Document in support of the Peer Review of the validation of the update of the TG 407 (repeat dose 28-day oral toxicity study)

This document was prepared by the Secretariat to present the current status of the validation of the updated TG 407, within the context of validation criteria outlined in Guidance Document 34. In particular, each of the 8 criteria required for successful validation of a new test method are listed along with details of how these criteria have been addressed during the OECD exercise and how the available information supports the validity of this method. This document is intended to guide the updated TG 407 Peer Review Panel in their review of supporting material and provide a framework for their discussions.

References to the different documents of the peer review package are identified as presented below and followed by the paragraph number:

- Validation report: VR
- Draft test Guideline: DTG

GD34 - Criteria 1/8

a)	**The rationale for the test method should be available.** This should include a clear statement of the scientific basis, regulatory purpose and need for the test.

Scientific basis

Ambient levels of natural and industrial chemicals may interact with the endocrine system and as a consequence possibly elicit reproductive, developmental, and other adverse effects in humans and wildlife. The supporting evidence for these concerns have been expressed in number of expert reviews. This leads then to the need to ensure that endpoints sensitive to these mechanisms are evaluated for the usefulness and that appropriate endpoints are incorporated into toxicity tests such as the TG 407 (VR, 3).

Regulatory purpose and need for the test

The regulatory need for this assay stems from the report that existing Test Guidelines were generally insufficient to identify certain endocrine mechanisms (oestrogen, androgen, and thyroid) and might not be adequate to fully characterize the hazards of these mechanisms (VR, 4).

The OECD initiative to develop and validate *in vitro* and *in vivo* assays for the detection of chemicals that may interfere with the endocrine response was taken following the recommendations of a number of national, regional and international workshops and following a detailed OECD review of the status of existing test and research methods.

The proposed updates to the TG 407 are intended to potentially address both (anti)estrogens and (anti)androgens as well as thyroid toxicants.

In the assessment and evaluation of the toxic characteristics of a chemical, the determination of oral toxicity using repeated doses is central to hazard identification and characterization. The 28-day repeat dose study (TG 407) provides information on the possible health hazards likely to arise from repeated exposure over a relatively limited period of time. The current version of TG 407 provides information on a wide range of toxic effects and indicates target organs, but is not designed to identify particular mechanisms or modes of action. The study also serves as a dose range finder for follow-up studies and provides an estimate of a no-observed-adverse-effect level (NOAEL) of exposure for certain toxic effects, and potentially a maximum tolerated dose. This information can be used in selecting dose levels for chronic studies, if needed.

As a result of its broad spectrum of endpoints and observations, TG 407 is frequently part of regulatory data requirements. This pivotal study for the hazard assessment of chemicals is commonly performed for most high production volume chemicals, thus update of this Test Guideline suggested greater efficiency and lower animal use than constructing a battery of individual assays to address different mechanisms (VR, 2 to 5).

GD34 - Criteria 2/8

b) **The relationship between the test method's endpoint(s) and the (biological) phenomenon of interest should be described.**

This should include a reference to scientific relevance of the effect(s) measured by the test method in terms of their mechanistic (biological) or empirical (correlative) relationship to the specific type of effect/toxicity of interest. Although the relationship may be mechanistic or correlative, test methods with biological relevance to the effect/toxicity being evaluated are preferred.

In order to identify possible estrogens, antiestrogens, androgens, antiandrogens, and thyroid toxicants during the course of the 28-day repeat dose toxicity study, the principle for TG 407 updates is that the protocol incorporates the necessary array of endpoints sensitive to these mechanisms (VR, 6).

Thus the updates of TG 407 rely on the measurement of the following parameters (see table below for more details):

- Weight and histopathology of hormone-dependant tissues:
  - male and female reproductive tracts
  - thyroid and pituitary
- Thyroid hormones (others not feasible)
- Sperm and estrous parameters

Comparison of the current TG407 endpoints with the proposed update endpoints investigated during the phase-2 studies (VR, table 2)

Endpoint/ effects	Current TG 407 Endpoints	Proposed updates to TG 407
Organ/tissue weights	liver, kidney, adrenals, testes, epididymides, thymus, spleen, brain, heart	1. testes (each weighed separately) 2. seminal vesicles + coagulating glands 3. prostate (possible dissection and separate weights for ventral and dorsolateral prostate), ovaries 4. thyroid 5. uterus
Histopathology	brain, spinal cord, stomach, small and large intestines, liver; kidneys, adrenals, spleen; heart, thymus, trachea, lungs, gonads, accessory sex organs (i.e., uterus, prostate), thyroid, urinary bladder, lymph nodes, peripheral nerve, bone marrow, all gross lesions	1. pituitary 2. vagina 3. one epididymidis, seminal vesicles + coagulation glands 4. mammary gland,
Thyroid Hormones	none	1. circulating levels of T_3 and T_4 2. circulating levels of TSH
Spermatology	none	1. epididymal sperm number 2. sperm morphology
Estrous cycle	none	Daily vaginal smears to assess oestrous cycling via epithelial cytology for at least five days to ensure necropsy during diestrus

GD34 - Criteria 3/8

c) **A detailed protocol for the test method should be available.**

The protocol should be sufficiently detailed and should include, *e.g.*, a description of the materials needed, such as specific cell types or construct or animal species that could be used for the test (if applicable), a description of what is measured and how it is measured, a description of how data will be analysed, decision criteria for evaluation of data and what are the criteria for acceptable test performance.

A detailed protocol is provided, including guidance on the test materials needed: rat (age, weight, number of animals used), administration of doses, clinical observations, haematology and clinical biochemistry (emphasizing on conditions of sampling) and pathology.

General decision criteria for the evaluation of data are similar to existing OED Test guidelines, based on observation and statistical significance of measures.

In addition, combination of TG 407 data with that from other studies (e.g. OECD reproduction toxicity screen) may be necessary in special cases (depending on exposure, use pattern etc.). A detailed guidance for histopathological evaluation of subtle effects and quality control procedures may also be needed to obviate increase in animal number (VR, 24, 276, 277, 338).

GD34 - Criteria 4/8

d) **The intra-, and inter-laboratory reproducibility of the test method should be demonstrated.**

Data should be available revealing the level of reproducibility and variability within and among laboratories over time. The degree to which biological variability affects the test method reproducibility should be addressed.

In phase-2, for each chemical, duplicate independent studies have been performed in 2 laboratories, for a total of 20 studies (VR, 317).

Overall, the TG 407 data and findings generally were in good agreement between the two laboratories conducting studies on the same substance. The findings in these studies support the reproducibility and reliability of the traditional use of the TG 407 in flagging possible systemic and target organ toxicity. For each of the ten test substances, changes in body weights, haematology and clinical chemistry parameters, and current absolute and relative tissue weight endpoints, and the histopathology endpoints have been compiled into comparative tables in each of the test substance chapters of the validation report. In addition, where other toxicological data were available on the test substance or a close mechanistic relative, these same endpoints have been compared with TG 407 data at the end of each test substance chapter.

These comparisons show that, within the TG 407 studies and with other toxicological data, these endpoints are fundamentally reproducible and that the TG 407 is basically a reliable predictor for other studies. The changes in body weight were a reproducible effect between the TG 407 studies, noting that, based on the power calculations, the percent change in body weight often had to be > 10% to achieve significance even with the low CV for this measurement. Similar body weight changes were observed in longer and more detailed toxicological assays. For haematological and clinical chemistry parameters, four basic classifications were entertained: 1) the parameters was statistically significant in both studies, 2) the parameter was statistically significant in one study and the absolute trend in the other study was in agreement, 3) the parameters was statistically significant in one study and the absolute trend in the other study was unchanged or moved in the opposite direction, and 4) the direction of statistical significance was in the opposite direction between the two studies. Within the TG 407 studies, the bulk of the changes fell into the first two categories. These data were often not published for other toxicological assays, so no comparison here was consistently made and no conclusion is offered. For tissue weights, such as liver, kidneys, and adrenals, the TG 407 studies were fundamentally reproducible and matched findings in longer and more detailed toxicological assays. Histopathological findings were also similar both among the TG 407 studies for a test substance and with longer and more detailed histopathological studies. Due to the longer time of administration, LOEL doses in these non-TG 407 studies were typically lower, as would be expected (VR, 336, 337).

The CV values of the proposed update tissue weights were slightly higher than a number of current tissues. For example, current tissues, that are larger and easier to dissect, such as liver and testes, had lower mean CVs ranging from about 8 to 11. As tissues became smaller and more difficult to dissect, CVs increased. Mean CVs were 12-14 for paired adrenals, 18-19 for pituitary, and 20-21 for the thyroid. Fluid-filled male accessory tissues that are difficult to dissect, such as the prostate lobes, had higher mean CVs approaching 24. The CVs for the uterus and for the male accessory tissues were consistent with those observed in the uterotrophic and Hershberger validation programs. In addition, as also with the uterotrophic and Hershberger studies, CV values varied and appeared to be related with the individual

performing laboratories. This suggests that laboratory technique is a possible variable and could be connected to the ability to detect weakly acting substances (VR 338).

GD34 - Criteria 5/8

e) **Demonstration of the test method's performance should be based on the testing of reference chemicals representative of the types of substances for which the test method will be used.**

A sufficient number of the reference chemicals should have been tested under code to exclude bias (see paragraphs on "Coding and Distribution of Test Samples").

Representativity

For phase-2 of the validation, consultation by the Lead Laboratory with the participating laboratories and other experts led to the proposal for a battery of ten test substances, ranging from pharmaceuticals to substances with broad environmental release and human exposure. These substances, covering a large range of modes of action (see table below) include 6 strongly acting substances and 4 substances that are considered to be weakly acting. The six strongly acting substances were ethinyl estradiol, tamoxifen, methyl testosterone, flutamide, l-Thyroxine, and propylthiouracil, and the four additional weak substances were CGS 18320B, *p,p'*-DDE, genistein, and nonylphenol.

Chemical	**Mode of action**	**Type of chemical**	**Testing phase of the updated TG 407 validation study**
Ethinyl estradiol	Potent oestrogen	Pharmaceutical	Tested in phase-1 and 2
Genistein	Weak oestrogen	Natural isoflavone	Tested in phase-2
Nonylphenol	Weak oestrogen	Industrial chemical	Tested in phase-2
Tamoxifen	Potent antioestrogen	Pharmaceutical	Tested in phase-1 and 2
CGS 18320B	Antioestrogen - Aromatase inhibitor	Metabolic inhibitor	Tested in phase-2
Methyltestosterone (MT)	Potent androgen (aromatisable)	Pharmaceutical	Tested in phase-1 and 2
Flutamide (FLU)	Potent anti-androgen	Pharmaceutical	Tested in phase-1 and 2 (reference chemical in phase-1)
p,p'-DDE (DDE)	Weak anti-androgen	Pesticide	Tested in phase-2
Methoxychlor	Weak estrogen - antiandrogen	Pesticide	Tested in phase-2
Propylthiouracil	Thyroid toxicant	Pharmaceutical	Tested in phase-1 and 2
l-Thyroxine	Thyroid agonist	Natural hormone, Pharmaceutical	Tested in phase-2

Coding

In Phase-1 and Phase-2, none of the substances tested has been coded. The purpose of coded substance is to avoid bias in the results due to prior knowledge of the expected responses. Guidance Document 34 says that it is “preferable” but it is not a requirement.

GD34 - Criteria 6/8

f) **The performance of the test method should have been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data.**

In the case of a substitute test method adequate data should be available to permit a reliable analysis of the performance and comparability of the proposed substitute test method with that of the test it is designed to replace.

In the case of the updated TG 407, the new parameters incorporated to the TG are intended to enable an initial screening of effects (endocrine disruption potential) that are not investigated in the current TG 407. Thus a comparison between current TG 407 and updated TG 407 would not be relevant.

However, in phase-2, the laboratories were requested to conduct the full updated TG 407 so as to assess whether any interference or compromise would be encountered with the functional observation battery or any other current protocol requirements (VR, v). The validation study has proved that the ability of laboratories to perform the entire TG 407 protocol was not negatively impacted by the updates. Laboratories were able to conduct the functional observation and motor activity batteries without interference (VR, xii)

A comparison has been made in most cases between the results of the updated TG 407 and other toxicological assays with longer and or in utero exposures, with the same or similar endocrine active test substances. At least some data were found for all substances except methyl testosterone and l-thyroxine. These comparisons support the ability of the TG 407 to detect systemic, target organ and endocrine-related toxicities for potent chemicals. Where chronic studies or reproductive and development studies easily detected effects, concordant results were typically found in the current TG 407 studies for systemic, target organ, and endocrine toxicities, but where only marginal effects were observed in longer term studies, the updated TG 407 may not be able to detect clear endocrine-related effects.

As expected, the chronic circumstances in the reproductive and developmental studies usually resulted in the findings at lower doses than in the TG 407 studies, but not in all cases such as with PTU. These comparisons are summarised at the end of each test substance chapter in the validation report (VR, 343).

GD34 - Criteria 7/8

g) **Ideally, all data supporting the validity of a test method should have been obtained in accordance with the principles of GLP.**

Aspects of data collection not performed according to GLP should be clearly identified and their potential impact on the validation status of the test method should be indicated.

The laboratories were encouraged to perform the updated TG 407 protocol in compliance with Good Laboratory Practices (GLP). From the reports received from the various laboratories, it appeared that 13 of the 20 studies were carried out in full compliance with GLP. Five of the studies were carried out under procedures that could be considered as generally complying with GLPs, but that would require a quality assurance audit for one or more aspects in order to bring the studies into full, confirmed compliance. Two studies were apparently not carried out in accordance with GLPs.

The GLP status of the studies is summarized in Annex, and additional comments have been provided as to whether the test substance dosing solutions were analyzed to confirm that the nominal doses were indeed administered. This important study element was confirmed to have been done for 18 of the 20 updated TG 407 studies (VR, 24, 25, table 8).

GD34 - Criteria 8/8

h) **All data supporting the assessment of the validity of the test method should be available for expert review.**

The detailed test method protocol should be readily available and in the public domain. The data supporting the validity of the test method should be organised and easily accessible to allow for independent review(s), as appropriate. The test method description should be sufficiently detailed to permit an independent laboratory to follow the procedures and generate equivalent data. Benchmarks should be available by which an independent laboratory can itself assess its proper adherence to the protocol.

It is intended that the materials documenting the protocol development, validation and associated supporting documents will be made freely available by promulgation through the OECD or by publication in the peer reviewed literature.

The validation report has been declassified by the Joint Meeting and is now available on the OECD public website. Additionally, a publication to compare the results obtained with the enhanced 407 with data from standard tests of longer or in utero exposure has been written on behalf of CEFIC Endocrine Modulators Steering Group (CEFIC-EMSG). It has now been accepted for publication in a scientific journal.

A draft of the Test guideline is also available to the Peer reviewers on the OECD public website.

Annex - Good Laboratory Practice (GLP) Compliance of the 407 Studies.

Chemical	Lab	GLP Compliance		Comments on GLP and substance and/or dosing sample analyses
		Yes	No	
Ethinyl Oestradiol	2	✓		EE stock and dosing solution analyses were conducted and reported. For technical reasons, the dosing solutions were analysed at an outside, non-GLP compliant laboratory.
	5	✓		EE stability and stock and dosing sample analyses were stated to have been done, but specific results were not reported.
Genistein	4	✓		Genistein stock and dosing sample analyses were conducted to confirm levels, and the results were reported.
	12	✓		Genistein stock and dosing sample analyses were conducted to confirm levels, and the results were reported.
Nonylphenol	1	✓		Nonylphenol stock and dose sample analyses were conducted during the study, and results were reported.
	6	✓		Nonylphenol stock and dosing samples analyses were conducted during the study, and the analytical results were reported.
Tamoxifen	3		✓	According to the final report: "This study was not performed in compliance with Good Laboratory Practice in that it was not subjected to specific Quality Assurance inspections. It was performed according to standard operating procedures which were previously accepted and periodically inspected by Quality Assurance Unit."
	10		✓	The report did not state whether tamoxifen dosing sample analyses were done, and no analytical results were reported.
CGS 18320B	8	✓		CGS 18320B analyses for the test substance stability and dosing samples were conducted to confirm dosage levels, and the analytical results were reported.
	13	✓		CGS 18320B stock and dosing samples analyses were conducted to confirm dosage levels, and the analytical results were reported.
Methyl Testosterone	3		✓	According to the final report: "This study was not performed in compliance with Good Laboratory Practice in that it was not subjected to specific Quality Assurance inspections. It was performed according to standard operating procedures which were previously accepted and periodically inspected by Quality Assurance Unit."
	12	✓		Methyl testosterone stock and dosing sample analyses were conducted to confirm dosage levels, and the analytical results were reported.
Flutamide	2	✓		Flutamide stock and dosing solution analyses were conducted to confirm dosage levels and the analytical results were reported.
	11	✓		The report did not state whether Flutamide dosing sample analyses were done, and no analytical results were reported.
p,p'-DDE	6	✓		DDE test substance dosing samples were analysed for homogeneity and stability during the study, and the analytical results were reported.
	7		✓	DDE stability and test substance dosing samples were analysed, and the results were reported.
Propyl-thiouracil	1	✓		PTU test substance dosing samples were analysed during the study, and the results were reported.
	10		✓	The report did not state whether PTU dosing sample analyses were done, and no analytical results were reported.
l-Thyroxine	9		✓	Thyroxine test substance dosing samples were analysed, and the analytical results were reported.
	13	✓		Thyroxine analyses for the test substance dosing samples were conducted to confirm dosage levels, and the analytical results were reported.