European Commission

Validation Study Report
Performance assessment of the AR-CALUX® *in vitro* method

to support the development of an international test guideline for Androgen Receptor Transactivation Assays (ARTA) for the detection of compounds with (anti)androgenic potential

Joint Research Centre

EUR 30044 EN

This publication is a Validated Methods, Reference Methods and Measurements report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The validation study of the AR-CALUX® *in vitro* test method was coordinated by Anne Milcamps on behalf of the JRC's EU Reference Laboratory for alternatives to animal testing (EURL ECVAM).

The AR-CALUX® test method is included in OECD Test Guideline 458 on Androgen Receptor TransActivation assays.

Contact information
European Commission, Joint Research Centre (JRC), Chemical Safety and Alternative Methods Unit (F3)
Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
JRC-F3-ENQUIRIES@ec.europa.eu

EU Science Hub
https://ec.europa.eu/jrc

JRC116993

EUR 30044 EN

PDF	ISBN 978-92-76-15037-4	ISSN 1831-9424	doi:10.2760/60314
Print	ISBN 978-92-76-15036-7	ISSN 1018-5593	doi:10.2760/89293

Luxembourg: Publications Office of the European Union, 2020

How to cite this report:
Milcamps, A., Liška, R., Langezaal, I., Casey, W., Dent, M. and Odum., J., *Validation Study Report: Performance assessment of the AR-CALUX® in vitro method*, EUR 30044 EN, Publication Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-15037-4, doi:10.2760/60314, JRC116993.

Acknowledgements

We thank Tom Cole and Salvador Fortaner Torrent for the management of the chemicals and solubility testing; Jan de Lange and Emilio Mendoza for laboratory assistance; Elisabeth Joossens for structural input; Julien Burton for the initiation of the chemical selection; David Asturiol Bofill for the mapping of the chemical space; Harrie Besselink (BDS) for assistance with the SOP, DAFs and troubleshooting; Sandra Coecke for support as EU-NETVAL coordinator; Dave Allen and Patricia Ceger (NIH) for assistance during the VMG decisions; Sylvia Casati, João Barroso and Sharon Munn for reviewing the validation study report; all contributors of the four participating laboratories: Emma Peetersen, Kristina Fant, Jenny Johansson, Lovisa Bodin, Therese Andersson (RISE)*, Mylène Valin, Rachel Lercier, Mégane Auvray, Kelly Chevalier, Pascal Bogdanowicz, Baptiste Coulon, Marion Godefroy, Cédric Gerbeix (CitoxLAB)*, Jo Hubbard, Jo Kelsall, Leslie Akhurst, John Carter (ENVIGO)*, Harrie Besselink, Matthijs Naderman (BDS).

Warren Casey, Matthew Dent and Jenny Odum were part of the AR-CALUX® validation management group and provided expert input, Ingrid Langezaal was responsible for the SOP and the DAFs, Roman Liska provided data analysis and drafted the statistical report, Anne Milcamps coordinated the validation study and drafted the main body of the validation study report.

We also thank the members of the AR-CALUX® ESAC working group for the independent peer review: Ian Cotgreave, Rebecca Clewell, Miriam Jacobs, Hajime Kojima, Jose Navas Anton, Aldert Piersma.

*Note: During the course of the validation study, the participating laboratories changed name: SP became RISE, Huntingdon became ENVIGO. After the validation study (2019) ENVIGO became Covance, CitoxLAB became Charles River Labs.

Version control

Version	Date	Expert panel
V01	10/04/2019	Approved by VMG AR-CALUX®
V02	28/06/2019	ESAC WG AR-CALUX®
		Revisions after peer-review comments and questions:

		• Indication of the name change of the participating laboratories during the validation study
		• Overview of the time lines of the studies per laboratory
		• Comparison of obtained variability values to Japan ARTA % CV values
		• Potency ranking of the tested chemicals
		• Indication % CV of 26 additional chemicals
		• Mapping of the tested chemicals vs regulatory chemicals (REACH)
		• Reformulation of the definition and description of R^2
		• Additional information from the test method developer:
		  o Cell line characterisation regarding receptor expression and metabolism
		  o In light of the guiding principles on good practices for the availability of protected elements in OECD TGs, adopted at OECD WNT April 2019, information on parental cell line origin and authorisation for commercialisation

List of abbreviations

AR: Androgen Receptor

ARTA: Androgen Receptor TransActivation

Japan ARTA: developed by Japan, using the AR-EcoScreen™ Chinese hamster ovary cell line. TG 458 adopted

Korea ARTA: developed by Korea, using the 22Rv1/MMTV GR$^-$ Human prostate cancer cell line. Validation finalised and currently under peer-review (05/2019).

AR-CALUX® method: the transactivation *in vitro* method to measure (anti)androgenic potential of chemicals using the AR-CALUX® cells

ARE: Androgen Responsive Elements

BLR: Between Laboratory Reproducibility

CV: Coefficient of Variation

DMSO: DiMethyl SulfOxide

EC: European Commission

$EC_{10\ and}\ EC_{50}$: 10% and 50% effective concentration

ED: Endocrine Disruptor

EJ: Expert Judgement

ER: Estrogen Receptor

ERTA: Estrogen Receptor TransActivation

ESAC: EURL ECVAM's Scientific Advisory Committee

EU-NETVAL: European Union Network of Laboratories for the Validation of Alternative Methods

EURL ECVAM: European Union Reference Laboratory for Alternatives to Animal Testing

GD: Guidance Document

GLP: Good Laboratory Practice

IC_{50}: half maximal inhibitory concentration

ICATM: International Cooperation on Alternative Test Methods

ICCVAM: Interagency Coordinating Committee on the Validation of Alternative Methods

LDH: Lactate DeHydrogenase

MSDS: Material Safety Data Sheet

NCP: National Contact Point

OECD: Organisation for Economic Cooperation and Development

PBTG: Performance Based Test Guideline

PC_{10} and PC_{50}: Concentration of the test item at which there is 10% or 50% response relative to the reference item

RI: relative induction

REF RPC_{10} and REF RPC_{80}: Response of the reference item at 10% or 80% induction

SC: Solvent Control

Y_c: Response of the test item at concentration c

S_c: Response of the test item's specificity control at concentration c

R^2: the square of the correlation coefficient

DHT: DiHydroTestosterone

STTA: Stably Transfected Transactivation Assay

SOP: Standard Operating Procedure

STR: Short Tandem Repeats

TA: TransActivation

TG: Test Guideline

ToR: Terms of Reference

UVCBs: chemical substances of Unknown or Variable composition, Complex reaction products and Biological materials

VC: Vehicle Control

VMG: Validation Management Group

VMG-NA: OECD Validation Management Group Non Animal

WLR: Within Laboratory Reproducibility

TABLE OF CONTENTS

Contents

1 SUMMARY

The AR-CALUX® *in vitro* method was submitted by the Dutch company BioDetectionSystems (BDS) to EURL ECVAM to be considered for a EURL ECVAM coordinated validation study. The method is applied to the detection of compounds with (anti)androgenic potential. The AR-CALUX® cells are osteosarcoma U2OS cells which are transfected with the cDNA of a human androgen receptor and a luciferase encoding reporter gene preceded by androgen receptor (AR) response elements (ARE), hence responding to chemicals with androgenic activity. The method allows the detection of chemicals with an agonist or/and an antagonist behaviour.

The objectives of the study included assessing the reproducibility (within and between laboratories) and the relevance of the *in vitro* method, leading to the drafting of an OECD Test Guideline. EURL ECVAM is both sponsor and coordinator for this validation study. Three test facilities were selected from the European Union Network of Validation Laboratories for alternative methods (EU-NETVAL) to participate in this validation study: RISE, ENVIGO and CitoxLAB. The test method submitter BDS produced an additional data set for the assessment of the between lab reproducibility.

Reproducibility was evaluated on a set of 20 coded test chemicals based on concordance of classifications. Classification was carried out by the 4 laboratories, applying faithfully the classifier which led to the conclusion of "Positive", "Negative" and "Inconclusive". The latter classification was given when the test chemical displayed activity for only one tested concentration above (agonist) or below (antagonist) threshold values as determined in the classifier. This occurred with an overall frequency of 0.9 % in the agonist assay, and, 7.4% in the antagonist assay. Analysis of the data generated by the 4 laboratories showed a within lab reproducibility (WLR) of 89%, 95%, 100% and 100% for the agonist assay, and, 75%, 80%, 85% and 90% for the antagonist assay. The overall between lab reproducibility (BLR) scored via concordance of classifications was 100% for the agonist assay and 87.5% for the antagonist assay.

In order to arrive at a clear classification of "Positive" or "Negative" for a given test chemical, the classifier was reformulated and guidance included in the Standard Operating Procedure (SOP) for handling borderline situations of only the highest tested concentration displaying an activity passing the threshold values. An approximation of the results, when applying this new guidance in the SOP and the new classifier to the results of the validation study, is shown and discussed. This would lead to increased WLR (94.7% to 100%) and BLR (100%) for both agonist and antagonist testing.

The variability within the measurements (% CV of log of EC_{50}, PC_{10}, IC_{50}, PC_{80}) was less than 2.5% and comparable to what is reported for the ER-CALUX® *in vitro* method validation and the AR STTA (cell line EcoScreen™) validation.

A comparison was made of the classifications in the AR-CALUX® validation study for 46 tested chemicals with reliable published or publicly available ARTA classifications. For 23 tested chemicals a 94.6% concordance was found with the ICCVAM AR-Reference list (2017). Performance values were calculated versus this AR-Reference list. For the additional 23 tested chemicals, an overall comparison was performed with classifications from two Tox21 assays and the AR-pathway computational model. This revealed that for all chemicals tested with the AR-CALUX® method there is an identical classification with at least one other ARTA.

Having reviewed the data, the VMG concluded that the AR-CALUX® method is a reliable test method. It is nevertheless opportune to provide a warning in the SOP for potential interferences. The response to the vehicle control should be assessed to ensure no interference from glass or plastic ware before running studies.

The VMG is of the opinion that the AR-CALUX® method merits proposal to OECD for the development of a test guideline.

Important note:

This validation report is best read together with the statistical report. Whereas the structure of the validation report follows the modules of the modular approach to validation, the statistical report gives the overall analysis of all data obtained in the study. By clicking on a specific test chemical in the index of the statistical report, one is guided immediately to the corresponding data analysis and graphs of the dose responses obtained in the studies.

Due to size constraints, the statistical report and the final version of the SOP (version 07) are not part of this report. They can be found as individual files at EURL ECVAM's Tracking system for alternative methods towards regulatory acceptance (TSAR) (https://tsar.jrc.ec.europa.eu/test-method/tm2010-07)

2 INTRODUCTION AND OBJECTIVES

2.1 General introduction

Endocrine disruptors (EDs) are a high priority topic on the agenda of several national and international governmental institutions given the observed and documented endocrine-related adverse effects on human and animal health (UNEP WHO, 2013). These substances impact development and reproduction by disturbing the functioning of the endocrine (hormone) system.

National and international governments are in the process of establishing testing programmes and strategies to assess the safety of currently used chemicals with regard to their potential to interfere with the endocrine system. Several pieces of European legislation address EDs: chemicals Regulation 1907/2006 'REACH'; Regulation 1107/2009 on Plant Production Products (PPP); Regulation 1223/2009 on Cosmetic Products; Regulation 528/2012 on Biocidal Products (BP). The European Commission launched a work programme entitled "Community strategy for endocrine disruptors" (EC, 1999) addressing several actions, e.g. to establish criteria to identify EDs for further evaluation, to develop and validate test methods to assess EDs, to fund research for understanding the ED mechanisms and to adapt present EU legislation to take account of ED effects. Scientific criteria to identify an ED under the PPP and BP regulations were established and published in 2017 and 2018 (EC, 2017; EC, 2018) and a guidance document for the implementation of the criteria was published by the European Chemicals Agency (ECHA) and the European Food Safety Authority (EFSA) (EFSA, 2018). The United States Environmental Protection Agency (EPA) developed the Endocrine Disruptor Screening Program (EDSP) as one of the first national programmes. In Japan, the Japan Environment Agency initiated the Strategic Programs on Endocrine Disruptors (SPEED) to promote e.g. test method development, while in the republic of Korea, the relevant ministries developed mid and long term research plans mainly dealing with environmental monitoring (Hecker and Holler, 2011; WHO, 2014).

The Organisation for Economic Co-operation and Development (OECD) recognised the potential impact of ED substances on human health and the environment. Since 1996, effort has been made in developing Test Guidelines (TG) and other tools to support member countries' needs for testing and assessment of chemicals for endocrine disruption. In terms of providing overview and guidance, a number of important documents have been drafted, e.g. Guidance document No. 150 on the assessment of chemicals for ED was developed in 2012 as a tool to support regulatory bodies by helping to interpret assay results and suggesting additional studies for reducing uncertainty. It was updated in 2018 (OECD, 2018). GD 150 also contains the Conceptual Framework for screening and testing EDs (CF) which was adopted in 2002, and revised in 2011 and 2017. It lists OECD TGs and standardized test methods available, under development or proposed to be used to evaluate chemicals for ED. It is structured over different levels where level 2 includes *in vitro* assays (data about selected endocrine mechanism) and levels 3, 4 and 5 include *in vivo* assays (data about selected endocrine mechanisms and/or adverse effects on endocrine-relevant endpoints).

The OECD CF as well as the US EPA have adopted *in vitro* assays as regulatory test guidelines, in order to measure hormone receptor binding and transactivation, for the fast screening of putative EDs for prioritisation purposes and for identifying endocrine activity. Within the set of transactivation assays, the Androgen Receptor Transactivation Assays (ARTAs) incorporate androgen receptors (AR) that, once activated, initiate transcription.

AR-CALUX® cell lines were identified within the EU funded project ReProTect (LSHB-CT-2004-503257) which aimed at optimizing an integrated set of tests as a basis for a reproductive/developmental battery, in order to provide detailed understanding of the main tissues or biological mechanisms which could be targeted and disrupted by toxicants across different stages of reproduction. The AR-CALUX® cell lines were considered as addressing a critical biological mechanism (androgen receptor interaction) and therefore a relevant test system for the development of a method for ED screening. The test developer of the AR-CALUX® method carried out a pre-validation (van der Burg *et al*, 2010) and the *in vitro* method was subsequently submitted to EURL ECVAM (Dec. 2011) for a validation process. EURL ECVAM reviewed the submission in 2012 with a favourable outcome for entering into a formal validation study.

2.2 Goal and Objectives

EURL ECVAM launched a validation study of the AR-CALUX® method in 2014 with the overall goal, provided the validation study would be successful, of proposing the test method to OECD to become a test guideline (TG). The European Commission submitted in 2012 a Standard Project Submission Form (SPSF) to OECD for the development of a Performance Based Test Guideline (PBTG) on ARTAs. This was accepted and inserted in the OECD 2013 work plan. Several ARTAs would be considered to be annexed to the PBTG:

- the AR-STTA of Japan, using the AR-EcoScreen™ Chinese hamster ovary cell line (validation finalised and TG 458 adopted)
- the ARTA of Korea, using the 22Rv1/MMTV GR$^-$ Human prostate cancer cell line (validation finalised, report under review)
- the ARTA of the Netherlands, AR-CALUX®, subject of this validation report

The objectives of this validation study were as follows:

- to evaluate the transferability and reliability (reproducibility within and between laboratories)
- to evaluate the relevance of the test method by comparison of the classifications by this test method to reported classifications of the ICCVAM AR-reference list and of other ARTAs.

3 MANAGEMENT OF THE VALIDATION STUDY

The organisation and conduct of the study was performed in compliance with the principles laid down in the OECD guidance document on test method validation GD 34 (OECD, 2005).

3.1 Sponsor

EURL ECVAM was the sponsor and coordinator of the validation study.

Table 01: Sponsor	Address
Maurice Whelan Email: Maurice. Whelan@ec.europa.eu	The EU Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) European Commission Joint Research Centre Directorate F3 Via E. Fermi, 2749 I-21027 Ispra, Italy

The EU-NETVAL laboratories had the possibility to ask for support to their respective governments under Directive 2010/63/EU on "The protection of animals used for scientific purposes". EURL ECVAM provided the test system, test chemicals and training.

The overall study coordination was conducted by EURL ECVAM. This included the organisation of all necessary VMG meetings and teleconferences, SOP technical and experimental assessment, acting as contact point for the participating laboratories for day to day work and troubleshooting, contact point for the method developer, the maintenance of the data sharing platform (CIRCABC) for all data and document storage and retrieval, all data handling and analysis, as well as producing the draft statistical report and the draft validation report.

3.2 Validation Management Group (VMG)

The VMG was established by EURL ECVAM and consisted of three external experts in the field, the validation study coordinator and the biostatistician from EURL ECVAM. VMG's role was to provide oversight on the validation study. Its responsibilities, defined in a Terms of Reference (ToR), included the following:

- To review and approve the validation project plan in all its components (objectives, validation study design, organisation, statistical analysis methods, list of chemicals to be tested, the SOP to be implemented);
- To monitor progress through setting key milestones and reviewing the results of the test facilities and to provide assistance in troubleshooting when need be;
- To manage deviations to the validation study;
- To interpret the validation results and formulation of conclusions;
- To assist, review and approve the validation report;
- To assist in the drafting of the EURL ECVAM recommendation and the TG for ARTAs.

Table 02: Composition of the AR-CALUX® VMG		
Name	**Role and expertise**	**Affiliation**
EURL ECVAM members		
Anne Milcamps	Coordinator	EURL ECVAM, Ispra, Italy Email: anne.milcamps@ec.europa.eu
Roman Liska	Data analysis	EURL ECVAM, Ispra, Italy Email: roman.liska@ec.europa.eu
External members		
Warren Casey	Director NICEATM Expertise in toxicology, EDs, validation	NIEHS/NICEATM, Research Triangle Park, North Carolina, USA Email: warren.casey@nih.gov
Matthew Dent	Safety Science Leader Expertise in general/reproductive toxicology, risk assessment	Unilever, Colworth Science Park, Bedford, UK Email: matthew.dent@unilever.com
Jenny Odum	Independent consultant toxicologist Expertise in toxicology, EDs, validation	Stockport, UK Email: jenny.odum@regulatoryscience.com

3.3 Participating Test Facilities

Three laboratories were identified from EU-NETVAL, EURL ECVAM's network of expert laboratories that was established in January 2014. The network is comprised of 37 laboratories from the EU member countries as well as EURL ECVAM's own laboratory. Its tasks and responsibilities are formulated within a ToR with primary focus on participation in validation studies. In the beginning of 2014, an invitation was launched to the EU-NETVAL members for participation in the AR-CALUX® validation study. This Call for expression of interest included a description of the different tasks to be undertaken when participating to the validation study. Essential requirements had to be met in order for a laboratory to participate. Several test facilities met the requirements and were ranked on the basis of a communicated selection procedure. The 3 highest ranked facilities were approved by the EU Member States via the National Contact Points. For details of the Call see Annex 13.7.11. Commitment of the laboratories was ensured through a Collaboration Agreement / Licence Agreement (+) (January 2015 / July 2018(+)).

Table 03: The participating laboratories

1	**RISE*** EU-NETVAL member / GLP laboratory	Sweden
	Address: RISE Research Institutes of Sweden, Brinellgatan 4, SE-501 15 Boras	
	Laboratory manager: Benny Lyven Study director: Emma Pedersen/Kristina Fant Study personnel: Jenny Johansson, Therese Andersson, Lovisa Bodin	Email: benny.lyven@ri.se Email: emma.pedersen@ri.se; kristina.fant@ri.se
2	**CitoxLAB*** EU-NETVAL member / GLP laboratory	France
	Address: CitoxLAB, BP 563, 27005 Evreux Cedex	
	Laboratory manager: Olivier Foulon Study director: Mylene Valin/ Kelly Chevalier/ Cedric Gerbeix Study personnel: Rachel Lercier, Megane Auvray, Pascal Bogdanowicz, Baptiste Coulon, Marion Godefroy	Email:Mylene.Valin@fr.CitoxLAB.com; Cedric.Gerbeix@fr.CitoxLAB.com; Kelly.Chevallier@fr.CitoxLAB.com
3	**ENVIGO*** EU-NETVAL member / GLP laboratory	UK
	Address: ENVIGO, Woolley Road, Alconbury, Huntingdon, Cambridgeshire	
	Laboratory manager: L. Akhurst, J. Carter Study director: Joanne Hubbard Study personnel: Joanne Kelsall	Email: leslie.akhurst@ENVIGO.com, john.carter@ENVIGO.com Email: joanne.hubbard@ENVIGO.com
4	**BDS** Test method developer (non EU-NETVAL member) (+)	Netherlands
	Address: BioDetection Systems BV (BDS), Science Park 406, 1098 XH Amsterdam	
	Laboratory manager: Harrie Besselink Study director: Harrie Besselink Study personnel: Matthijs Naderman	Email: Harrie.Besselink@bds.nl

*Note: During the course of the validation study, the participating laboratories changed name: SP became RISE, Huntingdon became ENVIGO. In 2019, ENVIGO became Covance, CitoxLAB became Charles River Labs.

The laboratory BDS (test method developer) was included in June 2018 as a participating laboratory. The facility CitoxLAB observed frequently high RLUs, first during Study 1 (Transfer phase) which were thought to be resolved, but continued during Study 2 (BLR). The test method developer visited the laboratory (2018) to assist in discovering the source of the high RLUs but a concrete indication was not found and the technical issues remained (Report on the technical issues can be found in Annex 13.7.9). These issues did not occur in the other laboratories and could be considered as specific to one laboratory only (see section 6.5). Given the uncertainty of achieving a full and valid third set of data to determine BLR, the VMG agreed to ask the laboratory BDS (test method developer) to deliver the third required set of data to evaluate BLR. CitoxLAB continued investigating the technical issues and succeeded in producing a full set of acceptable data within the deadline. For the following modules, data sets of both laboratories BDS and CitoxLAB are included.

The 3 EU-NETVAL laboratories were GLP facilities and were asked to perform GLP compliant studies for the testing of the coded test chemicals. BDS is an ISO17025 accredited laboratory and performed the study following the GLP principles. This would entail qualified personnel and facilities, study planning, data to be recorded in the provided Data Analysis Files (DAFs), reporting of deviations and amendments during the study and quality control to confirm raw data are accurately reflected in the report.

3.4 Structure of the Validation Study & Validation Project Plan

The validation study was organised according to EURL ECVAM's Modular Approach to validation (Hartung *et al*, 2004), to generate information relevant to the modules 1 to 6 (Module 1: Test definition; Module 2: WLR; Module 3: Transferability; Module 4: BLR; Module 5: Predictive capacity (PC); Module 6: Applicability domain).

Table 04: **Overview of studies, purpose of the study, modules and laboratories**

Studies	Assessment of	Modules	Test chemicals	Participating Laboratories
Training		3		RISE, ENVIGO, CitoxLAB
Study 1	Transfer	3	6 non coded	RISE, ENVIGO, CitoxLAB
Study 2A Study 2B	WLR and BLR and PC	2, 4, 5	10 coded 10 coded	RISE, ENVIGO, CitoxLAB, BDS
Study 3	WLR and PC	2 and 5	26 coded	RISE

Prior to the start of the validation study, a Validation Project Plan was drafted by EURL ECVAM, reviewed and approved by the VMG. This document was updated whenever a modification to the validation study was necessary. Detailed information of the management of the study, responsibilities and tasks of the participating laboratories and of EURL ECVAM, overall set-up of the study and the organisation into 3 studies covering Modules 2 to 5, are provided in this Validation Project Plan (see Annex 13.2).

3.5 Test Chemicals and Test System

3.5.1 Selection procedure for the test chemicals

Compilation of data for a set of chemicals

A list of 83 chemicals, and their reported classification as positive or negative androgen as well as positive or negative anti-androgen, was compiled on the basis of the following data sources.

Literature

- ICCVAM recommendations for ARTAs (2003): list of 78 chemicals recommended for validation of *in vitro* ARTA agonist and antagonist assays that use mammalian cell reporter gene systems
- Publication on AR-CALUX® (B. van der burg, *et al*, 2010): AR-CALUX® cells tested with 11 chemicals for agonism, 9 chemicals for antagonism
- Publication on PALM (A. Freyberger *et al*, 2012): human prostate cancer cells (PC-3) transformed with the cDNA for a human androgen receptor, tested with 6 chemicals for agonism, 6 chemicals for antagonism
- Publication on AR-STTA (N. Araki *et al*, 2005): Chinese hamster ovary cells (EcoScreen™), transformed with the cDNA for a human androgen receptor, tested with 40 chemicals

Tox21 ARTA assay data

- Tox21 AR-BLA assay (no cytotoxicity data): human embryonic kidney cells (HEK293T) transformed with the cDNA for a human androgen receptor
- Tox21 AR-luc assay (no cytotoxicity data): human breast carcinoma cells (MDA-kb2)

QSAR

The Pass AR agonist model (http://www.pharmaexpert.ru/passonline/) was applied. This model relies on structural similarity towards a reference sample of AR agonists and non-agonists. Positive classification is achieved when the probability of being an agonist is higher than the probability of being inactive.

Expert consultation (VMG-NA, ICATM)

The list was provided to all members of the VMG-NA in 2013, for review and input. The list was also shared with ICATM. A few additional chemicals were suggested.

In addition, the chemicals were tested in-house via high throughput screening for a first evaluation (no cytotoxicity data).

Selection of a subset for testing in the AR-CALUX® validation study

A subset of 45 chemicals was selected, aiming at a balanced set of agonist, antagonist and negative chemicals (~15 from each class). Detailed information of the selection procedure can be found in the report of J. Burton (2014) (Annex 13.3) The criteria listed below were considered:

- Dose responses and classifications.
- Availability and price.
- Solubility as stock solution and as working solution (in cell medium).
- Potency.
- Known properties of the chemical.
- Structural diversity.
- Glucocorticoid receptor crosstalk.

A selection was carried out as follows:

- Few of the chemicals had restrictions for access while others were quite expensive leading to the elimination of 15 potential candidates.
- Solubility of each chemical was tested in-house and based on insolubility observations at the lowest concentration admitted for the validation study (a priori determined). A few candidate chemicals were eliminated.
- Diversity in terms of potency, chemical properties, structural space, glucocorticoid crosstalk were addressed.
- Values for the potency of the chemicals were retrieved from publications and/or databases. Chemicals were selected with the aim of representing low, medium and high potency values.
- The structural space was investigated based on structural similarity and cluster analysis (Avalon structural fingerprint, Tanimoto similarity). Substances were selected on the basis of maximum structural diversity as far as possible.
- AR-CALUX® cells are reported to have a highly specific selective response to low levels of different natural and synthetic androgens, and an insignificant response to other nuclear hormone receptor ligands such as estrogens, progestins, and glucocorticoids. The inclusion of one chemical with glucocorticoid binding properties was therefore considered in order to challenge the test system (Corticosterone).

The requests from the OECD's VMG-NA at the meeting of December 2014 were taken into consideration. Amongst the chemicals with antagonistic response, an inclusion of false competitive antagonists was suggested.

Overlap with the test chemicals of the Japanese ARTA and the Korean ARTA, both test methods under validation at the time of assembling the chemicals' list for the AR-CALUX®, was evaluated. During the course of both validation studies, the identity of the chemicals tested in these 2 ARTAs became available. Those that were not present yet in the list of 45 were added (see Annex 13.3). In 2017, a new ICCVAM list of AR-reference chemicals became available (Kleinstreuer *et al*, 2017). Given that the EURL ECVAM list of chemicals to be tested had only a small number of chemicals with reported agonist activity, it was discussed with the VMG to add 5 chemicals with agonist behaviour from the ICCVAM list. The two lists have 30 chemicals in common (overlap of 60%).

The VMG decided on the following:

- the reference chemicals DHT and Flutamide would not be used as coded test chemicals but the PC and NC chemicals could be used given that they had been used, in their capacity as controls, only at fixed concentrations.
- the 6 chemicals used in the transfer phase would not be used again as coded test chemicals given that the laboratories had gained already experience with these chemicals. The exception was SoAz, included as coded chemical for the reproducibility phase because 1) this is the only chemical soluble in water 2) challenges were experienced by some laboratories during the transfer phase.
- the chemical Disulfiram was proposed by the VMG as a putative control for the specificity control test i.e. a false competitive antagonist. This chemical had been tested in the Tox21 project with the assay Tox21 AR-luc. It was reported to have antagonist activity at concentrations that were not cytotoxic.
- one chemical tested in the reproducibility phase, Spironolactone, was once more included in the predictive capacity phase to test repeatability.

The complete list of the chemicals used in the AR-CALUX® validation is shown in Annex 13.3 and comprises in total 53 chemicals of which 46 were used as coded test chemicals.

3.5.2 Procurement and coding of the test chemicals

EURL ECVAM was responsible for the acquisition of all chemicals (including reference and control chemicals) and the preparation, labelling and storage at EURL ECVAM's chemical repository. 46 test chemicals were coded (for studies 2 and 3) and distributed to the 4 laboratories. In the event of an accident in the laboratory with these test chemicals, an emergency procedure was foreseen to obtain adequate information on the specific chemical. The laboratories had been instructed to treat all coded test chemicals as potential endocrine disrupters. A detailed description of the chemical coding and distribution procedure is provided in Annex 13.4.

The list with all the codes was provided to the statistician when all data had been received and analysed.

3.5.3 Test system

The test system consists of the AR-CALUX® cell line developed by the Dutch company BDS. EURL ECVAM prepared a cell bank of this test system for distribution to the test facilities. Prior to distribution, EURL ECVAM had the cell line tested for purity: 1) the absence of Hepatitis B and C, and HIV 1; 2) the absence of mycoplasma; 3) authenticity: the absence of cross contamination by other cell lines (STR profiling). Each laboratory received at the onset of the validation study 6 vials of frozen cells and more vials were supplied when so needed. For each study, a fresh vial of cells was used.

During the validation study, all laboratories had been asked to send aliquots of the last passage of the cell cultures to EURL ECVAM for a final verification of authenticity (see Annex 13.7.10). In summary, during the period of the validation study, the AR-CALUX® cell lines kept their identity and remained free of mycoplasma in all 4 laboratories.

3.6 Experimental Study Design

The VMG had reviewed and agreed on the following experimental set-up, per laboratory and across the laboratories.

Each laboratory had received a set of coded test chemicals, for which maximal solubility had to be determined starting from a concentration not higher than 50 mg/ml. A test chemical had to be tested with both the agonist assay and antagonist assay. The testing regime for each test chemical consisted of one (or more if needed) pre-screen experiments (with dilution factor 10) combined with a cytotoxicity test (LDH test) to 1) determine if the test chemical displayed a significant positive response according to the instructions in the SOP, 2) determine if and which concentrations were cytotoxic, 3) conclude on the proper dose range, both non-cytotoxic and soluble, for a test chemical showing a significant response (full or partial dose response). In order to achieve 3 valid runs (3 biologically independent replicates), the pre-screen test would be followed by either

- More pre-screen tests in case the test chemical did not display a positive response, leading to a total number of 3 valid runs of which two would have a cytotoxicity test. Visual checking of cytotoxicity was mandatory for all runs.
- Comprehensive tests (with a closer dose spacing) in order to obtain better resolution for calculating the parameters) in case the test chemical did display a positive response, leading to a total number of 3 valid runs, of which the first valid run would have a cytotoxicity test. In total, a test chemical with a positive response would have been tested twice with the cytotoxicity test (once in the pre-screen and once in the comprehensive test). Visual checking of cytotoxicity was mandatory for all runs.

The determination of a response being significantly positive was imbedded in the first series of Data Analysis Files (DAFs) and relied on an ANOVA Test. A classifier was not included in the first SOP versions used by the laboratories given that it was under development. With the introduction of the classifier in the final version of the SOP (V06), the criteria were also included in the updated DAFs leading to the removal of the ANOVA test.

The laboratories were informed that if it was practically difficult to meet the acceptance criteria (e.g. for a problematic test chemical), leading to invalidity of experiments, it would suffice to perform a maximum of six biological replicates irrespective of their validity.

For each pre-screen test and each comprehensive test, 8 concentrations per test chemical were tested. Each concentration was tested 3 times as defined by the plate layout for both agonist and antagonist testing (3 technical replicates).

In case of an agonist response, the testing proceeded as described above. In case of an antagonist response (full or partial dose response) each comprehensive test had to be accompanied by a specificity control (see section 4.7).

The 4 laboratories were asked to test 20 coded test chemicals for the assessment of reproducibility (WLR and BLR). The laboratory RISE was asked to test an additional 26 coded test chemicals for the assessment of predictive capacity on a total number of 46 test chemicals. The VMG considered this number sufficiently large to allow such assessment.

3.7 Data Collection and Analysis

3.7.1 Data collection

EURL ECVAM provided validated Data Analysis Forms (DAFs) to the participating laboratories for them to collect and analyse their data. These forms had embedded (and locked) calculations for the determination of the requested parameters (e.g. acceptance criteria, specificity control criterion, concentration points above or below the thresholds set for classifications). The test facilities received these forms with a data set to verify the correct functioning of the forms at the test facilities premises. In addition, the software *Graphpad Prism* was recommended for generating dose responses. During the Training phase of the validation study, the laboratories had received these forms in order to familiarise themselves with their use.

The following DAFs were used throughout the study:

- Form DAT02-ASY06 for the agonist assay (pre-screen and comprehensive testing)
- Form DAT04-ASY06 for the antagonist assay (pre-screen)
- Form DAT05-ASY06 for the antagonist assay (comprehensive testing and specificity control)
- Form DAT06-ASY06 for cytotoxicity data recording

The forms were modified and updated during the course of the validation study (see section 4.8 and Table 10).

EURL ECVAM retrieved all quality controlled data (completed DAFs, (Draft) final reports) of all test facilities via CIRCABC. Upon receipt of the DAFs from each laboratory, EURL ECVAM verified that 1) all data of valid and invalid runs (reported in DAFs and final reports) were submitted 2) the parameters calculated in the DAFs corresponded to those reported in the final reports.

The overall statistical analysis of the data reported in the DAFs was performed by EURL ECVAM statistician with the statistical software Matlab. The data analysis was performed according to the SOP, i.e. the evaluation of all acceptance criteria, the re-scalement of raw data into relative induction, the visualisation of concentration responses, estimation of parameters such as EC_{50}, PC_{10}, IC_{50}, PC_{80}, R^2 etc. In addition, for the initial part of the study where the classifier was not yet available, classification was applied on the reported data. These outcomes (dose responses, the measured values for the criteria, and final conclusions) for each test chemical were compared to the final conclusions and outcomes provided in the reports from the 4 participating laboratories.

3.7.2 Acceptance of data sets

Each laboratory was required to report all obtained data, being either valid or invalid. "Valid" data sets are defined as data that are in accordance with the acceptance criteria of the AR-CALUX® method. "Invalid" data sets are defined as the data from failed experiments (not meeting the acceptance criteria). The data were submitted to EURL ECVAM via DAFs on a regular basis. The laboratories were requested to report the measures that had been taken to overcome any failure to meet the acceptance criteria. Solubility data were submitted either as separate reports or as part of the Final report.

3.7.3 Data analysis

A statistical evaluation of all studies was performed on the basis of the following criteria:

- The number of valid/invalid runs (acceptance criteria met or not) and the reasons for invalidity
- Similarity of the obtained patterns (dose responses)
- EC_{50}/IC_{50} estimates of reference and test chemicals (where possible) and its variability

For the Transfer phase, the data generated by the laboratories had been compared to data generated during EURL ECVAM's GLP study ST57 (Annex 13.7.7)

The data that fulfilled the Acceptance Criteria (data from valid runs) were used for further data analysis. The determination of reproducibility within laboratories (WLR) and between laboratories (BLR) was based on concordance of classifications. A classifier to determine agonist and antagonist behaviour was developed by EURL ECVAM, in collaboration with the test method submitter, and included a criterion for the specificity control to be applied in the antagonist assay. This classifier was introduced in the SOP V06 (used for studies 2 and 3).

Once all data had been obtained from all 4 laboratories and analysed by the statistician for the assessment of WLR and BLR, the test chemicals were decoded. A final review of the data was performed by the VMG.

3.8 Time line for the studies

Time line for the studies (experimental part)				
	RISE	ENVIGO	CitoxLAB	BDS
Transfer phase (Study 1) (6 chemicals)	1/09/2015 – 5/12/2015 (6 chemicals)	03/11/2015 – 25/05/2016 (6 chemicals)	22/07/2015 – 12/04/2016 (6 chemicals) 08/11/2016 – 13/07/2017 (additional tests)	NA
Reproducibility phase (Study 2)	25/10/2016 – 14/12/2016 (10 chemicals)	17/01/2017 – 27/04/2017 (10 chemicals) 16/05/2017 – 22/02/2018 (10 chemicals)	03/07/2018 – 07/11/2018 (20 chemicals)(*)	15/07/2018 – 28/11/2018 (20 chemicals)
Additional 36 chemicals (Study 3)	16/05/2017 – 23/02/2018	NA	NA	NA

Note: The indicated time slots may not reflect the actual time spent to the validation study given that all participating laboratories provided services to their clients. (*) investigative tests took place during Study 2 (10/10/2017 – 16/05/2018). NA = not applicable

4 TEST DEFINITION (MODULE 1)

4.1 Description of the *in vitro* Method

The AR-CALUX® cell based assay provides information on the endocrine activity of chemicals, and more specifically the (anti-)androgenic activity, when the AR-CALUX® cells are exposed to substances. This *in vitro* method is a transactivation assay where the reporter gene *luc* (encoding luciferase) is activated by the androgen receptor but only when bound to a ligand, i.e. a chemical with androgen receptor affinity. This receptor-ligand complex enters the nucleus where it will bind to specific recognition sequences in the promoter region of a target gene (so called androgen responsive elements or ARE). Hence, the target gene will be transcribed. When the target gene expresses the reporter (luciferase), *in vitro* hormonal activity of chemicals can be quantified as well as the agonistic or antagonistic mode of action. Such assays are called Androgen Receptor Transactivation Assays (ARTA).

The AR-CALUX® cell line was created via transfection of the human osteosarcoma cell line U2-OS (ATCC HTB 96) with 2 constructs: pSG5-neo-hAR carrying the cDNA of a human androgen receptor under a constitutive promoter, and, a luciferase reporter gene which is preceded by a triple tandem of AREs in front of a TATA box (3x ARE Luc).

This cell line had been reported to stably express the human androgen receptor, to be highly selective in its response to low levels of different androgens (due to the multimerized ARE and a minimal promoter – TATA box only), and to have an insignificant response to other nuclear hormone receptor ligands such as estrogens and glucocorticoids (due to the cells not expressing other steroid receptors that can activate transcription via the same ARE as the androgen receptor) (Sonnenveld *et al*, 2004).

The cell line has low metabolic activity as was shown via RNA sequencing where major classes of metabolic genes were targeted and found to have no or low expression (personal communication). By combining the test method with a S9 fraction, the impact of metabolism on test chemical activity can be studied (van Vught-Lussenburg *et al*, 2018).

The assay has been used for high throughput screening (van der Burg *et al*, 2015)

The name "CALUX" has been registered at a national trade mark office. This trade mark, owned by Abraham Brouwer and BDS, is for "Conducting chemical, biochemical and biological analyses; preparing cell lines, tissue cultures, culture media and supplements therefore and products thereof". The cell lines, the protocol, training and technical support are available through a license agreement. The parental cell line U2OS was obtained from ATCC and approval for commercialisation was given in 2002.

4.2 Purpose and Regulatory context of the *in vitro* Method

The AR-CALUX® method is intended to be used for screening purposes because of an easy and time efficient application. Both the OECD Conceptual Framework and the US EPA have recommended transactivation assays as an important tool for fast screening of chemicals with possible endocrine disrupting properties. OECDs Conceptual Framework has identified several type of methods classified over levels, e.g. level 2 involves *in vitro* assays providing mechanistic data. Validated ERTAs are included at this level, but there is still a lack of validated ARTAs. The proposed AR-CALUX® method, once validated, could be inserted at this level 2.

4.3 Principle of the *in vitro* Method

The test method is described by the test method submitter to measure the ability of a chemical to activate AR dependent transcription (i.e. act as an agonist) and to suppress AR dependent transcription (i.e. act as an antagonist). Hence, the test method is composed of an agonist and an antagonist assay.

Both assays include a pre-screen for determining the appropriate dose range, followed by comprehensive testing. To determine the agonist or antagonist nature of a test chemical, it will be tested in the following manner:

1) A dilution series of the chemical is prepared in solvent (e.g. DMSO) and applied to the cells in assay medium. When the luminescent signal increases in a concentration dependent way in comparison to the solvent control, the chemical has an agonist response.

2) A dilution series of the chemical is prepared in solvent (e.g. DMSO) and applied to the cells in assay medium supplemented with the EC_{50} concentration of Dihydroxytestosterone (DHT). When the luminescent signal decreases in a concentration dependent way in comparison to the solvent control, the chemical has an antagonist response unless there is a non-specific response. In order to rule out a false antagonist response, the chemical is tested with an EC_{50} and 100X EC_{50} concentration of DHT within the same plate.

In order to label a test chemical as an agonist or antagonist, a classification scheme is applied.

4.4 Reference Chemicals and Control Chemicals

The agonist and antagonist assays each have a reference chemical for which the dose response is measured, and EC_{50} \ IC_{50} values, the induction factor and the Z factor (see Tables 05 and 06) calculated. It is also the chemical to which the response of a test chemical is compared (normalisation).

The positive control and negative control consist of the addition of a chemical to the test medium (including DMSO) for which respectively a response or no response is expected. Both assays have a vehicle control VC which is the assay medium including DMSO (0.1%) while the antagonist testing includes also a solvent control SC which is the vehicle control plus EC_{50} of DHT.

Table 05: Proposed reference and control chemicals for the agonist assay

	Name	CAS No.
Reference	Dihydroxytestosterone (DHT)	521-18-6
Positive control	Methyl testosterone	58-18-4
Negative control	Corticosterone	50-22-6

Table 06: Proposed reference and control chemicals for the antagonist assay

	Name	CAS No.
Reference	Flutamide	13311-84-7
Positive control	Linuron	330-55-2
Negative control	Levonorgestrel	797-63-7

4.5 Acceptance Criteria

Criteria for the reference chemical, positive and negative control were established by both the test method submitter and EURL ECVAM during the assessment of the SOP. An experiment is considered valid and will be accepted when all of these acceptance criteria are met (Tables 07 and 08).

Table 07: Acceptance criteria in the agonist assay

No.	Acceptance criterium	Value
	Reference chemical DHT	
1	Curve fitting	Sigmoidal
2	EC_{50} range	1.0 E-10-1.0 E-9 M
3	CV of estimated log(EC_{50})	< 1.5%
4	Induction factor	> 20
5	Z-factor	> 0.5
	Positive control	
6	RI Methyl testosterone	> 30%
	Negative control	
7	RI Corticosterone	< 10%

Table 08: Acceptance criteria in the antagonist assay

No.	Acceptance criterium	Value
	Reference chemical Flutamide	
1	Curve fitting	Sigmoidal
2	IC_{50} range	1.1 E-7-1.1 E-6 M
3	CV of estimated log(IC_{50})	< 3%
4	Inhibition factor	> 10
5	Z-factor	> 0.5
	Positive control	
6	RI Linuron	< 60%
	Negative control	
7	RI Levonorgestrel	> 85%
	Reference chemical Flutamide specificity control	
8(*)	R^2 between the RI of Y_c and S_c for Flutamide	≤ 0.7

RI = Relative Induction

(*) To be applied for assessment of the specificity response (S_c) of Flutamide, Y_c being the standard response.

4.6 Classifier

EURL ECVAM and the test method developer worked together to assemble a classifier for both agonist and antagonist assay. When the specificity control criterion was developed (see section 4.7), the VMG decided to include this criterion into the classifier. The classifier, as introduced in SOP V06, is shown below.

Agonism: For each run, a test item is considered

A. **Positive** when the relative induction (Y_c) of the test item is equal or exceeds 10% (REF RPC_{10}) for two or more consecutive concentrations.

B. **Negative** when the relative induction (Y_c) of the test item does not exceed 10% (REF RPC_{10}) for any concentration.

C. **Inconclusive** in all other cases.

Antagonism: For each run, a test item is considered

A. **Positive** when the following two conditions are met:

- the relative induction (Y_c) of the test item is less or equal to 80% (REF RPC80) for two or more consecutive concentrations and
- the correlation coefficient (R^2) is less or equal to 0.9 between the relative induction of the test item (Y_c) and the relative induction of its specificity control (S_c).

B. **Negative**

Either

- when the relative induction (Y_c) of the test item is greater than 80% (REF RPC_{80}) at all concentrations;

or

- when the relative induction (Y_c) of the test item is less or equal to 80% (REF RPC_{80}) for at least 2 consecutive concentrations and the correlation coefficient (R^2) is greater than 0.9 between the relative induction of the test item (Y_c) and the relative induction of its specificity control (S_c).

C. **Inconclusive** in all other cases.

During the review of all obtained data at the end of the validation study, the SOP and the classifier were modified by the VMG. This included guidance in the SOP regarding the observation of one concentration passing the thresholds of 10% (for agonism) and 80% (for antagonism), and, a modified classifier (see section 10 Discussion).

4.7 Development of a Specificity Control for the Antagonist Assay

An antagonist assay is performed by spiking the assay medium with a specific concentration of a known ligand. Usually this concentration will be the EC_{50} of the ligand. The antagonist assay leads typically to dose responses with a sigmoid shape (see figure 02, response Y_c) which is normally caused due to increasing competition at the receptor binding site between the agonist ligand (at fixed concentration) and the antagonist test chemical (with increasing concentration). Such response however can also be caused by other effects of which cytotoxicity is the most occurring cause. Interferences along the pathway of receptor-ligand binding, dimerization, binding to the UREs, transcription, translation up to protein stability have also been reported.

A control to identify (true) competitive antagonists (i.e. competition for the same binding site on the AR) and false positive antagonists (e.g. due to cytotoxicity) was introduced. It consists of testing the chemical with two different concentrations of stimulating ligand (DHT $3x10^{-10}$ M for a standard response (=EC_{50}), and, $3x10^{-8}$ M for the specificity response (=100X EC_{50})), leading to two different dose responses and subsequently a potency shift (see figure 02).

4.7.1 Concentration for spiking

The 100X EC_{50} concentration was defined on the basis of the complete dose response for DHT when administered to the cells (see figure 01). A concentration was chosen that was as high as possible in order to have full saturation of the receptor binding places but not leading to cytotoxicity.

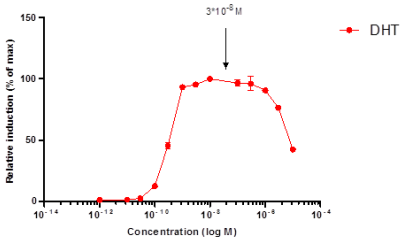


Figure 01:
Dose response of DHT

4.7.2 Criterion R^2

A criterion was sought to quantify the shift of the two dose responses, based on the assumption that the decrease of the test chemicals response which is <u>not</u> due to competitive antagonism is proportionally the same as the decrease of the specificity control response at all (non-cytotoxic) concentration. The VMG agreed on the criterion R^2 : the square of the correlation coefficient between the Relative Induction (RI) of the test chemical at concentration c (Y_c) and the Relative Induction of its specificity control at concentration c (S_c).

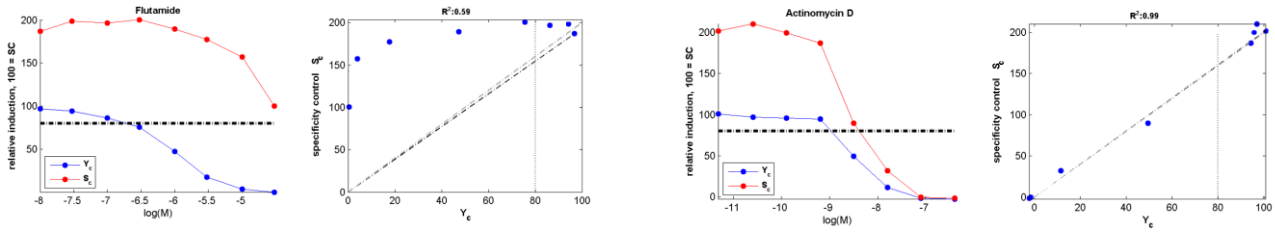


Figure 02: Dose responses generated by Flutamide and Actinomycin D. The curve in blue is the standard response (Y_c), the curve in red is the specificity response (S_c).

The threshold value for R^2, to distinguish (true) competitive antagonist from false positive antagonist, was based on historical data obtained from several data sets generated at EURL ECVAM and BDS (see Annex 13.7.8). The following values were applied in the validation study:

- Competitive antagonist: $R^2 \leq 0.9$.
- False positive antagonist: $R^2 > 0.9$

This measurement was included as a criterion in the experimental design. To verify the proper functioning of the specificity control, an acceptance criterion for Flutamide was introduced when tested with the specificity control. The threshold value was determined from historical data (see Annex 13.7.8.). This criterion was added in the list of Acceptance Criteria leading to the decision of a valid run (see section 4.5): R^2 between Y_c and S_c for Flutamide ≤ 0.7.

4.8 Protocol of the Test Method

In the test submission of the test method developer BDS to EURL ECVAM, 6 SOPS had been provided that included instructions for both ER and AR-CALUX®. BDS in discussion with EURL ECVAM modified the SOPs, assembling them to one SOP, removing references to ER-CALUX®, adding additional acceptance criteria. EURL ECVAM drafted SOP V01 for experimental testing in the laboratory, leading to versions 02 and 03. Modified versions of the SOP have been generated throughout the validation study based on observations and feedback from the 3 laboratories. The different SOP versions have been used as follows:

- Version 03 for the Training
- Version 04 for the Transfer phase (6 test chemicals)
- Version 05 for the Reproducibility phase (10 test chemicals by RISE and ENVIGO)
- Version 06 for the Reproducibility phase (10 test chemicals by ENVIGO, 20 test chemicals by CitoxLAB and BDS), and, for the Predictive capacity phase (26 test chemicals by RISE)

Version 06 included the classifier as well as a criterion for the specificity control. Version 07 was prepared and released by the VMG at the end of the validation study. It included a warning for the usage of glass ware and plastic ware, to be carefully verified for contaminants prior to initiating experiments.

The major modifications to the SOP included the following:

- Completion and modification of acceptance criteria for both agonist and antagonist assay
- Development and inclusion of a specificity control for the antagonist assay, including a quantitative criterion R^2 for the reference chemical Flutamide and for test chemicals
- Adding a classifier (with inclusion of the criterion for the specificity control)

In summary, for 20 coded test chemicals, to be evaluated for WLR and BLR, two SOPs were handled that differed only by having the classifier included in the last version V06. The additional 26 coded test chemicals evaluated by RISE were all tested with the SOP version V06 (with classifier). The use of these two different versions is due to the laboratories initiating and finishing the studies at different times. Once the SOP version V06 was ready, it was provided to those laboratories that still needed to start with part (ENVIGO) or whole (CitoxLAB and BDS) of the study 2 (20 coded test chemicals).

A summary overview of the modifications made to the set of SOPs is given in Table 09.

Data analysis files (DAFs), for recording and analysing data, had been received from the test method developer and were updated during the course of the validation study in line with the updates in the SOP versions. A summary overview of the changes introduced to those DAFs is shown in Table 10. The major changes included:

- replacement of the ANOVA test with the classifier
- inclusion of the specificity control criterion
- modifiable dilution factor instead of a fixed dilution of 3.33 in the comprehensive test
- inclusion of VC level of 5% for monitoring too high RLUs

DAT 06 for recoding data of cytotoxicity testing remained unchanged.

The most recent forms were provided to the laboratories RISE, CitoxLAB and BDS during the validation study in order to make use of a flexible dilution factor. ENVIGO did not use this form because it had finished the experimental part.

Table 09: Overview of major changes introduced in the SOP versions used in the validation studies: ASY06 versions 3, 4, 5 and 6 and 7	
Version 03 (09/02/2015) - used for the Training Changes from SOP BDS versions	
General	• Merged 6 BDS SOPs into one EURL ECVAM SOP-ASY06. • Highest concentration of a chemical to be tested shall be 50 mg/ml instead of 100 mM. This was changed because molecular weight of test chemicals will be unknown. • Inclusion of a specificity control for antagonist assay (experimental part only, no criterion) • Inclusion of pictures for confluency and cytotoxicity visual checking • Removal of BDS solubility section • No classifier
Technical aspects	• Change to a fixed sub-culturing seeding density of AR-CALUX® cells instead of splitting 1:5 or 1:7 • Inclusion of option to use also commercially available solutions for luciferase expression, lysis buffer, trypsin, geneticin and charcoal stripped FCS • Change in sample preparation: 24 hours after seeding 100 µl assay medium is not removed from the wells. 100 µl of test chemical with double concentration is added to the wells. • Maximum DMSO concentration reduced from 1% to 0.1% on basis of EURL ECVAM in-house experiments. • Increase of linuron concentration (PC Antagonist) from 1 x 10^{-6} M to 1 x 10^{-5} M to ensure sufficient antagonistic (positive) response. Higher linuron concentrations are not recommended because signs of insolubility were observed at 15 µg/ml in assay medium. • Increase of concentration of Levonorgestrel (NC Antagonism) from 1 x 10^{-7} M to 1 x 10^{-6} M (giving the maximum agonist response). • Increase of concentration of 17α-Methyltestosterone (PC Agonist) from 1 x 10^{-9} M to 1 x 10^{-7} M (giving the highest agonist response). • Instruction for cytotoxicity determination via both LDH leakage and visual inspection. • Inclusion of a cytotoxicity control: Triton X to be used in column 1 of the plates. • Established the concentration of DHT to be used for Antagonism specificity control at 100 x EC_{50} DHT concentration. • Induction factors are calculated per plate to confirm validity.
Plate layout	• Plate layout changed for Agonist and Antagonist assay. Cytotoxicity positive control added; C4 concentration of FLU and DHT replaced by their C8 concentration to be able to calculate the IF. The plate layout of the Antagonist assay had an additional change to one test chemical per plate only to accommodate the inclusion of the specificity control within the same plate.
Acceptance Criteria	• Removal of the tolerated variability of triplicate samples: % SD of triplicate wells for reference chemical C1 to C8, test chemical samples, PC, NC, CO and SC • Induction factors (IF) are calculated for each plate using SC and highest concentration C8 of the reference chemical, to confirm validity of each separate plate, instead of comparing C4 and C8 samples with those on plate 1 • Inclusion of CV of estimated log(EC_{50}) and estimated log(IC_{50})for the reference chemicals DHT and Flutamide • Inclusion of the Z-factor to encompass tolerated variability of all samples

Data analysis	• Included the calculations of the existing parameters • Added the calculations for the additional parameters Z-factor and induction factors (IF) per plate • Data analysis section revised and calculations moved to the Annex • Inclusion of an ANOVA test to distinguish significant differences from the solvent control (VC in Agonist testing or SC in Antagonist testing) and simplify the selection of proper concentrations for Agonist and Antagonist comprehensive testing after the pre-screen

Version 04 (16/06/2015) – used for the Transfer phase (6 chemicals) and for EURL ECVAM study ST57 (comparison date for transfer) Changes from v03:	
General	• Re-organisation of the SOP to better separate Agonist and Antagonist data analysis from concentration selection and acceptance criteria. • No classifier
Technical aspects	• Addition of procedure for manual preparation of trypsin-EDTA solution starting from powder • Addition of schemes for dilution of reference and control chemical stock solutions • Inclusion of extra control step to microscopically confirm successful lysis of the cells for luminescence sample preparation.
Plate layout	• Solvent control of the reference chemical C0 is renamed to SC as it is identical to the SC of the test chemical • Included plate layout for the Antagonism specificity control.
AC	Not changed
Data analysis	Not changed

Version 05 (24/10/2016) – used for the reproducibility phase (10 chemicals for RISE, 10 chemicals for ENVIGO) Changes from V04:	
General	• Inclusion of reference to the SOP for solubility testing: SOP-ASY15 • No classifier
Technical aspects	• Monitoring instruction included for VC: to monitor closely the RI for unexplainable increase • Monitoring instruction included for SC: to monitor closely that RI of C1 samples is approximately at SC level in order to be able to calculate the parameters • Laboratories own historical DHT EC_{50} value to be used in the Antagonist tests instead of the default EC_{50} value
Plate layout	Not changed
AC	Not changed
Data analysis	Correction of formulas for PC_{50} and PC_{10}

Version 06 (30/03/2017) – used for the reproducibility phase / predictive capacity (10 chemicals ENVIGO, 37 chemicals RISE, 20 chemicals CitoxLAB and BDS)

Changes from V05:	
General	Inclusion of the classifier for Agonism and Antagonism and visual explanation of the specificity control.
Technical aspects	Not changed
Plate layout	Not changed
AC	Inclusion of the R^2 criterion (criterion 8) for Flutamide that shows the proper functioning of the specificity control in the Antagonist assay.
Data analysis	• Removal of the calculation of ANOVA test that distinguished SC from C1 concentration and replaced it by the threshold values 10% for agonism and 80% for antagonism. • Inclusion of the calculation of R^2 • Introduction of parameter S_c for specificity response and Y_c for normal response in the calculations for Antagonist responses. Calculating these parameters and using the ratio between S_c and Y_c it can be determined if a test chemical is a true competitive Antagonist or not.

Version 07 – FINAL SOP for dissemination (available at TSAR)	
	• Warning for verification of assay performance with the usage of material and reagents such as glass ware, plastic ware and solvents. Include the material and reagents that worked well as examples • Classifier modified • Change the test chemical concentration from mg to mM • Guidance for testing with dilution factor 2 when only the highest tested concentration is found to pass the threshold values for classifications • Flexibility for the use of fresh or frozen stock solutions of the two reference chemicals • Flexibility for the use of plastic 24 well plates or glass tubes for preparing working solutions

Table 10: Overview of major changes introduced to the DAFs			
DAT02 **Agonist PRE and COM**	**DAT04** **Antagonist PRE**	**DAT05** **Antagonist COM and specificity control**	**Forms used for**
Draft V04	Draft V03	Draft V03	
• Revised BDS DAF conform SOP V03	• Revised BDS DAF conform SOP V03	• Revised BDS DAF conform SOP V03	Training
V04 (validated)	V03 (validated)	V03 (validated)	
No change	No change	No change	Study 1: 6 chemicals (Transfer phase) Study 2A: 10 coded chemicals RISE and ENVIGO
V05 (validated)	V04 (validated)	V04 (validated)	
• ANOVA replaced by classification threshold 10%	• ANOVA replaced by classification threshold 80%	• ANOVA replaced by classification threshold 80% • Inclusion R^2 calculation for reference ítem Flutamide • Inclusion R^2 for calculation for test chemical • Inclusion graph plotting dose responses S_c versus Y_c	Study 2B: 10 coded chemicals ENVIGO Study 2: 20 coded chemicals CitoxLAB and BDS Study 3: 36 coded chemicals RISE
V06	V05	V05	
• Dilution factor made flexible • Formulas changed for calculation of average RLU and RI, to ensure that non-numerical values are ignored for calculation of parameters and visualisation of graphs	• Dilution factor remained to be 10X • Inclusion IF calculation for the VC in antagonist assay • Formulas changed for calculation of average RLU and RI, to ensure that non-numerical values are ignored for calculation of parameters and visualisation of graphs	• Dilution factor made flexible • Inclusion IF calculation for the VC in antagonist assay • Formulas changed for calculation of average RLU and RI, to ensure that non-numerical values are ignored for calculation of parameters and visualisation of graphs	For some coded chemicals RISE, CitoxLAB, BDS in Study 2 and 3

4.9 Known Technical Limitations and Drawbacks of the Test Method

The test method developer reported no specific limitations of the test method except for the general limitations of transcription assays: direct extrapolation to the *in vivo* complex network of signalling and regulation should not be made; information is gained on the parent molecule while *in vivo* other molecules may be generated due to the cells' metabolism.

4.10 Conclusion of the VMG

The need of the test method in the context of current OECD regulatory requirements is evident. The proposed use of the test method as a screening method for EDs is relevant.

The AR-CALUX® SOP was considered sufficiently developed for the purposes of this study. The acceptance criteria and their values were considered as adequate.

The introduction of a test to verify the true positive nature of an antagonist response (this is the specificity control) was considered of added value (see further section 10 Discussion).

During the course of the validation study some minor modifications were made to the SOP. At the end of the validation study, recommendations were made by the VMG for further modifications to the SOP e.g. additional instructions for the choice of the dilution factor and modifications of the classifier (see Table 9 and section 10 Discussion). These changes will be introduced in the final version of the SOP.

5 TRANSFER (MODULE 2 / STUDY 1)

Reference documents:

- Statistical report (Annex 13.1)
- SOP version V03 and V04 (Annex 13.7.3)
- Agenda and Planning of the training (Annex 13.7.1)
- EURL ECVAM report GLP compliant study on 6 test chemicals SR-ST57 (Annex 13.7.7)
- JRC technical report on " Technical meeting on the Implementation of the AR-CALUX® *in vitro* method" (Annex 13.7.4)
- JRC technical report "Transfer Evaluation report" (Annex 13.7.5)
- Study plans and study reports of the participating laboratories (Annex 13.7.2)

5.1 General Aspects

The first phase of the validation study consisted of the Transfer phase where the laboratories implemented the test method in their own laboratory. Their results were evaluated to verify if the laboratories were sufficiently trained with the application of this test method. EURL ECVAM's GLP laboratory had carried out a GLP study with the test method (SOP version 03) and generated GLP compliant test data which would serve as a reference data set when assessing the transferability of the test method to the test facilities (see Annex 13.7.7).

The Transfer phase was initiated by EURL ECVAM on June 15, 2015. The 3 laboratories started at different times within a time span of 3 months and data were received in April 2016 (RISE) and June 2016 (CitoxLAB and ENVIGO). The data analysed in this report are the verified (QC) data from the 3 laboratories. During the course of the Transfer phase, EURL ECVAM visited the 3 laboratories to discuss the ongoing validation study and help out with issues encountered. CitoxLAB had a transfer phase which consisted of two parts (July 2015 - July 2016, and, November 2016 - June 2017) due to technical issues encountered. Additional training was given by EURL ECVAM in April 2017.

Prior to the initiation of the Transfer phase, the 3 participating laboratories were given training on the usage of the test method. Training was provided by EURL ECVAM in collaboration with the test method developer BDS at the premises of the Joint Research Centre in Ispra.

5.2 Training

SOP V03 was used for the training. The training included theoretical components such as detailed review and discussion of the SOP, the DAFs, study plans and reports. Hands-on training was offered for the usage of the DAFs with a set of fictive data and the use of CIRCABC for retrieving and uploading files. CIRCABC is the European Commission platform for data exchange and was the tool for data and documentation transfer between each laboratory and EURL ECVAM. Experimental training was provided where the participants either observed or performed the critical elements of the SOP such as plating of cells, preparing dilutions of few test chemicals and treating the cells, luciferase measurements.

At the end of the course, the participants were presented with a questionnaire about the theory/practical sessions covered during the previous days, to challenge what they learned during the course.

5.3 Transfer of the Test Method to the Naïve Laboratories

SOP V04 was used for the Transfer assessment. The testing comprised out of pre-screen testing to deduce the correct dose response, followed by comprehensive testing. For the antagonist assay, a specificity control test was also required.

The 3 participating laboratories had been asked to test 6 test chemicals with both agonist and antagonist assay (see Table 11). Classification was not required given that the SOP did not include yet the classifier. For each test, cytotoxicity was measured. Each laboratory was required to achieve 3 valid runs. Laboratories had been given DAFs to record the RLUs. These forms had imbedded calculations for the Induction factor, Z-factor, etc. Dose responses, EC and IC values for the test chemicals had to be obtained by using GraphPad or other similar statistical software.

Table 11: Test chemicals for the Transfer phase					
Test chemical	CAS No.	AR reference list		EURL ECVAM's ARTA data list	
		AGONIST	ANTAGONIST	AGONIST	ANTAGONIST
Testosterone	58-22-0	Strong	Negative		
4-Androstenedione	63-05-8	Moderate	Negative		
Procymidone	32809-16-8	NA	Very weak		
p,p'-Methoxychlor	72-43-5	NA	Weak		
Sodium azide	26628-22-8			N, N, N	N, N, N
Di-n-butyl phthalate	84-74-2			N, N, N	N, N, N

N= negative response
AR reference list: ICCVAM list of AR Reference chemicals (Kleinstreuer *et al*, 2017)
EURL ECVAM's ARTA data list is described in Annex 13.3

A summary overview of the obtained results is given in Table 12. More information and details can be found in Annex 13.7.5.

Transfer to RISE

RISE observed few times high values for the Vehicle Control (VC) in the antagonist assay (2 runs out of 5 valid runs) though this did not affect acceptance of the run. The VC did not have an Acceptance criterion. In discussion with the laboratory the reason for these high values was not discovered. The high values for the VC however were of concern given that also CitoxLAB obtained high values (see further). Therefore EURL ECVAM discussed with the VMG to introduce a monitoring guidance for the VC.

Transfer to ENVIGO

ENVIGO reported early on about difficulties with the luminescence readings which they thought was due to cross contamination in the plates. They therefore proposed to use the Steady Glo kit for luminescence measurements because of previous good experience from ERTA measurements. Using this particular kit does not require an injection luminometer. The usage of such luminometer however had been one of the requirements in the invitation to participate in this validation study as it was the equipment proposed by the test method submitter. ENVIGO carried out the experiments for the agonist assay with the Steady Glo kit (and without double injection luminometer) and provided data of good quality.

EURL ECVAM visited ENVIGO (April 2016) to understand the issues with the luminescence measurements. It was discovered that ENVIGO's equipment was calibrated with very sensitive settings. While faithfully following the example setting indicated in the SOP V04 (4 seconds integration time), this time span was too high for the given settings. With lower integration time, the readings were fine. Given that ENVIGO had already carried out the agonist part with the Steady Glo mix (and had obtained good data), EURL ECVAM allowed them to continue with this kit for the antagonist part. The data nevertheless would be carefully analysed for acceptability and subject to review and approval by the VMG. EURL ECVAM considered that, under the condition that the generated data would be similar to the data of the other laboratories, the usage of another kit and a luminometer without injection by one laboratory could be of benefit for the test method, for the validation study and for the future TG on ARTAs. The current SOP V04 restricted the luminometer to those with double injection which can be a limiting factor in the application of the test method in many laboratories.

When comparing the number of valid/invalid runs across the 3 laboratories, ENVIGO had the lowest number of invalid runs: 1 out of 8 agonist runs, and, 1 out of 7 antagonist runs.

Table 12: Overview of some results obtained in the 3 laboratories and EURL ECVAM									
AGONIST ASSAY					**ANTAGONIST ASSAY**				
	RISE	**CitoxLAB**	**ENVIGO**	**EURL ECVAM**		**RISE**	**CitoxLAB**	**ENVIGO**	**EURL ECVAM**
# of total runs	8	11	4	8	# of total runs	8	15	7	6
# of valid runs	3	3	3	7	# of valid runs	3	4	6	4
# of partially valid runs	1	2	0	1	# of partially valid runs	2	1	1	0
Average across runs					Average across valid runs				
DHT EC_{50} (M)	2.5E-10	2.7E-10	3.8E-10	3.1E-10	Flutamide IC_{50} (M)	6.2E-07	7.3E-07	5.2E-07	6.6E-07
CV($logEC_{50}$)	0.41%	0.53%	0.60%	0.47%	CV($logIC_{50}$)	0.84%	0.98%	0.90%	0.61%
IF	47	83	78	76	IF	31	44	30	32
ZF	0.84	0.83	0.72	0.84	ZF	0.79	0.79	0.74	0.82
Variability of plate triplicates CV (test chemicals only, RLU where RI>2.5%)					Variability of plate triplicates CV (test chemicals only, averaged over runs, RLU where RI>2.5%)				
DHT	7.9%	8.4%	9.8%	7.4%	Flutamide	8.5%	8.9%	10.2%	6.8%
Methoxychlor	not reported, in absence of agonist response				Methoxychlor	6.2%	6.8%	9.0%	5.7%
Procymidone					Procymidone	6.7%	5.2%	8.8%	7.6%
Dibutylphtalate					Dibutylphtalate	5.7%	6.4%	9.6%	9.1%
Androstenedione	9.6%	5.7%	9.8%	8.0%	Androstenedione	5.4%	4.2%	8.0%	6.2%
Testosterone	7.3%	5.4%	11.0%	7.1%	Testosterone	4.9%	5.1%	9.3%	6.0%
Sodium azide	not reported, in absence of agonist response				Sodium azide	6.5%	5.5%	9.0%	9.2%
average CV of all triplicates	9.3%	6.8%	10.2%	7.5%	average CV of all triplicates	6.3%	6.0%	9.1%	7.2%
DHT Within and Between runs variability					Flutamide Within and Between runs variability				
within run variability [average CV($log EC_{50}$)]	0.36%	0.46%	0.51%	0.43%	**within** run variability [average CV($log IC_{50}$)]	0.74%	0.92%	0.83%	0.56%
between runs variability [CV($log EC_{50}$)]	0.61%	0.82%	1.44%	0.76%	**between** runs variability [CV($log IC_{50}$)]	1.45%	1.56%	1.03%	1.89%
Main reasons for invalid runs	* DHT curve failed (1run) * SC too high (3 runs) * plate reader error (1 run)	* triton incorrect (5 runs) *wrong plates used (2runs) *DMSO incorrect in solvent (3runs) *DMSO incorrect as solvent (1 run)	all criteria failed in one run	* No NaOH injected into the plate (part of 1 run) * DHT dilution series incorrectly prepared (1 run)	**Main reasons for invalid runs**	* IC_{50} Flutamide out of range (1run) * IF out of range (1run) *plate reader error (1run) * NC an PC not on plate (1run)	* triton incorrect (3 runs) *wrong plates used (2runs) *DMSO incorrect in solvent (2runs) *incorrect preparation, criteria not met(2 runs)	Z-factor not met in one plate of one run	Relative induction of control chemicals criteria not met

Transfer to CitoxLAB

Similar to RISE, CitoxLAB obtained few times high values for the Vehicle Control (VC) in the antagonist assay (2 runs out of 5 valid runs).

A larger number of runs to arrive to the required 3 valid runs for agonist and antagonist assays had to be carried out, mainly due to the usage of a too high concentration of Triton-X (the positive control for cytotoxicity in all plates, placed in row 1). Such had influenced the cells in the neighbouring row (row 2 with SC). This was obvious in the antagonist assay as the wells with SC (assay medium with DHT, normally leading to high RLUs) had too low RLU values. It was not immediately picked up in the agonist assay as the Triton-X wells were bordering the SC wells (assay medium only, normally leading to low values). Several invalid runs for the agonist assay can be noticed (see Table 12; first 6 out of 11 runs) due to a relabelling as invalid because of the discovery of issues with the use of Triton-X. On EURL ECVAM's recommendation, all experiments where Triton-X may have been used at an incorrect concentration were repeated. The laboratory purchased ready to use 10% Triton-X solution what solved the issue. The laboratory reported on not clear enough instructions in the SOP V04 leading to different interpretations in terms of pipet use and calculations.

CitoxLAB reported for few antagonist runs SC data that increased with increasing plate number. The SC is present on each plate in the second row (6 wells). EURL ECVAM visited the facility (May 2016) and advised on strictly following the instructions of the SOP V04 for thawing the luciferase mix until room temperature before usage. CitoxLAB's practice of placing the mix from the freezer in a beaker with tap water of which the temperature was not controlled may have led to the observed variability in the results. The mix could have been used at temperatures lower than room temperature, and, slowly warmed up in the instrument during the measurements. This may explain why the SC data increased over time (with increasing plate number). When the laboratory changed its practice of thawing, the issue was remediated.

CitoxLAB provided graphs for the antagonist assay (for the reference chemical Flutamide and for the test chemicals) which sometimes displayed a deviating dose response (2 runs out of 5). In particular, the lowest tested concentrations (C1 and higher) should normally lead to responses at the 100% SC level (the plateau phase). This laboratory obtained responses that decreased up to 50% and not all parameters could be calculated anymore. The reason was not known. EURL ECVAM proposed a monitoring guidance for the next version of the SOP.

The VMG reviewed all data from all 3 laboratories (September 2016) and recommend CitoxLAB to perform some additional tests for the implementation of the antagonist assay. Seven additional tests were performed (within the time span November 2016 - June 2017) of which the last 2 could be concluded as valid. The technical issues that were observed in the Transfer phase could be remediated during the course of the testing. Investigative tests that were run alongside the test of the test chemicals seemed to indicate that the usage of a certain type of glass tube (big glass tubes) increased the RLU values. Irregularities in the RLU values were occasionally still noted in the usage of plastic tubes.

5.4 Discussion and Conclusion of the VMG

The VMG reviewed the data of all 3 laboratories in September 2016 and concluded on the following.

The analysis of the Transfer data showed for the Agonist protocol

- Agonist positive and negative responses were correctly obtained
- Concentration responses were comparable across runs and laboratories, some differences were observed in few runs in 1 or two concentrations.
- Estimates of DHT EC_{50} were around the value $3x10^{-10}$ M except for RISE where all the values were slightly lower but within the acceptance range.
- Within run variability was measured via CV of DHT $logEC_{50}$ and was comparable across laboratories except slightly higher variability at ENVIGO
- Between run variability was measured via CV of DHT $logEC_{50}$ with similar observation.

The analysis of the Transfer data showed for the Antagonist protocol

- Antagonist positive and negative responses were correctly obtained
- Concentration responses were comparable across runs and laboratories, some differences were observed in few runs, mainly by CitoxLAB, where the responses started well below or above the solvent control (SC) level.
- Estimates of Flutamide IC_{50} were comparable (with slightly lower values for ENVIGO) and all within the acceptance range.
- Within run variability was measured via CV of Flutamide $logIC_{50}$ and was comparable across laboratories except slightly higher variability at CitoxLAB.
- Between run variability was measured via CV of Flutamide $logIC_{50}$ and was comparable across laboratories with EURL ECVAM having slightly higher values.

The data presented by ENVIGO, using Steady Glo mix for luciferase measurements and a luminometer without injectors, were quite similar to the data obtained by the other 2 laboratories. The variability (within runs and between

runs) was slightly higher for certain runs but was considered not to pose a problem. Therefore, ENVIGO was allowed to continue with the Steady Glo mix for the remainder of the validation study.

The VMG decided to have monitoring guidance in the SOP for the RLUs of VC and SC values. The guidance consists of observing the RI of the VC and when higher than 5%, the cause and the impact on the data shall be identified. In addition, the RI of the Flutamide C1 concentration (the lowest concentration) shall be monitored. If its value is above 120% or below 80%, the cause shall be identified, e.g. too high SC response, and the impact on the calculation of the parameters investigated.

The laboratories ENVIGO and RISE had demonstrated a successful transfer of the test method in their facility and were allowed to continue with the next study (Module 4). For CitoxLAB, a successful transfer of the agonist assay was concluded. To be considered fully successful with the transfer of the antagonist assay, the laboratory was recommended to investigate and to try to remediate the issue with the SC level in order to swiftly progress in the following study. The laboratory was recommended to perform 2 to 3 additional antagonist experiments. Upon receiving the data from these additional runs, the VMG reconvened in July 2017 to review these data and discuss the observed influences of glass ware on the results. CitoxLAB seemed to have remediated the technical issue of high SC values by changing the glass tubes for preparing chemical's working solutions, leading to acceptable dose responses in the two last tests. It was concluded that apparently the usage of certain glass tubes resulted in complications and that the change to other tubes had improved the implementation of the test method. The laboratory could continue the validation study. The laboratory was recommended to keep investigating its tubes (glass or plastic) for proper use with ED methods (see further information in section 6.5).

On the basis of the Transfer phase results, and observations from the 3 laboratories, few modifications were recommended to be taken up in the AR-CALUX® SOP as follows:

- Broadening up the SOP for the usage of other luminometers and not only the double injector luminometer
- An additional guidance for the antagonist assay to monitor carefully the VC response and the SC response
- Minor rewording and correction of certain sections in order to make the instructions clearer and less prone to different interpretations.

6 WITHIN LABORATORY REPRODUCIBILITY (MODULE 3 / STUDY 2)

Reference documents:

- Statistical report (Annex 13.1)
- SOP versions V05 and V06, and, SOP for solubility ASY15-V01 (Annex 13.6)
- Study plans and study reports of the participating laboratories for Study 2 (Reproducibility) (Annex 13.7.2)

6.1 General Aspects

WLR was assessed on the data obtained from 20 coded test chemicals, tested in 3 independent and valid runs. Before applying the AR-CALUX® SOP, the laboratories were required to assess the solubility of each of the test chemicals, both in the solvent and in the assay medium. Hereto, the SOP ASY15 V01 had been provided (visual inspection). The results of all laboratories are summarised and discussed in section 6.2.

WLR of the AR-CALUX® method was assessed on the concordance of classification between the 3 independent runs. Additionally, a reproducibility analysis of the EC_{50} and IC_{50} values was performed.

The classifier proposed in the AR-CALUX® method SOP V06 allowed besides "Negative" and "Positive" classification also the option of "Inconclusive" or "I" (see section 4.6). This option was introduced by the VMG for cases where the test chemical displays an activity for only one concentration point above the threshold of 10% (agonist) or below the threshold of 80% (antagonist).

The VMG had introduced for the antagonist assay a specificity control. The criterion R^2 had been introduced immediately in the classifier of SOP V06. A $R^2 > 0.9$ would indicate a false positive and would therefore be classified as a negative.

During the final review of the obtained data in the validation study, the VMG opted to modify the classifier and the SOP. The application of the modified SOP/classifier to the data of the validation study is discussed in section 10.

Table 13: Test chemicals for assessment of reproducibility (20 coded test chemicals)

TEST CHEMICAL	CAS No.	AR reference list		EURL ECVAM'S ARTA data list	
		AGONIST	ANTAGONIST	AGONIST	ANTAGONIST
Fluoxymestrone	76-43-7	Strong/moderate	NA		
17β-Trenbolone	10161-33-8	Strong	NA		
Medroxyprogesterone acetate	71-58-9	Moderate/weak	NA		
Stanozolol	10418-03-8	Moderate	NA		
Spironolactone	52-01-7	NA	Strong/moderate		
Hydroxyflutamide	52806-53-8	NA	Strong		
Bisphenol A	80-05-7	NA	Moderate/weak		
Vinclozolin	50471-44-8	NA	Moderate/weak		
Prochloraz	67747-09-5	Negative	Moderate/weak		
Bicalutamide	90357-06-5	NA	Strong		
Butylbenzyl phthalate	85-68-7	Negative	NA		
Tamoxifen	10540-29-1	Negative	NA		
Atrazine	1912-24-9	Negative	Negative		
Sodium azide	26628-22-8			N, N, N	N, N, N
Methyldihydrotestosterone	521-11-9			P, P, P, P	N, N, N
Propylthiouracil	51-52-5			N, N, N, N	N, N, N, N, N
Diethylhexyl phthalate	117-81-7			N, N, N, N, N	N, N, N, N, N
17β-Estradiol	50-28-2			P, P, P, P, P	P, P, FP, N
17α-Ethinyl estradiol	57-63-6			N, P, N, P, N	P, P, P, P
Disulfiram	97-77-8			N, P, N	FP, FP, FP

N= negative response
P= positive response
AR reference list: ICCVAM list of AR Reference chemicals (Kleinstreuer *et al*, 2017)
EURL ECVAM's ARTA data list is described in Annex 13.3

6.2 Solubility Data of the 4 Laboratories for 20 coded Test Chemicals

The data reported by each of the 4 laboratories for 20 test chemicals is summarised below. The laboratories had been asked to find for each of the test chemicals the appropriate solvent, to determine the highest solubility in the chosen solvent starting from a concentration of 50 mg/ml, to determine the highest solubility in assay medium. The measurements are reported in mg/ml and µg/ml.

All 4 laboratories reported slightly different soluble concentrations but the variation was always between 50 and 1.5 mg/ml or µg/ml. The laboratory RISE was more conservative than the other laboratories in its observations and determinations. The test chemical Sodium Azide was found to be soluble in water by all 3 EU NETVAL laboratories while BDS choose for DMSO.

In few cases, the differences in determined solubility between the labs influenced the classification in the antagonist assay, i.e. inconclusive classification was obtained.

- Test chemical 1, scored "I" 3 times by ENVIGO, "I" one time by BDS, due to a higher tested concentration
- Test chemical 18, scored "P" by 3 labs and "I" for one run in RISE, due to lower concentrations tested
- Test chemical 19 where BDS started with one higher concentration point (150 µg/ml) in its comprehensive testing with the antagonist assay which led to the different classifications.

For the purpose of the validation study, the laboratories had received instructions to start from 50 mg/ml as the maximal concentration. Converted in molar, the range for the chemicals tested would be between 40 mM and 330 mM with few at higher concentrations e.g. 769 mM for Sodium Azide.

Table 14: Solubility results of all 4 laboratories

Solubility in DMSO [mg/ml]						Solubility in medium [ug/ml]				
Chem ID	CitoxLAB	ENVIGO	RISE	BDS		Chem ID	CitoxLAB	ENVIGO	RISE	BDS
1	50	50	50	50		1	5	50	15	50
2	50	50	5	15		2	5	5	5	2
3	50	50	15(15)	50		3	50	50	15(15)	50
4	5	15	1.5	15		4	5	5	1.5	15
5	50	50	50	50		5	50	50	15	50
6	50	50	50	50		6	50	50	50	50
7	50	50	50	50		7	15	15	5	15
8	15	15	5	15		8	5	15	5	15
9	5	50	15	50		9	5	15	15	15
10	50	50	15	50		10	15	15	15	15
11	15	15	15	15		11	5	15	15	15
12	50	50	15	50		12	50	15	15	50
13	15	5	5	15		13	5	5	5	5
14	50	50	50	50		14	15	15	5	15
15	50	50	50	50		15	50	50	50	50
16	50	5	50	50		16	15	5	5	5
17	50	50	50	50		17	5	15	5	15
18	50	50	50	50		18	15	50	1.5	15
19	50	50	50	50		19	50	50	50	50
20	50	50	1.5	50		20	50	50	1.5	50

Green: identical values
Yellow: lowest reported value(s)
Values in bracket for chemical 3 means it was tested twice (in Study 2 and 3)

6.3 Reproducibility (Concordance of classifications)

The data reported by the labs, including dose responses for 20 coded test chemicals for each of the 4 laboratories, can be found in the Statistical report. Tables 15 and 16 (below) show the overview of the classifications as reported by all 4 laboratories except for the test chemicals 1 to 10 analysed by RISE and ENVIGO. The SOP used by these 2 laboratories did not contain the classifier yet and classification has been assigned by EURL ECVAM's statistician according to the classifier of SOP V06 (indicated in light green shade in Tables 15 & 16).

Verification of all other classifications was carried out by EURL ECVAM's statistician by re-applying the classifier on the reported data. This re-analysis led to few changes in the antagonist classifications due to few laboratories not applying the "Inconclusive" option of the classifier. These changes are denoted in the tables with * and + and explained briefly under the table.

WLR of the AR-CALUX® method was assessed on the concordance of classification between the 3 independent runs.

AGONIST assay

The classification "I" for "Inconclusive" was reported 4 times by only 2 of the 4 laboratories. Each time it was just one of the 3 runs within a lab resulting in such "I" (see Table 15A). Despite this one concentration slightly above the threshold, the dose responses per test chemical showed very good reproducibility (see dose response figures of test chemicals 3, 8, 17 and 20 in the Statistical report – Annex 13.1).

Including the "I" classification for the WLR assessment (e.g. NIN is considered as not concordant classification), the evaluation for the 4 laboratories resulted in **89%, 95%, 100% and 100%.**

Table 15: **Within Laboratory Reproducibility**

AGONIST													
TEST CHEMICAL		RISE			ENVIGO			BDS			CitoxLAB		
1	17β-Trenbolone	P	P	P	P	P	P	P	P	P	P	P	P
2	Stanozolol	P	P	P	P	P	P	P	P	P	P	P	P
3	Spironolactone(#)	N	N	N	N	I (1)	N	N	N	N	N	N	N
4	Medroxyprogesterone acetate	P	P	P	P	P	P	P	P	P	P	P	P
5	Bisphenol A	N	N	N	N	N	N	N	N	N	N	N	N
6	Bicalutamide	N	N	N	N	N	N	N	N	N	N	N	N
7	Disulfiram	N	N	N	N	N	N	N	N	N	N	N	N
8	Tamoxifen	N	N	N	** (2)	N	N	N	N	N	N	N	N
9	Atrazine	N	N	N	N	N	N	N	N	N	N	N	N
10	17α-Ethynyl estradiol	N	N	N	N	N	N	N	N	N	N	N	N
11	Sodium azide	N	N	N	N	N	N	N	N	N	N	N	N
12	Diethylhexyl phthalate	N	N	N	N	N	N	N	N	N	N	N	N
13	Methyldihydrotestosterone	P	P	P	P	P	P	P	P	P	P	P	P
14	Vinclozolin	N	N	N	N	N	N	N	N	N	N	N	N
15	Prochloraz	N	N	N	N	N	N	N	N	N	N	N	N
16	Fluoxymesterone	P	P	P	P	P	P	P	P	P	P	P	P
17	17β-Estradiol	P	P	I (3)	P	P	P	P	P	P	P	P	P
18	Benzylbutyl phthalate	N	N	N	N	N	N	N	N	N	N	N	N
19	Propylthiouracil	N	N	N	N	N	N	N	N	N	N	N	N
20	Hydroxyflutamide	N	N	N	N	I (4)	N	N	N	N	N	N	N
	WLR	**95%**			**89%**			**100%**			**100%**		
	Concordant/total classifications	19/20			17/19			20/20			20/20		

P = Positive response, N= Negative response, I=Inconclusive, as defined by the application of the classifier in SOP V06.
** = disqualified (chemical not included for WLR assessment)
(#): was tested twice by RISE under two different codes and resulted in the same outcome
Green left upper corner section: classification by EURL ECVAM as defined by the classifier in SOP V06

Summary notes:
(1) Border case, the second highest tested concentration has a RI of 10.3%
(2) Data points with lots of variability between the technical replicates, leading to 3 mean values > 10% threshold. Disqualified for analysis. Run would have been best repeated.
(3) Though 3 runs started with the same highest concentration, in only one run this concentration was scored as cytotoxic
(4) Border case, the highest tested concentration has a RI of 11.7%
All inconclusive cases are described in detail in the Statistical report page 18.

ANTAGONIST assay

As was observed in the agonist assay, the classification "I" was obtained few times by all the laboratories, due to the highest tested concentration below the 80% threshold (for the *test chemicals 1, 4, 8, 9, 18 and 19*). For details see Notes below Table 15B and the Statistical report – Annex 13.1).

Including the "I" classification for the WLR assessment of the 4 laboratories, WLR resulted in **75%, 80%, 85% and 90%.**

Test chemical 1, 17β-Trenbolone, was consistently scored as "I" for ENVIGO but not so in the other laboratories. This was due to one higher tested concentration by ENVIGO which was recorded at the other laboratories as cytotoxic and excluded from further testing. At BDS, it was scored as "I" once due to the usage of also one higher test concentration only in this particular run. This concentration was scored cytotoxic in the other 2 runs.

Test chemicals 4, 8 and 9, displayed RI values close to the 80% threshold value for one or two runs

Test chemical 19, Propylthiouracil, yielded a mix of classifications between all 4 laboratories due to its response at the highest concentration around the 80% threshold.

The lower WLR in RISE (75%) is due to the *test chemical 4* (Medroxyprogesterone acetate) and *test chemical 15* (Prochloraz) where different classifications were scored amongst the 3 runs.

Table 15B: Within Laboratory Reproducibility													
ANTAGONIST													
TEST CHEMICAL		RISE			ENVIGO			BDS			CitoxLAB		
1	17β-Trenbolone	N	N	N	I (1)	I	I	I (2)+	N	N	N	N	N
2	Stanozolol	N	N	N	N	N	N	N	N	N	N	N	N
3	Spironolactone(#)	P	P	P	P	P	P	P	P	P	P	P	P
4	Medroxyprogesterone acetate	N(14)	P	P	I (3)	N	N	N	N	N	N	N	N
5	Bisphenol A	P	P	P	P	P	P	P	P	P	P	P	P
6	Bicalutamide	P	P	P	P	P	P	P	P	P	P	P	P
7	Disulfiram	N	N	N	N	N	N	N	N	N	N	N	N
8	Tamoxifen	N	N	N	I(4)	N	N	N	I (5)+	N	N	N	N
9	Atrazine	I (6)	N	N	N	N	N	N	N	N	N	N	I (7)
10	17α-Ethynyl estradiol	P	P	P	P	P	P	P	P	P	P	P	P
11	Sodium azide	N	N	N	N	N	N	N	N	N	N	N	N
12	Diethylhexyl phthalate	N	N	N	N	N	N	N	N	N	N	N	N
13	Methyldihydrotestosterone	N	N	N	N	N	N	N	N	N	N	N	N
14	Vinclozolin	P	P	P	P	P	P	P	P	P	P	P	P
15	Prochloraz	P	N(15)	P	P	P	P	P	P	P	P	P	P
16	Fluoxymesterone	N	N	N	N	N	N	N	N	N	N	N	N
17	17β-Estradiol	P	P	P	P	P	P	P	P	P	P	P	P
18	Benzylbutyl phthalate	P	P	I (8)	P	P	P	P	P	P	P	P	P
19	Propylthiouracil	I (9)	I	N	I (10)*	N	I*	I (11)+	P	I+	I (12)	N	I
20	Hydroxyflutamide	P	P	P	P	P	P	P	P	P	P	P	P
	WLR	**75%**			**80%**			**85%**			**90%**		
	Concordant/total classifications	15/20			16/20			17/20			18/20		

P = Positive response, N= Negative response, I=Inconclusive, as defined by the application of the classifier in SOP V06
N(FP)= Negative classification due to false positive (FP) outcome by $R^2 > 0.9$
(#): was tested twice by RISE under two different codes and resulted in the same outcome
Green left upper corner section: classification by EURL ECVAM as defined by the classifier in SOP V06
* Was reported by the lab as "N" due to lab not applying the Inconclusive option
+ Was reported by the lab as "N" due to lab not applying the Inconclusive option
All inconclusive cases are described in detail in the Statistical report page 18.

Summary notes:
(1) Highest tested concentration in 3 runs with a RI of 19.9%, 12.7%, 7% at ENVIGO. This concentration point was not considered in 2 other laboratories due to cytotoxic scoring except by BDS (see below in (2)).
(2) Run 1 with one higher tested concentration with a RI 28.7% (this concentration point was scored as cytotoxic in the other 2 runs)
(3) Border case of being "N" due to 3rd highest concentration with a RI of 77.5%.
(4) Border case of highest tested concentration with a RI of 77.2%
(5) Border case of highest tested concentration with a RI of 76.7%
(6) Border case of highest tested concentration with a RI of 77.6%
(7) Border case of highest tested concentration with a RI of 77.9%
(8) Border case of being "P" due to only highest tested concentration with RI of 48.1% and 2nd highest with a RI of 83%
(9) Border case in 2 runs of highest tested concentration with a RI of 78.8% and 76.4 %
(10) Two runs with highest tested concentration with a RI of 66.3% and 71.6%
(11) Two runs with highest tested concentrations with a RI of 70.2% and 63.7%
(12) Two runs with highest tested concentration with a RI of 69.1% and 77.5%
(14) One run yielded a "P" conclusion due to 2 consecutive concentration points just below 80% (79.2% and 79.8%) which are the 3rd and 4th highest tested concentrations. The second run yielded a RI between 70.2 and 77.3% for the 4 highest tested concentrations
(15) The "N" is due the specificity control response with $R^2 = 0.91$ which is > 0.9 and hence indicates a false positive FP, leading to N.

6.4 Reproducibility of EC_{50} and IC_{50} values

EC_{50}, PC_{10}, IC_{50} and IC_{80} values and the mean, SD and CV's per item and per lab can be found in the Statistical report (see section of Potency statistics, page 29 and further, and, Appendix B, page 61 and further). Shown below are the calculated average CV's of all data in this study for the reference chemicals and test chemicals. These values show a good reproducibility of the method within each lab.

Table 16A: Average coefficients of variation of the reference chemicals for EC_{50} and IC_{50} (in M)

AGONIST						ANTAGONIST				
	CitoxLAB	ENVIGO	RISE	BDS			CitoxLAB	ENVIGO	RISE	BDS
DHT						Flutamide				
CV log(EC_{50})	0.96%	1.37% 2.54%	0.72%	0.96%		CV log(IC_{50})	1.97%	2.72% 1.47%	0.67%	1.09%

The following values are reported: for ENVIGO: study2A and 2B, for RISE, BDS and CitoxLAB: study 2

Table 16B: Average coefficients of variation of the test chemicals where EC_{50}/IC_{50} could be calculated (in mg/ml)

AGONIST - CV log(EC_{50})						ANTAGONIST - CV log(IC_{50})			
	CitoxLAB	ENVIGO	RISE	BDS		CitoxLAB	ENVIGO	RISE	BDS
1	0.80%	0.54%	0.50%	0.32%					
2	1.14%	1.54%	0.37%	0.84%					
3						0.32%	0.78%	0.31%	0.37%
4	0.57%	6.42%	1.21%	1.01%					
5						0.97%	0.82%	0.15%	0.32%
6						1.64%	0.30%	2.77%	1.26%
7									
8									
9									
10						1.45%	2.12%	0.90%	1.69%
11									
12									
13	0.82%	0.47%	0.44%	0.28%					
14						0.32%	2.02%	0.47%	3.93%
15						1.08%	1.44%	0.95%	0.62%
16	1.46%	1.82%	0.74%	1.00%					
17						2.22%	2.30%	0.96%	0.98%
18						1.97%	1.87%	1.56%	1.79%
19									
20						0.87%	0.57%	0.46%	0.83%

Table 16C: Average coefficients of variation of all test chemicals where EC_{50}/IC_{50} could be calculated (in M)

AGONIST						ANTAGONIST				
	CitoxLAB	ENVIGO	RISE	BDS			CitoxLAB	ENVIGO	RISE	BDS
All test items						All test items				
CV log(EC_{50})	0.96%	2.16%	0.65%	0.69%		CV log(IC_{50})	1.20%	1.28%	0.95%	1.31%

6.5 Conclusion of the VMG

Visual inspection of the dose responses of all tested chemicals, and a first inspection of the concordance of the classifications as reported by the labs, led already to the conclusion that the results are reproducible.

The classifier proposed by the test method developer had originally no conclusion for cases where the test chemical would display only one concentration passing the threshold. The option "Inconclusive" was introduced by the VMG for these cases. Such conclusion would trigger further testing though this had not been required for the validation study. From the study, it can be concluded that "I" was obtained for 4 chemicals in the agonist assay and 9 chemicals in the antagonist assay, mostly in only one of the 12 runs performed by the 4 labs. Most of the time, the "I" classification was due to the RI value of the highest tested concentration (C8) close to the threshold and a repeated test could likely result in "P" or "N". The labs had been asked to test 50 mg/ml as the maximal concentration which corresponds to a range of 39.8 mM to 769 mM for all chemicals tested in this validation study. For about half of the observed "I" conclusions, the concentration of test chemical used was higher than 100 mM which is the maximal concentration recommended in TG 455. The VMG considered that 100 mM as highest stock concentration would be sufficient for testing.

The WLR for all 4 laboratories, with inclusion of the "I", could be calculated to be 89%, 95%, 100% and 100% for the agonist testing. For antagonist testing, the WLR is 75%, 80%, 85% and 90%.

The VMG agreed that the WLR results were sufficient though impacted by the "I" option of the classifier. It was concluded that this classification of "I" would benefit from some modifications. Reviewing all data obtained at the end of the validation study, the VMG decided on providing instructions in the SOP on how to handle cases where the test chemical would display activity only at the highest tested concentration and pass the threshold values of 10% (agonist testing) and 80% (antagonist testing). In addition, a reformulation of the classifier was suggested. Application of this new guidance in the SOP/classifier to the validation study results resolved the situations where chemicals were classified as "I". The WLR values would increase to 95%, 100%, 100% and 100% for agonist testing, and, 94.7%, 100%, 100% and 100% for antagonist testing (see further in section 10).

The measurements of the parameters (EC_{50} and IC_{50}) within the laboratories were very reproducible. The overall CV's for the reference chemicals were between 0.69% and 2.72% and for all the test items between 0.65% and 2.16% for the agonist assay, and, between 0.95% and 1.28% for the antagonist assay.

Within the 20 coded test chemicals, 9 showed a positive antagonist response even when cytotoxic concentrations had been removed from the dose response. The specificity control, introduced to identify competitive antagonists, proved to be helpful. Dose responses were obtained where already visually one could deduce a clear shift between the dose responses of lower and higher supplemented DHT concentration. The introduction of the criterion R^2 made the decision objective. Among the collection of 20 coded test chemicals, only one test chemical was reported false positive in only one of the 3 technical replicates (test chemical 15, Prochloraz). This will be further discussed in section 10.

The chemical Disulfiram had been proposed by the VMG as a possible false positive given that it was scored as such in the Tox21 luc assay. This test chemical tested with the AR-CALUX® method was scored by all laboratories as cytotoxic at the higher concentrations leading to dose responses above the 80% RI and resulting in the conclusion of "N". The inconsistency of the response between the Tox21 luc assay and the AR-CALUX® assay is very likely due to the usage of a different test system (osteosarcoma AR-CALUX® cells versus breast cancer MDA-kb2 cells).

At the start of Study 2 (testing 20 coded chemicals), one laboratory encountered technical issues which were also observed during the transfer phase. The cause of these technical issues could not be clearly identified but seemed to be linked to the usage of certain glass and plastic ware. Such issues were not reported by the other 3 laboratories. Remediation in this particular situation had been possible due to using plastic 24-well plates instead of glass tubes for the preparation of the test chemical working solutions. The VMG therefore suggested that the final version of the SOP would instruct to use either glass ware or 24-well plates. In addition, a warning would be issued to verify all material from interference or contamination. Hereto, the response to the vehicle control should be assessed to ensure no interference from glass or plastic ware before running studies.

7 BETWEEN LABORATORY REPRODUCIBILITY (MODULE 4 / STUDY 2)

Reference documents:

- Statistical report (Annex 13.1)
- SOP versions V05 and 06, and, SOP for solubility (Annex 13.7.3)
- Study plans and study reports of the participating laboratories for Study 2 (Reproducibility) (Annex 13.7.2)

7.1 General Aspects

BLR was assessed on the same set of data that was evaluated for WLR with inclusion of the "I" (see Statistical report page 27). It was assessed based on the concordance of classifications: this includes that the classifications from the 3 valid runs per test chemical will lead to one classification per test chemical based on the mean of the 3 (majority rule, see Table 17). For example, a classification of NNN resulted in N; a classification of NPN resulted in N. Concordance of the classifications was subsequently evaluated for 3 labs by making all possible combinations.

In addition, the reproducibility of the EC_{50} and IC_{50} values for the reference chemicals and the tested chemicals was evaluated.

7.2 Reproducibility (concordance of classifications)

Evaluating concordance of classification, an overall BLR of **100%** was observed for agonist testing and an overall BLR of **87.5%** for antagonist testing (Table 17 and 18). Two chemicals did not have a consistent classification: *17β-Trenbolone*, due to one "I" classification in the antagonist assay in one lab, and, *Medroxyprogesterone acetate* due to one "P" classification in one lab. *Propylthiouracil* yielded "I" in the antagonist assay in each lab. It was considered for BLR evaluation as displaying non concordant classifications (i.e. non conclusive result).

Table 17: Between laboratory reproducibility with concordance of classifications (majority rule)

TEST CHEMICAL		AGONIST				ANTAGONIST			
		RISE	ENVIGO	BDS	Citox LAB	RISE	ENVIGO	BDS	Citox LAB
1	17β-Trenbolone	P	P	P	P	N	**I**	N	N
2	Stanozolol	P	P	P	P	N	N	N	N
3	Spironolactone	N	N	N	N	P	P	P	P
4	Medroxyprogesterone acetate	P	P	P	P	**P**	N	N	N
5	Bisphenol A	N	N	N	N	P	P	P	P
6	Bicalutamide	N	N	N	N	P	P	P	P
7	Disulfiram	N	N	N	N	N	N	N	N
8	Tamoxifen	N	N	N	N	N	N	N	N
9	Atrazine	N	N	N	N	N	N	N	N
10	17α-Ethynyl estradiol	N	N	N	N	P	P	P	P
11	Sodium azide	N	N	N	N	N	N	N	N
12	Diethylhexyl phthalate	N	N	N	N	N	N	N	N
13	Methyldihydrotestosterone	P	P	P	P	N	N	N	N
14	Vinclozolin	N	N	N	N	P	P	P	P
15	Prochloraz	N	N	N	N	P	P	P	P
16	Fluoxymesterone	P	P	P	P	N	N	N	N
17	17β-Estradiol	P	P	P	P	P	P	P	P
18	Benzylbutyl phthalate	N	N	N	N	P	P	P	P
19	Propylthiouracil	N	N	N	N	**I**	**I**	**I**	**I**
20	Hydroxyflutamide	N	N	N	N	P	P	P	P

Table 18: BLR calculations									
	AGONIST					ANTAGONIST			
Lab 1 Lab 2 Lab 3	CitoxLAB ENVIGO RISE	CitoxLAB ENVIGO BDS	ENVIGO RISE BDS	CitoxLAB RISE BDS		CitoxLAB ENVIGO RISE	CitoxLAB ENVIGO BDS	ENVIGO RISE BDS	CitoxLAB RISE BDS
Majority rule									
# test chemicals	20	20	20	20		20	20	20	20
# Concordant classifications	20	20	20	20		17	18	17	18
BLR	100%	100%	100%	100%		85%	90%	85%	90%
Overall BLR	**100%**					**87.5%**			

7.3 Reproducibility of the EC_{50} and IC_{50} Values

The reference chemicals

The statistical report (Appendix B, page 61 and further) shows all results obtained per laboratory with mean, SD and CV.

Table 19 (below) shows average calculations of the results obtained for the acceptance criteria by all 4 laboratories. Each lab obtained few invalid runs/plates (acceptance criteria not met). Absolute numbers cannot be compared as the number of test chemicals included per run could be different between runs and between labs. For details, see Statistical report in Annex 13.1.

The values for the acceptance criteria are quite comparable. It can be observed that CitoxLAB obtained higher values for the negative control in both agonist and antagonist assay but still within the criterion values set: < 10% in the agonist assay, and, > 85% in the antagonist assay. The vehicle control (VC) in the antagonist assay with CitoxLAB revealed higher values than with the other labs.

Table 19: Overview of obtained results for number of runs and acceptance criteria in the 4 laboratories for all chemicals tested

AGONIST ASSAY					ANTAGONIST ASSAY				
	RISE 46 chemicals	CitoxLAB 20 chemicals	ENVIGO 20 chemicals	BDS 20 chemicals		RISE 46 chemicals	CitoxLAB 20 chemicals	ENVIGO 20 chemicals	BDS 20 chemicals
# of total runs	24	17	13	10	# of total runs	27	24	19	17
# of valid runs	20	15	13	10	# of valid runs	20	15	16	17
Average across valid runs					Average across valid runs				
DHT EC_{50} (M)	2.7 E-10	3.5 E-10	3.23 E-10	3.57 E-10	Flutamide IC_{50} (M)	4.56 E-07	5.11 E-07	6.01 E-07	4.91 E-07
CV($logEC_{50}$)	0.31%	0.41%	0.58%	0.42%	CV($logIC_{50}$)	0.68%	0.61%	**1.02%**	0.57%
Induction factor (IF)	71	**129**	69	84	IF	33.7	51.5	31.2	29.6
ZF	0.85	0.86	0.79	0.86	ZF	0.78	0.8	0.7	0.81
Positive control (PC) relative induction	78.9	78.2	76.9	77.1	Positive control (PC) relative induction	13.4	11.9	13.1	12.4
Negative control (NC) relative induction	0.1	**6.2**	0.9	0.2	Negative control (NC) relative induction	144	**177.3**	123.6	139
					Vehicle control (VC) Relative induction	0.1	**5.7**	1.6	5.4 *

Indicated in blue: highest value obtained

* Overall average with elimination of one (high) value. After receipt of the DAFs and the final report, it was noted that the VC values of one run were exceptionally high, resulting in an overall average VC of 11%. While inquiring about this high VC value, BDS informed about an operator error made.

The Test Chemicals

Where it was possible, **EC_{50} and IC_{50} values** were calculated, as well as **PC_{10} and PC_{80}** for the tested chemicals (see Statistical report section of Potency statistics, page 29 and further). The values were very reproducible within the laboratories as well as between the laboratories.

Table 20 (below) shows the average of all EC_{50} and IC_{50} values obtained per test chemical per lab (20 coded chemicals for ENVIGO, BDS and CitoxLAB and RISE).

Table 20 A: Average EC_{50} values of the test chemicals in agonist testing (in mg/ml)								
		CitoxLAB	ENVIGO	RISE	BDS			
No.	TEST CHEMICAL	Log(EC_{50}) average of all runs				Mean	SD	CV
1	17β-Trenbolone	-7.06	-7.23	-7.26	-7.18	-7.18	0.09	1.2%
2	Stanozolol	-6.15	-6.31	-6.26	-6.27	-6.25	0.07	1.1%
3	Spironolactone(#)	NA	-5.16	NA	NA	-5.16		
4	Medroxyprogesterone acetate	-5.22	-6.01	-5.52	-5.53	-5.57	0.32	5.8%
5	Bisphenol A							
6	Bicalutamide							
7	Disulfiram							
8	Tamoxifen							
9	Atrazine							
10	17α-Ethynyl estradiol							
11	Sodium azide							
12	Diethylhexyl phthalate							
13	Methyldihydrotestosterone	-6.96	-7.03	-6.98	-7.00	-6.99	0.03	0.4%
14	Vinclozolin							
15	Prochloraz							
16	Fluoxymesterone	-4.91	-5.47	-5.28	-5.50	-5.29	0.27	5.1%
17	17β-Estradiol	NA	-3.39	NA	NA	-3.39		
18	Benzylbutyl phthalate							
19	Propylthiouracil							
20	Hydroxyflutamide							

Average log(EC_{50}) value result from the log(EC_{50}) from 3 valid and independent runs (in mg/ml). Mean, SD and CV are taken from all values per test chemical
NA = not available, value could not be calculated

Table 20 B: Average IC_{50} values of the test chemicals in antagonist testing (in mg/ml)								
		CitoxLAB	ENVIGO	RISE	BDS			
No.	TEST CHEMICAL	Log(IC_{50}) average of all runs				Mean	SD	CV
1	17β-Trenbolone							
2	Stanozolol							
3	Spironolactone(#)	-4.82	-4.81	-4.84	-4.82	-4.82	0.02	0.3%
4	Medroxyprogesterone acetate							
5	Bisphenol A	-3.51	-3.48	-3.52	-3.49	-3.50	0.02	0.5%
6	Bicalutamide	-4.29	-4.21	-4.27	-4.50	-4.32	0.12	2.9%
7	Disulfiram							
8	Tamoxifen							
9	Atrazine							
10	17α-Ethynyl estradiol	-4.93	-4.94	-4.95	-4.94	-4.94	0.01	0.1%
11	Sodium azide							
12	Diethylhexyl phthalate							
13	Methyldihydrotestosterone							
14	Vinclozolin	-4.57	-4.39	-4.52	-4.73	-4.55	0.14	3.1%
15	Prochloraz	-3.06	-3.13	-3.08	-3.07	-3.08	0.03	1.0%
16	Fluoxymesterone							
17	17β-Estradiol	-5.82	-5.40	-5.63	-5.61	-5.62	0.17	3.1%
18	Benzylbutyl phthalate	-2.75	-2.69	-3.26	-2.77	-2.87	0.26	9.1%
19	Propylthiouracil							
20	Hydroxyflutamide	-5.20	-5.13	-5.28	-5.29	-5.23	0.08	1.5%

Average log(IC_{50}) value result from the log(IC_{50}) from 3 valid and independent runs (from mg/ml). Mean, SD and CV are taken from all values per test chemical.

Relative potency measures, i.e. log(EC_{50})/log(DHT EC_{50}), log(PC_{10})/log(DHT PC_{10}), log(IC_{50})/log(FLUTAMIDE IC_{50}) and log(PC_{80})/log(FLUTAMIDE PC80) were calculated and compiled in tables 21A and B below. Values above 1 are referring to more potent chemicals than reference item (DHT or Flutamide) whereas values below 1 to less potent chemicals than reference item. The test chemicals are ordered vertically in the tables from lowest potency to highest potency. The order is identical when assessing EC_{50} or PC_{10} values, or, IC_{50} or PC_{80} values.

Table 21A	log(EC_{50}) / log(DHT EC_{50}) **averaged values**				log(PC_{10}) /log (DHT PC_{10}) **averaged values**			
Test chemical	CitoxLAB	ENVIGO	RISE	BDS	CitoxLAB	ENVIGO	RISE	BDS
17β-Estradiol					0.58	0.61	0.60	0.63
Fluoxymesterone	0.79	0.84	0.82	0.84	0.81	0.84	0.83	0.85
Medroxyprogesterone acetate	0.83	0.90	0.84	0.85	0.83	0.90	0.83	0.86
Stanozolol	0.91	0.93	0.91	0.93	0.91	0.93	0.91	0.92
Methyldihydrotestosterone	1.00	1.01	1.00	0.99	1.00	1.00	1.00	1.01
17β-Trenbolone	0.99	1.02	1.01	1.01	1.00	1.02	1.00	1.02
Norethindrone acetate			0.85				0.83	
Norethindrone			0.85				0.85	
Levonorgestrel			0.92				0.92	
Methyltestosterone			0.94				0.95	
Nandrolone			0.98				0.99	
Methyltrienolone (R1881)			0.99				0.99	

Table 21B	log(IC_{50}) / log(FLUTAMIDE IC_{50}) **averaged values**				log(PC_{80}) / log(FLUTAMIDE PC_{80}) **averaged values**			
Test chemical	CitoxLAB	ENVIGO	RISE	BDS	CitoxLAB	ENVIGO	RISE	BDS
Benzylbutyl phthalate	0.83	0.83	0.91	0.84	0.86	0.84	0.86	0.83
Prochloraz	0.91	0.92	0.89	0.89	0.95	0.90	0.89	0.90
Bisphenol A	0.94	0.93	0.94	0.93	0.95	0.94	0.96	0.93
Vinclozolin	1.12	1.04	1.10	1.13	1.13	1.04	1.09	1.12
Bicalutamide	1.11	1.09	1.10	1.12	1.10	1.08	1.11	1.14
Spironolactone(#)	1.19	1.18	1.19	1.19	1.19	1.20	1.19	1.15
17α-Ethynyl estradiol	1.20	1.18	1.18	1.19	1.25	1.18	1.20	1.18
Hydroxyflutamide	1.22	1.22	1.21	1.22	1.24	1.22	1.22	1.21
17β-Estradiol	1.33	1.28	1.27	1.27	1.33	1.26	1.27	1.24
Ketoconazole								
2-tert-Butylanthraquinone			0.89				0.84	
Linuron			0.91				0.91	
Diethylstilbestrol			0.91				0.94	
Finasteride			0.92				0.95	
2-sec-Butylphenol			0.93				0.95	
Arochlor1254			0.97				0.97	
Cycloheximide			0.97				0.99	
o,p'-DDT			0.99				1.01	
Corticosterone			1.00				1.03	
Mifepristone			1.19				1.16	
Progesterone			1.26				1.27	
Cyproterone acetate			1.27				1.27	
Actinomycin D			1.37				1.29	

7.4 Conclusion of the VMG

The main focus of the BLR evaluation for 20 coded test chemicals was on the concordance of the classifications. Overall BLR was 100% for agonist testing, and, 87.5% for antagonist testing. By applying the new guidance in the SOP/classifier to the obtained data, the BLR values would increase to 100% for both agonist and antagonist testing (see further in section 10).

When evaluating the standard deviations and the CVs of the average EC_{50} and IC_{50} values for the 20 coded test items in all labs tested, the VMG agreed that these values were low. The highest CV% was noted for *Benzylbutyl phthalate* in the antagonist assay, being 9.1%. All other tested chemicals had CVs of lower than 4 %.

The VMG agreed that the BLR results were very good for the AR-CALUX® method.

8 PREDICTIVE CAPACITY FOR 46 CODED TEST CHEMICALS (MODULE 5 / STUDY 2 AND 3)

Reference documents:

- Statistical report (Annex 13.1)
- SOP V06 and SOP for solubility V01 (Annex 13.7.3)
- Study plan and study report of RISE for Study 3 (Annex 13.7.2)

8.1 Additional Data set on 26 Test Chemicals

In addition to the data of 20 coded test chemicals (assessed for BLR), a set of data was generated for 26 test chemicals by one laboratory only (RISE), to be considered as well for predictive capacity. Solubility data and classifications can be found below and in the statistical report.

Solubility results

Table 22: Solubility results

Chem ID		Solubility in DMSO [mg/ml]	Solubility in medium [µg/ml]
21	Levonorgestrel	15	1.5
22	Cyproterone acetate	50	15
23	2-tert-Butylanthraquinone	15	5
24	Arochlor1254	50	5
25	Nandrolone	50	50
26	o,p'-DDT	50	1.5
27	Phenolphthalin	15	15
28	2,4,5-T	0.15	0.15
29	Methyltrienolone (R1881)	50	50
30	Actinomycin D	5	5
31	Diethylstilbestrol	50	15
32	L-Thyroxine	50	50
33	Haloperidol	50	5
34	Norethindrone acetate	5	5
35	Pimozide	5	5
36	Progesterone	50	15
37	Linuron	50	15
38	Methyltestosterone	50	15
39	2-sec-Butylphenol	15	15
40	Corticosterone	50	50
41	Ketoconazole	15	15
42	Finasteride	50	50
43	Fulvestrant	50	1.5
44	Cycloheximide	50	50
45	Norethindrone	50	5
46	Mifepristone	50	5

Classifications

Table 23: Classifications								
No.	TEST CHEMICAL	AGONIST				ANTAGONIST		
21	Levonorgestrel	P	P	P		N	N	N
22	Cyproterone acetate	N	N	N		P	P	P
23	2-tert-Butylanthraquinone	N	N	N		P	P	P
24	Arochlor1254	N	N	N		P	P	P
25	Nandrolone	P	P	P		N	N	N
26	o,p'-DDT	N	N	N		P	P	P
27	Phenolphthalin	N	N	N		N	N	N
28	2,4,5-T	N	N	N		N	N	N
29	Methyltrienolone (R1881)	P	P	P		N	N	N
30	Actinomycin D	N	N	N		I (1)	I	I
31	Diethylstilbestrol	N	N	N		P	P	P
32	L-Thyroxine	N	N	N		N	N	N
33	Haloperidol	N	N	N		N	N	N
34	Norethindrone acetate	P	P	P		I(2)	P	I
35	Pimozide	N	N	N		N	N	N
36	Progesterone	N	N	N		P	P	P
37	Linuron	N	N	N		P	P	P
38	Methyltestosterone	P	P	P		N	N	N
39	2-sec-Butylphenol	N	N	N		N(FP)(3)	P	P
40	Corticosterone	N	N	N		P	P	P
41	Ketoconazole	N	N	N		N(FP)	N(FP)	N(FP)
42	Finasteride	N	N	N		P	P	P
43	Fulvestrant	N	N	N		N(FP)	I(4)	P
44	Cycloheximide	N	N	N		N(FP)	N(FP)	N(FP)
45	Norethindrone	P	P	P		N	N	N
46	Mifepristone	N	N	N		P	P	P
	WLR	**100%**				**84.6%**		
	Concordant/total classifications	26/26				22/26		

P = Positive response, N= negative response, I=Inconclusive, as defined by the application of the classifier in SOPV06
N(FP)= Negative classification due to false positive (FP) outcome by $R^2 > 0.9$

Summary notes:

(1) Chemical 30: only the highest tested concentration has RI < 80% in 3 runs (57.2%, 39%, 37.5%). R^2 is > 0.9 in all 3 runs what indicates FP but with the classifier it gets classified as "I" due to only one concentration below the threshold value.

(2) Chemical 34 has a dose response with the specificity control that declines before the standard response (shift in the opposite direction). "I" is scored in 2 runs due to the highest tested concentration with RI of < 80%: 54.2% and 56.2%.

(3) Chemical 39: classified as N(FP) in the 1st run due to $R^2 = 0.95$

(4) Chemical 43: yielded 3 different classifications though the dose responses are quite comparable (see Index in the statistical report). The 2nd run (run 7) had one concentration point with RI of 111.5% and is likely an outlier.

WLR and variability for the 26 additional chemicals

For the chemicals tested under the agonist assay,
WLR was found to be 100 % and
% CV for log(EC_{50}) was 0.92.

For the chemicals tested under the antagonist assay,
WLR was found to be 84.6% and
% CV for log(IC_{50}) was 1.55.

8.2 Comparison of Classifications

The predictive capacity was evaluated on the basis of comparing the mean classifications of each laboratory in this validation study (as was done for the assessment of the BLR) for the 46 coded test chemicals with reported classifications from

1) ICCVAM AR-Reference chemical list published in 2017 (Kleinstreuer *et al*, 2017) (AR REF)

2) EURL ECVAM's list of publicly available data for ARTAs (for detailed info, see Annex 13.3). This list includes the classifications of 2 Tox21 ARTA assays for which the data were revisited in 2018 given the availability of cytotoxicity values, the results of the AR pathway computational model (Kleinstreuer *et al*, 2017) and the results of the 2 validated ARTAs (Japan ARTA of OECD TG 458, 2016; Korean ARTA, personal communication). Given the limited data sources in this list, caution has to be taken with the comparison and interpretation e.g. the chemicals Arochlor and 2-tert-butylanthraquinone were added in 2015 to the list because they were proposed as test chemicals for the validation study of the Korean ARTA but finally were not retained for generating data. Korea provided a provisional classification for these chemicals.

An overview of the comparison of the classifications in the validation study with the two lists is presented in Tables 25 and 26. It includes the results from the 20 test chemicals that were tested for BLR by all 4 laboratories, and, the additional 26 test chemicals tested by the laboratory RISE.

Comparison to the AR-Reference list (AR REF)

For 23 chemicals tested in the AR-CALUX® validation study, the classification could be compared to the AR-Reference list. An identical classification was observed for all tested chemicals where information was available in the AR-reference list with exception of *Cyproterone acetate* in the agonist assay (scored as "N" with the AR-CALUX® method and reported as "P" in the AR- Reference list). With the AR-CALUX® method (tested by one lab), the dose response is below the 10% threshold with a very slight increase at the highest concentration tested (4 µM). It is reported in the AR-Reference list that this chemical has a weak positive behaviour. Possibly, the reported values in literature, used for the AR-reference list, may have resulted from higher tested concentrations.

Performance values

Concordance of classifications were calculated versus the AR-Reference list only given that this list is the most trustable source of *in vitro* data. The values are listed in the Tables 24A and B below. Of the 23 test chemicals that could be compared, 13 were tested for (ant)agonist activity in all 4 laboratories and an additional 10 in one laboratory.

The AR-Reference list does not always provide info on both agonist and antagonist behaviour, leading to 8 chemicals that could be scored for agonist behaviour, and, 7 for antagonist behaviour for 3 of the 4 laboratories. The lab RISE tested more chemicals (16 for agonist, 12 to for antagonist, to be compared with the AR-Reference list). Only one chemical had a different classification (*Cyproterone acetate*, see also Table 25).

Table 24A: AGONIST

AR Reference	RISE		ENVIGO		BDS		CitoxLAB	
	P	N	P	N	P	N	P	N
P	10	1	4	0	4	0	4	0
N	0	5	0	4	0	4	0	4
Positive concordance	90.9%		100%		100%		100%	
Negative concordance	100%		100%		100%		100%	
Overall concordance	93.8%		100%		100%		100%	

Table 24B: ANTAGONIST

AR Reference	RISE		ENVIGO		BDS		CitoxLAB	
	P	N	P	N	P	N	P	N
P	10	0	6	0	6	0	6	0
N	0	2	0	1	0	1	0	1
Positive concordance	100%		100%		100%		100%	
Negative concordance	100%		100%		100%		100%	
Overall concordance	100%		100%		100%		100%	

Table 25: Comparison to the 1st set of 23 tested chemicals with the AR-Reference chemical list and other ARTA classifications

No	TEST CHEMICAL	AGONIST										ANTAGONIST									
		RISE	ENVIGO	BDS	Citox LAB	AR REF	ARTA JAPAN	ARTA KOREA	Tox21 Luc	Tox21 Bla	AR pathway	RISE	ENVIGO	BDS	Citox LAB	AR REF	ARTA JAPAN	ARTA KOREA	Tox21 Luc	Tox21 Bla	AR pathway
1	17β-Trenbolone	P	P	P	P	**P**			P	P	P	N	I[N]	N	N				N	P	N
2	Stanozolol	P	P	P	P	**P**						N	N	N	N				N		
3	Spironolactone	N	N	N	N				P	P	N	P	P	P	P	**P**			P	P	FP?
4	Medroxyprogesterone acetate	P	P	P	P	**P**	P	P	P	P		N	N	N	N			N	N	P	
5	Bisphenol A	N	N	N	N		N	N	N	N	N	P	P	P	P	**P**	P	P	P	P	P
6	Bicalutamide	N	N	N	N			N	N	N	N	P	P	P	P	**P**		P	P	P	P
8	Tamoxifen	N	N	N	N	**N**			N	P	N	N	N	N	N				P	?	N
9	Atrazine	N	N	N	N	**N**		N	N	N	N	N	N	N	N	**N**	N	N	N	N	N
14	Vinclozolin	N	N	N	N			N	N	N	N	P	P	P	P	**P**	P	P	P	P	P
15	Prochloraz	N	N	N	N	**N**		N	N	N	N	P	P	P	P	**P**	P	P	P	FP	P
16	Fluoxymesterone	P	P	P	P	**P**			P	P		N	N	N	N				N	N	
18	Benzylbutyl phthalate	N	N	N	N	**N**	N	N	N	N	N	P	P	P	P			P	P	N	N
20	Hydroxyflutamide	N	N	N	N		N	P	P		N	P	P	P	P	**P**	P	P	P	P	P
21	Levonorgestrel	P				**P**		P	P	P	P	N						N	N	P	N
22	Cyproterone acetate	N				**P weak**			P	N	P	P				**P**			P	P	P
25	Nandrolone (19-Nortestosterone)	P				**P**			P	P		N							N	N	
26	o,p'-DDT	N				**N**		N	N	N	N	P				**P weak**		P	P	FP	P
29	Methyltrienolone (R1881)	P				**P**						N									
34	Norethindrone acetate	P				**P**			P	P		I[P]							?	P	
37	Linuron	N						N	N	N	N	P				**P**		P	P	P	P
38	Methyltestosterone	P				**P**		P	P	P	P	N				**N**		N	N	N	N
45	Norethindrone	P				**P**			P	P	P	N							N	P	N
46	Mifepristone	N							P	N	N	P				**P**			P	P	FP?

Grey shading: AR-CALUX® classification consistent with at least one other ARTA classification; yellow shading: the inconclusive classification
Black box: AR-CALUX® classification not consistent with the AR-reference chemicals list (AR REF)
Blank: not tested in the ARTAs or not available in the AR REF
In straight brackets []: plausible classification with modified SOP/classifier as discussed in section 10
FP?: A shift of the specificity dose response in the opposite direction
?: classification could not be assigned due to incomplete data
AR REF: ICCVAM list of AR- Reference chemicals (Kleinstreuer *et al*, 2017)
ARTA Japan: classifications from validated ARTA; ARTA Korea: classifications from validated ARTA

In comparison to AR REF: all identical but not no. 22 in the agonist; in comparison to ARTA JP: all identical; in comparison to ARTA KR: all identical but not no. 20 in agonist.

Comparison to EURL ECVAM's list of ARTA classifications

The remaining 23 tested chemicals (for which there was no info in the AR-reference list) were compared with all the classifications compiled in the EURL ECVAM list. For 19 tested chemicals, an identical classification could be found with one or more reported ARTA classification in agonist and antagonist assay (see Table 26 below).

Few chemicals tested with the AR-CALUX® method were found to have consistent behaviour across all the ARTAs and the AR pathway model: *Diethylhexyl phthalate* (N) in the agonist and antagonist assay, *17β-Estradiol* (P) in the agonist assay, *Propylthiouracil* (N) in the antagonist assay. Given that more comparisons could be made with the Tox21 assays and the AR pathway model, 13 chemicals were found with identical classifications in agonist assay and 9 chemicals in the antagonist assay.

For few chemicals, the AR-CALUX® classification was different from all other classifications: no. 24, 36 and 40 for agonist testing; no 39 and no 43 in antagonist testing.

Arochlor (no 24): scored clearly "N" for the agonist testing with the AR-CALUX® method. Only a provisional classification was provided by Korea as "P".

Progesterone (no 36): scored as "N" for the agonist testing with the AR-CALUX® method and "P" in 3 other ARTAs and the AR-pathway model. Inspection of the dose responses shows that there is a slight increase at the highest tested concentration (4 µM) but still below the 10% threshold. The higher tested concentrations were found to be cytotoxic. Possibly, the chemical may have been tested at higher concentrations in the other ARTAs.

Corticosterone (no 40): scored as "N" in the agonist testing with the AR-CALUX® method and "P" in 3 other ARTAs and the AR-pathway model. AR-CALUX® cells are reported to have no GR activity due to absence of the GR receptor. Testing *Corticosterone* for agonist properties confirmed this. In other ARTAs, this chemical may display a "P" response for which a possible GR interference cannot be excluded. Moreover, this chemical is also used as the negative control item in the agonist assay of the AR-CALUX® method. Inspection of the dose responses shows that there is a slight increase at the highest tested concentration (171 µM).

2 sec Butylphenol (no 39) scored as a clear "P" in the agonist testing with the AR-CALUX® method though seems to be negative in 2 other ARTAs and the AR-pathway.

Fulvestrant (no. 43) scored a FP with the AR-CALUX® method though the classifications in the Tox21 ARTAs and AR pathway are reported as "P" and "N" respectively.

Disulfiram (no 7) was included as a potential FP antagonist for the AR- CALUX® method due to such classification in the Tox21 luc assay. It was observed in this assay that the chemical was active (dose response declining) at concentrations that were not cytotoxic. With the AR-CALUX® method, this chemical displayed cytotoxicity already at high concentrations, leading to a negative response. Non conformity in the classification is due to cell line species differences (osteosarcoma cell line versus breast cancer cell line).

17β-Estradiol (no 17), scored in the AR-CALUX® as "P" for both agonist and antagonist response by all 4 laboratories. It shows the same dual response in most of the other ARTAs and the AR pathway model (agonist).

Three (possibly 4) FP classifications were observed due to $R^2 > 0.9$*: Actinomycin D (no 30), Ketoconazole (no 39), Fulvestrant (no 43) and Cycloheximide (no 44).* This is further discussed in section 10.3. Only for *Ketoconazole*, the FP response is also observed with the other Tox21 assays and the AR-pathway model.

Table 26: Comparison of 2nd set of 23 tested chemicals with the EURL ECVAM's list of ARTA classifications (no AR-Reference chemical list info available)

No.	TEST CHEMICAL	AGONIST									ANTAGONIST								
		RISE	ENVIGO	BDS	Citox LAB	ARTA Japan	ARTA Korea	Tox21 Luc	Tox21 Bla	AR Pathway	RISE	ENVIGO	BDS	Citox LAB	ARTA Japan	ARTA Korea	Tox21 Luc	Tox21 Bla	AR Pathway
7	Disulfiram	N	N	N	N			N	P	N	N	N	N	N			FP	FP	FP
10	17α-Ethynyl estradiol	N	N	N	N	N	P	N	P	N	P	P	P	P		P	P	P	P
11	Sodium azide	N	N	N	N			N	N	N	N	N	N	N			N	N	N
12	Diethylhexyl phthalate	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
13	Methyldihydrotestosterone	P	P	P	P	P	P	p	P		N	N	N	N	N	N	?	N	
17	17β-Estradiol	P	P	P	P	P	P	P	P	P	P	P	P	P		P	P	FP	N
19	Propylthiouracil	N	N	N	N		N	N	N	N	I[N]	I[N]	I[N]	I[N]	N	N	N	N	N
23	2-tert-Butylanthraquinone	N					(N)				P					(P)			
24	Arochlor1254	N					(P)				P					(P)			
27	Phenolphthalin	N						N	N	N	N						N	N	N
28	2,4,5-Trichlorophenoxyacetic acid	N						N	N	N	N						N	N	N
30	Actinomycin D	N						N	N		I [FP]						FP	P	
31	Diethylstilbestrol	N						N	N	N	P						P	FP	P
32	L-Thyroxine	N						N	N		N						N	N	
33	Haloperidol	N						N	N	N	N						FP	N	FP?
35	Pimozide	N						N	N		N						?	FP	
36	Progesterone	N					P	P	P	P	P					P	FP	N	FP
39	2-sec-Butylphenol	N						N	N	N	P						N	N	N
40	Corticosterone	N					P	P	P	P	P					P	N	N	N
41	Ketoconazole	N						N	P	N	FP						FP	FP	FP
42	Finasteride	N						N	N	N	P						P	P	P
43	Fulvestrant	N						N	N	N	I[FP]						P	P	N
44	Cycloheximide	N						N	N	N	FP						FP	N	FP

Grey shading: AR-CALUX® classification consistent with at least one other ARTA classification; yellow shading: the inconclusive classification
Blank: not tested in the ARTAs or not available in the AR REF
(): Provisional classification reported by Korea
?: classification could not be assigned due to incomplete data
Black box: AR-CALUX® classification different from all others
In straight brackets []: plausible classification with modified SOP/classifier as discussed in section 10
AR REF: ICCVAM list of AR-Reference chemicals (Kleinstreuer *et al*, 2017)
ARTA Japan: classifications from validated ARTA
ARTA Korea: classifications from validated ARTA

In comparison to ARTA JAPAN (validated): all identical
In comparison to ARTA KOREA (validated): all identical but not no. 10, 24, 36 and 40 in agonist

8.3 Conclusion of the VMG

To investigate potential androgenic and anti-androgenic effects of chemicals *in vivo*, the rodent Hershberger assay had been developed and standardised. However, a high degree of variability in the results of Hershberger studies, including disagreements between the results for the same chemical, have been reported (Browne *et al*, 2017). In addition, the Hershberger assay is capable of detecting several modes of action, whereas the AR-CALUX® method assesses only directly acting AR (ant)agonists. A comparison between these assays may therefore not be helpful.

In the absence of good *in vivo* reference data, a comparison was made in this validation study to the published AR-reference list (Kleinstreuer *et al*, 2017). This list results from a targeted literature search for AR *in vitro* reference data, including AR binding data and transactivation data, for which a range of quality criteria were applied. The comparison showed that of 26 chemicals tested with the AR-CALUX® method, all but one of the chemicals tested displayed an identical classification.

9 APPLICABILITY DOMAIN (MODULE 6)

Limitations were not reported by the test submitter other than the generic ones related to working with *in vitro* systems. Poor solubility of the chemical or physico-chemical properties can be incompatible with the standard serum-containing tissue culture media. Metabolism of a chemical cannot be evaluated with this test system.

The method allows testing of liquids and solid chemicals as long as they can be solubilised in a solvent. DMSO and water were both used in this validation study as solvents.

The chemicals tested in this validation study have a spectrum of chemical classes that cover pharmaceutical usage (e.g. cancer treating drugs and antibiotics), industrial usage (e.g. plasticizers, lubrificants) and agricultural usage (pesticides, fungicides). When compared to the REACH chemical space (structural diversity) the ARCALUX® validation set covers a rather large area (see Annex 13.03).

10 DISCUSSION

10.1 Qualitative assessment of the results

10.1.1 Classification of the 20 tested chemicals (for WLR and BLR assessment)

During the final review of the data, the VMG decided on the following points regarding classification of the test chemicals:

- Expert judgement would be applied in those cases where the RI values obtained were border line close to the threshold values taking into account the value of the RI and the shape of the dose response (see Notes under tables 27 and 28).
- By applying the classifier of SOP V06, several "I" classifications were observed for test chemicals displaying activity for only one concentration above the threshold of 10% in the agonist assay or below the 80% in the antagonist assay. This situation occurred mostly at the highest tested concentration (C8). A modified classification was considered that would not include "I" given that such conclusion would not be preferable for regulatory testing. In addition, a closer dose range testing at the highest test range (C8 to C4) was considered. Therefore, instructions in the final version of the SOP were recommended for when a tested chemical displays in the pre-screen test (where dilution factor 10 is used) an activity at only the highest tested concentration passing the threshold of 10% or 80%. In such case, comprehensive testing would be performed with a dilution factor 2 instead of the default dilution factor 3.3. This would generate more data points at the right end of the dose response curve, i.e. for the highest tested concentrations.
- The criterion R^2 was found to have some limitations (as discussed in section 10.2). In order to improve the application of the specificity control, a normalisation of the specificity control values was carried out, with an evaluation of the tested concentration points above or below the threshold of 80%. If a chemical would display activity at all tested concentrations > 80%, it would be classified as "P"; if the activity at the highest tested concentration (C8) would be < 80%, the R^2 criterion would be applied. This would lead to an additional modification of the classifier.
- The maximal highest concentration to be tested was 50 mg/ml which would correspond to concentrations in the range of 39.8 mM to 769 mM. In view of the TG to be drafted, it was decided to recommend 100 mM as maximal test concentration. This concentration was also used for the ER-CALUX® method (TG 455). It was verified that lowering the highest concentration(s) tested to 100 mM would not affect the "P" classifications. At least two consecutive concentrations above or below the threshold values (as indicated by the classifier) remained below the 100 mM. For few test chemicals the maximal concentration of 50 mg/ml corresponded to values below 100 mM (39.8 mM to 94.1 mM). The increase to 100 mM as start concentration would not affect the "N" classification because the 50 mg/ml was found to be already insoluble and/or cytotoxic. The change to the 100 mM as maximal concentration would eliminate for some chemicals tested in the validation study the highest tested concentrations, leading to "I" becoming "N".

The revised classifier, as proposed by the VMG, is as follows:

Agonism: For each run, a test item is considered
A. **Positive** when the relative induction (Y_c) of the test item is ≥ 10% (REF RPC_{10}) for two or more consecutive concentrations.
B. **Negative** in all other cases

Antagonism: For each run, a test item is considered
A. **Positive** (competitive antagonist) when the relative induction (Y_c) of the test item is ≤ 80% (REF RPC_{80}) for two or more consecutive concentrations and

Either

- the relative induction of the test items normalised specificity control s_c^n > 80% at all concentrations

or when the following two conditions are met:

- the relative induction of the test items normalised specificity control at the highest concentration $s_{c_8}^n$ is ≤ 80%,
- the square of the correlation coefficient (R^2) is ≤ 0.9 between the relative induction of the test item (Y_c) and its specificity control (S_c)

B. **Negative** in all other cases

Explanation in the text under the classifier:

Negative classification for antagonism would include that the relative induction (Y_c) of the test item is > 80% (REF RPC80) at all concentrations or only one concentration is < 80%. Negative classification would also be applied when the following 2 conditions are met:

- the relative induction (Y_c) of the test item is ≤ 80% (REF RPC_{80}) for at least 2 consecutive concentrations and the relative induction of the test item's specificity control (S_c) at the highest tested concentration $s^n_{c_8}$ is ≤ 80%,
- the square of the correlation coefficient (R^2) is > 0.9 between the relative induction of the test item (Y_c) and its specificity control (S_c) (false positive)

WLR assessment with application of the modified SOP/classifier

The outcome of applying expert judgement (EJ) and applying the modified SOP instruction (dilution factor 2 for chemicals displaying in the pre-screen only the highest tested concentration passing the classification threshold value) as well as the modified classifier is shown in Tables 27 and 28 alongside the reported classifications by the lab (as discussed in the previous sections).

Given that the data reported in the validation study are the result of using the default dilution factor 3.3 when a comprehensive test was carried out, an approximation of what could be the second highest tested concentration with the dilution factor 2, was calculated. Hereto, the intermediate concentration between the 2 highest concentrations was calculated by interpolation. This concentration point cannot be considered as reflecting the real situation but gives nevertheless a good view on the possible outcome. It is envisioned that the application of the modified instructions and the modified classifier will lead to an objective conclusion of "P" and "N".

Detailed info per test chemical where this modified approach has been applied is provided below. In the statistical report, the interpolated values for the second highest non cytotoxic concentration by applying a dilution factor 2 can be found (see tables per test chemical stating interp value (x,y) and graphs showing a green square sign).

The assessment by applying the modified SOP and the classifier, resulted in higher levels of WLR and BLR.

Detailed information for re-assessment of tested chemicals in the AGONIST assay:

WLR assessment after expert judgement and application of the modified instruction in the SOP / classifier would be **95%, 95%, 100% and 100%.**

Test chemical 3, *Spironolactone*, led to N, I, N. VMG expert judgement changed the "I" to "N" due to the borderline response of the second highest tested concentration of 10.3 % RI.

Test chemical 8, *Tamoxifen*, as discussed before, yielded in one lab a run that shall be disqualified. This run would have been best repeated due to much variability within the triplicate samples.

Test chemical 17, *17β-Estradiol*, led to P, P, I in one lab. "I" is due to the highest tested concentration being scored as cytotoxic. With the modified SOP/classifier it could become "N" but the RI of the approximated intermediate concentration between C8 and C7 (equalling the concentration point resulting from DF 2) has a borderline value of 9.7 %).

Test chemical 20, *Hydroxyflutamide*, led to N, I, N in one lab. With the modified SOP/classifier, it would become "N" (approximated intermediate concentration with RI of 9.5%). In addition, the highest tested concentration is above the 100 µM (171 µM). With 100 µM as maximal concentration, the classification would be immediately "N".

Detailed information for re-assessment of tested chemicals in the ANTAGONIST assay:

WLR calculation after expert judgement and application of the modified instruction in the SOP/ classifier would be **90%, 95%, 95% and 100%.** If 100 µM would be the maximal concentration to be used, the WLR would increases to 100% for 3 labs and 90% for one lab.

Test chemical 1, *17β-Trenbolone*, consistently scored as "I" at ENVIGO, would become with the modified SOP/classifier **N, FP, FP** (approximate intermediate concentration with RI of 97.7% in 1st run, 53.1%, 60.8% in 2nd and 3rd run; R^2 of 0.94in 2nd run and 0.99 3rd run). At BDS, the one "I" could become "P" (approximated intermediate concentration with RI of 78.8%) or maybe N(FP) which cannot be assessed due to lack of the specificity control. The chemical was tested at a maximal concentration of 50µg/ml or 184 µM. With 100 µM as maximal test concentration, the classification would be immediately "N".

Test chemical 4, *Medroxyprogesterone acetate*, was scored as N, P, P for the results of RISE. The 3nd run yielded a "P" conclusion due to 2 consecutive concentration points just below 80% (79.2% and 79.8%) which are the 3rd and 4th highest tested concentrations. VMG expert judgement changes the "P" to "**N**" given the borderline response which occurs in the middle of the response. The 2nd run yielded a RI between 70.2 and 77.3% for the 4 highest tested concentrations. The response however starts with C1 at 81% (instead of 100%) what renders the analysis difficult. This

chemical was therefore not classified. The result would be **N, +, N**. At BDS, the chemical was scored once as I. VMG expert judgement is "**N**" due to the high variability in all the triplicate data points and trend of the response is "**N"**.

Test chemical 8, *Tamoxifen*, was scored as "I" by 2 labs due to RI % just below 80%. With the modified SOP/classifier it would lead to "**N**" in both cases (approximation of intermediate concentration with RI of 84.6% and 83.3 %).

Test chemical 9, *Atrazine*, was scored as "I" by 2 labs. VMG expert judgement led to "**N**" for the 3rd run in CitoxLAB due to variability in the technical replicates. With the new SOP/classifier, the 2nd run for RISE would lead to "**N**" (approximated intermediate concentration RI of 84.1%)

Test chemical 15, *Prochloraz*, was scored as P, N(FP), P. The FP is due to $R^2 = 0.91$. It would not change with the new SOP/classifier (see further section 10.2).

Test chemical 18, *Benzylbutyl phthalate, was* scored once as "I" by RISE. The response has clearly a negative trend. This lab was conservative with its solubility observations leading to less high tested concentrations than the other labs. With the modified SOP/classifier, it would likely result in "**P**" (approximated intermediate concentration with RI of 68.2%)

Test chemical 19, *Propylthiouracil*, yielded a mix of classifications between all 4 laboratories due to its response that are border line for the highest tested concentration. With the modified SOP/classifier, the "I" results could become all "**N**" (approximated intermediate concentration with RI of 83.6% and 83.4% for CitoxLAB; 87.7% and 83% for RISE; 80.5% for BDS) except for one run at ENVIGO where it could be "**P**" (RI 74.1%). This "P" could possibly also be N(FP) if results of a specificity control test would be available. With 100 µM as maximal concentration, the classification would be immediately "N" for all labs given that concentrations up to 293 µM were tested. For BDS, the highest tested concentration of 881 µM was excluded given that such high concentration was not required to be tested.

Table 27A and B: Within Laboratory Reproducibility of 20 coded test items

		(A) AGONIST- reported by the laboratories												(B) AGONIST – classifications with modified SOP/classifier and EJ (in red)											
	TEST CHEMICAL	RISE			ENVIGO			BDS			CitoxLAB				RISE			ENVIGO			BDS			CitoxLAB		
1	17β-Trenbolone	P	P	P	P	P	P	P	P	P	P	P	P		P	P	P	P	P	P	P	P	P	P	P	P
2	Stanozolol	P	P	P	P	P	P	P	P	P	P	P	P		P	P	P	P	P	P	P	P	P	P	P	P
3	Spironolactone(#)	N	N	N	N	I (1)	N	N	N	N	N	N	N		N	N	N	N	N EJ	N	N	N	N	N	N	N	N
4	Medroxyprogesterone acetate	P	P	P	P	P	P	P	P	P	P	P	P		P	P	P	P	P	P	P	P	P	P	P	P	P
5	Bisphenol A	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
6	Bicalutamide	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
7	Disulfiram	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
8	Tamoxifen	N	N	N	**(2)	N	N	N	N	N	N	N	N		N	N	N	**	N	N	N	N	N	N	N	N	N
9	Atrazine	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
10	17α-Ethynyl estradiol	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
11	Sodium azide	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
12	Diethylhexyl phthalate	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
13	Methyldihydrotestosterone	P	P	P	P	P	P	P	P	P	P	P	P		P	P	P	P	P	P	P	P	P	P	P	P	P
14	Vinclozolin	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
15	Prochloraz	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
16	Fluoxymesterone	P	P	P	P	P	P	P	P	P	P	P	P		P	P	P	P	P	P	P	P	P	P	P	P	P
17	17β-Estradiol	P	P	I (3)	P	P	P	P	P	P	P	P	P		P	P	N$	P	P	P	P	P	P	P	P	P	P
18	Benzylbutyl phthalate	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
19	Propylthiouracil	N	N	N	N	N	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
20	Hydroxyflutamide	N	N	N	N	I (4)	N	N	N	N	N	N	N		N	N	N	N	N	N	N	N	N	N	N	N	N
	WLR	**95%**			**89%**			**100%**			**100%**				**95%**			**100%**			**100%**			**100%**		
	Concordant/total classifications	19/20			17/19			20/20			20/20				19/20			19/19			20/20			20/20		

P = Positive response, N= Negative response, I=Inconclusive, as defined by the application of the classifier in SOP V06. Green left upper corner section: classification by EURL ECVAM according to classifier in SOP V06.
(#): was tested twice by RISE under two different codes but with same outcome
(**): disqualified
(red colour): Application of new classifier for the "I" with approximation of the intermediate additional concentration point between C8 and C7
($): Application of new classifier but border line value for the approximation of the intermediate concentration (9.7%) when DF 2 would have been applied
EJ stands for VMG expert judgement.
Grey highlight: Tested at higher concentration than 100 µM.

Summary notes:

(1) *Spironolactone*: Border case, the second highest tested concentration has a RI of 10.3%. EJ led to "N" due to borderline value and shape of the curve.

(2) *Tamoxifen*: Data points with lots of variability between the technical replicates, leading to 3 concentration points > 10% threshold. Disqualified run.

(3) *17 β -Estradiol*: Though 3 runs started with the same highest concentration, in only one run this concentration was scored as cytotoxic leading to only one concentration point > 10%.

(4) *Hydroxyflutamide*: Border case, the highest tested concentration has a RI of 11.7%. (Note: highest tested concentration is 171 µM. With 100 µM as maximal concentration, classification is immediately N)

Table 28A and B: Within Laboratory Reproducibility																									
(A) ANTAGONIST- reported by the laboratories														(B) ANTAGONIST –classifications with modified SOP/classifier and EJ (in red)											
TEST CHEMICAL		RISE			ENVIGO			BDS			CitoxLAB			RISE			ENVIGO			BDS			CitoxLAB		
1	17β-Trenbolone	N	N	N	I(1)	I	I	I(2)*	N	N	N	N	N	N	N	N	N	N (FP)	N (FP)	$ P/FP	N	N	N	N	N
2	Stanozolol	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
3	Spironolactone(#)	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
4	Medroxyprogesterone acetate	N	P(3)	P(3)	I (4)	N	N	N	N	N	N	N	N	N	+	N EJ	N EJ	N	N	N	N	N	N	N	N
5	Bisphenol A	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
6	Bicalutamide	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
7	Disulfiram	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
8	Tamoxifen	N	N	N	I(5)	N	N	N	I (6)*	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
9	Atrazine	I (7)	N	N	N	N	N	N	N	N	N	N	I (8)	N	N	N	N	N	N	N	N	N	N	N	N EJ
10	17α-Ethynyl estradiol	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
11	Sodium azide	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
12	Diethylhexyl phthalate	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
13	Methyldihydrotestosterone	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
14	Vinclozolin	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
15	Prochloraz	P	N(FP) (9)	P	P	P	P	P	P	P	P	P	P	P	N(FP)	P	P	P	P	P	P	P	P	P	P
16	Fluoxymesterone	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
17	17β-Estradiol	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
18	Benzylbutyl phthalate	P	P	I (10)	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
19	Propylthiouracil	I (11)	I	N	I (12)*	N	I*	I (13)*	P	I*	I (14)	N	I	N	N	N	$ P/FP	N	N	N	N	N	N	N	N
20	Hydroxyflutamide	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
	WLR	**75%**			**80%**			**85%**			**90%**			**94.7%**			**100%**			**100%**			**100%**		
	Concordant/total classifications	15/20			16/20			17/20			18/20			18/19			19/19			19/19			20/20		

P = Positive response, N= negative response, I=Inconclusive, N(FP) = Negative classification due to false positive (FP) outcome by $R^2 > 0.9$, according to classifier SOPV06
Green left upper corner section: classification by EURL ECVAM according to classifier SOP V06.
(#): was tested twice by RISE under two different codes and with the same outcome
* Was reported by the lab as "N" due to lab not applying the Inconclusive option
+ Excluded from classification due to data difficult to analyse
(red colour): Application of new classifier for the "I" with approximation of the intermediate additional concentration point between C8 and C7
($): Application of new classifier but FP call cannot be made due to lack of specificity control. Excluded from classification.
Grey highlight: Tested at higher concentration than 100 µM.

Summary notes:
17β-Trenbolone: (1) Highest tested concentration with a RI of 7 - 17% for 3 runs. This concentration point was not considered in 2 other laboratories due to cytotoxic scoring except by BDS. (2) Run 1 with one higher tested concentration with a RI 28.7% (this concentration point was scored as cytotoxic in the other 2 runs). (Note: highest tested concentration is 171 µM. With 100 µM as maximal concentration, classification is immediately "N").

Medroxyprogesterone acetate: (3) Three runs with borderline RI values. 2nd run would have best been repeated due to C1 at 81%. Results are difficult to interpret. 3rd run with EJ is "N" (dose response shape and variability triplicate samples). (4) Border case due to 3rd highest concentration with a RI of 77.5%. EJ is "N" (mid dose response plus variability in the triplicate samples).

Tamoxifen: (5) Border case of highest tested concentration with a RI of 77.2%. (6) Border case of highest tested concentration with a RI of 76.7%.

Atrazine: (7) Border case of highest tested concentration with a RI of 77.6%. (8) Border case of highest tested concentration with a RI of 77.9% EJ is "N" due to borderline value and variability data points

Prochloraz: (9) FP due to R2.

Benzylbutyl phthalate: (10) One run is "I" though there is a clear decline of the response.

Propylthiouracil: (11) Border case in 2 runs of highest tested concentration with a RI of 78.8% and 76.4 %. (12) Border case in 2 runs of highest tested concentration with a RI of 66.3% and 71.6%. (13) Border case of 1 or 2 highest tested concentrations with a RI of 70.2% and 63.7%. (14) Border case in 2 runs of highest tested concentration with a RI of 69.1% and 77.5%. (Note: highest tested concentration is 293 µM for 2 labs, 881 µM for one lab. With 100 µM as maximal concentration, classification is immediately "N" for all labs)

BLR assessment with application of the modified SOP/classifier

For the agonist testing, the BLR (already 100%) would not change by the re-classification.

For the antagonist testing, 5 classifications would be influenced by the re-classification, as shown below: test chemical 1 for one lab and test chemical 19 for four labs. The BLR would increase to 100%.

Table 29: Between laboratory reproducibility with concordance of classifications (majority rule)

TEST CHEMICAL		AGONIST				ANTAGONIST			
		RISE	ENVIGO	BDS	Citox LAB	RISE	ENVIGO	BDS	Citox LAB
1	17β-Trenbolone	P	P	P	P	N	**N(FP)**	N	N
2	Stanozolol	P	P	P	P	N	N	N	N
3	Spironolactone	N	N	N	N	P	P	P	P
4	Medroxyprogesterone acetate	P	P	P	P	N	N	N	N
5	Bisphenol A	N	N	N	N	P	P	P	P
6	Bicalutamide	N	N	N	N	P	P	P	P
7	Disulfiram	N	N	N	N	N	N	N	N
8	Tamoxifen	N	N	N	N	N	N	N	N
9	Atrazine	N	N	N	N	N	N	N	N
10	17α-Ethynyl estradiol	N	N	N	N	P	P	P	P
11	Sodium azide	N	N	N	N	N	N	N	N
12	Diethylhexyl phthalate	N	N	N	N	N	N	N	N
13	Methyldihydrotestosterone	P	P	P	P	N	N	N	N
14	Vinclozolin	N	N	N	N	P	P	P	P
15	Prochloraz	N	N	N	N	P	P	P	P
16	Fluoxymesterone	P	P	P	P	N	N	N	N
17	17β-Estradiol	P	P	P	P	P	P	P	P
18	Benzylbutyl phthalate	N	N	N	N	P	P	P	P
19	Propylthiouracil	N	N	N	N	**N**	**N**	**N**	**N**
20	Hydroxyflutamide	N	N	N	N	P	P	P	P
	WLR	**100%**				**100%**			
	Concordant/total classifications	20/20				20/20			

Summary notes:
For chemical 1: ENVIGO's WLR was previously 3 times "I". With the modified SOP/classifier, a "N" and two N(FP) are obtained, resulting in a mean of N(FP).

For chemical 19: All labs had previously WLR of "I". With the modified SOP/classifier it becomes "N" in all labs.

10.1.2 The Specificity Control and Criterion R^2

For several chemicals, the specificity control was helpful in the designation of a classification as competitive antagonist on the basis of the criterion R^2.

For 3 tested chemicals the specificity control led to conclude on a false positive (FP) for only one of the 3 runs, see Table 30 below: *Prochloraz, 2-sec Butylphenol* and *Fulvestrant*.

For 2 chemicals FP was scored in all 3 runs: *Cycloheximide* and *Ketoconazole*.

Norethinodrone acetate (34) is also discussed in this section due to its interesting dose response(s).

Table 30: Classification of false positive antagonist

No.	TEST CHEMICAL	ANTAGONIST Classification under SOPV06			R^2			Modified SOP/classifier with S_c^n > or < 80%		
15	Prochloraz	P	N(FP)	P	0.62	0.91	0.8	P	N(FP)	P
39	2-sec-Butylphenol	N(FP)	P	P	0.95	0.81	0.34	P	P	P
43	Fulvestrant	N(FP)	P	I	0.97	0.9	0.7	N(FP)	P[1]	N(FP)[2]
44	Cycloheximide	N(FP)	N(FP)	N(FP)	0.95	0.99	0.99	N(FP)	N(FP)	N(FP)
30	Actinomycin D	I	I	I	0.93	0.98	0.98	N	N(FP)	N(FP)
41	Ketoconazole	N(FP)	N(FP)	N(FP)	0.97	0.93	0.96	N(FP)	N(FP)	N(FP)
34	Norethindrone acetate	I	P	I	0.85	0.73	0.62	P but questionable		

[1] Due to borderline value of R^2, the outcome could be P or N(FP)
[2] With removal of the outlier, R^2 would be 0.98 leading to N(FP)

From the data analysis, it was observed that the value of R^2 can be influenced by 1) an outlier in the dose response with the lower concentration of DHT (the standard response) (chemical 43); 2) a dose response of the specificity control that is slightly declining but not more than 20% (chemical 39). Moreover, it was noted that for a chemical with a shift of the dose response from the specificity control in the opposite direction, the R^2 is < 0.9 (chemical 34).

For several tested chemicals, it was observed that the lowest tested concentration of the specificity control dose response C1 did not start around the 200 % RI. To ease visual inspection of the dose response shift and also to verify the decline the specificity control dose response, the VMG agreed to rescale (normalisation) this dose response to the 100% (see figure 03).

Prochloraz (no. 15) with $R^2 = 0.91$. This chemical was the only chemical tested by all 4 labs but only one lab obtained one run out of 3 with a FP outcome. It deserves to be noted that the R^2 has a borderline value (0.91).

2-sec-Butylphenol (no. 39) was classified as FP for one out of 3 runs with a R^2 clearly above the threshold value of 0.9. This is due to the shape of the specificity control curve which is slightly declining with increasing concentrations. This outcome shows that the application of the R^2 criterion has some limitations. Normalisation of the specificity control values (rescaling to 100%) and visual evaluation of the dose response shift show it is "P" (see figure 03). All concentrations are above the 20% threshold. On the basis of this response, the VMG agreed to modify the classifier with inclusion of the normalised specificity control (S_c^n) (see Discussion section 10.1).

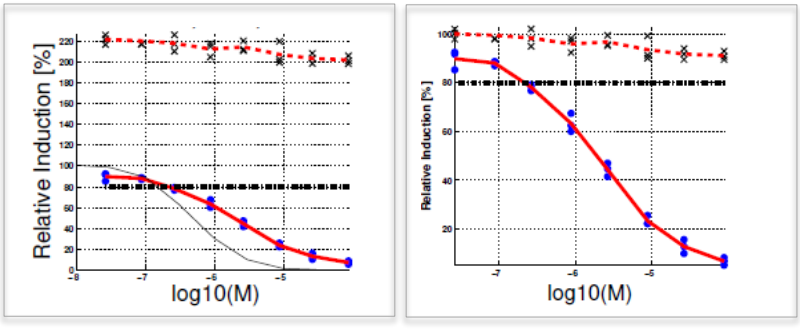


Figure 03: Test chemical 39. Display of the dose responses when tested with the lower (standard response, red solid curve) and the higher (specificity control response, red dashed curve) spiking of DHT. Black curve indicates the reference Flutamide response. Figure at the right with re-scalement of the specificity control dose response (S_c^n)

Fulvestrant (no. 43) was classified as FP, P, I though the dose responses are quite comparable. The 1st run has 2 concentration points below 80% and R^2 of 0.97. The 2nd run has a borderline value of 0.9 for the R^2. The 3rd run had only one concentration point below 80% (RI of 42.5%) though the second concentration seems to have yielded an outlier (see figure 04). R^2 is 0.76 and will be impacted by this outlier. VMG expert judgement led to interpolate the intermediate concentration between the highest and the third highest concentration (exclusion of the outlier). Such would lead to a R^2 of 0. 98 and change the "I" to N(FP). Application of the modified SOP/classifier and EJ to all 3 runs would lead to **N(FP), P, N(FP).** Given the borderline value of 0.9 for R^2 in the 2nd run, it may be a **N(FP)**.

This chemical is used in prostate cancer treatment and is thought to have its action by downregulating the expression of the androgen receptor (Bhattacharyya *et al*, 2019). This could possibly explain the observed result of FP. With increasing concentrations of Fulvestrant, AR expression would decrease, and less AR available for binding with the ligand DHT.

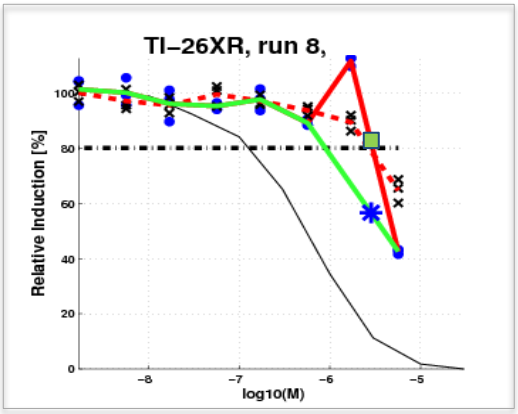


Figure 04: Test chemical 43 (run 8). Display of the dose responses when tested with the lower (standard response, red solid curve) and the higher (specificity control response, red dashed curve) spiking with DHT. Black curve indicates the reference Flutamide response. Green curve indicates possible dose response when C7 (outlier) is excluded. Blue asterisk indicates the interpolated value between the highest (C8) and the 3rd highest concentration (C6) of what could be the result by applying dilution factor 2. Green square shows the interpolated value between C8 and C7.

Actinomycin D *(no. 30)* and *Cycloheximide (no. 44)* are both antibiotics. *Cycloheximide* yielded 2 to 3 concentrations below the 80% in the comprehensive test for the 3 runs and R^2 of 0.95, 0.99, 0.99, leading to **FP**. For *Actinomycin* D only one concentration was found to be below 80% in 3 runs, leading to "I". The modified classifier/SOP would instruct to perform dilution 2 after the pre-screen. This may lead to 2 concentrations below 80% for possibly 3 runs (approximated intermediate concentration of 64% and 65.3% and 80%) but the specificity control would still conclude to **FP** for each run ($R^2 > 0.9$). The false positive behaviour of both tested chemicals may be due to the cytotoxicity that indeed was observed at higher concentrations (with the LDH test and with visual inspection) but could not be observed at the lower tested concentrations. Early events that lead to cell death are not picked up by the LDH test nor by visual inspection.

Ketoconazole (no. 41) is reported to be a fungicide but also as a pharmaceutical (prostate cancer treating drug). It displayed an antagonist activity response in 3 runs, likely due to cytotoxicity. The specificity control results in **FP (**$R^2 > 0.9$).

Norethinodrone acetate (no. 34) displays 1 concentration point below the 80% threshold value in 2 of the 3 runs (RI of 54.2% and 56.2%) (representative example, see figure 05). With the new SOP, a dilution factor of 2 would have been taken, leading to 2 concentrations in both runs below 80% and the S_c^n of C8 below 80%, with R^2 is < 0.9, resulting in 2 times "**P**" (approximated intermediate concentrations with RI of 73.1% and 79.1%). R^2 values are below 0.9 in all 3 runs indicating a competitive antagonist. This result however is questionable. The shift of the dose response curve of the specificity control is to the opposite direction (left shift) because it declines before the standard assay response. This situation reveals that the application of the criterion R^2 has some limitations and expert judgement shall be made. Chemicals displaying this type of response have been described as well for the Tox21 lux assay. 65 chemicals were found to have a potency shift in the opposite direction. However, the chemical of this study was not tested with the Tox21 assay (Kleinstreuer *et a*l, 2017).

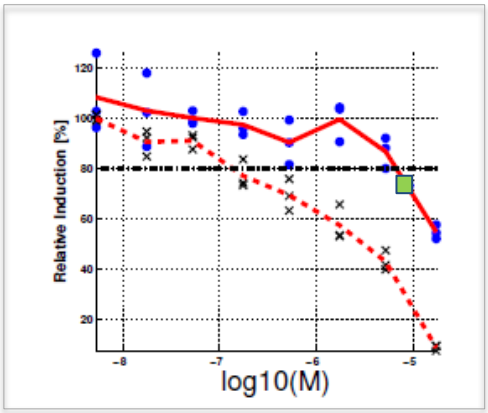


Figure 05: Test chemical 34. Display of the dose responses when spiked with DHT: the lower (standard response, red solid curve) and the higher concentration (specificity control response, red dashed curve).

10.1.3 Classifications of the additional 26 tested chemicals

Shown below (Table 31) are the classifications as reported by the laboratory for agonist and antagonist testing (as discussed in section 6). For the agonist testing no "I" results were found and WLR is 100%. For the antagonist testing, the plausible outcome of the application of the modified instructions in the SOP / classifier is also shown in Table 31, alongside the reported classifications: WLR increases from 84.4% to 96.1%.

Table 31: Classifications

No.	TEST CHEMICAL	AGONIST Reported by the labs			ANTAGONIST Reported by the labs			ANTAGONIST Modified SOP/classifier		
21	Levonorgestrel	P	P	P	N	N	N	N	N	N
22	Cyproterone acetate	N	N	N	P	P	P	P	P	P
23	2-tert-Butylanthraquinone	N	N	N	P	P	P	P	P	P
24	Arochlor1254	N	N	N	P	P	P	P	P	P
25	Nandrolone	P	P	P	N	N	N	N	N	N
26	o,p'-DDT	N	N	N	P	P	P	P	P	P
27	Phenolphthalin	N	N	N	N	N	N	N	N	N
28	2,4,5-T	N	N	N	N	N	N	N	N	N
29	Methyltrienolone (R1881)	P	P	P	N	N	N	N	N	N
30	Actinomycin D	N	N	N	I (1)	I	I	N	N(FP)	N(FP)
31	Diethylstilbestrol	N	N	N	P	P	P	P	P	P
32	L-Thyroxine	N	N	N	N	N	N	N	N	N
33	Haloperidol	N	N	N	N	N	N	N	N	N
34	Norethinodrone acetate	P	P	P	I(2)	P	I	P?&	P?&	P?&
35	Pimozide	N	N	N	N	N	N	N	N	N
36	Progesterone	N	N	N	P	P	P	P	P	P
37	Linuron	N	N	N	P	P	P	P	P	P
38	Methyltestosterone	P	P	P	N	N	N	N	N	N
39	2-sec-Butylphenol	N	N	N	N(FP)(3)	P	P	P	P	P
40	Corticosterone	N	N	N	P	P	P	P	P	P
41	Ketoconazole	N	N	N	N(FP)	N(FP)	N(FP)	N(FP)	N(FP)	N(FP)
42	Finasteride	N	N	N	P	P	P	P	P	P
43	Fulvestrant	N	N	N	N(FP)	P	I(4)	N(FP)	P	N(FP)
44	Cycloheximide	N	N	N	N(FP)	N(FP)	N(FP)	N(FP)	N(FP)	N(FP)
45	Norethindrone	P	P	P	N	N	N	N	N	N
46	Mifepristone	N	N	N	P	P	P	P	P	P
	WLR	**100%**			**84.4%**			**96.1%**		
	Concordant/total classifications	26/26			22/26			25/26		

Summary notes for the antagonist classifications (reported by the labs):
(1) Chemical 30: only the highest tested concentration has RI < 80% in 3 runs (57.2%, 39%, 37.5%). R^2 is > 0.9 in all 3 runs what indicates FP but with the classifier it gets classified as "I" due to only one concentration below the threshold value.
(2) Chemical 34 has dose responses with the specificity control that declines before the standard response. "I" is scored in 2 runs due to only the highest tested concentration with RI of < 80%: 54.2% and 56.2%. &: Questionable result. Shift of the specificity control dose response curve in the opposite direction. Criterion $R^2 > 0.9$ would not be applicable in this case.
(3) Chemical 39 was reported as "P" but shall be FP due to $R^2 = 0.91$
(4) Chemical 43: second highest tested concentration has RI of 111.5% which is likely an outlier.
Grey highlight: Tested at higher concentration than 100 µM.

The tested chemicals with "I" or N(FP) result are discussed in detail in the previous section 10.2. A brief summary per chemical can be found below.

Test chemical 30, Actinomycin D: classified as "I" under the classifier of SOP V06 but with the modified SOP/classifier the testing could result in 2 concentrations under the threshold of 80% for 2 of the 3 runs, leading to "P" (approximated intermediate concentrations with RI of 64 % and 65%). The specificity control result however would lead to FP due to $R^2 > 0.9$ what would result in **N(FP)**.

Test chemical 34, Norethinodrone acetate, is a positive agonist that displays an interesting dose response in the antagonist assay when tested with excess of DHT (specificity control). The specificity control response declines before the standard response leading to shift in the opposite direction. Applying the classification would lead to "P" but this is rather questionable given the left shift of the dose response.

Test chemical 39, 2-sec-Butylphenol, has one N(FP) classification due to $R^2 = 0.95$. By applying the new classifier (with the evaluation of normalised values of the specificity control) this would result in a **P**.

Test chemical 43, Fulvestrant, showed 3 different classifications though the dose responses are quite comparable. VMG expert judgement and application of the modified SOP/classifier would lead to **N(FP), P, P or FP**.

10.2 Quantitative assessment of the results

Measurements of the EC_{50} and PC_{10} (agonism) and IC_{50} and PC_{80} values (antagonism) were very reproducible within the laboratories and between the laboratories (for overview tables of the values see Statistical report).

Based on all the log(EC_{50}), log(PC_{10}) (agonism), and, log(IC_{50}) and log(PC_{80}) (antagonism) obtained by each of the 4 laboratories, the average % CV's per laboratory was less than 2.5%. These values are comparable to what is reported for the ER-CALUX® method (less than 4%) which was adopted as part of TG 455, and, comparable to what is reported for the Japan ARTA (less than 2%) which was adopted as TG 458.

Table 32: Range of average % CV of all tested chemicals for all labs from different validation studies

Validation Study	AGONIST testing	ANTAGONIST testing
AR-CALUX®	Log(EC_{50}): 0.65% to 2.16% (5 test chemicals)	Log(IC_{50}): 0.95% to2.09% (9 test chemicals)
ER-CALUX®	Log(EC_{50}): 1.2% to 3.1% (17 test chemicals)	Log(IC_{50}): 0.5% to 1.6% (4 test chemicals)
Japan ARTA	Log(PC_{50}): 0.38% to 1.53% (3 test chemicals)	Log(IC_{50}): 0.84% to 1.15% (3 test chemicals)

Values taken from validation study reports: % CV taken from calculations expressed in molar.

10.3 Additional observations from the validation study

10.3.1 Usage of different luminescence kits and substrates and luminometers

Amongst the 4 laboratories, 2 have used in-house prepared illuminate mix, one laboratory opted for the Promega luciferase kit while another laboratory chose the Promega Steady Glo mix kit. The last kit did not require a luminometer with double injection.

Comparison of the potency measures (EC_{50}, IC_{50}, etc.) across the different laboratories did not show substantial differences (see variability measures in the Statistical report, page 38-39).

10.3.2 Usage of frozen stock solutions for the reference chemicals

The two reference chemicals DHT (for the agonist testing) and Flutamide (for the antagonist testing) are used for each run. Preparing these chemicals always fresh implies a significant work load. When the laboratory BDS joined the study, the VMG agreed that the two reference chemicals could be prepared up front and aliquots stored at - 20ºC. The test method developer reported on stability of these reference chemicals for a period up to 3 months. Once an aliquot is thawed, it can be stored at - 4ºC and used for up to 4 weeks.

Comparison of the potency measures (EC_{50}, IC_{50}, etc.) of DHT and Flutamide did not show a significant difference (see variability measures in the Statistical report, page 38-39).

10.3.3 Usage of plastic plates and glass tubes

The labs BDS and CitoxLAB used 24 well plastic plates for preparing the work solutions of the test chemicals while the other two labs worked with glass tubes.

Comparison of the potency measures (EC_{50}, IC_{50}, etc.) obtained by BDS and CitoxLAB did not show a significant difference with the other 2 laboratories (see variability measures in the Statistical report, page 38-39).

10.3.4 Invalid runs and reasons for rejection

Table 33 shows a summary of the reasons for invalidity of runs as reported by the laboratories. Most invalid runs occurred in the antagonist assay. The 2 most reported criteria, leading to failure, were having a too low Inhibition factor and too low Plate Z factor.

Table 33: Invalid runs

AGONIST		ANTAGONIST	
Acceptance criterion	**# Reported failures**	**Acceptance criterion**	**# Reported failures**
1 - Sigmoidal shape	0	1 - Sigmoidal shape	0
2 - DHT EC_{50} out of range	1	2 - FLU IC_{50} out of range	5
3 - DHT CV ($LogEC_{50}$) ≥ 1.5%	1	3 - FLU CV ($LogEC_{50}$) ≥ 3%	0
4 - RI PC ≤ 30%	0	4 - RI PC ≥ 60%	6
5 - RI NC ≥ 10%	3	5 - RI NC ≤ 85%	8
6 - Induction factor ≤ 20	0	6 - Inhibition factor < 10	12
7 - Plate Z-factor < 0.5	3	7 - Plate Z-Factor < 0.5	23
		8 - FLU R^2 > 0.7	3
Other reported issues	**# Reported**	**Other reported issues**	**# Reported**
DHT C1 > 10%	1	RI VC > 5%	4
Test Item CV ($LogEC_{50}$) > 3%	9	FLU C1 > 120% or < 80%	2
Variable RLU values	1	Test item C1 > 120% or < 80%	2
Test item EC_{50} inaccurate	1	Variable RLU values	2
Test item too cytotoxic	2		

11 VMG OVERALL CONCLUSIONS AND RECOMMENDATIONS

The primary aim of this validation study was to assess the transferability, within laboratory (WLR) and between laboratory reproducibility (BLR) of the AR-CALUX® method with a number of relevant coded chemicals that were judged by the VMG to be suitable and sufficiently challenging to permit conclusions to be drawn.

The VMG concludes that the test method can be transferred among properly equipped and staffed laboratories, including those having no prior experience in similar test methods. Experienced personnel can readily be trained for the test method and the necessary equipment and supplies can be readily obtained. Caution is however needed in handling chemicals that may be potent endocrine disruptors and the material used during the testing (e.g. glass tubes or plastic tubes) which should be free of contamination. The VMG recommends that such caution shall be included in the SOP.

The SOP is considered robust, the acceptance criteria values adequate. It allows flexibility in the usage of type of luminescence substrates or kits, and, type of luminometers. It is clearly written and the testing and analysis of results can be easily performed.

The WLR based on concordance of classifications within the laboratories is **89%, 95%, 100%** and **100%** for agonist testing, **75%, 80%, 85%** and **90%** for antagonist testing. Overall BLR is **100%** for agonist testing and **87.5%** for antagonist testing. These values are influenced by the occurrence of the classification "I", especially in the antagonist testing.

In order for an end user to arrive immediately to a classification of "N" and "P", without the option of "I", further instructions in the SOP are suggested (as detailed in the previous section). Applying these instructions in the SOP to the validation study data, by approximating the second highest concentration point that would result from applying a dilution factor 2, the reported classifications could be improved. This would lead to WLR values of **95%, 100%, 100%** and **100%** for agonist testing, and, **94.7%, 100%, 100%** and **100%** for antagonist testing; BLR of **100%** for both agonist and antagonist testing.

The application of a specificity control in the antagonist assay, to identify competitive antagonists, has proven to be useful, both for gaining more confidence that the positive classification is indeed correct, and, for defining false positives. The criterion R^2 is a good measure but some caution is advisable as it was shown not be 100% reliable. A further amendment of the classifier, allowing normalisation of the specificity control values with evaluation of the values being above or below the threshold value of 80%, would improve the classification. Expert judgement nevertheless will be required.

Comparison of the classifications of 23 tested chemicals shows a very good concordance with the classifications reported in the AR-reference list.

Amendments and deviations to the studies, reported by the laboratories, are minor and have no impact on the results. For example, one laboratory substituted during the transfer phase the recommended luciferase kit by another, leading to data of similar quality to the other laboratories.

This validation study was conducted with a gravimetric method using 50 mg/ml as the highest stock concentration. In view of the drafting of a TG, the VMG recommends to use 100 mM as the highest stock concentration. When the molecular weight of a test chemical cannot be calculated such as for multi constituent substances, polymers, mixtures, UVCBs etc., the gravimetric method should be used starting from 50 mg/ml.

Overall, the VMG concludes that the information generated in this validation study shows that the AR-CALUX® method is a reliable test method that can contribute to the determination of (anti)-androgen potential of substances.

The recommendations by the VMG after the transfer phase were as follows, and were introduced in the following versions of the SOP used in the validation study:

- Broadening up the SOP for the usage of other luminometers and not only the double injector luminometer
- An additional guidance for the antagonist assay to monitor carefully the VC response and the SC response

The recommendations by the VMG after evaluation all data, and to be introduced in the final version of the SOP, are as summarised below:

- After the section of the pre-screen, to identify which dose range is appropriate and also which dilution factor to use for preparing the working solutions. Dilution factor 2 is recommended for chemicals displaying activity at only the highest tested concentration above the 10% threshold (agonist) or below the 80% threshold (antagonist).
- Highest stock concentration to be used for testing of chemicals: 100 mM.
- Modification of the classifier to arrive to a "P" or "N" classification.
- Normalisation of the specificity control values.

- A warning for the use of plastic/glass before running studies (the response to the vehicle control should be assessed to ensure no interference).

In addition, the VMG suggests that end users can choose to use glass tubes or plastic plates for the preparation of working solutions for the test chemicals, and, to prepare fresh or to freeze down stock solutions of the reference chemicals.

12 REFERENCES

Araki, N., Ohno, K.,Takeyoshi, M., Iida, M. Evaluation of rapid *in vitro* androgen receptor transcriptional activation assay using AR-Ecoscreen cells. Toxicology in Vitro (2005), 19, 335-352.

Bhattacharyya, R.S., Krishnan, A.V., Swami, S., Feldman, D. Fulvestrant (ICI 182,780) downregulates androgen receptor expression and dimishes androgenic responses in LNCaP human prostate cancer cells (2006). Mol Cancer Ther 5(6), 1539-1549.

Browne, P., Kleinstreuer, N.C, Ceger, P., Deisenroth, C., Baker, N., Markey, K., Thomas, R.R, Judson, R.J., Casey, W. Development of a curated Hershberger database (2018), Reproductive toxicology, 81, 259-271.

Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on "the protection of animals used for scientific purposes" OJ L 276, 20.10.2010

EC, 1999: Community strategy for endocrine disruptors, a range of substances suspected of interfering with the hormone systems of humans and wildlife (COM (1999) (706) EUR-Lex - 51999DC0706.

EC, 2017: Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council OJ L 301, 17.11.2017, p. 1–5.

EC, 2018: Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties OJ L 101, 20.4.2018, p. 33–36.

EFSA, 2018. Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. https://doi: 10.2903/j.efsa.2018.5311.

OECD (1999), OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 5. Compliance of Laboratory Suppliers with GLP Principles, OECD Publishing, Paris. Available at http://search.oecd.org/officialdocuments ENV/JM/MONO(99)21.

Freyberger, A., Witters, H., Weimer, H., Lofink, W., Berckmans, P., Ahr, H-J. Screening for (anti)androgenic properties using a standard operation protocol based on the human stably transfected androgen sensitive PALM cell line, first step towards validation (2012). Reproductive Toxicology 30, 9-17.

Hartung T1, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, Prieto P, Sabbioni E, Scott L, Worth A, Zuang V., A modular approach to the ECVAM principles on test validity, Altern Lab Anim. (2004), 32(5):467-72.

Hecker, M. and Holler, H. Endocrine disruptor screening: regulatory perspectives and needs. Env. Sciences Europe (2011), 23, 15.

Kleinstreuer, N.C, Ceger, P., Watt, E.D., Martin, M., Houck, K., Patience Browne, P., Russell S. Thomas, R.S, Casey, W.M, Dix, D.J., Allen, D., Sakamuru, S., Xia, M., Huang, R., Judson, R. Development and validation of a computational model for androgen receptor activity (2017). Chemical research in toxicology, 30, 946-964.

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

OECD (2016), Test No. 458: Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals, OECD Publishing, https://doi.org/10.1787/9789264264366-en.

OECD (2018), Revised Guidance Document No 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption, OECD Series on Testing and Assessment, No. 150, OECD Publishing, Paris, https://doi.org/10.1787/9789264304741-en.

Sonneveld, E., Jansen, H.J., Riteco, J.A., Brouwer, A., van der Burg, B. Development of androgen and estrogen responsive bioassays, members of a panel of human cell line based highly selective steroid responsive bioassays (2004). Tox. Sciences 83, 136-148.

Van der Burg, B., Pieterse, B., Buist, H., Lewin, G., van der Linden, S.C., Man, H.Y., Rorije, E., Piersma, A.H., Mangelsdorf, I., Wolterbeek, A.P., Kroese, E.D., van Vught-Lussemburg, B. (2015a). A high throughput screening system for predicting chemically-induced reproductive organ deformities. Reprod Toxicol 55, 95-103.

Van der Burg, B., Winter, R., Mana, H., Vangenechten, C., Berckmans, P., Weimer, M., Witters, H., van der Linden, S. Optimization and prevalidation of the *in vitro* AR CALUX® method to test androgenic and antiandrogenic activity of compounds(2010). Reproductive Toxicology 30, 18–24.

van Vugt-Lussenburg, B.M.A., van der Lee, R.B., Man, H.Y., Middelhof, I., Brouwer, A., Besselink, H., van der Burg, B. (2018) Incorporation of metabolic enzymes to improve predictivity of reporter gene assay results for estrogenic and anti-androgenic activity. Reprod Toxicol 75:40-48.

13 ANNEXES

13.1 Statistical report (see separate document, in TSAR)

13.2 Project plan version 05

13.3 List of reference, control and test chemicals and their properties

13.4 Chemical coding and distribution procedure

13.5 SOP final version V07 (see separate document, in TSAR)

13.6 Solubility data for all 46 tested chemicals

13.7 List of additional documents filed for the study and available on request

13.7.1 Training documents (Agenda, Planning of the training)

13.7.2 Study plans and study reports of the 4 participating laboratories

13.7.3 SOP versions V03, V04, V05 and V06, and solubility SOP

13.7.4 JRC technical report on "Technical meeting on the Implementation of the AR-CALUX® *in vitro* method"

13.7.5 JRC technical report on "Transfer Evaluation Report"

13.7.6 JRC technical report on "Data of study 2"

13.7.7 EURL ECVAM report GLP compliant study SR-ST57

13.7.8 Supporting data for development of the specificity control criterion R^2

13.7.9 Technical issues during implementation of the AR-CALUX® method at CitoxLAB (report of BDS)

13.7.10 Quality control (Identity verification) of the cell lines used in the participating laboratories

13.7.11 Call for expression of interest "Proposal to EU-NETVAL members for participation in a multi-study validation trial"

ANNEX 13.1

STATISTICAL REPORT

Available at EURL ECVAM's

Tracking system for alternative methods towards regulatory acceptance (TSAR)

(https://tsar.jrc.ec.europa.eu/test-method/tm2010-07)

ANNEX 13.2

VALIDATION PROJECT PLAN V05

VALIDATION PROJECT PLAN

Performance assessment of the AR-CALUX *in vitro* method

to support the development of an international test guideline and associated performance standards for Androgen Receptor Transactivation Assays (ARTA) for the detection of compounds with (anti)androgenic potential

Version	**Author**	**Reviewer**	**Sponsor (EURL ECVAM) Approval**	**Validation Study Coordinator for VMG Approval**
01	Anne Milcamps	Ingrid Langezaal Roman Liska Sandra Coecke Raphaella Corvi Pilar Prieto Peraita Sylvia Casati	Maurice Whelan	Anne Milcamps 15/11/2015
Document history				
Version	**Drafted by**	**Date**	**Amendment**	
02	Anne Milcamps	21-12-2015	Added Annex 1 Selection of test chemicals for study 2A; solubility testing protocol	
03	Anne Milcamps	27-02-2017	Added Annex 2 Reorganisation of the studies; solubility testing for all test chemicals	
04	Anne Milcamps	24-03-2017	Added Annex 3 Introduction of criterion for specificity control and classifier in the SOP; update of the test chemicals; selection of 36 test chemicals for studies 2B and 3	
05	Anne Milcamps	26-06-2018	Added Annex 4 A 4th laboratory was added to the validation study	

List of abbreviations

ARTA: Androgen Receptor Transactivation Assay

AR-CALUX method: the transactivation *in vitro* method to measure (anti)androgenic potential of chemicals using the AR-CALUX® cells

ARE: Androgen Responsive Elements

BDS. BioDetectionSystems

BLR: Between Laboratory Reproducibility

CV: Coefficient of variation

DB-ALM: EURL ECVAM DataBase service on ALternative Methods to animal experimentation

DMSO: dimethyl sulfoxide

EC: European Commission

EC_{50}: half maximal effective concentration

ED: Endocrine Disrupter

ER TA: Estrogen Receptor Transactivation Assay

ESAC: ECVAM's Scientific Advisory Committee

EU-NETVAL: European Union Network of Laboratories for the Validation of Alternative Methods

EURL ECVAM: European Union Reference Laboratory for Alternatives to Animal Testing

GLP: Good Laboratory Practice

IATA: International Air Transport Association

IC_{50}: half maximal inhibitory concentration

ICATM: International Cooperation on Alternative Test Methods

ICCVAM: Interagency Coordinating Committee for the Validation of Alternative Methods

LDH: Lactate DeHydrogenase

MSDS: Material Safety Data Sheet

NCP: National Contact Point

OECD: Organisation for Economic Cooperation and Development

OHS: Occupational Health and Safety

PALM: PC-3 human prostate carcinoma cells

PBTG: Performance Based Test Guideline

SARM: Selective Androgen Responsive Modulator

STTA: Stably Transfected Transactivation Assay

SOP: Standard Operating Procedure

STR: Short Tandem Repeats

TA: Transactivation Assay

TG: Test Guideline

ToR: Terms of Reference

VMG: Validation Management Group

VMG-NA: OECD Validation Management Group Non Animal

WLR: Within Laboratory Reproducibility

YAS: Yeast Androgen Screening assay

TABLE OF CONTENTS

1 SUMMARY

The European Commission (EURL ECVAM) proposed to OECD to develop a Performance Based Test Guideline (PBTG) and related Performance Standards for Androgen Receptor Transactivation Assays (ARTAs). Such PBTG and standards will be drafted on the basis of ARTAs for which the validation is ongoing, and the EURL ECVAM coordinated validation study on the AR-CALUX method.

The AR-CALUX *in vitro* method was submitted by the Dutch company BioDetectionSystems (BDS) to EURL ECVAM to be considered for a EURL ECVAM coordinated validation process. The method is applied for the detection of compounds with (anti)androgenic potential. The AR-CALUX® cells are osteosarcoma U2OS cells which are transfected with the cDNA of a human androgen receptor and a luciferase encoding reporter gene preceded by AR response elements (ARE), hence responding to chemicals with endocrine disrupting activity.

EURL ECVAM is both sponsor and coordinator for this validation study. Three test facilities have been selected from the recently established European Union Network of Validation laboratories for alternative methods (EU-NETVAL) to participate in this validation study. A Validation Management Group (VMG) has been established consisting of EURL ECVAM internal staff and external experts in the field with as main task the oversight of activities.

Within this validation study, the following will be addressed: experimental definition of the *in vitro* method following OECD principles of GLP by EURL ECVAMs GLP test facility, assessment of the transfer of the method to each of the 3 laboratories (Study 1 – Transferability), the reproducibility within each laboratory and between laboratories (Study 2 – Reproducibility), as well as the predictive capacity and the applicability domain of the AR-CALUX method (Study 3 – Predictive capacity). Hereto, the 3 test facilities are requested to test 3 sets of chemicals (2 sets will be blinded) with the SOP version provided by EURL ECVAM and to report the data according to predefined Data Analysis Forms. A list of chemicals to be tested has been compiled on the basis of a weight of evidence approach as well as in-house (EURL ECVAM) generated data on the dose response and solubility in DMSO/assay medium. The list is composed of a balanced number of agonists, antagonists and those without any response.

This validation project plan details the objectives of the validation study, the managerial aspects, the overall content of the tasks for the test facilities, the expected deliverables and the respective time periods to be respected.

2 INTRODUCTION AND OBJECTIVES

2.1 General introduction

Endocrine disruptors (EDs) are a high priority topic on the agenda of several national and international governmental institutions given the observed and documented adverse effects on humans and animals' health (UNEP WHO, 2013). These substances impact development and reproduction by disturbing the functioning of the endocrine (hormone) system. The definition of an endocrine disrupting substance has been put forward in 2002 by WHO as "An endocrine disruptor is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub) populations" (WHO/IPCS, 2002). The criteria for defining a substance as an ED are still under debate.

National and international governments are in the process of establishing testing programs and strategies to assess the safety of currently used chemicals with regard to their potential to interfere with the endocrine system. Several pieces of European legislation are addressing EDs, reflecting also the need of scientific criteria to identify EDs: the REACH regulation No 1907/2006; the Regulation No 1107/2009 on Plant Production Products; the Regulation No 1223/2009 on Cosmetic Products; Regulation 528/2012 on Biocidal Products. The European Commission launched in 1991 a work program "Community strategy for endocrine disruptors" (EC, 1997) addressing several actions to establish criteria and identify EDs for further evaluation, to develop and validate test methods to assess EDs, to fund research for understanding the ED mechanisms and to adapt present EU legislation to take account of ED effects. The United States Environmental Protection Agency (EPA) developed i.e. the Endocrine Disruptor Screening Program (EDSP) as one of the first national programs. In Japan, the Japan Environment Agency initiated the Strategic Programs on Endocrine Disruptors (SPEED) to promote e.g. test method development while in the republic of Korea, the relevant ministries developed mid and long term research plans mainly dealing with environmental monitoring (Hecker and Holler, 2011; WHO, 2014).

The Organisation for Economic Co-operation and Development (OECD) recognised the impact of ED substances on human health and environment. Since 1996, effort has been spent in developing Test Guidelines and other tools to support countries' needs for testing and assessment of chemicals for endocrine disruption. In terms of providing overview and guidance, a number of important documents have been drafted, e.g. a Conceptual Framework for testing and assessment of EDs was adopted in 2002, and revised in 2011. It lists OECD Test Guidelines and standardized test methods available, under development or proposed to be used to evaluate chemicals for ED. It is structured over different levels where level 2 includes *in vitro* assays (data about selected endocrine mechanism) and levels 3, 4 and 5 handle *in vivo* assays (data about selected endocrine mechanism and on adverse effects on endocrine relevant endpoints) (OECD, 2012a).

The OECD Conceptual Framework as well as the EPA have recommended *in vitro* assays, and more precisely hormone receptor binding assays and transactivation assays, as important tools for the fast screening of putative EDs and prioritisation purposes. Within the transactivation assays, the Estrogen Receptor Activation Assays (ERTAs) have ligand bound estrogen receptors (ER) leading to transactivation while the Androgen Receptor Activation Assays (ARTAs) have activated androgen receptors (AR) initiating the transcription. Since few years, OECD has initiated the concept of Performance Based Test Guidelines (PBTGs) where similar methods can be included in the same Test Guideline. Performance Standards accompany such PBTG, including a set of Reference Chemicals, accuracy and reliability performance values, applicable for all annexed *in vitro* methods. For ERTAs, such PBTG has been already developed and includes the assays STTA and BG1luc (OECD, 2012b, 2012c). A PBTG for ARTAs is currently under development (see further).

AR-CALUX cells were identified within the EU funded project ReProTect (LSHB-CT-2004-503257) which aimed at optimizing an integrated set of tests as a basis for a reproductive/developmental battery, in order to provide detailed understanding of the main chemical target tissues or biological mechanisms in reproduction. The AR-CALUX cells were considered as addressing a critical biological mechanism (androgen receptor interaction) and therefore a relevant test system for the development of a method for ED screening. The test developer of the AR-CALUX method carried out a pre-validation (van der Burg et al, 2010) and the *in vitro* method was subsequently submitted to EURL ECVAM for a validation process. EURL ECVAM considered the *in vitro* method as sufficiently developed for entering a validation.

2.2 Goal and Objectives

The European Commission submitted in 2012 a Standard Project Submission Form (SPSF) to OECD for the development of a PBTG on ARTAs and this proposal was accepted and inserted in the OECD 2013 work plan. Several ARTAs are currently undergoing a validation process, and when successful, they can be considered to be annexed to the PBTG:

- the AR-STTA assay of Japan, using the AR-EcoScreen Chinese hamster ovary cell line (validation study finalized and validation report under review by OECD VMG-NA)
- the ARTA assay of Korea, using the 22Rv1/MMTV human prostate cancer cell line (validation ongoing)
- the YAS assay of BASF (Germany), using yeast cells (validation study finalized and considered for peer-review)

While EURL ECVAM will coordinate the validation of the AR-CALUX assay of BioDetectionSystems (BDS) (the Netherlands), using the human osteosarcoma based cell line AR-CALUX, it will investigate simultaneously the existing validation data of the Japanese ARTA and follow closely the ongoing validation study of the Korean ARTA. The data of all 3 ARTA validation studies will be analysed to conclude on a set of Performance Standards for ARTAs and hence the drafting of a PBTG to which the 3 methods will be annexed. If the YAS assay will be submitted for peer-review and has a successful outcome, also this assay will be considered in the drafting of the Performance Standards.

The objective of this validation study is to test the AR-CALUX method for its reliability and its relevance, concluding on Performance Standards.

For this purpose, the following tasks will be carried out according to the Terms of Reference (ToR) for EU NETVAL:

- ToR Task i: Definition and description of *in vitro* methods
- ToR Task ii (Study 1): Transfer of *in vitro* methods between laboratories
- ToR Task iii (Study 2): Assessment of the reproducibility of *in vitro* methods
- ToR Task iv (Study 3): Assessment of the predictive capacity and applicability domain of *in vitro* methods

EURL ECVAM will take care of the first task, while 3 European Union test facilities (selected from the European Union Network of Laboratories for the validation of alternative methods (EU NETVAL) will carry out the following 3 tasks with an increasing number of chemicals. With the first study, the successful transfer of the method to the 3 test facilities will be evaluated while studies 2 and 3 will lead to the generation of data to assess within and between lab reproducibility (reliability) as well as predictive capacity and the applicability domain (relevance).

Data generated by the 3 test facilities will be compiled in Data Analysis Forms and Final reports which will be collected, analysed and evaluated by a Validation Management Group (VMG). The data analysis, the conclusions on the method's performance and deduced Performance Standards will be gathered in a validation report. After a positive evaluation by ECVAM's Scientific Advisory Committee (ESAC), a EURL ECVAM recommendation will be formulated.

This validation study will be carried out according to the EURL ECVAM modular approach (Hartung et al, 2004). The modules to be covered include Module 1 (test method definition), Module 2 (within-laboratory reproducibility), Module 3 (transferability between laboratories), Module 4 (between-laboratory reproducibility) and Module 5 (predictive capacity). The data will also be used for Module 6 (applicability domain) and Module 7 (minimal performance standards) for this class of method.

3 THE TEST METHOD TO BE VALIDATED

3.1 Description of the *in vitro* method

The AR-CALUX cell based assay provides information on the endocrine activity of chemicals, and more specifically the (anti-)androgenic activity, when the AR-CALUX cells are exposed to substances. This *in vitro* method is a transactivation assay where the reporter gene *luc* (encoding luciferase) is activated by the androgen receptor but only when bound to a ligand, i.e. a chemical with androgen receptor affinity. This receptor-ligand complex enters the nucleus where it will bind to specific recognition sequences in the promoter region of a target gene (so called androgen responsive elements or ARE). Hence, the target gene will be transcribed. When the target gene expresses the reporter (luciferase), *in vitro* hormonal activity of chemicals can be quantified as well as the agonistic or antagonistic mode of action. Such assays are called Androgen Receptor Transactivation Assays (ARTA).

The AR-CALUX cell line was created via transfection of the human osteosarcoma cell line U2-OS (ATCC HTB 96) with 2 constructs: pSG5-neo-hAR carrying the cDNA of a human androgen receptor under a constitutive promoter, and, pGL3-3XAREtataLuc carrying the luciferase reporter gene which is preceded by a triple tandem of AREs in front of a TATA box.

This cell line has been reported to stably express the human androgen receptor, to be highly selective in its response to low levels of different androgens (due to the multimerized ARE and a minimal promoter – TATA box only), and to have an insignificant response to other nuclear hormone receptor ligands such as estrogens and glucocorticoids (due to the cells not expressing other steroid receptors that can activate transcription via the same ARE as the androgen receptor).

There are no specific limitations of the method reported except for the general limitations of TAs (being direct extrapolation to the *in vivo* complex network of signalling and regulation should not be made; information is gained on the parent molecule while *in vivo* other molecules may be generated due to the cells' metabolism). It can be performed in any laboratory with *in vitro* method expertise and does not require expensive equipment. The assay is a proprietary method of BDS and the cells, the protocol, training and technical support are available through a license agreement.

3.2 Purpose and regulatory context of the *in vitro* method

The AR-CALUX method is intended to be used for screening purposes due to an easy and time efficient application.

Both the OECD Conceptual Framework and the US EPA have recommended transactivation assays as an important tool for fast screening of chemicals with expected endocrine disrupting properties. OECDs Conceptual Framework has identified several type of methods classified over levels, e.g. level 2 involves *in vitro* assays providing mechanistic data. Validated ERTAs are included at this level, but there is still a lack of validated ARTAs. The proposed AR-CALUX method, once validated, could be inserted at this level 2. Moreover, as detailed in section 2.2., this method has potential to be annexed to a PBTG for ARTAs.

3.3 Principle of the *in vitro* method

The method is described by the test submitter to measure the ability of a chemical to activate AR dependent transcription (i.e. act as an agonist) and to suppress AR dependent transcription (i.e. act as an antagonist). Hence, the method is composed of an agonist and an antagonist assay.

Both assays include a pre-screen for determining the appropriate dose range, followed by comprehensive testing. To determine the agonistic or antagonistic nature of a test chemical, it will be tested in the following manner:

1) A dilution series of the chemical is prepared in solvent (e.g. DMSO) and applied to the cells in assay medium. When the luminescent signal increases in a concentration dependent way in comparison to the solvent control, the chemical has an agonistic response.

2) A dilution series of the chemical is prepared in solvent (e.g. DMSO) and applied to the cells in assay medium supplemented with the EC_{50} concentration of Dihydroxytestosterone (DHT). When the luminescent signal decreases in a concentration dependent way in comparison to the solvent control, the chemical has an antagonistic response unless there is a non-specific response. Such is the case when:

- the chemical provokes cytotoxicity which leads to decrease of the luminescent signal. Therefore, a cytotoxicity test must be performed.
- the chemical interferes with the generation of the luminescent signal (e.g. at the level of receptor-ligand binding to the AREs, the transcription of the reporter gene, the translation of the reporter gene, stability of the reporter gene product). Therefore, a specificity control must be performed (based on competition of agonist and antagonist chemical for the receptor) during antagonism comprehensive testing

In order to label a test chemical as an agonist or antagonist, a classification scheme will be applied. To classify a chemical as an antagonist, the outcome of the application of the specificity control is important. EURL ECVAM, while assessing the method (see section 3.6), proposed to include such control. A criterion for the specificity control and the classifier will be introduced in the SOP towards the end of the validation study, before the initiation of Study 3.

3.4 Reference chemicals and control chemicals

Agonist and antagonist assay have each a reference chemical for which the dose response is measured, EC_{50} values, the induction factor and the Z factor (see tables 3 and 4) calculated. It is also the chemical to which the response of a test chemical is compared (normalisation). For both assays, the maximal response of the reference chemical is set at 100% (this is in the agonist assay the highest concentration of the reference chemical and in the antagonist assay the lowest concentration).

The positive control and negative control include the addition of a chemical to the test medium (including DMSO) for which respectively a response or no response is expected.

Both assays have a vehicle control which is the test medium including DMSO (0.1%) while the antagonist testing includes also a solvent control which is the vehicle control plus EC_{50} of DHT.

Table 1: Proposed reference and control chemicals for the agonist assay

	Name	CAS
Reference	Dihydroxytestosterone (DHT)	521-18-6
Positive control	Methyl testosterone	58-18-4
Negative control	Corticosterone	50-22-6

Table 2: Proposed reference and control chemicals for the antagonist assay

	Name	CAS
Reference	Flutamide	13311-84-7
Positive control	Linuron	330-55-2
Negative control	Levonorgestrel	797-63-7

3.5 Acceptance criteria

Criteria for the reference chemical, positive and negative control have been established. An experiment is considered valid and will be accepted when all of these acceptance criteria are met.

Table 3: Acceptance criteria in the agonist assay

No	Acceptance criterium	Value
	Reference chemical DHT	
1	Curve fitting	Sigmoidal
2	EC_{50} range	1.10^{-10} – 1.10^{-9} M
3	CV of estimated log(EC_{50})	< 1.5%
4	Induction factor	> 20
5	Z-factor	> 0.5
	Positive control	
6	Relative induction for Methyl testosterone	> 30%
	Negative control	
7	Relative induction for Corticosterone	< 10%

Table 4: Acceptance criteria in the antagonist assay

No	Acceptance criterium	Value
	Reference chemical Flutamide	
1	Curve fitting	Sigmoidal
2	IC_{50} range	1.10^{-7} – 1.10^{-6} M
3	CV of estimated log(IC_{50})	< 3%
4	Inhibition factor	> 10
5	Z-factor	> 0.5
	Positive control	
6	Relative induction for Linuron	< 60%
	Negative control	
7	Relative induction for Levonorgestrel	> 85%

3.6 Assessment of the AR-CALUX method by EURL ECVAM's GLP test facility

Prior to the start of the validation study, EURL ECVAM has analysed, evaluated and modified in collaboration with the test submitter, the SOP of the AR-CALUX method. This included a technical assessment (paper based) of the SOP(s) in terms of their scientific basis, completeness and clarity, as well as for their suitability to be implemented within a GLP environment. EURL ECVAM's GLP test facility has carried out a GLP study with the method and has generated GLP compliant test data which will serve as a reference data set when assessing the transferability of the method to the test facilities. The experimental assessment and the GLP compliant study have led to a number of changes in the SOP e.g. modified acceptance criteria, inclusion of statistical tools, inclusion of a specificity control for the antagonist assay, etc. (see Annex 3).

The SOP version to be used for the training phase: SOP-ASY06-v03

The SOP version to be used for the transfer phase (Study 1): SOP-ASY06-v04

4 MANAGEMENT OF THE VALIDATION STUDY

4.1 Sponsor

EURL ECVAM accepted the AR-CALUX method, submitted by BDS, for carrying out a validation study and initiated the organisation of the study. Hence, EURL ECVAM is the sponsor.

Sponsor	Address
Maurice Whelan Email: Maurice. Whelan@ec.europa.eu	The EU Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) European Commission Joint Research Centre Institute for Health and Consumer Protection Via E. Fermi, 2749 I-21027 Ispra, Italy

EURL ECVAM will take responsibility for:

- **Provision of relevant documentation:** For proper implementation of the test method EURL ECVAM will provide a complete SOP for all parts of the method, raw data recording forms, respective data analysis forms, as well as an aid for planning and reporting (Study plan templates, Final report templates)
- **Training** on the *in vitro* method for the 3 participating test facilities at Ispra premises (conducted February 10-13, 2015)
- Distribution of the **test system** to the participating test facilities (provided May 19 and June 16, 2015)
- Purchase and distribution of all **reference chemicals, control chemicals and test chemicals** to the 3 test facilities; coding of the test chemicals (provided for Study 1 May 19 and June 16, 2015)
- **Access to CIRCABC**: Access will be provided to the test facilities for the EC database CIRCABC for exchange of all documents (provided February 10-13, 2015).
- An independent **data analysis** and statistical support based on the Final reports generated by the 3 participating test facilities
- **Publication of the method:** At the end of the validation study, the final version of the SOP will be prepared in a format fit for public dissemination through EURL ECVAM's database on alternative methods, DB-ALM.
- **Harmonisation and standardisation of *in vitro* methods:** When the outcome of the validation study is successful, the validation report will be sent to ESAC who provides an opinion on the fitness of the method. Based on this input, EURL ECVAM will formulate a recommendation on the AR-CALUX *in vitro* method and the outcome will be further elaborated in support of the development of an OECD PBTG for ARTAs and Performance Standards

The 3 test facilities that participate in this validation study are responsible for covering the costs of their staff, costs of travel and subsistence for training and meeting(s), and costs of all experimentally related activities (e.g. cell culture medium, reagents, kits).

4.2 Coordination

EURL ECVAM takes care of the coordination of the AR-CALUX validation study and has appointed a validation study coordinator for this particular trial. The coordinator will be supported by an EURL ECVAM biostatistician on a permanent basis and can at any moment, and when deemed necessary, asks for additional support from specific persons with appropriate knowledge in certain topics (e.g. the EURL ECVAM study director who assessed the test method, the Chemical Selection Team, the test method submitter etc.).

The tasks of the coordinator are the following:

- Establishing of and interacting with the VMG (see section 4.3);
- Establishing of and interacting with the Chemical Selection Team;
- Contact point for the managers of the 3 participating test facilities;
- Central point for receipt and/or distribution of all information, documentation and data between all parties involved in the validation study;
- Drafting of the validation project plan with assistance of the biostatistician for the validation study design;
- Review, with assistance of the biostatistician, the Study plans and the Final reports from the test facilities prior to providing them to the external members of the VMG;
- As part of the VMG, assessing and documenting the impact of any amendments and/or deviations in the Study Plans on the quality and integrity of the ring trial;
- Preparing the draft validation report for review and approval by all VMG members;
- Preparing with all members of VMG the draft Performance Standards, draft EURL ECVAM recommendations and a draft PBTG for ARTAs;
- Preparing the final version of the SOP and ensuring its publication in DB-ALM

A Chemicals Selection Team has been established including EURL ECVAM staff. The tasks of the Chemical Selection team are the following:

- Compilation of a list of relevant test chemicals
- Solubility and stability assessment of the test chemicals
- Purchase, coding and distribution of all reference/control/test chemicals to the participating test facilities

4.3 Validation Management Group (VMG)

The VMG encompasses collective expertise with similar test systems and test methods, within the field of developmental and reproductive toxicology, with the validation process of test methods and with management and evaluation of a validation study.

The VMG for the AR-CALUX validation study will consist of external experts in the field, the validation study coordinator, the biostatistician from EURL ECVAM and representatives of the ICATM validation bodies. The latter ones do not participate in the decision-making of the VMG.

The VMG will provide oversight on the validation study. Its responsibilities are defined in a Terms of Reference, and include:

- To review and approve the validation project plan in all its components (objectives, validation study design, organisation, statistical analysis methods, list of chemicals to be tested, the SOP to be implemented);
- To monitor progress through setting key milestones and reviewing the results of the test facilities and to provide assistance in troubleshooting when need be;
- To manage deviations to the validation study;
- To interpret the validation results and formulation of conclusions;
- To assist, review and approve the validation report;
- To assist in the drafting of the EURL ECVAM recommendation and the PBTG for ARTAs.

Composition of the AR-CALUX VMG

Name	Role and expertise	Affiliation
EURL ECVAM members		
Anne Milcamps	Coordinator	EURL ECVAM, Ispra, Italy Email: anne.milcamps@ec.europa.eu Phone: +39 0332785244
Roman Liska	Data analysis and preparation of biostatistics dossier	EURL ECVAM, Ispra, Italy Email: roman.liska@ec.europa.eu
External members		
Warren Casey	Director NICEATM Expertise in toxicology, EDs, validation	NIEHS/NICEATM, Research Triangle Park, North Carolina, USA Email: warren.casey@nih.gov
Matt Dent	Safety science leader Expertise in toxicology and risk assessment	Unilever, Colworth Science Park, Bedford, UK Email: matthew.dent@unilever.com
Jenny Odum	Independent consultant toxicologist Expertise in toxicology, EDs, validation	Stockport, UK Email: jenny.odum@regulatoryscience.com

Representatives of ICATM bodies: to be nominated.

4.4 Lines of communication

In order to have a uniform distribution of information to all parties, each test facility is requested not to engage in any contact with the other test facilities neither with the test submitter nor with the VMG or any other third party during the course of the validation study. Such has also been stipulated in the signed collaboration agreements. All flow of information, data, inquiries, requests from each of the 3 participating test facilities, the VMG or the test submitter shall therefore go through one central contact point: the validation study coordinator.

For the daily business, a designated email address for the validation study has been established: JRC-ECVAM-NETVAL@ec.europa.eu. Details of the coordinator for phone calls have been provided in section 4.3. Video-conferences can be arranged on a need be basis.

The AR-CALUX validation study will make use of the European Commission's data application CIRCABC, a free and open source software for the creation of a collaborative workspace between geographically dispersed teams, that allows viewing, uploading, downloading, updating and copying of information. Test facility and the VMG members will be granted access to a specific workspace in CIRCABC (via password protected access) to which only the coordinator, the side manager and the EU-NETVAL coordinator have access. Each test facility and the VMG will have its own private space. This tool will be used for exchange of documents and data between the coordinator and the manager of each of the 3 participating test facilities, and, between the coordinator and the VMG. Whenever a document is uploaded, notice of insertion is automatically provided to the coordinator.

5 PARTICIPATING EU-NETVAL TEST FACILITES

5.1 EU-NETVAL and procedure of selection

EU-NETVAL was established in January 2014. It comprises 25 labs of the EU and EURL ECVAM's GLP test facility. Its tasks and responsibilities are formulated within a ToR with primary focus on the participation in a validation study.

Beginning 2014, an invitation to the EU-NETVAL members for participation in the AR-CALUX validation study was launched. Several test facilities met the requirements and were ranked on the basis of a communicated selection procedure. The 3 highest ranked facilities were approved by the EU Member States via the National Contact Points. Collaboration agreements have been put in place between the JRC and each of the 3 participating test facilities.

5.2 The 3 participating EU-NETVAL members

	EU-NETVAL member / GLP test facility		Country
1	SP Technical Research Institute		Sweden
	Address: SP Sveriges Tekniska Forskningsinstitut Medicinteknik (KMm) Brinellgatan 4, 501 15 Boras		
	Test facility manager: Benny Lyven Study director: Emma Pedersen(*) Study personnel: Kristina Fant Study personnel: Lovisa Ringstad Study personnel: Jenny Johansson (*) Will be leaving October 2015 and replaced by K. Fant	Email: benny.lyven@sp.se Email: emma.pedersen@sp.se Email: kristina.fant@sp.se	
2	CiToxLAB		France
	Address: CiToxLAB, BP 563, 27005 Evreux Cedex		
	Test facility manager: in replacement Study director: Mylene Valin Study personnel: Rachel Larcier Study personnel: Megane Auvray	Email:mylene.valin@fr.citoxlab.com	
3	Huntingdon		UK
	Address: Huntingdon Life Sciences Limited, Woolley Road, Alconbury, Huntingdon, Cambridgeshire, PE28 4HS		
	Test facility manager: Leslie Akhurst Study director: Joanne Hubbard Study personnel: Joanne Kelsall	Email: AkhurstL@UKOrg.Huntingdon.com Email: HubbardJ@UKOrg.Huntingdon.com Email: KelsallJ@UKOrg.Huntingdon.com	

6 TEST CHEMICALS AND TEST SYSTEM

6.1 Selection procedure of the test chemicals

A list of 83 chemicals, and their reported classification as positive or negative (anti)androgen, was compiled on the basis of several data sources: literature (Araki et al, 2005; Freyburger et al, 2012; Van der Burg et al, 2010; ICCVAM, 2003), Tox21 ARTA assay data, in-house high throughput screening data for the AR-CALUX method, QSAR. The list was shared for an expert consultation and provided to all members of the VMG-NA in 2013, ICATM contact points and other experts. Details on the composition of this list can be found in the Report on the Chemical Selection (see Annex 2).

From the list of 83 chemicals, a subset of 45 chemicals was selected for the AR-CALUX validation study, aiming at a balanced set of agonist, antagonist and negative chemicals (about 15 for each class). The criteria listed below were considered:

- Dose responses and classifications
- Availability and price
- Solubility as stock solution and as working solution (in cell medium).
- Potency
- Structural diversity
- Glucocorticoid receptor crosstalk
- Requests from the VMG-NA

Details on the selection strategy are outlined in the Report on the Chemical Selection (see Annex 2).

The requests from VMG-NA at the meeting of December 2014 were taken into consideration. Amongst the chemicals with antagonistic response, an inclusion of pure (true) antagonists, unspecific (false) antagonists and SARMs (selective androgen receptor modulator) could be aimed for.

- Chemicals with unspecific antagonist behaviour are those that do not compete for binding at the androgen receptor but interfere with the generation of the reporter enzyme (at the level of transcriptional activation, transcription, translation, stability). Few such chemicals have been described for ERTAs though for ARTAs not much (yet) is reported. Possible candidates will be discussed with the AR-CALUX VMG.
- SARMs are molecules with a tissue selective activation of androgenic signalling. Examples of selective estrogen receptor modulator are known, though few information exists for SARM. In the original list of 83 chemicals, few could be found reported as agonist in certain assays (cell lines) while antagonist in other assays (cell lines) and hence could be potential SARMS. At the VMG-NA meeting of December 2014, the members proposed to include this type of chemicals. Possible candidates will be discussed with the AR-CALUX VMG.

Overlap with the test chemicals of the Japanese ARTA (under validation) and the Korean ARTA (for which data were presented at the VMG-NA meeting Dec. 2014) was evaluated. During the course of drafting this validation project plan, the identity of the chemicals tested in these 2 ARTAs became available.

- Most of the 15 chemicals from the Japanese ARTA list were already part of the selected 45 subset for AR-CALUX. Four have been added of which one is currently analysed in-house to check if its solubility is sufficiently high to be tested. The chemical R1881 was excluded (drug banned in several countries).
- Of the 20 chemicals from the Korean presented study, several were already included in the subset for AR-CALUX. Three of that list could be interesting to be tested (described as agonist and dual behaviour) though information (data) are needed as well as solubility testing.

The list of the subset of chemicals for the AR-CALUX validation is shown in Annex 1. This list currently includes 49 chemicals (including the reference and positive control chemicals for the AR-CALUX assay), and, 3 additional chemicals to be considered for inclusion.

6.2 Procurement and solubility testing of the test chemicals

EURL ECVAM will be responsible for the acquisition of all chemicals (including reference and control chemicals). Previous validation studies at EURL ECVAM have shown that in many occasions the participating test facilities have solubility issues what leads to extra work, more consumption of chemicals, and significant time loss. Moreover, solubility testing usually relies on visual inspection which is rather subjective. For the AR-CALUX validation study, it was decided to measure in an objective manner (via nephelometry) the solubility of all the test chemicals at a maximal concentration of 50 mg/ml in the solvent DMSO or water and at the working concentration (1000 x diluted) in assay medium. Also the stability of the solubilised chemical over time was investigated (24 hours of incubation mimicking the test conditions of the SOP).

For Study 2, the 3 labs will perform a solubility test with a SOP to be prepared by EURL-ECVAM. On the basis of the obtained results, a decision will be taken if further solubility testing for the test items of Study 3 are needed. Chemicals will be tested at 50, 15 or 5 µg/ml assay medium. Alternatively, the participating test facilities could be provided with information for the type of solvent to use and the maximal concentration to be prepared for their test chemicals

6.3 Distribution and handling of the test/reference and control chemicals

EURL-ECVAM will be responsible for chemical distribution to the participating laboratories (including reference and control items) assigning a unique random identity code to replicate test chemical aliquots. The vial labels will also include a cautionary *toxic* indication.

In compliance with IATA (International Air Transport Association) regulations, the chemicals will be sent with MSDS copies enclosed in a single envelope, including a list of corresponding codes, indicated *for customs use only*.

Relevant to remedial procedures at a test facility in case of accident or emergency, duplicate MSDS copies will also be included, sealed individually in opaque envelopes identified by code. Consultation of a particular MSDS, for reasons of safety only, would be reported as a study deviation. At the end of the validation study, the Safety Officer shall return the unopened MSDSs to the validation study coordinator.

The test chemicals will be addressed to nominated personnel, informing them of imminent arrival (e-mail) and requesting acknowledgement of receipt confirming (or otherwise) integrity of the MSDS envelopes.

The consignment will include a checklist of the test chemicals, including expiry dates, storage conditions, and material weight. The list will also indicate appropriate stock solution concentrations (and solvent) based on solubility determinations at EURL ECVAM, with summary preparation procedure. In addition, there will be a Chemicals Receipt Form to be signed and returned (via CIRCABC).

Test Facility personnel shall be instructed to treat all coded test chemicals as ***potential EDs.***

6.4 Test system

The test system constitutes of AR-CALUX® cells developed by the Dutch company BDS. EURL ECVAM has prepared a cell bank of this test system for distribution to the test facilities. Each test facility will receive at the onset of the trial 6 vials of frozen cells: one vial to be used to practice with the cells and the method, one vial for the cultivation of cells for Study 1, one vial for Study 2, two vials for Study 3, and an extra vial.

The cells have been tested for purity by EURL ECVAM: 1) the absence of Hepatitis B and C, and HIV 1 and 2; 2) the absence of mycoplasma; 3) authenticity: the absence of cross contamination from other cell lines (STR profiling). A Certificate of Analysis shall be provided with the shipment of the cells. Upon receipt, each test facility is requested to complete and return a Test System Receipt Form (via CIRCABC).

Upon arrival of the cells, the internal procedures related to test system/mycoplasma testing prior to the initiation of a study and at completion of a study, should be followed.

6.5 Return of the test system and test chemicals to EURL ECVAM

Each test facility commits to freeze down 2 vials of cells from the highest passage number for each of the cell vials used for the experimental testing phases. At the end of each study (Study 1, Study 2 and Study 3) the test facilities are requested to send these 2 vials to EURL ECVAM where they will be used for characterisation purposes.

The test facility agrees not to use the test system for any other purposes than described in the validation project plan and to destroy the remaining test system (frozen or in culture) not later than 4 weeks after the completion of the validation study. The coordinator will communicate when the validation study is officially completed.

Each test facility is also requested to keep the remainder, if any, of the test chemicals until the end of the validation study and all data have been analysed. In case of problematic results for a specific test chemical, EURL ECVAM may request to return the test chemical. At the end of the validation study, all chemicals should be disposed of according to the test facilities internal procedures.

7 VALIDATON STUDY DESIGN

The following sections describe the experimental activities to be undertaken by the 3 participating test facilities.

7.1 Participation in training

Training for the 3 test facilities was carried out in February 10-13, 2015. This training was provided by EURL ECVAM, assisted by the test submitter, and aimed at an appropriate implementation of the AR-CALUX method by all participating test facilities in order to achieve a standardised and harmonised application of the method across the facilities.

Staff of each participating test facility (Study Director and Study Personnel) was invited to receive training. It is mandatory that the staff trained will be those managing and executing the experimental work as described below.

During the training, each test facility received all necessary documentation from EURL ECVAM for implementation of the AR-CALUX method (e.g. Standard Operating Procedure (SOP), related Data Analysis Forms). The training covered practical and theoretical sessions, e.g. hands-on execution of critical steps in the SOP, a correct compilation of the Data Analysis Forms, guidance on data analysis and acceptance criteria, information and awareness on critical steps, sharing of experiences and trouble shooting, usage of CIRCABC. At the end of the training, the required experience was tested via a questionnaire and a certificate was handed out.

In the undesirable event that a test facility changes staff during the ring trial, the training and demonstration of competence of the new staff member(s) to correctly implement the method will be under the responsibility of the test facility. EURL ECVAM will need to be informed immediately and all relevant details should be reported in the Study plan and/or Study report.

After the training, the test facilities have been asked for critical observations and comments on the SOP. The comments received have been considered for the updated version of the original SOP.

7.2 Study 1: Transfer of the AR-CALUX method to the EU-NETVAL test facility

The AR-CALUX method will be implemented by each test facility in its own laboratory. For this Study, 6 non-blinded test chemicals will be tested (defined by EURL ECVAM). Prior to the initiation of the experimental work, a Study Plan template will be provided by EURL ECVAM to facilitate the drafting of the Draft Study plan by each test facility. This Draft Study plan will be provided to the coordinator for review and feedback. The finalised and signed Study plan shall be sent to the coordinator.

The testing regime for the transfer phase will be identical to the regime required for Study 2 and Study 3 (described in section 7.5).

After completion of the experimental work, the results shall be reported with Data Analysis Forms and with a Draft Final report, to be send to the coordinator. The VMG will evaluate the results. Feedback on the Draft Final Report will be provided and the Final report can be completed by the test facility, and returned to the coordinator.

Success of the transfer of the AR-CALUX method will be concluded on the basis of the criteria defined by VMG for a successful transfer.

VMG will evaluate the data from the transfer phase from all 3 test facilities, as well as the observations made by the test facilities when implementing the *in vitro* method. Such will lead to conclusions regarding the robustness of the SOP. If deemed necessary, amendments may be made to the SOP. The test facilities will be informed regarding:

- Continuation (or not) of the validation study and to proceed to Study 2.
- Any minor amendments to the SOP if deemed necessary and a new version of the SOP will be provided.

The test facilities are requested to wait for the feedback from the coordinator before proceeding to Study 2.
In case the results of the transfer phase would be exceptionally out of range, a face-to-face meeting of the test facilities with EURL ECVAM may be needed to discuss the data. Such meeting would last a maximum of one day.

7.3 Study 2: Generation of data sets for the assessment of reproducibility (within and between laboratory) of the AR-CALUX method

Assessment of reproducibility involves the generation of data sets under blinded conditions by applying the *in vitro* method to coded chemicals. Study 2 will consist of testing about 10 test chemicals (defined by the VMG). The exact number of chemicals to be tested will be communicated to the test facility upon completion of Study 1.

Prior to the initiation of the experimental phase, the test facility will prepare a Draft Study plan to be provided to the coordinator for verification. Experimental work shall be carried out in accordance with the SOP and according to the testing regime as described in section 7.5.

After finalisation of the experimental work, the following shall be provided to the coordinator: completed Data Analysis Forms and a Draft Final report. Feedback on the Draft report will be provided and the Final report can be completed by the test facility, and returned to the coordinator.

The results (per test facility and between test facilities) will be discussed within the VMG. On the basis of the VMG conclusions, each test facility will be informed regarding:

- Continuation (or not) of the validation study and to proceed to Study 3.
- Any minor amendments to the SOP if deemed necessary.

The test facilities are requested to wait for the feedback from the coordinator before proceeding to Study 3

7.4 Study 3: Generation of data sets for the assessment of the predictive capacity and applicability domain of the AR-CALUX method

Assessment of predictive capacity and applicability domain essentially involves the testing of a sufficiently large number of test chemicals to determine the accuracy of the method to detect agonist/antagonist activity of a test chemical. This study will consist of the testing of 30 to 50 chemicals (defined by the VMG). The exact number of chemicals to be tested will be communicated to the facility upon completion of Study 2.

Prior to the initiation of the experimental phase, the test facility will prepare a Draft Study plan, to be provided to the coordinator for verification. Experimental work shall be carried out as described for Study 2. The reporting will be similar as described in Study 2.

The results (per test facility and between test facilities) will be discussed within the VMG and conclusions will be drawn. When Study 3 is finalised by the 3 participating test facilities, all test facilities will be invited to a meeting at the JRC (Ispra, Italy, in order to present their results to the VMG. The overall outcome of the work, issues that arose, conclusions to be drawn from the validation study will be discussed.

This meeting will last a maximum of one and a half days.

7.5 General work flow (testing regime) for Study 1, Study 2 and Study 3

Each test chemical shall be tested with both agonist assay and antagonist assay.

The testing regime for each test chemical typically consists out of pre-screen experiments followed by comprehensive testing for the generation of 3 valid data sets. This regime, including biological replicates and decisions to proceed to comprehensive testing, is presented in figure 1 and applies to each Study in this ring trial. For each study, the testing regime will be detailed in the Study plan.

Pre-screen

Pre-screen experiments are carried out to determine 1) if the test chemical displays a significant response 2) to determine the proper dose range, non-cytotoxic and soluble, for a test chemical showing a response (full or partial). The determination of a response being significant is imbedded in the Data Analysis Forms and relies on a Anova Test. For the dose range finding, the pre-screen will be carried out once (or more) to determine the appropriate range to be tested in the subsequent comprehensive tests. There is no requirement for 3 valid runs at the stage of pre-screens. The data generated in the pre-screen need to fulfil the acceptance criteria for reference and control chemicals as described in the SOP (valid run).

Exceptions on the above regime:

- If in the pre-screen (agonist or antagonist assay) the chemical displays no response, there is no need to proceed to comprehensive testing but a total of 3 valid pre-screen experiments is required. In the event that the second or the third pre-screen experiment would however show response, proceed with comprehensive testing.
- If in a pre-screen experiment for an agonist assay a full dose response can be captured and all parameters can be estimated, proceed nevertheless to comprehensive testing (refinement of the calculations). If however the comprehensive test would reveal that part of the dose response cannot be captured, pre-screen experiments are sufficient (3 valid runs).
- If in a pre-screen experiment for an antagonist assay, the data show a full dose response and all parameters can be estimated, comprehensive testing is mandatory.
- In certain cases where a pre-screen experiment (leading to a comprehensive test) reveals the dose range to be chosen but it is not a valid run, the pre-screen test can be accepted when the invalidity is unmistakably due to manual errors (e.g. pipetting error leading to a single outlier) and can be clearly evidenced.

Comprehensive testing

The comprehensive testing (for test chemicals that show a response) needs to include 3 valid experiments per test chemical. An experiment is valid when all acceptance criteria have been met. If it is practically difficult to meet the acceptance criteria (e.g. for a problematic test chemical), leading to invalidity of experiments, it will suffice to perform a maximum of six biological replicates irrespective of their validity. For each comprehensive test, 8 concentrations per test chemical are tested. Each concentration is tested 3 times as defined by the plate layout for both agonist and antagonist testing.

In case of an agonistic response, the testing will proceed as described above. In case of an antagonistic response (full or partial) each comprehensive test will be accompanied by a specificity control. This specificity control will give additional information not only for the specificity of the response, but also (indirectly) about the cytotoxic effect.

Cytotoxicity tests (LDH) are included for all plates in the pre-screen experiments. For comprehensive testing it is sufficient to perform this only for the first run. In case of only pre-screen runs, it suffices to perform 2 LDH tests. Visual inspection will be applied for all runs regardless if they are pre-screen or comprehensive.

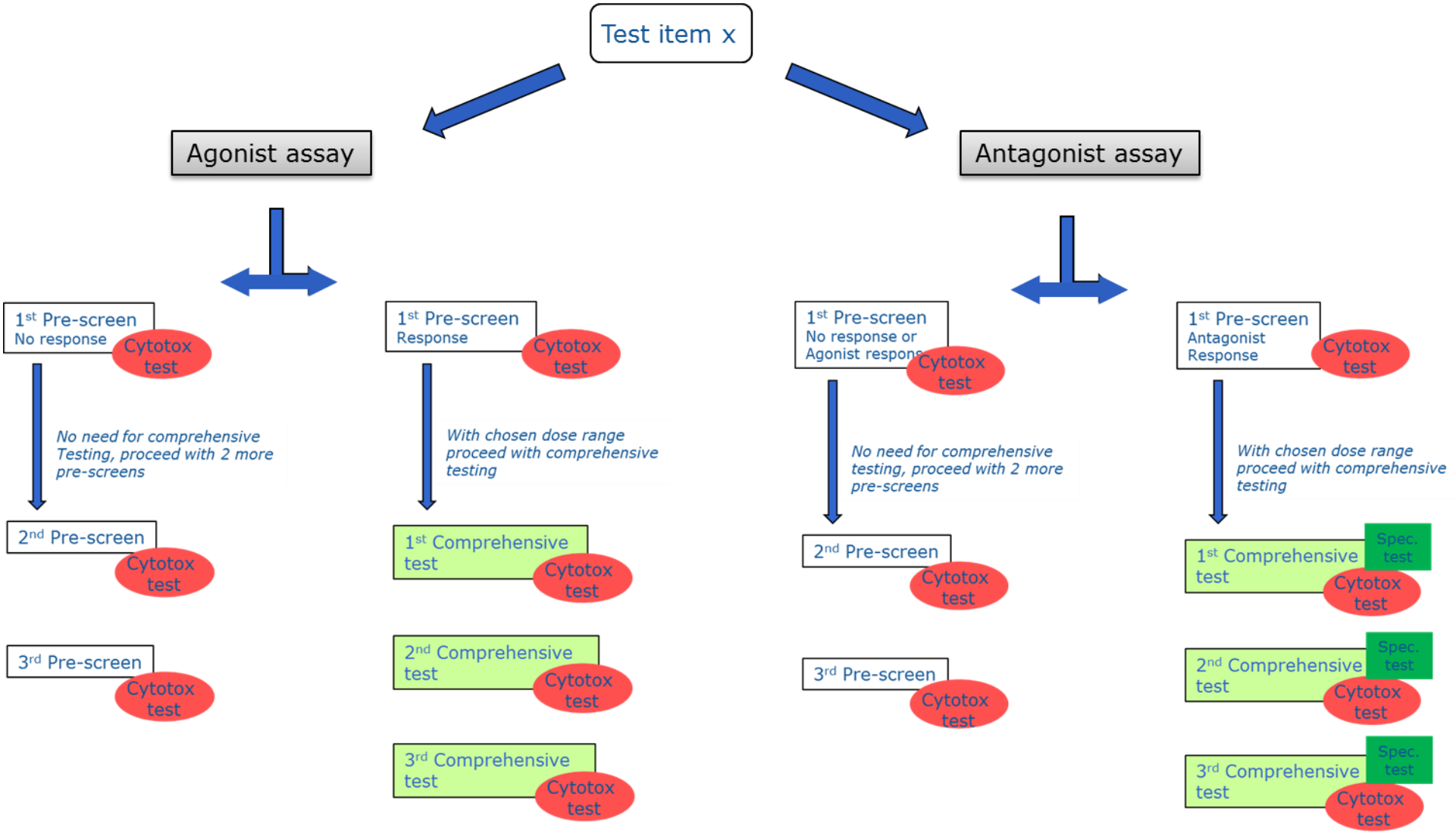


Figure 1: Testing regime for the AR-CALUX method

8 DATA COLLECTION AND ANALYSIS

8.1 Data collection

EURL ECVAM has prepared and will distribute validated Data Analysis Forms for the collection and analysis of the data:

- Form DAT02-ASY06 for the agonist assay (pre-screen and comprehensive testing)
- Form DAT04-ASY06 for the antagonist assay (pre-screen)
- Form DAT05-ASY06 for the antagonist assay (comprehensive testing and specificity control)
- Form DAT06-ASY06 for cytotoxicity data recording

The test facilities will receive the validated forms with a data set that must be used to verify the correct functioning of the Data Analysis Forms at the test facilities premises. The test facilities will have procedures in place for quality control of correctly reporting the generated data.

The validation study coordinator and the EURL ECVAM biostatistician will retrieve all data in a controlled manner (completed Data Analysis Forms, Draft Final reports and Final reports) of all test facilities via CIRCABC.

8.2 Acceptance of data sets

The test facility is required to provide all the obtained data, being either valid or invalid. "Valid" data sets are defined as data that are in accordance with the acceptance criteria of the AR-CALUX method. "Non valid" data sets are defined as the data from failed experiments (not meeting the acceptance criteria). The test facility is requested to report the measures that have been taken to overcome any failure to meet the acceptance criteria. Such can be reported in the Final report.

8.3 Data analysis

Upon receipt of the Final report and the completed Data Analysis Forms from each test facility, a completeness check will be carried out by EURL ECVAM with which will be verified that all Data Analysis Forms for all valid and invalid experiments are submitted with the Final report, and, that the parameters reported in the Final report correspond to the parameters calculated in the Data Analysis Forms.

The EURL ECVAM biostatistician will provide statistical evaluation of all 3 studies on the basis of the following criteria:

- The number of valid/invalid runs (acceptance criteria met or not) and the reasons for invalidity
- Similarity of the obtained patterns (dose responses)
- EC_{50}/IC_{50} estimates of reference and test chemicals (where possible) and its variability
- Comparison to data generated during EURL ECVAM's GLP study ST57 (especially for Study 1)

The determination of reproducibility within laboratory (WLR) and between laboratories (BLR) will be based on a classifier to determine agonist and antagonist behaviour. The classifier is developed by EURL ECVAM, in collaboration with the test submitter, and includes a criterion for the specificity control to be applied in the antagonist assay. This classifier will be introduced in the SOP before Study 3 will start.

The evaluation made by the biostatistician will be in the form a statistical report for each study (Study 1, 2, and 3). The report will be mainly based on

- The results presented per test facility in their Final reports
- An analysis of the data per study merged from the data provided by the 3 test facilities

The statistical reports will be discussed with VMG.

A final statistical report will be assembled, based on the statistical reports made for Study 1, 2 and 3. The main purpose will be to provide overall information about WLR, BLR and predictive performance of the AR- CALUX method.

9 RECORDS, OUTPUT, AND ARCHIVING

9.1 The SOP

EURL ECVAM will provide the SOP version to be used in the validation study. A GLP test facility may want to use its own in house template for the SOP. Based on the results of Study 1, the SOP may be subject to modifications in which case a new version will be issued. At the end of the validation study, and as a result of the observations during the validation, the SOP may be further modified. A final SOP version will be prepared by the coordinator for publication in DB-ALM.

9.2 Study Plans

Before starting Study 1, 2 or 3, a Draft Study plan has to be prepared by each test facility for each Study and be sent to the validation study coordinator for review. An example of a Study plan will be provided by EURL ECVAM. A GLP test facility may want to use its own in house template. After approval of the Draft Study plan, the test facility can finalise its Study plan (dated, signed and uploaded in CIRCABC) and may commence the experimental part.

9.3 Raw Data Recording Forms

EURL ECVAM will provide raw data recording forms together with the AR-CALUX SOP-ASY06. Test facilities may use their own developed forms, as long as the raw data requested in the SOP are recorded. These forms should be stored and archived by the test facility according to the provisions of their quality system.

9.4 Data Analysis Forms

Data Analysis Forms are the main manner of recording the obtained data. All information on the test items and the experiment, including test item ID, concentrations, plate numbers, cytotoxicity, non-solubility, RLU values, observations etc. are collected in these forms. Validated forms are provided by EURL ECVAM. Each test facility will validate the received Data Analysis Forms according to their internal procedures. For this purpose, a dedicated set of data will be provided by EURL ECVAM.

For each study, the Data Analysis Forms need to be completed and to be sent, together with the Draft Final report, to the validation study coordinator.

9.5 Final Reports

If requested, an example of a Final report will be made available by EURL ECVAM. The GLP test facilities may want to use their own in house templates. At the end of each study, a Draft Final report will be prepared and provided to the validation study coordinator. VMG will provide comments in due time, hence a Final report can be prepared and returned to the coordinator.

The results from the experiments (valid and invalid) shall be reported as a summary based on the Data Analysis Forms. Any observation during the planning and implementation of the studies shall be reported as well as, and if appropriate, recommendations on the performance of the *in vitro* method and the suitability of the protocols from the participating test facility perspective. In addition, any deviations and/or amendments from the original Study plan or the EURL ECVAM SOP(s) have to be reported.

9.6 Validation Report

A draft Validation report is prepared by the coordinator, based on the Final reports of the 3 test facilities and complemented with the overall data analysis from the biostatistician. Such report will be analysed and

reviewed by VMG and needs final approval from the VMG. The finalised Validation report will be shared with the test facilities and the test submitter.

9.7 Archiving

Raw and processed data produced in each GLP test facility shall be stored and archived in the individual GLP test facilities according to their GLP procedures.

The documentation provided by the 3 GLP test facilities in CIRCABC (verified electronic copies of Draft Study plans, Study plans, Draft Final report, Final reports, completed Data Analysis Forms) will be retrieved by EURL ECVAM and safely stored. All SOP versions, the Validation project plan and the Validation report will also be archived.

10 TIME SCHEDULE AND OVERVIEW DELIVERABLES

The timeframe for the completion of all experimental tasks is **18 months**. The **indicative** time schedule is as follows:

Months	Tasks	EURL ECVAM Deliverables/Actions	Test facility Deliverables
1	**Training**	• Training Certificates	
2– 6	**Study 1.** Transfer of the *in vitro* method to the EU-NETVAL test facility	• Shipment of test system and chemicals • Validation project plan • Template Study plan • Template Final report • Data analysis report on of the 3 test facilities • Assessment of transferability and admission to Study 2	• Approved Study plan on the transferability • Completed Data Analysis Forms • Final report verified by test facilities QA
7 - 10	**Study 2.** Generating data for the assessment of the reproducibility (within- and between laboratory) of the AR-CALUX method	• Shipment of coded chemicals • Data analysis report on Study 2 of the 3 test facilities • Assessment of the reproducibility and admission to Study 3	• Approved Study plan • Completed Data Analysis Forms • Final report verified by test facilities QA
11 – 18	**Study 3.** Generating data for the assessment of the reproducibility (within- and between laboratory) and for the assessment of the predictive capacity and applicability domain of the AR-CALUX method	• Shipment of coded chemicals • Data analysis report of the 3 test facilities • Assessment of the reproducibility (within- and between laboratory) and predictive capacity of all data generated in studies 1, 2 and 3	• Approved Study plan • Completed Data Analysis Forms • Final report verified by test facilities QA
19-22		• AR-CALUX Validation report	

11 PERFORMANCE OF WORK AND QUALITY ASSURANCE

The GLP test facility should make all effort to ensure that the experimental work for the Studies 1 to 3 is performed using the same equipment, the same reagents (i.e. of the same supplier and the same batch) and the same staff for testing of all the test chemicals during the entire duration of the validation trial. Change of equipment, reagents or/and staff has to be reported in the Study plan or Final report(s) of the studies concerned.

Each GLP test facility shall work according to their quality control procedures relating to the correct data reporting.

Any deviations or amendments that the study director considers critical to the progress of the study should be reported to the coordinator.

EURL ECVAM will inform the test facility in due time if GLP compliant studies are needed. In case of GLP compliant studies, a copy of the QA statement must be provided.

12 LIST OF RELEVANT DOCUMENTS

Araki, N., Ohno, K.,Takeyoshi, M., Iida, M. Evaluation of rapid in vitor androgen receptor transcriptional activation assay using AR-Ecoscreen cells. Toxicology in Vitro (2005), 19, 335-352)

Freyberger, A., Witters, H., Weimer, H., Lofink, W., Berckmans, P., Ahr, H-J. Screening for (anti)androgenic properties using a standard operation protocol based on the human stably transfected androgen sensitive PALM cell line, first step towards validation. Reproductive Toxicology (2012), 30, 9-17

Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Haider, M., Hoffmann, S., Roi A.J., Prieto, P., Sabbioni, E., Scott, L., Worth, A. and Zuang, V. A Modular Approach to the ECVAM Principles on Test Validity. ATLA (2004), 32, 467-72.

Hecker, M. and Holler, H. Endocrine disruptor screening: regulatory perspectives and needs. Env. Sciences Europe (2011), 23, 15

Van der Burg, B., Winter, R., Mana, H., Vangenechten, C., Berckmans, P., Weimer, M., Witters, H., van der Linden, S. Optimization and prevalidation of the *in vitro* AR CALUX method to test androgenic and antiandrogenic activity of compounds. Reproductive Toxicology (2010), 30, 18–24.

Biocidal Products Regulation (EU) No 528/2012 concerning the making available on the market and use of biocidal products. OJEE L167/1

Cosmetic Products Regulation (EC) No 1223/2009. OJEE L342/59

EC, 1997: Community strategy for endocrine disruptors, a range of substances suspected of interfering with the hormone systems of humans and wildlife (COM (1997) 706)

ICCVAM, 2003. Evaluation of *in vitro* test methods for detecting potential endocrine disruptors: estrogen receptor and androgen receptor binding and transcriptional activation assays (2003). NIH publication No:03-4503

OECD (2012a): Conceptual Framework for testing and assessment of EDs. Guidance document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Series on Testing and Assessment No. 150. ENV/JM/MOMO(2012)22

OECD (2012b): OECD Guideline for the testing of Chemicals, TG 455. Performance based test guideline for stably transfected transactivation *in vitro* assays to detect estrogen and androgen receptor agonists. To replace the current TG 455 on estrogen receptor agonists

OECD (2012c): Performance Standards for stably transfected transactivation *in vitro* assays to detect estrogen agonists for TG 455. Series on Testing and Assessment. No. 173. ENV/JM/MONO(2012)18

Plant Production Products Regulation (EC) No 1107 /2009 concerning the placing of plant protection products on the market and repealing council directives 79/117/EEC and 91/414/EEC. OJEE L 309/1

REACH Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJEE L396

UNEP WHO, 2013: State of the science of endocrine disrupting chemicals, 2012. Available from http://unep.org/pdf/9789241505031_eng.pdf

WHO/IPCS, 2002: Global assessment of the state-of-the-science of endocrine disruptors. World Health organization International Progamme on Chemical Safety. WHO/PCS/EDC/02.2

Annex 1: Update of the validation project plan

Date of discussion with VMG: 21-12-2015

1. Test chemicals for Study 2A (10 coded test chemicals)

Seq. #	Chemical name
1	17β-Trenbolone
2	Stanozolol
3	Spironolactone
4	Medroxyprogesterone acetate
5	Bisphenol A
6	Bicalutamide
7	Disulfiram
8	Tamoxifen
9	Atrazine
10	17α-Ethynyl estradiol

Disulfiram was included as a potential false positive antagonist. Such behaviour was observed in the Tox21 assay luc where it displayed its antagonist activity at concentrations that were not scored as cytotoxic.

2. Solubility testing: Protocol

A simple protocol was developed by EURL ECVAM (see below) based on visual inspection, to be used for the testing of the 10 test items for study 2.

Solubility Determination by Visual Inspection

applicable to '*Transactivation assay for the detection of compounds with (anti)androgenic potential using AR-CALUX® cells*'

Standard Operating Procedure

ID	Version	Date
SOP-ASY15	Version 1	23/05/2016

		Signature	Date
Author (Lead laboratory)	Dr. Thomas Cole		
Review (Lead laboratory)	Dr. Ingrid Langezaal		
Approval (Representative of Validation Management Group)	Dr. Anne Milcamps		

1. PURPOSE AND APPLICABILITY

This standard operating procedure (SOP) describes chemical solubility determination based on visual inspection, applicable to '*Transactivation assay for the detection of compounds with (anti)androgenic potential using AR-CALUX® cells*'.

The purpose is to assess stock solution solubility in solvent, with observation of stability on dilution (500-fold interim and 1000-fold complete) in medium, including 24 hours incubation equivalent to the conditions of assay cell culture/test item exposure.

The solvent is DMSO (commonly applicable) with water as alternative (for inorganic salts).

Solubility determination in solvent is based on visual inspection for signs of turbidity due to solid particulate or liquid droplet suspension.

Stability determination in medium, where observation of dissolution may be obscured by foaming, is assisted by centrifugation to detect any insoluble suspension as a sediment deposit. Solubility in medium is also checked by microscope examination of a droplet.

An overview (flowchart) of the procedure is included in Annex (section 3).

• Definitions and Abbreviations

Three trial concentrations are prescribed[1] (abbreviated as C50, C15, C5):

C50	Solvent stock solution concentration: 50mg/mL Medium 1000-fold dilution concentration: 50µg/mL
C15	Solvent stock solution concentration: 15mg/mL Medium 1000-fold dilution concentration: 15µg/mL
C5	Solvent stock solution concentration: 5mg/mL Medium 1000-fold dilution concentration: 5µg/mL

[1] in event of insolubility at C5, further 3-fold dilutions may be considered (i.e., C1.5, C0.5, C0.15)

2. PROCEDURE

• Materials

- Balance (1 decimal place, mg) (with printer, optional).
- Vortex mixer.
- Ultrasonic water bath.
- Thermal water bath (37°C).
- Incubator (37°C, 5% CO_2).
- Centrifuge (fixed head for spin at 10,000g).
- Microscope, for observation of clarity/turbidity of medium droplets.
- Microplate (96-well flat-bottom, preferably seal wrapped individually to minimise interference from dust etc.) for observation of sample medium droplets.
- Clear glass vials (7mL size) with caps, for test item weighing and stock solutions.
- Test tubes (clear plastic, 15mL size) with caps, for medium dilutions and incubation.
- Test tubes (clear plastic, 50mL size) with caps, for medium handling.
- Centrifuge tubes (conical, clear plastic, 1.5mL size).
- Micro-pipettes (ranges: 2 – 20µL; 10 – 100µL; 100 – 1000µL).
- Pipette (5mL).
- Polyfoam floating tube racks (suitable for the 7mL glass vials and 50mL test tubes).
- FRM01-ASY15: Stock Solution Solubility: Reporting Form.
- FRM02-ASY15: Medium Dilution Stability: Reporting Form.

• Solubility in solvent (stock solutions)

Note: Soluble stock solutions should be used for medium stability testing within 24 hours.

1. Add DMSO solvent to a weighed amount of test item in a clear glass vial (7mL size) sufficient for visual inspection of solubility, starting at the upper concentration, C50 (minimum weight: 25mg, minimum volume: 0.5mL). Usually, several test items (e.g., 6) would be prepared together as a series.

   The chemical weight and solvent volume are calculated according to:

$$\text{Volume solvent [µL]} = \frac{\text{Weight chemical [mg]} * 1000}{\text{Concentration required [mg/mL]}}$$

2. Record the test item weight(s) and solvent volume(s) on respective Form(s) FRM01-ASY15.
3. Vortex mix

   3.1 Vortex mix for 1 minute (repeating, if appropriate) with visual check for dissolution against a suitable background illumination/contrast. A black background is recommended for effective observation of white suspension.

   3.2 If the test item is already soluble, indicate the vortex time and result (FRM01-ASY15) and set the solution aside for stability determination in medium (section 2.3).

4. Ultrasonic immersion (for crystal disaggregation)

   4.1 If not completely soluble after vortex mixing, immerse the vials in the ultrasonic water bath for 15 minutes, supported in a polyfoam floating tube rack.

   4.2. Repeat the vortex mix for 10 seconds to ensure homogeneity, with visual check for dissolution against a suitable background illumination/contrast.

   4.3. If the test item is now soluble, indicate the sonication time and result (FRM01-ASY15) and set the solution aside for stability determination in medium (section 2.3).

5. Thermal immersion (to accelerate kinetic delay)

   5.1. If not completely soluble after sonication, immerse the vials in the thermal water bath at 37°C for 30 minutes, retained in the same polyfoam rack. NB: Also allow the mixture to cool for at least 30 minutes, checking for possible recrystallization.

   5.2. Repeat the vortex mix for 10 seconds to ensure homogeneity, with visual check for dissolution against a suitable background illumination/contrast.

   5.3. If the test item is now soluble, indicate the warming time and result (FRM01-ASY15) and set the solution aside for stability determination in medium (section 2.3).

6. Solution standing (to ensure complete dissolution)

   6.1. If the test item is evidently soluble but visible traces of undissolved material remain, allow the solution to stand at room temperature for 1 hour (approx.) to complete the dissolution.

   6.2. Repeat the vortex mix for 10 seconds to ensure homogeneity, with visual check for dissolution against a suitable background illumination/contrast.

   6.3. If the test item is now satisfactorily soluble, indicate the standing time (approx.) and result (FRM01-ASY15) and set the solution aside for stability determination in medium (section 2.3).

7. Stock solution insoluble at C50

   7.1 If the test item persists as insoluble at the upper concentration C50, then solubility is attempted at the intermediate concentration C15, preparing fresh stock solution.

   7.2 Repeat the above procedure (steps 1 – 6) using re-weighed test item, appropriate for concentration C15, recording weight(s) and volume(s) on respective Form(s) FRM01-ASY15.

8. Stock solution insoluble at C15

   8.1 If the test item persists as insoluble at the intermediate concentration C15, then solubility is attempted at the lower concentration C5, preparing fresh stock solution.

   8.2 Repeat the above procedure (steps 1 – 6) using re-weighed test item, appropriate for concentration C5, recording weight(s) and volume(s) on respective Form(s) FRM01-ASY15.

9. If the test item is not soluble in DMSO, even at C5, repeat the above steps using water as alternative solvent (generally applicable for inorganic salts). In this case, compile another Form FRM01-ASY15.

10. FRM01-ASY15 provides raw data for stock solution preparation and solubility (including, if applicable, confirmation of the duration of vortex mixing, sonication, warming and standing). FRM01-ASY15 also allows space for conclusion (e.g., C50 / C15 / C5 soluble, or C5 insoluble) and comment.

• Stability in medium (with incubation)

Note 1: Stability in medium is determined for interim (500-fold) and full (1000-fold) dilutions, including two time points for the latter:

1) pre-incubation (time zero)
2) post-incubation (24 hours).

The observation at 500-fold dilution is for information only.

Effective solubility is the highest concentration (C50, C15, C5) at 1000-fold dilution where no precipitation is observed at either time point.

Note 2: Although test items are generally soluble in solvent at C50, precipitation frequently occurs on aliquot addition to medium (dilution 500/1000-fold, pre/post-incubation).

Therefore, medium stability testing of <u>all</u> soluble stock solution concentrations (in parallel) is standard procedure.

Thus:
- if stock solution is C50 soluble: medium stability tested at C50, C15 and C5;
- if stock solution is C15 soluble: medium stability tested at C15 and C5;
- if stock solution is C5 soluble only: medium stability tested at C5 only.

1. Warm a sufficient volume of medium (10mL per test item, per concentration, and a blank) from its refrigerated storage temperature to 37°C in the thermal water bath, ready for use.
2. Pipette 10µL stock solution (highest available soluble concentration, determined according to section 2.2) to 5mL medium in 15mL clear plastic tubes (500-fold dilution) and vortex mix for 10 seconds.
3. Dilute the remaining stock solution (by simple addition of solvent) to the next lower concentration(s):

   For C15 stock solution, dilute the C50 stock 3.33-fold.

   For C5 stock solution, dilute the C15 stock 3-fold.

Dilution volumes are calculated according to:

$$\text{Volume}^{2} \text{ of solvent to add } [\mu\text{L}] = \text{Final Volume } [\mu\text{L}] - \text{Initial Volume } [\mu\text{L}]$$

where:

$$\text{Final Volume } [\mu\text{L}] = \text{Dilution Factor} * \text{Initial Volume } [\mu\text{L}]$$

4. Record the total and added solvent volumes (FRM01-ASY15).
5. Repeat the 500-fold medium dilution (step 2) for the lower concentration(s) of stock solution, prepared consecutively (steps 3 – 4) as applicable.
6. Arrange an extra tube containing 10mL medium only, as reference blank.
7. Transfer 950µL aliquot samples (including a blank) to 1.5mL clear plastic conical vials, for solubility determination (500-fold medium dilution) assisted by centrifugation.
8. Pipette 50µL aliquot samples (including a blank) to a 96-well flat-bottom microplate, for solubility determination (500-fold medium dilution) assisted by microscopy.

Note 3: For convenience, it is recommended to prepare the samples for centrifugation and microscopy together, enabling simultaneous observation.

9. Centrifuge the vials (950µL aliquots) at 10,000g for 5 minutes at room temperature.
10. Examine the microplates (50µL aliquots) under the microscope (e.g., 20X magnification) with reference to the blank, checking for occurrence of undissolved material, recording the solubility observations on respective Forms FRM02-ASY15.
11. Check the centrifuge vials for occurrence of deposited precipitate (visible as a small speck or pellet) indicative of insolubility, alongside the blank for comparison, recording the solubility observations on respective Forms FRM02-ASY15. For effective observation of a white deposit, a contrasting black background is recommended.

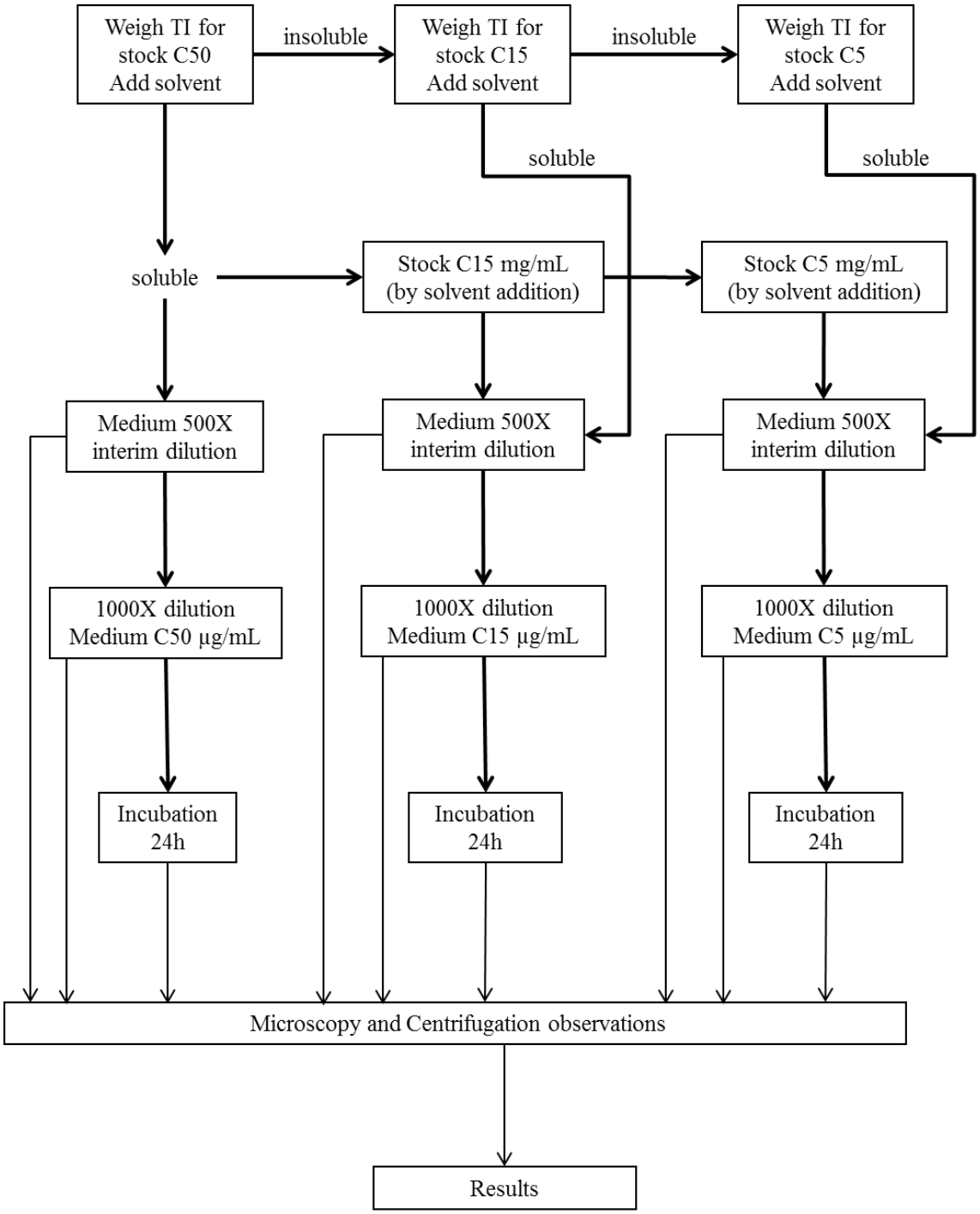
---

[2] rounded to nearest integer (µL).

12. Complete the dilution (1000-fold) by doubling the medium volume (4mL) remaining in the tube and vortex mix again for 10 seconds.
13. Repeat steps 7 – 11 for pre-incubation solubility determination (1000-fold medium dilution, time zero).
14. Incubate (37°C, 5% CO_2) the medium dilutions (with tube caps loosened, blank included) for 24 hours (approx.) equivalent to the conditions and duration of assay cell culture/test item exposure,
15. Repeat the vortex mixing (10 seconds) and centrifugation (1000µL aliquots) with check for deposited precipitate, as post-incubation solubility determination (24 hours) recording the results (FRM02-ASY15).
16. Repeat the microscope examination (50µL aliquot) as supplementary final solubility determination (24 hours) recording the results (FRM02-ASY15).
17. From the two observations, pre- and post-incubation, note the effective solubility result (FRM02-ASY15). Effective solubility is the highest concentration (C50, C15, C5) at 1000-fold dilution where no precipitation is observed at either time point.
18. FRM02-ASY15 provides raw data for medium dilution stability with incubation. FRM02-ASY15 also allows space for comment / overall conclusion (e.g., explicit soluble concentration).

3. ANNEX

Flowchart: solubility determination by visual inspection



Annex 2: Update of the validation project plan

Date of discussion with VMG: 27-02-2017

1. The experimental organisation of the studies was modified as indicated below, still following the modular approach as described by Hartung et al. Module 5 suggests assessing predictive capacity (PC) of the test method in one lab only instead of in all 3 laboratories. Such modification would benefit the time line of the validation study. BLR would be assessed on a total number of 20 test chemicals.

As a consequence, the studies 1, 2 (10 coded chemicals) and 3 (36 coded chemicals) will be renamed as study 1 (no change), Study 2A (10 coded chemicals) and 2B (10 coded chemicals), Study 3 (26 coded chemicals).

Study 1 (6 chemicals, transfer): to be carried out by all 3 laboratories ENVIGO, CitoxLAB and RISE.

Study 2A (10 coded chemicals, WLR, BLR and PC):
to be carried out by all 3 laboratories ENVIGO, CitoxLAB and RISE.

Study 2B (10 coded chemicals, WLR, BLR and PC):
to be carried out by all 3 laboratories ENVIGO, CitoxLAB and RISE.

Study 3 (26 coded chemicals, WLR, PC):
to be carried out by RISE

2. Solubility testing

The testing shall not be restricted to the chemicals of study 2A but shall be carried out for all test chemicals in this validation study.

Annex 3: Modification to the validation project plan

Date of discussion with VMG: 21-12-2016 and 24-03-2017

1. Modification of the SOP

1.1. Introduction of the criterion for the specificity control R^2

- True competitive antagonist: the coefficient of determination (R^2) is less or equal to 0.9 for the linear regression of relative induction of the test item's specificity control (S_c) versus the relative induction (Y_c).
- False competitive antagonist: the coefficient of determination (R^2) is greater than 0.9 for the linear regression of the relative induction of the test item's specificity control (S_c) versus the relative induction (Y_c).

1.2. Introduction of an acceptance criterion for the reference item flutamide when applying the specificity control. Table 4 becomes as follows:

Table 4: Acceptance criteria in the antagonist assay

No	Acceptance criterium	Value
	Reference chemical Flutamide	
1	Curve fitting	Sigmoidal
2	IC_{50} range	1.10^{-7} – 1.10^{-6} M
3	CV of estimated log(IC_{50})	< 3%
4	Inhibition factor	> 10
5	Z-factor	> 0.5
	Positive control	
6	Relative induction for Linuron	< 60%
	Negative control	
7	Relative induction for Levonorgestrel	> 85%
8(*)	R^2 for the linear regression of Sc on Yc for Flutamide	≤ 0.7

(*) To be applied for assessment of the specificity response (Sc) of Flutamide

1.3. Inclusion of the classifier in the SOP

As decided at the onset of the validation study, a classifier was introduced in the last version of the SOP to be used by the laboratories. Hence a set of data would be generated where the laboratories could apply the classifier. The specificity control criterium R^2 was included in the antagonist part of the classifier. For both agonist and antagonist parts, the option Inconclusive was included. The classifier was included in an updated version of the SOP (SOP V06).

1.4. Progress of the validation study

Given that the 3 laboratories have a different progress in the study, SOP V06 will be applied as follows:

Study 2B 10 coded chemicals ENVIGO

Study 2A and 2B 20 coded chemicals CitoxLAB

Study 3 46 coded chemicals RISE

SOP V06 Section 2.5.4 Classification for agonist and antagonist properties

Agonism

For each run, a test item is considered

A. **Positive** when the relative induction (Y_c) of the test item is equal or exceeds 10% (REF RPC_{10}) for two or more consecutive concentrations.
B. **Negative** when the relative induction (Y_c) of the test item does not exceed 10% (REF RPC_{10}) for any concentration.
C. **Inconclusive** in all other cases.

Antagonism

For each run, a test item is considered

A. **Positive** when the following two conditions are met:
- the relative induction (Y_c) of the test item is less or equal to 80% (REF RPC80) for two or more consecutive concentrations and
- the coefficient of determination (R^2) is less or equal to 0.9 for the linear regression of relative induction of the test item's specificity control (S_c) versus the relative induction (Y_c).

B. **Negative**

Either
- when the relative induction (Y_c) of the test item is greater than 80% (REF RPC80) at all concentrations;

or
- when the relative induction (Y_c) of the test item is less or equal to 80% (REF RPC80) for at least 2 consecutive concentrations and the coefficient of determination (R^2) is greater than 0.9 for the linear regression of the relative induction of the test item's specificity control (S_c) versus the relative induction (Y_c).

C. **Inconclusive** in all other cases.

2. Update of the test chemicals for the validation study.

The original list was updated with few chemicals as discussed and suggested by VMG. This included the addition of few chemicals with reported agonist behaviour from the recently published ICCVAM A-reference list: Cyproterone acetate, Methyltrienolone (R1881), Norethinodrone, Norethinodrone acetate, 19-Nortestosterone (Nandrolone). Few other chemicals have been omitted from the original list.

3. Selection of the 36 coded test chemicals (for studies 2B and 3)

Seq. #	Chemical name
11	Sodium azide
12	Diethylhexyl phthalate
13	Methyldihydrotestosterone
14	Vinclozolin
15	Prochloraz
16	Fluoxymesterone
17	17β-Estradiol
18	Benzylbutyl phthalate
19	Propylthiouracil
20	Hydroxyflutamide
21	Levonorgestrel
22	Cyproterone acetate
23	2-tert-Butylanthraquinone
24	Arochlor1254
25	Nandrolone
26	o,p'-DDT
27	Phenolphthalin
28	2,4,5-T
29	Methyltrienolone (R1881)
30	Actinomycin D
31	Diethylstilbestrol
32	L-Thyroxine
33	Haloperidol
34	Norethindrone acetate
35	Pimozide
36	Progesterone
37	Linuron
38	Methyltestosterone
39	2-sec-Butylphenol
40	Corticosterone
41	Ketoconazole
42	Finasteride
43	Fulvestrant
44	Cycloheximide
45	Norethindrone
46	Mifepristone

Annex 4: Modification to the validation project plan

Date of discussion with VMG: 26-06-2018

One of the laboratories is experiencing technical issues on a regular basis. This challenges the timely finalisation of the validation study. In order to assess BLR, a 4th laboratory is added to the validation study to produce a set of data on 20 coded test chemicals.

BioDetection Systems BV (BDS), Science Park 406, 1098 XH Amsterdam, The Netherlands	
Laboratory manager: Harrie Besselink Study director: Harrie Besselink Study personnel: Matthijs Naderman	Email: Harrie.Besselink@bds.nl

Annex 13.3

List and properties of reference, control and test chemicals used in the AR-CALUX® validation study

1. List of the 53 chemicals used in the AR-CALUX® validation study

2. Source of the chemicals that led to the final list of a total of 53 chemicals

3. Compilation of properties for 53 chemicals used in the AR-CALUX® validation study

4. Selection strategy of 45 chemicals for the AR-CALUX® validation study (Report by J. Burton) including mapping of the AR-CALUX® chemicals versus the REACH chemicals.

5. Compilation of publicly available data on ARTA's and the AR-pathway model, for 46 coded test chemicals to be used in the AR-CALUX® validation study

1. List of the 53 chemicals used in the AR-CALUX® validation study

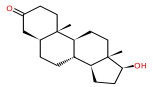
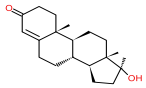
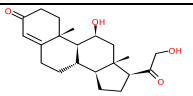
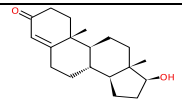
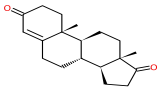
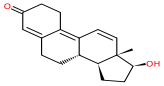
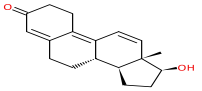
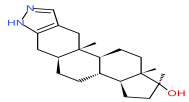
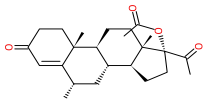
Chemical name	CASNR	Study 1 (Transfer)	Study 2A	Study 2B	Study 3
5α-Dihydrotestosterone	521-18-6	REF Agonist			
Flutamide	13311-84-7	REF Antagonist			
Corticosterone	50-22-6	NC Agonist			x
Levonorgestrel	797-63-7	NC Antagonist			x
Methyl testosterone	58-18-4	PC Agonist			x
Linuron	330-55-2	PC Antagonist			x
Testosterone	58-22-0	x			
4-Androstenedione	63-05-8	x			
Procymidone	32809-16-8	x			
p,p'-Methoxychlor	72-43-5	x			
Di-n-butyl phthalate	84-74-2	x			
Sodium azide	26628-22-8	x		x	
17β-Trenbolone	10161-33-8		x		
Medroxyprogesterone acetate	71-58-9		x		
Stanozolol	10418-03-8		x		
Spironolactone	52-01-7		x		X
Bisphenol A	80-05-7		x		
17α-Ethinyl estradiol	57-63-6		x		
Bicalutamide	90357-06-5		x		
Tamoxifen	10540-29-1		x		
Atrazine	1912-24-9		x		
Disulfiram	97-77-8		x		
Fluoxymestrone	76-43-7			x	
Methyldihydrotestosterone	521-11-9			x	
17β-Estradiol	50-28-2			x	
Hydroxyflutamide	52806-53-8			x	
Vinclozolin	50471-44-8			x	
Prochloraz	67747-09-5			x	
Propylthiouracil	51-52-5			x	
Diethylhexyl phthalate	117-81-7			x	
Butylbenzyl phthalate	85-68-7			x	
2-tert-butylanthraquinone	84-47-9				x
Arochlor1254	11097-69-1				x
Progesterone	57-83-0				x
Mifepristone	84371-65-3				x
Cyproterone acetate	427-51-0				x
Methyltrienolone (R1881)	965-93-5				x
Norethinodrone	68-22-4				x
Norethinodrone acetate	51-98-9				x
19-Nortestosterone	434-22-0				x
ICI 182,780 (Fulvestrant)	129453-61-8				x
Pimozide	2062-78-4				x
Actinomycin D	50-76-0				x
Diethylstilbestrol	56-53-1				x
Ketoconazole	65277-42-1				x
Cycloheximide	66-81-9				x
o,p'-DDT	789-02-6				x
Finasteride	98319-26-7				x
L-Thyroxine	51-48-9				x
Haloperidol	52-86-8				x
Phenolphthalin	81-90-3				x
2-sec-Butylphenol	89-72-5				x
2,4,5-Trichlorophenoxyacetic acid	93-76-5				x

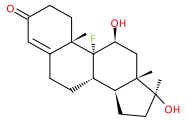
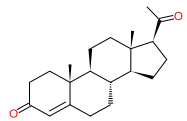
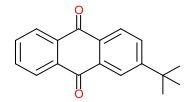
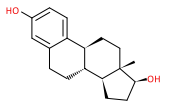
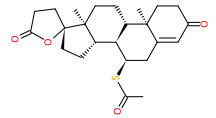
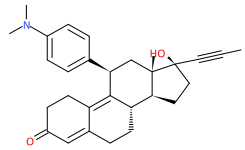
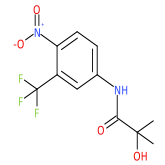
2. Source of the test chemicals that lead to the final list of a total of 53 chemicals

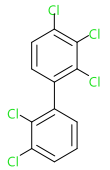
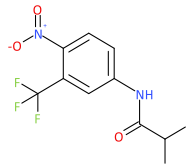
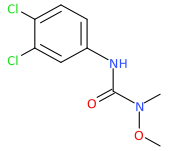
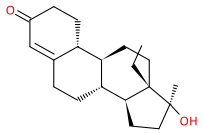
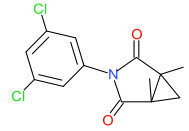
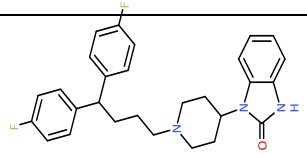
Source of test items	Items added
Publications, ICCVAM list (2003), 2 Tox21 assays, consultations with VMG-NA and ICCATM.	A first list of 83 chemicals was prepared from which a selection was made to lead to a list of 45 chemicals (see Report J.Burton in this Annex). From this list of 45 some were deleted to encompass the ones that were added later (see below).
ARTA Japan validation study	Hydroxyflutamide Diethylhexyl phthalate Butylbenzyl phthalate 17α-Ethinyl estradiol
ARTA Korea validation study	Bicalutamide Fluoxymestrone, 2-tert-butylanthraquinone, Arochlor1254: included as they were initially proposed for the Korean validation study but during the course of the validation study they were not retained in the Korean ARTA final list of testing chemicals
ICCVAM AR reference list	Cyproterone acetate Methyltrienolone (R1881) Norethinodrone Norethinodrone acetate 19-Nortestosterone (Nandrolone)
VMG proposal	Disulfiram

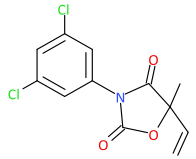
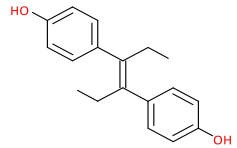
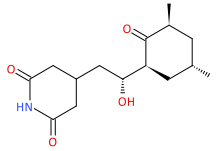
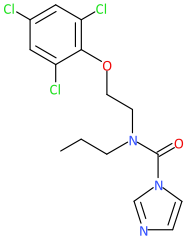
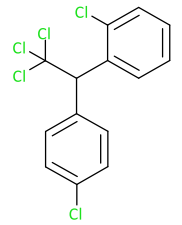
3. Compilation of properties for 53 chemicals used in the AR-CALUX® validation study

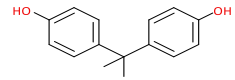
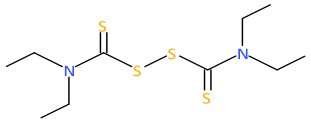
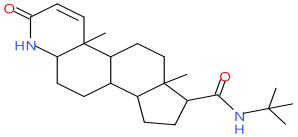
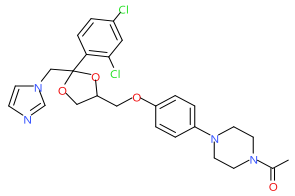
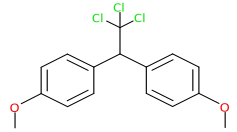
See below

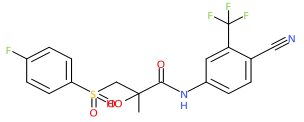
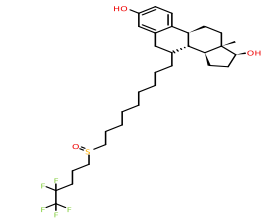
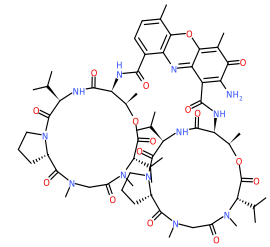
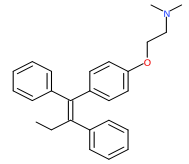
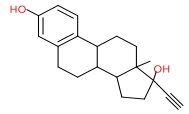
Nr.	Name	CAS Nr.	Use	State	Source	Toolbox 4.1 Organic functional groups, Norbert Haider (checkmol)	Toolbox 4.1 OECD HPV Chemical Categories	Toolbox 4.1 US-EPA New Chemical Categories	Structure
1	5α-Dihydrotestosterone	521-18-6	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Carbonyl compound|Hydroxy compound|Ketone|Secondary alcohol	Not categorized	Neutral Organics	
2	Methyl testosterone	58-18-4	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Hydroxy compound|Tertiary alcohol	Not categorized	Not categorized	
3	Corticosterone	50-22-6	Biomedical research	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Carbonyl compound|Hydroxy compound|Ketone|Secondary alcohol	Not categorized	Neutral Organics	
4	Testosterone	58-22-0	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alcohol|Hydroxy compound|Secondary alcohol	Not categorized	Not categorized	
5	4-Androstenedione	63-05-8	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Carbonyl compound|Ketone	Not categorized	Neutral Organics	
6	Methyldihydro testosterone	521-11-9	Pharmaceutical ingredient (obsolete/restricted)	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Carbonyl compound|Hydroxy compound|Ketone|Tertiary alcohol	Not categorized	Neutral Organics	
7	17β-Trenbolone	10161-33-8	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Hydroxy compound|Secondary alcohol	Not categorized	Not categorized	
8	Stanozolol	10418-03-8	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Aromatic compound|Heterocyclic compound|Hydroxy compound|Tertiary alcohol	Not categorized	Not categorized	
9	Medroxyprogesterone acetate	71-58-9	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Carbonyl compound|Carboxylic acid derivative|Carboxylic acid ester|Ketone	Not categorized	Esters (Acute toxicity)	

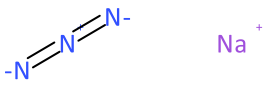
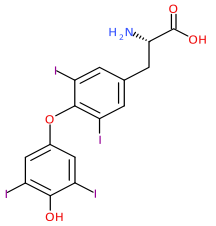
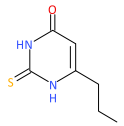
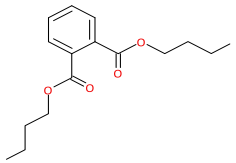
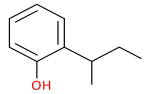
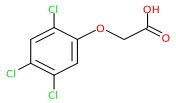
10	Fluoxymestrone	76-43-7	Pharmaceutical ingredient	solid (powder) white, crystalline	LGC Standards (UK/Italy)	Alcohol|Alkyl fluoride|Alkyl halide|Halogen derivative|Hydroxy compound|Secondary alcohol|Tertiary alcohol	Not categorized	Not categorized	
11	Progesterone	57-83-0	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Carbonyl compound|Ketone	Not categorized	Neutral Organics	
12	2-tert-butylanthraquinone	84-47-9	Industrial intermediate (dyes/pigments)	solid (powder) white, crystalline	TCI Europe (Belgium)	Aromatic compound|Carbonyl compound|Ketone	Not categorized	Neutral Organics	
13	17β-Estradiol	50-28-2	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alcohol|Aromatic compound|Hydroxy compound|Phenol|Secondary alcohol	Not categorized	Phenols (Acute toxicity)	
14	Spironolactone	52-01-7	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Carboxylic acid derivative|Carboxylic acid ester|Heterocyclic compound	Not categorized	Not categorized	
15	Mifepristone	84371-65-3	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Alkyne|Amine|Aromatic compound|Hydroxy compound|Tertiary amine|Tertiary mixed amine	Not categorized	Not categorized	
16	Hydroxyflutamide	52806-53-8	Flutamide metabolite	solid (powder) white, crystalline	Chemos (Germany) and Carbosynth (UK)	Alcohol|Alkyl fluoride|Alkyl halide|Anion|Aromatic compound|Carboxylic acid amide|Carboxylic acid derivative|Carboxylic acid sec. amide|Cation|Halogen derivative|Hydroxy compound|Nitro compound	Not categorized	Not categorized	

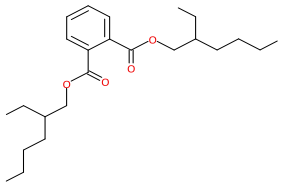
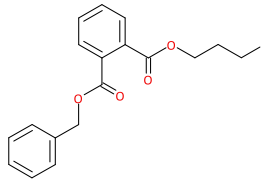
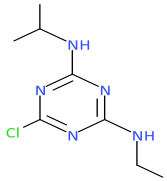
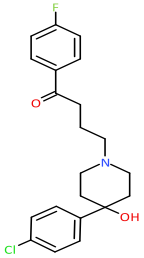
17	Arochlor1254	11097-69-1	PCB mixture (lubricant/coolant)	Liquid (oily) colourless, viscous	Sigma/Merck (Italy/Germany)	Aromatic compound|Aryl chloride|Aryl halide|Halogen derivative	Not categorized	Persistent, Bioaccumulative and Toxic (PBT) Chemicals	
18	Flutamide	13311-84-7	Pharmaceutical ingredient	solid (powder) pale yellow, crystalline	TCI Europe (Belgium)	Alkyl fluoride|Alkyl halide|Anion|Aromatic compound|Carboxylic acid amide|Carboxylic acid derivative|Carboxylic acid sec. amide|Cation|Halogen derivative|Nitro compound	Not categorized	Not categorized	
19	Linuron	330-55-2	Herbicide	solid (powder) white, crystalline	Chemos (Germany)	Aromatic compound|Aryl chloride|Aryl halide|Carbonic acid derivative|CO2 derivative (general)|Halogen derivative	Not categorized	Neutral Organics	
20	Levonorgestrel	797-63-7	Pharmaceutical ingredient	solid (powder) white, crystalline	Chemos (Germany)	Alcohol|Hydroxy compound|Tertiary alcohol	Not categorized	Not categorized	
21	Procymidone	32809-16-8	Fungicide (agricultural)	solid (powder) white, crystalline	Chemos (Germany)	Aromatic compound|Aryl chloride|Aryl halide|Carboxylic acid derivative|Halogen derivative|Heterocyclic compound	Not categorized	Imides (Acute toxicity)	
22	Pimozide	2062-78-4	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	(N/A)	(N/A)	(N/A)	

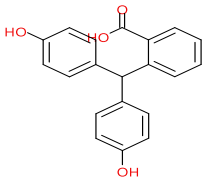
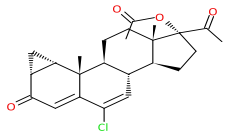
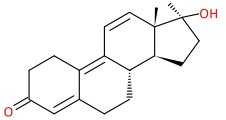
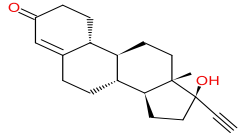
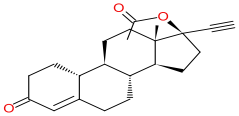
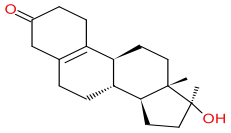
23	Vinclozolin	50471-44-8	Fungicide (agricultural)	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alkene|Aromatic compound|Aryl chloride|Aryl halide|Carbonic acid derivative|Carboxylic acid derivative|CO2 derivative (general)|Halogen derivative|Heterocyclic compound	Not categorized	Neutral Organics	
24	Diethylstilbestrol	56-53-1	Pharmaceutical ingredient (obsolete/research)	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alkene|Aromatic compound|Hydroxy compound|Phenol	Not categorized	Phenols (Acute toxicity)	
25	Cycloheximide	66-81-9	Biomedical research (antibacterial and antifungal)	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alcohol|Carbonyl compound|Carboxylic acid derivative|Carboxylic acid imide|Carboxylic acid unsubst. imide|Heterocyclic compound|Hydroxy compound|Ketone|Secondary alcohol	Not categorized	Imides (Acute toxicity)	
26	Prochloraz	67747-09-5	Fungicide (agricultural)	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alkylarylether|Aromatic compound|Aryl chloride|Aryl halide|Carbonic acid derivative|CO2 derivative (general)|Ether|Halogen derivative|Heterocyclic compound	Tertiary Amines	Neutral Organics	
27	o,p'-DDT	789-02-6	Insecticide (agicultural) (obsolete)	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alkyl chloride|Alkyl halide|Aromatic compound|Aryl chloride|Aryl halide|Halogen derivative	Not categorized	Neutral Organics	

28	Bisphenol A	80-05-7	Industrial intermediate (plastics manufacture)	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Aromatic compound|Hydroxy compound|Phenol	Not categorized	Phenols (Acute toxicity)	
29	Disulfiram	97-77-8	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	CO2 derivative (general)|Thiocarbonic acid derivative	Tertiary Amines	Dithiocarbamates (Acute toxicity)	
30	Ketoconazole	65277-42-1	Pharmaceutical ingredient (antifungal)	solid (powder) white, crystalline	TCI Europe (Belgium)	Alkylarylether|Amine|Aromatic compound|Aryl chloride|Aryl halide|Carboxylic acid amide|Carboxylic acid derivative|Carboxylic acid tert. amide|Ether|Halogen derivative|Heterocyclic compound|Tertiary amine|Tertiary mixed amine	Tertiary Amines	Neutral Organics	
31	Finasteride	98319-26-7	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Carboxylic acid amide|Carboxylic acid derivative|Carboxylic acid sec. amide|Heterocyclic compound	Not categorized	Not categorized	
32	p,p'-Methoxychlor	72-43-5	Insecticide (agicultural) (obsolete)	solid (powder) white, crystalline	TCI Europe (Belgium)	Alkyl chloride|Alkyl halide|Alkylarylether|Aromatic compound|Ether|Halogen derivative	Not categorized	Not categorized	

33	Bicalutamide	90357-06-5	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Alkyl fluoride|Alkyl halide|Aromatic compound|Aryl fluoride|Aryl halide|Carboxylic acid amide|Carboxylic acid derivative|Carboxylic acid sec. amide|Halogen derivative|Hydroxy compound|Nitrile|Sulfone	Not categorized	Not categorized	
34	ICI 182,780 (Fulvestrant)	129453-61-8	Pharmaceutical ingredient	solid (powder) white, crystalline	Chemos (Germany)	Alcohol|Alkyl fluoride|Alkyl halide|Anion|Aromatic compound|Cation|Halogen derivative|Hydroxy compound|Phenol|Secondary alcohol|Sufoxide	Not categorized	Not categorized	
35	Actinomycin D	50-76-0	Pharmaceutical ingredient (antibiotic)	solid (powder) dark red, crystalline	Chemos (Germany) and Carbosynth (UK)	(N/A)	(N/A)	(N/A)	
36	Tamoxifen	10540-29-1	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alkene|Alkylarylether|Amine|Aromatic compound|Ether|Tertiary aliphatic amine|Tertiary amine	Tertiary Amines	Aliphatic Amines	
37	17α-Ethinyl estradiol	57-63-6	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alcohol|Alkyne|Aromatic compound|Hydroxy compound|Phenol	Not categorized	Phenols (Acute toxicity)	

38	Sodium azide	26628-22-8	Vehicle manufacture Airbag detonator and gas (nitronen) source	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Anion|Cation	Not categorized	Undefined	
39	L-Thyroxine	51-48-9	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alpha-aminoacid|Amine|Aromatic compound|Aryl halide|Aryl iodide|Carboxylic acid|Carboxylic acid derivative|Diarylether|Ether|Halogen derivative|Hydroxy compound|Phenol|Primary aliphatic amine|Primary amine	Not categorized	Aliphatic Amines|Phenols (Acute toxicity)	
40	Propylthiouracil	51-52-5	Pharmaceutical ingredient	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	(N/A)	(N/A)	(N/A)	
41	Dibutyl phthalate	84-74-2	Industrial intermediate (plasticizer)	Liquid (oily) colourless, viscous	Sigma/Merck (Italy/Germany)	Aromatic compound|Carboxylic acid derivative|Carboxylic acid ester	Not categorized	Esters (Acute toxicity)	
42	2-sec-Butylphenol	89-72-5	Industrial intermediate (organic syntheses)	Liquid colourless	Sigma/Merck (Italy/Germany)	Aromatic compound|Hydroxy compound|Phenol	Not categorized	Phenols (Acute toxicity)	
43	2,4,5-Trichlorophenoxyacetic acid	93-76-5	Herbicide (agricultural)	solid (powder) white, crystalline	Sigma/Merck (Italy/Germany)	Alkylarylether|Aromatic compound|Aryl chloride|Aryl halide|Carboxylic acid|Carboxylic acid derivative|Ether|Halogen derivative	Not categorized	Neutral Organics	

44	Diethylhexyl phthalate	117-81-7	Industrial intermediate (plasticizer)	Liquid (oily) colourless, viscous	Sigma/Merck (Italy/Germany)	Aromatic compound|Carboxylic acid derivative|Carboxylic acid ester	High molecular weight phthalate esters	Not categorized	
45	Butylbenzyl phthalate	85-68-7	Industrial intermediate (plasticizer)	Liquid (oily) colourless, viscous	Sigma/Merck (Italy/Germany)	Aromatic compound|Carboxylic acid derivative|Carboxylic acid ester	Not categorized	Esters (Acute toxicity)	
46	Atrazine	1912-24-9	Herbicide (agricultural) (EU: obsolete)	solid (powder) white, crystalline	TCI Europe (Belgium)	Amine|Aromatic compound|Aryl chloride|Aryl halide|Halogen derivative|Heterocyclic compound|Secondary amine|Secondary mixed amine (aryl, alkyl)	Not categorized	Substituted Triazines (Acute toxicity)	
47	Haloperidol	52-86-8	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Amine|Aromatic compound|Aryl chloride|Aryl fluoride|Aryl halide|Carbonyl compound|Halogen derivative|Heterocyclic compound|Hydroxy compound|Ketone|Tertiary alcohol|Tertiary aliphatic amine|Tertiary amine	Tertiary Amines	Aliphatic Amines	

48	Phenolphthalin	81-90-3	Laboratory chemical (indicator dye)	solid (powder) white, crystalline	TCI Europe (Belgium)	Aromatic compound|Carboxylic acid|Carboxylic acid derivative|Hydroxy compound|Phenol	Not categorized	Phenols (Acute toxicity)	
49	Cyproterone acetate	427-51-0	Pharmaceutical ingredient	solid (powder) white, crystalline	Carbosynth (UK)	Carbonyl compound|Carboxylic acid derivative|Carboxylic acid ester|Halogen derivative|Ketone	Not categorized	Esters (Acute toxicity)	
50	Methyltrienolone (R1881)	965-93-5	Pharmaceutical ingredient (obsolete/research)	solid (powder) white, crystalline	Chemos (Germany)	Alcohol|Hydroxy compound|Tertiary alcohol	Not categorized	Not categorized	
51	Norethinodrone	68-22-4	Pharmaceutical ingredient	solid (powder) white, crystalline	Carbosynth (UK)	Alcohol|Alkyne|Hydroxy compound	Not categorized	Not categorized	
52	Norethinodrone acetate	51-98-9	Pharmaceutical ingredient	solid (powder) white, crystalline	Carbosynth (UK)	Alkyne|Carboxylic acid derivative|Carboxylic acid ester	Not categorized	Esters (Acute toxicity)	
53	19-Nortestosterone (Nandrolone)	434-22-0	Pharmaceutical ingredient	solid (powder) white, crystalline	TCI Europe (Belgium)	Alcohol|Alkene|Carbonyl compound|Hydroxy compound|Ketone|Tertiary alcohol	Not categorized	Neutral Organics	

4. Selection strategy of 45 chemicals for the AR-CALUX validation study (Report by J. Burton) and mapping of the 53 chemicals versus the REACH chemicals

Chemical selection for AR CALUX validation study

Julien Burton - 18 Nov.2014

Introduction

This document aims at presenting the chemical selection strategy followed for the ARCALUX validation study.

The strategy relied on a knowledge-driven weight of evidence approach, using previously generated results to select chemicals with consistent behaviours. The collected data on 83 chemicals was compiled from expert opinions, AR in vitro assays, ARCALUX assays and high-throughput screening.

The behaviour (agonist, antagonist, negative) of the chemicals was assessed via a scoring methods and the actual selection was performed choosing compounds with robust observations while covering a satisfying diversity.

Finally, we proposed a scheme for collecting further data on chemicals with the objective of defining the applicability domain of the validation study.

Data sources

The chemical selection strategy relied on data collection from several ARTA experiments, in-house testings, literature findings and expert judgements coming from several reliable sources:

Literature:

- ***ICCVAM recommendations***: ICCVAM Evaluation of In Vitro Test Methods for Detecting Potential Endocrine Disruptors: Estrogen and Androgen Receptor Binding and Transcriptional Activation Assays (Addendum, Tables 6-1 and 6-2)
- ***AR Ecoscreen*** assay results: Evaluation of rapid in vitro androgen receptor transcriptional activation assay using AR-Ecoscreen cells (N.Araki, K.Ohno, M.Takeyoshi, M.Iida, *Toxicology in Vitro* (**2005**), *19*, 335-352)

- ***AR CALUX*** assay results: Optimization and prevalidation of the in vitro AR CALUX method to test androgenic and antiandrogenic activity compounds (B.van der Burg, R.Winter, H.Man, C.Vangenechten, P.Berckmans, M.Weimer, H.Witters, S.van der Linden, *Reproductive Toxicology* (**2010**), *30*, 18-24)
- ***PALM*** assay results: Screening for (anti)androgenic properties using a standard operation protocol based on the human stably transfected androgen sensitive PALM cell line. First step towards validation (A.Freyberger, H.Witters, M.Weimer, W.Lofink, P.Berckmans, H.-J. Ahr, *Reproductive Toxicology* (**2012**), *30*, 9-17)
  Tox21 assays (qualitative data provided by collaborators):

- ***Tox21 AR-BLA assay***
- ***Tox21 AR-luc assay***
  In-house HTS results on CALUX (by Jean-Michel Gineste and Roman Liska):

- Runs (biological replicates) **CALUX 01**, **02**, and **03** for agonism
- Runs (biological replicates) **CALUX 04**, **05**, and **06** for antagonism

QSAR

- **PASS AR agonist model** (http://www.pharmaexpert.ru/passonline/). Model relying on structural similarity toward a reference sample of AR agonists and non-agonists. Positive classification is achieved when the probability of being an agonist is higher than the probability of being inactive.

Table 1 shows a comparison between the different experimental sources, including the cell line, biological construct, number of replicates, concentration range and classification criteria.

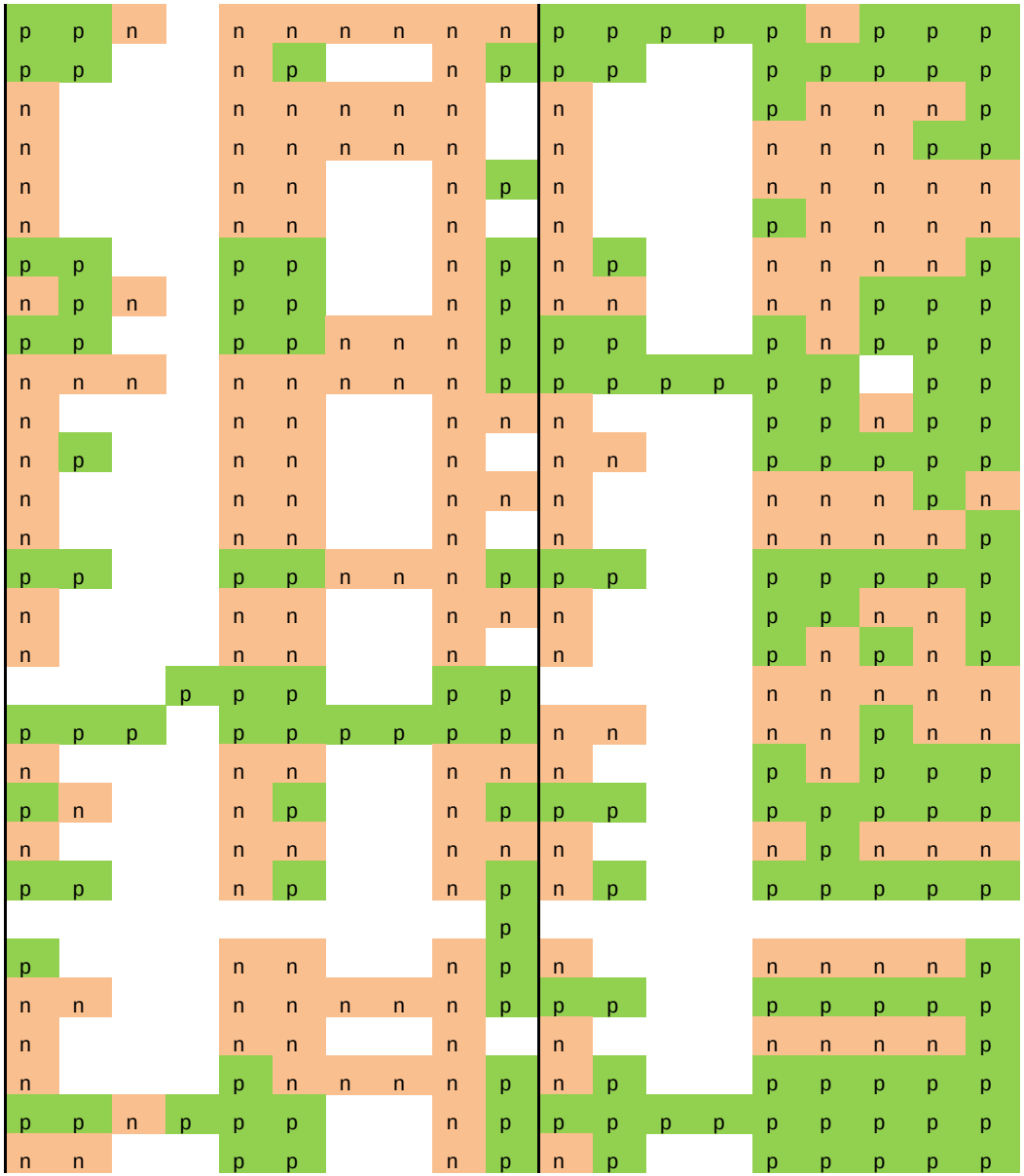
Experimental source	Cell line	Construct	Replicates (technical)	Concentration range (M)	POS/NEG classification
AR Ecoscreen	Chinese Hamster Ovary	cDNA human androgen receptor / firefly luciferase gene preceded by 4 AREs of prostate C3 gene heat shock protein / renilla (frog) luciferase gene constitutively expressed	3	10^{-9} – 10^{-4}	From table 5. Agonist: fold induction (FI) > 1.7 → Positive Antagonist: DHT-induced relative luciferase activity (RLA) < 80% + relative cell viability >80% at the same concentration tested → Positive
AR CALUX	U2-OS human osteosarcoma	full length cDNA human androgen receptor / luciferase gene preceded by 3 ARREs coupled to a minimal adenovirus TATA promoter	Min. 3 (up to 8)	Chemical-dependant (10^{-10} – 10^{-5}?)	2 consecutive values below/above a given threshold
PALM	human prostate cancer	human prostate cancer cell line PC-3 stably transfected with human AR and a Luciferase gene under the control of mouse mammary tumour virus (MMTV)	3	Chemical-dependant (10^{-9} – 10^{-5}?)	Agonist: a sigmoid concentration response should be observed for MDHT. A maximally stimulating concentration of MDHT (100nM or above) should result in an induction factor of ≥4.0, whereas the EC50 of MDHT should be in the range 0.1–1 nM. Antagonist: 0.2nM MDHT should result in an induction factor of ≥2.5, flutamide should concentration-dependently reduce the agonistic effect of 0.2nM MDHT, and at 10_Mflutamide should reduce the net increase of luciferase activity (brought about by 0.2nM MDHT) by ≥70%.
Tox21 AR-BLA	Human Embryonic Kidney HEK293T	stably transfected with human ER-α ligand-binding domain and a β-lactamase reporter gene	3	10^{-8} – 10^{-4}	As provided by collaborators
Tox21 AR-luc	Human breast carcinoma MDA-kb2	human breast carcinoma cell line stably transfected with luciferase reporter gene under control of MMTV promoter containing response elements for both androgen receptor (AR) and glucocorticoid receptor (GR).	3	10^{-8} – 10^{-4}	As provided by collaborators
In-house HTS	See AR CALUX	See AR CALUX	1	Chemical-dependant 10^{-13} – 10^{-5}	Same ARCALUX

Table 1. Comparison of the experimental sources supporting the chemical selection

Data collection

Using the above-mentioned sources, supplemented with internal discussion with collaborators, data could be gathered for 83 chemicals. Table 2 summarises the collected data with chemical identifiers (name and CAS number) and the associated classification (positive or negative, following the internal protocol of each study/document) for both agonism and antagonism tests.

Name	*CAS*	AGONISM										ANTAGONISM								
		ICCVAM	*Ecoscreen*	*CALUX*	*PALM*	*Tox21 AR-BLA*	*Tox21 AR-luc*	*HTS CALUX01*	*HTS CALUX02*	*HTS CALUX03*	*AR agonist PASS*	*ICCVAM*	*Ecoscreen*	*CALUX*	*PALM*	*Tox21 AR-BLA*	*Tox21 AR-luc*	*HTS CALUX04*	*HTS CALUX05*	*HTS CALUX06*
17β-Trenbolone	10161-33-8	p				p	p			p	p	n				p	n	n	n	n
Stanozolol	10418-03-8									p	p							n	n	n
p-n-Nonylphenol	104-40-5	n	n			n	n			n	p	n	p			p	n	n	p	p
Tamoxifen	10540-29-1	n				n	n			n	n	n				p	n	n	n	n
4,4'-(1H-imidazol-1-ylmethanediyl)dibenzonitrile	112808-99-8	n										n								
Diethylhexyl phthalate	117-81-7	n	n			n	n	n	n	n	n	n	n			p	n	n	n	n
Ethyl paraben	120-47-8	n				n	n			n	n	n				n	n	n	n	n
Anastrazole	120511-73-1	n				n	n			n	n	n				n	n	n	n	n
ICI 182,780	129453-61-8	n				n	n	n	n	n	p	n				p	p	p	p	p
Flutamide	13311-84-7	n	n	n		n	n	n	n	n	p	p	p	p	p	p	p	p	p	p
4-tert-Octylphenol	140-66-9	n	n			n	n				p	p	p			p	p			
Zearalenone	17924-92-4	n				n	n			n	p	n				p	p	n	p	p
Atrazine	1912-24-9	n	n			n	n	n	n	n	p	n	n			n	n	n	n	n
Pimozide	2062-78-4	n				n	n			n		n				p	p	p	p	p
Fluoranthene	206-44-0	n	p			n	n			n	p	p	p			n	n	n	p	p
Sodium azide	26628-22-8	n	n			n	n	n	n	n		n	n			n	p	n	n	n
Procymidone	32809-16-8	n	n			n	n	n	n	n	p	p	p			p	p		p	p

Linuron	330-55-2	
Cyproterone acetate	427-51-0	
Genistein	446-72-0	
Coumestrol	479-13-0	
Morin	480-16-0	
Daidzein	486-66-8	
Dexamethasone	50-02-2	
Corticosterone	50-22-6	
17β-Estradiol	50-28-2	
Vinclozolin	50471-44-8	
Reserpine	50-55-5	
Actinomycin D	50-76-0	
L-Thyroxine	51-48-9	
Propylthiouracil	51-52-5	
Spironolactone	52-01-7	
Kaempferol	520-18-3	
Apigenin	520-36-5	
Methyldihydrotestosterone/Mestanolone	521-11-9	
5α-Dihydrotestosterone	521-18-6	
Flavone	525-82-6	
Hydroxyflutamide	52806-53-8	
Haloperidol	52-86-8	
Estrone	53-16-7	
Androsterone	53-41-8	
Dibenzo[a,h]-anthracene	53-70-3	
Diethylstilbestrol	56-53-1	
Phenobarbital	57-30-7	
17α-Ethinyl estradiol	57-63-6	
Progesterone	57-83-0	
17α-Estradiol	57-91-0	

Apomorphine	58-00-4	n				n	n				n	n				p	p			
Methyl testosterone	58-18-4	p	p	p	p	p	p			p	p	n	n			n	p	n	n	n
Testosterone	58-22-0	p	p			p	p	p	p	p	p	n	n			n	n	n	n	n
4-androstanedione	5982-99-0										p									
4-Cumylphenol	599-64-4	n				n	n			n	p	n				p	p	n	p	p
Fenarimol	60168-88-9	n				n	n			n	n	p				p	p	n	n	p
Oxazepam	604-75-1	n				n	p				n	n				n	p			
Tetrahydrogestrinone	618903-56-3										p									
4-Androstenedione	63-05-8	p	p			p	p	p	p	p	p	n	n			n	n	n	n	n
Nilutamide	63612-50-0	p				n	n			n	p	p				p	p	p	p	p
Ketoconazole	65277-42-1	p	n			n	n			n	n	n	p			p	p	n	p	p
Cycloheximide	66-81-9	n				n	n			n	n	n				p	p	p	p	p
Prochloraz	67747-09-5					n	n			n	n			p	p	p	p	n	n	p
4-Hydroxytamoxifen	68047-06-3	n	n			n	n	n	n	n	n	n	n			p	p	n	n	n
Norethynodrel	68-23-5			p	p	p	p	p	p	p	p			n	n	p	n		n	p
Medroxyprogesterone acetate	71-58-9	p	p			p	p	p	p	p	p	n	n			p	n	p	n	n
p,p'-Methoxychlor	72-43-5	n	n			n	n	n	n	n	n	p	p			p	p	p	p	p
p,p'-DDE	72-55-9	p	n			n	n			n	n	p	p			n	n	p	p	p
Fluoxymestrone	76-43-7	p	p			p	p	p	n	p	p	n	n			n	n	p	n	n
Bisphenol B	77-40-7	n				n	n	n	n	n	p	n				p	p	p	p	p
Ammonium perchlorate	7790-98-9	n				n	n			n		n				n	n	n	n	n
o,p'-DDT	789-02-6	n	n	n	n	n	n	n	n	n	n	p	p	p	p	p	n		p	p
Levonorgestrel	797-63-7			p	p	p	p			p	p			n		p	n	n	n	n
Bisphenol A	80-05-7	n	n			n	n			n	p	p	p			p	p	p	p	p
Phenolphthalin	81-90-3	n				n	n			n	n	n				n	n	n	p	p
meso-Hexestrol	84-16-2	n				n	n	n	n	n	p	n				p	p	p	p	p
Mifepristone	84371-65-3	p	p			n	p			n	p	p	p			p	p	p	p	p
Boldenone	846-48-0										p									
Di-n-butyl phthalate	84-74-2	n	n	n	n	n	n	n	n	n	n	n	p	n	n	n	n	n	n	n
Butylbenzyl phthalate	85-68-7	n				n	n		n	n	n	n				n	n	n	p	n

2-sec-Butylphenol	89-72-5	n		n	n	n	p	n		n	n	n	n	n		
Bicalutamide	90357-06-5	p		n	n	n	p	p		p	p	p	p	p		
Clomiphene	911-45-5	n		n	n		n	n		p	n					
2,4,5-Trichlorophenoxyacetic acid	93-76-5	n	n	n	n	n	n	n	p	n	n	n	n	n		
Methyltrienolone(R1881)	965-93-5	p	p				p	n	n							
Finasteride	98319-26-7	n	n	n	n	n	p	n	p	p	n	p	p	p		

Table 2. Summary of the data gathered on ARTA for 83 candidate compounds for the validation study. "p" and "n" stand for positive and negative responses, respectively.

Compounds without actual data in the table come from an expert consultation advising the inclusion of such compounds. They were placed on hold list in case some more test chemicals were needed

Class definition (scoring)

The number of observed behaviour in the different source was summed up and divided by the total number of observation (associated ratio). From there it is possible to define four classes of compounds: agonists, antagonists, negatives, dual behaviour, and the rest (named here inconclusive) (see Table 3.). A first cleaning step was performed, removing unavailable, legally banned or expensive chemicals (15 chemicals eliminated). After solubility tests, 4 insoluble compounds (Diethylhexyl phthalate, Estrone, Dibenzo[a,h]-anthracene, and p,p'-DDE) were also discarded (64 chemicals remaining). Note that, in this count, the AR CALUX in-house HTS accounts for a weight of three since it was performed on 3 biological replicates.

Class	Ratio of positive agonist	Ratio of positive antagonist	Total ratio of positive (both agonist and antagonist)	Number of chemicals (original list of 83)	Number of chemicals (after cleaning, 64 remaining)
Pure agonist	>70%	<25%		10	8
Pure antagonist	<25%	>70%		19	17
Negative			<25%	23	17
Dual behaviour	>50%	>50%		8	7
Inconclusive				23	15
			Total	***83***	***64***

Table 3. Chemicals list break down by classes based on the number of observation of agonist and antagonist behaviour in the different data sources

Chemical selection

The goal of chemical selection was to select 15 agonists, 15 antagonists, and 15 negative compounds. From Table 3, it is clear that negative and antagonist pools have to be downsized from 17 to 15 while the agonist set has to be supplemented with, for example, the chemicals showing a dual behaviour

Negatives

17 compounds were showing negative in both agonist and antagonist tests (total ratio of positive response in all assays < 25%). To reach the target number of 15, we had to discard two compounds. We decided to cluster the 17 negative candidates according to their chemical structure (Avalon structural fingerprint, Tanimoto similarity). Figure 1 shows the result of the complete linkage hierarchical clustering. Two pairs of similar compound appear quite obviously and could be considered as redundant.

Those two pairs are:

- Daidzein (486-66-8) and Genistein (446-72-0), both isoflavones. Genistein was considered more interesting to keep because it is more data rich than Daidzein (e.g. 417 PubChem citations *versus* 141).
- Butylbenzyl phthalate (85-68-7) and Di-n-butyl phthalate (84-74-2), both belonging to the phthalate family. Di-n-butyl phthalate is kept because we could gather more observation and have more confidence in its behaviour (19 observations *versus* 12)

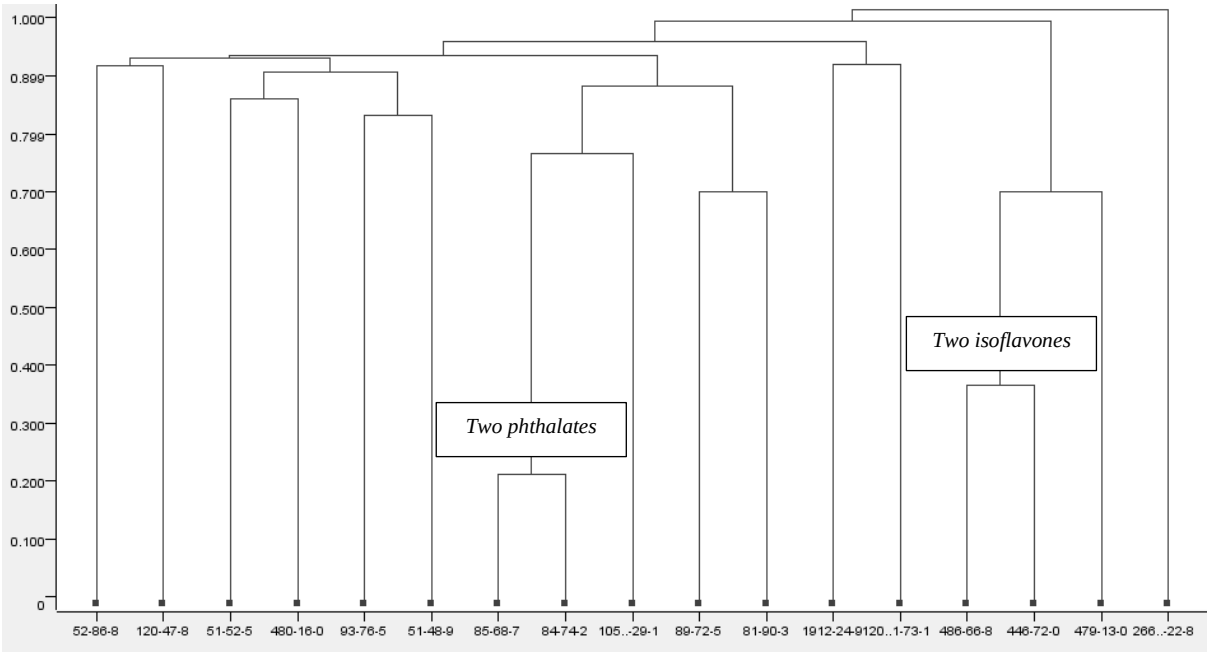


Figure 1. Clustering dendrogram obtained on the Tanimoto similarity matrix calculated on Avalon fingerprints for the 17 negative candidates.

The proposed list of 15 negatives to be tested is then:

Name	CASNR	Name	CASNR	Name	CASNR
Tamoxifen	10540-29-1	Genistein	446-72-0	Haloperidol	52-86-8
Ethyl paraben	120-47-8	Coumestrol	479-13-0	Phenolphthalin	81-90-3
Anastrazole	120511-73-1	Morin	480-16-0	Di-n-butyl phthalate	84-74-2
Atrazine	1912-24-9	L-Thyroxine	51-48-9	2-sec-Butylphenol	89-72-5
Sodium azide	26628-22-8	Propylthiouracil	51-52-5	2,4,5-Trichlorophenoxyacetic acid	93-76-5

Table 4. Proposed list of 15 negative chemicals

Antagonists

After data collection and cleaning, 17 compounds showed pure antagonist behaviour.

When the same clustering strategy as for the negatives is applied, one can found 3 pairs of similar compounds (clustering below a distance value of 0.55) as seen in Figure 2 (clustering dendrogram) and Table 5 (table of pairs).

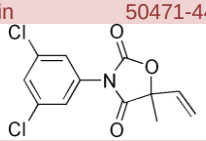
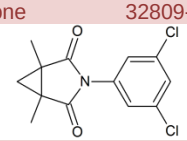
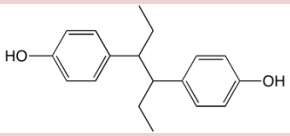
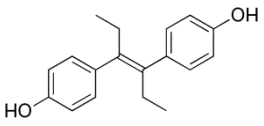
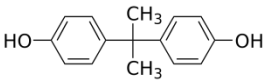
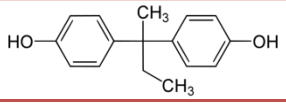
#	Name	CASNR	Name	CASNR	Commonalities
1	Vinclozolin	50471-44-8	Procymidone	32809-16-8	Dicarboximide (fungicides)
					
2	Meso-Hexestrol	84-16-2	Diethylstilbestrol	56-53-1	Structure differing only by one double bond
					
3	Bisphenol A	80-05-7	Bisphenol B	77-40-7	Bisphenols
					

Table 5. Three pairs of similar antagonist compounds (Tanimoto distance on Avalon fingerprint < 0.55) detected by clustering.

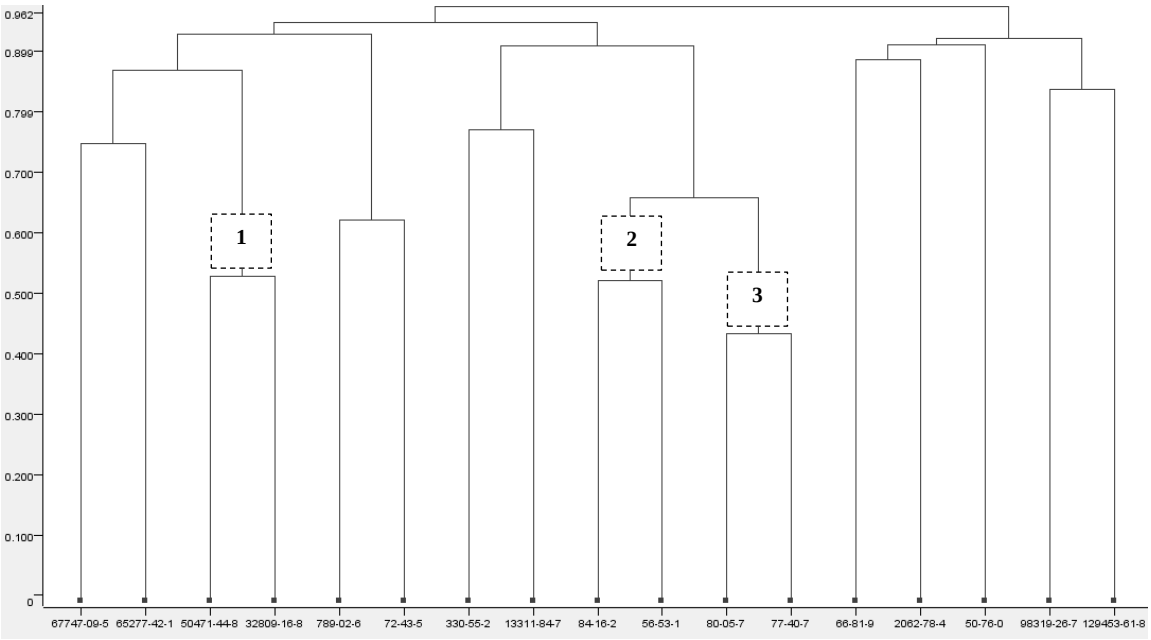


Figure 2. Clustering dendrogram obtained on the Tanimoto similarity matrix calculated on Avalon fingerprints for the 17 antagonists' candidates. Three pairs of similar compounds are detected (distance < 0.55).

If we apply a criteria of data richness within each pair (i.e. number of antagonist response observed), we select Vinclozilin (pair 1), Diethylstilbestrol (pair), and Bisphenol A (pair 3). However, since we only had to discard 2 compounds (and not 3), Procymidone (pair 1) is also kept in the selection as it was part of the initial selection used in the GLP study. That selection was performed previously with less data than this version of the strategy.

The proposed 15 antagonists are then:

Name	CASNR	Name	CASNR	Name	CASNR
ICI 182,780	129453-61-8	Vinclozolin	50471-44-8	Cycloheximide	66-81-9
Flutamide	13311-84-7	Actinomycin D	50-76-0	Prochloraz	67747-09-5
Pimozide	2062-78-4	Diethylstilbestrol	56-53-1	p,p'-Methoxychlor	72-43-5
Procymidone	32809-16-8	o,p'-DDT	789-02-6	Bisphenol A	80-05-7
Linuron	330-55-2	Ketoconazole	65277-42-1	Finasteride	98319-26-7

Table 6. Proposed list of 15 antagonists

Agonists and dual behaviour

Based on the classes defined, we collected 8 pure agonists. In order to reach the target number of 15, we decided to investigate the "inconclusive" and "dual behaviour" chemicals for which no clear behaviour could be observed, based on the defined criteria.

Two compounds (Corticosterone and Dexamethasone) are added because of their potential to interact with the Glucocorticoid receptor (GR). Those two chemicals indeed show up positive agonists in several AR assays due to a crosstalk mechanism between AR and GR. One of the specificity of ARCALUX is to be able to avoid this phenomenon and, in order to demonstrate this characteristic, the two compounds are included in the list and should show up negatives in the AR CALUX agonists test.

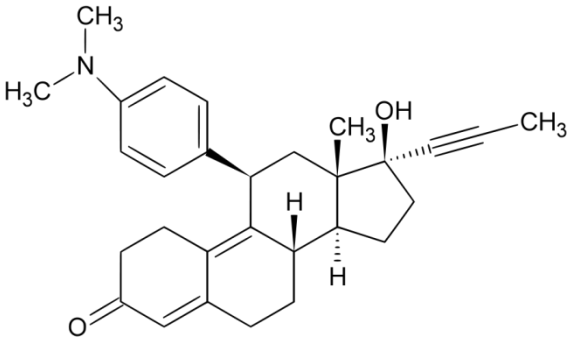
For the 5 spots left in the agonist lists, we simply include the compounds for which we have the higher number of positive agonist responses in the remaining list of "inconclusive" and "dual behaviour". Table 7 shows the top 6 "inconclusive"/"dual behaviour" compounds ranked by the number of positive agonist response observed in the data sources. All these compounds (with the exception of Medroxyprogesterone acetate) show a substantial antagonist activity, the reason they were considered as showing a "dual behaviour" in the first place.

Name	CAS	Positive agonist	Total agonist test	Positive antagonist	Total antagonist tests
Medroxyprogesterone acetate	71-58-9	8	8	2	7
Progesterone	57-83-0	6	8	9	9
17β-Estradiol	50-28-2	5	8	6	7
Spironolactone	52-01-7	5	8	7	7
Cyproterone acetate	427-51-0	4	6	7	7
Mifepristone	84371-65-3	4	6	7	7

Table 7. Six compounds, first tagged as "inconclusive"/"dual behaviour", with the highest number of agonist responses

To make the call between Cyproterone acetate and Mifepristone (showing the same profile in the data sources), we let the chemistry speak by selecting the structure that would bring more diversity to the set of agonist. Observing the already selected agonists, only the steroid family is represented, not surprisingly. Mifepristone was then selected over Cyproterone acetate because the former is decorated by an additional aromatic ring and an interesting alkyne moiety (see Figure 3).

Mifepristone (CAS 84371-65-3)



Cyproterone acetate (CAS 427-51-0)

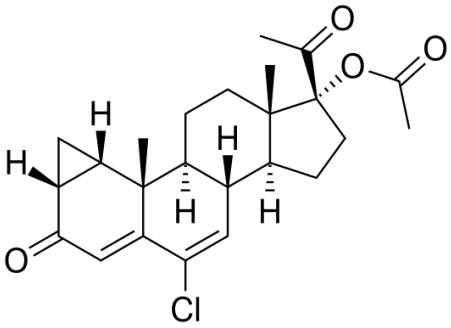


Figure 3. Structures of the Mifepristone and Cyproterone acetate

Name	CASNR	Name	CASNR	Name	CASNR
17β-Trenbolone	10161-33-8	Testosterone	58-22-0	Medroxyprogesterone acetate	71-58-9
Stanozolol	10418-03-8	4-Androstenedione	63-05-8	Progesterone	57-83-0
Methyldihydrotestosterone	521-11-9	Levonorgestrel	797-63-7	17β-Estradiol	50-28-2
5α-Dihydrotestosterone	521-18-6	Dexamethasone	50-02-2	Spironolactone	52-01-7
Methyl testosterone	58-18-4	Corticosterone	50-22-6	Mifepristone	84371-65-3

Table 8. Proposed list of 15 agonists

Summary

Class	Agonist	Dual	Antagonist	Negative	Inconclusive	Total
Selected	8	4	15	15	3	**45**
Considered		3	2	2	12	**19**
Insoluble		1	1	1	1	**4**
Unavailable	2		1	5	7	**15**
Total	**10**	**8**	**19**	**23**	**23**	**83**

Table 9. Summary table of the selection process.

The outcome of the presented selection process is presented in Table 9 where 45 chemicals were retained among negatives, antagonists, agonists, dual behaviour, and inconclusive compounds.

In term of chemical space covered, one can see in Figure 4 that the 45 selected chemicals cover well the initial space of 83 pre-selected compounds, thanks to the diversity-based strategy employed to select antagonists and negatives.

It is obvious that the agonists cover a very restricted space of structure since only certain fine structural features are needed to trigger an agonist response from the AR. Antagonists cover a wider space and “dual behaviour” compounds lay between the two types of response. Negatives show also a satisfying diversity.

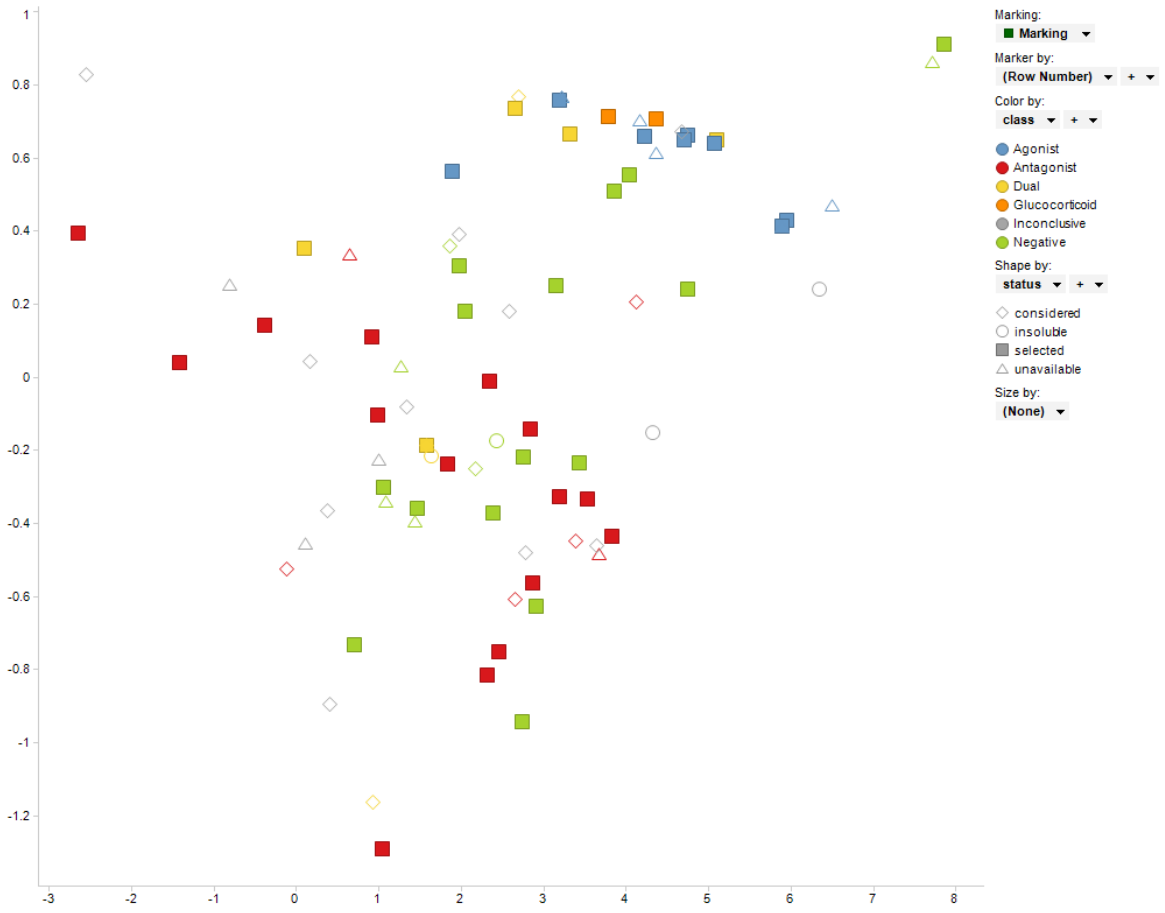


Figure 4. PCA (first 2 PCs) calculated on the similarity profiles (Tanimoto index on the standard RDkit structural fingerprint) for the 83 pre-selected chemicals. Colour accounts for the behaviour class and shape for the selection status.

In an effort to rationalize the chemical space covered by the selection, six main properties were computed (Figure 5). As seen on the histograms, the selection (red) represents well the space covered initially by the 83 pre-selected chemicals (red + grey).

In a validation study revolving around one particular receptor (AR), a substantial portion of the selection is centred around the activity domain (in term of properties of the binding ligands) linked to that receptor. Interestingly, a significant number of the selected chemicals show a profile similar to the natural binding ligands for the AR, testosterone (MW = 288; LogP=3.88; TPSA=37.3; HBD=1; HBA=2; rotatable bonds = 0)

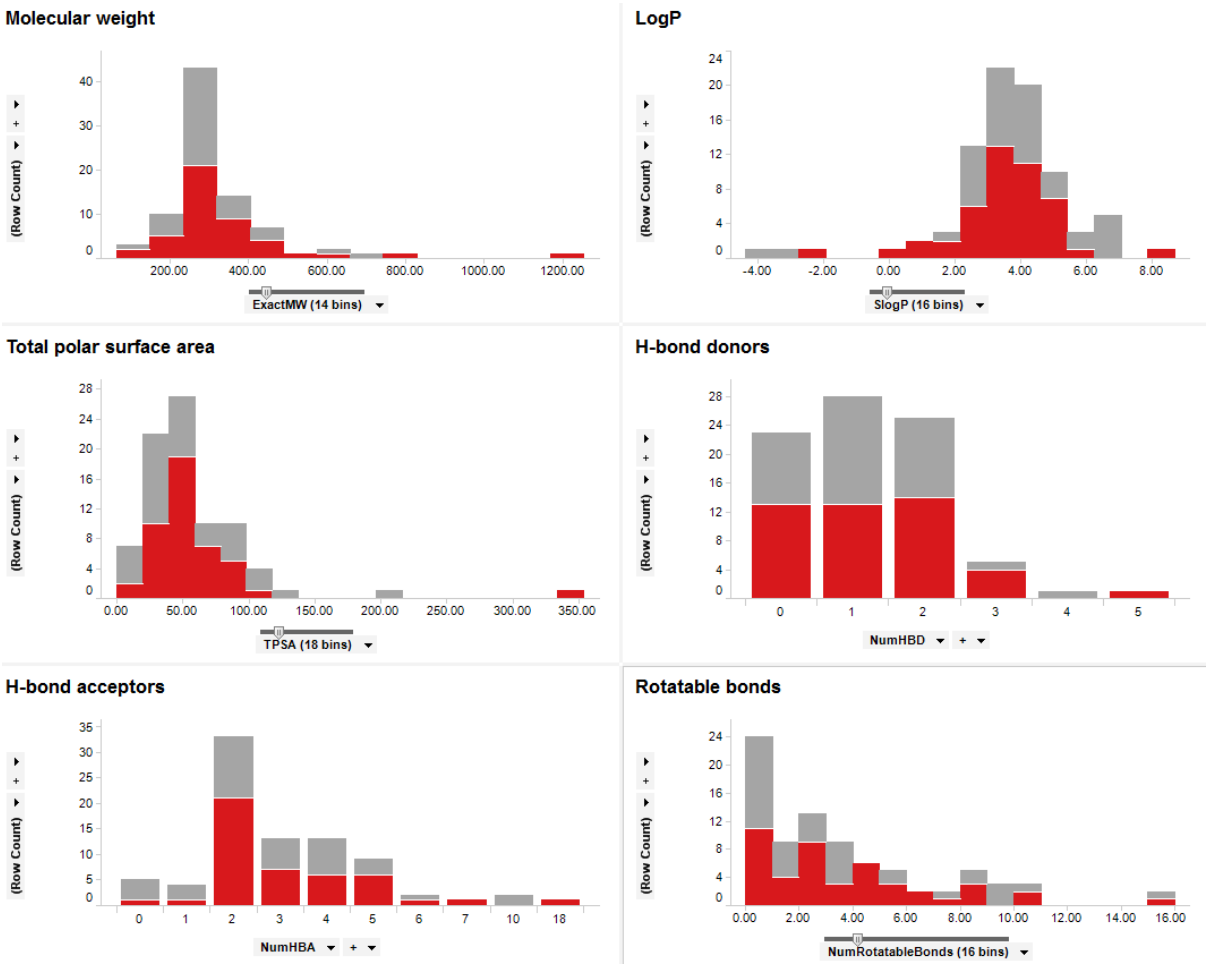


Figure 5. Histograms for 6 physical/chemical properties for the 45 selected (red) compared to the non-selected (grey) chemicals.

Information on chemicals

In order to accurately describe the diversity space covered by the validation study, we would like to gather an identity file for each of the chemicals tested, including information on chemical class, properties, type of use, types of toxicities, kinetics, and potency data relevant for the AR CALUX validation. A template is proposed in Table 10.

CASNR	Name	ARCALUX class	Chemical class	Use	MW	LogP	ED related behaviour	Other toxicities	Kinetics (metabolism, clearance)	Potency (EC50 for agonists, IC50 for antagonists, in M)
120-47-8	Ethyl paraben	Negative	Paraben, Benzoic acid	Fungicide Microbiocide Preservative Food additive (E214)	136.1	0.77	Weak estrogenic			N/A
13311-84-7	Flutamide	Antagonist	Anilide	Drug (prostate cancer)	276.0	3.21	Anti-androgen	Hepatotoxicity	Hydroxylated to hydroxiflutamide. Fast clearance (urine)	$3.99\ 10^{-07}$ (AR CALUX) $1.84\ 10^{-06}$ (PALM) $1.66\ 10^{-06}$ (Ecoscreen) $7.38\ 10^{-06}$ (Tox21 BLA) $4.56\ 10^{-05}$ (Tox21 LIC)
10161-33-8	17β-Trenbolone	Agonist	Steroid	Growth hormone	270.4	3.33	Androgen		HL 48/72h Urinary	$1.35\ 10^{-09}$ (Tox21 BLA) $1.02\ 10^{-10}$ (Tox21 LUC)

Table 10. Proposition of identity file to be gathered for each compounds selected for the validation study

Mapping of the 53 chemicals used in the validation study versus the REACH chemicals

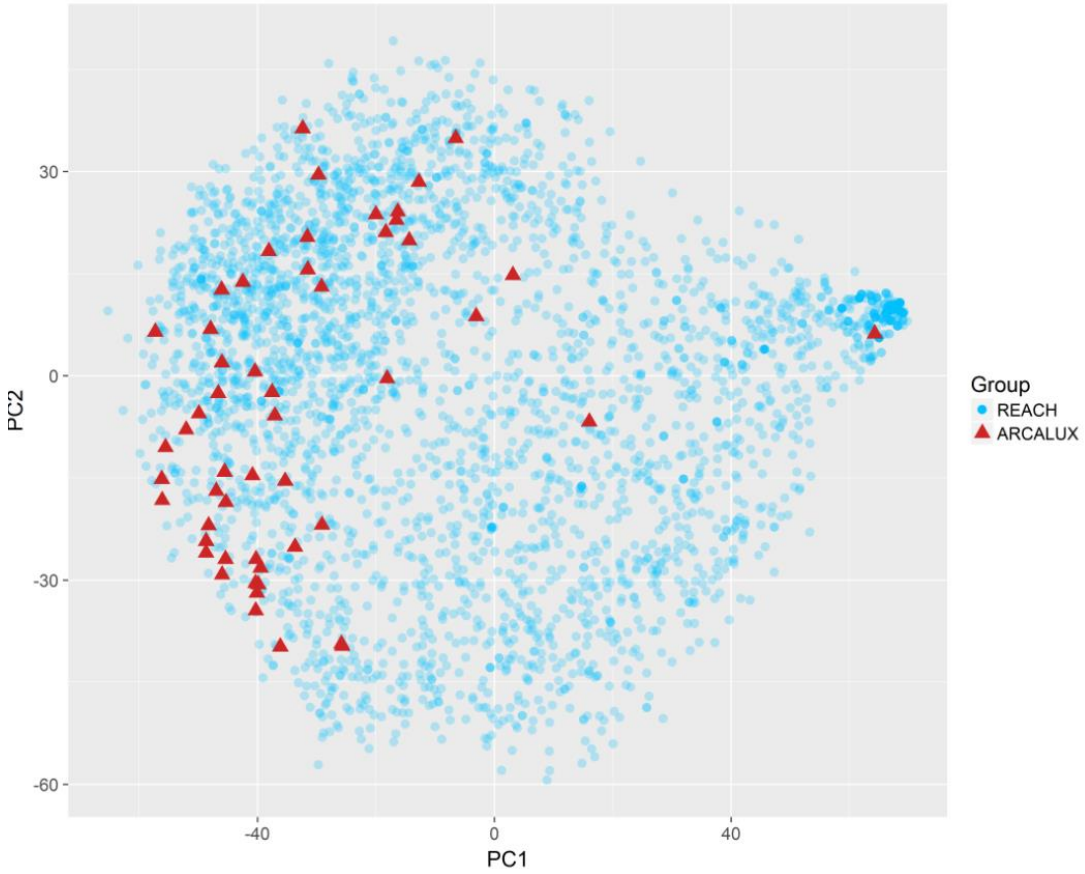


Figure 1: PCA calculated on the structural profiles (Tanimoto index) of the REACH chemicals versus AR-CALUX® validation set.

The axis and positions of the chemicals correspond to the first two principal components of the similarity matrix of the chemicals built using the RDKit (Landrum G. RDKit: Open-source informatics. 2015. http://www.rdkit.org) atomic pairs fingerprints. In blue: REACH chemicals; in red: AR-CALUX® validation set.

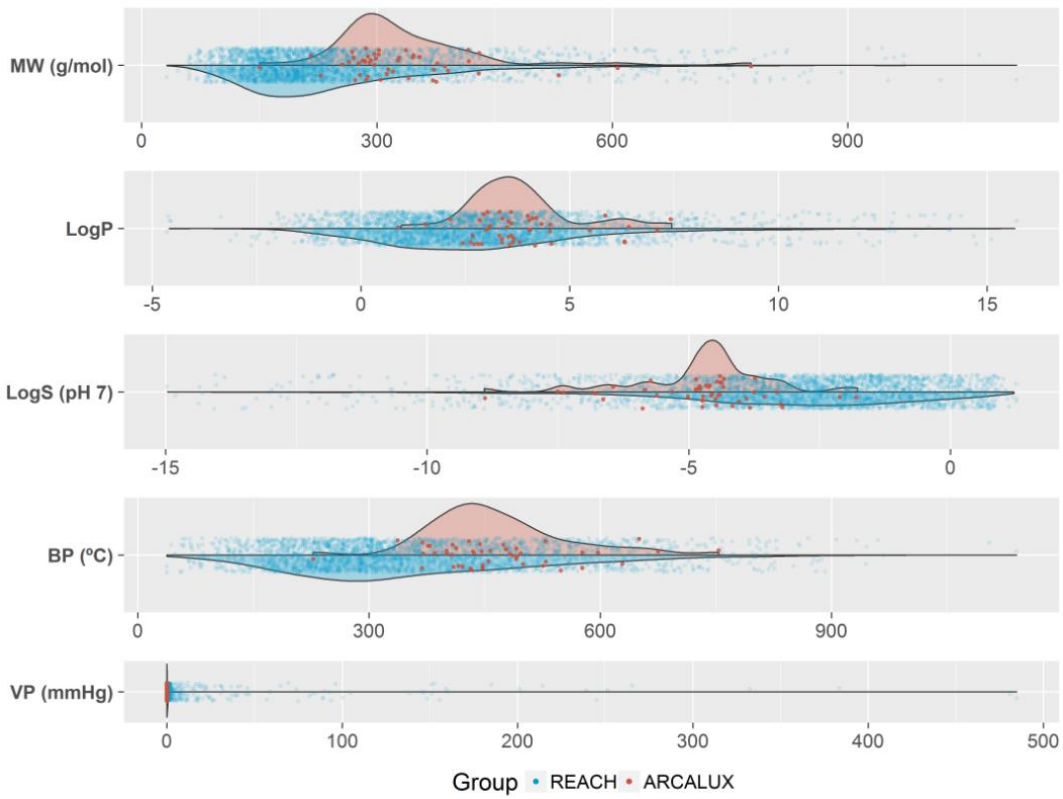


Figure 2: Distribution of physicochemical properties of the AR-CALUX® validation set versus the REACH chemicals

In blue: REACH chemicals, in red: AR-CALUX® validation set. The physicochemical predictions were carried out with ACDLabs/Perceptra.

5. Compilation of publicly available data on ARTA's and the AR-pathway model, for 46 coded test chemicals to be used in the AR-CALUX® validation study

The data sources were as follows. For the Tox21 assays and the AR-pathway computational model, the assignment of the classification is described.

- ARTA Japan validation study
  - Compiled from OECD TG 458

- ARTA Korea validation study
  - Compiled from validation report under review (2019)

- Tox21 assay luc data (Run Antagonist 2 with spiking of ligand (standard response), Run Antagonist 1 with higher concentration of ligand (specificity control)). The classification was carried out as follows:
  - Agonist
    - **Positive:** When Call of TOX21_AR_LUC_MDAKB2_Agonist is **Active**
    - **Negative** : When Call of TOX21_AR_LUC_MDAKB2_Agonist is **Inactive**

  - Antagonist
    - **Positive**
      - Call of [Antagonist2, Antagonist] = [Active, Active] and Log10(ACC Antagonist) - Log10(ACC Antagonist2) is => 0.5 (exception Diethylstilbestrol and p,p'-Methoxychlor with diff slightly below 0.5, both classified as P)
      - Call of [Antagonist2, Antagonist] = [Active, Inactive]
    - **Negative**
      - Call of [Antagonist2, Antagonist] = [Inactive, Inactive]
    - **False positive**
      - Call of [Antagonist2, Antagonist] = [Active, Active] and Log10(ACC Antagonist) - Log10(ACC Antagonist2) is < 0.5 (exception Diethylstilbestrol and p,p'-Methoxychlor with diff slightly below 0.5, both classified as P)
      - Call of [Antagonist2, Antagonist] = [Inactive, Active]
    - **no classification**
      - Call of [Antagonist2, Antagonist] = [not specified , Active
      - When Flag-Omit value is in Call

- Tox21 assay bla data
  - Agonist
    - **Positive** : When Call of TOX21_AR_BLA_Agonist_ratio is **Active**
    - **Negative** : When Call of TOX21_AR_BLA_Agonist_ratio is **Inactive**

  - Antagonist
    - **Positive**
      Call of Antagonist_ratio = [Active] and Log10(ACC Antagonist_viability) - Log10(ACC Antagonist_ratio) is => 0.5
    - **Negative**
      Call of Antagonist_ratio = [Inactive]
    - **False Positive**
      Call of Antagonist_ratio = [Active] and Log10(ACC Antagonist_viability) - Log10(ACC Antagonist_ratio) is < 0.5
    - **no classification**
      When Flag-Omit value is in Call

- AR pathway model (Kleinstreuer et al, 2017; Supplemental file 4)

  - Agonist
    - **Negative**: when AUC AG < 0.1
    - **Positive**: AUC AG => 0.1
  - Antagonist
    - **Negative**: when "Tox21 Confirmation Assay Flag" is
      - NA or AUC ANT < 0.1
    - **False Positive**: when "Tox21 Confirmation Assay Flag" is
      - FLAG: Antagonist shift, but CI overlap, AUC ANT => 0.1
      - FLAG: Wrong direction shift (Hit/Hit), AUC ANT => 0.1
    - **Positive**: when "Tox21 Confirmation Assay Flag" is
      - True antagonist shift (Hit/Hit), AUC ANT => 0.1
      - True antagonist shift (No hit/Hit), AUC ANT => 0.1

Annex 13.4

Chemical coding and distribution procedure

1. Sample preparation and distribution

The chemicals to be used in the validation study were distributed to the participating laboratories as sample aliquots in vials prepared by EURL-ECVAM (about 300mg for test items and about 500mg for reference and control items). The individual weights of the aliquots were also recorded, and provided to the respective recipients for information. In general, the chemicals were sent as coherent sets according to the project phase (study number).

The vials were labelled as follows:

- Chemical name and CAS number (or identity coded)

- Storage temperature (with inert gas, if indicated by supplier/SDS)

- Expiry date

- Hazard statement (H number) according to safety data sheet (SDS)

- Recipient laboratory acronym, to ensure correct package allocation for shipment.

Together with the aliquot weights, this information was also provided to the respective recipients as a sample list. For internal record keeping at EURL-ECVAM, lot numbers were also noted for all chemicals, with corresponding certificates of analysis (CoA) indicating material purity (available from the commercial supplier) retained on file.

Together with printed SDS, the samples were packed for shipping in fibreboard boxes, able to withstand dropping and stacking. When coded items were sent, two sets of SDS were included in the package:

- Bundled together in one sealed envelope, with all codes listed on the exterior (corresponding to the vial labels and marked SDS copies inside) and indicated '*for customs use only*'.

- Individually sealed in separate envelopes, each marked with respective code, and each indicated '*for emergency use only*'.

To preserve the integrity of the test item identity codes, the recipient was given instructions in advance to discard the customs set on arrival (unopened) and ensure the individual envelopes were retained (also unopened) for emergency reference only.

2. Coding of the test items

Random order binary letter codes were generated via *Random.org* string generator (on-line application) producing 676 (26x26) different combinations (or 576, omitting O and I to avoid confusion with numerals 0 and 1).

For the distribution of test items as unidentified chemicals, a three part code was adopted:

- laboratory acronym (e.g., BDS for BioDetection Systems)

- study reference (sequential study number: 1, 2, 3)

- sample aliquot code (alphabet letter pairs).

Thus, example codes were: BDS-2-AB, ENV-3-XY.

The identity of the chemicals to which the codes were assigned remained confidential from the VMG and the statistician until the end of the experimental part of the validation study The codes corresponding to the same test item were disclosed to the statistician for un-biased statistical analysis of reproducibility. Once the data analysis was complete, the identity of the test items was released to VMG.

3. List of the codes

Seq. #	chemical name	lab/code			
1	17β-Trenbolone	**CIT-2-CB**	**ENV-2-VP**	**SPT-2-WT**	**BDS-2-SQ**
2	Stanozolol	**CIT-2-XF**	**ENV-2-XN**	**SPT-2-NM**	**BDS-2-GG**
3	Spironolactone	**CIT-2-AE**	**ENV-2-HD**	**SPT-2-HX** **SPT-3-AR**	**BDS-2-CG**
4	Medroxyprogesterone acetate	**CIT-2-NX**	**ENV-2-ET**	**SPT-2-KX**	**BDS-2-LN**
5	Bisphenol A	**CIT-2-RD**	**ENV-2-YA**	**SPT-2-MZ**	**BDS-2-CW**
6	Bicalutamide	**CIT-2-YY**	**ENV-2-DF**	**SPT-2-WS**	**BDS-2-QX**
7	Disulfiram	**CIT-2-RK**	**ENV-2-VZ**	**SPT-2-YP**	**BDS-2-MU**
8	Tamoxifen	**CIT-2-XK**	**ENV-2-QF**	**SPT-2-NS**	**BDS-2-WW**
9	Atrazine	**CIT-2-LK**	**ENV-2-ZR**	**SPT-2-SH**	**BDS-2-YN**
10	17α-Ethynyl estradiol	**CIT-2-GQ**	**ENV-2-GK**	**SPT-2-LR**	**BDS-2-SD**
11	Sodium azide	**CIT-2-ZT**	**ENV-3-AT**	**SPT-3-AM**	**BDS-2-XH**
12	Diethylhexyl phthalate	**CIT-2-DQ**	**ENV-3-FJ**	**SPT-3-SJ**	**BDS-2-BU**
13	Methyldihydrotestosterone	**CIT-2-TJ**	**ENV-3-UJ**	**SPT-3-NY**	**BDS-2-VF**
14	Vinclozolin	**CIT-2-NN**	**ENV-3-GU**	**SPT-3-RU**	**BDS-2-AJ**
15	Prochloraz	**CIT-2-JB**	**ENV-3-BH**	**SPT-3-ZZ**	**BDS-2-AN**
16	Fluoxymesterone	**CIT-2-YW**	**ENV-3-DX**	**SPT-3-FA**	**BDS-2-BG**
17	17β-Estradiol	**CIT-2-ML**	**ENV-3-PU**	**SPT-3-BL**	**BDS-2-UE**
18	Benzylbutyl phthalate	**CIT-2-PY**	**ENV-3-KU**	**SPT-3-QT**	**BDS-2-CM**
19	Propylthiouracil	**CIT-2-HZ**	**ENV-3-CK**	**SPT-3-WZ**	**BDS-2-NZ**
20	Hydroxyflutamide	**CIT-2-TF**	**ENV-3-EH**	**SPT-3-ZV**	**BDS-2-MK**
21	Levonorgestrel	**SPT-3-RY**			
22	Cyproterone acetate	**SPT-3-TH**			
23	2-tert-Butylanthraquinone	**SPT-3-YC**			
24	Arochlor1254	**SPT-3-SB**			
25	Nandrolone	**SPT-3-VK**			
26	o,p'-DDT	**SPT-3-GZ**			
27	Phenolphthalin	**SPT-3-RX**			
28	2,4,5-T	**SPT-3-DC**			
29	Methyltrienolone (R1881)	**SPT-3-VE**			
30	Actinomycin D	**SPT-3-FK**			
31	Diethylstilbestrol	**SPT-3-JA**			
32	L-Thyroxine	**SPT-3-TM**			
33	Haloperidol	**SPT-3-BR**			
34	Norethindrone acetate	**SPT-3-XY**			
35	Pimozide	**SPT-3-QM**			
36	Progesterone	**SPT-3-JQ**			
37	Linuron	**SPT-3-NQ**			
38	Methyltestosterone	**SPT-3-RV**			
39	2-sec-Butylphenol	**SPT-3-UA**			
40	Corticosterone	**SPT-3-WJ**			
41	Ketoconazole	**SPT-3-MG**			
42	Finasteride	**SPT-3-FX**			
43	Fulvestrant	**SPT-3-XR**			
44	Cycloheximide	**SPT-3-DS**			
45	Norethindrone	**SPT-3-VR**			
46	Mifepristone	**SPT-3-LU**			

Annex 13.5

SOP version V07
(final version)

Available at EURL ECVAM's

Tracking system for alternative methods towards regulatory acceptance (TSAR)

(https://tsar.jrc.ec.europa.eu/test-method/tm2010-07)

Annex 13.6

Solubility data on 46 coded test items

Solubility data on 46 coded test items from the 4 participating laboratories

Chemical name	Stock solution mg/ml					Medium solution µg/ml			
	CiToxLab	ENVIGO	RISE (SPT)	BDS		CiToxLab	ENVIGO	RISE (SPT)	BDS
Trenbolone	50	50	50	50		5	50	15	50
Stanozolol	50	50	5	15		5	5	5	5
Spironolactone	50	50	15	50		50	50	15	50
Medroxyprogesterone	5	15	1.5	15		5	5	1.5	15
Bisphenol A	50	50	50	50		50	50	15	50
Bicalutamide	50	50	50	50		50	50	50	50
Disulfiram	50	50	50	50		15	15	5	15
Tamoxifen	15	15	5	15		5	15	5	15
Atrazine	5	50	15	50		5	15	15	15
Ethynyl estradiol	50	50	15	50		15	15	15	15
Fluoxymesterone	50	5	50	50		15	5	5	5
Prochloraz	50	50	50	50		50	50	50	50
Benzylbutyl phthalate	50	50	50	50		15	50	1.5	15
o,p'-DDT			50					1.5	
Methyldihydrotestosterone	15	5	15	15		5	5	5	5
17β-Estradiol	50	50	50	50		5	15	5	15
Hydroxyflutamide	50	50	1.5	50		50	50	1.5	50
Vinclozolin	50	50	50	50		15	15	5	15
Propylthiouracil	50	50	50	50		50	50	50	50
Diethylhexyl phthalate	50	50	15	50		50	15	15	50
Sodium azide	15	50	15	15		5	15	15	15
Linuron			50					15	
Levonorgestrel			15					1.5	
Corticosterone			50					50	
Methyltestosterone			50					15	
Progesterone			50					15	
Nandrolone			50					50	
Methyltrienolone			50					50	
Norethindrone			50					5	
Norethindrone acetate			5					5	
Cyproterone acetate			50					15	
Diethylstilbestrol			50					15	
Finasteride			50					50	
Mifepristone			50					5	
Haloperidol			50					5	
Ketoconazole			15					15	
Pimozide			5					5	
L-Thyroxine			50					50	

Fulvestrant			50					1.5	
Actinomycin D			5					5	
Cycloheximide			50					50	
2,4,5-T			0.15					0.15	
2-sec-Butylphenol			15					15	
tert-Butyl anthraquinone			15					5	
Arochlor 1254			50					5	
Phenolphthalin			15					15	

Annex 13.7

List of additional documents filed for the study and available on request at EURL ECVAM

- Training documents (Agenda, Planning of the training)
- Study plans and study reports of the 4 participating laboratories
- SOP versions V03, V04, V05 and V06, and solubility SOP
- JRC technical report on "Technical meeting on the Implementation of the AR-CALUX® *in vitro* method"
- JRC technical report on "Transfer Evaluation Report"
- JRC technical report on "Data of study 2"
- EURL ECVAM report GLP compliant study SR-ST57
- Supporting data for development of the specificity control criterion R^2
- Technical issues during implementation of the AR-CALUX® method at CitoxLAB (report of BDS)
- Quality control (Identity verification) of the cell lines used in the participating laboratories
- Call for expression of interest "Proposal to EU-NETVAL members for participation in a multi-study validation trial"

GETTING IN TOUCH WITH THE EU

In person
All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email
Europe Direct is a service that answers your questions about the European Union. You can contact this service:
- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online
Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-30044-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi: 10.2760/60314
ISBN 978-92-76-15037-4