Regulatory Toxicology and Pharmacology 70 (2014) 629–640

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA)

Knut Erik Tollefsen [a], Stefan Scholz [b], Mark T. Cronin [c], Stephen W. Edwards [d], Joop de Knecht [e], Kevin Crofton [d], Natalia Garcia-Reyero [f], Thomas Hartung [g], Andrew Worth [h], Grace Patlewicz [i,*]

[a] Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, N-0349 Oslo, Norway
[b] UFZ – Helmholtz Centre for Environmental Research, Department of Bioanalytical Ecotoxicology, Permoserstr. 15, 04318 Leipzig, Germany
[c] Liverpool John Moores University, School of Pharmacy and Biomolecular Sciences, Byrom Street, Liverpool L3 3AF, UK
[d] U.S. Environmental Protection Agency, Office of Research and Development (ORD), Research Triangle Park (RTP), NC 2771, USA
[e] Environment Health and Safety Division, Environment Directorate, Organisation for Economic Co-operation and Development (OECD), 2 rue André Pascal, 75775 Paris Cedex 16, France
[f] Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Starkville, MS, USA
[g] Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), 615 N Wolfe St, Baltimore, MD 21205, USA
[h] European Commission-Joint Research Centre, Institute for Health & Consumer Protection, Systems Toxicology Unit, Via E. Fermi, Ispra, Varese, Italy
[i] DuPont Haskell Global Centers for Health and Environmental Sciences, Stine-Haskell 320/212, 1090 Elkton Road, Newark, DE 19711, USA

ARTICLE INFO

Article history:
Received 4 September 2014
Available online 27 September 2014

Keywords:
Adverse Outcome Pathway (AOP)
Integrated Approaches to Testing and Assessment (IATA)
(Q)SAR
Read-across
High throughput screening (HTS)
High content screening (HCS)
Cross species extrapolation

ABSTRACT

Chemical regulation is challenged by the large number of chemicals requiring assessment for potential human health and environmental impacts. Current approaches are too resource intensive in terms of time, money and animal use to evaluate all chemicals under development or already on the market. The need for timely and robust decision making demands that regulatory toxicity testing becomes more cost-effective and efficient. One way to realize this goal is by being more strategic in directing testing resources; focusing on chemicals of highest concern, limiting testing to the most probable hazards, or targeting the most vulnerable species. Hypothesis driven Integrated Approaches to Testing and Assessment (IATA) have been proposed as practical solutions to such strategic testing. In parallel, the development of the Adverse Outcome Pathway (AOP) framework, which provides information on the causal links between a molecular initiating event (MIE), intermediate key events (KEs) and an adverse outcome (AO) of regulatory concern, offers the biological context to facilitate development of IATA for regulatory decision making. This manuscript summarizes discussions at the Workshop entitled "Advancing AOPs for Integrated Toxicology and Regulatory Applications" with particular focus on the role AOPs play in informing the development of IATA for different regulatory purposes.

1. Introduction

1.1. AOPs to support IATA in regulatory decision-making

Within the last decade, the global economy has witnessed a significant shift in the regulatory environment coupled with the volume and diversity of industrial chemicals being manufactured. Despite some regional differences, regulatory management in general comprises hazard identification/characterization, an exposure assessment and a risk assessment as its main steps. In some cases, the identification of hazards is prior to market approval and certain hazards e.g. carcinogenicity (C), mutagenicity (M), or reproductive (R) effects (CMRs) may lead to restrictions on use irrespective of any subsequent risk assessment.

The hazard identification step is driven by a desire to identify all the hazards of potential concern and assign the appropriate hazard classification (i.e. classification and labelling requirements) regardless of the relevance of these hazards as a consequence of exposure. The hazard characterization step is often associated with extensive *in vivo* toxicity testing using standardized guidelines or protocols. The time, cost and animal use to generate such hazard data are significant and difficult to achieve in practice given the

* Corresponding author at: DuPont Haskell Global Centers, Newark, DE 19711, USA. Fax: +1 302 451 4531.
E-mail addresses: knut.erik.tollefsen@niva.no (K.E. Tollefsen), stefan.scholz@ufz.de (S. Scholz), M.T.Cronin@ljmu.ac.uk (M.T. Cronin), Edwards.Stephen@epa.gov (S.W. Edwards), Joop.DEKNECHT@oecd.org (J. de Knecht), Crofton.Kevin@epa.gov (K. Crofton), natalia@icnanotox.org (N. Garcia-Reyero), thartung@jhsph.edu (T. Hartung), Andrew.WORTH@ec.europa.eu (A. Worth), patlewig@hotmail.com (G. Patlewicz).

http://dx.doi.org/10.1016/j.yrtph.2014.09.009
0273-2300/

large number of chemicals that need to be evaluated. Furthermore, the classical *in vivo* testing approach is based on apical endpoints, which typically provide minimal information on the mode or mechanism of action. This limits the development and application of new *in vitro* assays, read-across approaches or inter-species extrapolation, all of which could facilitate an initial hazard assessment. In addition, the societal demand to use (vertebrate) animal tests only as a last resort for obtaining hazard information coupled with the increasing number of different regulatory programs worldwide calls for a re-consideration of traditional assessment strategies and incorporation of alternative approaches. At the same time, substantial advances have been made in the use of high throughput (HT) and high content (HC) screening assays to quantify and characterize molecular and cellular responses to chemicals (Kavlock et al., 2007; Judson et al., 2014; Kleinstreuer et al., 2014). A shift towards more mechanistically-based alternative approaches represents a promising opportunity for assessing hazards of regulatory concern. To that end the Adverse Outcome Pathway (AOP) framework provides the biological context and supporting weight of evidence (WoE) to facilitate the interpretation of such alternative data. An AOP represents the existing knowledge concerning the causal linkages between the molecular initiating event (MIE) and the cascade of intermediate or key events (KEs) at the subcellular, cellular, tissue and organ level that lead to a specific adverse outcome (AO) at the individual or population level (Ankley et al., 2010; OECD, 2013). This conceptual framework enables information and data from different chemicals, different levels of biological organization, and different taxonomic domains relevant for one AOP to be assembled. Well-developed AOPs may therefore be expected to help guide identification of experimental testing (e.g. *in vivo*, *in vitro*, *in chemico*) and non-testing (*in silico*) approaches to support regulatory decision making. There is now a need for an objective framework to interpret the results from novel test methods and their prediction models in order to facilitate their application in regulatory decision making. Such a framework will conceivably consist of three main elements: the AOP itself, non-animal (alternative) test methods and *in silico* approaches targeting key components of the AOP, and their associated prediction models for a particular regulatory context. The synthesis and integration of these elements form the basis for developing Integrated Approaches to Testing and Assessment (IATA) that may be used in regulatory applications.

This manuscript summarizes discussions from the Workshop entitled "Advancing AOPs for Integrated Toxicology and Regulatory Applications" held in Somma Lombardo, Italy on the 2–7th March, 2014 (https://aopkb.org/saop/workshops/somma.html). Specifically it captures the discussions and insights derived within the workgroup that discussed the role that AOPs can play in informing the development of IATA for regulatory purposes. The next section defines IATA and related terms. Following that, the main elements or components that make up IATA are described including considerations, (e.g. scientific confidence), that are associated with these different elements and their integration. Once the components have been defined, the overall applicability and limitations of IATA for different regulatory purposes are considered. These concepts are then illustrated by way of three examples that are supported by specific AOPs at different levels of development. A final summary considers how the proposed conceptual framework may impact different regulatory applications.

2. IATA and related concepts

Integrated Approaches to Testing and Assessment (IATA) are structured approaches that integrate and weigh different types of data for the purposes of performing hazard identification (i.e. the potential to cause a hazard), hazard characterization (e.g. the toxic potency) and/or safety assessment (i.e. the potential/toxicity potency related to exposure) of a chemical or group of chemicals. For the purposes of this paper, IATA will be generally referred to in a singular form to represent a specific case rather than a collective approach. An IATA should be viewed as an iterative process that includes efficiency analyses to determine whether more data, and what type of data, are required to make effective regulatory decisions while reducing reliance on animal testing. An IATA is not a novel concept per se, indeed it has been discussed at a special OECD workshop on IATA in 2007 (OECD, 2008) and described by the US EPA as part of a FIFRA[1] Scientific Advisory Panel document in 2011 (US EPA, 2011).

An IATA initially gathers and weighs relevant existing information to derive an initial conclusion. If the existing information is insufficient to address the regulatory or safety decision under consideration, it guides the generation of new data using a hypothesis-driven approach with the goal of addressing the residual uncertainty preventing a regulatory decision. The benefit of an IATA lies in the potential breadth of information that can be used in the assessment, as it may exploit both non-testing (*in silico*) and experimental (*in vivo*, *in vitro* and *in chemico*) approaches. The IATA is considered a generic approach and may encompass testing strategies such as integrated testing strategies (ITS), sequential testing strategies (STS), as well as weight of evidence (WoE) considerations (OECD, 2014a). Both ITS, i.e. the fixed and structured integration and weighing of relevant information to support the final decision (Ahlers et al., 2008; Hartung et al., 2013a,b), and STS, i.e. the fixed stepwise approach involving interim decision steps to reach a decision, represent structured and formal processes to derive a conclusion (OECD, 2014a). In contrast, WoE considerations, i.e. the structured, systematic, independent and transparent review of existing and available data without the use of experimental or computational efforts, aim to perform a reliable and relevant compilation of knowledge intended for a certain regulatory purpose (Balls et al., 2006; OECD, 2014a). Whilst an IATA provides a structure for data integration and a means for targeting testing for particular uses, it is not necessarily framed by any mechanistic rationale. There is growing support for using AOPs to provide such a mechanistic basis (OECD, 2013). Thus, AOP-informed IATA development may drive the development of *in silico*, *in vitro*, or *in chemico* approaches that are anchored in well-developed knowledge as captured within an AOP. Exposure considerations and the use of exposure assessment tools may also form an integral part of an IATA.

AOPs are expected to provide insight into the biological relevance, reliability, and uncertainties associated with the results from *in silico*, *in chemico* and *in vitro* approaches for regulatory use. AOPs also have substantial merit over traditional assessment strategies. For instance, they can assist manufacturers and regulators to identify whether a potential hazard can be expected that justifies subsequent detailed testing. Furthermore, in environmental hazard and risk assessment, they show great promise in species-to-species extrapolation critical for protection of endangered species (Perkins et al., 2013). AOPs could also help to design ITS, which ideally cover the relevant key events of an AOP. AOPs are intended to provide a transparent evaluation of available evidence and relevant data, scientific confidence is envisioned to be evaluated through approaches akin to the "Bradford Hill Considerations" developed originally in epidemiology (Hill, 1965; OECD, 2013). Briefly, by examining: (1) biological concordance, (2) essentiality of Key Events, (3) concordance of empirical observation (encompasses dose response and temporal concordance and beyond), (4) consistency (among different biological contexts) and (5) analogy

[1] Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA).

(consistency across chemicals), a clear statement regarding the supporting evidence for the AOP can be developed (Meek et al., 2014a,b). Depending on the outcomes for these considerations, a given AOP may differ in its level of scientific rigor and confidence, which in turn will drive its practical suitability in addressing different regulatory applications (Perkins et al., submitted; Patlewicz et al., submitted).

The practical implementation of an AOP-informed IATA for a given chemical or group of chemicals considers problem formulation based on the risk management scope and goals, the selection and evaluation of suitable AOPs to inform the IATA and existing information that is available for the chemical(s) of interest. All these considerations will influence the makeup of an IATA in terms of the different types of testing (e.g. *in chemico*, *in vitro* and *in vivo*), non-testing (e.g. *in silico*), or data integration approaches (e.g. ITS, STS, WoE or other IATA strategies) that can be exploited (Fig. 1.). Fig. 1 outlines a proposed framework to guide how existing information (e.g. hazard and exposure information) needs to be evaluated and what new data, if any, needs to be generated, so that the IATA can lead to a regulatory decision.

3. Elements for developing AOP-informed IATA

Non-testing and testing approaches as well as data integration strategies form the elements or building blocks that are necessary to derive IATA. These elements are described in more detail in terms of the applicability and limitations in the following sections.

3.1. Non-testing approaches

Non-testing or *in silico* approaches serve two functions within an IATA, they either provide a way to organize existing information or they are used to make predictions of molecular initiating events (MIE) or other key events (KEs) as defined in an associated AOP. The breadth of non-testing approaches is extensive. They range from the search and retrieval of existing data, to the identification of structural fragments to indicate activity and assist grouping (e.g. structure–activity relationships (SARs), read-across), to quantitative models (e.g. quantitative structure–activity relationships (QSARs). Some of these SARs or QSARs may be housed in software tools known as expert systems for ease of use. A summary of non-testing approaches that may be useful in the development of IATA are described in more detail in Cronin and Madden (2010).

Within IATA, non-testing approaches will most likely be exploited to characterize the MIE within AOPs either qualitatively or quantitatively (Table 1). A number of different types of MIEs, and thus AOPs, may be identified for a given IATA in order to generate sufficient information for the decision to be made.

3.1.1. Confidence factors for in silico models in IATA

Assuring scientific confidence in the validity of *in silico* models and their outcomes are key considerations for their application. For (Q)SARs, the OECD has developed validation principles which provide a framework for assuring the scientific validity (relevance and reliability) of a (Q)SAR model (OECD, 2004, 2007). The (Q)SAR model reliability is a relative concept, depending on the context in which the model is applied; meeting each and all of the OECD principles is not necessarily warranted. It is worth noting that the OECD principles only focus on the scientific validity of a given (Q)SAR model, and not on the prediction it generates. The adequacy of a (Q)SAR result for a given compound also needs to be considered before use. In the context of the European Regulation for Registration, Evaluation, Authorisation and restriction of CHemicals (REACH) (EC, 2006; ECHA, 2008) the following specific conditions are considered when evaluating *in silico* models:

1. the estimate should be generated by a valid (relevant and reliable) model;
2. the model should be applicable to the chemical of interest with the necessary level of reliability;
3. the model endpoint should be relevant for the regulatory purpose.

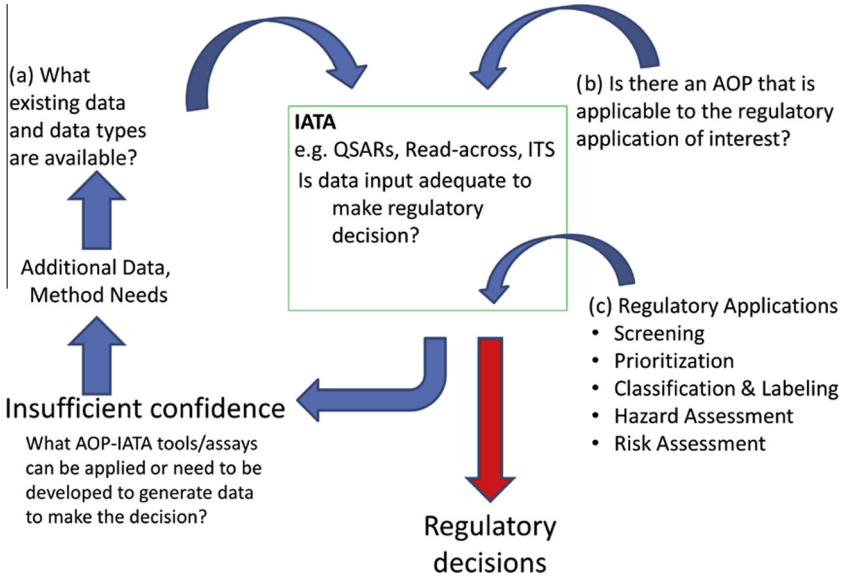

Fig. 1. Conceptual framework for an AOP-informed IATA to support regulatory decisions. The framework is driven by the problem formulation, which involves a consideration of the risk management scope, the data requirements and the level of acceptable uncertainty associated with the decision being made. The regulatory application will also provide an indication of the level of AOP confidence, ideally needed. The framework, which comprises different elements (testing and non-testing approaches, etc.), will evaluate the existing information that is available for the chemical(s) of interest (a), the type of information that might be required as defined by the AOP itself (b), and other relevant information that is pertinent in making a regulatory decision (c). If the outcome generated based on the framework is of sufficient confidence for the regulatory purpose of interest, no further action is warranted. If the outcome derived from the framework is of insufficient confidence, then additional data might need to be generated through new testing and assessment. The new information derived will then be passed back into the framework for re-evaluation. Indeed a decision outcome could result in more thorough regulatory follow up or implementation of measures to reduce use and/or exposure. Any new information generated will also be used to augment the corresponding AOP.

Table 1
Examples of MIEs within AOPs that may be relevant to IATA endpoints derived from non-testing approaches.

MIE	Effect	Examples of *in silico* tool(s)
Unspecific – no definable single molecular site of action	General accumulation in cellular membranes leading to e.g. narcosis, basal cytotoxicity etc	Classification schemes e.g. Verhaar implemented in the OECD QSAR Toolbox or Toxtree. QSARs based on hydrophobicity
Non-specific covalent binding (and formation of radicals)	Irreversible binding to cellular protein and/or DNA which may lead to a variety of effects; e.g. fibrosis	OECD QSAR Toolbox profilers for protein and DNA binding. Quantum chemical calculations
Redox cycling leading to disruption of specific pathways	Mitochondrial toxicity	Structural alerts e.g. Nelms et al., in press
Receptor mediated effects on signalling pathways	A wide variety of acute and non-lethal effects e.g. estrogen receptor binding	3-D molecular modelling. Toxicophores/alerts e.g. EPA ER binding expert system as encoded in the OECD QSAR Toolbox, DART system (Wu et al., 2013)
Physical effects	Skin corrosion	Structural alerts, physicochemical properties e.g. pH
Unknown or very poorly defined MIE	Idiosyncratic drug toxicity	Structural alerts

Whilst these were outlined specifically for REACH, the conditions could be conceivably adapted to address other regulatory purposes.

For chemical categorization (OECD, 2014b) (e.g. read-across), no such principles have been formalized. Systematic frameworks to aid in the evaluation of read-across and identify associated uncertainties are in development by European Chemicals Agency (ECHA) (known as the Read-Across Assessment Framework (RAAF)) as well as by Industry (Blackburn and Stuard, 2014). To date these frameworks do not specifically consider the role of AOPs or how alternative data characterizing MIEs or other KEs may be conceivably used to address uncertainties. Work underway within the SEURAT program (ChemWatch, 2014 – see: http://chemicalwatch.com/19594/seurat-1-homes-in-on-test-chemicals-for-read-across) and independently by DECO-2, a Cefic-LRI AIMT-4 project (Patlewicz et al., in preparation) are both aiming to investigate the feasibility of enhancing read-across by using the AOP concept.

It is noteworthy to mention that there will be clear instances when *in silico* approaches will not provide meaningful information in the context of an IATA, for example if there is no direct linkage to the MIE because the MIE is unknown or ill-defined. Predictions from *in silico* approaches will also be inappropriate, when the target substance is outside of the applicability domain of the model.

3.2. Testing approaches

There are many testing approaches that can form key elements within an IATA – from *in chemico*, *in vitro* to *in vivo* experimental efforts. Testing elements such as toxicogenomics, high content/high-throughput screening (HC/HT) in particular will play a crucial role in shifting IATA away from a reliance on *in vivo* information addressing one or multiple adverse outcomes.

3.2.1. In chemico tests

Biological effects of chemicals can be provoked by an initial covalent modification of a biological macromolecule. The covalent modification of DNA leading to mutagenesis or the reaction with immunoproteins resulting in immunosuppression represent prominent examples (Cronin et al., 2009). *In chemico* tests are experimental measurements that address these covalent modifications without involving biological organisms (reviewed in Schwöbel et al., 2011). These assays are usually used to identify, and in some cases estimate, the intrinsic reactivity of substances to a specific biological target and in that respect are best suited to target the MIE within an AOP. Most *in chemico* tests relevant to toxicity prediction have investigated the reaction of an electrophilic molecule (normally assumed to be the toxicant) with a model nucleophile (representing a surrogate for the target biological macromolecule) (e.g. Roberts et al., 2008; Aptula and Roberts, 2006; Schultz et al., 2005; Thaens et al., 2012). Also included in this type of data could be the assessment of oxidizing behavior and the role of other reactive species (nucleophiles, reactive oxygen species, radicals) principally amenable to *in chemico* testing (Cronin et al., 2009).

3.2.2. In vitro and alternative test systems

Cellular *in vitro* systems, lower vertebrate embryos and invertebrates are proposed and used as alternative test systems to indicate toxic potential to various organisms. Relevant information on the toxic potential of a chemical can be obtained via e.g. comparison of the toxicity to baseline toxicity as an indicator of a non-narcotic or specific mode of action (Escher and Schwarzenbach, 2002). By including appropriate endpoints, it is possible to target MIEs or KEs relevant for an AOP-informed assessment. Extrapolations from alternative test systems, however, have to consider that the toxicokinetic properties may greatly differ and result in deviating effect concentrations between e.g. *in vitro* and *in vivo* tests. Fish/amphibian embryos or invertebrates – despite their evolutionary distance to e.g. mammalians or other vertebrate classes – may provide in some cases a higher predictive capacity than *in vitro* systems given that they represent/accommodate the complexity of a whole organism (Perkins et al., 2013).

3.2.3. High throughput screening assays

High throughput screening assays (HT) comprise *in chemico* and certain *in vitro* test methods such as receptor binding or receptor transactivation assays (Romanov et al., 2008), cellular reporter assays (Romanov et al., 2008; Kleinstreuer et al., 2014), assays using invertebrate (e.g. *Caenorhabditis elegans*, *Drosophila*, algae, crustaceans, see Perkins et al., 2013) or fish embryos (Truong et al., 2014). Toxicogenomic (transcriptomics, proteomics, and metabolomics), utilizing non-biased screening approaches may play a more important role in the future within IATA, since they allow more detailed insights into mechanisms of action and can be applied to survey the breadth of molecular/cellular effects relevant for a wide variety of AOPs (Garcia-Reyero et al., 2014a,b).

Assays targeted towards MIEs, can be very specific for a distinct target (e.g., receptor, enzyme) that leads to an AO. This specificity will also provide the chemical structure and bioactivity data needed to foster development of *in silico* models (as described in Section 3.1). Assays that target downstream KEs such as more generic stress responses (Simmons et al., 2009) may not have this specificity, but may provide an approach to integrate multiple MIEs (Miller et al., 2009). More importantly, analysis of downstream KEs provide the opportunity to predict an AO even in cases where the precise MIE is not known or is not fully understood. Table 2 presents several examples of alternative experimental testing approaches that may be relevant to predict AOs within an IATA.

3.2.4. Confidence factors and limitations for testing approaches in IATA

The use of alternative testing approaches provide higher confidence when they are scientifically and technically valid for use. Validation of alternative assays in particular HT/HC assays has

been subject of several publications such as Judson et al. (2013), Hartung et al. (2013a,b), Patlewicz et al. (2013). In the latter, scientific confidence was discussed in the context of the existing validation frameworks for (Q)SARs and biomarkers (Institute of Medicine, 2010). In Cox et al. (2014), a scientific confidence framework was proposed comprising three inter-related elements to facilitate the systematic, transparent and objective evaluation and documentation of HT/HC assays and their associated prediction models. The elements comprise analytical validation, qualification and utilization. Analytical validation would entail an assessment of the biological basis and analytical performance of the assays. This would involve a consideration of what events within the AOP the assay(s) were mapped to – whether they target the MIE or other downstream KEs. The applicability domain of the assay in terms of the chemical coverage and the typical performance statistics – sensitivity, specificity, accuracy, would be considered as well. The qualification step would involve an assessment of the associated prediction models derived from such assays and utilization would consider the intended regulatory application based on the previous 2 steps.

Even when assays have been scientifically and technically validated, they may exhibit certain limitations. Most assays do not consider the impact of potential metabolic transformation, which can lead to reduced sensitivity (in case of *in vivo* metabolic activation) or to a high number of false positives (in case of *in vivo* inactivation) or false negatives (in case of *in vivo* bioactivation). Furthermore, certain compounds are difficult or impossible to test using *in vitro* systems, for example due to their poor solubility in the culture medium, aggregation potential, volatility, or partitioning behavior (tendency to adsorb onto plastic). In such cases *in silico* methods could provide a more appropriate approach (Zaldivar et al., 2010, 2011).

3.3. Data-integration strategies

Whilst there has been a tendency to define one "definitive" test for hazard assessment in the past, increasingly the need for more than one piece of evidence for hazard assessment has become evident. This need is fundamental in both the AOP concept and the AOP-informed IATA. Therefore, data integration strategies are needed to integrate *in silico*, *in chemico*, *in vitro*, *in vivo*, and available epidemiological or clinical data which:

1. Allow for the combination of low-cost (sensitive) screening assays with more sophisticated (specific) confirmatory assays.
2. Consider the incomplete coverage of one assay in the chemical universe (applicability domain), severity classes or modes of action.
3. Compensate for the insufficient reliability of a single test.
4. Combine kinetic and exposure information, with (quantitative) *in vitro* to *in vivo* extrapolation.

Testing and non-testing outcomes can be manually integrated together to derive an outcome for specific regulatory purposes. This is relatively straightforward for a simple linear AOP with a limited number of KEs, such as that for skin sensitization (OECD, 2012a,b). As more AOPs are developed, and KEs are identified that cut across different AOPs into networks of interlinked AOPs, the complexity of data integration supporting an IATA will increase. Manual integration of a myriad of KEs may not be feasible to do. Moreover, some of the assay outcomes or prediction models derived may require interpretation, a translation step to convert the raw test outcome into a form that addresses the information need for the regulatory purpose under consideration (see Weinberg, 1971 for a detailed discussion). Note this interpretation step is not specific to IATA, but as the complexity of IATA increases, more formalized systematic and transparent translation approaches will be required. Integration of many information sources can be addressed in different ways from:

1. Battery approaches, i.e. all results are collected and then interpreted.
2. Sequential/tiered approach, i.e. in a given sequence results are collected stopping when sufficient information is available through to
3. Result-driven further testing, e.g. determination of next most valuable test or branching of test strategies depending on previous test results (prioritization).

Integration of results derived from these information sources in turn occurs on different levels, from the raw data level to the summary (categorical) level where certain information is lost. Examples of data integration approaches include:

1. Boolean AND/OR/NOT combinations of categorized results (e.g. overall call is denoted as positive if any of the test outcomes are positive).
2. Scoring approaches (e.g. various tests contributing to an overall score).
3. Decision trees (typically sequential with branching).
4. Deterministic, i.e. a point of departure for assessments is derived (e.g. lowest active concentration) possibly combined with assessment factors to derive a threshold value.

Table 2
Summary and examples of the different types of experimental testing approaches in AOP-informed IATAs.

Approach	Usage	AOP target	Example(s) of HT/HC compatible assays	Adverse outcome
In chemico	Indicate reactivity or covalent interaction with a biomolecule	MIE	GSH (Schultz et al., 2005); dNTP adduct formation (Zhao et al. 2002)	Unspecific (excess toxicity), genotoxicity/mutagenicity, immunosuppression, skin sensitization
In vitro (cellular)	Confirm toxicity pathway Confirm the (absence of) need for higher-tier testing Can be HT/HC compatible	MIE, KE	Cell lines, transactivation and reporter cell assays, subcellular assays, e.g. HTS assays for endocrine disruption (Cox et al., 2014; Murk et al., 2013)	Through targeting specific toxicity pathways, a wide range of AOs can be targeted (Bhattacharya et al. 2011) e.g. for endocrine disruption sexual development, reproductive disorders. Many different endpoints are possible through targeting specific toxicity pathways
Invertebrates	Replace (vertebrate) animal tests	MIE, KE	*C. elegans* (Leung et al. 2008)	e.g. Acute toxicity, developmental toxicity, neurotoxicity, genotoxicity,
Fish or amphibian embryos	Replace (adult vertebrate) animal tests	MIE, KE	*D. rerio* embryos (Truong et al., 2014)	Acute and chronic fish toxicity, hepatoxicity neurotoxicity, teratogenicity, endocrine disruption (reviewed in Scholz, 2013; Scholz et al., 2013)

Notes: GSH = reduced glutathione, dNTP = deoxyribonucleotide triphosphate, HT/HC = high-throughput/high-content.

5. Probabilistic, i.e. probabilities are assigned as a function of different information leading typically to distributions of probabilities/uncertainties.
6. Prediction based on machine learning (e.g. PCA, random forest, multiple regression) applied to a training set of compounds.

IATA does extend beyond hazard information and will often also include kinetics and exposure data, which in turn augments the complexity of the data integration approaches applied. At this stage, no general guidance can be proposed, although it is envisaged that a learning-by-doing is necessary and the advantages (and possible disadvantages) of formally integrated data will emerge and can be resolved.

4. Applicability of AOP-informed IATA for regulatory purposes

Any non-standard approach needs to be fit for purpose whether it will be used for prioritization, hazard identification, classification and labelling and/or risk assessment. This is true for IATA as a whole, as well as the respective IATA elements themselves; the latter of which have already been discussed in the previous sections. Specific criteria to define fitness for specific regulatory applications have not been defined but guiding principles are being proposed. Becker et al. (2014) outlined a scientific confidence framework first proposed for HT/HC screening assays (Patlewicz et al., 2013) and their prediction models (Cox et al., 2014) but adapted it to help in the evaluation of AOPs for different purposes including IATA. Specific guidance for the assessment of IATA is not currently available but recent initiatives taken up by the OECD Task Force for Hazard Assessment (TFHA) are aiming to develop general principles for the evaluation and documentation of IATA using skin sensitization as an initial case study (Worth and Patlewicz, submitted). The initial principles proposed are framed by a clear identification of the regulatory requirement as well as the applicability domain of the IATA itself:

(a) Define the endpoint of regulatory concern being assessed.
(b) Define the purpose/application for which the IATA is proposed.
(c) Describe the rationale, including mechanistic basis (e.g. AOP), according to which the IATA is constructed.
(d) Describe the individual information sources constituting the IATA.
(e) Characterize the predictive performance and applicability domain of the IATA, or IATA subcomponent(s) that can be expressed as a prediction model(s).

5. Examples of AOP-informed IATAs in regulatory decision-making

There are many potential regulatory applications for IATA. In this section, we highlight three case study examples, which target different regulatory scenarios and hence are characterized by differing levels of scientific confidence.

5.1. Identification of chemicals disrupting estrogen, androgen, and thyroid hormone pathways

Endocrine disruption, particularly disruption of estrogen, androgen and thyroid pathways, is considered as an endpoint of high regulatory concern, given the potential adverse impact on human and environmental health, particularly sexual differentiation, reproduction and population development. AOPs linked to endocrine disruption of these three hormonal pathways represent examples where links between the MIE and KEs and the final AO have been reasonably established (Ankley et al., 2005; Miller et al., 2009; Volz et al., 2011). The OECD has already provided a conceptual framework describing the assays that would be available to target the different MIE and KE for endocrine disruption (OECD, 2012c). Recent suggestions for developing scientific criteria for identification of an endocrine disrupting chemical (EDC) also conform to the principle of providing evidence of causality between mechanistic information (e.g. KEs) and AOs for endocrine disruption (Munn and Goumenou, 2013). Principally there is no single AOP for endocrine disruption. Depending on the targeted hormonal pathway or whether it is applied in the environmental or human health context, multiple AOPs could be defined. However, they share great commonalities at the different levels of biological complexity and are therefore described here (see Table 3).

In the US, the Endocrine Disruption Screening Program (EDSP) was established in an effort to identify substances with the potential to interact with components of the endocrine system. The Program comprises two Tiers; Tier 1 consists of a battery of *in vitro* and *in vivo* assays that are intended to determine the potential of a chemical to interact with the estrogen (E), androgen (A), or thyroid (T) hormone pathways whereas Tier 2 comprises multigenerational reproductive and developmental toxicity tests in several species to determine whether a substance can cause adverse effects resulting from effects on the E, A, or T pathways. In Tier 2, the tests to be run are selected by a WoE evaluation of Tier 1 results. The Tier 1 battery itself is expensive, time consuming and does not lend itself to the assessment of large numbers of chemicals (each Tier 1 costs of the order of 1 million US dollars). Furthermore, it still relies to a large extent on *in vivo* assays. Hence, more cost-efficient processes relying on *in silico* (QSAR and Expert Systems) and HT screening data for prioritizing large numbers of chemicals for hazard assessment purposes are being developed (Fig. S1, supplementary information outlines the use of the framework for prioritizing substances for their potential E, A and T effects). When coupled with exposure predictions (from ADME and exposure models), such a combination of non-testing and resource-efficient testing approaches could provide sufficient confidence in prioritization decisions for subsequent testing requirements. The EDSP represents an opportunity where relevant HTS assays can be mapped to associated AOPs that are already well understood (e.g. Schmieder et al., 2003; Crofton and Zoeller, 2005) and where confidence in the HTS predictive power for higher levels of the AOP are well established.

While the goal of the US EDSP and application of corresponding HTS assays is clearly one of prioritization and directing of testing, there is also scope to apply a tiered approach for defined testing schemes such as required by European Union regulations. Substances with endocrine disrupting capacity are conditionally exempted from exposure criteria, i.e. higher tier assays for these compounds are required also at lower production volumes. Similarly, Tier 1 *in vivo* assays to analyze the endocrine disruption potential are required in environmental hazard assessment for the regulation of pesticides, biocides and pharmaceuticals (Scholz, 2013; Scholz et al., 2013). It is however, not yet clear how the endocrine disrupting potency will be identified but HT assays may provide a cost-effective and reliable approach.

HT screening assays for determination of interference with hormone production, hormone receptor binding and activation are currently available for a number of hormone pathways. Of these, assays to target interference with the estrogen pathway seems to be best developed with HTS methods for steroidogenesis and aromatase inhibition (Villeneuve et al., 2007; Vinggaard et al., 2000), estrogen receptor (ER) binding and activation (Legler et al., 1999; Tollefsen and Nilsen, 2008; Tollefsen et al., 2008) and *in silico* (QSARs and docking models) for interaction with the ER

Table 3
Examples of MIEs and KEs relevant for different levels of the AOPs for endocrine disruption. Given the large number of assays available for the different MIEs, KEs, and AOs, only selected examples are presented. For further assays descriptions refer to OECD (2012c).

AOP level (MIE and KE not in sequential order)	Description (examples)	Level of biological organization	Test/non test method examples
MIE1	Hormone receptor binding and activation	Molecular level	Receptor-ligand binding assays (Tollefsen and Nilsen, 2008); Transactivation assays (Legler et al., 1999); QSARs for hormone receptor binding (Lo Piparo and Worth, 2010; Novic and Vracko, 2010)
MIE2	Interference with hormone synthesis	Molecular level	Steroidogenesis *in vitro* (OECD TG 456); *In vitro* assays for induction and inhibition of enzymes for TH metabolism (Murk et al., 2013); Zebrafish embryo assay for disruption of thyroid gland function (Raldua and Babin, 2009)
KE 1	Cell proliferation	Cellular	MCF7 cell proliferation assay (Körner et al., 1998)
KE 2	Increased vitellogenin production	Cellular	*In vitro* fish hepatocyte vitellogenin production (Tollefsen et al., 2008)
KE3	Proliferation of uterus Metamorphosis	Organ	Uterotrophic assay (OECD TG 440) Amphibian metamorphosis assay (OECD TG 231)
KE4	Vitellogenin induction, 2nd sex characteristics, fecundity, gonad development	Organ Organism	Fish Reproductive Screening Assay (OECD TG 229)
AO	Reproduction	Population	Reproductive toxicity studies Fish full life cycle assays (TG 415, 443)

(Schmieder et al., 2003; Mombelli, 2012). A similar suite of assays exists for androgen signaling pathways, although the role of androgen agonists or antagonists in endocrine disruption is not as well developed. Nevertheless, assays such as a transcriptional activation assay for the detection of the androgenic and anti-androgenic activity of chemicals have been developed to support the assessment of disruption of the androgen axis (Rostkowski et al., 2011). HT assays for detecting thyroid receptor agonists and antagonists also exist (Murk et al., 2013), however, the majority of thyroid disruptors act via a variety of MIEs that alter cellular TH signaling pathways via modulation of the TH levels. Thus, for thyroid disrupting compounds the most relevant KE with respect to AO is the reduction of thyroid hormone synthesis and homeostasis (Capen, 1997; Crofton, 2008). Appropriate thyroid hormone-relevant assays are missing for many of the targets, and development of appropriate assays that cover relevant MIEs and KEs are strongly needed (Murk et al., 2013). As an interim approach, TR transcription assays such as ToxCast and Tox21-TR assays can be applied. Assays of fish embryos targeting reduced T4-levels (Thienpont et al., 2011; Opitz et al., 2012) can be employed to identify goitrogens. Despite the remaining high uncertainty for thyroid hormone disruption, a significant reduction of higher tier testing could be achieved by including exposure modelling into the screening approach. Wambaugh et al. (2013) have developed a high-throughput exposure model that uses data on production and use of chemicals, in combination with a Bayesian statistical approach to describe the degree of uncertainty, to provide exposure estimates for thousands of chemicals. Combining this with hazard data allows for a rapid estimate of margins of exposure and prioritization of further testing using both exposure and hazard data. Whilst the IATA framework has been illustrated for prioritization per se, it could be refined for other purposes such as classification and labelling, or hazard assessment both of which would be pertinent for registration of chemicals in Europe.

5.2. Skin sensitization

Skin sensitization has been well studied over many decades. The chemical and biological pathway driving the induction and elicitation of allergic contact dermatitis is relatively well understood (see Lepoittevin et al., 1997; Smith Pease, 2003; Adler et al., 2011) and this knowledge has helped shape the development of alternative non-animal test methods. Most recently the knowledge has been structured and documented in an AOP construct and published by the OECD (2012a,b). The OECD documentation for this AOP summarizes the scientific evidence and assesses the overall WoE supporting the AOP. There is strong evidence for the qualitative sequence of events from the MIE to AO. Indeed empirical evidence from various elements of the AOP has value in assessing the *potential* of a chemical to be a skin sensitizer but, with few exceptions, it is insufficient to predict the relative *potency* of a chemical. As such, animal methods, in particular the Local Lymph Node Assay (LLNA) are at present still needed to provide a quantitative measure of relative sensitizing potency, which is critical for risk assessment applications.

In order for the AOP for skin sensitization to be applied in practice, available test/non test approaches that characterize each of the KEs need to be mapped to the AOP. This mapping provides a perspective of what practical testing/non testing strategies could be derived as IATA. For skin sensitization, there has been considerable progress in developing specific test methods that target MIEs and many of the KEs relevant for the AO (see Table 4 for examples of appropriate assays).

A specific framework for the assessment of skin sensitization potential was adapted from Fig. 1 (shown in Fig. S2 of the supplementary information). In applying the framework, two outcomes can be envisaged – either the evaluation of the model/assay outcomes will result in a consistent profile enabling an assessment of skin sensitization hazard to be made (i.e. the substance is (not) a skin sensitizer with high confidence) or the outcomes are insufficient to conclude with any great certainty that the substance is (not) a skin sensitizer. The latter could be due to inadequacies in the model/assay domains of applicability either on the basis of the underlying training sets or due to technical limitations in the assays themselves (volatility, solubility, metabolic competence). These insufficiencies however inform the development or refinement of new test assays or refinement/extension of the *in silico* models. Any new information then generated can be passed back to refine and improve the original AOP for sensitization. A more detailed example for this IATA for skin sensitization has been discussed in a separate manuscript (see Patlewicz et al., 2014).

5.3. AChE inhibition leading to lethality

Organophosphate and carbamate insecticides, which are widely used for agricultural and residential purposes, have frequently

Table 4
A summary of *in silico* and experimental testing approaches targeting MIEs and KEs of skin sensitization.

AOP level	Description	Level of biological organization	Test/non test method
Dermal exposure	Dermal metabolism, epidermal disposition	Chemical structure and properties	(Q)SARs
MIE	Covalent binding between electrophile and skin protein	Molecular level	DRPA (Gerberick et al., 2004, 2007), GSH depletion assay (Schultz et al., 2005), QSARs/read-across
KE 1	Activation of inflammatory cytokines	Cellular response	KeratinoSens™ (Emter et al., 2010, 2013), read-across
KE 2	Maturation and mobilization of dendritic cells	Cellular response	MUSST (Python et al., 2007), h-CLAT (Sakaguchi et al., 2007), read-across
KE 3	T-cell proliferation	Organ response	LLNA (OECD Test Guideline (TG) 429), QMM, read-across
Adverse outcome (AO)	Allergic contact dermatitis	Organism response	GPMT (OECD TG 406); HRIPT

Notes: DRPA = Direct peptide reactivity assay, GSH = Reduced glutathione, MUSST = Myeloid U937 *skin sensitization* test, h-CLAT = Human cell line activation test, LLNA = - Mouse local lymph node assay, QMM = Quantitative mechanistic model, GPMT = Guinea pig maximization test, HRIPT = Human repeat insult patch test.

been reported to cause toxicity to organisms ranging from invertebrates to vertebrates and mammals (McHenery et al., 1997; Fulton and Key, 2001). The toxicity of these compounds is mainly due to the selective inhibition of acetylcholinesterase (AChE), leading to accumulation of acetylcholine (ACh) in the synaptic cleft, subsequent overstimulation, and the disruption of nerve impulses ultimately leading to ataxia, central respiratory paralysis, seizures, coma and death (Costa, 2006; Bradbury et al., 2008). The well-developed knowledge on how these chemicals cause lethality has led to the development of an AOP for acetylcholinesterase inhibition leading to acute mortality (Russom et al., 2014). This AOP is characterized by a clear mechanistic understanding of the MIE, KEs and AOs (Table 5) for a number of species (Russom et al., 2014). The available information on relevant chemical structures, the overall weight of evidence and the broad taxonomic applicability domain of this AOP are of particular value to inform and provide input to IATAs, particularly for cross-species extrapolations.

Since Acetylcholine esterase (AchE) inhibition is a well-established AOP, it can support a variety of regulatory uses. The WoE supporting this AOP is strong (Russom et al., 2014), and there is extensive toxicity data for a number of chemicals in a variety of species that is consistent with mechanistic knowledge assembled in the AOP (http://www.epa.gov/ecotox/). Information from *in vitro* results could potentially be used under certain circumstances, but the use of *in vitro* AChE inhibition alone may not be sufficient (Knudsen et al., 2011) possibly due to lack of these assays accounting for bioactivation of certain chemicals such as Diazinon by metabolism (Aylward and Hays, 2011) or mitigation of effects by metabolic degradation such as observed for malathion (de Bruijn and Hermens, 1993). *In silico* approaches might be sufficient for some uses (Fukuto, 1990; El Yazal et al., 2001; Wong et al., 2012), but should be used with caution particularly in cases where metabolic activation is required (de Bruijn and Hermens, 1993). Extensive *in vivo* data exist with reasonable concordance seen between sequence similarity among AChE enzymes and *in vivo* activity across non-vertebrate species (Russom et al., 2014). For animals, including humans, determination of AchE inhibition in both the central and peripheral nervous systems are considered crucial for a thorough evaluation of potential hazard (http://www.epa.gov/pesticides/trac/science/cholin.pdf). However, blood cholinesterase inhibition is accepted as a surrogate parameter in humans, when data for AchE inhibition in peripheral and central nervous system are not available. Recommendations on surrogate parameters in wildlife have currently not been developed sufficiently to support a WoE approach to identify potential hazard.

To illustrate how this AOP could be used in IATA, consider the classification of a pesticide known to act via AChE inhibition as a potential application. A particular concern in this case is the biological impact on non-target organisms (see Fig. S3 in supplementary information). If this is a crop use that is expected to result in minimal exposure through either application or ingestion, the species of concern might be restricted to non-target organisms that would be exposed during the application or via interactions with the treated crops and possibly aquatic organisms from runoff following application. Demonstration of low level of exposure in combination with low sensitivity for AchE in vertebrates, would be expected to limit potential hazards to non-target invertebrates. If toxicity data from the target species (e.g. insects for use of insecticides) exist, hazard assessment could be facilitated by sequence alignments to predict cross-species susceptibility to non-target species where exposure is considered relevant (Lalone et al., 2013; Russom et al., 2014). Documentation of potential risk scenarios (e.g. small margin of safety between exposure and potential effects) based on the non-testing approaches proposed herein, may lead to a decision to generate additional testing data using *in vivo* studies with the appropriate species or relevant surrogate species in cases where testing is not feasible (e.g. endangered species, lack of appropriate laboratory strains etc.).

This hypothetical case study illustrates how a well-defined AOP could be used for certain regulatory purposes independent of chemical specific information at the intermediate key events. The weight of evidence incorporates over 50 years of research including basic biochemistry as well as toxicology. Given the strong support and conservation of the AOP across taxa, a wealth of toxicological data at the organism level can be leveraged for the decision at hand. This allows the use of *in silico* predictions for cross-species extrapolations in combination with use of data from experimentally tractable species to limit the need for additional studies to characterize intermediate events of well-developed AOPs. If this were not the case, other approaches such as *in vitro* screening and *in vivo* measurement of intermediate KE (Fig. S3 in supplementary information) would likely be required to safeguard against adversely impacting non-target species.

6. Implications for Integrated Toxicology and Regulatory Applications

Development and application of AOP-informed IATA represents a new way to evaluate and generate information to meet different regulatory purposes. A conceptual framework for applying IATA has been proposed that considers existing information (from a hazard and exposure perspective) in the context of an AOP to make an informed decision based on the regulatory context. Frameworks to characterize the scientific confidence of an AOP that are required to meet different regulatory needs are in development (Becker et al., 2014; Patlewicz et al., submitted; Perkins et al., submitted). These will shape the structure of the IATA and its elements in terms of the test methods and non-testing approaches. Establishing scientific confidence is critical for both the elements making

Table 5
A summary of *in silico* and experimental testing approaches relevant for different levels of the AOP – Acetylcholine esterase (AchE) inhibition leading to lethality (Russom et al., 2014). References represent examples only. See Russom et al. (2014) for a more extensive review of the literature supporting this AOP.

AOP level	Description	Level of biological organization	Test/non test method
MIE	Inhibition of AchE activity. Inhibition caused by non-reversible or reversible inhibition	Molecular level	QSARs/read-across Inhibition of AchE activity (*in vitro*) (Garcia-Reyero et al., 2014b; Holth and Tollefsen, 2012)
KE1	Accumulation of acetylcholine (ACh) in the synaptic cleft	Cellular level	No direct test-method available; biological plausibility well established; many studies linking MIE with downstream KEs & AOs across a variety of species (Bianco et al., 2013); Brain ACh levels can serve as a surrogate biomarker for associated KEs (Kobayashi et al., 1985)
KE2	Excitatory responses in muscle and brain	Organ level	Electrophysiology in isolated neurons (Oyama et al., 1989); Contractile response in muscle (Kobayash et al., 1994); Altered response in brain (biological plausibility well established)
AO	Neurotoxic symptomology (increased respiration, bradycardia, seizures) leading to death	Organism	Respiratory/cardiovascular responses (McKim et al., 1987); Altered photomotor or locomotor response (Kokel et al. 2010; Irons et al. 2010; Garcia-Reyero et al. 2014b);
AO	Population decline	Population	Inferred based on measured effects on mortality (Barata et al., 2004) and feeding behavior (Hunt et al., 1991)

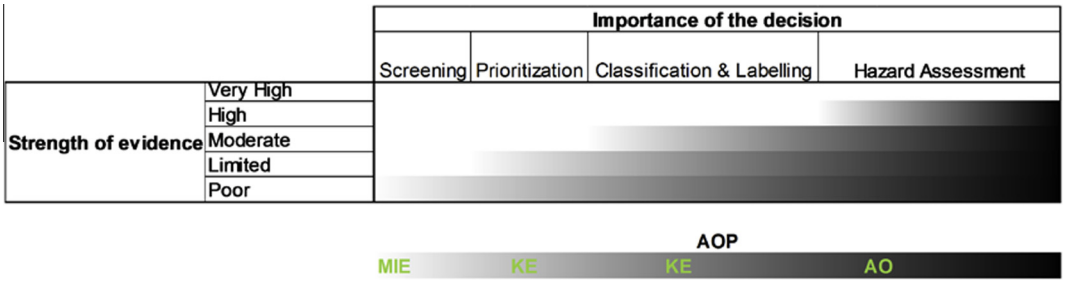

Fig. 2. Relationship between strength of evidence (reliability, relevance and concordance etc.) for the IATA supporting the AO of regulatory concern and the importance of the regulatory decision to be made. The figure shows the possible combination of AOP based information and available data, and how the use of one could be strengthened by the other. The color of the cells represent the amount of additional information from other sources needed for a decision (the darker the color, the more additional information is needed to reach a decision with confidence). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

up the IATA as well as the IATA as a whole. Three case studies have been described in detail to illustrate how the conceptual framework proposed in Section 1 can be adapted to meet different regulatory purposes (e.g. prioritization, hazard assessment, classification and labelling and other applications such as cross-species extrapolations).

The EDC example shows how a battery of AOPs and associated HT assays can be used in a prioritization scenario. This addresses the first 4 principles for IATA development and application as outlined in Section 4. The skin sensitization example, which is aimed at addressing hazard assessment, arguably addresses all 5 principles. The AchE inhibition example illustrates how an established AOP can be used for classification and labelling in certain regulatory contexts despite a lack of properly developed testing and non-testing methods spanning the full AOP continuum. In the regulatory context considered for that example, the need for explicit tests of intermediate KEs is avoided by the wealth of data available. A well-developed AOP and by demonstration of phylogenetically-conserved MIEs across taxa enable identification of susceptible species being particularly relevant or tractable to cost-efficient *in vivo* testing (e.g. invertebrates). Clearly the degree to which these principles need to be characterized can and will differ based on the level of uncertainty that can be tolerated for the regulatory purpose under consideration. Scientific confidence of the AOP and its associated IATA will be strongest where there is a close link between the MIE and KEs to the AO.

There is a desire to exploit *in silico* and HTS testing tools to populate an IATA. One starting point for such AOP-informed IATAs could be to apply i*n silico* methods or HT approaches for providing information about the MIE to determine what data if any would need to be generated for different KEs or the AO for a given chemical. The stronger the evidence coming from non-testing or alternative testing approaches, the less additional information would, in theory, need to be generated for a given decision. Thus, a moderate level of confidence might be sufficient for a prioritization purpose, but in order to make a decision related to hazard and risk assessment, assays or a combination of assays closely linked to the MIE and with high predictivity of the AO may be required (Fig. 2.). Additional information on one or more KEs along the pathway generated from *in vitro*, *in chemico* or HT/HC assays would serve to provide increasing confidence for a given decision.

Of course implicit in such a confidence determination, is the WoE evaluation of the AOP itself. The OECD AOP guidance entails completing a template, using evolved and tailored Bradford Hill (BH) considerations, in which each KEs and KE relationships (KERs) in an AOP are evaluated and are scored as high, moderate and low (OECD, 2013). The outcomes of these WoE determinations for the KEs and KERs of an AOP help in making a determination of whether different decisions are feasible based on the outcomes of MIEs or other KEs and the extent to which they are predictive of the AO.

The case studies presented could in theory be applied in practice now, although the number of well-developed AOPs is currently limiting the practical applicability for larger scale regulatory deployment. Furthermore, consideration needs to be given to the analytical validation of testing and non-testing approaches in order to better characterize their applicability domain i.e. the types of chemicals that can be reliably assessed. A detailed description of AOPs of regulatory relevance and the establishment of qualitative and quantitative links between MIEs, KEs and AOs will additionally help foster application for different regulatory decisions. While

qualitative links are already established for a number of the AOPs so far developed and supported by visualization and description tools such as the AOP Knowledge Base (https://aopkb.org), appropriate quantitative approaches for confidence evaluation by WoE assessments of KERs are currently being critically assessed (Barton-Maclaren et al., in preparation). Recent initiatives to provide quantitative assessment of the role of MIE and KE proximity to the AO for the confidence of predictions to regulatory-relevant endpoints will likely also assist in developing pragmatic tools for IATA development. Additional improvements of IATAs by including toxicokinetics and reverse dosimetry into extrapolations to regulatory-relevant endpoints would further increase the applicability of IATAs for practical use.

Although not necessarily applicable to the case studies highlighted here, many of the AOPs in development have been data-rich and based on historical *in vivo* data. Thus the body of evidence to justify the essentiality of KEs and the linkages has facilitated different use scenarios including risk assessment where the KEs proximal to the AO are better defined. Going forward, the challenges foreseen will be to identify the data gaps and assay needs, to integrate different AOPs together to provide a more holistic assessment of likely effects. The latter is a major issue as an AOP by its nature assumes that adversity can be described by a relevant assembly of MIEs and KEs, whereas the question remains of how many AOPs need to be integrated into IATA to assure that there is no important hazard or adversity overlooked.

Conflict of interest

None declared.

Acknowledgments

The content of this paper is the result of an international expert workshop on *Advancing Adverse Outcome Pathways (AOPs) for Integrated Toxicology and Regulatory Applications*. Support for the workshop was provided by the American Chemistry Council (ACC), BioDetection Systems, European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), Environment Canada, European Commission Directorate General Joint Research Centre, Human Toxicology Project Consortium, International Life Sciences Institute – Health and Environmental Science Institute, The Research Council of Norway (Grant No. 221455), Society for Environmental Toxicology and Chemistry (SETAC), US Army Engineer Research and Development Center, and the U.S. Environmental Protection Agency. The authors acknowledge the workshop organizing committee, and all workshop participants for both plenary and informal discussions that influenced the content presented here. The views expressed are those of the authors, and do not necessarily represent the views of the organizations the authors are affiliated with or the sponsors. Mention of trade names does not imply endorsement by any of the author organizations or sponsors.

The participation of Tatiana Netzeva and Alberto Martin Aparicio from the European Chemicals Agency (ECHA) in the discussions of the workshop is also acknowledged. The views expressed at the workshop represented the personal opinion of the experts as individual scientists and regulatory professionals and do not constitute an official position of ECHA.

The opinions expressed and arguments employed herein are those of the authors and do not necessarily reflect the official views of the OECD or the governments of its member countries.

The information in this document has been funded wholly (or in part) by the U.S. Environmental Protection Agency. It has been subjected to review by the National Health and Environmental Effects Research Laboratory (NHEERL) and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.yrtph.2014.09.009.

References

Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V., Andersen, K.E., Angers-Loustau, A., Aptula, A., Bal-Price, A., Benfenati, E., Bernauer, U., Bessems, J., Bois, F.Y., Boobis, A., Brandon, E., Bremer, S., Broschard, T., Casati, S., Coecke, S., Corvi, R., Cronin, M., Daston, G., Dekant, W., Felter, S., Grignard, E., Gundert-Remy, U., Heinonen, T., Kimber, I., Kleinjans, J., Komulainen, H., Kreiling, R., Kreysa, J., Leite, S.B., Loizou, G., Maxwell, G., Mazzatorta, P., Munn, S., Pfuhler, S., Phrakonkham, P., Piersma, A., Poth, A., Prieto, P., Repetto, G., Rogiers, V., Schoeters, G., Schwarz, M., Serafimova, R., Tähti, H., Testai, E., van Delft, J., van Loveren, H., Vinken, M., Worth, A., Zaldivar, J.M., 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Arch. Toxicol. 85, 367–485.

Ahlers, J., Stock, F., Werschkun, B., 2008. Integrated testing and intelligent assessment—new challenges under REACH. Environ. Sci. Pollut. Res. 15, 565–572.

Ankley, G.T., Jensen, K.M., Durhan, E.J., Makynen, E.A., Butterworth, B.C., Kahl, M.D., Villeneuve, D.L., Linnum, A., Gray, L.E., Cardon, M., Wilson, V.S., 2005. Effects of two fungicides with multiple modes of action on reproductive endocrine function in the fathead minnow (*Pimephales promelas*). Toxicol. Sci. 86, 300–308.

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrrano, J.A., Tietge, J.E., Villeneuve, D.L., 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29, 730–741.

Aptula, A.O., Roberts, D.W., 2006. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. Chem. Res. Toxicol. 19, 1097–1105.

Aylward, L.L., Hays, S.M., 2011. Consideration of dosimetry in evaluation of ToxCast data. J. Appl. Toxicol. 31, 741–751.

Balls, M., Amcoff, P., Bremer, S., Casati, S., Coecke, S., Clothier, R., Combes, R., Corvi, R., Curren, R., Eskes, C., Fentem, J., Gribaldo, L., Halder, M., Hartung, T., Hoffman, S., Schechtman, L., Scott, L., Spielmann, H., Stokes, W., Tice, R., Wagner, D., Zuang, Z., 2006. The principles of weight of evidence validation of test methods and testing strategies: the report and recommendations of ECVAM workshop 58. Altern. Lab. Anim. 34, 603–620.

Barata, C., Solayan, A., Porte, C., 2004. Role of B-esterases in assessing toxicity of organophosphorus (chlorpyrifos, malathion) and carbamate (carbofuran) pesticides to Daphnia magna. Aquat. Toxicol. 66, 125–139.

Barton-Maclaren, T., Becker, R., van der Burg, B., Kennedy, S., Meek, M. E., Ankley, G., Linkov, I., Segner, H., Watanabe, K., Sachana, M., Villeneuve, D., Edwards, S. Weight of evidence evaluation to define uncertainties associated with predictive relationships represented in an AOP (Unpublished-in preparation).

Becker, R.A., Rowlands, J.C., Patlewicz, G., Simon, T., 2014. Enhancing the utility of T21 assessment methods by employing a scientific confidence framework and exposure: activity profiling. Presented at ICCA-JRC Workshop, June 17–18, 2014 Lugano, Switzerland.

Bhattacharya, S., Zhang, Q., Carmichael, P.L., Boekelheide, K., Andersen, M.E., 2011. Toxicity testing in the 21st century: defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. PLoS One 6, e20887.

Bianco, K., Yusseppone, M.S., Otero, S., Luquet, C., Rios de Molina, M.C., Kristoff, G., 2013. Cholinesterases and neurotoxicity as highly sensitive biomarkers for an organophosphate insecticide in freshwater gastropod (*Chilina gibbosa*) with low sensitivity carboxylesterases. Aquat. Toxicol. 26–35, 144–145. http://dx.doi.org/10.1016/j.aquatox.2013.09.025.

Blackburn, K., Stuard, S.B., 2014. A framework to facilitate consistent characterization of read across uncertainty. Reg. Toxicol. Pharmacol. 68, 353–362.

Bradbury, S.P., Carlson, R.W., Henry, T.R., Padilla, S., Cowden, J., 2008. Toxic responses of the fish nervous system. In: Di Giulio, R.T., Hinton, D.E. (Eds.), The Toxicology of Fishes, CRC Press, pp. 417–455.

Capen, C.C., 1997. Mechanistic data and risk assessment of selected toxic end points of the thyroid gland. Toxicol. Pathol. 25, 39–48.

Costa, L.G., 2006. Current issues in organophosphate toxicology. Clin. Chim. Acta 366, 1–13.

Cox, L.A., Douglas, D., Marty, S., Rowlands, J.C., Patlewicz, G., Goyak, K.O., Becker, R.A., 2014. Applying a scientific confidence framework to a HTS-derived prediction model for endocrine endpoints: lessons learned from a case study. Reg. Toxicol. Pharmacol. 69, 443–450.

Crofton, K.M., 2008. Thyroid disrupting chemicals: mechanisms and mixtures. Int. J. Androl. 31, 209–223.

Crofton, K.M., Zoeller, R.T., 2005. Mode of action: neurotoxicity induced by thyroid hormone disruption during development–hearing loss resulting from exposure to PHAHs. Crit. Rev. Toxicol. 35, 757–769.

Cronin, M.T.D., Madden, J.C., 2010. In Silico Toxicology: Principles and Applications, Royal Society of Chemistry, Cambridge.

Cronin, M.T.D., Bajot, F., Enoch, S.J., Madden, J.C., Roberts, D.W., Schwöbel, J., 2009. The in chemico–in silico interface: challenges for integrating experimental and computational chemistry to identify toxicity. Altern. Lab. Anim. 37, 513–521.

de Bruijn, J., Hermens, J., 1993. Inhibition of acetylcholinesterase and acute toxicity of organophosphorous compounds to fish: a preliminary structure-activity analysis. Aquat. Toxicol. 24, 257–274.

EC, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off. J. Eur. Union. L396/1 of 30.12.2006. Commission of the European Communities.

ECHA, 2008. Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. Available at: <http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf> (last accessed 16.09.14).

El Yazal, J., Rao, S.N., Mehl, A., Slikker Jr., W., 2001. Prediction of organophosphorus acetylcholinesterase inhibition using three-dimensional quantitative structure-activity relationship (3D-QSAR) methods. Toxicol. Sci. 63, 223–232.

Emter, R., Ellis, G., Natsch, A., 2010. Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. Toxicol. Appl. Pharmacol. 245, 281–290.

Emter, R., van der Veen, J.W., Adamson, G., Ezendam, J., van Loveren, H., Natsch, A., 2013. Gene expression changes induced by skin sensitizers in the KeratinoSens™ cell line: discriminating Nrf2-dependent and Nrf2-independent events. Toxicol. In Vitro 27, 2225–2232.

Escher, B., Schwarzenbach, P., 2002. Mechanistic studies on baseline toxicity and uncoupling of organic compounds as a basis for modeling effective membrane concentrations in aquatic organisms. Aquat. Sci. 64, 20–35.

Fukuto, T.R., 1990. Mechanism of action of organophosphorus and carbamate insecticides. Environ. Health Perspect. 87, 245–254.

Fulton, M.H., Key, P.B., 2001. Acetylcholinesterase inhibition in estuarine fish and invertebrates as an indicator of organophosphorus insecticide exposure and effects. Environ. Toxicol. Chem. 20, 37–45.

Garcia-Reyero, N., Kennedy, A.J., Escalon, B.L., Habib, T., Laird, J.G., Rawat, A., Wiseman, S., Hecker, M., Denslow, N., Steevens, J.A., Perkins, E.J., 2014a. Differential effects and potential adverse outcomes of ionic silver and silver nanoparticles in vivo and in vitro. Environ. Sci. Technol. 48, 4546–4555.

Garcia-Reyero, N., Escalon, B.L., Prats, E., Stanley, J.K., Thienpont, B., Melby, N.L., Barón, E., Eljarrat, E., Barceló, D., Mestres, J., Babin, P.J., Perkins, E.J., Raldúa, D., 2014b. Effects of BDE-209 contaminated sediments on zebrafish development and potential implications to human health. Environ. Int. 63, 216–223.

Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., Lepoittevin, J.-P., 2004. Development of a peptide reactivity assay for screening contact allergens. Toxicol. Sci. 81, 332–343.

Gerberick, G.F., Vassallo, J.D., Foertsch, L.M., Price, B.B., Chaney, J.G., Lepoittevin, J.-P., 2007. Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. Toxicol. Sci. 97, 417–427.

Hartung, T., Luechtefeld, T., Maertens, A., Kleensang, A., 2013a. Integrated testing strategies for safety assessments. ALTEX 30, 3–18.

Hartung, T., Hoffmann, S., Stephens, M., 2013b. Mechanistic validation. ALTEX 30, 119–130.

Hill, A.B., 1965. The environment and disease: association or causation? Proc. R. Soc. Med. 58, 295–300.

Holth, T.F., Tollefsen, K.E., 2012. Acetylcholine esterase inhibitors in effluents from oil production platforms in the North Sea. Aquat. Toxicol. 112–113, 92–98.

Hunt, K.A., Bird, D.M., Mineau, P., Shutt, L., 1991. Secondary poisoning hazard of fenthion to American kestrels. Arch. Environ. Contam. Toxicol. 21, 84–90.

Institute of Medicine (IOM), 2010. Evaluation of biomarkers and surrogate endpoints in chronic disease. ISBN: 978-0-309-15129-0.

Irons, T.D., MacPhail, R.C., Hunter, D.L., Padilla, S., 2010. Acute neuroactive drug exposures alter locomotor activity in larval zebrafish. Neurotoxicol. Teratol. 32, 84–90.

Judson, R., Kavlock, R.J., Martin, M.T., Reif, D., Houck, K.A., Knudsen, T.B., Richard, A.M., Tice, R.R., Whelan, M., Xia, M., Huang, R., Austin, C.P., Daston, G.P., Hartung, T., Fowle III, J., Wooge, W., Tong, W., Dix, D.J., 2013. Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. ALTEX 30, 51–66.

Judson, R., Houck, K., Martin, M., Knudsen, T., Thomas, R.S., Sipes, N., Shah, I., Wambaugh, J., Crofton, K., 2014. In vitro and modelling approaches to risk assessment from the U.S. environmental protection agency ToxCast program. Basic Clin. Pharmacol. Toxicol. 115, 69–76.

Kavlock, R.J., Dix, D.J., Houck, K.A., Judson, R.S., Martin, M.T., Richard, A.M., 2007. ToxCast™: developing predictive signatures for chemical toxicity. ALTEX 14, 623–627.

Kleinstreuer, N.C., Yang, J., Berg, E.L., Knudsen, T.B., Richard, A.M., Martin, M.T., Reif, D.M., Judson, R.S., Polokoff, M., Dix, D.J., Kavlock, R.J., Houck, K.A., 2014. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. Nat. Biotechnol. 32, 583–591.

Knudsen, T.B., Houck, K.A., Sipes, N.S., Singh, A.V., Judson, R.S., Martin, M.T., Weissman, A., Kleinstreuer, N.C., Mortensen, H.M., Reif, D.M., Rabinowitz, J.R., Woodrow Setzer, R., Richard, A.M., Dix, D.J., Kavlock, R.J., 2011. Activity profiles of 309 ToxCast™ chemicals evaluated across 292 biochemical targets. Toxicology 282, 1–15.

Kobayashi, H., Yuyama, A., Kajita, T., Shimura, K., Ohkawa, T., Satoh, K., 1985. Effects of insecticidal carbamates on brain acetylcholine content, acetylcholinesterase activity and behavior in mice. Toxicol. Lett. 29, 153–159.

Kobayashi, H., Sato, I., Akatsu, Y., Fujii, S.I., Suzuki, T., Matsusaka, N., Yuyama, A., 1994. Effects of single or repeated administration of a carbamate, propoxur, and an organophosphate, DDVP, on jejunal cholinergic activities and contractile responses in rats. J. Appl. Toxicol. 14, 185–190.

Kokel, D., Bryan, J., Laggner, C., White, R., Cheung, C.Y., Mateus, R., Healey, D., Kim, S., Werdich, A.A., Haggarty, S.J., Macrae, C., Shoichet, B., Peterson, R., 2010. Rapid behavior-based identification of neuroactive small molecules in the zebrafish. Nat. Chem. Biol. 6, 231–237.

Körner, W., Hanf, V., Schuller, W., Bartsch, H., Zwirner, M., Hagenmaier, H., 1998. Validation and application of a rapid in vitro assay for assessing the estrogenic potency of halogenated phenolic chemicals. Chemosphere 37, 2395–2407.

Lalone, C.A., Villeneuve, D.L., Burgoon, L.D., Russom, C.L., Helgen, H.W., Berninger, J.P., Tietge, J.E., Severson, M.N., Cavallin, J.E., Ankley, G.T., 2013. Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk of chemicals with known modes of action. Aquat. Toxicol. 144–145, 141–154.

Legler, J., van den Brink, C., Brouwer, A., Murk, A., van der Saag, P., Vethaak, A., van der Burg, B., 1999. Development of a stably transfected estrogen receptor-mediated luciferase reporter gene assay in the human T47D breast cancer cell line. Toxicol. Sci. 48, 55–66.

Lepoittevin, J.-P., Basketter, D.A., Goossens, A., Karlberg, A.-T., 1997. Allergic contact dermatitis. The molecular basis. Springer, Heidelberg.

Leung, M.C.K., Williams, P.L., Benedetto, A., Au, C., Helmcke, K.J., Aschner, M., Meyer, J.N., 2008. *Caenorhabditis elegans*: an emerging model in biomedical and environmental toxicology. Toxicol. Sci. 106, 5–28.

Lo Piparo, E., Worth, A., 2010. Review of QSAR models and software tools for predicting developmental and reproductive toxicity. JRC report EUR 24522 EN. Publications Office of the European Union.

McHenery, J.G., Linley-Adams, G.E., Moore, D.C., Rodger, G.K., Davies, I.M., 1997. Experimental and field studies of effects of dichlorvos exposure on acetylcholinesterase activity in the gills of the mussel, *Mytilus edulis* L. Aquat. Toxicol. 38, 125–143.

McKim, J.M., Schmieder, P.K., Niemi, G.J., Carlson, R.W., Henry, T.R., 1987. Use of respiratory-cardiovascular responses of rainbow trout (*Salmo gairdneri*) in identifying acute toxicity syndromes in fish: Part 2. Malathion, carbaryl, acrolein, and benzaldehyde. Environ. Toxicol. Chem. 6, 313–328.

Meek, M.E., Boobis, A., Cote, I., Dellarco, V., Fotakis, G., Munn, S., Seed, J., Vickers, C., 2014a. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J. Appl. Toxicol. 34, 1–18.

Meek, M.E., Palermo, C.M., Bachman, A.N., North, C.M., Lewis, R.J., 2014b. Mode of action human relevance (species concordance) framework: evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. J. Appl. Toxicol. 34, 595–606.

Miller, M.D., Crofton, K.M., Rice, D.C., Zoeller, R.T., 2009. Thyroid-disrupting chemicals: interpreting upstream biomarkers of adverse outcomes. Environ. Health Perspect. 2117, 1033–1041.

Mombelli, E., 2012. Evaluation of the OECD (Q)SAR application toolbox for the profiling of estrogen receptor binding affinities. SAR QSAR Environ. Res. 23, 37–57.

Munn, S., Goumenou, M., 2013. Key scientific issues relevant to the identification and characterisation of endocrine disrupting substances. Report of the Endocrine Disrupters Expert Advisory Group. JRC Scientific and Policy Report EUR 25919 EN. Publications Office of the European Union, Luxembourg. Available from: http://publications.jrc.ec.europa.eu/repository/.

Murk, A.J., Rijntjes, E., Blaauboer, B.J., Clewell, R., Crofton, K.M., Dingemans, M.M., Furlow, J.D., Kavlock, R., Kohrle, J., Opitz, R., Traas, T., Visser, T.J., Xia, M., Gutleb, A.C., 2013. Mechanism-based testing strategy using in vitro approaches for identification of thyroid hormone disrupting chemicals. Toxicol. In Vitro 27, 1320–1346.

Nelms, M.D, Ates, G., Madden, J.C, Vinken, M., Cronin, M.T, Rogiers, V., Enoch, S.J., 2014. Proposal of an in silico profiler for categorisation of repeat dose toxicity data of hair dyes. Arch. Toxicol. (in press). http://dx.doi.org/10.1007/s00204-014-1277-8.

Novic, M., Vracko, M., 2010. QSAR models for reproductive toxicity and endocrine disruption activity. Molecules 15, 1987–1999.

Organisation for Economic Cooperation and Development, (OECD), 2004. OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure–Activity Relationship Models. Available at: <http://www.oecd.org/env/ehs/risk-assessment/37849783.pdf>.

Organisation for Economic Cooperation and Development, (OECD). 2007. Guidance Document on the Validation of (Q)SAR Models. OECD Series on Testing and Assessment No. 69. Organisation for Economic Co-operation and Development, Paris, France.

Organisation for Economic Cooperation and Development, (OECD), 2008. Workshop on Integrated Approaches to Testing and Assessment. OECD Series on Testing and Assessment No. 88. Organisation for Economic Co-operation and Development, Paris, France.

Organisation for Economic Cooperation and Development, (OECD), 2012a The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins Part 1: Scientific Evidence. Series on Testing and Assessment No. 168 ENV/JM/MONO(2012) 10/PART1.

Organisation for Economic Cooperation and Development, (OECD), 2012b. The Adverse Outcome Pathway for Skin Sensitisation by Covalent Binding to Proteins. Part 2: Use of the AOP to Develop Chemical Categories and Integrated Testing and Assessment Approaches. Organisation for Economic Co-operation and Development Paris, France.

Organisation for Economic Cooperation and Development, (OECD), 2012c. OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters. Organisation for Economic Co-operation and Development Paris, France.

Organisation for Economic Cooperation and Development (OECD), 2013. Guidance Document on Developing and Assessing Adverse Outcome Pathways. Organisation for Economic Co-operation and Development Paris, France.

Organisation for Economic Cooperation and Development, (OECD), 2014a. New Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation. Series on Testing and Assessment No. 203. Organisation for Economic Co-operation and Development, Paris, France.

Organisation for Economic Cooperation and Development, (OECD), 2014b. Guidance on grouping of chemicals, second edition. Series on Testing and Assessment No. 194. Organisation for Economic Co-operation and Development, Paris, France.

Opitz, R., Maquet, E., Huisken, J., Antonica, F., Trubiroha, A., Pottier, G., Janssens, V., Costagliola, S., 2012. Transgenic zebrafish illuminate the dynamics of thyroid morphogenesis and its relationship to cardiovascular development. Dev. Biol. 372, 203–216.

Oyama, Y., Hori, N., Evans, M.L., Allen, C.N., Carpenter, D.O., 1989. Electrophysiological estimation of the actions of acetylcholinesterase inhibitors on acetylcholine receptor and cholinesterase in physically isolated Aplysia neurons. Br. J. Pharmacol. 96, 573–582.

Patlewicz, G., Simon, T., Goyak, K., Phillips, R.D., Rowlands, J.C., Seidel, S., Becker, R.A., 2013. Use and validation of HT/HC assays to support 21st century toxicity evaluations. Regul. Toxicol. Pharmacol. 65, 259–268.

Patlewicz, G., Kuseva, C., Kesova, A., Popova, I., Zhechev, T., Pavlov, T., Roberts, D.W., Mekenyan, O.M., 2014. Towards AOP application – implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. Reg. Toxicol. Pharmacol. 69, 529–545.

Patlewicz, G., Simon, T., Rowlands, J.C., Budinsky, R.A., Becker, R.A. Using a scientific confidence framework to support application of adverse outcome pathways for regulatory purposes. Reg. Toxicol. Pharmacol. (Unpublished – submitted).

Patlewicz, G., Ball, N., Blackburn, K., Aptula, A, Lampi, M., Boogaard, P., Becker, R.A., Hubesch, B. Building scientific confidence in the development and evaluation of read-across. (Unpublished – in preparation).

Perkins, E.J., Ankley, G.T., Crofton, K.M., Garcia-Reyero, N., LaLone, C.A., Johnson, M.S., Tietge, J.E., Villeneuve, D.L., 2013. Current perspectives on the use of alternative species in human health and ecological hazard assessments. Environ. Health Perspect. 121, 1002–1010.

Perkins, E.J., Antczak, P., Burgoon, L., Falciani, F., Gutsell, S., Hodges, G., Kienzler, A., Knapen, D., McBride, M., Willett, C., 2014. Using adverse outcome pathways for regulatory applications. Toxicol. Sci. (Unpublished – submitted).

Python, F., Goebel, C., Aeby, P., 2007. Assessment of the U937 cell line for the detection of contact allergens. Toxicol. Appl. Pharmacol. 220, 113–124.

Raldua, D., Babin, P.J., 2009. Simple, rapid zebrafish larva bioassay for assessing the potential of chemical pollutants and drugs to disrupt thyroid gland function. Environ. Sci. Technol. 43, 6844–6850.

Roberts, D.W., Aptula, A.O., Patlewicz, G., Pease, C., 2008. Chemical reactivity indices and mechanism-based read across for non-animal based assessment of skin sensitization potential. J. Appl. Toxicol. 28, 443–454.

Romanov, S., Medvedev, A., Gambarian, M., Poltoratskaya, N., Moeser, M., Medvedeva, L., Gambarian, M., Diatchenko, L., Makarov, S., 2008. Homogeneous reporter system enables quantitative functional assessment of multiple transcription factors. Nat. Methods 5, 253–260.

Rostkowski, P., Horwood, J., Shears, J.A., Lange, A., Oladapo, F.O., Besselink, H.T., Tyler, C.R., Hill, E.M., 2011. Bioassay-directed identification of novel antiandrogenic compounds in bile of fish exposed to wastewater effluents. Environ. Sci. Technol. 45, 10660–10667.

Russom, C.L., LaLone, C.A., Villeneuve, D.L., Ankley, G.T., 2014. Development of an adverse outcome pathway for acetylcholinesterase inhibition leading to acute mortality. Environ Toxicol Chem. 33, 2157–2169. http://dx.doi.org/10.1002/etc.2662.

Sakaguchi, H., Ashikaga, T., Kosaka, N., Sono, S., Nishiyama, N., Itagaki, H., 2007. The in vitro skin sensitization test; human cell line activation test (h-CLAT) using THP-1 cells. Toxicol. Lett. 172, S93.

Schmieder, P.K., Ankley, G., Mekenyan, O., Walker, J.D., Bradbury, S., 2003. Quantitative structure-activity relationship models for prediction of estrogen receptor binding affinity of structurally diverse chemicals. Environ. Toxicol. Chem. 22, 1844–1854.

Scholz, S., 2013. Zebrafish embryos as an alternative model for screening of drug-induced organ toxicity. Arch. Toxicol. 87, 767–769.

Scholz, S., Sela, E., Blaha, L., Braunbeck, T., Galay-Burgos, M., Garcia-Franco, M., Guinea, J., Kluver, N., Schirmer, K., Tanneberger, K., Tobor-Kaplon, M., Witters, H., Belanger, S., Benfenati, E., Creton, S., Cronin, M.T., Eggen, R.I., Embry, M., Ekman, D., Gourmelon, A., Halder, M., Hardy, B., Hartung, T., Hubesch, B., Jungmann, D., Lampi, M.A., Lee, L., Leonard, M., Kuster, E., Lillicrap, A., Luckenbach, T., Murk, A.J., Navas, J.M., Peijnenburg, W., Repetto, G., Salinas, E., Schuurmann, G., Spielmann, H., Tollefsen, K.E., Walter-Rohde, S., Whale, G., Wheeler, J.R., Winter, M.J., 2013. A European perspective on alternatives to animal testing for environmental hazard identification and risk assessment. Regul. Toxicol. Pharmacol. 67, 506–530.

Schultz, T.W., Yarbrough, J.W., Johnson, E.L., 2005. Structure-activity relationships for reactivity of carbonyl compounds with glutathione, SAR QSAR. Environ. Res. 16, 313–322.

Schwöbel, J.A., Koleva, Y.K., Enoch, S.J., Bajot, F., Hewitt, M., Madden, J.C., Roberts, D.W., Schultz, T.W., Cronin, M.T., 2011. Measurement and estimation of electrophilic reactivity for predictive toxicology. Chem. Rev. 111, 2562–2596.

Simmons, S.O., Fan, C.-Y., Ramabhadran, R., 2009. Cellular stress response pathway system as a sentinel ensemble in toxicological screening. Toxicol. Sci. 111, 202–225.

Smith Pease, C.K., 2003. From xenobiotic chemistry and metabolism towards better risk assessment in skin allergy. Toxicology 192, 1–22.

Thaens, D., Heinzelmann, D., Böhme, A., Paschke, A., Schüürmann, G., 2012. Chemoassay screening of DNA-reactive mutagenicity with 4-(4-nitrobenzyl)pyridine – application to epoxides, oxetanes and sulfur heterocycles. Chem. Res. Toxicol. 25, 2092–2102.

Thienpont, B., Tingaud-Sequeira, A., Prats, E., Barata, C., Babin, P.J., Raldua, D., 2011. Zebrafish eleutheroembryos provide a suitable vertebrate model for screening chemicals that impair thyroid hormone synthesis. Environ. Sci. Technol. 45, 7525–7532.

Tollefsen, K.E., Nilsen, A.J., 2008. Binding of alkylated phenols and non-phenolics to the hepatic estrogen receptor in rainbow trout (*Oncorhynchus mykiss*). Ecotoxicol. Environ. Saf. 69, 163–172.

Tollefsen, K.E., Eikvar, S., Finne, E.F., Fogelberg, O., Gregersen, I.K., 2008. Estrogenicity of alkylphenols and alkylated non-phenolics in a rainbow trout (*Oncorhynchus mykiss*) primary hepatocyte culture. Ecotoxicol. Environ. Saf. 71, 370–383.

Truong, L., Reif, D.M., St Mary, L., Geier, M.C., Truong, H.D., Tanguay, R.L., 2014. Multidimensional in vivo hazard assessment using zebrafish. Toxicol. Sci. 137, 212–233.

US EPA, 2011. Integrated approaches to testing and assessment strategy: use of new computational and molecular tools us environmental protection agency. FIFRA Scientific Advisory Panel Consultation May 24-26, 2011. Available at: <http://yosemite.epa.gov/sab/sabproduct.nsf/373C1DB0E0591296852579F2005BECB3/$File/OPP+SAP+document-May2011.pdf> (accessed 16.09.14).

Villeneuve, D.L., Ankley, G.T., Makynen, E.A., Blake, L.S., Greene, K.J., Higley, E.B., Newsted, J.L., Giesy, J.P., Hecker, M., 2007. Comparison of fathead minnow ovary explant and H295R cell-based steroidogenesis assays for identifying endocrine-active chemicals. Ecotoxicol. Environ. Saf. 68, 20–32.

Vinggaard, A.M., Hnida, C., Breinholt, V., Larsen, J.C., 2000. Screening of selected pesticides for inhibition of CYP19 aromatase activity in vitro. Toxicol. In Vitro 14, 227–234.

Volz, D.C., Belanger, S., Embry, M., Padilla, S., Sanderson, H., Schirmer, K., Scholz, S., Villeneuve, D., 2011. Adverse outcome pathways during early fish development: a conceptual framework for identification of chemical screening and prioritization strategies. Toxicol. Sci. 123, 349–358.

Wambaugh, J.F., Setzer, R.W., Reif, D.M., Gangwal, S., Mitchell-Blackwood, J., Arnot, J.A., Joliet, O., Frame, A., Rabinowitz, J., Knudsen, T.B., Judson, R.S., Egeghy, P., Vallero, D., Cohen Hubal, E.A., 2013. High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environ. Sci. Technol. 47, 8479–8488.

Weinberg, A.M., 1971. Science and trans-science. Ciba Found. Symp. 1, 105–122.

Wong, K.Y., Duchowicz, P.R., Mercader, A.G., Castro, E.A., 2012. QSAR applications during last decade on inhibitors of acetylcholinesterase in Alzheimer's disease. Mini Rev. Med. Chem. 12, 936–946.

Worth, A.P., Patlewicz., G., 2014. Integrated approaches to testing and assessment. In: Whelan, M., Chantra E., (Eds.), Validation of Alternative Methods, Springer. (Unpublished – submitted).

Wu, S., Fisher, J., Naciff, J., Laufersweiler, M., Lester, C., Daston, G., Blackburn, K.B., 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem. Res. Toxicol. 26, 1840–1861.

Zaldívar, J.M., Mennecozzi, M., Marcelino Rodrigues, R., Bouhifd, M., 2010. A biology-based dynamic approach for the modelling of toxicity in cell-based assays. Part I: Fate modelling. Available from: <http://publications.jrc.ec.europa.eu/repository/>.

Zaldívar, J. M., Menecozzi, M., Macko, P., Rodriguez, R., Bouhifd, M., Baraibar Fentanes, J., 2011. A Biology-Based Dynamic Approach for the Modelling of Toxicity in Cell Assays: Part II: Models for Cell Population Growth and Toxicity. Available from: http://publications.jrc.ec.europa.eu/repository/.

Zhao, S., Narang, A., Gierthy, J., Eadon, G., 2002. Detection and characterization of DNA adducts formed from metabolites of the fungicide ortho-phenylphenol. J. Agric. Food Chem. 50, 3351–3358.