Food and Chemical Toxicology 123 (2019) 50–56

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

Pre-validation study of alternative developmental toxicity test using mouse embryonic stem cell-derived embryoid bodies

Jae-Hwan Lee[a], Seon Young Park[a], Changhwan Ahn[a], Cho-Won Kim[b], Ji-eun Kim[c], Na Rae Jo[d], Hee Young Kang[a], Yeong-Min Yoo[a], Eui-Man Jung[a], Eun Mi Kim[c], Ki-Suk Kim[c], Kyung-Chul Choi[b], Sung Duck Lee[d], Eui-Bae Jeung[a,*]

[a] Laboratory of Veterinary Biochemistry and Molecular Biology, College of Veterinary Medicine, Chungbuk National University, Cheongju, Chungbuk, 28644, Republic of Korea
[b] Laboratory of Biochemistry and Immunology, College of Veterinary Medicine, Chungbuk National University, Cheongju, Chungbuk, 28644, Republic of Korea
[c] R&D center for Advanced Pharmaceuticals & Evaluation, Korea Institute of Toxicology, Daejeon, 34114, Republic of Korea
[d] Department of Information and Statistics, College of Natural Sciences, Chungbuk National University, Cheongju, Chungbuk, 28644, Republic of Korea

ARTICLE INFO

Keywords:
Validation
Hanging drop
Embryoid body's area
Developmental toxicity

ABSTRACT

The embryoid body test (EBT) is a developmental toxicity test method that assesses the half inhibitory concentrations of substances in the area of embryoid bodies (EBs), and in the viability of mouse embryonic stem cells (ESCs) and fibroblasts (3T3 cells) following chemical exposure for three and four days, respectively. In the previous study, the EBT showed more advanced than the embryonic stem cell test (EST) from the European Centre for the Validation of Alternative Methods (ECVAM) applying cardiac differentiation of mouse ESCs, because the EBT greatly reduced the exposure time, labor, and amount of materials required, and misclassification of embryotoxic potential. This pre-validation study evaluated the predictive accuracy of the EBT using 26 coded test substances by two steps: intra-laboratory and inter-laboratory reproducibility tests. Since some substances have different embryotoxic potentials at different pregnancy periods, in this study, a new prediction model consisting of non-toxic and toxic classes was used, instead of the existing prediction model assessing embryotoxicants in four classes. The results of the intra- and inter-laboratory tests were highly accurate (above 80%) when substances were classified using the predictive model. In conclusion, EBT can accurately classify various embryotoxicants in a short time with less effort and greater validation.

1. Introduction

Embryotoxicants inhibit the development of the fetus, causing growth retardation or teratogenesis. Therefore, it is important to evaluate drug side effects during drug development. There are various OECD test guidelines (TG), such as prenatal developmental toxicity (TG 414), one generation reproduction toxicology (TG 415), two-generation reproduction toxicology (TG 416), reproduction/developmental toxicity screening test (TG 421), and combined repeated dose toxicity study along with a reproduction/developmental toxicity screening test (TG 422). These TG protocols are generated after paying the costs associated with sacrificing many laboratory animals during test trials, which are typically high cost and time to consume (Scholz et al., 1999).

Alternative test methods using ESCs include the EST method of ECVAM and the Hand 1-Luc EST method of JaCVAM (Le Coz et al., 2015; Nagahori et al., 2016; Seiler and Spielmann, 2011; Tandon and Jyoti, 2012). The EST testing method proposed by ECVAM can be used to evaluate cytotoxicity in mouse ESCs and fibroblasts (3T3 cells) as well as assessing the inhibition of cardiac differentiation of ESCs (Genschow et al., 2004). The Hand 1-Luc EST method of JaCVAM involves transfection of ES cells (ES-D3) with heart and neural crest derivatives expressing transcript 1 (Hand1), and the expression of the Hand1 gene is monitored through the luciferase reporter gene to confirm its toxicity (Le Coz et al., 2015). These methods consume a lot of time and effort moreover there can be large errors depending on the experimenter performing the test.

The embryoid bodies test (EBT), a developmental toxicity testing method, is a cytotoxicity test that evaluates reductions in embryoid body (EB) area in order to determine drug toxicity (Ohnuki and Kurosawa, 2013; Zhou et al., 2010). The EBT measures the reduction of EB area as an endpoint after 3 days of chemical exposure and avoids problems associated with conventional cytotoxicity tests. In the EBT,

* Corresponding author.
E-mail address: ebjeung@chungbuk.ac.kr (E.-B. Jeung).

https://doi.org/10.1016/j.fct.2018.10.044
Received 26 June 2018; Received in revised form 22 September 2018; Accepted 15 October 2018
Available online 16 October 2018
0278-6915/

the morphology of the EBs in a toxic chemical-treated group is uneven and their area is smaller; moreover, the change in EB area depends on the degree of toxicity of the chemical. Therefore, the EBT method can assess the developmental toxicity of chemicals through evaluation of the cross-sectional area of the EBs. The area of mouse EBs (mEBs) has the potential to be used as an endpoint to evaluate embryonic toxic chemicals. The mEB-EST has shown identical prediction accuracy to that of the ECVAM-EST. Since the mEB-EST can reduce time and labor involved in testing, it is useful for assessing developmental toxicity. Prediction models were proposed to assess and classify developmental toxicants (Kang et al., 2017).

In this study, to establish the reliability and feasibility of the EBT, experiments were conducted with identity-blind, coded chemicals in three laboratories. The usefulness of the EBT as an alternative for assessing developmental toxicity was confirmed by assessing intra-laboratory reproducibility, inter-laboratory transferability, and inter-laboratory reproducibility of results for 26 identity-blinded test chemicals. The classification of the tested embryotoxicant was determined by applying a biostatistical prediction model.

2. Materials and methods

2.1. Organization for EBT validation

Members of the Validation Management Team, Chemical Management Group, and Data Analysis Group (including a biostatistician) were divided according to the South Korean Centre for the Validation of Alternative Methods (KoCVAM). The Validation Management Team was organized into the lead laboratory and participatory laboratories 1 and 2 (Fig. 1a). The three laboratories are as follows: Lead laboratory, Laboratory of Veterinary Biochemistry and Molecular Biology, College of Veterinary Medicine, Chungbuk National University; Participatory laboratory 1: Laboratory of Biochemistry and Immunology, College of Veterinary Medicine, Chungbuk National University; and Participatory laboratory 2: Predictive Model Research Centre, Korea Institute of Toxicology.

The identities of the chemicals were encrypted by KoCVAM and the code-named chemicals distributed to the Validation Management Team. Each laboratory conducted the experiments, measured the parameters, entered the appropriate values into the data sheets, and analyzed the results. The analysis results file was then submitted to KoCVAM, and KoCVAM opened the identity-encrypted codes. The analysis files and codes were then submitted to the Data Analysis Group's biostatistician, and statistical analyses were carried out by using predictive models. Also, intra-laboratory reproducibility, inter-laboratory transferability, and inter-laboratory reproducibility proceeded as shown in the following Fig. 1b.

2.2. Chemicals

In this pre-validation study, 26 chemicals were selected along with *in vivo* classification judged by an external expert group (Table 1) and were classified into two groups, including non-toxic and toxic, rather than classifying them as four classes, including non, weak, moderate, and strong, because some chemicals have shown the different toxic effects during embryonic development. Most chemicals had already been used during pre-validation of other alternative developmental toxicity test methods, for which results have been reported by the ECVAM (Marx-Stoelting et al., 2009; Rohwedel et al., 2001) and the Japanese Centre for the Validation of Alternative Methods (JaCVAM) (Le Coz et al., 2015; Nagahori et al., 2016). The test chemicals used were encoded by the chemical manager of KoCVAM and the participating laboratories did not know the identity of any chemical.

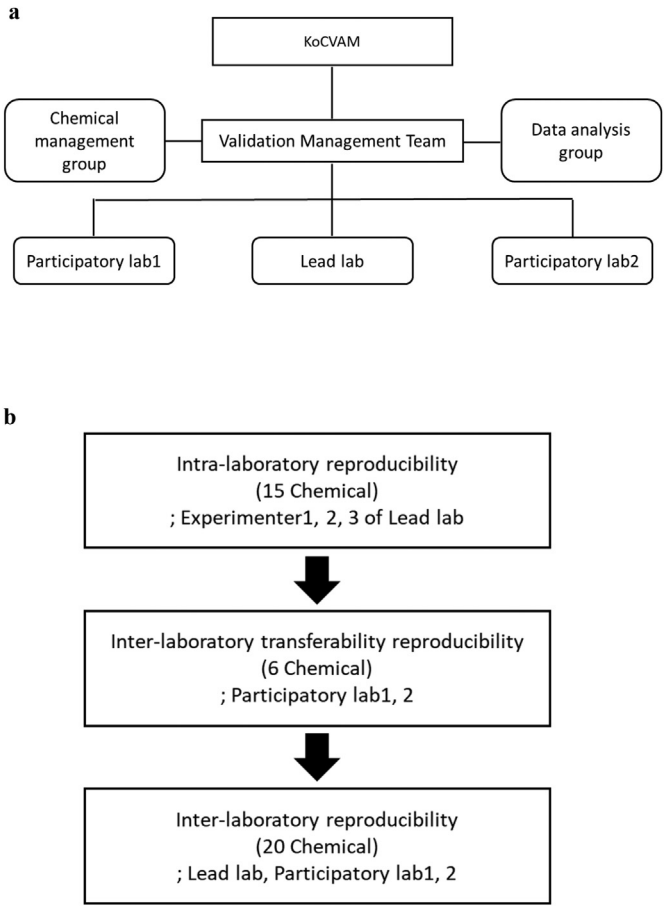

Fig. 1. a. Organization of EBT validation management Team. **b.** Organizing groups according to experiment.

Table 1
Information on the 26 test chemicals including chemical name, CAS number, molecular weight, and *in vivo* class (non-toxic or toxic). All chemicals were purchased from Sigma-Aldrich, USA.

No	Chemicals	CAS No	Solvent	MW	In vivo class
1	Ascorbic acid	134-03-2	DMSO	198.11	Non-toxicity
2	Aspartame	22839-47-0	Media	294.3	Non-toxicity
3	Penicillin G	113-98-4	Media	372.48	Non-toxicity
4	Saccharin	82385-42-0	Media	205.17	Non-toxicity
5	Sodium bicarbonate	144-55-8	Media	84.01	Non-toxicity
6	Sodium gluconate	527-07-1	Media	218.14	Non-toxicity
7	Sorbitol	50-70-4	Media	182.17	Non-toxicity
8	Sucralose	56038-13-2	Media	397.63	Non-toxicity
9	Xylitol	87-99-0	Media	152.15	Non-toxicity
10	Verapamil-HCl	152-11-4	DMSO	491.06	Toxicity
11	5,5 diphenylhydantoin	57-41-0	DMSO	252.27	Toxicity
12	Caffeine	58-08-02	DMSO	194.19	Toxicity
13	Clomiphene citrate	125-85-9	DMSO	325.87	Toxicity
14	Dexamethasone	50-02-02	DMSO	392.46	Toxicity
15	Diphenhydramine	147-24-0	DMSO	291.82	Toxicity
16	Doxylamine succinate salt	562-10-7	DMSO	388.46	Toxicity
17	D-Penicillamine	52-67-5	DMSO	149.21	Toxicity
18	Indomethacin	53-86-1	DMSO	357.79	Toxicity
19	Letrozol	112809-51-5	DMSO	285.3	Toxicity
20	Methimazole	60-56-0	DMSO	114.17	Toxicity
21	Papaverine hydrochloride	61-25-6	DMSO	375.85	Toxicity
22	Paroxetine hydrochloride	110429-35-1	DMSO	374.83	Toxicity
23	Acetazolamide	59-66-5	DMSO	222.25	Toxicity
24	Retinoic acid	302-79-4	DMSO	300.44	Toxicity
25	Ribavirin	36791-04-5	DMSO	244.2	Toxicity
26	Tetracycline hydrochloride	64-75-5	DMSO	480.9	Toxicity

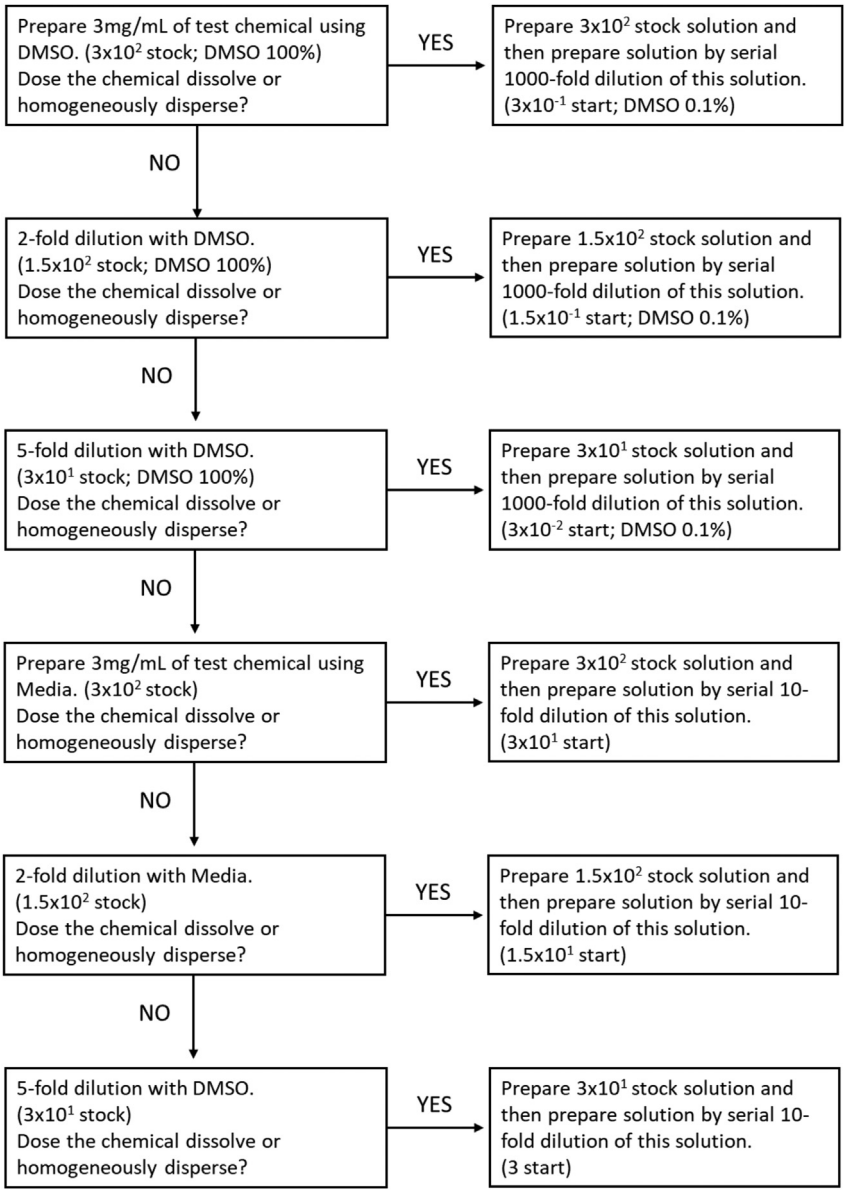

Fig. 2. Solvent selection and sample preparation. Two solvents were used in the EBT. Selection of solvent was dependent on solubility, DMSO or media and was selected based on the process described in this figure (Kojima et al., 2013).

2.3. Solvent selection and sample preparation

The workflow including the period from solvent selection to sample preparation is described in Fig. 2 (Kojima et al., 2013).

Initially, a 300 mg/mL solution of a test chemical was prepared by using DMSO (Sigma-Aldrich) as the solvent, and the dissolution pattern of the chemical was observed. If the chemical dissolved or homogeneously dispersed (Notes 1 and 2), DMSO was chosen as the solvent for this chemical.

When the chemical did not dissolve or homogeneously disperse in DMSO, the use of culture media as the solvent was attempted. If the chemical dissolved or homogeneously dispersed in the culture medium, the media became the solvent for this chemical.

Note (1) If maintained for more than 5 min, it is considered to be uniformly dispersed.

Note (2) Dissolution was accomplished by vortexing (1–60 min) or warming (5–10 min), as appropriate.

2.4. Cell line and culture

Mouse ESCs (ES-E14TG2a) and 3T3 cells (Clone A31) were purchased from the American Type Culture Collection (ATCC; Manassas, VA, USA). The ESCs were grown on mitomycin C-treated mEFs and were cultured in DMEM/F-12 (Gibco, Logan, UT, USA) culture medium supplemented with 10% heat-inactivated fetal bovine serum (FBS; Gibco), non-essential amino acids (NEAA, 1×; Gibco), 10^{-4} M 2-mercaptoethanol, 100 U/mL of penicillin, and 100 μg/mL of streptomycin along with mouse leukemia inhibitory factor (mLIF, 10 ng/mL; Millipore, Darmstadt, Germany) at 37 °C in a 5% CO_2 humidified tissue culture incubator (Sanyo, San Diego, CA, USA). All experiments have used the ESCs between passage numbers 20 and 30. The 3T3 cells were cultured in DMEM (Gibco) supplemented with 10% FBS, 100 U/mL of penicillin, and 100 μg/mL of streptomycin.

2.5. Measurement of three endpoints

The EBT consists of three endpoints: the half inhibition

concentration for cell viability of mouse ESCs (IC_{50} E14) and 3T3 fibroblasts (IC_{50} 3T3) measured by the CCK assay after treating the chemical for 4 or 10 days, and the half inhibition concentration for EB area (ID_{50} EB) determined by analyzing images taken by phase contrast microscope after forming EBs in the medium containing the chemical for 3 days. Refer to our previous study for detailed protocols for measuring the three endpoints (Marx-Stoelting et al., 2009).

2.6. Statistical analysis and prediction model (PM)

Concentration-response curves were generated for each test chemical and a one-site fit to a three-parameter logistic function was obtained by using GraphPad Prism (v. 6.01). The IC_{50} ESC, IC_{50} 3T3, and ID_{50} EB values were used to classify the compounds as non-toxic or toxic. A linear discriminant analysis (LDA) was used to identify endpoints via stepwise discriminant analysis (SPSS) (Seiler et al., 2004). LDA is a method of classifying given predictive variables into a class with the largest posterior probability. Discriminant analysis approaches assume that the conditional probability density functions $p(x|y = 1)$ and $p(x|y = 2)$ are both normally distributed with mean and covariance parameters (μ_1, Σ_1) and (μ_2, Σ_2), respectively. The LDA classifier assumes that the observations in the kth class are drawn from a multivariate Gaussian distribution $N(\mu_k, \Sigma_k)$ where μ_k is a class-specific mean vector, and Σ_k is a covariance matrix that is common to all k classes. By plugging the density function for the kth class, $f_k(X = x)$ into Bayes' theorem and performing a little bit of algebra, discriminant function,

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k{}^T\Sigma^{-1}\mu_k + \log\pi_k$$

is derived. LDA results are classified into classes in which the value of the discriminant function is largest for a new observation.

Model performance was evaluated by calculating the sensitivity, specificity, and accuracy for predicting outcomes. These metrics were calculated by using the following formulae based on Table 2:

$$\text{accuracy} = \frac{TP + TN}{TP + FN + TN + FP}$$

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

$$\text{specificity} = \frac{TN}{TN + FP}$$

We fitted an LDA model in order to predict toxicity using $\log_{10}IC_{50}(ESC)$, $\log_{10}IC_{50}(EB)$, and $\log_{10}ID_{50}(3T3)$. A toxicity result was recorded whether a chemical is toxic or non-toxic. Model building was implemented using the MASS packages in the R statistical analysis software for Windows version 3.4.2. We used a training set of 21 observations to build the LDA model. The training set consists of 15 toxic chemicals (71% or 15/21) and 6 non-toxic chemicals (29% or 6/21). It assumed that the observations in the toxic (or non-toxic) class are drawn from a multivariate Gaussian distribution, where μ_k is a class-specific mean vector, and Σ is a covariance matrix that is common. LDA was performed using the number of the training set as an option of prior probability by the R program.

The estimated linear discriminant function is as follows.

Table 2
Confusion table.

		Predict	
		Non-toxic	Toxic
original	Non-toxic	True Negative; TN	False Positive; FP
	Toxic	False Negative; FN	True Positive; TP

Table 3
Comparison of 10-day and 4-day treatments of chemical exposure.

No	Chemicals	IC_{50} mES(E14) (μM)		IC_{50} 3T3 (μM)	
		10-day	4-day	10-day	4-day
1	Sodium bicarbonate	34410	30450	39860	38470
2	Sodium gluconate	31210	38160	41970	42940
3	Saccharin	25330	21160	16720	5196
4	Penicillin G	9471	5082	5261	2048
5	Isoniazid	6282	3865	3011	2524
6	Ascorbic acid	1073	1355	140.7	812.6
7	Doxylamine succinate	317.5	499.6	275.1	421.5
8	Pravastatin	235.7	341.5	509.6	193.2
9	Caffeine	2147	1521	2294	1114
10	Aspirin	1210	3833	1618	982.9
11	Diphenhydramine	43.65	54.63	339.6	51.81
12	Diphenylhydantoin	88.83	403.2	528	304.7
13	Indomethacin	79.18	124.1	93.92	411.7
14	Dexamethasone	44.34	316.4	57.83	0.04369
15	Papaverine	26.17	54.98	134.7	30.79
16	Lovastatin	1.219	5.87	30.79	5.679
17	Verapamil-HCl	37.88	35.64	40.62	31.87
18	Methotrexate	0.0137	0.03982	0.0415	0.02424
19	D-Penicillamine	345.5	2113	3325	285.3
20	Ochratoxin A	0.6744	8.806	35.44	10.44
21	Retinoic Acid	0.0024	0.009318	0.2819	6.486

$$\begin{aligned}\text{discriminant function} = {} & 0.4852526 \times \log_{10} IC_{50}(\text{ESC}) \\ & - 0.5251880 \times \log_{10} \text{IC50(3T3)} \\ & + 0.1700371 \times \log_{10} ID_{50}(\text{EB})\end{aligned}$$

3. Results

3.1. The exposure time to chemicals

We shortened the duration of the drug treatment to shorten the experimental time. Our previously developed EBT methods need a chemical exposure condition of 10 days to measure cell viability (Kang et al., 2017). To optimize the test method for use in this study, we decreased the exposure period to 4 days. Cell viability was measured after exposure to the chemicals for both 10 days and 4 days, and there was no notable difference in IC_{50} measurements between the 10-day and 4-day results (Table 3).

3.2. Prediction model change

In this study, we created a new prediction model. For a new observation, the model calculates the score function, which is then classified by using an allocation rule; *i.e.*, if the calculated discriminant function value is ≤ -2.353739, it is classified as non-toxic, otherwise, it is classified as toxic. The confusion table and performance indicators for the training set and test data are shown in Table 4.

Table 4
Confusion table and performance indicators for training set.

		Predict	
		Non-toxic	Toxic
original	Non-toxic	5	1
	Toxic	0	15
Accuracy (%)	95.2		
Sensitivity (%)	100		
Specificity (%)	83.3		

Table 5
Intra-laboratory reproducibility test evaluation classifications obtained by applying the predictive model.

Chemicals	In vivo Class	Experimenter 1					Experimenter 2					Experimenter 3				
		Code	IC_{50} mES	IC_{50} 3T3	ID_{50} EB	Class	Code	IC_{50} mES	IC_{50} 3T3	ID_{50} EB	Class	Code	IC_{50} mES	IC_{50} 3T3	ID_{50} EB	Class
Saccharin	N	A103	26437	18019	5922	O	E331	17464	14101	22918	O	F340	23688	14447	24565	O
Penicillin G	N	A108	18232	3079	3229	O	E323	3802	2429	8097	O	F312	5751	2075	7941	O
Sodium bicarbonate	N	A109	71003	47280	11903	O	E327	17010	26913	45685	O	F343	20450	32948	43007	O
Sodium gluconate	N	A104	44508	273448	5189	O	E308	46301	37737	20565	O	F325	39951	61153	20909	O
Ascorbic acid	N	A106	1582	3190	32891	**X**	E322	1374	753	7107	**X**	F305	1576	914.6	6431	**X**
Verapamil-HCl	T	A112	69.24	120.3	3.617	O	E304	18.45	20.69	10.63	O	F334	22.56	26.53	10.36	O
Dexamethasone	T	A110	128.1	0.0085	287.7	O	E313	161.9	0.0571	137	O	F330	308.1	0	22.94	O
Diphenhydramine	T	A113	53.18	206.8	161.5	O	E319	56.47	51.37	91.6	O	F306	47.94	89.64	86.7	O
Retinoic acid	T	A101	13.8	10604	8.498	O	E310	0.0775	674	16.41	O	F328	0.1897	82.95	3.19	O
Caffeine	T	A107	1470	11514	1488	**X**	E332	1079	7093	2364	O	F303	1290	878	769.9	O
Indomethacin	T	A105	683.4	6227	956.1	**X**	E315	149.4	183.7	827.9	O	F336	153.4	408.1	343.5	O
Papaverine hydrochloride	T	A102	314	786	308.6	O	E326	35.97	29.67	31.87	O	F339	55.95	39.96	25.4	O
Doxylamine succinate salt	T	A100	526.4	466.7	290.6	O	E307	250.7	314.58	456.2	O	F317	566.6	608.6	612.9	O
D-Penicillamine	T	A114	4813	2833	61.56	**X**	E318	4691	681.6	2079	O	F320	6890	808.3	360.7	**X**
5,5 Diphenylhydantoin	T	A111	169.5	59.46	601.3	O	E314	250.9	174.2	98.9	O	F344	226.1	229.7	19.52	O

N, Non-toxic; T, Toxic; O, Match; X and Bold, Mismatch.
IC_{50} & ID_{50} (Unit; μM).

3.3. Intra-laboratory reproducibility test

Three experimenters of the lead laboratory conducted testing with a coded chemical (15 substances). Table 5 shows the mean 3T3 cell viability, mESC viability, and mEB area obtained as well as the toxicity classification derived by using the prediction model. Accuracy is expressed as the percentage of the actual toxicity status classified as toxic and the actual non-toxicity status classified as non-toxic. Sensitivity is the percentage of actual toxicity classified as toxic. Specificity is the percentage of actual non-toxicity classified as non-toxic. A high level of consistency was observed. Experimenter 1 had an accuracy of 73%, whereas experimenter 2 had an accuracy of 93% and experimenter 3 had an accuracy of 87%. The results of our statistical analysis of the intra-laboratory reproducibility test indicate that intra-laboratory accuracy, sensitivity, and specificity were 84%, 87%, and 80%, respectively.

3.4. Inter-laboratory transferability test

Experiments were conducted to confirm that the test method is applicable to other laboratories. Based on the results of previous experiments, we selected a chemical that was completely non-toxic and toxic. In order to assess the inter-laboratory transferability test, participatory laboratories 1 and 2 performed testing using the coded chemicals (6 substances). Table 6 shows the mean 3T3 cell viability, mESC viability, and mEB area obtained and the toxicity classifications derived by using the prediction model. High inter-laboratory consistency was observed. Participatory laboratory 1 had an accuracy of 100%, and participatory laboratory 2 had an accuracy of 83%. The results of the statistical analysis of inter-laboratory transferability test results indicated that accuracy, sensitivity, and specificity were 92%, 100%, and 75%, respectively.

3.5. Inter-laboratory reproducibility test

The accuracy was high in the participatory laboratory, and I judged that the test method transfer as well. There, inter-laboratory reproducibility testing, including the lead laboratory, participatory laboratory 1, and participatory laboratory 2, was performed using the coded chemicals (20 substances). In the case of the lead laboratory, the chemical tested in the inter-laboratory reproducibility test was used in the previous Intra-laboratory reproducibility test. Except for duplicate chemical, statistics were processed. Table 7 shows the mean 3T3 cell viability, mESC viability, and mEB area obtained and the chemicals' toxicity classifications derived by using the prediction model. The lead laboratory had an accuracy of 82%, while participatory laboratory 1 had an accuracy of 70%, and participatory laboratory 2 had an accuracy of 90%. The results of the statistical analysis of the inter-laboratory reproducibility test indicated that reproducibility accuracy, sensitivity, and specificity were 80%, 79%, and 83%, respectively.

4. Discussion

Toxicity assessment has been used in research and development as well as for regulatory testing for purposes such as launching a new drug or registering a new chemical substance. The purpose of developing an *in vitro* developmental toxicity test method is to reduce the number of experimental animals used when assessing developmental toxicity. *In vitro* toxicity testing is faster than animal studies and is a more ethical method because it does not use animals (Spielmann, 2005; Thompson

Table 6
Inter-laboratory transferability test evaluation classifications derived by the predictive model.

Chemicals	In vivo Class	Participation Lab 1					Participation Lab 2				
		Code	IC_{50} mES	IC_{50} 3T3	ID_{50} EB	Class	Code	IC_{50} mES	IC_{50} 3T3	ID_{50} EB	Class
Saccharin	N	C105	22791	9714	23317	O	J590	11644	7950	25096	O
Penicillin G	N	C111	4816	1670	2323	O	J585	2200	1245	6615	**X**
Verapamil-HCl	T	C110	28	16.99	12.43	O	J560	184.2	38.83	23.95	O
Dexamethasone	T	C106	106.97	2.33	140.42	O	J559	460.9	0.0878	69.05	O
Diphenhydramine	T	C108	37.08	657.26	73.3	O	J552	200.88	452.68	56.23	O
Retinoic acid	T	C109	0.3352	0.8155	0.5399	O	J524	0.0082	154.24	35.91	O

N, Non-toxic; T, Toxic; O, Match; X and Bold, Mismatch.
IC_{50} & ID_{50} (Unit; μM).

Table 7
Inter-laboratory reproducibility test chemical evaluation classification by predictive model.

Chemicals	In vivo Class	Lead Lab					Participation Lab 1					Participation Lab 2				
		Code	IC_{50} mES	IC_{50} 3T3	ID_{50} EB	Class	Code	IC_{50} mES	IC_{50} 3T3	ID_{50} EB	Class	Code	IC_{50} mES	IC_{50} 3T3	ID_{50} EB	Class
Sodium bicarbonate	N	–				–	H473	3165	338293	2951	O	G437	4377	17284	43828	O
Sodium gluconate	N	–				–	H455	2585	68992	20184	O	G417	7651	25653	22078	O
Ascorbic acid	N	–				–	H451	251	952	850	**X**	G445	274.9	661.2	962.1	**X**
Aspartame	N	I406	22358	21363	7496	O	H476	9395	9395	32052	O	G424	9840	9834	22117	O
Xylitol	N	I489	97733	135656	157936	O	H474	51075	154321	31653	O	G414	62031	163128	27604	O
Sucralose	N	I495	3194	4560	1261	O	H471	1271	1687	2734	**X**	G415	5724	2862	6368	O
Sorbitol	N	I460	67245	90410	186968	O	H456	14882	60987	29511	O	G442	28314	98260	52160	O
Caffeine	T	–				–	H454	410.3	2367	1.725	**X**	G436	265.3	1106	913.5	O
Indomethacin	T	–				–	H452	59.72	283.1	49.92	O	G418	57.18	395.4	221.6	O
Papaverine hydrochloride	T	–				–	H461	12.89	36.98	43.82	O	G433	12318	17.84	210.3	O
Doxylamine succinate salt	T	–				–	H475	251.6	556.6	814.5	O	G435	460.5	278.3	235.9	O
D-Penicillamine	T	–				–	H465	1369	2339	1543	**X**	G412	1408	212.3	3913	O
5,5 diphenylhydantoin	T	–				–	H453	306.1	2086	120.1	O	G440	41.46	488.8	286.8	O
Paroxetine hydrochloride	T	I492	7.865	6.638	3.06	O	H467	4.135	21.86	5.37	O	G434	4.53	5.75	84.43	O
Methimazole	T	I422	1024	4329	1570	**X**	H462	725.2	1826	1396	O	G441	925	1770	842	O
Letrozol	T	I449	567.5	1018	356.5	O	H472	1703	11023	1760	**X**	G416	628	1737	1513	O
Tetracycline hydrochloride	T	I415	254.7	239.3	30.48	O	H459	147.4	189.6	43.38	O	G410	138.1	319.8	39.34	O
Clomiphene citrate	T	I469	1.1	8.517	0.9321	O	H478	0.7733	4.7895	1.998	O	G413	0.6283	1.5	1.236	O
Ribavirin	T	I451	3.282	15.99	5.512	O	H464	2.21	13.48	7.66	O	G429	0.3343	12.17	88.8	O
Acetazolamide	T	I465	9813	6772	6907	**X**	H466	6065	7847	3461	**X**	G430	34236	62272	20846	**X**

N, Non-toxic; T, Toxic; O, Match; X and Bold, Mismatch.
IC_{50} & ID_{50} (Unit; μM).

et al., 2015).

The ECVAM-EST has been scientifically assessed and is widely used. To predict the developmental toxicity of the test substance in the ECVAM-EST, three endpoints are used: cytotoxic effects on stem cells, cytotoxic effects on 3T3 fibroblasts, and inhibition of ESC differentiation into beating cardiomyocytes (Seiler and Spielmann, 2011). A disadvantage of that test is that testing can take a long time to perform because it is necessary to identify the cell beating ratio after differentiation into myocardial cells. Our EBT method can determine the toxicity of a chemical in the mEB area before ESCs differentiate into cardiomyocytes. Unlike other developmental toxicity tests, the EBT takes a relatively short time. The EST has not sufficiently identified the mechanism of developmental toxicity; however, the EBT can confirm the mechanism of the developmental toxicity as it is an advanced developmental toxicity test method based on embryoid body's area (Kang et al., 2017). Moreover, the EBT can reflect the cytotoxicity of undifferentiated cells and fibroblasts and can be used to assess growth retardation and embryo mortality based on the cross-sectional area of the EBs. In addition, there is a high correlation between beating ratio and EB area (Kang et al., 2017).

In our developed EBT method, the test chemical is exposed to mES and 3T3 cells for 10 days in the same manner as in the EST. Because the chemical needed to be replaced in the middle of the test, additional time and labor were required. By changing the number of times media was changed, the period of drug exposure has been changed to 4 days, which reduces errors and costs associated with media changes. Our results show that the IC_{50} data for the 4-day and 10-day treatments are similar. Thus, the experiments in this study were conducted on mES and 3T3 cells over a 4-day treatment period. In addition to changing the chemical treatment time and to enhance the reliability of the test, both inside and outside of the laboratory, the testing protocol was modified and optimized as needed. For example, the solubility test content was added or modified and was changed to a time unit (96 h) that is not a date unit (4 days) to increase accuracy and reduce errors. In the case of conventional experiments, toxicity is divided into four classes: non-toxic, weak, moderate, and strong. To create a new prediction model for use with this EBT method, we divided toxicity into only two classes: non-toxic and toxic. The new prediction model is described in Materials and Methods. 2-step classification is more accurate than 4-step classification. The drug toxicity class may be different for each pregnancy stage, so they are classified as toxic and non-toxic.

To verify the reliability of a test method, it is necessary to determine whether the same result is obtained when the experiment is conducted by another laboratory and/or person. Thus, to identify the reliability of the EBT method, intra-laboratory reproducibility, inter-laboratory mobility, and inter-laboratory reproducibility tests were performed. For the reliability assessments in the present study, the tests were conducted using identity-blinded test substances. Statistical analysis of the intra-laboratory reproducibility test revealed an accuracy of 84%. Experimenter 1 had an accuracy of 73% and misclassified ascorbic acid, caffeine, indomethacin, and D-penicillamine. Experimenter 2 had an accuracy of 93% and misclassified ascorbic acid. Experimenter 3 had an accuracy of 87% and misclassified ascorbic acid and D-penicillamine. Intra-laboratory reproducibility test chemical classifications were derived by using the prediction model. Interestingly, non-toxic ascorbic acid was misclassified as toxic by all three experimenters. Non-toxic ascorbic acid was also misclassified as toxic in a participatory laboratory as well as at the lead laboratory. It has been reported that ascorbic acid is non-toxic *in vivo* but is toxic *in vitro* (Malhotra et al., 2016). In the case of D-penicillamine, both experimenters 1 and 2 incorrectly classified it as non-toxic. D-penicillamine has weak toxicity and only very weak action when compared with that of the L isomer (Gibbs and Walshe, 1969). Because our approach classifies toxicity at three endpoints, if one of those three results is wrong, the chemical is misclassified. Results from Table 5 and Table 7, none of the experimenters matched the ascorbic acid as a non-toxic chemical. It might be some unveiled mechanism of ascorbic acid to EB formation. In contrast, most of the other chemicals which showed inaccurate data are having weak-toxicity compare to other strong embryotoxicants. This result means we should test and validate with more chemicals to make the equation for classifying the toxicity screening.

Statistical analysis of the inter-laboratory transferability test showed an accuracy of 92%. Participatory laboratory 1 had an accuracy of 100%, whereas participatory laboratory 2 had an accuracy of 83% with

penicillin being misclassified. The accuracy of the transferability test was high, so the EBT can be considered an easy and accurate test method.

Statistical analysis of the inter-laboratory reproducibility test showed an accuracy of 80%. The lead laboratory had an accuracy of 82%, but methimazole and acetazolamide were misclassified. Participatory laboratory 1 had an accuracy of 70% with ascorbic acid, sucralose, caffeine, D-penicillamine, letrozole, and acetazolamide misclassified. Participatory laboratory 2 had an accuracy of 90% with ascorbic acid and acetazolamide misclassified. Non-toxic ascorbic acid was misclassified as toxic at the participatory laboratory as well as at the lead laboratory, and, as mentioned, ascorbic acid is toxic *in vitro* but non-toxic *in vivo* (Malhotra et al., 2016). Caffeine is a drug at the boundary between toxic and non-toxic, and the results differ among experimenters (Ahir and Pratten, 2016; Fazeli et al., 2017). Toxic acetazolamide was incorrectly classified as non-toxic at all three laboratories. Acetazolamide is a developmentally toxic drug that can cause deformity but must be treated at high concentrations to cause toxicity (Ruokoniemi et al., 2009). It may have been classified as non-toxic because it causes toxicity only at high concentrations.

For some chemical, the experimenter misclassified toxicity class. The reason is that the three values (3T3 IC_{50}, mES IC_{50}, mEB ID_{50}) are classified using the prediction model. Even if one of the three values is different, it can be misclassified. Therefore, accurate experimentation is required. If the correct experimental results are obtained, then the tests will be highly accurate.

In the present study, using developmental toxicity test method (termed EBT) applying area of embryoid bodies (EBs), the total of 26 chemicals was evaluated. The results show that inter-laboratory reproducibility accuracy was as high as 80%, indicating that an EB-area-based test can be considered a good alternative test method for predicting a chemical's developmental toxicity.

Conflicts of interest

The authors declare no conflict of interest.

Funding

This research was supported by a grant (17182MFDS487) from Ministry of Food and Drug Safety in 2017.

Transparency document

Transparency document related to this article can be found online at https://doi.org/10.1016/j.fct.2018.10.044.

References

Ahir, B.K., Pratten, M.K., 2016. The impact of caffeine on connexin expression in the embryonic chick cardiomyocyte micromass culture system. J. Appl. Toxicol. 36, 903–913.

Fazeli, W., Zappettini, S., Marguet, S.L., Grendel, J., Esclapez, M., Bernard, C., Isbrandt, D., 2017. Early-life exposure to caffeine affects the construction and activity of cortical networks in mice. Exp. Neurol. 295, 88–103.

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, 209–244.

Gibbs, K., Walshe, J.M., 1969. Interruption of the tryptophan-nicotinic acid pathway by penicillamine-induced pyridoxine deficiency in patients with Wilson's disease and in experimental animals. Ann. N. Y. Acad. Sci. 166, 158–169.

Kang, H.Y., Choi, Y.K., Jo, N.R., Lee, J.H., Ahn, C., Ahn, I.Y., Kim, T.S., Kim, K.S., Choi, K.C., Lee, J.K., Lee, S.D., Jeung, E.B., 2017. Advanced developmental toxicity test method based on embryoid body's area. Reprod. Toxicol. 72, 74–85.

Kojima, H., Hayashi, K., Sakaguchi, H., Omori, T., Otoizumi, T., Sozu, T., Kuwahara, H., Hayashi, T., Sakaguchi, M., Toyoda, A., Goto, H., Watanabe, S., Ahiko, K., Nakamura, T., Morimoto, T., 2013. Second-phase validation study of short time exposure test for assessment of eye irritation potency of chemicals. Toxicol. Vitro 27, 1855–1869.

Le Coz, F., Suzuki, N., Nagahori, H., Omori, T., Saito, K., 2015. Hand1-Luc embryonic stem cell test (Hand1-Luc EST): a novel rapid and highly reproducible in vitro test for embryotoxicity by measuring cytotoxicity and differentiation toxicity using engineered mouse ES cells. J. Toxicol. Sci. 40, 251–261.

Malhotra, S., Welling, M.N., Mantri, S.B., Desai, K., 2016. In vitro and in vivo antioxidant, cytotoxic, and anti-chronic inflammatory arthritic effect of selenium nanoparticles. J. Biomed. Mater. Res. B Appl. Biomater. 104, 993–1003.

Marx-Stoelting, P., Adriaens, E., Ahr, H.J., Bremer, S., Garthoff, B., Gelbke, H.P., Piersma, A., Pellizzer, C., Reuter, U., Rogiers, V., Schenk, B., Schwengberg, S., Seiler, A., Spielmann, H., Steemans, M., Stedman, D.B., Vanparys, P., Vericat, J.A., Verwei, M., van der Water, F., Weimer, M., Schwarz, M., 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.

Nagahori, H., Suzuki, N., Le Coz, F., Omori, T., Saito, K., 2016. Prediction of in vivo developmental toxicity by combination of Hand1-Luc embryonic stem cell test and metabolic stability test with clarification of metabolically inapplicable candidates. Toxicol. Lett. 259, 44–51.

Ohnuki, Y., Kurosawa, H., 2013. Effects of hanging drop culture conditions on embryoid body formation and neuronal cell differentiation using mouse embryonic stem cells: optimization of culture conditions for the formation of well-controlled embryoid bodies. J. Biosci. Bioeng. 115, 571–574.

Rohwedel, J., Guan, K., Hegert, C., Wobus, A.M., 2001. Embryonic stem cells as an in vitro model for mutagenicity, cytotoxicity and embryotoxicity studies: present state and future prospects. Toxicol. Vitro 15, 741–753.

Ruokoniemi, P., Tertti, R., Paalosmaa-Puusa, P., Kareranta, H., Laine, K., 2009. Acetazolamide may provoke cyclosporine toxicity-a case report. NDT Plus 2, 298–299.

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

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.

Seiler, A.E., Spielmann, H., 2011. The validated embryonic stem cell test to predict embryotoxicity in vitro. Nat. Protoc. 6, 961–978.

Spielmann, H., 2005. Predicting the risk of developmental toxicity from in vitro assays. Toxicol. Appl. Pharmacol. 207, 375–380.

Tandon, S., Jyoti, S., 2012. Embryonic stem cells: an alternative approach to developmental toxicity testing. J. Pharm. BioAllied Sci. 4, 96–100.

Thompson, P.A., Hulme, C., Nash, H.M., Gooch, D., Hayiou-Thomas, E., Snowling, M.J., 2015. Developmental dyslexia: predicting individual risk. J. Child Psychol. Psychiatry 56, 976–987.

Zhou, J., Zhang, Y., Lin, Q., Liu, Z., Wang, H., Duan, C., Wang, Y., Hao, T., Wu, K., Wang, C., 2010. Embryoid bodies formation and differentiation from mouse embryonic stem cells in collagen/Matrigel scaffolds. J Genet Genomics 37, 451–460.