Case Study

DATAHOW

Hybrid Model Development to Support Process Understanding: An Industrial mRNA In Vitro Transcription Case Study

This case study was performed in collaboration with Janssen as part of a published work

Introduction

In vitro transcription (IVT) for large mRNA production is a complex and sensitive process, where transcription efficiency and product integrity are strongly influenced by parameters such as temperature, reagent ratios, material variability, and process timing.

Exploring these interactions experimentally is time-consuming and costly, particularly when lot-to-lot variability and unmeasured effects complicate purely mechanistic modelling approaches.

AI-enabled hybrid modelling offers a powerful framework to address these challenges by combining fundamental process understanding with data-driven learning. By accurately capturing key process dynamics, including mRNA yield, fragmentation, NTP consumption, and pH evolution, across a range of operating conditions, the approach enables deeper insight into critical drivers of performance and supports more efficient process development and scale-up strategies.

Objective

OBJ. 1: Demonstrate the hybrid model's ability to accurately predict IVT process performance across diverse experimental conditions.

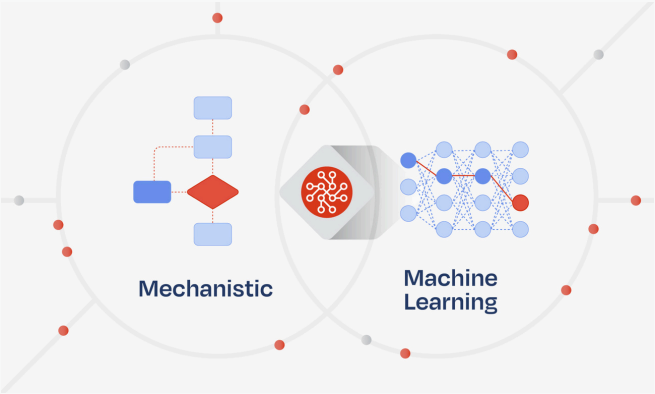
OBJ. 2: Guide process development through iterative model-driven experimentation to meet mRNA yield and integrity targets.

OBJ. 3: Derive deeper process understanding through Shapley value analysis and what-if simulations to support process optimization.

Approach

The hybrid model was trained on 33 experiments and validated against 9 unseen experiments, achieving prediction errors comparable to natural batch-to-batch variation.

An active learning strategy was applied across five experimental iterations: model predictions guided the design of each campaign, and new data was used to retrain the model, progressively converging on optimal process conditions.



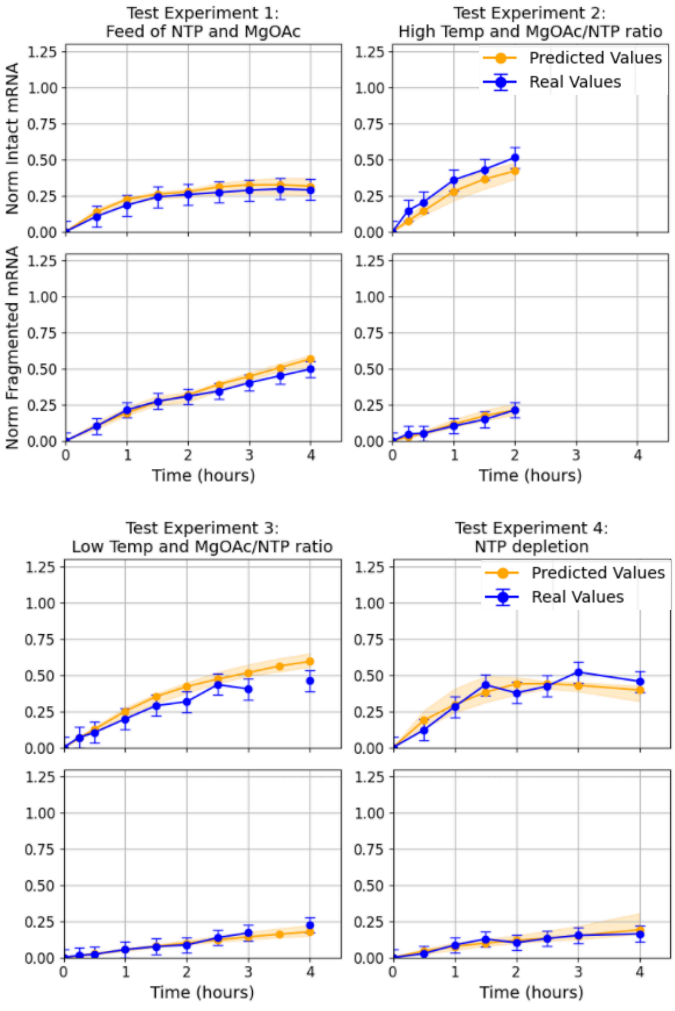
Hybrid Models

DataHow has pioneered the application of hybrid models to bioprocessing. The mechanistic backbone transports the domain's scientific knowledge to describe process dynamics, while the machine learning captures unknown, complex, and nonlinear behaviors.

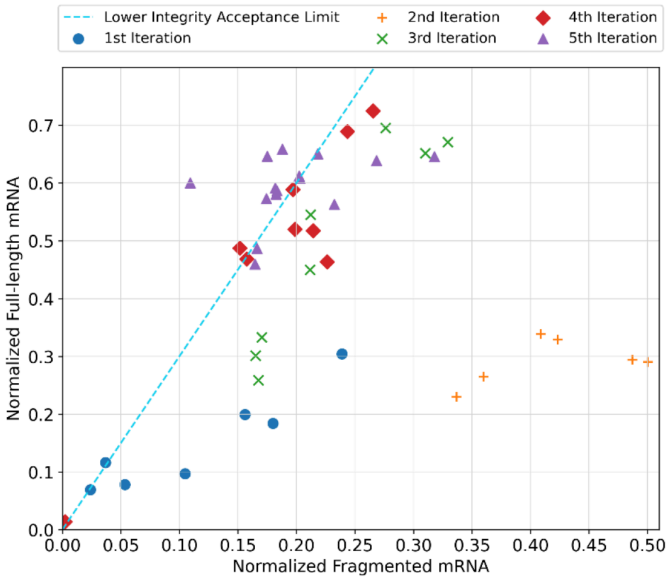
Key Results and Insights

OBJ. 1: The model successfully captured the complex effects of temperature and MgOAc/NTP ratio variations, interactions too intricate for simple mechanistic equations alone, by leveraging its machine learning component to learn these non-linear relationships from data.

Additionally, the mechanistic knowledge embedded within the hybrid framework enabled accurate predictions in challenging scenarios involving NTP depletion and dynamic NTP feeding, demonstrating how combining data-driven learning with mechanistic knowledge yields a more robust model for understanding the mRNA IVT process.

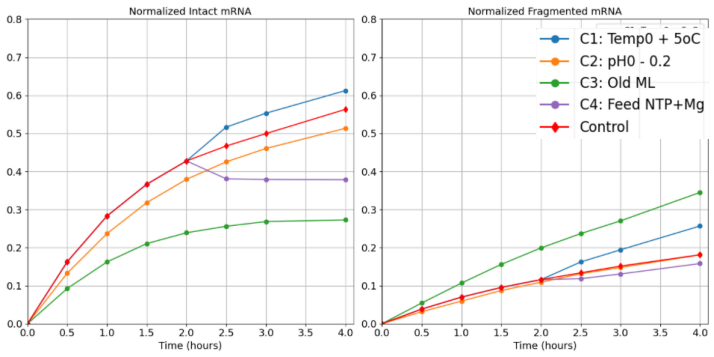


OBJ. 2: Alterations 1 and 2 established the baseline and highlighted the Mg^{2+}/NTP ratio and temperature as key process drivers. From iteration 3 onwards, model-guided design produced a major jump in yield and integrity, with a new DNA lot driving a 24% integrity increase. Iteration 4 delivered the first run exceeding the acceptance limit, with initial pH identified as an additional sensitivity. By iteration 5, all runs exceeded the integrity acceptance limit.



OBJ. 3: Shapley values were used to identify which process variables most drive differences in transcription rate between runs, revealing MgOAc concentration as the key differentiator between high and low yield outcomes, and DTT preparation time as a practically relevant variable. For full details refer to the publication (1). At each iteration, the updated hybrid model was also used to run what-if simulations, exploring the impact of process parameter changes in silico and generating informed suggestions for what to test next.

The simulated conditions shown in the figure are representative examples obtained using the final model trained on all available data as described in OBJ. 2. The temperature ramp (C1) delivered the largest productivity gain, though with a simultaneous increase in fragmentation, a trade-off that would have been costly to discover purely experimentally. A pH reduction (C2) selectively decreased intact mRNA, switching to an older DNA lot (C3) reduced integrity while increasing fragmentation, and a bolus NTP and MgOAc feed (C4) counterintuitively decreased both species, a result later confirmed experimentally.



1. https://www.sciencedirect.com/science/article/abs/pii/S1369703X25003729?dgcid=author