An Integrated Data-Driven Framework for Predictive Safety and Sustainability Assessment

Harry Sarimveis

National Technical University of Athens

ANTHOS'26 Solutions Session 4:
Projects solutions to the needs of (scientific) Implementors

11 March 2026

THE PINK PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON EUROPE RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 101137809.

ASSOCIATED PARTNERS (I.E. (A) SWISS PARTNERS AND (B) UK PARTNERS) HAVE RECEIVED NATIONAL FUNDING FROM (A) THE SWISS STATE SECRETARIAT FOR EDUCATION, RESEARCH AND INNOVATION (SERI), AND (B) INNOVATE UK.

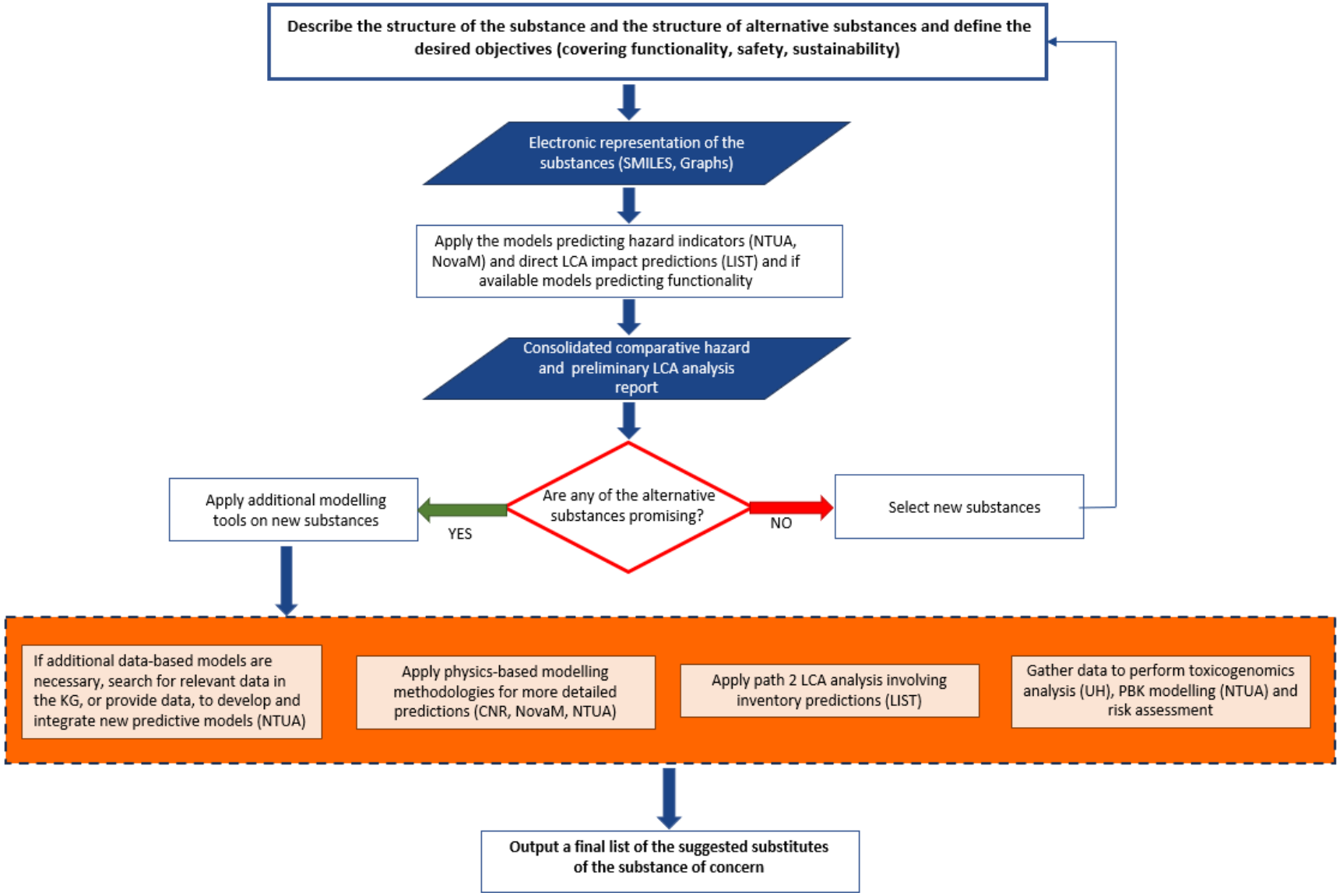
PINK's high-level objectives

- Develop innovative modelling and simulation approaches addressing industrial SSbD needs
- Make the software accessible to SMEs and industry through an open innovation platform
- Validate the platform on PINK Developmental Case Studies and Industrial Demonstrators provided by SMEs and large industries
- Fully implement open science and FAIR principles contributing to establishing a European chemicals and materials data, modelling and software ecosystem
- Strengthen knowledge transfer through collaboration and exploiting synergies
- Boost the innovative capacity of SMEs and industry and make them more agile to respond to external and internal influences

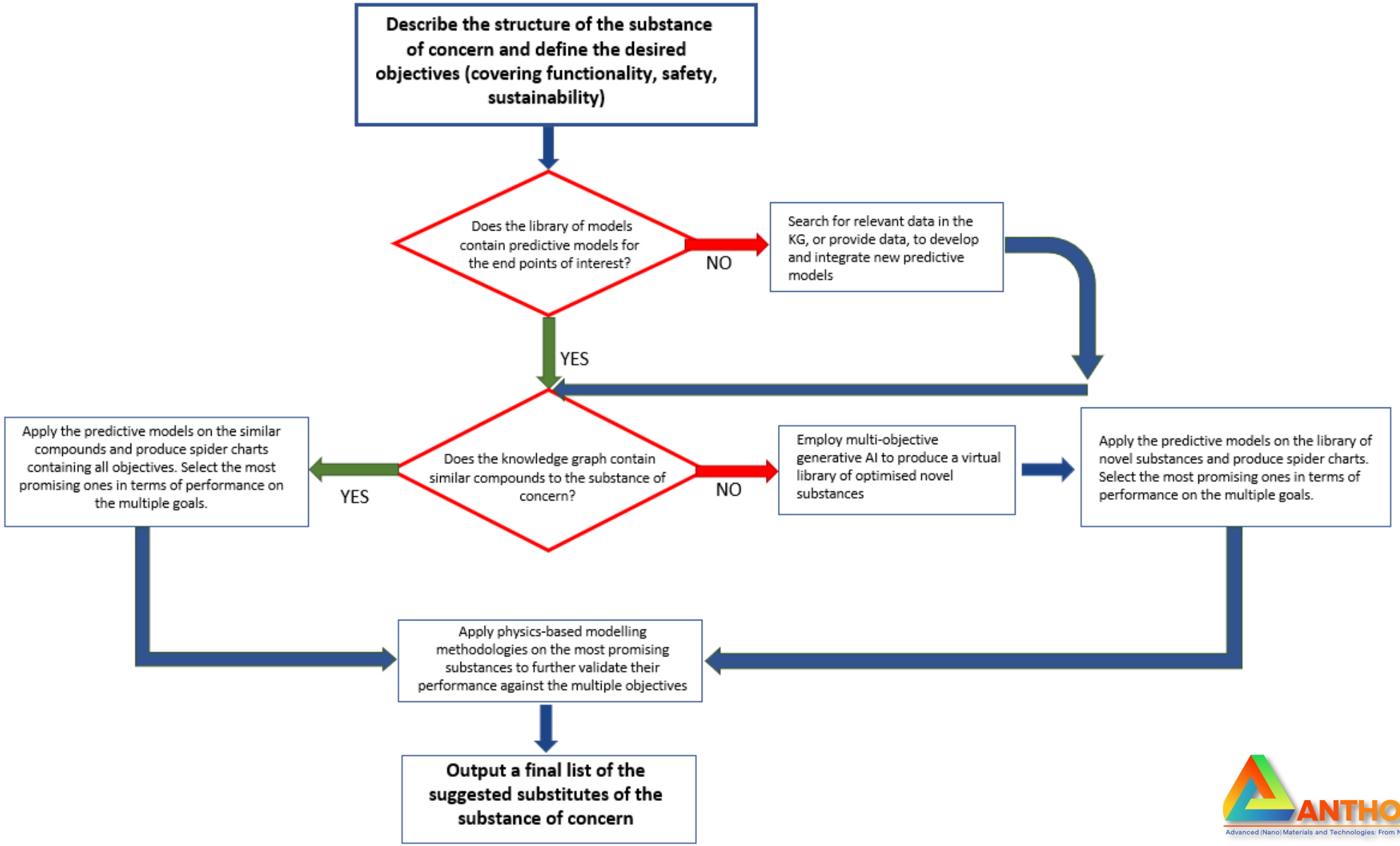
Modelling objectives

The goal is to develop a **comprehensive toolset that will generate the input for the PINK Decision-Support System (DSS)**, merging both data-driven and physics-based modelling approaches with specialised cheminformatics, bioinformatics, and Life-Cycle Assessment (LCA) tools and biokinetics simulations and integrated with data search and AI technologies.

Development of Decision Support Systems

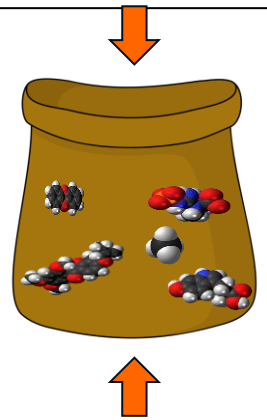


Development of Decision Support Systems



Stage-gated SSbD assessment

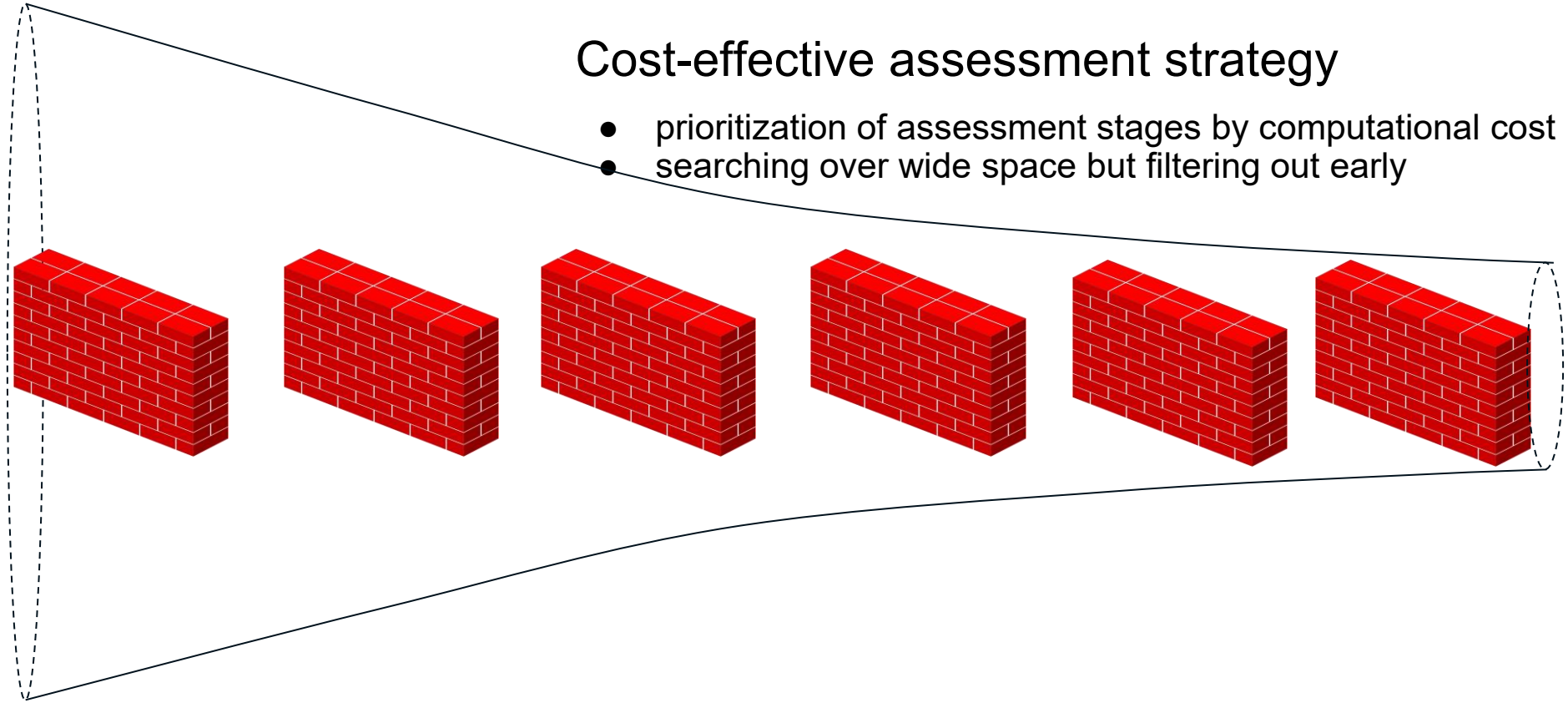
Similarity search, Literature search, Mol. generation (Generative AI)



Expert-guided compound recommendations

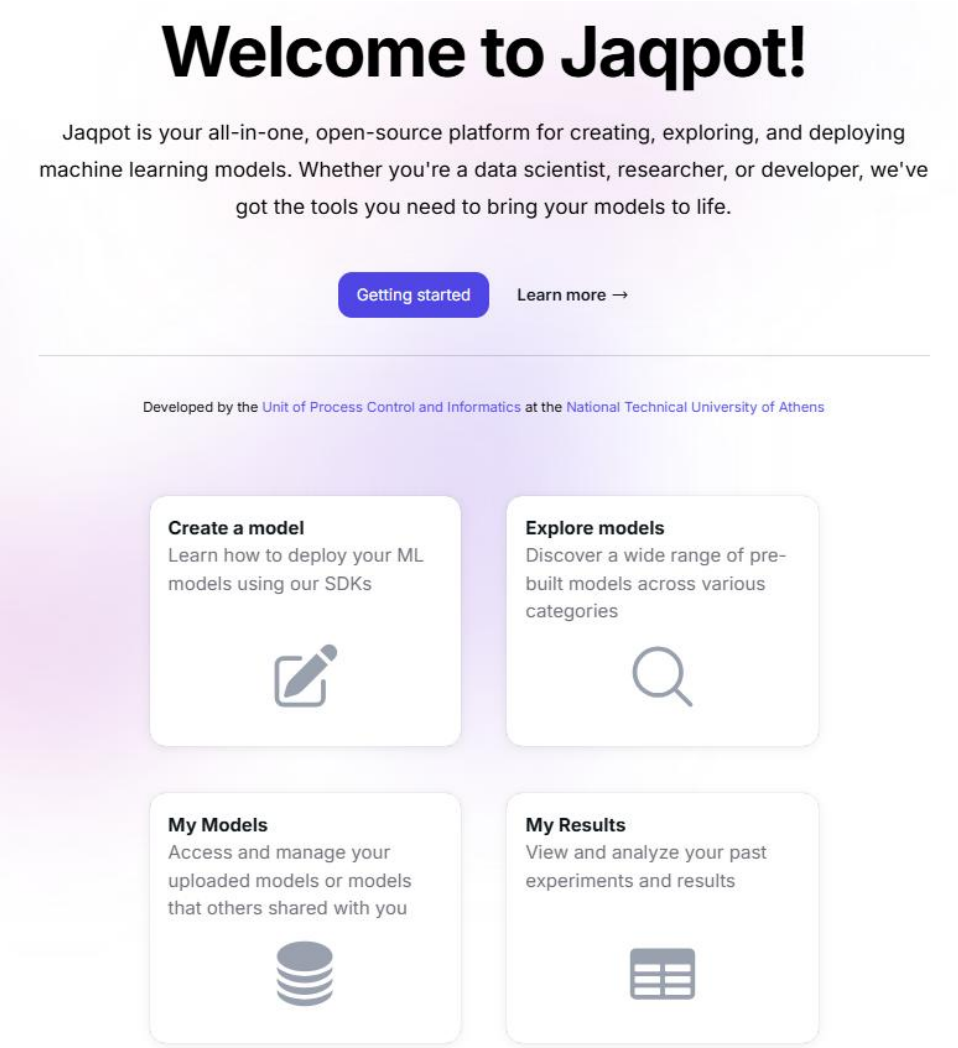
Cost-effective assessment strategy

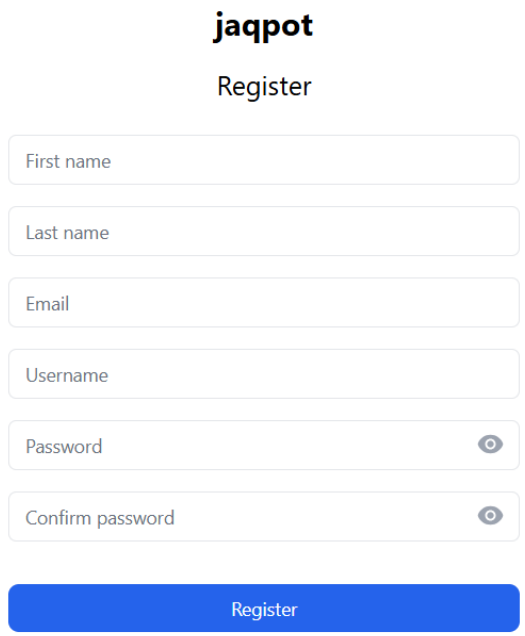
- prioritization of assessment stages by computational cost
- searching over wide space but filtering out early



Shortlist for expert evaluation

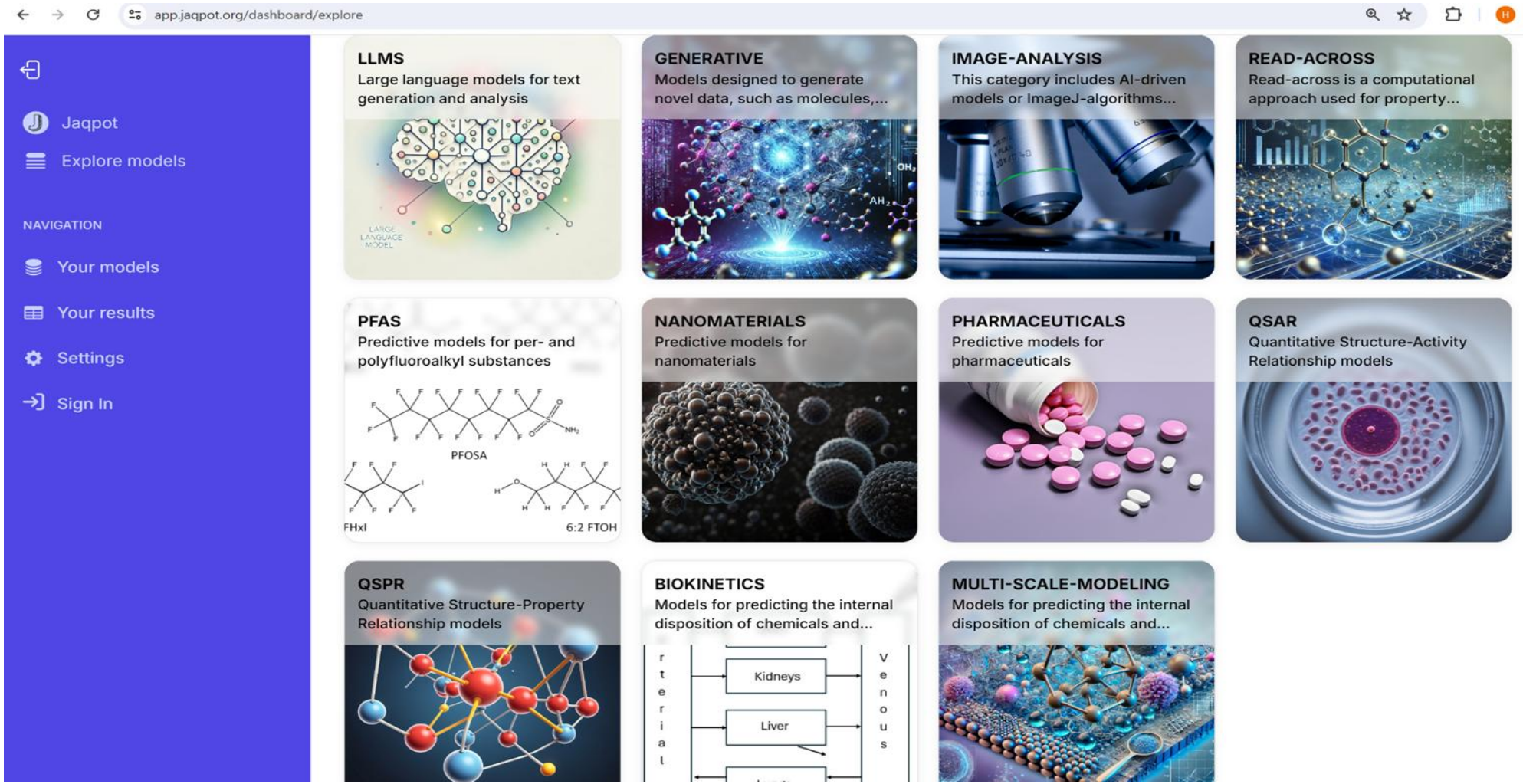
Jaqpot cloud platform





https://app.jaqpot.org/dashboard

Jaqpot cloud platform



https://app.jaqpot.org/dashboard/explore

PINK Organisation on the Jaqpot cloud platform

PINK

Harry Sarimveis
@hsarimv@mail.ntua.gr

Description | People | Shared | Edit

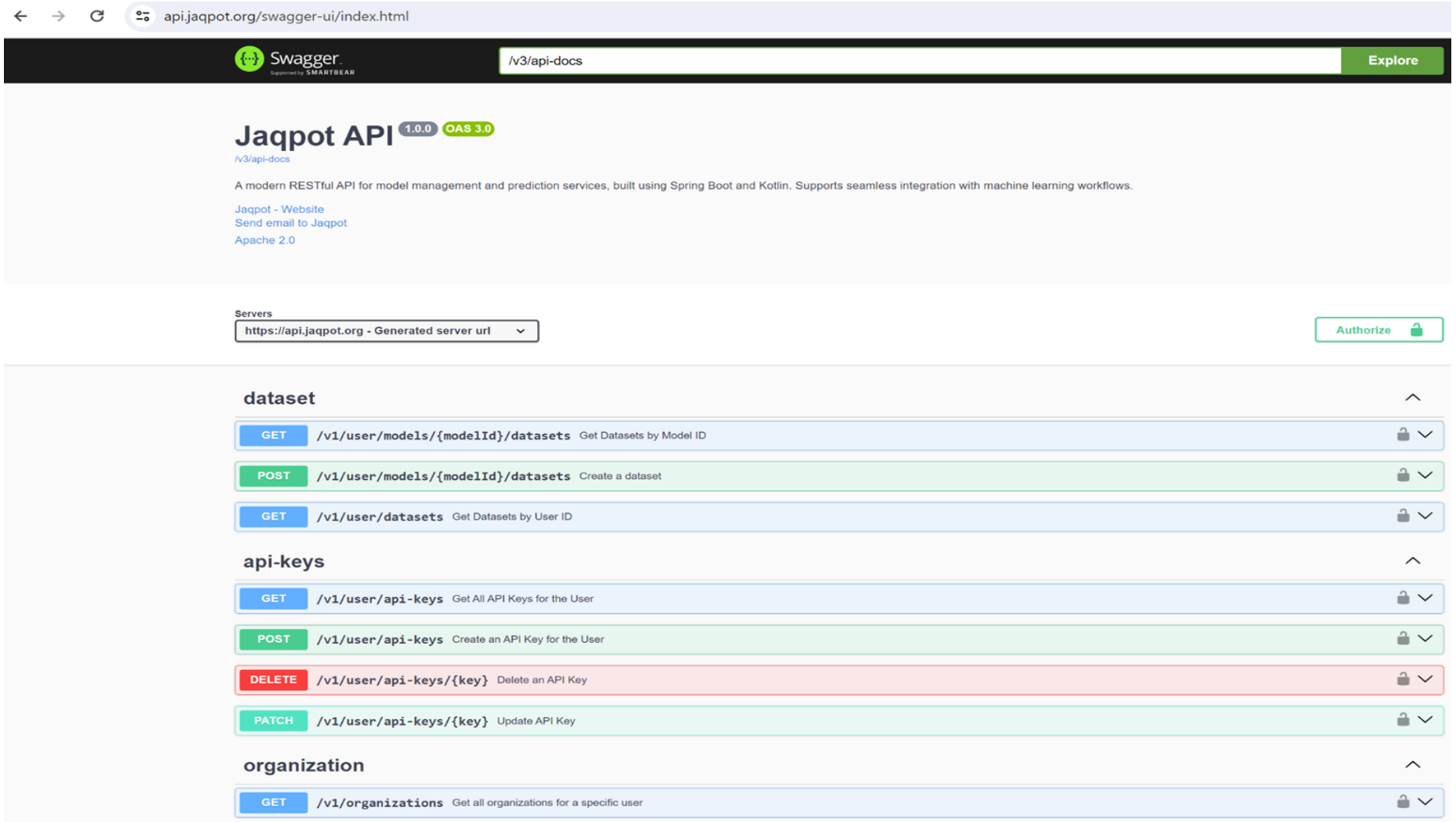
The PINK EU project aims at providing integrated computational approaches for the introduction of safe-and-sustainable-by-design chemicals and materials into the market. PINK will develop a Tiered Approach, where each development cycle will start with mining and creating data on the functionality, safety, and sustainability, as well as cost/economic feasibility, which will be provided simultaneously as input for the decision support workflow. The PINK Tiered Approach will categorise all computational methods into three tiers that parallel the material development TRLs, with respect to their data needs, throughput, and confidence levels. At earlier development stages (ideation/design phase of material development), where less data is available, high-throughput computational approaches of low tiers will be applied to avoid limiting the exploration and missing promising design routes due to time constraints (human and computer) and costs.

For more information visit https://pink-project.eu/

Title	Model Description
QSAR toolbox Calculator integration	This service provides access to the calculators included in the QSAR toolbox
QSAR Toolbox Model Integration	This service provides access to the QSAR models included in the QSAR toolbox
QSAR Toolbox Profiler integration	This service provides access to the profilers included in the QSAR toolbox
DILI Prediction Model	This model predicts if a substance induces liver injury
IndusChemFate PBTK	IndusChemFate is a PBTK model designed to estimate blood and urine concentrations of multiple chemicals and their metabolites given specific exposure scenarios.
IndusChemFate Lite PBTK	A lite version of IndusChemFate PBTK model, limiting metabolism to liver-only and single metabolites, reducing complexity and parameter requirements compared to the original model

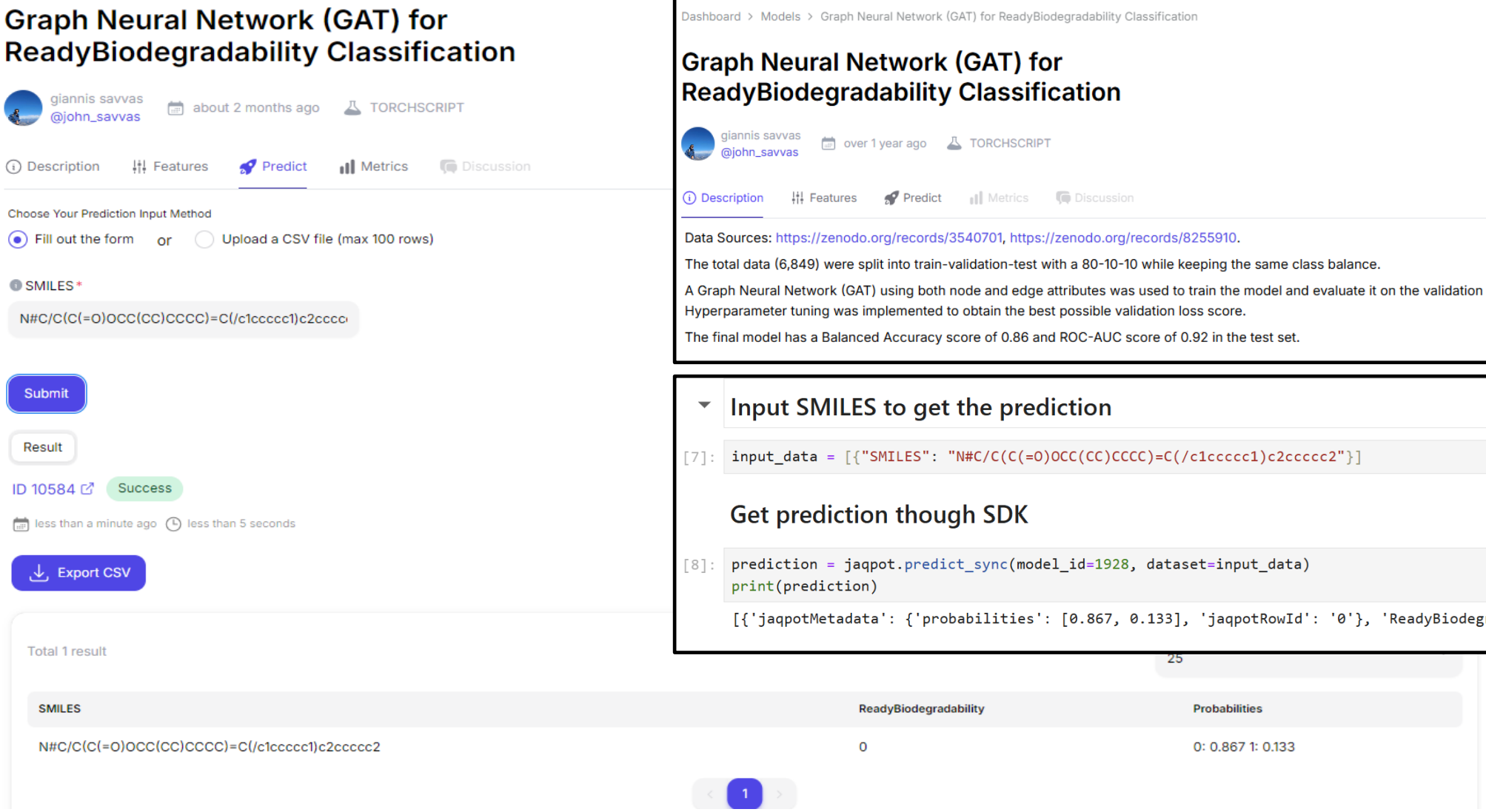
https://app.jaqpot.org/dashboard/organizations/PINK/description

Application Programming Interfaces



https://api.jaqpot.org/swagger-ui/index.html

Example: Ready Biodegradability Classification Model



Dashboard > Models > Graph Neural Network (GAT) for ReadyBiodegradability Classification

Graph Neural Network (GAT) for ReadyBiodegradability Classification

giannis savvas
@john_savvas
over 1 year ago
TORCHSCRIPT

Description | Features | Predict | Metrics | Discussion

Data Sources: https://zenodo.org/records/3540701, https://zenodo.org/records/8255910.

The total data (6,849) were split into train-validation-test with a 80-10-10 while keeping the same class balance.

A Graph Neural Network (GAT) using both node and edge attributes was used to train the model and evaluate it on the validation set. Hyperparameter tuning was implemented to obtain the best possible validation loss score.

The final model has a Balanced Accuracy score of 0.86 and ROC-AUC score of 0.92 in the test set.

Input SMILES to get the prediction

```
[7]: input_data = [{"SMILES": "N#C/C(C(=O)OCC(CC)CCCC)=C(/c1ccccc1)c2ccccc2"}]
```

Get prediction though SDK

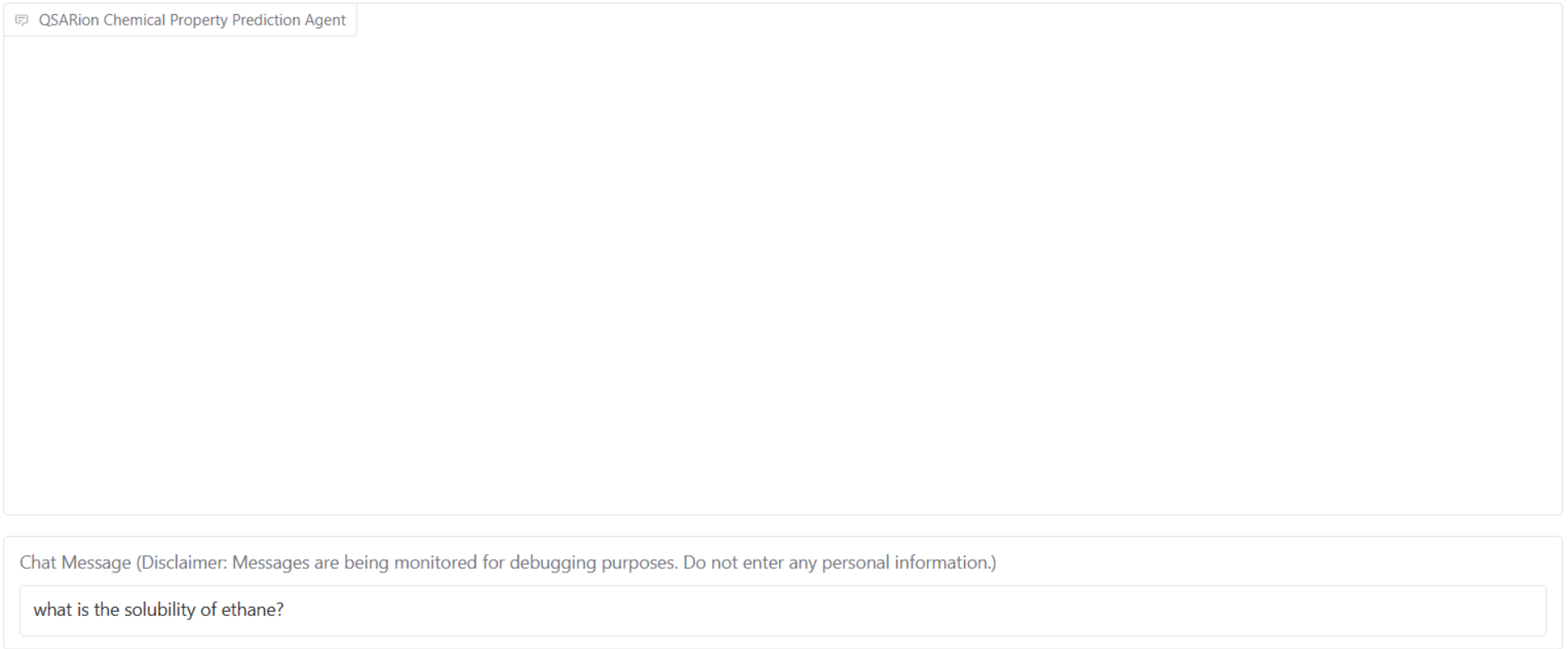
```
[8]: prediction = jaqpot.predict_sync(model_id=1928, dataset=input_data)
     print(prediction)

[{'jaqpotMetadata': {'probabilities': [0.867, 0.133], 'jaqpotRowId': '0'}, 'ReadyBiodegradability': 0.0}]
```

https://app.jaqpot.org/dashboard/models/1928/description

AI agents for SSBD assessment

Verification Required: Always verify predictions through experimental validation or consultation with domain experts before practical application.

Supported by: This work is supported by the PINK Project - a Horizon Europe initiative for Safe-and-Sustainable-by-Design computational approaches for chemicals and materials.



Step 1 – user input

Step 1

Thought: To find the solubility of Ethane, I will first retrieve the SMILES string for Ethane using the `retrieve_smiles_tool`. Then, I will use the `qsartoolbox_get_property_value_tool` with the appropriate model to predict the solubility.

Code:

```
# Retrieve the SMILES string for Ethane
ethane_smiles = retrieve_smiles_tool(compound_name="Ethane")
print("SMILES for Ethane:", ethane_smiles)
```

Used tool python_interpreter

```
# Retrieve the SMILES string for Ethane
ethane_smiles = retrieve_smiles_tool(compound_name="Ethane")
print("SMILES for Ethane:", ethane_smiles)
```

Execution Logs

SMILES for Ethane: CC
Last output from code snippet:
None

Step 1 | Input-tokens:8,272 | Output-tokens:107 | Duration: 13.61

Step 2 – trigger appropriate models

Step 2

Thought: I have the SMILES string for Ethane, which is "CC". Now, I will use the `qsartoolbox_get_property_value_tool` to predict the solubility of Ethane using the "Water Solubility (fragments) (EPISUITE)" model.

Code:

```
# Predict the solubility of Ethane using the Water Solubility (fragments) (EPISUITE) model
solubility_result = qsartoolbox_get_property_value_tool(smiles="CC", qsartoolbox_qsar_model="Water Solubility (fragments) (EPISUITE)")
print("Solubility result:", solubility_result)
```

Calling tools:

[{'id': 'call_2', 'type': 'function', 'function': {'name': 'python_interpreter', 'arguments': '# Predict the solubility of Ethane using the Water Solubility (fragments) (EPISUITE) model\nsolubility_result = qsartoolbox_get_property_value_tool(smiles="CC", qsartoolbox_qsar_model="Water Solubility (fragments) (EPISUITE)")\nprint("Solubility result:", solubility_result)'}}]

Used tool python_interpreter

```
# Predict the solubility of Ethane using the Water Solubility (fragments) (EPISUITE) model
solubility_result = qsartoolbox_get_property_value_tool(smiles="CC", qsartoolbox_qsar_model="Water Solubility (fragments) (EPISUITE)")
print("Solubility result:", solubility_result)
```

Execution Logs

Solubility result: {'DomainResult': 'Undefined', 'DataType': 'Qsar prediction.', 'RigidPath': 'Physical Chemical Properties#Water solubility', 'Endpoint': 'Water Solubility', 'MetaData': ['Prediction approach=External SAR/QSAR', 'Domain=Undefined (Water Solubility (fragments) (EPISUITE))', 'Endpoint=Water Solubility', 'SAR/(Q)SAR=Water Solubility (fragments) (EPISUITE) (v1.0)', 'Domain status=Undefined', 'Primary group=N/A', 'Primary group with metabolism=N/A', '(Q)SAR options='], 'Value': '383.29', 'Qualifier': '', 'MinQualifier': '', 'MaxQualifier': '', 'Unit': 'mg/L', 'Family': 'Mass concentration', 'Name': 'Ethane', 'ChemId': '3c6c8350-508d-1765-d8b2-f0c9e8f71c12', 'jaqpotMetadata': {'jaqpotRowId': '0'}}

Last output from code snippet:

None

Step 2 | Input-tokens:8,489 | Output-tokens:263 | Duration: 23.99

Step 3 - results

Step 3

Thought: The solubility of Ethane has been predicted to be 383. 29 mg/L using the "Water Solubility (fragments) (EPISUITE)" model from the QSAR Toolbox.

Code:

```
final_answer(answer="383. 29 mg/L")
```

Used tool python_interpreter

Step 3 | Input-tokens:9,140 | Output-tokens:72 | Duration: 12.52

Final answer:

383. 29 mg/L

QSAR Toolbox and EPA CCTE

Experimental and calculated data points are provided from the QSAR toolbox API and the EPA's CCTE API.

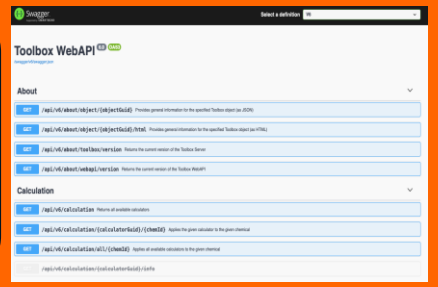
They follow a different rationale and their output is formatted to a different schema.

A common schema was structured to join information while retaining information from both sources.

QSAR TOOLBOX

OECD QSAR Toolbox
NTUA-hosted API server

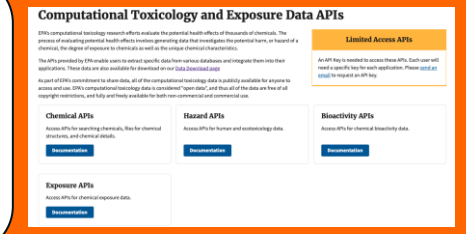
http://54.155.131.85/swagger/index.html



EPA

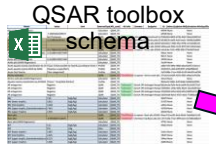
Computational Toxicology and Exposure Data APIs

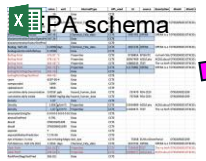
https://api-ccte.epa.gov/docs/index.html



QSAR toolbox API: QSAR predictions, calculators, profilers, data

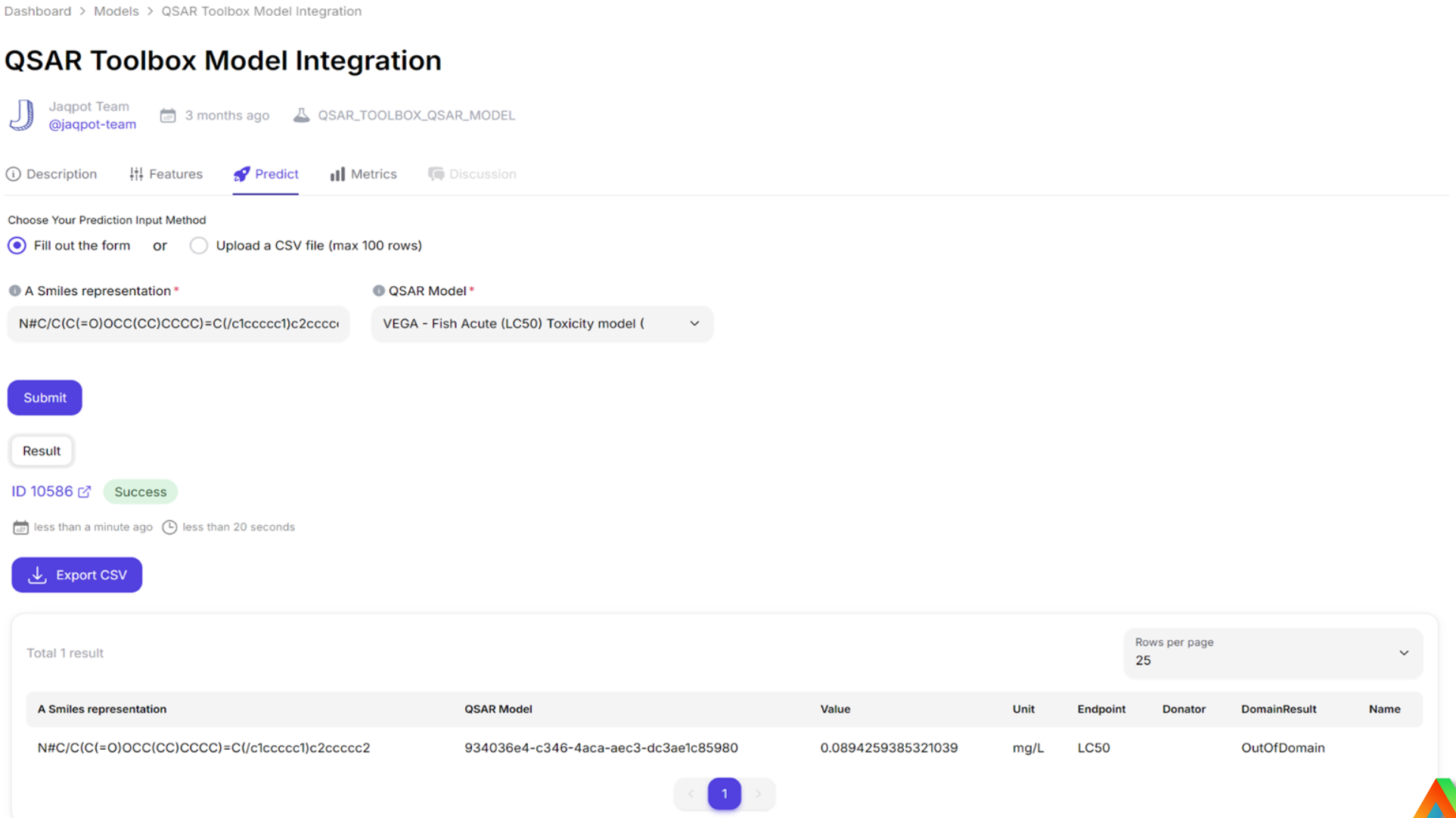
CCTE API: Predictions, Properties, data





All data and predictions collected under a common schema

Example of generating a QSAR Toolbox prediction through Jaqpot



QSAR: performance, applicability, uncertainty and explainability

1 Performance

Model-level (fit & predictivity)

Regression: R^2, RMSE
Classification: AUC, MCC
CV + external test

Output: R^2=0.72, RMSE=0.35

2 Applicability

Compound-level (in vs out)

kNN similarity/coverage
Leverage check
In/out badge + reason

Output: In-domain (sim=0.81)

3 Uncertainty

Prediction-level (how sure?)

Prediction interval (e.g., 90%)
Conformal (optional)
Used in ranking

Output: Pred=3.2 (PI: 2.6–3.9)

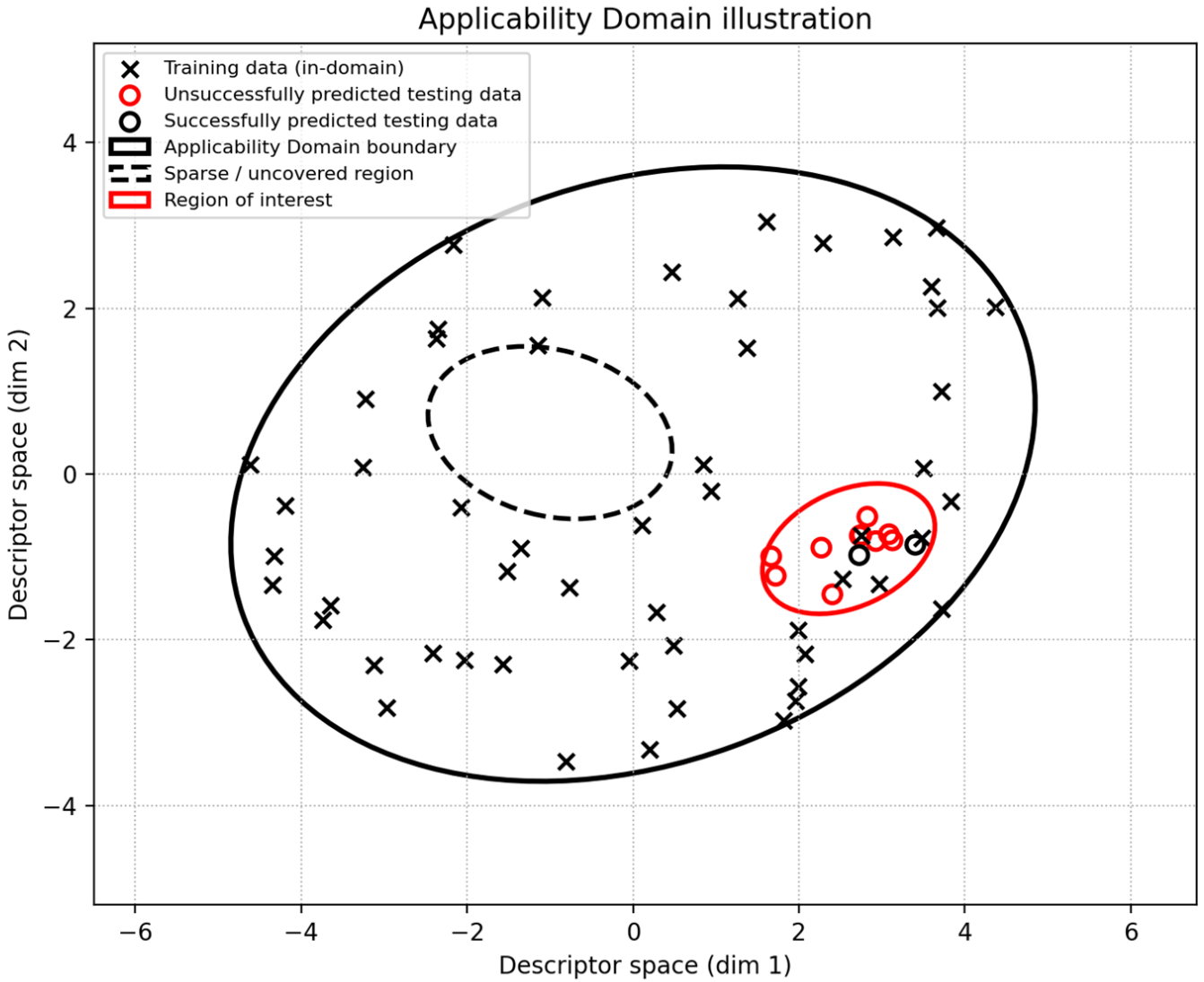
4 Explainability

Why did the model say this?

Top drivers/features
Simple ‘reason’ sentence

Output: Key drivers

Applicability domain illustration

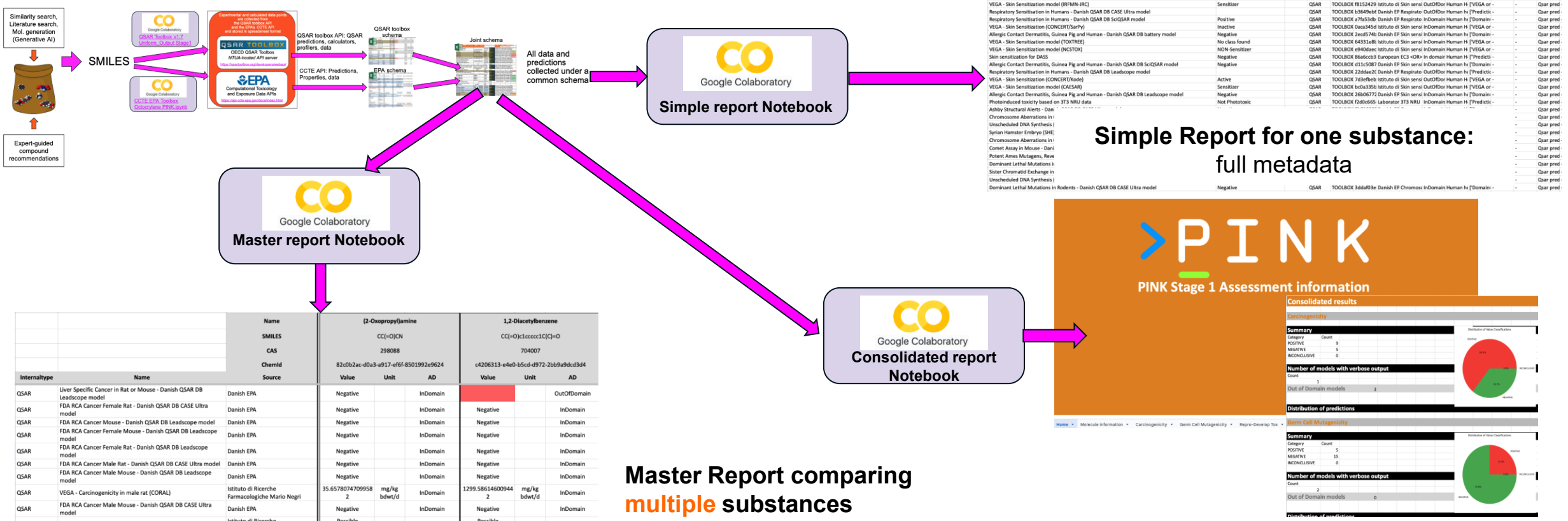


The PINK matrix

SSbD Framework - Indicator	Type of Hazard	Type of model	Expressed via	More info about readout	Inputs to the model	Reference	Modeller	Ready for Modelling Data File	Source of data	Model URL	Report (QMRF, MODA etc.)	Additional information	Id
Carcinogenicity													
Carcinogenicity (genotox and nongenotox) alerts by ISS	Human Health	QSAR - Categorical	Categories include: ['Aliphatic halogens (Genotox); Alkyl		SMILES	https://qsartoolbox.org/	ISS, Istituto Superiore di			.org/dashboard/models/1842/predic	ile/d/1A-PrNYn-XXgJyLs_c9xaiTGxbC5qHoB		2256b12d-5ef8-49a2-af9a-
FDA RCA Cancer Female Mouse - Danish QSAR DB CASE Ultra	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative', 'Equivocal']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1UfVlkwcNl9Nil3T1g9D463njkqScE5Gr/view?us		e2078536-1427-4d94-88dc-
FDA RCA Cancer Female Mouse - Danish QSAR DB Leadscope model	Human Health	QSAR - Categorical	No units, 1 for positives and 0 for negatives.		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1pxYfEJEMego3NmXu-TdkUF0V-		12b85d79-c468-4da6-8e2e-
FDA RCA Cancer Female Rat - Danish QSAR DB CASE Ultra	Human Health	QSAR - Categorical	Categories include: ['Equivocal', 'Positive',		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1fsxTKkRx2ITCoacSMfUou84jbOpnlZcN/view?us		eb6861f5-178d-4109-8bc3-
FDA RCA Cancer Female Rat - Danish QSAR DB Leadscope model	Human Health	QSAR - Categorical	No units, 1 for positives and 0 for negatives.		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1zVAfqI89v0Fj3r8YOnbYk-		892add19-9efa-4011-b8dc-
FDA RCA Cancer Male Mouse - Danish QSAR DB CASE Ultra	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative', 'Equivocal']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1BUJe3Sm02P9pJLXxYOVDfzGpXOtRPw5G/view		adc64575-36e0-4eb5-86df-
FDA RCA Cancer Male Mouse - Danish QSAR DB Leadscope model	Human Health	QSAR - Categorical	No units, 1 for positives and 0 for negatives.		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1mQwCqkix8Ch0xAgpnGMCU-		41b3e7be-562a-44a2-82f6-
FDA RCA Cancer Male Rat - Danish QSAR DB CASE Ultra	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/14iwdyHTiDBRAETU0hAMPKSTcmUKym0L-		f7395d96-ddfb-4b68-b8cd-
FDA RCA Cancer Mouse - Danish QSAR DB CASE Ultra model	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative', 'Equivocal']		SMILES	https://qsartoolbox.org/	Danish QSAR Group at DTU			.org/dashboard/models/1837/predic	ile/d/1ufxFyOXzqR5YxxQutfhqqLc3-		9bb94df1-4788-4f17-bbd9-
FDA RCA Cancer Mouse - Danish QSAR DB Leadscope model	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1WgdOMBe8MaL78HMypVaw6vMz7pov6wYR		c11ced21-23d1-4574-aa17-
FDA RCA Cancer Rat - Danish QSAR DB CASE Ultra model	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative', 'Equivocal']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1k8ui3cGds6z0Kk3prODpHazvk6zHcCsg/view?		3fa56226-315a-4e18-9021-
FDA RCA Cancer Rat - Danish QSAR DB Leadscope model	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1KY4PhlceQvPvKHPDhFm63DLMJlqHa_nE/view		7ef3e1a1-d10a-4278-bd90-
FDA RCA Cancer Rodent - Danish QSAR DB CASE Ultra model	Human Health	QSAR - Categorical	Carcinogenicity in rodent liver (exclusively), positive or		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ocument/d/1-2-9lRc51yfmL_-		1e7d360f-85da-4957-84cf-
FDA RCA Cancer Rodent - Danish QSAR DB Leadscope model	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1L-zbrx-k7N6hXi2TPpYGNJaxfTZ-		1d743a96-53b9-428f-bae0-
Liver Specific Cancer in Rat or Mouse - Danish QSAR DB battery	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1evay6urEoY-dqnTi3556PdJNa8SGEpSB/		c40d6a42-80bd-4917-8878-
Liver Specific Cancer in Rat or Mouse - Danish QSAR DB CASE	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1Z-4OmG9cccNIlPHvWtTchSg		0a590899-12a7-466c-a8dc-
Liver Specific Cancer in Rat or Mouse - Danish QSAR DB	Human Health	QSAR - Categorical	Categories include: ['Positive', 'Negative']		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1XLxFuC3xT0QUH8Q1rgkIx2hknRlVyPFz/view?		a329e3be-5b40-4fd0-b5ad-
Liver Specific Cancer in Rat or Mouse - Danish QSAR DB SciQSAR	Human Health	QSAR - Categorical	Positive or Negative		SMILES	https://qsartoolbox.org/	Danish EPA			.org/dashboard/models/1837/predic	ile/d/1BWfPS_5G0NAtlOJdAIK32hiT1RhVcNFH/view?		a86720fe-34f5-4bbc-a3f8-
VEGA - Carcinogenicity inhalation classification model (IRFMN)	Human Health	QSAR - Categorical	0 (non-carcinogen), 1 (carcinogen)		SMILES	https://qsartoolbox.org/	Istituto di Ricerche			.org/dashboard/models/1837/predic	ile/d/1tH-jr9KOxe9UPx_QDQ4Vqp6V	Appers in Acute Toxicity:Inhalation and Carcinogenicity	494572f1-eaca-49ba-95d4-
VEGA - Carcinogenicity model (CAESAR)	Human Health	QSAR - Categorical	0 (non-carcinogen), 1 (carcinogen)		SMILES	https://qsartoolbox.org/	Istituto di Ricerche			.org/dashboard/models/1837/predic	ile/d/1ByfOIHJbOtpddnZxDFi8ldOTY01Vbyf2/view?		edaafce9-0b6d-49a3-9801-
VEGA - Carcinogenicity model (IRFMN-Antares)	Human Health	QSAR - Categorical	0 (non-carcinogen), 1 (carcinogen)		SMILES	https://qsartoolbox.org/	Istituto di Ricerche			.org/dashboard/models/1837/predic	ocument/d/1e00ZJi8pDSj4Ur8On_skdnxCi-		0665aa46-d259-4194-9f4b-
VEGA - Carcinogenicity model (IRFMN-ISSCAN-CGX)	Human Health	QSAR - Categorical	0 (non-carcinogen), 1 (carcinogen)	TOX 7.7. Carcinogenicity (in	SMILES	https://qsartoolbox.org/	Istituto di Ricerche			.org/dashboard/models/1837/predic	ocument/d/1bdAr03gfzUFhJ4yuF34ah4-		1e68bcdf-b2b8-4d84-b45e-
VEGA - Carcinogenicity model	Human	QSAR -	0 (non-carcinogen), 1			https://qsartoolbox.org/	Istituto di			.org/dashboard/m	ile/d/1ZVQvuvP37IvW1h		e9b1d082-4358-

1.Hazard Assessment | 2.Production&processing phase | 3.Final Application Phase | 4.Environmental sustainability | 5.Social-Economic Sustainab | Functionality | +

Workflow for SSbD assessment



Name	Value	Unit	Internaltype	Api_Used	Id	Source	Description	Ad Info	Rigidpath	Metadata	Dtxsid	Dtxcid	Proptype
Allergic Contact Dermatitis, Guinea Pig and Human - Danish QSAR DB CASE Ultra model			QSAR		TOOLBOX 16251e73	Danish EP	Skin sensi	OutOfDor	Human h	["Predictic	-	-	Qsar pred
Respiratory Sensitisation in Humans - Danish QSAR DB battery model			QSAR		TOOLBOX 27ab0aa7	Danish EP	Respirato	OutOfDor	Human h	["Predictic	-	-	Qsar pred
VEGA - Skin Sensitization model (IRFMN-JRC)	Sensitizer		QSAR		TOOLBOX f8152429	Istituto di	Skin sensi	OutOfDor	Human h	["VEGA or	-	-	Qsar pred
Respiratory Sensitisation in Humans - Danish QSAR DB CASE Ultra model			QSAR		TOOLBOX b3649ebf	Danish EP	Respirato	OutOfDor	Human h	["Predictic	-	-	Qsar pred
Respiratory Sensitisation in Humans - Danish QSAR DB SciQSAR model	Positive		QSAR		TOOLBOX a7fa53db	Danish EP	Respirato	InDomain	Human h	["Domain:	-	-	Qsar pred
VEGA - Skin Sensitization (CONCERT/SarPy)	Inactive		QSAR		TOOLBOX 0aca345d	Istituto di	Skin sensi	InDomain	Human H	["VEGA or	-	-	Qsar pred
Allergic Contact Dermatitis, Guinea Pig and Human - Danish QSAR DB battery model	Negative		QSAR		TOOLBOX 2ecd574b	Danish EP	Skin sensi	InDomain	Human h	["Domain:	-	-	Qsar pred
VEGA - Skin Sensitization model (TOXTREE)	No class found		QSAR		TOOLBOX 64331e8E	Istituto di	Skin sensi	InDomain	Human H	["VEGA or	-	-	Qsar pred
VEGA - Skin Sensitization model (NCSTOX)	NON-Sensitizer		QSAR		TOOLBOX e940daec	Istituto di	Skin sensi	InDomain	Human H	["VEGA or	-	-	Qsar pred
Skin sensitization for DASS	Negative		QSAR		TOOLBOX 86a6ccb3	European	EC3 <OR>	In domair	Human H	["Predicti	-	-	Qsar pred
Allergic Contact Dermatitis, Guinea Pig and Human - Danish QSAR DB SciQSAR model	Negative		QSAR		TOOLBOX d11c5087	Danish EP	Skin sensi	InDomain	Human h	["Domain:	-	-	Qsar pred
Respiratory Sensitisation in Humans - Danish QSAR DB Leadscope model			QSAR		TOOLBOX 22ddae2C	Danish EP	Respirato	OutOfDor	Human h	["Predictic	-	-	Qsar pred
VEGA - Skin Sensitization (CONCERT/Kode)	Active		QSAR		TOOLBOX 7d3efbeb	Istituto di	Skin sensi	OutOfDor	Human H	["VEGA or	-	-	Qsar pred
VEGA - Skin Sensitization model (CAESAR)	Sensitizer		QSAR		TOOLBOX bc0a335b	Istituto di	Skin sensi	OutOfDor	Human H	["VEGA or	-	-	Qsar pred
Allergic Contact Dermatitis, Guinea Pig and Human - Danish QSAR DB Leadscope model	Negative		QSAR		TOOLBOX 26b06772	Danish EP	Skin sensi	InDomain	Human h	["Domain:	-	-	Qsar pred
Photoinduced toxicity based on 3T3 NRU data	Not Phototoxic		QSAR		TOOLBOX f2d0c665	Laboratori	3T3 NRU	InDomain	Human H	["Predictic	-	-	Qsar pred
Ashby Structural Alerts - Dani												-	Qsar pred
Chromosome Aberrations in												-	Qsar pred
Unscheduled DNA Synthesis												-	Qsar pred
Syrian Hamster Embryo (SHE)												-	Qsar pred
Chromosome Aberrations in												-	Qsar pred
Comet Assay in Mouse - Dani												-	Qsar pred
Potent Ames Mutagens, Reve												-	Qsar pred
Dominant Lethal Mutations i												-	Qsar pred
Sister Chromatid Exchange in												-	Qsar pred
Unscheduled DNA Synthesis												-	Qsar pred
Dominant Lethal Mutations in Rodents - Danish QSAR DB CASE Ultra model	Negative		QSAR		TOOLBOX 3ddaf03e	Danish EP	Chromosc	InDomain	Human h	["Domain:	-	-	Qsar pred

Simple Report for one substance:
full metadata

	Name	(2-Oxopropyl)amine			1,2-Diacetylbenzene		
	SMILES	CC(=O)CN			CC(=O)c1ccccc1C(C)=O		
	CAS	298088			704007		
	ChemId	82c0b2ac-d0a3-a917-ef6f-8501992e9624			c4206313-e4e0-b5cd-d972-2bb9a9dcd3d4		
Internaltype	Name	Source	Value	Unit	AD	Value	Unit	AD
QSAR	Liver Specific Cancer in Rat or Mouse - Danish QSAR DB Leadscope model	Danish EPA	Negative		InDomain			OutOfDomain
QSAR	FDA RCA Cancer Female Rat - Danish QSAR DB CASE Ultra model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	FDA RCA Cancer Mouse - Danish QSAR DB Leadscope model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	FDA RCA Cancer Female Mouse - Danish QSAR DB Leadscope model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	FDA RCA Cancer Female Rat - Danish QSAR DB Leadscope model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	FDA RCA Cancer Male Rat - Danish QSAR DB CASE Ultra model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	FDA RCA Cancer Male Mouse - Danish QSAR DB Leadscope model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	VEGA - Carcinogenicity in male rat (CORAL)	Istituto di Ricerche Farmacologiche Mario Negri	35.6578074709958 2	mg/kg bdwt/d	InDomain	1299.58614600944 2	mg/kg bdwt/d	InDomain
QSAR	FDA RCA Cancer Male Mouse - Danish QSAR DB CASE Ultra model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	VEGA - Carcinogenicity model (IRFMN-ISSCAN-CGX)	Istituto di Ricerche Farmacologiche Mario Negri	Possible NON-Carcinogen		OutOfDomain	Possible NON-Carcinogen		InDomain
QSAR	VEGA - Carcinogenicity inhalation classification model (IRFMN)	Istituto di Ricerche Farmacologiche Mario Negri	NON-Carcinogen		InDomain	NON-Carcinogen		InDomain
QSAR	FDA RCA Cancer Mouse - Danish QSAR DB CASE Ultra model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	Liver Specific Cancer in Rat or Mouse - Danish QSAR DB CASE Ultra model	Danish EPA			OutOfDomain	Negative		InDomain
QSAR	VEGA - Carcinogenicity model (ISS)	Istituto di Ricerche Farmacologiche Mario Negri	NON-Carcinogen		OutOfDomain			InDomain
QSAR	FDA RCA Cancer Rodent - Danish QSAR DB Leadscope model	Danish EPA			OutOfDomain			OutOfDomain
QSAR	VEGA - Carcinogenicity oral Slope Factor model (IRFMN)	Istituto di Ricerche Farmacologiche Mario Negri	2.90967136940088 58	mg/kg bdwt/d	InDomain	274.140963918774 73	mg/kg bdwt/d	InDomain
QSAR	FDA RCA Cancer Female Mouse - Danish QSAR DB CASE Ultra model	Danish EPA	Negative		InDomain	Negative		InDomain
QSAR	VEGA - Carcinogenicity model (IRFMN-Antares)	Istituto di Ricerche	Possible		OutOfDomain	Possible		InDomain

Master Report comparing multiple substances

Consolidated Report for one substance

Consolidated report

Consolidated results

Carcinogenicity

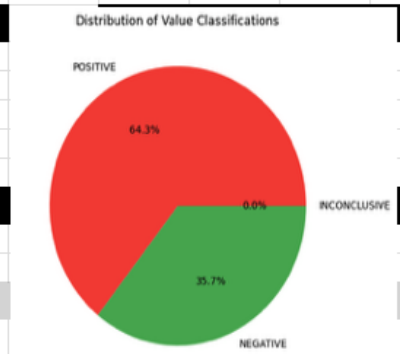
Summary

Category	Count
POSITIVE	9
NEGATIVE	5
INCONCLU	0

Number of models with verbose output

Count
1

Out of Domain model: 2



Distribution of predictions

Germ Cell Mutagenicity

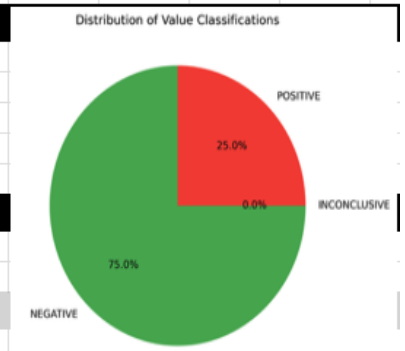
Summary

Category	Count
POSITIVE	5
NEGATIVE	15
INCONCLU	0

Number of models with verbose output

Count
2

Out of Domain model: 0



Distribution of predictions

Reproductive Developmental Toxicity

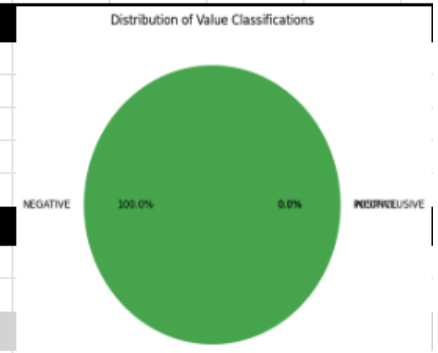
Summary

Category	Count
POSITIVE	0
NEGATIVE	7
INCONCLU	0

Number of models with verbose output

Count
1

Out of Domain model: 2



Distribution of predictions

Endocrine Disruption

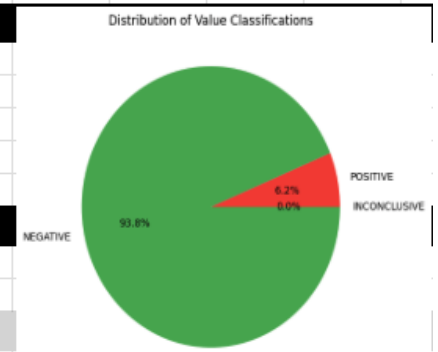
Summary

Category	Count
POSITIVE	1
NEGATIVE	15
INCONCLU	0

Number of models with verbose output

Count
1

Out of Domain model: 2



Distribution of predictions

The Jaqpotpy library for the development and integration of models

jaqpotpy is a Python library that helps you build models and use the Jaqpot platform. You can use it to create, train and deploy machine learning models based on **Scikit-learn**, **PyTorch** and **PyTorch Geometric**.

README MIT license

Build and test passing Publish to PyPI passing

Jaqpotpy

The jaqpotpy library enables you to upload and deploy machine learning models to the Jaqpot platform. Once uploaded, you can manage, document, and share your models via the Jaqpot user interface at https://app.jaqpot.org. You can also make predictions online or programmatically using the Jaqpot API.

Getting Started

Prerequisites

- Python 3.10
- An account on https://app.jaqpot.org

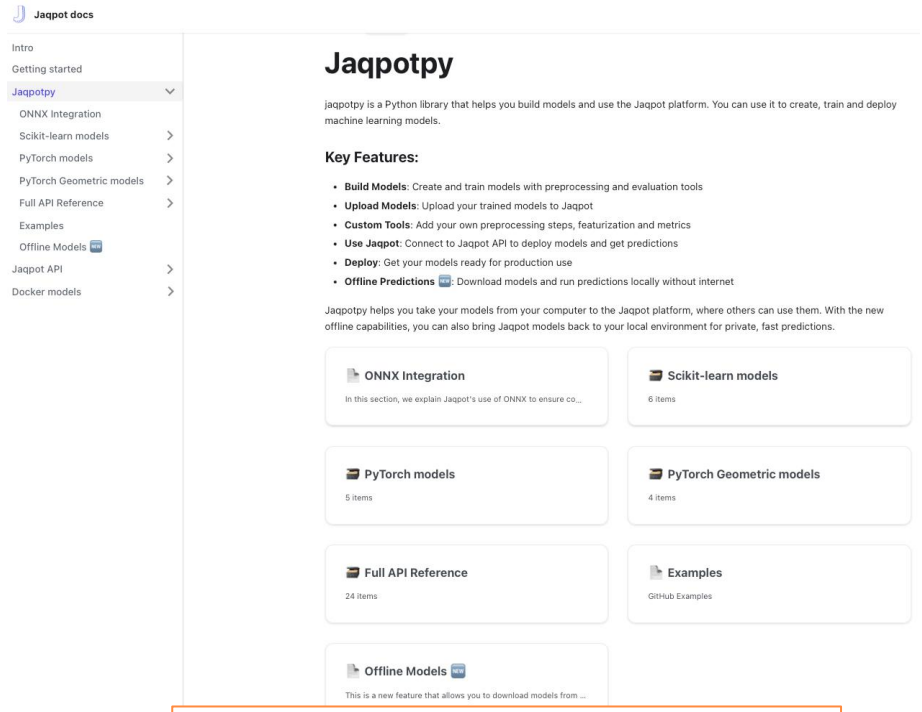
Installation

Install jaqpotpy using pip:

```
pip install jaqpotpy
```

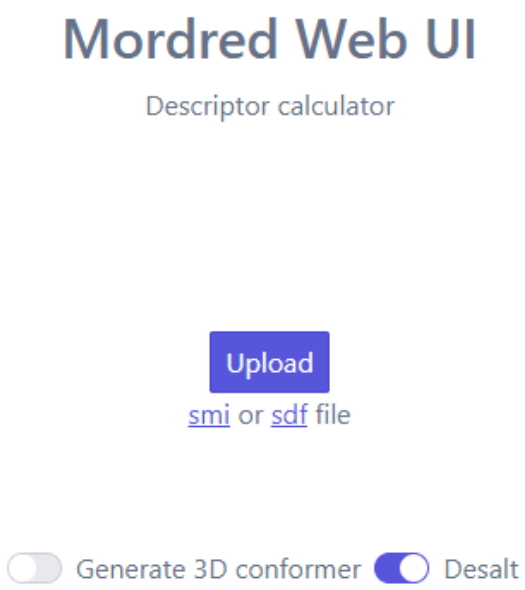
Model Training and Deployment

Source code here: https://github.com/ntua-unit-of-control-and-informatics/jaqpotpy



Tutorials and documentation is available here: https://jaqpot.org/docs/jaqpotpy/

Automatic calculation of descriptors



https://mordred-descriptor.github.io/documentation/master/descriptors.html

The Mordred library includes more than 1800 descriptors which are grouped into 50 categories.

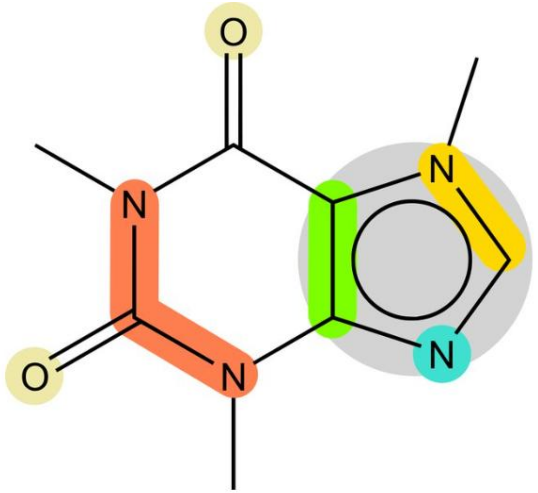
Summary

#	NAME	
1819	WPol	113.0
1820	Zagreb1	320.0
1821	Zagreb2	399.0
1822	mZagreb1	19.46
1823	mZagreb2	11.37

Moriwaki, H., Tian, YS., Kawashita, N. *et al.* Mordred: a molecular descriptor calculator. *J Cheminform* **10,** 4 (2018).
Molecular descriptor guide https://www.epa.gov/sites/default/files/2015-05/documents/moleculardescriptorsguide-v102.pdf

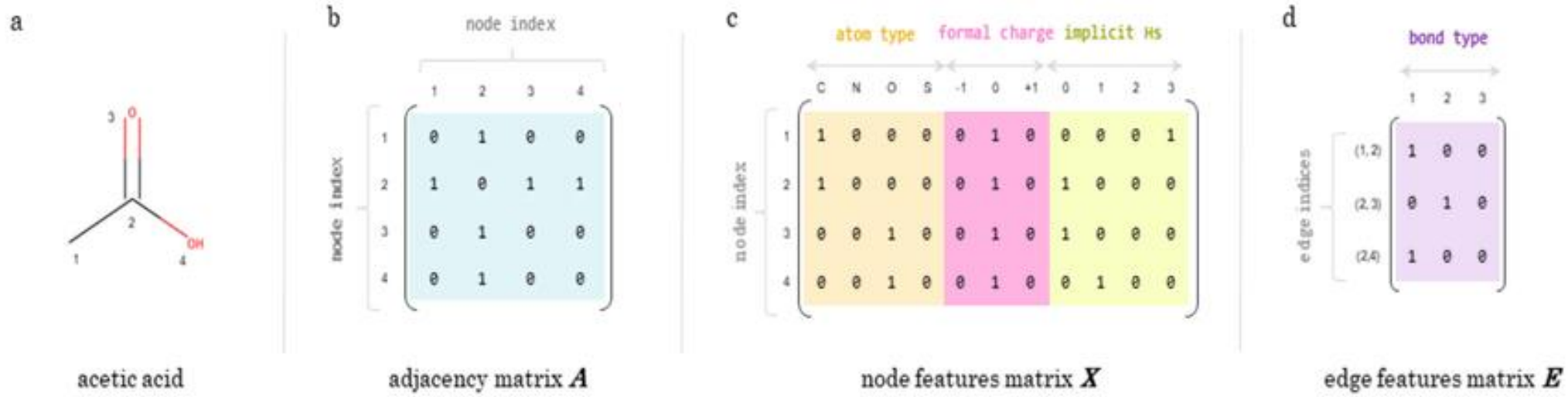
Automatic calculation of fingerprints

The MACCS key (Molecular ACCess System) is a binary fingerprint (a string of 166 positions where each position contains 0 or 1). Each bit position represents the presence (=1) or absence (=0) of a predetermined structural feature. The feature definitions for MACCS keys are available at the following address:
https://github.com/rdkit/rdkit/blob/master/rdkit/Chem/MACCSkeys.py



Key	Description
65	N in aromatic bonds with C
77	N separated by 2 bonds
83	heteroatoms in 5 ring
89	O separated by 4 bonds
99	C in C=C
162	aromatics

Graph Representations of chemical substances



acetic acid

adjacency matrix A

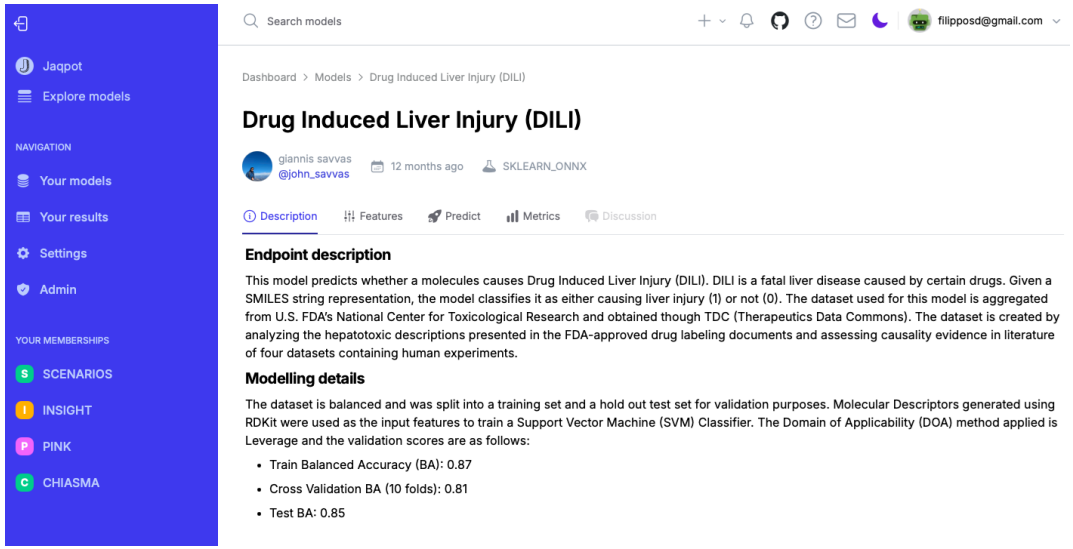
node features matrix X

edge features matrix E

The Jaqpotpy library for the development and integration of models

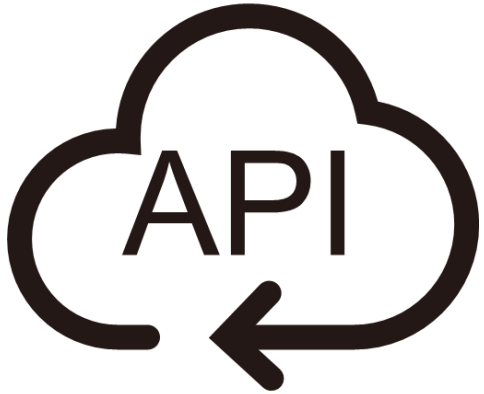
jaqpotpy assists users during models' lifecycle in feature preprocessing, model evaluation and DoA calculations.

```python
import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestRegressor
from jaqpotpy import Jaqpot
from jaqpotpy.datasets import JaqpotTabularDataset
from jaqpotpy.models import SklearnModel

# Creating a Simulated Dataset for Model Training
np.random.seed(42)
X1 = np.random.rand(100)
X2 = np.random.rand(100)
ACTIVITY = 2 * X1 + 3 * X2 + np.random.randn(100) * 0.1
df = pd.DataFrame({"X1": X1, "X2": X2, "ACTIVITY": ACTIVITY})
y_cols = ["ACTIVITY"]
x_cols = ["X1", "X2"]

# Step 1: Create a Jaqpotpy dataset
dataset = JaqpotTabularDataset(df=df, y_cols=y_cols, x_cols=x_cols, task=ModelTask

# Step 2: Build a model
rf = RandomForestRegressor(random_state=42)
myModel = SklearnModel(dataset=dataset, model=rf)
myModel.fit()

# Step 3: Upload the model on Jaqpot
jaqpot = Jaqpot()
jaqpot.login()  # log in to Jaqppt
myModel.deploy_on_jaqpot(
    jaqpot=jaqpot,
    name="Demo: Regression",
    description="This is a description",
    visibility=ModelVisibility.PRIVATE
)
```



https://app.jaqpot.org/dashboard/models/2019/description

Deployment leads to a model accessible both over its dedicated webpage and over API



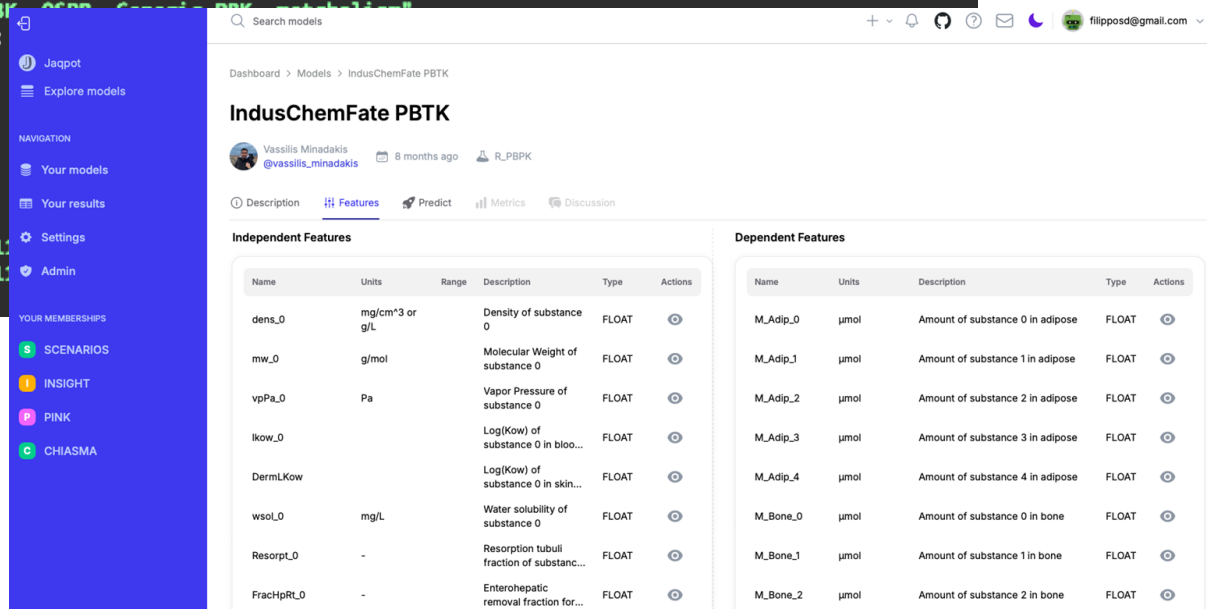
https://github.com/ntua-unit-of-control-and-informatics/jaqpot-google-colab-examples/blob/main/Scikit-learn-models/develop_deploy_a_model_with_chemical_descriptors.ipynb

Generating and integrating new PBK models

The model creator can provide information that will accompany the model as metadata from its creation. Fields can be customised to the PINK annotation schema, for example the type of impact covered.


```
{
  "name": "IndusChemFate PBTK",
  "type": "R_PBPK",
  "libraries": [
    {
      "name": "deSolve",
      "version": "1.40",
      "id": 62462,
      "createdAt": "2025-05-13T11:18:48.819025Z",
      "updatedAt": "2025-05-13T11:18:48.819025Z"
    }
  ],
  "dependentFeatures": [
    {
      "key": "M_Adip_0",
      "name": "M_Adip_0",
      "featureType": "FLOAT",
      "id": 41782,
      "units": "µmol",
      "range": null,
      "description": "Amount of substance 0 in adipose",
      "featureDependency": "DEPENDENT",
      "visible": null,
      "possibleValues": null,
```

```
    "username": "vassilis_minadakis",
    "firstName": "Vassilis",
    "lastName": "Minadakis",
    "email": "vassilisminadakis@gmail.com",
    "emailVerified": true,
    "avatarUrl": "https://d2zoqz4gyxc03g.cloudfront.net/avatars/8ba56047-97db-4954-affb-b4f78bab0c05.jpg",
    "canEdit": null,
    "createdAt": null
  },
  "canEdit": false,
  "isAdmin": false,
  "selectedFeatures": null,
  "tags": "IndusChemFate, PBK
  "legacyPredictionService":
  "scores": {
    "train": [],
    "test": [],
    "crossValidation": []
  },
  "rPbpkConfig": null,
  "dockerConfig": null,
  "createdAt": "2025-05-13T1
  "updatedAt": "2025-05-13T1
}
```

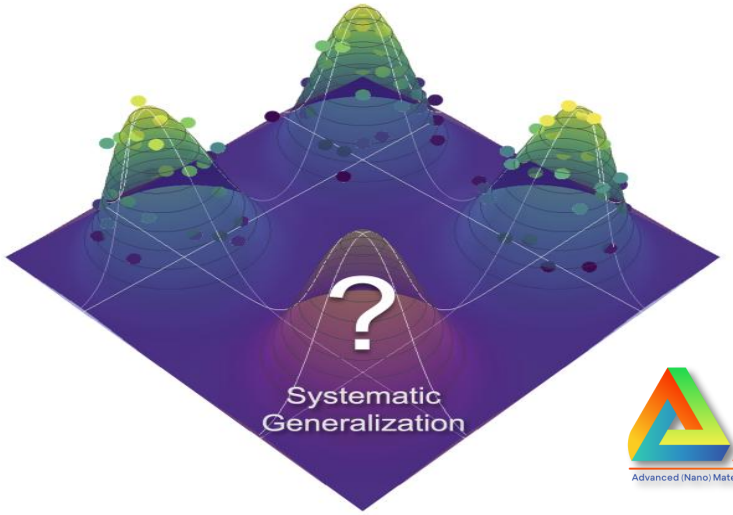



Key metadata are communicated over the UI

https://app.jaqpot.org/dashboard/models/2124/description

Generative modelling approach

- The goal is to generate molecules with **optimized properties**.
- Molecules should be **valid** and **diverse**
- **Generative Flow Network (GFlowNet)**[1] is used .
- This particular generative framework is chosen due to it's **explorative** nature to generate a diverse set of high-quality solutions with desired properties.

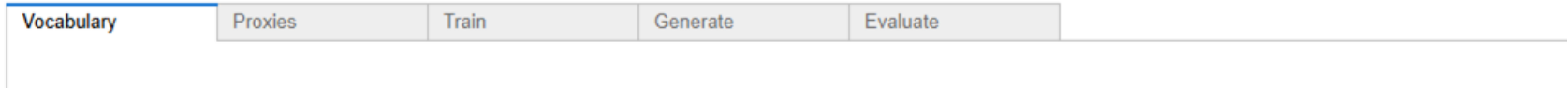


1. https://arxiv.org/abs/2106.04399

GenAI workflow

3. Train a GFlowNet (molecule generator using fragments, reward guided by proxy models):

- Open **gflownet_trainer.ipynb** notebook inside the repo and Run All Cells.
- Navigate through different tabs:

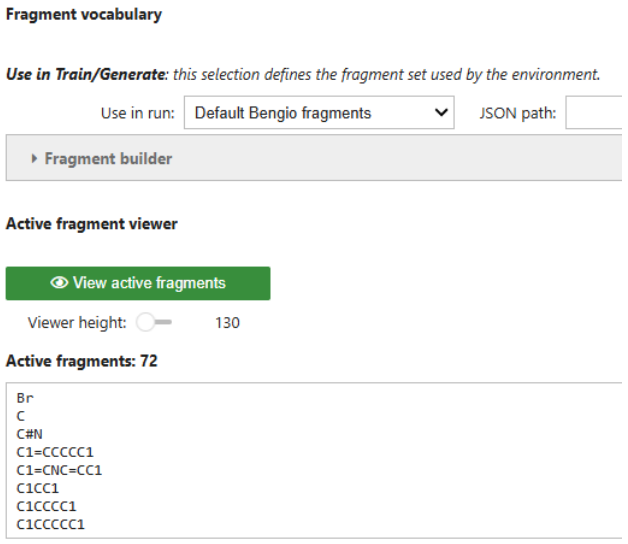


Example for Multi-Objective Property Optimization:

- ➢ Target specific range for maximum absorbance wavelength (**λmax**): target = 300-360 nm
- ➢ Maximize absorbance factor (**f**)

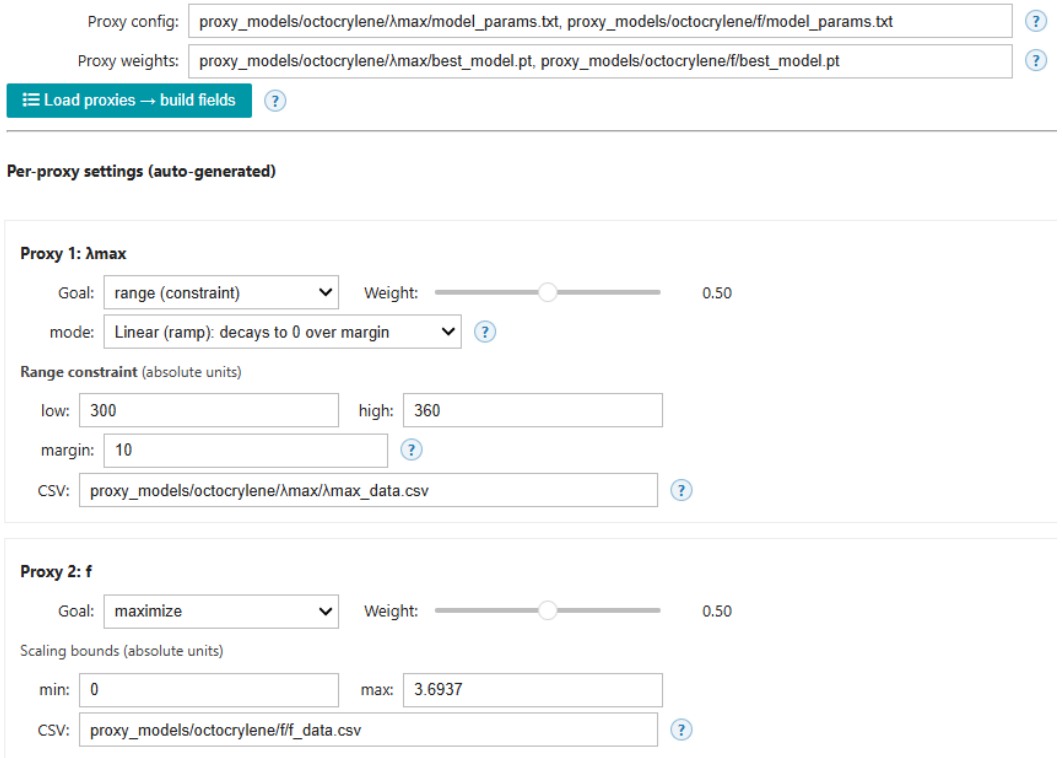
(1) Vocabulary:

Build or load the fragment set that defines the 'building blocks' for your molecules.



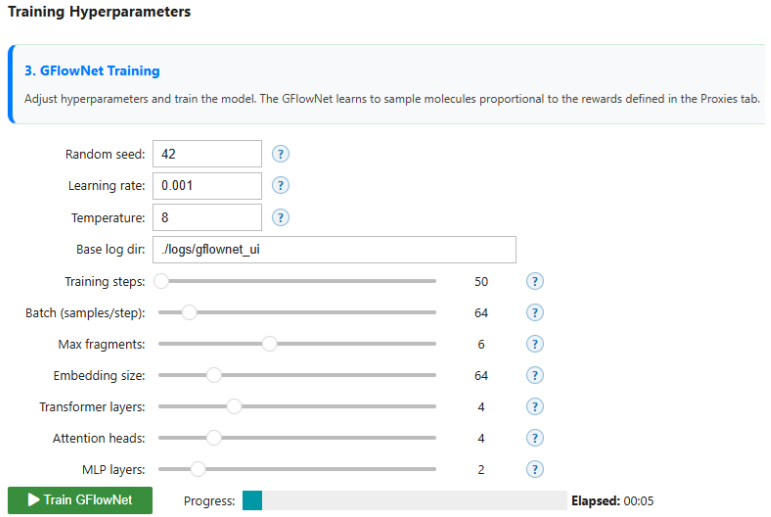
(2) Proxies:

Load your pre-trained proxy models and configure objectives for each property (minimization, maximization, targeted range, scaling bounds).



(3) Train:

Adjust hyperparameters and train the model.
The model is automatically saved in the indicated folder.
The GFlowNet learns to sample molecules proportional to the rewards defined in the Proxies tab.



>PINK

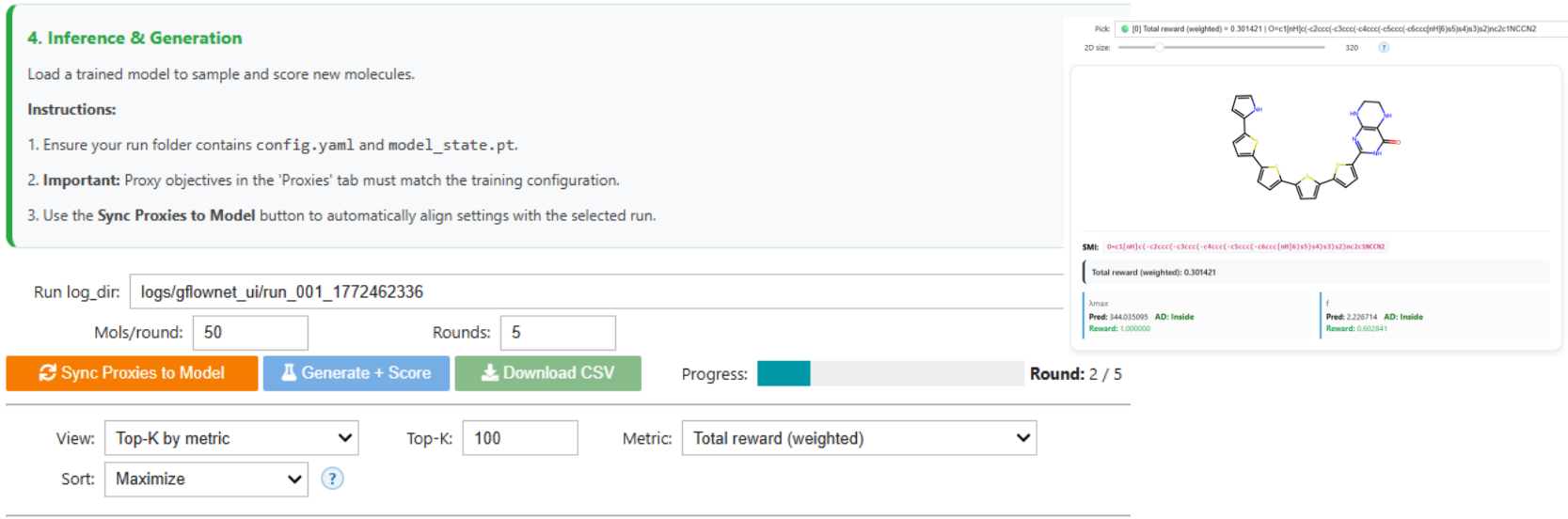
ANTHOS'26
Advanced (Nano) Materials and Technologies: From Needs to Solutions

GenAI workflow

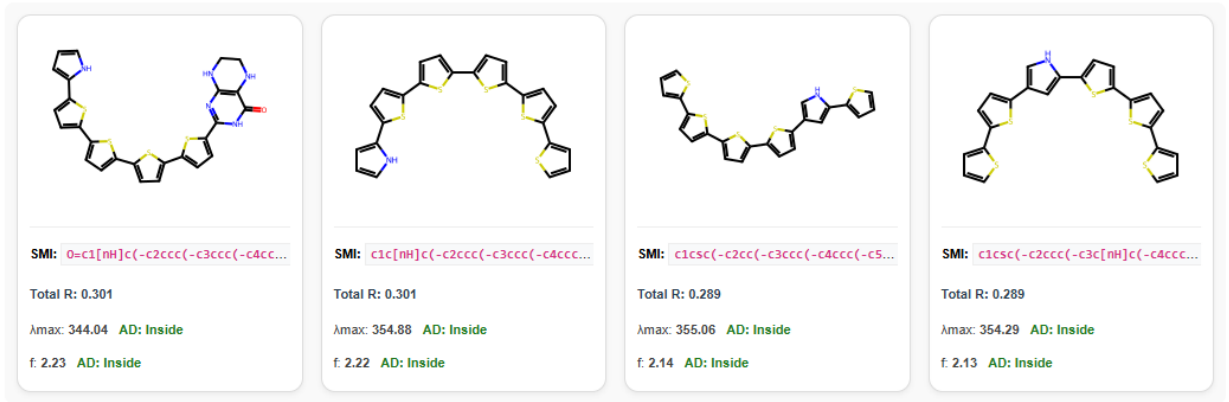
(4) Generate:

- Use the model you just trained or load a pre-trained model in Run log_dir.
- Sample molecules with Generate+Score.
- Download designs and visualize them.

Molecule cards collect all info about a specific design (including AD data)



Top 4 Designs (Sorted by Total Reward)



Visualize top candidates

Example for Multi-Objective Property Optimization:

- Target specific range for maximum absorbance wavelength (**λmax**): target = 300-360 nm
- Maximize absorbance factor (**f**)

Gather statistics

Top-100 UNIQUE summary (Total reward (weighted))

Summary (All selected)

	count	mean	std	min	max
Total reward (weighted)	100.0	0.259995	0.016565	0.233797	0.301421
λmax prediction	100.0	331.718048	10.569038	309.345154	357.026520
λmax reward	100.0	1.000000	0.000000	1.000000	1.000000
f prediction	100.0	1.920690	0.122370	1.727154	2.226714
f reward	100.0	0.519991	0.033129	0.467595	0.602841

Summary (Inside AD for all proxies) — 100/100

	count	mean	std	min	max
Total reward (weighted)	100.0	0.259995	0.016565	0.233797	0.301421
λmax prediction	100.0	331.718048	10.569038	309.345154	357.026520
λmax reward	100.0	1.000000	0.000000	1.000000	1.000000
f prediction	100.0	1.920690	0.122370	1.727154	2.226714
f reward	100.0	0.519991	0.033129	0.467595	0.602841

Summary (Outside AD for ≥1 proxy) — 0/100

No rows.

>PINK

ANTHOS'26
Advanced (Nano) Materials and Technologies: From Needs to Solutions

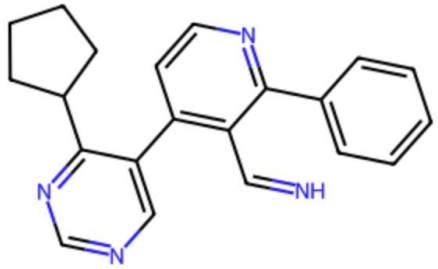
Illustrative example: Generating molecules with high logP values

Model-Framework is available on jaqpot: https://app.jaqpot.org/dashboard/models/2032/description

QSAR ToolBox LogP predictions

Proxy GNN model used in GFlowNet to predict LogP

Example of a generated molecule

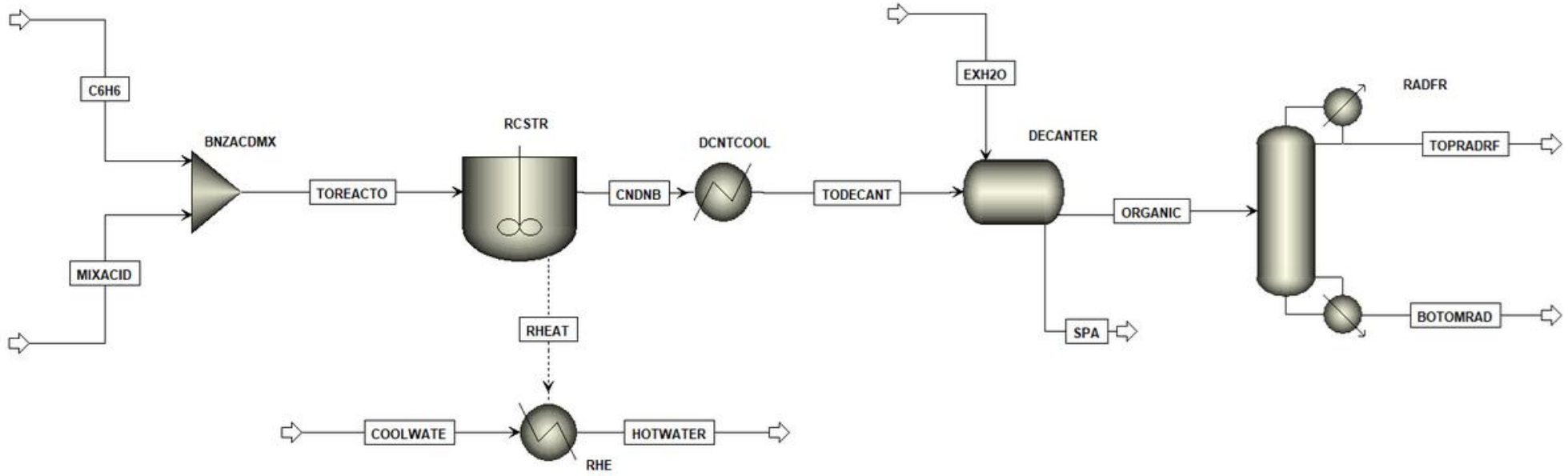


QSAR ToolBox LogP predictions	Prediction	SMILES
9.6566	8.277	Oc1cc(C2=CC(C3C=CCCC3)CCC2)cc(-c2ccc3ccccc3c2)c1
10.7758	9.521	CC(C)(C)C1C(C2=CCCCC2)CCCC1c1cccc(-c2cccs2)c1
8.7837	8.299	c1ccc(-c2ccnc(-c3ccc(C4CCC(c5ccccc5)CC4)cc3)c2)cc1
7.3031	6.107	c1cc(C2CC2)c(-c2ccc3ccccc3c2)c(C2COCC2C2CCCO2)c1
11.1945	9.373	C1CCC(C2CC(C3CCC(C4CCCC4)CC3)CC(C3CCOC3)C2)CC1
7.7512	6.581	CCc1ccc(C2OCCC2C2CCC(c3ccccc3)CC2)s1
6.8858	5.097	c1n[nH]c(C2CCCCC2)c1C1CCCCC1C1CCCCC1N1CCNCC1
4.9659	2.868	N=Cc1c(-c2cncnc2C2CCCC2)ccnc1-c1ccccc1

>PINK

ANTHOS'26 Advanced (Nano) Materials and Technologies: From Needs to Solutions

MOO Example – nitration of aromatic (benzene)



Input variables:

Variables	(units)	Limits
Reactor Temperature	(C)	[65.0,95.0]
Residence time	(hr)	[6.0,13.0]
Benzene to acids ratio	(w/w)	[0.0,1.0]

4 Objectives:

1. NB_Mol_Percentage_Yield (% mol NB/MOl C6H6) ⇨ Process efficiency (production capacity)
2. DNB_Mol_Yield (mol DNB/hr) ⇨ Safety related (toxic byprod.)
3. CO2e per kg NitroC6h6 (kg CO2/hr) ⇨ LCA related (emissions)
4. TOTALCOST ($) ⇨ LCC related (production costs)

Production process MOO - Workflow 3

Sensitivity analysis

Process Simulation (Aspen)

Input variables

Output variables

Matlab interface

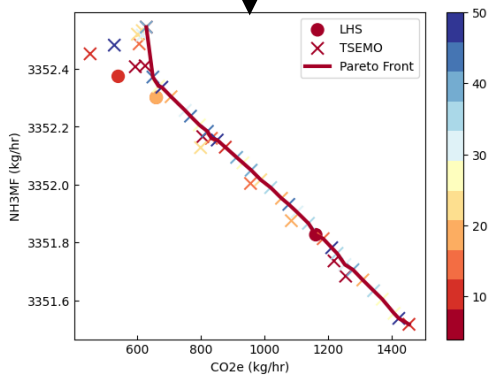
Input output dataset

Summit Emulator

Data-driven Process model

Input output bounds

TSEMO



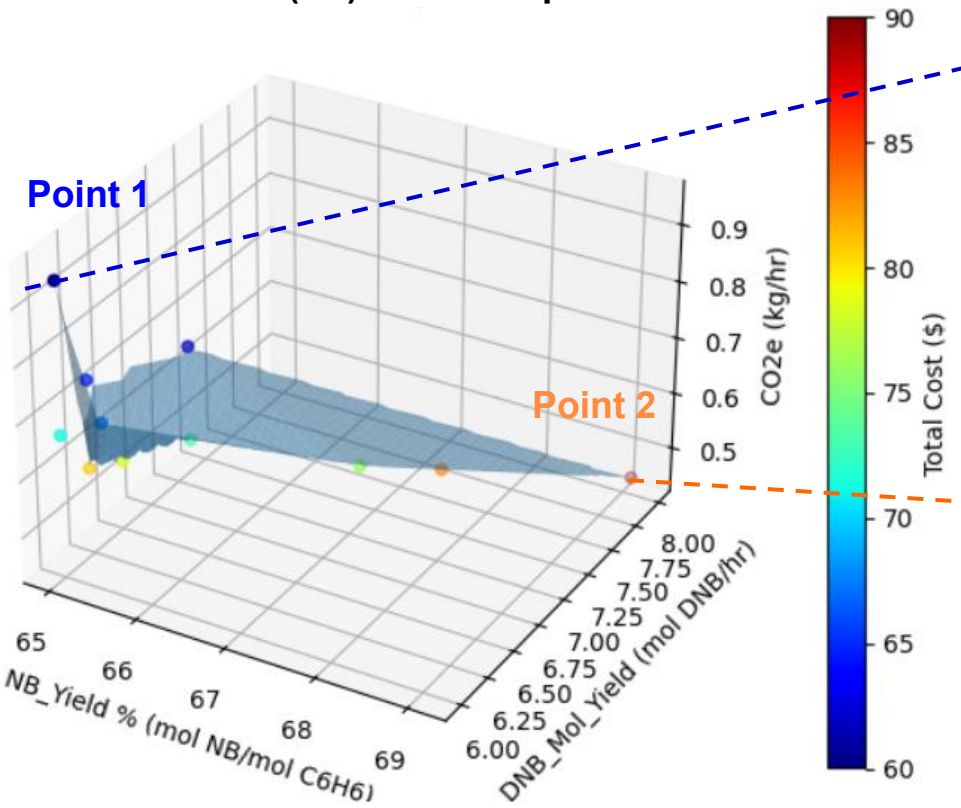
Pareto front

MOO Example – nitration of aromatic (benzene)

Reading the results

Values of the input parameters are decided by the designer

Pareto surface (2D) + heat map for Total cost



R_T DATA	ResTime DATA	Benz2AcidRatio DATA	NByield DATA	DNByield DATA	CO2e DATA	TOTALCOST DATA
65.0	6.0	0.940553	60.0	0.0	11.0	13811.0
65.724664	6.563721	0.770414	61.987709	0.000173	11.413369	14276.293945
65.030944	6.206928	0.738339	64.898796	0.000194	12.32363	14536.286133
65.480091	6.001707	0.717673	66.898422	0.000226	12.935511	14658.143555
65.0	6.0	0.673955	71.166206	0.000314	14.402148	15174.744141
65.988574	6.371139	0.650601	73.401672	0.000393	15.111883	15502.807617
67.528135	6.660242	0.627405	75.91481	0.000495	15.888021	15788.65918
65.393521	6.229343	0.607747	78.379135	0.000528	16.9161	16062.123047
65.054719	6.189286	0.586421	80.922569	0.000603	17.849514	16366.34668
67.931696	7.095091	0.566099	83.320786	0.000746	18.505066	16815.646484
69.108916	7.980441	0.455444	90.0	0.001396	21.0	18933.447266

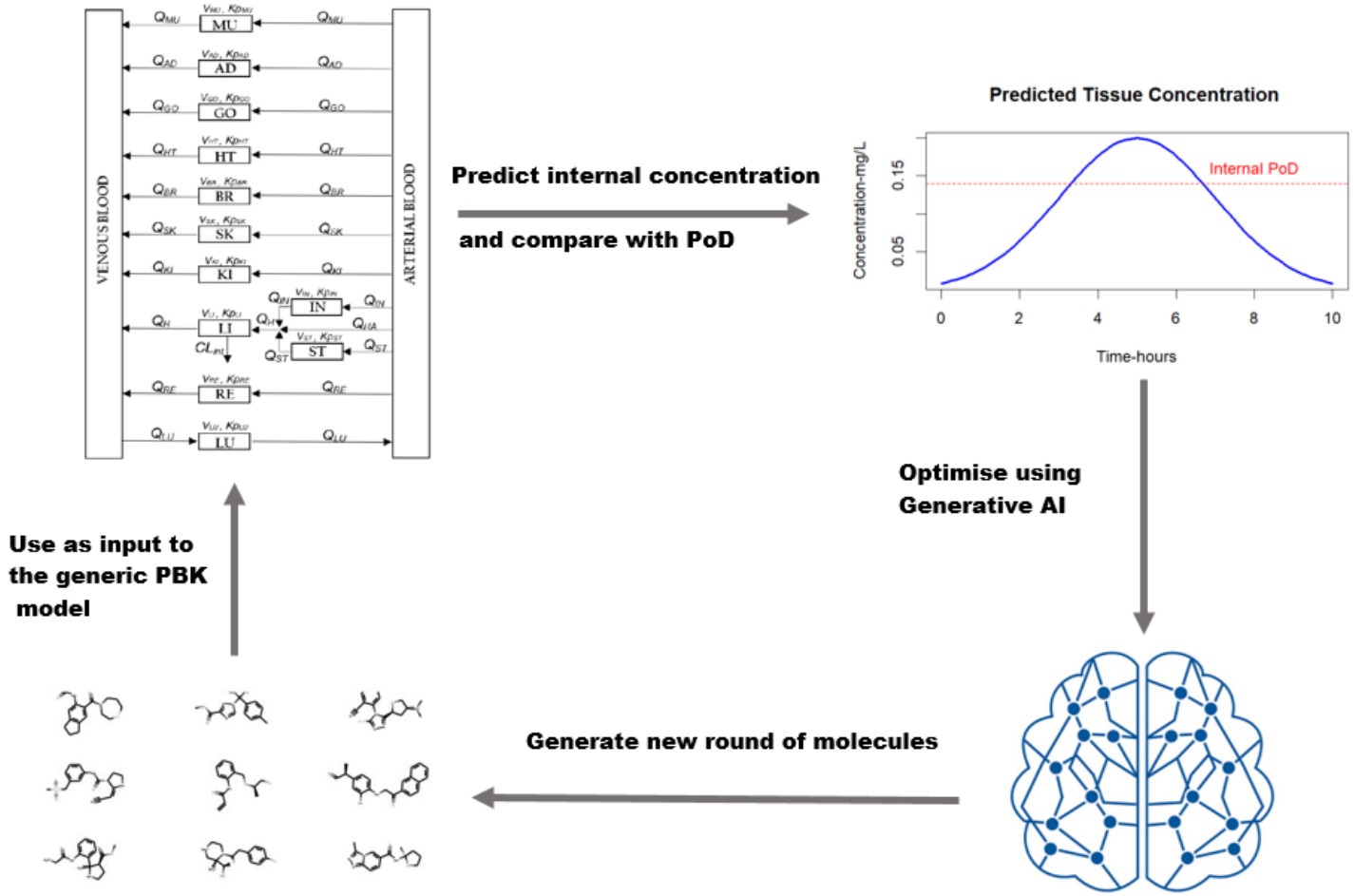
Example design options:

Point 1: minimum total cost, CO_2e, DNB **but** also minimum NB

Point 2: maximum total cost, CO_2e, NB **but** also maximum DNB

Integrating biokinetics in the PINK DSS

- Develop and integrate PBK models for small molecules with an embedded QSAR for predicting partition coefficients from chemical structures
- Utilise the PBK model **for forward dosimetry and estimate the internal burden at the target organ** (linked to the KE) following a toxicologically-relevant exposure
- Compute risk by comparing internal concentrations to **internal Points of Departure** (PoD) connected to specific Key Events (KE) in AOPs
- Include this analysis in an iterative learning process, penalise molecules that accumulate in organs and the resulting concentration exceeds the defined internal PoD



IndusChemFate PBTK model - A Generic PBTK model

- Originally developed in VisualBasic (Available through Excel Application)
- Structured by 11 compartments
- Exposure paths: Inhalation and Skin
- Elimination paths: Urinary, Exhalation, Metabolism
- Metabolism for up to 4 metabolites can be simulated
- Parameterization through QSARs

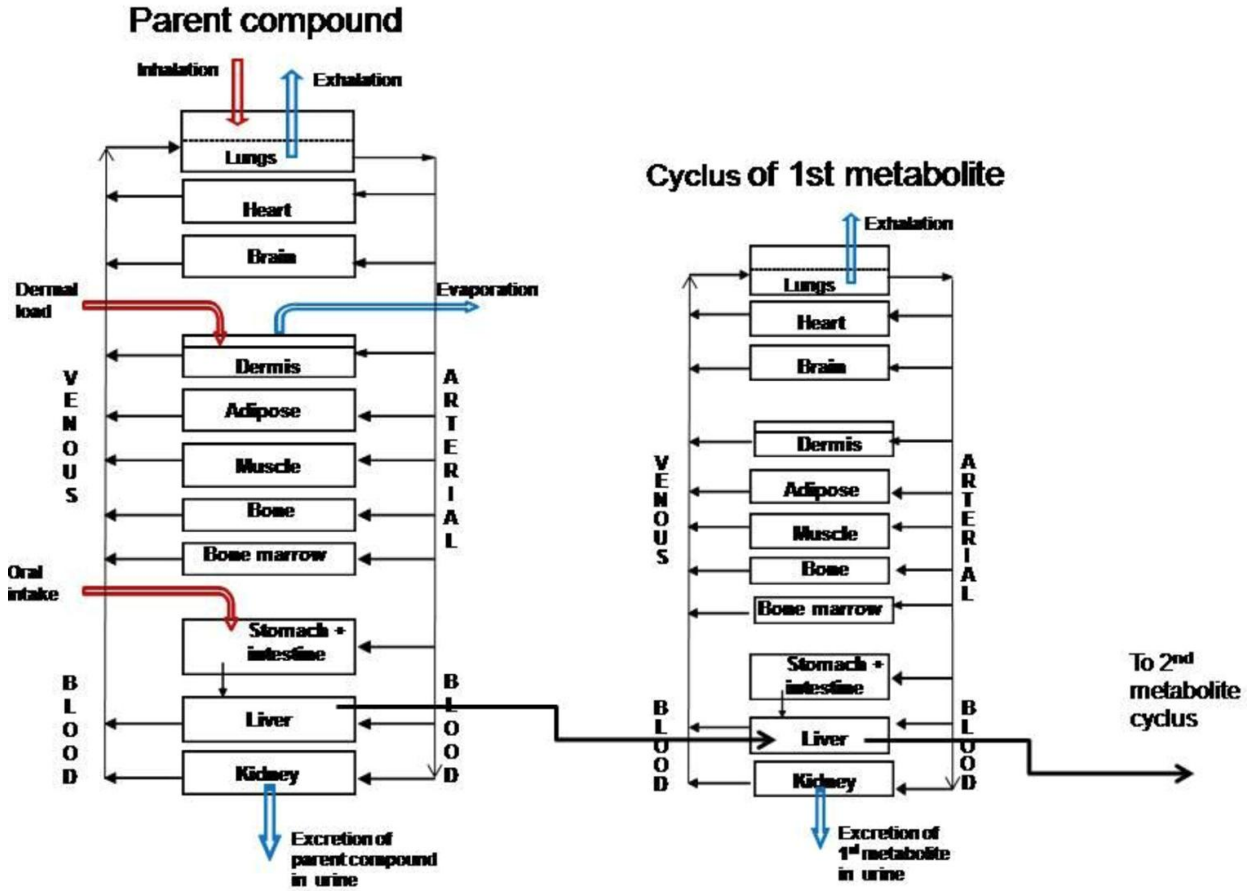
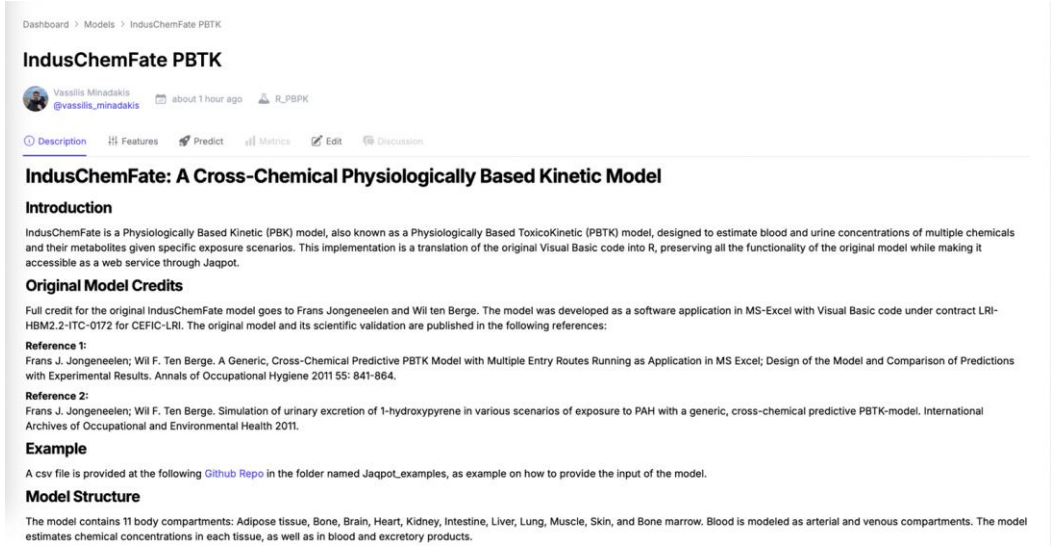


Figure PBTK_1: Structural representation of IndsuChemFate (sourced by Jongeneelen and ten Berge (2011).

Integration of IndusChemate on Jaqpot

Full Version of IndusChemFate Model

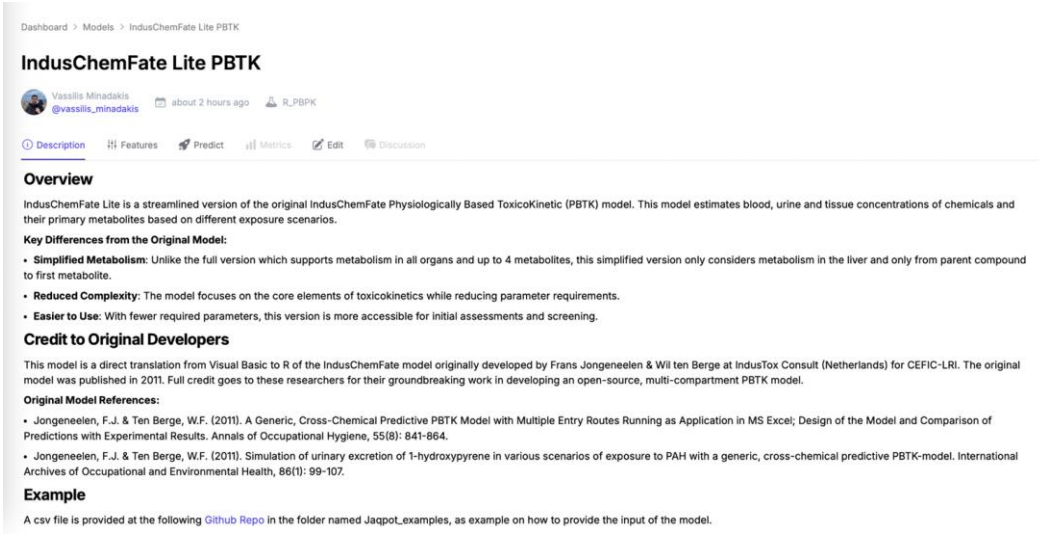
- Simulate up to 4 metabolites
- Applies metabolism processes to any compartment
- Large number of input parameters required as input
- Available User Interface and API



https://app.jaqpot.org/dashboard/models/2124/description

Lite Version of IndusChemFate Model

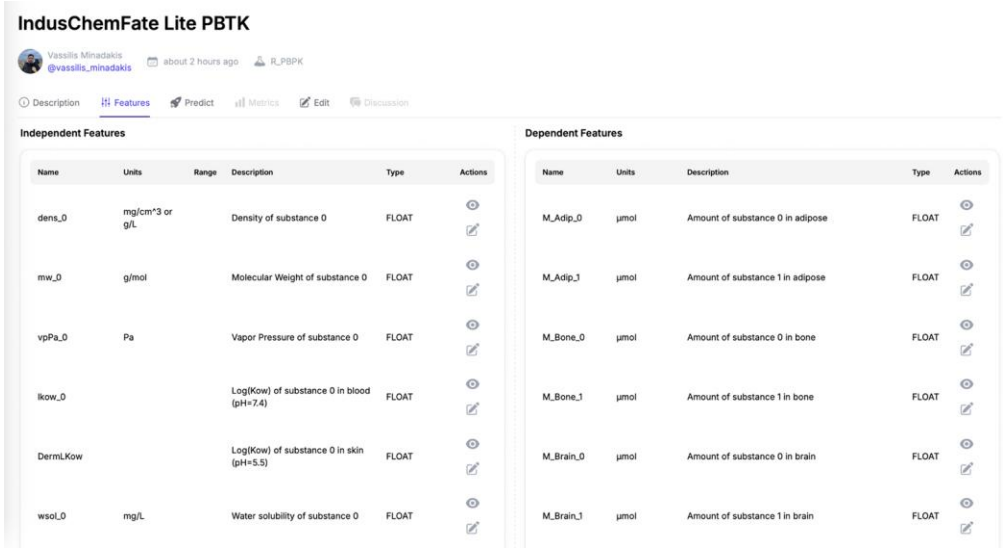
- Simulate up to 1 metabolite
- Applies metabolism processes only in Liver compartment
- Much less input parameters are required (more user-friendly).
- Available User Interface and API



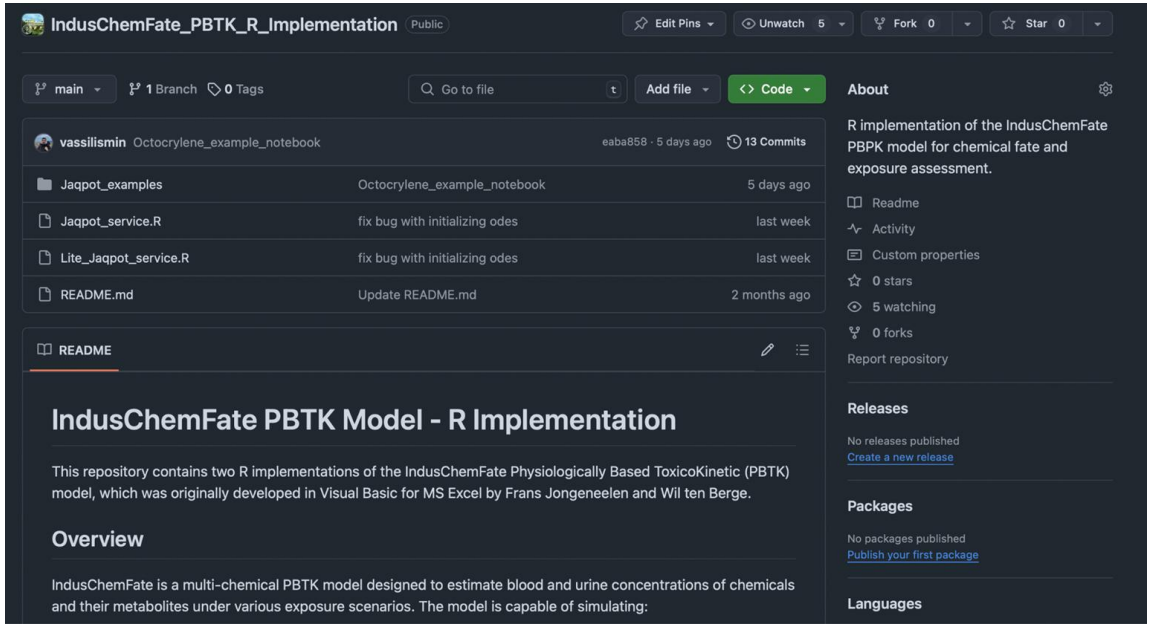
https://app.jaqpot.org/dashboard/models/2123/description

Integration of IndusChemate on Jaqpot

Fully documented: Descriptions and Units are provided for all parameters of the model through GUI.



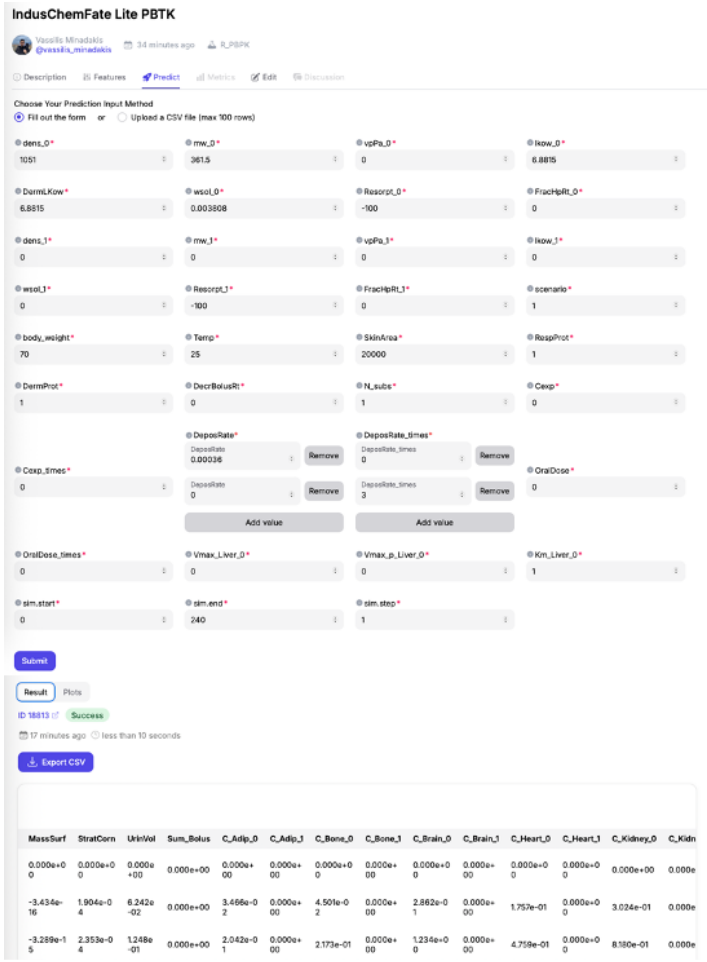
The R code of both models is provided in Github repo, along with CSV files to run demo simulations.



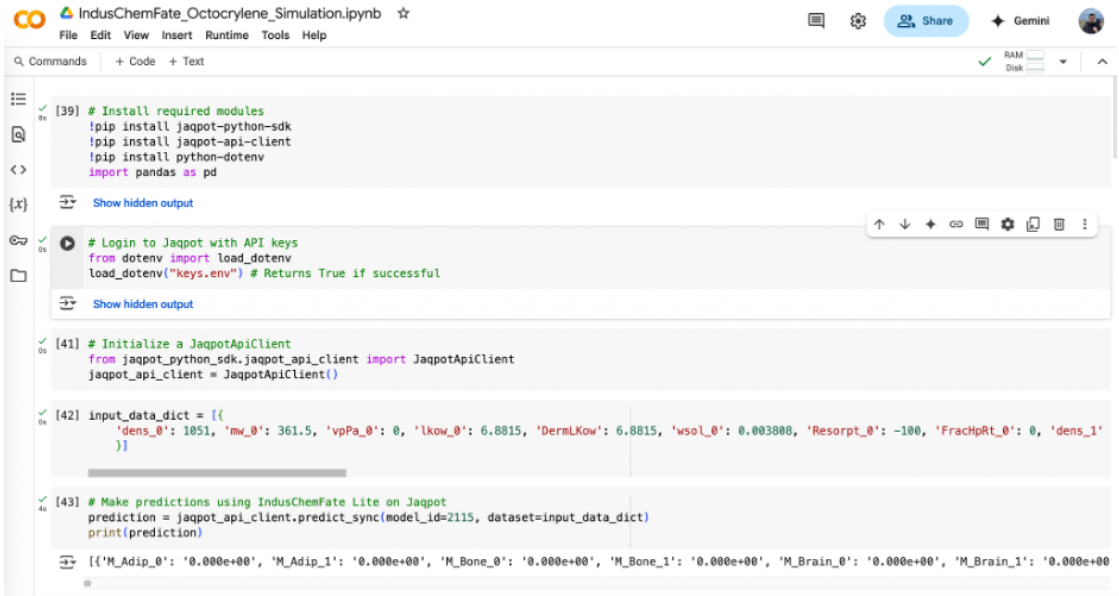
Available at: https://github.com/ntua-unit-of-control-and-informatics/IndusChemFate_PBTK_R_Implementation/tree/main

Octocrylene Example

Predictions using Jaqpot GUI

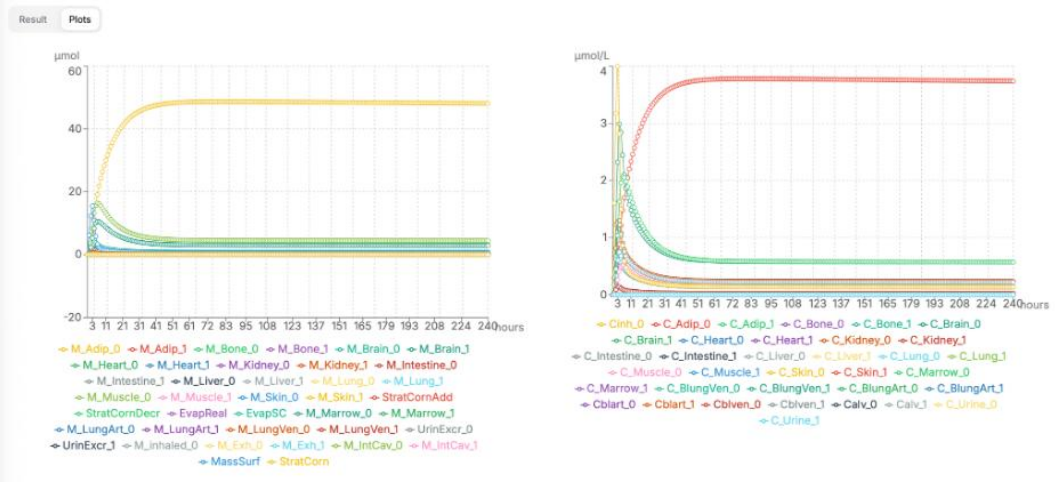


Predictions using Python SDK of Jaqpot

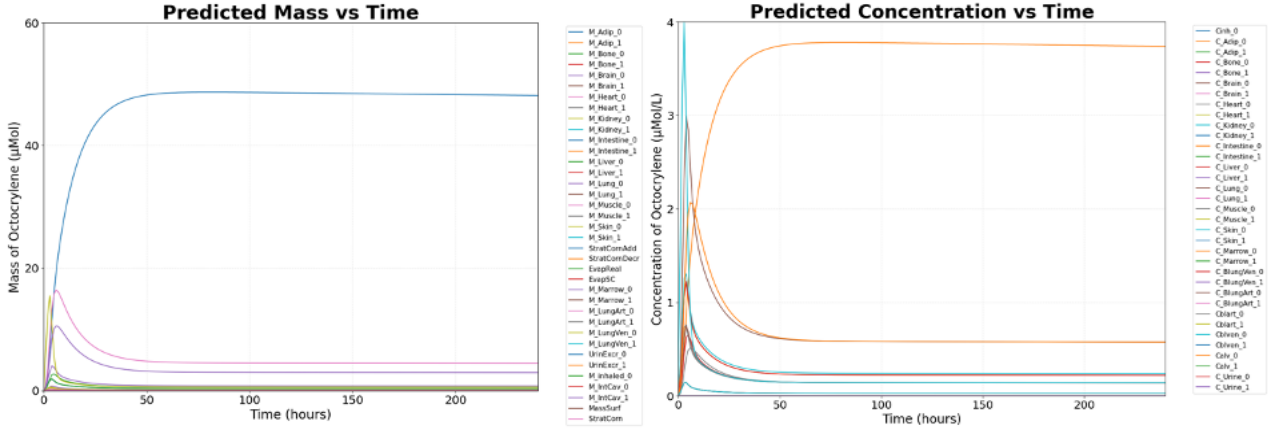


Biokinetics simuations, Octocrylene Example

Automatically created plots by Jaqpot GUI.



Plots created with python code through Google Colab notebook.



The PINK team

1 Harry Sarimveis

National Technical University of Athens (NTUA)

>THANK YOU!

WWW.PINK-PROJECT.EU

THE PINK PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON EUROPE RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 101137809.

ASSOCIATED PARTNERS (I.E. (A) SWISS PARTNERS AND (B) UK PARTNERS) HAVE RECEIVED NATIONAL FUNDING FROM (A) THE SWISS STATE SECRETARIAT FOR EDUCATION, RESEARCH AND INNOVATION (SERI), AND (B) INNOVATE UK.