

6.0 **IN VITRO ER TA TEST METHOD PERFORMANCE ASSESSMENT**

6.1 **Introduction**

The ICCVAM Submission Guidelines (ICCVAM, 1999) request information supporting the assessment of test method performance (i.e., accuracy, sensitivity, specificity, positive and negative predictivity, and false positive and false negative rates¹). The ability of the new test method to predict the effect of interest is typically compared to the reference test method currently accepted by regulatory agencies. Where feasible, an assessment is made of the ability of the new method to directly predict adverse health outcomes in the species of interest (e.g., humans, wildlife). Currently, there are no methods accepted by regulatory authorities to assess ER-induced TA, and data on endocrine disruption in humans or wildlife are too limited to be used for this purpose. Thus, the approach taken to evaluating the performance of such assays is to compare the data from existing *in vitro* ER TA assays against each other with regard to their ability to detect ER agonists and antagonists.

6.2 **Quantitative Assessment of Assay Performance**

A quantitative analysis of the relative performance of the approximately 113 *in vitro* ER TA assays (yeast reporter gene assays = ~13 yeast strains with at least 22 assay variants; mammalian cell reporter gene assays = 12 mammalian cell lines with at least 81 assay variants; mammalian cell proliferation assays = three cell lines with at least 10 assay variants) considered in this BRD was not conducted (see ***In Vitro* ER Binding Assay BRD, Section 6**). The major reason was the almost unlimited permutations among *in vitro* ER TA assays in regard to the mammalian cell line or yeast strain used, the nature and source of the ER, the nature and type of the reporter gene (for reporter gene assays), the type of transfection (stable or transient), the experimental protocol for detecting agonistic or antagonistic activity, and the numerous and varied approaches used by the various investigators to express *in vitro* ER TA assay results. These factors, combined with

¹ Accuracy is defined as the proportion of correct outcomes of a method, often used interchangeably with concordance; Sensitivity is defined as the proportion of all positive substances that are correctly classified as positive in a test; Specificity is defined as the proportion of all negative substances that are correctly classified as negative in a test; Positive predictivity is defined as the proportion of correct positive responses among substances testing positive; Negative predictivity is defined as the proportion of correct negative responses among substances testing negative; False positive rate is defined as the proportion of all negative substances that are falsely identified as positive; False negative rate is defined as the proportion of all positive substances that are falsely identified as negative (ICCVAM, 1997).

the relatively few substances tested in multiple assays, precluded a quantitative analysis. The numbers of substances tested for agonism and antagonism activity in various *in vitro* ER TA mammalian cell/yeast reporter gene assays and the mammalian cell proliferation assays are tabulated in **Table 6-1**.

6.3 Qualitative Assessment of Assay Performance

A qualitative comparative assessment of assay performance that considered the relative abilities of the various *in vitro* ER TA assays to identify substances that induced or inhibited TA was conducted. The qualitative assessment was performed separately for *in vitro* ER TA agonism and antagonism test methods. In conducting this assessment, it was assumed that there were no false positive study results. Inspection of the *in vitro* ER TA database (**Appendix D**) suggests that negative calls for some substances in some assays could be the result of limitations in protocol design (i.e., the highest dose tested might have been inadequate) rather than due to intrinsic differences in assay sensitivity. However, no effort was made to account for this possible limitation in the qualitative assessments of assay performance.

Combining the results obtained in different *in vitro* ER TA assays was not possible because of differences in the ability of the cell lines to metabolize hormones and xenobiotics, the source of the ER and type of reporter genes, as well as the possible differences in ER-induced TA levels depending on whether the ER was endogenous or transiently or stably transfected. It has been reported that the intracellular concentration of ER molecules is higher in transiently transfected cell lines than in cell lines expressing the receptor either endogenously or after it has become stabilized. The ERs transfected into the mammalian cell lines and yeast strains were derived from humans, rats, mice, or rainbow trout. In addition, ER that coded for a fusion protein of the binding domain or the ER and ER from human and mouse were used by some investigators. An additional difference between the cell lines is their intracellular concentration of other hormone receptors (e.g., glucocorticoid [GR])

Table 6-1 Number of Substances Tested in Multiple *In Vitro* Mammalian Cell/Yeast ER TA Reporter Gene Assays

	Number of Assays																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
Number of Substances Tested for Agonism	317	127	32	16	12	5	7	2	6	2	2	1	0	1	0	1	2	1	534
Percentage	59.4	23.7	6	3	2.2	0.9	1.3	0.4	1.1	0.4	0.4	0.2	0	0.2	0	0.2	0.4	0.2	100
Number of Substances Tested for Antagonism	76	79	7	4	4	1	0	1	0	0	0	1	1	0	0	0	0	0	174
Percentage	43.7	45.1	4	2.3	2.3	0.6	0	0.6	0	0	0	0.6	0.6	0	0	0	0	0	100

(Table 6-3), which can modulate the level of ER-induced TA. Modulation can occur in a cell if the test substance can bind to the GR and if MMTV response elements are used in the reporter vector, since it has been reported that the ligand-GR complex can interact with these elements.

Table 6-2 Number of Substances Tested in Multiple *In Vitro* ER TA Cell Proliferation Assays

	Number of Assays				
	1	2	3	4	Total
Number of Substances Tested for Agonism	265	43	1	3	312
Percentage	84.9	13.8	0.32	0.96	100
Number of Substances Tested for Antagonism	59	7	1	0	67
Percentage	88.1	10.4	1.5	0	100

Table 6-3 Characteristics of Cells Used in *In Vitro* ER TA Assays

Cell Line	Stable Transfection		Transient Transfection		Steroid Metabolizing Enzymes	Other receptors
	EXP	REP	EXP	REP		
BG-1	*	Yes	No	No		
CHO-K1†	Yes	Yes	Yes	Yes	Metabolize vinclozolin	
COS-1	No	No	Yes	Yes		
ELT-3	No	No	Yes	Yes		Progesterone
HEC-1	No	No	Yes	Yes		Progesterone
HEK293†	Yes	No	Yes	Yes		
HeLa†	Yes	Yes	Yes	Yes		
HepG2	No	No	Yes	Yes		No ER or ER
Ishikawa	No	No	Yes	Yes		Progesterone
MCF-7	*	No	No	Yes		Progesterone
MDA-MB-231	No	No	Yes	Yes		
T47D	*	No	No	Yes		Progesterone
Yeast	Yes	Yes	No	No		
ZR-75	*	Yes	No	No		Progesterone

Abbreviations: EXP = Expression plasmid; REP = Reporter plasmid.

*The ER is endogenous in this cell line.

†In some cases the expression plasmid has been stably transfected and in others, it is transient.

Based on this rationale, the studies were organized by the cell line/type used. The assays were further divided into whether they harbored the hER or hER gene or one of the other ERs used in various studies. The majority of reporter gene assays used luciferase synthesis to assess ER-induced TA; CAT activity was used in the other studies. The data, separated by agonism and antagonism assays, are provided in **Appendix D**.

Since very few substances were tested in multiple assays, test substance responses in mammalian cell reporter gene assays, yeast reporter gene assays, and mammalian cell proliferation assays were grouped separately and a comparison made between these three broad groupings to determine whether there were obvious differences in performance (i.e., positive, negative) based on the type of assay (reporter gene versus cell proliferation) or the target cell population (mammalian versus yeast) (**Appendix E**). For this qualitative analysis, agonism and antagonism responses were considered separately. Only substances tested in at least two of these broad groups of assays were included in the analysis. In terms of testing for ER agonism, a total of 99 substances were tested in at least one mammalian cell and at least one yeast ER TA reporter gene assay; 105 substances were tested in at least one mammalian cell ER TA reporter gene assay and at least one ER TA cell proliferation assay; and 98 substances were tested in at least one yeast ER TA reporter gene assay and in at least one ER TA cell proliferation assay. With regard to testing for ER antagonism, only 12 substances were tested in at least one mammalian cell and at least one yeast ER TA reporter gene assay; 26 substances were tested in at least one mammalian cell ER TA reporter gene assay and at least one ER TA cell proliferation assay; and 19 substances were tested in at least one yeast ER TA reporter gene assay and in at least one ER TA cell proliferation assay. In conducting this qualitative assessment, it was assumed that there were no false positive calls, even in situations where multiple tests were conducted and the number of positive calls was in the minority. This approach was used because of possible limitations in some assays associated with testing substances at relatively low concentrations only, which might have led to false negative results. The agonism assay results are presented in **Table 6-4**; the antagonism assay results in **Table 6-5**.

Table 6-4 Concordance of *In Vitro* ER TA Agonism Assays

	Agonism Assay Response*				Total
	Negative/ Negative	Negative/ Positive	Positive/ Negative	Positive/ Positive	
Mammalian Cell ER TA vs Yeast ER TA Reporter Gene Assays	11	12	9	67	99
Percentage	11.1	12.1	9.1	67.7	100
Mammalian Cell ER TA Reporter Gene vs Cell Proliferation Assays	23	8	25	49	105
Percentage	21.9	7.6	23.8	46.7	100
Yeast ER TA Reporter Gene vs Mammalian Cell Proliferation Assays	11	1	25	61	98
Percentage	11.2	1.0	25.5	62.2	100

*The first and second classifications (negative, positive) in each column refer to the first and second sets of assays, respectively, listed in each row.

When *in vitro* mammalian cell and yeast ER TA reporter gene agonism assays are compared, the results for 21 (21.2%) of the 99 substances tested in common were discordant, with about the same proportion of substances classified as negative/positive or positive/negative in both sets of assays. In contrast, 33 (31.4%) of the 105 substances tested in both mammalian cell ER TA reporter gene and cell proliferation assays were discordant, with the majority of discordant results being associated with a positive response in mammalian cell ER TA reporter gene assays and a negative response in the mammalian cell proliferation assays. A similar finding was observed for yeast ER TA reporter gene assays compared to mammalian cell proliferation assays. Based on this approach, the mammalian cell ER TA proliferation assays do not perform as well as either the mammalian cell or yeast ER TA reporter gene assays, while the two sets of reporter gene assays appear to have about equal performance.

The available data for *in vitro* ER TA antagonism assays are too limited for any conclusion about relative performance to be made.

Table 6-5 Concordance of *In Vitro* ER TA Antagonism Assays

	Antagonism Assay Response*				Total
	Negative/ Negative	Negative/ Positive	Positive/ Negative	Positive/ Positive	
Mammalian Cell vs Yeast ER TA Reporter Gene Assays	3	2	2	5	12
Percentage	25	16.7	16.7	41.7	100
Mammalian Cell ER TA Reporter Gene vs Cell Proliferation Assays	1	3	4	18	26
Percentage	3.8	11.5	15.4	69.2	100
Yeast ER TA Reporter Gene vs Mammalian Cell Proliferation Assays	7	6	0	6	19
Percentage	36.8	31.6	0	31.6	100

*The first and second classifications (negative, positive) in each column refer to the first and second sets of assays, respectively, listed in each row.

This qualitative assessment is confounded by a number of limitations, including:

- The large number of assay permutations used in each category of assays;
- The lack of replicate assay test data for most of the substances considered;
- The lack of a common set of substance tested in multiple assays; and
- The assumption that positive results were more accurate than negative results.

6.4 Performance of *In Vitro* ER TA Assays

The *in vitro* ER TA assays that would be the most useful as screening tests for endocrine disrupting substances are those that are the most sensitive (i.e., have the ability to detect weak agonists and antagonists) and the most reliable within and among laboratories (see **Section 7**). In addition, it might be anticipated that assays that use ER derived from the species of interest (e.g., human for predicting human-related effects, wildlife species for predicting effects in wildlife, experimental animals for predicting the results of *in vivo* test methods) might be the most informative. Since none of these assays required the use of animals, animal welfare is not a

consideration. Finally, when taking human health and safety issues into consideration, assays that do not use radioactivity would have the greatest utility. Only a few investigators that measured CAT activity to assess ER-induced TA utilized radioactivity. However, an ELISA assay for this enzyme is now available, eliminating the need for radioactivity if this reporter gene system is used.

A qualitative evaluation of the responses of the same substances tested in the same laboratory but in different assays indicated that generally the same outcome was obtained (**Appendix F**). The major discordant results are described below.

An agonism study by Connor et al. (1997) suggested that a HeLa cell line harboring a stable hER and a stable luciferase reporter construct was more sensitive to certain hydroxylated polychlorinated biphenyls than an MCF-7 cell line in which the plasmids were transiently transfected. However, it must be noted that besides the difference in transient and stably transfected plasmids, the cell lines and reporter genes differed in these assays. Thus, it is not possible to discern which of these differences contributed to the difference in outcomes.

In an agonism study by Klotz et al (1997), three carbamate pesticides elicited a positive response in MCF-7 cells that had been transiently transfected with the Luc reporter gene but a negative response in Ishikawa cells that had been transiently transfected with the same reporter gene. This finding suggests that MCF-7 cells are more sensitive to these compounds than are Ishikawa cells.

Six xenoestrogen compounds (formononetin, tectoridin, sissotorin, 5-methoxygenistein, 7-methoxygenistein, irisolidone) induced a positive ER TA response in yeast Y190 cells transfected with the hER but not when the same strain was transfected with hER (Morito et al., 2001b). In contrast to these findings, Meyers et al. (1999) reported that the hER was more sensitive to certain chrysene derivatives than was the hER .

With only these few qualitative differences in responses between different assays within a laboratory, it is not possible to identify which assays are consistently more sensitive. Thus,

based on the very limited data available, there is no single reporter gene assay that can be concluded to perform better than any other assay. However, it might be anticipated that mammalian cell reporter gene assays would be preferred over similar yeast assays, simply because of differences in the increased ability of test substances to cross the mammalian cell membrane compared to the yeast cell wall (Krall and Yamamoto, 1996; Gray et al., 1997).

The mammalian cell lines used in the various *in vitro* ER TA reporter gene assays differ from each other in a number of characteristics (**Tables 2-1** and **6-2**). One important difference is whether the cell line contains expression and/or reporter genes that are stable or whether these constructs have to be transfected into the cells prior to each experiment. The COS-1, HEC-1, HeLa, and HepG2 cells used by a number of investigators were all transiently transfected with both expression and reporter plasmids prior to each experiment (**Table 2-1**). Although the MCF-7 cell line contains an endogenous ER, some investigators transfected an intact ER (Ramamoorthy et al., 1997a,b; Charles et al., 2000a,b) or the sequences coding for the binding (def) domain (Fertuck et al, 2001a,b) into this cell line. However, since the cells in which the "def" domains were transfected lacked other regions of the protein that contribute to TA, assays using only the binding domain may not be as useful as those assays using the entire protein.

6.5 Strengths and Limitations of *In Vitro* ER TA Assays

Data from *in vitro* ER TA assays indicate whether a substance can interact with the target receptor which, in turn, binds to responsive elements in DNA that initiate transcription of genes related to hormone-stimulated events in the cell. In contrast to binding assays, the TA assays provide sufficient evidence to conclude whether a substance is an ER agonist or antagonist. However, neither the binding nor the TA assay takes into consideration other cellular or organismic mechanisms that may lead to endocrine disruption (Zacharewski, 1998).

The *in vitro* ER TA reporter gene assays can be important components of a battery of screening tests because they:

- Use eukaryotic cells, many of which are derived from human tissues;
- Are cost-effective;
- Are rapid and relatively easy to perform;

- Are based on an easily quantitated, well-elucidated mechanism of action (i.e., binding to a specific protein and initiating the transcription of ER-responsive genes);
- Can be performed using small amounts of test substances;
- Can be used to test multiple substances simultaneously; and
- Can be easily standardized among laboratories.

The limitations of these assays include:

- The efficiency of transfection for transiently transfected cells can vary from assay to assay and laboratory to laboratory;
- The responsiveness of transiently transfected cells lasts for only a few days (Terouanne et al., 2000); and
- Inability to distinguish between the regulation of gene transcription by binding to DNA versus other mechanisms.

For yeast-based assays, additional limitations include:

- Transfected yeast lines are more prone to genetic drift over time than mammalian cells (Joyeux et al., 1997);
- Transport of test substances through the yeast cell wall might be more difficult than transport through a mammalian cell membrane, increasing the likelihood of false negative results; and
- Yeast cells may have steroid metabolic pathways that differ from mammalian cells (Gaido et al., 1997).

False positive results could occur if the cells are unable to detoxify chemicals that are usually detoxified *in vivo*, or for antagonism studies, by test substance-induced cytotoxicity that is not taken into consideration. False negative results could occur if the cell line used lacks the enzymes present *in vivo* that would normally activate the test substance to a reactive intermediate that then binds to the ER. The metabolic competency of the various cell lines (except for HepG2) is not very well characterized. The addition of the enzymes and cofactors required for metabolic activation to the assay can help to eliminate this limitation. This approach has been used in three studies in which ER-induced TA was assessed (Charles et al., 2000b; Elsby et al., 2001; Sumida et al., 2001).

Other reasons for obtaining a false negative response would be incomplete solubility of the test substance in the medium, or the presence of different coactivators in the different cell lines. Not known or addressed by any investigator is the role of coactivators in the assessment of ER-induced TA using these artificial systems. If these coactivators affect the rate of TA in these systems, the response elicited by the same substance might differ among cell lines.

6.6 Summary, Conclusions, and Recommendations

Relatively few substances have been tested for ER agonism or antagonism by more than one investigator in the same *in vitro* ER TA assay, or in multiple *in vitro* ER TA assays. Consequently, much of the published data are of limited value in terms of a relative analysis of assay performance. This prevents an accurate assessment of the effectiveness and limitations of *in vitro* ER TA assays.

Based on the limited data available, there is no single *in vitro* ER TA assay that can be concluded to perform better than any other assay. However, it might be anticipated that mammalian cell assays would be preferred over yeast assays, simply because of differences in the ability of test substances to cross the mammalian cell membrane compared to the yeast cell wall. Taking various factors into consideration, it would seem that a cell line stably transfected with both hER expression and luciferase reporter plasmids (e.g., T47D) would offer the greatest utility in terms of eliminating the need to continuously prepare multiple batches of transiently transfected cells, thereby eliminating one potential source of interlaboratory variability.

Formal validation studies should be conducted using appropriate substances, covering the range of expected responses for ER agonists and antagonists from strong to weak to negative. Testing of substances encompassing a wide range of agonist or antagonist responses are needed to adequately demonstrate the performance characteristics of any *in vitro* ER TA test method recommended as a screening assay. A list of potential test substances for use in validation efforts is provided in **Section 12**.

[This page intentionally left blank]