

## 6.0 *IN VITRO* AR TA TEST METHOD PERFORMANCE ASSESSMENT

### 6.1 Introduction

The ICCVAM Submission Guidelines (ICCVAM, 1999) request a description of what is known about the performance (i.e., accuracy, sensitivity, specificity, positive and negative predictivity, and false positive and false negative rates<sup>1</sup>) of the proposed test method. The extent to which the new test method predicts or measures the effect of interest is 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 predict adverse health outcomes in the species of interest (e.g., humans, wildlife). Currently, there are no methods accepted by regulatory authorities to assess AR-induced transcriptional activation, and data on endocrine disruption in humans or wildlife are too limited to be used for this purpose. The approach taken to evaluate the performance of AR TA assays in this BRD is a comparison of the data from existing *in vitro* AR TA assays against each other with regard to their ability to detect AR agonists and antagonists.

### 6.2 Quantitative Assessment of Assay Performance

For a number of reasons, a quantitative analysis of the relative performance of the 18 *in vitro* AR TA assays considered in this BRD could not be conducted (see *In Vitro ER Binding Assay BRD, Section 6*). The reasons included the limited number of substances tested within and across different assays, the lack of quantitative data for substances that had been tested, and the numerous and varied approaches used by different investigators to express *in vitro* AR TA assay results, particularly from agonism studies. Agonism data was reported as the maximum fold increase compared to the concurrent control, relative activity compared to the reference androgen, or the EC<sub>50</sub> value. Antagonism data was reported as relative activity compared to the reference androgen alone, or as an IC<sub>50</sub> value. The numbers of compounds tested for agonism

---

<sup>1</sup> 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).

**Table 6-1** Number of Substances Tested in Multiple *In Vitro* AR TA Assays

	Number of Assays										
	1	2	3	4	5	6	7	8	9	10	11
Number of Substances Tested for Agonism Activity <sup>a</sup>	81	20	7	3	3	6	0	0	0	1	1
Percentage of substances	65.3	16.1	5.6	2.4	2.4	4.8	0	0	0	0.8	0.8
Number of Substances Tested for Antagonism Activity	59	16	3	5	2	1	0	0	1	0	0
Percentage of Substances	67.8	18.4	3.4	5.7	2.3	1.1	0	0	1.1	0	0

	Number of Assays								Total
	12	13	14	15	16	17	18		
Number of Substances Tested for Agonism Activity <sup>a</sup>	1	0	0	1*	0	0	0	124	
Percentage of substances	0.8	0	0	0.8	0	0	0	100	
Number of Substances Tested for Antagonism Activity	0	0	0	0	0	0	0	87	
Percentage of Substances	0	0	0	0	0	0	0	100	

\*This substance is DHT.

<sup>a</sup>Includes the cell proliferation assay performed by Sonnenschein et al. (1989).

and antagonism are tabulated in **Table 6-1**. The type of reference androgen (DHT, mibolerone, R1881, testosterone) used for the antagonism assays was not considered in compiling the number of substances tested in each assay.

### 6.3 Qualitative Assessment of Assay Performance

A qualitative comparative assessment of assay performance was conducted that considered the relative ability of the various *in vitro* AR TA assays to identify substances that induced or inhibited transcriptional activation. In conducting this assessment, it was assumed that there were no false positive study results. The qualitative assessment was performed separately for AR TA agonism and antagonism test methods. Inspection of the *in vitro* AR 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 assessment of assay performance.

To maximize the numbers of substances available for consideration during the qualitative assessment, data from different studies were combined where possible after taking into account the cell line, the source of the AR, the specific ARE on the reporter construct, and the reporter gene used. The major difference between the various *in vitro* AR TA assays used in the different studies was the cell line, and this criterion was used as the primary basis for combining or not combining data from different laboratories. Within each of the different cell lines, the AR was either transiently or stably transfected or was endogenous. Various cell lines differ in their ability to metabolize hormones and xenobiotics, as well as in their intracellular concentration of other hormone receptors (e.g., glucocorticoid, progesterone) (**Table 6-2**). These receptors can interfere with the binding of certain substances (e.g., medroxyprogesterone acetate) to the AR and subsequent transcriptional activation (Poulin et al., 1991). Differences in the metabolic capabilities of the cell lines are exemplified by the presence of two enzymes, 17 $\alpha$ -oxidase and 5 $\alpha$ -reductase, which metabolize testosterone and testosterone-like compounds, and are present in CV-1 and HeLa cells but not in CHO cells (Deslypere et al., 1992).

**Table 6-2 Characteristics of Cell Lines Used in *In Vitro* AR TA Assays**

Cell Line	Stable Transfection		Transient Transfection		Level of AR (fmol/mg protein)	Steroid Metabolizing Enzymes	Other Receptors
	EXP	REP	EXP	REP			
CHO	No	No	Yes	Yes		Metabolize vinclozolin	
CV-1	No	No	Yes	Yes	30	17 -oxidase 5 -reductase	No GR
HepG2	No	No	Yes	Yes			No ER and ER
MDA-MB-453	Yes*	No	No	Yes**	240		GR present; Very low level of ER
MDA-MB-453-kb2	Yes	Yes	No	No	240		GR present; Very low level of ER
PC-3	No	No	Yes	Yes	1200		
PALM	Yes	Yes	No	No	1200		
Yeast	Yes	Yes	No	No			

Abbreviations: EXP = Expression plasmid; REP = Reporter plasmid; GR = Glucocorticoid receptor; ER = Estrogen receptor.

\*The AR is endogenous in this cell line. \*\*Reporter introduced into cells by viral transduction.

Qualitative analysis was performed on each of the assays described in **Section 5**:

CHO hAR(S)+Luc(S); CHO hAR(T)+Luc(T); CHO-K1 hAR(S)+Luc(S); CHO-K1 hAR(T)+Luc(T)+EGFP(T); CHO hAR(T)+CAT(T)+ -gal(T); CV-1 hAR(T)+Luc(T)\*(transduced); CV-1 hAR(T)+CAT(T); CV-1 hAR(T)+Luc(T); MDA-MB-453 hAR(E)+Luc(T)\*(transduced); MDA-MB-453-kb2 hAR(E)+Luc(S); HepG2hAR(T)+Luc(T)+ -gal(T); PALM hAR(S)+Luc(S); PC-3 hAR(T)+Luc(T); Yeast (*S.cer*) hAR(S) + -gal(S); and LnCaP-FGC hAR(E)+CP. Excluded from the qualitative analysis were the two studies that did not use the hAR (Van Dort et al., 2000; Takeo and Yamashita, 2000), the HeLa cell-based assay used by Wang and Fondell (2001), and any substance not tested in at least two different assays. The HeLa cell-based assay was excluded because only four substances had been tested in one laboratory. The resulting data, separated by agonism and antagonism assays, are provided in **Appendix E**.

A total of 43 substances were tested for agonism activity in at least two of the fifteen *in vitro* AR TA assays considered during the qualitative assessment. In conducting this assessment, it was

assumed that there were no false positive calls in the published literature even in situations where multiple tests were conducted and the number of positive calls was in the minority. The primary limitation associated with this approach is that the substance might truly be negative for AR agonist or antagonist activity (i.e., the positive call was incorrect). Based on this approach, the results obtained using the CHO assays were the most frequently discordant (i.e., a negative response was obtained for nine substances that tested positive in another assay(s), which is 25.7% of the 35 substances tested in this assay). The LnCaP-FGC hAR(E)+CP cell proliferation assay was discordant for one of 10 substances (10%) that tested positive in at least one other assay, and the yeast-based assays were discordant for one of 17 substances (5.8%) that tested positive in at least one other assay. There was no discordance among the responses obtained for substances tested in common among the CV-1, HepG2, MDA-MB-453, and PC-3 cell-based assays.

A total of 28 substances were tested for antagonism in at least two of the 11 *in vitro* AR TA assays considered for the qualitative assessment. The yeast-based assay was discordant for one of three substances (33%) that tested positive in at least one other assay, the CHO-K1 hAR(T)+Luc(T)+EGFP(T) was discordant for one of three substances (33%), the HepG2 assay was discordant for one of 12 substances (8%), and the CV-1 hAR(T)+Luc(T) assay was discordant for one of 17 substances (6%).

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

- The very limited size of the database;
- The lack of replicate 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* AR TA Assays**

The *in vitro* AR 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 acting agonists and antagonists) and the most reliable (i.e., exhibit the lowest variance) within and across laboratories (see **Section 7**). In addition, it might be anticipated that assays that use

AR derived from the species of interest (e.g., human for predicting human-related effects, wildlife species for predicting effects in wildlife) might be the most informative. Since none of these assays use animals, animal welfare is not a consideration. Finally, when taking human health and safety issues into consideration, assays that do not use radioactivity might have the greatest utility. Only the CAT assay, which was used much less frequently than luciferase-based assays, utilized radioactivity. However, an ELISA assay for this enzyme is now available, eliminating the need for radioactivity if this reporter gene system is used.

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

The cell lines used in the various *in vitro* AR TA 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. Except for two of the CHO-based assays, the remaining CHO, CV-1, and HepG2 cell lines used in the majority of *in vitro* AR TA assays were all transiently transfected with expression and reporter plasmids prior to each experiment (**Table 2-1**). The LnCaP-FGC and MDA-MB-453-kb2 cell lines contain a functional endogenous hAR gene (Sonnenschein et al., 1989; Lambright et al, 2000), while a stably transfected cell line (PALM) was developed from the PC-3 cell line (Terouanne et al., 2000; Schrader and Cooke, 2000).

Two different approaches were used to incorporate the reporter construct into the MDA-MB-453 cell line. In one approach, cells were transduced before each experiment with a reporter gene (luciferase) by integrating the reporter and ARE into an infective but nonreplicative adenovirus (Hartig et al., 2002). In a different approach, cells were transfected with a reporter construct and, following antibiotic selection, a clone (MDA-MB-453-kb2) with a stably transfected reporter gene was isolated (Wilson et al., 2002). This cell line has both the expression and reporter

constructs stably integrated into the genome. Selection and expansion of the clone resulted in a cell line that could be used for many passages to measure AR TA. From passages 1-10, luciferase induction by 1 nM of DHT was 10-fold compared to control (Wilson et al., 2002). Over 30 to 40 passages, the fold induction decreased to 5 to 6 fold but then stabilized and remained at this level out to 80 passages.

### **6.5 Strengths and Limitations of *In Vitro* AR TA Assays**

Data from *in vitro* AR TA assays indicate whether a substance can interact with the target receptor which, in turn, binds to responsive elements in the 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 agonist or an antagonist. However, neither assay takes into consideration other mechanisms of action that may lead to endocrine disruption (Zacharewski, 1998). The TA 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 AR-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 potential generation of false positive and false negative results;
- The efficiency of transfection for transiently transfected cells can vary from assay to assay; and
- The responsiveness of transiently transfected cells lasts for only a few days (Terouanne et al., 2000).

For yeast-based assays, additional limitations include:

- 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 accounted for. Another reason for false positives is induction of the reporter by a mechanism not involving AR activation. This could occur if the MMTV is used as the promoter in the reporter gene construct and the cells used for the assay contain a glucocorticoid, progesterone, or unknown receptor that can activate the ARE. 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 AR. The metabolic competency of the various cell lines (except for HepG2) is not very well characterized. The addition of the enzymes and co-factors required for metabolic activation to the assay can help to eliminate this limitation. This approach has been used in two studies in which ER-induced transcriptional activation was assessed (Charles et al., 2000; Sumida et al., 2001). Another reason for obtaining a false negative response would be incomplete solubility of the test substance in the medium.

## 6.6 Summary and Conclusions and Recommendations

Relatively few substances have been tested in more than one laboratory using the same *in vitro* AR TA assay. Also, few of the same substances have been tested for agonism or antagonism in different *in vitro* AR TA assays. Furthermore, because the primary focus of many of the investigations using *in vitro* AR TA assays has been to understand the process of AR-induced transcriptional activation and not to identify substances that act as AR agonists or antagonists, 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* AR TA assays.

Based on the limited data available, there is no single *in vitro* AR TA assay that can be concluded to perform better than any other assay. However, it might be anticipated that mammalian cell-based assays would be preferred over yeast-based assays, simply because of differences in the ability of test substances to cross the mammalian cell membrane compared to the yeast cell wall. Taking other factors into consideration, it would seem that a cell line with endogenous hAR and stably transfected with a luciferase reporter plasmid (e.g., MDA-MB-453-kb2 hAR(E)+Luc(S)) would offer the greatest utility in terms of eliminating the need to continuously prepare multiple batches of transiently transfected cells, while being the most relevant and sensitive. Due to patents held by a private company, some of the CV-1 cell lines transfected with the AR as described in this BRD may not be available to testing laboratories and, thus, they cannot be recommended for use in a screening assay.

Formal validation studies should be conducted using appropriate substances, covering the range of expected responses for agonist and antagonist from strong to weak to negative. Testing of substances encompassing a wide range of agonist/antagonist responses are needed to adequately demonstrate the performance characteristics of any *in vitro* AR 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]*