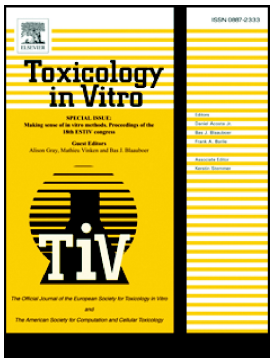
Integrate mechanistic evidence from new approach methodologies (NAMs) into a read-across assessment to characterise trends in shared mode of action

Escher, S.E.; Aguayo-Orozco, A.; Benfenati, E.; Bitsch, A.; Braunbeck, T.; Brotzmann, K.; ... ; Fisher, C.

A read-across case study on chronic toxicity of branched carboxylic acids (1): Integration of mechanistic evidence from new approach methodologies (NAMs) to explore a common mode of action



Sylvia E. Escher, Alejandro Aguayo-Orozco, Emilio Benfenati, Annette Bitsch, Thomas Braunbeck, Katharina Brotzmann, Frederic Bois, Bart van der Burg, Jose Castel, Thomas Exner, Domenico Gadaleta, Iain Gardner, Daria Goldmann, Oliver Hatley, Nazanin Golbamaki, Rabea Graepel, Paul Jennings, Alice Limonciel, Anthony Long, Richard Maclennan, Enrico Mombelli, Ulf Norinder, Sankalp Jain, Liliana Santos Capinha, Olivier T. Taboureau, Laia Tolosa, Nanette G. Vrijenhoek, Barbara M.A. van Vugt-Lussenburg, Paul Walker, Bob van de Water, Matthias Wehr, Andrew White, Barbara Zdrazil, Ciarán Fisher

PII: S0887-2333(21)00194-6

DOI: https://doi.org/10.1016/j.tiv.2021.105269

Reference: TIV 105269

To appear in: *Toxicology in Vitro*

Received date: 13 August 2021

Revised date: 17 October 2021

Accepted date: 27 October 2021

Please cite this article as: S.E. Escher, A. Aguayo-Orozco, E. Benfenati, et al., A read-across case study on chronic toxicity of branched carboxylic acids (1): Integration of mechanistic evidence from new approach methodologies (NAMs) to explore a common mode of action, *Toxicology in Vitro* (2021), https://doi.org/10.1016/j.tiv.2021.105269

process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A read-across case study on chronic toxicity of branched carboxylic acids (1): integration of mechanistic evidence from new approach methodologies (NAMs) to explore a common mode of action.

Sylvia E. Escher[1], Alejandro Aguayo-Orozco[2], Emilio Benfenati[3], Annette Bitsch[1], Thomas Braunbeck[4], Katharina Brotzmann[4], Frederic Bois[10], Bart van der Burg[5], Jose Castel[6], Thomas Exner[9], Domenico Gadaleta[3], Iain Gardner[10], Daria Goldmann [18], Oliver Hatley[10], Nazanin Golbamaki[11], Rabea Graepel[12], Paul Jennings[13], Alice Limonciel[13], Anthony Long[11], Richard Maclennan[7], Enrico Mombelli[6], Ulf Norinder[8], Sankalp Jain[16], Liliana Santos Capinha[13], Olivier T. Taboureau[14], Laia Tolosa[6], Nanette G. Vrijenhoek[12], Barbara M. A. van Vugt-Lussenburg[5], Paul Walker[7], Bob van de Water[12], Matthias Wehr[1], Andrew White[15], Barbara Zdrazil[16], Ciarán Fisher[10]

1 Fraunhofer Institute for Toxicology and Experimental Medicine, Chemical Safety and Toxicology, Germany

2 Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark

3 Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milano, Italy

4 Aquatic Ecology and Toxicology Group, Center for Organismal Studies, University of Heidelberg, Heidelberg, Germany

5 BioDetection Systems, Amsterdam, The Netherlands

6 Instituto de Investigación Sanitaria La Fe, Valencia, Spain

7 Cyprotex, Cheshire, United Kingdom

8 MTM Research Centre, Örebro University, Sweden

9 Edelweiss Connect GmbH, Basel, Switzerland

10 Certara UK Ltd, Simcyp Division, Sheffield, United Kingdom

11 Lhasa Limited, Leeds, United Kingdom

12 Leiden Academic Centre for Drug Research (LACDR), Leiden University, Leiden, The Netherlands

13 Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

14 Université de Paris, France

15 Unilever Safety and Environmental Assurance Centre, Sharnbrook, Bedfordshire, United Kingdom

16 University of Vienna, Department of Pharmaceutical Sciences, Division of Pharmaceutical Chemistry, Vienna, Austria

Abbreviations

ADME- absorption, distribution, metabolism and excretion; AOP - adverse outcome pathway; DEGs - differentially expressed genes; DST - Dempster-Shafer Theory; ECHA - European Chemical Agency; EFSA - European Food Safety Authority; FBS - fetal bovine serum; GLP - good laboratory practice; GSH - glutathione; hOED - human oral equivalent dose; KE – key event; MEC - minimum effective concentration; MIE - molecular initiating event; MMP - mitochondrial membrane potential; MoA - mode-of-action; PI - propidium iodide; NAM – new approach methodology; PBK - physiologically based pharmacokinetic modelling; PHH – primary human hepatocyte; QIVIVE - quantitative *in vitro* to *in vivo* extrapolation; RAx- read across; ZFET - Zebrafish embryo test.

Abstract

This read-across case study characterises thirteen, structurally similar carboxylic acids demonstrating the application of in vitro and in silico human-based new approach methods, to determine biological similarity. Based on data from in vivo animal studies, the read-across hypothesis is that all analogues are steatotic and so should be considered hazardous.

Transcriptomic analysis to determine differentially expressed genes (DEGs) in hepatocytes served as first tier testing to confirm a common mode-of-action and identify differences in the potency of the analogues. An adverse outcome pathway (AOP) network for hepatic steatosis, informed the design of an *in vitro* testing battery, targeting AOP relevant MIEs and KEs, and Dempster-Shafer decision theory was used to systematically quantify uncertainty and to define the minimal testing scope.

The case study shows that the read-across hypothesis is the critical core to designing a robust, NAM-based testing strategy. By summarising the current mechanistic understanding, an AOP enables the selection of NAMs covering MIEs, early KEs, and late KEs. Experimental coverage of the AOP in this way is vital since MIEs and early KEs alone are not confirmatory of progression to the AO. This strategy exemplifies the workflow previously published by the EUTOXRISK project driving a paradigm shift towards NAM-based NGRA.

Introduction

Read-across (RAx) is one of the most frequently applied alternative approaches in chemical hazard assessment, in particular for complex *in vivo* endpoints like (sub)chronic or reproductive toxicity (Ball et al., 2016; ECHA, 2017). Within the RAx assessment, the *in vivo* endpoint data of source compounds (SCs) are used to predict the *in vivo* endpoint of a similar target compound (TC), based on structure or biological activity. One of the most challenging aspects in this process is the identification of source compounds, which will elicit the same toxicological response in humans or follow a predictable trend.

The starting point of the similarity assessment is usually a list of SCs, which share structural and physicochemical properties (ECHA, 2017). To justify this initial selection, sufficient evidence must be provided that these shared chemical properties will result either in similar toxicodynamic properties, or a consistent, predictable trend within the grouped compounds. An understanding of toxicokinetics can be used as supporting evidence to explain variations across a category of grouped compounds. Recently, an analysis of several case studies revealed that the uncertainty in RAx assessments is highly dependent on the justification of similarity (Schultz and Cronin, 2017).

The analysis of the toxicodynamic properties would ideally infer a shared mode-of-action (MoA) across a category of grouped compounds.

This conclusion, however, is typically not possible for endpoints such as chronic toxicity, as the observed apical findings/effects, as seen in in vivo repeated exposure studies, are generally not informative of the underlying biological mechanisms. Furthermore, multiple mechanisms or adverse outcome pathways (AOPs) can lead to the same observed macroscopic alterations (tissue degeneration, atrophy etc.) or histopathological changes (inflammatory responses, hypertrophy,

hyperplasia, necrosis etc.). The uncertainty of the analysis of shared toxicodynamic properties is likely to increase with a decreasing number of source compounds. Such a similarity assessment is impaired in a scenario with one source compound (i.e. analogue approach). In practice, risk assessment often faces this data-sparse situation, since the number of analogues is restricted to those with appropriate high-quality *in vivo* endpoint studies.

Human in vitro and in silico approaches have the potential to elucidate common toxicodynamic properties and thus may contribute to overcoming this hurdle in read-across assessments.

The application of human *in vitro* and *in silico* approaches goes hand in hand with new challenges such as predictive performance, and robustness and reproducibility of the *in vitro* models (Krebs et al., 2019). In addition to the experimental challenges, it is particularly important to demonstrate the relevance of the new models, along with a strategy for integrating them into human risk assessment.

To date, the incorporation of new approach methodologies (NAMs) into regulatory decision making has been slow. This is primarily because risk assessors do not have the same confidence in their use as they do in traditional in vivo guideline studies (Patterson et al., 2021). One reason for this lack of confidence is that most NAMs have not yet reached a status of formal validation e.g. as being achieved within the OECD (OECD, 2005) and ECVAM (Hartung et al., 2004) validation process. However, regulatory agencies have expressed a vision to replace and/or reduce animal testing with Next Generation Risk Assessment (NGRA) as much as possible in the near future (EFSA, 2019; US EPA, 2018). Thus, new concepts are needed here to show the applicability and validity of non-validated NAMs and to address regulators reservations.

This paper describes a RAx case study using NAM-based data to characterize a shared MoA within the grouped compounds. The approach is developed using 13 branched carboxylic acids to illustrate the testing and evaluation approach for the typical human endpoint of chronic toxicity. Traditionally, the RAx would be based on few source compounds with high-quality *in vivo* endpoint data, namely subchronic rodent toxicity studies with oral exposure. In the following, we demonstrate the use of NAMs, in particular human *in vitro* and *in silico* models, to characterize the toxicodynamic properties of each analogue in the category, with the aim of demonstrating trends/(dis)similarities across the grouped compounds more clearly. Based on the mechanistic insight gained, the most appropriate source compounds per target compound are selected, thus reducing the uncertainty of the RAx extrapolation.

The hazard characterization applies i) a high-level biological similarity assessment of the initially selected structural analogues using transcriptome data, ii) an AOP network to inform targeted testing of mechanistically relevant molecular initiating events (MIEs) and key events (KEs) in selected human *in vitro* test systems, ii) additional non-AOP related read-outs which cover a broader mechanistic space or like the zebrafish embryo assay test a whole organism response, and iv) Dempster-Shafer decision theory to integrate different types of *in vitro* data.

Material and Methods

Chemicals

Chemicals were purchased at the highest purity available. 2-ethylheptanoic acid (2-EHP, purity 99.9%, VBP00053) was purchased from Endeavour Speciality Chemicals Limited; 2-propylhexanoic acid (2-PHA, purity 96%, ZCA0360) was purchased from Finetech Industry Limited; 4-ene valproic acid

(4-ene VPA, purity 98%, sc-209255) and 2-ethylpentanoic acid (2-EPA, purity 98%, sc-496785) were purchased from Santa Cruz Biotechnology; 2-methylhexanoic acid (2-MHA, purity >99%, W319104); 2-methylpentanoic acid (2-MPA, purity ≥98%, W275409); 2-ethylbutyric acid (2-EBA, purity 99%, 109959); 2-methylbutyric acid (2-MBA, purity 98%, 245526); 2-propylheptanoic acid (2-PHP, purity NA, VBP00053); 2-ethylhexanoic acid (2-EHA, purity ≥99%, 538701); valproic acid (VPA, certified reference material for analytical application, PHR1061); pivalic acid (PVA, purity 99%, T71803) and 2,2-dimethyl valeric acid (DMVA, purity >97%, 41740) were purchased from Sigma Aldrich. The chemicals were purchased centrally and distributed to the partners as pure compounds. Each partner prepared stocks according to the specific requirements of their test methods.

Development of an AOP network for liver steatosis

The hepatic steatosis AOP network primarily builds on the following existing AOPs (AOPWiki, https://aopwiki.org and sAOP, http://saop.cpr.ku.dk/ (Aguayo-Orozco et al., 2019), status April 2020):

- AOP 34- LXR activation leading to hepatic steatosis
- AOP 36- Peroxisomal Fatty Acid Beta-Oxidation Inhibition Leading to Steatosis
- AOP 57 - AhR activation leading to hepatic steatosis
- AOP 58 - NR1I3 (CAR) suppression leading to hepatic steatosis
- AOP 59 - HNF4alpha suppression leading to hepatic steatosis
- AOP 60 - NR1I2 (Pregnane X Receptor, PXR) activation leading to hepatic steatosis
- AOP 61 - NFE2L2/FXR activation leading to hepatic steatosis
- AOP 62 - AKT2 activation leading to hepatic steatosis
- AOP 318 - Glucocorticoid receptor activation leading to hepatic steatosis

The involvement of THRSP (thyroid hormone responsive) was added on the basis of an analysis of the toxicogenomic database TG-GATES (Igarashi et al., 2015). This gene has been shown to be expressed in liver and adipocytes, particularly in lipomatous modules. The AOP network comprises some MIEs and KEs which play a role in the MoA of the analogue VPA. MIEs and KE related to VPA are color coded as follows: yellow – MIE/KE is known for VPA; red: MIE/KE is known for VPA and tested in this case study (Figure 2).

The gene expression of CPT1A, HMGCS2, CD36, HSD17B10, NRF2, SHP, LXR, PPARγ, SREBF1, and SRXN1 and ECI1 were significantly altered following VPA exposures in primary human hepatocytes (PHHs) as part of TG-GATES (Aguayo-Orozco et al., 2018; Grinberg et al., 2014). ECI1 knockdown mice suffer from steatosis and the Comparative Toxicogenomics Database (Davis et al., 2019) also indicates that ECI1 is affected by VPA. ECI1 encodes a member of the hydratase/isomerase superfamily, a key mitochondrial enzyme involved in beta-oxidation of unsaturated fatty acids (according to NCBI, National Center for Biotechnology Information). The sequestration of co-enzyme A (e.g. by VPA) can impair mitochondrial β-oxidation (Aires et al., 2010; Schumacher and Guo, 2015). The action of VPA on PPAR-γ, CD36, SCD1, ChREBP, SREBF1, and FAS was documented by (van Breda et al., 2018), who also confirmed its activity on PPAR-α, LXR, PXR and AhR testing inPHHs. One mechanism involves the binding of VPA to coenzyme A (CoA), which cause a reduction of mitochondrial β-oxidation of fatty acids and in consequence lipid accumulation in the vacuoles of the cytoplasm (Aires et al., 2010; Schumacher and Guo, 2015).

The resultant AOP network is in most parts congruent with the ones recently proposed by (Mellor et

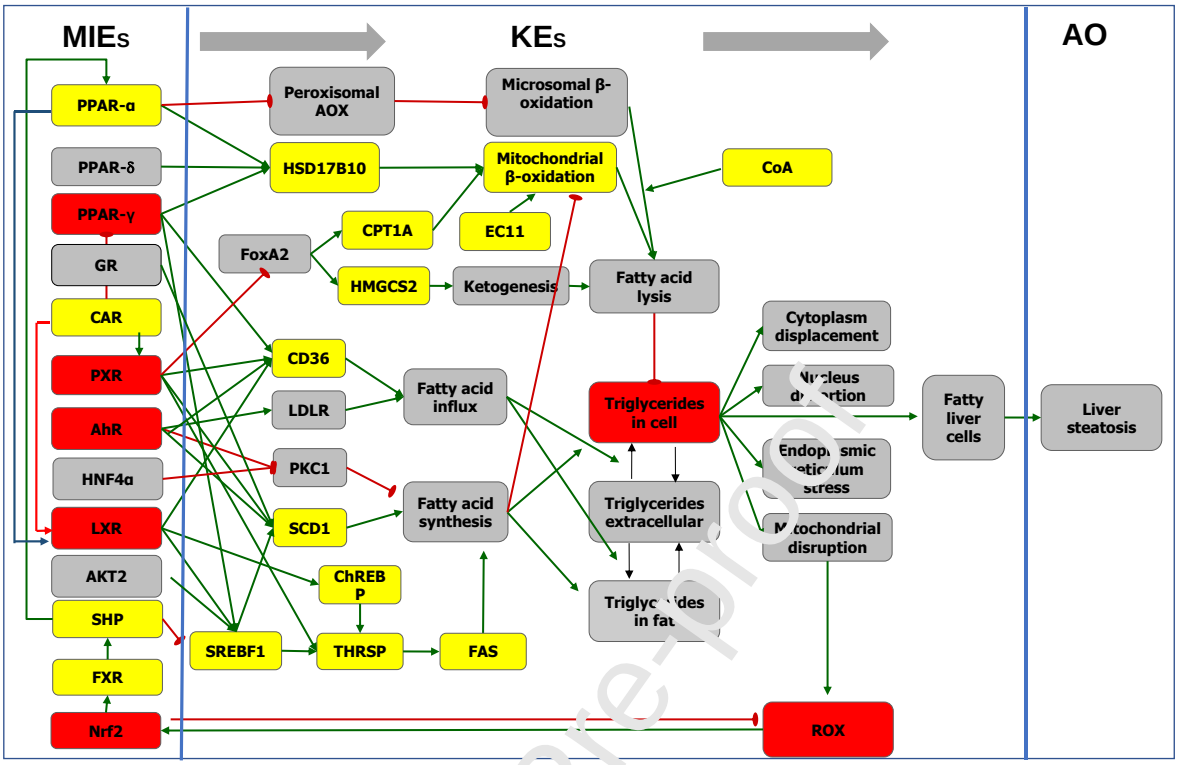


Figure 1: The AOP network for microvesicular liver steatosis illustrates the activation and inhibition of molecular initiation events (MIEs) progressing to key events (KEs) and resulting in the adverse outcome (AO) liver steatosis; red lines indicate inhibition, green lines activation; coloured MIEs/KEs are obtained from studies testing VPA (red/yellow). Red coloured MIEs/KE were selected for AOP specific testing in this case study. In addition endoplasmatic reticulum stress was measured. Names and abbreviations are given

al., 2016; van Breda et al., 2018) (Figure 1, abbreviations described in Supplemental data).

Tiered testing strategy

Our tiered testing strategy classified NAMs into seven separate effect associated data classes. The first class comprised an analysis of transcriptomics data generated in HepG2 cells and identified shared differentially expressed genes (DEGs) and altered signalling pathways. This evidenced a biological rationale for a shared MoA within the thirteen structurally similar, grouped category compounds (Table 1; class 1).

The remaining 6 effect classes comprise more specific functional read outs, representative of MIEs and KEs belonging to the AOP network, and additional MIEs, KEs, and cellular processes covering a broader mechanistic space (Table 1, class 2 to 7).

As an initial assessment of general toxicity, the cytotoxicity of all compounds was measured in two high-throughput reporter assays (CALUX and GFP), two human hepatocyte cell lines (HepG2, HepaRG), PHHs, and the human RPTEC/TERT1 (renal proximal tubule epithelial cell) cell line (class 2).

Nine MIEs and early KEs from the AOP network were selected to characterize the mechanisms across the category leading to liver steatosis (class 3, Figure 1). The CALUX reporter panel tested the activation of PPARγ, PXR, AhR, GR and LXR. Oxidative stress was tested using the Nrf2-CALUX and SRXN1-GFP reporter assays. Endoplasmatic reticulum (ER) stress was measured using BIP-GFP and ESRE-CALUX reporter assays.

The late KE "accumulation of triglycerides" was measured in PHH and two liver cell lines (HepG2, HepaRG). Lipid accumulation was assessed semi-quantitatively after single (treatment duration 24 h or 72 h) or repeated exposure (five times in 10 days, class 4) using fluorescence-based methods. This key event is close to the endpoint steatosis and was used for in vitro to in vivo extrapolation (IVIVE) to derive a human relevant threshold.

Within classes 5 and 6, reporter genes and NAMs assessing general cellular processes were included to cover a broader mechanistic space. This broader testing aimed to reveal potential differences between the grouped category compounds. Class 5 comprised several other 'orphan' MIEs and early KEs, from the CALUX reporter panel and GFP reporter assays, not included within the AOP network. These included the nuclear hormone receptor assays ERα, Anti-ERα, AR, anti-AR, PR, Anti-PR, Anti-GR, TRβ, Anti-TRβ (all: endocrine activity), RAR, and the stress- and other pathway assays Hif1α (hypoxia), TCF (wnt-pathway), AP-1 (cell cycle control), NFκB (inflammatory response) and p21 and p53 (DNA damage response). Though PPARα and PPARδ are MIEs within the steatosis AOP network, all reporter assays were performed in agonist mode. The role of PPARα and PPARδ as MIEs within the steatosis AOP network is characterized through interaction with antagonists and so data from these NAMs were included in class 3 (van Breda et al., 2018).

Class 6 gave an overview on the perturbation mitochondrial processes in HepG2 cells, in particular glutathione (GSH) depletion, disruption of the mitochondrial membrane potential (MMP), formation of mitoxsuperoxide (MitoSOX) or phospholipids. The role of mitochondrial dysfunction, including the interference with β-oxidation and oxidative stress, represented by altered GSH homeostasis and/or enhanced reactive oxygen species (ROS) formation, have been suggested as mechanisms involved in liver injury by different drugs including inducers of steatosis (Chang and Abbott, 2006; Silva et al., 2008).

The observations from the human-based *in vitro* assays, were aligned to effects seen in the liver of zebrafish embryos (ZFET, class 7). More specifically, the histopathological alterations of the liver include the detection and amount of storage materials (fats and glycogen) as well as the assessment of overall changes in hepatocellular structure and size. Comparisons of zebrafish and mammals showed remarkable similarities in hepatic lipid metabolism between both systems (Anderson et al 2011, Brand et al 2002) as well as conserved processes including fundamental development, cellular composition and liver functionality (Goessling and Sadler 2015, Hill 2012, Hölttä-Vuori et al. 2010, Tao and Peng 2009, van Wijk et al. 2016, Wallace et al 2005, Wilkins and Pack 2013). Since these alterations are not directly indicative of a steatotic response, they are considered within this category assessment as supporting information. Since these alterations are not directly indicative of a steatotic response, they are considered within this category assessment as supporting information.

Table 1: Overview on selected MIEs and KEs tested in different human in vitro models in this case study

ID	Class	Measured readouts (KEs/MIEs)	In vitro model
1	Transcriptome analysis	Differentially Expressed Gene	HEPG2 – 24h exposure, single application
2	Viability/Cytotoxicity	IC50	HepG2, HepaRG, PHH, RPTEC/TERT1, CALUX, GFP
3	AOP-specific MIEs and early KEs	PPARγ, PXR, AhR , GR, LXR, Nrf2, ESRE	CALUX reporter assay
		SRXN1, BIP	HepG2-GFP reporter assay
4	AOP-specific late KE	Triglyceride accumulation	HepG2 – 24h or 72h exposure, both with single application
			HepaRG - 24h exposure single application, 10 days exposure, repeated application
			PHH- 24h exposure, single application
5	MIEs and early KEs not related to AOP	ERα, Anti-ERα, AR, anti-AR, GR, Anti-PR, Anti-GR, TRβ, Anti-TRβ, PPAR-α, PPAR-δ, RAR, Hif1α, TCF, AP-1, NFκB, p21 and p53	CALUX reporter assay
		p21	HepG2-GFP reporter assay
6	Mitochondrial dysfunction	GSH, mitochondrial membrane potential, superoxide, phospholipids	HepG2- 24h exposure, single application
7	Non-human evidence synonymous with late KEs	Histopathological alterations of the liver	Zebrafish embryos

NAMs – New approach methodologies

All test compounds are weak organic acids (pKa ≈ 4.8). pH indicators included in cell culture medium showed that high concentrations of test compound decreased the pH of the medium. In addition to the possible cytotoxic effects of this pH shift, *in silico* distribution predictions using the virtual in in vitro distribution (VIVD) model implemented in the Simcyp™ *in vitro* analysis (SIVA) toolkit (v3.0) (Certara UK Ltd, Sheffield, UK, Fisher et al. 2019) showed that a shift in the ionized fraction of solubilized test compound would impact their distribution cells and so the effective *in vitro* concentration (data not shown). Therefore, before application to human cells, all treatment medium were stored for 2h at 37°C in a 5% CO_2 atmosphere, buffering the medium to a physiological pH (~7.4).

As far as possible, the range of test concentrations was standardised across the NAM test battery, informed by *in vivo* relevant concentrations of the source compound VPA through reverse-translation of NOELs determined in rat dosing studies and therapeutic dosing in humans. Physiologically-based pharmacokinetic (PBPK) modelling and simulation of rat NOEL studies predicted a corresponding maximum unbound plasma concentration of 2.5 mM (Fisher, 2019). In humans, VPA is used for the

treatment of epilepsy and at therapeutic doses reaches total plasma concentrations of ~1 mM (Turnbull et al., 1983).

MIE screening for compounds using the CALUX reporter assay was performed in agonist mode across a concentration range of 1 nM to 1 mM testing incrementally by 0.5 log units. Eight test concentrations, ranging from 0.062 to 8 mM were applied in the HepG2, HepaRG and RPTEC assays. The GFP reporter and HepG2 assays additionally tested at 16 mM for inclusion in transcriptomic analyses; PHH were tested up to 31.6 mM. A maximum DMSO solvent concentration of 0.1% was allowed for all test systems.

Specific minimum effective concentration (MEC) values were derived per assay given the dependence on detection limits and signal to noise ratio inherent to each assay exact definitions are detailed below, but broadly the MEC was defined as the compound concentration at which an effect was observed that significantly exceeded the background signal.

CALUX reporter gene assays

From the CALUX® (BioDetection Systems) battery of *in vitro* reporter gene assays a panel of 27 human cell-based assays was used, each able to measure chemical interactions between a test chemical and a specific nuclear receptor or cell signalling pathway (van der Burg et al., 2013). Exposure to the test compounds, dissolved at 0.1 % in DMSO, was performed for 24 h and at 1% (v/v) according to the assay procedure as described in DB-ALM protocol 197 "Automated CALUX reporter gene assay procedure". The analysis consisted of technical triplicates, and was performed twice as independent biological replicates. MEC values were derived per assay based on the background responses. For nuclear receptor agonist assays, the MEC was defined as the PC_{10} concentration in LogM, which is the concentration where the test compound causes an activation effect equal to 10% of the maximum effect elicited by the test's reference compound. For nuclear receptor antagonist assays, the MEC was defined as the PC_{20} concentration, which is the concentration where the test compound causes an antagonist effect equal to 20% of the maximum antagonist effect elicited by the test's reference compound. For the stress pathway related assays which typically do not show sigmoidal dose-response curves, the MEC was defined as the FI 1.5 concentration, which is the concentration where the test compound elicits pathway activation 1.5-fold above baseline.

GFP reporter assay in HepG2 cells

HepG2 cells (clone HB8065, acquired from ATCC) carrying reporters were previously generated according to (Poser et al., 2008) and characterized (Wink et al., 2017). The BAC-GFP reporter cell lines were maintained and exposed in high-glucose Dulbecco's Modified Eagle Medium (DMEM, Thermo Fisher; No. 11504496) supplemented with 10% (v/v) FBS, 25 U/ml penicillin and 25 µg/ml streptomycin. For the reporter assay, cells grow in 384-well plates until 80% confluence at start of imaging. Hoechst33342 (100 ng/ml) was added overnight prior to imaging to visualise cell nuclei. Cells were exposed to analogues and cell permeability stain propidium iodide (PI) (an indicator of cell viability) for 24 h and imaged live at 37°C and 5% CO_2, capturing 2 images per well at 20x magnification using a Nikon TiE2000 confocal laser microscope.

GFP intensities (relative fluorescent units, rfu) in reporter cell lines were quantified as previously described (Wink et al., 2018). Nuclei were segmented and cytoplasm boundaries were determined

after which GFP intensity was measured in either the nuclei (P21) or cytoplasm (SRXN1, BIP). MEC values were calculated from the fraction of GFP-positive cells. A non-mechanistic fitting through the data points was applied (R Core Team, 2020). The MEC was determined at the concentration where a GFP-value of DMSO treated control cells (+ 2x SD) crosses the fit. All data are from 3 biological replicates.

HepG2 cells and High-content imaging assay

HepG2 cells were seeded in 96-well plates (5000 cells/well) and cultured for 24h. Prior to exposure to the test compounds, HepG2 cells were preincubated with a 62 µM mixture of oleate and palmitate (2:1 ratio) in medium supplemented with lipid depleted FBS (Biowest (S181L-100)) for 14 h, followed by a change to fatty-acid free medium containing different concentrations of the test chemicals for a further 24 or 72 h. Stock solutions of compounds were prepared in DMSO, and were diluted in the culture medium to obtain the desired final concentrations. . Cells were simultaneously loaded with a combination of fluorescent probes (Hoechst 33342 for nuclei identification, PI for cell viability, monochlorbimane for GSH detection, TMRM for the detection of changes in MMP and MitoSOX Red for mitochondrial superoxide production) for 30-min incubated at 37°C as previously described (Donato et al., 2012; Tolosa et al., 2015). Labelled cells were then imaged using the INCELL6000 Analyser (GE Healthcare, USA) (Tolosa et al., 2012) and analysed in the INCELL Workstation. Dose response curves for cell viability data, lipid or phospholipid accumulation were generated using GraphPad Prism Software (version 8). The MEC was determined as the concentration able to an elicit 20% change in a respective endpoint relative to non-treated cells. All data are from 3 biological replicates.

HepaRG Spheroids and High-content imaging assays

Cryopreserved differentiated HepaRG cells were obtained from Caltag Medsystems (UK) and were used for formation of 3D spheroid cultures. (Bell et al., 2016) Cells were seeded into ultra-low attachment (ULA) 96-well plates (Corning, USA) at a density of 2,000 viable cells per well and left overnight. HepaRG spheroids were seeded in 100 µl DMEM 31053-044 (ThermoFisher, UK) containing hepatocytes bullet kit (Lonza, UK) supplemented with 2 mM ultra glutamine (ITS), 25 mM HEPES (Sigma), and 10% (v/v) FBS. HepaRG cells spontaneously self-aggregate in to spheroids. Five days after seeding, visible compact spheroids had formed, 50% of the medium was exchanged daily for serum-free medium (Williams E medium E W1878 (Sigma-Aldrich) supplemented with 2 mM L-glutamine (ITS), 100 units/ml penicillin, 100 µg/ml streptomycin, 10 µg/ml insulin, 5.5 µg/ml transferrin, 6.7 ng/ml sodium selenite, 100 nM dexamethasone (Sigma-Aldrich), and 10% (v/v)FBS; Bell et al., 2016). Cells were maintained in a 5% CO_2 humidified atmosphere at 37°C.

Cell imaging with fluorescence analysis was performed with a Cellomics® ArrayScan XTI Infinity High Content Imaging (HCI) platform (ThermoFisher, UK), which utilised HCS StudioTM 2.0 software (ThermoFisher, UK) and the compartmental analysis bioapplication for image analysis. Test compounds were prepared as stock solutions at 200x higher concentration than the maximum test concentration (solvent concentration maintained at 0.5%). Cells were treated in triplicates at eight different concentrations of each test compound for either 24 h or 5 days. The culture medium was removed and cells were stained with cellular dyes HCS LipidTOX™ Green Neutral Lipid Stain (ThermoFisher H34475) and Hoechst 33342. Cells were washed three times with phosphate buffered saline (PBS) and fluorescence image acquisition was performed. A single HepaRG spheroid per well

was imaged in confocal mode using a 70 pin hole size, using 11 steps at 11.6 µm distance. Cell nuclei or spheroid size were detected analysing the Hoechst 33342 (Sigma) fluorescence signal (λex.360 to 400 nm; λem. 410 to 480 nm); each endpoint was analysed for changes in fluorescent intensity signal (rfu) in either the cytoplasmic or nuclear regions of each cell or spheroid region and compared against the vehicle control wells.

Cellular ATP was determined in HepaRG cells and HepaRG spheroids following dosing, using luminescence following the manufacturers guidelines (2D; CellTiterGlo, 3D; CellTiter-Glo 3D Cell Viability Assay, Promega); luminescence was determined using a BioTek Synergy 2 (BioTek). Raw fluorescence intensity values (rfu) were normalised to vehicle control wells in all cases and expressed as fold changes in assay signal. Data was normalized to vehicle control and for each compound dose-response curve. The lowest concentration exceeding the vehicle control limits (0.85 to 1.15 of the vehicle control values) were defined as the MEC. All data are from 3 biological replicates.

Primary human hepatocytes (PHH) and lipid accumulation

Cytotoxicity: PHHs were cultivated in clear William's E medium (PAN-biotech P04-29510) supplemented with 100 U/ml penicillin/streptomycin, 10 µg/ml gentamycin, 100 nM dexamethasone, 2 mM glutamine and 2 ng/ml insulin (ITS). For attachment (3 h) 10% (v/v) FBS (PAN-biotech P30-3701) was added to the medium. The cells were cultivated in a 96-well format, 50,000 cells/well in 200 µl medium/well on collagen (0.25 mg/ml) monolayer coated wells. The next day after seeding, the cells were incubated with 62 µM 2:1 mixture of oleate and palmitate complexed to bovine serum albumin for 24 h. Following preincubation the cells were incubated for 48 hours with respective test e compounds in medium with or without fatty acids. Five concentrations and vehicle matched controls are prepared for each compound applying a dilution factor of 3.16, doses ranged from 0.316 to 31.6 mM. Cell viability was determined with the CellTiterBlue kit (Promega).

Cryopreserved PHHs were treated with test compounds with (day 5) and without (day 1) preincubation of fatty acids and stained to determine the lipid accumulation after treatment. The five tested dose ranged from 0.1 to 10 mM. PHHs were cultivated in clear William's E medium (PAN-biotech P04-29510) supplemented with 100 U/ml penicillin/streptomycin, 10 µg/ml gentamycin, 100 nM dexamethasone, 2 mM glutamine and 2 ng/ml insulin (ITS). For attachment (3 h) 10% (v/v) FBS (PAN-biotech P30-3701) was added to the medium and for the fatty acid treatment a 62 µM mixture of oleate and palmitate (2:1 ratio). The cells were cultivated in a 24-well format, 250,000 cells/well in 500 µl medium/well, collagen (0.25 mg/ml) monolayer coated wells plus collagen (1 mg/ml) gel on top of the cells. One day after seeding (day 0) the cells were treated with the test compounds.

After 24 h cells were fixed with ROTI-Histofix (4% PFA) for 20 minutes at 37°C, washed with PBS and then permeabilized with 0.3% Triton-X-100 in PBS for 10 minutes at room temperature. Afterwards the cells were stained with phalloidin (actin), BODIPY (lipids), DAPI (DNA) and covered with FluorPreserve Reagent. The glass slides were processed with an Axio Scan.Z1 (ZEISS) and the accompanying software. The generated images were evaluated by semi-quantitative manual inspection. Three human donors were tested.

Human kidney cells (RPTEC/TERT1), resazurin reduction and supernatant lactate assay for membrane integrity

The telomerase immortalised human renal proximal tubule epthelial cells, RPTEC/TERT1 (Wieser et al., 2008), were routinely cultured in a 1:1 mixture of DMEM (Gibco 11966-025) and Ham's F12 (Gibco 21765-029), with a final concentration of 5 mM glucose supplemented with 2 mM Glutamax, 10 ng/ml epidermal growth factor, 36 ng/ml hydrocortisone, 5 µg/ml insulin, 5 µg/ml transferrin, 5 ng/ml selenium, 100 U/ml penicillin, 100 µg/ml streptomycin (Jennings et al., 2009) and supplemented with final concentration of 0.5% (v/v) FBS. Cells were cultured in a humidified environment (37 °C, 5% CO_2) and were routinely passaged once a week detatched through trypsin EDTA treatment. Cells were cultured and differentiated in 96-well plates by allowing them to reach confluence and remain in a confluent state for at least 7 days before chemical exposure as previously described (Aschauer et al., 2013). MEC values were calculated as the first tested concentration that shows a significantly higher value than vehicle control (1-way ANOVA).

After 24 h chemical exposure, cells were washed with 100 µl PBS per well and further incubated with 100 µl of 44 µM resazurin solution for 1.5 to 2 h at 37°C. The fluorescent product resorufin generated through metabolic activity of viable cells was detected using a CLARIOstar plate reader (BMG Labtech) (λex. 540 nm; λem. 590 nm).

Supernatant lactate was quantified against a standard curve in a 96-well plate with 10 µl supernatant medium incubated with 90 µl lactate reagent buffer (86 mM Triethanolamine HCl, 8.6 mM EDTA.Na2, 33 mM $MgCl_2$, 326 µM N-methylphenazonium methyl sulphate (PMS), 790 µM p-iodonitrotetrazolium violet (INT), 3.37 mM β-NAD, 7% (v/v) ethanol, 0.4% (v/v) Triton-X-100, 4 U/ml Lactate Dehydrogenase) for approximately 7 minutes at room temperature, as previously described (Limonciel et al., 2011). Optical density (λabs. 490 nm) was measured using a CLARIOstar plate reader.

Zebrafish embryo test (ZFET)

Zebrafish (*Danio rerio*) embryos were treated with concentrations ranging from 2 to 1000 µM for 120 hours post-fertilization (hpf). ZFET assays were conducted according to the extended Fish Embryo Acute Toxicity (FET) test described in detail in OECD guideline 236 (OECD, 2013), which is still within the developmental phase defined as non-protected (EU, 2010) according to Strähle et al. (2012). In accordance with OECD 236, pH was adjusted using hydrogen chloride and sodium hydroxide to 7.75 ± 0.02 prior to addition of the test compounds; no further adjustment was made after addition of the compounds. For all substances, 10 embryos per test concentration were analyzed.

At the end of exposure, embryos were anesthetized in crushed ice (Wilson et al., 2009) and fixed in Davidsons's fluid (Johnson et al., 2010; Mulisch and Welsch, 2015) at 4°C for 24 h. Sections were cut at 4 µm thickness with a Reichert-Jung HN 40 microtome (Reichert-Jung, Heidelberg, Germany) and transferred onto microscope slides coated with glycerin albumin (Serva, Heidelberg, Germany). Hematoxylin-eosin straining was performed following the approach from Mulisch and Welsch (2015).

Histopathological evaluation of zebrafish liver slides was carried out with a Nikon ECLIPSE 90i microscope (Nikon Instruments, Düsseldorf, Germany) and the Nikon NIS Elements AR 64-bit software, v. 4.00.05. Embryos showing alterations of the liver (predominantly changes in the size and overall structure of hepatocytes as well as the amount of storage materials (lipids and glycogen)) were recorded for each test concentration. While slides containing liver sections of each embryo were reviewed, slides showing the biggest cross sections of the liver, excluding those divided by the gut, were selected for the evaluation. Based on this data, an EC_{20} value was calculated for each

compound using ToxRat® v. 2.10.03 (ToxRat Solutions, Alsdorf, Germany). EC_{20} defines the concentration causing a liver effect in 20 % of all zebrafish embryos tested with a compound.

Statistical analyses

Analysis and derivation of differentially expressed genes (DEGs)

HepG2 cells (clone HB8065, acquired from ATCC) were maintained and exposed in high-glucose DMEM supplemented with 10% (v/v) FBS, 25 U/ml penicillin and 25 µg/ml streptomycin. Cells (passage 14 to 15) were plated in 96 well plates (Cellstar, Greiner), with 50,000 cells per well and exposed the next day with seven concentrations per compound, ranging from 0.2 to 16 mM for 24 h. Cells were washed once with PBS and lysed for 15 min at room temeprature with 50 µl 2x BioSpyder lysis buffer diluted to 1x with PBS. Samples were stored at -80 °C and subsequently shipped on dry ice to Biospyder technologies (Bioclavis, UK). All data is from 3 biological replicates.

Transcriptome analyses were carried out with the human TempO-Seq S1500+ assay, which comprises 3,565 genes (Mav et al., 2018). DEGswere derived following four general steps: 1) Quality control: samples with low total read counts were removed (threshold: 500,000 total reads per sample). 2) Increase statistical power: genes with low variance across all concentrations and replicates per compounds were removed (threshold = 1). 3) Normalize read counts: Counts per million (CPM) were used. 4) Derive DEGs: DEGs were derived with the R-package DESeq2 (version R-3.6.3 DESeq2 1.26.0, Lover et al., 2014) considering an adjusted p-value of 0.05 (Benjamini-Hochberg method) and a log2fold change of 1 or higher. This analysis identified 2,197 DEGs across this category taking into account all DEGS per concentration for all compounds tested. For the analysis of category specific DEGs, we selected 249 consistently expressed DEGs from the seven most responsive compounds (2-PHP to 2-EPA). Consistency was assumed for genes being differentially expressed in at least 20 out of a total of 49 testing conditions.

From the DEGs a group profile was derived to show commonalities in differential expression between grouped category compounds. The criteria for selection of a DEG for this profile was defined as the frequency in which it was significantly differentially expressed in the conditions (individual concentrations) of the seven longer chain compounds (formerly regarded as active). For heatmap visualisation the profile was joined with significantly DEGs of rotenone in HepG2 treated at 0.08 mM (151 DEGs). Criteria for these DEGs was defined through p-value adjusted by Benjamini-Hochberg method ($P < 0.05$, absolute log2fold ≥2). Hierarchical clustering was conducted within concentration groups, based on the log2fold change of the resulting 337 genes.

Visualization of data

ToxPi (v2.3) (Marvel et al., 2018) was used to illustrate the biological similarity of compounds (Figure 6 and 7). Linear scaling of the MEC (in –logM units) was performed to standardise the values between different assays. MECs were set to zero for compounds being inactive up to the highest *in vitro* tested dose. ToxPi scores were calculated for all assays in the seven effect classes (Table1). The hierarchical clustering of all experimental data applied Euclidean distances and complete linkage. MEC values shown in -log(M) units were clustered with a Morpheus widget created in R (Morpheus, https://software.broadinstitute.org/morpheus) (Figure 8).

Dempster-Shafer Theory (DST)

DST is an extension of generalized Bayesian statistical inference in which evidence can be associated with multiple sources of information (Dempster, 2008; Shäfer, 1976). DST was used to combine the evidence from different *in vitro* assays for the source compounds in order to provide a weight-of-evidence (WoE) estimate for the target compound with respect to steatotic *in vivo* outcome.

The DST analysis used assay results in the form of binary data; compounds for which a MEC could be derived were classified as active; compounds being not active up to the highest *in vitro* tested dose were classified as inactive (0).

Sets of assays were identified, which gave good validation results from a leave-one-out (LOO) cross-validation for compounds with known *in vivo* outcomes (steatotic or non-steatotic, see Supplemental data). This LOO validation enables the calculation of reliability, i.e. balanced accuracy in this study, positive prediction accuracy (PPV) and negative prediction accuracy (NPV) for the assay. The predicted outcome from DST for compounds with known *in vivo* data was performed using a LOO cross validation. For the remaining compounds without known outcome the analysis was performed using data from all compounds with known outcome.

The DST software used in this work was developed in the EU-ADR project (ICT-215847). The DST combines the available information and quality thereof into two prediction outcomes – belief (BEL) and plausibility with respect to a 'proposition (p)'. The proposition in this analysis is that all analogues are steatotic. The BEL part indicates the strength of the evidence in support of p, on a scale of 0 (no certainty) to 1 (full certainty) and difference (plausibility - BEL) represents the level of uncertainty based on the evidence from the data.

Results

Selection of category members based on structural similarity

The grouped category compounds share a common core-structure (Figure 2). They differ with regard to the length of the two linear alkyl side chains at carbon atom C2. In total, the category comprises eleven branched aliphatic analogue carboxylic acids with linear side chains ranging from methyl (C1) to pentyl (C5, Figure 2). One analogue in this group, 4-ene VPA, comprises one alkene side chain. In addition to these eleven structurally similar category members, two further aliphatic carboxylic acids, namely pivalic acid (PVA) and dimethyl valeric acid (DMVA), were included into the testing and assessment approach. Both have a third methyl substitutent at C2 (R3, Figure 2). This results in in a total of 13 structurally related members of the category (Figure 3). The physicochemical properties in this category show that lipophilicity and molecular weight increases slightly with longer side chains. Taking lipophilicity as a first indicator for a potential bioaccumulation in human tissues, these data do not alert for a bioaccumulation in humans (i.e. logPow > 4.0).

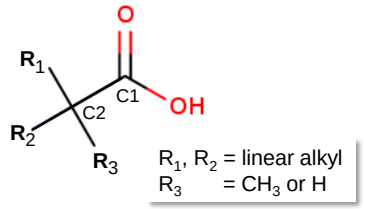


Figure 2: Generalised chemical structure of the category members.

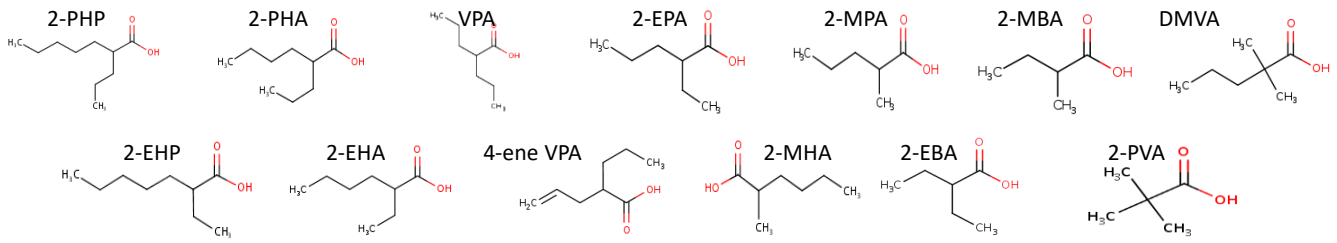


Figure 3: Overview on case study members in order of decreasing side chain and molecular weight: 2-Propylheptanoic acid (2-PHP); 2-Ethylheptanoic acid (2-EHP); 2-Propylhexanoic acid (2-PHA; 2-Ethylhexanoic acid (2-EHA); Valproic acid (VPA); 4-ene Valproic acid (4-ene VPA); 2-Ethyl pentanoic acid (2-EPA); 2-Methyl-hexanoic acid (2-MHA); 2-Methyl-pentanoic acid (2-MPA); 2-Ethylbutyric acid (2-EBA); 2-Methylbutyric acid (2-MBA), pivalic acid (PVA) and dimethyl-valeric acid (DMVA).

Read-across hypothesis

This read-across addresses the endpoint chronic toxicity after repeated exposure, which usually requires a subchronic in vivo study per source compound. The source compound 2-EHA has *in vivo* oral dosing endpoint studies with subchronic duration (BG Chemie, 2000; Juberg et al., 1998). Supporting information is available from shorter-term *in vivo* studies for the four analogues VPA, 4-ene VPA, 2-EBA and PVA (Abdel-Dayem et al., 2014; Espandiari et al., 2008; Ibrahim, 2012; Knapp et al., 2008; Loscher et al., 1992; Sugimoto et al., 1987; Tong et al., 2005; Zhang et al., 2014). 2-EHA, VPA and 4-ene VPA induced liver steatosis as the most sensitive observed adverse effect, whereas 2-EBA and PVA did not show any adverse liver effect up to the highest in vivo tested dose.

For precautionary reasons, a worst-case assumption is used and thus the read-across hypothesis is that all category members will potentially cause hepatic steatosis in vivo. It can be justifiably assumed that it would be difficult to conclude that the subchronic *in vivo* studies of only 2-EHA could be read across all analogues of the category given the sparsity of the *in vivo* data matrix.

Biological similarity based on transcriptome data to support category inclusion

Structural similarity is generally regarded as a good starting point to identify an initial list of source compounds (ECHA, 2017). Follow-up analyses are needed to confirm common toxicological properties. Transcriptome data in the human liver cell line HepG2 were evaluated for a preliminary biological similarity assessment.

A category specific profile of differentially expressed genes (DEGs) was defined, consisting of 249 DEGs, dysregulated within most of the test conditions for the seven most active compounds in the category. The contribution of the different dose-groups per compounds to the group profile is indicated in Figure 4. Although the number of DEGs per compound increases with increasing test

concentration, the majority of the 249 shared DEGs are derived from low and mid test concentrations (0.5 to 4 mM). The highest tested doses (8 mM, 16 mM) did not contribute many additional shared DEGs to the group profile, indicating minimal unspecific high dose effects in this profile.

To learn to what extent transcriptome data can be used for biological similarity assessment, an additional compound, rotenone, was included. Rotenone is a complex-I inhibitor and showed a distinct expression profile relative to the profile of the carboxylic acids (Figure 5).

Hierarchical clustering of the 249 DEGs results in two clearly separated clusters (Figure 5). From 0.5 mM to 2mM, the long-chain analogues show the highest and comparable activity at the transcriptome level (2-PHP; 2-PHA, VPA; 2-EHP and 4-ene VPA). At doses of 4mM, 8mM and above, 2-EHA and 2-EPA also join this cluster. Compounds that result in relatively fewer DEGs fall into a second cluster, consisting of 2-EBA, 2-MPA and 2-MBA, 2-MHA, PVA and DMVA.

The transcriptome signature of the grouped category compounds becomes increasingly homogenous at higher concentrations suggesting that these compounds act through similar mechanisms but with varying potency across the category. Further analyses have been done to better understand the extent and nature of the shared mechanisms and their relation to steatosis and other biological processes (Vrijenhoek et al., 2021).

The initial category of source compounds seems to be justified based on this preliminary biological similarity assessment.

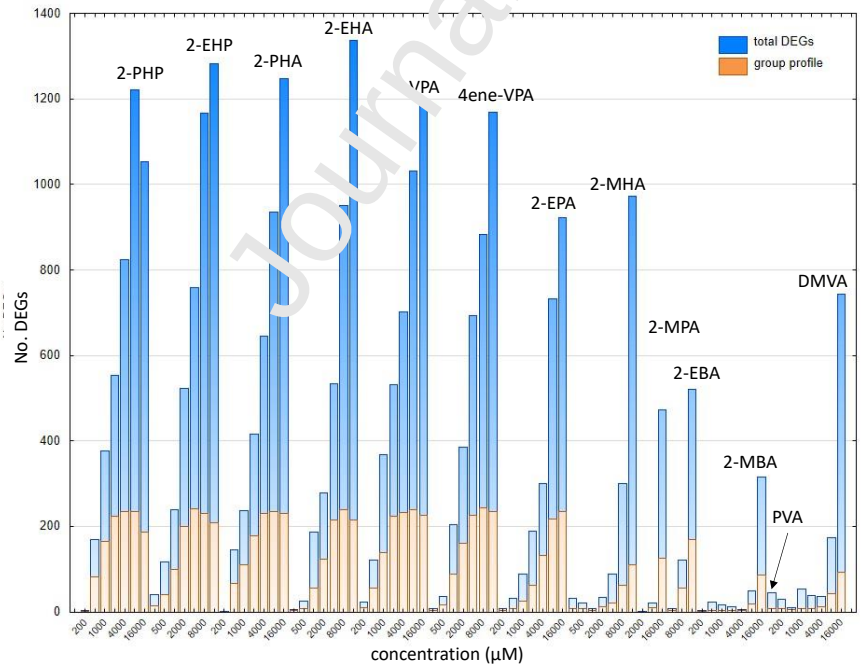


Figure 4: Group specific DEGs – the total number of DEGs per concentration and test compound is shown (blue bars) as well as the amount of DEGs per concentration contributing to the group

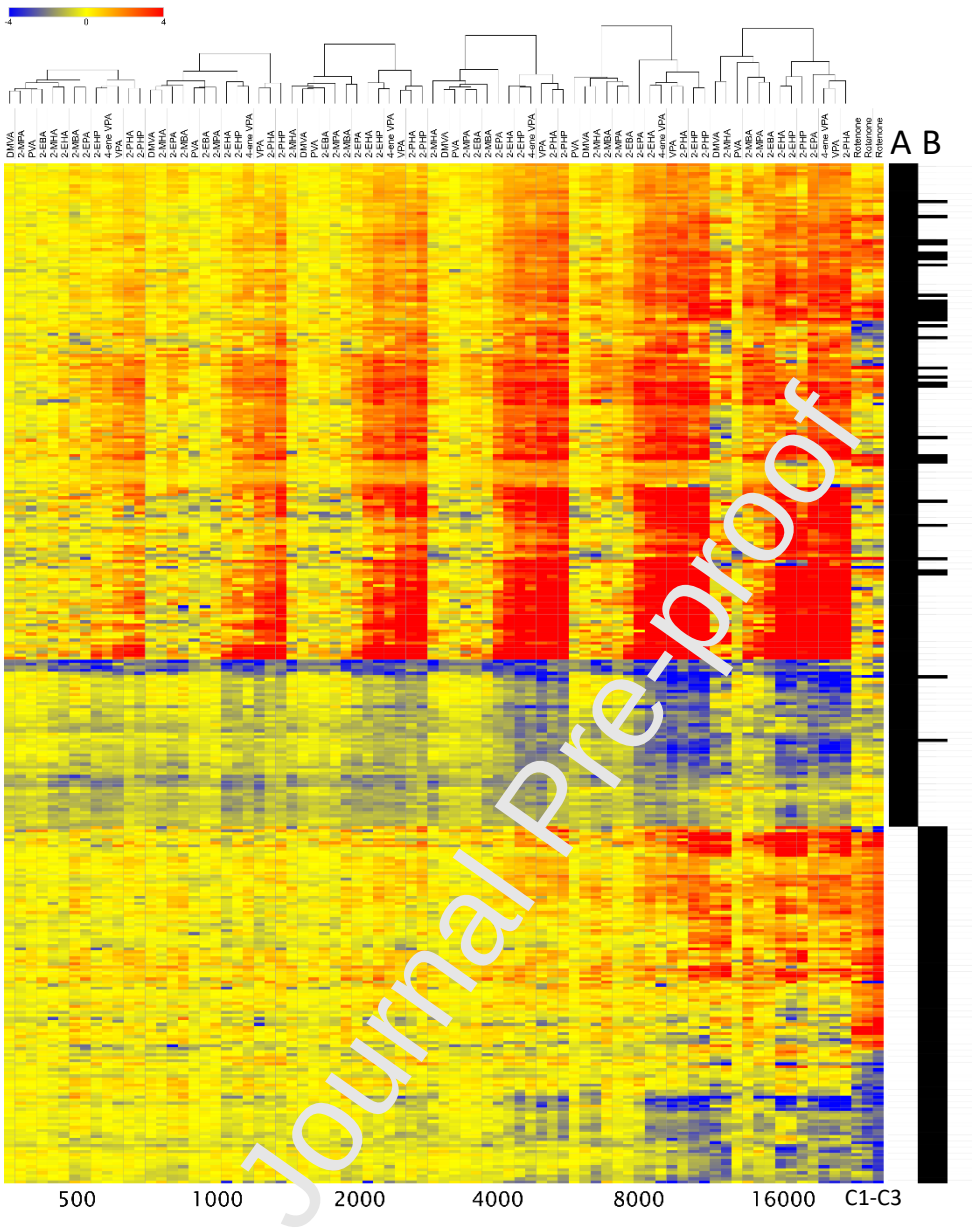


Figure 5: Unsupervised hierarchical clustering of DEGs in HepG2 cells – category members are shown together with the complex-I inhibitor rotenone (C1: 0.0032 µM; C2: 0.016µM; C3: 0.08µM). The category members are clustered within concentration using hierarchical clustering (Euclidean distances, complete linkage). The number of shared up and down regulated DEGs increase with dose and side chain length of the investigated compounds. The origin of a gene is annotated, A: category; B: rotenone. For better visualization, the log2fold change values ≥ +4 were coloured deep red, all values ≤ -4 deep blue per compound and concentration.

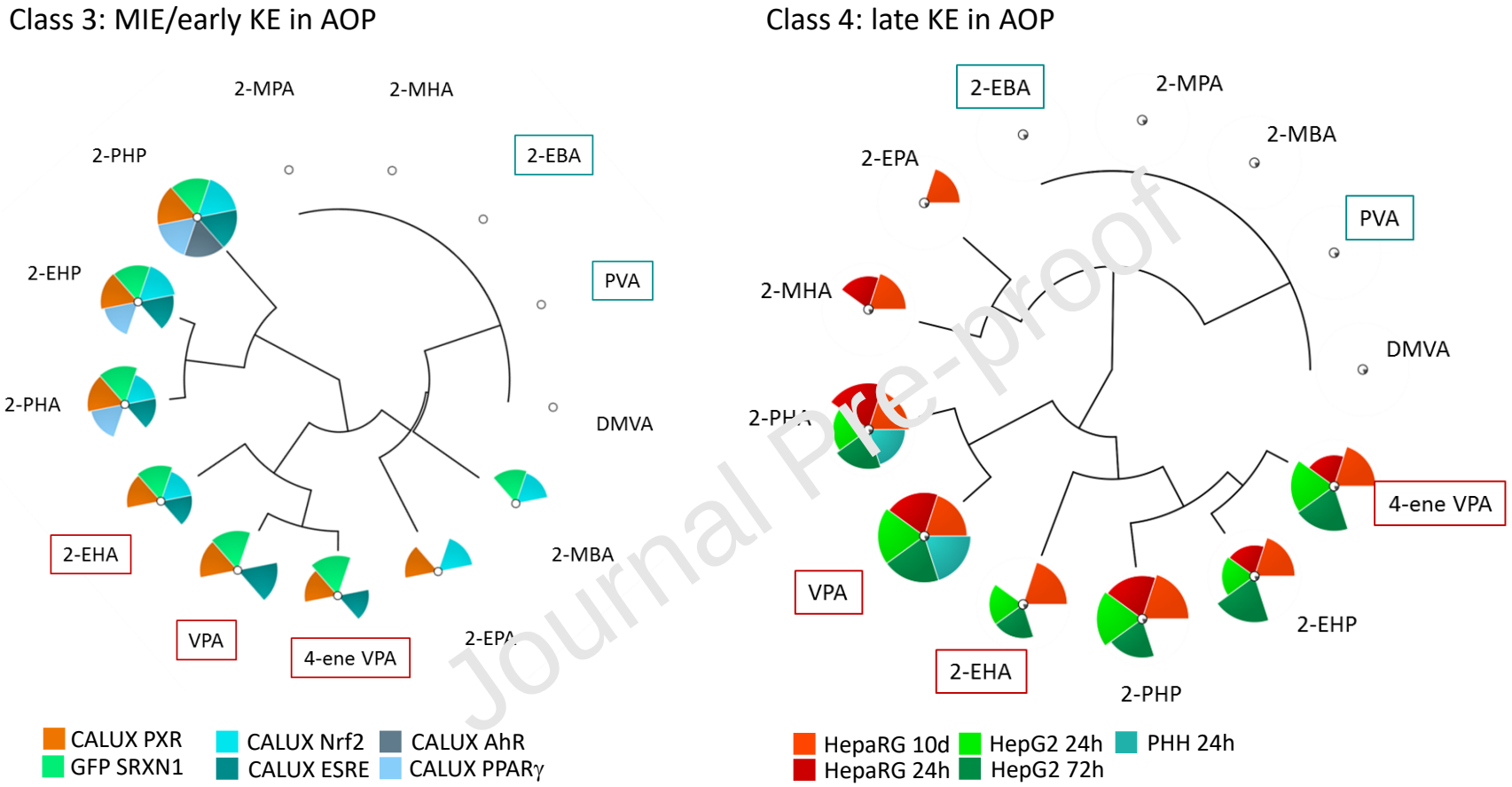


Figure 6: Comparison of the biological activity of all analogues in the AOP relevant classes 2 (MIE and early KE in AOP) and 3 (late KE triglyceride accumulation in HepG2 cells). *In vivo* positive compounds (showing induction of liver steatosis primarily in preclinical rodent studies) are marked in red; *in vivo* negative compounds are marked in green. Minimal effect concentrations in -logM units were normalized and hierarchical clustering was achieved using complete linkage.

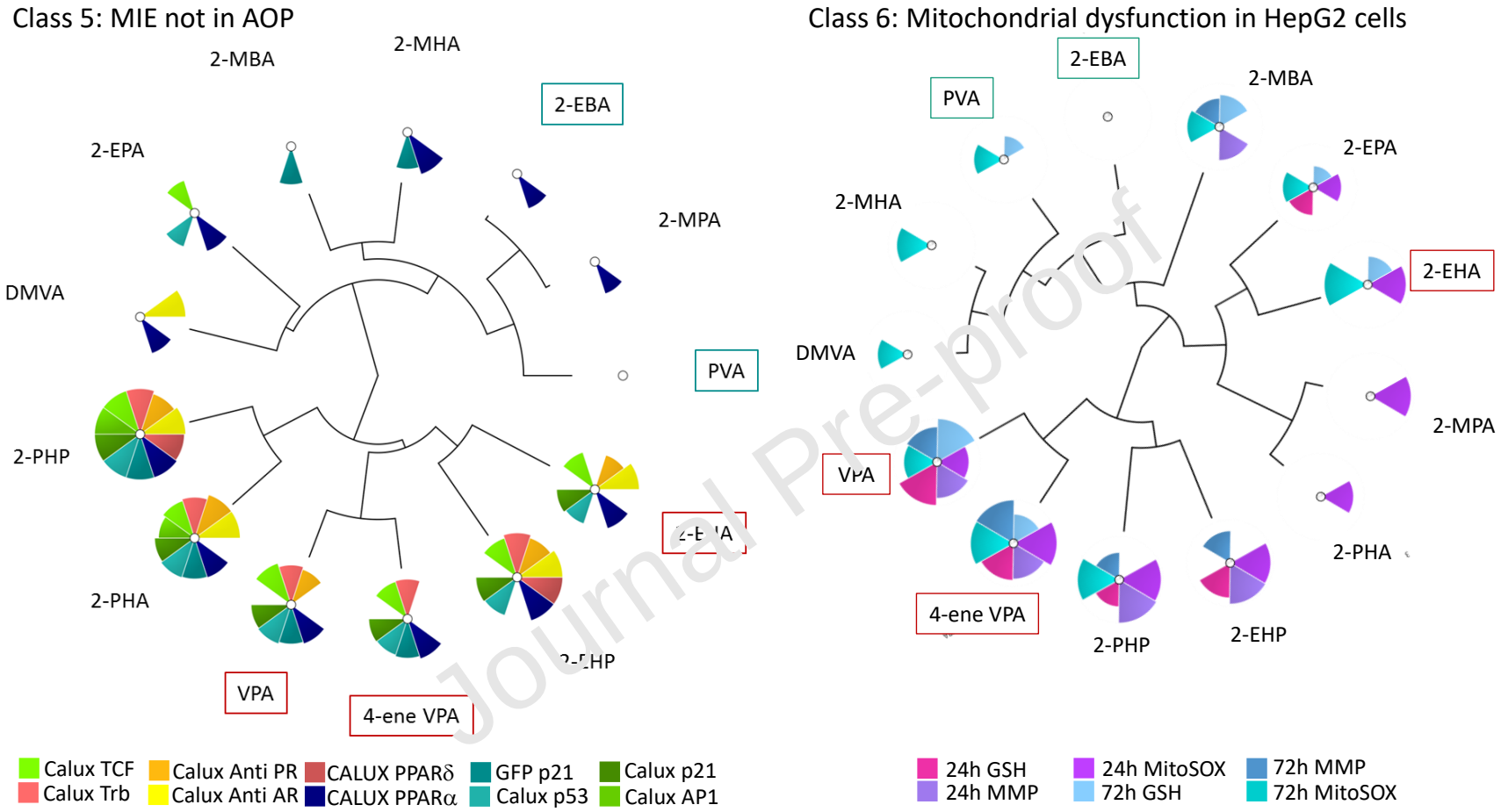


Figure 7: Comparison of the biological activity of all analogues in the not AOP relevant classes 4 (MIE not in AOP) and 5 (Mitochondrial dysfunction in HepG2 cells). *In vivo* positive compounds (showing induction of liver steatosis primarily in preclinical rodent studies) are marked in red; *in vivo* negative compounds are marked in green. Minimal effect concentrations in -logM units were normalized and hierarchical clustering was achieved using complete linkage.

Tiered Testing for Hazard Identification

Overall category compounds showed low cytotoxicity in CALUX, GFP and liver models relative to controls. Viability was also not compromised in RPTEC/TERT1 kidney cells based on resazurin metabolism and supernatant lactate as indicators (see Supplemental data, Limonciel et al. (2011); (2012)). Only 4-ene VPA induced resazurin reduction in RPTEC/TERT1 kidney cells at a MEC of 125 µM (see Supplemental Data).

The hazard charaterization continued with the testing of MIEs and KEs from the AOP network on hepatic steatosis (Figure 1).

Testing of nine MIEs and early KEs from the steatosis AOP network reveals that the number of activated MIEs/KEs decrease with decreasing side chain length (Figure 6, class 3). The shorter chain analogues, 2-MPA, 2-MHA, 2-EBA, as well as DMVA and PVA are inactive up to the highest in vitro tested dose and can be clearly separated from the broader clusters of active compounds. Long-chain analogues activate several MIEs with clusters comprising 2-PHP, 2-EHP, 2-PHA as well as 2-EHA, VPA; 4-ene VPA and a less active cluster comprising 2-EPA and 2-MBA. Appreciable differences in potency are not evident between the active tested analogues in this assay. None of the analogues showed any activity on the GR/LXR CALUX or GFP-BIP reporters (data not shown). However, the analogues 2-PHP to 4-ene VPA, showed activity in the ESRE CALUX assay, which indicates activation of the unfolded protein stress response in contrast to the results of the GFP-BIP reporter assay.

Accumulation of lipids is seen for all long-chain analogues from 2-PHP to 2-MHA in at least one tested liver model (class 4, Figure 6). Within this cluster, analogues with longer side chains induced lipid accumulation in several tested liver models, whereas the shorter-chain analogue 2-EPA showed this activity only in HepaRG cells after repeated exposure (5 treatments in 10 days) at relative high concentrations compared to the other analogues (MEC = 5290 µM). In the case of 2-MHA, lipid accumulation was also observed in HepaRG cells at the highest tested dose after single exposure (MEC = 2380 µM; single exposure). These data suggest 2-MHA to be a relatively more potent inducer of lipid accumulation compared to 2-EPA. 2-MPA, 2-EBA and 2-MBA did not induce accumulation of triglycerides up to the highest dose tested *in vitro*, comparable to DMVA and the in vivo negative compound PVA.

The comparison of MECs in classes 3 and 4 shows that MIEs and early KEs measured in the CALUX assays are activated at lower concentrations compared to the late KE lipid accumulation (supplemental data). Potency differences were much more evident from the testing of the late KE lipid accumulation in hepatocytes.

As might be expected, a less defined trend is seen for the activation of MIEs and early KEs non-specific to the steatosis AOP network (class 5, Figure 7; MIEs/KEs being inactive for all tested compounds are not shown). Overall, the number of activated MIEs decreased with decreasing side chain length. All long chain analogues, including 2-EPA, showed activity in several MIEs, including activation of the p53 CALUX assay. This activity was no longer observed on treatment with S9 incubated compound, an indication for detoxification through hepatic metabolism. DMVA was inactive for the majority of reporters tested in this category but showed activity on the anti-AR CALUX assay in the same range as 2-EHA, 2-PHA, 2-EHP and 2-PHP.

The side chain dependent trend on activity is not evident within category 6 assays (Figure 7). GSH depletion after 24 h was only observed for the analogues 2-EHP, VPA and 4-ene VPA; 4 ene-VPA

showed the greatest potency. After 72 h, 2-EHA and 2-EPA induced GSH depletion, whereas 2-EHP showed no effect on GSH levels (2-PHP was not determined). Disruption of mitochondrial membrane potential (MMP) showed decreasing potency with decreasing side chain length up to 4-ene VPA (2-PHA and 2-EHA inactive).In an exception to the trend, 2-MBA showed some comparable activity to analogues with longer-side chains. Mitochondrial superoxide formation was observed up to 2-EPA and in addition for 2–MPA after 24 h. After 72h, all compounds except of 2-EHP and 2-PHA showed formation of mitochondrial superoxide, including DMVA and PVA. None of the tested compounds induced the formation of phospholipids.

In addition to the AOP network informed targeted testing, a whole organism assay was tested to explore its predictivity. To date, ZFET assays have been used in the context of teratogenicity assessments. In an exploratory study with eight out of the 13 analogues, liver histology from ZFET revealed cellular alterations between untreated and treated embryos (Supplemental data). The livers of untreated zebrafish were characterized by high amounts of storage materials and a homogenous arrangement of relatively large hepatocytes comprising the liver parenchyma. Overall, the liver size of treated zebrafish embryos was reduced, and only regions of the livers still contained storage materials while other areas seemed empty and nuclei appeared less regular (Brotzmann et al. 2021, in preparation). Based on these observations, liver development seemed at least partly inhibited. In zebrafish embryos, the short-chain analogues 2-MPA, 2-MHA, 2-EBA and DMVA proved less potent than the most active long-chain analogues 2-EHA, VPA and 2-PHA. In quantitative terms, 2-EHA, VPA and 2-PHA were more potent by a factor of 10 than 4-ene VPA, identifying similar trends to those found in the human *in vitro* assays.

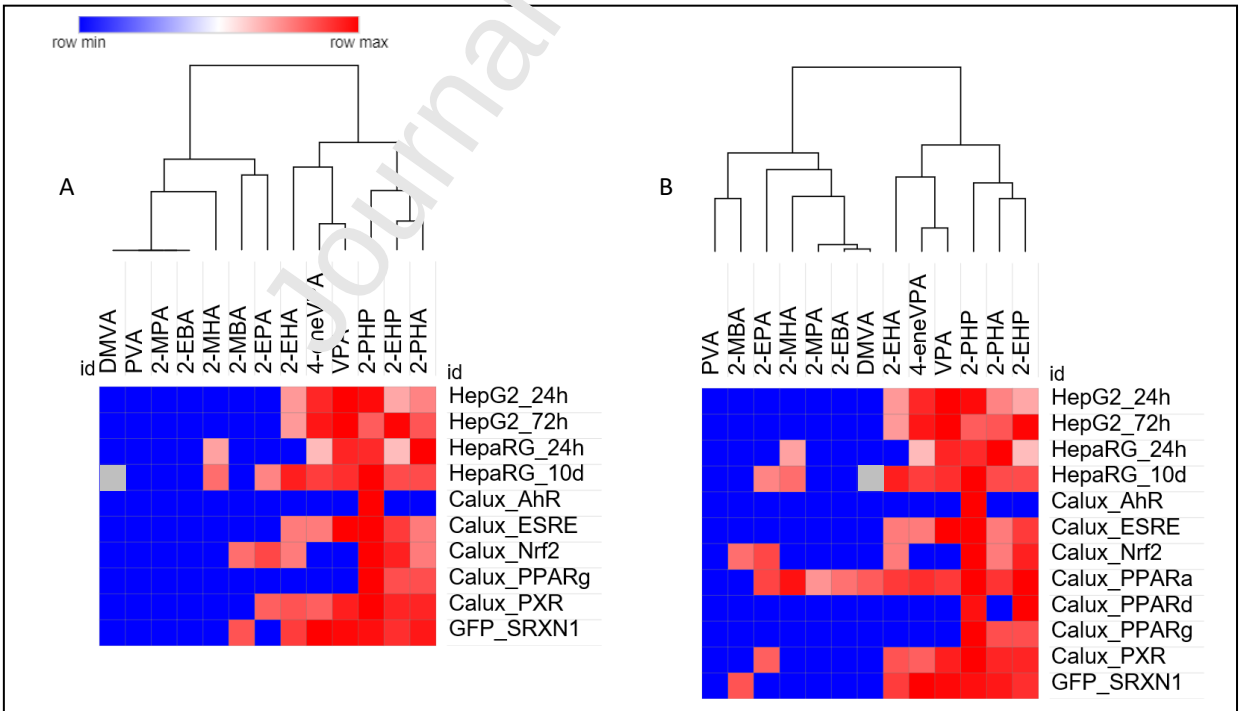


Figure 8: Unsupervised hierarchical clustering of MEC values (in –logM units) for AOP specific assays (Euclidean distances, complete linkage). A: MIEs and KES as described in the AOP (Figure 1); B: AOP enriched with PPAR-α, PPAR-δ. Assays without a response to any of the tested compounds are not included. Red - maximal response in assay, blue - minimal response in assay; grey – no data. PHH data are not included into the clustering as only 3 out of 13 analogues were tested.

Hierarchical clustering of the MEC values of early MIEs/KEs and the late KE lipid accumulation (class 3 and 4, Table 1) discriminates two main groups of most similar compounds, long chain analogues are distinguished from shorter-chain analogues (Figure 8 A). Long-chain analogues show generally more activity on MIEs and KEs, with 2-EHA, 4-ene VPA, VPA forming a subcluster, as well as 2-PHP, 2-EHP, and 2-PHA. VPA and 2-PHP were most potent with regard to the induction of lipid accumulation in the different human liver models, both deriving the lowest MEC after treatment of HepG2 cells after 1d exposure (supplemental data).

In the second cluster of less active compounds, 2-MHA was able to induce lipid accumulation, but did not activate any of the early MIEs and KEs. 2-MBA and 2-EPA cluster together, as well as all inactive compounds, namely DMVA, PVA, 2-MPA and 2-EBA.

AOPs represent the current knowledge and putative mechanisms. Here the AOP network was used to inform the in vitro testing battery. MIEs, like PPARs have been recognized as playing an important role in the regulation of lipid metabolism (Cariello et al., 2021; Wahli and Michalik, 2012). Three isoforms are known, of which PPAR-α is predominantly found in the liver.

PPAR-α is a transcriptional regulator of genes involved in peroxisomal and mitochondrial β-oxidation and fatty acid transport. PPAR-α also coordinates different pathways of de novo lipid synthesis, to supply fatty acid for storage as hepatic triglycerides (Pawlak et al., 2015). In a similar way, PPAR-δ controls fatty acid oxidation by regulating genes involved in fatty acid transport, beta-oxidation, and mitochondrial respiration (Tanaka et al., 2003). Our steatosis AOP network currently represents PPAR-α and δ as an MIE resulting from interaction with antagonists. However, the regulation of fatty-acid metabolism is a complex process and the role for PPARs as MIEs preceding steatosis is not fully elucidated (Maldonado et al., 2018). As such, we incorporated data for PPAR agonism from the CALUX assays into the clustering (Figure 8 B). PPAR-α is activated by all analogues except for PVA and 2-MBA and potency decrease with side chain length across the category. These data suggest a role for PPAR-α agonism in the regulation of steatosis. The integration of PPAR-δ did not add much to the category, as it is only activated by the two longest chain analogues at high concentrations tested. Overall, the inclusion of PPAR-α and δ into the clustering did not alter the overall clustering identifying two groups defined by long and short chain analogues.

In summary, the *in vitro* results obtained from MIEs and KE being present in the AOP (classes 3 and 4) correspond well to the activity seen in the available *in vivo* reference data, in which VPA, 4-ene VPA and 2-EHA induced steatosis related effects (indicated by red boxes in Figure 6), whereas 2-EBA and PVA were inactive (indicated by green boxes in Figure 6). The analysis identified a trend of decreasing biological activity with decreasing side chain length, supported by 'whole-organism' ZFET data. Overall, long-chain analogues are more potent and promiscuous in their activation of MIEs and KEs in comparison to short-chain analogues.

Quantification of uncertainty

A rigorous decision-theory approach, Dempster-Shafer theory (DST) (Dempster, 2008; Rathman et al., 2018; Shäfer, 1976), was applied for the combination of assay data (Table 2). The DST approach presented in this section used binary data instead of minimal effect concentrations. The uncertainty

of the resulting prediction is quantified by belief and plausibility functions with respect to the proposition that all category compounds are steatotic.

The predicted outcomes from DST for compounds with *in vivo* data (termed true class) have a belief of 1 and a plausibility of 1 for positive compounds and the reverse (0 and 0 for negative compounds). Thus, there is no uncertainty associated to these predictions based on the applied *in vitro* data e.g. using the data from class 2 and 3 (ID1), all data (ID2), or only the CALUX dataset (ID4). The analogues without *in vivo* data were also always classified with a belief and plausibility of 1 by all applied *in vitro* datasets. Only 2-EPA shows ambiguous results. 2-EPA is predicted with high belief and plausibility to be steatotic, using all assays (ID2) or non-steatotic, using the AOP specific data (ID1). A restriction to assays, that predicted best the five analogues with *in vivo* data (ID3, Table 1), as indicated by the balanced accuracy greater 90%, did not improve the picture, but resulted in both a belief and plausibility of 50%. This result shows that the DST analysis does provide a correlation, not a causation.

Table 2: Results of Dempster-Shafer theory using binary data: Compounds are listed in order of decreasing molecular weight and side chain length. Compounds with in vivo data (true class) were predicted with 100% belief and plausibility using the different assays in this case study. The predictions for the other analogues give the same results. 2-EPA shows ambiguous results.

Compound	True class*	RAx hypothesis	Belief RAx hypothesis	Plausibility RAx hypothesis	NAM Predicted class	Dataset ID**
2-EHA	steatotic	steatotic	1	1	steatotic	1 to 6
VPA	steatotic	steatotic	1	1	steatotic	1 to 6
4-eneVPA	steatotic	steatotic	1	1	steatotic	1 to 6
2-EBA	non-steatotic	steatoti	0	0	non-steatotic	1 to 6
PVA	non-steatotic	steatotic	0	0	non-steatotic	1 to 6
2-PHP	no data	steatotic	1	1	steatotic	1 to 6
2-EHP	no data	steatotic	1	1	steatotic	1 to 6
2-PHA	no data	steatotic	1	1	steatotic	1 to 6
2-EPA	no data	steatotic	unsure			1 to 6
2-MHA	no data	steatotic	0	0	non-steatotic	1 to 6
2-MPA	no data	steatotic	0	0	non-steatotic	1 to 6
2-MBA	no data	steatotic	0	0	non-steatotic	1 to 6
DMVA	no data	steatotic	0	0	non-steatotic	1 to 6
2-EPA	no data	steatotic	0	0	non-steatotic	1
2-EPA	no data	steatotic	0.84	0.84	steatotic	2
2-EPA	no data	steatotic	0.5	0.5	unsure	3
2-EPA	no data	steatotic	1	1	steatotic	4
2-EPA	no data	steatotic	1	1	steatotic	5
2-EPA	no data	steatotic	0.81	0.81	steatotic	6

*classification based on available in vivo studies with repeated exposure; **different in vitro assay datasets included into the DST analysis; ID1- AOP specific MIs and KEs; ID2 - all assays; ID3 – all assays with balanced accuracy (BA) >9; ID4 - all CALUX assays; ID5 - all CALUX assays with BA >9; ID 6 – CALUX assays belonging to the AOP (see Supplemental data).

Discussion

This case study on branched aliphatic carboxylic acids illustrates the use of NAM data within a RAx approach for the characterization of a shared mechanism of action. For this purpose, a tiered testing strategy was developed, which integrates the evidence from transcriptome data, AOP specific human in vitro and in silico NAM data, and non-AOP specific NAM data. The NAM data were generated to test a read-across hypothesis based on the most critical in vivo effects of source compounds with in vivo endpoint data. An AOP network guided the selection and testing of MIEs as well as early and late KEs, an approach following the RAx schema recently developed through the EUToxRisk project (Escher et al., 2019).

Firstly, confirmation that the initial list of structurally similar source compounds shared biological similarity was achieved by analysing differentially expressed genes in a treated human liver model. Within the category, the analogues cluster into two subgroups comprising the long-chain and short chain analogues, respectively. The expression pattern of the putatively steatotic grouped carboxylic acids differed clearly from those of rotenone, which is a mitochondrial complex-I inhibitor.

Thereafter, MIEs and, early and late KEs were used for hazard identification and biological similarity assessment of the grouped category compounds. The analysis of the resultant biological effect patterns illustrates the usefulness of anchoring the RAx hypothesis to an AOP network. Targeted testing of selected MIEs and KEs belonging to the AOP network supported a shared MoA and identified a trend of decreasing in potency with decreasing side chain length with respect to the lead endpoint of concern, lipid accumulation in hepatocytes. The known in vivo positive compounds (2-EHA, VPA and 4-ene VPA) can be generally well distinguished from the known negative compounds (PVA and 2-EBA). As demonstrated, the combination of selected early MIEs and KE together with one late KE was sufficient to identify this trend. Given these results testing of all MIEs and KEs of the AOP network can be considered unnecessary, provided the *in vitro* assay panel gives sufficient coverage of early MIEs and KEs, and late stage KEs close to adverse outcome.

In the applied human reporter gene assays, MIEs and early KEs responded at lower concentrations compared to those needed to induce accumulation of lipids in human hepatocytes. The higher susceptibility of the reporter assays tested can partially be explained by the artificial nature of the assay system which has been optimized for sensitive responses. In this case study, it has also been shown, that the activation of MIEs/early KEs does not inevitably progress to a later KE as observed for e.g., 2-MBA. Thus, MIEs or early KEs are not solely predictive of progression to an adverse outcome. In this case study, the late KE “triglyceride accumulation in hepatocytes”, is very closely related to the adverse outcome steatosis, defined as lipid accumulation in the liver accounting for ≥ 5% liver weight. It seems, thus, appropriate to use the late KE, in closer proximity to the apical adverse outcome, for the derivation of a human threshold.

The analysis of other MIEs and cellular processes, not directly associated with the AOP network, did not reveal major differences between the grouped compounds. This is also true for the testing of a human kidney model, which was included into the testing battery, because of unspecific kidney weight changes being observed in *in vivo* studies of 2-EHA at predominantly higher doses. All analogues, except of 4-ene VPA, were not cytotoxic up to the highest *in vitro* tested dose, and this finding served as indication that kidney effects are not of primary concern for this category. It can, however, not be excluded that kidney is a critical target organ of 4-ene VPA.

An open question in the generation of NAM data is the testing scope. NAM data can be generated for many compounds at relatively low cost, but biological effect patterns can be complex and difficult to interpret, especially in case of conflicting data.

A decision theory like DST provides an indication about the correlation of the *in vitro* and *in vivo* data and characterises the remaining uncertainty of the obtained predictions. In this case study, DST estimated for the majority of short-chain analogues to be non-steatotic based on the applied *in vitro* datasets, up to the highest *in vitro* tested dose. The classification was already possible based a reduced set of endpoints e.g. using only the results of the CALUX test battery.

However, DST is inconclusive for 2-EPA, which induces lipid accumulation in HepaRG cells following repeated exposures at relatively high concentrations. 2-MHA, is predicted to be non-steatotic by DST but induce lipid accumulation in HepaRG cells following single and repeated exposures at relatively high concentrations. Following the precautionary principles of toxicology, a risk for lipid accumulation at the corresponding human *in vivo* dose cannot be excluded or is even likely. In the present analysis, DST used binary data, so that trends in potency could not be captured. Further, a higher weighting of the late KE "lipid accumulation" was not factored into this DST analysis. Nonetheless, the decision theory gives an objective result which might be helpful in more complex situations and might lead to new hypotheses and iterative testing to support hazard assessment. DST is therefore a useful tool supporting hazard assessment.

A comprehensive uncertainty assessment is not presented. This would have to take into account the uncertainty of each assessment element (AE) in the read-across process and its overall impact on the obtained prediction. This is beyond the scope of this article. Semi-quantitative assessment schemas have been proposed (Blackburn and Stuard, 2014) as well as a template for RAx assessment using NAMs (OECD - Integrated Approaches to Testing and Assessment (IATA) Case Studies Project).

NAM data can be used in two different ways in the read-across assessment: qualitatively to support evidence on biological similarity (substantiating the hazard characterization), or in a quantitative way deriving human threshold values. Here, the biological similarity analysis indicated that long-chain analogues are more homogenous in their activity than short-chain analogues, and this information can be used within a read-across assessment to better define relevant source compounds for a given target compound.

Most similar with regard to AOP-related MIEs were 2-PHP, 2-EHP, and 2-PHA as well as 2-EHA, VPA (Figure 8), with VPA and 2-PHP being the two most potent analogues in this group. In a read-across assessment, the in vivo data of 2-EHA and VPA could be used to derive reference doses for the remaining analogues without preclinical in vivo data. The shorter chain analogues 2-MHA, 2-MBA and 2-EPA were less active compared to 2-EHA and VPA. However, all three also induced lipid accumulation and/or activated early KEs in human cell lines and this would justify a worst-case approach extrapolating the in vivo data from more active analogues 2-EHA and VPA. 2-EBA, 2-MPA, PVA and DMVA failed to induce lipid accumulation and were generally clustered together based on the biological profiles obtained. Since, from a chemical and biological point of view, 2-EBA is most analogues to 2-MPA the *in vivo* data for 2-EBA could rationally be used to predict the toxicity of 2-MPA.

Alternatively, to derive a human threshold, quantitative in vitro to in vivo extrapolation (qIVIVE) could be used. In this case, the in vitro benchmark concentration of the late KE lipid accumulation

would be used to derive a human equivalent dose (hOED) by reverse dosimetry. This approach is described and compared to the thresholds obtained from in vivo data in Fisher et al. 2021.

To date, there are few published RAx studies which use biological similarity based on human *in vitro* models to strengthen the hazard evaluation in a RAx assessment (Gadaleta et al., 2020; Low et al., 2013; Petrone et al., 2012; Shah et al., 2016; Zhu et al., 2016). A commonality of these previously published studies is that they were supported by pre-existing NAM data (Pestana et al., 2021). Following the same concept, a generalized read-across framework (GenRa) has been proposed (Helman et al., 2019; Patlewicz et al., 2018). The use of an AOP to design a targeted NAM testing battery to generate de novo data, driven by a read-across hypothesis and a regulatory problem formulation distinguishes this study from these data mining approaches.

This case study did not address the impact of metabolites. 4-ene VPA is for example one out of several metabolites of VPA observed in patients (Kreher et al. 2002). In this case study, 4-ene VPA induced liver steatosis comparable to its parent VPA. The biotransformation kinetics of VPA in the tested three human *in vitro* assays are however, not known and also not their relation to the human *in vivo* situation. Further work needs to be done to integrate the metabolism into the assessment. So far, it can only be speculated that the formation of an active metabolite like 4-ene VPA explains to some extent the slightly higher potency of VPA compared to the other analogues in this category.

Despite the increasing use of RAx for regulatory purposes and the development of guidance and assessment documents like the read-across assessment framework (ECHA, 2017), it is still a challenge to provide convincing evidence about the robustness of the RAx hypothesis and to quantify the remaining uncertainty of the estimated toxicity for the target compound. New approach methodologies like human *in vitro* and *in silico* models have a great potential to substantiate the RAx assessment through increased weight of evidence. Case studies, as presented here, provide further confidence in the application of these new approaches by illustrating current benefits and limitations. As is typical for read-across evaluations, *in vivo* guideline studies are available for some analogues, allowing a direct comparison of the NAM data with the already well-accepted *in vivo* results, which can be considered as *in situ* validation. This study is an example for other RAx analyses and a step towards the definition of a common strategy integrating structural, physicochemical and biological reasoning in next-generation risk assessment.

Acknowledgement

This work was supported by the European Union's Horizon 2020 research and innovation programme under Grant agreement no. 681002 (EU-ToxRisk). This work was further supported by the Institute of Health Carlos III and cofinanced by FEDER through grants PI18/00993 and CP16/00097. Parts of this work contributed to the IATA Case Studies project (ENV/JM/WRPR(2020)31 - Case Study On The Use Of Integrated Approaches To Testing And Assessment For Prediction Of A 90 Day Repeated Dose Toxicity Study (OECD 408) For 2-Ethylbutyric Acid Using A Read-Across Approach From Other Branched Carboxylic Acids).

References

Abdel-Dayem, M.A., Elmarakby, A.A., Abdel-Aziz, A.A., Pye, C., Said, S.A., El-Mowafy, A.M., 2014. Valproate-induced liver injury: modulation by the omega-3 fatty acid DHA proposes a novel anticonvulsant regimen. Drugs R D 14, 85-94.

Aguayo-Orozco, A., Audouze, K., Siggaard, T., Barouki, R., Brunak, S., Taboureau, O., 2019. sAOP: linking chemical stressors to adverse outcomes pathway networks. Bioinformatics 35, 5391-5392.

Aguayo-Orozco, A., Bois, F.Y., Brunak, S., Taboureau, O., 2018. Analysis of Time-Series Gene Expression Data to Explore Mechanisms of Chemical-Induced Hepatic Steatosis Toxicity. Front Genet 9, 396.

Aires, C.C., Ijlst, L., Stet, F., Prip-Buus, C., de Almeida, I.T., Duran, M., Wanders, R.J., Silva, M.F., 2010. Inhibition of hepatic carnitine palmitoyl-transferase I (CPT IA) by valproyl-CoA as a possible mechanism of valproate-induced steatosis. Biochem Pharmacol 79, 792-799.

Aschauer, L., Gruber, L.N., Pfaller, W., Limonciel, A., Athersuch, T.J., Cavill, R., Khan, A., Gstraunthaler, G., Grillari, J., Grillari, R., Hewitt, P., Leonard, M.O., Wilmes, A., Jennings, P., 2013. Delineation of the key aspects in the regulation of epithelial monolayer formation. Mol Cell Biol 33, 2535-2550.

Ball, N., Cronin, M.T., Shen, J., Blackburn, K., Booth, E.D., Bouhifd, M., Donley, E., Egnash, L., Hastings, C., Juberg, D.R., Kleensang, A., Kleinstreuer, N., Kroese, E.D., Lee, A.C., Luechtefeld, T., Maertens, A., Marty, S., Naciff, J.M., Palmer, J., Pamies, D., Penman, M., Richarz, A.N., Russo, D.P., Stuard, S.B., Patlewicz, G., van Ravenzwaay, B., Wu, S., Zhu, H., Hartung, T., 2016. Toward Good Read-Across Practice (GRAP) guidance. ALTEX 33, 149-166.

Bell, C.C., Hendriks, D.F., Moro, S.M., Ellis, E., Walsh, J., Renblom, A., Fredriksson Puigvert, L., Dankers, A.C., Jacobs, F., Snoeys, J., Sison-Young, R.L., Jenkins, R.E., Nordling, A., Mkrtchian, S., Park, B.K., Kitteringham, N.R., Goldring, C.E., Lauschke, V.M., Ingelman-Sundberg, M., 2016. Characterization of primary human hepatocyte spheroids as a model system for drug-induced liver injury, liver function and disease. Sci Rep 6, 25187.

BG Chemie, 2000. 2-Ethylhexanoic acid 06/2000.

Blackburn, K., Stuard, S.B., 2014. A framework to facilitate consistent characterization of read across uncertainty. Regulatory Toxicology and Pharmacology 68, 353-362.

Cariello, M., Piccinin, E., Moschetta, A., 2021. Transcriptional Regulation of Metabolic Pathways via Lipid-Sensing Nuclear Receptors PPARs, FXR, and LXR in NASH. Cellular and Molecular Gastroenterology and Hepatology 11, 1519-1539.

Chang, T.K.H., Abbott, F.S., 2006. Oxidative Stress as a Mechanism of Valproic Acid-Associated Hepatotoxicity. Drug Metabolism Reviews 38, 627-639.

Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., McMorran, R., Wiegers, J., Wiegers, T.C., Mattingly, C.J., 2019. The Comparative Toxicogenomics Database: update 2019. Nucleic Acids Res 47, D948-D954.

Dempster, A.P., 2008. Upper and Lower Probabilities Induced by a Multivalued Mapping, Classic Works of the Dempster-Shafer Theory of Belief Functions, pp. 57-72.

Donato, M.T., Tolosa, L., Jimenez, N., Castell, J.V., Gomez-Lechon, M.J., 2012. High-content imaging technology for the evaluation of drug-induced steatosis using a multiparametric cell-based assay. J Biomol Screen 17, 394-400.

ECHA, 2017. Read-Across Assessment Framework (RAAF), ECHA-17-R-01-EN ed. European Chemical Agency,.

Escher, S.E., Kamp, H., Bennekou, S.H., Bitsch, A., Fisher, C., Graepel, R., Hengstler, J.G., Herzler, M., Knight, D., Leist, M., Norinder, U., Ouédraogo, G., Pastor, M., Stuard, S., White, A., Zdrazil, B., van de Water, B., Kroese, D., 2019. Towards grouping concepts based on new approach methodologies in chemical hazard assessment: the read-across approach of the EU-ToxRisk project. Arch Toxicol 93, 3643-3667.

Espandiari, P., Zhang, J., Schnackenberg, L.K., Miller, T.J., Knapton, A., Herman, E.H., Beger, R.D., Hanig, J.P., 2008. Age-related differences in susceptibility to toxic effects of valproic acid in rats. Journal of Applied Toxicology 28, 628-637.

EU, 2010. Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. . OJ, pp. 33-79.
European Food Safety, A., Bronzwaer, S., Kass, G., Robinson, T., Tarazona, J., Verhagen, H., Verloo, D., Vrbos, D., Hugas, M., 2019. Food Safety Regulatory Research Needs 2030. EFSA J 17, e170622.
Fisher, C., 2019. Abstracts of the 55th Congress of the European Societies of Toxicology (EUROTOX 2019) TOXICOLOGY SCIENCE PROVIDING SOLUTIONS. Toxicology Letters 314, S1-S309.
Gadaleta, D., Golbamaki Bakhtyari, A., Lavado, G.J., Roncaglioni, A., Benfenati, E., 2020. Automated integration of structural, biological and metabolic similarities to improve read-across. ALTEX.
Grinberg, M., Stober, R.M., Edlund, K., Rempel, E., Godoy, P., Reif, R., Widera, A., Madjar, K., Schmidt-Heck, W., Marchan, R., Sachinidis, A., Spitkovsky, D., Hescheler, J., Carmo, H., Arbo, M.D., van de Water, B., Wink, S., Vinken, M., Rogiers, V., Escher, S., Hardy, B., Mitic, D., Myatt, G., Waldmann, T., Mardinoglu, A., Damm, G., Seehofer, D., Nussler, A., Weiss, T.S., Oberemm, A., Lampen, A., Schaap, M.M., Luijten, M., van Steeg, H., Thasler, W.E., Kleinjans, J.C., Stierum, R.H., Leist, M., Rahnenfuhrer, J., Hengstler, J.G., 2014. Toxicogenomics directory of chemically exposed human hepatocytes. Arch Toxicol 88, 2261-2287.
Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Roi, A.J., Prieto, P., Sabbioni, E., Scott, L., Worth, A., Zuang, V., 2004. A modular approach to the ECVAM principles on test validity. Altern Lab Anim 32, 467-472.
Helman, G., Shah, I., Williams, A.J., Edwards, J., Dunne, J., Patlewicz, G., 2019. Generalized Read-Across (GenRA): A workflow implemented into the EPA CompTox Chemicals Dashboard. ALTEX 36, 462-465.
Ibrahim, M.A., 2012. Evaluation of Hepatotoxicity of Valproic acid in albino mice, Histological and Histoistochemical studies. Life Science Journal-Acta Zhengzhou University Overseas Edition 9, 153-159.
Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., Yamada, H., 2015. Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res 43, D921-927.
Jennings, P., Aydin, S., Bennett, J., McBride, R., Weiland, C., Tuite, N., Gruber, L.N., Perco, P., Gaora, P.O., Ellinger-Ziegelbauer, H., Ahr, H.J., Kooten, C.V., Daha, M.R., Prieto, P., Ryan, M.P., Pfaller, W., McMorrow, T., 2009. Inter-laboratory comparison of human renal proximal tubule (HK-2) transcriptome alterations due to Cyclosporine A exposure and medium exhaustion. Toxicol In Vitro 23, 486-499.
Johnson, R., Wolf, J., Braunbeck, T., 2010. Guidance document on the diagnosis of endocrine-related histopathology in fish gonads.
Juberg, D.R., David, R.M., Katz, G.V., Bernard, L.G., Gordon, D.R., Vlaovic, M.S., Topping, D.C., 1998. 2-ethylhexanoic acid: Subchronic oral toxicity studies in the rat and mouse. Food and Chemical Toxicology 36, 429-436.
Knapp, A.C., Todesco, L., Beier, K., Terracciano, L., Sagesser, H., Reichen, R., Krahenbuhl, S., 2008. Toxicity of valproic acid in mice with decreased plasma and tissue carnitine stores. Journal of Pharmacology and Experimental Therapeutics 324, 568-575.
Krebs, A., Waldmann, T., Wilks, M.F., Van Vugt-Lussenburg, B.M.A., Van der Burg, B., Terron, A., Steger-Hartmann, T., Ruegg, J., Rovida, C., Pedersen, E., Pallocca, G., Luijten, M., Leite, S.B., Kustermann, S., Kamp, H., Hoeng, J., Hewitt, P., Herzler, M., Hengstler, J.G., Heinonen, T., Hartung, T., Hardy, B., Gantner, F., Fritsche, E., Fant, K., Ezendam, J., Exner, T., Dunkern, T., Dietrich, D.R., Coecke, S., Busquet, F., Braeuning, A., Bondarenko, O., Bennekou, S.H., Beilmann, M., Leist, M., 2019. Template for the description of cell-based toxicological test methods to allow evaluation and regulatory use of the data. ALTEX 36, 682-699.
Limonciel, A., Aschauer, L., Wilmes, A., Prajczer, S., Leonard, M.O., Pfaller, W., Jennings, P., 2011. Lactate is an ideal non-invasive marker for evaluating temporal alterations in cell stress and toxicity in repeat dose testing regimes. Toxicol In Vitro 25, 1855-1862.
Limonciel, A., Wilmes, A., Aschauer, L., Radford, R., Bloch, K.M., McMorrow, T., Pfaller, W., van Delft, J.H., Slattery, C., Ryan, M.P., Lock, E.A., Jennings, P., 2012. Oxidative stress induced by potassium

bromate exposure results in altered tight junction protein expression in renal proximal tubule cells. Archives of Toxicology 86, 1741-1751.
Loscher, W., Wahnschaffe, U., Honack, D., Wittfoht, W., Nau, H., 1992. Effects of valproate and E-2-en-valproate on functional and morphological parameters of rat liver. I. Biochemical, histopathological and pharmacokinetic studies. Epilepsy Res 13, 187-198.
Low, Y., Sedykh, A., Fourches, D., Golbraikh, A., Whelan, M., Rusyn, I., Tropsha, A., 2013. Integrative chemical-biological read-across approach for chemical hazard classification. Chem Res Toxicol 26, 1199-1208.
Maldonado, E.M., Fisher, C.P., Mazzatti, D.J., Barber, A.L., Tindall, M.J., Plant, N.J., Kierzek, A.M., Moore, J.B., 2018. Multi-scale, whole-system models of liver metabolic adaptation to fat and sugar in non-alcoholic fatty liver disease. npj Systems Biology and Applications 4, 33.
Marvel, S.W., To, K., Grimm, F.A., Wright, F.A., Rusyn, I., Reif, D.M., 2018. ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models. BMC Bioinformatics 19, 80.
Mav, D., Shah, R.R., Howard, B.E., Auerbach, S.S., Bushel, P.R., Collins, J.B., Gerhold, D.L., Judson, R.S., Karmaus, A.L., Maull, E.A., Mendrick, D.L., Merrick, B.A., Sipes, N.S., Svoboda, D., Paules, R.S., 2018. A hybrid gene selection approach to create the S1500+ targeted gene sets for use in high-throughput transcriptomics. PLoS One 13, e0191105.
Mellor, C.L., Steinmetz, F.P., Cronin, M.T., 2016. The identification of nuclear receptors associated with hepatic steatosis to develop and extend adverse outcome pathways. Crit Rev Toxicol 46, 138-152.
Mulisch, M., Welsch, U., 2015. Romeis – Mikroskopische Technik. Springer Verlag, Heidelberg.
OECD, 2005. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. , ENV/JM/MONO (2005)14 ed. Organisation for Economic Co-operation and Development.
OECD, 2013. Test No. 236: Fish Embryo Acute Toxicity (FET) Test. Organisation for Economic Co-operation and Development,.
Patlewicz, G., Cronin, M.T.D., Helman, G., Lambert, J.C., Lizarraga, L.E., Shah, I., 2018. Navigating through the minefield of read-across frameworks: A commentary perspective. Computational Toxicology 6, 39-54.
Patterson, E.A., Whelan, M.P., Worth, A.P., 2021. The role of validation in establishing the scientific credibility of predictive toxicology approaches intended for regulatory application. Comput Toxicol 17, 100144.
Pawlak, M., Lefebvre, P., Staels, B., 2015. Molecular mechanism of PPARα action and its impact on lipid metabolism, inflammation and fibrosis in non-alcoholic fatty liver disease. Journal of Hepatology 62, 720-733.
Pestana, C.B., Firman, J.W., Cronin, M.T.D., 2021. Incorporating lines of evidence from New Approach Methodologies (NAMs) to reduce uncertainties in a category based read-across: A case study for repeated dose toxicity. Regulatory Toxicology and Pharmacology 120, 104855.
Petrone, P.M., Simms, B., Nigsch, F., Lounkine, E., Kutchukian, P., Cornett, A., Deng, Z., Davies, J.W., Jenkins, J.L., Glick, M., 2012. Rethinking molecular similarity: comparing compounds on the basis of biological activity. ACS Chem Biol 7, 1399-1409.
Poser, I., Sarov, M., Hutchins, J.R., Heriche, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M.M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K.P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A.F., Peters, J.M., Buchholz, F., Hyman, A.A., 2008. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods 5, 409-415.
R Core Team, 2020. The R Project for Statistical Computing.
Rathman, J.F., Yang, C., Zhou, H., 2018. Dempster-Shafer theory for combining in silico evidence and estimating uncertainty in chemical risk assessment. Computational Toxicology 6, 16-31.
Schultz, T.W., Cronin, M.T.D., 2017. Lessons learned from read-across case studies for repeated-dose toxicity. Regul Toxicol Pharmacol 88, 185-191.

Schumacher, J.D., Guo, G.L., 2015. Mechanistic review of drug-induced steatohepatitis. Toxicol Appl Pharmacol 289, 40-47.
Shäfer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press
Shah, I., Liu, J., Judson, R.S., Thomas, R.S., Patlewicz, G., 2016. Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information. Regul Toxicol Pharmacol 79, 12-24.
Silva, M.F., Aires, C.C., Luis, P.B., Ruiter, J.P., L, I.J., Duran, M., Wanders, R.J., Tavares de Almeida, I., 2008. Valproic acid metabolism and its effects on mitochondrial fatty acid oxidation: a review. J Inherit Metab Dis 31, 205-216.
Strähle, U., Scholz, S., Geisler, R., Greiner, P., Hollert, H., Rastegar, S., Schumacher, A., Selderslaghs, I., Weiss, C., Witters, H., Braunbeck, T., 2012. Zebrafish embryos as an alternative to animal experiments—a commentary on the definition of the onset of protected life stages in animal welfare regulations. Reprod Toxicol 33, 128-132.
Sugimoto, T., Woo, M., Nishida, N., Takeuchi, T., Sakane, Y., Kobayashi, Y., 1987. Hepatotoxicity in rat following administration of valproic acid. Epilepsia 28, 142-146.
Tanaka, T., Yamamoto, J., Iwasaki, S., Asaba, H., Hamura, H., Ikeda, Y., Watanabe, M., Magoori, K., Ioka, R.X., Tachibana, K., Watanabe, Y., Uchiyama, Y., Sumi, K., Iguchi, H., Ito, S., Doi, T., Hamakubo, T., Naito, M., Auwerx, J., Yanagisawa, M., Kodama, T., Sakai, J., 2003. Activation of peroxisome proliferator-activated receptor δ induces fatty acid β-oxidation in skeletal muscle and attenuates metabolic syndrome. Proceedings of the National Academy of Sciences 100, 15924-15929.
Tolosa, L., Carmona, A., Castell, J.V., Gomez-Lechon, M.J., Donato, M.T., 2015. High-content screening of drug-induced mitochondrial impairment in hepatic cells: effects of statins. Arch Toxicol 89, 1847-1860.
Tolosa, L., Pinto, S., Donato, M.T., Lahoz, A., Castell, J.V., O'Connor, J.E., Gomez-Lechon, M.J., 2012. Development of a multiparametric cell-based protocol to screen and classify the hepatotoxicity potential of drugs. Toxicol Sci 127, 187-198.
Tong, V., Teng, X.W., Chang, T.K., Abbott, F.S., 2005. Valproic acid I: time course of lipid peroxidation biomarkers, liver toxicity, and valproic acid metabolite levels in rats. Toxicol Sci 86, 427-435.
Turnbull, D.M., Rawlins, M.D., Weightman, D., Chadwick, D.W., 1983. Plasma concentrations of sodium valproate: Their clinical value. Annals of Neurology 14, 38-42.
US EPA, 2018. Strategic plan to promote the development and implementation of alternative test methods within the TSCA program.
van Breda, S.G.J., Claessen, S.M.H., van Herwijnen, M., Theunissen, D.H.J., Jennen, D.G.J., de Kok, T.M.C.M., Kleinjans, J.C.S., 2018. Integrative omics data analyses of repeated dose toxicity of valproic acid in vitro reveal new mechanisms of steatosis induction. Toxicology 393, 160-170.
van der Burg, B., van der Linden, S., Man, H.-y., Winter, R., Jonker, L., van Vugt-Lussenburg, B., Brouwer, A., 2013. A Panel of Quantitative Calux® Reporter Gene Assays for Reliable High-Throughput Toxicity Screening of Chemicals and Complex Mixtures, High-Throughput Screening Methods in Toxicity Testing, pp. 519-532.
Vrijenhoek, N.G., Wehr, M.M., Kunnen, S., Moné, M.J., Escher, S.E., Water, B.v.d., 2021. Application of high-throughput transcriptomics for mechanism-based biological read-across of short-chain carboxylic acid analogues of valproic acid. ALTEX submitted.
Wahli, W., Michalik, L., 2012. PPARs at the crossroads of lipid signaling and inflammation. Trends Endocrinol Metab 23, 351-363.
Wieser, M., Stadler, G., Jennings, P., Streubel, B., Pfaller, W., Ambros, P., Riedl, C., Katinger, H., Grillari, J., Grillari-Voglauer, R., 2008. hTERT alone immortalizes epithelial cells of renal proximal tubules without changing their functional characteristics. Am J Physiol Renal Physiol 295, F1365-1375.
Wilson, J.M., Bunte, R.M., Carty, A.J., 2009. Evaluation of rapid cooling and tricaine methanesulfonate (MS222) as methods of euthanasia in zebrafish (Danio rerio). J Am Assoc Lab Anim Sci 48, 785-789.

Wink, S., Hiemstra, S., Herpers, B., van de Water, B., 2017. High-content imaging-based BAC-GFP toxicity pathway reporters to assess chemical adversity liabilities. Arch Toxicol 91, 1367-1383.
Wink, S., Hiemstra, S.W., Huppelschoten, S., Klip, J.E., van de Water, B., 2018. Dynamic imaging of adaptive stress response pathway activation for prediction of drug induced liver injury. Arch Toxicol 92, 1797-1814.
Zhang, L.F., Liu, L.S., Chu, X.M., Xie, H., Cao, L.J., Guo, C., A, J.Y., Cao, B., Li, M.J., Wang, G.J., Hao, H.P., 2014. Combined effects of a high-fat diet and chronic valproic acid treatment on hepatic steatosis and hepatotoxicity in rats. Acta Pharmacol Sin 35, 363-372.
Zhu, H., Bouhifd, M., Donley, E., Egnash, L., Kleinstreuer, N., Kroese, E.D., Liu, Z., Luechtefeld, T., Palmer, J., Pamies, D., Shen, J., Strauss, V., Wu, S., Hartung, T., 2016. Supporting read-across using biological data. ALTEX 33, 167-182.

A read-across case study on chronic toxicity of branched carboxylic acids (1): integration of mechanistic evidence from new approach methodologies (NAMs) to explore a common mode of action.

Highlights:

- Integration of new approach methods into human hazard characterization
- Read-across on systemic toxicity assessment supported by New Approach Methods
- AOP informed testing strategy for hazard characterization
- Uncertainty characterization
- Case studies as alternative tools for model validation