TOXICOLOGICAL SCIENCES **120(2)**, 360–370 (2011)
doi:10.1093/toxsci/kfr001
Advance Access publication January 11, 2011

The Embryonic Stem Cell Test as Tool to Assess Structure-Dependent Teratogenicity: The Case of Valproic Acid

Christian Riebeling,* Ralph Pirow,* Klaus Becker,† Roland Buesen,*,[1] Daniel Eikel,‡,[2] Johanna Kaltenhäuser,*,†,[3] Frauke Meyer,§ Heinz Nau,‡ Birgitta Slawik,* Anke Visan,* Jutta Volland,§ Horst Spielmann,*,[4] Andreas Luch,* and Andrea Seiler*,[5]

*German Federal Institute for Risk Assessment (BfR), Center for Alternative Methods to Animal Experiments—ZEBET, 12277 Berlin, Germany; †Bayer Schering Pharma AG, Nonclinical Drug Safety, Genetic Toxicology, 13342 Berlin, Germany; ‡Institute of Food Toxicology and Chemical Analysis, University of Veterinary Medicine Hanover, 30173 Hanover, Germany; and §Nycomed GmbH, Institute for Pharmacology and Preclinical Drug Safety (IPAS), Hamburg, 22885 Barsbüttel, Germany

[1]Present address: BASF SE, 67056 Ludwigshafen, Germany.
[2]Present address: Advion BioSystems, Product Application Laboratory, Ithaca, NY 14850.
[3]Present address: German Federal Institute for Risk Assessment, Center for Alternative Methods to Animal Experiments—ZEBET, 12277 Berlin, Germany.
[4]Present address: Department of Biology, Chemistry, and Pharmacy, The Free University of Berlin, 14195 Berlin, Germany.
[5]To whom correspondence should be addressed at German Federal Institute for Risk Assessment, Center for Alternative Methods to Animal Experiments—ZEBET, Diedersdorfer Weg 1, 12277 Berlin, Germany. Fax: +49-30-8412-2958. E-mail: Andrea.Seiler@bfr.bund.de.

Received October 25, 2010; accepted January 3, 2011

Teratogenicity can be predicted *in vitro* using the embryonic stem cell test (EST). The EST, which is based on the morphometric measurement of cardiomyocyte differentiation and cytotoxicity parameters, represents a scientifically validated method for the detection and classification of chemicals according to their teratogenic potency. Furthermore, an abbreviated protocol applying flow cytometry of intracellular marker proteins to determine differentiation into the cardiomyocyte lineage is available. Although valproic acid (VPA) is in worldwide clinical use as antiepileptic drug, it exhibits two severe side effects, i.e., teratogenicity and hepatotoxicity. These limitations have led to extensive research into derivatives of VPA. Here we chose VPA as model compound to test the applicability domain and to further evaluate the reliability of the EST. To this end, we study six closely related congeners of VPA and demonstrate that both the standard and the molecular flow cytometry-based EST are well suited to indicate differences in the teratogenic potency among VPA analogs that differ only in chirality or side chain length. Our data show that identical results can be obtained by using the standard EST or a shortened protocol based on flow cytometry of intracellular marker proteins. Both *in vitro* protocols enable to reliably determine differentiation of murine stem cells toward the cardiomyocyte lineage and to assess its chemical-mediated inhibition.

Key Words: **valproic acid; congeners; embryonic stem cell test; structure-activity relationships; embryotoxicity; teratogenicity.**

Testing for reproductive and developmental toxicity of drugs and other chemical compounds *in vitro* is an attractive alternative procedure to time-consuming and expensive *in vivo* or *ex vivo* experiments (Spielmann, 2009). The embryonic stem cell test (EST) employs the blastocyst-derived embryonic stem cell line D3 from mice that spontaneously differentiates into contracting cardiomyocytes in cell aggregates, termed embryoid bodies (EBs). This morphological feature is used as an endpoint for differentiation together with the measurement of cytotoxicity in D3 cells and 3T3 fibroblasts in the standard EST (Seiler *et al.*, 2006; Spielmann *et al.*, 1997). In an international validation study supported by the European Center for the Validation of Alternative Methods (ECVAM), this method proved reliable for the prediction of embryotoxicity *in vivo* (Genschow *et al.*, 2004). For a set of 20 reference compounds with different embryotoxic potencies *in vivo* of nonembryotoxic, weakly embryotoxic, and strongly embryotoxic, the EST provided correct classification in 78% of all cases. Remarkably, a predictivity of 100% was obtained for strong embryotoxicants. However, in a second evaluation study set up to enlarge the database, only 2 out of 13 previously untested compounds were classified correctly, thus highlighting the limitations of the prediction model (PM, Marx-Stoelting *et al.*, 2009). A detailed review of these data is currently in progress at ZEBET. Preliminary analysis suggests that some of the misclassification might be because of different classification rules in the validation and postvalidation study applied by the experts to assign the chemicals to the embryotoxicity classes according to their *in vivo* potencies.

After validation of the EST, many research groups were interested in refining the EST protocol toward a more rapid and molecular-based assay to allow for quantification of drug effects on the cellular level as well as for higher throughput analysis. Recently, we reported on an abbreviated, molecular-based EST

protocol in which highly predictive markers of cellular differentiation were selected to substitute for visual evaluation of contracting cell areas as a toxicological endpoint. In this improved EST protocol (referred to as molecular FACS-EST), both viability and the expression of the sarcomeric myosin heavy chain (αMHC) protein are quantified in differentiating cells while being continuously exposed to the test chemical (Buesen *et al*., 2009; Seiler *et al*., 2004). Intracellular flow cytometry of αMHC expression and concurrent assessment of cell viability is scheduled at day 7 of differentiation.

In drug development, late-stage attrition of newly developed drug candidates partly because of unexpected safety reasons is a critical issue (Kola and Landis, 2004). Methods that can be implemented in early-stage drug research and development to assess safety issues such as developmental toxicity would significantly decrease the financial burden of discovering and introducing new drugs to the market. In recent years, the EST has been evaluated by the pharmaceutical industry for testing during research and development and considerable efforts have been made to improve the EST for this purpose (Augustine-Rauch *et al*., 2010; Paquette *et al*., 2008; Whitlow *et al*., 2007). To develop more efficient or less adverse derivatives of a lead compound, small modifications such as stereochemical isomerization or side chain variation are introduced to the structure of the parent compound. Ideally, the resulting substance library would be tested for possible adverse effects in addition to its intended pharmacological action to identify the most promising candidates for further development.

To evaluate the applicability of the EST to assess such structure-activity relationships (SAR), we chose valproic acid (VPA) as a model compound because it is a teratogenic drug with a large number of tested congeners (Eikel *et al*., 2006b; Isoherranen *et al*., 2003). VPA is one of the most frequently used antiepileptic drugs worldwide. Additionally, it has recently been approved for migraine prophylaxis (Calabresi *et al*., 2007) and is used in the treatment of bipolar disorders (Bowden, 2009). Moreover, it is being investigated for treatment of cancer (Batty *et al*., 2009), Alzheimer's disease (Tariot *et al*., 2002), and latent human immunodeficiency virus infections (Lehrman *et al*., 2005). Two severe side effects of VPA, hepatotoxicity (Lheureux and Hantson, 2009), and teratogenicity (Ornoy, 2009) have led to research into derivatives of VPA. The embryotoxic effect of VPA has been ascribed to malformation of the heart and defects in neurulation (Ornoy, 2009). In addition, craniofacial and skeletal anomalies and behavioral deficits have been reported. VPA has been shown to inhibit histone deacetylases *in vitro* (Phiel *et al*., 2001). Histone deacetylases are crucial enzymes involved in chromatin remodeling, and this may explain the multiple effects of VPA on embryonic development and why it may be a promising anticancer drug.

We chose six closely related analogs of VPA whose teratogenic potency has been determined *in vivo* in the Naval Medical Research Institute (NMRI)-exencephaly-mouse model (Fig. 1) (Eikel *et al*., 2006b). For this *in vivo* assay, the compounds are injected into pregnant NMRI mice, and after 10 days, the animals are sacrificed, the uteri removed, and the number of implants and resorptions are counted. Individual fetuses are weighed and inspected for external malformations, and the exencephaly rate is calculated relative to the number of live fetuses (Eikel *et al*., 2006b).

In a collaborative study supported by the German Ministry for Education and Research (BMBF), with two partners from the pharmaceutical industry, Bayer Schering Pharma AG and Nycomed GmbH, we evaluate the applicability domain of both the EST and the FACS-EST using these six closely related compounds (Fig. 1). These VPA congeners exhibit varying teratogenic potencies resulting from introduction of minute changes at the chiral center, change of position of a methyl group, addition of an ethyl group, and introduction of a triple carbon bond (acetylenic moiety) (Fig. 1) (Eikel *et al*., 2006b). Here, we demonstrate the potency of the EST and FACS-EST to distinguish among these similar structures.

MATERIALS AND METHODS

VPA derivatives. 2-Ethyl-4-methylpentanoic acid (2-EMP, 144.2 g/mol), 2-isobutyl-4-pentynoic acid (2-IP, 154.2 g/mol), (*R*)-2-propyl-4-pentynoic acid (RProP, 140.2 g/mol), (*S*)-2-propyl-4-pentynoic acid (SProP, 140.2 g/mol), (*R*)-

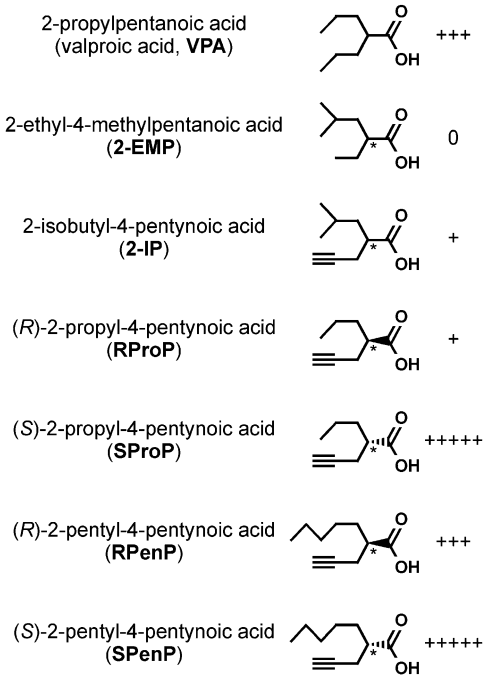

FIG. 1 Chemical structures of the VPA derivatives investigated in this study and their teratogenic potencies *in vivo* determined in the NMRI-exencephaly-mouse model. Teratogenicity rating (0, +, +++, and +++++) according to Eikel *et al*. (2006b). 2-EMP and 2-IP were applied as mixtures of both stereoisomeric forms (racemates), whereas all alkyne derivatives used were enantiomerically pure compounds; ''*'' Denotes a chiral center.

2-pentyl-4-pentynoic acid (RPenP, 168.2 g/mol), and (*S*)-2-pentyl-4-pentynoic acid (SPenP, 168.2 g/mol) (Fig. 1) were synthesized as described in detail elsewhere (Bojic *et al*., 1996 and references therein). Their teratogenic potencies *in vivo* have been determined in the NMRI-exencephaly-mouse model and rated as described (Eikel *et al*., 2006b; Hauck and Nau, 1992). VPA (144.2 g/mol) was purchased from Sigma (Taufkirchen, Germany). All chemicals were dissolved in dimethyl sulfoxide (Sigma). Final solvent concentrations applied showed no undesired background effect in differentiation and viability assays (Seiler *et al*., 2006).

Cell culture. The undifferentiated murine embryonic stem cell line D3 (Doetschman *et al*., 1985) was cultured in high glucose (4.5 g/l) Dulbecco's modified Eagle's medium (DMEM; Gibco Invitrogen, Karlsruhe, Germany) containing 15% fetal calf serum (different sources and batches of embryonic stem cell-qualified fetal bovine serum were screened for suitability for the EST), 2mM glutamine, 50 U/ml penicillin, 50 µg/ml streptomycin, 1% nonessential amino acids (Gibco Invitrogen), 0.1mM β-mercaptoethanol (Sigma), and 1000 U/ml murine leukemia inhibitory factor (mLIF; Chemicon, Hofheim, Germany). Only cells in passages 6–24 after thawing were used in the experiments. BALB/c 3T3 fibroblasts (clone A31, ICN Flow, Eschwege, Germany) were grown in DMEM supplemented with 10% fetal calf serum, 4mM glutamine, 50 U/ml penicillin, and 50 µg/ml streptomycin. Cells were maintained in 5% CO_2 and 95% humidity at 37°C and were routinely passaged every 2–3 days.

Differentiation assay. Differentiation assays were performed to detect compound-induced adverse effects on the differentiation of D3 cells into contracting cardiomyocytes. The method has been described in detail for the validated differentiation assay (Seiler *et al*., 2006; Spielmann *et al*., 1997). The modified protocol using intracellular flow cytometry of selected marker proteins as molecular endpoints of differentiation is described elsewhere (Buesen *et al*., 2009; Seiler *et al*., 2006) and in brief below under "Flow cytometry" section.

Briefly, 750 D3 cells in 20 µl of supplemented DMEM without mLIF (to allow differentiation) were placed as hanging drops on the lid of a dish filled with PBS (Sigma) and incubated for 3 days at 37°C, 5% CO_2, and 95% humidity in the presence of test chemicals at various concentrations. During this period, the cells form aggregates referred to as EBs. Subsequently, the EBs were transferred to uncoated dishes (Greiner, Frickenhausen, Germany) to avoid attachment, containing the appropriate concentration of the test chemical for another 2 days. On day 5, EBs were plated separately into the wells of a 24-multiwell tissue culture plate (BD Falcon, Erembodegem, Belgium), containing the appropriate concentration of test chemical, to allow adherence and outgrowth of the EBs and development of spontaneously beating cardiac muscle cells. On day 10 of the assay, differentiation into contracting myocardial cells is determined using direct phase contrast microscopy. Each well of the 24-well plate is inspected, and the number of wells containing any spontaneously contracting cells is recorded. Individual experiments were considered valid when solvent control plates exhibited beating cardiomyocytes in at least 21 out of 24 wells. From the concentration-response relationship, the half maximum inhibitory concentration of differentiation of D3 cells (ID_{50}(D3)) value was derived as the concentration of 50% reduction in the number of differentiated EBs compared with solvent controls (see "Data analysis and statistics" section below).

Flow cytometry. Cell dissociation and intracellular staining for flow cytometry analyses were carried out as previously described (Buesen *et al*., 2009; Seiler *et al*., 2004, 2006). Briefly, EB formation was induced as described above, and EBs were transferred into uncoated dishes after 3 days. On day 7 of differentiation, EBs were dissociated using trypsin/EDTA. The cells were fixed with 2% paraformaldehyde in PBS (pH 7.4), permeabilized with 0.15% saponin in PBS, and treated with 10% normal goat serum to block nonspecific antigens. For immunostaining, the cells were incubated for 1 h on ice with primary monoclonal antibodies against sarcomeric αMHC (clone MF20 concentrate, Hybridoma Bank, University of Iowa, see also Bader *et al*., 1982) diluted 1:1600 in blocking solution (PBS/10% goat serum/1% bovine serum albumin/0.15% saponin). After washing, cells were incubated for 30 min on ice with biotin-conjugated goat anti-mouse immunoglobulin G (1:1000, Dianova, Hamburg, Germany) followed by incubation on ice with *R*-phycoerythrin–conjugated streptavidin (1:600, Dianova) for 15 min. The cells were analyzed in a flow cytometer at 585 nm. For each experiment, 10^4 viable cells were analyzed; nonviable cells were excluded by appropriate gating. On day 7 of differentiation, 10–20% of the untreated viable cell populations were αMHC positive. Untreated cells and cells lacking primary antibody were used as controls. All data analyses were carried out using CellQuest software (Becton Dickinson, Erembodegem, Belgium). From the concentration-response relationships, the ID_{50}(D3) value was derived as the concentration of 50% reduction in the number of αMHC-positive cells compared with solvent controls (see below).

Determination of cytotoxicity. Cytotoxic effects on D3 cells and 3T3 fibroblasts were determined using the 3-(4,5-di-methylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) assay (Mosmann, 1983) either at experimental day 10 of the validated EST or at day 7 of the FACS-EST, as described previously (Seiler *et al*., 2004, 2006). Briefly, 500 cells were seeded into each well of a 96-well microtiter plate (BD Falcon) and grown in supplemented DMEM without mLIF in the presence of the test compound. A negative control with solvent diluted in medium was included. After 7 and 10 days of culture and two changes of medium (containing the appropriate concentration of test chemical) at days 3 and 5, MTT (Sigma) was added followed by incubation for 2 h. The conversion into a dark blue formazan product was quantified spectrophotometrically by measuring the absorbance at a wavelength of 570 nm and subtraction of the absorbance at 630 nm using a microplate ELISA reader. From the concentration-response relationships, the half maximum inhibitory concentration of growth of D3 cells (IC_{50}(D3)) and 3T3 cells (IC_{50}(3T3)) values were derived as the concentration of 50% reduction in the absorbance compared with solvent controls (see below).

Data analysis and statistics. Endpoints for differentiation and cytotoxicity were determined for each compound in the three different laboratories with at least two independent runs each. The relationship between the substance concentration x (millimolar) and the blank-subtracted response y was analyzed in terms of a three-parameter logistic function

$$y(x) = \frac{d}{1 + \exp[b(\ln x - e)]},$$

where d represents the upper asymptote and e is the natural log-transformed concentration producing a response halfway between the upper asymptote and zero. The parameter b denotes the relative slope around e. The concentration x_{50} (millimolar) at which y decreased to 50% of the mean response under control condition (y_{control}) was derived from the model parameters according to

$$x_{50} = \exp(e)\left(\frac{d - 0.5y_{\text{control}}}{0.5y_{\text{control}}}\right)^{\frac{1}{b}}.$$

Depending on the experimental context, the x_{50} represented the concentration for a 50% reduction in the number of differentiated EBs [ID_{50}(D3)], in αMHC expression [ID_{50}(D3)], or of the viability of D3 cells [IC_{50}(D3)] and 3T3 cells [IC_{50}(3T3)]. For each endpoint, a two-way ANOVA was performed using the $\log_{10}$-transformed x_{50} as response variable and the "compound" and "laboratory" as factors. An interaction term "compound × laboratory" was additionally included into the model. The fitted parameter values were used to derive estimates and 95% confidence intervals for each compound. Multiple comparisons between the compound-specific half maximum effective concentration (ID_{50} or IC_{50}) estimates ($\log_{10}$ scale) were performed by Tukey's method for all pairwise differences in combination with robust procedure that accounted for heterogeneous variances and unbalanced group sizes (Herberich *et al*., 2010). Differences between means were regarded as significant at a value of $p < 0.05$.

The curve fitting and subsequent statistical analyses were performed in the R statistical computing environment (R Development Core Team, 2009) in

combination with the drc package (Ritz and Streibig, 2005). Multiple comparisons were performed by using the multcomp package (Hothorn *et al*., 2008).

The effect concentrations for the three endpoints were entered into the PM of the EST (Seiler *et al*., 2006) after appropriate conversion from molar to mass concentration. The PM was originally expressed in terms of Fisher's classification functions (functions I, II, and III; Scholz *et al*., 1999), which permit the assignment to one of the three toxicity classes. In order to obtain a gradual ranking of the VPA congeners within the three predicted toxicity classes and also to associate this ranking with the substances of the prevalidation and validation study (Genschow *et al*., 2004; Scholz *et al*., 1999), we have instead chosen to express the PM in terms of the two canonical discriminant functions (Green, 1979), LD1 and LD2,

$$\mathrm{LD1} = -0.879\log_{10}\mathrm{IC}_{50}^{3T3} - 0.819\log_{10}\mathrm{IC}_{50}^{D3} + 0.967\frac{\mathrm{IC}_{50}^{3T3}-\mathrm{ID}_{50}}{\mathrm{IC}_{50}^{3T3}} + 2.31$$

$$\mathrm{LD2} = 0.265\log_{10}\mathrm{IC}_{50}^{3T3} - 0.595\log_{10}\mathrm{IC}_{50}^{D3} - 0.947\frac{\mathrm{IC}_{50}^{3T3}-\mathrm{ID}_{50}}{\mathrm{IC}_{50}^{3T3}} + 0.264.$$

The classification and gradual ranking of the VPA congeners are subsequently visualized in a so-called canonical plot. The coefficients of the two canonical discriminant functions were calculated in the R statistical computing environment (R Development Core Team, 2009) by using the lda function of the MASS package (Venables and Ripley, 2002) and the published EST data of the prevalidation set (Genschow *et al*., 2004). The decision boundaries in the canonical plot were calculated as described (Venables and Ripley, 2002).

Study design. This collaborative study was performed in three independent laboratories, ZEBET, Bayer Schering Pharma AG, and Nycomed GmbH. Initially, the interlaboratory variance was evaluated using three substances, 5-fluorouracil, 6-aminonicotinamide, and VPA, which had been previously tested in the validation study (Genschow *et al*., 2004). No significant difference between the laboratories was found for any given endpoint.

Similarly, for the subsequent testing of the VPA congeners, the overall interlaboratory variance (σ) of the $\log_{10}$-transformed effect concentration was $\sigma < 0.18$ for the morphological determination of differentiation and $\sigma < 0.11$ for cytotoxicity measurements of all substances. The intralaboratory variance was higher with $\sigma < 0.25$ for the differentiation assay, whereas the variance in cytotoxicity was $\sigma < 0.16$.

RESULTS

Structure and In Vivo *Potency of VPA Derivatives*

The structures of the compounds tested are given in Figure 1. VPA, a branched short-chain carboxylic acid with two propyl chains at the α-carbon, is the lead compound and was graded as exhibiting medium teratogenic potency (+++) in the NMRI-exencephaly-mouse model (Eikel *et al*., 2006b). *In vitro*, it was correctly predicted as weakly embryotoxic by the EST in the validation study (Genschow *et al*., 2002, 2004), in line with its *in vivo* embryotoxicity (Brown, 2002).

The teratogenic potency *in vivo* mentioned hereafter for all compounds refers to results obtained by the NMRI-exencephaly-mouse model as reported by Eikel *et al*. (2006b). 2-EMP has one isobutyl and one ethyl chain at the α-carbon and was used as racemic mixture. It was rated having no detectable teratogenic potency (0). Similarly, 2-IP, containing one isobutyl and one propynyl chain, was also used as racemate but exhibited weak teratogenicity (+). The alkyne derivatives originally were developed to abolish the hepatotoxicity of VPA caused by the metabolic intermediates 4-ene-VPA and 2,4-diene-VPA (Kassahun and Abbott, 1993). RProP and SProP have one propyl and one propynyl chain. The (*R*)-enantiomer was graded as of low teratogenicity (+), whereas the (*S*)-enantiomer exhibited very strong teratogenicity (+++++). RPenP and SPenP have one pentyl and one propynyl chain, with the (*R*)-enantiomer graded as medium teratogenic (+++), whereas the (*S*)-enantiomer exhibited again very high teratogenicity (+++++).

Inhibition of Differentiation

For the EST, concentration-dependent inhibition of differentiation was determined using cardiomyocyte contraction in EB outgrowths as an endpoint. D3 cells, initially grown in a "hanging drop" and then replated as individual EBs, were treated with different concentrations in the range of 1–1000 μg/ml of the respective compound for 10 days. The ID_{50}(D3) was derived from logistic regression analysis of the concentration-dependent reduction of beating EBs. Figures 2A and 2B show representative concentration-response curves for the seven compounds tested. Mean ID_{50}(D3) values are summarized in Figure 3A, and all ID_{50} and IC_{50} values of the EST together with the 95% confidence intervals are listed in Table 1. Differentiation was most potently inhibited by SPenP and SProP, and their ID_{50}(D3) values were more than 30-fold lower when compared with the least potent compound, 2-EMP (Table 1 and Fig. 3A). Comparing the differentiation endpoint with the rating of Eikel *et al*. (2006b) showed that this is in agreement with the high potency of SPenP and SProP and the inactivity of 2-EMP in the NMRI-exencephaly-mouse model. RPenP was 11-fold less potent than its (*S*)-enantiomer, and RProP was 17-fold less potent than its (*S*)-enantiomer (Table 1). This is in good correlation with their *in vivo* ranking of

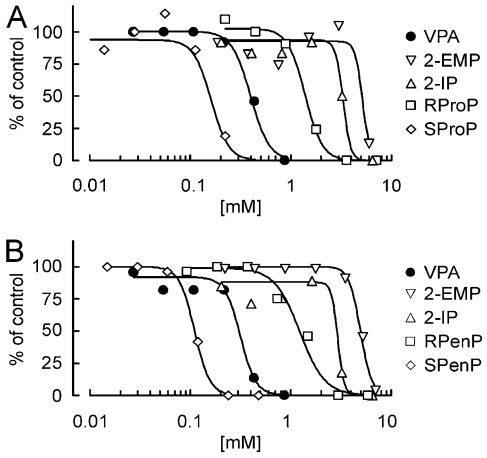

FIG. 2 Effects on cardiomyocyte differentiation of D3 cells in response to exposure against VPA or one of its congeners. Representative curves of concentration-dependent inhibition of differentiation after 10 days of treatment are shown in (A) for SProP, VPA, RProP, 2-IP, and 2-EMP and in (B) for SPenP, VPA, RPenP, 2-IP, and 2-EMP.

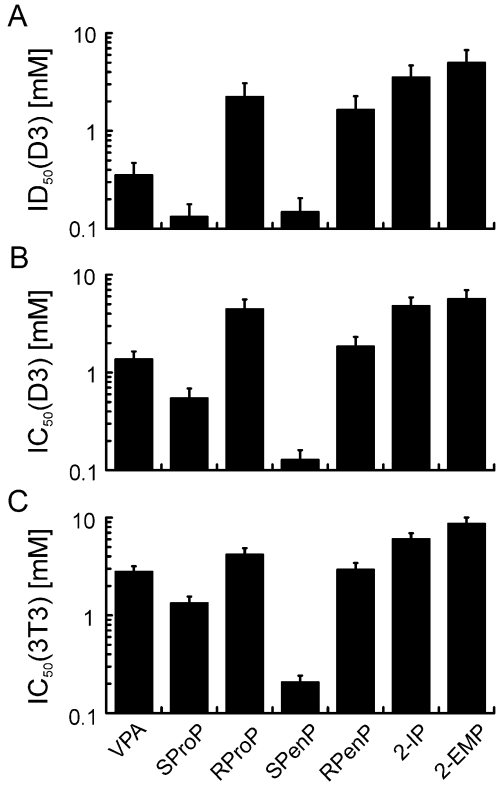

FIG. 3 Graphical presentation of ID_{50} and IC_{50} values of VPA and its congeners obtained in the EST. (A) Cardiomyocyte differentiation of D3 cells assessed on day 10 of differentiation by morphological evaluation. Viability was measured using the MTT assay after 10 days of treatment of (B) D3 cells and (C) 3T3 cells. Compound-specific ID_{50} and IC_{50} values (means ± 95% confidence interval) were derived from experiments performed by three different laboratories in at least two independent experiments each.

(+++++) for SPenP versus (+++) for RPenP and (+++++) for SProP versus (+) for RProP. The *in vitro* potency of 2-IP lies between those of RProP and 2-EMP, in line with its (+) rating *in vivo* (Eikel *et al*., 2006b). VPA and RPenP had both been ranked (+++) *in vivo* (Eikel *et al*., 2006b); however, in the differentiation assay, VPA was closer to the (*S*)-enantiomers and RPenP closer to RProP resulting in a fivefold difference in their ID_{50}(D3) values (Fig. 3A). The mean values for VPA were similar to previously reported ID_{50}(D3) values obtained with the same cell line (Buesen *et al*., 2009; Hettwer *et al*., 2010; Peters *et al*., 2008). Consequently, the ID_{50}(D3) values follow a ranking of SProP < SPenP < VPA << RPenP < RProP < 2-IP < 2-EMP (Table 1).

Cytotoxicity

To test for cytotoxicity, D3 and 3T3 cells were grown in a monolayer and treated with different concentrations in the range of 4–2000 μg/ml of the respective compound for 10 days. Viability was determined using the MTT assay, and IC_{50}(D3) and IC_{50}(3T3) values were derived from logistic regression analysis of the concentration-dependent reduction of MTT conversion.

TABLE 1
Mean Values and 95% Confidence Interval (Upper and Lower Bound) of the EST Derived from Experiments Performed in Three Independent Laboratories with At Least Two Independent Runs Each. The Rightmost Column Lists the Teratogenicity Ranking According to Eikel *et al.* (2006b)

Endpoint	Chemical	Mean (mM)	Lower bound (mM)	Upper bound (mM)	*In vivo*
ID_{50}(D3)					
	SProP	0.13	0.09	0.24	+++++
	SPenP	0.15	0.09	0.24	+++++
	VPA	0.35	0.24	0.53	+++
	RPenP	1.6	1.0	2.6	+++
	RProP	2.2	1.4	3.6	+
	2-IP	3.5	2.4	5.2	+
	2-EMP	5.0	3.3	7.6	0
IC_{50}(D3)					
	SPenP	0.13	0.10	0.17	+++++
	SProP	0.55	0.41	0.73	+++++
	VPA	1.4	1.1	1.7	+++
	RPenP	1.9	1.4	2.5	+++
	RProP	4.5	3.3	6.0	+
	2-IP	4.8	3.7	6.2	+
	2-EMP	5.7	4.4	7.3	0
IC_{50}(3T3)					
	SPenP	0.21	0.17	0.25	+++++
	SProP	1.3	1.1	1.6	+++++
	VPA	2.8	2.4	3.2	+++
	RPenP	2.9	2.5	3.5	+++
	RProP	4.2	3.5	5.0	+
	2-IP	6.0	5.1	7.1	+
	2-EMP	8.7	7.3	11	0

Note. Note that endpoint values are expressed in millimolar concentrations; the calculation of the mean values ± 95% confidence interval was based on their logarithmic endpoint values and was transformed to the original values in the table. The overall interlaboratory variance (σ) of the $\log_{10}$-transformed effect concentration was $\sigma < 0.18$ for the ID_{50}(D3) and $\sigma < 0.11$ for IC_{50}(D3) and IC_{50}(3T3). The intralaboratory variance was $\sigma < 0.25$ for the ID_{50}(D3) and the variance in IC_{50}(D3) and IC_{50}(3T3) values was $\sigma < 0.16$.

The viability of D3 cells in the presence of the seven compounds is summarized in Figure 3B, and individual IC_{50}(D3) values are listed in Table 1. Here, the strongest congener SPenP was more than 40-fold more potent than the weakest, 2-EMP. Similar to the differentiation endpoint, (*S*)-enantiomers were always more potent than the corresponding (*R*)-enantiomers. SPenP was 15-fold more potent than RPenP, and SProP was eightfold more potent than RProP. In addition, the longer chain derivatives SPenP and RPenP were four- and twofold more potent than their propyl chain counterparts SProP and RProP, respectively; a difference that was not evident in the differentiation assay. Moreover, VPA had an IC_{50}(D3) value similar to RPenP, and RProP was similar to 2-IP (Table 1). Overall, the following ranking according to the IC_{50}(D3) values emerged: SPenP << SProP << VPA < RPenP << RProP < 2-IP < 2-EMP (Table 1).

For the 3T3 cells, the IC_{50}(3T3) values are summarized in Figure 3C and Table 1. Most of the IC_{50}(3T3) values were higher compared with the respective IC_{50}(D3) values, and ranking of the substances according to their cytotoxicity in 3T3 cells yields a similar ranking as observed with D3 cells: SPenP << SProP < VPA < RPenP < RProP < 2-IP < 2-EMP (Table 1). This similarity in ranking was expected because all compounds belong to the same substance class.

Multiple Endpoint Comparisons

Multiple pairwise comparisons among the ID_{50} and IC_{50} estimates revealed that many but not all observed differences were significant (Fig. 4). For the differentiation endpoint, the ID_{50}(D3) values of SPenP, SProP, and VPA were not significantly different from each other. In a second group comprising RPenP, RProP, 2-IP, and 2-EMP, the ID_{50}(D3) values again were not significantly different from each other. However, each ID_{50}(D3) value from the first group was significantly different to any ID_{50}(D3) value in the second group, demonstrating that the (*S*)-enantiomers are significantly more potent than corresponding (*R*)-enantiomers ($p < 0.004$). Taken together, for the compound-inherent potency to inhibit D3 differentiation *in vitro*, a ranking according to significant differences follows (SPenP, SProP, and VPA) > (RPenP, RProP, 2-IP, and 2-EMP). With regard to cytotoxicity in D3 cells, VPA and RPenP were not significantly different and in a second group of RProP, 2-IP, and 2-EMP compounds again were not significantly different from each other. All other IC_{50}(D3) values were significantly separated. Therefore, (*S*)-enantiomers are significantly more cytotoxic than corresponding (*R*)-enantiomers ($p < 0.001$) and long-chain derivatives significantly more cytotoxic than corresponding short-chain congeners ($p < 0.001$). Thus, taking into account the significance of separation, the ranking according to cytotoxicity reads: SPenP > (SProP and VPA) > (VPA and RPenP) > (RProP, 2-IP, and 2-EMP).

A comparison of the inhibition of differentiation with the respective cytotoxicity in D3 cells reveals that the longer chain compounds RPenP and SPenP act in both assays at similar concentrations (Fig. 5), suggesting that there is a significant cytotoxic effect contributing to the reported *in vivo* teratogenicity. The shorter chain analogs RProP and SProP as well as VPA and to a lesser extend 2-IP exhibit lower ID_{50}(D3) than IC_{50}(D3) values, arguing for a specific effect on developmental differentiation of these compounds (Fig. 5). No systematic bias was observed.

Prediction Model

The PM in the form of the Fisher functions I, II, and III can only be used to classify substances into the three classes it was designed for. Application of the PM to our data classifies 2-EMP as Class 1 nonembryotoxic, and 2-IP, RProP, RPenP, VPA, SProP, and SPenP as Class 2 weakly embryotoxic. To allow for a ranking of the congeners according to their embryotoxic potency, we used the form of the two canonical discriminant functions LD1 and LD2 (Fig. 6). SPenP and SProP, the two (+++++) assigned congeners, are found in the right, Class 3–facing half of the 95% confidence region of Class 2, whereas VPA roughly in the middle (Fig. 6). RPenP, RProP, 2-IP, and 2-EMP are found in the intersection of the 95% confidence regions of Class 1 and Class 2 (Fig. 6). As is already evident from the ID_{50}(D3), VPA is also separate from the other two (+++) assigned congeners RPenP and RProP in the PM. 2-IP and 2-EMP are found close to the decision boundary of Class 2 and Class 1, respectively.

FACS-EST

Recently, we established an abbreviated, molecular-based protocol to the validated EST, entitled FACS-EST (Seiler *et al.*,

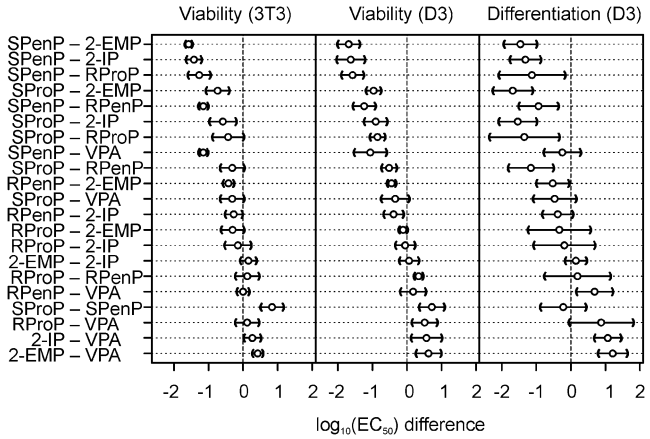

FIG. 4 Simultaneous confidence intervals for all pairwise comparisons of ID_{50} and IC_{50} mean values of VPA and its congeners obtained in the EST. The intervals were computed by a robust procedure accounting for heterogeneous variances. Confidence intervals that do not cross zero indicate statistically significant differences. Data are given as log_{10}-transformed millimolar concentrations.

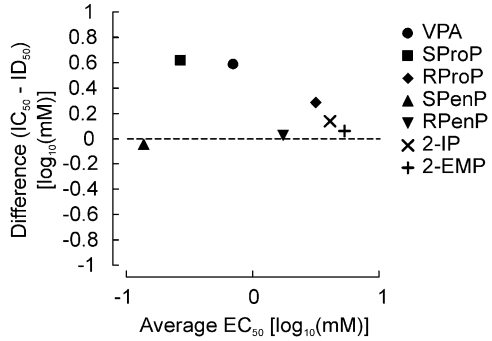

FIG. 5 Comparison of differentiation and cytotoxicity of VPA and its congeners in D3 cells. The difference of the log_{10}-transformed ID_{50} and IC_{50} estimates derived by morphological assessment of differentiation of D3 cells and cytotoxicity in D3 cells for a given substance is plotted against the average of the two estimates. Values above the dashed zero-difference line represent substances with a stronger effect on differentiation than on viability.

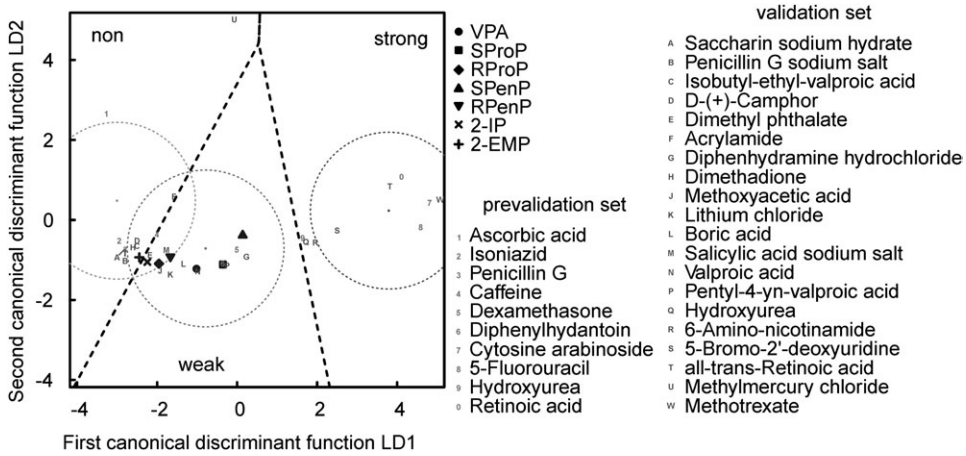

FIG. 6 Canonical plot of the VPA data on the two linear discriminant axes. In addition to the VPA congeners (symbols), the substances of the EST prevalidation set (numbers) and of the validation set (letters) are plotted for comparison. The small points and the large circles represent the group means and the 95% confidence regions for the three toxicity classes, respectively. The dashed lines represent the decision boundary that is used for classification into Class 1, nonembryotoxic (non), Class 2, weakly embryotoxic (weak), and Class 3, strongly embryotoxic (strong). Chemicals of the prevalidation set and the validation study were classified as follows: nonembryotoxic, 1–3 and A-G; weakly embryotoxic, 4–6 and H-P; strongly embryotoxic, 7–0 and Q-W.

2004). All measurements of the FACS-EST are performed after 7 days of treatment, considerably shortening the experimental protocol. Instead of cardiomyocyte contraction, expression of the cardiomyocyte marker protein sarcomeric αMHC is employed as the differentiation endpoint. In a previous study, we compared the FACS-EST to the validated EST using 10 unrelated compounds (Buesen *et al.*, 2009). Here, we test if the SAR of the VPA congeners demonstrated by the validated EST can be also obtained by the FACS-EST. The agreement between the three endpoints of the two methods was assessed using Bland-Altman plots (Bland and Altman, 2010) (Fig. 7). Indeed, a good agreement was found comparing the molecular endpoint versus the morphological endpoint as indicated by the inclusion of the zero-difference line by the 95% confidence interval (Fig. 7A), as well as in the viability measurements on day 7 versus day 10 of 3T3 cells (Fig. 7C). A nonsignificant common bias of 0.1 ($\log_{10}$ scale) was detected for the two differentiation endpoints (Fig. 7A) and the measurement of viability of 3T3 cells on day 7 versus day 10 (Fig. 7C). This means that for the seven congeners, the endpoints have a tendency to be more sensitive on day 7, i.e., the ID_{50}(D3) and the IC_{50}(3T3) values are on average 21% lower. Viability of D3 cells exhibited a significant common bias of 0.2 ($\log_{10}$ scale) on day 7 versus day 10 (Fig. 7B). This means that for the compounds used, this endpoint is more sensitive on day 7 with a 37% lower IC_{50}(D3) on average. Analysis of our previous data of 10 different compounds using Bland-Altman plots did not show a significant bias in any of the endpoints (Buesen *et al.*, 2009 and not shown). This suggests that the observed common bias of a higher sensitivity on day 7 for the VPA congeners is a specific effect of this substance class.

DISCUSSION

As part of an effort to comply with the recommendation to expand the database and thus to increase confidence for adoption of the EST by regulatory agencies (Marx-Stoelting *et al.*, 2009), we used the known embryotoxicant VPA and six of its congeners. All seven compounds are expected to exert their teratogenic effect directly, i.e., no metabolic activation is required. This is a prerequisite because the EST in its current form exhibits no or only little metabolic capacity. The application of closely related chemical structures in the present study was also meant to challenge the EST and its modified version and to test whether its applicability domain can be expanded to structural or stereochemical isomers of one particular compound. In good correlation with *in vivo* data (Eikel *et al.*, 2006b; Hauck and Nau, 1992), the SPenP and SProP exhibited the lowest ID_{50}(D3) and IC_{50}(D3) values, and 2-EMP was the least effective inhibitor of differentiation and the least cytotoxic compound. In general, the (*S*)-enantiomers proved more potent compared with their (*R*)-stereoisomers in all three assays used. This result is in excellent agreement with their relative *in vivo* teratogenic potency and similar to their potency of inhibition of histone deacetylases (Eikel *et al.*, 2006b). With increasing length of the alkyl side chain, VPA congeners exert increasing antiproliferative activity (Bojic *et al.*, 1996). Our data reproduce this effect with significantly different IC_{50}(D3) values of both pentyl derivatives compared with their propyl counterparts in the MTT tests, whereas the effect of these congeners on differentiation (i.e., the ID_{50}(D3)) was not significantly different (Fig. 4). This was expected and demonstrates that the differentiation endpoint is not directly affected by cytotoxicity. A comparison of the effects on differentiation and cytotoxicity in D3 cells revealed that VPA,

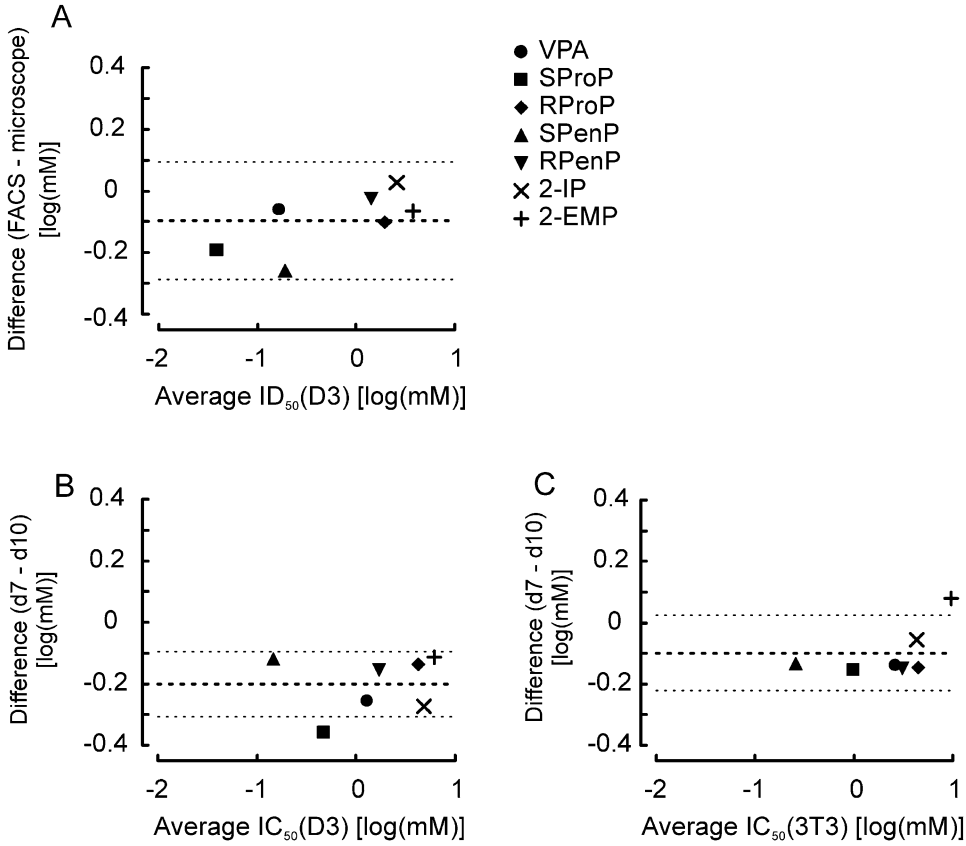

FIG. 7 Bland-Altman plots for inter-method comparisons between the validated EST and the molecular FACS-EST. (A) Inhibition of differentiation as assessed by morphology versus flow cytometry, (B) cytotoxicity in D3 cells on day 10 versus day 7, and (C) cytotoxicity in 3T3 cells on day 10 versus day 7. The difference of the $\log_{10}$-transformed ID_{50} and IC_{50} estimates derived by the respective assays for a given substance is plotted against the average of the two estimates. The dashed and dotted lines represent the mean and the 95% confidence interval (mean ± 1.96 × SD) of the difference, respectively.

SProP, and RProP exhibit a specific effect on differentiation, whereas SPenP and RPenP reached their half-maximal effect at similar concentrations in both assays, and therefore their effect on differentiation could be primarily because of cytotoxicity (Fig. 5).

In the cardiomyocyte differentiation assay, VPA was nearly as potent as SPenP and SProP (Fig. 2), and their ID_{50}(D3) values were not significantly different from each other (Fig. 4). In addition, RPenP exhibited an ID_{50}(D3) value similar to RProP but not to VPA as in the *in vivo* ranking (Eikel *et al.*, 2006b). Recently, a stem cell–based reporter cell line was generated that exploits a T-cell factor/lymphoid enhancer factor (LEF)–dependent luciferase reporter to monitor the Wnt/β-catenin signaling pathway (Uibel *et al.*, 2010). In this so-called ReProGlo assay, VPA induced reporter activity at lower concentrations than SProP (Uibel *et al.*, 2010). RPenP was found to be a fivefold less potent teratogen than VPA in the cardiomyocyte differentiation assay (Table 1) and similarly elicited no significant response in the ReProGlo system (Uibel *et al.*, 2010). Taken together, the ID_{50}(D3) values of VPA and RPenP are significantly different, and the two compounds also elicited different responses in two independent *in vitro* assays. A possible explanation for different responses in these *in vitro* assays could be differences either in the cell systems used or in the kinetic behavior of the corresponding compounds. To address the latter issue further and to adjust for such differences, methods to predict pharmacokinetic behavior *in silico* are currently under development (Verwei *et al.*, 2006). In line with this suggestion, VPA has been found to have a threefold shorter systemic half-life *in vivo* compared with the alkyne derivatives (Eikel *et al.*, 2006a; Löscher, 1999). However, it should be noted that the (+++) rating of VPA and RPenP by Eikel *et al.* (2006b) comprises an effect of 25–60% exencephaly rate and thus allows for considerable differences in the formation of this endpoint.

A small number of studies have addressed SAR *in vitro*. When a group of related glycol ether alkoxy acid metabolites was compared with *in vivo* data using the EST, the potency ranking was found to be the same, whereas the relative differences of *in vitro* results were smaller than those obtained *in vivo* (de Jong *et al.*, 2009). Other studies have investigated substance families such as heavy metal ions (Imai and Nakamura, 2006) and nanoparticles of different sizes (Park *et al.*, 2009); however, no *in vivo* data were presented. In the present study, the results cover a somewhat larger

concentration range of nearly two magnitudes compared with about one magnitude (0.25–3 mmol/kg) for the *in vivo* ranking (Eikel *et al*., 2006b). However, because of the variance in the differentiation assay, some values do not separate significantly. The ranking according to their respective ID_{50} and IC_{50} values nevertheless provides a valuable tool to monitor the relative effects of the congeners, as it resembles the ranking according to their teratogenic activity obtained from *in vivo* experiments.

The three functions of the PM of the EST have been designed to classify substances according to the criteria defined for the ECVAM international validation study (Brown, 2002; Scholz *et al*., 1999), and their application to an unrelated set of criteria does not yield satisfying results (Marx-Stoelting *et al*., 2009). To allow for visualization as well as gradual ranking of the PM data, we used its canonical form (Fig. 6). The congeners distribute in good agreement with the *in vivo* ranking. 2-EMP (0) is nonembryotoxic in the PM, and 2-IP (+) is weakly embryotoxic with its data point falling closely to the decision boundary of Class 1 and Class 2 (Fig. 6). RPenP and RProP fall more to the right, indicating higher potency, but separate from the other (+++) ranked substance VPA. The possible reasons for this separation have been discussed for the ID_{50}(D3) above. The two (+++++) ranked congeners SProP and SPenP clearly separate to the right, which is in good agreement with their high potency. The classification by Brown does not classify a substance based solely on the severity or breadth of its embryotoxic effects (Brown, 2002). Thus, it can be explained that Eikel *et al*. (2006b) ranked these two congeners as of very high teratogenicity based on their effects in a single mouse strain, whereas the PM predicts them as Class 2, weakly embryotoxic. It should be stressed that the near linear separation of the congeners on the first canonical discriminant function LD1, caused by the similarity in the values of the three endpoints, is in consequence of the shared mode of action. The substances of the prevalidation set and the validation study exhibit a wide spread on the second discriminant function, and this affects their classification especially at the decision boundaries (Fig. 6), underscoring the importance of all three endpoints for a balanced decision.

It is explicitly stated in the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACh) legislature, Annex XI.1.2, that *in vitro* methods can be included in a weight of evidence approach, and in Annex XI.1.4, it is stated that *in vitro* methods can be used for identifying a certain dangerous property. Furthermore, Annex XI.1.5 allows for grouping of substances whose properties are likely following a regular pattern. The validated EST is a suitable method in the sense of the REACh legislature as it has been validated by ECVAM (Genschow *et al*., 2004). However, the lack of metabolic activity of the validated EST means that the prediction of Class 1 (nonembryotoxic) for a substance is strictly valid only for the substance at hand. Metabolism of the substance and the effects of metabolites have to be tested separately. Conversely, a Class 2 or Class 3 predicted substance might be rapidly metabolized to a Class 1 substance *in vivo* or might not be able to cross the placental barrier. Thus, the validated EST cannot function as a stand-alone method to assess embryonic toxicity.

Recent efforts have focused on adapting the validated EST to the requirements of high-throughput screening applications. We investigated the use of molecular endpoints by flow cytometry and identified αMHC to be a valuable biomarker of cardiomyocyte differentiation (Seiler *et al*., 2004). Using a set of 10 well-known embryotoxicants that encompasses the full range of embryotoxic potencies, the FACS-EST exhibited the same sensitivity as the validated EST (Buesen *et al*., 2009). Here, the seven VPA congeners were tested in the FACS-EST, and the resulting ID_{50} and IC_{50} values were compared to the results of the validated EST (Fig. 7). Our results confirm that the FACS-EST is in good agreement with the validated EST. Because the FACS-EST is adaptable to other stem cell differentiation protocols, its applicability domain could be easily expanded to other cell lineages. To this end, transcriptomics and proteomics approaches can reveal new markers suitable as molecular endpoints (Groebe *et al*., 2010; van Dartel *et al*., 2010). Another approach was recently developed by Uibel *et al*. (2010), which takes advantage of a reporter stem cell line in the so-called ReProGlo assay (cf., above). By testing a similar set of molecules as in this study, however, only SPenP, SProP, and VPA yielded a significant effect on reporter gene activity. Conversely, RPenP, RproP, and 2-EMP were without effect on luciferase expression, and among all VPA derivatives tested, only SPenP revealed a weak effect on cell viability. The lower sensitivity of the ReProGlo system compared with the EST or FACS-EST is most probably because of the much shorter incubation time of 24 h.

During the validation of the EST and the development of the FACS-EST, the test compounds chosen were recruited from a wide range of different chemical classes to avoid bias because of structure-function relationships. In the present study, we demonstrate that the EST is suited to distinguish teratogenic potencies within a group of closely related VPA derivatives. The resulting ranking is in excellent agreement to the ranking obtained *in vivo* using the NMRI-exencephaly-mouse model. However, it is important to evaluate both the chemical-mediated effects on cardiomyocyte differentiation and cell viability, as was demonstrated by the differential effects of chirality and chain length on differentiation and cytotoxicity. Similar results to the validated EST were obtained using the FACS-EST. Our data thus strengthen the applicability of the EST and further confirm the FACS-EST as an efficient and reliable replacement of the EST toward a higher throughput.

FUNDING

German Federal Ministry for Education and Research (BMBF grant # 0313070).

ACKNOWLEDGMENTS

The authors declare that there are no conflicts of interest.

REFERENCES

Augustine-Rauch, K., Zhang, C. X., and Panzica-Kelly, J. M. (2010). In vitro developmental toxicology assays: a review of the state of the science of rodent and zebrafish whole embryo culture and embryonic stem cell assays. *Birth Defects Res. C Embryo Today* **90,** 87–98.

Bader, D., Masaki, T., and Fischman, D. A. (1982). Immunochemical analysis of myosin heavy chain during avian myogenesis in vivo and in vitro. *J. Cell Biol.* **95,** 763–770.

Batty, N., Malouf, G. G., and Issa, J. P. (2009). Histone deacetylase inhibitors as anti-neoplastic agents. *Cancer Lett.* **280,** 192–200.

Bland, J. M., and Altman, D. G. (2010). Statistical methods for assessing agreement between two methods of clinical measurement. *Int. J. Nurs. Stud.* **47,** 931–936.

Bojic, U., Elmazar, M. M., Hauck, R. S., and Nau, H. (1996). Further branching of valproate-related carboxylic acids reduces the teratogenic activity, but not the anticonvulsant effect. *Chem. Res. Toxicol.* **9,** 866–870.

Bowden, C. L. (2009). Anticonvulsants in bipolar disorders: current research and practice and future directions. *Bipolar Disord.* **11**(Suppl. 2), 20–33.

Brown, N. A. (2002). Selection of test chemicals for the ECVAM international validation study on in vitro embryotoxicity tests. *Altern. Lab. Anim.* **30,** 177–198.

Buesen, R., Genschow, E., Slawik, B., Visan, A., Spielmann, H., Luch, A., and Seiler, A. (2009). Embryonic stem cell test remastered: comparison between the validated EST and the new molecular FACS-EST for assessing developmental toxicity in vitro. *Toxicol. Sci.* **108,** 389–400.

Calabresi, P., Galletti, F., Rossi, C., Sarchielli, P., and Cupini, L. M. (2007). Antiepileptic drugs in migraine: from clinical aspects to cellular mechanisms. *Trends Pharmacol. Sci.* **28,** 188–195.

de Jong, E., Louisse, J., Verwei, M., Blaauboer, B. J., van de Sandt, J. J., Woutersen, R. A., Rietjens, I. M., and Piersma, A. H. (2009). Relative developmental toxicity of glycol ether alkoxy acid metabolites in the embryonic stem cell test as compared with the in vivo potency of their parent compounds. *Toxicol. Sci.* **110,** 117–124.

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W., and Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J. Embryol. Exp. Morphol.* **87,** 27–45.

Eikel, D., Hoffmann, K., Zoll, K., Lampen, A., and Nau, H. (2006a). S-2-pentyl-4-pentynoic hydroxamic acid and its metabolite S-2-pentyl-4-pentynoic acid in the NMRI-exencephaly-mouse model: pharmacokinetic profiles, teratogenic effects, and histone deacetylase inhibition abilities of further valproic acid hydroxamates and amides. *Drug Metab. Dispos.* **34,** 612–620.

Eikel, D., Lampen, A., and Nau, H. (2006b). Teratogenic effects mediated by inhibition of histone deacetylases: evidence from quantitative structure activity relationships of 20 valproic acid derivatives. *Chem. Res. Toxicol.* **19,** 272–278.

Genschow, E., Spielmann, H., Scholz, G., Pohl, I., Seiler, A., Clemann, N., Bremer, S., and Becker, K. (2004). Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern. Lab. Anim.* **32,** 209–244.

Genschow, E., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., *et al.* (2002). The ECVAM international validation study on in vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. European Centre for the Validation of Alternative Methods. *Altern. Lab. Anim.* **30,** 151–176.

Green, B. F. (1979). The two kinds of linear discriminant functions and their relationship. *J. Educ. Behav. Stat.* **4,** 247–263.

Groebe, K., Hayess, K., Klemm-Manns, M., Schwall, G., Wozny, W., Steemans, M., Peters, A. K., Sastri, C., Jaeckel, P., Stegmann, W., *et al.* (2010). Unexpected common mechanistic pathways for embryotoxicity of warfarin and lovastatin. *Reprod. Toxicol.* **30,** 121–130.

Hauck, R. S., and Nau, H. (1992). The enantiomers of the valproic acid analogue 2-n-propyl-4-pentynoic acid (4-yn-VPA): asymmetric synthesis and highly stereoselective teratogenicity in mice. *Pharm. Res.* **9,** 850–855.

Herberich, E., Sikorski, J., and Hothorn, T. (2010). A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. *PLoS One* **5,** e9788.

Hettwer, M., Reis-Fernandes, M. A., Iken, M., Ott, M., Steinberg, P., and Nau, H. (2010). Metabolic activation capacity by primary hepatocytes expands the applicability of the embryonic stem cell test as alternative to experimental animal testing. *Reprod. Toxicol.* **30,** 113–120.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* **50,** 346–363.

Imai, K., and Nakamura, M. (2006). In vitro embryotoxicity testing of metals for dental use by differentiation of embryonic stem cell test. *Congenit. Anom. (Kyoto)* **46,** 34–38.

Isoherranen, N., Yagen, B., and Bialer, M. (2003). New CNS-active drugs which are second-generation valproic acid: can they lead to the development of a magic bullet? *Curr. Opin. Neurol.* **16,** 203–211.

Kassahun, K., and Abbott, F. (1993). In vivo formation of the thiol conjugates of reactive metabolites of 4-ene VPA and its analog 4-pentenoic acid. *Drug Metab. Dispos.* **21,** 1098–1106.

Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **3,** 711–715.

Lehrman, G., Hogue, I. B., Palmer, S., Jennings, C., Spina, C. A., Wiegand, A., Landay, A. L., Coombs, R. W., Richman, D. D., Mellors, J. W., *et al.* (2005). Depletion of latent HIV-1 infection in vivo: a proof-of-concept study. *Lancet* **366,** 549–555.

Lheureux, P. E., and Hantson, P. (2009). Carnitine in the treatment of valproic acid-induced toxicity. *Clin. Toxicol. (Phila)* **47,** 101–111.

Löscher, W. (1999). Valproate: a reappraisal of its pharmacodynamic properties and mechanisms of action. *Prog. Neurobiol.* **58,** 31–59.

Marx-Stoelting, P., Adriaens, E., Ahr, H. J., Bremer, S., Garthoff, B., Gelbke, H. P., Piersma, A., Pellizzer, C., Reuter, U., Rogiers, V., *et al.* (2009). A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect Workshop. *Altern. Lab. Anim.* **37,** 313–328.

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods.* **65,** 55–63.

Ornoy, A. (2009). Valproic acid in pregnancy: how much are we endangering the embryo and fetus? *Reprod. Toxicol.* **28,** 1–10.

Paquette, J. A., Kumpf, S. W., Streck, R. D., Thomson, J. J., Chapin, R. E., and Stedman, D. B. (2008). Assessment of the embryonic stem cell test and application and use in the pharmaceutical industry. *Birth Defects Res. B Dev. Reprod. Toxicol.* **83,** 104–111.

Park, M. V., Annema, W., Salvati, A., Lesniak, A., Elsaesser, A., Barnes, C., McKerr, G., Howard, C. V., Lynch, I., Dawson, K. A., *et al.* (2009). In vitro developmental toxicity test detects inhibition of stem cell differentiation by silica nanoparticles. *Toxicol. Appl. Pharmacol.* **240,** 108–116.

Peters, A. K., Steemans, M., Hansen, E., Mesens, N., Verheyen, G. R., and Vanparys, P. (2008). Evaluation of the embryotoxic potency of compounds

in a newly revised high throughput embryonic stem cell test. *Toxicol. Sci.* **105,** 342–350.

Phiel, C. J., Zhang, F., Huang, E. Y., Guenther, M. G., Lazar, M. A., and Klein, P. S. (2001). Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. *J. Biol. Chem.* **276,** 36734–36741.

R Development Core Team. (2009). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria. Available at: http://www.R-project.org.

Ritz, C., and Streibig, J. C. (2005). Bioassay analysis using R. *J. Stat. Softw.* **12,** 1–22.

Scholz, G., Genschow, E., Pohl, I., Bremer, S., Paparella, M., Raabe, H., Southee, J., and Spielmann, H. (1999). Prevalidation of the embryonic stem cell test (EST)—a new in vitro embryotoxicity test. *Toxicol. In Vitro* **13,** 675–681.

Seiler, A., Visan, A., Buesen, R., Genschow, E., and Spielmann, H. (2004). Improvement of an in vitro stem cell assay for developmental toxicity: the use of molecular endpoints in the embryonic stem cell test. *Reprod. Toxicol.* **18,** 231–240.

Seiler, A. E., Buesen, R., Visan, A., and Spielmann, H. (2006). Use of murine embryonic stem cells in embryotoxicity assays: the embryonic stem cell test. *Methods Mol. Biol.* **329,** 371–395.

Spielmann, H. (2009). The way forward in reproductive/developmental toxicity testing. *Altern. Lab. Anim.* **37,** 641–656.

Spielmann, H., Pohl, I., Döring, B., Liebsch, M., and Moldenhauer, F. (1997). The embryonic stem cell test, an in vitro embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. *In Vitr Mol. Toxicol.* **10,** 119–127.

Tariot, P. N., Loy, R., Ryan, J. M., Porsteinsson, A., and Ismail, S. (2002). Mood stabilizers in Alzheimer's disease: symptomatic and neuroprotective rationales. *Adv. Drug Deliv. Rev.* **54,** 1567–1577.

Uibel, F., Muhleisen, A., Kohle, C., Weimer, M., Stummann, T. C., Bremer, S., and Schwarz, M. (2010). ReProGlo: a new stem cell-based reporter assay aimed to predict embryotoxic potential of drugs and chemicals. *Reprod. Toxicol.* **30,** 103–112.

van Dartel, D. A., Pennings, J. L., van Schooten, F. J., and Piersma, A. H. (2010). Transcriptomics-based identification of developmental toxicants through their interference with cardiomyocyte differentiation of embryonic stem cells. *Toxicol. Appl. Pharmacol.* **243,** 420–428.

Venables, W. N., and Ripley, B. D. (2002). In *Modern Applied Statistics with S*. Springer, New York, NY.

Verwei, M., van Burgsteden, J. A., Krul, C. A., van de Sandt, J. J., and Freidig, A. P. (2006). Prediction of in vivo embryotoxic effect levels witha combination of in vitro studies and PBPK modelling. *Toxicol. Lett.* **165,** 79–87.

Whitlow, S., Burgin, H., and Clemann, N. (2007). The embryonic stem cell test for the early selection of pharmaceutical compounds. *ALTEX* **24,** 3–7.