EUROPEAN COMMISSION
JOINT RESEARCH CENTRE
Institute for Health and Consumer Protection
I.5 Systems Toxicology Unit
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Report on the Test Presubmission Assessment

Test method:

devTOX quickPredict

TM2014-04

(version 2)

Report on the Test Presubmission Assessment

Test Method: *devTOX quickPredict*

21st May 2015

This report presents a revised version of the EURL ECVAM pre-submission assessment on the devTOXquickPredict test method. It is based on the revised pre-submission provided by the Submitter on March 20th 2015 upon request for clarifications by EURL ECVAM. EURL ECVAM appreciates the efforts of the submitter to answer all questions listed in the previous assessment report.

1. Completeness of the Information Provided

The information provided in the test pre-submission form is sufficient to understand the toxicological endpoint that is intended to be measured, the principle of the test method, and the biological and general mechanistic relevance. However, in order to gain a complete understanding of the test method and its performance it is necessary to study all the additional information provided by the submitter in the attachments (1-10).

2. Test Method Summary

The described test method is an *in vitro* developmental toxicity assay that evaluates whether a compound may have the potential to induce teratogenicity in humans. The assay is based on the measurement of changes in the ornithine/cystine ratio (metabolic biomarkers) present in the media after exposure of undifferentiated human induced pluripotent stem cells to a test item as well as assessment of cell viability. The submitter claims that these two biomarkers provide predictive information on a test item's developmental toxicity potential. Predictive capacity of this assay is based on the assumption that any perturbation in the metabolite ratio past the established teratogenicity threshold (0.85) will have the potential to cause developmental toxicity.
The established 8-point dose response curve enables the prediction of the exposure level that could be teratogenic. Only concentrations that are above a defined teratogenicity threshold based on the ornithine/cystine ratio are considered to have teratogenic potential. However, the relevance of this predictive value to the real exposure could only be evaluated if the C_{max} value is known. The assay is designed for binary predictions, meaning teratogens and non-teratogens regardless of their classification as weak or strong.

The submitter claims that based on previous metabolomic studies the ratio of these two metabolites showed the best correlation with the teratogenicity of the tested compounds (training set). It is important to note that the sensitivity of the assay allows the identification of teratogenic potential at non-cytotoxic concentrations of the tested compound. The assay is currently used to assess developmental toxicity risk potential in drug discovery and preclinical development programs for prioritisation of the development of new compounds using a human ES cell version of the assay (not iPS cells). Besides the prediction of human risk potential the assay with human ES cells has been shown to

have concordance (>75%) with existing *in vivo* models suggesting that it could be used to reduce the number of animals used in *in vivo* developmental toxicity testing. The submitter claims that this human test system has a higher accuracy than the current (validated) animal cell based alternative methods; the mouse embryonal stem cell assay (mEST), the whole embryo culture (WEC) and the zebrafish embryotoxicity (ZET) tests. No other environmental or biological effects are covered by this assay. Based on information provided in an attached table (Attachment #3), the accuracy, sensitivity and specificity of the iPS based assay is: 0.85, 0.78 and 0.91, respectively, meaning a false negative rate of 0.22 and a false positive rate of 0.09.

The devTOX^{qP} assay based on hES cells is already used in the assessment of developmental toxicity in discovery and pre-clinical development in order to facilitate early decision-making. However, the test method does not need to enter a formal validation process for such purpose. Furthermore, the submitter attempts to frame this test method within the current regulatory testing requirements speculating that the assay has the potential to replace studies in a second species, which are required for example for developmental toxicity testing under REACH. In line with the requested clarifications, the submitter suggested that there is potential to eliminate the second species for developmental toxicity testing by combining devTOX^{qP} with complementary *in vitro* endpoints but fails to provide any specific scenarios. To understand the feasibility of this approach and gain regulatory acceptance, it is suggested that data from *in vivo* models along with the proposed *in vitro* screening approach could be compared. The submitter proposes to combine devTOX^{qP} with the zebra fish embryo assay to examine changes in cell morphology in a longer window of development, adding additional confidence. The submitter suggests that ultimately, a single *in vivo* model along with complementary *in vitro* endpoints could provide a weight of evidence based approach to supplant the need for *in vivo* testing in a second species but does not provide any evidence to back up this hypothesis.
This test cannot be used as standalone assay and has to be combined with other complementary assays. Even after the requested further clarifications, it is still not clear what complementary *in vitro* endpoints and what data from *in vivo* studies are required in order to use this test method as part of a reliable testing strategy. The regulatory use of this test method is still unclear and this is important missing information that has to be clearly defined prior to possible formal validation.
Further to EURL ECVAM's request and in view of PARERE consultation, the test submitter provided the following additional information on 26 June 2015: *"The devTOXqP assay reduces animal testing as part of a wider testing strategy in two ways. Firstly, the Assay is applied as a screening tool to identify candidate molecules for further development. Molecules that are identified as having undesirable teratogenic potential at an early stage in the development cycle can be eliminated from further development, and higher-tier animal testing of these molecules will be eliminated. Secondly, it is envisaged that the test will be used to substantiate read-across or category justification for regulatory submissions (e.g. under REACH). By comparing responses between substances for which in vivo data are available (source) and for which such data are lacking (target) it is expected that, as part of a weight-of-evidence approach within the relevant guidance, the application of the available data will be extended, to the benefit of eliminating animal testing for the target molecule(s)".*

The submitter compared the devTOX^{qP} assay to three other *in vitro* test methods for developmental toxicity, i.e. WEC, ZET, and mEST. The devTOX^{qP} assay appears to be a novel approach since it focuses on *human* pluripotent stem cells that remain undifferentiated throughout the testing period (48h). Thus they are able to detect early changes that may indicate teratogenic potential of a compound. In contrast, the other, validated stem cell test, mEST, is carried out in murine stem cells that have differentiated into cardiomyocytes and may be limited as to their ability to predict teratogenic effects that are specific for these cell types. Both the WEC test and the zebrafish embryo assays rely on the use of embryos that are cultured in the presence of test compounds from gestational day 9.5 – 11.5 or 4h post-fertilisation – 6 days, respectively. These methods mostly rely on subjective scoring of morphology or functional outcomes whereas the devTOX^{qP} assay allows objective, quantitative and non-invasive measurement of metabolites in cell culture media.

The submitter also compared the predictive ability of the four *in vitro* methods for the 66 compounds that were used in this submission. Data was incomplete for the WEC, mEST and zebrafish assay albeit, it showed that the devTOX^{qP} assay using iPS cells had higher accuracy and specificity than the WEC, mEST and zebrafish assay but lower sensitivity. However, the reference data used to verify predictions from the *in vitro* assays was mixed meaning that human reference data was used when available but otherwise predictions from rodent (non specified) developmental toxicity assays were used as a reference. EURL ECVAM suggests that a better approach would be to separate the data set and to compare only to human and separately to rodent predictions in addition to the mixed set of reference data.

The submitter pointed out that iPS cells are extremely sensitive to environmental changes affecting the cell's response to a treatment and the internal quality controls that will allow evaluation of the cells' stability. Three different iPS cell lines were evaluated in proof of concept studies. The submitter clarified that quality control criteria for each step in the assay (referring to model and endpoint) are defined in standard operating procedures, which will be disclosed to EURL ECVAM under confidentiality at a possible next step of the submission process.

The submitter further explained that the use of iPS cells can be limited when metabolic activation is required for a compound to show teratogenic potential. It is recommended that biologically active products of biotransformation are tested independently in the assay. Another limitation of the assay reported in this TPF is the maximum tolerated DMSO concentration (0.1%) that is used as a solvent.

The devTOX^{qP} assay is patent protected under United States Patent 8,703,424 (April 22, 2014). The assay can be performed for research purposes only without a license under Stemina's patents. In such cases, if the test method is to be taken up in an OECD Test guideline, performance standards will need to be defined to avoid a monopolistic situation and to allow test method developers to develop similar methods.

3. Preliminary Assessment of the Protocol's Optimisation Level

The submitter referred to several peer-reviewed publications which included protocol description, the cell system used, as well as negative and positive controls. The description of the method to derive results from raw data and description of the quality criteria for assay performance are adequately described. The submitter provided a satisfying description of the prediction model of human developmental effects.

According to the submitter this test method could refine and reduce animal use in two ways: (1) when the assay is applied as a screening tool to take suitable compounds forward into the current required in *vivo* testing and (2) when devTOX^{qP} has gained wider acceptance it could be used in place of a second species or used in read-across or weight of evidence approaches. However, again these scenarios as of now are only hypothetical.

Based on the pre-submission and the additional peer-reviewed publications the submitter referred to, it can be concluded that the assay development is sufficiently described. However, a detailed protocol for assay implementation was not provided and would be needed if this request for validation was to proceed any further. Upon request the submitter further clarified that TPF does not contain confidential information but the standard operating procedures (SOPs) describing assay execution at *Stemina* are considered confidential. The assay methodology has been disclosed in peer reviewed journal articles and in poster format at international meetings. However, Stemina quality checkpoints and internal processes described in SOPs have not been publicly disclosed. Stemina will share relevant SOPs with EURL-ECVAM under confidentiality at the next stage of the evaluation.

4. Preliminary Assessment of the Reliability of the Test Method

The submitter evaluated the within laboratory reproducibility of the assay and upon request it is now described in some more detail. Assay reproducibility was tested on a limited number of compounds (10 chemicals) which is too small for reproducibility assessment. *Sensu stricto* what is presented in Attachment 1 is the variability of the response rather than reproducibility. The EURL ECVAM agrees that such assessment is useful and desirable, however in parallel, the reproducibility as percentages of concordant predictions ("Teratogen" and "Non-Teratogen") within the testing set of chemicals should also be conducted within- and between laboratories. The submitter has offered to supply the raw data to EURL ECVAM if required. No information on transferability and between laboratory reproducibility is provided yet.

5. Preliminary Assessment of the Predictive Capacity of the Test Method

From the information provided to EURL ECVAM, the *devTOX^{qP}* assay using human iPS cells appears to be a promising new method to predict the potential of chemicals to induce developmental toxicity by teratogenicity. The assay was designed around the premise that manifestation of developmental toxicity is dependent to a large degree on compound dose, referring to C_{max} values if available. The prediction model is based on

the evaluation of the ornithine/cystine ratio in response to test compounds, measured over an eight-point dose-response curve. The concentration where the curve crosses the teratogenicity threshold (a threshold of metabolic perturbation that is associated with the potential for teratogenesis; 0.85) is interpolated to give the teratogenicity potential concentration. The submitter claims that exposure levels greater than the teratogenicity potential concentration are predicted to have developmental toxicity potential.

Both the C_{max} and if that was unavailable, the cut-off threshold of 65 µM have been used to put into context the performance of the assay using information available for the test compounds. The C_{max} has been used to evaluate the concentration at which the compound is considered a teratogen where the C_{max} value is known. For those where information regarding the physiological concentration is not known, an arbitrary 65 µM cut-off was used. However, assay results must be put into context using additional knowledge such as estimated human exposure levels extrapolated from pre-clinical data.

The predictive capacity of the assay is based on testing 66 compounds (32 classified as teratogens and 34 as non-teratogens) including agrochemicals, environmental, industrial and pharmaceutical compounds. In comparison with other alternative assays (mEST, WEC and Zebrafish) this assay showed higher accuracy and specificity but lower sensitivity for the test compounds (bearing in mind the limitation of using a mixed human/ rodent reference data set mentioned above).

Overall Conclusion:

Based on the information provided in the revised pre-submission, EURL ECVAM recognises the potential of the use of devTOX quickPredict measurements as means to detect xenobiotics that may have potential to induce developmental toxicity in humans through teratogenicity mechanisms. This test method is useful for non-regulatory applications such as drug discovery, screening and prioritization for further testing but such applications do not require formal validation.

Before proceeding to a full submission for validation, it is essential to determine the usefulness of this assay in a regulatory context which, in the current submission, is still not clearly defined.

The test submitter is invited to reply to the following questions once the regulatory relevance of the devTOXqp has been clarified with PARERE:

1) The submitter needs to clearly distinguish between the results obtained from the test method based on the hES cells type and those obtained on the basis of the iPS cell type. In the pre-submission form results obtained with the two different cell types were being referenced which created some confusion since it was not always immediately clear, which version of the test method was being referred to. Importantly, the performance of these two cell types in the test method is not the same. The submitter claims that the response of the iPS and hES cells are similar on the basis of *Attachments 1 and 2*. In attachment 1 correlation is presented for only 17 chemicals. The R2=0.84 shows good correlation for these chemicals only. What is the correlation for the entire set of 66 chemicals? Therefore there is no sufficient evidence that hES and iPS cells perform similarly. An additional way to do it would be to calculate Cohen's kappa coefficient (for measuring inter-rater agreement for the 66 chemicals). *Attachment #2* presents results for

iPS cells; as no results are presented in this *Attachment* for the other version, it does not show that hES and iPS cells perform similarly.

2) As already pointed out in the assessment report (see paragraphs 2 and 4), further clarifications and analyses are needed to evaluate the performance and reproducibility of the devTOXqP assay using iPS and hES cells. The final outcome of the devTOXqP are binary predictions performed as "Teratogen" or "Non-Teratogen". Therefore, the predictive capacity of the devTOXqP assay, expressed as sensitivity and specificity values, is a key parameter that will ultimately determine the fate of the assay in the regulatory context (especially the rate of false negatives). Ideally, the submitter should provide these sensitivity and specificity values according to three modalities: (a) against human data only; (b) against animal data only; (c) against human and animal studies mixed (considered globally). In each of these three cases it is recommended to indicate the numbers of chemicals that were used to derive these values. This is because human and animal data present different levels of information where animal is rather a proxy for human data; therefore splitting up the values of sensitivity and specificity (and then compare to the mixed dataset) is a way to ensure that no marked difference is indeed observed i.e., that performance remain in the same range of values.
In paragraph 4 of the assessment report it is described that "Assay reproducibility was tested on a limited number of compounds (10 chemicals) which is too small for reproducibility assessment. Sensu stricto what is presented in *Attachment 1* is variability of the response rather than reproducibility. EURL ECVAM agrees that such assessment is useful and desirable, however in parallel, the reproducibility as percentages of concordant predictions ("Teratogen" and "Non-Teratogen") within the testing set of chemicals should also be conducted within- and between laboratories." From the revised submission it was not possible to determine whether within- and/or between laboratory reproducibility was evaluated, and this should be clarified.

3) Page 11 of the TPF, section 4.2. The submitter explained the rationale behind choosing the same compound, methotrexate (MTX) at different concentrations as a negative and positive control."(…) a compound's teratogenic potential is directly related to the exposure level", which is the basic principle behind the devTOXqP assay. However, EURL ECVAM would advise to use a non-teratogen as a negative control as such an approach is widely used and accepted.

4) Page 4 of the pre-submission text, in the clarification provided, the submitter mentions that the US EPA has chosen the hES version of the assay for the ToxCast™ program. Why did EPA choose this version only?

5) Confidentiality of SOPs is an issue if the test method is meant to be adopted in international guidelines since regulators and users request transparency on important mechanisms that enable them to understand how test results are generated and interpreted, i.e. avoiding a ‘black box’ test system.