COMMITTEE ON TOXICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT

Environmental, health and safety alternative testing strategies: Development of methods for potency estimation

Background

1. In October 2019 we introduced the topic of developing methods for potency estimation in an introductory paper (TOX/2019/61) which discussed the development/technologies of the current methodologies for prediction models.

2. The paper also introduced the planned combined workshop with physiologically based pharmacokinetic (PBPK) modelling[1] on 11th March 2020 in which experts/organisations will be invited to speak, review in-house case studies and have roundtable discussions on the topic.

Introduction

3. Advances in molecular biology, biotechnology, and other fields are paving the way for major improvements in how scientists evaluate the health risks posed by potentially toxic chemicals. These advances would make toxicity testing quicker, less expensive, and more directly relevant to human exposures. They could also reduce the need for animal testing by substituting these for more laboratory tests based on human cells.

4. One of the major state-of-the-art methods is potency estimation via a collective multidisciplinary approach. Potency is a measure of the chemical activity expressed in terms of the amount required to produce an effect of given intensity. Potency estimates can be used to directly compare chemical profiles and prioritize compounds for confirmation studies or employed as input data for prediction modelling and association mapping.

5. Many methods have been developed to predict the toxicity of chemicals. In the following paper we discuss the paradigm shift in toxicity testing, the development of the current methodologies for prediction models.

6. In 2009, The Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (COT) held a workshop on 21st century toxicology[2]. The workshop addressed the US National Academy report called *Toxicity Testing in the 21st Century: A Vision and a Strategy*[3]. The report called for accelerated

[1] Physiologically based pharmacokinetic (PBPK): a mathematical modelling technique for predicting the absorption, distribution, metabolism and excretion (ADME) of synthetic or natural chemical substances in humans and other animal species. PBPK modelling is used in pharmaceutical research and drug development, and in health risk assessment for cosmetics or general chemicals.
[2] https://cot.food.gov.uk/cotmtgs/cotmtsem/cotwrkshop11feb09
[3] https://www.nap.edu/catalog/11970/toxicity-testing-in-the-21st-century-a-vision-and-a

development and adoption of human cell *in vitro* and *in silico* methods for the prediction of hazards, the determination of mechanistic information, and the integration of data.

7. The National Academy report set out a 10-20 year strategy in which the goal would be to develop and validate toxicological protocols that move away from testing in animals through use of *in vitro* and computer-based (*in silico*) assessments of toxicity and mechanisms. The aim is to enable predictions of human *in vivo* responses to chemicals in a high throughput and cost-effective manner, with less use of experimental animals. Among other things, this might facilitate toxicological assessment of combined exposure to multiple chemicals, which has been an area of increasing interest in recent years.

8. As we are half way through the vision and strategy (10 years) it would be apt to review the current methodologies available.

Food Standards Agency (FSA) requirement for potency estimation

9. The FSA have previously put forward a business case for potency estimation to aid in risk assessment. When responding to food incidents[4] we regularly have chemicals, particularly novel foods and sports/dietary supplements such as selective androgen receptor modulators where certain ingredients have very little or no toxicological information. For certain novel ingredients, a lot of which tend to be from plants and have a history of medical use in certain parts of the world, again there is very little toxicological information and sometimes it is not possible to give any risk advice to our FSA Policy colleagues.

10. The possible toxicological values for the chemical can be estimated by *in silico* models from chemicals with a similar structure or in the same group. A method or approach which could provide a means of estimating the potency of these chemicals could improve the accuracy of the information and confidence in the risk assessment. An *in vitro/in silico* approach that can provide information on the relative potencies would provide essential information for toxicity prediction, where information is only available on 1 or 2 compounds from the group. This will allow the identification of the level of risk from a given chemical and give greater confidence in risk assessments that individual compounds can be assessed, not just assuming that all compounds have the same toxicological potency.

11. This will be fundamental in risk assessment scenarios where limited to no information is available on the toxicity of a chemical.

Chemical Landscape

12. Thousands of chemicals are in common use, but only a portion of them have undergone significant toxicologic evaluation, and as more emerge it is important to prioritize the remainder for targeted testing (Judson *et al*., 2009). This is specifically important for chemicals (found in food and in the environment) where little or no toxicological information is available.

[4] https://www.food.gov.uk/business-guidance/food-incidents-product-withdrawals-and-recalls

13. To address this issue, organizations are developing chemical screening and prioritization programmes. As part of these efforts, it is important to catalogue, from widely dispersed sources, the toxicological information that is available (Richard *et al*., 2016).

14. The main objective of these programmes is to define a list of chemicals that are candidates for screening and prioritization process, and to catalogue the available toxicological information.

15. The combined advances in discovery and clinical sciences, data science and technology[5] has resulted in toxicity testing reaching a pivotal transformation point. The advances in the technology and science sector are taking advantage of the 4th industrial revolution (4IR)[6].

Toxicology Testing in the 21st Century (Tox21)

16. The phrase '21st century toxicology' (Tox 21) (Hartung, 2010) refers to 'the transformation underway in the tools and approaches used to evaluate chemical substances for possible effects on human health'[7]. Tox21 focuses on toxicity pathways (Bhattacharya *et al*., 2011) mechanisms, modes of action, and adverse outcome pathways (AOP) (Tollefsen *et al*., 2014) in humans.

17. Another related concept is the 3Rs (Hartung, 2010) which was proposed 50 years ago in the publication of Russell and Burch (1959)[8]:

- Replace: Methods which avoid or replace the use of animals
- Reduce: Methods which minimise the number of animals used per experiment
- Refine: Methods which minimise the number of animals used per experiment

18. The principles of the 3Rs is providing a framework for performing more humane animal research.

19. In 2004, The National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) was set up in the United Kingdom (UK). NC3Rs[9] is the national organisation for the 3Rs. Their strategy is to advance the 3Rs by focusing on their scientific impacts and benefits. They have re-defined the standard 3Rs definitions so that they are more reflective of contemporary scientific practice and developments.

20. Several strategies have been proposed to implement Tox21. 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

[5] http://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Health_Council_Report.pdf
[6] The Fourth Industrial Revolution (4IR) is the fourth major industrial era since the initial Industrial Revolution of the 18th century. It is characterized by a fusion of technologies that is blurring the lines between the physical, digital and biological spheres, collectively referred to as cyber-physical systems
[7] National Research Council, 2007. Toxicity testing in the 21st century: a vision and a strategy. National Academies Press.
[8] https://www.nc3rs.org.uk/the-3rs
[9] https://www.nc3rs.org.uk/

predominantly predictive science focused upon a broad inclusion of target-specific, mechanism-based, biological observations'.[10]

21. Tox 21 is a federal collaboration among the Environmental Protection Agency (EPA)[11], the National Institutes of Health (NIH)[12], including the National Center for Advancing Translational Sciences (NCATS)[13] and the National Toxicology Program at the National Institute of Environmental Health Sciences (NIEHS)[14], and the Food and Drug Administration (FDA)[15]. Tox21 researchers aim to develop better toxicity assessment methods to quickly and efficiently test whether certain chemical compounds have the potential to disrupt processes in the human body that may lead to negative health effects. One of the EPA's contributions to Tox21 is the chemical screening results from the Toxicity Forecaster (ToxCast).

22. Using a high-throughput screening system[16] housed at NCATS, researchers are testing 10,000 environmental chemicals (called the "Tox21 10K library"[17]) for their potential to disrupt biological pathways that may result in toxicity. Screening results help the researchers prioritize chemicals for further in-depth investigation.

Toxicity Forecasting

23. The EPA needs rapid and efficient methods to prioritize, screen and evaluate thousands of chemicals. The EPA's Toxicity Forecaster (ToxCast[18]) generates data and predictive models on thousands of chemicals of interest to the EPA. ToxCast[19] uses high throughput screening methods and computational toxicology approaches to rank and prioritize chemicals. As a result, the EPA's Endocrine Disruption Screening Program (EDSP)[20] is working to use ToxCast to identify priority chemicals.

24. ToxCast has data for approximately 1,800 chemicals from a broad range of sources including industrial and consumer products, food additives, and potentially green chemicals[21] that could be safer alternatives to existing chemicals. ToxCast screens chemicals in more than 700 high-throughput assay endpoints that cover a range of high-level cell responses. Part of the EPA's contribution to the "Toxicology in the 21st century" federal agency collaboration is some of the ToxCast data.

[10] https://ntp.niehs.nih.gov/ntp/about_ntp/ntpvision/ntproadmap_508.pdf
[11] https://www.epa.gov/
[12] https://www.nih.gov/
[13] https://ncats.nih.gov/
[14] https://www.niehs.nih.gov/
[15] https://www.fda.gov/home
[16] High-throughput screening (HTS) is a method for scientific experimentation especially used in drug discovery and relevant to the fields of biology and chemistry. Using robotics, data processing/control software, liquid handling devices, and sensitive detectors, high-throughput screening allows a researcher to quickly conduct millions of chemical, genetic, or pharmacological tests. Through this process one can rapidly identify active compounds, antibodies, or genes that modulate a particular biomolecular pathway. The results of these experiments provide starting points for drug design and for understanding the noninteraction or role of a particular location.
[17] https://ntp.niehs.nih.gov/whatwestudy/tox21/toolbox/index.html
[18] https://www.epa.gov/chemical-research/toxicity-forecasting
[19] ToxCast: program within the U.S. Environmental Protection Agency (EPA) employs high-throughput in vitro assays to efficiently screen large numbers of chemicals to support the development of improved toxicity prediction models, particularly to be applied to environmental chemicals for which limited or no in vivo animal toxicity data are available.
[20] https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-overview
[21] Green chemicals/chemistry, also called sustainable chemistry, is an area of chemistry and chemical engineering focused on the designing of products and processes that minimize or eliminate the use and generation of hazardous substances.

25. Another one of the goals of the Tox21 collaboration is to establish *in vitro* signatures of *in vivo* human and rodent toxicity (*i.e. in vitro* to *in vivo* extrapolation[22]) which include cytotoxicity, cellular pathway assays and computer modelling. Some examples include adverse outcome pathways[23] cardiotoxicity[24], skin sensitisation[25] and organ on a chip (Maschmeyer *et al*., 2015).

26. In 2007, the National Research Council (NRC) published a report "*Toxicity Testing in the 21st Century: a Vision and a Strategy*"[26], which proposed using computational methods *i.e. in silico* 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.[27] This might also facilitate toxicological assessment of combined exposure to multiple chemicals, which has been an area of increasing interest in recent years.

Integrated Approaches to Testing and Assessment

27. Integrated approaches to testing and assessment (IATAs) provide a means by which all relevant and reliable existing information about a chemical can be used to answer a defined hazard characterization question. Information considered can include toxicity data, exposure routes, use cases, and production volumes. This information is used to characterize outcomes that can inform regulatory decision-making.

28. The drawbacks of traditional toxicity testing approaches using laboratory animals may be overcome by the use of human cell-based, biochemical, and/or computational methods to predict chemical toxicity. Due to the complexity of toxicity mechanisms, data from several methods usually need to be considered in combination to adequately predict toxic effects. IATAs provide a means by which these data can be considered in combination. When necessary, IATAs can guide generation of new data, preferably using non-animal approaches, to inform regulatory decision-making[28].

Legislation and Laws

USA

Frank R. Lautenberg Chemical Safety for the 21st Century Act

29. The Frank R. Lautenberg Chemical Safety for the 21st Century Act[29] is a law passed by the 114th United States Congress and signed into law by US President Barack Obama in 2016. Administered by the US EPA, which regulates the introduction of new or already existing chemicals, the Act amends and updates the

[22] https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/comptox/ct-ivive/ivive.html
[23] https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/comptox/ct-aop/aop.html?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=niceatm-aop
[24] https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/cardio/index.html
[25] https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/immunotoxicity/nonanimal/index.html
[26] https://www.nap.edu/catalog/11970/toxicity-testing-in-the-21st-century-a-vision-and-a
[27] https://www.nap.edu/read/11970/chapter/1#iii
[28] https://ntp.niehs.nih.gov/whatwestudy/niceatm/integrated-testing-strategies/index.html
[29] https://www.congress.gov/bill/114th-congress/house-bill/2576

Toxic Substances Control Act[30] (TSCA) [31] that went into force in 1976, the nation's primary chemicals management law.

30. Among the implementation strategies they issued a Strategic Plan to promote the development and implementation of alternative test methods and strategies to reduce, refine or replace vertebrate animal testing[32].

31. The TSCA, as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act, directs the EPA to:

- Reduce and replace, to the extent practicable and scientifically justified, the use of vertebrate animals in the testing of chemical substances or mixtures.

- Promote the development and timely incorporation of alternative test methods or strategies that do not require new vertebrate animal testing.

32. The TSCA also requires the EPA to develop a strategic plan on this topic and provide a progress report on the implementation of the plan to Congress every five years since the date of the enactment of the Lautenberg Chemical Safety Act, *i.e.* beginning in 2021.

33. In 2018, the EPA published its Strategic Plan to Promote the Development and Implementation of Alternative Test Methods within the TSCA Program. The Strategic Plan incorporated input from two public meetings and written comments submitted on the draft strategic plan.

34. The Strategic Plan has three core components: (1) identifying, developing and integrating New Approach Methodologies (NAMs) for TSCA decisions; (2) building confidence that the NAMs are scientifically reliable and relevant for TSCA decisions; and (3) implementing the reliable and relevant NAMs for TSCA decisions.

35. On September 10th 2019, the U.S. EPA signed a directive that prioritizes efforts to reduce animal testing[33]. It stated that the EPA's plan would identify tangible steps to ensure that the Agency's regulatory, compliance and enforcement activities, including chemical and pesticide approvals and agency research, remain fully protective of human health and the environment while pursing these reduction goals. Among them, "*demonstration that NAMs are applicable for use in risk assessment and that new decision-making approaches are as protective of human health and the environment as existing approaches".*

[30] https://www.epa.gov/laws-regulations/summary-toxic-substances-control-act
[31] The Toxic Substances Control Act (TSCA) is a United States law, passed by the United States Congress in 1976 and administered by the United States Environmental Protection Agency (EPA), that regulates the introduction of new or already existing chemicals.
[32] https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/alternative-test-methods-and-strategies-reduce
[33] https://www.epa.gov/research/administrator-memo-prioritizing-efforts-reduce-animal-testing-september-10-2019

European Union (EU)

Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)

36. REACH is an EU regulation dating from 18 December 2006[34]. REACH addresses the production and use of chemical substances, and their potential impacts on both human health and the environment. It is the strictest law to date regulating chemical substances and will affect industries throughout the world. REACH entered into force on 1 June 2007, with a phased implementation over the next decade. The regulation also established the European Chemicals Agency (ECHA)[35], which manages the technical, scientific and administrative aspects of REACH.

Animal testing under REACH

37. REACH aims to ensure a high level of protection of human health and the environment from effects of hazardous chemicals. It strives for a balance: to increase our understanding of the possible hazards of chemicals while at the same time avoiding unnecessary testing on animals. Learning more about chemicals sometimes requires testing them on animals as a last resort. Registrants may only carry out new tests using animals when they have exhausted all other relevant and available data sources[36].

38. On January 2017, REACH updated its information requirements in which they made non-animal testing the default method for skin corrosion/irritation, serious eye damage/eye irritation and skin sensitisation. In addition, the guidance introduces a new possibility, to use a weight-of-evidence approach for acute toxicity.

39. The guidance update expands on the information on alternative methods[37] and clarifies when and how to use them for REACH purposes.

40. The Cosmetics Regulation[38] prohibits the placing on the market of cosmetic products, or products containing ingredients, which have been tested on animals to meet the requirements of that regulation using a method other than a validated alternative method.

Potency Estimation

41. Potency measures can be applied for rapid identification of pharmacoactive hits or toxicological assessment and used as input data for prediction modelling or association mapping[39].

[34] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02006R1907-20140410
[35] https://echa.europa.eu/home
[36] https://echa.europa.eu/animal-testing-under-reach
[37] https://echa.europa.eu/support/testing-methods-and-alternatives
[38] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32009R1223
[39] Association mapping, also known as "linkage disequilibrium mapping", is a method of mapping quantitative trait loci that takes advantage of historic linkage disequilibrium to link phenotypes to genotypes, uncovering genetic associations.

42. 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*. LD_{50}: lethal dose to 50% of tested individuals) or qualitative, such as binary (*e.g.* toxic or non-toxic) or ordinary (*e.g*. low, moderate, or high toxicity).

43. 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 (repeat dose) (Raies and Bajic 2016).

44. The toxicity of a substance usually depends on the following factors[40]:

- Form and innate chemical activity
- Dosage, especially dose-time relationship
- Exposure route
- Species
- Life stage, such as infant, young adult, or elderly adult
- Gender
- Ability to be absorbed
- Metabolism
- Distribution within the body
- Excretion
- Health of the individual, including organ function and pregnancy, which involves physiological changes that could influence toxicity
- Nutritional status
- Presence of other chemicals
- Circadian rhythms (the time of day a drug or other substance is administered)

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

46. Using the results of animal tests to predict human health effects involves a number of assumptions and extrapolations that remain a topic of debate.

47. This paper will outline some of the current methodologies available.

In silico

48. *In silico* is an expression meaning "performed on computer or via computer simulation" in reference to biological experiments.

[40] https://toxtutor.nlm.nih.gov/03-002.html

49. *In silico* toxicity modelling is carried out using computational resources (*i.e.* methods, algorithms, software, data) to organise, analyse, model, simulate, visualize or predict toxicity of chemicals (Deeb *et al*., 2012; Valerio, 2009). Computational methods aim to complement *in vitro*/*in vivo* toxicity testing to potentially minimize the need for animal testing, reduce the cost and time of toxicity tests and improve toxicity prediction.

50. However, *in silico* techniques are used in a wide variety of scenarios within and between industries including, but not limited to, screening, prioritisation, classification and labelling, risk assessment, and product development.

51. As an example, within the pharmaceutical industry, knowledge-based systems and Quantitative Structure Activity Relationships (QSAR)s are used to predict mutagenicity of impurities as part of the ICH Harmonised Guideline M7 scheme[41] (Amberg *et al*., 2018).

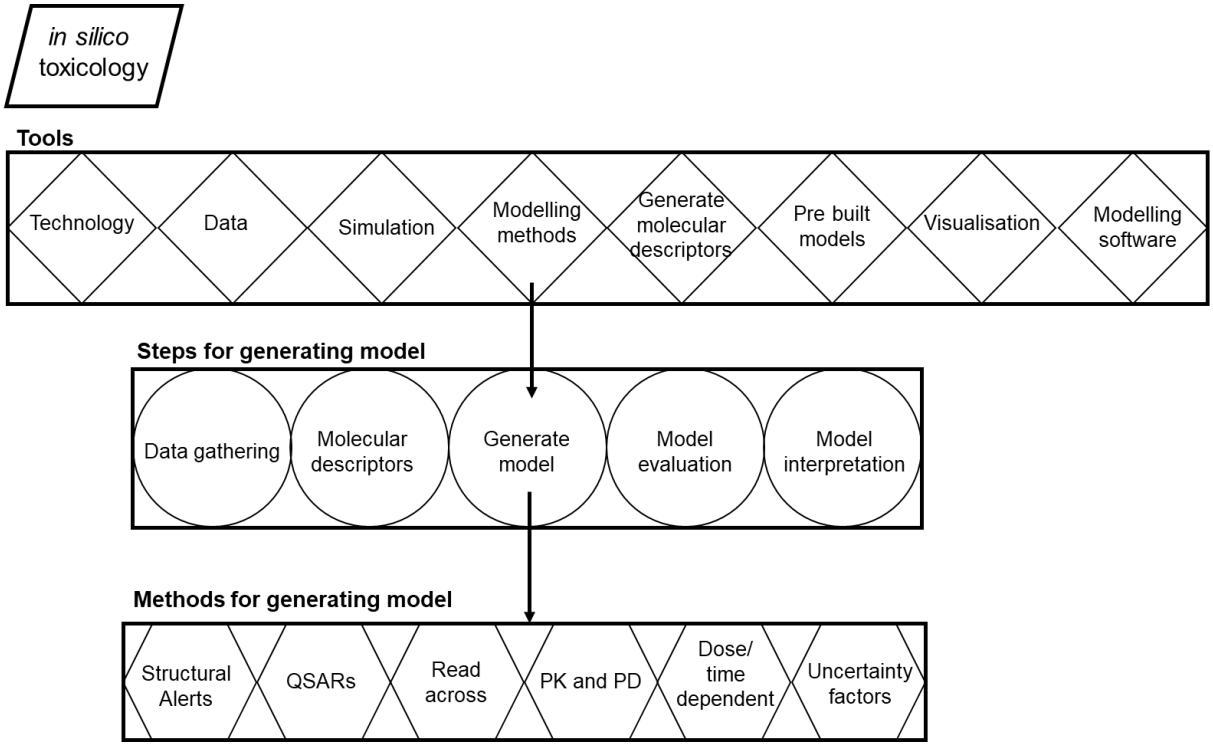
52. The cosmetics industry, foresee the use of *in silico* techniques as part of an *ab initio*[42] approach to assess the overall impact of a chemical. The assessment typically includes information on mechanisms of action, exposure, and uses case scenarios, as well as the more traditional and accepted use for toxicity prediction (Berggren *et al*., 2015).

Overview of in silico toxicology

53. *In silico* toxicology encompasses a wide variety of computational tools (Figure 1): databases for storing data about chemicals, their toxicity, and chemical properties; software for generating molecular descriptors; simulation tools for systems biology and molecular dynamics; modelling methods for toxicity prediction; modelling tools such as statistical packages and software for generating prediction models; expert systems that include pre-built models in web servers or standalone applications for predicting toxicity; and visualization tools. In general, modelling methods include the following steps while developing prediction models (Figure 1): gathering biological data that contain associations between chemicals and toxicity endpoints, calculating molecular descriptors of the chemicals, generating a prediction model, evaluating the accuracy of the model, and interpreting the model.

[41] http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_R1_Addendum_Step_4_31Mar2017.pdf

[42] *Ab initio*: a Latin term meaning "from the beginning".



Key
QSARs: Quantitative Structure Activity Relationships
PK: Pharmacokinetic
PD: Pharmacodynamic

Figure 1. Overview of *in silico* toxicology. Tools, steps for generating model and methods for generating model (Figure adopted from Raies and Bajic 2016).

Models

Structural alerts and rule-based models

54. Structural alerts (SAs) also known as toxicophores[43] and “expert rules”, are molecular structures that indicate or associate to toxicity (Roncaglioni *et al*., 2013). SAs can consist of only one atom or several connected atoms (Lepailleur *et al*., 2013). Alerts have been used since a series of studies published on chemical carcinogenicity and mutagenicity (Ashby *et al*., 1991; Ashby *et al*., 1988, Ashby *et al*., 1985).

55. A combination of SAs may contribute to toxicity more than a single SA. 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, as illustrated in the following example:

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

[43] A toxicophore is a chemical structure or a portion of a structure (*e.g.*, a functional group) that is related to the toxic properties of a chemical. Toxicophores can act directly (*e.g.*, dioxins) or can require metabolic activation (*e.g.*, tobacco-specific nitrosamines).

56. There are two main types of rule-based models: human-based rules (HBRs) and induction-based rules (IBRs) (Venkatapathy *et al*., 2013).

57. HBRs are derived from human knowledge of subject matter expertise or from literature, but IBRs are derived computationally (Valerio *et al*., 2009). HBRs are limited to human knowledge that could be incomplete or biased but tend to be more accurate. However, updating HBR can be challenging to review and require extensive literature analysis.

58. 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. 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. These are used in drug design and help to determine how drugs should be altered to reduce their toxicity.

59. Using structure to predict toxicity allows for identifying the structure of potential metabolites too (Toropov *et al*., 2014). However, SAs use only binary features (*e.g*. chemical structures are either present or absent) and only qualitative endpoints (*e.g*. carcinogenic or non-carcinogenic) (Venkatapathy *et al*., 2013).

60. 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 (Milan *et al*., 2011).

61. The list of SAs and rules may be incomplete, which may cause a large number of false negatives predictions (Roncaglioni *et al*., 2013).

62. It has been discussed in the scientific community that most alerts represent functional groups or substructures that can be found in many compounds, both toxic and non-toxic, leading to predictions with overly high sensitivity.

63. Some SAs and rule-based models examples include:

- Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism (s) of action (Hewitt *et al*., 2013).

- Interaction of cytotoxic agents: a rule-based system for computer-assisted cell survival analysis (Gentile *et al*., 1992).

- Evaluation of SARs for the prediction of skin irritation/corrosion potential–structural inclusion rules in the BfR decision support system (Gallegos *et al*., 2007).

- 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 (Gerner *et al*., 2004).

64. Several systems provide pre-built rule-based/knowledge base and SAs lists, for example:

Toxtree[44] : open source application, which is able to estimate toxic hazard by applying a decision tree approach.

Derek Nexus[45] : knowledge-based toxicology predictions for various endpoints *e.g.* carcinogenicity, mutagenicity or skin sensitization.

HazardExpert[46] : initial estimation of toxic symptoms of organic compounds in humans and in animals through rule-based system. HazardExpert can also consider the bioavailability of the compounds.

Meteor[47] : knowledge base prediction for metabolic fate of chemical.

65. Other tools are available that can extract SAs from datasets that contains toxic or non-toxic compounds such as:

CASE: Computer-assisted structure elucidation is the technique of using software to generate all possible molecular structures that are consistent with a particular set of spectroscopic data (Jaspars *et al*., 1999).

PASS: prediction of activity spectra for substances is a web-based application that predicts the biological activity spectrum of a compound based on its structure (Parasuraman *et al*., 2011).

Cat-SAR: categorical-structure activity relationship expert system that analyses categorical data and two-dimensional fragments has been successfully used in the analysis of chemical compounds that cause toxicity (Kumar *et al*., 2014)

Chemical Category

66. A chemical category is a group of chemicals whose physicochemical and human health and/or ecotoxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern, usually as a result of structural similarity.

67. The OECD Guidance on Grouping of Chemicals[48] lists several methods for grouping, such as chemical identity and composition, physicochemical and Adsorption Distribution Metabolism and Excretion (ADME) properties, mechanism of action (MoA), and chemical/biological interactions.[49]

[44] http://toxtree.sourceforge.net/
[45] https://www.lhasalimited.org/products/derek-nexus.htm
[46] https://www.compudrug.com/hazardexpertpro
[47] https://www.lhasalimited.org/products/meteor-nexus.htm
[48] https://www.oecd.org/chemicalsafety/risk-assessment/groupingofchemicalschemicalcategoriesandread-across.htm
[49] https://read.oecd-ilibrary.org/environment/guidance-on-grouping-of-chemicals-second-edition_9789264274679-en#page1

68. The similarities may be based on the following:

- A common functional group (*e.g.* aldehyde, epoxide, ester, specific metal ion).
- Common constituents or chemical classes, similar carbon range numbers.
- An incremental and constant change across the category (*e.g.* a chain-length category).
- The likelihood of common precursors and/or breakdown products, via physical or biological processes, which result in structurally similar chemicals (*e.g.* the metabolic pathway approach of examining related chemicals such as acid/ester/salt).

Read-Across

69. Read-across is a method of predicting unknown toxicity of a chemical using similar chemicals (called chemical analogues) with known toxicity from the same chemical category (Dimitrov and Mekenyan 2010).

70. In the OECD guideline on *Grouping of Chemicals: Chemical Categories and Read-Across* it states that: In the read-across approach, endpoint information for one chemical (the source chemical) is used to predict the same endpoint for another chemical (the target chemical), which is considered to be "similar" in some way (usually on the basis of structural similarity or on the basis of the same mode or mechanisms of action). In principle, read-across can be used to assess physicochemical properties, toxicity, environmental fate and ecotoxicity. For any of these endpoints, it may be performed in a qualitative or quantitative manner.

71. Qualitative read-across is similar to the use of a SAR, and the process involves:

- the identification of a chemical substructure or mode or mechanism of action that is common to two substances (which are considered to be analogues).
- the assumption that the presence (or absence) of a property/activity for a substance can be inferred from the presence (or absence) of the same property/activity for the analogous substance.

The main application of qualitative read-across is in hazard identification.

72. Quantitative read-across involves:

- the identification of a chemical substructure or mode or mechanism of action that is common to two substances (which are considered to be analogues).

- the assumption that the known value of a property for one substance can be used to estimate the unknown value of the same property for another substance.

In both cases, expert judgement is needed and some justification should be provided.

Trend Analysis

73. Trend analysis is a method of predicting toxicity of a chemical by analysing toxicity trends (increasing, decreasing, or constant) of tested chemicals. An example of trend analysis shows that when carbon chain length (CCL) increases, acute aquatic toxicity increases (Figure 2) (Jeliazkova *et al*., 2010)

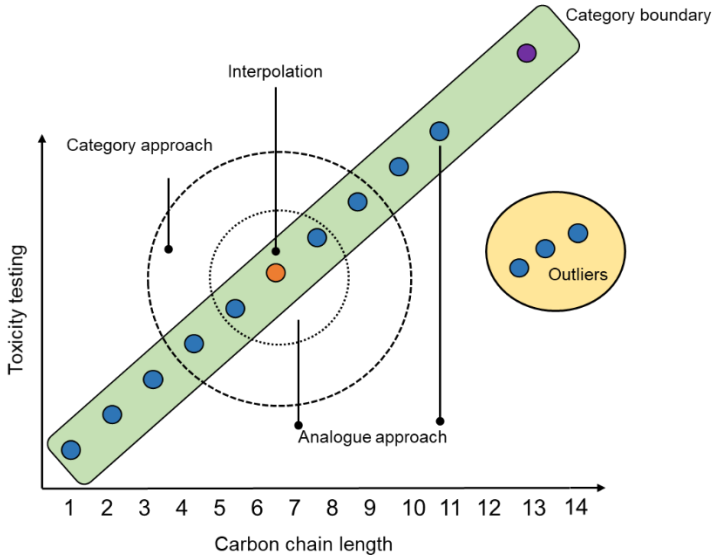


Figure 2. Different approaches of read-across: analogue versus category approaches, interpolation versus extrapolation, category boundary and outliers. (Figure adopted from Raies and Bajic 2016).

74. Trend analysis and read across are developed in similar ways.

75. There are two ways to develop a read-across method (Venkatapathy *et al*., 2013, Vink *et al*., 2010), analogue approach (AN) (called one-to one), which uses one or few analogues, and a category approach (CA) (called many-to-one), which uses many analogues. AN may be sensitive to outliers because two analogues may have different toxicity profiles (Venkatapathy *et al*., 2013).

76. Using many analogues for CA is useful to detect trends within a category and may increase confidence in the toxicity predictions (Venkatapathy *et al.,* 2013., Modi *et al.,* 2012). CA requires defining a category boundary to determine if a chemical belongs to the category (Venkatapathy *et al*., 2013) and implementing a 'combination of predictions' method for analogues 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 (Dimitrov *et al.*, 2010).

77. Read-across can be qualitative if the toxicity endpoint is qualitative; otherwise, read-across is quantitative (Valerio *et al*., 2009; Benigni *et al*., 2013) Also, interpolation using source chemicals surrounding the target chemical (see Figure 2) is better than extrapolation from one side (Worth *et al*., 2013). 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.

78. Identifying similar chemicals can be done in two steps: representing chemicals as feature vectors of chemical properties, and then calculating similarity of chemicals.

79. The first step is implemented using either binary or holographic fingerprints.

80. Molecular fingerprints are widely used in several areas of chemoinformatics including diversity analysis and similarity searching (Fernández de Gortari *et al*., 2017).

81. 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) (Civjan, 2012).

82. A holographic fingerprint uses frequency of properties (*e.g.* number of methyl groups) (Igaki *et al*., 1992).

83. Continuous chemical properties (*e.g*. melting point) can be used as well.

84. 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.

85. 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.

86. An example of hierarchal categories is "*Hierarchical cluster analysis of environmental pollutants through P450 induction in cultured hepatic cells: Indications for a toxicity screening test"* (Dubois *et al*., 1996).

87. Statistical similarity of two chemicals can be calculated using different types of distances, such as Hamming[50], Euclidean[51], Cosine[52], Mahalanobis[53], Tanimoto (or Jaccard) distance[54] coefficient, or linear or nonlinear relationships of the features.

88. Read across has several advantages and disadvantages. Read-across is transparent (Cronin *et al*., 2011) easy to interpret and implement (Enoch *et al*., 2010). Read-across can model quantitative and qualitative toxicity endpoints, and it allows for a wide range of types of similar measurements to be used to express similarity between chemicals.

89. However, statistical similarity measurements do not provide biological insight of toxicity (Dimitrov *et al*., 2010). Moreover, complex similarity measures may complicate model interpretation. In reality, read-across uses small datasets compared to other approaches such as QSAR because there are usually only a few analogues for a given chemical. Additionally, accuracy depends on the number and choice of analogues, similarity metrics, strength in chemicals' similarity, chemical properties, and category boundaries (Dimitrov *et al*., 2010).

90. These parameters are very subjective, mutually dependent, endpoint-specific, and may require expert opinions. Moreover, this approach could be inapplicable or inaccurate if analogues have conflicting toxicity profiles or the number of analogue chemicals is insufficient. In such cases, the QSAR approach can be used[55] (Modi *et al*., 2012).

Dose-Response and Time-Response

91. 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). 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' (El-Masri, 2013).

92. The time element can be the time to produce a response or the time for recovery (Bliss and Stevens, 1937). Exposure time can be continuous, intermittent, or random, and exposure can be acute, short-term, sub-chronic, or chronic (El-Masri, 2013).

93. Time-dose models describe the relationship between time and dose for a constant response (Brown and Foureman, 2005). Figure 3 shows different types of dose/time-response models.

[50]Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different.
[51] Euclidean distance or Euclidean metric is the "ordinary" straight-line distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space.
[52] Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.
[53] Mahalanobis distance is a measure of the distance between a point P and a distribution D, introduced by P. C. Mahalanobis in 1936. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D.
[54]Jaccard similarity index (sometimes called the Jaccard similarity coefficient) compares members for two sets to see which members are shared and which are distinct. It's a measure of similarity for the two sets of data, with a range from 0% to 100%. The higher the percentage, the more similar the two populations.
[55] https://www.oecd.org/publications/guidance-on-grouping-of-chemicals-second-edition-9789264274679-en.htm

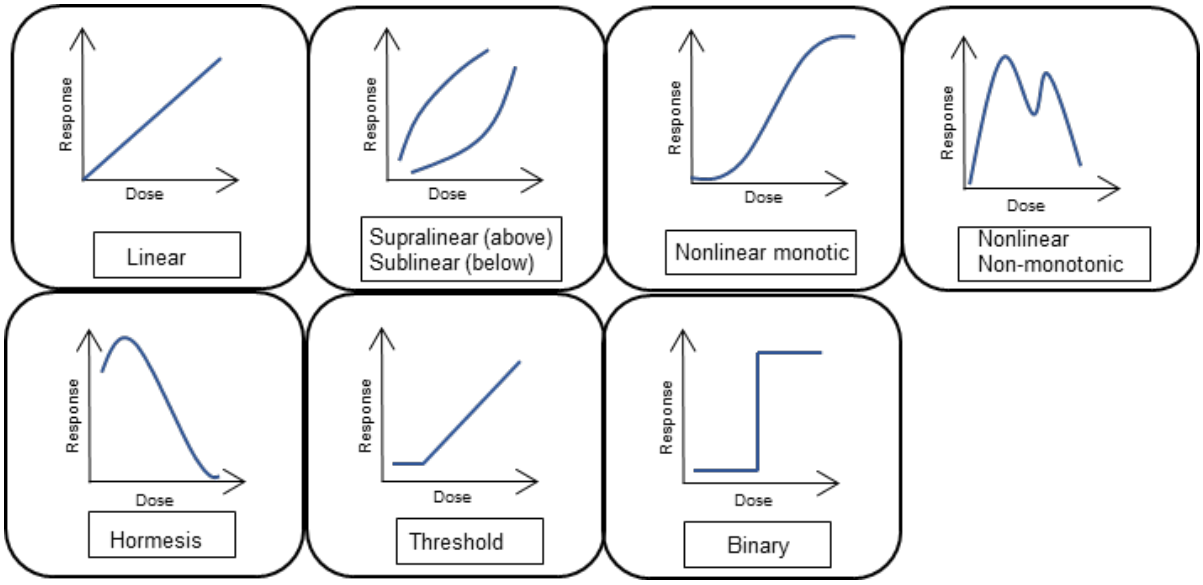


Figure 3. Different types of relationships for dose-response models. (Figure adopted from Rais and Bajaic 2016)

94. The first dose-response model relates to concentration (C) and time (t) with response (K), which is:

Haber's law (law of toxicity[56]) (El-Masri, H., 2013. Miller *et al*., 2000):

$C \times t = K$

95. On the other hand, Haber's law does not hold in many situations and 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.

96. 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.

97. Several well-known generalizations of Haber's law are shown below:

- Ostwald: $(C - C_0)^n t = K$ that emphasizes concentration (Bliss,1940)
- Druckery: $C \times t^n = K$ that emphasizes time (Miller *et al*., 2000)
- Miller *et al.,* (2000) : $(C - C_0)^n t^m = K$ that emphasizes both concentration and time

98. One of the frequently measured responses is mortality (the number of deceased individuals).

[56] Haber's rule or Haber's law is a mathematical statement of the relationship between the concentration of a poisonous gas and how long the gas must be breathed to produce death, or other toxic effect. The rule was formulated by German chemist Fritz Haber in the early 1900s.

99. The Bliss method (Bliss, (1935)) (or Probit model) (Figure 4) 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 units) and express dose or time on a logarithmic scale. Probits are inferred doses (or time) that correspond to a given mortality percentage.

100. Bliss devised a special table called 'probits table' to calculate the probits (Bliss, 1935).

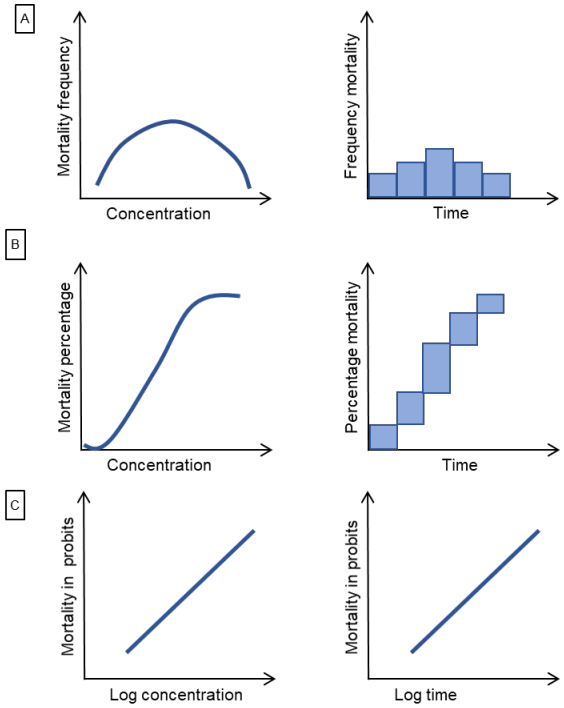


Figure 4. Bliss method (A) Plot mortality frequency (the number of dead subjects) versus dose or time (B) Convert frequency to percentages (percentage of deceased subjects) (C) Transform percentages to probits and transform dose or time to logarithms. (Figure adopted from Rais and Bajaic, 2016).

101. 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 (Bliss, 1935; 1937).

102. 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 (Bliss, 1937).

103. 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 (Bliss, 1937).

104. Miller *et al*. (2000) 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.

105. Brown and Foureman (2005) used a time-concentration-response model to generalize the concentration-response models using time as a parameter.

106. 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 (Bliss, 1937; Miller *et al.,* 2000) using dose-response models, and interpolation between different exposure times for the same toxicant and dose within the range of experimental data (Bliss, 1937; Miller *et al.,* 2000) using time-response models.

107. There are some limitations to the models, such as, they cannot extrapolate to other chemicals (El Masri, 2013). In addition, time-response models cannot extrapolate to other doses of the same chemical. Time-response models require that tested individuals have uniform susceptibility levels, 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. The three models do not take into consideration target tissue, biological process, ADME, toxicokinetics, toxicodynamics, detoxification, damage or repair, or chemical properties.

108. 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:

- Chemical Effects in Biological Systems (CEBS)[57] is an integrated public repository for toxicogenomics data, including the study design and timeline, clinical chemistry and histopathology findings and microarray and proteomics data. CEBS contains data derived from studies of chemicals and of genetic alterations, and is compatible with clinical and environmental studies.

- PubChem[58] is a database of chemical molecules and their activities against biological assays. The system is maintained by the National Center for

[57] https://manticore.niehs.nih.gov/cebssearch
[58] https://pubchem.ncbi.nlm.nih.gov/

Biotechnology Information, a component of the National Library of Medicine, which is part of the US NIH.

- ToxRefDB[59] provides detailed chemical toxicity data in a publicly accessible searchable format. As an example, ToxRefDB contains mammal toxicity information that when combined with other sources of information, such as exposure and metabolism, form the basis for pesticide risk assessments.

109. Some usage examples of these models include modelling rectal cancer (Appelt, 2015), mutagenicity (Pottinger *et al*., 2009) and developmental toxicity (Hunt *et al*., 2008).

Pharmacokinetic Models and Pharmacodynamic Models

110. Pharmacokinetic[60] (PK) models relate chemical concentration in tissues to time, estimate the amount of chemicals in different parts of the body, and quantify ADME processes (Jack *et al*., 2012, Sung *et al*., 2014).

111. Toxicokinetic (TK) 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 (El Masri, 2013, Sung *et al*., 2014).

112. A compartment is the whole or part of an organism in which the concentration is uniform (Ingrisch and Sourbron 2013). Compartmental models consist of one or more compartments, and each compartment is usually represented by differential equations (Sung *et al*., 2014).

113. In PK, compartmental models are in widespread use for describing the concentration-time curves of a drug concentration following administration. This gives a description of how long it remains available in the body, and is a guide to defining dosing regimens, method of delivery, and expectations for its effects (Bassingthwaighte *et al*., 2012).

114. 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:

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

where C_0 is the initial concentration and k is the elimination constant. A plot of log of concentration versus time results in a straight line of slope $-k$ (Sung *et al*., 2014).

115. However, these models do not consider the distribution time of chemicals. Additionally, concentrations in some organs reach equilibrium faster than in others.

[59] https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NCCT&dirEntryId=227139
[60] Pharmacokinetics: Process of the uptake of drugs by the body, the biotransformation they undergo, the distribution of the drugs and their metabolites in the tissues, and the elimination of the drugs and their metabolites from the body over a period of time.

116. 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:

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

117. These models, however, cannot extrapolate between species or provide a mechanistic insight (Sung *et al*., 2014).

118. 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, metabolism and/or excretion (Jack *et al*., 2013; El Masri 2013).

119. PBPK models represent each organ as a compartment, represented by a differential equation that includes PK parameters (Jack *et al*., 2013, El Masri, 2013, Sung *et al*., 2014, Mager *et al*., 2014).

120. 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 (Sung *et al*., 2014). A general PBPK model to calculate plasma concentration (C_P) uses a feature vector of PK parameters (θ_{PK}), time (t), and dose (X) as follows:

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

where f is a function that models the relationship.

121. Due to the equation structure and because the physiological parameters are tissue specific, PBPK models allow for interspecies extrapolation and provide a mechanistic basis of ADME (Modi *et al*., 2012, El Masri 2013, Sung *et al*., 2014).

122. PBPK models can convert administered doses to tissue dosimetry, which is ‘the amount of chemical that is distributed to a tissue or part of a tissue,’ (Jack *et al*., 2013) and generate concentration versus time models (Modi *et al*., 2012).

123. Pharmacodynamic (PD) models relate a biological response to the concentration of chemical in the tissue (Sung *et al*., 2014). 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 (Andersen *et al*., 2001).

124. 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 (Sung *et al*., 2014). Similar to PBPK models, PBPD can be described by differential equations.

125. 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*) (Mager *et al*., 2014) can be represented as:

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

where *f* is a function that models the relationship. PD models can be combined with PK models (Sung *et al*., 2014). The resulting model is called biologically based dose-response models (BBDR) and can be used to relate doses with responses. Jack *et al*., 2013; El Masri 2013; Sung *et al*., 2014; Andersen *et al*., 2001).

126. In addition to PK and PD parameters, BBDR may include biological parameters such as cell division rates, mortality rates, or production rates of hormones. 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 (Crump *et al*., 2010).

127. There are many advantages for these models. Determining internal doses rather than administered doses and key metabolites allows for a more direct relationship with the response (El Masri, 2013). 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 (El Masri, 2013). 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 (Crump *et al*., 2010).

128. 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 (Sung *et al*., 2014). Otherwise, QSAR modelling could be more appropriate because it depends only on molecular descriptors (Modi *et al*., 2012; El Masri, 2013). Additionally, if biological data is not available, empirical dose-response models are used instead of BBDR.

129. Using BBDR for extrapolation between species assumes that the relationship between dose and response in animals is the same in humans (Crump *et al*., 2010; Haber *et al*., 2001).

130. The same problem applies when using animal studies to estimate PK or PD parameters for modelling toxicity in humans (Egorov *et al*., 2013).

131. Although BBDR models were 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 modelled 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 Crump *et al* (2010).

132. Expert knowledge is required for defining MoA, toxicity pathways and chemical interactions that cause the response.

133. Different types of PK and PD models are reviewed in Mager *et al.* (2003). Also, methods for estimating PK parameters are reviewed in Avent *et al*. (2013).

134. Examples of PK and PD modelling tools are WinNonlin[61],[62] Kinetica[63] and ADAPT 5[64]. Some examples include route-to-route extrapolation (Chiu and White, 2006), toxicity and risk assessment (Andersen, 1995) and carcinogenicity assessment (Clewell *et al.,* 2007).

COT consideration of PBPK modelling

135. In July 2019, a discussion paper "*Review of physiologically-based pharmacokinetic (PBPK) modelling used for human health risk assessment*" (TOX/2019/3) was presented to COT.

136. The COT recognised that PBPK modelling could be used to verify the appropriateness of test concentrations used for *in vitro* assays through their comparison with estimates of human internal exposure. Furthermore, it was considered that the values generated by high-throughput and *in silico* methods for some model parameters (*e.g.* partition coefficients and transporter activity) can be associated with varying degrees of uncertainty.

137. The COT said it was necessary to assess how realistic and reliable these parameter values are. In addition, the Committee considered that further guidance on the use and application of PBPK models developed for nanomaterials would be helpful.

138. A deficiency of human pharmacokinetic data was often noted for those xenobiotics for which PBPK models are developed and assessed by the Committee. This was central to the discussion held in 2003 when PBPK modelling was last brought to the Committee.

139. Approaches that were considered to assess model reliability in this context included use of the read-across approach and conducting interspecies extrapolations to animal species other than humans. Thus, it was noted that in-house expertise in the field of PBPK modelling will be needed increasingly in the future for the interpretation of these models.

140. The Committee agreed it would be useful to have further information in the form of case studies, for example where *in vitro* data have been successfully extrapolated to *in vivo*, or cases where risk assessments considered in retrospect

[61] Phoenix® WinNonlin® is the industry standard for non-compartmental analysis (NCA), pharmacokinetic/pharmacodynamic (PK/PD), and toxicokinetic (TK) modeling.
[62] https://www.certara.com/software/pkpd-modeling-and-simulation-2/phoenix-winnonlin-2/?ap%5B0%5D=PKPD&ap%5B1%5D=PKPD
[63] http://tools.thermofisher.com/content/sfs/brochures/KIN410.pdf
[64] https://bmsr.usc.edu/software/adapt/

may have benefitted from PBPK modelling. It was also noted a workshop on PBPK would be beneficial since the last one hosted by the COT was in 2003.

141. Henceforth, a paper on PBPK case studies (TOX/2019/73) is also presented at this meeting and a joint workshop with potency estimation models is planned for next year.

Uncertainty Factor Models Uncertainty factors

142. Uncertainty factors (UFs) (also called assessment/extrapolation/risk factors) are used in the assessment of risk from chemical exposure or the recommended daily intake of chemicals (Falk-Filipsson *et al*., 2007).

143. UFs are used to compensate for a deficiency in knowledge concerning the accuracy of test results and the difficulty in estimating the health effects in a different species and/or in different exposure conditions. UFs date back to 1954 when Lehman and Fitzhugh (1954) proposed a 100-fold factor (which they referred to as a "margin of safety") for extrapolating from animal toxicity data to safe levels of human exposure to food additives and pesticide residues.

144. An 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 foetuses), or exposure duration extrapolation (*e.g.* from short exposure to long exposure). It requires two main factors (Martin *et al*., 2013): no observed adverse effect levels (NOAEL), which is the highest dose not exhibiting observable toxicity and an 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 the NOAEL by the UF.

145. However, there are two limitations for using the NOAEL approach (Martin *et al*., 2013): (1) the definition of NOAEL indicates the absence of the 'appreciable risk' of toxicity, but it does not indicate a zero-effect threshold; (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 a higher NOAEL.

146. On the other hand, it is possible to use a least observable adverse effect level (LOAEL, which is the lowest 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 a NOAEL is not available (Falk-Filipsson *et al*., 2007; Martin *et al*., 2013).

147. 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.

148. Although, existing UFs account for intra-species variability, the use of additional factors for child safety is recommended.

149. The values of MFs and SFs cannot exceed 10 (Falk-Filipsson *et al*., 2007).

150. 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) MoA (El Masri, 2013").

151. The reference values are calculated as:

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

where POD is the point of departure (*e.g.,* NOAEL, LOAEL, or BMDL) (El Masri, 2013). A default UF of 100 was first proposed in 1954 (Falk-Filipsson *et al*., 2007).

152. 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.

153. Therefore, several factors have been calculated by different agencies as explained in Falk-Filipsson *et al* (2007) and Martin *et al* (2013).

154. There are several advantages of UF models such as it's easy to implement and understand them and they provide adequate safety levels for a single chemical and mixtures of chemicals (Martin *et al*., 2013). Additionally, they account for inter-species and inter-individual as well as PK and PD differences.

155. 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 (Martin *et al*., 2013).

156. 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 (Falk-Filipsson *et al*., 2007).

Quantitative Structure–Activity Relationship

157. Structure-activity relationships represent a core aspect of medicinal chemistry. The fact that a small change in structure (usually) leads to a small change in biological activity, allows chemists to rationalize substitutions at specific positions, giving them the freedom to modify a molecule to improve various properties such as lipophilicity, bioavailability without sacrificing potency (to a large extent) (Guha, 2012).

158. Quantitative structure-activity relationship (QSAR) is a family of models that uses molecular descriptors to predict chemicals' toxicity.

159. The models relate a set of "predictor" variables (X) to the potency of the response variable (Y), while classification QSAR models relate the predictor variables to a categorical value of the response variable.

160. It is assumed that chemicals that fit the same QSAR model may work through the same mechanism (Toropov *et al*., 2014). 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 θP is:

$$T = f(\theta_P)$$

161. A local QSAR is generated from congeneric chemicals (*i.e.* similar chemicals); a global QSAR is made from diverse chemicals.

162. 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.

163. Local QSARs can also provide insight on the MoA of specific chemicals, which global QSARs may overlook.

164. Quantitative Structure Toxicity/Property Relationship (QSTR/QSPR) models are QSAR models that predict toxicity and chemical properties, respectively (Toropov *et al*., 2014).

165. Structure activity relationships (SARs) are used for categorical endpoints (Devillers, 2013).

166. Several types of molecular descriptors can be used to describe chemicals:

- 1D QSAR: 1D descriptors represent the structure of the chemicals such as atoms and functional groups (Devillers, 2013).

- 2D QSAR: 2D descriptors represent the physico-chemical, physico-biological properties (Kortagere and Ekins 2010) and topological indices (Devillers, 2013).

- 3D QSAR: 3D descriptors represent filed properties in 3D such as energy fields: steric, electrostatic and hydrophobic (Devillers, 2013).

- 3D QSAR: CoMSIA (Comparative Molecular Similarity Indices Analysis) (Crammer *et al*., 1988).

- 4D QSAR: CoMFA (Comparative Molecular Field Analysis) (Devillers, 2013).

- COREPA (Common Reactivity Pattern) Conformational distribution of chemical across local and global reactivity parameters that are linked to the biological activity (Devillers, 2013).

- Pseudo 3D: Eigen values derived from IR and Raman range molecular vibrational frequencies. Weighted holistic invariant molecular (WHIM) descriptors (Devillers, 2013).

- QSIIR (Quantitative structure *in vitro in vivo* relationship)
  Chemical and biological descriptors (from high throughput screening *in vitro* data) (Zhu, H., 2013).

167. There are two main steps to develop a QSAR model: generating molecular descriptors and then generating models to fit the data.

168. Therefore, feature selection algorithms based on, for example, simulated annealing, genetic algorithm, or principal component analysis can be used (Deeb, O. and Goodarzi, 2012; Devillers, 2013). 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 (Devillers, 2013) (Figure 5).

169. 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 (Deeb and Goodarzi, 2015) (Devillers, 2013); nonlinear models such as artificial neural networks or support vector machines (Deeb and Goodarzi, 2015; Devillers, 2013); and data-driven models such as those based on decision trees, clustering, Naïve Bayes, and K-nearest neighbour (Zhu, H., 2013).

170. 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 as illustrated in Figure 5.

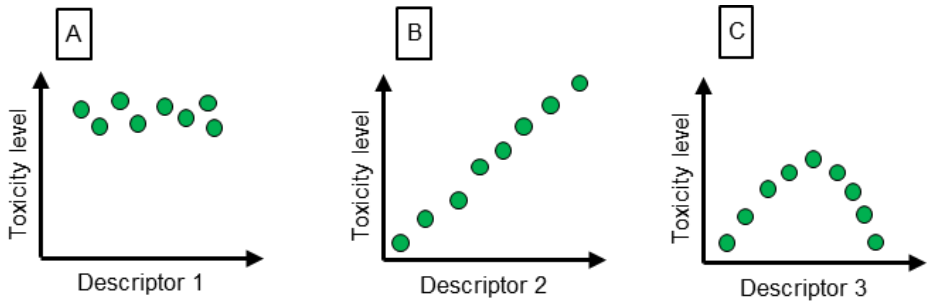


Figure 5. 2D scatter plots of molecular descriptors and toxicity levels. (a) no correlation between molecular descriptor 1 and the toxicity endpoint. (b) and (c) linear and nonlinear relationships between the molecular descriptors 2 and 3,

respectively, with the toxicity endpoint. (b) and (c) can be modelled with linear and nonlinear algorithms, respectively. (Figure adopted from Rais and Baijic, 2016).

171. 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 6 shows a hypothetical example of a SAR landscape. The smooth region 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).

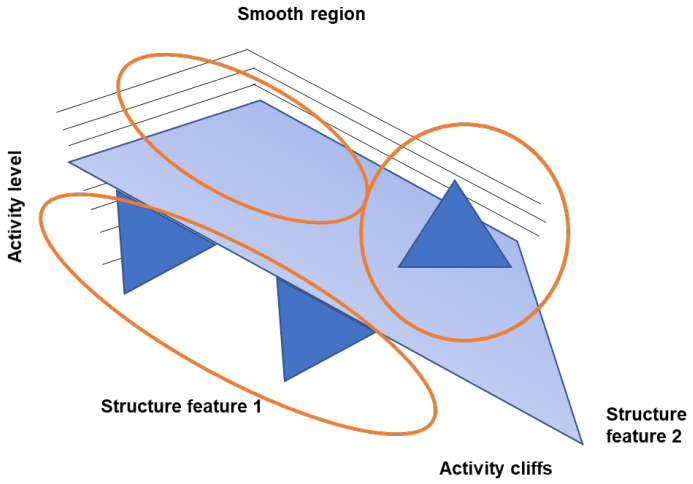


Figure 6. SAR landscapes (Figure adopted from Rais and Baijic, 2016)

172. The activity cliffs are the most interesting part of the SAR landscape (Guha, 2013). They encode structural relationships in which small chemical modifications lead to large potency variations. Accordingly, if activity cliffs are encountered during compound optimization in drug discovery-reveal substitution sites and chemical changes that determine structure-activity relationships (SARs) in compound series (Bajorath, 2017).

173. 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:

Scaffold hops: Low structural similarity and high activity similarity
Smooth regions: High structural similarity and high activity similarity
Nondescript: Low structural similarity and low activity similarity
Activity cliffs: High structural similarity and low activity similarity

174. Moreover, a structure activity landscape index (SALI) and a structure activity relationship index (SARI) can be used to analyse SAR landscapes.

175. Gramatica (2012) elucidated on the development and validation of QSAR Models.

176. Examples/case studies of QSARs:

- Dearden (2003) describes the *in silico* prediction of drug toxicity.

- Ellison *et al.* (2011) wrote a review related to the use of *in silico* methods to predict the chemistry of molecular initiating events related to drug toxicity.

- Netzeva *et al.* (2008) paper reviews the quantitative structure activity relationships for acute aquatic toxicity, as well as different methods described in the literature for calculating the aquatic toxicity of chemical substances.

- Pasha *et al.* (2009) discussed the *in silico* quantitative structure toxicity relationship of aromatic nitro compounds.

- Carlsen *et al.* (2008) studied the impact on environmental health by residuals of the rocket fuel 1,1-dimethyl hydrazine (heptyl) and its transformation products.

- Thakur and Thakur (2009) developed a QSTR model based on cytotoxic concentration for the set of 19 (tetrahydromidazo [4,5,1-jk][1, 4] benzodizepin-2(1H)-one) (TIBO) derivatives.

- Cronin and Madden (2010) on skin sensitization.

177. There are many tools that provide pre-built QSAR models such as: OECD QSAR Toolbox[65], TopKat[66] and METEOR[67]. Case studies on combining the results of different prediction tools are available in Milan *et al.* (2011) and Worth *et al.* (2013).

178. However, specialized software tools for generating QSAR models such as ADAPT and TOPKAT include databases for toxicity data and can calculate molecular descriptors (Venkatapathy, 2013).

Quantitative structure activity relationship (QSAR) in toxicology (QSTR)

179. QSTRs are based on the hypothesis that the structure of a molecule must enclose the features responsible for its physical, chemical and biological properties, and on the capability to characterize the chemical by one, or more, numerical descriptor(s) or properties. The toxicity of substances is governed by their properties, which in turn are determined by their chemical structure. As a result, there are interrelationships between structure, properties, and toxicity. QSTR is a statistically

[65] https://www.oecd.org/chemicalsafety/oecd-qsar-toolbox.htm
[66] https://www.3dsbiovia.com/products/datasheets/ds_topkat.pdf
[67] https://www.lhasalimited.org/products/meteor-nexus.htm

resulting equation that quantitatively describes a molecular property in terms of descriptors of compound structure (Deeb and Goodarzi, 2015).

180. QSTR can be divided into four crucial steps:

1. conversion of structures into descriptors (parameters).
2. descriptor selection to minimize the risk of chance correlations.
3. deriving the relationship between the molecular descriptors and the toxicological data.
4. validating the QSTR model and assessing its predictivity.

181. The QSTR models are developed using a variety of chemometric tools such as multiple linear regression (MLR)[68], partial least squares (PLS)[69], artificial neural network (ANN)[70], support vector machine (SVM)[71] and others.

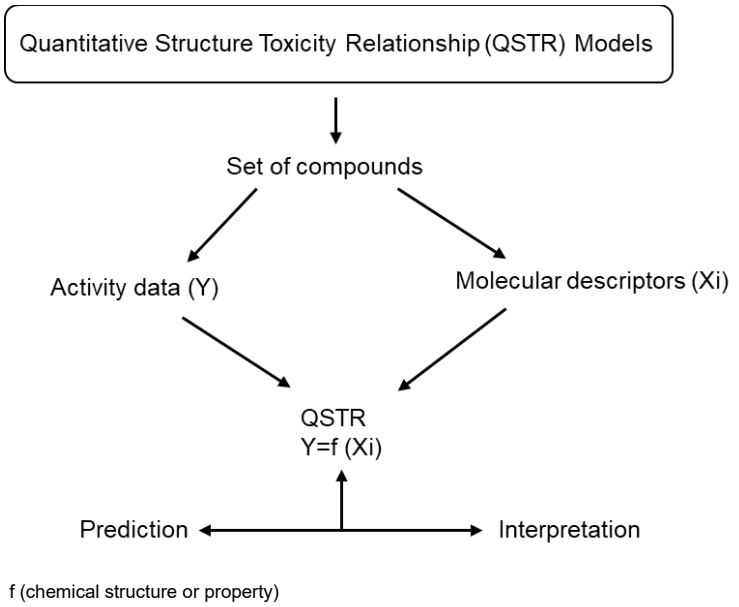


Figure 7. Scheme summarizing these QSTR steps (Figure adapted from Deeb and Goodarzi, 2015).

182. The advantages of QSAR models include: They are easy to interpret if the descriptors are meaningful. They can model categorical and continuous toxicity endpoints, molecular descriptors and toxic and non-toxic chemicals. Using different types of descriptors allows for modelling complex endpoints (Modi *et al*., 2012).

[68] Multiple regression is the statistical procedure to predict the values of a response (dependent) variable from a collection of predictor (independent) variable values. For example, if scores on multiple predictors and one criterion are available, multiple regression may be used to develop a single equation to predict criterion performance from the set of predictors.
[69] Partial least squares regression (PLS regression) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of maximum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space.
[70] Artificial neural networks (ANN) or connectionist systems are computing systems that are inspired by, but not identical to, biological neural networks that constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with task-specific rules.
[71] In machine learning, support-vector machines (SVMs, also support-vector networks are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis.

183. However, QSARs may not always be applicable. QSARs require a large number of chemicals in model development to achieve statistical significance. 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 (Weaver and Gleeson, 2008).

184. The disadvantages include: 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.

QSAR Applications for Chemical Screening, Prioritization, and Regulatory and Corporate Decision-making

185. QSAR predictions are used by regulatory authorities, private corporations, and institutions in three major contexts: priority-setting, hazard classification and labelling, and screening for health and ecological risks of chemicals (Pittinger and Mohapatra, 2009.)

186. Regulatory uses of QSARS include:

(1) supporting priority setting of chemicals

(2) guiding experimental design of regulatory tests or testing strategies

(3) providing mechanistic information

(4) grouping of chemicals into categories based on similarity

(5) filling a data gap needed for classification and labelling

(6) filling a data gap needed for risk assessment.

187. Each application carries unique considerations for QSAR, with the most stringent considerations placed upon QSARs used for 'high regulatory impact,' for example, risk assessments under mandated regulatory programmes.

Special types of substances or toxicity endpoints that require new prediction or analysis methods

Nanoparticles

188. Nanotechnology[72] is a known field of research since last century. Since "nanotechnology" was presented by Nobel laureate Richard P. Feynman during his famous 1959 lecture "*There's Plenty of Room at the Bottom*"[73] (Feynman, 1960), there have been made various revolutionary developments in the field of nanotechnology (Khan *et al*., 2019).

[72] Nanotechnology is manipulation of matter on an atomic, molecular, and supramolecular scale.
[73] https://en.wikipedia.org/wiki/There%27s_Plenty_of_Room_at_the_Bottom

189. Nanotechnology is one of the key emerging technologies identified in the EU 2020 Strategy[74]. The heterogeneity of engineered nanoparticles (NPs), with respect to physicochemical properties and observed (eco)toxicological effects, makes their case-by-case testing for risk assessment unsustainable in terms of costs, time, and number of test animals (Basei *et al*., 2019).

190. The large number of NPs and the high complexity associated with their interactions in biological and environmental systems (Mirshafiee, Osborne, Sun and Xia, 2018) have raised the call for Amendments of the REACH Annexes (EC, 2017[75]) to require additional information for the safety assessment of NPs.

191. As stated in the introduction, to address these challenges, it has been widely agreed by regulators, industries and scientists that the way forward is to develop robust IATA that should be compliant with the 3R principles for reducing animal testing (Burden *et al*., 2017). These IATA would involve both experimental and modelling tools (Hristozov *et al*., 2016) to facilitate intelligent testing, grouping and read-across to inform both regulatory risk assessment and safer design of quality products.

192. The application of *in silico* methods for grouping and read-across is subject to REACH Annex XI (Rudén and Hansson 2009). A group or category, according to REACH, is the arrangement of substances based on similar physicochemical and (eco)toxicological, toxicokinetic and/or environmental fate properties (ECHA, 2017a[76]).

193. OECD goes beyond identification of toxicological properties to support Grouping and identifies also mode of toxicological action as a relevant principle of similarity (OECD, 2014[77]).

194. Other possible ways to group NPs are based on commercial importance and volume of production, composition/chemistry (*e.g*. carbon-based; metal and metalloid oxides; metals, metal salts and metalloids; semiconductor quantum dots; organics and other classes) (RCC, 2013[78,79]), on properties (such as dimension, shape, morphology, complexity and surface functionalization), or based on synthetic and biological identity (Lynch *et al*., 2014).

195. Grouping can serve several purposes such as informing targeted testing for Risk Assessment, facilitating Safe-by-Design[80] practices, and filling data gaps (Mirshafiee, Osborne, Sun and Xia, 2018).

196. The physicochemical characteristics such as size (Osborne *et al*., 2015; Zhu *et al*., 2012), surface area (Monteiller *et al*., 2007), surface reactivity (Duffin *et al*., 2007), surface functionality (Pelaz *et al*., 2015), charge, aspect ratio (Lin *et al*, 2014),

[74] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/future-and-emerging-technologies
[75] EC, 2017. Amendments of the Annexes to REACH for Registration of Nanomaterials. European Commission.
[76] https://op.europa.eu/en/publication-detail/-/publication/492b96dc-5bde-11e7-954d-01aa75ed71a1/language-en
[77] https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm
[78] https://www.oekopol.de/wp-content/uploads/2014/06/Work-Element-2.pdf
[79] http://science.gc.ca/eic/site/063.nsf/vwapj/nano3_e.pdf/$file/nano3_e.pdf
[80] Safe by design is the concept of applying methods to minimize occupational hazards early in the design process, with an emphasis on optimizing employee health and safety throughout the life cycle of materials and processes.

protein corona[81] (Gebauer *et al*., 2012) and hydrophobicity (Li *et al*., 2008) could significantly affect material interactions with biological systems at the nanoscale (Nel *et al*., 2009). The latter is typically achieved through Read-Across, an established approach used to predict properties and/or effects for a "target" substance by using information from analogous "source" substance(s) (ECHA, 2017[82]).

197. Furthermore, Grouping for Read-Across is not only accepted under REACH, but is also applicable under different chemicals regulatory frameworks (Mech *et al*., 2019): for instance, for risk assessment of NPs in the food and feed chain, the EFSA supports grouping for read-across from (other) NPs or non-NPs (Hardy *et al*., 2018).

198. As an example, the H2020 GRACIOUS project[83] is currently building upon these developments to generate a highly innovative science-based framework to enable practical application of Grouping, leading to Read-Across and classification of NPs.

199. For classical chemical compounds, a number of descriptors (directly derived or computed starting from physicochemical properties are available and can be modelled, however the greater number of these descriptors are not applicable or not useful with respect to nanomaterials (NMs) (Puzyn *et al*., 2011; Winkler *et al*., 2013; Oksel *et* al., 2015).

200. Indeed, recently ECHA provided a guidance on how to identify and characterize NMs and nanoforms (NFs) (ECHA, 2017[84]): in addition to "substance identity" parameters specified in Section 2 of Annex VI of REACH regulation (European Commission, 2006[85]), which include composition of the substance, of impurities or additives, as well as information of surface chemistry and crystalline structure, other parameters that are relevant to characterize NPs are physical parameters such as size, shape and surface area. Other relevant parameters include solubility, hydrophobicity, zeta potential, dispersibility, dustiness, as well as biological (re)activity and photoreactivity.

201. Examples include:

- Grouping of nano-titanium dioxide (TiO_2) to read-across genotoxicity according to the ECHA guidance including application of chemoinformatic approaches, by using measured physicochemical descriptors and nanospecific descriptors (mainly related to size distribution, zeta potential, and dispersibility) (Lamon *et al*., 2018).

- Classification NanoSAR Development for Cytotoxicity of Metal Oxide Nanoparticles (Liu *et al*, 2011).

- Development of SAR for metal oxide nanoparticles (Liu *et al*., 2013).

[81] Protein corona consists of proteins adsorbed from physiological fluids on NPs forming
[82] https://op.europa.eu/en/publication-detail/-/publication/841c5a3a-2981-11e7-ab65-01aa75ed71a1/language-en
[83] https://www.h2020gracious.eu/about
[84] https://op.europa.eu/en/publication-detail/-/publication/492b96dc-5bde-11e7-954d-01aa75ed71a1/language-en
[85] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02006R1907-20140410

202. Like other chemical compounds, to be able to predict the adverse (eco)toxicological effects of NPs by means of *in silico* tools, it is indeed fundamental to have access to high quality (meta)data[86]. Curation of the data is essential. Powers *et al.* (2015) proposed a data curation workflow: (i) assessing the quality and completeness of the selected data, (ii) extracting and annotating the data, (iii) contacting authors for any missing data, (iv) formatting the data for inclusion into the databases, (v) reviewing the data, (vi) releasing the curated data to target communities, and (vii) updating the curated data as new information is provided by the authors.

203. Examples of nano repositories of data:

- eNanoMapper[87]: Contains primary research data from various nano-EHS projects and from literature.

- OCHEM[88]: Contains experimental data on nano and non-nanomaterials. Allows building and validating computational models from the available data using ML techniques.

- NanoDatabank[89]: Includes data on NMs toxicity, characterization, fate and transport.

- Nanowerk[90]: Contains physicochemical information and details on manufacturers of 4000 NMs from more than 150 suppliers worldwide.

- SUN[91]: Includes physicochemical, release, exposure, *in vitro* and *in vivo* toxicological data.

Nano-QSAR

204. Nano-QSARs are QSAR models that use NPs-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 (Winkler *et al.*, 2013; Epa *et al*., 2012, Gajewicz *et al*., 2014).

205. Examples of studies include:

- QSAR model for the prediction of the cellular uptake of NMs in pancreatic cancer cells using SMILES descriptors (Melagraki and Afantitis, 2014).

[86] Metadata is "data that provides information about other data". In short, it's data about data.
[87] http://www.enanomapper.net/
[88] https://ochem.eu/home/show.do
[89] http://nanoinfo.org/nanodatabank/
[90] https://www.nanowerk.com/
[91] http://sun.iom-world.co.uk/

- Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles (Puzyn *et al*., 2011).

- Nano-QSAR modelling for ecosafe design of heterogeneous TiO_2-based nano-photocatalysts (Mikolajczyk *et al*., 2018).

206. PBPK modelling of nanoparticles is an emerging field, 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 (Li *et al*., 2012).

207. Some examples include a study which used PBPK modelling for dietary risk assessment of TiO_2 NMs (Bachler *et al*., 2015) and a PBPK model for ionic silver and silver NMs (Bachler *et al*., 2013).

Mixtures

208. Toxicity of chemicals is affected by interactions with other chemicals (Hamelink *et al*., 1994). Mixtures may exhibit adverse effects at NOAEL doses of each chemical separately. Assessing toxicity of chemicals separately may underestimate or overlook the adverse effects of mixtures.

209. Therefore, 'cumulative risk assessment' was developed to study toxicity of mixtures (Løkke *et al*., 2013). However, there is lack of experimental datasets for toxicity of mixtures due to a large number of different combinations of chemicals (Kar and Leszczynski, 2019) such as exposure patterns, and complex interactions. It is challenging to test all combinations of these factors.

210. Furthermore, predictive models must address concurrent and sequential exposure to mixtures. A recently developed database by NoMiracle (Novel Methods for Integrated Risk Assessment of Cumulative Stressors in Europe)[92] contains mixtures' toxicity datasets for eco-toxicological test species and human cell lines (Løkke *et al*., 2010).

211. Methods for single chemicals may not be applicable for mixtures due to difficulty in determining the combined effect (Sarigiannis and Hansen, 2012). For example, dose-response models for mixtures vary depending on the dose ratios of chemicals in the mixture (Løkke *et al*., 2013).

212. Additionally, co-administration of chemicals may alter their ADME properties, which should be taken into consideration when developing PBPK models for mixtures (Sarigiannis and Hansen, 2012).

213. Unlike single chemicals, mixtures have no or very few experimental datasets for toxicity.

[92] https://ec.europa.eu/jrc/en/scientific-tool/novel-methods-integrated-risk-assessment-cumulative-stressors

214. The reasons mentioned below make database preparation work difficult and multifaceted (Kar and Leszczynski 2019):

a) toxicity data vary with different combinations of the same chemicals in a mixture

b) form of exposure

c) identification of each chemical in a specific mixture is also difficult due to the presence of very small quantities

d) complex interactions among chemicals

215. The assessment of a mixture's toxicity is much more complex than toxicity evaluation of a single chemical. Interactions of chemicals in a mixture can be the reason for complex and significant changes in the apparent properties of its components.

216. Bliss (1939) classified the joint action of mixtures into distinctive categories:

- Joint action: Chemicals act independently and have different modes of action. The combined effect is calculated using the effects of constituents and their interactions.

- 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.

- Concentration addition (Ca): If chemicals in a mixture showed same mechanism of action for a specific response and act on same site of action, then there are chances of dilution of the response.

- Independent action (IA): If chemicals in mixtures act on different sites of action with dissimilar MoA, this may disclose statistically independent responses without interaction.

217. If chemicals are interactive in nature, then they may show synergistic or antagonistic effects. Toxicity of synergistic action is greater than that of the constituent chemicals, while antagonistic action has lower toxicity than that of the 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.

218. 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) (Hardup *et al*., 2013) Therefore, a generalized concentration addition (GCA) model was developed to address these shortcomings (Howard and Webster 2009). GCA

calculates the combined effect of a mixture using the potency and efficacy of the mixture's constituents (Hardup *et al*., 2013).

219. 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 (Baas *et al*., 2010).

220. In ecotoxicology[93] there is a growing interest in effects of mixtures as in the environment there is a plethora of substances creating a cocktail effect (Relyea, 2009).

221. When the chemical composition of a mixture is known and constant, effects of mixtures can be described in a bottom-up approach, starting from the effects of the individual compounds that make up the mixture.

222. An example of a process-based model is a study that performed simulations with a known mixture of 14 individual polycyclic aromatic hydrocarbons (PAHs) and predicted the no-effect concentration (NEC) for fathead minnow[94] (*P. promelas)* of 490 nM (Baas *et al*., 2010).

223. 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) (Løkke *et al*., 2010). Examples include air contaminants (Astel, 2010) as well as wildlife and human exposure modelling (Loos *et al*., 2010).

224. 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; chemical interaction models such as the interaction-based hazard index and isobole method; and statistical models such as tree-based clustering and weighted quartile score regression (Sarigiannis and Hansen, 2012).

225. EuroMix[95] has been aiming to establish novel testing and assessment strategies for chemical mixtures found in humans, as well as the relevant tests and models to go alongside them[96].

226. Recently artificial intelligence (AI) has been used in mixture models. Cipullo *et al*. (2019) employed two machine learning (ML) models, including random forest (RF) and artificial neural networks (NN) to predict temporal bioavailability followed by toxicity prediction employing predicted bioavailability features as the input of complex chemical mixtures.

[93] Ecotoxicology is the study of the effects of toxic chemicals on biological organisms, especially at the population, community, ecosystem, and biosphere levels. Ecotoxicology is a multidisciplinary field, which integrates toxicology and ecology.
[94] Fathead minnow (*Pimephales promelas*) is a species of temperate freshwater fish belonging to the genus *Pimephales* of the cyprinid family.
[95] https://www.euromixproject.eu/
[96] https://library.wur.nl/WebQuery/wurpubs/fulltext/409001

Summary of in silico models

227. As described above, *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

(3) customize these methods for each problem if necessary.

228. In Table 1 a summary of *in silico* model elucidating approaches, advantages / disadvantages, limitations as well as existing software or databases. In Figure 8 a flow chart provides a practical guideline for choosing a method for certain types of features and toxicity endpoints.

Table 1 Table summary of in silico model elucidating approaches, advantages/disadvantages, limitations as well as existing software or databases (Table 1 taken from Raies and Bajic (2016)

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 effect at a given concentration	• One-compartment models • Two-compartment models • PBPK, PBPD and BBDR models	1) PK models determine internal doses rather than administered doses. 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.	1) PK and PD parameters may be unavailable or inaccurate. 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.	• WinNonlin • Kinetica • ADAPT
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

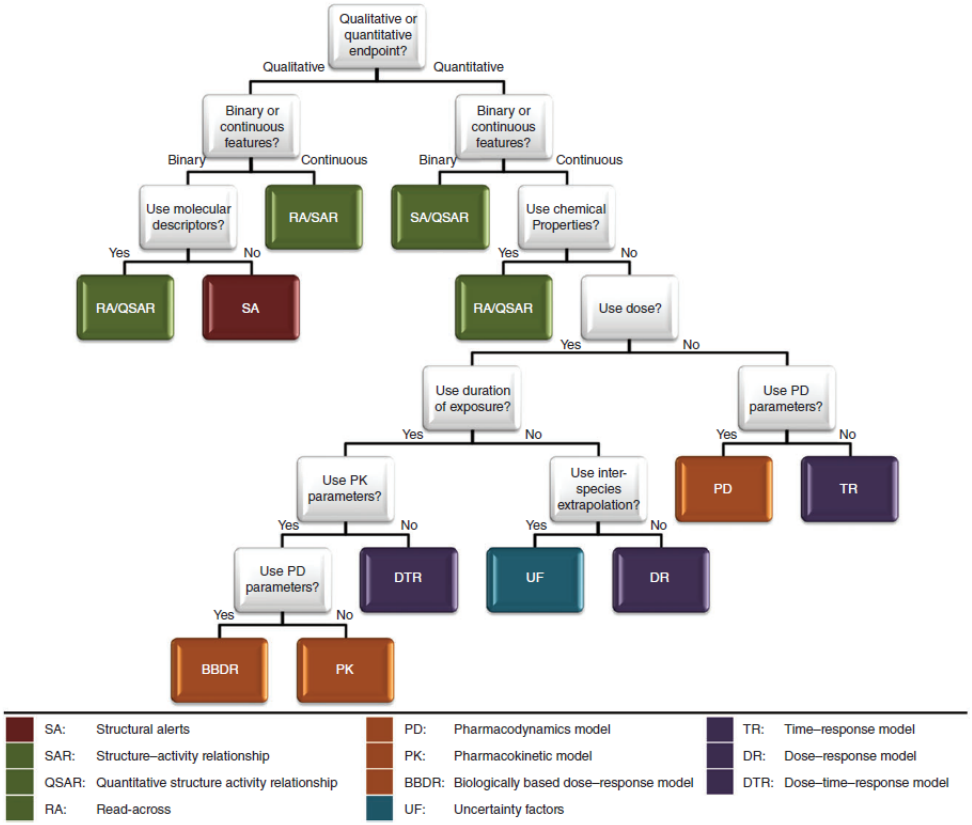


Figure 8. Overview of *in silico* methods. Flow chart practical guideline for choosing a method for certain types of features and toxicity endpoints. Figure taken from Raies and Bajic (2016).

Adverse Outcome Pathways

229. Adverse outcome pathways (AOPs) are novel tools in toxicology and human risk assessment with broad potential. AOPs span multiple levels of biological organisation (Delrue *et al*., 2016). AOPs are designed to provide a clear-cut mechanistic representation of critical toxicological effects that span over different layers of biological organization. AOPs share a common structure consisting of a molecular initiating event, a series of intermediate steps and key events, and an adverse outcome.

230. Development of AOPs ideally complies with OECD guidelines. In general, AOP development includes 3 consecutive steps, namely the identification of the main information blocks, the data summation and the evaluation (OECD, 2012[97]).

231. This also holds true for AOP evaluation, which includes consideration of the Bradford Hill criteria for weight of evidence (WoE) assessment and meeting a set of key questions defined by the OECD.

232. Elaborate AOP frameworks have been proposed for chemical-induced skin sensitization, cholestasis, liver fibrosis and liver steatosis. These newly postulated AOPs can serve a number of ubiquitous purposes, including the establishment of

[97] http://www.oecd.org/chemicalsafety/testing/49963554.pdf

(quantitative) SARs, the development of novel *in vitro* toxicity screening tests and the elaboration of prioritization strategies.

233. AOPs are a conceptual framework that portrays existing knowledge concerning the linkage between some molecular initiating event (MIE) and an adverse outcome (AO) (Groh *et al*., 2015) that occurs at a level of biological organization considered relevant to regulatory decision-making (Ankley *et al*., 2010).

234. Individual AOPs are represented as sequences of measurable key event (KE) nodes that reflect a causal progression from an initial perturbation of normal biology, caused through direct interaction with a chemical, to a series of system failures at higher levels of biological organization (Groh *et al*., 2015).

235. KEs are linked via Key Event Relationships (KERs) that define both the structural and functional relationship between a given pair of KEs and compile specific empirical evidence that supports the idea that if the upstream KE is altered to a sufficient degree, predictable changes (qualitative or quantitative) can be expected in the downstream event in the sequence. AOPs are described using modular assemblies of KE and KER descriptions.

236. These modular descriptions, properly structured and connected in the AOP knowledge base (KB)[98], provide the foundation for construction and analysis of AOP networks that can provide a more comprehensive, integrated, and biologically realistic synthesis of available knowledge concerning the ways chemicals can adversely impact organisms.

237. Overall, AOPs and AOP networks provide structure for our knowledge of how a molecular initiating event (MIE) (or MIEs) can lead to deviations from normal healthy function of a biological system (*i.e.* adverse outcome (s)) (Wittwehr *et al*., 2017).

238. From a modelling perspective, structuring of knowledge is extremely informative for model design and development. In particular, AOPs can help reduce an initially overwhelmingly complex biology to the essentials necessary for a predictive model, avoiding model overload.

239. A summary of an AOP represented in the form of a box and arrow diagram that identifies the KEs and KERs (Figure 9) provides an overall conceptual model that bounds the modelling challenge within a specific biological domain. It defines, for example, the key chemical biological interaction that triggers a toxicologically relevant biological perturbation by identifying the MIE.

[98] https://aopkb.oecd.org/index.html

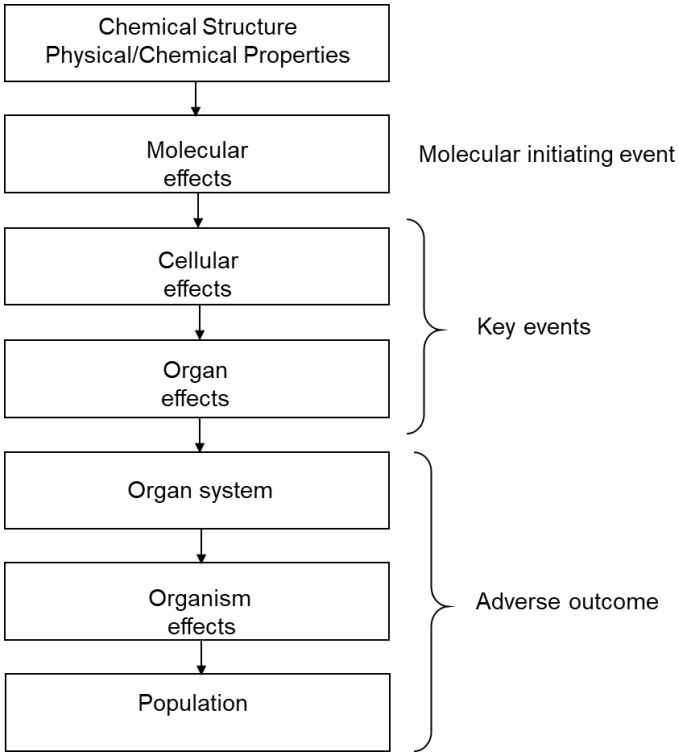


Figure 9. Flow diagram depicting an adverse outcome pathway representing common chemicals triggering molecular initiating events leading to a sequential series of higher order effects to produce an adverse outcome.

240. This can immediately inform the development of QSAR models and chemical categories useful for defining the chemical space for which the AOP is likely to have relevance. It then identifies the key biological pathways, functions and compartments (*i.e*. cell types, tissues, organs) in which the biology to be modelled operates. Each KE is defined at a particular level of biological organization, the AOP also provides a road map of the biological scales at which a single model, or series of models, must operate. In this way, the AOP suggests the heuristic[99] domain and biological scope/space in which the prediction models should function. A second level of information in the AOP description that guides the development of prediction models is the description of the KEs (OECD, 2016[100]). In many cases, those with the computational modelling expertise may not be familiar with the biology represented in the AOP.

241. The biological description of the KEs presented in the AOP-WIKI[101] module provides the entry point or gateway that can introduce the modeler to the biology encompassed by the AOP and its events. Whereas this description may not be sufficient to fully support model design and formulation, it can suggest the appropriate biological subject matter experts with which the modeler may want to consult and partner.

242. Additionally, it may be helpful in identifying the type of modelling approaches, mathematical, formalisms, and parameters that could be employed. For example, a

[99] Heuristic: enabling a person to discover or learn something for themselves.
[100] https://one.oecd.org/document/ENV/JM/MONO(2016)12/en/pdf
[101] https://aopwiki.org/

KE involving enzyme inhibition may require identification of the type of inhibition (*e.g.* competitive, irreversible) and related kinetic constants. A KE involving cell proliferation or selective cell death may indicate the need to employ an agent-based modelling approach, whereas one involving an increased risk of disease may require a probabilistic approach. KE descriptions associated with an AOP also contain useful information regarding how a KE is measured (Wittwehr *et al*., 2017)

243. This information helps define the kinds of data likely to be available, or which could be generated to inform model development. Identification of specific assays may, in some cases, provide useful information regarding data sources that the modeler(s) could utilize for model development purposes. For example, for KEs measured in ToxCast assays (Kavlock *et al*., 2012), association of the KE with an assay identifies a database[102] of relevant data that may be useful for model development. Depending on the experimental method(s) used, additional information might be required to translate the raw output of the method to *in vivo* relevant data.

244. Future modules of the AOP-KB (*i.e*. Effectopedia[103]) aim to provide standardized summaries of the data itself along with meta-information describing the test methods and transformation functions need to put those data into appropriate *in vivo* context. The identification of specific approaches used to measure a given KE can suggest the types of data that may serve as inputs to the model, and parameters that may be useful to simulate from an interpretive standpoint. For example, if the AOP involves enzyme inhibition that leads to a decrease in a circulating hormone followed by a loss of function in a particular cell type, one might want to design a model that can take a standardized measure of a chemical's potency to inhibit the enzyme and predict the dose-response and time-course behaviour of the circulating hormone concentration, subject to feedback regulation and other modulating factors represented in the model (Wittwehr *et al*., 2017).

245. Finally, identification of the methods used to measure the KEs can provide insights into the time-scales over which the variables represented as KEs can be measured. This provides information regarding the level of temporal resolution that the models should be designed to predict.

246. KER descriptions (OECD, 2016[104]) are similarly useful. The KER description gives a summary of the WoE that establishes the causal nature of the relationship between 2 measurable biological events (Becker *et al*., 2015). The structure of the KER immediately defines key input and output parameters relevant for model simulation. Defining the biological plausibility of the relationship between the pair of KEs highlights the important biological context and the processes that need to be captured in the relationship model. Furthermore, the empirical evidence summarized in the KER description, provides references that can provide data for model parameterization, fitting, and/or testing.

[102] https://comptox.epa.gov/dashboard
[103] https://www.effectopedia.org/
[104] https://www.oecd-ilibrary.org/environment/users-handbook-supplement-to-the-guidance-document-for-developing-and-assessing-adverse-outcome-pathways_5jlv1m9d1g32-en;jsessionid=ULi86IAy_Zk5_LEWGy_edG8a.ip-10-240-5-4

Case Study 1: Skin sensitisation

247. An AOP which involves a MIE of covalent modification of cellular proteins in the skin by electrophilic chemicals, which results in an AO of sensitization of the skin to allergens[105].

248. Intermediate KEs in the AOP capture processes related to the induction of inflammatory cytokines by dendritic cells and keratinocytes, and activation and proliferation of T-cells that ultimately cause sensitization (Kimber *et a*l., 2012).

249. The scientific support for this AOP is considered to be strong (Patlewicz *et al*., 2015; Perkins *et al*., 2015) so it also offers an excellent basis for development of quantitative models relating the MIE to the AO (Maxwell *et al*., 2014). In particular, there is a strong regulatory and scientific interest in applying mechanistic understanding captured in the AOP to help reduce and replace the need for animal testing associated with the hazard characterization and risk assessment of skin sensitizing chemicals for use in cosmetic and other consumer care products (*e.g.* soaps, lotions).

250. The skin sensitization AOP has enabled a clearer dialogue with regulatory authorities and risk assessors on the mechanistic relevance of each of the *in vitro* approaches either when applied in isolation or when these datasets are combined using integrated testing strategies (ITS)/data integration procedures (DIP).

251. Twelve skin sensitization DIPs have been identified and discussed as case studies within the OECD “Skin Sensitization IATA guidance working group” (including Bauch *et al*., 2012; Hirota *et al*., 2015; MacKay *et al*., 2013).

Case Study 2: Activation of oestrogen receptor-a leading to diverse adverse outcome

252. For some applications, quantitative modelling may only need to capture the MIE and/or very early KEs of an AOP. An example of this is illustrated by ongoing activities through the US EPA’s EDSP, the objective of which is to identify chemicals with potential to cause adverse effects through alteration of pathways associated with hypothalamic-pituitary-thyroidal (HPT) and hypothalamic-pituitary-gonadal (HPG) function (US EPA, 2014[106]).

253. One MIE of concern is activation of the oestrogen receptor-alpha (ERα). Oestrogenic chemicals have been associated with a large number of different AOs involving reproduction and development in vertebrates (WHO/IPCS, 2002[107]).

254. The development of a network model is to predict the potential for chemicals to act as oestrogens *in vivo* based on a chemical’s ability to elicit responses in high throughput *in vitro* assays that capture multiple aspects of the MIE, including binding

[105] https://aopwiki.org/wiki/index.php/Aop:40
[106] https://www.epa.gov/sites/production/files/2015-08/documents/edsp_comprehesive_management_plan_021414_f.pdf
[107] https://www.who.int/ipcs/publications/en/toc.pdf?ua=1

to ERα, receptor dimerization, chromatin binding, transcriptional activation, and ER dependent cell proliferation (Browne *et al*., 2015; Judson *et al*., 2015).

255. The quantitative model provides potency values for test chemicals relative to E2 (the ERα endogenous ligand) and was evaluated/validated by comparing model output to results from the uterotrophic assay, an *in vivo* pathway-based system considered to be a "gold standard" for identifying ERα agonists (Browne *et al*., 2015; Kleinstreuer *et al*., 2015).

256. Whereas this particular quantitative model only reflects early portions of AOPs relevant to interaction with the ERα, it nonetheless has substantial utility for addressing one of the challenges faced by EDSP. Specifically, it is being utilized by the US EPA to prioritize 10,000-plus chemicals for more resource intensive *in vivo* testing necessary to assess potential risks, based on their predicted estrogenic potency.

257. A recent "proof of concept" study conducted through the EDSP indicates that the quantitative model predictions, in conjunction with a rapid exposure assessment, provide a reasonable basis for test chemical prioritization based on agreement between the *in vitro*-based predictions and *in vivo* results available for a reference set of estrogenic compounds (US EPA, 2014[94]).

258. In this hazard-based scenario, the AOP provides a toxicological "anchor" for election/use of the high-throughput assays in the context of application of the computational model to hazard assessment.

Case example 3: Evaluating pesticide toxicity to pollinators

259. Key pollinator species, such as honeybees, have experienced significant worldwide declines, resulting in concerns for possible effects on global food production. In the USA, for example, a national strategy has been developed to assess the significance and causes of pollinator declines (White House, 2015)[108].

260. A number of chemical and nonchemical stressors have been proposed as contributing to declines, one of the more prominent of which are neonicotinoid pesticides (Godfray *et al*., 2015; EPA, 2016[109]).

261. However, significant uncertainties exist as to the biological plausibility of a link between the MIE of neonicotinoids-activation of the nicotinic acetylcholine-receptor and impacts on honeybee colonies.

262. To help assess the veracity of hypothesized effects of neonicotinoids on honeybees, LaLone *et al*. (2017) assembled an AOP network based on molecular, biochemical, physiological, behavioural, and population data from more than 220 papers in the open literature. This demonstrated a plausible linkage between perturbation of nicotinic acetylcholine receptor signalling and adverse effects in

[108] https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/Pollinator%20Health%20Strategy%202015.pdf
[109] https://19january2017snapshot.epa.gov/newsreleases/epa-releases-first-four-preliminary-risk-assessments-insecticides-potentially-harmful_.html

honeybees, but the analysis highlighted areas of uncertainty that would benefit from focussed research and/or monitoring (LaLone *et al.*, 2017).

263. In this example, the AOP framework supported integration of a complex, biologically-diverse dataset, in the context of evaluating causal relationships among endpoints at different levels of organization and served as a basis for generating hypotheses to test these interactions.

Toxicity tiered testing

264. Toxicity tiered testing is a set of biologically based toxicity testing decision triggers, developed and analysed within a tiered testing and decision-making framework for evaluating potential human health hazards and risks associated with chemical exposures.

265. The proposed three-tiered toxicity testing approach (Figure 10) starts from a base set of toxicity studies (acute toxicity, *in vitro* genetic toxicity, *in vitro* cytogenetics, repeat dose/sub-chronic toxicity, developmental toxicity, reproductive toxicity) and then uses the toxicity triggers to identify which specific additional tests are needed to adequately characterize a substance's hazard potential (Becker *et al*., 2007).

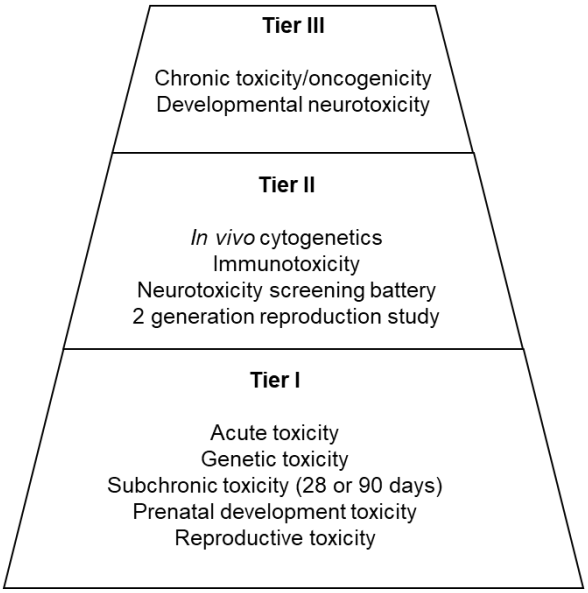


Figure 10. Example of tiered testing grouped into three tiers.

266. These tiered and sequential testing approaches have been used to obtain the desired hazard information in an organized, efficient, and readily interpretable manner.

267. The proposed tiered testing strategy begins with a battery of toxicity tests in Tier I, which is comprised of the OECD-Screening Information Data Set (SIDS)[110] base set.

[110] A screening information dataset (SIDS) is a study of the hazards associated with a particular chemical substance or group of related substances, prepared under the auspices of the Organisation for Economic Co-operation and Development (OECD).

268. After arraying the results of this base set, the specific toxicity data are evaluated using the toxicity triggers to determine which specific higher tiered tests would be warranted, due to the presence of an observed toxicity in accordance with the toxicity triggers and their associated decision criteria (Botham, 2004).

269. The Tier I battery consists of the standard OECD-SIDS endpoints for human health evaluation for high production volume (HPV) chemicals (Federal Register, 2000[111]). These Tier I tests provide both general toxicological information and specific information relevant to possible mode of action (MoA) that can be used to assess potential effects on human health and to determine when more extensive testing is warranted (Becker *et al*., 2007).

270. Specific toxicity triggers were developed to be used in evaluating the results of the specific tests in Tier I and guiding decisions as to which (if any) specific additional toxicity tests (Tier II or III tests) may be needed in order to reduce uncertainties about the potential hazards of a given chemical (Becker *et al*., 2007).

271. These toxicity-based decision triggers were developed based on collective extensive experience with toxicity testing and evaluation. It is intended that the toxicity decision criteria be applied along a designated pathway. In a given pathway, toxicity endpoints are linked together in such a way that specific findings from a test, such as a Tier I test for prenatal developmental toxicity, are used to trigger more in-depth investigation of relevant endpoints in a more complex and definitive test, such as a Tier II 2-generation reproduction study or a Tier III developmental neurotoxicity test (Becker *et al*., 2007).

272. Examples and case studies:

- A Tiered Approach to Systemic Toxicity Testing for Agricultural Chemical Safety Assessment (Doe *et al*., 2006).

  Summary: A proposal has been developed by the Agricultural Chemical Safety Assessment (ACSA) Technical Committee of the ILSI Health and Environmental Sciences Institute (HESI) for an improved approach to assessing the safety of crop protection chemicals. The goal is to ensure that studies are scientifically appropriate and necessary, and that tests emphasize toxicological endpoints and exposure durations that are relevant for risk assessment. The ACSA Systemic Toxicity Task Force proposes an approach to systemic toxicity testing as one part of the overall assessment of a compound's potential to cause adverse effects on health. The approach is designed to provide more relevant data for deriving reference doses for shorter time periods of human exposure. This includes fewer studies for deriving longer term reference doses-that is, neither a 12-month dog study nor a mouse carcinogenicity study is recommended. All available data, including

The substances studied are high production volume (HPV) chemicals, which are manufactured or imported in quantities of more than 1000 tonnes per year for any single OECD market.

[111] https://www.federalregister.gov/documents/2000/12/26/00-32498/data-collection-and-development-on-high-production-volume-hpv-chemicals

TK and metabolism data and life stages information, are taken into account. The proposed tiered testing approach has the potential to provide new risk assessment information for shorter human exposure durations, while reducing the number of animals used and without compromising the sensitivity of the determination of longer-term reference doses.

- Case Study: Incorporating new approach methodologies in toxicity testing and exposure assessment for tiered risk assessment using the RISK21 approach: Case studies on food contact chemicals (Turtley *et al*., 2019).

  Summary: Two indirect food additive chemicals where ToxCast data were compared with *in vivo* toxicity data using the RISK21 approach. Two food contact substances, sodium (2-pyridylthio)-N-oxide and dibutyltin dichloride, were selected, and available exposure data, toxicity data, and model predictions were compiled and assessed. Oral equivalent doses for the ToxCast bioactivity data were determined by *in-vitro in-vivo* extrapolation (IVIVE). For sodium (2-pyridylthio)-Noxide, bioactive concentrations in ToxCast assays corresponded to LOAELs and NOAELs in animal studies. For dibutyltin dichloride, the ToxCast bioactive concentrations were below the dose range that demonstrated toxicity in animals; however, this was confounded by the lack of toxicokinetic data, necessitating the use of conservative toxicokinetic parameter estimates for IVIVE calculations. This study highlights the potential utility of the RISK21 approach for interpretation of the ToxCast HTS data, as well as the challenges involved in integrating *in vitro* HTS data into safety assessments.

- Case Study: Use of a Pro-Fibrogenic Mechanisms-Based Predictive Toxicological Approach for Tiered Testing and Decision Analysis of Carbonaceous Nanomaterials (Wang *et al*., 2015).

  Summary: Engineered carbonaceous nanomaterials (ECNs), including single-wall carbon nanotubes (SWCNTs), multiwall carbon nanotubes (MWCNTs), graphene, and graphene oxide (GO), are potentially hazardous to the lung. With incremental experience in the use of predictive toxicological approaches, seeking to relate ECN physicochemical properties to adverse outcome pathways (AOPs), it is logical to explore the existence of a common AOP that allows comparative analysis of broad ECN categories. We established an ECN library comprising three different types of SWCNTs, graphene, and graphene oxide (two sizes) for comparative analysis according to a cell-based AOP that also plays a role in the pathogenesis of pulmonary fibrosis. SWCNTs synthesized by Hipco, arc discharge and Co-Mo catalyst (CoMoCAT) methods were obtained in their as-prepared (AP) state, following which they were further purified (Pd) or coated with Pluronic F108 (PF108) or bovine serum albumin (BSA) to improve dispersal and colloidal stability. GO was prepared as two sizes, GO-small (S) and GO-large (L), while the graphene samples were coated with BSA and PF108 to enable dispersion in

aqueous solution. *In vitro* screening showed that AP- and Pd-SWCNTs, irrespective of the method of synthesis, as well as graphene (BSA) and GO (S and L) could trigger interleukin-1β (IL-1β) and transforming growth factor-β1 (TGF-β1) production in myeloid (THP-1) and epithelial (BEAS-2B) cell lines, respectively. Oropharyngeal aspiration in mice confirmed that AP-Hipco tubes, graphene (BSA-dispersed), GO-S and GO-L could induce IL-1β and TGF-β1 production in the lung in parallel with lung fibrosis. Notably, GO-L was the most pro-fibrogenic material based on rapid kinetics of pulmonary injury. In contrast, PF108-dispersed SWCNTs and graphene failed to exert fibrogenic effects. Collectively, these data indicate that the dispersal state and surface reactivity of ECNs play key roles in triggering a pro-fibrogenic AOP, which could prove helpful for hazard ranking and a proposed tiered testing approach for large ECN categories.

Data

273. *In silico* modelling requires access to data on the investigated endpoint(s) that are sufficient both in terms of quantity and quality.

274. Toxicological data and information are crucial to *in silico* safety assessment not only in terms of availability, but also their abundance and quality. The current status is an increasing number of data relating to the adverse effects of chemicals which range from the results of high content analyses to historical toxicity data across a number of publicly and commercially available databases.

275. There is also a wealth of (potentially high quality) toxicological data in the archives of business, and pharmaceutical companies in particular, which would offer great opportunities if exploited (Cronin *et al*., 2019). This has brought forward the concept of data sharing to enable access to otherwise untapped resources.

276. One key factor identified from the literature is the requirement for detailed recording of the data, model and supporting documentation to enable the validity of the model and its applicability for a given purpose to be ascertained. A checklist-style reporting format has previously been developed by the European Commission's Joint Research Centre (JRC, Ispra), known as the (Q)SAR Model Reporting Format (QMRF)[112]. This provides a template for recording key information about QSAR models and associated validation studies. The format was designed with adherence to the OECD Principles in mind.

Data Quality

277. The inherent quality of the data upon which a model is built is arguably the most important characteristic of any model. Data quality here refers to the accuracy and completeness of the information on the chemicals studied as well as the adequacy and reliability of the experimental data (Young *et al*., 2008; Beck and Geppert, 2014).

[112] https://publications.jrc.ec.europa.eu/repository/bitstream/JRC107491/kjna28713enn.pdf

Data/Model Validation

278. Unlike *in vitro* alternatives, which have a distinct protocol for validation by organisations such as the European Centre for the Validation of Alternative Methods (ECVAM)[113] *in silico* models have not yet been verified in such a formalised manner. Given the degree of diversity seen in available *in silico* models (model architecture, statistical analyses used, dataset size and composition, etc.) developing a universal approach is difficult (Hewitt *et al*., 2015).

279. Another significant factor that impedes greater acceptance of models is not that the model itself lacks validity but the level of detail by which the model is recorded is insufficient to allow judgement of model quality; this again means the model cannot be used with confidence.

280. Increased acceptance and uptake of *in silico* modelling approaches will only be possible where confidence in the applicability and usefulness of a model to provide a given prediction can be assured. In a recent scientific report on modern methodologies and tools for human hazard assessment of chemicals, EFSA highlighted the need for validation of predictive models as an important step in their utilisation for chemical risk assessment[114].

281. It has been stated that some factors should be considered when assessing the validity of an *in silico* model, appropriate recording of model details and a pragmatic scheme that can be applied for model verification (Hewitt *et al*., 2015).

282. Several factors were considered to be of key importance in developing the assessment scheme to be used for model verification (Figures 11 & 12) (Hewitt *et al*., 2015):

(i) Carrying out an assessment of a model had to be a realistic task both in terms of the required expertise of the individual and the time needed to conduct such an assessment.

(ii) The assessment criteria had to be presented in a format which would be compatible with a wide range of operating systems and software.

(iii) The verification process had to be transparent, scientifically justifiable and the results readily accessible to end users.

(iv) The introduction of a peer review *i.e.* the model builders themselves prepare and submit all required documentation and supporting data, such that an external verification can be readily carried out. The role of the model verifier is then to check the submission for accuracy, completeness and reproducibility.

[113] https://eurl-ecvam.jrc.ec.europa.eu/
[114] https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2014.3638

(v) model documentation *i.e.* a model's identity and its developers, the endpoint investigated, training/test set data (including source), model algorithm/ summary statistics, external predictivity, mechanistic information, applicability domain and interpretation of prediction.

(vi) Data used to build the model and consistency of model output *i.e.* the nature and quality of data used to build a model is a major determining factor in model acceptability. It was therefore considered prudent to include an assessment of the dataset(s) used to develop a model as part of the verification process.

(vii) Implementation of the model *i.e.* the final component of the assessment criteria relates to how a model is implemented. This includes: model stability, robustness to input files, consistency of output.

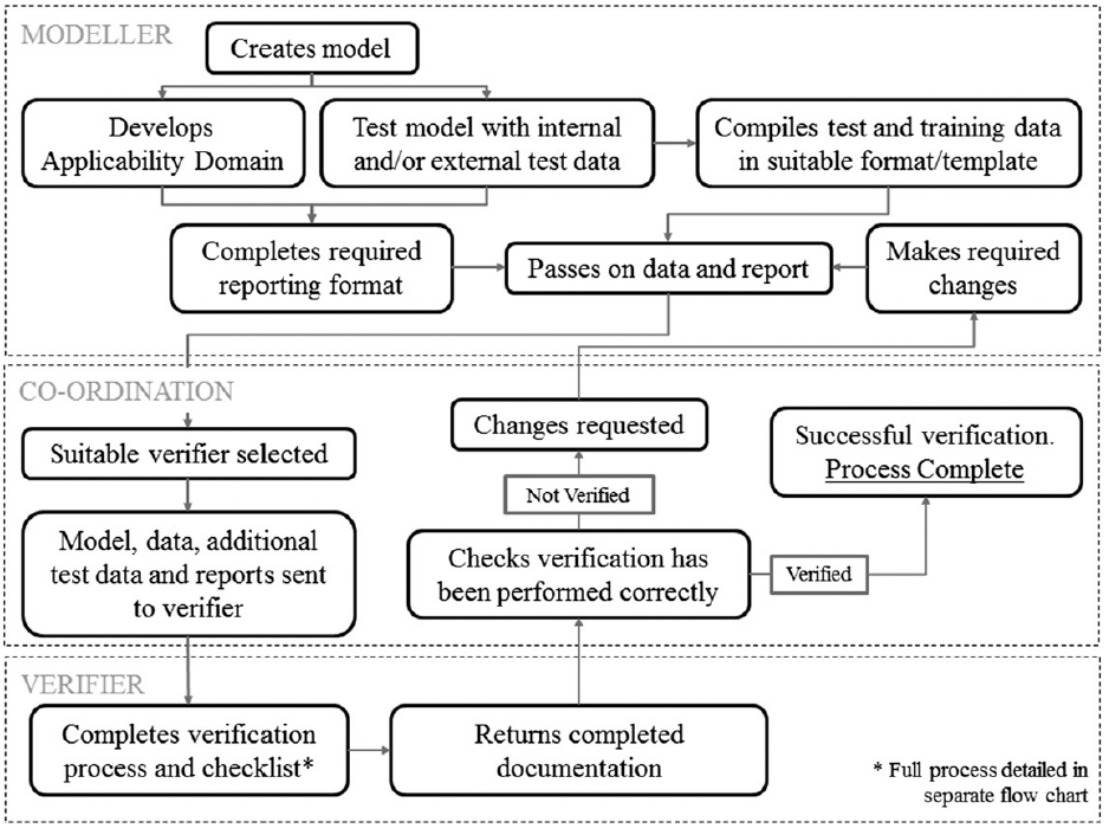


Figure 11. Overall process of model development and verification (Figure taken from Hewitt *et al.*, 2015):

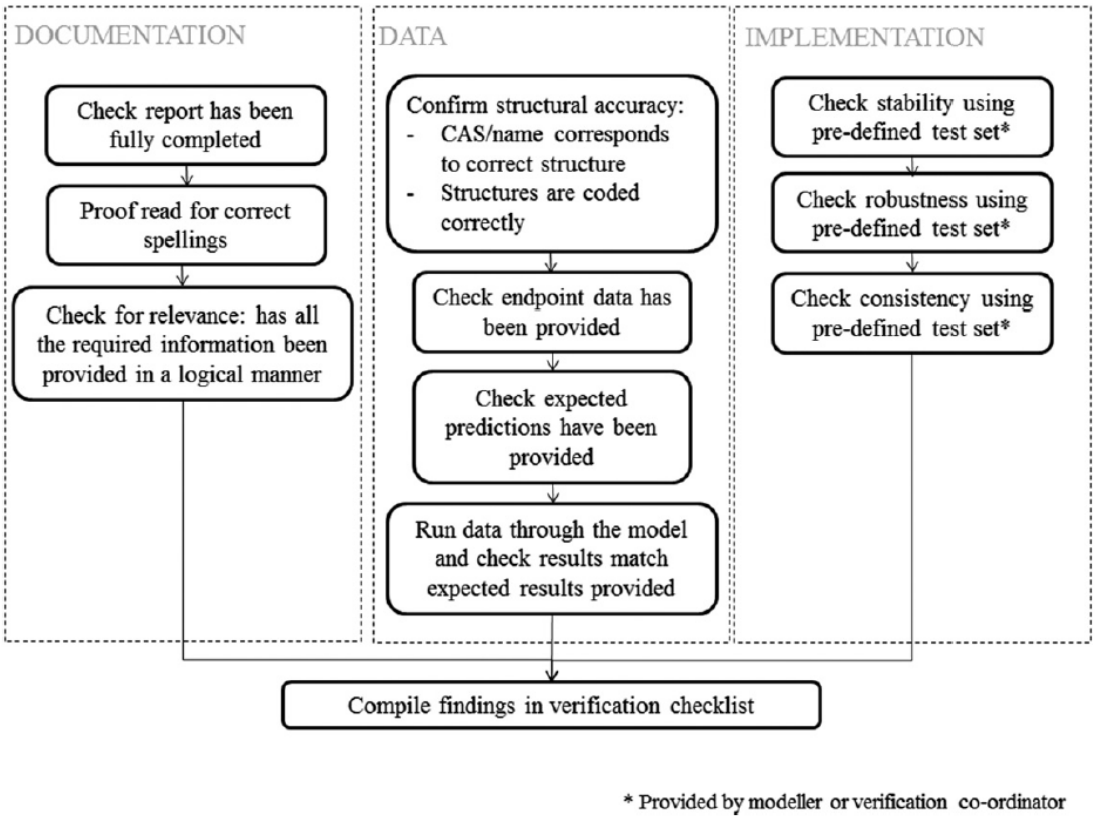


Figure 12. Specific requirements of the verifier (Figure taken from Hewitt *et al.*, 2015).

Databases

283. It has been stated that the nature and quality of data used to build a model is a major determining factor in model acceptability.

284. Data sharing projects, such as PubChem[115] (Wang *et al.*, 2009) have made chemical "big data" publicly available, which advanced modern toxicology studies into a big data era (Zhu *et al.*, 2014) and even using them in risk assessment (Luechtefeld *et al.*, 2018).

285. The term "big data" refers to data sets, structured or unstructured, that multiply quickly and are so large and multifaceted that they are impossible to treat using personal computers and traditional computational approaches (Gandomi and Haider 2015).

286. Data sets with big data require advanced tools such as heterogeneous and cloud computing (Schadt *et al*, 2011) that have capabilities beyond those of conventional data processing and handling techniques as well as dynamic data curation and sharing using algorithms such as those used to handle data streams (Liu *et al.*, 2007; Charikar *et al.*, 2003).

[115] https://pubchem.ncbi.nlm.nih.gov/

287. Publicly available databases store much of the data obtained from the toxicology community, including data from HTS programs such as the ToxCast and Tox21 programs (Ciallella and Zhu 2019).

288. A selection of significant sources representing publicly available big data in the toxicology field (Ciallella and Zhu 2019):

- ACToR[116]: EPA's Aggregated Computational Toxicology Online Resource (ACToR) aggregates data from thousands of public sources on over 500,000 chemicals. It is searchable by chemical name and other identifiers. ACToR is also the data and web applications warehouse for EPA's computational toxicology information which includes high-throughput screening, chemical exposure, sustainable chemistry (chemical structures and physicochemical properties) and virtual tissues data.

- CEBS[117]: Chemical Effects in Biological Systems (CEBS) database. CEBS (Chemical Effects in Biological Systems) is an integrated public repository for toxicogenomic data, including the study design and timeline, clinical chemistry and histopathology findings and microarray and proteomics data. CEBS contains data derived from studies of chemicals and of genetic alterations, and is compatible with clinical and environmental studies. CEBS is designed to permit the user to query the data using the study conditions, the subject responses and then, having identified an appropriate set of subjects, to move to the microarray module of CEBS to carry out gene signature and pathway analysis (Waters *et al*., 2007).

- RepDose[118]: A database on repeated dose toxicity studies of commercial chemicals (Bitsch *et al*., 2006)

Sharing data

289. One of the most well knowns databases is ChemSpider[119]. It is a free chemical structure database owned by the Royal Society of Chemistry providing access to over 67 million structures, properties, and associated information. It also integrates and links compounds from hundreds of high-quality data sources.

290. Another example is the Kyoto Encyclopaedia of Genes and Genomes (KEGG) Pathway database[120] which is a valuable collection of metabolic pathway maps for metabolism, genetic information processing and other functions.

291. Two international initiatives, the eTOX[121] and the eTRANSAFE[122] projects, have shown how sharing data (previously considered to be commercially sensitive) could be achieved, on a with cost basis, with the former project showing

[116] https://actor.epa.gov/actor/home.xhtml
[117] https://manticore.niehs.nih.gov/cebssearch/
[118] https://repdose.item.fraunhofer.de/
[119] http://www.chemspider.com/
[120] https://www.genome.jp/kegg/
[121] http://www.etoxproject.eu/
[122] https://etransafe.eu/

demonstrable success and promise for the future (Sanz *et al*., 2017; Piñero *et al*., 2018).

292. The development of these databases from in-house data complement other activities, such as the freely available COSMOS[123], which have focussed on sharing data for non-pharmaceutical compounds such as cosmetics ingredients and fragrances.

293. The DSSTox Database incorporates state-of-the-art cheminformatics workflows, provides the chemical infrastructure for EPA's Safer Chemicals Research, including the ToxCast and Tox21 high-throughput toxicology efforts.

294. The DSSTox project has the following major elements (Richard and Williams 2002):

(1) to adopt and encourage the use of a common standard file format (structure data file (SDF)) for public toxicity databases that includes chemical structure, text and property information, and that can easily be imported into available CRD applications.

(2) to implement a distributed source approach, managed by a DSSTox Central Website, that will enable decentralized, free public access to structure–toxicity data files, and that will effectively link knowledgeable toxicity data sources with potential users of these data from other disciplines (such as chemistry, modelling, and computer science).

(3) to engage public/commercial/academic/industry groups in contributing to and expanding this community-wide, public data sharing and distribution effort.

295. More recently the distributed structure-searchable toxicity (DSSTox)[124] public database has been established. DSSTox provides a high-quality public chemistry resource for supporting improved predictive toxicology. A distinguishing feature of this effort is the accurate mapping of bioassay and physicochemical property data associated with chemical substances to their corresponding chemical structures.

296. The DSSTox project's overall aims are to effect the closer association of chemical structure information with existing toxicity data, and to promote and facilitate structure-based exploration of these data within a common chemistry-based framework that spans toxicological disciplines.

297. These projects have helped identify and resolve a number of problems; for instance, integration of data from different sources requires the development and implementation of ontologies and other standards eTOX being an example where effort was made to create standardised ontologies (Cronin *et al*., 2019).

[123] http://www.cosmostox.eu/home/welcome/
[124] https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database

298. Currently, there is still no universal criterion to select modelling approaches for big data sets (Ciallella and Zhu 2019).

Methodologies to obtain data

High throughput screening

299. High throughput screening (HTS) is an automated experimental platform for rapidly identifying a small number of molecular entities or conditions with unique biological properties from a large number of tests (Figure 13). A screen is generally considered high throughput if it can assay > 10,000 assays (wells) per day. HTS allows a researcher to quickly conduct millions of tests and to rapidly identify relevant modifier genes, proteins, or compounds involved in a specific biological pathway (Zhong *et al*., 2015).

300. Robotic fluid handling and microspotting[125] facilitate high-throughput studies of different immobilized factors in microarrays, but require time-dependent stimulation to be applied simultaneously (Lanza *et al*., 2011).

301. Many of the technological advances aided HTS advancement, specifically in the following areas (Hertzberg and Pope, 2000):

1. Assay methods and detection (bioware).

2. Liquid handling and robotics (hardware).

3. Process flow and information management (software).

302. The first HTS methods were developed by the pharmaceutical industry (Hertzberg and Pope, 2000) and were *in vitro* assays measuring molecular interactions by fluorescence, luminescence, or absorbance readouts (Inglese *et al*., 2007; Macarron, 2006; Macarron and Hertzberg, 2009).

303. A smooth transition from hits generated *in vitro* to efficacious compounds in more complex disease models in cells, tissues, and most critically in animal models has often been hard to accomplish (Houston and Galetin, 2008; Zhang *et a*l., 2000).

[125]A contact-based method of transferring nucleic acids and peptides, which utilises pins or capillaries to deposit biomolecules on a solid surface

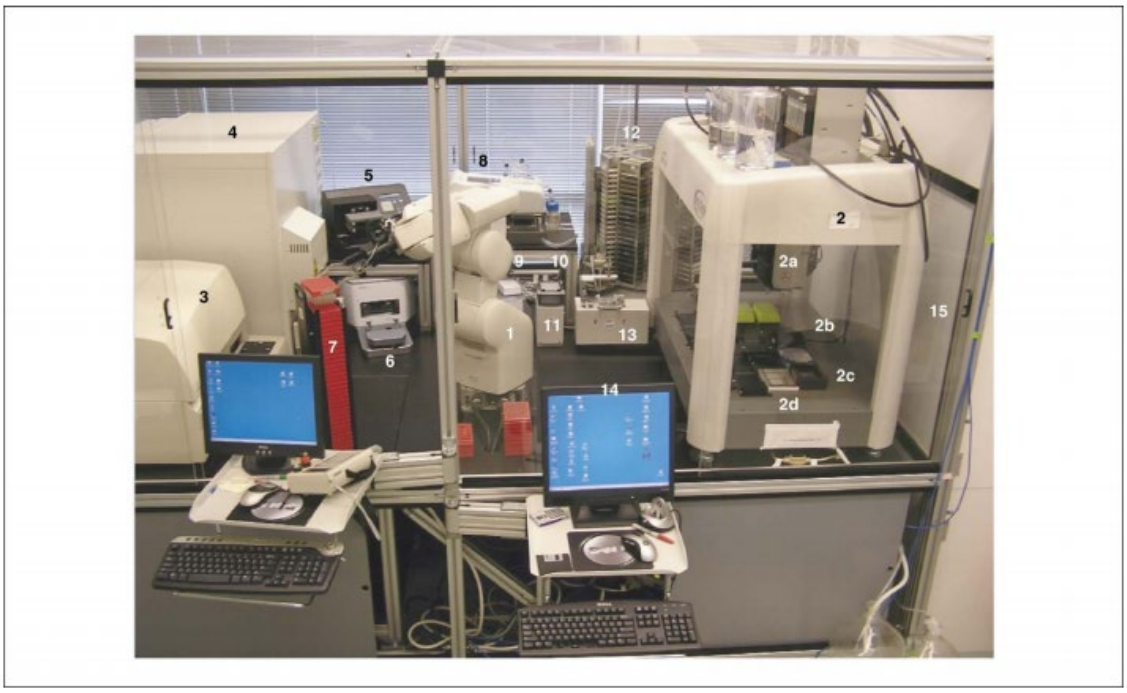


Figure 13. A fully integrated multifunctional robotic screening system. The system is fully enclosed and comprises the following components: (1) Mitsubishi MELFA RV2A 6-axis robot; (2) Caliper Sciclone ALH3000 with interchangeable 96-or 384-tip pipetting head, an independent 8-channel pipettor, two bulk-reagent dispensers, and plate gripper (2a). The following accessories are integrated into the Sciclone: microtiter plate shaker (2b); positive-pressure filtration system (2c); and ultrasonic tip-wash station (2d). (3) PerkinElmer Fusion with 11-mode detection, which includes absorbance, fluorescence, fluorescence polarization, timeresolved fluorescence, time-resolved fluorescence–resonance energy transfer, AlphaScreenTM, etc. (4) Kendro Cytomat6001 with humidity, temperature and CO2 controls, and 189 normal microtiter plate storage capacity. (5) Biotek ELX-405 plate washer, which can be used for 96-well and 384-well plates. (6) Volecity11 Vspin centrifuge can be used for normal and deep well plates. (7) Thermo CRS high-capacity stacker is used to store up to 32 stacked tip boxes. (8) PerkinElmer Flexdrop equipped with four individual dispensing heads that can dispense four bulk reagents in a broad volume range for each head (from 200 nl to 2 ml). (9) Velocity11 Vcode automatic barcode labeler. (10) MicroScan MS-3 barcode reader. (11) Caliper plate regrip station that changes the plate orientation to facilitate the interaction between the robot arm and individual components. (12) Kendro room temperature incubator that stores 189 regular microtiter plates. (13) Caliper plate-lid-handling station. (14) Six Variomag shaker station that provides an independent plate-shaking operation (behind the monitor, not visible). (15) Liberty Industry air purifier provides ultra-dust-free conditions for the enclosed system and prevents the introduction of contaminants, from the surrounding air, to the work area. The reader table is modular and can be swapped for another reader if necessary. (Figure taken from Wu and Doberstein 2006).

304. Cell-and animal-based assays have the advantage of identifying compounds within the complex environments of cells and tissues but are more costly, difficult to miniaturize, and tend to have lower throughput due to their complexity. Diseases that affect the brain add another layer of complications. Mature neurons differentiated *in vivo* must be derived from primary sources and are difficult to transfect. Thus, most primary large-scale screens use neuroblastoma cell lines.

305. Better culturing and transfection protocols, however, now make HTS with primary neurons more feasible, and the increased biological and clinical relevance (Daub *et al*., 2009; Nolan, 2007) is worth the extra effort and expense.

306. HTS data were already proposed as a means by which to prioritize chemicals for further testing (U.S. EPA, 2014[126]). A comparison of HTS data with animal data for predictivity assessments of human toxicity is ongoing (Liu *et al*., 2017).

307. The goal is to use *in vitro* data to predict *in vivo* human effects, reducing the reliance on animal testing (NCATS, 2016[127]; NTP, 2016[128]). The EPA's ToxCast programme, which is part of the larger Tox21 interagency collaboration, has provided a wealth of HTS data to the toxicology community, generating data on more than 3000 chemicals across 1000 assay endpoints, with the goal of generating screening data that could be used for prioritizing chemicals for further testing (Dix *et al*., 2007; Richard *et al*., 2016).

308. Nel *et al.* (2012) stated key ingredients for developing predictive toxicology through HTS approaches towards NPs:

- A well-characterized nanoparticle library.

- The second infrastructure requirement for a predictive toxicological approach is the development of appropriate HTS approaches to quantitatively assess dose- and time-dependent cellular injury responses that are predictive of *in vivo* adverse outcomes through assays.

- Mechanistic injury pathways such as: oxidative stress, dissolution and release of toxic metal ions; cationic injury to surface membrane and organelles; pro-fibrogenic responses to CNT; inflammasome activation by long aspect ratio materials; photoactivation and influence of bandgap; zebrafish embryo hatching interference; cell membrane lysis by surface reactivity.

- Development of a data analysis framework that includes *in silico* tools for data analysis, data transformation, machine learning and modelling of HTS data.

309. HTS can facilitate collecting data for biological heat maps (Figure 14). Biology heat maps are typically used in molecular biology to represent the level of expression of many genes across a number of comparable samples (*e.g.* cells in

[126] https://www.epa.gov/chemical-research/rapid-chemical-exposure-and-dose-research
[127] https://ncats.nih.gov/tox21
[128] https://ntp.niehs.nih.gov/whatwestudy/tox21/index.html

different states, samples from different patients) as they are obtained from DNA microarrays.

310. High-density data sets with hundreds to thousands of data points for each tested subject. In the field of toxicology, genomics technologies have been used to investigate how different stresses alter gene expression. For large data sets, including large human clinical/epidemiological studies, it can be problematic to effectively evaluate the phenotypic anchor, due to the sheer number of data points to consider (Auman *et al*., 2007).

311. Clinical chemistry data are often viewed in a data table or a bar graph, where one can examine the changes that occur for one analyte across the groups of interest. For a study involving only one or a few compounds, these types of visualizations help investigators determine how subjects in each group react to the given stressor. However, for large animal data sets involving multiple compounds, dose groups, and time points, it is very difficult to give a meaningful visual representation of the data with traditional bar graphs due to the number of data points that exist in these types of experiments (Auman *et al*., 2007).

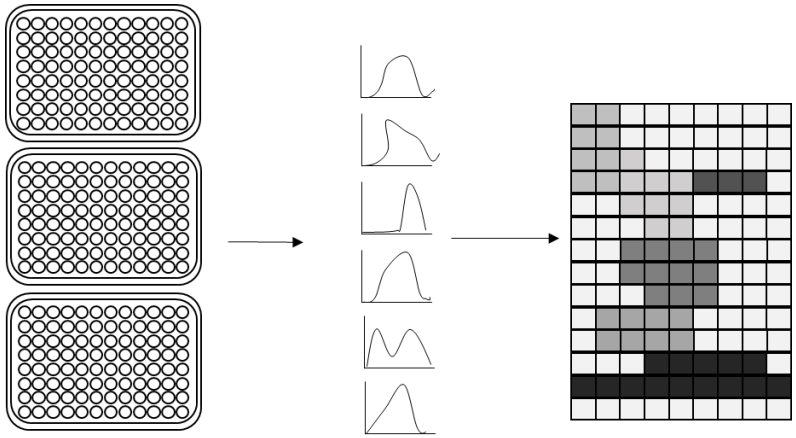


Figure 14. How data from heatmaps is obtained through HTS. Heatmaps help to represent the level of expression of many genes across a number of comparable samples (*e.g.* cells in different states, samples from different patients) from large data sets obtained through HTS.

312. Another example of the utility of HTS is gathering data for self-organising maps to visualise and analyse the diversity of databases. (Kohonen, 1991; Oja and Kaski 1999).

313. Examples of heatmaps and self-organisation:

- Heat map visualization of high-density clinical chemistry data (Auman *et al*., 2007).

- Use of an HTS approach coupled with *in vivo* zebrafish embryo screening to develop hazard ranking for engineered nanomaterials (George *et al*., 2011).

Emerging Technologies

Microfluidics

314. Recent advances in microfluidic platforms, which facilitate studies of diffusible growth factors with spatial and temporal control, have historically been difficult to scale-up into larger, individually manipulated culture chambers (Bhumiratana, 2014). However, versatile, fully automated microfluidic cell culture systems that create arbitrary media formulations in independent cell culture chambers can help address these requirements (Bhumiratana, 2014).

315. Based on cell handling techniques, microfluidics has been widely applied in the fields of Polymerase Chain Reaction (PCR), immunoassays, organ-on-chip, stem cell research, and analysis and identification of circulating tumour cells. As a major step in drug discovery, HTS allows rapid analysis of thousands of chemical, biochemical, genetic or pharmacological tests in parallel.

316. Culture conditions can be customized in terms of cell seeding density, composition of culture medium, and feeding schedule, and each culture chamber can be imaged with time-lapse microscopy. Multiplexers and mixers (controlled by actuators and valves) allow generation of broad ranges of medium compositions and accurately control the temporal feeding/washing patterns (Bhumiratana, 2014).

317. Recently, microfluidic devices have been proposed as a potential platform for HTS technology because of their properties of low sample consumption, low analysis cost, easy handling of nanoliter-volumes of liquids and being suitable for cell-based assays (Dittrich and Manz, A 2006; Neužil *et al*., 2012; Hong *et al*., 2009).

318. There are many applications making use of these properties of microfluidics, with a plethora in drug discovery. Some representative examples include:

319. Ye *et al*. (2007) presented a high content multiparametric screening method (plasma membrane permeability, nuclear size, mitochondrial transmembrane potential and intracellular redox states) for human liver carcinoma (HepG2) responding to multiple anti-cancer drugs with different concentrations by integrating 8 such gradient generators with parallel cell culture chamber (Qin, et al., 2005) (Figure 15).

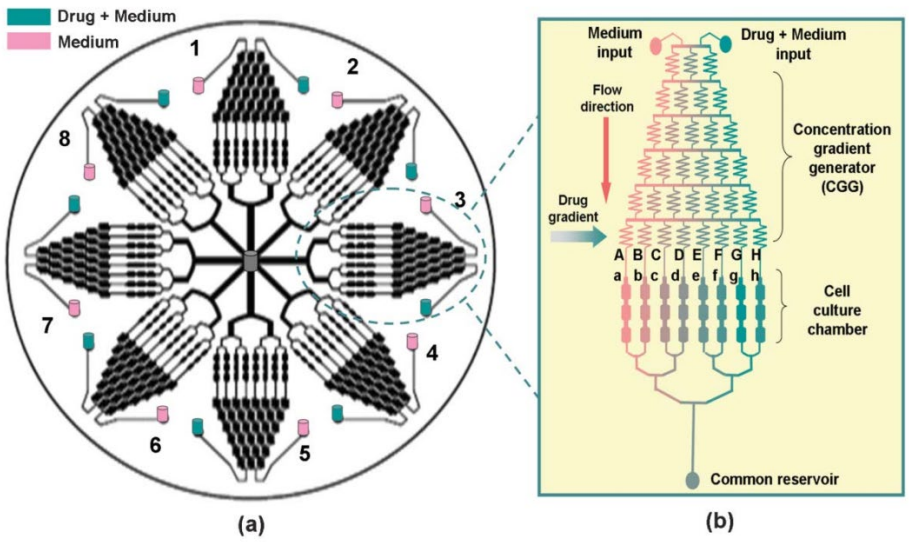


Figure 15. Schematic of the integrated microfluidic device for cell-based high content screening (HCS). (a) The device consists of eight uniform structure units and each unit is connected by a common reservoir in the centre of the device. (b) Magnified section of the single structure unit containing an upstream concentration gradient generator (CGG) and downstream parallel cell culture chambers (Figure taken from Yen *et al*., 2007).

320. Chen *et al*. (2012) reported a platform based on stable isotope labelling carried out in a microfluidic chip with electrospray ionization mass spectrometry for qualitative and quantitative analysis of the metabolism of cells treated by drugs. This platform has integrated cell culture chambers, on-chip sample preparation (*i.e.* a solid phase extraction (SPE) module) and ESI-MS (Figure 16). This platform has the potential to be used as an on-line multiparameter cell metabolism analysis platform for high throughput drug screening.

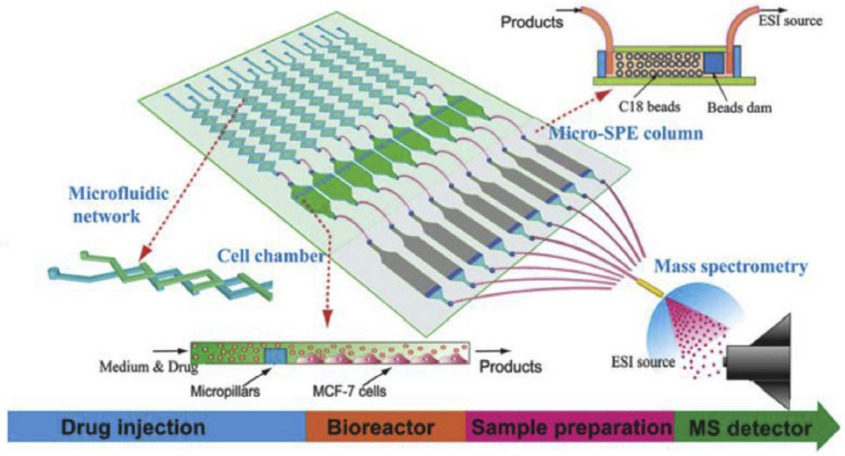


Figure 16. Schematic diagram of the chip ESI-MS platform for qualitative and quantitative analysis of the metabolic activity of cells exposed to drugs (Figure taken from Chen *et al*., 2012).

321. The use of hanging drops on the underside of culture plate lids is a typical method to generate 3D cellular spheroids. 3D cell spheroid culture allows for cellular

self-organization and enables straightforward monitoring. As a result, the method can provide valuable information that is physiologically more relevant than 2D cell culture.

322. Tsung *et al*. (2011) achieved a hanging droplet microarray system for drug testing in cellular spheroid formation (Figure 17). This platform significantly simplified the experimental process for cell culture and cellular formation in hanging droplets.

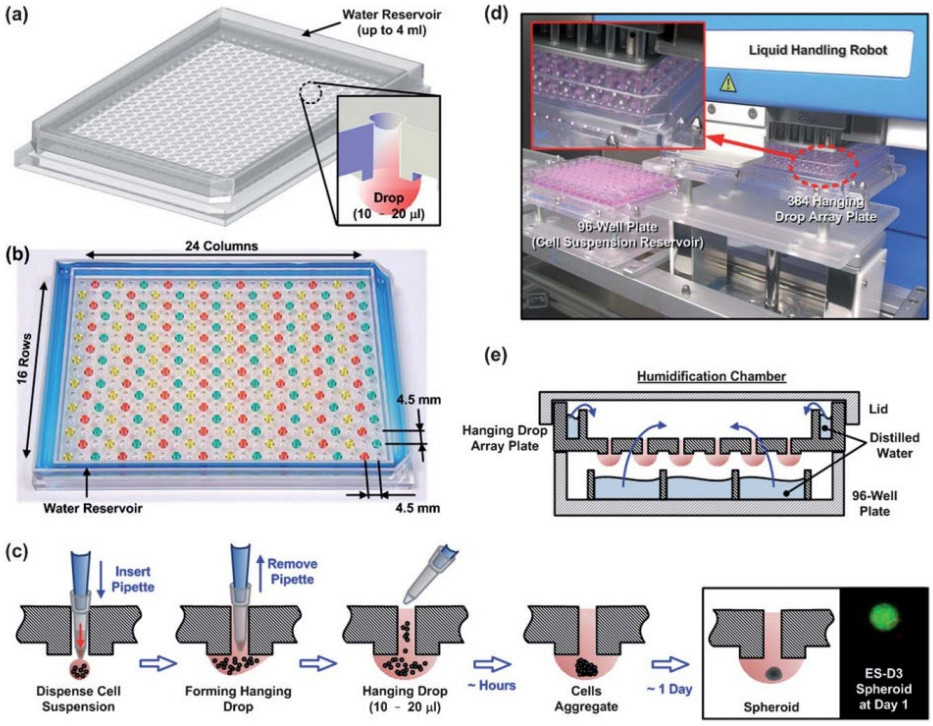


Figure 17. (a) Illustration of the designed 384 hanging drop spheroid culture array plate, and its cross-sectional view. (b) Photo and key dimensions of the array plate. (c) Cartoon of the hanging drop formation process in the array plate. The pipette tip is first inserted through the access hole to the bottom surface of the plate, and cell suspension is subsequently dispensed. Cell suspension is quickly attracted to the hydrophilic plate surface and a hanging drop is quickly formed and confined within the plateau. Within hours, individual cells start to aggregate and eventually form into a single spheroid around 1 day. (d) Photo of the 384 hanging drop array plate operated with liquid handling robot capable of simultaneously pipetting 96 cell culture sites. (e) Cartoon of the final humidification chamber used to culture 3D spheroids in the hanging drop array plate. The 384 hanging drop array plate is sandwiched between a 96-well plate filled with distilled water and a standard-sized plate lid. Distilled water from the bottom 96-well plate and the peripheral water reservoir prevent serious evaporation of the small volume hanging drops (Figure taken from Tsung *et al*. (2011)).

Organ on a chip

323. Organs on chips are microengineered biomimetic systems that represent key functional units of living human organs. They often consist of transparent 3D polymeric microchannels lined by living human cells and replicate three important aspects of intact organs: the 3D microarchitecture defined by the spatial distribution of multiple tissue types; functional tissue-tissue interfaces; and complex organ-specific mechanical and biochemical microenvironments (Huh *et al*., 2011, 2010).

324. These systems could be used as specialized *in vitro* models that permit simulation, mechanistic investigation and pharmacological modulation of complex biological processes.

325. In recent years, this biomimetic microsystem approach has been used to establish microengineered models that recapitulate the structural and functional complexity of human organs such as the liver, heart, lung, intestine, kidney, brain and bone (Huh *et al*., 2012)

326. A representative example is the lung on-a-chip microdevice that reconstitutes the mechanically active alveolar-capillary barrier in the human lung (Figure 18). This model is created in a compartmentalized 3D microfluidic system in which human alveolar epithelial cells are cultured in close apposition with human pulmonary microvascular endothelial cells on a thin porous elastomeric membrane to form a barrier tissue that resembles the *in vivo* alveolar-capillary interface. This microfluidic cell culture system is integrated with a biologically inspired mechanical actuation system that uses computer-controlled negative pressure to cyclically stretch the alveolar-capillary barrier to mimic physiological breathing motions (Huh *et al*., 2012).

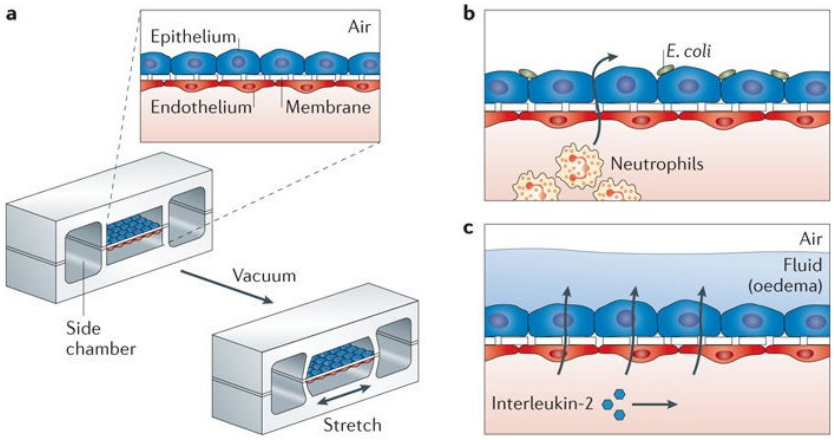


Figure 18. a) A human breathing lung-on-a-chip was created by co-culturing human alveolar epithelial cells and pulmonary microvascular endothelial cells on opposite sides of a stretchable porous membrane to replicate the alveolar-capillary boundary of the breathing human lung. A vacuum was applied to mimic the tissue stretch that occurs during normal breathing. b) this system was used to reconstitute integrated organ-level functions such as inflammatory responses to intra-alveolar pathogenic bacteria such as Escherichia coli that are mediated by endothelial recruitment of circulating neutrophils, transmigration through the alveolar-capillary interface and

subsequent bacterial phagocytosis. c) The lung-on-a-chip was used to model human lung diseases such as pulmonary oedema. Administration of interleukin-2 into the microvascular channel resulted in fluid leakage into the alveolar compartment, recapitulating the pulmonary oedema induced by acute toxicity of interleukin-2 that is observed in patients with cancer (Figure taken from Huh *et al*., 2012)).

327. Bioprinting is a revolutionary technology to assemble scaffolds for growing tissue. It has been applied to fabricate organ-on-a-chip models owing to its ability to print multiple materials and cell types simultaneously with good spatial resolution and reproducibility. This enables the creation of a biomimetic microenvironment with heterogeneous 3D structures. Functional vascularized tissue structure can be printed directly enabling fluid flow for transport of nutrition, gaseous exchange and removal of waste (Yu and Choudhury, 2019).

328. Bioprinting allows automated fabrication of reproducible tissue constructs with precise control over spatial parameters. Organ-on-a-chip platforms offer the ability to mimic physiological, mechanical and chemical cues *in vitro*. Integration of the two technologies presents a promising direction for high-throughput drug validation and testing as an alternative to animal and human models. Bioprinted cell models on microfluidic chips that mimic the microenvironment, spatial distribution and vasculature represent a promising strategy for future 4-D bioprinting, where smart materials are printed (Yu and Choudhury, 2019).

Plant scaffolds

329. A recent study (Gershlal *et al*., 2017) discussed using plant tissues as scaffolds for regenerating large volume vascularized tissue. By taking advantage of the similarities in the vascular structure of plant and animal tissues, they developed decellularized plant tissue as a pre-vascularized scaffold for tissue engineering. Perfusion-based decellularization was modified for different plant species, providing different geometries of scaffolding. After decellularization, plant scaffolds remained patent and able to transport microparticles. Plant scaffolds were recellularized with human endothelial cells that colonized the inner surfaces of plant vasculature. Human mesenchymal stem cells and human pluripotent stem cell derived cardiomyocytes[129] adhered to the outer surfaces of plant scaffolds. Cardiomyocytes demonstrated contractile function and calcium handling capabilities over the course of 21 days.

Artificial Intelligence

330. Experts from environmental health sciences had a workshop (NIEHS-funded workshop was sponsored by the National Academies of Science, Engineering, and Medicine (NASEM)) earlier this year on "*Leveraging Artificial Intelligence and Machine Learning to Advance Environmental Health Research and Decisions*"[130]. It was stated that artificial intelligence (AI) may revolutionize environmental

[129] Cardiac muscle cells or cardiomyocytes (also known as myocardiocytes or cardiac myocytes are the muscle cells (myocytes) that make up the cardiac muscle (heart muscle).
[130] http://nas-sites.org/emergingscience/meetings/ai/

epidemiology, exposure and toxicity assessment, and other studies, with careful attention to data quality.

331. AI is the simulation of the human intelligence process by computers. The process includes acquiring information, developing rules for using the information, drawing approximate or definite conclusions and self-correction (Mak and Pichika 2018).

332. The use of Machine Learning (ML) methods trained on empirical data could be advantageous to make predictions on the potential degradation and reduction in toxicity occurring during remediation. ML models are able to learn the relationships between input variables (*e.g.* soil amendment, soil type) and output variables (e.g. long-term changes in contaminants 'bioavailability) from a training dataset, these relationships can then be generalised to make informed decisions in new cases (Wu *et al*., 2013).

333. ML is categorised into supervised, unsupervised and reinforcement learning (Mak and Pichika 2018).

334. Supervised learning comprises classification and regression methods where the predictive model is developed based upon the data from input and output sources. Output from supervised ML entails disease diagnosis under the subgroup classification; and drug efficacy and ADME Tox[131] prediction under the subgroup regression (Gunčar *et al*., 2018).

335. Unsupervised learning comprises clustering and feature-finding methods by grouping and interpreting data based solely on input data (Koohy, 2017). Through unsupervised ML, outputs such as disease subtype discovery from clustering and disease target discovery from feature-finding methods can be attained (Young *et al*., 2017).

336. Reinforcement learning is largely driven by decision making in a given environment and its execution to maximise its performance. The outputs from this type of ML include de novo drug design under decision making and experimental designs under execution-where both can be achieved via modelling and quantum chemistry (Chen *et al*., 2018).

337. A further subfield of ML called deep learning (DL) uses artificial neural networks that adapt and learn from the vast amount of experimental data (Lee et al., 2017; Grys *et al*., 2017).

338. Some examples include:

- Xu *et al*. 2017, reported three neural network models developed to predict acute oral toxicity end points based on a training set of 8080 compounds. All three models (*i.e*., a regression model for LD_{50} values, a multiclassification model for US EPA hazard categories, and a multitask model to

[131] ADME-Tox: absorption, distribution, metabolism, and excretion - toxicity in pharmacokinetics

simultaneously predict both of these end points) simultaneously outperformed previously.

- Wen *et al*. 2017 also reported a deep learning model developed to predict interactions between drugs and their biological targets based on 15 524 drug-target pairs obtained from the DrugBank database[132]. This model employed a pretraining feature extraction step to predict whether specific drug-target pairs will interact and overall outperformed classic QSAR approaches.

- Pu *et al*., 2019 elucidated on eToxPred[133], which employs machine learning algorithms trained on molecular fingerprints to evaluate drug candidates products, and synthetic bioactive compounds.

Moving from research to risk assessment to regulatory testing *and beyond*

339. Biological models of metabolism, pharmacokinetics, and risk estimation have been prominent in toxicology for a few decades; however, new graphical and analytical tools and methods are needed in order to "decode the toxicological blueprint of active substances that interact with living systems" (Sturla *et al*. 2014).

340. The needs of risk assessment are context-dependent and can vary from simple classification of a substance for hazard (*e.g.* is it genotoxic or not) to prioritization by the nature and severity of hazard for further investigation to quantitative estimates of risk to determine the urgency and nature of any risk-management action (Sturla *et al*. 2014).

341. It was suggested that systems toxicology[134] can provide a deep mechanistic understanding of toxicological effects, permitting prediction of responses to chemicals. If adequately described, a systems description should enable prediction of responses for which experimental data were not available (*i.e*. the system will exhibit emergent properties entailing novel patterns and properties arising from the inherent structure of the system) (Sturla *et al*. 2014).

342. Systems toxicology has an ultimate potential for extrapolating from early and highly sensitive quantifiable molecular and cellular events to medium and long term outcomes at the organism level, and its application could be part of a new paradigm for risk assessment (Figure 19).

[132] https://www.drugbank.ca/
[133] https://github.com/pulimeng/etoxpred
[134] Systems Toxicology is the integration of classical toxicology with quantitative analysis of large networks of molecular and functional changes occurring across multiple levels of biological organization.

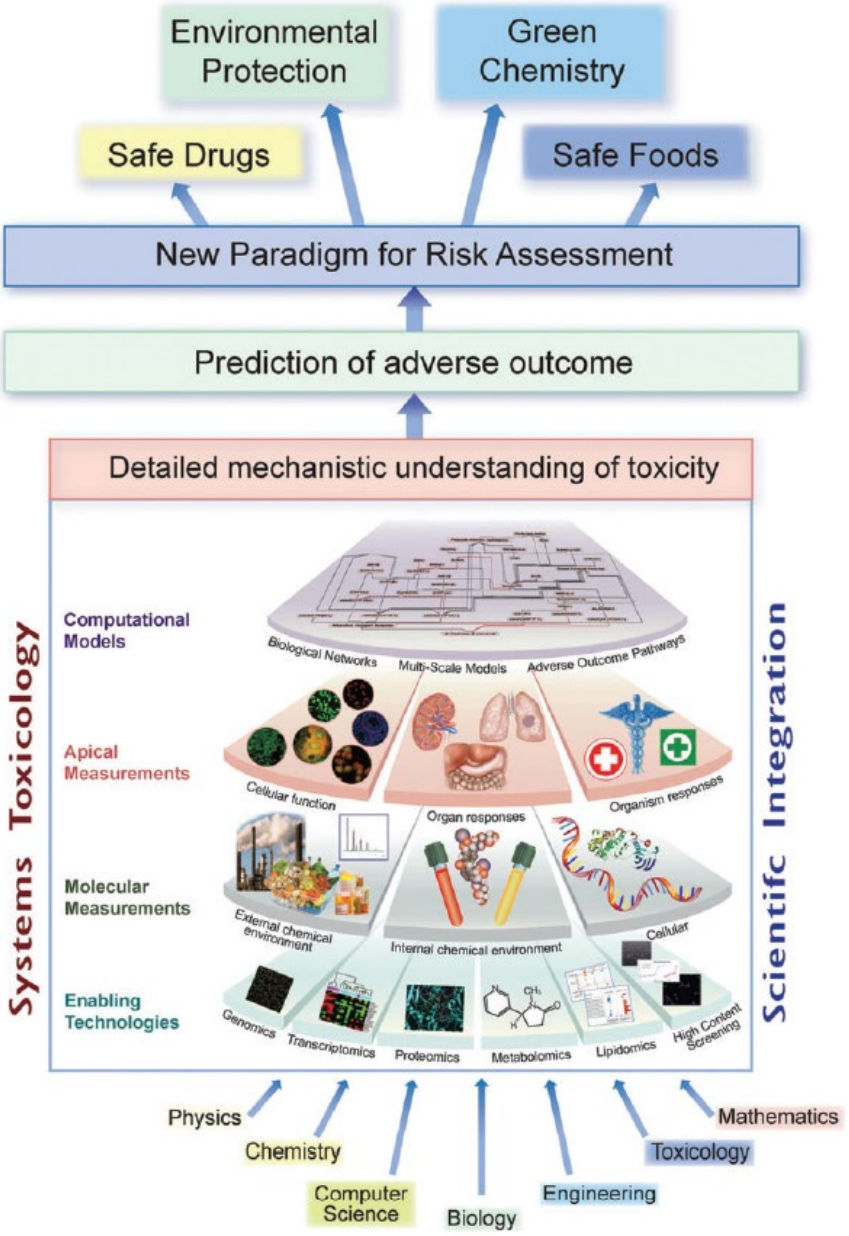


Figure 19. What is Systems Toxicology? Systems Toxicology is aimed at decoding the toxicological blueprint of active substances that interact with living systems. It resides at the intersection of Systems Biology with Toxicology and Chemistry. It integrates classic toxicology approaches with network models and quantitative measurements of molecular and functional changes occurring across multiple levels of biological organization. The multidisciplinary Systems Toxicology approach combines principles of chemistry, computer science, engineering, mathematics, and physics with high content experimental data obtained at the molecular, cellular, organ, organism, and population levels to characterize and evaluate interactions between potential hazards and the components of a biological system. It is aimed at developing a detailed mechanistic as well as quantitative and dynamic understanding of toxicological processes, permitting prediction and accurate simulation of complex (emergent) adverse outcomes. Thereby, the approach provides a basis for translation between model systems (*in vivo* and *in vitro*) and

study systems (*e.g*. human, ecosystem). Systems Toxicology, therefore, has an ultimate potential for extrapolating from early and highly sensitive quantifiable molecular and cellular events to medium and long term outcomes at the organism level, and its application could be part of a new paradigm for risk assessment (Figure taken from Sturla *et al*., 2014).

Regulatory setting

343. With the goal of faster, less expensive and more predictive assessment approaches, many new technologies have been proposed. Although some of these new assessment approaches have gained considerable attention, few if any have been universally accepted as available to replace existing testing paradigms (Sikkler *et al*., 2018).

344. In order to build confidence in new methods their value needs to be proven, and this is best accomplished by applying them in a “fit-for-purpose” manner to address areas of uncertainty that are difficult to address using conventional toxicology methods.

345. For example, the assessment of mixtures, read-across based on mechanism of action, and the role of human genetic variability are all areas for which new methods offer unique contributions toward reducing uncertainty. Acceptance of test methods by regulatory authorities is difficult to achieve and usually occurs gradually (Knudsen *et al*., 2017).

346. Three major opportunities exist for improving the current human risk assessment paradigm (Zeis *et al*., 2012):

(1) Derivation of probabilistically based human-specific toxicity estimates to replace deterministic estimates based on rodent models.

(2) Genetically diverse experimental animal systems to assess phenotypic variation for adverse outcomes.

(3) Integrated quantitative analysis of human variability and susceptibility.

347. Opportunities exist to incorporate modelling and more quantitative estimates of data values and variability into all areas of risk assessment, from *in vitro* to PK to animal data to human estimates.

348. Ultimately, probabilistic toxicology will be able to break down artificial dichotomies, moving the scientific community toward a broader conception of population health, where toxicological responses are modelled using a continuum from ‘no effect’ to ‘effect’ (*i.e*. non-toxic to toxic), based on probabilistic measures, such as the chemical potency confidence interval and the benchmark dose confidence interval in reference to a genetically diverse population with a characteristic distribution of susceptible subpopulations (Knudsen *et al*., 2017).

Limitations of the predictive toxicological approach

349. New approaches, new types of data, and new technologies for human risk assessment have been sometimes perceived with reluctance (Knudsen *et al*., 2017).

350. This is due to well-recognized limitations of *in vitro* systems such as: limited metabolic capacity; inability to account for ADME and pharmacokinetics; uncertainty in cell-cell network interactions; life stage considerations; reductionism in the biological complexity of systems being tested and biological diversity in general (Chiu *et al*., 2013; Kavlock *et al*., 2012; Osborne *et al*., 2017; Sturla *et al*., 2014).

351. Extrapolating *in vitro* effects to *in vivo* prediction faces the general problem of false-positive (*in vitro* positive, *in vivo* negative) and false-negative (*in vitro* negative, *in vivo* positive) results that may arise for many reasons: pharmacokinetic issues that impact biotransformation and/or clearance *in vivo*; incomplete assay coverage of molecular pathways and biological processes; physical limitations of complex multi-cellular networks and interactions between diverse cell types; statistical power in analysing diverse, multidimensional data sets; and the potential for *in vivo* adaptation through homeostatic mechanisms (Knudsen *et al*., 2017).

352. *In vitro* to *in vivo* modelling does hold promise in predictive toxicology when *in vitro* assays can plausibly be used to link specific molecular endpoint perturbations to an AOP in humans or ecological populations. Before this vision becomes a reality, AOPs themselves must be established with enough precision and detail to enable an understanding of false positive and negative predictions (Knudsen *et al*., 2017).

353. A lot would depend on real world scenarios and fate, transport and life cycle analysis would be beneficial in the assessment.

Wider implementations of a predictive toxicological approach

354. Predictive toxicological approach can speed up hazard ranking and decision making (Nel *et al*., 2011). Furthermore, it can be included in a chemical life cycle analysis (Fanke *et al.,* 2018).

355. An example of this of life cycle impact assessment: toxicity estimate by USETox[135] (Schupp *et al*, 2017) USEtox is a scientific consensus model endorsed by the United Nations Environment Programme (UNEP)/ Society of Environmental Toxicology and Chemistry (SETAC) Life Cycle Initiative for characterizing human and ecotoxicological impacts of chemicals. Main output is a database of recommended and interim characterization factors including fate, exposure, and effect parameters.

356. Fanke *et al* 2018 elucidated on the advancements in life cycle human exposure and toxicity characterization (Figure 20).

[135] https://usetox.org/

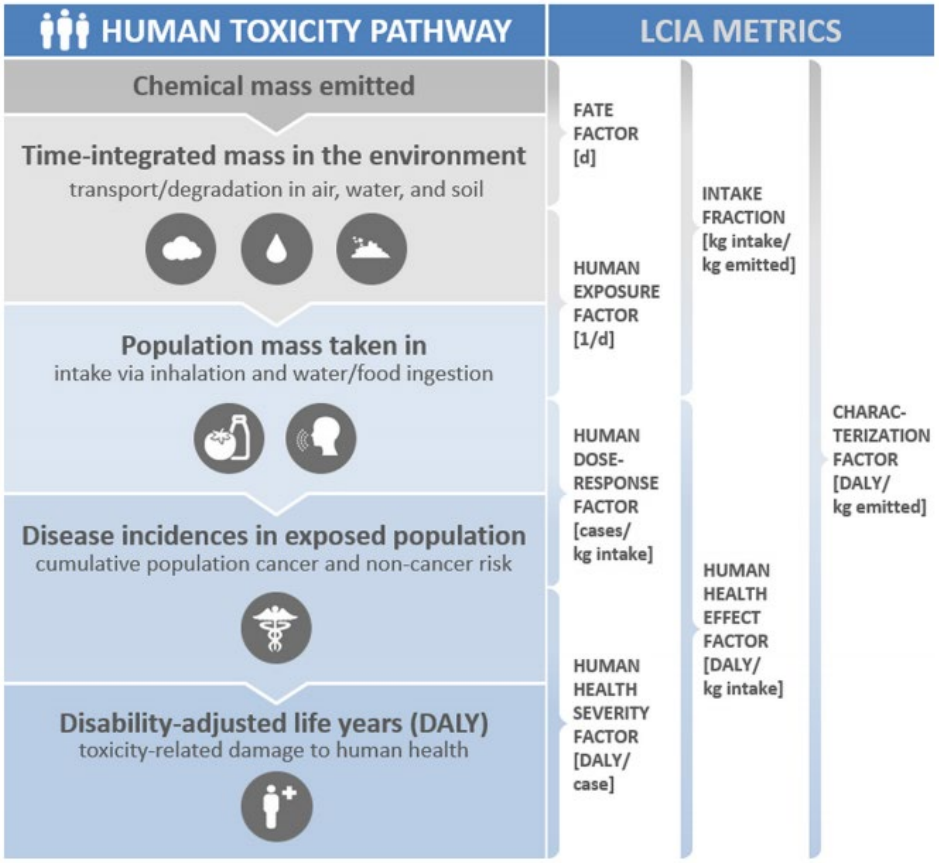


Figure 20. Generalized illustrative representation of the existing life cycle human toxicity source-to-damage characterization framework. Units of metrics and impact pathways considered may differ between methods (Figure taken from Franke *et al*., 2018).

Collaboration

357. The focus of the 7th annual Global Summit on Regulatory Science (GSRS17) was Emerging Technologies for Food and Drug Safety[136]. GSRS17 is an international conference held under the auspices of the Global Coalition for Regulatory Science Research (GCRSR), with the goal of discussing innovative technologies and developing partnerships to enhance translation of basic science into regulatory applications within the global context. The conference provided an interactive platform for scientists from government, industry, and academic-research communities to objectively assess the utility of emerging technologies (such as nanotechnology, imaging, omics for translational science, precision medicine, bioinformatic approaches, medical product safety, and food safety) for addressing regulatory research questions and to discuss the best way to translate these technologies into real-world applications. GSRS17 also allowed the opportunity to exchange views and practices that can assist the regulatory research community in harmonizing educational and training opportunities world-wide (Sikkler *et al*., 2018).

358. Food authorities should strive to incorporate the best scientific methods available. These include activities such as source attribution, risk assessment

[136] https://www.fda.gov/about-fda/science-research-nctr/global-summit-regulatory-science-brasilia-brazil-09182017-09202017

(identification in gene tests), and a focus on developing rapid diagnosis testing. Next generation sequencing methods are replacing traditional molecular tests used to inform on these activities, but this process presents several significant challenges that must be addressed before these new methods can be approved for application in a food regulatory environment. It is important to invite all the stakeholders to facilitate an improved probability of standardization to organize metadata. In support, a consortium was formed with a focus on three ontologies[137], such as FoodOns[138] (Sikkler *et al*., 2018).

359. The Chemical and Food Ingredient Safety Program works actively with international regulatory and research agencies around the world to address key current gaps in chemical safety assessment and regulation. There are ongoing collaborations with the US EPA, Health Canada, ECHA, and EFSA, including an EPA-led international case study to examine the utility of *in vitro* bioactivity data as a conservative estimate of point-of-departure (POD) for chemical risk assessments, as well as collaborative projects with the EPA focused on the development of *in vitro* assays for organ-specific and developmental toxicity.

360. In the GSRS17 meeting, it was said that moving forward toward greater integration of emerging data and novel methodologies for chemicals risk assessment in Canada will need continuous efforts on capacity building. This will be accomplished through increased data accessibility and sharing, the maintenance and establishment of key partnerships, technical workshops and training sessions with international experts, and ongoing focus on data analysis tools development to address regulatory questions. It is also important to demonstrate proof of concept through various case studies and work collaboratively on the interpretation and application of new data for use in regulatory applications. This is currently being done at an international level under the OECD and as the focus of the Accelerating the Pace for Chemical Risk Assessment initiative co-lead by the US EPA, the ECHA and Health Canada (Kavlock *et al*., 2018).

361. The future direction of safety assessment science will depend heavily on the evolution of the regulatory landscape. A key challenge, though, is whether the regulatory framework can keep pace with the increasing speed of scientific and technological developments (Worth *et al*., 2019).

362. This implies close collaboration between chemists, toxicologists, informaticians and risk assessors to develop, maintain and utilise appropriate models. Not only must the different disciplines come together, but also those scientists from industry, academia and regulatory agencies must recognise the commonalities (Cronin *et al*., 2018). The challenge is to respond to the growing need for adaptable, flexible and even bespoke computational workflows that meet the demands of industry and regulators, by exploiting the emerging methodologies of Tox21 and risk assessment.

[137] In computer science and information science, an ontology encompasses a representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many or all domains of discourse.
[138] https://foodon.org/

Conclusions

363. The combined advances in discovery and clinical sciences, data science and technology has resulted in toxicity testing reaching a pivotal transformation point taking advantage of the 4IR.

364. Many different types of *in silico* methods have been developed to characterize and predict toxic outcomes in humans and environment.

365. These *in silico* methods include databases, different kinds of QSAR methods, AOPs, HTS, pharmacophores, homology models and other molecular modelling approaches, machine learning, data mining, network analysis tools, and data analysis tools using AI.

366. The improved *in silico* technologies presents an opportunity in toxicology to bridge the communication gap and collaboration with scientists from industry, academia and regulatory agencies to develop, maintain and utilise appropriate models.

Questions for the COT

i) Are there any models the Members think would be applicable to risk assessment for chemicals in food?

ii) Are there any specific areas that Members think should be/would like covered in the workshop next year?

iii) Any other possible experts/speakers that we may want to consider inviting to the workshop?

iv) Any other comments?

Secretariat November 2019

References

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K. and Serrrano, J.A., 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry: An International Journal*, 29(3), pp.730-741.

Amberg, A., Beilke, L., Bercu, J., Bower, D., Brigo, A., Cross, K.P., Custer, L., Dobo, K., Dowdy, E., Ford, K.A. and Glowienke, S., 2016. Principles and procedures for implementation of ICH M7 recommended (Q) SAR analyses. *Regulatory Toxicology and Pharmacology*, 77, pp.13-24.

Andersen, M.E. and Dennison, J.E., 2001. Mode of action and tissue dosimetry in current and future risk assessments. *Science of the Total Environment*, 274(1-3), pp.3-14.

Andersen, M.E., 1995. Development of physiologically based pharmacokinetic and physiologically based pharmacodymamic models for applications in toxicology and risk assessment. *Toxicology letters*, 79(1-3), pp.35-44.

Ashby, J. and Tennant, R.W., 1991. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the US NTP. Mutation *Research/Reviews in Genetic Toxicology*, 257(3), pp.229-306.

Ashby, J. and Tennant, R.W., 1988. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the US NCI/NTP. *Mutation Research/Genetic Toxicology*, 204(1), pp.17-115.

Ashby, J., 1985. Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. *Environmental mutagenesis*, 7(6), pp.919-921.

Astel, A.M., 2010. Air contaminants modelling by use of several receptor-oriented models. *International Journal of Environment and Pollution*, 42(1-3), pp.32-57.

Auman, J.T., Boorman, G.A., Wilson, R.E., Travlos, G.S. and Paules, R.S., 2007. Heat map visualization of high-density clinical chemistry data*. Physiological genomics*, 31(2), pp.352-356.

Avent, M.L., Vaska, V.L., Rogers, B.A., Cheng, A.C., Van Hal, S.J., Holmes, N.E., Howden, B.P. and Paterson, D.L., 2013. Vancomycin therapeutics and monitoring: a contemporary approach. *Internal medicine journal*, 43(2), pp.110-119.

Appelt, A.L., Bentzen, S.M., Jakobsen, A. and Vogelius, I.R., 2015. Dose-response of acute urinary toxicity of long-course preoperative chemoradiotherapy for rectal cancer. *Acta Oncologica*, 54(2), pp.179-186.

Bachler, G., von Goetz, N. and Hungerbuhler, K., 2015. Using physiologically based pharmacokinetic (PBPK) modeling for dietary risk assessment of titanium dioxide (TiO2) nanoparticles. *Nanotoxicology*, 9(3), pp.373-380.

Bachler, G., von Goetz, N. and Hungerbühler, K., 2013. A physiologically based pharmacokinetic model for ionic silver and silver nanoparticles. *International journal of nanomedicine*, 8, p.3365.

Bajorath, J., 2017. Representation and identification of activity cliffs. Expert Opinion on Drug Discovery, 12 (9), pp 879–883

Basei, G., Hristozov, D., Lamon, L., Zabeo, A., Jeliazkova, N., Tsiliki, G., Marcomini, A. and Torsello, A., 2019. Making use of available and emerging data to predict the hazards of engineered nanomaterials by means of in silico tools: A critical review. *NanoImpact*.

Baas, J., Jager, T. and Kooijman, B., 2010. A review of DEB theory in assessing toxic effects of mixtures. *Science of the Total Environment*, 408(18), pp.3740-3745.

Baas, J., Jager, T. and Kooijman, S.A.L.M., 2009. A model to analyze effects of complex mixtures on survival. *Ecotoxicology and environmental safety*, 72(3), pp.669-676.

Bassingthwaighte, J.B., Butterworth, E., Jardine, B. and Raymond, G.M., 2012. Compartmental modeling in the analysis of biological systems. *Computational toxicology* (pp. 391-438). Humana Press, Totowa, NJ.

Bauch, C., Kolle, S.N., Ramirez, T., Eltze, T., Fabian, E., Mehling, A., Teubner, W., van Ravenzwaay, B. and Landsiedel, R., 2012. Putting the parts together: combining *in vitro* methods to test for skin sensitizing potentials. *Regulatory toxicology and pharmacology*, 63(3), pp.489-504.

Beck, B. and Geppert, T., 2014. Industrial applications of in silico ADMET. *Journal of molecular modeling*, 20(7), p.2322.

Becker, R.A., Ankley, G.T., Edwards, S.W., Kennedy, S.W., Linkov, I., Meek, B., Sachana, M., Segner, H., Van Der Burg, B., Villeneuve, D.L. and Watanabe, H., 2015. Increasing scientific confidence in adverse outcome pathways: application of tailored Bradford-Hill considerations for evaluating weight of evidence. *Regulatory Toxicology and Pharmacology*, 72(3), pp.514-537.

Becker, R.A., Plunkett, L.M., Borzelleca, J.F. and Kaplan, A.M., 2007. Tiered toxicity testing: evaluation of toxicity-based decision triggers for human health hazard characterization. *Food and Chemical Toxicology*, 45(12), pp.2454-2469.

Benigni, R., Battistelli, C.L., Bossa, C., Colafranceschi, M. and Tcheremenskaia, O., 2013. Mutagenicity, carcinogenicity, and other end points. *In Computational Toxicology* (pp. 67-98). Humana Press, Totowa, NJ.

Berggren, E., Amcoff, P., Benigni, R., Blackburn, K., Carney, E., Cronin, M., Deluyker, H., Gautier, F., Judson, R.S., Kass, G.E. and Keller, D., 2015. Chemical safety assessment using read-across: assessing the use of novel testing methods to strengthen the evidence base for decision making. *Environmental health perspectives*, 123(12), pp.1232-1240.

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

Bitsch, A., Jacobi, S., Melber, C., Wahnschaffe, U., Simetska, N. and Mangelsdorf, I., 2006. REPDOSE: a database on repeated dose toxicity studies of commercial chemicals—a multifunctional tool. *Regulatory Toxicology and Pharmacology*, 46(3), pp.202-210.

Bliss, C.I., 1939. The toxicity of poisons applied jointly 1. *Annals of applied biology*, 26(3), pp.585-615.

Bliss, C.I. and Stevens, W.L., 1937. The calculation of the time-mortality curve. *Annals of Applied Biology*, 24(4), pp.815-852.

Bliss, C.I., 1940. The relation between exposure time, concentration and toxicity in experiments on insecticides. *Annals of the Entomological Society of America*, 33(4), pp.721-766.

Bliss, C.I., 1935. The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22(1), pp.134-167.

Brown, K.G. and Foureman, G.L., 2005. Concentration–time–response modeling for acute and short-term exposures. *Regulatory Toxicology and Pharmacology*, 43(1), pp.45-54.

Browne, P., Judson, R.S., Casey, W.M., Kleinstreuer, N.C. and Thomas, R.S., 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. *Environmental science & technology*, 49(14), pp.8804-8814.

Bhumiratana, S., Bernhard, J., Cimetta, E. and Vunjak-Novakovic, G., 2014. Principles of bioreactor design for tissue engineering. *Principles of tissue engineering* (pp. 261-278). Academic Press.

Botham, P.A., 2004. Acute systemic toxicity—prospects for tiered testing strategies. *Toxicology in vitro*, 18(2), pp.227-230.

Burden, N., Aschberger, K., Chaudhry, Q., Clift, M.J., Doak, S.H., Fowler, P., Johnston, H., Landsiedel, R., Rowland, J. and Stone, V., 2017. The 3Rs as a framework to support a 21st century approach for nanosafety assessment. *Nano Today*, 12, pp.10-13.

Carlsen, L., Kenessov, B.N. and Batyrbekova, S.Y., 2008. A QSAR/QSTR study on the environmental health impact by the rocket fuel 1, 1-dimethyl hydrazine and its transformation products. *Environmental health insights*, 1, pp.EHI-S889.

Ciallella, H.L. and Zhu, H., 2019. Advancing computational toxicology in the big data era by artificial intelligence: data-driven and mechanism-driven modelling for chemical toxicity. *Chemical research in toxicology*, 32(4), pp.536-547.

Cipullo, S., Snapir, B., Prpich, G., Campo, P. and Coulon, F., 2019. Prediction of bioavailability and toxicity of complex chemical mixtures through machine learning models. *Chemosphere*, 215, pp.388-395.

Charikar, M., O'Callaghan, L. and Panigrahy, R., 2003, June. Better streaming algorithms for clustering problems. *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing* (pp. 30-39). ACM.

Chiu, W.A., Euling, S.Y., Scott, C.S. and Subramaniam, R.P., 2013. Approaches to advancing quantitative human health risk assessment of environmental chemicals in the post-genomic era. *Toxicology and applied pharmacology*, 271(3), pp.309-323

Chen, Q., Wu, J., Zhang, Y. and Lin, J.M., 2012. Qualitative and quantitative analysis of tumor cell metabolism via stable isotope labeling assisted microfluidic chip electrospray ionization mass spectrometry. *Analytical chemistry*, 84(3), pp.1695-1701.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. and Blaschke, T., 2018. The rise of deep learning in drug discovery. *Drug discovery today*, 23(6), pp.1241-1250.

Chiu, W.A. and White, P., 2006. Steady-state solutions to PBPK models and their applications to risk assessment I: route-to-route extrapolation of volatile chemicals. *Risk Analysis*, 26(3), pp.769-780.

Civjan, N., 2012. Chemical biology: approaches to drug discovery and development to targeting disease. John Wiley & Sons.

Clewell, H.J., Thomas, R.S., Gentry, P.R., Crump, K.S., Kenyon, E.M., El-Masri, H.A. and Yager, J.W., 2007. Research toward the development of a biologically based dose response assessment for inorganic arsenic carcinogenicity: A progress report. *Toxicology and applied pharmacology*, 222(3), pp.388-398.

Cramer, R.D., Patterson, D.E. and Bunce, J.D., 1988. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society, 110(18), pp.5959-5967.

Cronin, M.T., 2011. In silico tools for toxicity prediction. In New Horizons in Predictive Toxicology: Current Status and Application (pp. 9-25). The Royal Society of Chemistry Cambridge.

Cronin, M.T.D. and Madden, J.C., 2010. Using *in silico* toxicity predictions: case studies for skin sensitisation. *In Silico Toxicology* (pp. 606-623).

Cronin, M.T., Madden, J.C., Yang, C. and Worth, A.P., 2019. Unlocking the potential of in silico chemical safety assessment–A report on a cross-sector symposium on current opportunities and future challenges. *Computational Toxicology*, 10, pp.38-43.

Crump, K.S., Chen, C., Chiu, W.A., Louis, T.A., Portier, C.J., Subramaniam, R.P. and White, P.D., 2010. What role for biologically based dose–response models in estimating Low-dose risk?. *Environmental health perspectives*, 118(5), pp.585-588.

Dearden, J.C., 2003. In silico prediction of drug toxicity. *Journal of computer-aided molecular design*, 17(2-4), pp.119-127

Delrue, N., Sachana, M., Sakuratani, Y., Gourmelon, A., Leinala, E. and Diderich, R., 2016. The adverse outcome pathway concept: A basis for developing regulatory decision-making tools. *Alternatives to Laboratory Animals*, 44(5), pp.417-429.

Deeb, O. and Goodarzi, M., 2012. *In silico* quantitative structure toxicity relationship of chemical compounds: Some case studies. *Current drug safety*, 7(4), pp.289-297.

Dimitrov, S. and Mekenyan, O., 2010. An introduction to read-across for the prediction of the effects of chemicals. In Silico Toxicology: Principles and Applications. Cambridge: RSC Publishing, pp.372-384.

Dittrich, P.S. and Manz, A., 2006. Lab-on-a-chip: microfluidics in drug discovery. *Nature Reviews Drug Discovery*, 5(3), p.210

Devillers, J., 2013. Methods for building QSARs. *Computational Toxicology* (pp. 3-27). *Humana Press*, Totowa, NJ.

Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W. and Kavlock, R.J., 2006. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicological Sciences*, 95(1), pp.5-12.

Doe, J.E., Boobis, A.R., Blacker, A., Dellarco, V., Doerrer, N.G., Franklin, C., Goodman, J.I., Kronenberg, J.M., Lewis, R., McConnell, E.E. and Mercier, T., 2006. A tiered approach to systemic toxicity testing for agricultural chemical safety assessment. *Critical reviews in toxicology*, 36(1), pp.37-68.

Dubois, M., Plaisance, H., Thomé, J.P. and Kremers, P., 1996. Hierarchical cluster analysis of environmental pollutants through P450 induction in cultured hepatic cells: Indications for a toxicity screening test. *Ecotoxicology and environmental safety*, 34(3), pp.205-215

Duffin, R., Tran, L., Brown, D., Stone, V. and Donaldson, K., 2007. Proinflammogenic effects of low-toxicity and metal nanoparticles in vivo and in vitro: highlighting the role of particle surface area and surface reactivity. *Inhalation toxicology*, 19(10), pp.849-856.

El-Masri, H., 2013. Modeling for regulatory purposes (risk and safety assessment). In *Computational Toxicology* (pp. 297-303). Humana Press, Totowa, NJ.

Egorov, A.I., Dalbokova, D. and Krzyzanowski, M., 2013. Biomonitoring-based environmental public health indicators. *Computational Toxicology* (pp. 275-293). Humana Press, Totowa, NJ.

Ellison, C.M., Enoch, S.J. and Cronin, M.T., 2011. A review of the use of in silico methods to predict the chemistry of molecular initiating events related to drug toxicity. *Expert opinion on drug metabolism & toxicology*, 7(12), pp.1481-1495

Enoch, S.J., 2010. Chemical category formation and read-across for the prediction of toxicity. *Recent advances in QSAR studies: Methods and applications*, pp.209-219.

Epa, V.C., Burden, F.R., Tassa, C., Weissleder, R., Shaw, S. and Winkler, D.A., 2012. Modeling biological activities of nanoparticles. *Nano letters*, 12(11), pp.5808-5812.

Fantke, P., Aylward, L., Bare, J., Chiu, W.A., Dodson, R., Dwyer, R., Ernstoff, A., Howard, B., Jantunen, M., Jolliet, O. and Judson, R., 2018. Advancements in life cycle human exposure and toxicity characterization. *Environmental health perspectives*, 126(12), p.125001.

Falk-Filipsson, A., Hanberg, A., Victorin, K., Warholm, M. and Wallén, M., 2007. Assessment factors—applications in health risk assessment of chemicals. *Environmental research*, 104(1), pp.108-127.

Fernández-de Gortari, E., García-Jacas, C.R., Martinez-Mayorga, K. and Medina-Franco, J.L., 2017. Database fingerprint (DFP): an approach to represent molecular databases. *Journal of cheminformatics*, 9(1), p.9.

Feynman, R.P., 1960. There's plenty of room at the bottom. California Institute of Technology, Engineering and Science magazine.

Gallegos Saliner, A., Tsakovska, I., Pavan, M., Patlewicz, G. and Worth, A.P., 2007. Evaluation of SARs for the prediction of skin irritation/corrosion potential–structural inclusion rules in the BfR decision support system. SAR and QSAR in Environmental Research, 18(3-4), pp.331-342.

Gajewicz, A., Cronin, M.T., Rasulev, B., Leszczynski, J. and Puzyn, T., 2014. Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: nano-read-across. *Nanotechnology*, 26(1), p.015701.

Gandomi, A. and Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), pp.137-144.

Gentile, F.P., Chiatti, L., Mauro, F., Briganti, G., Floridi, A. and Benassi, M., 1992. Interaction of cytotoxic agents: a rule-based system for computer-assisted cell survival analysis. *Anticancer research*, 12(3), pp.637-643.

George, S., Xia, T., Rallo, R., Zhao, Y., Ji, Z., Lin, S., Wang, X., Zhang, H., France, B., Schoenfeld, D. and Damoiseaux, R., 2011. Use of a high-throughput screening approach coupled with *in vivo* zebrafish embryo screening to develop hazard ranking for engineered nanomaterials. *ACS nano*, 5(3), pp.1805-1817.

Gerner, I., Barratt, M.D., Zinke, S., Schlegel, K. and Schlede, E., 2004. 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. *Alternatives to laboratory animals*, 32(5), pp.487-509.

Gebauer, J.S., Malissek, M., Simon, S., Knauer, S.K., Maskos, M., Stauber, R.H., Peukert, W. and Treuel, L., 2012. Impact of the nanoparticle–protein corona on colloidal stability and protein structure. *Langmuir*, 28(25), pp.9673-9679.

Gershlak, J.R., Hernandez, S., Fontana, G., Perreault, L.R., Hansen, K.J., Larson, S.A., Binder, B.Y., Dolivo, D.M., Yang, T., Dominko, T. and Rolle, M.W., 2017.

Crossing kingdoms: using decellularized plants as perfusable tissue engineering scaffolds. Biomaterials, 125, pp.13-22.

Godfray, H.C.J., Blacquiere, T., Field, L.M., Hails, R.S., Potts, S.G., Raine, N.E., Vanbergen, A.J. and McLean, A.R., 2015. A restatement of recent advances in the natural science evidence base concerning neonicotinoid insecticides and insect pollinators. *Proceedings of the Royal Society B: Biological Sciences*, 282(1818), p.20151821.

Gramatica, P., 2013. On the development and validation of QSAR models. *Computational toxicology* (pp. 499-526). Humana Press, Totowa, NJ.

Grys, B.T., Lo, D.S., Sahin, N., Kraus, O.Z., Morris, Q., Boone, C. and Andrews, B.J., 2017. Machine learning and computer vision approaches for phenotypic profiling. Journal of Cell Biology, 216(1), pp.65-71.

Groh, K.J., Carvalho, R.N., Chipman, J.K., Denslow, N.D., Halder, M., Murphy, C.A., Roelofs, D., Rolaki, A., Schirmer, K. and Watanabe, K.H., 2015. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. *Chemosphere*, 120, pp.764-777.

Guha, R., 2013. On exploring structure–activity relationships. *In Silico Models for Drug Discovery* (pp. 81-94). Humana Press, Totowa, NJ.

Guha, R., 2012. Exploring structure–activity data using the landscape paradigm. Wiley Interdisciplinary Reviews: *Computational Molecular Science*, 2(6), pp.829-841.

Gunčar, G., Kukar, M., Notar, M., Brvar, M., Černelč, P., Notar, M. and Notar, M., 2018. An application of machine learning to haematological diagnosis. *Scientific reports*, 8(1), p.411.

Haber, L.T., Maier, A., Zhao, Q., Dollarhide, J.S., Savage, R.E. and Dourson, M.L., 2001. Applications of mechanistic data in risk assessment: the past, present, and future. *Toxicological sciences*, 61(1), pp.32-39.

Hamelink, J., Landrum, P.F., Bergman, H. and Benson, W.H., 1994. Bioavailability: physical, chemical, and biological interactions. CRC press.

Hadrup, N., Taxvig, C., Pedersen, M., Nellemann, C., Hass, U. and Vinggaard, A.M., 2013. Concentration addition, independent action and generalized concentration addition models for mixture effect prediction of sex hormone synthesis in vitro. *PloS one*, 8(8), p.e70490.

Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, H.K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A. and Rychen, G., 2018. Guidance on risk assessment of the application of nanoscience and nanotechnologies in the food and feed chain: Part 1, human and animal health. *EFSA Journal*, 16(7).

Hartung, T., 2010. Lessons learned from alternative methods and their validation for a new toxicology in the 21st century. *Journal of Toxicology and Environmental Health*, Part B, 13(2-4), pp.277-290.

Hewitt, M., Enoch, S.J., Madden, J.C., Przybylak, K.R. and Cronin, M.T.D., 2013. Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism (s) of action. *Critical reviews in toxicology*, 43(7), pp.537-558.

Hewitt, M., Ellison, C.M., Cronin, M.T., Pastor, M., Steger-Hartmann, T., Munoz-Muriendas, J., Pognan, F. and Madden, J.C., 2015. Ensuring confidence in predictions: a scheme to assess the scientific validity of in silico models. *Advanced drug delivery reviews*, 86, pp.101-111.

Hertzberg, R.P. and Pope, A.J., 2000. High-throughput screening: new technology for the 21st century. *Current opinion in chemical biology*, 4(4), pp.445-451.

Hirota, M., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M. and Toyoda, A., 2015. Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. *Journal of Applied Toxicology*, 35(11), pp.1333-1347.

Howard, G.J. and Webster, T.F., 2009. Generalized concentration addition: a method for examining mixtures containing partial agonists. Journal of theoretical biology, 259(3), pp.469-477.

Houston, J.B. and Galetin, A., 2008. Methods for predicting in vivo pharmacokinetics using data from in vitro assays. *Current drug metabolism*, 9(9), pp.940-951.

Hristozov, D., Gottardo, S., Semenzin, E., Oomen, A., Bos, P., Peijnenburg, W., van Tongeren, M., Nowack, B., Hunt, N., Brunelli, A. and Scott-Fordsmand, J.J., 2016. Frameworks and tools for risk assessment of manufactured nanomaterials. *Environment international*, 95, pp.36-53.

Hong, J., Edel, J.B. and Demello, A.J., 2009. Micro-and nanofluidic systems for high-throughput biological screening. *Drug discovery today*, 14(3-4), pp.134-146.

Hunt, D.L., Rai, S.N. and Li, C.S., 2008. Summary of dose-response modeling for developmental toxicity studies. *Dose-Response*, 6(4), pp.dose-response.

Huh, D., Hamilton, G.A. and Ingber, D.E., 2011. From 3D cell culture to organs-on-chips. *Trends in cell biology*, 21(12), pp.745-754.

Huh, D., Matthews, B.D., Mammoto, A., Montoya-Zavala, M., Hsin, H.Y. and Ingber, D.E., 2010. Reconstituting organ-level lung functions on a chip. *Science*, 328(5986), pp.1662-1668.

Huh, D., Torisawa, Y.S., Hamilton, G.A., Kim, H.J. and Ingber, D.E., 2012. Microengineered physiological biomimicry: organs-on-chips. *Lab on a Chip*, 12(12), pp.2156-2164.

Ingrisch, M. and Sourbron, S., 2013. Tracer-kinetic modeling of dynamic contrast-enhanced MRI and CT: a primer. *Journal of pharmacokinetics and pharmacodynamics*, 40(3), pp.281-300.

Inglese, J., Johnson, R.L., Simeonov, A., Xia, M., Zheng, W., Austin, C.P. and Auld, D.S., 2007. High-throughput screening assays for the identification of chemical probes. *Nature chemical biology*, 3(8), p.466.

Igaki, S., Eguchi, S., Yamagishi, F., Ikeda, H. and Inagaki, T., 1992. Real-time fingerprint sensor using a hologram. *Applied Optics*, 31(11), pp.1794-1802.

Jack, J., Wambaugh, J. and Shah, I., 2013. Systems toxicology from genes to organs. *Computational Toxicology* (pp. 375-397). Humana Press, Totowa, NJ.

Jaspars, M., 1999. Computer assisted structure elucidation of natural products using two-dimensional NMR spectroscopy. *Natural Product Reports*, 16(2), pp.241-248.

Jeliazkova, N., Jaworska, J. and Worth, A., 2010. Open source tools for read-across and category formation. *In Silico Toxicology: Principles and Applications*, pp.408-445.

Judson, R., Richard, A., Dix, D.J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P. and Tan, S., 2008. The toxicity data landscape for environmental chemicals. *Environmental health perspectives*, 117(5), pp.685-695.

Judson, R.S., Magpantay, F.M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., Xia, M., Huang, R., Rotroff, D.M., Filer, D.L. and Houck, K.A., 2015. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicological Sciences*, 148(1), pp.137-154.

Khan, I., Saeed, K. and Khan, I., 2019. Nanoparticles: Properties, applications and toxicities. *Arabian Journal of Chemistry,* 12 (7), pp. 908-931

Kar, S. and Leszczynski, J., 2019. Exploration of computational approaches to predict the toxicity of chemical mixtures. *Toxics*, 7(1), p.15.

Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D. and Richard, A., 2012. Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. *Chemical research in toxicology*, 25(7), pp.1287-1302.

Kleinstreuer, N.C., Ceger, P.C., Allen, D.G., Strickland, J., Chang, X., Hamm, J.T. and Casey, W.M., 2015. A curated database of rodent uterotrophic bioactivity. *Environmental health perspectives*, 124(5), pp.556-562.

Kimber, I., Maxwell, G., Gilmour, N., Dearman, R.J., Friedmann, P.S. and Martin, S.F., 2012. Allergic contact dermatitis: a commentary on the relationship between T lymphocytes and skin sensitising potency. *Toxicology*, 291(1-3), pp.18-24.

Koohy, H., 2017. The rise and fall of machine learning methods in biomedical research. F1000Research, 6.

Kortagere, S. and Ekins, S., 2010. Troubleshooting computational methods in drug discovery. *Journal of pharmacological and toxicological methods*, 61(2), pp.67-75.

Kohonen, T., 1991. Self Organizing maps: Ophmization approaches. *Artificial neural networks* (pp. 981-990). North-Holland.

Knudsen, T.B., Keller, D.A., Sander, M., Carney, E.W., Doerrer, N.G., Eaton, D.L., Fitzpatrick, S.C., Hastings, K.L., Mendrick, D.L., Tice, R.R. and Watkins, P.B., 2015. FutureTox II: *in vitro* data and *in silico* models for predictive toxicology. *Toxicological Sciences*, 143(2), pp.256-267.

Kumar, P., Carrasquer, C.A., Carter, A., Song, Z.H. and Cunningham, A.R., 2014. A categorical structure–activity relationship analysis of GPR119 ligands. *SAR and QSAR in Environmental Research*, 25(11), pp.891-903.

LaLone, C.A., Villeneuve, D.L., Wu-Smart, J., Milsk, R.Y., Sappington, K., Garber, K.V., Housenger, J. and Ankley, G.T., 2017. Weight of evidence evaluation of a network of adverse outcome pathways linking activation of the nicotinic acetylcholine receptor in honey bees to colony death. *Science of The Total Environment*, 584, pp.751-775

Lamon, L., Asturiol, D., Richarz, A., Joossens, E., Graepel, R., Aschberger, K. and Worth, A., 2018. Grouping of nanomaterials to read-across hazard endpoints: from data collection to assessment of the grouping hypothesis by application of chemoinformatic techniques. *Particle and fibre toxicology*, 15(1), p.37.

Lanza, R., Langer, R. and Vacanti, J.P. eds., 2011. *Principles of tissue engineering*. Academic press.

Lin, S., Wang, X., Ji, Z., Chang, C.H., Dong, Y., Meng, H., Liao, Y.P., Wang, M., Song, T.B., Kohan, S. and Xia, T., 2014. Aspect ratio plays a role in the hazard potential of CeO2 nanoparticles in mouse lung and zebrafish gastrointestinal tract. *ACS nano*, 8(5), pp.4450-446

Liu, W., Schmidt, B., Voss, G. and Muller-Wittig, W., 2007. Streaming algorithms for biological sequence alignment on GPUs. *IEEE transactions on parallel and distributed systems*, 18(9), pp.1270-1281.

Lehman, A.J. and Fitzhugh, O.G., 1954. 100-fold margin of safety. Assoc. Food Drug Off. USQ Bull, 18, pp.33-35.

Lee, J.G., Jun, S., Cho, Y.W., Lee, H., Kim, G.B., Seo, J.B. and Kim, N., 2017. Deep learning in medical imaging: general overview. *Korean journal of radiology*, 18(4), pp.570-584.

Lepailleur, A., Poezevara, G. and Bureau, R., 2013. Automated detection of structural alerts (chemical fragments) in (eco) toxicology. *Computational and structural biotechnology journal*, 5(6), p.e201302013.

Li, Y., Chen, X. and Gu, N., 2008. Computational investigation of interaction between nanoparticles and membranes: hydrophobic/hydrophilic effect. *The Journal of Physical Chemistry B*, 112(51), pp.16647-16653.

Liu, J., Patlewicz, G., Williams, A.J., Thomas, R.S. and Shah, I., 2017. Predicting organ toxicity using in vitro bioactivity data and chemical structure. *Chemical research in toxicology*, 30(11), pp.2046-2059.

Liu, R., Rallo, R., George, S., Ji, Z., Nair, S., Nel, A.E. and Cohen, Y., 2011. Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles. *Small*, 7(8), pp.1118-1126.

Liu, R., Zhang, H.Y., Ji, Z.X., Rallo, R., Xia, T., Chang, C.H., Nel, A. and Cohen, Y., 2013. Development of structure–activity relationship for metal oxide nanoparticles. *Nanoscale*, 5(12), pp.5644-5653.

Li, M., Panagi, Z., Avgoustakis, K. and Reineke, J., 2012. Physiologically based pharmacokinetic modeling of PLGA nanoparticles with varied mPEG content. *International journal of nanomedicine*, 7, p.1345.

Loos, M., Schipper, A.M., Schlink, U., Strebel, K. and Ragas, A.M., 2010. Receptor-oriented approaches in wildlife and human exposure modelling: a comparative study. *Environmental Modelling & Software*, 25(4), pp.369-382.

Løkke, H., Ragas, A.M. and Holmstrup, M., 2013. Tools and perspectives for assessing chemical mixtures and multiple stressors. *Toxicology*, 313(2-3), pp.73-82.

Løkke, H., 2010. Novel methods for integrated risk assessment of cumulative stressors—Results from the NoMiracle project. *Science of the Total Environment*, 408(18):3719-24

Luechtefeld, T., Rowlands, C. and Hartung, T., 2018. Big-data and machine learning to revamp computational toxicology and its use in risk assessment. *Toxicology research*, 7(5), pp.732-744.

Lynch, I., Weiss, C. and Valsami-Jones, E., 2014. A strategy for grouping of nanomaterials based on key physico-chemical descriptors as a basis for safer-by-design NMs. *Nano Today*, 9(3), pp.266-270.

MacKay, C., Davies, M., Summerfield, V. and Maxwell, G., 2013. From pathways to people: applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. *ALTEX-Alternatives to animal experimentation*, 30(4), pp.473-486.

Mager, D.E., Wyska, E. and Jusko, W.J., 2003. Diversity of mechanism-based pharmacodynamic models. *Drug Metabolism and Disposition*, 31(5), pp.510-518.

Mak, K.K. and Pichika, M.R., 2018. Artificial intelligence in drug development: present status and future prospects. *Drug discovery today,* 24 (3), pp. 773-778.

Martin, O.V., Martin, S. and Kortenkamp, A., 2013. Dispelling urban myths about default uncertainty factors in chemical risk assessment–sufficient protection against mixture effects? *Environmental health*, 12(1), p.53.

Macarron, R., 2006. Critical review of the role of HTS in drug discovery. *Drug discovery today*, 7(11), pp.277-279.

Macarrón, R. and Hertzberg, R.P., 2011. Design and implementation of high throughput screening assays. *Molecular biotechnology*, 47(3), pp.270-285.

Maxwell, G., MacKay, C., Cubberley, R., Davies, M., Gellatly, N., Glavin, S., Gouin, T., Jacquoilleot, S., Moore, C., Pendlington, R. and Saib, O., 2014. Applying the skin sensitisation adverse outcome pathway (AOP) to quantitative risk assessment. *Toxicology In Vitro*, 28(1), pp.8-12.

Mech, A., Rasmussen, K., Jantunen, P., Aicher, L., Alessandrelli, M., Bernauer, U., Bleeker, E.A.J., Bouillard, J., Di Prospero Fanghella, P., Draisci, R. and Dusinska, M., 2019. Insights into possibilities for grouping and read-across for nanomaterials in EU chemicals legislation. *Nanotoxicology*, 13(1), pp.119-141.

Melagraki, G. and Afantitis, A., 2014. Enalos InSilicoNano platform: an online decision support tool for the design and virtual screening of nanoparticles. *RSC Advances*, 4(92), pp.50713-50725.

Mikolajczyk, A., Gajewicz, A., Mulkiewicz, E., Rasulev, B., Marchelek, M., Diak, M., Hirano, S., Zaleska-Medynska, A. and Puzyn, T., 2018. Nano-QSAR modeling for ecosafe design of heterogeneous TiO 2-based nano-photocatalysts*. Environmental Science: Nano*, 5(5), pp.1150-1160

Mirshafiee, V., Osborne, O.J., Sun, B. and Xia, T., 2018. Safety Concerns of Industrial Engineered Nanomaterials. Handbook of Nanomaterials for Industrial Applications (pp. 1063-1072). Elsevier.

Modi, S., Hughes, M., Garrow, A. and White, A., 2012. The value of *in silico* chemistry in the safety assessment of chemicals in the consumer goods and pharmaceutical industries. *Drug discovery today*, 17(3-4), pp.135-142.

Monteiller, C., Tran, L., MacNee, W., Faux, S., Jones, A., Miller, B. and Donaldson, K., 2007. The pro-inflammatory effects of low-toxicity low-solubility particles, nanoparticles and fine particles, on epithelial cells *in vitro*: the role of surface area. *Occupational and environmental medicine*, 64(9), pp.609-615.

Milan, C., Schifanella, O., Roncaglioni, A. and Benfenati, E., 2011. Comparison and possible use of *in silico* tools for carcinogenicity within REACH legislation. *Journal of Environmental Science and Health*, Part C, 29(4), pp.300-323.

Miller, F.J., Schlosser, P.M. and Janszen, D.B., 2000. 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. *Toxicology*, 149(1), pp.21-34.

Netzeva, T.I., Pavan, M. and Worth, A.P., 2008. Review of (quantitative) structure–activity relationships for acute aquatic toxicity. *QSAR & Combinatorial Science*, 27(1), pp.77-9

Nel, A., Xia, T., Meng, H., Wang, X., Lin, S., Ji, Z. and Zhang, H., 2012. Nanomaterial toxicity testing in the 21st century: use of a predictive toxicological approach and high-throughput screening. *Accounts of chemical research*, 46(3), pp.607-621

Neužil, P., Giselbrecht, S., Länge, K., Huang, T.J. and Manz, A., 2012. Revisiting lab-on-a-chip technology for drug discovery. Nature reviews Drug discovery, 11(8), p.620.

Oksel, C., Ma, C.Y., Liu, J.J., Wilkins, T. and Wang, X.Z., 2015. (Q) SAR modelling of nanomaterial toxicity: A critical review. *Particuology*, 21, pp.1-19.

Oja, E. and Kaski, S. eds., 1999. Kohonen maps. Elsevier.

Osborne, O.J., Lin, S., Chang, C.H., Ji, Z., Yu, X., Wang, X., Lin, S., Xia, T. and Nel, A.E., 2015. Organ-specific and size-dependent Ag nanoparticle toxicity in gills and intestines of adult zebrafish. *ACS nano,* 9(10), pp.9573-9584.

Osborne, O.J., Lin, S., Jiang, W., Chow, J., Chang, C.H., Ji, Z., Yu, X., Lin, S., Xia, T. and Nel, A.E., 2017. Differential effect of micron-versus nanoscale III–V particulates and ionic species on the zebrafish gut. *Environmental Science: Nano*, 4(6), pp.1350-1364.

Parasuraman, S., 2011. Prediction of activity spectra for substances. *Journal of pharmacology & pharmacotherapeutics*, 2(1), p.52.

Pasha, F.A., Neaz, M.M., Cho, S.J., Ansari, M., Mishra, S.K. and Tiwari, S., 2009. In silico quantitative structure–toxicity relationship study of aromatic nitro compounds. *Chemical biology & drug design*, 73(5), pp.537-544.

Patlewicz, G., Simon, T.W., Rowlands, J.C., Budinsky, R.A. and Becker, R.A., 2015. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. *Regulatory Toxicology and Pharmacology*, 71(3), pp.463-477.

Pelaz, B., del Pino, P., Maffre, P., Hartmann, R., Gallego, M., Rivera-Fernandez, S., de la Fuente, J.M., Nienhaus, G.U. and Parak, W.J., 2015. Surface functionalization of nanoparticles with polyethylene glycol: effects on protein adsorption and cellular uptake. *ACS nano*, 9(7), pp.6996-7008.

Perkins, E.J., Antczak, P., Burgoon, L., Falciani, F., Garcia-Reyero, N., Gutsell, S., Hodges, G., Kienzler, A., Knapen, D., McBride, M. and Willett, C., 2015. Adverse outcome pathways for regulatory applications: examination of four case studies with different degrees of completeness and scientific confidence. *Toxicological Sciences*, 148(1), pp.14-25.

Pittinger, C. and Mohapatra, A., 2009. Software Tools for Toxicology and Risk Assessment. *Information Resources in Toxicology* (pp. 631-638). Academic Press.

Piñero, J., Furlong, L.I. and Sanz, F., 2018. In silico models in drug development: where we are. *Current opinion in pharmacology*, 42, pp.111-121.

Pottenger, L.H., Schisler, M.R., Zhang, F., Bartels, M.J., Fontaine, D.D., McFadden, L.G. and Gollapudi, B.B., 2009. Dose–response and operational thresholds/NOAELs for in vitro mutagenic effects from DNA-reactive mutagens, MMS and MNU. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 678(2), pp.138-147.

Powers, C.M., Mills, K.A., Morris, S.A., Klaessig, F., Gaheen, S., Lewinski, N. and Hendren, C.O., 2015. Nanocuration workflows: Establishing best practices for identifying, inputting, and sharing data to inform decisions on nanomaterials. *Beilstein journal of nanotechnology*, 6(1), pp.1860-1871.

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T.P., Michalkova, A., Hwang, H.M., Toropov, A., Leszczynska, D. and Leszczynski, J., 2011. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature nanotechnology*, 6(3), p.175.

Pu, L., Naderi, M., Liu, T., Wu, H.C., Mukhopadhyay, S. and Brylinski, M., 2019. e ToxPred: a machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacology and Toxicology*, 20(1), p.2.

Richard, A.M., Judson, R.S., Houck, K.A., Grulke, C.M., Volarath, P., Thillainadarajah, I., Yang, C., Rathman, J., Martin, M.T., Wambaugh, J.F. and Knudsen, T.B., 2016. ToxCast chemical landscape: paving the road to 21st century toxicology. *Chemical research in toxicology*, 29(8), pp.1225-1251.

Richard, A.M. and Williams, C.R., 2002. Distributed structure-searchable toxicity (DSSTox) public database network: a proposal. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 499(1), pp.27-52.

Relyea, R.A., 2009. A cocktail of contaminants: how mixtures of pesticides at low concentrations affect aquatic communities. *Oecologia*, 159(2), pp.363-376.

Sarigiannis, D.A. and Hansen, U., 2012. Considering the cumulative risk of mixtures of chemicals–a challenge for policy makers. *Environmental Health*, 11(S1), p.S18.

Sanz, F., Pognan, F., Steger-Hartmann, T., Díaz, C., Cases, M., Pastor, M., Marc, P., Wichard, J., Briggs, K., Watson, D.K. and Kleinöder, T., 2017. Legacy data sharing to improve drug safety assessment: the eTOX project. Nature reviews Drug discovery, 16(12), p.811.

Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L. and Nolan, G.P., 2011. Cloud and heterogeneous computing solutions exist today for the emerging big data problems in biology. *Nature Reviews Genetics*, 12(3), p.224.

Schupp, T., Georg, P.A. and Kirstein, G., 2017. Toxicological risk at workplace and toxicity as life cycle assessment impact category: substitution of solvents as an example. *EXCLI journal*, 16, p.40.

Slikker Jr, W., de Souza Lima, T.A., Archella, D., de Silva Junior, J.B., Barton-Maclaren, T., Bo, L., Buvinich, D., Chaudhry, Q., Chuan, P., Deluyker, H. and Domselaar, G., 2018. Emerging technologies for food and drug safety. *Regulatory Toxicology and Pharmacology*, 98, pp.115-128.

Silva, T., Pokhrel, L.R., Dubey, B., Tolaymat, T.M., Maier, K.J. and Liu, X., 2014. Particle size, surface charge and concentration dependent ecotoxicity of three organo-coated silver nanoparticles: comparison between general linear model-predicted and observed toxicity. *Science of the Total Environment*, 468, pp.968-976.

Sturla, S.J., Boobis, A.R., FitzGerald, R.E., Hoeng, J., Kavlock, R.J., Schirmer, K., Whelan, M., Wilks, M.F. and Peitsch, M.C., 2014. Systems toxicology: from basic research to risk assessment. *Chemical research in toxicology*, 27(3), pp.314-329.

Sung, J.H., Srinivasan, B., Esch, M.B., McLamb, W.T., Bernabini, C., Shuler, M.L. and Hickman, J.J., 2014. Using physiologically-based pharmacokinetic-guided "body-on-a-chip" systems to predict mammalian response to drug and chemical exposure. *Experimental biology and medicine*, 239(9), pp.1225-1239.

Thakur, A. and Thakur, M., 2009. Chem-bioinformatics: QSTR studies on TIBO derivatives. *Journal of biochemical and molecular toxicology*, 23(3), pp.166-171.

Tollefsen, K.E., Scholz, S., Cronin, M.T., Edwards, S.W., de Knecht, J., Crofton, K., Garcia-Reyero, N., Hartung, T., Worth, A. and Patlewicz, G., 2014. Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). *Regulatory Toxicology and Pharmacology*, 70(3), pp.629-640.

Toropov, A.A., Toropova, A.P., Raska Jr, I., Leszczynska, D. and Leszczynski, J., 2014. Comprehension of drug toxicity: Software and databases. *Computers in biology and medicine*, 45, pp.20-25

Turley, A.E., Isaacs, K.K., Wetmore, B.A., Karmaus, A.L., Embry, M.R. and Krishan, M., 2019. Incorporating new approach methodologies in toxicity testing and exposure assessment for tiered risk assessment using the RISK21 approach: Case studies on food contact chemicals. *Food and Chemical Toxicology*, 134, p.110819.

Tung, Y.C., Hsiao, A.Y., Allen, S.G., Torisawa, Y.S., Ho, M. and Takayama, S., 2011. High-throughput 3D spheroid culture and drug testing using a 384 hanging drop array. *Analyst*, 136(3), pp.473-478.

Roncaglioni, A., Toropov, A.A., Toropova, A.P. and Benfenati, E., 2013. In silico methods to predict drug toxicity. *Current opinion in pharmacology*, 13(5), pp.802-806.

Rudén, C. and Hansson, S.O., 2009. Registration, evaluation, and authorization of chemicals (REACH) is but the first step–how far will it take us? Six further steps to improve the European chemicals legislation. *Environmental health perspectives*, 118(1), pp.6-10.

Valerio Jr, L.G., 2009. In silico toxicology for the pharmaceutical sciences. *Toxicology and applied pharmacology*, 241(3), pp.356-370.

Venkatapathy, R. and Wang, N.C.Y., 2013. Developmental toxicity prediction. *In Computational toxicology* (pp. 305-340). Humana Press, Totowa, NJ.

Vink, S.R., Mikkers, J., Bouwman, T., Marquart, H. and Kroese, E.D., 2010. Use of read-across and tiered exposure assessment in risk assessment under REACH–A case study on a phase-in substance. *Regulatory Toxicology and Pharmacology*, 58(1), pp.64-71.

Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J. and Bryant, S.H., 2009. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic acids research*, 37, pp.623-633.

Waters, M., Stasiewicz, S., Alex Merrick, B., Tomer, K., Bushel, P., Paules, R., Stegman, N., Nehls, G., Yost, K.J., Johnson, C.H. and Gustafson, S.F., 2007. CEBS—Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. *Nucleic acids research*, 36, pp.D892-D900.

Weaver, S. and Gleeson, M.P., 2008. The importance of the domain of applicability in QSAR modeling. Journal of Molecular Graphics and Modelling, 26(8), pp.1315-1326.

Wen, M., Zhang, Z., Niu, S., Sha, H., Yang, R., Yun, Y. and Lu, H., 2017. Deep-learning-based drug–target interaction prediction. *Journal of proteome research*, 16(4), pp.1401-1409.

Wittwehr, C., Aladjov, H., Ankley, G., Byrne, H.J., de Knecht, J., Heinzle, E., Klambauer, G., Landesmann, B., Luijten, M., MacKay, C. and Maxwell, G., 2017. How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. *Toxicological Sciences*, 155(2), pp.326-336.

Winkler, D.A., Mombelli, E., Pietroiusti, A., Tran, L., Worth, A., Fadeel, B. and McCall, M.J., 2013. Applying quantitative structure–activity relationship approaches to nanotoxicology: current status and future potential. *Toxicology*, 313(1), pp.15-23.

Worth, A.P., Lapenna, S. and Serafimova, R., 2013. QSAR and metabolic assessment tools in the assessment of genotoxicity. *Computational Toxicology* pp. 125-162 Humana Press, Totowa, NJ.

Worth, A.P., 2019. The future of in silico chemical safety… and beyond. *Computational Toxicology*, (10) pp 60-62

Wu, G. and Doberstein, S.K., 2006. HTS technologies in biopharmaceutical discovery. *Drug Discovery Today*, 11(15-16), pp.718-724.

Xu, Y., Pei, J. and Lai, L., 2017. Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction. *Journal of chemical information and modeling*, 57(11), pp.2672-2685.

Ye, N., Qin, J., Shi, W., Liu, X. and Lin, B., 2007. Cell-based high content screening using an integrated microfluidic device. *Lab on a Chip*, 7(12), pp.1696-1704.

Young, D., Martin, T., Venkatapathy, R. and Harten, P., 2008. Are the chemical structures in your QSAR correct? *QSAR & combinatorial science*, 27(11-12), pp.1337-1345.

Young, J.D., Cai, C. and Lu, X., 2017. Unsupervised deep learning reveals prognostically relevant subtypes of glioblastoma. *BMC bioinformatics*, 18(11), p.381.

Yu, F. and Choudhury, D., 2019. Microfluidic bioprinting for organ-on-a-chip models. *Drug discovery today*.

Qin, J., Ye, N., Liu, X. and Lin, B., 2005. Microfluidic devices for the analysis of apoptosis. *Electrophoresis*, 26(19), pp.3780-3788.

Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. 2012. Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environmental health perspectives,* 19; 121(1):23-31.

Zhang, J.H., Chung, T.D. and Oldenburg, K.R., 2000. Confirmation of primary active substances from high throughput screening of chemical and biological populations: a statistical approach and practical considerations. *Journal of combinatorial chemistry*, 2(3), pp.258-265.

Zhong, Y., Guo, Z. and Che, J., 2015. Data Quality Assurance and Statistical Analysis of High Throughput Screenings for Drug Discovery. *Frontiers in Computational Chemistry (*pp. 389-425). Bentham Science Publishers.

Zhu, H., 2013. From QSAR to QSIIR: searching for enhanced computational toxicology models. *Computational Toxicology* (pp. 53-65). Humana Press, Totowa, NJ.

Zhu, H., Zhang, J., Kim, M.T., Boison, A., Sedykh, A. and Moran, K., 2014. Big data in chemical toxicity research: the use of high-throughput screening assays to identify potential toxicants. *Chemical research in toxicology*, 27(10), pp.1643-1651.

Abbreviations

4IR 4th industrial revolution
AI artificial intelligence
ADME Adsorption Distribution Metabolism and Excretion
ACSA Agricultural Chemical Safety Assessment
AN analogue approach
AO adverse outcome
AOPs adverse outcome pathways
AOP-KB adverse outcome pathway knowledge base
AOP adverse outcome pathways
AP as-prepared
BMDL benchmark dose level
BSA bovine serum albumin
Ca concentration addition
CA category approach
CCL carbon chain length
COT Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment
CoMoCAT Co-Mo catalyst
DL deep learning
DIP data integration procedures
ERα oestrogen receptor-alpha
ECNs engineered carbonaceous nanomaterials
ECHA European Chemicals Agency
ECVAM European Centre for the Validation of Alternative Methods
EDSP Endocrine Disruption Screening Program
EPA Environmental Protection Agency
EU European Union
FDA Food and Drug Administration
FSA Food Standards Agency
GCA generalized concentration addition
GO graphene oxide
GSRS17 7th annual Global Summit on Regulatory Science
HBRs human-based rules
HCS high content screening
HTS High throughput screening
HPT hypothalamic-pituitary-thyroidal
HPV high production volume
HepG2 human liver carcinoma
IA independent action
IATAs Integrated approaches to testing and assessment
IBRs induction-based rules
IL-1β interleukin-1β
IVIVE in-vitro in-vivo extrapolation
JRC Joint Research Centre

KE key event
KERs Key Event Relationships
LOAEL least observable adverse effect level
MIE molecular initiating event
ML machine learning
MLR multiple linear regression
MFs modifying factors'
MoA mode of action
MWCNTs multiwall carbon nanotubes
NAMs New Approach Methodologies
NMs nanomaterials
NFs nanoforms
NCATS National Center for Advancing Translational Sciences
NIEHS National Toxicology Program at the National Institute of Environmental Health Sciences
NIH National Institutes of Health
NN neural networks
NPs nanoparticles
NRC National Research Council
NC3Rs National Centre for the Replacement, Refinement and Reduction of Animals in Research
NOAEL no observed adverse effect levels
NTP National Toxicology Program
OECD Organisation for Economic Co-operation and Development
PLS partial least squares
PCR Polymerase Chain Reaction
PD pharmacodynamic
PBPK physiologically based pharmacokinetic
PK pharmacokinetic
RfD reference dose
RfC reference concentration
REACH Registration, Evaluation, Authorisation and Restriction of Chemicals
QSAR Quantitative Structure Activity Relationships
QSTR Quantitative structure activity relationship in toxicology
QSPR Quantitative structure property relationship
SALI structure activity landscape index
SARMs selective androgen receptor modulators
SAR structure activity relationship
SARI structure activity relationship index
SAs structural alerts
SIDS Screening Information Data Set
SWCNTs single-wall carbon nanotubes
SVM support vector machine
SPE solid phase extraction
SFs safety factors
TiO_2 titanium dioxide

TK	toxicokinetic
TGF-β1	transforming growth factor-β1
Tox 21	21st century toxicology
ToxCast	Toxicity Forecaster
TSCA	Toxic Substances Control Act
UFs	uncertainty factors
UK	United Kingdom
WoE	weight of evidence