Toxicology and Applied Pharmacology 247 (2010) 91–97

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap

# Comparing three novel endpoints for developmental osteotoxicity in the embryonic stem cell test

Nicole I. zur Nieden [a,b,c,*,1], Lesley A. Davis [a], Derrick E. Rancourt [a,1]

[a] Institute of Maternal and Child Health, University of Calgary, Calgary, AB, Canada T2N 4N1
[b] Department of Cell Biology and Neuroscience and Stem Cell Center, University of California Riverside, Riverside, CA 92521, USA
[c] Fraunhofer Institute for Cell Therapy and Immunology, Perlickstrasse 1, 04103 Leipzig, Germany

ARTICLE INFO

*Article history:*
Received 2 February 2010
Revised 21 April 2010
Accepted 15 May 2010
Available online 26 May 2010

*Keywords:*
Embryonic stem cells
Osteogenesis
Embryotoxicity
Embryonic stem cell test
Retinoic acid
Developmental osteotoxicity

ABSTRACT

Birth defects belong to the most serious side effects of pharmaceutical compounds or environmental chemicals. *In vivo*, teratogens most often affect the normal development of bones, causing growth retardation, limb defects or craniofacial malformations. The embryonic stem cell test (EST) is one of the most promising models that allow the *in vitro* prediction of embryotoxicity, with one of its endpoints being bone tissue development. The present study was designed to describe three novel inexpensive endpoints to assess developmental osteotoxicity using the model compounds penicillin G (non-teratogenic), 5-fluorouracil (strong teratogen) and all-trans retinoic acid (bone teratogen). These three endpoints were: quantification of matrix incorporated calcium by (1) morphometric analysis and (2) measurement of calcium levels as well as (3) activity of alkaline phosphatase, an enzyme involved in matrix calcification.

To evaluate our data, we have compared the concentration curves and resulting $ID_{50}$s of the new endpoints with mRNA expression for osteocalcin. Osteocalcin is an exclusive marker found only in mineralized tissues, is regulated upon compound treatment and reliably predicts the potential of a chemical entity acting as a bone teratogen. By comparing the new endpoints to quantitative expression of osteocalcin, which we previously identified as suitable to detect developmental osteotoxicity, we were ultimately able to illustrate IMAGE analysis and $Ca^{2+}$ deposition assays as two reliable novel endpoints for the EST. This is of particular importance for routine industrial assessment of novel compounds as these two new endpoints may substitute previously used molecular read-out methods, which are often costly and time-consuming.

Published by Elsevier Inc.

## Introduction

The teratogenic potential of a compound can non-reversibly interfere with the normal development of an embryo, creating malformations and other birth defects. For the last two to three decades research and industry have been in urgent need for a practical *in vitro* embryotoxicity model able to substitute animal testing, since complete toxicological data sets are missing for most environmental toxicants, chemicals and new pharmaceutical candidate compounds. Over a decade ago, the currently most promising *in vitro* embryotoxicity test, the embryonic stem cell test (EST) was introduced by Spielmann et al. (1995, 1997). In contrast to other *in vitro* embryotoxicity tests, the EST uses two permanent cell lines of mouse origin, embryonic stem cells (ESCs) and differentiated 3T3 fibroblasts and thus does not mandate the killing of pregnant animals. ESCs are the only cells currently known that go through all developmental stages when differentiated *in vitro*, from a pluripotent cell to a fully mature somatic cell/tissue. Thus, this *in vitro* model comprises most stages of embryonic development. Treating these cells during *in vitro* differentiation with the compound of interest may be very useful to predict unwanted negative effects on embryonic development. Moreover, ESCs possess the ability to differentiate under proper culture conditions into not only one cell type, but into cell types of all three germ layers including those relevant to study toxicological effects (Wobus et al., 1995; Kramer et al., 2000; zur Nieden et al., 2003, 2005), thereby providing researchers with a system to study differentiation into a variety of distinct lineages starting with the same uncommitted cell.

The EST takes advantage of these properties by assessing the degree of inhibition that a substance causes on differentiation processes. The classic EST originally described by Spielmann et al. (1995) assesses two different endpoints, the inhibition of growth (cytotoxicity) for both cell lines, whereby ESCs represent embryo-like effects, and adult 3T3 fibroblasts mimic effects on maternal tissues and organs (Scholz et al., 1999a,b; Peters et al., 2008). The second endpoint that was evaluated is the differentiation of ESCs into contracting cardiomyocytes. In the

* Corresponding author. 1113 Biological Science Building, Department of Cell Biology and Neuroscience, University of California Riverside, Riverside, CA 92521, USA. Fax: +1 951 827 3087.
*E-mail addresses:* nicole.zurnieden@ucr.edu, nicole.zurnieden@izi.fraunhofer.de (N.I. zur Nieden).
[1] Authors contributed equally as senior authors.

0041-008X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.taap.2010.05.010

classic EST, this is done by counting the proportion of differentiating embryoid bodies containing contracting cardiomyocytes in the presence of a test substance. Previously, we demonstrated that the EST could be enhanced through inclusion of a more reliably quantifiable molecular endpoint (zur Nieden et al., 2001). Instead of contraction counting, the expression levels of the cardiac marker gene myosin heavy chain (MHC) were measured by quantitative real-time RT-PCR under influence of a test compound.

Despite its validation in an EU-wide study (Genschow et al., 2004), the original EST was limited in its ability to identify developmental inhibition. Specifically, the analysis of cardiomyocyte differentiation alone as an endpoint does not suffice for the determination of embryotoxic effects. Recently, we and others further improved the EST to include molecular endpoints for bone, cartilage, as well as tissues of neuronal and endodermal origin (zur Nieden et al., 2004; Festag et al., 2007). Further investigations are on the way to replace subjective assessment by objective measurements. As such, genetic modifications and molecular markers such as the ones detected by FACS and quantitative gene expression analyses offer significant advantages in developing a new test approach (Bremer et al., 1999; Kolossov et al., 1998; Seiler et al., 2004; reviewed in Huuskonen, 2005).

Here, we propose to utilize IMAGE analysis to quantify osteogenic impairment caused by compound treatment. In this technique, cultures are photographed and the percentage of black pixels is calculated (Purpura et al., 2003). This is only possible because osteocalcin-expressing mineralized osteoblasts appear black in phase contrast microscopy (zur Nieden et al., 2003). We analyzed osteoblast yield by measuring: (1) the degree of calcification in the cultures using morphometric IMAGE analysis, (2) the amount of mineralized calcium incorporated into the matrix, and (3) alkaline phosphatase (ALP) enzyme activity. In order to confirm the utility of these three macromolecular markers as endpoints for osteotoxicity, the osteotoxic potential of three test compounds penicillin G (non-teratogenic), 5-fluorouracil, and all-trans retinoic acid (strong teratogens) were studied.

## Materials and methods

*Cell culture and differentiation of ESCs into osteoblasts.* NIH 3T3 cells were expanded in DMEM (Invitrogen) with 10% FCS (Invitrogen), 2 mM glutamine, 50 U/ml penicillin and 50 µg/ml streptomycin and passaged when reaching 80% confluency. Cells of the mouse ESC line, D3, were cultured with the addition of LIF (1000 U/ml) as described (zur Nieden et al., 2001). Differentiation of ESCs into osteoblasts was induced in hanging drops (zur Nieden et al., 2003, 2004). Briefly, a single cell suspension of $3.75\times10^4$ cells per ml was generated by trypsinization. ESCs were forced to aggregate into embryoid bodies (EBs) in drops of 20 µl volume for three days followed by transfer of EBs into suspension culture for an additional two days. EBs were then plated onto tissue-culture adhesive substrate and supplemented with 10 mM β-glycerophosphate, 50 µg/ml ascorbic acid and $5\times10^{-8}$ M dihydroxy vitamin $D_3$ from day 5 onwards (zur Nieden et al., 2003). The medium was changed every two days after day 5.

*Test compounds.* The test chemicals were chosen from a list that was published for the validation of *in vitro* embryotoxicity tests by the American Society for Teratology (Smith et al., 1983). Penicillin G and 5-fluorouracil were used in the original validation study of the classic EST and the molecular multiple endpoint EST (Scholz et al., 1998; Spielmann et al., 2001; zur Nieden et al., 2004) with penicillin G (PenG) being included as a negative control since it is not known to have adverse effects on reproduction (Boucher and Delost, 1964; Heinonen et al., 1977). All-trans retinoic acid (RA) was chosen from the validation of the molecular multiple endpoint EST as being a prototypic teratogenic substance with specific concentration-dependent effects on osteoblast differentiation (zur Nieden et al., 2004).

Penicillin was dissolved in phosphate buffered saline, while 5-fluorouracil and RA were dissolved in DMSO. For all endpoint assays control cultures containing vehicle were included. Final concentrations of DMSO in the culture were kept below 0.01%, which is a decade below the 0.1% that has been specified as the highest non-toxic DMSO concentration allowable (Adler et al., 2006). Since sensitivity of ESCs to toxic agents is related to their differentiation state (Laschinski et al., 1991), compound treatment was performed throughout the entire culture time. To avoid photodegradation of the light sensitive compounds RA and PenG, cultures were microscopically monitored before media changes were performed.

*Cytotoxicity assay.* The cytotoxic effects of test substances in both cell models were determined using an MTT assay as described (zur Nieden et al., 2001). Cells were seeded at a density of 3125 cells/$cm^2$ and cultivated in the presence of a dilution series of test chemicals over the course of 10 days with media changes on day 3 and day 5. Both the test substance and control (solvent only) were tested in six independent experiments. At day 10, cells were incubated with MTT solution (Sigma) for 2 h at 37 °C. The change in absorbance reflecting the decrease in mitochondrial dehydrogenase activity was read in a Benchmark Plus spectrophotometer (Biorad) at 570/630 nm. The cytotoxicity of the test substance was determined from a dose–response curve at that concentration, which reduced the growth of the cells in comparison to the solvent control by 50% ($IC_{50}$).

*IMAGE analysis.* We have previously established that mineralized osteocalcin-expressing osteoblasts show a black coloration in phase contrast optics (zur Nieden et al., 2003). This feature may be applied to quantify the degree of mineralization of the cultures using morphometric IMAGE analysis (zur Nieden et al., 2007a). Grey-scale photographs of EBs were taken on day 30 of the differentiation with the SPOT Advanced Imaging system (Diagnostic Instruments) mounted to an Olympus IX70 inverted microscope ($n=6$ ea, 3 independent experiments) and a 2× objective with a consistent gamma adjustment of 1.8 and consistent exposure times. The pixel bit depth was set to a monochromatic value of 8 bpp. An intensity value in the range between 0 (white) and 255 (black) was assigned to each pixel with the IMAGE J 1.33u software available from the website of the National Institutes of Health (http://www.rsb.info.nih.gov/ij/). The mean black value was calculated for each compound concentration and expressed as *n*-fold up-regulation above the vehicle control, which was set as 100%. Values for the half-maximal inhibition of differentiation ($ID_{50}$) were determined from dose–response curves by linear interpolation.

*Quantitative PCR.* RNA was isolated from 20 EBs per treatment group ($n=3$) on day 30 of *in vitro* culture, a time point at which cells were previously shown to exhibit a peak expression for osteocalcin (zur Nieden et al., 2003). RNA concentrations of samples were measured using RiboGreen (Invitrogen) and cDNA synthesized as described (zur Nieden et al., 2001). Quantitative real-time PCR was performed in a Biorad iCycler using a SYBR green PCR master mix (Biorad) as described (Cormier et al., 2006) with a specific primer set for osteocalcin (zur Nieden et al., 2004). Specificity of reaction product was checked with post-run melting curves. Target gene $C_T$ values were normalized to the respective $C_T$ values obtained for GAPDH. Raw data obtained for experimental groups were compared to time matched vehicle controls, which were set as 100%. Values for the half-maximal inhibition of differentiation ($ID_{50}$) were determined from dose–response curves by linear interpolation.

*$Ca^{2+}$ determination.* Matrix incorporated calcium of the cultures was quantified on day 30 of differentiation using the purple substrate Arsenazo III (DCL, Toronto, Canada) as described previously (zur Nieden et al., 2005). Prior to lysis in RIPA buffer (150 mM NaCl, 10 mM

Tris, pH 7.2, 0.1% SDS, 1% Triton X-100, 1% Deoxycholate, 5 mM EDTA containing 1 mM phenylmethylsulfonyl fluoride, 10 mM benzamidine, 2 μg/ml leupeptin, 100 μM sodium orthovanadate and 10 mM p-nitrophenylphosphate), cultures were washed with PBS to remove remaining traces of medium. Values were calculated from a $CaCl_2$ standard, which was measured along with the samples at 650 nm in a Benchmark Plus microplate spectrophotometer (Biorad). Calcium content was normalized to the total protein content of the samples, measured with DC protein assay reagent (Biorad) as described (zur Nieden et al., 2007b). Values obtained for compound treated cultures (5 wells ea concentration, three independent experiments) were normalized to solvent control cultures, which were set as 100%.

*ALP assay*. Alkaline phosphatase activity was determined from the same protein lysates using p-nitrophenyl phosphate (Sigma) as a substrate. The production of yellow p-nitrophenol was stopped with 3 N NaOH after a 20 min incubation at 37 °C. Absorbance was read at $t=0$ and $t=20$ min at 405 nm. Enzyme activity was calculated as described before (zur Nieden et al., 2003).

*Classification of teratogenicity*. In order to classify the three test compounds into their *in vivo* teratogenicity classes, three different models were employed, a biostatistical prediction model based on linear discriminant functions, which was developed by the ZEBET (Center for Documentation and Evaluation of Alternative Methods to Animal Experiments) (Spielmann, 1997; Genschow et al., 2000, 2002), a linear regression model (zur Nieden et al., 2001) and a logarithmic calculation previously described by Bremer et al. (2001). The prediction model distinguishes three classes of teratogenicity: strong, weak and non-embryotoxic, taking into account the $IC_{50}$ of a compound (measured on ESCs and 3T3) and the $ID_{50}$ of the same compound. Linear regression analysis contrasts the $IC_{50}$ found for ESCs and the $ID_{50}$ only. In the logarithmic classification, a compound falls into the strong embryotoxic category, if $\log ID_{50} \leq 1$ and in the non-strong category, if $\log ID_{50} > 1$.

*Statistical analysis*. All data are presented as average ± standard deviation. Significance of differences in $IC_{50}$ or $ID_{50}$ values was determined with a student's *t*-test using the web-based GraphPad QuickCalcs data analysis and biostatistics software available at http://graphpad.com/quickcalcs/.

## Results

### *Control substances for the embryonic stem cell test*

Penicillin G and 5-fluorouracil are the classical negative and positive substances routinely used to characterize new endpoints for the EST. As expected for a non-teratogenic substance (Boucher and Delost, 1964; Heinonen et al., 1977), penicillin G displayed no cytotoxicity in a concentration range up to 210 μg/ml (Fig. 1A). The no-effect level (reduction to 80%, NoEL) is reached at a concentration of around 500 μg/ml with an $IC_{50}$ value of 800 ± 10.3 μg/ml. The reduction in osteogenic differentiation yield is phenotypically noticeable only in the highest concentration tested (Fig. 1B). The inhibition of osteogenic differentiation by this compound can be attributed to its general cytotoxicity, as the shape of the concentration–response curves for the osteogenic endpoints IMAGE analysis and OCN expression closely resemble that of the MTT-curve (Fig. 1C). For the other two osteogenic endpoints, the concentration–response curve did not reach the 50% levels even at a concentration of 1000 μg/ml, which is the cut-off concentration for the EST. However, penicillin G displayed its cytotoxicity towards ESCs at concentrations that are beyond its therapeutic range.

Being a strong teratogen *in vivo*, 5-fluorouracil (5-FU) showed an intense general cytotoxicity leading to $IC_{50}$ and $ID_{50}$ values at very low concentrations in this and previous studies (Table 1). Undifferentiated cells were susceptible to the toxic effects of 5-FU at very low dosage with an $IC_{50}$ of 0.022 ± 0.0049 μg/ml (Fig. 2A, Table 1). Judging from the presence of black mineralized cells, a concentration-dependent decrease in the osteogenic yield is expected for all endpoints (Fig. 2B). The cells were more sensitive in all differentiation endpoints than in the MTT assay suggesting an osteotoxic effect of 5-FU, which exceeds its cytotoxicity.

Furthermore, we noted that all three new differentiation endpoints were as sensitive if not more sensitive as the gene expression analysis (Fig. 2C). The only endpoint for which the concentration–response curve did not show an exponential decrease, but rather showed a wavy pattern, is ALP activity. At the NoEL for OCN expression and IMAGE analysis, ALP activity was still at 112%. Although the curve for the calcium endpoint was also at over 110% at the NoEL of the gene expression and IMAGE analysis curve, the shape of the curve followed the exponential trend expected for a classical concentration–response curve. Although these specific differences were noted in the endpoints, when comparing the resulting $ID_{50}$ values, FU seemed to affect all endpoints similarly as the values were in the same concentration range (Table 1).

### *New endpoints evaluated for retinoic acid*

Consistent with being a strong teratogen *in vivo*, cytotoxicity of all-trans retinoic acid (RA) was seen at a half-maximal inhibition concentration of $9\times10^{-5} \pm 6.6\times10^{-6}$ μg/ml (Fig. 3A). In the specific concentration range that was chosen for this study, RA is known to induce osteogenesis (zur Nieden et al., 2004). Congruent with this, a mild increase in the number of calcified cells was noticeable in the lowest RA concentration tested (Fig. 3A). Although the classical embryotoxic

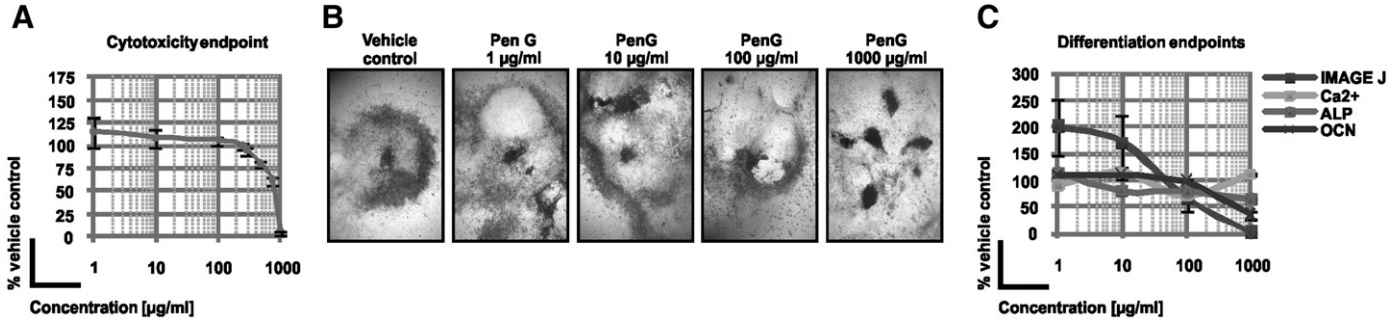


**Fig. 1.** Endpoints of toxicity assessed for penicillin G (Pen G). (A) Concentration–response curve for cytotoxicity measured with the MTT assay, $n=6\pm$SD. (B) Phenotypic appearance of osteogenic cultures under different concentrations of test chemical compared to non-compound treated vehicle controls. (C) Concentration–response curves for morphometric analysis (IMAGE J), calcium assay ($Ca^{2+}$), alkaline phosphatase activity (ALP) and osteocalcin gene expression (OCN). Data were averaged from three independent runs with varying replicates (see Materials and methods) ± SD. Quantitative PCR data for OCN were normalized to GAPDH and values were standardized to values obtained for non-compound treated vehicle controls, which were set as 100%.

**Table 1**
$ID_{50}$ values for test compounds as indicated taken from all assays conducted. Grey shaded cells indicate significantly lower $ID_{50}$ compared to the $IC_{50}$. PenG = penicillin G, 5-FU = 5-fluorouracil, RA = all-trans retinoic acid. $^{*}P<0.1$, $^{**}P<0.001$, $^{***}P<0.0001$, student's *t*-test.

| | MTT D3 | EB | | | |
|---|---|---|---|---|---|
| | | Image | $Ca^{2+}$ | ALP | OCN |
| PenG | 800 ± 10.3 | 333 ± 198* | > 1000 | > 1000 | 593 ± 100.7* |
| 5-FU | 0.022 ± 0.0049 | 0.0038 ± 0.0023** | 0.0089 ± 0.0045* | 0.0051 ± 0.009** | 0.0069 ± 0.0023** |
| RA | $9 \times 10^{-5} \pm 6.6 \times 10^{-6}$ | $5.85 \times 10^{-3} \pm 0.0004$*** | 0.03 ± 0.0005*** | 0.0025 ± 0.0006** | 0.00051 ± 0.0009 |

potential of a compound is usually represented by it causing a decrease in differentiation, any disturbance of normal embryonic development may cause maldevelopment *in vivo*. Therefore, all differentiation endpoints give higher $ID_{50}$ values than the $IC_{50}$ for this substance (Table 1). The three novel endpoint protocols yielded almost identical half-maximal inhibition concentrations and similar dose–response curves, consistent with a correlation between the three endpoints (Fig. 3C): the expression of the bone-specific marker gene osteocalcin and the deposition of a mineralized extracellular matrix that underlies the calcium measurement as well as the morphological examination by IMAGE analysis.

### *Classification of teratogenicity*

In a next step, a linear correlation chart was then generated for all four endpoints to analyze the relationship between cytotoxicity and embryotoxicity (Fig. 4), as applied by Bigot et al. (1999) and our lab (zur Nieden et al., 2001) for cardiac toxicity. Compounds were classified as embryotoxic if a significant alteration in differentiation was seen without cytotoxic effects at the same concentration or if the cytotoxic effect was clearly less pronounced than the differentiation effect. Based on these criteria, none of the four tested endpoints showed a correct classification of the three compounds tested. Retinoic acid was always classified as non-embryotoxic as its $IC_{50}$ value in ESCs was significantly higher than all of the $ID_{50}$s. However, when embryotoxicity is evaluated based on the concentration–response curves only, it is noticeable that the osteotoxic effect of RA seen *in vivo* may be based on increased bone target gene expression and also an increase in mineralization, as we have suggested in an earlier study (zur Nieden et al., 2004).

The classical EST employs a biostatistical prediction model based on linear discriminant functions, which was developed by the ZEBET (Center for Documentation and Evaluation of Alternative Methods to Animal Experiments) in order to classify the embryotoxic potential of chemicals into the three classes: non-, weakly/moderate or strongly embryotoxic (Spielmann, 1997; Genschow et al., 2000, 2002). The $IC_{50}$ and the $ID_{50}$ values generated from the concentration–response curves of OCN expression and the three novel differentiation endpoints were then applied to this prediction model (Table 2). With this prediction model, all three test compounds were classified correctly in all four endpoints. In order to graphically represent these categorization results, we then applied a logarithmic calculation for the specified $ID_{50}$ values to graph the compounds into the appropriate embryotoxicity categories (Fig. 5), a method that was used in a similar manner for cardiac embryotoxicity (Bremer et al., 2001). With this method, two categorizations are possible: non-strong and strong embryotoxic. A compound falls into the strong embryotoxic category, if $\log ID_{50} \leq 1$ and in the non-strong category, if $\log ID_{50} > 1$. With the logarithmic chart, all three compounds were classified correctly according to their known *in vivo* potential to cause developmental osteotoxicity. Therefore, the logarithmic evaluation of $ID_{50}$ values seems to be a useful tool in predictive toxicology.

## Discussion

Due to new legislation, up to 30,000 chemicals that are currently on the market will have to be re-evaluated over the next ten years within the European Union with an estimated use of 10 million animals for *in vivo* teratogenicity testing. Therefore, there is a need for more rapid and especially economical *in vitro* tests that are suitable for high-throughput screenings. The embryonic stem cell test (EST) is one of three *in vitro* embryotoxicity tests, which has recently been evaluated in an ECVAM (European Centre for the Validation of Alternative Methods) validation study of the European Union (Scholz et al., 1999a,b). Developmental toxicity is currently assessed in animal experiments performed according to OECD guideline no. 643. Evaluating developmental osteotoxicity with the EST *in vitro* has previously been shown to detect tissue-specific embryotoxic effects (zur Nieden et al., 2004) and therefore has the potential to substitute animal experiments in this specific endpoint. A small set of six chemicals with known *in vivo* teratogenic potential was tested in a pilot study in order to evaluate the feasibility of using quantifiable gene expression data to determine the embryotoxic potential of compounds in the molecular multiple endpoint EST (mme-EST) (zur Nieden et al., 2004). Although quantitative real-time PCR has proven useful to predict developmental osteotoxicity, these types of molecular analyses require the use of expensive assay reagents. Sample

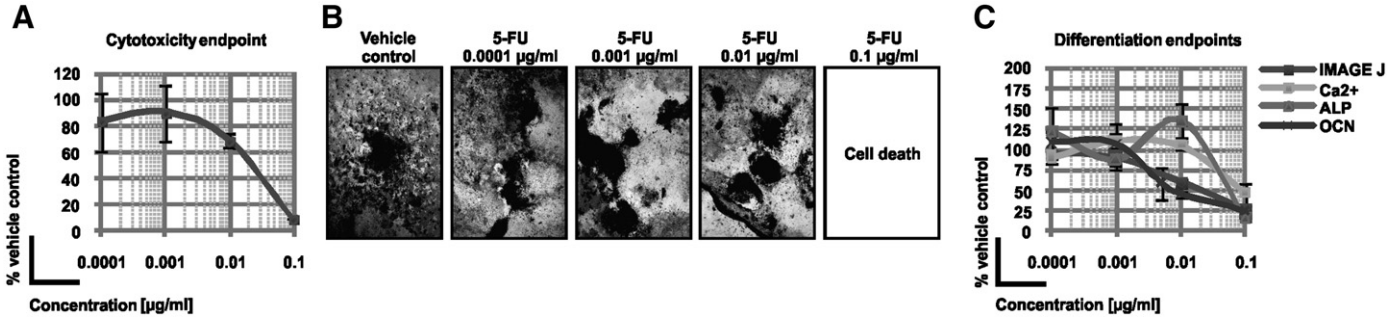


**Fig. 2.** Endpoints of toxicity assessed for 5-fluorouracil (5-FU). (A) Concentration–response curve for the cytotoxicity endpoint measured with the MTT assay, $n = 6 \pm$ SD. (B) Phenotypic appearance of osteogenic cultures under different concentrations of test chemical compared to non-treated solvent controls. Excessive cell death was noted in the highest concentration of 5-FU tested. (C) Concentration–response curves for morphometric analysis (IMAGE J), calcium assay ($Ca^{2+}$), alkaline phosphatase activity (ALP) and osteocalcin gene expression (OCN). Data were averaged from three independent runs with varying replicates (see Materials and methods) and graphed with SD. Quantitative PCR data for OCN were normalized to GAPDH and values were standardized to values obtained for non-treated solvent controls, which were set as 100%.

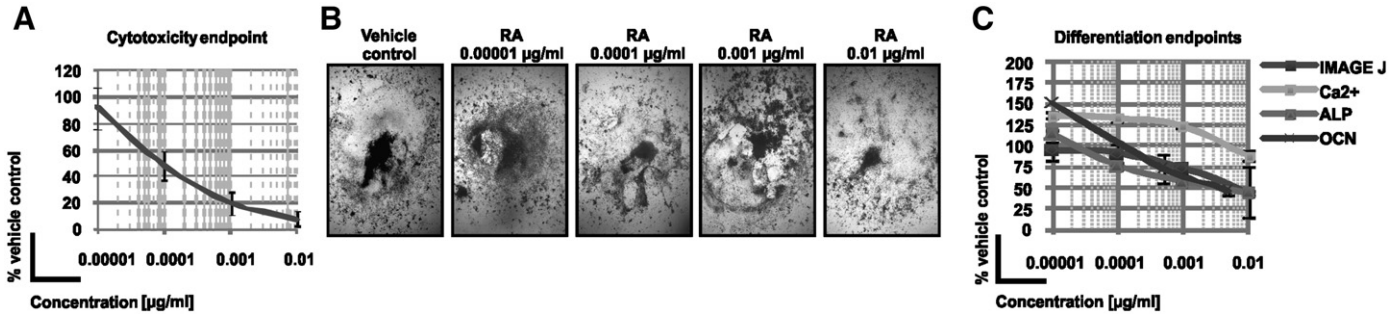


**Fig. 3.** Endpoints of toxicity assessed for all-trans retinoic acid (RA). (A) Concentration–response curve for cytotoxicity measured with the MTT assay after an assay duration of 10 days, $n = 6 \pm$ SD. (B) Phenotypic appearance of osteogenic cultures under different RA concentrations. (C) Concentration–response curves for morphometric analysis (IMAGE J), calcium assay ($Ca^{2+}$), alkaline phosphatase activity (ALP) and osteocalcin gene expression (OCN). Data are expressed as mean ± SD (3 independent runs with varying replicates (see Materials and methods)). Quantitative PCR data for OCN were normalized to GAPDH and values were standardized to values obtained for non-RA treated osteogenic cultures, which were set as 100%.

preparation includes RNA isolation, quantification, quality assessment of resulting RNA, cDNA synthesis and lastly reagents to conduct quantitative PCR. The analysis takes about two days, which introduces a time component, in addition to the significant cost of the endpoint assay. The goal of this study was therefore to identify assays that would be more affordable in high-throughput, yet be as effective as molecular read-outs for predictability and reliability.

We have tested here three novel endpoints that may be used instead of tissue-specific gene expression profiling on a set of chemicals that are classically used as positive and negative controls for the EST (zur Nieden et al., 2001). Out of these three endpoints, two correctly evaluated the degree of osteogenic differentiation under compound influence, therefore allowing prenatal bone toxicity to be detected *in vitro*. Penicillin G is typically categorized as a non-teratogenic substance and thus shows no developmental toxicity in the endpoint bone with none of the four endpoint assays tested. In addition, 5-FU, a strong *in vivo* teratogen, is classified correctly according to its *in vivo* teratogenicity with the novel osteotoxicity endpoints. Indeed, reduced ossification has been described in the skeleton of rat embryos after maternal exposure to 5-FU (Shuey et al., 1994). Numerous malformations, such as cleft palate and deformation of hind limbs, are induced by 5-FU in developing embryos from a variety of species (Dagg, 1960; Stephens et al., 1980; Shuey et al., 1994).

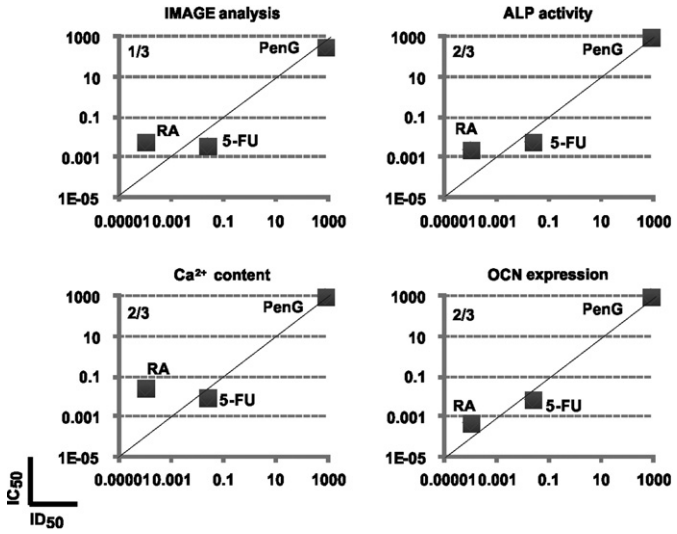


**Fig. 4.** Linear regression chart for all endpoints. Half-maximal inhibition of cell growth (cytotoxicity, $IC_{50}$) is graphed versus the half-maximal inhibition of differentiation ($ID_{50}$) for each test compound in the four potential endpoints for developmental osteotoxicity. PenG = penicillin G, 5-FU = 5-fluorouracil, RA = all-trans retinoic acid, OCN = osteocalcin, ALP = alkaline phosphatase activity.

All-trans retinoic acid (RA) was chosen as a substance with specific inhibitory or supportive effects on bone development depending on the concentration range tested (zur Nieden et al., 2004). *In vivo*, RA triggers multiple effects via distinct molecular mechanisms inducing ectodermal, mesodermal and endodermal differentiation in a time and concentration-dependent manner thus being a strong *in vivo* teratogen. The *in vitro* effects seen specifically on osteogenesis represent the capacity of RA to induce abnormal developmental and growth of cartilage and skeletal elements *in vivo* as well as *in vitro* (Eichele, 1989; Raouf and Seth, 2000). RA has been described as a potent inducer of bone formation (zur Nieden et al., 2004), which is potentially based on its enhancing influence on ALP activity and the expression of the Vitamin $D_3$ receptor, as was described in clonal osteoblastic cell lines (Grigoriadis et al., 1986). RA can also stimulate the secretion of osteopontin (Denhardt and Noda, 1998), a protein that binds to calcified matrices and links organic phases with inorganic molecules, another potential cause for its osteogenic effect. RA treatment at a concentration of $10^{-7}$ M has been shown to delay expression of T-brachyury, a mesodermal gene normally expressed at around day 4 of embryoid body development (Pellizzer et al., 2004). Since osteoblasts arise from the mesodermal lineage, this delay in T-brachyury expression may very well account for the inhibitory effects of RA on osteoblast differentiation in certain concentration ranges. In turn, our group has shown that RA treatment at the same concentration, but applied at the time of osteoprogenitor commitment, has resulted in a higher number of osteoblasts through down-regulation of nuclear activity of beta-catenin, a gene that plays a critical role in stem cell differentiation (zur Nieden et al., 2007a). Together, these studies suggest that an inhibition of differentiation as well as an enhancement of differentiation seen in the ESC *in vitro* model may mirror the teratogenic *in vivo* effects of test compounds.

One of the endpoint assays chosen in this study, the IMAGE analysis, had been previously applied for the quantitative measurement of embryonic toxins in a slightly different context, but has been qualified as a valid endpoint read-out. Utilizing a genetically modified ESC line that expresses the green fluorescent protein under the control of the endodermal alpha-fetoprotein promotor Paparella et al. (2002) quantified the number of green pixels in images taken from cultures incubated with test compounds. We show here that we can apply a similar procedure to accurately measure the effect of embryotoxic agents on bone development by evaluating the number of black pixels corresponding to the presence of mineralized osteoblasts. Likewise, measuring the levels of calcium that the cells incorporate into their matrix during differentiation,

**Table 2**
Classification of compounds according to the prediction model of the classic EST (as per Genschow et al. (2000)).

| | IMAGE | $Ca^{2+}$ | ALP | OCN |
|---|---|---|---|---|
| PenG | I | I | I | I |
| 5-FU | III | III | III | III |
| RA | III | III | III | III |

A

| Osteotoxic class | Mean log [μg/ml] | | | |
|---|---|---|---|---|
| | IMAGE J | $Ca^{2+}$ | ALP | OCN |
| Non-strong (PenG) | 2.52 | 3.00 | 3.00 | 3.00 |
| Strong (5-FU) | -1.42 | -1.05 | -1.29 | -2.16 |
| Strong (RA) | -2.23 | -1.52 | -2.60 | -3.29 |

B

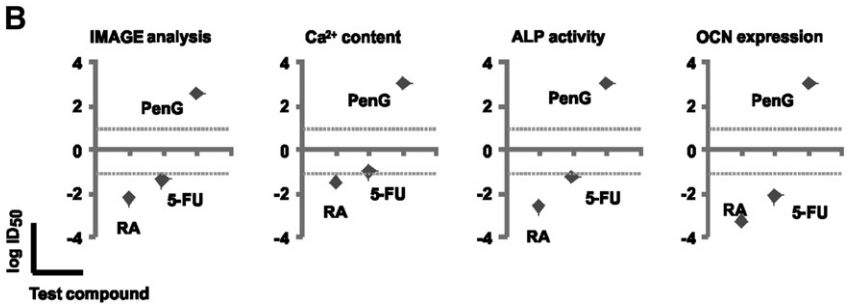


**Fig. 5.** Logarithmic evaluation for all endpoint assays conducted. (A) Descriptive statistics for $ID_{50}$s (as per Bremer et al. (2001)). (B) Classification chart based on logarithmic values of $ID_{50}$s. PenG = penicillin G, 5-FU = 5-fluorouracil, RA = all-trans retinoic acid, OCN = osteocalcin, ALP = alkaline phosphatase activity.

all three test substances were classified correctly according to their known *in vivo* teratogenicity class. These absorbance based assays do not require the use of expensive reagents, but are promising in reliably predicting the osteotoxic potential of a compound thereby substituting for molecular endpoints. The test assay that we could not recommend for its use in the prediction of osteotoxicity was alkaline phosphatase activity as this endpoint measurement did not give reproducible enough data points.

## Conclusion

Undoubtedly, there is an urgent need for more rapid and economical *in vitro* tests that reliably evaluate developmental osteotoxicity. By using the new endpoints described here, the assessment of osteotoxicity becomes more affordable as compared to the previously identified molecular endpoint evaluation.

## Acknowledgments

Contributions: NzN and LAD carried out cell culture and assays based on NzN's design of the study. NzN drafted the manuscript. DER provided supplies and approved the final manuscript.

This study was supported by the Alberta Heritage Foundation for Medical Research (AHFMR) through a fellowship to NzN. LAD acknowledges funding support from the Canadian Stem Cell Network. DER is an AHFMR Senior Scholar. The authors thank Dr. Brad Thomas for making us aware of the freely available IMAGE J software.

## References

Adler, S., Pellizzer, C., Paparella, M., Hartung, T., Bremer, S. (2006). The effects of solvents on embryonic stem cell differentiation. Toxicol. In Vitro 20 (3), 265–271.

Bigot, K., De Lange, J., Archer, J., Clothier, R., Bremer, S. (1999). The relative semi-quantification of mRNA expression as a useful toxicological endpoint for the identification of embryotoxic/teratogenic substances. Toxicol. In Vitro 13, 619–623.

Boucher, D., Delost, P. (1964). Developpement post-natal des descendents issus de meres traitees par la penicilline au cours de la gestion chez la souris. C.R. Soc. Biol. 158, 528–532.

Bremer, S., van Dooren, M., Paparella, M., Kolossov, E., Fleischmann, B., Hescheler, J. (1999). Establishment of an *in vitro* embryotoxicity assay with green-fluorescent protein-expressing embryonic cell-derived cardiomyocytes. ATLA 27, 471–484.

Bremer, S., Worth, A.P., Paparella, M., Bigot, K., Kolossov, E., Fleischmann, B.K., Hescheler, J., Balls, M. (2001). Establishment of an *in vitro* reporter gene assay for developmental cardiac toxicity. Toxicol. In Vitro 15 (3), 215–223 Jun.

Cormier, J.T., zur Nieden, N.I., Rancourt, D.E., Kallos, M.S. (2006). Expansion of undifferentiated murine embryonic stem cells as aggregates in suspension culture bioreactors. Tissue Eng. 12 (11), 3233–3245.

Dagg, C.P. (1960). Sensitive stages for the production of developmental abnormalities in mice with 5-fluorouracil. Am. J. Anat. 106, 89–96.

Denhardt, D.T., Noda, M. (1998). Osteopontin expression and function: role in bone remodeling. J. Cell. Biochem. Suppl 30–31, 90–102.

Eichele, G. (1989). Retinoids and vertebrate limb pattern formation. Trends Genet. 5, 246–251.

Festag, M., Viertel, B., Steinberg, P., Sehner, C. (2007). An *in vitro* embryotoxicity assay based on the disturbance of the differentiation of murine embryonic stem cells into endothelial cells. II. Testing of compounds. Toxicol. In Vitro 21 (8), 1631–1640.

Genschow, E., Scholz, G., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S., Becker, K., Spielmann, H. (2000). Development of prediction models for three *in vitro* embryotoxicity tests in an ECVAM validation study. In Vitro Mol. Toxicol. 13 (1), 51–66.

Genschow, E., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S., Becker, K. (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 (2), 151–176 Mar-Apr.

Genschow, E., Spielmann, H., Scholz, G., Pohl, I., Seiler, A., Clemann, N., Bremer, S., 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 (3), 209–244.

Grigoriadis, A., Petkovich, P.M., Rosenthal, E.E., Heersche, J.N.M. (1986). Modulation by retinoic acid of 1, 25-dihydroxyvitamin $D_3$ effects on alkaline phosphatase activity and parathyroid hormone responsiveness in an osteoblast-like osteosarcoma cell line. Endocrinology 119, 932–939.

Heinonen, O.P., Slone, D., Shapiro, S., Kaufman, D.W. (1977). Birth Defects and Drugs in Pregnancy. John Wright Publishing Sciences Group, Littleton, MA.

Huuskonen, H. (2005). New models and molecular markers in evaluation of developmental toxicity. Toxicol. Appl. Pharmacol. 207 (2 Suppl), 495–500.

Kolossov, E., Fleischmann, B.K., Liu, Q., Bloch, W., Viatchenko-Karpinski, S., Manzke, O., Ji, G.J., Bohlen, H., Addicks, K., Hescheler, J. (1998). Functional characteristics of ES cell-derived cardiac precursor cells identified by tissue-specific expression of the green fluorescent protein. J. Cell Biol. 143, 2045–2056.

Kramer, J., Hegert, C., Guan, K., Wobus, A.M., Müller, P.K., Rohwedel, J. (2000). Embryonic stem cell-derived chondrogenic differentiation *in vitro*: activation by BMP-2 and BMP-4. Mech. Dev. 92 (2), 193–205.

Laschinski, G., Vogel, R., Spielmann, H. (1991). Cytotoxicity test using blastocyst-derived euploid embryonal stem cells: a new approach to *in vitro* teratogenesis screening. Reprod. Toxicol. 5, 57–64.

Paparella, M., Kolossov, E., Fleischmann, B.K., Hescheler, J., Bremer, S. (2002). The use of quantitative image analysis in the assessment of *in vitro* embryotoxicity endpoints based on a novel embryonic stem cell clone with endoderm-related GFP expression. Toxicol. In Vitro 16, 589–597.

Pellizzer, C., Adler, S., Corvi, R., Hartung, T., Bremer, S. (2004). Monitoring of teratogenic effects *in vitro* by analyzing a selected gene expression pattern. Toxicol. In Vitro 18, 325–335.

Peters, A.K., Steemans, M., Hansen, E., Mesens, N., Verheyen, G.R., Vanparys, P. (2008). Evaluation of the embryotoxic potency of compounds in a newly revised high throughput embryonic stem cell test. Toxicol. Sci. 105 (2), 342–350.

Purpura, K.A., Aubin, J.E., Zandstra, P.W. (2003). Two-color image analysis discriminates between mineralized and unmineralized bone nodules *in vitro*. Biotechniques 34 (6), 1188–1192.

Raouf, A., Seth, A. (2000). Ets transcription factors and targets in osteogenesis. Oncogene 19, 6455–6463.

Scholz, G., Pohl, I., Seiler, A., Bremer, S., Brown, N.A., Piersma, A.H., Holzhütter, H.G., Spielmann, H. (1998). Results of the first phase of the ECVAM project "prevalidation and validation of three *in vitro* embryotoxicity tests". ALTEX 15 (1), 3–8 Article in German.

Scholz, G., Pohl, I., Genschow, E., Klemm, M., Spielmann, H. (1999a). Embryotoxicity screening using embryonic stem cells *in vitro*: correlation to *in vivo* teratogenicity. Cells Tissues Organs 165 (3–4), 203–211.

Scholz, G., Genshow, E., Pohl, I., Bremer, S., Paparella, M., Raabe, H., Southee, J., Spielmann, H. (1999b). 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., 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.

Shuey, D.L., Lau, C., Logson, T.R., Zucker, R.M., Elstein, K.H., Narotsky, M.G., Setzer, R.W., Kavlock, R.J., Rogers, J.M. (1994). Biologically based dose–response modeling in developmental toxicology: biochemical and cellular sequelae of 5-fluorouracil exposure in the developing rat. Toxicol. Appl. Pharmacol. 126, 129–144.

Smith, M.K., Kimmel, G.L., Kochhar, D.M., Shepard, S.P., Spielberg, S.P., Wilson, J.G. (1983). A selection of candidate compounds for *in vitro* teratogenesis test validation. Teratogen. Carcinogen. Mutagen. 3, 461–480.

Spielmann, H. (1997). Animal alternatives in Germany. Science 276 (5309), 19.

Spielmann, H., Pohl, I., Döring, B., Moldenhauer, F. (1995). *In vitro* embryotoxicity assay using two permanent cell lines: mouse embryonic stem cells and 3T3 fibroblasts. Abstracts of the 23 ETS conference 1995, Dublin. Teratology 51 (6), 31A–32A.

Spielmann, H., Genschow, E., Scholz, G., Brown, N.A., Piersma, A.H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S., Becker, K. (2001). Preliminary results of the ECVAM validation study on three *in vitro* embryotoxicity tests. Altern. Lab. Anim. 29 (3), 301–303.

Stephens, J.D., Golbus, M.S., Miller, T.R., Wilber, R.R., Epstein, C.J. (1980). Multiple congenital anomalies in a fetus exposed to 5-fluorouracil during the first trimester. Am. J. Obstet. Gynecol. 137, 747–749.

Wobus, A.M., Rohwedel, J., Maltsev, V., Hescheler, J. (1995). Development of cardiomyocytes expressing cardiac-specific genes, action potentials, and ionic channels during embryonic stem cell-derived cardiogenesis. Ann. NY Acad. Sci. 752, 460–469.

zur Nieden, N.I., Ruf, L.J., Kempka, G., Hildebrand, H., Ahr, H.J. (2001). Molecular markers in embryonic stem cells. In vitro Toxicol. 15, 455–461.

zur Nieden, N.I., Kempka, G., Ahr, H.J. (2003). *In vitro* differentiation of embryonic stem cells into mineralized osteoblasts. Differentiation 71 (1), 18–27.

zur Nieden, N.I., Kempka, G., Ahr, H.J. (2004). Molecular multiple endpoint embryonic stem cell test — a possible approach to test for the teratogenic potential of compounds. Toxicol. Appl. Pharmacol. 194, 257–269.

zur Nieden, N.I., Kempka, G., Rancourt, D.E., Ahr, H.J. (2005). Induction of chondro-, osteo- and adipogenesis in embryonic stem cells by bone morphogenetic protein-2: effect of cofactors on differentiating lineages. BMC Dev. Biol. 5, 1.

zur Nieden, N.I., Price, F.D., Davis, L.A., Everitt, R.E., Rancourt, D.E. (2007a). Gene profiling on mixed embryonic stem cell populations reveals a biphasic role for beta-catenin in osteogenic differentiation. Mol. Endocrinol. 21 (3), 674–685.

zur Nieden, N.I., Cormier, J.T., Kallos, M.S., Rancourt, D.E., Rancourt, D.E. (2007b). Murine embryonic stem cells maintain pluripotency after long-term culture in suspension bioreactors. J. Biotechnol. 129 (3), 421–432.