Journal of
Pharmaceutical and BioTech Industry

MDPI

Article

MorphADMET Links Chemical Structure and Cellular Morphology for Context-Dependent ADMET Assessment

Md Wasi Ul Kabir [1,2], Asif Mahmud [1,2], Sayanto Pal [1,2], Ryan King [1,2], Thomas R. Caulfield [1,2,*] and Naeyma N. Islam [1,2,*]

[1] Digital Ether Computing, Jacksonville, FL 32250, USA
[2] Digital Ether Computing, Miami, FL 33131, USA
* Correspondence: thomas@digitalethercomputing.com (T.R.C.); naeyma@digitalethercomputing.com (N.N.I.)

Abstract

Background: Cellular morphology captures compound-induced changes in cell state, but the associations between response magnitude and absorption, distribution, metabolism, excretion, and toxicity (ADMET) depend on the endpoint and assay context. We developed MorphADMET to characterize these associations and evaluate prediction and reference-compound retrieval. **Methods:** We analyzed six imaging datasets and seven morphological representations across eleven binary ADMET endpoints and aqueous solubility. Median morphological responses were compared between label groups using 95% bootstrap confidence intervals from 2000 resamples. Models predicted strong morphological responses from chemical structure and ADMET outcomes from measured morphology or chemical structure. Reference-compound label agreement was compared between morphology-based and random retrieval. Exploratory analyses compared compartment-level features between label groups. **Results:** Positive compounds showed greater morphological responses in A549–Ames, HUVEC–CYP1A2, HUVEC–CYP3A4, and HRCE–CYP2D6, whereas hERG-negative compounds showed greater responses in U2OS cells. Structure-based ensembles predicted strong morphological responses with average precision (AP) of 0.353–0.472, outperforming a compact descriptor baseline in four of five contexts. Discrimination using measured morphology varied by endpoint, with hERG in A549 cells yielding an AP of 0.601 and an area under the receiver operating characteristic curve of 0.786. Exploratory analyses identified group differences in nuclear texture, cytoplasmic organization, and cell shape. In the structure-based ADMET benchmark, MorphADMET achieved the highest observed AP for four of eleven endpoints and the lowest solubility mean absolute error (0.660 $\log_{10}$ mol/L). In the hERG retrieval example, morphology-selected references had higher class-balanced label agreement than random references in all six cellular contexts. An example CYP2D6 neighborhood showed shared physicochemical properties and partial scaffold similarity. **Conclusions:** These findings support endpoint-specific use of chemical structure and cellular morphology in compound assessment. Response magnitude, predictive discrimination, and retrospective reference-label agreement capture distinct aspects of that assessment.

Keywords: cellular morphology; Cell Painting; ADMET; molecular descriptors; phenotypic similarity; machine learning

Check for updates

Academic Editor: Yaw Sing Tan

Received: 21 October 2026
Revised: 17 November 2026
Accepted: 23 November 2026
Published: 27 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Assessing compound properties requires information about both molecular structure and biological response. Structure-based models can estimate ADMET properties before

experimental profiling, whereas high-content imaging measures changes in cells after compound exposure. Cell Painting uses multiplexed fluorescent stains to quantify diverse cellular features, providing a broad description of compound-induced phenotypes [1]. These measurements can complement structure-derived descriptors by capturing responses that depend on the cellular system and exposure conditions.

Previous work has shown that morphological profiles and chemical structures contain useful information for predicting compound activity [2]. Morphological descriptors have also been evaluated against molecular fingerprints for cytotoxicity- and proliferation-related assays [3]. In parallel, structure-based methods such as ADMET-AI and fingerprint-based approaches provide predictions across multiple ADMET endpoints [4,5]. Recent compound-development studies have also incorporated computational ADMET and drug-likeness assessments alongside molecular design, synthesis, and experimental characterization [6]. These studies highlight the importance of computational ADMET assessment in evaluating compound properties during early-stage drug discovery. These approaches motivate a closer examination of when cellular morphology contains endpoint-relevant information and how that information can support assessment alongside chemical structure.

An important distinction is the difference between the magnitude of a morphological response and its pattern. A compound may cause a large overall morphological response without being positive for a particular endpoint. Conversely, a compound with a modest overall response may produce a specific combination of nuclear, cytoplasmic, and whole-cell changes associated with an endpoint label. Endpoint positivity also has different meanings across the ADMET panel: it may indicate an inhibitory activity, a toxicity annotation, or blood–brain barrier penetration. Consequently, stronger morphological change cannot be assumed to imply greater toxicity or uniformly greater endpoint activity.

Here, we introduce MorphADMET, a multimodal framework trained jointly on compounds labeled with ADMET endpoints and on the cellular morphological changes those same compounds produce. Because both modalities are learned in a shared representation, the trained model can be applied in either direction: from chemical structure it predicts the morphological profile a compound would produce together with its ADMET endpoints, and from a measured morphological profile it recovers compound-level chemical information and the associated endpoint annotations. We first compare morphological response magnitudes between endpoint-positive and endpoint-negative compounds across six cellular contexts. We then ask whether chemical structure predicts strong morphological responses and whether measured morphological profiles discriminate endpoint labels. Selected compartment-level features provide an exploratory view of the cellular changes associated with those labels. We next evaluate structure-based endpoint prediction against established approaches and examine whether nearby compounds in morphological space share endpoint annotations and chemical features, the property on which the framework's morphology-to-compound readouts depend. This sequence connects descriptive analysis with predictive evaluation and reference-based interpretation without treating these distinct forms of evidence as interchangeable.

2. Materials and Methods

2.1. Study Design

The study included descriptive comparisons of ADMET endpoints and four analyses: structure-based prediction of strong morphological responses, morphology-based endpoint prediction, structure-based ADMET benchmarking, and morphology-based reference retrieval. Datasets and evaluation procedures were defined for each task. Figure 1 summarizes data partitioning, grouped cross-validation, model selection, and held-out test evaluation for the supervised analyses.



Figure 1. Study design, grouped validation, and model selection. (**a**) Eligible compounds were assigned to a development pool and a reserved test set using compound and scaffold groups, and development data were split into three grouped cross-validation folds with preprocessing fitted within training folds only. (**b**) Model selection used development data alone for each of the three supervised analyses: structure-based prediction of strong morphological response (six model families, equal-weight ensemble built by adding a model only when averaging improved validation AP); structure-based endpoint prediction (tabular and neural models over fingerprints, descriptors, and graphs, combined by greedy weighted ensemble selection); and morphology-based endpoint prediction (model and preprocessing configuration with the best development performance). (**c**) Selected models were refitted on the full development set, frozen, and evaluated once on the reserved test set using AP and AUROC for classification or MAE for solubility. Test scores were not used for selection.

Figure 2 illustrates the MorphADMET architecture. Each compound is represented by two parallel inputs: its chemical structure, labeled with an ADMET endpoint, and a Cell Painting morphological profile acquired for the same compound. The Cell Painting morphological profile datasets are reported in Table 1 and Table S23. A structure encoder (a graph neural network or chemical language model) and a morphology encoder (a CNN image model or profile MLP) project these inputs into a shared multimodal latent space, in which structure and morphology for the same compound map to the same region and the space is organized by ADMET endpoint. Because placement is driven by morphology as well as structure, unlabeled compounds are positioned alongside labeled positives and negatives. For each ADMET endpoint in the datasets listed in Table 2, Table S22 reports the

total number of compounds with endpoint labels and the number of labeled compounds present in each Cell Painting database.

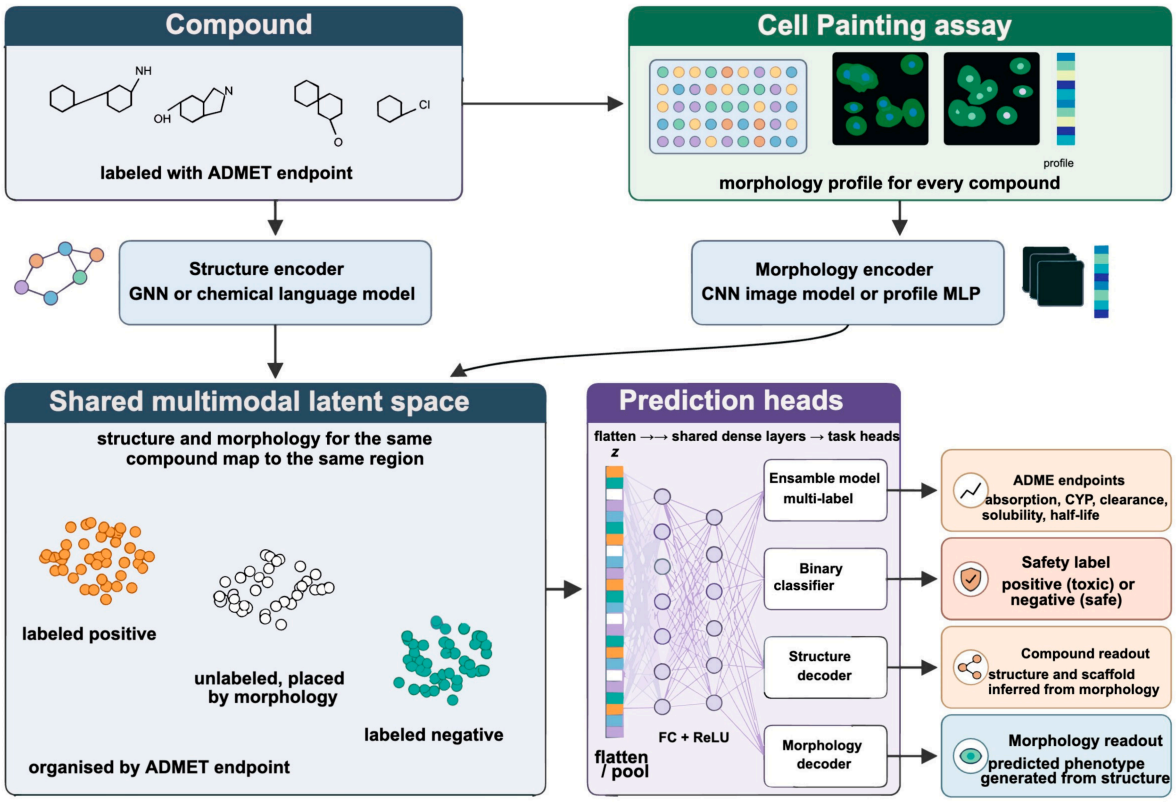


Figure 2. Overview of the MorphADMET framework of chemical structure and Cell Painting profiles for endpoint prediction. Separate encoders map molecular structures and Cell Painting profiles into a shared latent space, where embeddings from the same compound are aligned. ADMET labels help organize the space by endpoint, while morphological profiles allow compounds without endpoint labels to be embedded. The shared representation supports prediction of ADME properties and toxicity, inference of chemical structure and scaffolds from morphology, and prediction of cellular phenotypes from chemical structure. Arrows indicate the direction of information flow. Orange, green, and white circles represent positive, negative, and unlabeled compounds, respectively.

Table 1. Summary of imaging datasets, cellular contexts, compound concentrations, and morphological representations.

Dataset	Cellular Context	Dose (μM)	Feature Extraction Method
JUMP	U2OS, bone cancer	0.625	CellProfiler
OASIS	Primary human hepatocytes	11	CellProfiler and DINOv2
LINCS	A549, lung cancer	10	CellProfiler
EU-OS	HepG2, liver cancer	10	CellProfiler
RxRx3-core	HUVEC, primary vascular endothelial	10	CellProfiler
RxRx19a	HRCE, primary kidney epithelial	3	Cell-DINO

Table 2. Endpoint panel evaluated in this study.

Endpoint	Task	Positive Label or Target
hERG	Binary classification	hERG inhibition
DICTrank	Binary classification	Cardiotoxicity category under the study mapping
DILI	Binary classification	Drug-induced liver injury annotation
DIRIL	Binary classification	Drug-induced renal injury annotation
Ames	Binary classification	Ames mutagenicity
BBB	Binary classification	Blood–brain barrier penetration
CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP3A4	Five binary classifications	Inhibition of the specified CYP isoform
Aqueous solubility	Regression	Measured log10 solubility in mol/L

The latent representation is flattened or pooled and passed through shared dense layers that feed four task-specific heads: an ensemble multi-label head predicting ADMET endpoints (absorption, CYP, clearance, solubility, half-life); a binary classifier producing a safety label (toxic or safe); a structure decoder recovering compound-level chemical features and scaffold from morphology; and a morphology decoder generating a predicted phenotype from structure.

At inference, the trained model operates on either modality alone. From chemical structure, it predicts a morphological profile and ADMET endpoints. From a measured morphological profile, it predicts compound-level chemical features and ADMET endpoints.

For supervised prediction, development data were used for model training, cross-validation, and selection, and reserved test data were used for final evaluation. Analyses of median response differences and selected cellular features were descriptive. Reference retrieval evaluated agreement between experimentally annotated compounds and their morphological neighbors. The unit of interpretation was the compound within a defined assay context, rather than an individual imaged cell.

2.2. Imaging Datasets and Morphological Representations

We used six imaging datasets spanning U2OS bone cancer cells, primary human hepatocytes, A549 lung cancer cells, HepG2 liver cancer cells, human umbilical vein endothelial cells (HUVECs), and human renal cortical epithelial cells (HRCEs). The datasets were JUMP, OASIS, LINCS, EU-OS, RxRx3-core, and RxRx19a, respectively [7–13]. Table 1 lists the selected concentrations and representations. These cell labels describe the experimental systems and do not imply that every endpoint was measured directly in the corresponding tissue.

Morphological representations included engineered CellProfiler measurements and learned image embeddings from DINOv2 and Cell-DINO (cell_dino_cp_vits8) [14–16]. OASIS was analyzed using both CellProfiler and DINOv2, giving seven representations for morphology-based endpoint prediction. The endpoint-comparison and reference-retrieval figures each show six cellular contexts, with one primary-hepatocyte representation.

Well-level profiles were linked to plate, treatment, compound, and control metadata. Where normalization was required, a feature value x was transformed relative to plate-matched DMSO controls as $z = (x - \text{medianDMSO})/(1.4826 \times \text{MADDMSO})$, where MAD

is the median absolute deviation. Profiles supplied in normalized form were retained without repeated normalization.

2.3. Endpoint Annotations and Compound Matching

The hERG Karim dataset defines channel blockade below 10 micromolar as positive and at or above 10 micromolar as negative [17,18]. Ames labels describe bacterial reverse-mutation mutagenicity (positive = mutagenic) [19]. BBBP labels describe blood–brain barrier penetration (positive = penetrating) [20]. Aqueous solubility is the continuous AqSolDB endpoint in log_10 mol/L [21]. Repeated continuous labels were aggregated by the median during structure curation.

For DICTrank, Most-concern and Less-concern categories were mapped to positive and No-concern to negative [22]. DIRIL retained the curated binary annotations from the FDA source [23]. We used the DILI Goldstandard dataset released by Seal et al. [24], which combines DILIst [25] and DILIrank [26]. In the source curation, standardized structures were deduplicated using the first 14 characters of their InChIKeys. When the two sources assigned conflicting toxicity labels to the same compound, the toxic annotation was retained. The resulting source dataset contains 1111 compounds: 716 positive and 395 negative.

The five CYP endpoints retained the isozyme-specific binary inhibition labels from the TDC Veith datasets [17,27]. Additional source checks considered measured Active/Inactive calls and source-specific activity thresholds; these criteria do not imply independent re-curation of all original benchmark labels. Missing annotations were not treated as negative labels. Differences in assay conditions, source thresholds, and annotation criteria may affect label consistency and predictive performance.

Our study presents an integrated framework linking chemical structure, ADMET endpoint annotations, and cellular morphology, enabling the prediction and characterization of compound properties using either chemical information or Cell Painting morphological profiles as independent inputs. Within this framework, RDKit (versions 2026.03.5) molecular descriptors and physicochemical properties were calculated for all compounds with available chemical structures, regardless of Cell Painting data availability. For compounds with experimentally available ADMET endpoint annotations, these data were incorporated into model development and evaluation. In parallel, morphological profiles provided an additional source of information for investigating and predicting ADMET-related outcomes. Importantly, the developed framework supports endpoint prediction from either chemical structure-derived features alone or cellular morphology-derived features alone, while integrating both sources of information to better characterize relationships among molecular properties, morphological responses, and ADMET outcomes.

One important challenge was the limited overlap between the six Cell Painting datasets and the 12 experimentally annotated ADMET endpoint datasets. Restricting the analysis exclusively to compounds with both available morphological profiles and experimental endpoint annotations would substantially reduce the number of eligible compounds, potentially affecting sample sizes, class distributions, and model performance. Therefore, for compounds lacking experimental ADMET annotations, additional computational predictions were generated using QikProp (Schrödinger, LLC, New York, NY, USA) [28] and SwissADME (SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland) [29] to expand the characterization of their ADMET-related properties. This approach allowed us to investigate a broader compound population while distinguishing experimentally available annotations from computationally predicted properties.

To clarify the composition and relationships of these datasets, Tables S14 and S15 report the compound counts and class distributions associated with morphology-derived and chemical structure-derived model inputs, respectively. Table S22 provides a detailed

breakdown of the compounds associated with each of the 12 ADMET endpoints, including the number of compounds overlapping with each of the six Cell Painting datasets. Table S23 summarizes the total number of compounds available in each of the six Cell Painting datasets and Figure S5 shows overlap of the compounds across the six imaging datasets.

Furthermore, the compounds overlapping between the Cell Painting and ADMET endpoint datasets were investigated to characterize median morphological response differences across safety-related endpoints (Figure 3) and ADME-related endpoints (Figure 4). These analyses establish a connection between compound-induced cellular phenotypic changes and their associated ADMET characteristics, highlighting the complementary information provided by chemical structure, physicochemical properties, and cellular morphology. Collectively, this integrated framework enables a more comprehensive characterization of compound behavior while supporting endpoint prediction from either molecular information or cellular morphological features, depending on data availability.

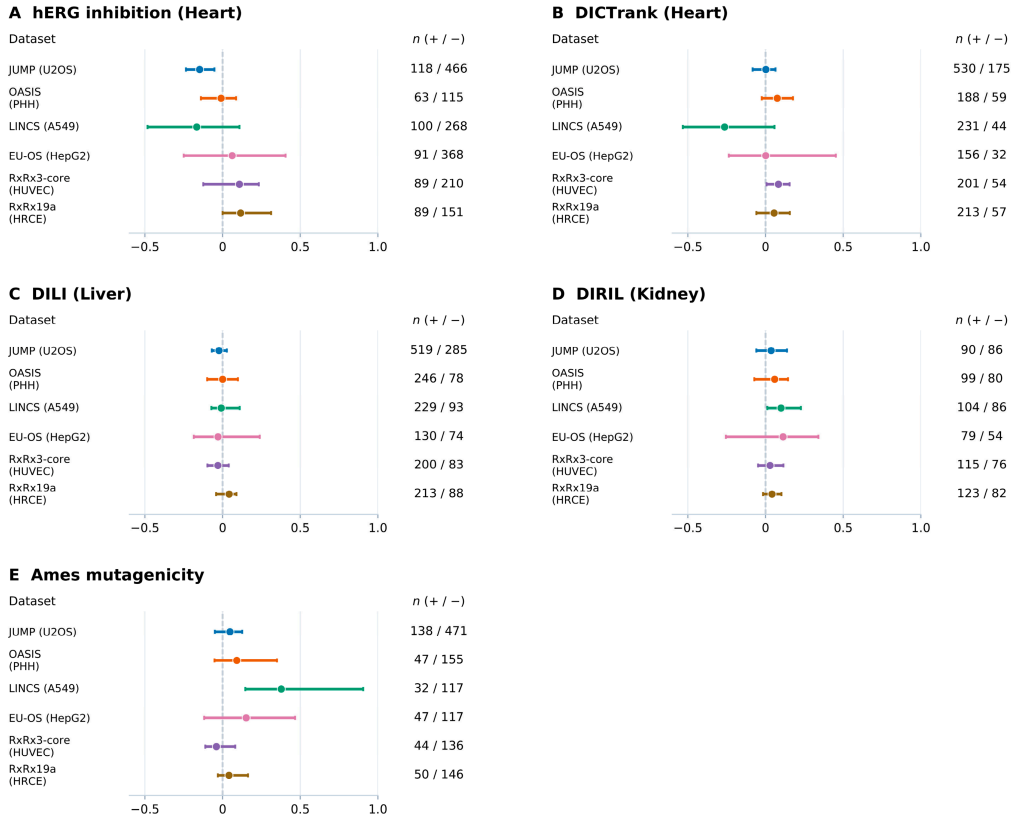


Figure 3. Median morphological response differences across safety endpoints. Panels show (**A**) hERG inhibition, (**B**) DICTrank cardiotoxicity, (**C**) DILI, (**D**) DIRIL, and (**E**) Ames mutagenicity across six cellular contexts, labeled by cell description: bone cancer cells (U2OS, JUMP), primary liver cells (primary human hepatocytes, OASIS), lung cancer cells (A549, LINCS), liver cancer cells (HepG2, EU-OS), primary vascular endothelial cells (HUVEC, RxRx3-core), and primary kidney epithelial cells (HRCE, RxRx19a). Each point represents the median response among endpoint-positive compounds minus the median response among endpoint-negative compounds. Values to the right of zero indicate greater median response among positive compounds; values to the left indicate greater response among negative compounds. Horizontal bars are pointwise 95% bootstrap confidence intervals calculated using 2000 resamples. An interval crossing zero leaves the direction uncertain. Counts beside each row give positive/negative compound numbers. Colors identify cellular contexts. These descriptive comparisons do not measure classifier performance or establish multiplicity-adjusted significance. Response-magnitude distributions for each endpoint are shown in Supplementary Figure S1a–l.

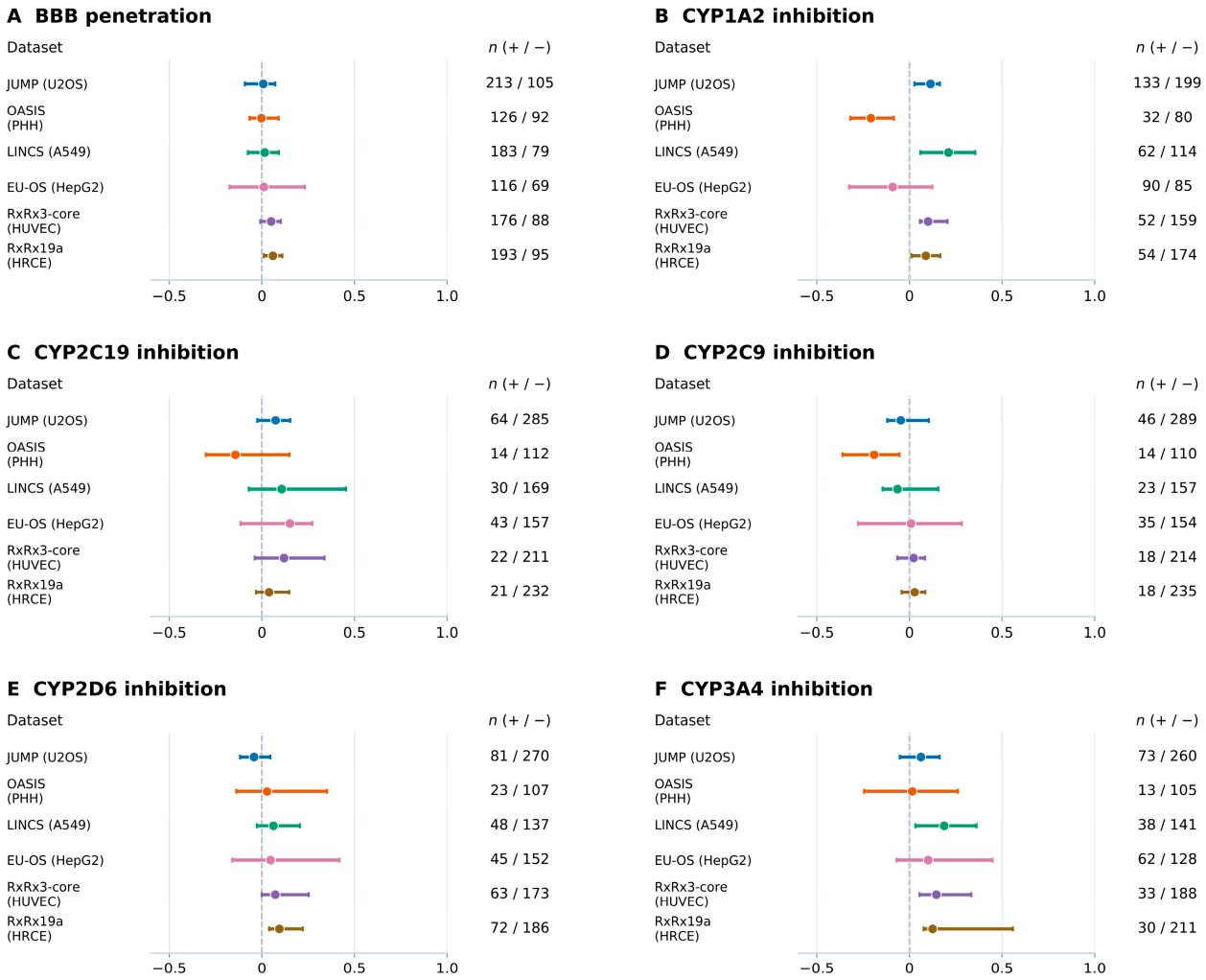


Figure 4. Median morphological response differences across distribution and metabolism endpoints. Panels show (**A**) BBB penetration and inhibition of (**B**) CYP1A2, (**C**) CYP2C19, (**D**) CYP2C9, (**E**) CYP2D6, and (**F**) CYP3A4. Points are positive-minus-negative differences in median morphological response across the same six cellular contexts and labeling convention as Figure 3. Horizontal bars are pointwise 95% bootstrap confidence intervals from 2000 resamples, and the dashed vertical line marks zero. Each color represents a distinct cellular context. Positive values indicate greater response among positive compounds. Counts give positive/negative compound numbers. Endpoint positivity denotes penetration or inhibition, as appropriate, and does not have a uniform interpretation as toxicity. The panel does not include solubility.

2.4. Morphological Response and Endpoint Comparisons

Morphological response magnitude summarized the overall deviation of a profile from its normalized reference. For profile i with included features S_i, normalized values were clipped to the interval $[-10, 10]$ and combined as $R_i = \mathrm{sqrt}\{\Sigma_{j \in S_i} [\mathrm{clip}(z_{ij}, -10, 10)]^2 / |S_i|\}$. This root-mean-square summary limits the influence of extreme feature values while retaining information about the magnitude of deviations in either direction. Because the underlying features and normalization differ across assays, response magnitudes were interpreted within each context.

For each binary endpoint and cellular context, we calculated $\Delta = \mathrm{median}(\mathrm{Rpositive}) - \mathrm{median}(\mathrm{Rnegative})$. A positive Δ indicates greater median morphological response among endpoint-positive compounds, whereas a negative Δ indicates greater median morphological response among endpoint-negative compounds. Pointwise 95% bootstrap confidence intervals were calculated using 2000 resamples. An interval spanning zero

leaves the direction of the median difference uncertain. The response-magnitude distributions stratified by endpoint label are shown for each endpoint and cellular context in Supplementary Figure S1a–k. These findings are exploratory because evaluating multiple endpoint–cellular-context combinations increases the risk of false-positive findings. The reported confidence intervals are pointwise and unadjusted for multiple comparisons; therefore, they characterize uncertainty for individual comparisons without establishing multiplicity-adjusted statistical significance across the endpoint panel. Furthermore, potential dependence among related compounds, chemical families, plates, and batches remains a limitation of the compound-level analysis. These considerations should be taken into account when interpreting the observed morphological response differences.

For response prediction, a separate binary target identified compounds above the assay-specific 90th percentile of the development-set response distribution. With threshold T, a strong response was defined as $R_i > T$ and other responses as $R_i \leq T$. This target describes a comparatively strong morphological response within an assay and is distinct from each experimental ADMET label.

2.5. Molecular Representations and Model Evaluation

Chemical inputs included RDKit molecular descriptors, two-dimensional Mordred (mordredcommunity version 2.0.7) descriptors, and structural fingerprints [30–32]. Fingerprint families included Morgan binary and count fingerprints, atom-pair, topological-torsion, RDKit topological, MACCS, Avalon count, and extended reduced graph fingerprints. The structure-based endpoint analysis additionally considered learned molecular representations and graph models. RDKit descriptors describe properties such as molecular weight, lipophilicity, polarity, hydrogen-bonding capacity, and ring counts; the model inputs therefore extended beyond physicochemical descriptors alone.

Chemical identity and scaffold groups were defined within each supervised modeling dataset. In the full-data structure benchmark, compound standardization included fragment-parent selection, uncharging, and tautomer canonicalization. Compound counts during structure curation are reported in Supplementary Table S21. Stereoisomers were grouped by parent connectivity, and structural relatedness was addressed using nonchiral Murcko scaffolds [33]. Acyclic compounds were assigned to one conservative scaffold group. Model-development compounds were excluded if their connectivity parents or scaffold groups overlapped with those represented in the reserved morphology-test populations. Across all 84 final morphology-based model datasets, development and test sets were confirmed to be disjoint in full InChIKey, connectivity-parent block, and the stored scaffold group. Models were fitted and evaluated separately for each imaging dataset.

For the structure benchmark, we made 2000 paired scaffold-bootstrap attempts per endpoint using the original normalized scaffold groups. Each sampled group retained all its compounds, and the same sampling multiplicities were applied to every model. All 2000 draws were valid for every benchmark comparison. We report pointwise 95% percentile intervals for 92 paired differences against four other models: 48 AP/MAE comparisons and 44 AUROC comparisons [34]. Differences were defined as MorphADMET minus the other model for AP/AUROC and the other model minus MorphADMET for MAE, so positive values consistently favor MorphADMET. An interval entirely above zero supports an advantage for MorphADMET in that comparison; an interval entirely below zero supports the other model. An interval containing zero leaves the direction unresolved and does not establish equivalence. The intervals are unadjusted for multiple comparisons and therefore do not provide simultaneous coverage across the benchmark. For the final morphology models, we made 2000 scaffold-bootstrap attempts per test set, retaining the scaffold grouping used for model evaluation and the exact reported model selection. Table

S17 reports the resulting intervals. These intervals are conditional on the fitted predictions and, for binary endpoints, on draws containing both classes.

2.6. Prediction of Strong Morphological Response from Structure

Response models were developed separately for JUMP, EU-OS, LINCS, RxRx3-core, and RxRx19a. OASIS was excluded from response ensemble selection because too few strong responders were available in the training and validation folds (Supplementary Table S2). A histogram gradient boosting classifier using fourteen chemical descriptors served as the compact baseline.

Candidate response classifiers included logistic regression (scikit-learn version 1.9.0), histogram gradient boosting, CatBoost (version 1.2.10), XGBoost (version 2.1.4), LightGBM (version 4.7.0), and ExtraTrees, combined with alternative chemical feature sets and hyperparameters. Models were ranked by development cross-validation AP. Starting with the highest-ranked candidate, additional models were retained when averaging their validation predictions improved AP. Selected models were refitted on all development compounds, and their test predictions were averaged with equal weights. Supplementary Tables S1–S3 provide development scores, ensemble membership, and test performance.

2.7. Endpoint Prediction from Measured Morphology

For each of the twelve endpoints and seven morphological representations, we evaluated regularized linear models, CatBoost, LightGBM, XGBoost, ExtraTrees, and multilayer perceptrons. Logistic regression was used for binary linear models and ridge regression for solubility. Candidate configurations combined model settings with either ExtraTrees feature-importance selection or principal component analysis (PCA), using 16, 32, or 64 features or components where supported by the data.

For each endpoint–representation pair, we considered seventy-eight configurations within six model families. Model selection used the highest mean development cross-validation AP for classification or the lowest mean MAE for regression. Test scores were not used to select the winning configuration. Each selected configuration was refitted on the development data before evaluation on reserved test compounds. Supplementary Tables S7–S13 report the selected configurations and cross-validation results, and Supplementary Table S14 reports test-set sizes and positive counts. Table S18 provides the candidate search settings, and Table S19 records the training environment.

2.8. Exploratory Analysis of Cellular Compartments

To examine whether endpoint labels were associated with specific morphological patterns, we analyzed selected features in A549 cells for hERG, HepG2 cells for CYP1A2, and HUVECs for hERG. Three features per context were selected using development-set labels to represent distinct measurements and reduce redundant displays. The selected features described nuclear texture, cytoplasmic organization, and whole-cell or cytoplasmic shape. Distributions were displayed as individual compound points and violins, with medians and interquartile ranges.

For the same three endpoint–context pairs and development compounds, we summarized RDKit-derived physicochemical properties separately for endpoint-positive and endpoint-negative groups. These properties comprised molecular weight, calculated logP, TPSA, hydrogen-bond donor and acceptor counts, and aromatic-ring counts. The aim was to ask whether compounds sharing an endpoint label also share recognizable bulk chemical characteristics, and whether those characteristics track with the morphological changes the same compounds produce in a given cellular context. For a safety endpoint, for example, this asks what compounds labeled positive for toxicity look like physicochemically and how that

chemical profile corresponds to the phenotypic changes they induce in the corresponding cell type. Displaying the property distributions alongside the morphological comparisons therefore allowed us to examine whether label-associated morphological differences were accompanied by differences in bulk chemical properties, and whether compounds producing similar morphological responses converge on a common chemical description.

Because the features were selected and displayed using development compounds, this analysis is exploratory. It neither independently validates the selected features nor identifies the variables driving the full predictive model. Any correspondence between chemical properties, morphological response, and endpoint label is descriptive: the observed differences do not establish that the displayed features cause the endpoint, and these comparisons do not validate the displayed features as biomarkers.

2.9. Structure-Based Endpoint Models and Benchmark Comparison

For structure-based endpoint prediction, candidate families included CatBoost, LightGBM, XGBoost, ExtraTrees, regularized linear models, multilayer perceptrons, mixture-of-experts networks, and directed message-passing neural networks (D-MPNNs). We also considered endpoint-specific fine-tuning of CheMeleon (checkpoint: chemeleon_mp.pt) and MapLight-based candidates [5,35,36]. These candidates used complementary descriptor, fingerprint, or learned molecular representations.

Greedy ensemble selection combined candidate predictions according to development performance [37]. At each step, the candidate producing the greatest improvement in the combined prediction was added, allowing repeated selection to increase its weight. Selection continued for up to fifty steps, and the best development ensemble was retained. Supplementary Tables S4 and S5 summarize candidate scores and ensemble membership.

The final MorphADMET ensemble was compared with MapLight, D-MPNN, fine-tuned CheMeleon, and ADMET-AI using the same endpoint-specific development and test compounds and labels (Supplementary Tables S6 and S15). ADMET-AI was trained using our study's training data and its modeling procedure (ADMET-AI 1.4.0 and Chemprop 1.6.1). The methods differed in their model architectures and training settings. Therefore, the benchmark compares their performance under the study's evaluation protocol.

2.10. Morphological Reference Retrieval and Chemical Characterization

Within each cellular context, compounds were ranked by Euclidean distance in a dataset-specific PCA representation of measured morphology, and the ten nearest eligible compounds were retrieved for each query. For binary endpoints, agreement was the fraction of retrieved compounds sharing the query's experimental label. Class-balanced agreement assigned equal weight to the mean agreement for positive and negative queries. The random baseline was the exact expected value of this statistic under uniform selection from the same eligible reference pool. We report morphology-selected minus random agreement in percentage points for all six cellular contexts. Pointwise intervals for this difference were obtained from 2000 query-scaffold-block bootstrap draws, with reference selection and representation held fixed. Table S20 reports the six hERG intervals.

Reference retrieval used measured morphology and experimental annotations, rather than predicted endpoint probabilities. For a selected CYP2D6-positive HUVEC query, the ten nearest references were examined for shared substructures and full Murcko scaffolds. RDKit-derived molecular weight, calculated logP, topological polar surface area (TPSA), hydrogen-bond donor and acceptor counts, and aromatic-ring counts summarized the neighborhood's chemical properties. This was a retrospectively selected example.

3. Results

The supervised analyses followed the development/test assignments and grouped model-selection procedures summarized in Figure 1. Candidate configurations were selected using development-fold performance, and selected models were evaluated on their corresponding reserved test populations. Figure 2 illustrates the distinct prediction heads within the integrated framework, including endpoint prediction, morphology-based readout, and morphology-based reference compound retrieval.

3.1. Endpoint-Associated Response Differences Depend on the Cellular Context

We first examined whether endpoint-positive and endpoint-negative compounds differed in median morphological response (Figures 3 and 4). Across safety, distribution, and metabolism endpoints, both the direction and uncertainty of the difference varied among cellular contexts. Many pointwise confidence intervals crossed zero. Thus, the comparisons did not support a general rule that positive compounds induce stronger morphological changes.

For the safety-related endpoints, Ames-positive compounds in A549 lung cancer cells showed greater median morphological response than Ames-negative compounds, with the displayed confidence interval lying to the right of zero (Figure 3E). This comparison included 32 positive and 117 negative compounds. By contrast, hERG-positive compounds in U2OS bone cancer cells showed a lower median response than hERG-negative compounds, with the interval lying to the left of zero (Figure 3A; 118 positive and 466 negative compounds). These opposing patterns show that the direction of association depends on both endpoint and cellular context.

Among the metabolism endpoints, HUVEC profiles showed greater responses for CYP1A2-positive and CYP3A4-positive compounds, with displayed intervals on the positive side of zero (Figure 4B,F). These comparisons included 52 positive and 159 negative compounds for CYP1A2 and 33 positive and 188 negative compounds for CYP3A4. HRCE profiles similarly showed greater responses for CYP2D6-positive compounds (Figure 4E; 72 positive and 186 negative compounds). These examples were among several context-specific shifts in the plots rather than an exhaustive list of associations.

The median comparisons provide a starting point for modeling but do not measure classification performance. A difference in central tendency can coexist with substantial overlap between the two compound groups. Conversely, multivariate morphological patterns may contain discriminative information even when the median overall response differs little. We therefore evaluated response magnitude prediction and endpoint prediction as separate tasks.

3.2. Chemical Structure Predicts Strong Morphological Responses

Structure-based response ensembles achieved AP values of 0.353–0.472 and AUROC values of 0.807–0.876 across the five evaluated contexts (Figure 5; Table 3). AP exceeded the corresponding test-set strong-response prevalence in each context, which ranged from 7.5% to 13.6%. This indicates that structural features contained information for ranking compounds by their likelihood of producing comparatively strong morphological responses.

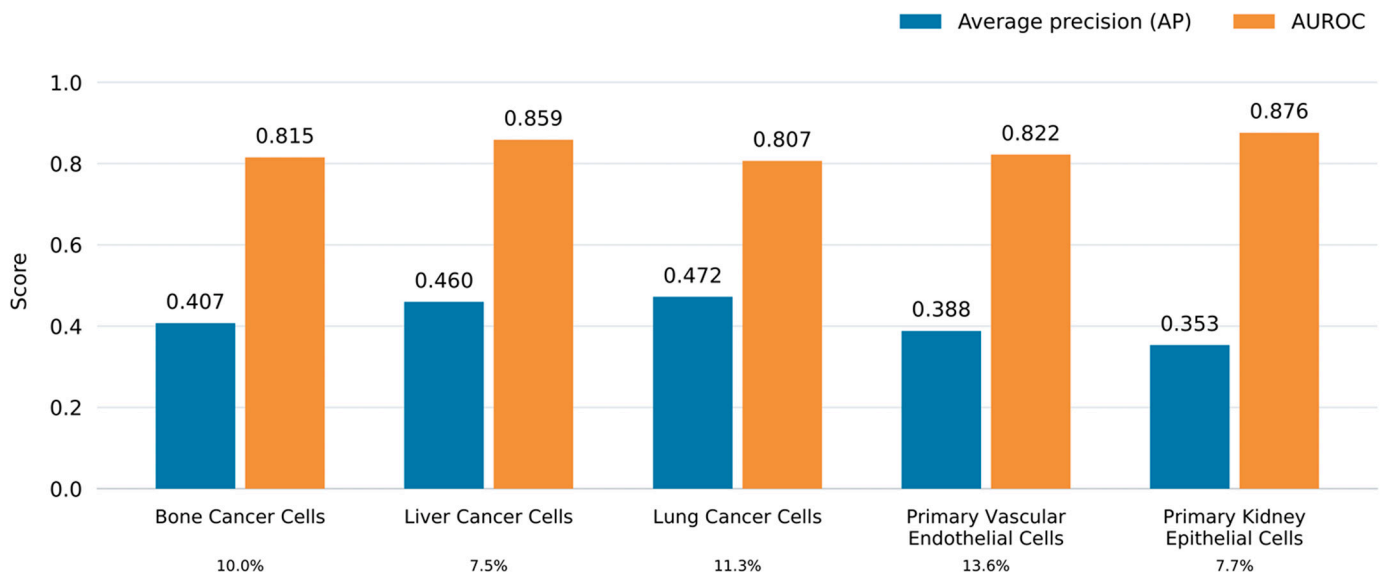


Figure 5. Prediction of strong morphological response from chemical structure. Reserved-test AP and AUROC of the selected MorphADMET response ensembles in bone cancer cells (U2OS), liver cancer cells (HepG2), lung cancer cells (A549), primary vascular endothelial cells (HUVEC), and primary kidney epithelial cells (HRCE). Strong responses exceed the assay-specific development-set 90th-percentile threshold. Blue and orange bars show AP and AUROC, respectively. Percentages below the context labels indicate strong-response prevalence in the test set (10.0%, 7.5%, 11.3%, 13.6%, and 7.7%), which provides the AP reference. Random-ranking AUROC is 0.50. OASIS was excluded from this analysis because of insufficient strong responders in development folds. Baseline comparisons and compound counts are reported in Supplementary Table S3.

Table 3. Reserved test performance for strong morphological response prediction.

Context	Test *n* (+)	Baseline AP	Ensemble AP	Ensemble AUROC
U2OS/JUMP	862 (86)	0.306	0.407	0.815
HepG2/EU-OS	453 (34)	0.326	0.460	0.859
A549/LINCS	230 (26)	0.274	0.472	0.807
HUVEC/RxRx3-core	154 (21)	0.466	0.388	0.822
HRCE/RxRx19a	169 (13)	0.230	0.353	0.876

Relative to the fourteen-descriptor baseline, ensemble AP increased from 0.306 to 0.407 in U2OS, from 0.326 to 0.460 in HepG2, from 0.274 to 0.472 in A549, and from 0.230 to 0.353 in HRCE. The largest absolute increase was 0.198 in A549. HUVEC was the exception: AP decreased from 0.466 to 0.388, although AUROC increased from 0.813 to 0.822. Ensemble AUROC was higher than baseline AUROC in all five contexts (Supplementary Table S3).

The selected ensembles used different chemical representations and model families across contexts (Supplementary Table S2). These results support the feasibility of predicting strong morphological response from structure while showing that a more complex ensemble does not uniformly improve positive-class ranking. Because the baseline and ensemble differ in features, model families, and selection procedure, this comparison does not isolate the contribution of any one component.

3.3. Measured Morphology Supports Endpoint Prediction in Selected Settings

Models using measured morphology showed substantial variation across endpoints and representations (Figure 6). For hERG, AP and AUROC were 0.601 and 0.786 in A549, 0.666 and 0.760 in HUVEC, and 0.599 and 0.762 in HRCE. The corresponding positive-class prevalences were 20.2%, 25.9%, and 32.1%. In primary hepatocytes, the DINOv2 representation yielded hERG AP of 0.615 and AUROC of 0.739, compared with 0.491 and 0.632 for CellProfiler on the matched test population. This example shows that the representation can affect predictive performance even within the same cellular context.

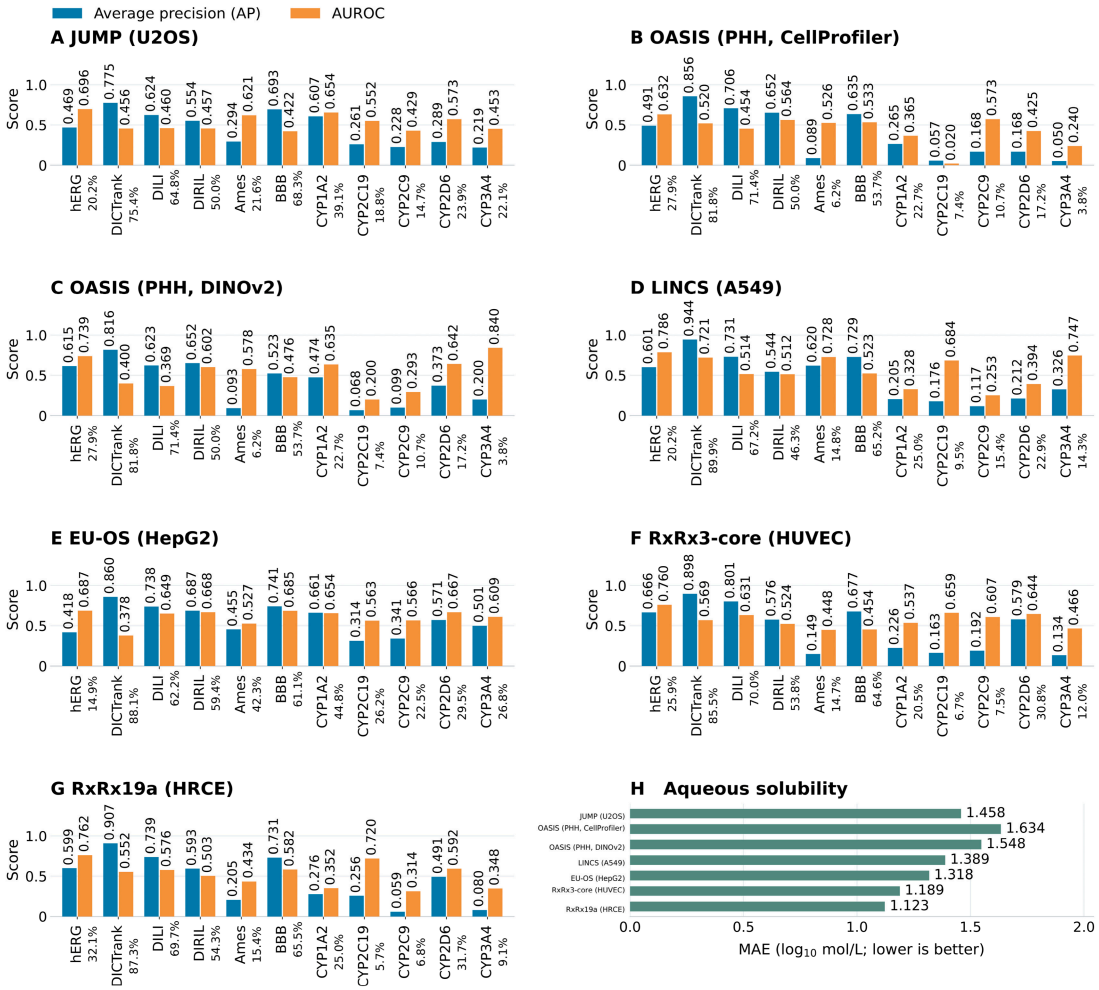


Figure 6. Endpoint prediction from measured cellular morphology. Panels (**A**–**G**) show reserved-test AP and AUROC for eleven classification endpoints using bone cancer cell (U2OS) CellProfiler, primary human hepatocyte CellProfiler, primary human hepatocyte DINOv2, lung cancer cell (A549) CellProfiler, liver cancer cell (HepG2) CellProfiler, primary vascular endothelial cell (HUVEC) CellProfiler, and primary kidney epithelial cell (HRCE) Cell-DINO profiles, respectively. Each endpoint–representation model was selected using development cross-validation. Percentages denote positive test-compound prevalence, the reference for AP; random-ranking AUROC is 0.50. Panel (**H**) shows solubility MAE in log_10 mol/L, represented by green bars. Lower MAE indicates better performance. Endpoint-specific populations differ between contexts, and some have few positives. Test counts are provided in Supplementary Table S14.

Other favorable examples included Ames prediction in A549, with AP of 0.620 and AUROC of 0.728 against 14.8% positive prevalence, and CYP2D6 prediction in HUVEC, with AP of 0.579 and AUROC of 0.644 against 30.8% prevalence. These results indicate that morphology can contain endpoint-relevant information beyond the scalar response magnitude.

Higher AP values did not consistently imply strong discrimination. For U2OS DICTrank, AP was 0.775 against a positive prevalence of 75.4%, while AUROC was 0.456. Similarly, HepG2 DICTrank AP of 0.860 was accompanied by 88.1% positive prevalence and AUROC of 0.378. Some metabolism tasks also showed weak discrimination despite positive response-magnitude differences. In HUVEC, CYP1A2 and CYP3A4 AUROCs were 0.537 and 0.466, respectively, although both endpoints showed greater median responses among positive compounds in Figure 4. Group-level response differences therefore did not guarantee useful prediction on held-out compounds.

Solubility MAE ranged from 1.123 log_10 mol/L for HRCE to 1.634 log_10 mol/L for primary-hepatocyte CellProfiler profiles (Figure 6H). Because these scores describe different test populations, they do not provide a controlled ranking of cell types or a direct comparison with compound's solubility performance. Several classification test sets also contained few positive compounds: OASIS CYP3A4 had one positive among 26 test compounds, and LINCS Ames had four among 27 (Supplementary Table S14). Such sparse positive classes limit the stability of performance estimates. We calculated conditional scaffold-bootstrap intervals for all 84 selected models, yielding 161 AP/AUROC/MAE intervals. Table S17 reports the scores and intervals, and Table S14 provides the corresponding sample sizes and class counts.

3.4. Compartment-Level Features Reveal Endpoint-Associated Morphological Patterns

We next examined selected features to describe the cellular patterns associated with endpoint labels (Figure 7). In A549, the hERG comparison included 288 development compounds, comprising 203 negatives and 85 positives. The displayed nuclear DNA texture, cytoplasmic endoplasmic-reticulum distribution, and cytoplasmic shape features each showed higher medians in the hERG-positive group, together with differences in distribution shape and dispersion (Figure 7A).

For HepG2 CYP1A2, 69 negative and 77 positive development compounds were examined. Positive compounds had a lower median for the displayed nuclear DNA texture feature, a higher median for the cytoplasmic ER–mitochondrial correlation feature, and a lower median for the cell-shape feature (Figure 7B). The opposite directions across features emphasize that an endpoint-associated profile can involve redistribution across measurements rather than a uniform increase in every cellular feature.

In HUVEC, the hERG comparison comprised 169 negative and 76 positive development compounds. Positive compounds showed higher median nuclear texture and cell-shape values and a lower median cell-texture value for the selected measurements (Figure 7C). Although the group distributions differed, they overlapped substantially. These exploratory examples identify measurable cellular correlates of endpoint labels, but they do not establish a common mechanism, causal organelle injury, or independent validation of a morphological biomarker.

3.5. Structure-Based ADMET Performance Varies by Endpoint

We then evaluated endpoint prediction directly from chemical structure (Figure 8; Supplementary Table S6). MorphADMET achieved the highest observed AP among the five compared methods for hERG (0.682), DICTrank (0.914), BBB penetration (0.913), and CYP1A2 inhibition (0.929). For solubility, MorphADMET achieved MAE of 0.660 log10 mol/L, compared with 0.722 for MapLight, 0.863 for D-MPNN, 0.690 for CheMeleon, and 0.723 for ADMET-AI.

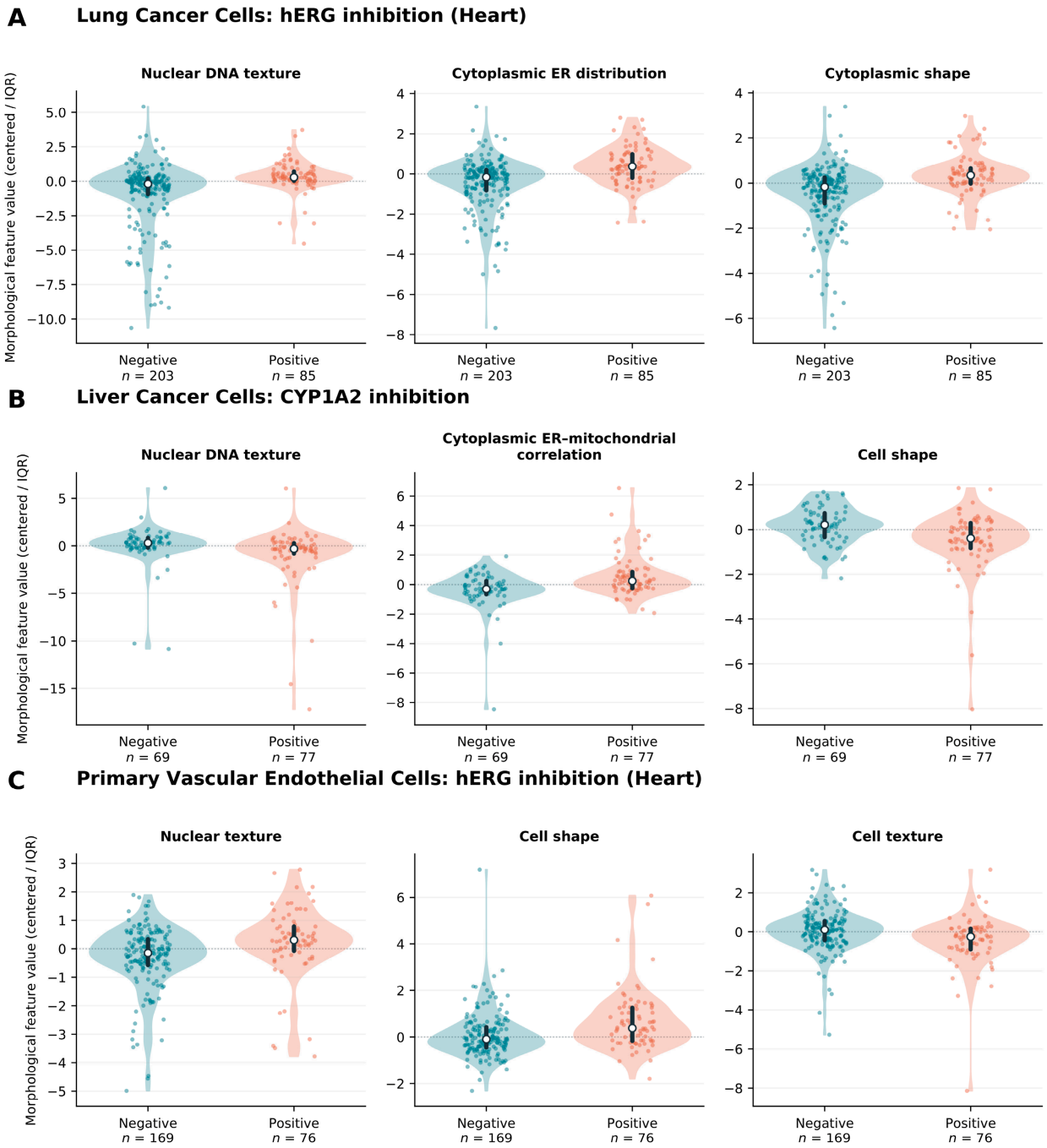


Figure 7. Exploratory compartment-level features associated with endpoint labels. Selected morphological features are shown for (**A**) lung cancer cells (A549), hERG inhibition, with 203 negative and 85 positive development compounds; (**B**) liver cancer cells (HepG2), CYP1A2 inhibition, with 69 negative and 77 positive compounds; and (**C**) primary vascular endothelial cells (HUVEC), hERG inhibition, with 169 negative and 76 positive compounds. The panels show the indicated nuclear texture, cytoplasmic organization, and shape or texture measurements. Each dot represents a compound; violins show distributions, white dots mark medians, and dark bars show interquartile ranges. The vertical scale shows centered features divided by IQR. Features were selected and displayed using development compounds, so these are exploratory associations rather than independent validation or causal evidence. Blue and red represent endpoint-negative and endpoint-positive compounds, respectively.

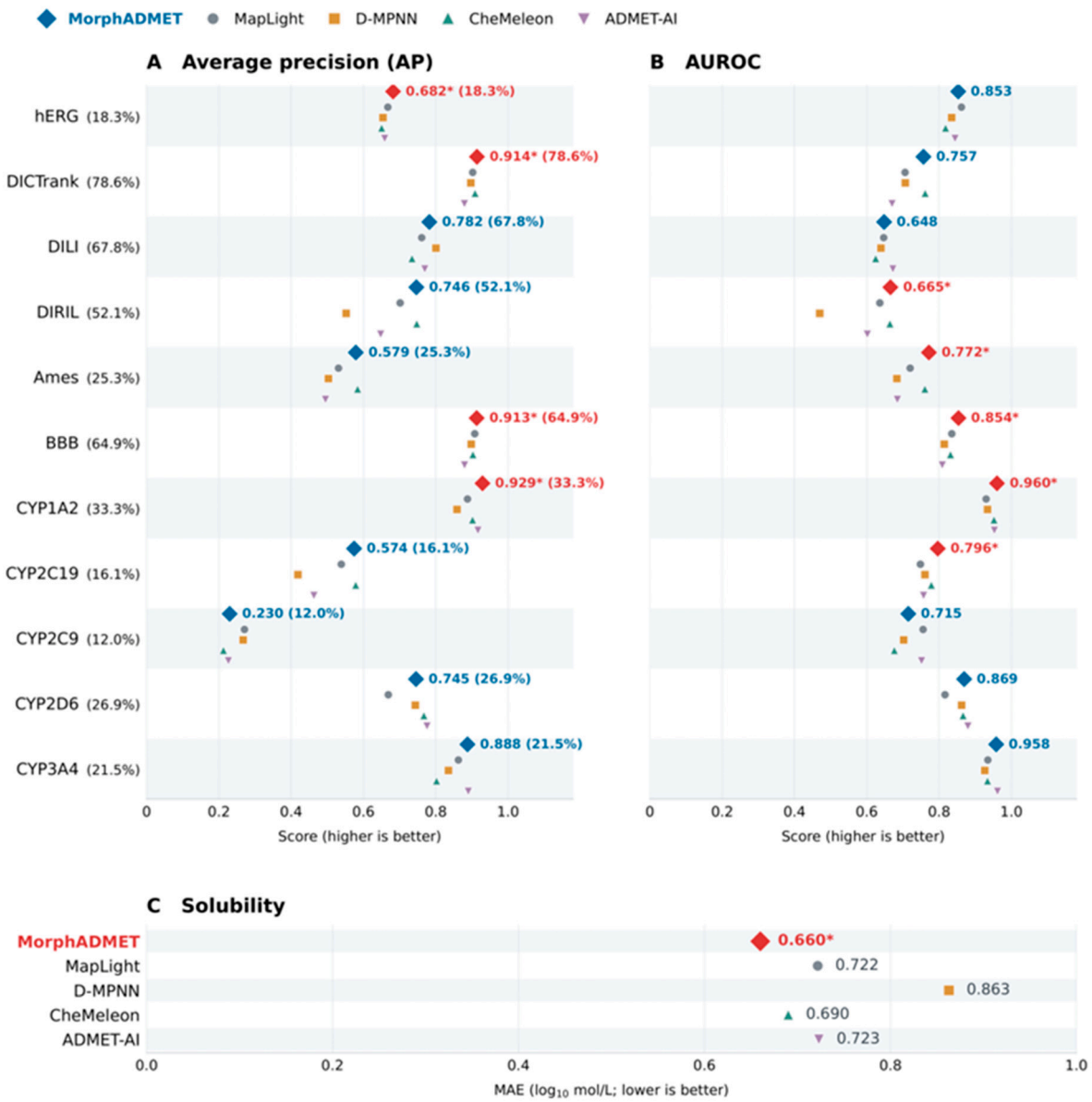


Figure 8. Structure-based ADMET benchmark on matched test compounds. MorphADMET is compared with MapLight, D-MPNN, CheMeleon, and ADMET-AI using (**A**) AP, (**B**) AUROC, and (**C**) solubility MAE. Each endpoint comparison uses the same endpoint-specific development and test compounds. Higher AP/AUROC and lower MAE indicate better performance. Percentages beside classification endpoints indicate test-set positive prevalence. Red diamonds and asterisks mark the best observed score among the five methods for that endpoint and metric, not statistical significance. Solubility error is expressed in log10 mol/L. Numerical scores and sample sizes are provided in Supplementary Table S6.

Using AP for classification and MAE for regression, MorphADMET had better point estimates than MapLight on eleven of twelve endpoints and ADMET-AI on ten of twelve. These pairwise summaries do not mean that MorphADMET led the complete five-method comparison on every endpoint. D-MPNN achieved higher DILI AP (0.801 versus 0.782). CheMeleon achieved slightly higher AP for DIRIL (0.747 versus 0.746), Ames (0.584 versus 0.579), and CYP2C19 (0.579 versus 0.574). MapLight had higher CYP2C9 AP (0.271 versus 0.230), while ADMET-AI had higher AP for CYP2D6 (0.776 versus 0.745) and CYP3A4 (0.890 versus 0.888).

Benchmark rankings depended on both the metric and the endpoint. Across 2000 paired scaffold-bootstrap resamples, 24 of 92 pointwise 95% confidence intervals favored MorphADMET on nine endpoints: BBB penetration, CYP1A2, CYP2C19, CYP2D6, CYP3A4, DICTrank, DILI, DIRIL and solubility. Comparisons with ADMET-AI favored MorphADMET for BBB penetration AP and AUROC, CYP2C19 AP, DICTrank AP and AUROC, and DIRIL AP. For example, the DICTrank AUROC difference was 0.087 (95% CI: 0.020–0.154).

Other favorable comparisons included CYP2D6 AP against MapLight (difference 0.077; 95% CI: 0.020–0.129) and solubility MAE against D-MPNN (reduction 0.203 log10 mol/L; 95% CI: 0.059–0.263). Table S16 reports all differences and intervals (Figures S2–S4). Two intervals favored MapLight on CYP2C9 AP and AUROC, while 66 included zero. No interval favored MorphADMET for Ames or hERG.

3.6. Physicochemical Properties of Endpoint Label Groups in the Compartment-Level Contexts

Because morphological differences between label groups could coincide with differences in bulk chemical properties, we summarized RDKit-derived descriptors for the same three endpoint–context pairs and development compounds used for the compartment-level comparison (Figure 9). In A549 cells, hERG-positive compounds showed higher calculated logP and lower topological polar surface area than hERG-negative compounds, with broadly similar molecular weight and aromatic-ring counts and slightly lower hydrogen-bond donor and acceptor counts (Figure 9A).

In HepG2 cells, CYP1A2-positive compounds showed a narrower molecular-weight distribution, modestly higher calculated logP, lower TPSA, and lower hydrogen-bond donor and acceptor counts than CYP1A2-negative compounds, together with a higher median aromatic-ring count (Figure 9B). In HUVECs, hERG-positive compounds again showed higher calculated logP and lower TPSA, with lower hydrogen-bond donor and acceptor counts and comparable aromatic-ring counts (Figure 9C).

These property shifts are consistent with the lipophilicity and polarity trends commonly reported for hERG and CYP inhibition and indicate that the endpoint-associated morphological differences in Figure 7 occur in compound groups that also differ in bulk physicochemical properties. The distributions overlap substantially in every comparison, and the groups were defined by experimental labels rather than matched on chemistry. These displays therefore describe the chemical composition of the label groups and do not separate chemical from phenotypic contributions to the observed morphological differences.

3.7. Morphological Neighbors Show Enriched hERG Label Agreement

We examined whether compounds with similar morphological profiles shared experimental endpoint annotations. Morphology-selected references showed higher class-balanced hERG label agreement than random references across all six cellular contexts (Figure 10A). The improvements ranged from 2.1 to 8.2 percentage points, with the largest gains observed in A549 and HRCE. The pointwise 95% confidence intervals were above zero in all six contexts, although they were not adjusted for multiple comparisons. Detailed agreement values, query counts, and confidence intervals are provided in Supplementary Table S20.

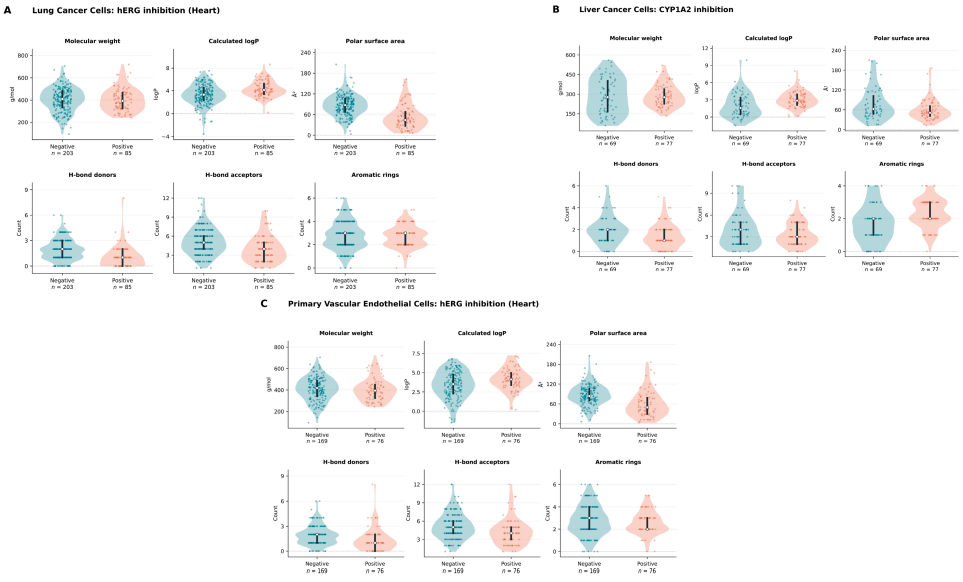


Figure 9. Physicochemical properties of endpoint label groups in the compartment-level contexts. RDKit-derived molecular weight, calculated logP, topological polar surface area, hydrogen-bond donor and acceptor counts, and aromatic-ring counts are shown for endpoint-negative and endpoint-positive development compounds in (**A**) lung cancer cells (A549), hERG inhibition, 203 negative and 85 positive; (**B**) liver cancer cells (HepG2), CYP1A2 inhibition, 69 negative and 77 positive; and (**C**) primary vascular endothelial cells (HUVEC), hERG inhibition, 169 negative and 76 positive. Compound sets match those in Figure 7. Each dot represents a compound; violins show distributions, white dots mark medians, and dark bars show interquartile ranges. The groups were defined by experimental endpoint labels and were not matched on chemical properties, so these displays describe the chemical composition of the label groups rather than isolating chemical from phenotypic contributions.

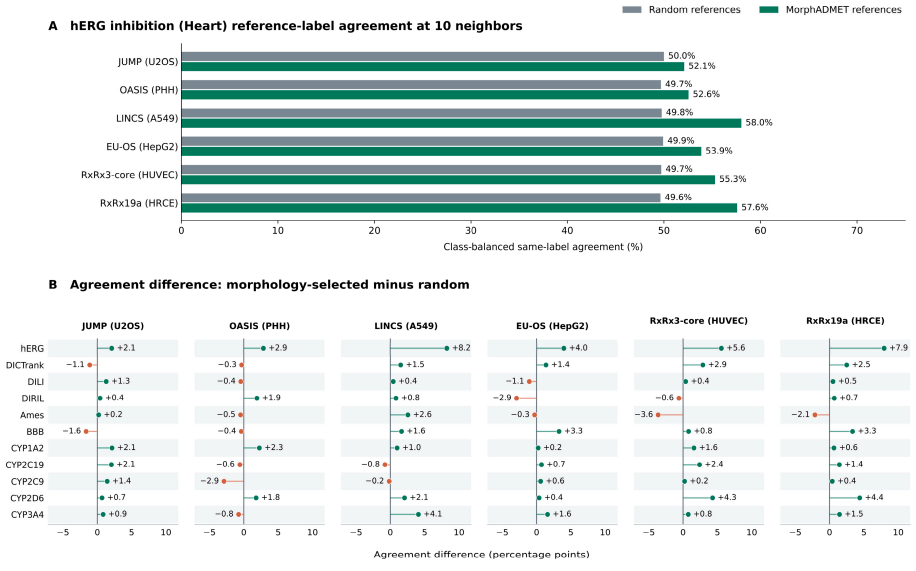


Figure 10. Experimental endpoint agreement among morphological neighbors. (**A**) Class-balanced hERG same-label agreement for ten morphology-selected references and ten random references in six cellular contexts. Class balancing gives equal weight to mean agreement for positive and negative queries. (**B**) Agreement differences across eleven binary endpoints, calculated as morphology-selected minus random agreement in percentage points. Lines connect point estimates to zero and are not confidence intervals. Positive values indicate higher agreement for morphological neighbors; negative values indicate lower agreement. The analysis uses experimental query and reference annotations, not model-predicted labels. Rounded percentages in panel A may not reproduce panel B differences exactly.

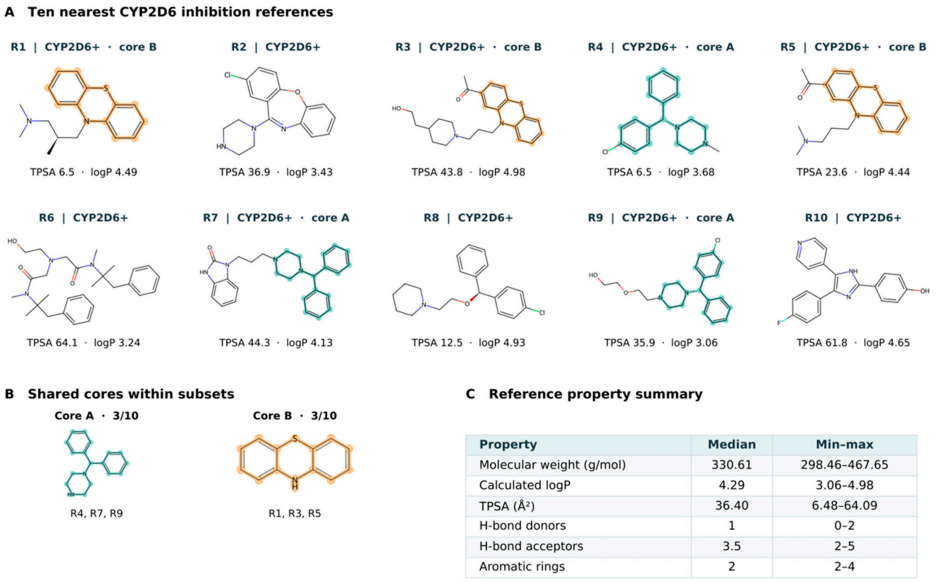
Chemical scaffold similarity and shared physicochemical properties may also contribute to the observed agreement among morphological neighbors. Future studies will investigate the independent contribution of morphological similarity using scaffold-restricted retrieval and physicochemical property-matched comparisons.

The pattern was less consistent for other endpoints (Figure 10B). HUVEC and HRCE showed CYP2D6 agreement gains of 4.3 and 4.4 percentage points, respectively, while A549 showed a 4.1-point gain for CYP3A4. Negative differences also occurred, including Ames in HUVEC (−3.6 points) and DIRIL in HepG2 (−2.9 points). Morphological proximity therefore enriched endpoint agreement in some settings but could be uninformative or less useful than random selection in others.

This analysis supports the use of measured morphology to retrieve contextual experimental evidence. Because the score measures neighbor-label agreement, it is distinct from the held-out endpoint-classification scores in Figure 6. The enrichment also does not show that all nearby compounds share a molecular target or adverse-effect mechanism.

3.8. A CYP2D6 Neighborhood Combines Partial Scaffold Similarity with Shared Properties

To illustrate the chemical context of a retrieved neighborhood, we selected one CYP2D6-positive HUVEC query after examining the retrieval results and inspected its ten morphology-selected references (Figure 11). All ten references were experimentally CYP2D6-positive. Their calculated logP values ranged from 3.06 to 4.98, with a median of 4.29, and each compound contained at least two aromatic rings. Median molecular weight was 330.61 g/mol (range 298.46–467.65), and median TPSA was 36.40 Å2 (range 6.48–64.09). Hydrogen-bond donors ranged from zero to two and acceptors from two to five.



Property	Median	Min-max
Molecular weight (g/mol)	330.61	298.46–467.65
Calculated logP	4.29	3.06–4.98
TPSA (Å²)	36.40	6.48–64.09
H-bond donors	1	0–2
H-bond acceptors	3.5	2–5
Aromatic rings	2	2–4

Figure 11. Chemical characteristics of a morphology selected CYP2D6 neighborhood. (**A**) The ten nearest morphological references for a CYP2D6-positive primary vascular endothelial cell (HUVEC) query, displayed in retrieval rank order. All ten references are experimentally CYP2D6-positive. Structures are annotated with TPSA and calculated logP. (**B**) Recurring cores highlighted in subsets: core A occurs in R4, R7, and R9 and core B in R1, R3, and R5, each in three of the ten references. Identical full Murcko scaffolds occur in R1/R5 and R4/R9; no full scaffold is shared by all ten references. (**C**) Median and range of RDKit-derived physicochemical descriptors across the ten references. All references have calculated logP ≥3 and at least two aromatic rings. The retrospectively selected example provides structural context and does not establish a shared mechanism or prospective retrieval performance.

Two highlighted structural cores each occurred in three references. Core A was present in R4, R7, and R9, and core B in R1, R3, and R5. Within the displayed neighborhood, each core occurred in three of the ten retrieved references (Figure 11B). Identical full Murcko scaffolds were restricted to the R1/R5 and R4/R9 pairs, and no full scaffold spanned all ten compounds. Thus, morphological retrieval recovered a neighborhood with shared endpoint labels and broadly similar physicochemical characteristics while retaining structural differences.

The example is compatible with both structural contributions to phenotypic similarity and convergence among different structures. It cannot establish scaffold-independent retrieval, because related chemical motifs occur within the neighborhood and no scaffold-matched control is shown. As a retrospectively selected case, it illustrates interpretation of a reference set rather than the expected accuracy of retrieval for an arbitrary query.

4. Discussion

The relationship between morphology and ADMET is conditional on both the endpoint and the cellular assay. Initial median comparisons showed greater morphological response among positive compounds for selected endpoint–context pairs and the opposite pattern for others, for both toxicity (Figure 3) and ADME endpoints (Figure 4). Subsequent prediction (Figures 5 and 6) and retrieval (Figure 10) analyses reinforced this heterogeneity. MorphADMET was designed with this heterogeneity in mind: because structure and morphology are embedded in a shared latent space (Figure 2), the model learns the correspondence between a compound's chemistry, the phenotypic changes it produces across cell types, and the endpoint labels assigned to it. Each prediction head then addresses a different aspect of compound assessment, and the same trained representation can be queried from either modality—predicting phenotypic consequences from structure, or recovering compound-level chemical features and endpoint annotations from a measured profile.

The distinction between response magnitude and response pattern is particularly important for what the shared representation can encode. A single response-magnitude score summarizes how far a profile deviates from its normalized reference but discards the identities and directions of individual feature changes; the latent space retains them. This helps explain why a favorable median difference may not translate into strong endpoint prediction, as seen for HUVEC CYP1A2 and CYP3A4, where the positive median shifts in Figure 4B,F are accompanied by AP values close to the corresponding prevalence in Figure 6F. Conversely, the hERG models in several contexts discriminated labels (Figure 6A–G) even though greater overall morphological response was not a consistent property of hERG-positive compounds (Figure 3A)—the discriminating information lay in the pattern rather than the magnitude of the response. The compartment-level plots in Figure 7 provide examples of the heterogeneous feature shifts a multivariate profile can retain. They support further investigation of specific cellular measurements without establishing that these selected features explain model predictions.

Structure-based prediction of strong response is the forward direction of the framework: anticipating how a compound will perturb a given cell type before it is imaged, which offers a potential route to prioritizing compounds for experimental screening. Across five contexts, response AP exceeded the corresponding positive prevalence (Figure 5), and four ensembles improved on the compact descriptor baseline. However, the lower AP in primary vascular endothelial cells (0.388 at 13.6% prevalence; Figure 5) demonstrates that increased model complexity can fail to improve the ranking most relevant to positive-compound prioritization. Moreover, a strong-response prediction concerns an assay-specific percentile threshold. It does not establish endpoint positivity, cytotoxicity, or clinical risk. Prospective evaluation would be needed to determine whether these rankings improve experimental selection.

The reverse direction—inferring endpoint labels from measured morphology—was supported in selected settings, with hERG providing several favorable examples across all six cellular contexts (Figure 6A–G). This is the inference path in which the morphological response itself is the evidence, and the model assigns the endpoint annotation that compounds producing comparable phenotypes tend to carry. Because morphological profiles were retained only for compounds with an available endpoint annotation, each compound in these analyses contributes a structure, a measured profile and a label, so the two modalities are matched at the compound level within every dataset. However, high AP on positive-dominated tasks could coexist with weak AUROC, as for DICTrank in several contexts (Figure 6A–G), and some apparently favorable results were based on very few positive test compounds (Figures 3 and 4, class counts). Interpretation therefore requires both class counts and complementary performance metrics. The variation between CellProfiler and DINOv2 representations for primary-hepatocyte hERG (Figure 6B versus Figure 6C) further suggests that representation choice can influence the information available to a predictive model. It does not demonstrate an overall advantage of learned embeddings across endpoints.

The structure-based benchmark evaluates prediction when imaging data are unavailable, the setting in which the structure encoder must carry the shared representation alone. Paired scaffold-bootstrap intervals supported 24 favorable MorphADMET differences across nine endpoints (Figure 8 and Figures S2–S4), including six comparisons with ADMET-AI across BBB penetration, CYP2C19, DICTrank and DIRIL. Favorable AP (Figure S2) and AUROC (Figure S3) differences indicate better ranking performance; the favorable solubility difference (Figure 8C and Figure S4) indicates lower absolute prediction error. The advantage nevertheless depends on the endpoint, metric and model compared: two intervals favored MapLight on CYP2C9 (Figures S2 and S3), and 66 comparisons remained unresolved.

Morphological reference retrieval makes the morphology-to-compound direction interpretable by returning an annotated neighborhood rather than a point estimate—the compounds whose morphological responses most resemble the query, together with their chemical features and endpoint labels. The consistent direction of hERG agreement improvement across all six contexts (Figure 10A, +2.1 to +8.2 percentage points; Figure 10B) suggests that morphological similarity can enrich for related annotations, although the observed gains were modest and variable, with negative differences for several other endpoints (Figure 10B). The CYP2D6 example in Figure 11 shows how such a neighborhood can be examined for shared chemical properties and structural motifs, with two scaffold cores each recurring in three of ten references (Figure 11B) and a narrow property range across the set (Figure 11C), illustrating the chemistry–phenotype–endpoint correspondence the latent space is intended to capture. References may help formulate hypotheses or choose follow-up experiments, but their labels are not independent measurements of the query. Related scaffolds, general cell stress, exposure, and shared off-target activities remain alternative explanations for similarity.

As with any retrospective benchmarking study, several aspects warrant further study. The datasets differ in cell type, exposure conditions, and representation, and compound overlap between them is limited relative to their individual sizes (Figure S5), so cross-dataset comparisons are best treated as descriptive; the illustrative compartment-level (Figure 7) and chemical-property (Figure 9) examples are likewise exploratory and are intended to show the kinds of signals the framework surfaces. Although structure and morphology are matched within each dataset, the two inference directions were evaluated separately rather than jointly, so the degree to which they agree on the same compound—and whether combining them improves on either alone—remains to be established. Per-compound

calibration and uncertainty estimates and prospective experimental confirmation are further natural next steps.

Future evaluation will build on this work in several directions. Joint-modality fusion experiments on the matched compounds used here, repeated scaffold-grouped or temporal validation, and prospective compound selection would strengthen the performance estimates reported in Figures 5, 6 and 8. Testing whether morphology adds predictive signal beyond structure on the same compounds would characterize complementarity more directly, while scaffold-restricted and property-matched controls would help resolve how much of the retrieval signal in Figure 10 reflects shared phenotypic information rather than structural similarity, given the scaffold recurrence visible in Figure 11B. Evaluating the generated morphology readout (Figure 2) against held-out measured profiles for the same compounds would test the forward path directly, rather than through its downstream endpoint predictions. Replicating selected compartment-level patterns (Figure 7) in independent compounds and concentration-response experiments would further support their biological interpretation.

5. Conclusions

MorphADMET embeds chemical structure and cellular morphology in a shared latent space, so the same trained representation can be queried in either direction: predicting the morphological response and ADMET endpoints of a compound from its structure, or recovering chemical features and endpoint annotations from a measured morphological profile. The two directions were trained and evaluated as separate tasks on non-matched compound sets, so every result reported here characterizes a single modality; the framework is designed to accommodate joint fusion, and the comparative benefit of combining the two modalities is defined here as the question the next phase of work is built to answer.

Across six imaging datasets, seven morphological representations, and twelve ADMET endpoints, the relationship between morphology and ADMET proved conditional rather than general, with greater morphological change accompanying endpoint positivity for some endpoint–context pairs and the opposite pattern for others; response magnitude alone is therefore not an indicator of endpoint liability. Structure-based ensembles ranked strong morphological responders above the corresponding positive prevalence in every context evaluated, measured morphology discriminated endpoint labels in selected settings and most consistently for hERG, and the structure-only benchmark placed MorphADMET competitively against established ADMET models across several endpoints and metrics. Morphological reference retrieval raised class-balanced hERG label agreement above random selection in all six cellular contexts and returned an interpretable neighborhood with shared physicochemical properties and recurring scaffold cores.

These findings are established retrospectively and within modality, and reference retrieval reports the informativeness of a morphological neighborhood rather than an independent measurement of a query compound, so the advantages observed hold conditional on endpoint, representation, and evaluation protocol. The next phase extends the framework along exactly these lines: joint-modality fusion with explicit single-modality ablations on compounds profiled under matched conditions, scaffold-restricted and property-matched retrieval controls, and prospective selection and testing. MorphADMET therefore supports context-specific use of structural and phenotypic evidence while preserving the distinction between morphological response, endpoint prediction, and reference retrieval.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jpbi3040032/s1, Figure S1a–l. Distributions of morphological response magnitude stratified by endpoint label or solubility category across six cellular contexts; Figure S2: Paired differences in benchmark average precision (AP); Figure S3: Paired differences in

benchmark AUROC; Figure S4: Paired differences in solubility prediction error; Figure S5: Compound overlap across six imaging datasets; Table S1: Cross-validation performance for structure-based prediction of strong morphological response; Table S2: Response-prediction ensembles selected using development data; Table S3: Reserved-test performance of MorphADMET response ensembles; Table S4: Development-set cross-validation performance of candidate structure-based ADMET models and the selected ensemble; Table S5: Selected members of the structure-based ADMET ensemble for each endpoint; Table S6: Reserved-test performance of structure-based ADMET models; Table S7: Morphology-based ADMET model selection for JUMP; Table S8: Morphology-based ADMET model selection for OASIS CellProfiler; Table S9: Morphology-based ADMET model selection for OASIS DINOv2; Table S10: Morphology-based ADMET model selection for LINCS; Table S11: Morphology-based ADMET model selection for EU-OS; Table S12: Morphology-based ADMET model selection for RxRx3; Table S13: Morphology-based ADMET model selection for RxRx19a; Table S14: Development and test compound counts for morphology-based endpoint prediction; Table S15: Development and test compound counts for each endpoint; Table S16: Paired performance differences between MorphADMET and four benchmark models; Table S17: Performance and uncertainty of 84 morphology-based ADMET models; Table S18: Hyperparameter search settings for morphology-based ADMET models; Table S19: Software versions used for morphology-based ADMET model training on Linux; Table S20: hERG label agreement for morphology-selected and random references across six imaging contexts; Table S21: Compound counts during structure curation; Table S22: Endpoint panel evaluated in this study; Table S23: Compound counts for the six imaging datasets.

Author Contributions: M.W.U.K. and N.N.I. contributed to methodology, software, data curation, formal analysis, visualization, and writing. A.M. and S.P. contributed to data collection. R.K. provided technical expertise. T.R.C. and N.N.I. contributed to conceptualization, supervision, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted as part of the internal research and development activities of Digital Ether Computing.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study utilized existing datasets from JUMP, OASIS, LINCS, EU-OS, RxRx3-core, and RxRx19a. ADMET endpoint data were obtained from the Therapeutics Data Commons (TDC) and published toxicity resources, including the DILIst/DILIrank-derived gold-standard dataset, DICTrank, and FDA DIRIL. The original datasets and their associated publications are cited in the manuscript. The source code for MorphADMET is publicly available on GitHub at https://github.com/DigitalEtherComputing-AI/MorphADMET (accessed on 15 August 2026).

Acknowledgments: The authors gratefully acknowledge the researchers and organizations responsible for generating, curating, and sharing the imaging and ADMET datasets utilized in this study. We also recognize the developers and contributors of the open-source software tools and computational resources that supported this research. Special thanks are extended to Valon Llabjani and Revivocell for their valuable collaboration and partnership.

Conflicts of Interest: All authors are employed by Digital Ether Computing, a for-profit company that funded this research.

References

1. 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*, 1757–1774. [CrossRef] [PubMed]
2. 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*, 1967. [CrossRef] [PubMed]

3. Seal, S.; Yang, H.; Vollmers, L.; Bender, A. Comparison of Cellular Morphological Descriptors and Molecular Fingerprints for the Prediction of Cytotoxicity- and Proliferation-Related Assays. *Chem. Res. Toxicol.* **2021**, *34*, 422–437. [CrossRef] [PubMed]
4. Swanson, K.; Walther, P.; Leitz, J.; Mukherjee, S.; Wu, J.C.; Shivnaraine, R.V.; Zou, J. ADMET-AI: A machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics* **2024**, *40*, btae416. [CrossRef] [PubMed]
5. Notwell, J.H.; Wood, M.W. ADMET property prediction through combinations of molecular fingerprints. *arXiv* **2023**. [CrossRef]
6. Serdaroğlu, G.; Uludağ, N.; Üstün, E.; Erten, G.; Çolak, N. Design, synthesis, DNA, and BSA-binding of 5-mercapto-1,3,4-thiadiazol and 5-methylisoxazol derivatives: DFT, molecular dockings, ADMET, and drug-likeness profiles. *J. Mol. Struct.* **2026**, *1351*, 144183. [CrossRef]
7. Chandrasekaran, S.N.; Ackerman, J.; Alix, E.; Ando, D.M.; Arevalo, J.; Bennion, M.; Boisseau, N.; Borowa, A.; Boyd, J.D.; Brino, L.; et al. JUMP Cell Painting dataset: Morphological impact of 136,000 chemical and genetic perturbations. *bioRxiv* **2023**. [CrossRef]
8. Ewald, J.D.; Titterton, K.L.; Bäuerle, A.; Beatson, A.; Boiko, D.A.; Cabrera, Á.A.; Cheah, J.; Cimini, B.A.; Gorissen, B.L.; Harrill, J.; et al. Cell Painting for cytotoxicity and mode-of-action analysis in primary human hepatocytes. *Cell Syst.* **2026**, *17*, 101566. [CrossRef] [PubMed]
9. Ewald, J. Axiom OASIS Profiles and Metadata. 2025. Zenodo. Available online: https://zenodo.org/records/17067683 (accessed on 15 August 2026).
10. Way, G.P.; Natoli, T.; Adeboye, A.; Litichevskiy, L.; Yang, A.; Lu, X.; Caicedo, J.C.; Cimini, B.A.; Karhohs, K.; Logan, D.J.; et al. Morphology and gene expression profiling provide complementary information for mapping cell state. *Cell Syst.* **2022**, *13*, 911–923.e9. [CrossRef] [PubMed]
11. Wolff, C.; Neuenschwander, M.; Beese, C.J.; Sitani, D.; Ramos, M.C.; Srovnalova, A.; Varela, M.J.; Polishchuk, P.; Skopelitou, K.E.; Škuta, C.; et al. Morphological profiling data resource enables prediction of chemical compound properties. *iScience* **2025**, *28*, 112445. [CrossRef] [PubMed]
12. Kraus, O.; Comitani, F.; Urbanik, J.; Kenyon-Dean, K.; Arumugam, L.; Saberian, S.; Wognum, C.; Celik, S.; Haque, I.S. RxRx3-core: Benchmarking drug-target interactions in High-Content Microscopy. *arXiv* **2025**, arXiv:2503.20158.
13. Heiser, K.; McLean, P.F.; Davis, C.T.; Fogelson, B.; Gordon, H.B.; Jacobson, P.; Hurst, B.; Miller, B.; Alfa, R.W.; Earnshaw, B.A.; et al. Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of human cells infected with SARS-CoV-2. *bioRxiv* **2020**, 2020.04.21.054387. [CrossRef]
14. 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 Bioinform.* **2021**, *22*, 433. [CrossRef] [PubMed]
15. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.V.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. DINOv2: Learning robust visual features without supervision. *arXiv* **2023**, arXiv:2304.07193. [CrossRef]
16. Moutakanni, T.; Couprie, C.; Yi, S.; Doron, M.; Chen, Z.S.; Moshkov, N.; Gardes, E.; Caron, M.; Touvron, H.; Joulin, A.; et al. Cell-DINO: Self-supervised image-based embeddings for cell fluorescent microscopy. *PLoS Comput. Biol.* **2025**, *21*, e1013828. [CrossRef]
17. Huang, K.; Fu, T.; Gao, W.; Zhao, Y.; Roohani, Y.; Leskovec, J.; Coley, C.W.; Xiao, C.; Sun, J.; Zitnik, M. Artificial intelligence foundation for therapeutic science. *Nat. Chem. Biol.* **2022**, *18*, 1033–1036. [CrossRef] [PubMed]
18. Karim, A.; Lee, M.; Balle, T.; Sattar, A. CardioTox net: A robust predictor for hERG channel blockade based on deep learning meta-feature ensembles. *J. Cheminform.* **2021**, *13*, 60. [CrossRef] [PubMed]
19. Xu, C.; Cheng, F.; Chen, L.; Du, Z.; Li, W.; Liu, G.; Lee, P.W.; Tang, Y. In silico prediction of chemical Ames mutagenicity. *J. Chem. Inf. Model.* **2012**, *52*, 2840–2847. [CrossRef] [PubMed]
20. Martins, I.F.; Teixeira, A.L.; Pinheiro, L.; Falcao, A.O. A Bayesian approach to in silico blood-brain barrier penetration modeling. *J. Chem. Inf. Model.* **2012**, *52*, 1686–1697. [CrossRef] [PubMed]
21. Sorkun, M.C.; Khetan, A.; Er, S. AqSolDB, a curated reference set of aqueous solubility and 2D descriptors for a diverse set of compounds. *Sci. Data* **2019**, *6*, 143. [CrossRef] [PubMed]
22. Qu, Y.; Li, T.; Liu, Z.; Li, D.; Tong, W. DICTrank: The largest reference list of 1318 human drugs ranked by risk of drug-induced cardiotoxicity using FDA labeling. *Drug Discov. Today* **2023**, *28*, 103770. [CrossRef] [PubMed]
23. Connor, S.; Li, T.; Qu, Y.; Roberts, R.A.; Tong, W. Generation of a drug-induced renal injury list to facilitate the development of new approach methodologies for nephrotoxicity. *Drug Discov. Today* **2024**, *29*, 103938. [CrossRef] [PubMed]
24. Seal, S.; Williams, D.; Hosseini-Gerami, L.; Mahale, M.; Carpenter, A.E.; Spjuth, O.; Bender, A. Improved Detection of Drug-Induced Liver Injury by Integrating Predicted In Vivo and In Vitro Data. *Chem. Res. Toxicol.* **2024**, *37*, 1290–1305. [CrossRef] [PubMed]
25. Thakkar, S.; Li, T.; Liu, Z.; Wu, L.; Roberts, R.; Tong, W. Drug-induced liver injury severity and toxicity (DILIst): Binary classification of 1279 drugs by human hepatotoxicity. *Drug Discov. Today* **2020**, *25*, 201–208. [CrossRef] [PubMed]
26. Chen, M.; Suzuki, A.; Thakkar, S.; Yu, K.; Hu, C.; Tong, W. DILIrank: The largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discov. Today* **2016**, *21*, 648–653. [CrossRef] [PubMed]

27. Veith, H.; Southall, N.; Huang, R.; James, T.; Fayne, D.; Artemenko, N.; Shen, M.; Inglese, J.; Austin, C.P.; Lloyd, D.G.; et al. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat. Biotechnol.* **2009**, *27*, 1050–1055. [CrossRef] [PubMed]
28. Schrödinger, LLC. *QikProp*; Schrödinger, LLC: New York, NY, USA, 2024.
29. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [CrossRef] [PubMed]
30. RDKit: Open-Source Cheminformatics. Available online: https://www.rdkit.org (accessed on 17 October 2026).
31. Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A molecular descriptor calculator. *J. Cheminform.* **2018**, *10*, 4. [CrossRef] [PubMed]
32. Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef] [PubMed]
33. Bemis, G.W.; Murcko, M.A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893. [CrossRef] [PubMed]
34. Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef] [PubMed]
35. Heid, E.; Greenman, K.P.; Chung, Y.; Li, S.-C.; Graff, D.E.; Vermeire, F.H.; Wu, H.; Green, W.H.; McGill, C.J. Chemprop: A Machine Learning Package for Chemical Property Prediction. *J. Chem. Inf. Model.* **2024**, *64*, 9–17. [CrossRef] [PubMed]
36. Burns, J.W.; Zalte, A.S.; Abreu, C.R.A.; Sieg, J.; Feldmann, C.; Mathea, M.; Green, W.H. Deep Learning Foundation Models for Low-Data Regimes from Classical Molecular Descriptors. *J. Chem. Inf. Model.* **2026**, *66*, 10692–10701. [CrossRef] [PubMed]
37. Caruana, R.; Niculescu-Mizil, A.; Crew, G.; Ksikes, A. Ensemble selection from libraries of models. In *Proceedings of the Twenty-First International Conference on Machine Learning*; Association for Computing Machinery: New York, NY, USA, 2004; p. 18.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.