DRAFT PERFORMANCE STANDARDS FOR NEW HUMAN CYTOCHROME P450 (CYP) N-FOLD INDUCTION *IN VITRO* TEST METHOD FOR TG XXX

Series on Testing and Assessment

No XXX

INTRODUCTION

1. The following Performance Standards (PS) accompanies the draft proposal for a new performance based test guideline on human cytochrome P450 (CYP) n-fold induction *in vitro* test method. This document is intended as a guide to developers of new test methods that are analogous to existing, fully validated test methods in that they are based on similar scientific principles and predict the same effect (colloquially referred to as "me too" tests) (1). Prior to the acceptance of a new test method for (TG XXX) regulatory testing applications, validation studies are conducted using scientifically sound principles to establish its reliability (i.e., the extent of intra- and inter-laboratory reproducibility over time when performed using the standardized protocol), and its relevance (i.e., the ability of the test method to correctly predict or measure the biological effect of interest) (1; 2; 3; 4). The purpose of the PS is to communicate the basis by which new proprietary (i.e. copyrighted, trademarked, registered) or non-proprietary test methods can be determined to have sufficient accuracy (i.e., agreement between a test method result and an accepted reference value) and reliability (i.e., extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol) for a specific testing purpose. Thus, this provides an avenue to demonstrate that a newly developed test method based on similar scientific principles has comparable or better performance capabilities than those from which the existing PS were derived, and may allow a more timely use of the new test method. New test methods ("me too" tests) can be added to TG XXX after OECD review and agreement that performance standards are met. A new test method developed under this PS will be covered by TG XXX only after TG XXX has been updated to add the new test method.

2. Performance standards are based on an adequately validated test method(s) and provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar (1; 2). The three elements of performance standards are:

- Essential test method components: these consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed test method that is considered to be mechanistically and functionally similar to the validated method. Essential test method components include unique characteristics of the test method, critical procedural details, and quality control measures.

- A list of reference substances: Reference substances are used to assess the accuracy and reliability of a proposed mechanistically and functionally similar test method. These substances are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method, and are the minimum number that should be used to evaluate the performance of a proposed mechanistically and functionally similar test method.

- Accuracy and reliability performance values: These are the standards for accuracy (i.e., sensitivity, specificity, false positive/negative rates) and reliability (i.e., degree to which the test method can be

performed reproducibly within and among laboratories over time) that the proposed test method should meet or exceed when evaluated using the minimum list of reference substances.

3. The fully validated reference test methods that provide the basis for this PS are:

- The cryopreserved human HepaRG cells CYP n-fold induction *in vitro* test method (CryoHepaRG)
- The cryopreserved human primary hepatocytes CYP n-fold induction *in vitro* test method (Cryoheps)

ESSENTIAL TEST METHOD COMPONENTS AND OTHER VALIDATION CONSIDERATIONS

4. Certain principles are important in delineating the essential test method components that determine whether the CYP n-fold induction tests are functionally and mechanistically similar. Human *in vitro* CYP n-fold induction tests are designed to identify test substances that induce CYP1A2, CYP2B6 and CYP3A4, involving several cellular processes such as xenobiotic-receptor binding (CAR, PXR and AhR), *de novo* protein synthesis or protein stabilization. This information can be integrated into a test strategy that could help to assign a test substance to a particular Adverse Outcome Pathway, where CYP induction is identified as a key event.

5. The *following test method components may vary*, so this PBTG does apply to test methods that may differ in

- human CYP1A2, CYP2B6 and CYP3A4 metabolic competent cell type
- culture conditions
- plating density
- plate layout (including how controls are incorporated)
- method of determining solubility
- method of determining cytotoxicity

These elements should be clearly described in each test method.

6. Essential test method components for *in vitro* CYP n-fold induction protocols should include:

- The use of prototypical inducers β-naphthoflavone (25 µM), phenobarbital (500 µM), and rifampicin (10 µM). Exposure to reference substances must lead to a > 2-fold increase of enzymatic activity (of statistical significance) at the defined concentrations.

- The use of a negative (vehicle) solvent-treated control, i.e., samples without test chemical (or control inducers) but with organic solvent (e.g. 0.1% v/v DMSO) to determine the generation of specific products without inducing effects. Basal values of the respective negative controls are used for calculation of the n-fold induction potential of test and reference substances (n-fold induction). The organic solvent should be miscible with cell culture media at concentrations that are not cytotoxic and do not otherwise interfere with the test system.

- A quantitative evaluation of cytotoxicity and how it is applied to the test method should be included in each study. Concentrations of test substances that clearly reduce viability should not be considered in the analysis of the data. Adequate negative and positive controls should be included in the cytotoxicity assay.

- All concentrations of the controls (e.g., vehicle, positive(s), or prototypical inducer(s)), and the test substance should be tested at least in triplicates.

- In the absence of solubility or cytotoxicity restraints, the maximum concentration may be 40 mg/ml or even up to the limit of solubility, if appropriate.

7. For analyzing the data obtained from the CYP n-fold induction methods the 2-fold induction factor should be used to identify efficacious inducers. The decision rule for a substance to be called an inducer is that it should lead to ≥ 2-fold increase of enzymatic activity for at least two soluble non-cytotoxic concentrations. In addition, the use Anova with post-hoc Dunnett test ($p<0.05$) for multiple comparisons to vehicle control and to exclude non-significant findings is recommended.

8. CYPs' basal activity in control wells (in presence of 0.1% DMSO): based on data generated in the validation study the variability of CYP activities is markedly higher in cryoheps compared to cryoHepaRG cells, reflecting the heterogeneity and inter-donor variability of CYP expression in human liver (5). This is in agreement with the characterization CYP activity of cryoHepaRG cells and primary human hepatocytes as untreated cells and in response to several prototypic inducers (6).

9. The following values were observed in the validation study:
- Basal and induced CYP enzyme activities in different batches of cryoHepaRG and cryoheps cells.
- CYP enzyme activity average values over plates and laboratory are expressed as pmol specific probe substrate metabolite /mg protein/min.

	Basal Activity CryoHepaRG		
	Batch 1	**Batch 2**	**Batch 3**
CYP1A2	0.8 *	0.7	0.8
CYP2B6	0.9	1.1	0.9
CYP3A4	4.2	2.8	4.4
	Basal Activity Cryoheps		
	Batch 1	**Batch 2**	**Batch 3**
CYP1A2	7.6*	14.7	6.1
CYP2B6	1.3	3.1	2.0
CYP3A4	5.9	20.5	3.8

Based on data generated with the two test methods, new test systems should fall in the following basal activity ranges:

CYP	Basal activity
CYP1A2	0.7-14.7
CYP2B6	0.9-3.1
CYP3A4	2.8-20.5

10. To ensure that a proposed *in vitro* human CYP n-fold induction test method possesses characteristics similar to other validated test methods, the reference substances for testing n-fold CYP induction listed in Table 1 and 2 should be used to demonstrate the reliability and accuracy of the new test method. The reference substances were tested in both the cryoHepaRG and cryoheps test methods. Reference substances should be classified as inducers or non-inducers according to Table 3.

Table 1. List of reference substances (9) for evaluation of human CYP n-fold induction

	substances	Use and Mode of Action	CAS #	MW g/mole	physical state	solubility in water
1	Omeprazole	proton pump inhibitor, acid reducer for treatment of active duodenal ulcer	73590-58-6	345.42	white to off-white powder	slightly
2	Carbamazepine	anticonvulsant and specific analgesic for trigeminal neuralgia	298-46-4	236.27	white to off-white powder	no
3	Phenytoin	anticonvulsant, antiepileptic drug	630-93-3	274.3	White crystalline	yes
4	Penicillin G	Narrow spectrum antibiotic for gram positive aerobic organisms (e.g. Streptococcus)	69-57-8	356.4	crystalline	yes
5	Sulfinpyrazone	uricosuric drug used to reduce the serum urate levels in gout therapy	57-96-5	404.5	white crystalline powder	slightly
6	Bosentan	endothelin receptor antagonist used for treatment of pulmonary arterial hypertension	157212-55-0	569.6	white to yellowish powder	Poorly in water and in aqueous solutions at low pH (0.1 mg/100 ml at pH 1.1 and 4.0; 0.2 mg/100 ml at pH 5.0). Solubility increases at higher pH values (43 mg/100 ml at pH 7.5).
7	Rifampicin	inhibits DNA-dependent RNA polymerase activity. It is a very broad spectrum antibiotic against most gram-positive and gram-negative organisms (including Pseudomonas aeruginosa) and specifically Mycobacterium tuberculosis	13292-46-1	822.9	red-orange powder	Very slightly
8	Metoprolol	cardioselective β1-adrenergic blocking agent used for acute myocardial infarction	51348-51-1	267.4	white crystalline powder	yes
9	Sotalol HCl	non-selective competitive β-adrenergic receptor blocker used as antiarrhythmic	959-24-0	308.8	white, crystalline solid	yes

Table 2: Decision classification of proficiency substances

Proficiency substance	Test concentration range [µM]	LOEL [µM][a]			Decision classification		
		CYP1A2	CYP2B6	CYP3A4	CYP1A2	CYP2B6	CYP3A4
Omeprazole	0.48-115.8	12.9	38.6	38.6	Induction	Induction	Induction
Carbamazepine	0.70-169.3	54.6	2.09	6.27	Induction	Induction	Induction
Phenytoin	0.45-145.8	16.2	0.60	1.80	Induction	Induction	Induction
Rifampicin	0.20-48.6	0.2	0.2	0.2	Induction	Induction	Induction
Sulfinpyrazone	0.41-98.9	3.66	3.66	1.22	N-fold induction	N-fold induction	N-fold induction
Bosentan	0.07-70.2	0.29	0.29	0.07	N-fold induction	N-fold induction	N-fold induction
Metoprolol	0.62-149.6	NA	NA	NA	No induction	No induction	No Induction
Sotatol	0.53-129.5	NA	NA	NA	No induction	No induction	No induction
Penicillin G	0.46-112.2	NA	NA	NA	No n-fold induction	No n-fold induction	No n-fold induction

[a]Lowest F2 value from either CryoHepaRG or Cryoheps.
NA = applicable because the CYP isoform was not induced.
F2 = the concentration (µM) of the test chemical leading to a statistically significant increase of the base line activity by 2-fold of the baseline levels.

11. New similar test methods should be developed on the basis of at least the 9 reference substances. These reference substances should be used to determine equivalence of performance compared to the validated reference test methods.

12. All substances should be tested in a coded/blinded manner. When evaluated using the nine reference substances, the reliability and accuracy (i.e. sensitivity, specificity) of the proposed CYP n-fold induction test method should approximate the following:

Defined reliability and accuracy values

13. For the purposes of establishing the reliability and accuracy of the proposed test method when transferred between laboratories, all 9 reference substances (Table 1, 2 and 7) should be tested in two or (preferably) three laboratories. In each laboratory, all 9 reference substances should be tested in triplicates in three independent runs. Within-laboratory reproducibility and between-laboratory reproducibility should be evaluated on the call for a single curve (at least 2 concentrations with ≥2 and Dunnett test significant), not on one individual concentrations.

Within-laboratory reproducibility

14. For the assessment of within-laboratory reproducibility during the validation study, the concordance of classifications (CYP inducer/CYP non-inducer) obtained in three batches of cryopreserved human primary hepatocytes and cryopreserved human HepaRG cells, when using the 9 reference substances in Table 1, should be based on data generated during the validation study for the two test systems (cryopreserved human primary hepatocytes and the cryopreserved human HepaRG cell line). The reproducibility was based on the following classification rule: if a reference substance

causes n-fold induction above 2 at least 2 of the concentrations tested, the reference substance is classified as positive.

Table 3: Within-laboratory reproducibility (Between-batch reproducibility) based on 2-fold induction threshold classification in cryoheps

	Lab 1	Lab 2	Lab 3
CYP1A2	66%	82%	77%
CYP2B6	78%	60%	60%
CYP3A4	78%	82%	74%

Based on data generated during the validation study on cryopreserved human primary hepatocytes the minimum within-laboratory reproducibility values should be as follows:

CYP1A2 ≥ 66%
CYP2B6 ≥ 60%
CYP3A4 ≥74%

Table 4: Within-laboratory reproducibility (Between-batch reproducibility) based on 2-fold induction threshold classification in cryoHepaRG

	Lab 1	Lab 2	Lab 3
CYP1A2	82%	72%	85%
CYP2B6	75%	75%	78%
CYP3A4	87%	92%	88%

Based on data generated during the validation study on cryopreserved human HepaRG cells the minimum within-laboratory reproducibility values should be as follows:

CYP1A2 ≥ 72%
CYP2B6 ≥ 75%
CYP3A4 ≥87%

The lower reproducibility with cryopreserved human primary hepatocytes is unsurprising given its high variation in expression across individuals and the use of a 2-fold cut-off to define CYP induction (which increases sensitivity to background noise). A higher cut-off (e.g., 5-fold) would probably increase the reproducibility for this enzyme.

Between-laboratory reproducibility

15. To assess between-laboratory reproducibility, the 9 reference substances should be tested in a minimum of three laboratories. The concordance of classifications (CYP inducer/CYP non-inducer) obtained when using the 9 reference substances in Table 1, should be based on data generated during the validation study for the two test systems (cryopreserved human primary hepatocytes and

cryopreserved human HepaRG cell line). The reproducibility was based on the following classification rule: if a reference substance causes induction ≥ 2 at least 2 tested concentrations, the reference substance is classified as positive.

Table 5: Between-laboratory reproducibility based on 2-fold induction threshold classification in cryoheps

	batch 1	batch 2	batch 3
CYP1A2	80%	58%	74%
CYP2B6	67%	37%	63%
CYP3A4	71%	55%	61%

Based on data generated during the validation study on cryopreserved human primary hepatocytes the minimum between-laboratory reproducibility values should be as follows:

CYP1A2 ≥ 58%
CYP2B6 ≥ 37%
CYP3A4 ≥55%

Table 6: Between-laboratory reproducibility based on 2-fold induction threshold classification in cryoHepaRG

	batch 1	batch 2	batch 3
CYP1A2	95%	83%	68%
CYP2B6	82%	75%	70%
CYP3A4	90%	95%	90%

Based on data generated during the validation study on cryopreserved human HepaRG cells the minimum between-laboratory reproducibility values should be as follows:

CYP1A2 ≥ 68%
CYP2B6 ≥ 70%
CYP3A4 ≥90%

In the validation study, reproducibility was lower for the cryopreserved human primary hepatocytes (except batch 3 for CYP1A2) derived from three donors and higher for cryopreserved human HepaRG cells derived from one donor but tested on three different batches.

Predictive capacity (accuracy)

15. The accuracy (sensitivity, specificity, and overall accuracy) of the proposed test method should be comparable to that demonstrated for the fully validated test methods, the cryopreserved human HepaRG cells CYP n-fold induction *in vitro* test method (CryoHepaRG) and the cryopreserved human primary hepatocytes CYP n-fold induction *in vitro* test method (Cryoheps). On the basis of the performance values (sensitivity / specificity) of the validated reference methods (see Table 7) for the 9 reference substances the performance of the new test method should be assessed.

Table 7: Comparison of predictivity to human *in vivo* exposure of two *in vitro* systems (HepaRG and Cryoheps cells) related to CYP enzyme potency: positive and negative is the results from *in vitro*, if it is true or false is whether the results is in agreement with what the *in vivo* results show.

Definition: Sensitivity=True Positive/(True Positive+False Negative)
Specificity = True Negative/(True Negative+False Positive)

CYP3A4	HepaRG				Cryoheps		
Sensitivity	5/5+0=1				5/5+0=1		
Specificity	5/5+0=1				5/5+0=1		
	In vivo Cmax (µM)	Induction *in vitro*	F2 (µM)	Cmax/F2 and classification	Induction *in vitro*	F2 (µM)	Cmax/F2 and classification
Omeprazole	0.68-3.5	++	38,6	<2 TN	++	38.7-117	<2 TN
Carbamazepine	39	++	18,8	2 TP	++	6.27 – 18.	>2 TP
Phenytoin	40-80	++	12,2	>2 TP	++	1.80 – 16.2	>2 TP
Sulfinpyrazone	45	++	11	>2 TP	++	3.66	>2 TP
Bosentan	5.8	++	0.29	>2 TP	++	0.65 - 1.95	>2 TP
Rifampicin	8.0-12.0	++	0.2-0.6	>2 TP	++	0.2	>2 TP
Metoprolol	0.14-0.38	-	nv	TN	+	nv	TN
Penicillin G	36	-	nv	TN	-	112	<2 TN
Sotalol	2	-		TN	-	nv	TN

CYP2B6	HepaRG				Cryoheps		
Sensitivity	4/4+0=1				4/4+0=1		
Specificity	4/4+0=1				4/4+0=1		
	In vivo Cmax (µM)	Induction *in vitro*	F2 (µM)	Cmax/F2 and classification	Induction *in vitro*	F2 (µM)	Cmax/F2 and classification
Omeprazole	0.68-3.5	++	38.6	<2 TN	++	38.6 - 116	<2 TN
Carbamazepine	39	++	18,8	2 TP	++	2.09 – 6.27	>2 TP
Phenytoin	40-80	++	1.35	>2 TP	++	0.60 – 48.6	>2 TP
Sulfinpyrazone*	45	++	11.0-33.0	nv	++	33	nv
Bosentan*	5.8	++	2.60-7.80	nv	++	5.85	nv
Rifampicin	8.0-12.0	++	0.6-1,80	>2 TP	++	0,2	>2 TP
Metoprolol	0.14-0.38	-	nv	TN	-	nv	TN
Penicillin G	36	-	nv	TN	-	nv	TN
Sotalol	2	-	nv	TN	-	nv	TN

CYP1A2	HepaRG	Cryoheps

Sensitivity	3/3+0=1				2/1+2=0.66		
Specificity	5/5+0=1				4/4+1=0.8		
CYP1A2 HepaRG	In vivo Cmax (µM)	Induction in vitro	F2 (µM)	Cmax/F2and classification	Induction in vitro	F2 (µM)	Cmax/F2 and classification
Omeprazole	0.68-3.5	++	12.9	<2 TN	++	38.6-118	<2 TN
Carbamazepine	39	++	54.6	<2 TN	+	No value	FN
Phenytoin	40-80	++	36.5	>2 TP	++	16.2 – 48.6	>2 TP
Sulfinpyrazone	45	++	11	2 TP	+	33.0 – 98.9	2 TP
Bosentan	5.8	++	2.60 – 7.80	nv	+	?	nv
Rifampicin	8.0-12.0	++	0.20 – 0.60	>2 TP	+	No value	<2FN
Metoprolol	0.14-0.38	-	nv	TN	-	nv	TN
Penicillin G	36	-	nv	TN	-	nv	TN
Sotalol	2	-	nv	TN	-	nv	TN

True Positive predictivity (TP), True Negative predictivity (TN), False Negative predictivity (FN)
*No information on *in vivo effects*
F2 = the concentration (µM) of the test chemical leading to a statistically significant increase of the base line activity by 2-fold of the baseline levels. F2 is an *in vitro* value.
Cmax, the maximum plasma (serum) concentration of the drug (here a test chemical, or a reference inducer).
Cmax/F2, a ratio between the maximum plasma concentration of a test chemical and the (lowest) concentration of the test chemical leading to a statistically significant >2-fold increase of the base line enzyme activity (A value >0.5 is considered to predict an induction *in vivo*).

16. Although it is not realistic to expect test methods to perform identically, discordant results should be discussed in terms of the ability of the test method to detect a similar range of potencies and substance/product classes, as demonstrated by the fully validated test methods using human cryohepaRG and human cryoheps as test systems.

LITERATURE

1. OECD (2005), Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment , Series on Testing and Assessment No.34,,OECD, Paris. Available at: http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=ENV/JM/MONO(2005)14&doclanguage=en.

2. ICCVAM (1997), Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM),. NIH Publication No. 97-3981, National In stitute of Environmental Health Sciences, Research Triangle Park, NC.

3.ICCVAM (2003), ICCVAM Guidelines for the No mination and Submission of New, Revised, and Alternative Test Methods,NIH Publication No. 03-4508, Research Triangle Park, NC, National Institute of Environmental Health Sciences

4. OECD(1996), Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods, ENV/MC/CHEM/TG(96)9, OECD, Paris.
Available at:
http://www.oecd.org/document/24/0,3746,en_2649_37465_47858904_1_1_1_37465,00.html
5. Hewitt NJ, Lechón MJ, Houston JB, Hallifax D, Brown HS, Maurel P, Kenna JG, Gustavsson L, Lohmann C, Skonberg C, Guillouzo A, Tuschl G, Li AP, LeCluyse E, Groothuis GM, Hengstler JG. (2007) Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. Drug Metab Rev. 1:159-234
6. Gerets HH, Tilmant K, Gerin B, Chanteux H, Depelchin BO, Dhalluin S, et al. (2012) Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins. Cell Biol Toxicol. 28(2):69-87.