



Review

# Computational Strategies for Assessing Adverse Outcome Pathways: Hepatic Steatosis as a Case Study

Rita Ortega-Vallbona <sup>1</sup>, Martina Palomino-Schätzlein <sup>1</sup>, Laia Tolosa <sup>2,3</sup>, Emilio Benfenati <sup>4</sup>,  
Gerhard F. Ecker <sup>5</sup>, Rafael Gozalbes <sup>1,6</sup> and Eva Serrano-Candelas <sup>1,\*</sup>

- <sup>1</sup> ProtoQSAR S.L., Calle Nicolás Copérnico 6, Parque Tecnológico de Valencia, 46980 Paterna, Spain; rortega@protoqsar.com (R.O.-V.); mpalomino@protoqsar.com (M.P.-S.); rgozalbes@protoqsar.com (R.G.)
- <sup>2</sup> Unidad de Hepatología Experimental, Instituto de Investigación Sanitaria La Fe (IIS La Fe), Av. Fernando Abril Martorell 106, 46026 Valencia, Spain; laia\_tolosa@iislafe.es
- <sup>3</sup> Biomedical Research Networking Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Instituto de Salud Carlos III, C/Monforte de Lemos, 28029 Madrid, Spain
- <sup>4</sup> Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy; emilio.benfenati@marionegri.it
- <sup>5</sup> Department of Pharmaceutical Sciences, University of Vienna, Josef-Holaubek Platz 2, 1090 Wien, Austria; gerhard.f.ecker@univie.ac.at
- <sup>6</sup> MolDrug AI Systems S.L., Olimpia Arozena Torres 45, 46108 Valencia, Spain
- \* Correspondence: eserrano@protoqsar.com; Tel.: +34-962021811

**Abstract:** The evolving landscape of chemical risk assessment is increasingly focused on developing tiered, mechanistically driven approaches that avoid the use of animal experiments. In this context, adverse outcome pathways have gained importance for evaluating various types of chemical-induced toxicity. Using hepatic steatosis as a case study, this review explores the use of diverse computational techniques, such as structure–activity relationship models, quantitative structure–activity relationship models, read-across methods, omics data analysis, and structure-based approaches to fill data gaps within adverse outcome pathway networks. Emphasizing the regulatory acceptance of each technique, we examine how these methodologies can be integrated to provide a comprehensive understanding of chemical toxicity. This review highlights the transformative impact of *in silico* techniques in toxicology, proposing guidelines for their application in evidence gathering for developing and filling data gaps in adverse outcome pathway networks. These guidelines can be applied to other cases, advancing the field of toxicological risk assessment.

**Keywords:** adverse outcome pathway; hepatic steatosis; computational toxicity; new generation risk assessment; molecular mechanisms



**Citation:** Ortega-Vallbona, R.; Palomino-Schätzlein, M.; Tolosa, L.; Benfenati, E.; Ecker, G.F.; Gozalbes, R.; Serrano-Candelas, E. Computational Strategies for Assessing Adverse Outcome Pathways: Hepatic Steatosis as a Case Study. *Int. J. Mol. Sci.* **2024**, *25*, 11154. <https://doi.org/10.3390/ijms252011154>

Academic Editor: Giulio Vistoli

Received: 20 September 2024

Revised: 10 October 2024

Accepted: 11 October 2024

Published: 17 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The evolving landscape of risk assessment is leading to the development of hypothesis-driven, tiered, and iterative approaches to prioritize chemicals. These new early toxicity detection strategies, incorporating New Approach Methodologies (NAMs), offer an ethical alternative to traditional *in vivo* experimentation, which faces challenges in ethics, time, and cost [1,2]. NAMs include tools such as *in silico* models, artificial intelligence, machine learning (ML) systems, toxicogenomic tests, and read-across approaches [1,3,4]. These human-relevant tools can transform safety assessments by offering insights into key biological pathways and targets [1,4,5]. NAMs mimic human biology and provide mechanistic information on chemical toxicity [1], which can be organized into Adverse Outcome Pathways (AOPs) and AOP networks, establishing relationships between the interaction of a chemical with a biological system and the resulting adverse outcome (AO) [5–7].

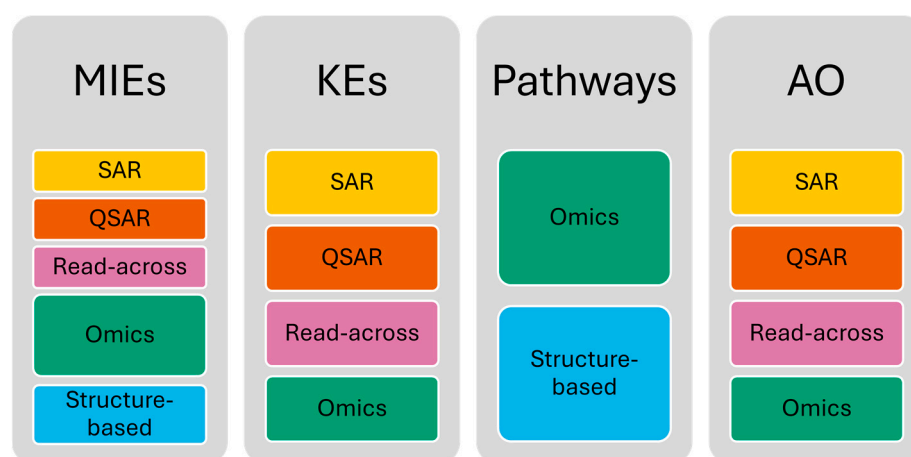
AOPs go beyond the older mode-of-action concept by considering effects at varying levels of biological organization [2,8,9]. A molecular initiating event (MIE) is always the first step of an AOP, where the chemical modulates with a specific molecule in the biological

system [7]. The AO is the apical toxicological endpoint, referring to an acute or local adverse effect or systemic outcome. The pathway from an MIE to an AO consists of a series of steps involving different biological elements that can lead to adverse health effects. Some of those steps are considered key events (KEs), which are measurable changes in a biological state that are essential but not sufficient for the AO to occur [8,10,11] and are connected by KE relationships. A KE characterizes the system's condition at a specific stage, whereas a KE relationship scientifically explains how a system transitions from one upstream KE to another downstream KE. It establishes a connection between the two events, facilitating the inference or extrapolation of the downstream KE's state from the known, measured, or predicted state of the upstream KE [5,11]. In addition, a quantitative AOP involves computational models that quantitatively characterize the relationships between MIEs and KEs, as well as among KEs and the AO [11]. A quantitative AOP provides quantitative, dose-responsive, and time-course predictions [2,12,13].

An AOP is not a test method, computational model, or risk assessment but rather a structured description of collective knowledge in toxicology supported by scientific evidence. AOPs are designed to aid regulatory decision making using mechanistic data rather than solely outcome data from traditional whole-organism toxicity tests [1,8,14]. However, it is important to note that AOPs themselves do not include exposure evaluation, but they can be used in conjunction with exposure data to inform risk assessment processes [14]. AOPs facilitate the prioritization of chemicals for further testing through screening processes to determine if they trigger any biological response. The substances can be ranked according to their toxicological risk to guide detailed testing of higher-risk compounds. In the same way, AOPs can be used to label chemicals based on their potency to activate an AO [1,5].

To improve chemical safety assessment through AOPs, it is necessary to gather and integrate biological process knowledge into a cohesive network [1]. AOPs can be combined into networks that describe common KEs activated by different MIEs to provide a complete view of the mechanisms leading to pathology [10,11]. Developing an AOP requires collecting information on each building block from the literature, in vivo or in vitro assays, or computational methods and associating evidence and testing methods [7]. AOPs are "living documents" and should be continually refined as new data become available [11,15]. Despite ongoing efforts to link test methods and data with AOPs, current initiatives have limitations, relying on publicly available knowledge without sufficient curation for fidelity. These initiatives often overlook evidence from in silico predictive systems and struggle to associate diverse data formats from various NAMs [1].

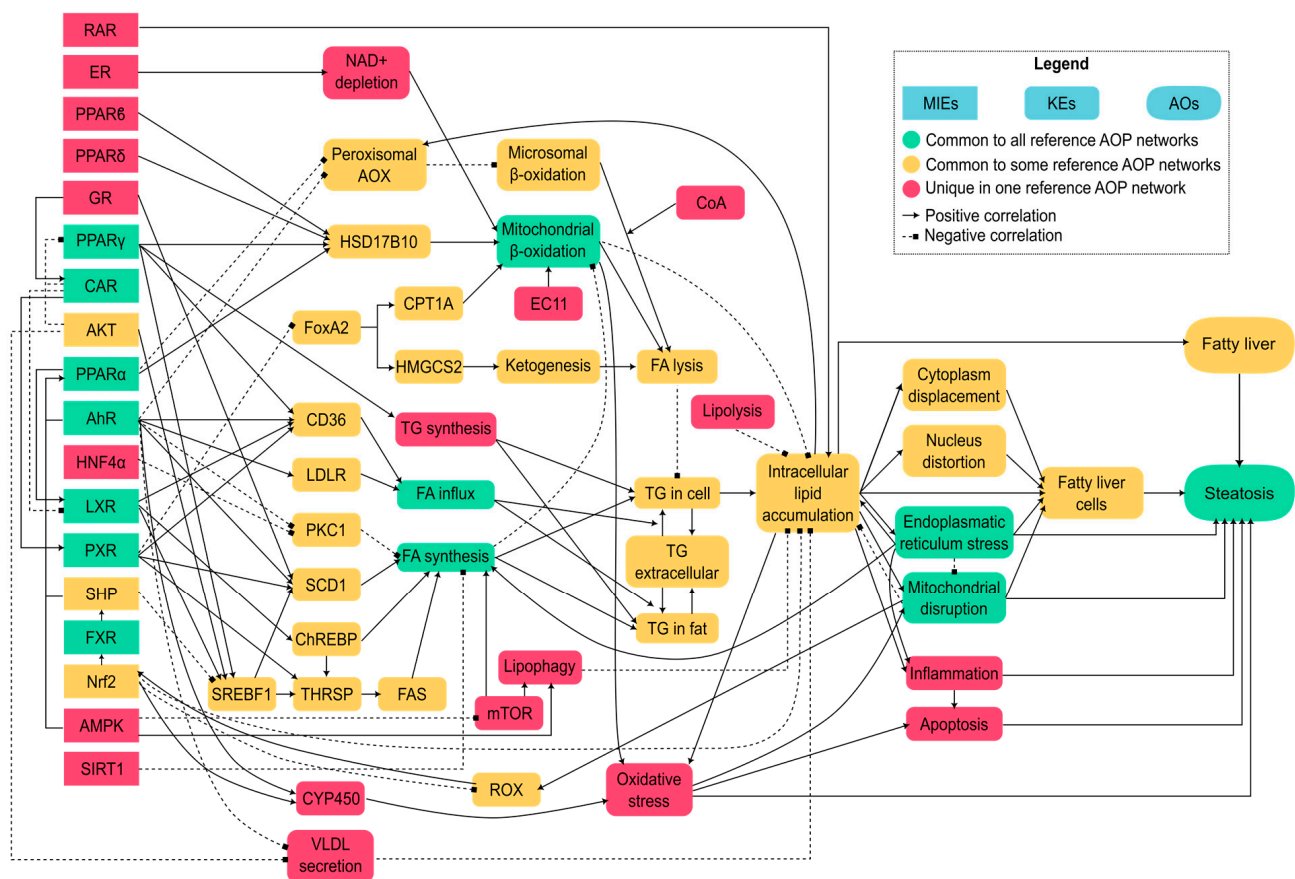
Diverse computational methods can be employed to identify AOP components, predict the effects of chemicals on each node at the molecular level, and fill knowledge gaps within an AOP network. Figure 1 shows schematically which computational methods can be applied in each case.



**Figure 1.** Representation of potential use of computational methods in AOP elements.

Fatty liver disease or hepatic steatosis is defined as the accumulation of fatty acids (FA) in the liver beyond 5% of its wet weight due to an abnormality in the synthesis, transportation, elimination of lipids, or a combination of those factors [16]. Steatosis can be defined as increased fat accumulation in the liver without hepatocellular necrosis and no, or minimal, inflammation. Although it can be a mild condition with a good short-term prognosis, it can lead to steatohepatitis lesions characterized by hepatocellular injury and inflammation [17–19]. Liver steatosis exhibits two histological patterns: macrovesicular steatosis, marked by large vacuoles causing peripheral nucleus displacement, and microvesicular steatosis, characterized by numerous smaller vacuoles without nucleus displacement [19,20]. Several causes can lead to hepatic steatosis, such as obesity, insulin resistance, and exposure to alcohol and chemicals, including pharmaceuticals [19,21,22].

Drug-induced steatosis has the potential to directly cause metabolic-associated steatotic liver disease (MASLD, formerly known as non-alcoholic fatty liver disease). Even though drugs are not typically a primary cause of MASLD, it has been demonstrated that they might underlie, mimic, or exacerbate MASLD [23]. It is known that certain drugs commonly used in clinical practice, such as amiodarone, tamoxifen, or certain chemotherapeutic agents, may contribute to hepatic steatosis. Drug-induced steatosis is a well-studied condition, and the mechanisms behind it have been extensively described in various AOPs. These AOPs have been combined into a comprehensive AOP network, which is depicted in Figure 2 [11,24–26].



**Figure 2.** AOP network for liver steatosis. It integrates MIEs and KE collected from four reference AOP networks [11,24–26]. The network depicts the progression from MIEs through KEs to the AO of liver steatosis. The nodes are color-coded to indicate their presence in the reference AOP networks: green for nodes common to all reference AOP networks, yellow for nodes common to some reference AOP networks, and red for nodes unique to one reference AOP network. Solid arrows represent positive correlations, while dashed arrows represent negative correlations. The names and abbreviations used in this figure are available in the Supplementary Table S1.

FA accumulation, which is a major feature of this condition, is caused by an increase in de novo FA synthesis and a decrease in  $\beta$ -oxidation, both of which indicate a disruption in FA metabolism, as well as an increase in FA influx [22,27].

The primary MIEs of this AOP are believed to result from the interaction between chemicals and various nuclear receptors. The activation of the liver X receptor (LXR), a transcriptional regulator of genes linked to cholesterol uptake, transport, and excretion, is an integral part of the AOP [20,28]. LXR activation upregulates four genes involved in de novo FA synthesis and increases hepatic FA uptake via the CD36 (Cluster of Differentiation 36, also known as Fatty Acid Translocase (FAT)) [19,20,22].

Other receptors have been described as causing the accumulation of triglycerides by sharing one or both mechanistic pathways described for LXR. These receptors include the aryl hydrocarbon receptor (AHR), the constitutive androstane receptor (CAR), the farnesoid X receptor (FXR), the glucocorticoid receptor (GR), the peroxisome proliferator-activated receptors (PPARs), and the Pregnane X receptor (PXR) [19,22]. CAR, FXR, LXR, PXR, and PPARs share some characteristics: they are nuclear receptors of the subfamily 1 and heterodimerize with retinoid X receptor (RXR). It has been reported that RXR competition could play a functional role in lipid metabolism [22,28,29].

Mitochondria play a pivotal role in maintaining lipid homeostasis, and disruptions in their function can lead to the accumulation of lipids within cells. Some hepatotoxicants induce mitochondrial dysfunction, inhibiting mitochondrial  $\beta$ -oxidation and resulting in the intracellular accumulation of FA. Activation of PPAR $\alpha$  by FA and fibrate drugs regulates FA oxidation pathways, while antagonistic binding inhibits fatty acyl-CoA oxidase, suppressing microsomal  $\beta$ -oxidation and promoting triglyceride accumulation [19,22]. The estrogen receptor (ER) also regulates appetite and adipose biology, which has a close relationship with FA uptake by the hepatocytes [22].

The ASPIS cluster, created in the EU Horizon 2020 Research and Innovation programme, strives to advance the safety assessment of chemicals while minimizing the reliance on animal testing (<https://aspis-cluster.eu/>, accessed on 20 September 2024). ASPIS is a collaboration between the projects ONTOX (<https://ontox-project.eu/>, accessed on 20 September 2024), PrecisionTox (<https://precisiontox.org/>, accessed on 20 September 2024), and RISK-HUNT3R (<https://www.risk-hunt3r.eu/>, accessed on 20 September 2024). The ASPIS cluster creates a network of NAMs using case studies to predict various AOs, including drug-induced hepatic steatosis.

This paper reviews the use of computational techniques to fill in the missing data gaps in the AOP network for drug-induced hepatic steatosis. The results of this study provide valuable insights into the potential of these techniques to improve our understanding of AOPs and their role in risk assessment. This could have significant implications in the toxicology field by enabling more accurate predictions of chemical-induced toxicity and the development of targeted risk management strategies. We propose establishing guidelines for using in silico strategies to gather the weight of evidence, which can be applied to other cases where an AOP network has been developed.

## 2. Computational Methods

### 2.1. Automated SAR

The structure–activity relationship (SAR) is a technique based on the connection between the chemical structure of a molecule and its physicochemical, biological, or toxicological activity. The main idea behind the SAR is that certain structural features determine significant changes in its biological effects. The molecular substructures responsible for the toxicological effects of a compound are considered “structural alerts” (SAs) [30,31]. SAs are typically selected based on scientific publications, expert knowledge, and understanding of structure–activity relationships, toxicological pathways, and metabolism [32,33]. Alternatively, various statistical pattern recognition techniques can automatically mine SAs from a dataset of molecules with clear structural information and associated bioactivity values [32,34].

For automatic SA generation (automatized SAR), substructures can be identified using different approaches. Table 1 shows a range of algorithms for the automatic generation of SAs that use different approaches, such as utilizing predetermined fragments, embedding molecules into graphs, or employing sets of fingerprints [30,35]. Some researchers have used strategies based on established fingerprints, like PubChem's 881-bit fingerprint indicating unique substructures [36] and Klekota's fingerprint encoding 4860 substructures [37].

**Table 1.** Programs for automatic SA generation. All links were accessed on 20 September 2024.

Approach	Program	URL	Reference
Fingerprints	Bioalerts	<a href="https://github.com/isidro/bioalerts">https://github.com/isidro/bioalerts</a>	[38]
Fragments	CASE (Computer-Automated Structure Evaluation), MultiCASE	<a href="https://multicase.com/case-ultra">https://multicase.com/case-ultra</a>	[39,40]
	PASS (Prediction of Activity Spectra for Substances)	<a href="https://genexplain.com/pass/">https://genexplain.com/pass/</a>	[41]
	SARpy	<a href="https://sarp.py.sourceforge.net/readme.html">https://sarp.py.sourceforge.net/readme.html</a>	[42]
Graphs	Gaston		[43]
	MoSS		[44]

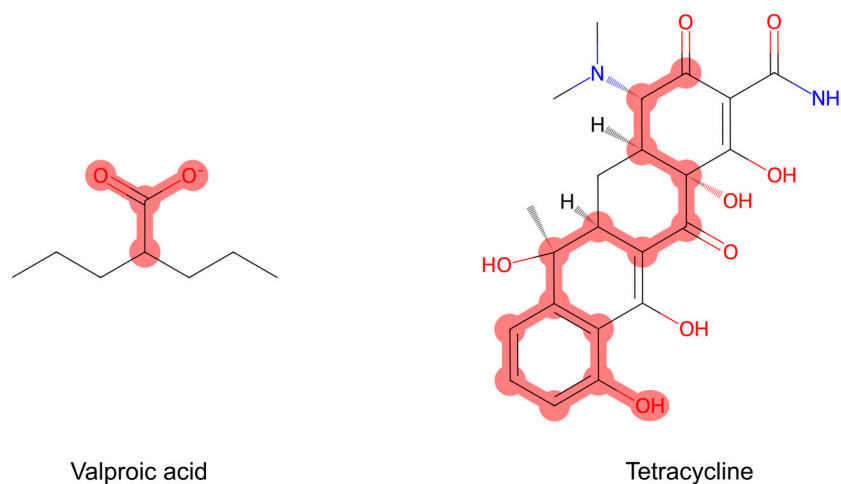
Once the set of SAs has been established, statistical metrics such as precision, enrichment factor, and p-value are used to assess their performance [30]. Additionally, as is characteristic in computational models based on a chemical structure, SAR models have an applicability domain (AD) [45] that defines the chemical space in which the model's predictions are reliable, and different methods can be applied to estimate it, including leveraging the range of descriptors, calculating leverage values, defining descriptors subspaces, or establishing confidence indices based on nearest neighbours. These methods have been extensively discussed elsewhere [46,47].

An essential advantage of SAR analysis lies in the rapid computational detection of these rules in a new set of molecules [33]. Additionally, SAR analysis enables mechanistic interpretation, associating SAs with specific biochemical mechanisms such as enzyme activation or ion channel opening [48,49]. This association helps to identify chemicals sharing a common mechanism of action [31]. The use of molecular fragments or fingerprints as SAs contributes to visualizing the SAR in terms of actual chemical structure. A simple 2D representation of a molecule with the relevant SAs highlighted (Figure 3) makes it easier for the user to understand complex chemical interactions. This facilitates the interpretation and enhances accessibility [50].

Nevertheless, SAs cannot give quantitative results; they can only classify compounds into discrete groups. To cope with this issue, QSARpy has been introduced, associating with the fragment a quantitative value related to the potency of the effect [51]. On top of that, they often give inconclusive results for negative predictions, as the absence of an SA in a compound is not enough proof to classify it as negative [17,31]. Therefore, they are a weak option when trying to generate data for constructing a quantitative AOP. Furthermore, expertise in cheminformatics and toxicology is required to create SAs and understand their underlying mechanisms [17], which can make developing and updating SAR models time-consuming [31].

To illustrate the application of SAR analysis in understanding hepatic steatosis, this section explores examples where researchers have worked to relate SAs with different points of the steatosis AOP. Mellor and colleagues characterized 214 SAs associated with binding to 10 nuclear receptors recognized as MIEs in steatosis AOPs: AHR, CAR, ER, FXR, GR, LXR, PPAR, PXR, retinoic acid receptor (RAR), and RXR [52]. Information on agonists for these receptors was gathered from ChEMBL and PDB. First, chemical descriptors were calculated for the ligands to define the chemical space covered by them. Next, structural features associated with the protein–ligand interactions were identified and encoded into SMARTS strings, setting the group of SAs that would be applied to an *in silico* workflow.

For most nuclear receptors, the substructural features generally follow a consistent pattern of examining key scaffolding structures (e.g., ring structures) and subsequent screening for essential functional groups. Exceptions include RXR and PPAR, which have specific exclusion rules, as well as GR, AR, ER, and PXR, which implement high molecular weight filters to accommodate larger ligands with distinct receptor binding patterns compared to those with lower molecular weights.



**Figure 3.** Representation of two steatotic compounds, valproic acid and tetracycline, showing two structural alerts for PPAR binding from Mellor et al., 2016 [52]. The carboxylic acid group (SMARTS: C(=O)O) is highlighted in valproic acid. A retinoid substructure is highlighted in tetracycline (SMARTS: [#6]~\*~[#6]~\*~[#6]~\*~[#6]~\*~a1aa([O,Cl,F,I,Br,N\*])aaa1).

Although not focused solely on steatosis, a publication by Ivanov and colleagues demonstrated the use of PASS software (<https://www.way2drug.com/>, accessed on 20 September 2024) to identify protein targets related to drug-induced liver injury (DILI) [53]. They analyzed a set of 699 drugs classified into three DILI severity classes (no DILI, moderate, and severe). Using the PASS Targets version, they predicted interactions of these drugs with a total of 1534 human proteins. Among these, they detected 61 proteins associated with DILI. Notably, out of these 61 proteins, 10 were found to be related to lipid disorders, which involve a pathological mechanism closely linked to steatosis.

Some of the proteins related to the lipid disorders they identified are particularly interesting: AHR is an MIE for the steatosis AOP; fructose-1,6-bisphosphatase takes part in the gluconeogenesis pathway, which, when it is excessive, can promote de novo lipid synthesis, leading to steatosis [54]; dihydroorotate dehydrogenase is involved in the de novo pyrimidine biosynthesis pathway and is also coupled to the mitochondrial respiratory chain [55]; some other proteins identified are related to MASLD—and not exclusively to drug-induced steatosis. This study is an excellent example of how SAR models are useful for identifying SAs and characterizing possible MIEs for steatosis.

Mitochondrial dysfunction is one of the KEs in the AOP network that has been most largely studied. In a book chapter published in 2018, Enoch, Mellor, and Nelms compiled a list of 22 SAs detected on different datasets of chemicals classified as active or inactive for mitochondrial toxicity [56]. According to their findings, these SAs primarily address two specific biological mechanisms contributing to mitochondrial dysfunction: the uncoupling of oxidative phosphorylation and the inhibition of the electron transport chain. The inhibition of the electron transport chain occurs when chemicals act as alternate electron acceptors, facilitating direct electron transport from complex I to complex IV (or oxygen). These chemicals, exemplified by the 2-nitroaniline moiety, form SA “pairs” where either the oxidized or reduced form can cause mitochondrial toxicity. A similar rationale applies to SAs for hydroxynaphthoquinones/naphthoquinones and hydroquinones/quinones. The uncoupling of oxidative phosphorylation is caused by compounds transporting hydro-

gen ions across the inner mitochondrial membrane into the mitochondrial matrix. The mechanistic chemistry behind these alerts stems from their role as weak acids in the intermembrane space. In this process, the deprotonated forms scavenge free protons, migrate into the mitochondrial matrix, and undergo proton dissociation in the alkaline environment. This cycle, facilitated by a pH gradient between the intermembrane space (pH  $\approx$  7) and the mitochondrial matrix (pH  $\approx$  8), increases oxygen consumption and heat production and decreases the electrochemical gradient and ATP production. The 13 SAs in this category exhibit moieties capable of resonance stabilization, which is crucial for proton cycling.

In a follow-up study two years later, Hemmerich and colleagues compiled a dataset of 5761 compounds from various sources, examining different endpoints related to mitochondrial toxicity [57]. By utilizing fragmentation techniques and alerts proposed by previous studies [58,59], they identified 17 highly specific alerts that could provide a further understanding of the potential mode of action. These alerts were combined with a QSAR predictive model, which was then used to screen the DrugBank dataset, which consists of 2278 approved or withdrawn drugs. A total of 52 compounds were flagged with the identified alerts, and further investigation revealed that there is experimental evidence for 6 of those molecules: amrubicin, tolvaptan, epirubicin, thiothixene, teriflunomide, and ponatinib.

Shifting the focus to another KE, namely, the accumulation of FAs, we have recently developed a SAR model using a reduced set of carboxylic acids. The aim was to identify substructures associated with a steatotic response [60]. We found two relevant substructures presenting a carboxyl moiety and an aliphatic chain comprising a minimum of three carbon atoms. While the presence of the carboxyl group was expected, given its ubiquity in the dataset, the distinguishing factor arose from the aliphatic chain. This characteristic could differentiate steatotic molecules from their non-steatotic counterparts, which typically exhibit a ring in their structure.

In 2021, Jain and colleagues analysed enriched chemical substructures in the steatosis-positive class within their dataset [21]. They identified three significantly enriched scaffolds and utilized the ChemoTyper program [61] to pinpoint enriched structural fragments in the positive class. Of note, eight of the substructural patterns they discovered contained halogens, which is coherent with the fact that halo-alkanes and halo-alkenes can potentially induce toxicity and cancer based on their ability to form a stable carbon-centred radical. As the number of halide groups increases, so does the likelihood of radical formation due to the growing electron-withdrawing capacity from the carbon centre. Reactive intermediates can bind to macromolecules and disrupt lipid metabolism, ultimately leading to steatosis.

Focusing on the prediction of general hepatotoxicity, in 2010, a group of researchers performed extensive literature research to identify SAs associated with DILI [62]. They identified 22 SAs through visual analysis and the consideration of therapeutic endpoints after collecting data on more than 1200 chemicals. The researchers mentioned tetracyclines as an example of steatogenic compounds. Tetracyclines prevent the movement of triglycerides from the liver by inhibiting protein synthesis and causing mitochondrial damage. They have been widely demonstrated in both humans and in vitro models to produce steatosis [63].

A team of researchers employed SARpy in 2016 to automate the extraction of SAs to characterize DILI [17]. They gathered information from several human datasets to compile their dataset comprising 950 compounds labelled either hepatotoxic or non-hepatotoxic. Using the similarity index, the researchers utilized a manual clustering process within the VEGA platform to identify SAs. This process led to the identification of 13 SAs, 11 of which were hepatotoxic. They also utilized SARpy to automatically extract 75 SAs, 40 of which were related to hepatotoxicity. While the authors were able to provide a mechanistic explanation for some of the SAs they identified, they were unfortunately unable to directly relate any of the SAs to steatosis.

In conclusion, SAR analysis facilitates the identification of SAs, and, with the input of experts' knowledge, it enables the explanation of the underlying biological mechanisms

related to the interaction of compounds with specific substructures and the prediction of the potential implication of specific proteins or enzymatic cascades in the occurrence of an AO.

## 2.2. QSAR

Quantitative structure–activity relationships (QSARs) are statistical models that explore the correlation between molecular descriptors encoding the intrinsic nature of the chemicals studied and a determined property or biological activity [64]. These models, often developed through ML algorithms, rely on a clear structural definition of molecules associated with the property or activity of interest [65,66], as happens with a SAR analysis. However, unlike SAR models, which are typically qualitative, QSAR models can provide both qualitative classifications (e.g., toxic vs. non-toxic) and quantitative predictions (e.g., IC50 values, binding affinities, etc.) [67].

The process of constructing a QSAR model follows a systematic workflow comprising several key steps, as illustrated in Figure 4: data gathering and curation to ensure that the available information (chemical structures and property of study) is suitable for modelling; the calculation of molecular descriptors; feature selection to identify the most relevant descriptors; model training using ML algorithms; validation to ensure its predictive performance and reliability; and determination of the model's AD [47,68]. The quality of a QSAR model depends on factors such as the quality of input data, choice of descriptors, and statistical methods for modelling and validation [69].

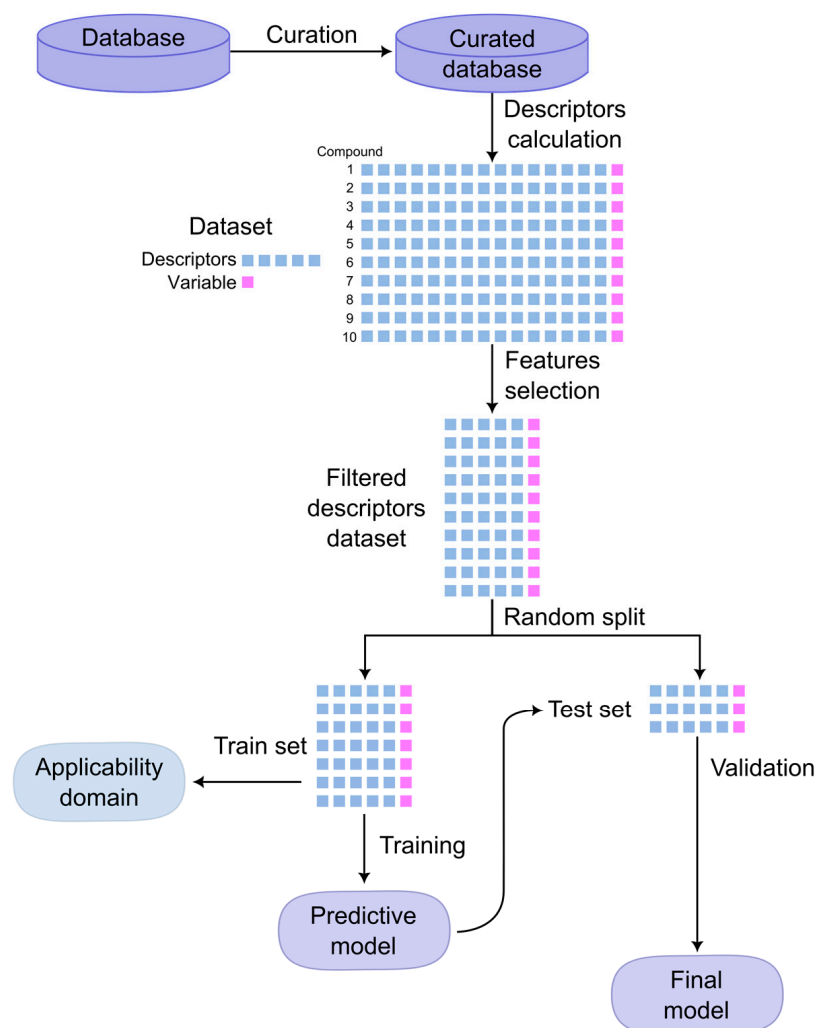


Figure 4. QSAR model generation workflow.



In QSAR modelling, several types of molecular descriptors exist, such as physicochemical (e.g., molecular weight, solubility, etc.); graph-theoretical descriptors, which involve counts of atoms and fragments; mathematical descriptors that contribute complexity metrics; fragment-based descriptors that use predefined fragments, chemical groups, or molecular fingerprints; and 3D descriptors that consider volumetric aspects and molecular shape [33,66,70]. These descriptors accurately represent molecular structure or properties and are crucial in establishing a correlation between biological activity or physicochemical properties in QSAR models. In recent years, biology-related fingerprints were introduced as descriptors for QSAR models [71], such as, e.g., compound/pathway interaction fingerprints, as outlined in Füzi et al. [72].

In the last two decades, ML techniques have revolutionized QSAR modelling, enabling more accurate predictions of chemical properties and activities. Traditional QSAR methods have relied on congeneric series of compounds and predefined molecular descriptors, which limited their ability to predict complex biological phenomena [73]. With the advent of ML, QSAR modelling has broadened its scope [64,73]. Techniques such as random forest and Support Vector Machine have become popular for their robust predictive capabilities.

More recently, deep learning, particularly deep neural networks, has shown superior performance in QSAR applications. Deep neural networks can model chemical properties directly from molecular structures, like SMILES strings, without relying on pre-extracted descriptors [67,70]. This data-driven approach departs from traditional ML, offering greater flexibility. Deep neural networks can also perform multitask learning, which is especially beneficial when biological endpoints share a similar mode of action [64,73].

However, deep neural networks face challenges, including high computational demands, interpretability issues, and the need for large datasets. In some cases, simpler models, such as single-layer neural networks, may perform comparably to traditional ML methods, especially when data scarcity or experimental errors limit accuracy [64]. Despite these challenges, ongoing advancements in deep learning are expected to enhance QSAR model accuracy as more sophisticated architectures and larger datasets become available [73].

Ensuring the reliability and predictive power of QSAR models requires robust validation strategies. The first step is the evaluation of the statistical parameters (e.g., coefficient of determination, Matthews correlation coefficient, accuracy, sensitivity, and specificity). Typically, the dataset is partitioned into training and test sets, with approximately 25–35% of molecules allocated to the test set, which is a significant proportion of the total dataset. Model performance can be further assessed using leave-one-out cross-validation and k-fold cross-validation, which offer insights into goodness-of-fit and robustness [69]. On the other hand, external validation evaluates the model's predictive ability on independent datasets not used during model building, demonstrating its capability to predict outcomes for new compounds [68]. To avoid chance correlations, overtraining, or weak reproducibility, additional strategies, such as using additional independent datasets and the y-randomization test, are implemented [66]. Moreover, the model's AD must be established based on the training set or the model's characteristics, as described elsewhere [46,47]. Together, these validation steps play a crucial role in establishing the credibility and applicability of QSAR models in predicting molecular activities.

Once a QSAR model has been developed and validated, it can be utilized to predict the biological activities or properties of a new molecule whose chemical structure is known, and which enters the AD. To have a reliable SAR or QSAR prediction, we must ensure that the molecule is within the AD of the model and that the probability of the prediction is acceptable for those models where the probability can be measured. In the desirable scenario, the prediction unequivocally meets these two criteria.

On the other side, several scenarios can give unreliable predictions:

- For those molecules that fall outside the AD, the prediction cannot be considered.
- For those chemicals that are located near the boundaries of the AD, the prediction will be associated with higher uncertainty.

- For models where the relationship to the AD cannot be unequivocally established—such as those based on physicochemical descriptors, where the reliability of each input is challenging to be determined—the analysis should focus on non-experimental descriptors.
- For those predictions with a probability between 0.3 and 0.7, the results of the model will be considered unreliable or inaccurate.

In this context, models built from large training sets and broader chemical space will be more capable of generating predictions for a wider variety of chemical structures than smaller, more structurally and mechanistically homogeneous models [74–76].

QSAR models are highly efficient and can be easily automated, enabling the evaluation of a large number of molecules in a short amount of time [65,66,77]. These models can predict not only categorical outcomes but also quantitative values, making them highly versatile tools in cheminformatics.

While QSAR models offer high predictive capabilities, they also present inherent limitations. The main challenge lies in mechanistic interpretation due to the non-intuitive nature of molecular descriptors [49,70]. Additionally, the relationship between bioactivity and descriptors can be challenging to comprehend, lacking clear intelligibility.

QSAR models can predict biological activities, enabling them to predict MIEs, KEs, and AOs. Several studies have addressed the QSAR modelling of different nuclear receptors that are identified as MIEs for steatosis [78–81]. Additionally, in 2018, Gadaleta and colleagues developed a set of nine QSAR models for *in vitro* high-throughput screening assays related to MIEs in the AOPs leading to liver steatosis: up- and downregulation of PXR, AHR, and LXR, as well as upregulation of PPAR $\alpha$ , PPAR $\gamma$ , and NRF2 (nuclear factor (erythroid-derived 2) like 2) [25]. They used data provided by the ToxCast program [82] and employed strategies to address the issue of unbalanced datasets.

On the other hand, several publications have shown the development of QSAR models to predict mitochondrial toxicity as an essential KE in steatosis. For example, Tang and co-authors used Tox21 data [83] to generate discriminant QSAR models for mitochondrial toxicity [84]. The authors found that a consensus model that combines the results of individual models using different types of molecular fingerprints as descriptors can improve the accuracy of the predictions. Substructure analysis revealed that phenol, carboxylic acid, nitro, and arylchloride are informative substructures for mitochondrial toxicity discrimination. Furthermore, researchers at Bayer published a paper on developing a suite of machine-learning models to predict mitochondrial toxicity, leveraging both mechanistic and cell-painting data [85]. They collected mechanistic assay results for membrane potential, respiratory chain, and mitochondrial function from diverse sources. They built individual mechanistic models using continuous and data-driven molecular descriptors [67] and, in parallel, developed models trained on morphological features identified in cell painting assays, independently or in conjunction with the molecular descriptors, and notably, including cell painting profiles demonstrated a marked enhancement in the predictive performance for mitochondrial toxicity.

In a different matter, we previously developed two QSAR models to predict FA and phospholipid accumulation biomarkers, which can aid in differentiating between steatosis and phospholipidosis [60]. To build these models, we utilized data from high-content screening *in vitro* assays conducted on HepG2 cells exposed to a set of structurally similar carboxylic acids. Our analysis revealed that changes in chain length, flexibility, and molecular weight were important factors linked to FA accumulation.

Additionally, there have been attempts to develop QSAR models to predict hepatic steatosis as the primary endpoint. In 2020, researchers working on the EuroMix project (<https://www.euromixproject.eu/>, accessed on 20 September 2024) attempted to predict steatosis using QSAR models for nuclear receptors related to steatosis. However, these models did not perform as well as expected. Consequently, they developed a QSAR model to predict steatosis using a dataset of 207 pharmaceuticals and pesticides identified from *in vivo* assays on rodents [86]. The model's descriptors included average molecular

weight and SlogP, which is a hybrid atomic calculation of the partition coefficient between n-octanol and water.

That same year, another research study focused on developing binary classification models for four mechanisms that lead to drug-induced liver injuries, including steatosis [87]. The study used data on 257 drugs collected from DILIrank [88] and PharmaPendium [89] datasets, as well as 318 drug metabolites obtained from the ADME database [90].

The study found that including drug metabolite sets improved the performance of the models, highlighting the importance of drug metabolism in the prediction of hepatotoxicity. The researchers utilized 2D fingerprints to identify significant structural characteristics more commonly found in the DILI-positive class. They compared the positive and negative sets and the drug and drug metabolite datasets to study the metabolic activation of compounds. They revealed that moieties such as phenyl and hydroxyl attached to aromatic carbons were more frequently observed in positive drug metabolites than in positive drugs. Conversely, carbon–nitrogen bonds were observed to be less frequent in positive drug metabolites than in positive drugs. The different bioactivation and detoxification processes in hepatocytes can cause these differences. However, these structural differences suggest that it is incorrect to assume that metabolites derived from a steatotic drug would necessarily be steatotic themselves. The metabolic transformations in the liver can significantly alter the toxicity profile of the metabolites, meaning that each metabolite must be individually assessed for its hepatotoxic potential.

A year later, Jain and coworkers explored the development of QSAR models for predicting hepatic steatosis using different combinations of chemical descriptors and predictions of transporter interactions and nuclear receptor activation/deactivation [21]. They used an unbalanced database of *in vivo* studies on rodents with repeated oral exposure, consisting of 120 active and 921 inactive compounds. Consequently, they explored the use of unbalanced learning techniques like bagging with stratified under-sampling and confidence predictors to train their models. While both meta-classifiers proved beneficial, the study underscores the specific utility of confidence predictors, emphasizing their potential relevance to predictive toxicology. Additionally, despite statistically insignificant performance differences, the researchers recommend further inspection of the role of transporters in hepatic steatosis and advocate for their inclusion into newer AOPs.

The performance metrics of the QSAR models discussed in this section, including accuracy, sensitivity, specificity, and correlation coefficient in the case of Al Sharif's model for PPAR $\gamma$  [78], are provided in Supplementary Table S2.

In summary, the versatility and predictive power of QSAR models, coupled with their ability to analyse large datasets efficiently, make QSAR models a valuable computational tool, particularly in the context of AOP data gap filling.

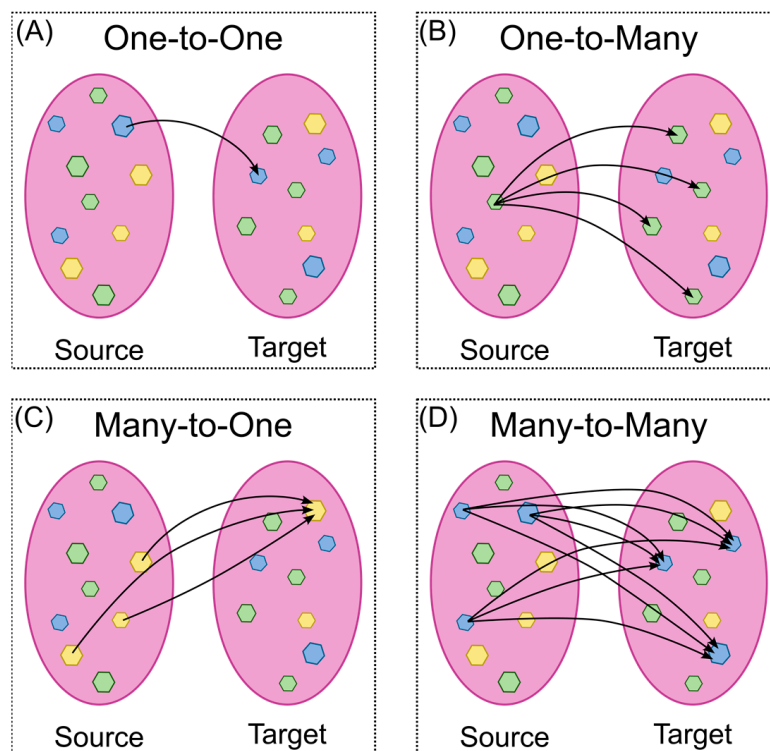
### 2.3. Read-Across

Read-across is founded on the principle that chemicals with comparable molecular structures are likely to share similar chemical, toxicokinetic, and toxicodynamic properties [91,92]. This approach uses information from chemical analogues with known toxicity within the same chemical group or "category" to assess unknown chemicals [93,94].

A chemical category is a group of chemicals that exhibit a similar biological action pattern and comparable toxicity or properties [93,95]. Similarity can be determined by shared functional groups, with common precursors leading to structurally similar metabolites or a consistent pattern in the variation of properties across the group [95,96]. However, it is important to note the inadequacy of relying solely on structural similarities to predict substance properties. Comprehensive scientific justification integrating various properties is recommended to address the potential toxicological impacts of structural variations [97]. This includes multiple lines of evidence, such as chemical structure, molecular properties, ADME (Absorption, Distribution, Metabolism, and Excretion) profiles, and biological assays [92,98]. Enoch and colleagues highlighted that defining chemical similarity based on a

clear MIE yields the most accurate prediction [56], emphasizing the importance of a holistic evaluation of biological and physicochemical properties in read-across assessments.

In the context of a read-across assessment, source chemicals are reference points for toxicity estimation within a category, while target chemicals are the compounds being estimated [99]. Read-across can be applied in four ways: first, “one-to-one”, where a single similar chemical is used to estimate the target chemical (Figure 5A); second, “one-to-many”, where a single analogue is used to make estimates for multiple target chemicals (Figure 5B); third, “many-to-one”, where multiple source chemicals are used to estimate the target chemical (Figure 5C); and, finally, “many-to-many”, where multiple source analogues are used to make estimates for multiple target chemicals (Figure 5D) [92,100,101].



**Figure 5.** Illustration of read-across assessment approaches. (A) One-to-one. (B) One-to-many. (C) Many-to-one. (D) Many-to-many.

There are two ways to conduct a read-across: the analogue and the category approaches. The analogue approach is used when only a few structurally similar substances are available with no apparent pattern in their properties. In this approach, the results of studies conducted on the source substance are directly applied to the target substance for all parameters, often using a worst-case scenario. This approach is generally susceptible to uncertainty [95,97,101]. On the other hand, the category approach is used when several substances share structural similarities and can be grouped based on defined characteristics and differences. Toxicological properties within the group are expected to be similar or exhibit a regular pattern, and predictions for target substance(s) can be made based on this pattern [97]. If no regular pattern is evident for a specific property among category members, a conservative read-across may be made from a category member with relevant information using a worst-case approach [95]. In this context, a worst-case approach ensures that the predicted effects for the target substance are not underestimated, with scientific explanations potentially rooted in kinetic or potency considerations. One notable advantage of this approach is that it enhances confidence in the reliability of results for individual substances within the category. This is due to identifying consistent patterns of effects within the category, which is a more robust and comprehensive evaluation compared to an assessment conducted on a substance-by-substance basis [101].

The workflow for read-across involves seven key steps: defining substance identity, providing comprehensive documentation, selecting the appropriate scenario, developing a hypothesis, gathering supporting evidence, evaluating assessment elements, and concluding the scientific acceptability of the read-across. These steps have been extensively described in the European Chemicals Agency (ECHA) Read-Across Assessment Framework (RAAF) [95]. Several publicly available tools can be used at different stages of the read-across workflow, as illustrated in Table 2. The characteristics and applicability of these tools have been thoroughly analysed and compared in prior works [97–99,102].

**Table 2.** Publicly available tools for the read-across workflow. All links were accessed on 20 September 2024.

Tool	Remarks	Link	Reference
AMBIT	Used for substance identification and similarity assessment.	<a href="https://cefic-lri.org/toolbox/ambit/">https://cefic-lri.org/toolbox/ambit/</a>	[103]
QSAR Toolbox	Utilized to develop hypotheses and evaluate assessment elements by providing mechanistic and endpoint-specific data.	<a href="https://qsartoolbox.org/">https://qsartoolbox.org/</a>	[104]
ToxRead	Provides toxicity predictions and read-across justification based on structural and mechanistic similarities.	<a href="https://www.vegahub.eu/download/toxread-download/">https://www.vegahub.eu/download/toxread-download/</a>	[105]
ToxMatch	Useful for identifying substances and scenarios by finding similar compounds and creating matrices.	<a href="https://toxmatch.sourceforge.net/">https://toxmatch.sourceforge.net/</a>	[106]
GenRA	Supports hypothesis development and evaluation by offering a read-across methodology that predicts toxicity based on chemical similarity and existing data.	<a href="https://comptox.epa.gov/genra/">https://comptox.epa.gov/genra/</a>	[107]
ToxGPS	Used for collecting evidence and evaluating assessments by combining data sources to offer detailed toxicity profiles and predictions.	<a href="https://mn-am.com/products/chemtunestoxgps/">https://mn-am.com/products/chemtunestoxgps/</a>	[108]
VERA	Uses multiple similarity metrics and clusters of analogues	<a href="https://www.vegahub.eu/portfolio-item/vera/">https://www.vegahub.eu/portfolio-item/vera/</a>	[109]

When a read-across prediction is performed, the following can enhance the confidence of the prediction:

- Scientific justification that the target chemical belongs to the category of the source chemicals.
- Experimental data from source chemicals that fall within the predicted values for the target chemical, indicating interpolation rather than extrapolation.
- Consideration of multiple source chemicals in the prediction process (a many-to-one approach).
- Integration of the prediction as part of a weight of evidence alongside data from other sources, such as ADME properties or relevant *in vitro* studies.

This evaluation of a read-across-based prediction is usually performed a posteriori, and, finally, it requires an expert judgment evaluation, which is very difficult to automatize.

In an attempt to assess this confidence, Myatt et al. introduced a scoring method, not only to read-across, to evaluate the reliability of *in silico* methods in general. This scoring considers the reliability, relevance and coverage of information [110]. Related to that, the ECHA RAAF also provides relevant guidelines to evaluate these aspects [95]. Nevertheless, as far as we know, there is no available tool to evaluate the confidence assessment in an automatic way.

Read-across is a practical alternative to other *in silico* techniques, such as SAR and QSAR models, especially when the toxicity data are limited. Unlike SAR and QSAR model

development, where a large dataset of compounds with known activity is required, read-across allows experts to use prior knowledge to identify suitable source chemicals. For this reason, it is possible to infer the activity of the target chemical even when only a few chemicals have toxicity data. One notable advantage of read-across is that it allows the interpretability of the association between chemical structure and activity, which is a feature often lacking in QSARs. Additionally, contrasting with the use of SAs, it is possible to predict the absence of a property or effect by using read-across from a negative result. This method is considered valid and convincing if the test design is adequate. Moreover, read-across in toxicology is a versatile approach that can model both qualitative and quantitative data, offering simplicity and transparency in its application [93,94,101,111].

Although read-across has proved to be a very useful tool for risk assessment, it also has several limitations that must be recognized. One significant challenge is the limited number of available source chemicals for a target chemical in small datasets. The effectiveness of read-across depends on data availability and quality for similar chemicals, which poses a limitation in cases where information is scarce or of insufficient quality. Additionally, the potential for inaccurate predictions arises when dealing with source chemicals that exhibit conflicting toxicity profiles. Another significant drawback is the reliance on expert knowledge, which introduces subjectivity and variability into the methodology. The current read-across methods rely heavily on the expertise of the individual conducting the study, which can lead to a lack of reproducibility and objectivity. This subjectivity appears in the selection of source chemicals and the interpretation of data, creating uncertainties and hindering the development of standardized practices [93,94,111].

Read-across has been used in several case studies related to hepatic steatosis. In the EU-ToxRisk project context, a microvesicular liver steatosis case study was developed. A group of 19 aliphatic carboxylic acids was evaluated *in vitro* for their potential to cause MIEs or KEs described in the AOP for liver steatosis [112]. From this case study, a set of publications were made with different read-across approaches. First, the Organization for Economic Co-operation and Development (OECD) published a read-across case study to assess the potential for 2-ethylbutyric acid to cause liver toxicity [92,113]. They used a group of aliphatic carboxylic acids, some known to cause liver steatosis, as source substances for the approach. Despite activating a single MIE (PPAR $\alpha$ ), *in vitro* and *in silico* data contradicted the hypothesis that 2-ethylbutyric acid might cause liver steatosis after 90 days of treatment. This finding aligns with the trend observed in source chemicals, where MIE activation and the potential to induce lipid accumulation decrease with shorter side chain length. Afterwards, the team published a study to enhance the application of chemical read-across in risk assessment by investigating transcriptomic responses of primary human hepatocytes. The focus was on 18 structurally different analogues of valproic acid known to cause hepatic steatosis in rodents [114]. The study revealed that compounds highly structurally related to valproic acid exhibited similar transcriptomic responses regarding gene expression patterns and potency. They highlighted the importance of integrating biological and physicochemical data to enhance read-across approaches. Later, they conducted a case study on 13 branched aliphatic carboxylic acids that were structurally similar [26]. They aimed to investigate the use of human-based NAMs to establish biological similarity. The study found a correlation between side chain length and the biological activity of the carboxylic acids. Longer-chain chemicals showed more significant activity in terms of MIEs and KEs. The study also highlighted the importance of considering late KEs for confirmation, as progression from MIEs to AOs was not guaranteed.

A different study investigated the limitations of using structural similarity alone for read-across in hepatotoxicity research and proposed integrating *in vitro* omics technologies to improve accuracy [115]. The investigation focused on three structural analogues of *p*-dialkoxy chlorobenzenes and identified shared biological pathways aligned with observed hepatotoxicity in the source chemical 1,4-dichloro-2,5-dimethoxybenzene. The study found that exposure to these chemicals induced cytochrome P450 (CYP) 4 family genes, linking them to peroxisome proliferators and emphasising the importance of considering metabolic

predictions. They identified several biological pathways altered by these chemicals, including mitochondrial dysfunction and FA  $\beta$ -oxidation. The study demonstrated the potential utility of in vitro toxicogenomics for grouping chemicals in read-across.

In summary, while read-across presents some limitations, it is a valuable tool in toxicology and has a high regulatory acceptance.

#### 2.4. Omics

“Omics” is a term used to describe the set of technologies that generate high-content data sets with measurements of gene transcripts, DNA methylation, proteins, metabolites, and lipids [116]. These assays produce large, diverse data sets, requiring bioinformatics tools for processing. Advanced computational techniques, including data mining, ML, and statistical methods, are essential for analysing these data and building prediction models [117]. Analysing omics data is essential to identify significant differences between experimental conditions. This analysis is critical for functional enrichment, which helps to identify altered biological pathways after exposure to a compound. Pathway databases such as Gene Ontology [118] or Kyoto Encyclopaedia of Genes and Genomes [119] can be used for this purpose. While processing vast volumes of data presents challenges, database technologies and new statistical algorithms offer solutions [120].

Recently, omics technologies have given birth to a new branch of toxicology known as toxicogenomics [121], which, by integrating omics data analysis and computational toxicology, investigates toxicity pathways and cellular responses to exposure to chemical substances [9,122,123]. Within the AOP framework, this integration facilitates the identification of downstream effects of MIEs and supports KEs by detecting molecular changes [5,9,14,116]. Most importantly, toxicogenomics provides highly valuable mechanistic information and is crucial in constructing AOPs by offering information on KEs, pathway enrichment, and biomarkers for toxicity testing [14,124].

A growing number of omics and toxicology databases useful for AOP development are available in the public domain, reducing the need for further experimentation through data reuse. Resources such as PubChem [125], ChEMBL [126], ToxNet [127], DrugMatrix [128], TG-GATEs [129], Connectivity Map [130], LINCS [131], and the Comparative Toxicogenomics Database [132] can be used to retrieve data from in vitro and in vivo experiments for AOP development [14,116,133]. Additionally, Table 3 shows examples of omics-based databases that provide relevant data for AOP studies.

**Table 3.** Omics data repositories. All links were accessed on 20 September 2024.

Omics	Data Repositories	Website	Reference
Genomics	Cell Collective	<a href="https://cellcollective.org/#">https://cellcollective.org/#</a>	[134]
	EGA (European Genome-phenome Archive)	<a href="https://ega-archive.org/">https://ega-archive.org/</a>	[135]
	ENA (European Nucleotide Archive)	<a href="https://www.ebi.ac.uk/ena/browser/home">https://www.ebi.ac.uk/ena/browser/home</a>	[136]
	EVA (European Variation Archive)	<a href="https://www.ebi.ac.uk/eva/">https://www.ebi.ac.uk/eva/</a>	[137]
Transcriptomics	ArrayExpress	<a href="https://www.ebi.ac.uk/biostudies/arrayexpress">https://www.ebi.ac.uk/biostudies/arrayexpress</a>	[138]
	dbGaP (database of Genotypes and Phenotypes)	<a href="https://www.ncbi.nlm.nih.gov/gap/">https://www.ncbi.nlm.nih.gov/gap/</a>	[139]
	ExpressionAtlas	<a href="https://www.ebi.ac.uk/gxa/home">https://www.ebi.ac.uk/gxa/home</a>	[140]
	GEO (Gene Expression Omnibus)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	[141,142]
	SRA (Sequence Read Archive)	<a href="https://www.ncbi.nlm.nih.gov/sra">https://www.ncbi.nlm.nih.gov/sra</a>	[143]
Proteomics	JPOST (Japan Proteome Standard) Repository	<a href="https://jpostdb.org/">https://jpostdb.org/</a>	[144]
	MassIVE (Mass Spectrometry Interactive Virtual Environment)	<a href="https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp">https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp</a>	[145]
	PaxDb (Protein Abundance Database)	<a href="https://pax-db.org/">https://pax-db.org/</a>	[146]
	PeptideAtlas	<a href="https://peptideatlas.org/">https://peptideatlas.org/</a>	[147]
	PRIDE	<a href="https://www.ebi.ac.uk/pride/archive/">https://www.ebi.ac.uk/pride/archive/</a>	[148]
	ProteomeXchange	<a href="https://www.proteomexchange.org/">https://www.proteomexchange.org/</a>	[149]
	ProteomicsDB	<a href="https://www.proteomicsdb.org/">https://www.proteomicsdb.org/</a>	[149]
Metabolomics	HMDB (Human Metabolome Database)	<a href="https://hmdb.ca/">https://hmdb.ca/</a>	[150]
	MetaboLights	<a href="https://www.ebi.ac.uk/metabolights/">https://www.ebi.ac.uk/metabolights/</a>	[151]
	Metabolomics Workbench	<a href="https://www.metabolomicsworkbench.org/">https://www.metabolomicsworkbench.org/</a>	[152]

The following section provides more details about the four main omics and how they can be applied in AOP development.

#### 2.4.1. Genomics

Genomics involves studying an organism's complete set of DNA, including all its genes and their interactions with each other and the environment. It encompasses technologies like gene sequencing and genotype analysis [153]. Genomics can be used for risk assessment by analysing multiple variants across the entire genome and their impact on susceptibility to compound exposure [123]. The individual's susceptibility to toxicity from exposure to a compound can be determined at the genome level by considering both pharmacokinetic (distribution, metabolism, and excretion) and pharmacodynamic (response of the tissues to the chemical) aspects [154], although it is not possible to detect effects of the exposure to chemical compounds on the individual, as genomics are not dynamic processes [155].

Genomic studies can be particularly beneficial in cases of idiosyncratic DILI, which involves unexpected adverse reactions in susceptible individuals at safe doses for the general population. Genetic polymorphisms, particularly those related to CYP and conjugating enzymes, are among the risk factors associated with idiosyncratic DILI. These polymorphisms are believed to contribute to variations in drug metabolism, which can lead to the accumulation of the compound or production of toxic metabolites [156,157] or alterations of immune-related genes, such as the human leukocyte antigen gene or genes related to cytokines [158]. For example, the influence of genetic variation of PXR on drug metabolism and clearance has been studied to reveal that it influences the expression of multiple CYPs and plays an important role in glucose homeostasis [159].

In 2011, researchers collected information on a list of genes whose hereditary alteration causes hepatic steatosis [160]. They pointed out that although the errors of metabolism observed in these disorders are observed in a minority of patients, they provide valuable information on the metabolic pathways involved in hepatic steatosis. A more recent study by the Genome-Wide Association used 1483 cases of MASLD and more than 17,000 controls [161]. This study identified genetic signals at four locations (chromosomes 2, 4, 19, and 22), with *PNPLA3* confirmed as a risk factor for the full spectrum of MASLD. Additionally, novel signals such as *PYGO1* were associated with steatosis, implicating Wnt signalling pathways in MASLD pathogenesis. Studies on different patient cohorts have identified *TM6SF2* and *GCKR* as key genes in the progression of MASLD [155]. Although MASLD is not directly caused by drug exposure, identifying these polymorphisms can help identify patients vulnerable to developing steatosis when undergoing pharmacological treatment.

Genomics can also be applied to cross-species extrapolation to help discern the differences in response between animal and human models [123,154]. A 2018 review article focused on AHR and analysed differences among tissues, organs, and species, considering AHR ligands as modulators of this nuclear receptor [162]. Analysing data from 20 studies performed on human, rat, and mouse cell lines, they detected differences in the toxic equivalency factors of several ligands dependent on the species.

#### 2.4.2. Transcriptomics

Transcriptomics involves the global measurement of RNA transcripts in a biological system, reflecting the simultaneous transcription of all genes at one point in time [153]. Different technologies are used for transcriptomics analysis, starting with quantitative polymerase chain reaction (qPCR) arrays, suitable for a few dozen genes, while targeted transcriptomic panels (microarrays) are used for hundreds to thousands of genes. RNA sequencing (RNA-Seq) enables comprehensive transcriptome analysis, whereas targeted RNA-Seq (TempO-seq) balances specificity with broader investigation [9,163]. RNA-Seq has garnered particular interest in the study of liver injury due to its wider dynamic range, making it more adept at detecting genes relevant to hepatotoxicity than microarrays [158]. However, this type of data is also difficult to analyse and interpret and requires a high level of genomic annotation.



For the statistical analysis of transcriptomics data to identify significant differences between experimental conditions, two different approaches can be applied: those analysing individual gene expression (e.g., fold change, rank product, etc.) and those considering the entire distribution (e.g., Bayesian, counting methods, etc.) [133,164]. Subsequently, the data undergo functional analysis to assign biological significance to the results.

The analysis of transcriptomics data helps characterize specific changes in the transcriptome, generating “transcriptomic signatures” that can play a crucial role in developing an AOP. Thus, transcriptomic data provides a solid basis for defining MIEs and KEs. Additionally, they can help identify a set of biomarkers associated with the AO, which can guide toxicity testing and hazard identification [8,165].

Transcriptomic techniques have been used to characterize the downstream effects of the interaction of certain compounds with proteins identified as MIEs of the AOP for hepatic steatosis. For example, microarrays were used to identify upregulated genes when human hepatocytes were treated with rifampicin, which is a well-known PXR agonist [166]. It was observed that the expression of genes such as *CYP3A4*, *CYP2B6*, multidrug resistance gene, *ALAs1* and thyroid hormone-responsive *SPOT14* homolog was increased.

Transcriptomic responses have also been used to identify specific genes and pathways associated with KEs related to hepatic lipid metabolism. For example, a study used transcriptome analysis to investigate three steatogenic compounds (amiodarone, valproic acid, and tetracycline) [63]. Amiodarone and valproic acid upregulate lipid metabolism, while tetracycline downregulates mitochondrial functions and fibrosis. The study suggests that all three compounds may affect PPAR signalling. In 2022, a set of 18 valproic acid analogues was assessed using primary human hepatocytes, which revealed differences in potency and highly similar expression patterns for structurally related compounds [114].

Some authors have taken advantage of the large amount of transcriptomics data available in public databases such as those listed in Table 3. For example, Abedini and colleagues used data from DrugMatrix [128] to perform a differentially expressed gene identification and pathway analysis [167]. The pathways identified through statistical analysis included glycolysis/gluconeogenesis, steroid hormone biosynthesis, retinol metabolism, metabolism of xenobiotics by CYP, PXR/RXR activation, AHR signalling, and xenobiotic metabolism signalling.

AbdulHameed and colleagues used pre-existing transcriptomic data on 18 known steatogenic chemicals to identify genes and pathways associated with drug-induced steatosis [168]. They found that pathways like the retinol metabolism were consistent across species and identified mitochondrial toxicity as a KE for steatosis.

Other studies chose a broader perspective and studied DILI using ML approaches on transcriptomic data. In the book Artificial Intelligence in Drug Design, chapter 18 presents an example model built using data from LINCS L1000 [131] and offers a guide on each step to build the ML model from transcriptomics data [169]. Some years prior, information on 77 compounds from the same database was used to identify the pathways altered by each compound [170]. They incorporated the transcriptomic data into a genome-wide metabolic model of hepatocytes to estimate the overall effects of chemical perturbation on the cell. Using this approach, they could identify, on average, 40 metabolic reactions activated under chemical perturbations.

#### 2.4.3. Proteomics

Proteomics is a multidisciplinary field that studies the functional outcomes encoded by genes by studying dynamic collections of proteins, known as proteomes, in living systems [9,120,153]. Proteomic experiments generally collect data on three properties of proteins in a sample: location, abundance/turnover, and post-translational modifications [9,153,171].

The most common technique for protein detection is mass spectrometry (MS), which has recently evolved to more advanced techniques like Tandem MS and imaging MS [149,172].

Isotope labelling techniques, like isotope-coded affinity tags and stable isotope labelling with amino acids in cell culture, can be applied to facilitate protein detection [117,120,173,174].

Proteomics is a powerful tool in toxicology, and two types of proteomic studies are commonly used. The first type focuses on understanding the molecular responses of a biological system exposed to a toxicant, which can provide knowledge of the cascade of events leading to changes in protein expression and function [153,174,175]. The second type of study is dedicated to identifying the molecular targets of the toxicant itself due to proteomics' broad scope and in-depth analysis capabilities [175]. These molecular targets could be identified as MIEs in the context of an AOP.

Even though proteomics is a valuable technique for toxicological assessment, it has not been extensively used for studying steatosis [155]. Some proteomic studies have generated data on a few compounds that cause steatosis using in vivo or in vitro models. Tetracycline was demonstrated to affect mitochondrial proteins involved in  $\beta$ -oxidation of FAs, especially acyl-CoA dehydrogenase [176]. Braeuning et al. conducted an in vitro study using proteomic analysis to investigate the hepatotoxicity of methapyrilene in rat hepatocytes [177]. They found that methapyrilene caused significant changes in amino acid and ammonia metabolism, branched-chain amino acid metabolism, urea and tricarboxylic cycle enzymes, and lipid metabolism. On top of that, even at non-toxic concentrations, methapyrilene triggered inflammatory responses.

#### 2.4.4. Metabolomics

Metabolomics involves the comprehensive analysis of small molecule intermediates and end products resulting from various biological processes. Categorized into intracellular, extracellular, microbial, and xenometabolome, they reflect the actions of proteins in biochemical pathways, which are influenced by the genome, transcriptome, proteome, environmental factors, drugs, and underlying diseases [117,153,178,179]. Two complementary analytical platforms are used for metabolomics: nuclear magnetic resonance (NMR), standing out for its robustness and non-destructibility, and MS, which provides very high sensitivity [153,178–181].

Metabolomics can play a role in developing AOPs by contributing to the understanding of toxic agents' mechanisms through a subfield known as toxicometabolomics [178]. This approach involves studying metabolites as downstream elements of toxicological mechanisms, providing structural and quantitative data to identify new biomarkers and target organs [9]. Toxicometabolomics not only offers insights into the mechanisms underlying KEs but also directly links specific metabolites to KEs within an AOP, thereby enhancing the causal understanding of the pathway components leading to AOs [165,182].

Metabolomics techniques have effectively characterized steatosis in various studies. For instance, a study employed MS to characterize steatosis induction in HepaRG cells through exposure to sodium valproate at four concentrations [183]. The research team discerned distinct patterns between the exposed and control cells, selecting up to 200 features via multivariate analyses that serve as potential biomarkers for steatosis. Among the identified biomarkers were diacylglycerol, triglyceride accumulation, and carnitine deficiency. Initial toxic responses revealed heightened levels of S-adenosylmethionine and mono-acetylspermidine, coupled with a moderate increase in triglycerides. Moreover, the study uncovered novel specific toxicity markers, including spermidines, creatine, and acetylcholine.

A study used a different approach to develop a technique for early detection of drug-induced steatosis [184]. This group opted to identify urinary metabolite changes associated with tetracycline-induced hepatic steatosis in rats at different disease stages. Combining NMR and MS, they revealed that exposure to 125 mg/kg of this compound potentially induced steatosis, as confirmed by histological and triglyceride content analyses. The integrated analysis of these metabolomic platforms, alongside multivariate statistical methods, which identified six urinary metabolite changes across six metabolic pathways. Notably, lysine concentration correlated with liver damage, providing insights into the pathogenesis

of steatosis development. The combination of metabolomics techniques demonstrated their suitability for studying the early stages of steatosis before the presentation of morphological alterations.

Focusing on general hepatotoxicity, a series of untargeted metabolomics studies investigated the impact of hepatotoxic and non-hepatotoxic substances on HepG2 cells. A first work revealed significant metabolic changes triggered by toxic insults and facilitated the construction of a discriminant model capable of distinguishing between non-toxic and hepatotoxic compounds [185]. Furthermore, distinct metabolite alterations could be related to oxidative stress, steatosis, and phospholipidosis, unravelling changes in underlying biochemical pathways [186]. A third study was focused on discerning potentially hepatotoxic drugs, quantifying associated alterations and unravelling the underlying toxicity mechanisms, including steatosis [187]. They trained their models with 25 hepatotoxic and 4 non-hepatotoxic compounds to predict toxicity mechanisms. They then tested 87 chemicals to create a “toxicity index” and identified unique pathways associated with steatosis, including the nicotinate and nicotinamide pathways. Correlation between NAD decrease, a key feature of this pathway, and the onset of steatosis was also highlighted.

#### 2.4.5. Multiomics

Multiomics refers to integrating various omics datasets to comprehensively analyse biological systems [188]. In toxicology, analysing single-type data is deemed insufficient, as this provides only a partial view of the biological system [117,188]. Hence, the integration of multiomics data at a higher organizational level is encouraged to enhance a comprehensive mechanistic understanding of toxicological responses in the context of an interconnected network that leads to determined phenotypic characteristics [117,179,189,190].

Recent advancements in computational models enable the integrative analysis of multiomics datasets for a more holistic approach in biological research. Integrative analysis employs three main methods to analyse diverse data types: concatenation-based integration, transformation-based integration, and model-based integration [117,188,191,192], which are summarized in Table 4.

**Table 4.** Characteristics of different omics data integration strategies.

Integration Strategy	Description	Processing	Main Drawback
Concatenation-based	Combines different datasets into a large matrix, increasing variables but not samples.	Applies feature selection or dimensionality reduction.	Makes the data matrix complex and noisy.
Transformation-based	Transforms each dataset into a less dimensional and noisy graph or kernel matrix.	Transformation into graph, kernel matrix, or deep learning architectures.	Can lead to information loss and distorted data.
Model-based	Applies separate ML models to each dataset and combines predictions.	Supervised: Bagging or voting to combine models. Unsupervised: Aggregates clustering results based on optimization criteria.	Prevents learning features shared across different omics data, limiting understanding of underlying mechanisms.

Multiomics data are complex and heterogeneous, making it difficult to analyse using ML approaches [191,193]. ML models require larger sample sizes for a correct fitting and can suffer from overfitting challenges if dimensionality problems arise from having too many omics variables associated with insufficient samples [117,188,191].

Another challenge is missing data, which can be omics-wise or sample-wise [117,188,191]. Omics-wise missing values occur when some samples have data for only certain omics, while sample-wise missing values can occur randomly or due to technical and experimental limitations [188,191]. To deal with this, the datasets must be harmonized, and missing values can be imputed; however, this can introduce bias [117,188].

Beyond the technical challenges of integrating diverse omics samples, which leads to the discussion on how distinct omics layers should be assessed at different temporal points [194], temporal data pose challenges for ML algorithms, which assume static inputs [117]. Addressing these issues is crucial for accurately interpreting time-dependent changes in omics data and optimizing the application of ML algorithms in dynamic contexts.

Some studies have performed experiments on multiple omics in the context of steatosis. For example, Ruepp and colleagues performed transcriptomic and proteomic analyses of acetaminophen toxicity in mouse liver [195]. The study revealed acute induction of *GM-CSF* mRNA (*GM-CSF* is a granulocyte-specific gene) from Kupffer cells and decreased mitochondrial chaperone proteins (Hsp10 and Hsp60) before observable morphological changes.

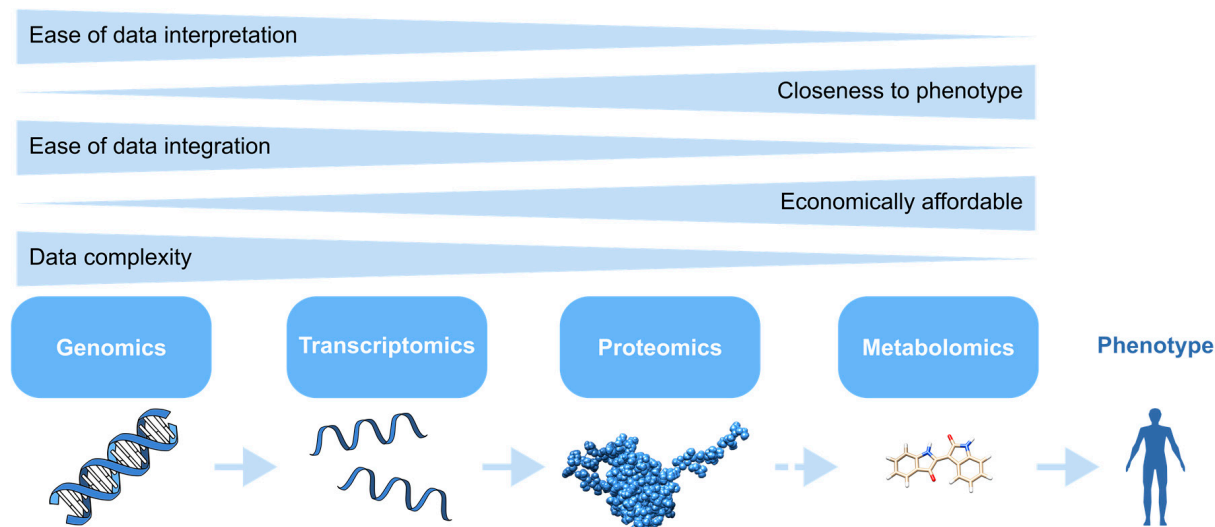
Other multiomics studies, particularly dual-omics studies, take a more comprehensive approach by conducting integrative analyses, delving into pathway analysis to uncover mechanistic insights associated with a steatotic response.

In 2018, van Breda and colleagues conducted an in-depth exploration of the impact of repetitive valproic acid treatment on primary human hepatocytes, aiming to elucidate the mechanisms leading to steatosis [196]. Employing microarrays and integrated data analysis, the team analysed whole-genome gene expression, DNA methylation, and miRNA changes in primary human hepatocytes exposed to a non-cytotoxic dose of the drug for five consecutive days, with a subsequent 3-day untreated period to assess the persistence of omics changes. The study identified significant alterations in pathways related to FA metabolism, cell signalling, bile acid metabolism, nuclear receptor pathways, and biotransformation. Integrative data analyses, including a washout period, revealed persistent changes, particularly on the epigenome level. Surprisingly, valproic acid treatment inhibited key transcription factors (*HNF1A* and *ONECUT1*), disrupting FA and glucose metabolism. The study also uncovered novel genes and networks associated with exposure to this compound and steatosis in a human liver model, shedding light on the molecular mechanisms involved. Notably, the research highlighted the substantial role of valproic acid-induced epigenetic changes, particularly in nuclear receptors such as PPAR $\alpha$ , PPAR $\gamma$ , AHR, and CD36, providing new insights into the intricate molecular pathways underlying steatosis induced by this drug.

The same workgroup published a second multiomics investigation to understand the intricate molecular events associated with mitochondrial dysfunction induced by the same compound in primary human hepatocytes [197]. They employed epigenomic, transcriptomic, and proteomic data to evaluate the cellular response over time. This study was focused on elucidating alterations in pathways related to mitochondrial function, particularly mitochondrial complex I and III–V genes, citric acid cycle, and  $\beta$ -oxidation. Employing time-resolved supervised data integration, the team utilized a network-based approach to reveal dynamic gene interactions during valproic acid treatment. Surprisingly, the study proposed a novel mitochondria–nucleus signalling axis (*MT-CO2–FN1–MYC–CPT1*) and challenged the conventional notion that oxidative stress solely triggers the mitochondria–nucleus signalling pathway. The results unveiled transient methylation changes in mitochondrial DNA and highlighted *MYC* and *FN1* as pivotal players in the persistent mitochondrial dysfunction caused by this compound. They also emphasized the importance of temporal resolution in unravelling complex biological responses.

#### 2.4.6. Omics Strengths and Limitations

Omics technologies, including transcriptomics, proteomics, genomics, and metabolomics, each bring unique strengths and limitations to the development of AOPs. Careful consideration of the research question and the desired depth of understanding is crucial for selecting the most appropriate omics technology in AOP development. Figure 6 shows a visual representation of the strengths and limitations of omics technologies based on their characteristics.



**Figure 6.** Visual comparison of four omics technologies: genomics, transcriptomics, proteomics, and metabolomics.

With its high-throughput sequencing techniques, genomics allows cost-efficient sequencing of complete genomes and simultaneous gene expression analysis. However, biological effects from exposure to a toxicant cannot be evaluated solely based on genomic analysis due to post-transcriptional, post-translational, and epigenetic changes [178].

Transcriptomics provides the most comprehensive platform, allowing for the measurement of individual gene expression, isoforms, and post-transcriptional modifications of the complete genome. While this technology provides abundant data, it faces challenges in the robustness of signals and processing of high-dimensional and noisy data [8,124,133].

Proteomics is a technique that analyses thousands of proteins simultaneously, linking proteomic profiles to phenotypes. It is distinct for its focus on post-translational modifications, which are crucial for understanding protein functions and signalling pathways [175]. This focus enhances the precision of pathway analyses and is key for studying protein networks and making quantitative comparisons across biological states [120,198,199]. Despite its advantages, proteomics faces challenges such as being labour-intensive, having a limited dynamic range—which complicates detecting less abundant proteins—and requiring expensive equipment [153,155,199]. Additionally, the proteome's complexity and instability, along with limited reproducibility, present significant obstacles [178].

Metabolomics is centred on examining the outcome of biological activities, providing a deeper understanding of the biochemical foundations of different phenotypes. It offers advantages such as cost-effectiveness, the simultaneous analysis of a large set of metabolites, and the non-invasive analysis of biofluids. Moreover, shared metabolic pathways and universal metabolite structures enable the translation of metabolic analyses between animal models and humans. However, they also have drawbacks, such as the need for analytical experts and specialized computational skills and the difficulty of extracting and analyzing metabolites with a single analytical platform. Additionally, sample collection, preparation, and analysis can limit metabolite detection in metabolomics research, underscoring the need for careful consideration when employing metabolomics [153,178,180,181,200].

Common challenges within omics technologies collectively impact their robustness and applicability in toxicology studies. Until very recently, the field has faced challenges such as a lack of standardization and regulations for data management, analysis, and interpretation, leading to variability between measurement platforms and bioinformatics pipelines [8,116,201]. Furthermore, omics technologies have intrinsic technical limitations. The data generated typically provide relative concentrations rather than absolute values, introducing uncertainties related to background noise, sample quality, and variability between MS runs and hindering reproducibility [155,201].

The choice of bioinformatics methods influences the outcomes, making data interpretation pipeline-dependent and hindering cross-platform correlation [124]. Furthermore, in certain instances, the annotation of biomolecule functions is constrained, complicating the interpretation of their contributions to the progression of an AO [9].

Despite the challenges, omics data analysis offers a comprehensive insight into an organism's biology, generating vast amounts of biological data across different complexity levels. This wealth of information becomes a valuable resource within the AOP framework, capable of contributing to virtually all information blocks in AOPs. Recognized for its role as a valuable source of mechanistic information, omics data play a crucial role in unravelling the intricate molecular details that underlie biological responses [8,121,124].

It is important to understand that pre-existing omics datasets cannot be used as predictive computational models. However, the information in these datasets can be utilized to conclude specific omics experiments, and these findings can be applied to novel compounds that belong to the same chemical space. The criteria for evaluating the AD used for read-across and predictive in silico techniques, such as SAR and QSAR, which can be applied to determine whether new compounds belong to the same chemical space.

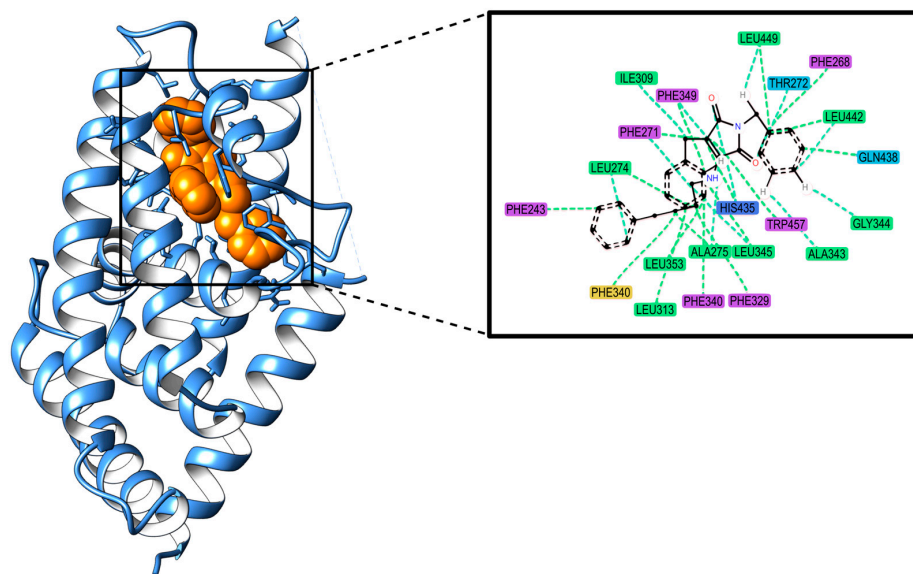
### 2.5. Structure-Based Approaches

Structure-based approaches are computational techniques that utilize the three-dimensional structures of molecules to study their interactions. In toxicology, these methods can be used to understand the interaction between toxicants and biological systems. Their applications include predicting toxicity outcomes, studying the mechanisms of action of toxicants, screening potential toxins, and facilitating the design of chemicals with enhanced safety profiles. Structure-based molecular modelling techniques primarily encompass molecular docking and molecular dynamics (MD) simulations [65]. These methods allow for the detailed analysis of molecular interactions, binding affinities, and dynamic behaviours of toxicants within biological targets, providing valuable insights into their potential toxic effects.

Molecular docking is defined as the prediction of the conformation and orientation of a ligand within the binding site of a target protein [202]. It is used to understand the modulation of biochemical processes related to a ligand's activity [203]. The docking process involves the spatial and energy matching between ligand and receptor, allowing the ranking of chemical databases according to complementarity to a given target [204,205]. This in silico method has gained popularity since the early 1980s and consists of two steps: considering the ligand's conformational state and evaluating its interactions with the target using scoring functions [202–204]. Figure 7 illustrates the result of docking a ligand to the LXR $\beta$  and shows the specific interactions.

Molecular docking is used in drug discovery for hit identification and optimization, drug repurposing, and multitarget ligand design [66]. Elucidating the interactions between ligands and molecular targets provides insights into disease mechanisms. Advances in docking algorithms and the availability of 3D structures of potential targets, facilitated by the evolution of techniques such as X-ray crystallography and cryo-electron microscopy, have made large-scale screenings possible. These advancements allow for the identification of protein binding sites, novel molecular targets of known ligands, and potential adverse drug reactions [202,204].

The docking process highly depends on the availability of a three-dimensional structure of the target with sufficient resolution [65,202,204]. When such structures are unavailable, in silico homology modelling is often used, which characterizes an unknown protein structure using a related homologous protein whose experimental structure is known [204,206]. Recently, thanks to AlphaFold, DeepMind's artificial intelligence model [207], structures from the AlphaFold Protein Structure Database can also be used for structure-based drug discovery. However, there are limitations associated with this strategy, such as the absence of a ligand in the structure and the incomplete model that is obtained due to the lack of water molecules, metal ions, and co-factors [206].



**Figure 7.** LXR ligand docked onto LXR $\beta$  structure. The figure on the right shows a representation of the ligand–protein interactions.

According to the “induced-fit” theory created by Koshland [208], both proteins and ligands are inherently flexible and are reshaped by their interactions. Different strategies are used for docking purposes regarding flexibility, going from exclusively rigid docking to semi-flexible and flexible docking. Rigid docking is the fastest but less precise, while flexible docking is the most accurate but computationally expensive. Semi-flexible docking is popular when computer resources are limited [203,205].

On the other hand, for simplicity and faster calculations, the participation of water molecules in protein–ligand interactions is often ignored [202]. Furthermore, when docking is applied alone, the kinetics of possible ligands are not considered [209]. This kind of approximation in the scoring functions is one of the sources of a potential lack of correlation with the experimentally obtained binding affinities [204].

Multiple molecular docking and molecular dynamics tools are available, utilizing distinct algorithms and serving different purposes [65]. Table 5 presents a compilation of several of these programs.

**Table 5.** Molecular docking and molecular dynamics most used tools. All links were accessed on 20 September 2024.

Technique	Tool	Website	Reference
Molecular Docking	AutoDock	<a href="https://autodock.scripps.edu/">https://autodock.scripps.edu/</a>	[210]
	CSAlign-Dock	<a href="https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=CSALIGN">https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=CSALIGN</a>	[211]
	DOCK	<a href="https://dock.compbio.ucsf.edu/">https://dock.compbio.ucsf.edu/</a>	[212]
	DockingServer	<a href="https://www.dockingserver.com/web">https://www.dockingserver.com/web</a>	[213]
	GalaxyDock	<a href="https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=DOCK">https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=DOCK</a>	[214]
	Gold	<a href="https://www.ccdc.cam.ac.uk/solutions/software/gold/">https://www.ccdc.cam.ac.uk/solutions/software/gold/</a>	[215]
	Glide	<a href="https://newsite.schrodinger.com/platform/products/glide/">https://newsite.schrodinger.com/platform/products/glide/</a>	[216]
	HADDOCK	<a href="https://wenmr.science.uu.nl/">https://wenmr.science.uu.nl/</a>	[217]
	Medusa Dock	<a href="https://dokhlab.med.psu.edu/cpi/#/MedusaDock">https://dokhlab.med.psu.edu/cpi/#/MedusaDock</a>	[218]
	PyDock	<a href="https://life.bsc.es/pid/pydock/">https://life.bsc.es/pid/pydock/</a>	[219]
	Rosetta	<a href="https://www.rosettacommons.org/software">https://www.rosettacommons.org/software</a>	[220]
SwissDock	<a href="http://www.swissdock.ch/">http://www.swissdock.ch/</a>	[221]	
Molecular Dynamics	GROMACS	<a href="https://www.gromacs.org/">https://www.gromacs.org/</a>	[222]
	AMBER	<a href="https://ambermd.org/">https://ambermd.org/</a>	[223]
	NAMD	<a href="http://www.ks.uiuc.edu/Research/namd/">http://www.ks.uiuc.edu/Research/namd/</a>	[224]
	CHARMM	<a href="https://www.charmm.org/">https://www.charmm.org/</a>	[225]
	Desmond	<a href="https://www.schrodinger.com/platform/products/desmond/">https://www.schrodinger.com/platform/products/desmond/</a>	[226]

Moreover, beyond the traditional docking approach, scientists leverage other cutting-edge techniques, such as MD. This computer-based methodology emulates the natural movements of atoms and molecules within chemical and biological systems. Using MD simulations, researchers can delve into how chemicals interact with biological systems, such as proteins and cell membranes [227,228]. This, in turn, can offer valuable insights into molecular mechanisms and enable the development of safer chemicals. Given their versatility, MD simulations are extensively employed in computational studies.

A variety of MD simulation software has been developed over time. The most well-known software packages are listed in Table 5. Some MD simulation packages come with their own force fields, while others require external force fields as input for the simulation algorithms [227].

MD simulations can be used to gain insights into the quality of homology-modelled protein structures, investigate the structural or conformational changes of proteins, understand important interactions between a protein and small molecules, and estimate protein–protein and protein–ligand binding affinity [227]. Utilizing MD simulations for virtual screening proves to be a daunting task, primarily due to their high computational demands and extensive time requirements [228], surpassing those of molecular docking.

Docking and other structural approaches are the most useful when the study's objective is the detection of substrates or inhibitors of determined proteins. Therefore, they can be used to predict changes in the MIEs. Molecular modelling approaches have been employed to study different proteins marked as MIEs in the steatosis AOP network.

For example, LXR was characterized using different molecular modelling approaches and a QSAR model to predict potential binders [79]. The objective of this study was to characterize LXR's ligand binding domain by defining the essential features leading to binding. For this, the researchers used ensemble docking, which consists of docking a ligand into an ensemble of protein structures corresponding to the same protein. The algorithm automatically selects the ligand–protein pair with the best docking score. In addition, they integrated this approach with identifying key features from a potent binder compound and calculating fingerprints for chemical similarity assessment, comparing with two reference ligands as templates.

Researchers employed a multifaceted approach to explore hepatic steatosis induced by novel replacement flame retardants [229]. In addition to conducting a series of *in vitro* and omics assays to delineate KEs within the steatosis AOP, they performed molecular docking studies on PXR and PPAR $\gamma$ . The selection of PXR and PPAR $\gamma$  as focal points stemmed from examining the ToxCast database, revealing that most of the novel replacement flame retardants exhibited activity specifically on these two receptors.

Several studies have investigated the mode of action of different compounds as ligands of PPAR $\gamma$  in relation to liver steatosis. In 2014, Tsakovska and colleagues used molecular modelling techniques to analyse PPAR $\gamma$  complexes with full agonists [230]. They characterized the receptor binding pocket and the specific ligand–receptor interactions of the most active ligands. They determined the key residues with which agonists interact and pointed out that the binding pocket presents certain flexibility depending on the bound ligand. Additionally, they developed a model that collected the three-dimensional arrangement of chemical features essential for the interaction with PPAR $\gamma$  from these ligands. In a second study, the same team performed a virtual screening based on docking filtering to screen for PPAR $\gamma$  full agonists [78]. When validating their virtual screening protocol, they obtained an 85% sensitivity, indicating the protocol's ability to correctly identify PPAR $\gamma$  full agonists. They also achieved a 77% specificity, demonstrating the protocol's effectiveness in correctly identifying PPAR $\gamma$  decoys (compounds that mimic binder's properties but differ topologically to reduce the chance of binding).

Still focusing on the same MIE, several studies have investigated the effect of concrete groups of compounds on that receptor. Alemán-González-Duhart and colleagues worked on developing thiazolidinediones derivatives that are effective in treating diabetes by activating PPAR $\gamma$  but would reduce the tendency to cause hepatic steatosis [231]. They



screened 130 derivatives and analysed how the differences in their structures changed the interaction with the ligand binding domain of PPAR $\gamma$ . Another study centred on the hepatotoxic effect of bisphenol S used docking to determine the interactions of this compound with PPAR $\gamma$  [232]. They detected interactions with four key residues in the ligand binding domain. The authors concluded that PPAR $\gamma$  is a potential target of bisphenol S, which would explain the steatogenic effect of this compound.

On the other hand, a recent publication used molecular docking in combination with QSAR models to predict mitochondrial dysfunction, which, as mentioned, is considered a KE for steatosis. In this case, the enzyme succinate–cytochrome c reductase activity was taken as a single molecular activity related to mitochondrial dysfunction [233]. This unusual approach can be taken to broaden the applicability field of molecular docking to generate predicted data that could fill knowledge gaps in an AOP network. Another study combined screening with structure-based pharmacophore models with ML models to identify mitochondrial toxicants [234].

An example of automatization of structure-based techniques is the online tool called Endocrine Disruptome (<http://endocrinedisruptome.ki.si>, accessed on 20 September 2024) that encompasses an array of 14 nuclear receptors, seven of which have been identified as MIEs of the steatosis AOP: ER $\alpha$ , ER $\beta$ , GR, LXR $\alpha$ , LXR $\beta$ , PPAR $\alpha$ , and PPAR $\gamma$ . It is a docking server that utilizes AutoDock Vina for automated molecular docking to validated crystal structures of these nuclear receptors [235].

### 3. Regulatory Acceptance

In the context of risk assessment, several attempts have generated regulatory guidelines to standardize these computational approaches. Organizations like the OECD and the ECHA have worked to generate several documents that delineate principles ensuring the transparency, consistency, and acceptability of results derived from various *in silico* approaches.

Over the past two decades, legislative measures such as the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) regulation have embraced computational approaches like SAR, QSAR, and read-across as valid tools for risk assessment. Under this regulation, manufacturers are obligated to provide detailed information on the chemicals they manufacture, market, or import [77,236–239].

Comprehensive guidance documents on read-across have been published by the ECHA [100] and the OECD [101]. These documents outline general considerations for conducting read-across [100,101], including the following:

- Evaluating the relevance and reliability of data from source chemicals.
- Ensuring that the data align with current OECD test methods.
- Assessing the impact of multifunctional compounds on read-across reliability.
- Analysing the purity and impurity profiles of both target and source chemicals to identify impurities that may influence overall toxicity.
- Comparing physicochemical properties such as molecular weight and water solubility to gauge similarity.
- Understanding likely toxicokinetics and metabolic pathways.
- Using QSARs for decision making on additional testing.

In their 2008 guidelines, the ECHA establishes the RAAF under the REACH regulation [100]. Furthermore, the use of NAMs has been explored to support read-across in a consecutive report that gathered several cases of framework application [3].

On the other hand, the OECD has outlined general rules for assessing the suitability of SAR and QSAR models for regulatory purposes. The agreed OECD principles for SAR and QSAR models determine that, to be considered for regulatory purposes, a model should include the following information: a defined endpoint; an unambiguous algorithm; a defined domain of applicability; appropriate measures of goodness-of-fit, robustness and predictivity; and a mechanistic interpretation, if possible [239]. Since 2004, the OECD has

published a complete guideline document for validating SAR and QSAR models used for regulatory purposes, where specific steps and requirements are stated [76,239].

The OECD recently developed a guidance document called the OECD Omics Reporting Framework (OORF) to report omics data for regulatory purposes [240]. The framework aims to promote data sharing, increase transparency in data processing, enable the assessment of study quality, and encourage reproducibility for omics toxicology experiments. The OORF resulted from previous efforts to standardize omics technologies in the regulatory field, such as the Transcriptomics Reporting Framework, which was proposed in 2017. The Transcriptomics Reporting Framework provided a detailed list of items that should be reported in a transcriptomics experiment, including experimental design, platforms, and data transformations [124,164]. The new OORF is focused on transcriptomics and metabolomics, leaving work on proteomics for the future. They harmonized modules for the reporting of the study summary and the toxicology experiment, as well as different data analysis methods, and they developed separate reporting modules for data acquisition and processing in transcriptomics assays (microarrays, RNA-Seq or Targeted RNA-Seq, and qPCR arrays) and metabolomics assays (MS and NMR) [240]. The regulation and standardization of omics data can ultimately lead to a more straightforward implementation of analysis workflows to extract any relevant mechanistic information.

In the case of structure-based approaches, although they are widely used for drug discovery purposes and can also be applied in the field of toxicology [78,227,241,242], there have been no efforts thus far to develop regulatory frameworks in the context of risk assessment.

The increasing interest of regulatory agencies in incorporating NAMs, particularly computer-based methodologies, into new-generation risk assessment strategies is expected to drive the establishment of novel regulatory frameworks. These frameworks will standardize the workflows of *in silico* methodologies, ensuring their reliability for integration into regulatory risk assessment processes.

#### 4. Integration of the *In Silico* Methodologies to Fill Knowledge Gaps in the AOP

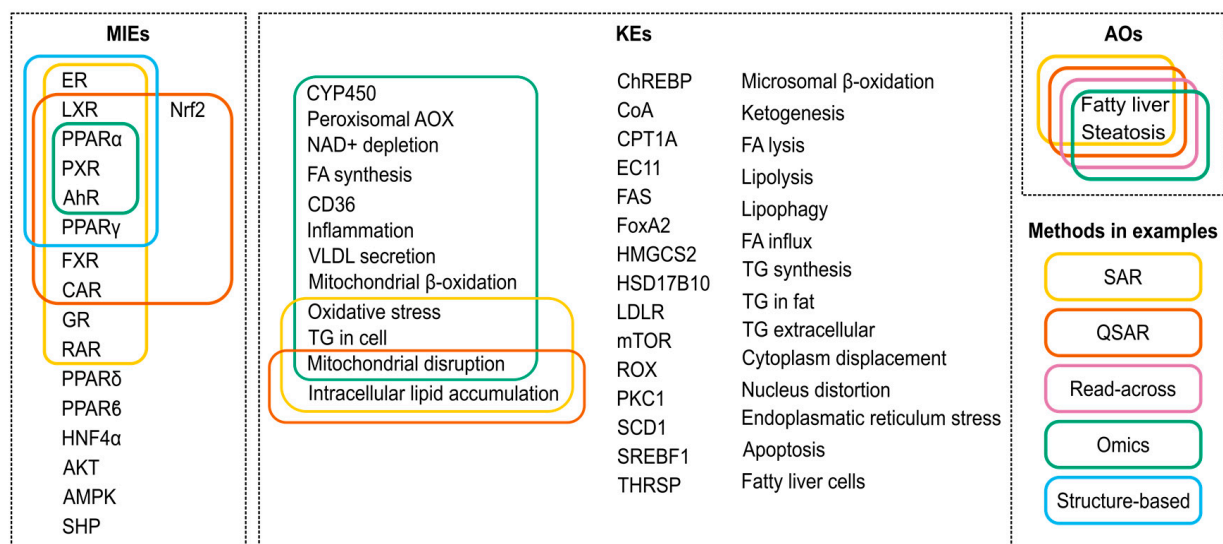
During this review, we have examined various computational methods that have been utilized to provide evidence for the activation of different points in the AOP for hepatic steatosis by chemicals. Figure 8 offers a visual representation of AOP points covered by the examples gathered in this review. Our findings are promising, as we could find evidence for many points contemplated in the AOP. We are pleased to report that the majority of MIEs have been covered with SAR, and many have specific QSAR models. We also identified structure-based studies exploring binding to ER, GR, LXR, PPARs, and PXR. We also encountered several examples of SAR, QSAR, read-across, and omics approaches that tackle steatosis as a whole endpoint. However, we observed that only a few KEs identified in the AOP have been addressed in the reviewed examples.

When searching for evidence for each point, we prioritized the key events KEs that are not unique to one of the references AOPs. For this reason, although we are aware of numerous QSAR models for predicting activation and inhibition of CYP450 isoforms [243,244], we did not consider them for this review.

We have found more evidence for the KE of mitochondrial disruption, along with mitochondrial  $\beta$ -oxidation inhibition and dysfunction. This endpoint has not been exclusively studied in the context of hepatic steatosis, as it is an essential indicator of different types of toxicity resulting from exposure to drugs.

In our analysis of existing literature, we utilized the combined AOP as a framework to identify instances of computational approaches employed for predicting specific endpoints. However, it came to our attention that several KEs within the AOP were not addressed in the studies examined. While the upregulation of several genes such as *ChREBP*, *SREBP-1c*, *FAS*, and *SCD1* is highlighted in the AOP and has been extensively characterized, we did not discover any examples in our analysis demonstrating a clear association between the upregulation of these genes and drug exposure. We believe further exploration of omics

studies in this area would yield valuable insights into the mechanisms underlying the alteration of these genes. However, this was beyond the scope of our review, which aimed to provide a comprehensive overview of the applicability of computational approaches in addressing data gaps in an AOP.



**Figure 8.** Hepatic steatosis AOP points covered by the different computational approaches reviewed in this work.

Moreover, we observed that the elevation in FA influx was marked as a KE. However, we could not identify any computational study that examined the effect of exposure to compounds that directly impact FA influx. While several studies explored changes in lipid metabolism, we do not consider FA influx synonymous with FA metabolism *per se*. Therefore, we conclude that these studies did not address changes in the transportation of FA.

We have observed that while many changes at the cellular level can be identified through in vitro studies (such as cytoplasm displacement or nucleus distortion) these KEs are less commonly used for computational research than mitochondrial disruption. These changes occur because of the accumulation of lipid droplets in the cytoplasm, which displace organelles and compress the nucleus. However, measuring lipid accumulation seems to be the most convenient way to evaluate changes in cell phenotype, and this information has been used as an endpoint for some computational studies to predict cellular changes associated with steatosis.

The question arises of whether new computational models are necessary to predict the specific KEs not covered by currently available models. Firstly, we must consider whether enough data are available for those specific endpoints to build a model. Additionally, we should assess whether having specific computational models for those endpoints provides valuable information. We aim to use all sources of evidence to obtain a more accurate prediction of the steatotic AO. Perhaps focusing our efforts on improving the prediction of the activation of a few KEs and having more specific models for groups of chemicals could be a more practical approach.

On the other hand, we have found some significant changes at the molecular and cellular levels that are not covered in the AOP. For instance, Ivanov and colleagues' SAR study identified some proteins associated with lipid disorders, but only AHR is mentioned in the steatosis AOP [53]. However, some other proteins they identified, such as the Niemann–Pick C1 protein, crucial in transporting cholesterol and other lipids within cells, are not included. Additionally, other listed proteins have roles in metabolic processes, such as the 1-acylglycerol-3-phosphate O-acyltransferase  $\beta$  in CoA production, fructose-1,6-

bisphosphatase in gluconeogenesis, and glucagon-like peptide 1 receptor in the regulation of insulin.

In this review, we have covered multiple omics approaches that offer significant insight into pathways that may not be represented in the AOP network we worked with. For instance, the genes upregulated by exposure to rifampicin in Moreau's transcriptomics study were absent in the AOP [166]. The proteomics studies we gathered in this work identified several potential biomarkers of oxidative stress that are not currently included in the AOP and could help in the precise characterization of molecular-level alterations. Among the metabolomics studies we reviewed, only Martínez-Sena's study identified NAD depletion [187]. In contrast, the others identified alterations of other relevant metabolites related to lipid metabolism, such as diacylglycerol and carnitine, which are not reflected in the AOP. These examples illustrate the importance of omics approaches in developing AOPs, particularly in identifying biomarkers that could play a key role as MIE or KE for a specific AO and in aiding in connecting the points in the AOP through integrating molecular pathways.

However, some of the assays used to generate data for *in silico* modelling discussed in this review present a significant challenge: extrapolating *in vivo* toxicity from *in vitro* data. On one side, there is a lack of cellular models to cover all the MIEs and KEs critical for triggering a specific AO. On the other side, *in vitro* models do not reflect the complexity of complex physiological systems, such as the input from other signalling pathways driven by hormones or cytokines that may influence the outcome. This is especially important if toxicity is caused by metabolites, as some *in vitro* assays ignore metabolic activation. In the context of the analysis of omics data, there is an additional limitation, as the transcriptome, proteome and metabolome shift in the majority of cells under certain culture conditions, which is evident when different batches are compared [245,246]. Thus, the transparency of protocols and the rigorous and robust reproducibility analysis are essential to understand the limitations and interpret the conclusions.

Building upon the insights gained from the reviewed examples, structure-based approaches, especially molecular docking, provide a comprehensive understanding of protein targets that require precise interaction modes for ligand activation or inhibition. However, in the case of protein targets with promiscuous binding pockets, such as PPAR $\alpha$ , more complex methodologies may be necessary to determine interaction specificity. This can include analysing protein–ligand interactions in-depth or performing molecular dynamics simulations. Regardless of the complexity, structure-based techniques can be combined with other *in silico* methodologies to accurately predict protein activation or inhibition for MIEs.

Based on the information we have gathered, it is apparent that several *in silico* methodologies, such as SAR, QSAR, read-across, and structure-based approaches, can be used to generate predictions of how novel compounds act on different endpoints in the AOP network. It is important to note that the knowledge generated from omics techniques can be used to develop *in silico* models to predict the activity of new compounds on those specific endpoints and extrapolate conclusions on novel compounds using read-across approaches. However, linking simplistic *in silico* models with complex regulatory-relevant *in vivo* health outcomes remains a major challenge [2]. When multiple sources of evidence are integrated, they must be assessed for their reliability, relevance, and consistency. The weight of evidence assigned to each source should be determined based on these factors. After that, the sources of evidence should be organized and filtered accordingly [99,247,248].

When multiple lines of evidence are available for a single endpoint in an AOP, such as the inhibition of PPAR $\alpha$ , it is essential to assess the biological significance of the available predictive models [247]. This evaluation should also consider the biological AD of the MIE within the AOP, including its relevance to taxonomic, age, and sex factors [249]. Additionally, we should assess whether the results obtained for the same chemical using different available models are compatible, thereby assigning a consistency evaluation to the predictions [247]. Finally, we should also evaluate the AD, accuracy and precision of

individual computational models used to make predictions for the specific MIE to assess the reliability of predictions made for new chemicals [247,248].

When we encounter inconsistent results for a particular substance using various prediction techniques, we must carefully evaluate all the evidence by examining the specific criteria used to classify the effect [248]. However, establishing a hierarchy of prioritization of techniques can be challenging. Some authors believe that performing a read-across to extrapolate predictions from experimental values of similar compounds can yield more reliable results than those obtained from QSAR models [248]. Nonetheless, the conflicting concept of similarity evaluation discussed in the read-across section can make the criteria for extrapolation unclear, especially when data is scarce.

Conversely, some frameworks have been developed to establish a hierarchy of predictions based on reliability, such as the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) M7 guideline for mutagenicity [250,251]. In these cases, even when different predictions come from the same technique, the final decision is made by experts. Therefore, the best strategy to follow remains an ongoing discussion.

Nonetheless, using computational models presents itself as a promising approach to predict the AO more accurately. By employing this strategy, we can obtain a complete matrix of data that incorporates all the biological knowledge present in the AOP. This approach may be considered more favourable than experimental data, which necessitate the conduction of *in vitro* or *in vivo* assays, as well as the associated time and economic costs and ethical implications in the case of *in vivo* assays.

## 5. Conclusions

In this comprehensive review, we have compiled the characteristics and potential applications of various computational methodologies, including SAR, QSAR, read-across, omics data analysis, and structure-based approaches, in addressing data gaps within an AOP framework or constructing new AOPs. Utilizing a combination of four established AOP networks for drug-induced hepatic steatosis as our focal point, we have critically analysed existing studies, elucidating the strengths and limitations of each explored technique.

The case study on hepatic steatosis demonstrates the versatility of computational techniques in generating a more complete and interconnected data matrix. Our analysis reveals that computational models can effectively complement AOPs by filling critical data gaps and identifying new KEs. The examples gathered in this paper highlight the transformative impact of *in silico* techniques in toxicology, providing a blueprint for integrating these strategies into new-generation risk assessment methodologies. By leveraging the strengths of computational tools, we can achieve more efficient, cost-effective, and predictive toxicological assessments, ultimately enhancing our ability to safeguard public health and the environment.

There is a growing acceptance by regulatory agencies like OECD and ECHA of these computational methodologies, which underscores their practical applicability in toxicological risk assessments. However, full implementation into regulatory decision making remains a challenge, requiring standardized workflows and consistent validation protocols for computational models. Regulatory guidelines like the OECD principles for QSAR model validation provide a strong foundation but need further refinement to accommodate emerging methods. To achieve that, continuous collaboration between regulatory bodies, industry, and academic research is essential. In conclusion, while hepatic steatosis serves as a pertinent case study, the principles and guidelines proposed in this review are broadly applicable to other types of chemical-induced toxicity. The continued development and integration of computational approaches within the AOP framework is necessary to address existing data deficiencies and to support the establishment of novel AOPs. This integrated approach holds great promise for the future of toxicological risk assessment, providing a more mechanistically driven and evidence-based foundation for regulatory decision making.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms252011154/s1>.

**Author Contributions:** Conceptualization, E.S.-C., L.T. and R.O.-V.; writing—original draft preparation, R.O.-V.; writing—review and editing, E.S.-C., M.P.-S., R.G., E.B., L.T. and G.F.E.; visualization, R.O.-V.; supervision, E.S.-C. and R.G.; funding acquisition, R.G., E.B. and G.F.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was performed as part of the ASPIS cluster (<https://aspis-cluster.eu/>). The ASPIS cluster has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 963845 (ONTOX—<https://ontox-project.eu/>), No 965406 (PrecisionTox—<https://precisiontox.org/>) and No 964537 (RISK-HUNT3R—<https://www.risk-hunt3r.eu/>). We gratefully acknowledge the financial support of the Agencia Valenciana de la Investigación (AVI, <https://innoavi.es/en/>) through its program Innodocto (Reference number INNNTA3/2022/13), financed by the European Union through the European Regional Development Fund (ERDF) Program 2021–2027 Comunitat Valenciana.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analysed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** The authors thank the members of the ASPIS Computational Methods Working Group for their valuable contributions and engaging discussions that shaped this review.

**Conflicts of Interest:** Authors Rita Ortega-Vallbona, Martina Palomino-Schätzlein, Rafael Gozalbes and Eva Serrano-Candelas were employed by the company ProtoQSAR S.L., author Rafael Gozalbes was employed by the company MolDrug AI Systems S.L. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. This output reflects only the authors’ view, and the European Union cannot be held responsible for any use that may be made of the information contained therein.

## Abbreviations

AD: Applicability Domain; ADME: Absorption, Distribution, Metabolism, Excretion; AHR: Aryl hydrocarbon receptor; AO: Adverse Outcome; AOP: Adverse Outcome Pathway; CAR: Constitutive androstane receptor; CD36: Cluster of Differentiation 36, also known as Fatty Acid Translocase (FAT); CYP: Cytochrome P450; DILI: Drug-Induced Liver Injury; ECHA: European Chemicals Agency; ER: Estrogen receptor; FA: Fatty Acid; FXR: Farnesoid X receptor; GR: Glucocorticoid receptor; KE: Key Event; LXR: Liver X receptor; MASLD: Metabolic dysfunction-associated steatotic liver disease; MD: Molecular dynamics; MIE: Molecular Initiating Event; ML: Machine Learning; MS: Mass spectrometry; NAMs: New Approach Methodologies; NMR: Nuclear magnetic resonance; NRF2: nuclear factor (erythroid-derived 2) like 2; OECD: Organisation for Economic Co-operation and Development; OORF: OECD Omics Reporting Framework; PASS: Prediction of Activity Spectra for Substances; PPAR: Peroxisome proliferator-activated receptor; PXR: Pregnane X receptor; qPCR: Quantitative polymerase chain reaction; QSAR: Quantitative Structure–Activity Relationship; RAAF: Read-Across Assessment Framework; REACH: Registration, Evaluation, Authorization, and Restriction of Chemicals; RNA-Seq: RNA sequencing; RXR: Retinoid X receptor; SAR: Structure-Activity Relationship; SA: Structural Alert; TempO-seq: Targeted RNA sequencing.

## References

1. Brescia, S.; Alexander-White, C.; Li, H.; Cayley, A. Risk assessment in the 21st century: Where are we heading? *Toxicol. Res.* **2023**, *12*, 1–11. [[CrossRef](#)]
2. Bajard, L.; Adamovsky, O.; Audouze, K.; Baken, K.; Barouki, R.; Beltman, J.B.; Beronius, A.; Bonefeld-Jørgensen, E.C.; Cano-Sancho, G.; de Baat, M.L.; et al. Application of AOPs to assist regulatory assessment of chemical risks—Case studies, needs and recommendations. *Environ. Res.* **2023**, *217*, 114650. [[CrossRef](#)]
3. ECHA. New Approach Methodologies in Regulatory Science. In Proceedings of the Scientific Workshop, Helsinki, Finland, 19–20 April 2016; European Chemicals Agency: Helsinki, Finland, 2016.
4. Kleinstreuer, N.; Hartung, T. Artificial intelligence (AI)—it's the end of the tox as we know it (and I feel fine). *Arch. Toxicol.* **2024**, *98*, 735–754. [[CrossRef](#)]
5. Krewski, D.; Andersen, M.E.; Tyshenko, M.G.; Krishnan, K.; Hartung, T.; Boekelheide, K.; Wambaugh, J.F.; Jones, D.; Whelan, M.; Thomas, R.; et al. Toxicity testing in the 21st century: Progress in the past decade and future perspectives. *Arch. Toxicol.* **2020**, *94*, 1–58. [[CrossRef](#)]
6. Burgoon, L.D.; Angrish, M.; Garcia-Reyero, N.; Pollesch, N.; Zupanic, A.; Perkins, E. Predicting the Probability that a Chemical Causes Steatosis Using Adverse Outcome Pathway Bayesian Networks (AOPBNs). *Risk Anal.* **2020**, *40*, 512–523. [[CrossRef](#)]
7. Nelms, M.D.; Mellor, C.L.; Enoch, S.J.; Judson, R.S.; Patlewicz, G.; Richard, A.M.; Madden, J.M.; Cronin, M.T.D.; Edwards, S.W. A mechanistic framework for integrating chemical structure and high-throughput screening results to improve toxicity predictions. *Comput. Toxicol.* **2018**, *8*, 1–12. [[CrossRef](#)]
8. Vinken, M. Omics-based input and output in the development and use of adverse outcome pathways. *Curr. Opin. Toxicol.* **2019**, *18*, 8–12. [[CrossRef](#)]
9. Portugal, J.; Mansilla, S.; Piña, B. Perspectives on the Use of Toxicogenomics to Assess Environmental Risk. *Front. Biosci.* **2022**, *27*, 294. [[CrossRef](#)]
10. Vinken, M. The adverse outcome pathway concept: A pragmatic tool in toxicology. *Toxicology* **2013**, *312*, 158–165. [[CrossRef](#)]
11. Arnesdotter, E.; Gijbels, E.; Rodrigues, B.D.S.; Vilas-Boas, V.; Vinken, M. Chapter 20: Adverse Outcome Pathways as Versatile Tools in Liver Toxicity Testing. In *In Silico Methods for Predicting Drug Toxicity*; Benfenati, E., Ed.; Springer Science+Business Media: New York, NY, USA, 2022; pp. 521–535.
12. Conolly, R.B.; Ankley, G.T.; Cheng, W.; Mayo, M.L.; Miller, D.H.; Perkins, E.J.; Villeneuve, D.L.; Watanabe, K.H. Quantitative Adverse Outcome Pathways and Their Application to Predictive Toxicology. *Environ. Sci. Technol.* **2017**, *51*, 4661–4672. [[CrossRef](#)]
13. Perkins, E.J.; Ashauer, R.; Burgoon, L.; Conolly, R.; Landesmann, B.; Mackay, C.; Murphy, C.A.; Pollesch, N.; Wheeler, J.R.; Zupanic, A.; et al. Building and Applying Quantitative Adverse Outcome Pathway Models for Chemical Hazard and Risk Assessment. *Environ. Toxicol. Chem.* **2019**, *38*, 1850–1865. [[CrossRef](#)]
14. Leist, M.; Ghallab, A.; Graepel, R.; Marchan, R.; Hassan, R.; Bennekou, S.H.; Limonciel, A.; Vinken, M.; Schildknecht, S.; Waldmann, T.; et al. Adverse outcome pathways: Opportunities, limitations and open questions. *Arch. Toxicol.* **2017**, *91*, 3477–3505. [[CrossRef](#)]
15. Villeneuve, D.L.; Crump, D.; Garcia-Reyero, N.; Hecker, M.; Hutchinson, T.H.; LaLone, C.A.; Landesmann, B.; Lettieri, T.; Munn, S.; Nepelska, M.; et al. Adverse outcome pathway (AOP) development I: Strategies and principles. *Toxicol. Sci.* **2014**, *142*, 312–320. [[CrossRef](#)]
16. Donato, M.T.; Gomez-Lechon, M.J. Drug-induced Liver Steatosis and Phospholipidosis: Cell-Based Assays for Early Screening of Drug Candidates. *Curr. Drug. Metab.* **2012**, *13*, 1160–1173. [[CrossRef](#)]
17. Pizzo, F.; Lombardo, A.; Manganaro, A.; Benfenati, E. A New Structure-Activity Relationship (SAR) Model for predicting drug-induced liver injury, based on statistical and expert-based structural alerts. *Front. Pharmacol.* **2016**, *7*, 442. [[CrossRef](#)]
18. Donato, M.T.; Martínez-Romero, A.; Jiménez, N.; Negro, A.; Herrera, G.; Castell, J.V.; O'Connor, J.-E.; Gómez-Lechón, M.J. Cytometric analysis for drug-induced steatosis in HepG2 cells. *Chem. Biol. Interact.* **2009**, *181*, 417–423. [[CrossRef](#)]
19. Luckert, C.; Braeuning, A.; De Sousa, G.; Durinck, S.; Katsanou, E.S.; Konstantinidou, P.; Machera, K.; Milani, E.S.; Peijnenburg, A.A.C.M.; Rahmani, R.; et al. Adverse Outcome Pathway-Driven Analysis of Liver Steatosis in Vitro: A Case Study with Cyproconazole. *Chem. Res. Toxicol.* **2018**, *31*, 784–798. [[CrossRef](#)]
20. Vinken, M. Adverse Outcome Pathways and Drug-Induced Liver Injury Testing. *Chem. Res. Toxicol.* **2015**, *28*, 1391–1397. [[CrossRef](#)]
21. Jain, S.; Norinder, U.; Escher, S.E.; Zdravil, B. Combining in Vivo Data with In Silico Predictions for Modeling Hepatic Steatosis by Using Stratified Bagging and Conformal Prediction. *Chem. Res. Toxicol.* **2021**, *34*, 656–668. [[CrossRef](#)]
22. Angrish, M.M.; Kaiser, J.P.; McQueen, C.A.; Chorley, B.N. Tipping the balance: Hepatotoxicity and the 4 apical key events of hepatic steatosis. *Toxicol. Sciences* **2016**, *150*, 261–268. [[CrossRef](#)]
23. Liebe, R.; Esposito, I.; Bock, H.H.; vom Dahl, S.; Stindt, J.; Baumann, U.; Luedde, T.; Keitel, V. Diagnosis and management of secondary causes of steatohepatitis. *J. Hepatol.* **2021**, *74*, 1455–1471. [[CrossRef](#)]
24. Verhoeven, A.; van Ertvelde, J.; Boeckmans, J.; Gatzios, A.; Jover, R.; Lindeman, B.; Lopez-Soop, G.; Rodrigues, R.M.; Rapisarda, A.; Sanz-Serrano, J.; et al. A Quantitative Weight-of-Evidence Method for Confidence Assessment of Adverse Outcome Pathway Networks: A Case Study on Chemical-Induced Liver Steatosis. *Toxicology* **2024**, *505*, 153814. [[CrossRef](#)]
25. Gadaleta, D.; Manganelli, S.; Roncaglioni, A.; Toma, C.; Benfenati, E.; Mombelli, E. QSAR Modeling of ToxCast Assays Relevant to the Molecular Initiating Events of AOPs Leading to Hepatic Steatosis. *J. Chem. Inf. Model.* **2018**, *58*, 1501–1517. [[CrossRef](#)]

26. Escher, S.E.; Aguayo-Orozco, A.; Benfenati, E.; Bitsch, A.; Braunbeck, T.; Brotzmann, K.; Bois, F.; van der Burg, B.; Castel, J.; Exner, T.; et al. Integrate mechanistic evidence from new approach methodologies (NAMs) into a read-across assessment to characterise trends in shared mode of action. *Toxicol. In Vitro* **2022**, *79*, 105269. [CrossRef]
27. Arnesdotter, E.; Spinu, N.; Firman, J.; Ebbrell, D.; Cronin, M.T.D.D.; Vanhaecke, T.; Vinken, M. Derivation, characterisation and analysis of an adverse outcome pathway network for human hepatotoxicity. *Toxicology* **2021**, *459*, 152856. [CrossRef]
28. Hong, C.; Tontonoz, P. Liver X receptors in lipid metabolism: Opportunities for drug discovery. *Nat. Rev. Drug Discov.* **2014**, *13*, 433–444. [CrossRef]
29. Xiao, L.; Xie, X.; Zhai, Y. Functional crosstalk of CAR-LXR and ROR-LXR in drug metabolism and lipid metabolism. *Adv. Drug Deliv. Rev.* **2010**, *62*, 1316–1321. [CrossRef]
30. Yang, H.; Lou, C.; Li, W.; Liu, G.; Tang, Y. Computational Approaches to Identify Structural Alerts and Their Applications in Environmental Toxicology and Drug Discovery. *Chem. Res. Toxicol.* **2020**, *33*, 1312–1322. [CrossRef]
31. Métivier, J.P.; Lepailleur, A.; Buzmakov, A.; Poezevara, G.; Crémilleux, B.; Kuznetsov, S.O.; Goff, J.L.; Napoli, A.; Bureau, R.; Cuissart, B. Discovering structural alerts for mutagenicity using stable emerging molecular patterns. *J. Chem. Inf. Model.* **2015**, *55*, 925–940. [CrossRef]
32. Chakravarti, S. Augmenting Expert Knowledge-Based Toxicity Alerts by Statistically Mined Molecular Fragments. *Chem. Res. Toxicol.* **2022**, *36*, 848–858. [CrossRef]
33. Gini, G. QSAR Methods. In *In Silico Methods for Predicting Drug Toxicity*; Benfenati, E., Ed.; Springer Science+Business Media: New York, NY, USA, 2016; pp. 1–20.
34. Ahlberg, E.; Carlsson, L.; Boyer, S. Computational derivation of structural alerts from large toxicology data sets. *J. Chem. Inf. Model.* **2014**, *54*, 2945–2952. [CrossRef]
35. Lepailleur, A.; Poezevara, G.; Bureau, R. Automated detection of structural alerts (chemical fragments) in (eco)toxicology. *Comput. Struct. Biotechnol. J.* **2013**, *5*, e201302013. [CrossRef]
36. PubChem. PubChem Substructure Fingerprint. Available online: [https://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem\\_fingerprints.pdf](https://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.pdf) (accessed on 2 January 2024).
37. Klekota, J.; Roth, F.P. Chemical substructures that enrich for biological activity. *Bioinformatics* **2008**, *24*, 2518–2525. [CrossRef]
38. Cortes-Ciriano, I. Bioalerts: A python library for the derivation of structural alerts from bioactivity and toxicity data sets. *J. Cheminform* **2016**, *8*, 13. [CrossRef]
39. Klopman, G. Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *J. Am. Chem. Soc.* **1984**, *106*, 7315–7321. [CrossRef]
40. Chakravarti, S.K.; Saiakhov, R.D.; Klopman, G. Optimizing Predictive Performance of CASE Ultra Expert System Models Using the Applicability Domains of Individual Toxicity Alerts. *J. Chem. Inf. Model.* **2012**, *52*, 2609–2618. [CrossRef]
41. Parasuraman, S. Prediction of activity spectra for substances. *J. Pharmacol. Pharmacother.* **2011**, *2*, 52–53. [CrossRef]
42. Ferrari, T.; Cattaneo, D.; Gini, G.; Golbamaki Bakhtyari, N.; Manganaro, A.; Benfenati, E. Automatic knowledge extraction from chemical structures: The case of mutagenicity prediction. *SAR QSAR Environ. Res.* **2013**, *24*, 365–383. [CrossRef]
43. Nijssen, S.; Kok, J.N. A quickstart in frequent structure mining can make a difference. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2004; pp. 647–652.
44. Borgelt, C.; Berthold, M.R. Mining molecular fragments: Finding relevant substructures of molecules. In *Proceedings of the 2002 IEEE International Conference on Data Mining, 2002. Proceedings, Maebashi, Japan, 9–12 December 2002*; pp. 51–58.
45. Enslein, K. An overview of structure-activity relationships as an alternative to testing in animals for carcinogenicity, mutagenicity, dermal and eye irritation, and acute oral toxicity. *Toxicol. Ind. Health* **1988**, *4*, 479–498. [CrossRef]
46. Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *Altern. Lab. Anim.* **2005**, *33*, 445–459. [CrossRef]
47. Tropsha, A.; Golbraikh, A. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Curr. Pharm. Des.* **2007**, *13*, 3494–3504. [CrossRef]
48. Benigni, R.; Bossa, C.; Tcheremenskaia, O.; Giuliani, A. Alternatives to the carcinogenicity bioassay: In silico methods, and the in vitro and in vivo mutagenicity assays. *Expert Opin. Drug Metab. Toxicol.* **2010**, *6*, 809–819. [CrossRef]
49. Ferrari, T.; Gini, G.; Golbamaki Bakhtyari, N.; Benfenati, E. Mining toxicity structural alerts from SMILES: A new way to derive structure activity relationships. In *Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining, Paris, France, 11–15 April 2011*; pp. 120–127. [CrossRef]
50. Guha, R. Chapter 6 On Exploring Structure—Activity Relationships. In *In Silico Models for Drug Discovery*; Kortagere, S., Ed.; Humana Press: Totowa, NJ, USA, 2013; Volume 993. [CrossRef]
51. Ferrari, T.; Lombardo, A.; Benfenati, E. QSARpy: A new flexible algorithm to generate QSAR models based on dissimilarities. The log Kow case study. *Sci. Total Environ.* **2018**, *637–638*, 1158–1165. [CrossRef]
52. Mellor, C.L.; Steinmetz, F.P.; Cronin, M.T.D. Using Molecular Initiating Events to Develop a Structural Alert Based Screening Workflow for Nuclear Receptor Ligands Associated with Hepatic Steatosis. *Chem. Res. Toxicol.* **2016**, *29*, 203–212. [CrossRef]
53. Ivanov, S.; Semin, M.; Lagunin, A.; Filimonov, D.; Poroikov, V. In Silico Identification of Proteins Associated with Drug-induced Liver Injury Based on the Prediction of Drug-target Interactions. *Mol. Inform.* **2017**, *36*, 1–13. [CrossRef]
54. Onyango, A.N. Excessive gluconeogenesis causes the hepatic insulin resistance paradox and its sequelae. *Heliyon* **2022**, *8*, e12294. [CrossRef]



55. Le, T.T.; Ziemba, A.; Urasaki, Y.; Hayes, E.; Brotman, S.; Pizzorno, G. Disruption of uridine homeostasis links liver pyrimidine metabolism to lipid accumulation. *J. Lipid Res.* **2013**, *54*, 1044–1057. [[CrossRef](#)]
56. Enoch, S.; Mellor, C.; Nelms, M. Mitochondrial Structure and Function. In *Mitochondrial Dysfunction by Drug and Environmental Toxicants*; Wiley: Hoboken, NJ, USA, 2018; pp. 25–34.
57. Hemmerich, J.; Troger, F.; Füzi, B.F.; Ecker, G. Using Machine Learning Methods and Structural Alerts for Prediction of Mitochondrial Toxicity. *Mol. Inform.* **2020**, *39*, 1–16. [[CrossRef](#)]
58. Nelms, M.D.; Mellor, C.L.; Cronin, M.T.D.; Madden, J.C.; Enoch, S.J. Development of an In Silico Profiler for Mitochondrial Toxicity. *Chem. Res. Toxicol.* **2015**, *28*, 1891–1902. [[CrossRef](#)]
59. Naven, R.T.; Swiss, R.; Klug-Mcleod, J.; Will, Y.; Greene, N. The development of structure-activity relationships for mitochondrial dysfunction: Uncoupling of oxidative phosphorylation. *Toxicol. Sci.* **2013**, *131*, 271–278. [[CrossRef](#)]
60. Ortega-Vallbona, R.; Méndez, R.; Tolosa, L.; Escher, S.E.; Castell, J.V.; Gozalbes, R.; Serrano-Candelas, E. Uncovering the toxicity mechanisms of a series of carboxylic acids in liver cells through computational and experimental approaches. *Toxicology* **2024**, *504*, 153764. [[CrossRef](#)]
61. Yang, C.; Tarkhov, A.; Marusczyk, J.; Bienfait, B.; Gasteiger, J.; Kleinoeder, T.; Magdziarz, T.; Sacher, O.; Schwab, C.H.; Schwoebel, J.; et al. New Publicly Available Chemical Query Language, CSRML, To Support Chemotype Representations for Application to Data Mining and Modeling. *J. Chem. Inf. Model.* **2015**, *55*, 510–528. [[CrossRef](#)]
62. Greene, N.; Fisk, L.; Naven, R.T.; Note, R.R.; Patel, M.L.; Pelletier, D.J. Developing structure-activity relationships for the prediction of hepatotoxicity. *Chem. Res. Toxicol.* **2010**, *23*, 1215–1222. [[CrossRef](#)]
63. Szalowska, E.; Van Der Burg, B.; Man, H.Y.; Hendriksen, P.J.M.; Peijnenburg, A.A.C.M. Model steatogenic compounds (amiodarone, valproic acid, and tetracycline) alter lipid metabolism by different mechanisms in mouse liver slices. *PLoS ONE* **2014**, *9*, e86795. [[CrossRef](#)]
64. Muratov, E.N.; Bajorath, J.; Sheridan, R.P.; Tetko, I.V.; Filimonov, D.; Poroikov, V.; Oprea, T.I.; Baskin, I.I.; Varnek, A.; Roitberg, A.; et al. QSAR without borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564. [[CrossRef](#)]
65. Carpio, L.E.; Sanz, Y.; Gozalbes, R.; Barigye, S.J. Computational strategies for the discovery of biological functions of health foods, nutraceuticals and cosmeceuticals: A review. *Mol. Divers.* **2021**, *25*, 1425–1438. [[CrossRef](#)]
66. Serrano-Candelas, E.; Carpio, L.E.; Gozalbes, R. Chapter 10: Computational Modeling of DYRK1A Inhibitors as Potential Anti-Alzheimer Agents. In *Computational Modeling of Drugs Against Alzheimer's Disease*; Roy, K., Ed.; Springer Science+Business Media: Berlin/Heidelberg, Germany, 2023; pp. 295–324.
67. Winter, R.; Montanari, F.; Noé, F.; Clevert, D.A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci.* **2019**, *10*, 1692–1701. [[CrossRef](#)]
68. Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* **2010**, *29*, 476–488. [[CrossRef](#)]
69. Toropov, A.A.; Toropova, A.P. QSPR/QSAR: State-of-Art, Weirdness, the Future. *Molecules* **2020**, *25*, 1292. [[CrossRef](#)]
70. Chakravarti, S.K.; Alla, S.R.M. Descriptor Free QSAR Modeling Using Deep Learning With Long Short-Term Memory Neural Networks. *Front. Artif. Intell.* **2019**, *2*, 1–18. [[CrossRef](#)]
71. Helmke, P.; Füzi, B.; Ecker, G.F. Bioactivity descriptors for in vivo toxicity prediction: Now and the future. *Expert Opin. Drug Metab. Toxicol.* **2024**, *20*, 541–543. [[CrossRef](#)]
72. Füzi, B.; Malik-Sheriff, R.S.; Manners, E.J.; Hermjakob, H.; Ecker, G.F. KNIME workflow for retrieving causal drug and protein interactions, building networks, and performing topological enrichment analysis demonstrated by a DILI case study. *J. Cheminform.* **2022**, *14*, 37. [[CrossRef](#)]
73. Hemmerich, J.; Ecker, G.F. In silico toxicology: From structure-activity relationships towards deep learning and adverse outcome pathways. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, e1475. [[CrossRef](#)]
74. North American Free Trade Agreement (NAFTA) Technical Working Group on Pesticides (TWG). *(Quantitative) Structure Activity Relationship [(Q)SAR] Guidance Document*; U.S. Environmental Protection Agency: Washington, DC, USA, 2012.
75. Worth, A.; Lapenna, S.; Piparo, E.L.; Serafimova, R. *A Framework for Assessing In Silico Toxicity Predictions: Case Studies with Selected Pesticides*; Publications Office of the European Union: Luxembourg, 2011.
76. OECD. *(Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure Activity Relationship Models and Predictions*; OECD Publishing: Paris, France, 2023. [[CrossRef](#)]
77. Gozalbes, R.; Julián-Ortiz, J.V. Applications of Chemoinformatics in Predictive Toxicology for Regulatory Purposes, Especially in the Context of the EU REACH Legislation. *Int. J. Quant. Struct.-Prop. Relatsh.* **2018**, *3*, 1–24. [[CrossRef](#)]
78. Al Sharif, M.; Tsakovska, I.; Pajeva, I.; Alov, P.; Fioravanzo, E.; Bassan, A.; Kovarich, S.; Yang, C.; Mostrag-Szlichtyng, A.; Vitcheva, V.; et al. The application of molecular modelling in the safety assessment of chemicals: A case study on ligand-dependent PPAR $\gamma$  dysregulation. *Toxicology* **2017**, *392*, 140–154. [[CrossRef](#)]
79. Palczewska, A.; Kovarich, S.; Ciacci, A.; Fioravanzo, E.; Bassan, A.; Neagu, D. Ranking strategies to support toxicity prediction: A case study on potential LXR binders. *Comput. Toxicol.* **2019**, *10*, 130–144. [[CrossRef](#)]
80. Goya-Jorge, E.; Giner, R.M.; Sylla-Iyarreta Veitia, M.; Gozalbes, R.; Barigye, S.J. Predictive modeling of aryl hydrocarbon receptor (AhR) agonism. *Chemosphere* **2020**, *256*, 127068. [[CrossRef](#)]
81. Kurosaki, K.; Wu, R.; Uesawa, Y. A toxicity prediction tool for potential agonist/antagonist activities in molecular initiating events based on chemical structures. *Int. J. Mol. Sci.* **2020**, *21*, 1–20. [[CrossRef](#)]

82. Dix, D.J.; Houck, K.A.; Martin, M.T.; Richard, A.M.; Setzer, R.W.; Kavlock, R.J. The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. *Toxicol. Sci.* **2007**, *95*, 5–12. [[CrossRef](#)]
83. Thomas, R.S.; Paules, R.S.; Simeonov, A.; Fitzpatrick, S.C.; Crofton, K.M.; Casey, W.M.; Mendrick, D.L. The US Federal Tox21 Program: A strategic and operational plan for continued leadership. *ALTEX* **2018**, *35*, 163–168. [[CrossRef](#)]
84. Tang, W.; Chen, J.; Hong, H. Discriminant models on mitochondrial toxicity improved by consensus modeling and resolving imbalance in training. *Chemosphere* **2020**, *253*, 126768. [[CrossRef](#)]
85. Garcia de Lomana, M.; Marin Zapata, P.A.; Montanari, F. Predicting the Mitochondrial Toxicity of Small Molecules: Insights from Mechanistic Assays and Cell Painting Data. *Chem. Res. Toxicol.* **2023**, *36*, 1107–1120. [[CrossRef](#)]
86. Cotterill, J.; Price, N.; Rorije, E.; Peijnenburg, A. Development of a QSAR model to predict hepatic steatosis using freely available machine learning tools. *Food Chem. Toxicol.* **2020**, *142*, 111494. [[CrossRef](#)]
87. Shin, H.K.; Kang, M.-G.; Park, D.; Park, T.; Yoon, S. Development of Prediction Models for Drug-Induced Cholestasis, Cirrhosis, Hepatitis, and Steatosis Based on Drug and Drug Metabolite Structures. *Front. Pharmacol.* **2020**, *11*, 67. [[CrossRef](#)]
88. Chen, M.; Suzuki, A.; Thakkar, S.; Yu, K.; Hu, C.; Tong, W. DILIrank: The largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discov. Today* **2016**, *21*, 648–653. [[CrossRef](#)]
89. Elsevier: Pharmapendium. 2008. Available online: <https://www.pharmapendium.com> (accessed on 5 August 2024).
90. Fujitsu. ADME Database. Available online: <https://www.fujitsu.com/global/solutions/business-technology/tc/sol/admedatabase/> (accessed on 5 August 2024).
91. Schultz, T.W.; Przybylak, K.R.; Richarz, A.N.; Mellor, C.L.; Escher, S.E.; Bradbury, S.P.; Cronin, M.T.D. Read-across of 90-day rat oral repeated-dose toxicity: A case study for selected n-alkanols. *Comput. Toxicol.* **2017**, *2*, 12–19. [[CrossRef](#)]
92. Rovida, C.; Escher, S.E.; Herzler, M.; Bennekou, S.H.; Kamp, H.; Kroese, D.E.; Maslankiewicz, L.; Moné, M.J.; Patlewicz, G.; Sipes, N.; et al. NAM-Supported Read-Across: From Case Studies to Regulatory Guidance in Safety Assessment. *ALTEX* **2021**, *38*, 140–150. [[CrossRef](#)]
93. Chatterjee, M.; Roy, K. Computational Modeling of Mixture Toxicity. In *In Silico Methods for Predicting Drug Toxicity*; Benfenati, E., Ed.; Humana: New York, NY, USA, 2022; pp. 561–587.
94. Gajewicz, A.; Jagiello, K.; Cronin, M.T.D.; Leszczynski, J.; Puzyn, T. Addressing a bottle neck for regulation of nanomaterials: Quantitative read-across (Nano-QRA) algorithm for cases when only limited data is available. *Environ. Sci. Nano* **2017**, *4*, 346–358. [[CrossRef](#)]
95. ECHA. *Read-Across Assessment Framework (RAAF)*; European Chemicals Agency: Helsinki, Finland, 2017.
96. Smith, G.F. Artificial Intelligence in Drug Safety and Metabolism. *Methods Mol. Biol.* **2022**, *2390*, 483–501. [[CrossRef](#)]
97. Patlewicz, G.; Helman, G.; Pradeep, P.; Shah, I. Navigating through the minefield of read-across tools: A review of in silico tools for grouping. *Comput. Toxicol.* **2017**, *3*, 1–18. [[CrossRef](#)]
98. Yang, C.; Rathman, J.F.; Mostrag, A.; Ribeiro, J.V.; Hobocienski, B.; Magdziarz, T.; Kulkarni, S.; Barton-Maclaren, T. High Throughput Read-Across for Screening a Large Inventory of Related Structures by Balancing Artificial Intelligence/Machine Learning and Human Knowledge. *Chem. Res. Toxicol.* **2023**, *36*, 1081–1106. [[CrossRef](#)]
99. Astuto, M.C.; Di Nicola, M.R.; Tarazona, J.V.; Rotais, A.; Devos, Y.; Liem, A.K.D.; Kass, G.E.E.; Bastaki, M.; Schoonjans, R.; Maggiore, A.; et al. In Silico Methods for Environmental Risk Assessment: Principles, Tiered Approaches, Applications, and Future Perspectives. In *In Silico Methods for Predicting Drug Toxicity*; Benfenati, E., Ed.; Springer Science+Business Media: Berlin/Heidelberg, Germany, 2022; pp. 589–634.
100. ECHA. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.6: QSARs and Grouping of Chemicals. 2008. Available online: <https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment> (accessed on 16 January 2024).
101. OECD. *Guidance on Grouping of Chemicals*; Series on Testing and Assessment; OECD Publishing: Paris, France, 2014.
102. Jeliaskova, N.; Jaworska, J.; Worth, A.P. Open Source Tools for Read-Across and Category Formation. In *In Silico Toxicology*; The Royal Society of Chemistry: London, UK, 2010; pp. 408–445.
103. Cefic-LRI. Ambit—Cheminformatics Data Management System. Available online: <https://cefic-lri.org/toolbox/ambit/> (accessed on 12 December 2023).
104. 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.; et al. QSAR Toolbox—Workflow and major functionalities. *SAR QSAR Environ. Res.* **2016**, *27*, 203–219. [[CrossRef](#)]
105. Gini, G.; Franchi, A.M.; Manganaro, A.; Golbamaki, A.; Benfenati, E. ToxRead: A tool to assist in read across and its use to assess mutagenicity of chemicals. *SAR QSAR Environ. Res.* **2014**, *25*, 999–1011. [[CrossRef](#)]
106. Gallegos-Saliner, A.; Poater, A.; Jeliaskova, N.; Patlewicz, G.; Worth, A.P. Toxmatch—A chemical classification and activity prediction tool based on similarity measures. *Regul. Toxicol. Pharmacol.* **2008**, *52*, 77–84. [[CrossRef](#)]
107. Helman, G.; Shah, I.; Williams, A.J.; Edwards, J.; Dunne, J.; Patlewicz, G. Generalized read-across (GENRA): A workflow implemented into the EPA comptox chemicals dashboard. *ALTEX* **2019**, *36*, 462–465. [[CrossRef](#)]
108. Yang, C.; Rathman, J.F.; Bienfait, B.; Burbank, M.; Detroyer, A.; Enoch, S.J.; Firman, J.W.; Gutsell, S.; Hewitt, N.J.; Hobocienski, B.; et al. The role of a molecular informatics platform to support next generation risk assessment. *Comput. Toxicol.* **2023**, *26*, 100272. [[CrossRef](#)]

109. Colombo, E.; Luca Viganò, E.; Raitano, G.; Lombardo, A.; Manganaro, A.; Sommovigo, A.; Benfenati, E. The VERA software: Implementation of the acute fish toxicity endpoint and its application to pharmaceutical compounds. *Chemosphere* **2024**, *358*, 142232. [[CrossRef](#)]
110. Myatt, G.J.; Ahlberg, E.; Akahori, Y.; Allen, D.; Amberg, A.; Anger, L.T.; Aptula, A.; Auerbach, S.; Beilke, L.; Bellion, P.; et al. In silico toxicology protocols. *Regul. Toxicol. Pharmacol.* **2018**, *96*, 1–17. [[CrossRef](#)]
111. Shah, I.; Liu, J.; Judson, R.S.; Thomas, R.S.; Patlewicz, G. Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information. *Regul. Toxicol. Pharmacol.* **2016**, *79*, 12–24. [[CrossRef](#)]
112. Escher, S.E.; Kamp, H.; Bennekou, S.H.; Bitsch, A.; Fisher, C.; Graepel, R.; Hengstler, J.G.; Herzler, M.; Knight, D.; Leist, M.; et al. Towards grouping concepts based on new approach methodologies in chemical hazard assessment: The read-across approach of the EU-ToxRisk project. *Arch. Toxicol.* **2019**, *93*, 3643–3667. [[CrossRef](#)]
113. OECD. *Case Study on the Use of Integrated Approaches to Testing and Assessment for Prediction of a 90-Day Repeated Dose Toxicity Study (OECD 408) for 2-Ethylbutyric Acid Using a Read-Across Approach to Other Branched Carboxylic Acids*; Series on Testing & Assessment No. 324. ENV/JM/MONO(2020)20; OECD Publishing: Paris, France, 2020.
114. Vrijenhoek, N.G.; Wehr, M.M.; Kunnen, S.J.; Wijaya, L.S.; Callegaro, G.; Moné, M.J.; Escher, S.E.; Van de Water, B. Application of high-throughput transcriptomics for mechanism-based biological read-across of short-chain carboxylic acid analogues of valproic acid. *ALTEX* **2022**, *39*, 207–220. [[CrossRef](#)]
115. Nakagawa, S.; Okamoto, M.; Nukada, Y.; Morita, O. Comparison of the potential mechanisms for hepatotoxicity of p-dialkoxy chlorobenzenes in rat primary hepatocytes for read-across. *Regul. Toxicol. Pharmacol.* **2020**, *113*, 104617. [[CrossRef](#)]
116. Brockmeier, E.K.; Hodges, G.; Hutchinson, T.H.; Butler, E.; Hecker, M.; Tollefsen, K.E.; Garcia-Reyero, N.; Kille, P.; Becker, D.; Chipman, K.; et al. The role of omics in the application of adverse outcome pathways for chemical risk assessment. *Toxicol. Sci.* **2017**, *158*, 252–262. [[CrossRef](#)]
117. Kaur, P.; Singh, A.; Chana, I. Computational Techniques and Tools for Omics Data Analysis: State-of-the-Art, Challenges, and Future Directions. *Arch. Comput. Methods Eng.* **2021**, *28*, 1–37. [[CrossRef](#)]
118. Blake, J.A.; Christie, K.R.; Dolan, M.E.; Drabkin, H.J.; Hill, D.P.; Ni, L.; Sitnikov, D.; Burgess, S.; Buza, T.; Gresham, C.; et al. Gene ontology consortium: Going forward. *Nucleic Acids Res.* **2015**, *43*, D1049–D1056. [[CrossRef](#)]
119. Kanehisa, M.; Goto, S.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res.* **2014**, *42*, 199–205. [[CrossRef](#)]
120. Aslam, B.; Basit, M.; Nisar, M.A.; Khurshid, M.; Rasool, M.H. Proteomics: Technologies and Their Applications. *J. Chromatogr. Sci.* **2017**, *55*, 182–196. [[CrossRef](#)]
121. Verheijen, M.C.; Meier, M.J.; Asensio, J.O.; Gant, T.W.; Tong, W.; Yauk, C.L.; Caiment, F. R-ODAF: Omics data analysis framework for regulatory application. *Regul. Toxicol. Pharmacol.* **2022**, *131*, 105143. [[CrossRef](#)]
122. Bourdon-lacombe, J.A.; Moffat, I.D.; Deveau, M.; Husain, M.; Krewski, D.; Thomas, R.S.; Bushel, P.R.; Williams, A.; Yauk, C.L.; Bureau, A.Q.; et al. Technical guide for applications of gene expression profiling in human health risk assessment of environmental chemicals. *Regul. Toxicol. Pharmacol.* **2015**, *72*, 292–309. [[CrossRef](#)]
123. Sikakana, P.; Roberts, R.A. A decade of toxicological trends: What the papers say. *Toxicol. Res. (Camb)* **2020**, *9*, 676–682. [[CrossRef](#)]
124. Verheijen, M.; Tong, W.; Shi, L.; Gant, T.W.; Seligman, B.; Caiment, F. Towards the development of an omics data analysis framework. *Regul. Toxicol. Pharmacol.* **2020**, *112*, 104621. [[CrossRef](#)]
125. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109. [[CrossRef](#)]
126. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [[CrossRef](#)]
127. Fowler, S.; Schnall, J.G. TOXNET: Information on toxicology and environmental health. *Am. J. Nurs.* **2014**, *114*, 61–63. [[CrossRef](#)]
128. Ganter, B.; Snyder, R.D.; Halbert, D.N.; Lee, M.D. Toxicogenomics in drug discovery and development: Mechanistic analysis of compound/class-dependent effects using the DrugMatrix® database. *Pharmacogenomics* **2006**, *7*, 1025–1044. [[CrossRef](#)]
129. Igarashi, Y.; Nakatsu, N.; Yamashita, T.; Ono, A.; Ohno, Y.; Urushidani, T.; Yamada, H. Open TG-GATES: A large-scale toxicogenomics database. *Nucleic Acids Res.* **2015**, *43*, D921–D927. [[CrossRef](#)]
130. Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.-P.; Subramanian, A.; Ross, K.N.; et al. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* **2006**, *313*, 1929–1935. [[CrossRef](#)]
131. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171*, 1437–1452.e17. [[CrossRef](#)]
132. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; McMorran, R.; Wieggers, J.; Wieggers, T.C.; Mattingly, C.J. The Comparative Toxicogenomics Database: Update 2019. *Nucleic Acids Res.* **2019**, *47*, D948–D954. [[CrossRef](#)]
133. Alexander-Dann, B.; Pruteanu, L.L.; Oerton, E.; Sharma, N.; Berindan-Neagoe, I.; Módos, D.; Bender, A. Developments in toxicogenomics: Understanding and predicting compound-induced toxicity from gene expression data. *Mol. Omics* **2018**, *14*, 218–236. [[CrossRef](#)]
134. Helikar, T.; Kowal, B.; McClenathan, S.; Bruckner, M.; Rowley, T.; Madrahimov, A.; Wicks, B.; Shrestha, M.; Limbu, K.; Rogers, J.A. The Cell Collective: Toward an open and collaborative approach to systems biology. *BMC Syst. Biol.* **2012**, *6*, 96. [[CrossRef](#)]

135. Freeberg, M.A.; Fromont, L.A.; D'Altri, T.; Romero, A.F.; Ciges, J.I.; Jene, A.; Kerry, G.; Moldes, M.; Ariosa, R.; Bahena, S.; et al. The European Genome-phenome Archive in 2021. *Nucleic Acids Res.* **2022**, *50*, D980–D987. [[CrossRef](#)]
136. Harrison, P.W.; Ahamed, A.; Aslam, R.; Alako, B.T.F.; Burgin, J.; Buso, N.; Courtot, M.; Fan, J.; Gupta, D.; Haseeb, M.; et al. The European Nucleotide Archive in 2020. *Nucleic Acids Res.* **2021**, *49*, D82–D85. [[CrossRef](#)]
137. Cezard, T.; Cunningham, F.; Hunt, S.E.; Koylass, B.; Kumar, N.; Saunders, G.; Shen, A.; Silva, A.F.; Tsukanov, K.; Venkataraman, S.; et al. The European Variation Archive: A FAIR resource of genomic variation for all species. *Nucleic Acids Res.* **2022**, *50*, D1216–D1220. [[CrossRef](#)]
138. Parkinson, H.; Kapushesky, M.; Shojatalab, M.; Abeygunawardena, N.; Coulson, R.; Farne, A.; Holloway, E.; Kolesnykov, N.; Lilja, P.; Lukk, M.; et al. ArrayExpress—A public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* **2007**, *35*, D747–D750. [[CrossRef](#)]
139. Tryka, K.A.; Hao, L.; Sturcke, A.; Jin, Y.; Wang, Z.Y.; Ziyabari, L.; Lee, M.; Popova, N.; Sharopova, N.; Kimura, M.; et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res.* **2014**, *42*, D975–D979. [[CrossRef](#)]
140. Moreno, P.; Fexova, S.; George, N.; Manning, J.R.; Miao, Z.; Mohammed, S.; Muñoz-Pomer, A.; Fullgrabe, A.; Bi, Y.; Bush, N.; et al. Expression Atlas update: Gene and protein expression in multiple species. *Nucleic Acids Res.* **2022**, *50*, D129–D140. [[CrossRef](#)]
141. Clough, E.; Barrett, T. The Gene Expression Omnibus database. *Methods Mol. Biol.* **2016**, *1418*, 93–110. [[CrossRef](#)]
142. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2012**, *41*, D991–D995. [[CrossRef](#)]
143. Kodama, Y.; Shumway, M.; Leinonen, R.; International Nucleotide Sequence Database Collaboration. The sequence read archive: Explosive growth of sequencing data. *Nucleic Acids Res.* **2012**, *40*, D54–D56. [[CrossRef](#)]
144. Okuda, S.; Watanabe, Y.; Moriya, Y.; Kawano, S.; Yamamoto, T.; Matsumoto, M.; Takami, T.; Kobayashi, D.; Araki, N.; Yoshizawa, A.C.; et al. jPOSTrepo: An international standard data repository for proteomes. *Nucleic Acids Res.* **2017**, *45*, D1107–D1111. [[CrossRef](#)]
145. Choi, M.; Carver, J.; Chiva, C.; Tzouros, M.; Huang, T.; Tsai, T.-H.; Pullman, B.; Bernhardt, O.M.; Hüttenhain, R.; Teo, G.C.; et al. MassIVE.quant: A community resource of quantitative mass spectrometry-based proteomics datasets. *Nat. Methods* **2020**, *17*, 981–984. [[CrossRef](#)]
146. Huang, Q.; Szklarczyk, D.; Wang, M.; Simonovic, M.; von Mering, C. PaxDb 5.0: Curated Protein Quantification Data Suggests Adaptive Proteome Changes in Yeasts. *Mol. Cell. Proteom.* **2023**, *22*, 100640. [[CrossRef](#)]
147. Jones, P.; Côté, R.G.; Martens, L.; Quinn, A.F.; Taylor, C.F.; Derache, W.; Hermjakob, H.; Apweiler, R. PRIDE: A public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res.* **2006**, *34*, D659–D663. [[CrossRef](#)]
148. Vizcaino, J.A.; Deutsch, E.W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J.A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223–226. [[CrossRef](#)]
149. Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A.M.; Lieberenz, M.; Savitski, M.M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-spectrometry-based draft of the human proteome. *Nature* **2014**, *509*, 582–587. [[CrossRef](#)]
150. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. [[CrossRef](#)]
151. Kale, N.S.; Haug, K.; Conesa, P.; Jayseelan, K.; Moreno, P.; Rocca-Serra, P.; Nainala, V.C.; Spicer, R.A.; Williams, M.; Li, X.; et al. MetaboLights: An Open-Access Database Repository for Metabolomics Data. *Curr. Protoc. Bioinform.* **2016**, *53*, 14.13.1–14.13.18. [[CrossRef](#)]
152. Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi, R.; Nair, K.S.; et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* **2016**, *44*, D463–D470. [[CrossRef](#)]
153. National Research Council. *Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment*; National Academies Press: Washington, DC, USA, 2007.
154. Thomas, R.S.; Rank, D.R.; Penn, S.G.; Zastrow, G.M.; Hayes, K.R.; Hu, T.; Pande, K.; Lewis, M.; Jovanovich, S.B.; Bradfield, C.A. Application of genomics to toxicology research. *Environ. Health Perspect.* **2002**, *110*, 919–923. [[CrossRef](#)]
155. Perakakis, N.; Stefanakis, K.; Mantzoros, C.S. The role of omics in the pathophysiology, diagnosis and treatment of non-alcoholic fatty liver disease. *Metabolism* **2020**, *111*, 154320. [[CrossRef](#)]
156. Tolosa, L.; Jiménez, N.; Pérez, G.; Castell, J.V.; Gómez-Lechón, M.J.; Donato, M.T. Customised in vitro model to detect human metabolism-dependent idiosyncratic drug-induced liver injury. *Arch. Toxicol.* **2018**, *92*, 383–399. [[CrossRef](#)]
157. Fontana, R.J. Pathogenesis of idiosyncratic drug-induced liver injury and clinical perspectives. *Gastroenterology* **2014**, *146*, 914–928.e1. [[CrossRef](#)]
158. Kralj, T.; Brouwer, K.L.R.; Creek, D.J. Analytical and Omics-Based Advances in the Study of Drug-Induced Liver Injury. *Toxicol. Sci.* **2021**, *183*, 1–13. [[CrossRef](#)]
159. Zhang, B.; Xie, W.; Krasowski, M.D. PXR: A xenobiotic receptor of diverse function implicated in pharmacogenetics. *Pharmacogenomics* **2008**, *9*, 1695–1709. [[CrossRef](#)]
160. Hooper, A.J.; Adams, L.A.; Burnett, J.R. Thematic review series: Genetics of human lipid diseases—Genetic determinants of hepatic steatosis in man. *J. Lipid Res.* **2011**, *52*, 593–617. [[CrossRef](#)]

161. Anstee, Q.M.; Darlay, R.; Cockell, S.; Meroni, M.; Govaere, O.; Tiniakos, D.; Burt, A.D.; Bedossa, P.; Palmer, J.; Liu, Y.L.; et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort<sup>☆</sup>. *J. Hepatol.* **2020**, *73*, 505–515. [[CrossRef](#)]
162. Safe, S.; Han, H.; Goldsby, J.; Mohankumar, K.; Chapkin, R.S. Aryl hydrocarbon receptor (AhR) ligands as selective AhR modulators: Genomic studies. *Curr. Opin. Toxicol.* **2018**, *11*, 10–20. [[CrossRef](#)]
163. Harrill, J.; Shah, I.; Setzer, R.W.; Haggard, D.; Auerbach, S.; Judson, R.; Thomas, R.S. Considerations for strategic use of high-throughput transcriptomics chemical screening data in regulatory decisions. *Curr. Opin. Toxicol.* **2019**, *15*, 64–75. [[CrossRef](#)]
164. Gant, T.W.; Sauer, U.G.; Zhang, S.D.; Chorley, B.N.; Hackermüller, J.; Perdichizzi, S.; Tollefsen, K.E.; van Ravenzwaay, B.; Yauk, C.; Tong, W.; et al. A generic Transcriptomics Reporting Framework (TRF) for ‘omics data processing and analysis. *Regul. Toxicol. Pharmacol.* **2017**, *91*, S36–S45. [[CrossRef](#)]
165. Perkins, E.J.; Chipman, J.K.; Edwards, S.; Habib, T.; Falciani, F.; Taylor, R.; Van Aggelen, G.; Vulpe, C.; Antczak, P.; Loguinov, A. Reverse engineering adverse outcome pathways. *Environ. Toxicol. Chem.* **2011**, *30*, 22–38. [[CrossRef](#)]
166. Moreau, A.; Téruel, C.; Beylot, M.; Albalea, V.; Tamasi, V.; Umbdenstock, T.; Parmentier, Y.; Sa-Cunha, A.; Suc, B.; Fabre, J.M.; et al. A novel pregnane X receptor and S14-mediated lipogenic pathway in human hepatocyte. *Hepatology* **2009**, *49*, 2068–2079. [[CrossRef](#)]
167. Abedini, J.A.; Handa, S.; Edwards, S.; Chorley, B.; El-Masri, H. Identification of differentially expressed genes and networks related to hepatic lipid dysfunction. *Toxicol. Appl. Pharmacol.* **2019**, *382*, 114757. [[CrossRef](#)]
168. AbdulHameed, M.D.M.; Pannala, V.R.; Wallqvist, A. Mining Public Toxicogenomic Data Reveals Insights and Challenges in Delineating Liver Steatosis Adverse Outcome Pathways. *Front. Genet.* **2019**, *10*, 1–14. [[CrossRef](#)]
169. Rex, R. Machine Learning from Omics Data. In *Artificial Intelligence in Drug Design*; Heifetz, A., Ed.; Springer Science+Business Media, LLC, part of Springer Nature: Berlin/Heidelberg, Germany, 2022; pp. 421–431.
170. Carbonell, P.; López, O.; Amberg, A.; Pastor, M.; Sanz, F. Hepatotoxicity Prediction by Systems Biology Modeling of Disturbed Metabolic Pathways Using Gene Expression Data. *ALTEX* **2017**, *34*, 219–234. [[CrossRef](#)]
171. Zhang, J.; Yang, Y.; Wang, Z.; Zhang, X.; Zhang, Y.; Lin, J.; Du, Y.; Wang, S.; Si, D.; Bao, J.; et al. Integration of Metabolomics, Lipidomics, and Proteomics Reveals the Metabolic Characterization of Nonalcoholic Steatohepatitis. *J. Proteome Res.* **2023**, *22*, 2577–2592. [[CrossRef](#)]
172. Han, X.; Aslanian, A.; Yates, J.R. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **2008**, *12*, 483–490. [[CrossRef](#)]
173. Timp, W.; Timp, G. Beyond mass spectrometry, the next step in proteomics. *Sci. Adv.* **2020**, *6*, eaax8978. [[CrossRef](#)]
174. Collins, B.C.; Lau, T.Y.K.; Pennington, S.R.; Gallagher, W.M. Differential Proteomics Incorporating iTRAQ Labeling and Multi-dimensional Separations. In *Drug Safety Evaluation: Methods and Protocols*; Gautier, J., Ed.; Springer Science+Business Media: New York, NY, USA, 2011; pp. 369–383.
175. Rabilloud, T.; Lescuyer, P. Proteomics in mechanistic toxicology: History, concepts, achievements, caveats, and potential. *Proteomics* **2015**, *15*, 1051–1074. [[CrossRef](#)]
176. Deng, Z.; Yan, S.; Hu, H.; Duan, Z.; Yin, L.; Liao, S.; Sun, Y.; Yin, D.; Li, G. Proteomic profile of carbonylated proteins in rat liver: Discovering possible mechanisms for tetracycline-induced steatosis. *Proteomics* **2015**, *15*, 148–159. [[CrossRef](#)]
177. Braeuning, A.; Oberemm, A.; Heise, T.; Gundert-Remy, U.; Hengstler, J.G.; Lampen, A. In vitro proteomic analysis of methapyrirene toxicity in rat hepatocytes reveals effects on intermediary metabolism. *Arch. Toxicol.* **2019**, *93*, 369–383. [[CrossRef](#)]
178. Araújo, A.M.; Carvalho, F.; De Pinho, P.G.; Carvalho, M. Toxicometabolomics: Small molecules to answer big toxicological questions. *Metabolites* **2021**, *11*, 692. [[CrossRef](#)]
179. Kosmidis, A.K.; Kamisoglu, K.; Calvano, S.E.; Corbett, S.A.; Androulakis, I.P. Metabolomic fingerprinting: Challenges and opportunities. *Crit. Rev. Biomed. Eng.* **2013**, *41*, 205–221. [[CrossRef](#)]
180. Reily, M.D.; Tymiak, A.A. Metabolomics in the pharmaceutical industry. *Drug Discov. Today Technol.* **2015**, *13*, 25–31. [[CrossRef](#)]
181. Bouhifd, M.; Hartung, T.; Hogberg, H.T.; Kleensang, A.; Zhao, L. Review: Toxicometabolomics. *J. Appl. Toxicol.* **2013**, *33*, 1365–1383. [[CrossRef](#)]
182. Bernhard, A.; Poulsen, R.; Brun Hansen, A.M.; Hansen, M. Toxicometabolomics as a tool for next generation environmental risk assessment. *EFSA J.* **2023**, *21*, 1–12. [[CrossRef](#)]
183. Cuykx, M.; Claes, L.; Rodrigues, R.M.; Vanhaecke, T.; Covaci, A. Metabolomics profiling of steatosis progression in HepaRG<sup>®</sup> cells using sodium valproate. *Toxicol. Lett.* **2018**, *286*, 22–30. [[CrossRef](#)]
184. Yong, H.Y.; Larrouy-Maumus, G.; Zloh, M.; Smyth, R.; Ataya, R.; Benton, C.M.; Munday, M.R. Early detection of metabolic changes in drug-induced steatosis using metabolomics approaches. *RSC Adv.* **2020**, *10*, 41047–41057. [[CrossRef](#)]
185. García-Cañaveras, J.C.; Jiménez, N.; Gómez-Lechón, M.J.; Castell, J.V.; Donato, M.T.; Lahoz, A. LC-MS untargeted metabolomic analysis of drug-induced hepatotoxicity in HepG2 cells. *Electrophoresis* **2015**, *36*, 2294–2302. [[CrossRef](#)]
186. García-Cañaveras, J.C.; Castell, J.V.; Donato, M.T.; Lahoz, A. A metabolomics cell-based approach for anticipating and investigating drug-induced liver injury. *Sci. Rep.* **2016**, *6*, 1–12. [[CrossRef](#)]
187. Martínez-Sena, T.; Moro, E.; Moreno-Torres, M.; Quintás, G.; Hengstler, J.; Castell, J.V.; Martínez-sena, T.; Moro, E.; Moreno-Torres, M.; Quintas, G.; et al. Metabolomics-based strategy to assess drug hepatotoxicity and uncover the mechanisms of hepatotoxicity involved. *Arch. Toxicol.* **2023**, *97*, 1723–1738. [[CrossRef](#)]
188. Kang, M.; Ko, E.; Mersha, T.B. A roadmap for multi-omics data integration using deep learning. *Brief. Bioinform.* **2022**, *23*, 1–16. [[CrossRef](#)]

189. Yahya, F.A.; Hashim, N.F.M.; Israf Ali, D.A.; Chau Ling, T.; Cheema, M.S. A brief overview to systems biology in toxicology: The journey from in to vivo, in-vitro and -omics. *J. King Saud. Univ. Sci.* **2021**, *33*, 101254. [[CrossRef](#)]
190. Subramanian, I.; Verma, S.; Kumar, S.; Jere, A.; Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform. Biol. Insights* **2020**, *14*, 7–9. [[CrossRef](#)]
191. Flores, J.E.; Claborne, D.M.; Weller, Z.D.; Waters, K.M.; Bramer, L.M. Missing data in multi-omics integration: Recent advances through artificial intelligence. *Front. Artif. Intel.* **2013**, *6*, 1098308. [[CrossRef](#)]
192. Picard, M.; Scott-Boyer, M.P.; Bodein, A.; Périn, O.; Droit, A. Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3735–3746. [[CrossRef](#)]
193. Chicco, D.; Cumbo, F.; Angione, C. Ten quick tips for avoiding pitfalls in multi-omics data integration analyses. *PLoS Comput. Biol.* **2023**, *19*, 1–15. [[CrossRef](#)]
194. Canzler, S.; Schor, J.; Busch, W.; Schubert, K.; Rolle-Kampczyk, U.E.; Seitz, H.; Kamp, H.; von Bergen, M.; Buesen, R.; Hackermüller, J. Prospects and challenges of multi-omics data integration in toxicology. *Arch. Toxicol.* **2020**, *94*, 371–388. [[CrossRef](#)]
195. Ruepp, S.U.; Tonge, R.P.; Shaw, J.; Wallis, N.; Pognan, F. Genomics and proteomics analysis of acetaminophen toxicity in mouse liver. *Toxicol. Sci.* **2002**, *65*, 135–150. [[CrossRef](#)]
196. van Breda, S.G.J.; Claessen, S.M.H.; van Herwijnen, M.; Theunissen, D.H.J.; Jennen, D.G.J.; de Kok, T.M.C.M.; Kleinjans, J.C.S. Integrative omics data analyses of repeated dose toxicity of valproic acid in vitro reveal new mechanisms of steatosis induction. *Toxicology* **2018**, *393*, 160–170. [[CrossRef](#)]
197. Wolters, J.E.J.J.; van Breda, S.G.J.J.; Grossmann, J.; Fortes, C.; Caiment, F.; Kleinjans, J.C.S.S. Integrated 'omics analysis reveals new drug-induced mitochondrial perturbations in human hepatocytes. *Toxicol. Lett.* **2018**, *289*, 1–13. [[CrossRef](#)]
198. Zhang, X.; Li, X.; Xiong, X. Applying proteomics in metabolic dysfunction-associated steatotic liver disease: From mechanism to biomarkers. *Clin. Res. Hepatol. Gastroenterol.* **2023**, *47*, 102230. [[CrossRef](#)]
199. Wetmore, B.A.; Merrick, B.A. Toxicoproteomics: Proteomics applied to toxicology and pathology. *Toxicol. Pathol.* **2004**, *32*, 619–642. [[CrossRef](#)]
200. Kell, D.B.; Oliver, S.G. The metabolome 18 years on: A concept comes of age. *Metabolomics* **2016**, *12*, 1–8. [[CrossRef](#)]
201. Hartung, T. Utility of the adverse outcome pathway concept in drug development. *Expert. Opin. Drug Metab. Toxicol.* **2017**, *13*, 1–3. [[CrossRef](#)]
202. Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949. [[CrossRef](#)]
203. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput.-Aided Drug Design. Curr. Comput. Aided Drug Des.* **2011**, *7*, 146–157. [[CrossRef](#)]
204. Pinzi, L.; Rastelli, G. Molecular docking: Shifting paradigms in drug discovery. *Int. J. Mol. Sci.* **2019**, *20*, 4331. [[CrossRef](#)]
205. Tao, X.; Huang, Y.; Wang, C.; Chen, F.; Yang, L.; Ling, L.; Che, Z.; Chen, X. Recent developments in molecular docking technology applied in food science: A review. *Int. J. Food Sci. Technol.* **2020**, *55*, 33–45. [[CrossRef](#)]
206. Scardino, V.; Di Filippo, J.I.; Cavasotto, C.N. How good are AlphaFold models for docking-based virtual screening? *ChemRxiv* **2022**, *26*, 105920. [[CrossRef](#)]
207. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
208. Koshland, D.E. The Key–Lock Theory and the Induced Fit Theory. *Angew. Chem. Int. Ed. Engl.* **1995**, *33*, 2375–2378. [[CrossRef](#)]
209. Steinmetz, F.P.; Mellor, C.L.; Meinel, T.; Cronin, M.T.D. Screening Chemicals for Receptor-Mediated Toxicological and Pharmacological Endpoints: Using Public Data to Build Screening Tools within a KNIME Workflow. *Mol. Inform.* **2015**, *34*, 171–178. [[CrossRef](#)]
210. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [[CrossRef](#)]
211. Kwon, S.; Seok, C. CSAlign and CSAlign-Dock: Structure alignment of ligands considering full flexibility and application to protein–ligand docking. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 1–10. [[CrossRef](#)]
212. Allen, W.J.; Balias, T.E.; Mukherjee, S.; Brozell, S.R.; Moustakas, D.T.; Lang, P.T.; Case, D.A.; Kuntz, I.D.; Rizzo, R.C. DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015**, *36*, 1132–1156. [[CrossRef](#)]
213. Bikadi, Z.; Hazai, E. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminform.* **2009**, *1*, 1–16. [[CrossRef](#)]
214. Yang, J.; Baek, M.; Seok, C. GalaxyDock3: Protein–ligand docking that considers the full ligand conformational flexibility. *J. Comput. Chem.* **2019**, *40*, 2739–2748. [[CrossRef](#)]
215. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748. [[CrossRef](#)]
216. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196. [[CrossRef](#)]
217. Dominguez, C.; Boelens, R.; Bonvin, A.M.J.J. HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. [[CrossRef](#)]

218. Wang, J.; Dokholyan, N. V. MedusaDock 2.0: Efficient and Accurate Protein–Ligand Docking With Constraints. *J. Chem. Inf. Model.* **2019**, *59*, 2509–2515. [CrossRef]
219. Jiménez-García, B.; Pons, C.; Fernández-Recio, J. pyDockWEB: A web server for rigid-body protein–protein docking using electrostatics and desolvation scoring. *Bioinformatics* **2013**, *29*, 1698–1699. [CrossRef]
220. Lyskov, S.; Gray, J.J. The RosettaDock server for local protein–protein docking. *Nucleic Acids Res.* **2008**, *36*, W233–W238. [CrossRef]
221. Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein–small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **2011**, *39*, W270–W277. [CrossRef]
222. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef]
223. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688. [CrossRef]
224. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [CrossRef]
225. Brooks, B.R.; Brooks, C.L.; Mackerell, A.D.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [CrossRef]
226. Bowers, K.J.; Sacerdoti, F.D.; Salmon, J.K.; Shan, Y.; Shaw, D.E.; Chow, E.; Xu, H.; Dror, R.O.; Eastwood, M.P.; Gregersen, B.A.; et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, Tampa, FL, USA, 11–17 November 2006; p. 43.
227. Sakkiah, S.; Kusko, R.; Tong, W.; Hong, H. Applications of Molecular Dynamics Simulations in Computational Toxicology. In *Advances in Computational Toxicology*; Hong, H., Ed.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 181–212. [CrossRef]
228. Selvaraj, C.; Sakkiah, S.; Tong, W.; Hong, H. Molecular dynamics simulations and applications in computational toxicology and nanotoxicology. *Food Chem. Toxicol.* **2018**, *112*, 495–506. [CrossRef]
229. Negi, C.K.; Bajard, L.; Kohoutek, J.; Blaha, L. An adverse outcome pathway based in vitro characterization of novel flame retardants-induced hepatic steatosis. *Environ. Pollut.* **2021**, *289*, 117855. [CrossRef]
230. Tsakovska, I.; Sharif, M.A.; Alov, P.; Diukendjieva, A.; Fioravanzo, E.; Cronin, M.T.D.; Pajeva, I. Molecular modelling study of the PPAR $\gamma$  receptor in relation to the mode of action/adverse outcome pathway framework for liver steatosis. *Int. J. Mol. Sci.* **2014**, *15*, 7651–7666. [CrossRef]
231. Alemán-González-Duhart, D.; Tamay-Cach, F.; Correa-Basurto, J.; Padilla-Martínez, I.I.; Álvarez-Almazán, S.; Mendieta-Wejebe, J.E. In silico design, chemical synthesis and toxicological evaluation of 1,3-thiazolidine-2,4-dione derivatives as PPAR $\gamma$  agonists. *Regul. Toxicol. Pharmacol.* **2017**, *86*, 25–32. [CrossRef]
232. Mornagui, B.; Rezg, R.; Neffati, F.; Najjar, M.F.; Rejeb, A. Postnatal exposure to Bisphenol S induces liver injury in mice: Possible implication of PPAR $\gamma$  receptor. *Toxicol. Ind. Health* **2023**, *39*, 237–247. [CrossRef]
233. Rosell-Hidalgo, A.; Moore, A.L.; Ghafourian, T. Prediction of drug-induced mitochondrial dysfunction using succinate-cytochrome c reductase activity, QSAR and molecular docking. *Toxicology* **2023**, *485*, 153412. [CrossRef]
234. Troger, F.; Delp, J.; Funke, M.; van der Stel, W.; Colas, C.; Leist, M.; van de Water, B.; Ecker, G.F. Identification of mitochondrial toxicants by combined in silico and in vitro studies—A structure-based view on the adverse outcome pathway. *Comput. Toxicol.* **2020**, *14*, 100123. [CrossRef]
235. Kolšek, K.; Mavri, J.; Sollner Dolenc, M.; Gobec, S.; Turk, S. Endocrine disruptome—An open source prediction tool for assessing endocrine disruption potential through nuclear receptor binding. *J. Chem. Inf. Model.* **2014**, *54*, 1254–1267. [CrossRef]
236. Burello, E. Review of (Q)SAR models for regulatory assessment of nanomaterials risks. *NanoImpact* **2017**, *8*, 48–58. [CrossRef]
237. Benfenati, E. The specificity of the QSAR models for regulatory purposes: The example of the DEMETRA project. *SAR QSAR Environ. Res.* **2007**, *18*, 209–220. [CrossRef]
238. Limban, C.; Nuță, D.C.; Chiriță, C.; Negreș, S.; Arsene, A.L.; Goumenou, M.; Karakitsios, S.P.; Tsatsakis, A.M.; Sarigiannis, D.A. The use of structural alerts to avoid the toxicity of pharmaceuticals. *Toxicol. Rep.* **2018**, *5*, 943–953. [CrossRef]
239. OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*; OECD Publishing: Paris, France, 2014.
240. OECD. OECD Omics Reporting Framework (OORF): Guidance on Reporting Elements for the Regulatory Use of Omics Data from Laboratory-Based Toxicology Studies. 2023. Available online: <https://doi.org/10.1787/6bb2e6ce-e> (accessed on 14 January 2024).
241. Ankley, G.T.; Cureton, P.; Hoke, R.A.; Houde, M.; Kumar, A.; Kurias, J.; Lanno, R.; McCarthy, C.; Newsted, J.; Salice, C.J.; et al. Assessing the Ecological Risks of Per- and Polyfluoroalkyl Substances: Current State-of-the Science and a Proposed Path Forward. *Environ. Toxicol. Chem.* **2021**, *40*, 564–605. [CrossRef]
242. Cavaliere, F.; Lorenzetti, S.; Cozzini, P. Molecular modelling methods in food safety: Bisphenols as case study. *Food Chem. Toxicol.* **2020**, *137*, 111116. [CrossRef]
243. Holmer, M.; de Bruyn Kops, C.; Stork, C.; Kirchmair, J. CYPstrate: A Set of Machine Learning Models for the Accurate Classification of Cytochrome P450 Enzyme Substrates and Non-Substrates. *Molecules* **2021**, *26*, 4678. [CrossRef]
244. Plonka, W.; Stork, C.; Šicho, M.; Kirchmair, J. CYPlebrity: Machine learning models for the prediction of inhibitors of cytochrome P450 enzymes. *Bioorg. Med. Chem.* **2021**, *46*, 116388. [CrossRef]
245. Novoradovskaya, N.; Whitfield, M.L.; Basehore, L.S.; Novoradovsky, A.; Pesich, R.; Usary, J.; Karaca, M.; Wong, W.K.; Aprelikova, O.; Fero, M.; et al. Universal reference RNA as a standard for microarray experiments. *BMC Genom.* **2004**, *5*, 1–13. [CrossRef]

246. Lu, S.; Lu, H.; Zheng, T.; Yuan, H.; Du, H.; Gao, Y.; Liu, Y.; Pan, X.; Zhang, W.; Fu, S.; et al. A multi-omics dataset of human transcriptome and proteome stable reference. *Sci. Data* **2023**, *10*, 1–12. [[CrossRef](#)]
247. Hardy, A.; Benford, D.; Halldorsson, T.; Jeger, M.J.; Knutsen, H.K.; More, S.; Naegeli, H.; Noteborn, H.; Ockleford, C.; Ricci, A.; et al. Guidance on the use of the weight of evidence approach in scientific assessments. *EFSA J.* **2017**, *15*, 4971. [[CrossRef](#)]
248. Benfenati, E.; Chaudhry, Q.; Gini, G.; Dorne, J.L. Integrating in silico models and read-across methods for predicting toxicity of chemicals: A step-wise strategy. *Environ. Int.* **2019**, *131*, 105060. [[CrossRef](#)]
249. OECD. Users' Handbook Supplement to the Guidance Document for Developing and Assessing Adverse Outcome Pathways. 2018. Available online: <https://doi.org/10.1787/5jlv1m9d1g32-en> (accessed on 14 January 2024).
250. Ahlberg, E.; Amberg, A.; Beilke, L.D.; Bower, D.; Cross, K.P.; Custer, L.; Ford, K.A.; Van Gompel, J.; Harvey, J.; Honma, M.; et al. Extending (Q)SARs to incorporate proprietary knowledge for regulatory purposes: A case study using aromatic amine mutagenicity. *Regul. Toxicol. Pharmacol.* **2016**, *77*, 1–12. [[CrossRef](#)]
251. Barber, C.; Antonucci, V.; Baumann, J.-C.; Brown, R.; Covey-Crump, E.; Elder, D.; Elliott, E.; Fennell, J.W.; Gallou, F.; Ide, N.D.; et al. A consortium-driven framework to guide the implementation of ICH M7 Option 4 control strategies. *Regul. Toxicol. Pharmacol.* **2017**, *90*, 22–28. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.