Fishing for Teratogens: A Consortium Effort for a Harmonized Zebrafish Developmental Toxicology Assay

Jonathan S. Ball,1,2 Donald B. Stedman,1,2 Jedd M. Hilegass,1 Cindy X. Zhang,§ Julie Panzica-Kelly,¶ Aleasha Coburn,† Brian P. Enright,‡ Belen Tornesi,§ Hamid R. Amouzadeh,‡ Malcolm Hetheridge,* Anne-Lee Gustafson,‡,3 and Karen A. Augustine-Rauch†

1 To whom correspondence should be addressed at Brixham Environmental Laboratory, Freshwater Quarry, Brixham, Devon TQ5 8BA, United Kingdom. Fax: +44 (0)1803 882974. E-mail: jonathan.ball2@astrazeneca.com.
2 These authors contributed equally to this work.
3 Present address: Swedish Chemicals Agency, PO Box 2, SE-172 12 Sundbyberg, Sweden.

Received September 20, 2013; accepted January 28, 2014

A consortium of biopharmaceutical companies previously developed an optimized Zebrafish developmental toxicity assay (ZEDTA) where chorionated embryos were exposed to nonproprietary test compounds from 5 to 6 h post fertilization and assessed for morphological integrity at 5 days post fertilization. With the original 20 test compounds, this achieved an overall predictive value for teratogenicity of 88% of mammalian in vivo outcome [Gustafson, A. L., Stedman, D. B., Ball, J., Hilegass, J. M., Flood, A., Zhang, C. X., Panzica-Kelly, J., Cao, J., Coburn, A., Enright, B. P., et al. (2012). Interlaboratory assessment of a harmonized Zebrafish developmental toxicity assay—Progress report on phase I. Reprod. Toxicol. 33, 155–164]. In the second phase of this project, 38 proprietary pharmaceutical compounds from four consortium members were evaluated in two laboratories using the optimized method using either pond-derived or cultivated-strain wild-type Zebrafish embryos at concentrations up to 100 μM. Embryo uptake of all compounds was assessed using liquid chromatography-tandem mass spectrometry. Twenty eight of 38 compounds had a confirmed embryo uptake of >5%, and with these compounds the ZEDTA achieved an overall predictive value of 82% and 65% at the two respective laboratories. When low-uptake compounds (≤5%) were retested with logarithmic concentrations up to 1000 μM, the overall predictivity across all 38 days was 79% and 62% respectively, with the first laboratory achieving 74% sensitivity (teratogen detection) and 82% specificity (non-teratogen detection) and the second laboratory achieving 63% sensitivity (teratogen detection) and 62% specificity (non-teratogen detection). Subsequent data analyses showed that technical differences rather than strain differences were the primary contributor to interlaboratory differences in predictivity. Based on these results, the ZEDTA harmonized methodology is currently being used for compound assessment at lead optimization stage of development by 4/5 of the consortium companies.

Key words: Zebrafish; embryo; interlaboratory assessment; developmental toxicity assay; teratogenicity; teratogenicity screen; pharmaceutical compounds.

Developmental toxicants have traditionally been identified via in vivo studies in rodent and nonrodent species (usually the rabbit). Although this in vivo testing paradigm is recognized as the gold standard in terms of assessing candidate drugs for human safety (and is required by regulatory authorities), there are limitations in terms of assay throughput, test material requirements, timing, and cost. Therefore, development of in vitro and in vivo screening assays (e.g., mouse embryonic stem [ES] cell test, FETAX [Frog Embryo Teratogenesis Assay Xenopus], Zebrafish development, and rat or rabbit whole embryo culture [WEC]) to prioritize mammalian in vivo testing has been pursued by a number of groups (Augustine-Rauch et al., 2010; Ellis-Hutchings and Carney, 2010; Fort et al., 2004; Paquette et al., 2008). In general, screening assays are capable of differentiating potent teratogens from non-teratogens, but tend to encounter difficulty with compounds whose teratogenic effects are equivocal (Genschow et al., 2004). This has limited broad implementation of developmental toxicity screening assays and may be related to the fact that not all mechanisms of in vivo mammalian developmental toxicity can be readily mimicked using in vivo or in vitro screens (e.g., maternal metabolism, physiological differences, and/or the attributes of drug exposure levels at critical times of gestation; van der Laan et al., 2012). The Zebrafish developmental toxicity assay (ZEDTA) is promising due to its simplicity of maintaining breeding pairs, rapid organogenesis, and transparency of embryos. Importantly, Zebrafish contain orthologues of 86% of drug targets and 70% of human...
genes, which increases its potential as a model for early identification of teratogenicity hazard (Gunnarsson et al., 2008, Howe et al., 2013). ZEDTA models in vivo mammalian embryo-fetal development (EFD) studies in that intact embryos at the gastrulation stage are exposed to a concentration range of compound, and then allowed to develop through the remainder of organogenesis and into an equivalent of a fetal stage. Larvae at 5 days post fertilization (dpf) are assessed for overall viability and morphological integrity using a numerical score system (Panzica-Kelly et al., 2010).

A consortium of biopharmaceutical companies was formed in 2009 to develop a harmonized ZEDTA and to test the suitability of such an assay in early pharmaceutical development. In phase I of this assessment, 10 known teratogenic and 10 non-teratogenic non-proprietary compounds were assessed in a standard protocol in four different laboratories. The Zebrafish embryo source reflected assay conditions across laboratories. Following the initial testing, the overall concordance to mammalian data across laboratories was between 60 and 70%. Subsequent optimization of the assay which focused on formulation and concentration adjustments and number of replicates resulted in improved concordance to 85% (Gustafson et al., 2012).

In phase II of this ZEDTA assessment (reported within this manuscript) 38 proprietary pharmaceutical compounds (19 mammalian teratogenic and 19 mammalian non-teratogenic) were supplied by four members of the consortium. By systematically assessing the Zebrafish assay with non-proprietary and then proprietary compounds, the overall goal of this consortium was to develop a harmonized methodology as a means to determine the degree of overall concordance to in vivo mammalian data. The proprietary compounds (sourced from R&D facilities with corresponding in vivo data [Supplementary tables 1 and 2]) differ from previous interlaboratory validation efforts where test compound selection was more limited (Genschow et al., 2002). The pharmaceutical compounds included molecules that were structurally diverse with a variety of intended targets (Supplementary table 3). Full details of target and therapy area were still retained by each consortium member.

The mammalian in vivo data for each compound in this study was based upon International Conference on Harmonisation guideline on tests for reproductive toxicity (ICH S5) and dose range finding embryo-fetal developmental toxicity studies in two species. The Zebrafish studies described in this manuscript were conducted in two laboratories of which one used a pond-derived strain and the other a cultivated Zebrafish strain, both using the optimized method to assess concordance across laboratories. When neither decreased viability nor developmental effects were observed following treatment with experimental compounds, uptake analysis of the compound was conducted to confirm appropriate exposure to the test material. Subsequently, this was followed by uptake analysis of all remaining compounds.

The phase I and phase II results were assessed to determine the degree of concordance to in vivo mammalian data and also considered its utility as part of a battery of in vivo or in vitro screening assays to facilitate early-stage assessment for teratogenicity of pharmaceutical compounds under development.

**MATERIALS AND METHODS**

**Study Design**

This study evaluated a chorionated Zebrafish embryo developmental toxicity assay previously described by the consortium (Gustafson et al., 2012). Briefly, gastrulation stage embryos (5–6 h post fertilization [hpf]) were exposed to a range of concentrations of the test compound and allowed to develop through the remainder of organogenesis period and into the larval stage. At 5 dpf, the larvae were assessed for overall viability and morphological integrity using a numerical score system (Panzica-Kelly et al., 2010). Briefly, a score of 5 to 0.5 was assigned to each morphological structure/organ to delineate malformed embryos from normal embryos. The morphological endpoints included somites, notochord, tail, fins, heart, facial structures (a score that includes assessment of eye, sacculi/otoliths, and olfactory region), neural tube, and pharyngeal arches/jaw. A score of 5 was assigned to represent normal morphology, 4 was assigned to represent a slight variation from normal, 3 was assigned to represent a mild abnormality, 2 was assigned to represent a moderate malformation, 1 was assigned to represent a severe malformation, and a score of 0.5 was assigned to represent missing structures (Panzica-Kelly et al., 2010). Thirty eight anonymous proprietary blinded compounds were run at two laboratories [Laboratories B and D; using laboratory nomenclature from Gustafson et al. (2012)] with pond-derived and proprietary, cultivated Zebrafish strain wild types. Positive and negative controls were included in each experiment (saccharin and all-trans-retinoic acid, respectively). Results were considered acceptable for inclusion in the interlaboratory assessment if the following criteria were met:

1. Negative control treatments produced background malformation rates that were ≤ historical control background malformation rates (≤ 16% or ≤ 2 affected embryo in negative control treatment groups from 12 embryos).
2. The lethality rate for media and DMSO controls was ≤ 10%.
3. All experimental and formulation procedures were in alignment with the optimized protocol.

Experiments that did not meet the above criteria were repeated. Embryos exposed to the positive control at 0.01 μM responded with a very consistent pattern of malformations with most if not all endpoints affected. So for phase II we only indicated whether each embryo was affected or not affected with the RA phenotype, as the main concern was for controlling for poor clutches as indicated by the media, solvent, and saccharine controls.
**Compound Test Set Selection**

The 38 proprietary pharmaceutical compounds (19 teratogens and 19 non-teratogens) were provided by 4 of the pharmaceutical companies in the consortium. All of the compounds were previously assessed in EFD and/or dose range finder (DRF) EFD studies *in vivo* in at least two mammalian species (Supplementary tables 1 and 2 detail maternal toxicity, fetal toxicity, and fetal malformations). Compounds that were clearly teratogenic in either the DRF EFD studies or where an unequivocal signal was present in one EFD study were only assessed in one species. A compound was considered a teratogen if it induced visceral or skeletal structural abnormalities in the fetus at doses that caused no or only minimal to mild (maternal) toxicity. A compound was considered a non-teratogen if it was reported not to cause malformations in either of the two species evaluated. Compound 22 was accepted as an exception as it is an active ingredient of a prenatal vitamin supplement which was assumed to be negative. The test set of 38 proprietary compounds had an equal distribution of teratogens and non-teratogens. This equal distribution was a means of balancing the positive and negative predictive values of the ZEDTA. The true distribution of teratogenic to non-teratogenic compounds in a drug discovery portfolio is not known. The compounds were structurally divergent and the intended human target was present in the Zebrafish.

Animal care and egg production. Two laboratories (laboratories B and D, nomenclature from Gustafson et al., 2012) participated in the assessment of the compounds for teratogenicity in this phase II study. Laboratory B used wild-type pond fish (Segrest Farms, Gibsonton, FL) whereas Laboratory D used a proprietary-cultivated Zebrafish strain generated internally and originally obtained from Carolina Biological Supply Company (Burlington, NC). Uptake assessment of the 38 compounds was carried out at Laboratory C using a cultured WIK strain (Wild-type India Kolkata). All three laboratories showed strong alignment in phase I of the consortium effort. Due to timing and potential import restrictions, the uptake was conducted on the strain previously tested in phase I which was present at Laboratory C.

Breeding stocks of healthy, unexposed mature adult Zebrafish were used for the production of fertilized eggs. Spawning Zebrafish were maintained in a recirculating husbandry system that varied by laboratory (Aquaneering, San Diego, CA: Laboratory B; Tecniplast, Germany: Laboratory C; and Pharmaceutical Research Laboratory, Nuagatuck, CT: Laboratory D). Water was maintained at approximately pH 7.35 (± 0.65) and 28 (± 1) °C. Fish were cultured in the aquarium facility with 14 h light: 10 h dark light cycle. Spawning and egg collection procedures were conducted as described previously by Gustafson et al. (2012).

**Experimental procedures.** All procedures for the use of the optimized protocol including the culture of media and reagent, compound administration, viability assessment, and morphological scoring are described in detail in Supplemental Materials I, Gustafson et al. (2012), and an optimized protocol reprinted here (Supplemental Materials I optimized protocol for Ring II 9[1].1.11). Briefly, 5 hpf embryos were exposed to test compounds at 0.1, 1, 10, and 100µM, positive control (retinoic acid) at 0.01µM, negative control (saccharin) at 100µM, and solvent (DMSO at 0.5% v/v) and media controls. Details of solubility, pH, and embryo clutch information were also recorded. Exposed embryos (one per well in Corning Costar 24-well plates [3527], USA) were then incubated at 28 (± 1) °C without solution change until assessment at 5 dpf in facility with 14 h light: 10 h dark light cycle.

**Teratogenic Classification Procedures**

A “teratogenic index,” defined as the LC25/No Adverse Effect Level (NOAEL) ratio, was used to obtain a teratogenic classification of a compound. In earlier studies, LC25 of 5 dpf Zebrafish was selected as the optimum general toxicity value for the teratogenic index and the compound’s half maximal inhibitory concentration (IC50) value in NIH3T3 cells, a fibroblast cell line used as surrogate cell line for adult toxicity in developmental toxicology assays validated by European Union Reference Laboratory for alternatives to animal testing (EC-VAM) (Buesen et al., 2004; LC25 studies reviewed in Brannen et al., 2010 Supplemental materials). Subsequent work demonstrated that the LC25 for viability in Zebrafish was sufficient for supporting the teratogenic index so the cytotoxicity evaluations in the NIH3T3 surrogate cell line were not incorporated in the final teratogenic index supporting the dechorinated Zebrafish assay (Brannen et al., 2010). Likewise, the LC25 represents the concentration of test compound capable of causing 25% lethality in 5 dpf embryos treated with a concentration range of compound and determined by curve fitting analysis (detailed in Supplemental Materials Ia in Gustafson et al., 2012). An LC25/NOAEL ratio ≥ 10 was considered a positive classification for *in vivo* teratogenic potential, whereas a ratio <10 was considered a negative classification. This ratio was identified as optimal for the correct classification of teratogenic potential of 31 definitive teratogens and non-teratogens drawn from a range of industrial and consumer chemicals, pharmaceuticals, biocides, and vitamins (Brannen et al., 2010). Determination of the Lowest Adverse Effect Level (LOAEL) and NOAEL involved assessing the concentration-response relationship of adverse morphological effects using the morphological score system described by Panzica-Kelly et al. (2010). Consideration was given to the fact that some variation in morphology may occur in the controls. Therefore, to be considered compound-related, an adverse effect must exhibit both a
found that all compounds that caused 100% lethality through the 0.1 μM concentration were in vivo teratogens (Brannen et al., 2010, Gustafson et al., 2012, Ball and Augustine-Rauch, personal communication). In addition, none of the false positive (FP) compounds displayed this profile. However, it is possible that other compounds (perhaps reactive intermediates, industrial and consumer chemicals, or chemical agents used as pesticides and biocides, etc.) may have different profiles than pharmaceuticals and may not fit the criteria used in the present study. In cases where a compound produced a steep toxicity curve that led to limited concentrations to assess NOAEL, a subsequent experiment was performed with a refined dose range (see Supplemental Materials I optimized protocol for Ring II 9[1].1.11 reprinted from Gustafson et al. (2012)).

Compounds classified as non-teratogenic and which demonstrated less than 5% uptake were retested [using methods described above with the same compound administration procedures as described in Gustafson et al. (2012)] at a higher concentration range (100, 500, and 1000 μM) to maximize the exposure of embryos to the test compounds and determine the final teratogenic classification. It should be emphasized that the assay design and prediction model described in this paper has evolved during work with pharmaceutical small molecules, and before adapting these procedures to other types of compounds (industrial and consumer chemicals or chemical agents used as pesticides and biocides, etc.) one first needs to assess the performance of a larger repertoire of these chemicals in this assay.

Compound uptake measurement in embryos. Following score assessment and before decoding, the compounds were assessed for uptake analysis by liquid chromatography-tandem mass spectrometry method as described in Gustafson et al. (2012). Uptake chemical analysis was performed on extracted effluent from ~30 hpf embryos which had been exposed to the test compound for 24 h. Uptake was assessed at 1, 10, and 100 μM in embryo media containing 0.5% DMSO for all compounds with the exception of compound numbers 19, 20, 27, and 38 which were assessed at 0.1, 1, and 10 μM due to previously demonstrated overt toxicity. Additionally, compound numbers 2 and 31 were only assessed at 1 and 10 μM due to unexpected overt toxicity to the embryos at 100 μM.

Compound uptake into the embryo was expressed as a percentage of nominal exposure concentration. This value was determined against the standard curve for each individual compound. The amount of compound present within a defined amount of solution (extracted from a defined number of embryos) was calculated against the calibration curve. This gave a concentration as a percentage of the test solution. As a general rule, results are expressed as high (>100%), medium (5–100%), or low (<5%) uptake relative to the test solution concentration.

**Assay Performance Analysis**

An assay’s performance was evaluated by generating contingency tables for phase I and phase II of the project. True positive (TP) and true negative (TN) compounds were compounds that had in vitro Zebrafish teratogenic classification that agreed with the in vivo mammalian data. FP and false negative (FN) were compounds that had in vitro teratogenic classification that did not agree with in vivo mammalian data. The analyses included determining the following endpoints:

- Sensitivity for detecting teratogens: TP/(TP + FN) × 100.
- Specificity for detecting non-teratogens: TN/(TN + FP) × 100.
- Positive predictive value (PPV) = sensitivity for detecting teratogens: TP/(TP + FP) × 100.
- Negative predictive value (NPV) = specificity for detecting non-teratogens: TN/(TN + FN) × 100.
- Overall predictivity: ((PPV + NPV)/2).
- Overall concordance: total number of compounds correctly classified/total number of compounds × 100.

The assay’s overall performance was evaluated in relation to the positive predictive value and the negative predictive value for correct classification of teratogens and non-teratogens, respectively.

**Data Analysis**

Prior to statistical analysis, the control average value dataset was tested for normality (Anderson-Darling test) and homogeneity of variance (Bartlett’s or Levene’s tests) using Minitab. Nonparabolic data was tested using a Kruskal-Wallis test and compared using a Mann-Whitney test. Statistical significance between groups was indicated by a significance level of at least p < 0.05.
RESULTS

Interlaboratory Control Assessment of Background Abnormalities and Lethality

The negative control (DMSO, saccharine, and untreated) score values at both testing laboratories were comparable. The group average morphological score values of the controls over >50 experiments ranged between 4.0 and 5.0 for both laboratories, with exception of one experiment at Laboratory D (Fig. 1). Overall average lethality in negative control groups was ≤ 1.15% for both laboratories. The majority of the 38 compounds had a LC$_{25}$ value of greater than 100 μM (Fig. 2).

Phase II Results—Comparison Between Laboratories

Laboratory B returned classifications for 38 compounds whereas Laboratory D returned classifications for 37 compounds (Table 1 and Supplementary table 4). The results for each compound from both laboratories were classified as TP, TN, FP, or FN in relation to the mammalian classification. Cross-comparison of results between the two laboratories indicated that there was 76% agreement in the teratogenic classification for each of the compounds; however, overall concordance varied between the two laboratories. Laboratory B had an overall concordance of 71% to mammalian in vivo teratogenic outcome. This included 16 of the 19 TN compounds and 11 of the 19 TP compounds (Tables 1 and 2). Laboratory D had an overall concordance of 57% for in vivo teratogenic outcome. This included 13 of the 19 TN compounds and 9 of the 18 TP compounds (Tables 1 and 2).

Uptake analysis. Uptake of compounds into embryos was measured for all 38 compounds (Table 1). Twenty six of the 38 compounds demonstrated a presence within the embryo at or above 5%; seven demonstrated uptake above the exposure concentration (>100%) up to a maximum of over 20,000% (200× the exposure concentration). Of the remaining 12 compounds, one showed a very high background level and the uptake in the embryo could not be determined; one could not be assessed for uptake because there was no appropriate ion for analysis (no ion detectable [NID]); the remaining 10 compounds were defined as low-uptake and showed either a low percentage uptake (≤ 5%; n = 5) or the compound could not be detected (ND; n = 5) within the embryo after 24 h of exposure (Table 1). Of the 11 low-uptake compounds (ND, NID, and ≤5%), four demonstrated morphologic abnormalities after a 5-day exposure and were classified as teratogenic by one or both laboratories. There was no observed grouping between the mammalian teratogenic classification and the uptake observed in the Zebrafish (Fig. 3). Excluding the compounds which had either low uptake or were not detectable in the embryo after 24 h of exposure, the predictivity and overall concordance of the assays in relation to mammalian in vivo classification conducted at laboratories B and D increased to 82% and 65%, respectively (Table 2).

Assay Performance

Five compounds (four teratogens and one non-teratogen [mammalian classification]) that presented no teratogenic signal and with <5% uptake were anonymized a second time and retested at a higher concentration range (100, 500, and 1000 μM) in order to maximize uptake and determine the added value of inclusion of higher test concentrations. Previously, the 100 μM concentration was designated as a top concentration in order to maintain reasonable compound requirement and due to frequent precipitation and pH issues encountered when at or exceeding this concentration. Both laboratories correctly reclassified one compound as a teratogen and the others were classified as being non-teratogenic (Table 1). Inclusion of the high-concentration experimental data improved the overall concordance to 74% and 60% for laboratories B and D, respectively (Table 2). Laboratory B demonstrated improved sensitivity to 63% although the specificity was not affected. Laboratory D also showed an increase in sensitivity and no effect on the specificity (Table 2).

Because of limited compound supply, four additional low-uptake compounds classified as non-teratogenic in the first study were retested at a separate laboratory [at 100, 500, and 1000 μM; Laboratory A using nomenclature from Gustafson et al. (2012)]. These compounds included one teratogen and three non-teratogens (footnoted on Table 1). All of these compounds were classified as non-teratogenic, although the teratogenic compound exhibited signs of growth delay (yolk ball retention) at the highest concentration.

Further analysis of the high-concentration results suggested the 100, 500, and 1000 μM range may potentially limit sensitivity due to the reduced chance for LC$_{25}$/NOAEL ratios to achieve ≥ 10. The entire test set, including the original 0.1–100 μM experiments, were re-evaluated using logarithmic concentrations (0.1, 1, 10, 100, and 1000 μM). With this approach, overall concordance to mammalian in vivo classification of Laboratory B improved to 79% with 74% sensitivity and 84% specificity (Table 2). Overall concordance of Laboratory D improved to 62% with 63% sensitivity and 62% specificity (Table 2).

As a means to determine whether strain differences in the LC$_{25}$ values were a cause of the disparity between the laboratories, Laboratory D was supplied with pond-derived breeder fish from Laboratory B. The two strains were assessed simultaneously for both LC$_{25}$ value and morphological change for the eight mismatched compounds. There were no observed Zebrafish strain differences when tested at this one facility (data not shown).

DISCUSSION

The finalized ZEDTA assay incorporating replicate standards, logarithmic dose range (0.1–100 μM) and compound uptake assessment, demonstrated optimum concordance for the evaluation of a pharmaceutical compound’s teratogenicity potential in relation to mammalian in vivo assessment. Higher...
FIG. 1. Consistency of control data between Laboratory B and Laboratory D. The averaged score values of the DMSO (square), saccharin (circle), and untreated (triangle) controls are plotted for both Laboratory B (light gray) and Laboratory D (dark gray) from all experiments run for the assessment of the proprietary compound experimentation (Laboratory B: \( N = 31 \), Laboratory D: \( N = 45 \)). Plots of the different average values for each control generated from the values scored for arches/jaw, face, fin, heart, neural tube, notochord, somite, and tail from each control. Plots are stuttered to facilitate visualization. The mean score values for all DMSO, saccharin, and untreated for Laboratory B and Laboratory D for each control’s treatment ranged between 4.88 and 4.95, 4.91 and 4.93, and 4.91 and 4.95, respectively. The average score values ranged between 5 and 4.54 for all control treatments across both laboratories.

FIG. 2. Comparison of lethality of the 38 proprietary compounds between Laboratory B and Laboratory D. Proprietary compounds identified as either teratogen (dark gray circle) or non-teratogen (light gray square) by mammalian classification. The LC\(_{25}\) value averaged over the independent replicates (\( N = 2–4 \)) for each compound in each laboratory. Black dashed box indicates area of greatest difference between Laboratory B and Laboratory D; this included the non-teratogenic compounds numbered 3 and 4 and the teratogenic compounds numbered 16, 17, and 26. Twenty seven compounds (5 teratogens and 22 non-teratogens) demonstrated a LC\(_{25}\) value >100 \( \mu \)M in both laboratories.
dosing concentrations could be employed with compounds that demonstrated no lethality, morphological malformations, and low uptake (<5%).

Overall, the results showed that the ZEDTA had an overall concordance with mammalian in vivo classification of 65% (Laboratory D) and 82% (Laboratory B) when compound uptake of >5% could be demonstrated (Table 2). In addition, when considering the logarithmic concentration, analysis of the entire test set achieved a 62% (Laboratory D) and 79% (Laboratory B) concordance. Two limiting aspects of this Zebrafish developmental toxicology assay include compound solubility limitations in the aqueous-based culture media and low uptake associated with some compounds. The results of uptake analyses indicate that uptake or lack of it affects the predictivity of the assay.

### Table 1: Teratogenic Classifications of Test Set Between Laboratories and Compound Uptake

| Compound | In vivo mammalian classification | Laboratory B prediction | Classification | Laboratory D prediction | Classification | Uptake (%) | Laboratory B prediction | Laboratory D prediction |
|----------|---------------------------------|-------------------------|----------------|-------------------------|----------------|------------|-------------------------|-------------------------|
| 2        | Teratogen                       | Non-teratogen           | FN             | Non-teratogen           | FN             | 14         | Non-teratogen           | Non-teratogen           |
| 5        | Teratogen                       | Non-teratogen           | FN             | Non-teratogen           | FN             | 3          | Non-teratogen           | Non-teratogen           |
| 8        | Teratogen                       | Non-teratogen           | FN             | Non-teratogen           | FN             | 4.8        | Teratogen               | Teratogen               |
| 9        | Teratogen                       | Non-teratogen           | FN             | Non-teratogen           | FN             | ND         | Non-teratogen           | Non-teratogen           |
| 10       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | 20.9       | Non-teratogen           | Non-teratogen           |
| 12       | Teratogen                       | Non-teratogen           | FN             | Non-teratogen           | FN             | ND         | Non-teratogen           | Non-teratogen           |
| 14       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | ND         | Non-teratogen           | Non-teratogen           |
| 15       | Teratogen                       | Non-teratogen           | FN             | Non-teratogen           | FN             | 3.2        | Non-teratogen           | Non-teratogen           |
| 16       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | 20574      | Non-teratogen           | Non-teratogen           |
| 17       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | 11         | Non-teratogen           | Non-teratogen           |
| 20       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | HB         | Non-teratogen           | Non-teratogen           |
| 26       | Teratogen                       | Non-teratogen           | TP             | Not returned            | N/A            | 305        | Non-teratogen           | Non-teratogen           |
| 27       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | 1136       | Non-teratogen           | Non-teratogen           |
| 31       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | 114        | Non-teratogen           | Non-teratogen           |
| 32       | Teratogen                       | Non-teratogen           | FN             | Non-teratogen           | FN             | 15.1       | Non-teratogen           | Non-teratogen           |
| 33       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | 141        | Non-teratogen           | Non-teratogen           |
| 35       | Teratogen                       | Teratogen               | TP             | Teratogen               | TP             | 67         | Non-teratogen           | Non-teratogen           |
| 37       | Teratogen                       | Non-teratogen           | FN             | Non-teratogen           | FN             | 37.5       | Non-teratogen           | Non-teratogen           |
| 38       | Teratogen                       | Teratogen               | TP             | Non-teratogen           | FN             | 378        | Non-teratogen           | Non-teratogen           |
| 1        | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 8.4        | Non-teratogen           | Non-teratogen           |
| 3        | Non-teratogen                   | Teratogen               | FP             | Teratogen               | FP             | <6         | Non-teratogen           | Non-teratogen           |
| 4        | Non-teratogen                   | Teratogen               | FP             | Non-teratogen           | TN             | 363        | Non-teratogen           | Non-teratogen           |
| 6        | Non-teratogen                   | Non-teratogen           | TN             | Teratogen               | FP             | 23.4       | Non-teratogen           | Non-teratogen           |
| 7        | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 5          | Non-teratogen           | Non-teratogen           |
| 11       | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 7.7        | Non-teratogen           | Non-teratogen           |
| 13       | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 703        | Non-teratogen           | Non-teratogen           |
| 18       | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 21         | Non-teratogen           | Non-teratogen           |
| 19       | Non-teratogen                   | Non-teratogen           | TN             | Teratogen               | FP             | 61         | Non-teratogen           | Non-teratogen           |
| 21       | Non-teratogen                   | Non-teratogen           | TN             | Teratogen               | FP             | 77         | Non-teratogen           | Non-teratogen           |
| 22       | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | ND         | Non-teratogen           | Non-teratogen           |
| 23       | Non-teratogen                   | Teratogen               | FP             | Non-teratogen           | TN             | 6.3        | Non-teratogen           | Non-teratogen           |
| 24       | Non-teratogen                   | Teratogen               | TN             | Non-teratogen           | TN             | ND         | Non-teratogen           | Non-teratogen           |
| 25       | Non-teratogen                   | Non-teratogen           | TN             | Teratogen               | FP             | ND         | Non-teratogen           | Non-teratogen           |
| 28       | Non-teratogen                   | Non-teratogen           | TN             | Teratogen               | FP             | 12         | Non-teratogen           | Non-teratogen           |
| 29       | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 6.5        | Non-teratogen           | Non-teratogen           |
| 30       | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 7.6        | Non-teratogen           | Non-teratogen           |
| 34       | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 7.1        | Non-teratogen           | Non-teratogen           |
| 36       | Non-teratogen                   | Non-teratogen           | TN             | Non-teratogen           | TN             | 14.5       | Non-teratogen           | Non-teratogen           |

Note. ND: No compound detected with uptake analysis. NID: No ion detectable for analysis. HB: High background in detection readings caused uptake to not be accurately determined. N/A: not applicable.

*Due to limited compound supply, compounds were retested at the high-concentration range at a separate laboratory (Laboratory A) and all were classified as non-teratogenic.

**During phase I of this consortium effort, Laboratory B tested the optimized protocol in a training set of 10 teratogens and 10 non-teratogens that achieved 85% overall concordance (Gustafson et al., 2012). Using this larger test set of 38 compounds as a form of cross-validation against the small non-proprietary training set, the results suggested a good stability between the two sets even though the test set was almost twice as large as the training set but still produced a concordance within 10% of the original training set.**
TABLE 2
Summary of Assay Performance

| Assessment of assay result                                      | Non-teratogens | Teratogens | Sensitivity$^a$ | Specificity$^a$ | Positive predictive value$^a$ | Negative predictive value$^a$ | Overall predictive value$^a$ | Overall concordance$^a$ |
|----------------------------------------------------------------|---------------|------------|-----------------|-----------------|-------------------------------|-------------------------------|----------------------------|------------------------|
| All compounds (Laboratory B)                                   | 19            | 19         | 58              | 84              | 79                            | 67                            | 73                        | 71                     |
| All compounds (Laboratory D)                                   | 19            | 18         | 44              | 68              | 57                            | 57                            | 57                        | 57                     |
| Excluding low-uptake compounds (Laboratory B)                 | 14            | 13         | 77              | 86              | 83                            | 80                            | 82                        | 82                     |
| Excluding low-uptake compounds (Laboratory D)                 | 14            | 12         | 58              | 71              | 64                            | 67                            | 65                        | 65                     |
| Including low-uptake compounds tested at higher concentrations (Laboratory B) | 19           | 19         | 63              | 84              | 80                            | 70                            | 75                        | 74                     |
| Including low-uptake compounds tested at higher concentrations (Laboratory D) | 19           | 18         | 50              | 68              | 60                            | 59                            | 60                        | 60                     |
| Logarithmic concentration range including higher concentrations (Laboratory B) | 19           | 19         | 74              | 84              | 82                            | 76                            | 79                        | 79                     |
| Logarithmic concentration range including higher concentrations (Laboratory D) | 19           | 18         | 63              | 62              | 56                            | 68                            | 62                        | 62                     |

$^a$Values are expressed as percentage.

FIG. 3. Percentage uptake of proprietary compounds into developing Zebrafish embryos after 24-h exposure. Proprietary compounds identified as either teratogen (dark gray circle) or non-teratogen (light gray square) by mammalian classification. Uptake percentage plotted on log scale and plots are stuttered to facilitate visualization.

Large variation between the observed LC$_{25}$ averages between the two laboratories (Fig. 2). Of the remaining compounds, when the lethality and scores were compared, the data were comparable for the most part, with only slight differences in lethality or mild malformation rates. Such differences may be due to Zebrafish strain and/or technical variations between laboratories. An assessment of the mismatched compounds using both strains at Laboratory D showed no differences in both the LC$_{25}$ and the percentage of larvae displaying normal morphology (data not shown). This indicates a potential attribute of the assay in that it can achieve repeatable results using distinct Zebrafish strains. Some of the differences may be related to slight formulation variations at one laboratory and not at another. An example of this is that four of the compounds had differences in observed precipitation between the labs on certain experiments, where slight differences in media temperature or pH might influence solubility. Differences between the laboratories may also be related to biological variations (such as the quality of the clutches used in the assessment) or variations in how individual technicians assign scores, particularly when delineating an anomaly.
(score of 4) from a mild malformation (score of 3). Such variations in score assignment could potentially shift assignment of NOAEL to a different concentration. These slight differences can influence the LC25/NOAEL ratios to either a positive or negative teratogenic classification. In the future, simplification of the score system may reduce interlab variability, although it remains to be determined whether such a change would have an adverse effect on overall concordance. As a means to have a comprehensive dataset for analysis of the 38 compounds, we maintained the original 0.5–5 score system as in the training set. Both the training and test sets employed in the assessment of the ZEDTA assay were equally balanced between teratogenic and non-teratogenic effects observed using the mammalian in vivo assessment. Due to the late stage positioning of mammalian in vivo teratogenicity assessment, the real incidence of teratogenic compounds during drug development stages is not known. Loong (2003) demonstrated that depending on the incidence of an event (teratogenicity) and using the specificity and sensitivity values, the positive and negative predictive values of an assay can change. Using the specificity and sensitivity of Laboratory B (including low-uptake compounds tested at higher concentrations 63% and 84%, respectively, Table 2) with a hypothesized incidence of teratogenicity at 10%, the negative predictive value increased to 95% (from 70%) and the positive predictive value decreased to 31% (from 70%). These alterations indicated that you would be 95% sure of a negative result but only 30% sure of a positive result if the incidence was 10% compared with an incidence of 50%. These alterations demonstrate the importance of a balanced training set and an understanding of the baseline incidence of the effect or endpoint you are assessing for.

Overall, this Zebrafish assay presents comparable sensitivity to other in vitro assays. For instance, the standard battery of assays used to test genotoxicity of a range of chemicals for identification of carcinogens ranges from 59% (Ames test) to 79% (Micronucleus test; Kirkland et al., 2005). Additionally, in vivo EFD studies also present similar sensitivity to identifying human teratogens, where the standard regulatory test species correctly identify about 61% (rat) or 41% (rabbit) human teratogens (Bailey et al., 2005). In both cases, multiple assays are utilized to increase sensitivity in correct identification of carcinogens and teratogens. As such, combined use of rat and rabbit teratogenicity studies correctly identified about 74% of human teratogens (Supplementary table 5).

Currently, other in vitro assays may provide useful alternatives to assess compounds with solubility or uptake limitations. For instance, the streamlined WEC assay (Zhang et al., 2012) has similar predictivity to the more extensive (ECVAM) assay (~80%). That assay uses a prediction model algorithm using data collected from only one test concentration (1μM), at which precipitation is rarely encountered. In addition, the culture media contain 70% serum that can better represent the compounds’ protein binding properties encountered in vivo and facilitate both solubility and delivery to the embryo (Augustine-Rauch et al., 2004; Kratz and Beyer, 1998). However, in some cases, certain compounds may not be stable in the presence of serum or become entirely protein-bound, leading to inhibited delivery and/or pharmacological activity in the embryo. In such cases, the ZEDTA is advantageous because of it being a serum-free assay.

The Zebrafish assay may present its greatest strengths when it is applied in conjunction with other assays, where a final teratogenic classification is made following assessment of multiple assay data. Various in vitro developmental toxicity assays are available and may be suitable for assessing compounds for specific types of teratogenic liability or to be used routinely in a platform. Details associated with these assays are summarized in Supplementary table 6 (Augustine-Rauch et al., 2010, Chapin et al. 2008 and Oberemm 2000). This Zebrafish assay could be compared with those achieved from other assays with a final classification based on a consideration of all results, with risk of teratogenic liability set according to the degree of teratogenic signal observed across multiple assays, which may help to rank different compounds from the same series.

The Environmental Protection Agency’s ToxCast effort (for the identification and prioritization of potential toxic chemicals) uses such an approach where multiple assay data are integrated into a computational systems biology approach for making classifications of various organ-based toxicity and teratogenic liabilities of chemicals and environmental agents (Dix et al., 2007). Additionally, frequency analysis of teratogenic classifications ascertained from multiple developmental toxicology assay data has been found to enhance predictivity, and assay tiering may have potential to improve results, turnaround, and overall predictivity (Stedman and Augustine-Rauch, personal communications; Stedman and Augustine-Rauch, unpublished results). For example, if uptake is not adequate, an additional higher concentration (1000μM) could be included and if the final result presents no adverse effects on viability or morphology, other assays may be included to enhance predictivity in teratogenic classification; these could include rat WEC and/or ES cell culture assessment (Panzica-Kelly et al., 2013; Seiler et al., 2004; van der Laan et al., 2012; Zhang et al., 2012).

The ZEDTA’s culture work would benefit from automation, particularly image analysis; future efforts may also simplify the assay score assignments and applying statistical algorithms requiring fewer morphologic endpoints or reduced test concentrations. The current assay design achieved good predictivity and concordance. The assay has promising utility either as a stand-alone assay or in combination with other in vitro developmental toxicity assays such as ES cell assays and rat and/or rabbit embryo culture to enhance overall predictivity of teratogenic liability for pharmaceutical compounds.

‘Supplemental Materials I optimized protocol for Ring II 9[1].1.11’—Reprinted from Reproductive Toxicology 33, Gustafson, A-L., Stedman, D. B., Ball, J., Hillegass, J., Flood, A., Panzica-Kelly, J., Zhang, C., Coburn, A., Enright, B., Amouzadeh, H., Tornesi, M. B., Hetheridge, M. J., and Augustine-Rauch, K. A. Inter-Laboratory Assessment of
a Harmonized Zebrafish Developmental Toxicology Assay—Progress Report on Phase I, (155–164), copyright (2012) with permission from Elsevier.

SUPPLEMENTARY DATA

Supplementary data are available online at http://toxsci.oxfordjournals.org/.

ACKNOWLEDGMENTS

We would like to acknowledge the steadfast support of Hugo M. Vargas (Amgen) on this project as well as his critical review of this manuscript. We would also like to acknowledge Jane Stewart (AstraZeneca) for her critical review of the manuscript. The following statements are provided in compliance with the International Committee of Medical Journal Editors policy on Conflict of Interest:

1. Co-author Jedd M. Hillegass, Ph.D., conducted scientific work and experimentation with respect to the Manuscript while in the employ of Lampire Biological Labs, Inc. ("Lampire").
2. Dr Hillegass continues in the employ of Lampire as of the date herein above.
3. Lampire received remuneration for performance of the work described in the Manuscript.
4. Lampire is organized as a "for profit" corporation, and actively markets the assay and techniques described in the Manuscript to third parties as a "fee-for-service" contract research organization.

REFERENCES

Augustine-Rauch, K. A., Zhang, C. X., and Panzica-Kelly, J. (2010). In vitro developmental toxicity assays: A review of the state of the science of rodent and zebrafish whole embryo culture and embryonic stem cell assays. Birth Def. Res. Part C: Embryo Today, 89, 87–98.

Augustine-Rauch, K. A., Zhang, Q., Kleinman, M., Lawton, R., and Welsh, M. J. (2004). A study of vehicles for dosing rodent whole embryo culture with non aqueous soluble compounds. Reprod. Toxicol. 18, 391–398.

Bailey, J., Knight, A., and Balcombe, J. (2005). The future of teratology research is in vitro. Biogenic Amines 19, 97–145.

Brannen, K. C., Panzica-Kelly, J. M., Danberry, T. L., and Augustine-Rauch, K. A. (2010). Development of a zebrafish embryo teratogenicity assay and quantitative prediction model. Birth Def. Res. B Dev. Reprod. Toxicol. 89, 66–77.

Buesen, R., Visan, A., Genschow, E., Slawik, B., Spielmann, H., and Seiler, A. (2004). Trends in improving the embryonic stem cell test (EST): An overview. Altex 21, 15–22.

Chapin, R., Augustine-Rauch, K., Beyer, B., Daston, G., Finnell, R., Flynn, T., Hunter, S., Mirkes, P., O’Shea, K. S., Piersma, A., et al. (2008). State of the art in developmental toxicity screening methods and a way forward: A meeting report addressing embryonic stem cells, whole embryo culture, and zebrafish. Birth Def. Res. (Part B) 83, 446–456.

Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W., and Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol. Sci. 95, 5–12.

Ellis-Hutchings, R. G., and Carney, E. W. (2010). Whole embryo culture: A “New” technique that enabled decades of mechanistic discoveries. Birth Def. Res. B Dev. Reprod. Toxicol. 89, 304–312.

Fort, D. J., Rogers, R. L., Thomas, J. H., Buzzard, B. O., Noll, A. M., and Spaulding, C. D. (2004). Comparative sensitivity of Xenopus tropicalis and Xenopus laevis as test species for the FETAX model. J. Appl. Toxicol. 24, 443–457.

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.

Gustafson, L., Jauhiainen, A., Kristiannson, E., Nerman, O., and Larsson, J. (2008). Evolutionary conservation of human drug targets in organisms used for environmental risk assessments. Environ. Sci. Technol. 42, 5807–5813.

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498–503.

Kirkland, D., Aardema, M., Henderson, L., and Muller, L. (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non carcinogens. I. Sensitivity specificity and relative predictivity. Mutat. Res. 1-2, 1–256.

Kratz, F., and Beyer, U. (1998) Serum proteins as drug carriers of anticancer agents: A review. Drug Deliv. 5, 281–299.

van der Laan, J. W., Chapin, R. E., Haenen, B., Jacobs, A. C., and Piersma, A. (2012). Testing strategies for embryo-fetal toxicity of human pharmaceuticals. Animal models vs. in vitro approaches: A workshop report. Regul. Toxicol. Pharmacol. 63, 115–123.

Loong, T-W. (2003). Understanding sensitivity and specificity with the right side of the brain. BMJ 327, 716–719.

Oberemm, A. (2000) The Use of a Refined Zebrafish Embryo Bioassay for the Assessment of Aquatic Toxicity. Lab Anim. 29, 32–40.

Panzica-Kelly, J., Brannen, K. C., Ma, Y., Zhang, C. X., Flint, O. P., Lehmann-Metzner, L., and Augustine-Rauch, K. A. (2013). Establishment of a molecular embryonic stem cell developmental toxicity assay. Toxicol. Sci. 131, 447–457.

Panzica-Kelly, J. M., Zhang, C. X., Danberry, T. L., Flood, A., DeLan, J. W., Brannen, K. C., and Augustine-Rauch, K. A. (2010). Morphological score assignment guidelines for the dechorionated zebrafish teratogenicity assay. Birth Def. Res. (Part B) 89, 382–395.

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 Def. Res. B Dev. Reprod. Toxicol. 83, 104–111.

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.

Zhang, C. X., Cao, J., Kenyon, J., Panzica-Kelly, J., Gong, L., and Augustine-Rauch, K. A. (2012). Development of a streamlined rat whole embryo culture assay for classifying teratogenic potential of pharmaceutical compounds. Toxicol. Sci. 127, 535–546.