Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2026)20

Unclassified

English - Or. English

29 June 2026

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Peer Review Report of the Human Glucocorticoid Receptor Transactivation (GRTA)

OECD Series on Testing and Assessment

E-mail: ehscont@oecd.org.

JT03590131

This work was approved and declassified by the Chemicals and Biotechnology Committee on [02/07/2026].

Please cite this publication as:

OECD (2026), *Peer Review Report of the Human Glucocorticoid Receptor Transactivation (GRTA)*, OECD Series on Testing and Assessment, No. 442, OECD Environment, Health and Safety, Paris, link to ONE M&P document https://one.oecd.org/document/ENV/CBC/MONO(2026)20/en/pdf

Contact us

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

E-mail: ehscont@oecd.org

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 countries in North and South America, Europe and the Asia and Pacific region, as well as the European Union, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several Partner countries and from interested international organisations attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (https://www.oecd.org/en/topics/chemical-safety-and-biosafety.html).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank, Basel, Rotterdam and Stockholm Conventions and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

Foreword

This document contains the Peer review report of the human Glucocorticoid Receptor TransActivation (GRTA) assay, developed for the detection of glucocorticoid receptor agonist and antagonist activity of chemicals.

The scientific validity of the GRTA assay was assessed by an international Peer Review Panel in 2024. The peer review report supports the development of the corresponding test method for inclusion in a new Test Guideline, TG No 454A. The Peer review report and the Validation report of the GRTA assay were circulated for comments to the Working Party of the National Coordinators of the Test Guidelines Programme (WNT) in December 2025. The WNT approved the peer review report at its 38th meeting in April 2026. The WNT also approved the Validation report of the GRTA assay.

This report is published under the responsibility of the Chemicals and Biotechnology Committee.

Summary of the Peer Review of the Validation Study for the Glucocorticoid receptor *trans*-activation (GRTA) method

17 March 2025

Table of Content:

Summary

1. This review report presents the summary of the independent review of the validation study report for the Glucocorticoid Receptor *trans*-activation (GRTA) test method by an independent Peer Review Panel (PRP). Reviewers were selected from a pool of experts nominated via the OECD Working Party of the National Coordinators to the Test Guideline Programme (WNT) to ensure balanced representation of scientific and regulatory expertise, as well as geographical and institutional representation (NGO, industry, academic, governmental).

2. The GRTA test method is based on the HeLa-derived HMLN-hGR cell line that contains the Luciferase (Luc) gene under the control of the Glucocorticoid Receptor (GR) responsive MMTV promoter and stably overexpresses the human glucocorticoid receptor alpha (GRα). GR activity is evaluated by measuring the level of expression of luciferase based on its activity to convert luciferin to the luminescent oxyluciferin.

3. The test system contains two modes, an agonist mode for detecting substances that activate the GR and an antagonist mode where the ability of a substance to block or interfere with the GR activity induced by a known GR agonist is evaluated. In both modes, the reference activator/stimulant is the synthetic pharmaceutical dexamethasone (DEX). In addition, to distinguish between "true" antagonism and interference due to other (unspecific to GR ligand binding) mechanisms, an Interreference Mode of the protocol was used. In the Interference mode, the data interpretation procedure applied a Specificity Control Test based on statistical criterion known as R^2 (square of the correlation coefficient) for comparing two concentration responses generated from competition of a test substance with the activity of DEX concentrations nominally producing 100% (specificity control response) and 80% (test response) of the Luc reporter signal. This criterion is based on the assumption that the concentration-dependent decrease of the Luc signal, generated by increasing non-toxic concentrations of the test substance, which is not due to competitive antagonism with DEX, is proportionally the same ($R^2 \geq 0.9$ between the two dose responses) when tested in the presence of two concentrations of the specific control agonist DEX generating 100% and 80% Luc signal.

4. The validation ring trial for the GRTA test method was performed between 2021 and 2023. It was organised by PEPPER[1], supported by an external consultant for the initial literature search on reference chemicals, a statistician and data management expert, as well as a validation management team (VMT). The PEPPER Secretariat co-authored the report titled: "Validation report on the Glucocorticoid receptor transactivation method" (Attachment A) which was made available, together with the raw data, on the OECD public site in October 2024.

5. The independent expert review was conducted between October 2024 and February 2025. The PRP evaluated the validation study considering eight charge questions based on the principles of the OECD Guidance Document (GD) 34 "*Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment"(OECD 2005)*, in the context of the objectives and goals outlined in the validation study: (i) to evaluate the transferability and reliability of the GRTA method (repeatability within a laboratory and reproducibility between laboratories) and (ii) to evaluate the relevance of the test method through comparison of the classifications of the compounds by this test method to reported/expected classifications from other GR relevant test methods, with the ultimate goal to support the development of an OECD Test Guideline for assessment of agonists and antagonists of the activity of the GR.

6. Overall, the peer review panel considers that the validation study of the GRTA method is consistent with the validation principles of the OECD GD 34 and aligns well with the approach being proposed in revisions to OECD GD 34. Specifically, the PRP considered that the technical transferability, the intra- and inter-laboratory reproducibility of the GRTA method was high

[1] (https://ed-pepper.eu/en/organisation/)

and adequately demonstrated. The PRP also considered that the test protocol adequately describes the methodology but recommended several clarifications and modifications that are likely to further improve reproducibility. Specifically, a simplification of the prediction model for the antagonist mode (i.e. omission of the R^2 criterion) and inclusion of cortisol as a reference 'stimulant'. Use of cortisol as a reference agonist, alongside DEX as a positive control, in the agonist mode is also recommended.

7. Given that the GRTA method addresses an endpoint that has not been previously considered for an OECD test guideline and for which there are no well-established reference data, a set of 36 reference chemicals for the validation study was selected by a literature search for chemicals associated with GR activity *in vitro* and *in vivo*. The specific strategy for the search is not reported in the validation study report but the majority of the active chemicals (agonists as well as antagonists) are well known steroid pharmaceuticals.

8. The PRP highlights that the validation study primarily focused around identification of strong GR binders, activators and antagonists that parallel and compete with the activity of the synthetic glucocorticoid DEX. Under these circumstances, the broad utility of the current GRTA method as an OECD test guideline for identifying interaction with the GR by non-steroidal pharmaceutical substances is uncertain.

9. The PRP recognises that the choice of reference chemicals for the validation was a difficult task since only a few chemicals (mainly pharmaceutical steroids) have convincing and reliable data to aid assessment of predictive capacity of GR activity. Given the available evidence, the validation principle 5 and 6 are considered largely met. However, in order to evaluate and ensure good predictive performance for substances across the wider spectrum of responses of the GRTA method, further analysis and testing of relevant substances will be useful. The review panel has provided an initial analysis from the ToxCast database (Annex 5) to facilitate identification of additional substances for reference testing in the future.

10. The GRTA method and its validation study address a significant gap at Level 2 of the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters. The scientific rationale of the assay is well described but primarily focuses on GR activity resulting from ligand binding. PRP recommends strengthening the discussion on the wider context of GR signalling pathways. In addition, more explicit highlighting of the limitations of the assays is recommended to avoid future over-interpretation of the results of the assay in the regulatory context.

11. Generally, it is not expected that this cell-based system should cover all possible mechanisms of GR activity/modulation/disruption. It is understood that transactivation assays are intended to be used together with other assays in the context of tiered and/or integrated testing strategies to assess a potential link between endocrine activity and endocrine disruption. However, better characterisation of the cell line is needed as well as adequate controls, including properly addressing the level of a decrease of Luc signal and the allowable level of cytotoxicity in antagonist mode. This limitation is not unique to the GRTA assay and it is similar to other TA assays used in the context of assessment of endocrine active substances.

12. It was a concern that the current protocol using the particularly strong agonist DEX as a stimulant in the antagonist mode would reduce the sensitivity of the assay to detect relatively weak antagonists. This led to the recommendation to use cortisol rather than DEX as the stimulant in the antagonist mode, and/or use two nominal lower concentrations of DEX (e.g., generating 50% and 80% Luc activity).

13. It was concluded that there is not a well-recognized candidate method for a reliable statistical approach that can distinguish "true" antagonism from interference and that the interference mode of the antagonism assay as described in the current protocol should be omitted.

14. Considering the need for a non-animal mechanistic, easily transferable and scalable method to target GR activation pathways in the assessment of endocrine active substances, the Validation Study (revised as per the recommendations 1-4 and 7) is appropriate to initially support the development an OECD guidance document based on the simplified protocol for the GRTA test method recommended in this review. The Guidance Document would make the methodology available to the regulatory and research communities to conduct a mechanistic assay to help inform integrated assessment and identification of GR active substances that could potentially lead to adversity in intact organism. The 2002 WHO/IPCS definition of an endocrine disruptor requires establishing a link between an endocrine mechanism and adversity in an intact organism. Importantly, high quality new data generated according to a standardised protocol in the Guidance Document would allow for more rapid refinement and updates to the methodology to support the development of a future Test Guideline.

Recommendations:

1. The introduction of the Validation Study should include further discussion on: (i) the wider scientific context, i.e. the GR activation/inhibition pathway beyond substance binding to the GR; (ii) examples of adverse outcomes related to overactivation of GR (hypercortisolism); (iii) listing all relevant AOPs and literature (can be done in an Appendix) to illustrate how the results of the GRTA method could be linked to relevant AOs and/or a higher level testing strategy context; (iv) the limitations of the assay (see next).

2. Discussion and addressing limitations:
    a. Consider the limitations for detecting weaker agonists and antagonists by using DEX versus cortisol as a reference agonist (agonist mode of the protocol) and stimulant (antagonist mode) in the protocol under review.
    b. Discuss and to the extent possible characterise the level and ratio of GR to relevant co-factors in the context of this HeLa-based cell system
    c. Discuss and to the extent possible characterise the availability and function of cellular transporters for different tested substances
    d. Discuss other aspects of interference with the GR activated response beyond binding to the receptor, including posttranslational modifications of the receptor
    e. Discuss the potential limitations of using one particular promoter response element, i.e. the MMTV promoter

3. Descriptive/Editorial modifications of the protocol
    a. Represent data as fold increase over negative control in addition to presenting normalized activity as a percent of control. Add second Y-axis for fold induction and second X -axis to report the test concentrations in micromolar concentration.
    b. Specify cell seeding number in addition to cell confluency.
    c. Specify a highest concentration limit for testing in the protocol. PRP discussed 100 µM and even 30 µM to suit most environmentally relevant concentrations.
    d. Clarify and consistently use the wording for the criterion for positive agonist throughout the validation study. PRP recommends .."above 10% of the maximum effect of the positive control" (see also discussion under CQ3).
    e. Provide more guidance and background on why recommending the 4-parameter over the 3-parameter logistic regression for concentration response modelling and the rationale for including the 4th parameter (i.e., y-intercept). It was noted in the report that the 4-parameter was chosen to get more consistent results from the R^2 analysis for the interference mode. If the interference mode is no longer part of the protocol (as recommended below) the 3-parameter may be a better choice for modelling and consistency with concentration-responses.

f. Potentially modify the current Acceptance Criteria (AC) for basal activity (now "≤ 4% of maximum luminescence" based on:
    - the typical background luciferase activity for the solvent control and what labs typically measure for mean maximal fold induction for DEX (current protocol) and/or cortisol (updated protocol – see Recommendation 5 below) relative to the solvent control.
    - a Z-factor (https://en.wikipedia.org/wiki/Z-factor)

4. Modifications of the current protocol regarding the prediction model – additional analysis without additional testing
    a. Omit the interference mode from the protocol. Consider using cortisol as a reference stimulant (recommendation 5) or initially two DEX concentrations generating sub-saturating activity levels (e.g. 50% and 80% activity) to confirm antagonism and/or inhibition of GR signalling by a different (unknown) mode of action.
    b. Antagonism mode: **cytotoxicity threshold** for positive antagonist should be lowered from 20% to 10% unless the IC30 instead of IC20 is used to define positive antagonist with the current protocol where DEX is used as a stimulant.
    c. Agonism mode: Consider adjusting the **threshold for positive agonist** (currently: >10% of the maximum effect of the positive control) to help identify weaker agonists. Given that there is a limited number of tested substances that have intermediate activity in the validation study, such analysis would best be done in the context of the consideration of cortisol as a reference "calibrator' of the test system (see recommendation 5 below).
    d. Use the 3-parameter logistic regression (instead of 4-parameter) to increase consistency between different labs analysing the same data and to avoid not correctly modelling "hanging curves" where there is an incomplete concentration response curve in antagonism mode.

5. Updating the protocol to enhance the potential for detecting weaker agonists and antagonists - may require additional testing, selection of additional test substances, or analysis of results with cortisol alongside DEX in an updated validation study
    a. Agonist mode: index/calibrate the responses from zero (vehicle) to 100% (maximum response) using cortisol as the response reference and use DEX as the confirmatory control.
    b. Antagonist mode: both DEX and cortisol could be used as the stimulant but at sub saturating concentrations (e.g. starting stimulation for the stimulant at EC50 and EC80, for DEX, 80% for cortisol)
    c. Consider appropriate thresholds for positive agonist and cytotoxicity in antagonist mode using cortisol as a reference agonist and stimulant

6. Further analysis of relevant literature (published after the initial search and also some recommended by the panel under CQ5) and databases (e.g. entire ToxCast database) to identify additional reference chemicals for testing in the intermediate activity range of the GRTA method.

7. Include a better representation of the participants in the Validation Study and their responsibilities i.e. provide the Validation Plan as an attachment to the Validation study in which the distinct roles of the VMG, consultants and the authors of the Validation Study report is clearly outlined in each stage of the validation, including the selection of relevant chemicals for testing.

8. In addition to addressing the recommendations above, the authors should ensure thorough review of clarity of language and inclusion of all relevant data in Tables and Figures in the revised Validation Study.

Background

15. There is a growing concern of possible harmful consequences of exposure to chemicals that are capable of modulating or disrupting the endocrine system. The endocrine system is a complex network of glands and organs and uses hormones to control and coordinate the physiological processes in the body. Therefore, the evaluation of chemicals that interfere with endocrine system requires a large suit of relevant assays. Traditional *in vivo* testing methods can be time-consuming, costly, and raise ethical concerns regarding animal welfare. *In vitro* assays offer a faster, in some cases more cost-effective approach to assess the potential for interaction of chemical substances with endocrine pathways, allowing for moderate- to high-throughput screening and prioritization of numerous chemicals for further testing.

16. The OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters provides a guide to the available tests which can provide information for assessment of potential endocrine modulators or disruptors, including validated and developing *in silico*, *in vitro* and *in vivo* methods (OECD 2018). Level 2 of the OECD conceptual framework lists several test guidelines for *in vitro* assays relevant to the assessment of modulators of the estrogen and androgen receptors signalling systems (including steroidogenesis), and some assays undergoing validation to address modulators of the thyroid system. However, assays for evaluating other endocrine signalling pathways, in particular for glucocorticoid receptor (GR) activity modulators and/or disruptors, are lacking.

17. The GR plays a critical role in stress response, metabolism, and immune function. Therefore, it is important to develop and validate assays for evaluating the ability of chemicals to interfere with the activity of this specific nuclear receptor. The GRTA method is based on the HeLa-derived HMLN-hGR cell line that contains the Luciferase (Luc) gene under the control of the GR responsive MMTV[2] promoter and stably overexpresses the human glucocorticoid receptor alpha (GRα) described in Grimaldi et al., 2019 and Toso et al. 2023. In this test system, substances that activate or block/interfere with the activation of the GR can be detected by measuring the level of expression of Luciferase based on its activity to convert luciferin to the luminescent oxyluciferin.

18. The method was developed in the laboratory of Professor Patrick Balaguer at INSERM in Montpellier (Grimaldi et al., 2019). In this validation study the INSERM laboratory had major responsibility for the development of the SOP and transfer of knowledge and the other labs.

19. In 2021, the GRTA test method was proposed to the OECD Test Guideline Programme (TPG) by France and Canada. It was accepted on the TPG workplan (Project 4.161) in 2022 with the outline to perform transferability and validation testing work followed by an expert peer review to assess the validation status of the assay and the readiness of the GRTA method for development of an OECD Test Guideline based on the results from the Validation Study Report (Attachment A).

20. The experimental testing part of the validation study was performed between 2021 and 2023 in 4 participating laboratories in France, 3 naïve laboratories and the developer's laboratory at INSERM. It proceeded in 3 phases: 1st: transferability phase (5 chemicals tested in 4 laboratories), 2d blind ring-trial phase (34 chemicals tested in 4 laboratories) and, a supplementary 3d phase (11/34+1 new chemicals were tested under slightly modified SOP in 2 laboratories). In 2024, the PEPPER Secretariat performed additional detailed analysis of the literature used to select test/reference chemicals for the validation (Attachment A, Annex 1).

21. The validation study was organised by PEPPER, the French innovative public-private platform dedicated to the validation of methods relevant for the characterisation of endocrine disruptors (https://ed-pepper.eu/en/organisation/). PEPPER's permanent team/secretariat is supported by

[2] Mouse Mummery Tumour Virus

a Scientific Council and a Relevance Committee that provide high level strategic advice on assay needs, choice of substances and validation questions in general. Based on these PEPPER organises a call for tests labs to join the ring trial of the GRTA method, and also helped to finance the trial, including contracting support from data manger, biostatistician, technical support for anonymisation and dispatching of test substances, and for investigation of literature on GR pathway-relevant substances. In addition, a GRTA method Validation Management Group was engaged including external experts that governed important decisions about the study design, choice of reference chemicals for the ring trial of the GRTA method validation, protocols, timelines, data analysis and acceptance of results during the study. PEPPER's secretariat authored the Validation Study Report.

22. The choice of test chemicals was a particularly challenging task since most substances (other than the well-known steroid GR agonist and antagonists) that have been examined in various assays or models relevant to the GR activation pathway, were only supported by one or two publications and often with inconsistent results. However, the Validation Study Report highlights that the inclusion of all these substances was important to evaluate further their GR activity with a reliable test and also to 'challenge' the specificity of the GRTA test method i.e. whether it would show a high number of GR active substances, potentially giving rise to false positives.

23. The objectives of the validation study, as outlined in the study report, were to (i) evaluate the transferability and reliability of the GRTA test method (repeatability within a given lab and reproducibility between laboratories) and (ii) to evaluate its relevance through comparison of the classifications of the compounds tested by the GRTA to classifications reported/expected based on the results of other existing GR activity test methods. Ultimately the results of the validation study would support development of an OECD Test Guideline for testing of GR active substances and expand the domain of coverage of Level 2 of the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters (OECD 2018[1]).

The peer review process

24. The Peer Review Panel (PRP) was established in October 2024 in order to provide an independent review of the validation of the GRTA test method according to the principles of the OECD GD 34 (OECD, 2005) considering the purpose and goals of the validation study (see above paragraph 23 of this report and Section 1.4 of the Validation Study Report in Attachment A).

25. The review experts were nominated by the OECD Working Party of the National Coordinators of the Test Guidelines Programme (WNT) and were then selected by an independent Peer Review Manager (PRM) who was contracted by Pepper to coordinate the work. The members of the Panel are listed in Annex 1.

26. The PRP was asked to review the document: '*Validation report on the Glucocorticoid receptor transactivation method, 24th of September 2024"* (made available on the OECD public website in October 2024), and respond to charge questions (outlined in Annex 2) based on the OECD GD 34: *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment*. The raw data and the results from the 2^{d} phase of the validation study were made publicly available.

27. Each Panel member provided written responses to the charge questions to the PRM by the end of November 2024. PRM provided the PRP with the compilation of the written comments (Annex 3) and highlighted points requiring further discussion or clarification at a teleconference (TC) on 4 December 2024. PEPPER representatives and some of the supporting experts involved in the GRTA validation study (listed in Annex 4) were also invited at this TC to provide specific clarifications requested by the review panel. End-of-review teleconference was held on 14 January 2024 with the presence of the review panel and the review manager.

28. Based on the teleconference discussions, the PRM prepared a draft report that was reviewed and agreed by the PRP via written communication and bilateral clarifying discussions during March 2025.

29. This report presents the summary of the assessment of the validation study report on the GRTA test method and the resulting agreed responses of the PRP to each of the charge questions.

Scientific basis and Relevance to an endpoint of interest (Validation principles 1 and 2)

Charge Question 1: *Do you consider that the rationale for the GRTA test method is clearly elaborated in terms of its scientific basis, regulatory purpose and need?*

30. The written comments reflect agreement that the need for assays which help identification of modulators and disruptors of the GR activated pathways is well justified. At the end-of-review teleconference all reviewers commended the efforts by PEPPER and the GRTA method developers to fill a gap in the regulatory space for endocrine assessment.

31. The rationale for the test method is provided but needs further elaboration. Namely, the scientific support for the antagonist mode of the test (e.g. evidence related to the genetic loss of GR function in rodents and humans) is well elaborated. However, the agonist mode needs further support with examples of health effects arising from overexposure to glucocorticoids (e.g. cases of hypercortisolism and Cushing's syndrome). Reviewers also noted that the authors discuss the rationale in the context of the AOPs available in the AOP wiki, which are all related to GR overactivation rather than inactivation. In this context it was noted that the information from the exiting AOPs on GR activation, but also on other AOPs linked to specific endpoints related to metabolic disorders or cardiovascular disease, is not sufficiently used and could be highlighted better in the validation study.

32. In relation to the rationale for the validation study design, the reviewers find that it is primarily focused around identification of strong GR binders, activators and antagonists that parallel and compete with the synthetic glucocorticoid dexamethasone. Under these circumstances the broad utility of the current GRTA method for identifying interaction with non-pharmaceutical substances is uncertain. As the authors point out at the end of section 7, the main applicability of the GRTA method is expected to be for pharmaceuticals particularly in the steroid family.

33. Related to the above and to the rationale for the scientific basis and the regulatory need, further identification and discussion of the limitations of the GRTA test method is needed to avoid inappropriate expectations and conclusions when using this test to assess potential interaction with the GR pathway for substances with the current protocol.

34. Nevertheless, considering the need for a mechanistic, non-animal, easily transferable and scalable assay to target the GR activation pathway in the assessment of endocrine active substances, the rationale is appropriate to support developing a guidance document based on the current or slightly modified (see below) protocol for the GRTA test method that would lead to the adoption of an OECD standardised test guideline. However, further discussion is needed in the validation study to illustrate the wider scientific context or the GR activation/inhibition pathway beyond substance binding to the GR and more explicit link to relevant AOPs and literature relating to adverse outcomes resulting from both activation and inhibition of GR signalling pathway.

Overall, the peer review panel considers that the Validation Principle 1 has been largely met. The scientific basis of the assay (induction of Luc reporter gene via the GR response promoter element following the cell system's exposure to chemicals activating the GR response pathway) is elaborated. However, the validation study needs to include further discussion on: (i) the wider scientific context i.e. the GR activation/inhibition pathway beyond substance binding to the GR; (ii) examples of adverse outcomes related to overactivation of GR (hypercortisolism (iii) listing all relevant AOPs and literature to illustrate how can the results of the GRTA method be linked to the AOs and/or a higher level testing strategy context; (iii) the limitations of the assay (see CQ2).

Charge Question 2: *Is the relationship between the endpoint(s) measured with the GRTA method (activation of GR or antagonism/interference of GR activation) and the (biological) phenomenon of interest (endocrine activity/disruption) described adequately and using relevant scientific information/references?*

35. The relationship between the endpoint(s) measured with the GRTA method and the (biological) phenomenon of interest is generally well-described. However, more in-depth discussion is needed to support the relevance of the GRTA test method for regulatory purposes. For this, it is necessary to expand the description of the test system and better identify, represent and potentially address the limitations of the assay. To do this it was recommended to expand the background discussion of the Validation Study with:

    - More extensive discussion is needed regarding the limitation of the current assay in relation to identifying weaker agonists given that the system is validated with the strong agonist DEX to set the level of maximal fold of induction rather than the physiological agonist cortisol. The panel also discussed at length the possibility of considering cortisol as a reference agonist/stimulant in the antagonist mode. This change from DEX to cortisol however would require additional testing and validation work. To support the use of cortisol as a good calibrator for both agonist and antagonist mode in this test system, specific examples were discussed with other test systems where cortisol has been used at physiological relevant concentrations with good dynamic response and close to maximal stimulation levels[3].

    - Discussion regarding characterization of the levels and ratios of the expression of the GR in the context of this HeLa-based cell system to relevant co-factors and potential dimerization partners (e.g. oestrogen receptors). This may potentially help interpret some results (e.g. consistent inhibition effect of 17β-estradiol in the antagonist/interference mode).

    - Discussion regarding the availability and function of cellular transporters for different tested substances in different cell test systems.

    - Discussion of other aspects of interference with the GR activated response beyond binding to the receptor, including posttranslational modifications of the receptor.

    - Discussion of the potential limitations of using one particular glucocorticoid response element (MMTV derived) in the Luc promoter.

36. The panel also discussed the uncertainty about the potential cross talk with the Androgen Receptor (AR) activators in this test system. It was noted that the set of reference chemicals did not contain any known AR agonists. After consulting the literature, the reviewers found convincing evidence that HeLa cells do not express AR[4], thus further characterising the GRTA test method by including a specific AR agonist seems not indicated.

37. One reviewer specifically noted that the GRTA assay identifies 17b-estradiol as a potential disruptor of the GR signalling (probably via interaction/crosstalk with ER which is expressed in this HeLa cell-derived HMLN-hGR cell line[5]), while other assays targeting ER antagonist activity could not, thus supporting the GRTA method's utility in the wider context of tiered or IATA approach for assessment of endocrine disrupting chemicals. This also highlights the

[3] Rebuffat et al.(2004) doi: 10.1016/j.mce.2003.11.027.

[4] Krycer JR, (2011) doi 10.1074/jbc.M111.227082 // Mac Namara et al. (1998) doi 10.1007/s002239900487 //Govindan MV and Warriar N (1998) doi 10.1074/jbc.273.38.24439.

[5] Monje and Boland (2002) doi: 10.1002/jcb.10193. PMID: 12112024.

importance of good characterisation of the cell line used in the test system in terms of the expression of the different receptors that may cross-talk in the test system.

The peer review panel agreed that the Validation Principle 2 has been largely met. However, it is important to highlight the limitations discussed above to strengthen the validation study and avoid potential over-interpretations of the results of the GRTA method in the regulatory context in the future.

Technical Characterisation
(Validation principles 3 and 4)

Charge Question 3: *Do you consider that the protocol description in the Validation Study (including introductory sections and SOP) is sufficiently detailed, providing information on all the materials and procedures for data generation and interpretation?*

38. The protocol described in Annex 2 and used for generating results in the Phase 2 is considered the SOP for the assay under validation. The use of drospirenone as a substance expected to decrease GR activity (specifically activated by DEX) in the first transferability phase was questioned as it has not been known to strongly bind to GR.

39. This protocol in Annex 2 was considered elaborate and sufficient to reproduce, however several clarifications were requested and specific modifications were suggested. Some were discussed at the first teleconference of the review panel with the representatives of the participants in the validation study. Special emphasis was put on the SOP/prediction model for antagonist/interference mode and what is interpreted as "interference" in this assay.

40. It was agreed that the following descriptive modifications would improve clarity of the protocol:

    - Represent agonist data as fold increase over negative control in addition to normalized RLU (Relative Light Units) as a percent of maximal responses. In fact, having graphs with second Y (fold induction) and X -axis (micromolar concentration) would greatly improve clarity and comparison of results for different chemicals.

    - Specify cell seeding number (e.g. "Cells should be seeded at 100,000 cells per well in 100 µL assay medium and incubated 16-18 hours at 37°C at which cells should be 80-90% confluent.") to ensure more consistent cell density at start of testing.

    - Specify a highest concentration limit for testing. PRP considered 100 µM or even 30 µM to minimise nonspecific effects/interference that can be induced by high concentrations of test chemicals. It was emphasized that realistic environmental concentrations should be considered. In addition, it should be taken into account that many of the lipophilic compounds are not soluble at 100 µM and this may even result in underestimation of cytotoxicity in the range finding runs.

    - I agree with Steven, 30 uM will be hard to sell, but why 100 uM? For every substance we would need to first look at expected exposure scenario. I would suggest 30 uM, and see what happens.

    - Consistently use the wording for the criterion for positive agonist (differed across the different sections of the validation study) and clarification whether the comparison point is basal level or maximal level produced by DEX. Reviewers recommend clearly specifying criteria in relation to the maximal value of stimulation (currently 100% of DEX response), i.e. 10% or above the maximal DEX response. This may need to be adjusted in the context of using cortisol as a reference agonist but would require further validation work to support a new criterion for classification of a positive response.

    - Modify the Acceptance Criteria (AC) for basal activity (now defined as "≤ 4% of maximum luminescence" where the average luciferase activity of the positive control (DEX 10-7 M) must be at least 10 times that of the solvent control), based on:

- the typical background luciferase activity for the solvent control and what labs typically measure for mean max fold induction for DEX relative to the solvent control.
- a z-factor (https://en.wikipedia.org/wiki/Z-factor) which uses the central tendency (mean, median) and the variation (standard deviation, median absolute deviation) around both the basal and maximal responses to formulate a single, unitless value to evaluate the assay's ability to effectively distinguish positive and negative responses in each experimental run.

At the first teleconference PEPPER informed that they are currently performing such analyses and will be able to provide the results to inform appropriate modification of the criteria.

- Provide more guidance and background on using the 4-parameter (or 3-parameter) logistic regression for concentration response modelling. Consider additional references (e.g. De Lean et al.[6]). It was clarified by PEPPER that parametric or non-parametric tests are not required even in cases when the Hill function cannot be fitted.

41. The following modifications regarding the prediction model are suggested to potentially improve predictivity, particularly for weak agonists and antagonists which are currently underrepresented in the validation study due to limited published information. These suggestions require additional analysis without additional testing and would strengthen the validation study:

- Regarding cytotoxicity: In their written comments, reviewers noted the solid effort of the developers and the comprehensive detailing of the protocol in terms of determining testing concentrations to avoid effects of cytotoxicity. The combination of visual inspection with the MTT assay for evaluating cell viability/cytotoxicity was appreciated. At the first TC it was agreed that further analysis on the variability of cytotoxicity with the vehicle controls is needed to determine if the current threshold <20% cytotoxicity is appropriate or needs to be lowered particularly for antagonism mode. PEPPER clarified that they are undertaking this analysis and will provide it in the updated Validation Study after the review. Following additional discussion by the review panel it is recommended that the cytotoxicity threshold for positive antagonist should be lowered to 10% unless the IC30 is used instead of IC20 to define positive antagonist using the current protocol where DEX is used as a stimulant. These changes should help better delineate the specific signal decrease due to antagonism from any confounding cytotoxicity.

- Regarding the criteria for positive agonist: The review panel discussed whether the particular criterion for positive agonist: "above 10% of the maximal DEX activity" (where basal activity should not exceed 4% for qualifying run)" may be too high for weak agonists. At the first TC it was concluded that under current circumstances, with the reference chemicals used (mostly strong agonists/pharmaceuticals) generating results where most observations were either flat (non-responsive) or well above 50% of maximal DEX induction, there is not sufficient information to recommend changing this criterion at present. However, as more chemicals are tested (some recommended under CQ5) this criterion could be adjusted in the future. However, it would be necessary to examine the consistency of response for weak agonists with the assay. In a situation of limited availability of known weak agonists, this could be done retrospectively by looking back at the consistency of responses for low concentrations of strong agonist that produce a threshold (>10% but < 20%) effects. Such analysis could also support the current criterion for a positive in the agonist assay.

[6] De Leon et al., (1987) doi: 10.1152/ajpendo.1978.235.2.E97.

42. The use of the 4-parameter logistic regression analysis and how it was implemented in the data processing system was well covered by the presentation on data integrity and this approach was generally considered as adequate (see discussion under CQ7 and 8). However, upon further discussion at the second TC the panel considers that it is more appropriate to use a the 3-parameter logistic regression to potentially increase consistency between different labs. The 4th parameter is likely not needed considering the low level of basal activity that is permitted (4%) and for more accurate EC/ICx estimates. The concern is that for hanging-curves the 4-parameter logistic regression will set the y-intercept far above where it should be[7].

43. The consideration about the implications of maximum testable concentration and also about the acceptance criteria for cytotoxicity in antagonist mode, initiated a robust discussion about what is meant by "interference" with the GRTA test method. At the first TC, PEPPER clarified that the interference mode testing was particularly introduced to distinguish decrease of luminescence signal as a result of processes (be it indirect, unspecific, off target interferences) other than those due to competitive antagonism, because in the literature many substances may have been "called/assigned" antagonists even though the corresponding assays could not distinguish the precise mechanism.

44. Reviewers asked whether methods other than the R^2 method were considered to distinguish antagonism from interference. PEPPER clarified that the subjective methods of fold difference (transferability phase) and "parallel/not parallel curves were considered but despite the unpopularity of the R^2 method due to its assumptions and limitations recognised by statistical experts, it was reluctantly accepted as an objective and clear-cut method for the validation study.

45. Discussion included examples of other similar TA assays where interference with the assay (not the reporter protein function) was a result of "off target" processes that are being initiated within the cell test system by the test chemical (e.g. genistein for ERTA, and other lipophilic substances) and which are (indirectly) relevant to the activity of the test substance, even though not by the mechanism primarily pursued by the tests system (e.g., agonism and/or antagonism of the target receptor). Reviewers also informed on an existing analysis of a data set of over 600K drug-like compounds tested for interference with firefly luciferase enzyme activity and found that only 0.1% of substances did show some interference[8]. The analysis strongly suggests that in the set of about 40 tested chemicals in the GRTA method validation study, it is unlikely that any of the observed decrease of luminescence is a result of interference with luciferase function itself, but more likely linked to some other aspect of the GR activity/signalling or perturbances of the cell system related to the exposure to test chemical and reflected in the activity of the GR responsive promoter element used.

46. Taking all into account, one reviewer suggested that as long as the cell viability issue is properly addressed (which the protocol does adequately) the conclusion of active/positive antagonist response would be acceptable as a result from this test system when clear dose dependent decrease in the luciferase signal is observed in the absence of any overt cytotoxicity. Even though the exact mechanism of assay signal decrease would not be known, the information would be valuable in the context of integrated assessment of endocrine active substances. The test system as used now is probably not optimal to identify competitive antagonists from other mechanisms of interference with GR activity. A note was made that under the current protocol using the particularly strong agonist DEX, the decrease of the assay signal is most likely due to competitive antagonism, so the additional interference mode testing is not adding significant additional information.

[7] https://www.graphpad.com/guides/prism/latest/curve-fitting/reg_50_of_what__relative_vs_absolu.htm
[8] Auld et al. (2009) doi: 10.1021/jm8014525

47. The issue was further discussed at the second TC of the review panel and it was agreed to not recommend that the R^2 method for distinguishing positive antagonists from nonspecific interference with the Luc signal and the interference mode should be omitted from the protocol.

48. If additional testing could be done, the antagonism mode of the assay could be performed using two concentrations of dexamethasone (current 80% saturation and another lower saturation concentration, e.g. 50%) or the physiological agonist cortisol.

49. Overall, it was concluded that there is no statistical approach under the current experimental design could distinguish the different types of mechanisms involved in potential interference/modulation/disruption of GR signalling but it is important to have a transferable, reproducible and robust assay that can detect and distinguish activation/interference/modulation/disruption of GR signalling for a diverse set of chemical substances. For this it is important to have appropriate orthogonal assay information such as cytotoxicity and crosstalk with other receptors expressed in the test system that bind to the particular reporter promoter.

50. The use of an AR agonist (e.g. 17 methyl testosterone) was discussed again as potential specificity control regarding crosstalk with AR receptor. It was noted that authors do not characterise the status of the AR receptor in the HMLN-hGR cells so such a control would be informative, given that both AR and GR bind nearly identical response elements, and that GR activation is a known confounder in ARTA methods. However, it was pointed out that HeLa cells do not express AR[4] and such control is not considered necessary for the validation study.

Overall, the peer review panel agreed that the Validation Principle 3 has been met. However, several recommendations are provided for specific clarifications and potential improvement of the protocol. It is understood that this kind of assays are intended to be used together with other assays in the context of tired and/or integrated testing strategies to assess and link endocrine activity and potential endocrine disruption. Therefore, it is not expected that this cell-based system should cover all possible mechanisms of GR activity/modulation/disruption. However, good characterisation of the cell line is needed as well as adequate controls, including properly addressing the ratio between the decrease of Luc signal and cytotoxicity, which this protocol can provide with the suggested modifications.

Charge Question 4: *Do you consider that the intra-, and inter-laboratory reproducibility of the GRTA test method results is adequately demonstrated? Please consider the degree to which biological variability affects the test method reproducibility.*

51. The intra- and inter-laboratory reproducibility of the GRTA test method is adequately demonstrated in the validation report, with high consistency across multiple runs and laboratories. The Coefficients of Variation (CV) is mostly below the acceptable limits (<15%), ensuring the OECD GD34 standards for reproducibility. These results suggest that the provided protocol is robust and could be utilized across laboratories.

52. The validation study demonstrates that the reproducibility is very good for detecting agonists. Although it was noted by some that interlaboratory variability of EC50 is greater for some agonists (e.g. aldosterone, budesonide, cortisol).

53. The variability of the results is higher for the antagonist/interference mode. This is probably due to the current methodology applied to the prediction model (R^2 criterion) and the difficulty in making the final call for interference versus "true" antagonism using the R^2 approach, which reviewers suggest should be omitted from the protocol (discussion under CQ3). It was noted that the literature regarding the R^2 criterion was limited and not popular even with PEPPER's statistical experts (discussed at the first TC).

54. At the first TC additional analysis of the variability of the background control and variability of the fold induction/inhibition over time in the different labs was requested. PEPPER informed that the analysis is undergoing and will be included in the updated validation study following the review.

55. It is highlighted that the validation study includes mostly strong agonists and antagonists and substances with clear negative results so the variability in the area of the weakly active substances may have higher variability. Thus, it is important to evaluate more such substances in the future (further discussed under CQ5).

The peer review panel agreed that the Validation Principle 4 has been met. Reproducibility is expected to be further improved after implementation of the modifications of the protocol recommended under CQ3.

Fitness for Purpose
(Validation principles 5 and 6)

Charge Question 5: *Is the assessment of the test method's performance based on the testing of relevant reference chemicals that are representative of the types of substances for which the test method is intended to be used?*

56. In the written comments the panel had diverging opinion on the balance of positive versus negative reference compounds used to evaluate the method's performance. Some thought that the ratio of negative chemicals should be increased, others that there should be a greater proportion of positive chemicals and with greater range of potencies (addition of weak and moderate potency chemicals). There was general agreement though that the clearly positive substances (both agonists and antagonists) were only strong pharmaceutical chemicals related to steroidogenic chemicals.

57. At the first TC reviewers again recognised that finding a perfectly balanced set of reference chemicals for GR activity is difficult. The importance of including more moderate and weak agonists and antagonists was emphasized, particularly for distinguishing true antagonists from substances interfering with the GR signalling by another mechanism, but also for the activity spectrum of the assay in general.

58. The representatives for the validation (PEPPER and VMG) clarified that the search for GR active substances was based on a wide literature search, as there are really not many known GR active substances, outside of the pharmaceutical steroids. PEPPER emphasized that the decision to include a GRTA method in their work plan was made precisely taking into account this issue which in a way may have led to the obvious regulatory gap for assays addressing GR active/modulating substances. Therefore, the first action of the validation study was a literature search on data for GR activity of different substances to generate the list for testing substances i.e. reference substances for the study. For the substances on the initial list, additional analysis was performed to examine ToxCast data and QSAR predictions (Section 6 of the validation study).

59. It is pointed out that the initial search for reference chemicals was done more than 2 years ago and updating the literature search may be useful for identifying additional relevant substances to test.

60. A wide analysis of the ToxCast data from three agonist mode (human) GR Trans Activation assays (ATG_GR_TRANS_up; ATG_GRE_CIS_up; and TOX21_GR_BLA_Agonist_ratio) and one antagonist mode assay (TOX21_GR_BLA) was provided by one reviewer. List of substances not active in any of these ToxCast assays was also provided. The panel agrees that this analysis may help identify potential new candidate substances to test within the mid-range activity level in future optimisations of the method. The results are available in the xlsx file (Annex 5) and can be used in next steps of the progress of the GRTA method project.

61. In addition, considering the evidence[9] for some complex substances (e.g. Genistein and Compound A) and potentially including them in future testing was recommended. Even though these substances are not known to specifically/directly induce GR activity by binding to the receptor, the evidence may be informative in interpreting results/"calls" with the reference substances or other test chemicals in the future.

[9] Rengfei Shi et al. (2014) https://doi.org/10.1016/j.neuint.2013.12.002 // Whirledge S, et al. (2015) doi: 10.1289/ehp.1408437 // De Bosscher K, et al. (2005) doi: 10.1073/pnas.0505554102 // Beck IM, et al.: (2013) doi: 10.1371/journal.pone.0069115 // Rebuffat AG, et al (2004) doi: 10.1016/j.mce.2003.11.027.

Overall, the peer review panel agreed that the Validation Principle 5 has been largely met particularly considering the scarcity and complexity of the evidence related to GR activity of non-steroid and non-pharmaceutical substances. However, given that currently there is no established reference data set of environmentally relevant GR active substances, the GRTA test method would need to be periodically reviewed. For the near future, additional analysis and testing of relevant substances over a range of potencies is needed. To facilitate identification of additional substances, the review panel has provided an initial analysis (Annex 5) of potentially relevant substances to analyse and test. Further analysis and testing of relevant substances over a range of potencies will also help to support the fitness for purpose of the assay beyond pharmaceutical and steroid substances.

Charge Question 6: *Has the performance of the test method been evaluated in relation to relevant information from the species of concern, and all existing relevant toxicity testing data for relevant reference chemicals? The panel is encouraged to advise on any other relevant data and/or analysis.*

62. The reviewers acknowledge that the toxicity data used to evaluate performance of the GRTA test method is focused only on human relevance. However, they, as the authors of the validation study, recognise that the results reported in literature are highly heterogeneous due to the use of different cell systems, different media, concentrations, and assay protocols. More research and analyses are needed to determine the sources of the heterogeneity of results. Additionally, one reviewer recommended to assess homology between functional domains of the GR in different species and a review of the literature comparing relative ligand binding affinities between the receptor in the GRTA and aquatic vertebrates. Nuclear receptors, including the GR, are highly conserved across different species sharing a significant degree of structural and functional similarity, particularly in their DNA-binding domain and ligand-binding domain, which are the most conserved regions within the protein family[10]. The largest differences exist in the AF1 and AF2 regions where the coactivator and corepressor proteins are binding. Therefore, it would be useful to determine if the results from this assay are amenable to cross species extrapolation[11]. This may also provide some information for explaining differences in the different testing systems in the literature used to derive the reference chemicals and the reference activity 'calls' in the different studies. Though it was recognised that this may be out of the scope of the validation study, it may ultimately be useful to better define the applicability domain for the test method for future Test Guideline.

63. Short term and for updating the Validation Study, it required that results obtained in the validation study should be clearly compared with the starting hypothesis for the reference chemicals outlined in Table 7, particularly after finalising results based on the approach for identifying antagonists (see discussion under CQ3).

64. In addition, investigation of most recent (past the initial literature search) and relevant literature for the chemicals identified by the ToxCast analysis provided by reviewers (see CQ5) would be useful as well as additional examination of epidemiological studies to compare the outcomes of the GRTA method and observed endocrine-related health effects in humans. Better link to AOP context discussed under CQ1 would also be helpful in this respect.

65. It was also suggested that employing *in silico* QSAR models (e.g. in OECD QSAR Toolbox, US EPA Endocrine Disruptor Knowledge Base (EDKB), and SwissADME/SwissTargetPrediction) can provide valuable predictive insights into GR binding and activation for chemical classes underrepresented in the validation study. However, it was also noted that despite the rapid improvement in the molecular modelling approaches, their use in predicting GR activity so far has been limited to date.

The peer review panel agreed that the Validation Principle 6 has been largely met. However, additional investigation of the most recent literature (past two years) and literature specifically relevant to the chemicals identified by the additional ToxCast analysis (see CQ5) would augment the value of the validation study and the usefulness of the GRTA assay in the wider regulatory context. Other relevant evidence (e.g. epidemiological evidence, read-across approach, pathway analyses and novel QSARs), could also be explored to identify additional relevant substances for testing and optimising the GRTA test method in the future.

[10] Robinson-Rechavi et al., (2001) doi: 10.1016/s0168-9525(01)02417-9 // Bertrand et al., (2004) doi.org/10.1093/molbev/msh200
[11] LaLone et al., (2018) doi: 10.1021/acs.est.8b04587

Data integrity and transparency (Validation principles 7 and 8)

Charge Question 7: *Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP. Assess the goodness of the experimental/validation practices (for example: the coding of test chemicals for blind testing and analysis of test results, the acceptance criteria, the qualitative and quantitative comparison in and between labs, or any others you considered relevant) under which the data have been generated for the validation study of the GRTA test method*

66. The validation study does not specify whether the participating labs applied GLP and during the initial review the reviewers had different perceptions as to whether data have been obtained according to GLP. Based on the review comments a representative from PEPPER and the Data Manager for the validation study were invited to a teleconference to provide more information at on quality control measures that were used for data generation and analysis (e.g., auditing the raw data and the methodology for dose response modelling, calibration standards, reagent lot tracking, and personnel training).

67. Data Manger and PEPPER director presented on the processes (Annex 6). PEPPER director also informed that the labs and the data manger analysed the data independently. The results were the same, providing additional control of the integrity of the data in the process. In addition, it was clarified that the alternative non-parametric approaches for response modelling mentioned in the report, were not applied to data analysis in the validation study as they were not needed. They were provided as optional alternatives if data distribution requires.

68. The review panel appreciated the presentation and acknowledged that the system employed is robust and allows traceability of the data. In addition, the test chemicals in Phase 2 were coded and blind tested in accordance with the GD34 principles.

69. However, better description of the process and the specific roles and responsibilities of the participants (validation plan) was requested for the updated Validation Study.

70. Ultimately, it was also noted that the high BLR reproducibility demonstrates the robustness of the assay system.

The peer review panel agreed that the Validation Principle 7 has been met.

Charge Question 8: *Do you consider that all the data supporting the assessment of the validity of the test method are easily available for expert review? These include: a detailed and readily available test method protocol to the public and independent laboratories; organised and easily accessible data to permit independent review(s); benchmarks by which an independent laboratory can itself assess its proper adherence to the protocol?*

71. All data were made publicly available within Excel file. However, it was not easy to perform independent re-analyse without additional guidance. In addition, a there was a lot of missing data in the Figures and Tables for phase II (mainly in interference data tables) that may need to be provided in the revised validation study.

72. At the first TC Data Manager and PEPPER Director presented how the data were collected and analysed (Annex 6). Presentation will also be attached to the updated validation study to further ease the independent analysis of the data.

73. Additional analysis of the variability of the background control was requested and variability of the fold induction/inhibition over time in the different labs was requested. (discussed also under also CQ3).

74. A comment was made that long term and affordable availability of the cell line should be ensured.

The peer review panel agreed that the Validation Principle 8 has been partially met.

Overall Conclusions

75. The review panel was also asked to consider the validation study overall addressing the following charge question: *Taking into account your overall review, do you consider that the validation study and the supporting references provide appropriate evidence to inform testing strategies, guidance and/or guidelines for application of the GRTA method to assessment of ED toxicity of chemicals. Reviewers are encouraged to provide any specific recommendations that they may have for future near- and long-term application and potential development of the GRTA method.*

76. The panel considered that the validation study was conducted in accordance with the principles and requirements of OECD GD 34 and aligns well with the approach for its ongoing revisions[12].

77. Specifically, the protocol is well described, even though some modifications/corrections are recommended (see CQ3). Most notably a simplification of the prediction model for the antagonist mode and inclusion of cortisol as a reference 'stimulant' and consistent accurate language for the identification of a positive chemical in the agonist mode are proposed. Use of cortisol as a reference active chemical in the agonist and antagonist modes, alongside DEX as a positive control in the agonist assay, is also recommended.

78. The assay, even using the current protocol's prediction model, is highly reproducible. However, only potent positive substances were generally tested, questioning the performance for substances with intermediate activity in the agonist mode. Thus, despite the fact that performance cannot be well evaluated given the recognised uncertainties for the GR activity potential for most non-pharmaceutical substances studied in the literature, the GRTA test method may be useful for testing in a regulatory context as a screening level assay and as part of testing strategies based on the tiered approaches described in other guidance documents (e.g. GD150). This limitation is not unique to the GRTA test method and it is similar to other TA assays used in the context of assessment of endocrine active substances.

79. It was recognised that the choice of reference chemicals for the validation was a difficult task since only a few chemicals (mainly pharmaceutical steroids) have convincing and reliable data to aid assessment of predictive capacity of GR activity in the agonist and antagonist modes.

80. Specific recommendations for investigating the literature and potential testing of additional substances are provided based on the analysis performed by a review panel member.

81. Overall, the assay addresses gap in the context of the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters and development of a Guidance Document over a Test Guideline at this time. A Test Guideline can be supported by a revised Validation Study after including all the recommended modifications and additional analyses. PRP is of the opinion that it is important to enable active use of the assay according to standardised protocol within a format of a Guidance Document and generate more real-world data that would support future development of a Test Guideline for wider regulatory context.

[12] van der Zalm et al. (2022) https://doi.org/10.1007/s00204-022-03365-4

References

Auld et al. (2009) A Basis for Reduced Chemical Library Inhibition of Firefly Luciferase Obtained from Directed Evolution. Journal of Medicinal Chemistry 52 (5), 1450-1458, DOI: 10.1021/jm8014525

Beck IM, et al.: Compound A, a selective glucocorticoid receptor modulator, enhances heat shock protein Hsp70 gene promoter activation. PLoS One. 2013 Jul 30;8(7):e69115. doi: 10.1371/journal.pone.0069115. PMID: 23935933; PMCID: PMC3728325.

De Bosscher K, et al.: A fully dissociated compound of plant origin for inflammatory gene repression. Proc Natl Acad Sci U S A. 2005 Nov 1;102(44):15827-32. doi: 10.1073/pnas.0505554102. Epub 2005 Oct 21. PMID: 16243974; PMCID: PMC1276063.

De Leon et al., (1987) Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curved. American Journal of Physiology, 235, E97-102.

Krycer JR, J Biol Chem 2011, p20637, doi 10.1074/jbc.M111.227082. ; Mac Namara P and Lougrey HC Calcif Tissue Int 1998, p39, doi 10.1007/s002239900487.; Govindan MV and Warriar N, J Biol Chem 1998 p24439, doi 10.1074/jbc.273.38.24439.

LaLone CA, Villeneuve DL, Doering JA, Blackwell BR, Transue TR, Simmons CW, Swintek J, Degitz SJ, Williams AJ, Ankley GT. Evidence for Cross Species Extrapolation of Mammalian-Based High-Throughput Screening Assay Results. Environ Sci Technol. 2018 Dec 4;52(23):13960-13971. doi: 10.1021/acs.est.8b04587. Epub 2018 Nov 13. PMID: 30351027; PMCID: PMC8283686.

Monje P, Boland R. Expression and cellular localization of naturally occurring beta estrogen receptors in uterine and mammary cell lines. J Cell Biochem. 2002;86(1):136-44. doi: 10.1002/jcb.10193. PMID: 12112024.

OECD. (2005). "OECD Series on Testing and Assessment, Number 34. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment." Available at: https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-gd34.pdf

OECD (2018) Revised Guidance Document 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.

Rebuffat AG, Tam S, Nawrocki AR, Baker ME, Frey BM, Frey FJ, Odermatt A. The 11-ketosteroid 11-ketodexamethasone is a glucocorticoid receptor agonist. Mol Cell Endocrinol. 2004 Feb 12;214(1-2):27-37. doi: 10.1016/j.mce.2003.11.027. PMID: 15062542.

Shi R, Wang S, Qi X, Chen S, Chen P, Zhang Q. Lose dose genistein inhibits glucocorticoid receptor and ischemic brain injury in female rats. Neurochem Int. 2014 Jan;65:14-22. doi: 10.1016/j.neuint.2013.12.002. Epub 2013 Dec 12. PMID: 24334056.

Robinson-Rechavi M, Carpentier AS, Duffraisse M, Laudet V. How many nuclear hormone receptors are there in the human genome? Trends Genet. 2001 Oct;17(10):554-6. doi: 10.1016/s0168-9525(01)02417-9. PMID: 11585645.

Stéphanie Bertrand, Frédéric G. Brunet, Hector Escriva, Gilles Parmentier, Vincent Laudet, Marc Robinson-Rechavi, Evolutionary Genomics of Nuclear Receptors: From Twenty-Five Ancestral Genes to Derived Endocrine Systems, Molecular Biology and Evolution, Volume 21, Issue 10, October 2004, Pages 1923–1937, https://doi.org/10.1093/molbev/msh200

van der Zalm, A.J., Barroso, J., Browne, P. et al. A framework for establishing scientific confidence in new approach methodologies. Arch Toxicol 96, 2865–2879 (2022). https://doi.org/10.1007/s00204-022-03365-4

Whirledge S, Senbanjo LT, Cidlowski JA. 2015. Genistein disrupts glucocorticoid receptor signaling in human uterine endometrial Ishikawa cells. Environ Health Perspect 123:80–87; https://doi.org/10.1289/ehp.1408437

Annex 1 – Peer review Panel

Expert Name	Affiliation	Representing
Alex Odermatt, PhD	University of Basel, Basel, Switzerland	Switzerland
Hajime Kojima, PhD	National Institute of Health Sciences (NIHS) & Sanyo-Onoda City University	Japan
Shagun Krishna, PhD	The Physicians Committee for Responsible Medicine - PCRM	ICAPO
Steven L. Levine, PhD	Bayer Crop Science, Chesterfield, USA	BIAC
Steven O. Simmons, PhD	US EPA Office of Research and Development	USA

Peer Review Manager Julija Filipovska, PhD, ERT - Independent Consultant

Annex 2 – Charge Questions

Charge Questions for the Peer Review of the Validation Report on the GRTA assay: a Glucocorticoid receptor Transactivation Assay

Questions based on the validation principles and criteria in OECD GD34 (link)

Scientific basis and Relevance to an endpoint of interest
CQ1. Do you consider that the rationale for the GRTA test method is clearly elaborated in terms of its scientific basis, regulatory purpose and need?
[Your review comments]
CQ2. Is the relationship between the endpoint(s) measured with the GRTA (activation of GR or antagonism/interference of GR activation) and the (biological) phenomenon of interest (endocrine activity/disruption) described adequately and using relevant scientific information/references?
[Your review comments]
Technical Characterisation
Q3. Do you consider that the protocol description in the Validation Study (including introductory sections and SOP) is sufficiently detailed, providing information on all the materials and procedures for data generation and interpretation?
[Your review comments]
Q4. Do you consider that the intra-, and inter-laboratory reproducibility of the GRTA test method results is adequately demonstrated? Please consider the degree to which biological variability affects the test method reproducibility.
[Your review comments]
Fitness for Purpose
For the following two CQs you may wish to consider the particular aspect of GD34 (Para 33) that is applicable to GRTA test method: *"For cases in which a new test method addresses an endpoint that has not been previously considered or for which there are no or inadequate established reference data, a rigorous assessment of the predictive capacity may not be possible, however, the relevance should be assessed using all available information. Such tests that have shown to be reliable, but whose relevance has only been partially assessed by comparison with related reference data (e.g., TG 414 (14) and TG 424 (15)), should preferably have their validation status periodically reviewed, with full consideration of all relevant available data."* Q5. Is the assessment of the test method's performance based on the testing of relevant reference chemicals that are representative of the types of substances for which the test method is intended to be used?
[Your review comments]
Q6. Has the performance of the test method been evaluated in relation to relevant information from the species of concern, and all existing relevant toxicity testing data for relevant reference chemicals? The panel is encouraged to advise on any other relevant data and/or analysis.
[Your review comments]
Data integrity and transparency
Q7. Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP. Assess the goodness of the experimental/validation practices (for example: the coding of test chemicals for blind testing and analysis of test results, the acceptance criteria, the qualitative and quantitative comparison in and between labs, or any others you considered relevant) under which the data have been generated for the validation study of the GRTA test method.

[Your review comments]
Q8. Do you consider that all the data supporting the assessment of the validity of the test method are easily available for expert review? These include: a detailed and readily available test method protocol to the public and independent laboratories; organised and easily accessible data to permit independent review(s); benchmarks by which an independent laboratory can itself assess its proper adherence to the protocol?
[Your review comments]

Overall Review
Taking into account your overall review, do you consider that the validation study and the supporting references provide appropriate evidence to inform testing strategies, guidance and/or guidelines for application of the GRTA test assay to assessment of ED toxicity of chemicals. Reviewers are encouraged to provide any specific recommendations that they may have for future near- and long-term application and potential development of the GRAT assay.
[Your review comments]

Annex 3 – Compilation Responses

Peer Review of the Validation Report on the GRTA Assay

September-November 2024

Scientific basis and Relevance to an endpoint of interest	
CQ1. Do you consider that the rationale for the GRTA test method is clearly elaborated in terms of its scientific basis, regulatory purpose and need?	**PR1** The rational to apply this test to identify substances that bind to the GR ligand binding pocket and compete with the synthetic glucocorticoid dexamethasone is clearly described. However, the discussion of the scientific basis and regulatory value should be extended. What is missing is a critical description of the limitations of this assay and that this assay will cover only a very small fraction of the substances that interfere with GR activity. These limitations are important to consider in order to avoid inappropriate expectations and conclusions when using this test to assess potential toxicity of substances.
	PR2 I agree the needs for the evaluation of GR activation as one of the endocrine screenings for chemicals.
	PR3 The first question this reviewer addressed was whether the validation process in OECD GD 34 was followed? Generally, the validation process in OECD GD 34 was followed. The current steps in the validation process outlined in GD 34, are as follows. Develop assay →optimize assays →standardize and refine SOPs → relevance evaluation → intra- and inter-lab transferability study → pre-validation study → independent review of results → another relevance evaluation → fully assess reliability and relevance (validation study) → finalization of test protocol → another independent review (this OECD review of the validation report) → Recommendations for against proposed regulatory use → develop TG for use under MAD. Note: "Phase 2" is the term used in the report to refer to the validation study. The scientific rationale for the GRTA is clearly elaborated in terms of its scientific basis in the Introduction. However, after reading the validation report this reviewer questions the broad application (relevance) and utility of the GRTA for regulatory purposes particularly in antagonist mode. Very few substances tested in the validation were shown to be agonists and antagonists except for a few pharmaceutical or pharmaceutical-like substances Nevertheless, this reviewer sees the utility of this assay for early non-regulatory screening purposes. As the authors point out at the end of section 7, the main applicability of the GRTA is expected to be for pharmaceuticals particularly in the steroid family. Therefore, I question broad relevance of this assay for adoption as an OECD test guideline. Rather, I can see the validated methodology for this assay being captured in an OECD guidance document rather than an adopted OECD test guideline.

	This would still allow the methodology to be available to the regulatory and research communities when there is scientific rationale to conduct this mechanistic assay to help inform an endocrine assessment. Additionally, as a guidance document, it would allow for more rapid updates to the methodology once those are recommended.
	PR4 A clear rationale was provided supporting the antagonist test mode only. The health effects described in Section 1 distilled to (1) the genetic loss of GR in knockout mice, and (2) inactivating GR gene (NR3C1) mutations underlying Chrousos syndrome, both of which involve loss of GR function akin to that which might occur with anti-glucocorticoid exposure. No health effects case was made in Section 1 (or elsewhere) to support the agonist test mode, which is strange given the ample literature on hypercortisolism and Cushing's syndrome that arises from overexposure to glucocorticoids.
	PR5 The GRTA test method is well-documented in the validation report. It clearly addressing the need for alternative methods to identify endocrine-disrupting chemicals (EDCs). Its scientific foundation is based on the essential biological roles of the human glucocorticoid receptor (hGR), which include development, metabolism, stress response, and cardiovascular regulation. This assay provides a mechanistic approach of evaluating drug toxicity by directly targeting pathways involved in endocrine disruption, by analyzing agonistic and antagonistic activity. The GRTA is applicable for global regulatory initiatives, for example EU REACH and the OECD EDSP as, from a regulatory standpoint, it supports global initiatives under OECD standards to substitute human-relevant, high-throughput techniques for animal testing. By using stable transfected human GR cell lines, this assay circumvents the major problems of traditional animal models, improving scalability and biological relevance. Though, it would be beneficial to show a clear association between the results of the GRTA analysis and the adverse outcome pathways (AOPs), especially those related to cardiovascular toxicity and/or metabolic disruption. Furthermore, its role within tiered regulatory strategies might also be supported by including chemical prioritization and decision-making frameworks.
CQ2. Is the relationship between the endpoint(s) measured with the GRTA (activation of GR or antagonism/interference of GR activation) and the (biological) phenomenon of interest (endocrine activity/disruption) described adequately and using relevant scientific information/references?	**PR1** The endpoints measured with the GRTA is well described (agonism/antagonism/interference). The description of interferences of substances with GR activity that might lead to endocrine disruption is limited. This system has several limitations that narrow down the value for its use for regulatory purposes if these are not mentioned. 1) As ligand, one of the most potent glucocorticoid dexamethasone is used. This is a good positive control for maximal activation. To detect antagonists, the use of the physiological substrate cortisol may have been more adequate. It binds slightly different, has a higher Kd and would require lower concentrations of antagonist X to compete with the ligand. 2) This assay is very well described and easy to use; however, it depends on one cell line, with a cancer background. Whilst steroids obviously are able to enter the cells and reach the ligand binding

	pocket, we cannot know whether this is also the case for substance X. Each cell line has a different set of transport proteins that might either efflux a given compound or take it up. We cannot estimate the intracellular concentration of substance X. There may be efflux in one cell type but not another. 3) One could assume that most substances that interfere with GR activity will not directly bind to the ligand binding pocket but affect GR activity by altering post-translational modifications, thereby altering the sensitivity of the receptor. This is mostly not covered by this assay or then dependent on the pathways expressed and active in HeLa cells. 4) Due to overexpression of GR, the ratio of receptor to coactivators and corepressors is not representing physiology. Also formation of heterodimers, for example with MR, is not covered. 5) MMTV promoter differs from other promoters. Substance X may cause cell-type specific effects on GR-mediated gene transcription and this is not covered by any of the TAs that are available. The ultimate goal for regulators should be to have a set of assays that have a high predictivity for substances causing endocrine disruption by altering GR activity. This assay is a very useful addition and it is correctly concluded that mainly steroids and pharmaceuticals with high affinity to bind GR will be detected. It would be useful to add a paragraph on the limitations of the assay to avoid too high expectations. On page 5 "TA assays…they represent the cellular context of the hormonal system", this statement should be questioned.
	PR2 I think this assay is valuable for the evaluation of GR activation, because the 17b-estradiol is able to evaluate regardless of the other ER assays couldn't.
	PR3 The report provides a cursory overview between the endpoint(s) measured with the GRTA and the (biological) phenomenon of interest (endocrine activity/disruption) using scientific information/references. It would have been more impactful for the report to: (1) have a more in-depth discussion justifying the relevance of the GRTA for regulatory purposes and (2) provide appendices for the most relevant AOPs (rather than having to look them up) to provide the reviewers with a more in-depth understanding between GRTA agonism and/or antagonism and adverse outcomes. Nore none of the AOPs cited are OECD endorsed AOPs. From my review the cited AOPs refer to GR activation not antagonism – questioning the utility of the antagonist assay.
	PR4 The MMTV promoter used in the luciferase reporter has response elements that can be bound by other steroid hormone receptors including estrogen receptor (ER) and androgen receptor (AR) in addition to GR. While it is true that ER is not expressed in HeLa cells, AR is expressed at low, but detectable levels in HeLa cells. This creates the possibility of false positive results in agonist test mode. The possibility could have been addressed in the study with the addition of a strong AR agonist such as 5α-dihydrotestosterone (DHT) as a test item. The lack of any AR agonist in the test item set was curious as it the effort was made to include nearly every other hormone agonist

	class including estrogen, progesterone, and mineralocorticoid. This issue primarily impacts the relevance of agonist test mode.
	PR5 The provided report effectively summarizes the relationship between the endpoints measured in the GRTA assay i.e. activation of GR or antagonism/interference of GR activation and the biological phenomena of endocrine activity and disruption. It discusses the role of GR as a nuclear receptor, which is involved in regulating crucial physiological processes, spanning from metabolism, immune response, cardiac function, to stress adaptation. For example, glucose homeostasis and lipid metabolism are affected by GR activation, however different metabolic syndromes and cardiovascular diseases are linked with GR dysregulation. Several stress response pathways are also affected by GR inhibition, and it could lead to adverse effects on immune function and endocrine signalling. These functions of GR also play a major role in endocrine disruption, as chemicals that modulate GR activity often initiates downstream biological effects that overlaps with known adverse outcome pathways (AOPs) linked to metabolic disorders or cardiovascular toxicity. While the validation report appropriately references scientific literature to validate these mechanisms, its regulatory context could be enhanced by linking GRTA endpoints to key regulatory frameworks, such as OECD guidance on endocrine disruptor testing, and by elaborating on how assay results integrate into broader chemical risk assessments and/or tiered testing strategies. This would enhance the assay's relevance and utility in regulatory decision-making.
Technical Characterisation	
Q3. Do you consider that the protocol description in the Validation Study (including introductory sections and SOP) is sufficiently detailed, providing information on all the materials and procedures for data generation and interpretation?	**PR1** The protocol is nicely described and should be easy to reproduce by any laboratory with some expertise in TAs.
	PR2 I welcome the revision SOP at the end of phase 2. In addition, I have some concerns for the protocol. 1) Abbreviations and glossary: I cannot find “DEX”. 2) Acceptance criteria : Regarding “The cell must have 100 % confluence when exposed.” , it is difficult to check 100% confluence. I recommend you “almost 100%” or approximately 100%”. 3) Acceptance criteria : Regarding “Chemicals must be tested in the concentration ranges not inducing cytotoxicity” , I propose you from “cytotoxicity” to “exceeding 20% cytotoxicity”. 4) For interference assay, which case for antagonist assay is necessary or not unclear. Please add the acceptance criteria.
	PR3 Generally, yes. However, more guidance and background should be included on using the 4-parameter logistic regression for concentration response modelling. I can elaborate in our first

	teleconference. The protocol requires more specific guidance regarding concentration response modelling other than to simply use the 4-parameer Hill equation. Considering one are normalizing the data for concentration-response modelling, should the parameters be constrained at the top and the bottom of the curve to bottom = 0 and top = 100%. In addition, it would be valuable to share with the group De Lean et al. 1978 where the 4th parameter was added to the standard 3-parameter Hill equation. The 4th parameter was added to derive or constrain the bottom of the curve to a specific y-intercept. Based on how the data is analysed it seems the 3-paramter Hill equation would be a more sensible choice. De Lean et al. 1978. Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curved. American Journal of Physiology, 235, E97-102. There is an area to highlight in the protocol in Annex 2. The criterion/prediction model for a positive agonist is as follows in section 6.7 “The basal activity of the cells is < 4% of the maximal activity. A chemical is considered to have an agonistic effect when it increases luciferase activity above 10% of the basal activity of the plate (which is maximum 4%) for at least two consecutive concentrations or the highest concentration”. First, it is stated <4% in the first sentence but ≤4 in the last sentence. “<” in the first sentence should be changed to “≤”. Second, the for the criterion for a positive agonist used in the “Phase 2” validation study is different than the wording in Annex 2 and could be interpreted differently. The wording used in phase 2 was an increase above 10% of the maximum effect of the positive control, which is more descriptive. The wording in the protocol should be updated for clarity and to be consistent with the wording used in Phase 2. There are two criteria that need to be met for a valid positive response: 1. Basal activity of the cells must be < 4% of the maximal activity 2. There must be at least a 10-fold induction with the positive control DEX relative to the solvent control. If mean basal activity is 4 light units, and basal activity of the cells must be < 4% of the maximal activity with ≥10-fold induction with the positive control for a valid assay, the mean maximally inducing concentration of DEX would need to have an activity of 100 light units, which reflects a 25-fold induction relative to the solvent control, to meet the criteria.

	For perspective, it would be helpful to know the typical background luciferase activity for the solvent control and what labs typically measure for mean max fold induction for DEX relative to the solvent control. In the validation study (Phase 2), we have four positive test substances (Cortisol, Budesonide, Fluticasone, Aldosterone). Cortisol, Budesonide, and Fluticasone are relatively strong agonists with aldosterone as a relatively weak agonist that did not produce a full concentration-response. There are some concerns with the validity of the criterion for a positive agonist of ≥10% of maximum DEX induction with a minimum 10-fold criteria for only the highest concentration because of the inherent variability of the assay and this scenario was not rigorously evaluated with weak agonists in the validation phase.
	PR4 Given that different laboratories use different methods for counting cells which can yield varied results, it would be helpful to include a confluency standard for cell seeding to ensure labs can adjust the cell number based on their counting method. For example, "Cells should be seeded at 100,000 cells per well in 100 µL assay medium and incubated 16-18 hours at 37°C at which cells should be 80-90% confluent." The acceptance criteria (AC) for basal and maximal activity seem to rely upon circular reasoning. The AC for maximal activity is "the average luciferase activity of the positive control (DEX 10-7 M) must be at least 10 times that of the solvent control." The AC for basal activity is "must be ≤ 4% of maximum luminescence." If we are to assume no solvent effects, then the responses of the medium control wells (i.e., "basal") should mirror those of the solvent control wells. And if the basal activity "must be ≤ 4% of maximum luminescence", then it stands to reason it will always be ≥ 25 times that of the solvent control. This gives the appearance of being two independent AC but actually distils to a single AC. A preferable AC would be a z-factor (https://en.wikipedia.org/wiki/Z-factor) which uses the central tendency (mean, median) and the variation (stdev, mad) around both the basal and maximal responses to formulate a single, unitless value (theoretical maximum of 1) to determine whether sufficient dynamic range and invariability exists to delineate active and inactive response in an assay. Typically, an AC of z-factor ≥ 0.5 is used to qualify each assay plate (or in this case perhaps half-plate). There are several concerns with the method used to discriminate antagonist responses from those resulting from assay interference as diagrammed in Figure 1 and described in section 2.1 and revised in section 3.4. The revised method uses the root square (R2) of the Pearson correlation between two datasets dervied from different concentrations of DEX and cited the validation of AR-CALUX as precedent. First, this method even if successfully executed could only hope to discriminate interfering chemicals from reversible, competitive GR inhibitors. There are other "true antagonists" (p.7) of nuclear receptors besides reversible, competitive inhibitors including irreversible competitive and allosteric inhibitors which can be

	noncompetitive or uncompetitive. This problem is illustrated with the confused phase 2 results for 21-hydroxyprogesterone, spironolactone, 17β-estradiol, and fulvestrant. With the possible 21-hydroxyprogesterone, the other three compounds have been used as control chemicals for nearly identical firefly luciferase TA assays for AR or ER agonist and antagonist test modes, meaning it is inconceivable these compounds somehow indiscriminately interfere with a firefly luciferase TA assay that measures GR activity, but not AR or ER. Moreover, firefly luciferase is remarkably resistant to interference from drug-like chemicals with a rate of 0.1-0.9% (Auld et al., 2009). Using the higher figure (0.9%) the probability of identifying 4+ firefly inhibitors from a set of 35 test items is 0.026%. As for 21-hydroxyprogesterone, it is highly similar to progesterone, which has considerable binding affinity for GR. It is much more probable that these compounds are acting as GR antagonists and the R2 method is insufficient to properly classify these chemicals. Second, this method has a very narrow window to identify putative interferants. The antagonist test mode utilizes 80% of the assay's maximal dynamic range, leaving only the remaining 20% as room for shifting due to assay interference. This narrow window, coupled with the natural variation in replicate experiments may explain why likely inhibitors such as 21-hydroxyprogesterone, spironolactone, 17β-estradiol, and fulvestrant have R2 > 0.9 and are thus misclassified. Lastly, the major concern for interference in a cell-based assay such as HMLN-GR is cytotoxicity, i.e., signal decrease due to a loss of viability. The protocol and study as executed does an admirable job of addressing cytotoxicity during the range finding phase ensuring chemicals are not tested for GR activity (or loss) at concentrations that imperil cell viability. That said, it is not clear why the R2 method is being superimposed to further interpret loss of signal. It is unclear as to why drospirenone was selected as a chemical "expected… to decrease the receptor activity" since this compound has little to no affinity for GR (Fuhrmann et al., 1996).
	PR5 The protocol provided in the report is comprehensive. It includes detailed explanations of the materials used, applied procedures, and approaches of data interpretation and it's criteria. Standard operating procedures (SOPs) given in the report are well-documented, ensuring replicability of the procedures. However, to enhance the assays' robustness, further elaboration could be added regarding the handling of cytotoxicity effects and the criteria for weakly active compounds. For example, weakly active chemicals such as Cortivazol, a weak agonist with borderline EC50 values, and Mifepristone (RU-486), a weak antagonist with high variability in IC50 across laboratories, highlight the need for additional discussion in handling such cases. The assay protocol describes using co-exposure with a reference agonist at a full response concentration, to address the interference phenomenon exhibited by few chemicals that inhibit luminescence in antagonist assays. This approach explains the potential signal interference in such cases. However, the assay's reliability and relevance could be benefited by using similar controls to agonist assays and providing detailed procedures for identifying interference, such as autofluorescence or quenching. This aligns with OECD GD34's emphasis on addressing confounders.

Q4. Do you consider that the intra-, and inter-laboratory reproducibility of the GRTA test method results is adequately demonstrated? Please consider the degree to which biological variability affects the test method reproducibility.	**PR1** Yes, looks very solid.
	PR2 Yes. I think the reliability of this assay is high. However, I hope to describe the following reasons. 1) P79, INSERN and Tame-Water did observe a slight increase of the signal at the highest concentrations tested at butyl paraben. 2) P85, Tame-Water did observe a slight increase of the signal at the highest concentrations tested at B(a)P. 3) P88a and 89, Do you think a large variation of three runs tested at Budesonide and Fluticason propionate in Tame-Water. This information is described at Table 24. 4) P90, Toxen did observe a increase of the signal at the highest concentrations tested at Fluticason propionate. 5) P106, One labs obtained a correlation coefficient between antagonist and interference runs of 0.89. Which lab is not cleared? Interis did observe a high increase of the signal at the highest concentrations tested at Sodium azide. 6) P116, Toxmen did observe a high increase of the signal at the highest concentrations tested at Buthyl paraben. 7) P120, The labs did not provide any explanation for these results. Please collect the explanation form labs.
	PR3 Generally, yes for the agonist mode. However, there were a fair number of inconsistencies for inter-laboratory reproducibility of the GRTA test method results in the antagonist/interference modes where there were different calls in some cases across the labs. This is likely related to the methodology used to assess interference, which is generally recognized by experts in the field to be a problematic approach, particularly for weak agonists, based on work done with the ARTA assay. There are concerns with prescribed approach to assess the shift of curves between the antagonism and interference modes for weak antagonists. There is no scientific literature cited for the R2 criterion except in the ARTA Calux guideline. The R2 has been prone to technical variability and has limitations which is evident when testing weak agonists. Bio Detection Systems (BDS) is now providing an alternative method for assessing the shift on top of the R2 e.g., ratio IC20s and IC50s or AUCs when IC20 can't be estimated. BDS has contacted OECD to discuss this alternative method. It would be a meaningful exercise to evaluate this approach as a supplemental procedure if it improves the calls. At the end of section 6.7 the protocol asks to calculate an IC30 for antagonist chemicals. Should this be changed to an IC20 considering the criterion for a positive? This looks to be a hold over. Additionally, there is a paragraph on guidance to determine if a parametric or non-parametric test should be in case the Hill function

	cannot be fitted. It is not clear if this analysis is for the agonist or antagonist assay or both? A nonparametric test is recommended (e.g., Kruskal-Wallis) but a parametric test is not mentioned, nor the alpha level, etc. What is the purpose of the statistical test in terms of making agonist and antagonist calls? The cell line can be obtained by an MTA with ISERM. However, the cost to obtain the cell line from INSERM should not be prohibitive for research laboratories.
	PR4 The reproducibility both within and across laboratories was amply demonstrated, both with respect to classification (positive/negative) and potency (EC50).
	PR4 The intra- and inter-laboratory reproducibility of the GRTA test method is adequately demonstrated in the validation report, with high consistency across multiple runs and laboratories. The values of Coefficients of variation (CV) are mostly below the acceptable limits (<15%), ensuring the OECD GD34 standards for reproducibility. These results suggest that the provided protocol is robust and could be utilized across laboratories. However, for weak chemicals biological variability is also observed, and it is challenging issue in this report that warrants further analysis. For example, Hydrocortisone, a partial agonist, is exhibiting inconsistent dose-response relationships at lower concentrations, and Fluticasone, a weak antagonist, demonstrates variability in IC50 values between different laboratories due to sensitivity near the detection threshold. These chemicals highlight the need for enhanced methodologies to reduce variability and ensure consistent results. Biological variability in weakly active chemicals could be affected by assay sensitivity, dynamic range limitations, and differences in laboratory practices, such as reagent quality or operator expertise. This could potentially increase the likelihood of false positives or negatives, especially for chemicals exhibiting activity close to the detection threshold. To address this, the assay could benefit from additional validation steps, including expanded testing of weak chemicals, refinement of concentration ranges, and the implementation of advanced statistical methods to identify and mitigate variability sources. Further, incorporating more detailed protocols for handling borderline activity, such as normalization techniques or additional controls, would improve assay robustness and reliability across a broader applicability domain. While reproducibility for most chemicals is robust, addressing the variability observed with weakly active compounds is essential to fully meet regulatory expectations and ensure confidence in the assay's predictive capacity across diverse chemical classes.

Fitness for Purpose

For the following two CQs you may wish to consider the particular aspect of GD34 (Para 33) that is applicable to GRTA test method:

"For cases in which a new test method addresses an endpoint that has not been previously considered or for which there are no or inadequate established reference data, a rigorous assessment of the predictive capacity may not be possible, however, the relevance should be assessed using all available information. Such tests that have shown to be reliable, but whose relevance has only been partially assessed by comparison with related reference data (e.g., TG 414 (14) and TG 424 (15)), should preferably have their validation status periodically reviewed, with full consideration of all relevant available data."

Q5. Is the assessment of the test method's performance based on the testing of relevant reference chemicals that are representative of the types of substances for which the test method is intended to be used?	**PR1** If the test is used to identify steroidal compounds or compounds directly binding to the ligand binding domain, then clearly yes. If the test is used to identify substances that cause endocrine disrupting effects by altering GR activity, then no because several limitations and levels of disturbances of GR activity cannot be detected.
	PR2 Based on the information of suggested effect in Table 7, I classified 17 agonist, 19 antagonist and 4 non-chemicals. Please add the information of DEHP (not clear effect) in Table 7. I think complete negative chemicals are too little. I prefer 15 - 20 % complete negative chemicals in 40 chemicals.
	PR3 There were only a limited number of known agonists and antagonists tested. Typically, a larger number and range of potencies of agonists and antagonists are tested for an OECD validation. However, there were enough negatives tested.
	PR4 The reference chemical was adequate to assess the method performance, especially given the limited number of well-studied GR agonists and antagonist. The VMG did an excellent job of mining available resources (ToxCast assays, QSAR models, and open literature) to identify a relatively balanced set of putative actives and inactives for both agonist and antagonist test modes.

	PR4 The validation study includes a well-selected and diverse set of reference chemicals, ranging both agonists and antagonists, that are representative of potential endocrine disruptors. This selection ensures that the GRTA test method addresses a broad applicability domain, including chemicals with varying activity profiles, structural diversity, and relevance to regulatory needs. For example, the inclusion of potent GR agonists such as Dexamethasone and antagonists like Mifepristone (RU-486) demonstrates the assay's ability to capture strong GR modulation. Additionally, substances with minimal or borderline activity, such as Fluticasone or Prednisolone, are included to analyze the assay's sensitivity and dynamic range. However, there is room for improvement by incorporating more weakly active or borderline chemicals to refine the predictive capacity and evaluate variability at detection limits. While the reference chemical selection aligns well with the test's intended purpose, the study could enhance its performance assessment by including additional chemicals that test the boundaries of the assay's applicability domain. For example, expanding the dataset to include structurally novel or underrepresented classes of chemicals would further strengthen the assay's robustness and relevance. Additionally, periodic review and updates to the reference chemical set, as recommended by OECD GD34, would ensure the method remains aligned with emerging regulatory requirements and chemical landscapes. The study adequately assesses the relevance of the GRTA method by comparing assay results with related reference data and established benchmarks. However, given the evolving landscape of endocrine disruptor testing, periodic re-evaluation of the assay's validation status with all available data is essential. Incorporating mechanistic insights and linking assay outcomes to adverse outcome pathways (AOPs) would also enhance the biological relevance and regulatory applicability of the GRTA test method.
Q6. Has the performance of the test method been evaluated in relation to relevant information from the species of concern, and all existing relevant toxicity testing data for relevant reference chemicals? The panel is encouraged to advise on any other relevant data and/or analysis.	**PR1** The test is based on human GR. Results of reference chemicals reported in literature were mainly taken from TAs performed with human GR, bu they are highly heterogeneous due to the use of different cell systems, different media, concentrations, and assay protocols used. For avoiding this, a standardized protocol and validated GRTA system is clearly needed. If differences to other cell systems can be convincingly shown, then the underlying mechanism should be elucidated but this is not the goal of this assay but rather a need for safety assessment.
	PR2 I think relevant of this assay is necessary for regulatory use. Based on the information of suggested effect in Table 7, please compared to both results. I think the section 6 of results analyses on compound is not results in the validation study and they are discussion.
	PR3 There is not a discussion of the utility of this assay for ecological species (e.g., aquatic vertebrates like Xenopus and rainbow trout). This analysis can be performed with SeqAPASS to assess homology between functional domains and a review of the literature comparing

	relative ligand binding affinities between the receptor in the GRTA and aquatic vertebrates.
	PR4 Given the relative scarcity of GR reference chemicals for which strong toxicity data exists, the test method performance was evaluated accordingly.
	PR5 The performance evaluation of the GRTA test method leverages available toxicological and biological data, providing a solid foundation for assessing its relevance. The assay's focus on the human GR provides a biologically and species-relevant approach, which aligns with regulatory requirements for endocrine disruptor testing. Data from reference chemicals, including both agonists and antagonists, are used to validate the assay's predictive performance and reliability. For instance, the use of compounds like Dexamethasone (a strong agonist) and Fluticasone (a weak antagonist) provides insight into the assay's capacity to detect GR modulation across a range of activities. However, the study could benefit from integrating additional human-relevant data sources to strengthen its regulatory applicability. Including in silico modeling, such as quantitative structure-activity relationship (QSAR) predictions, could help contextualize assay results and expand the applicability domain. For example, QSAR models could predict GR activity for structurally diverse or poorly characterized chemicals, supplementing experimental data. Similarly, coupling GRTA results with adverse outcome pathway (AOP) frameworks could link molecular interactions with higher-order biological effects, such as metabolic disorders or cardiovascular toxicity, improving the assay's relevance to human health. The evaluation also highlights the assay's potential to complement existing toxicity testing frameworks, but further comparison with in vivo and epidemiological data could provide a more comprehensive validation. This includes exploring correlations between GRTA outcomes and observed endocrine-related health effects in humans. Additionally, assessing the assay's performance for underrepresented chemical classes or novel structures would address gaps in current data and strengthen the assay's predictive power. By integrating human-relevant tools like in silico models and AOPs, and periodically reviewing the assay's validation status with emerging data, the GRTA method can significantly enhance its application for regulatory decision-making.
Data integrity and transparency	
Q7. Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP. Assess the goodness of the experimental/validation	**PR1** In my opinion this was very well done. All of the compounds reported in the literature led to the same result by the different labs analysing them. The conflicting data compared to literature finding activity of the chosen compounds may be due to inappropriate assays, use of different cell systems (for example when using reporter genes regulated by endogenously expressed receptors. The fact that all labs obtained the same results (with some differences for antagonist vs interfering where more data points may be needed) demonstrates the robustness of the assay.

practices (for example: the coding of test chemicals for blind testing and analysis of test results, the acceptance criteria, the qualitative and quantitative comparison in and between labs, or any others you considered relevant) under which the data have been generated for the validation study of the GRTA test method	**PR2** This validation study is out of GLP study and it is not clear in accordance with the principles of GLP. In the VMG, who is checked all the data and worksheet? Reviewer at page 1 is clarified? Please make a clear the organisation and role of the VMG and attach the study plan, because I cannot recognize the correlation among authors, reviewers and directors at page 1, the data manager and biostatistician at page 9 and overview of the VMG at page 10. In addition, please share the certification and FRAND declaration of the cells.
	PR3 Data were generated non-GLP. Therefore, it would be good to have more information on quality control measures that were used for data generation e.g., auditing the raw data and the methodology for dose response modelling to make sure there was alignment in the methodology among the participating labs.
	PR4 All elements of this study were conducted in accordance with GLP principles.
	PR5 The validation study demonstrates adherence to Good Laboratory Practices (GLP), ensuring robust data quality and integrity. Test chemicals were coded for blind testing, and inter-laboratory reproducibility assessments were performed, aligning with OECD GD34 requirements. The report highlights general consistency across laboratories, particularly for strong and medium-active chemicals, with coefficients of variation (CV) within acceptable limits (<15%). However, variability was noted for weakly active chemicals, where responses approached the detection threshold, though specific examples were not detailed. While the report emphasizes the use of standardized protocols and coded samples to mitigate biases, it does not explicitly detail how inter-laboratory discrepancies, if any, were identified or resolved. Including this information would strengthen the assay's validation. Furthermore, a dedicated section summarizing quality control measures, such as calibration standards, reagent lot tracking, and personnel training, would enhance the report's comprehensiveness. Specifying how discrepancies in weak chemical performance, if observed, were addressed would further illustrate efforts to maintain consistency across laboratories. By addressing these points and elaborating on quality control practices, the GRTA test method could enhance confidence in its data integrity and reliability, bolstering its credibility for regulatory applications.
Q8. Do you consider that all the data supporting the assessment of the validity of the test method are easily available for expert	**PR1** Fully agree. Detailed and very clear assay description. Should be straightforward to apply and yield reproducible results.
	PR2 In case of antagonist assay with interference assay, please add the tables as show in Table 5. Nothing of Table of cortisol and DEHP

review? These include: a detailed and readily available test method protocol to the public and independent laboratories; organised and easily accessible data to permit independent review(s); benchmarks by which an independent laboratory can itself assess its proper adherence to the protocol?	at phase I. I found the terms of dose-response curve. I think concentration-response curve is correct in phase I. For phase II, 1) I think missing data of Triclosan at Tame-Water in Table 16. 2) I guess Fig 58 is made a mistake and not including "INTER". 3) The title is missing from 21-hydroxyprogesteron to RU486 in Table 19. 4) The Interference data at Interis is missing in Table 19. 5) For Keteconazol, please add the Table, because I think no needs of INTER of this chemical at almost labs. 6) The Interference data at all the labs is missing in Table 20. 7) Fig.68 A)-D) are broken. 8) The subtitle of Fig.78, E) Lithium chloride ANTAG-INTER is correct. 9)For Sodium azide, please add the Table, because I think no needs of INTER of this chemical at Ineris. For supplemental phase, 1) Please share the data for BisphenolA form Toxem for in Table 29. 2) Please share the data for Propylparaben form Toxem in Table 32. 3) Please share the data for Methoxychlor form Toxem in Table 33. 4) Please share the data for Keraconazole form Toxem in Table 34. 5) Please share the data for Sodium arsenite form Toxem in Table 35. Thank you.
	PR3 Yes and No. The no refers to the way the data were presented in the report (normalized data only) and the difficulty understanding how the results were presented in the Excel file. Perhaps we have a primer on the Excel sheet on how the data were entered on our first teleconference.
	PR4 The data needed to review the validity of the test method are publicly available as is the benchmarks/AC needed to determine whether a naïve laboratory has successfully conducted the assay.
	PR5 The validation study provides a solid foundation for assessing the GRTA test method's validity, with most supporting data being well-organized and accessible for expert review. The test method protocol and the provided raw data is detailed and includes information on assay components, procedures, and acceptance criteria, ensuring that independent laboratories can replicate the method. The report provides transparency by outlining the coding of test chemicals for blind testing and specifying the benchmarks used for evaluating assay performance, such as Z'-factors and coefficients of variation. These benchmarks are critical for assessing an independent laboratory's adherence to the protocol and ensuring reproducibility across different enviroments. However, while the data is generally accessible, certain aspects could benefit from additional clarity. For example, including a public repository for data and protocols, such as an online platform or supplementary document, could further enhance accessibility for broader regulatory and scientific communities. In alignment with OECD GD34 principles, ensuring that data supporting the validation of test methods is fully transparent and organized not only promotes confidence in the method but also enables independent laboratories to effectively assess and adhere to the protocol. Enhancing data accessibility through centralized

	documentation and repository systems would strengthen the GRTA method's credibility and usability in regulatory contexts.
Overall Review	
Taking into account your overall review, do you consider that the validation study and the supporting references provide appropriate evidence to inform testing strategies, guidance and/or guidelines for application of the GRTA test assay to assessment of ED toxicity of chemicals. Reviewers are encouraged to provide any specific recommendations that they may have for future near- and long-term application and potential development of the GRAT assay.	**PR1** This assay is nicely validated, with a well-described protocol and should yield highly reproducible data between labs. The assay should be used to identify substances binding to the ligand binding domain, thereby activating or inhibiting GR. This is one of several mechanisms by which substances can alter GR activity and cause ED toxicity. In order to further assess ED toxicity of substances altering GR activity, several other levels will need to be taken into account. This may need GRTA in cell systems expressing endogenous levels of GR and where signalling pathways are active that are relevant for specific cells/tissues. The same applies for ARTA, ERTA and other nuclear receptor TAs and may explain the low predictivity of TAs for ED toxicity seen in living organisms.
	PR2 I think this validation report incomplete. Based on the reviewer' s comment including myself, the VMG team will revise the report and re-submit the amended report soon.
	PR3 As discussed above, the validation work completed to date informs testing strategies and the development of guidance, but this reviewer questions the relevance and utility of this method for broad application as an adopted test guideline to assess an endocrine modality. Relative to other TA validations a low number of true positives for the agonist and antagonist assays were tested, which reflects the low likelihood of identifying agonists or antagonists in the environmental outside of pharmaceuticals. The choice of chemicals for the validation was a difficult task since only a few chemicals had convincing data that would allow for an assessment of predictability. Therefore, another supplementary phase is recommended to increase the number of weak agonists and antagonists tested to further interrogate the performance of the assay against the established criteria for agonism and antagonism. Interpreting the results of the report for this review would have benefitted by also expressing the data in terms of increases or decreases in fold-induction rather than as normalized to a percent of control for the y-axis and concentrations in uM rather than in mg/ml on the x-axis. This would have put the results into a more transparent format for a validation review. The later would have made for an easier comparison of relative potencies across the substances tested. The criteria for a positive in the antagonism protocol is two consecutive concentrations with >20% inhibition for a level of 65% to 80% of maximal induction. However, the assay allows for up to 20% cytotoxicity, which could create a scenario where most of the reduction in luciferase activity is from cytotoxicity. Therefore, recommend consider changing the criterion back to the original 30% or correcting for cytotoxicity in antagonism assessment. Also, for the antagonism assay would recommend that the maximum test

	concentration should not exceed 100 uM to reduce the likelihood of observing interference.
	PR4 The HMLN-GR assay is well-positioned to play a role in the future ED assessment of chemicals. There are some technical/protocol concerns to be addressed that have been detailed above, namely assurance that the assay does not respond to androgens and a recommendation to jettison the statistical analysis regarding assay interference that should be addressed before a finalized version is approved.
	PR5 The validation study provides strong evidence supporting the relevance and reliability of the GRTA assay for detecting endocrine-disrupting chemicals (EDCs). The study aligns with OECD GD34 principles, demonstrating robust intra- and inter-laboratory reproducibility, comprehensive protocol documentation, and appropriate use of reference chemicals. The inclusion of both agonists and antagonists, along with blind testing and standardized acceptance criteria, highlights the assay's strength in addressing regulatory needs. Despite these strengths, there are opportunities for further enhancement. For near-term improvements, expanding the chemical set to include additional weakly active or borderline chemicals would provide a more rigorous assessment of the assay's dynamic range and predictive capacity. Refining protocols to handle cytotoxicity effects and incorporating advanced statistical approaches to assess variability in weak chemical responses would enhance reliability. For long-term developments, the GRTA assay would benefit from integration with Adverse Outcome Pathway (AOP) frameworks, which could link molecular-level findings with higher-order biological effects, such as metabolic or cardiovascular outcomes. Additionally, incorporating computational models, such as quantitative structure-activity relationship (QSAR) tools or machine learning algorithms, would complement the assay results and expand its applicability domain. These integrations would make the assay more versatile and align it with the evolving landscape of New Approach Methodologies (NAMs). Lastly, enhancing data accessibility through centralized repositories and increasing transparency in inter-laboratory variability resolution would bolster confidence in the method. By addressing these points, the GRTA assay can serve as a critical tool in endocrine disruptor testing strategies and contribute significantly to regulatory decision-making frameworks.

Annex 4 – Representatives from the Validation Study present at the TC on 4 December 2024

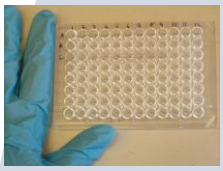
Name	Role in the Validation Study	Country
Philippe HUBERT	Director, PEPPER	France
Torben ÖSTERLUND	Chief Scientific Officer, PEPPER	France
François Sermier	Data Manager, Independent Expert	France
Sebastian Hoffman	Biostatistician, Independent Expert	Germany
Ella Atlas	VMG member, Health Canada Expert on ED Assays, regulatory application, validation	Canada

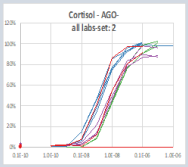
Annex 5 – Initial analysis of ToxCast database for additional potentially relevant substances for the GRTA test method

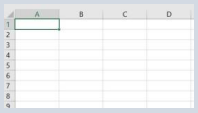
See separate Annex

Annex 6– Presentations from teleconference with VMG and PEPPER

PEPPER – GR-TA Excel file and data management







1

Data flow upstream Excel file

Three actors

Laboratories/ Pepper/Data Manager

+ downstream : biostatistician (advises, uses data for analysis…)

Basic tools

Excel template file for data collection
one per run and per substance – or pair of substances - and per laboratory)
in standardized files

Excel tools (compilation file) to
compile templates (VBA macro)
store data
analyze data (mostly PivotTables PT)
and display results (charts using palette selectors to navigate)

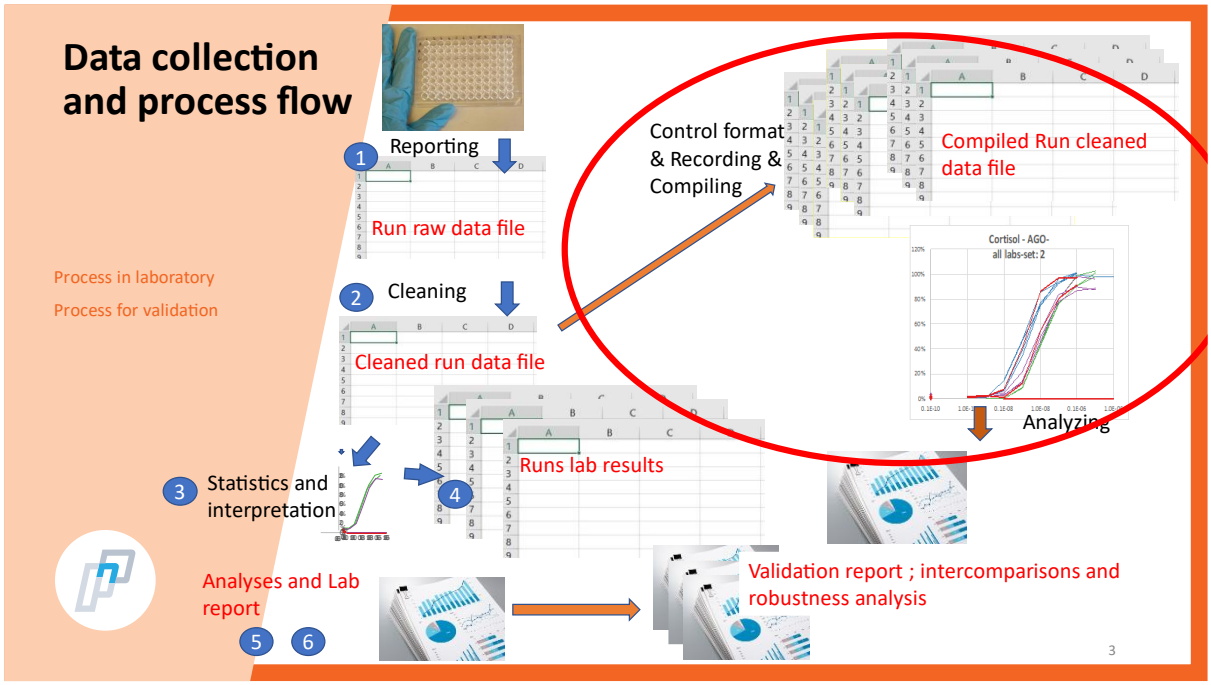
Basic responsibilities

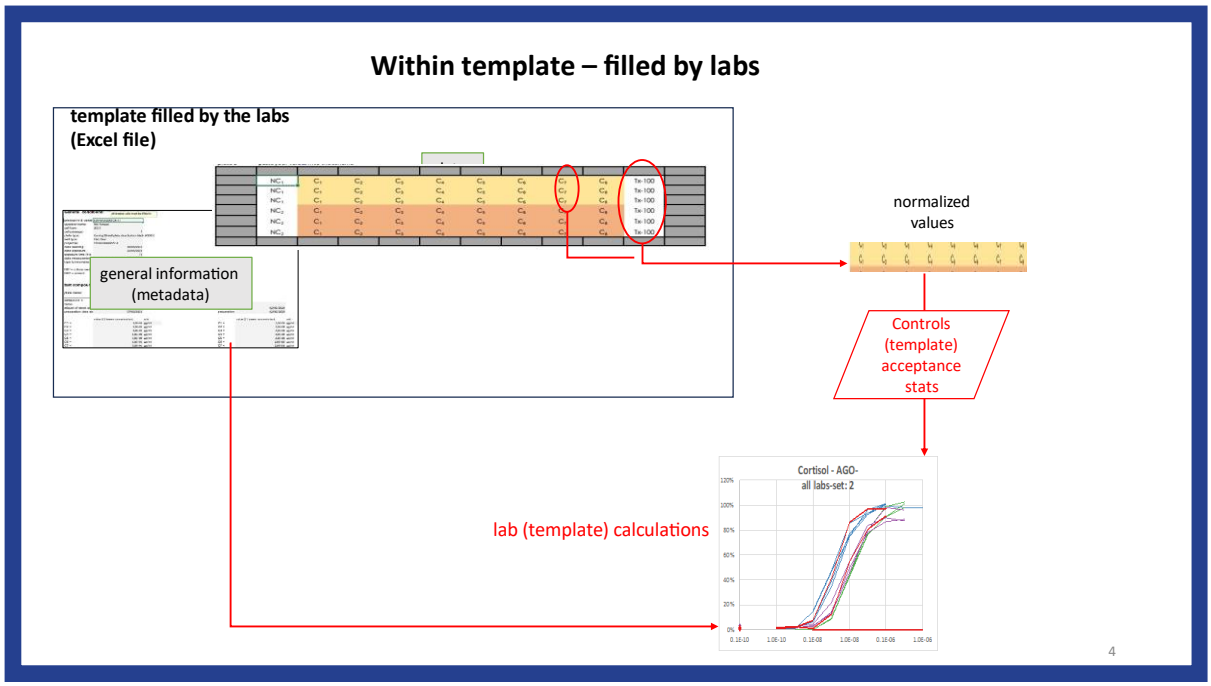
Laboratories : generate data, fill in the template file
responsible for reporting fidelity

Pepper : checks with a level 2 quality control
gives the green light

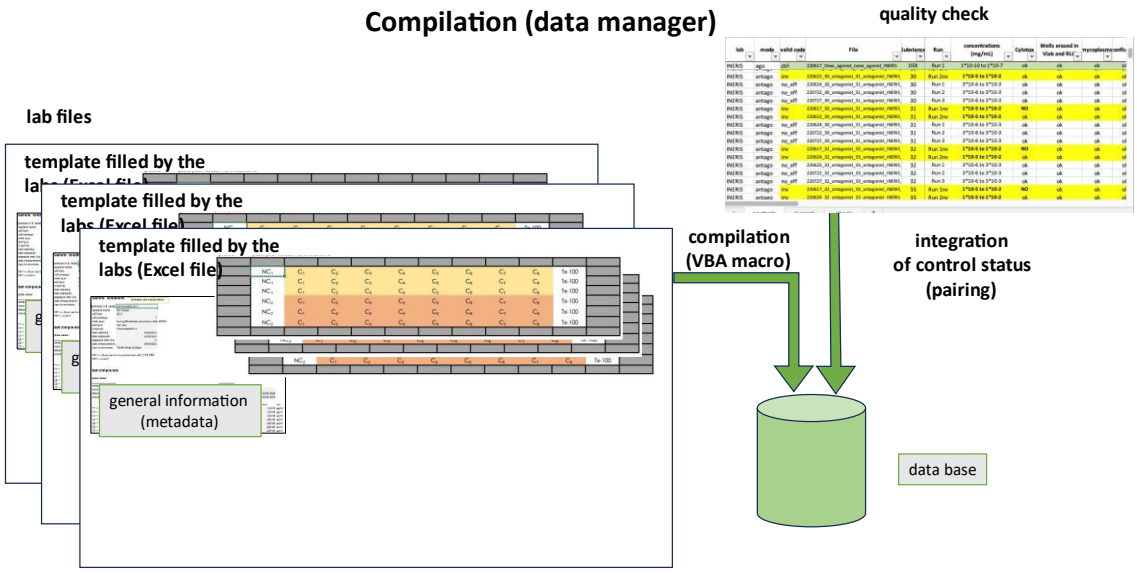
Data manager : compiles, stores and treats data

2



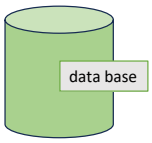


Compilation (data manager)



data structure – results navigation

data tables
always present
as XL sheets



note : all data sheets
will be locked
in read-only access

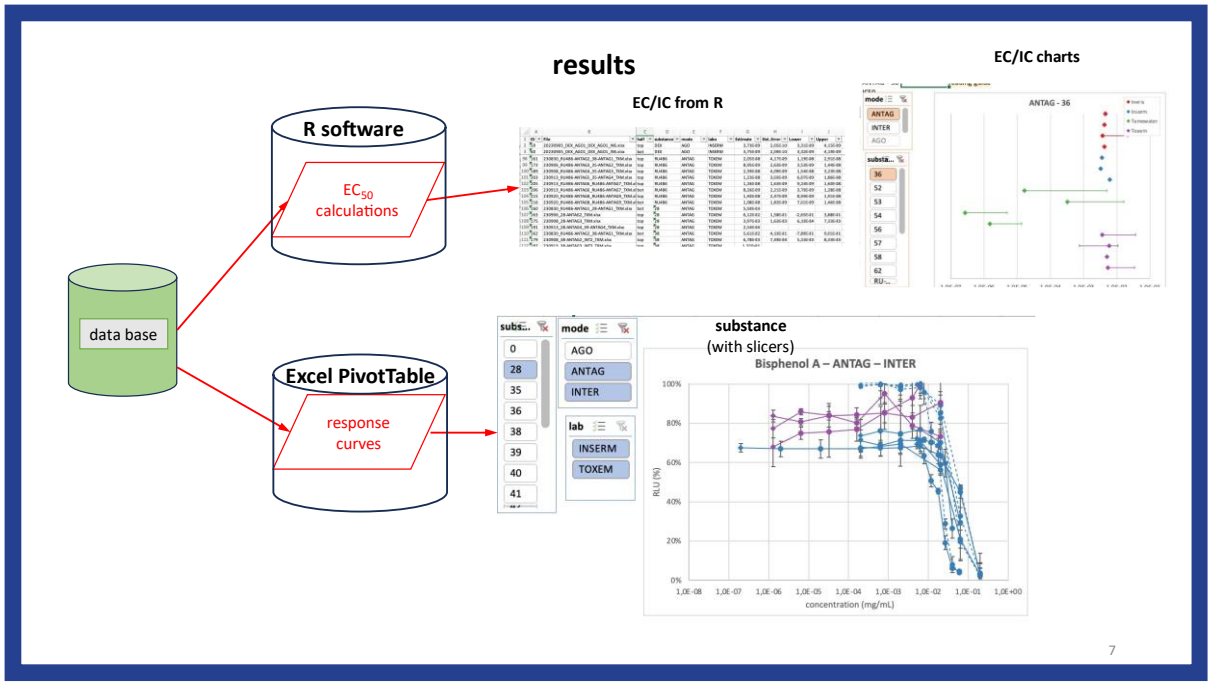
In "long form"
1 value (measure) per row
all other info as attributes
(ex. plate row, plate col.)

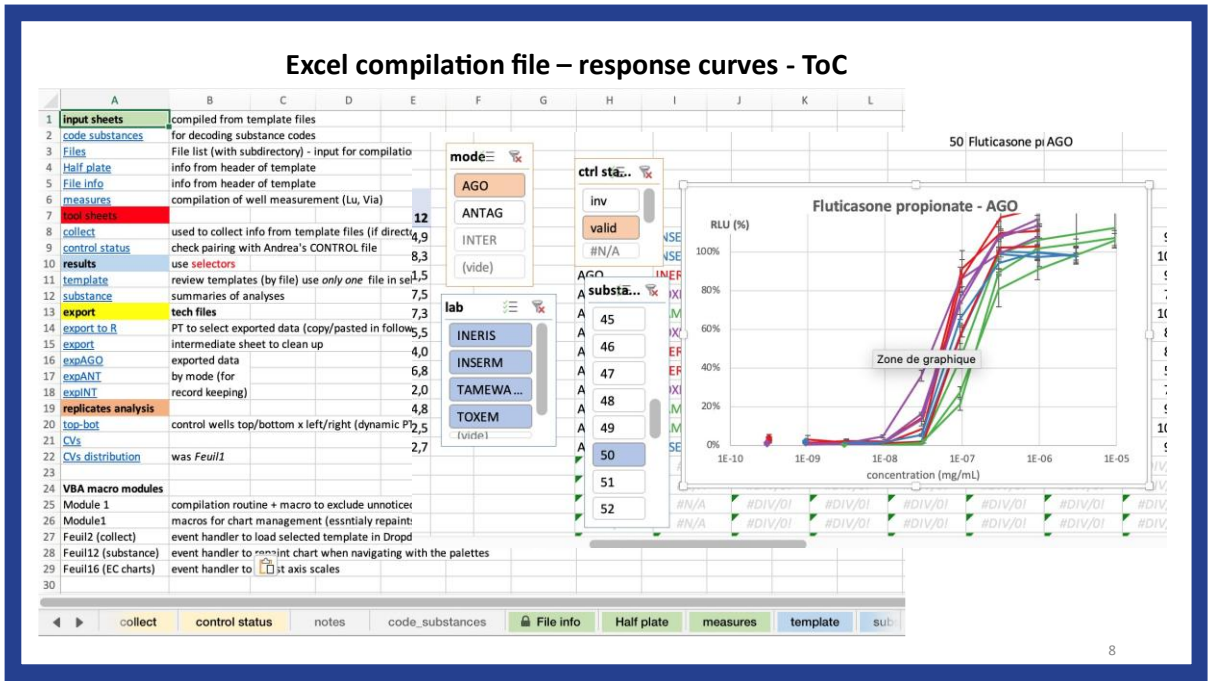
almost all calculations are
done using XL **PivotTables** (PT)

select among the results and displays
using XL **Slicers** (palettes)

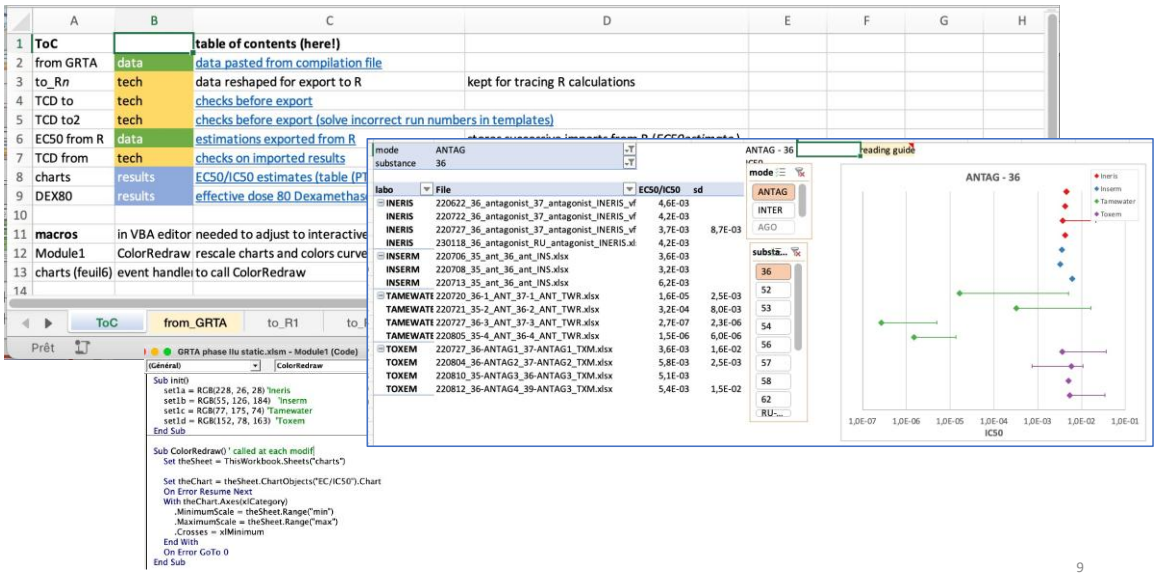
macros are used *only*
- in compilation phase
- for cosmetics in chart display
  (color by lab, yaxis scale setting)

6



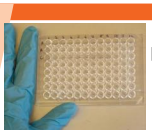


Excel file – EC50/IC50- ToC



9

Roles and responsablity in quality assurance



Lab Staff

reporting quality and format

Run raw data file

cleaning quality check, tracing changes
Sending clean and formated files

Cleaned run data file

Control Statistics and interpretation

Cleaned run lab results

Lab report and analyses

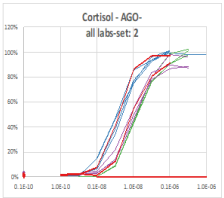
Compiled Run raw data file

Check reject unformatted files

Check quality criteria 2 nd level

Compiled Run cleaned data file

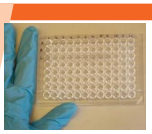
Check quality of compilation
Check analysis process and tools



The tools for data management

Process in laboratory

Process for validation



Reporting

Formated template
Tracing experiment conditions
Rules for cleaning
Acceptance criteria

Run raw data file

Cleaning

Automated control tools

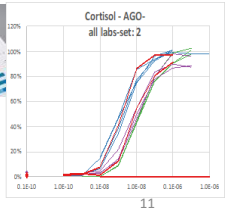
Compiled Run cleaned data file

Cleaned run data file

change tracing with Excel comments

Statistics and interpretation

Cleaned run lab results



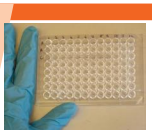
Lab report and analyses

Common routines for analysis

11

Traceability

What is stored, what is not



Reporting

Compiled run raw data file
Excel databases
@Pepper @DM

File storage for
Controlled raw data files
In lab, in Pepper, in DM

Run raw data file

Cleaning

File storage
Controlled cleaned data files
+ change tracing
In lab , in Pepper

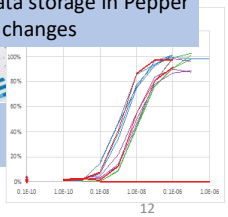
Compiled Run cleaned data file

Cleaned run data file

Routines storage
And/or
Processed data storage in Pepper
Versioning if changes

Statistics and interpretation

Report storage



Lab report and analyses

12