OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

Case Studies

The OpenRiskNet Consortium

OpenRiskNet: Open e-Infrastructure to Support Data Sharing, Knowledge Integration and *in silico* Analysis and Modelling in Risk Assessment
Project Number 731075

Case Studies: What is a case study ?

A particular instance of something used or analysed in order to illustrate a thesis or principle.

Is meant as:

- an in depth study of a particular situation
- method used to narrow down a very broad field of research into a researchable topic
- a guide to allow further elaboration and hypothesis creation on a subject
- an exercise to facilitate different disciplines to combine forces

Is not meant as

- a sweeping statistical survey
- a complete answer to a particular question completely

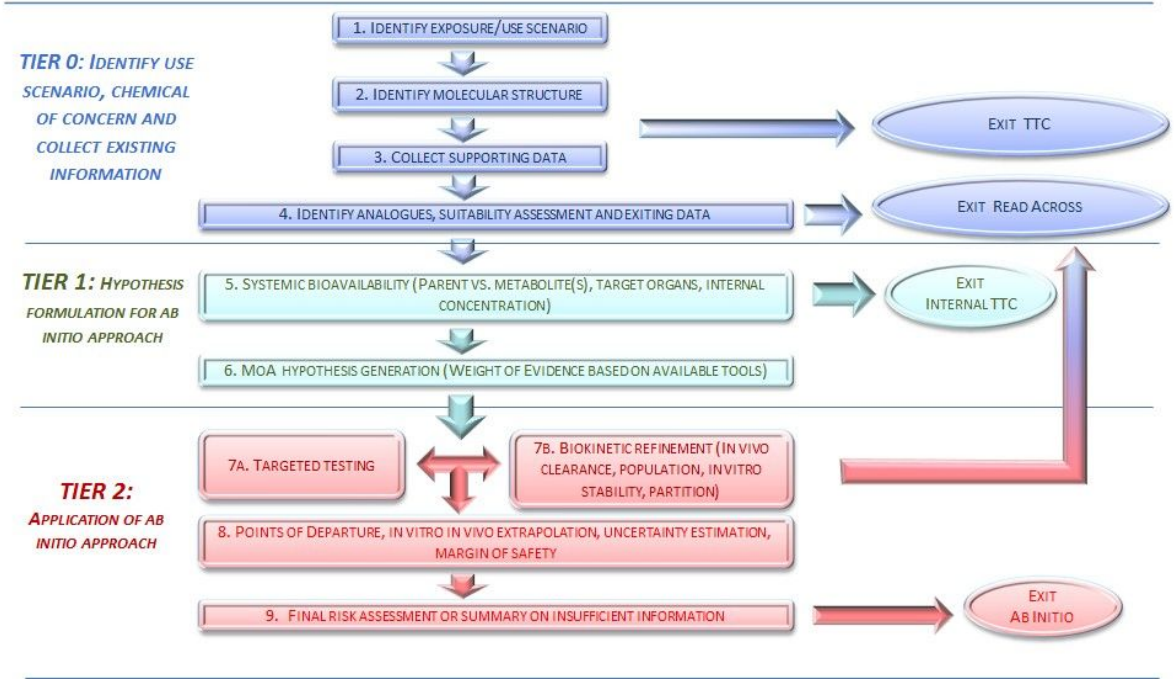
SEURAT-1

Towards the Replacement of in vivo Repeated Dose Systemic Toxicity Testing

Comput Toxicol. 2017 Nov; 4: 31–44.
Published online 2017 Nov. doi: 10.1016/j.comtox.2017.10.001
PMCID: PMC5695905
PMID: 29214231

Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods

Elisabet Berggren,[a,*] Andrew White,[b] Gladys Ouedraogo,[c] Alicia Paini,[a] Andrea-Nicole Richarz,[a] Frederic Y. Bois,[d] Thomas Exner,[e] Sofia Leite,[f] Leo A. van Grunsven,[f] Andrew Worth,[a] and Catherine Mahony[g]

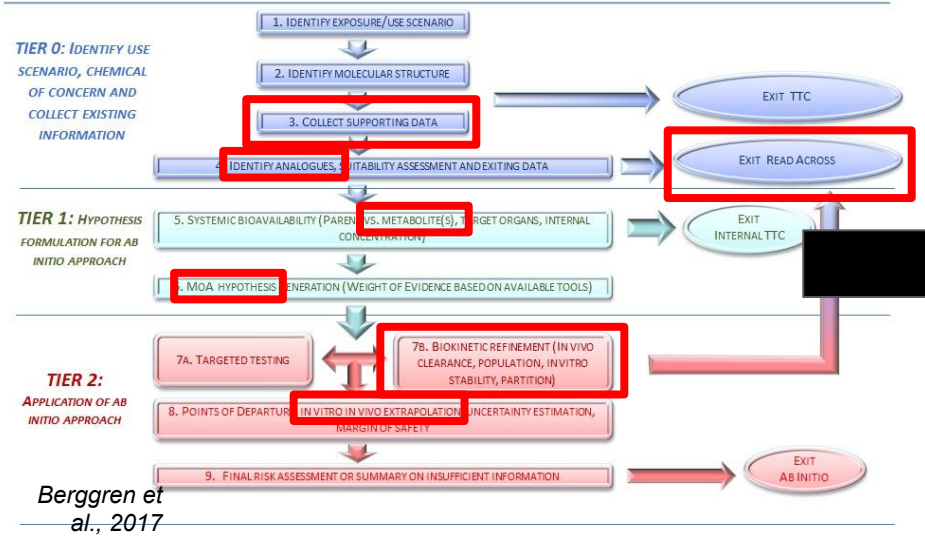


CONCLUSION: This general "ab initio" workflow was developed as a means of structuring knowledge and data in a **logical sequence** for an integrated safety assessment applying non animal methods. **Workflow could be the basis for a full risk assessment and is aiming to provide a tool to guide the evaluation through the different steps to be considered and enable and gain confidence in decision making.**

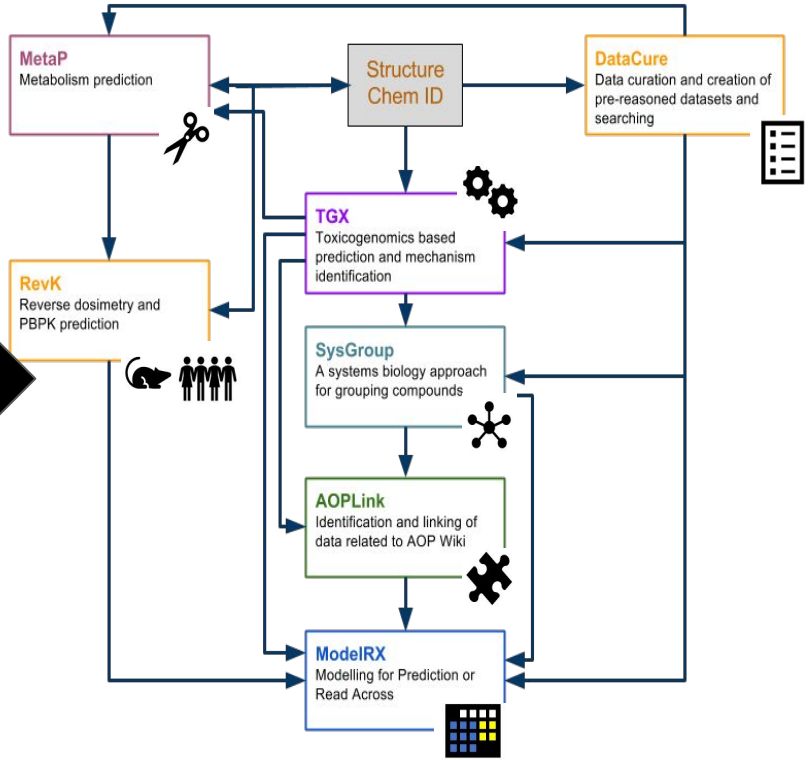
The workflow is general enough to cover different types of chemicals, endpoints and exposure scenarios.

www.openrisknet.org

Case studies based on risk assessment framework



Berggren et al., 2017



Data curation and creation of pre-reasoned datasets and searching [DataCure]

CS leader: Noffisat Oki (DC), **Involved:** DC, IM, NTUA, Fraunhofer, UoB

AIM: Users access different ORN data sources via curation services, which gets re-submitted to the data source.

e.g. nomenclature, structure (SMILES), cheminformatics; constants (logP, Henrys), kinetic and dynamic data etc..

Objectives

- Entry point of curation of all data sources
- Semantic annotation and API definition for the selected databases will also be carried out in this use case.

In close contact with the **Ontology task force** which is reviewing and adapting ontology design concepts.

Risk Assessment Framework
Tier 0.2-0.4 (data collection), 1.5 (biokinetics)

Databases

- Cheminformatic sources (PubChem)
- diXa (UM)
- ToxCast/Tox21 (DC)
- FDA EADB (DC, NTUA)

Tools / APIs

- ID converter service (DC)
- Physchem, toxicological and omics databases: RDKit, CDK (NTUA), RDKit (IM), Data Explorer (DC)
- Ontology/terminology/annotation: SCAIView / JProMiner / BELIEF (Fraunhofer), openBEL (Fraunhofer)

Metabolism Prediction [MetaP]

CS leader: Daan Geerke (VU), **Involved:** UU, JGU

Risk Assessment Framework
Tier 0.1 (mol. structure), 1.5 (biokinetics), 1.6 (MoA)

AIM: Integration of tools for site-of-metabolism (SOM) prediction and metabolite prediction

Ligand-based metabolite predictors (e.g. MetPred) and incorporate protein-structure and -dynamics based approaches to predict the site of metabolism (SOM) by **Cytochrome P450**, which metabolize ~75% of the currently marketed drugs.

Objectives: Integration, comparison and combination of tools for metabolism prediction

- Ligand-based Site-Of-Metabolism (SOM) prediction using reaction SMARTS, circular fingerprints and/or atomic reactivities
- QSBR (quantitative-structure biotransformation relationship) modelling of microbial biotransformation
- Protein-structure and -dynamics based prediction of CYP450 isoform specific binding and SOMs
- Predicting probabilities for specific reaction type events

Databases During method development, model calibration and validation we will use data from XMetDB and other open-access databases for drugs, xenobiotics and their respective metabolites.

XMetDB, SMARTcyp, ZINC, ChEMBL

Tools / APIs

- Metaprint 2D (UU)
- MetPred (UU) / MetPred2 (UU)
- UM-PPS (JGU)
- enviPath (JGU)
- SMARTCyp (external service, integration by VU)
- Plasticity tools (VU) - not yet started

Service integration

To facilitate combining metabolite prediction approaches and using MetaP outcomes as input for other predictors, we will take advantage of ongoing development in workflow management systems (Nextflow, Squonk, MDStudio) and we will explore integration into/with and use of these platforms.

https://openrisknet.org/development/case-studies/case-study-metap/

Toxicogenomics based prediction and mechanism identification [TGX]

CS leader: Danyel Jennen (UM), **Involved:** UM, VU, CRG

Risk Assessment Framework
Tier 0.3-0.4 (data collection), 1.6 (MOA)

AIM: To provide a transcriptomics-based hazard prediction model for identification of specific molecular initiating events (MIE)

The foreseen transcriptomics-based hazard prediction model will be applied based on:

(A) top-down Creation of prediction models based on differentially regulated genes

(B) bottom-up Using knowledge of stress response pathways to integrate data sets for their activation or inhibition (bottom-up approach).

The MIEs can include, but are not limited to:

(1) Genotoxicity (p53 activation), (2) Oxidative stress (Nrf2 activation), (3) Endoplasmic Reticulum Stress (unfolded protein response), (4) Dioxin-like activity (AhR receptor activation), (5) HIF1 alpha activation and (6) Nuclear receptor activation (e.g. for endocrine disruption).

Databases

- diXa / BioStudies (UM)
- TG-GATEs
- EU-ToxRisk (nascent)
- HeCaToS (nascent)
- ArrayExpress / GEO

Tools / APIs

- top-down: Data normalisation tools, prediction tools such as Caret;
- bottom-up: ToxPi
-

Service integration

- Service integration will be needed for the omics databases; knowledge bases and data mining; processing and analysis.
-

Activities

- First top-down case study based on Magkoufopolou et al 2012 paper is ready to be converted into NextFlow pipeline

https://openrisknet.org/development/case-studies/case-study-sysgroup/

Reverse dosimetry and PBPK prediction [RevK]

Risk Assessment Framework
Tier 1.5 (biokinetics)

CS leader: Frederic Bois (INERIS) **Involved:** NTUA

AIMs: Integrate PBK models in the ORN e-infrastructure. PBK models provide valuable information for risk assessment because they introduce the time dimension, simulate the full ADMET processes and can be used for in-vitro in-vivo (IVIVE) and species-to-species extrapolations.

Objectives:

- Reverse dosimetry (finding the exposure dose of a chemical leading to a “prescribed” * blood plasma concentration);
- Forward predictions of plasma and tissue concentrations following a prescribed exposure dose or concentration. In this case, Monte Carlo simulations of inter-individual variability will also be performed.

* determined

Databases Open systems Pharmacology, httk

Tools / APIs

- PKSim through Jaqpot (NTUA)
  https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/#/biokinetics
- Httk through Jaqpot (in progress, NTUA)
  https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/#/biokinetics
- R httk client (in progress, INERIS)

Service integration

Two PK (pharmacokinetics) modelling environments have been integrated in OpenRiskNet e-infrastructure through the NTUA *Jaqpot* modelling platform: The extensive, industry strength, *PKSim* PBPK open-source software, and the US EPA developed R package *httk*. With *httk* the user can request concentration-time profiles at any dose and for any length of time following exposure to about 500 chemicals, in rats, rabbits, dogs, mice or humans. A client is being developed by INERIS to present the user with an interactive interface automatizing the requests and presentation of the results.

https://openrisknet.org/development/case-studies/case-study-revk/

Identification and Linking of Data related to AOPWiki [AOPLink]

CS leader: Marvin Martens, Egon WIllighagen, Chris Evelo (UM)
Involved: DC, UoB, CRG

AIM: Linking data sources to support AOP- and omics-based risk assessment.

An AOP comprises a number of events and the adverse outcome: a molecular initiation event (MIE) is followed by one or more key events (KEs), leading to the adverse outcome (AO). Part of the development of AOPs is the search for data that supports the occurrence and **biological plausibility** of KEs and their relationships (KERs). This data can be found in literature and increasingly in databases. This case study focuses on establishing links between AOPs and data that supports a particular AOP.

Objectives:

- FAIR version of AOPWiki and WikiPathways;
- Identifier mappings for MIEs, KEs, and biological and chemical entities (genes, proteins, metabolites);
- Establish links between MIEs and KEs to biological assays;
- Establish links between assays and biological and chemical entities;
- Establish interoperable databases.

Risk Assessment Framework
Tier 0.3, 0.4 (Support Data), 1.6 (MOA)

Databases

- AOPWiki, AOP knowledgebase (AOPKB);
- WikiPathways, Reactome: biological pathway database;
- diXa, BioStudies, ArrayExpress, etc: experimental data;
- PubMed, Wikidata: literature for key events and their molecular processes.

Tools / APIs

- BridgeDb: identifier mapping;
- Effectopedia, PathVisio: pathway analysis;
- ContentMine: text mining.

Results

- BridgeDb service operational in VRE
- Report on AOPWiki<>WikiPathways linking options

Activities

- Continued development identifier mapping databases
- Semantification of AOPWiki
- AOP Portal (http://aop.wikipathways.org)
- Exploration of APIs around semantic web technologies

https://openrisknet.org/development/case-studies/case-study-aoplink/

A systems biology approach for grouping compounds [SysGroup]

CS leader: Danyel Jennen (UM), **Involved:** UM, Fraunhofer, CRG

Risk Assessment Framework
Tier 0.2-0.4 (data collection, read across)

AIM: To provide the services for **improved grouping** of compounds by integrating chemoinformatics and omics data.

Will use the approach of the diXa / DECO2 (Cefic-LRI AIMT4) projects to reproduce and extend the results obtained on the identification of hepatotoxicant groups based on similarity in mechanisms of action (omics-based) and chemical structure using services from OpenRiskNet.

Objectives

- This case study will implement an integrated analysis approach using chemoinformatics and omics data for improved grouping of compounds with similar toxicity and/or mode of action

Databases

- diXa /BioStudies (UM)
- PubChem
- ChEMBL

Tools / APIs

- PubChem
- ChEMBL or PIDGIN
- (pre)processing tools for gene expression data (e.g. microarray data) → e.g. arrayanalysis.org (UM)
- iClusterPlus

Required steps

- Chemical similarity calculated by 2D or 3D Tanimoto coefficient (PubChem)
- Protein target prediction (ChEMBL/PIDGIN)
- Interface to diXa for obtaining gene expression data
- Integration of the multiple data sources and grouping by iClusterPlus

Service integration

- Integration with other case studies is needed. SysGroup acquires information from the DataCure case study and can feed into AOPLink and ModelRX.

Modelling for Prediction or Read Across [ModelRX]

CS leader: Harry Sarimveis (NTUA) **Involved:** JGU, UU

AIMs: Given a dataset with a defined end-point to be predicted, develop a **predictive model**, validate it using OECD principles and integrate it in risk assessment workflows for **filling gaps** in available experimental information.

A training data set will be obtained from an ORN data source. The model has then to be trained with ORN modelling tools and the resulting model has to be packaged into a container, documented and ontologically annotated. The model will be validated using OECD guidelines. Finally, a prediction can be run.

Objectives:

- Support similarity identification in the DataCure case study (by providing tools for calculating theoretical descriptors of substances);
- Fill gaps in incomplete datasets and use in silico predictive modelling approaches (read across, QSAR) to support final risk assessment.

Risk Assessment Framework
Tier 0.4 (read-across), 1.5, 1.6

Databases ChEMBL-eNanoMapper-Gene Ontology (GO) - Any ORN data providing service

Tools / APIs

- Jaqpot (NTUA)/https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/
- Lazar (JGU)/ https://lazar.prod.openrisknet.org/
- WEKA Rest Service (JGU)/ https://jguweka.prod.openrisknet.org/

Service integration: **Jaqpot** is a language agnostic, extensible modelling platform, able to integrate, in the form of microservices, practically any open-source algorithm or tool that can be used for developing predicting toxicology models. The platform has integrated many popular statistical and machine learning algorithms implemented in R, Python and WEKA. Integrated modelling and analysis services also include, scaling/normalisation of data, variable selection, model validation, algorithms for defining the domain of applicability, and tools for the automatic generation of model validation reports and model prediction reports, such as the QSAR prediction reporting formats (QPRF). The **Lazar** services are available and more general modelling services. Lazar (lazy structure–activity relationships) is a modular framework for predictive toxicology. Similar to the *read across* procedure in toxicological risk assessment, lazar creates local QSAR (quantitative structure–activity relationship) models for each compound to be predicted. The TU of Munich modelling services based on **WEKA** were brought into the new technological framework. This web-service provides an OpenRiskNet compliant REST interface to the machine learning algorithms from the WEKA Java Library.

Acknowledgements

OpenRiskNet (Grant Agreement 731075) is a project funded by the European Commission within Horizon2020 Programme

Project partners:

P1 Douglas Connect GmbH, Switzerland (DC)
P2 Johannes Gutenberg-Universitat Mainz, Germany (JGU)
P3 Fundacio Centre De Regulacio Genomica, Spain (CRG)
P4 Universiteit Maastricht, Netherlands (UM)
P5 The University Of Birmingham, United Kingdom (UoB)
P6 National Technical University Of Athens, Greece (NTUA)
P7 Fraunhofer Gesellschaft Zur Foerderung Der Angewandten Forschung E.V., Germany (Fraunhofer)
P8 Uppsala Universitet, Sweden (UU)
P9 Medizinische Universitat Innsbruck, Austria (MUI)
P10 Informatics Matters Limited, United Kingdom (IM)
P11 Institut National De L'environnement Et Des Risques INERIS, France (INERIS)
P12 Vrije Universiteit Amsterdam, Netherlands (VU)