Environment International 154 (2021) 106574

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Identification of non-validated endocrine disrupting chemical characterization methods by screening of the literature using artificial intelligence and by database exploration

Elias Zgheib [a], Min Ji Kim [b], Florence Jornod [a], Kévin Bernal [a], Céline Tomkiewicz [a], Sylvie Bortoli [a], Xavier Coumoul [a], Robert Barouki [a], Kelly De Jesus [c], Elise Grignard [c], Philippe Hubert [c], Efrosini S. Katsanou [d], Francois Busquet [e], Karine Audouze [a,*]

[a] Université de Paris, T3S, Inserm UMR S-1124, F-75006 Paris, France
[b] Université Sorbonne Paris Nord, Bobigny, INSERM UMR-S 1124, Paris, France
[c] PEPPER, Paris, France
[d] Benaki Phytopathological Institute, Athens, Greece
[e] Altertox, Brussels, Belgium

ARTICLE INFO

Handling Editor: Heather Stapleton

Keywords:
Bioassays
Assay validation
Endocrine disruptors
Guidelines
OECD
AOP-helpFinder
PEPPER

ABSTRACT

Background: Exposure to endocrine disrupting chemicals (EDCs) represents a critical public health threat. Several adverse health outcomes (*e.g.,* cancers, metabolic and neurocognitive/neurodevelopmental disorders, infertility, immune diseases and allergies) are associated with exposure to EDCs. However, the regulatory tests that are currently employed in the EU to identify EDCs do not assess all of the endocrine pathways.
Objective: Our objective was to explore the literature, guidelines and databases to identify relevant and reliable test methods which could be used for prioritization and regulatory pre-validation of EDCs in missing and urgent key areas.
Methods: Abstracts of articles referenced in PubMed were automatically screened using an updated version of the AOP-helpFinder text mining approach. Other available sources were manually explored. Exclusion criteria (computational methods, specific tests for estrogen receptors, tests under validation or already validated, methods accepted by regulatory bodies) were applied according to the priorities of the French Public-privatE Platform for the Pre-validation of Endocrine disRuptors (PEPPER) characterisation methods.
Results: 226 unique non-validated methods were identified. These experimental methods (*in vitro* and *in vivo*) were developed for 30 species using diverse techniques (*e.g.,* reporter gene assays and radioimmunoassays). We retrieved bioassays mainly for the reproductive system, growth/developmental systems, lipogenesis/adipogenicity, thyroid, steroidogenesis, liver metabolism-mediated toxicity, and more specifically for the androgen-, thyroid hormone-, glucocorticoid- and aryl hydrocarbon receptors.
Conclusion: We identified methods to characterize EDCs which could be relevant for regulatory pre-validation and, ultimately for the efficient prevention of EDC-related severe health outcomes. This integrative approach highlights a successful and complementary strategy which combines computational and manual curation approaches.

1. Introduction

Over the past decades, the impact of natural and synthetic substances on human and animal health has been studied increasingly. Some of these substances are referred to as endocrine disrupting chemicals (EDCs) because of their potential to disrupt hormonal homeostasis and the endocrine system thus leading to toxicity (Diamanti-Kandarakis et al. 2009).

In recent years, the number of scientific studies (experimental or computational) which establish a link between exposure to EDCs and human chronic diseases has increased significantly and the level of evidence supporting such associations has been growing steadily

* Corresponding author at: Université de Paris, T3S, Inserm UMR-S 1124, 45 rue des Saints-Pères, F-75006 Paris, France.
E-mail address: karine.audouze@u-paris.fr (K. Audouze).

https://doi.org/10.1016/j.envint.2021.106574
Received 18 November 2020; Received in revised form 5 April 2021; Accepted 8 April 2021
Available online 23 April 2021

(Taboureau et al., 2013; Wu, 2020). Several health outcomes such as hormone-dependent cancers, infertility, immune and metabolic disorders, allergies, as well as neurocognitive and neurodevelopmental diseases have been linked to EDC exposure (Heindel et al. 2017). Moreover, the health-related economic burden resulting from exposure to some EDCs has been estimated to be around €150 billion per year in Europe alone (Trasande et al., 2015). It is due, mostly, to adverse outcomes for which validated assays are not yet available such as decreased IQ.

The mechanisms of action of EDCs and EDC mixtures in biological systems are complex and not well defined. An EDC can alter the activity of the endocrine system either directly, by binding to hormonal receptors (*e.g.* estrogen (α or β), androgen, thyroid, *etc.*) (Gaido et al. 2000; Gore et al. 2015; Lemaire et al. 2006; Wang et al. 2017), or indirectly, by interfering with the activity of enzymes and signaling pathways that control hormone synthesis, metabolism or degradation (*e.g.* Aromatase, Thyroid hormone deiodinase *etc.*) (Mrema et al. 2013), by modulating gene expression (*e.g.* CYP11A, CYP19 *etc.*) (Gore et al. 2015), or even through epigenetic effects (*e.g.* histone modification, DNA methylation of *e.g.* estrogen receptor *etc.*) (Collotta et al. 2013; Kang et al. 2011; Song et al. 2010; Zama and Uzumcu 2009). Nevertheless, additional efforts are needed to fill gaps in knowledge particularly concerning non-EATS (Estrogen, Androgen, Thyroid and Steroidogenesis) related impacts. Moreover, the regulatory requirements to identify EDCs are challenging due to the adaptive nature of the endocrine system, the absence of a single method to define endocrine disruption, the latency between exposure to EDCs during sensitive life stages and the manifestation of adverse responses (Browne et al. 2020) and the relatively limited specificity of endocrine disruption.

Given the importance of all these issues, new and effective methods to test for EDCs are warranted in order to reduce the level of exposure in humans and wildlife. To reach these goals, there is a need to develop a coherent integrated testing strategy (ITS) to identify EDCs and to understand their role in potential health outcomes (Audouze et al. 2020). An ITS will consist of a combination of test methods destined to identify the biological effects of chemical exposures. Although regulatory measures already have been enacted in the EU, critical issues remain to be addressed to ensure an adequate protection of the human population and the environment. Among the research and regulatory programs which have been initiated are the EU Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL) (https://ec.europa.eu/jrc/en/eurl/ecvam/alternative-methods-toxicity-testing/eu-netval) which supports the EU Reference Laboratory European Center for the Validation of Alternative Methods (EURL ECVAM) for the validation of a battery of tests for thyroid disruption (https://ec.europa.eu/jrc/en/science-update/vitro-methods-detection-thyroid-disruptors), the ED guidance document (GD) on characterization of endocrine disruptors developed by the European Chemical Agency (ECHA) and the European Food Safety Authority (EFSA) with the support of the JRC (ECHA, EFSA, 2018), the Swedish Academic Consortium on Chemical Safety (SwACCS, https://www.swaccs.se/), as well as efforts at the international level such as the OECD dedicated programs (OECD Work Related to Endocrine Disrupters, https://www.oecd.org/env/ehs/testing/oecdworkrelatedtoendocrinedisrupters.htm), the United States (US) Environmental Protection Agency (EPA), and the Endocrine Disrupter Screening Program (EDSP, https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-overview). While various testing endpoints (*e.g.* estrogen, androgen and thyroid) were included by the EDSP, few are actually used and, for example, only the estrogen agonism is required by EPA. Recently, a Horizon 2020 EU cluster, which consists of eight research projects, called EURION (https://eurion-cluster.eu) was established. EURION will propose pre-validated test methods to improve some under-investigated and yet important outcomes (*e.g.*, thyroid, female infertility, metabolic disorders *etc.*). One of these projects, OBERON (https://oberon-4eu.com/) will develop, improve, and pre-validate a battery of test methods to detect EDC-related metabolic disorders based on the concept of an integrated approach for testing and assessment (IATA) (Audouze et al. 2020).

The process of validating existing and novel test methods into internationally recognized OECD test guidelines (TG) must be accelerated (Demeneix, 2019). In 2019, PEPPER (Public-privatE Platform for the Pre-validation of Endocrine disRuptors characterisation methods, https://ed-pepper.eu) was created. Its aim is to accelerate the validation of tests in key areas of concern for regulatory purposes by funding pre-validation studies. PEPPER is a project that is prioritized by the French National Strategy on endocrine disruptors and is conducted by the Institut National de l'Environnement Industriel et des Risques (INERIS), accompanied by professional organizations (*e.g.* France Chimie, Fédération des Entreprises de la Beauté – FEBEA), individual businesses, the Maison de la Chimie foundation and the relevant French ministries involved in the National Strategy on EDCs (INERIS, 2019). PEPPER activity is organized into three stages: (1) the identification, compilation and prioritization of existing non-validated EDC test methods, (2) ascertainment of assay repeatability and reproducibility and, (3) support of the validation process of prioritized test methods that are to be sent to international guidelines bodies. PEPPER is the first platform dedicated to the acceleration of the validation of EDC test methods at the European and international level and to the promotion of the utilization of these methods (AFSSI, 2020). Pre-validation, here, refers to the practical operations, such as transferability, repeatability, ring tests that are required for the validation by an international body (typically OECD), which lead to mutual international recognition.

The main objective of this study, initiated in the context of the PEPPER platform, was to generate, through the use of diverse data sources, a preliminary list of relevant existing test methods (*in vitro* and *in vivo)* for EDC characterization, which have not yet been validated. To this end, a computational strategy was developed, and then used to explore the scientific literature by text-mining (TM). Therefore, we took advantage of a recently developed tool called AOP-helpFinder based on artificial intelligence (AI), which uses available text information to automatically identify and extract links between items from dictionary lists (i.e. EDC and test methods). This bioinformatics approach was combined with a manual screening of databases (DBs) and of existing guideline documents in order to capture as much as possible existing knowledge.

2 Material and methods

To assess the number of existing *in vitro* and *in vivo* EDC test methods (validated and non-validated), we utilized a strategy consisting of two steps: (1) AOP-helpFinder, a new artificial intelligence (AI) TM-based approach to screen the scientific literature, was used after updating, and (2) a manual exploration of the DB and guideline websites was performed. The selected test methods (compiled in a structured table) were annotated according to (1) the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters level (from level 2 to level 5: Level 2 refers to *In vitro* mechanistic screening assays which reveal selected endocrine mechanism(s) applicable to humans and/or wildlife. Level 3 indicates *in vivo* screening assays which reveal specific endocrine-mediated mechanisms. Level 4 covers *in vivo* assays which focus on endocrine-relevant adverse effects as well as on multiple modes of action (MoA). Level 5 includes developmental and reproductive toxicity studies which comprehensively assess endocrine-relevant adverse effects extensively covering the life cycle of the organism) (2) the year of publication (3) the country of origin, (4) the data source, and (5) the species investigated.

2.1 Automatic literature screening using an AI bioinformatics TM tool

The literature search was performed using a multi-step procedure. A workflow shows the full strategy (Fig. 1).

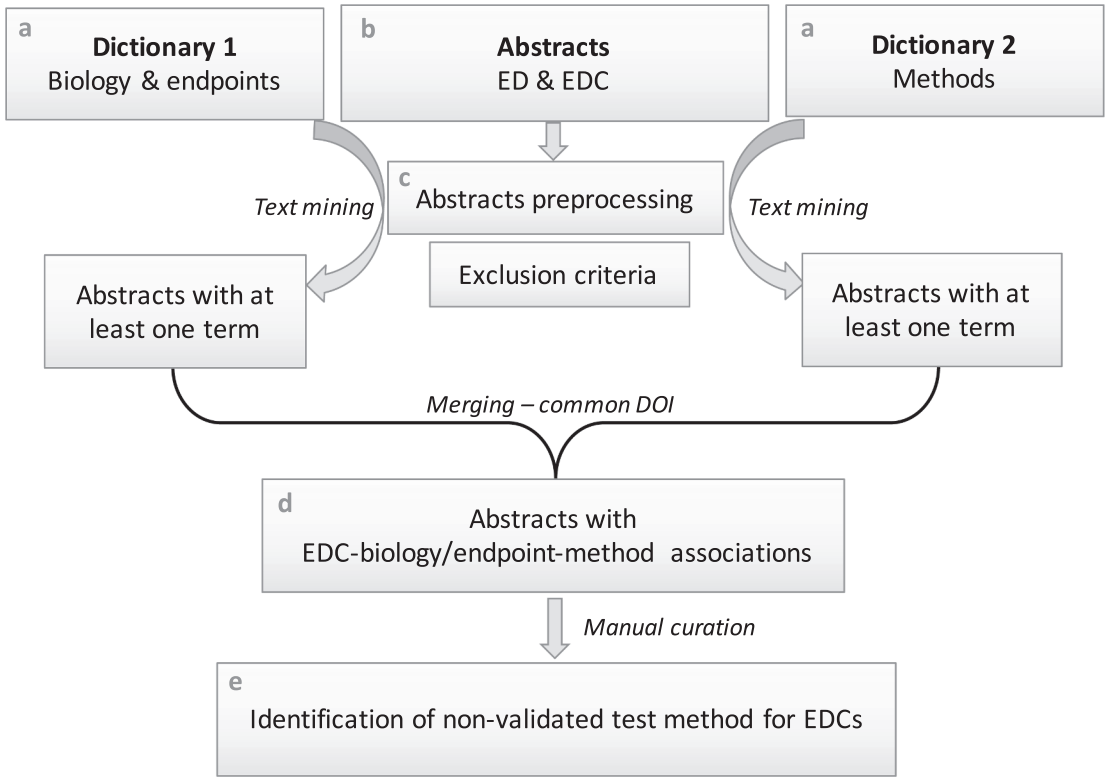

Fig. 1. Workflow of the text mining procedure used to identify EDC characterization test methods: (a) development of the dictionaries of terms; (b) extraction of relevant abstracts for EDCs from the literature, (c) preprocessing of abstracts for automatic screening, (d) identification of co-occurrences between EDCs and terms present in the developed dictionaries, and (e) manual curation by experts for validation.

2.1.1. Development of the dictionaries and exclusion criteria list

To identify test methods that focus on the characterization and identification of EDCs, two dictionaries were generated and refined by experts from different disciplines (biology, (eco)toxicology, regulatory science). The first dictionary focuses on biological endpoints and contains 133 keywords (*e.g.,* thyroid peroxidase). The second dictionary includes 63 technical terms related to assays, tests and tools (*e.g.,* transactivation assay). A list of 9 terms designating exclusion criteria was also developed in order to eliminate validated tests, estrogen-related tests for which sufficient test methods are available, *in silico* methods etc. (Table S1). During the manual curation, exclusion criteria were refined and EDC-related items were excluded if they were not describing an assay, test or method, or if they exclusively linked to the abiotic environment (air, residues, bioaccumulation, biodegradation), microorganisms (except yeast) or plants. Further, if the assay was level 1 (kinetics, physical–chemical properties, *in silico*, mechanistic, bioinformatics) or did not have an endocrine disruption-related endpoint (genotoxicity, mutagenicity, immunotoxicity, neurotoxicity, phototoxicity, eye irritation or corrosion), it was also removed.

2.1.2. Extraction of scientific abstracts related to EDCs

The AOP-helpFinder tool was used to screen the PubMed DB (more than 30 million scientific publications). All published EDC-related abstracts ('endocrine disrupting chemicals' and 'endocrine disruptors') were saved in .xml format files (as of May 6, 2020). Considering the terms used for the searches, the first article was published in 1992.

2.1.3. The AOP-helpFinder tool for the identification of EDC tests

AOP-helpFinder, a recently developed tool which is based on artificial intelligence (AI), uses available text information to automatically identify and extract links between items from dictionary lists. AOP-helpFinder is a hybrid approach that combines the Natural Language ToolKit procedure (NLTK) and Dijkstra graph theory (Carvaillo et al. 2019). Initially, AOP-helpFinder was developed to identify relevant associations between a given substance (or list of substances) and biological events present in Adverse Outcome Pathways (AOP) (Jornod et al. 2020; Rugard et al. 2020). In the present study, the AOP-helpFinder version 2 (Jornod et al. 2020), which was designed, initially, to preprocess and screen scientific abstracts extracted from the PubMed DB, was adapted to automatically retrieve reliable associations between EDCs and test methods.

The preprocessing step maximized the probability of identifying publications related to EDCs and test methods. The compiled abstracts were cleaned and simplified as described previously (Carvaillo et al. 2019). Briefly, this step consists of filtering out the noise (negation, space, punctuation, *etc.*) while keeping only the stemming words. For example, 'increase androgen receptor activity' was kept as 'increas androgen receptor activit', thus allowing the tool to retrieve either 'increase' or 'increased' or 'increasing' in the text. As a result, all the abstracts that were selected were ready for screening against the previously described dictionaries.

The TM tool was then applied independently to both the dictionaries to identify co-occurrences between EDCs and the biological test terms. Only the abstracts that matched with the stemmed terms and EDCs were kept. In the next step, the TM results from both dictionaries were merged using the digital object identifier (DOI) of the publications. This step, thus, preserves abstracts that mention at least one ED or EDC, one biological term of interest and one method term. The TM tool was executed several times to optimize and refine the dictionary terms. This quality control step permitted an assessment of the importance of each dictionary word that was used during the selection process (for example the identification of terms too broad or having an unspecific stem).

2.1.4. Manual curation

A manual curation, by scientific experts, was performed on all the abstracts selected by TM in order to keep only the relevant ones. To categorize the identified tests in the kept relevant abstracts, the full publications were read to complete the results table.

2.2 Exploration of guideline documentation and databases

2.2.1. Identification of EDC testing methods in guidance documents

The validation study of a bioassay is only the first step of a long process with the aim to become an official guideline with a reference code (i.e., OECD TG, EPA 'Official Chemical Safety and Pollution Prevention' (OCSPP) and European test method regulation 440/2008) which can then be used for regulatory testing because it offers mutual international recognition. Validation steps encompass, but are not limited to, the definition of scientific principles, a ring trial or round-robin study to assess reproducibility of the bioassay, its prediction model and its applicability domain. Once validated, the publication describing the results of the bioassay has to be peer-reviewed following which it can be recommended by recognized national or international institutions such as: EURL ECVAM in Europe, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in the US and the Japanese Center for the Validation of Alternative Methods (JaCVAM) in Japan. In order to identify and keep only non-validated bioassays, all the data sources that were used were manually screened (Table S2). Synergies exist between validation bodies under the International Cooperation on Alternative Test Methods (ICATM). Since the OECD (OECD TG) and the US-EPA (EPA OCSPP) have online platforms which are regularly updated, in which all approved guidelines are listed, we adopted a non-exhaustive methodology exploration strategy which listed the totality of items present in these sources.

2.2.2. Screening of databases

Several data sources were considered and checked in order to identify as many assays as possible (validated or not) for EDCs (Table S2). Some DBs, which are more chemical based (CompTox, OpenFoodTox, T3DB and ToxNet), were not explored further due to their website structures which were not compatible with and/or were not relevant to our study. Thus, eight DBs were screened (see Table S2 for the full list and the specificities of the data sources evaluated). Each compiled bioassay (collected information and reference codes from different sources) was manually checked and cleaned when necessary to avoid redundancies in the final table.

2.3 Annotation of the selected test methods according to OECD level and validation status

Regardless of the source, whether from the literature or DB exploration, all of the selected bioassay documents were checked manually by experts to confirm their relevance to the study target (exclusion or not), to define their OECD level and to define their validation status (validated, not validated, validation in progress, *etc.*). A decision tree was developed and used to attain these goals (a workflow of the complete multi-step procedure is presented in Fig. S1). First, exclusion criteria were checked *e.g.* whether the selected item was an experimental test method/bioassay or not. If the test did not pass, the bioassay was excluded with the exclusion reason being mentioned (*e.g.* OECD level 1, no EDC endpoint, non-chemical exposure *etc.*). If no exclusion criteria were identified, information about the test method and its validation status were compiled. Then, the OECD levels were defined using four queries in the following order: (1) Is it an *in vitro* or *in vivo* test? (2) Is the test linked to any endocrine-related endpoint? (3) Is it a toxicity test that might impact endocrine endpoints? and (4) Does it concern population, generational or intergenerational levels?

2.4 A structured table for the non validated test methods that were identified

All of the methods identified from the literature, guideline documentations and DBs were compiled into one unique table (Table S3). This table was organized into three parts: (1) 'Excluded' (903 tests that referred to an exclusion criteria), (2) 'Included – Validated, Ongoing' (306 tests that are included in our study but were already validated or are declared to be under validation), and (3) 'Included – Not Validated' (226 tests that are included in our study and are not yet validated). A note was added when the test method was changed or discarded by regulatory authorities. For all of the categories, the compiled information was preserved using several columns for data source details. For both, 'excluded' and 'validated and on-going' tests, the first 19 columns were dedicated to the codes and the dates of the bioassay source(s). An assay was often found in various sources and under different versions. Our results tables were designed to clearly display the optimal amount of source information that we were able to gather. The first column contains the OECD TG number if the bioassay is a standardized test from the OECD Test Guideline Program, the second column mentions the number of associated OECD GD(s) (if available), the third column mentions the dates of previous versions of the OECD TG and the fourth column the date of the final version. If indexed elsewhere, as, for example, from the EPA OCSPP guideline (number and date), contributions and recommendations by US (NICEATM & ICCVAM) and EU (EURL ECVAM) institutions, tracking system for alternative methods towards regulatory acceptance (TSAR) references and from other national validation bodies (JaCVAM, KoCVAM and BraCVAM), it is so indicated. Using the same logic, the source of information about other DBs and/or laboratories (DB-ALM, PubChem, Tox21, ToxCast, Watchfrog) also is reported. Retrieved references from the literature abstracts mentioning the assays are listed under two columns (PubMed ID and year). Finally, the last column is reserved for the country where the test was developed considering first author's affiliation. In the 'Exclusion' sheet, we included the 'reason for exclusion'. To further simplify, in the 'Included – Not Validated' sheet, all these columns are summarized in four columns: 'Main Reference', 'Other References', 'Year' and 'Country'. Then, for included tests, this is followed by a title assay column, six assay details columns and a validation status column. Finally, the 'Included – Not Validated' sheet is completed by analysis columns that evaluate, using a binary code (1: yes, 0: no), the 'Site of Action' (22 columns) and 'Assay Method' (13 columns) (Table S3).

3. Results

3.1. Screening of the test guideline documents and databases

We manually explored available open access databases such as test guidelines (i.e. OECD & EPA), nationally recognized institutions documents (*e.g.* JaCVAM, BraCVAM, KoCVAM etc.) and specialized databases (*e.g.* DB-ALM, TSAR etc.). The decision tree workflow used for the DB exploration, as well as for the analysis of articles during manual curation is described on Fig. S1. Then, all identified tests were checked against three guideline sources (i.e. OECD test guidelines (OECD TG), OECD guidance documents (OECD GD) and EPA OCSPP guidelines (all last accessed as of May 11, 2020), define their validation status (i.e. 'validated assays', 'assays under ongoing validation', 'assays with withdrawn validation' and 'non-validated assays'). A total of 690 excluded assays was identified (corresponding to 859 unique references), 121 non-validated tests (120 unique references), 20 under ongoing validation (26 unique references), 4 assays with withdrawn validation (5 unique references) and 88 validated methods (179 unique

references) (Fig. S1).

3.2. Literature screening

3.2.1. Data preparation

A total of 10,651 and 11,238 scientific abstracts related to 'Endocrine Disruptors' and 'Endocrine Disrupting Chemicals', respectively, were identified using both keywords and extracted from the 30 million publications in the PubMed DB (as of May 6, 2020). As expected, even if the first article on ED was published in 1992, most of the publications were more recent for both terms (only 95 articles for EDs before 2000 and 142 for EDCs). Using the DOI of the selected abstracts, the two lists were merged to remove duplicates. A large number of abstracts overlapped (9715 articles). 12,175 unique abstracts were retained for further analysis. In addition, as the main objective was to identify non-validated experimental test methods, abstracts which contained exclusion criteria terms (*e.g.,* validated, *in silico etc.*) were flagged.

Then the AOP-helpFinder tool was run independently against the two dictionaries. Only abstracts having co-mentioned terms (ED or EDC and at least one term from a dictionary) were kept. As a next step, TM results from both dictionaries were merged using the DOIs of the publications.

3.2.2. Optimization and refinement of the dictionary terms

A total of three runs was performed to refine and optimize the TM analysis. From the 12,175 abstracts compiled, a total of 8384 abstracts (run 1), 3750 (run 2) and 1198 (run 3) were identified. The optimization phases (run 2 and 3) correspond to the improvement of the dictionaries to be used with the TM tool. For example, in the initial method dictionary (run 1) the terms 'screen' and 'screening' were used. As the TM tool uses the stem of the word, 779 common abstracts were identified for 'screening' and 'screen'. Therefore, only one term 'screen' was kept (run 2). Moreover, some terms that were initially part of the method dictionary were moved to the biological/endpoint dictionary as these terms fit better into biology rather than methods (*e.g.,* adipocyte differentiation) (run 3). Consequently, the biological/endpoint revised dictionary contained 133 terms and the method dictionary 63 terms (Table S1).

3.2.3. Manual curation

Since the abstracts related to ED and EDC are relatively recent (mostly after 2010), we decided to analyze primarily the most recent publications since their probability of being validated or under validation would be rather low. In an initial feasibility study covering years 1992–2020, we found that the older publications were, indeed, less relevant than the most recent ones. This is because some published methods became 'obsolete', maybe due to high technical progress and development of advanced computing methods. These improvements included the development of large scale or High Throughput Screening (HTS) tests. Moreover, some of the tests that fit with our selection criteria either were already validated or were re-used in recent studies and published. Therefore, this reduced the number of PubMed abstracts to be manually curated to 1198 for the period 2011–2020. All these 1198 abstracts were manually checked by scientific experts to confirm the validity of the established links according to human language semantics. This step allowed to prioritize 523 unique abstracts (referenced in Table S3). All the 523 corresponding articles were downloaded and fully read in order to categorize the mentioned tests, and to determine specific information such as OECD classification level, species, substances *etc.* Some of these 523 publications were referring to several tests, some being validated, some other not validated. To define the validation status of these tests, we used the same three guidelines as for the DB, that are the OECD TG, the OECD GD and EPA OCSPP guidelines. A total of 885 excluded assays was identified (from 217 unique PMIDs), 113 non-validated tests (96 unique PMIDs), 195 under ongoing validation (184 unique PMIDs) and 20 validated methods (35 unique PMIDs) (Fig. S1).

3.3 Characterization of non-validated test methods

As a result of the screening, using both sources (literature and DBs) and the information obtained from the test guidelines and guidance documents, only 113 and 121 tests were not yet validated from the literature and the DBs, respectively. Half of the tests were identified from the PubMed DB using TM (113 out of 226), illustrating the complementary of using both approaches.

Further analyses were performed on these 226 test methods (i.e., test method without exclusion criteria). Very few overlaps were found (8, between the literature and different DBs) which highlights the need to explore various data sources as they appear to be complementary. Among DBs, ToxCast (64 selected assays) and DB-ALM (30 selected assays) were the major sources of non-validated test methods, followed by PubChem 13), TSAR (9), Watchfrog (4) and ISO (3) (Fig. 2).

According to the OECD EDC conceptual framework, more than 90% of the selected tests were OECD level 2 (*in vitro*), and no tests were OECD level 5, which require long development periods due to inter/trans-generational effects. All others tests were split between OECD level 3 and OECD level 4 (Fig. 3a). With respect to the species used for the assays, the majority of the selected test methods were based on human cells/tissues (125 assays of the 226), followed by assays on aquatic environment organisms (30) and from biological materials from rodents (26) (Fig. 3b). In total, we were able to identify test methods based on 30 species (Table 1). Most of the non-validated bioassays that were selected were developed in the USA (95), followed by Europe (59) and then Asia (33) (Fig. 4). The number of test methods continues to increase. More than the half of the non-validated tests were developed over the last five years (102) (Fig. 5).

3.4. Biological outcomes of the non-validated bioassays

A majority of the identified bioassays were based on receptor regulation and were classified accordingly (Fig. 6). Most of the assays involved proteins belonging to the nuclear receptor superfamily with endocrine or metabolic functions. Among them, 47 test methods were related to thyroid receptors (8 specifically for TRα, for 10 for TRβ), 20 for different PPARs (4 for each of PPARα and PPARβ and 10 for PPARγ), one for the RAR receptor family, 11 for the liver X receptor-like family (*e.g.,* farnesoid X receptor, FXR (8)) and 3 for the vitamin D receptor-like family (VDR). Eight tests for three retinoid X receptor-like (RXR) were found. Different types of estrogen receptor-like proteins such as 3-keto-steroid receptor-like types (25 tests for glucocorticoid receptors, 5 for mineralocorticoid receptors, 29 for progesterone receptors and 60 for androgen receptors) also were identified (Fig. 6). No data were provided regarding estrogen receptors, since these were excluded according to the PEPPER criteria. In addition, 18 assays related to aryl hydrocarbon receptor (AhR) were identified and 59 others were not linked to a specific receptor but rather to outcomes (reproduction, development, adipogenicity and liver/metabolism-mediated toxicity) (not shown in Fig. 6). It is important to notice that the manual curation of the selected abstracts allowed us to positively enrich the results. Indeed, terms such as nuclear receptors (RXR, AhR, RAR and FXR) that were not prioritized by the experts, therefore not present in the list of query terms, were identified after full text reading. We decided to add these finding in the result table (Table S3) as they are extremely important to endocrine disrupting assessment. The test methods could be classified by assay type (Fig. 6). The most frequent assay types were distributed in 10 categories (*e.g.,* reporter gene assay (RGA) which include luciferase and CALUX, radioimmunoassay, fluorescence *etc.*) (Fig. 6). Fig. 6 does not distinguish between receptors that were screened using AOP-helpFinder (biological endpoints listed in Table S1) and those that were sole byproducts (RXR, AhR, RAR and FXR) that happened to be found through searching for other key words.

4 Discussion

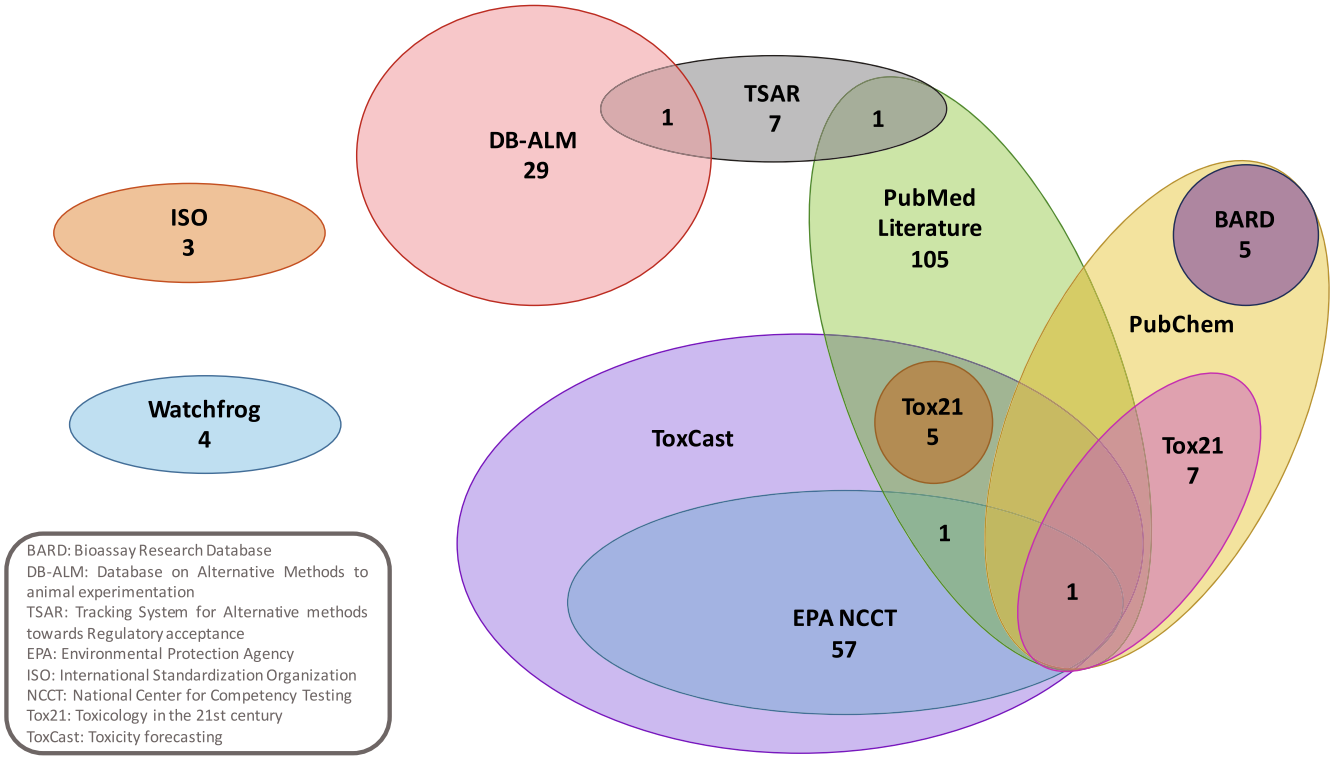

Fig. 2. Distribution of the identified non-validated included test methods for EDCs from the different data sources.

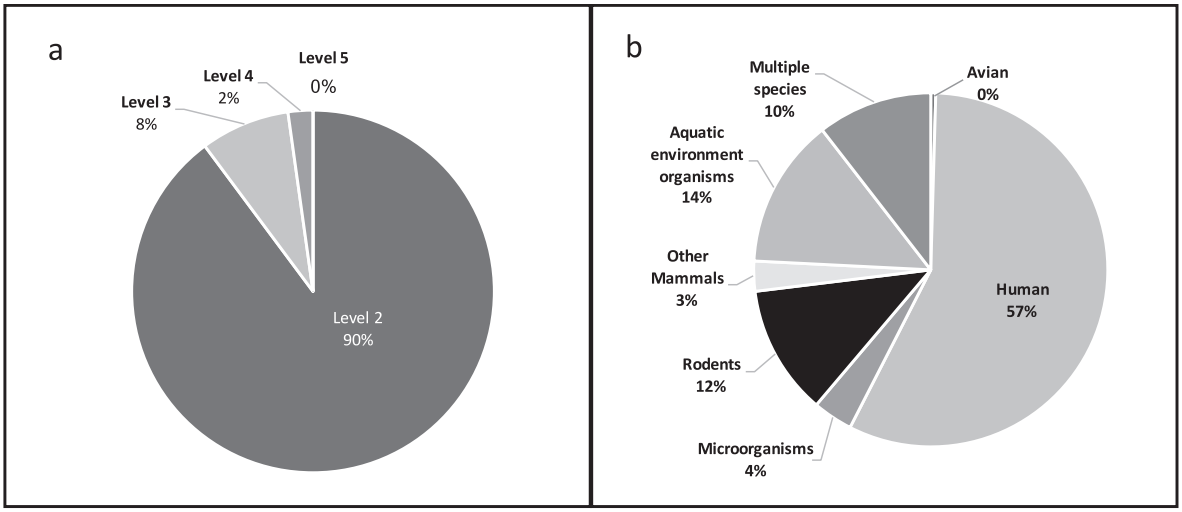

Fig. 3. Distribution of the selected tests according to: (a) OECD EDC conceptual framework level, and (b) species. Multiple species mean that more than one specie is concerned by the test.

We developed an integrative approach by exploring several data sources to identify EDC test methods that have been developed but not yet pre-validated. Combining a computational strategy to automatically explore the available information from the published literature along with manual curation, we built a list of the most relevant EDC test methods which will be used by the PEPPER platform to prioritize assays for pre-validation. This should accelerate the response to the need for validated tests in key areas of concern for regulatory purposes.

The strength of the present study is the screening of various data sources which appear to be complementary to each other since different test methods were identified from the literature and the DBs with very few overlaps. In order to increase knowledge identified with such study, other data sources such as scopus for the scientific literature, might be considered in the future. Another point could be to screen the full text of open access publication as performed by Jensen and al., who use frequent item-set mining to extract fragmented information in full text (Jensen et al. 2012).

Our method goes beyond traditional methods to identify EDC testing methods but it has its limitations and it is not meant to be comprehensive at this stage. This approach is highly dependent on the available information (*e.g.,* what is reported in the publication and the documentation, keywords used when submitting manuscripts, *etc.*). Although the available literature allowed the identification non-validated tests (>100), this information may be limited since papers often are not centered on assays and the description of assays frequently are not extensive. Databases are much more informative as to the description of

Table 1
List of all 30 species studied with the 226 non-validated test methods.

Category	Species	Number of tests
Avian	Chicken (*Gallus gallus domesticus*), Duck (*Anas platyrhynchos*), Quail (*Coturnix coturnix*), Turkey (*Meleagris gallopavo*)	1
Human	*Homo sapiens*	122
Micro-organisms	Yeast (*Arxula adeninivorans*, *Saccharomyces cerevisiae*)	8
Rodents	Mouse (*Mus musculus*), Rat (*Rattus*), Chinese Hamster (*Cricetulus griseus*)	24
Other Mammals	Bovine, Monkey (*Chlorocebus sabaeus*), Porcine (*Cavia porcellus*)	6
Aquatic environment organisms	Amphibians (*Lithobates sylvaticus, Xenopus laevis*), Crustacea (*Acartia tonsa, Americamysis bahia, Amphiascus tenuiremis, Daphnia magna, Neocaridina davidi*), Fish (Fathead minnow (*Pimephales Promelas*), *Gobiocypris rarus*, Goldfish (*Carassius auratus*), Gudgeon (*Gobio gobio*), Medaka (*Oryzias latipes*), *Pleuronectes Platessa*, Rainbow Trout (*Oncorhynchus mykiss*), Zebrafish (*Danio rerio*)) and others (*Potamopyrgus antipodarum, Thais clavigera*)	31
Multiple species	Human & Rodents	5
	Human, Rodents & other Mammals	3
	Human, Rodents & Aquatic environment organisms	2
	Human, & Aquatic environment organisms	1
	Rodents & other Mammals	9
	All types	2
NA	–	12

assays. Furthermore, the TM procedure is limited by the terms used for the searches and the structure of the abstracts. Therefore, it is critical to have multi-disciplinary experts involved in the creation of the dictionaries in order to be able to retrieve as much information as possible.

The main outcomes present in the non-validated EDC test methods that we identified in the different DBs were related to the reproductive system, the growth/developmental systems, lipogenesis/adipogenicity, thyroid, steroidogenesis and liver metabolism-mediated toxicity, and frequently involved the androgen receptor, thyroid hormone receptors, glucocorticoid receptors and the aryl hydrocarbon receptor.

At present, several validated regulatory tests for EATS exists which can be used in combination to provide evidence linking EDCs to health effects. However, many rely on mammalian testing and are costly and time consuming (Browne et al. 2020). Among the EDC test methods identified here, some belong to new approach methodologies (NAM) which are alternative test methods which are supported by diverse organizations including the OECD, ECVAM and ECHA. These NAMs, such as tests in zebrafish, could be implemented in the regulatory process to help regulatory decisions regarding identification of EDCs.

Our study allowed also to identify gaps in knowledge as well as the absence of or the insufficient reporting for some ED modalities and pathways for which the development of test methods should be prioritized in the future. In addition to the use of test methods for regulatory purposes, an inventory of existing methods that provide data at different levels of the biological organization is very important for the development and improvement of adverse outcome pathways (AOP) and quantitative AOPs (qAOPs). Indeed, AOPs constitute a linear organization of the existing biological knowledge and are constructed using different assays. AOPs start from a molecular initiating event which leads to an adverse outcome through several key events. The AOP framework already is very valuable for evaluating potential EDCs and has been applied by the US EPA endocrine disruptor screening program (Browne et al. 2017). Several studies demonstrate that AOPs are widely accepted as a new tool for toxicological safety assessment and that they enable improved use of mechanistic knowledge for regulatory purposes. Recently, ten putative AOPs relevant for female reproductive disorders were proposed to improve testing and regulation of EDCs (Johansson et al. 2020).

5. Conclusion

This study provides, to our knowledge, the first characterization of existing relevant test methods to assess suspected EDCs. The identified

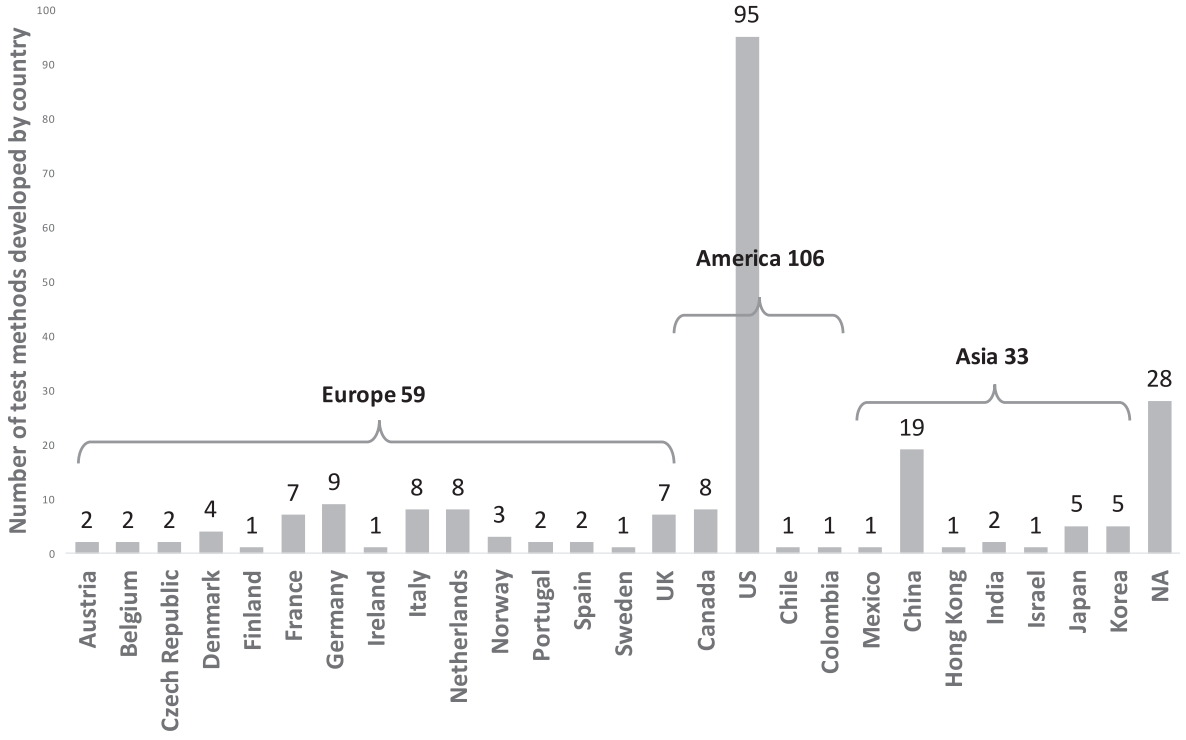

Fig. 4. Distribution of the selected tests by country.

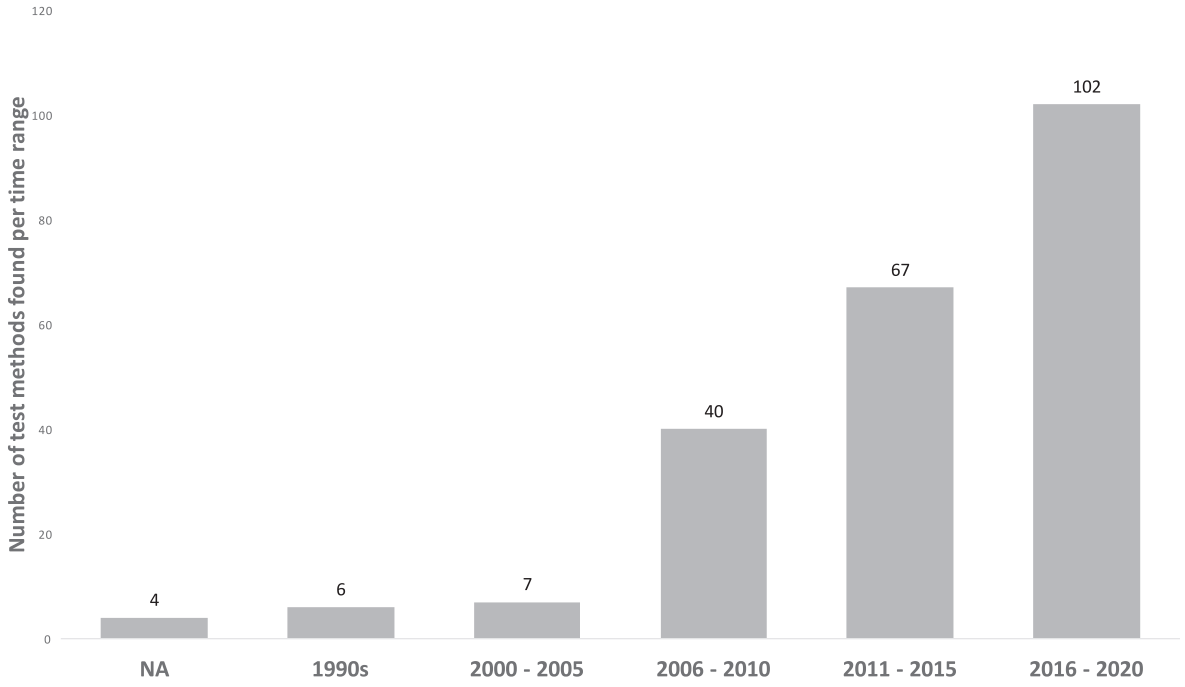

Fig. 5. Distribution of the selected tests by year.

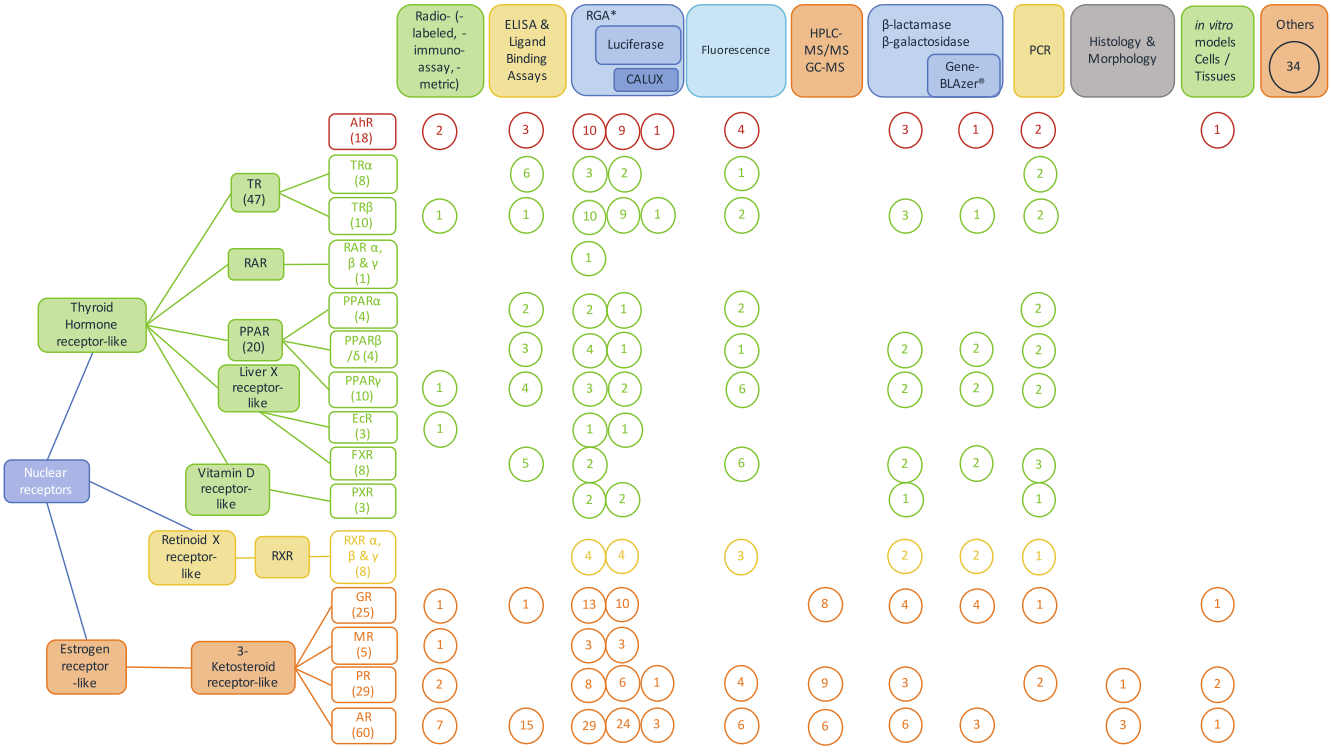

Fig. 6. Representation of included non-validated EDC test methods identified by text mining of the literature and by screening of databases. Methods were classified by hormonal receptor types or biological categories (boxes to the left) across the different types of assays (boxes at the top). The numbers in each receptor box represent the number of identified tests/methods (for example AhR was investigated with 18 tests). The number in each circle represents the number of assay types. One method/test can include one or several assay types. * Reporter Gene Assay.

list is a preliminary step, that will need to pass through other filtering in order to prioritize the most relevant tests for pre-validation, with a final aim of application in regulation. Using diverse data sources, a computational strategy combined with manual curation allowed the retrieval of 259 test methods for EDCs that are not yet validated. These findings can be used to select and then prioritize methods to be pre-validated and subsequently validated for EDC characterization. This approach will assist in filling in the gaps in knowledge concerning the modes of action or adverse effects of these substances. In addition, it will accelerate and improve EDC testing as developed in the EURION EU cluster and,

consequently, support future regulatory measures in the European Union and in Member States to ensure sufficient protection of EU citizens and the environment.

CRediT authorship contribution statement

Elias Zgheib: Resources, Investigation, Visualization, Writing - original draft. **Min Ji Kim:** Data curation. **Florence Jornod:** Methodology. **Kévin Bernal:** Data curation. **Céline Tomkiewicz:** Data curation. **Sylvie Bortoli:** Data curation. **Xavier Coumoul:** Writing - review & editing. **Robert Barouki:** Writing - review & editing. **Kelly De Jesus:** Resources. **Elise Grignard:** Writing - review & editing. **Philippe Hubert:** Resources. **Efrosini S. Katsanou:** Conceptualization, Writing - review & editing. **Francois Busquet:** Conceptualization, Writing - review & editing. **Karine Audouze:** Conceptualization, Methodology, Visualization, Writing - original draft, Supervision.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

The authors would like to acknowledge PEPPER. This work was also supported by the University of Paris, INSERM and by OBERON (https://oberon-4eu.com/), a project funded by the European Union's Horizon 2020 research and innovation program under grant agreement no. 825712 The authors would like to thank Dr Lon Aggerbeck for critical reading of the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2021.106574.

References

AFSSI, 2020. Naissance de la plateforme PEPPER sur les perturbateurs endocriniens. AFSSI Sciences de la Vie. URL https://www.afssi.fr/blog/la-plateforme-pepper-les-perturbateurs-endocriniens/ (accessed 7.16.20).

Audouze, K., Sarigiannis, D., Alonso-Magdalena, P., Brochot, C., Casas, M., Vrijheid, M., Babin, P.J., Karakitsios, S., Coumoul, X., Barouki, R., 2020. Integrative Strategy of Testing Systems for Identification of Endocrine Disruptors Inducing Metabolic Disorders—An Introduction to the OBERON Project. Int. J. Mol. Sci. 21 (8), 2988. https://doi.org/10.3390/ijms21082988.

Browne, P., Noyes, P.D., Casey, W.M., Dix, D.J., 2017. Application of Adverse Outcome Pathways to U.S. EPA's Endocrine Disruptor Screening Program. Environ. Health Perspect. 125 (9), 096001. https://doi.org/10.1289/EHP1304.

Browne, P., Van Der Wal, L., Gourmelon, A., 2020. OECD approaches and considerations for regulatory evaluation of endocrine disruptors. Mol. Cell. Endocrinol. 504, 110675. https://doi.org/10.1016/j.mce:2019.110675.

Carvaillo, J.-C., Barouki, R., Coumoul, X., Audouze, K., 2019. Linking Bisphenol S to Adverse Outcome Pathways Using a Combined Text Mining and Systems Biology Approach. Environ. Health Perspect. 127 (4), 047005. https://doi.org/10.1289/EHP4200.

Collotta, M., Bertazzi, P.A., Bollati, V., 2013. Emerging health issues from chronic pesticide exposure: Innovative methodologies and effects on molecular cell and tissue level. Epigenetics Pesticides Toxicol. 307, 35–41. https://doi.org/10.1016/j.tox.2013.01.017.

Demeneix, B., 2019. Endocrine Disruptors: from Scientific Evidence to Human Health Protection 132.

Diamanti-Kandarakis, E., Bourguignon, J.-P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T., Gore, A.C., 2009. Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. Endocr. Rev. 30, 293–342. https://doi.org/10.1210/er.2009-0002.

Gaido, K.W., Maness, S.C., McDonnell, D.P., Dehal, S.S., Kupfer, D., Safe, S., 2000. Interaction of Methoxychlor and Related Compounds with Estrogen Receptor α and β, and Androgen Receptor: Structure-Activity Studies. Mol. Pharmacol. 58 (4), 852–858. https://doi.org/10.1124/mol.58.4.852.

Gore, A.C., Chappell, V.A., Fenton, S.E., Flaws, J.A., Nadal, A., Prins, G.S., Toppari, J., Zoeller, R.T., 2015. Executive Summary to EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. Endocr. Rev. 36, 593–602. https://doi.org/10.1210/er.2015-1093.

Heindel, J.J., Blumberg, B., Cave, M., Machtinger, R., Mantovani, A., Mendez, M.A., Nadal, A., Palanza, P., Panzica, G., Sargis, R., Vandenberg, L.N., vom Saal, F., 2017. Metabolism disrupting chemicals and metabolic disorders. Reprod. Toxicol. Develop. Origins Dis. 68, 3–33. https://doi.org/10.1016/j.reprotox.2016.10.001.

INERIS, 2019. The birth of the PEPPER platform | Ineris [WWW Document]. URL https://www.ineris.fr/en/ineris/news/birth-pepper-platform (accessed 7.16.20).

Jensen, P.B., Jensen, L.J., Brunak, S., 2012. Mining electronic health records: towards better research applications and clinical care. Nat. Rev. Genet. 13 (6), 395–405. https://doi.org/10.1038/nrg3208.

Johansson, H.K.L., Damdimopoulou, P., van Duursen, M.B.M., Boberg, J., Franssen, D., de Cock, M., Jääger, K., Wagner, M., Velthut-Meikas, A., Xie, Y., Connolly, L., Lelandais, P., Mazaud-Guittot, S., Salumets, A., Draskau, M.K., Filis, P., Fowler, P.A., Christiansen, S., Parent, A.-S., Svingen, T., 2020. Putative adverse outcome pathways for female reproductive disorders to improve testing and regulation of chemicals. Arch. Toxicol. 94 (10), 3359–3379. https://doi.org/10.1007/s00204-020-02834-y.

Jornod, F., Rugard, M., Tamisier, L., Coumoul, X., Andersen, H.R., Barouki, R., Audouze, K., 2020. AOP4EUpest: Mapping of pesticides in Adverse Outcome Pathways using a text mining tool. Bioinformatics btaa545. https://doi.org/10.1093/bioinformatics/btaa545.

Kang, E.-R., Iqbal, K., Tran, D.A., Rivas, G.E., Singh, P., Pfeifer, G.P., Szabó, P.E., 2011. Effects of endocrine disruptors on imprinted gene expression in the mouse embryo. Epigenetics 6 (7), 937–950. https://doi.org/10.4161/epi.6.7.16067.

Lemaire, G., Mnif, W., Mauvais, P., Balaguer, P., Rahmani, R., 2006. Activation of α- and β-estrogen receptors by persistent pesticides in reporter cell lines. Life Sci. 79 (12), 1160–1169. https://doi.org/10.1016/j.lfs.2006.03.023.

Mrema, E.J., Rubino, F.M., Brambilla, G., Moretto, A., Tsatsakis, A.M., Colosio, C., 2013. Persistent organochlorinated pesticides and mechanisms of their toxicity. Toxicology 307, 74–88. https://doi.org/10.1016/j.tox.2012.11.015.

Rugard, M., Coumoul, X., Carvaillo, J.-C., Barouki, R., Audouze, K., 2020. Deciphering Adverse Outcome Pathway Network Linked to Bisphenol F Using Text Mining and Systems Toxicology Approaches. Toxicol. Sci. 173, 32–40. https://doi.org/10.1093/toxsci/kfz214.

Song, C., Kanthasamy, A., Anantharam, V., Sun, F., Kanthasamy, A.G., 2010. Environmental Neurotoxic Pesticide Increases Histone Acetylation to Promote Apoptosis in Dopaminergic Neuronal Cells: Relevance to Epigenetic Mechanisms of Neurodegeneration. Mol. Pharmacol. 77 (4), 621–632. https://doi.org/10.1124/mol.109.062174.

Taboureau, O., Jacobsen, U.P., Kalhauge, C., Edsgärd, D., Rigina, O., Gupta, R., et al., 2013. HExpoChem: a systems biology resource to explore human exposure to chemicals. Bioinformatics 29, 1231–1232. https://doi.org/10.1093/bioinformatics/btt112.

Trasande, L., Zoeller, R.T., Hass, U., Kortenkamp, A., Grandjean, P., Myers, J.P., et al., 2015. Estimating Burden and Disease Costs of Exposure to Endocrine-Disrupting Chemicals in the European Union. The Journal of Clinical Endocrinology Metabolism 100, 1245–1255. https://doi.org/10.1210/jc.2014-4324.

Wang, H., Ding, Z., Shi, Q.-M., Ge, X., Wang, H.-X., Li, M.-X., Chen, G., Wang, Q., Ju, Q., Zhang, J.-P., Zhang, M.-R., Xu, L.-C., 2017. Anti-androgenic mechanisms of Bisphenol A involve androgen receptor signaling pathway. Toxicology 387, 10–16. https://doi.org/10.1016/j.tox.2017.06.007.

Wu, Q., 2020. Computational systems biology as an animal-free approach to characterize toxicological effects of persistent organic pollutants. ALTEX. https://doi.org/10.14573/altex.1910161.

Zama, A.M., Uzumcu, M., 2009. Fetal and Neonatal Exposure to the Endocrine Disruptor Methoxychlor Causes Epigenetic Alterations in Adult Ovarian Genes. Endocrinology 150, 4681–4691. https://doi.org/10.1210/en.2009-0499.