From *in silico* predictions to regulatory submissions

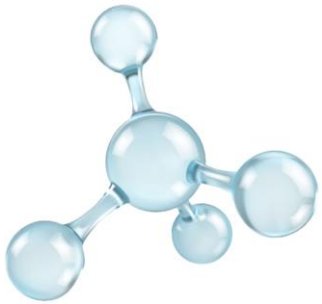
AiChemist Paris School, March 10th, 2026

Marco Evangelista

CEHTRA and Simply Predict

CEHTRA (Consultancy for Environmental & Human Toxicology and Risk Assessment)

Is an international leader in technical and regulatory support for the safety of chemical products in multiple markets: industrial chemicals, biocides, plant protection products, cosmetics, pharmaceuticals, food products, etc.

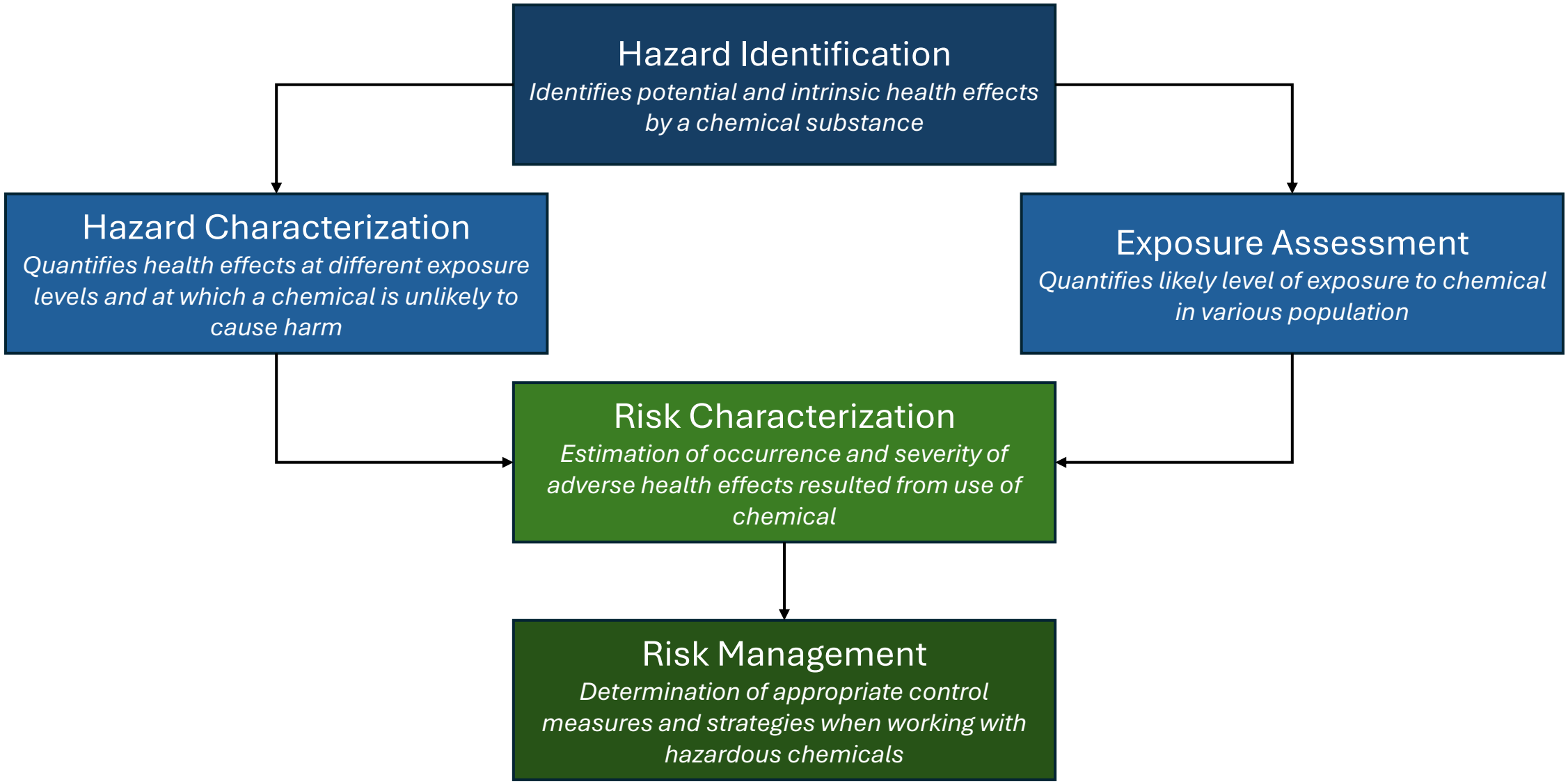


simply predict®

Powered by CEHTRA

- Service fully powered by CEHTRA;
- We promote and apply *in silico* New Approach Methodologies (NAMs), notably Read-Across, SARs and QSARs for R&D and regulatory needs of our clients: physicochemical and e-fate properties, ecotoxicological and toxicological endpoints.

Background: Chemical Risk Assessment



Gold Standard: (Vertebrate) Animal-Based Testing

Fails to keep pace

> 350,000 chemicals on the global market
~ 500 chemicals with hazard characterization

Resource demanding

- Financial costs
- Number of living organisms
- Time

Interspecies extrapolations

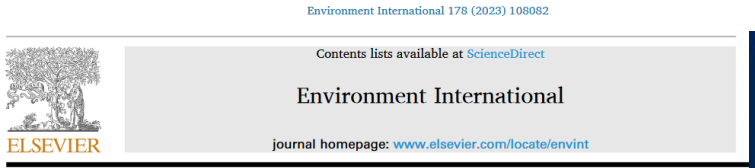
Limitations and challenges to accurately predict hazards across different species.

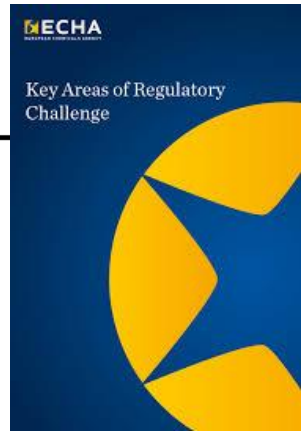
Ethical concern

3 Rs principles: Refine, Reduce, Replace

Need to develop, validate and integrate
New Approach Methodologies (NAMs)
in chemical hazard assessment.

New Approach Methodologies (NAMs)



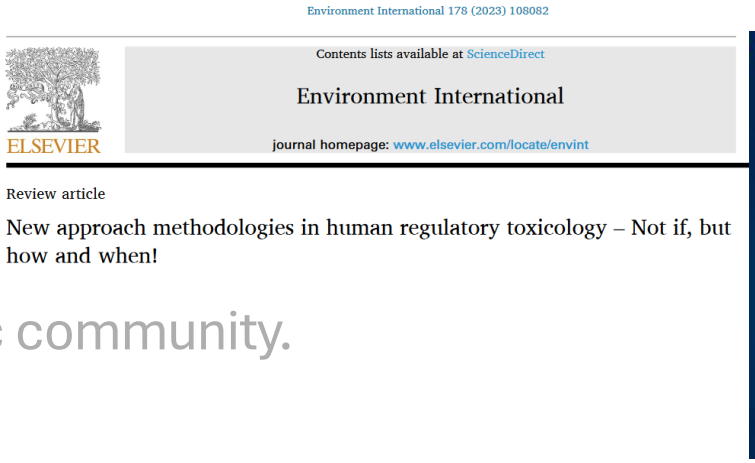


NAMs as alternatives to animal testing:

1. Include *in silico*, *in vitro*, *in chemico* methods;
2. Actively promoted by regulatory bodies and scientific community.
3. Embedded in major EU chemical regulations, e.g.:
    - REACH promotes the use of alternatives "whenever possible" and explicitly states that testing on animals should only occur as a last resort;
    - BPR and PPPR encourage minimization of animal-based tests thereby promoting the use of alternative methods;
    - Testing ban on finished cosmetic products and cosmetic ingredients.
4. The EC is preparing a "Roadmap Towards Phasing Out Animal Testing for Chemical Safety Assessments".

New Approach Methodologies (NAMs)

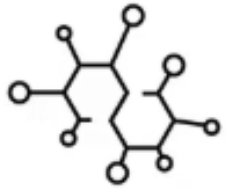
NAMs as alternatives to animal testing:



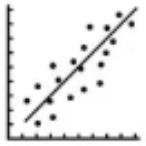
1. Include ***in silico***, *in vitro*, *in chemico* methods;
2. Actively promoted by regulatory bodies and scientific community.
3. Embedded in major EU chemical regulations, e.g.:
    - REACH promotes the use of alternatives “whenever possible” and explicitly states that testing on animals should only occur as a last resort;
    - BPR and PPPR encourage minimization of animal-based tests thereby promoting the use of alternative methods.
4. The EC is preparing a “Roadmap Towards Phasing Out Animal Testing for Chemical Safety Assessments”.

In silico NAMs

Different types of *in silico* NAMs…



Structure-Activity Relationship (SARs) or Profilers



Quantitative Structure-Activity Relationship (QSAR) Models

Read-Across

Endpoints

… to predict a wide variety of endpoints of interest

Physicochemical and e-fate properties

- Dissociation constant (pKa)
- Flash point
- Melting point
- Boiling point
- Vapour pressure
- Water solubility
- Kow
- Koc
- Viscosity
- Ready Biodegradability
- Bioaccumulation (BAF)
- Bioconcentratiion (BCF)
- …

Human Health endpoints

- Dermal absorption
- Skin sensitization
- Irritation/corrosion (skin/eye)
- Acute (dermal, oral, inhalation) toxicity
- Genotoxicity & Mutagenicity
- Carcinogenicity
- Developmental toxicity
- Reproductive toxicity
- Repeated dose toxicity
- Endocrine Disruption (EATS)
- …

Ecotoxicological endpoints

- Acute toxicity to Fish
- Chronic toxicity to Fish
- Acute toxicity to aquatic invertebrates
- Chronic toxicity to aquatic invertebrates
- Acute toxicity to Algae
- Chronic toxicity to Algae
- …

In silico NAMs – SARs / Profilers

- Structure-Activity Relationships (SARs);

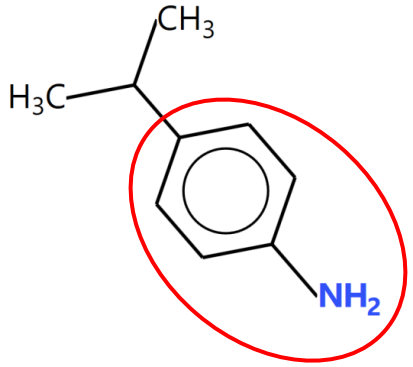
- Chemical substructures (sub-structural moieties, functional groups) known or suspected to be associated with a certain (eco)toxicological effect, based on expert knowledge or historical data. The presence of a structural alert flags a potential hazard;

- Commonly applied in the initial stage of R&D screening to phase out potentially toxic chemicals.

- Absence of an alert does not imply absence of toxicity!
- Presence of an alert does not imply presence of toxicity!

In silico NAMs – SARs / Profilers – Skin sensitization

CAS N°	99-88-7
Name	4-Isopropylaniline
Molecular formula	$C_9H_{13}N$
MW (g/mol)	135.21
SMILES	CC(C)C1=CC=C(C=C1)N



No experimental test data on skin sensitization.

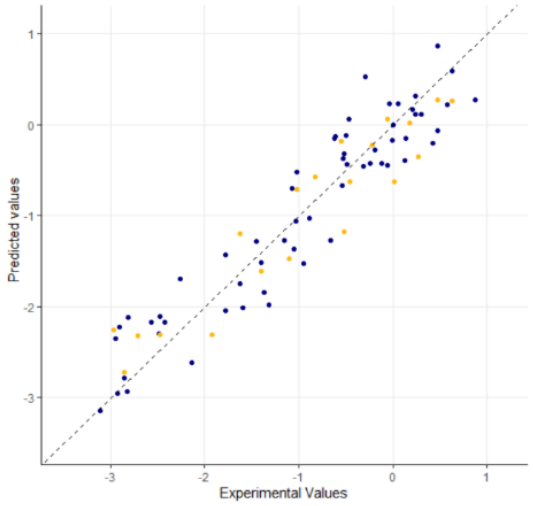
Multiple SARs (e.g., OECD QSAR Toolbox): Structural alerts for protein binding (molecular initiating event for skin sensitization).

In silico NAMs – QSAR models

- Quantitative Structure-Activity Relationships (QSARs);
- Trained using statistical algorithms;
- From a defined training set of multiple chemical structures, correlate a set of molecular descriptors (e.g., measured or computed physical-chemical properties, theoretical descriptors) with a measured (experimental) biological activity or toxicity;
- Statistical QSAR model can be regression-based (i.e., predict continuous values) or classification-based (i.e., predict categorical values).

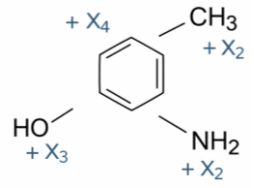
CHEMICAL STRUCTURE
(molecular descriptors – independent variables – X)

Y = f (X)

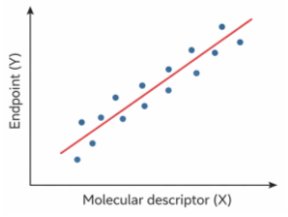


PHYSICOCHEMICAL or BIOLOGICAL PROPERTY
(endpoint – dependent variable – Y)

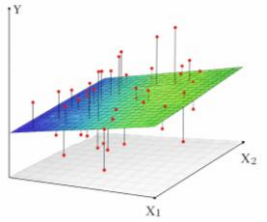
QSAR models – Different Modelling Approaches



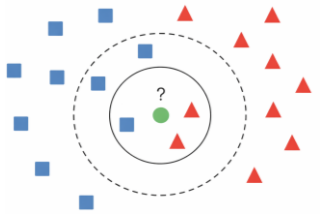
Fragments-based approach: also known as group contribution, each fragment have a specific value/contribution to the endpoint of interest.



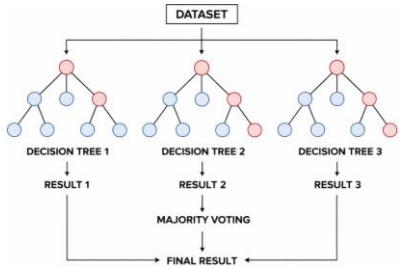
Simple linear regression: the endpoint of interest is directly correlated and predicted by a single descriptor.



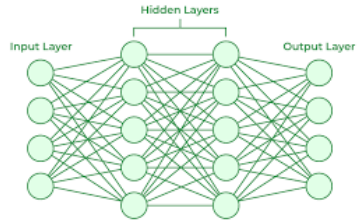
Multiple linear regression: The endpoint of interest is directly correlated and predicted by several descriptors simultaneously.



k Nearest Neighbors (kNN): The parameter is predicted from the molecules identified as the nearest to the target molecule.



Random Forest: The prediction is made by aggregating the results from multiple decision trees, where each tree is trained on a random subset of the data. The final prediction is determined through majority voting (for classification) or averaging (for regression) the outcomes of all trees in the forest.



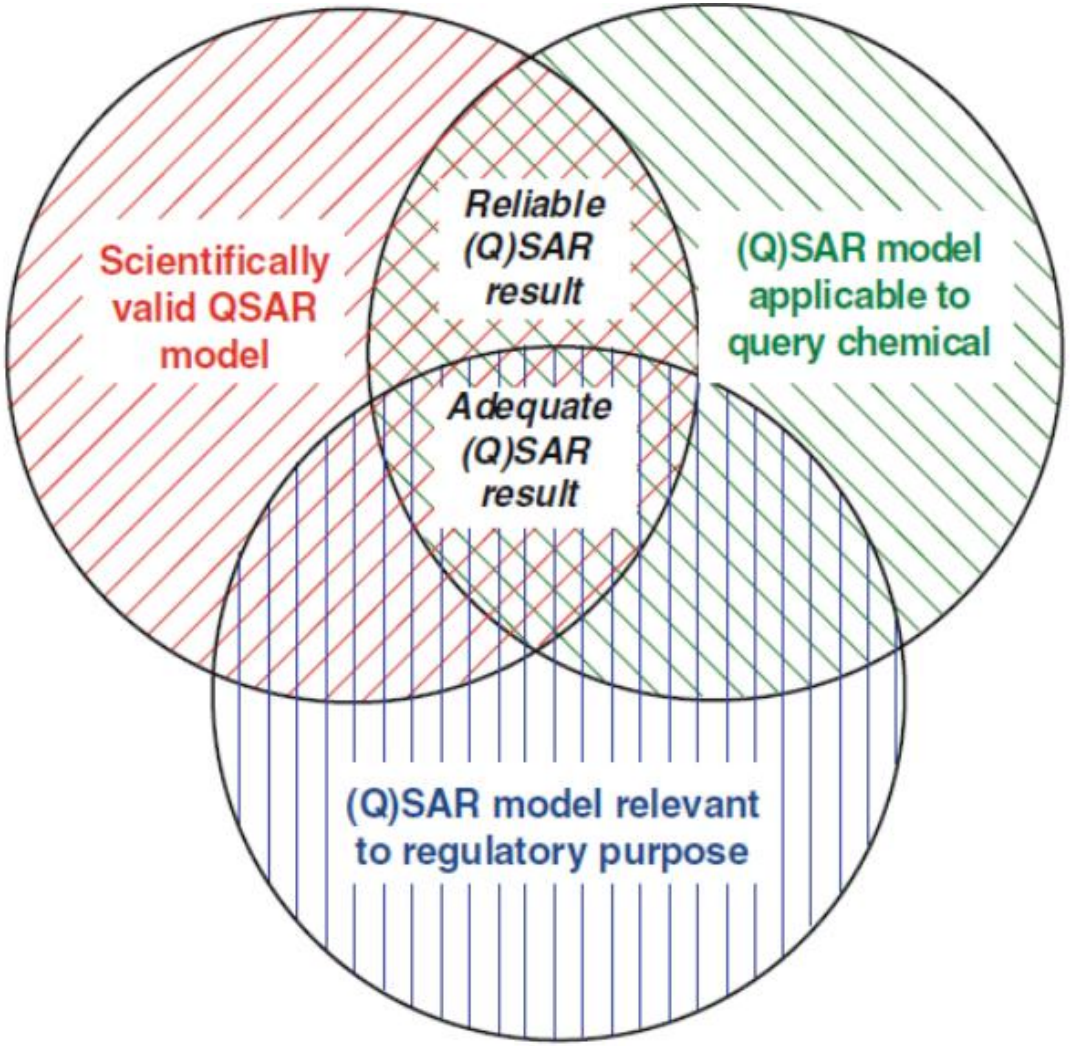
Artificial Neural Networks: System working as the neurals of a human brain.

For more information on methodologies in QSARs development:
Pirhadi et al., 2015, https://doi.org/10.1039/C5RA10729F
Muratov et al., 2020, https://doi.org/10.1039/D0CS00098A

QSAR models – Applicability Domain (AD)

- Limitations in terms of the types of chemical structures, physicochemical properties and mechanisms of action for which the models can generate reliable predictions;
- "*The applicability domain of a (Q)SAR model is the response and chemical structure space in which the model makes predictions with a given reliability*" (Netzeva et al., 2005);
- This space is determined by the characteristics of the data used to build the model, thus by the structural, physico-chemical, and experimental response information of the training set;
- Predictions are considered reliable if the substance falls within the AD. Predictions falling outside the AD are considered extrapolations, and likely being unreliable;
- AD definition is critical in QSAR models development: third OECD principle for the validation of QSARs;
- There is not a universally accepted definition of AD: multiple definitions, each with its own level of restrictiveness (e.g, range-based methods, distance-based methods, etc).

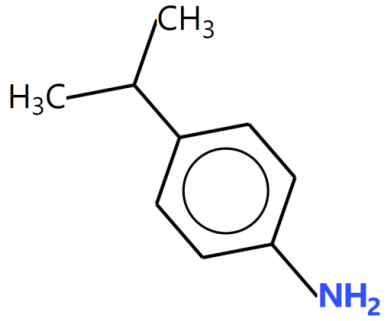
QSAR models – Validity, Applicability, Relevance



Worth AP (2010) The Role of QSAR Methodology in the Regulatory Assessment of Chemicals. T. Puzyn et al. (eds.), Recent Advances in QSAR Studies, pp 367–382. Springer

In silico NAMs – QSARs – Skin sensitization

CAS N°	99-88-7
Name	4-Isopropylaniline
Molecular formula	$C_9H_{13}N$
MW (g/mol)	135.21
SMILES	CC(C)C1=CC=C(C=C1)N



VEGA

CAESAR model: Skin sensitizer
AD analysis suggests reliable prediction.

Danish (Q)SAR Models

Skin sensitizer
AD analysis suggests reliable prediction.

QSAR models – Limitations

Data quantity and quality

- Obtaining sufficient experimental data for model training can be challenging (external validation).
- Garbage In Garbage Out: the quality of the data used to train the model significantly impacts its predictive power.

Applicability domain and structural similarity

- QSAR can provide reliable predictions for compounds within the domain of the training set;
- The model may struggle to predict accurately for compounds with radically different structural features compared to the training set.

Simplification of biological systems

- QSAR models do not consider the complex nature of biological systems, e.g., ADME processes are not evaluated;
- Prediction of hazard and MIE in AOPs, rather than risk quantification and complex endpoint prediction

Overfitting

- QSAR models can fit noise or random fluctuations in the training data rather than capturing the underlying trends, leading to poor model generalization and making the model unreliable for new, unseen data;
- Quantification of prediction uncertainty.

Interpretability and transparency

- They often lack the ability to provide clear mechanistic insights into the relationship between molecular structure and activity, thereby limiting their trust among non-expert users and use of predictions for decision-making.

Difficult to model substances and high-tier endpoints

- Currently available models are developed based on organic, mono-constituent substances. Inorganics, multi-constituents, mixtures, stereoisomers, salts, UCVBs (unknown or variable composition, complex reaction products or of biological materials)?
- Need for more robust models for high-tier endpoints.

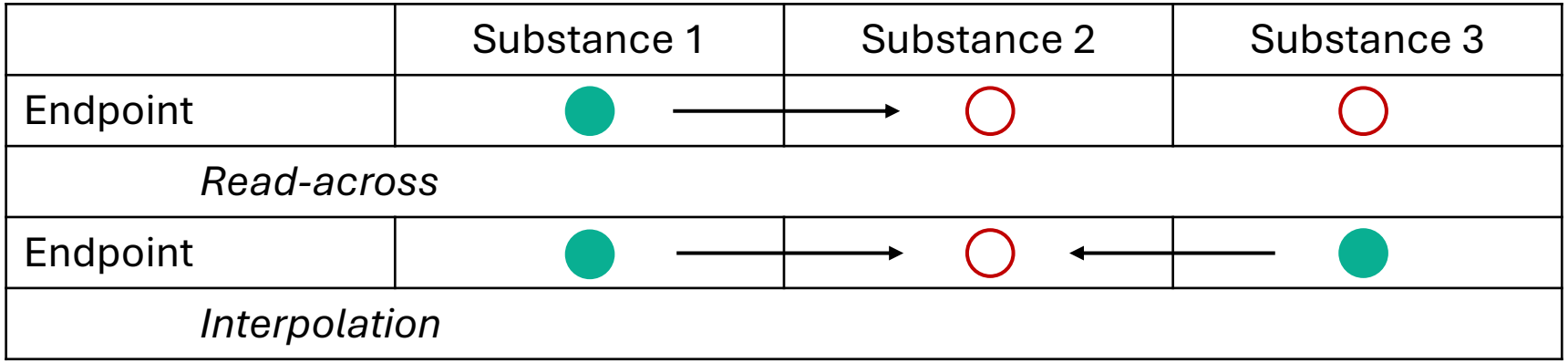
In silico NAMs – Read-Across

Methodology to predict the endpoint for one substance (target substance) by using data from one or more substances (source substance(s)) for which experimental data are available.

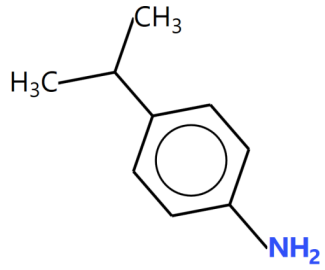
- Categories: interpolation
- Analogues

Structural similarity is a pre-requisite for any grouping and read-across approach.

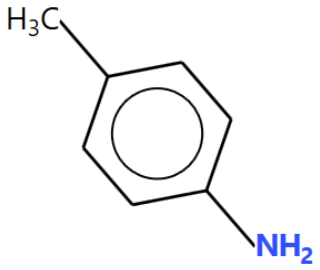
Additionally, analogy should be evaluated also through physicochemical, ADME and mechanistic similarities.



In silico NAMs – Read-across – Skin sensitization



Target substance	
CAS N°	99-88-7
Name	4-Isopropylaniline



Source substance	
CAS N°	106-49-0
Name	*p*-Toluidine

Selection criteria for analogues:

- Structural similarity: ~ 70%
- Aniline
- Substitution in the *para* position
- Similar mechanistic and metabolic properties: e.g., aniline derivatives are metabolically activated by *N-oxidation* in the skin
- Experimental data on skin sensitization

Positive in *in vivo* GPMT (Guinea Pig Maximisation Test)

Read-Across – Limitations

- Assumption that similar chemicals show similar properties and/or toxicities;
- Structural or mechanistic differences might be subtle but critical for the endpoint under scrutiny;
- Need to identify one or even more analogues, since all potential toxicophores in the target substance must be considered and evaluated;
- High-quality data must be available: need to evaluate validity and reliability of the study;
- For novel or uncommon target substances, it may be impractical to find appropriate source substances for Read-Across;
- Definition of similarity:
  - Common functional groups
  - Common precursor or degradation products/metabolites
  - Common physico-chemical and/or biological properties

Battery of *in silico* NAMs

Applying multiple *in silico* methods provides a more robust and scientifically sound assessment than relying on a single approach (Weight of Evidence, WoE).

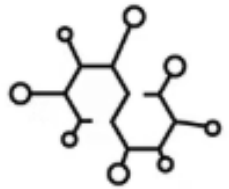
- **SARs / Profilers**
  Identification of structural alerts and mechanistic features linked to potential hazards.
- **QSARs**
  Generate predictions using validated statistical algorithms.
- **Read-Across**
  Leverage experimental data from structurally similar analogues to support predictions.

Weight of Evidence

- Integration of mechanistic insights, statistical predictions, and experimental analogue data;
- Reduces uncertainty by cross-validating independent lines of evidence;
- Enhances transparency and scientific credibility.

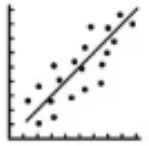
When results are coherent across SARs, QSARs, and Read-Across, confidence in the prediction is significantly increased, supporting more reliable regulatory and risk assessment decisions.

Weight of Evidence (WoE)



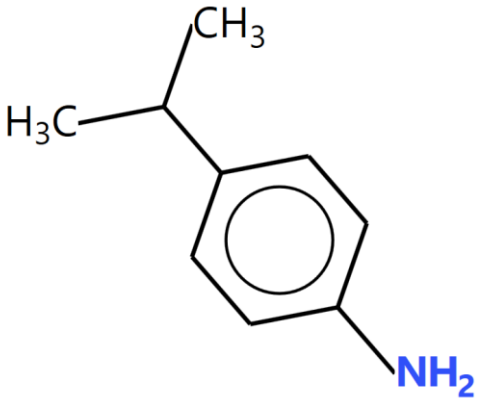
Structure-Activity Relationship (SARs) or Profilers

Alerts found



Quantitative Structure-Activity Relationship (QSAR) Models

Predicted as positive



Read-Across

Close analogue tested as positive

A WoE analysis of the experimental data on structural analogue, and (Q)SAR models predictions, gives a strong indication that **4-isopropylaniline would very likely be a skin sensitizer**.

Regulatory frameworks

Support for the safety of chemical products in multiple markets: industrial chemicals, biocides, plant protection products, cosmetics, pharmaceuticals, food products, etc.

Different regulations, different scopes, different requirements;

The same substance may be assessed separately under different legislation;

Acceptance of *in silico* NAMs can vary across sectors.

One Substance, One Assessment (**OSOA**) as part of the CSS:

- Each dataset is submitted once and reused across all EU regulatory frameworks, ensuring consistency and avoiding duplication of evaluations;
- A NAM accepted once is recognized across sectors to inform multiple regulatory decisions without additional testing;
- Greater regulatory alignment strengthens trust and uptake of integrated, non-animal assessment approaches.

Strategy selection

SARs and QSARs

- There is not a single model able to provide reliable predictions for the entire chemical universe;
- Commercial vs free-to-use models; Online vs standalone tools;
- Model implementation in software facilitates their use, analysis, and acceptance.

Read-Across

- Selection of appropriate source chemicals often involves expert judgment, which introduce subjectivity and inconsistencies in how chemicals are selected, leading to biased or questionable predictions;
- Different regulatory agencies may have different criteria for acceptable Read-Across approaches. The absence of standardized protocols can lead to inconsistencies in how Read-Across is applied and interpreted across different contexts.

Example: REACH

- REACH Regulation is the main EU law to protect human health and the environment from the risks that can be posed by chemicals;

- Substances placed on the market in quantities >= 1 tonne per year must be registered in the form of a dossier. The registration requires information on uses, toxicological and ecotoxicological properties, etc;

- New (Q)SAR Assessment Framework (QAF) to provide a systematic and harmonised framework for the regulatory assessment of (Q)SAR models, predictions, and results based on multiple predictions:
    - Newly defines the principles for the assessment of (Q)SAR predictions and results based on multiple predictions
    - (Q)SAR Model Reporting Format (QMRF)
    - (Q)SAR Prediction Reporting Format (QPRF)

OECD Principles for the validation, for regulatory purposes, of QSAR models

1. **A defined endpoint**: every QSAR model must predict a clearly defined endpoint;

2. **An unambiguous algorithm**: need for transparency and reproducibility of the process used to generate predictions;

3. **A defined applicability domain**: need to define the space where the model is expected to provide reliable predictions for new chemical substances;

4. **Appropriate measures of goodness-of-fit, robustness, and predictivity**: a valid QSAR must be supported by strong statistical evidence of its performance;

5. **A mechanistic interpretation, if possible**: definition of a plausible mechanistic link between the molecular descriptors selected in the model and the predicted response.

QSAR model and prediction checklists (QAF)

Is a QSAR model valid?

Model 1

when more than one model is considered, add a comment here to identify to which model the checklist refers to (e.g. model name)

Principle	Assessment element	Outcome	Comments
Defined endpoint			
1.1	Clear scientific and regulatory purpose		
1.2	Transparency of the underlying experimental data		
1.3	Quality of the underlying experimental data		
Unambiguous algorithm			
2.1	Description of the algorithm and/or software		
2.2	Inputs and other options		
2.3	Model accessibility		
Defined domain of applicability			
3.1	Clear definition of the applicability domain and limitations of the model		
Appropriate measures of goodness-of-fit, robustness and predictivity			
4.1	Goodness-of-fit, robustness		
4.2	Predictivity		
Mechanistic interpretation			
5.1	Plausibility of the mechanistic interpretation		

Is a QSAR prediction valid?

Prediction 1

when more than one prediction is considered, add a comment here to identify to which prediction the checklist refers to (e.g. model name and/or predicted structure)

Principle	Assessment element	Weight	Outcome	Uncertainty	Comments
		Default values precompiled			
Correct input(s) to the model					
1.1	Clear and complete description of the input and model settings	High			
1.2	Input representative of the substance under analysis	High			
1.3	Reliable input (parameters)	Medium			
Substance within the applicability domain of a valid model					
2.1	Substance within the applicability domain	High			
2.2	Any other limitation of the model is considered	High			
Reliable prediction					
3.1	Reproducibility	High			
3.2	Overall performance of the model	Medium			
3.3	Fit within the physicochemical, structural and response spaces of the training set of the model	Medium			
3.4	Performance of the model for similar substances	High			
3.5	Mechanistic and/or metabolic considerations	High			
3.6	Consistency of information	High			
Outcome is fit for the regulatory purpose					
4.1	Compliance with additional requirements	High			
4.2	Correspondence between predicted property and property required by the regulation	High			
4.3	Decidability within the specific framework	High			
Conclusion on the individual prediction					
Uncertainty					
Outcome of the assessment (individual prediction)					
Comments					

IUCLID

IUCLID 6

IUCLID supports organizations to manage scientific data on chemicals in a regulatory context

>	1 General information*	5
>	2 Classification & Labelling and PBT assessment*	3
>	3 Manufacture, use and exposure*	16
>	4 Physical and chemical properties*	35
>	5 Environmental fate and pathways*	8
>	6 Ecotoxicological information*	10
>	7 Toxicological information*	48
	8 Analytical methods	
>	11 Guidance on safe use*	1
	12 Literature search	
>	13 Assessment reports*	3
>	14 Information requirements	
	Inherited templates	

Endpoint
biodegradation in water: ready biodegradability
Type of information
(Q)SAR
Adequacy of study
key study

Endpoint
biodegradation in water: ready biodegradability
Type of information
(Q)SAR
Adequacy of study
supporting study

Endpoint
biodegradation in water: ready biodegradability
Type of information
(Q)SAR
Adequacy of study
weight of evidence

Endpoint
biodegradation in water: ready biodegradability
Type of information
(Q)SAR
Adequacy of study
disregarded due to major methodological deficiencies

Example: ICH-M7

- International Conference on Harmonization (ICH) M7 for the assessment and control of DNA reactive (genotoxic) impurities in pharmaceutical products.

- Use of *in silico*-based systems as part of the hazard identification strategy to predict bacterial mutation (Ames test).

- The guideline requires the use of two complementary approaches:
  - an expert rule-based method (SAR), and
  - a statistical-based model (QSAR)

"the absence of structural alerts from two complementary (Q)SAR approaches (expert rule-based and statistical) is sufficient to conclude that an impurity is of no mutagenic concern and no further testing is recommended"

Example: Cosmetic Regulation

- Regulation on cosmetic products is the main regulatory framework for finished cosmetic products when placed on the EU market aiming to strengthen the safety of cosmetic products;
- Animal testing ban on finished cosmetic products and ingredients: use of NAMs;
- Less stringent documentation, but *in silico* approaches are advised to be used alongside other alternatives, e.g., *in vitro* and *in chemico* (WoE approach);
- Botanical extracts: complex mixtures showing inter-batch variations in composition;

Key areas for advancement

Can explainable AI support this transition?

- Development of models for difficult to model substances (salts, mixtures, UVCBs, stereoisomers);
- Development of models for high-tier endpoints (e.g., DART);
- Integration of QSARs within the AOP framework: simultaneous modelling of multiple MIEs to achieve higher human relevance;
- Increase reliability of model predictions: transparency across non-expert users, mechanistic interpretations, quantification of prediction uncertainty;
- Effective modelling use large pool of data;
- Broader adoption of open-source tools and models implementation in software.

Thank you for your attention!

marco.evangelista@cehtra.com