In silico toxicology: computational methods for the prediction of chemical toxicity

Arwa B. Raies and Vladimir B. Bajic*

Determining the toxicity of chemicals is necessary to identify their harmful effects on humans, animals, plants, or the environment. It is also one of the main steps in drug design. Animal models have been used for a long time for toxicity testing. However, *in vivo* animal tests are constrained by time, ethical considerations, and financial burden. Therefore, computational methods for estimating the toxicity of chemicals are considered useful. *In silico* toxicology is one type of toxicity assessment that uses computational methods to analyze, simulate, visualize, or predict the toxicity of chemicals. *In silico* toxicology aims to complement existing toxicity tests to predict toxicity, prioritize chemicals, guide toxicity tests, and minimize late-stage failures in drugs design. There are various methods for generating models to predict toxicity endpoints. We provide a comprehensive overview, explain, and compare the strengths and weaknesses of the existing modeling methods and algorithms for toxicity prediction with a particular (but not exclusive) emphasis on computational tools that can implement these methods and refer to expert systems that deploy the prediction models. Finally, we briefly review a number of new research directions in *in silico* toxicology and provide recommendations for designing *in silico* models. © 2016 The Authors. *WIREs Computational Molecular Science* published by John Wiley & Sons, Ltd.

INTRODUCTION

Toxicity is a measure of any undesirable or adverse effect of chemicals. Specific types of these adverse effects are called toxicity endpoints, such as carcinogenicity or genotoxicity, and can be quantitative (e.g., LD50: lethal dose to 50% of tested individuals)[1] or qualitative, such as binary (e.g., toxic or non-toxic) or ordinary (e.g., low, moderate, or high

toxicity).[2] Toxicity tests aim to identify harmful effects caused by substances on humans, animals, plants, or the environment through acute-exposure (single dose) or multiple-exposure (multiple doses).[3] Several factors determine the toxicity of chemicals, such as route of exposure (e.g., oral, dermal, inhalation), dose (amount of the chemical), frequency of exposure (e.g., single versus multiple exposure), duration of exposure (e.g., 96 h), ADME properties (absorption, distribution, metabolism, and excretion/elimination), biological properties (e.g., age, gender), and chemical properties.[4]

Animal models have been used for a long time for toxicity testing.[3] However, *in vitro* toxicity tests became plausible due to the advances in high throughput screening.[3] *In silico* toxicology (computational toxicology) is one type of toxicity assessment that uses computational resources (i.e., methods,

*Correspondence to: vladimir.bajic@kaust.edu.sa

King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Centre (CBRC), Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), Thuwal, Saudi Arabia

algorithms, software, data, etc.) to organize, analyze, model, simulate, visualize, or predict toxicity of chemicals.[5,6] It is intertwined with *in silico* pharmacology, which uses information from computational tools to analyze beneficial or adverse effects of drugs for therapeutic purposes.[5,6]

Computational methods aim to complement *in vitro* and *in vivo* toxicity tests to potentially minimize the need for animal testing, reduce the cost and time of toxicity tests, and improve toxicity prediction and safety assessment. In addition, computational methods have a unique advantage of being able to estimate chemicals for toxicity even before they are synthesized.[7] *In silico* toxicology encompasses a wide variety of computational tools (Figure 1): (A) databases for storing data about chemicals, their toxicity, and chemical properties; (B) software for generating molecular descriptors; (C) simulation tools for systems biology and molecular dynamics; (D) modeling methods for toxicity prediction; (E) modeling tools such as statistical packages and software for generating prediction models; (F) expert systems that include pre-built models in web servers or standalone applications for predicting toxicity; and (G) visualization tools.

The purpose of this study is to provide a comprehensive overview of existing modeling methods and algorithms for toxicity prediction (element D above), with a particular (but not exclusive) emphasis on computational tools that can implement these methods (element E), and expert systems that deploy the prediction models (element F). Due to the nature of this expanding field, this study cannot provide an exhaustive overview of all the seven *in silico* components mentioned above. Therefore, the reader is encouraged to refer to existing literature to get more information about toxicity databases,[6,8–11] molecular descriptors generation software,[12] toxicology simulation tools,[13,14] statistical modeling packages,[12] expert systems[6,9,11,12,15–17], and visualization tools.[18]

Generally, modeling methods include five major steps while developing prediction models[19] (Figure 1): (1) gathering biological data that contain associations between chemicals and toxicity endpoints, (2) calculating molecular descriptors of the chemicals, (3) generating a prediction model, (4) evaluating the accuracy of the model, and (5) interpreting the model.

The scope of this review covers the third step, generating prediction models. We focus on using computational methods to predict toxicity of different types of substances such as drugs, other chemicals, mixtures, and nanomaterials both quantitatively and qualitatively. There are various methods to solve
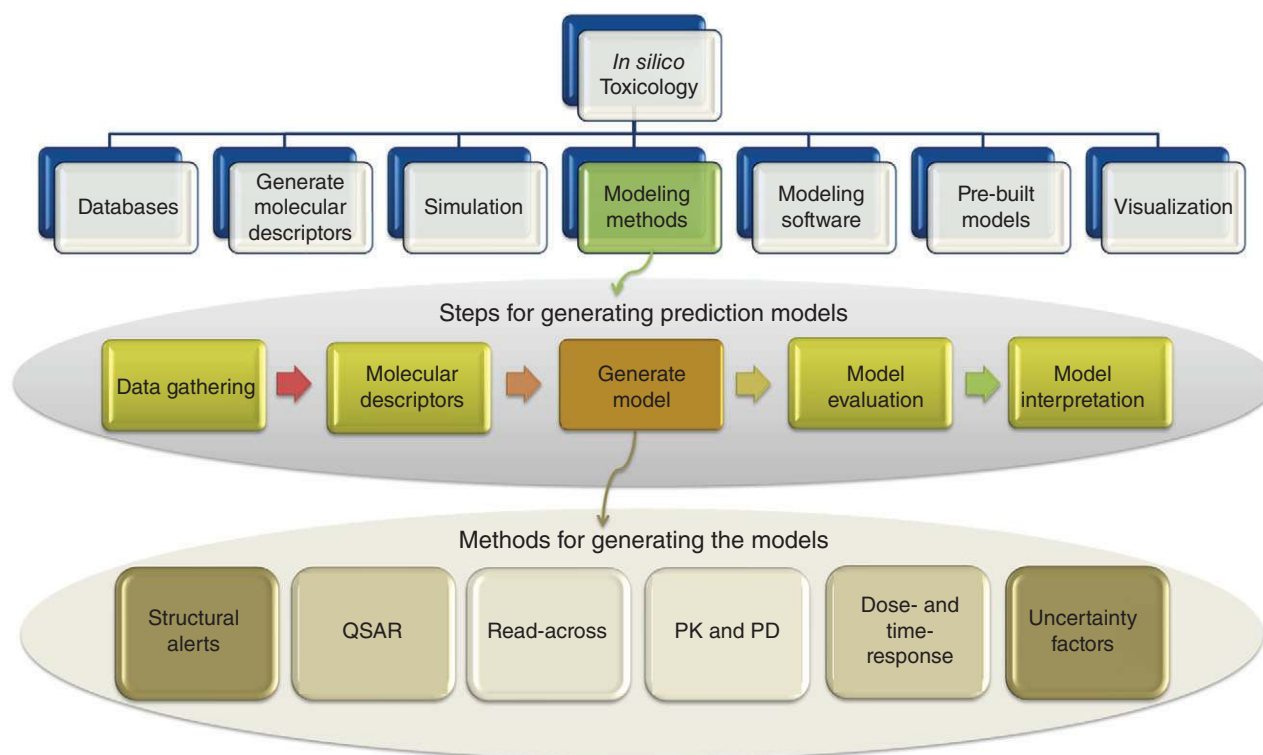
FIGURE 1 | *In silico* toxicology tools, steps to generate prediction models, and categories of prediction models.

such problems, and each method has its strengths, limitations, scope of application, and specificity of interpretation. The goal is to find the most effective method for the problem at hand. However, all the five steps mentioned above are inter-related. Therefore, we discuss the remaining steps whenever necessary. We divided this review into four main sections. First, we provide an explanation and discussion of the existing *in silico* methods. Then, we briefly discuss two special types of chemicals: mixtures and nanomaterials. Subsequently, we offer recommendations on how to develop and apply toxicity prediction models. Finally, we provide an overview of 21st century toxicology.

IN SILICO MODELING METHODS

Many *in silico* methods have been developed to predict the toxicity of chemicals. The methods we discuss here are chosen either because they illustrate the historical development of *in silico* toxicology or they represent the state-of-the-art method for predicting toxicity. For each method, we provide (if applicable) a mathematical description, discussion of strengths and limitations, recommendations about when and why to use the method, and existing tools that implement the method. Additionally, for the sake of clarity, we keep equations and visual representations of models as general as possible.

Structural Alerts and Rule-based Models

Structural alerts (SAs)[12,20] (also called toxicophores/toxic fragments[17]) are chemical structures that indicate or associate to toxicity.[6,12] SAs can consist of only one atom or several connected atoms.[21] A combination of SAs may contribute to toxicity more than a single SA.[21] SAs are often used in rules defined in the form 'if A is B then T,' where A is an SA, B is the value of the SA, and T is the toxicity prediction with assigned certainty level,[6] as illustrated in the following example:

IF (chemical_substructure) IS (*present*) THEN (skin_sensitizer IS *certain*)

There are two main types of rule-based models that we will consider: human-based rules (HBRs) and induction-based rules (IBRs).[12] HBRs are derived from human knowledge of field experts or from literature, but IBRs are derived computationally.[6,12] HBRs are more accurate but are limited to human knowledge that could be incomplete or biased.[6,12] Moreover, updating HBRs is often impractical as it requires detailed literature analysis.[21] On the contrary, IBRs can be generated efficiently from large datasets. IBRs may propose hypotheses about associations between chemical structural properties (or their combinations) and toxicity endpoints, which may not be identified through human insights.[6,21] IBRs are implemented using probabilities to determine if SAs correspond to the toxic or non-toxic class. It is possible to have hybrid-based rules systems that contain IBRs and HBRs, with new rules being generated computationally.[12]

It is easy to interpret and implement SAs.[15] They are useful in drug design to determine how drugs should be altered to reduce their toxicity. Using structure to predict toxicity allows identifying the structure of potential metabolites.[10] However, SAs have a number of limitations. SAs use only binary features (e.g., chemical structures are either present or absent) and only qualitative endpoints (e.g., carcinogenic or non-carcinogenic).[12] SAs do not provide insights into the biological pathways of toxicity and may not be sufficient for predicting toxicity. Depending on the concurrent absence or presence of other chemical properties, toxicity may decrease or increase.[15] The list of SAs and rules may be incomplete, which may cause a large number of false negatives (i.e., toxic chemicals predicted as non-toxic) in predictions.[12,15,20]

The last point is particularly important. It is necessary to understand how to interpret the output of SA models. If a chemical does not include SAs or does not match any toxicity rules, this does not indicate non-toxicity.[6] This is especially true for HBRs that usually include SAs or rules that indicate toxicity but do not include SAs or rules that indicate non-toxicity.[6] Therefore, in developing such models, it is necessary to ensure that the list of SAs and rules are comprehensive and that they are refined when more experimental data becomes available. However, there should be a balance between the list of SAs and rules, their comprehensiveness, and predictive power. If SAs and rules are diverse, they can be applied to a large number of chemicals, but this may increase false positives (i.e., non-toxic chemicals predicted as toxic). However, if they are too narrow, they can be applied only to a small group of chemicals, and this may increase false negatives (i.e., toxic chemicals predicted as non-toxic).

An example of SA list for skin sensitization was published in 1982 by Dupuis and Benezra.[22] Another SA list was proposed by Ashby and Tennant[19,21,23] in 1988 to predict carcinogenicity and mutagenicity. One of the most developed lists of carcinogenic SAs was proposed by Benigni and Bossa[19,21,24] in

the Organisation for Economic Co-operation and Development (OECD) Quantitative Structure-Activity Relationship (OSAR) Toolbox[25] and in Tox-tree.[19,21,26] Recently, Benigni and Bossa published a new list of non-genotoxic carcinogenic SAs.[27] Other lists of SAs and rule-based models were developed for hepatotoxicity,[28] cytotoxicity,[29] irritation/corro-sion of skin,[30] and eye[31] and skin sensitization.[19,32]

Several systems (also called 'expert systems'[20]) provide pre-built rule-based and SAs lists, for exam-ple, Oncologic Cancer Expert System (OCES),[33] Toxtree,[15,26] Derek Nexus,[15,34] HazardExpert[15,35] and Meteor.[36] Other tools that can extract SAs from datasets that contains toxic or non-toxic chemicals are reviewed in[21], such as computer assisted structure elucidation (CASE),[37] prediction of activity spectra for substances (PASS)[38], and categorical-structure activity relationship (cat-SAR).[39]

Additionally, there are several approaches for extracting the longest frequent molecular substruc-tures such as Apriori (based on breadth-first search) and pattern growth (based on depth-first search).[21] Examples of algorithms that implement the pattern growth approach are reviewed in[21], such as molecu-lar fragment miner (mofa),[40] graph-based substruc-ture pattern mining (gSpan),[41] fast frequent subgraph mining (FFSM),[42] and Graph/Sequence/Tree Extrac-tion (gaston).[43] Significant substructures capable of discriminating between toxic and non-toxic chemi-cals can be extracted using an emerging chemical pat-tern approach as explained in ref. 21.

Chemical Category, Read-Across, and Trend Analysis

A chemical category[44] is a group of chemicals whose properties and toxicity effects are similar or follow a similar pattern.[11,45] Chemicals in the category are also called source chemicals. The *OECD Guidance On Grouping Of Chemicals* lists several methods for grouping, such as chemical identity and composition, physicochemical and ADME properties, mechanism of action (MoA), and chemical/biological interac-tions.[46] Structural similarity is described in the OECD guidelines as the starting point for grouping, but it is also criticized for lacking a 'scientifically sup-portable basis' for grouping, and it can be used if impurities or other constituents in the chemical com-position would not change toxicity.[46]

Read-across is a method of predicting unknown toxicity of a chemical using similar chemicals (called chemical analogs) with known toxicity from the same chemical category.[9,11,12,45,47] Trend analysis is a method of predicting toxicity of a chemical by ana-lyzing toxicity trends (increase, decrease, or constant) of tested chemicals.[9] A hypothetical example of trend analysis shows that when carbon chain length (CCL) increases, acute aquatic toxicity increases (Figure 2).[45]

Here, we focus on the read-across method. A summary of different parameters that must be con-sidered when designing a read-across model is depicted in Figure 3 and explained later. Note,
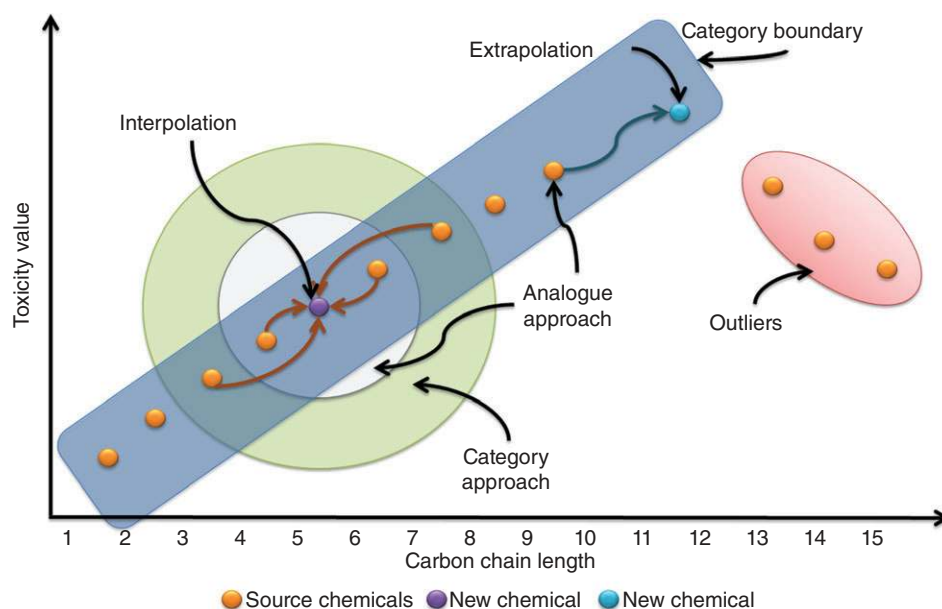
FIGURE 2 | Different approaches of read-across: analog versus category approaches, interpolation versus extrapolation, category boundary and outliers.
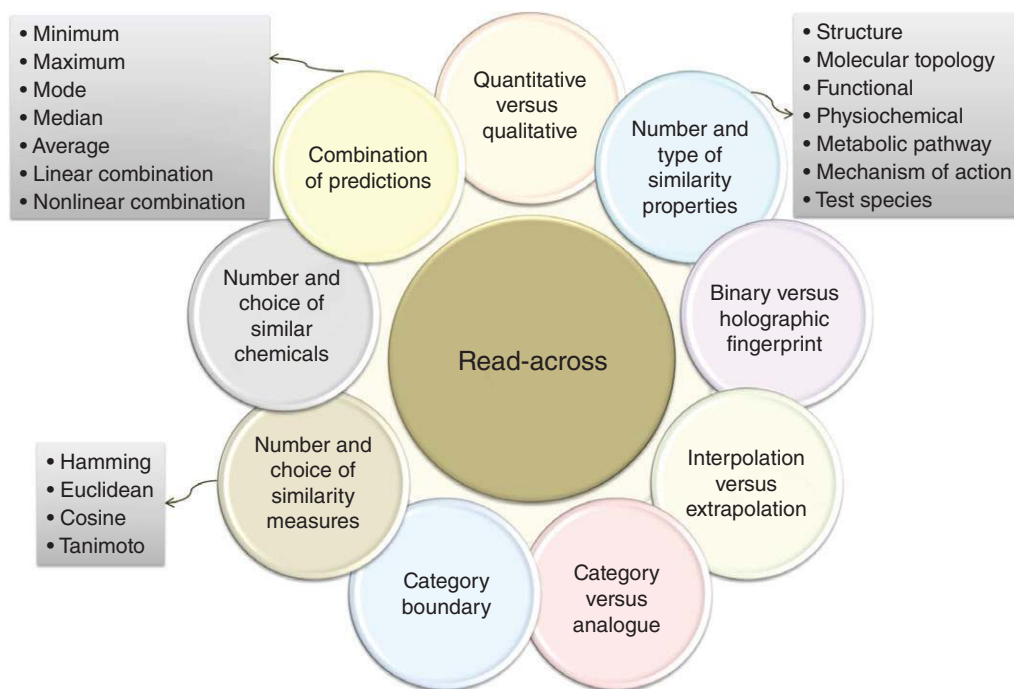
FIGURE 3 | Different properties of read-across models.

however, that the points discussed are similarly applicable to trend analysis.

There are two ways to develop a read-across method[12,45,48]: analog approach (AN) (called one-to-one), which uses one or few analogs, and a category approach (CA) (called many-to-one), which uses many analogs. AN may be sensitive to outliers because two analogs may have different toxicity profiles.[12] Using many analogs for CA is useful to detect trends within a category and may increase confidence in the toxicity predictions.[11,45,48] CA requires defining a category boundary to determine if a chemical belongs to the category[45] and implementing a 'combination of predictions' method for analogs that have conflicting toxicity profiles. A combination of predictions can be done using (if applicable) minimum, maximum, mode, median, average, linear, quadratic, or other nonlinear combinations of the predictions.[47]

Read-across can be qualitative if the toxicity endpoint is qualitative; otherwise, read-across is quantitative.[6,9,12] Also, interpolation using source chemicals surrounding the target chemical (see Figure 2) is better than extrapolation from one side.[17] In Figure 2, interpolation is used with the chemical that has CCL of length 6, but extrapolation is used with a chemical that has CCL of length 12.

Identifying similar chemicals can be done in two steps: representing chemicals as feature vectors of chemical properties, and then calculating similarity of chemicals. The first step is implemented using either binary or holographic fingerprints. A binary fingerprint is a feature vector of binary bits representing presence (1) or absence (0) of a property (e.g. presence of a methyl group).[44,47] However, a holographic fingerprint uses frequency of properties (e.g. number of methyl groups). Continuous chemical properties (e.g., melting point) can be used as well. A hierarchy of categories and subcategories can be better than a single feature vector. At each level of the hierarchy, a property is applied for category formation. Subsequently, categories are divided using another property to generate subcategories and so on. The hierarchy can allow for investigating the significance of properties and can simplify model interpretation.[47] An example of hierarchal categories is provided in ref. 47. Statistical similarity of two chemicals can be calculated using different types of distances, such as Hamming, Euclidean, Cosine, Mahalanobis, Tanimoto distance, or linear or nonlinear relationships of the features.[45,47]

There are several advantages of read-across. Read-across is transparent,[16] easy to interpret and implement.[44] Read-across can model quantitative and qualitative toxicity endpoints, and it allows for a wide range of types of descriptors and similarity measures to be used to express similarity between chemicals.[47]

However, there are also limitations. Statistical similarity measures do not provide biological insight of toxicity.[47] Moreover, complex similarity measures may complicate model interpretation.[47] In reality, read-across uses small datasets compared to other approaches such as QSAR because there are usually only a few analogs for a given chemical.[47] Additionally, accuracy depends on the number and choice of analogs, similarity metrics, strength in chemicals' similarity, chemical properties, and category boundaries.[47] These parameters are very subjective, mutually dependent, endpoint-specific, and may require expert opinions.[17,44–47] Moreover, this approach could be inapplicable or inaccurate if analogs have conflicting toxicity profiles[11] or the number of analog chemicals is insufficient. In these cases, the QSAR approach can be used.[11,12,44,45]

Read-across was applied to predict carcinogenicity,[49] hepatoxicity,[28] aquatic toxicity,[50] reproductive toxicity,[51] skin sensitization,[52] and environmental toxicity.[53] Examples of tools implementing read-across are The OECD QSAR Toolbox,[25] Toxmatch,[54] ToxTree,[26] AMBIT,[55] AmbitDiscovery,[56] AIM,[57] DSSTox,[58] or ChemIDplus.[59] A detailed explanation of some of these tools is available in refs. 6,9,11,16,17,44,45,48.

Dose–Response and Time–Response Models

Dose–response (or time–response) models are relationships between doses (or time) and the incidence of a defined biological effect (e.g., toxicity or mortality).[60] A dose is 'the total quantity of a substance administered to, taken up, or absorbed by an organism, organ, or tissue and can be measured with *in vitro* or *in vivo* experiments.'[60] Time can be the time to produce a response or the time for recovery.[61] Exposure time can be continuous, intermittent, or random, and exposure can be acute, short-term, sub-chronic , and chronic exposure.[60] Time–dose models describe the relationship between time and dose for a constant response.[62] Figure 4 shows different types of dose/time–response models. These models that describe relationships between response versus dose or time can be generated by regression to fit the data.

The first dose–response model relates concentration (C) and time (t) with response (K), which is Haber's law (law of toxicity)[60,63]:

$$C \times t = K$$
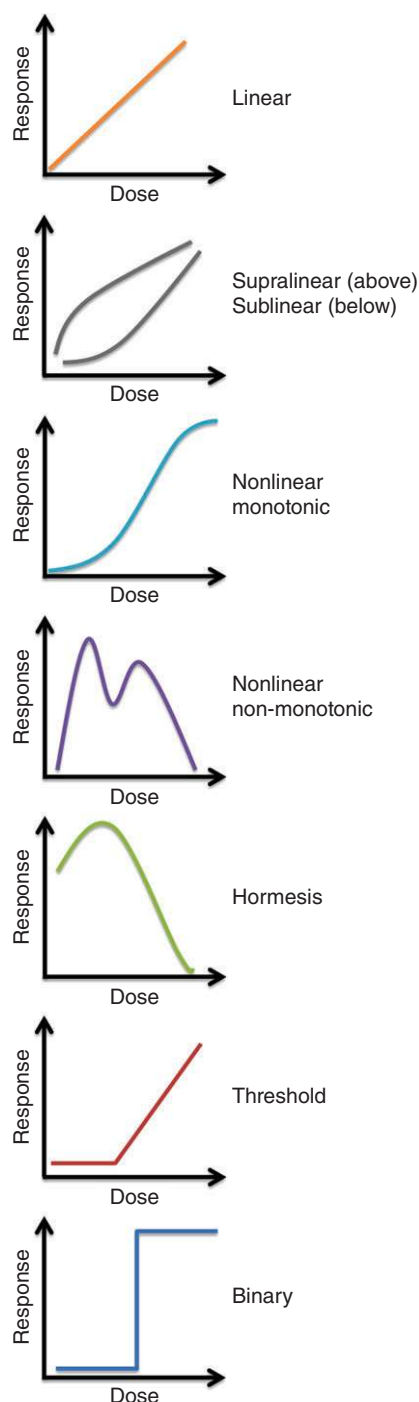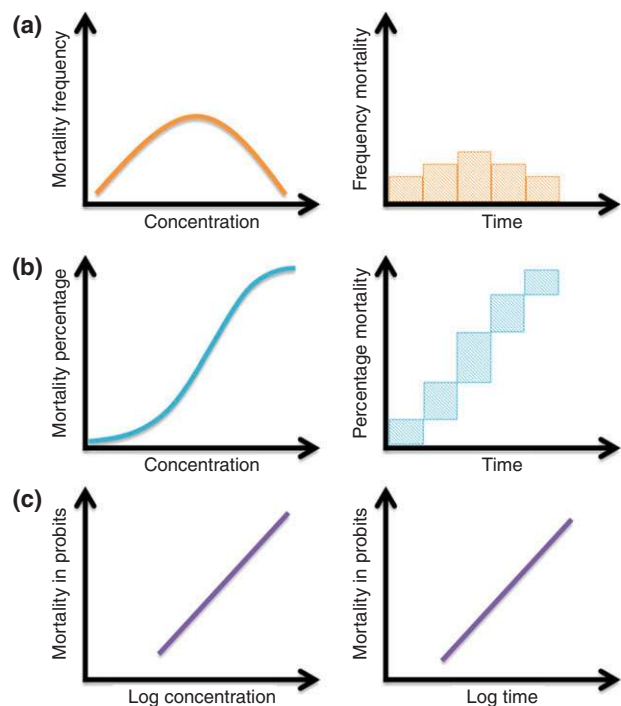

FIGURE 4 | Different types of relationships for dose–response models. Similar relationships can be generated for time–response models.

However, Haber's law does not hold in many situations,[63] and it does not take detoxification into consideration. The law assumes that any combination of concentration and time that has the same $C \times t$ product should produce the same level of

toxicity. However, in reality, this is not the case. Toxicity of some chemicals can be more dependent on concentration than time. Subsequently, Haber's law was generalized. Let C_0 denote a threshold concentration, and n and m are constants. Several well-known generalizations of Haber's law are shown below:

- Ostwald: $(C - C_0)^n \, t = K$ that emphasizes concentration[64];
- Druckery: $C \times t^n = K$ that emphasizes time[63]; and
- Miller *et al.*: $(C - C_0)^n \, t^m = K$ that emphasizes both concentration and time.[63]

One of the frequently measured responses is mortality (the number of deceased individuals). The Bliss method (or Probit model) (Figure 5)[61,63,65] transforms time–mortality and dose–mortality relationships into linear relationships. This transformation follows the next steps: (a) link mortality frequency (the number of deceased subjects) to dose or time; (b) convert frequency to percentages (percentage of deceased subjects); (c) transform percentages to probits (probability unites) and express dose or time in on logarithmic scale. Probits are inferred doses (or time) that correspond to a given mortality percentage. Bliss devised a special table called 'probits table' to calculate the probits as explained in ref. 65. This method takes into consideration the variation of an individual's susceptibility to toxic agents. For example, a certain dose (or time exposure) can cause the mortality of some individuals but not others.[61,65]

There are many inherited differences between dose–mortality and time–mortality models. Time–mortality curves are based on the same individuals whose susceptibility is measured at specific intervals. The percentage of mortality at a given interval cannot be less than that of the preceding interval, and the susceptibility of individuals in successive time intervals are correlated. However, dose–mortality curves are based on different individuals for each dose. Therefore, susceptibility of individuals at successive doses is unrelated especially if there are individuals who have a high toxicity resistance.[61]

The effectiveness of time–mortality curves depends on the 'whole' distribution of susceptibilities and their relationship to the response. Time–mortality curves that measure the response time can be incomplete for small doses due to individuals who have a high resistance and fail to show the measured response. Similarly, time–response curves that measure the recovery time may be incomplete for large doses if some individuals fail to recover. Bliss explained how to estimate the truncated distribution of time–mortality models.[61]

Miller et al.[63] proposed a three-dimensional model for concentration–time–response that can reliably interpolate within the scope of experimental data, and they provided an estimation of error when extrapolating outside the scope. Recently, Brown and Foureman[62] used a time–concentration–response model to generalize the concentration–response models using time as a parameter.

There are many advantages of time–response, dose–response, and dose–time–response models: ease of interpretation and implementation, consideration of dose and time of exposure, interpolation of effects between different doses of the same chemical within the range of experimental data[61,63] using dose–response models, and interpolation between different exposure times for the same toxicant and dose within the range of experimental data[61,63] using time–response models.

However, there are many limitations. The three models cannot extrapolate to other chemicals.[60] Additionally, time–response models cannot extrapolate to

FIGURE 5 | Bliss method. (a) Plot mortality frequency (the number of dead subjects) versus dose or time. (b) Convert frequency to percentages (percentage of deceased subjects). (c) Transform percentages to probits and transform dose or time to logarithms.

other doses of the same chemical.[61,63] Time–response models require that tested individuals have uniform susceptibility levels,[61] or these models may be unreliable if some individuals have an extremely low or high resistance. If time intervals are long, time–response models may overestimate or underestimate the response at a given moment.[61] The three models do not take into consideration target tissue, biological process, ADME, toxicokinetics, toxicodynamics, detoxification, damage or repair, or chemical properties.[60,63]

These time–response and dose–response models are complementary to one another and must be used together to achieve reliable conclusions. Several databases include dose–response data such as CEBS,[6,66] PubChem,[18,67] and ToxRefDB.[8,68] These models were used, for example, for modeling rectal cancer,[69] mutagenicity,[70] and developmental toxicity.[71]

Pharmacokinetic Models and Pharmacodynamic Models

Pharmacokinetic (PK) models relate chemical concentration in tissues to time, estimate the amount of chemicals in different parts of the body, and quantify ADME processes.[13,72] Toxicokinetic models are PK models used to relate chemical concentration in tissues to the time of toxic responses. PK models can be compartmental and non-compartmental.[60,72] A compartment is the whole or part of an organism in which the concentration is uniform.[73] Compartmental models consist of one or more compartments, and each compartment is usually represented by differential equations.[60,72,74]

One-compartment models represent the whole body as a single compartment, assume rapid equilibrium of chemical concentration within the body after administration, and do not consider the time to distribute the chemical. The concentration C at a given time t is computed by[72]

$$C(t) = C_0 \times e^{-kt}$$

where C_0 is the initial concentration and k is the elimination constant. The plotting log of concentration versus time results in a straight line of slope $(-k)$.[72] However, these models do not consider the distribution time of chemicals. Additionally, concentrations in some organs reach equilibrium faster than in others. Two-compartment models consist of two compartments: central (for rapidly-perfused tissues e.g., liver or kidney) and peripheral (for slowly

perfused tissues e.g., muscle or skin). Each compartment is represented by a differential equation similar to the one-compartment models. After solving the coupled equations, the concentration is the sum of two exponential terms of time (interpreted as distribution phase with initial concentration C_a and slope $-a$ and elimination phase with initial concentration C_b and slope $-b$). The concentration C based on this model is represented by[72]

$$C(t) = C_a \times e^{-at} + C_b \times e^{-bt}$$

These models, however, cannot extrapolate between species or provide a mechanistic insight.[72] On the other hand, physiologically based pharmacokinetic (PBPK) models include, in addition to concentration and time, physiological descriptors of tissues and ADME processes such as volumes, blood flows, chemical binding/partitioning, metabolisms, or excretions.[13,60] PBPK models represent each organ as a compartment, represented by a differential equation that includes PK parameters.[13,60,72,74] An organ can be split into several compartments if there is a high variability in organ tissue. Also, one compartment can represent several similar organs.[72] A general PBPK model to calculate plasma concentration (C_P) uses a feature vector of PK parameters (θ_{PK}), time (t), and dose (X)[74] as follows:

$$C_P = f(\theta_{PK}, X, t)$$

where f is a function that models the relationship. Because equation structure and the physiological parameters are tissue specific, PBPK models allow for interspecies extrapolation and provide a mechanistic basis of ADME.[11,60,72] PBPK models can convert administered doses to tissue dosimetry,[60] which is 'the amount of chemical that is distributed to a tissue or part of a tissue,'[13] and generate concentration versus time models.[11]

Pharmacodynamic (PD) models relate a biological response to the concentration of chemical in tissue.[72] Toxicodynamic models are PD models that relate toxicity to the concentration of the chemical. PD models that are based on anatomy, physiology, biochemistry, and biology are called physiologically based pharmacodynamic (PBPD) models.[75] Similar to dose–response models, PD models can be linear or nonlinear. Linear models should be used with caution because they do not consider the upper limit of responses and assume that responses always increase when concentrations increase.[72] Similar to PBPK models, PBPD can be described by differential equations. A general PBPD model calculates the response

(R) using a feature vector of PD parameters (θ_{PD}), plasma concentration (C_P), which is calculated using the PBPK model given above, or biophase concentration (C_e), and chemical-independent system parameters (Z)[74] can be represented as:

$$R = f(\theta_{PD}, C_P \, or \, C_e, Z)$$

where f is a function that models the relationship. PD models can be combined with PK models.[72] The resulting model is called biologically based dose–response models (BBDR) and can be used to relate doses with responses.[13,60,72,75] In addition to PK and PD parameters, BBDR may include biological parameters such as cell division rates, mortality rates, or production rates of hormones.[13,60,76] BBDR models are more powerful than dose–response models because the former consider time-dependent changes of concentration and can extrapolate at low doses and between species.[60,75]

There are many advantages for the models discussed in this section. Determining internal doses rather than administered doses and key metabolites allows for a more direct relationship with the response.[60] Additionally, using ADME, PK, and PD properties permits route-to-route and species-to-species (e.g., animal-to-human) extrapolations and *in vitro*-to-*in vivo* extrapolation.[60] BBDR is useful for extrapolating at low doses. Such low doses provide realistic estimates for human toxicity as human exposure to toxicants is at much lower doses than those tested on animals.[75]

However, there are a number of disadvantages. PK and PD parameters may be unavailable or inaccurate. In such cases, the parameters are estimated using *in vitro*-to-*in vivo* or species-to-species extrapolation.[72] Otherwise, QSAR modeling could be more appropriate because it depends only on molecular descriptors.[11,60] Additionally, if biological data is not available, empirical dose–response models are used instead of BBPD. Using BBDR for extrapolation between species assumes that the relationship between dose and response in animals is the same in humans.[76,77] The same problem applies when using animal studies to estimate PK or PD parameters for modeling toxicity in humans.[78] Although BBDR models have been proposed more than 20 years ago as a tool to minimize uncertainty for low-dose and interspecies extrapolation, it was recently shown that BBDR has not progressed to reach such expectations due to uncertainty in modeled parameters and data, limited applicability of BBDR models to a small group of chemicals, or inherited complexity of BBDR models or toxicity mechanisms as discussed in ref. 76. Expert knowledge is required for defining MoA, toxicity pathways and chemical interactions that cause the response.

Different types of PK and PD models are reviewed in[74] and summarized in Supplementary Table S1. An example of developing a PBPK model is available in ref. 79. Also, methods for estimating PK parameters are reviewed in ref. 80. Examples of PK and PD modeling tools are WinNonlin,[81] Kinetica,[82] and ADAPT 5.[83] For example, PBPK was used for route-to-route extrapolation,[84] toxicity and risk assessment,[85] and carcinogenicity assessment.[86]

Uncertainty Factor Models

Uncertainty factors (UFs) (also called assessment/extrapolation/risk factors) are used for assessing risk from chemical exposure or the recommended daily intake of chemicals.[87] A UF model is the simplest form of model for inter-species extrapolation (e.g., from animals to humans), intra-species extrapolation (e.g., from healthy people to special groups of the population such as elderly people, pregnant women, children, and fetuses), or exposure duration extrapolation (e.g., from short exposure to long exposure). It requires two main factors[88]: no observed adverse effect levels (NOAEL), which is the highest dose not exhibiting observable toxicity and a UF, which is a numerical value to account for variability in inter-species, intra-species, exposure duration, or exposed dose. Extrapolation is done by dividing NOAEL by UF.

However, there are two limitations for using NOAEL[88]: (1) the definition of NOAEL indicates the absence of the 'appreciable risk' of toxicity, but it does not indicate a zero-effect threshold; and (2) NOAEL values are not constants and can vary depending on experimental designs such as the number of tested animals, number of doses, and toxicity endpoints. It was shown that low statistical power (e.g., a small number of tested animals or a small number of tested doses) would result in higher NOAEL. However, it is possible to use a least observable adverse effect level (LOAEL, which is the least dose or concentration that causes the observed effect) or to use a benchmark dose level (BMDL, which is 'the lower statistical confidence limit of the dose resulting in a predetermined response') if NOAEL is not available.[87,88]

In addition to UFs, 'modifying factors' (MFs) are used to account for uncertainties in the data and the database. Additionally, 'safety factors' (SFs) are used for irreversible effects, such as teratogenicity and non-genotoxic carcinogenicity. Although,

existing UFs account for intra-species variability, the use of additional factors for child safety is recommended. The values of MFs and SFs cannot exceed 10.[87]

UFs are necessary to estimate reference dose (RfD) and reference concentration (RfC). RfD or RfC 'provide quantitative information for use in risk assessments for health effects known or assumed to be produced through a nonlinear (presumed threshold) mode of action.[60,] The reference values are calculated as

$$RfD \ or \ RfC = \frac{POD}{UF \times MF}$$

where *POD* is the point of departure (e.g., NOAEL, LOAEL, or BMDL).[60] A default UF of 100 was first proposed in 1954.[87] However, this default value does not account for the quality of the database, the nature of the effect, the duration of the exposure, route-to-route extrapolation, and consideration of special groups of the population. Therefore, several factors have been calculated by different agencies as explained in[87,89] and summarized in Supplementary Table S2.

There are several advantages of UF models. It is easy to implement and understand them. They provide adequate safety levels for a single chemical and mixtures of chemicals.[88] Additionally, they account for inter-species and inter-individual as well as PK and PD differences. However, there are some limitations of UF models. Default UFs or sub-factors are not conservative nor do they assume the worst-case scenario. Therefore, extrapolated safety levels of chemicals are not always below the realistic safety threshold for humans.[88] These models cannot be used to extrapolate toxicity levels of genotoxic carcinogens as these chemicals always cause toxicity effects that are proportional to the dose, even at small doses.[87]

Quantitative Structure–Activity Relationship

Quantitative structure–activity relationship (QSAR) is a family of models that uses molecular descriptors to predict chemicals' toxicity. It is assumed that chemicals that fit the same QSAR model may work through the same mechanism.[10] A general QSAR model to predict toxicity (*T*) using a feature vector of chemical properties (θ_P) and a function *f* that calculates *T* given θ_P is

$$T = f(\theta_P)$$

A local QSAR is generated from congeneric chemicals (i.e., similar chemicals); otherwise, the model is a global QSAR[6] if it was made from diverse chemicals. Local QSARs are more accurate as they are customized for specific chemicals. However, there is an overhead to develop a local QSAR for each type of chemical. Therefore, global QSARs are more practical but may be less accurate. Local QSARs can also provide insight on the MoA of specific chemicals, which global QSARs may overlook.

Quantitative Structure Toxicity/Property Relationship (QSTR/QSPR) models are QSAR models that predict toxicity and chemical properties, respectively.[10,19] Structure activity relationships (SAR) are used for categorical endpoints.[19] There are different types of models in the QSAR family as summarized in Supplementary Table S3.

There are two main steps to develop a QSAR model: generating molecular descriptors and then generating models to fit the data. Several types of molecular descriptors can be used to describe chemicals as summarized in Supplementary Table S3. Therefore, feature selection algorithms based on, for example, simulated annealing, genetic algorithm, or principal component analysis can be used.[5,19] If there are a small number of descriptors, using two-dimensional scatter plots of each descriptor versus the biological activity can help identify significant descriptors[19] (Figure 6).

There are several types of algorithms to generate QSAR models: linear models such as those based on linear regression analysis, multiple linear regression and partial least squares for continuous endpoints, and linear discriminant analysis for categorical endpoints[5,19]; nonlinear models such as artificial neural networks or support vector machines[5,19]; and data-driven models such as those based on decision trees, clustering, Naïve Bayes, and K-nearest neighbor.[7]

A comparison of different machine learning and regression models is provided in ref. 7. Linear models are simpler and, in general, require tuning fewer parameters than nonlinear models. However, many relationships between chemicals and toxicity are nonlinear. Therefore, nonlinear models are commonly used for developing QSARs. The two-dimensional scatter plots can help identify the type of regression models[19] as illustrated in Figure 6.

Additionally, SAR landscapes are three-dimensional plots through which one can visualize structure–activity relationships. The X–Y plane represents the molecular descriptors, and the Z-axis represents response. Figure 7 shows a hypothetical example of a SAR landscape. The smooth region
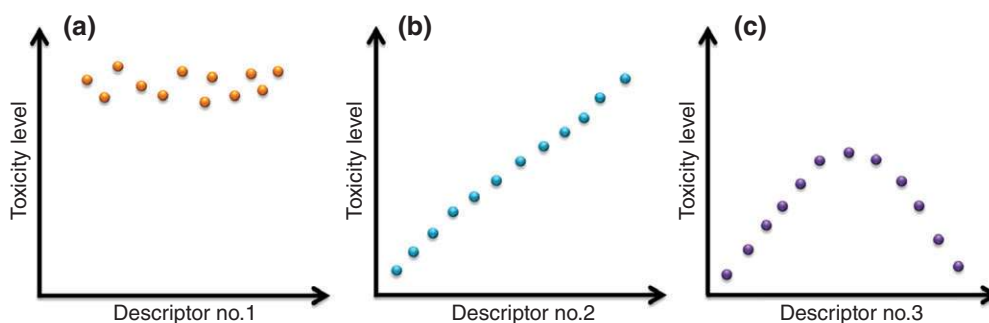
FIGURE 6 | 2D scatter plots of molecular descriptors and toxicity levels. (a) no correlation between molecular descriptor 1 and the toxicity endpoint. (b) and (c) linear and nonlinear relationships between the molecular descriptors 2 and 3, respectively, with the toxicity endpoint. (b) and (c) can be modeled with linear and nonlinear algorithms, respectively.
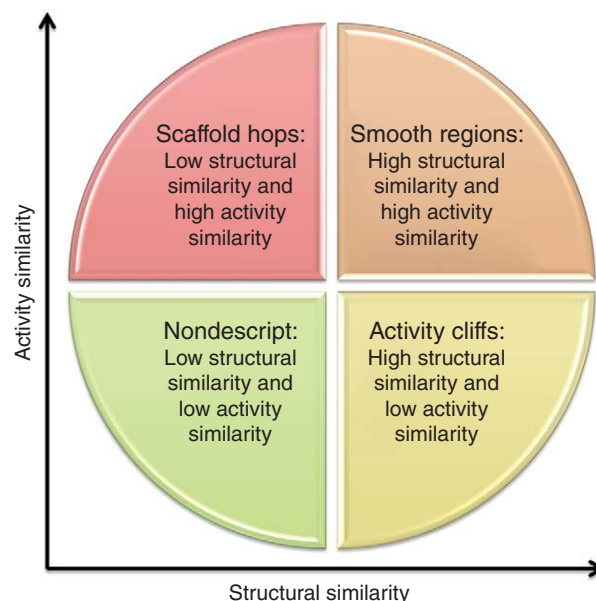

FIGURE 7 | SAR landscapes.
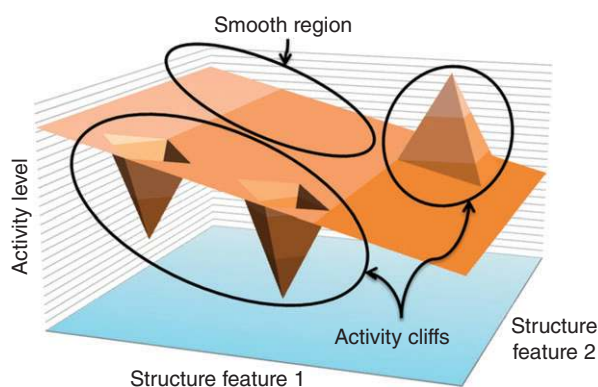
FIGURE 8 | Four regions in SAR maps are scaffold hops, smooth region, nondescript, and activity cliffs.

corresponds to chemicals that have a similar structure and similar activity. However, the ragged region corresponds to chemicals that have a similar structure but different activity levels (also called activity cliffs). The activity cliffs are the most interesting part of the SAR landscape.[18] They show that small structural changes correspond to huge changes in activity. Additionally, they affect the performance of machine learning models, either because these regions are discarded as outliers, cause over-fitting, complicate the prediction models, or increase the prediction error while generating the model.

SAR landscapes can be visualized using SAR maps. SAR maps are two-dimensional plots of activity similarity versus structure similarity that characterize SAR landscapes through four regions[18] as shown in Figure 8. Moreover, a structure activity landscape index (SALI) and a structure activity index (SARI) can be used to analyze SAR landscapes as explained in ref. 90.

Historically, one of the early QSAR models was developed in 1962 by Hansch et al.[19] in which the log of chemical concentration (C) was estimated using the octanol/water partition coefficient (π) and the Hammett constant (σ)[91]:

$$\mathrm{Log}\left(\frac{1}{C}\right) = 4.08\,\pi - 2.14\,\pi^2 + 2.78\,\sigma + 3.36$$

If the coefficient of a descriptor is positive, there is a positive relationship between the toxicity endpoint and the descriptor; otherwise, there is a negative relationship.[92]

Examples of QSARs for predicting toxicity of aromatic nitro compounds, nitrobenzene compounds, cytotoxicity of TIBO derivatives, and carcinogenicity of sulfa drugs are explained in ref. 5. A discussion of the performance of QSAR models to predict

carcinogenicity, mutagenicity, reproductive, and developmental toxicity endpoints are available in ref. 6. Case studies on applications of QSAR to skin sensitization and developmental toxicity are available in refs. 93 and 12, respectively.

There are many tools that provide pre-built QSAR models such as OECD QSAR Toolbox,[25] TopKat,[94] Derek Nexus,[34] HazardExpert,[35] VEGA,[95] and METEOR.[36] Their characteristics are summarized together with those of other QSAR-based tools in.[10,12,17,96] Case studies on combining the results of different prediction tools are available in refs. 15,17. However, specialized software tools for generating QSAR models such as ADAPT and TOPKAT include databases for toxicity data and can calculate molecular descriptors.[12] Additionally, several stand-alone databases have been compiled to provide toxicity data and/or molecular descriptors as summarized in refs. 6,7,10.

There are several advantages of QSAR models. They are easy to interpret if the descriptors are meaningful. They can model categorical and continuous toxicity endpoints and molecular descriptors and toxic and non-toxic chemicals. Using different types of descriptors allows for modeling complex endpoints.[6,11]

However, QSARs may not be always applicable. QSARs require a large number of chemicals in model development to achieve statistical significance.[5] Additionally, QSARs require using feature selection to identify the most significant and independent molecular descriptors, and a large number of descriptors makes the multidimensional space complex and fragmented.[97] QSARs cannot be used for extrapolation between species, routes of exposure, or doses unless biological data is used. Moreover, QSARs may not be biologically interpretable, and QSARs do not take dose, duration, or metabolites into consideration.

A brief description of all the tools mentioned in the *IN SILICO* MODELING METHODS section is available in Table 1.

SPECIAL CASES

This section focuses on special types of chemicals or toxicity endpoints that require new prediction or analysis methods.

Nanotoxicity

Nanotoxicity is the study of adverse effects caused by nanomaterials. Nanomaterials are small particles on the nanoscale (10^{-9} m) size range. When a particle

size is decreased within the nanoscale size range, its physical and chemical properties are changed, affecting its toxicity. It was found that nanoparticles cause different or worse toxicity effects than the larger particles of the same substance.[98,99] A nanoparticle can be toxic even if the particle is not toxic at a larger size. The small size of nanomaterials facilitates cell membrane penetration and biodistribution.[98,99] Other properties that affect toxicity of nanomaterials are as follows[100]:

- The shape of nanomaterials affects toxicological responses. For example, isolated long fiber carbon nanotubes are more inflammogenic in the outer regions of the lung than non-fibrous nanotubes.

- Large surface areas of nanomaterials increase the contact area with the biological environment and their chemical reactivity.

- Surface coating material of the nanomaterials can affect biological functions. It was found that toxicity of nanomaterials that have the same metallic core could be predicted by using the properties of the coating material.

- Other physicochemical properties such as electrostatic interactions between nanomaterials and biological targets can influence toxicity.

There are several mechanisms by which nanoparticles induce toxicity:

- interaction and binding of the nanoparticle's surface with a biological environment (e.g., protein or cells),[100]

- cellular entry: nanoparticles potential to enter cells,[101]

- release of ions from the surface: ionic forms of metals can be more active,[102,103] and

- generation of reactive oxygen species (ROS): overproduction of ROS can cause oxidative stress and inflammation, which disrupt normal biological functions and damage DNA and proteins.[104]

Quantitative structure nanotoxicity relationships (QSNR)[100] (also called nano-QSAR[105]) are QSAR models that use nanomaterial-specific descriptors such as size, shape, surface area, relaxivities, solubility, zeta potential, corona composition, biodistribution, bioavailability, and surface charge in addition to structural and physicochemical properties.[100,106] However, toxicity of nanomaterials is affected by

TABLE 1 | Summary of *In Silico* Modeling Methods

Method	Definition	Approaches	Advantages	Limitations	Existing Software or Databases
Structural alerts (SAs) and rule-based models	SAs are chemical structures that indicate or associate to toxicity.	• Human-based rules • Induction-based rules • Apriori (based on breadth-first search) • Pattern growth (based on depth-first search) such as mofa, gSpan, FFSM, and gaston.	1) It is easy to interpret and implement SAs. 2) SAs allow determining how chemicals should be altered to reduce their toxicity. 3) SAs allow identifying the structure of potential metabolites.	1) This method can indicate only the presence or absence of SAs. 2) SAs do not provide insight into the biological pathways of toxicity. 3) The list of SAs may be incomplete, which may increase false negatives.	• OECD QSAR • Toxtree • OCES • Derek Nexus • HazardExpert • Meteor • CASE • PASS • cat-SAR
Read-across (RA)	A method of predicting unknown toxicity of a chemical using similar chemicals with known toxicity from the same chemical category	• Analog approach (one-to-one) • Category approach (many-to-one) • Qualitative and quantitative RA • Interpolation and extrapolation RA	1) RA is transparent, easy to interpret, and implement. 2) RA can model quantitative and qualitative toxicity endpoints, and uses many types of descriptors and similarity measures.	1) RA uses small datasets. 2) Accuracy depends on the number and choice of analogs, similarity metrics, strength in chemicals' similarity, chemical properties, and category boundary. 3) RA may be inaccurate if analogs have conflicting toxicity profiles.	• OECD QSAR • Toxmatch • ToxTree • AMBIT • AmbitDiscovery • AIM • DSSTox • ChemIDplus
Dose–response (DR) and time–response (TR) models	Dose–response (or time–response) models are relationships between doses (or time) and the incidence of a defined biological effect.	• Haber's law and its generalizations • Bliss method (Probit model) • 3D time–dose–response models	1) Ease of interpretation and implementation 2) Consideration of dose and time of exposure 3) Interpolation of effects between different doses and exposure times	1) DR and TR models cannot extrapolate to other chemicals. 2) TR models require tested individuals to have uniform susceptibility levels. 3) DR and TR models do not consider target tissue, or chemical properties.	• CEBS • PubChem • ToxRefDB
Pharmacokinetic (PK) and pharmacodynamic (PD) models	PK models calculate concentration at a given time. PD models calculate internal doses rather than administered doses.	• One-compartment models • Two-compartment models	1) PK models determine internal doses rather than administered doses.	1) PK and PD parameters may be unavailable or inaccurate.	• WinNonlin • Kinetica • ADAPT

TABLE 1 | Continued

Method	Definition	Approaches	Advantages	Limitations	Existing Software or Databases
	effect at a given concentration	• PBPK, PBPD and BBDR models	2) PK and PD models permit route-to-route, species-to-species, and *in vitro*-to-*in vivo* extrapolation. 3) BBDR is useful for extrapolating at low doses.	2) Extrapolation between species assumes that the relationship between dose and response in certain species is the same as in the other. 2) Expert knowledge is required for defining MoA, toxicity pathway and chemical interactions.	
Uncertainty factors (UFs) models	UF is a numerical value to account for variability in inter-species, intra-species, exposure duration, or exposed dose	• Extrapolation using NOAEL, LOAEL, or BMDL. • RfD and RfC models • Modifying factors and safety factors	1) It easy to implement and understand UF models. 2) They provide adequate safety levels for single chemical and mixtures of chemicals. 3) They account for inter–species and inter–individual, PK and PD differences.	Default UFs or sub-factors are not conservative nor do they assume the worst-case scenario.	
Quantitative structure-activity relationship (QSAR) models	QSAR is a family of models that use molecular descriptors to predict chemicals' toxicity.	• Local and global QSAR • SAR, QSTR, and QSPR • SAR landscapes and maps	1) QSARA models are easy to interpret if the descriptors are meaningful. 2) They can model categorical and continuous toxicity endpoints, and toxic and non-toxic chemicals. 3) Using different types of descriptors allows for modeling complex endpoints.	1) QSARs require large datasets. 2) QSARs may require using features selection. 3) QSARs cannot be used for extrapolation between species, routes of exposure or doses, unless biological data are used. 4) QSARs do not take dose, or duration into consideration.	• OECD QSAR • TopKat • Derek Nexus • HazardExpert • VEGA • METEOR

properties of their core and surface coating material, and different combinations of core and surface compositions may cause different effects. Moreover, the coating material can be modified by the biological environment.[100]

An example of a linear QSNR model for predicting EC_{50} (effective concentration for 50% enzyme activity inhibition) for organo-coated silver nanoparticles using size and surface charge is available in ref. 107. Additionally, QSNR models that use quantum-chemical descriptors have been developed to predict binding affinity for a set of fullerene-C_{60} derivatives[108] and predict cytotoxicity of metal oxide nanoparticles.[109] Another model to predict metal oxide nanoparticles based on random forests classification is explained in ref. 103. Case studies on developing QSNR models are available in ref. 110.

There are some challenges in modeling nanomaterials. Nanomaterials can be composed of organic, metal, metal oxides, silica or carbon-based nanoparticles. Therefore, it is difficult to compile datasets of congeneric nanomaterials.[105] Additionally, nanomaterials are structurally diverse and act upon different MoA. However, QSNRs (similarly to QSARs) assume a common MoA for the modeled chemicals.[105] The scarcity of nanomaterial experimental data and descriptors[100] affects progress in this field.

However, a nano-read-across model (read-across for nanomaterials) can be used because read-across methods do not require large datasets to generate groups of sufficiently similar nanomaterials.[105] An application of a nano-read-across model to predict cytotoxicity of metal oxide nanoparticles is provided in ref. 105.

Additionally, PBPK models may predict biodistribution, which is essential for assessing toxicity of nanomaterials.[106] Customized PBPK models should be developed for nanomaterials because ADME properties, some physiological processes of nanomaterials, and transportation mechanisms are different from those of small molecules. Moreover, there are some processes that are not involved with small molecules such physical property changes, enzymatic degradation, cellular recognition, and internalization and opsonization in the blood.[106] PBPK may include nanomaterial-related descriptors such as traffic within tissues and cells, interaction with blood and tissue cells, tissue/blood partition coefficients, tissue concentration, and permeability through membranes.[106] An example of a kinetic model to describe mechanisms of releasing silver ions of citrate-coated silver nanoparticles in aqueous environments is available in ref. 102.

Mixtures

Toxicity of chemicals is affected by interactions with other chemicals. For example, mixtures may exhibit adverse effects at NOAEL doses of each chemical separately.[111] Assessing toxicity of chemicals separately may underestimate or overlook the adverse effects of mixtures.[111] For instance, it was found that toxicity of lead metal increases with the co-administration of higher levels of other metals. Therefore, 'cumulative risk assessment' was developed to study toxicity of mixtures.[112]

However, there is lack of experimental datasets for toxicity of mixtures due to a large number of different combinations of chemicals,[111] exposure patterns, and complex interactions. It is impossible to test all combinations of these factors. Furthermore, predictive models must address concurrent and sequential exposure to mixtures. However, a recently developed database by NoMiracle (Novel Methods for Integrated Risk Assessment of Cumulative Stressors in Europe) contains mixtures' toxicity datasets for eco-toxicological test species and human cell lines.[112]

Methods for single chemicals may not be applicable for mixtures due to difficulty in determining the combined effect.[113] For example, dose–response models for mixtures vary depending on the dose ratios of chemicals in the mixture.[112] Additionally, co-administration of chemicals may alter their ADME properties,[111] which should be taken into consideration when developing PBPK models for mixtures.[113]

Bliss generated dose–mortality curves for mixtures by changing the doses of mixtures while preserving the constituents' ratios. He identified three categories of the joint action of mixtures.[114] The first category is *independent joint action*. Chemicals act independently and have different modes of action. The combined effect is calculated using the effects of constituents and their interactions.[114]

The second category is *similar joint action*. Chemicals act independently and have similar MoA. The combined effect is calculated using the dose–mortality curves of constituents. This category assumes that an ingredient in the mixture can be substituted for any proportion of another ingredient without changing the combined effect.[114]

The third category includes *synergistic* and *antagonistic actions*. Toxicity of synergistic action is greater than that of constituents, while antagonistic action has lower toxicity than that of constituents. Synergistic effects depend upon the proportion of constituents in mixtures unlike the first two categories in which chemicals act independently, and

therefore, their proportions do not alter their combined effect. Bliss developed two models (also called independent action (IA)[115]) to analyze synergism as explained in ref. 114.

Another additive model is the concentration addition (CA) (also called dose addition) that was developed for chemicals with a similar MoA.[115] The concentration of mixtures that produces a certain effect is calculated using the proportion of a constituent and its concentration that produces the same effect.[115]

IA and CA models have been criticized for being ineffective for chemicals that have high potency (dose to produce a given effect) but low efficacy (maximum effect).[115] Therefore, a generalized concentration addition (GCA) model was developed to address these shortcomings.[116] GCA calculates the combined effect of a mixture using the potency and efficacy of the mixture's constituents.[115] A comparison of CA, IA, and GCA models is available in ref. 115.

Process-based models, however, are mechanistic models that usually use dynamic energy budgets theory such that the combined effect is calculated using the effects of constituents in addition to exposure time, toxicokinetic, and biological parameters, which allow for extrapolation between different species, chemicals, or exposure duration.[117] A discussion and an example of process-based models are available in[117] and,[118] respectively.

Another mechanistic model is the receptor-oriented model, which is based on the premise that the toxicity of mixtures is caused by many chains of reactions that converge at the exposed receptor (i.e., an individual or population).[112] A discussion of receptor-oriented models is available in ref. 112. Other methods to assess toxicity of mixtures include numerical additive models such as hazard index, point of departure index, margin of exposure and cumulative risk index[113]; chemical interaction models such as the interaction-based hazard index and isobole method[113]; and statistical models such as tree-based clustering and weighted quartile score regression.[111]

RECOMMENDATIONS

The previous sections discuss advantages and limitations of different models. The flow chart in Figure 9 provides a practical guideline for choosing a method for certain types of features and toxicity endpoints, and a summary of the methods is depicted in Figure 10. In this section, we provide some

recommendations for developing *in silico* toxicity prediction models.

Datasets Quality

Large datasets from one resource must be used to achieve statistical significance and structural diversity.[92,119] While large datasets may not be readily available, it is possible to gather data from various public resources[120]: prepared datasets suitable for QSAR modeling such as DSSTox (www.epa.gov/comptox/dsstox/), databases that contain toxicity data such as ECOTOX (http://cfpub.epa.gov/ecotox/), global search engines that retrieve data from several databases such as ToxNet (http://toxnet.nlm.nih.gov/), and scientific literature. A discussion of advantages and limitations of these resources in addition to case studies on obtaining toxicity data are available in Ref 120. Examples and descriptions of many toxicity databases are provided in Refs 6,8–11.

However, several factors must be considered when using data from multiple resources. The experimental design must be identical or compatible.[19] Additionally, the datasets should preferably be balanced (equal number of active and inactive chemicals) as some statistical methods are sensitive to unbalanced datasets.[19] This limitation, however, is not absolute as there are methods that can handle imbalanced data. Moreover, experimental and computed values of the same molecular descriptor should not be mixed, and calculated molecular descriptors must be generated using the same tools.[19] In addition, redundant records and outliers should be removed if necessary.[119]

The readers are referred to a recent review[121] that discusses formal scoring methods to assess quality of datasets; chemical, biological, and endpoint-specific factors in data variability; a practical checklist to guide the process of data assessment; data integration; and many other considerations of dataset quality.

Data Transformation

In some cases, biological data must be transformed. Continuous toxicity endpoints can be expressed in logarithmic scales to avoid statistical problems when using models such as regression.[19] Additionally, continuous endpoints, which come from different resources or are generated using different experimental procedures, can be transformed to categorical values (the number of categories are problem dependent).[19] Features must be normalized when using statistical similarity measures because these measures are

biased toward features that have large values.[47] Moreover, frequency descriptors (e.g., frequency of atoms) must be transformed to binary descriptors if the frequencies are not high enough to achieve reliable coefficients when deriving regression models.[19] In addition, values must be expressed in the same units.[119]

Feature Selection

The number of descriptors should be as small as possible.[92] Feature selection is necessary to ensure that descriptors are independent, orthogonal, uncorrelated, and non-redundant. This will increase chances that the models generated are closer to the optimal as this should remove correlation between descriptors. The descriptors should be meaningful to simplify model interpretation. Some methods for feature selection were previously explained in the QSAR sub-section. Additionally, it is preferable to use features that are suitable for toxicity endpoints (e.g., using environmental-related descriptors such as acidity of the water for ecotoxicity), chemicals type (e.g., using nanomaterials-specific features for nanomaterials), and model types (e.g., using PK descriptors for PBPK models).

Statistical Models

Simple accurate models are preferred to the complex ones. However, more sophisticated algorithms can be used to achieve better performance as long as interpretability of the model is preserved and overfitting is avoided. Interpretability of the models depends on
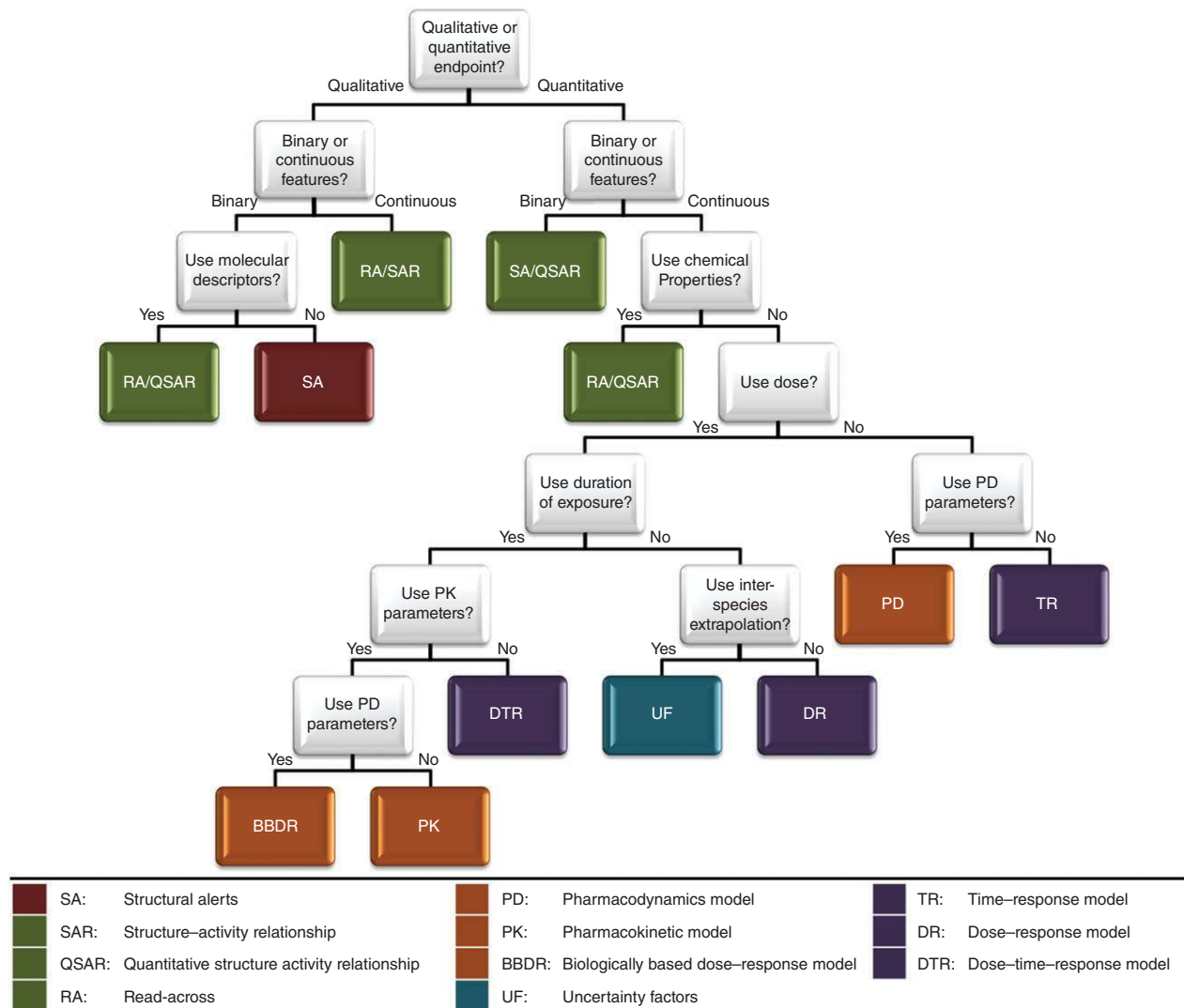
FIGURE 9 | Flowchart of *in silico* prediction models.
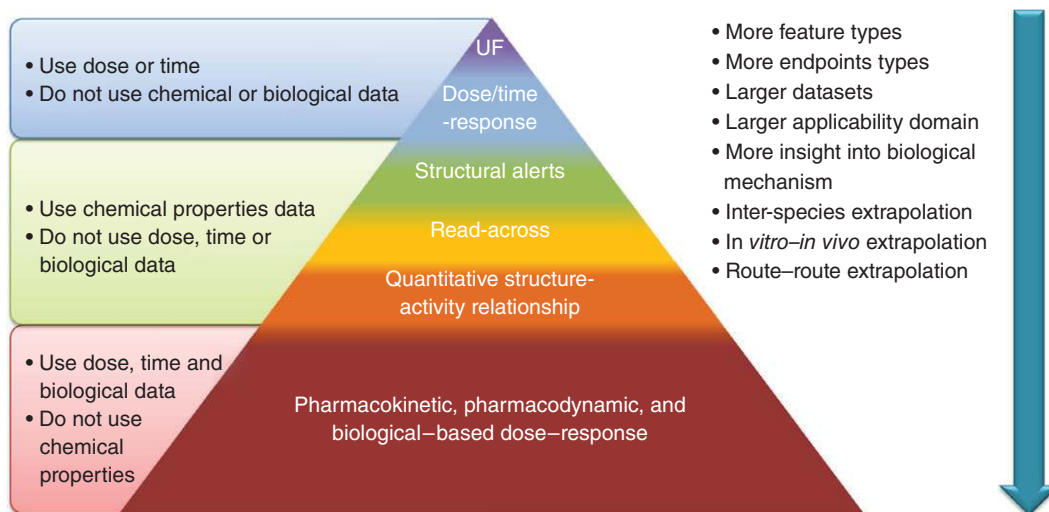
FIGURE 10 | Summary of methods to predict toxicity of chemicals.

the meaning of the descriptors and the relationship between them.[90]

There are two types of models[12]: correlative models, which use statistical algorithms that relate molecular descriptors to toxicity endpoints without providing a mechanistic insight, and mechanistic models, which explain the underlying mechanism of toxicity. Mechanistic models can be generated by statistical algorithms or designed by human experts.

However, models that are not mechanistically interpretable are still useful if they produce meaningful predictions, which is usually the case if they were correctly developed from large datasets.[19] Generally, models must meet users' expectations. If users expect models that provide mechanistic insight of toxicity, developers must try to meet these expectations.

Overfitting

Overfitting occurs when a model fits the training data extremely but cannot describe the external data well. It can be detected when models achieve high performance on training data but low performance on new data. Overfitting can be caused when modelers develop models that cover all cases in the training set or use complex models or the model includes more descriptors than necessary or wrong descriptors.[12] It is customary to have at least five chemicals per descriptor. Otherwise, an overfit model may become meaningless, be overly complicated, find random correlations, or have a limited applicability domain ref. 97.

There are two main statistical parameters to assess goodness of fit of models (e.g., QSARs generated using regression): coefficient of multiple determination (R^2) and standard error of estimate (s).[92,122] R^2 estimates how successful the regression line is in explaining variation in the biological data (e.g., measured response), while s estimates deviation of the measured response from the regression line.[92] A well-fit model should have large values of R^2 (close to 1) but low values of s (close to 0).[92,122] However, an overfitted model can be generated if its statistical fit is greater than the experimental variation (caused by errors in measured response or calculated descriptors).[122] It is possible to increase R^2 by adding more (possibly irrelevant) molecular descriptors. This can be avoided by using adjusted R^2 ($R^2_{(adj)}$). Unlike R,[2] $R^2_{(adj)}$ does not increase when adding irrelevant descriptors. Mathematical definitions of these and other statistical parameters for fitting are explained in,[92] and a discussion of sources of experimental variability is available in ref. 122.

Applicability Domain

The applicability domain (AD) is 'a theoretical region in physicochemical space (the response and chemical structure space) for which a QSAR model should make predictions with a given reliability.'[92] The AD determines types of molecules and toxicity endpoints to which the model can be applied. For example, global and local QSARs have large and small ADs, respectively.[6] Interpolation within the domain is more reliable than extrapolation outside the domain.[19] AD determines uncertainty in predicted toxicity by measuring the difference between the new molecules and the training set.[97] The model may not be reliable or perform well if the new molecule is sufficiently different than the training set.[6,90]

A stepwise approach to defining ADs was developed by Dimitrov et al.[123] and further expanded by Hewitt and Ellison.[124] Five categories of AD definition approaches were explained and compared in[124] descriptor, structural fragments, structural similarity, mechanistic, and metabolism-based approaches.

The descriptor-based category includes four types of methods: range, geometric, distance, and probability density-based methods, which were compared in ref. 124. Range-based methods require determining the range (i.e., minimum and maximum) of each descriptor. However, geometric methods such as convex hull define the smallest convex area containing the whole training set. Probability density distribution-based methods identify probability density of the dataset and then determine the highest density region. Distance-based methods use distance measures (e.g., Euclidean, Hamming, or Tanimoto). Then, the average distance to k-nearest neighbors or the distance to the centroid is taken to determine how similar/dissimilar the new chemical is from the training set. A threshold can be used to determine if the new chemical is within the AD. A comparison of different distance measures to define AD is available in ref. 97. Another application using cluster analysis is illustrated in ref. 125.

Structural fragment-based ADs require that all structural fragments in the new chemical be present in the training set.[124] However, the structural similarity-based ADs determine structural similarity of the whole chemical rather than fragments.[124] Mechanistic ADs have been discussed in refs. 123–126. Another AD method using simplified molecular-input line entry system (SMILES) attributes is illustrated in ref. 127. Moreover, examples of tools for defining ADs (e.g., The OECD QSAR Toolbox and ToxTree, etc.) are explained in ref [124].

Model Evaluation

Cross-validation on training sets is insufficient for a realistic evaluation because models are usually applied to new chemicals. Models must be evaluated using testing sets that are large enough to achieve statistical significance, diverse, within the AD, and different than the training sets. Evaluating models' accuracy is discussed in ref. 92.

Model Application

Model development processes and AD must be transparent to users to ensure appropriate application of the models. For example, molecular descriptors of new chemicals must be calculated and expressed in the same units as in the training set, and the new data must be transformed as the training set. Additionally, the output of the models must be clearly explained. Moreover, models preferably should provide a rationale (e.g., a list of the rules that were used to form the prediction). Also, models should provide a confidence score of the prediction. This will help users decide whether to accept the predictions of the models. Consensus can be used to determine reliability of the prediction as explained in ref. 15. For example, if a chemical was predicted to be carcinogenic by several tools, there might be higher confidence in the prediction than if it was predicted by a single tool. However, if the tools provide conflicting predictions, then it is necessary to resolve such conflicting situations, which is usually done by an additional machine learning model.

TWENTY-FIRST CENTURY TOXICOLOGY

The phrase '21st century toxicology' (Tox-21c[128]) refers to 'the transformation underway in the tools and approaches used to evaluate chemical substances for possible effects on human health.[129,] Tox-21c focuses on toxicity pathways,[130] mechanisms/modes of action, and adverse outcome pathways (AOP)[131] in humans. Another related concept is the 3Rs (replace, reduce, and refine), which was proposed in 1958.[128] Tox-21c and 3Rs overlap, but a discussion of their differences is presented in ref. 128.

Several strategies have been proposed to implement Tox-21c. In 2004, the National Toxicology Program (NTP) published its report *A National Toxicology Program for the 21st century*, which aims 'to support the evolution of toxicology from a predominantly observational science at the level of disease-specific models to a predominantly predictive science focused upon a broad inclusion of target-specific, mechanism-based, biological observations.[132,] In 2007, the National Research Council (NRC) published another report *Toxicity Testing in the 21st Century: a Vision and a Strategy*, which proposed using computational methods to decrease the number of tested animals, make toxicity testing more relevant to humans by using human cells, and make toxicity testing cheaper and faster.[133]

Several approaches have been developed to implement these visions. Integrated Testing Strategy (ITS) aims to combine information from testing and non-testing methods.[134] Another approach is Pathways of Toxicity (PoT), which aims to produce a

comprehensive list of all human PoT.[135] However, Integrated Approaches to Testing and Assessment (IATA) focuses on using hypotheses to prioritize chemicals for testing.[131]

Several projects implement the above approaches. For example, ReProTect is an integrated testing project specializing in reproductive toxicity.[136] The ToxCast (Toxicity Forecaster) project uses quantitative high throughput screening to test chemical toxicity on many biological pathways,[137] and it is part of the Tox21 project that follows NTP and NRC guidelines.[138] Additionally, the Human Toxicology Project Consortium implements the NRC guidelines and focuses on implementing AOP .[139]

Conclusion

The field of *in silico* toxicology has been in a continuous development through the introduction of new methods, improvement of the existing ones, or discarding of some of them. Unfortunately, a method that is suited for certain types of toxicity endpoints or chemicals may not work properly (or not work at all) for others. If used correctly, *in silico* tools can be very effective in assessing chemicals' toxicity. Therefore, to ensure accurate and effective application of *in silico* models, it is necessary to (1) understand the methods' strengths, limitations, scope of application, and interpretation; (2) choose the most effective method for the problem at hand; and (3) customize these methods for each problem if necessary. Users of toxicity prediction models can follow these three steps only if the data and processes to develop the model are transparent, applicability domains are well defined, the outputs of the models are clearly explained, and models are simplified.

The Tox-21c stresses replacing animal testing with human-relevant testing methods, either *in vitro* or *in silico*. With the increasing number and variety of alternative testing methods, it is necessary to apply strategies (e.g., ITS) to intelligently combine and use this information for toxicity assessment and decision making. Clearly, *in silico* toxicology is a useful component of the toxicity assessment process. Looking to the future, computational methods are likely to expand to include models for special and new types of toxicity endpoints and chemicals, provide insight into toxicological pathways, combine and compare results from different models, customize models to meet users' expectations, and refine models as new data become available.

FURTHER READING

Balls M, Combes RD, Bhogal N, eds. *New Technologies for Toxicity Testing*. New York: Springer-Verlag; 2012.

Davis A, Ward SE, eds. *The Handbook of Medicinal Chemistry: Principles and Practice*. Cambridge, UK: Royal Society of Chemistry; 2015.

Dearden JC. In silico prediction of drug toxicity. *J Comput Aided Mol Des* 2003, 17:119–127.

Derelanko MJ, Auletta CS, eds. *Handbook of Toxicology*. Boca Raton, FL: CRC Press; 2014.

Kapetanovic IM. Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chem Biol Interact* 2008, 171:165–176.

Marchant CA, Briggs KA, Long A. In silico tools for sharing data and knowledge on toxicity and metabolism: Derek Nexus, Meteor, and Vitic. *Toxicol Mech Methods* 2008, 18:177–187.

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF. An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. *Regul Toxicol Pharmacol* 2006, 44:97–110.

Merlot C. In silico methods for early toxicity assessment. *Curr Opin Drug Discov Dev* 2008, 11:80–85.

Pfannkuch F, Suter-Dick L, eds. *Predictive Toxicology: From Vision to Reality*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2015.

Salem H, Katz SA, eds. *Alternative Toxicological Methods*. Boca Raton, FL: CRC Press; 2003.

Terstappen GC, Reggiani A. In silico research in drug discovery. *Trends Pharmacol Sci* 2001, 22:23–26.

Vedani A, Dobler M, Lill MA. The challenge of predicting drug toxicity in silico. *Basic Clin Pharmacol Toxicol* 2006, 99:195–208.

Hvd W, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* 2003, 2:192–204.

Wilson AG, White AC, Mueller RA. Role of predictive metabolism and toxicity modeling in drug discovery – a summary of some recent advancements. *Curr Opin Drug Discov Dev* 2003, 6:123–128.

Xu JJ, Urban L, eds. *Predictive Toxicology in Drug Safety*. New York: Cambridge University Press; 2011.

REFERENCES

1. Rowe PH. Statistical methods for continuous measured endpoints in in silico toxicology. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 228–251.

2. Rowe PH. Statistical methods for categorised endpoints in in silico toxicology. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 252–274.

3. AltTox. Toxicity testing overview. Available at: http://alttox.org/mapp/toxicity-testing-overview/. (Accessed May 5, 2015).

4. Canadian Centre for Occupational Health and Safety. What makes chemicals poisonous. Available at: http://www.ccohs.ca/oshanswers/chemicals/poisonou.html. (Accessed May 5, 2015).

5. Deeb O, Goodarzi M. In silico quantitative structure toxicity relationship of chemical compounds: some case studies. *Curr Drug Saf* 2012, 7:289–297.

6. Valerio LG Jr. In silico toxicology for the pharmaceutical sciences. *Toxicol Appl Pharmacol* 2009, 241:356–370.

7. Madan AK, Bajaj S, Dureja H. Classification models for safe drug molecules. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 99–124.

8. Zhu H. From QSAR to QSIIR: searching for enhanced computational toxicology models. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 53–65.

9. Benigni R, Battistelli CL, Bossa C, Colafranceschi M, Tcheremenskaia O. Mutagenicity, carcinogenicity, and other end points. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 67–98.

10. Toropov AA, Toropova AP, Raska I Jr, Leszczynska D, Leszczynski J. Comprehension of drug toxicity: software and databases. *Comput Biol Med* 2014, 45:20–25.

11. Modi S, Hughes M, Garrow A, White A. The value of in silico chemistry in the safety assessment of chemicals in the consumer goods and pharmaceutical industries. *Drug Discov Today* 2012, 17:134–142.

12. Venkatapathy R, Wang NCY. Developmental toxicity prediction. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 305–340.

13. Jack J, Wambaugh J, Shah I. Systems toxicology from genes to organs. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 375–397.

14. Cannata N, Corradini F, Merelli E, Tesei L. Agent-based models of cellular systems. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 399–426.

15. Milan C, Schifanella O, Roncaglioni A, Benfenati E. Comparison and possible use of in silico tools for carcinogenicity within REACH legislation. *J Environ Sci Health C Carcinog Ecotoxicol Rev* 2011, 29:300–323.

16. Cronin MTD. In silico tools for toxicity prediction. In: Wilson AGE, ed. *New Horizons in Predictive Toxicology: Current Status and Application*. Cambridge, UK: Royal Society of Chemistry; 2011.

17. Worth AP, Lapenna S, Serafimova R. QSAR and metabolic assessment tools in the assessment of genotoxicity. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 125–162.

18. Guha R. On exploring structure–activity relationships. In: Kortagere S, ed. *In Silico Models for Drug Discovery*, vol. 993. New York: Humana Press; 2013, 81–94.

19. Devillers J. Methods for building QSARs. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 3–27.

20. Roncaglioni A, Toropov AA, Toropova AP, Benfenati E. In silico methods to predict drug toxicity. *Curr Opin Pharmacol* 2013, 13:802–806.

21. Lepailleur A, Poezevara G, Bureau R. Automated detection of structural alerts (chemical fragments) in

(eco)toxicology. *Comput Struct Biotechnol J* 2013, 5:1–8.

22. Dupuis G, Benezra C. *Allergic Contact Dermatitis to Simple Chemicals: A Molecular Approach*. Marcel Dekker Inc: New York; 1982.

23. Ashby J, Tennant RW. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. *Mutat Res* 1988, 204:17–115.

24. Benigni R, Bossa C. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. *Mutat Res* 2008, 659:248–261.

25. OECD. The OECD QSAR toolbox. 2015. Available at: http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm. (Accessed August 3, 2015).

26. Ideaconsult Ltd. Toxtree. 2013. Available at: http://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxtree. (Accessed August 3, 2015).

27. Benigni R, Bossa C, Tcheremenskaia O. Nongenotoxic carcinogenicity of chemicals: mechanisms of action and early recognition through a new set of structural alerts. *Chem Rev* 2013, 113:2940–2957.

28. Hewitt M, Enoch S, Madden J, Przybylak K, Cronin M. Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action. *Crit Rev Toxicol* 2013, 43:537–558.

29. Gentile F, Chiatti L, Mauro F, Briganti G, Floridi A, Benassi M. Interaction of cytotoxic agents: a rule-based system for computer-assisted cell survival analysis. *Anticancer Res* 1992, 12:637–643.

30. Saliner A, Tsakovska I, Pavan M, Patlewicz G, Worth A. Evaluation of SARs for the prediction of skin irritation/corrosion potential: structural inclusion rules in the BfR decision support system. *SAR QSAR Environ Res* 2007, 18:331–342.

31. Tsakovska I, Saliner A, Netzeva T, Pavan M, Worth A. Evaluation of SARs for the prediction of eye irritation/corrosion potential: structural inclusion rules in the BfR decision support system. *SAR QSAR Environ Res* 2007, 18:221–235.

32. Gerner I, Barratt M, Zinke S, Schlegel K, Schlede E. Development and prevalidation of a list of structure-activity relationship rules to be used in expert systems for prediction of the skin-sensitising properties of chemicals. *Altern Lab Anim* 2004, 32:487–509.

33. Woo YT, Lai DY. Oncologic: a mechanism based expert system for predicting the carcinogenic potential of chemicals. In: Helma C, ed. *Predictive Toxicology - The Book*. Boca Raton, FL: CRC Press; 2005.

34. Lhasa Limited. Derek Nexus. 2014. Available at: http://www.lhasalimited.org/products/derek-nexus.htm. (Accessed on August 3, 2015).

35. CompuDrug Ltd. HazardExpert Pro. 2013. Available at: http://www.compudrug.com/hazardexpertpro. (Accessed August 3, 2015).

36. Lhasa Limited. Meteor Nexus. 2014. Available at: http://www.lhasalimited.org/products/meteor-nexus.htm. (Accessed August 3, 2015).

37. 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.

38. Poroikov VV, Filimonov DA, Borodina YV, Lagunin AA, Kos A. Robustness of biological activity spectra predicting by computer program PASS for non-congeneric sets of chemical compounds. *J Chem Inf Model* 2000, 40:1349–1355.

39. Cunningham AR, Moss ST, Iype SA, Qian G, Qamar S, Cunningham SL. Structure-activity relationship analysis of rat mammary carcinogens. *Chem Res Toxicol* 2008, 21:1970–1982.

40. Borgelt C, Berthold MR. Mining molecular fragments: finding relevant substructures of molecules. In: Kumar V, Tsurnoto S, Zhong N, Yu PS, Wu X, eds. *IEEE International Conference on Data Mining*, Maebashi City, Japan. Piscataway, NJ: IEEE; 2002, 51–58.

41. Yan X, Han J. gSpan: graph-based substructure pattern mining. In: Kumar V, Tsurnoto S, Zhong N, Yu PS, Wu X, eds. *IEEE International Conference on Data Mining*, Maebashi City, Japan. Piscataway, NJ: IEEE; 2002, 721–724.

42. Huan J, Wang W, Prins J. Efficient mining of frequent subgraph in the presence of isomorphism. In: Wu X, Tuzhilin A, eds. *Proceedings of the Third IEEE International Conference on Data Mining*, Melbourne, FL. Washington, DC: IEEE Computer Society; 2003, 549–552.

43. Nijssen S, Kok JN. A Quickstart in frequent structure mining can make a Difference. In: Kohavi R, Gehrke J, DuMouchel W, Ghosh J, eds. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA. New York, NY: ACM; 2004, 647–652.

44. Enoch SJ. Chemical category formation and read-across for the prediction of toxicity. In: Puzyn T, Leszczynski J, Cronin MT, eds. *Recent Advances in QSAR Studies*, vol. 8. Dordrecht, Netherlands: Springer; 2009, 209–219.

45. Jeliazkova N, Jaworska J, Worth AP. Open source tools for read-across and category formation. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 408–445.

46. OECD. *Guidance on Grouping of Chemicals*. 2nd ed. Paris, France: OECD; 2014.

47. Dimitrov S, Mekenyan O. An introduction to read-across for the prediction of the effects of chemicals. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 372–384.

48. Vink SR, Mikkers J, Bouwman T, Marquart H, Kroese ED. Use of read-across and tiered exposure assessment in risk assessment under REACH – a case study on a phase-in substance. *Regul Toxicol Pharmacol* 2010, 58:64–71.

49. Piparo EL, Maunz A, Helma C, Vorgrimmler D, Schilter B. Automated and reproducible read-across like models for predicting carcinogenic potency. *Regul Toxicol Pharmacol* 2014, 70:370–378.

50. Rorije E, Aldenberg T, Peijnenburg W. Read-across estimates of aquatic toxicity for selected fragrances. *Altern Lab Anim* 2013, 41:77–90.

51. Fabjan E, Hulzebos E, Mennes W, Piersma A. A category approach for reproductive effects of phthalates. *Crit Rev Toxicol* 2006, 36:695–726.

52. Enoch S, Cronin M, Schultz W, Madden J. Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem Res Toxicol* 2008, 21:513–520.

53. Koleva Y, Madden J, Cronin M. Formation of categories from structure-activity relationships to allow read-across for risk assessment: toxicity of alpha,beta-unsaturated carbonyl compounds. *Chem Res Toxicol* 2008, 21:2300–2312.

54. Patlewicz G, Jeliazkova N, Saliner AG, Worth A. Tox-match-a new software tool to aid in the development and evaluation of chemically similar groups. *SAR QSAR Environ Res* 2008, 19:397–412.

55. Ideaconsult Ltd. AMBIT. 2005. Available at: http://ambit.sourceforge.net/index.html. (Accessed August 3, 2015).

56. Ideaconsult Ltd. AMBIT discovery. 2005. Available at: http://ambit.sourceforge.net/download_ambitdiscovery.html. (Accessed August 3, 2015).

57. EPA Office of Pollution Prevention and Toxics (OPPT). Analog identification methodology (AIM). 2013. Available at: http://www.epa.gov/oppt/sf/tools/aim.htm. (Accessed August 3, 2015).

58. Richard A, Yang C, Judson R. Toxicity data informatics: supporting a new paradigm for toxicity prediction. *Toxicol Mech Methods* 2008, 18:103–118.

59. U.S. National Library of Medicine. ChemIDplus. 2015. Available at: http://chem.sis.nlm.nih.gov/chemidplus/. (Accessed August 3, 2015).

60. El-Masri H. Modeling for regulatory purposes (risk and safety assessment). In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 297–303.

61. Bliss CI, Steven WL. The calculation of the time mortality curve. *Ann Appl Biol* 1937, 24:815–852.

62. Brown KG, Foureman GL. Concentration–time–response modeling for acute and short-term exposures. *Regul Toxicol Pharmacol* 2005, 43:45–54.

63. Miller FJ, Schlosser PM, Janszen DB. Haber's rule: a special case in a family of curves relating concentration and duration of exposure to a fixed level of response for a given endpoint. *Toxicol Appl Pharmacol* 2000, 149:21–34.

64. Bliss CI. The relation between exposure time, concentration and toxicity in experiments on insecticides. *Ann Entomol Soc Am* 1940, 33:721–766.

65. Bliss CI. The calculation of the dosage-mortality curve. *Ann Appl Biol* 1935, 22:134–167.

66. Waters M, Stasiewicz S, Merrick A, Tomer K, Bushel P, Paules R, Stegman N, Nehls G, Yost K, Johnson C, et al. CEBS—chemical effects in biological systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. *Nucleic Acids Res* 2008, 36:D892–D900.

67. Bolton EE, Wang Y, Thiessen PA, Bryant SH. Pub-Chem: integrated platform of small molecules and biological activities. In: Wheeler RA, Spellmeyer DC, eds. *Annual Reports in Computational Chemistry*, vol. 4. Amsterdam, Netherlands: Elsevier B.V.; 2008, 217–241.

68. Judson RS, Martin MT, Egeghy P, Gangwal S, Reif DM, Kothiya P, Wolf M, Cathey T, Transue T, Smith D, et al. Aggregating data for computational toxicology applications: the U.S. Environmental Protection Agency (EPA) aggregated computational toxicology resource (ACToR) system. *Int J Mol Sci* 2012, 13:1805–1831.

69. Appelt A, Bentzen S, Jakobsen A, Vogelius I. Dose–response of acute urinary toxicity of long-course pre-operative chemoradiotherapy for rectal cancer. *Acta Oncol* 2015, 54:179–186.

70. Pottenger L, Schisler M, Zhang F, Bartels M, Fontaine D, McFadden L, Gollapudi BB. Dose–response and operational thresholds/NOAELs for in vitro mutagenic effects from DNA-reactive mutagens, MMS and MNU. *Mutat Res* 2009, 678:138–147.

71. Hunt D, Rai S, Li C. Summary of dose–response modeling for developmental toxicity studies. *Dose Response* 2008, 6:352–368.

72. Sung JH, Srinivasan B, Esch MB, McLamb WT, Bernabini C, Shuler ML, Hickman JJ. Using PBPK guided "Body-on-a-Chip" systems to predict mammalian response to drug and chemical exposure. *Exp Biol Med* 2014, 239:1225–1239.

73. Ingrisch M, Sourbron S. Tracer-kinetic modeling of dynamic contrast-enhanced MRI and CT: a primer. *J Pharmacokinet Pharmacodyn* 2013, 40:281–300.

74. Mager DE, Wyska E, Jusko WJ. Diversity of mechanism-based pharmacodynamic models. *Drug Metab Dispos* 2003, 31:510–519.

75. Andersen ME, Dennison JE. Mode of action and tissue dosimetry in current and future risk assessments. *Sci Total Environ* 2001, 274:3–14.

76. Crump KS, Chen C, Chiu WA, Louis TA, Portier CJ, Subramaniam RP, White PD. What role for biologically based dose–response models in estimating low-dose risk? *Environ Health Perspect* 2010, 118:585–588.

77. Haber LT, Maier A, Zhao Q, Dollarhide JS, Savage RE, Dourson ML. Applications of mechanistic data in risk assessment: the past, present, and future. *Toxicol Sci* 2001, 61:32–39.

78. Egorov AI, Dalbokova D, Krzyzanowski M. Biomonitoring-based environmental public health indicators. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 275–293.

79. Bois FY. Bayesian inference. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 597–636.

80. Avent ML, Vaska VL, Rogers BA, Cheng AC, van Hal SJ, Holmes NE, Howden BP, Paterson DL. Vancomycin therapeutics and monitoring: a contemporary approach. *Intern Med J* 2013, 43:110–119.

81. Certara. Phoenix WinNonlin. 2015. Available at: https://www.certara.com/products/pkpd/phx-wnl. (Accessed August 3, 2015).

82. Thermo Scientific. Kinetica. 2007. Available at: http://www.thermoscientific.com/content/tfs/en/product/kinetica-software-pk-pd-data-analysis-simulation-reporting.html. (Accessed August 3, 2015).

83. D'Argenio DZ, Schumitzky A, Wang X. ADAPT 5 user's guide: pharmacokinetic/pharmacodynamic systems analysis software. 2009. Available at: http://bmsr.usc.edu/software/adapt/. (Accessed August 3, 2015).

84. Chiu WA, White P. Steady-state solutions to PBPK models and their applications to risk assessment I: route-to-route extrapolation of volatile chemicals. *Risk Anal* 2006, 26:769–780.

85. Andersen M. Development of physiologically based pharmacokinetic and physiologically based pharmacodynamic models for applications in toxicology and risk assessment. *Toxicol Lett* 1995, 79:35–44.

86. Clewell HJ, Thomas RS, Gentry PR, Crump KS, Kenyon EM, El-Masri HA, Yager JW. Research toward the development of a biologically based dose response assessment for inorganic arsenic carcinogenicity: a progress report. *Toxicol Appl Pharmacol* 2007, 222:388–398.

87. Falk-Filipsson A, Hanberg A, Victorin K, Warholm M, Wallén M. Assessment factors – applications in health risk assessment of chemicals. *Environ Res* 2007, 104:108–127.

88. Martin OV, Martin S, Kortenkamp A. Dispelling urban myths about default uncertainty factors in chemical risk assessment – sufficient protection against mixture effects? *Environ Health* 2013, 12:53.

89. Hasegawa R, Hirata-Koizumi M, Dourson ML, Parker A, Sweeney LM, Nishikawa A, Yoshida M, Ono A, Hirose A. Proposal of new uncertainty factor application to derive tolerable daily intake. *Regul Toxicol Pharmacol* 2010, 58:237–242.

90. Guha R. The ups and downs of structure-activity landscapes. In: Bajorath J, ed. *Chemoinformatics and Computational Chemical Biology*, vol. 672. New York: Humana Press; 2011, 101–117.

91. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. *Nature* 1962, 194:178–180.

92. Gramatica P. On the development and validation of QSAR models. In: Reisfeld B, Mayeno AN, eds. *Computational Toxicology*, vol. 930. New York: Humana Press; 2013, 499–526.

93. Cronin MTD, Madden JC. Using in silico toxicity predictions: case studies for skin sensitisation. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 606–621.

94. Accelrys. TOPKAT (TOxicity Prediction by Komputer Assisted Technology). 2015. Available at: http://accelrys.com/products/discovery-studio/admet.html. (Accessed August 3, 2015).

95. IRCCS - Istituto di Ricerche Farmacologiche Mario Negri, CRS4, KnowledgeMiner Software, Kemijski inštitut, Kode S.R.L., Politecnico di Milano, US EPA, UK FERA. VEGA. 2015. Available at: http://www.vega-qsar.eu/. (Accessed August 3, 2015).

96. Greene N, Judson PN, Langowski JJ, Marchant CA. Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. *SAR QSAR Environ Res* 1999, 10:299–314.

97. Weaver S, Gleeson MP. The importance of the domain of applicability in QSAR modeling. *J Mol Graph Model* 2008, 26:1315–1326.

98. Warheit DB. How meaningful are the results of nanotoxicity studies in the absence of adequate material characterization? *Toxicol Sci* 2008, 101:183–185.

99. Wani MY, Hashim MA, Nabi F, Malik MA. Nanotoxicity: dimensional and morphological concerns. *Adv Phys Chem* 2011, 2011:450912.

100. Winkler DA, Mombelli E, Pietroiusti A, Tran L, Worth A, Fadeel B, McCall MJ. Applying quantitative structure–activity relationship approaches to nanotoxicology: current status and future potential. *Toxicology* 2013, 313:15–23.

101. Beddoes CM, Case CP, Briscoe WH. Understanding nanoparticle cellular entry: a physicochemical perspective. *Adv Colloid Interface Sci* 2015, 218:48–68.

102. Zhang W, Yao Y, Sullivan N, Chen Y. Modeling the primary size effects of citrate-coated silver nanoparticles on their ion release kinetics. *Environ Sci Technol* 2011, 45:4422–4428.

103. Sizochenko N, Rasulev B, Gajewicz A, Kuz'min V, Puzyn T, Leszczynski J. From basic physics to mechanisms of toxicity: the "liquid drop" approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles. *Nanoscale* 2014, 6:13986–13993.

104. Fu PP, Xia Q, Hwang HM, Ray PC, Yu H. Mechanisms of nanotoxicity: generation of reactive oxygen species. *J Food Drug Anal* 2014, 22:64–75.

105. Gajewicz A, Cronin MTD, Rasulev B, Leszczynski J, Puzyn T. Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: nano-readacross. *Nanotechnology* 2015, 26:015701.

106. Li M, Reineke J. Physiologically based pharmacokinetic modeling for nanoparticle toxicity study. In: Reineke J, ed. *Nanotoxicity*. Humana Press: New York; 2012, 369–382.

107. Silva T, Pokhrel LR, Dubey B, Tolaymat TM, Maier KJ, Liu X. Particle size, surface charge and concentration dependent ecotoxicity of three organo-coated silver nanoparticles: comparison between general linear model-predicted and observed toxicity. *Sci Total Environ* 2014, 468–469:968–976.

108. Ahmed L, Rasulev B, Turabekova M, Leszczynska D, Leszczynski J. Receptor- and ligand-based study of fullerene analogs: comprehensive computational approach including quantum-chemical, QSAR and molecular docking simulations. *Org Biomol Chem* 2013, 11:5798–5808.

109. Pathakoti K, Huang MJ, Watts JD, He X, Hwang HM. Using experimental data of Escherichia coli to develop a QSAR model for predicting the photo-induced cytotoxicity of metal oxide nanoparticles. *J Photochem Photobiol B* 2014, 130:234–240.

110. Puzyn T, Gajewicz A, Leszczynska D, Leszczynski J. Nanomaterials – the next great challenge for QSAR modelers. In: Puzyn T, Leszczynski J, Cronin MT, eds. *Recent Advances in QSAR Studies*, vol. 8. Dordrecht, Netherlands: Springer; 2009, 209–219.

111. Henn BC, Coull BA, Wright RO. Chemical mixtures and children's health. *Curr Opin Pediatr* 2014, 26:223–229.

112. Løkke H, Ragas AMJ, Holmstrup M. Tools and perspectives for assessing chemical mixtures and multiple stressors. *Toxicology* 2013, 313:73–82.

113. Sarigiannis DA, Hansen U. Considering the cumulative risk of mixtures of chemicals – a challenge for policy makers. *Environ Health* 2012, 11:S18.

114. Bliss CI. The toxicity of poisons applied jointly. *Ann Appl Biol* 1939, 26:585–615.

115. Hadrup N, Taxvig C, Pedersen M, Nellemann C, Hass U, Vinggaard AM. Concentration addition, independent action and generalized concentration addition models for mixture effect prediction of sex hormone synthesis in vitro. *PLoS One* 2013, 8:e70490.

116. Howard GJ, Webster TF. Generalized concentration addition: a method for examining mixtures containing partial agonists. *J Theor Biol* 2009, 259:469–477.

117. Baas J, Jager T, Kooijman B. A review of DEB theory in assessing toxic effects of mixtures. *Sci Total Environ* 2010, 408:3740–3745.

118. Baas J, Jager T, Kooijman SA. A model to analyze effects of complex mixtures on survival. *Ecotoxicol Environ Saf* 2009, 72:669–676.

119. Kortagere S, Ekins S. Troubleshooting computational methods in drug discovery. *J Pharmacol Toxicol Methods* 2010, 61:67–75.

120. Cronin MTD. Finding the data to develop and evaluate (Q)SARs and populate categories for toxicity prediction. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 31–58.

121. Nendza M, Aldenberg T, Benfenati E, Benigni R, Cronin MTD, Escher S, Fernandez A, Gabbert S, Giralt F, Hewitt M, et al. Data quality assessment for in silico methods: a survey of approaches and needs. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 59–117.

122. Cronin MTD, Schultz WT. Pitfalls in QSAR. *J Mol Struc-Theochem* 2003, 622:39–51.

123. Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J, Mekenyan O. A stepwise approach for defining the applicability domain of SAR and QSAR models. *J Chem Inf Model* 2005, 45:839–849.

124. Hewitt M, Ellison CM. Developing the applicability domain of in silico models: relevance, importance and methods. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 301–333.

125. Schultz TW, Hewitt M, Netzeva TI, Cronin MTD. Assessing applicability domains of toxicological QSARs: definition, confidence in predicted values, and the role of mechanisms of action. *QSAR Comb Sci* 2007, 26:238–254.

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

127. Toropov AA, Benfenati E. Additive SMILES-based optimal descriptors in QSAR modelling bee toxicity:

using rare SMILES attributes to define the applicability domain. *Bioorg Med Chem* 2008, 16:4801–4809.

128. Hartung T. Lessons learned from alternative methods and their validation for a new toxicology in the 21st century. *J Toxicol Environ Health B Crit Rev* 2010, 13:277–290.

129. Human Toxicology Project Consortium. Human Toxicology Project Consortium. Available at: http://www.thehamner.org/content/HTOX_flyer_10.pdf. (Accessed October 17, 2015).

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

131. Tollefsen KE, Scholz S, Cronin MT, Edwards SW, de Knecht J, Crofton K, Garcia-Reyero N, Hartung T, Worth A, Patlewicz G. Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). *Regul Toxicol Pharmacol* 2014, 70:625–640.

132. National Toxicology Program. A National Toxicology Program for the 21st Century: roadmap to achieve the NTP vision. 2004. Available at: http://ntp.niehs.nih.gov/ntp/about_ntp/ntpvision/ntproadmap_508.pdf. (Accessed October 17, 2015).

133. National Research Council. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: The National Academies Press; 2007.

134. Balls M. Integrated testing strategies and the prediction of toxic hazard. In: Cronin MTD, Madden JC, eds. *In Silico Toxicology: Principles and Applications*. Cambridge, UK: The Royal Society of Chemistry; 2010, 584–605.

135. Kleensang A, Maertens A, Rosenberg M, Fitzpatrick S, Lamb J, Auerbach S, Brennan R, Crofton KM, Gordon B, Fornace AJ Jr, et al. t4 workshop report: pathways of toxicity. *ALTEX* 2014, 31:53–61.

136. Hareng L, Pellizzer C, Bremer S, Schwarz M, Hartung T. The integrated project ReProTect: a novel approach in reproductive toxicity hazard assessment. *Reprod Toxicol* 2005, 20:441–452.

137. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* 2007, 95:5–12.

138. Shukla SJ, Huang R, Austin CP, Xia M. The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. *Drug Discov Today* 2010, 15:997–1007.

139. Stephens ML, Barrow C, Andersen ME, Boekelheide K, Carmichael PL, Holsapple MP, Lafranconi M. Accelerating the development of 21st-century toxicology: outcome of a human toxicology project consortium workshop. *Toxicol Sci* 2012, 125:327–334.