

List of 15 SPSFs submitted in November 2025 for consideration in April 2026 at WNT-38

Deadline for returning comments to the OECD Secretariat: Thursday 15 January 2026

Please also indicate by one of the following letters your country's expected involvement ("Inv. "in the table) in the project, should it be included in the work plan:

- A. None,
- B. Only NC comments,
- C. NC & experts comments,
- D. NC, expert comments and participation in expert meetings,
- E. NC, expert comments, participation in expert meetings & practical co-drafting/validation work

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
SECTION 2 – Biotic effects on environmental species			
# 1 New TG on Semi-Field Study Design to Investigate Potential Impacts of Crop Protection Products on cavity-nesting solitary bees (Megachilidae)			
Lead: GER/CH			
BIAC	No		1. The methodology and regulatory frame is not well developed enough to be Test Guidance and that a Guidance Document (like OECD 75 – honey bee brood

Indicate in that column whether you are favourable to inclusion of the project on the work plan of the TG Programme (yes), not in favour (No), or if you require further information to make your decision (FIR).

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			test) would be more appropriate at this time. There are uncertainties, particularly for the EU, about data requirements, validity and also how such studies will be interpreted in a regulatory context given that a new bee guidance doc, data requirements and uniform principles are currently under review or pending adoption. Any new method would need take these unknowns into account. In addition: 2. Clarify scope and test organisms: confirm that the TG targets <i>Osmia bicornis</i>, justify any extension to <i>Megachile rotundata</i> (including sanitary/import constraints and availability), and list potential sources/limitations for obtaining cocoons. 3. Define validity and reproducibility criteria: specify environmental/foraging thresholds (e.g., weather triggers or minimum foraging activity), formalize quality/validity criteria used in ring tests, and propose a follow-up ring test to demonstrate consistent, reproducible results across labs. 4. Specify the toxic reference(s): state the recommended reference compound(s) (and fallback options), address dimethoate availability in the EU, and confirm whether the same standard applies to <i>M. rotundata</i>.

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			5. Explain regulatory fit and geographic need: link the method's endpoints to SPGs/risk assessment use (how data will be used quantitatively and in a tiered approach), and justify the study's relevance for regions outside the EU (e.g., US, Japan), accounting for pending revisions to data requirements.
# 2 New TG on Transgenic Eleuthero-embryonic THYroid-Specific (TETHYS) assay Lead: FR/SP			
BIAC	Yes	D	1. The proposed project would it fill a gap in the CF level 3 to 5 testing battery for the T-modality in fish, providing mechanistic information that cannot be delivered by existing TGs and further contributing to reduce animal testing. 2. We would appreciate more clarity on why two models (Medaka and ZF) are proposed for the one guideline – they both feature the same promotor as sensor (Tg) but different reporters. Its not clear whether are there differences in the performance or which aspects are to be compared in the validation. It does carry a risk of diluting the robustness of the validation exercise to spread the effort across the two models unless there is a compelling reason to progress both. 3. Hopefully the test chemical selection for the validation exercise will be mindful of chemicals tested in XETA and the response of the XETA to those, to fully leverage the validation effort in characterisation the respective applicability domains of TETHYS vs XETA wrt thyroid

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			active substances. The proposal is careful to explain that TETHYS is seen as complementary but ideally would replace XETA – as the proposers point out fish assays are available at levels 3-5 of the CF (ie reproduction, a full life cycle can be done in fish, unlike amphibians). So adding a mechanistically informative transgenic fish NAM to the toolbox would be a good thing and ideally would pave the way to using fish models for EAS and T modalities, and potentially reduced animal testing in ED assessment.
SECTION 3 – Environmental fate and behaviour			
# 3 New GD on biodegradation testing of difficult-to-test substances			
Lead: EC/ECHA			
BIAC	FIR		While some industry stakeholders see value in the envisioned guidance, there are many details/questions to be addressed before approving the proposal as written. Further information is required to assess the rationale, relevance, and added value of the proposed project. While guidance on biodegradation testing for difficult-to-test substances is recognised as necessary, the current proposal would require more analysis to ensure it encompasses all relevant work. In particular, the development of such a guidance document should be supported by a comprehensive literature review and extensive input from experts routinely conducting biodegradation assessments on difficult-to-test substances.

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			In addition, the proposed timeline and work plan do not seem to allow sufficient opportunity for (i) the integration of emerging or alternative methodologies, and (ii) thorough expert review and alignment with the current state of the science and best practices. More broadly, it is important to ensure that new GD provides clear added value and does not inadvertently introduce additional complexity or uncertainty for industry. In this context, it may also be noted that there is ongoing work that is supporting this topic on some of the TGs of interest, namely Cefic-LRI funded project (LRI ECO70: Expanding the applicability domain of the OECD 309 Test Guideline for biodegradation) that could support the GD development. This project aims inter alia to “Re-evaluate and establish data-driven boundaries for the physico-chemical applicability domain (volatility, solubility and hydrophobicity)” and “identify suitable methods for dosing difficult-to-test substances and needed adjustments to the test system configuration”. The project will end in Q4 2027/Q1 2028 with preliminary results in Q1 2027.
# 4 New TG Bioaccumulation of Nanomaterials in Hyalella azteca: Aqueous and dietary exposure (nanoHYBIT) Lead: GER			
BIAC	N/FIR		Hyalella azteca will accumulate all inorganic materials and could confound results. Understanding data generated using the TG could be quite difficult based on the natural

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			background noise of inorganics in the sediments. Again, the characterization will be crucial to differentiate natural noise from the intentionally added materials, and I would like to be sure that for the bioaccumulation of inorganics (which represent a large amount of nanomaterials) we are not creating a guideline that cannot recognize background or may require radioactive materials. Additional concerns with the proposed work are: 1. The proposed validation dataset, limited to three test substances, appears insufficient to support a robust and scientifically reliable validation effort. 2. Tested chemistries should have high quality fish bioaccumulation data for comparison purposes. In addition, the proposed work aims to develop a method incorporating both aqueous and dietary exposure. In this context, it may be more appropriate to first establish and validate the approach using a broader range of well-studied conventional substances that have historically been assessed via dietary exposure in fish (<i>e.g.</i> substances with very high logKow and/or low aqueous solubility). Several such substances exist and could provide a stronger empirical basis for method development and performance evaluation.

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
# 5 Revision of OECD GD 280 (2018): 'Guidance Document on the Determination of in vitro intrinsic clearance using cryopreserved hepatocytes (RT-HEP) or liver S9 subcellular fractions (RT-S9) from rainbow trout and extrapolation to in vivo intrinsic clearance' Lead: UK			
BIAC	Yes		We support the revision of the GD 280 to align with the state of the science IVIVE approaches for converting 319A/B results to estimated BCF values. As the authors note, clear, practical, and transparent tools and guidance are needed to support regulatory acceptance and application of 319 data. The authors will be developing tools and analysis for neutral organic and ionizable organic chemistries. This is encompasses the necessary scope of work. We agree that revalidation is not required.
SECTION 4 – Effects on human health			
# 6 Revised TG 492B: EpiOcular time-to-toxicity test method for hazard identification and labelling of eye irritating chemicals Lead: SK			
BIAC	Yes		The addition of a second protocol and a second tissue model from a second supplier promotes a healthy competition, it mitigates the risk that the industry needs to rely on a single proprietary tissue from one supplier as a single point of failure for the entire TG, it increases global availability and accessibility of the tissue and balances a monopolistic position. Besides these rather practical considerations, the addition of EpiOcular to OECD TG 492B will also increase scientific

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			confidence and robustness in the testing approach as it can help to better understand the nuances of the method.
# 7 Reporting Guidance for Biotransformation Systems: Addition to GIVIMP Lead: UK/CZ/SE			
BIAC	Yes		We see value in formalizing reporting of biotransformation systems included in GIVIMP to ensure better transparency and consistent documentation of which biotransformation systems is used, and a checklist is a useful, actionable, yet flexible tool to do that. Increasing transparency and ensuring reproducibility through appropriate documentation is valuable, especially where considerable variability may be expected - e.g. with biotransformation steps/ systems derived from different sources/ in different ways. It is noteworthy that the associated publication/ systematic review seems rigorous and robust. The proposal seems to indicate that the intention is to add the checklist to GIVIMP as published and provided, without further contextualisation in the main document or discussion of impact of such recommendations with relevant stakeholders at the OECD. Though we support going forward, we also encourage some refinement and discussion as part of the process to include contextualisation for use in regulatory testing in the main text of the GIVIMP

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			document. A few suggestions for consideration should this go forward. The following concern the BTS reporting checklist: • Some clarification about the checklist becoming a requirement or being required retrospectively (in cases where e.g., in vitro tests are expected to follow GIVIMP principles). E.g., the number of individuals (box ID) contributing to a preparation of S9 may not have been recorded, but is (currently) considered sufficiently characterised based on species, strain, sex, protein content, and chemical used to induce enzyme activity. • ID. The gender is definitely an important aspect in BTS characterisation. However, I do not see the importance of the number of individuals used in the pooling procedure. Can the authors explain the rationale? • red section/: The first small box ("II") is not actually a box to be checked, but a header (that is followed by the sub-boxes IIA-IIE). This is slightly confusing, and contains duplicate information from the sub-boxes IIA-IIE. • IIA: what is meant by "buffer system in which the BTS rection takes place"? The buffering for, e.g., the S9 itself, or final concentrations - e.g., in the culture?

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			- IIB: utilised solvents for chemical exposure - if this suggest the solvents affecting the biotransformation reaction, this might not be too relevant given the usually shorter-term incubation for biotransformation. If solvents are expected to affect the enzymatic activity/ profile of enzymes in the biotransformation preparation, then the vehicle used during induction (e.g., vehicle for administering b-naphthoflavone and/or phenobarbital to live rats) may have greater impact. This is also at least partially duplicative to box II: "clearly define all reaction components and their respective concentrations within the final BTS reaction mixture". Also, solvents as "reaction components" - I am not sure if one would always expect solvents to react/ be a reaction component. Would maybe vehicle be a better word than solvent, to avoid misunderstanding? - IIC: We agree that characterisation of S9 or microsomal fraction by protein content is more relevant than a volume fraction. However, would it be sufficient to report protein concentration for the fraction preparation, or in the final reaction mixture? Would it be acceptable to have protein content for the preparation, then refer to, e.g., "the S9 content in the final reaction mixture was 10%"? (In my view, this would still support the transparent accurate documentation.)

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			• IID: again, duplication to boxes II, and potentially IIA - cofactors are probably included in the buffer system to enable catalysis. • IIE: This point may not always be technically possible/ required, and in its current form of "define, report, or measure" is neither clear nor specific, therefore not useful. It also is not a reaction component, therefore not appropriate in section II. • Same: The BTS report often confirms its activity in an Ames test using two different pro-mutagens. This is more of a semi-quantitative approach. Measuring the BTS activity is a quantitative approach. Which enzymatic activity should be assessed? We think this should be left open for the individual intention of the BTS usage. • Same. The BTS report often confirms its activity in an Ames test using two different pro-mutagens. This is more a semi quantitative approach. Measuring the BTS activity is a quantitative approach. Which enzymatic activity should be assessed. I think this should be left open for the individual intention of the BTS usage. • Quantification of enzymatic activity (e.g., with probe substrates) would often require analytical methods that may not be always available, and could be excessive for demonstrating certain cases/ activities. This would need

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			clarification if it is good enough to demonstrate the biotransformation system is working (as e.g., currently done for the Bacterial Reverse Mutation (Ames) Test), of whether quantification of metabolism is needed - if so, with which substrate. Potentially alternative wording for point IIE could be: "Were appropriate positive controls used for enzymatic activity, or enzymatic activity measured (under control conditions)?" • IIIB: might be duplicative to IIE. Further: are controls meant on an experiment or BTS batch-basis?
# 8 Revision of TG 497: Proposal for Incorporating a Novel in silico Model to Predict Skin Sensitization Hazard and Potency Lead: JP			
BIAC	FIR/N		Responses to charge questions (outcome of WNT discussion on (Q)SAR models for DATG) to be added. Proprietary model with unclear licensing conditions using additional commercial models and the OECD toolbox while no versions are mentioned. SPSFs#8 and #9 and the proposed tools should be merged if consensus of Yes is reached at WNT-38
# 9 Proposal to Include the in silico RSRKE1 as an Alternative to the in chemico Method in OECD TG 497 Lead: JP			

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
BIAC	No		Responses to charge questions (outcome of WNT discussion on (Q)SAR models for DATG) to be added. Unpublished proprietary model with unclear licensing conditions. SPSFs#8 and #9 and the proposed tools should be merged if consensus Yes is reached at WNT-38.
# 10 New TG on In vitro reconstructed human respiratory epithelium (RHRE) test method for identifying respiratory tissue irritation Lead: US/ICAPO			
BIAC	Y/FIR		While there is some support for this project, there is an important regulatory question. It is not clear to us whether the predicted point of departure will also be translated into classification needs (STOT-SE Category 3). There are concerns that surfactants and many other typical drivers, for e.g. eye irritation, will likely result in a positive and potent response when tested for cellular effects, such as barrier integrity and cell health, after 24 hours of exposure.
# 11 New TG on in vitro gH2AX/pH3 method: a multiplexed biomarker approach that provides information on genotoxic mode of action. Lead: FR/GER			
BIAC	FIR		There were mixed opinions among BIAC genotoxicity experts, and we respectfully request that the project sponsors help clarify some of the questions and concerns raised in the review. Responses that can be characterized as FIR: How would the test guideline address the fact that the endpoints are not

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			“apical”? In other words, mutations are not directly assessed; DNA damage can be repaired, mitotic effects are indicated, etc. Another question is, if the test guideline is adopted, how are regulatory agencies likely to decide where and how the new test guideline fits into their testing strategies? Are they likely to use it to replace existing test guidelines, run concurrently with an existing test guideline, or discount it due to testing complexity and limited CROs that can run the test? This second question may be more speculative but expert team members may have a sense of how different global regulatory agencies would treat a new NAMs-based test. Responses that can be characterized as “No” (from here to the end, including the general and specific recommendations below): • The proposal is narrowly focused on a single way of measuring gH2AX and pH3 despite the existence of multiple recognized methodologies (e.g., Litron’s MultiFlow™, etc.). A guideline limited to one method would be restrictive. • If a guideline is needed for gH2AX and pH3 specifically, the scope should be expanded. Multiple well-recognized methods exist for measuring gH2AX and pH3 (e.g., flow cytometrybased multiplex assays) should be included. The guideline should be method-agnostic, encompassing validated approaches and performance criteria rather than prescribing a single assay. Expanding the scope will require input from additional experts.

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			- Moreover, several key scientific concepts are misstated, and the proposal overstates the gH2AX and pH3 potential to predict genotoxicity. The document should also differentiate screening utility from definitive genotoxic endpoints and address confounding factors with appropriate additional measurements. For instance, gH2AX is not a direct measurement of genotoxicity; it is a biomarker commonly used as a reporter of clastogenic DNA. <i>General Recommendations</i> - Broaden scope: Include all recognized methods for measuring gH2AX and pH3, and engage additional experts. - Clarify role of biomarkers: gH2AX and pH3 are indicators of mode of action (clastogenic/aneugenic) and useful for screening or follow-up, but they are not definitive genotoxic endpoints. These biomarkers do not measure mutagenicity. - Strengthen interpretation context: many confounding factors could influence outcomes and produce false positives. The proposal should strengthen measurements and interpretation of confounders (e.g., cell-cycle progression, DNA content, and robust apoptosis/cytotoxicity markers) to mitigate false results. gH2AX activation can occur under non-genotoxic stress conditions (e.g., protein stress–induced gH2AX activation without true genotoxicity); therefore, interpretation requires complementary markers and context.

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			<i>Specific sections recommendations:</i> - Essential Information Paragraph 1 • “an important requirement for a comprehensive characterization of mutagens that is not generated by other existing Test Guidelines.” (Issue: “Mutagens” is not the right term here). - Essential Information Paragraph 3 • “There are currently no existing Test Guidelines that cover both genotoxic endpoints in one assay. Clastogenicity is assessed in TG 487 and TG 473, while aneugenicity is only assessed in TG 487.” (Issue: this is erroneous, as TG 487 covers both clastogenicity and aneugenicity, contradicting the first sentence. The proposal should acknowledge TG 487’s coverage and clearly define the incremental value of gH2AX/pH3.) - Essential Information Paragraph 3: • “the in vitro gH2AX/pH3 method is the first method that enables simple and reliable discrimination of genotoxic mode of action (clastogens/mutagens, aneugens and misleading cytotoxic chemicals).” (Issues: Overstates reliability on gH2AX/pH3 and incorrectly mixes clastogens with mutagens.) Paragraph 6 • The assay’s limitations extend beyond Aurora kinase inhibitors; other mechanisms can perturb mitosis or induce replication stress (e.g., kinesin inhibitors, proteasome inhibitors, CDK modulators). The guideline should specify which additional

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			biomarkers and controls (e.g., phospho–Aurora A/B, spindle integrity markers, mitotic indices) are recommended to address these cases. There are many examples where gH2AX activation arises from non-genotoxic mechanisms. Paragraph 7 “While gH2AX/pH3 can detect signals from diverse mechanisms (e.g., oxidative stress, bulky adducts, topoisomerase inhibition, crosslinks), not all such signals should be interpreted as evidence of genotoxicity. The proposal should differentiate primary genotoxic mechanisms from secondary stress responses, outline confirmatory testing, and caution against overinterpretation, especially where in vivo relevance may be limited.”
# 12 DRP on in vitro renal cytotoxicity assays for future creation of new TG Lead: JP			
BIAC	No		The proposal says it will focus on a specific test, which appears to be a standard cytotoxicity assay that happens to use kidney cells which is quite far from an in vitro model that predicts in vivo kidney toxicity. Thus we question the utility of the assay. The project only considers the 3D RPTEC model established in Japan and no other country. Assay is only patented in Japan severely limiting accessibility. More transferability efforts should be established prior to seeking guideline status.

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			Furthermore, frozen cells are not yet established and would be needed for international use. • It is not clear what the predictive accuracy of the model is in predicting renal toxicity with all known human nephrotoxics. • The model does not include ureter cells, vascular cells, urinary bladder cells, glomerular cells and immune cells all of which are capable of demonstrating nephrotoxicity. • The model cannot predict crystal formation. • The model demonstrates renal toxicity by indicating a decrease in intracellular ATP. This is unusual. One would expect a creatinine reading. • No performance standards have been established.
# 13 New TG Zebrafish Embryo Developmental Neurotoxicity Screening Test (ZEDNT) Lead: KO			
BIAC	FIR		The proposal for OECD Test Guideline: Zebrafish embryo developmental neurotoxicity screening test (ZEDNT) is interesting and could be potentially valuable as a tool for DNT assessment. Particularly, higher-tiered options ranging several DNT categories (e.g., sensory, neuromuscular) and could be useful. However, from a regulatory perspective

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			(particularly in the agrochemical space), several considerations have been raised. We kindly ask that the following points be considered as a request for further information: 1. There is no mention of the OECD DNT in vitro testing battery (IVB) activities under question 3 “If there are existing Test Guidelines or projects in the work plan of the Test Guidelines Programme covering the same endpoint, please refer to these and describe the added value and usability of the proposed new test method.” a. What is the added value of the zebrafish assay in the context of the already published OECD initial recommendations on the Evaluation of the Data from the DNT IVB (2023)? For example, has there been discussion or investigation on the alignment across endpoints (e.g., do the results from the tail coiling assay correlate with those of the neurite outgrowth assay?) Which existing data gaps in the DNT IVB are addressed by the zebrafish assay? How could the zebrafish assay be integrated into a tiered DNT testing approach? b. Has there been alignment/discussion with the OECD DNT in vitro Zebrafish Expert Group? A zebrafish assay (light-dark transition locomotor assay) was originally planned for inclusion in the OECD DNT IVB (e.g. Sachana

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			et al. 2021 Toward a Better Testing Paradigm for Developmental Neurotoxicity: OECD Efforts and Regulatory Considerations). The assay principle is the same as the locomotor assay suggested in this submission, but with some differences including exposure paradigm and age of assessment. Discussion with the OECD DNT in vitro Zebrafish Expert Group would be especially helpful here as it has worked extensively on issues regarding methodological standardization and reproducibility across laboratories internationally. c. Inclusion of the zebrafish light-dark transitioning test into the OECD DNT IVB was postponed due to lacking robustness and limited information about species differences, toxicokinetics and predictive value. In addition, it was raised that more mechanistic information would be needed to contextualize zebrafish data with DNT IVB outcomes. How do the project submitters plan to address these points within their project? d. As this is proposed to be considered an addition to the OECD TG 236 FET test, has integration of the data been considered in relation to the DNT IVB? Exposure parameters and experimental design between ecotoxicology and human health toxicology experiments

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			are sometimes not compatible as they are addressing fundamentally different questions. This difference in experimental design may make data interpretation in the context of the DNT IVB difficult. 2. Can the authors provide more detailed information in the SOP? Including but not limited to: a. Additional detail on analysis is necessary. The “statistical analysis” section mentions several tests “Probit analysis, Independent Samples T-test, Analysis of Variance”; however it is not mentioned which are used for each assay. For example, for the light-dark transition test, is the total distance moved binned by 15min section, and analyzed via ANOVA (of which assumptions must be met- and are often not met for this type of data, usually resulting in the use of a non-parametric test being performed), and then post-hoc T-test in comparison to vehicle? b. Additional details on data collection are necessary. For example, for both imaging assays, the endpoints collected simply state “abnormal central nervous system formation in the brain and spinal cord, the presence or absence of fluorescence, and the intensity of fluorescence”. Details such as defined criteria of “abnormal” vs. “normal” are necessary (e.g., is this

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			analyzed via imaging analysis program (and what are the parameters measured) or manually via blind scoring (and what are the categories with clear scoring guidance). c. Details on final “DNT scoring” across assays are necessary. How is the “final DNT” score derived, and what are the cutoffs to distinguish a “positive” vs. a “negative” DNT result? i. Clear definitions within each assay will be necessary as well (e.g., if one light period at one concentration shows significantly increased movement with no other significant changes in the assay, is that locomotor assay considered positive?) ii. In the final scoring, is there distinction between general neurotoxicity and developmental toxicity? d. Can the authors please provide more detailed information regarding assay performance and proposed validation testing? What are the 40 chemicals used to optimize the assay? Are these aligned with the test set used for the DNT IVB? What were the criteria for selection of a “positive” and “negative” compound? e. What are the 20 performance standards mentioned?

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			f. Which laboratories will conduct the inter-laboratory validation? Please consider an international approach to assess transferability if not already. Consultation with the OECD zebrafish working group (as mentioned above) would be valuable here. 3. Can the authors expand on how data from this test can be used relative to current DNT assays? a. Is this intended to replace or supplement the OECD 424 and 426? See also question 1 regarding integration of the DNT IVB. b. “mechanistically relevant ED data” is mentioned as a need to be met (section 4). These assays are not mechanistically informative ED assays- can the authors expand on this? c. Can the authors expand on the uncertainties around metabolism (section 6)? 4. Additional information related to the transferability as well as the inter- and intra-lab reproducibility (specifically for the larvae swimming assessment) would be important to know in order to assess the feasibility of developing and implementing this zebrafish embryo DNT SOP. In particular, further clarification is needed regarding the robustness and variability of the larval swimming assessment, which appears to be a key

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			endpoint in the proposed approach. Given the known sensitivity of behavioural endpoints, the development of a detailed, harmonised, and well-standardised protocol (including clear acceptance criteria and quality control measures) should be considered a high priority. 5. The proposal states that ZEDNT would be sufficient on its own to “meet current regulatory needs for hazard identification”. This statement appears premature and is not supported by evidence within the SPSF (e.g. supporting data, peer-reviewed publications, or demonstrated performance metrics). Based on current regulatory practice, it is unlikely that a short-term zebrafish embryo screening approach could replace established long-term mammalian studies, such as OECD TG 426, without a strong and transparent validation demonstrating predictive capacity and reliability. Therefore, the method should currently be considered, as a screening tool whose reliability and regulatory applicability remain to be demonstrated.
# 14 New TG on a Human Mineralocorticoid Receptor Transactivation Assay: MRTA Lead: FR/GER			
BIAC	FIR/N		1. Ahead of validating the MRTA assay, it would be worthwhile for OECD to first consider a project to develop a DRP (Detailed Review Paper). The DRP would help to justify formally validating the MRTA assay and going through the validation, validation review process, and lengthy TG development process. For example, OECD developed a comprehensive DRP for the retinoid

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			system ahead of PEPPER validating and submitting a RARTA assay into the OECD TG process. A DRP comprehensively assess the regulatory need for the MRTA and help the VMG select test substances for a future validation. 2. Currently, the only AOP that this method is linked to is <u>AOP 580</u> which is titled mineralocorticoid receptor activation (agonism) leading to increased body mass index. However, the assay is proposed to be validated for both agonism and antagonism. A DRP would help to inform the need to validate a MRTA assay for the antagonist mode, which has proven to be the more difficult mode to interpret results from in validation studies. 3. The adverse outcome for AOP 580 of increased BMI. BMI is not a true adverse effect but rather a high-level general description of an adverse effect (like e.g., cardiovascular disease and metabolic syndrome) and not accurate for hazard identification. Having a clearly identified adverse effect would be a good starting point to justify the need for a MRTA assay. 4. Taken together, preparing a DRP as a starting point would help to inform the regulatory need and for OECD

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			to support development of a validated TG for the MRTA assay.
# 15 New TG on Estrogen receptor dimerization(α-α/ α-β/β-β) assays to detect chemicals with estrogenic potential Lead: KO			
BIAC	No		1. It is unclear how regulators would find utility with ER dimerization data when ER binding data and ERTA results or just ERTA results are available. 2. Competitive ER binding assays are considered the gold standard for assessing ER agonism/antagonism and to characterize relative potencies between ER ligands. The current EPA OCSPP TG 890.1250 TG uses rat uterine cytosol and utilizes all cytosolic estrogen receptor subtypes that are expressed in this tissue including ERα and ERβ. Along with OECD TG 455, these existing approaches are considered sufficiently comprehensive for identifying and characterising ER-mediated effects in vitro, and they are supported by extensive experience in regulatory contexts. 3. The suggested assay only uses a single dose to detect activity (dimerization). However, the derivation of dose-response curves are required for data quality purposes (e.g., Does dimerization follow a dose response? To derive estimates of relative potency and potential thresholds for ER dimer activity).

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			4. OECD TG 455 currently describes the methodology of mechanistically and functionally equivalent ERTA assays that can reliably detect ER agonist and antagonists. The VM7Luc ERTA assay within OECD 455 uses the VM7Luc4E2 cell line that predominately expresses hERα but has some contribution from hERβ. 5. Part of the rationale for including the AR BRET assay as an appendix to OECD 455 is that the assay does not require use of a radioactive isotope. However, it should be recognized that the OECD ERTA assays are also non-radioactive assays and use luciferase reporter genes that are simple to run and analytically measure. 6. Although the proposed assay results can provide supporting information in the context of an AOP, this information would generally not be considered critical information by a regulator when ER binding and ERTA results are available. Additionally, the ER BRET methodology is captured in a peer-reviewed publication and the cell line can be obtained with an MTA. It is also worth noting that other well-established methodology provides similar information (ER dimerization and ERE binding) has been used for decades in the peer-reviewed literature (e.g., gel-shift assays). In addition, nuclear translocation, and the level of nuclear translocation of ER homodimers, can be assessed as was

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			done by the authors to help illustrate the proof of concept for the ER BRET assay. 7. Previously, the VMG-NA recommended that the similar AR BRET assay should not be a standalone method. Experts on the VMG-NA that did not support the need for the AR BRET assay stating that the AR binding and ARTA assays provide sufficient mechanistic screening information in support of regulatory assessment programs. The same comments largely apply to the proposed ER BRET assay(s). 8. The results of the ER β-β BRET dimerization assay only achieved 67% specificity which is an unacceptably low value for specificity. 9. OECD TG 250 is another NAM that enables the detection of substances that activate or inactivate ERs in medaka embryos. 10. More broadly, the SPSFs appear to be positioned as new TG to expand the available in vitro toolbox for the identification of ED properties, by increasing mechanistic coverage across additional modes of action. While the continued development of mechanistically informative in vitro methods is acknowledged, it is important to ensure that new assays provide clear incremental value in terms of predictive capacity, biological relevance, and regulatory utility. Otherwise, there is a risk of introducing

Comment. Country/ organisation	Position Yes/No/FIR ¹	Involvement	Comments
			additional testing requirements without a proportional improvement in hazard characterisation or decision-making. 11. In this context, it should be recognised that many substances may already trigger a “positive” result in at least one <i>in vitro</i> assay within an endocrine screening battery, particularly when multiple mechanistic endpoints are combined (<u>Burden et al., 2024</u>). This is especially relevant in light of recent discussions on endocrine disruptor information requirements under REACH, where broader screening batteries are being considered. Continuous expansion of the <i>in vitro</i> assay battery without clear performance criteria and interpretative frameworks may increase the likelihood of detecting isolated mechanistic activity of uncertain biological significance. This could lead to a higher rate of “positive” hits that are not necessarily indicative of adversity or endocrine disruption <i>in vivo</i>, thereby increasing follow-up testing and regulatory uncertainty. 12. Therefore, before proceeding, it would be important to clarify the specific regulatory question(s) the proposed assay is intended to address and to demonstrate how it improves upon existing validated methods in terms of reliability, relevance, and added decision-making value.