



## Evaluation of an alternative *in vitro* test battery for detecting reproductive toxicants in a grouping context

E. Dinant Kroese<sup>a,\*</sup>, Sieto Bosgra<sup>a</sup>, Harrie E. Buist<sup>a</sup>, Geertje Lewin<sup>e</sup>, Sander C. van der Linden<sup>f</sup>, Hai-yen Man<sup>f</sup>, Aldert H. Piersma<sup>b,c</sup>, Emiel Rorije<sup>c</sup>, Sjors H.W. Schulpen<sup>b</sup>, Michael Schwarz<sup>d</sup>, Frederik Uibel<sup>d</sup>, Barbara M.A. van Vugt-Lussenburg<sup>f</sup>, Andre P.M. Wolterbeek<sup>a</sup>, Bart van der Burg<sup>f</sup>

<sup>a</sup> TNO, Zeist, The Netherlands

<sup>b</sup> Institute for Risk Assessment Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

<sup>c</sup> Center for Health Protection, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

<sup>d</sup> Institute of Experimental and Clinical Pharmacology and Toxicology, Department of Toxicology, University of Tübingen, Tübingen, Germany

<sup>e</sup> Fraunhofer Institute for Toxicology and Experimental Medicine (ITEM), Hannover, Germany

<sup>f</sup> BioDetection Systems BV, Amsterdam, The Netherlands

### ARTICLE INFO

#### Article history:

Received 1 May 2014

Received in revised form

17 September 2014

Accepted 1 October 2014

Available online 14 October 2014

#### Keywords:

Alternative model

Developmental toxicity

Read across

Grouping

CALUX

Embryonic stem cell test

Zebrafish embryotoxicity assay

ReProGlo

### ABSTRACT

Previously we showed a battery consisting of CALUX transcriptional activation assays, the ReProGlo assay, and the embryonic stem cell test, and zebrafish embryotoxicity assay as 'apical' tests to correctly predict developmental toxicity for 11 out of 12 compounds, and to explain the one false negative [7]. Here we report on applying this battery within the context of grouping and read across, put forward as a potential tool to fill data gaps and avoid animal testing, to distinguish *in vivo* non- or weak developmental toxicants from potent developmental toxicants within groups of structural analogs. The battery correctly distinguished 2-methylhexanoic acid, monomethyl phthalate, and monobutyltin trichloride as non- or weak developmental toxicants from structurally related developmental toxicants valproic acid, monoethylhexyl phthalate, and tributyltin chloride, respectively, and, therefore, holds promise as a biological verification model in grouping and read across approaches. The relevance of toxicokinetic information is indicated.

© 2014 Elsevier Inc. All rights reserved.

### 1. Introduction

Testing strategies in toxicology are constantly being challenged, especially in the past decade when important regulatory changes occurred with the introduction of the EU legislation REACH [1]. This regulation enforces hazard and risk assessment of over 30 thousand preregistered chemicals before 2018. It is foreseen that this will require great numbers of test animals, especially for reproductive and developmental toxicity testing [2]. Therefore, innovative strategies that replace and/or reduce animal testing are urgently needed. Such strategies are also stimulated by REACH that prescribes animal testing only as a last resort [1].

Substantial efforts have already been undertaken to develop alternative methods for the assessment of reproductive and development toxicity [3,4]. However, none of these assays alone can cover the whole mammalian reproductive cycle due to its inherent complexity [3,4]. Therefore, recent studies have attempted to combine several *in vitro* assays into a test battery instead of applying individual assays. A feasibility study published by Schenk et al. [5] studied 10 compounds in a battery of 14 assays [5]. This battery was able to detect all reproductive toxicants for which the modes of action were actually represented in at least one of the assays. Another example is represented by the US EPA's ToxCast program, which used a large group of high throughput alternative assays to analyze the developmental toxicity of some 300 chemicals [6]. They showed >70% balanced accuracy of detecting developmental toxicants across some 650 assays reduced to a multivariate signature for chemicals identified as developmental toxicants in rat or rabbit guideline studies. This relatively low prediction potential might be

\* Corresponding author at: TNO, PO Box 360, 3700 AJ, Zeist, The Netherlands.

Tel.: +31 088 866 1508.

E-mail address: [dinant.kroese@tno.nl](mailto:dinant.kroese@tno.nl) (E.D. Kroese).

due to the fact that most assays applied were not intended to focus on developmental toxicity, and had only limited representation of assays for morphogenesis and differentiation. In a previous study of the European Framework Programme 7 project ChemScreen, therefore, the zebrafish embryo test (ZET) and the embryonic stem cell test (EST) were included as more apical assays [7], to detect effects on development of a whole egg from fertilization until the hatching stage 72 h later [8], and to detect effects on cellular differentiation of cardiomyocytes [9], respectively. Also the medium throughput screen ReProGlo, monitoring interference with the WNT pathway [10], and assays for CYP17 and CYP19, to detect effects on steroidogenesis, enzymes essential for reproductive hormone homeostasis [11,12] were included. Finally, a panel of 24 high throughput CALUX assays were added that measure changes in activity of key transcription factors, varying from nuclear receptors, including reproductive hormone receptors, to transcription factors involved in cellular signaling [13,14]. The approach also encompassed toxicokinetic modeling to reveal whether effective *in vitro* concentrations observed in the battery are in the range expected from the *in vivo* reproductive toxicity data, in consideration of suggestions by Daston et al. [15]. As previously reported [7], this battery approach (including the toxicokinetic model) successfully identified eleven out of twelve compounds with varying mechanisms of action, with the unidentified compound, glufosinate ammonium, having a mechanism not covered by the battery. This result encourages to further optimize this battery into one ultimately able to detect all reprotoxic compounds.

The use and interpretation of battery results very much depends on the purpose of testing and the information that may already be available. In the absence of any *in vivo* test information relevant to potential reproductive toxicity and/or in the absence of any structural alerts pointing to such effects, the battery could be applied as a filter optimizing and/or reducing the testing of potential reproductive toxicants in animal studies [16]. The battery could also help by prioritizing chemicals for further investigation and/or by selecting candidate (pharmaceutical) compounds for further development [17]. On the other hand, when there are clear indications for potential reproductive toxicity based on close structural similarity of a query chemical to a reproductive toxicant, the battery could be used to confirm any reproductive toxicity, and to avoid any further *in vivo* studies. Ideally, this battery should also be capable to correctly distinguish reproductive toxicants from non-reproductive toxicants, even when both have structural similarity. The purpose of this investigation was to explore this discriminating capability of the battery *via* testing three groups consisting of structurally related chemicals, differing in their reproductive toxicity. Therefore, as a follow-up of our previous paper, we have tested two valproic acid (VPA) analogs: 2-ethylhexanoic acid (EHA), and 2-methylhexanoic acid (MHA), two analogs of monoethylhexyl phthalate (MEHP): monobenzyl phthalate (MBzP), and monomethyl phthalate (MMP), and three organotin analogs of dioctyltin dichloride: tributyltin chloride, dibutyltin dichloride (TBTC), dibutyltin chloride dibutyltin dichloride (DBTC) and monobutyltin trichloride (MBTC). Of these analogs to be tested, MHA, and MMP were not considered reproductive toxicants, while MBTC is considered a weak reproductive toxicant, if at all. The outcome of this investigation will be discussed within the context of the purpose of the battery within an integrated testing strategy.

## 2. Materials and methods

### 2.1. General experimental set-up

As described in our previous manuscript [7], tests were selected at a special meeting of the ChemScreen consortium, based on

the following criteria: (1) Relevance to known mechanisms and endpoints involved in reproductive toxicity. (2) Availability at partner institutes. (3) Overlap with ReProTest feasibility study [5], to allow comparison, and 4. with preference to higher throughput assays that can be automated, but also including more apical lower throughput assays. The battery consisted of the cardiac embryonic stem cell test (EST), the zebrafish embryotoxicity test (ZET), the ReProGlo assay (ReProGlo), and a panel of CALUX assays (see Table 2). The individual tests were performed, according to previously standardized protocols established in the various collaborating centers as further detailed below.

#### 2.1.1. The embryonic stem cell test (EST)

Pluripotent mouse D3 embryonic stem cells (ESC; ATCC, Rockville, MD) were routinely subcultured every 2–3 days and grown as a monolayer in complete medium, consisting of DMEM (Gibco BRL, Gaithersburg, MD) supplemented with 20% fetal bovine serum (Hyclone, Logan, UT), 1% non-essential amino acids (Gibco BRL, Gaithersburg, MD), 1% penicillin/streptomycin (Gibco BRL, Gaithersburg, MD), 2 mM L-glutamine (Gibco BRL, Gaithersburg, MD) and 0.1 mM  $\beta$ -mercapto-ethanol (Sigma–Aldrich, Zwijndrecht, The Netherlands). Leukemia inhibitory factor (LIF; Chemicon, Temecula, CA) was added directly to the culture disk in a final concentration of 1000 units/ml. The cells were maintained in a humidified atmosphere at 37 °C and 5% CO<sub>2</sub>. Embryoid body (EB) formation was used as the initial step for the EST differentiation assay. EB were obtained *via* hanging drop culture in complete medium without LIF [18]. In brief, stem cell suspensions ( $3.75 \times 10^4$  cells/ml) were placed on ice before the set up of the culture. Drops (20  $\mu$ l) containing 750 cells were placed onto the inner side of the lid of a Petri dish filled with phosphate buffered saline (PBS; Gibco BRL, Gaithersburg, MD) and incubated at 37 °C, 90% relative humidity and 5% CO<sub>2</sub>. After 3 days of hanging drop culture EB had formed and these were subsequently transferred to bacterial Petri dishes (Greiner Bio-one, Frickenhausen, Germany). On day 5, 24 EB were plated one per well into 24 well tissue culture plates (TPP, Trasadingen, Switzerland). Differentiation was determined microscopically at day 10 of differentiation by inspection of EB outgrowths into contracting myocardial cells. EB were considered as cardiomyocyte positive if at least one contracting focus was present. The number of positive EB was expressed as fraction of total EB examined. Data are expressed as follows: ‘cytotox’ values are IC<sub>50</sub> (concentration inhibiting cell viability by 50%), and ‘diff010’ values are ID<sub>50</sub> (concentration inhibiting differentiation by 50%), respectively, after 10 days of exposure in the assay. The test was considered positive if ID<sub>50</sub>  $\leq$  IC<sub>50</sub>, or if IC<sub>50</sub>  $\leq$  100  $\mu$ M ( $\geq 4$  in the table). If kinetic data were available for a compound, these cut-off values could be shifted up or downwards according to the calculated relevant *in vivo* plasma concentrations.

#### 2.1.2. The zebrafish embryotoxicity test (ZET)

The ZET was performed as described previously [19]. In brief, fertilized batches of eggs with a fertilization rate of at least 90% were collected within 30–60 min after spawning using 800  $\mu$ m mesh and were rinsed with MilliQ water to remove impurities. The fertilized eggs were then directly transferred into different Petri dishes containing the test compounds at selected concentrations. Subsequently, embryos at the 8–16 cells stage were selected and transferred to a 12-well plate containing 3 mL of the test medium per well. Embryos were kept in an incubator at  $28.0 \pm 1$  °C with a photoperiod of 12 h light:12 h dark. Morphological changes were evaluated at 24, 48, 72 and 96 hpf, as described recently [19]. Experiments were considered valid if survival rates in controls were  $>90\%$  [20]. The morphological scores for each experiment were normalized and expressed as a percentage compared to controls, combined and presented as mean  $\pm$  standard error of the mean (SEM). The

EC50 (cytotox in the table) was defined as the concentration at which there was a 50% decrease in survival in comparison with the control. The ED50 was defined as the concentration at which there was a 50% increase (or decrease) in the incidence of a given developmental parameter in comparison with the control. The test was considered positive if  $ED50 \leq EC50$ , or if  $EC50 \leq 100 \mu\text{M}$  ( $\geq 4'$  in the table). If kinetic data were available for a compound, these cut-off values could be shifted up or downwards according to the calculated relevant *in vivo* plasma concentrations.

### 2.1.3. The ReProGlo assay

The previously described protocol for the ReProGlo assay [10] was adapted to an automated liquid handling system. In brief, murine embryonic stem cells that were stably transfected with a WNT-responsive luciferase reporter vector were seeded manually in DMEM high glucose medium supplemented with 15% fetal calf serum, 2 mM L-glutamine, 100  $\mu\text{M}$  non-essential amino acids (all purchased from Invitrogen, Karlsruhe, Germany) 100  $\mu\text{M}$   $\beta$ -mercaptoethanol (Sigma, Munich, Germany), 0.5% penicillin/streptomycin (40 U/ml penicillin, 40  $\mu\text{g}/\text{ml}$  streptomycin, Biochrom, Berlin, Germany) and 100  $\mu\text{g}/\text{ml}$  geneticin (G418, Sigma) on 96 well plates (Falcon, BD Biosciences, Heidelberg, Germany). Cells were treated with the test chemicals on day 2 using the liquid handling system (Freedom EVO, Tecan, Crailsheim, Germany). Pre-dilutions of test chemicals were prepared in water (if possible) or DMSO, and then mixed with DMEM without phenol red and geneticin. Concentration of DMSO in the assay was 0.2%. On day 3, cell viability was determined by applying the Alamar Blue assay. Cells were lysed using the liquid handling system, and a luciferase assay was performed to determine reporter activity (see [7] for details). Cell viability values were normalized for each well relative to the mean of all solvent control values of the respective experiment yielding relative viability values. Then, the absolute luciferase values of each well were normalized to the mean of the luciferase values of the solvent controls to obtain relative luciferase values for each well. Finally, these latter values were divided by the relative viability values for normalization. As endpoints for each assay, we define benchmark-concentrations. For the nonspecific effect of chemicals on cell viability, we define a benchmark of 80% viability termed  $BMC_{0.8}$ . For the specific effect on WNT pathway up- or downregulation, we define a doubling ( $BMC_2$ ) or bisection ( $BMC_{0.5}$ ) in reporter activity. The test was considered positive if  $BMC_2$  or  $BMC_{0.5}$  are less than twice as low as  $BMC_{0.8}$ . If  $BMC_{0.8}$  are less than twice as high as  $BMC_2$  or  $BMC_{0.5}$  the result was called 'unspecific'. If kinetic data were available for a compound, the calculated relevant *in vivo* plasma concentrations were considered in the judgment.

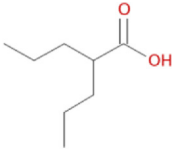
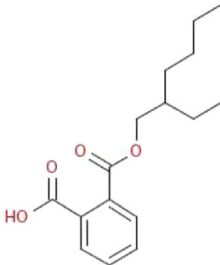
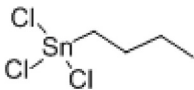
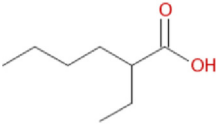
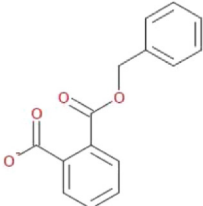
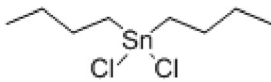
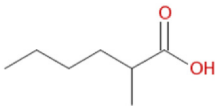
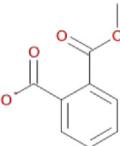
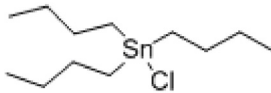
### 2.1.4. CALUX battery

The CALUX battery of stable reporter gene assays [14] comprised assays and test conditions that were selected from a larger panel of CALUX cells because of their possible relevance to reproductive toxicity testing. The selected panel consisted of the human T47D cell based ER CALUX line containing an estrogen responsive reporter gene and endogenous expression of multiple steroid receptors providing a model that also accounts for receptor interactions [21]. Two additional lines were included using rat H4IIE liver cells expressing the dioxin receptor endogenously (the DR CALUX and PAH CALUX line, respectively; [22–24]. In addition a range of highly selective human U2OS cell based lines were used, including doubly (*i.e.*, reporter gene and receptor) transfected ER $\alpha$ -, PR (progesterone receptor)-, GR (glucocorticoid receptor)-, AR (androgen receptor) CALUX lines [13], and an extension of the panel with peroxisome proliferator activated receptor PPAR $\alpha$ - and PPAR $\gamma$  CALUX assays [25,26], TR (thyroid receptor)  $\beta$ -, RAR (retinoic acid receptor)-, ER (estrogen receptor)  $\beta$ -CALUX [27]. This screening

panel was completed with a range of U2OS-based singly transfected lines expressing a reporter gene only, which are designed to selectively measure the activity of main intracellular signaling pathways. This included the assays to assess transcriptional activation by the p53 protein, p53 CALUX line [28]; transcriptional activation of the cyclin-dependent kinase inhibitor p21 promoter region; p21 CALUX, the oxidative stress responsive nrf-2 pathway, Nrf2 CALUX; the endoplasmic reticulum stress pathway, ESRE CALUX. The Wnt signaling pathway, TCF CALUX; the activator protein 1 pathway, AP1 CALUX; the nuclear factor  $\kappa\text{B}$  pathway, NF $\kappa\text{B}$  CALUX; and the hypoxia-inducible factor induced pathway, HIF1 $\alpha$  CALUX [27]. In general, tests were carried out in a way to measure agonistic effects of compounds, while in several cases deemed relevant, tests were carried out in the additional presence of reference agonist at an EC50 level in order to assess antagonism [13]. These assays were additionally labeled with '(anti)' adjunct. Only those CALUX assays that gave at least one positive outcome in this study are included in Table 2. The CALUX cells were cultured as described before [13]. U2OS- and T47D-CALUX cells were routinely subcultured every 3–4 days in growth medium consisting of DMEM (Gibco) supplemented with 7.5% fetal calf serum, 1 $\times$  non-essential amino acids (Gibco) and 10 U/ml penicillin and 10  $\mu\text{g}/\text{ml}$  streptomycin. H4IIE-CALUX cells were routinely subcultured every 3–4 days in growth medium consisting of  $\alpha$ MED (Gibco) supplemented with 10% fetal calf serum. All cell types were maintained at 37 °C and 5% CO<sub>2</sub> at all times. All CALUX assays were performed in assay medium, consisting of DMEM without phenol-red indicator (Gibco) supplemented with 5% DCC-stripped fetal calf serum, 1 $\times$  non-essential amino acids (Gibco) and 10 U/ml penicillin and 10  $\mu\text{g}/\text{ml}$  streptomycin.

For the CALUX assay, a cell suspension in assay medium was made of  $1 \times 10^5$  cells/ml (U2OS and T47D) or  $4 \times 10^5$  cells/ml (H4IIE), and the 384-well plates were seeded with 30  $\mu\text{l}/\text{well}$  cell suspension using a MicroFlo Select dispenser (BioTek). After 24 h, exposure medium was prepared by adding 2% of test compound dilution series in DMSO to a 96-wells plate with assay medium. Of this exposure mixture, 30  $\mu\text{l}$  was added to the assay plates containing the CALUX cells, resulting in final DMSO concentrations of 1%. The final concentrations of the compounds in the wells were  $3 \times 10^{-11}$  to  $10^{-3}$  M in 0.5 log unit increments. Additionally, two DMSO blanks and a full dose response curve of the relevant reference compound were included on each plate. All samples were tested in triplicate. The preparation of the compound dilution series as well as the exposure of the cells were performed on a Hamilton Starlet liquid handling robot coupled to a Cytomat incubator. After 24 h exposure (PAH-CALUX: 4 h exposure) the exposure medium was removed and 10  $\mu\text{l}/\text{well}$  triton-lysis buffer was added by the MicroFlo Select. Subsequently, the luciferase signal was measured in a luminometer (Berthold), as described before [13]. Data were processed as follows: after background (DMSO only) subtraction, for each plate the maximum signal induced by the reference compound was set at 100%. The signals induced by the test compounds were expressed in % of maximum reference signal. Then, the concentration where the test compound reached 10% of the maximum effect of the reference compound was determined with Graphpad Prism software (log agonist binding curve fit, extrapolate unknowns). This concentration, designated EC10, was used to rank the potencies of the compounds. Results in which the cytotox CALUX was found positive (*i.e.* activities equal to or higher than the EC10 level) were excluded. Since CALUX tests generally are more responsive in comparison to the other assays used in this study, a threshold for positivity was used. To avoid false-positives, in the absence of a validated threshold for the reporter gene assays this threshold was arbitrarily set at  $10^{-5}$  M ('5'), close to the threshold that is used in OECD guideline TG455 for estrogen reporter gene assays [29]. If kinetic data were available for a compound,

**Table 1**  
Chemical structures of investigated groupings.

Alkyl alcanoic acids	Phthalates	Organotin chlorides
 <p>VPA</p>	 <p>MEHP</p>	 <p>MBTC</p>
 <p>EHA</p>	 <p>MBzP</p>	 <p>DBTC</p>
 <p>MHA</p>	 <p>MMP</p>	 <p>TBTC</p>

this cut-off value could be shifted up or downwards according to the calculated relevant *in vivo* plasma concentrations.

## 2.2. Test chemicals

The three groups of chemicals were selected on the basis of the following main criteria: (1) At least three structural similar within the groups and (2) one of the structural similar chemicals having a different *in vivo* toxicity result. Once the final list was decided upon, BioDetection Systems (BDS) created a chemical repository consisting of weighed aliquots in glass Teflon vials and distributed the chemicals to the testing labs. The selected groups of structurally related chemicals were: (1) sodium valproate (VPA; CAS no 1069-66-5), 2-ethylhexanoic acid (EHA; CAS no 149-57-5), and 2-methylhexanoic acid (MHA; CAS no 4536-23-6); (2) monoethylhexylphthalate (MEHP; CAS no 4376-20-9), monobenzyl phthalate (MBzP; CAS no 2528-16-7), and monomethyl phthalate (MMP; CAS no 4376-18-5), and (3) dioctyltin dichloride/dichlorodioctylstannane (DOTC; CAS no 3542-36-7), tributyltin chloride (TBTC; CAS no 1461-22-9), dibutyltin dichloride (DBTC; CAS no 683-18-1), and monobutyltin trichloride (MBTC; CAS no 1118-46-3), that are all depicted in Table 1. All chemicals were obtained from Sigma–Aldrich (Zwijndrecht, The Netherlands).

## 2.3. In vitro–in vivo comparison

Maternal plasma concentration–time data of VPA, EHA and MHA, given as single doses of 6.25, 12.5, and 14.1 mmol/kg, respectively, at day 12 of gestation, are presented in Scott et al. [30]. For the compounds added to the groups of phthalates (MBzP, MMP) and organotins (TBTC, DBTC, MBTC), no kinetic data were available to

perform a dosimetry-based comparison between *in vitro* and *in vivo* effects. The toxicokinetics (TK) of the group of VPA analogs were described by a generic TK model as reported in detail by Bosgra and Westerhout [31]. In short, the model contained compartments representing stomach, intestinal tract and body, with elimination from the body compartment returning to the GI compartment to accommodate enterohepatic circulation for VPA and EHA. Eliminated MHA was not returned to the intestinal compartment, as the data showed no evidence of enterohepatic circulation for MHA. The model was fitted to the data, and used to calculate the maximum and average plasma concentration over the exposure duration in the study at doses that led to developmental effects for VPA and EHA, but not for MHA [30]. The resulting *in vivo* calculated concentrations were compared with effective concentrations in *in vitro* assays, after correction for protein binding and lipid partitioning between plasma and test medium [31].

## 3. Results

### 3.1. Test battery results

Table 2 shows the detailed results of the individual assays of the *in vitro* test battery (published results for VPA, and MEHP [7] were included for facilitating comparison). Remarkably, the di- and tributyltin-compounds were most potent across all assays.

VPA was considered positive based on the EST, ZET, ReProGlo, and on CALUX assay ERA, ERb. EHA was considered positive based on the EST, ZET, and ReProGlo. MHA was only positive in ZET, though at a higher concentration as compared to VPA and EHA.

MEHP was considered positive based on the EST, ZET, and on CALUX assays PPAR- $\alpha$ , and weak responses in PPAR- $\gamma$  and AR(anti).



**Table 2**

Detailed results of screening 9 chemicals in the ChemScreen test battery: numbers represent the negative logarithm of the critical effective molar concentration.

Assays	EST		ZET		ReProGlo	
Compound	cytotox	diff 010	cytotox	ED50	BMC2	BMC0.5
VPA	2.6	3.6 <sup>a</sup>	<sup>b</sup>	4.1	3.5	
EHA		3.1		4.2	3.4	
MHA	2.9			3.5		
MEHP	2.9	3.7 <sup>a</sup>		3.4		
MBzP	2.6	3.1	3.4		– <sup>c</sup>	–
MMP	>2.6 <sup>d</sup>		3.2		–	–
TBTC	6.9	7.0		7.0		7.2
DBTC	7.0	7.1		5.0	4.4	
MBTC	3.4	3.8		3.5		2.8

Assays	CALUX													
Compound	ERa	ERb	AR(anti)	PR(anti)	GR(anti)	TRb	PPAR-α	PPAR-γ	DR	AP1	ESRE	Nrf2	p21	p53
VPA	4.8	4.1		3.5		3.2	4.0	3.2	3.7		3.5		3.2	3.5
EHA							4.0					3.0		
MHA			3.5				3.8							
MEHP			4.5				5.5	4.7						
MBzP						4.0	4.5	4.1				4.0	4.0	
MMP														
TBTC					8.5		8.5	8.5				7.2		
DBTC					7.5			6.5		6.7	6.8	6.8		
MBTC			4.5									5.0		

Numbers represent cytotox dose or EC50, etc. expressed as negative logarithm of the molar concentration: e.g.  $10^{-4}$  M is shown as '4'. Chemical name abbreviations and critical effective concentration as defined for each assay is detailed in Section 2. All chemicals were tested in all assays (unless indicated otherwise).

<sup>a</sup> Data taken from [7].

<sup>b</sup> Empty fields indicate absence of effect.

<sup>c</sup> '–': Not measured.

<sup>d</sup> '>': not observed at this highest tested concentration.

Note that of the 24 assays performed (see Section 2) only positively responding assays are depicted here; not responding: T47D-ER, ER, AR, PR, GR, RAR, PAH, Hif1a, TCF, and NFkB.

MBzP was considered positive based on the EST, and on a weak response in CALUX assay PPAR-α, MMP was negative in all assays.

Both TBTC and DBTC responded positively in EST, ZET, and ReProGlo assays at relatively low concentrations. At similar concentrations, both compounds triggered a number of CALUX assays: TBTC triggered GR(anti), PPAR-α, PPAR-γ, and Nrf2, while DBTC triggered GR(anti), PPAR-γ, AP1, ESRE, and Nrf2. MBTC was also positive in EST, ZET, and ReProGlo, but at clearly higher concentrations as compared to TBTC and DBTC. The CALUX assays AR(anti) and Nrf2, were triggered also at clearly higher concentrations as compared to TBTC and DBTC.

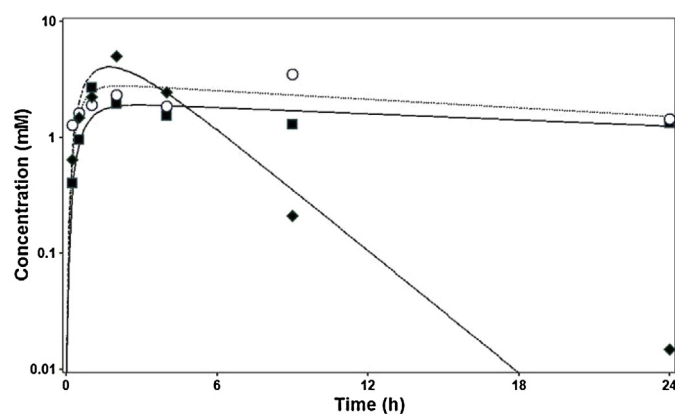
### 3.2. In vitro–in vivo comparison

As indicated, designing and interpreting *in vitro* tests ideally includes toxicokinetic modeling. Here we had that data for VPA and its analogs only. The results of the parameter estimation of the generic valproate analog PK model are described elsewhere [31], and the results are summarized and depicted in Fig. 1 and Table 3.

The estimated maximum and average plasma concentrations were corrected for differences in protein binding and lipid partitioning by a method described earlier [7,35,36], and detailed in [31]. The VPA fraction bound in adult rat plasma of 85% [37] was also assumed for EHA and MHA. Albumin concentrations in fetal plasma were approximately 10% of adult values on GD12, increasing to 40% at term [38]. The fraction bound was assumed to be proportional to the albumin concentration. Table 3 lists the resulting maximum and average total concentrations *in vivo* and equivalent nominal concentrations *in vitro* (i.e., nominal medium concentrations with an equal unbound concentration as *in vivo*). Please, note that at the doses indicated in this table (at which pregnant rats were treated), the embryotoxic effects of VPA were more severe than those of EHA, and no embryotoxicity of MHA was observed [30]. The corresponding maximum concentrations ( $C_{max}$ ) in maternal

plasma of VPA, EHA and MHA were 1.9, 2.8 and 4.0 mM, respectively, while the average concentration between exposure at GD12 and GD21 ( $C_{avg}$ ) was 0.45, 0.48 and 0.086 mM, respectively. Free concentrations in maternal and fetuses plasma were considered in equilibrium, with relatively little variation over time and across compounds [31]. Because of the inherent uncertainty within these calculations responses at concentrations one order of magnitude higher than calculated are considered as relevant as well.

When for the alkanolic acids VPA, EHA, and MHA,  $C_{max}$  concentrations were compared with calculated predicted systemic concentrations following the indicated *in vivo* exposures the conclusions for VPA, EHA, and MHA with regard to EST, ZET, and ReProGlo responses do not change. With regard to the CALUX assays, the TK data suggest to take all other recorded responses in Table 1 as weak positive responses for these chemicals as well.



**Fig. 1.** Measurements [34] and model simulations of maternal plasma concentrations (mM) of VPA (filled squares and solid curve), EHA (open circles and dotted curve) and MHA (filled diamonds and dashed curve) against time (h).

**Table 3**  
Estimated maximum and average plasma concentrations corresponding to observations of developmental toxicity in Scott et al. [31] and equivalent nominal concentrations in *in vitro* assays (depicted as  $-\log^{10}$  values).

	Dose <sup>a</sup> (mg/kg)		Concentration in plasma mM	Equivalent nominal concentration in <i>in vitro</i>			
				EST mM	ZET mM	ReProGlo mM	CALUX mM
VPA	1039	$C_{\max}$	1.9	3.3	4.1	3.4	3.7
		$C_{\text{avg}}$	0.45	3.9	4.7	4.0	4.3
EHA	1803	$C_{\max}$	2.8	3.1	3.8	3.2	3.5
		$C_{\text{avg}}$	0.48	3.9	4.6	4.0	4.3
MHA	1836	$C_{\max}$	4.0	2.9	3.4	3.0	3.2
		$C_{\text{avg}}$	0.086	4.5	5.1	4.6	4.9

<sup>a</sup> Doses VPA and EHA represent LOAEL values, MHA dose represents NOAEL value (as highest tested dose).

However, in case  $C_{\text{avg}}$  concentrations are more representative for inducing *in vivo* effects the observed effect of MHA in ZET is to be interpreted as a negative response. This also holds for the CALUX assay responses, and herewith the conclusion for MHA for the whole battery turns into negative.

In literature TK model data were only available for MEHP [34], and not for the other two phthalates. As we cannot make a comparison on this basis between these phthalates, we have not included TK considerations for MEHP in our conclusions for these substances.

Table 4 summarizes the conclusions based on results derived from Table 2 and, for VPA and its analogs, from Table 3, and it also shows the major *in vivo* reproductive toxicity profiles of the tested chemicals (first column). The conclusion for the battery is ‘(weak) positive’ if one or more assays show a (weak) positive response that is considered relevant for reproductive and developmental toxicity. In general, within the three different structural groups of alkanolic acids, phthalates, and organotin compounds, the test battery correctly identified the developmental toxicants VPA, EHA, MEHP, MBzP, TBTC, and DBTC. As can be seen from Table 4, none of the conclusions leading to a ‘positive’ prediction depended upon just a single assay. At least two of the battery assays contributed to this conclusion, though one of the two may have been weakly positive. For developmental toxicity EST, and ZET performed with comparable concordances of 6 and 5 out of 6, respectively, *i.e.*, correctly distinguishing developmental toxicants. The CALUX assays within the battery are expected to individually detect fewer embryotoxicants based on their more limited applicability domain in terms of

modes of action represented in the assays. Nonetheless, the CALUX PPAR- $\alpha$  assay had a comparable concordance score as the EST and ZET. The battery correctly identified MMP as a non-developmental toxicant, but also falsely identified MHA and MBTC as developmental toxicants, be it of weaker potency as compared to their developmental toxic analogs.

#### 4. Discussion

In a previous paper by Piersma et al. [7] the rationale behind constructing a specific battery of complementary *in vitro* toxicity assays, as well as the detection accuracy of this battery toward a set of compounds representing a variety of reproductive effects and a corresponding variety of underlying modes or mechanisms of action, has been outlined. In that study, a chemically diverse group of reproductive toxicants was used. The aim of the present study, however, was to investigate whether this battery could also be applied in cases of read across to determine the reproductive toxicity of an untested chemical that shares structural features with an already tested chemical. Read across has been put forward as a potential tool to fill data gaps and avoid animal testing, based on the idea that similar chemical structures will very probably induce similar toxicological effects [39]. However, the application of chemical groups supposedly having similar toxicological responses does have its limitations if a structural analog unexpectedly deviates from a presumed shared biological response. For

**Table 4**  
Summarized interpretation of ChemScreen test battery outcome in terms of prediction of reproductive and developmental toxicity.

Compound	Major <i>in vivo</i> toxicity profile <sup>a</sup>	EST diff	Zebrafish	ReProGlo	CALUX panel	Battery Prediction
VPA	Neurodevelopmental toxicant	Differentiation effect	Developmental toxicant	Inducer	Positive in PPAR- $\alpha$ , Era, and ER $\beta$ , weak in many others	Positive
EHA	Neurodevelopmental toxicant	Differentiation effect	Developmental toxicant	Inducer	Positive in PPAR- $\alpha$ and Nrf2	Positive
MHA	Not developmental toxicant	No effect	Developmental toxicant	No effect	Positive in PPAR- $\alpha$ and AR(anti)	Positive/ $C_{\max}$ <sup>c</sup>
MEHP	Male reproductive organ	No effect Differentiation effect	No effect Developmental toxicant	No effect No effect	No effect Positive in PPAR- $\alpha$ and (weak) in PPAR- $\gamma$ and AR(anti)	Negative/ $C_{\text{avg}}$ <sup>c</sup> Positive
MBzP	Male reproductive organ, developmental toxicant	Differentiation effect	No effect	– <sup>b</sup>	Weak positive in PPAR- $\alpha$	Positive
MMP	Not developmental toxicant	No effect	No effect	–	Negative	Negative
TBTC	Developmental (immuno)toxicant	Differentiation effect	Developmental toxicant	Repressor	Positive in PPAR- $\alpha$ , PPAR- $\gamma$ , GR(anti), and Nrf2	Positive
DBTC	Neurodevelopmental toxicant	Differentiation effect	Developmental toxicant	Inducer	Positive in PPAR- $\gamma$ , GR(anti), AP1, ESRE, and Nrf2	Positive
MBTC	Not developmental toxicant	Differentiation effect	Developmental toxicant	No effect	Positive in Nrf2 and AR(anti)	Positive

<sup>a</sup> See Section 4 for references.

<sup>b</sup> Not measured.

<sup>c</sup> Conclusion depends on whether  $C_{\max}$  or  $C_{\text{avg}}$  is taken as valid *in vitro* concentration.

example, VPA, EHA, and MHA could be grouped on the basis of their structural similarity (see Table 1). VPA is a well-known neurodevelopmental toxicant *in vivo* [33,40] and based on structural similarity, one would expect EHA and MHA to be reproductive toxicants as well. However, while EHA causes VPA-like effects on rat development, e.g., reduced pup weight, mortality, extra presacral vertebrae, fused ribs, delayed parturition, and increase in gestation length, MHA appears devoid of all these effects despite observed maternal toxicity [41]. Two explanations have been offered for the lack of embryotoxicity of MHA *in vivo* [30]. Either it is intrinsically less potent than VPA and EHA (despite its structural similarity), or it is more rapidly eliminated resulting in lower fetal exposure. The observed maximum MHA concentration was higher than that of EHA and VPA, but the estimated AUC from exposure to term was 5-fold lower [31]. Contrary to the effects observed *in vivo* our test battery showed MHA as an embryotoxic compound, though less potent as compared to VPA and EHA, confirming earlier observations in ZET by Hermann [32]. The absence of embryotoxic effects in *in vivo* studies, therefore, will be partly due to this lower intrinsic potency, but most likely it is due to its lower predicted fetal concentrations. This also suggests that average concentrations are to be taken as more critical for induction of the observed effects than peak concentrations (see Fig. 1). Therefore, read across based solely on chemical similarity would for MHA lead to wrong conclusions, and biological verification with a relevant test battery, with inclusion of toxicokinetic data will generate relevant additional pieces of information for predicting a chemical's toxicity.

Similarly, the structural group of phthalates is well known for its reproductive toxicity [42,43]. In a previous study, MEHP was identified as a reprotoxic chemical *via* this battery approach [7], and on the basis of structural similarity one might expect MBzP and MMP to also be reprotoxicants (see Table 1). MBzP was positive in EST, and PPAR- $\alpha$  only, while MMP was inactive in all the battery assays. On this basis MBzP may be classified as a reprotoxicant, while MMP appears inactive in this respect. Comparable conclusions for these substances can be derived when evaluating their *in vivo* data: MBzP affects fertility *via* damage to testes, prostate, and seminal vesicles, leading to reduced sperm counts, reduced epididymal sperm motility, sharing this activity with MEHP [42,44]. MBzP also induced developmental effects, such as increased embryo lethality and, at higher dose levels, various malformations, such as cleft palate, fusion of sternebrae, fusion of cervical or thoracic vertebral arches, absence of lumbar vertebral arches and fusion or absence of ribs [42,45,46]. MMP, on the other hand, had no effect on sperm counts and epididymal sperm motility [44,47], and its parent compound, dimethyl phthalate, is an accepted non-embryotoxic phthalate model compound [48,49].

A third group of compounds we have investigated is the organotin chlorides MBTC, DBTC, TBTC, all structurally related to the previously investigated reproductive toxicant DOTC [7]. Organotin chlorides potentially may interfere with sex hormone homeostasis by binding to various enzymes regulating the male–female sex hormone balance *via* the tin moiety of the molecule [50–52]. Exposure to TBTC induced post-implantation loss, cleft palate and other skeletal malformations [55,56], while DBTC consistently showed dose-dependent developmental toxicity, such as fetal death, birth defects and decreased fetal weights [54]. MBTC, however, was concluded to be not developmentally toxic, based on a series of prenatal developmental toxicity studies in rats showing no effects on fetuses or only at maternally toxic dose levels [53]. The outcome of the battery was positive for all three chemicals, with MBTC showing by far the least potent: at least two orders of magnitude less potent than DBTC and TBTC. For MBTC, and DBTC a similar conclusion on relative developmental toxic potency was noted before in the ZET [19]. In the absence of *in vivo* toxicokinetic data it remains speculative to explain the battery's false positive response toward MBTC

as due to too high *in vitro* test concentrations, at which already non-developmental (maternal) effects are expected to dominate *in vivo*.

For the groups of phthalates and organotin compounds, insufficient data were available for *in vitro*–*in vivo* comparison based on TK modeling. The TK data for VPA and analogs show their importance for interpreting of *in vitro* observations. Generic physiology-based TK models based on high throughput *in vitro* kinetic assays and quantitative property–property relationships may be of help in such situations. For instance, Rotroff et al. [57] used a simple one compartment model taking as input only *in vitro* hepatic intrinsic clearance and plasma fraction unbound for reverse dosimetry of a large number of compounds. The performance of this and more complex models to predict TK is currently under investigation [58]. It will certainly be worthwhile to add such high throughput TK assays to future toxicity test batteries to facilitate *in vitro*–*in vivo* extrapolation.

The above examples show that read across of developmental toxicity data based solely on structural similarity arguments cannot be fully trusted. To verify whether an untested chemical that shares structural features with an already tested chemical does in fact also show similar toxicological responses, as is intrinsically assumed, some biological model or battery relevant for the endpoint interest should be applied to both query and tested source chemicals. In this way those parts of the molecules actually responsible for the biological response observed *in vivo*, and thus the most relevant part to assess structural similarity, can be identified. The *in vitro* battery developed within ChemScreen was successful in identifying analogs that have either much lower potency or would not be classified at all for developmental toxicity. Obviously, the battery also works the other way around: for example, if MHA would have been the source chemical, and its *in vivo* data were proposed to be read across to EHA or VPA (based on their structural similarity), the battery would have identified EHA and VPA as developmental toxicants, despite the absence of this property for MHA.

It is clear that more examples like this need to be generated in order to fully trust this battery approach within a grouping and read across concept as a validated alternative to *in vivo* studies. Also, besides this qualitative aspect of whether or not a substance has reproductive toxicological properties, the quantitative aspect of the assessment, i.e., what are the corresponding relative potencies, needs to be addressed. But this battery approach clearly is a promising concept for replacing animal experiments for untested chemicals that have (a) tested structural analog(s). A similar response in a battery like the one presented here for structurally related chemicals clearly reinforces their application for grouping and read across within a regulatory context like REACH. In this way, the battery may already contribute to reducing *in vivo* studies, despite the fact that its use as a stand alone (i.e., in the absence of tested structural analogs), is not expected to be accepted before 2020 [3,4]. Apart from this, this test battery may already be applied as a filter for selecting potential reproductive toxicants for *in vivo* testing, or for prioritizing chemicals for further investigation (safe design) and/or by (de)selecting candidate (pharmaceutical) compounds for further development [17]. Also, the biological verification of grouping by this battery approach may help in identifying structural elements that are critical to the observed effect. This may facilitate the design of less toxic molecules, identify potential biological targets *in vivo*, and consequently help build adverse outcome pathways.

## Conflict of interest

The authors declare that there are no conflicts of interest.

## Transparency document

The [Transparency document](#) associated with this article can be found in the online version.

## Acknowledgement

This work was partially financed by the Commission of the European Communities, the collaborative project ChemScreen (GA244236).

## References

- [1] European Commission (EC). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/105/EC and 2000/21/EC. Off J Eur Union L 2007;136:3–280.
- [2] Costanza R, Hartung T. Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals – a report by the transatlantic think tank for toxicology. ALTEX 2009;26:187–208.
- [3] Adler S, Basketter D, Creton S, Pelkonen O, Van Benthem J, Zuang V, et al. Alternative (non-animal) methods for cosmetics testing: current status and future prospects – 2010. Arch Toxicol 2011;85:367–485 [Reproductive toxicity: 447–463].
- [4] Leist M, Hasigawa N, Novida C, Daneshmand M, Basketter D, Kiber I, et al. Consensus report on the future of animal-free systemic toxicity testing. ALTEX 2014;31:341–56 [Reproductive toxicity: 350–351].
- [5] Schenk B, Weimer M, Bremer S, van der Burg B, Cortvriendt R, Freyberger A, et al. The ReProTect Feasibility Study, a novel comprehensive in vitro approach to detect reproductive toxicants. Reprod Toxicol 2010;30:200–18.
- [6] Sipes NS, Martin MT, Reif DM, Kleinstreuer NC, Judson RS, Singh AV, et al. Predictive models of prenatal developmental toxicity from ToxCast high throughput screening data. Toxicol Sci 2011;124(November (1)):109–27.
- [7] Piersma AH, Bosgra S, van Duursen MBM, Hermesen SAB, Jonker LRA, Kroese ED, et al. Evaluation of an alternative in vitro test battery for detecting reproductive toxicants. Reprod Toxicol 2013;38:53–64.
- [8] Hermesen SA, van den Brandhof EJ, van der Ven LT, Piersma AH. Relative embryotoxicity of two classes of chemicals in a modified zebrafish embryotoxicity test and comparison with their in vivo potencies. Toxicol In Vitro 2011;25(April (3)):745–53.
- [9] Scholz G, Pohl I, Genschow E, Klemm M, Spielmann H. Embryotoxicity screening using embryonic stem cells in vitro: correlation to in vivo teratogenicity. Cells Tissues Organs 1999;165(3–4):203–11.
- [10] Uibel F, Mühleisen A, Köhle C, Weimer M, Stummann TC, Bremer S, et al. ReProGlo: a new stem cell-based reporter assay aimed to predict embryotoxic potential of drugs and chemicals. Reprod Toxicol 2010;30(August (1)):103–12.
- [11] van Duursen MB, Nijmeijer SM, Ruchirawat S, van den Berg M. Chemo-preventive actions by enterolactone and 13 VIOXX-related lactone derivatives in H295R human adrenocortical carcinoma cells. Toxicol Lett 2010;192(February (3)):271–7.
- [12] Hecker M, Hollert H, Cooper R, Vinggaard AM, Akahori Y, Murphy M, et al. The OECD validation program of the H295R steroidogenesis assay: phase 3 final inter-laboratory validation study. Environ Sci Pollut Res Int 2011;18(March (3)):503–15.
- [13] Sonneveld E, Jansen HJ, Riteco JAC, Brouwer A, Van der Burg B. Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid responsive bioassays. Toxicol Sci 2005;83:136–48.
- [14] Van der Burg B, Van der Linden SC, Man HY, Winter R, Jonker L, Lussenburg B, et al. A panel of quantitative CALUX® reporter gene assays for reliable high throughput toxicity screening of chemicals and complex mixtures. In: Steinberg P, editor. High throughput screening methods in toxicity testing. New York: John Wiley and Sons, Inc.; 2013. p. 519–32.
- [15] Daston GP, Chapin RE, Scialli AR, Piersma AH, Carney EW, Rogers JM, et al. A different approach to validating screening assays for developmental toxicity. Birth Defects Res 2010;89:526–30.
- [16] Wu S, Fisher J, Nacliff J, Laufersweiler M, Lester C, Daston G, et al. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem Res Toxicol 2013;26:1840–61.
- [17] Van der Burg B, Kroese ED, Piersma AH. Towards a pragmatic alternative testing strategy for the detection of reproductive toxicants. Reprod Toxicol 2011;31(May (4)):558–61.
- [18] van Dartel DA, Zeijen NJ, de la Fonteyne LJ, van Schooten FJ, Piersma AH. Disentangling cellular proliferation and differentiation in the embryonic stem cell test, and its impact on the experimental protocol. Reprod Toxicol 2009;28(September (2)):254–61.
- [19] Beker van Woudenberg A, Wolterbeek A, Ter Brake L, Snel C, Menke A, Rubingh C, et al. A category approach to predicting the developmental (neuro)toxicity of organotin compounds: the value of the zebrafish (*Danio rerio*) embryotoxicity test ZET. Reprod Toxicol 2013;41:35–44.
- [20] OECD. Guideline for the testing of chemicals. Draft proposal for a new guideline Fish Embryo Toxicity (FET) test; 2006.
- [21] Legler J, Van den Brink CE, Brouwer A, Murk AJ, Van der Saag PT, Vethaak AD, et al. Development of a stably transfected estrogen receptor-mediated luciferase reporter gene assay in the human T47-D breast cancer cell line. Toxicol Sci 1999;48:55–66.
- [22] Garrison PM, Tullis K, Aarts JMMJG, Brouwer A, Giesy JP, Denison MS. Fund. Appl Toxicol 1996;30:194–203.
- [23] Gauggel S, Derreza-Greven C, Wimmer J, Wingfield M, Van der Burg B, Dietrich DR. Characterization of biologically available wood combustion particles in cell culture medium. ALTEX 2012;29(2):183–200.
- [24] Pieterse B, Felzel E, Winter R, van der Burg B, Brouwer A. PAH-CALUX, an optimized bioassay for carcinogenic hazard identification of polycyclic aromatic hydrocarbons (PAHs) as individual compounds and in complex mixtures. Environ Sci Technol 2013;47(20):11651–9.
- [25] Gijssbers L, van Eekelen H, de Haan LH, Swier JM, Heijink NL, Kloet SK, et al. Induction of peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ )-mediated gene expression by tomato (*Solanum lycopersicum* L.) extracts. J Agric Food Chem 2013;61:3419–27.
- [26] Gijssbers L, Man HY, Van Eekelen H, De Haan L, Keijer J, Aarts JM, et al. Induction of peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ )-mediated gene expression by extracts of tomato. J Agric Food Chem 2013;61(14):3419–27.
- [27] Van Vugt-Lussenburg B, Besselink HT, van der Burg B, Brouwer A. Dr-Calux®: a high-throughput screening assay for the detection of dioxin and dioxin-like compounds in food and feed. In: Steinberg P, editor. High-throughput screening methods in toxicity testing. New Jersey, US: John Wiley & Sons Inc.; 2013. ISBN 978-111806563-1 p. 533–46 [chapter 29].
- [28] Van der Linden SC, von Bergh ARM, Van Vugt-Lussenburg B, Jonker L, Brouwer A, Teunis M, et al. Development of a panel of high throughput reporter gene assays to detect genotoxicity and oxidative stress. Mutat Res Genet Toxicol Environ Mutagen 2013;760:23–32.
- [29] OECD. Test guideline No.455. Stably transfected human estrogen receptor-transcriptional activation assay for detection of estrogenic agonist-activity of chemicals. Paris: Organisation for Economic Cooperation and Development; 2009.
- [30] Scott WJ, Collins MD, Nau H. Pharmacokinetic determinants of embryotoxicity in rats associated with organic acids. Environ Health Perspect 1994;102:97–101.
- [31] Bosgra S, Westerhout J. Interpreting in vitro developmental toxicity test battery results: the consideration of toxicokinetics. Reprod Toxicol 2015;55:73–80.
- [32] Hermann K. Effects of the anticonvulsant drug valproic acid and related substances on the early development of the zebrafish (*Brachydanio rerio*). Toxicol In Vitro 1993;7:41–54.
- [33] Robert E. Valproic acid and spina bifida. Lancet 1982;2:1096.
- [34] Kavlock R, Boekelheide K, Chapin R, Cunningham M, Faustman E, Foster P, et al. NTP Center for the Evaluation of Risks to Human Reproduction: phthalates expert panel report on the reproduction and developmental toxicity of di(2-ethylhexyl) phthalate. Reprod Toxicol 2002;16:529–653.
- [35] Krug AK, Kolde R, Gaspar JA, Rempel E, Balmer NV, Meganathan K, et al. Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. Arch Toxicol 2013;87(1):123–43.
- [36] Zimmer B, Pallocca G, Dreser N, Foerster S, Waldmann T, Westerhout J, et al. Profiling of drugs and environmental chemicals for functional impairment of neural crest migration in a novel stem cell-based test battery. Arch Toxicol 2014;88(5):1109–26.
- [37] Mandula H, Parepally JM, Feng R, Smith QR. Role of site-specific binding to plasma albumin in drug availability to brain. J Pharmacol Exp Ther 2006;317(2):667–75.
- [38] Semmes RL, Shen DD. Nonlinear binding of valproic acid (VPA) and E-delta 2-valproic acid to rat plasma proteins? Pharm Res 1990;7(5):461–7.
- [39] OECD Environment Health and Safety Publications, Series on Testing and Assessment No. 80, Guidance on grouping of chemicals; 2007.
- [40] Binkerd PE, Rowland JM, Nau H, Hendrickx AG. Evaluation of valproic acid (VPA) developmental toxicity and pharmacokinetics in Sprague–Dawley rats. Fundam Appl Toxicol 1988;11:485–93.
- [41] Narotsky MG, Kavlock EZ. Developmental toxicity and structure–activity relationships of aliphatic acids, including dose–response assessment of valproic acid in mice and rats. Fundam Appl Toxicol 1994;22:251–65.
- [42] Fabian E, Hulzebos E, Mennes W, Piersma AH. A category approach for reproductive effects of phthalates. Crit Rev Toxicol 2006;36:695–726.
- [43] Lyche JL, Gutleb AC, Bergman A, Eriksen GS, Murk AJ, Ropstad E, et al. Reproductive and developmental toxicity of phthalates. J Toxicol Environ Health B Crit Rev 2009;12:225–49.
- [44] Kwack SJ, Kim KB, Kim HS, Lee BM. Comparative toxicological evaluation of phthalate diesters and metabolites in Sprague–Dawley male rats for risk assessment. J Toxicol Environ Health 2009;72:1446–54.
- [45] Ema M, Miyawaki E. Effects on development of the reproductive system in male offspring of rats given butyl benzyl phthalate during late pregnancy. Reprod Toxicol 2002;16:71–6.
- [46] Saillenfait AM, Sabate JP, Gallissot F. Comparative embryo-toxicities of butyl benzyl phthalate, mono-n-butyl phthalate and mono-benzyl phthalate in mice and rats: in vivo and in vitro observations. Reprod Toxicol 2003;17:575–83.



- [47] Gray LE, Ostby J, Furr J, Pricwe M, Veeramachaneni DNR, Parks L. Exposure to the phthalates DEHP, BBP, and DINP, but not DEP, DMP, or DOTP, alters sexual differentiation of the male rat. *Toxicol Sci* 2000;58: 350–65.
- [48] Schwetz BA, Harris MW. Developmental toxicology: status of the field and contribution of the National Toxicology Program. *Environ Health Perspect* 1993;100:269–82.
- [49] Genschow E, Spielmann H, Scholz G, Seiler A, Brown N, Piersma A, et al. The ECVAM international validation study on *in vitro* embryotoxicity tests: results of the definitive phase and evaluation of prediction models. *ATLA* 2002;30:151–76.
- [50] Heidrich DD, Steckelbroeck S, Klingmuller D. Inhibition of human cytochrome P450 aromatase activity by butyltins. *Steroids* 2001;66(10):763–9.
- [51] Cooke GM. Effect of organotins on human aromatase activity *in vitro*. *Toxicol Lett* 2002;126(2):121–30.
- [52] Lo S, Alléra A, Albers P, Heimbrecht J, Jantzen E, Klingmüller D, et al. Dithioerythritol (DTE) prevents inhibitory effects of triphenyltin (TPT) on the key enzymes of the human sex steroid hormone metabolism. *J Steroid Biochem Mol Biol* 2003;84(5):569–76.
- [53] World Health Organization (WHO). Mono- and disubstituted methyltin, butyltin, and octyltin compounds. In: Concise international chemical assessment document; 2006. p. 73.
- [54] Ema M, Arima A, Fukunishi K, Matsumoto M, Hirata-Koizumi M, Hirose A, et al. Developmental toxicity of dibutyltin dichloride given on three consecutive days during organogenesis in cynomolgus monkeys. *Drug Chem Toxicol* 2009;32:150–7.
- [55] Noda T, Morita S, Yamano T, Shimizu M, Nakamura T, Saitoh M, et al. Teratogenicity study of tri-*n*-butyltin acetate in rats by oral administration. *Toxicol Lett* 1991;55:109–15.
- [56] Ema M, Harazono A, Miyawaki E, Ogawa Y. Effect of the day of administration on the developmental toxicity of tributyltin chloride in rats. *Arch Environ Contam Toxicol* 1997;33:90–6.
- [57] Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, et al. Incorporating human dosimetry and exposure into high-throughput *in vitro* toxicity screening. *Toxicol Sci* 2010;117(2):348–58.
- [58] Wambaugh J, Strobe C, Nicolas C, Pearce R, Judson R, Goldsmith R. High throughput pharmacokinetics for environmental chemicals. In: Poster presentation, society of toxicology's 53rd annual meeting. 2014 [in preparation].