

Supplementary manuscript relating to the project comprising:

Main article: Cross-species bioanalytical assessment of the oxidative stress and xenobiotic metabolism modes of action for environmental water samples – a multi-layered approach to testing and modelling

Sebastian Lungu-Mitea (a,b,c,d), Geeta Mandava (a), Jana Horackova (f,g), Oksana Golovko (e), Zuzana Tousova (b), David Bednar (f,g), Lutz Ahrens (e), Klara Hilscherova (b), Johan Lundqvist (a)

Data in Brief: A dataset on the multi-species and multi-endpoint bioanalytical assessment of environmental samples derived from Swedish wastewater treatment plants

Sebastian Lungu-Mitea (a,b,c,d), Geeta Mandava (a), Oksana Golovko (e), Kim Frieberg (a), Sara Alsayfi (a), Stefan Örn (a), Lutz Ahrens (e), Johan Lundqvist (a)

Data in Brief: Human and zebrafish cross-species comparison on ligand affinity towards the aryl hydrocarbon receptor and Keap1 oxidative stress sensor: a molecular docking dataset

Sebastian Lungu-Mitea (a,b,c,d), Jana Horackova (f,g), David Bednar (f,g), Klara Hilscherova (b)

Affiliations

(a) Department of Animal Biosciences, Molecular Toxicology, Swedish University of Agricultural Sciences (SLU), SE-75007 Uppsala, Sweden

(b) RECETOX, Faculty of Science, Masaryk University (MUNI), CZ-62500 Brno, Czech Republic

(c) Department of Pharmaceutical Biosciences, Drug Safety and Toxicology, Uppsala University (UU), SE-75124 Uppsala, Sweden

(d) Department of Biology and Environmental Science, Linnaeus University (LNU), SE-39182 Kalmar, Sweden

(e) Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences (SLU), SE-75007 Uppsala, Sweden

(f) Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Masaryk University (MUNI), CZ-62500 Brno, Czech Republic

(g) International Clinical Research Centre, St. Anne's University Hospital Brno, CZ-60200 Brno, Czech Republic

*Corresponding author: sebastian.lungu@slu.se

ORCID IDs

DB: <https://orcid.org/0000-0002-6803-0340>

GM: <https://orcid.org/0000-0003-2834-2259>

JH: <https://orcid.org/0000-0001-7119-121X>

JL: <https://orcid.org/0000-0001-5693-9007>

KF: <https://orcid.org/0000-0002-0043-4031>

KH: <https://orcid.org/0000-0001-6320-8093>

LA: <https://orcid.org/0000-0002-5430-6764>

OG: <https://orcid.org/0000-0003-3256-3054>

36 SA: -
 37 SLM: <https://orcid.org/0000-0001-8192-9134>
 38 SÖ: <https://orcid.org/0000-0002-0086-2124>
 39 ZT: <https://orcid.org/0000-0001-7069-0471>

40 INDEX

41	1. Concept, summary list, and abbreviations.....	5
42	Tab. S1: Description of supplementary information (SI) in the data repository.....	6
43	Tab. S2: List of abbreviations.....	7
44	2. Supplementary material and methods.....	9
45	2.1 Environmental sample preparation and extraction.....	9
46	Fig. S1: A map indicating WWTP sites' A to C approximate locations in central and western	
47	Sweden.....	10
48	Tab. S3: Environmental sample and WWTP-related characteristics in detail.....	10
49	2.2 Instrumental chemical assessment.....	12
50	2.3 Bioanalytical assessment: reporter gene bioassays.....	12
51	2.3.1 Oxidative stress response reporter assay in zebrafish cell line.....	13
52	2.3.2 Oxidative stress reporter assay in human cell line.....	13
53	2.3.3 Xenobiotic metabolism response reporter assay in zebrafish cell line.....	14
54	2.3.4 Xenobiotic metabolism response reporter assay in human cell line.....	15
55	2.3.5 Xenobiotic metabolism response reporter assay in mouse cell line.....	16
56	2.3.6 Cell handling and plating.....	16
57	2.3.7 Transient transfection.....	16
58	2.3.8 Exposure.....	17
59	2.3.9 Viability and reporter-related readouts.....	18
60	2.3.10 MTS-based cellular viability assay.....	18
61	2.3.11 Luciferase-based reporter readouts.....	18
62	2.4 Bioanalytical assessment: the FET assay.....	19
63	2.5 In silico-mediated effect-directed analysis (ISEDA).....	19
64	2.5.1 tcpl and MySQL R coding packages.....	19
65	2.5.2 OCHEM AhR-NR QSAR.....	20
66	2.6 Molecular docking: assessing receptor-ligand affinity.....	20
67	2.6.1 Ligands.....	20
68	Fig. S2: 2D representation of all selected ligands.....	21
69	2.6.2 Tertiary protein structure selection and prediction: Keap1 and AhR PAS-b domain.....	21
70	2.6.3 Amino acid sequence alignment.....	22

71	2.6.4 Topology, mapping, and nomenclature.....	22
72	Fig. S3: Topology of hKeap1.....	23
73	Fig. S4: Topology of hAhR.....	24
74	2.6.5 Molecular docking procedure.....	24
75	2.7 Quantification of bioavailable concentrations via mass-balance modelling.....	25
76	2.8 Data processing and statistical analyses.....	25
77	2.8.1 Reporter gene assays – raw data handling and normalisation.....	25
78	2.8.2 Reporter gene assays – regression analyses.....	26
79	2.8.3 Iceberg modelling.....	26
80	2.8.4 Removal efficiencies.....	27
81	2.8.5 Statistical analysis of molecular docking-derived data.....	27
82	2.8.6 Ratios.....	28
83	3. Supplementary results.....	28
84	3.1 Supplementary results: DIB_B co-submission.....	28
85	Fig. S5: Molecular docking (tBHQ).....	28
86	Fig. S6: Molecular docking (tBHQ).....	29
87	Fig. S7: Molecular docking (tBHQ).....	30
88	Fig. S8: Molecular docking (tBHQ).....	30
89	Fig. S9: Molecular docking (tBHQ).....	31
90	Fig. S10: Molecular docking (tBHQ).....	31
91	Fig. S11: Molecular docking (tBHQ).....	32
92	Fig. S12: Molecular docking (tBHQ).....	32
93	Fig. S13: Molecular docking (MZC).....	33
94	Fig. S14: Molecular docking (MZC).....	33
95	Fig. S15: Molecular docking (MZC).....	34
96	Fig. S16: Molecular docking (MZC).....	34
97	Fig. S17: Molecular docking (MZC).....	35
98	Fig. S18: Molecular docking (MZC).....	35
99	Fig. S19: Molecular docking (MZC).....	36
100	Fig. S20: Molecular docking (MZC).....	36
101	Fig. S21: Molecular docking (MZC).....	37
102	Fig. S22: Molecular docking (MZC).....	38
103	Fig. S23: Molecular docking (MZC).....	38
104	Fig. S24: Molecular docking (CLB).....	39
105	Fig. S25: Molecular docking (DAI).....	40
106	Fig. S26: Molecular docking (TBDZ).....	41

107	Fig. S27: Molecular docking (TCDD).....	42
108	Fig. S28: Molecular docking summary.....	43
109	3.2 Supplementary results: Main article – Summarised cytotoxicity/cellular viability.....	44
110	Fig. S29: Cytotoxicity/cellular viability impact assessment of the environmental samples from	
111	WWTP sites A to C.....	44
112	Fig. S30: Cytotoxicity/cellular viability impact assessment of reference control and ISEDA-	
113	identified compounds.....	45
114	3.3 Supplementary results: Main article – Optimised data fitting strategies for the oxidative stress	
115	MoA.....	46
116	Fig. S31: Alternative data fitting (tBHQ, MZC).....	48
117	Fig. S32: Alternative data fitting (tBHQ, MZC).....	50
118	Fig. S33: Alternative data fitting (tBHQ, MZC).....	51
119	3.4 Supplementary results: Main article – Optimised data fitting strategies for the xenobiotic	
120	metabolism MoA.....	52
121	Fig. S34: Alternative data fitting (TCDD).....	53
122	Fig. S35: Alternative data fitting (TCDD).....	54
123	Fig. S36: Alternative data fitting (TCDD).....	55
124	Tab. S4: Effect estimates (EC10 or EC20) ratios, as derived from various DRE data model fits..	56
125	Fig. S37: Alternative data fitting (TCDD).....	57
126	Fig. S38: Alternative data fitting (TCDD).....	59
127	Fig. S39: Alternative data fitting (TCDD).....	60
128	3.5 Supplementary results: Main article – other supplementary results.....	61
129	Tab. S5: Summary of all effect point estimates ($EC_{IR1.5}$), their variation metrics, calculated BEQ	
130	values, and chemical removal efficiencies, as derived from the oxidative stress MoA	
131	assessment.....	61
132	Tab. S6: Summary of all effect point estimates (EC10), calculated BEQ values, and chemical	
133	removal efficiencies, as derived from the xenobiotic metabolism MoA assessment.....	61
134	Tab. S7: EC50 estimates computed herein vs. reported EC50 estimates within the literature .	62
135	Tab. S8: Percentage [%] of BEQ_Bio explained by BEQ_Chem.....	63
136	Tab. S9: Summary of all tested ISEDA compounds' concentration estimates (AC50 and EC10),	
137	including their variation metrics.....	63
138	Fig. S40: Alternative data fitting (ISEDA).....	65
139	3.6 Supplementary results: Main article – Bioavailability assessment.....	66
140	Fig. S41: Outline of the bioavailability assessment.....	67
141	4. Supplementary discussion.....	67
142	4.1 The Nrf2-Keap1-ARE toxicity pathway – Canonical reactive cysteine residues and mechanistic	
143	insights on the “cysteine code”	67
144	4.1.1 Background and objectives of this project concerning Keap1.....	67

145	Fig. S42: Amount of total conserved cysteines within all Keap1 variants.....	69
146	4.1.2 The Nrf2-Keap1-ARE signalling pathway: nomenclature and mechanisms.....	69
147	4.1.3 Molecular docking results in the context of the “cysteine code” (see DIB_B for results)....	71
148	4.2 Supplementary discussion: Comment on zfKeap1a/b functional divergence and structural	
149	variability.....	72
150	4.3 Supplementary discussion: “CALUX” assay nomenclature and identification.....	72
151	4.4 Supplementary discussion: Cross-species assessment in context with regulatory metrics – BEQs	
152	and EBTs.....	72
153	Fig. S43: Hypothetical TCDD-BEQ scenario.....	74
154	4.5 Supplementary discussion: TCDD as a reference compound, reasoning and scrutiny.....	74
155	4.6 Supplementary discussion: validity of molecular docking approach.....	76
156	4.7 Supplementary discussion: AhR crosstalk and non-canonical pathways.....	77
157	4.8 Supplementary discussion: outlook and further technicalities regarding multi-species reporter	
158	gene assays.....	77
159	Fig. S44: Outlook on short to long-term challenges.....	80
160	5. References.....	80
161		

162 1. Concept, summary list, and abbreviations

163 The main article is supported by two *Data in Brief* co-submissions. As such, the co-submissions are
164 self-reliant, but they supplement the primary article by providing raw data, specific methodologies,
165 and first-tier analyses. Higher-tier analyses are exclusive to the main article. Readers are directed to
166 these co-submissions for an in-depth review of datasets to avoid overburdening the main article. This
167 supplementary manuscript at hand serves as a link between the main article and the co-submissions,
168 describing the shared methodology in detail, as recommended by the *Data in Brief* guidelines.
169 Moreover, the supplementary manuscript provides additional results, analyses, and discussions.

170 The first *Data in Brief* co-submission (hereafter, “DIB_A”; for abbreviations, see tab. S2) encompasses
171 raw data and preliminary analysis (LOEC/NOEC) pertinent to the bioanalytical assessment of the
172 investigated environmental samples. This includes additional samples beyond the scope of the main
173 article and employs further test frameworks, specifically the fish embryo acute toxicity test (FET,
174 OECD TG 236 [1]). The data in DIB_A are related to the SI files SI1-SI8.

175 The second *Data in Brief* co-submission (hereafter, “DIB_B”) incorporates comprehensive molecular
176 docking models along with protein sequence alignment studies. It also features confirmatory
177 analyses that compare AlphaFold-predicted structures with those derived from cryo-EM and X-ray

178 crystallography. DIB_B addresses the consequential implications and is associated with SI files SI9
179 through SI11.

180 All raw and computed metadata across the manuscripts are consolidated in a shared data repository
181 (<https://doi.org/10.5878/zwnq-qb88>), which contains 16 Supplementary Information files (hereafter,
182 “SI”) and this supplementary manuscript (hereafter, “SM”). These files are numbered chronologically,
183 starting with SI12, which contains data explicitly related to the main article (tab. S1).

Tab. S1: Description of supplementary information (SI) in the data repository. (https://doi.org/10.5878/zwnq-qb88)		
Name	Content	Associated Article
SI1	Raw and analysed data of all experiments conducted with the ZF4 assay	DIB_A
SI2	Raw and analysed data of all experiments conducted with the MCF7/AREc32 assay	DIB_A
SI3	Raw and analysed data of all experiments conducted with the ZFL assay	DIB_A
SI4	Raw and analysed data of all experiments conducted with the HepG2 assay	DIB_A
SI5	Raw and analysed data of all experiments conducted with the DREcoScreen assay	DIB_A
SI6	Raw, analysed data, and statistical interference of the mFET assay	DIB_A
SI7	Transformation and statistical interference conducted in all cellular bioassays	DIB_A
SI8	Transformation and statistical interference conducted for all pairwise comparisons between assays/species	DIB_A
SI9	Summarised binding energies	DIB_B
SI10	PDBQT and FASTA-type files of ligands and receptors	DIB_B
SI11	TCDD-energy-equivalent calculations	DIB_B
SI12	Physico-chemical properties of all targeted compounds detected in chemical analysis; measured chemical concentration of targeted compounds per environmental sample; calculated REP values, extracted ToxCast data; OCHEM AhR QSAR outputs; bioanalytics for AhR-QSAR positive compounds; BEQ-Chem calculations.	Main article SM
SI13	Bioavailability (C _{free} concentrations) data generated via linear solvation energy relationships QSPRs from UFZ-LSER	Main article SM
SI14	Metazachlor reference compound standard curves, cytotoxicity data, and all respectively raw data generated	Main article SM
SI15	Raw and metadata from the bioanalytical assessment of ISEDA-identified chemical compounds	Main article SM
SI16	BEQ-Bio calculations relating to estimates derived via 3PNL regression analysis (main article) and all other applied data-fitting scenarios (supplementary manuscript)	Main article SM

184

Tab. S2: List of abbreviations	
2PNL	Two-parametric non-linear regression
30%SL	Refitted simple linear regression using the first 30% curve estimates of the 3PNL model
3PNL	Three-parametric non-linear regression
4PNL	Four-parametric non-linear regression
A/B/C_X	Coding of environmental samples
AA	Amino acid
AC50	Concentration at which a compound produces 50% of its maximal observed activity in a specific assay system
AhR	Aryl hydrocarbon receptor
ARE	Antioxidant response element
ARNT	Aryl hydrocarbon receptor nuclear translocator protein
BEQ	Bioanalytical equivalents
BTB	Broad complex, tramtrack, and bric-a-brac domain
CI	Confidence interval
CLB	Climbazole
CRC	Concentration-response curve
cryo-EM	Cryogenic electron microscopy
Ctrl	Solvent control
CysX	Cysteine residue no. X
DAI	Daidzein
DIB_A	Data in Brief co-submission A
DIB_B	Data in Brief co-submission B
DLC	Dioxin-like compounds
DLR	Dual luciferase assay
DMSO	Dimethyl sulfoxide
EBT	Effect-based trigger value
EC10	Effective concentration of a sample causing 10% of the effect in relation to the reference compound
EC50	Effective concentration of a sample causing 50% of the effect in relation to the reference compound

ECIR1.5	Effect concentration causing a 1.5x induction ratio (vs. Control)
EROD	Ethoxyresorufin-O-deethylase activity (CYP1A activity)
FBS	Fetal bovine serum
FET	Fish embryo acute toxicity test
Fig. Sx	Supplementary figure X
Fluc	Firefly luciferase
GSR	Glutathione reductase activity
hpf	Hours post-fertilisation
ISEDA	In silico-mediated effect-directed analysis
IVR	Intervening region domain
Keap1	Kelch-like ECH-associated protein 1; human (h) or zebrafish (zf) variant
Kelch domain	Five to seven kelch tandem repeats that form a β -propeller tertiary structure
LOEC	Lowest observed effect concentration
LR	(single) luciferase assay
mFET	Modified zebrafish embryo test
MoA	Mode of action
MTS	Dimethyl-thiazol-sulfophenyl tetrazolium-based cytotoxicity assay
MZC	Metazachlor
NC	Negative control (solvent control)
NOEC	No observed effect concentration
Nrf2	Nuclear factor erythroid 2-related factor 2
PAS	Per-Arnt-Sim domain
PDB ID 1ZGK, 7EXI, 7ZUB	Protein database identifier of crystal or cryo-EM structures
PPAR γ	Peroxisome Proliferator-Activated Receptor gamma
PXR	Pregnane X receptor
REF	Relative enrichment factor
Rluc	Renilla luciferase
SEM	Standard error of mean

SI	Supplementary information
SL	simple linear regression
SM	Supplementary manuscript
TBDZ	Thiabendazole
tBHQ	tert-Butylhydroquinone
TCDD	2,3,7,8-tetrachlorodibenzodioxin
WWTP	Wastewater treatment plant
XRE	Xenobiotic response element, also known as DRE - dioxin response element

185

186 2. Supplementary material and methods

187 2. 1 Environmental sample preparation and extraction

188 Water samples were collected from three wastewater treatment plants (WWTPs) in central and
189 western Sweden, with downstream catchments identified as areas of anthropogenic (drinking water
190 production, leisure, and food consumption) and ecological (water protection areas and Natura 2000
191 sites) interest (fig. S1). Additionally, the three WWTP sites were selected based on a prior commercial
192 bioanalytical screening campaign (unpublished data due to confidentiality), in which these samples
193 showed comparatively strong induction of Nrf2/Keap1/ARE- and AhR/ARNT/XRE-mediated
194 responses. Site selection was therefore intentionally targeted toward bioactive wastewater samples
195 to enable potential quantitative cross-species assay comparison across both investigated MoAs,
196 rather than to provide a generally representative survey of Swedish WWTPs.

197



Fig. S1: A map indicating WWTP sites' A to C approximate locations in central and western Sweden, including ecological and anthropogenic concern protection areas.

Influent and effluent samples were collected as one- to seven-day composite samples during the summer (June). Surface water samples were collected as grab samples 10 cm below the water surface, contemporaneous with the WWTP samples. The latter samples were collected into 1 L high-density polyethylene containers, snap-frozen, and stored until usage or analysis at -20°C. Tab. S3 gives further details on the respective environmental samples. The precise geographic locations are disclosed due to legal confidentiality.

Environmental sample extraction was performed using an automatic solid-phase extraction (SPE) system (SPE-DEX, Horizon Technology, Salem, NH, USA) with HLB extraction disks (Atlantic HLB-H Disks, diameter 47 mm; Horizon Technology, Salem, NH, USA). HLB extraction disks were conditioned with 280 mL of MeOH and 420 mL of Millipore water. After conditioning, approximately 1 L of the environmental samples was loaded onto the disks at a flow rate of 50 mL/min. Loaded HLB disks were washed twice with 24 mL of 5% (v/v) MeOH/95% Millipore water, dried under vacuum for 30 min, and eluted three times with 25 mL of MeOH, for a total volume of 75 mL. Furthermore, eluates were evaporated under a gentle nitrogen stream at 35 °C. The residue was dissolved in 0.5 mL of anhydrous dimethyl sulfoxide (DMSO; ≥99.9% purity, sterile; Sigma-Aldrich, Steinheim, Germany), yielding a 2000-fold enrichment factor. The same volume of Millipore water was concentrated by solid-phase extraction to serve as an operational control.

The overall loading volume for the environmental samples was 1 L. However, for sample IDs A_I and B_I, only 900 and 980 mL could be recovered, respectively, resulting in slightly changed relative enrichment factors (REFs) in the later bioassays (tab. S3). Enrichment and dilution factors were calculated according to Escher et al. [2].

Relative enrichment factor (REF) = $enrichment\ factor_{SPE} * dilution\ factor_{bioassay}$

Enrichment factor_{SPE} = $V_{water}/V_{extract}$

Dilution factor_{bioassay} = $V_{extract\ in\ bioassay}/V_{total\ in\ bioassay}$

A REF >1 implies that the water sample is concentrated, while a REF <1 implies that the water sample is diluted compared to the original environmental sample.

Tab. S3: Environmental sample and WWTP-related characteristics in detail. Samples were collected in June 2018.

ID	WWTP site	Sample type	Catchment	Sampling method	Waste source	Person equivalent	Inflow/outflow [m ³ /day]	Treatment	SPE enrichment factor
A_RU	A	Recipient upstream	Mjörn (water protection area, Natura 2000 site)	Grab	90% households, 10 % industries	30000	-	-	2000
A_I	A	Influent	Mjörn (water protection area, Natura 2000 site)	5-day composite	90% households, 10 % industries	30000	400	-	1800
A_E	A	Effluent	Mjörn (water protection area, Natura 2000 site)	5-day composite	90% households, 10 % industries	30000	400	Roughing, 1 st sand filter, 1 st clarifier, aeration/bioreactor, 2 nd clarifier, digester, reject water influent to WWTP	2000
A_RD	A	Recipient downstream	Mjörn (water protection area, Natura 2000 site)	Grab	90% households, 10 % industries	30000	-	-	2000
B_RU	B	Recipient upstream	Mälaren (water protection area, Natura 2000 site)	Grab	92% households, 8 % industries	111105	-	-	2000
B_I	B	Influent	Mälaren (water protection area, Natura 2000 site)	7-day composite	92% households, 8 % industries	111105	39888	-	1960
B_E	B	Effluent	Mälaren (water protection area, Natura 2000 site)	7-day composite	92% households, 8 % industries	111105	39888	Roughing, 1 st sand filter, 1 st clarifier, aeration/bioreactor, 2 nd clarifier	2000

B_RD	B	Recipient downstream	Mälaren (water protection area, Natura 2000 site)	Grab	92% households, 8 % industries	111105	-	-	2000
C_RU	C	Recipient upstream	Ekoln (water protection area, Natura 2000 site)	Grab	Not defined	160000	-	-	2000
C_I	C	Influent	Ekoln (water protection area, Natura 2000 site)	1-day composite	Not defined	160000	37100	-	2000
C_E	C	Effluent	Ekoln (water protection area, Natura 2000 site)	1-day composite	Not defined	160000	37100	Roughing, 1 st sand filter, 1 st clarifier, aeration/bioreactor, 2 nd clarifier, iron chloride treatment, eventually chlorine dosing	2000
C_RD	C	Recipient downstream	Ekoln (water protection area, Natura 2000 site)	Grab	Not defined	160000	-	-	2000

226

227 2.2 Instrumental chemical assessment

228 The environmental samples were analysed on a DIONEX UltiMate 3000 ultra-high pressure liquid
229 chromatography (UPLC) system (Thermo Scientific, Waltham, MA, USA) coupled to a triple
230 quadrupole mass spectrometer (MS/MS) (TSQ QUANTIVA, Thermo Scientific, Waltham, MA, USA).
231 For method performance on linearity, limits of quantification, precision, blanks, and additional
232 physicochemical properties of the targeted compounds, please consult Golovko et al. [3] and SI12.

233 Notably, technical replicates of the same environmental samples, amongst others, were spiked with
234 isotopically labelled standards and target-analysed for 165 chemicals of emerging concern. The
235 results have been previously published by Golovko et al. [3] and are described in detail there.

2.3 Bioanalytical assessment: reporter gene bioassays

From the three investigated WWTPs (A to C), influent and effluent samples were analysed by five reporter gene assays in cell cultures derived from three different species (zebrafish, human, and mouse). These assays cover two modes of action (MoA; activation of oxidative stress defence via Nrf2-Keap1-ARE interaction (see, e.g., [4]) and activation of the xenobiotic metabolism via AhR-ARNT-XRE interaction (consult, e.g., [5])). For the interested reader, section 4.1 of the SM (below) offers a comprehensive review of the Nrf2-Keap1-ARE-mediated cellular stress response in the context of the molecular docking results (see sections 2.6 and 3.1).

2.3.1 Oxidative stress response reporter assay in zebrafish cell line

The oxidative stress response assay is based on the embryonic zebrafish fibroblast cell line ZF4 (CVCL_3275) [6]), which is transiently transfected with the pGL4.37[luc2P/ARE/Hygro] (Promega, Madison, USA) and pRL-CMV plasmid vectors. The basic ZF4 cell line was purchased from ATCC (Manassas, USA). The pGL4.37[luc2P/ARE/Hygro] plasmid vector bears four copies of a consensus Nrf2-responsive antioxidant response element (ARE) driving transcription of the primary Firefly (*Photinus pyralis*) luciferase reporter gene luc2P. The pRL-CMV plasmid vector drives the constitutive expression of the secondary Renilla luciferase reporter gene Rluc (*Renilla reniformis*), under the control of a cytomegalovirus (CMV) promoter. The secondary RLuc reporter normalises transfection efficiency within the transient setup (DLR, “dual luciferase assay”). Detailed descriptions of the assay's development and format are given in previous publications [7], [8].

Transient transfection was carried out 24h after initial cell plating on 96-well plates after the cells reached a confluency of about 80% (see further details in the sections 2.3.6 and 2.3.7 below) at a 2 µg transfection reagent to 1 µg plasmid DNA ratio per reaction using FugeneHD (Promega, Madison, USA). The FugeneHD transfection reagent increases cellular competence via lipid fusion. Transfection preconditioning was performed in OptiMEM medium (Gibco, Paisley, UK) according to the manufacturer's instructions. Primary and secondary reporter plasmids were co-transfected in a 10:1 ratio (0.9 µg reporter plasmid, 0.1 µg control plasmid). Transfection optimisation experiments were conducted according to the manufacturer's guidelines and have been reported previously [7], [8].

ZF4 cells were cultured in Dulbecco's Modified Eagle's Medium/Ham's Nutrient Mixture F-12 (DMEM: F12, nr. 21331020) containing phenol red (Gibco, Paisley, UK), supplemented with 10% (v/v) fetal bovine serum (Gibco, Paisley, UK), 1% (v/v) penicillin-streptomycin 100 U/mL (Gibco, Paisley, UK), 2.5 mM L-glutamine (Lonza, Basel, Swiss), 15 mM HEPES (Gibco, Paisley, UK), 0.5 mM sodium pyruvate (Sigma-Aldrich, Steinheim, Germany), and 1200 mg/L sodium bicarbonate (Gibco, Paisley, UK). The cells were cultured in a humidified environment at 28°C and with 5% CO₂. The cells were passaged

269 weekly at a 1:10 subculture ratio, using phosphate-buffered saline (PBS; pH 7.4; Gibco, Paisley, UK)
270 for washing and 0.25% (w/v) trypsin (Sigma-Aldrich, Steinheim, Germany) for detachment.

271 2.3.2 Oxidative stress reporter assay in human cell line

272 The MCF7/AREc32 (AREc32) reporter line (CVCL_1D32) was obtained from Ximbio (London, UK). The
273 reporter line is based on the human mammary carcinoma cell line MCF7 [9] (CVCL_0031), which was
274 stably transfected with the pGL3-8xARE plasmid vector and named AREc32 [10]. The plasmid vector
275 drives the expression of a Firefly (*Photinus pyralis*) luciferase reporter gene luc+, controlled by eight
276 copies of the rat GST gene's antioxidant response element (gstrARE).

277 MCF7 cells were cultured in high-glucose (4.5 g/L D-glucose) Dulbecco's Modified Eagle's Medium
278 (DMEM GlutMAX-I, nr.32430027) (Gibco, Paisley, UK), supplemented with 10% (v/v) fetal bovine
279 serum (Gibco, Paisley, UK), 1% (v/v) penicillin-streptomycin 100 U/mL (Gibco, Paisley, UK), and 0.8
280 mg/mL geneticin (G418) selective antibiotic (Invivogen, CA, USA). The cells were cultured in the
281 above medium, and whilst plating, they were grown in the same composition medium without the
282 antibiotic G418. The cells were maintained in an incubator at 37 °C under a humidified 5% CO₂
283 atmosphere and passaged twice a week, with medium changes every three days. Phosphate-buffered
284 saline (PBS; at pH 7.4; Gibco, Paisley, UK) was used for washing and 0.05% (w/v) trypsin-EDTA (Gibco)
285 for detachment.

286 2.3.3 Xenobiotic metabolism response reporter assay in zebrafish cell line

287 The reporter assay is based on the zebrafish hepatocyte cell line ZFL [11] (CVCL_3276), transiently
288 transfected with the pGudLuc7.5 and pRL-null[zfEF1aPro] plasmids [12]. The basic ZFL cell line was
289 purchased from ATCC (Manassas, USA). The pGudluc7.5 plasmid vector was donated by Michael S.
290 Denison (University of California, Davies, USA; *in memoriam*:
291 [https://toxchange.toxicology.org/blogs/pamela-lein1/2022/06/30/in-memory-of-michael-steven-](https://toxchange.toxicology.org/blogs/pamela-lein1/2022/06/30/in-memory-of-michael-steven-denison)
292 [denison](https://toxchange.toxicology.org/blogs/pamela-lein1/2022/06/30/in-memory-of-michael-steven-denison)). The latter consists of a pGL3 backbone, 20 xenobiotic response element (XRE) copies from
293 the genomic promoter region of the mCyp1A1 gene upstream of an MMTV viral promoter, and the
294 Firefly (*Photinus pyralis*) luciferase Fluc reporter gene [13], which was employed as the primary
295 reporter. The zebrafish-optimised pRL-null[zfEF1aPro] plasmid vector drives the constitutive
296 expression of the secondary Renilla luciferase reporter gene (*Renilla reniformis*) Rluc, regulated by
297 the endogenous zebrafish translation elongation factor 1 alpha promoter (zfEF1aPro). The secondary
298 RLuc reporter normalises transfection efficiency within the transient setup (DLR). Further details on
299 the assay are provided by Lungu-Mitea et al. [12].

300 Transient transfection was carried out 24h after initial cell plating in 96-well plates, when the cells
301 reached about 80% confluency (see further details in sections 2.3.6 and 2.3.7 below), at a 3 µg

transfection reagent to 1 µg plasmid DNA ratio per reaction using jetPRIME (Polyplus, Illkirch, France). The jetPRIME transfection reagent increases cellular competence via a proprietary disclosed mechanism. Transfection pre-conditioning was conducted in the kit-specific reaction buffer according to the manufacturer's instructions. Primary and secondary reporter plasmids were co-transfected in a 10:1 ratio (0.9 µg reporter plasmid, 0.1 µg control plasmid). Transfection optimisation experiments were conducted according to the manufacturer's guidance, as previously reported [7]. Note that we changed our original protocol from using a competitor's transfection reagent to jetPRIME, as the latter does not induce systemic toxicity in the affected cells (data not shown).

The ZFL cell line was cultured in a medium made from powder, consisting of 50% (v/v) Leibovitz's L-15 (Sigma-Aldrich, Steinheim, Germany), 35% (v/v) Dulbecco's modified Eagle's medium (Gibco, Paisley, UK), 15% (v/v) Ham's Nutrient Mixture F-12 (Gibco, Paisley, UK), and phenol red. Additionally, 150 mg/L sodium bicarbonate (Gibco, Paisley, UK), 15 mM HEPES (Gibco, Paisley, UK), 10 µg/mL bovine insulin (Sigma-Aldrich, Steinheim, Germany), 50 ng/mL mouse EGF (Sigma Aldrich, Steinheim, Germany), and 5% (v/v) fetal bovine serum (Gibco, Paisley, UK) were supplemented. The cells were cultured in a humid environment at 28°C and at atmospheric CO₂. The cells were passaged weekly at a 1:20 ratio, using PBS (pH 7.4; Gibco, Paisley, UK) for washing and 0.25% (w/v) trypsin-EDTA (Sigma-Aldrich, Steinheim, Germany) for detachment.

2.3.4 Xenobiotic metabolism response reporter assay in human cell line

The reporter assay is based on the human hepatoma cell line HepG2 [14] (CVCL_0027), transiently transfected with the pGudLuc7.5 (see section 2.3.3 above for details) and the pGL4.70 (Promega, Madison, USA) plasmid vectors. As such, the transient reporter assay is identical to the stably transfected human AhR CALUX-type clone HG2L7.5c1 [15], with 2,3,7,8-tetrachlorodibenzodioxin (TCDD)-induced EC50 within the same order of potency (2e-10 and 5e-10 M, respectively). Note that several versions of the AhR CALUX-type assays are derived from clones of different species (rat, mouse, human, and guinea pig). The Tox21 dataset is based on the human clone HG2L7.5c1 (<https://tripod.nih.gov/tox/assays>). The pGL4.70 plasmid vector drives the constitutive expression of the secondary “humanised” Renilla luciferase reporter gene (*Renilla reniformis*) hRLuc, which is optimised for expression within human cell lines and regulated via a minimal artificial promoter (minP). The secondary hRLuc reporter normalises transfection efficiency within the transient setup (DLR).

Transient transfection was carried out 48h after initial cell plating on 96-well plates when the cells reached a confluency of about 50% (see details in the sections 2.3.6 and 2.3.7 below) at a 0.3 µl transfection reagent to 100 ng plasmid DNA ratio per reaction using Lipofectamine 2000 (Invitrogen,

Thermo Fisher Scientific, USA). The Lipofectamine 2000 transfection reagent increases cellular competence via lipid fusion. Transfection preconditioning was performed in OptiMEM medium (Gibco, Paisley, UK) according to the manufacturer's instructions. Primary and secondary reporter plasmids were co-transfected in a 10:1 ratio (0.9 µg reporter plasmid, 0.1 µg control plasmid). Transfection optimisation experiments were conducted according to the manufacturer's guidance, as reported previously [16].

HepG2 cells were cultured in high-glucose Dulbecco's Modified Eagle Medium (DMEM, nr.31053028) (Gibco, Paisley, UK), supplemented with 10% (v/v) fetal bovine serum (Gibco, Paisley, UK), 1% (v/v) penicillin-streptomycin 100 U/mL (Gibco, Paisley, UK), and 2 mM L-glutamine (Gibco, Paisley, UK). The cells were maintained in an incubator at 37 °C in a humidified atmosphere of 5% CO₂. The medium was changed every 3 days, and cells were passaged twice a week. PBS (pH 7.4; Gibco, Paisley, UK) was used for washing, and 0.05% (w/v) trypsin-EDTA (Gibco) for detachment.

2.3.5 Xenobiotic metabolism response reporter assay in mouse cell line

The stably transfected reporter line DREcoScreen (DRE) [17] (CVCL_EG66) was obtained from Hiro Biotech (Singapore) via the Japanese Collection of Research Bioresources (Ibaraki, Osaka, Japan) and is based on the mouse hepatoma cell line Hepa1c1c7 [18] (CVCL_0328). DRE has been stably transfected with the pIND-DCDR7 plasmid vector driving the expression of a Firefly (*Photinus pyralis*) luciferase reporter gene luc+, under the control of seven copies of the mouse CYP1A1 gene's xenobiotic response element (cypmXRE).

DRE cells were cultivated in α-Minimum Essential Media (α-MEM, nr.12561) (Gibco, Paisley, UK), supplemented with 5% (v/v) fetal bovine serum (Gibco, Paisley, UK), 1% (v/v) penicillin-streptomycin 100 U/mL (Gibco, Paisley, UK), and 150 µg/mL Hygromycin B gold (InvivoGen, USA). The cells cultured in flasks were grown in the above medium, while cells plated for assays were grown in the composition medium without the antibiotic Hygromycin. The cells were maintained in an incubator at 37 °C in a humidified atmosphere of 5% CO₂. The medium was changed every three days, and cells were passaged twice a week. PBS (pH 7.4; Gibco, Paisley, UK) was used for washing, and 0.05% (w/v) trypsin-EDTA (Gibco) for detachment.

2.3.6 Cell handling and plating

Cells of running cultures were seeded on the first day of the experiment onto white, clear-bottom 96-well microtiter plates (Corning, New York, USA), in 100 µL/well culture medium, at different cell concentrations, regarding the used cell line; ZF4, 2.5*10⁵ cells/mL; AREc32, 1.5*10⁵ cells/mL; ZFL, 1.5*10⁵ cells/mL; HepG2, 1.5*10⁵ cells/mL; DRE, 10⁵ cells/mL. Afterwards, plated cells were returned to the incubator for 24h.

368 2.3.7 Transient transfection

369 After 24h, cultures of ZF4 and ZFL reached approximately 80% confluency, whereas 48h cultures of
370 HepG2 reached 50% confluency and were subjected to transient transfection according to the
371 manufacturer's instructions and specific optimisations (see sections above for particular assays).
372 After assay-specific pre-transfection treatment, 10 µL/well transfection mix in reduced medium was
373 added to the cells, containing 100 ng/well pooled plasmid DNA and the respective ratio amount of
374 transfection reagent. Subsequently, the cells were returned to the incubator for 24 h. Stably
375 transfected lines AREc32 and DRE did not follow this procedure.

376 2.3.8 Exposure

377 The following day, the culture medium was removed, and the cells were exposed to the respective
378 regimen for 24 h at 100 µL/well. After 24h of transfection in ZFL, ZF4, and HepG2 cells, the cells were
379 exposed, whereas DRE and AREc32 cells were treated directly after 24h of plating. For the oxidative
380 stress assays (ZF4 and AREc32), cells were exposed to four concentrations of the respective
381 environmental samples at 1:1 dilutions, starting at REF 20 (REF 20, 10, 5, and 2.5). For the xenobiotic
382 metabolism assays (ZFL, HepG2, and DRE), cells were exposed to four concentrations of the
383 respective environmental samples at 1:1 dilution ratios, starting at REF 2 (REF 2, 1, 0.5, and 0.25).
384 Cells were exposed in triplicate or quadruplicate per concentration. Each experiment was replicated
385 at least three times. DRE cells were exposed only to influent samples from the three WWTP sites of
386 interest. Note that A_I and B_I sample concentrations differed slightly due to initially varying SPE
387 enrichment factors.

388 Additionally, a standard reference compound control was applied to every plate. tert-
389 butylhydroquinone (tBHQ; CAS 1948-33-0; 97% purity; 100 µM stock in DMSO) (Sigma-Aldrich, St.
390 Louis, USA) was utilised as a reference compound of the oxidative stress MoA, whereas 2,3,7,8-
391 tetrachlorodibenzodioxin (TCDD; CAS 1746-01-6; 10 µg/ mL in toluene; Supelco analytical standard;
392 32 µM stock in DMSO) (Sigma- Aldrich, Steinheim, Germany) for the xenobiotic metabolism. Stock
393 solutions were stored at -20 °C in a desiccator. For every assay, cells were exposed to five increasing
394 concentrations of the reference compounds. tBHQ was diluted in 1:1 titration steps, starting at 100
395 µM (100, 50, 25, 12.5, and 6.25 µM). The exposure concentrations were identical for both assays (ZF4
396 and AREc32). TCDD was titrated in half-log steps of different potency ranges in regards to the used
397 assay, giving differences in species-specific sensitivity; ZFL: 30, 10, 3, 1, and 0.3 nM; HepG2: 3, 1, 0.3,
398 0.1, and 0.03 nM; DRE: 300, 100, 30, 10, 3, 1, 0.3, 0.1 pM. Cells were exposed in triplicate or
399 quadruplicate at each concentration. Each experiment was replicated at least three times.

The exposure medium was spiked with 0.1% (v/v) DMSO as the solvent vehicle, and the highest exposure concentrations across the specific setups were adjusted to yield the same final solvent concentration. A negative control (vehicle control) consisting of six to nine well-replicates was tested on every plate. Among the latter, a triplicate control was performed without solvent and pooled. Non-solvent controls consisted of wells including cells but non-spiked medium. A potential statistically significant difference between the two controls was assessed using two-sided Mann-Whitney U tests (alpha level = 0.05). Pilot studies, the first conducted experiments (see SI1), and wherever plate space was provided (see SI5), also included background controls. Background controls consisted of empty wells to determine background noise levels for the specific lumino- or colourimetric endpoint measurements at the end of each assay. Pilot and historical data from multiple high-throughput screenings did not show statistically significant differences between solvent and non-solvent controls. Background measures were historically stable, comprising approximately 70 raw luminescence units for the luciferase readout and 0.05 raw absorbance units for the MTS readout. The latter comprises 0.01% (luminescence) and 5% (absorbance) of the control values. To save plate space, background controls were omitted, and historical values were subtracted instead. Thus, we could fit all environmental samples, reference compounds, and vehicle controls onto one 96-well plate per assay and per experiment (e.g., raw data in SI1-5 for plate setup). We considered reducing inter-plate variability to be the better trade-off, given the exhaustive historical experience with the previously mentioned controls.

2.3.9 Viability and reporter-related readouts

After 24h exposure, cellular viability and the reporter-related luciferase readouts were both acquired from the same plate by multiplexing a [3-(4,5-dimethyl thiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium]-based ("MTS"; CellTiter 96® AQueous One Solution Cell Proliferation Assay; Promega, Madison, USA) viability assay with a luciferase or Dual-Luciferase® ("DLR"; Promega, Madison, USA) reporter assay. The MTS assay was conducted according to the manufacturer's protocol. The latter measures the NADPH-dependent turnover of a formazan-salt product in a rapid, colourimetric reaction, thereby indirectly revealing the activity of viable cells. Because the tetrazolium salt is water-soluble, cell lysis is not required. Furthermore, the reagent is non-toxic to cells and enables subsequent readout of the luciferase signal from the same cell population.

2.3.10 MTS-based cellular viability assay

After 24 h of exposure to environmental samples and controls, the exposure medium was discarded, the cells were washed once with 200 µL/well of PBS, and 100 µL/well of Earle's Balanced Salt Solution (EBSS; Gibco, Paisley, UK) was added. For ZFL cells cultured at atmospheric CO₂ levels, 10 mM HEPES was added to the EBSS medium. To start the viability readout, 17% (v/v) MTS reagent was added to

every well. Given their specific cultivation needs, the plates were returned to the incubator for 30 min (mammalian cell lines) or 60 min (fish cell lines). The absorbance of formazan product turnover was measured at 490 nm using a Tecan Spark (Tecan, Männedorf, Switzerland) plate reader.

2.3.11 Luciferase-based reporter readouts

Following the MTS viability readout, cells were washed once in 200 µL/well PBS and lysed in 20 µL/well passive lysis buffer ("PLB"; Promega, Madison, USA) for 30 min on an orbital plate shaker at room temperature. Total raw luminescence units were measured on a Tecan Spark plate reader using a flash luminescence protocol and an auto-injection setup. For stably transfected, single-luciferase reporter assays (AREc32 and DRE), 10 µL/well Fluc substrate (Luciferase Assay System; Promega, Madison, USA) was injected, and the reaction was recorded for 2s. For the DLR reporter assays (ZF4, ZFL), 50 µL/well and (HepG2) 45µL/well Fluc substrate were injected, the reaction was recorded for 2s, quenched by addition of 50/45 µL/well, respectively, Rluc "stop&glow" substrate (Dual-Luciferase®; Promega, Madison, USA), and the secondary signal was recorded for 2s.

2.4 Bioanalytical assessment: the FET assay

Further bioanalytical assessment of the environmental samples has been conducted using a modified version (mFET [19]) of the FET (OECD TG 236 [1]), which also includes sublethal endpoints up to 144 hpf, such as movements per minute, heart rate, and hatching time. In this setup, we also investigated the upstream and downstream recipient samples associated with the respective WWTPs (refer to DIB_A).

2.5 In silico-mediated effect-directed analysis (ISED)

2.5.1 tcpl and MySQL R coding packages

The R package ToxCast (TM) Analysis Pipeline (tcpl) [20], in connection with the MySQL ToxCast database interface [21] was utilised in an *in silico*-mediated effect-directed analysis (ISED) to discover the primary drivers of toxicity pathway induction for the xenobiotic metabolism MoA with regard to the compounds identified in the targeted chemical analysis.

As a reference, the total AhR CALUX-type [15] concentration-response data ("TOX21_AhR_LUC_Agonist"; human AhR-CALUX clone HG2L7.5c1) were extracted using the following parameters in the command line: assay IDs 543 and 544 (acid), assay acronyms "TOX21_AhR_LUC_Agonist" and "TOX21_AhR_LUC_Agonist_viability" (acnm/aenm), and data IDs 806 and 807 (aeid). This assay was chosen because it is technically closely related to the HepG2 reporter assays used in this study. EC₂₀ point estimates were derived from the extracted concentration-response relationships within the constructed database (see SI12). Given the scenario that multiple datasets were available per compound and assay, the geometric mean of the derived

467 EC₂₀s per compound was computed. Relative effect potencies (REPs) were derived as a ratio of EC₂₀ of
468 TCDD to the EC₂₀ of the inquired compound.

469 Additionally, using the extracted cytotoxicity data (“TOX21_AhR_LUC_Agonist_viability”), cytotoxicity
470 thresholds were set either above 50 µM for each compound or at each compound's specific cut-off
471 level. Such compounds were flagged within the database for further analysis. Finally, the total
472 database was compared with the identified compounds from the targeted chemical analysis using
473 the “VLOOKUP” function in MS Excel to retrieve CAS numbers and names. Given that a chemically
474 detected compound could not be found in the extracted ToxCast data, the process was repeated for
475 its respective salt.

476 For compounds flagged as cytotoxic, the data extraction process was independently repeated to
477 ensure accuracy using the “CLD_CYP1A1_24hr”, “CLD_CYP1A2_24hr”, “CLD_CYP1A1_48hr”, and
478 “CLD_CYP1A2_48hr” command line parameters (see details in SI12). Given the scenario in which EC₂₀
479 estimates below the initial cytotoxicity cut-off could be obtained, these values were forwarded for
480 further analysis.

481 2.5.2 OCHEM AhR-NR QSAR

482 All positive hits and cytotoxicity-flagged compounds were forwarded to the OCHEM AhR nuclear
483 receptor QSAR [22], [23], where potential receptor interference was estimated from SMILES inputs.
484 This step aimed to infer potential receptor activation based on SMILES inputs, as depicted in SI12.
485 Compounds flagged as “cytotoxic but active” underwent further testing as pure compounds in the
486 DRE reporter assay, as outlined in section 2.3. Only compounds depicting activity with the DRE assay
487 were forwarded to subsequent analysis.

488 2.6 Molecular docking: assessing receptor-ligand affinity

489 2.6.1 Ligands

490 As employed in the bioanalytical assessment (see section 2.3, above), the reference compounds
491 tBHQ, MZC, TCDD, and the three identified compounds from ISEDA (climbazole, daidzein,
492 thiabendazole) were utilised as *in silico* ligands for molecular docking. Ligand chemical structures and
493 PubChem chemical identifiers (CIDs) are depicted below in Figure S2.

494 3D structures of all ligands were downloaded from the PubChem database
495 (<https://pubchem.ncbi.nlm.nih.gov> [24], [25]) and optimised using the Avogadro 1.2.0 software
496 (<https://avogadro.cc> [26]). Optimisation was conducted as follows. The multiplicity of bonds and
497 hydrogen atoms was checked, and the structures were minimised using the steepest descent
498 algorithm in the “Auto Optimise tool” of Avogadro, utilising the Universal Force Field (UFF)

parameter set. Next, the ligands were uploaded to the RESP ESP charge Derive (RED) Server Development 2.0 [27] to derive the restrained electrostatic potential (RESP) charges. Finally, AutoDock atom types were added, and PDBQT-type files were generated via MGLTools (<https://ccsb.scripps.edu/mgltools> [28]–[30]). All PDBQT-type files are available in the supplementary folder S110.

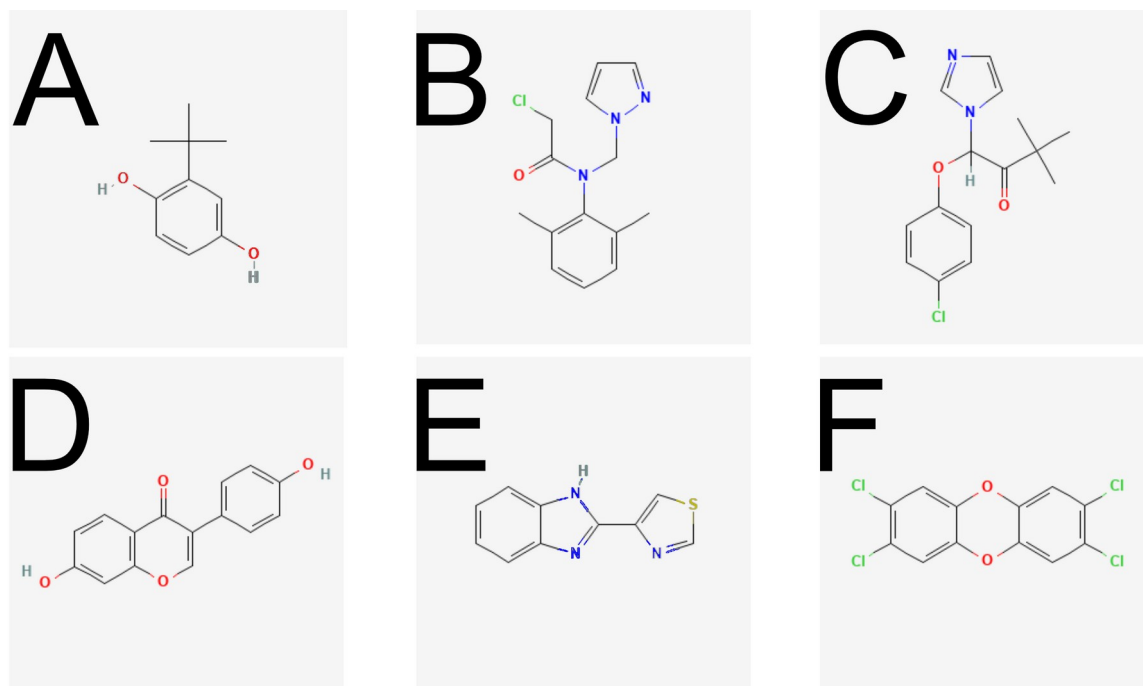


Fig. S2: 2D representation of all selected ligands. Structures were retrieved from the PubChem database and modified. (A) tert-butylhydroquinone (tBHQ, CID: 16043), (B) Metazachlor (MZC, CID: 49384), (C) Climbazole (CLB, CID: 37907), (D) Daidzein (DAI, CID: 5281708), (E) Thiabendazole (TBDZ, CID: 5430), (F) 2,3,7,8-Tetrachlorodibenzodioxin (TCDD, CID: 15625).

2.6.2 Tertiary protein structure selection and prediction: Keap1 and AhR PAS-b domain

As there are no experimental full-sequence protein structures available for Keap1 (fig. S3), the structures were predicted by the AlphaFold2 tool (<https://alphafold.ebi.ac.uk> [31], [32]) from human and zebrafish Keap1 protein sequences available at UniProt (<https://www.uniprot.org> [33]; hKeap1: Q14145, zfKeap1a: Q1ECZ2, zfKeap1b: A0A2R8Q1W5). For comparison, two experimental hKeap1 structure fragments were retrieved from the RCSB Protein Data Bank (<https://www.rcsb.org> [34], [35]). PDB ID 7EXI covers residues 50 to 179 with a resolution of 1.82 Å, and PDB ID 1ZGK covers residues 322 to 609 with a resolution of 1.35 Å.

For AhR (fig. S4), there are no full-sequence experimental structures either. For the human AhR variant (hAhR), a partial cryo-EM structure, PDB ID 7ZUB chain D [36], was selected as the receptor and retrieved from the RCSB Protein Data Bank. This structure covers residues 271 to 427 of hAhR, the PAS-b ligand-binding domain, with a resolution of 2.85 Å. For zebrafish AhR (zfAhR), there are three variants available at UniProt (zfAhR1a: A0A8N7UZF6, zfAhR1b: Q4U3K9, and zfAhR2: Q9YGV3).

Since there are no experimental structures available for these sequences, the PAS-b domains of the three zebrafish AhR variants were modelled with AlphaFold2, corresponding to residues 271 to 385 in zfAhR1a, 281 to 395 in zfAhR1 b, and 289 to 403 in zfAhR2.

The predicted full protein structures of Keap1 were retrieved from the AlphaFold database (<https://alphafold.ebi.ac.uk> [32]) via UniProt, the PAS-b domains of zfAhR variants were predicted using AlphaFold2 in ColabFold (<https://colab.research.google.com/github/sokrypton/ColabFold> [37]), and the X-ray and cryo-EM structures were downloaded from RCSB PDB. All zebrafish structures were aligned to the predicted full-length human sequence structure in PyMOL (PyMOL Molecular Graphics System, v. 1.8.4.) and stripped of all HETATM records (non-protein atoms). All positions mentioned throughout the project correspond to the hKeap1 and hAhR amino acid sequences (see figs. 6, 7, and 13 in DIB_B). The structure PDB ID 1ZGK contains seven modified selenomethionine residues, which were converted to methionine residues by replacing the selenium atoms with sulphur. The structures were protonated using the H++ web server v. 4.0 (<http://newbiophysics.cs.vt.edu/H++/> [38], [39]) with pH = 7.0, salinity = 0.1 M, internal dielectric = 10, and external dielectric = 80. AutoDock atom types and Gasteiger charges were added to the receptors by MGLTools, and the corresponding PDBQT files were generated (supplementary folder SI10). All receptors were prepared without any ligands.

2.6.3 Amino acid sequence alignment

FASTA files of the human and zebrafish sequences named in section 2.6.2 above were acquired from the National Centre for Biotechnology Information's (NCBI) webpage (<https://www.ncbi.nlm.nih.gov/> [25]) and imported into the Unipro Ugene alignment software (<https://ugene.net/> [40]). All FASTA-type files used are provided in SI10. AA multiple sequence alignment was conducted via the ClustalW algorithm, utilising the following parameters: gap open penalty 10, gap extension penalty 0.2, weight matrix BLOSUM, gap separation distance 4, weight transition no, hydrophilic gaps yes. Sequence alignment and distance matrices were computed in Unipro Ugene, exported, and graphically aligned in Affinity Designer (Serif, Nottingham, UK). Simplified, linear domain maps were designed in IBS Illustrator (<https://ibs.biocuckoo.org/> [41]).

2.6.4 Topology, mapping, and nomenclature

This section gives 3D structures and domain maps for Keap1 (Fig. S3) and AhR (Fig. S4) to facilitate orientation. Section 4.1 below provides a summary of the Nrf2-Keap1-ARE pathway for the interested reader, in relation to the docking modelling conducted. We refer to relevant literature for more details about the AhR-ARNT-XRE pathway (e.g., [5]).

Notably, humans have only one receptor/sensor variant, whereas fish have multiple ones due to the teleost genome duplication event [42].

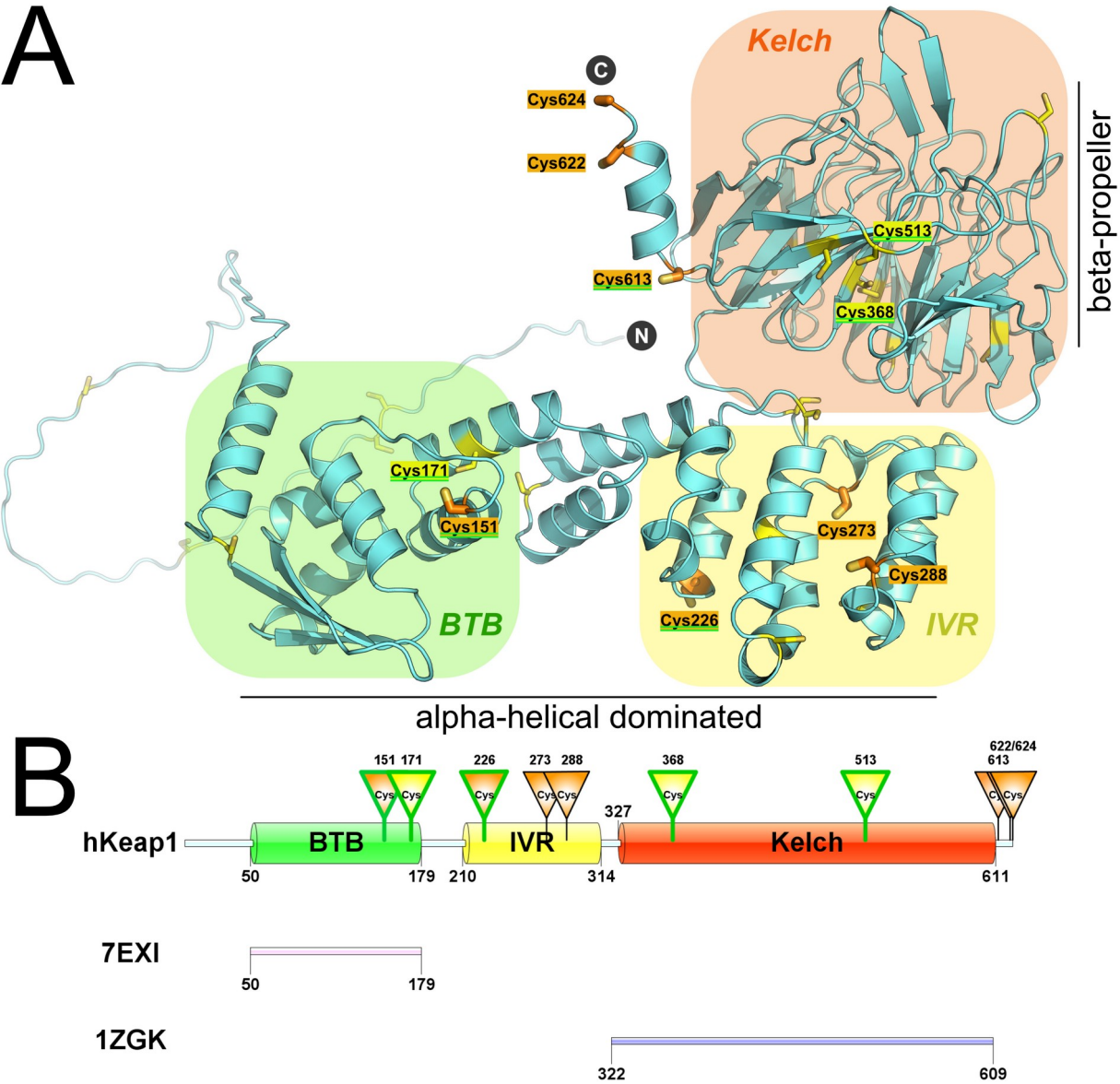
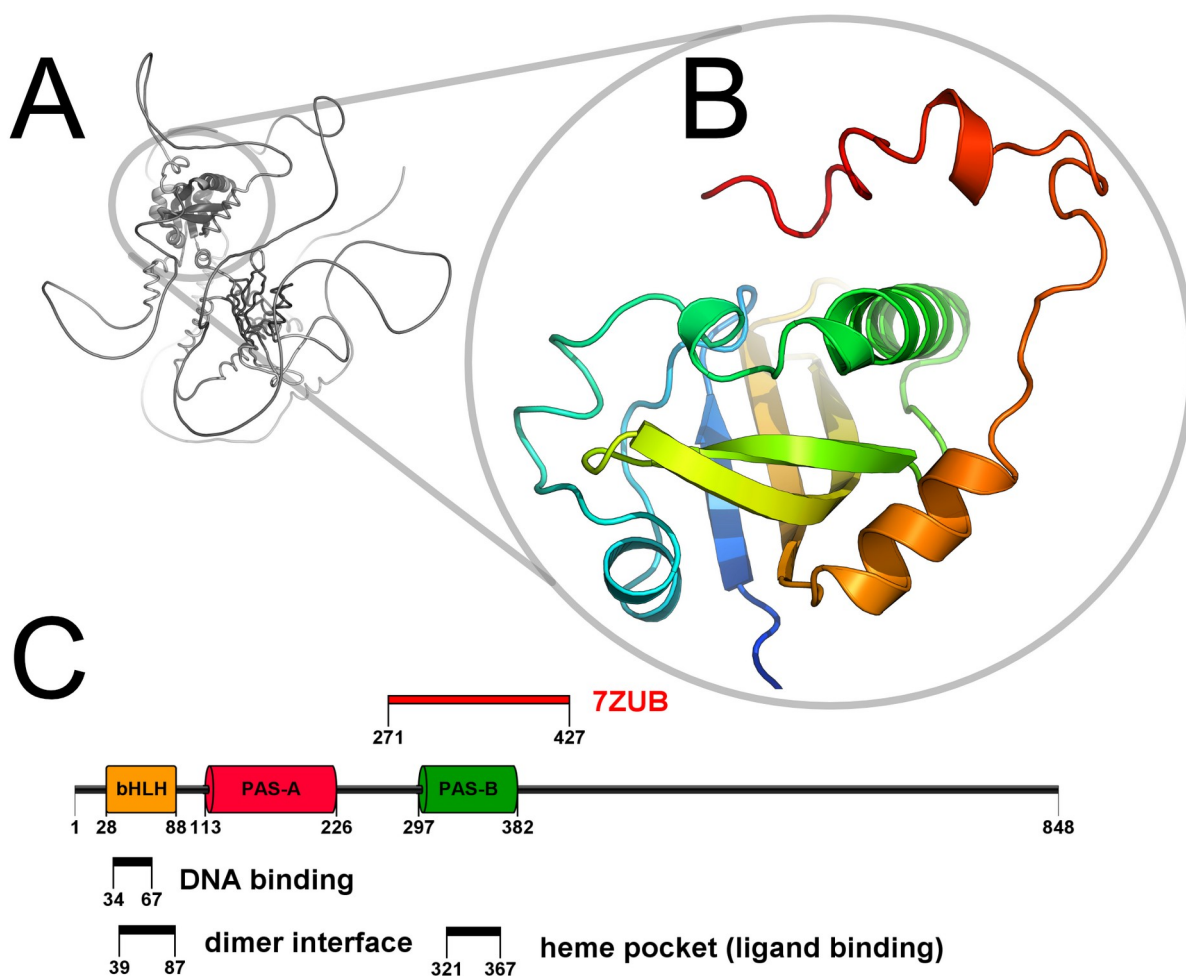


Fig. S3: Topology of hKeap1. The 3D structure is shown as a cyan cartoon (panel A). A simplified sequence structure is given below (panel B), to which partial crystal structures 7EXI and 1ZGK were aligned. Canonical reactive cysteines are highlighted in orange, and other cysteines are highlighted in yellow. Cysteines identified within this study are marked in green. Domains are highlighted as rectangles concerning their region. Alpha-helical dominated region: Broad complex, tram-track, and Bric a brac domain (BTB, residues 50 to 179, green) and the intervening region (IVR, residues 210 to 314, yellow). Beta-propeller region: Kelch domain (residues 327 to 611, orange). The N- and C-termini are illustrated as black-and-white dots labelled N and C, respectively. Illustrations were created in PyMOL and the IBS Illustrator.



563

564 **Fig. S4: Topology of hAhR.** (A) The 3D structure of the entire hAhR, as predicted by AlphaFold, is illustrated in grey. (B) Zoom in
 565 on the PAS-B ligand binding domain of hAhR as a rainbow-coloured cartoon. (C) A simplified sequence structure is given in the
 566 lower left corner, to which the partial cryo-EM structure 7ZUB is aligned. Domains are illustrated as rectangles and cylinders
 567 with regard to their location within the AA sequence: basic helix-loop-helix (bHLH, residues 28 to 58), PAS-A (Per-Arnt-Sim-A,
 568 residues 113 to 226), and PAS-B (Per-Arnt-Sim-B, residues 297 to 382). The respective protein features are given below the
 569 simplified sequence. Illustrations were created in PyMOL and the IBS Illustrator.

570 2.6.5 Molecular docking procedure

571 The AutoDock Vina 1.1.2 algorithm [43] was utilised for molecular docking. As Keap1 is a large
 572 protein with distinct domains, two overlapping grid boxes covering the whole protein were set up in
 573 PyMOL (Schrödinger Inc., New York, US) using the Autodock/Vina plugin [44]. The first docking grid
 574 dimensions were $x = 65 \text{ \AA}$, $y = 68 \text{ \AA}$, $z = 72 \text{ \AA}$, with the center at $x = -10.74$, $y = -32.27$, $z = -7.88$
 575 (Box1). The second docking grid was $53 \times 102 \times 61 \text{ \AA}$ in size, with the centre at $x = -24.74$, $y = -49.27$,
 576 $z = -57.88$ (Box2). Regarding site-directed docking to the reactive Cys151 residue, a $15 \times 15 \times 15 \text{ \AA}$
 577 box was used, centred at $x = -19$, $y = -35$, and $z = -57$ (Box3). For AhR, site-directed docking to the
 578 PAS-b domain was set up by placing a docking grid to cover the binding site, with dimensions 18×18

x 18 Å and centre coordinates at x = 13, y = 6, and z = -24. The flag “exhaustiveness = 100” was used to thoroughly sample the possible conformational space. The number of output conformations of the docked ligand was set to 10 poses. Finally, PyMOL was used to visualise the predicted poses, measure distances, and render pictures. The raw data are presented in section 3 of the DIB_B article and in SI9. Additional box plots of binding energies were created in GraphPad Prism 9 (GraphPad Software, Boston, Massachusetts, USA).

2.7 Quantification of bioavailable concentrations via mass-balance modelling

Bioavailable (or free cellular concentrations, C_{free}) of the utilised reference compounds and the three compounds identified via ISEDA were computed via the UFZ-LSER QSPR [45], employing a method by Fischer et al. [46] that is based on several mass-balance-equation models [47], [48]. For experimental and calculated descriptors, mass-balance equations, biomaterials, model output, summary, and further references, please consult SI13.

The following input parameters were chosen: I) Concentration of analyte, 3 µM for tBHQ and MZC, 30 nM for TCDD and the compounds identified via ISEDA (additionally, AC_{50} point estimates of the ISEDA-identified compounds, as derived via log-logistic regression (see the following section 2.8.1), were transformed into free cellular concentrations, as well). II) Cell culture medium volume: 100 µL. III) Type of base medium: DMEM. IV) Fraction of base medium: 95% for ZFL and DRE, 90% for ZF4, MCF7AREc32, and HepG2. V) Type of serum: FBS. VI) Fraction of serum: 5% for ZFL and DRE, 10% for ZF4, MCF7AREc32, and HepG2 cell lines. VII) Cell number: according to concentrations named in section 2.3.6 above, in 100 µL.

The C_{free} geometric mean was calculated and forwarded for further analysis in the scenario where we obtained multiple free cellular concentration predictions from various mass-balance equations, descriptors, and database queries.

2.8 Data processing and statistical analyses

2.8.1 Reporter gene assays – raw data handling and normalisation

For the initial raw data handling and normalisation procedures, please consult SI1 to SI8, SI14, and SI15. To summarise, in the first step, raw luminescence units from the single-luciferase reporter assay (LR) were used for analysis, as measured by the AREc32 and DRE assays. Similarly, raw absorbance units recorded in the MTS viability assay were not further normalised in the first step. DLR raw luminescence units were normalised to cell number and transfection efficiency by dividing the primary Firefly luciferase (Fluc) signal by the constitutive Renilla luciferase (Rluc) signal for every technical replicate. In the second step, technical replicates within each exposure group were pooled

611 to obtain experimental means, which were then forwarded to subsequent data transformation and
612 analysis procedures, as described below.

613 2.8.2 Reporter gene assays – regression analyses

614 Data-fitting techniques for simple linear and nonlinear regression were conducted as described by
615 Escher et al. [49], [50], with modifications by Weimer et al [51]. Alternative data-fitting procedures
616 are discussed further below in the SM sections 3.3 and 3.4, along with their respective
617 supplementary results.

618 For the main article, values obtained from the oxidative stress MoA bioassays (ZF4 and
619 MCF7/AREc32) were expressed as induction ratios (IR) in relation to the negative controls. Both
620 environmental sample and reference compound values were fitted to simple linear regressions,
621 anchored at the origin ($f(x)=1$). ECIR_{1.5} point estimates for calculating BEQs were derived from the line
622 fit.

623 Values obtained from the xenobiotic MoA bioassays (ZFL, HepG2, and DRE) were expressed as a
624 percentage of maximum TCDD induction (%min.max.). Data obtained from the environmental
625 samples were fitted to a simple linear regression, anchored at the origin ($f(x)=0$). In contrast, TCDD-
626 reference and ISEDA-compound concentration-response curves were fitted using Hill or log-logistic
627 regression with different parameterisations. Alternative parametrisation and curve fit are further
628 discussed in sections 3.3 and 3.4. The 3PNL model was applied throughout the main article, as it
629 aligns with the guidance provided by Escher et al [50]. EC₁₀ values for BEQ calculations were derived
630 from the line and curve fits, respectively.

631 To compare general effect differences between reporter lines within the same exposure setup, the
632 above-stated values were log₁₀- or Box-Cox-transformed and subjected to a 2-way ANOVA with
633 “reporter line” and “exposure concentration” as factors. For tests involving three different reporter
634 lines (xenobiotic metabolism MoA, all influent samples), the 2-way ANOVA was followed by Tukey's
635 multiple comparison test. Normality of residuals via Shapiro-Wilk and Kolmogorov-Smirnov tests,
636 homoscedasticity via Spearman's rank correlation test, and optical inspection of residuals via QQ-plot
637 and residual plot (predicted vs. fitted residuals) were conducted to ensure that ANOVA criteria were
638 met. A $p<0.05$ was considered statistically significant, while a $p<0.1$ indicated a trend. All calculations
639 in the main article were performed using GraphPad Prism 9 (GraphPad Software, Boston, USA).

640 2.8.3 Iceberg modelling

641 Iceberg modelling was conducted as described by Neale et al. [52]–[54]. Relative effect potencies
642 (REPs) of the identified compounds within the targeted chemical analysis were calculated by dividing
643 the EC₂₀ of TCDD by the geometric mean point estimates (EC₂₀) derived from the ISEDA, as obtained

644 from ToxCast AhR CALUX-type bioassay (TOX21_AhR_LUC_Agonist; human AhR-CALUX clone
645 HG2L7.5c1) data (see section 2.5).

$$646 \quad REP = \frac{EC_{20;RC}}{\prod_{i=1}^n EC_{20;i}}$$

647 With RC: reference compound (here TCDD), i: compound identified in ISEDA.

648 Chemical equivalents (BEQ_{chem}) were calculated by summing up the respective positive ISEDA hits for
649 every WWTP sample (see SI12).

$$650 \quad BEQ_{Chem} = \sum_{i=1}^n \frac{REP_i * c_i * MW_{RC}}{MW_i}$$

651 With RC: reference compound (here TCDD), i: compound identified by chemical analysis, MW:
652 molecular weight, c: concentration, REP: relative effect potency.

653 Biological equivalents (BEQ_{bio}) were calculated by dividing the computed effect point estimates of the
654 reference compound ($EC_{IR1.5}$ for tBHQ and MZC, EC_{10} for TCDD) by the point estimates derived from
655 the WWTP sample regression fitting (see SI16).

$$656 \quad BEQ_{Bio; ox. stress} = \frac{EC_{IR1.5;RC}}{EC_{IR1.5;i}}$$

$$657 \quad BEQ_{Bio; xen. met.} = \frac{EC_{10;RC}}{EC_{10;i}}$$

658 With RC: reference compound (here tBHQ, MZC, and TCDD), i: environmental sample.

659 Percentage contribution of BEQ_{chem} to BEQ_{bio} :

$$660 \quad [\%] = \frac{BEQ_{Chem}}{BEQ_{Bio}} * 100$$

661 2.8.4 Removal efficiencies

662 Removal efficiency (%) from chemical analyses:

$$663 \quad [\%] = \frac{\sum_{i=1}^{n;influent} c_i - \sum_{i=1}^{n;effluent} c_i}{\sum_{i=1}^{n;influent} c_i} * 100$$

664 Removal efficiency (%) from bioanalytical assessment:

665

$$[\%] = \frac{BEQ_{inf} - BEQ_{eff}}{BEQ_{inf}} * 100$$

666 2.8.5 Statistical analysis of molecular docking-derived data

667 To compare differences between receptor/sensor binding energies for all investigated zebrafish and
 668 human variants, up to ten binding poses (where possible) per binding site were subjected to non-
 669 parametric Kruskal-Wallis tests, followed by Dunn's multiple comparison test. A $p < 0.05$ was regarded
 670 as statistically significant. All calculations were conducted in GraphPad Prism 9.

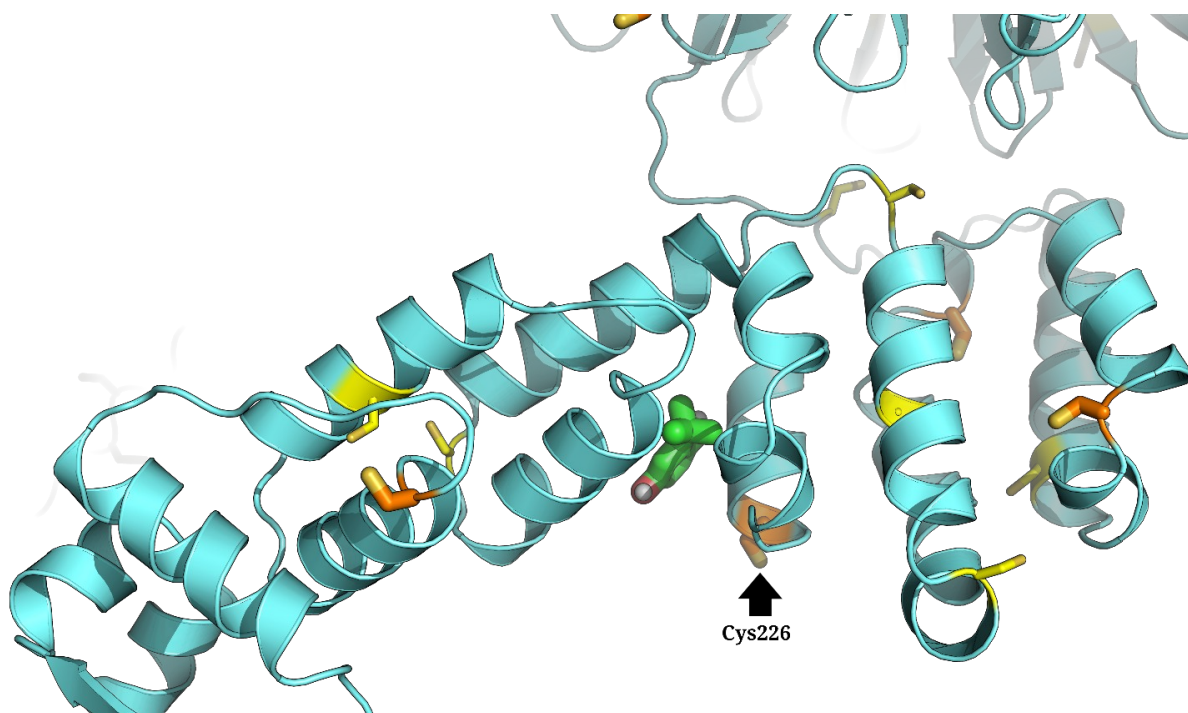
671 2.8.6 Ratios

672 To compare regression fit estimates, BEQ values, and bioavailable concentrations for species
 673 sensitivity, the zebrafish estimates were divided by the corresponding human or mouse estimates.

674 3. Supplementary results

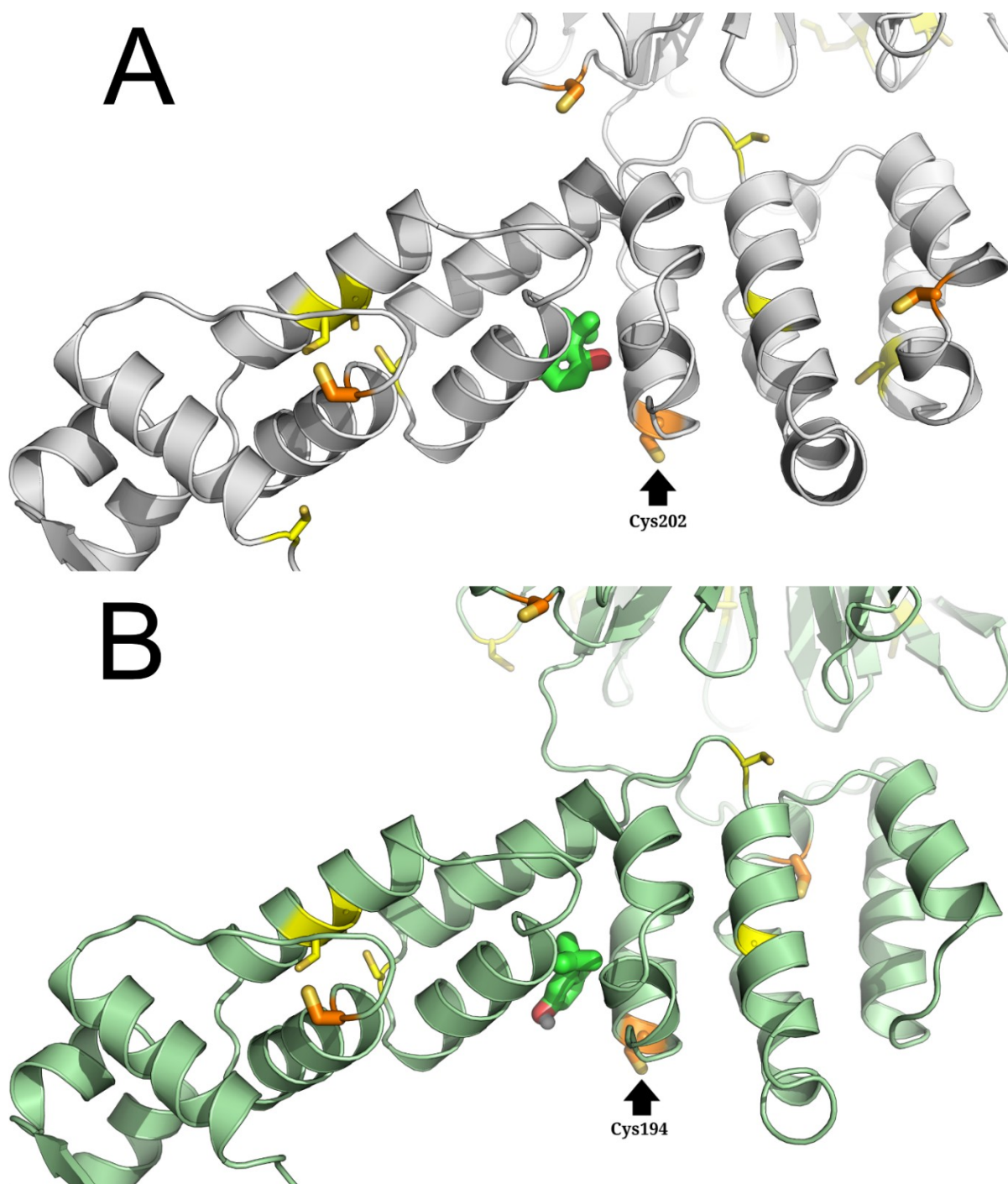
675 3.1 Supplementary results: DIB_B co-submission

676 More detailed and zoomed-in docking poses from blind and site-directed docking to the investigated
 677 ligands



678

679 **Fig. S5: Molecular docking (tBHQ).** Best energy pose (site A, -6.6 kcal/mol) of tBHQ (green ligand) from blind docking to a
 680 predicted full-sequence structure of hKeap1 (cyan cartoon). Here, the alpha-helical region (IVR domain) is emphasised.
 681 Canonical cysteine residues 151, 226, 273, and 288 are highlighted as orange sticks; the remaining ones are in yellow.



682

683 **Fig. S6: Molecular docking (tBHQ).** Best binding pose (site A, -6.3 and -6 kcal/mol, respectively) of tBHQ (green ligands) in the
 684 alpha-helical region (IVR domain) of predicted structures of zfKeap1a (light grey cartoon, A) and zfKeap1b (light green
 685 cartoon, B). Canonical cysteine residues are shown as orange sticks, and the remaining cysteine residues are shown as yellow
 686 lines. The enumerated zebrafish cysteines correspond to Cys226 in hKeap1.

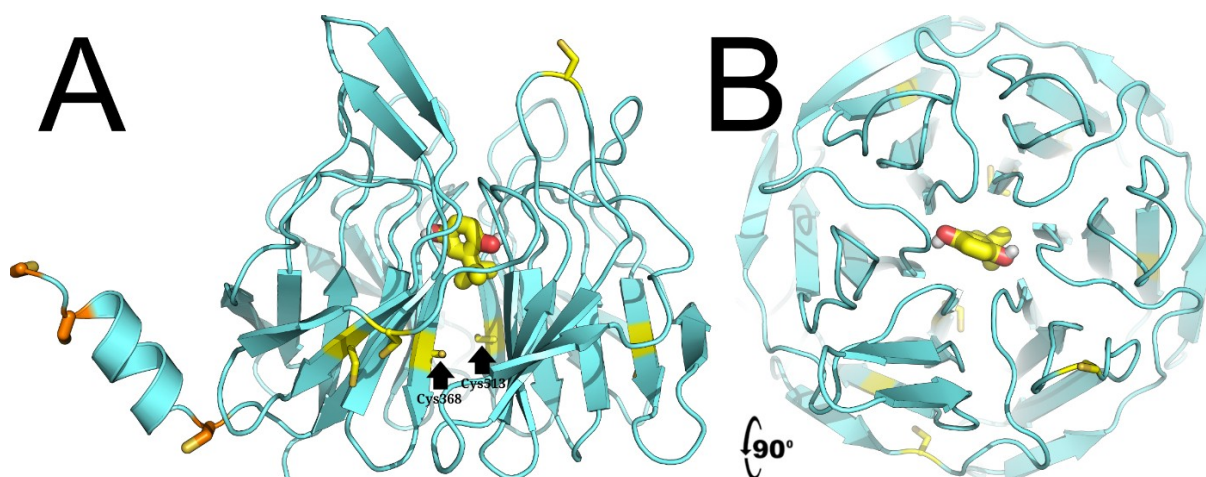


Fig. S7: Molecular docking (tBHQ). Two different views of the second-best pose (site B, -6.2 kcal/mol) of tBHQ (yellow ligand) situated in the central cavity of the Kelch domain from blind docking to the AlphaFold-predicted structure of hKeap1 (cyan cartoon). Canonical cysteine residues are highlighted as orange sticks branching from the protein's backbone, while the remaining cysteines are in yellow.

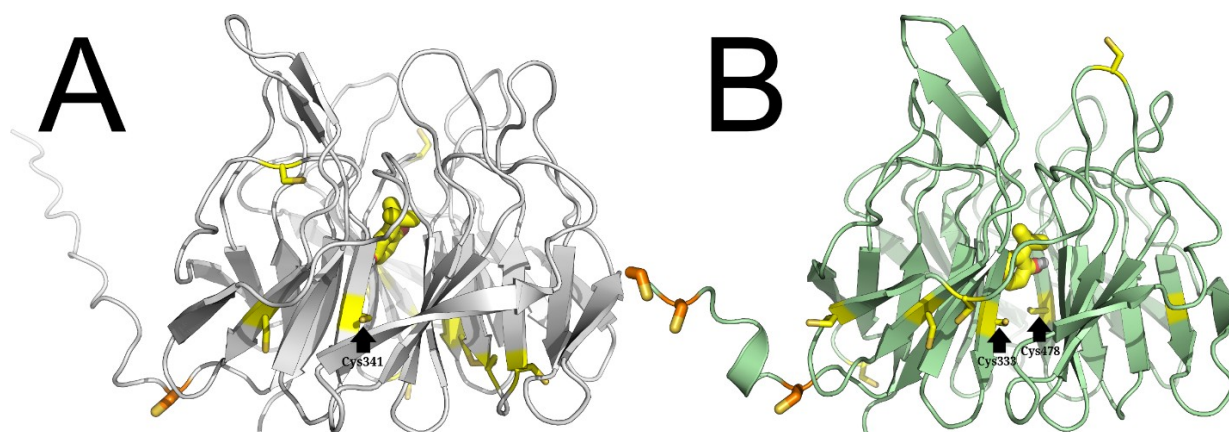
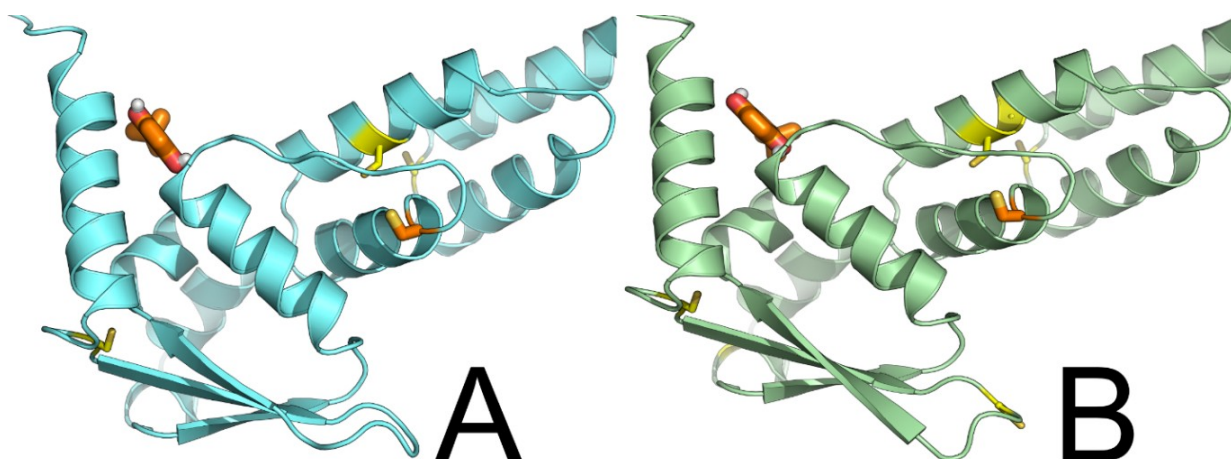
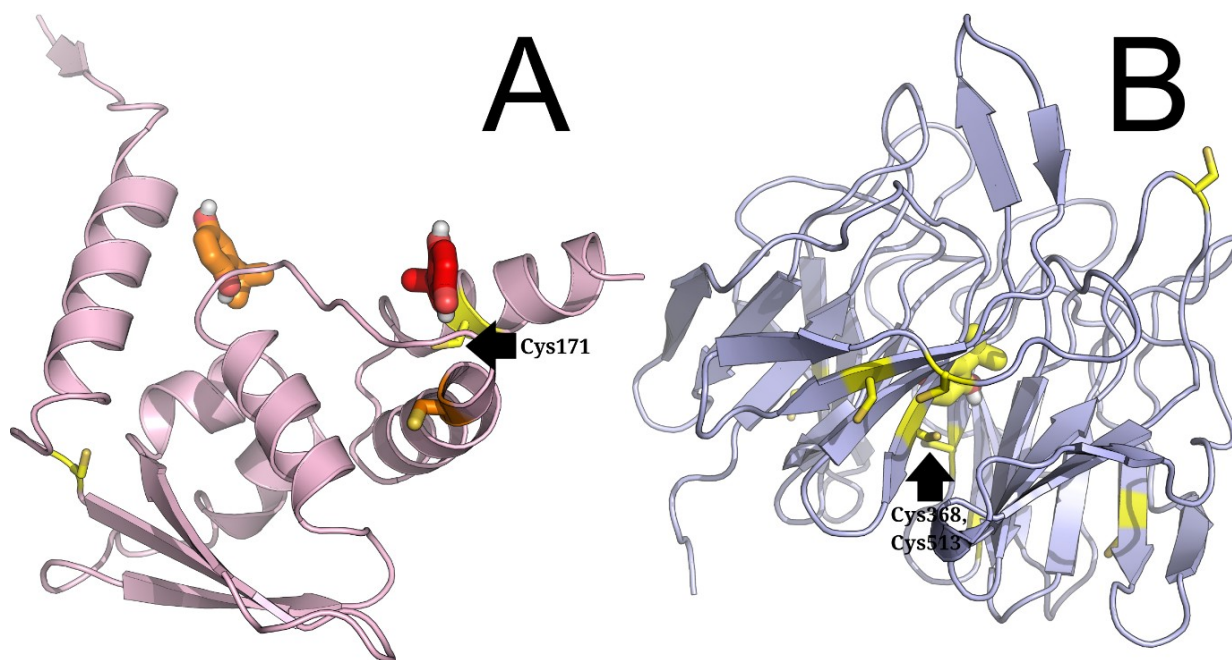


Fig. S8: Molecular docking (tBHQ). The poses of tBHQ (yellow ligands) inside the Kelch domain (site B) of zfKeap1a (light grey cartoon, -6 kcal/mol, A) and zfKeap1b (light green cartoon, -6.1 kcal/mol, B). Canonical cysteine residues are highlighted as orange sticks branching from the protein's backbone, while the remaining cysteines are in yellow. The enumerated zebrafish cysteines correspond to Cys368 and 513 in hKeap1.



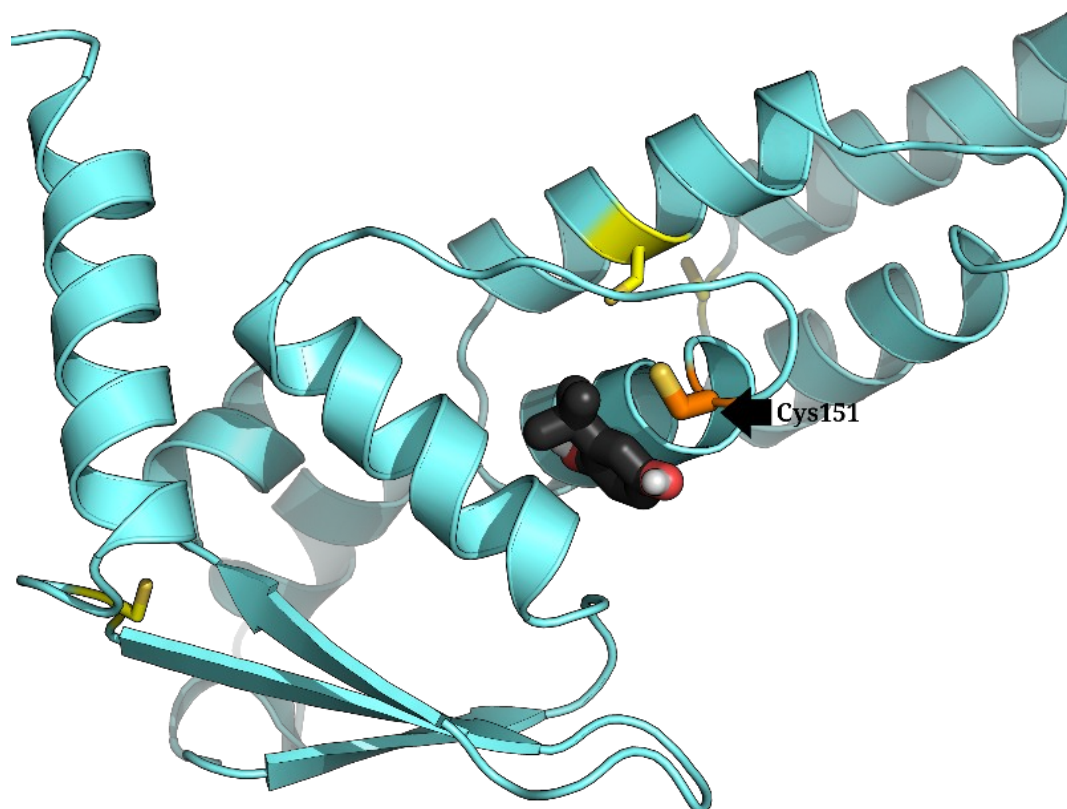
698

699 **Fig. S9: Molecular docking (tBHQ).** tBHQ (orange ligands) in the third-best binding site (site C) situated in the BTB domain from
700 blind docking to predicted structures of hKeap1 (cyan cartoon, -5.5 kcal/mol, A) and zfKeap1b (light green cartoon, -5.3
701 kcal/mol, B). Canonical cysteine residues are shown as orange sticks, the remaining ones as yellow lines.



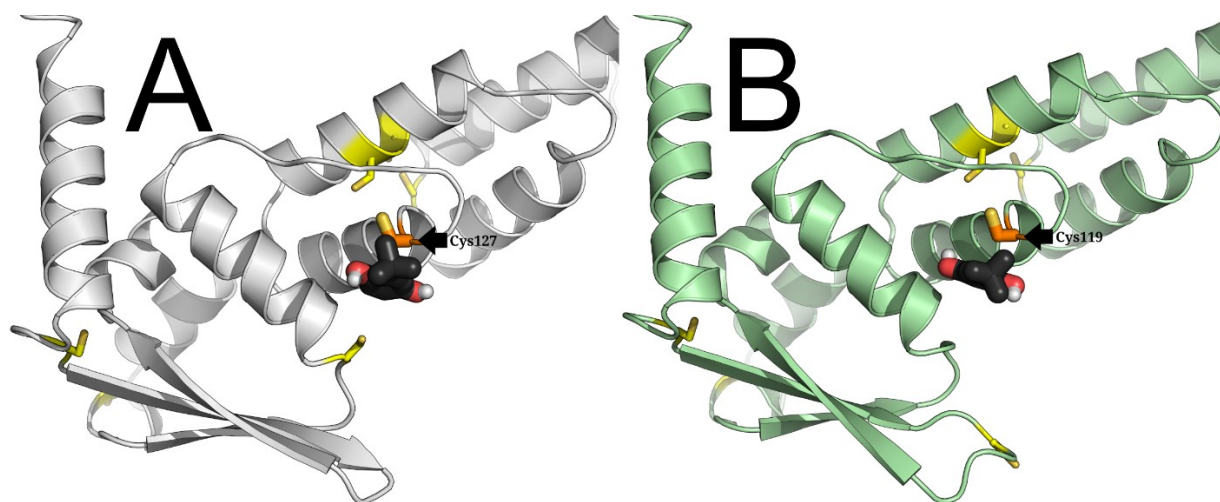
702

703 **Fig. S10: Molecular docking (tBHQ).** Best energy poses of tBHQ (coloured ligands) per site from blind docking to X-ray
704 structures PDB ID 7EXI (light pink cartoon, A) and 1ZGK (light purple cartoon, B). The ligand poses are colour-coded, with
705 yellow corresponding to the best predicted binding energy (site B, -6.3 kcal/mol) and red to the worst (site E, -5.1 kcal/mol).
706 Canonical cysteine residues are highlighted as orange sticks branching from the protein's backbone, while the remaining
707 cysteines are in yellow. All illustrative data is summarised quantitatively in tabs. 3 and 4 (DIB_B article).



708

709 **Fig. S11: Molecular docking (tBHQ).** The best-energy pose for tBHQ (site D, black ligand, -4.6 kcal/mol) from site-directed
 710 docking to the proximity of reactive residue Cys151 in hKeap1 (cyan cartoon). Canonical cysteine 151 is shown as orange
 711 sticks, while the remaining cysteines are in yellow. All illustrative data is summarised quantitatively in tabs. 3 and 4 (DIB_B
 712 article).



713

714 **Fig. S12: Molecular docking (tBHQ).** The best-energy poses of tBHQ (site D, black ligands) from site-directed docking to the
 715 proximity of the canonical Cys151 residue of AlphaFold-predicted structures of zfKeap1a (light grey cartoon, -3.5 kcal/mol, A)
 716 and zfKeap1b (light green cartoon, -3.9 kcal/mol, B). Reactive cysteines 127 and 119 (corresponding to Cys151 in hKeap1) are
 717 shown as orange sticks, while the remaining cysteines are yellow. All illustrative data is summarised quantitatively in tabs. 3
 718 and 4 (DIB_B article).

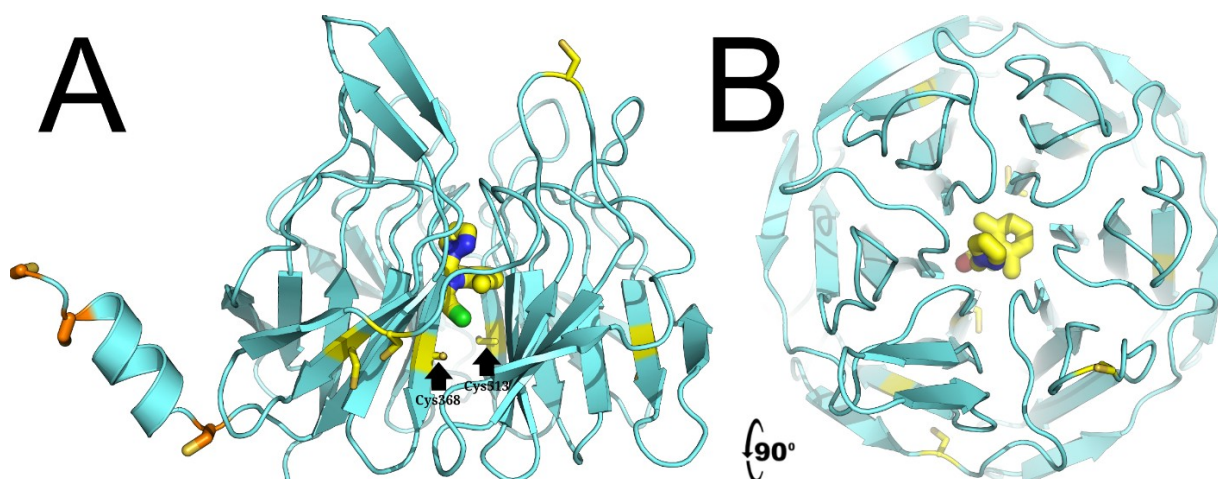


Fig. S13: Molecular docking (MZC). Two different views of the best-energy pose (site A, -6.6 kcal/mol) of MZC (yellow ligands) situated in the central cavity of the Kelch domain of hKeap1 from blind docking to AlphaFold-predicted structure (cyan cartoon). Canonical cysteine residues are highlighted as orange sticks branching from the protein's backbone, while the remaining cysteines are in yellow.

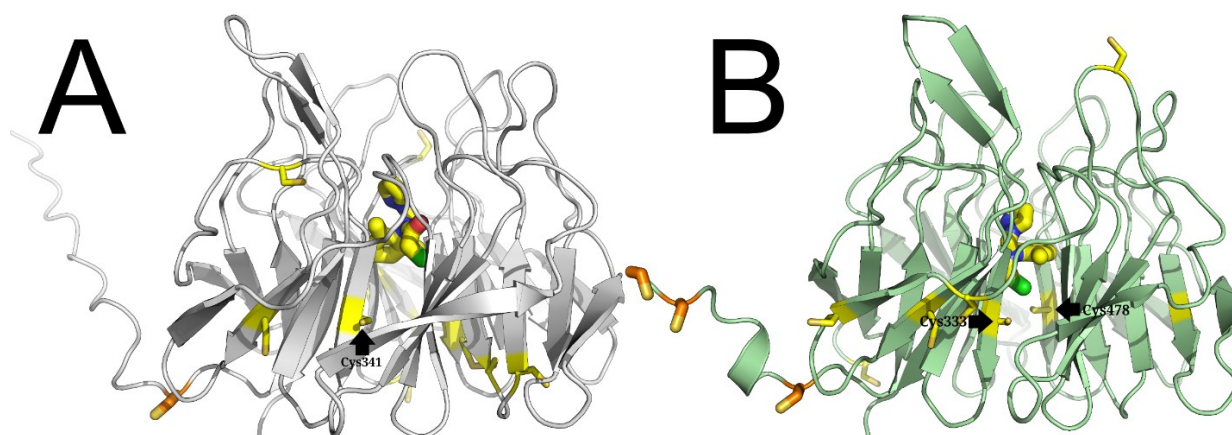


Fig. S14: Molecular docking (MZC). Best-energy poses (site A) of MZC (yellow ligands) inside the Kelch domain of zfKeap1a (light grey cartoon, -6.9 kcal/mol, A) and zfKeap1B (light green cartoon, -6.8 kcal/mol, B). Canonical cysteine residues are highlighted as orange sticks branching from the protein's backbone, while the remaining cysteines are in yellow. Enumerated cysteine residues 333, 341, and 478 correspond to Cys 368 and 513 in hKeap1.

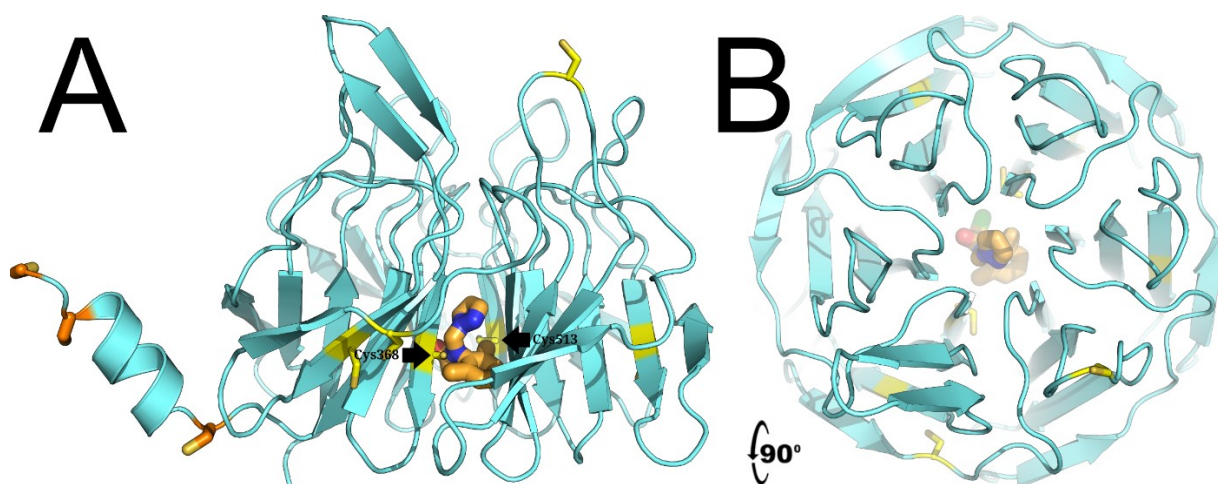


Fig. S15: Molecular docking (MZC). Two different views of the second best-energy site (site B, -6.5 kcal/mol), including a pose of MZC (light orange ligand) situated in the central cavity of the Kelch domain of hKeap1 from blind docking to the AlphaFold-predicted structure (cyan cartoon). Canonical cysteine residues are highlighted as orange sticks branching from the protein's backbone, while the remaining cysteines are in yellow.

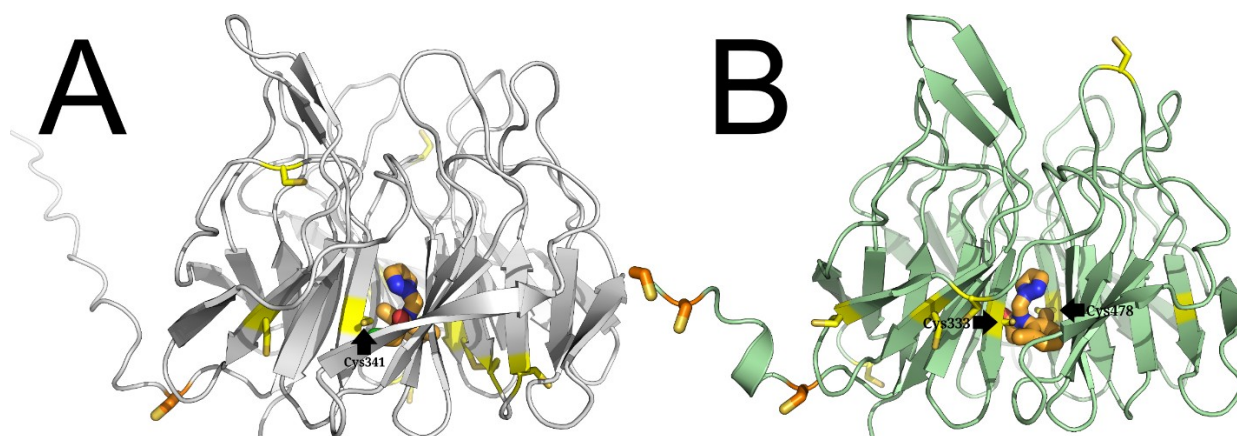
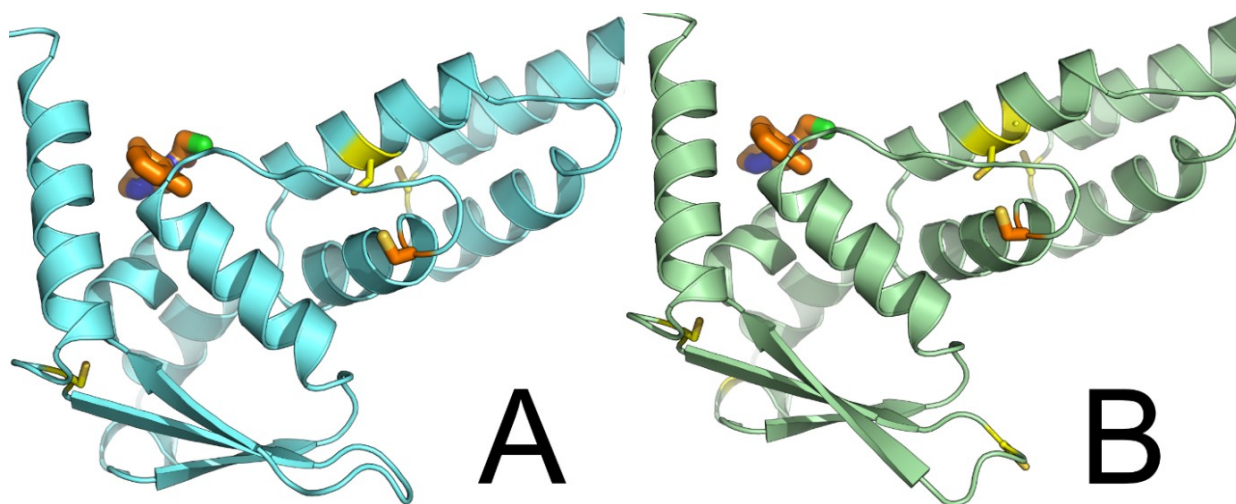
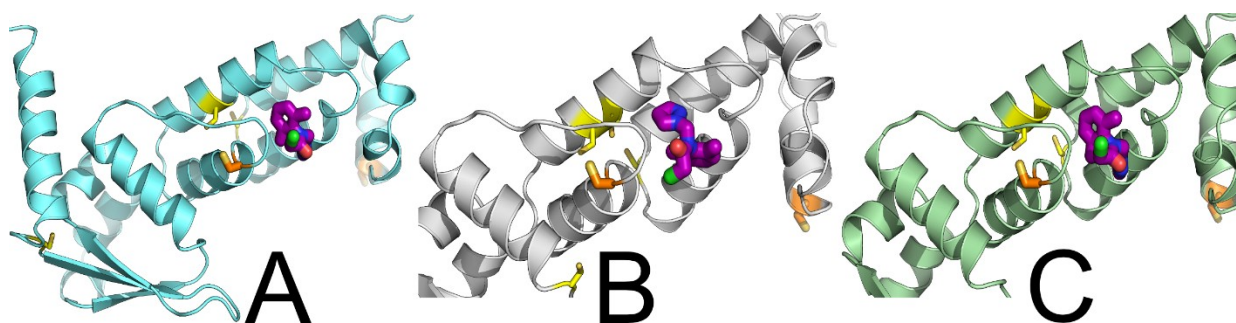


Fig. S16: Molecular docking (MZC). Best-energy poses (for site B) of MZC (light orange ligands) inside the Kelch domain of zfKeap1a (light grey cartoon, -6.4 kcal/mol, A) and zfKeap1b (light green cartoon, -6.5 kcal/mol, B). Canonical cysteine residues are highlighted as orange sticks branching from the protein's backbone, while the remaining cysteines are in yellow. Enumerated cysteine residues 333, 341, and 478 correspond to Cys368 and 513 in hKeap1.



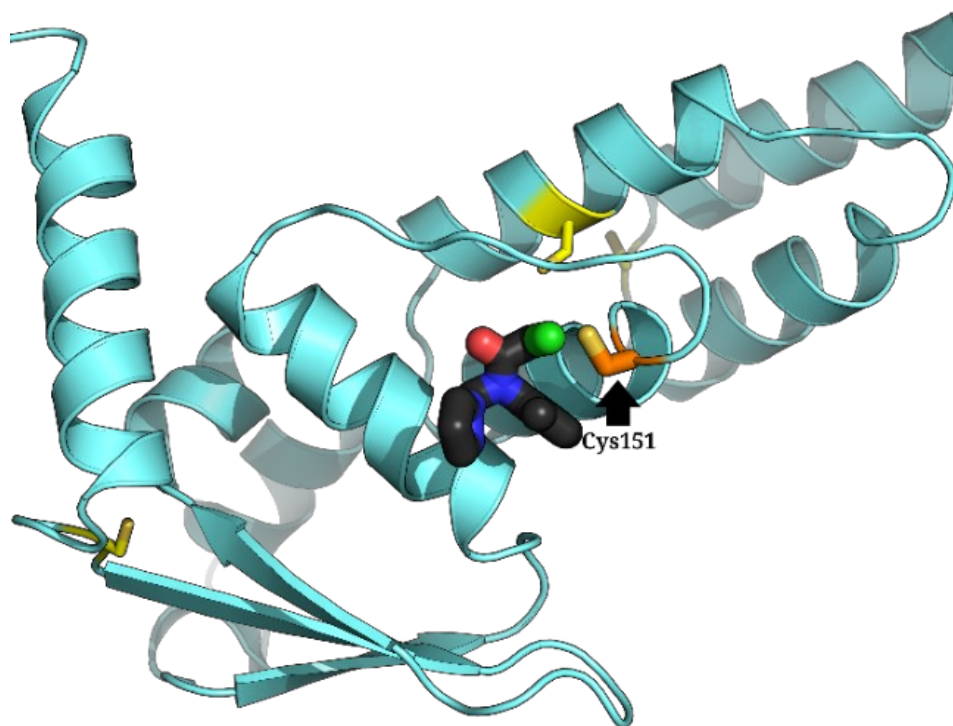
740

741 **Fig. S17: Molecular docking (MZC).** MZC poses (orange ligands) in the third best binding site situated within the BTB domain
 742 (site C) of hKeap1 (cyan cartoon, -5 kcal/mol, A) and zfKeap1b (light green cartoon, -5.1 kcal/mol, B) predicted structures.
 743 Canonical cysteine residues are shown as orange sticks, the remaining ones as yellow lines.



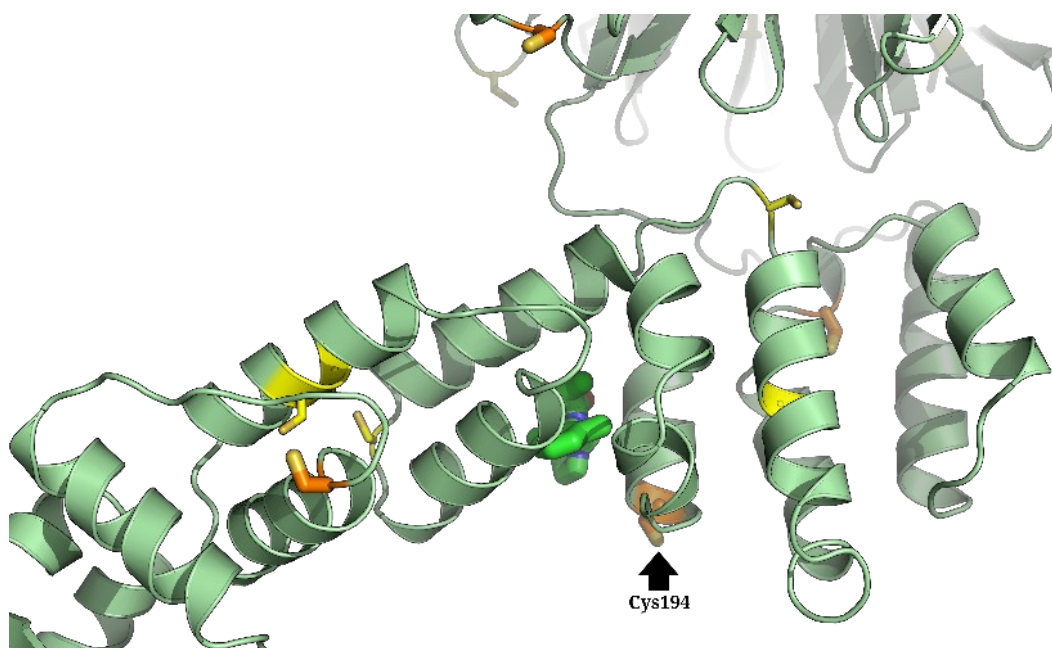
744

745 **Fig. S18: Molecular docking (MZC).** MZC poses (purple ligands) in the fourth best binding site of the alpha-helical region (site D)
 746 of hKeap1 (cyan cartoon, -5 kcal/mol, A), zfKeap1a (light grey cartoon, -5.6 kcal/mol, B), and zfKeap1b (light green cartoon,
 747 -5.1 kcal/mol, C) predicted structures. Canonical cysteine residues are shown as orange sticks, the remaining ones as yellow
 748 lines.



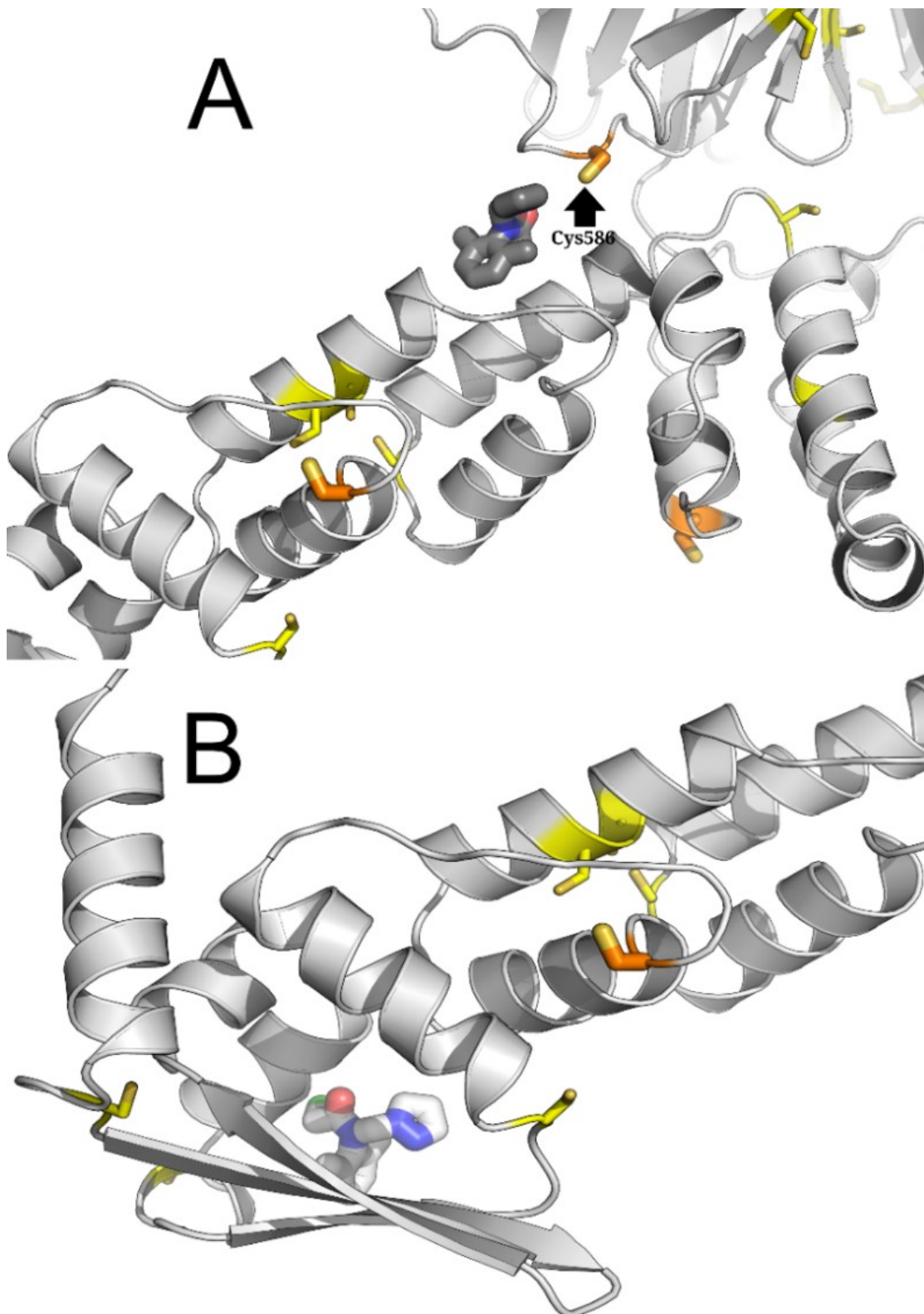
749

750 **Fig. S19: Molecular docking (MZC).** A MZC pose (black ligand) within the fifth best binding site inside the BTB domain (site E,
 751 -4.8 kcal/mol) of hKeap1 predicted structure (cyan cartoon). Canonical cysteine residues are shown as orange sticks, the
 752 remaining ones as yellow lines.



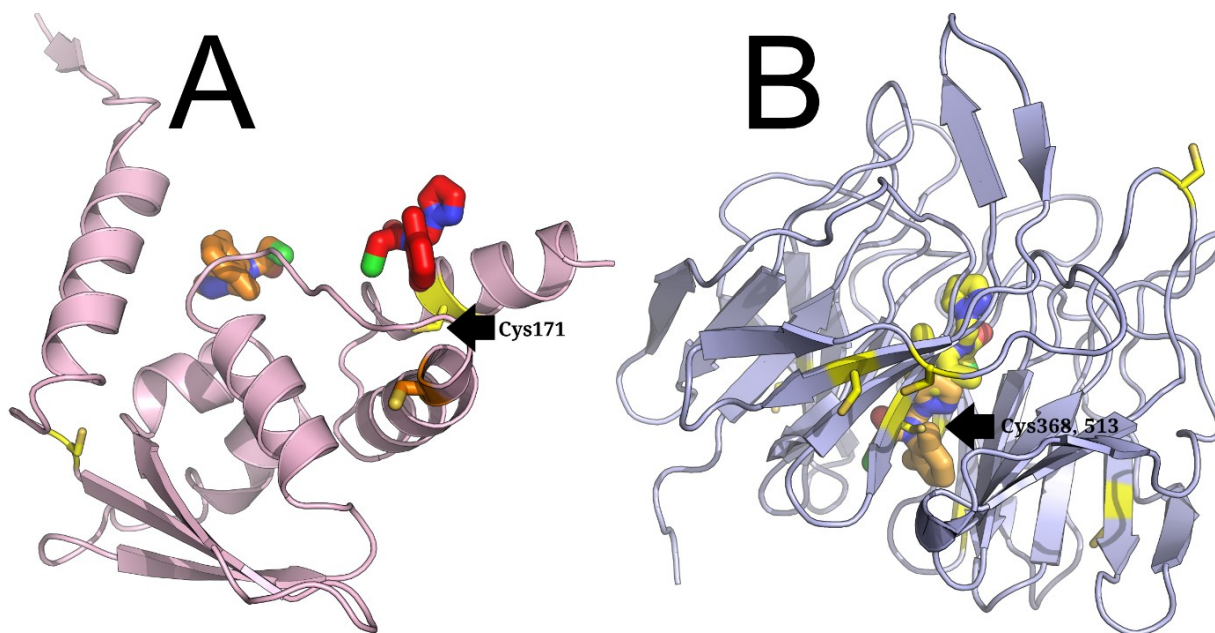
753

754 **Fig. S20: Molecular docking (MZC).** MZC (green ligand) within the best-energy binding site (site F, -5.6 kcal/mol) of the
 755 zfKeap1b IVR domain (light grey cartoon). Canonical cysteine residues are shown as orange sticks, the remaining ones as
 756 yellow lines. The enumerated cysteine 194 corresponds to Cys226 in hKeap1.



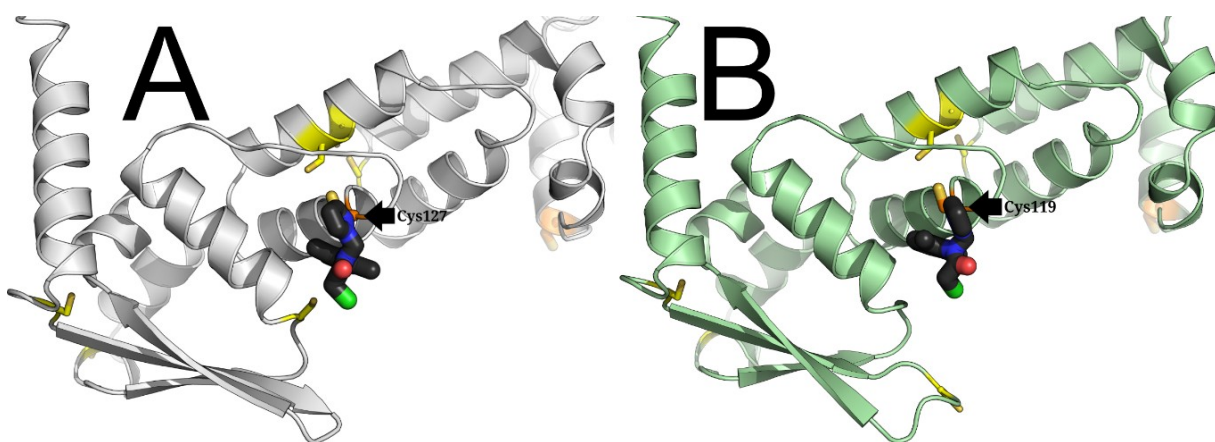
757

758 Fig. S21: Molecular docking (MDC). MZC (coloured ligands) in other binding sites (site G, -5 kcal/mol, A, and site H, -4.9
 759 kcal/mol, B) of zfKeap1a (light grey cartoon). Canonical cysteine residues are shown as orange sticks, the remaining ones as
 760 yellow lines. The enumerated cysteine 588 corresponds to Cys613 in hKeap1



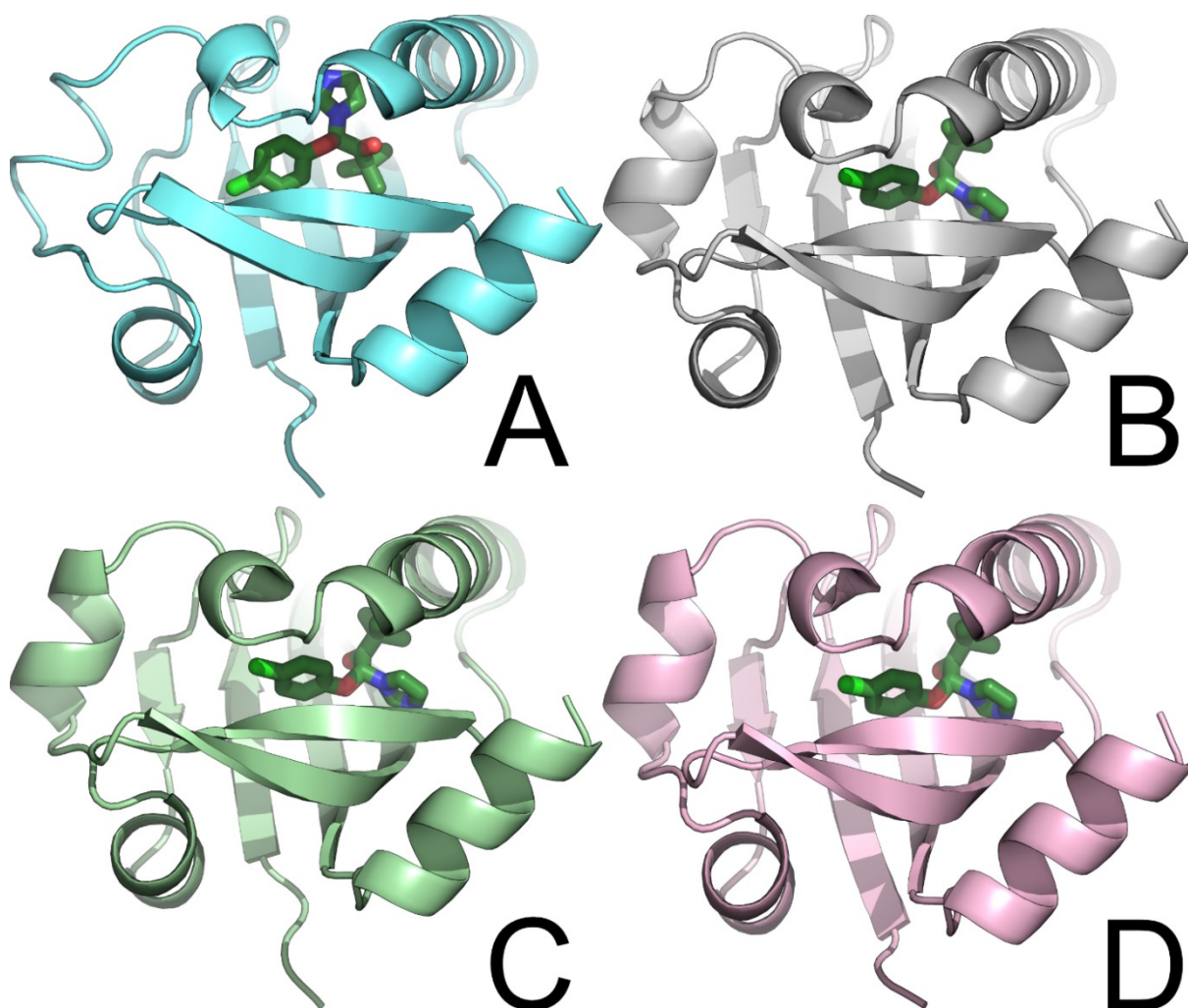
761

762 **Fig. S22: Molecular docking (MZC).** Best energy poses of MZC (coloured ligands) per site from blind docking to X-ray structures
 763 PDB ID 7EXI (light pink cartoon, A) and 1ZGK (light purple cartoon, B). The poses and binding sites are highlighted in colour,
 764 with yellow indicating the best-predicted binding energy (site A, -6.3 kcal/mol) and red the worst (site I, -4.9 kcal/mol).
 765 Canonical cysteine residues are highlighted as orange sticks branching from the protein's backbone, while the remaining
 766 cysteines are in yellow. All illustrative data is summarised quantitatively in tabs. 5 and 6 (DIB_B article).



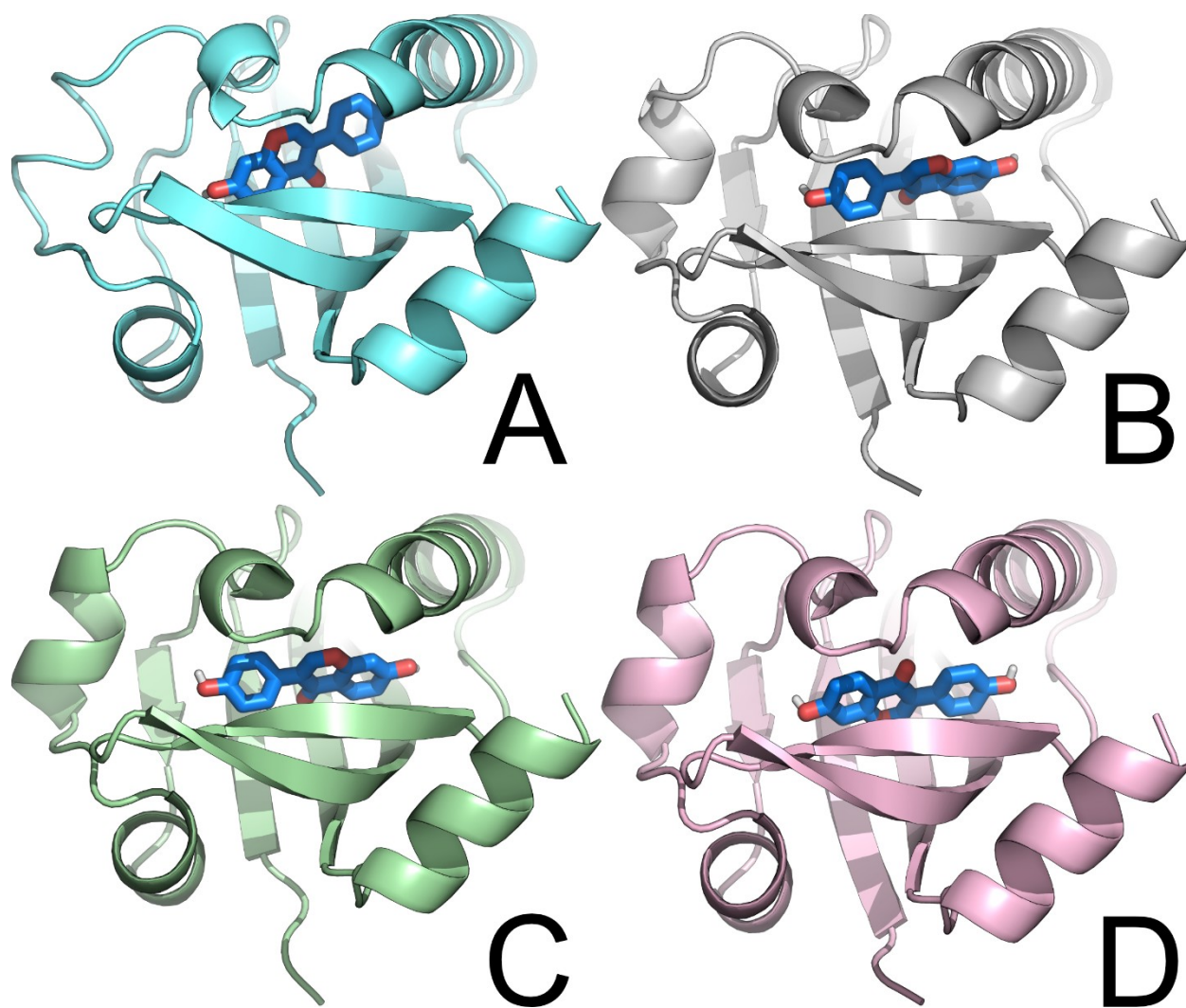
767

768 **Fig. S23: Molecular docking (MZC).** The best-energy poses of MZC (site E, black ligands) from site-directed docking to the
 769 proximity of the canonical Cys151 residue of AlphaFold-predicted structures of zfKeap1a (light grey cartoon, -3.6 kcal/mol, A)
 770 and zfKeap1b (light green cartoon, -4.1 kcal/mol, B). Reactive cysteines 127 and 119 (corresponding to Cys151 in hKeap1) are
 771 shown as orange sticks, while the remaining cysteines are in yellow. All illustrative data is summarised quantitatively in tabs. 5
 772 and 6 (DIB_B article).



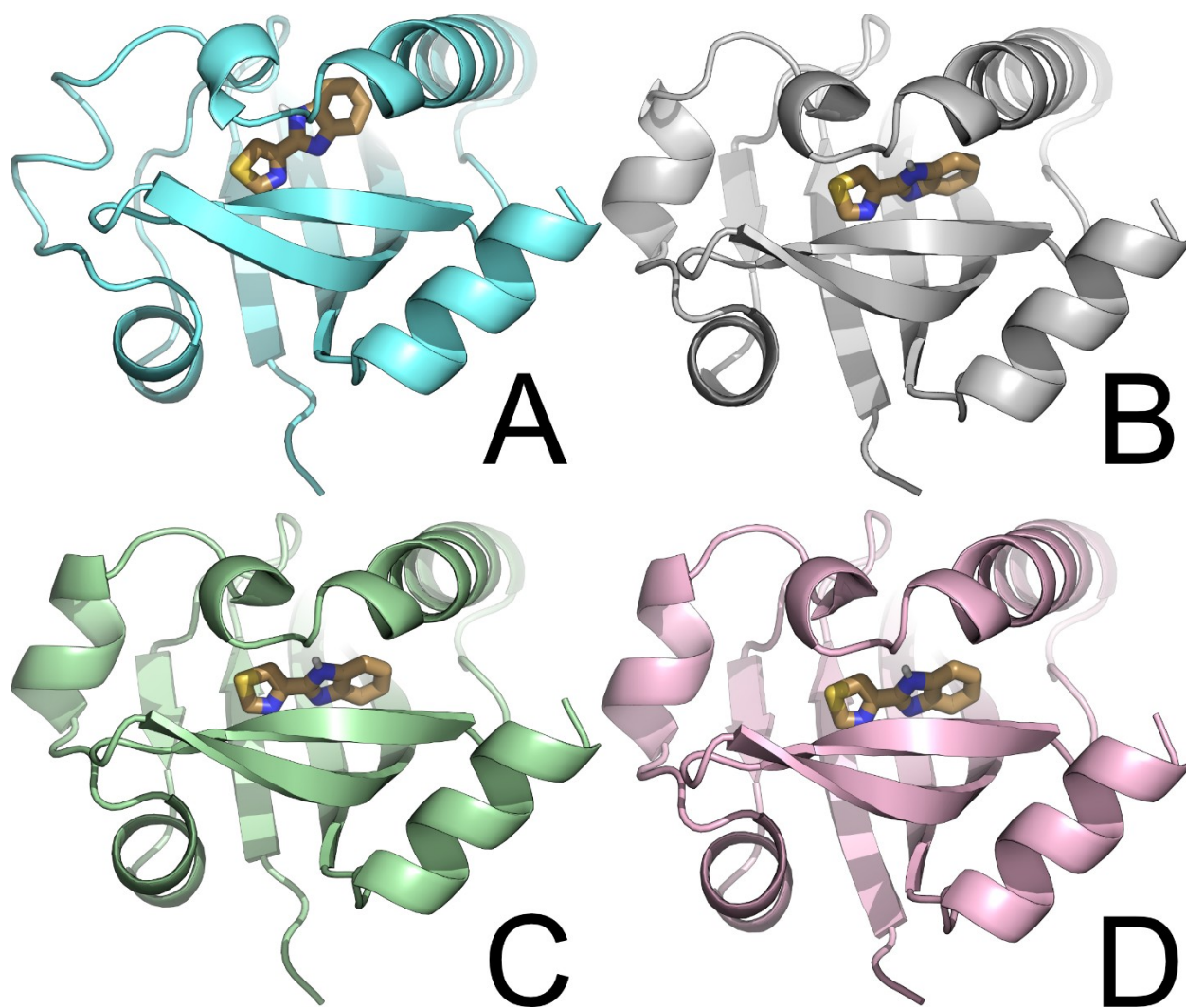
773

774 **Fig. S24: Molecular docking (CLB).** The best-energy poses of climbazole (CLB, green ligands) from site-directed docking to the
 775 PAS-B domain of human (hAhR, A) and zebrafish (zfAhR1a (B), zfAhR1b (C), zfAhR2 (D)) aryl hydrocarbon receptor variants. Up
 776 to ten best poses were computed; the quantitative data are given in tab. 8 (DIB_B article).



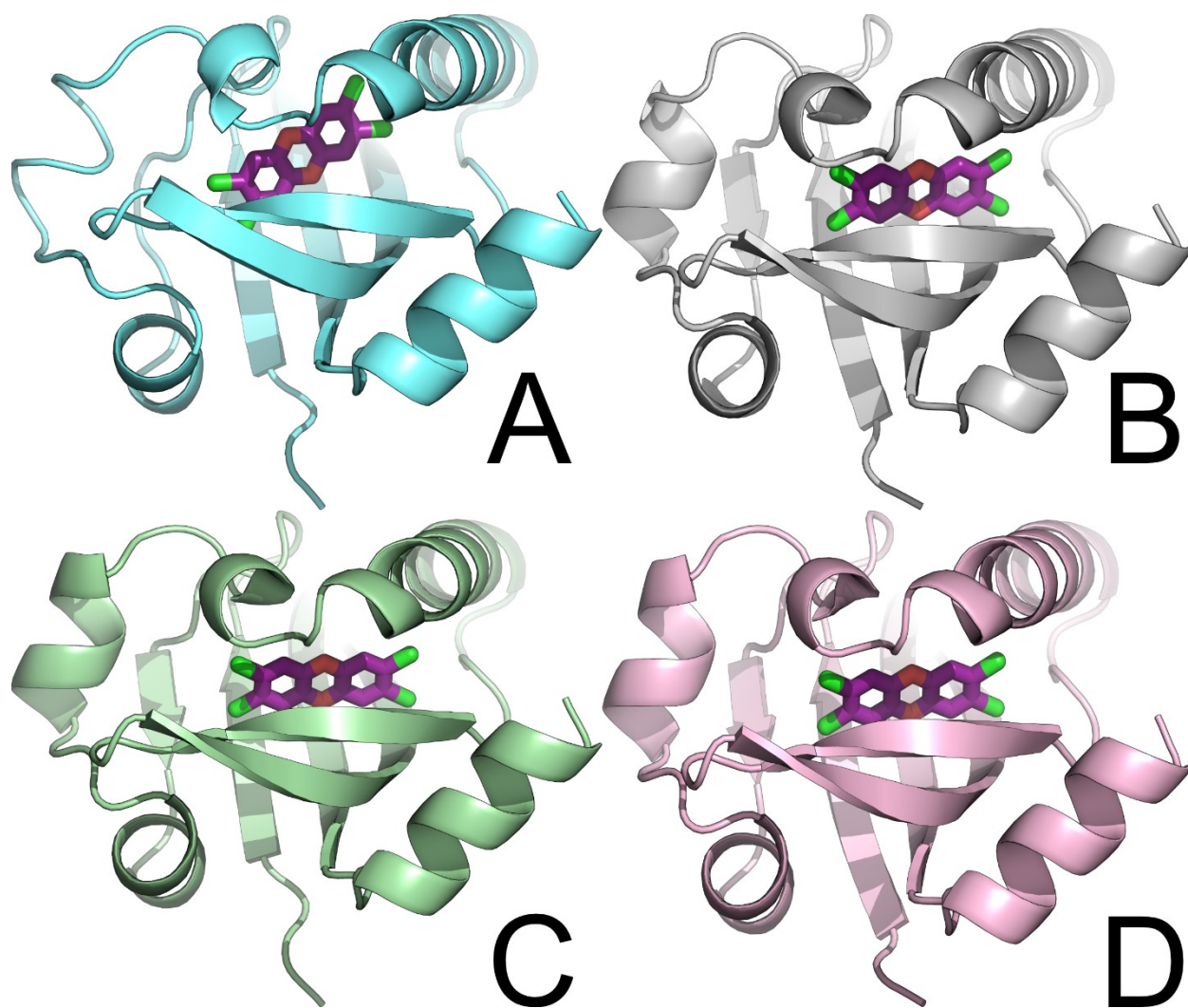
777

778 **Fig. S25: Molecular docking (DAI).** The best-energy poses of daidzein (DAI, blue ligands) from site-directed docking to the PAS-B
 779 domain of human (hAhR, A) and zebrafish (zfAhR1a (B), zfAhR1b (C), zfAhR2 (D)) aryl hydrocarbon receptor variants. Up to ten
 780 best poses were computed; the quantitative data are given in tab. 8 (DIB_B article).



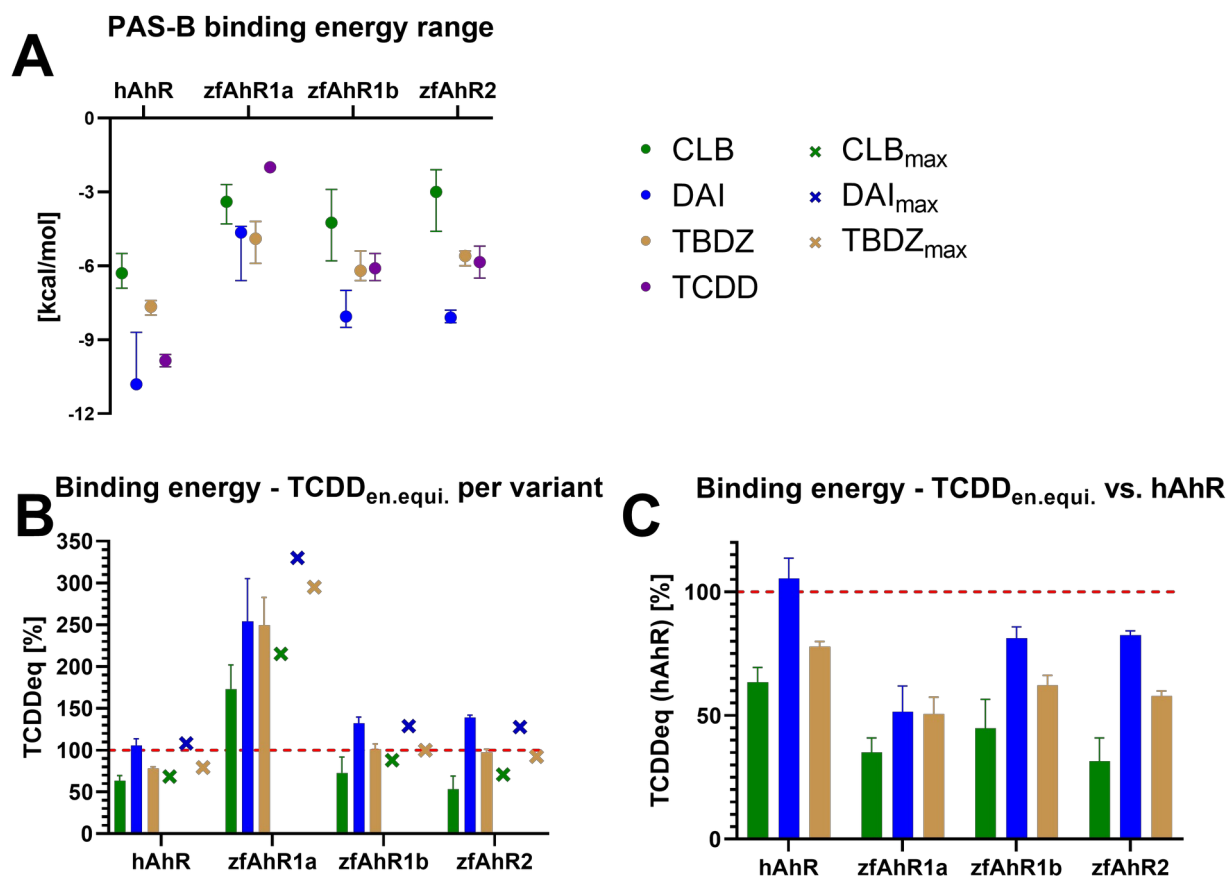
781

782 Fig. S26: Molecular docking (TBDZ). The best-energy poses of thiabendazole (TBDZ, ochre ligands) from site-directed docking
 783 to the PAS-B domain of human (hAhR, A) and zebrafish (zfAhR1a (B), zfAhR1b (C), zfAhR2 (D)) aryl hydrocarbon receptor
 784 variants. Up to ten best poses were computed; the quantitative data are given in tab. 8 (DIB_B article).



785

786 Fig. S27: Molecular docking (TCDD). The best-energy poses of 2,3,7,8-Tetrachlorodibenzodioxin (TCDD, purple ligands) from
 787 site-directed docking to the PAS-B domain of human (hAhR, A) and zebrafish (zfAhR1a (B), zfAhR1b (C), zfAhR2 (D)) aryl
 788 hydrocarbon receptor variants. Up to ten best poses were computed; the quantitative data are given in tab. 8 (DIB_B article).



789

790 Fig. S28: Molecular docking summary. (A) Range of derived binding energies in kcal/mol per compound (Climbazole (CLB),
 791 Daidzein (DAI), Thiabendazole (TBDZ), and 2,3,7,8-tetrachlorodibenzodioxin (TCDD)) for every respective receptor variant.
 792 Dots indicate the medians, and whiskers extend to 10 poses. (B) TCDD binding energy equivalents in per cent, normalised to
 793 the respective variant's mean TCDD binding energy value. Boxes indicate the mean energy equivalents with standard
 794 deviations. Crosses illustrate the equivalent of the best energy pose of the respective compound, normalised to the best
 795 energy pose of TCDD. (C) TCDD binding energy equivalents in per cent, normalised to the mean hAhR TCDD binding energy
 796 value. Boxes indicate the mean energy equivalents with standard deviations. All data and calculations are given in SI11.

3.2 Supplementary results: Main article – Summarised cytotoxicity/cellular viability

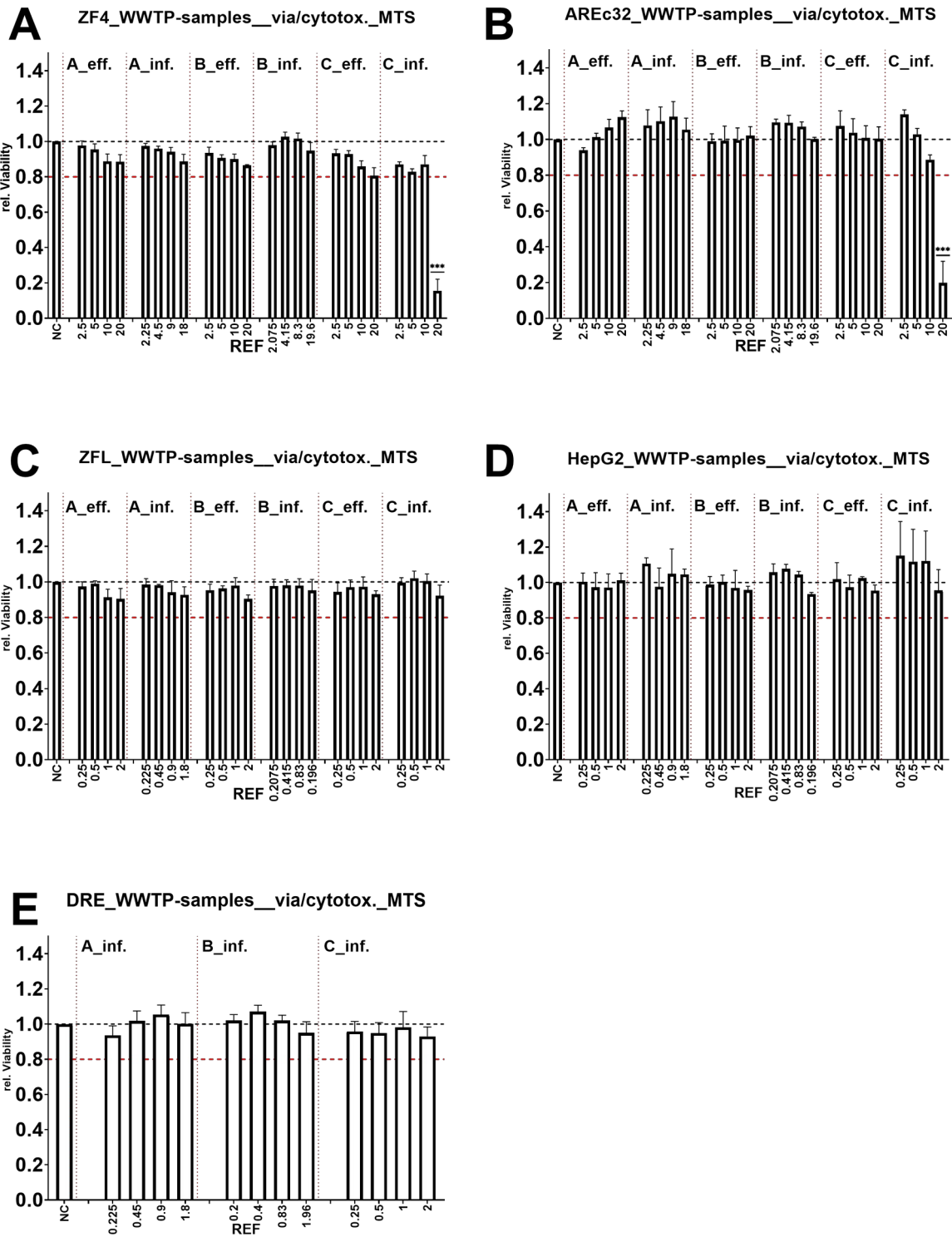


Fig. S29: Cytotoxicity/cellular viability impact assessment of the environmental samples from WWTP sites A to C, conducted in the zebrafish (ZF4 and ZFL), human (MCF7/AREc32 (AREc32) and HepG2), and mouse (DRE) reporter lines. Viability corresponds to control-normalised mitochondrial metabolism, as measured by the MTS assay. Each bar represents the mean, including SEM (experimental units n = 3-4; observational units N = 9-12). All statistical comparisons versus the

negative vehicle control (NC) were conducted using one-sided Dunnett's tests (alpha level = 0.05; ***p<0.001) on pooled raw data (see SI1 to SI8 for raw data and details on calculations).

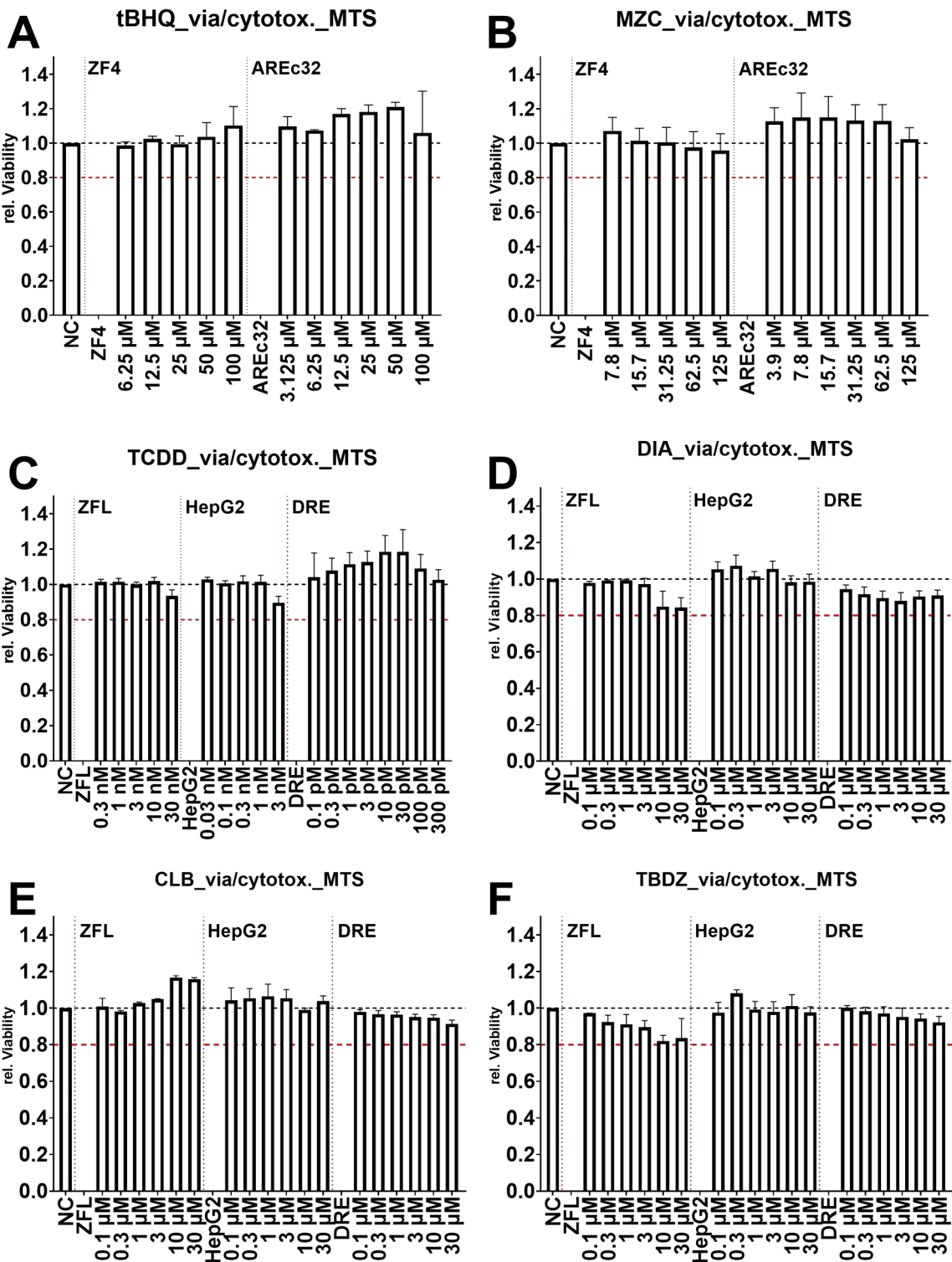


Fig. S30: Cytotoxicity/cellular viability impact assessment of reference control and ISEDA-identified compounds, conducted in the zebrafish (ZF4 and ZFL), human (MCF7AREc32 (AREc32) and HepG2), and mouse (DRE) reporter lines. Viability corresponds to control-normalised mitochondrial metabolism, as measured by the MTS assay. Each bar represents the mean, including SEM (experimental units n = 3-4; observational units N = 9-12). All statistical comparisons versus the

negative vehicle control (NC) were conducted using one-sided Dunnett's tests (alpha level = 0.05) on pooled raw data (see SI1 to SI8, SI14, and SI15 for raw data and details of calculations).

3.3 Supplementary results: Main article – Optimised data fitting strategies for the oxidative stress MoA

In the main article (section 3.1), we recognised that simple linear regression is suboptimal for modelling the CRC relationship of the MZC reference compound due to its hyperbolic profile (fig. S31B). Conversely, a linear model adequately described the tBHQ data (fig. S31A). To achieve an optimal fit for both datasets, we implemented a four-parameter log-logistic nonlinear regression (4PNL) with modifications described by Weimer et al. [51]. The formulation of the 4PNL model is delineated in section 3.4, below.

However, calculating BEQ values from nonlinear regressions is problematic due to variable slope values [55]–[57], potentially leading to significant inaccuracies [57]. Escher et al. recommended estimating parameters only from the initial 30% of the response curve, where the function is approximately linear. This guidance is considered the canonical approach and has been adopted within the main article. Despite this, enhanced parametric flexibility (four parameters and beyond) provided by the 4PNL model could invalidate this assumption. A data-driven, multi-parametric model might be more robust, particularly for the oxidative stress MOA, without a receptor-saturation plateau, where effects likely transition into cytotoxicity. An alternative solution to this challenge is proposed herein, as follows.

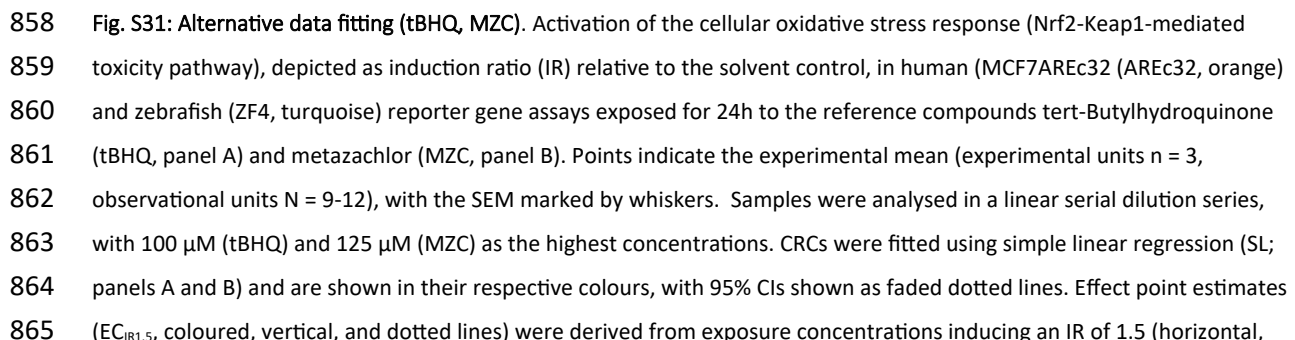
Both tBHQ and MZC reference compound data were fitted to the named 4PNL model (fig. S31C+D). Most statistical software, such as GraphPad Prism or respective R packages, generates a curve consisting of many estimates that can be interrogated and retrieved. We retrieved the 4PNL curve estimates for tBHQ and MZC up to 30% of the maximum response, which showed no cytotoxicity (see SI16, tab “Nrf2-30%fit”), and utilised them to recalibrate a simple linear regression (designated “30%SL” hereafter; see fig. S31E+F).

This revised method rectifies the inadequate fits of the MZC data by linear regression (fig. S31F), maintaining mathematical integrity for BEQ estimations and conserving mechanistic induction discrepancies. While the fit-adjustment processes slightly alter the ratios between zebrafish and human estimates (fig. S33), the ratios for tBHQ consistently exceed unity (4PNL: 1 to 2; 30%SL: 1.6 to 3.4), suggesting mechanistic discrepancies. Moreover, differences in response to tBHQ are less noticeable at an induction level of 1.5 after normalisation to vehicle control.

For the main article, all non-cytotoxic concentrations were retained for regression fitting, provided that the corresponding concentration-response data exhibited approximately semi-linear behaviour

844 within the tested exposure range. Thus, no fixed IR4/IR5 (induction ratio 4 or 5) cut-off was applied
845 to the oxidative stress MoA datasets. Instead, cytotoxicity was assessed in parallel (see above), and
846 cytotoxic concentrations were excluded from subsequent analyses, consistent with recent AREc32-
847 based applications [58]. Similarly, for the xenobiotic metabolism MoA, concentrations exceeding 30%
848 of the maximum TCDD-induced response were not automatically excluded if the corresponding
849 concentration-response relationship remained approximately semi-linear and non-cytotoxic within
850 the tested range (one incidence, influent sample C at REF 2 in the HepG2 assay).

851 This approach was chosen because the study's primary aim was to compare mechanistic response
852 patterns across species and assays, rather than to optimise each individual concentration-response
853 relationship using a canonical fitting procedure. Retaining all valid, non-cytotoxic data points
854 maintained a consistent data-analysis workflow across samples and assays, avoiding the need for
855 sample- and assay-specific exclusion criteria that would require separate downstream statistical
856 procedures.



866 dotted line). The respective values are depicted in the graphs. A two-way ANOVA was conducted to compute differences
867 between effect sizes of the respective reporter line pairs (ns = statistically non-significant, ***p < 0.001). The respective
868 samples' cellular viability/cytotoxicity data are given in the SM, figs. S29 and S30, and detailed in DIB_A. The respective raw
869 and metadata are provided in SI1, SI2, and SI14. Alternatively, CRCs were fitted to a non-linear, log-logistic regression model
870 (4PNL, panels C and D; see equations in section 3.4 below). EC_{IR5} estimates were also derived. Total CRC estimates from the
871 4PNL model were retrieved up to the point that induced 30% of the maximum effect, and the model was refitted using a
872 simple linear regression (30%SL model; panels E and F; see sections 3.3 and 3.4 below for details).

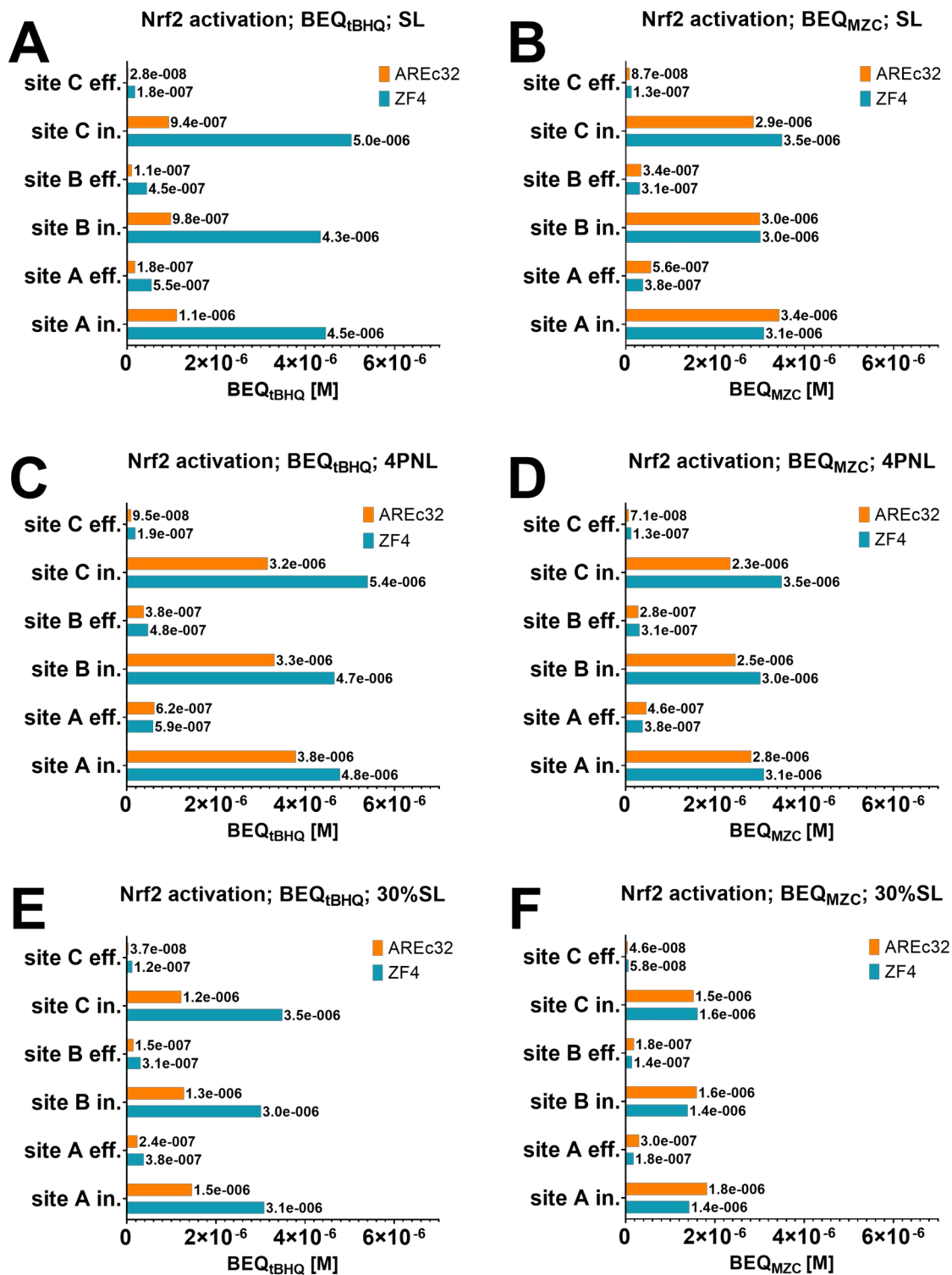
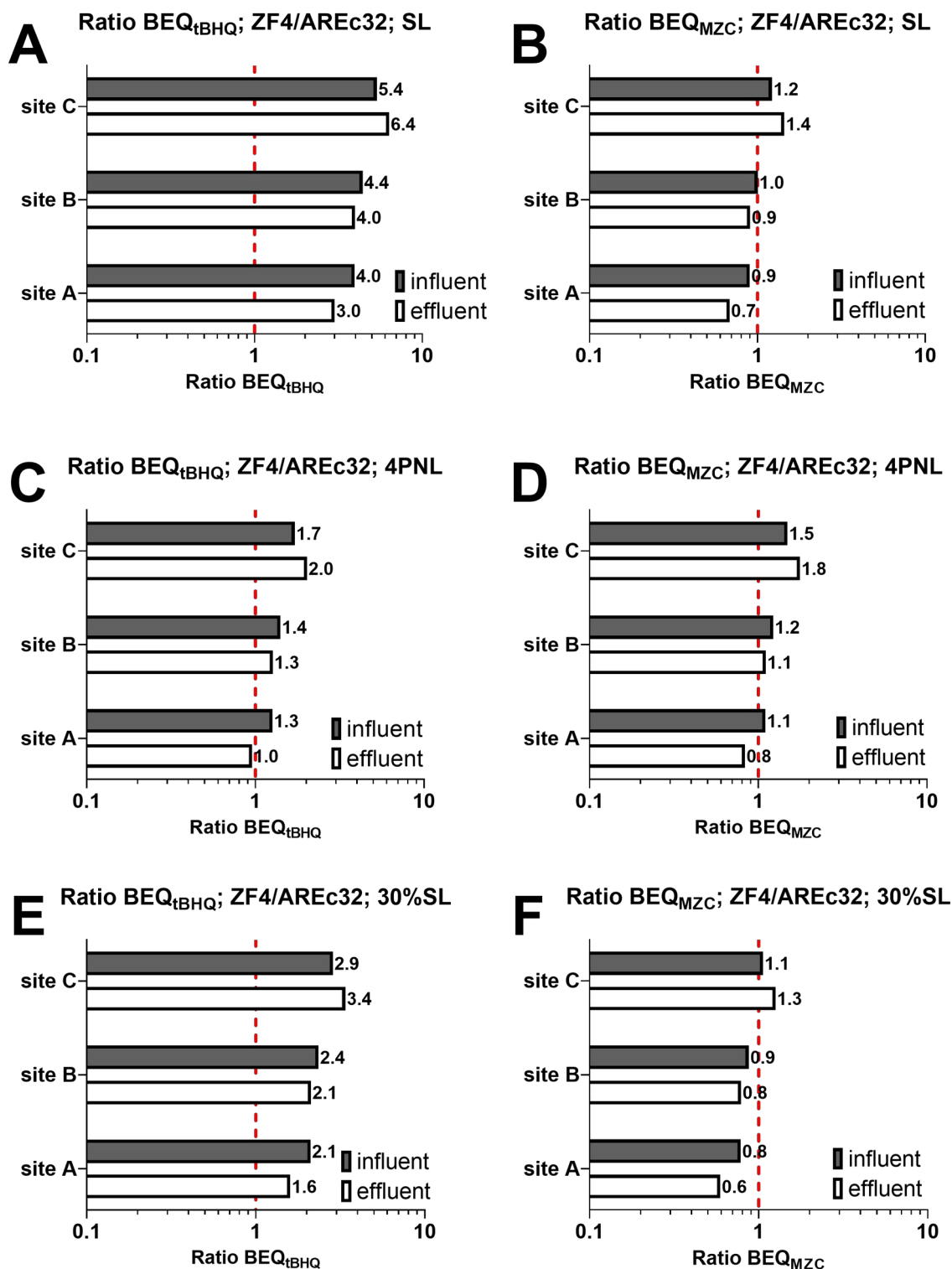


Fig. S32: Alternative data fitting (tBHQ, MZC). Outline of derived tBHQ-BEQ standardisations (panels A, C, and E) and MZC-BEQ standardisations (panels B, D, and F), regarding the used regression model (SL, 4PNL, or 30%SL, see sections 3.3 and 3.4 below).



877

878 Fig. S33: Alternative data fitting (tBHQ, MZC). Outline of derived zebrafish-to-human BEQ ratios, regarding the used
 879 regression model: SL (panels A and B), 4PNL (C and D), and 30%SL (E and F); see details in sections 3.3 and 3.4 below. The
 880 unity ratio is illustrated as a dotted red line.

3.4 Supplementary results: Main article – Optimised data fitting strategies for the xenobiotic metabolism MoA

Similar to section 3.3 above, we explored several multi-parameter nonlinear regression models for data fitting. Direct interpolation of BEQ values from sigmoidal curves with variable slopes is deemed incorrect [49], [56], [57], a scenario particularly likely for environmental samples. The alternative is to fit a simple linear regression to the environmental data, given that the response is unlikely to exceed a high percentage increase relative to the utilised reference compound's maximum response. A point estimate from the CRC within the initial 30% is selected, where the response exhibits a quasi-linear trend. Nevertheless, the choice of model and the nature of the empirical data ultimately determine the CRC's form, and the above scenario might not always hold. We fitted xenobiotic metabolism MoA data to elaborate on our point, as delineated in the main article, with the following models:

2PNL:

$$f(x) = \min \frac{+max - min}{1 + 10^{(logEC50 - x)}}$$

With upper (max.) and lower (min) asymptotes defined by normalisation to vehicle controls (min.) and maximum TCDD response (max.)

3PNL:

$$f(x) = \min \frac{+max - min}{1 + 10^{HillSlope * (logEC50 - x)}}$$

With upper (max.) and lower (min) asymptotes defined by normalisation to vehicle controls (min.) and maximum TCDD response (max.). As such, the model aligns with the recommendations made by Escher et al [50].

4PNL:

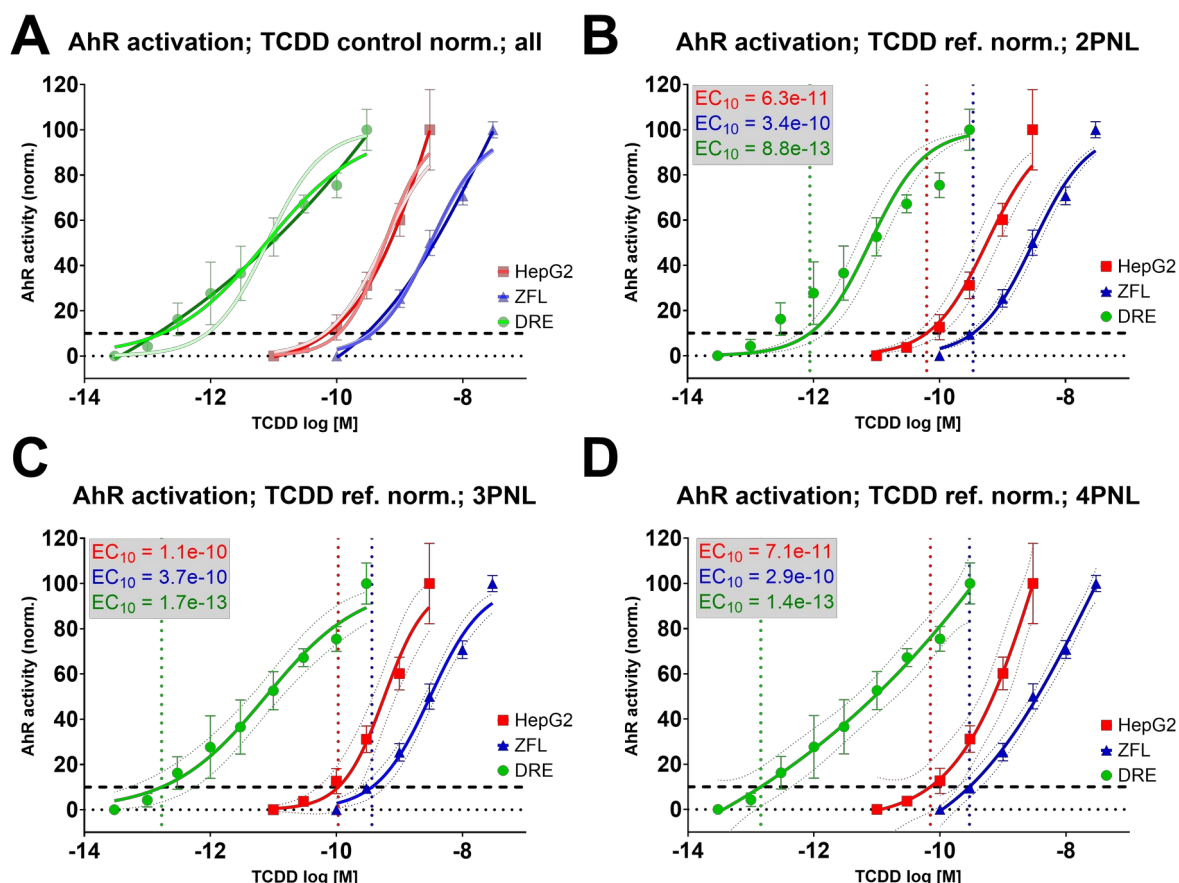
$$f(x) = d \frac{+a - d}{1 + 10^{HillSlope * (logEC50 - x)}}$$

With upper (a) and lower (d) asymptotes refitted to the data by the model, as recommended by Weimer et al. [51].

30%SL: Modelled according to section 3.3, using the first 30% of data from the 3PNL model.

The inadequacy of a simple linear regression for deriving EC₁₀ values for BEQ calculation is evident from the fig. S36B, D, and F, necessitating a sigmoidal fit. Application of the 2PNL, 3PNL, and 4PNL models to ZFL and HepG2 data yielded satisfactory fits for all models (fig. S34). However, the 2PNL

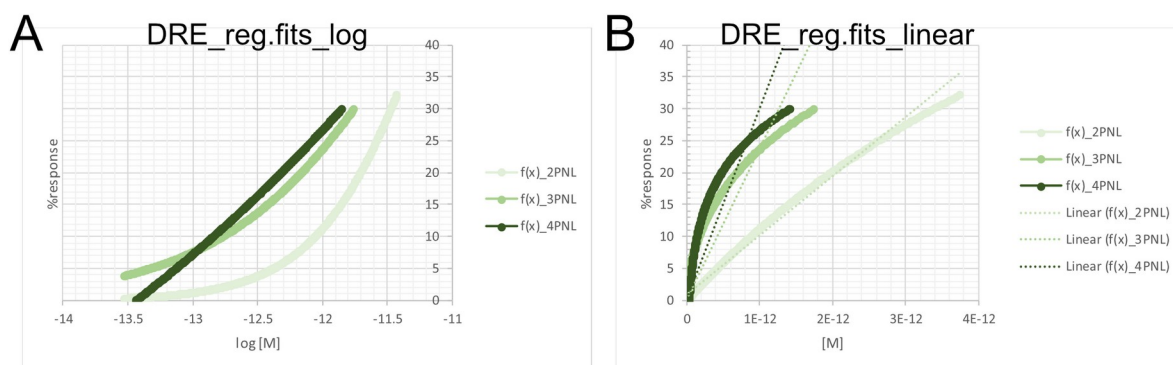
909 model inaccurately estimates the upper and lower bounds of the empirical curve for DRE data (fig.
 910 S34B), leading to a substantial underestimation of EC_{10} . For example, the EC_{10} estimate in question is
 911 misjudged by almost half an order of magnitude.



912

913 Fig. S34: Alternative data fitting (TCDD). Activation of the cellular xenobiotic metabolism (AhR-ARNT-mediated toxicity
 914 pathway), depicted as a minimum (solvent control background) to maximum TCDD-induced response normalisation, in
 915 human (HepG2, red), zebrafish (ZFL, blue), and mouse (DRE, green) reporter gene assays exposed for 24 h to the reference
 916 compound TCDD. Points indicate the experimental mean (experimental units $n = 3-5$, observational units $N = 9-22$), with the
 917 SEM marked by whiskers. The data were analysed via various non-linear, log-logarithmic regression models (panel B: 2PNL,
 918 panel C: 3PNL, panel D: 4PNL, panel A: merge of the latter) and are given in their respective colours, with 95% CIs as faded,
 919 dotted lines. Effect point estimates (EC_{10} ; coloured, vertical, and dotted lines) were derived from the concentrations that
 920 cause 10% of the maximum response (horizontal dotted line). The respective values are depicted within the graphs.

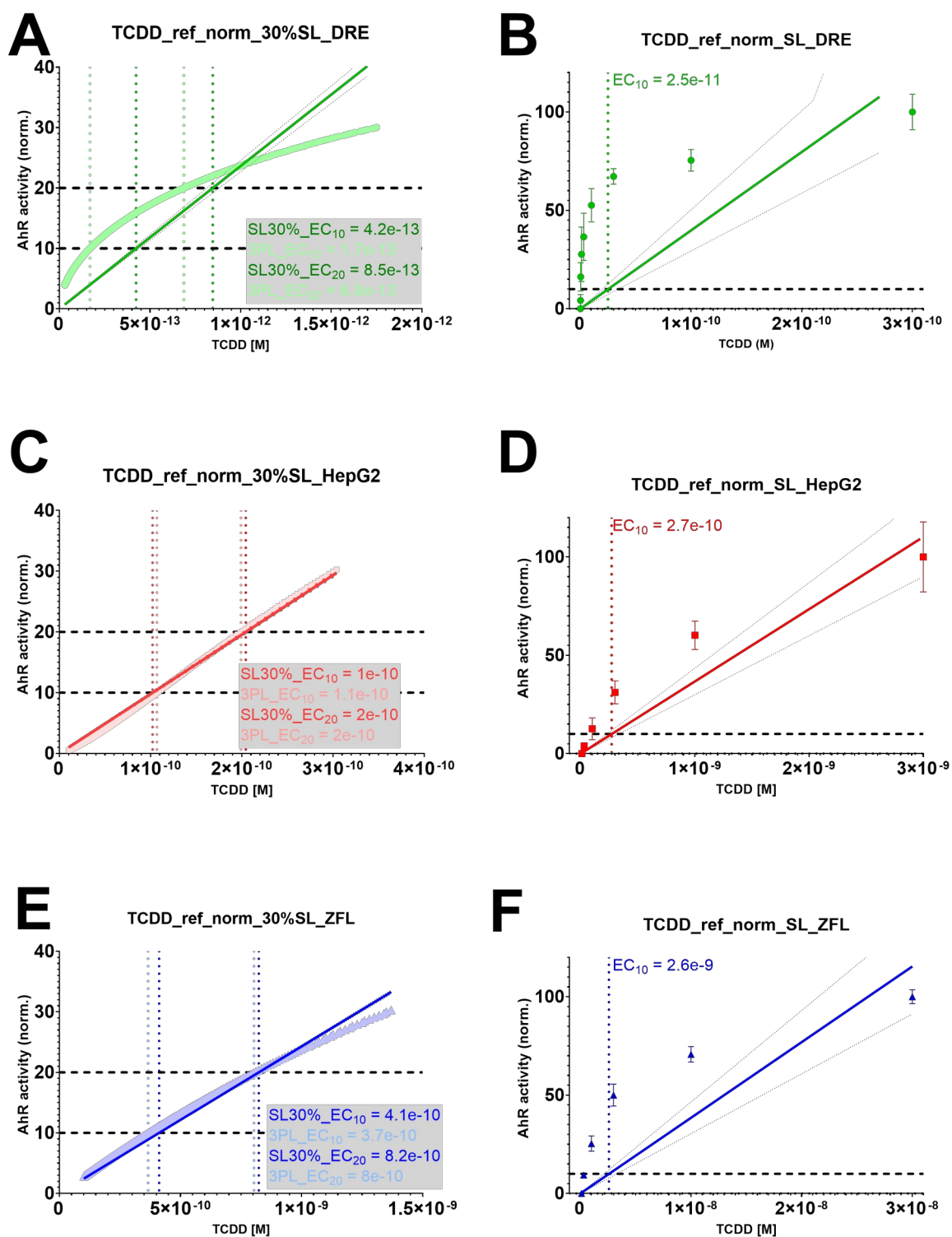
921 Examining the lower 30% of the DRE reporter line response across 2PNL, 3PNL, and 4PNL models on
 922 logarithmic and linear scales revealed differing behaviours within these bounds (fig. S35). Only the
 923 2PNL model exhibited a semi-linear characteristic. At the same time, the 3PNL and 4PNL fits were
 924 relatively hyperbolic (fig. S35B), presenting a dilemma between choosing a suboptimal 2PNL fit or
 925 disregarding the linearity assumption for BEQ extrapolation. Instead, we adopted the 30%-refitting
 926 approach from the 3PNL data, indicating a near-linear response up to a 10% threshold.



927

928 **Fig. S35: Alternative data fitting (TCDD).** The lower 30% of the 2PNL, 3PNL, and 4PNL model fits; displayed on a logarithmic
 929 (panel A) and linear (panel B) scale. The data were taken from the DRE assay results.

930 Figures S36C and E show that the 30%SL refitting yields estimates comparable to those of the 3PNL
 931 model. While unnecessary in this instance, as the data already exhibited a semi-linear character, it
 932 suggests that this approach does not introduce distortion. Although not ideal, the refitting of DRE
 933 data (fig. S36A) reduces the error inherent in the 2PNL model by more than half and closely matches
 934 the 3PNL at EC_{20} (see fig. S36A).



935

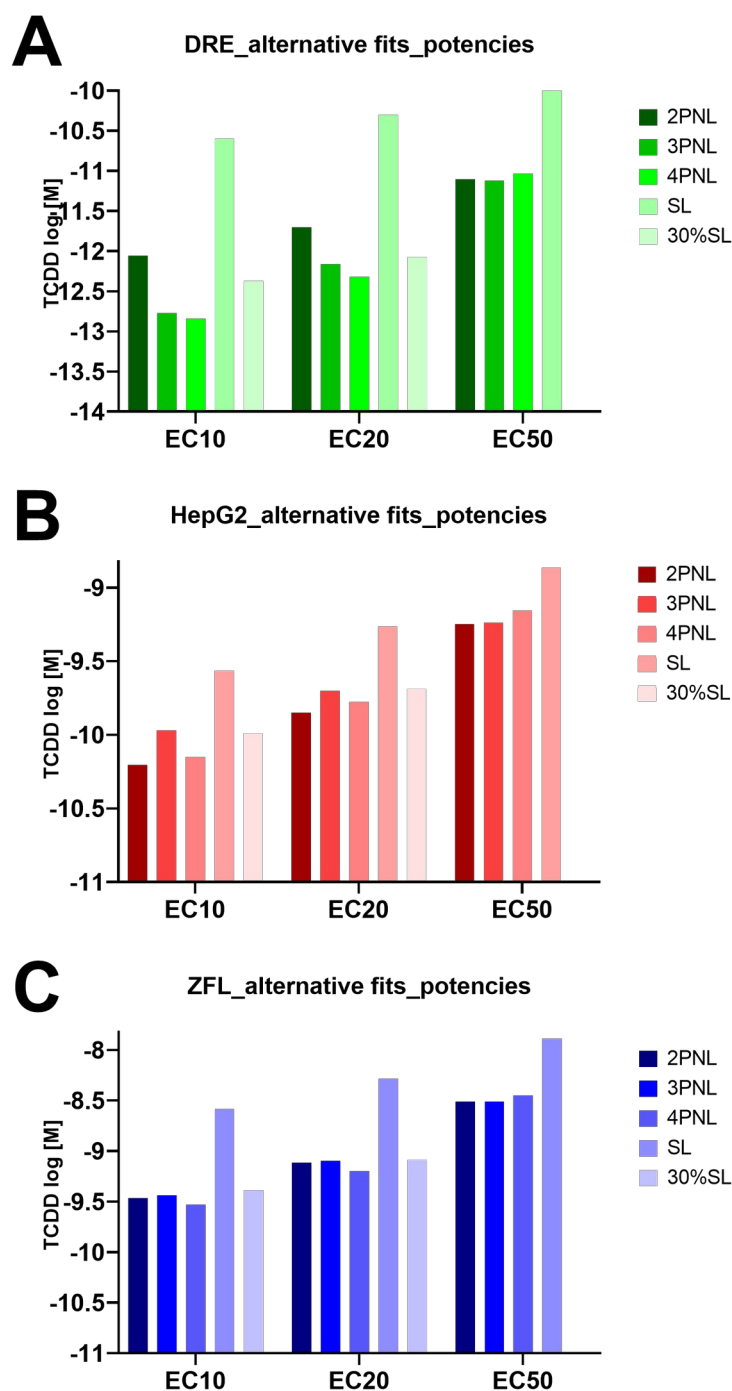
936 Fig. S36: Alternative data fitting (TCDD). Activation of the cellular xenobiotic metabolism (AhR-ARNT-mediated toxicity
 937 pathway), depicted as a minimum (solvent control background) to maximum TCDD-induced response normalisation, in
 938 human (HepG2, red), zebrafish (ZFL, blue), and mouse (DRE, green) reporter gene assays exposed for 24 h to the reference
 939 compound TCDD. Points indicate the experimental mean (experimental units n = 3-5, observational units N = 9-22), with the
 940 SEM marked by whiskers. CRCs were fitted using simple linear regression (panels B, D, and E) and are shown in their
 941 respective colours, with 95% CIs shown as faded dotted lines. Alternatively, total curve estimates derived from the 3PNL

942 model were refitted to a simple linear regression (panels A, C, and E). Effect point estimates (EC_{10} and EC_{20} ; coloured,
 943 vertical, and dotted lines) were derived from the concentrations that cause 10 or 20% of the maximum response (horizontal
 944 dotted line).

Tab. S4: Effect estimates (EC_{10} or EC_{20}) ratios, as derived from various DRE data model fits.	
Pair	Ratio
2PNL/3PNL	5.2
2PNL/4PNL	6.3
2PNL/30%SL	2.1
30%SL/3PNL	2.5
30%SL/4PNL	3
30%SL/3PNL at EC_{20}	1.2

945

946 Comparing different model-derived CRC estimates (EC_{10} , EC_{20} , and EC_{50}), the 30%-refitting method appears
 947 consistent with standard sigmoidal fits (fig. S37), both when data naturally aligns with semi-linearity
 948 (fig. S37B+C) and when the sigmoidal curve inaccurately characterises the empirical data or breaches
 949 the semi-linearity assumption (fig. S37A). Therefore, this novel strategy merits consideration as a
 950 standard evaluative procedure, given its safeguard against misinterpretation and its preservation of
 951 linearity. Nonetheless, more extensive, additional validation is required before drawing definitive
 952 conclusions about its broader applicability.



953

954 Fig. S37: Alternative data fitting (TCDD). Comparison of point-effect estimate performance among the indicated models. The
 955 applied regression models are described in section 3.4.

956 Subsequent meta-analyses mirrored those in the main article, converting model estimates to TCDD-
 957 BEQ values and comparing zebrafish and mammalian estimates and ratios (figs. S38 and S39). A
 958 simple linear regression (SL) distorts the data, exaggerating differences between ZFL and HepG2
 959 while underestimating those between ZFL and DRE (fig. S39A). Except for the site C influent, all
 960 nonlinear regression models (2PNL, 3PNL, 4PNL) predict a half- to an entire-order-of-magnitude
 961 difference in TCDD-BEQs between the zebrafish and the human reporter line. We conclude that,

962 except for the site C influent sample, the TCDD-BEQ standardisation procedure does not sufficiently
963 account for differences in species sensitivity.

964 Zebrafish-to-mouse estimates and potency differences depend on the chosen model. For 2PNL, the
965 observed ratios were the lowest, ranging from 319 (site A influent) to 784 (site B influent); keeping in
966 mind that the respective estimate was derived from a semi-linear curve but insufficiently
967 represented empirical data (fig. S39B). Both 3PNL and 4PNL yielded satisfactory fits yet breached the
968 semi-linearity premise, resulting in a threefold disparity in potencies (3PNL ratios: 1773 to 4359;
969 4PNL: 1675 to 4118; fig. S39C+D). The 30%SL adjustment aligns potencies more closely, although a
970 significant magnitude difference persists (ratios 795 to 1952, fig. S39E). In conclusion, for
971 environmental mixture effects, particularly between zebrafish (ZFL) and mouse (DRE) reporter lines,
972 the TCDD-BEQ normalisation does not adequately correct for species-specific differences in
973 sensitivity.

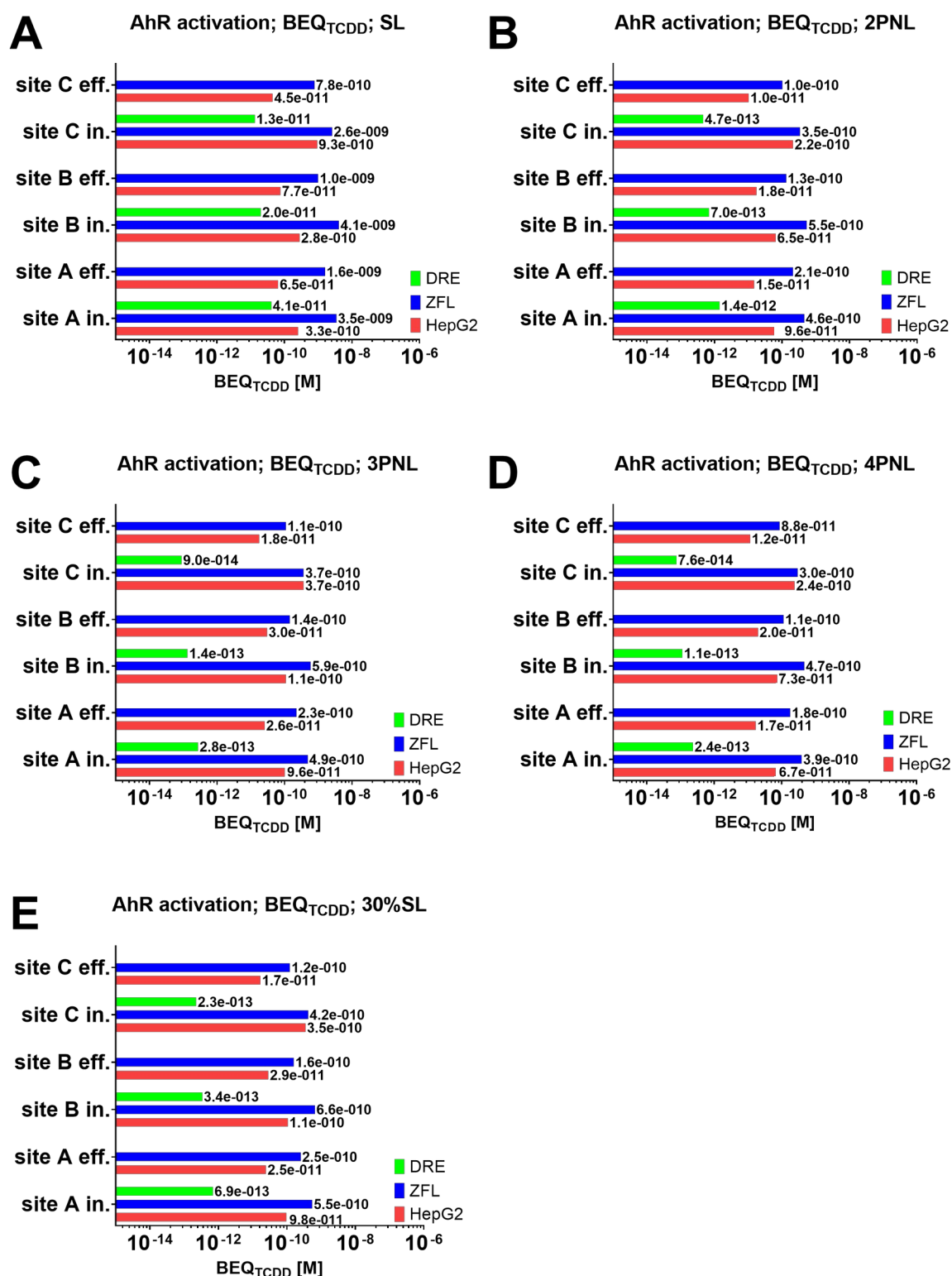


Fig. S38: Alternative data fitting (TCDD). Outline of derived TCDD-BEQ standardisations, regarding the used regression model (panel A: SL, panel B: 2PNL, panel C: 3PNL, panel D: 4PNL, and panel E: 30%SL, see details in section 3.4).

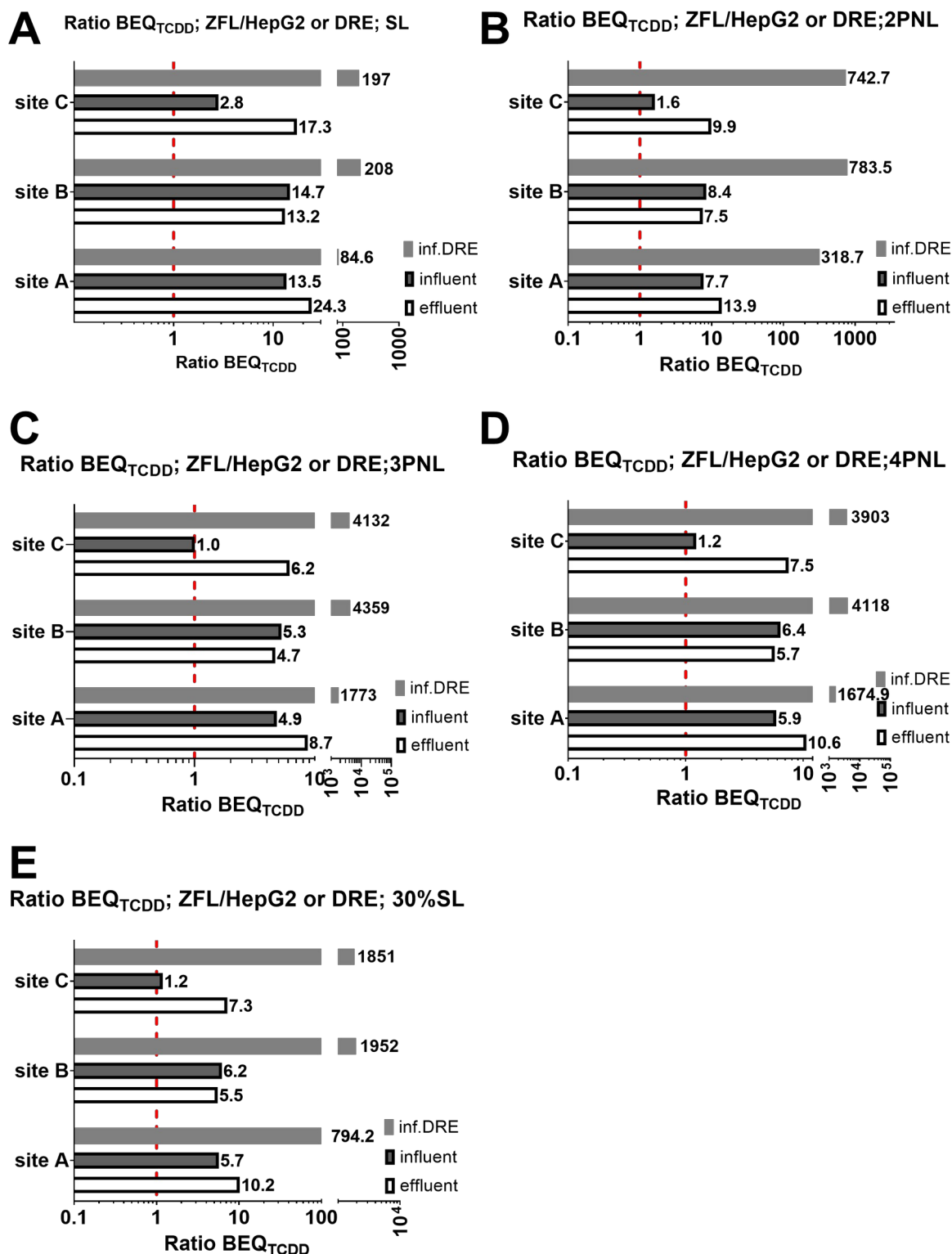


Fig. S39: Alternative data fitting (TCDD). Outline of derived zebrafish-to-mammal BEQ ratios, regarding the used regression model: SL (panel A), 2PNL (panel B), 3PNL (panel C), 4PNL (panel D), and 30%SL (panel E), see details in section 3.4. The unity ratio is illustrated as a dotted red line.

3.5 Supplementary results: Main article – other supplementary results

Tab. S5: Summary of all effect point estimates ($EC_{IR1.5}$), their variation metrics, calculated BEQ values,

and chemical removal efficiencies, as derived from the oxidative stress MoA assessment (Ctrl = reference control, inf = influent, eff = effluent, end = endpoint, hu = human, zf = zebrafish, SE = standard error, CIs = confidence intervals, Ch. = removal efficiency from chemical assessment data (SI12, tab. "REP_BEQ_calculations") given in parentheses). Effect concentration estimates computed beyond the empirically tested exposure concentrations, as well as consequential BEQ values and biological removal efficiency, are marked with a double-cross (‡).

Site	Sample	Spec	Cell line	EC _{IR1.5} [REF] [M] Including SE	95% CIs [REF] [M]	BEQ_tBHQ [M]	BEQ_MZC [M]	Removal [%] (Ch.)
A	inf	hu	AREc32	1.35 ± 0.15	1.02;1.68	1.12E-06	3.43E-06	
A	eff	hu	AREc32	8.27 ± 1.45	5.08;11.45	1.83E-07	5.60E-07	84 (73)
A	inf	zf	ZF4	1.39 ± 0.10	1.18;1.61	4.45E-06	3.09E-06	
A	eff	zf	ZF4	11.25 ± 2.53	5.68;16.82	5.49E-07	3.82E-07	88 (73)
B	inf	hu	AREc32	1.54 ± 0.07	1.40;1.69	9.83E-07	3.00E-06	
B	eff	hu	AREc32	13.48 ± 4.00	4.70;22.28	1.12E-07	3.44E-07	89 (85)
B	inf	zf	ZF4	1.43 ± 0.14	1.13;1.73	4.34E-06	3.02E-06	
B	eff	zf	ZF4	13.82 ± 3.72	5.65;22.00	4.47E-07	3.11E-07	90 (85)
C	inf	hu	AREc32	1.62 ± 0.30	0.93;2.30	9.38E-07	2.87E-06	
C	eff	hu	AREc32	53.48 ± 48.32 [‡]	-52.9;160	2.83E-08 [‡]	8.66E-08 [‡]	97 [‡] (92)
C	inf	zf	ZF4	1.23 ± 0.14	0.90;1.56	5.03E-06	3.50E-06	
C	eff	zf	ZF4	34.35 ± 21.59 [‡]	-13.2;81.9	1.80E-07 [‡]	1.25E-07 [‡]	96 [‡] (92)
Ctrl	tBHQ	hu	AREc32	1.52E-06 ± 5.95E-08	1.39E-06; 1.64E-06			
Ctrl	tBHQ	zf	ZF4	6.18E-06 ± 5.55E-07	4.99E-06; 7.37E-06			
Ctrl	MZC	hu	AREc32	4.63E-06 ± 4.01E-07	3.78E-06; 5.48E-06			
Ctrl	MZC	zf	ZF4	4.30E-06 ± 6.12E-07	2.99E-06; 5.61E-06			

982

Tab. S6: Summary of all effect point estimates (EC₁₀), calculated BEQ values, and chemical removal efficiencies, as derived from the xenobiotic metabolism MoA assessment (Ctrl = reference control, inf = influent, eff = effluent, end = endpoint, hu = human, zf = zebrafish, m = mouse, SE = standard error, CIs = confidence intervals, Ch. = removal efficiency from chemical assessment data (SI12, tab. "REP_BEQ_calculations") given in parentheses). Effect concentration estimates computed beyond the empirically tested exposure concentrations, as well as consequential BEQ values and biological removal efficiency, are marked with a double-cross (‡).

Site	Sample	Spec.	Cell line	EC ₁₀ [REF] [M]	95% CIs [REF] [M]	BEQ_TCDD	Removal
------	--------	-------	-----------	----------------------------	-------------------	----------	---------

				Including SE		[M]	[%] (Ch.)
A	inf	hu	HepG2	1.06 ± 0.19	0.76; 1.72	1.01E-10	
A	eff	hu	HepG2	4.13 ± 0.45 [†]	3.14; 5.11	2.59E-11 [†]	74 [†] (73)
A	inf	zf	ZFL	0.75 ± 0.08	0.57; 0.93	4.90E-10	
A	eff	zf	ZFL	1.62 ± 0.23	1.11; 2.14	2.26E-10	54 (73)
A	inf	m	DRE	0.61 ± 0.07	0.46; 0.76	2.77E-13	
B	inf	hu	HepG2	0.96 ± 0.13	0.68; 1.24	1.11E-10	
B	eff	hu	HepG2	3.53 ± 0.73 [†]	1.91; 5.15	3.03E-11 [†]	73 [†] (85)
B	inf	zf	ZFL	0.62 ± 0.05	0.52; 0.72	5.88E-10	
B	eff	zf	ZFL	2.57 ± 0.28 [†]	1.95; 3.19	1.43E-10 [†]	76 [†] (85)
B	inf	m	DRE	1.25 ± 0.12	1.00; 1.51	1.35E-13	
C	inf	hu	HepG2	0.29 ± 0.03	0.22; 0.36	3.67E-10	
C	eff	hu	HepG2	6.01 ± 1.61 [†]	2.48; 9.54	1.78E-11 [†]	95 [†] (92)
C	inf	zf	ZFL	0.99 ± 0.11	0.75; 1.22	3.71E-10	
C	eff	zf	ZFL	3.33 ± 0.79 [†]	1.59; 5.06	1.10E-10 [†]	70 [†] (92)
C	inf	m	DRE	1.88 ± 0.31	1.23; 2.52	8.99E-14	
Ctrl	TCDD	hu	HepG2	1.07E-10 ± 4.43E-11	1.52E-11; 1.99E-10		
Ctrl	TCDD	zf	ZFL	3.66E-10 ± 8.45E-11	1.91E-10; 5.42E-10		
Ctrl	TCDDI	m	DRE	1.69E-13 ± 1.06E-13	4.58E-14; 3.83E-13		

983

Tab. S7: EC50 estimates computed herein vs. reported EC50 estimates within the literature for identical assays or reporter constructs.

	EC ₅₀ measured [M] ± SE	EC ₅₀ reported [M] ± SE	Reference	Ratio
HepG2	5.77E-10 ± 1.10E-10	2.2E-10 ± 3E-11	[15]	2.64
ZFL	3.08E-09 ± 3.21E-10	1.2E-09 ± 4.7E-12	[12]	2.58
DRE	7.58E-12 ± 4.36E-12	2.8E-12 (SE not given)	[17]	2.71

984

Tab. S8: Percentage [%] of BEQ_Bio explained by BEQ_Chem, as derived from the iceberg modelling approach (see S112, tab. "REP_BEQ_calculations"), regarding the utilised regression model (see section 3.4 herein).

Reg. Mod.	Cell line	site A inf.	site A eff.	site B inf.	site B eff.	site C inf.	site C eff.
-----------	-----------	-------------	-------------	-------------	-------------	-------------	-------------

SL	HepG2	0.0017	0.0038	0.0022	0.0040	0.0011	0.0067
	ZFL	0.0001	0.0002	0.0001	0.0003	0.0004	0.0004
	DRE	0.0109		0.0309		0.0752	
2PNL	HepG2	0.008	0.017	0.009	0.017	0.005	0.029
	ZFL	0.001	0.001	0.001	0.002	0.003	0.003
	DRE	0.310		0.881		2.142	
3PNL	HepG2	0.004	0.010	0.006	0.010	0.003	0.017
	ZFL	0.001	0.001	0.001	0.002	0.003	0.003
	DRE	1.610		4.571		11.121	
4PNL	HepG2	0.007	0.015	0.008	0.015	0.004	0.026
	ZFL	0.001	0.001	0.001	0.003	0.003	0.003
	DRE	1.894		5.377		13.084	
30%SL	HepG2	0.005	0.010	0.006	0.011	0.003	0.018
	ZFL	0.001	0.001	0.001	0.002	0.002	0.002
	DRE	0.642		1.822		4.433	

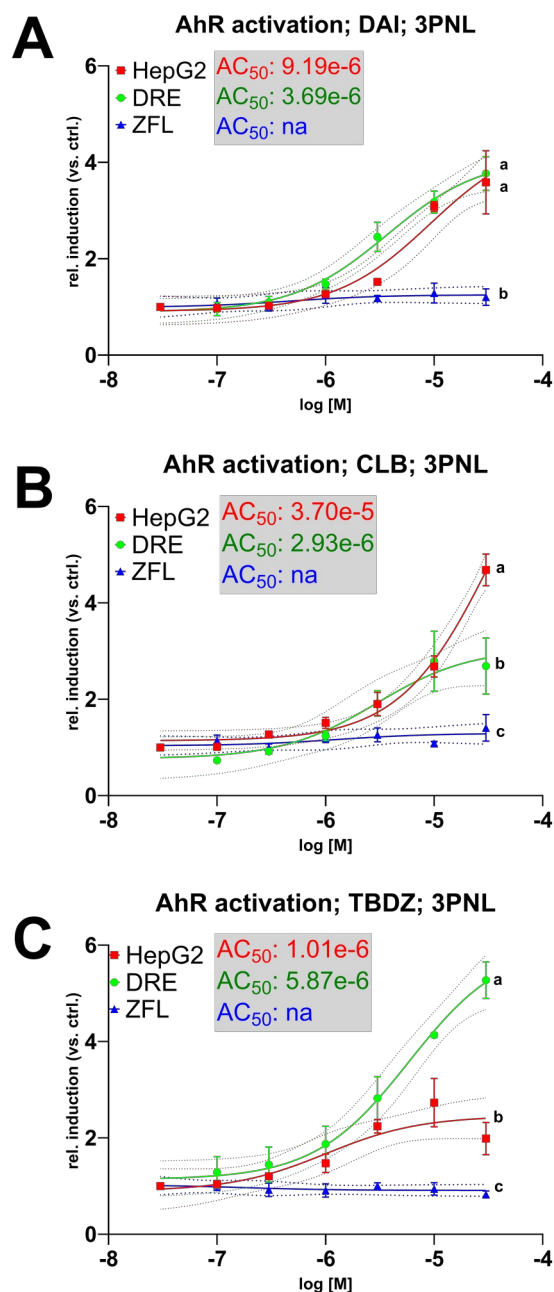
985

Tab. S9: Summary of all tested ISEDA compounds' concentration estimates (AC₅₀ and EC₁₀), including their variation metrics. Effect concentration estimates computed beyond the empirically tested exposure concentrations are marked with a double-cross (#). Non-computable estimates and respective variation metrics are marked with "na" (not applicable).

AC ₅₀			
ISEDA-identified compound	Reporter cell line	AC ₅₀ [M]	SE [M]
Daidzein (DAI)	HepG2	9.19E-06	±4.51E-06
	DRE	3.69E-06	±1.19E-06
	ZFL	na	na
Climbazole (CLB)	HepG2	3.70E-05	±1.85E-05
	DRE	2.93E-06	±2.08E-06
	ZFL	na	na
Thiabendazole (TBDZ)	HepG2	1.01E-06	±8.23E-07
	DRE	5.87E-06	±2.11E-06
	ZFL	na	na

EC ₁₀			
ISEDA-identified compound	Reporter cell line	EC ₁₀ [M]	SE [M]
Daidzein (DAI)	HepG2	2.39E-05	±3.76E-06
	DRE	4.09E-06	±9.20E-07
	ZFL	na	na
Climbazole (CLB)	HepG2	1.59E-05	±1.09E-06
	DRE	5.75E-06	±3.62E-06
	ZFL	na	na
Thiabendazole (TBDZ)	HepG2	2.14E-04‡	±2.61E-04‡
	DRE	1.62E-06	±4.18E-07
	ZFL	na	na

986



987

988

989

990

991

992

993

994

995

996

997

998

999

Fig. S40: Alternative data fitting (ISEDA). Activation of the cellular xenobiotic metabolism (AhR-ARNT-mediated toxicity pathway), depicted as relative induction in relation to the solvent control, in human (HepG2, red), zebrafish (ZFL, blue), and mouse (DRE, green) reporter gene assays exposed for 24 h to the ISEDA-identified compounds daidzein (DAI, panel A), climbazole (CLB, panel B), and thiabendazole (TBDZ, panel C). Points indicate the experimental mean (experimental units $n = 3-5$, observational units $N = 9-15$), with the SEM marked by whiskers. Samples were analysed in a logarithmic dilution series with 30 μM as the highest concentration. CRCs were fitted using non-linear, log-log regressions (model “3PNL”; see SM, section 3.4) and are shown in their respective colours, with 95% CIs as faded, dotted lines. Effect point estimates (AC₅₀; coloured, vertical, and dotted lines) were derived from the concentrations causing 50% of the response. The respective values are depicted in the graphs. A two-way ANOVA, followed by Tukey’s test for multiple comparisons, was conducted to compute differences between effect sizes of the respective reporter line comparisons (alternating lowercase letters indicate statistical significance, $p < 0.05$). The respective samples’ cellular viability/cytotoxicity data are given in the SM, fig. S30. The raw data and metadata are provided in SI15.

3.6 Supplementary results: Main article – Bioavailability assessment

To evaluate the potential impact of bioavailability differences within the *in vitro* test systems, the UFZ-LSER QSPR tool [45] was applied, employing mass-balance equation models to estimate free cellular concentrations [46]–[48]. DMEM was chosen as the standard culture medium input parameter because all cultures rely on DMEM or a variant thereof, and mass-balance parameters are not defined for more specific cultivation media. Similarly, HepG2 was chosen as the sole cellular parameter in the mass-balance equation, given that none of the other cell lines used in this study has established parameters. As such, only the FBS amount and cellular concentrations (at seeding) were accounted for in the model. From previous experience, these parameters describe most of the compartmentalisation, and specific differences introduced by the utilised cellular systems are relatively minuscule [46], [59].

First, the reference compounds tBHQ and MZC were assessed at a nominal concentration of 3 μ M, close to the computed $EC_{IR1.5}$ estimates. Because the same amount of FBS was used in both the ZF4 and AREc32 assays and the number of cells was comparable, we also calculated free cellular concentrations, which were nearly identical (fig. S41A).

Second, free cellular concentrations of the reference compound TCDD and the three identified ISEDA compounds (CLB, DAI, and TBDZ) were computed for the ZFL, HepG2, and DRE bioassays. A fixed nominal concentration of 30 nM was chosen as a standard of comparison, and the nominal EC/AC_{50S} were transformed into free EC/AC_{50S} . For the fixed concentration, free cellular concentrations were consistent across the ZFL and DRE assays, whereas the HepG2 assay showed lower concentrations, reflecting its higher FBS content (fig. 41B).

Finally, further comparison of TCDD-relative effect potency (REP) values of the ISEDA compounds between nominal and free cellular concentrations was performed for the HepG2 and DRE assays (fig. 41C). ZFL-specific calculations were also derived but are not presented within the articles due to the instability of the point estimates (see fig. S40, or fig. 8 in the main article). The latter values and all raw data are given in SI13. The comparison (fig. S41D) indicated negligible differences in TCCD-REP nominal or free cellular concentration ratios between HepG2 and DRE, suggesting that bioavailability differences were minuscule and unlikely to account for the observed discrepancies in activation, reinforcing the conclusion that the observed effect differences are predominantly mechanistic in nature.

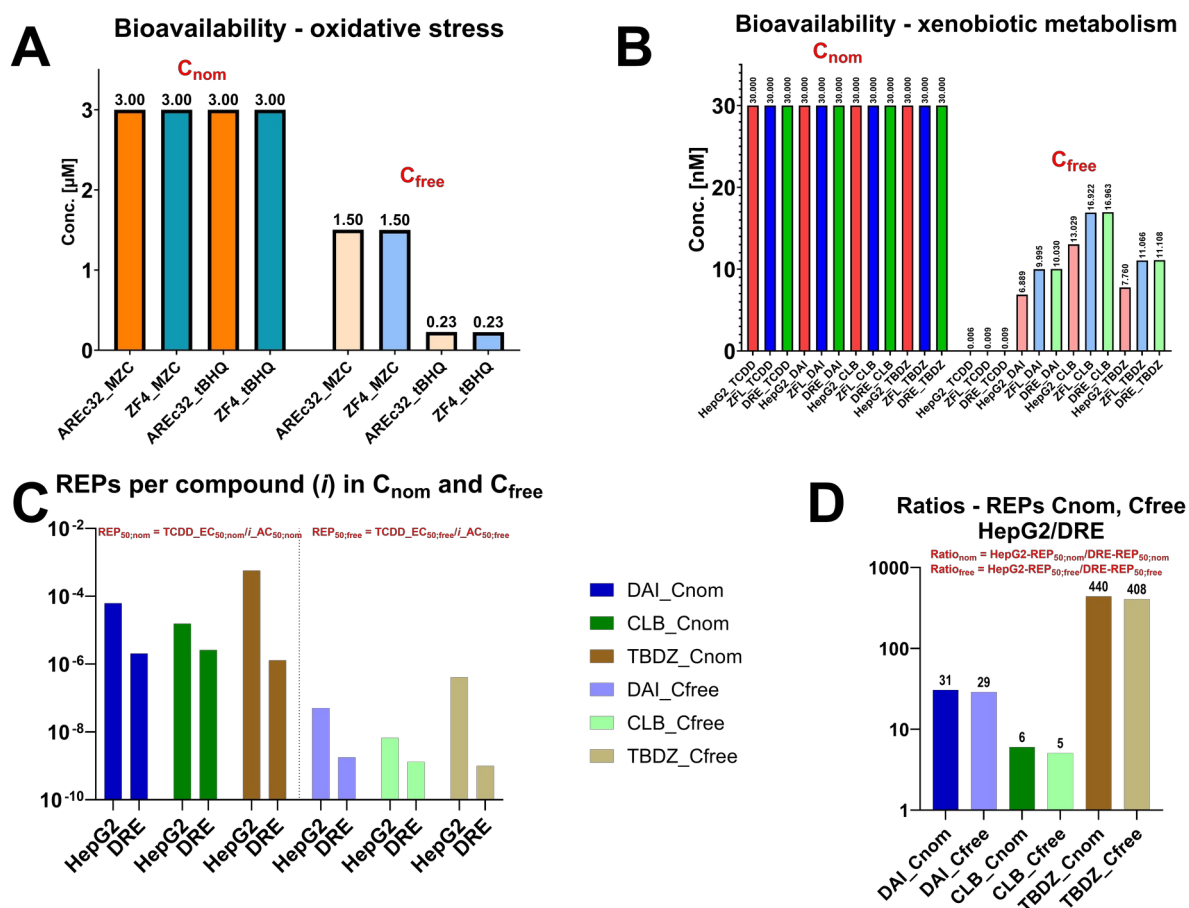


Fig. S41: Outline of the bioavailability assessment regarding exposure concentrations and derived effect point estimates. Free cellular concentrations (C_{free}) were derived from nominal concentrations (C_{nom}) utilising the mass-balance models in the UFZ-LSER QSPR tool. Panel A: bioavailability assessment for the oxidative stress MoA exposure concentrations. Panel B: bioavailability assessment for the xenobiotic metabolism MoA exposure concentrations. Panel C: relative effect potencies (REPs) derived from nominal and free cellular concentrations. Panel D: potency ratios between HepG2 and DRE-derived nominal and free cellular concentration estimates. Abbreviations: tert-Butylhydroquinone (tBHQ), metazachlor (MZC), 2,3,7,8-tetrachlorodibenzodioxin (TCDD), daidzein (DAI), climbazole (CLB), and thiabendazole (TBDZ). Additional data are given in SI13 and SI16.

4. Supplementary discussion

4.1 The Nrf2-Keap1-ARE toxicity pathway – Canonical reactive cysteine residues and mechanistic insights on the “cysteine code”

4.1.1 Background and objectives of this project concerning Keap1

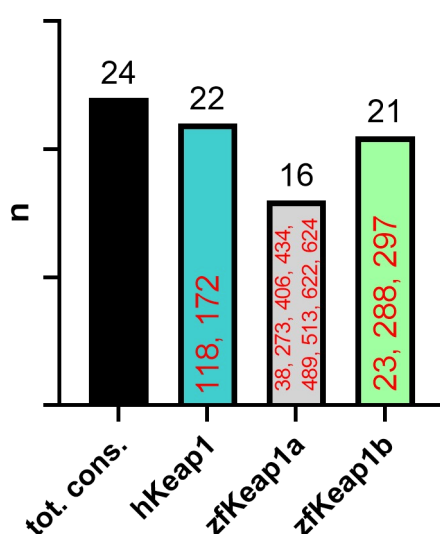
Keap1 (kelch-like ECH-associated protein 1) is an oxidative stress sensor protein. Electrophiles react with the cysteine residues of this protein, most prominently Cys151, 226, 273, 288, 613, 622, and 624, triggering a conformational change and the release of the antioxidant transcription factor Nrf2 (nuclear factor erythroid-derived 2-like), which regulates the cellular response to oxidative stress. In zebrafish, there are two paralogs of the protein – zfKeap1a and zfKeap1b with slightly different functionality [60].

1048 Also, the canonical cysteine residues are alternatively distributed between the two paralogs; 1a is
 1049 missing Cys273, 622, and 624, whereas 1b lacks Cys288 (beyond other non-canonical cysteines). See
 1050 also fig. S42 below for a comparison of conserved cysteine residues among the investigated zebrafish
 1051 (zfKeap1a, zfKeap1b) and human (hKeap1) sensor variants.

1052 The interaction of Keap1 with two reference electrophiles, tert-butylhydroquinone (tBHQ) and
 1053 metazachlor (MZC), was studied in both biological (*in vitro* reporter gene assays, main article) and
 1054 computational (*in silico* molecular docking, “data in brief” co-submission, DIB_B) terms. While tBHQ
 1055 interacts primarily with Cys151 [60], [61] and is thereby considered a Cys151-dependent activator, the
 1056 reaction mechanism with MZC is unknown. Within the main article, tBHQ was initially chosen as a
 1057 reference compound to derive bioequivalent values (BEQs) for environmental samples, given its
 1058 previous utilisation in that role [57]. We decided on the environmental pollutant MZC as an alternative,
 1059 as others have associated it with the induction of oxidative stress [62], and we have previously
 1060 demonstrated its ability to induce the Nrf2-Keap1 pathway [7], [8].

1061 The goals of docking the reference electrophilic compounds to different structures of human and
 1062 zebrafish Keap1 were: (i) to gain insight into where binding might occur, (ii) to observe if the poses are
 1063 near any cysteine residues, especially around canonical ones, e.g. Cys151, (iii) to compare the
 1064 predicted poses in the human variant and two zebrafish paralogous variants, and (iv) to compare
 1065 binding poses in crystallographic and AlphaFold-predicted structures.

Amount (n) of total conserved cysteines per variant



1066

Fig. S42: Amount of total conserved cysteines within all Keap1 variants. The total number of conserved cysteine residues per variant is given on top of each bar. Missing conserved cysteines per variant are provided inside the bar, highlighted in red. Residue enumeration corresponds to the hKeap1 AA sequence (see DIB_B, fig. 6).

4.1.2 The Nrf2-Keap1-ARE signalling pathway: nomenclature and mechanisms

The Nrf2-Keap1-ARE signalling pathway is one principal regulator of oxidative stress detoxification, metabolism, and induction of related phase-II enzymes (e.g., glutathione S-transferase) [4], [63], [64]. Further, it plays a principal role in lipid, carbohydrate, nucleotide, and AA metabolism (reviewed in [65]). Crosstalk and reciprocal regulation between ARE-Nrf2-Keap1 and other central signalling cascades, such as MAPK/ERK, etc., have been reported [66]–[69].

The antioxidant-responsive element (ARE) sequence [70], [71] is found in the regulatory region of the mentioned phase-II metabolising enzymes, often coinciding with the xenobiotic response element (XRE) of the AhR-ARNT signalling pathway [72].

The nuclear factor erythroid-derived 2-like (Nrf2) is a basic region leucine zipper (bZip)-type transcription factor of the “Cap’n’Collar” (CNC) family that binds to the ARE and transactivates expression of related genes [63], [64], [73]–[75]. Nrf2 occurs ubiquitously in the cytosol under normal physiological conditions, bound to its negative regulator, Kelch-like ECH-associated protein 1 (Keap1).

Keap1, an actin-bound protein, serves as an adaptor for the Cullin 3-based E3-ligase ubiquitin complex (CUL-E3) and thereby marks Nrf2 for further ubiquitination and subsequent degradation [73]. Via the broad complex, tramtrack, and bric a brac (BTB) domain, Keap1 homodimerises with itself and also binds to CUL-E3 (see fig. 4 in DIB_B and fig. S3 for domain mapping). Nrf2 attachment is modulated via the Kelch domain (fig. 3 in DIB_B) [76], [77]. One of each Keap1 homodimers’ Kelch domains anchors to DLG and ETGF motifs within Nrf2 (reviewed in [65], [78]). The Nrf2-Keap1 heteromer complex is described by a so-called “hinge-and-latch” apparatus, in which the DLG/Kelch interaction represents the latch and the ETGF/Kelch interaction the hinge. Thereby, the binding affinity of the latter complex is approximately two orders of magnitude higher [77].

Turnover and half-life of Nrf2 are short-term under unstressed conditions [79]. However, upon exposure to oxidative stress, Nrf2 is released from Keap1, thereby avoiding proteolysis. It translocates and accumulates within the nucleus, initiating gene expression [80].

Electrophilic interactions mediate Nrf2 release. As a common attribute, electrophiles can covalently modify sulfhydryl groups by alkylation, oxidation, or reduction [81], [82]. As such, 22 cysteine residues can be identified within the hKeap1 amino acid backbone. Some are more susceptible to reaction (Cys151, 226, 273, 288, 613, 622, and 624) and can be termed “canonical cysteines” [82]–[84]. Thus, certain classes of compounds seem to prefer interactions with specific cysteine residues and can be

1100 categorised accordingly [60]. However, the overall recruitment of a cysteine pattern appears unique
1101 for every compound, a phenomenon termed “cysteine code” (reviewed in [85]).

1102 *In vivo* and *in vitro* analyses revealed Cys151, 226, and 288 as the primary sensors due to increased
1103 reactivity (reviewed in [78]). Especially Cys151 is considered highly reactive, given that nearby basic
1104 residues (H129, K131, K150, and H154) lower its pKa, rendering it a thiolate anion even at physiological
1105 pH. Further research has shown that certain compounds are Cys151-preferable inducers, whereas
1106 others are Cys151-independent [61]. As such, common inducers tBHQ and sulforaphane are Cys151-
1107 preferable, implying that a deletion/mutation of Cys151 still renders Keap1 activation by these
1108 compounds, just to a lesser extent. Hence, the Nrf2-Keap1 interplay cannot be reduced to a single
1109 cysteine interaction but depends on multiple sensors (“cysteine-code”).

1110 Cys151 is located in the previously mentioned BTB domain, which coordinates the interaction with the
1111 CUL-E3. Cys273 and Cys288 are situated within the intervening region (IVR) domain, which links BTB to
1112 Kelch. The geometry of the polypeptide chain changes when electrophiles covalently bind to these
1113 cysteine residues. Regarding location, interaction with CUL-E3 is prevented, or Kelch dissociation is
1114 mediated by the respective DLG and ETGF motifs, leading to Nrf2 release or, at least, Keap1
1115 inactivation. More recent studies suggest a cycling mechanism (discussed in [78]). Per se, Nrf2 is not
1116 released from Keap1 once bound, even upon conformational changes within Keap1 due to electrophilic
1117 interactions [86]. Moreover, ETGF/Kelch interactions remain persistent, prevalent Keap1 is only
1118 rendered non-functional, and the oxidative stress response is mediated by *de-novo* synthesised Nrf2.
1119 Additionally, the latest results indicate an additional role of protein-protein interaction inhibitors, such
1120 as p62, regarding the hinge and latch mechanisms. Apparently, DLG/Kelch latch dissociation is also
1121 dependent on p62-mediated Nrf2 activation and not solely on electrophilic interactions [87].

1122 The Nrf2-Keap1-ARE signalling pathway is highly conserved within vertebrates and beyond [75], [88],
1123 [89]. However, structural differences occur in fish due to the teleost genome duplication [42]. For both
1124 Nrf2 and Keap1, duplicated co-orthologs (Nrf2a/Nrf2b and Keap1a/Keap1b, respectively) were
1125 identified in zebrafish [90], [91]. Beyond duplication, subfunction partitioning occurred with
1126 differences in expression and spatiotemporal characteristics [91]. Structural differences were
1127 reported, with Nrf2b lacking certain transactivation domains and Keap1a and Keap1b being Cys273- or
1128 Cys288-deficient (beyond other cysteine residues). Nevertheless, both variants retain functionality
1129 [88]. In more detail, studies have pointed out that zfKeap1b is an authentic ortholog of mammalian
1130 Keap1 [92]. As such, Cys151-preferable compounds show minor mitigation of the oxidative stress
1131 response for zfKeap1a compared to zfKeap1b.

1132 4.1.3 Molecular docking results in the context of the “cysteine code” (see DIB_B for results)
1133 We concluded from the molecular docking analyses that tBHQ showed qualitatively similar affinity
1134 patterns and binding energies between human and zebrafish Keap1 variants regarding blind docking
1135 (DIB_B, section 3.1.1). Notably, there were minuscule interaction differences, with zfKeap1a preferring
1136 binding to the IVR (Cys226) and zfKeap1b to the Kelch domain (Cys368/513). Site-directed docking to
1137 the Cys151 proximity revealed cognisable binding energy differences for tBHQ between variants with
1138 the following hierarchy: hKeap1 > zfKeap1b > zfKeap1a.

1139 For MZC (DIB_B, section 3.1.2), we found comparable best-energy binding poses across all three
1140 variants within the Kelch domain (Cys368/513). However, for intermediate- to low-energy binding
1141 poses, MZC affinities diverged across specific variants. As such, MZC preferred binding to Cys151 in
1142 hKeap1, Cys613 in zfKeap1a, and Cys226 in zfKeap1b. Site-directed docking to the Cys151 proximity
1143 revealed the same binding energy hierarchy as for tBHQ: hKeap1 > zfKeap1b > zfKeap1a.

1144 Comparison of AA sequence homology between variants solidifies the above observations (DIB_B,
1145 section 3.1.1.6, figs. 6 and 7), as hKeap1 and zfKeap1b share higher overall sequence homology (73%),
1146 especially in the vicinity of Cys151 (90%). Nevertheless, substituting only 3 AA residues between
1147 hKeap1 and zfKeap1b within the Cys151 proximity leads to a binding energy reduction of 0.7 Δ
1148 kcal/mol. As such, sequence homology coincides with the encountered Cys151 binding energy
1149 hierarchy: hKeap1 > zfKeap1b > zfKeap1a.

1150 The biological data presented in this project, the literature, and the molecular docking results support
1151 the concept that tBHQ is a Cys151-dependent inducer. According to the docking data, the primary
1152 interactions occur at Cys226, 368, and 513. However, Cys151 recruitment is necessary for complete
1153 engagement (“cysteine code”). Because zebrafish variants show lower susceptibility to tBHQ in the
1154 Cys151 vicinity, this reduces the cellular stress response. For tBHQ, the cysteine code seems to be
1155 conserved across variants.

1156 Conversely, MZC is likely a Cys151-independent inducer that interacts via the Kelch domain
1157 (Cys368/513) in all variants. Differences in Cys151 interaction do not impact the measured *in vitro*
1158 endpoint. Interestingly, however, beyond the Kelch domain, MZC preferentially interacts with different
1159 cysteine residues in the observed Keap1 variants. The latter might indicate a diversion in functionality
1160 and susceptibility within the zebrafish variants to compensate for missing cysteines (compared to the
1161 human AA sequence).

4.2 Supplementary discussion: Comment on zfKeap1a/b functional divergence and structural variability

Multiple sequence alignments (see DIB_B, fig. 7) resolved 55% and 90% homology for zfKeap1a and zfKeap1b, respectively, to hKeap1 near cysteine 151 (amino acid residues 125-156). Despite minimal amino acid differences for zfKeap1b, tertiary structural changes affect tBHQ affinity at this site. In line with these results, studies have pointed out that zfKeap1b is an authentic ortholog of mammalian Keap1 [92]. As such, cysteine 151-preferable ligands show minor mitigation of the oxidative stress response for zfKeap1a compared to zfKeap1b. Nevertheless, both variants retain functionality [88]. Structural differences within fish variants occur due to the teleost genome duplication [42], resulting in duplicated co-orthologs (Nrf2a/Nrf2b and Keap1a/Keap1b) in zebrafish [90], [91]. Structural differences were reported: zfNrf2b lacks certain transactivation domains, whereas zfKeap1a and zfKeap1b are deficient in specific cysteine residues (see fig. S42). However, in our experimental setup, it remains unclear how signal transduction is sequestered among different zebrafish Nrf2 and Keap1 variants, as we can only measure the final interaction with the ARE regulatory unit. The extent to which stochastic processes between zebrafish Nrf2 and Keap1 variants impact transactivation in a cysteine 151-dependent manner remains enigmatic and warrants further investigation.

4.3 Supplementary discussion: “CALUX” assay nomenclature and identification

As the main article's discussion suggests, we want to highlight a significant gap in methodological reporting in the literature, particularly regarding the AhR/DRE CALUX reporter bioassays. Iterations of this assay arise from different generations of the reporter construct (pGudLuc1 to 7.5 [13], [93]–[95]) and their transfection into diverse species' cell lines [15], [96], [97], leading to numerous AhR/DRE CALUX clones with varied sensitivities to TCDD [15]. Eichbaum et al. [98] reviewed the use of AhR-responsive reporter bioassays for detecting DLCs and noted difficulties in identifying specific clones used in studies. Consistent with our findings, we recommend explicitly specifying the cell line and clone used when referencing the AhR/DRE CALUX-type assay. Noteworthy are the complications associated with extrapolating AhR-mediated effects across different species using reporter assays, which have been noted since their inception in environmental testing [99].

4.4 Supplementary discussion: Cross-species assessment in context with regulatory metrics – BEQs and EBTs

In the main article's discussion, we elaborate on the complexities involved in BEQ normalisation, particularly how it skews cross-species comparisons. This issue becomes apparent with TCDD-BEQ normalisation, due to species- and ligand-specific modulation of the AhR-ARNT-XRE pathway. To address these concerns, we conducted a thought experiment (fig. S43) comparing all our derived

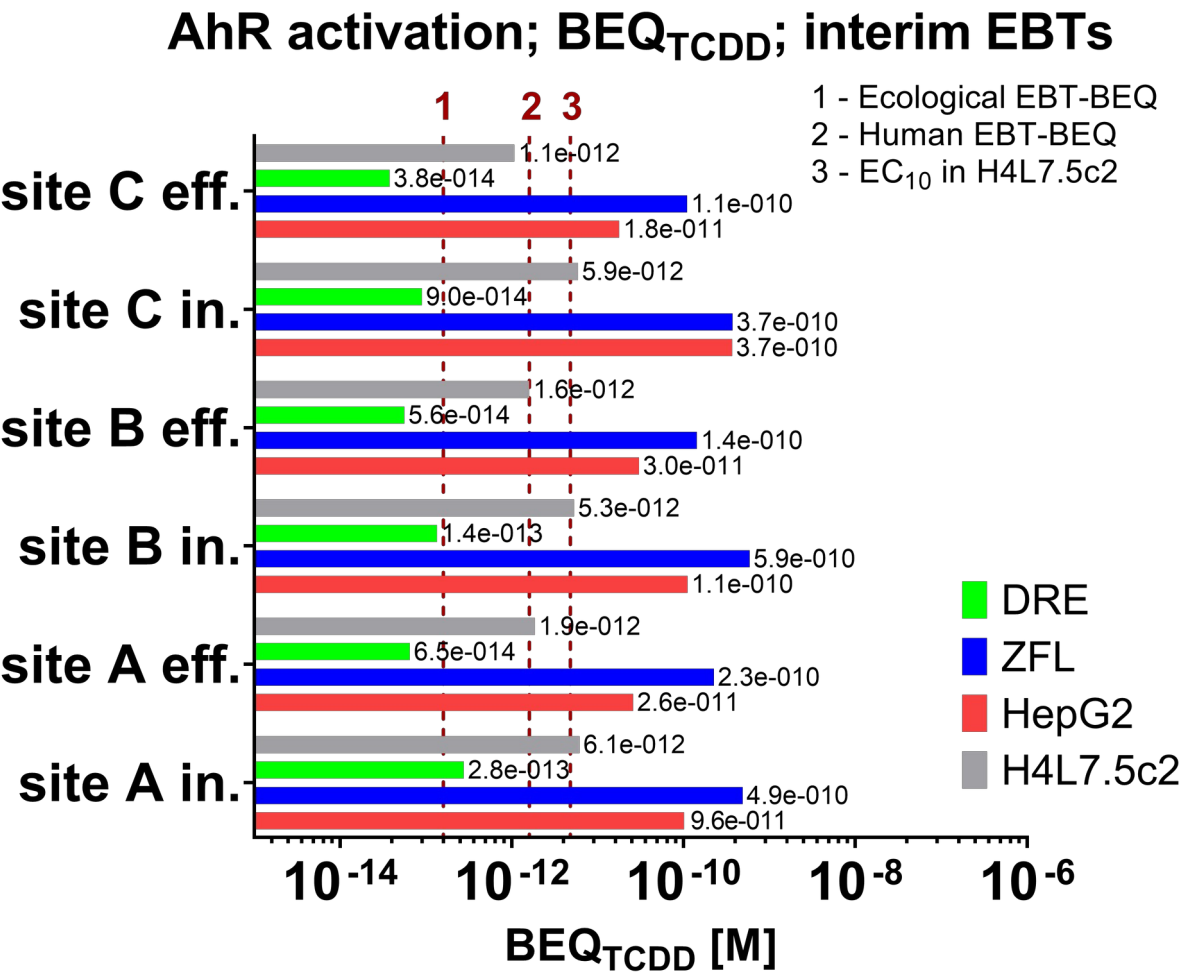
1195 TCDD-BEQ values from different species reporter assays with published interim tier 1 effect-based
1196 trigger (EBT) values.

1197 Rat TCDD-BEQs (“H4L7.5c2” in fig. S43) were computed for comparison. The AhR CALUX-type
1198 bioassay clone H4L7.5c2 is based on the rat hepatocellular carcinoma (H4-II-E, CVCL_0284) cell line,
1199 and is often employed in the bioanalytical assessment of environmental samples. Given that the
1200 human HepG2-based CALUX-type assay clone (HG2L7.5c1) can be easily confused with the rat clone
1201 (sometimes clone definitions are even completely omitted in the literature; only “CALUX” is
1202 mentioned) and that the two clones differ by one order of magnitude in terms of responsiveness to
1203 TCDD potency [15] we wanted to depict potential implications on the derivation of safety levels. Rat
1204 TCDD-BEQs were computed from the geometric means of our data. Given the variability in our
1205 biological estimates and the H4L7.5c2 CALUX-type assay's responsiveness to TCDD, we deemed this
1206 approach a feasible compromise. The EC₅₀ value for H4L7.5c2 was derived from the literature [15],
1207 while the EC₁₀ was computed assuming a Hill slope of one.

1208 Further, we consulted interim EBTs, as presented and discussed in Neale et al. [100], [101] to assess
1209 our data from a hypothetical regulatory perspective. The ecological EBT-BEQ was initially derived
1210 from van der Oost et al. [102] and the human EBT-BEQ from the State of California water monitoring
1211 program [103]. The EBT-BEQs by Van der Oost were derived from *in vivo* exposure studies on the rare
1212 minnow (*Gobiocypris rarus*) and employ a species-sensitivity distribution approach. However, the
1213 source of the State of California's EBT-BEQs could not be found. These interim EBTs are intended to
1214 be broadly applicable (cross-species and -assay); however, their limitations are acknowledged.
1215 According to Carere et al. [104], these interim values constitute tier 1 EBTs. At the same time,
1216 previous work by fellow researchers and co-authors already demonstrated the need for assay-
1217 specific EBTs [105], and the hypothetical cross-species scenario presented here further illustrates
1218 why such values will be important for future regulatory applications, as broadly applied interim EBTs
1219 may shift interpretation depending on assay-specific reference-compound responsiveness rather
1220 than on environmental-sample bioactivity alone.

1221 Our hypothetical, comparative scenario, as illustrated in fig. S43, indicates that mouse DRE assay-
1222 derived BEQs are generally lower than the ecological and human EBT-BEQ. Rat AhR CALUX-type-
1223 derived BEQs in effluent samples are below the human EBT-BEQ, while human HepG2 and zebrafish
1224 ZFL BEQs exceed these thresholds. Interestingly, this is not necessarily due to increased sensitivity of
1225 the bioassays to environmental samples, but rather to their lower sensitivity to TCDD compared with
1226 rodent models. This suggests that general, broad-application, interim EBTs of the AhR response can

1227 lead to either overestimation or underestimation of effects, depending on the assessment
1228 perspective and assay used.



1229
1230 Fig. S43: Hypothetical TCDD-BEQ scenario derived from the cross-species bioanalytical assessment. Published tier 1/interim
1231 AhR EBT benchmark values are illustrated as dotted red lines for comparison; these values were not recalculated as assay-
1232 specific EBTs for each reporter assay.

1233 4.5 Supplementary discussion: TCDD as a reference compound, reasoning and
1234 scrutiny

1235 In their comprehensive research, Escher et al. [106], [107] delineated tier 2 EBTs, emphasising
1236 environmental quality standards (EQS) as a point of departure. These EQS inherently necessitate an
1237 inter-species sensitivity assessment. However, a limitation arises in their application: these tier 2
1238 EBTs are explicitly defined only for the reference compound benzo[a]pyrene (BaP). Our choice of
1239 TCDD as the reference compound in this study is based on several factors. While PAHs are more
1240 commonly found internationally due to their origin in combustion processes, in our geographical
1241 context, DLCs appear more relevant. This significance is due to the legacy contamination from paper
1242 mills and trash incineration in Scandinavia [108], [109].

1243 While analysing DLCs and PAHs might have been more effective for explaining a larger fraction of the
1244 biological response, we aimed for a broader chemical analysis targeting emerging pollutants. We
1245 regard this scenario as more realistic within a standard screening approach. Following the usual
1246 *modus operandi*, we employed TCDD as a reference compound. Regarding iceberg modelling and the
1247 subsequent ISEDA, our objectives were threefold: to assess the cross-species applicability of iceberg
1248 modelling, to identify critical drivers of AhR-ARNT-XRE pathway induction from the ISEDA, and to
1249 evaluate specific compounds across species, potentially revealing variations in sensitivity to these
1250 pollutants.

1251 Moreover, DL-PCBs are known for their propensity to accumulate in the hydrosphere [110], [111].
1252 Another crucial aspect is that TCDD is a more widely used reference compound, enabling the
1253 application of a standard *modus operandi*. The choice of TCDD was further influenced by the
1254 complexities of using BaP. The majority of PAHs require biotransformation to trigger AhR-mediated
1255 toxicity [112]–[114]. The extent and comparability of these biotransformation capacities in stable
1256 hepatocyte cell lines, like those used in our study, remain poorly documented and understood. We
1257 will address the issue of biotransformation in a section below (4.8).

1258 When considering the environmental behaviour of these compound classes, both PAHs and DLCs
1259 exhibit poor water solubility [115]. The aqueous solubility of PAHs decreases logarithmically with
1260 increasing molecular mass, but the smaller, two- and three-ringed PAHs are more soluble [115]. In
1261 contrast, AhR-mediated toxicity is primarily associated with PAHs that have three or more aromatic
1262 rings [116]. TCDD and BaP have been shown to interact similarly with the AhR PASb domain [117],
1263 emphasising the need for a careful selection of reference compounds in environmental toxicology
1264 studies. Ultimately, the decision for the reference compound hinges on identifying the most potent
1265 inducer or the primary driver of the observed biological effect in the environment [50], [104].

1266 However, how representative is AhR activation by TCDD or BaP of this specific MoA in an
1267 environmental water context? Both are unlikely contaminants in running water [118]–[120], whereas
1268 other classes of compounds are more likely to activate the AhR [121]. This activation acts more as a
1269 signal modulator, especially in fish with multiple AhR variants. As discussed in the main article,
1270 mixtures might elicit a general cellular stress response mediated by the AhR. The observed
1271 differences in BEQ values are primarily due to variability in the potency of the reference compound
1272 and to the BEQ calculation method. Many AhR-related effects in environmental mixture samples may
1273 be facilitated not only by receptor-mediated mechanisms but also by multidimensional pathway
1274 crosstalk.

Generally, for bioassays indicative of integrative effects (non-specific, reactive, and adaptive stress responses), known and identified chemicals explain only a small to intermediate fraction of the effect [49]. In contrast, for bioassays indicative of hormonal receptor-mediated effects, known and identified chemicals can more often explain a larger percentage of the effect [50], [122], [123], with some deviations observed between human and yeast-based assays. Contextually, the choice of the reference compound is critical, with TCDD and BaP being common choices due to their strong induction of AhR-related effects. Yet one must consider whether these are the most representative compounds of the MoA in aquatic contexts. Sediment, biota, and feed/food samples may be more suitable for targeting PAHs and DLCs due to the tendency of hydrophobic compounds to bind to organic matter, bioaccumulate, and biomagnify, as previously demonstrated by others [124]–[130]. In light of the results and discussion, it may be worthwhile to consider alternative reference compounds for assessing the MoA of xenobiotic metabolism in environmental water samples.

The BEQ concept (see also section 4.4 above) is based on toxic equivalency factors (TEFs) or toxic equivalents (TEQs) approaches [131]–[133]. TEFs index structurally related compounds to a well-studied reference compound based on shared characteristics, thereby hypothetically aggregating signals of a specific MoA. This approach requires meeting seven criteria: demonstrated need, defined chemical grouping, a broad toxicological database, endpoint consistency, additivity of response, a common MoA, and consensus [131]. This approach has recently been reevaluated for TCDD in the context of dioxin-like polychlorinated dibenzo-p-dioxins, polychlorinated dibenzofurans, and polychlorinated biphenyls [134]. Notably, this comparison was made between several fish species, using TCDD-TEFs derived from rainbow trout. Given the multidimensional modulation of AhR-related toxicity, it is questionable whether a single compound, such as TCDD or BaP, can meet these criteria. Similarly, van Duursen et al. [135] noted that the AHR-mediated risk for DL-PCBs in humans is likely overestimated in the current TEF concept when extrapolated from rodent data. Given that we aim to extend environmental monitoring to encompass the entire biosphere, this exercise becomes increasingly challenging.

4.6 Supplementary discussion: validity of molecular docking approach

An earlier cross-species molecular docking study [136] reported findings that differ from ours: mAHR and zAhR2 were modelled at approximately -4 kcal/mol, and hAhR's affinity towards TCDD was estimated at only -2 kcal/mol (compared to approximately -10 kcal/mol in our study). However, Bisson *et al.*'s study was not corroborated by biological data, unlike ours, as elaborated in the main article. Moreover, as discussed by Tagliabue et al. [117], this earlier study had limitations, particularly its reliance on apo-template structures for homology modelling and its oversight to account for the ligand-induced fit effect on protein conformation.

In contrast, we utilised the most recent cryo-EM structure of the human PASb domain [36] for molecular docking. Additionally, recent advancements in computational tools, such as AlphaFold and AutoDock Vina, have significantly improved the accuracy of protein structure prediction and ligand docking [31], [137]. While not a substitute for experimental validation, these methods now serve as powerful complements to traditional techniques. The agreement between our docking results and biological data further supports this.

4.7 Supplementary discussion: AhR crosstalk and non-canonical pathways

As mentioned in the main article, AhR crosstalk and the induction of non-canonical pathways may be due to several reasons:

Firstly, the existence of multiple XRE regulatory units has been established [138]; some of which are located near genes involved in the oxidative stress response, such as NQO1. This suggests a complex interaction between the AhR and cellular stress defence pathways [139], [140], where AhR may regulate genes downstream of the Nrf2-Keap1-ARE pathway, necessitating coordinated activation [72], [141], [142] or, conversely, being influenced by these pathways [143], [144].

Secondly, various non-canonical pathways have been identified apart from the well-known ligand to AhR-ARNT-XRE pathway [5], [145]. In these pathways, AhR forms dimers with various translocators and transcriptional co-factors, leading to the activation of distinct downstream genes. Specific promiscuous ligands are known to activate these non-canonical pathways [146]. A microarray analysis revealed that only about one-third of the genes differentially expressed by AhR activation were associated with XRE binding, suggesting the involvement of mechanisms beyond the canonical pathway [147]. Unlike typical nuclear receptors, such as hormonal receptors, AhR uniquely responds to a wide range of structurally diverse chemicals, producing varied ligand-selective toxicological and biological effects through multiple AhR-dependent mechanisms.

The non-canonical involvement of AhR can manifest at various levels of pathway regulation, including the ligand-binding domain (LBD), sequences adjacent to the LBD that potentially influence PAS-PAS motif interactions, variations in signal transduction through non-canonical translocators and co-activators, and differential interactions with varying genomic response elements [148], [149].

4.8 Supplementary discussion: outlook and further technicalities regarding multi-species reporter gene assays

From the synopsis of the main article's discussion, we provide an outlook and propose further implementation steps to enhance the regulatory acceptance of reporter bioassays as a standard tool

1340 in environmental monitoring (fig. S44). While multi-species reporter bioassays boast accessibility,
 1341 technical know-how, and scalability [150], other fields still need improvement.

1342 From a short- to mid-term perspective, improving technical aspects of bioavailability and species
 1343 sensitivity in receptor-ligand interaction modelling is most promising. As elaborated in this study, the
 1344 computation of actual bioavailable concentrations in *in vitro* test systems via mass-balance equations
 1345 [47], [48] has been shown to be accurate [46], [59], at least for certain compound classes, and is now
 1346 supported by tools like the UFZ-LSER [45]. Based on our experience, an empirical assessment of non-
 1347 standard reporter assays and cell culture medium compositions, in the context of their partition
 1348 coefficient parameters, would be highly beneficial for making these applications more robust and less
 1349 variable across multiple species.

1350 Additionally, molecular docking has demonstrated good accuracy [43], [151] in reproducing
 1351 experimental ligand-binding poses, such as the recently revealed hAhR cryo-EM structure [36], as
 1352 well as in predicting AlphaFold2-predicted structures [137]. Extending the state of the science to
 1353 environmentally or technically relevant species, such as trout, zebrafish, mice, and rats, would
 1354 further enrich our understanding.

1355 Long-term challenges include addressing limited biotransformation capacities in *in vitro* systems, a
 1356 persistent issue in *in vitro* toxicology [152], [153]. As such, all employed hepatocyte cell lines,
 1357 including ZFL [154], [155], HepG2 [156], [157], and Hepa-1c1c7/DRE [158], [159], have been reported
 1358 to preserve at least some biotransformation capacity, either directly or indirectly. On the other hand,
 1359 we are unaware of any retained biotransformation capacities in the cell lines used for the oxidative
 1360 stress MoA, ZF4, and MCF-7. As such, this aligns with general accounts indicating that non-hepatic,
 1361 stable cell lines completely lose biotransformation capabilities over prolonged culture.

1362 Indeed, the observed variance in induction patterns for the xenobiotic metabolism MoA could be
 1363 attributed to differences in biotransformation capacity. Understanding these disparities is crucial, as
 1364 they can significantly affect induction patterns and activity. Regrettably, there is no universally
 1365 accepted standard for comparing biotransformation capacities across different cell lines [160], [161].
 1366 Furthermore, many stable hepatocyte-like cell lines have not been evaluated in this context.

1367 Unfortunately, conducting a thorough assessment of the biotransformation capacities of the cell lines
 1368 utilised in this study would have exceeded its scope and necessitated a dedicated investigation.

1369 However, we emphasise the importance of establishing standardised parameters for such
 1370 assessments, particularly for future cross-species evaluations of hepatocyte-like cultures.

1371 Another aspect we discussed throughout the article and supplementary material is signal
 1372 transduction, especially in fish, with a focus on non-specific toxicity pathways. Primarily due to

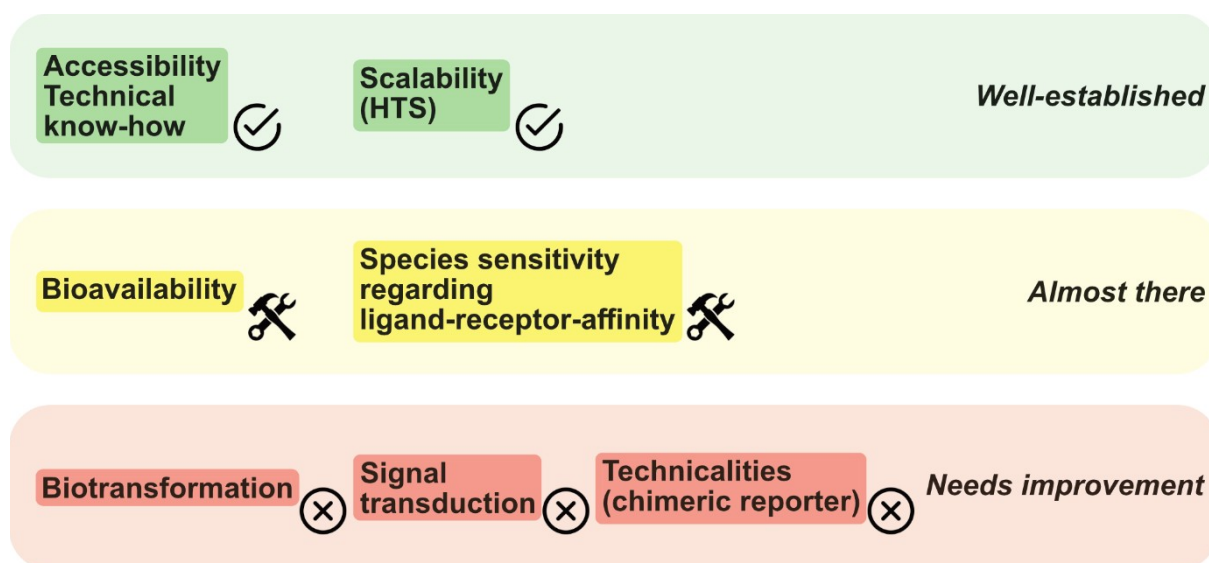
1373 genome duplication events [42], multiple transcription factors, translocators, co-factor isoforms, and
1374 the resulting stochastic sequestration of the ligand, the details of ligand-to-signal modulation remain
1375 an enigma. Potentially, CRISPR-mediated knockout studies could provide insights into these intricate
1376 mechanisms.

1377 Finally, the aspect of the chimeric reporter gene assays. Fundamentally, none of the commonly
1378 deployed reporter gene bioassays can be purely defined as a single-species assay, since all reporter
1379 constructs contain chimeric elements. Typically, reporter gene design employs one of two strategies.
1380 The first utilises the cell's endogenous signal transduction, transcription, and translation machinery,
1381 introducing only a foreign reporter construct. The second approach involves using cell lines deficient
1382 in specific components of these machineries, which are subsequently supplemented (e.g., [162]–
1383 [164]). Regardless of the strategy, the resulting reporter lines are chimeric.

1384 All reporter gene assays in this study leverage the cells' innate signal transduction, transcription, and
1385 translation mechanisms. However, the interaction between the transcription factor complex and the
1386 reporter gene construct is facilitated by chimeric gene-regulatory elements. Specifically, the
1387 pGudLuc7.5 reporter construct [13], [95] is regulated by 20 XRE sequences from the murine CYP1A1
1388 gene inserted into a modified viral MMTVLTR promoter. Similarly, our ZFL assay, although
1389 incorporating a zebrafish-derived genomic promoter (zfEF1aPro) to enhance performance, remains
1390 chimeric [12]. The DRE assay, driven by seven XREs from the C57BL/6 mouse strain's CYP1A1 gene,
1391 utilises a minimal heat shock promoter (mHSP) of uncertain origin [17], traditionally sourced from
1392 *Drosophila* or *E. coli* [165], [166]. The ZF4 assay employs a synthetic ARE consensus sequence and a
1393 synthetic minimal promoter in the pGL4.37[luc2P/ARE/Hygro] reporter vector. The AREc32-related
1394 construct, PGL8x-ARE, combines eight AREs from rat *gsta2* and mouse *gsta1* genes, driven by a viral
1395 SV40 promoter [10].

1396 We refer to the existing literature for a detailed discussion on the issues and pitfalls associated with
1397 transient or stable chimeric reporter constructs [8], [12], [167]. While some reports show the
1398 functional interchangeability of chimeric constructs across species due to mechanistic conservation
1399 [97], [168], others document specific experimental failures [163]. Determining the impact of chimeric
1400 constructs on the validity of species-related data remains challenging, particularly given the indistinct
1401 origins of some cloning components. Instead, we propose reevaluating the concept of species-
1402 specific reporter bioassays and advocating the adoption of novel tools for assay development. One
1403 such promising approach is CRISPR-Cas-mediated knock-ins, which allow site-directed integration of
1404 luminescent or fluorescent reporters adjacent to target genes (Lungu-Mitea et al., in preparation).
1405 This method ensures that reporter gene activation is entirely endogenous, potentially circumventing

1406 unintended effects caused by classical transgenesis and by random integration of reporter constructs
1407 into the host genome via bacterial plasmid vectors.



1408

1409 Fig. S44: Outlook on short to long-term challenges for fully integrating multi-species reporter bioassays into environmental
1410 monitoring.

1411 5. References

- 1412 [1] OECD, *Test No. 236: Fish Embryo Acute Toxicity (FET) Test*, no. July. OECD Publishing, 2025.
- 1413 [2] B. I. Escher *et al.*, “Benchmarking Organic Micropollutants in Wastewater, Recycled Water and
1414 Drinking Water with In Vitro Bioassays,” *Environ. Sci. Technol.*, vol. 48, no. 3, pp. 1940–1956,
1415 Feb. 2014, doi: 10.1021/es403899t.
- 1416 [3] O. Golovko *et al.*, “Occurrence and removal of chemicals of emerging concern in wastewater
1417 treatment plants and their impact on receiving water systems,” *Sci. Total Environ.*, vol. 754, p.
1418 142122, Feb. 2021, doi: 10.1016/j.scitotenv.2020.142122.
- 1419 [4] K. Itoh, K. I. Tong, and M. Yamamoto, “Molecular mechanism activating Nrf2-Keap1 pathway
1420 in regulation of adaptive response to electrophiles,” *Free Radic. Biol. Med.*, vol. 36, no. 10,
1421 pp. 1208–13, May 2004, doi: 10.1016/j.freeradbiomed.2004.02.075.
- 1422 [5] E. J. Wright, K. Pereira De Castro, A. D. Joshi, and C. J. Elferink, “Canonical and non-canonical
1423 aryl hydrocarbon receptor signaling pathways,” *Curr. Opin. Toxicol.*, vol. 2, pp. 87–92, Feb.
1424 2017, doi: 10.1016/j.cotox.2017.01.001.
- 1425 [6] W. Driever and Z. Rangini, “Characterization of a cell line derived from zebrafish (*Brachydanio*
1426 *rerio*) embryos,” *In Vitro Cell. Dev. Biol. Anim.*, vol. 29A, no. 9, pp. 749–54, Sep. 1993,
1427 [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8407719>.
- 1428 [7] S. Lungu-Mitea, A. Oskarsson, and J. Lundqvist, “Development of an oxidative stress in vitro
1429 assay in zebrafish (*Danio rerio*) cell lines,” *Sci. Rep.*, vol. 8, no. 1, p. 12380, Dec. 2018, doi:
1430 10.1038/s41598-018-30880-1.
- 1431 [8] S. Lungu-Mitea and J. Lundqvist, “Potentials and pitfalls of transient in vitro reporter
1432 bioassays: interference by vector geometry and cytotoxicity in recombinant zebrafish cell
1433 lines,” *Arch. Toxicol.*, no. 123456789, May 2020, doi: 10.1007/s00204-020-02783-6.

- 1434 [9] H. D. Soule, J. Vazquez, A. Long, S. Albert, and M. Brennan, "A Human Cell Line From a Pleural
1435 Effusion Derived From a Breast Carcinoma 2," *JNCI J. Natl. Cancer Inst.*, vol. 51, no. 5, pp.
1436 1409–1416, Nov. 1973, doi: 10.1093/jnci/51.5.1409.
- 1437 [10] X. J. Wang, J. D. Hayes, and C. R. Wolf, "Generation of a Stable Antioxidant Response
1438 Element–Driven Reporter Gene Cell Line and Its Use to Show Redox-Dependent Activation of
1439 Nrf2 by Cancer Chemotherapeutic Agents," *Cancer Res.*, vol. 66, no. 22, pp. 10983–10994,
1440 Nov. 2006, doi: 10.1158/0008-5472.CAN-06-2298.
- 1441 [11] C. Ghosh, Y. L. Zhou, and P. Collodi, "Derivation and characterization of a zebrafish liver cell
1442 line," *Cell Biol. Toxicol.*, vol. 10, no. 3, pp. 167–176, 1994, doi: 10.1007/BF00757560.
- 1443 [12] S. Lungu-Mitea, Y. Han, and J. Lundqvist, "Development, scrutiny, and modulation of transient
1444 reporter gene assays of the xenobiotic metabolism pathway in zebrafish hepatocytes," *Cell*
1445 *Biol. Toxicol.*, vol. 39, no. 3, pp. 991–1013, Jun. 2023, doi: 10.1007/s10565-021-09659-0.
- 1446 [13] G. He *et al.*, "Third-Generation Ah Receptor–Responsive Luciferase Reporter Plasmids:
1447 Amplification of Dioxin-Responsive Elements Dramatically Increases CALUX Bioassay
1448 Sensitivity and Responsiveness," *Toxicol. Sci.*, vol. 123, no. 2, pp. 511–522, Oct. 2011, doi:
1449 10.1093/toxsci/kfr189.
- 1450 [14] D. P. Aden, A. Fogel, S. Plotkin, I. Damjanov, and B. B. Knowles, "Controlled synthesis of HBsAg
1451 in a differentiated human liver carcinoma-derived cell line," *Nature*, vol. 282, no. 5739, pp.
1452 615–616, Dec. 1979, doi: 10.1038/282615a0.
- 1453 [15] J. C. Brennan *et al.*, "Development of Species-Specific Ah Receptor-Responsive Third
1454 Generation CALUX Cell Lines with Enhanced Responsiveness and Improved Detection Limits,"
1455 *Environ. Sci. Technol.*, vol. 49, no. 19, pp. 11903–11912, Sep. 2015, doi:
1456 10.1021/acs.est.5b02906.
- 1457 [16] A. K. Rosenmai *et al.*, "In vitro bioanalysis of drinking water from source to tap," *Water Res.*,
1458 vol. 139, pp. 272–280, 2018, doi: 10.1016/j.watres.2018.04.009.
- 1459 [17] S. Takeuchi, M. Iida, H. Yabushita, T. Matsuda, and H. Kojima, "In vitro screening for aryl
1460 hydrocarbon receptor agonistic activity in 200 pesticides using a highly sensitive reporter cell
1461 line, DR-EcoScreen cells, and in vivo mouse liver cytochrome P450-1A induction by propanil,
1462 diuron and linuron," *Chemosphere*, vol. 74, no. 1, pp. 155–165, Dec. 2008, doi:
1463 10.1016/j.chemosphere.2008.08.015.
- 1464 [18] H. P. Bernhard, G. J. Darlington, and F. H. Ruddle, "Expression of liver phenotypes in cultured
1465 mouse hepatoma cells: Synthesis and secretion of serum albumin," *Dev. Biol.*, vol. 35, no. 1,
1466 pp. 83–96, Nov. 1973, doi: 10.1016/0012-1606(73)90008-0.
- 1467 [19] G. Carlsson, J. Patring, J. Kreuger, L. Norrgren, and A. Oskarsson, "Toxicity of 15 veterinary
1468 pharmaceuticals in zebrafish (*Danio rerio*) embryos," *Aquat. Toxicol.*, vol. 126, pp. 30–41, Jan.
1469 2013, doi: 10.1016/j.aquatox.2012.10.008.
- 1470 [20] D. L. Filer, P. Kothiya, R. W. Setzer, R. S. Judson, and M. T. Martin, "tcpl: the ToxCast pipeline
1471 for high-throughput screening data," *Bioinformatics*, vol. 33, no. 4, pp. 618–620, Feb. 2017,
1472 doi: 10.1093/bioinformatics/btw680.
- 1473 [21] O. Jeroen, D. James, S. DebRoy, H. Wickham, and J. Horner, "RMySQL: Database Interface and
1474 'MySQL' Driver for R." 2025, [Online]. Available: <https://github.com/r-dbi/rmysql>.
- 1475 [22] A. Abdelaziz, H. Spahn-Langguth, K.-W. Schramm, and I. V. Tetko, "Consensus Modeling for
1476 HTS Assays Using In silico Descriptors Calculates the Best Balanced Accuracy in Tox21
1477 Challenge," *Front. Environ. Sci.*, vol. 4, no. FEB, pp. 1–12, Feb. 2016, doi:

1478 10.3389/fenvs.2016.00002.

1479 [23] I. Sushko *et al.*, "Online chemical modeling environment (OCHEM): web platform for data
1480 storage, model development and publishing of chemical information," *J. Comput. Aided. Mol.*
1481 *Des.*, vol. 25, no. 6, pp. 533–554, Jun. 2011, doi: 10.1007/s10822-011-9440-2.

1482 [24] S. Kim *et al.*, "PubChem in 2021: new data content and improved web interfaces," *Nucleic*
1483 *Acids Res.*, vol. 49, no. D1, pp. D1388–D1395, Jan. 2021, doi: 10.1093/nar/gkaa971.

1484 [25] E. W. Sayers *et al.*, "Database resources of the national center for biotechnology information,"
1485 *Nucleic Acids Res.*, vol. 50, no. D1, pp. D20–D26, Jan. 2022, doi: 10.1093/nar/gkab1112.

1486 [26] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison,
1487 "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform," *J.*
1488 *Cheminform.*, vol. 4, no. 1, p. 17, Dec. 2012, doi: 10.1186/1758-2946-4-17.

1489 [27] E. Vanqualef *et al.*, "R.E.D. Server: a web service for deriving RESP and ESP charges and
1490 building force field libraries for new molecules and molecular fragments," *Nucleic Acids Res.*,
1491 vol. 39, no. suppl_2, pp. W511–W517, Jul. 2011, doi: 10.1093/nar/gkr288.

1492 [28] G. M. Morris *et al.*, "AutoDock4 and AutoDockTools4: Automated docking with selective
1493 receptor flexibility," *J. Comput. Chem.*, vol. 30, no. 16, pp. 2785–2791, Dec. 2009, doi:
1494 10.1002/jcc.21256.

1495 [29] M. F. Sanner, "Python: a programming language for software integration and development.,"
1496 *J. Mol. Graph. Model.*, vol. 17, no. 1, pp. 57–61, Feb. 1999, [Online]. Available:
1497 <http://www.ncbi.nlm.nih.gov/pubmed/10660911>.

1498 [30] M. F. Sanner, A. J. Olson, and J.-C. Spehner, "Reduced surface: An efficient way to compute
1499 molecular surfaces," *Biopolymers*, vol. 38, no. 3, pp. 305–320, Mar. 1996, doi:
1500 10.1002/(SICI)1097-0282(199603)38:3<305::AID-BIP4>3.0.CO;2-Y.

1501 [31] J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol.
1502 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.

1503 [32] M. Varadi *et al.*, "AlphaFold Protein Structure Database: massively expanding the structural
1504 coverage of protein-sequence space with high-accuracy models," *Nucleic Acids Res.*, vol. 50,
1505 no. D1, pp. D439–D444, Jan. 2022, doi: 10.1093/nar/gkab1061.

1506 [33] A. Bateman *et al.*, "UniProt: the universal protein knowledgebase in 2021," *Nucleic Acids Res.*,
1507 vol. 49, no. D1, pp. D480–D489, Jan. 2021, doi: 10.1093/nar/gkaa1100.

1508 [34] H. M. Berman, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, Jan.
1509 2000, doi: 10.1093/nar/28.1.235.

1510 [35] S. K. Burley *et al.*, "RCSB Protein Data Bank: powerful new tools for exploring 3D structures of
1511 biological macromolecules for basic and applied research and education in fundamental
1512 biology, biomedicine, biotechnology, bioengineering and energy sciences," *Nucleic Acids Res.*,
1513 vol. 49, no. D1, pp. D437–D451, Jan. 2021, doi: 10.1093/nar/gkaa1038.

1514 [36] J. Gruszczyk *et al.*, "Cryo-EM structure of the agonist-bound Hsp90-XAP2-AHR cytosolic
1515 complex," *Nat. Commun.*, vol. 13, no. 1, p. 7010, Nov. 2022, doi: 10.1038/s41467-022-34773-
1516 w.

1517 [37] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, "ColabFold:
1518 making protein folding accessible to all," *Nat. Methods*, vol. 19, no. 6, pp. 679–682, Jun. 2022,
1519 doi: 10.1038/s41592-022-01488-1.

- 1520 [38] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev, "H++ 3.0: automating pK prediction and the
1521 preparation of biomolecular structures for atomistic molecular modeling and simulations,"
1522 *Nucleic Acids Res.*, vol. 40, no. W1, pp. W537–W541, Jul. 2012, doi: 10.1093/nar/gks375.
- 1523 [39] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, and A. Onufriev, "H++: a server for
1524 estimating pKas and adding missing hydrogens to macromolecules," *Nucleic Acids Res.*, vol.
1525 33, no. Web Server, pp. W368–W371, Jul. 2005, doi: 10.1093/nar/gki464.
- 1526 [40] K. Okonechnikov, O. Golosova, and M. Fursov, "Unipro UGENE: a unified bioinformatics
1527 toolkit," *Bioinformatics*, vol. 28, no. 8, pp. 1166–1167, Apr. 2012, doi:
1528 10.1093/bioinformatics/bts091.
- 1529 [41] W. Liu *et al.*, "IBS: an illustrator for the presentation and visualization of biological
1530 sequences," *Bioinformatics*, vol. 31, no. 20, pp. 3359–3361, Oct. 2015, doi:
1531 10.1093/bioinformatics/btv362.
- 1532 [42] S. M. K. Glasauer and S. C. F. Neuhauss, "Whole-genome duplication in teleost fishes and its
1533 evolutionary consequences," *Mol. Genet. Genomics*, vol. 289, no. 6, pp. 1045–1060, Dec.
1534 2014, doi: 10.1007/s00438-014-0889-2.
- 1535 [43] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a
1536 new scoring function, efficient optimization, and multithreading," *J. Comput. Chem.*, vol. 31,
1537 no. 2, pp. 455–461, Jan. 2010, doi: 10.1002/jcc.21334.
- 1538 [44] D. Seeliger and B. L. de Groot, "Ligand docking and binding site analysis with PyMOL and
1539 Autodock/Vina," *J. Comput. Aided. Mol. Des.*, vol. 24, no. 5, pp. 417–422, May 2010, doi:
1540 10.1007/s10822-010-9352-6.
- 1541 [45] Ulrich *et al.*, "UFZ-LSER database v 3.2 [Internet]." Leipzig, Deutschland, Helmholtz Zentrum
1542 für Umweltforschung - UFZ, 2017, [Online]. Available: <http://www.ufz.de/lserd>.
- 1543 [46] F. C. Fischer *et al.*, "Modeling Exposure in the Tox21 in Vitro Bioassays," *Chem. Res. Toxicol.*,
1544 vol. 30, no. 5, pp. 1197–1208, May 2017, doi: 10.1021/acs.chemrestox.7b00023.
- 1545 [47] S. Endo, B. I. Escher, and K.-U. Goss, "Capacities of Membrane Lipids to Accumulate Neutral
1546 Organic Chemicals," *Environ. Sci. Technol.*, vol. 45, no. 14, pp. 5912–5921, Jul. 2011, doi:
1547 10.1021/es200855w.
- 1548 [48] S. Endo and K.-U. Goss, "Serum Albumin Binding of Structurally Diverse Neutral Organic
1549 Compounds: Data and Models," *Chem. Res. Toxicol.*, vol. 24, no. 12, pp. 2293–2301, Dec.
1550 2011, doi: 10.1021/tx200431b.
- 1551 [49] B. I. Escher, C. van Daele, M. Dutt, J. Y. M. Tang, and R. Altenburger, "Most Oxidative Stress
1552 Response In Water Samples Comes From Unknown Chemicals: The Need For Effect-Based
1553 Water Quality Trigger Values," *Environ. Sci. Technol.*, vol. 47, no. 13, pp. 7002–7011, Jul. 2013,
1554 doi: 10.1021/es304793h.
- 1555 [50] B. I. Escher, P. Neale, and F. Leusch, *Bioanalytical tools in water quality assessment Beate*
1556 *Escher, Peta Neale and Frederic Leusch*, 2nd editio. London: IWA Publishing, 2021.
- 1557 [51] M. Weimer, X. Jiang, O. Ponta, S. Stanzel, A. Freyberger, and A. Kopp-Schneider, "The impact
1558 of data transformations on concentration–response modeling," *Toxicol. Lett.*, vol. 213, no. 2,
1559 pp. 292–298, Sep. 2012, doi: 10.1016/j.toxlet.2012.07.012.
- 1560 [52] P. A. Neale *et al.*, "Linking in Vitro Effects and Detected Organic Micropollutants in Surface
1561 Water Using Mixture-Toxicity Modeling," *Environ. Sci. Technol.*, vol. 49, no. 24, pp. 14614–
1562 14624, Dec. 2015, doi: 10.1021/acs.est.5b04083.

- 1563 [53] P. A. Neale *et al.*, "Integrating chemical analysis and bioanalysis to evaluate the contribution
1564 of wastewater effluent on the micropollutant burden in small streams," *Sci. Total Environ.*,
1565 vol. 576, pp. 785–795, Jan. 2017, doi: 10.1016/j.scitotenv.2016.10.141.
- 1566 [54] P. A. Neale *et al.*, "Assessing the Mixture Effects in In Vitro Bioassays of Chemicals Occurring in
1567 Small Agricultural Streams during Rain Events," *Environ. Sci. Technol.*, vol. 54, no. 13, pp.
1568 8280–8290, Jul. 2020, doi: 10.1021/acs.est.0c02235.
- 1569 [55] B. I. Escher *et al.*, "Water quality assessment using the AREc32 reporter gene assay indicative
1570 of the oxidative stress response pathway," *J. Environ. Monit.*, vol. 14, no. 11, pp. 2877–2885,
1571 2012, doi: 10.1039/c2em30506b.
- 1572 [56] B. I. Escher, P. A. Neale, and D. L. Villeneuve, "The advantages of linear concentration-
1573 response curves for in vitro bioassays with environmental samples," *Environ. Toxicol. Chem.*,
1574 vol. 37, no. 9, pp. 2273–2280, Sep. 2018, doi: 10.1002/etc.4178.
- 1575 [57] M. Wagner, E. L. M. Vermeirssen, S. Buchinger, M. Behr, A. Magdeburg, and J. Oehlmann,
1576 "Deriving bio-equivalents from in vitro bioassays: Assessment of existing uncertainties and
1577 strategies to improve accuracy and reporting," *Environ. Toxicol. Chem.*, vol. 32, no. 8, pp.
1578 1906–1917, Aug. 2013, doi: 10.1002/etc.2256.
- 1579 [58] E. Hommel *et al.*, "Following the Mixtures of Organic Micropollutants with In Vitro Bioassays
1580 in a Large Lowland River from Source to Sea," *ACS Environ. Au*, vol. 5, no. 4, pp. 363–375,
1581 2025, doi: 10.1021/acsenvironau.4c00059.
- 1582 [59] S. Lungu-Mitea *et al.*, "Modeling Bioavailable Concentrations in Zebrafish Cell Lines and
1583 Embryos Increases the Correlation of Toxicity Potencies across Test Systems," *Environ. Sci.*
1584 *Technol.*, vol. 55, no. 1, pp. 447–457, Jan. 2021, doi: 10.1021/acs.est.0c04872.
- 1585 [60] M. Kobayashi *et al.*, "The antioxidant defense system Keap1-Nrf2 comprises a multiple sensing
1586 mechanism for responding to a wide range of chemical compounds," *Mol. Cell. Biol.*, vol. 29,
1587 no. 2, pp. 493–502, Jan. 2009, doi: 10.1128/MCB.01080-08.
- 1588 [61] K. Takaya *et al.*, "Validation of the multiple sensor mechanism of the Keap1-Nrf2 system,"
1589 *Free Radic. Biol. Med.*, vol. 53, no. 4, pp. 817–827, Aug. 2012, doi:
1590 10.1016/j.freeradbiomed.2012.06.023.
- 1591 [62] A. Slaninova, M. Smutna, H. Modra, and Z. Svobodova, "A review: oxidative stress in fish
1592 induced by pesticides," *Neuro Endocrinol. Lett.*, vol. 30 Suppl 1, no. April, pp. 2–12, 2009, doi:
1593 NEL300709R01 [pii].
- 1594 [63] K. Itoh, K. Igarashi, N. Hayashi, M. Nishizawa, and M. Yamamoto, "Cloning and
1595 characterization of a novel erythroid cell-derived CNC family transcription factor
1596 heterodimerizing with the small Maf family proteins," *Mol. Cell. Biol.*, vol. 15, no. 8, pp. 4184–
1597 93, Aug. 1995, [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7623813>.
- 1598 [64] P. Moi, K. Chan, I. Asunis, A. Cao, and Y. W. Kan, "Isolation of NF-E2-related factor 2 (Nrf2), a
1599 NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1
1600 repeat of the beta-globin locus control region," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 21,
1601 pp. 9926–30, Oct. 1994, doi: 10.1073/pnas.91.21.9926.
- 1602 [65] J. D. Hayes and A. T. Dinkova-Kostova, "The Nrf2 regulatory network provides an interface
1603 between redox and intermediary metabolism," *Trends Biochem. Sci.*, vol. 39, no. 4, pp. 199–
1604 218, Apr. 2014, doi: 10.1016/j.tibs.2014.02.002.
- 1605 [66] K. Itoh, P. Ye, T. Matsumiya, K. Tanji, and T. Ozaki, "Emerging functional cross-talk between
1606 the Keap1-Nrf2 system and mitochondria," *J. Clin. Biochem. Nutr.*, vol. 56, no. 2, pp. 91–7,

- 1607 Mar. 2015, doi: 10.3164/jcbrn.14-134.
- 1608 [67] M. Kobayashi and M. Yamamoto, "Nrf2-Keap1 regulation of cellular defense mechanisms
1609 against electrophiles and reactive oxygen species.," *Adv. Enzyme Regul.*, vol. 46, no. 1, pp.
1610 113–40, 2006, doi: 10.1016/j.advenzreg.2006.01.007.
- 1611 [68] T. Nguyen, P. J. Sherratt, and C. B. Pickett, "Regulatory mechanisms controlling gene
1612 expression mediated by the antioxidant response element.," *Annu. Rev. Pharmacol. Toxicol.*,
1613 vol. 43, no. 1, pp. 233–60, 2003, doi: 10.1146/annurev.pharmtox.43.100901.140229.
- 1614 [69] A. K. Sharma, G. Taneja, D. Khanna, and S. K. Rajput, "Reactive oxygen species: friend or foe?,"
1615 *RSC Adv.*, vol. 5, no. 71, pp. 57267–57276, 2015, doi: 10.1039/C5RA07927F.
- 1616 [70] T. H. Rushmore, M. R. Morton, and C. B. Pickett, "The antioxidant responsive element.
1617 Activation by oxidative stress and identification of the DNA consensus sequence required for
1618 functional activity.," *J. Biol. Chem.*, vol. 266, no. 18, pp. 11632–9, Jun. 1991, [Online].
1619 Available: <http://www.ncbi.nlm.nih.gov/pubmed/1646813>.
- 1620 [71] W. W. Wasserman and W. E. Fahl, "Functional antioxidant responsive elements.," *Proc. Natl.*
1621 *Acad. Sci. U. S. A.*, vol. 94, no. 10, pp. 5361–6, May 1997, [Online]. Available:
1622 <http://www.ncbi.nlm.nih.gov/pubmed/9144242>.
- 1623 [72] C. Köhle and K. W. Bock, "Coordinate regulation of Phase I and II xenobiotic metabolisms by
1624 the Ah receptor and Nrf2.," *Biochem. Pharmacol.*, vol. 73, no. 12, pp. 1853–62, Jun. 2007, doi:
1625 10.1016/j.bcp.2007.01.009.
- 1626 [73] K. Itoh *et al.*, "Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2
1627 through binding to the amino-terminal Neh2 domain.," *Genes Dev.*, vol. 13, no. 1, pp. 76–86,
1628 Jan. 1999, doi: 10.1101/gad.13.1.76.
- 1629 [74] K. Itoh, T. Ishii, N. Wakabayashi, and M. Yamamoto, "Regulatory mechanisms of cellular
1630 response to oxidative stress.," *Free Radic. Res.*, vol. 31, no. 4, pp. 319–24, Oct. 1999, doi:
1631 10.1080/10715769900300881.
- 1632 [75] M. Kobayashi *et al.*, "Identification of the interactive interface and phylogenetic conservation of
1633 the Nrf2-Keap1 system.," *Genes Cells*, vol. 7, no. 8, pp. 807–20, Aug. 2002, doi:
1634 10.1046/j.1365-2443.2002.00561.x.
- 1635 [76] M. Xiang, A. Namani, S. Wu, and X. Wang, "Nrf2: bane or blessing in cancer?," *J. Cancer Res.*
1636 *Clin. Oncol.*, vol. 140, no. 8, pp. 1251–9, Aug. 2014, doi: 10.1007/s00432-014-1627-1.
- 1637 [77] T. Yamamoto *et al.*, "Physiological significance of reactive cysteine residues of Keap1 in
1638 determining Nrf2 activity.," *Mol. Cell. Biol.*, vol. 28, no. 8, pp. 2758–70, Apr. 2008, doi:
1639 10.1128/MCB.01704-07.
- 1640 [78] A. T. Dinkova-Kostova, R. V. Kostov, and P. Canning, "Keap1, the cysteine-based mammalian
1641 intracellular sensor for electrophiles and oxidants," *Arch. Biochem. Biophys.*, vol. 617, pp. 84–
1642 93, Mar. 2017, doi: 10.1016/j.abb.2016.08.005.
- 1643 [79] S. Fourquet, R. Guerois, D. Biard, and M. B. Toledano, "Activation of NRF2 by nitrosative
1644 agents and H2O2 involves KEAP1 disulfide formation.," *J. Biol. Chem.*, vol. 285, no. 11, pp.
1645 8463–71, Mar. 2010, doi: 10.1074/jbc.M109.051714.
- 1646 [80] K. Igarashi, K. Kataokata, K. Itoh, N. Hayashi, M. Nishizawa, and M. Yamamoto, "Regulation of
1647 transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins," *Nature*,
1648 vol. 367, no. 6463, pp. 568–572, Feb. 1994, doi: 10.1038/367568a0.
- 1649 [81] a T. Dinkova-Kostova, M. a Massiah, R. E. Bozak, R. J. Hicks, and P. Talalay, "Potency of

1650 Michael reaction acceptors as inducers of enzymes that protect against carcinogenesis
 1651 depends on their reactivity with sulfhydryl groups.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no.
 1652 6, pp. 3404–9, Mar. 2001, doi: 10.1073/pnas.051632198.

1653 [82] N. Wakabayashi *et al.*, "Protection against electrophile and oxidant stress by induction of the
 1654 phase 2 response: fate of cysteines of the Keap1 sensor modified by inducers.," *Proc. Natl.*
 1655 *Acad. Sci. U. S. A.*, vol. 101, no. 7, pp. 2040–5, Feb. 2004, doi: 10.1073/pnas.0307301101.

1656 [83] R. Saito *et al.*, "Characterizations of Three Major Cysteine Sensors of Keap1 in Stress
 1657 Response.," *Mol. Cell. Biol.*, vol. 36, no. 2, pp. 271–84, Nov. 2015, doi: 10.1128/MCB.00868-
 1658 15.

1659 [84] D. D. Zhang and M. Hannink, "Distinct cysteine residues in Keap1 are required for Keap1-
 1660 dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and
 1661 oxidative stress.," *Mol. Cell. Biol.*, vol. 23, no. 22, pp. 8137–51, Nov. 2003, doi:
 1662 10.1128/MCB.23.22.8137-8151.2003.

1663 [85] K. Taguchi, H. Motohashi, and M. Yamamoto, "Molecular mechanisms of the Keap1-Nrf2
 1664 pathway in stress response and cancer evolution," *Genes to Cells*, vol. 16, no. 2, pp. 123–140,
 1665 Feb. 2011, doi: 10.1111/j.1365-2443.2010.01473.x.

1666 [86] H. K. Bryan, A. Olayanju, C. E. Goldring, and B. K. Park, "The Nrf2 cell defence pathway: Keap1-
 1667 dependent and -independent mechanisms of regulation," *Biochem. Pharmacol.*, vol. 85, no. 6,
 1668 pp. 705–717, Mar. 2013, doi: 10.1016/j.bcp.2012.11.016.

1669 [87] Y. Horie *et al.*, "Molecular basis for the disruption of Keap1–Nrf2 interaction via Hinge & Latch
 1670 mechanism," *Commun. Biol.*, vol. 4, no. 1, pp. 1–11, 2021, doi: 10.1038/s42003-021-02100-6.

1671 [88] Y. Fuse and M. Kobayashi, "Conservation of the Keap1-Nrf2 System: An Evolutionary Journey
 1672 through Stressful Space and Time.," *Molecules*, vol. 22, no. 3, pp. 1–22, Mar. 2017, doi:
 1673 10.3390/molecules22030436.

1674 [89] M. Kobayashi and M. Yamamoto, "Molecular mechanisms activating the Nrf2-Keap1 pathway
 1675 of antioxidant gene regulation.," *Antioxid. Redox Signal.*, vol. 7, no. 3–4, pp. 385–94, 2005,
 1676 doi: 10.1089/ars.2005.7.385.

1677 [90] L. Li, M. Kobayashi, H. Kaneko, Y. Nakajima-Takagi, Y. Nakayama, and M. Yamamoto,
 1678 "Molecular evolution of Keap1. Two Keap1 molecules with distinctive intervening region
 1679 structures are conserved among fish.," *J. Biol. Chem.*, vol. 283, no. 6, pp. 3248–55, Feb. 2008,
 1680 doi: 10.1074/jbc.M708702200.

1681 [91] A. R. Timme-Laragy *et al.*, "Nrf2b, novel zebrafish paralog of oxidant-responsive transcription
 1682 factor NF-E2-related factor 2 (NRF2).," *J. Biol. Chem.*, vol. 287, no. 7, pp. 4609–27, Feb. 2012,
 1683 doi: 10.1074/jbc.M111.260125.

1684 [92] V. T. Nguyen *et al.*, "Generation and characterization of keap1a- and keap1b-knockout
 1685 zebrafish," *Redox Biol.*, vol. 36, no. July, p. 101667, Sep. 2020, doi:
 1686 10.1016/j.redox.2020.101667.

1687 [93] P. M. GARRISON, K. TULLIS, J. M. M. J. G. AARTS, A. BROUWER, J. P. GIESY, and M. S. DENISON,
 1688 "Species-Specific Recombinant Cell Lines as Bioassay Systems for the Detection of 2,3,7,8-
 1689 Tetrachlorodibenzo- p -dioxin-like Chemicals," *Toxicol. Sci.*, vol. 30, no. 2, pp. 194–203, 1996,
 1690 doi: 10.1093/toxsci/30.2.194.

1691 [94] P. M. Garrison, K. Tullis, J. M. Aarts, A. Brouwer, J. P. Giesy, and M. S. Denison, "Species-
 1692 specific recombinant cell lines as bioassay systems for the detection of 2,3,7,8-
 1693 tetrachlorodibenzo-p-dioxin-like chemicals.," *Fundam. Appl. Toxicol.*, vol. 30, no. 2, pp. 194–

- 1694 203, Apr. 1996, [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8812265>.
- 1695 [95] D. Han, S. R. Nagy, and M. S. Denison, "Comparison of recombinant cell bioassays for the
1696 detection of Ah receptor agonists," *BioFactors*, vol. 20, no. 1, pp. 11–22, 2004, doi:
1697 10.1002/biof.5520200102.
- 1698 [96] A. Murk, "Chemical-Activated Luciferase Gene Expression (CALUX): A Novel *In Vitro* Bioassay
1699 for Ah Receptor Active Compounds in Sediments and Pore Water," *Fundam. Appl. Toxicol.*,
1700 vol. 33, no. 1, pp. 149–160, Sep. 1996, doi: 10.1006/faat.1996.0152.
- 1701 [97] C. A. Richter, V. L. Tieber, M. S. Denison, and J. P. Giesy, "An *in vitro* rainbow trout cell
1702 bioassay for aryl hydrocarbon receptor-mediated toxins," *Environ. Toxicol. Chem.*, vol. 16, no.
1703 3, pp. 543–550, 1997, doi: 10.1897/1551-5028(1997)016<0543:AIVRTC>2.3.CO;2.
- 1704 [98] K. Eichbaum *et al.*, "In vitro bioassays for detecting dioxin-like activity — Application
1705 potentials and limits of detection, a review," *Sci. Total Environ.*, vol. 487, no. 1, pp. 37–48, Jul.
1706 2014, doi: 10.1016/j.scitotenv.2014.03.057.
- 1707 [99] M. H. El-Fouly, C. Richter, J. P. Giesy, and M. S. Denison, "Production of a novel recombinant
1708 cell line for use as a bioassay system for detection of 2,3,7,8-tetrachlorodibenzo- P -dioxin-like
1709 chemicals," *Environ. Toxicol. Chem.*, vol. 13, no. 10, pp. 1581–1588, Oct. 1994, doi:
1710 10.1002/etc.5620131006.
- 1711 [100] P. A. Neale *et al.*, "Application of Effect-Based Methods to Water Quality Monitoring:
1712 Answering Frequently Asked Questions by Water Quality Managers, Regulators, and Policy
1713 Makers," *Environ. Sci. Technol.*, vol. 57, no. 15, pp. 6023–6032, Apr. 2023, doi:
1714 10.1021/acs.est.2c06365.
- 1715 [101] P. A. Neale, B. I. Escher, M. L. de Baat, J. Enault, and F. D. L. Leusch, "Effect-Based Trigger
1716 Values Are Essential for the Uptake of Effect-Based Methods in Water Safety Planning,"
1717 *Environ. Toxicol. Chem.*, vol. 42, no. 3, pp. 714–726, Mar. 2023, doi: 10.1002/etc.5544.
- 1718 [102] R. van der Oost, G. Sileno, M. Suárez-Muñoz, M. T. Nguyen, H. Besselink, and A. Brouwer,
1719 "SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality
1720 assessment: Part I—model design and effect-based trigger values," *Environ. Toxicol. Chem.*, vol.
1721 36, no. 9, pp. 2385–2399, Sep. 2017, doi: 10.1002/etc.3836.
- 1722 [103] State Water Resources Control Board, "Water Quality Control Policy for Recycled Water State
1723 Water Resources Control Board California Environmental Protection Agency," 2018. [Online].
1724 Available: <http://www.waterboards.ca.gov>.
- 1725 [104] M. Carere, A.-S. Wernersson, T. Lettieri, and S. Buchinger, "Technical Proposal for Effect-Based
1726 Monitoring and Assessment under the Water Framework Directive," 2021. doi:
1727 10.13140/RG.2.2.19954.27847.
- 1728 [105] B. I. Escher *et al.*, "Effect-based trigger values for *in vitro* and *in vivo* bioassays performed on
1729 surface water extracts supporting the environmental quality standards (EQS) of the European
1730 Water Framework Directive," *Sci. Total Environ.*, vol. 628–629, pp. 748–765, 2018, doi:
1731 10.1016/j.scitotenv.2018.01.340.
- 1732 [106] B. I. Escher, P. A. Neale, and F. D. L. Leusch, "Effect-based trigger values for *in vitro* bioassays:
1733 Reading across from existing water quality guideline values," *Water Res.*, vol. 81, pp. 137–148,
1734 Sep. 2015, doi: 10.1016/j.watres.2015.05.049.
- 1735 [107] B. I. Escher *et al.*, "Effect-based trigger values for *in vitro* and *in vivo* bioassays performed on
1736 surface water extracts supporting the environmental quality standards (EQS) of the European
1737 Water Framework Directive," *Sci. Total Environ.*, vol. 628–629, pp. 748–765, Jul. 2018, doi:

- 10.1016/j.scitotenv.2018.01.340.
- [108] F. Tamaddon and W. Hogland, "Characterization and co-incineration of waste in a paper mill case study," *Linnaeus Eco-Tech*, pp. 699–708, Dec. 2007, doi: 10.15626/Eco-Tech.2007.076.
- [109] M. Almemark, G. Finnveden, and B. Frostell, "Treatment Technologies for Organochlorine-Containing Sludges and Concentrates from External Treatment of Pulp and Paper Wastewaters," *Water Sci. Technol.*, vol. 24, no. 3–4, pp. 319–329, Aug. 1991, doi: 10.2166/wst.1991.0488.
- [110] L. Morales *et al.*, "Oceanic Sink and Biogeochemical Controls on the Accumulation of Polychlorinated Dibenzo- p -dioxins, Dibenzofurans, and Biphenyls in Plankton," *Environ. Sci. Technol.*, vol. 49, no. 23, pp. 13853–13861, Dec. 2015, doi: 10.1021/acs.est.5b01360.
- [111] J. Castro-Jiménez *et al.*, "Atmospheric input of POPs into Lake Maggiore (Northern Italy): PCDD/F and dioxin-like PCB profiles and fluxes in the atmosphere and aquatic system," *Chemosphere*, vol. 73, no. 1, pp. S122–S130, Aug. 2008, doi: 10.1016/j.chemosphere.2007.06.097.
- [112] S. Lee *et al.*, "Measured and predicted affinities of binding and relative potencies to activate the AhR of PAHs and their alkylated analogues," *Chemosphere*, vol. 139, pp. 23–29, 2015, doi: 10.1016/j.chemosphere.2015.05.033.
- [113] M. M. Lam, R. Bülow, M. Engwall, J. P. Giesy, and M. Larsson, "Methylated PACs are more potent than their parent compounds: A study of aryl hydrocarbon receptor-mediated activity, degradability, and mixture interactions in the H4IIE-luc assay," *Environ. Toxicol. Chem.*, vol. 37, no. 5, pp. 1409–1419, May 2018, doi: 10.1002/etc.4087.
- [114] R. Lille-Langøy, K. B. Jørgensen, A. Goksøyr, D. M. Pampanin, M. O. Sydnes, and O. A. Karlsen, "Substituted Two- to Five-Ring Polycyclic Aromatic Compounds Are Potent Agonists of Atlantic Cod (*Gadus morhua*) Aryl Hydrocarbon Receptors Ahr1a and Ahr2a," *Environ. Sci. Technol.*, vol. 55, no. 22, pp. 15123–15135, Nov. 2021, doi: 10.1021/acs.est.1c02946.
- [115] A. R. Johnsen, L. Y. Wick, and H. Harms, "Principles of microbial PAH-degradation in soil," *Environ. Pollut.*, vol. 133, no. 1, pp. 71–84, Jan. 2005, doi: 10.1016/j.envpol.2004.04.015.
- [116] L. Goedtke *et al.*, "Polycyclic aromatic hydrocarbons activate the aryl hydrocarbon receptor and the constitutive androstane receptor to regulate xenobiotic metabolism in human liver cells," *Int. J. Mol. Sci.*, vol. 22, no. 1, pp. 1–17, 2021, doi: 10.3390/ijms22010372.
- [117] S. Giani Tagliabue, S. C. Faber, S. Motta, M. S. Denison, and L. Bonati, "Modeling the binding of diverse ligands within the Ah receptor ligand binding domain," *Sci. Rep.*, vol. 9, no. 1, p. 10693, Jul. 2019, doi: 10.1038/s41598-019-47138-z.
- [118] S. C. Monteriolo, A. di Domenico, V. Silano, G. Viviano, and G. Zapponi, "2,3,7,8-TCDD LEVELS AND DISTRIBUTION IN THE ENVIRONMENT AT SEVESO AFTER THE ICMESA ACCIDENT ON JULY 10th, 1976," in *Chlorinated Dioxins & Related Compounds*, Elsevier, 1982, pp. 127–136.
- [119] A. R. Batterman, P. M. Cook, K. B. Lodge, D. B. Lothenbach, and B. C. Butterworth, "Methodology used for a laboratory determination of relative contributions of water, sediment and food chain routes of uptake for 2,3,7,8-TCDD bioaccumulation by lake trout in Lake Ontario," *Chemosphere*, vol. 19, no. 1–6, pp. 451–458, Jan. 1989, doi: 10.1016/0045-6535(89)90351-2.
- [120] B. S. Giri, I. . Karimi, and M. . Ray, "Modeling and Monte Carlo simulation of TCDD transport in a river," *Water Res.*, vol. 35, no. 5, pp. 1263–1279, Apr. 2001, doi: 10.1016/S0043-1354(00)00379-1.

- 1782 [121] J. K. H. de Schepper, T. Slootweg, P. Behnisch, E. Felzel, and C. J. Houtman, "Beyond the
1783 Drinking Water Directive: The use of reporter gene assays as an added tool for effect-based
1784 monitoring of polycyclic aromatic hydrocarbons and polychlorinated biphenyls in drinking
1785 water sources," *Sci. Total Environ.*, vol. 941, no. May, p. 173366, Sep. 2024, doi:
1786 10.1016/j.scitotenv.2024.173366.
- 1787 [122] P. Y. Kunz *et al.*, "Effect-based tools for monitoring estrogenic mixtures: Evaluation of five in
1788 vitro bioassays," *Water Res.*, vol. 110, pp. 378–388, Mar. 2017, doi:
1789 10.1016/j.watres.2016.10.062.
- 1790 [123] L. Gómez *et al.*, "Estrogenicity of chemical mixtures revealed by a panel of bioassays," *Sci.*
1791 *Total Environ.*, vol. 785, p. 147284, Sep. 2021, doi: 10.1016/j.scitotenv.2021.147284.
- 1792 [124] E. B. Reiter, B. I. Escher, U. Siebert, and A. Jahnke, "Activation of the xenobiotic metabolism
1793 and oxidative stress response by mixtures of organic pollutants extracted with in-tissue
1794 passive sampling from liver, kidney, brain and blubber of marine mammals," *Environ. Int.*, vol.
1795 165, no. June, p. 107337, 2022, doi: 10.1016/j.envint.2022.107337.
- 1796 [125] M. Brinkmann *et al.*, "Multiple lines of evidence to assess risk from dioxins and dioxin-like
1797 chemicals in sediment and fish from waterbodies along a large prairie river," *J. Soils*
1798 *Sediments*, no. Kondolf 1997, Aug. 2023, doi: 10.1007/s11368-023-03620-w.
- 1799 [126] K. Hilscherova *et al.*, "Characterization of dioxin-like activity of sediments from a Czech River
1800 Basin," *Environ. Toxicol. Chem.*, vol. 20, no. 12, pp. 2768–2777, Dec. 2001, doi:
1801 10.1002/etc.5620201216.
- 1802 [127] S. Keiter *et al.*, "Activities and identification of aryl hydrocarbon receptor agonists in
1803 sediments from the Danube river," *Anal. Bioanal. Chem.*, vol. 390, no. 8, pp. 2009–2019, Apr.
1804 2008, doi: 10.1007/s00216-007-1652-x.
- 1805 [128] G. Liberatori *et al.*, "Occurrence and spatial distribution of dioxin and dioxin-like compounds
1806 in topsoil of Taranto (Apulia, Italy) by GC-MS analysis and DR-CALUX® bioassay,"
1807 *Chemosphere*, vol. 279, p. 130576, Sep. 2021, doi: 10.1016/j.chemosphere.2021.130576.
- 1808 [129] C. Budin *et al.*, "Detection of high PBDD/Fs levels and dioxin-like activity in toys using a
1809 combination of GC-HRMS, rat-based and human-based DR CALUX® reporter gene assays,"
1810 *Chemosphere*, vol. 251, p. 126579, Jul. 2020, doi: 10.1016/j.chemosphere.2020.126579.
- 1811 [130] P. A. Behnisch, "Screening of Dioxin-Like Toxicity Equivalents for Various Matrices with
1812 Wildtype and Recombinant Rat Hepatoma H4IIE Cells," *Toxicol. Sci.*, vol. 69, no. 1, pp. 125–
1813 130, Sep. 2002, doi: 10.1093/toxsci/69.1.125.
- 1814 [131] D. Delistraty, "Toxic equivalency factor approach for risk assessment of polycyclic aromatic
1815 hydrocarbons," *Toxicol. Environ. Chem.*, vol. 64, no. 1–4, pp. 81–108, Nov. 1997, doi:
1816 10.1080/02772249709358542.
- 1817 [132] S. P. Bhavsar, E. J. Reiner, A. Hayton, R. Fletcher, and K. MacPherson, "Converting Toxic
1818 Equivalents (TEQ) of dioxins and dioxin-like compounds in fish from one Toxic Equivalency
1819 Factor (TEF) scheme to another," *Environ. Int.*, vol. 34, no. 7, pp. 915–921, Oct. 2008, doi:
1820 10.1016/j.envint.2008.02.001.
- 1821 [133] B. I. Escher, N. Bramaz, J. F. Mueller, P. Quayle, S. Rutishauser, and E. L. M. Vermeirssen,
1822 "Toxic equivalent concentrations (TEQs) for baseline toxicity and specific modes of action as a
1823 tool to improve interpretation of ecotoxicity testing of environmental samples," *J. Environ.*
1824 *Monit.*, vol. 10, no. 5, p. 612, 2008, doi: 10.1039/b800949j.
- 1825 [134] J. A. Doering, D. E. Tillitt, and S. Wiseman, "Reevaluation of 2,3,7,8-Tetrachlorodibenzo- p

- 1826 -Dioxin Equivalency Factors for Dioxin-Like Polychlorinated Dibenzo- p -Dioxins,
1827 Polychlorinated Dibenzofurans, and Polychlorinated Biphenyls for Fishes," *Environ. Toxicol.*
1828 *Chem.*, vol. 42, no. 10, pp. 2215–2228, Oct. 2023, doi: 10.1002/etc.5690.
- 1829 [135] M. B. M. van Duursen, K. I. van Ede, and M. van den Berg, "One TEF concept does not fit all:
1830 The case for human risk assessment of polychlorinated biphenyls," *Curr. Opin. Toxicol.*, vol. 2,
1831 pp. 103–108, Feb. 2017, doi: 10.1016/j.cotox.2017.01.005.
- 1832 [136] W. H. Bisson *et al.*, "Modeling of the Aryl Hydrocarbon Receptor (AhR) Ligand Binding Domain
1833 and Its Utility in Virtual Ligand Screening to Predict New AhR Ligands," *J. Med. Chem.*, vol. 52,
1834 no. 18, pp. 5635–5641, Sep. 2009, doi: 10.1021/jm900199u.
- 1835 [137] A. P. Plonski and S. M. Reed, "Assessing protein homology models with docking
1836 reproducibility," *J. Mol. Graph. Model.*, vol. 121, no. February, p. 108430, Jun. 2023, doi:
1837 10.1016/j.jmglm.2023.108430.
- 1838 [138] G. ZeRuth and R. S. Pollenz, "Functional analysis of cis-regulatory regions within the dioxin-
1839 inducible CYP1A promoter/enhancer region from zebrafish (*Danio rerio*)," *Chem. Biol.*
1840 *Interact.*, vol. 170, no. 2, pp. 100–113, Nov. 2007, doi: 10.1016/j.cbi.2007.07.003.
- 1841 [139] J. V. Schmidt and C. A. Bradfield, "AH RECEPTOR SIGNALING PATHWAYS," *Annu. Rev. Cell Dev.*
1842 *Biol.*, vol. 12, no. 1, pp. 55–89, Nov. 1996, doi: 10.1146/annurev.cellbio.12.1.55.
- 1843 [140] S. O. Simmons, C.-Y. Fan, and R. Ramabhadran, "Cellular Stress Response Pathway System as a
1844 Sentinel Ensemble in Toxicological Screening," *Toxicol. Sci.*, vol. 111, no. 2, pp. 202–225, Oct.
1845 2009, doi: 10.1093/toxsci/kfp140.
- 1846 [141] L. A. Van Tiem and R. T. Di Giulio, "AHR2 knockdown prevents PAH-mediated cardiac toxicity
1847 and XRE- and ARE-associated gene induction in zebrafish (*Danio rerio*)," *Toxicol. Appl.*
1848 *Pharmacol.*, vol. 254, no. 3, pp. 280–287, Aug. 2011, doi: 10.1016/j.taap.2011.05.002.
- 1849 [142] A. Ulin *et al.*, "Developmental Regulation of Nuclear Factor Erythroid-2 Related Factors (nrfs)
1850 by AHR1b in Zebrafish (*Danio rerio*)," *Toxicol. Sci.*, vol. 167, no. 2, pp. 536–545, Feb. 2019,
1851 doi: 10.1093/toxsci/kfy257.
- 1852 [143] S. Shin *et al.*, "NRF2 Modulates Aryl Hydrocarbon Receptor Signaling: Influence on
1853 Adipogenesis," *Mol. Cell. Biol.*, vol. 27, no. 20, pp. 7188–7197, Oct. 2007, doi:
1854 10.1128/MCB.00915-07.
- 1855 [144] S. Kalthoff, U. Ehmer, N. Freiberg, M. P. Manns, and C. P. Strassburg, "Interaction between
1856 Oxidative Stress Sensor Nrf2 and Xenobiotic-activated Aryl Hydrocarbon Receptor in the
1857 Regulation of the Human Phase II Detoxifying UDP-glucuronosyltransferase 1A10," *J. Biol.*
1858 *Chem.*, vol. 285, no. 9, pp. 5993–6002, Feb. 2010, doi: 10.1074/jbc.M109.075770.
- 1859 [145] M. S. Denison and S. C. Faber, "And now for something completely different: Diversity in
1860 ligand-dependent activation of Ah receptor responses," *Curr. Opin. Toxicol.*, vol. 2, pp. 124–
1861 131, Feb. 2017, doi: 10.1016/j.cotox.2017.01.006.
- 1862 [146] A. A. Soshilov and M. S. Denison, "Ligand Promiscuity of Aryl Hydrocarbon Receptor Agonists
1863 and Antagonists Revealed by Site-Directed Mutagenesis," *Mol. Cell. Biol.*, vol. 34, no. 9, pp.
1864 1707–1719, May 2014, doi: 10.1128/MCB.01183-13.
- 1865 [147] N. Josyula, M. E. Andersen, N. E. Kaminski, E. Dere, T. R. Zacharewski, and S. Bhattacharya,
1866 "Gene co-regulation and co-expression in the aryl hydrocarbon receptor-mediated
1867 transcriptional regulatory network in the mouse liver," *Arch. Toxicol.*, vol. 94, no. 1, pp. 113–
1868 126, Jan. 2020, doi: 10.1007/s00204-019-02620-5.

- 1869 [148] E. Vazquez-Rivera *et al.*, "The aryl hydrocarbon receptor as a model PAS sensor," *Toxicol.*
1870 *Reports*, vol. 9, no. June 2021, pp. 1–11, 2022, doi: 10.1016/j.toxrep.2021.11.017.
- 1871 [149] W. Zhang *et al.*, "The aryl hydrocarbon receptor: A predominant mediator for the toxicity of
1872 emerging dioxin-like compounds," *J. Hazard. Mater.*, vol. 426, no. November 2021, p. 128084,
1873 Mar. 2022, doi: 10.1016/j.jhazmat.2021.128084.
- 1874 [150] F. D. L. Leusch and S. A. Snyder, "Bioanalytical tools: half a century of application for potable
1875 reuse," *Environ. Sci. Water Res. Technol.*, vol. 1, no. 5, pp. 606–621, 2015, doi:
1876 10.1039/C5EW00115C.
- 1877 [151] G. L. Warren *et al.*, "A Critical Assessment of Docking Programs and Scoring Functions," *J.*
1878 *Med. Chem.*, vol. 49, no. 20, pp. 5912–5931, Oct. 2006, doi: 10.1021/jm050362n.
- 1879 [152] S. Coecke *et al.*, "Metabolism: A Bottleneck in In Vitro Toxicological Test Development,"
1880 *Altern. to Lab. Anim.*, vol. 34, no. 1, pp. 49–84, Feb. 2006, doi: 10.1177/026119290603400113.
- 1881 [153] S. Lungu-Mitea *et al.*, "On the Utilization and Characterization of External Biotransformation
1882 Systems in In Vitro Toxicology: A Critical Review of the Scientific Literature with Guidance
1883 Recommendations," *ACS Environ. Au*, vol. 6, no. 1, pp. 21–45, Jan. 2026, doi:
1884 10.1021/acsenvironau.5c00096.
- 1885 [154] M. Eide, M. Rusten, R. Male, K. H. M. Jensen, and A. Goksøyr, "A characterization of the ZFL
1886 cell line and primary hepatocytes as in vitro liver cell models for the zebrafish (*Danio rerio*),"
1887 *Aquat. Toxicol.*, vol. 147, pp. 7–17, 2014, doi: 10.1016/j.aquatox.2013.11.023.
- 1888 [155] V. Le Fol *et al.*, "Cell-Specific Biotransformation of Benzophenone-2 and Bisphenol-S in
1889 Zebrafish and Human in Vitro Models Used for Toxicity and Estrogenicity Screening," *Environ.*
1890 *Sci. Technol.*, vol. 49, no. 6, pp. 3860–3868, Mar. 2015, doi: 10.1021/es505302c.
- 1891 [156] G. Williamson *et al.*, "Induction of glutathione S-transferase activity in hepG2 cells by extracts
1892 from fruits and vegetables," *Food Chem.*, vol. 60, no. 2, pp. 157–160, 1997, doi:
1893 10.1016/S0308-8146(95)00255-3.
- 1894 [157] A. Gronert, O. Zierau, D. Thieme, and A. M. Keiler, "Effect of HepG2 cell 3D cultivation on the
1895 metabolism of the anabolic androgenic steroid metandienone," *Drug Test. Anal.*, vol. 15, no.
1896 11–12, pp. 1319–1328, Nov. 2023, doi: 10.1002/dta.3455.
- 1897 [158] E. L. Schaefer and J. K. Selkirk, "Metabolism of benzo(a)pyrene by variant mouse hepatoma
1898 cells," *Cancer Res.*, vol. 45, no. 8, pp. 3487–92, Aug. 1985, [Online]. Available:
1899 <http://www.ncbi.nlm.nih.gov/pubmed/4016732>.
- 1900 [159] E. B. Froyen and F. M. Steinberg, "Soy isoflavones increase quinone reductase in hepa-1c1c7
1901 cells via estrogen receptor beta and nuclear factor erythroid 2-related factor 2 binding to the
1902 antioxidant response element," *J. Nutr. Biochem.*, vol. 22, no. 9, pp. 843–848, Sep. 2011, doi:
1903 10.1016/j.jnutbio.2010.07.008.
- 1904 [160] S. Wilkening, F. Stahl, and A. Bader, "COMPARISON OF PRIMARY HUMAN HEPATOCYTES AND
1905 HEPATOMA CELL LINE HEPG2 WITH REGARD TO THEIR BIOTRANSFORMATION PROPERTIES,"
1906 *Drug Metab. Dispos.*, vol. 31, no. 8, pp. 1035–1042, Aug. 2003, doi: 10.1124/dmd.31.8.1035.
- 1907 [161] M. J. Gómez-Lechón, L. Tolosa, I. Conde, and M. T. Donato, "Competency of different cell
1908 models to predict human hepatotoxic drugs," *Expert Opin. Drug Metab. Toxicol.*, vol. 10, no.
1909 11, pp. 1553–1568, Nov. 2014, doi: 10.1517/17425255.2014.967680.
- 1910 [162] J. A. Doering, R. Farmahin, S. Wiseman, S. W. Kennedy, J. P. Giesy, and M. Hecker,
1911 "Functionality of Aryl Hydrocarbon Receptors (AhR1 and AhR2) of White Sturgeon (*Acipenser*

- 1912 transmontanus) and Implications for the Risk Assessment of Dioxin-like Compounds,”
1913 *Environ. Sci. Technol.*, vol. 48, no. 14, pp. 8219–8226, Jul. 2014, doi: 10.1021/es502054h.
- 1914 [163] R. L. Tanguay, C. C. Abnet, W. Heideman, and R. E. Peterson, “Cloning and characterization of
1915 the zebrafish (*Danio rerio*) aryl hydrocarbon receptor,” *Biochim. Biophys. Acta - Gene Struct.*
1916 *Expr.*, vol. 1444, no. 1, pp. 35–48, Jan. 1999, doi: 10.1016/S0167-4781(98)00252-8.
- 1917 [164] R. L. Tanguay, E. Andreasen, W. Heideman, and R. E. Peterson, “Identification and expression
1918 of alternatively spliced aryl hydrocarbon nuclear translocator 2 (ARNT2) cDNAs from zebrafish
1919 with distinct functions,” *Biochim. Biophys. Acta - Gene Struct. Expr.*, vol. 1494, no. 1–2, pp.
1920 117–128, Nov. 2000, doi: 10.1016/S0167-4781(00)00225-6.
- 1921 [165] D. W. Cowing, J. C. A. Bardwell, E. A. Craig, C. Woolford, R. W. Hendrix, and C. A. Gross,
1922 “Consensus sequence for *Escherichia coli* heat shock gene promoters.,” *Proc. Natl. Acad. Sci.*,
1923 vol. 82, no. 9, pp. 2679–2683, May 1985, doi: 10.1073/pnas.82.9.2679.
- 1924 [166] M. Dreano *et al.*, “High-level, heat-regulated synthesis of proteins in eukaryotic cells,” *Gene*,
1925 vol. 49, no. 1, pp. 1–8, Jan. 1986, doi: 10.1016/0378-1119(86)90380-X.
- 1926 [167] A. A. Stepanenko and H. H. Heng, “Transient and stable vector transfection: Pitfalls, off-target
1927 effects, artifacts,” *Mutat. Res. Mutat. Res.*, vol. 773, pp. 91–103, Jul. 2017, doi:
1928 10.1016/j.mrrev.2017.05.002.
- 1929 [168] M. J. Carvan 3rd, W. A. Solis, L. Gedamu, and D. W. Nebert, “Activation of transcription
1930 factors in zebrafish cell cultures by environmental pollutants,” *Arch Biochem Biophys*, vol. 376,
1931 no. 2, pp. 320–327, 2000, doi: 10.1006/abbi.2000.1727.
- 1932