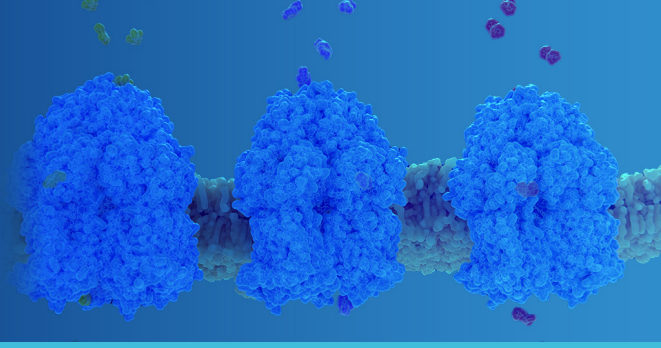
tetrascience

Accelerating ADME/Tox testing with data science and AI



CASE STUDY

Drug development failures often stem from poor absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) properties.[1] To detect these issues at an early stage, biopharma companies employ ADME/Tox screening in the initial phases of lead optimization. However, current *in vitro* and *in vivo* approaches for evaluating ADME/Tox properties are costly and time consuming. Therefore, ADME/Tox screening predictions based on *in silico* modeling techniques have become a promising approach in early drug discovery and development.[2,3]

This case study showcases how Charles River Laboratories Hungary (SOLVO) plans to streamline ADME/Tox screening with artificial intelligence (AI). The science-driven, data-centric approach reduces costs and accelerates *in vitro* experiments for biopharma companies.

The Challenge

One key safety measure in ADME/Tox screening is the inhibition of pharmacologically relevant drug transporters by a drug candidate. Inhibitory potency is often expressed as the half-maximal inhibitory concentration (IC_{50}), that is, the concentration of a drug needed to decrease the activity of a given drug transporter by 50%. Inhibition is typically assayed at multiple dilutions of the drug candidate to establish a dose-response curve, and the results are compared with a known reference inhibitor.

Current methods to determine IC_{50} of drug candidates often use rigid experimental schemes with twice the minimum number of sampling points required by FDA guidelines. The power of prior knowledge and massive scientific tests is underutilized.

The Solution

We present a novel approach to determining IC_{50} values in ADME/Tox screening. It provides an optimal concentration sampling algorithm to predict a drug candidate's half-maximal inhibitory concentration. Utilizing the Tetra Scientific Data and AI Cloud, we developed an integrated solution that seamlessly combines data replatforming and engineering with iterative model development and validation (Figure 1). The end-to-end process comprises five steps (Figure 2).

charles river

Challenge:

Scientists performing drug transporter assays follow a rigid, inefficient sampling scheme for dose-response curves. They test more concentrations of a drug candidate than required by regulations to determine its IC_{50} value.

Solution:

By replatforming and engineering data, the Tetra Scientific Data and AI Cloud™ supports the development and validation of an *in silico* model to optimize sampling and improve IC_{50} calculations.

Outcomes:

- Cut the number of sampling points by half while improving IC_{50} accuracy
- Reduce wet lab experiments by 50%
- Decrease manual data curation by 1.5 FTE
- Developed the *in silico* model in only 5 months

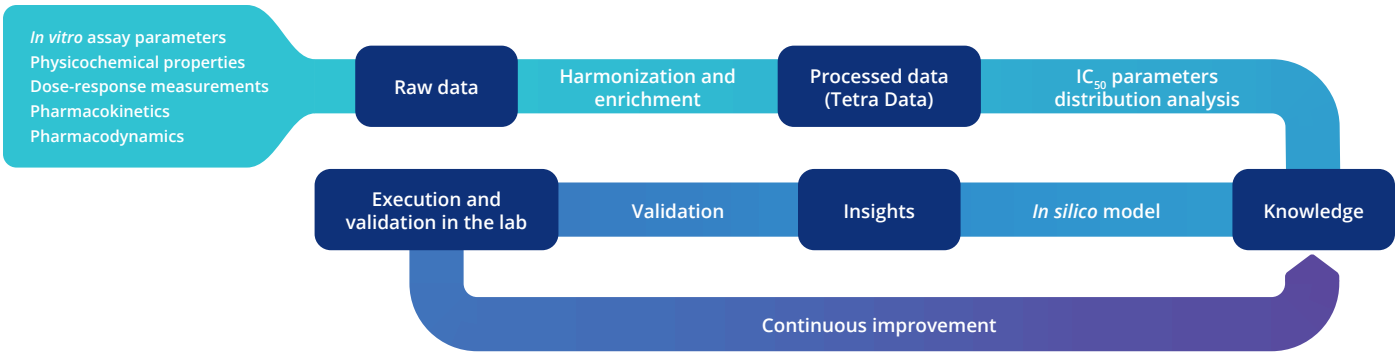


Figure 1. Overview of the science-driven, data-centric AI journey

Step 1: Data replatforming

To optimize the generation of IC_{50} curves through mathematical modeling and machine learning (ML), we needed to consider a myriad of scientific and experimental parameters. These parameters can significantly influence the interaction between the drug candidate and the transporter. Consequently, we undertook a comprehensive data replatforming initiative to integrate such variables into our model. This involved migrating diverse data sources, including legacy reports from the customer's network drives and raw outputs from plate readers and scintillation counters, into the Tetra Scientific Data and AI Cloud. The process preserved the data's scientific context, ensuring data integrity and relevance.

To protect sensitive data, we coordinated the replatforming of legacy reports with the subsequent data engineering step. The customer's recent adoption of the Tetra Scientific Data and AI Cloud catalyzed their laboratory data automation, leading to a richer, multimodal scientific dataset. This shift to a cloud-centric data flow minimized manual data curation, proving critical to our iterative approach and underscoring the essential role of data replatforming in AI/ML initiatives.

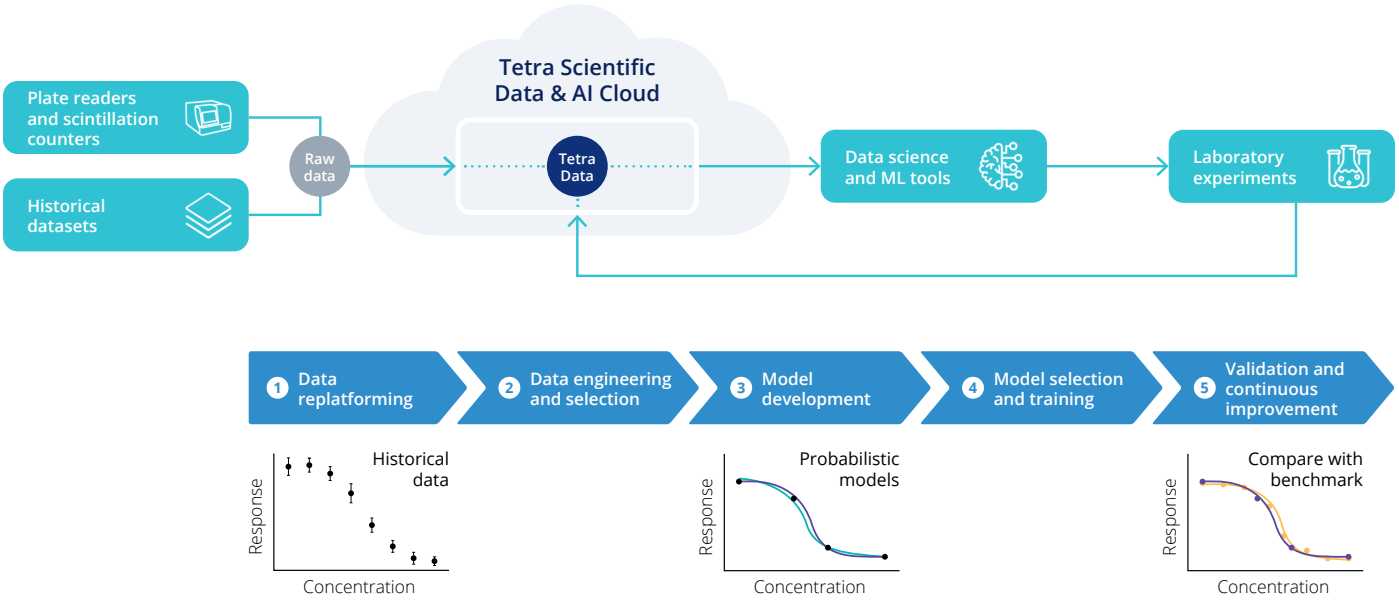


Figure 2. A novel approach to predict compound-transporter interactions

Step 2: Data engineering and selection

This step consisted of two distinct stages: data transformation and scientific collaboration.

Data transformation: We first transformed heterogeneous datasets into a unified, consistent format. This harmonization prepares the data for comprehensive analysis and facilitates the development of predictive models. The resulting data—Tetra Data—is replatformed, contextualized, harmonized, and ready for Scientific AI.

Scientific collaboration: Working with SOLVO's scientists, we identified and selected datasets that are most relevant and valuable for the project. This collaboration ensured that the data we processed and analyzed aligned with scientific objectives and research needs.

Step 3: Model development

Next, we embarked on a comprehensive feature engineering process within the Tetra Scientific Data and AI Cloud. We used an array of prevalent tools for data science and ML development, such as Posit and Google Colab, to craft and refine features that capture the intricate nuances of our datasets. This critical phase involved not only the selection and transformation of relevant variables but also the exploration of novel feature combinations that could unveil deeper insights. After this detailed feature engineering, we performed a series of calibration runs. These runs were instrumental in fine-tuning our models and ensured that the selected features were robust and accurately reflected the distributions of the curves in the uptake inhibition datasets.

> ***"Adopting the Tetra Scientific Data and AI Cloud for lab automation has opened new perspectives. We see great promise in moving from rigid, data-agnostic experimental designs to advanced, optimal techniques that effectively leverage our historical datasets through Tetra Data."***
>
> — Péter Tátrai, Ph.D.
> Senior Scientist
> Charles River Laboratories Hungary (SOLVO)

Step 4: Model selection and training

Following the initial three steps, we collaborated with SOLVO's scientists to select and train predictive models. We applied state-of-the-art optimization and modeling techniques to develop a hybrid approach that combines the adaptability of the ML techniques and the reliability of the information-theoretic mathematical optimization.

Step 5: Validation and continuous improvement

Lastly, we validated the experimental simulation by comparing its results with those from the customer's current rigid IC_{50} sampling scheme. In each experimental iteration, our model suggested the statistically optimal next concentration to sample. These recommendations were guided by information-theoretic principles and informed by historical tests and data from prior iterations. The model also offered an uncertainty estimate for key parameter evaluation, represented as a probability distribution variance. This capability enhances model transparency and enables scientists to end experiments early when success criteria are met. Scientists can also override the model's suggestions and adjust the sampling schedule for the next iteration at their discretion. The continuous incorporation of experimental data, supplemented by periodic feedback from scientists, significantly improved the model's development.

The Results

A focused 5-month collaboration between TetraScience and SOLVO delivered an *in silico* model for optimal iterative sampling of dose-response curves. Compared to the standard dilution series, the *in silico* model improved the accuracy of IC_{50} calculations while reducing the number of sampling points by half. The initial success of this project promises to substantially improve operational efficiency of ADME/Tox assays. We anticipate 50% fewer wet lab experiments.

Other achievements include:

- **Reduced manual data curation:** Automated data replatforming and engineering have reduced the time spent on manual data curation by 1.5 full-time employee equivalents.
- **Fewer concentration sampling points:** The novel approach requires only two sampling points, in addition to a concentration at the drug's solubility limit and a negative control.
- **Higher IC_{50} evaluation accuracy:** Despite using half as many sampling points, the new method estimated IC_{50} values with greater accuracy than the conventional approach.
- **Rapid model development:** Milestones were reached in under 5 months with a two-person team, whereas similar modeling initiatives often require several years and larger teams.
- **Leveraging of historical data:** The *in silico* model used two years' worth of experimental results to learn how to reduce and optimize the testing range for measuring IC_{50} values.

This study demonstrates how AI/ML, powered by the Tetra Scientific Data and AI Cloud, can accelerate and improve ADME/Tox screening. It paves the way for more rapid, scalable, and cost-effective evaluation of drug candidates.

The Next Steps

Our *in silico* experimental simulation yielded striking gains in both the accuracy of estimating IC_{50} values and the efficiency of choosing sampling points. However, since the approach changes how scientists collect and process data, rigorous validation in the lab is necessary.

Further improvement to the model is possible by adopting the following two approaches:

- Identifying scientifically relevant clusters in historical datasets to improve the accuracy of initial curve parameter estimations
- Leveraging molecular fingerprints of drug candidates and fine-tuning pre-trained Multivariate Graph Neural Network (MT-GNN) models

References

[1] Li AP. (2001) Screening for human ADME/Tox drug properties in drug discovery. Drug Discovery Today, 6, 357-366.

[2] Alqahtani S. (2017) *in silico* ADME/tox modeling: progress and prospects. Expert Opinion on Drug Metabolism & Toxicology, 1147-1158.

[3] Ekins S, Lane TR, Urbina F, Puhl AC. (2023) *in silico* ADME/tox comes of age: twenty years later. Xenobiotica, 1-7.