Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

19-Dec-2016

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

GUIDANCE DOCUMENT FOR THE USE OF ADVERSE OUTCOME PATHWAYS IN DEVELOPING INTEGRATED APPROACHES TO TESTING AND ASSESSMENT (IATA)

Series on Testing & Assessment
No. 260

JT03407308

ENV/JM/MONO(2016)67
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing & Assessment

No. 260

GUIDANCE DOCUMENT FOR THE USE OF ADVERSE OUTCOME PATHWAYS IN DEVELOPING INTEGRATED APPROACHES TO TESTING AND ASSESSMENT (IATA)

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris, 2016

ABOUT THE OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 35 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in 11 different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

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

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

This publication is available electronically, at no charge.

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

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

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

OECD member countries have been making efforts to expand the use of alternative methods in assessing chemicals. The OECD has been developing guidance documents and tools for the use of alternative methods such as (Q)SAR, chemical categories and Adverse Outcome Pathways (AOPs) as a part of Integrated Approaches for Testing and Assessment (IATA). However, there is a need for a systematic framework to characterise the individual biological and toxicological relevance of alternative methods in assessing chemicals in predicting toxicological endpoints. This framework could also inform their potential use in combination with other tools and methods to benefit from an integrated approach by applying mechanistic knowledge and understanding.

This document outlines an approach for the use of the AOP concept in developing IATA. It builds upon the workshop held in 2014 on a framework for the development and use of IATA (ENV/JM/MONO(2015)22) and experience to date with the development of IATA.

This document was prepared by the Secretariat and was endorsed by the Task Force on Hazard Assessment and is published under the responsibility of the Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology.

TABLE OF CONTENTS

1. INTRODUCTION

Current regulatory toxicity testing and assessment approaches remain to a large extent based on a checklist of *in vivo* tests, conducted in accordance with standardised test guidelines or protocols such as OECD Test Guidelines. While this approach has evolved over the past half century, it is unlikely to efficiently meet legislative mandates that require increased numbers of chemical assessments to be undertaken without a concomitant increase in the use of animals and resources. New approaches are necessary to close the gap between the number of chemicals in use and the number assessed to date. Significant advances in high throughput (HT) and high content (HC) methods offer new opportunities for gathering relevant information which quantify and characterise molecular and cellular responses to substances. For some endpoints, progress has been made in developing *in vitro* test methods; OECD Test Guidelines using *in vitro* techniques are available for skin/eye corrosion and irritation, skin sensitisation, genotoxicity and endocrine disruption. In recent years, these alternative test methods have influenced regulatory decision-making, especially when coupled with *in silico* approaches and grouping of substances into chemical categories. Thus, a shift is already occurring from a scheme basing toxicity assessment largely on *in vivo* test results to one incorporating results from alternative approaches (e.g. *in silico*, *in chemico*, *in vitro*, including HT/HC test methods).

At present, many testing approaches, irrespective of the particular methodology employed, do not result in a mechanistic understanding of the induced toxicity. This is particularly the case with non-animal testing approaches and understanding the relationship between what is tested and the apical toxicity endpoint being predicted. This is one of the reasons why results from novel approaches are not yet widely and consistently used for regulatory decision-making. Therefore, an objective and systematic framework is needed to characterise the individual biological and toxicological relevance of novel methods in predicting an adverse effect. The same framework could also inform their potential use in combination with other tools and methods to benefit from an integrated approach.

2. FRAMEWORK FOR DEVELOPING AND USING INTEGRATED APPROACHES FOR TESTING AND ASSESSMENT (IATA)

Integrated approaches to testing and assessment (IATA) are pragmatic, science-based approaches for chemical hazard or risk characterization that rely on an integrated analysis of existing information in a weight of evidence assessment coupled with the generation of new information using testing strategies. IATA follow an iterative approach to answer a defined question in a specific regulatory context, taking into account the acceptable level of uncertainty associated with the decision context.

There is a range of IATA - from more flexible, non-formalised judgment based approaches (e.g. grouping and read-across) to more structured, prescriptive, rule based approaches [e.g. Integrated Testing Strategy (ITS)]. IATA can include a combination of methods and can be informed by integrating results from one or many methodological approaches [(Q)SAR, read-across, *in chemico*, *in vitro*, *ex vivo*, *in vivo*] or omic technologies (e.g. toxicogenomics). The understanding of the likelihood of effects at lower levels of biological organisation (e.g. initiation of a toxicity pathway based on structure-activity relationships (SAR) and *in vitro* models), can help inform, in combination with other types of information in the IATA (e.g., exposure), on whether more resource intensive testing is warranted. This then contributes to an increased efficiency in the amount and type of hazard testing. This implies that there are potentially many different ways of applying an IATA. Even in cases where the workflow or decision logic of an IATA is documented (e.g. in a Guidance Document), the final approach taken will depend on the nature and level of existing information and, being generally underpinned by a weight of evidence approach (judgement-based approach), the decision-making process may not be fully harmonised.

While an IATA necessarily includes a degree of expert judgement (for example, in the choice of information sources and their weighting) some elements within an IATA can be standardised (i.e. rule-based). Particularly in certain areas of toxicology (e.g. skin sensitisation, skin corrosion and irritation), progress has been made in the development of defined approaches to testing and assessment, in which data generated by non-animal methods are evaluated by means of a fixed data interpretation procedure (OECD, 2014d; OECD, 2016b). Defined approaches could be standardised with a set of chemicals with available animal or human data for the hazard endpoint in question, before the methodology is applied for substances with data gaps. However, further experience needs to be gained in this area. When such approaches are clearly defined they can also be harmonised between countries to ensure consistency in how information is used in regulatory decision making. For this purpose, the development of testing guidelines (e.g. *in vitro* tests) or guidance documents (e.g. on application of *in silico* models) for these components of defined approaches is highly desirable. To standardise the evaluation of IATA in regulatory decision-making, guidance is being developed to provide principles for describing and evaluating defined approaches to testing and assessment (OECD, 2016a). In addition, reporting templates have been developed for different elements of IATA, such as read-across, so that the same documentation format for describing and evaluating IATA and its elements is used (OECD, 2016c).

While flexibility is foreseen in the construction of IATA, depending on the regulatory need and assessment context, IATA should ideally be mechanistically informed (Tollefsen et al., 2014). In other words, they should be based on knowledge of the mechanisms through which chemicals exert their toxicity. Mechanistic understanding provides a frame for the organisation and analysis of information from

methods that target different levels of biological organisation, enabling the contribution of these test results in deciding on the likelihood of the adverse outcome of interest (Tollefsen et al., 2014; OECD, 2015; Patlewicz et al., 2015; Perkins et al., 2015). Such mechanistic understanding can be provided by Adverse Outcome Pathways as outlined below.

2.1. The Adverse Outcome Pathway concept

An adverse outcome pathway (AOP) is a logical sequence of key events (KEs) triggered by chemical exposure and occurring at the molecular, cellular, organ, whole organism or population level (Figure 1). These KEs are causally linked to the adverse outcome (AO) under consideration and they are measurable. The AOP is anchored at one end by a molecular initiating event (MIE), which represents the direct interaction of a chemical with a biological target (Figure 2). At the other end, the AOP is anchored by an AO at the organism or population level. The AOs are often the reported endpoints from testing conducted using standard *in vivo* OECD Test Guidelines, or may be observations in other toxicological or epidemiological investigations.

The link between an upstream KE and a downstream KE in an AOP is called the key event relationship (KER). The KERs include the available evidence supporting the causal relationship between a pair of KEs (Villeneuve et al., 2014b; Edwards et al., 2016). KERs can also contain a quantitative description of the relationship between KEs (i.e., the level of change in the downstream KE that would be expected given a measured/predicted level of change in the upstream KE) and factors known to modulate that relationship. KERs also contain detailed mechanistic information of biological processes (named biological plausibility) that are involved and connect the upstream KE to the downstream KE.

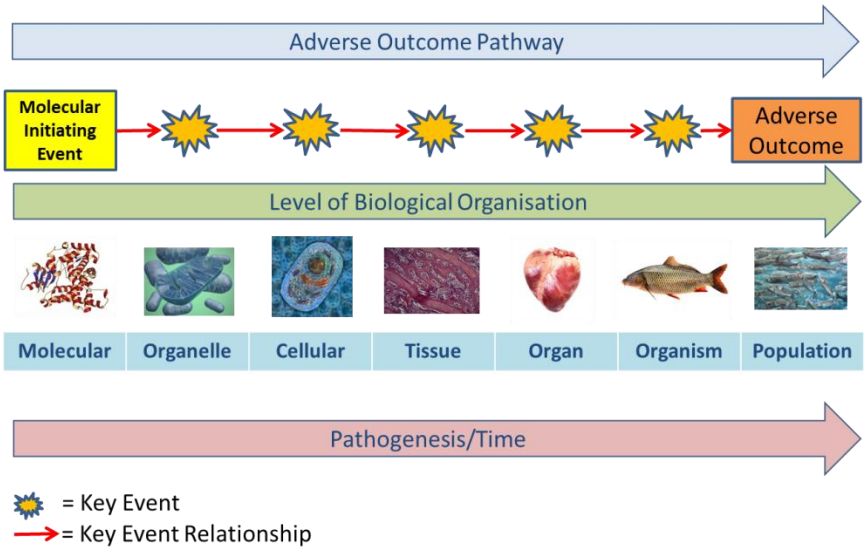

Figure 1: An AOP is a sequence of key events (KEs) linking a molecular initiating event (MIE) to an adverse outcome (AO) through different levels of biological organisation.

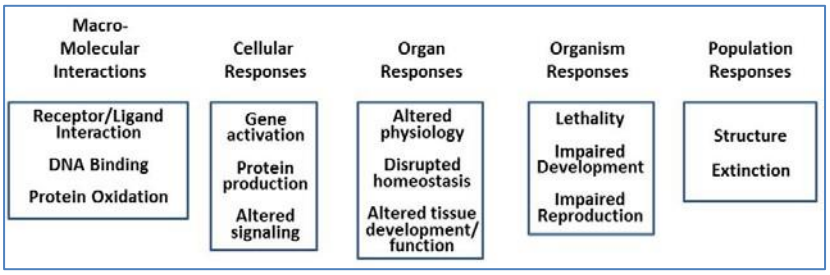

Figure 2. Examples of types of events that could be identified at different levels of biological organisation within AOPs (adapted from ENV/JM/MONO(2013)6).

Describing an AOP as a simple linear pathway beginning with a MIE and ending with an AO facilitates structured and clear organisation and evaluation of mechanistic information. This assumes that the simple linear model can indeed capture the essential elements of the perturbations of a biological system in order to adequately represent the critical phenomena that occur in vivo. However, AOPs that share KEs and KERs can form a larger AOP network that better represents the complexity of the pathways leading to an adverse outcome (Knapen et al., 2015; Edwards et al., 2016).

During AOP development, the evaluation of the underlying evidence linking KEs to one another can be based on the evolved Bradford Hill considerations, which are already used for the analysis of weight of evidence in the Mode of Action (MoA) context (Meek et al., 2014a; Meek et al., 2014b). However, these have been modified to be applicable to chemical agnostic (i.e. not chemical specific) AOPs (OECD, 2014a; Becker et al., 2015). The three primary considerations are:

- The biological plausibility of KERs
  - Is there a mechanistic (i.e., structural or functional) relationship between the upstream KE and the downstream KE consistent with established biological knowledge?
- The essentiality of KEs
  - Are downstream KEs and/or the AO prevented if an upstream KE is blocked?
- The empirical support of KERs
  - Does the empirical evidence support the hypothesis that a change in an upstream KE leads to an appropriate change in a downstream KE?
  - Does the upstream KE occur at lower doses and earlier time points than the downstream KE and is the incidence of the upstream KE more pronounced than that for the downstream KE?

Based on the weight of evidence (composed of biological plausibility & empirical support) for each KER and the essentiality for each KE, AOPs can be divided into the three operationally defined stages of development, which are described in more detail by Villeneuve et al. 2014a (see Table 1).

Table 1. Three operationally defined stages or phases of AOP development (adapted from Villeneuve et al. 2014a)

Operationally Defined Stage/Phase of development	Characteristics
Putative AOP	Assembly of a hypothesized set of KEs and KERs supported primarily through biological plausibility and/or statistical inference. Assembly of partial AOP with incomplete linkage between the MIE and AO as a result of known gaps and uncertainties.
Qualitative AOP	Assembly of KEs supported by descriptions of how the KEs can be measured and KERs supported by empirical evidence in addition to plausibility or statistical inference, along with qualitative evaluation of the overall weight of evidence supporting the AOP.
Quantitative AOP	Assembly of KEs supported by descriptions of how the KEs can be measured and the accuracy and precision with which the measurements are made along with KERs supported by quantitative understanding of what magnitude and/or duration of change in the upstream KE is needed to evoke some magnitude of change in the downstream KE.

2.2. IATA based on the AOP concept

The AOP concept can be applied as a framework to develop IATA as it allows one to: (a) evaluate in a structured way the existing information that is available for the chemical(s) of interest (see Figure 3) and possibly conclude on the hazard based on existing information; (b) identify and generate the type of information that might be required to increase the confidence level concerning evidence of a particular hazard; and (c) iteratively suggest which information is required to make a regulatory decision (see Figure 4). By evaluating existing information, an AOP allows for the mapping, organisation and integration of various types of information, ranging from *in silico* and *in chemico* data to field study data, around the MIE, KEs and the AO (see Figure 3).

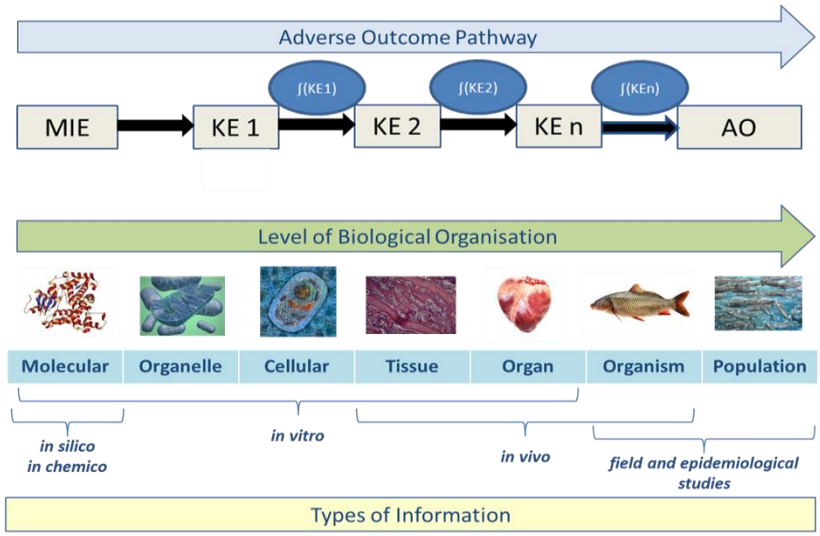

Figure 3: An AOP allows for the mapping, organisation and integration of various types of information, ranging from *in silico* and *in chemico* data to field study data, around the MIE, KEs and the AO.

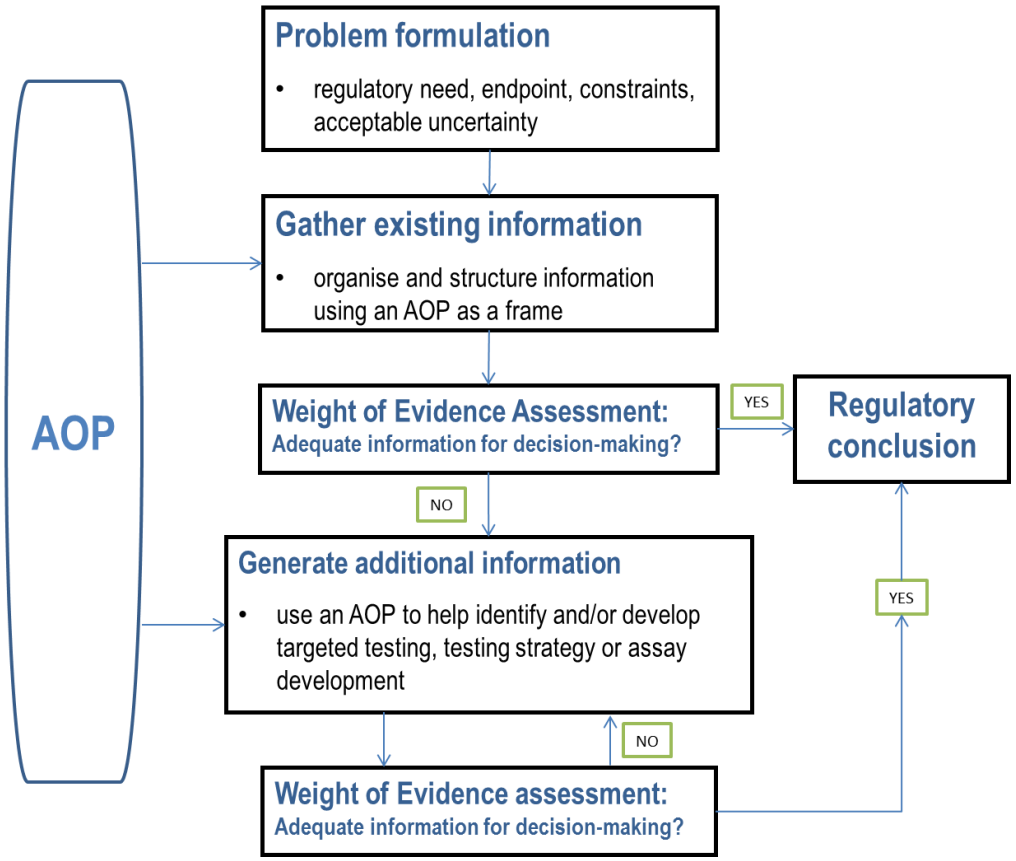

Figure 4: Framework for how an AOP can be applied to inform and structure IATA in a decision context

A high level of confidence and specificity of an AOP is important to derive test methods and defined approaches to be applied in a regulatory context. In some cases a quantitative AOP would be needed. Further defining how the level of evidence supporting a KE or KER should be classified during AOP development (OECD 2014a) will inform the level of confidence in the reliability and applicability of AOPs in various situations. In some cases a lower degree of confidence might be sufficient for an AOP to be used in an IATA context. This will be determined by the regulatory purpose. For example, a structural and mechanistic understanding from a putative AOP may be sufficient to interpret non-standard test results in a meaningful manner to prioritise substances for further assessment or testing. Qualitative AOPs, for which documented empirical evidence from one or more chemicals are available, can also inform the development of structure–activity relationships (SARs) that can even be quantifiable (QSARs) when a response – response relationship between the KEs is known (see section 3.1 and 3.2). Putative AOPs may also be valuable for the interpretation of data derived from high throughput (HT) and high content (HC) methods or omics technologies (see section 3.5).

To target the type of information that would be needed for regulatory decision making, AOPs can inform the design of testing strategies (see section 3.3). The extent of coverage of KEs within a testing strategy will very much depend on the assays developed for individual KEs and the mechanistic understanding derived from an AOP. To inform the development of testing strategies, at least the early and late KEs within an AOP need to be known and a qualitative understanding of the KER is needed. In

principle, the higher the confidence in an AOP the smaller the number of KEs will need to be covered in a testing strategy.

It is noted, that due to the network of pathways that AOPs may create it cannot be excluded that KEs other than those known and described in the AOP are in fact also leading to the same final AO. Therefore, an AOP informed IATA can only be used to identify substances with a likely AO and generally not to identify substances with no effects. It is important to derive a clear plausible relationship between the KE and the AO, in order to be able to conclude on which AOP is leading to an AO and whether there are interacting (networks) AOPs.

In an IATA it is also important to consider toxicokinetics since an AOP starts with the MIE only and therefore does not comprise ADME (i.e., absorption, distribution, metabolism and elimination) data. To determine the likelihood that a chemical and/or its metabolite(s) can reach the target organ(s) in the species of interest, however, toxicokinetics have to be taken into account. The toxicokinetics determine the relevant structural moiety (i.e., parent compound and/or metabolite(s) and site of the molecular initiating event(s) of the toxic action). Examining the physicochemical properties and structural features, the potential activation or detoxification processes, as well considering available toxicokinetic data or generating such data within IATA is crucial for supporting the validity of the prediction.

An important step in making AOPs useful for regulatory use is the development of reliable assays for KEs. Section 3.5 describes how to use AOP for the selection of the most essential KE(s) for the further assay and test guideline development.

The following sections aim to illustrate how AOPs may inform the development of different IATA for different purposes. Much of the information provided in section 3 derives from the report of a 2014 workshop on a framework for the development and use of IATA (OECD, 2015).

3. EXAMPLES ON HOW AOPS CAN BE USED IN THE DEVELOPMENT OF IATA

3.1 Development of (Q)SARs

As the MIE in each AOP involves a rather specific interaction of chemicals with biological systems, it may be used for generating mechanistically based structure–activity relationships (SARs) that can be used to predict whether a chemical can trigger an AOP. The SAR can also be used for chemical grouping to facilitate associated read-across or testing strategies (OECD, 2014c). If *in vitro*, *ex vivo* or *in vivo* assays have been developed for the MIE and/or one or more KEs along the AOP and have been tested for a certain number of chemicals, these results can be used to develop SARs, or QSARs (categorical or continuous when quantifiable) that can be used as a prediction or to confirm or refine the grouping of chemicals. This concept has been implemented within the OECD QSAR Toolbox for skin sensitisation (Dimitrov et al., 2016). SARs (called Profilers in the OECD QSAR Toolbox) have been developed that identify chemicals that trigger the MIE, i.e. the covalent binding with skin proteins. (Q)SARs that give continuous predictions or predict different categories (e.g. non, weak, moderate, strong) have also been developed based on substances that have been experimentally tested in assays characterising the different MIEs/KEs. Thus, a substance can be predicted to trigger the MIE/KE based on either a structural alert (SAR) or a QSAR and may be categorised as moderately reactive based on another SAR/QSAR derived from results of an assay measuring the MIE/KE.

When the sequence of KEs leading to a specific (adverse) effect is known at a sufficient level of detail, and the response- response relationships between the MIE, the KEs and the AO are well characterised based on results from *in chemico*, *in vitro*, *ex vivo* and/or *in vivo* assays, the toxicity of many other chemicals acting through the same AOP may be practically determined by predicting the MIE or any of the KEs, as illustrated in Figure 5. As mentioned earlier, it will be important to know the comparative kinetics and metabolism of the chemicals in question (i.e. both for the source and target substances).

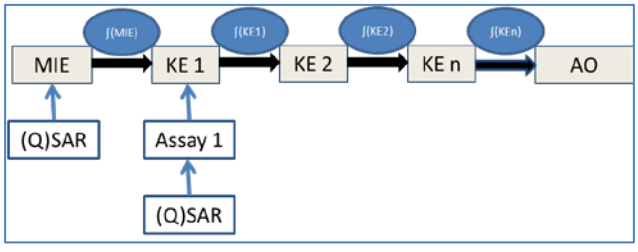

Figure 5. Use of an AOP to develop QSARs

3.2. Grouping of chemicals into chemical categories

AOPs can inform chemical grouping and subsequent data gap filling by read-across or trend analysis. Chemicals that are shown to activate the same AOP based on results of assays or predictions of the MIE or KEs can be grouped together, thereby improving the robustness of the data gap filling approach for the AO, compared to grouping chemicals solely based on their structural similarity. AOPs thus provide an opportunity to group chemicals based on their intrinsic chemical properties as well as their biological activity at different levels of biological organisation.

Whilst a complete knowledge of the AOP from the MIE to the final AO is not considered critical for the purposes of grouping substances around a common MIE or KE, establishing the linkages between the MIE or KEs and the AO will be needed to justify the data gap filling (such as read-across) performed. Figure 6 illustrates how a category of chemicals presumed to trigger the same AOP can be used for a read-across.

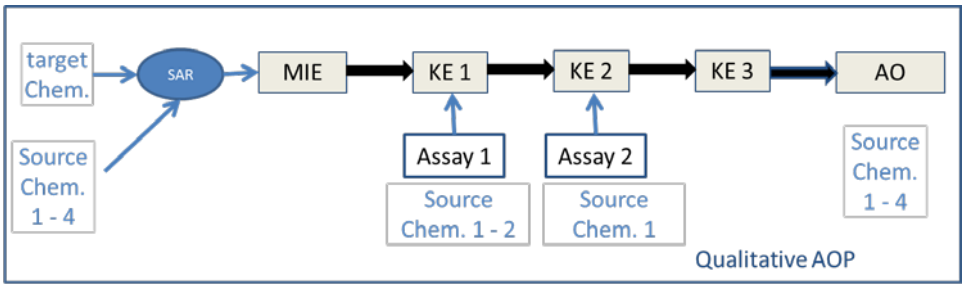

Figure 6. Use of the AOP concept to categorise chemicals for a specific endpoint

In the example outlined in Figure 6, it is predicted that exposure to four source substances (1-4) result in the same AO. The same information is lacking for a target substance which is structurally similar. An AOP has been developed where three KEs have been identified. In addition, a SAR has been developed that predicts the MIE (e.g. protein binding). For two specific KEs, identified *in vitro* assays are also available. Based on the SAR, it can be shown that both the source and the target substance will trigger the MIE. Based on the commonality in the MIE, it can be hypothesised that exposure to the target substance will result in a similar AO. For two of the source substances, *in vitro* test results show that they elicit KE 1, while one of these two substances also triggers KE 2. Based on these observations, it is likely that all four source substance exert their effects through this common AOP. This suggests that the target substance will also follow the same pathway resulting in the same AO thereby strengthening the read-across between the source substances and the target substance by using this structured mechanistic information that derives from the AOP. Depending on the potential use of this read-across prediction, the confidence could be strengthened by testing the target substance in assays which measure KE 1 or KE 2. Additional data and consideration of toxicokinetic aspects may permit an even more robust conclusion to be reached and even a relative ranking of potency amongst the substances.

3.3. Development of testing strategies

The AOP concept can be used to develop testing strategies for endpoints of interest by combining assays or prediction models that evaluate specific KEs along a particular AOP. However, the assays or prediction models and their combinations should be well characterised in terms of their applicability domain, their performance characteristics and combined in a transparent manner so that conclusions can be independently verified. For the use of AOPs in the development of testing strategies, it is of major

importance that quality criteria (strong KEs, strong KERs, sufficient examples of relevant chemicals, relevant toxicological endpoints) to ensure confidence in the use of AOPs in developing a testing strategy.

Once the available hazard information for a chemical has been identified and considered, the aim of a testing strategy is to gather information from a combination of tests that address different KEs along the AOP in a tiered-approach. Information from each tier is used to decide what test systems will generate the most relevant information in the next tier for the decision to be taken.

When developing a testing strategy, the level of confidence in an AOP can be used when deciding how many and which of the assays or prediction models developed for particular KEs need to be included in the testing strategy.

Figure 7 illustrates an example of how an AOP can potentially be used to inform a sequential testing strategy for the identification of a discriminant (positive or negative) endpoint. In this example, the MIE and two KEs are well characterised and *in silico*, *in chemico* and *in vitro* approaches are available. In addition, the individual performance of the non-animal tests has been compared to a standard *in vivo* test.

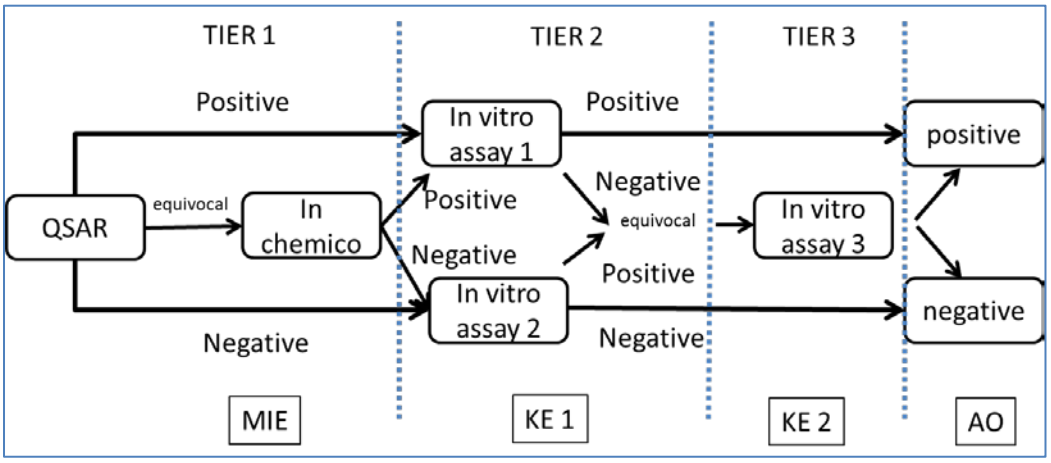

Figure 7. Use of an AOP in a testing strategy

In Figure 7, the MIE is known and can be characterised using a QSAR approach. The prediction made determines what subsequent testing is warranted. A positive prediction from the QSAR (Tier 1) triggers testing (Tier 2) with an assay that addresses KE 1 and has high a positive predictivity (low false positives), whereas a negative prediction from the QSAR triggers testing (Tier 2) with an assay that addresses KE 1 and has high negative predictivity (low false negatives). The final decision for the substances with a definitive positive or negative prediction in the Tier 1 analysis can be made in Tier 2 if the results in Tier 1 and 2 are concordant.

Substances for which the QSAR cannot generate an unambiguous prediction can be resolved in Tier 1 by testing in an assay that addresses the MIE. A positive or negative result from this assay determines which type of KE 1 assay should be used in Tier 2, namely one with a high positive or high negative prediction rate. Substances with conflicting results from Tier 1 and 2 are tested in Tier 3 by an assay addressing KE 2 and a weight of evidence approach is used to arrive at a final decision.

3.4. Interpretation of results from non-standard test methods

Linking a non-standard test method to a KE in an AOP provides context for understanding how to interpret these types of results and link them to an AO. Omic data (including toxicogenomics, transcriptomics, proteomics, and metabolomics) allow for more detailed insights into mechanisms of action, and can be applied to more efficiently survey the breadth of molecular/cellular effects elicited (*in vivo* or *in vitro*) by specific substances. Omic data could serve as either direct markers or indirect evidence of triggering a particular KE along an AOP leading to an adverse effect in the whole organism. Any omic dataset could potentially be associated with a KE, depending on the actual design of the experiment that was used to generate such data.

HT and HC data generated through *in chemico* methods, receptor binding or receptor transactivation assays, cellular reporter assays, may also serve to enhance identification of the chemical space associated with a particular KE. HT approaches have the potential to provide data on large numbers of chemicals in a cost efficient manner (Judson et al., 2013). In a prioritisation approach aiming at screening thousands of chemicals, HTS could be well positioned to identify new/novel chemicals that would be expected to initiate specific molecular targets or perturbation of cellular response pathways within AOPs. HTS or *in vitro* methods closely linked to a KE within a well characterised AOP would have high value in predicting an AO. It is noted, however, that such screening can only be used to identify substances with a likely (adverse) effect and generally not to identify substances with no effects because it cannot be excluded that KEs other than those known and described in the AOP are in fact also leading to the same final AO. This might then motivate the development of qualitative or quantitative AOPs and the creation of AOP networks that could eventually reduce subsequent higher tier testing. If the AOP consists of a clearly, quantitatively linked sequence of events (i.e., a chain of causative KEs), HTS assays might only need to target one of these events to be predictive.

3.5. Selection of methods for Test Guideline development/refinement

By linking KEs in an AOP to *in vitro* test methods and, when relevant, kinetic/ADME information (or refined *in vivo* methods with integrated kinetic information), the relationship between the results of the methods to hazard endpoints can be established. In practice, it makes most sense to develop test methods for a KE, or a set of KEs, that are sufficient to infer that an AO will occur following chemical exposure. In principle, triggering all KEs along the AOP is necessary for the final AO to occur, but none of them individually is sufficient. In practice, for predictive purposes, not all KEs need to be represented in a predictive model of the AO. Identifying KEs that are essential to induce the AO and have an established relationship with the AO will allow those who develop alternative methods to direct resources to the development of testing methods targeted to these specific informative KEs. This will also decrease the overall number of assays required for hazard identification. By reference to a (semi)quantitative AOP, Figure 8 aims to illustrate how the most appropriate assays can be selected for test guideline development.

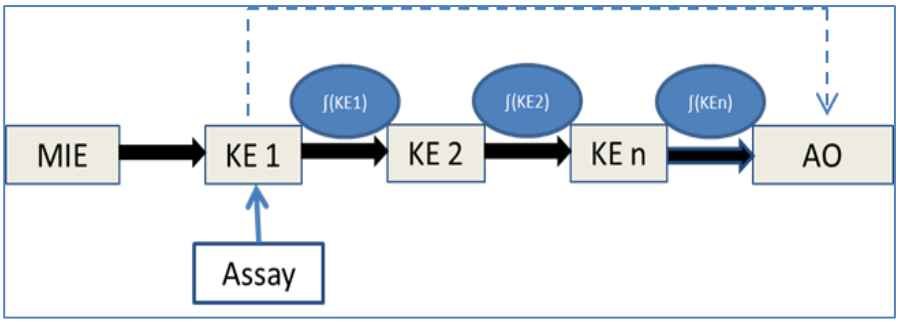

Figure 8. Illustrative example to show that a (semi)quantitative AOP can be used to target a KE, for which a Test Guideline could be developed or refined to predict an AO.

4. AREAS OF UNCERTAINTY IN THE DEVELOPMENT OF IATA BASED ON AOPS

In an AOP-informed IATA, the level of uncertainty is partly related to the limited development of the AOP, and partly to the reliability and coverage of methods used to measure or predict the MIE or the KEs. The following paragraphs investigate the sources of uncertainty associated with AOPs and the assays that are used in the development of IATA.

The amount and quality of data that ascertain the biological plausibility and empirical support of KERs, as well as the essentiality of KEs in an AOP or network of AOPs influence the confidence in the IATA conclusion. While the categorisation of AOPs in different stages of development (Table 1), as mentioned above, will be helpful to characterise AOPs, each AOP needs to be evaluated individually to determine its application in IATA based on the evidence provided. In this evaluation, the available evidence and associated confidence in the AOP is used to decide which KEs or KERs should be included in IATA. As confidence grows in a given AOP, it should allow more decisions to be made based on only a selected number of KEs. However, depending on the endpoint of interest, more than one AOP within an interconnected network might be needed in order to fully account for the biological processes that may influence the final AO. In this case, common KEs can emerge that should be considered during IATA development. However, many common KEs will be in close distance to the AO, but special attention should be given to also include KEs in the IATA that are higher up (closer to the MIE) and which may play a more fundamental role in the AOP network.

There are still challenges in determining how the weight of evidence supporting the KERs, and the causal support for the entire AOP, should be applied within an IATA. While biological plausibility is generally weighted more heavily than empirical support, there might be cases where the empirical evidence is quite strong whereas the biological plausibility has not been firmly established (Edwards et al., 2016). Both the AOPs for narcosis and hepatocellular proliferation are incomplete in the identification of essential causal KEs, but there is strong empirical evidence for the AO (Perkins et al., 2015). Consequently, during the construction and reporting of IATA, the combined considerations of biological plausibility and empirical support related to KERs or the whole AOP are required. It must also be emphasized that AOPs will exist on a continuous gradient from poorly-defined to extensively documented. And for this reason, evaluation of weight of evidence can facilitate the justification for a chosen KER or AOP to play a role in an IATA.

If an AOP-based testing strategy is developed, special attention should be given to the identification or possible development of assays that are needed for the measurement of the KEs. In some cases validation of the assays for the chosen KEs may be necessary, whereas in other cases reporting following the non-guideline *in vitro* methods guidance (OECD, 2014b) may be considered sufficient. In the reporting of the individual information sources used in a testing strategy, the status of development, standardisation or validation of an assay needs to be captured, indicating if the information source is: a) an officially adopted (standard) test method (e.g. an OECD Test Guideline); b) a validated but non-standard test method; c) a test method undergoing formal evaluation (e.g. prevalidation, validation, others); d) a non-validated test method widely in use; or e) a non-validated test method implemented by a small number of users (OECD, 2016a). An advanced example of this process is the evaluation of assays measuring the KEs from the skin sensitisation AOP (Reisinger et al., 2015). The case studies on skin sensitisation

(ENV/JM/MONO(2016)29/ANN1) indicate that even though single methods can be developed and validated for the different key events, it might be difficult to standardise and harmonise their regulatory application. Hence their validity and usability in testing strategies or defined approaches remains to be elucidated.

The same considerations are valid for (Q)SARs when they are used in an AOP-informed IATA. The limitations and the uncertainties related to the (Q)SAR models and their predictivity as well as the information and assays on which they are based, need to be taken into account. Furthermore, the validity of the (Q)SAR and its predictions should be evaluated with reference to the OECD principles of validation, for regulatory purposes, of (Q)SARs (OECD 2007).

5. THE WAY FORWARD

As the number of documented AOPs increases and more examples of AOPs at different stages of development (i.e. putative, quantitative etc.) are available, further demonstrations can be made, and guidance developed, of their application in IATA and also their use in various regulatory contexts. The current focus of the OECD is to continue the development of AOPs and to review case-studies on the practical application of IATAs. As experience grows, it is expected that guidance documents for the development of IATAs as well as harmonised IATAs will be developed.

REFERENCES

Becker, R.A., Ankley, G.T., Edwards, S.W., Kennedy, S., Linkov, I., Meek, B., Sachana, M., Segner, H., Van Der Burg, B., Villeneuve, D.L., Watanabe, H., Barton-Maclaren, T.S. (2015) Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence. *Regul. Toxicol. Pharmacol*. 72: 514-537.

Dimitrov, S.D., Diderich, R., Sobanski, T., Pavlov, T.S., Chankov, G.V., Chapkanov, A.S., Karakolev, Y.H., Temelkov, S.G., Vasilev, R.A., Gerova, K.D., Kuseva, C.D., Todorova, N.D., Mehmed, A.M., Rasenberg, M., Mekenyan, O.G. (2016) QSAR Toolbox - workflow and major functionalities. *SAR QSAR Environ Res*. 19: 1-17.

Edwards, S.W., Tan, Y.M., Villeneuve, D.L., Meek, M.E., McQueen, C.A. (2016). Adverse Outcome Pathways- Organizing Toxicological Information to Improve Decision Making. *J Pharmacol Exp Ther*.356: 170-181.

Judson, R.J., Kavlock, R., Martin, M., Reif, D., Houck, K., Knudsen, T., Richard, A., Tice, R.R., Whelan,M., Xia, M., Huang, R., Austin, C., Daston, G., Hartung, T., Fowle III, J.R., Wooge, W., Tong, W. Dix, D. (2013) Perspectives on validation of high-throughput assays supporting 21th Century Toxicity Testing. *ALTEX* 30(1): 51-56.

Knapen, D., Vergauwen, L., Villeneuve, D. L., Ankley, G.T. (2015). The potential of AOP networks for reproductive and developmental toxicity assay development. *Reprod. Toxicol*. 56: 52-55.

Meek, M. E., Boobis, A., Cote, I., Dellarco, V., Fotakis, G., Munn, S., Seed, J. and 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., Jeffrey Lewis, R. (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.

Organisation for Economic Co-operation and Development. (2007). Guidance document on the validation of (quantitative)structure-activity relationships [(Q)SAR] models. ENV/JM/MONO(2007)2.

Organisation for Economic Co-operation and Development. (2013). Guidance document on Developing and Assessing Adverse Outcome Pathways, (Series on Testing and Assessment No. 184) ENV/JM/MONO(2013)6.

Organisation for Economic Co-operation and Development. (2014a). Users' handbook supplement to the guidance document for developing and assessing AOPs https://aopkb.org/common/AOP_Handbook.pdf

Organisation for Economic Co-operation and Development. (2014b) Guidance document for describing non-guideline in vitro test methods (Series on Testing and Assessment No. 211) ENV/JM/MONO(2014)

Organisation for Economic Co-operation and Development. (2014c). Guidance on grouping of chemicals, second edition (Series on Testing & Assessment No. 194). ENV/JM/MONO(2014)4.

Organisation for Economic Co-operation and Development. (2014d). Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation, (Series on Testing and Assessment No. 203) ENV/JM/MONO(2014)19.

Organisation for Economic Co-operation and Development. (2015). Report of the workshop on a framework for the development and use of integrated approaches to testing and assessment, (Series on Testing and Assessment No. 215) ENV/JM/MONO(2015)22.

Organisation for Economic Co-operation and Development. (2016a). Guidance Document on the Reporting of Defined Approaches to be used within Integrated Approaches to Testing and Assessment. ENV/JM/MONO(2016)28.

Organisation for Economic Co-operation and Development. (2016b). Guidance Document on the reporting of Defined Approaches and Individual Information Sources to be used within Integrated Approaches to Testing and Assessment for Skin Sensitisation. ENV/JM/MONO(2016)29, ENV/JM/MONO(2016)29/ANN1 and ENV/JM/MONO(2016)29/ANN2.

Organisation for Economic Co-operation and Development. (2016c). Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA) - First Review Cycle (2015) - Case Studies on Grouping Methods as a Part of IATA.

Patlewicz, G., Simon, T.W., Rowlands, J.C., Budinsky, R.A., and Becker, R.A. (2015). Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regul. Toxicol. Pharmacol*. 71: 463-477.

Perkins, E.J., Antczak, P., Burgoon, L. D., Falciani, F., Garcia-Reyero, N., Gutsell, S., Hodges, G., Kienzler, A., Knapen, D., McBride, M. and Willett, C. (2015). Adverse outcome pathways for regulatory applications: examination of four case studies with different degrees of completeness and scientific confidence. *Toxicol. Sci.* 148: 14-25.

Reisinger, K., Hoffmann, S., Alépée, N., Ashikaga, T., Barroso, J., Elcombe, C., Gellatly, N., Galbiati, V., Gibbs, S., Groux, H., Hibatallah, J., Keller, D., Kern, P., Klaric, M., Kolle, S., Kuehnl, J., Lambrechts, N., Lindstedt, M., Millet, M., Martinozzi-Teissier, S., Natsch, A., Petersohn, D., Pike, I., Sakaguchi, H., Schepky, A., Tailhardat, M., Templier, M., van Vliet, E., Maxwell, G. (2015) Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. *Toxicol. In Vitro*. 29: 259-270.

Tollefsen, K.E., Scholz, S., Cronin, M.T., Edwards, S.W., de Knecht, J., Crofton, K., Garcia-Reyero, N., Hartung, T., Worth, A., Patlewicz, G. (2014). Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). *Regul. Toxicol. Pharmacol*. 70: 629-640.

Villeneuve, D., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landersmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M. (2014a) Adverse outcome pathway (AOP) development I: Strategies and principles *Toxicol. Sci.*142: 312– 320.

Villeneuve, D., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landersmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M. (2014b). Adverse outcome pathway development II: best practices. *Toxicol. Sci.* 142: 321-330.