EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)

ESAC OPINION

on an ECVAM-led validation study on two *in vitro* hepatic human-derived test methods for assessing induction of Cytochrome P450 enzymes (CYPs)

ESAC Opinion Nr.	2014-02
Relevant ESAC request Nr.	2013-01
Date of opinion	21.11.2014

TABLE OF CONTENTS

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

21 November 2014

Summary of the ESAC opinion

The ESAC was asked to provide an opinion on an EURL-ECVAM coordinated study concerning a validation study for cytochrome P450 induction providing a reliable human metabolic competent standard model or method using the human cryopreserved primary hepatocytes and the human cryopreserved HepaRG cell line.

The main objective of this study was to assess the transferability, the reproducibility (within and between laboratories) and the predictive capacity of two *in vitro* methods, each of which evaluated the induction of enzymatic activity of four Cytochrome P450 enzymes (CYP1A2, CYP2B6, CYP2C9, and CYP3A4). The two CYP induction *in vitro* methods used different metabolically competent *in vitro* test systems: cryopreserved human HepaRG cells and cryopreserved human primary hepatocytes. Predictive capacity was assessed using exclusively human CYP induction *in vivo* reference data, which meant a restriction to chemicals used for pharmaceutical purposes.

The ESAC believes that the study findings satisfy the requirements for test definition, within laboratory reproducibility, transferability, and between laboratory reproducibility, but only partially satisfy the requirements for assessment of predictive capacity. The ESAC does not believe that the study data are sufficient to conclude on the readiness of the test for regulatory use in the context of chemical safety assessments. In particular the ESAC recommends that additional CYP induction studies be conducted with rodent hepatocytes to further investigate the applicability domain and predictive power of the assay for chemicals that have challenging physico-chemical properties such as persistence/bioaccumulation (typically highly lipophilic compounds), rapid metabolism, and poor water solubility. Despite the limitations of using rodent hepatocytes with regard to human relevance, such studies would be valuable for assessing the broader applicability of induction assays to chemicals in general. Nevertheless, the ESAC agrees that there may be a potential role for a human CYP induction assay as a marker of possible receptor activation within an integrated testing strategy for a particular Adverse Outcome Pathway, particularly if the assay includes mRNA measurement to confirm increased transcription of the associated CYP isozyme. The human cell assay also has potential use for evaluation of the human relevance of animal test results, whether *in vivo* or *in vitro*, that suggest the activation of a receptor for which the human cyp induction assay is designed may be a key element in a toxicity pathway.

The ESAC feels that the current study data are in good agreement with other existing information regarding hepatocyte assays, and that it provides evidence that reliable hepatocyte assays for other important purposes, including identification of metabolites and quantification of metabolic clearance, are feasible. The ESAC strongly encourages ECVAM to continue to conduct studies with human hepatic models to develop methods for characterization of other kinetic data, including clearance, metabolic profiling, and inhibition. The importance of developing *in vitro* to *in vivo* extrapolation methods cannot be overemphasized.

1. Mandate of the ESAC

1) DESIGN & CONDUCT OF STUDY: The ESAC is requested to review whether the study was conducted appropriately in view of the objective of the study. The study objective was to assess:

(1) reproducibility of the test method within laboratories (WLR)

(2) transferability of the test methods to other laboratories

(3) reproducibility of the test methods between laboratories (BLR)

(4) relevance/predictive capacity of the *in vitro* test methods for biotransformation of substances as compared to human clinical data from pharmaceuticals.

(5) the applicability domain and possible limitations of the test method. The selection of the test chemicals and analyses of possible reasons results *in vitro* not matching the human reference data should be carefully reviewed.

When reviewing the design and conduct of the study, the following issues should be addressed in particular:

(a) Clarity of the test definition (module 1)

(b) Clarity of the definition of the study objective and study management

(c) Appropriateness of the study design & execution in view of the study objectives, inter alia:

- Is the number of tested chemicals sufficient for the purposes of the study?

- Are the reference data (which are associated with the test chemicals) appropriate and of good quality in view of assessing in particular the predictive capacity? Should additional reference data (potentially also for the same substances) but from other reference sources been included? Where there selection criteria? Was the selection scientifically justified?

- In case of gaps (chemical class etc.) – are these justified?

- Is the number of laboratories sufficient?

(d) Appropriateness of the study execution (e.g. were there pre-defined test acceptance criteria, were these respected? How were exceptions / deviations handled? Were provisions specified for retesting? Was the number of repetitions sufficient? etc.)

(e) Appropriateness of the statistical analysis used for analysing WLR, transferability, BLR and relevance / predictive capacity.

2) CONCLUSIONS OF STUDY: The ESAC is requested to assess whether the conclusions, as presented in the material made available to ESAC are substantiated by the information generated in the study and are plausible with respect to existing information and current views (e.g. literature).

In particular:

(a) Are the conclusions on reproducibility (WLR and BLR) as well as transferability justified and plausible?

(b) Are the conclusions on relevance / predictive capacity justified and plausible with respect to the reference data, other existing information and with respect to the intended use of the test method.

(c) Are there possible gaps between study design and study conclusions which remain to be addressed in view of the suggested conclusions / use (see also point 3)?

(d) Do the data generated with the validation set of chemicals together with possible available existing data provide sufficient information on the applicability and possible limitations of the test method, in particular in view of its potential use within integrated approaches to support assessment of biotransformation / toxicokinetics.

3) SUGGESTED USE OF THE TEST METHOD: The ESAC is requested

(a) to evaluate, on the basis of the data summarised in the Validation Study Report (VSR), the intended use of the test method and its readiness to serve as a reference point for defining performance criteria for routine assessment, i.e. for developing a Performance-Based Test Guideline (PBTG).

(b) to make additional recommendations (as required) on the proper scientific use of the test method, possibly within integrated approaches taking specific aspects of this method into account (e.g. applicability, technical limitations etc.),

(c) to identify possible further information required (i.e. are there data gaps or gaps with regard to mechanistic understanding ?) to be able to determine the potential use and usefulness of the test method within integrated approaches, duly considering the need to text chemicals and mixtures/products.

2. Detailed opinion of the ESAC

The ESAC was asked to provide an opinion on an EURL-ECVAM coordinated study concerning a validation study on cytochrome P450 induction providing a reliable human metabolic competent standard model or method using the human cryopreserved primary hepatocytes and the human cryopreserved HepaRG cell line.

2.1 Background, regulatory and scientific rationale

The scientific rationale for the assay is stated in the context of the study objectives, as well as under Module 1 (which provides a short description of the intended purpose). The rationale is described in more detail under the secondary objectives, which state that the assay is intended to contribute to knowledge regarding:

- CYP induction as a toxicity event

- Elucidation of Mode of Action (MoA)
- Biomarkers of exposure to chemicals (if CYPs are induced – there may have been significant exposure)
- Potential for effects on mixture toxicity
- Indication of a role of metabolism in a compound's toxicity

The ESAC believes that the clarity of the section outlining the scientific rationale would have benefited from restricting the explanations to what the test method measures, i.e. identification of chemicals that lead to induction of four selected CYP enzyme in human hepatic test systems. The ESAC feels that the text on the potential contribution of the assay for toxicokinetics distracts from the explanations on the scientific rationale. While part of the purpose of the assay may be to identify potential effects of induction by the studied chemical on the metabolism of other xenobiotics and endogenous compounds to which an individual may be co-exposed, the assay does not characterize metabolism of the studied xenobiotic itself. Also, in future descriptions of this assay, care should be taken not to imply that CYP induction, *per se*, is directly linked to toxicity (there is only very limited information suggesting that chronic induction may be linked, in some cases, to hepatomegaly) or as providing a measure for potential for metabolite formation.

A regulatory rationale is provided but it refers primarily to pharmaceuticals. In order to facilitate inclusion in a OECD PBTG and support EU legislations (REACH, Cosmetics, Animal Welfare), as stated in section 4.2 of the Validation Study Report (VSR), the regulatory rationale/usefulness should have been expanded also to other chemicals of concern for environmental/cosmetic exposure. Limitations in the scientific and regulatory usefulness of the assay are not sufficiently addressed, e.g., the spectrum of isoforms is limited to those affecting pharmaceuticals and the variability of CYP induction profiles in different tissues/cells is not addressed. The latter point is important since exposure routes to xenobiotics are not only oral (where hepatic induction is very relevant), but can also be inhalation or dermal.

Most of the background is more relevant to pharmaceuticals than to other xenobiotics The justification for the use of human test systems for induction is well presented, but the resulting limitations in the evaluation of the assay are not adequately discussed. Application of this assay to different classes of chemicals (e.g., volatiles, lipophilics) is suggested but not demonstrated, and indeed is considered problematic by the ESAC.

2.2 Design and conduct of the study

2.2.1 Definition of the study objectives

According to the VSR, the objective of this study was to assess the transferability, the reproducibility (within and between laboratories) and the predictive capacity of two Cytochrome P450 induction *in vitro* methods, each of which evaluated the induction of enzymatic activity of four CYP enzymes (CYP1A2, CYP2B6, CYP2C9, and CYP3A4). The two CYP induction *in vitro* methods used different metabolically competent *in vitro* test systems: cryopreserved human HepaRG cells and cryopreserved human primary hepatocytes. Predictive capacity was assessed using exclusively human CYP induction *in vivo* reference data.

2.2.2 Study design

The number of laboratories in this study is considered appropriate and the number of replicates (3) is considered adequate. The number of substances is considered to be too small for a confirmative data analysis of predictive capacity. A sensitivity, e.g., of 5/5 (Table M5.1) corresponds to a point estimate of 100%, but an exact 95%-Confidence interval ranging from 48% to 100%. This shows that proof of good predictive capacity can only be achieved by a strong increase in number of tested compounds. However, the ESAC also acknowledges that it is not always feasible in the context of validation studies to assess a sufficiently high number of chemicals. Practical constraints such as the availability of good reference data (and hence test chemicals with accompanying high quality data), the cost and time factors to be considered when organising practical testing as well as other factors, may impact on the final sample size that, realistically, can be studied. The ESAC considers that these factors (in particular availability of reference data) impacted on the sample size used in the CYP induction validation study. The ESAC nevertheless holds that the sample size, despite being insufficient for confirmative analysis, provides a good indication on the suitability of the chosen test systems for studying induction of CYP enzymes.

Reference chemicals are well chosen, but documentation of the search strategy and selection criteria was not found in the report. The acceptance criteria for determining the specific studies to be used to characterize the *in vivo* induction are critical to assure that the points of comparison with the *in vitro* assay are correct, so in future these criteria should be documented. It would have been valuable to include more compounds (non-prototypical, weaker inducers), with particular emphasis on representative chemicals from different classes with difficult properties (e.g., persistent / highly lipophilic compounds, rapidly metabolized compounds, poorly soluble compounds). A broader diversity of chemical properties would increase the level of understanding of the applicability domain for this assay. The ESAC recognizes that finding human *in vivo* data for such a wide variety of chemicals would be problematic and therefore suggests consideration of a rat hepatocyte CYP induction assay to further explore the applicability domain (see recommendations).

The choice of IC_{30} as the cutoff for toxicity is not considered optimal by the ESAC. It is probably responsible for the non-monotonic dose response behaviours observed in the current study. The ESAC would suggest no more than an IC_{10} as point of departure.

In future studies, measurement of mRNA expression for the CYPS in question should be considered as a parallel endpoint together with activity and cytotoxicity to assess the induction potential of the

compounds, in order to address the potential for inhibition, suppression, or toxicity that may confound induction assessments based on enzymatic activity. In addition, mRNA is more proximal to receptor activation and is required for pharmaceutical submissions by regulatory agencies.

2.2.3 Statistical analysis

The ESAC agrees with the use of a factor of two to identify efficacious inducers. The decision rule for a substance to be called an inducer is then that at least in one concentration the factor of two is exceeded. The ESAC suggests combining in the future the calculation of this factor with appropriate statistical hypothesis testing. The appropriate test for the comparison of the means of a quantitative variable in several treatment groups to control is to use ANOVA (to test whether there is any difference in means between all treatment groups and control) and, in case of a significant result of the ANOVA, to perform post-hoc Dunnett tests for pairwise comparison of treatment groups versus vehicle control to identify which treatment groups are different from control. Only if the factor exceeds two and the corresponding Dunnett test yields a significant difference should the substance be called an inducer.

Calculation of the factor together with ANOVA and post-hoc Dunnett tests should be performed separately for every batch in every laboratory, thus leading to the call of inducer/non-inducer for every batch and every laboratory. This call should then be the basis for evaluation of WLR and BLR.

In future it would be preferable to state exact 95%-confidence intervals for sensitivity and specificity.

2.3 Study results and conclusions

2.3.1 Standardised use of the test method protocol

Quality assurance systems

Overall, the ESAC considers that the Quality Assurance system adopted during the study is acceptable, although it should have been more completely described in the report. The study is described as being carried out 'according to GLP principles' in test facilities certified as compliant to GLP principles; therefore, it is expected that the personnel of the QAU were independent from the laboratory staff generating the data (as well as from the Study director). It is stated in the minimum requirements for non GLP test facilities: 'Quality Assurance should be performed in accordance with the principles of GLP (for GLP compliant laboratories)'. From the parenthetical text, it is not clear whether this bullet point applied to the non GLP test facilities as well. Although the Quality Assurance responsibilities are also described in the Study Plan for non GLP test facilities, the ESAC could not determine whether the QA activity was performed by a staff member independent from the laboratory staff generating the data.

Nevertheless, the ESAC is satisfied that quality systems were used. Pharmacelsus GmbH, Janssen Pharmaceutica and EURL ECVAM test facilities are certified as compliant to the GLP OECD principle. It is worth of note that EURL-ECVAM is certified for the 'validation of *in vitro* methods' (included in OECD category 9): although the topic is out of the scope of the EU regulation for the application of

GLP, it was included as an area of interest by the GLP Italian Monitoring Authority, since the request came from a DG of the Commission. Indeed, there is no clear written regulation about the need for validation studies being conducted in full compliance with GLP. Therefore it is fully correct, that the test facilities carried out the studies 'according to' the GLP principles. For the analytical part performed at EURL-ECVAM, although not fully compliant to GLP, it can be considered again carried out 'according to the principle'; in addition the control/ maintenance of the instrument was under the ISO 17025 accreditation, which is considered sufficiently reliable. For the non-GLP laboratories participating in the validation project, the minimum set of quality assurance requirements was considered appropriate. Overall, the ESAC is comfortable that quality assurance was acceptable. Once the method is further validated and adopted as a guideline for a test to be used in the regulatory frame of safety assessment of chemicals, the test could be carried out in compliance with GLP principle (i.e. OECD category 2: toxicity testing or category 9 under TK studies).

Standard Operating Procedure (SOP)

The SOPs made available to the ESAC (provided not as originals, without signatures and date) describing the method as a whole (including activities to be performed in different laboratories) are sufficiently detailed. However, they cannot be used as SOPs as intended in a GLP environment. They are not user-friendly: a document of more than 80 pages does not help consultation for the operators and it is very likely that other working documents were generated (indicated also in the study plan as 'home documents') and used in the daily laboratory work. Therefore to name them as SOPs could generate confusion, which is not compliant with a quality system. The system of drafting and managing SOPs could be improved. In addition the SOPs failed in describing a harmonized format for data reporting. Moreover, The suggested data reporting could be improved by standardizing it with a clear description in the SOPs, which is missing. The CYP induction SOPs contained a set of acceptance criteria for the evaluation of runs to determine whether the obtained results are valid. However, despite the length of the SOPs, no indication about the reporting of data is given.

2.3.2 Within- and between laboratory reproducibility

Based on a proposal of the ESAC WG, both WLR and BLR were recalculated. The WG had suggested that a >2 fold induction at *any* concentration would be sufficient to decide on a positive for a given batch. Thus, one single concentration with >2 induction is sufficient to classify a batch as positive=1. Thus the data matrices were reduced from six values (=six concentrations) to a single classifier (0=negative/1=positive) per batch in each laboratory.

Based on this approach

- WLR was calculated by assessing the **concordance of predictions between batches** used in a given laboratory.
- BLR was calculated by assessing the **concordance of predictions between laboratories** for a given batch.

Within-laboratory reproducibility (WLR):

The ESAC is comfortable with the results of this re-evaluation of WLR. Table 1 summarised the results for the two test systems by providing the ranges of concordances (in percent) between batches obtained in the laboratories: the ESAC considers the WLR to be representative of what can realistically be achieved by these test systems. The low reproducibility for CYP1A2 with hepatocytes is unsurprising given its high variation in expression across individuals and the use of a two-fold cut-off to define induction. Use of a higher fold cut-off (e.g., five-fold) would decrease sensitivity to background noise and probably increase the reproducibility for this enzyme.

Table 1: Ranges of observed WLR (based on concordance in % of predictions between batches within each laboratory) for the two test systems studied. The values are based on twelve test chemicals.

CYP isoform	cryoHep	HepaRG
CYP1A2	25-50%	60-100%
CYP2B6	58-83%	50-70%
CYP2C9	67-83%	40-80%
CYP3A4	67-75%	80-90%

Between-laboratory reproducibility (BLR):

The ESAC considers the BLR for all CYPS to be representative of what can realistically be achieved with the test systems given the inherent variability in the functionality of hepatocyte cultures. Although the reproducibility is lower for the cryohepatocytes, a reproducibility of about 60% or above is considered as being in line with the current state of the art for primary cells from multiple individuals. Table 2 summarised the results by showing the ranges of concordant predictions obtained between laboratories for the two test systems.

Table 2: Ranges of observed BLR (based on concordance in % of predictions between laboratories for a given batch) for the two test systems studied. The values are based on twelve test chemicals.

CYP isoform	cryoHep	HepaRG
CYP1A2	42-58%	70-90%
CYP2B6	67-75%	50-80%
CYP2C9	58-83%	60-70%
CYP3A4	75-83%	80-90%

2.3.3 Conclusions on predictive capacity

The analysis shown in Tables 5M.1-6 of the VSR is for the most part appropriate. For the small number of compounds considered, it demonstrates reasonably good predictive capacity. Due to the small number of compounds, however, the observed sensitivities and specificities are associated with large variation. A sensitivity, e.g., of 5/5 (Table M5.1) corresponds to a point estimate of 100%, but an exact 95%-Confidence interval ranging from 48% to 100%. This shows that proof of good predictive capacity can only be achieved by a strong increase in number of tested compounds.

The ESAC would suggest, however, that the comparison with *in vivo* C_{max} should not have been used to change the *in vitro* call (e.g., for omeprazole and artemisinin) since in the future application of this test, these data will not be available on the test chemicals, unless they have been clinically studied. The assay as described in the VSR is intended to identify chemicals with the potential to cause induction and is not described as a predictor of induction at environmentally relevant concentrations.

Incorporation of predicted AUC (Area Under the Curve) changes from activity and mRNA concentration-response data would improve the ability to assess the predictivity of observed results for human or rodent models.

2.3.4 Applicability and possible limitations

The test method is biologically relevant for the endpoint described in the report: induction of CYP activity for CYPs that may be associated with receptor-activated pathways. This information could be integrated into a test strategy that could help to assign a chemical to a particular AOP. The use of a minimum two-fold increase in metabolite production as the definition of induction is consistent with current practice, but a more robust definition that encompasses both potency and efficacy would have greater biological relevance.

The applicability domain for these assays is uncertain since all the test chemicals were pharmaceuticals with similar properties (low volatility/metabolism/lipophilicity, high bioavailability/solubility). The application of this test to xenobiotics with a much wider range of physical-chemical and other properties remains to be established, including, importantly, persistent and bioaccumulative substances for which CYP induction data may be useful but which may be challenging to test in *in vitro* systems. Ideally, additional receptor-related CYPs would also be included, such as CYP4A, and several other CYP2 isozymes such as 2E1 and 2C19.

2.3.5 Identified gaps between study design and study conclusions

One of the primary issues highlighted by the ESAC was the need to determine the broader applicability of these induction models and assay systems to xenobiotics with challenging physic-chemical properties. However, this issue would be difficult to assess with the relatively small number of chemicals that could be examined using human data. Part of the rationale for the selection of a relatively small set of reference compounds was the limited number of compounds, mostly pharmaceuticals, for which there are *in vivo* data on induction of human xenobiotic metabolizing enzymes. To more broadly evaluate the application of these human assays for xenobiotics could be

particularly challenging since human *in vivo* induction data is unlikely to be available for non-pharmaceuticals.

Measurement of mRNA expression for the CYPS in question should have been included as a parallel endpoint together with activity and cytotoxicity to assess the induction potential of the compounds, in order to address the potential for inhibition, suppression, or toxicity that may confound induction assessments based on enzymatic activity. This would provide greater confidence that the assay is identifying receptor based induction rather than other possible modulation of CYP activity.

2.4 Potential regulatory use of the test method

This test should eventually be useful as part of an integrated testing strategy to assign a chemical to a receptor mediated AOP. However, additional studies are needed to characterize the applicability domain, as discussed above. Neverthless, the BLR and WLR results in this study could serve as a reference point for defining performance criteria for routine assessment, i.e., for developing a Performance-Based Test Guideline (PBTG) for assays based on measurement of metabolism by CYPs 1A2, 2B6, 2C9, and 3A4 in cells that are cultivated in submerged cell culture systems. ESAC recommends carefully evaluating the type of culturing method (e.g. submerged hepatocyte monolayer cultures versus 3-dimensional, flow-based, or multicellular cultures) that would be included in a PBTG, since it is known that the culturing method may have consequences on cellular properties including to which extent cells are metabolically active (Schyschka et al., 2013; Godoy et al., 2013).

2.5 Recommendations

The ESAC strongly encourages ECVAM to continue to conduct studies with human hepatic models to develop methods for characterization of other kinetic data, including clearance, metabolic profiling, and inhibition. The importance of developing *in vitro* to *in vivo* extrapolation methods cannot be overemphasized. The study reported in the VSR specifically addresses the availability, transferability and reliability of metabolically competent hepatocellular test systems for *in vitro* testing, using induction of metabolism as a case study. As such, the reproducibility in this study could be a useful indicator of the potential for other uses. The current study design could be adapted for these purposes while concurrently addressing the objectives of this study, for example induction and mode of action. Generally, the ESAC is of the opinion that the indications of WLR and BLR (measured as between batch and batch-to-batch reproducibility; c.f. sections 2.3.2) represent the current state of the art of what can be realistically achieved with the test systems used. Recommendations reg. donor number (in case of hepatocytes) are being made below.

Since most *in vivo* metabolism studies will be conducted in rodents (usually rats) and species differences in the endogenous and induced levels of various CYP enzymes can be expected, a key component in the extrapolation from *in vitro* to *in vivo* and from rodent to human would be *in vitro* data obtained in rodent cells. The ESAC therefore suggests that a possible path forward for broadening the applicability domain of this type of approach to xenobiotics in general would be to conduct a similar study using primary rat hepatocyte culture models or rat hepatocyte cell lines, and to include induction assays for key enzymes reflective of AhR (e.g. CYP1A2), CAR (e.g. CYP2B1/2B2), PXR (e.g. CYP3A1/3A23), while possibly adding PPARα (CYP4A1), which is known to be linked to

rodent carcinogenicity, as well as CYP2E1 (although it is not related to activation of a pathway). This would allow a wider range of xenobiotics to be studied in concentration response and afford the opportunity to make use of rodent *in vivo* induction data for evaluation of predictions. While these results would be less directly translatable to human health, they could avoid the practical impossibility of intentionally exposing humans to higher concentrations of xenobiotics to assess induction potential. Such a rat study would support assessment of limitations of the applicability domain associated with chemical properties. In addition, the rodent hepatocyte induction assay could provide useful information for interpreting toxicity test results from *in vitro* or *in vivo* rodent studies and could contribute to a mechanistic understanding of the toxicity observed towards the elucidation of an AOP. The human hepatocyte induction assay could then be used in parallel for and evaluation of human relevance.

Additional suggestions for consideration in future CYP induction studies:

a) In future studies, measurement of mRNA expression for the CYPS in question should be considered as an endpoint together with activity and cytotoxicity to assess the induction potential of the compounds, in order to address the potential for inhibition, suppression, or toxicity that may confound induction assessments based on enzymatic activity.

b) While the use of hepatocytes from three individual donors is consistent with current practice, it would be preferable to have a larger number of individual donor preparations or develop approaches that use pooled cells from a larger number (e.g., 10) of donors. It is acknowledged, however, that there may be practical and economical constraints with regard to increasing the donor number. Moreover, given the fact that important functional single nucleotide polymorphisms (SNPs) have been identified in the CYP isoforms investigated and that inter-individual differences in drug disposition are important causes for adverse drug/chemical reactions, prior screening of high frequency SNPs is desirable.

c) The concentration of phenobarbital (that gives a two-fold induction) used in the current study was insufficient to saturate induction (to get an accurate EC_{50}), and a greater fold-change for positive controls may have been preferable. Also, acceptance of a batch would preferably be based on both fold-change in positive controls as well as metabolite generation in controls for cocktail exposures.

d) The Alamar Blue assay is primarily for mitochondrial activity, which can be perturbed by inducers without toxicity. Other biomarkers (LDH leakage, AST, ALT) would provide better markers for toxicity in hepatic cells.

e) The choice of IC_{20} as the cut-off for toxicity is not considered optimal by the ESAC. It is probably responsible for the non-monotonic dose response behaviours observed in the current study. ESAC suggests reanalysing the study data based on the use of the IC_{10}.

f) Prior to designing any additional studies, the ESAC recommends that ECVAM consider re-analysing the activity data from the current study without normalization for protein content (only normalize on plated cell number). Normalisation to protein content can add additional variability due in part to dying cells.

g) In future studies, use of a higher fold cut-off (e.g., five-fold) would decrease susceptibility to background noise and probably increase the reproducibility of the assay, especially for some CYPs with highly variable expression levels (e.g. CYP1A2). ESAC suggests reanalysing the

extent of background noise encountered for each CYP isoform in view of setting, possibly, cut-offs adapted to individual CYP isoforms, i.e. CYP-specific cut-offs based on their inducibility.

h) The ESAC recommends that in future evaluations the comparison with *in vivo* Cmax not be used to change the *in vitro* call (e.g., for omeprazole and artemisinin in the current study) since in the application of this test, these data will not be available on the test chemicals unless they have been studied clinically. The assay as described in the VSR is intended to identify chemicals with the potential to cause induction and is not described as a predictor of induction at environmentally relevant concentrations. Instead, selection of test concentrations should be based on cytoxicity range-finding data, with final concentrations in the CYP assays chosen to be below those inducing significant cytotoxicity.

i) The ESAC suggests in future to use ANOVA with post-hoc Dunnett test for multiple comparison to vehicle control and to exclude non-significant findings. WLR and BLR should be evaluated on the call for a single curve (at least 1 concentration with >2 and Dunnett test significant), not on the individual concentrations. Exact 95%-confidence intervals for sensitivity and specificity should be provided.

j) Test chemicals for this study were selected on the basis of availability of human data which meant a restriction to pharmaceuticals. Moreover, selection criteria for determining the specific published studies to be used to characterize the *in vivo* induction reference data were not provided in the validation study report. However such criteria are critical to assure that the points of comparison with the *in vitro* assay are correct, so in future these criteria should be documented.

k) Based on the notion that the SOPs were rather long and over-detailed, which may distract from proper execution of the assay procedure, ESAC suggests to consider revising the system of drafting and managing SOPs, including a harmonized format for data reporting (i.e. by devising a reporting template).

3. References

Schyschka L (2013) Hepatic 3D cultures but not 2D cultures preserve specific transporter activity for acetaminophen-induced hepatotoxicity. *Arch Toxicol. 87(8):*1581-93.

Godoy P et al. (2013) Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal livercells and their use in investigating mechanisms of hepatotoxicity, cell signalling and ADME. *Arch Toxicol.* 87(8):1315-530.

Annex 1 ESAC and ESAC Working Group charged with the scientific review

ECVAM Scientific Advisory Committee

- Prof. Jürgen BORLAK
- Dr. Neil CARMICHAEL
- Dr. Edward CARNEY
- Dr. Harvey CLEWELL
- Prof. Lucio G. COSTA
- Dr. Kristina KEJLOVÁ
- Prof. David John KIRKLAND
- Prof. Annette KOPP-SCHNEIDER
- Dr. Renate KRÄTKE
- Prof. Claus-Michael LEHR
- Dr. José Maria NAVAS
- Prof. Aldert PIERSMA
- Dr. Jonathan RICHMOND
- Dr. Erwin L ROGGEN
- Dr. Dorothea SESARDIC

ESAC Working Group (WG)

The ESAC WG had the following members:

ESAC members:

- Harvey Clewell (WG Chair)
- Claus-Michael Lehr
- Jürgen Borlak
- Annette Kopp-Schneider (focusing on statistical aspects)

External experts:

- Emanuela Testai, ISS, Rome, Italy (proposed by WG Chair)
- Stephen Ferguson (nominated by NICEATM/ICCVAM and supported by ECVAM).

ESAC Coordination:

- Dr. Claudius Griesinger (ESAC Coordinator)
- Dr. Michael Schäffer