A combination of high-throughput *in vitro* and *in silico* new approach methods (NAMs) for ecotoxicology hazard assessment for fish

Abstract

Fish acute toxicity testing is used to inform environmental hazard assessment of chemicals. *In silico* and *in vitro* approaches have the potential to reduce the number of fish used in testing and increase the efficiency of generating data for assessing ecological hazards. Here, two *in vitro* bioactivity assays were adapted for use in high-throughput chemical screening. First, a miniaturized version of the OECD TG 249 plate reader-based acute toxicity assay in RTgill-W1 cells was developed. Second, the Cell Painting (CP) assay was adapted for use in RTgill-W1 cells along with an imaging-based cell viability assay. Then 225 chemicals were tested in each assay. Potencies and bioactivity calls from the plate-reader and imaging-based cell viability assays were comparable. The CP assay was more sensitive than either cell viability assay in that it detected a larger number of chemicals as bioactive, and phenotype altering concentrations (PACs) were lower than concentrations that decreased cell viability. An *in vitro* disposition (IVD) model that accounted for sorption of chemicals to plastic and cells over time, was applied to predict freely dissolved PACs and compare to *in vivo* fish toxicity data. Adjustment of PACs using IVD modeling improved concordance of *in vitro* bioactivity and *in vivo* toxicity data. For the 65 chemicals where comparison of *in vitro* and *in vivo* values was possible, 59 % of adjusted *in vitro* PACs were within one order of magnitude of *in vivo* toxicity LC50s. *In vitro* PACs were protective for 73 % of chemicals. This combination of *in vitro* and *in silico* approaches has the potential to reduce or replace the use of fish for *in vivo* toxicity testing.

Abbreviations

3,4-DCA	3,4-dichloroaniline
AGP	actin skeleton, Golgi, plasma membrane
AIC	Akaike information criterion
BMC	benchmark concentration
BMR	benchmark response
CASRN	Chemical Abstract Services Registry Number
CP	Cell Painting
CV	Cell viability
CV-IB	imaging-based cell viability assay
CV-PR	plate reader-based cell viability assay
CCTE	Center for Computational Toxicology and Exposure
DTXSID	Distributed Structure-Searchable Toxicity (DSSTox) Database Substance Identifier
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EICO	5,8,11-Eicosatrynoic acid
ER	endoplasmic reticulum
FHM	fathead minnow
GR	glucocorticoid receptor
HTPP	high-throughput phenotypic profiling
IRIS	Integrated Risk Information System
IARC	International Agency for Research on Cancer
IVD	*in vitro* disposition
LOEC	lowest observed effect concentration
LSER	linear solvation energy relationships
MAD	median absolute deviation
Mito	mitochondria

NAMs	New Approach Methodologies
NOEC	No observed effect concentration
nCC	normalized cell count
OECD	Organisation for Economic Cooperation and Development
OPERA	OPEn structure-activity/property Relationship Application
PAC	phenotype altering concentration
PAH	polycyclic aromatic hydrocarbon
PBS	phosphate buffered saline
QSAR	quantitative structure activity relationship
RAR	retinoic acid receptor
RMSE	root mean square error
RBT	rainbow trout
RNA	ribonucleic acid
SMILES	simplified molecular-input line-entry system

Introduction

Fish acute toxicity testing is typically conducted to inform environmental hazard assessment of chemicals. According to Organisation for Economic Cooperation and Development (OECD) test guideline 203, a minimum of 42 fish are required to test a single chemical (OECD 2019). Accordingly, the use of New Approach Methods (NAMs) could help to both reduce the number of fish used in testing as well as the time needed to generate relevant hazard data for chemicals. To date, only one cell-based *in vitro* alternative to *in vivo* toxicity testing in fish has been validated by the OECD. In test guideline 249 (OECD TG249), rainbow trout (RBT, *Oncorhynchus mykiss*) gill cells (RTgill-W1) are exposed to test chemicals for 24 hours. Cell viability is then assessed using a plate reader-based assay (OECD 2021). Tanneberger et al (2013) found the effective concentration for 50% cytotoxicity (EC50) of the *in vitro* assay was predictive of the lethal concentration for 50% of the test organisms (LC50) in fathead minnow (FHM, *Pimephales promelas*) for a diverse set of 35 chemicals (Tanneberger et al. 2013). As encouraging as this finding was, the assay is low throughput, requiring a substantial number of cells, labor and mass of test chemical when run in 24-well plates. Miniaturization of the assay to a smaller well format could expand testing of additional chemical libraries and require smaller quantities of test material.

Three cell viability endpoints that evalute cellular membrane integrity or metabolic activity are measured in OECD TG249. However, in recent years, profiling methods have emerged that can identify chemical bioactivity at sub-cytotoxic concentrations. One such method is imaging-based high-throughput phenotypic profiling (HTPP) using the Cell Painting assay (Gustafsdottir et al. 2013; Bray et al. 2016; Nyffeler et al. 2020). In Cell Painting, a set of fluorescent probes is used to visualize the effect of chemicals on various cellular organelles (nucleus, nucleoli, endoplasmic reticulum (ER), actin cytoskeleton, Golgi bodies, plasma membrane and mitochondria). The chemical-induced changes in cellular phenotypes are then quantified using automated microscopy and image analysis. We have shown previously that in human cells HTPP with Cell Painting is more sensitive than simple cell viability measurements for detecting bioactivity of environmental chemicals (Nyffeler et al. 2020; Nyffeler et al. 2023). Moreover, the bioactivity threshold values derived from HTPP using a human cell line were lower or comparable to *in vivo* effect values for 68% of the 462 chemicals tested (Nyffeler et al. 2020).

Although HTPP has mostly been used in human cells, there have been a few applications in other mammals (Nelson et al. 2023). The labeling reagents, which consist of small molecules and lectins, are not specific for mammalian biology (Manalo et al. 2016). Non-mammalian cells are also amenable to HTPP

(Hecker et al. 2024), which motivated the present study. We hypothesized that the HTPP assay could be conducted with RTgill-W1 cells and that HTPP would be a more sensitive measure of bioactivity than cell viability alone based on previous HTPP findings using human cell lines. In the present study, the OECD TG249 assay was miniaturized to a 384-well format. We refer to this plate-reader based cell viability assay as CV-PR. We then applied two additional assays to the RTgill-W1 cells: an imaging-based cell viability assay (CV-IB) as described (Nyffeler et al. 2020; Willis et al. 2020) and HTPP. We then tested 225 unique chemicals with the aim of comparing the results from this study to (a) existing *in vitro* cell viability assay results from RTgill-W1 and (b) existing *in vivo* fish toxicity data. For the latter comparison, the *in vitro* disposition (IVD) of the chemicals in 384-well format was taken into account by applying an *in silico* IVD model that accounts for time-dependent sorption of chemicals to the plastic of the culture vessel (Fischer, Cirpka, et al. 2018), in addition to cellular uptake kinetics. The model can be run via a spreadsheet for low-throughput (single chemical) applications or in batch mode for high-throughput processing involving many chemicals. Combining the high-throughput *in vitro* data with the *in silico* IVD model allowed for a prediction of a lower bound for fish lethality. In the future, this combination of NAMs could potentially replace or reduce *in vivo* toxicity testing in fish.

Materials and Methods

Materials & Chemical Selection

Materials were acquired from various suppliers and are listed in Supplementary Table S1.

Experiments were conducted in two parts: (1) a pilot experiment for proof-of-concept and to identify suitable phenotype altering reference chemicals for tracking assay performance, and (2) screening a chemical test set.

In the pilot experiment, a set of 20 candidate phenotype altering reference chemicals and two negative controls, referred to as the RefChem20 set (Culbreth et al. 2022) as well as 3,4-dichloroaniline (3,4-DCA) as a viability positive control, were tested (Suppl. Table S2).

For screening, a list of 238 chemicals was assembled taking into consideration availability of existing toxicity and exposure information. Chemicals were selected based on (1) presence of toxicity data in rainbow trout as captured in the US EPA Ecotoxicology (ECOTOX) knowledgebase (Olker et al. 2022), (2) published data in RTgill-W1 cells, (3) overlap with chemicals tested in a larval fathead minnow high throughput transcriptomics assay (Villeneuve et al. 2023; Flynn et al. 2024) and (4) availability in our chemical inventory. Additionally, 5 piscicides and 10 putative inert compounds (based on being tested in at least 500 ToxCast *in vitro* assays and being active in less than 2% of those) were selected. As an internal quality control, nine chemicals (1,2-dichlorobenzene, 2,4,6-trichlorophenol, aniline hydrochloride, benzo[a]pyrene, carbendazim, disulfoton, hexachlorophene, pentachlorophenol and permethrin) were screened in duplicate. These were manually selected to favor bioactive chemicals. Of 238 chemical samples tested (229 unique chemicals), one chemical sample could not be dispensed during the preparation of dose plates and was replaced with another replicate of 3,4-DCA (at the same test concentration as the reference chemical stock). Additionally, 3,4-DCA was also present as a blinded test chemical sourced from the chemical vendor, resulting in 238 chemical samples (228 unique chemicals). Three test chemicals were inorganic salts (copper (II) sulfate, nickel (II) sulfate anhydrous, zinc sulfate heptahydrate), therefore were dissolved in water. While conducting the experiments, it was noted that these water-soluble chemicals were not accurately dispensed by our laboratory automation, so they were subsequently excluded from further evaluation. Ultimately, 235 chemical samples (225 unique chemicals) were assessed for bioactivity in RTgill-W1 cells (Suppl. File S1, https://comptox.epa.gov/dashboard/chemical-lists/HTPP2024_RTGILLW1).

Cell Culture

One vial of RTgill-W1 cells (CRL-2523™, ATCC) was thawed and expanded to internal passage number 8 (P8) using RTgill-W1 Expansion Medium (L-15 with 2 mM L-glutamine and phenol red, 5% heat inactivated fetal bovine serum, 1% HyClone Penicillin at 10,000 units/mL and Streptomycin at 10 mg/mL) (Suppl. Table S3) in a 19 °C incubator with ambient CO_2 atmosphere in the dark. At 4-6 days in culture, the cells were inspected under a light microscope, washed with 1X phosphate buffered saline (PBS) and medium was refreshed. Cells reached 90% or higher confluency after 8-10 days, at which point they were passaged. All passages were conducted using the following procedure: Flasks were washed with 1X PBS before trypsinization with 1X LifeTech™ TrypLE™ Select . Cells were incubated for 2 to 3 minutes with mild shaking until detachment occurred as confirmed via microscopy. Expansion medium was then added to the flask to inactivate the TrypLE™, and the cell suspension was transferred to a conical tube and centrifuged for 10 minutes at 250 relative centrifugal force. The resulting supernatant was aspirated, and cells were resuspended in the RTgill-W1 Expansion Medium. Trypan blue exclusion was used to count cells and assess cell viability. Cells were seeded at no less than 55,000 cells/cm^2, with the average cell density being approximately 60,000 cells/cm^2. At passage 8 (P8), cells were cryopreserved using RTgill-W1 Cryopreservation Medium (Suppl. Table S3) at 5.0x10^6 cells/mL.

For all experiments, vials of P8 RTgill-W1 cells were thawed and passaged on a seven-day schedule (Fig 1A) in Expansion Medium. As a result, flasks were seeded at a slightly higher density than during the initial expansion (approximately 70,000 cells/cm^2) to achieve 90% confluency after seven days. The formulation for Experimental Medium was identical to Expansion Medium except L-15 without phenol red was used (see Suppl.Table S3). RTgill-W1 Experimental Medium was used from the thawing of the cells onwards. On the fourth day of each culture cycle, cells were washed with 1X PBS and medium was refreshed as described.

Cells were maintained in culture for two passages after thaw. Plating of P10 cells for chemical screening occurred on the second passage, two weeks after the initial thaw (Fig 1A). Cell plating followed the standard procedure for passaging and counting described earlier. Revvity PhenoPlate 384-well plates were prepared for plating by dispensing 20 µL of Experimental Medium into each well using a Certus Flex Liquid Dispenser equipped with a 0.45/0.15 Gyger microvalve. The cell suspension from passaging was diluted to achieve a concentration which would yield the desired number of cells per well in 20 µL, 17,500 cells per well for the pilot experiments (875,000 cells/mL) or 22,500 cells per well for screening experiments (1,125,000 cells/mL). Twenty microliters of cell suspension were then added to each well

designated for cells using the Certus Flex Liquid Dispenser, for a total volume of 40 µL for each well. For the screening experiments, four wells received media only to serve as blank controls. Plates were then returned to the incubator until the next day.

A minimal medium, L-15/ex, was prepared for exposure, as described previously (see Suppl. Table S3) (Schirmer et al. 1997; OECD 2021). Approximately 1 to 2 hours before chemical exposure, the Experimental Medium was removed and replaced with 40 µL of L-15/ex minimal medium with the CyBio FeliX Automated Liquid Handler.

Chemical Treatment & Experimental Design

Chemicals were either purchased as neat chemicals and solubilized in dimethyl sulfoxide (DMSO) or received as solutions in DMSO from the US EPA ToxCast chemical library management contractor (Evotec, Princeton, NJ), typically as 20 mM stocks. Dilution series were prepared by dispensing with a LabCyte Echo 555 acoustic dispenser and backfilling with DMSO using a Certus Flex Liquid Dispenser and 0.10/0.03 Gyger microvalve as described previously (Nyffeler et al. 2020). Dilution series consisted of half-$\log_{10}$ spacing at 200x the desired well concentration in DMSO and typically consisted of eight concentrations. Dose plates were stored at -80°C and thawed on the day of dosing for each experiment. Cells were dosed 24 hours after plating with a LabCyte Echo 550 acoustic dispenser. Treatment position on each cell culture plate was randomized using a custom R script. Each cell culture well received 200 nL of chemical, and plates were transferred back to the 19°C incubator after dosing.

The dose plate for the pilot experiment contained eight concentrations for the 20 candidate phenotype altering reference chemicals and two HTPP inactive controls, saccharin and sorbitol, as well as six concentrations of 3,4-DCA as a cell viability positive control and pure DMSO as a vehicle control (Suppl. Fig S1A). Each concentration of candidate phenotype altering reference chemicals and inactive controls were tested using two technical replicate wells per plate, 3,4-DCA with one technical replicate well for all but the highest dose level which had 3 technical replicate wells, and 24 technical replicate wells of DMSO vehicle control. The pilot experiment was repeated in 4 independent cultures.

For the chemical screening, a total of six dose plates were prepared. Each dose plate contained 40 test chemicals at eight concentrations, three reference chemicals (cucurbitacin I, docetaxel, and 5,8,11-eicosatrynoic acid (EICO)) at eight concentrations and the viability positive control 3,4-DCA at 12 concentrations (Suppl. Fig S1B). The dilution series for 3,4-DCA was comprised of six concentrations at two-fold dilution (120 – 3.75 mM) covering the range proposed in the OECD TG249 followed by six

concentrations at approximately three-fold dilution (1.2 – 0.0036 mM) to cover the onset of phenotypic changes. Different design layouts were used for Cell Painting (CP) and cell viability (CV) plates to accommodate the appropriate control treatments for each assay. For CV assay plates, 16 wells were dosed with solvent control, triplicate wells for each of the six highest 3,4-DCA concentrations, and triplicate wells of the highest EICO concentration, with the remainder of reference chemical and test chemicals tested in one replicate well. For CP assay plates, 22 wells were dosed with solvent control, the highest concentration of EICO in triplicate wells, with all twelve 3,4-DCA concentrations, all test chemicals and the remainder of reference chemicals in one replicate well. The screening experiment was repeated in four independent cultures.

Imaging-based Cell Viability Assay (CV-IB)

Labeling and Fixation

To assess cytotoxicity, an imaging-based cell viability assay was used as described (Nyffeler et al. 2020; Willis et al. 2020) (Fig 1B). This assay had previously been applied to human-derived cell lines, but not RTgill-W1 cells. To each well, 2 µL of cell viability stain (202.5 µM Hoechst 33342 and 75.6 µM propidium iodide in L-15/ex) (Suppl. Table 4A) were added to the 40 µL of media in each assay well using a Certus Flex Liquid Dispenser and a 0.10/0.03 Gyger microvalve. After a 30-minute incubation in the 19°C incubator, the cells were fixed with direct addition of 12 µL of 16% paraformaldehyde using a MultiFlo™ FX liquid dispenser and incubated for 10 minutes at room temperature in the dark. Wells were finally washed 2 times with 1X PBS, the final volume of which was kept as a storage buffer. Plates were sealed with optical film and stored at 4°C until the day of imaging.

Image Acquisition

Plates were imaged using a PerkinElmer Opera Phenix Plus High Content Screening System and Harmony software (v. 5.0.2157.133). Images were obtained using a 20x water immersion objective in non-confocal mode with five 650 µm x 650 µm fields per well, using 2 x 2 pixel binning. The Hoechst 33342 channel was imaged with an excitation wavelength of 405 nm and emission of 435-480 nm. The propidium iodide channel was imaged with an excitation wavelength of 561 nm and emission 570-630 nm. Z-offsets were adjusted for each channel to focus on the nuclei. Acquired images were analyzed by segmenting the nuclei using the Hoechst 33342 channel, followed by quantification of the intranuclear propidium iodide fluorescence. Analysis outputs included Hoechst 33342 and propidium iodide intensities as well as morphological nucleus properties for each individual nucleus.

Data Aggregation

Cell-level data were analyzed in R (v. 4.1.2) *base* package and *tidyverse* for data wrangling and visualization (Wickham et al. 2019). Data were first filtered to include nuclei with an area within 30-1,000 μm^2 and roundness of at least 0.3 to exclude imaging artifacts. Cytotoxicity was measured by determining the 95^{th} percentile of propidium iodide intensity in solvent control cells and categorizing cells above this threshold as propidium iodide positive, indicating cell death. Cell-level data was then aggregated into well-level data and two endpoints were calculated for analysis: 1) the percentage of cells in each well that were propidium iodide positive and 2) normalized cell count (nCC), which is the number of nuclei counted in a treatment well divided by the average number of nuclei counted in plate-matched vehicle control wells.

Cell Painting Assay (CP)

Fixation and Labeling

The CP assay was conducted exactly as described previously (Nyffeler et al. 2020; Willis et al. 2020) (Fig 1C). This assay had previously been applied to human-derived cell lines, but not RTgill-W1 cells. The CP sample preparation consists of two individual labeling portions: live-labeling at 24 hours after dosing, followed by fixation, after which point the cells were permeabilized and stained with a second fluoroprobe mixture. Twenty-four hours after dosing, each well received 2 µL of 10 µM MitoTracker DeepRed in L-15/ex).). After 30 minutes of incubation in the 19°C incubator, plates received 12 µL/well of 16% paraformaldehyde to fix the cells, dispensed with a MultiFlo™ FX liquid dispenser. Plates were fixed for 10 minutes at room temperature in the dark before being washed with 1X PBS, after which they were stored in a 4 °C refrigerator until further processing.

To finish staining the plates, they were removed from the refrigerator and allowed to come to room temperature for approximately one hour. First, the cells in each plate were permeabilized. Permeabilization solution consisted of freshly filtered 0.5% Triton X-100 in PBS. The plates first underwent a wash step on the MultiFlo™ FX to ensure the amount of 1X PBS in each well was 40 µL before adding 10 µL of permeabilization solution to each well using the Certus Flex Liquid Dispenser. Plates were then left to incubate for 30 minutes at room temperature before being washed once with 1X PBS, leaving 40 µL in each well.

Plates received 2 µL of the full stain labeling solution (Table S4B) dispensed with the Certus Flex, followed by incubation in the dark for 30 minutes at room temperature. Then plates were washed three

times with 1X PBS, brought up to a final storage volume of approximately 70 µL per well and sealed with an optical plate seal. Plates were stored at 4 °C until the day of imaging.

Image Acquisition

Image acquisition followed a modified procedure from Nyffeler et al. (2020) on the Revvity Opera Phenix Plus with Harmony software (v 5.0). Four channels with different excitation wavelengths were used: 405 nm (DNA), 488 nm (RNA and endoplasmic reticulum [ER]), 561 nm (actin cytoskeleton, Golgi bodies, plasma membrane [AGP]), and 640 nm (mitochondria [Mito]). Emission wavelengths were 435-480 nm (DNA), 500-550 nm (RNA and ER), 570-630 nm (AGP), and 650-760 nm (Mito). Images were acquired using a 20X water immersion objective in confocal mode, with three 650 µm x 650 µm fields imaged per well and 2 x 2 pixel binning. For optimal focus, seven Z-planes were imaged spanning 15 µm.

Feature Extraction

In the first step of feature extraction, Z-planes were combined into one maximum projection image per field. Then, nuclei were segmented by using the DNA channel to determine the outer border of each nucleus, followed by segmentation using the ER channel to define the outer border of the cell. Nuclei that did not meet a minimum threshold for roundness (> 0.7) and median intensity (> 800) were considered artifacts and were removed, as were any cells touching the border of the image. For valid objects, subcellular compartments were defined: membrane, cytoplasm, nuclei and perinuclear space (i.e. "ring"), as described (Nyffeler et al. 2020). For each channel in each compartment, as well as for the whole cell, phenotypic features were measured based on patterns of fluorescence signal. Morphological properties such as area and roundness were calculated, and multiple numeric representations of texture were computed. The result was 1,294 phenotypic features per cell, which were then exported to a tab-separated value file for further analysis in R. The feature extraction protocol is provided in the supporting information (https://doi.org/10.23645/epacomptox.26799382).

CP images were inspected manually to identify chemical fluorescing in one or more of the channels used for CP. These treatments were excluded (Table S5) from later concentration-response modeling steps in the same manner as cytotoxic conditions.

Data Aggregation

Data aggregation was performed as described previously (Nyffeler et al. 2023). Briefly, well-level aggregation was performed for each plate by calculating the median of each feature for every well, which was then normalized to the median and normalized median absolute deviation (nMAD) of the vehicle

control wells. Next, well-level data from all four plates belonging to the same plate group (i.e., plates that received the same set of test chemicals) was further normalized to the mean and standard deviation (SD) of all solvent control wells to obtain scaled well-level data, where +1 means that the value of the feature is +1 SD above the mean of solvent controls. Scaled well-level data were used to generate the heatmaps and for concentration-response modeling (described below).

Plate reader-based Cell Viability Assay (CV-PR)

Labeling and Fluorescence Measurement

For the plate reader-based cell viability assay (CV-PR), OECD TG 249 was followed with modifications necessary for adaptation to the smaller 384-well format and available robotic liquid-handling equipment (Fig 1D). To begin, the L15/ex medium was removed from the plates and replaced with 36 µL 1X PBS using a CyBio FeliX Automated Liquid Handler. This occurred approximately 10-15 minutes prior to the 24-hour exposure timepoint, at which point 4 µL of a 10X alamarBlue/CFDA-AM working solution (Table S4C) was applied to each well to obtain a final concentration of 5% V/V alamarBlue and 4 µM CFDA-AM. After 30 minutes of incubation at 19°C, fluorescence was measured using the Clariostar Plate Reader with 530 nm excitation / 595 nm emission for alamarBlue and 493 nm excitation / 541 nm emission for CFDA-AM, using the bottom optic and no aperture spoon. Focal height was adjusted based on well A1 of the first plate of each replicate, and gain values were kept consistent for all plates in each replicate. Gain was adjusted only if needed to ensure there were few to no fluorescence measurements above the maximum detection threshold of the plate reader.

The alamarBlue/CFDA-AM solution was removed and replaced with 36 µL of 1X PBS with a CyBio FeliX Automated Liquid Handler as described above. Each well then received 4 µL of 10X Neutral Red working solution (Table S4D) using a Certus Flex Liquid Dispenser with a 0.20/0.10 Gyger microvalve to obtain a final concentration of 0.005%. After 60 minutes of incubation at 19°C, the Neutral Red working solution was removed from each well with a CyBio FeliX Automated Liquid Handler. Then pipette tips were washed first with extraction solution and second with 1X PBS before 50 µL of fixation solution (0.25% V/V paraformaldehyde and 5 g/L calcium chloride in deionized water) was applied to each well and incubated for two minutes. After aspirating the fixation solution, the tips were washed as described above; then 20 µL of extraction solution (50% V/V ethanol, 1% V/V glacial acetic acid, 49% V/V deionized water) was applied to each well. The plates were wrapped in aluminum foil and placed on a plate shaker on its lowest

setting for 10 minutes, after which the fluorescence of the Neutral Red was measured using a Clariostar plate reader with excitation at 493 nm and emission at 541 nm.

Data Aggregation

Data were exported as comma separated value files and analyzed using R (v. 4.1.2). For each plate and assay, the median fluorescence of the background wells (wells without cells in the screening experiments, otherwise wells containing the lowest four fluorescence values from each plate in the pilot experiments) was subtracted from each well. Data were then normalized to the median fluorescence of vehicle control wells. Normalization was done per plate to account for plate-to-plate variations across the span of the assay. This normalized fluorescence value was then used for concentration response modeling to determine cytotoxic concentrations.

Concentration-response Modeling

All concentration-response modeling was performed using the R package *tcplfit2* 0.1.3 (Sheffield et al. 2022).

Cell Viability Data

Concentration-response analysis was performed on three endpoints of the CV-PR assay (alamarBlue, CFDA-AM and neutral red) and two endpoints of the CV-IB assay (nCC, % propidium iodide (PI) positive cells). Data were processed by plate group, i.e., for each endpoint, the noise level was estimated by calculating the median and nMAD of all solvent control wells of the four corresponding plates (typically 96 wells in total). Solvent control wells with nCC < 50% were removed as outliers. In addition, solvent control wells that contained less than 100 individual cells were removed as outliers for the PI endpoint. All solvent control wells passed these criteria and therefore were included in analysis.

Benchmark responses (BMRs) and *cutoff* values (i.e., the value a response must exceed for a response to be classified as active) were determined independently for each assay. For the CV-PR and the nCC endpoints, constant and hill models were fit to the data only if data exceeded the *cutoff* that was set at 2*nMAD of solvent controls. If the Hill model had the lowest Akaike information criterion (AIC), the benchmark concentration (BMC) at a BMR of 50% determined. For the PI endpoint, the *cutoff* was set at 5*nMAD of solvent controls and the constant, Hill and gain-loss model were fit to the data. If the Hill or the gain-loss model had the lowest AIC, the BMC was calculated using a BMR of 3*nMAD.

A chemical was considered active in a particular CV endpoint if the curve fit resulted in a BMC below the highest tested concentration and the curve was fit in a downward direction for CV-PR and nCC or upward direction for the PI endpoint. The no observed effect concentration (NOEC) was defined as the highest tested concentration below the BMC. The lowest observed effect concentration (LOEC) was defined as the lowest test concentration above the BMC.

For some analyses, all endpoints from a CV assay were combined and represented as one value. In this case, the chemical was considered active if it was active in any of the individual endpoints and the lowest BMC was selected as the representative potency for the corresponding assay. In some instances, the results from both CV assays were combined. In this case, the chemical was considered active if it was active in either assay and the lower BMC was selected as the CV potency.

Cell Painting Data

For concentration-response modeling of CP data, the 1,294 features were condensed to either one value or 49 values, for the 'Global Mahalanobis' and 'Category-level Mahalanobis' approaches, respectively, as described in (Nyffeler et al. 2021). All wells with at least 500 cells were used in the principal component analysis to calculate a rotation matrix. For the 'Global Mahalanobis' approach, 5 eigenfeatures covering >95% of the variance were retained in the analysis. For the 'Category-level Mahalanobis', eigenfeatures were retained that covered >95% of the variance in each feature category.

To estimate the noise level for the screen, synthetic 'null chemicals' were generated as described previously (Nyffeler et al. 2021; Nyffeler et al. 2023). For this, wells with assumed no (or low) bioactivity were selected and randomly assigned to individual 'null chemicals'. Wells of the lowest two test concentrations of test chemicals were eligible, if at least six test concentrations were below the NOEC for cell viability (determined above). Each cell culture plate had at least 64 eligible wells, thus 8 null chemicals with 8 concentrations each were generated per plate. The procedure was repeated once for all six plate groups, thus resulting in 96 'null chemicals'. Subsequently, those 'null chemicals' were analyzed analogously to the test chemicals and reference chemicals.

Curve-fitting was conducted using the Mahalanobis distances. First, data from highly cytotoxic conditions were removed by discarding data from concentrations above the LOEC for CV (determined above). Distance values from the lowest two concentrations of test chemicals were used to determine the noise level. To ensure that those values were indicative of no bioactivity, a Tukey's outer fence test was conducted and values that exceeded the 75^{th} percentile + 3 interquartile range of the distance distribution

were removed. The remaining values were used to calculate the mean and standard deviation of wells showing negligible bioactivity. A concentration-response curve was only fit to the data if values exceeded a cutoff of 1 SD. Nine models were fit to the data (constant, Hill, gain-loss, first and second order polynomial, power, four exponential models). The BMR was set at 1.349*SD.

A BMC was considered valid in the 'Global Mahalanobis' approach if the curve fit resulted in a BMC below the highest tested concentration. For the 'Category-level Mahalanobis' approach, BMCs were considered valid if the fit resulted in a BMC below the highest tested concentration with a continuous hitcall >= 0.9. If the global or at least one category endpoint resulted in a valid BMC, the chemical was considered active, and the lowest value of all valid BMCs was considered the phenotype altering concentration (PAC).

For some comparisons, the results of chemicals tested as multiple samples (i.e., reference chemicals and duplicated test chemicals) were combined. In this case, a chemical was considered active if it was active in any sample, and the lowest PAC from either sample was retained.

In vitro Disposition: Modeling

Collection of Physicochemical Properties

Experimental and predicted physicochemical properties, namely the water-octanol partition coefficient (K_{ow}) and air-water partition coefficient (K_{aw}) as well as the acid dissociation constants (pKa) were collected, and partitioning to proteins and lipids predicted as described in Supplementary Method 1.

Kinetic Mass Balance Model

To describe *in vitro* disposition, a kinetic mass balance model was implemented as a nested model consisting of three compartments: medium, cells and plastic (polystyrene). L15/ex medium did not contain proteins or lipids and was treated as a pure aqueous phase. Cells consist of proteins, lipids and water with associated volume fractions $Vf_{protein,cell}$, $Vf_{lip,cell}$ and $Vf_{water,cell}$. The partition of a chemical to the cells is described with (eq. 1).

$$D_{cell/w} = Vf_{protein,cell} \cdot D_{protein/w} + Vf_{lip,cell} \cdot D_{lip/w} + Vf_{w,cell} \quad (1)$$

Where D is the distribution coefficient, obtained from the weighted average of the partition coefficients of all species:

$$D = \sum^i f^i K^i \quad (2)$$

The uptake and diffusion of chemicals into polystyrene is slow and hence highly time-dependent. Therefore, uptake and diffusion were implemented as a kinetic process in the model as described in Fischer et al. (Fischer, Cirpka, et al. 2018). Only the neutral fraction was considered to be available for uptake into polystyrene (Seidensticker et al. 2018).

If medium is supplemented with FBS, we assume the freely dissolved concentration is hardly depleted over time, and kinetics of cellular uptake will not substantially influence the uptake kinetics into polystyrene. This is not the case if the medium is serum free as is the case for L15/ex medium. In this case, it is assumed that the uptake kinetics in the cells cannot be neglected as they deplete chemical in the medium over time, so that the initial uptake into polystyrene is higher initially and slows as the chemical concentration decreases in medium and increases in cells. In bioassays with human cell lines, the cellular uptake was determined to reach a steady state ($t_{95\%}$) in three to four hours, depending on the chemical, its charge, and the amount of serum in the medium (Fischer, Abele, et al. 2018). Uptake kinetics into fish cell lines cultured without serum supplements have not yet been measured, therefore a $t_{95\%}$ of six hours was chosen as a conservative assumption. Cellular uptake kinetics were described by a first-order one-compartment uptake kinetics model (eq. 3):

$$\frac{dc_{cell}}{dt} = k_{w\rightarrow cell}\, c_w - k_{cell\rightarrow w}\, c_{cell} \quad (3)$$

With $k_{w\rightarrow cell}$ describing the rate constant for uptake into cells and $k_{cell\rightarrow w}$, describing the rate constant for release of chemical from cells. We used the numerical solution of eq. 3 provided by Gobas et al. (1992) (eq. 4).

$$C_{cell}(t) = C_{cell}(t-\Delta t) + k_{w\rightarrow cell} C_{w,t-\Delta t} \Delta t - k_{cell\rightarrow w}\, C_{cell,t-\Delta t} \Delta t \quad (4)$$

As no rate constants $\log k_{cell\rightarrow w}$ were available we estimated them from $D_{cell/w}$ and the presumed $t_{95\%}$ of six hours.

$$D_{cell/w} = \frac{k_{w\rightarrow cell}}{k_{cell\rightarrow w}} \quad (5)$$

For $t_{95\%}$ = 6h, the $k_{cell\rightarrow w}$ was non-linearly correlated with log $D_{cell/w}$ but the $k_{wl\rightarrow cell}$ was pseudo-linearly correlated with log $D_{cell/w}$ (Fig S2, eq. 6) and $k_{cell\rightarrow w}$ was obtained by eq. 4.

$$\log k_{w\rightarrow cell} = 0.677 \times \log D_{cell/w} \quad (6)$$

Model Validation

In order to validate the *in vitro* disposition modeling, *in vitro* disposition in 96-well format was analytically measured for five test chemicals. Detailed procedures are described in Supplementary Method 2.

Model Implementation and Application

The model can be implemented in two formats: (1) as an Excel spreadsheet to run a single chemical, requiring no programming skills (Supplementary File S2), and (2) as an R script to facilitate rapid analysis of hundreds of chemicals (Supplementary File S3). The latter was used in this study to estimate the *in vitro* disposition of the 205 chemicals. It requires an input file with the physicochemical and biomaterial partition parameters for all chemicals, as well as a file describing the well dimensions, the biological model and modeling settings (i.e., time steps). Input files as well as the result file are provided as Supplementary Files S4, S5 and S6). One input parameter is the volume per cell, which we estimated to be 1.5 pL (see Supplementary Method 3). The main output of the model is the mass fraction of chemical in the water (MF_{water}). In this work, the timed average of the mass fraction ($MF_{water,avg}$) was used, as it was assumed that this is a better representative of the true exposure concentration, compared to using the mass fraction at the end of the 24 h exposure.

Comparison to Literature Data

Comparison to Existing RTgill-W1 *in vitro* Data

Nominal effect concentrations were manually collected from the literature resulting in 57 values for 53 unique chemicals from 10 studies. If more than one effect concentration was available, the minimal value was used. A list with all values is assembled in Supplementary File S7.

Comparison to ToxCast *in vitro* Data

EPA's Toxicity ForeCaster (ToxCast) program has generated bioactivity data for hundreds of chemicals using a variety of high-throughput screening assays (Richard et al. 2016; Feshuk et al. 2023). Most ToxCast assays are targeted and measure chemical bioactivity at a human- or rodent-derived protein target or biological process. In the present study, an overall *in vitro* point of departure (POD) was calculated for chemicals in the RTgill-W1 test set, as described previously (Nyffeler et al. 2023). A total of 207 test chemicals were tested in at least 200 ToxCast assays. Of those 207 chemicals, 169 were active in at least three assays, and the remainder were defined as inactive chemicals.

Mining for Existing *in vivo* Data

The ECOTOX Knowledgebase was mined for mortality data for RBT and fathead minnow (FHM, *Pimephales promelas*), resulting in ~ 67,000 records corresponding to 4284 unique DTXSID. The chemicals of these records were matched to the list of chemicals tested in this study. A total of 174 chemicals were matched by DTXSID. Additionally, the chemical structures (QSAR-ready SMILES representation) were matched to manually identify further records where the same chemical was tested, but the formula had a different counter ion, hydration form or was a different stereoisomer. Overall, 177 chemicals (i.e., unique DTXSIDs) from the screen had records in ECOTOX, which were represented by 227 unique DTXSID in the ECOTOX data set.

The records of matching chemicals were further filtered as described in Fig S13. Records were filtered for (1) the organism life stage, to not include very young fish; (2) length, weight and a minimum age, also aimed at excluding young fish; (3) records where the concentration type was “active ingredient” or “dissolved” (and not “formulation”); (4) records where the concentration unit contained mg/L; (5) records where the effect was mortality; and (6) records where the endpoint was a lethal concentration (no NOEL, BMC, BCF, etc.). This resulted in 2144 records corresponding to 130 tested chemicals. Of those 130 chemicals, 97 chemicals had at least 3 records, and 68 and 69 chemicals had at least 3 records for FHM and RBT, respectively.

Comparison to Existing *in vivo* Data

To compare the *in vivo* literature data with the *in vitro* potencies from this study, all *in vivo* effect values (in mg/L) corresponding to water concentrations were converted to the molecular formulation of the chemical tested in this study: (conc in mg/L matching this study) = ((conc in mg/L in literature) / (molecular weight of chemical in literature)) * (molecular weight of chemical in this study). The *in vivo* lethal concentration (LC) for each chemical was summarized from the individual records by calculating percentiles of the $\log_{10}$-transformed concentration values using the R function *quantile* with the default setting (*type* = 7). For chemicals with at least 20 records, the 5th percentile was used, for chemicals with 3 – 19 records the 25%-percentile (i.e., the first quartile) was used. Individual records and aggregated values are listed in Supplementary File S8.

The summarized *in vivo* LC was then directly compared to the (predicted) water concentration in the *in vitro* model. Assumptions underlying this comparison approach are that the concentration of the *in vivo* exposure water in fish toxicity tests is approximately equal to the unbound concentration in the

organism at the site of toxicity and that *in vivo* toxicity is not necessarily mediated by effects on the gills even though *in vitro* bioactivity estimates are based on effects observed in a gill cell line.

To explore whether certain groups of chemicals were less well predicted with the RTgill-W1 assay, chemicals were assigned to chemical groups (see Supplementary File S1). A Chi2-test was then used to test whether membership in a given group (yes/no) had a statistical association with the predictivity relative to *in vivo* (within 1 order of magnitude / larger than 1 order of magnitude). The R function *chisq.test* was used with options *simulate.p.value*=T and *B*=2000. Moreover, a two-sided Wilcoxon rank sum test was conducted by comparing the difference of the *in vitro* to *in vivo* value (in log-units) for chemicals in each group vs chemicals not in that group. For this the R function *wilcox.test* was used.

Results

Background Subtraction for CV-PR Endpoints

The OECD TG249 assay was miniaturized from 24-well format to 384-well format (CV-PR) and applied to the 23 chemicals in a pilot experiment. OECD TG249 involves a well filled with medium without cells for subtracting background fluorescence, which was not followed in our pilot experiment (Fig S1A) but was followed in subsequent screening experiments (Fig S1B). We intended to use the positive control 3,4-DCA as a cell viability control; however, inspection of the images indicated that 3,4-DCA reduced the cell number to about 60% of vehicle controls (Fig S3A) but did not reduce fluorescence in the CV-PR endpoints as expected for loss of cell viability (Fig S3B). We then investigated two other options for background subtraction in the pilot experiment: use of (1) the fluorescence values of the highest concentration of a chemical with marked cytotoxicity (250 µM EICO, > 75% reduction in cell count) or (2) the four lowest fluorescence values (per endpoint) of each plate. Comparing all three options indicated that the last option gave the least variable results (Fig S3B, 3C), and hence the CV-PR data of the pilot experiment were background subtracted by using the four lowest fluorescence values per plate. Subsequently, for the screening experiments, wells with no cells were included for background subtraction (Fig S1B).

Pilot Experiment with 23 Candidate Phenotype Altering Reference Chemicals

The pilot experiment was aimed at identifying phenotype altering reference chemicals for tracking assay performance in subsequent chemical screening studies. Of the 23 tested chemicals, 12 had no observable cytotoxic effects in the CV-IB assay (Fig S4). Effects were observable and quantifiable via CP for 13 of the tested chemicals at a concentration below the lowest cytotoxic concentration or in the complete absence of cytotoxicity within the tested range. No effect was observed for the putative inactive controls, saccharin and sorbitol, as well as six test chemicals (aphidicolin, berberine chloride, Ca-074-Me, cladribine, cytarabine, etoposide). The viability positive control 3,4-DCA had observable effects in the CP assay and was cytotoxic in the CV-IB assay near the lowest tested concentration. Thus, a PAC could not be determined for 3,4-DCA due to a lack of a sufficient number of non-cytotoxic dose levels for concentration-response modeling in CP.

Choice of Positive Control and Phenotype Altering Reference Chemicals for Screening

As the OECD TG 249 positive control 3,4-DCA was not found to be active in the CV-PR assay in the current study, it would not be a suitable positive control for screening purposes. Instead, we selected EICO as a positive control for the CV assays. EICO caused a concentration-dependent decrease in all CV-PR endpoints and in both CV-IB endpoints with BMCs between 50-65 µM (Fig 2A). Of note, the normalized median absolute deviation (used to measure data spread and calculate noise bands) was higher for CV-PR endpoints (ranging from 17-25%) than for CV-IB: 8.7% for nCC and 1.4% for the propidium iodide endpoint, respectively.

We selected EICO (DTXSID10159018), cucurbitacin I (DTXSID501015546) and docetaxel (DTXSID0040464) as phenotype altering reference chemicals for RTgill-W1 cells (Fig 2B-D). EICO affected cellular phenotypes at approximately one order of magnitude below the threshold for cytotoxicity, at about 4 µM, with changes in the mitochondrial channel being the most potent and most pronounced. In solvent control wells, MitoTracker was distributed throughout the cytoplasm. In EICO treated cells, bright, granular objects were frequently observed in proximity to the nucleus. Moreover, the cells tended to group into clusters, whereas in solvent control wells the cells formed an even monolayer. Cucurbitacin I induced strong changes in the AGP channel beginning at ~0.003 µM, with a striking increase in intensity in the perinuclear region (indicating possible effects on the Golgi), while also decreasing the staining of the actin cytoskeleton. Docetaxel produced a wide variety of subtle phenotypic effects beginning at a relatively low concentration (>4.0×10^{-5} µM). Only the lowest four test concentrations of docetaxel were retained for analysis of CP data due to cytotoxicity. The most potent phenotypic effects within the non-cytotoxic dose range were often cell shape and structure related, such as symmetry, compactness, and overall shape. While docetaxel affected a wide variety of channels across its concentration range, the most potent effects were observed in the DNA and endoplasmic reticulum channels.

The three phenotype altering reference chemicals thus affect different cell organelles and have different effect sizes, with EICO having the largest and docetaxel the smallest effect sizes (Fig 2D).

Adapting Cell Seeding Density

We noted that cells were not forming a confluent monolayer in the pilot experiment. As confluency affects the toxicological response of RTgill-W1 cells and a confluent cell layer is preferred

(personal communication, Kristin Schirmer, Eawag), we explored various seeding densities and monitored the confluency two days after seeding (Fig S5). Cells were not 100% confluent at 20,000 cells/well but were very dense at 25,000 cells/well. Therefore, we decided to increase the cell seeding density from 17,500 to 22,500 cells/well for subsequent chemical screening.

Quality Control Measures for Chemical Screening

The 235 chemical samples were tested in concentration-response format in four independent cultures, with six plates per culture and assay. On each plate, the three reference chemicals and 3,4-DCA were tested to monitor reproducibility of the assays. The five CV endpoints gave reproducible results, with standard deviations in potency values of 0.1 – 0.2 orders of magnitude (Fig S6). PACs were calculated with two approaches (i.e. global and category-level) and also had standard deviations of 0.1 – 0.2 orders of magnitude for each reference chemical and approach combination (Fig S7). The PAC calculated with the category-level Mahalanobis approach tended to be 0.3 to 1 order of magnitude lower than PACs calculated with the global Mahalanobis approach. This was consistent with previous observations in human cell lines (Nyffeler et al. 2021; Nyffeler et al. 2023). For all four reference chemicals, a reproducible phenotypic profile was observed across cell culture plates (Fig S8). The greatest variation was observed for EICO. The displayed profiles are at 75 µM, which is the first cytotoxic concentration. On some plates, this concentration caused small changes in cytotoxicity endpoints while on others those endpoints were moderately affected. Lastly, ten chemicals were tested in duplicate. There was near perfect agreement between the duplicates for CV endpoints, with only one bioactivity mismatch (active vs inactive call) for one chemical in one CV endpoint (Fig S9A). CP hit calls for eight out of ten chemicals were concordant in both the global and category-level approaches. Aniline hydrochloride was active near the highest tested concentration in one duplicate and inactive in the other duplicate; 3,4-DCA replicates were also discordant and had a potency difference of more than one order of magnitude (Fig S9B). For the seven chemicals active in both duplicates, the difference was less than one order of magnitude, and in most cases, less than half an order of magnitude.

Screening Results

The screening comprised 225 unique chemicals, of which 151 (67%) were active in the CP assay (Fig 3A). A subset of chemicals was also active in the CV-IB and/or CV-PR assays, but there were no chemicals active in the CV-IB or CV-PR assays that were inactive in CP. A total of 48 (21%) chemicals were active in all three assays. For those chemicals, the BMCs derived with the CV-IB and CV-PR assays were

highly correlated and were typically within half an order of magnitude of each other (Fig 3B). However, each assay also identified some chemicals as active that were deemed inactive with the other CV assay (Fig 3A). The most potent chemicals were detected with both assays. Of note, PACs were lower than BMCs derived from the CV assays, in some cases by several orders of magnitude (Fig 3C). Consistent with previous studies, the most sensitive BMC derived from the category-level Mahalanobis approach tended to be lower than the BMC derived from the global Mahalanobis approach (Fig 3D). The category-level Mahalanobis approach also identified more chemicals as actives, but also identified 1/96 "null chemicals" as active compared to 0/96 for the global Mahalanobis approach. Overall, CP identified more chemicals than either of the CV assays and the potency estimates were lower (i.e., more sensitive).

Results in Comparison to RTgill-W1 *in vitro* Data from the Literature

RTgill-W1 cells have been used for chemical screening in previous studies by other groups. The chemical set in the present study comprised 46 chemicals previously tested in RTgill-W1 cells. In most cases, toxicity was assessed with the OECD TG 249 assay in 24-well format and the EC50 was reported. As the CV-PR assay was not very sensitive in our hands in 384-well format, we compared the literature reported EC50 values with the PACs. The hypothesis was that the PAC should be equal or lower than the EC50 value. This was indeed the case for all but two chemicals (Fig 4 and Fig S10). Pentachlorophenol was more than one order of magnitude more potent in previous studies than in our study, and for rotenone the PAC could not be determined because this chemical was active at the lowest tested concentration. Of note, all chemicals inactive in the present study had literature EC50 of >50 µM up to over 10 mM – concentrations much higher than the upper range of concentrations tested in the present study (100 µM). Overall, these results indicate that the PAC can serve as a lower bound to the EC50 for cell viability.

Results in Comparison to ToxCast

RTgill-W1 results from 207 chemicals could be compared to results from ToxCast. There were 32 chemicals that were inactive in both ToxCast and the RTgill-W1 assays. There were 36 chemicals that were inactive in RTgill-W1 CP but were active in ToxCast (i.e., a ToxCast *in vitro* POD could be calculated; Fig. 5A). Conversely, there were six chemicals that were active in RTgill-W1 CP but inactive in ToxCast. Each of these six chemicals had a PAC between 10 and 100 µM. A total of 133 chemicals were active in both RTgill-W1 CP and ToxCast and 66/133 (49.6%) were more potent in ToxCast by at least one order of magnitude.

A small number of chemicals were identified where the PAC in RTgill-W1 cells was markedly less (i.e., more potent) than the ToxCast POD (Fig 5B). These chemicals included carbendazim (DTXSID4024729), codeine (DTXSID2020341), octinoxate (DTXSID1025302), etodolac (DTXSID9020615), indapamine (DTXSID7044633) and L-ascorbic acid (DTXSID5020106). Each of these chemicals has been tested in more than 200 ToxCast assays. The percentage of active ToxCast assay endpoints out of the total number of ToxCast assay endpoints where these chemicals had been tested was relatively small for indapamide (3 / 237 = 1.27%), octinoxate (5 / 341 = 1.47%), etodolac (4 / 234 = 1.71 %), codeine (5 / 237 = 2.11 %) and L-ascorbic acid (10 / 393 = 2.54 %). In contrast, carbendazim was active in 10.18 % (63 / 619) of ToxCast assay endpoints tested, which is equal to the median number of ToxCast assay endpoints that were active for all chemicals in the RTgill-W1 test set (median: 10.18 %, range: 0 – 56.9 %).

Polyaromatic hydrocarbons (PAHs) (Fig 5A, squares) were among the most potent chemicals observed in the RTgill-W1 CP assay. Other potent chemicals that were not PAHs included the piscicide rotenone (DTXSID6021248), 5,6-benzoflavone (DTXSID8030423), hexachlorophene (DTXSID6020690) and codeine (DTXSID2020341), L-ascorbic acid (DTXSID5020106) and carbendazim (DTXSID4024729) (Fig 5C). A table listing PACs and ToxCast *in vitro* PODs is provided in the Suppl. File 9.

Development of a New *in vitro* Disposition Model Suitable for RTgill-W1 Cells

In contrast to most mammalian cell lines, RTgill-W1 are exposed to the test chemicals in a minimal medium (L15/ex) without the presence of FBS. FBS has been shown to help keep hydrophobic chemicals in culture and decrease their sorption to plastic (F.C. Fischer et al. 2019). Hence, many *in vitro* disposition models do not consider sorption to plastic, which is less problematic if FBS is supplemented but not an appropriate assumption for serum-free medium.

We expanded an *in vitro* disposition model based on multiple previous partitioning-based models (Proença et al. 2021) with a kinetic model for uptake into plastic (Fischer, Cirpka, et al. 2018) that additionally accounted for the kinetics of cellular uptake (Fischer, Abele, et al. 2018). The model consists of three compartments: medium, cells, and plastic (polystyrene) (Fig 6A). Cells consist of defined fractions of water, proteins and lipids, and the L15/ex medium is considered as pure water for the purposes of this study. Uptake into the cells and sorption to plastic are both modeled as kinetic processes, whereas partitioning between water, protein and lipid fractions is assumed to be in equilibrium instantaneously.

The model requires information about the geometry (e.g., well dimension, number of cells), the biological system (e.g., lipid and protein content of cells and medium, pH) and modeling parameters (assay

duration and time steps). Furthermore, chemical-specific parameters, such as logK_{ow}, logK_{aw} and pK_a need to be provided (Fig 6B). As we aimed for a high-throughput approach, these values were obtained from databases with experimental or predicted values (Fig 6B, Suppl. Method 1).

We implemented the model in both an Excel spreadsheet for analysis of individual chemicals and as a *R* script to run in batch mode over many chemicals. Of the 225 chemicals, inorganic chemicals, chemicals without defined structure, and zwitterionic chemicals were excluded, resulting in 205 chemicals eligible for modeling (Fig 6C). The test chemicals were selected based on relevance for ecotoxicology, not constrained to particular physicochemical properties. Many chemicals were neutral at the tested pH of 6.8 (135/205, 65.9%), but 13 and 23 chemicals were greater than 50% positively or negatively ionized, respectively, at experimental pH (Fig 6D). The distribution of logK_{ow} values ranged from -4.7 to 10.3 (Fig 6E). Most chemicals (149/205, 72.7%) had logK_{aw} values between -9 and -4. However, the remaining 56 chemicals had values > -4, which is the empirical cutoff for loss of chemicals through partitioning from medium to air (Fig 6F) (Escher et al. 2019).

Upon application of the new IVD model to the 384-well RTgill-W1 model system, four chemicals were flagged as resulting in negative concentrations, all of which had logK_{ow} >7. For the remaining 201 chemicals, a broad range of chemical medium concentrations was predicted (Fig 6G). For 155/201 (77.1%) chemicals, the medium concentration was greater than 30% of the nominal concentration, while medium concentrations of up to 200-fold below the nominal concentration were predicted for the remaining chemicals. The medium concentration was strongly dependent on the logK_{ow} for chemicals with >3. Chemicals with a logK_{ow} of 4 had 10-30% of the chemical available in the medium, and chemicals with a logK_{ow} >6 had less than 3% of the chemical available in the medium.

The model was validated with experimental data with chemical quantification from medium, cells and plastic for four chemicals (see Supplemental Method 2). A fifth chemical, pyrene, was volatile (logK_{aw} = -3.31) and could not be detected in the samples after 24 h of exposure. Two chemicals (malathion and imidacloprid) were neutral at the pH of the system and had a logK_{ow} < 3 (hydrophilic chemical) and thus were expected not to partition to the plastic (malathion 5% in plastic, 93% in medium; imidacloprid 0.5% in plastic, 99% in medium), which was confirmed by the experiments (see Supplemental Methods 2). Triazophos has a logK_{ow} of 3.3 and 90% was detected in the medium, the remainder in plastic, which agreed well with the modelling results of 83% in medium, <1% in cells and 17% in plastic. Fluoxetine (logK_{ow} of 4) is positively charged at pH 6.8 but still has an experimental logD_{lipw} of 3.84 (Neuwoehner et al. 2009) and, accordingly, partitioned into cells with a fraction of 63% but was also found in medium

(26%) and plastic (12%) after 24h exposure, while the model predicted no uptake into plastic, 23% in cells and 75% in medium.

The kinetically-resolved *in vitro* disposition model captured these analytical results qualitatively (see Supplemental Methods 2). However, due to the low number of chemicals analyzed and the variability of the experimental data, a more thorough validation of the model was not possible.

Comparison to *in vivo* Literature Data

Finally, the PACs generated in this study were compared to *in vivo* mortality data from rainbow trout (RBT) and fathead minnow (FHM) extracted from the ECOTOX knowledgebase as described in Fig S11. For 93 unique chemicals, at least 3 records of *in vivo* test results were available. The chemical with the most records was pentachlorophenol with 165 records, which was present in the screen in two forms (pentachlorophenol and sodium pentachlorophenate) of different molecular weights. For some chemicals the *in vivo* effect results varied by several orders of magnitude (Fig S12). For example, trichlorfon effect data ranged from 0.1 to 100 mg/L (68 total records). On the other hand, fenitrothion effect data ranged from 1 to 10 mg/L (68 total records). For some chemicals, differences in species sensitivity were apparent. For example, RBT were more sensitive to trichlorfon and azinophos-methyl than FHM. Overall, for chemicals with potent effective concentrations (Fig S12, bottom) CP was typically less sensitive than the *in vivo* test results. On the top of the graph are chemicals with high effect values (i.e., low potency), many of which were not tested at sufficiently high concentrations in CP to elicit a phenotypic change.

For each chemical a single *in vivo* LC was derived by summarizing all records. For chemicals with at least 20 records, the 5^{th} percentile of the individual lethal effect values was calculated, while for chemicals with less records the first quartile (25^{th} percentile) was calculated. This summarized *in vivo* LC spanned eight orders of magnitude (from 10 ng/L for antimycin to 1000 mg/L for 2-methyl-1-propanol) (Fig 7A). PACs spanned 4.5 orders of magnitude (from 3 µg/L to 30 mg/L). Of the 93 chemicals, 66 were active in CP (71%). Inactive chemicals are represented at their highest tested concentration. None of the chemicals with low *in vivo* LC were inactive in CP; all inactive chemicals had *in vivo* LC > 100 µg/L. Accounting for *in vitro* disposition leads to a left shift of data points for chemicals with a low free concentration. Of note, chemicals with PACs greater than *in vivo* LC have improved predictions, indicated by a shift toward the unity line, indicating an improvement in prediction (Fig 7A).

To continue the analysis, only chemicals for which a PAC could be calculated were included (n=64). Rotenone and antimycin A were excluded because they were bioactive at the lowest tested concentration.

A substantial amount of test chemicals (17/64; 27.3%) were flagged for volatility concerns, and we hypothesized that the *in vitro* model would underestimate the bioactivity of volatile chemicals. The difference between the PAC and the *in vivo* LC for volatile chemicals was larger than for less volatile chemicals (Fig 7B). The effect was stronger for nominal PACs as compared to adjusted PACs but was not statistically significant. Overall, for the 59 active chemicals for which physicochemical properties were available and the adjusted PAC could be calculated, the RMSE on the $\log_{10}$-µM scale was 1.27, resulting in 59.3% and 86.4% of chemicals predicted within one and two orders of magnitude from the *in vivo* LC, respectively. For 72.9% of chemicals, the adjusted PAC was lower than, or within one order of magnitude of, the *in vivo* LC, which we considered here as "protective" consistent with previous work (Nyffeler et al. 2020).

The 16 chemicals with a poor prediction (PAC more than 1 order of magnitude higher than the *in vivo* LC) are labeled in Fig 7C. Of these, thirteen chemicals had more than ten *in vivo* records. They included four organophosphates, four organochlorides, three pyrethroids, a carbamate and copper chloride ($CuCl_2$). The three chemicals with less than ten *in vivo* records were the fungicide chlorothalonil (DTXSID0020319), dibutyl (2E)-but-2-enedioate (DTXSID7021865) and the insecticide abamectin (DTXSID8023892).

When categorizing the results by chemical groups (File S1), it became apparent that the prediction capacity varied strongly among the different chemical groups (Fig 7D). For example, many chemicals with a neurotoxic potential were among the poorly predicted chemicals. Indeed, for the two carbamates and the three pyrethroids, the PAC was higher than the *in vivo* LC. Pyrethroids were statistically significantly enriched in the "not protective" space (p-value of the Chi2 test: 0.016). For the four phthalates, the PAC was lower than the *in vivo* LC. Interestingly, the three fungicides, the three chlorophenols and the six herbicides were well predicted (less than 1 order of magnitude difference to the *in vivo* LC).

Discussion

About 35,000 to 50,000 fish are used in acute tests each year to assess ecotoxicological safety in the European Union alone (European Commission 2020). Tanneberger et al. (2013) demonstrated that an *in vitro* cell viability assay with RTgill-W1 cells was predictive of the LC50 for FHM, however the assay is currently low-throughput. A high-throughput assay could significantly reduce the number of fish used for toxicity testing in the future.

We aimed to miniaturize the existing OECD TG 249 assay to 384-well format (which we refer to as CV-PR), and additionally establish CP as a screening approach for chemical bioactivity in an ecological receptor species. We hypothesized that potencies derived from CP would be more sensitive than potencies from cell viability assays and also protective of acute lethality in fish.

Miniaturization and Selection of Reference Chemicals

To begin, we adapted the OECD TG249 assay to 384-well format and supporting microfluidics devices and refer to this version of the assay as CV-PR. We then tested 23 chemicals in the CV-PR, CV-IB and CP assays to select suitable reference chemicals for the subsequent screening. The criterium for a control chemical for CV-PR and CV-IB was a robust decrease in cell viability (ideally down to 0% viability) in a concentration-dependent manner. Criteria for phenotype altering reference chemicals included the magnitude of the observed response and the range of concentrations where phenotypic responses were observed in the absence of cytotoxicity. An aim was to ensure that the phenotype altering reference chemicals selected produced diverse phenotypic responses in multiple organelles. Out of the potential candidates, EICO, cucurbitacin I, and docetaxel were selected. Effects of the phenotype altering reference chemicals ranged from subtle to very pronounced and one chemical from the selected set (EICO) produced cytotoxicity. 3,4-DCA was also selected as a reference chemical to be consistent with OECD TG249, although it was not very cytotoxic to RTgill-W1 cells in the present study within the concentration ranges tested. The only CV endpoint with a BMC for 3,4-DCA was propidium iodide (BMC ~ 400 µM). The EC50 was not reached in any of the CV-PR endpoints, although according to OECD TG249, the EC50 for CV-PR ranges from 269 to 362 µM (OECD 2021). Of note, 3,4-DCA affected the cellular phenotype reproducibly at 1-10 µM (Fig S7). Thus, this set of reference chemicals, tailored to RTgill-W1 cells is comparable in function to previous phenotype altering reference chemical sets tailored to human-derived cell types

(Culbreth et al. 2022; Nyffeler et al. 2023) and includes the chemical used as an assay control within OECD TG429. .

RTgill-W1 Screening Results

QC metrics

Using the reference chemicals, performance of the assays was monitored during the screening phase (Fig S5-S7). Reproducibility of the BMCs for reference chemicals was high for all endpoints, with a standard deviation of < 0.2 orders of magnitude. For chemicals tested in duplicate, the concordance was high, and the BMCs were typically within ½ an order of magnitude. These results are consistent with our previous screens in U-2 OS cells (Nyffeler et al. 2020; Nyffeler et al. 2023) and demonstrate good assay reproducibility throughout the screening portion of this study.

Screening results

Previous studies in human-derived cell models have demonstrated that CP is more sensitive than cell viability assays for detecting biologically active chemicals (Nyffeler et al. 2020; Culbreth et al. 2022). In a study of human U-2 OS osteosarcoma cells, changes in cell morphology were observed in absence of cytotoxicity or apoptosis (Dahlin et al. 2023). Further, in concentration-response studies of U-2 OS cells, there were no chemicals that were active in a cell viability assay that were not also active in CP, and PACs were often much lower than the threshold for cytotoxicity (Nyffeler et al. 2020; Nyffeler et al. 2023). The same phenomenon was observed in the present study in RTgill-W1 cells. There were no chemicals that were active in the CV-IB or CV-PR assay that were not also active in the CP assay, although there was some discrepancy in the number and identity of chemicals that were identified as active in the CV-IB and CV-PR assays (Fig 3A). These chemicals tended to have cell viability potencies approximately, or above, 10 µM in the viability assay in which they were active (Fig 3B). PACs were always more sensitive than the most sensitive potency reported across the CV-IB and CV-PR assays (Fig 3C). The enhanced sensitivity of CP compared to conventional cytotoxicity assays was not unexpected, as it makes sense biologically for changes in cell morphology to occur at concentrations lower than those that trigger apoptosis, disrupt membrane integrity or are otherwise overtly cytotoxic. However, to our knowledge, this is the first demonstration of this phenomenon in the RTgill-W1 cell model. The implication for bioactivity screening of environmental chemicals is that the CP assay will identify a broader variety of chemicals as active, and at lower concentrations, compared to a conventional cytotoxicity assay and, thus, is a more sensitive (i.e., conservative) approach for detecting potential ecotoxicity hazards.

Of note, the false positive rate for the CP component of the present study was also evaluated using the "null chemical" modeling approach, as previously described (Nyffeler et al. 2021). The false positive rate was low; 1.0 % corresponding to 1 out of 96 null chemicals as active. Therefore, while CP in RTgill-W1 cells is sensitive for detecting chemical bioactivity, it does not appear to be overly sensitive for classifying chemicals as active. One false positive null chemical was classified as active using the category-level Mahalanobis distance approach. There were no false positives identified using the global Mahalanobis distance approach. It has been previously shown in the U-2 OS cell model that the category-level Mahalanobis distance approach is more sensitive (i.e., results in lower PACs and greater numbers of chemicals identified as active) compared to the global Mahalanobis distance approach, but it is also slightly more prone to identification of false positives (Nyffeler et al. 2021; Nyffeler et al. 2023). These same trends were observed in the RTgill-W1 model in the present study. The present study adds to the growing body of evidence that the category-level approach is more sensitive (i.e. conservative) than the global modeling approach for CP data.

Comparison of RTgill-W1 Results to Previous RTgill-W1 Studies (OECD TG 249)

Differences from the OECD TG 249 Procedure

RTgill-W1 cells have been used for toxicity testing for decades (Schirmer et al. 1997; Tanneberger et al. 2013; M. Fischer et al. 2019), and a 24-well plate-based OECD test guideline has been established recently (OECD 2021). In adapting the assay to higher-throughput format, we attempted to follow the test guideline where possible. However, while miniaturizing the assay to a 384-well plate format, a few modifications were required.

First, media replacement in 384-well format is more challenging than in 24-well format. We aimed to remove the complete medium using a CyBio Felix automated pipettor and replace it with L15/ex. We did not include a wash step with L15/ex as suggested in TG249 in order to minimize the possibility of disrupting RTgill-W1 cell monolayers. This modification could have lead to trace amounts of FBS being left in the exposure solution that could affect the amount of bioavailable test chemical, as well as the behavior of the cells. On a similar note, we used heat-inactivated FBS in contrast to the recommendation in TG249.

Second, 3,4-DCA did not impact cell viability to the extent described in OECD TG249 (Fig S3). The expected EC50 for CV-PR was 370 µM (or 60 mg/L) and the highest tested concentration (600 µM = 97 mg/L) should reduce metabolic activity to 2%. However, we did not observe any effect in the CV-PR assay. The cell number was about 60% of controls and the BMC for the PI endpoint was 400 µM, indicating that

3,4-DCA was cytotoxic to the cells at high concentrations. A possible explanation was that the 3,4-DCA was applied at an incorrect concentration or was of low quality. However, the CP results indicated that cells treated with 3,4-DCA at concentrations as low as 1 µM had altered phenotypes, indicating that there was chemical present. The chemical did affect the phenotype but did not exert cytotoxicity in the CV-PR assay up to 600 µM. Therefore, we chose EICO as a viability positive control for the screening phase.

Third, we observed high variability in the CV-PR assay results including solvent control wells, and we were not able to identify a cause. The high variability translated to large cutoff values for concentration-response modeling (Fig 2). In contrast, variability in the CV-IB assay was very low and cutoff values were smaller as compared to CV-PR endpoints. For example, if a treatment induced 20% cell loss, this would be within the noise band of the CV-PR assay, but above the cutoff for the CV-IB assay. Hence, the CV-IB assay would result in a BMC while the CV-PR assay would not. Of note, if both assays detected cytotoxicity, the results were comparable (Fig 3B). One advantage of the CV-IB is that fewer washing steps are required for endpoint detection, saving time. However, the CV-IB assay requires a fluorescence microscope for detection as opposed to a fluorescence plate reader, which is more expensive and might not be available in every lab. Based on our results (Fig 3B), we advocate that the CV-IB assay should be considered equivalent to the CV-PR assay. In our view, the CV-IB assay is more automation friendly and better suited for high-throughput applications.

Comparison to RTgill-W1 Literature

We compared the PACs from this study, conducted in 384-well format, with EC50 data from the literature, mostly conducted in 24-well format. We found good agreement with the literature data (Fig 4). Two chemicals (aniline and 1-hydroxybenzotriazole) were more than 20x more potent in our study compared to previous studies. However, it should be noted that in the current study we derived a PAC, not an EC50 for cytotoxicity, and hence the difference in potencies can be explained by phenotypic changes happing at sub-cytotoxic concentrations. Our results indicated that the PAC serves as a lower bound to the EC50.

Comparison of RTgill-W1 Results to ToxCast

Beginning in 2007, the US EPA's ToxCast program aimed to characterize the biological activity of hundreds of chemicals that were either present in the environment or of programmatic interest using a variety of *in vitro* cell-based and cell-free assays (Richard et al. 2016). Assay endpoints in ToxCast are mainly from mammalian (human and rodent) test systems (Dix et al. 2007; Judson et al. 2010). Even

though ToxCast is mainly focused on mammalian biology, the utility of the ToxCast assay data for ecological hazard assessment in non-mammalian species has recently been assessed by several groups (Schaupp et al. 2023; Rodea-Palomares and Bone 2024). Overall, correlations between PODs derived from ToxCast and PODs derived from *in vivo* ecotoxicity data (i.e., ECOTOX Knowledgebase) were low. However, the PODs derived from ToxCast were protective, being less than, or equivalent to, the *in vivo* ecotoxicity PODs in many cases. In the present study, for most chemicals, the ToxCast POD was more sensitive (i.e., lower µM concentration) compared to the RTgill-W1 PAC (Fig 5). This finding is consistent with previous studies comparing HTPP PACs from U-2 OS human osteosarcoma cells to ToxCast PODs (Nyffeler et al. 2023) and was not unexpected. The ToxCast assay collection contains assays for molecular targets whose homologues may not be expressed in RTgill-W1 cells. ToxCast PODs that are lower than PACs observed in RTgill-W1 cells may be due to the absence of those molecular targets in the RTgill-W1 model. Alternatively, the differences in ToxCast PODs and RTgill-W1 PACs may reflect species differences between mammalian and non-mammalian models and differences in assay technology sensitivities: e.g., a combination of cell-free and cell-based in ToxCast versus cell-based in the present study.

For six chemicals, the PAC was ≥ 1 order of magnitude lower than the corresponding ToxCast *in vitro* POD (Fig 5B). One of those chemicals was L-ascorbic acid (vitamin C, DTXSID5020106), an antioxidant frequently used as a food preservation additive, with an PAC of 66 nM. We did not expect this chemical to be bioactive at such a low concentration and hypothesized that it was a false active. While L-ascorbic acid was inactive according to the global Mahalanobis endpoint, BMCs could be determined for a few category-level Mahalanobis distance endpoints (https://doi.org/10.23645/epacomptox.26799382). The affected categories all comprise features of the RNA and ER channel. Inspection of the corresponding microscopy images confirmed that cells treated with L-ascorbic acid display small, bright spots in the green (i.e., RNA and ER) channel (Fig S13). Hence, L-ascorbic acid is a true active chemical, but the biological reason and relevance remains unclear at this time.

Another chemical that was more potent in RTgill-W1 than in ToxCast was octinoxate (DTXSID1025302) with a (nominal) PAC of ~1 µM. It was active in only 5/341 ToxCast assays but was cytotoxic at >= 30 µM in RTgill-W1 cells. The chemical is used in sunscreens, and its use has been limited in Hawaii due to concerns about its impact on coral reefs (Galamgam et al. 2018)(https://www.acs.org/molecule-of-the-week/archive/o/octinoxate.html). Only few toxicity studies in fish have been conducted with inconclusive results (Nataraj et al. 2020; Damiani et al. 2023). Considering that only about 5% of the chemical is freely available according to our IVD model, the adjusted PAC is ~ 50 nM.

Based on these results, further studies are warranted to investigate whether the observed *in vitro* bioactivity is relevant to fish toxicity.

Overall, the reasons underlying the greater sensitivity of the RTgill-W1 assay to specific chemicals are unclear but it does not appear to be due to lack of activity in a large number of ToxCast assays. There were other chemicals where the percentage of active ToxCast assay endpoints was low (ethanolamide and copper(II) chloride as examples), but the PAC and ToxCast POD were similar. Differential species sensitivity could be a factor, but additional studies would be needed to support or refute this hypothesis.

In Vitro Disposition (IVD) Model

To our knowledge, this is one of the first studies conducting high-throughput phenotypic profiling in a non-mammalian cell line. With the cell line being exposed to chemicals in medium without protein and lipid supplements, it was unclear whether existing IVD models were appropriate. The FBS that is typically present in cell culture medium acts to retain chemicals in the medium, not in a freely dissolved form but bound to serum proteins. This reduces the freely dissolved concentrations and theoretically may enhance the availability of chemical to the cells and keep the system more stable as aqueous concentrations depleted by cellular uptake are replenished by desorption from medium components including lipids and serum proteins. Serum-containing medium also leads to lower sorption to plastic than serum-free medium. In the absence of serum, chemicals can partition to plastic to a greater extent, whereas in the presence of serum, chemicals can bind to serum components, thereby decreasing the amount that would adsorb to the culture vessel (Fischer, Cirpka, et al. 2018).

One of the most widely used *in vitro* disposition models applicable to high-throughput screening with diverse chemicals (i.e., varying logK_{ow}, neutral and charged chemicals) was developed by Armitage et al. (2013; 2021). Unlike our model, which implemented kinetic processes, this previous model assumes that steady state is reached for uptake into cells, evaporation and binding to plastic instantaneously. Due to absence of proteins and lipids in the RTgill-W1 exposure media, the model developed by Stadnicka-Michalak et al. (2021) is more relevant to the RTgill-W1 culture settings. Their study was more focused on volatile chemicals, and hence they used a plate seal – a major difference to our setup. As we assumed that the cellular uptake takes up to 6 h, initially more chemical is available for sorption to the plastic. The Armitage and Stadnicka-Michalak models may underestimate the amount of plastic binding that occurs during the chemical exposure scenario (24 h) used in our assay.

For high-throughput screening experiments incorporating chemicals with a broad range of physicochemical properties, accounting for IVD becomes critical. The IVD model presented herein identified 25 chemicals for which the predicted free concentration was more than 10x lower than the nominal concentration. These deviations highlight the importance of accounting for IVD, as otherwise the predicted effective *in vivo* exposure concentration would be overestimated by approximately one order of magnitude.

Comparison of RTgill-W1 Results to *In vivo* Fish Lethality Data

Of the 225 tested chemicals, 93 chemicals had enough existing *in vivo* data for comparison. To our knowledge, this is the largest comparison of *in vivo* fish data with *in vitro*, cell line-based data. Consistent with previous studies (Tanneberger et al. 2013; Stadnicka-Michalak et al. 2021), accounting for the *in vitro* disposition improved the prediction capability of the model.

We noted that, in particular, chemicals acting through a neurotoxic mechanism (organophosphates, organochlorines, pyrethroids, carbamates) were not predicted well with the RTgill-W1 model. This has already been noted by Tanneberger et al. (2013), who found that while the three acetylcholinesterase inhibitors (malathion, parathion ethyl, disulfoton) were well predicted, the other three tested neurotoxic chemicals (permethrin, caffeine, lindane) were poorly predicted. In our study, parathion ethyl and disulfoton were also well predicted (in agreement with Tanneberger et al. 2013) but malathion was poorly predicted. Permethrin and lindane were also poorly predicted in our study; caffeine did not have enough *in vivo* records to derive an *in vivo* LC. On the other hand, fungicides, herbicides and chlorophenols were well predicted with the RTgill-W1 model. We conclude from this that the RTgill-W1 model is suitable for predicting the toxicity of chemicals that were not designed to act on a molecular target in fish but may be acting in a non-specific manner. However, the RTgill-W1 model is not protective for chemicals that are neuroactive or that may require metabolism to become neuroactive. Coupling the RTgill-W1 assay with another assay that can detect chemicals that are neuroactive may provide improved protection in the context of ecological risk assessment.

Uncertainties

Limitations of Comparison to *In Vivo* Data

The comparison to *in vivo* data is limited by the fact that there is considerable variability in the *in vivo* data themselves. In this study, we retrieved data from the ECOTOX Knowledgebase, which contains

studies of varying design and quality. While we restricted the data to only include fish of a certain size and did not consider chemical formulations, the records varied in exposure duration and temperature. Only a minority of records analytically verified the chemical concentration. Further, we included multiple types of endpoints (such as lethal concentrations, lowest observed effect concentration, etc.).

In a few cases, the chemical identity was different between the ECOTOX records and the current *in vitro* screen. For example, abamectin is a mixture of B1a [CASRN: 65195-55-3, MW= 873 g/mol] ((S)-sec-Butyl)- form and B1b [CASRN: 65195-56-4, MW= 859 g/mol] (isopropyl)- form (CompTox Chemicals Dashboard v.2.3.0). For the *in vivo* experiments, abamectin containing 93.3% B1a was used (Jencic et al. 2006). The composition of abamectin used for the *in vitro* study is not exactly known, but DTXSID8023892 is reported as a 4:1 mixture (CompTox Chemicals Dashboard v.2.3.0). Similarly, the pyrethroid fenvalerate (DTXSID101017940) is a mixture of four isomers. In this study, (R,R)-Fenvalerate (DTXSID3020621) was tested, while the *in vivo* records were obtained from fenvalerate (DTXSID101017940). For abamectin and fenvalerate, a difference in composition could contribute to the poor agreement between *in vitro* and *in vivo* results (Fig 5).

In this study, we used *in vivo* records from two species, RBT and FHM. This is in contrast to the (Tanneberger et al. 2013) study, which compared the RBT *in vitro* data to *in vivo* FHM data. We decided to include both species to have more records for the comparison, potentially reducing the uncertainty. However, differences in species sensitivities also increased the variability for some chemicals (Fig S14). Even within a species, for many chemicals the *in vivo* LC spanned one order of magnitude. This is in line with (Pham et al. 2020) who estimated the uncertainty in toxicity studies of mammal species (i.e., mouse, rat, dog, and rabbit) to amount to an RMSE of 0.47 to 0.63 orders of magnitude.

Uncertainties of the IVD Modelling

Unfortunately, to date, we have only measured five chemicals analytically to attempt to verify the IVD model predictions (Supplementary Method 2), and the chemicals did not represent the full physicochemical diversity present in the chemical set. For example, no negatively charged chemical was analyzed. A more thorough validation of the model was not possible as part of the present study and should be undertaken in the future.

Unlike the Armitage model, our model does not account for evaporation of test chemicals. This is because the experimental setup has no fixed air space; and loss to air is strongly dependent on temperature and geometry of the bioassay. Instead, we flag all chemicals with $\log K_{aw} > -4$ as possible

volatile to alert the experimenter. This "volatility cutoff" at a logK_{aw} of 4 was empirically determined for human cell lines incubated at 37°C (Escher et al. 2019). As the RTgill-W1 cells were exposed at 19°C, this cutoff might in reality be at a higher logK_{aw}. The model furthers assumes that plastic covered by cells is unavailable for chemical sorption; however, this assumption has not been evaluated experimentally. Moreover, the cell culture plates are "tissue culture-treated", i.e., they were oxidized, which led to hydroxylation and carboxylation of the surfaces, and, hence, the plate surface is negatively charged (Lerman et al. 2018) . As such, cationic chemicals, such as fluoxetine, are likely to bind stronger to the plate material. This is consistent with our findings (12% measured in the plastic vs 0.1% predicted). At present, we do not have sufficient data to incorporate this process into the mass balance model.

Further limitations come from the uncertainty of input parameters into the model. It was difficult to accurately measure the cell volume (Supplementary Method 3). The lipid and protein content of RTgill-W1 cells was measured previously (Stadnicka-Michalak et al. 2021), but its uncertainties were not reported. We assumed the protein and lipid concentration in the exposure medium to be 0, however, it is possible that small amounts of FBS remained during the exposure, affecting the amount of freely available chemical. Similarly, the pH affects the speciation of ionizable test chemicals, in particular those with a pKa close to the pH of the medium. The recommended pH for the L15/ex is 7.0-7.5 (OECD 2021). We retrospectively measured the pH of L15/ex in our laboratory and found it to be 6.8. Strong acidic or basic chemicals could also change the pH, as the L15/ex has virtually no buffer capacity.

Model prediction is also affected by uncertainty in the physicochemical parameters of the test chemicals. For 50 test chemicals, the predicted logK_{ow} was used. Binding to lipids and proteins was predicted using the LSER model, not directly measured experimentally. For example, for fluoxetine, the logD_{lipw} was predicted with our model to be 3.25, whereas it has been determined experimentally at pH 7.0 to be 3.84 (Neuwoehner et al. 2009). Experimentally derived logK_{ow} values would be valuable for improving predictions but in many cases are not available.

For Armitage et al.'s model, chemical-specific parameters are input as a list into an Excel sheet. We followed this approach, as we consider it to be user friendly and to make it easy for a user to switch between the previous model and our new model. The system and model specific parameters are also input as an Excel sheet. The model runs in an R script, which makes all modeling steps completely transparent. In the future, the code could be easily adapted to other experimental systems.

To summarize, our IVD model is in a user-friendly format and fully transparent. Model validation needs to be expanded. The open-source code allows adaptation to other experimental settings or the incorporation of other processes (i.e., evaporation, biotransformation) in the future.

Other Factors of Uncertainty

We also noted that our assay had a lower prediction capacity for volatile chemicals than for non-volatile chemicals. This could be due to the higher surface-to-water ratio in the 384-well plate as compared to a large fish tank. Additionally, the *in vivo* studies were conducted at various temperatures, and lower temperatures would also decrease the volatility of test compounds. Sealing the plates with adhesive foil during the exposure period (as done by M. Fischer et al. (2019)) could be considered in the future to reduce the loss of volatile chemicals.

Summary & Impact

We present here two NAMs: (1) a high-throughput *in vitro* assay for chemical bioactivity screening in RTgill-W1 cells and (2) a corresponding *in silico* IVD model, which, in combination, were developed for use in predicting a lower bound to the chemical concentration that would be lethal to fish. This approach led to an *in vitro* potency estimate within 10-fold of the *in vivo* LC for 59% of the tested chemicals (Table 1). For 73% of chemicals, the potency estimate was protective (i.e., within 10-fold of or lower than the *in vivo* LC) relative to the *in vivo* LC; this value was slightly lower than previously published effect concentrations from a smaller chemical set and a larger well format (Tanneberger et al. 2013) (Table 1). In the later study, *in vitro* values were compared to *in vivo* data from FHM only and FHM are, for many chemicals, less sensitive than RBT (Fig S12). Thus, the comparison between both studies is not fully quantitative. In both studies, accounting for the freely dissolved concentration improved the prediction capacity.

Table 1. Comparison of *In Vivo* Toxicity Predictions.

	Tanneberger 2013		**This study**	
Test system	24-well CV-PR assays		384-well Cell Painting	
In vivo data	Fathead minnow		Rainbow trout + Fathead minnow	
	Nominal	Analytical chemistry (C0, C24)	Nominal	IVD modeling
N chemicals	35	29	64	59
Difference < 5 fold [a]	18 (51.4%)	20 (69.0%)	26 (40.6%)	25 (42.3%)
Difference < 10 fold [a]	23 (65.7%)	22 (75.9%)	34 (53.1%)	35 (59.3%)

Protective (10 fold) [a]	28 (80.0%)	24 (82.8%)	41 (64.1%)	43 (72.9%)
RMSE (log_{10} µM)	**1.27**	**1.14**	**1.55**	**1.27**
[a] Values in parenthesis are the number of chemicals meeting the criteria divided by N chemicals.				

By miniaturizing the existing RTgill-W1 assay, we were able to generate *in vitro* data for >200 chemicals in a high-throughput manner, compared to less than 100 chemicals obtained previously (Schirmer et al. 1998; Tanneberger et al. 2013; Natsch et al. 2018). This illustrates the potential of these new NAMs for rapid generation of ecotoxicity hazard information.

In the future, this NAMs combination could be used to prioritize chemicals (Fig 8). For example, upon development of new plant protection products, chemical scaffolds with lower predicted toxicity to fish could be prioritized. This NAM combination could also be used as a pre-test to select the concentration(s) for *in vivo* tests or the OECD 249 guideline study in support of a regulatory submission. According to this study, the PAC divided by a factor of 10 has a 73% chance to be below the LC50 in fish. This may contribute to a reduction in the number of animals tested (less substances and concentrations tested) or a refinement in animal testing practices (setting a maximal tolerable dose based on the adjusted PAC).

Outlook

For 27% of the tested chemicals the predicted effect concentration was not protective for fish, which implies that not all possible toxicity mechanisms can be detected in the current test system. Additional tests are likely needed to expand coverage of biological space. In particular, many of the poorly predicted chemicals have a neurotoxic mode-of-action and some chemicals require metabolic activation to exert toxicity.

At present, we compare a measure for bioactivity (change in phenotype) to a lethal outcome (toxicity in fish). The hypothesis is that a chemical will first exert a phenotypic change in the cells that later could translate to cell stress, cytotoxicity, organ damage and finally death. In the future it would be interesting to compare the PAC to non-lethal endpoints from *in vivo* studies. Finally, the combined approach of HTPP and IVD modeling could be extended to other species or taxa that are difficult to test *in vivo* or are otherwise underrepresented in toxicity testing.

Data Availability

CV-PR, CV-IB and CP screening results and summary plots of results from each assay for each chemical are available at https://doi.org/10.23645/epacomptox.26799382.

Figure Legends

Figure 1: Graphical overview of experimental schedule and assay concepts.

(A) Graphical representation of cell culture and sampling schedule. Each arrow or arrowhead corresponds to one day. Cells are thawed from passage 8 (P8) vials and grown for seven days with a media change on day 4, then passaged to passage 9 (P9) with another seven-day growth period and a media change on the fourth day. Passage 10 (P10) cells are then plated into 384-well format in RTgill-W1 growth media and cultured for one day before media is replaced with L-15/ex and chemicals are applied. 24 hours post application the cells are then sampled for one of the three assays. (B) Graphical representation of the output for the imaging-based cell viability (CV-IB) assay, including two images of RTgill-W1 cells (top) which display high viability (left) and low viability (right). (C) Graphical representation of the four fluorescent channels measured in Cell Painting across seven vertical Z-planes. Z-stack images are combined into maximum projection images for analysis. (D) Graphical representation of the biological principles underlying the three endpoints for the plate-reader based cell viability (CV-PR) assay, describing the conditions that yield fluorescence signal for viable and non-viable cells.

Figure 2: Reference chemicals for cell viability and Cell Painting in RTgill-W1

(A) Concentration-response curves for the five cell viability endpoints after exposure to 5,8,11-eicosatriynoic acid. The shaded regions span the mean +/- nMAD of vehicle controls. Each data point is from a single well; 4 experiments with 2 technical replicates were conducted. Calculated benchmark concentration (BMC) values are represented by the solid vertical lines.(B) Concentration-response curves for the three selected reference chemicals fit to global Mahalanobis distance values from CP data. The shaded regions span the mean +/- 1 SD of vehicle controls. Calculated BMC values are represented by the solid vertical lines. Highly cytotoxic concentrations have been excluded from concentration-response analysis. (C) Representative images for the three selected phenotype altering reference chemicals at non-cytotoxic concentrations (right) and solvent control (left). The most notably perturbed channel is displayed for each chemical in addition to the DNA channel and is annotated to the side of each image. (D) Potency-magnitude plots for each of the three selected reference chemicals displaying all feature-level curve fits: each point represents an individual CP feature, with the *x*-value representing the feature-level BMC and the *y*-value representing the effect size (*top_over_cutoff* from *tcplfit2* modeling) for the feature. The gray shaded area indicates the range [-1, 1], which is considered the noise band. Features are color and shape coded by the fluorescent channel and cellular compartment from which they were derived,

respectively. Letters overlaid on the most potent points represent the type of feature: T = texture, C = compactness, P = profile, S = symmetry, I = intensity, R = radial distribution, M = morphology). Points to the far right of each plot are inactive features (hit call < 0.9).

Figure 3: Results from the 225 unique chemicals tested in the CV-PR, CV-IB and CP assays

(A) Venn diagram indicating the number of chemicals that were active in each of the three assays. (B) Scatter plot comparing the benchmark concentrations (in µM) from the image-based (CV-IB) and plate-reader based (CV-PR) cell viability assays. Each point corresponds to one chemical. Only chemicals active in at least one of the two assays are depicted. The root mean squared error (RMSE) was calculated for the 48 chemicals that were active in both assays. (C) Scatter plot comparing the CP potency with the potency from the two cell viability assays (in µM). The most potent BMC of the five cell viability endpoints was selected. Only chemicals active in at least one assay are depicted. (D) Insert: Venn diagram with the number of chemicals that were determined as active with each of two different approaches: global Mahalanobis versus category-level Mahalanobis. Scatter plot comparing the phenotype altering concentration (PAC) (in µM) calculated with the two different approaches. Only chemicals active in CP using at least one approach are depicted. In each panel, the solid line indicates unity; the dashed lines are at an offset of half an order of magnitude. In each panel, blue or purple arrows indicate that the true potency value is less than the value displayed in the direction that the arrow is pointing.

Figure 4: Comparison of RTgill-W1 PACs to existing RTgill-W1 EC50 values from the literature

Comparison of the nominal phenotype altering concentration (PAC) from this study with EC50 values from previous studies conducted in RTgill-W1 cells. Data were collected from multiple studies and comprised 47 data points from 46 unique chemicals. Chemicals that were inactive in the literature were depicted with an open triangle plotted at the highest tested concentration. The upward pointing arrow indicates that the true EC50 is higher than indicated in this graph. The left pointing arrow indicates that the PAC of the tested chemical was below the tested range and could not be defined. The solid line indicates unity and the dashed lines are at an offset of one order of magnitude from unity. Chemicals with EC50 ratios > 1.3 are labeled.

Figure 5: Comparison of RTgill-W1 Cell Painting screening results to the ToxCast POD

(A) Scatter plot comparing the phenotype altering concentration (PAC) (in µM) to the point-of-departure derived by summarizing data from ToxCast. For 207 chemicals, results from at least 200 ToxCast assays were available. Chemicals active in less than three ToxCast assays are depicted as inactive. The color of the point indicates the number of ToxCast assays for which a particular chemical was active. (B)

Table listing the eight chemicals with the largest different between the PAC and ToxCast POD. (C) Table listing the nine chemicals with the lowest PACs. Many of these are polyaromatic hydrocarbons, indicated with an asterisk.

Figure 6: A new *in vitro* disposition model accounting for sorption to plastic

(A) Schematic of the implemented *in vitro* disposition model, following the style used in Proença (2021). Chemicals partition to cellular lipids and proteins (solid arrows). Uptake into cells and sorption to the plastic are modeled as kinetic processes (dotted arrows). Only the neutral species can absorb to the plastic. (B) Input parameters required for the model. Chemical parameters were obtained or predicted using various methods (indicated in the last column). Abbreviations: OPERA: OPEn structure-activity/property Relationship Application; LSER: linear solvation energy relationships. (C) For 205 of the 225 tested chemicals, *in vitro* disposition could be predicted. (D) Chemical speciation of each of the 205 chemicals at the pH of the test system (6.8). On the left side are acids, on the right side are basic chemicals. (E) Distribution of molecular weight and logK_{ow} for all 205 chemicals. (F) Distribution of molecular weight and logK_{aw} for all 205 chemicals. The dotted line is at a logK_{aw} of -4. Chemicals above that line are potentially volatile. (G) Predicted chemical fraction that is freely available in the medium (timed average over the 24-hour exposure period) plotted as a function of the logK_{ow}. The histogram insert represents the number of chemicals for each mass fraction bin.

Figure 7: Comparison to existing *in vivo* lethality data from rainbow trout and fathead minnow

(A) Scatter plot comparing the phenotype altering concentration (PAC) (in mg/L) with the *in vivo* lethal concentration summarized from multiple ECOTOX Knowledgebase records. The solid line indicates unity, the dashed lines indicate an offset of one order of magnitude from unity. Light purple points indicate the nominal PAC; darker purple points the adjusted PAC, which are left shifted according to the timed average of the free mass fraction of a chemical. Chemicals inactive in HTPP are displayed with open circles. The left-pointing arrow indicate that the PAC was below the tested range and could not be defined. Crosshatched squares represent two different salt forms of sodium pentachlorophenate. (B) Representation of the same data plotted in panel A, but displayed as the difference between the *in vitro* PAC and *in vivo* lethal concentration. Chemicals are divided into non-volatiles and volatiles using the cutoff value of logK_{aw} of -4. The vertical, green, solid line indicates unity; the vertical, green, dashed line is at one, which means that HTPP was 10x less sensitive than the *in vivo* data. The bold numbers indicate the root-mean-squared-error (RMSE). The table contains the summary statistics for the data shown in the graph, with the following information: conc (concentration, either nominal or adjusted based on

modeling); group (whether chemicals were subset into volatile/non-volatile and which subset is represented); RMSE; n; within1, within2 (the percentage of PAC values within 1 or 2 order(s) of magnitude of the *in vivo* effect concentration); and protective1 (the percentage of PAC values which are within 1 order of magnitude or lower than *in vivo* lethal concentrations). (C) Scatter plot comparing the adjusted PAC with the *in vivo* effect concentration. Only chemicals active in HTPP are shown (n=69). The solid line indicates unity, the dashed line indicates an offset of 1 order of magnitude from unity. Chemicals where HTPP was more than 10x less sensitive than the *in vivo* effect concentration are labeled. (D) Representation of the same data plotted in panel C, but displayed as the difference of the *in vitro* PAC and the *in vivo* lethal concentration. Chemicals are divided into chemical groups based on structure and MoA (see Supplemental File 1). The vertical, green, solid line indicates unity; the vertical, green, dashed line is at 1, which means that HTPP was 10x less sensitive than *in vivo* data. For each chemical group, a two-sided Wilcoxon test was conducted to test whether the chemicals in the group have a different distribution than the remainder of the chemicals ($p < 0.05$).

Figure 8. Schema for use of *in vitro* and *in silico* NAMs for evaluating fish toxicity.

For any chemical of interest, the first step in this scheme is determining whether the chemical falls into a domain of applicability based on logK_{aw} and logK_{ow}. If so, test in the HTPP assay and conduct *in vitro* disposition modeling to arrive at an adjusted phenotypic altering concentration (PAC) (mg/L). In this study, comparison of the adjusted PAC to fish acute lethality data resulted in a protective value in > 73% of cases.

Bibliography

Armitage JM, Arnot JA, Wania F, Mackay D. 2013. Development and evaluation of a mechanistic bioconcentration model for ionogenic organic chemicals in fish. Enviro Toxic and Chemistry. 32(1):115–128. https://doi.org/10.1002/etc.2020

Armitage JM, Sangion A, Parmar R, Looky AB, Arnot JA. 2021. Update and Evaluation of a High-Throughput In Vitro Mass Balance Distribution Model: IV-MBM EQP v2.0. Toxics. 9(11):315. https://doi.org/10.3390/toxics9110315

Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C, Kost-Alimova M, Gustafsdottir SM, Gibson CC, Carpenter AE. 2016. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 11(9):1757–1774. https://doi.org/10.1038/nprot.2016.105

Culbreth M, Nyffeler J, Willis C, Harrill JA. 2022. Optimization of Human Neural Progenitor Cells for an Imaging-Based High-Throughput Phenotypic Profiling Assay for Developmental Neurotoxicity Screening. Front Toxicol. 3:803987. https://doi.org/10.3389/ftox.2021.803987

Dahlin JL, Hua BK, Zucconi BE, Nelson SD, Singh S, Carpenter AE, Shrimp JH, Lima-Fernandes E, Wawer MJ, Chung LPW, et al. 2023. Reference compounds for characterizing cellular injury in high-content cellular morphology assays. Nat Commun. 14(1):1364. https://doi.org/10.1038/s41467-023-36829-x

Damiani E, Sella F, Astolfi P, Galeazzi R, Carnevali O, Maradonna F. 2023. First In Vivo Insights on the Effects of Tempol-Methoxycinnamate, a New UV Filter, as Alternative to Octyl Methoxycinnamate, on Zebrafish Early Development. International Journal of Molecular Sciences. 24(7):6767. https://doi.org/10.3390/ijms24076767

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci. 95(1):5–12. https://doi.org/10.1093/toxsci/kfl103

Escher BI, Glauch L, König M, Mayer P, Schlichting R. 2019. Baseline Toxicity and Volatility Cutoff in Reporter Gene Assays Used for High-Throughput Screening. Chem Res Toxicol. 32(8):1646–1655. https://doi.org/10.1021/acs.chemrestox.9b00182

European Commission. 2020. Report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union in 2015-2017. Commission staff working document. [Internet]. [place unknown]. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020SC0010

Feshuk M, Kolaczkowski L, Dunham K, Davidson-Fritz SE, Carstens KE, Brown J, Judson RS, Paul Friedman K. 2023. The ToxCast pipeline: updates to curve-fitting approaches and database structure. Front Toxicol. 5:1275980. https://doi.org/10.3389/ftox.2023.1275980

Fischer FC, Abele C, Droge STJ, Henneberger L, König M, Schlichting R, Scholz S, Escher BI. 2018. Cellular Uptake Kinetics of Neutral and Charged Chemicals in *in Vitro* Assays Measured by Fluorescence Microscopy. Chem Res Toxicol. 31(8):646–657. https://doi.org/10.1021/acs.chemrestox.8b00019

Fischer FC, Cirpka OA, Goss K-U, Henneberger L, Escher BI. 2018. Application of Experimental Polystyrene Partition Constants and Diffusion Coefficients to Predict the Sorption of Neutral Organic Chemicals to Multiwell Plates in in Vivo and in Vitro Bioassays. Environ Sci Technol. 52(22):13511–13522. https://doi.org/10.1021/acs.est.8b04246

Fischer FC, Henneberger L, Schlichting R, Escher BI. 2019. How To Improve the Dosing of Chemicals in High-Throughput *in Vitro* Mammalian Cell Assays. Chem Res Toxicol. 32(8):1462–1468. https://doi.org/10.1021/acs.chemrestox.9b00167

Fischer M, Belanger SE, Berckmans P, Bernhard MJ, Bláha L, Coman Schmid DE, Dyer SD, Haupt T, Hermens JLM, Hultman MT, et al. 2019. Repeatability and Reproducibility of the RTgill-W1 Cell Line Assay for Predicting Fish Acute Toxicity. Toxicol Sci. 169(2):353–364. https://doi.org/10.1093/toxsci/kfz057

Flynn K, Le M, Hazemi M, Biales A, Bencic DC, Blackwell BR, Bush K, Flick R, Hoang JX, Martinson J, et al. 2024. Comparing Transcriptomic Points of Departure to Apical Effect Concentrations For Larval Fathead Minnow Exposed to Chemicals with Four Different Modes Of Action. Arch Environ Contam Toxicol [Internet]. [accessed 2024 May 15]. https://doi.org/10.1007/s00244-024-01064-y

Galamgam J, Linou N, Linos E. 2018. Sunscreens, cancer, and protecting our planet. The Lancet Planetary Health. 2(11):e465–e466. https://doi.org/10.1016/S2542-5196(18)30224-9

Gobas FAPC, Zhang X. 1992. Measuring bioconcentration factors and rate constants of chemicals in aquatic organisms under conditions of variable water concentrations and short exposure time. Chemosphere. 25(12):1961–1971. https://doi.org/10.1016/0045-6535(92)90035-P

Gustafsdottir SM, Ljosa V, Sokolnicki KL, Anthony Wilson J, Walpita D, Kemp MM, Petri Seiler K, Carrel HA, Golub TR, Schreiber SL, et al. 2013. Multiplex cytological profiling assay to measure diverse cellular states. PLoS One. 8(12):e80999. https://doi.org/10.1371/journal.pone.0080999

Hecker FA, Leggio B, König T, Kim V, Osterland M, Gnutt D, Niehaus K, Geibel S. 2024. Cell Painting unravels insecticidal modes of action on Spodoptera frugiperda insect cells. Pesticide Biochemistry and Physiology. 203:105983. https://doi.org/10.1016/j.pestbp.2024.105983

Jencic V, Cerne M, Erzen NK, Kobal S, Cerkvenik-Flajs V. 2006. Abamectin effects on rainbow trout (Oncorhynchus mykiss). Ecotoxicology. 15(3):249–257. https://doi.org/10.1007/s10646-006-0056-6

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, Reif DM, Rotroff DM, Shah I, Richard AM, Dix DJ. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect. 118(4):485–492. https://doi.org/10.1289/ehp.0901392

Lerman MJ, Lembong J, Muramoto S, Gillen G, Fisher JP. 2018. The Evolution of Polystyrene as a Cell Culture Material. Tissue Engineering Part B: Reviews. 24(5):359–372. https://doi.org/10.1089/ten.teb.2018.0056

Manalo T, May A, Quinn J, Lafontant DS, Shifatu O, He W, Gonzalez-Rosa JM, Burns GC, Burns CE, Burns AR, Lafontant PJ. 2016. Differential Lectin Binding Patterns Identify Distinct Heart Regions in Giant Danio (*Devario aequipinnatus*) and Zebrafish (*Danio rerio*) Hearts. J Histochem Cytochem. 64(11):687–714. https://doi.org/10.1369/0022155416667928

Nataraj B, Maharajan K, Hemalatha D, Rangasamy B, Arul N, Ramesh M. 2020. Comparative toxicity of UV-filter Octyl methoxycinnamate and its photoproducts on zebrafish development. Sci Total Environ. 718:134546. https://doi.org/10.1016/j.scitotenv.2019.134546

Natsch A, Laue H, Haupt T, Von Niederhäusern V, Sanders G. 2018. Accurate prediction of acute fish toxicity of fragrance chemicals with the RTgill-W1 cell assay. Enviro Toxic and Chemistry. 37(3):931–941. https://doi.org/10.1002/etc.4027

Nelson L, Veling M, Farhangdoust F, Cai X, Huhn S, Soloveva V, Chang M. 2023. Transcriptomics and cell painting analysis reveals molecular and morphological features associated with fed-batch production performance in CHO recombinant clones. Biotech & Bioengineering. 120(11):3177–3190. https://doi.org/10.1002/bit.28518

Neuwoehner J, Fenner K, Escher BI. 2009. Physiological Modes of Action of Fluoxetine and its Human Metabolites in Algae. Environ Sci Technol. 43(17):6830–6837. https://doi.org/10.1021/es9005493

Nyffeler J, Haggard DE, Willis C, Setzer RW, Judson R, Paul-Friedman K, Everett LJ, Harrill JA. 2021. Comparison of Approaches for Determining Bioactivity Hits from High-Dimensional Profiling Data. SLAS Discov. 26(2):292–308. https://doi.org/10.1177/2472555220950245

Nyffeler J, Willis C, Harris FR, Foster MJ, Chambers B, Culbreth M, Brockway RE, Davidson-Fritz S, Dawson D, Shah I, et al. 2023. Application of Cell Painting for chemical hazard evaluation in support of screening-level chemical assessments. Toxicol Appl Pharmacol. 468:116513. https://doi.org/10.1016/j.taap.2023.116513

Nyffeler J, Willis C, Lougee R, Richard A, Paul-Friedman K, Harrill JA. 2020. Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling. Toxicol Appl Pharmacol. 389:114876. https://doi.org/10.1016/j.taap.2019.114876

OECD. 2019. Test No. 203: Fish, Acute Toxicity Test [Internet]. [place unknown]: OECD; [accessed 2023 Jul 21]. https://doi.org/10.1787/9789264069961-en

OECD. 2021. Test No. 249: Fish Cell Line Acute Toxicity - The RTgill-W1 cell line assay [Internet]. [place unknown]: OECD; [accessed 2023 Jul 21]. https://doi.org/10.1787/c66d5190-en

Olker JH, Elonen CM, Pilli A, Anderson A, Kinziger B, Erickson S, Skopinski M, Pomplun A, LaLone CA, Russom CL, Hoff D. 2022. The ECOTOXicology Knowledgebase: A Curated Database of Ecologically Relevant Toxicity Tests to Support Environmental Research and Risk Assessment. Enviro Toxic and Chemistry. 41(6):1520–1539. https://doi.org/10.1002/etc.5324

Pham LL, Watford SM, Pradeep P, Martin MT, Thomas RS, Judson RS, Setzer RW, Paul Friedman K. 2020. Variability in in vivo studies: Defining the upper limit of performance for predictions of systemic effect levels. Computational Toxicology. 15:100126. https://doi.org/10.1016/j.comtox.2020.100126

Proença S, Escher BI, Fischer FC, Fisher C, Grégoire S, Hewitt NJ, Nicol B, Paini A, Kramer NI. 2021. Effective exposure of chemicals in in vitro cell systems: A review of chemical distribution models. Toxicology in Vitro. 73:105133. https://doi.org/10.1016/j.tiv.2021.105133

Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, Yang C, Rathman J, Martin MT, Wambaugh JF, et al. 2016. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chem Res Toxicol. 29(8):1225–1251. https://doi.org/10.1021/acs.chemrestox.6b00135

Rodea-Palomares I, Bone AJ. 2024. Predictive value of the ToxCast/Tox21 high throughput toxicity screening data for approximating *in vivo* ecotoxicity endpoints and ecotoxicological risk in eco-surveillance applications. Science of The Total Environment. 914:169783. https://doi.org/10.1016/j.scitotenv.2023.169783

Schaupp CM, Maloney EM, Mattingly KZ, Olker JH, Villeneuve DL. 2023. Comparison of in silico, in vitro, and in vivo toxicity benchmarks suggests a role for ToxCast data in ecological hazard assessment. Toxicological Sciences. 195(2):145–154. https://doi.org/10.1093/toxsci/kfad072

Schirmer K, Chan AGJ, Greenberg BM, Dixon DG, Bols NC. 1997. Methodology for demonstrating and measuring the photocytotoxicity of fluoranthene to fish cells in culture. Toxicology in Vitro. 11(1–2):107–119. https://doi.org/10.1016/S0887-2333(97)00002-7

Schirmer K, Dixon DG, Greenberg BM, Bols NC. 1998. Ability of 16 priority PAHs to be directly cytotoxic to a cell line from the rainbow trout gill. Toxicology. 127(1–3):129–141. https://doi.org/10.1016/s0300-483x(98)00030-4

Seidensticker S, Grathwohl P, Lamprecht J, Zarfl C. 2018. A combined experimental and modeling study to evaluate pH-dependent sorption of polar and non-polar compounds to polyethylene and polystyrene microplastics. Environ Sci Eur. 30(1):30. https://doi.org/10.1186/s12302-018-0155-z

Sheffield T, Brown J, Davidson S, Friedman KP, Judson R. 2022. tcplfit2: an R-language general purpose concentration–response modeling package.Mathelier A, editor. Bioinformatics. 38(4):1157–1158. https://doi.org/10.1093/bioinformatics/btab779

Stadnicka-Michalak J, Bramaz N, Schönenberger R, Schirmer K. 2021. Predicting exposure concentrations of chemicals with a wide range of volatility and hydrophobicity in different multi-well plate set-ups. Sci Rep. 11(1):4680. https://doi.org/10.1038/s41598-021-84109-9

Tanneberger K, Knöbel M, Busser FJM, Sinnige TL, Hermens JLM, Schirmer K. 2013. Predicting fish acute toxicity using a fish gill cell line-based toxicity assay. Environ Sci Technol. 47(2):1110–1119. https://doi.org/10.1021/es303505z

Villeneuve DL, Le M, Hazemi M, Biales A, Bencic DC, Bush K, Flick R, Martinson J, Morshead M, Rodriguez KS, et al. 2023. Pilot testing and optimization of a larval fathead minnow high throughput transcriptomics assay. Current Research in Toxicology. 4:100099. https://doi.org/10.1016/j.crtox.2022.100099

Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. 2019. Welcome to the Tidyverse. JOSS. 4(43):1686. https://doi.org/10.21105/joss.01686

Willis C, Nyffeler J, Harrill J. 2020. Phenotypic Profiling of Reference Chemicals across Biologically Diverse Cell Types Using the Cell Painting Assay. SLAS Discov. 25(7):755–769. https://doi.org/10.1177/2472555220928004

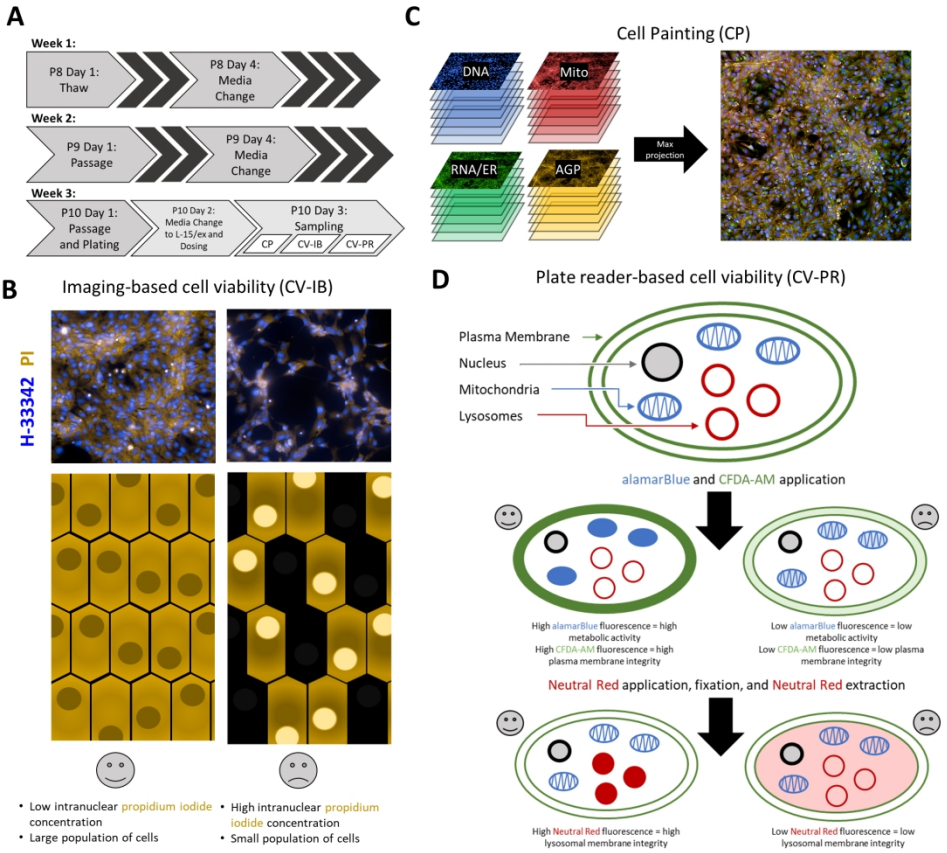


Figure 1: Graphical overview of experimental schedule and assay concepts.

190x165mm (300 x 300 DPI)

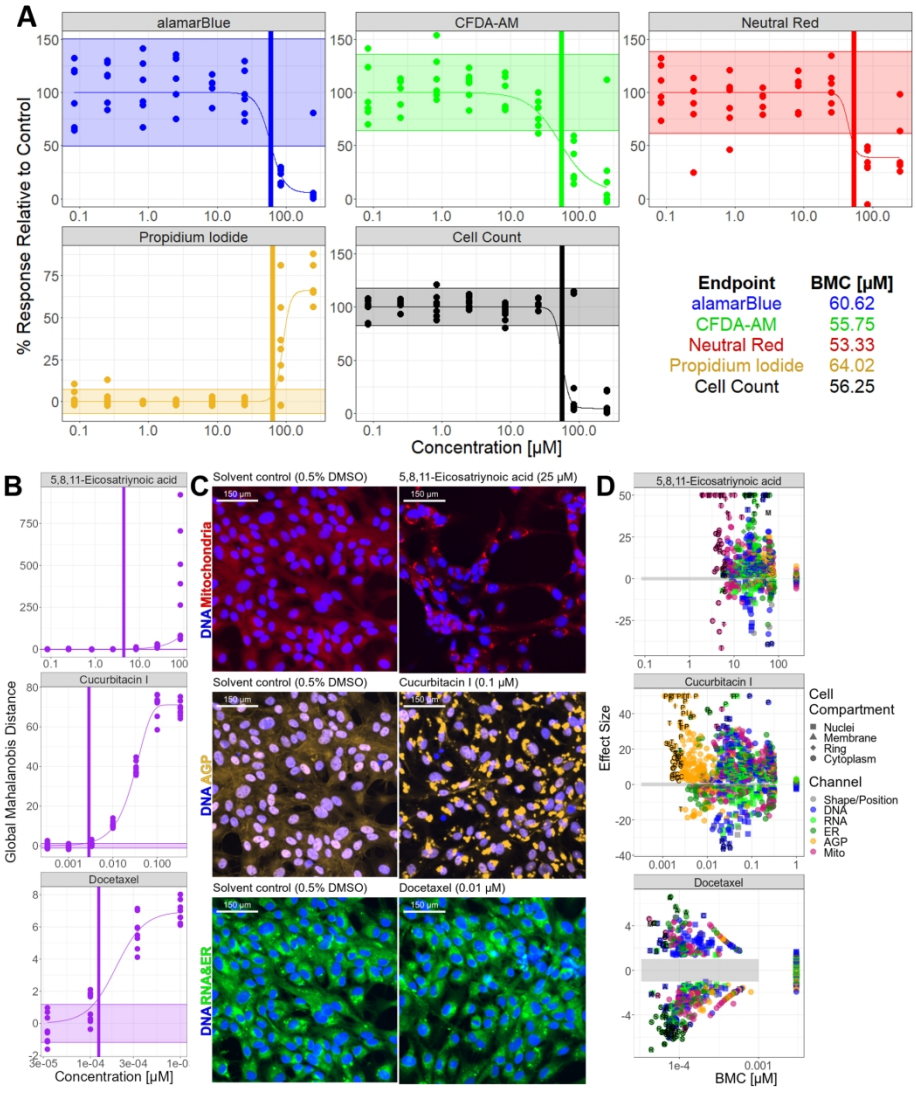


Figure 2: Reference chemicals for cell viability and Cell Painting in RTgill-W1

190x218mm (300 x 300 DPI)

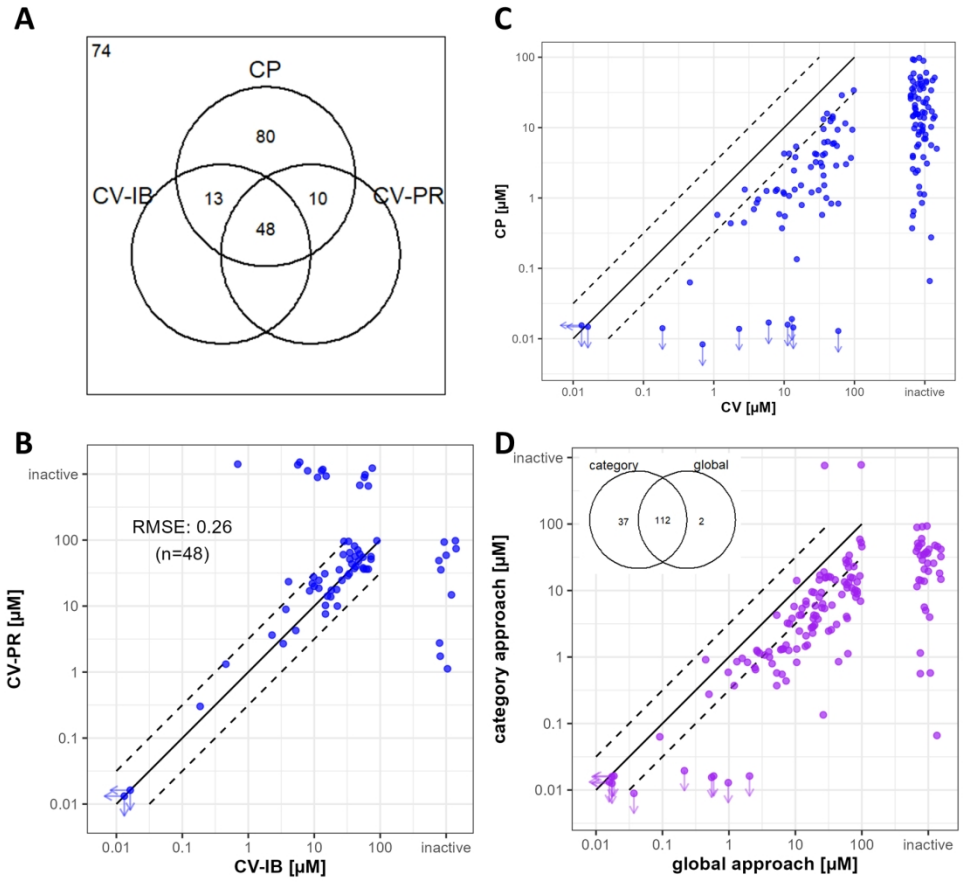


Figure 3: Results from the 225 unique chemicals tested in the CV-PR, CV-IB and CP assays

165x152mm (300 x 300 DPI)

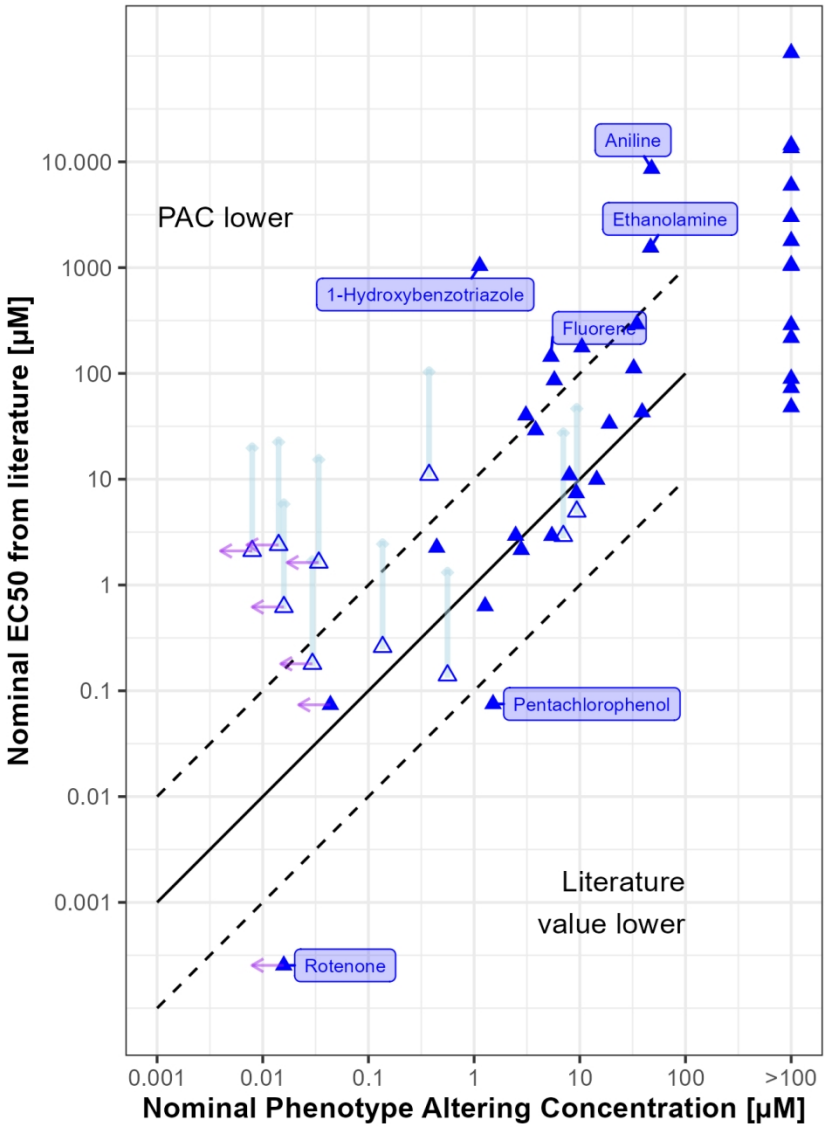
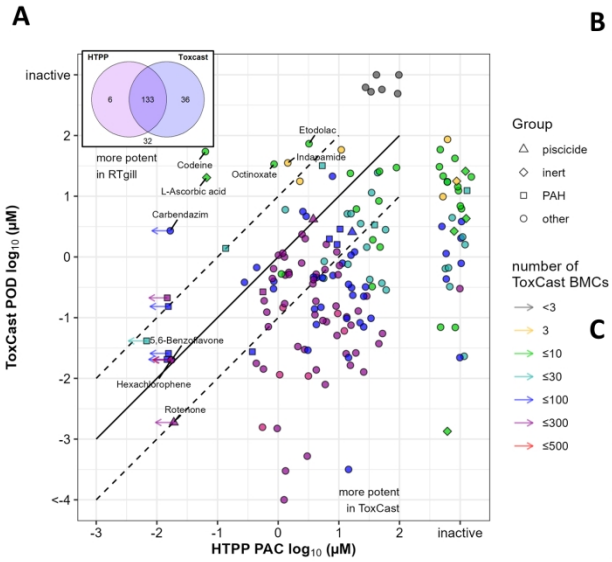


Figure 4: Comparison of RTgill-W1 PACs to existing RTgill-W1 EC50 values from the literature

120x159mm (300 x 300 DPI)



B

Largest difference	HTPP	ToxCast		
Chemical Name	PAC	# assays	# BMCs	POD
Codeine	-1.20	237	5	1.74
L-Ascorbic acid	-1.18	393	10	1.31
Carbendazim	-1.78	619	63	0.43
Octinoxate	-0.07	341	5	1.53
Indapamide	0.16	237	3	1.55
Nizatidine	1.62	319	0	-
Etodolac	0.51	234	4	1.87
Carisoprodol	1.44	237	2	-

C

Most potent	HTPP	ToxCast		
Chemical Name	PAC	# assays	# BMCs	POD
Indeno[1,2,3-cd]pyrene*	-2.17	311	30	-1.38
Benzo[k]fluoranthene*	-1.83	267	56	-1.69
Benzo(b)fluoranthene*	-1.83	624	162	-0.67
Benzo[a]pyrene*	-1.81	257	68	-0.81
Dibenz[a,h]anthracene*	-1.81	498	35	-1.60
Carbendazim	-1.78	619	63	0.43
Hexachlorophene	-1.77	546	311	-1.70
5,6-Benzoflavone	-1.75	334	66	-1.70
Rotenone	-1.72	768	281	-2.73

Figure 5: Comparison of RTgill-W1 Cell Painting screening results to the ToxCast POD

190x128mm (300 x 300 DPI)

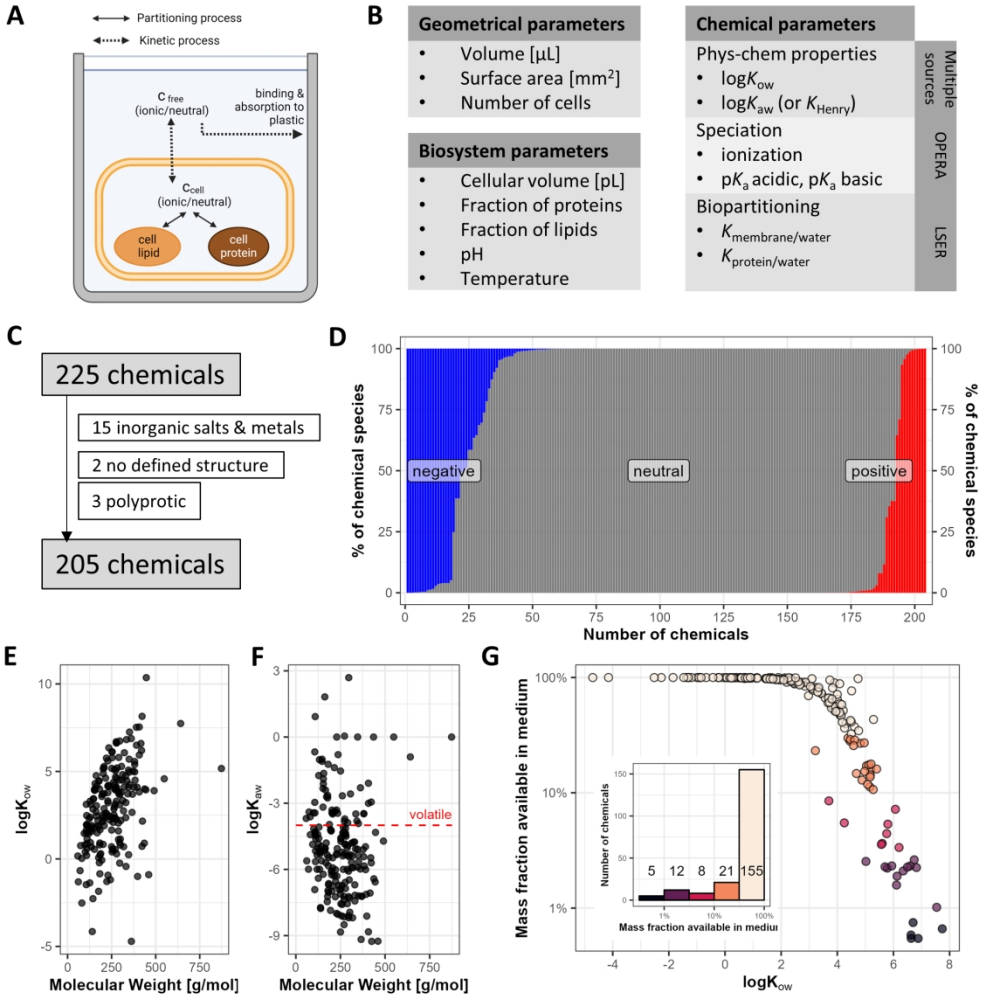
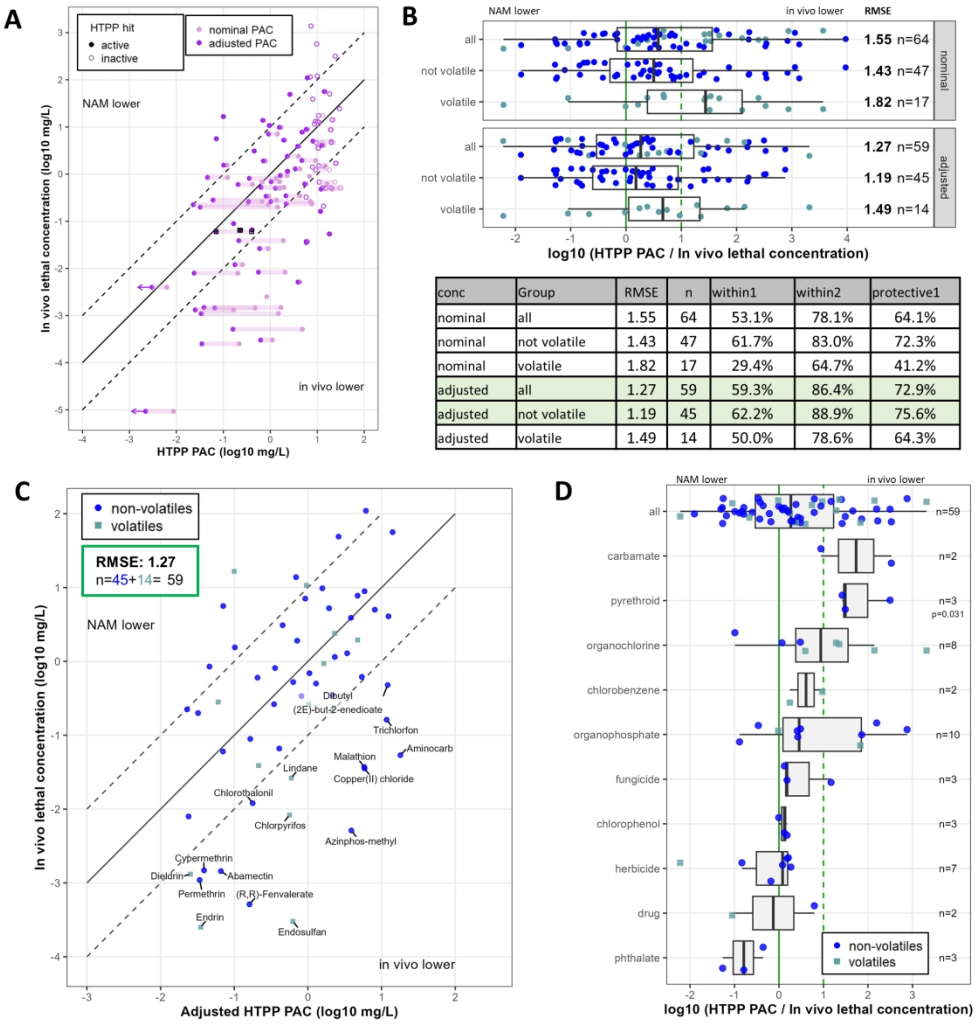


Figure 6: A new in vitro disposition model accounting for sorption to plastic

190x196mm (300 x 300 DPI)



conc	Group	RMSE	n	within1	within2	protective1
nominal	all	1.55	64	53.1%	78.1%	64.1%
nominal	not volatile	1.43	47	61.7%	83.0%	72.3%
nominal	volatile	1.82	17	29.4%	64.7%	41.2%
adjusted	all	1.27	59	59.3%	86.4%	72.9%
adjusted	not volatile	1.19	45	62.2%	88.9%	75.6%
adjusted	volatile	1.49	14	50.0%	78.6%	64.3%

Figure 7: Comparison to existing in vivo lethality data from rainbow trout and fathead minnow

190x198mm (300 x 300 DPI)

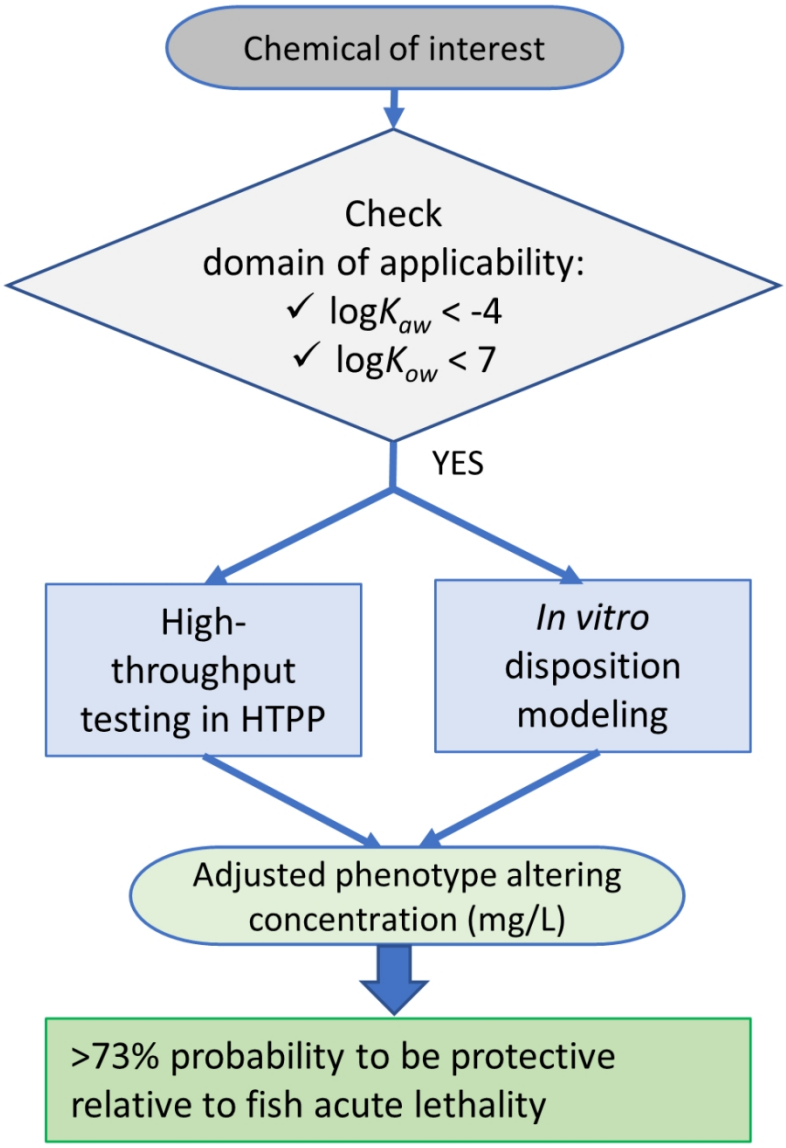


Figure 8. Schema for use of in vitro and in silico NAMs for evaluating fish toxicity.

85x125mm (300 x 300 DPI)