AFSA COLLABORATION

in silico tools & read-across

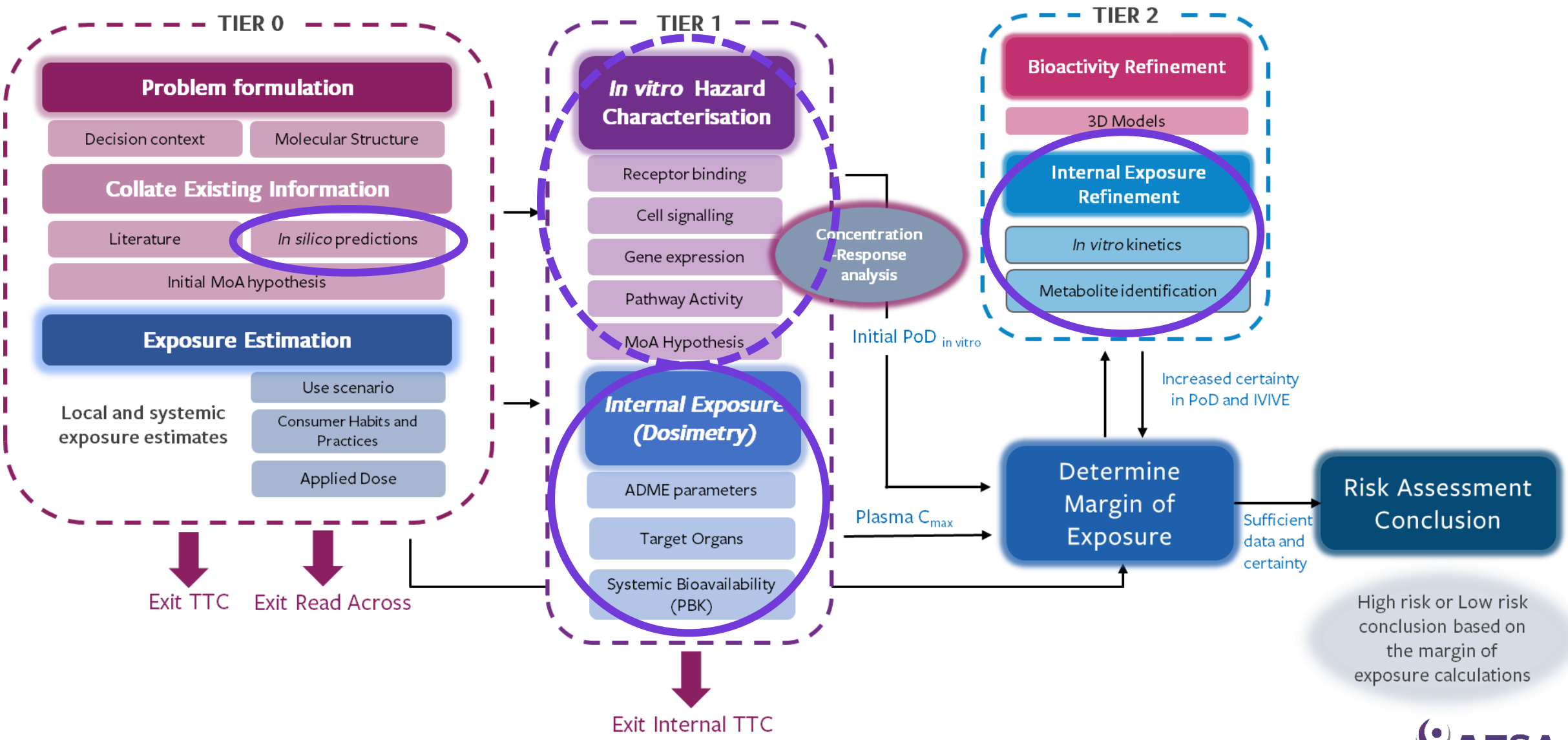
Jin Li

Unilever Safety & Environmental Assurance Centre

Unilever

16 Dec 2022

NGRA Framework

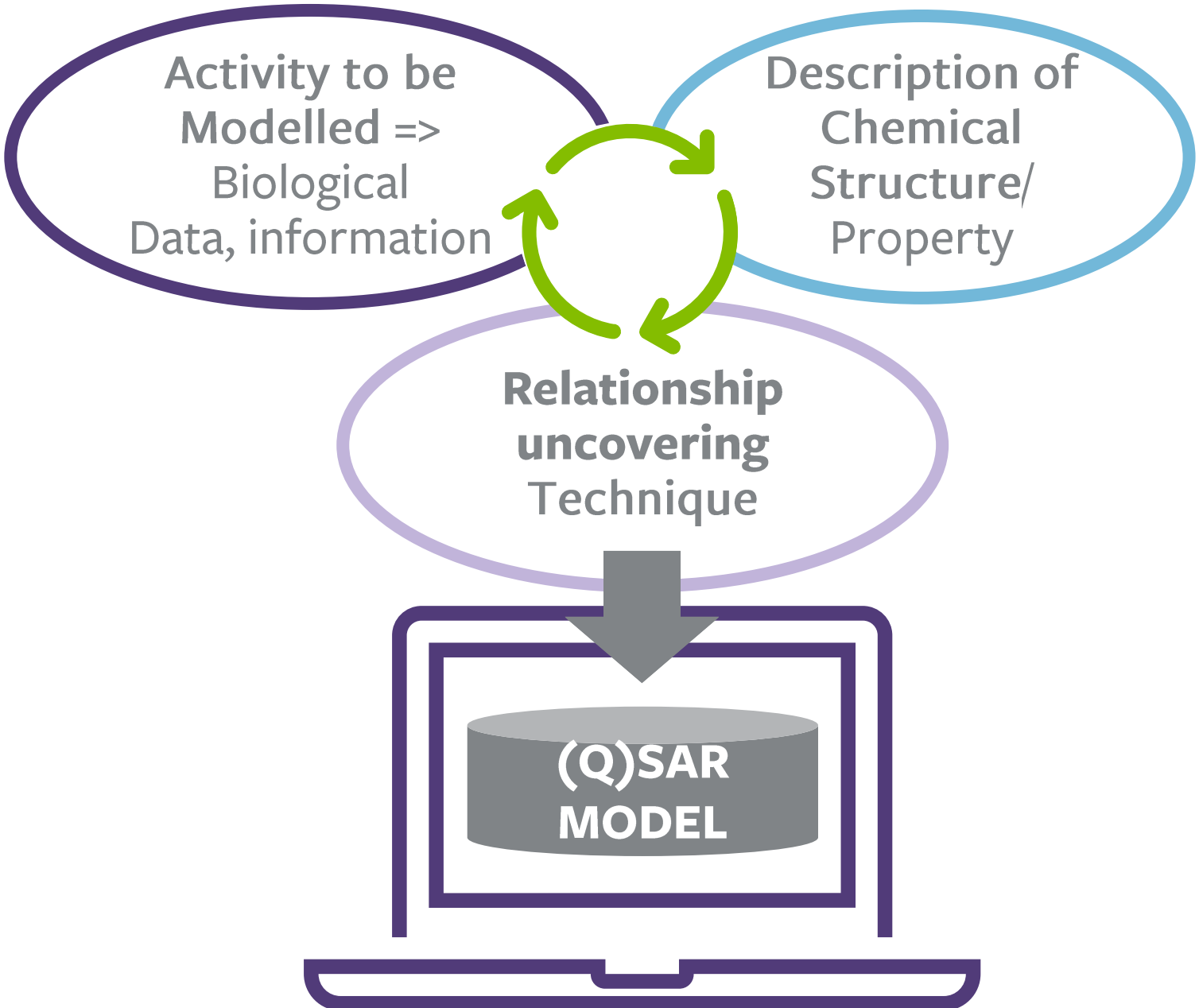


Outline

- In silico tools
  - Definition
  - Develop
  - Apply
- Read-across
  - Process/framework
  - Target profiling
  - Source ID and evaluation
  - R-A outcome

Part 1: In Silico (prediction) Tools

In silico prediction models



Typical process of in silico modelling

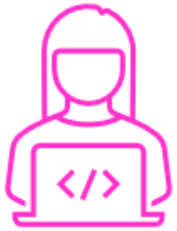
Find data

Curate data

Develop model using training set

Validate model (internal and external)

Run new data



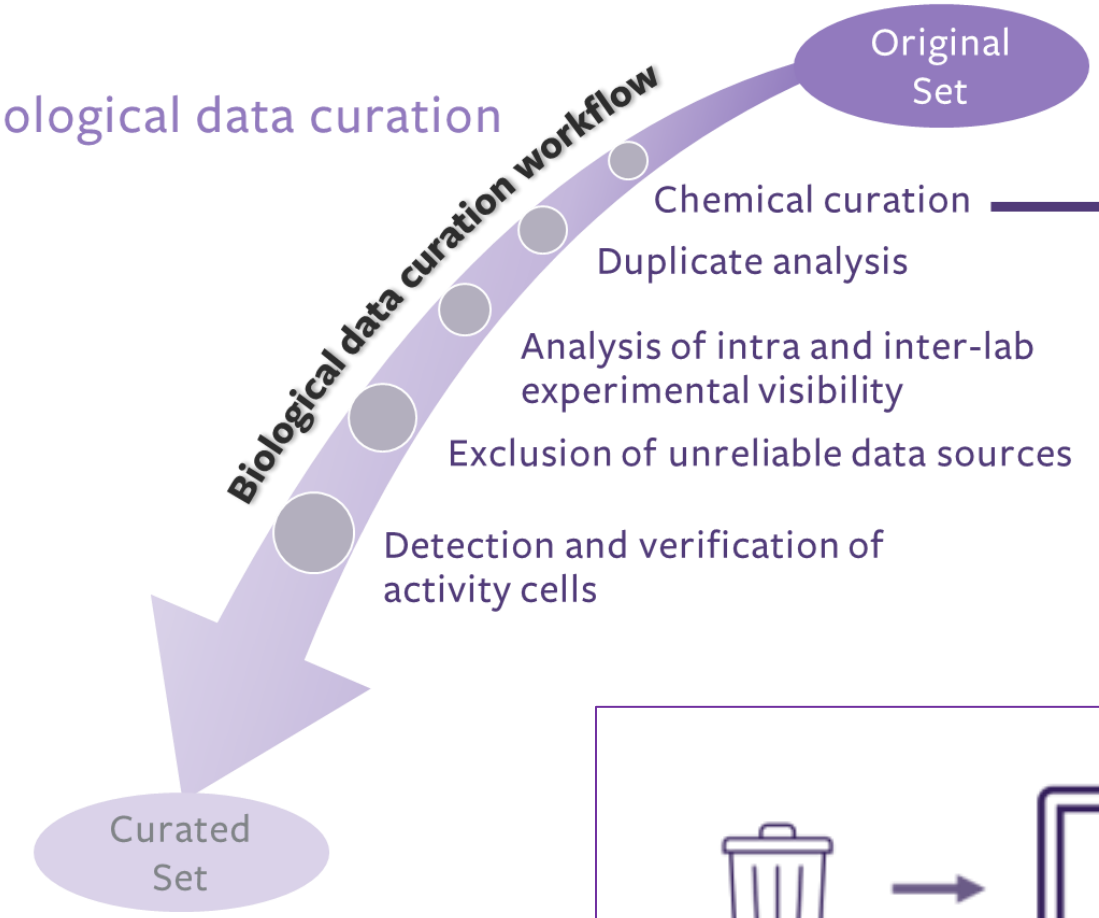
Use prediction

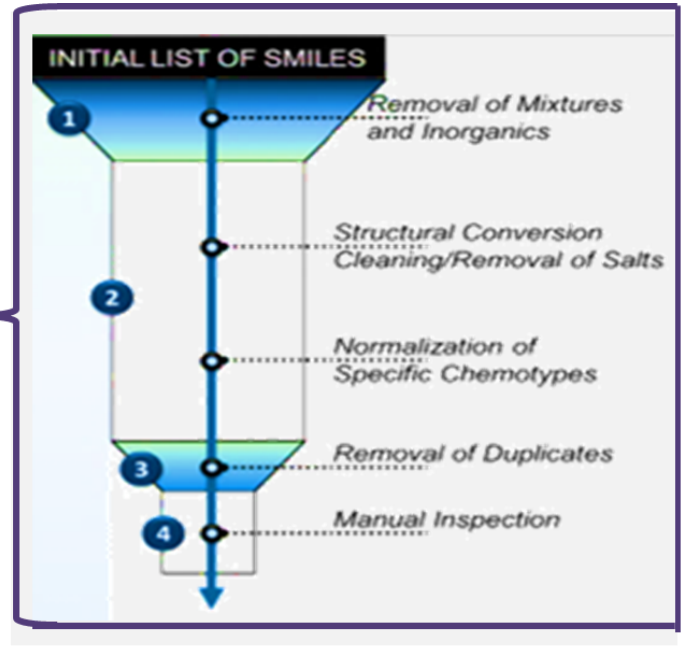
Develop

Apply

Develop : Data Curation

Biological data curation





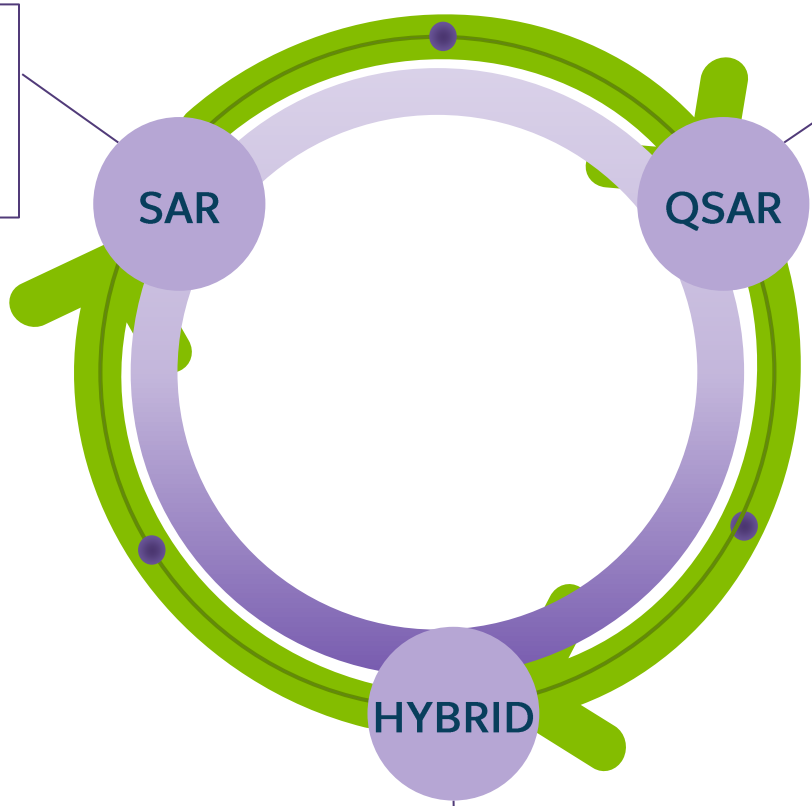
Chemical data curation workflow

Adapted from Fourches et al. 2015. Curation of chemogenomics data. Nat Chem Biol 11, 535. https://doi.org/10.1038/nchembio.1881

AFSA COLLABORATION

Develop : types (techniques) of in silico models

- Decision tree-based
- Expert knowledge-based



- Statistical models
- Machine learning

Combination of SAR and QSAR

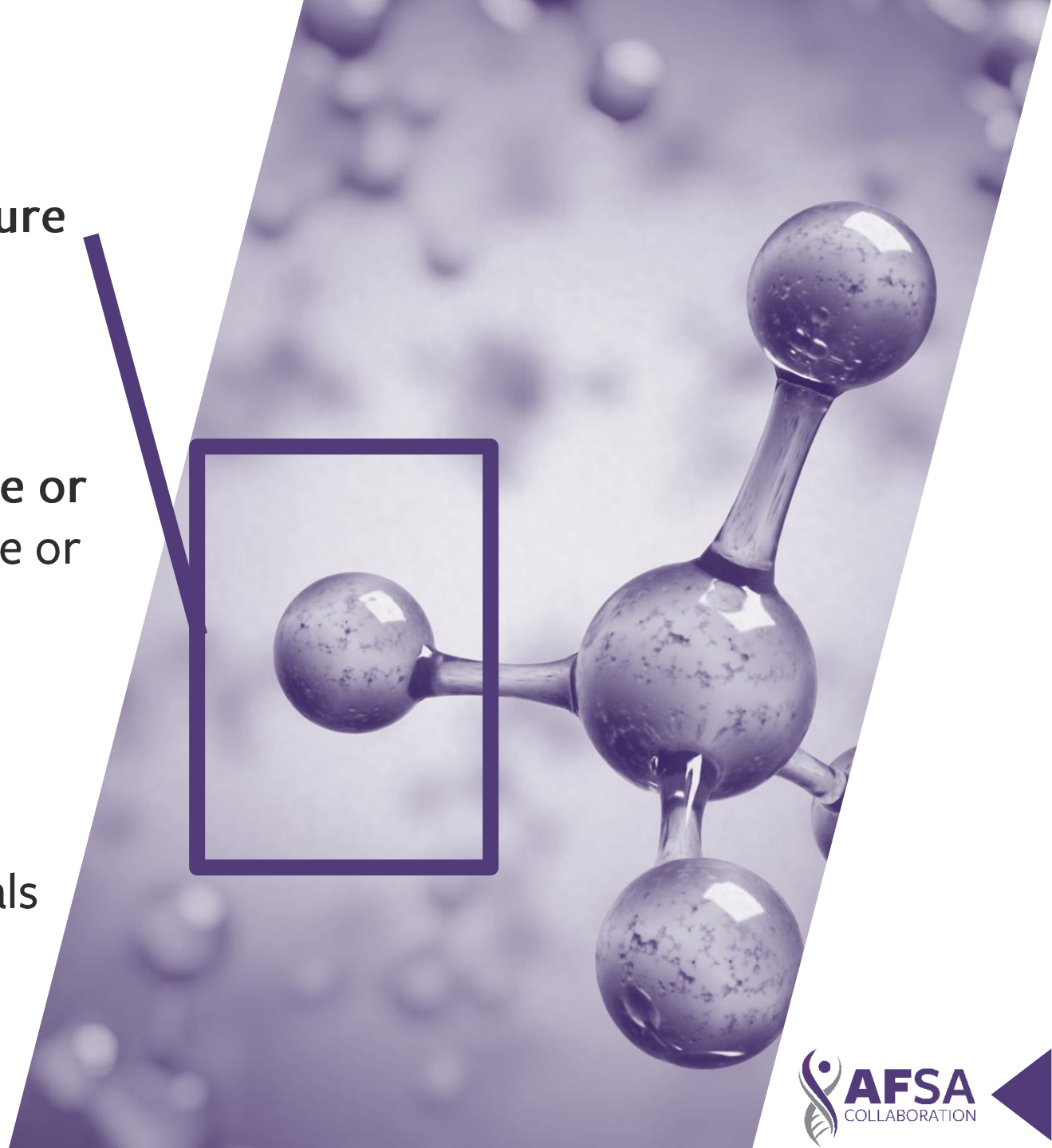
SAR Models

A SAR model uses a chemical's **(sub)-structure** to predict its (qualitative) biological activity-toxicity.

Very often based **on mechanistic knowledge or expert knowledge.** "Rules" relating presence or absence of activity to (a) specific chemical feature(s) are thus created and encoded.

⇒ also called "Alert models"

⇒ also widely used for grouping chemicals into categories which share the same mechanism of action ("Profilers")



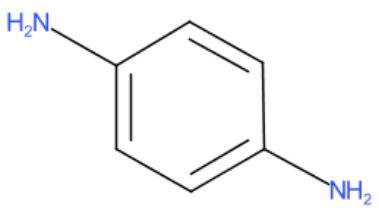
SAR - Decision tree-based expert system

ex. Toxtree Skin sensitization reactivity domain

Available structure attributes

Alert for Acyl Transfer age...	NO
Alert for Michael Acceptor i...	YES
Alert for SN2 identified.	NO
Alert for SNAr Identified.	NO
Alert for Schiff base forma...	NO
No skin sensitisation reacti...	NO
SMILES	Nc1ccc(N)cc1
cdk:Comment	Created from SMILES
cdk:Title	

Structure diagram



Toxic Hazard

Alert for SNAr Identified.

Alert for Schiff base formation identified.

Alert for Michael Acceptor identified.

Alert for Acyl Transfer agent identified.

Alert for SN2 identified.

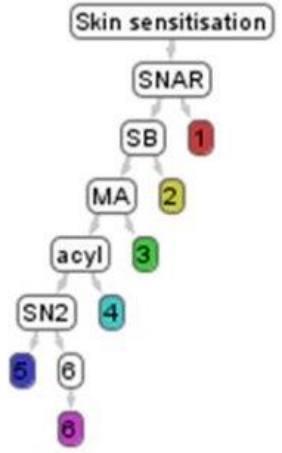
No skin sensitisation reactivity domains alerts identified.

☑ Verbose explanation

Skin sensitisation reactivity domains

- QSNAR.SNAr-Nucleophilic Aromatic Substitution **No** Nc1ccc(N)cc1
- QSB.Schiff Base Formation **No** Nc1ccc(N)cc1
- QMA.Michael Acceptor **Yes** Class Alert for Michael Acceptor identified. Nc1ccc(N)cc1
- Qacyl.Acyl Transfer Agents **No** Nc1ccc(N)cc1
- QSN2.SN2-Nucleophilic Aliphatic Substitution **No** Nc1ccc(N)cc1
- Q6.At least one alert for skin sensitisation? **Yes** Nc1ccc(N)cc1

- ToxTree applies the decision tree and alerts matched within the structural domains indicate skin sensitization potential



→ Para-phenylene diamine has matched the QMA Michael Acceptor mechanistic domain with at least one alert (Q6) so is considered to be a skin sensitiser

http://toxtree.sourceforge.net/index.html

AFSA COLLABORATION

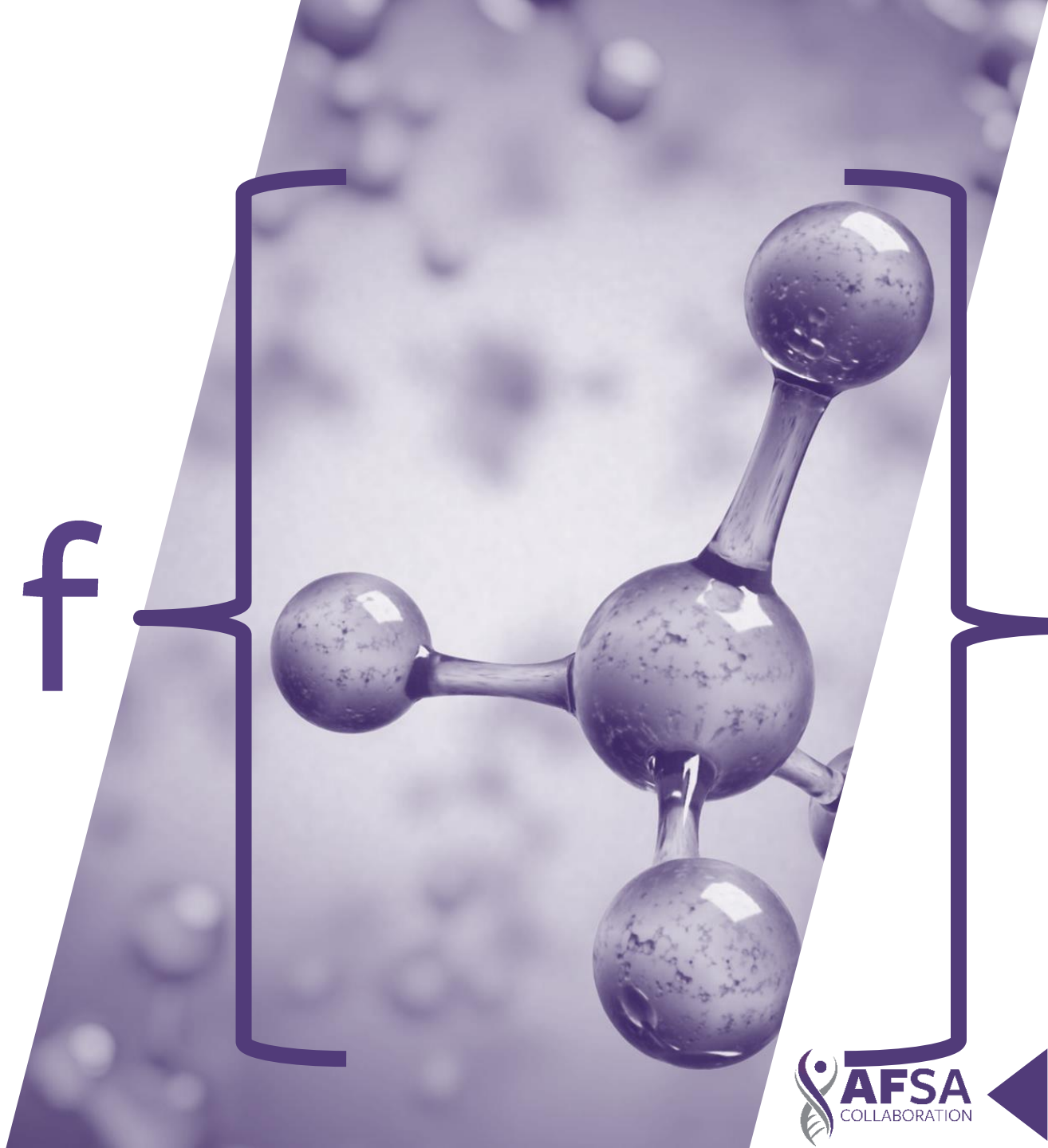
QSAR Models

QSAR differs from SAR in that they :

use quantitative measures of chemical structures (defined as **descriptors**)

+

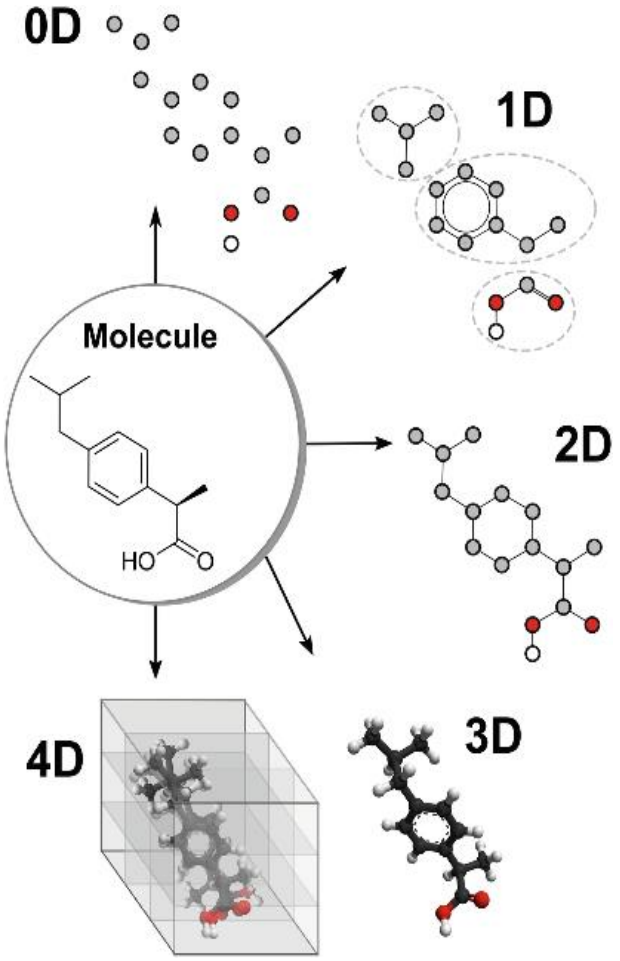
correlate one or more of these to a biological activity of interest using a **statistical technique**



Molecular descriptors

Molecular descriptors are a quantification of the various molecular properties of a chemical compound

Descriptor Types	Description	Examples
Constitutional	They represent a molecular structure, which take into account only chemical composition	molecular weight, the number of atoms and bonds, number of aromatic rings
Electrostatic	They represent properties related to electronic nature of the compound	atomic and partial charges
Topological	They are derived from the topological representation of molecular structures i.e., molecular graph	Wiener descriptors, Kappa shape
Geometric	They are derived from a 3-dimensional graph representation of the molecule, taking into account not only the positions of the atoms but also the connections among them	Geometry, Topology, and Atom-Weights Assembly (GETAWAY) descriptors
Quantum	They express all of the electronic and geometric properties of molecules and their interactions	highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO)



Grisoni et al., (2018) Impact of Molecular Descriptors on Computational Models. In: Brown J. (eds) Computational Chemogenomics. Methods in Molecular Biology, vol 1825.

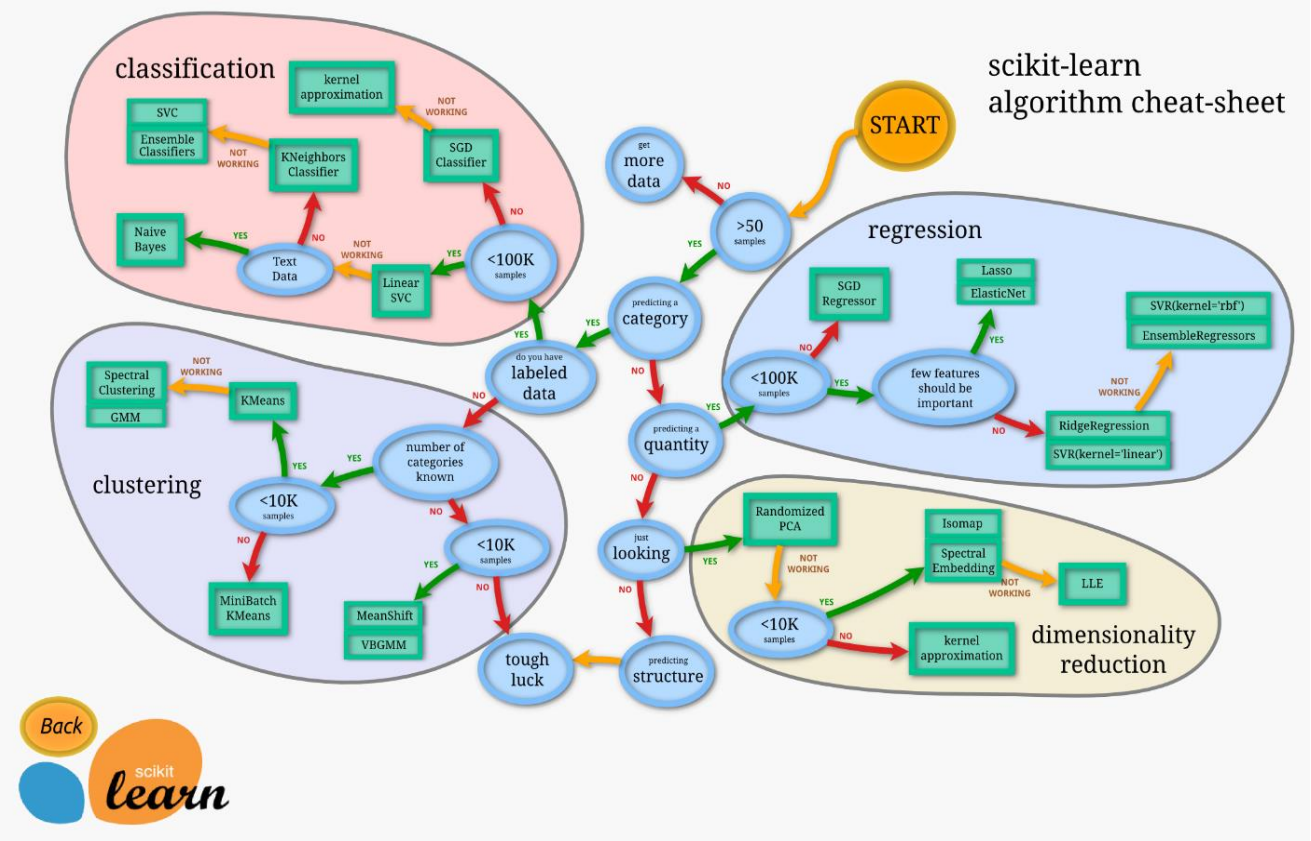
AFSA COLLABORATION

Statistical technique ?

- Some statistical techniques are more suitable for specific types of data or different sizes of datasets

- How to choose:
  - → Flowcharts
  - → Visualization of the data (e.g. using Principal Component Analysis)
  - → Literature for similar problems

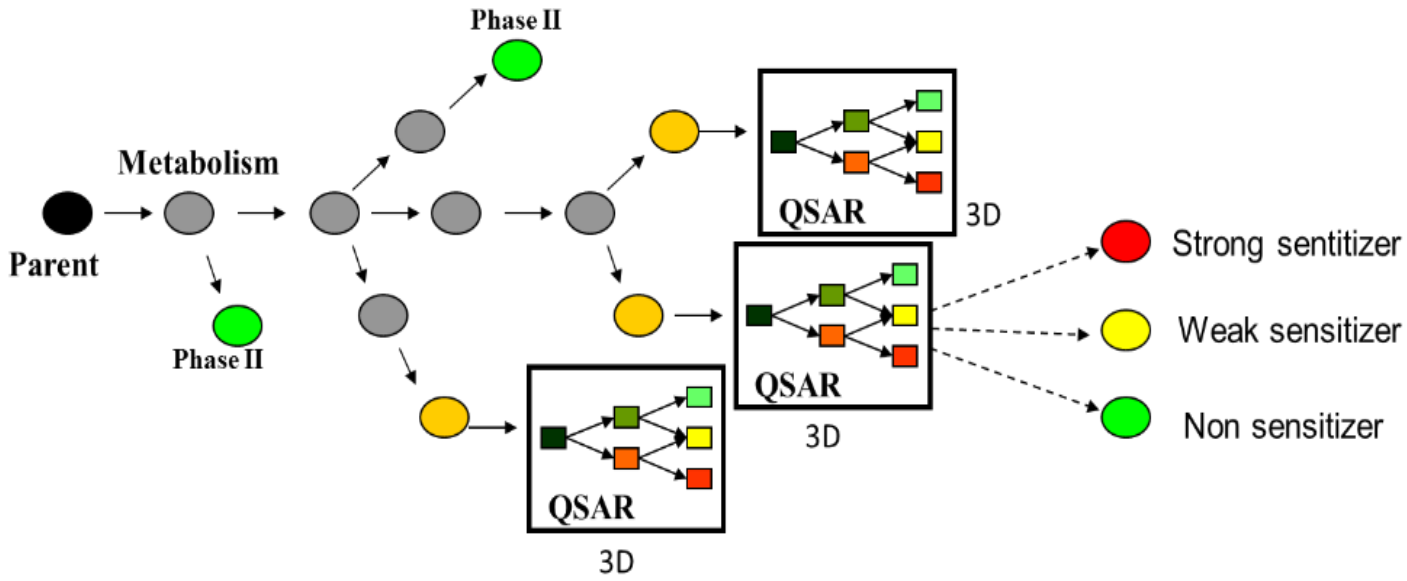
- Usually, several techniques would be tried and the best “performing” one chosen



Best of both : Hybrid Models

ex. TIMES Skin sensitization model with Autooxidation

This figure illustrates different interconnections between simulator of skin metabolism, classification and 3D-QSAR models in TIMES.



TIMES SS assessment

- Matching parent molecule against 420 hierarchical metabolic transformations
- For all matches, reactive or metabolic species and their respective protein (or Phase II) adducts are then generated
- The propagation of metabolism is stopped when protein conjugation reactions classifying the chemical as strong (or weak) sensitizer or Phase II reactions are applied.
- For some reactive species, additional information is required and 3D-QSARs are invoked to determine their sensitization effect.

Consensus models

Models that take the predictions of several (Q)SAR models and combine them to provide a single prediction.

Approaches that provide a consensus prediction include:

- Taking the predominant prediction
- Taking the average prediction
- Combining the predictions into a combined linear regression model

PRO

May provide more accurate and higher confidence predictions

CON

May put alert models at same level as prediction models, be too complex and lack transparency

AFSA COLLABORATION

Typical process of in silico modelling

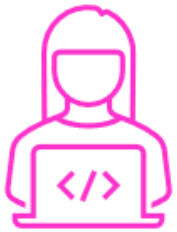
Find data

Curate data

Develop model using training set

Validate model (internal and external)

Run new data



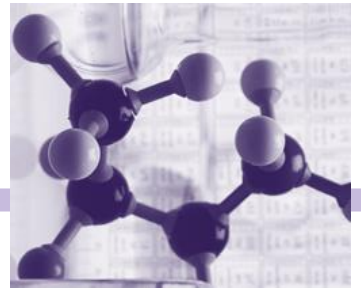
Use prediction

Develop

Apply

Applying in silico predictions

1 **Ensure you have the correct chemical structure for input into the model**

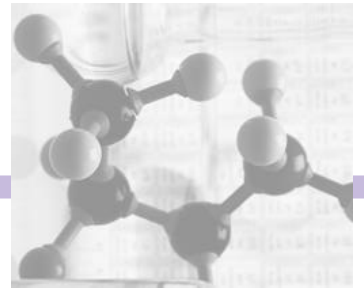


Ex.

- If you are not sure if metabolites may form and whether the parent or one of the metabolites is the active chemical, you can assess data for both the parent and the metabolite(s)
- *In silico* transformation (metabolism) simulators can predict likely metabolites

Applying in silico predictions

1 Ensure you have the correct chemical structure for input into the model



(Q)SAR MODEL

2 Choose a model that is applicable for your endpoint of interest

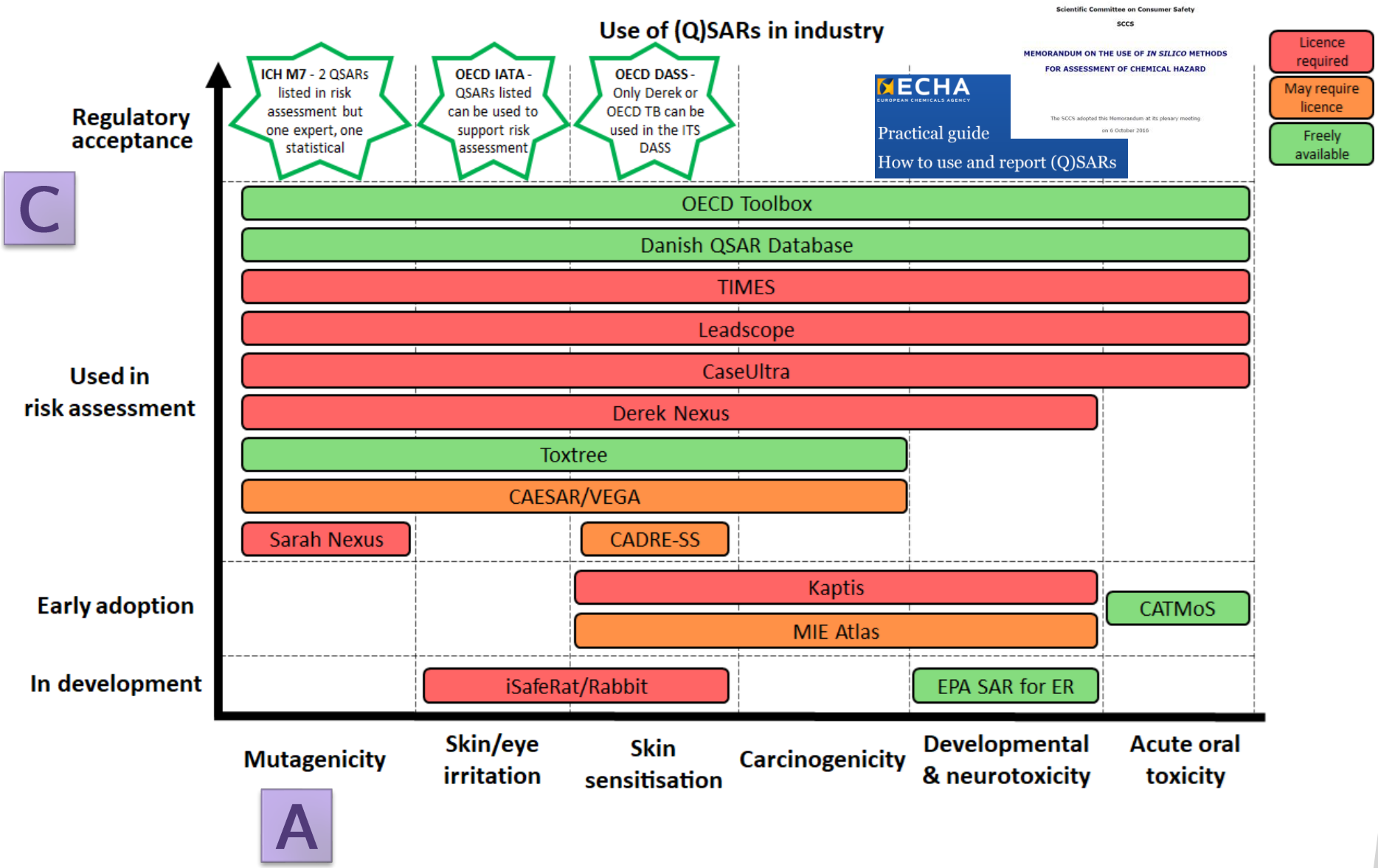
Generate a prediction using the protocol for the model

AFSA COLLABORATION

Choice of the model

C

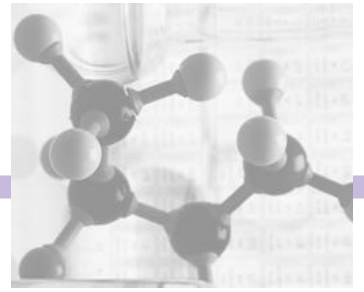
B



A

Applying in silico predictions

1 Ensure you have the correct chemical structure for input into the model



2 Choose a model that is applicable for your endpoint of interest

Generate a prediction using the protocol for the model

(Q)SAR MODEL

3 Assess reliability:

Understand whether the prediction is in the applicability domain of the model

Characterize and document uncertainty

Use the prediction!

AFSA COLLABORATION

Assessing reliability

Applicability Domain	Uncertainty	Variability	Validation	Reporting
Usually dependent on the training set used to develop the model.	Expresses the limitation in knowledge or lack of data. It can be reduced or eliminated.	Refers to inherent heterogeneity in the data. It cannot be reduced but it can be characterised.	Allows to evaluate the predictivity and reliability of the model. It can be internal or external.	(Q)SAR Model Reporting Format (QMRF)* is a harmonised template structured according to the OECD validation principles.

What the prediction can / cannot tell us

Be transparent about it

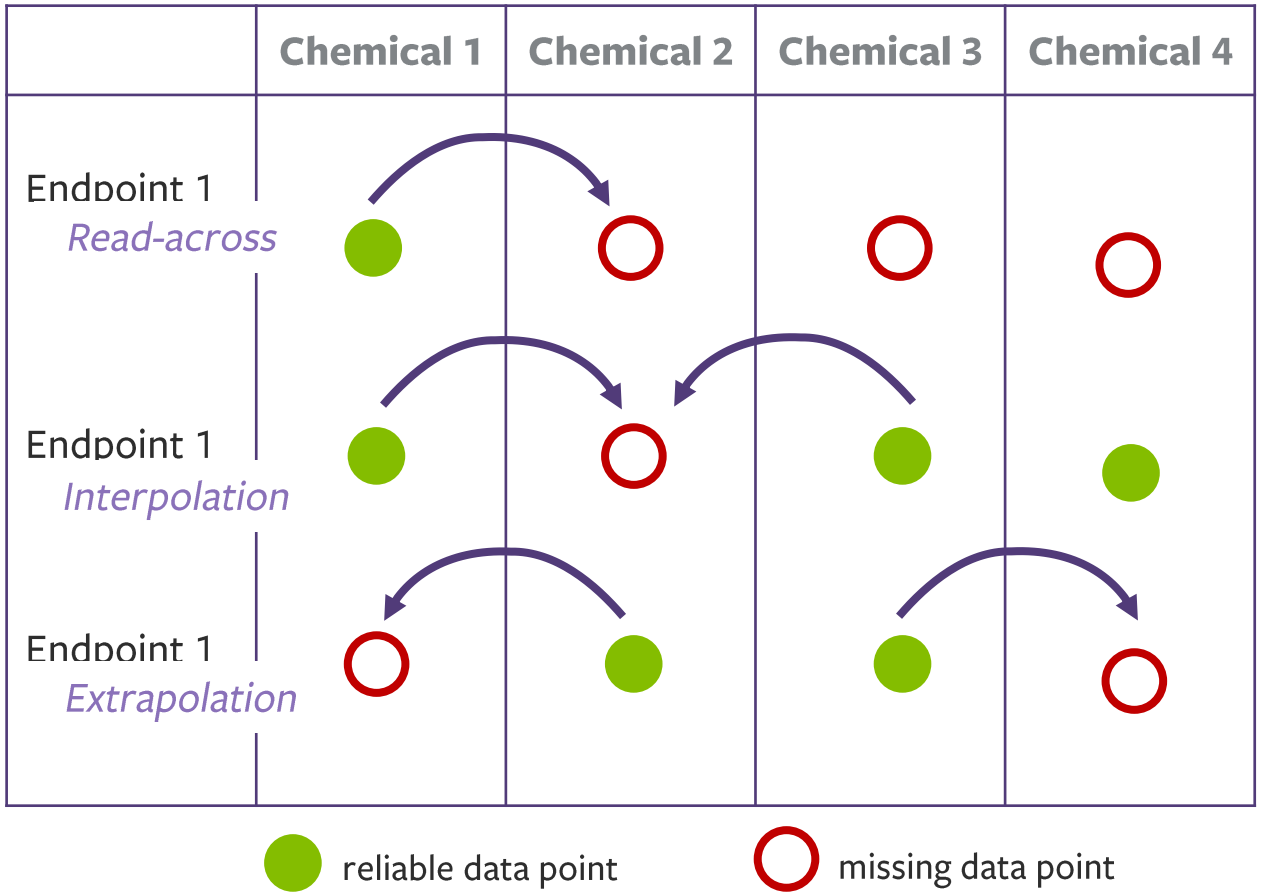
*http://ihcp.jrc.ec.europa.eu/our_databases/jrc-qsar-inventory

AFSA COLLABORATION

Part 2: Read-Across

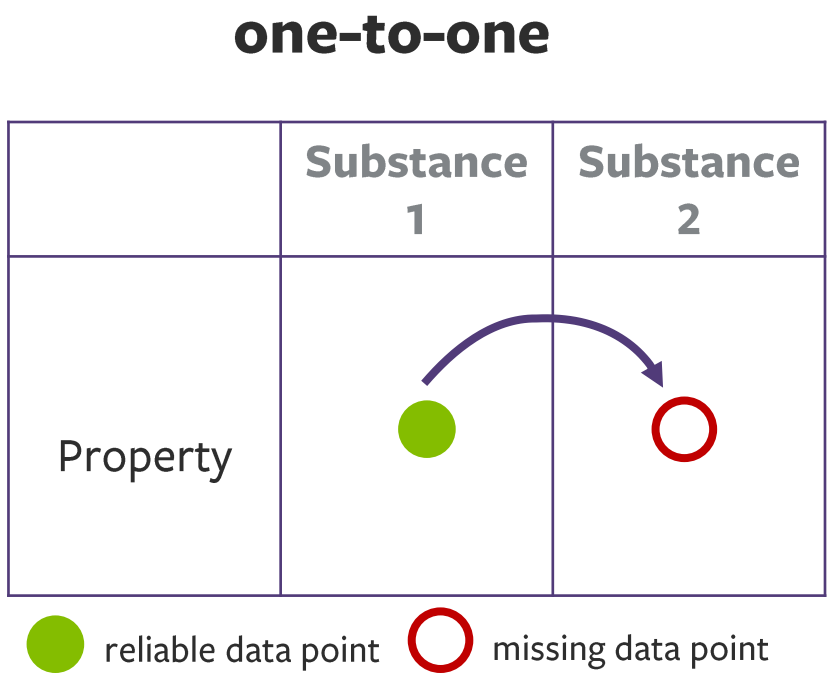
What is read-across?

Read-across is an alternative approach that is used to fill a data gap for a substance (the target), for a specific endpoint, by using the data from another structurally/ mechanistically similar substance (the source).



Read-across approaches

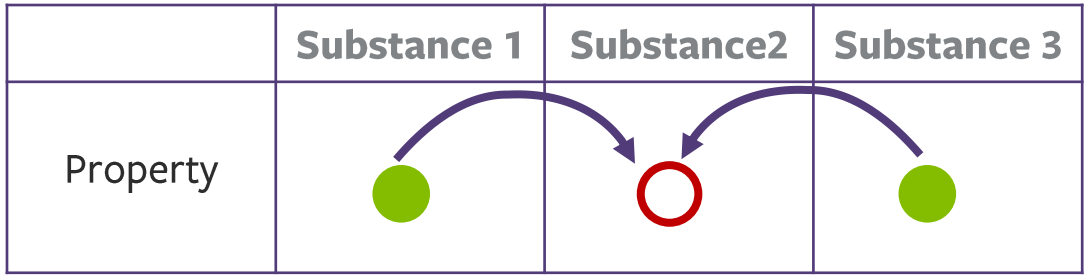
- The way in which source data are used in the read-across is dependent on the available data and the properties of the target and source substances.
- If there is only one source substance with data, this is a one-to-one read-across:



Read-across approaches

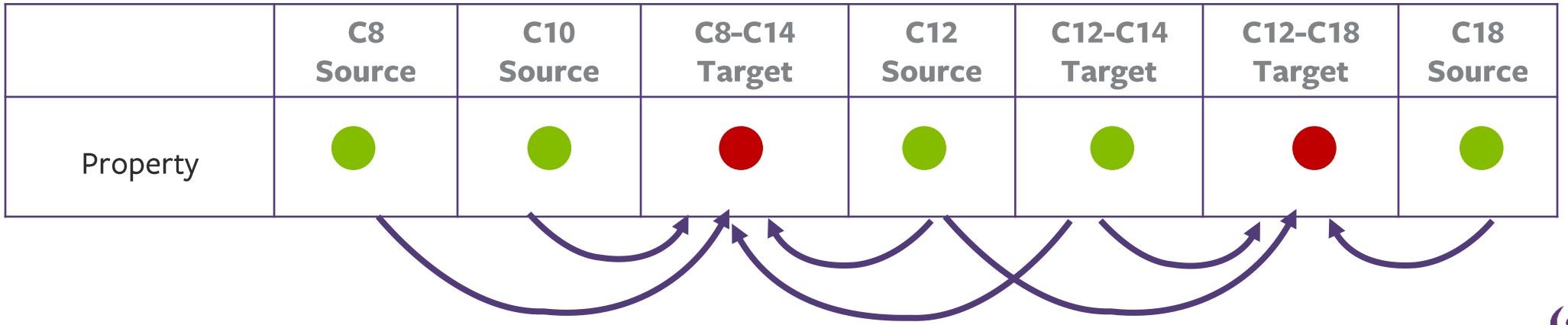
If there are multiple source substances, this is many-to-one:

many-to-one



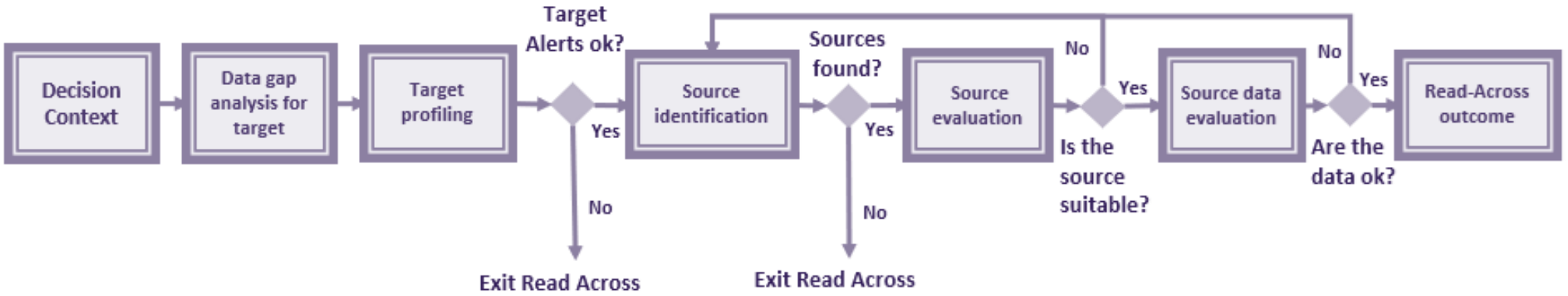
If there are multiple substances which are structurally similar, but which do not follow a trend or pattern in their properties, this is called an **Analogue** approach.

Where there are multiple substances that have similar properties, or which follow a pattern because of structural similarity, these may be considered as a **Group** (or **Category**).



The read-across process / framework

In summary, the key steps involved in read-across are:



- Define decision context
- Data gap analysis for the target
- Define hypothesis
- Target profiling
- Source identification
- Source evaluation
- Source data evaluation
- Read-across outcome

Defining the read-across hypothesis

For any read-across, there must be a hypothesis which describes why it is possible to use the data from a source substance to risk assess the target substance.

If the target and source substances are shown to have the same features, properties, and behavior, the hypothesis is that the target would exhibit the same biological response in an assay as the source substance. Therefore, justifying the use of the source data to support the target.

The hypothesis is supported by all the information gathered from the steps in the read-across and so develops as more information is collected.

The decision context

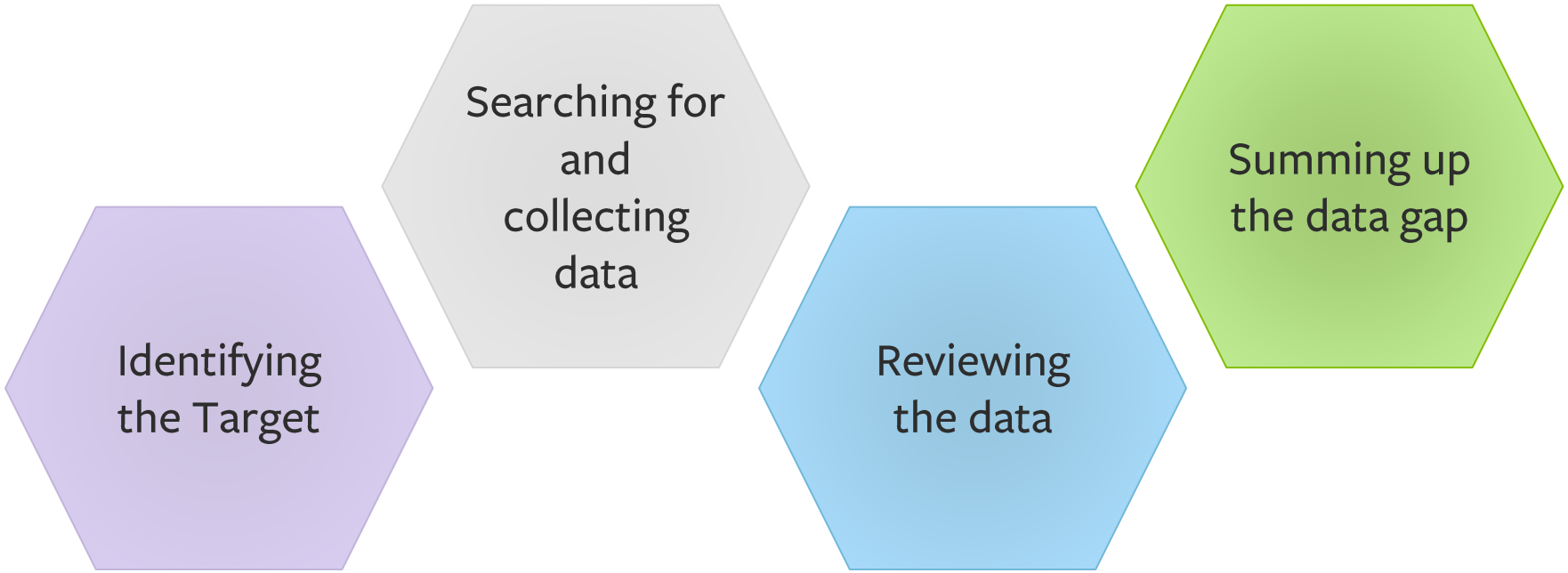
This step in the framework describes the problem and gives a reason why read-across is needed. At this step, it is important to know:

The Purpose	Is it for safety risk assessment or regulatory submission?
Target Details	The substance common name, synonyms, CAS, structure etc.
Exposure Scenario	How is the product containing the ingredient used? How often it is used? How is it administered (dermally, orally, inhaled?) How much is used per use? If it enters the body, how is it metabolised?

AFSA COLLABORATION

Data gap analysis for the target

Before a read-across is performed, it is important to know as much as possible about the target substance. This includes:



All the information collected about the Target can be stored in a series of tables in a data matrix.

Data gap analysis for the target: Summarising the data gaps

It is important to define the data gaps to be filled by read-across.

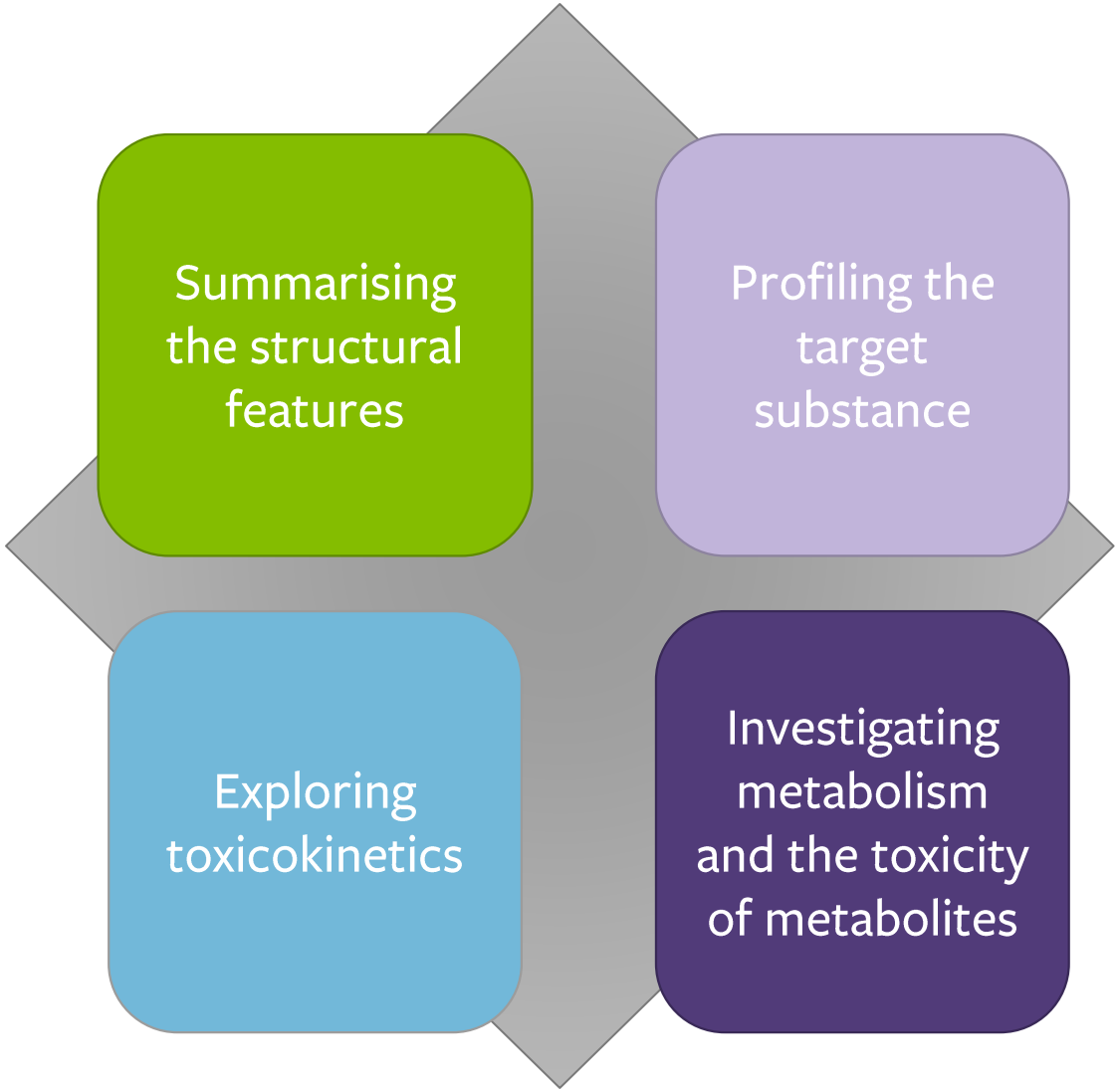
The data gap could be from missing endpoint studies, or from poor quality endpoint data which are not good enough to be used in the risk assessment.

As read-across is endpoint specific, read-across must be performed for each individual data gap / endpoint.

AFSA COLLABORATION

Target profiling

As well as creating a data matrix of existing data, further investigation is needed to completely understand the target substance. This includes:



Target Profiling: Summarising the target (and metabolite toxicity)

As already suggested, it is important to record the output for each stage of the read-across framework. Once all the information about the target is available, a summary can be prepared which gives an overall view of the target substance (and metabolites).

The data in the data matrix can be used to prepare the summary and can include a summary of:

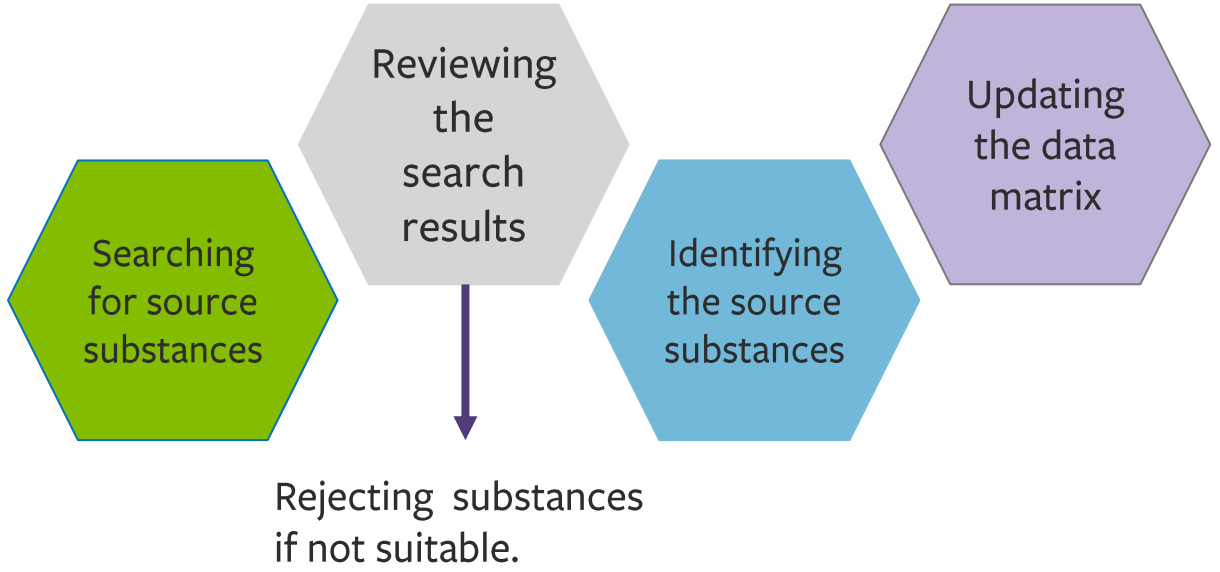
- Structural features.
- Existing toxicity data.
- Predicted toxicity alerts.
- Predicted physchem properties.
- Summary of toxicokinetics predictions.
- Key metabolic biotransformations.
- Metabolite toxicity alerts

AFSA COLLABORATION

Source identification

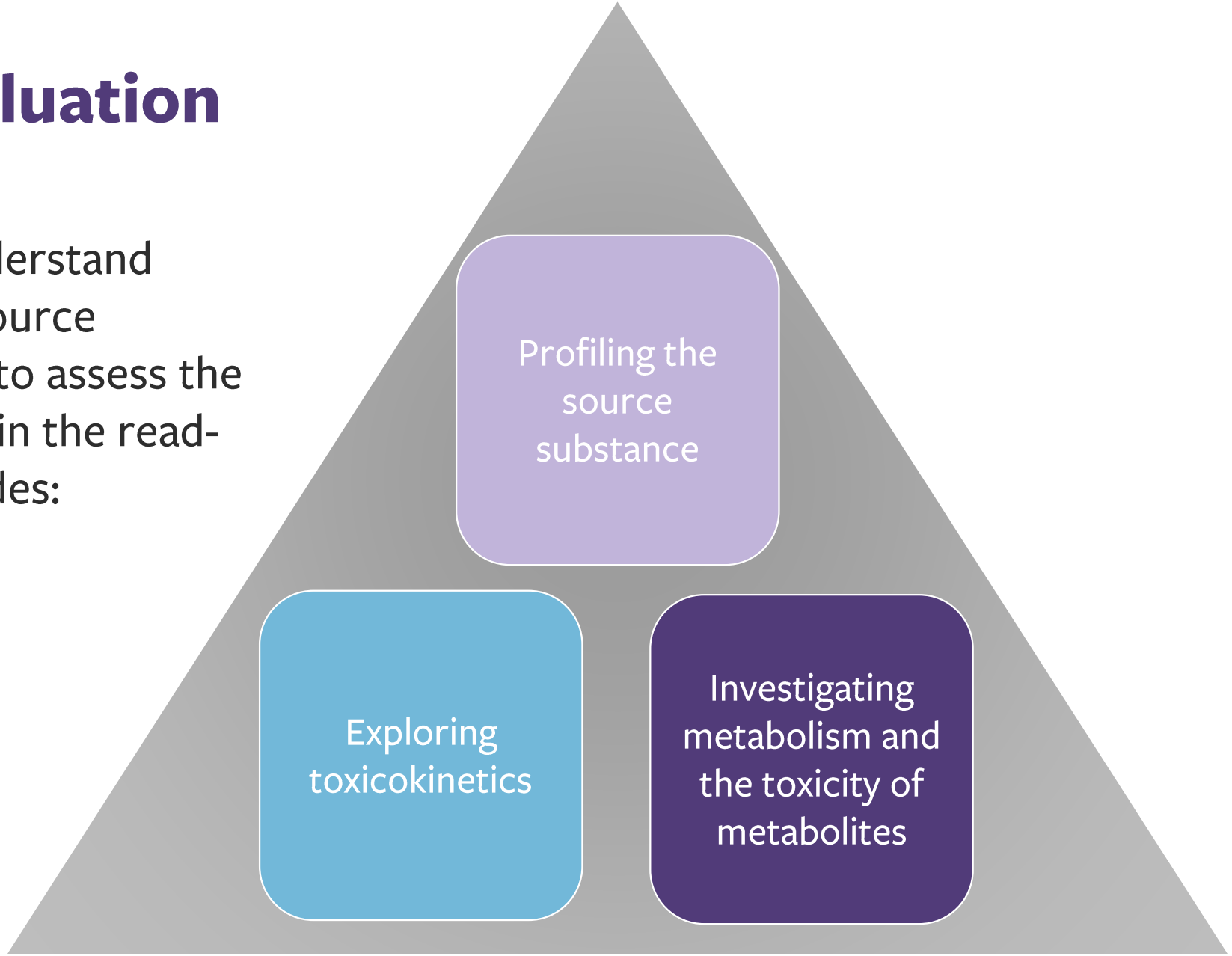
Source substances may be found in literature or regulatory dossiers. In these cases, the source substances are evaluated for their suitability for use in the read-across.

Suitable substances must be identified, and this includes:



Source evaluation

This step is to understand more about the source substance(s) and to assess the suitability for use in the read-across. This includes:



Opportunities to strengthen read-across

Use of New Approach Method (NAMs)

For example:

- Metabolism studies may confirm which metabolites are formed.
- Targeted in vitro studies may help to fill a data gap for a specific endpoint, and / or to confirm any toxicity predictions. For example:
  - → Tox Tracker from Toxys is a stem cell reporter assay which gives mechanistic insights into genotoxic properties of chemicals.
  - → ToxProfiler from Toxys uses human liver cells to quantitatively measure cell stress responses.
- High throughput assays (e.g. transcriptomics) may help to identify toxicity which may not have been identified by the available experimental data or in Silico tools.

AFSA COLLABORATION

Uncertainty

It is key to describe the type and degree of uncertainty in a read-across. Any areas of uncertainty must be recorded in the read-across documentation.

Sources of uncertainty can include:

- Context and relevance to risk assessment / regulation.
- Data for the endpoint under consideration. For example, the quality of the study data for the source substance(s).
- Argumentation of the read-across:
  - → Hypothesis.
  - → Plausibility of the mechanism.
  - → Weight of Evidence.
- Similarity between the target and source substances:
  - → Structure.
  - → PhysChem.
  - → Toxicodynamics.
  - → Toxicokinetics.

AFSA COLLABORATION

Documenting the read-across

It is important to document all stages of the read-across. The read-across must be transparent and it must be possible to understand which substances and data are used to support the hypothesis.

Several templates have been developed to support documenting a read-across. These include:

- A strategy for structuring and reporting a read-across prediction of toxicity, Schultz et al 2015. OECD doc
- ECHA Read Across Assessment Framework (RAAF).

However, in reality, it is more often necessary to adapt these or use an in-house designed reporting format.

Regulatory acceptance

Read-across can be used to inform a risk assessment or used to support a regulatory submission.

Read-across is one of the most applied alternative approaches (adaptations) for data filling in registrations submitted under the REACH Regulations. By using read-across, unnecessary animal testing may be avoided.

The conditions under which 'Read-across and grouping' can be used to adapt the standard testing regime for REACH are listed in Annex XI, 1.5 of the REACH Regulations:

1.5.: Grouping of substances and read-across approach
The similarities may be based on:
(1) a common functional group;
(2) the common precursors and/or the likelihood of common breakdown products via physical and biological processes, which result in structurally similar chemicals; or
(3) a constant pattern in the changing of the potency of the properties across the category.
In all cases results should:
—be adequate for the purpose of classification and labelling and/or risk assessment,
—have adequate and reliable coverage of the key parameters addressed in the corresponding test method referred to in Article 13(3),
—cover an exposure duration comparable to or longer than the corresponding test method referred to in Article 13(3) if exposure duration is a relevant parameter, and
—adequate and reliable documentation of the applied method shall be provided.

AFSA COLLABORATION

Summary: Read-across

- The following are steps of a read-across process:
  - → Hypothesis
  - → Decision Context
  - → Data gap analysis for the target
  - → Target profiling
  - → Source identification
  - → Source evaluation,
  - → Source data evaluation,
  - → Read-across framework/outcome
- There are different read-across approaches:
  - → Analogue
  - → Group (or category)
- When evaluating the read-across outcome, the following should be considered:
  - → Is there enough similarity between the target and source substance(s)?
  - → Is there enough evidence to support the hypothesis?
  - → Is the data of good quality?
- Sources of uncertainty should also be considered
- The read-across process must be appropriately documented.

AFSA COLLABORATION

Using both read-across and *in silico* tools in a weight of evidence approach

How do you apply a weight?

- Take into account the robustness and reliability of the different data sources
- Depends on factors such as:
  - → the quality of the data
  - → consistency of results
  - → nature and severity of effects (for in vivo/in vitro studies)
  - → relevance of the information

- The weight of evidence approach requires use of scientific judgment and as a general principle, the more information you provide, the stronger your weight of evidence is.

For more information:
See module 7 for more on WoE and integration of results in risk assessment

Thank You !

AFSA COLLABORATION

Cosmetics Workstream Partners

HUMANE SOCIETY INTERNATIONAL

THE HUMANE SOCIETY OF THE UNITED STATES

Unilever

L'ORÉAL

P&G

Firmenich

AVON

iff

Givaudan

symrise

LUSH

Delphic HSE
SAFETY & REGULATORY SOLUTIONS

Lhasa Limited

IIVS
Institute for In Vitro Sciences