Integrated Environmental Assessment and Management, 2026, **00**, 1–11
https://doi.org/10.1093/inteam/vjag067
Published: April 28, 2026
Special Series

A framework for chemical hazard assessments under "Safe and Sustainable by Design" using multiple in silico tools

Joanke van Dijk[1,2,*], Anna Agalliadou[3,4], Chiara Laura Battistelli[5], Emilio Benfenati[6], Cecilia Bossa[5], Maja Halling[7], Spyros Karakitsios[3,4], Achilleas Karakoltzidis[3,4], Fotini Nikiforou[3,4], Dimosthenis A. Sarigiannis[3,4,8,9], Gianluca Selvestrel[6], Anežka Sharma[10], Kirsi M. Siivola[11], and Ziye Zheng[7]

[1]Technology and Societa Laboratory, Empa—Swiss Federal Laboratories for Material Science and Technology, St. Gallen, Switzerland
[2]TEMAS Solutions GmbH, Hausen, Switzerland
[3]Department of Chemical Engineering, Environmental Engineering Laboratory, Aristotle University of Thessaloniki, Thessaloniki, Greece
[4]Center for Interdisciplinary Research and Innovation, HERACLES Research Center on the Exposome and Health, Thessaloniki, Greece
[5]Environment and Health Department, Istituto Superiore di Sanità, Rome, Italy
[6]Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milano, Italy
[7]IVL, Swedish Environmental Research Institute, Stockholm, Sweden
[8]National Hellenic Research Foundation, Athens, Greece
[9]School for Advanced Study (IUSS), Science, Technology and Society Department, Environmental Health Engineering, Pavia, Italy
[10]Recetox, Faculty of Science, Masaryk University, Brno, Czech Republic
[11]Finnish Institute of Occupational Health, Helsinki, Finland
*Corresponding author: Joanke van Dijk, Email: joanke@temasol.ch

Editor's Note: This article is part of the special series "Getting New Approach Methods (NAMs) into Environmental Assessment, Management, and Practice." The series presents a collection of articles on how NAMs are being developed to meet adopter needs and inform regulation, policy, and/or decision making, as well as their application in environment risk (and/or impact) assessment and management.

Abstract

The hazard identification of chemicals is a key step of the "Safe and Sustainable by Design" (SSbD) framework introduced by the European Commission, aiming to eliminate hazardous substances early in innovation. In this context, in silico methods such as (Quantitative) Structure–Activity Relationship ((Q)SAR) models offer rapid, cost-effective, and animal-free alternatives for early-stage hazard screening. The Partnership for the Assessment of Risks from Chemicals (PARC) is developing a toolbox to facilitate SSbD assessments containing numerous (Q)SAR models. Challenges, however, exist in using and combining multiple in silico tools. Here, we developed a workflow to assess chemical hazards using multiple in silico tools within the PARC toolbox. The workflow consists of three phases: (1) the preparation stage, (2) running the models, and (3) the evaluation stage. To demonstrate the approach, we applied it to a case study comparing bisphenol A, isosorbide, and bisphenol AP. Tools from the PARC toolbox were screened for relevance, transparency, and open access availability. Only models aligned with SSbD-required endpoints and adequately documented via (Q)SAR Model Reporting Formats were retained. The properties assessed in this study cover carcinogenicity, germ cell mutagenicity, reproductive toxicity, endocrine disruption, persistence, bioaccumulation, and aquatic toxicity. Predictions were filtered using applicability domain criteria and reliability scores. Next, three strategies were applied for integrating different model outputs. Model agreement varied across endpoints and integration methods. This emphasizes the possibility of different SSbD assessment outcomes and thus the need for transparent documentation of the chosen strategy and explicit handling of uncertainty. Our study demonstrates how multiple models can systematically and transparently be integrated via the developed workflow. Key areas for improvement are to refine integration strategies, harmonize the definition and communication of applicability domains across tools, expand in silico coverage for currently underrepresented endpoints, and develop approaches to consider data gaps in SSbD assessments.

Keywords: Safe and Sustainable by Design; in silico tools; (Q)SARs; hazard assessment; uncertainty

Received: September 1, 2025. **Revised:** April 7, 2026. **Accepted:** April 20, 2026

Introduction

The increasing number and volume of chemicals and materials released into the environment are a growing concern for both human health and the environment (Almroth et al., 2022; UNEP, 2019). To better address this issue, the "Chemicals Strategy for Sustainability" has been published in Europe (European Commission, 2020). A key component of this strategy is the concept of Safe and Sustainable by Design (SSbD), which promotes the integration of safety and sustainability considerations as early as possible in the innovation process and throughout the entire product lifecycle.

Safe and Sustainable by Design is intended to guide the development of chemicals and materials through iterative design and assessment cycles. In this framework, potential hazards should be identified at early stages of development so that problematic substances can be avoided or replaced before large investments are made in production or commercialization (Caldeira et al., 2022; Garmendia Aguirre et al., 2025). While minimizing intrinsic hazards is a central element of SSbD, effective chemicals management also relies on risk-based decision making that considers both hazard and exposure. Hazard identification, therefore, represents an early screening step within a broader decision-making framework that ultimately integrates exposure assessment and risk characterization.

Within the SSbD framework, hazard assessment focuses particularly on properties that are considered critical under European chemical regulations, such as the Classification, Labelling and Packaging (CLP) and the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). The "most hazardous substances"—which include substances that are carcinogenic, mutagenic, and/or reprotoxic (CMR), persistent, bioaccumulative, and toxic (PBT), and endocrine disrupting (ED)—should be filtered out early in the process as their use cannot be considered "safe and sustainable."

Traditionally, hazard assessments and further risk assessments required under CLP and REACH involve collecting experimental data (ECHA, 2011). At early stages of innovation, however, experimental data are often scarce because new substances have not yet undergone extensive testing. Under such conditions, predictive approaches are required to identify potential hazards and guide early design decisions. In silico methods, including (Quantitative) Structure–Activity Relationship ((Q)SAR) models, provide a promising solution for this purpose. (Quantitative) Structure–Activity Relationship models predict chemical properties and biological effects based on molecular structure. Although these models remain underused in current risk assessments (Thomas et al., 2019), (Q)SAR models are particularly well suited to identify areas of potential concern in the early innovation stages targeted by SSbD, since they are rapid, cost-effective, and do not require animal testing (Garmendia Aguirre et al., 2025).

Despite their advantages, the use of (Q)SAR models raises several methodological challenges. First, individual models often have limited applicability domains, meaning that predictions may only be reliable for certain types of chemicals. Second, multiple models may exist for the same hazard endpoint, which can result in conflicting predictions. Third, authorities are often promoting the use of integrated in silico models since this increases the confidence of the results (EMA, 2023), while integrating results from different models into a coherent interpretation requires transparent decision rules and careful consideration of uncertainty. Existing studies and guidance documents, such as general guidance on using in silico models (Benfenati et al., 2019), the studies and guidance on weight-of-evidence (WoE) approaches (Caldeira et al., 2023; EFSA Scientific Committee et al., 2017; Leder et al., 2015; Lorenz et al., 2021; van Dijk et al., 2022), and the OECD (Q)SAR Assessment Framework (QAF; OECD, 2023a, 2024), provide principles for evaluating and combining lines of evidence. However, practical workflows for applying multiple in silico tools within the SSbD context remain limited. Furthermore, interpreting predictions from multiple models requires expert judgment. Agreement between models does not necessarily guarantee that predictions are reliable, and experienced toxicologists and environmental scientists must evaluate model outputs in the context of modeling information (e.g., chemical structure, applicability domain (AD) of the models, and prediction uncertainties), biological mechanisms, and regulatory endpoints (Sarigiannis et al., 2023).

In this study, we therefore develop a workflow for conducting hazard assessments using multiple in silico tools in the context of SSbD. The workflow is designed to support transparent model selection, reliability evaluation, and integration of predictions. To illustrate its application, the workflow is applied to a case study involving bisphenol A (BPA), bisphenol AP (BPAP), and isosorbide, representing a substitution scenario relevant to SSbD assessments. The aims of this study are therefore to:

1) Develop a structured workflow for applying multiple in silico hazard prediction tools identified under the Partnership for the Assessment of Risks from Chemicals (PARC) (Marx-Stoelting et al., 2023) within the SSbD context and included in the PARC SSbD toolbox (Sarigiannis et al., 2024);
2) Demonstrate the workflow through a case study involving BPA and two potential alternative substances;
3) Explore strategies for evaluating, integrating, and communicating model outputs while explicitly addressing uncertainty based on previously published literature and the QAF (Benfenati et al., 2019; EFSA Scientific Committee et al., 2017).

The study focuses specifically on (Q)SAR-based approaches, including both traditional descriptor-based models and machine learning-driven models, as these represent some of the most mature and widely accepted in silico methods for hazard prediction. While (Q)SAR predictions cannot replace experimental data or expert evaluation, they can play a central role in early-stage screening and prioritization within SSbD assessments.

Methods

Framework design

To support the use of multiple in silico models in SSbD hazard assessments, a structured workflow was developed in this study. The workflow was based on the QAF and published documents outlining how uncertainty in (Q)SAR predictions can be

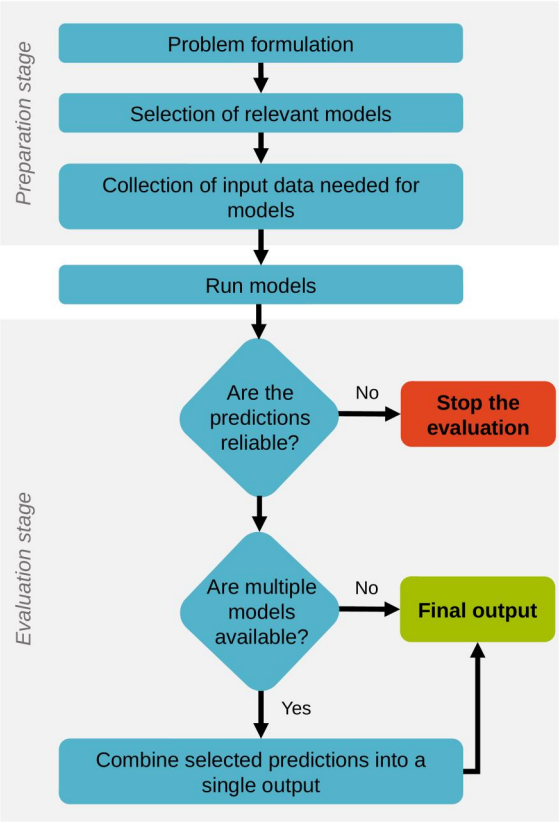

Figure 1 Developed workflow for the assessment of hazards using in silico tools based on Benfenati et al. (2019) and EFSA Scientific Committee (2017).

considered and how their results can be integrated (Benfenati et al., 2019; EFSA Scientific Committee et al., 2017; OECD, 2024).

The workflow aims to guide users through the selection, evaluation, and integration of predictions from multiple in silico tools in a transparent and reproducible manner. The workflow consists of three phases:

- Preparation stage, in which the assessment context (e.g., target chemicals) is defined, relevant hazard endpoints are identified, and suitable in silico models are selected;
- Model execution stage, in which the selected models are run and prediction outputs are collected;
- Evaluation stage, in which the reliability of model predictions is assessed and results from different models are integrated.

An overview of the workflow developed and applied in this study is presented in Figure 1.

Problem formulation

The present study demonstrates the workflow based on a case study conducted within the PARC project, for which the chemical bisphenol A (BPA, CAS: 80-05-7) and two potential alternative substances, bisphenol AP (BPAP, CAS: 1571-75-1) and isosorbide (CAS: 652-67-5), were selected to test a range of tools for SSbD assessments (PARC et al., 2024).

The selected substances represent a typical substitution scenario in which a well-known substance with recognized hazard concerns (BPA) is compared with potential alternative substances. Bisphenol A is widely studied and has known regulatory concerns, while BPAP and isosorbide represent structurally different alternatives. This combination allows the workflow to be tested across substances with different chemical structures and different levels of available background knowledge.

The hazard assessment focused on endpoints relevant to identifying substances of concern under the SSbD framework. These include properties associated with CMR effects, properties determining the environmental fate and behavior for PBT classification, and ED-related properties. The CMR and PBT properties are used to identify substances of very high concern and must be assessed under the SSbD framework to prevent the use of the most hazardous substances (Caldeira et al., 2022; Garmendia Aguirre et al., 2025), and ED are the recent-included endpoints in the CLP regulation.

Selection of relevant (Q)SAR models

A wide range of in silico tools are available to predict chemical hazards, and different platforms may provide models for different endpoints. Within the PARC project, an inventory of tools relevant for SSbD hazard assessments was previously compiled (Sarigiannis et al., 2023). From this inventory, models were selected based on three criteria:

- Relevance to hazard endpoints required in SSbD assessments,
- Availability through open-access platforms,
- Adequate documentation through (Q)SAR Model Reporting Formats (QMRF).

An overview of the platforms used in this study is presented in Table 1. These include VEGA (Benfenati et al., 2013), JANUS, EPI Suite (USEPA, 2012), the Danish (Q)SAR Database (DTU, n.d.), the Mistra SafeChem in silico toolbox (not openly available at time of publication), and the OECD QSAR Toolbox (OECD, 2023b). Not all platforms provide models for all hazard endpoints. Therefore, multiple platforms were used to ensure sufficient coverage across the endpoints considered in the assessment. The QMRFs were analyzed to filter out models that were developed for endpoints that are irrelevant for the SSbD assessment (such as anaerobic degradation) and avoid double-inclusion (for example, some models are included in both the Danish (Q)SAR database and OECD QSAR Toolbox). An overview of the different platforms used, and endpoints predicted, is presented in Table 1, and a list of all the individual models used in this work, and the reasoning why to include or exclude them, is presented in the online supplementary material. In some cases, platforms provide consensus models that integrate predictions from several individual models. In such cases, we used the outputs of the consensus model rather than those of the individual models to avoid double-counting predictions.

Running the models and collecting predictions

For each substance, predictions were generated using the selected models across all relevant hazard endpoints. The types of outputs produced by the models vary depending on the

Table 1 An overview of the different platforms used in this study to predict the studied hazard endpoints.

(Q)SAR platform	Carcinogenicity	Mutagenicity	Reproductive toxicity	Persistence	Bioaccumulation	Endocrine disruptions	Aquatic toxicity
VEGA 1.2.3	Yes	Yes	Yes	Yes	Yes	Yes	Yes
JANUS 1.0.3	Yes	Yes	Yes	Yes	Yes	No	No
EPI Suite v4.11	No	No	No	Yes	Yes	No	Yes
Mistra SafeChem in silico toolbox	Yes	Yes	Yes	Yes	Yes	Yes	No
Danish (Q)SAR Database	Yes	Yes	Yes	Yes	No	Yes[1]	Yes
QSAR Toolbox v4.6	No	No	Yes	No	Yes	Yes	Yes

[1]Included and run in the QSAR Toolbox v4.6.
((Q)SAR) = (Quantitative) Structure–Activity Relationship.
A list of all the individual models used in this work is available in the online supplementary material.

endpoint and modeling approach. Some models provide quantitative predictions (e.g., toxicity values), while others produce categorical outputs (e.g., positive or negative predictions for a specific hazard class). Detailed information for each model can be found in the online supplementary material.

Predictions from different models were first grouped according to the hazard endpoint they addressed. This grouping allowed predictions from multiple models to be evaluated collectively within each endpoint before integration across endpoints.

Evaluation of prediction reliability

After generating model predictions, the reliability of the predictions was evaluated. However, the prediction reliability information provided by different platforms varies. Most platforms (Except for EPI Suite) provide information regarding the QMRF and the way to evaluate the AD that can be used to evaluate prediction reliability. Some platforms provide additional measurements besides AD for reliability (for an overview, see online supplementary material). In this study, two evaluation approaches were considered.

In the simple assessment approach, predictions were retained only if they fell within the AD of the model. Predictions outside the applicability domain were excluded from further analysis. Depending on the platform, applicability domain information may be provided either as a binary indicator (in-domain or out-of-domain) or as a continuous reliability score. In the case of binary output, only the in-domain predictions were considered. In the case where the AD is assessed as a continuous value, only predictions with high reliability (>80%) are considered. For EPI Suite, basic recommendations provided by the platform (e.g., the molecular weight range) were used as a proxy of AD.

In the advanced assessment approach, additional expert evaluation was applied to examine predictions that fell outside the formal applicability domain. In such cases, additional information was considered, including structural similarity to compounds in the model training set, presence of mechanistic alerts, and supporting evidence from related models. This step allows experienced assessors to evaluate whether certain predictions may still be informative despite formal applicability domain limitations.

Integration of predictions

When multiple models are available for the same endpoint, different strategies can be used to integrate predictions into a single outcome. In this study, three integration strategies were evaluated based on approaches previously described in the literature (Benfenati et al., 2019):

- Majority vote, in which the most frequently occurring prediction among the models is selected;
- Mean approach, applied when quantitative predictions are available;
- Conservative approach, in which the most hazardous prediction is selected as a precautionary estimate.

These integration strategies represent different ways of addressing uncertainty and weighing evidence from multiple models. The choice of integration strategy can influence the final outcome of the hazard assessment and should therefore be transparently documented. In practice, the selection of the most appropriate integration approach requires expert judgment and may depend on the assessment context.

Results

Overview of the workflow application

The developed workflow was applied to three case study chemicals: Bisphenol A (BPA), bisphenol AP (BPAP), and isosorbide. These substances were evaluated across hazard endpoints relevant to the SSbD framework, including carcinogenicity, germ cell mutagenicity, reproductive toxicity, endocrine disruption, persistence, bioaccumulation, and aquatic toxicity.

Across these endpoints, predictions were generated using multiple (Q)SAR models available across different in silico platforms. The outputs of these models included both quantitative predictions (e.g., predicted toxicity values) and categorical outputs (e.g., positive or negative predictions for a hazard class). For each endpoint, predictions from different models were grouped before evaluating their reliability and integrating their results.

Model availability from studied platforms

Persistence

For persistence, the half-life in different environmental compartments and ready biodegradability were predicted. These are the required endpoints under European legislation. VEGA (www.vegahub.eu) offers both a classification model and a quantitative model that provides a continuous value for the three compartments: Water, sediment, and soil. These two models should be used in conjunction to ensure consistency and avoid conflicts. The models are integrated within the JANUS software, which refers to REACH requirements and thresholds. An example of how to apply and interpret these models has been published (Benfenati et al., 2025). Other platforms used for persistence predictions are EPI Suite and the Danish (Q)SAR platform, both of which have combination methods for the prediction of ready biodegradability, and the Mistra SafeChem in silico toolbox v1.1, which contains two classification models for persistence in soil and two classification models for ready biodegradability.

Bioaccumulation

For bioaccumulation, the bioconcentration factor (BCF) and $\log K_{ow}$ were considered. VEGA has four BCF models. Two of them, Meylan and Arnot-Gobas, are reimplementations of those available within the EPI Suite platform. VEGA automatically provides the evaluation of the AD to be used for the prediction, whereas this is not possible with EPI Suite. For these two added reasons, the BCF models from EPI Suite were not used in this exercise. The Mistra SafeChem in silico toolbox has two classification models for BCF in fish. As logKow plays a crucial role in the BCF models, it is important to understand the impact of this parameter. A detailed explanation of the consequences of the $\log K_{ow}$ on the BCF models is reported (Benfenati et al., 2025).

Carcinogenicity

To predict carcinogenic properties under the SSbD framework, all available models from the Danish (Q)SAR Database, VEGA, and Mistra SafeChem were taken into account. As the individual models can be based on different types of studies (i.e., animal vs. human data, data on different animal species and different genders), different results might be generated in these models for the same chemical structure. It is thus recommended to consider these indications within each platform, since the developers may have optimized the way to integrate the results. Therefore, for VEGA, the JANUS Carcinogenicity CONSENSUS model was run, and for the Danish (Q)SAR Database, the strategy to assign a positive or nonpositive overall call was applied (DTU Food, 2018). The strategy to integrate the results across platforms was equal to the one applied by EFSA to prioritize which botanicals found in vegetables available on the European market should require further studies in order to clarify issues related to their health effects (EFSA, 2025).

Germ cell mutagenicity

The CLP hazard class germ cell mutagenicity is concerned primarily with substances that may cause mutations in germ cells of humans that can be transmitted to the progeny. However, the results from mutagenicity or genotoxicity tests in vitro and in mammalian somatic and germ cells in vivo are also considered in the classification. Assessing the mutagenicity of a substance requires the coverage of several key endpoints, specifically gene mutation and chromosomal damage, including both clastogenic and aneugenic effects. Germ cell mutagenicity is generally evaluated in vivo by heritable germ cell mutagenicity tests, such as the rodent dominant lethal mutation test or mouse heritable translocation assay. For detecting gene mutations, the Ames test is widely used, which detects mutations in bacterial strains. On the other hand, the in vitro or in vivo micronucleus test, and the in vitro or in vivo chromosomal aberration assay, are used to evaluate structural and numerical chromosomal changes. Together, these complementary approaches provide a thorough understanding of a substance's potential to cause gene mutations and chromosomal damage. In general, evidence of germ cell toxicity in humans or mammals is required for classification in category 1 of CLP, unless there is supporting evidence that somatic cell mutagenicity or genotoxicity is also applicable in germ cells. Somatic cell genotoxicity tests in vivo, supported by in vitro tests or structure–activity relationships to known germ cell mutagens, warrant classification to category 2. It is important to note that although evidence of a single mechanism of germ or somatic cell mutagenicity and/or genotoxicity is sufficient for classification, a negative result in any individual test does not indicate a substance is safe.

For these endpoints, a large battery of models is present on different platforms. VEGA has several models for Ames, providing also a consensus tool. Moreover, VEGA and the Danish (Q)SAR Database have a list of models related to the in vitro and in vivo micronucleus test. Different strategies can be used for the application of these models, depending on the regulatory context. For SSbD purposes, all available models were used to predict mutagenic properties.

Reproductive toxicity and endocrine disruption

In the case of reproductive toxicity, the models could be divided into two main groups: Models specific for reproductive toxicity, which are based primarily on human and animal data, and models for ED. The ED models are included here as there is a significant overlap between reprotoxicity and ED-related endpoints. The ED models relate to reproductive toxicity by varying degrees, making it difficult to draw the line between models belonging to both groups and strictly ED models. Making such a differentiation between reprotoxicity and ED may leave out potentially useful information. The ED models include one general model based on human and wildlife data (VEGA) as well as many mechanistic in vitro models linked to reproduction to varying degrees, some directly and some indirectly. These mechanistic models can be used as supporting information and indicators of research needs. However, a positive outcome in these models alone should not override the results from models that are based on in vivo data, as their results are rarely quantitative and usually cover only a single key event. For classification purposes, it is currently necessary to show that effects are observed in vivo or in humans. In addition, the mechanistic models for ED reviewed in this case study are based on in vitro data that are subject to high variation due to differences in experimental conditions, resulting in variable biological relevance.

Aquatic toxicity

Aquatic toxicity refers to the potential of a chemical substance to cause harmful effects on aquatic organisms, such as fish, invertebrates, and algae. Under the REACH Regulation, aquatic toxicity is a key environmental hazard endpoint used to assess the intrinsic properties of substances and to inform risk management measures. For aquatic toxicity, we applied the battery of models implemented in the VEGA software. VEGA indeed contains 15 models for acute aquatic toxicity and three for chronic aquatic toxicity, covering the three trophic levels (fish, daphnia, algae). Other available models from the EPI Suite platform and the Danish (Q)SAR Database were used, too. Also, for aquatic toxicity, a detailed explanation on how to integrate the different models and apply a WoE approach is provided (Benfenati et al., 2025).

Evaluation of prediction reliability

Figure 2 illustrates the proportion of model predictions retained as reliable under the two evaluation approaches (the simple evaluation and advanced evaluation) for the three case chemicals. As expected, the advanced evaluation resulted in a larger number of predictions being considered reliable compared to the simple evaluation approach.

Integration of predictions

Table 2 summarizes the combined results for each hazard property across the three chemicals based on predictions retained after reliability evaluation. In general, BPA and BPAP showed potential hazard signals for all endpoints, whereas isosorbide was predicted to have fewer hazard concerns and only got flagged for reprotoxicity.

The general outcome per property is the same for the simple and advanced assessment; however, the prediction of the individual endpoints can be different. The full list of predicted endpoints for every substance and model as well as the combined output according to the three strategies can be found in the online supplementary material. For C, M, and R properties, all included models covered a different endpoint. Therefore, no summary of the results was created according to the three predefined strategies. Importance should be given to the type of endpoint flagged and by what model. These findings illustrate that integrating multiple in silico predictions can lead to different possible hazard assessment outcomes, and how this should inform chemical/material innovation under the SSbD framework is user-dependent. However, the transparent documentation of integration strategies and decision rules is important in the SSbD assessment.

Discussion

Selection of relevant models

The workflow for integrating the results for a certain property may vary depending on the specific case. Once the assessment context is defined, this should be used to identify the necessary endpoints. Here, the assessment context refers to the regulatory or decision-making purpose of the evaluation, including the type of hazard endpoint required, the stage of innovation, and the intended use of the assessment results. We have seen that multiple models may be available to address the same property or endpoint, and this is an advantage. For instance, in the CLP classification, mutagenicity classification requires assessing permanent changes in the amount or structure of genetic material in human germ cells. This definition implies covering germ cell mutations, while other regulations only require mutagenicity to detect mutations in somatic cells. Further differences exist not only for the regulations but also at the international level; thus, there are different procedures to evaluate a certain property depending on the geographical area. Furthermore, different toxicity classes and threshold values exist in different countries. This process should then be applied to the specific substance(s). The suitability of a certain model indeed depends on the specific substance. We have discussed the relevance of the AD to identify which prediction is reliable. Thus, it is not possible to identify a model that is always reliable and another one that fails. The results are substance-dependent. The user should refer to the QAF for proper guidance (OECD, 2024).

Regardless of the implications in different regulations and possible use of tools outside the SSbD assessment framework (Clift et al., 2013; Nohmi, 2018), the importance of the SSbD toolbox to allow users to select the most relevant properties for their specific case should be emphasized, rather than being limited to a single predefined workflow with a fixed set of (Q)SAR models. Flexibility in model and endpoint selection is critical to

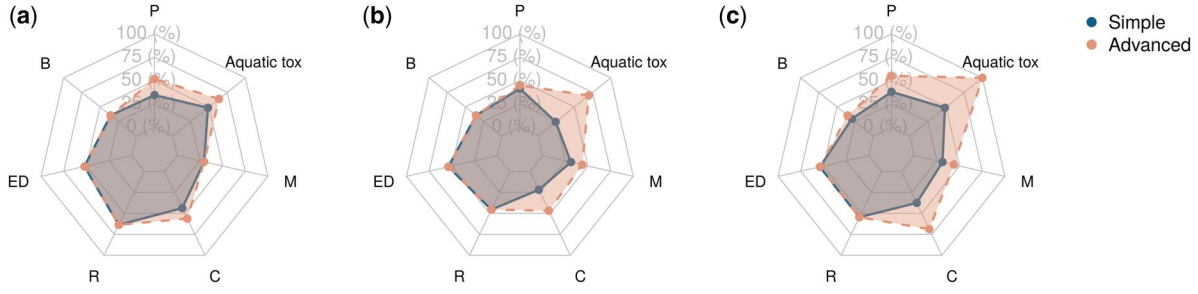

Figure 2 Share of (Q)SAR predictions, as a percentage out of all available (Q)SARs for that specific property, that were considered as reliable under the "simple" and "advanced" evaluation schemes for (A) BPA, (B) isosorbide, and (C) BPAP. The properties covered are persistency (P), bioaccumulation (B), carcinogenicity (C), mutagenicity (M), reproductive toxicity (R), endocrine disruption (ED), and aquatic toxicity. (Q)SAR = (Quantitative) Structure–Activity relationship; BPA = bisphenol A; BPAP = bisphenol AP.

Table 2 Combined results for each property for BPA, isosorbide, and BPAP.

Property	BPA	Isosorbide	BPAP
Persistency	Potential issue	Not flagged	Potential issue
Bioaccumulation	Potential issue	Not flagged	Potential issue
Carcinogenicity	Potential issue*,**	Not flagged*,**	Potential issue**
Mutagenicity	Potential issue**	Not flagged**	Potential issue**
Reprotoxicity	Potential issue**	Potential issue**	Potential issue**
Endocrine disruption	Potential issue	Not flagged	Potential issue*
Aquatic toxicity	Potential issue	Not flagged	Potential issue

* Inconclusive results.
** No results for one or more models. The full list of predicted endpoints and the combined output can be found in the online supplementary material.
BPA = bisphenol A; BPAP = bisphenol AP.

ensure relevance and applicability across a wide range of substances and regulatory contexts.

In addition, it must be acknowledged that not all endpoints can currently be addressed using in silico tools alone. This highlights the need for further research and development to expand the applicability domain and predictive power of existing models.

Uncertainty of in silico predictions

Each model prediction comes with a degree of uncertainty. To address this, it is recommended to use multiple models whenever possible. For example, consensus models can be seen as more reliable than individual models, so that in theory we can place more weight on these outcomes. However, the relative importance of outcomes is something that cannot be predetermined and needs to be defined at the start of every SSbD assessment. Also, agreement between multiple models does not necessarily guarantee that a prediction is correct, particularly if models rely on similar training data or modeling assumptions. Therefore, interpretation of model outputs should involve expert judgment from toxicologists or environmental scientists familiar with the endpoint and modeling approaches.

The approach of combining different models can help reduce the uncertainty of predictions by increasing confidence in the result. However, during this work, we observed that improvements could be made in how different platforms communicate information, particularly regarding predictions that fall outside the AD. Different tools convey whether a prediction falls inside or outside the AD in various ways. For instance, in EPI Suite, evaluating the AD is a manual process, requiring a time-consuming analysis of chemical moieties and their occurrence in the substance. In contrast, tools like VEGA and the Danish (Q)SAR Database automatically check several conditions to assess the AD. Moreover, it is highly beneficial if a tool provides information on why a prediction falls outside the AD. This allows for a manual assessment based on multiple lines of evidence, enabling users to include or exclude certain values by considering similar substances in the dataset and theoretical factors.

Within the perspective of using the results of in silico models for SSbD, the information about the uncertainty of the prediction should be incorporated within the process. Within this approach, it may be convenient to apply an evaluation of the compared hazard impact associated with the candidate substituent, compared to the substance to be replaced. In this scenario, the evaluation may not be an absolute one but rather a relative comparison. In other words, the in silico models should be applied both to the target and candidate substances. The information about the hazard of the target substance (which quite likely is available) to be replaced should be used to verify the correctness of the results of the in silico models. If the candidate substance is not very different from the target, a comparison of the correctness of the results can be made.

The granularity of the models for the same property

Several hazard properties need to be assessed, depending on the regulation and the purpose. As stated above, it is frequent that for the same property, several endpoints exist, which can differ according to defined guidelines. There is a general need to adopt a standardized ontology and refer to common definitions, such as is being developed at the OECD level with the Global Harmonized System (GHS) perspective. In Europe, the initiative called “One Substance-One Assessment” is addressing this need, too.

However, as observed by our analysis of the QRMFs, the development of several in silico models started with collections of data, which do not always adhere to a single protocol. Thus, there may be a misalignment between some in silico models and the official endpoint definitions. For instance, there are models for fish acute toxicity, which use data related to multiple fish species as a training set. The result of this model may refer to a “generic,” not defined, fish, whereas in other contexts, the identification of the fish species is requested. Similarly, it is usual and requested in certain regulatory frameworks to provide data on the Ames test generated using five different strains, with and without metabolic activation. However, for the Ames test in silico models, the common situation is that they provide a single outcome, since they are developed to merge results obtained within different conditions.

Even more complex is the case of the models for carcinogenicity. In this case, there is a range of models. Some models may relate to a single animal species, whereas others may relate to studies conducted with different animal species and different sexes. In several cases, assessments of substances are done by human experts who may consider epidemiological studies and in vitro tests, too. Thus, these models rely on different data, which in general can be called carcinogenicity but, in practice, look at different situations.

The case of the models for developmental toxicity and reprotoxicity is also challenging. In practice, the different models cannot be compared since they refer to different endpoints that cover developmental toxicity and impaired fertility. For classification purposes, it is necessary to demonstrate that the adverse effect is realized in humans or in experimental animals. Notably, these effects may manifest not only in the exposed individual but also in subsequent generations. As a result, there are many possible experimental methods that can be applied to identify reprotoxic effects, many of which are time-consuming and costly. Consequently, there are relatively few models based on animal/human data that are specific to this endpoint. This is something that the SSbD evaluator needs to be aware of, as in the end, this would result in a data gap. How data gaps can be addressed and taken into account in the evaluation approach of the SSbD assessment was not part of this work but should be addressed in the future.

Combining results

Based on the strategy to combine different predictions (i.e., mean, majority, or conservative approach), the final outcome may vary, meaning there are multiple SSbD outcomes possible for the same use case. For example, based on the BCF predictions, BPAP can potentially be nonbioaccumulative or bioaccumulative, depending on the strategy applied. This illustrates that integration strategies should be explicitly defined and transparently reported when using multiple in silico predictions. It is important to highlight that there is no single best way to combine the results, and the procedure also depends on practical considerations, such as time and tools available and the preference of the assessor for choosing either the most conservative value or to adopt a more weight-of-evidence approach. In practice, expert judgment is required to determine which integration strategy is most appropriate for the specific assessment context. In all cases, the procedure that is taken should be declared to increase transparency and to avoid bias.

When multiple models provide conflicting values, this is an indication of difficulties in the prediction, representing in a clear way the uncertainty of the assessment. When a single value is available, the evaluation becomes more difficult, and other lines of evidence should be considered, such as the presence of similar substances confirming the prediction (read-across approach, thus evidence-based) or the presence of a plausible mechanism confirming the assessment (theoretical supporting information). Furthermore, by considering additional models under the advanced assessment method, some unequivocal results under the majority and mean approaches can be resolved, emphasizing the added value of this step.

The QAF aims to develop a systematic and harmonized framework for the regulatory assessment of (Q)SAR models, predictions, and results based on multiple predictions. It is applicable irrespective of the modeling technique used to build the model, the predicted endpoint, and the intended regulatory purpose.

The assessment of (Q)SARs for regulatory purposes should not be limited to checking the validity of the model used because even a valid model can produce unacceptable predictions under certain conditions. Therefore, individual predictions and results from multiple predictions need dedicated assessments.

Practical implications for the implementation of SSbD

The findings of this study have different implications for stakeholders involved in the implementation of the SSbD framework, including regulators, industry actors, and researchers. In particular, they highlight both the opportunities and the current limitations of relying on in silico tools for early-stage hazard identification.

For regulators, the developed workflow presented in this study demonstrates how multiple in silico tools can be applied in a structured and transparent manner while explicitly addressing uncertainty. The results highlight that different strategies for integrating model outputs (e.g., majority vote, mean, or conservative approaches) can lead to different assessment outcomes, even when based on the same underlying predictions. This strengthens the importance of documenting not only the chosen tools but also the justification of the integration strategies and decision rules in line with EFSA's WoE Principles and the OECD's QAF (EFSA Scientific Committee et al., 2017; OECD, 2023a). Rather than aiming for a single definitive outcome, the workflow supports regulatory use of in silico tools as part of an iterative decision-making process, where uncertainty and data gaps are explicitly stated and can be re-examined as additional information becomes available.

For industry and innovators, particularly at early stages of chemical and material development, the workflow provides a practical means to screen potential hazards before significant resources are committed. The results show that in silico tools are well suited to identifying potential areas of concern across a broad range of SSbD-relevant endpoints, but also that conclusions may depend on model coverage and agreement. From a design perspective, this suggests that SSbD assessments should be used to guide informed choices and prioritization rather than to deliver binary "pass/fail" decisions. Early identification of potential hazard signals can support substitution, redesign, or targeted data generation, while acknowledging that uncertainty is an inherent feature of early innovation.

For researchers and tool developers, the analysis highlights several areas where further methodological development is needed to support SSbD implementation. These include improved harmonization of endpoint definitions across models, clearer and more consistent communication of AD information, and expanded model coverage for currently underrepresented endpoints. The observed variability between models and platforms also points to the need for further work on integration approaches that go beyond simple aggregation, particularly for complex endpoints such as reproductive toxicity and endocrine disruption. Addressing these challenges through harmonized reporting, improved applicability domain diagnostics, and the development of models for currently underrepresented SSbD-relevant endpoints would significantly strengthen the role of in silico tools in future SSbD assessment (Benfenati et al., 2019; OECD, 2023a).

Overall, in silico tools can play a central role in operationalizing SSbD, provided they are applied within a transparent, well-documented framework that explicitly accounts for uncertainty and context of use. Rather than replacing experimental data or expert judgment, such tools are most effective when used to

support early decision-making, learning, and iteration throughout the innovation process.

Conclusions and recommendations

In this study, we evaluated the use of in silico tools to predict chemical hazards under the SSbD assessment framework; however, the approach can be applied to different regulatory contexts, too. An SSbD assessment is an iterative process throughout the innovation path, where the flow of information gradually increases moving from one innovation stage to the next. In relation to the position of in silico tools, they are critical during the earlier stages of innovation when the information is very limited and the uncertainty higher. Even though these tools are commonly used for hazard predictions, they present limitations regarding their incorporation into the SSbD assessment process. Within the PARC SSbD toolbox, the previously discussed (Q)SAR tools, along with additional in silico tools, have been mapped across the innovation process and the SSbD framework assessments for safety and sustainability. One of the main objectives of the PARC SSbD toolbox is to integrate tools and data throughout an SSbD assessment pipeline. Therefore, the current study serves as a starting point for building an in silico hazard assessment workflow in the context of the PARC SSbD toolbox development. For a better-defined protocol for in silico hazard assessment under SSbD (and other regulatory contexts), a set of recommendations is provided.

Prediction reliability

In silico tools are generally easy to use and do not require extensive input data. However, to receive predictions that are not only rapid but also reliable, the hazard identification within the SSbD framework should incorporate the evaluation of the prediction's reliability. To accomplish this, a series of actions need to be implemented. First, a systematic approach to integrate reliability assessment into models that do not include it within their estimations should be structured. This would also facilitate the process of evaluating predictions of the same endpoint by different models. Moreover, assessors should be provided with comprehensive information regarding the AD status of each model, including the reason why a specific structure falls outside the AD. This would allow them to reach informed decisions on whether to consider a model prediction in their assessment.

Missing endpoints and data gap strategy

Within the iterative process of SSbD across innovation, data gaps are expected to occur, particularly in the earlier stages. These data gaps need to be considered during the decision-making process. In addition, from the current study, we identified hazard endpoints (e.g., Specific Target Organ Toxicity—Single Exposure, physical hazards) that are included in the EC SSbD framework for which predictive models are not available or very limited in applicability. As a result, the interpretation of results received from SSbD assessment, and particularly hazard assessment, should require a comprehensive and transparent approach to consider data gaps. In conjunction with the missing endpoints, there is a need for future (Q)SAR model developments that cover these types of endpoints and address potential data gaps within the assessment pipeline. Also, when (Q)SAR models are lacking, there are other in silico approaches—such as read-across and expert systems—that might be used to provide hazard predictions. These methods can leverage chemical similarity, mechanistic reasoning, or weight-of-evidence approaches to reduce uncertainty for specific endpoints or borderline cases. Future SSbD workflows could also benefit from integrated strategies combining (Q)SARs with complementary in silico methodologies to further strengthen confidence in hazard-based design decisions.

Complex endpoints

Another important aspect that needs to be considered is the complexity associated with some of the indicators, such as reprotoxicity. The prediction and evaluation of these indicators may be more challenging compared to simpler indicators. Therefore, a new scheme for the assessment of a substance should be structured that integrates in silico tools more effectively, which should be used not only to replicate a limited traditional scheme. Furthermore, the integration of results from different models into a single overall call should be investigated. For instance, the integration of outcomes resulting from models that predict the same endpoints, such as mutations in cells, but for different organisms (i.e., mouse and hamster) could be explored. Finally, an additional area of investigation could entail the integration of different endpoints (i.e., different assays) related to the same indicator (i.e., mutagenicity) into a single value. This would also involve the strategy needed for implementation and how this could be established efficiently.

Supplementary material

Supplementary material is available at *Integrated Environmental Assessment and Management* online.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Author contributions

Joanke van Dijk (Conceptualization, Data curation, Formal analysis, Investigation, Resources, Validation, Visualization, Writing—original draft, Writing—review & editing), Anna Agalliadou (Writing—original draft, Writing—review & editing), Chiara L. Battistelli (Data curation, Formal analysis, Investigation, Validation, Writing—review & editing), Emilio Benfenati (Software, Writing—original draft, Writing—review & editing), Cecilia Bossa (Data curation, Formal analysis, Investigation, Resources, Validation, Writing—review & editing), Maja Halling (Conceptualization, Project administration, Writing—review & editing), Spyros Karakitsios (Conceptualization, Supervision, Writing—review & editing), Achilleas Karakoltzidis (Investigation, Writing—original draft, Writing—review & editing), Fotini Nikiforou

(Conceptualization, Investigation, Writing—original draft, Writing—review & editing), Dimosthenis A. Sarigiannis (Funding acquisition, Supervision), Gianluca Selvestrel (Data curation, Formal analysis, Investigation, Resources, Validation, Writing—original draft, Writing—review & editing), Anežka Sharma (Conceptualization, Investigation, Writing—original draft, Writing—review & editing), Kirsi M. Siivola (Data curation, Formal analysis, Investigation, Resources, Validation, Writing—review & editing), and Ziye Zheng (Data curation, Formal analysis, Investigation, Resources, Validation, Writing—review & editing)

Funding

The European Partnership for the Assessment of Risks from Chemicals (PARC) has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101057014. JvD was also supported by the Swiss Federal Office for the Environment. AS was also supported by the RECETOX Research Infrastructure (No. LM2023069) financed by the Ministry of Education, Youth and Sports.

Conflicts of interest

There is no conflict of interest associated with this work.

Acknowledgments

The authors thank Fiorella Pitaro for the internal review.

References

Almroth, B. C., Cornell, S. E., Diamond, M. L., de Wit, C. A., Fantke, P., & Wang, Z. (2022). Understanding and addressing the planetary crisis of chemicals and plastics. *One Earth*, *5*, 1070–1074. https://doi.org/10.1016/j.oneear.2022.09.012

Benfenati, E., Chaudhry, Q., Gini, G., & Dorne, J. L. (2019). *Integrating in silico models and read-across methods for predicting toxicity of chemicals: A step-wise strategy*. https://doi.org/10.1016/j.envint.2019.105060

Benfenati, E., Manganaro, A., & Gini, G. (2013). VEGA-QSAR: AI inside a platform for predictive toxicology. *CEUR Workshop Proceedings*, *1107*, 21–28.

Benfenati, E., Selvestrel, G., Colombo, E., Viganò, E. L., Raitano, G., & Lombardo, A. (2025). In silico models: Assessment of the results of the VEGA models within an exercise of the PARC project. Zenodo. https://doi.org/10.5281/zenodo.15131130

Caldeira, C., Farcal, R., Garmendia Aguirre, I., Mancini, L., Tosches, D., Amelio, A., Rasmussen, K., Rauscher, H., Riego Sintes, J., & Sala, S. (2022). *Safe and sustainable by design chemicals and materials—Framework for the definition of criteria and evaluation procedure for chemicals and materials. EUR 31100 EN*. Publications Office of the European Union.

Caldeira, C., Garmendia Aguirre, I., Tosches, D., Mancini, L., Abbate, E., R., F., D., L., K., R., H., R., J., R. S., & S., S. (2023). *Safe and Sustainable by Design chemicals and materials—Application of the SSbD framework to case studies*. (KJ-NA-31-528-EN-N (online), KJ-NA-31-528-EN-C (print)). https://doi.org/10.2760/329423%2520(online),10.2760/769211%2520(print)

Clift, M. J. D., Raemy, D. O., Endes, C., Ali, Z., Lehmann, A. D., Brandenberger, C., Petri-Fink, A., Wick, P., Parak, W. J., Gehr, P., Schins, R. P. F., & Rothen-Rutishauser, B. (2013). Can the Ames test provide an insight into nano-object mutagenicity? Investigating the interaction between nano-objects and bacteria. *Nanotoxicology*, *7*, 1373–1385. https://doi.org/10.3109/17435390.2012.741725

DTU. (n.d.). *Danish (Q)SAR Database* [Computer software]. Research Group for Chemical Risk Assessment and GMO, National Food Institute, Technical University of Denmark. http://qsar.food.dtu.dk

DTU Food. (2018, December). *User Manual for the Danish (Q)SAR Database*. https://qsardb.food.dtu.dk/db/index.html

ECHA. (2011). *Guidance on information requirements and chemical safety assessment Part B: Hazard assessment, version 2.1 (ECHA-11-G-16-EN)*. European Chemicals Agency. https://echa.europa.eu/documents/10162/17235/information_requirements_part_b_en.pdf/7e6bf845-e1a3-4518-8705-c64b17cecae8

EFSA Scientific Committee, A., Benford, D., Halldorsson, T., Jeger, M. J., Knutsen, H. K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J. R., Silano, V., Solecki, R., Turck, D., Benfenati, E., Chaudhry, Q. M., Craig, P., Frampton, G., … Younes, M. (2017). Guidance on the use of the weight of evidence approach in scientific assessments. *EFSA Journal*, *15*, e04971. https://doi.org/10.2903/j.efsa.2017.4971

EFSA. (2025, December 5). *Compendium of botanicals: User guide and explanatory note on use of QSAR predictions*. https://www.efsa.europa.eu/en/data-report/compendium-botanicals

EMA. (2023). *ICH M7 Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk—scientific guideline*. https://www.ema.europa.eu/en/ich-m7-assessment-control-dna-reactive-mutagenic-impurities-pharmaceuticals-limit-potential-carcinogenic-risk-scientific-guideline

Hardy, A., Benford, D., Halldorsson, T., Jeger, M. J., Knutsen, H. K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J. R., Silano, V., Solecki, R., Turck, D., Benfenati, E., Chaudhry, Q. M., Craig, P., Frampton, G., … Younes, M.; EFSA Scientific Committee. (2017). Guidance on the use of the weight of evidence approach in scientific assessments. *EFSA Journal*, *15*, e04971. https://doi.org/10.2903/j.efsa.2017.4971

European Commission. (2020). Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. Chemicals Strategy for Sustainability—Towards a Toxic-Free Environment. COM(2020) 667 final.

Garmendia Aguirre, I., Abbate, E., Bracalente, G., Mancini, L., Cappucci, G. M., Tosches, D., Rasmussen, K., Sokull-Kluettgen, B., Rauscher, H., & Sala, S. (2025). *Safe and Sustainable by Design chemicals and materials. Revised framework (no. JRC143022)*. Publications Office of the European Union. https://doi.org/10.2760/5103785

Leder, C., Rastogi, T., & Kümmerer, K. (2015). Putting benign by design into practice-novel concepts for green and sustainable pharmacy: Designing green drug derivatives by non-targeted

synthesis and screening for biodegradability. *Sustainable Chemistry and Pharmacy*, *2*, 31–36. https://doi.org/10.1016/j.scp.2015.07.001

Lorenz, S., Amsel, A.-K., Puhlmann, N., Reich, M., Olsson, O., & Kümmerer, K. (2021). Toward application and implementation of in silico tools and workflows within benign by design approaches. *ACS Sustainable Chemistry & Engineering*, *9*, 12461–12475. https://doi.org/10.1021/acssuschemeng.1c03070

Marx-Stoelting, P., Rivière, G., Luijten, M., Aiello-Holden, K., Bandow, N., Baken, K., Cañas, A., Castano, A., Denys, S., Fillol, C., Herzler, M., Iavicoli, I., Karakitsios, S., Klanova, J., Kolossa-Gehring, M., Koutsodimou, A., Vicente, J. L., Lynch, I., Namorado, S., … Sanders, P. (2023). A walk in the PARC: Developing and implementing 21st century chemical risk assessment in Europe. *Archives of Toxicology*, *97*, 893–908. https://doi.org/10.1007/s00204-022-03435-7

Nohmi, T. (2018). Thresholds of genotoxic and non-genotoxic carcinogens. *Toxicological Research*, *34*, 281–290. https://doi.org/10.5487/TR.2018.34.4.281

OECD. (2023a). (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions (No. 386; OECD Series on Testing and Assessment). OECD Publishing, Paris. https://doi.org/10.1787/d96118f6-en

OECD. (2023b). QSAR Toolbox v4.6 [Computer software]. Organisation for Economic Cooperation and Development. https://qsartoolbox.org/

OECD. (2024). *(Q)SAR assessment framework: Guidance for the regulatory assessment of (quantitative) structure activity relationship models and predictions* (2nd ed.). OECD Publishing. https://doi.org/10.1787/bbdac345-en

Nikiforou, F., Halling, M., Rydberg, T., Telaretti Leggieri, R., Westra, J., Subramanian, V., Sarigiannis, D., Karakitsios, S., Karakoltzidis, A., Agalliadou, A., van Dijk, J., Iavicoli, I., Fontana, L., Leso, V., & D'Anna, A.; PARC. (2024). *D8.4—1st Report on the testing and uptake of the SSbD toolbox through use cases* [Public]. https://www.eu-parc.eu/sites/default/files/2025-09/PARC_D8.4.pdf

Sarigiannis, D., Karakitsios, S., Iavicoli, I., Westra, J., & Nowack, B. (2023). *D8.1: Report on the conceptual design of the SSbD toolbox. WP8—T8.1*. https://www.eu-parc.eu/sites/default/files/2024-03/PARC_D8.1.pdf

Sarigiannis, D., Nikiforou, F., Karakoltzidis, A., Agalliadou, A., Rydberg, T., Halling, M., Battistelli, C. L., Benfenati, E., Bossa, C., Bouman, E., Bourgé, É., Brouwer-Milovanovic, M., Hill, A., Iacovidou, E., Iavicoli, I., Kanerva, T., Kärnman, T., Leso, V., Linden, J., … Karakitsios, S. (2024). OS01-12 A computational toolbox supporting the development of Safe and Sustainable by Design chemicals and materials. *Toxicology Letters, Abstracts of the 58th Congress of the European Societies of Toxicology (EUROTOX 2024)*, *399*, S57. https://doi.org/10.1016/j.toxlet.2024.07.161

Thomas, P. C., Bicherel, P., & Bauer, F. J. (2019). How in silico and QSAR approaches can increase confidence in environmental hazard and risk assessment. *Integrated Environmental Assessment and Management*, *15*, 40–50. https://doi.org/10.1002/ieam.4108

UNEP. (2019). Global Chemicals Outlook II. From Legacies to Innovative Solutions: Implementing the 2030 Agenda for Sustainable Development – Synthesis Report. https://wedocs.unep.org/bitstream/handle/20.500.11822/27651/GCOII_synth.pdf? sequence=1&isAllowed=y

USEPA. (2012). *Estimation Programs Interface Suite for Microsoft Windows, v. 4.11 [Computer software]*. United States Environmental Protection Agency.

van Dijk, J., Flerlage, H., Beijer, S., Slootweg, J. C., & van Wezel, A. P. (2022). Safe and sustainable by design: A computer-based approach to redesign chemicals for reduced environmental hazards. *Chemosphere*, *296*, 134050. https://doi.org/10.1016/j.chemosphere.2022.134050