined. The first one, called "QSAR only compound set", was a set of 630 compounds; 109 were VAOT and 521 were NVAOT. The second one, called "Cell

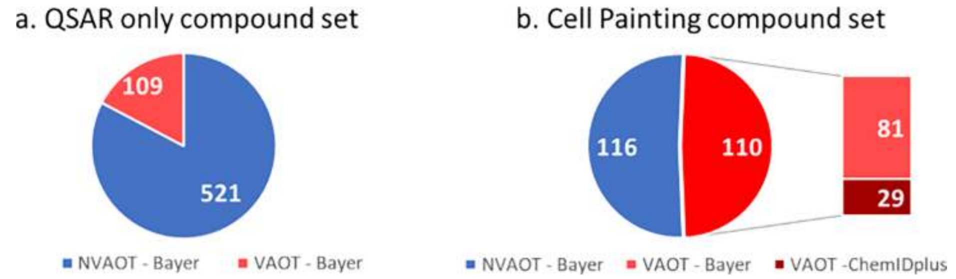


**Figure 1.** Composition of the two data sets by compound class: VAOT (very acutely oral toxic) and NVAOT (not very acutely oral toxic); and by source: Bayer et al. and ChemIDplus. (a) QSAR is the only compound set. Set used for QSAR classifiers only. (b) Cell Painting compound set was used for chemical structure and morphological profile classifiers.

Painting compound set", was a set of 226 compounds; 110 were VAOT, and 116 were NVAOT (Figure 1).

*Negative and Positive Controls.* For the Cell Painting assay, DMSO (dimethyl sulfoxide) only (0.1%) was used as the negative control.

To monitor the Cell Painting assay's performance and assess the quality of the experiment's replicates, a set of positive controls was used. These were compounds that inducibly produced reproducible and distinct morphological profiles in U2OS cells. The selection was based on published literature and pilot tests in our lab (Supporting Information, Table S2).

*CATMoS QSAR Model.* As a first attempt, the acute oral toxicity class of the compounds was predicted using the collaborative Acute Toxicity Modeling Suite (CATMoS) QSAR model, implemented in the OPERA (version 2.9)[5,20] QSAR suite. To align with the two classes presented in this paper (VAOT and NVAOT), three CATMoS predictions were needed: the EPA classification, the GHS classification, and the LD50 range estimation. All compounds being classified by CATMoS as EPA category 1 (LD50 ≤ 50 mg/kg), GHS category 1 (LD50 ≤ 5 mg/kg), or GHS category 2 (5 mg/kg < LD50 ≤ 50 mg/kg) were designated as VAOT. For compounds being classified as EPA category 2 (50 mg/kg < LD50 ≤ 500 mg/kg) or GHS category 3 (50 mg/kg < LD50 ≤ 300 mg/kg), the lower limit of the LD50 range predictions was considered: if the inferior limits were smaller than 60, the compounds were classified as VAOT. For all other predictions, the compounds were classified as NVOAT.

The Opera implementation of CATMoS provides three prediction reliability metrics[20] that were used to understand the predictions made by the model. A global applicability domain Boolean value is calculated, indicating if a compound falls within the training set's chemical space. Additionally, an applicability domain index is calculated, ranging from zero to one, revealing the proximity of the queried compound to the training set. This index is relative to the similarity of the query chemical to its five nearest neighbors.[20] In particular, a query chemical compound can belong to the CATMoS AD (global AD = 1) but can also be in a "gap" of the training chemical space (AD index < 0.6).[20] In such cases, the predictions should be considered with caution.[20] Finally, a confidence index is computed, indicating the accuracy of the prediction of the neighbors of the queried compound.

*Cell Painting Campaign.* A Cell Painting campaign was conducted in our laboratory to obtain the morphological profiles of our "Cell Painting compound set" (set of 226 compounds). The Cell Painting Protocol v3 of the Broad Institute was utilized on U2OS human osteosarcoma cells with four biological replicates.[21]

A previous Cell Painting pilot conducted at the unique dose of 10 $\mu$M demonstrated that few agrochemical compounds exhibited morphological changes compared to the negative control (see the Morphological Change Signal Measure section, and the Biological Response section). Therefore, in this campaign, to enhance the likelihood of capturing a morphological response, the compounds were screened at three concentrations: 10, 31.6, and 100 $\mu$M.

*Cell Culture and Seeding.* Human osteosarcoma cells U2OS have been purchased from ATCC (ref: HTB-96, lot: 70025046). The McCoy's 5A modified medium with GlutaMAX supplement (Thermo Fisher, ref: 36600021) supplemented with 10% fetal bovine serum (Gibco, ref: 16000044) and penicillin/streptomycin mix (Sigma-Aldrich, ref: P4458) was used for culturing cells in T75 or T175 flasks in a standard humidified incubator (37 °C and 5% $CO_2$). The passages were performed when the culture achieved about 80% confluency. Trypsin (Thermo Fisher, ref 25200056) was used to detach the cells during passage, and the number of live cells was calculated with an automatic cell counter (Countess II, Thermo Fisher) after staining the cells with trypan blue (Sigma, ref.: T8154). For the creation of a cell bank, the vial with frozen cells received from the supplier was thawed and expanded until internal passage 3 (P3). At this stage, the cells were cryopreserved in complete media supplemented with 10% DMSO in an ultralow temperature freezer (−150 °C) creating a master bank. One vial of the master bank was then thawed, expanded until internal passage no. 6 (P6), and cryopreserved as before to create a working bank. Vials of the working bank were then directly used for seeding the microplates. One vial of cells (containing 4 million cells) was removed from a −150 °C freezer and thawed in the water bath. The contents of the vial were immediately added to 10 mL of preheated complete media and centrifuged (5 min, 120 × *g*). After removal of the supernatant, the cell pellet was resuspended in 10 mL of complete medium through thorough pipetting. The cell suspension was then added to 150 mL of medium in a round bottle with a magnetic stirrer and immediately used for seeding the 384-well microplates (Greiner BioONE CELLSTAR $\mu$CLEAR; ref: 781091). Multidrop (Thermo Fisher) was used to automatically distribute 36 $\mu$L of cell suspension per well, resulting in a seeding density of around 900 cells/well. The cells were then incubated at 37 °C in an atmosphere of 5% $CO_2$ in an automatized incubator (Cytomat 2, Thermo Fisher). All experimental replicates were performed on a different day using a separate cell vial originating from the same working bank (P6).

*Chemical Treatment.* The test compounds were received in powder form in 96-well deep well plates. They were then dissolved in DMSO (dimethyl sulfoxide) to create 100 mM stock solutions, aliquoted in 96-well V-bottom plates (V96 PP Plate, Thermo Fisher), and frozen at −20 °C until the day of the treatment. Every biological replicate of the experiment originates from a separate aliquot of the stock solution plate so that the compounds undergo only 1 freeze−thaw cycle. On the day of the treatment (24 h postseeding), the plates containing stock solutions were thawed and the compounds were diluted in DMSO to create dose plates containing three concentrations per compound: 100, 31.6, and 10 mM. The dilutions were performed with the use of a Viper liquid handler (Synchron). The compound solutions from the dose plates were then administered to the cell plates in a two-step process. First, an intermediate dilution was prepared: 1 $\mu$L of the compound solution was diluted in 100 $\mu$L of complete cell medium (1:100 dilution), and next, 4 $\mu$L of the resulting intermediate solution was administered to the cell plate (4 $\mu$L of the diluted compound into 36 $\mu$L of cell media, 1:10 dilution). The final concentrations of compounds that the cells were exposed to were therefore 10, 31.6, and 100 $\mu$M, and the final vehicle (DMSO) concentration was 0.1%. The treated cell plates were subsequently incubated with the compounds for 48 h.

*Staining.* The staining and fixation were performed following the published protocol[21] with the use of PhenoVue JUMP kit (PerkinElmer, ref.: PING23). Briefly, 20 $\mu$L/well of the Mitotracker solution was distributed to the cell plates with Multidrop (final concentration: 500 nM). After 30 min of incubation at 37 °C, 20 $\mu$L/

well of 16% PFA solution (Thermo Fisher, ref 28908) was added. The fixation was performed at room temperature (25 °C). Two washes with HBSS buffer (Gibco, ref: 14065−056) were performed with the aid of a Mutlifo washer (BioTek). Twenty $\mu$L/well of the staining solution (HBSS, 1% BSA, 0.1% Triton X-100, 43.7 nM PhenoVue Fluor 555−WGA; 48 nM PhenoVue Fluor 488−Concanavalin A; 8.25 nM PhenoVue Fluor 568−Phalloidin; 1.62 $\mu$M PhenoVue Hoechst 33342 nuclear stain; 6 $\mu$M PhenoVue 512 nucleic acid stain) was added, and the plates were incubated for 30 min at room temperature before being washed again three times with HBSS. The plates were then sealed with aluminum foil, and images were recorded directly.

**Morphological Profile Generation.** *Image Acquisition.* ImageXpress Micro 4 epifluorescent microscope (Molecular Devices) with a 20× air objective was used for recording the fluorescent images (16-bit). The camera binning was set to 2 × 2. The total imaged area per well spanned 2163 $\mu$m × 2163 $\mu$m and consisted of 3 × 3 adjacent fields of view placed in the center of the well. For each field of view, images were recorded in five channels. The following filter sets were used: DAPI, GFP, Cy3, Texas Red, and Cy5. The Z-offset and exposure times were set separately for each channel. A total of 207,360 images were acquired in this campaign.

*Feature Extraction.* Morphological features were extracted using CellProfiler (version 4.2.1), the cell analysis software developed by the Broad Institute.[22] Two CellProfiler pipelines were used: one pipeline for image illumination correction and one pipeline for image analysis. The image illumination correction works at the plate level and averages the intensity of the images of each channel. The image analysis pipeline segmented objects on each image, they were labeled according to the channel they were segmented on, and thousands of measurements were made on those objects at the cell level. It also took measurements at the image level. A total of 4,761 features were measured and formed the morphological profile of a given cell.

*Aggregation and Normalization.* After extracting the cell morphological profiles with CellProfiler, features were aggregated at the well level by taking the means of each feature.

The features were then normalized, following the Broad Institute approach,[7] using the "mad robustized" method of the Python pycytominer package provided by the Broad Institute.[23] The normalization process involved calculating the median of all wells on a plate for each feature, subtracting this value from the median absolute deviation (MAD) of the wells on the plate, and then multiplying the result by 1.4826 to obtain an unbiased estimator. To avoid a null denominator when the MAD was null, a value of $10^{-18}$ was added to the MAD.

$$x_{\text{norm}} = \frac{x_{\text{well}} - \text{median}(x_{\text{plate}})}{\text{mad}(x_{\text{plate}}) + \varepsilon}$$

where $x_{\text{norm}}$ is the normalized value of a morphological profile feature.

$x_{\text{well}}$ is the value of a morphological profile feature.

$x_{\text{plate}}$ shows the values of a morphological profile feature of a plate.

$$\text{mad}(x_{\text{plate}}) = 1.4826 \times \text{median}(|x_{\text{plate}} - \text{median}(x_{\text{plate}})|)$$

$\varepsilon$ an infinitesimally small positive quantity ($10^{-18}$) to avoid a null denominator.

**Quality Check.** To assess the quality of the Cell Painting experiment, several metrics were calculated. First, the number of cells of the negative control treatment (DMSO) was monitored, which was an output of the Cell Profiler segmentation. The cell numbers should fall within the range of [1800; 3000] cells per well. The coefficient of variation of the number of cells for the negative control treatment should not exceed 15% per plate. All plates met the initial quality control standard.

To identify any other potential technical issues with the experiment, we calculated the Pearson correlations of positive controls across plates were calculated. The positive controls were selected to elicit very distinct and reproducible morphological profiles and were included in each plate. In a well-executed experiment, the replicates of these treatments should be well correlated. A correlation threshold of 0.8 was set for the Pearson correlation between replicates. Replicates were considered nonwell-correlated if the correlation fell below this threshold. No outlier plates were identified at this stage, and all of the plates passed this quality control step.

There were 10 missing feature values at the well level in the morphological profiles. The majority of these values originated from the "Cells_AreaShape_FormFactor" feature. This feature was removed along with two wells, in order to remove all missing values.

Further outliers were identified based on the number of cells within a group of replicates. Some wells exhibited a significant discrepancy in cell counts, with values exceeding 1800 cells, compared with other replicates of the same treatment. Twenty-two wells were identified as outliers and removed from the analysis.

*Unsupervised Feature Selection.* To reduce the dimensionality of the normalized morphological profiles, an unsupervised feature selection was performed with the "feature_select" function of the pycytominer Python package.[23]

This function performed several steps to select the features. First, highly correlated features were removed. For a pair of features with a Pearson correlation greater or equal to 0.9, the feature with the smallest sum of correlations with other features was removed. Second, features with low variances were removed. For a given feature, if the count of the second most common feature value divided by the count of the most common feature value was less than 0.05, the feature was removed. Furthermore, features with a ratio defined as the number of unique feature values divided by the number of samples, below 0.01, were excluded. Third, a list of features (contained in the package) that are known to be noisy and generally unreliable were removed. Fourth, features with at least one absolute value greater than 500, values considered as outliers, were not retained. Finally, within each treatment group, any features with a standard deviation greater than 1.2 were removed. This was done to identify and remove any noisy features. This process resulted in a total of 644 features that were then used for downstream analysis.

*Consensus Profiles.* For a given treatment, in our case a chemical compound at a given concentration, the consensus profiles were obtained by aggregating the replicates. This was done by taking the median values of the replicates for each of the remaining features after unsupervised feature selection.

*Morphological Change Signal Measure.* We utilized the grit score,[24,25] a metric developed by the Broad Institute, to measure the morphological changes in a treatment replicate relative to the negative control treatment (DMSO). To calculate this metric, different Pearson correlation coefficients were calculated. First, the correlations between the morphological profiles of a given treatment replicate and each of the negative control treatments were calculated. The distribution of those correlations was defined by its mean and standard deviation. Subsequently, the correlations between the morphological profiles of a given treatment replicate and those of other replicates of this treatment were calculated. Each of the previous correlation coefficients was z-transformed by using the distribution of the correlations with the negative controls. The mean was subtracted, and the results were divided by the standard deviation. The grit scores were then obtained by taking the mean of the transformed values.

The grit score indicated how much a given replicate profile deviated from the negative control profiles. A high grit score indicated a high deviation of the profile from the negative control profiles.

Median grit scores for a given treatment were also calculated, taking the median of the treatment replicate grit scores. This median grit score value allowed measuring how much a treatment impacts the morphology of U2OS cells compared to the negative controls (DMSO).

A threshold of 1 was set to indicate a treatment that induces a morphological change compared to that of the negative control. A grit score of 1 means that the correlation of the morphological profile of a treatment to its replicates is one standard deviation away from the mean of its correlation with the negative control profiles.

*Molecular Fingerprints.* The compound structures were extracted from the Bayer database in SMILES format (simplified molecular-input line-entry system). To reproduce the case when new chemical structures fall outside the applicability domain of chemical structure-based predictive models, the Morgan fingerprints were employed to

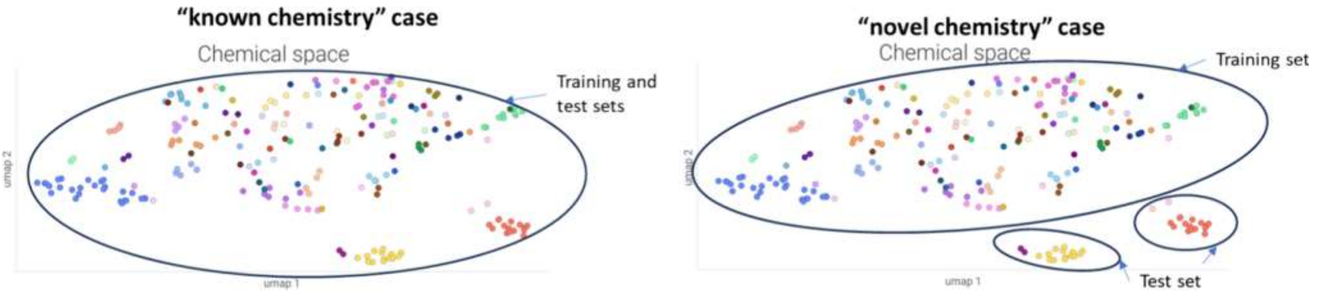


**Figure 2.** Example on UMAP of the chemical space of training and test sets from the two holdout strategies. For the "known chemistry" case, the training and test sets are originated from the same space. For the "novel chemistry" case, the training and test sets are originated from different Butina clusters. Each color on the UMAP corresponds to a Butina cluster.

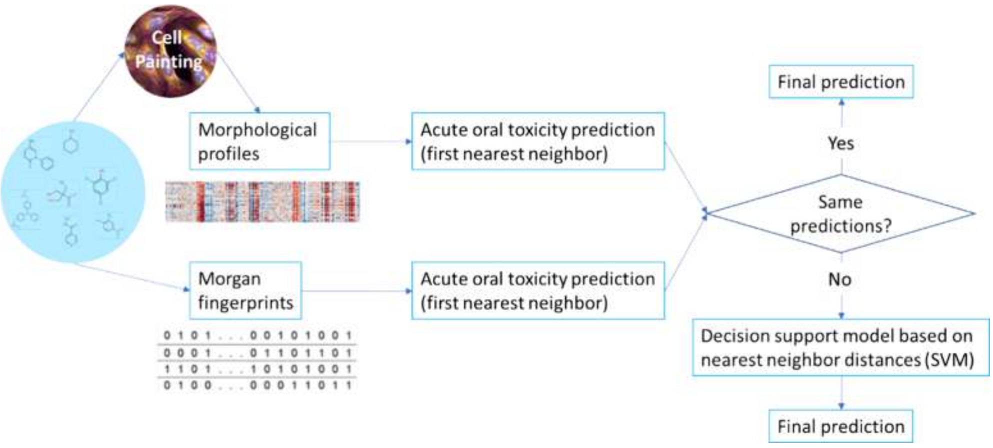


**Figure 3.** Decision support model.

describe the chemical structures of the compounds. To ensure sufficient Butina clusters for our data holdout strategy (see Chemical compounds Clustering with Butina and Data Set Splits), we selected a Morgan fingerprint on 1024 bits to provide the most detailed chemical structure description and thus differentiate more structural differences.[19,26,27] To obtain them from the SMILES, we performed the following steps using the RDKIT Python package.[28]

First, the SMILES were cleaned using the MolStandardize module of RDKIT. This involved removing hydrogens, disconnecting metal atoms, normalizing and ionizing the molecule, and keeping the parent fragments when several fragments of a compound existed. The molecule was then neutralized, and the canonical tautomer was returned. Finally, the cleaned SMILES were used to compute the Morgan fingerprints on 1024 bits, with a radius of three.

*Chemical Compounds Clustering with Butina.* The Butina clustering algorithm groups molecules based on their structural similarity.[19] The RDKit implementation of the Butina algorithm was used to cluster the chemical compounds.[28] The clustering was based on the Tanimoto similarity of the Morgan fingerprints of the molecules, with a cutoff value of 0.7.

*Data Set Splits.* To assess the performance of the binary classifiers, the data set was split several times into training and testing sets. Two types of splits were performed: a random one, which did not consider the chemical similarities of the compounds, and another split that aimed to create sets of structurally different chemicals. This was done to produce cases where the compounds to be classified were novel structures and therefore outside the applicability domain (Figure 2).

For the random split, called the "known chemistry case", a stratified 10-fold-cross-validation was performed to split the data set into 10 different training and testing sets. The scikit-learn Python package[29] was used to perform those splits with the StratifiedKFold function. The data set was split 10 times with a 10-fold cross-validation with each cross-validation having a different random state, resulting in a total of 100 different splits. Each testing set included 22 or 23 compounds.

For the splits based on the chemical structures of the molecules, called the "novel chemistry case", the compound structures were clustered using the Butina clustering algorithm.[19] A cluster number was then assigned to each compound. The StratifiedGroupK-Fold function of scikit-learn was used to make a 10-fold cross-validation based on the cluster number.[29] Indeed, this function assigned cluster numbers to the testing sets that differed from those in the training set. Additionally, it was attempted to maintain a consistent ratio of VAOT and NVAOT classes within each set. The data set was split in this manner 10 times, with different random states, resulting in a total of 100 unique splits. If any of these splits did not include compounds from both classes in the test sets, then they were discarded.

*Binary Classification Classifier.* To classify compounds as VAOT or NVOAT, several algorithms were tested. For this analysis, we decided to use a K nearest neighbors (KNN) algorithm,[16] as it showed good performances (Supporting Information). The KNN algorithm has also the advantage of being explainable and functions like a read-across, technique commonly used for toxicity prediction.[30]

We used the scikit-learn[29] implementation of the K nearest neighbors (KNN) classification algorithm. Several classifiers were built depending on the data that were used as the input. When using the chemical Morgan fingerprints, the Tanimoto (Jaccard) distance was used, and when using the morphological profiles, the Pearson correlation-based similarity measure (1-Pearson correlation) was used. For all classifiers, we set the number of neighbors to one. The choice of the distances and the number of neighbors were the results of benchmarking done on both data sets (supplementary data).

*Decision Support Model.* To aid the decision when the two types of classifiers (Morgan fingerprint and Cell Painting morphological profile classifiers) did not predict the same class, a model was built, similar to the Similarity-based merger model.[31] This ensemble model takes as input the predictions of the two KNN classifiers along with the distances of the nearest neighbors of each prediction (in total four values). A classifier was trained in each training set for the cases where

Table 2. QSAR only compounds set (set of 630 compounds)

a. Classification using CATMoS[a]

| | | Predicted Class | |
|---|---|---|---|
| | | VAOT | NVAOT |
| True class | VAOT | 5 | 104 |
| | NVAOT | 7 | 514 |
| Balanced Accuracy | MCC | Sensitivity | Specificity |
| 0.52 | 0.09 | 0.05 | 0.99 |

b. Prediction Reliabilities[b]

| Within Applicability Domain | Applicability Domain Index | Number of Compounds (percentage) | Average Confidence Index |
|---|---|---|---|
| no | all | 2 (0.3%) | NA |
| yes | <0.6 | 448 (71.1%) | 0.5 |
| yes | ≥0.6 | 180 (28.6%) | 0.57 |

c. Performance of the KNN classifiers trained on 630[c]

| Holdout Strategy | Balanced Accuracy | MCC | Sensitivity | Specificity |
|---|---|---|---|---|
| "known chemistry" case | 0.81 ± 0.07 | 0.60 ± 0.12 | 0.69 ± 0.13 | 0.92 ± 0.04 |
| "novel chemistry" case | 0.61 ± 0.16 | 0.17 ± 0.24 | 0.31 ± 0.29 | 0.92 ± 0.21 |

[a] Confusion matrix and metrics for the classification of CATMoS for 630 compounds from Bayer Crop Science. [b] Reliability of the predictions: Number of compounds outside the CATMoS applicability domain, number of compounds, and average confidence index for compounds within the CATMoS applicability domain and having an Applicability Domain index below or above 0.6. [c] Mean of 4 metrics assessing the performance of the KNN binary classifiers built out 630 Bayer CropScience agrochemical candidates, over the 100 splits of the "known chemistry" case where training and testing sets are split randomly, not taking into account chemical structure similarities, and over the 44 valid splits of the "novel chemistry" case where training and testing sets are split in order to have structurally different compounds over the two sets.

the two KNN classifiers did not agree on the predicted class. In the test sets, we used this model only when the two KNNs did not predict the same classes, otherwise the consensual predicted classes of the two classifiers were set as the final class (Figure 3).

For this decision support model, we used an SVM classifier[32] implementation of scikit-learn.[29]

In the "novel chemistry" case, the training set did not have enough examples with high distances. To remedy this, synthetic examples of distant structures were added. To do this, we subset the cases in each training set where the chemical structure-based predictions did not match the real class, and we updated the nearest neighbor distances with a random number between 0.7 and 0.9 and added these synthetic examples to the data set used to train the model. By doing so, the decision support model can learn not to favor the class of structurally distant chemical compounds.

*Model Performance Evaluation.* To evaluate the performance of the classifiers, we used different metrics. They were all based on the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which were the results of the model classification of a given testing set in a confusion matrix.

$$\text{sensitivity:}\ \ \text{SN} = \text{T}\frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{specificity:}\ \ \text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{balanced accuracy:}\ \ \text{BA} = \frac{\text{SN} + \text{SP}}{2}$$

$$\text{Matthews correlation coefficient:}\ \ \text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

$$\text{accuracy:}\ \text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}$$

When using cross-validation, those metrics were averaged over all testing sets and their standard deviations were calculated.

We performed a corrected *t* test[33] to compare the balanced accuracy values over the splits of classification models.

*Visualization of the Chemical and Biological Spaces.* To visualize on two-dimensional scatter plots the chemical and the biological spaces, UMAP embeddings were generated using the UMAP Python package.[34] For the chemical space, the chemical compound structure similarities embeddings were calculated with the Tanimoto distances of the compound Morgan fingerprints. As for the biological space, the morphological profile similarity embeddings were computed using their Pearson correlation-based similarity measures. The plots were visualized with TIBCO Spotfire software.

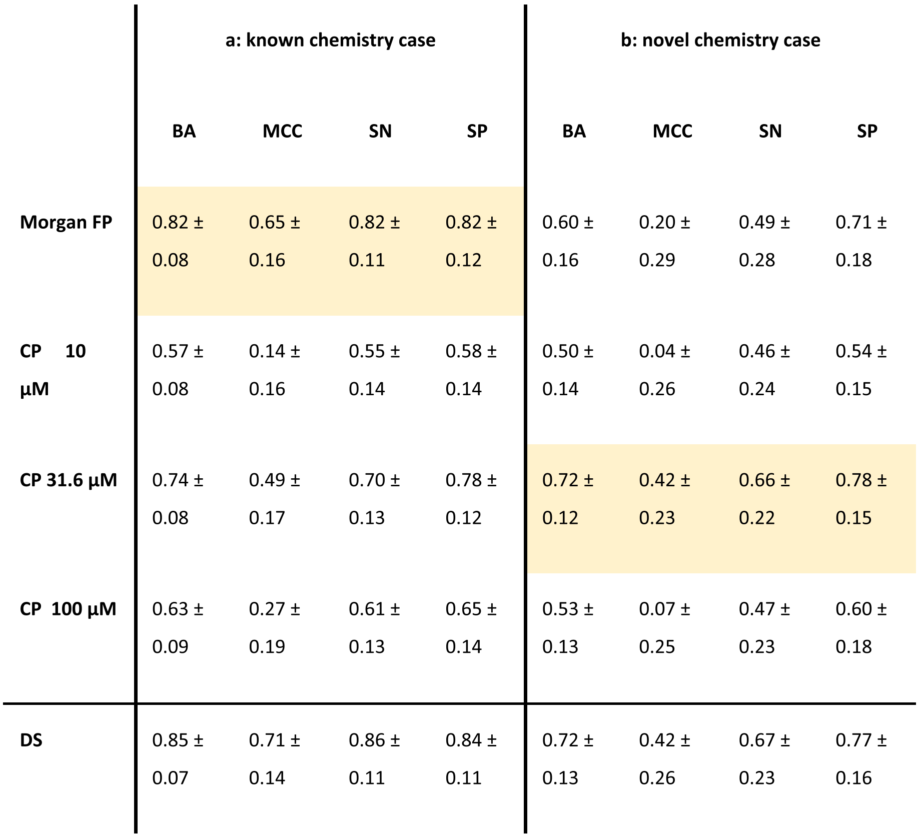
## ■ RESULTS

The objective of our analysis was to assess the predictive efficacy of read-across approaches for acute oral toxicity in rats comparing the use of chemical structural information, in vitro biological data derived from the Cell Painting profiling assay on U2OS cells, or the combination of both. Two distinct types of inputs were utilized to construct KNN models that classify compounds as Very acutely oral toxic (VAOT) or Non very acutely oral toxic (NVAOT).

Initially, we categorized 630 Bayer Crop Science (BCS) agrochemical compounds with known acute rat toxicity results using the public QSAR model CATMoS.[5] Additionally, we used this specific unbalanced set of 630 compounds to build a structure similarity-based classifier. Subsequently, KNN classifiers were used on a reduced but balanced set of 226 compounds using either the chemical structures or their morphological profiles in U2OS cells. A comprehensive analysis of both the chemical space (represented by the chemical structures) and biological space (revealed by the U2OS morphological profiles) was conducted to enhance our comprehension of the classifier results. Finally, we investigated whether combining the predictions of the two classifiers could enhance the accuracy of the predictions.

The analysis demonstrated that a simple read-across approach based on chemical structure information and biological data from the Cell Painting profiling assay on U2OS cells can be used to predict acute oral toxicity, even in the context of new chemical space exploration.

Table 3. Mean and Standard Deviations of Four Metrics: Balanced Accuracy (BA), Matthew's Correlation Coefficient (MCC), Sensitivity (SN), and Specificity (SP)[a]



| | a: known chemistry case | | | | b: novel chemistry case | | | |
|---|---|---|---|---|---|---|---|---|
| | BA | MCC | SN | SP | BA | MCC | SN | SP |
| Morgan FP | 0.82 ± 0.08 | 0.65 ± 0.16 | 0.82 ± 0.11 | 0.82 ± 0.12 | 0.60 ± 0.16 | 0.20 ± 0.29 | 0.49 ± 0.28 | 0.71 ± 0.18 |
| CP 10 µM | 0.57 ± 0.08 | 0.14 ± 0.16 | 0.55 ± 0.14 | 0.58 ± 0.14 | 0.50 ± 0.14 | 0.04 ± 0.26 | 0.46 ± 0.24 | 0.54 ± 0.15 |
| CP 31.6 µM | 0.74 ± 0.08 | 0.49 ± 0.17 | 0.70 ± 0.13 | 0.78 ± 0.12 | 0.72 ± 0.12 | 0.42 ± 0.23 | 0.66 ± 0.22 | 0.78 ± 0.15 |
| CP 100 µM | 0.63 ± 0.09 | 0.27 ± 0.19 | 0.61 ± 0.13 | 0.65 ± 0.14 | 0.53 ± 0.13 | 0.07 ± 0.25 | 0.47 ± 0.23 | 0.60 ± 0.18 |
| DS | 0.85 ± 0.07 | 0.71 ± 0.14 | 0.86 ± 0.11 | 0.84 ± 0.11 | 0.72 ± 0.13 | 0.42 ± 0.26 | 0.67 ± 0.23 | 0.77 ± 0.16 |

[a] Four different input data were used to classify compounds as VAOT and NVAOT: chemical structural data (Morgan FP) and Cell Painting (CP) morphological profiles of U2OS cells exposed to chemicals at three concentrations (MP 10 $\mu$M, MP 31.6 $\mu$M, and MP 100 $\mu$M). Orange highlights are the best average metric. The decision support (DS) model performance (combining the Morgan Fingerprint and the morphological profile 31.6 $\mu$M classifier predictions, supplemented with synthetic examples) is shown in the last row. Section (a) reports the performance of the binary classifiers, over the 100 splits of the "known chemistry" case where training and testing sets are split randomly, not considering chemical structure similarities. Section (b) reports the performance of the KNN classifiers, over the 99 valid splits of the "novel chemistry" case where training and testing sets are split to have structurally different compounds over the two sets

**Results of the QSAR Classifiers.** Initially, CATMoS was employed to classify BCS compounds as either VAOT or NVAOT. We used the Opera CATMoS implementation of the model on the "QSAR only compounds set" (Figure 1 a), with 630 compounds as an external test set. The CATMoS predictions were mapped to the two classes, with the majority of compounds being classified as NVAOT. Specifically, 5 of the 109 VAOT compounds, and 514 of the 521 NVAOT compounds were correctly predicted, resulting in a low sensitivity of 0.05, a high specificity of 0.99, a balanced accuracy of 0.52, and an MCC of 0.09 (Table 2a, Table S5). This outcome may be attributed to the fact that the CATMoS QSAR model was not trained on Bayer chemistry, but on mostly publicly available industrial chemical compounds. This indicates a possible mismatch in the applicability domain of the model for BCS chemistry.

For the predictions of the set of 630 BCS compounds, we could determine that most compounds, 628 (99,6%) were within the CATMoS Applicability Domain (Table 2b). Most of them, 448 (71%), had an Applicability Domain index below 0.6, suggesting that the predictions should be considered with caution (Table 2c). The remaining 180 (29%) predictions having an Applicability Domain index above 0.6, displayed an average confidence level of 0.57, indicating a relatively low level of confidence in the predictions (Table 2b).

Subsequently, we developed a classifier, based on our 630 compound set, using a KNN classifier on Morgan fingerprints of chemical compounds. The classifier's performance was evaluated through cross-validation with two data holdout strategies: the "known chemistry" case, where compounds from the test sets resemble those from the training sets, and the "novel chemistry" case, where compounds from the test sets differ structurally from those in the training sets.

In the "known chemistry" case, the classifiers achieved an average balanced accuracy of 0.81, a sensitivity of 0.70, a specificity of 0.92, and a MCC of 0.61 (Table 2c).

In the "novel chemistry" case, out of 100 theoretical cross-validation splits, only 44 of the 100 theoretical cross-validation splits included the two classes (VAOT and NVAOT) in both the training and testing sets, making them valid. The classifier demonstrated an average balanced accuracy of 0.61, a sensitivity of 0.31, a specificity of 0.92, and a MCC of 0.17 (Table 2c).

The "novel chemistry" case demonstrated that chemical structure-based classifiers perform less well when classifying compounds that are structurally distinct from those in the training set. In summary, the chemical structure similarity-based models demonstrated good performance in handling known

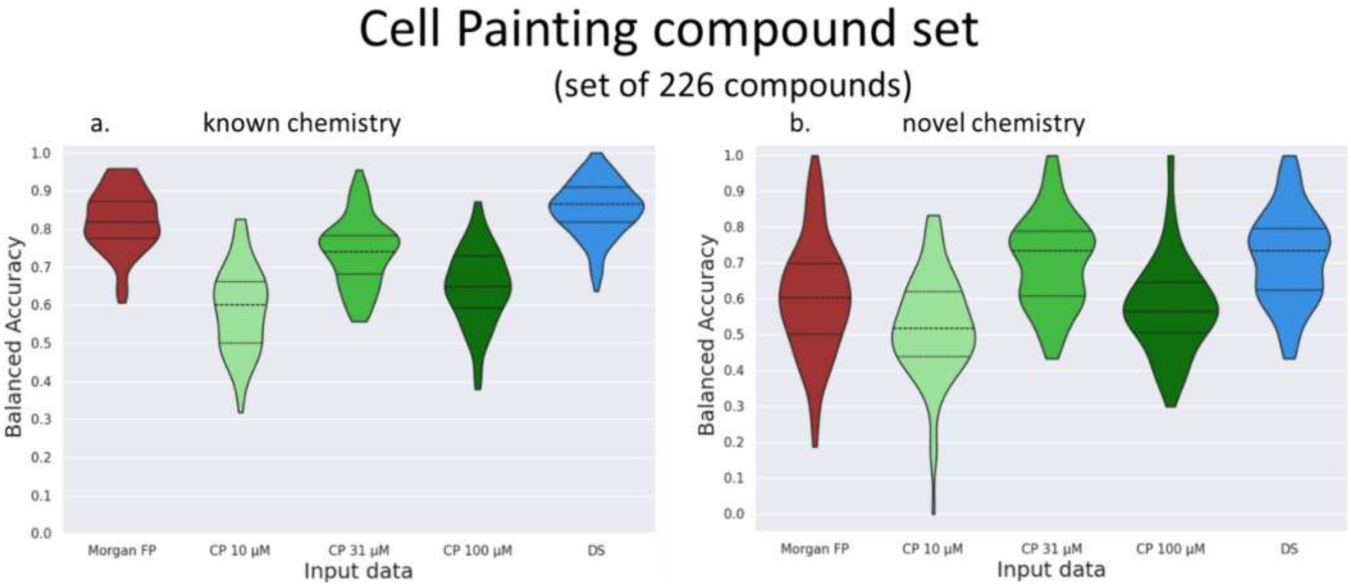


**Figure 4.** (a) Violin plots representing the balanced accuracies of the binary classifier for the 10 × 10-fold cross validation splits not considering the structure similarities (known chemical case). (b) Violin plots representing the balanced accuracies of the KNN binary classifier for 99 valid splits of the 10 × 10-fold cross validation that put in the testing set chemical structurally different from the training set (novel chemical case). Legend: in red, the classifier using the Morgan fingerprint; in light green, the classifier using the morphological profiles at 10 $\mu$M; in midgreen, the classifier using the morphological profiles at 31.6 $\mu$M; in dark green, the classifier using the morphological profiles at 100 $\mu$M; in blue, the decision support (DS) model. Inside each violin plot, the quartiles are indicated as dash lines.

chemistry. However, as expected by the design of the "novel chemistry" case, they exhibited a decrease in performance when confronted with unfamiliar chemical structures. This limitation becomes apparent when exploring new areas of chemical space. To address this limitation, we leveraged the biological effects of the chemical compounds for predictions. The subsequent section outlines our approach, employing a Cell Painting assay on U2OS cells to capture the biological effects of the chemical compounds.

**Comparison of Chemical Structure and Cell Painting Morphological-Based Classifiers.** The objective of our study was to compare two inputs for predicting acute oral toxicity, utilizing a data set called the "Cell Painting set", a subset of the 630 "QSAR only compound set", augmented with additional public chemical compounds. The Cell Painting set included a total of 226 compounds (Figure 1b). KNN classifiers were trained using both types of input and employing two data holdout strategies: "known chemistry" and "novel chemistry" cases.

Similarly to the previous chemical structure similarity-based classifier on the QSAR only compound set (630 compounds), KNN classifiers were trained on the Morgan fingerprints of the molecules.

For classifiers based on the morphological profiles obtained from Cell Painting, consensus profiles were utilized after normalization, unsupervised feature selection, and replicate profile aggregation at the treatment level. Regarding the chemical structure similarity-based classifiers, we used the KNN algorithm. Classifiers were built for each tested concentration (10, 31.6, and 100 $\mu$M).

**Results in the "Known Chemistry" Case.** In the "known chemistry" case, the chemical structure similarity-based classifier demonstrated superior performance compared to other classifiers, achieving a mean balanced accuracy of 0.82. This was followed by the 31.6 $\mu$M morphological profile classifier, with a mean balanced accuracy of 0.74 (Table 3a). The two other morphological profile classifiers at 10 and 100 $\mu$M demonstrated lower performance (Table 3a).

The distribution of the balance accuracy values over the 100 splits for each input type showed a narrow range (Figure 4a), which was confirmed by low standard deviations ranging from 0.08 to 0.09 (Table 3a).

A statistical analysis using Nadeau and Bengio's corrected *t* test to compare the balance accuracy values over the 100 splits of the two top classifiers indicated that the chemical structure similarity-based classifier significantly outperformed the 31.6 $\mu$M morphological profile classifier ($p$-value = 0.05).

**Results in the "Novel Chemistry" Case.** In the "novel chemistry" case, the 31.6 $\mu$M morphological profile classifier demonstrated superior performance, achieving a mean balanced accuracy of 0.72. This was followed by the chemical structure similarity-based classifier, with a mean balanced accuracy of 0.60 (Table 3b). The remaining two classifiers demonstrated lower performance (Table 3b).

The distributions of the balanced accuracies for each classifier showed that in certain splits, the chemical structure similarity-based classifier encountered challenges in making accurate predictions (Figure 4b). This was also the case, to a lesser extent, for the 31.6 $\mu$M morphological profile-based classifiers (Figure 4b).

The Nadeau and Bengio's corrected *t* test indicated that the 31.6 $\mu$M morphological profile classifier significantly outperformed the chemical structure similarity-based classifier ($p$-value = 0.045).

In summary, for the chemical structure similarity-based classifiers, we reproduced the results of the previous chemical structure similarity-based classifier, which was trained on approximately three times more compounds (630 compounds) with good performances in the "known chemistry" case, and a decrease in performance in the "novel chemistry" case (the balanced accuracy dropped from 0.82 to 0.60).

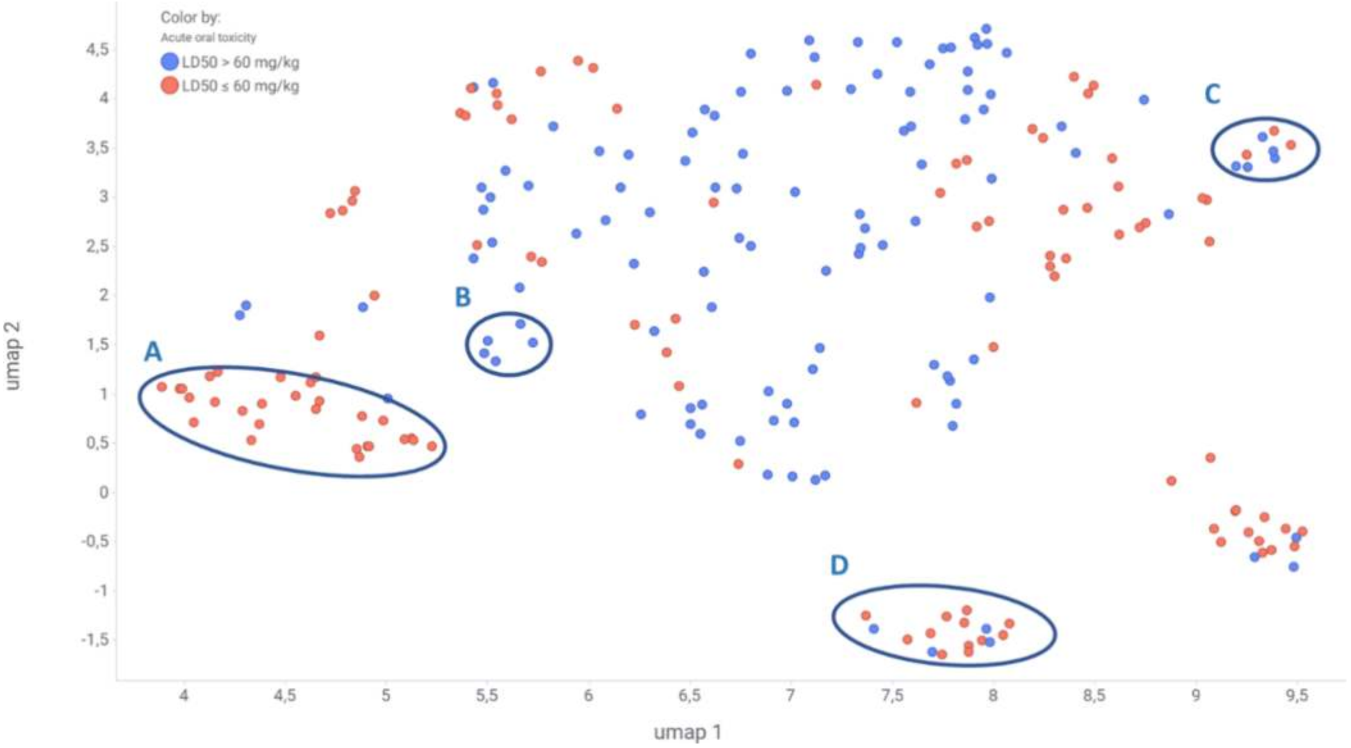


**Figure 5.** Scatter plot of the two-dimensional UMAP embedding of the chemical compound morgan fingerprints. In blue, the chemical compounds that are NVAOT. In red, chemical compounds that are VAOT. Four clusters of compounds are designated by the letters A, B, C, and D. Cluster A is an example of a cluster with only VAOT compound. Cluster B is an example of a cluster with only NVAOT compounds. B and C are two examples of clusters with a mix of VAOT and NVAOT compounds.

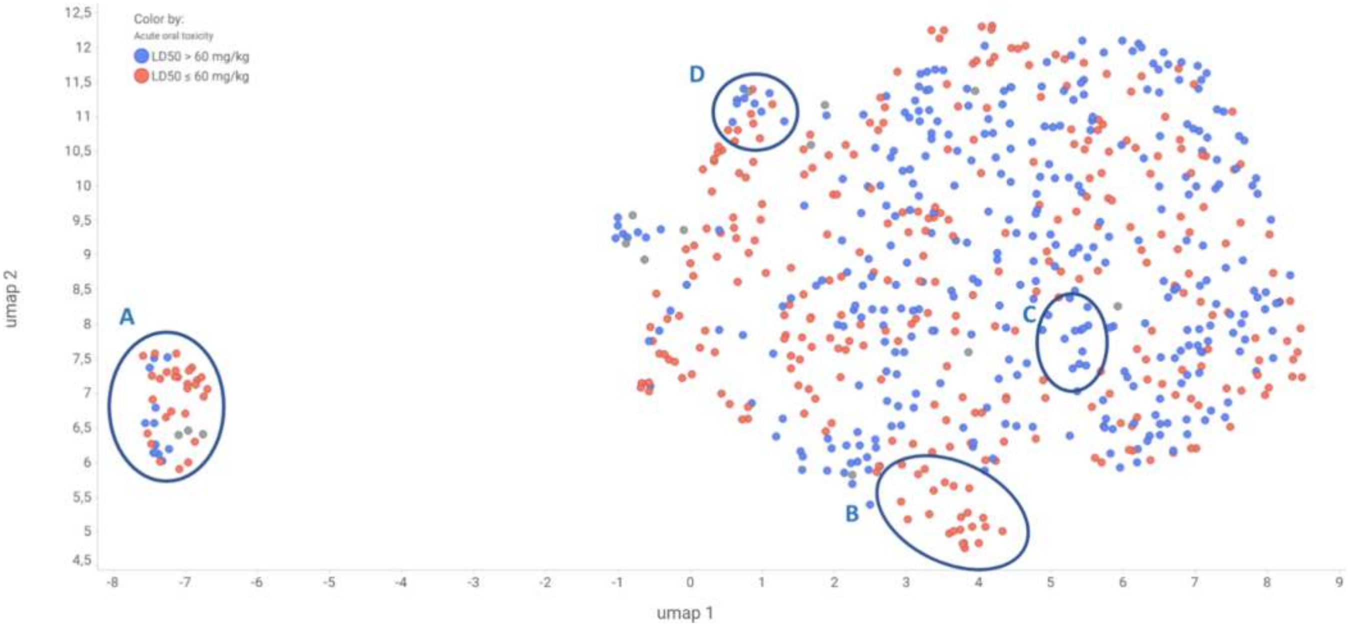


**Figure 6.** Two-dimensional representation of the consensus morphological profile similarities for all treatments and three concentrations, using uniform manifold approximation (UMAP) embedding on two components with the Pearson correlation-based similarity measure. In red, the morphological profiles of U2OS cells are perturbed by VAOT compounds. In blue, the morphological profiles of U2OS cells perturbated by NVAOT compounds. Four groups of compounds are designated by the letters A, B, C, and D. The group A of compounds corresponds to treatment with very low cell counts. The group B of compounds is an example of grouping with a large number of VAOT compounds. The group C of compounds is an example of grouping with a high number of NVAOT compounds. The group D of compounds is an example of grouping with a mixture of VAOT and NVAOT compounds.

Overall, our findings emphasize the superior performance of the chemical structure similarity-based classifier in the "known chemistry" case. However, the morphological profile-based classifier remains a valuable tool, particularly in the "novel chemistry" case, where the classifiers based on the 31.6 $\mu$M

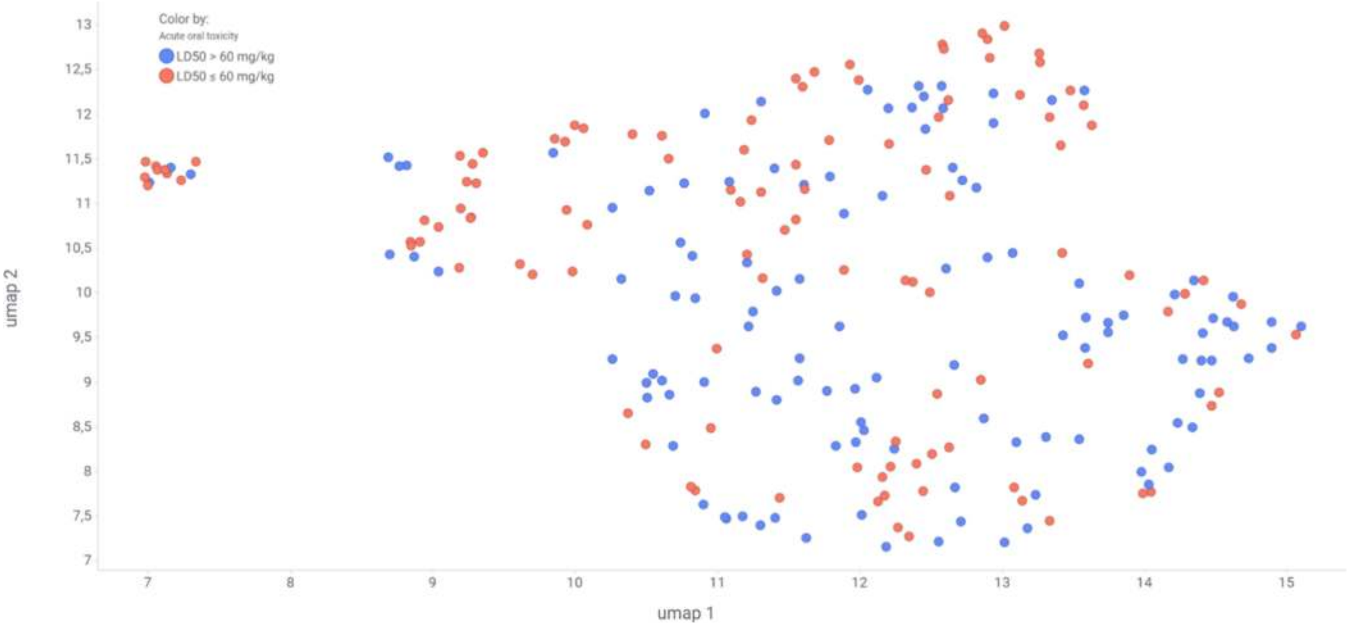


**Figure 7.** Two-dimensional representation of the consensus morphological profile similarities for all treatments at 31.6 $\mu$M, using uniform manifold approximation (UMAP) embedding on two components with the Pearson correlation-based similarity measure. In red, the morphological profiles of U2OS cells perturbated by VAOT compounds. In blue, the morphological profiles of U2OS cells perturbated by NVAOT compounds.

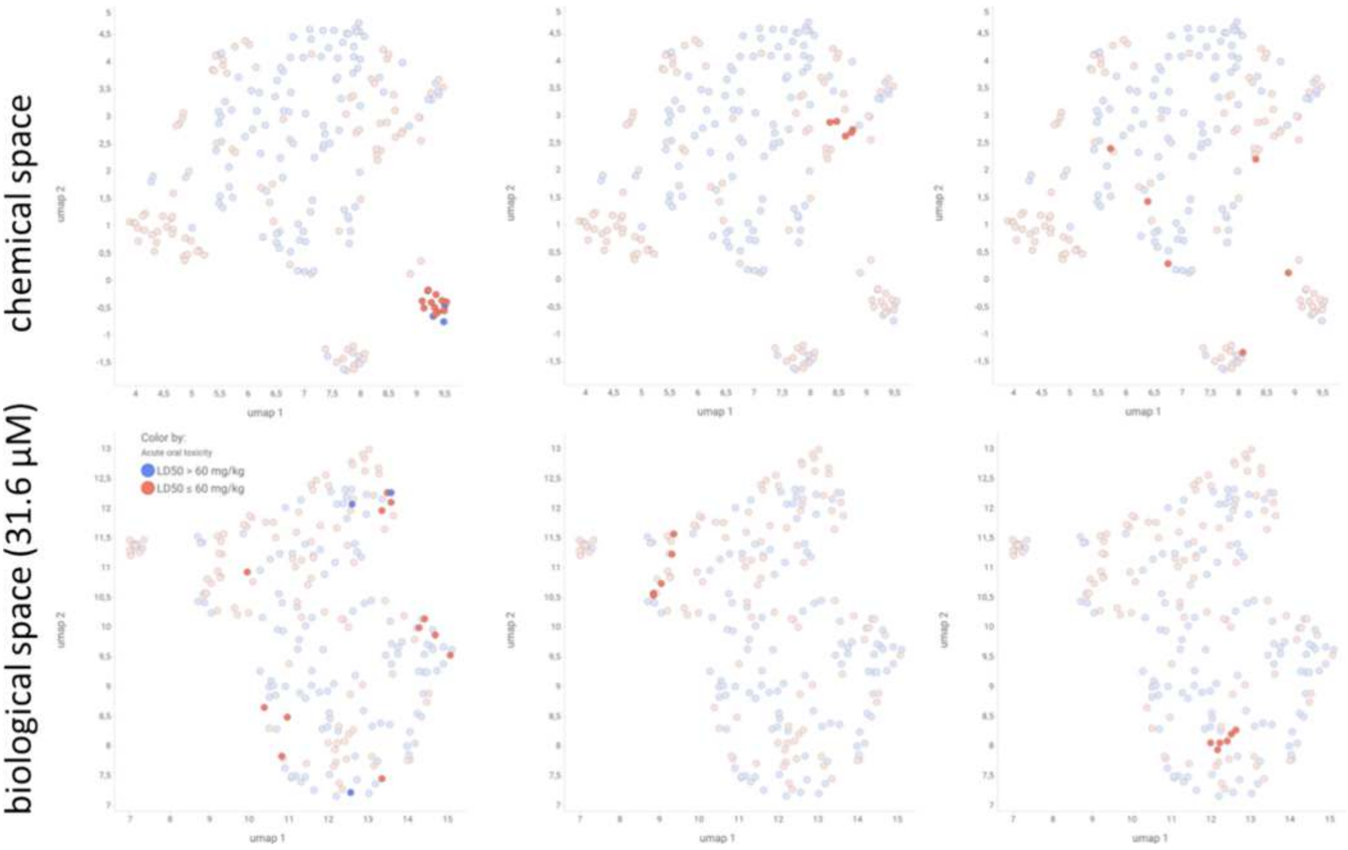


**Figure 8.** (top) Chemical space. Scatter plot of the two-dimensional UMAP embedding of the chemical compound morgan fingerprints. (bottom) Biological space. Scatter plot of the two-dimensional UMAP embedding of the Cell Painting morphological profiles of the chemical compounds at 31.6 $\mu$M. In red, the morphological profiles of U2OS cells perturbated by VAOT compounds; in blue, the morphological profiles of U2OS cells perturbated by NVAOT compounds. On each column, the chemical profiles and morphological profiles of the same compounds are selected. (left) Example of structurally similar compounds, inducing different morphological profiles. (center) Example of structurally similar compounds (carbamates), inducing similar morphological profiles. (right) Example of structurally different compounds, inducing similar morphological profiles.

morphological profile demonstrated the highest performance (balanced accuracy of 0.72).

### Performances of the Classifiers. Comparison of the Chemical and Biological Spaces.

To gain insight into the performance of the classifiers based on the chemical structures and the morphological profiles, we investigated the chemical and biological spaces for the Cell Painting set of compounds.

**Chemical Space.** Our data set comprises 226 compounds primarily originating from Bayer Crop Science chemistry and supplemented with 29 public compounds. The Butina algorithm, using the Tanimoto distance and a threshold of 0.7, identified 91 clusters. Of these, 61 were represented by a unique compound indicating structural diversity.

By plotting the similarity of the structures on a UMAP (Figure 5), using Morgan fingerprints and the Tanimoto distance, it is possible to visually identify distinct clusters. Notably, certain clusters exclusively comprised VAOT compounds (e.g., cluster A), while others were exclusively comprised of NVAOT compounds (e.g., cluster B). Additionally, several clusters contained a mixture of both (e.g., clusters C and D).

In summary, the chemical space exhibited diversity in the form of different clusters. Specific areas demonstrated a prevalence of either VAOT or NVAOT compounds, while others presented a combination of both classes.

**Biological Space.** Figure 6 illustrates the degree of similarity in the biological response of compounds, as measured by Pearson correlation-based similarities on a UMAP plot. In contrast to the chemical space, a limited number of clusters visually emerged with only two notable clusters observed. An isolated small cluster (cluster A) was clearly separated from the other profiles, and upon inspection, these profiles corresponded to instances with a notably low cell count.

The second cluster exhibited diverse areas: including regions with a high number of VAOT (e.g., group B), NVAOT compounds (e.g., group C), and areas with a mix of both classes (e.g., grouping D).

In Figure 7, we focused on the 31.6 $\mu$M concentration, which yielded optimal performance for the classifier using morphological profiles. Similar observations were made, with an isolated cluster corresponding to profiles with a very low number of cells. Additional distinct areas emerged, showcasing regions with a high number of VAOT or NVAOT compounds, as well as areas with a mixed representation of both classes.

**Comparison of the Chemical and Biological Spaces.** We compared different groups of compound structures and groups of morphological profiles to better understand the chemical and biological space interrelationship. Notably, we observed that chemicals clustering together in the chemical space could elicit a diversity of biological responses, emphasizing that structurally similar compounds may manifest distinct biological responses (Figure 8, left column). We illustrate a specific case involving a group of chemicals, the carbamates, that induce similar morphologies in U2OS cells (Figure 8, center column). Interestingly, similar morphological profiles induced by structurally different compounds could also be observed (Figure 8, right column). This comparison illustrates that (1) the biological effects of structurally similar molecules may not necessarily be identical, (2) structurally similar compounds could trigger different biological effects, and (3) conversely, compounds with different structures could result in comparable biological responses.

**Biological Response.** To analyze the biological response of U2OS cells to chemical compound perturbations, we employed morphological profiles, using two metrics: the grit score[23] and the number of cells. The analysis was conducted to better understand how U2OS cells reacted to our set of compounds, which, in turn, helped us to understand the results of the classifiers.

**Grit Score.** The grit score indicated the extent to which the average morphology of U2OS cells perturbed by a compound deviated from the average negative control morphology of nonperturbed U2OS cells. A high grit score indicated a cell morphology that was more distinct from that of the negative controls. For example, the average grit score of the positive controls was 4.8.

A Mann−Whitney U rank test on the grit scores, for the two compound groups, VAOT and NVAOT, demonstrated that VAOT compounds elicited a marginal though significantly stronger biological response compared to NVAOT (grit scores of 3 and 2.5, respectively, $p$-value of 0.004) (Table 4).

**Table 4. Average Grit Score and Number of Cytotoxic Treatments for Different Groups of Profiles: VAOT (Very Acutely Oral Toxic Compounds), NVAOT (Not Very Acutely Oral Toxic Compounds), 10 $\mu$M Treatment Profiles, 31.6 $\mu$M Treatment Profiles, and 100 $\mu$M Treatment Profiles**

| profiles | average grit score | number of cytotoxic treatments |
|---|---|---|
| negative control | NA | 0 |
| VAOT | 3 | 28 |
| NVAOT | 2.5 | 16 |
| 10 $\mu$M all | 1.9 | 11 |
| 10 $\mu$M-VAOT | 2.1 | 8 |
| 10 $\mu$M-NVAOT | 1.64 | 3 |
| 31.6 $\mu$M all | 2.6 | 22 |
| 31.6 $\mu$M-VAOT | 2.98 | 14 |
| 31.6 $\mu$M-NVAOT | 2.17 | 8 |
| 100 $\mu$M all | 3.7 | 44 |
| 100 $\mu$M-VAOT | 4.01 | 27 |
| 100 $\mu$M-NVAOT | 3.20 | 17 |

Regarding concentrations, on average, the 10 $\mu$M treatments had a grit score of 1.9, the 31.6 $\mu$M had a grit score of 2.6, and the 100 $\mu$M treatment had a grit score of 3.7. This aligns with our assumption that higher concentrations lead to increased biological responses, a consideration made when designing the Cell Painting campaign with three concentrations (Table 4).

Identifying compounds with no induced morphological changes, we set a grit score threshold of 1. Below this threshold, we considered a treatment that did not induce any morphological change. Of the 23 compounds falling below this threshold, 6 were VAOT.

**Number of Cells.** An additional output of the image analysis was the number of cells per well. For this analysis, the number of cells was not normalized, and the median number of cells per well for a given treatment was computed. The average number of cells for the negative controls was 2231. We arbitrarily set the number of cells that defines cytotoxicity as a cell count below 50% of the average negative control cell count, meaning a cell count per well of below 1115 defined a cytotoxic treatment.

In total, 44 compounds exhibited cytotoxicity: 11 compounds at 10 $\mu$M, 22 compounds at 31.6 $\mu$M, and 44 compounds at 100 $\mu$M (Table 4).

Categorizing by class, 28 VAOT compounds (25%) and 16 NVAOT compounds (14%) displayed cytotoxicity at least one concentration. A chi-square test of independence of variables with the null hypothesis that the number of cytotoxic compounds is independent of the class (VAOT and NVAOT) gave a $p$-value of 0.1. We could not conclude that there was a higher percentage of cytotoxicity for VAOT compounds (Table 4).

**Results of the Decision-Support Model.** The decision-support model aided the decision when the two KNN classifiers did not predict the same class. The model combined four pieces of information: predictions from the KNN classifier based on the chemical structure information and predictions from the KNN classifier based on the morphological profiles and distances to the nearest neighbor in each classifier.

In the "known chemistry" case, the model demonstrated an average balanced accuracy of 0.84, slightly above the chemical structure similarity-based classifier's average balanced accuracy of 0.82 but was not significantly different (Nadeau and Bengio's corrected $t$ test $p$-value of 0.47). In the "novel chemistry" case, the model had on average a balanced accuracy of 0.65, below the 31.6 $\mu$M morphological profile classifier's average balanced accuracy of 0.72.

To understand why in the "novel chemistry" case this model did not yield better performances, we computed the mean Morgan fingerprint Tanimoto distances between each chemical compound of the training set and its nearest neighbor in the training set, and the mean distances between each chemical compound of the testing set and its nearest neighbor in the training set.

In the "known chemistry" case, on average, in the training set, each compound has a distance to its nearest neighbor of 0.50 and in the testing set 0.48. For the "novel chemistry" case, on average, in the training set, each compound had a distance to its nearest neighbor of 0.49, and in the testing set 0.73.

In the "novel chemistry" case, the training set did not have enough examples of distant chemical structures. To help the model, we added synthetic examples of distant chemical structures in the training set. To do so, we subset in each training set the cases where the predictions of the chemical structure similarity-based did not match the real class, and we updated the distances of the nearest neighbors with a random number between 0.7 and 0.9 and added those as synthetic examples in the data set used to train the model.

By applying this approach, in the "known chemistry" case the model achieved an average balanced accuracy of 0.85, slightly above the chemical structure similarity-based classifier's average balanced accuracy of 0.82 but was not significantly different (Nadeau and Bengio's corrected $t$ test $p$-value of 0.12). In the "novel chemistry" case, the model achieved an average balanced accuracy of 0.72 (Table 3).

## ■ DISCUSSION

Our results showed that the classification of compounds as very acute oral toxic or not, using a similarity-based approach, was possible using chemical structure information, morphological profiles of U2OS cells, or the combination of both. When classifying compounds coming from the same chemical space as those in the classifier's training set, the chemical structure information was more predictive. Conversely, when the compounds to be classified came from a different chemical space than those in the classifier's training set, the morphological profiles of the U2OS cells were more predictive.

Initial attempts to use the publicly available QSAR model, CATMoS, for the prediction of acute oral toxicity on a set of 630 Bayer compounds did not yield good predictions.[5] The CATMoS performance is hindered, as Bayer Crop Science chemistry could be considered to be locally outside its training set (Table 2b). Although almost all compounds were globally within the CATMoS applicability domain, most resided in gaps of the training chemical space. It is well-known that QSAR models excel when the compounds to be classified fall within the applicability domain of the models, and can perform poorly when they do not.[35] In summary, CATMoS, which is a QSAR model trained on more than 10,000 compounds, has very good performance for the prediction of acute oral toxicity of those chemicals, but it does not work as effectively with the structurally diverse BCS chemistry.

To test our hypothesis, we trained a simple KNN classifier, resembling a similarity-based (or read-across) approach, on this set of 630 BCS compounds, using their chemical structure information. Working with two data-holdout strategies to simulate scenarios within and outside the applicability domain of a QSAR model, we evaluated our classifiers under two conditions: the "known chemistry" case, simulating scenarios within applicability domain case, and the "novel chemistry" case, attempting to simulate outside applicability domain case. In the "known chemistry" case, our classifier exhibited strong performance, comparable to CATMoS:CATMoS achieved a balanced accuracy of 0. 84, in classifying compounds as very toxic (VT) (LD50 < 50 mg/kg) whereas our classifier had a balanced accuracy of 0.81 (Table 2c) for the classification of compounds as VAOT (LD50 < 60 mg/kg).[5]

However, as designed, the performance of the classifier dropped in the "novel chemistry" case due to the data-holdout strategy, which placed Butina compound clusters, which are not present in the training sets, into the test sets. This effectively simulated scenarios outside of the model applicability domain, although the decrease in balanced accuracy (from 0.82 to 0.60) (Table 3) was not as drastic as that observed with the Bayer CropScience chemistry using CATMoS (from 0.84 to 0.52) (Table 2a).

To overcome this chemical applicability domain limitation, we explored whether using the compound biological effects could mitigate this problem. Compound-induced biological effects characterized by transcriptomics have previously been used to predict target activities, in association and comparison with QSAR models.[36,37] Here, we utilized Cell Painting to generate morphological profiles at a more reasonable cost compared to transcriptomics.

Using a smaller set of 226 compounds, (balanced with 49% of compounds where LD50 < 60 mg/kg; 51% where LD50 > 60 mg/kg), we trained KNN classifiers based on either chemical structure information or U2OS morphological profiles at three concentrations (10, 31.6, or 100 $\mu$M).

Morphological profiles at 31.6 $\mu$M concentration demonstrated better performance, compared to the other concentrations, in both the "known chemistry" and "novel chemistry" cases. With a balanced accuracy of 0.74 (Table 3) in the "known chemistry" case and 0.72 (Table 3) in the "novel chemistry" case, Cell Painting U2OS profiles demonstrated the ability to predict acute oral toxicity classes, interestingly, independent of the structural similarity of the tested compounds.

Cell Painting can indeed identify morphological patterns associated with specific mode of action (MoA) and molecular initiation event (MIE) of compounds.[38,39] Typically, acute toxicity involves a limited number of MIEs.[40] such as narcosis (activity at the lipid bilayer of the membrane), acetylcholinesterase inhibition, ion channel modulators and inhibitors of cellular respiration.[41] The Cell Painting experiment revealed morphological profiles (initiated by MIE) associated with acute oral toxicity, as evidenced by the grouping of the morphological profiles associated with VAOT compounds. For example, four carbamates (promecarb, methiocarb, propoxur, *m*-cumenyl methylcarbamate) known as acetylcholinesterase inhibitors, produced similar morphological profiles in U2OS cells (Figure 8, middle row). This partially explains why the morphological profile-based classifiers were able to correctly classify the compounds. In the "known chemistry" case, the performance of the classifiers based on the 31.6 $\mu$M morphological profile did not surpass the classifiers based on the chemical structure information but outperformed them in the "novel chemistry" case.

The comparison of the two chemical structure-based KNNs for the two data sets showed similar performances, but this comparison is limited because the two data sets are different. The QSAR data set consisting of 630 compounds is unbalanced (109 VAOT, 521 NVAOT), while the Cell Painting data set, made of 226 compounds, is balanced, but of smaller size.

Capturing the biological effects of compounds had limitations: the limitation of the cell system to reveal the effects causally related to acute toxicity, together with technical limitations of the laboratory experiment itself.

For the limitation of the cell system, we observed, by grit score analysis, that not all of the compounds induced a biological response in U2OS cells (10% of the tested compounds), regardless of the concentration used. Six known VAOT compounds did not elicit any morphological changes compared to those of the negative controls. Among these compounds, five were public compounds and some of them contained information on their possible mechanism of action. The warfarin, a vitamin K antagonist, and methamidophos, a potent acetylcholinesterase inhibitor, did not induce any biological response in U2OS cells. This suggests that U2OS cells have their own biological applicability domain and may not capture all of the bioactivities associated with acute oral toxicity observed in a whole organism such as a rat in our case study. Nevertheless, for our set of compounds, Cell Painting on U2OS managed to capture bioactivities for most of the VAOT compounds.

On the contrary, when analyzing the number of cells, we could also identify a limitation due to the cytotoxicity of the compounds: 44 compounds showed cytotoxicity at least at one concentration, and 12 exhibited cytotoxicity even at the lowest concentration. The morphological profiles of the cytotoxic treatments were not informative, as they consisted mainly of debris and dying cells. It appears that the 31.6 $\mu$M concentration represents a good compromise between inducing bioactivity and avoiding cytotoxicity. However, building a model using only this concentration is a limitation as morphological profiles coming from other concentrations could also have been associated with acute oral toxicity. Different rules for selecting the best concentration per compound, using the grit score to ensure that the compound was active or the number of cells to ensure that the compound was not cytotoxic, did not produce better models.

For the limitations of the experiments, several quality issues can arise when conducting an experiment in a laboratory. Experiments are technically demanding and are prone to variability and error. Seeding variability can affect the cell morphologies and thus the morphological profiles. There are other common problems that can occur in laboratory experiments, such as treatment errors, compounds with low purity, or precipitation at high concentrations. These problems can affect the quality of the morphological profiles and thus the performance of a classifier based on morphological profile similarities.

The chemical structural information did not suffer from these limitations because this information was not subject to quality issues, was not cell system dependent, and was not assay design dependent. This information was intrinsic to the description of a given compound. This may partly explain why chemical structure similarity-based classifiers, in the "known chemistry" case, performed better than biological-based classifiers: the full structural information is available, whereas the biological information is partially available and subject to quality issues, in particular reproducibility.

For both types of input data, the optimal number of neighbors for the KNN algorithm (Supporting Information, S4) was 1, indicating that few examples of identified profiles leading to high acute oral toxicity or not were present in the data set. A larger set of compounds, such as the CATMoS training set, would help to identify more examples of cell painting profiles associated with acute oral toxicity.

In the "novel chemistry" case, the morphological profile-based classifiers did not experience as large a performance drop as the chemical structure similarity-based classifiers, suggesting that the biological space did not cluster in the same way as the chemical space. This indicates that similar compounds did not consistently induce the same response in U2OS cells (for example due to activity cliffs) and vice versa. The presence of different Butina clusters in the training and test sets did not necessarily result in different morphological profiles, explaining why the morphological profile-based classifier performance did not drop drastically in the "novel chemistry" case.

The use of biological responses of compounds could also be an advantage with respect to enantiomers. The Morgan fingerprint used in this analysis does not take the chirality into account. Enantiomers may have different acute oral toxicity, and the classifier based on chemical structure will not distinguish between them, where morphological profiles may be different.

The decision support model combined both predictions along with the nearest neighbor distances to make the final predictions, slightly improving the classification performance in the "known chemistry" case but decreasing in the "novel chemistry" case. By adding a few synthetic examples in each training set with higher distances in the chemical spaces, it was possible to increase the classification accuracy in the "known chemistry" case, but not in the "novel chemistry" case, where the model performed like the 31.6 $\mu$M morphological-based classifier. Notably, in the "novel chemistry" case, the classifier preferred the predictions of the 31.6 $\mu$M morphological-based classifier predictions over the predictions of the chemical structure similarity-based classifier.

Further results could extend and refine these findings by employing a broader set of compounds covering additional molecular initiating events (MIE) associated with acute oral toxicity. Additionally, a larger set of compounds could facilitate the identification of additional morphological profiles associated with acute oral toxicity. A larger data set would also allow the isolation of a set of compounds as an external data set to further evaluate the performance of the classifier.

The choice of the KNN algorithm in this analysis was deliberate due to its simplicity and similarity to the read-across approach commonly used in toxicology. Since the amount of in vivo data is often limited, the read-across approach is often the only analysis that can be performed. For chemical structure-based classifiers, other algorithms yielded similar performances (Supporting Information).

To support the creation of public QSAR models with a wider applicability domain, representations of compound structures and results of acute toxicity studies for early candidates that failed to be placed on the markets could be shared by companies and organizations to expand the chemical space coverage.

In addition, in this analysis, the Morgan fingerprint was the only computed chemical fingerprint. The use of additional fingerprints or descriptors could help to achieve better QSAR and chemical structure similarity-based classifier performance.

Similarly, hand-crafted morphological features were used in this analysis. To capture a broader representation of morphological profiles, other deep learning-based representations could also be tested.[42]

The decision support model uses the nearest neighbor distance to decide which prediction to select. Other metrics, such as the "distance to model", which is used to estimate a prediction uncertainty could also be used to decide on which prediction to use.[43]

We have also seen the limitation of the U2OS cell line with not capturing all of the bioactivities of the compounds. We could assume that trying different cell lines could allow capturing more bioactivities linked to MIE leading to acute oral toxicity. Several cell lines have already been used with Cell Painting[44,45] and could help define a set of cell lines capable of capturing a maximum, if not all, MIE leading to acute oral toxicity.

Finally, absorption, distribution, metabolism, and excretion (ADME) properties of compounds were not taken into consideration in this study, but incorporating such data could enhance predictive models. We tried to use predicted maximum concentration in plasma and AUC from a predictive model,[46] but this information did not improve our results (data not shown).

In addition, preincubation of the compounds with liver S9 fractions (the 9000 g supernatant of a liver homogenate), containing phase I and II metabolic enzymes, to generate the possible metabolites of a parent compound, could be helpful when the toxicity is driven by a metabolite, as it is done for example in the Ames assay to test the mutagenic potential of chemical compounds.[47,48]

In conclusion, a combined approach utilizing chemical structure and Cell Painting morphological profiles-based classifiers based on chemical and biological space distances holds promise for predicting acute oral toxicity. These classifiers could be used in the context of early derisking and in the future serve in the context of Next Generation Risk Assessment (NGRA), which aims at refining if not replacing laboratory animal testing.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

GitHub (https://github.com/Bayer-Group/cellPainting_acuteTox) Consensus Cell Painting morphological profiles with oral acute toxicity label, and code are provided.

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemrestox.4c00169.

List of reference compounds, list of public compounds, UMAP of the chemical space with Butina clusters, other classifiers results, detailed CATMoS predictions, and Python package versions (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

**David Rouquié** − *Toxicology Data Science, Bayer SAS Crop Science Division, Sophia Antipolis 06906, France;* orcid.org/0000-0002-7796-7418; Email: david.rouquie@bayer.com

### Authors

**Fabrice Camilleri** − *Toxicology Data Science, Bayer SAS Crop Science Division, Sophia Antipolis 06906, France; I3S UMR 7271 du CNRS, Université Côte d'Azur, Sophia Antipolis 06903, France*

**Joanna M. Wenda** − *Early Toxicology, Bayer SAS Crop Science Division, Sophia Antipolis 06906, France*

**Claire Pecoraro-Mercier** − *Early Toxicology, Bayer SAS Crop Science Division, Sophia Antipolis 06906, France*

**Jean-Paul Comet** − *I3S UMR 7271 du CNRS, Université Côte d'Azur, Sophia Antipolis 06903, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrestox.4c00169

### Author Contributions

CRediT: **Fabrice Camilleri** conceptualization, data curation, formal analysis, writing - original draft; **Joanna M. Wenda** conceptualization, investigation, writing - review & editing; **Claire Pecoraro-Mercier** conceptualization, resources, supervision, writing - review & editing; **Jean Paul Comet** conceptualization, supervision, writing - review & editing; **David Rouquié** conceptualization, formal analysis, methodology, project administration, resources, supervision, writing - review & editing.

### Funding

F.C. holds a doctoral fellowship from the Association Nationale de la Recherche Technique (ANRT CIFRE PhD funding).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors thank Dr. Nina Hallmark and Dr. Oscar Mendes Lucio for their insightful suggestions to improve the article.

## ■ ABBREVIATIONS

AD: applicability domain
BA: balanced accuracy
BCS: Bayer Crop Science
DMSO: dimethyl sulfoxide
DS: decision support
GHS: Global Harmonized System
KNN: K nearest neighbor
LD50: median lethal dose
MAD: median absolute deviation
MCC: Matthews correlation coefficient
MIE: molecular initiative event
MoA: mode of action
NGRA: next generation risk assessment
NVAOT: not very acutely oral toxic (LD50 > 60 mg/kg)
QSAR: quantitative structure activity relationship
SMILES: simplified molecular-input line-entry system
SN: sensitivity

SP: specificity

TP, TN, FP, FN: true positives, true negatives, false positives, false negatives

VAOT: very acutely oral toxic (LD50 ≤ 60 mg/kg)

## ■ REFERENCES

(1) Khalak, Y.; Tresadern, G.; Hahn, D. F.; De Groot, B. L.; Gapsys, V. Chemical Space Exploration with Active Learning and Alchemical Free Energies. *J. Chem. Theory Comput.* **2022**, *18* (10), 6259−6270.

(2) Scannell, J. W.; Bosley, J.; Hickman, J. A.; Dawson, G. R.; Truebel, H.; Ferreira, G. S.; Richards, D.; Treherne, J. M. Predictive Validity in Drug Discovery: What It Is, Why It Matters and How to Improve It. *Nat. Rev. Drug Discovery* **2022**, *21*, 915.

(3) Henriquez, J. E.; Badwaik, V. D.; Bianchi, E.; Chen, W.; Corvaro, M.; LaRocca, J.; Lunsman, T. D.; Zu, C.; Johnson, K. J. From Pipeline to Plant Protection Products: Using New Approach Methodologies (NAMs) in Agrochemical Safety Assessment. *J. Agric. Food Chem.* **2024**, *72* (19), 10710−10724.

(4) Erhirhie, E. O.; Ihekwereme, C. P.; Ilodigwe, E. E. Advances in Acute Toxicity Testing: Strengths, Weaknesses and Regulatory Acceptance. *Interdisciplinary Toxicology* **2018**, *11* (1), 5−12.

(5) Mansouri, K.; Karmaus, A. L.; Fitzpatrick, J.; Patlewicz, G.; Pradeep, P.; Alberga, D.; Alepee, N.; Allen, T. E. H.; Allen, D.; Alves, V. M.; Andrade, C. H.; Auernhammer, T. R.; Ballabio, D.; Bell, S.; Benfenati, E.; Bhattacharya, S.; Bastos, J. V.; Boyd, S.; Brown, J. B.; Capuzzi, S. J.; Chushak, Y.; Ciallella, H.; Clark, A. M.; Consonni, V.; Daga, P. R.; Ekins, S.; Farag, S.; Fedorov, M.; Fourches, D.; Gadaleta, D.; Gao, F.; Gearhart, J. M.; Goh, G.; Goodman, J. M.; Grisoni, F.; Grulke, C. M.; Hartung, T.; Hirn, M.; Karpov, P.; Korotcov, A.; Lavado, G. J.; Lawless, M.; Li, X.; Luechtefeld, T.; Lunghini, F.; Mangiatordi, G. F.; Marcou, G.; Marsh, D.; Martin, T.; Mauri, A.; Muratov, E. N.; Myatt, G. J.; Nguyen, D.-T.; Nicolotti, O.; Note, R.; Pande, P.; Parks, A. K.; Peryea, T.; Polash, A. H.; Rallo, R.; Roncaglioni, A.; Rowlands, C.; Ruiz, P.; Russo, D. P.; Sayed, A.; Sayre, R.; Sheils, T.; Siegel, C.; Silva, A. C.; Simeonov, A.; Sosnin, S.; Southall, N.; Strickland, J.; Tang, Y.; Teppen, B.; Tetko, I. V.; Thomas, D.; Tkachenko, V.; Todeschini, R.; Toma, C.; Tripodi, I.; Trisciuzzi, D.; Tropsha, A.; Varnek, A.; Vukovic, K.; Wang, Z.; Wang, L.; Waters, K. M.; Wedlake, A. J.; Wijeyesakere, S. J.; Wilson, D.; Xiao, Z.; Yang, H.; Zahoranszky-Kohalmi, G.; Zakharov, A. V.; Zhang, F. F.; Zhang, Z.; Zhao, T.; Zhu, H.; Zorn, K. M.; Casey, W.; Kleinstreuer, N. C. CATMoS: Collaborative Acute Toxicity Modeling Suite. *Environ. Health Perspect* **2021**, *129* (4), No. 047013.

(6) Herman, D.; Kańduła, M. M.; Freitas, L. G. A.; Van Dongen, C.; Le Van, T.; Mesens, N.; Jaensch, S.; Gustin, E.; Micholt, L.; Lardeau, C.-H.; Varsakelis, C.; Reumers, J.; Zoffmann, S.; Will, Y.; Peeters, P. J.; Ceulemans, H. Leveraging Cell Painting Images to Expand the Applicability Domain and Actively Improve Deep Learning Quantitative Structure−Activity Relationship Models. *Chem. Res. Toxicol.* **2023**, *36*, 1028.

(7) Bray, M.-A.; Singh, S.; Han, H.; Davis, C. T.; Borgeson, B.; Hartland, C.; Kost-Alimova, M.; Gustafsdottir, S. M.; Gibson, C. C.; Carpenter, A. E. Cell Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes. *Nat. Protoc* **2016**, *11* (9), 1757−1774.

(8) Chandrasekaran, S. N.; Ceulemans, H.; Boyd, J. D.; Carpenter, A. E. Image-Based Profiling for Drug Discovery: Due for a Machine-Learning Upgrade? *Nat. Rev. Drug Discov* **2021**, *20* (2), 145−159.

(9) Lapins, M.; Spjuth, O. Evaluation of Gene Expression and Phenotypic Profiling Data as Quantitative Descriptors for Predicting Drug Targets and Mechanisms of Action. *Biorxiv* **2019**, No. 580654.

(10) Tian, G.; Harrison, P. J.; Sreenivasan, A. P.; Carreras-Puigvert, J.; Spjuth, O. Combining Molecular and Cell Painting Image Data for Mechanism of Action Prediction. *Artificial Intelligence in the Life Sciences* **2023**, 3, No. 100060.

(11) Simm, J.; Klambauer, G.; Arany, A.; Steijaert, M.; Wegner, J. K.; Gustin, E.; Chupakhin, V.; Chong, Y. T.; Vialard, J.; Buijnsters, P.; Velter, I.; Vapirev, A.; Singh, S.; Carpenter, A. E.; Wuyts, R.; Hochreiter, S.; Moreau, Y.; Ceulemans, H. Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. *Cell Chem. Biol.* **2018**, *25* (5), 611−618.e3.

(12) Nyffeler, J.; Willis, C.; Lougee, R.; Richard, A.; Paul-Friedman, K.; Harrill, J. A. Bioactivity Screening of Environmental Chemicals Using Imaging-Based High-Throughput Phenotypic Profiling. *Toxicol. Appl. Pharmacol.* **2020**, *389*, No. 114876.

(13) Garcia De Lomana, M.; Marin Zapata, P. A.; Montanari, F. Predicting the Mitochondrial Toxicity of Small Molecules: Insights from Mechanistic Assays and Cell Painting Data. *Chem. Res. Toxicol.* **2023**, *36* (7), 1107−1120.

(14) Seal, S.; Carreras-Puigvert, J.; Trapotsi, M.-A.; Yang, H.; Spjuth, O.; Bender, A. Integrating Cell Morphology with Gene Expression and Chemical Structure to Aid Mitochondrial Toxicity Detection. *Commun. Biol.* **2022**, *5* (1), 858.

(15) Lejal, V.; Cerisier, N.; Rouquié, D.; Taboureau, O. Assessment of Drug-Induced Liver Injury through Cell Morphology and Gene Expression Analysis. *Chem. Res. Toxicol.* **2023**, *36* (9), 1456−1470.

(16) Kowalski, B. R.; Bender, C. F. K-Nearest Neighbor Classification Rule (Pattern Recognition) Applied to Nuclear Magnetic Resonance Spectral Interpretation. *Anal. Chem.* **1972**, *44* (8), 1405−1411.

(17) Zhu, H. Supporting Read-across Using Biological Data. *ALTEX* **2016**, 167−182.

(18) ChemIDplus, 2023, https://pubchem.ncbi.nlm.nih.gov/source/ChemIDplus.

(19) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (4), 747−750.

(20) Mansouri, K.; Grulke, C. M.; Judson, R. S.; Williams, A. J. OPERA Models for Predicting Physicochemical Properties and Environmental Fate Endpoints. *J. Cheminform* **2018**, *10* (1), 10.

(21) Cimini, B. A.; Chandrasekaran, S. N.; Kost-Alimova, M.; Miller, L.; Goodale, A.; Fritchman, B.; Byrne, P.; Garg, S.; Jamali, N.; Logan, D. J.; Concannon, J. B.; Lardeau, C.-H.; Mouchet, E.; Singh, S.; Shafqat Abbasi, H.; Aspesi, P.; Boyd, J. D.; Gilbert, T.; Gnutt, D.; Hariharan, S.; Hernandez, D.; Hormel, G.; Juhani, K.; Melanson, M.; Mervin, L. H.; Monteverde, T.; Pilling, J. E.; Skepner, A.; Swalley, S. E.; Vrcic, A.; Weisbart, E.; Williams, G.; Yu, S.; Zapiec, B.; Carpenter, A. E. Optimizing the Cell Painting Assay for Image-Based Profiling. *Nat. Protoc* **2023**, *18* (7), 1981−2013.

(22) Stirling, D. R.; Swain-Bowden, M. J.; Lucas, A. M.; Carpenter, A. E.; Cimini, B. A.; Goodman, A. CellProfiler 4: Improvements in Speed, Utility and Usability. *BMC Bioinformatics* **2021**, *22* (1), 433.

(23) Serrano, E.; Chandrasekaran, S. N.; Bunten, D.; Brewer, K. I.; Tomkinson, J.; Kern, R.; Bornholdt, M.; Fleming, S.; Pei, R.; Arevalo, J.; Tsang, H.; Rubinetti, V.; Tromans-Coia, C.; Becker, T.; Weisbart, E.; Bunne, C.; Kalinin, A. A.; Senft, R.; Taylor, S. J.; Jamali, N.; Adeboye, A.; Abbasi, H. S.; Goodman, A.; Caicedo, J. C.; Carpenter, A. E.; Cimini, B. A.; Singh, S.; Way, G. P. Reproducible Image-Based Profiling with Pycytominer. *Arxiv* 2023. .

(24) Benchmarking Grit. *Benchmarking Grit.* https://github.com/broadinstitute/grit-benchmark.

(25) *Cytominer-eval: Evaluating quality of perturbation profiles.* https://github.com/cytomining/cytominereval.

(26) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5* (2), 107−113.

(27) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742−754.

(28) Landrum, G.; Tosco, P.; Kelley, B.; Rodriguez, R.; Cosgrove, D.; Vianello, R.; Sriniker; Gedeck, P.; Jones, G.; NadineSchneider; Kawashima, E.; Nealschneider, D.; Dalke, A.; Swain, M.; Cole, B.; Turk, S.; Savelyev, A.; Vaucher, A.; Wójcikowski, M.; Take, I.; Scalfani, V. F.; Walker, R.; Probst, D.; Ujihara, K.; Pahl, A.; Godin, G.; Lehtivarjo, J.; Bérenger, F.; Bisson, J.; Strets123. Rdkit/Rdkit: 2023_03_3 (Q1 2023) Release, 2023. .

(29) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.;

https://doi.org/10.1021/acs.chemrestox.4c00169

Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(30) Escher, S. E.; Bitsch, A. Read-Across Methodology in Toxicological Risk Assessment. In *Regulatory Toxicology*; Reichl, F.-X.; Schwenk, M., Eds.; Springer-Verlag GmbH: Berlin, Heidelberg, 2021; pp 1−14. .

(31) Seal, S.; Yang, H.; Trapotsi, M.-A.; Singh, S.; Carreras-Puigvert, J.; Spjuth, O.; Bender, A. Merging Bioactivity Predictions from Cell Morphology and Chemical Fingerprint Models Using Similarity to Training Data. *J. Cheminform* **2023**, *15* (1), 56.

(32) Cristianini, N.; Ricci, E. Support Vector Machines. In *Encyclopedia of Algorithms*; Kao, M.-Y., Ed.; Springer US: Boston, MA, 2008; pp 928−932. .

(33) Nadeau, C.; Bengio, Y. Inference for the Generalization Error. *Machine Learning* **2003**, *52* (3), 239−281.

(34) Sainburg, T.; McInnes, L.; Gentner, T. Q. Parametric UMAP Embeddings for Representation and Semisupervised Learning. *Neural Comput.* **2021**, *33* (11), 1−2907.

(35) Kar, S.; Roy, K.; Leszczynski, J. Applicability Domain: A Step Toward Confident Predictions and Decidability for QSAR Modeling. In *Computational Toxicology*; Nicolotti, O., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2018; Vol. *1800*, pp 141−169. .

(36) Baillif, B.; Wichard, J.; Méndez-Lucio, O.; Rouquié, D. Exploring the Use of Compound-Induced Transcriptomic Data Generated From Cell Lines to Predict Compound Activity Toward Molecular Targets. *Front. Chem.* **2020**, *8*, 296.

(37) Moshkov, N.; Becker, T.; Yang, K.; Horvath, P.; Dancik, V.; Wagner, B. K.; Clemons, P. A.; Singh, S.; Carpenter, A. E.; Caicedo, J. C. Predicting Compound Activity from Phenotypic Profiles and Chemical Structures. *Nat. Commun.* **2023**, *14* (1), 1967.

(38) Ljosa, V.; Caie, P. D.; Ter Horst, R.; Sokolnicki, K. L.; Jenkins, E. L.; Daya, S.; Roberts, M. E.; Jones, T. R.; Singh, S.; Genovesio, A.; Clemons, P. A.; Carragher, N. O.; Carpenter, A. E. Comparison of Methods for Image-Based Profiling of Cellular Morphological Responses to Small-Molecule Treatment. *J. Biomol Screen* **2013**, *18* (10), 1321−1329.

(39) Way, G. P.; Natoli, T.; Adeboye, A.; Litichevskiy, L.; Yang, A.; Lu, X.; Caicedo, J. C.; Cimini, B. A.; Karhohs, K.; Logan, D. J.; Rohban, M. H.; Kost-Alimova, M.; Hartland, K.; Bornholdt, M.; Chandrasekaran, S. N.; Haghighi, M.; Weisbart, E.; Singh, S.; Subramanian, A.; Carpenter, A. E. Morphology and Gene Expression Profiling Provide Complementary Information for Mapping Cell State. *Cell Systems* **2022**, *13* (11), 911−923.e9.

(40) Prieto, P. Investigating Cell Type Specific Mechanisms Contributing to Acute Oral Toxicity. *ALTEX* **2019**, *36* (1), 39−64.

(41) Leblanc, G. A. Acute Toxicity. In *A Textbook of Modern Toxicology*; 2004.

(42) Caicedo, J. C.; McQuin, C.; Goodman, A.; Singh, S.; Carpenter, A. E. Weakly Supervised Learning of Single-Cell Feature Embeddings. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Salt Lake City, UT, 2018; pp 9309−9318. .

(43) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Cherkasov, A.; Li, J.; Gramatica, P.; Hansen, K.; Schroeter, T.; Müller, K.-R.; Xi, L.; Liu, H.; Yao, X.; Öberg, T.; Hormozdiari, F.; Dao, P.; Sahinalp, C.; Todeschini, R.; Polishchuk, P.; Artemenko, A.; Kuz'min, V.; Martin, T. M.; Young, D. M.; Fourches, D.; Muratov, E.; Tropsha, A.; Baskin, I.; Horvath, D.; Marcou, G.; Muller, C.; Varnek, A.; Prokopenko, V. V.; Tetko, I. V. Applicability Domains for Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. *J. Chem. Inf. Model.* **2010**, *50* (12), 2094−2111.

(44) Cox, M. J.; Jaensch, S.; Van de Waeter, J.; Cougnaud, L.; Seynaeve, D.; Benalla, S.; Koo, S. J.; Van Den Wyngaert, I.; Neefs, J.-M.; Malkov, D.; Bittremieux, M.; Steemans, M.; Peeters, P. J.; Wegner, J. K.; Ceulemans, H.; Gustin, E.; Chong, Y. T.; Göhlmann, H. W. H. Tales of 1,008 Small Molecules: Phenomic Profiling through Live-Cell Imaging in a Panel of Reporter Cell Lines. *Sci. Rep* **2020**, *10* (1), 13262.

(45) Nyffeler, J. *Phenotypic Profiling for High-Throughput Chemical Bioactivity Screening at the U.S. EPA*. 2020, 11457644 Bytes. .

(46) Schneckener, S.; Grimbs, S.; Hey, J.; Menz, S.; Osmers, M.; Schaper, S.; Hillisch, A.; Göller, A. H. Prediction of Oral Bioavailability in Rats: Transferring Insights from in Vitro Correlations to (Deep) Machine Learning Models Using in Silico Model Outputs and Chemical Structure Parameters. *J. Chem. Inf. Model.* **2019**, *59* (11), 4893−4905.

(47) Hakura, A.; Suzuki, S.; Sawada, S.; Motooka, S.; Satoh, T. An Improvement of the Ames Test Using a Modified Human Liver S9 Preparation. *Journal of Pharmacological and Toxicological Methods* **2001**, *46* (3), 169−172.

(48) Hopperstad, K.; DeGroot, D. E.; Zurlinden, T.; Brinkman, C.; Thomas, R. S.; Deisenroth, C. Chemical Screening in an Estrogen Receptor Transactivation Assay With Metabolic Competence. *Toxicol. Sci.* **2022**, *187* (1), 112−126.