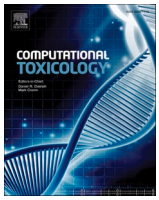
Computational Toxicology 31 (2024) 100325

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.sciencedirect.com/journal/computational-toxicology



Full Length Article

A developmental and reproductive toxicity adverse outcome pathway network to support safety assessments

Alun Myden *, Alex Cayley, Robert Davies, Jade Jones, Steven Kane, Daniel Newman, Martin P. Payne, Victor C. Ude, Jonathan D. Vessey, Emma White, Adrian Fowkes

Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS, United Kingdom

A B S T R A C T

Developmental and reproductive toxicity (DART) are key regulatory endpoints for the protection of human health. DART assessments require large numbers of animals, are expensive and often run at late stages of drug development. Therefore, new approach methodologies (NAMs) are being developed to transition away from animal testing. These NAMs (including *in silico* models) can be used to screen for DART hazards at the early stages of compound development and may in the future be used for regulatory DART assessments. Due to the implications of a mischaracterised developmental toxicant, both high confidence and understanding of the assessments made using NAMs will be required; it is likely that multiple NAMs will be needed in order to replace the current animal-based assessments. Adverse outcome pathways (AOPs) serve as a pragmatic tool for documenting mechanisms of toxicity. NAMs can be associated to key events (KEs) along an AOP, providing context to their outputs, and therefore increasing confidence in their use. It is likely that networks of pathways will be required for a specific toxicity endpoint in order to confidently apply an AOP-based approach to safety assessments. An insufficient number of DART AOPs are currently described within the public domain; therefore, using a literature-based approach, a network consisting of 340 KEs (including 68 MIEs) was developed. This foundation of pathways was made chemically aware through the association of relevant assays, data and expert rule-based structural alerts to appropriate KEs. The use of the network as a hazard screening tool was assessed, and the application of this to aid an ICH S5 workflow investigated. The knowledge captured within this AOP network can also guide the further development and use of DART-relevant NAMs and integrated approaches to testing and assessments (IATAs).

1. Introduction

Developmental and reproductive toxicity (DART) is an important set of toxicity endpoints that many industries use to determine a substance's potential to affect the ability to reproduce or impair the pre- and post-natal development of offspring. Regulatory guidance for DART testing came to prominence following the 1950 s-60 s thalidomide disaster, and have subsequently been refined [1]. In the current paradigm, the regulatory assessment of DART is undertaken using both rodent and non-rodent species [2]; as DART can impact the whole life cycle, a comprehensive assessment requires multiple studies to be performed at various life stages. Although these studies adequately identify potential toxicants, there are several limitations [3]. The requirement of multiple species and generations leads to high costs and large numbers of animals being used per study [4–6]. When considering industrial chemicals, it is thought that the number of animals required to test and register existing chemicals (under registration, evaluation, authorisation and restriction of chemicals (REACH) regulations) is impractically high (estimated to be up to 22 million vertebrates) [7,8]. Animal models provide limited insights into the mechanism of toxicity (e.g. through plasma hormone concentrations), as they are predominantly performed to identify gross/physical malformations (e.g. skeletal malformations) [9]. Finally, regulatory-relevant DART assessments are typically run at relatively late stages of chemical development [10,11] – therefore, depending on the intended use of the chemical, an adverse finding can result in the late termination of the development of a compound. As a result of these limitations, many new approach methodologies (NAMs) are being developed to align DART testing with the 3Rs of animal testing – reduction, refinement, and replacement [5] – and provide cheaper and earlier assessments of DART liabilities.

While it has been noted that no formally accepted definition of a NAM yet exists [12]; generally (and for the purposes of this work), NAMs are broadly accepted as alternatives to traditional mammalian methods, which can provide information regarding the hazards or risks posed by chemicals [13,14]. Examples of current NAMs include *in silico* and *in vitro* approaches [13,14]. Additionally within the field of DART, NAMs based on non-mammalian *in vivo* (alternative species) approaches have also been developed [5,15,16]. (Quantitative) structure–activity relationships ((Q)SARs) provide an example of an *in silico*-based NAM [14]. (Q)SAR models commonly predict for specific toxicity endpoints

* Corresponding author.
E-mail address: Alun.Myden@lhasalimited.org (A. Myden).

https://doi.org/10.1016/j.comtox.2024.100325
Received 30 January 2024; Received in revised form 30 July 2024; Accepted 5 August 2024
Available online 8 August 2024

(e.g. carcinogenicity or skin sensitisation) using historical data, and can be based on statistical correlations or expert-derived structure–activity relationships (SARs). Several DART-relevant (Q)SAR tools are available [17], including the commercially available SAR tool Derek Nexus which contains a suite of expert-derived SAR alerts for DART endpoints [18] (Fig. 1A). (Q)SAR tools provide a rapid means of identifying toxicity hazards without requiring any compound to perform a test. Also, depending on the level of curation of supporting information, these can provide insights into the mechanism of toxicity (Fig. 1A). Despite the benefits of DART (Q)SAR models, the coverage of chemical space in which confident predictions can be made has been shown to be quite limited [17,19]. *In vitro* NAMs focus on discrete biological events which could be indicative of mechanisms related to a toxicity endpoint (e.g. protein receptor binding or activation assays) [20], or represent fundamental processes such as the differentiation of embryonic stem cells [21]. Non-mammalian, whole organism NAMs (e.g. zebrafish toxicity assays) provide a wholistic approach to limiting the use of traditional animal models [16]. The use of either *in vitro* or non-mammalian, whole organism NAMs as part of a safety assessment will result in the generation of data which require interpretation and an understanding of the relevance of their outputs to mechanisms of toxicity. Additionally, data generated from these NAMs could be used to train *in silico* models, the outputs of which would also require the above contextualisation [22].

Adverse outcome pathways (AOPs) provide a formalised approach to documenting the mechanisms of toxicity in which an interaction of an exogenous substance with a biological system, through a molecular initiating event (MIE), causes a cascade of key events (KEs) to occur – eventually resulting in an adverse outcome (AO) [23]. Each KE should be measurable and can be associated with one or more relevant assays or *in silico* models. As a result, AOPs can facilitate the contextualisation of NAMs and the systematic integration of multiple observed or predicted effects along the whole AOP. This contextualisation of NAMs (as well as traditional assays), means that AOPs have the potential to be a useful aid in the transition to non-animal-based DART safety assessments; enhancing the development and use of NAMs for hazard identification, and aiding in the development of integrated approaches to testing and assessment (IATAs) for regulatory purposes – in turn, helping to progress the 3Rs of animal research [24,25].

It is likely that a network of many pathways will be required for AOPs to successfully support the use of NAMs in DART testing. Examples of well-documented DART-relevant AOPs exist within the AOP-Wiki and published manuscripts [26–29]. However, the coverage of pathways relating to known targets (e.g. MIEs) relevant to DART seems relatively low within the public domain [30]. A contributing factor for this may be that the development of AOPs can in itself be time- and resource-consuming.

The volume of evidence used to support an AOP is incredibly important as it not only influences the time required to develop the AOP, but also the confidence in its application for safety assessments. When developing a network of DART-relevant AOPs, different approaches can be used. At one extreme, a network of MIE-to-AO relationships could be automatically generated using pre-prepared datasets (through a statistical [30] or computer-generated approach); however, its application would require expertise, for example to inject mechanistic knowledge or address coincidental relationships resulting from biased training data. Such a network could prove useful to allow hazard screening, but this would provide little in the way of mechanistic understanding. At the other extreme, an in-depth approach to developing the AOPs requires vast volumes of evidence to be generated and curated into the pathways. While this would result in well-characterised AOPs, the approach would be very time- and labour-intensive which may be impractical for the development of a network of AOPs. A useful intermediate approach, balancing the time required to develop the network and the confidence in the use of the network, would be a literature-based approach focusing on repurposing existing knowledge and prioritising knowledge elements for AOP generation (Fig. 1B) [22]. AOPs developed using a literature-based approach would not only include the MIE-to-AO relationships (produced using a statistical approach), but also provide plausible mechanisms by which the MIE leads to the AO. The KEs describing the mechanisms of toxicity could then be used to associate existing NAMs and also to guide the development of new NAMs. These NAMs could then be integrated into the AOP network, further enhancing confidence in its application. This approach was taken for the work reported in this manuscript.

Herein we outline the development of a network of AOPs for DART endpoints and the subsequent association of assays, assay data and expert rule-based alerts to relevant KEs. We then investigated two potential use-cases that can be enhanced using the DART AOP framework: the first was the development of a screening tool for DART hazards, assessed against two datasets – a mammalian *in vivo* toxicity dataset, and a zebrafish developmental toxicity study dataset [31]; the second use-case focused on supporting a regulatory submission using NAM data, as described in the most recent iteration of the ICH S5 guidelines (ICH S5 (R3) Guideline on detection of reproductive and developmental toxicity for human pharmaceuticals) [2].

2. Methods

2.1. Target identification and AOP development

Biological targets (e.g. enzymes, receptors or signalling pathways) relevant to DART were identified from: alert comments within the expert rule-based SAR tool Derek Nexus [18], a publication by Wu *et al.* describing a chemical and biological class-based DART decision tree [32], and additional research and collaboration [33]. These biological targets were shortlisted and prioritised for further investigation and AOP development.

For the selected biological targets, evidence for their potential to impact DART endpoints, and the related mechanisms, was investigated. Where sufficient evidence suggested a biologically plausible link between perturbation of the target and the AO, an AOP was generated.

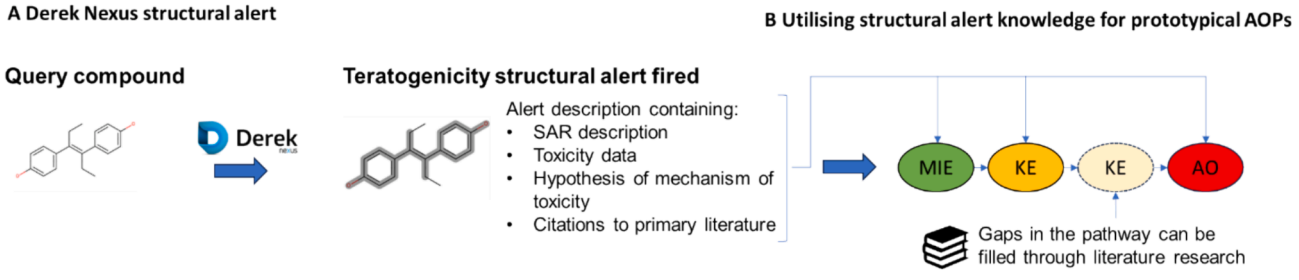


Fig. 1. A) Representation of how the SAR tool Derek Nexus can be used to make rapid predictions of toxic hazards – including examples of the supporting information contained within a Derek Nexus structural alert. B) An example of how mechanistic information within a Derek Nexus alert can be used to develop prototypical AOPs.

Biological targets related to enzymes or receptors often translated to MIEs within these pathways. When dealing with biological targets relating to signalling pathways, steps were undertaken during the literature review in order to identify the relevant proteins or enzymes to serve as MIEs and KEs within the AOP. The focus was to develop a network of mammalian-relevant pathways in which MIEs could be causally linked to DART outcomes. For each target under investigation, the initial aim was to develop a developmental toxicity pathway. If there was insufficient evidence available in the literature, then evidence relating to female fertility toxicity, male fertility toxicity or neurodevelopmental toxicity was reviewed and, where relevant, a pathway was developed for one of these other endpoints. Various resources were utilised when investigating each biological target. Initially, where possible, relevant mechanistic information contained within the expert rule-based software Derek Nexus [18] was extracted and used to develop a prototypical AOP. In addition, literature-based searches (utilising resources such as PubMed) were performed to identify relevant mechanistic information which could be used to enhance the prototypical AOPs or to aid in the full development of an AOP. These literature reviews were performed using key word searches using terms to identify evidence which could provide support for an AOP (e.g. [Biological target] + [Teratogenicity]). For each target, searches were undertaken for data relating to (1) DART studies for compounds known to interact with the target (e.g. a binder, inhibitor or agonist/antagonist), and (2) mechanistic studies examining the involvement of the target in DART. Available mammalian-relevant evidence from *in vitro* and *in vivo* stressor dosing studies, human cohort studies (where relevant), genetic animal knockout studies or human genetic data was considered and used to support the AOPs. Where available, the above evidence provided biological plausibility, empirical evidence and essentiality for the pathways, as described in the OECD 'Users' Handbook' AOP development document [34]. As a minimum, biological plausibility was required for the delineation of each KER within a pathway and additional empirical evidence and essentiality were included where available. Whilst searching the literature, if published DART AOPs relating to a relevant target were identified, these pre-existing pathways were reviewed and (where appropriate) integrated into the network. Integration of these pathways occurred after adaptation, in order to fit the terminology within the DART AOP network or to associate additional supporting evidence to the pathways.

Examples of the structure of the AOPs and the evidence used to support them can be found in the 2022 and 2023 Myden et al. publications [22,33]. Information regarding the KEs and KERs of each pathway was stored in an internal database [35]. This database provides a structured format to electronically capture AOPs and associated assay data.

2.2. Curation and mapping of assay data

Assays relevant to the AOP network were identified from the literature [31,36–46], and associated to specific KEs within the AOP network through relevant assay measurements [35,47]. These included assays that measured biological activity (e.g. enzyme inhibition or receptor binding assays); short-term *in vivo* assays measuring androgenic or oestrogenic effects (e.g. the uterotrophic or Hershberger bioassays [48,49]); and traditional and alternative developmental toxicity assays. High-quality curated databases (e.g. ChEMBL), EURL ECVAM validation studies and OECD test guidelines were reviewed to identify assays relevant to KEs within the AOP network.

In order to generate bioactivity data suitable for mining, data from each assay were grouped according to their assay type and the associated KE. Structure standardisation was performed using a pipeline in KNIME (version 4.5.2) to carry out validation (to check the validity of the structure, including valency or presence of unknown atoms) and normalisation processes (to generate a single tautomer of the structure). In addition to the structures, the biological data was also standardised across the database, including normalisation of activity call ('Positive', 'positive', '+', and 'active' were converted to 'Positive') and species information ('Human', 'Homo sapiens'). For *in vitro* potency data, a threshold of activity was assigned (e.g. 10 μM) – compounds which showed activity at or below the threshold were classified as 'Positive', and those above the threshold were classified as 'Negative'. If an author binary call was given and no potency data was available for a compound, then the binary call was used. Next the data were grouped, and a single binary call was generated for each compound where the assay protocol was the same (when a compound had multiple calls, a conservative call was taken). This resulted in a unique activity call per compound and assay protocol combination. Data from similar protocols were grouped into general assays, e.g. data from protocols which resulted in the measurement of aromatase inhibition were grouped under 'aromatase inhibition assay'. A contextualised chemical structure was also included for each datapoint – contextualisation refers to the further processing of chemical structures (e.g. through the removal of salts and stereochemistry). Finally, the data were added to the database containing the AOP knowledge.

2.3. Structural alert – Key event mapping

Structural alerts within the Derek Nexus knowledge base (Derek Nexus KB 2022, 2.0) were reviewed and, where relevant, associated to KEs within the AOP network [18]. The relevance was determined based on the endpoint the structural alert predicted for (e.g. teratogenicity) or the mechanisms described within the alert. In addition, structural alerts contained within an internal custom knowledge base were also associated to the AOP network. This custom knowledge base covered alerts for aromatase inhibition [22], glucocorticoid receptor binding, androgen receptor binding and oestrogen receptor binding.

2.4. AOP screening process

2.4.1. Screening of AOP-relevant assay data

A similarity searching method was utilised to query compounds in the AOP database. This employed a Tanimoto similarity method using a structural fragment-based fingerprint. Relevant datasets were processed through this method, using various thresholds of similarity (allowing for the identification of exact data for a query compound or for structurally similar compounds).

2.4.2. Screening of AOP-relevant structural alerts

In order to profile compounds against structural alerts, datasets were processed using Derek Nexus v6.2.1 and two knowledge bases: Derek KB 2022 2.0, with all toxicity endpoints considered, and the MIE alerts knowledge base described in Section 2.3.

2.4.3. Combining the results from the data-based AOP screen and structural alert-based AOP screen

Various combinations of similarity thresholds and structural alerts were evaluated to determine the performance of the AOP network in predicting the potential DART hazards that could be associated to compounds of interest. KNIME (version 4.5.2) was used for all data processing [50]. When considering data associated to the AOP network, reasoning was imposed to conclude that if a compound was associated with positive data, the output was considered to be positive; if no prediction was made by the network, then the output was assumed to be negative.

2.5. Derek Nexus developmental toxicity model

Derek Nexus structural alerts relating to the endpoints of teratogenicity and developmental toxicity were used as a performance benchmark to which the performance of the developmental toxicity-relevant AOPs within the network could be compared. This analysis was

undertaken to both evaluate the improved predictive coverage of the AOP network and to evaluate the use of the AOP network in an ICH S5 (R3) workflow.

3. Results and discussion

3.1. Overview of the AOP network

Literature review and curation of relevant evidence has led to the development of a DART AOP network (Fig. 2). In total, the network consists of 340 unique KEs, of which 68 are MIEs. As a result of the initial prioritisation focusing on the pathways of most concern, 301 of the 340 KEs within the AOP network relate to developmental toxicity pathways (encompassing teratogenic and embryo-foetal lethal pathways), whilst 51 relate to fertility toxicity pathways, and 33 relate to neurodevelopmental pathways. This grouping of pathways (into developmental toxicity, fertility toxicity or neurodevelopmental toxicity) was based on current regulatory guidelines, where developmental neurotoxicity is often addressed in separate assessments from developmental toxicity [9,51]. Already it can be observed that several KEs within the network have been utilised in more than one pathway, demonstrating interconnectivity within the network. For example, out of 12 MIEs within the network which lead to fertility toxicity outcomes, eight of these lead to both female and male fertility toxicity.

The biological roles and molecular functions of the MIEs within the network were profiled using the bioinformatics resource UniProt [52] (Fig. 3). This was achieved by identifying the relevant UniProt IDs for the protein described in each MIE and then using the associated 'UniProtKB Keywords' controlled terminology terms. Within UniProt, molecular functions were assigned to 59/68 MIEs, whilst biological processes were assigned to 52/68 MIEs. 45/68 MIEs were labelled with both a molecular function and biological process term, and only two MIEs within the network did not have a molecular function or biological process assigned to them. For each MIE within Fig. 3, the most commonly associated molecular functions and biological processes are reported. When reviewing the molecular functions associated with the MIEs within the network, it can be seen that the largest categories are oxidoreductases, DNA-binding, g-protein coupled receptors, and hydrolases – representing several classes of either enzyme or receptor (Fig. 3A). Other molecular functions (such as those related to transferase, lyase and the functioning of ion channels) are also represented within the network. Transport, transcription and lipid metabolism reflect the majority (approximately three quarters) of the biological processes related to MIEs within the network (Fig. 3B). Five MIEs relate to the biosynthesis of endogenous compounds/hormone (e.g. one-carbon metabolism, pyrimidine biosynthesis, thyroid hormone biosynthesis and nucleotide biosynthesis), three relate to angiogenesis, and six MIEs relate to a diverse range of biological processes. The two MIEs which were assigned with neither a function or a process term are both involved in metabolic processes.

To examine the sources of toxicity knowledge for the creation of the unified AOP network, the origins for each MIE hypothesis were examined (Fig. 4). Although the *in silico* model and decision tree were useful starting points, the additional research and collaborations resulted in the largest presence of MIEs within this network. The majority of the MIEs integrated into the AOP network via additional research were identified through one of two ways: 1) When investigating biological targets identified through Derek Nexus or the decision tree, often additional related targets were identified and incorporated into the network. 2) Targets were identified through collaborative research – one example of an AOP developed through collaboration can be found in the Myden et al. 2023 manuscript [33]. When comparing the MIEs stemming from Derek Nexus or the Wu et al. publication, 9 were identified from Wu et al., 11 from Derek Nexus and 17 of the 68 MIEs were identified from both Derek and Wu et al. This indicates both a reasonable overlap of biology and highlights that each resource was valuable in identifying relevant MIEs to develop into AOPs within the network. Fig. 4 also indicates that the biology described in the AOP network is broader than that described within DART-relevant Derek Nexus alerts, or biological categories captured in the Wu et al. publication. It should be noted that mechanistic drivers are not described for all chemical classes within the Wu et al. publication or DART Derek Nexus alerts.

The 'node degree' (i.e. the number of KERs leading to and from a KE) [53] was examined to determine how many KEs were re-used beyond a linear set of KERs (e.g. one KER leading into a KE, and one KER leading out of a KE). To do this, the total number of KERs per KE was examined – two KERs were deducted for a KE in the middle of a pathway, and one KER was deducted for an initial or terminal KE (e.g. an MIE). In doing so, it was found that 118/340 KEs were re-used at least once – demonstrating a high level of interconnectivity. The fact that the network was developed by a single research group allowed for careful control of language and curation, likely contributing to the high interconnectivity of the network. To ensure the development of networks of AOPs, each pathway related to teratogenicity and embryofoetal lethality terminates in the KE of 'developmental toxicity', and all pathways relating to male or female fertility toxicity terminate in the KE of 'fertility toxicity'. A schematic to show the fertility toxicity network is shown in Fig. 5.

To further examine the coverage of the DART AOP network, the AOPs within the network were compared to AOPs contained within the AOP-Wiki. When comparing against the AOP-Wiki, only pathways which were 'EAGMST approved', 'EAGMST under review', or 'WPHA/WNT endorsed' were considered – as these statuses reflect that the pathways are likely to be completed to a high standard or certified by an endorsement body. When performing this comparison, 16 AOPs within

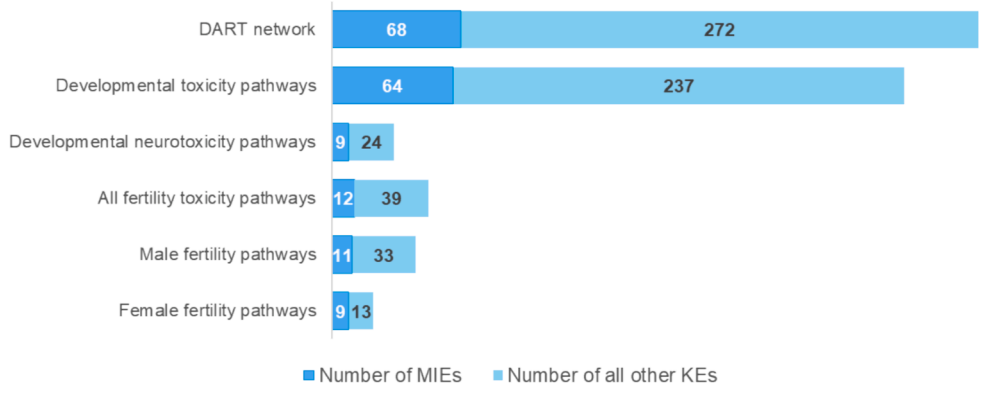


Fig. 2. Overview of the composition of the network. Male fertility and female fertility pathways are both sub-components of the 'All fertility toxicity pathways'. The developmental neurotoxicity pathways are not a sub-component of the developmental toxicity pathways. DART, developmental and reproductive toxicity; KEs, key events; MIEs, molecular initiating events.

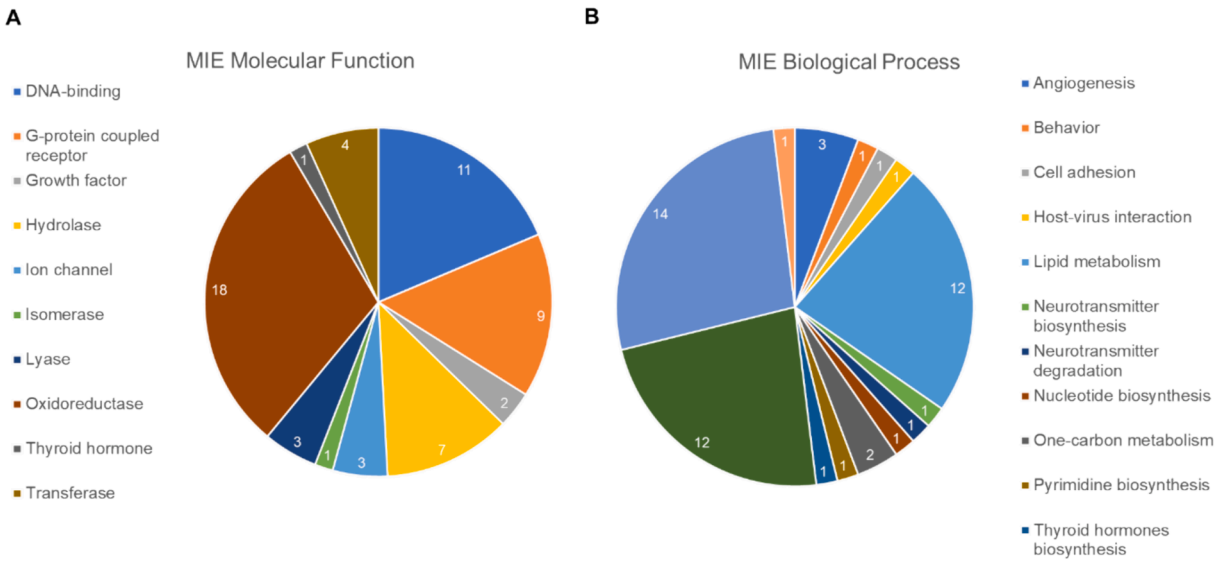


Fig. 3. A) Molecular functions associated with MIEs within the network, assigned using UniProt-controlled vocabulary terms [52]. B) Biological processes associated with MIEs within the network, assigned using UniProt-controlled vocabulary terms [52].

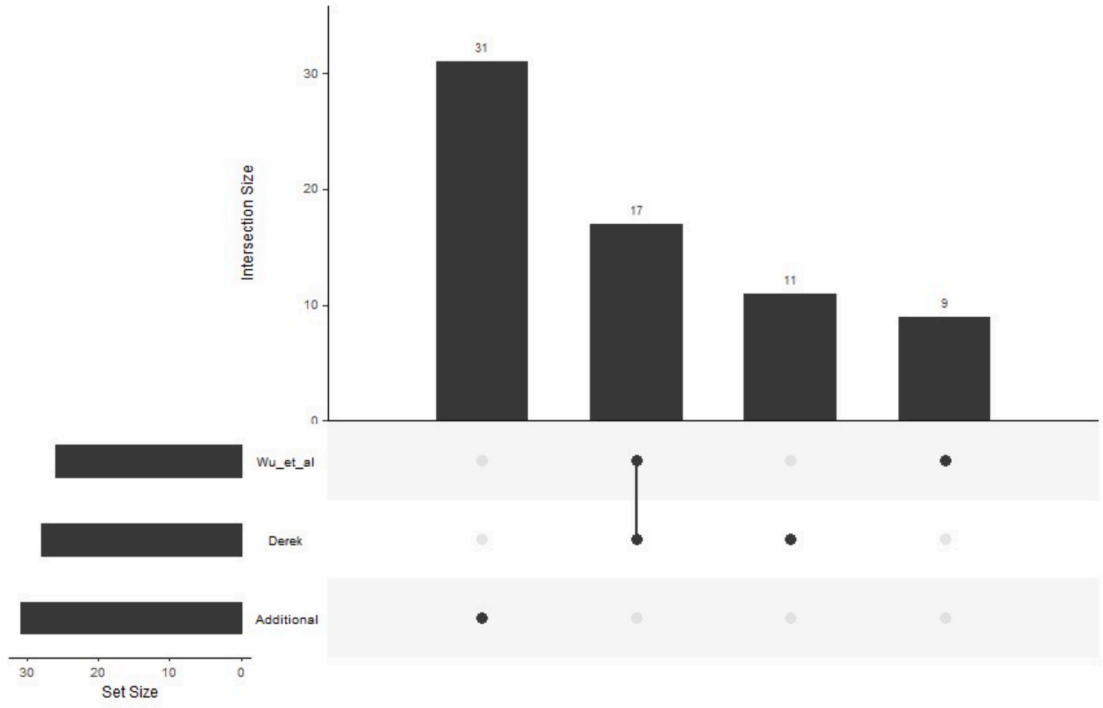


Fig. 4. The resources used to identify each MIE in the DART AOP network, along with the number of MIEs identified using each resource. The horizontal bar chart represents the total number of MIEs identified from each resource, whilst the vertical bar chart and below area represents the overlap in the number of MIEs identified within each resource.

the AOP-Wiki were identified as mammalian-relevant AOPs for DART endpoints. These 16 AOPs described 13 MIEs – three of which are not currently present in the DART AOP network described within this manuscript. One of the MIEs missing from the AOP network led to neurodevelopmental toxicity, whilst another two missing MIEs were relevant to genotoxicity. Pathways related to genotoxicity were deprioritised in the initial stages of development of the DART AOP network due to the creation of a complementary network for carcinogenicity-relevant AOPs [47], and the known association between genotoxicity and developmental toxicity [55].

3.2. *Creating a predictive framework from the AOP network*

Whilst developing the AOP network, structural alerts and assay data were annotated to relevant KEs (Fig. 6). Such annotations allow chemical-based queries to be performed, transforming the knowledge network into a resource which can support other aspects of chemical safety assessments. In total, 1,165,084 studies for 288,826 unique normalised compounds were associated to assays within the AOP network. Assay data was linked to 46 MIEs, 2 KEs, and 3 AOs within the network. In total, 240 Derek Nexus structural alerts were associated to 41 unique KEs within the AOP network – 112 to MIEs, 99 to KEs and 61 to AOs

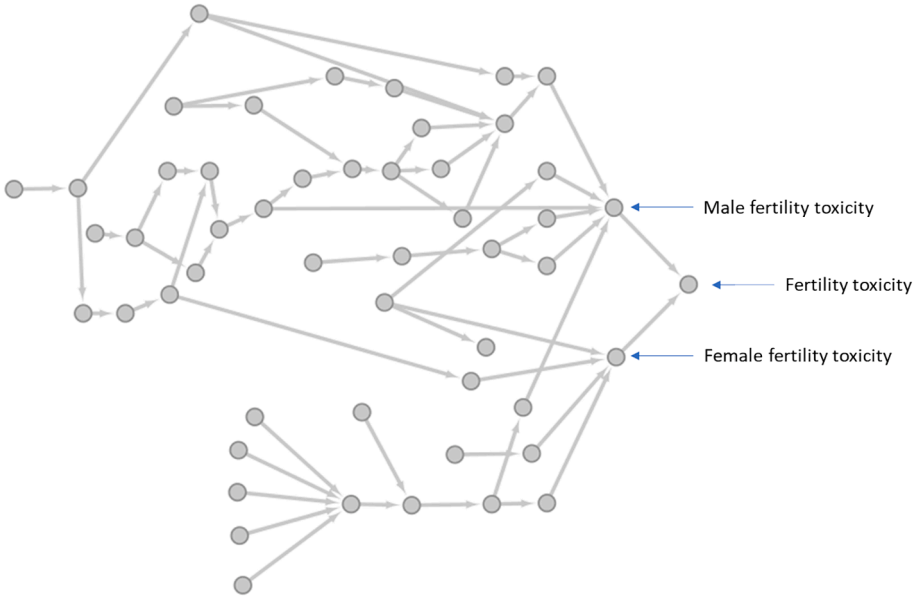


Fig. 5. Illustration of the network of pathways leading to fertility toxicity. Grey circles represent KEs, whilst the grey arrows represent KERs – the AOs of female fertility toxicity, fertility toxicity and male fertility toxicity are highlighted. Image developed in cystoscope, version 3.6.1 [54].

				Network total
Assays	46	2	5	**53**
Studies	1,159,224	2,375	3,485	**1,165,084**
Normalised compounds	288,315	281	1,794	**288,826**
Contextualised compounds	269,983	275	1,746	**270,308**
Derek alerts	112	99	61	**240**

Fig. 6. Distribution of data and structural alerts associated to the AOP network. AOs, adverse outcomes; KEs, key events; MIEs, molecular initiating events.

within the network. The mechanistic information documented within Derek Nexus structural alerts allowed for single alerts to be linked to multiple KEs. Hence, several alerts within the network are associated to multiple KEs (e.g. MIEs, KEs and/or AOs).

The majority of data associated to the network related to assays associated with MIEs. Relevant MIE data consisted of both binding and activity data for receptors, and inhibition data for enzymes. MIE data for each assay ranged from 1 to 345,581 studies (median: 2,693 studies) (Fig. 7). The prevalence of positive results in each assay dataset varied, ranging from 2–100 % with a mean of 61 %. Data for two KE assays were mapped to the network – a uterotrophic assay dataset contained 1,905 studies with a positive prevalence of 71 %, and a Hershberger assay dataset contained 470 studies with a positive prevalence of 57 % [45,46]. The distribution of the data, and prevalence of positive compounds per assay is important, as it will impact the density of chemical knowledge available across different areas of biological space (pathways) – this will influence how well a compound's relevance to certain pathways can be predicted.

The toxicity data associated to the network were grouped into five assay types: traditional *in vivo* assay data, zebrafish assay data, mouse embryonic stem cell data, rodent whole embryo culture assay data, and micromass assay data. The traditional *in vivo* assay dataset contained studies from ToxrefDB (version 1.3, August 2014 release), along with studies curated from the literature. This dataset contained 2,336 studies (with a positive prevalence of 50 %) for 1,111 normalised compounds [38]. A dataset of compounds tested in a zebrafish assay study [31], which measured multiple developmental endpoints (18 endpoints were measured at 120 h post-fertilisation), was also associated to the network. The dataset was extracted from the Environmental Protection Agency's 'CompTox Chemicals Dashboard' [37], and the extracted dataset contained an overall activity call for each tested compound. After validation and normalisation, the resulting dataset contained 1,038 unique compounds with a positive prevalence of 19 %.

Mouse embryonic stem cell data (67 studies for 54 compounds, with a positive prevalence of 68 %), rodent whole embryo culture assay data (21 studies for 21 compounds, with a positive prevalence of 79 %), and micromass assay data (23 studies for 23 compounds, with a positive prevalence of 65 %) – predominantly from published validation datasets – were also associated to AOs within the network [41–44].

The association of multiple different types of assays to the AOP network demonstrates how the AOP framework can contextualise the heterogenous evidence that is available to safety assessors.

3.3. *Using the predictive framework for hazard screening*

The use of the AOP network as a hazard screening tool for DART was assessed. The association of both assay data and structural alerts to the network provided several options for using the AOP network to screen

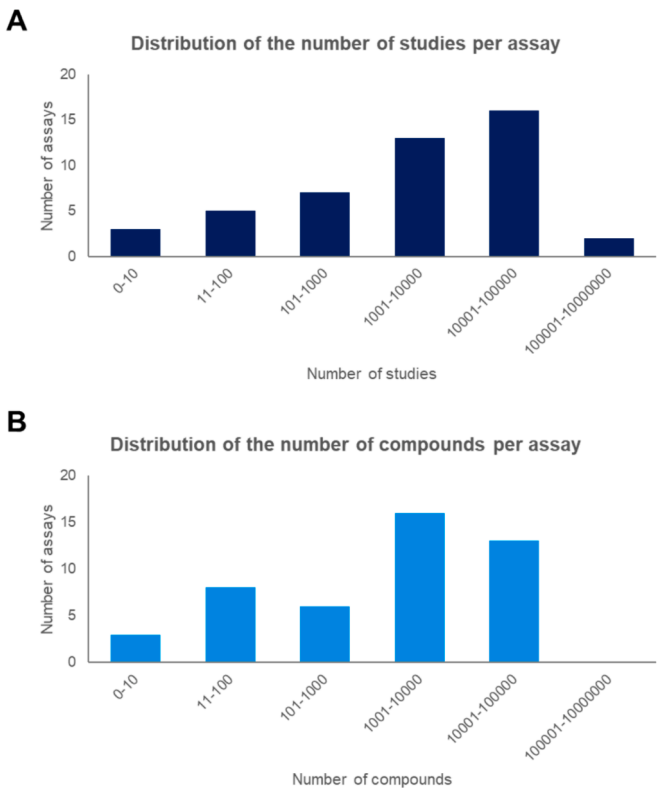


Fig. 7. Distribution of (A) the number of studies and (B) the number of normalised compounds associated to each relevant assay across the 46 MIEs in the DART network.

potential hazards (Fig. 8). In order to profile a query compound for DART liabilities, the network allows for a determination of toxicity based on matching existing data for the query compound of interest. In addition, the DART potential of a query compound can be predicted through an associated structural alert. Finally, through the use of similarity searching, the DART potential of a compound could be inferred based on the notion that structurally similar compounds will have similar biological properties. Each of these methods could be used in isolation, or the outputs of each assessment combined to potentially broaden the coverage of chemical space.

When evaluating the various methods, a conservative approach was taken so that any positive evidence (e.g. the presence of a Derek Nexus structural alert, an indication that a compound was able to initiate a KE [either for the exact compound or data for a similar compound]) would be treated as an indication of developmental toxicity. Contextualised structures were used to evaluate the various predictive methods, as this would help to ensure the matching of similar compounds. In addition, to the various outputs described in Fig. 8, the AOP network allows for predictions to be made for specific KEs. This distinction may be useful, as AO-based associations provide limited information about the mechanism but suggest that the compound of interest is a DART toxicant. In contrast, MIE/KE associations provide more insight about the mechanisms of toxicity, as they describe a biological event which is causally linked to a DART outcome through the described AOP – however, these may not always translate to the AO as they must reach the tipping point of each KER in between.

3.4. Evaluating the predictive framework against a traditional in vivo developmental toxicity dataset

In order to evaluate how well the AOP framework could identify mammalian developmental toxicants, the *in vivo* mammalian data associated to the AOP network was utilised. For the assessment, the *in vivo* mammalian dataset was removed from the AOP network in an attempt to evaluate the predictions in a setting where the concluding studies were yet to be performed. This experiment allows for the evaluation of whether the NAM data and models within the AOP predictive framework are capable of identifying developmental toxicants (according to traditional *in vivo* studies). The results would also indicate whether the AOP predictive framework could serve as a useful developmental toxicity hazard screening tool. When utilising the data associated to the network, three similarity thresholds were compared (100 %, 90 % and 80 %) (Fig. 8). The performance of the AOP network-based models were compared to predictions obtained using only Derek Nexus alerts for the endpoints of teratogenicity and developmental toxicity (DX-dev) (Fig. 9). Grouping of the contextualised *in vivo* dataset structures resulted in 874 unique compounds using the most conservative call (taking positive over equivocal, and equivocal over negative). This resulted in 474 positives and 400 negatives; compounds classed as equivocal were not included. For the DX-dev prediction method, a balanced accuracy of 72 % and a specificity of 94 % was observed; however, sensitivity was 51 %. The high specificity of the DX-dev structural alerts is not surprising, as these alerts were developed by experts to predict for *in vivo* developmental toxicity. Focusing on the performance of the AOP-SA method (alerts mapped to the AOP network), we see a similar balanced accuracy of 73 %, an 11 % increase in sensitivity, and a 9 % decrease in specificity when comparing these to the DX-dev structural alerts.

Focusing on the data associated to the AOP network, it was found that at a similarity of 100 % (the AOP-100 method), a balanced accuracy of 51 %, sensitivity of 38 % and specificity of 65 % is observed. Combining the exact-match data and AOP-relevant structural alerts (AOP-100-SA) provides a balanced accuracy of 65 % with a sensitivity of 74 % and specificity of 55 %. Incorporating the similarity searching methodologies with structural alerts (AOP-80-SA) gives a balanced accuracy of 63 %, and sensitivity increases to 78 %; however, specificity drops to 48 %.

The above analysis demonstrated that the structural alerts (e.g. DX-dev and AOP-SA) performed well against the dataset, with a high balanced accuracy and specificity. However, sensitivity could be greatly improved through considering both the AOP-based data and predictions. Sensitivity is an important metric for a screening tool, where the aim is to identify potential toxicants to better support prioritisation. In this scenario, the identification of false positives is tolerated, provided that a high number of true positives are identified. The corresponding trends of increasing sensitivity and decreasing specificity of the AOP models is not surprising, because a decrease in the similarity threshold results in additional positive predictions being made based on less reliable evidence. In an ideal world, data would be available or generated for each compound of interest – this would be expensive and likely unfeasible. Therefore, predictive models and/or read-across will be useful techniques to fill information gaps.

Despite the reasonable sensitivity output by the AOP-based methods, there were still 102 of the 474 teratogens in the dataset which were not predicted positive using the AOP-80-SA method. This could indicate one of two areas for improvement. It may be that relevant mechanisms of toxicity are yet to be integrated into the network, or it may be that suitable structural alerts or assay data need to be harvested and associated to relevant KEs within the AOP network. Even so, the Lhasa DART AOP network provides a foundation upon which to capture new knowledge or data as it is generated.

3.5. Evaluating the network against a zebrafish dataset

To determine the performance of the network against an alternative developmental toxicity assay, the screening methods were next tested against a zebrafish assay dataset. The zebrafish dataset was used to evaluate the performance of the AOP network for two reasons. The first is that, when evaluating the chemical space covered within the *in vivo*

A Model descriptions

Prediction method name	Description
AOP-SA	Inference of DART based on AOP-relevant Derek Nexus structural alerts which matched the input compound
AOP-100	Positive assay data linked to the network, for the exact input compound
AOP-100-SA	The combined results of AOP-100 and AOP-SA, using a conservative approach where any positive result indicated an overall positive outcome.
AOP-80*	An inference of DART based on positive assay data for compounds which have been identified as ≥80% similar to the input compound, associated to the AOP network
AOP-80-SA	The combined results of AOP-80 and AOP-SA, using a conservative approach where any positive result indicated an overall positive outcome.
DX-dev	Inference of DART based on structural alerts within Derek Nexus for the endpoints of developmental toxicity and teratogenicity. This prediction method was included to act as a performance benchmark, upon which we could assess the benefits of an AOP-based predictive model

B Workflow

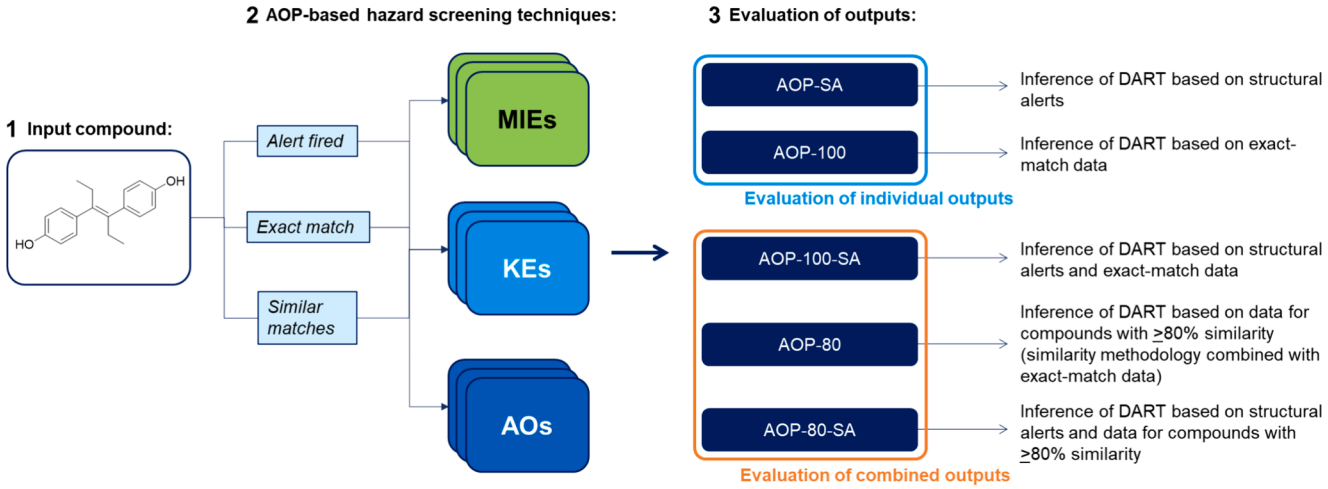


Fig. 8. A representation of how the structural alerts and data associated to the AOP network can be used to infer the DART potential of a compound of interest (different screening options). A) a description of some of the various methods assessed in the following sections. B) A depiction of the general workflow, B1) a compound of interest is identified. B2) structural alerts and data associated to the network is assessed to see whether a match can be found. Using a fragment-based fingerprinting method and Tanimoto similarity, data for similar compounds can also be determined. B3) Summary of the evaluation of the positive matches – many combinations are possible. *The similarity threshold of ≥ 90 % was also used: in this instance the assessed models were denoted as AOP-90 and AOP-90-SA.

dataset and the zebrafish datasets, it was found that the two datasets contained fairly chemically diverse structures (Fig. 10) and therefore screening the network against both datasets may provide further insights into the chemical coverage of the AOP network. Secondly, as described in Section 3.6, the use of the zebrafish dataset allows us to probe whether the Lhasa DART AOP network could be a useful aid to ICH S5 regulatory assessments based on alternative assay data.

The assay models embryo-foetal development, so the subset of developmental toxicity-related AOPs, data and predictions were assessed. In this instance, the traditional *in vivo* data were introduced back into the AOP network, and the zebrafish dataset was hidden from the network. This means that there were 327 exact matches between the zebrafish dataset and the traditional *in vivo* toxicity dataset; however, as described in Section 3.6, this allowed probing of the concordance between the zebrafish activity calls and traditional *in vivo* activity calls. It was also possible to determine the level of influence that the traditional data has on the predictive performance of the network. The structures in the zebrafish dataset were contextualised, and grouped based on contextualised structures – this led to 1,021 compounds (197 positives and 824 negatives). The resulting performance of each method tested can be found in Fig. 11.

Balanced accuracy of the DX-dev alerts and AOP-SA method with the zebrafish dataset were ~ 50 %, whilst all AOP-based models which utilised the data associated to the network had a balanced accuracy of ~ 60 %. Sensitivity of the DX-dev method was 7 % and that of the AOP-SA method was 15 %; whilst the sensitivity of the AOP models which incorporated relevant data ranged between 68 % (for the AOP-100 method) and 83 % (for the AOP-80-SA method). A converse trend can be seen in the specificity of each model, with DX-dev and AOP-SA being very high (91 % and 90 % respectively) and all other methods ranging from 33 % to 53 %. Only 34 of the 197 developmental toxicants were not identified using the AOP-80-SA method.

When evaluating the performance of the various methods against an alternative assay dataset (Fig. 11), we found that the structural alert-based methods (DX-dev and AOP-SA) did not perform as well, compared to their performance using the *in vivo* developmental toxicity dataset; however, specificity was still high. This contrast could be because the Derek Nexus-based models were trained on traditional animal data (predominantly from the pharmaceutical space), whilst the MIE/KE data may be more typical of that found in the zebrafish dataset. As mentioned, only 327 compounds overlapped between the zebrafish dataset and the *in vivo* toxicity dataset, and clustering analysis indicates

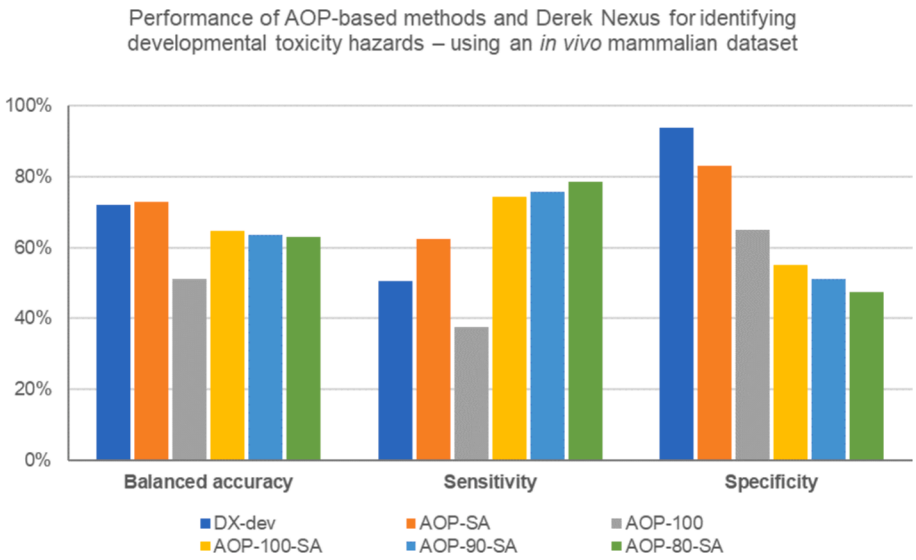


Fig. 9. The performance (balanced accuracy, sensitivity, and specificity) of various models evaluated using an *in vivo* developmental toxicity test dataset. A description of the models can be found in Fig. 8 – in brief, DX-dev relates to structural alerts contained within Derek Nexus for the endpoints of teratogenicity and developmental toxicity; all other models reflect combinations of AOP predictions, based on assay data or structural alerts mapped to KEs. AOP, adverse outcome pathway; DART, developmental and reproductive toxicity; DX, Derek Nexus; KEs, key events; SA, structural alerts.

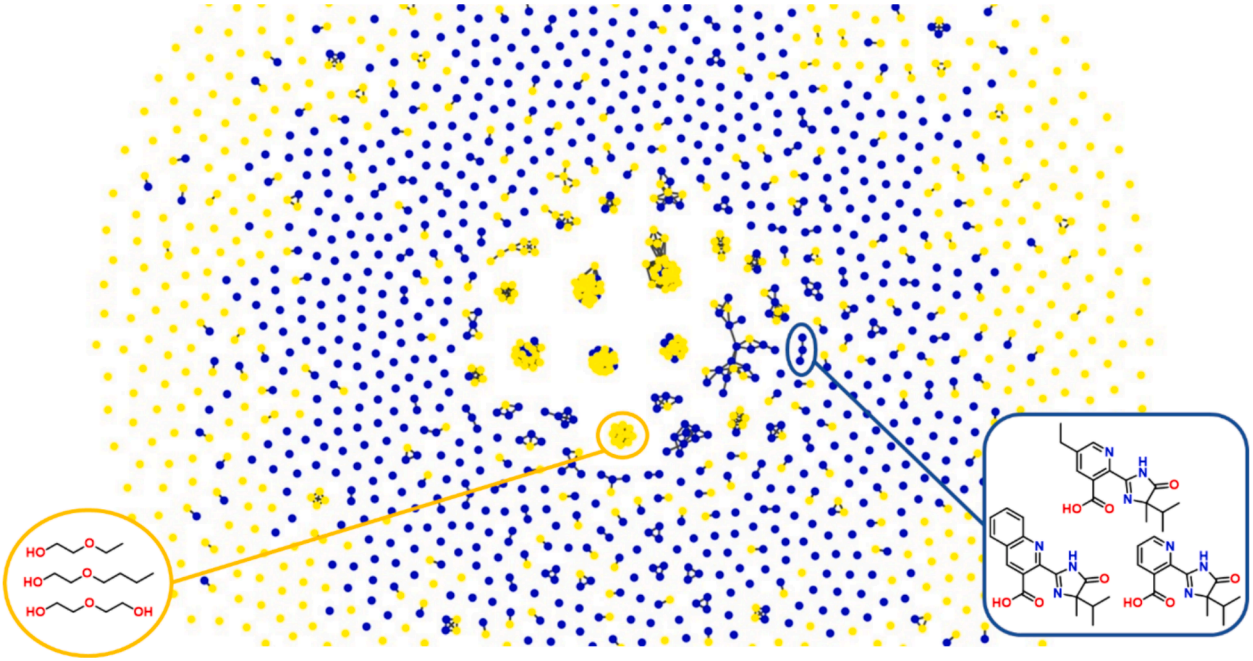


Fig. 10. Representation of the chemicals within the zebrafish dataset (yellow) and the traditional *in vivo* toxicity dataset (blue). The image was produced using a fingerprinting method, Tanimoto similarity and a similarity threshold of 80% to cluster similar chemicals. 327 of the blue to yellow interactions represent exact matches. Two clusters of chemical space are highlighted – these indicate an example of industrial chemicals (within the zebrafish dataset) and pharmaceutical-like chemicals (in the *in vivo* dataset). The largest clusters of chemically similar compounds are present in the centre of the image; most of these clusters are biased towards one dataset or the other. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that many of the compounds within these datasets appear to be structurally diverse (Fig. 10). Nevertheless, this finding highlights the complementarity of the two methods (structural alert- and data-driven) of screening, and that a consistent sensitivity can be achieved when combining both methods.

3.6. Using the AOP network to support an ICH S5 (R3) workflow

The ICH S5 (R3) guidelines propose scenarios in which NAMs could be used to classify compounds as developmental toxicants when a relevant mechanism has been identified [2]. One such scenario outlines how knowledge of a mechanism of action, in combination with a positive result in a suitable NAM (as defined in the guidelines), could be used in place of traditional mammalian animal testing (Fig. 12). The utility of the AOP network in identifying relevant mechanisms of toxicity to aid in this use-case was investigated.

To test this use-case, further examination of the outputs from the AOP-80-SA method described in Section 3.3 was undertaken. The

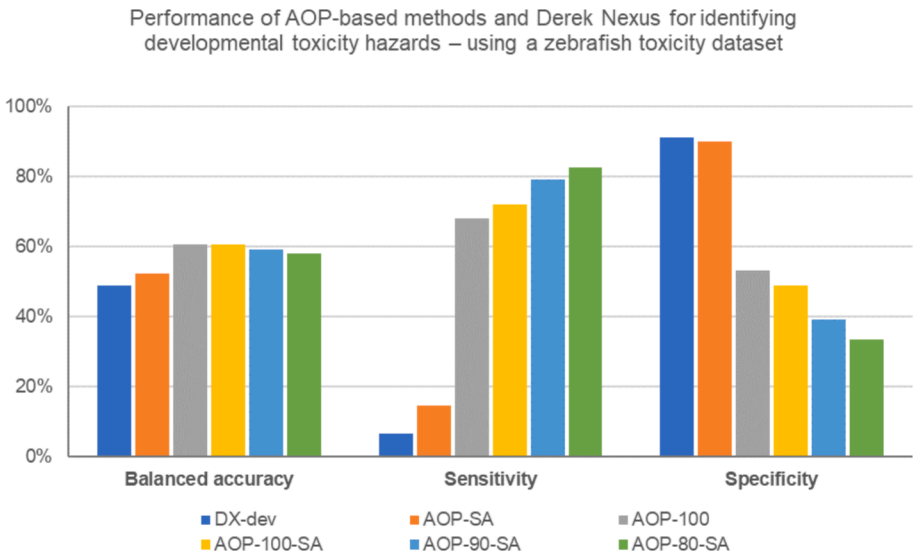


Fig. 11. The performance (balanced accuracy, sensitivity, and specificity) of various models evaluated using a zebrafish developmental toxicity dataset. A description of the models can be found in Fig. 8 – in brief, DX-dev relates to structural alerts contained within Derek Nexus for the endpoints of teratogenicity and developmental toxicity; all other models reflect combinations of AOP predictions, based on assay data or structural alerts mapped to KEs. AOP, adverse outcome pathway; DART, developmental and reproductive toxicity; DX, Derek Nexus; KEs, key events; SA, structural alerts.

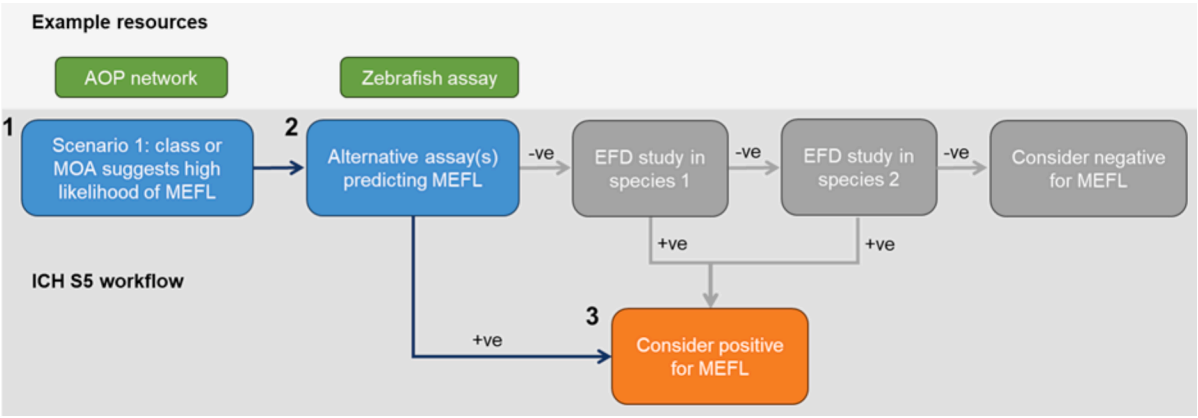


Fig. 12. Bottom panel – recreation of a workflow described in the ICH S5 (R3) guidelines [2]. A combined understanding of the mode of action (step 1) and a positive result from a NAM (step 2) would allow step 3 to be reached without the need for additional DART testing. Top panel – potential resources to fulfil steps 1 and 2. AOP, adverse outcome pathway; DART, developmental and reproductive toxicity; EFD, embryo-foetal development; MEFL, malformation or embryo-foetal lethality.

zebrafish assay was used to illustrate this use-case, as this is a reasonably well-known NAM, and a suitable dataset was available for this analysis. Potential mechanistic insights from the AOP network were the focus of this scenario, so research into the suitability of the zebrafish assay for the ICH S5 workflow (Fig. 12) and an in-depth analysis of zebrafish assay predictivity for mammalian developmental toxicity was not undertaken.

When using this workflow (Fig. 12), a researcher may either run the NAM assay in response to an indication of a mechanism of toxicity for a compound (which can be identified using the AOP network), or run the NAM assay first, identify a positive finding, and then wish to determine the mechanism of action. In either scenario, a combination of the mechanism of toxicity and a positive finding in a NAM assay could potentially be used to limit the need for traditional animal testing. For this use-case, predictions based on data or models associated with MIEs or KEs are most informative as they provide insights into the potential mechanisms of toxicity. In contrast, the AO-based predictions would only allow for hazard identification (Fig. 13).

First taking the scenario where the AOP network is used to identify potential toxicants and guide further testing, the mechanistic predictions and assay data identified potentially relevant mechanisms for 691 of the compounds in the zebrafish dataset (Fig. 13). Therefore, a researcher may wish to further investigate these signals by running the zebrafish assay – at which point they would find that 162 of the predictions were developmental toxicants in zebrafish. However, prior to running the zebrafish assay, they could use the AOP network to determine which other assays (i.e. MIE assays) could be run to confirm the mechanistic prediction – aiding the decision of whether to run a NAM assay or not. 205 of the compounds predicting positive for MIEs or KEs were identified based on data for similar compounds or from structural alerts; therefore, it may be useful to run confirmatory studies on these compounds prior to running the zebrafish assay. Within the AO-based predictions, 122 false positives are found (Fig. 13C). 83 of these false positives are based on experimental results for exact compounds, and these could indicate chemicals to which the zebrafish assay is less sensitive versus the traditional *in vivo* mammalian assays. For the workflow proposed in the ICH S5 (R3) guidelines (Fig. 12), it may be preferable for the NAM to have high specificity to limit the number of false positives identified as developmental toxicants. Subsequently, the compounds identified as non-developmental toxicants using the NAM will be tested using the traditional animal models.

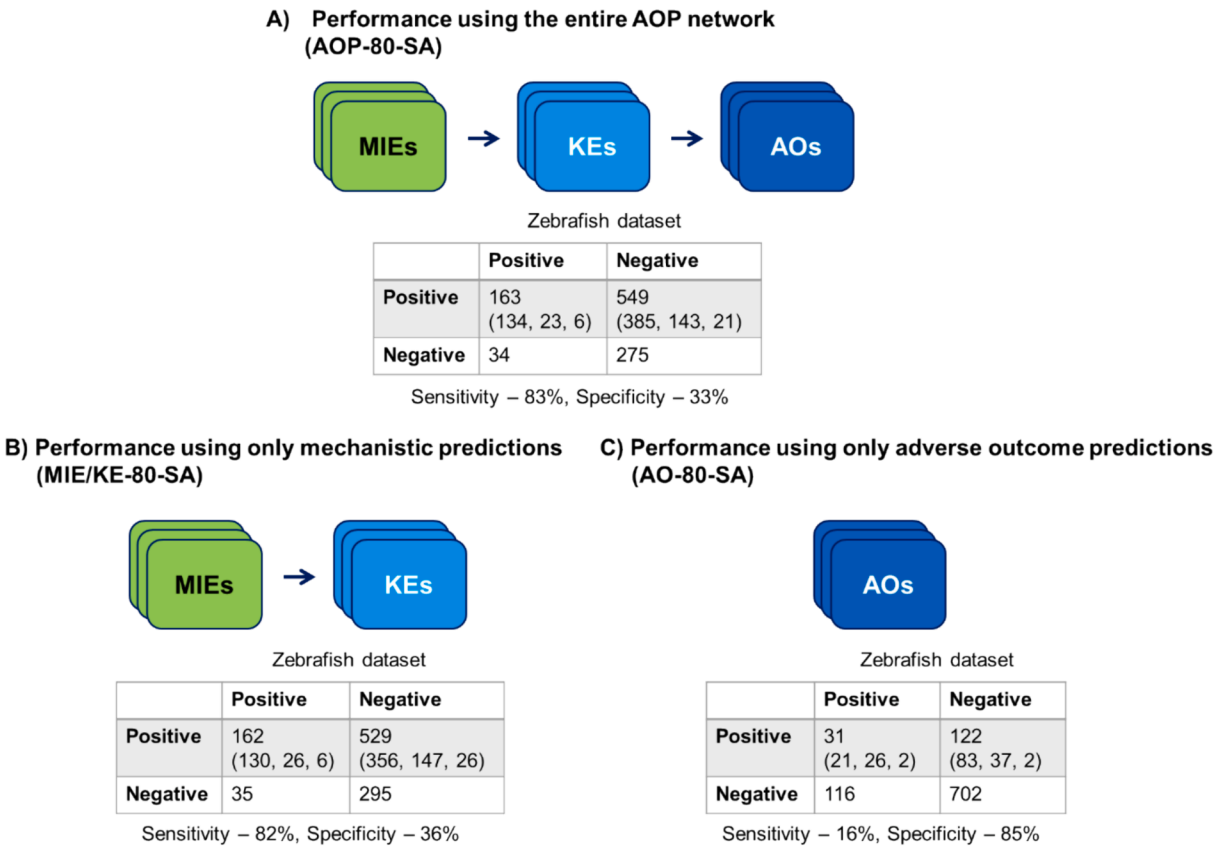


Fig. 13. Confusion matrices of the outputs from screening the zebrafish assay dataset against the AOP network using the AOP-80-SA method. A) The confusion matrix when considering all of the data and structural alerts mapped to the AOP network. B) the confusion matrix when only considering the MIE/KE-relevant data or structural alerts. C) The confusion matrix when only considering the AO-relevant data or structural alerts. The numbers in the brackets reflect how many compounds were classified using data for the exact compound, similar compound or Derek Nexus alert respectively. AO, adverse outcome; AOP, adverse outcome pathway; KE, key event; MIE, molecular initiating event.

In the second scenario, an investigator may first run the NAM assay and determine that 197 compounds were potential developmental toxicants – they may then wish to determine the relevant mechanism of action. In this scenario, the AOP network was able to propose mechanisms of action for 162 of the 197 positive results in the zebrafish assay dataset. The results associated to each of these proposed mechanisms of action should be reviewed by experts to ensure that they are applicable. However, taking the screening results at face value, the model could be used to support the submission of 82 % of the relevant zebrafish studies if the ICH S5 use-case (Fig. 12) were followed. Pleasingly, when comparing the mechanistic based predictions (Fig. 13B) to the predictions made by the entire network (Fig. 13A), we find that mechanisms can be associated to 162/163 positive compounds identified by the AOP network. When comparing the MIE/KE predictions (Fig. 13B) to the AO predictions (Fig. 13C), we find that the MIE/KE data and models associated with the network are driving the majority of the predictions. Therefore, for the 162 instances where mechanisms can be associated, the performance of a traditional animal study may not provide any additional value to the method of combining a positive result from a NAM with an AOP-based mechanistic hypothesis. The improvement achieved through the AOP-based approach is significant when compared to the DX-dev alerts which identified 13 of the positive zebrafish studies. It should be noted that the ICH S5 guidelines relate to pharmaceutical compounds, whilst the zebrafish dataset contained chemicals from a broad range of industries. Nevertheless, this still provides a useful illustration of how the Lhasa DART AOP network could complement a NAM-based regulatory submission. Further assessment of this workflow, with NAM data for pharmaceutical compounds, would be very useful – this may also allow the assessment of whether there is value in running two animal studies in the event of a negative result from the AOP network in combination with negative results from both a NAM and one traditional animal study.

Looking beyond the workflow described in Fig. 12, to a time where traditional animal testing is no longer required, the 122 false positives highlight the possible need for additional data from other NAMs (and a better understanding of the applicability domain of them) to ensure that non-animal methods are at least as predictive of the human scenario as the current traditional *in vivo* testing regimens. As AOPs provide mechanistic contextualisation for NAMs, the DART AOP network could be utilised to help reason between multiple NAMs. The biological coverage of NAMs could also be evaluated using the known pathways contained within the network. Furthermore, knowledge of the applicability domains of NAMs (or any assays) can be encoded within the AOP network. This future research could facilitate the development of AOP-based weight of evidence assessments and aid risk assessor's decision making.

4. Conclusion

To our knowledge, the Lhasa DART AOP network is the most comprehensive, expert-curated, mammalian-relevant DART AOP network currently available. The AOP network provides a foundation of mechanistic knowledge upon which to build, using additional relevant evidence when available. AOPs can aid safety assessments by providing a summary of evidence regarding specific mechanisms of toxicity. As well as being a useful reference source, this method of documenting a mechanism of toxicity allows for the contextualisation of mechanistic-based assay data and models.

The AOP network described within this manuscript allows for the screening of compounds in order to predict both the potential for DART and also the mechanism by which it may be caused. We demonstrated that sensitivity of the network can be very high when utilising both structural activity relationships and data associated to the network. The improvement in sensitivity is particularly apparent when comparing this to Derek Nexus developmental toxicity-related alerts. Therefore, the network and predictive framework could be a valuable resource for compound prioritisation and mechanistic elucidation. The ICH S5 use-case explored within this manuscript (where the identification of teratogens using a suitable NAM, combined with an understanding of the mechanism of toxicity, can be used in place of traditional animal models) also highlighted that the network may be a valuable aid to support the uptake of NAM-based assessments. As described in Section 3.3, a conservative approach was taken to the predictive framework outlined in this manuscript. Therefore, any positive signal was used as an indication of a compound's potential to cause developmental toxicity. Future research could focus on combining evidence along a pathway in order to enhance the confidence in a prediction. This may involve investigating how predictive of the endpoint specific KEs or assay are and then weighting those pieces of evidence accordingly. Through the mechanistic framing that it provides, the AOP network described in this article can facilitate the grouping and comparison of NAMs and traditional assay data. Such investigations could aid in the development and implementation of IATAs and tier-based safety assessments, as well as enhanced confidence in their outputs.

This work highlights a large step forward in terms of DART AOP knowledge curation and DART hazard identification. In its current form, the AOP network is demonstrated to be a valuable hazard screening tool with the potential to provide a valuable resource for DART-based safety assessments. Such a network has many potential uses, and exploration of these uses will be greatly enhanced through industry-wide collaboration (using the AOP network to bring together cross-industry experts to share evidence and determine the best way forward for animal-free DART assessments).

CRediT authorship contribution statement

Alun Myden: Writing – original draft, Investigation, Conceptualization. **Alex Cayley:** Writing – review & editing, Supervision, Conceptualization. **Robert Davies:** Data curation. **Jade Jones:** Writing – review & editing, Investigation. **Steven Kane:** Writing – review & editing, Data curation. **Daniel Newman:** Visualization, Software. **Martin P. Payne:** Investigation. **Victor C. Ude:** Writing – review & editing, Investigation. **Jonathan D. Vessey:** Writing – review & editing, Software. **Emma White:** Writing – review & editing, Investigation, Data curation. **Adrian Fowkes:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

We would like to thank the following people: Dr Antonio Anax F.de Olivera for his support, encouragement and insightful discussions, Dr Takashi Yamada and Dr Akihiko Hirose (from the National Institute of Health Sciences of Japan) for their collaborative efforts which helped to contribute to the knowledge within the AOP network, and Dr Stephane Werner for his work developing the force-directed chemical space visualisation tool used to create Fig. 10.

References

[1] T. Collins, History and evolution of reproductive and developmental toxicology guidelines, Curr. Pharm. Des. 12 (2006) 1449–1465, https://doi.org/10.2174/138161206776389813.
[2] EMA/CHMP/ICH/544278/1998, ICH S5 (R3) guideline on reproductive toxicology: Detection of toxicity to reproduction for human pharmaceuticals, 2020. https://www.ema.europa.eu/en/ich-s5-r3-guideline-reproductive-toxicology-detection-toxicity-reproduction-human-pharmaceuticals.
[3] E. Stallman Brown, A. Jacobs, S. Fitzpatrick, Reproductive and developmental toxicity testing: from in vivo to in vitro, ALTEX. 29 (2012) 333–339, https://doi.org/10.14573/altex.2012.3.333.
[4] K.L. Chapman, H. Holzgrefe, L.E. Black, M. Brown, G. Chellman, C. Copeman, J. Couch, S. Creton, S. Gehen, A. Hoberman, L.B. Kinter, S. Madden, C. Mattis, H. A. Stemple, S. Wilson, Pharmaceutical toxicology: Designing studies to reduce animal use, while maximizing human translation, Regul. Toxicol. Pharmacol. 66 (2013) 88–103, https://doi.org/10.1016/j.yrtph.2013.03.001.
[5] K.C. Brannen, R.E. Chapin, A.C. Jacobs, M.L. Green, Alternative models of developmental and reproductive toxicity in pharmaceutical risk assessment and the 3Rs, ILAR J. 57 (2016) 144–156, https://doi.org/10.1093/ilar/ilw026.
[6] C. Rovida, F. Longo, R.R. Rabbit, How are reproductive toxicity and developmental toxicity addressed in REACH dossiers? ALTEX 28 (2011) 273–294, https://doi.org/10.14573/altex.2011.4.273.
[7] A.R. Scialli, The challenge of reproductive and developmental toxicology under REACH, Regul. Toxicol. Pharmacol. 51 (2008) 244–250, https://doi.org/10.1016/j.yrtph.2008.04.008.
[8] J. Knight, 4.2 million and counting… The animal toll for REACH systemic toxicity studies, ALTEX. (2023). doi:10.14573/altex.2303201.
[9] OECD, Test No. 414: Prenatal Developmental Toxicity Study, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 2018, 10.1787/9789264070820-en.
[10] R.C. Garg, W.M. Bracken, A.M. Hoberman, B. Enright, B. Tornesi, Reproductive and developmental safety evaluation of new pharmaceutical compounds, Reprod. Dev. Toxicol. (2022) 211–237, https://doi.org/10.1016/B978-0-323-89773-0.00012-6. Elsevier.
[11] J.A. Kramer, J.E. Sagartz, D.L. Morris, The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates, Nat. Rev. Drug Discov. 6 (2007) 636–649, https://doi.org/10.1038/nrd2378.
[12] C. Westmoreland, H.J. Bender, J.E. Doe, M.N. Jacobs, G.E.N. Kass, F. Madia, C. Mahony, I. Manou, G. Maxwell, P. Prieto, R. Roggeband, T. Sobanski, K. Schütte, A.P. Worth, Z. Zvonar, M.T.D. Cronin, Use of new approach methodologies (NAMs) in regulatory decisions for chemical safety: Report from an EPAA Deep Dive Workshop, Regul. Toxicol. Pharmacol. 135 (2022) 105261, https://doi.org/10.1016/j.yrtph.2022.105261.
[13] European Chemicals Agency, New Approach Methodologies in Regulatory Science Proceedings of a scientific workshop Helsinki, 19–20 April 2016, 2016. doi:10.2823/543644.
[14] A.O. Stucki, T.S. Barton-Maclaren, Y. Bhuller, J.E. Henriquez, T.R. Henry, C. Hirn, J. Miller-Holt, E.G. Nagy, M.M. Perron, D.E. Ratzlaff, T.J. Stedeford, A. J. Clippinger, Use of new approach methodologies (NAMs) to meet regulatory requirements for the assessment of industrial chemicals and pesticides for effects on human health, Front. Toxicol. 4 (2022), https://doi.org/10.3389/ftox.2022.964553.
[15] J.C. Achenbach, C. Leggiadro, S.A. Sperker, C. Woodland, L.D. Ellis, Comparison of the zebrafish embryo toxicity assay and the general and behavioral embryo toxicity assay as new approach methods for chemical screening, Toxics. 8 (2020) 126, https://doi.org/10.3390/toxics8040126.
[16] A.M.J. Weiner, I. Irijalba, M.P. Gallego, I. Ibarburu, L. Sainz, F. Goñi-de-Cerio, C. Quevedo, A. Muriana, Validation of a zebrafish developmental defects assay as a qualified alternative test for its regulatory use following the ICH S5(R3) guideline, Reprod. Toxicol. 123 (2024) 108513, https://doi.org/10.1016/j.reprotox.2023.108513.
[17] A. Weyrich, M. Joel, G. Lewin, T. Hofmann, M. Frericks, Review of the state of science and evaluation of currently available in silico prediction models for reproductive and developmental toxicity: A case study on pesticides, Birth Defects Res. 114 (2022) 812–842, https://doi.org/10.1002/bdr2.2062.
[18] C.A. Marchant, K.A. Briggs, A. Long, In silico tools for sharing data and knowledge on toxicity and metabolism: Derek for Windows, Meteor, and Vitic, Toxicol. Mech. Methods. 18 (2008) 177–187, https://doi.org/10.1080/15376510701857320.
[19] M. Marzo, S. Kulkarni, A. Manganaro, A. Roncaglioni, S. Wu, T.S. Barton-Maclaren, C. Lester, E. Benfenati, Integrating in silico models to enhance predictivity for developmental toxicity, Toxicology. 370 (2016) 127–137, https://doi.org/10.1016/j.tox.2016.09.015.
[20] D.J. Dix, K.A. Houck, M.T. Martin, A.M. Richard, R.W. Setzer, R.J. Kavlock, The ToxCast program for prioritizing toxicity testing of environmental chemicals, Toxicol. Sci. 95 (2007) 5–12, https://doi.org/10.1093/toxsci/kfl103.
[21] A.E.M. Seiler, H. Spielmann, The validated embryonic stem cell test to predict embryotoxicity in vitro, Nat. Protoc. 6 (2011) 961–978, https://doi.org/10.1038/nprot.2011.348.
[22] A. Myden, E. Hill, A. Fowkes, Using adverse outcome pathways to contextualise (Q) SAR predictions for reproductive toxicity – A case study with aromatase inhibition,

Reprod. Toxicol. 108 (2022) 43–55, https://doi.org/10.1016/j.reprotox.2022.01.004.
[23] G.T. Ankley, R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrrano, J.E. Tietge, D.L. Villeneuve, Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment, Environ. Toxicol. Chem. 29 (2010) 730–741, https://doi.org/10.1002/etc.34.
[24] N. Burden, F. Sewell, M.E. Andersen, A. Boobis, J.K. Chipman, M.T.D. Cronin, T. H. Hutchinson, I. Kimber, M. Whelan, Adverse Outcome Pathways can drive non-animal approaches for safety assessment, J. Appl. Toxicol. 35 (2015) 971–975, https://doi.org/10.1002/jat.3165.
[25] D. Knapen, L. Vergauwen, D.L. Villeneuve, G.T. Ankley, The potential of AOP networks for reproductive and developmental toxicity assay development, Reprod. Toxicol. 56 (2015) 52–55, https://doi.org/10.1016/j.reprotox.2015.04.003.
[26] S. Tanabe, M. Sachana, R. FitzGerald, Adverse Outcome Pathways in reproductive and developmental toxicology, Reprod. Dev. Toxicol. (2022) 63–72, https://doi.org/10.1016/B978-0-323-89773-0.00004-7. Elsevier.
[27] Society for Advancement of AOPs, AOP-Wiki, (n.d.). aopwiki.org (accessed December 12, 2023).
[28] A. Rolaki, F. Pistollato, S. Munn, A. Bal-Price, AOP 54: Inhibition of Na+/I-symporter (NIS) leads to learning and memory impairment, AOPWiki. (n.d.). https://aopwiki.org/aops/54 (accessed June 30, 2021).
[29] K. Crofton, M. Gilbert, K. Paul Friedman, B. Demeneix, M.S. Marty, R.T. Zoeller, AOP 42: Inhibition of Thyroperoxidase and Subsequent Adverse Neurodevelopmental Outcomes in Mammals, AOPWiki. (n.d.). https://aopwiki.org/aops/42 (accessed June 30, 2021).
[30] X. Cendoya, C. Quevedo, M. Ipiñazar, F.J. Planes, Computational approach for collection and prediction of molecular initiating events in developmental toxicity, Reprod. Toxicol. 94 (2020) 55–64, https://doi.org/10.1016/j.reprotox.2020.03.010.
[31] L. Truong, D.M. Reif, L. St Mary, M.C. Geier, H.D. Truong, R.L. Tanguay, Multidimensional in vivo hazard assessment using zebrafish, Toxicol. Sci. 137 (2014) 212–233, https://doi.org/10.1093/toxsci/kft235.
[32] S. Wu, J. Fisher, J. Naciff, M. Laufersweiler, C. Lester, G. Daston, K. Blackburn, Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants, Chem. Res. Toxicol. 26 (2013) 1840–1861, https://doi.org/10.1021/tx400226u.
[33] A. Myden, S.A. Stalford, A. Fowkes, E. White, A. Hirose, T. Yamada, Enhancing developmental and reproductive toxicity knowledge: A new AOP stemming from glutathione depletion, Curr. Res. Toxicol. 5 (2023) 100124, https://doi.org/10.1016/j.crtox.2023.100124.
[34] OECD, "Users" Handbook Supplement to the Guidance Document for Developing and Assessing Adverse Outcome Pathways. OECD Series on Adverse Outcome Pathways No. 1, OECD Publishing, Paris, 2018, 10.1787/5jlv1m9d1g32-en.
[35] T. Ball, C.G. Barber, A. Cayley, M.L. Chilton, R. Foster, A. Fowkes, C. Heghes, E. Hill, N. Hill, S. Kane, D.S. Macmillan, A. Myden, D. Newman, A. Polit, S. A. Stalford, J.D. Vessey, Beyond adverse outcome pathways: making toxicity predictions from event networks, SAR models, data and knowledge, Toxicol. Res. (Camb) 10 (2021) 102–122, https://doi.org/10.1093/toxres/tfaa099.
[36] U.S. EPA, ToxCast & Tox21 Summary Files from invitrodb_v3.2., (2015). https://www.epa.gov/chemical-research/exploring-toxcast-data.
[37] EPA, Tanguay et al. zebrafish activity score, (n.d.). https://comptox.epa.gov/dashboard/assay-endpoints/Tanguay_ZF_120hpf_ActivityScore (accessed December 2, 2020).
[38] M.T. Martin, R. Judson, U.S. EPA, ToxRefDB - Release user-friendly web-based tool for mining ToxRefDB, (2010). https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NCCT&dirEntryId=227139.
[39] J. Sun, N. Jeliazkova, V. Chupakin, J.-F. Golib-Dzib, O. Engkvist, L. Carlsson, J. Wegner, H. Ceulemans, I. Georgiev, V. Jeliazkov, N. Kochev, T.J. Ashby, H. Chen, ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics, J. Cheminform. 9 (2017) 17, https://doi.org/10.1186/s13321-017-0203-5.
[40] A. Gaulton, A. Hersey, M.L. Nowotka, A. Patricia Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibrian-Uhalte, M. Davies, N. Dedman, A. Karlsson, M.P. Magarinos, J.P. Overington, G. Papadatos, I. Smit, A.R. Leach, The ChEMBL database in 2017, Nucl. Acids Res. 45 (2017) D945–D954, https://doi.org/10.1093/nar/gkw1074.
[41] E. Genschow, H. Spielmann, G. Scholz, I. Pohl, A. Seiler, N. Clemann, S. Bremer, K. Becker, Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests, Altern. Lab. Anim. 32 (2004) 209–244, https://doi.org/10.1177/026119290403200305.
[42] H. Spielmann, E. Genschow, N.A. Brown, A.H. Piersma, A. Verhoef, M.Q. I. Spanjersberg, H. Huuskonen, F. Paillard, A. Seiler, Validation of the rat limb bud micromass test in the international ECVAM validation study on three in vitro embryotoxicity tests, Altern. Lab. Anim. 32 (2004) 245–274, https://doi.org/10.1177/026119290403200306.
[43] A.H. Piersma, E. Genschow, A. Verhoef, M.Q.I. Spanjersberg, N.A. Brown, M. Brady, A. Burns, N. Clemann, A. Seiler, H. Spielmann, Validation of the postimplantation rat whole-embryo culture test in the international ECVAM validation study on three in vitro embryotoxicity tests, Altern. Lab. Anim. 32 (2004) 275–307, https://doi.org/10.1177/026119290403200307.
[44] J.A. Paquette, S.W. Kumpf, R.D. Streck, J.J. Thomson, R.E. Chapin, D.B. Stedman, Assessment of the embryonic stem cell test and application and use in the pharmaceutical industry, Birth Defects Res. B. Dev. Reprod. Toxicol. 83 (2008) 104–111, https://doi.org/10.1002/bdrb.20148.
[45] N.C. Kleinstreuer, P.C. Ceger, D.G. Allen, J. Strickland, X. Chang, J.T. Hamm, W. M. Casey, A curated database of rodent uterotrophic bioactivity, Environ. Health Perspect. 124 (2016) 556–562, https://doi.org/10.1289/ehp.1510183.
[46] P. Browne, N.C. Kleinstreuer, P. Ceger, C. Deisenroth, N. Baker, K. Markey, R. S. Thomas, R.J. Judson, W. Casey, Development of a curated Hershberger database, Reprod. Toxicol. 81 (2018) 259–271, https://doi.org/10.1016/j.reprotox.2018.08.016.
[47] A.N. Cayley, R. Foster, E. Hill, S. Kane, G. Kocks, A. Myden, D. Newman, S. A. Stalford, J.D. Vessey, R. Zarei, A.A.F. Oliveira, Development of a network of carcinogenicity adverse outcome pathways and its employment as an evidence framework for safety assessment, ALTEX 40 (2022) 34–52, https://doi.org/10.14573/altex.2201311.
[48] OECD, Test No. 441: Hershberger Bioassay in Rats: A Short-term Screening Assay for (Anti)Androgenic Properties, OECD Publishing, Paris, 2009, 10.1787/9789264076334-en.
[49] OECD, Test No. 440: Uterotrophic Bioassay in Rodents: A short-term screening test for oestrogenic properties, OECD Publishing, Paris, 2007, 10.1787/9789264067417-en.
[50] M.R. Berthold, N. Cebron, F. Dill, T.R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, KNIME: The Konstanz Information Miner, in: Stud. Classif. Data Anal. Knowl. Organ. (GfKL 2007), Springer, 2007, pp. 319–326, https://link.springer.com/chapter/10.1007/978-3-540-78246-9_38.
[51] OECD, Test No. 426: Developmental Neurotoxicity Study, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 2007 doi: 10.1787/9789264067394-en.
[52] A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E.H. Bowler-Barnett, R. Britto, H. Bye-A-Jee, A. Cukura, P. Denny, T. Dogan, T. Ebenezer, J. Fan, P. Garmiri, L.J. da Costa Gonzales, E. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, V. Joshi, D. Jyothi, S. Kandasaamy, A. Lock, A. Luciani, M. Lugaric, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, A. Mishra, K. Moulang, A. Nightingale, S. Pundir, G. Qi, S. Raj, P. Raposo, D.L. Rice, R. Saidi, R. Santos, E. Speretta, J. Stephenson, P. Totoo, E. Turner, N. Tyagi, P. Vasudev, K. Warner, X. Watkins, R. Zaru, H. Zellner, A.J. Bridge, L. Aimo, G. Argoud-Puy, A.H. Auchincloss, K.B. Axelsen, P. Bansal, D. Baratin, T.M. Batista Neto, M.-C. Blatter, J.T. Bolleman, E. Boutet, L. Breuza, B.C. Gil, C. Casals-Casas, K. C. Echioukh, E. Coudert, B. Cuche, E. de Castro, A. Estreicher, M.L. Famiglietti, M. Feuermann, E. Gasteiger, P. Gaudet, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, C. Hulo, N. Hyka-Nouspikel, F. Jungo, A. Kerhornou, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, V. Muthukrishnan, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, S. Poux, M. Pozzato, M. Pruess, N. Redaschi, C. Rivoire, C.J.A. Sigrist, K. Sonesson, S. Sundaram, C.H. Wu, C.N. Arighi, L. Arminski, C. Chen, Y. Chen, H. Huang, K. Laiho, P. McGarvey, D.A. Natale, K. Ross, C.R. Vinayaka, Q. Wang, Y. Wang, J. Zhang, UniProt: the universal protein knowledgebase in 2023, Nucleic Acids Res. 51 (2023) D523–D531, https://doi.org/10.1093/nar/gkac1052.
[53] D.L. Villeneuve, M.M. Angrish, M.C. Fortin, I. Katsiadaki, M. Leonard, L. Margiotta-Casaluci, S. Munn, J.M. O'Brien, N.L. Pollesch, L.C. Smith, X. Zhang, D. Knapen, Adverse outcome pathway networks II: Network analytics, Environ. Toxicol. Chem. 37 (2018) 1734–1748, https://doi.org/10.1002/etc.4124.
[54] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.
[55] World Health Organisation, Organization, International Labour Organization, UN Environmental Programme, Section 4.5 Genotoxicity, in: World Health Organization & Food and Agriculture Organization of the United Nations (Ed.), Princ. Methods Risk Assess. Chem. Food, Second, 2020. https://www.who.int/publications/i/item/9789241572408.