OSP Community Conference 2026

September 24 - 25, 2026; Johnson & Johnson Campus, Beerse, Belgium

”Combining machine learning and high throughput physiology-based pharmacokinetic modeling”

Eren Alkanat –
Goethe University, Frankfurt, Germany, PMx-R Comp. Bio AI/ML, Sanofi, Frankfurt, Germany

Supervisors:
Dr. Denise Feick & Dr. Henrik Cordes –
Pharmacometrics-Research (PMx-R) Computational Biology AI/ML, Sanofi, Frankfurt, Germany

From virtual compounds to early mechanistic PK

Goal
- Enable PK modeling in early drug discovery with clinical-phase mechanistic resolution

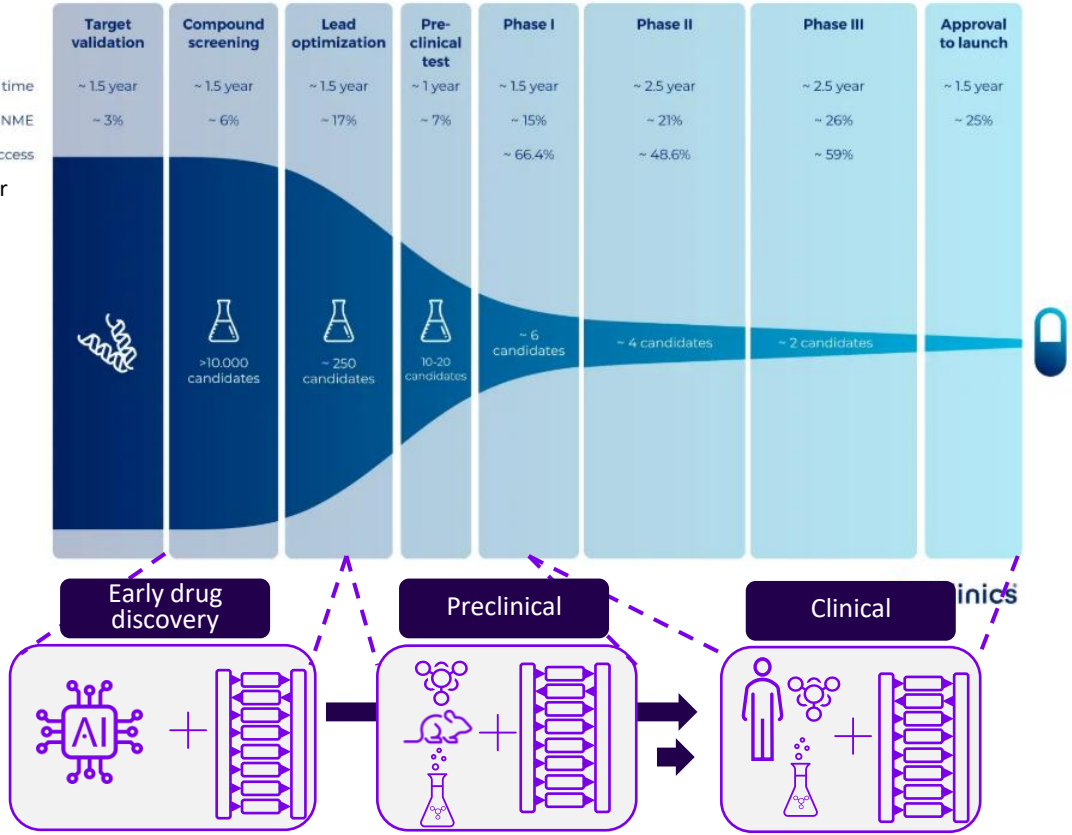
Clinical
- Data rich environment
- High quality and large volume on human *in vitro* & *in vivo* data to cover ADME processes mechanistically

Preclinical
- Data moderate environment
- Mainly animal, with limited *in vitro* human measurements available
- Estimation/scaling of PK behavior to human from animal & *in vitro* needed

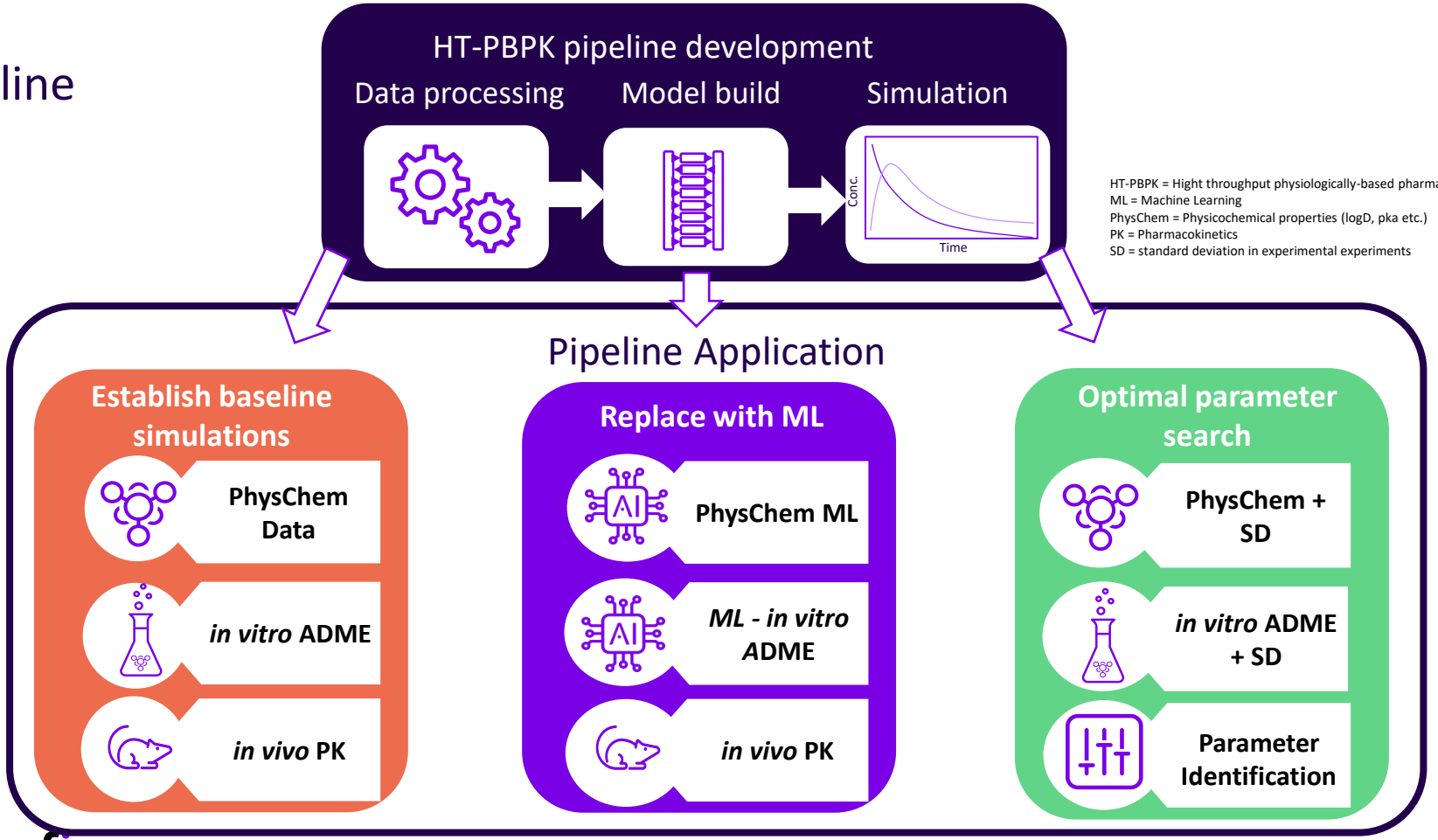
Early Drug discovery
- Scarce & heterogenous data environment
- Early in vitro screening assays
- Predicted PhysChem, *in vitro* & *in vivo* properties
- Large number of compounds

Prerequisites
- Data processing & mapping to PK-Sim processes
- Seamless integration of experimental & predicted input values
- Automated model building process (high throughput-PBPK)

Source: ZeClinics

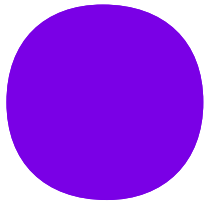


Outline



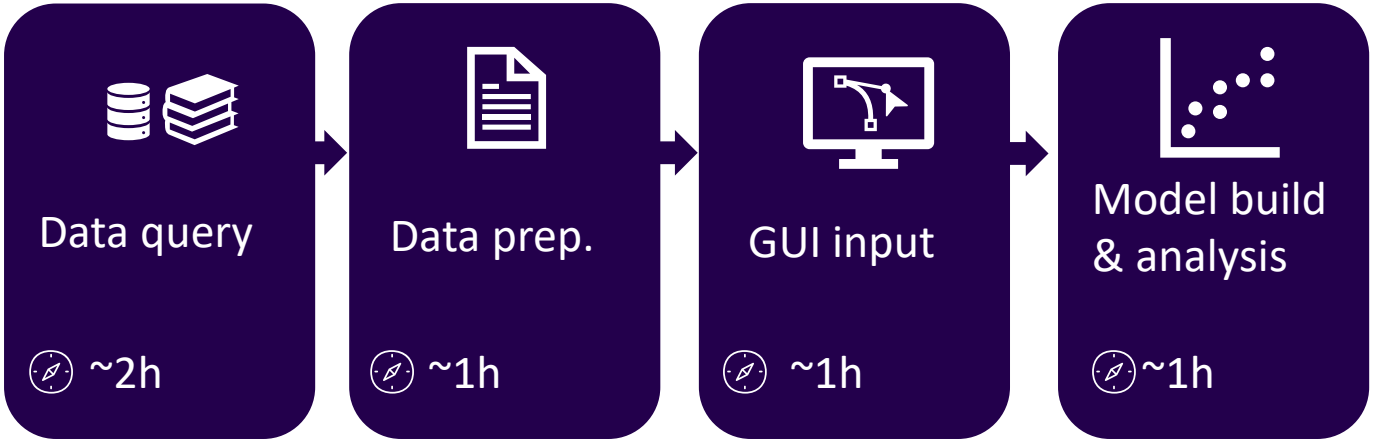
HT-PBPK = Hight throughput physiologically-based pharmacokinetic
ML = Machine Learning
PhysChem = Physicochemical properties (logD, pka etc.)
PK = Pharmacokinetics
SD = standard deviation in experimental experiments

HT-PBPK pipeline development



Traditional PBPK model building does not scale

Manual, sequential and repetitive tasks for every compound



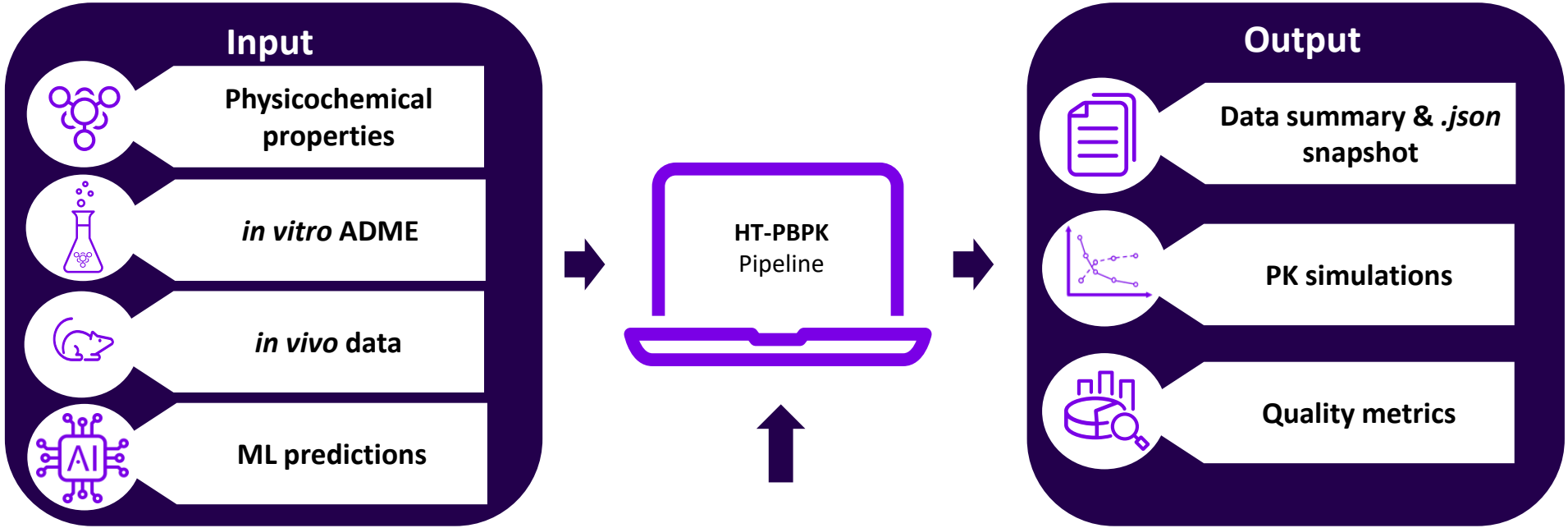
~5h per compound
Time spend scales linear with number of compounds

Burden:

- Repetitive tasks per compound
- Manual GUI-based parameter input
- Manual simulation one at a time
- Error prone

Disclaimer: Simplified workflow

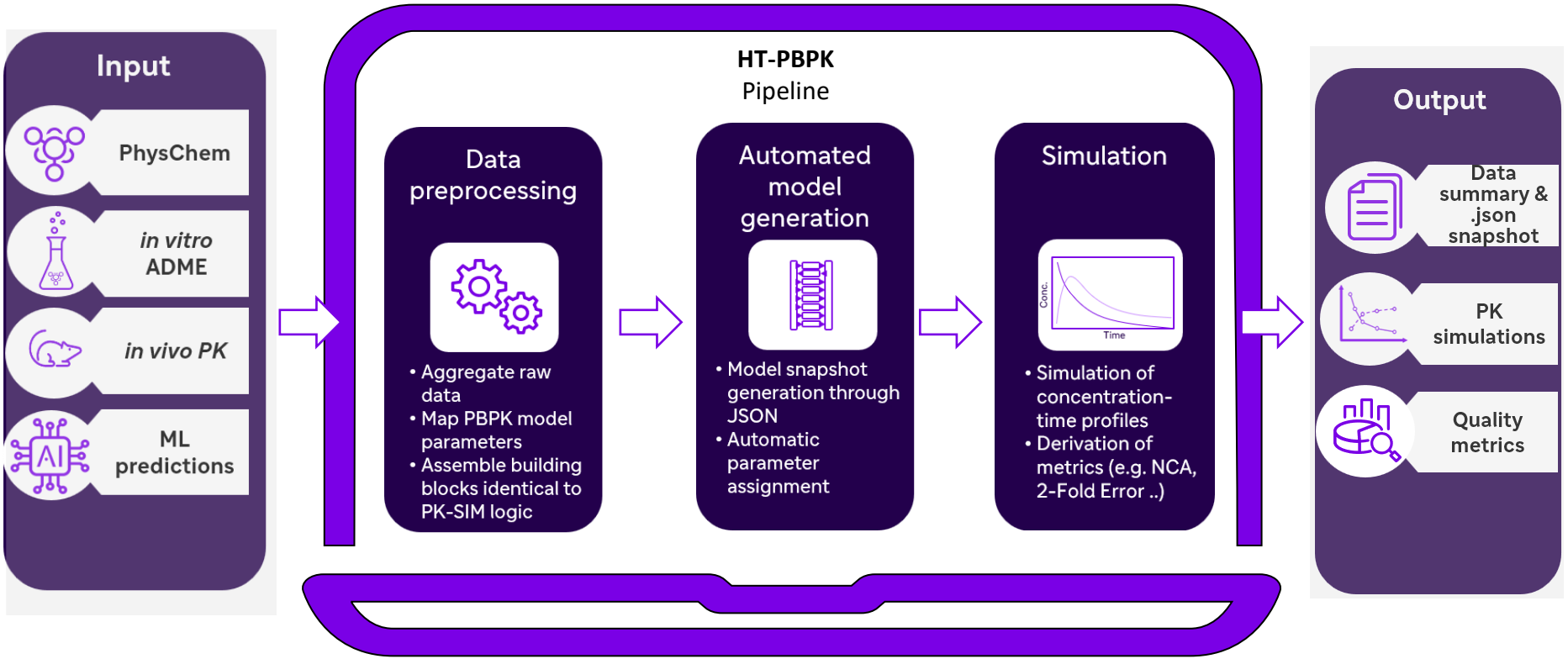
Scalable PBPK models building with HT-PBPK



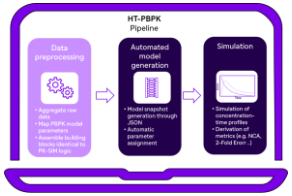
Key concepts:

- **Handling missing values**
- **Parameter heuristic**
- **Parameter & process mapping**

From data to model – fully automated



Data preprocessing



List of assays

- PhysChem properties
- Plasma binding
- Drug metabolism
- Drug permeability
- Enzyme & transporter DDI

Extract

PK-Sim Parameters

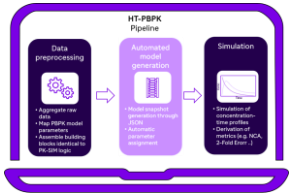
- Lipophilicity, solubility, pKa
- Fraction unbound
- Microsomal & Hepatocyte stability
- Caco-2 permeability

Map

PK-Sim Building Block

- Physico-chemistry
- ADME, total hepatic clearance
- Specific intestinal permeability

Source: ADME PK | Cyprotex ADME-Tox Solutions | Evotec



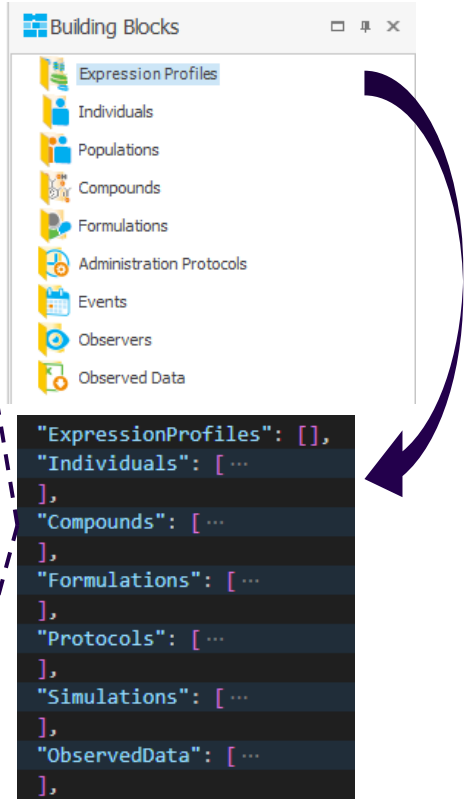
Model snapshot generation through JSON

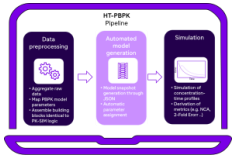
Technical Implementation: utilizing strength of JSON snapshot

- human-readable text files in JSON format
- Contain minimal information required to create a PKSim project & pkml files
- Populated with substance-specific parameters (molecular weight, lipophilicity, etc.) and PK experiments

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







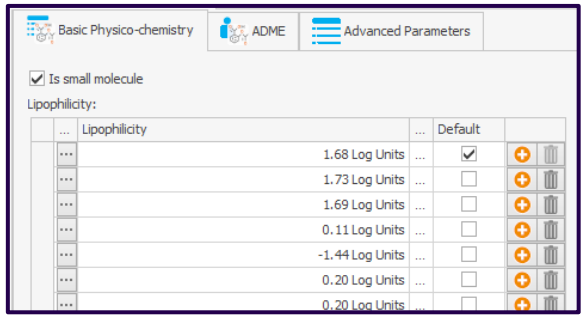
Automatic parameter assignment: Compound building block

Writing data into building blocks

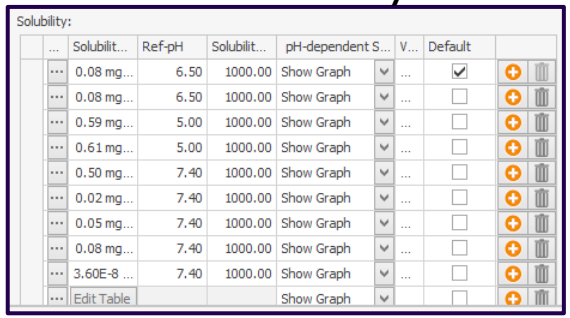
Model .json

```
"Compounds": [
  {
    "Name": "test Compound",
    "IsSmallMolecule": true,
    "Lipophilicity": [ ···
    ],
    "FractionUnbound": [ ···
    ],
    "Solubility": [ ···
    ],
    "IntestinalPermeability": [ ···
    ],
    "PkaTypes": [ ···
    ],
    "Processes": [ ···
    ],
    "CalculationMethods": [ ···
    ],
    "Parameters": [ ···
    ],
    "Description": "Smiles Code test"
  }
]
```

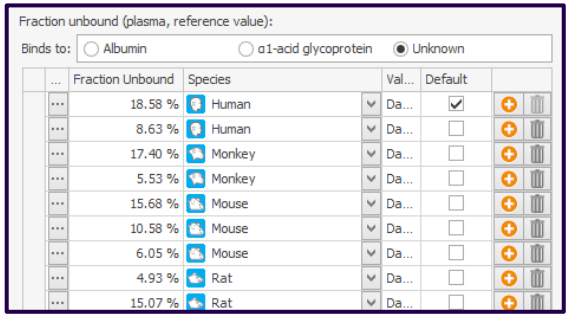
Lipophilicity



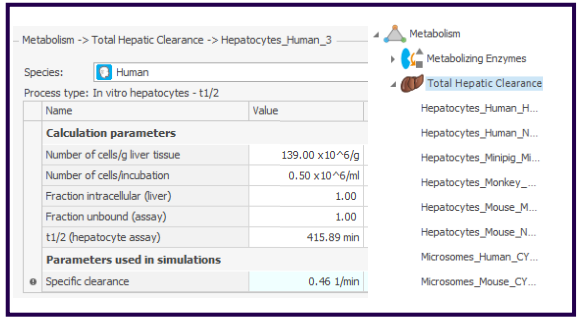
Solubility

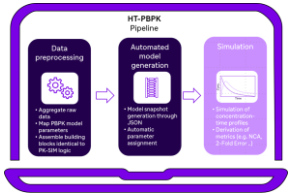


Plasma protein binding

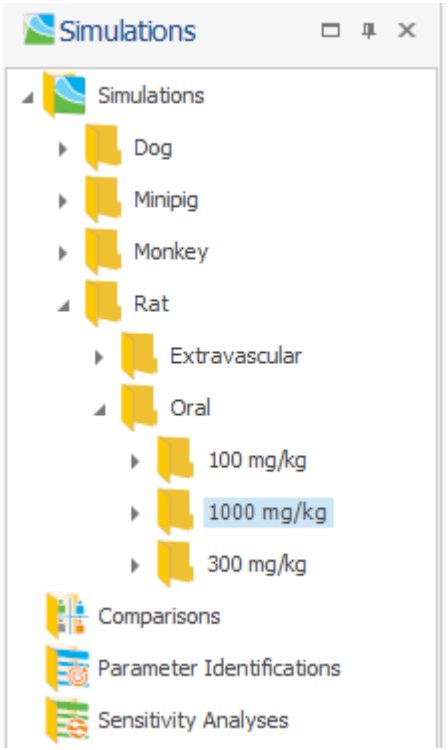


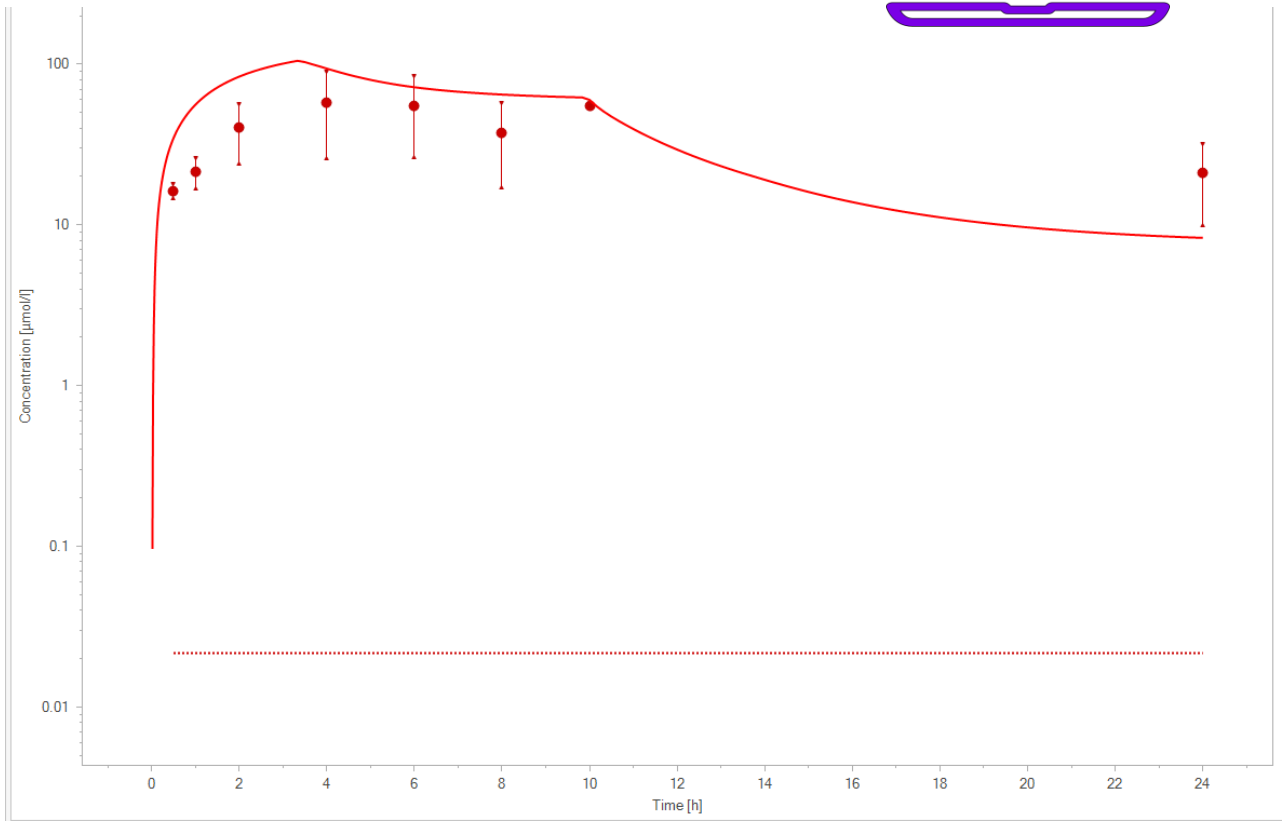
Hepatic clearance





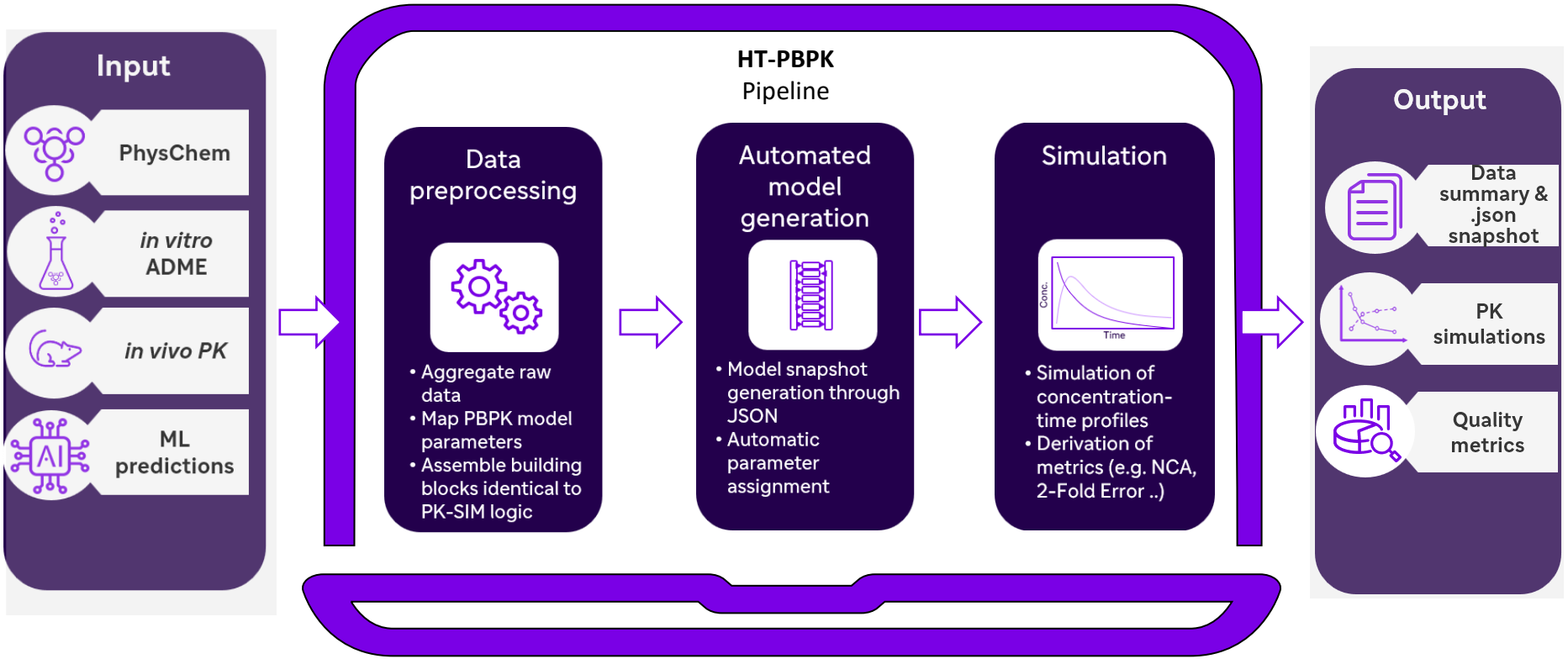
Simulations with default parameters



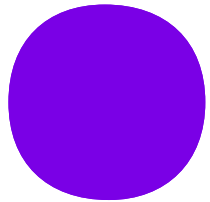


~from hours to mins per compound

From data to model – fully automated



HT-PBPK pipeline application



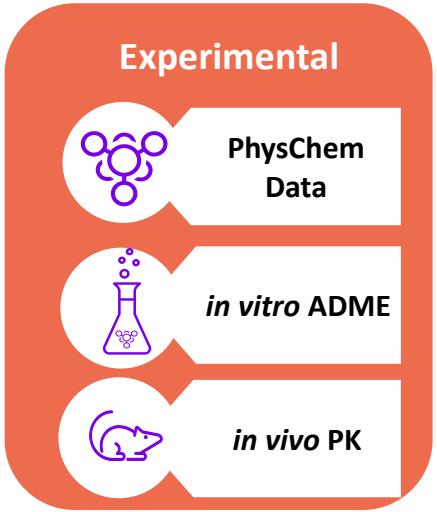
Establishing HT-PBPK prediction baseline

Can in vitro inputs reliably predict in vivo outputs?

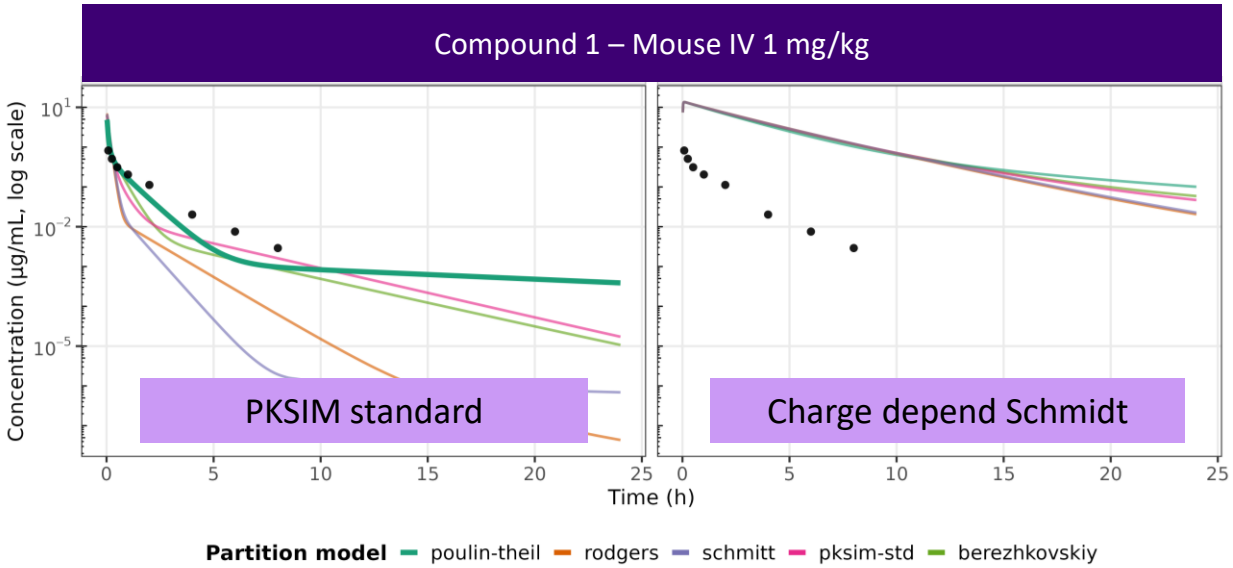
- in vitro measurements:
  - Lipophilicity (logD)
  - Plasma protein binding
  - Microsomal half-life

- in vivo observation:
  - Species: mouse, rat, rabbit, dog, minipig, monkey
  - Average IV concentration-time profiles

PhysChem = Physicochemical properties (logD, pka etc.)
PK = Pharmacokinetics
IV = Intravenous



Identification of optimal distribution model

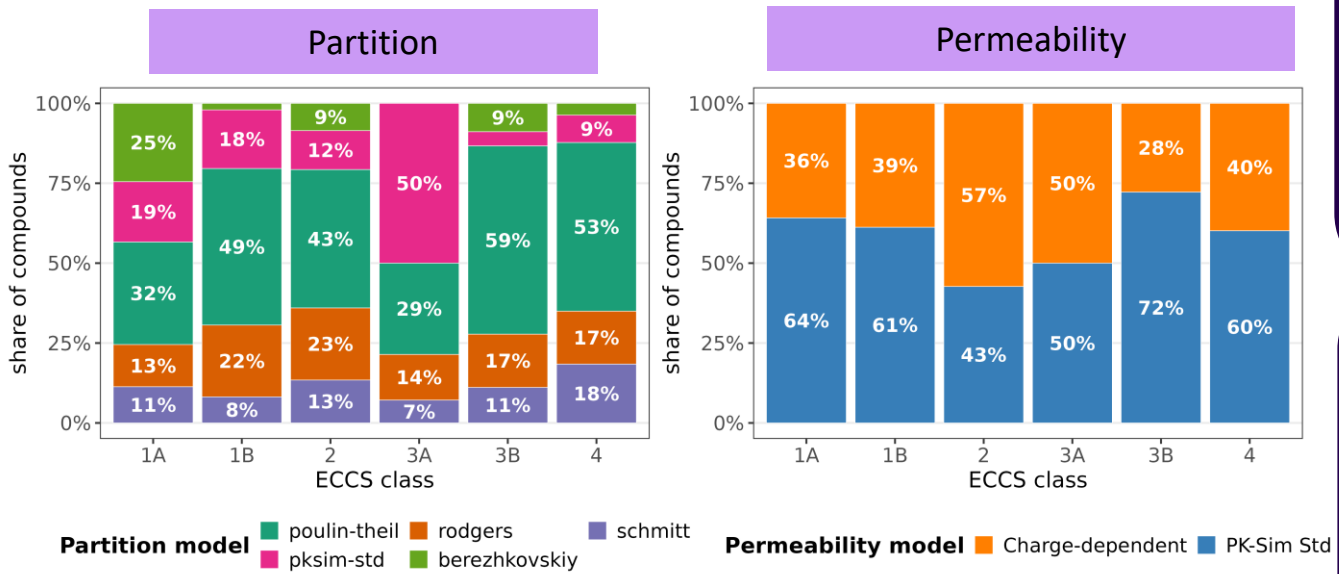


Summary

Best permeability & partitioning combination describing experimental data

Extended Clearance Classification System (ECCS)

From compound-level screening to class-level model-selection guidance

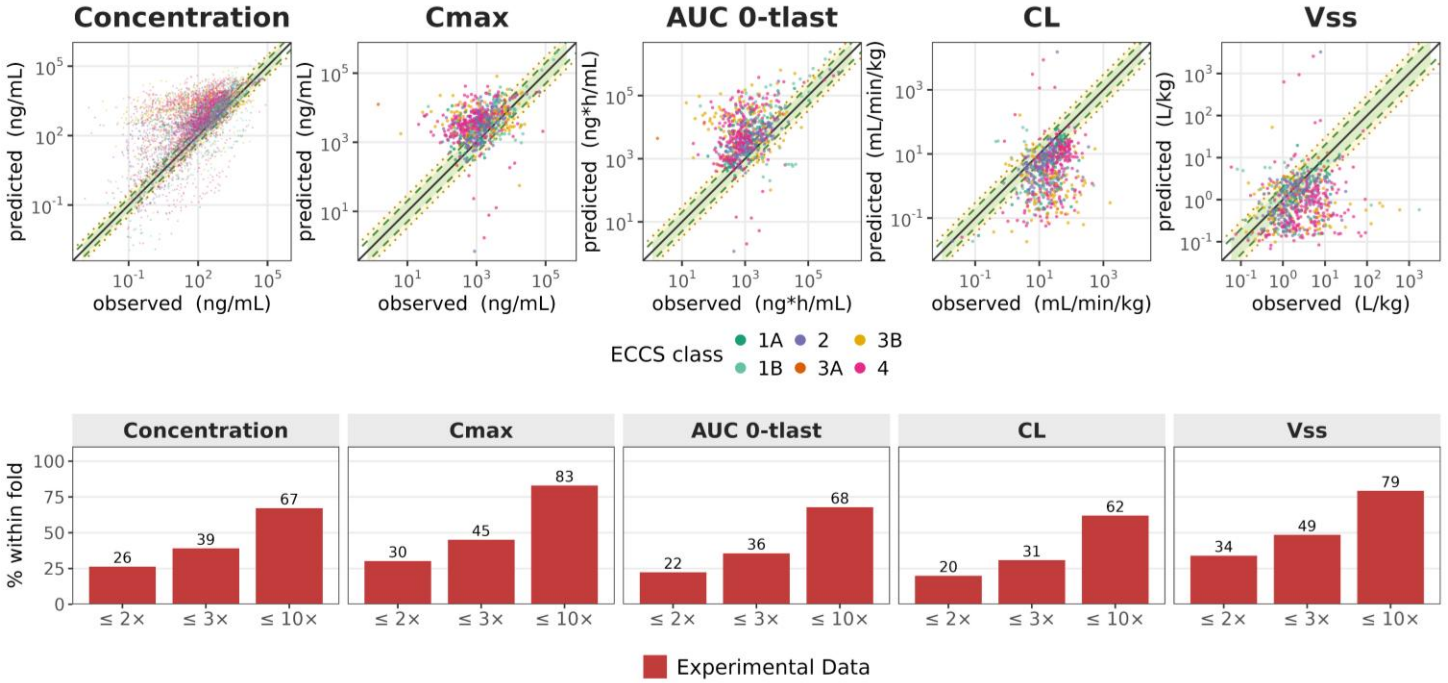


Goal:
Move from individual best fits toward generalizable model-selection rules

Summary:

No “one fits all” individual model preference for ECCS classes

Exposure predictions with experimental input data



Summary | Limited performance 30-50% within 3-fold error

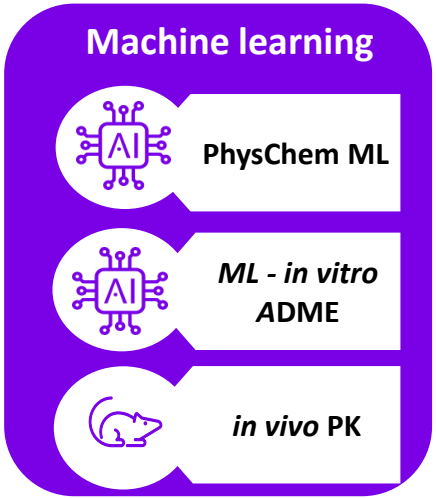
Exchanging experimental data with machine learning predictions

Can we achieve similar prediction performance with machine learning input parameters ?

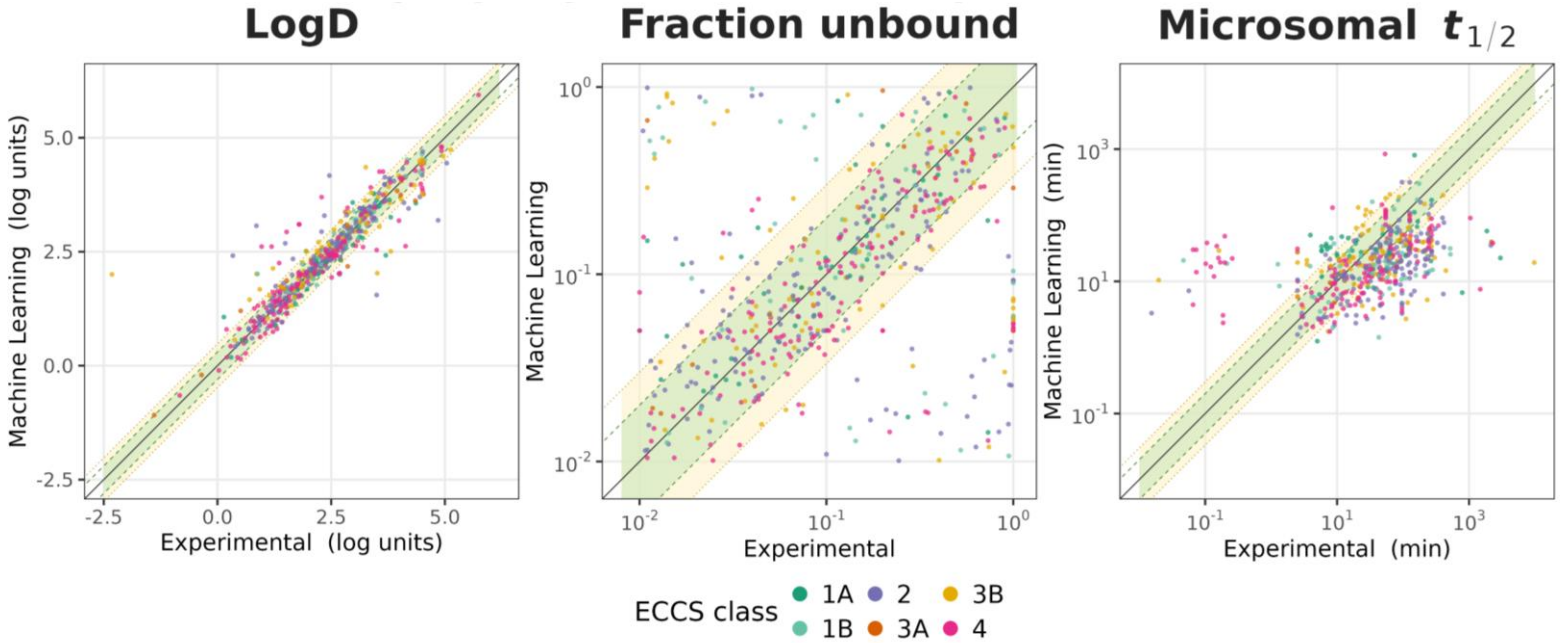
- in vitro ML predictions:
  - Lipophilicity (logD)
  - Plasma protein binding
  - Microsomal half-life

- in vivo observation:
  - Species: mouse, rat, dog, minipig, rabbit, monkey
  - Averaged IV concentration time profiles

ML = Machine Learning
PhysChem = Physicochemical properties (logD, pka etc.)
PK = Pharmacokinetics
IV = Intravenous

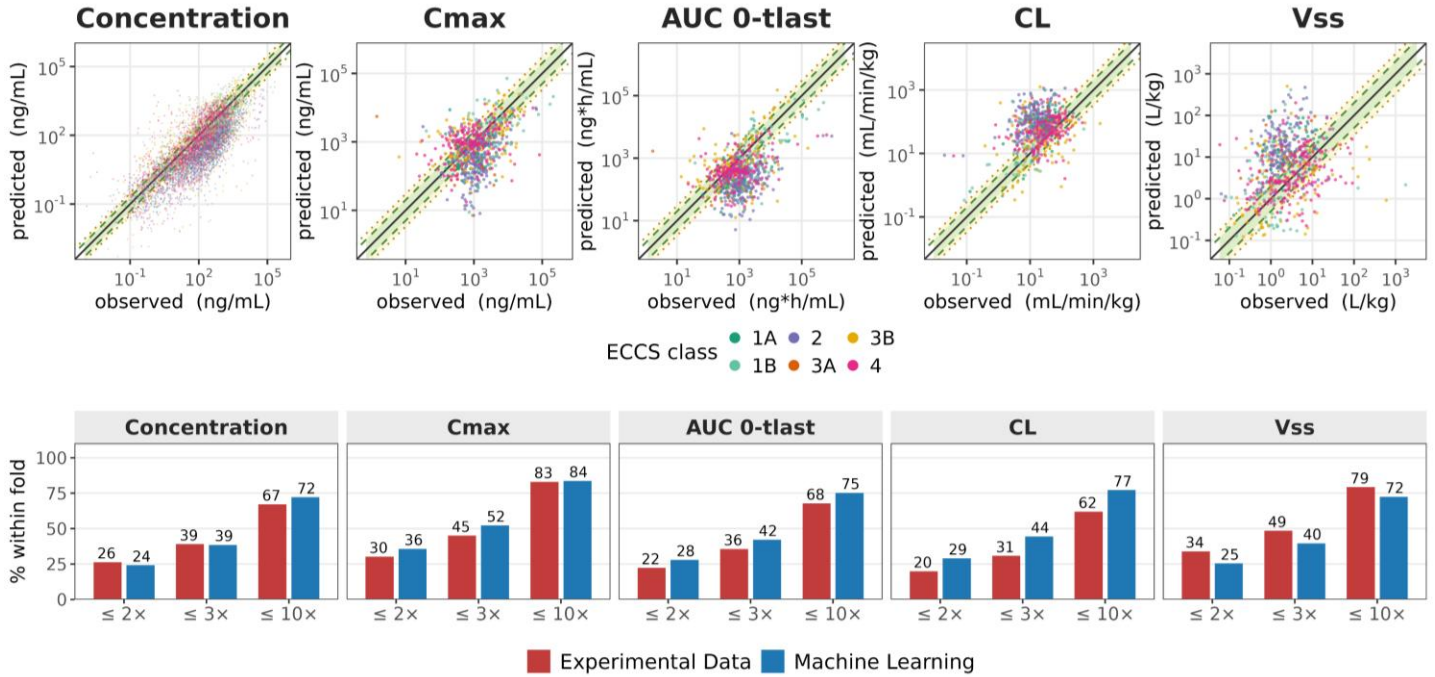


Experimental input parameters vs ML predicted



Summary Similar value for logD and microsomal half life but noisy for fraction unbound

Exposure predictions with machine learning predicted parameters



Summary | Similar performance to exp & ML input parameter

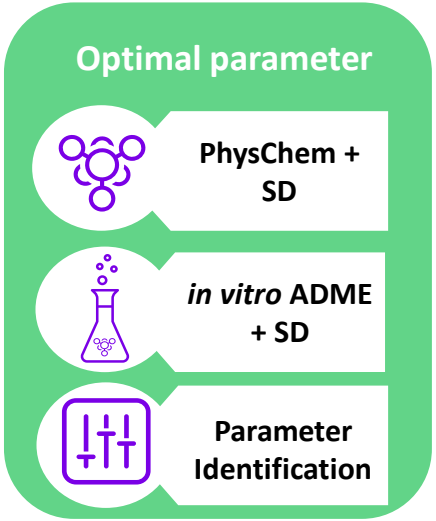
Identifying best model parameters

Given the default HT-PBPK input, what is the prediction performance we can achieve?

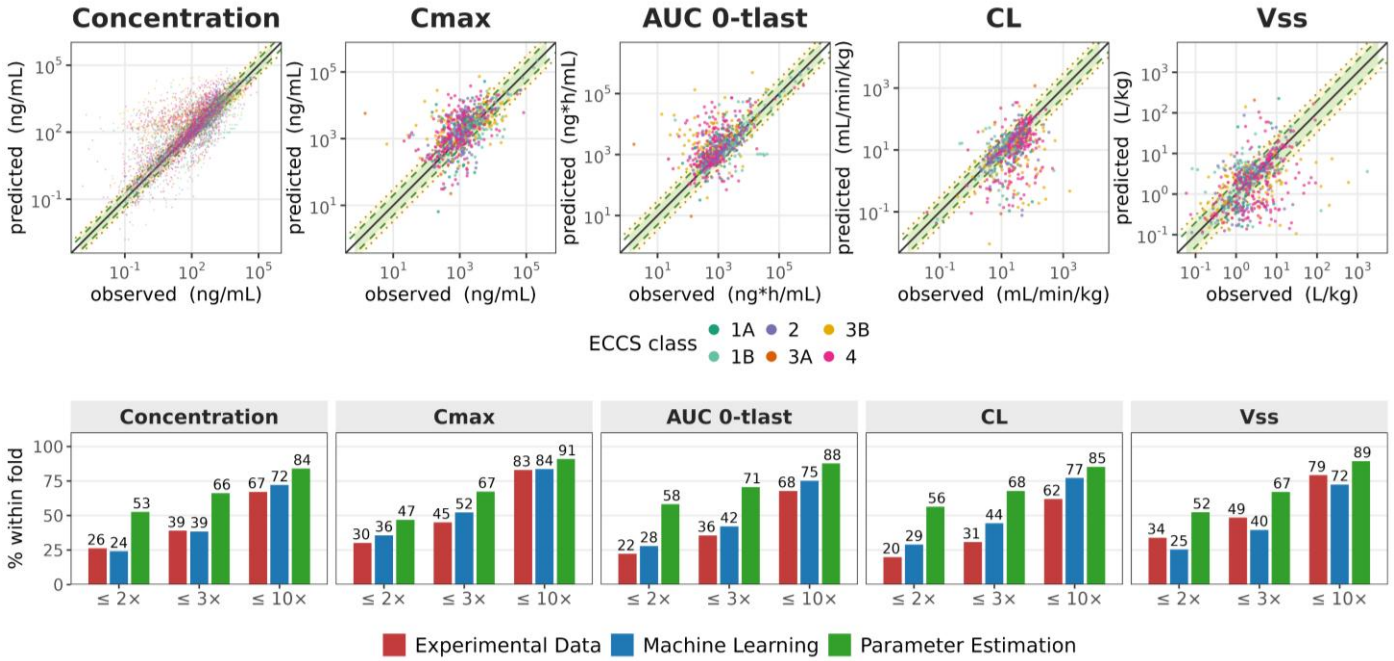
- in vitro as start values for fit, experimental SD as boundaries:
  - Lipophilicity (logD)
  - Plasma protein binding
  - Microsomal half-life

- in vivo observation:
  - Species: mouse, rat, dog, minipig, rabbit, monkey
  - Averaged IV concentration time profiles

PhysChem = Physicochemical properties (logD, pka etc.)
PK = Pharmacokinetics
SD = standard deviation in experimental experiments
IV = Intravenous



Exposure predictions with optimal parameters



Summary

~50 % → 2-fold error for every metric

~70 % → 3-fold error for every metric

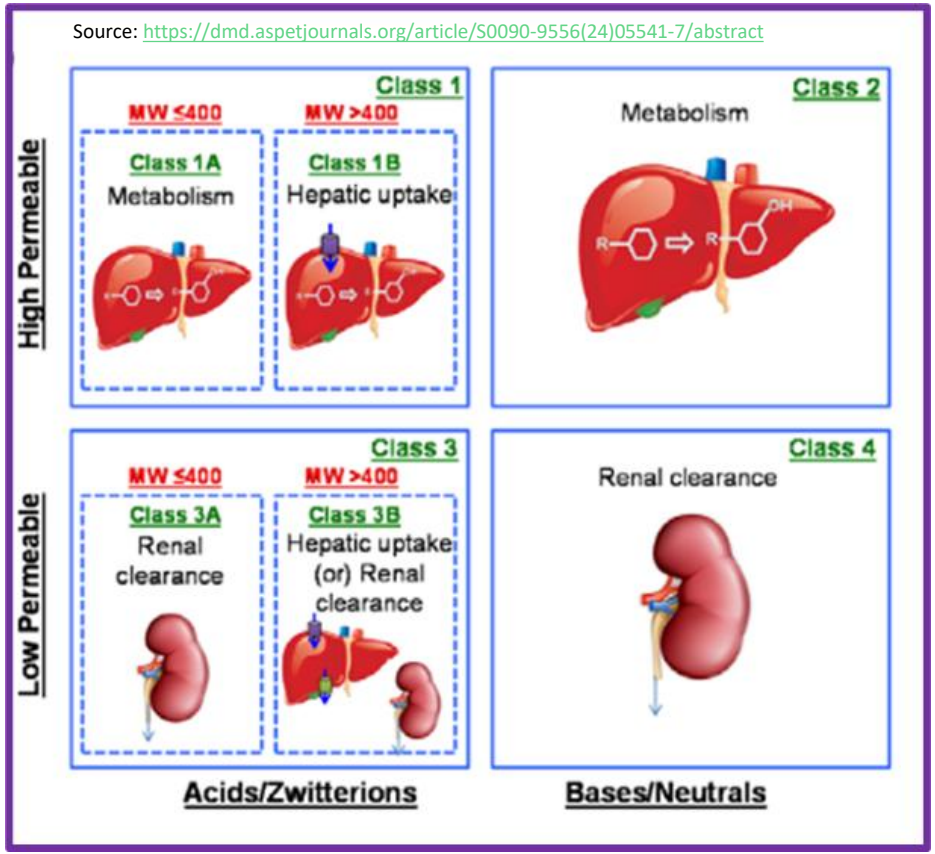
Extended Clearance Classification System of compounds

- Rule of thumb for primary clearance pathway based on:
  - Molecular weight
  - Ionization (pKa)
  - Membrane permeability

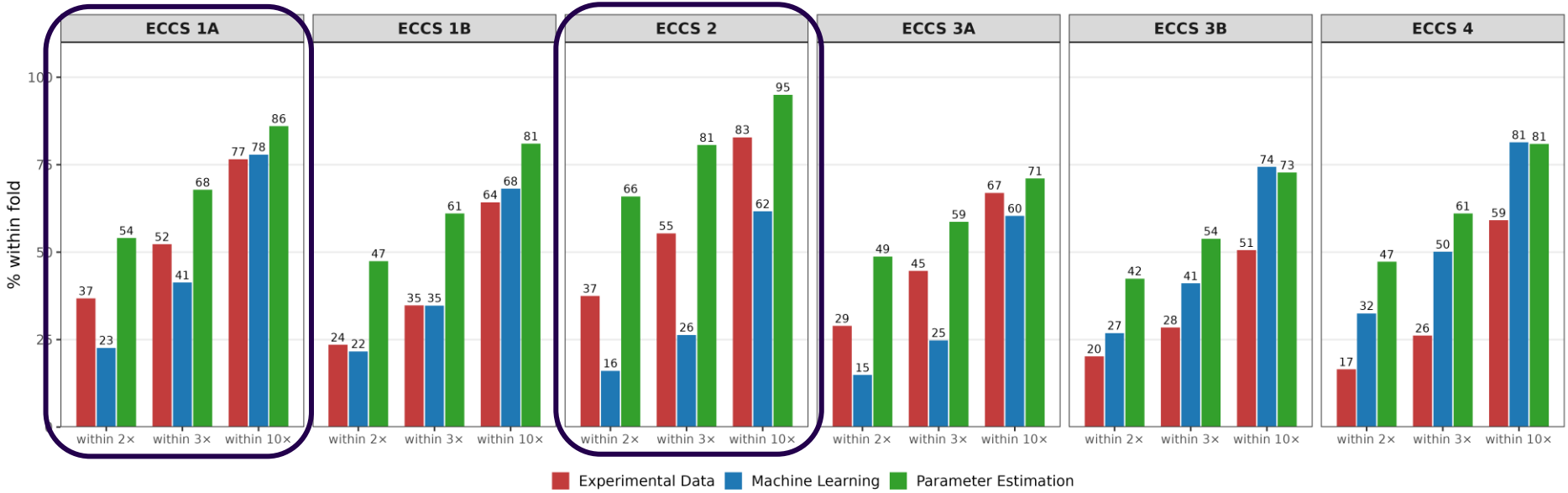
- Categorizes drug in:
  - Liver metabolism
  - Hepatic uptake (transporters)
  - Renal (kidney) excretion
  - A combination thereof

Model input:
Microsome half-life → mostly CYP mediated liver clearance

Source: https://dmd.aspetjournals.org/article/S0090-9556(24)05541-7/abstract



Predictive performance stratified for ECCS (cross species PK-Profile data)



Summary | ECCS class 1A & 2 show best performance; still room for improvement

Conclusion

PBPK for virtual compounds

- ML based pk prediction performance is similar to experimental
- Prediction of PhysChem and ADME inputs enable early PBPK predictions for virtual compounds

Limitations

- PK prediction performance ceiling set by level of mechanistic detail
  - Model assumption build on CYP & liver only clearance (microsomal)
- Uncertainty in ML-predicted parameters limits predictability on “new compounds”

Future perspective

- Evaluate alternative processes (total hepatic clearance, liver hepatocytes, liver microsomes)
- Evaluate alternative input parameter (Hepatocytes)

Summary



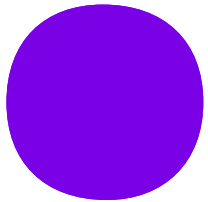
High-throughput PBPK modelling prepares mechanistic PK predictions at scale
Combining machine learning and HT-PBPK enables PK prediction for virtual compounds

Thank you for your Attention

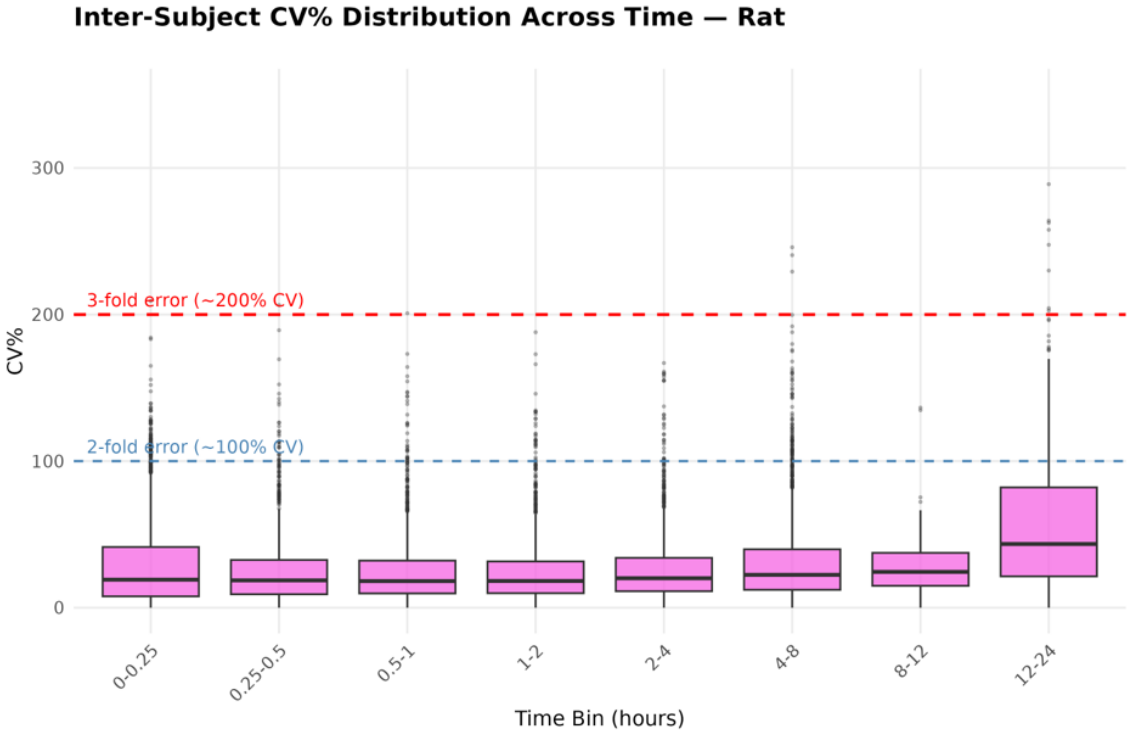


sanofi

Backup



Observed inter-individual variability provides context for performance



Why this matters
- Mean predictions
- Observed data vary between individuals

How variability was quantified
- Group by study, compound, species and time
- Calculate variance across individual concentrations

Key takeaway
Model performance should be interpreted relative to observed biological variability

Summary Observed variability sets the context for evaluating mean PBPK predictions