NOVARTIS

Accelerating Drug Discovery: systematic application of PBPK/PD modelling supporting candidate selection

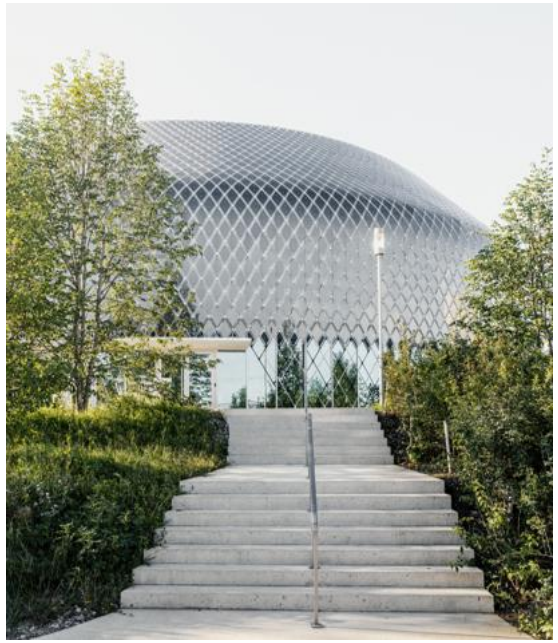
30-Sep-2025

Gregori Gerebtzoff

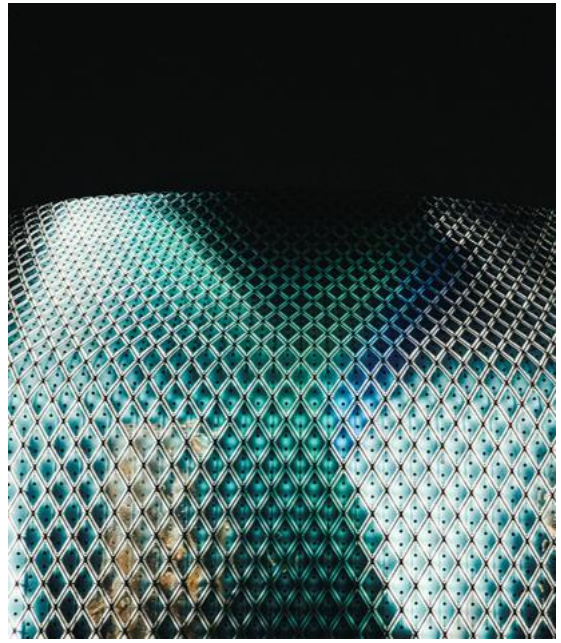
Agenda

1. Introduction

2. Leveraging PK-Sim snapshots



3. Data and application



4. Outlook

Introduction

NOVARTIS

Motivation

- At Novartis, early-stage PBPK modeling on low molecular weight compounds is rarely done due to lack of experimental data, lack of a streamlined process, and challenge to get and process all required data
- Our overall goal is to enable a better understanding of distribution on a project basis, and to couple multi-scale modeling to pharmacodynamic models
- We developed an integrated, high-throughput PBPK platform connected to our corporate database and in-house machine learning models to execute and visualize simulations at scale, leveraging the OSP suite
- The workflow can be applied at all stages of drug development, including to yet-to-be-synthesized virtual compounds, and supports portfolio-wide screening for complex mechanisms underlying disposition

Strengthening the Discovery PBPK strategy by developing a HT platform

List of compounds

```
["NOVARTIS-111",
 "NOVARTIS-222",
 "NOVARTIS-333",
 …]
```

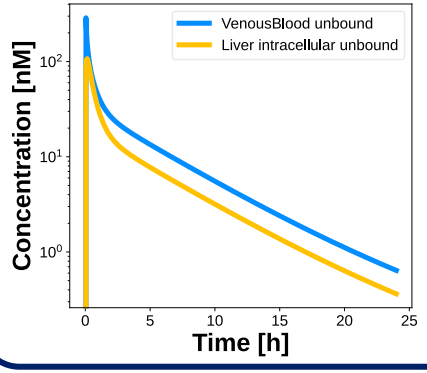
Automated NVP-specific data collection

High-throughput PBPK simulator

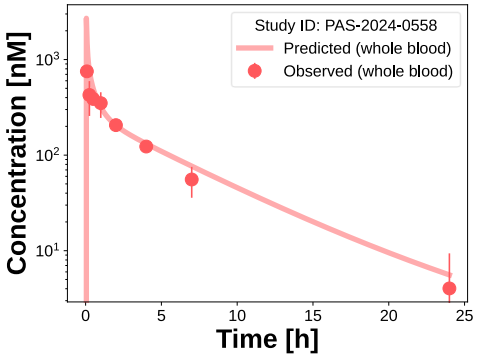
PK-Sim®

esqLABS
WE EMPOWER HEALTH CARE

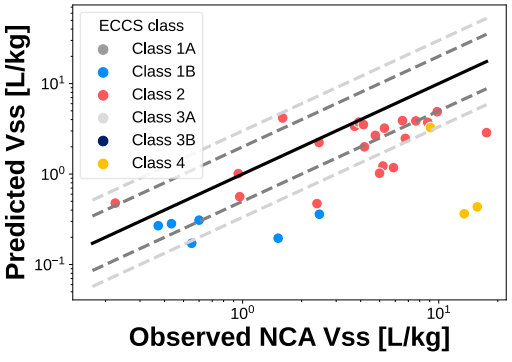
Tissue-specific PK predictions



Comparison of predicted vs. observed C-t profiles

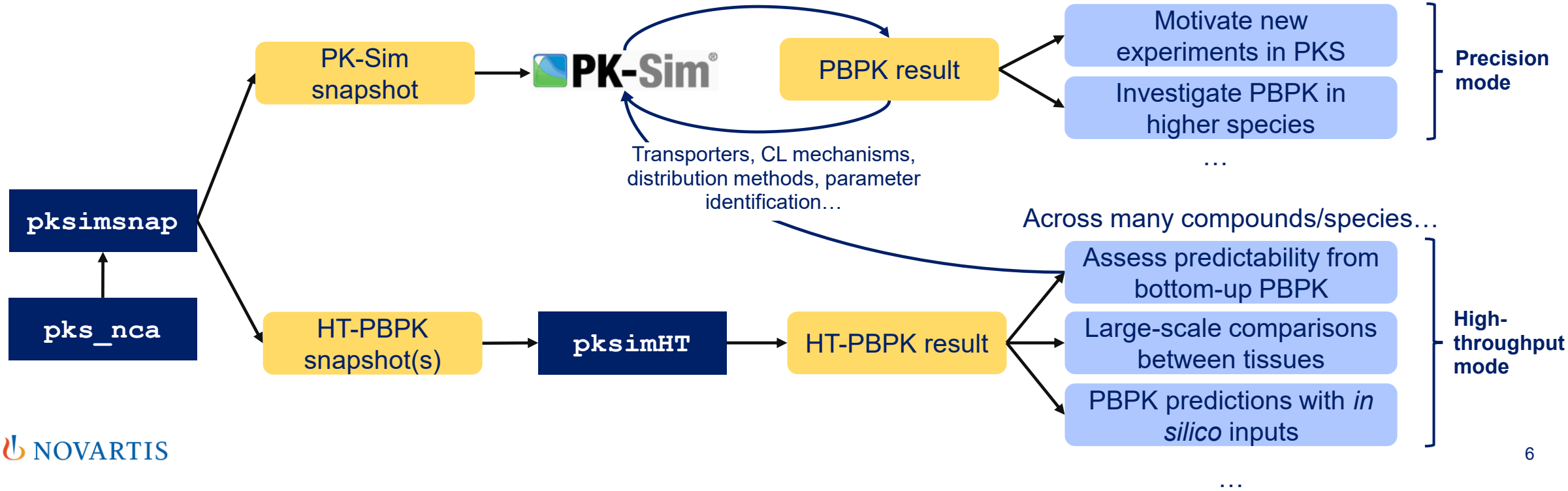


NCA of predicted (and observed) C-t profiles



High-throughput generation of PK-Sim simulations

- We developed a set of Python packages to collect data, run simulations, process and visualize results, and run non-compartmental analysis on simulations to compare with observed data
- To run simulations, we leverage PK-Sim through **`OSPSuite-R`** functions, developed in collaboration with ESQLabs. The **`pksimHT`** Python package is wrapping these R functions



Leveraging PKSim's Snapshots

NOVARTIS

Open Systems Pharmacology - Snapshot

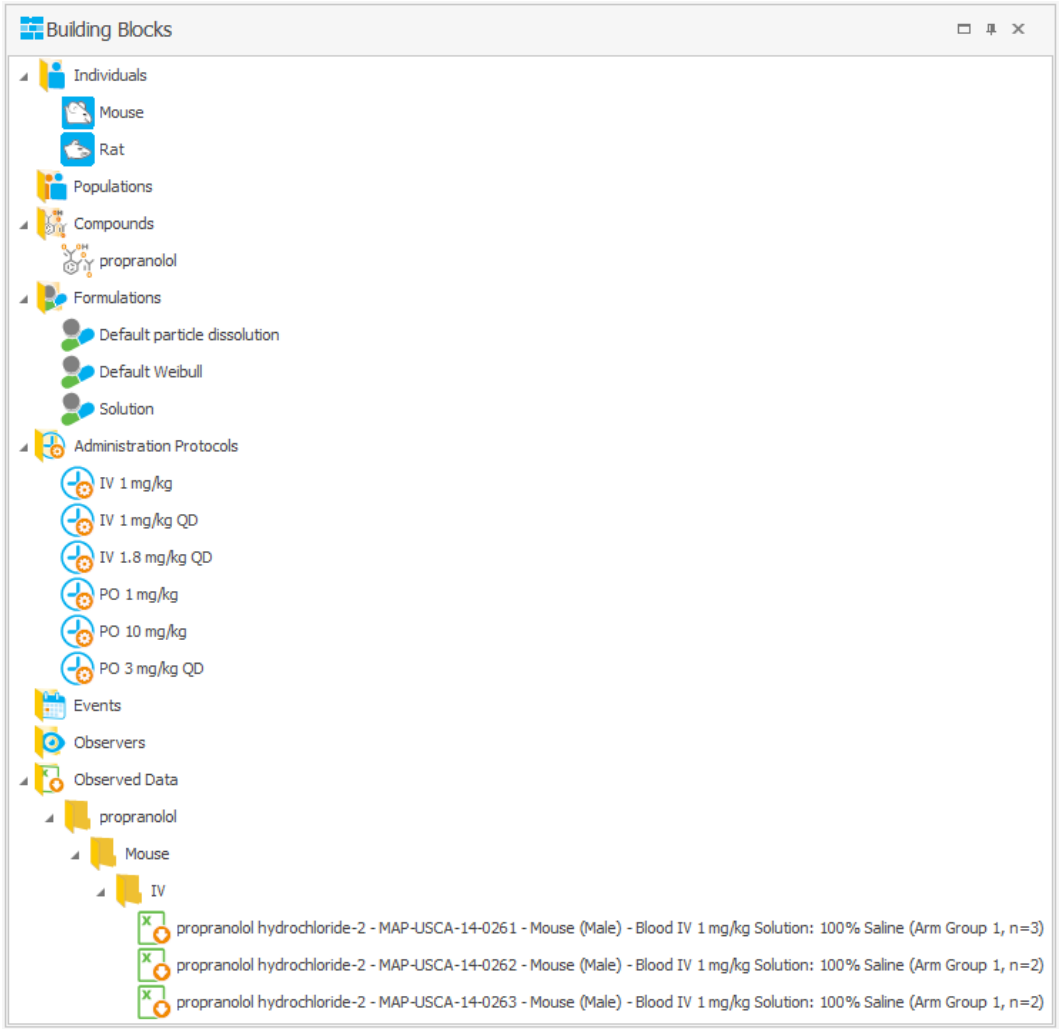
“A project snapshot contains the minimal amount of information required to recreate the project from scratch. This includes the information on primary substance specific input parameters (e.g. molecular properties like molecular weight, lipophilicity, etc.) and required inputs (e.g. demographic characteristics) for defining the system parameters.

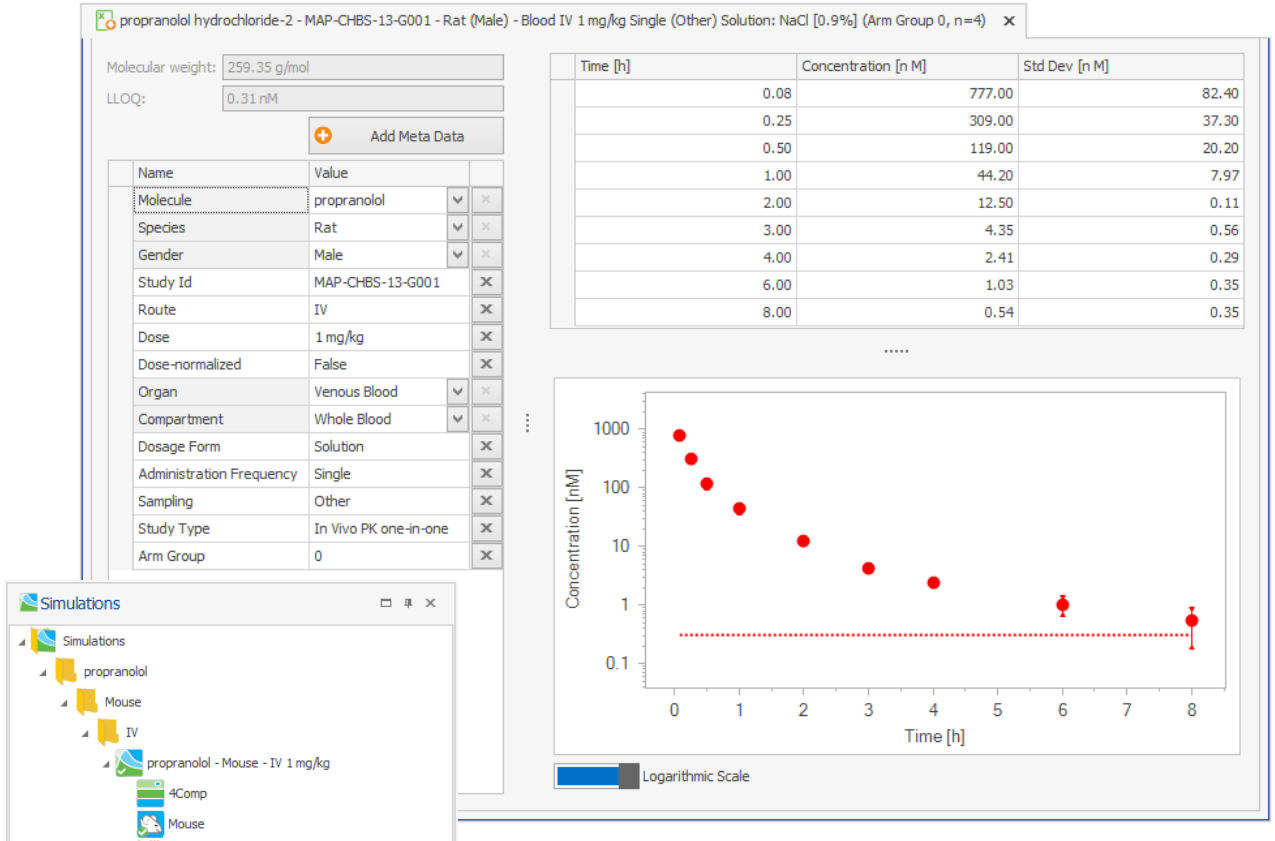
Project snapshots are human-readable text files in JSON format.”

Source: https://docs.open-systems-pharmacology.org/working-with-pk-sim/pk-sim-documentation/importing-exporting-project-data-models


```
{
  Version: 74,
  ▸ Individuals: [ … ],
  ▾ Populations: [
    ▾ {
        Name: "Healthy adults",
        Seed: 128991875,
      ▾ Settings: {
          NumberOfIndividuals: 100,
          ProportionOfFemales: 50,
        ▾ Age: {
            Min: 20.0,
            Max: 40.0,
            Unit: "year(s)"
          },
        ▸ Individual: { … }
      }
    }
  ],
  ▾ Compounds: [
    ▾ {
        Name: "SuperDrug",
        PlasmaProteinBindingPartner: "Albumin",
      ▸ Lipophilicity: [ … ],
      ▸ FractionUnbound: [ … ],
      ▸ Solubility: [ … ],
      ▾ CalculationMethods: [
          "Cellular partition coefficient method - PK-Sim Standard",
          "Cellular permeability - PK-Sim Standard"
        ],
      ▸ Parameters: [ … ]
    }
  ],
  ▾ Protocols: [
    ▾ {
        Name: "Bolus",
        ApplicationType: "IntravenousBolus",
        DosingInterval: "Single",
      ▸ Parameters: [ … ]
    }
  ],
  ▾ Simulations: [
    ▾ {
        Name: "S1",
        Model: "4Comp",
        Solver: { },
      ▸ OutputSchema: [ … ],
        Population: "Healthy adults",
      ▾ Compounds: [
        ▾ {
            Name: "SuperDrug",
          ▾ CalculationMethods: [
              "Cellular partition coefficient method - PK-Sim Standard",
              "Cellular permeability - PK-Sim Standard"
            ],
          ▾ Protocol: {
              Name: "Bolus"
            }
          }
        ],
        HasResults: false
    }
  ],
```

What do our snapshots contain?



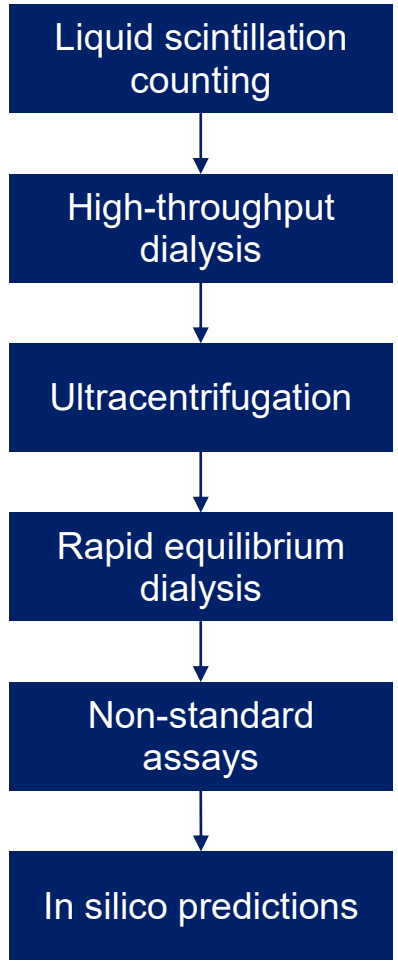


Getting *in vivo*, *in vitro* and *in silico* data from our corporate database

- *In vivo* concentration-time PK data are collected from 8 individual assays (various species in current and legacy systems)
- Physchem (pKa, solubility, lipophilicity) and ADME (plasma protein binding, intrinsic clearance, blood:plasma ratio, fu_{mic}) *in vitro* and *in silico* data are collected from 48 individual assays
- For each physchem and ADME property, a decision tree is implemented to return the "best" possible value
- The script can generate two flavors of the snapshot:
  - A compact version which only returns the "best" values
  - A complete version which return all available values, but the "best" value will be tagged accordingly

Experiment	Lipophilicity	Value Origin	Default
Lipophilicity determination in 1-octanol/water	3.21 Log Units	Database-In Vitro-Potentiometric titration, average of 3 experiments	☑
Direct logP/D in 1-octanol/buffer	2.80 Log Units	Database-In Vitro-Direct, from one experiment	☐
High-Throughput Octanol/Water Partition Coefficient	3.40 Log Units	Database-In Vitro-High-throughput, from one experiment	☐
NIBR logP and logD model (multi-task)	3.06 Log Units	Database-Other-In silico	☐

PPB decision tree



Generate PKSim Snapshot files

Arguments

Preferred Assays

You can select the preferred assays for the PKSim snapshot. If you leave it blank (None), all assays will be included, and best selected.

Cellular Partition Coefficient Method

CPC Method

Rodgers and Rowland

Species, for observered values

Species

Mouse × Rat × Dog × Monkey ×

Yes/No Options

- [x] **Add Human?** Adds human ADME data to the snapshot. Note that you still have to add human simulations under 'Default simulations' if desired.
- [ ] **Normalize Dose?** Pulls dose-normalized experimental C-t data and sets all simulations to 1mg/kg (other than simulations specified in Add simulations below)
- [x] **Use Legacy data?** Pulls data from three 'Legacy Assessment' assays and the Aragen Ct assay
- [ ] **Use Insilico data only?** Uses in silico options for all compound properties and doesn't pull observed data
- [ ] **Run NCA?** Calculates NCA clearance and other PK parameters from observed data (when available) and uses the observed NCA clearance to calculate hepatic clearance (other clearance option(s) will still be included in the snapshot)
- [x] **Substitute BQL?** Replaces BQL values with LLOQ/sqrt(2).

Simulation options

- [x] Human (1mpk IV 3mpk PO single dose)
- [ ] Mouse (1mpk IV 3mpk PO single dose)
- [ ] Rat (1mpk IV 3mpk PO single dose)
- [ ] Dog (0.1mpk IV 0.3mpk PO single dose)
- [ ] Monkey (0.1mpk IV 0.3mpk IV single dose)

NVP codes

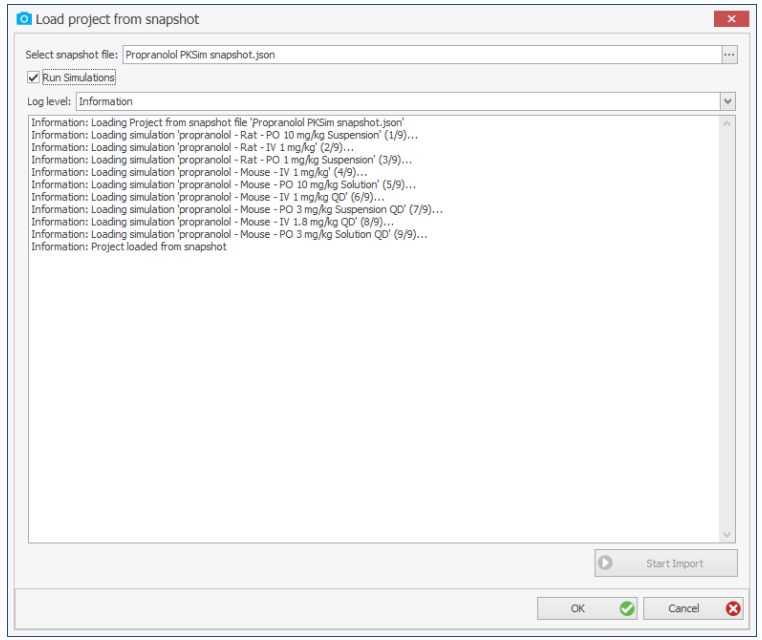
NVP Codes

propranolol

You can enter multiple NVP codes, separated by commas.

Example: NVP-XYZ123, NVP-XYZ123, ...

Submit NVP Codes



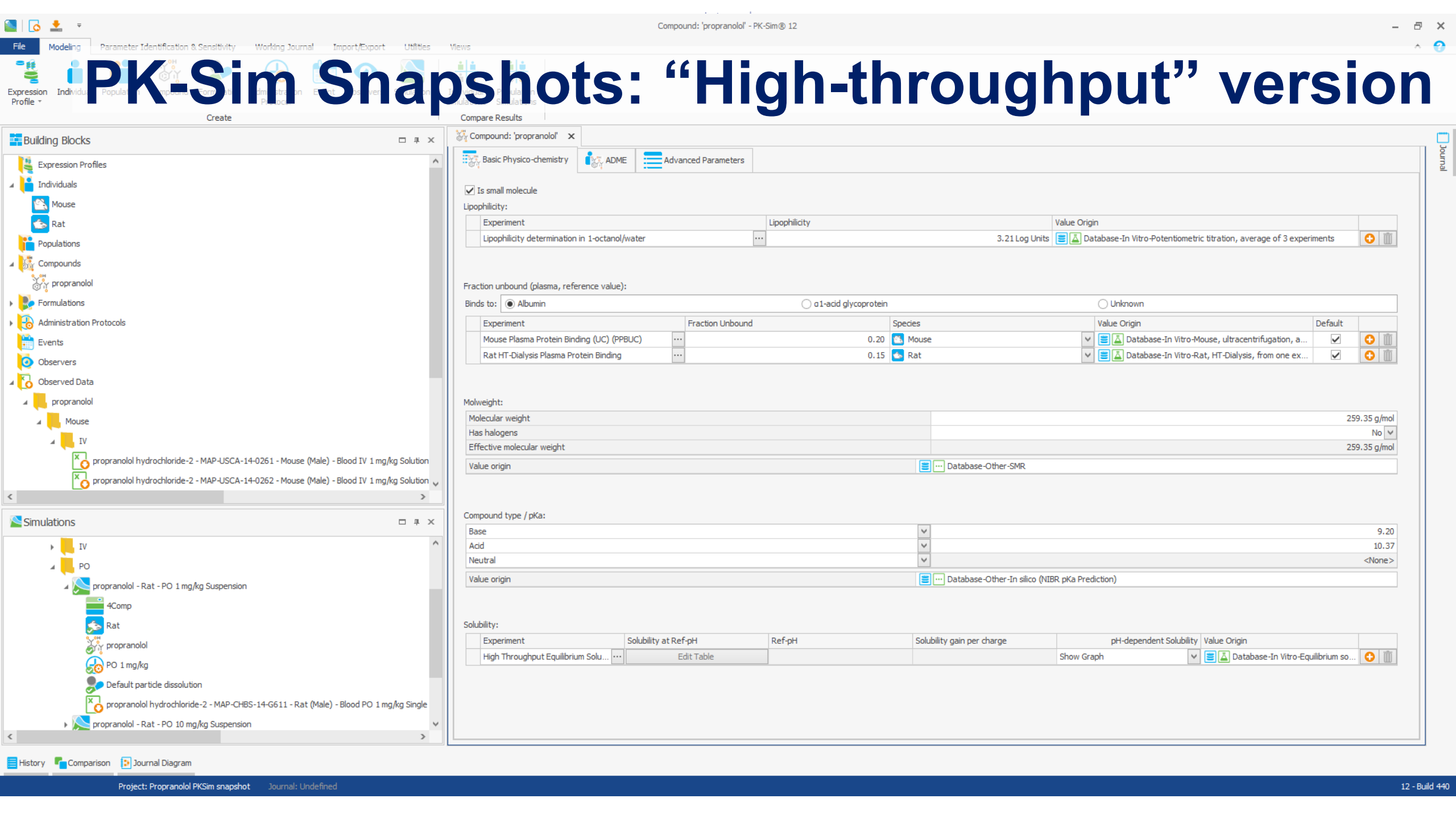
PKSim snapshot generated successfully!

PKSim Snapshot

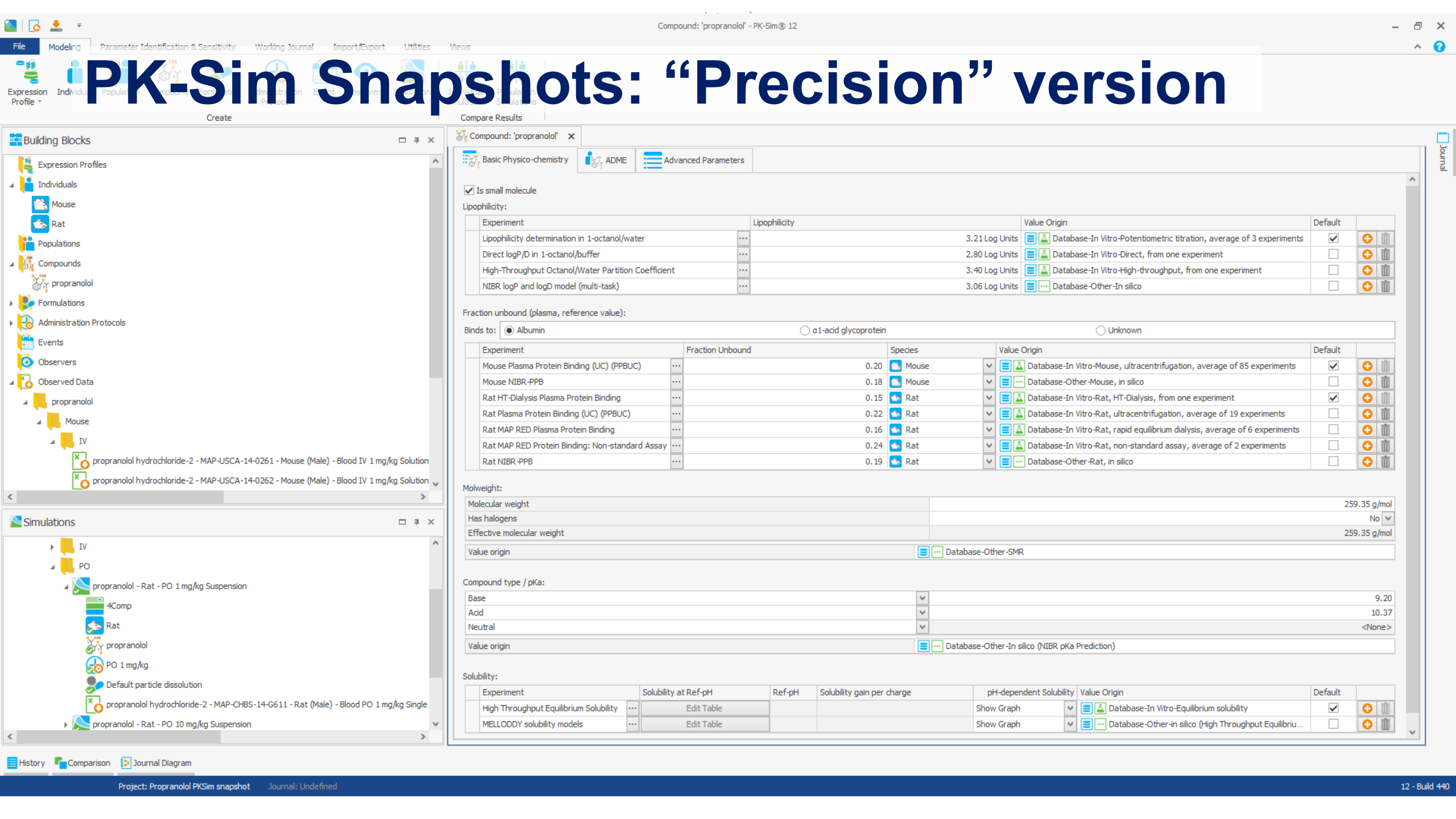
{"Version":79,"Description":"PKSim snapshot created for propranolol from CDF on 2025-09-23 10:49:48\r\nScript version 1.5.0","Individuals":[{"Name":"Rat","OriginData":{"CalculationMethods":["SurfaceAreaPlsInt_VAR1"],"Species":"Rat","Weight":{"Value":0.312,"Unit":"kg"},"ValueOrigin":{"Source":"Database","Method":"InVivo","Description":"Average of 2 animals from 1 study"}},"ExpressionProfiles":[]},{"Name":"Mouse","OriginData":{"CalculationMethods":["SurfaceAreaPlsInt_VAR1"],"Species":"Mouse"},"ExpressionProfiles":[]},{"Name":"Human","OriginData":{"CalculationMethods":["SurfaceAreaPlsInt_VAR1","Body surface area - Mosteller"],"Species":"Human","Population":"European_ICRP_2002","Gender":"MALE","Age":{"Value":30.0,"Unit":"year(s)"},"Weight":{"Value":70.0,"Unit":"kg"}},"ExpressionProfiles":[]}],"Compounds":[{"Name":"propranolol","IsSmallMolecule":true,"PlasmaProteinBindingPartner":"Albumin","Lipophilicity":[{"Name":"Lipophilicity determination in 1-octanol/water","Parameters":[{"Name":"Lipophilicity","Value":3.21,"Unit":"Log Units","ValueOrigin":{"Source":"Database","Method":"InVitro","Description":"Potentiometric titration, average of 3 experiments"}}]},{"Name":"Direct logP/D in 1-octanol/buffer","Parameters":[{"Name":"Lipophilicity","Value":2.8,"Unit":"Log Units","ValueOrigin":{"Source":"Database","Method":"InVitro","Description":"Direct, from one experiment"}}],"IsDefault":false},{"Name":"High-Throughput Octanol/Water Partition Coefficient","Parameters":[{"Name":"Lipophilicity","Value":3.4,"Unit":"Log Units","ValueOrigin":{"Source":"Database","Method":"InVitro","Description":"High-throughput, from one experiment"}}],"IsDefault":false},{"Name":"NIBR logP and logD model (multi-task)","Parameters":[{"Name":"Lipophilicity","Value":3.06,"Unit":"Log Units","ValueOrigin":{"Source":"Database","Method":"Other","Description":"In silico"}}],"IsDefault":false}],"FractionUnbound":[{"Name":"Mouse Plasma Protein Binding (UC) (PPBUC)","Species":"Mouse","Parameters":[{"Name":"Fraction unbound (plasma, reference value)","Value":0.1989,"ValueOrigin":{"Source":"Database","Method":"InVitro","Description":"Mouse, ultracentrifugation, average of 85 experiments"}}]},{"Name":"Mouse NIBR-PPB","Species":"Mouse","Parameters":[{"Name":"Fraction unbound (plasma, reference value)","Value":0.1735,"ValueOrigin":{"Source":"Database","Method":"Other","Description":"Mouse, in silico"}}],"IsDefault":false},{"Name":"Rat HT-Dialysis Plasma Protein Binding","Species":"Rat","Parameters":[{"Name":"Fraction unbound (plasma, reference value)","Value":0.153,"ValueOrigin":{"Source":"Database","Method":"InVitro","Description":"Rat, HT-Dialysis, from one experiment"}}]},{"Name":"Rat Plasma Protein Binding (UC) (PPBUC)","Species":"Rat","Parameters":[{"Name":"Fraction unbound (plasma, reference value)","Value":0.2201,"ValueOrigin":{"Source":"Database","Method":"InVitro","Description":"Rat,

Download PKSim Snapshot

PK-Sim Snapshots: “High-throughput” version



PK-Sim Snapshots: “Precision” version

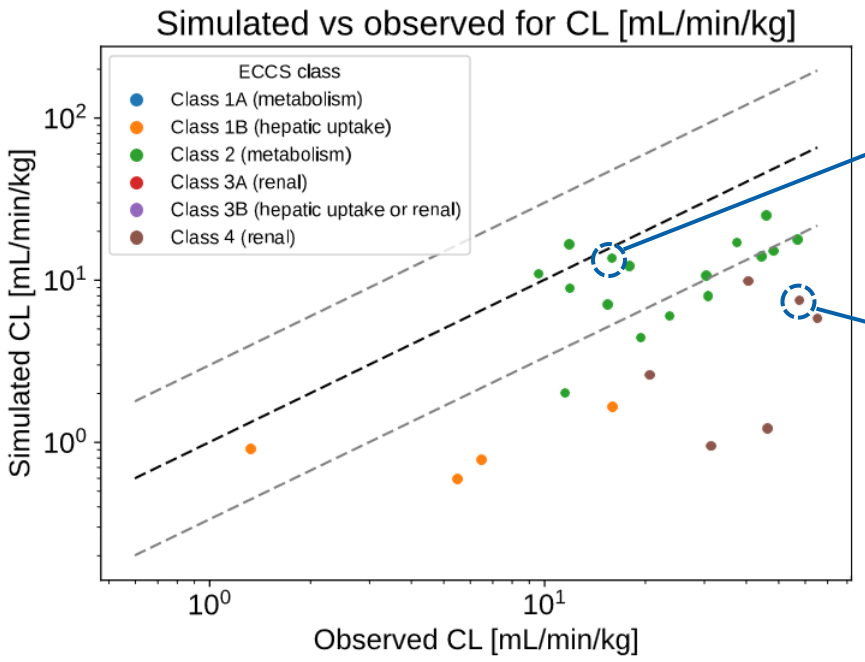


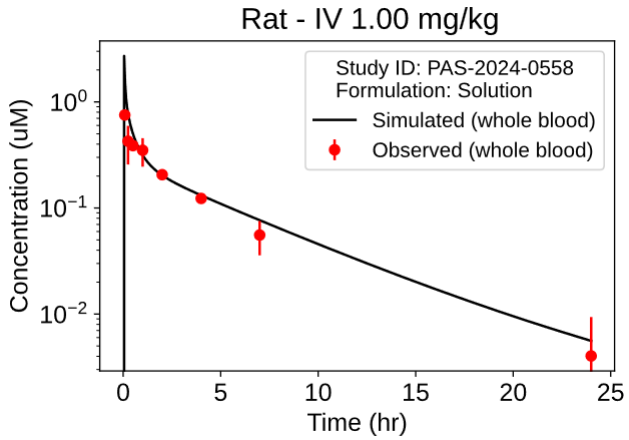
Data and application

NOVARTIS

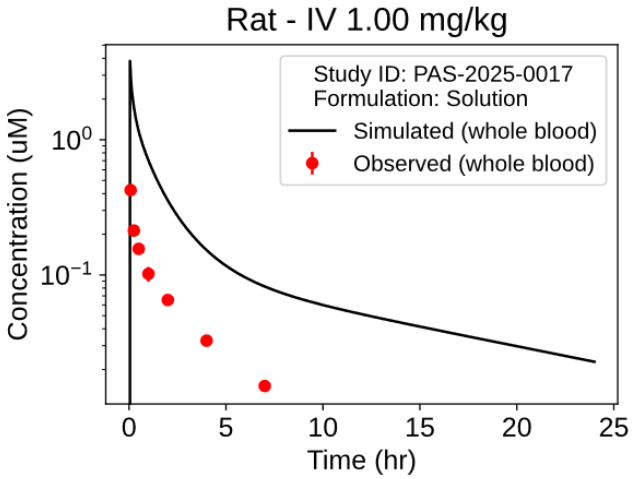
Small/medium-scale use cases

On a project basis, our workflow allows to perform quick checks of whether compounds are “well-behaved”





Can be predicted well by bottom-up Vss & scaled LM CLint

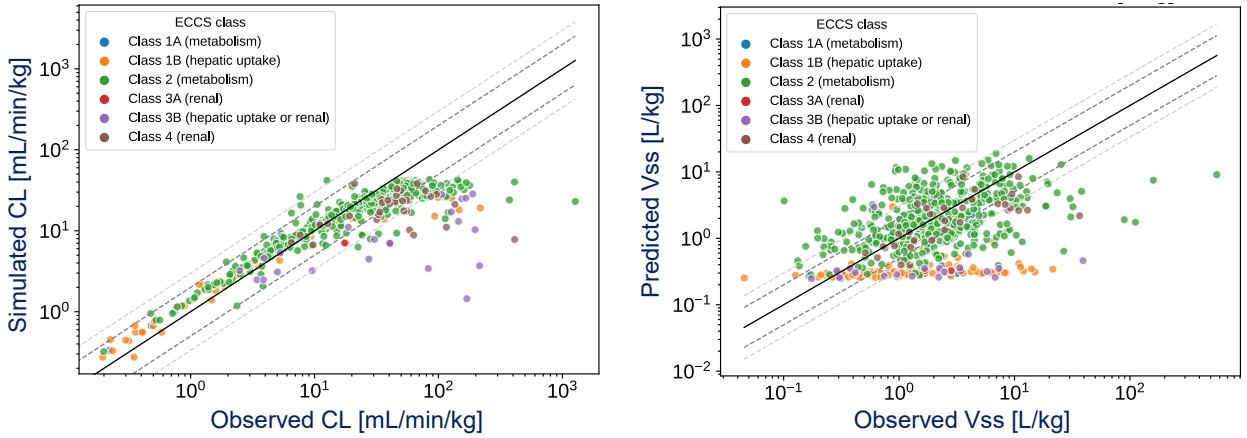


Requires further investigation in PK-Sim GUI (alternative CL & Vss mechanisms, transporters)

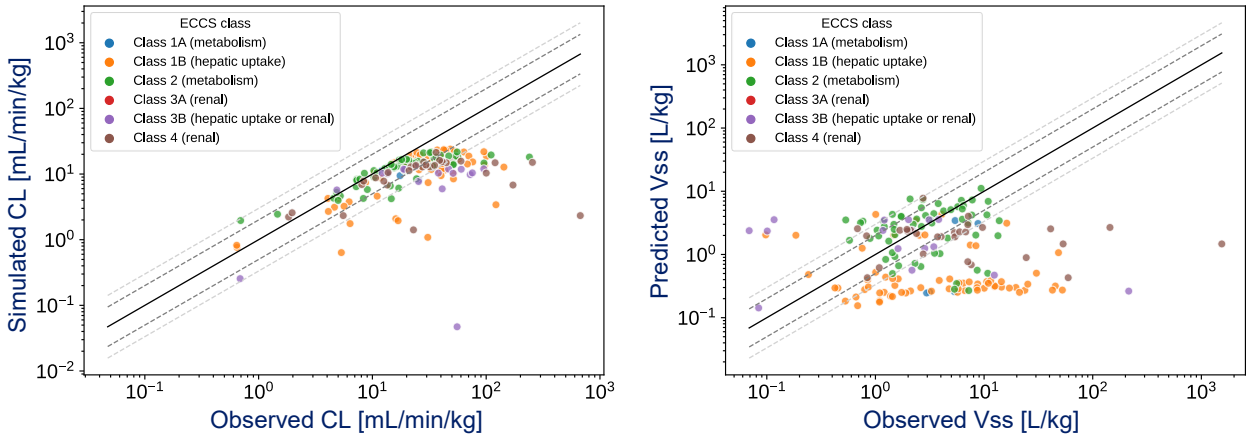
Large-scale use case

- Bottom-up PBPK simulations during early discovery (pre-synthesis) can be done using *in silico* predicted PhysChem and ADME properties
- We performed a large-scale assessments across our portfolio:
  how well can we assess Vss across compounds and species when we have IV data?
- Vss was well predicted for ECCS class 2 compounds, but class 3A/B were under-predicted

For rat IV studies: we surveyed ~575 compounds



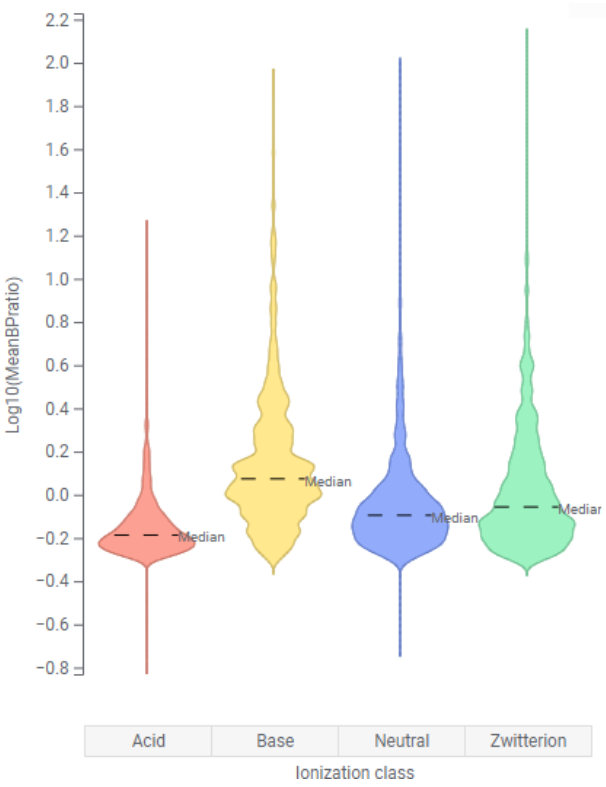
For monkey IV studies: we surveyed ~150 compounds



PhysChem and ADME in silico models as input parameters for PK-Sim

The simulations require these input parameters:

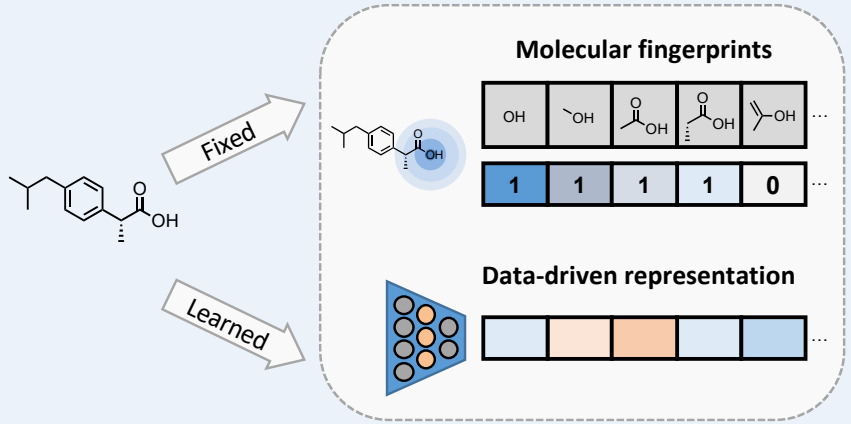
- Lipophilicity → multi-task (LogP / LogD) LASSO, RDKit 2D descriptors + Morgan FP (r=2, 4k bits)
- fu,p → multi-task (5 species + 7 auxiliary tasks) GNN (chemprop)
- Solubility → MELLODDY multi-task FF-NN (SparseChem), ECFP6, 32k bits
- Passive permeability → multi-task (5 end-points: MDCK-LE, PAMPA, Caco-2, MDR1) GNN (chemprop)
- Ionization → Molecular Discovery's MoKa, trained on in-house data
- Blood/plasma ratio → no models (yet!), using instead a default value per ionization class, based on an analysis of our in-house data
- Intrinsic clearance → multi-task (6 species) GNN (chemprop)



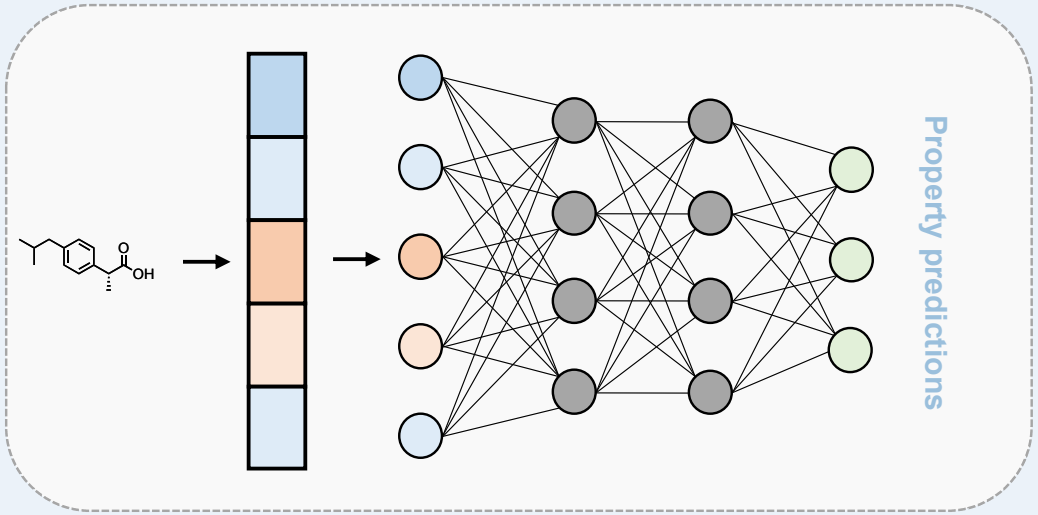
Machine learning models to predict PhysChem/ADME/PK properties

- **Machine learning (ML) models can relate molecular structures to properties**
  → Quantitative or qualitative structure-property relationship (Q)SPR

1. Molecular features to represent them numerically, e.g. descriptors, fingerprints, graph neural networks



2. ML algorithm that finds structural patterns that correlate with ADME properties



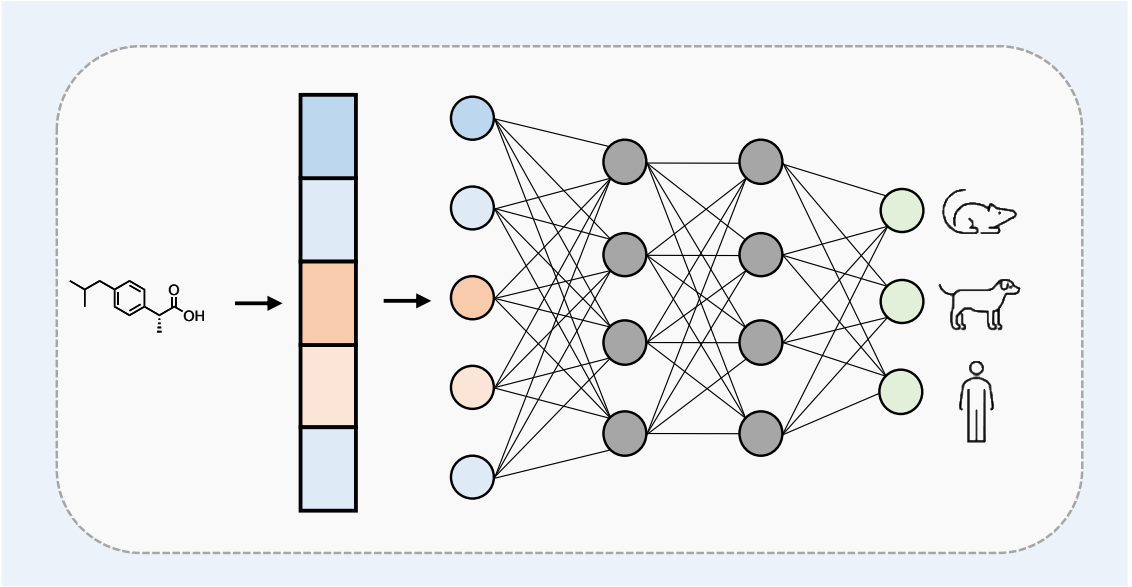
Predicting metabolic clearance

- Multi-task learning strategy where each species was considered as a separate task[1]
- A graph neural network (GNN) was utilized for representation learning[2]
- A multi-species GNN model was generated to predict metabolic clearance[1]

~ # Compounds

Rat	Human	Mouse	Dog	Monkey	Minipig
180,000	112,000	85,000	10,000	8,000	700

Example: Multi-species metabolic clearance model (Multi-task graph neural network, MT-GNN)



[1] Rodriguez-Perez et al. 2022 Mol Pharm, 20, 1, 383-394
[2] Yang et al. 2019 JCIM, 59, 3370-3388

Predicting metabolic clearance

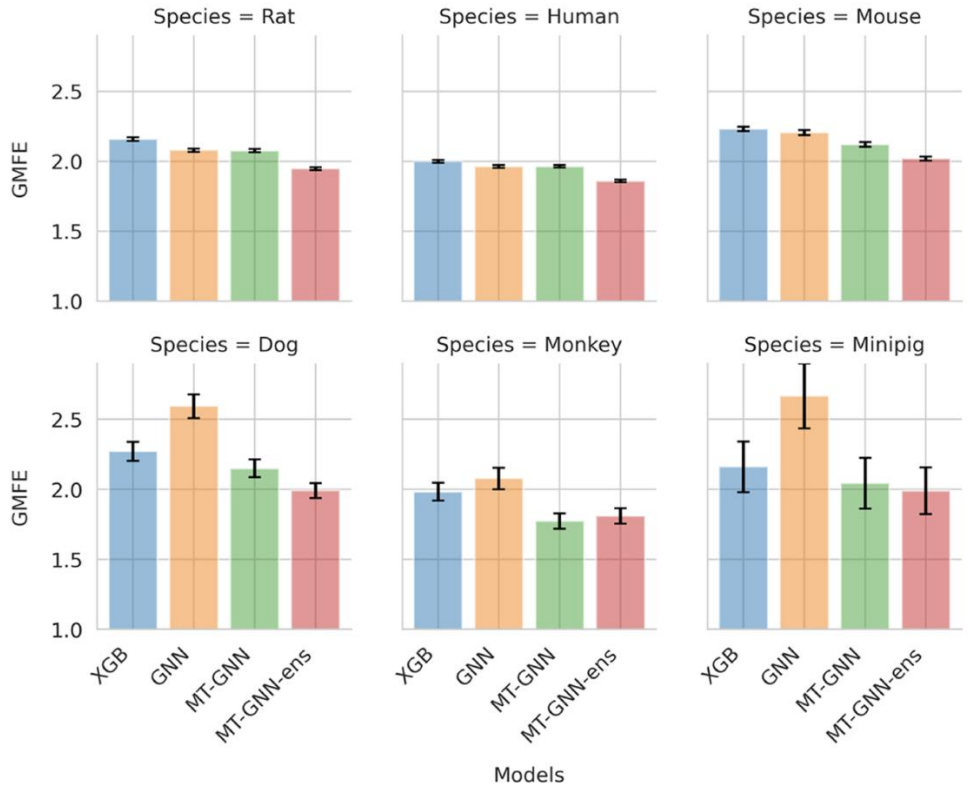
- **Multi-task learning strategy where each species was considered as a separate task[1]**
- **A graph neural network (GNN) was utilized for representation learning[2]**
- **A multi-species GNN model was generated to predict metabolic clearance[1]**

~ # Compounds

Rat	Human	Mouse	Dog	Monkey	Minipig
180,000	112,000	85,000	10,000	8,000	700

[1] Rodriguez-Perez et al. 2022 Mol Pharm, 20, 1, 383-394
[2] Yang et al. 2019 JCIM, 59, 3370-3388

Example: Multi-species metabolic clearance model (Multi-task graph neural network, MT-GNN)



Especially for species with less data points, e.g. dog and monkey, results were superior with MT-GNN

Metabolic clearance:
Model precision vs. Experimental reproducibility

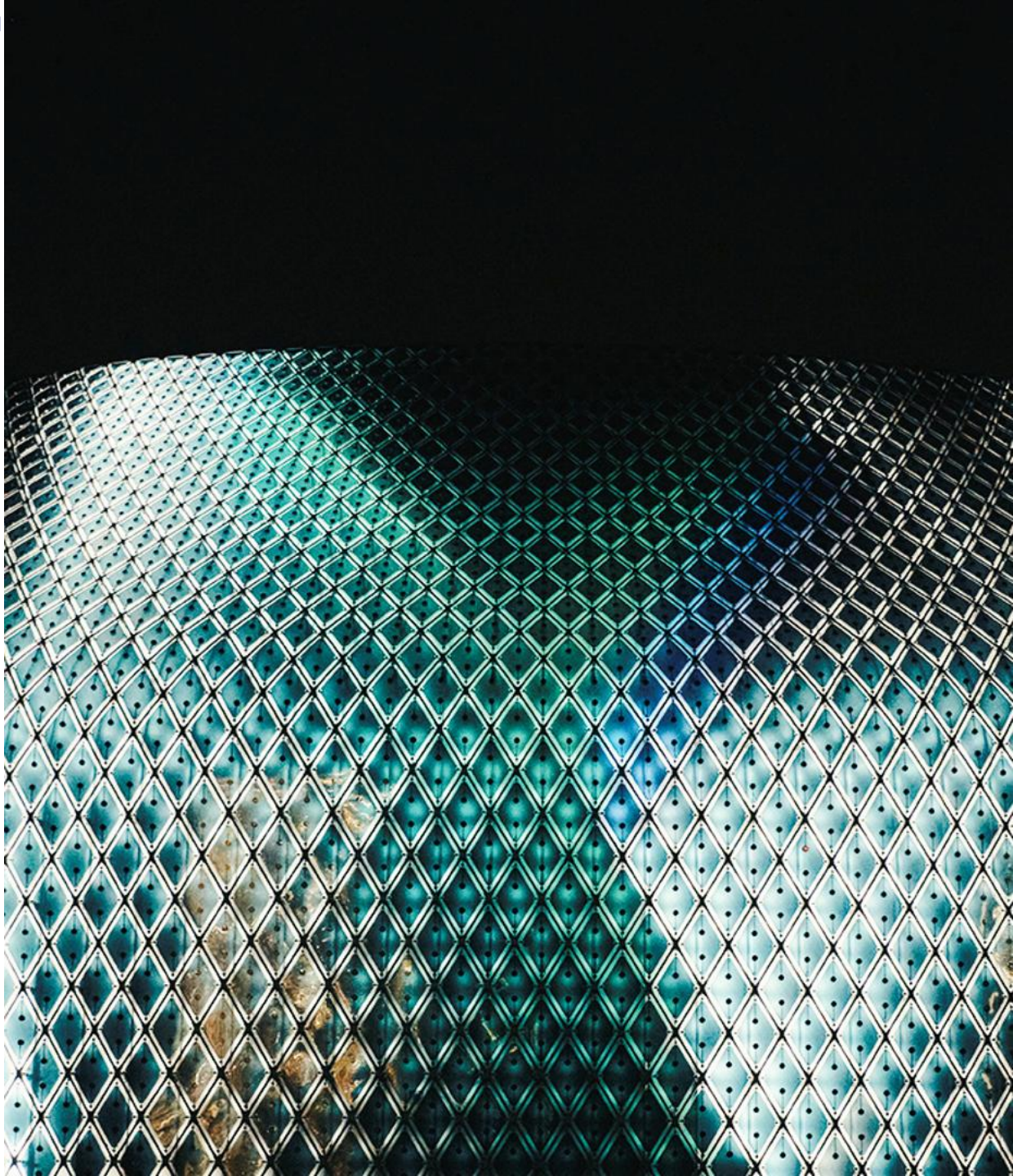
- **Model precision: From the compounds predicted "Low", how many are measured "Low"?**
- **Experimental reproducibility: From the ones measured "Low", how many were always measured "Low"?**

- **With the limitation of having some 'inconclusive' predictions (25-32%), model's precision is comparable to experimental reproducibility**

	Low CL_{int} (≤ 100µL/min/mg)		High CL_{int} (≥ 300µL/min/mg)	
Species	**Model precision**	**Experim. reproduc.**	**Model precision**	**Experim. reproduc.**
Rat	**84**	83	**75**	69
Human	**85**	88	**75**	64
Mouse	**80**	84	**78**	72
Dog	**92**	89	**97**	58
Monkey	**94**	87	**86**	71

Rodriguez-Perez et al. 2022 Mol Pharm, 20, 1, 383-394

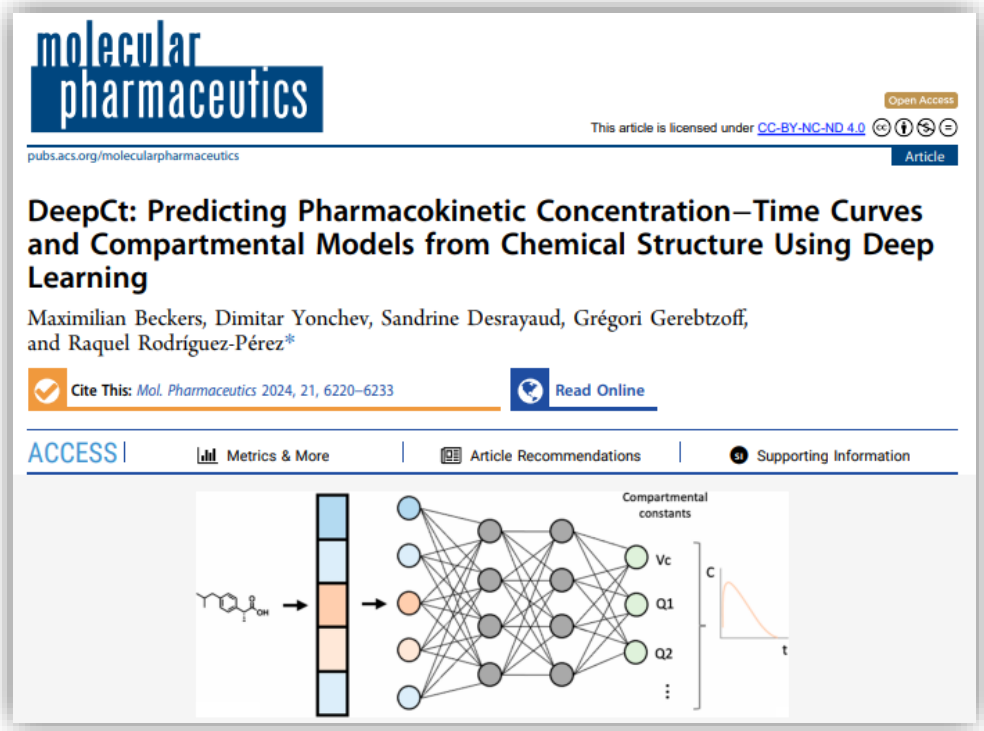
Outlook



NOVARTIS

What is coming next?

- Still some plumbing to do:
  - Storing of simulations
  - Centralizing the decision trees into a snowflake-based process, which will improve the performance (currently: ~3.8s/compound for data pre-processing and ~2.5s/compound to run a single species, single dose, single arm simulation)
- Automating parameter identification
- Learning from the data! E.g.
  - Which method works best for which type of compounds (depending on ECCS class, ionization, permeability, modality, …)
  - Investigating protein binding, transporters, clearance…
  - Assessing performance of unbound parameters
- Democratizing the process
  - Onboarding more projects
  - Enabling simulations in our design & analytics tools



Filters

Compound Id

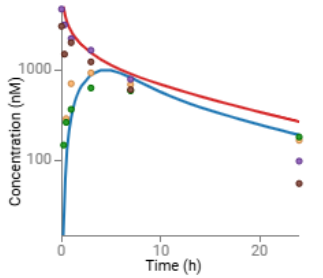
```
clonazepam
etoposide
gliquidone
tofacitinib
trametinib
vismodegib
```

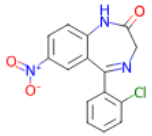
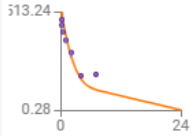
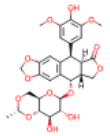
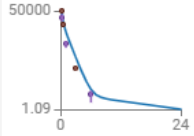
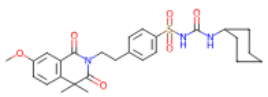
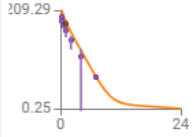
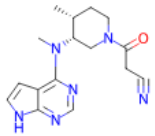
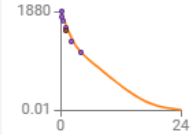
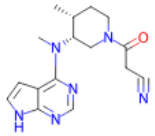
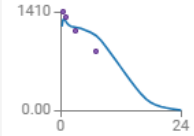
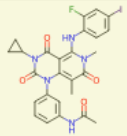
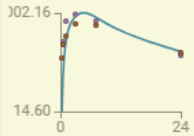
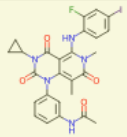
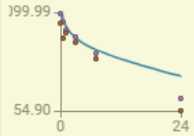
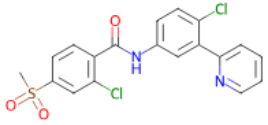
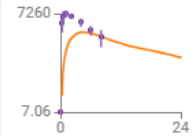
Dose

Route

- i.p. 0
- i.t. 0
- i.v. 7
- i.v. infusion 0
- o.p. 0
- other 0
- p.o. 3
- s.c. 0
- topical 0

MultiCurveViewer



	Common Name	Structure	HT-PBPK	Species	Matrix	Route	Vehicle	Dose	In vivo CL	Vss	Cmax	AUCinf	AUClast	%F
1	clonazepam			Rat Legacy	Plasma	i.v.	20%PG, 50(20%)Solutol, WFI	1.00	71.00	3.10		766.00	759.00	
2	etoposide			Mouse	Plasma	i.v.	PBS	10.00	8.97	0.16		31700.00	31700.00	
3	gliquidone			Rat	Blood	i.v.	NMP:PEG 200 (10:90)	0.30	6.42	0.38		1610.00	1600.00	
4	tofacitinib			Rat	Blood	i.v.	NMP:PEG200 (30:70)	1.00	53.00	1.70		1004.00	987.00	
5	tofacitinib			Mouse	Plasma	p.o.	0.5% MC/0.1% Tween	10.00			1409.00	1985.00		19.00
6	trametinib			Mouse	Plasma	p.o.	0.5% MC/0.5% Tween; suspension	10.00			635.00	13178.00		18.00
7	trametinib			Mouse	Plasma	i.v.	PEG300/D5W, 3:1; solution	2.00	3.63	1.19	3085.00	14938.00		18.00
8	vismodegib			Rat Legacy	Plasma	p.o.	10% NMP, 15% PEG300, 5 Solutol and WFI	3.00			7399.23	39911.27	32466.09	100.00

Reimagining medicine, together.

Novartis:
- Jan Schlender
- Sarah McFann
- Gloria Ha
- Churni Gupta
- Jimmy Kromann
- Peter Ashcroft
- Raquel Rodríguez-Pérez

ESQLabs:
- Stephan Schaller
- Pavel Balazki
- Laura Villain
- Diane Lefaudeux
- Susana Proença
- Mariana Guimaraes Sa Correia