Journal Pre-proof



Leveraging Ecotoxicity Parameters and Machine Learning to Rethink the Drug Discovery Pipeline

Daniele Aiello, Laura Bertarini, Reagon Karki, Sheraz Gul, Federica Pellati, Michele Tonelli, Maria Paola Costi

PII: S2095-1779(26)00190-5

DOI: https://doi.org/10.1016/j.jpha.2026.101734

Reference: JPHA 101734

To appear in: *Journal of Pharmaceutical Analysis*

Received Date: 4 November 2025

Revised Date: 22 July 2026

Accepted Date: 27 July 2026

Please cite this article as: D. Aiello, L. Bertarini, R. Karki, S. Gul, F. Pellati, M. Tonelli, M.P. Costi, Leveraging Ecotoxicity Parameters and Machine Learning to Rethink the Drug Discovery Pipeline, *Journal of Pharmaceutical Analysis*, https://doi.org/10.1016/j.jpha.2026.101734.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Published by Elsevier B.V. on behalf of Xi'an Jiaotong University.

Journal Pre-proof

Leveraging Ecotoxicity Parameters and Machine Learning to Rethink the Drug Discovery Pipeline

Daniele Aiello[1], Laura Bertarini[1,2,], Reagon Karki[3,4], Sheraz Gul[3,4], Federica Pellati[1], Michele Tonelli[5], Maria Paola Costi[1*]

[1] Department of Life Sciences, University of Modena and Reggio Emilia, Via G. Campi 103, 41125 Modena, Italy.

[2] Clinical and Experimental Medicine PhD Program. Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy.

[3] Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Discovery Research ScreeningPort, D-22525 Hamburg, Germany.

[4] Fraunhofer Cluster of Excellence for Immune-Mediated Diseases (CIMD), Theodor-Stern-Kai 7, 60596 Frankfurt am Main, Germany.

[5] Department of Pharmacy, University of Genoa, Viale Benedetto XV 3, 16132 Genoa, Italy.

Journal Pre-proof

Leveraging ecotoxicity parameters and machine learning to rethink the drug discovery pipeline

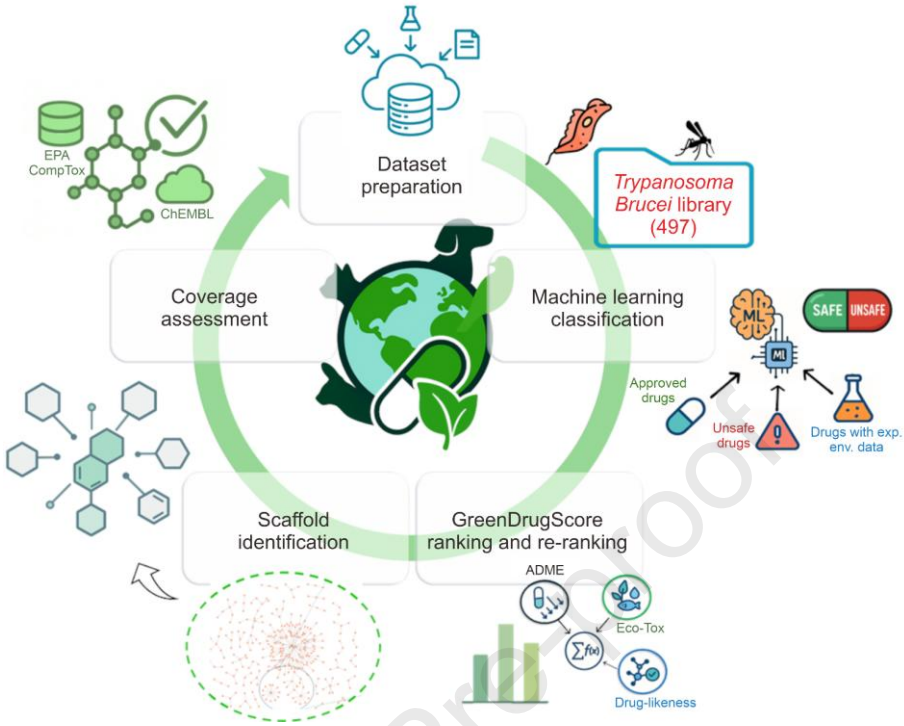
Abstract

This study introduces a modular computational workflow integrating ecotoxicological parameters into the early stages of antiparasitic drug discovery. Using a curated library of 497 *Trypanosoma brucei*-active compounds, we combined a machine learning-based SAFE/UNSAFE classifier with a composite scoring system, the GreenDrugScore (GDS), which merges pharmacokinetic (ADMETscore), ecotoxicological (ECOscore), and drug-likeness (DLscore) properties. The classifier, trained on 1,464 annotated molecules, achieved strong performance (ROC-AUC = 0.85) and identified 267 SAFE compounds. Incorporation of ecotoxicological endpoints, bioaccumulation (BCF), protozoan growth inhibition (IGC_{50}), and aquatic toxicity (LC_{50} for *Daphnia magna* and *Fathead minnow*), reshaped prioritization, with 28.5% of compounds being deprioritized due to unfavourable predicted ECO profiles. Scaffold analysis of top-ranked compounds identified 19 chemotypes with predicted environmentally favourable profiles, combining low predicted bioaccumulation with high aquatic LC_{50}/IGC_{50} values. Overall, these results support the utility of early ecotoxicological integration as a decision-oriented strategy for prioritizing candidate scaffolds under One Health perspective.

Keywords: green drug discovery; ecotoxicology; machine learning; *Trypanosoma brucei*; sustainable prioritization; One Health

- One Health–based workflow integrates ecotoxicology in antiparasitic discovery.
- Machine learning predicts molecular safety from structural information.
- GreenDrugScore changes hit prioritization via ADMET and ECO parameters integration.
- Nineteen prioritized scaffolds with favourable predicted ecotoxicological profiles proposed as antiparasitic candidates for follow-up.

Graphical abstract



1. Introduction

Drug discovery pipeline traditionally relies on a sequential process involving high-throughput screening (HTS), hit validation, lead optimization, and preclinical testing[1]. This well-established pipeline, while effective in optimizing pharmacological profiles, often overlooks key aspects of long-term ecological safety. Early-stage compound selection is typically driven by a combination of *in vitro* potency, target selectivity, and predicted pharmacokinetic and toxicity profiles (ADME-Tox)[2]. This workflow (Fig.1A) has been extensively applied to neglected tropical diseases (NTDs), such as *Leishmaniasis*, *Chagas disease*, and *African trypanosomiasis*, which collectively affect over one billion people worldwide and remain a major cause of morbidity not only in low-income tropical and subtropical regions, but also in certain European Mediterranean countries[3]. These diseases are caused by protozoan parasites of the *Trypanosomatidae* family, including *Trypanosoma brucei*, which is responsible for both the human African trypanosomiasis (HAT) and animal African trypanosomiasis (AAT). Despite increasing awareness of sustainability in pharmaceutical research and development (R&D), environmental risk factors remain

largely excluded from early-stage hit selection, especially in antiparasitic drug discovery. Large-scale use of antiparasitic agents in both human and veterinary medicine has revealed an often-overlooked side effect: many active pharmaceutical ingredients (APIs) are excreted essentially unmetabolized, persist in surface waters and soils, and can trigger chronic ecotoxic effects in non-target species[4–6]. Ivermectin (Fig. 1B), for instance, remains detectable in pasture runoff weeks after administration and lethally disrupts dung- and aquatic-invertebrate communities at nanomolar concentrations[7,8]. Such evidence has prompted regulatory bodies to tighten environmental-risk guidelines, yet early-stage discovery workflows still screen hits almost exclusively for potency, pharmacokinetics and host toxicity, leaving ecotoxicity to late-phase assessment[9].

This disconnect between pharmacological optimization, and environmental safety is exemplified by several widely used antiparasitic agents. Despite their efficacy and favourable pharmacokinetic properties in both human and veterinary medicine, compounds, such as ivermectin, are poorly metabolized and extensively excreted, leading to persistence in soils and surface waters. Environmental exposure to trace concentrations has been shown to cause severe toxicity in non-target organisms, particularly aquatic invertebrates and dung fauna, with documented long-term ecosystem disruption[10–12]. Similar concerns apply to other clinically relevant antiparasitic agents, including nitroheterocycles and diamidines, whose environmental fate and ecotoxicological profiles remain incompletely characterized[13,14]. These examples highlight how compounds that successfully pass conventional ADMET filters may nonetheless pose substantial ecological risks, underscoring the need for discovery strategies that explicitly account for environmental impact at the earliest stages of compound prioritization. This lack of integration represents a major limitation in the field of drug discovery, particularly within the therapeutic area of antiparasitic agents. The issue becomes even more pressing in the search for novel agents against *Trypanosoma brucei,* where existing therapies, including diamidines like diminazene aceturate and nitroheterocycles such as fexinidazole and benznidazole (Fig. 1B), combine sub-optimal pharmacology with poorly characterised environmental profiles[15–18]. Consequently, sustainable discovery must address two interconnected challenges: pharmacological efficacy and ecological safety from the very beginning.

Journal Pre-proof

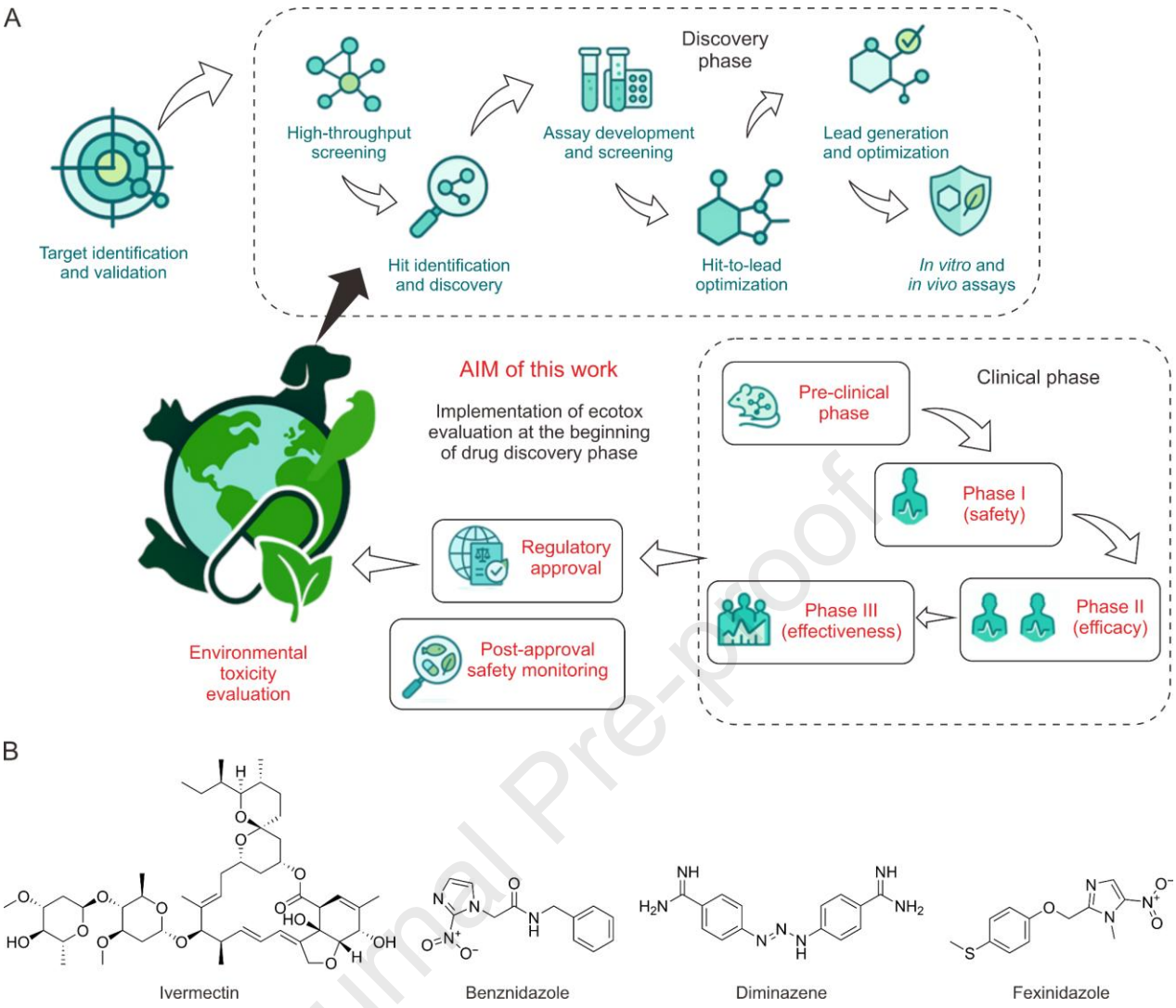


Fig. 1. Overview of the study rationale and representative antiparasitic drugs. (A) Schematic representation of the drug discovery and development pipeline integrating the One Health approach, where environmental toxicity evaluation is introduced early to ensure safer, sustainable compounds. (B) Chemical structures of selected antiparasitic agents with clinical relevance and known environmental concerns.

Despite growing recognition of the environmental risks posed by pharmaceutical compounds, ecotoxicological parameters remain largely disregarded in early-stage drug discovery, a gap that stands in contrast with the One Health perspective, which emphasizes the interdependence of human, animal, and environmental health. These parameters, such as bioaccumulation factor (BCF), acute aquatic toxicity (e.g. LC_{50} for fish or daphnia, IGC_{50} for protozoa), and biodegradability, are central components of environmental

risk assessment (ERA) frameworks enforced by regulatory agencies including the European Medicines Agency (EMA) and the US Environmental Protection Agency (EPA). However, these evaluations are typically performed late in the development pipeline[19], often after clinical candidates have been selected, and mainly for drugs intended for widespread use in either veterinary or human medicine. This delayed assessment creates a regulatory bottleneck, where compounds with otherwise favourable pharmacological profiles may be disqualified due to unacceptable environmental persistence or toxicity. One of the main reasons for this gap is that experimental ecotoxicology assays are resource-intensive, low-throughput, and technically complex, requiring dedicated test organisms, prolonged exposure conditions, and advanced analytical methods[20–22]. As a result, they are incompatible with early-stage compound screening workflows, which must process hundreds or thousands of candidates. In recent years, however, advances in cheminformatics and computational toxicology, particularly the development of machine learning (ML) models and quantitative structure-activity relationship (QSAR) approaches, have enabled the *in-silico* prediction of many ecotoxicological endpoints[23–26]. Tools such as ADMETlab, VEGA, and ECOSAR now allow researchers to estimate bioaccumulation, aquatic toxicity, and degradation profiles early in the drug discovery process, providing an opportunity to integrate ecological safety into virtual screening pipelines in a scalable and cost-effective manner[26–29]. Despite the availability of multiple computational tools for predicting pharmacokinetic and ecotoxicological endpoints, existing approaches remain largely fragmented[30]. Most ADMET-focused workflows prioritize human safety and developability while treating environmental risk as a late-stage regulatory concern. Conversely, ecotoxicological QSAR models are typically developed and applied in isolation, without integration into multiparametric drug discovery pipelines or explicit consideration of medicinal chemistry constraints[31]. As a result, current models rarely provide a unified framework capable of jointly balancing efficacy, human safety, environmental impact, and drug-likeness during early compound triage [32–41]. This lack of integration limits their utility for decision-making in hit selection, particularly in therapeutic areas such as antiparasitic drug discovery where large-scale environmental exposure is expected. To address this issue in a structured and coordinated manner, the European COST Action CA21111 “One Health drugs against parasitic vector borne diseases in Europe and beyond (OneHealthdrugs)” has mobilized a multidisciplinary network of research groups with expertise in parasitology, medicinal chemistry, cheminformatics, veterinary and ecotoxicology. Through

this collaborative approach, the initiative aims to develop integrated strategies to support the discovery of antiparasitic agents that are not only effective but also environmentally sustainable. Importantly, the novelty of the present work does not reside in the development of a new standalone QSAR model or in the prediction of individual ADMET or ecotoxicological endpoints, which are already addressed by several established tools [34,42]. Rather, it lies in the systematic integration of pharmacokinetic, toxicological, ecotoxicological, and drug-likeness blocks within a unified, modular framework explicitly inspired by a One Health perspective. By combining these dimensions into a composite analytical tool, the GreenDrugScore (GDS), the proposed approach moves beyond isolated endpoint prediction and enables informed, comparative decision-making during early-stage compound prioritization. While existing platforms provide valuable endpoint-specific predictions, they do not offer a unified decision-support metric capable of integrating pharmacokinetic, toxicological, environmental, and drug-likeness information within a single prioritization framework. By embedding ecotoxicological criteria directly into the ranking process, the proposed workflow enables comparative prioritization scenarios that are not accessible through conventional ADMET-centric pipelines, thereby addressing a critical gap in sustainable drug discovery.

As part of its coordinated activities, OneHealthdrugs proposes a two-step strategy to demonstrate how early-stage ecotoxicological screening can reshape hit prioritization workflows. The first phase was the development of a database, built through a data mining project on public compound libraries active against various parasites, including *Trypanosoma species*, *Leishmania*, *Babesia*, and *Schistosoma*. Compounds were selected based on explicit criteria of biological activity (phenotypic $IC_{50} < 10$ μM) and structural uniqueness, considering publications from the 2019–2024 timeframe. In the second phase, the data collected were used to develop a workflow aiming to assess whether the integration of *in-silico* predicted ecotoxicological parameters can reshape compound prioritization during hit selection compared to conventional workflow based only on ADME criteria and identify scaffolds with good environmental profile that could serve as suitable prototypes for the development of new, environmentally sustainable compounds. In this context, the primary aim of the present study was to integrate pharmacokinetic, toxicological, ecotoxicological, and drug-likeness dimensions into a single, coherent decision-support framework inspired by a One Health perspective. To achieve this integration, we developed a modular computational workflow for early stage antiparasitic drug discovery that incorporates predicted

ecotoxicological parameters alongside conventional developability criteria. The specific objectives were: (i) implement a machine learning-based classifier to flag compounds predicted as UNSAFE at the screening stage; (ii) to extend ADMET-driven prioritization, we introduced the GDS, a composite scoring function that integrates predicted ecotoxicological endpoints (ECOscore) with pharmacokinetic/toxicity criteria (ADMETscore) and drug-likeness descriptors (DLscore); (iii) to quantify how the inclusion of environmental parameters reshapes compound ranking; and (iv) to identify scaffolds (i.e., core structural frameworks shared by a compound series) with favourable predicted environmental profiles across the selected ecotoxicological endpoints. This work was applied to a subset of the whole collected library, on the 497 compounds active against *Trypanosoma brucei*. Overall, the proposed framework is intended to support early-stage decision-making by enabling comparative prioritization of compounds and chemotypes based on their predicted profiles across multiple dimensions, without implying confirmed environmental safety. In future work, experimental confirmation of antiparasitic activity together with targeted ecotoxicological testing under standardized aquatic assays will be required to translate prediction-based prioritization into evidence within a One Health-oriented development strategy.

2. Material and methods

2.1 Data collection and curation

A focused dataset of 497 compounds active against *Trypanosoma brucei* was assembled through a systematic literature review covering the period from 2019 to 2024 (Supplementary File 1). From each publication, the most potent molecules within each chemical series were selected to maximize both biological activity and structural diversity, applying a threshold of phenotypic IC_{50} or EC_{50} values equal to or lower than 10 µM. The dataset comprised 289 synthetic compounds and 208 natural products (or their derivatives). Chemical structures were standardized using RDKit (v2023.03)[43], with removal of salts, tautomers, and stereoisomers, and converted to canonical simplified molecular input line entry system (SMILES) format for downstream analysis. The analysis focused on key parameters including molecular weight (MW), calculated lipophilicity (ALogP), number of hydrogen bond acceptors (HBA) and donors (HBD), total polar surface area (TPSA), and the number of rotatable bonds. The compounds in the collection displayed molecular weights ranging from 134.2 to 1295.7 g/mol, with an average value of

441.4 g/mol. The ALogP values ranged from −16.8 to 10.3, with a mean of 3.8. The number of HBA and HBD varied from 0 to 18 (avg. 5.3), and 0 to 14 (avg. 2), respectively. Total polar surface area (TPSA) values spanned from 0 to 415.3 Å², with a mean of 88.4 Å². The number of rotatable bonds per molecule ranged from 0 to 33, with an average of 5.9. Based on Lipinski's Rule of Five (RO5)[44], which allows for a maximum of one violation to still be considered drug-like, 79.1% of the database compounds were found to be compliant. High levels of conformity were observed for individual parameters, with 96.4% of compounds meeting the HBA criterion, 98.0% for HBD, and 91.3% for TPSA. Compliance was somewhat lower for MW (70.6%), AlogP (70.0%), and the number of rotatable bonds (84.7%).

2.2 Computational environment and cheminformatics pipeline

All data processing and modelling steps were performed in Python 3.9, employing an RDKit-based framework for molecular standardization, descriptor calculation, fingerprint generation, scaffold extraction, and clustering[43]. Data manipulation was managed using Pandas[45] and Numpy[46], while visualization was carried out through Matplotlib[47] and Seaborn[48]. Molecular similarity matrices were computed using Tanimoto coefficients, and hierarchical clustering was conducted with functions from SciPy[49] and Scikit-learn[50]. The resulting clustering matrix was transformed into a tree file format and subsequently visualized using the iTOL online server[51]. The full workflow was modular and fully automated through custom Python scripts, allowing reproducibility and scalability for future applications.

2.3 Prediction of ADMET and ecotoxicological properties

Pharmacokinetic, toxicological, and ecotoxicological properties were predicted using the ADMETlab 3.0 platform[27]. Human-related endpoints included *h*ERG inhibition, drug-induced liver injury (DILI), and HEK293 cytotoxicity. Environmental impact was estimated using four ecotoxicological descriptors: bio-concentration factor (BCF), growth inhibition of *Tetrahymena pyriformis* (IGC_{50}), and the LC_{50} values for *Daphnia magna* and *Fathead minnow*. These predictions were obtained in log-transformed units and converted into mg/L values using the molecular weight of each compound and the equation:

$$mg/L = 1000 \times MW \times 10^{-\log\ value} \text{ (Eq. 1)}$$

These parameters were selected due to their regulatory relevance and capacity to capture key features of aquatic toxicity and bioaccumulation.

2.4 Machine learning-based safety classification

To classify compounds based on structural safety, a supervised machine learning model was trained to distinguish between “SAFE” and “UNSAFE” molecules using only two-dimensional (2D) molecular descriptors. The training dataset included 880 drugs approved by the US Food and Drug Administration (US FDA), 268 withdrawn agents, 264 phase II failures selected for structural similarity coverage, and 52 molecules with experimental ecotoxicological annotations, yielding a final set of 1464 (Supplementary File 2) unique structures[52,53]. A limited subset of compounds with experimentally determined ecotoxicological data was included to provide a robust reference for defining categorical SAFE/UNSAFE labels according to the Organisation for Economic Co-operation and Development (OECD) environmental hazard guidelines[54–57]. These annotations were used exclusively for training and validation of the machine learning classifier, allowing the model to capture general structural patterns associated with lower versus higher environmental risk. Noteworthy, continuous ecotoxicological values were not directly incorporated into the classification model. Quantitative assessment and prioritization of environmental risk were instead handled downstream by the ECOscore module, which integrates predicted ecotoxicological endpoints within a separate and conceptually distinct scoring block. This modular design ensures a clear separation between classification and ranking tasks and improves the robustness, interpretability, and applicability of the overall workflow. All compounds were standardized to canonical SMILES and described using over 200 RDKit-based molecular descriptors encompassing topological, electronic, and physicochemical features. The enrichment of the SAFE class with approved and clinically evaluated drugs reflects the design objective of aligning the classifier with clinical/regulatory safety standards by anchoring it to structural features associated with human-relevant safety. By prioritizing compounds that have successfully passed extensive preclinical and clinical toxicity filters, this design choice anchors the model decision boundary to real-world pharmacological standards rather than to the average chemical space of screening libraries. Importantly, this clinically informed bias did not result in an unbalanced classification

outcome, as the SAFE/UNSAFE distribution within the *T. brucei* dataset remained well balanced (53.7% SAFE, 46.3% UNSAFE), indicating that the classifier does not impose overly restrictive criteria.

A consensus-based feature selection strategy was adopted by benchmarking six classifiers—Random Forest (RF), CART, XGBoost, Extra Trees (ET), AdaBoost, and Gradient Boosting—across fivefold cross-validation and external test sets. Features consistently prioritized across models were retained. The best-performing classifier, XGBoost, achieved a ROC-AUC of 0.851, a balanced accuracy of 0.773, and a Matthews correlation coefficient (MCC) of 0.553, and was selected for the final classification task. The trained model was applied to the *T. brucei* compound library to assign SAFE/UNSAFE labels and associated probability scores. A principal component analysis-based applicability domain (AD) assessment was also performed to evaluate the reliability of predictions in the context of the model's learned chemical space.

2.5 Multiparametric scoring and compound prioritization

All compounds predicted as SAFE (i.e., with a classification probability above 0.5) were further evaluated using a composite scoring system named GDS, designed to integrate ADMET properties, environmental risk, and drug-likeness criteria (Table S1). The scoring framework consisted of three modules: ADMETscore, ECOscore, and DLscore. ADMETscore was derived from 21 binary endpoints related to pharmacokinetics and human toxicity, adapted from the original ADMETscore model proposed by Guan *et al.* [42]. ECOscore incorporated four ecotoxicological endpoints, BCF, IGC_{50}, LC_{50} for *Daphnia magna*, and LC_{50} for *Fathead minnow*, and applied regulatory thresholds to classify compounds into binary values. Each ecotoxicological endpoint was converted into a binary classification (0 or 1) based on regulatory thresholds and predicted concentrations expressed in mg/L. The IGC_{50}, LC_{50}DM, and LC_{50}FM values, originally provided in log-scale form by ADMETlab 3.0, were first transformed into mg/L using the molecular weight of each compound (Eq. 1). Compounds with predicted toxicity values above 1.5 mg/L were considered non-toxic (score = 1), while those at or below 1.5 mg/L were classified as toxic (score = 0). For BCF, values ≤ 1.5 were considered indicative of a favourable predicted profile (score = 1), and > 1.5 were designated as bioaccumulative (score = 0). This binarization schema was designed to align with

the European Chemicals Agency (ECHA) and OECD toxicity thresholds and enabled consistent integration into the ECOscore calculation.

DLscore was based on a combination of normalized continuous descriptors (such as quantitative estimate of drug-likeness (QED), logP, TPSA, Fsp^3, SAScore) and binary flags derived from rule-based filters like Lipinski's Rule of Five and PAINS. Each score was calculated as a weighted average, where the weight of each endpoint was computed as the product of its entropy-based informativeness, model accuracy (when applicable), and biological or regulatory relevance.

The GDS score was implemented in two configurations. The two-block version combined ADMETscore and ECOscore, while the three-block version included DLscore as an additional refinement to account for structural and physicochemical developability. This ranking allowed a direct comparison of prioritization outcomes with and without the inclusion of environmental parameters.

2.6 Scaffold extraction and ecotoxicological profiling

To evaluate chemical core structures and their associated environmental profiles, all SAFE compounds were subjected to scaffold decomposition using the Bemis–Murcko formalism as implemented in RDKit[43]. Scaffolds were simplified to remove side chains and generalized for consistency. For each scaffold, associated compounds were grouped, and mean values for the four ecotoxicological endpoints were computed. The structural similarity between scaffolds was analysed using Extended-Connectivity Fingerprints (ECFP4) [58] and Tanimoto similarity, and a scaffold network was built to explore the distribution of the chemotypes with favourable predicted ecotoxicological profiles.

Each scaffold was then assigned a composite risk score based on ten parameters, spanning ADME features (such as QED, plasma protein binding, logP), human-relevant toxicological endpoints (*h*ERG, HEK293, *h*ERG-10 µM), and the four ecotoxicological predictions. A traffic-light coding system (green = 2, yellow = 1, red = 0) was applied to each parameter based on thresholds from literature and regulatory guidelines. The average score across all parameters was used to classify each scaffold as low-risk (score ≥ 1.7), intermediate-risk (score between 1.0 and 1.6), or high-risk (score < 1.0). These scaffold-level assessments were visualized using annotated dendrograms, binary heatmaps, and ranked panels to support data interpretation and future optimization strategies.

2.7 Computational workflow and reproducibility

The logical sequence of the workflow is designed to mimic a conventional drug discovery pipeline, and it can be readily implemented or adapted using the modular Python scripts described herein.

All computational analyses were performed using the scripts and datasets provided in the associated repository, which contains the complete workflow required to reproduce the results reported in this study. Scripts are numbered according to their execution order, and they were run sequentially using the input and output files generated at each step. Molecular descriptors were calculated starting from SMILES representations using the script 1_OHD_Paper_2_Descriptors_calculator.py; however, for the present study, descriptor values were directly taken from Supplementary File 2, in which all descriptors were precomputed. Exploratory analysis of molecular properties and chemical space organization was performed using 2_OHD_paper_2_Descriptor_analysis.py and 3_OHD_paper_2_Clustering_analysis.py, enabling statistical characterization and unsupervised clustering of the compound library.

Machine learning-based safety classification was carried out using the pretrained XGBoost model provided in the folder 200625_XGB_5Epo/. The script 4_OHD_paper_2_Model_use.py applies this model to the full descriptor dataset and assigns SAFE or UNSAFE labels together with prediction probabilities. Only compounds classified as SAFE were retained for subsequent analyses. For these compounds, physicochemical, ADMET, and ecotoxicological endpoints were predicted externally using ADMETlab 3.0 and ADMETSAR 3.0. From the ADMETSAR output, the OCT2 and AOT_c endpoints were extracted and merged with the ADMETlab predictions. The resulting unified dataset was processed using 5_OHD_paper_2_Prep_dati_GDS.py, which performs endpoint harmonization, normalization, and binarization or semi-binarization according to predefined thresholds, generating the input required for score calculation.

ADMETscore, ECOscore, and drug-likeness score (DLscore) were then calculated sequentially using 6_OHD_paper_2_Admetscore.py, 7_OHD_paper_2_admetscoreparametrico.py, 8_OHD_paper_2_Ecoscore.py, and 9_OHD_paper_2_DLscore.py, respectively. Intermediate outputs from these scripts were merged and integrated using 10_OHD_paper_2_GDS.py to obtain the final GDS used for compound ranking and prioritization. Finally, scaffold-based analysis of prioritized compounds was performed using 11_OHD_paper_2_scaffold_analysis.py, which extracts Murcko scaffolds and generates a

scaffold similarity network for structural and topological interpretation. All input datasets, intermediate files, trained models, and final outputs required to reproduce the workflow are provided in the repository and corresponding Supplementary Files.

3. Results

As summarized in Fig. 2, the proposed modular workflow, a focused subset of *Trypanosoma brucei*-active compounds was selected and subjected to *in-silico* analysis of ecotoxicological parameters (bioaccumulation and aquatic LC_{50}/IGC_{50}), with a view to providing the necessary input for the subsequent implementation of the modular prioritization workflow. The first module involved the construction of a machine learning classifier trained on a curated dataset of compounds labelled as SAFE or UNSAFE, based on structural features. This model served as a preliminary classifier to identify compounds with high predicted toxicity. To refine the classifier, we performed a fine-tuning step using a curated dataset of 52 compounds with experimentally determined ecotoxicological endpoints[52,53]. This step allowed us to enhance the model ability to distinguish between structurally diverse SAFE and UNSAFE compounds based on real-world ecotoxicological data. The second module of the workflow adopted a multiparametric scoring approach inspired by the ADMETscore developed by Guan et al.[42]. While the original model focused on 18 ADMET endpoints, we expanded the concept by introducing an ECOscore block which accounts for four ecotoxicological parameters: bioaccumulation, aquatic LC_{50}, IGC_{50}. This extension led to the formulation of the GDS, a composite scoring function designed to rank compounds by integrating pharmacokinetic, toxicological, ecotoxicological, and drug-likeness properties. Following the GDS-based ranking, the top-performing compounds were subjected to scaffold extraction to identify recurring structural motifs with potential environmental safety. Two complementary molecular fingerprints were employed for scaffold analysis: ECFP4 to capture structural similarity and pharmacophore fingerprints (PHARM) to reflect pharmacophoric features. In summary, the modular workflow that simulates a conventional drug discovery pipeline started from the CA21111 database, from which we extracted a subset of compounds active against *T. brucei* to perform our analysis. This dataset was processed through a machine learning classifier whose results were then improved by applying the GDS scoring function. This

approach allowed us to compare prioritization outcomes based on ADMET endpoints alone versus the integrated GDS approach. Top-ranked compounds were subjected to scaffold extraction and further analysed using multiple fingerprinting methods, covering both structural and pharmacophoric features. Finally, ecotoxicological parameters were predicted for the most representative scaffolds, and those with the most favourable profiles were predicted as eco-friendly scaffolds.

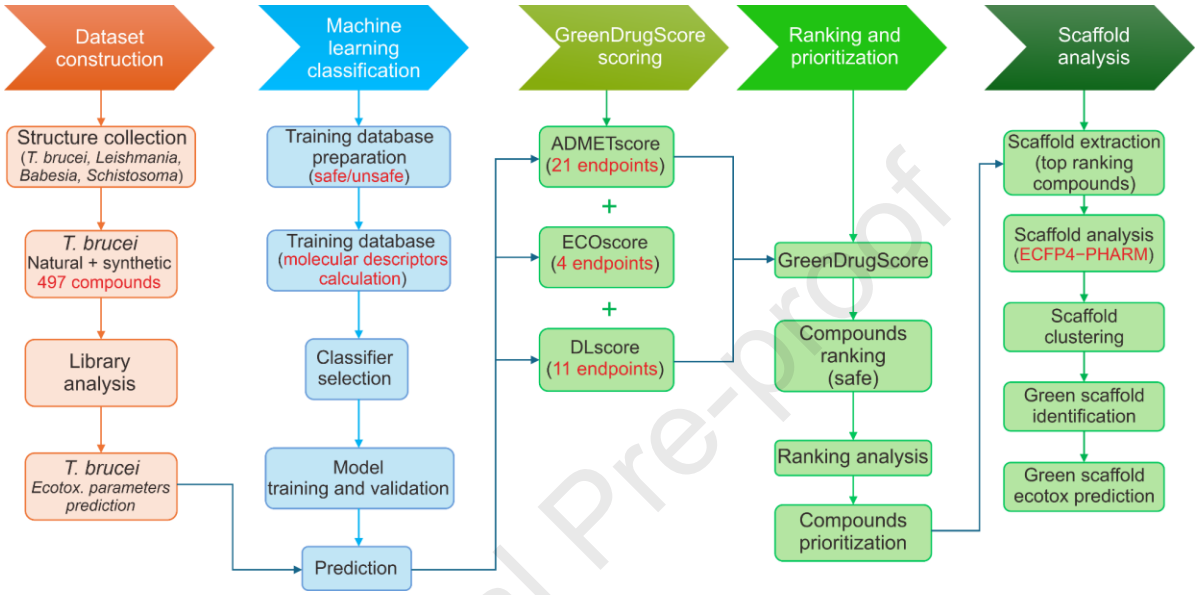


Fig. 2. Modular computational workflow for early-stage ecotoxicology-aware prioritization of antiparasitic compounds. The workflow schematic represents a modular, stepwise framework that outlines the sequential decision process of the proposed approach, from dataset curation and machine learning–based safety classification to multiparametric scoring, ranking-driven prioritization, and scaffold extraction. Each module corresponds to a distinct analytical step within the pipeline, collectively defining a coherent algorithm for early-stage compound selection under integrated pharmacological and ecotoxicological constraints. The pipeline consists of five interconnected modules: (1) Dataset construction: Compounds active against *Trypanosoma brucei*, including both synthetic and natural products (n = 497), were curated from the literature and associated with predicted ecotoxicological parameters. (2) Machine learning classification: A binary classifier was trained on molecular descriptors to predict the likelihood of compounds being SAFE or UNSAFE. (3) GreenDrugScore (GDS) scoring: SAFE compounds were scored using a multiparametric index composed of ADMETscore (21 endpoints), ECOscore (4 ecotoxicological endpoints), and DLscore (11 drug-likeness criteria). (4) Ranking and prioritization: Compounds were ranked based on their GDS values, and high-ranking entries were selected for further analysis. (5) Scaffold analysis: Core structures were extracted from top-ranking compounds and clustered using extended-connectivity fingerprint 4 (ECFP4) and pharmacophore fingerprints (PHARM). Identified "green" scaffolds were evaluated for structural recurrence and predicted ecotoxicological performance.

3.1 *Trypanosoma brucei* library description and preparation

The *Trypanosoma brucei* compound library (Tb-Library), consisting of 208 natural products and 289 synthetic compounds used in this study is a dataset developed through a literature-based data mining project. The dataset includes molecules reported as active (phenotypic $IC_{50} \leq 10$ μM) and was assembled from peer-reviewed publications spanning the 2019–2024 period.

The chemical space of the library was evaluated using a combination of structural (ECFP4)[58] and pharmacophore (PHARM[59]) fingerprints and it was represented using 2D and three-dimensional (3D) *t*-stochastic neighbour embedding (*t*-SNE) algorithm (Fig. 3A). For all the compounds, a preliminary *in-silico* chemoinformatic analysis was carried out to evaluate the drug-likeness and physicochemical properties (Fig. 3B). This assessment was performed using molecular descriptors generated with RDKit[43]. To further investigate the structural diversity of the compound library, we performed cluster analysis based on ECFP4 fingerprints, enabling us to identify recurring chemotypes and evaluate scaffold distribution as reported respectively in Figs. 4A and B.

To complement the structural and clustering analysis, the physicochemical properties of the Tb-Library compounds were evaluated to assess their overall drug-likeness and compliance with classical medicinal chemistry guidelines. As summarized in **Table 1**, the dataset shows a well-balanced distribution of molecular descriptors, with molecular weights predominantly below 500 Da and moderate lipophilicity (AlogP ≈ 3–4). The number of HBD and HBA, as well as the TPSA, fall within the ranges typically associated with favourable absorption and permeability. Moreover, 84% of the molecules meet Lipinski's RO5 criteria, confirming the overall suitability of the library for antiparasitic hit optimization. These results support the chemical diversity and developability of the compounds, providing a robust basis for subsequent ADMET and ecotoxicological profiling.

Journal Pre-proof

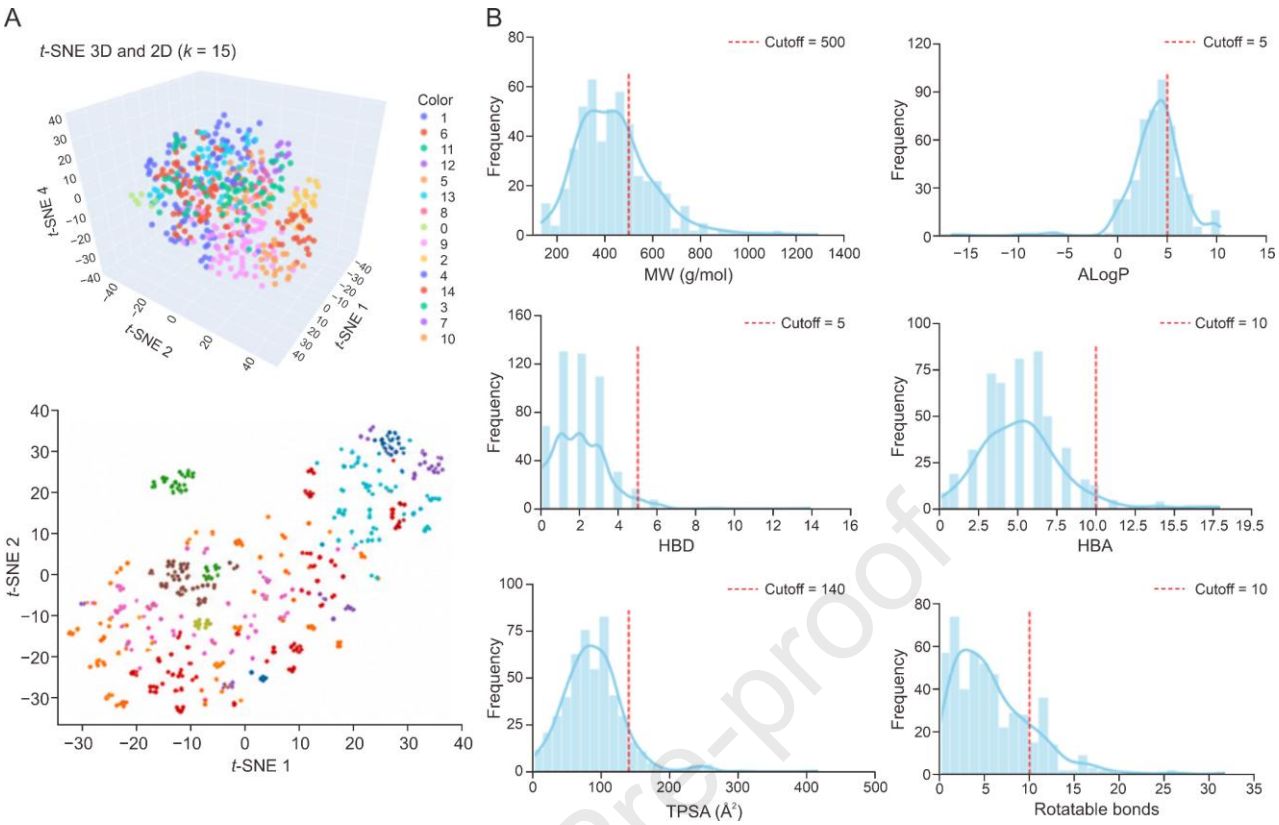


Fig. 3. Cheminformatics-based scaffold analysis and structural clustering of prioritized *T. brucei* compounds. (A) Hybrid t-distributed stochastic neighbour embedding (*t*-SNE) visualizations of the selected dataset based on combined extended-connectivity fingerprint 4 (ECFP4) and pharmacophore (PHARM) fingerprints, shown in three-dimensional (3D, top) and two-dimensional (2D, bottom) projections, reveal the distribution of structural clusters across the chemical space. Each colour represents a distinct cluster identified through unsupervised learning. (B) Distribution plots of key physicochemical descriptors for the compound library, including molecular weight, AlogP, hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), topological polar surface area (TPSA), and number of rotatable bonds.

Table 1. Physicochemical properties of the *Tb-Library* (*n* = 497), including Rule of Five (RO5) compliance.

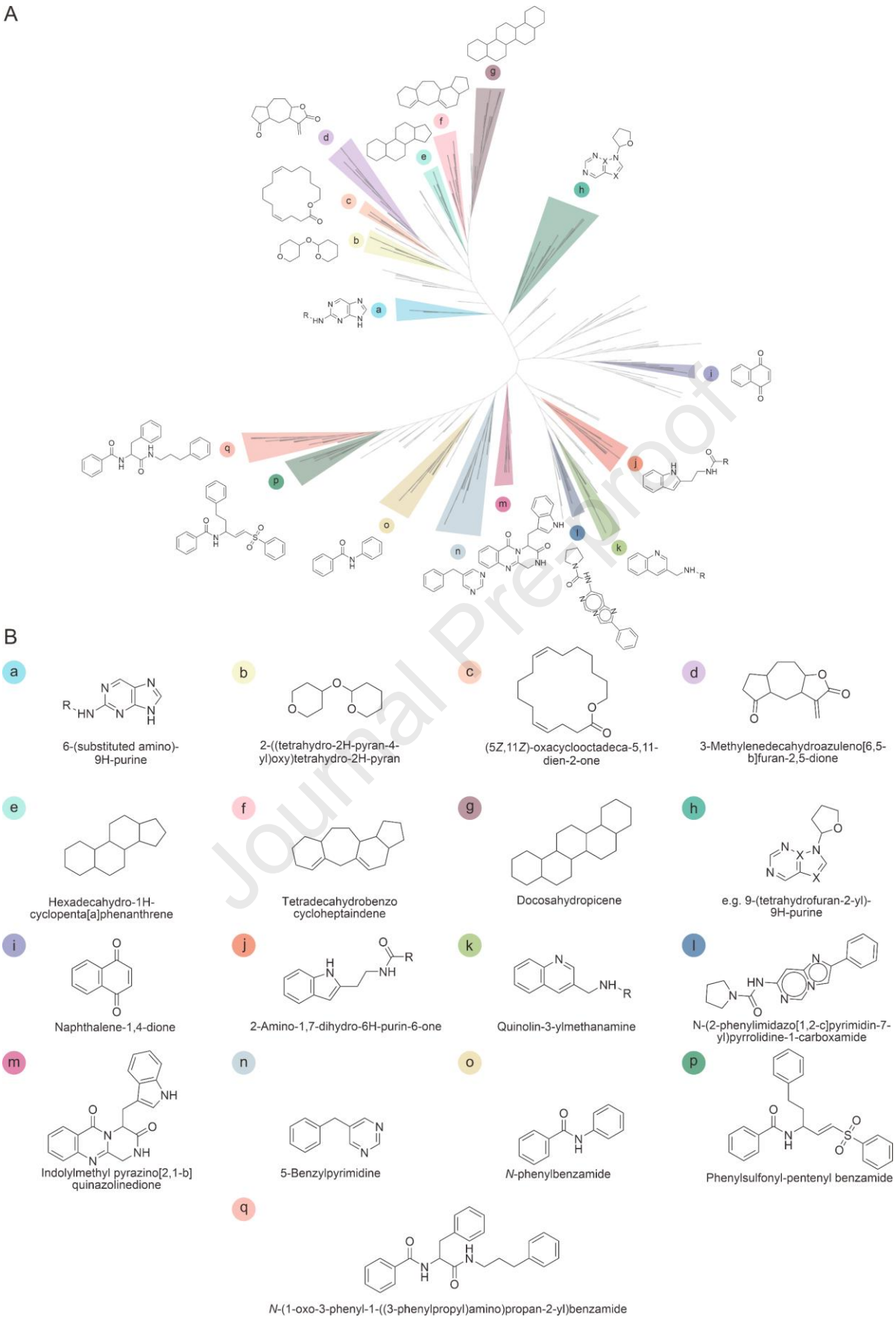
Physicochemical property	Minimum value	Maximum value	Average	Drug-likeness criteria	Compounds according to RO5 (%)
MW (g/mol)	134.2	1295.7	441.4	≤500	70.6
AlogP	−16.8	10.3	3.8	≤5	70.0

HBA	0	18	5.3	≤10	96.4
HBD	0	14	2	≤5	98.0
TPSA (Å^2)	0	415.3	88.4	≤140	91.3
No. of rotatable bonds	0	33	5.9	≤10	84.7
% of compounds passing the RO5 (assuming no more than one violation)					**79.1 %**

MW: molecular weight; HBA: hydrogen bond acceptors; HBD: hydrogen bond donors; TPSA: topological polar surface area; AlogP: calculated octanol/water partition coefficient.

The Analysis of the 17 obtained clusters revealed the presence of several recurring chemotypes (Fig. 4A and B), reflecting a balance between synthetic and natural product scaffolds. Among the most represented classes, purine derivatives (Fig. 4A and B, letters: a, h, j and l) stood out, including substituted 6-aminopurines and guanine-like structures, highlighting the positive impact of nucleobase-like frameworks on the antiparasitic activity. Polycyclic and steroid-like scaffolds (Figs. 4A and B, letters: e, f, g), such as cyclopenta[a]phenanthrene and docosahydropicene derivatives, were also prominent, suggesting a prevalence of hydrophobic and rigid cores typical of bioactive natural products. Additionally, N-aryl and branched N-aryl alkyl benzamide derivatives (Figs. 4A and B, letters: o, p, q), represented another frequent motif, possibly reflecting their versatility and modifiable physicochemical properties. Simpler aromatic heterocycles (Figs. 4A and B, letters: k, n), such as pyrimidines and quinolines, were also observed, confirming their continued relevance as privileged scaffolds in antiparasitic medicinal chemistry. Finally, more complex fused nitrogen-containing polycyclic scaffolds were disclosed (Figs. 4A and B, letter: m), thus offering a wide panorama of structural diversity, useful for identifying scaffolds with favourable safety profiles. A dataset with these characteristics justifies the application of predictive machine learning strategies to guide early-stage prioritization based on multiparametric properties.

Journal Pre-proof



Journal Pre-proof

Fig. 4. Circular scaffold dendrogram and representative chemotype scaffolds of the Tb-Library. (A) Circular scaffold dendrogram based on ECFP4 fingerprint similarity, showing structurally related clusters. Enlarged coloured wedges indicate representative chemotypes; the central node represents the common root of similarity expansion. (B) Representative molecular scaffolds from the major chemotype clusters. Each scaffold is labelled with a letter (a–q) corresponding to its position in panel B and coloured accordingly. Classes include purine derivatives (a, h), steroid-like polycyclic frameworks (e, f, g), imidazole- and triazole-containing heterocycles (l, p), benzamide analogues (o, q), and simple heteroaromatics such as quinolines and pyrimidines (k, n), illustrating the chemical diversity within the Tb-Library. ECFP4, extended-connectivity fingerprint 4.

3.2 Compounds classification by machine learning algorithm

The integration of bioinformatics and cheminformatics, together with recent advances in AI-powered models and agentic workflows, is increasingly important in modern drug discovery. These approaches enable efficient analysis of complex, high-dimensional data and support data-driven prioritization and design of new therapeutics. To obtain a preliminary, structure-based SAFE/UNSAFE classification, we implemented a supervised machine learning workflow designed to distinguish compounds predicted as “SAFE” or “UNSAFE” using molecular structure alone. This classifier was conceived as an early-stage prioritization layer, operating without reliance on experimental ADMET or ecotoxicological measurements. The overall workflow is summarized in Fig. 2 and consists of five main steps: (i) preparation of a labelled training dataset based on regulatory status and known toxicological outcomes, (ii) calculation of 2D molecular descriptors using RDKit, (iii) selection and benchmarking of multiple classifiers, (iv) feature selection based on importance rankings, and (v) model training, cross-validation, and performance evaluation on held-out test and external datasets. Six machine learning algorithms were tested in parallel, Random Forest, Decision Tree (CART), XGBoost, Extra Trees, AdaBoost, and Gradient Boosting, using descriptor matrices generated from a curated dataset of 1464 compounds. These compounds included approved drugs, withdrawn molecules, phase II failures, and environmentally hazardous substances, thus ensuring adequate chemical diversity and applicability domain coverage. The training dataset was compiled

from curated sources to cover both SAFE and UNSAFE compounds. An initial set of 880 US FDA-approved (SAFE) and 268 withdrawn (UNSAFE) drugs was sourced from the ADMETscore dataset[42]. To improve chemical space coverage, we added 264 ChEMBL compounds that failed phase II trials, selected based on molecular similarity to our compound library. This helped populate the intermediate zone between approved and high-risk scaffolds. A further set of 52 compounds with experimentally reported ecotoxicity were included, classified as SAFE or UNSAFE using literature-based lowest effect concentration (LEC), no observed effect concentration (NOEC), and predicted no-effect concentration (PNEC), or regulatory thresholds[52,53]. The final dataset contained 1464 unique structures (932 SAFE, 532 UNSAFE), standardized as canonical SMILES. Redundancy was removed via international chemical identifier key (InChIKey), and over 200 RDKit (v2023.03) molecular descriptors were computed, encompassing physicochemical, topological, and electronic features for ML analysis.

To support classification, we computed over 200 2D molecular descriptors per compound, encompassing physicochemical, topological, and electronic properties. From this initial set, a feature selection pipeline was applied to reduce dimensionality and retain only the most informative descriptors. We adopted a consensus approach using the aforementioned six machine learning algorithms (Table 2) each trained over multiple iterations. At each run, top-ranked features were recorded, and only descriptors consistently prioritized across classifiers were retained for downstream modelling. Subsequently, a benchmarking analysis was conducted to identify the best-performing classifier for SAFE/UNSAFE prediction. All models were trained on a balanced dataset using the reduced descriptor set and evaluated across five independent train/test splits.

Journal Pre-proof

Table 2. Comparison of classifier performance using synthetic minority oversampling technique (SMOTE), 30 descriptors, and 5 training epochs. Reported values include area under the receiver operating characteristic curve (ROC-AUC), balanced accuracy (BA), area under the precision-recall curve (PR AUC), and the Matthews correlation coefficient (MCC), the latter being especially suitable for evaluating models on imbalanced datasets.

Classifier	ROC-AUC	BA	PR AUC	MCC
XGBoost	**0.851**	0.773	0.837	**0.553**
AdaBoost	0.744	0.676	0.725	0.357
Gradient Boosting	0.759	0.688	0.724	0.357
Extra Trees	0.787	0.709	0.755	0.410
CART	0.739	0.665	0.726	0.318
Random Forest	0.810	0.735	0.783	0.464

Among the tested models, XGBoost emerged as the top-performing classifier (Fig. 5), achieving the highest ROC-AUC (0.851) (Fig. 5A), balanced accuracy (0.773), PR-AUC (0.837), and Matthews correlation coefficient (MCC = 0.553). Random Forest and Extra Trees also demonstrated strong performance, with ROC-AUC values of 0.810 (Fig. 5A) and 0.787, respectively, although with slightly lower MCC values (0.464 and 0.410). Gradient Boosting and AdaBoost reached comparable ROC-AUCs (0.759 and 0.744), but showed reduced predictive reliability, particularly in terms of MCC (both ~0.357). CART yielded the lowest scores across all metrics, confirming its limited generalizability in this context. Overall, XGBoost was thus selected as the final classifier for compound screening due to its superior and balanced performance across all evaluation criteria (Supplementary material Figs. S2–S7).

Journal Pre-proof

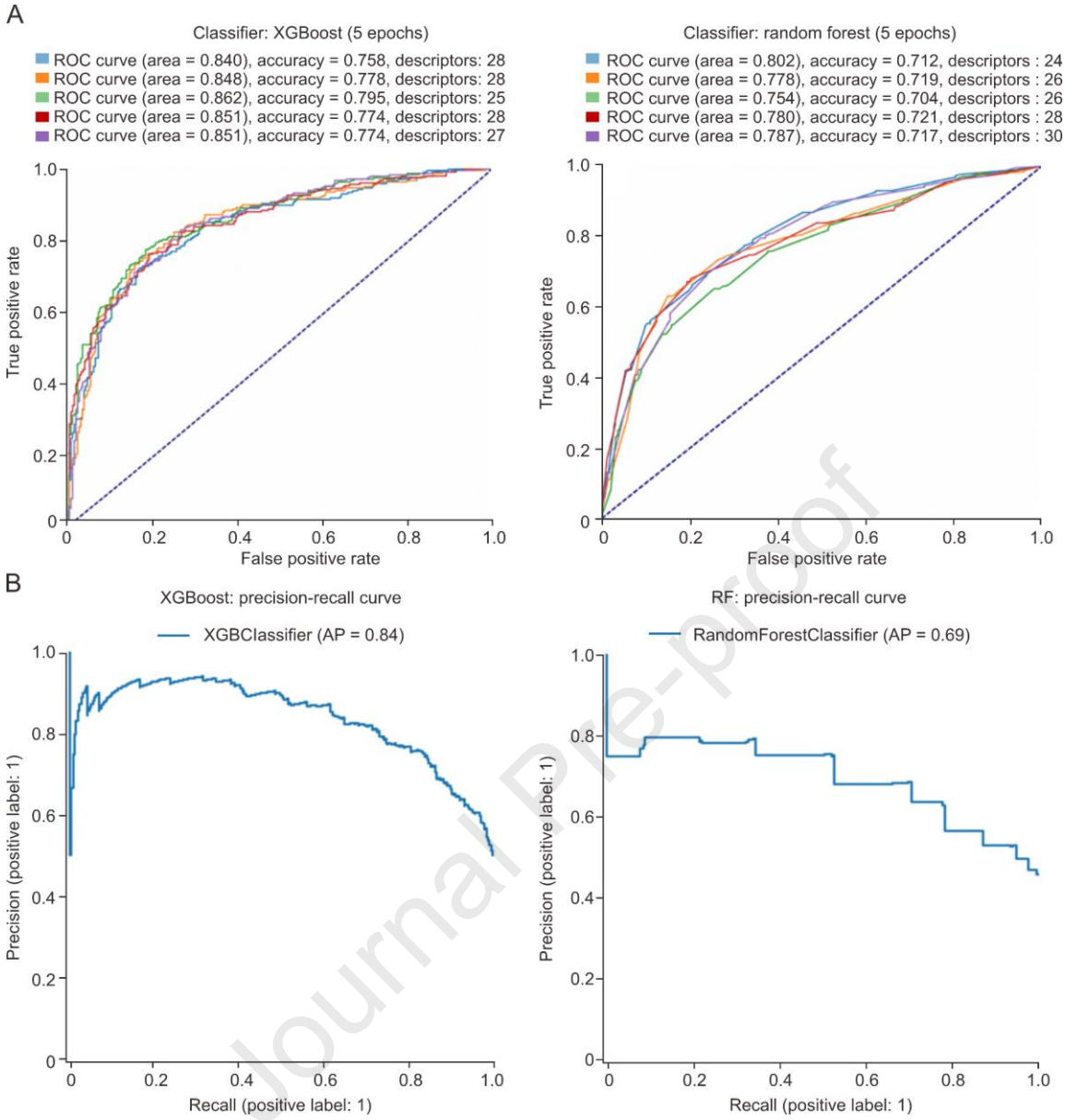


Fig. 5. Comparison of XGBoost and Random Forest classifiers for SAFE/UNSAFE prediction. (A) Receiver operating characteristic (ROC) curves from five training runs using optimized descriptor subsets. (B) Precision-recall curves for the corresponding models. XGBoost achieved higher and more stable performance than Random Forest, with mean area under the receiver operating characteristic curve (ROC-AUC) and average precision (AP) values of approximately 0.84 and 0.84, respectively, compared with approximately 0.80 and 0.69 for Random Forest.

XGBoost was trained using the top 24 descriptors selected by iterative feature ranking and achieved a ROC-AUC of 0.833 on the external test set. The overall classification accuracy was 77.4%, with a sensitivity of 84.8% for identifying UNSAFE compounds and a specificity of 69.9% for correctly classifying SAFE molecules. The confusion matrix confirmed this balanced performance, correctly identifying 218 out of 257 UNSAFE and 179 out of 256 SAFE compounds (Supplementary File 3).

The most relevant descriptors included LogP_VSA4, PEOE_VSA4, SMR_VSA3, BalabanJ, qed, and several EState and BCUT2D features, capturing key topological, hydrophobic, and electronic contributions associated with structural toxicity.

After training and validating the model, the XGBoost classifier was applied to Tb-Library to predict the chemical safety profile of each molecule which was then classified as either SAFE or UNSAFE.

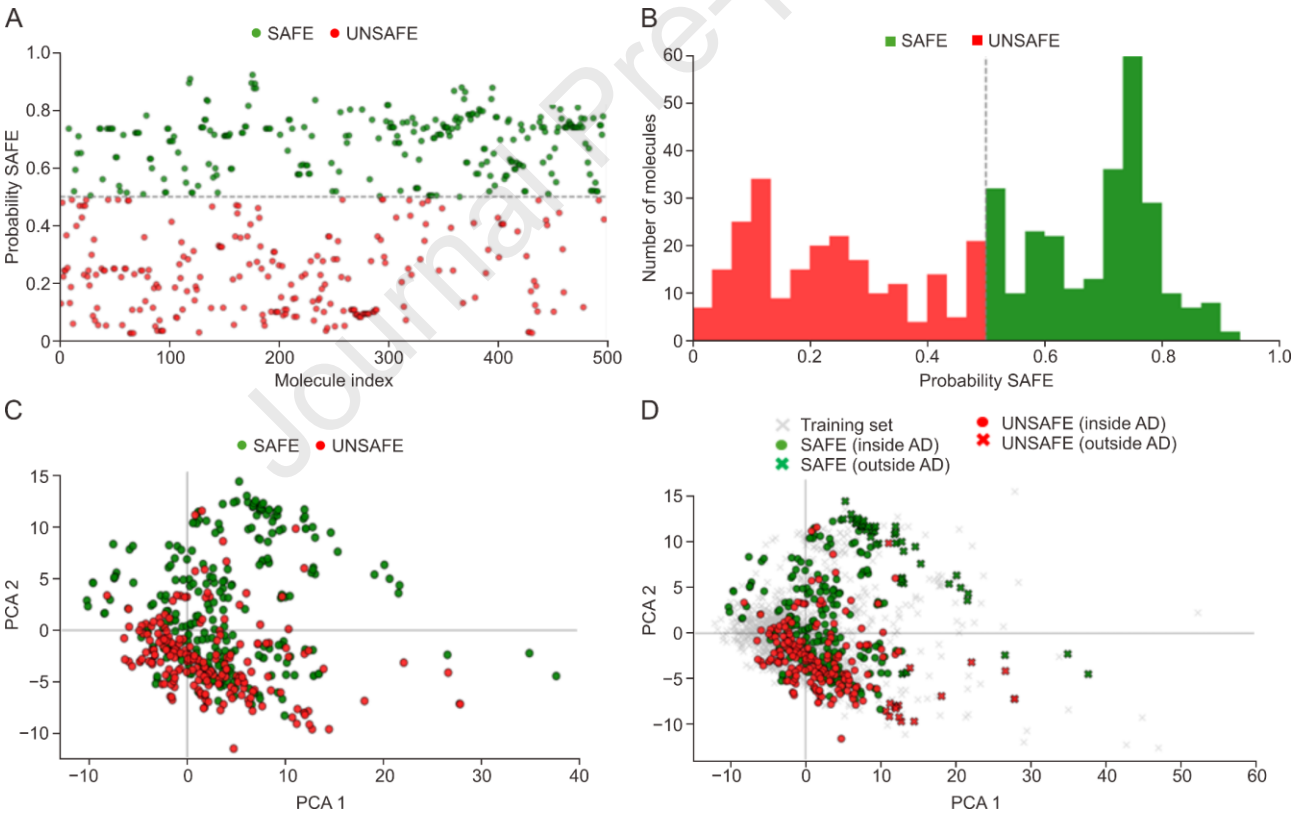


Fig. 6. Prediction probabilities and applicability domain assessment of the XGBoost classifier applied to the Tb-Library. (A) Predicted probabilities for all compounds in the test set, shown by molecule index. Green and red dots indicate compounds predicted as SAFE and UNSAFE, respectively, relative to the classification threshold (dashed line at 0.5). (B) Histogram of predicted SAFE probabilities by class,

showing the distribution of model confidence. (C) Principal component analysis (PCA) projection of predicted compounds coloured by predicted class (SAFE or UNSAFE). (D) PCA projection including the applicability domain (AD), defined by the chemical space of the training set (gray crosses). Predicted compounds are labelled according to both predicted class and AD status.

From a total of 497 compounds evaluated, the prediction results were slightly uneven toward the SAFE class, with 53.7% predicted as SAFE and 46.3% as UNSAFE. Importantly, the model produced high-confidence predictions, as illustrated in Fig. 6. Most compounds classified as SAFE had predicted probabilities well above the classification threshold of 0.5, while most compounds classified as UNSAFE clearly fell below this boundary. This sharp separation suggests the model operates with low uncertainty near the decision threshold, supporting its utility for virtual screening. To evaluate the reliability of these predictions in the context of the model chemical space, we performed an AD analysis using principal component analysis (PCA). Most predictions for both SAFE (80.8%) and UNSAFE (95.0%) compounds fell within the domain defined by the training data, indicating that most predictions were made under well-supported chemical conditions. However, 19.2% of the SAFE predictions were located outside the AD, highlighting the importance to exercise caution when interpreting confident predictions that lie at the periphery of the model's learned chemical space. Among the compounds predicted as SAFE (Probability > 0.5), 58.1% were natural products, while the remaining 41.9% were classified as synthetic, indicating a notable enrichment of natural-origin molecules within the top-ranked safe candidates.

3.3 Environmental toxicity and ADME-Tox parameters integration

To evaluate how the integration of environmental criteria influences compound prioritization, we designed and implemented the GDS, a modular scoring function that extends the previously reported ADMETscore[42]. The GDS scoring function is composed of three independent yet combinable blocks: ADMETscore, ECOscore, and DLscore each representing a critical dimension of compound evaluation spanning pharmacokinetic behaviour, toxicity risk, environmental hazard proxies, and drug-like properties. This modular architecture enables flexible adaptation to different screening objectives, ranging from risk-

aware filtering to comprehensive multiparametric optimization. The first component, ADMETscore (Eq. 2), is based on the framework originally proposed by Guan et al.[42] and was adapted to the present dataset by expanding the endpoint set to 21 binary predictions (see Table 3), including transporter activity (e.g., P-gp), cytochrome inhibition (e.g., CYP3A4), genotoxicity (Ames), cardiotoxicity (*h*ERG), and liver injury (DILI). Each endpoint contributes to the final score through a weight (w_i) defined as the product of three components (Eq. 3): w_1, an entropy-derived usefulness index (Eq. 4), w_2, the predictive accuracy of the model (e.g., AUC), w_3, the biological or regulatory relevance of the endpoint (Table 3), Supporting Information Tables S2–S7). The final score is computed as a normalized weighted average

$$ADMETscore = \frac{\sum_{i=1}^{n} w_i \times q}{\sum_{i=1}^{n} w_i} \quad \text{(Eq. 2)}$$

$$w_i = w_1 \times w_2 \times w_3 \quad \text{(Eq. 3)}$$

$$w_1 = [0.5 \times \left(1 - \left((-p) \times \log_2 p - (1-p) \times \log_2 (1-p)\right)\right)] + 0.5 \quad \text{(Eq. 4)}$$

Table 3: Endpoints used in the ADMETscore calculation and their accuracy of the prediction models. In the Table, weight (w_i) calculated as the product of three components (Eq. 3): w_1, an entropy-derived usefulness index (Eq. 4), w_2, the predictive accuracy of the model, and w_3, the biological or regulatory relevance of the endpoint.

No	Endpoints	w_1	w_2	w_3	w_i
1	Ames mutagenicity	0.500	0.796	1.000	0.398
2	Acute oral toxicity	0.581	0.832	1.000	0.483
3	Caco-2 permeability	0.501	0.743	0.800	0.298
4	Carcinogenicity	0.532	0.662	1.000	0.352
5	CYP1A2 inhibitor	0.541	0.864	0.500	0.233
6	CYP2C19 inhibitor	0.512	0.846	0.500	0.216
7	CYP2C9 inhibitor	0.563	0.853	0.500	0.240
8	CYP2C9 substrate	0.608	0.730	0.500	0.222
9	CYP2D6 inhibitor	0.712	0.880	0.500	0.313
10	CYP2D6 substrate	0.569	0.766	0.500	0.218
11	CYP3A4 inhibitor	0.545	0.835	0.800	0.364
12	CYP3A4 substrate	0.501	0.760	0.800	0.305

13	Drug-induced liver injury	0.578	0.825	1.000	0.476
14	HEK293	0.501	0.818	1.000	0.409
15	*h*ERG (10 μM)	0.500	0.698	1.000	0.349
16	*h*ERG	0.535	0.829	1.000	0.444
17	Human intestinal absorption	0.878	0.927	1.000	0.814
18	Organic cation transporter protein 2 inhibitor	0.827	0.808	0.500	0.334
19	P-gp inhibitor	0.508	0.841	0.800	0.341
20	P-gp substrate	0.598	0.831	0.800	0.397
21	Promiscuous	0.743	0.821	1.000	0.610

Caco-2: human colorectal adenocarcinoma cell line; CYP: cytochrome P450; HEK293: human embryonic kidney 293 cells; *h*ERG: human ether-à-go-go-related gene; P-gp: P-glycoprotein.

To integrate ecotoxicological criteria into the scoring framework, we developed the ECOscore (Eq. 5), a scoring block that incorporates four predicted ecotoxicological endpoints each capturing different aspects of aquatic toxicity and bioaccumulation across representative species. These parameters: bioaccumulation factor (BCF), 48 h *Tetrahymena pyriformis* (IGC_{50}), 48 h *Daphnia magna* ($LC_{50}DM$), and 96 h *fathead minnow* ($LC_{50}FM$), were chosen for their regulatory relevance, mechanistic complementarity, and availability as predictive outputs in platforms such as ADMETlab 3.0. As for ADMETscore, each endpoint contribution was weighted by combining an entropy-derived informativeness term, an endpoint-specific model-performance term, and a regulatory/biological relevance factor. (Table 4):

$$ECOscore = \frac{\sum_{i=1}^{n} w_i \times q}{\sum_{i=1}^{n} w_i} \quad \text{(Eq. 5)}$$

Table 4. Ecotoxicological endpoints used for ECOscore calculation, with associated accuracy (w_2), entropy (w_1), regulatory relevance (w_3), and final weight (w_i) used in the scoring formula.

No	Endpoints	w_1	w_2	w_3	w_i
1	Bioaccumulation factor (BCF)	0.553	0.786	0.350	0.152
2	48 h *Tetrahymena pyriformis* (IGC_{50})	0.928	0.890	0.150	0.124
3	48 h *Daphnia magna* ($LC_{50}DM$)	0.516	0.751	0.250	0.097
4	96 h *fathead minnow* ($LC_{50}FM$)	0.654	0.702	0.250	0.114

IGC_{50}: 50% growth inhibitory concentration; $LC_{50}DM$: median lethal concentration for Daphnia magna; $LC_{50}FM$: median lethal concentration for fathead minnow.

To further refine compound ranking, we implemented the DLscore (Eq. 6), which integrates a combination of normalized continuous descriptors (e.g., QED, logP, TPSA, SAScore, Fsp^3) and binary rule-based flags (Lipinski, PAINS, GASA). Continuous descriptors were scaled to the [0,1] range via min–max normalization and inverted when necessary to reflect ideal ranges. As these descriptors lack associated predictive models, w_2 was set manually to 1.0 or 0.8 depending on whether the value was computed or rule-based. The DLscore was calculated as:

$$DLscore = \frac{\sum_{i=1}^{n} w_i \times q}{\sum_{i=1}^{n} w_i} \quad \text{(Eq. 6)}$$

Table 5. DLscore components with weights for informativeness (w_1), confidence (w_2), relevance (w_3), and final contribution (w_i).

No	Endpoints	w_1	w_2	w_3	w_i
1	QED_norm	0.501	1.000	1.000	0.501
2	Fsp3_norm	0.508	1.000	0.800	0.406
3	TPSA_norm	0.626	1.000	0.800	0.501
4	logP_norm	0.508	0.950	0.800	0.386
5	SAScore_norm	0.528	1.000	0.800	0.422
6	gasa_norm	0.523	0.800	0.800	0.335
7	PAINS_bin	1.000	1.000	1.000	1.000
8	Lipinski_bin	1.000	1.000	1.000	1.000

QED: quantitative estimate of drug-likeness; Fsp3: fraction of sp3-hybridized carbon atoms TPSA: topological polar surface area; SAScore: synthetic accessibility score; GASA: genetic algorithm-based synthetic accessibility; PAINS: pan-assay interference compounds; norm: normalized value; bin: binary value.

The final GDS was computed in two configurations. The two-block version of the GDS scoring function integrates ADMET and ecotoxicological contributions as a weighted average:

$$GDS_{\{2-block\}} = \frac{w_{\{ADMET\}} \cdot ADMETscore + w_{\{ECO\}} \cdot ECOscore}{w_{\{ADMET\}} + w_{\{ECO\}}} \quad \text{(Eq. 7)}$$

To further refine compound prioritization by incorporating structural and physicochemical developability, we extended the model to a three-block version of the GreenDrugScore (GDS), integrating ADMETscore, ECOscore, and DLscore as follows:

$$GDS_{\{3-block\}} = \frac{w_{\{ADMET\}}.ADMETscore + w_{\{ECO\}}.ECOscore + w_{\{DL\}}.DLscore}{w_{\{ADMET\}} + w_{\{ECO\}} + w_{\{DL\}}} \quad \text{(Eq. 8)}$$

For all the safe predicted compounds (Probability >0.5) all the scoring functions were applied and ADMETscore, ECOscore and DLscore (Supplementary material S8–S10) alongside $GDS_{2\text{-block}}$ (ADMET + ECO) and $GDS_{3\text{-block}}$ (ADMET + ECO + DL) were calculated (Table S8). Among the 267 compounds with a SAFE probability above 0.5, the average ADMETscore was 0.57 ± 0.11, consistent with more favourable predicted pharmacokinetic and toxicity profiles within this subset. The ECOscore block yielded an average of 0.74 ± 0.28, indicating that compounds classified as SAFE also tended to display more favourable predicted ecotoxicological profiles across the selected endpoints. Finally, the DLscore showed an average of 0.48 ± 0.04, suggesting only moderate alignment with conventional drug-likeness criteria and highlighting the chemical heterogeneity of the SAFE subset. Overall, these patterns support the internal coherence of the three-block framework, while remaining predictive in nature and subject to experimental confirmation. (Supplementary File 4).

3.4 Ranking analysis and prioritization

In many conventional *in-silico* workflows, ADMET and ecotoxicological endpoints are typically evaluated independently but are not formally integrated into a single decision layer. Here, we enable a direct comparison between prioritization strategies with and without explicit environmental-aware criteria by contrasting a conventional ADMET-based ranking with a GreenDrugScore-driven ranking. Importantly, our goal is not to improve the predictive performance of individual endpoints, but to quantify how their formal integration changes prioritization decisions at both compound and scaffold levels. As shown below, incorporating the ECOscore can induce measurable re-ranking patterns that are less evident when endpoints are considered in isolation, providing a transparent framework to evaluate the practical consequences of

ecotoxicology-aware prioritization in early-stage drug discovery. Within the subset of compounds classified as SAFE by the machine learning model (predicted probability > 0.5), 58.1% were natural products and 41.9% were synthetic. This modest enrichment of natural compounds suggests that biologically favourable profiles may be more prevalent in natural scaffolds (Fig. 7A). The ADMETscore distribution further supports this trend: natural compounds exhibit a broader and right-shifted profile compared to synthetics, with higher mean and median values, indicating a generally favourable predicted pharmacokinetic and toxicity profile on average (Fig. 7B, and Table S9). Importantly, incorporating the ECOscore in $GDS_{2\text{-block}}$ introduced meaningful re-ranking. Approximately 28.5% of the SAFE compounds experienced a decrease in rank ($\Delta = GDS_{2\text{-block}} - ADMETscore < 0$), whereas 71.5% improved (Fig. 7C). The ADMETscore versus $GDS_{2\text{-block}}$ scatter plot showed a strong linear association (Pearson r = 0.989), yet with sufficient dispersion to reveal rank shifts for compounds with less favourable predicted ecotoxicological profiles (Fig. 7D).

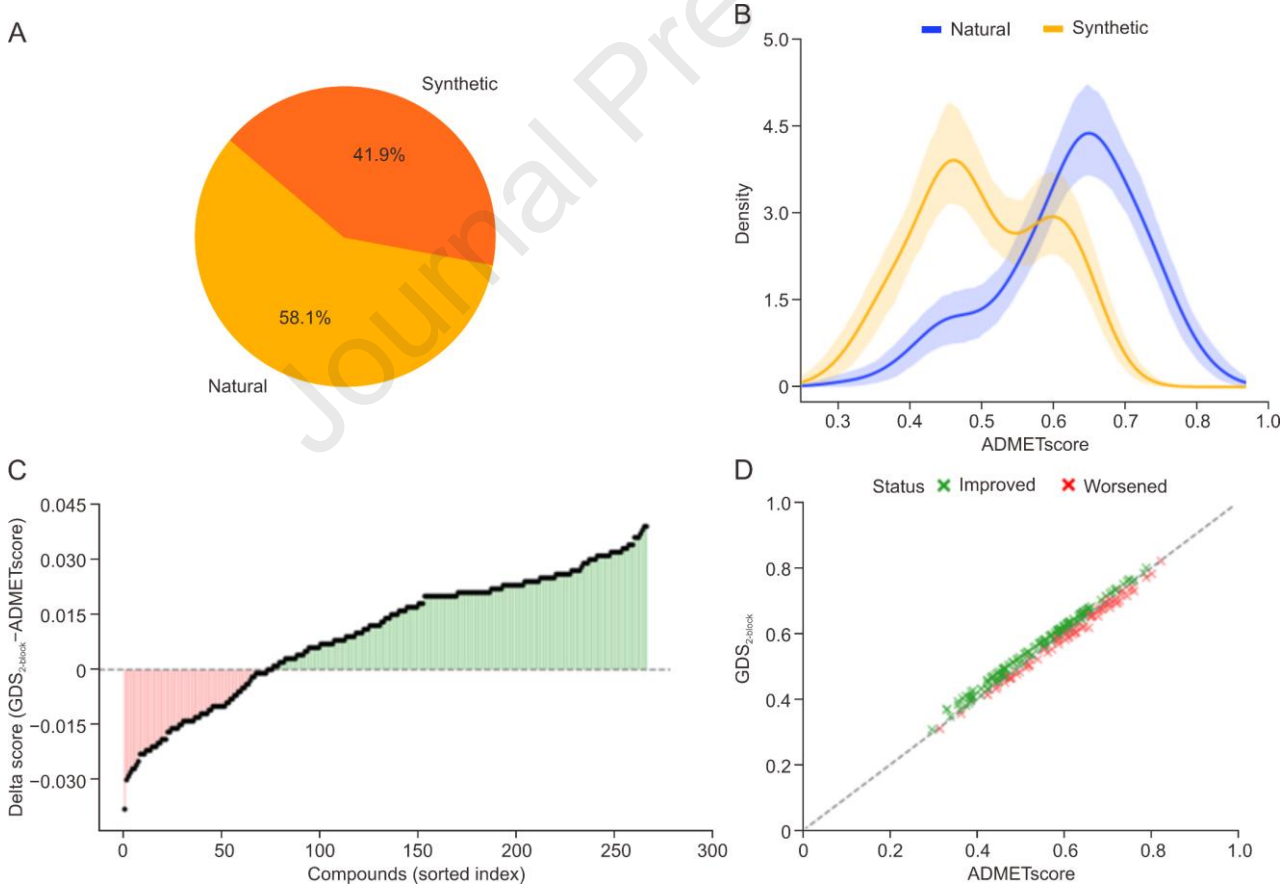


Fig. 7. Comparative analysis of ADMETscore and GreenDrugScore ADMETscore+ECOscore ($GDS_{2\text{-block}}$) ranking and the impact of ECOscore integration. (A) Distribution of SAFE compounds (probability > 0.5)

by origin, with a higher proportion of natural products. (B) ADMETscore distributions for natural vs. synthetic compounds with confidence intervals, showing higher scores on average for synthetics. (C) Lollipop plot of the score delta ($GDS_{2\text{-}block}$ − ADMETscore), highlighting compounds that are promoted (green) or penalized (red) after ECOscore inclusion. (D) Scatter plot comparing ADMETscore and $GDS_{2\text{-}block}$ values; green and red markers indicate compounds whose rank improved or worsened, respectively, when ecotoxicological criteria were added.

To assess whether the integration of ecotoxicological parameters (ECOscore) in the $GDS_{2\text{-}block}$ score affects compound prioritization, we applied the Mann–Whitney U test to the delta scores ($GDS_{2\text{-}block}$ − ADMETscore) across different subgroups. The analysis was performed on all compounds with a predicted Probability > 0.5, classified as SAFE by the model (Table S10). The comparison between natural and synthetic compounds revealed a statistically significant difference (U = 6572.5, P = 0.0007), suggesting that natural compounds are generally less penalized, or even favoured, by the ECOscore, while synthetic compounds tend to show a greater reduction in prioritization due to environmental concerns. A second comparison, based on high vs. low predicted safety probability (above or below the median), also showed a significant shift (U = 7306.0, P = 0.011) suggesting that compounds with stronger model confidence for the SAFE class can still experience ECOscore-driven deprioritization. Together, these results support the relevance of incorporating ecotoxicological criteria into early-stage compound selection, allowing a more balanced evaluation that accounts for both efficacy and environmental sustainability. From a practical standpoint, the integration of ECOscore into the prioritization process had direct consequences on compound selection. When examining the top 10% of compounds ranked by ADMETscore vs. $GDS_{2\text{-}block}$, only 22 out of 26 compounds overlapped, implying a 15% reshuffling of top-ranked hits due to the predicted ecotoxicological endpoints (Fig. 8A). Notably, we identified 76 compounds (Supplementary File 5) that were deprioritized due to an unfavourable predicted ECOscore, many of which are natural products with high ADMETscore and high predicted probability (Fig. 8B). Although these molecules appear promising based on ADMETscore and model confidence for SAFE class, their lower $GDS_{2\text{-}block}$ ranks reflect less favourable predicted ecotoxicological profiles across the selected endpoint. Overall, ECOscore integration adds an explicit ecotoxicology-aware decision layer beyond ADMET alone, enabling transparent

comparison of prioritization strategies and clearer trade-off analysis within a One Health-oriented screening framework consistent with increasing interest in incorporating environmental considerations earlier in pharmaceutical R&D.

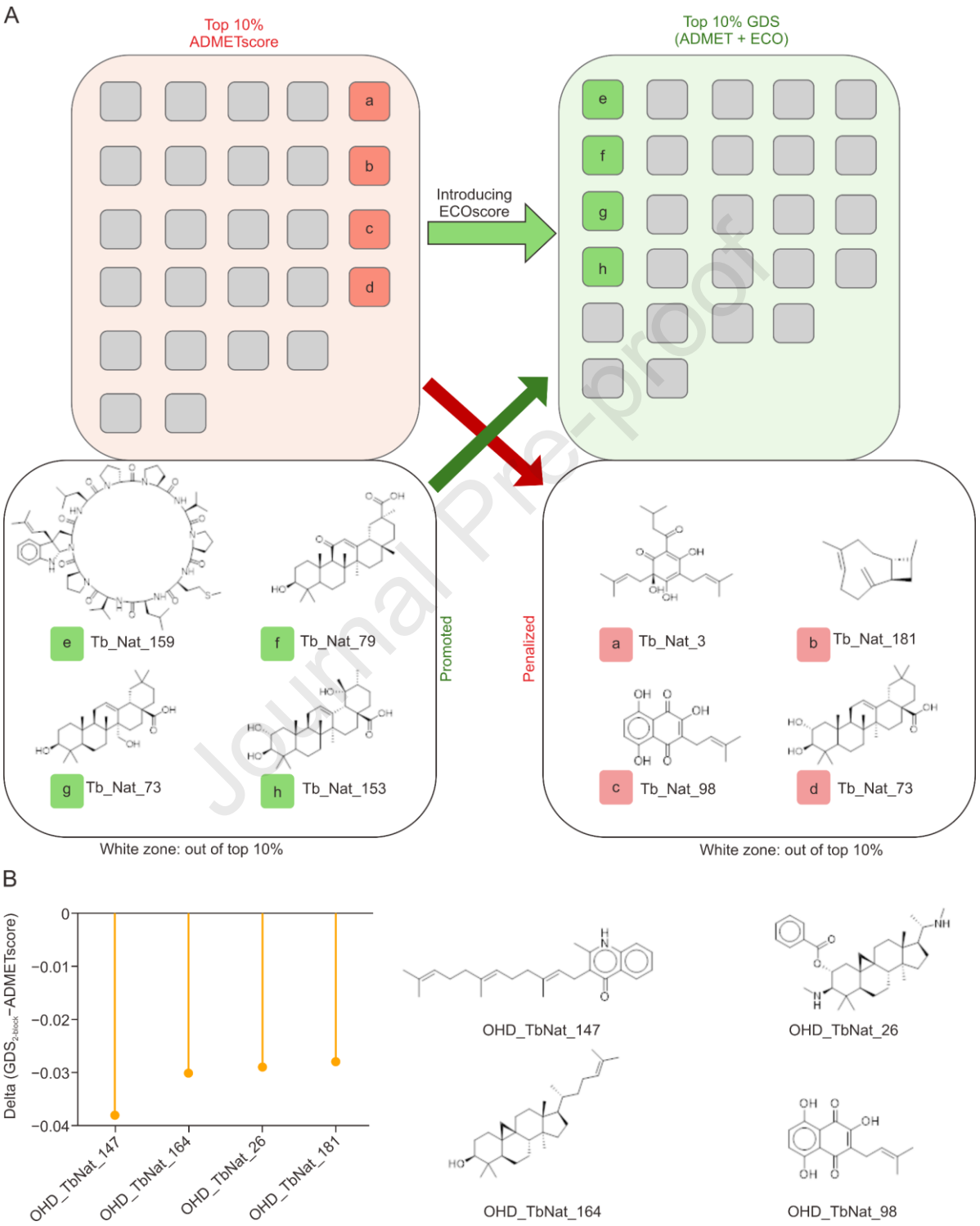


Fig. 8. Impact of ECOscore integration on compound prioritization and top 4 compounds most penalized by ECOscore. (A)The left panel shows the top 10% of compounds ranked by ADMETscore alone ($n = 26$),

while the right panel shows the top 10% after integrating ecotoxicological parameters via GreenDrugScore ADMETscore + ECOscore (GDS ADMET + ECO). Upon inclusion of the ECOscore, 4 compounds (a–d) are penalized and excluded from the top ranks due to unfavourable environmental predictions. Conversely, 4 new compounds (e–h) are promoted into the top 10% for their predicted eco-friendly profiles. Most compounds (n = 22) are shared between both rankings. (B) Delta values represent the difference between the $GDS_{2\text{-block}}$ and ADMETscore for each compound. Negative values indicate a reduction in overall prioritization due to unfavourable ecotoxicological predictions, highlighting the impact of ECOscore integration on hit selection.

3.5 Identification and analysis of prioritized scaffolds with favourable predicted ecotoxicological profiles

To identify chemotypes with safer predicted environmental properties within the prioritized chemical space, we selected the top 100 compounds ranked by the $GDS_{2\text{-block}}$ and subjected them to scaffold-level analysis. Initially, these compounds were encoded using ECFP4 fingerprints and clustered hierarchically based on Tanimoto similarity (Fig. S11). We then performed a scaffold decomposition using the Bemis–Murcko formalism and constructed a scaffold network in which nodes represented unique scaffolds and edges encoded structural relationships, including fragment, initialize, and generic connections (Fig. 9A, and Supplementary File 6). This network allowed us to explore the organization of core chemical structures and to identify scaffolds that recurrently appeared among those compounds exhibiting favourable ecotoxicological profiles. From this network, 26 chemically diverse scaffolds were identified (Table S11) based on topological relevance and structural representativeness. Notably, 19 of them stood out for their consistently favourable predicted ecotoxicological properties, qualifying them as good candidates for further prioritization. These scaffolds, labelled a to s in the (Fig. 9B), encompassed a broad range of chemical features. Several structures contained fused nitrogen-containing heterocycles and polycyclic systems such as tetrahydro-azepines (a), decahydronaphthalenes (b), and chromenones (j), while others included oxygenated rings like di-, tetrahydro- furans (d, f, i), and tetrahydropyrans (g) heteroaromatic scaffolds such as pyrrolo[2,3-d]pyrimidines (f, p), indoles (q), and N-phenyl benzamides (r) were also

present, highlighting their frequent occurrence in antiparasitic drug-like space. The set further included piperazines with electron-withdrawing substituents (k), as well as imidazolidines (d, l, o), isoquinolines (h), and simpler heteroaromatics, such as pyrrolines (m, n) and dihydropiridines (e). Each of the 19 scaffolds was independently evaluated for its predicted ecotoxicological impact (Fig. 9C) using the same criteria applied at the molecular level. Among them, 16 scaffolds (e.g., a, d, e, f, g, h, j, k, l, m, n, o, p, q, r, s) demonstrated predicted favourable profiles across all endpoints, exhibiting low predicted BCF values and high IGC_{50} and LC_{50} thresholds, consistent with reduced bioaccumulation and aquatic toxicity. The remaining three scaffolds (b, c and i) displayed moderate ecotoxicological concerns, associated primarily with borderline BCF or LC_{50} values, and were deprioritized accordingly. This scaffold-level dissection of the GDS-ranked compound space supports the identification of chemotypes with favourable predicted ecotoxicological profiles while retaining developability-relevant structural features. The resulting scaffold panel provides hypothesis-generating starting points for medicinal chemistry follow-up aimed at optimizing antiparasitic candidates under a One Health–oriented, ecotoxicology-aware prioritization framework.

Journal Pre-proof

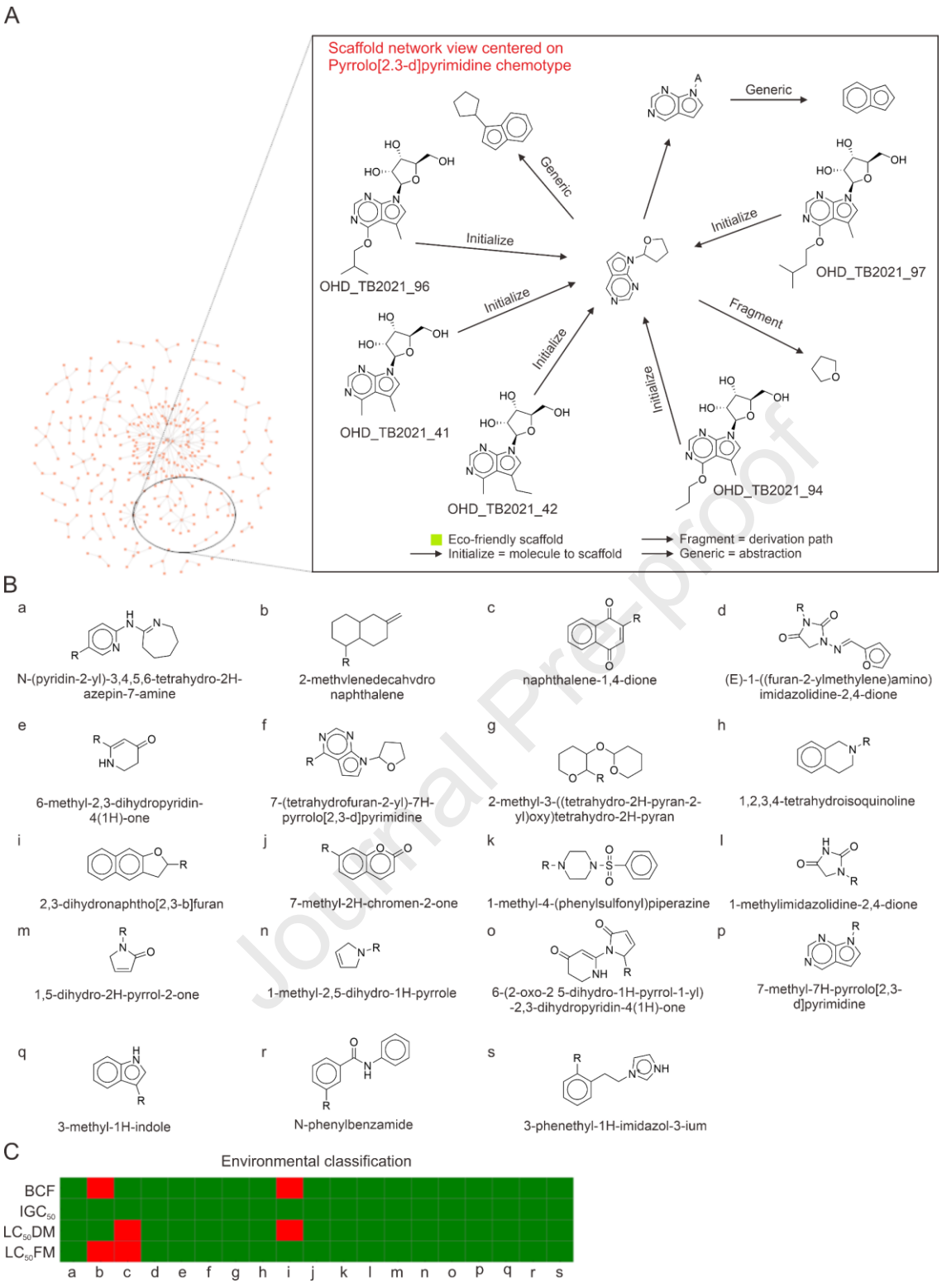


Fig. 9. Scaffold network focussed on the pyrrolo[2,3-d]pyrimidine chemotype identified from GDS-prioritized compounds. (A) This representative network illustrates the structural relationships between a scaffold (center, highlighted in green) and surrounding molecules. Arrows indicate the type of

transformation used in the scaffold network: Initialize (compound-to-scaffold mapping), Fragment (structural simplification) and Generic (scaffold abstraction). Several drug-like molecules (OHD_TB2021 series) converge on this core, supporting its relevance as a predicted green chemotype. (B) Chemical structures (a–s) represent the 19 scaffolds emerging from network analysis and decomposition of the top 100 GDS-ranked compounds. (C) The binary heatmap shows the ecotoxicological classification (green = favourable, red = unfavourable) across four endpoints: bioaccumulation factor (BCF), 50% growth inhibitory concentration (*Tetrahymena pyriformis*, IGC_{50}), median lethal concentration for *Daphnia magna* ($LC_{50}DM$), and median lethal concentration for fathead minnow ($LC_{50}FM$). Most scaffolds display consistent predicted environmental safety profiles, supporting their relevance as green chemotypes for future design efforts.

3.6 Exploration of prediction set in public curated databases

Next, the prediction set of 497 compounds (Tb-Library) was used as a query set to the EPA computational toxicology and exposure online resources (CompTox)[60]. Although experimental ecotoxicological measurements were not generated within this study, comparison with EPA CompTox classifications was employed as an external consistency check. Observed concordances and discrepancies reflect differences in model scope and endpoint definition rather than predictive failure and provide contextual support for the integration of ecotoxicological prediction at early discovery stages, particularly for compounds that remain poorly characterized experimentally. CompTox hosts several tools and dashboards for providing chemistry, toxicity and exposure data, thereby, assisting research activities and evaluation of chemical information. Upon searching CompTox with SMILES of the chemicals, we found that only 60 chemicals (38 SAFE and 22 UNSAFE) were represented in EPA database. Out of these, only 9 were synthetic chemicals, while the remaining 51 were natural products. Most chemicals were new as observed from missing CAS identifiers. This also suggested that these chemicals are not catalogued in public databases such as PubChem and ChEMBL and their corresponding data such as bioactivity, protein target, quantified values in assays including ecotoxicity are yet to be reported.

The 60 chemicals identified were subjected to generation of hazard profiles using CompTox chemoinformatic modules (Supplementary File 7). The analysis also simultaneously generated safety profiles, however, we do not present the results here considering the scope of this study. We found that the hazard profiles (i.e., very high, high, medium and low) were categorized into human health effects, ecotoxicity and QSAR predicted environmental fate endpoints. Again, considering the scope of this study, we explored only the ecotoxicity and fate endpoints of the chemicals. The ecotoxicity was sub-categorized as acute and chronic aquatic toxicity, where the information about acute toxicity was more prevalent than the chronic toxicity. To this, 13 and 15 SAFE chemicals were labelled as having high and very high acute toxicity, respectively. Moreover, only 1 and 6 SAFE compounds were labelled as medium and low acute toxicity. The labels for 22 UNSAFE compounds, however, were almost evenly distributed for acute aquatic toxicity. Likewise, 1, 4 and 7 SAFE chemicals were labelled as having medium, high and very high chronic toxicity, respectively, whereas no data was available for rest of the SAFE classified chemicals. In case of UNSAFE compounds, only 3 chemicals (i.e., 1 low and 2 very high) had the data for chronic aquatic toxicity. Interestingly, within QSAR fate endpoints we observed availability of bioaccumulation data, which was one of the properties considered in our prediction model. We found that the bioaccumulation for 21 SAFE chemicals was low with only 1 SAFE compound with very high bioaccumulation. Whereas 13 and 4 UNSAFE chemicals were labelled as having low and medium bioaccumulation, respectively.

We also observed for some chemicals, irrespective of being SAFE or UNSAFE, that the overall availability of hazard data (i.e., human health effects, ecotoxicity and QSAR fate endpoint) was better than the rest. Upon reviewing the number of patents and articles, we found that better data availability corresponds to a higher number of patents and article counts (Supplementary File 8). For instance, thymol and myrcene, both of which are SAFE, have a total of 86722 and 27601 patents and 12869 and 5618 articles, respectively. Likewise, apigenin, which is predicted as UNSAFE, has a total of 20907 patents and 11578 articles. Along these lines, we also discerned that the higher number of patents and articles also corresponds to the functional uses (antioxidants, preservatives, drugs, additives, etc.) of the chemicals. Lastly, we filtered the SAFE chemicals from the list and retrieved corresponding human biology-specific knowledge such as disease areas, target proteins, side effects, etc using the knowledge graph generator (KGG) workflow[61] (Fig. S12). We found that the knowledge for these chemicals in public databases, such as ChEMBL and

UniProt, is very sparse as indicated by 6 chemicals mapped to the databases. Out of these, only 3 chemicals were used in treating diseases whereas the remaining 3 chemicals still were pre-clinical. Dactolisib (CHEMBL1879463), with kinase inhibiting mechanism of action, targeted PIK3CA in several diseases including infections, SARS and various cancer sub-types. Borneol (CHEMBL486208) with no reported mechanism of action and target was used to treat stroke and cognitive dysfunction whereas thymol (CHEMBL29411) was used against gonorrhoea, tooth diseases and gingivitis.

4. Discussion

These results collectively illustrate how the proposed workflow integrates predictive modelling, multiparametric scoring, and structural analysis to support ecotoxicology-aware hit prioritization against *T. brucei*. The initial characterization of the Tb-Library showed that the dataset covers a broad and balanced chemical space, combining natural and synthetic scaffolds with physicochemical properties largely compliant with drug-likeness rules. This diversity provided a suitable foundation for developing predictive models capable of evaluating both pharmacological and ecological attributes.

The machine-learning step demonstrated that structural descriptors alone can effectively capture features related to chemical safety. Among the tested algorithms, XGBoost achieved the most reliable balance between sensitivity and specificity, validating its suitability for early virtual screening. The classifier produced consistent SAFE/UNSAFE predictions across a heterogeneous set of scaffolds, and the enrichment of natural products within the SAFE category suggested a relationship between scaffold rigidity, reduced lipophilicity, and lower predicted toxicity liabilities. These outcomes provided the computational basis for integrating environmental and pharmacokinetic dimensions within a unified prioritization strategy.

The introduction of the GDS expanded the prioritization framework beyond conventional ADMET evaluation. By coupling pharmacokinetic and toxicological descriptors (ADMETscore) with ecotoxicological endpoints (ECOscore) and drug-likeness parameters (DLscore), the GDS enables an explicit trade-off analysis between ADMET/drug-likeness criteria and predicted ecotoxicological endpoints within a single prioritization metric. The resulting multiparametric structure revealed that molecules with excellent pharmacological profiles could nonetheless be deprioritized when ecological factors are

considered, highlighting potential ecotoxicological concerns under the selected predicted endpoints, commonly overlooked in traditional discovery pipelines.

The comparison between ADMET-based and GDS-based rankings revealed the effect of integrating ecotoxicological criteria: nearly one-third of compounds changed position, and natural products tended to retain or even improve their prioritization, due to inherently lower predicted bioaccumulation and aquatic toxicity. This re-ranking highlights the importance of incorporating ecological indicators from the earliest phases of compound triage. Rather than conflicting with pharmacological optimization, environmental filters provide an additional layer of selectivity that aligns chemical prioritization with the One Health perspective.

Finally, scaffold-level analysis refined this integration by identifying recurrent chemotypes characterized by biological relevance and favourable predicted environmental profiles. The 19 scaffolds emerging from this process display favourable predicted ecotoxicological profiles across the selected endpoints, including heteroaromatic, polycyclic, and oxygenated frameworks, suggest that integrating ecotoxicological endpoints with ADMET constraints in a unified prioritization scheme can highlight chemotypes with environmentally favourable predicted profiles within established medicinal-chemistry space, without requiring wholesale scaffold redesign. Taken together, these findings reinforce the concept of sustainable molecular innovation and support the utility of the proposed workflow as an integrated, multiparametric decision framework in which pharmacological performance, developability, and predicted environmental impact are jointly considered. Overall, this workflow-level integration provides a practical basis for early-stage decision-making, enabling ecotoxicological considerations to be embedded alongside ADMET constraints during hit and scaffold prioritization.

In the final stage, we performed a cross-check against public repositories (EPA CompTox and ChEMBL) to contextualize dataset coverage and assess the availability of experimental annotations relevant to toxicity and bioaccumulation, rather than to validate predictive performance. Only 60 of the 497 compounds were found in CompTox, and most of these entries lacked experimental ecotoxicological measurements, highlighting the limited public availability of eco-annotated data for the chemotypes considered here and supporting the novelty/poor characterization of the library. Where discrepancies were observed (e.g., compounds classified as SAFE by our model that are associated with acute-toxicity hazard labels in

CompTox), these are expected given differences in scope and label definition: CompTox categories typically derive from broad QSAR/hazard schemes that are not equivalent to our SAFE/UNSAFE criteria, which are trained on a distinct reference set and optimized for early-stage prioritization. Overall, this cross-check provides external contextualization and reinforces the need for additional experimental ecotoxicological data to improve model learning, expand applicability-domain coverage, and progressively translate prediction-driven prioritization into evidence.

4.1 Limitations of the proposed approach

This prioritization strategy relies on predicted ecotoxicological endpoints and should be interpreted accordingly. As with all QSAR and machine learning–based approaches, the reliability of the predictions is inherently constrained by the quality, coverage, and representativeness of the underlying training data. Importantly, the aim of the present workflow is not to replace experimental testing, but to rationally reduce chemical space and support early-stage prioritization by identifying candidates with balanced pharmacological and environmental profiles. In this context, the computational identification of potentially eco-compatible scaffolds should be viewed as a hypothesis-generating step, intended to guide and inform subsequent *in vitro* and *in vivo* validation rather than to anticipate definitive safety assessments.

5. Conclusion

This study introduces a modular prioritization workflow (Fig. 10) that integrates ecotoxicological parameters into the early stage of the drug discovery process, exemplified through a curated library of compounds active against *T. brucei*. By coupling ML classification with the GreenDrugScore, a composite scoring function combining ADME, toxicity, ecotoxicity and drug-likeness, in comparison with ADMET-focused metrics classification approach, we showed prioritization that can change materially when ecotoxicological criteria are incorporated. We observed that 28.5% of compounds ranked highly by ADMET-focused metrics were deprioritized after ECOscore integration, indicating potential ecotoxicological liabilities among hits that would otherwise appear promising based on the selected predicted endpoints. From the top-ranked candidates, 19 structurally diverse scaffolds with favourable

predicted profiles across the selected ecotoxicological endpoints were identified as hypothesis-generating starting points for follow-up in medicinal chemistry program.

To contextualize data availability and external coverage, the dataset was cross-checked against public repositories, such as EPA CompTox and ChEMBL. Only 60 of 497 molecules were found in CompTox, and most lacked experimental ecotoxicological annotations, highlighting both the limited overlap with existing resources and the scarcity of publicly available eco-annotated data for these chemotypes. The partial mismatch between our SAFE/UNSAFE outputs and CompTox hazard labels is consistent with differences in scope and label definition (endpoint-specific predictions in this work versus broader hazard categorizations in CompTox) and suggests that the two resources can provide complementary decision support.

Importantly, the scaffolds discussed here represent recurring chemical cores among prioritized compounds rather than experimentally validated chemotypes. Their identification is therefore intended as a hypothesis-generating outcome that highlights chemotype-level starting points for medicinal chemistry campaigns. In future programs, systematic chemical elaboration around these cores will enable the generation of biologically active derivatives and will provide a practical basis for experimental confirmation of antiparasitic efficacy together with targeted ecotoxicological measurements. As a perspective, building on the hypothesis-generating nature of the workflow outlined above, to translate prediction-driven prioritization into evidence, we envisage a tiered validation strategy in which a focused subset of top-ranked and discordant-ranked compounds (e.g., high antiparasitic potency but unfavourable predicted ecotoxicity, and vice versa) is first selected to stress-test the ranking logic and evaluated in standardized acute aquatic assays aligned with the endpoints used here (algal growth inhibition, Daphnia immobilization, and fish acute toxicity). Confirmatory testing will then prioritize the most promising chemotypes and resolve borderline cases, complemented by exposure-relevant measurements (aqueous solubility, ionization state, and stability) to support an initial link between hazard proxies and realistic exposure scenarios. In parallel, curated eco-annotated datasets will be expanded to improve chemical-space coverage, enable applicability-domain assessment, and support uncertainty/confidence reporting. Importantly, GreenDrugScore is intended to be reused iteratively during scaffold optimization to provide a transparent, prediction-based readout of how structural modifications influence prioritization under a One Health–oriented set of

constraints. Overall, embedding ecotoxicology-aware filters early in discovery supports more explicit multi-objective decision-making while providing a realistic pathway toward experimental confirmation.

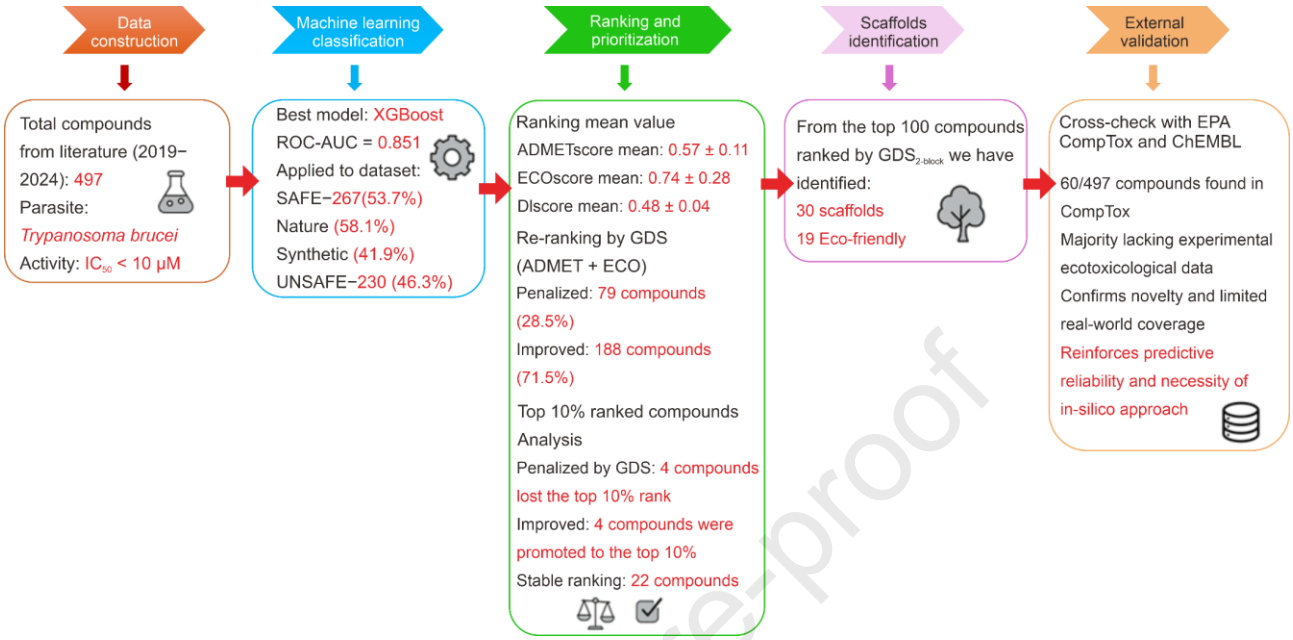


Fig. 10. Modular workflow for the early-stage prioritization of antiparasitic compounds with predicted environmentally favourable profiles. From 497 *T. brucei*-active molecules with half-maximal inhibitory concentration (IC_{50}) values below 10 µM, 267 were predicted as SAFE by the machine learning model, with an area under the receiver operating characteristic curve (ROC-AUC) of 0.851 for the extreme gradient boosting (XGBoost) classifier. SAFE and UNSAFE indicate the predicted safe and unsafe compound classes, respectively. SAFE compounds were ranked using the GreenDrugScore (GDS), which integrates the absorption, distribution, metabolism, excretion and toxicity score (ADMETscore), ecotoxicological score (ECOscore), and drug-likeness score (DLscore). Thirty scaffolds were extracted, including 19 with favourable predicted ecotoxicological profiles. External comparison with the US Environmental Protection Agency (EPA) CompTox Chemicals Dashboard and ChEMBL highlighted the limited public coverage of the dataset and supported contextual interpretation of the *in-silico* prioritization workflow.

Data availability

The curated *Trypanosoma brucei* compound library, calculated molecular descriptors, machine learning predictions, and prioritization results generated in this study are provided in the Supplementary material. In addition, all datasets and analysis files used to generate the reported

results are openly available at the following repository: https://github.com/BreakDan/OHD-Prioritization-Paper.git

The repository includes the following Supplementary Files:

- Supplementary File 1: Tb-Library
- Supplementary File 2: Machine learning training dataset
- Supplementary File 3: Tb-Library SAFE/UNSAFE classification
- Supplementary File 4: Tb-Library GreenDrugScore rankings
- Supplementary File 5: Compounds downgraded after ECOscore integration
- Supplementary File 6: Scaffold network analysis (top 100 compounds)
- Supplementary File 7: EPA CompTox query results
- Supplementary File 8: Compound hazard and profiling data

Custom Python scripts used for data processing, model training, scoring, and analysis are available within the same repository and are fully documented to ensure reproducibility.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

The authors acknowledge the use of ChatGPT (OpenAI) to assist in improving the readability and language quality of the manuscript. All generated text was carefully reviewed, revised, and approved by the authors, who take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the European Cooperation in Science and Technology (COST) under the COST Action CA21111 *OneHealthdrugs – One Health drugs against parasitic vector-borne diseases in Europe and beyond*.

Journal Pre-proof

Journal Pre-proof

6. References

[1] N. Singh, P. Vayer, S. Tanwar, et al., Drug discovery and development: Introduction to the general public and patient groups, Front. Drug Discov. 3 (2023), 1201419.
[2] H. Yu, A. Adedoyin, ADME-Tox in drug discovery: Integration of experimental and computational technologies, Drug Discov. Today 8 (2003) 852–861.
[3] J.C. Semenza, S. Paz, Climate change and infectious disease in Europe: Impact, projection and adaptation, Lancet Reg. Health Eur. 9 (2021), 100230.
[4] L. Gunnarsson, J.R. Snape, B. Verbruggen, et al., Pharmacology beyond the patient–The environmental risks of human drugs, Environ. Int. 129 (2019) 320–332.
[5] A.B.A. Boxall, J.L. Wilkinson, A. Bouzas-Monroy, Medicating nature: Are human-use pharmaceuticals poisoning the environment? One Earth 5 (2022) 1080–1084.
[6] C. Manyi-Loh, S. Mamphweli, E. Meyer, et al., Antibiotic use in agriculture and its consequential resistance in environmental sources: Potential public health implications, Molecules 23 (2018), 795.
[7] B.A. Halley, W.J.A. VandenHeuvel, P.G. Wislocki, Environmental effects of the usage of avermectins in livestock, Vet. Parasitol. 48 (1993) 109–125.
[8] B.A. Halley, R.J. Nessel, A.Y.H. Lu, Environmental aspects of ivermectin usage in livestock: General considerations. Ivermectin and Abamectin. Springer New York, (1989), pp 62–172.
[9] B. Martinengo, E. Diamanti, E. Uliassi, et al., Harnessing the 12 green chemistry principles for sustainable antiparasitic drugs: Toward the one health approach, ACS Infect. Dis. 10 (2024) 1856–1870.
[10] L. Mancini, I. Lacchetti, F. Chiudioni, et al., Need for a sustainable use of medicinal products: environmental impacts of ivermectin, Ann. Ist. Super. Sanita 56 (2020) 492–496.
[11] L.M. Mesa, I. Lindt, L. Negro, et al., Aquatic toxicity of ivermectin in cattle dung assessed using microcosms, Ecotoxicol. Environ. Saf. 144 (2017) 422–429.
[12] R.B. de Souza, J.R. Guimarães, Effects of avermectins on the environment based on its toxicity to plants and soil invertebrates: A review, Water Air Soil Pollut. 233 (2022), 259.
[13] M. Bilal, A.R. Bagheri, P. Bhatt, et al., Environmental occurrence, toxicity concerns, and remediation of recalcitrant nitroaromatic compounds, J. Environ. Manag. 291 (2021), 112685.
[14] L. Xu, Three-dimensional structural features and the toxicity of aminobenzenes and phenols, Sci. China Ser. B 46 (2003), 431.
[15] M.A. Ungogo, H.P. de Koning, Drug resistance in animal trypanosomiases: Epidemiology, mechanisms and control strategies, Int. J. Parasitol. Drugs Drug Resist. 25 (2024), 100533.
[16] K.I. Kasozi, E.T. MacLeod, I. Ntulume, et al., An update on African trypanocide pharmaceutics and resistance, Front. Vet. Sci. 9 (2022), 828111.
[17] J. Hidalgo, J.F. Ortiz, S.P. Fabara, et al., Efficacy and toxicity of fexinidazole and nifurtimox plus eflornithine in the treatment of African trypanosomiasis, Cureus (2021), e16881.
[18] G.L. da Silva Oliveira, A.P. dos Santos C.L. da Silva, Evaluation of the non-clinical toxicity of an antiparasitic agent: Diminazene aceturate, Drug Chem. Toxicol. 45 (2022) 2003–2013.
[19] Environmental risk assessment of medicinal products for human use - Scientific guideline | European Medicines Agency (EMA), (2006). https://www.ema.europa.eu/en/environmental-risk-assessment-medicinal-products-human-use-scientific-guideline.
[20] M. Breitholtz, C. Rudén, S. Ove Hansson, et al., Ten challenges for improved ecotoxicological testing in environmental risk assessment, Ecotoxicol. Environ. Saf. 63 (2006) 324–335.
[21] C.H. Walker, Ecotoxicity testing of chemicals with particular reference to pesticides, Pest Manag. Sci. 62 (2006) 571–583.
[22] Estimating Persistence, Bioaccumulation, and Toxicity using the PBT Profiler, (2012). https://www.epa.gov/sites/default/files/2015-05/documents/07.pdf.
[23] A. Tropsha, O. Isayev, A. Varnek, et al., Integrating QSAR modelling and deep learning in drug discovery:

The emergence of deep QSAR, Nat. Rev. Drug Discov. 23 (2024) 141–155.
[24] P. Peets, W.C. Wang, M. MacLeod, et al., MS2Tox machine learning tool for predicting the ecotoxicity of unidentified chemicals in water by nontarget LC-HRMS, Environ. Sci. Technol. 56 (2022) 15508–15517.
[25] J.P. Zubrod, N. Galic, M. Vaugeois, et al., Bio-QSARs 2.0: Unlocking a new level of predictive power for machine learning-based ecotoxicity predictions by exploiting chemical and biological information, Environ. Int. 186 (2024), 108607.
[26] X. Shi, Z. Wang, Y. Wang, et al., AquaticTox: A web-based tool for aquatic toxicity evaluation based on ensemble learning to facilitate the screening of green chemicals, Environ. Health 2 (2024) 202–211.
[27] L. Fu, S. Shi, J. Yi, et al., ADMETlab 3.0: An updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support, Nucleic Acids Res. 52 (2024) W422–W431.
[28] Y. Gu, Z. Yu, Y. Wang, et al., admetSAR3.0: A comprehensive platform for exploration, prediction and optimization of chemical ADMET properties, Nucleic Acids Res. 52 (2024) W432–W438.
[29] A. Roncaglioni, A. Lombardo, E. Benfenati, The VEGAHUB platform: The philosophy and the tools, Altern. Lab. Anim. 50 (2022) 121–135.
[30] S. Gowri, K.S. Balamurugan, A.G. Noorul Julaiha, et al., Comprehensive in vivo and in vitro platform for evaluating drug efficacy against cancer cells in patient treatment, 2024 IEEE 9th International Conference on Engineering Technologies and Applied Sciences (ICETAS). November 20-22, 2024. Bahrain, Bahrain. IEEE, (2024): 1-6. DOI:10.1109/icetas62372.2024.11120135.
[31] V.M. Alves, E.N. Muratov, S.J. Capuzzi, et al., Alarms about structural alerts, Green Chem. 18 (2016) 4348–4360.
[32] K. Ilbeigi, C. Barata, J. Barbosa, et al., Assessing environmental risks during the drug development process for parasitic vector-borne diseases: A critical reflection, ACS Infect. Dis. 10 (2024) 1026–1033.
[33] T.T. Wager, X. Hou, P.R. Verhoest, et al., Moving beyond rules: The development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties, ACS Chem. Neurosci. 1 (2010) 435–449.
[34] J. Dulsat, B. López-Nieto, R. Estrada-Tejedor, et al., Evaluation of free online ADMET tools for academic or small biotech environments, Molecules 28 (2023), 776.
[35] M.G. Bertram, B.B.M. Wong, K. Kümmerer, et al., Development of environmentally biodegradable drugs: What are the key challenges? Expert Opin. Drug Discov. 20 (2025) 1–4.
[36] K. Kümmerer, Green and sustainable medicinal chemistry. Green and sustainable medicinal chemistry: methods, tools and strategies for the 21st century pharmaceutical industry. The Royal Society of Chemistry, (2016), ch. 7, pp 73–81.
[37] E. Wynendaele, C. Furman, B. Wielgomas, et al., Sustainability in drug discovery, Med. Drug Discov. 12 (2021), 100107.
[38] C.M. Lima, E. Uliassi, E.S.J. Thoré, et al., Environmental impacts of drugs against parasitic vector-borne diseases and the need to integrate sustainability into their development and use, Open Res. Eur. 4 (2024), 207.
[39] R. Vidaurre, I. Bramke, N. Puhlmann, et al., Design of greener drugs: Aligning parameters in pharmaceutical R&D and drivers for environmental impact, Drug Discov. Today 29 (2024), 104022.
[40] T. Tamilvizhi, R. Thirumalini, M. Takshinyaa, et al., An AI-based framework for real-time patient monitoring and intelligent treatment recommendation in critical care units, 2025 2nd International Conference on Computing and Data Science (ICCDS). July 25-26, 2025. Chennai, India. IEEE, (2025), 1-5. DOI:10.1109/iccds64403.2025.11209640.
[41] W.B. Sherine, T. Tamilvizhi, S. Lincy Jemina, Graph enhanced transformers for predicting drug using drug synergies and antagonistic interactions, 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC). September 10-12, 2025. Coimbatore, India. IEEE, (2025): 902-907. DOI:10.1109/icesc65114.2025.11212498.
[42] L. Guan, H. Yang, Y. Cai, et al., ADMET-score–a comprehensive scoring function for evaluation of

chemical drug-likeness, Med. Chem. Commun. 10 (2019) 148–157.
[43] RDKit: Open-source cheminformatics. https://www.rdkit.org.
[44] C.A. Lipinski, Drug-like properties and the causes of poor solubility and poor permeability, J. Pharmacol. Toxicol. Meth. 44 (2000) 235–249.
[45] W. McKinney, Data structures for statistical computing in Python, Proceedings of the Python in Science Conference. Austin, Texas. SciPy, (2010). DOI:10.25080/majora-92bf1922-00a.
[46] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585 (2020), 357–362.
[47] J.D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. 9 (2007) 90–95.
[48] M. Waskom, Seaborn: Statistical data visualization, J. Open Source Softw. 6 (2021), 3021.
[49] Virtanen, P., Gommers, R., Oliphant, T.E. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17 (2020), 261–272.
[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[51] I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees, Nucleic Acids Res. 44 (2016) W242–W245.
[52] A. Bouzas-Monroy, J.L. Wilkinson, M. Melling, et al., Assessment of the potential ecotoxicological effects of pharmaceuticals in the world's rivers, Environ. Toxicol. Chem. 41 (2022) 2008–2020.
[53] J.L. Wilkinson, A.B.A. Boxall, D.W. Kolpin, et al., Pharmaceutical pollution of the world's rivers, Proc. Natl. Acad. Sci. U. S. A. 119 (2022) e2113947119.
[54] Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006 (Text with EEA relevance), 2008. http://data.europa.eu/eli/reg/2008/1272/oj.
[55] Understanding the Globally Harmonized System of Classification and Labelling of Chemicals (GHS), (2024).
https://unitar.org/sites/default/files/media/file/GHS%20Companion%20Guide%20-%2012.7.2024%20FINAL.pdf.
[56] United Nations economic commission for Europe, globally harmonized system of classification and labelling of chemicals (GHS). United Nations, (2023).
[57] European Chemicals Agency., Guidance on the application of the CLP Criteria: guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Part 4, Environmental hazards. Part 5, Additional hazards., Publications Office, LU, 2024. https://data.europa.eu/doi/10.2823/4895485.
[58] D. Rogers, M. Hahn, Extended-connectivity fingerprints, J. Chem. Inf. Model. 50 (2010) 742–754.
[59] G. Landrum, Generating Pharmacophore Fingerprints in RDKit-RDKit Documentation, chapter Chemical Features and Pharmacophores, (2013). https://www.rdkit.org.
[60] A.J. Williams, C.M. Grulke, J. Edwards, et al., The CompTox Chemistry Dashboard: A community data resource for environmental chemistry, J. Cheminf. 9 (2017), 61.
[61] R. Karki, Y. Gadiya, A. Zaliani, et al., KGG: A fully automated workflow for creating disease-specific knowledge graphs, Bioinformatics 41 (2025), btaf383.

Journal Pre-proof

Highlights

- One Health–based workflow integrates ecotoxicology in antiparasitic discovery.
- Machine learning predicts molecular safety from structural information.
- GreenDrugScore changes hit prioritization via ADMET and ECO parameters integration.
- Nineteen prioritized scaffolds with favorable predicted ecotoxicological profiles proposed as antiparasitic candidates for follow-up.

Journal Pre-proof

Journal Pre-proof

Declaration of Interest Statement

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof