Hybrid-modeling in mRNA manufacturing process development

Konstantinos Alexias
Scientist, API Process Development
Leiden, The Netherlands

Johnson&Johnson
Innovative Medicine

Speaker info

Konstantinos Alexias
Scientist API Process Development,
Leiden, The Netherlands

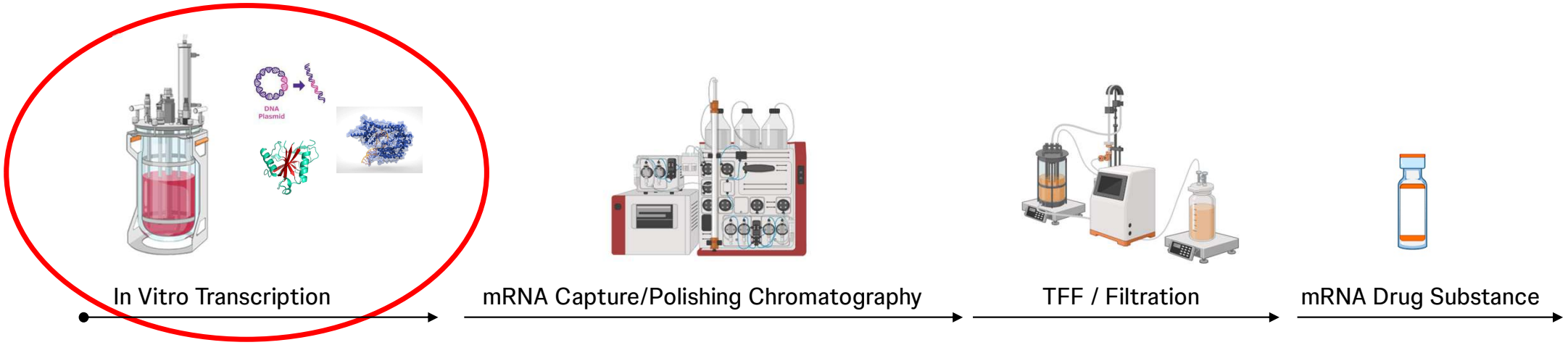
Contents

Introduction

mRNA DS manufacturing process

The In-Vitro transcription reaction

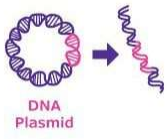
Unit operation general overview



mRNA DS manufacturing process

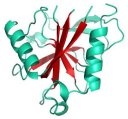
The In-Vitro transcription reaction

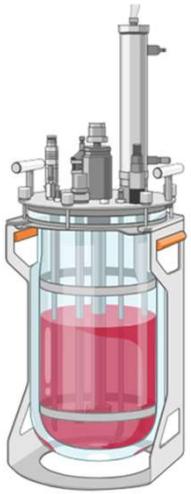
Plasmid DNA linearization



Reaction supporting materials

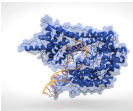
- Pyrophosphatase
- RNAse inhibitor
- Buffers

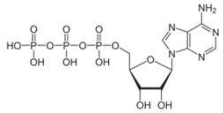




mRNA synthesis

- T7 RNA polymerase
- Nucleotides





Post-IVT Reaction clean-up

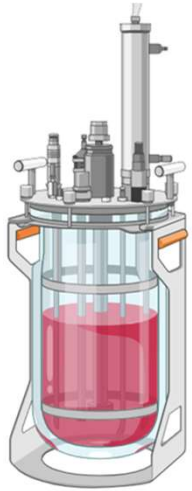
Template removal

mRNA DS manufacturing process

The In-Vitro transcription reaction

Typical IVT reaction process conditions:

- Batch unit operation
- Duration ≤ 4h
- Incubation temperature: 32 - 37°C
- Stirring to ensure homogenization during incubation

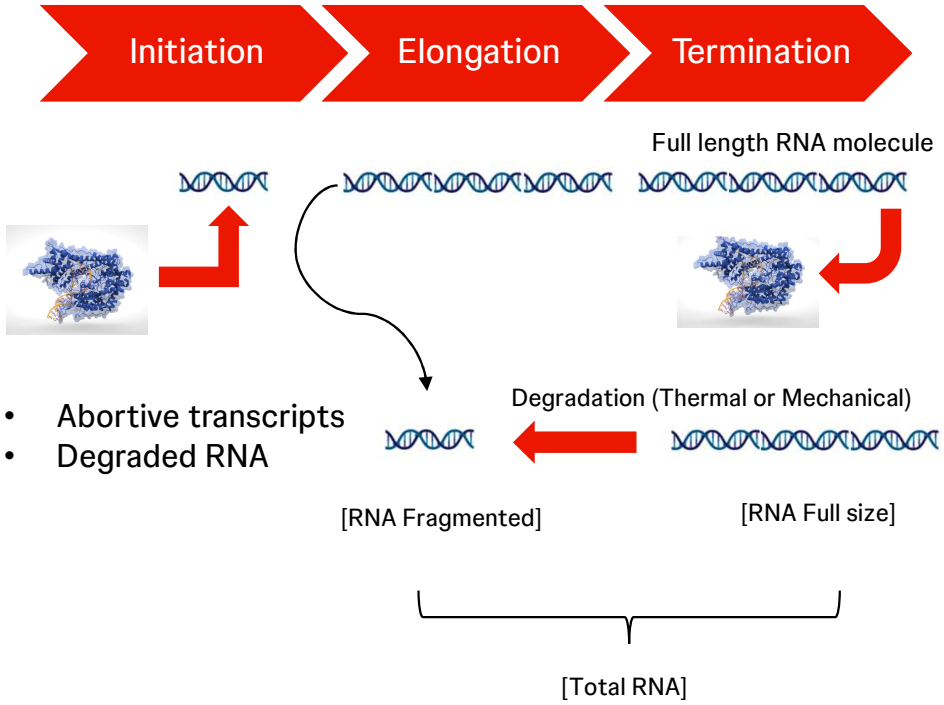


mRNA DS manufacturing process

The In-Vitro transcription reaction

Some IVT underlying mechanisms known from literature:

- Transcription initiation
- Elongation/Polymerization
- Transcription termination
- Aborted transcripts
- mRNA degradation over time (pH, Temperature)



Samnuan et al - Design-of-Experiments In Vitro Transcription Yield Optimization of Self-Amplifying RNA (2021)
Rosa, Sara Sousa et al. - Maximizing mRNA vaccine production with Bayesian optimization. Biotechnology and bioengineering vol. 119,11 (2022)
Satoru Akama et al. - A Multiphysics Model of In Vitro Transcription Coupling Enzymatic Reaction and Precipitation Formation. Biophysical Journal vol 102, Issue 2 (2012)

mRNA DS manufacturing process

The In-Vitro transcription reaction

Key process characteristics

- Absence of cells → well defined enzymatic kinetics instead of complex metabolic pathways
- Batch process mode → simplified mass balances, relative straightforward scale-up, easier overall process monitoring/control compared to other process modes
- Very fast reaction leading to short processing time (≤ 4h)

Major challenges

- Large complex molecules (>10k bp)
- Multiple buffers and other components introduced at specific conditions
- Assay availability and variability impact
- Variation of starting materials which are product specific (DNA Template) could lead to differences in process performance

mRNA DS manufacturing process

The In-Vitro transcription reaction

Goal: Develop a model that can guide process design and optimization of the IVT reaction. The resulting model will:

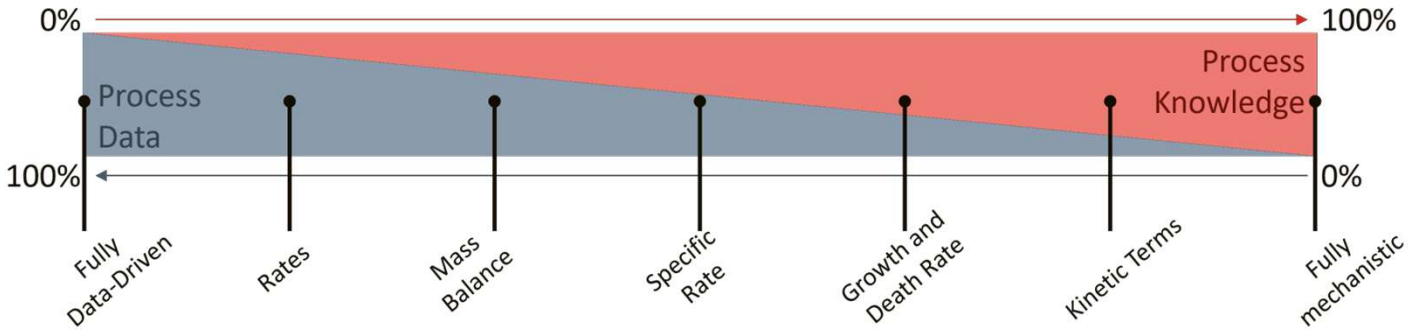
- ☐ **Improve process performance** by guiding process development using the determined process dynamics of the model
- ☐ **Integrate knowledge** by assimilating new information from experimental data throughout the lifecycle of process development into the model
- ☐ **Accelerate process development timelines** by minimizing the required experimental burden and resources by utilizing model predictions to gain insights and develop a process which meets acceptance criteria

Hybrid modeling

What is a hybrid model?

The combination of aspects from multiple approaches of mathematical modeling:

Data Driven Model	Hybrid Model – Combination of Mechanistic & Data driven models	Mechanistic Model
• Rely on machine learning & statistical techniques • Requires a large amount of process datapoints to develop • Disregards domain knowledge • Limited insight into process behavior	• Reduced data requirements (VS data-driven methods) • Integrating domain knowledge & data-driven insights • Retains interpretability while capturing complex behaviors	• Rely on deterministic and algebraic equations • Tedious to develop • Requires a high level of understanding of the underlying process mechanisms



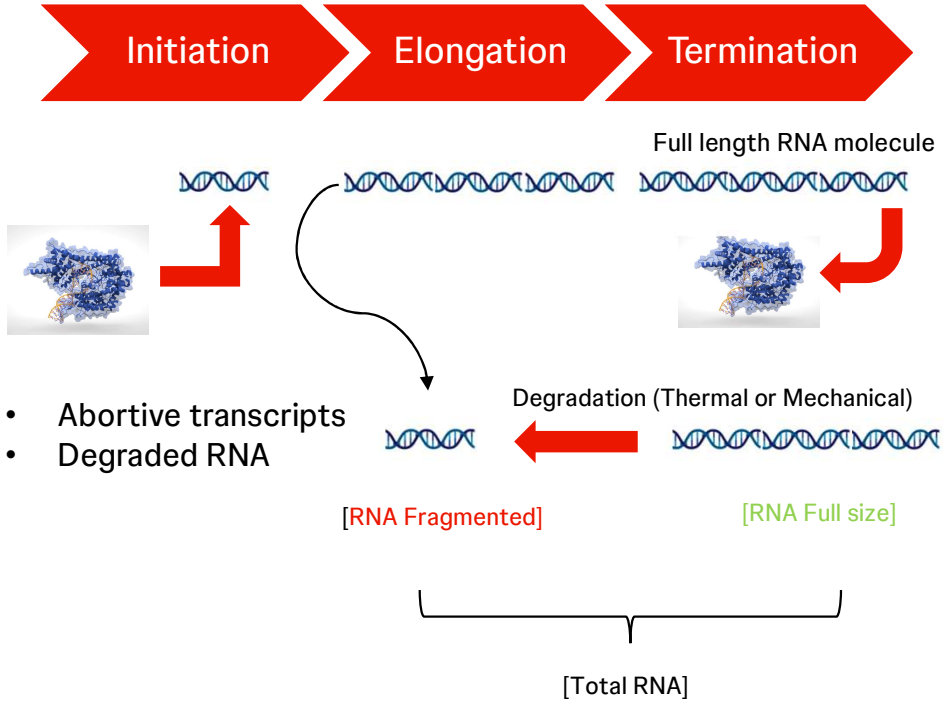
Model structure

What is the purpose of the model?

Define the optimization function of the model based on selected material attributes of the IVT process intermediate:

$$max\ \{RNA\ FullSize - \lambda * RNA\ Fragmented\}$$

→ *Using **total [RNA]** and **fragmentation%** of RNA molecules as process perfomance indicators we can determine distinct levels of optimal process settings*



Model structure

What is the purpose of the model?

- Define the optimization function of the model based on selected material attributes of the IVT process intermediate:

$$max\ \{RNA\ FullSize - \lambda * RNA\ Degraded\}$$

Model Outputs	
Measured	Added variables indicating process performance
1. Total [RNA] 2. Fragmentation (%) 3. [NTP] (ATP, GTP, CTP, UTP) 4. pH	1. Fullsize [RNA] = Total [RNA]*(1-Frag (%)) 2. Fragmented [RNA] = Total [RNA]*Frag (%)

Model Inputs	
Varried/Designed	Fixed/Pre-defined
1. Reaction Volume 2. Incubation duration 3. Incubation temperature 4. Stirring rate 5. DNA Template lot and conc. 6. [HEPES] 7. [Spermidine] 8. [NaOAc] 9. [MgOAc] 10. [DTT] 11. Initial [NTP] 12. T7 Polymerase lot and conc. 13. RNAse lot and conc. 14. iPPAse lot and conc.	1. DNA template Nucleotide distribution 2. Molecular weight of RNA molecule

Model structure

Mechanistic equations

$$\frac{dmRNA_t}{dt} = v_{transcription} - D \cdot mRNA_t$$

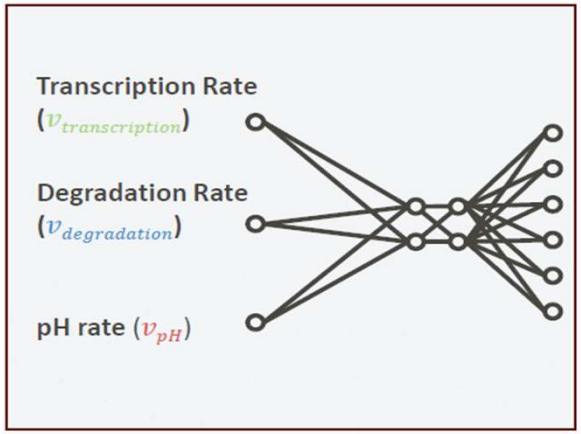
$$\frac{dmRNA_d}{dt} = v_{degradation} - D \cdot mRNA_d$$

$$\frac{dmRNA_i}{dt} = v_{transcription} - v_{degradation} - D \cdot mRNA_i$$

$$\frac{dNTP}{dt} = -S_{NTP} \cdot v_{transcripton} - D \cdot (NTP - NTP_f)$$

$$\frac{dpH}{dt} = v_{pH}$$

Machine Learning Model



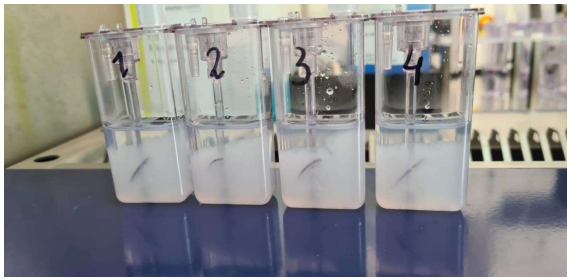
Inputs

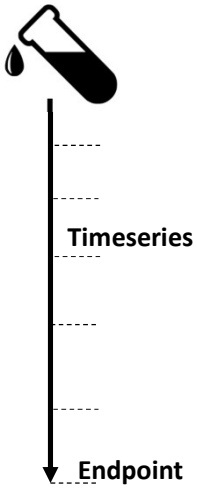
1. Reaction Volume
2. Incubation duration
3. Incubation temperature
4. Stirring rate
5. DNA Template lot and conc.
6. [HEPES]
7. [Spermidine]
8. [NaOAc]
9. [MgOAc]
10. [DTT]
11. Initial [NTP]
12. T7 Polymerase lot and conc.
13. RNAse lot and conc.
14. iPPAse lot and conc.
15. DNA template Nucleotide distribution
16. Molecular weight of RNA molecule

- Process insights gained from the combination of the fixed mechanistic knowledge and the flexible machine learning component of the model
- Characterization of parameters that would not be possible via a detailed mechanistic model, can be taken into account e.g. DNA lot-to-lot variation effect on the process

Experimental setup

Ambr® 15 is a high throughput, automated bioreactor system for 24 parallel cultivations at the 10 –15 mL microbioreactor scale





Material and buffers prepared 72h to 24h prior experimental execution

Multiple conditions per run, online monitoring of pH & Temperature

Automated sampling at multiple timepoints during the reaction with minimum variation introduced from the system

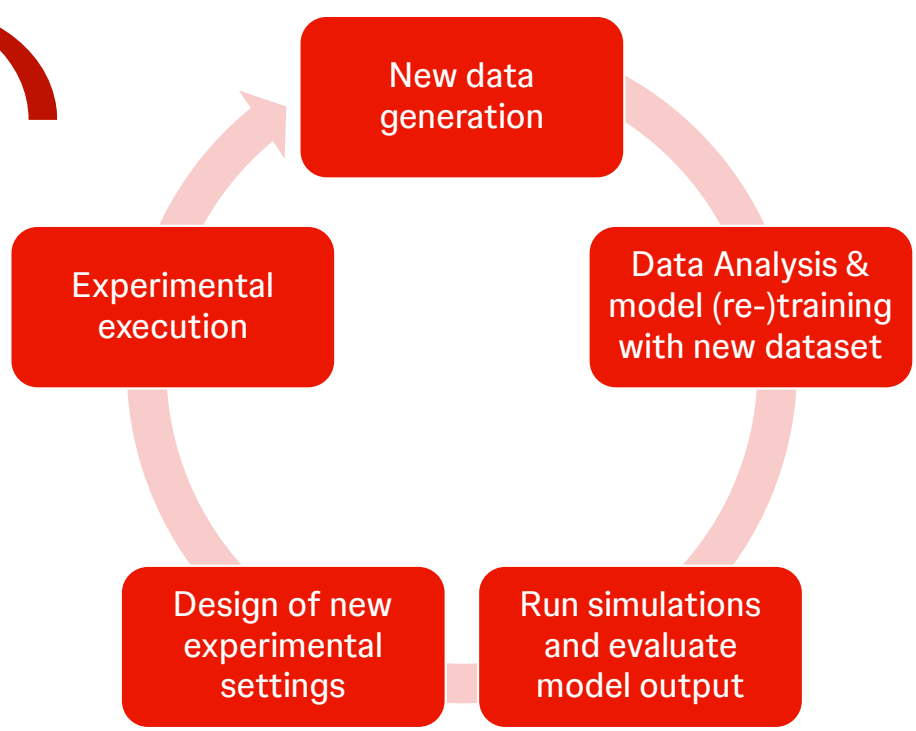
Results & model performance

Model-based Design of Experiments

Iterative IVT process step optimization approach

IVT process performance review and set of a new “golden standard” IVT reaction protocol with each model training iteration

IVTprotocol 1
.
.
.
IVTprotocol N



5 Iterations

Model-based Design of Experiments

Design runs to:

1. **Establish optimum process settings where, max $\{fullsize\ RNA - \lambda * Fragmented\ RNA\}$**
2. **Introduce variability to the dataset to train the model within the defined process design space**

Key process parameters selected as design variables based on knowledge from literature & "what-if" analysis using the model output from each iteration

Design Variables
Initial [NTP]
Incubation Temperature
Incubation duration
T7 RNA Polymerase
[MgOAc]
[HEPES]
[Spermidine]

1 *Samnuan et al - Design-of-Experiments In Vitro Transcription Yield Optimization of Self-Amplifying RNA (2021)*
2 *Rosa, Sara Sousa et al. - Maximizing mRNA vaccine production with Bayesian optimization. Biotechnology and bioengineering vol. 119,11 (2022)*
3 *Satoru Akama et al. - A Multiphysics Model of In Vitro Transcription Coupling Enzymatic Reaction and Precipitation Formation. Biophysical Journal vol 102, Issue 2 (2012)*

Model performance & output accuracy

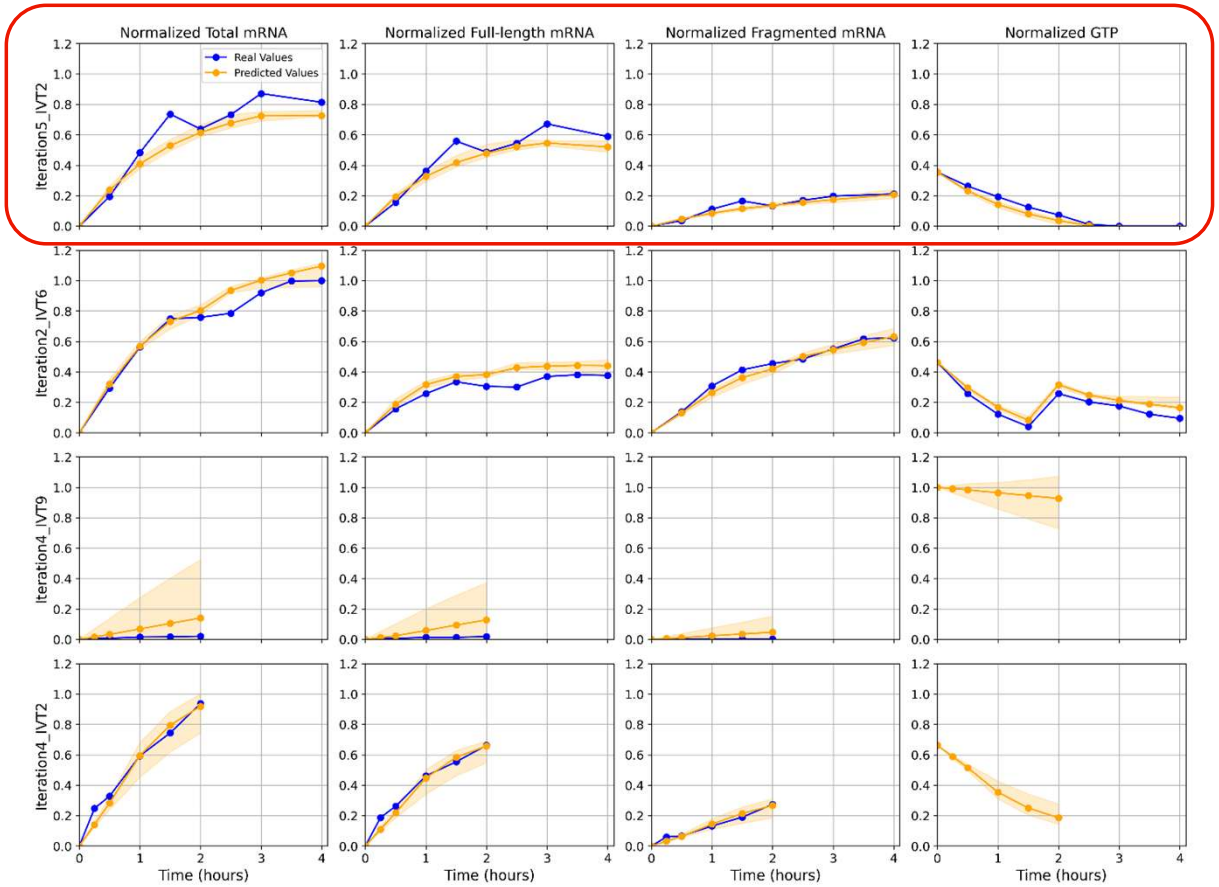
Comparative relative RMSE analysis for hybrid model versions with various train/test splits

			Hybrid Model V1	Hybrid Model V2	Hybrid Model V3	Hybrid Model V4
# of runs in train set			10	15	22	33
# of runs in test set			3	4	6	9
Relative RMSE	Total mRNA	Train	0.13	0.21	0.29	0.24
		Test	0.26	0.22	0.25	0.22
	Full-length mRNA	Train	0.16	0.22	0.31	0.23
		Test	0.33	0.29	0.28	0.27
	Fragmented mRNA	Train	0.16	0.19	0.26	0.23
		Test	0.18	0.18	0.25	0.18
	ATP	Train	0.14	0.31	0.30	0.15
		Test	0.27	0.21	0.19	0.24
	GTP	Train	-	-	-	0.18
		Test	-	-	-	0.26
	CTP	Train	-	-	-	0.22
		Test	-	-	-	0.25
	UTP	Train	-	-	-	0.21
		Test	-	-	-	0.27
	pH	Train	0.13	0.19	0.33	0.19
		Test	0.41	0.20	0.26	0.33

- Data from a previous iteration were included in the new train/test dataset to generate the next model version
- Each model version output guided the experimental design, and the conditions tested in the next iteration
- Relative low-test dataset RMSE observed for most of the model outputs
- To further reduce the RMSE of some model outputs (pH evolution) targeted DoE is needed

Model performance & output accuracy

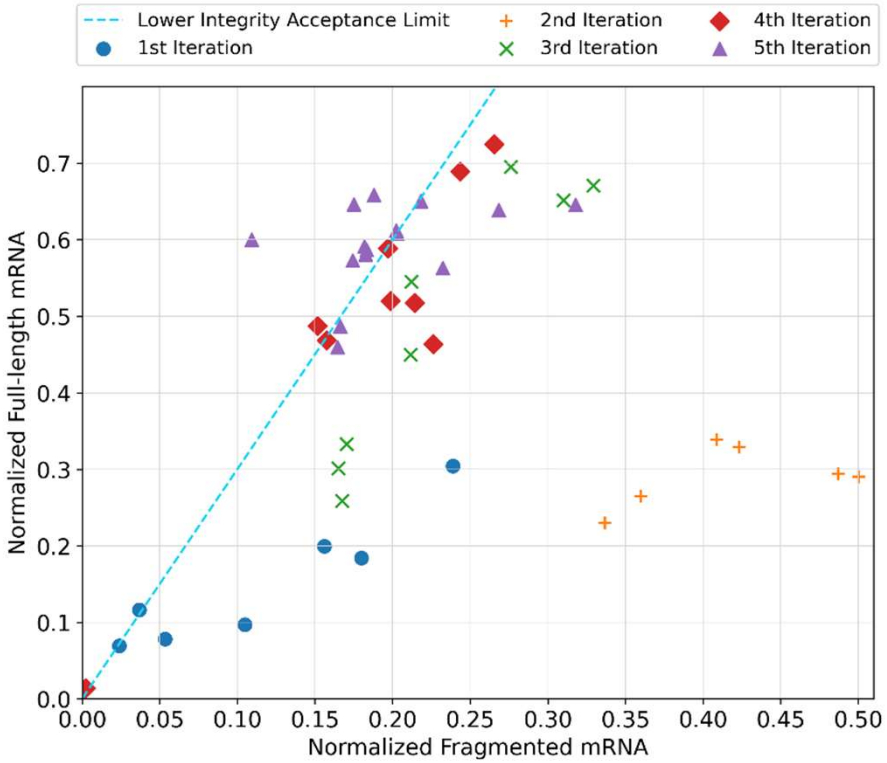
Model output (orange line) and measurements (blue line) of total, full-length and fragmented mRNA and GTP over time



- Shaded orange region represents the model output uncertainty.
- Iteration(X)_IVT(Y) → X iteration, Y specific IVT condition
- Each row corresponds to a unique combination of process settings for a specific iteration:
  - → *Iteration5_IVT2* presents a case where GTP is totally consumed before t=3h.
  - → *Iteration2_IVT6* is a case where NTP and MgOAc are fed at t=1.5h.
  - → *Iteration4_IVT9* is a case with a low MgOAc/NTP concentration ratio.
  - → *Iteration4_IVT2* represents a high Mg/NTP concentration ratio case. Outputs obtained with *Hybrid Model V4*

IVT process performance evolution in time

Result overview of fragmented vs full-length mRNA obtained across the iterations



- Increased molecule integrity for each new iteration – blue dashed line indicating the lower integrity acceptance limit
- For the 1st iteration, conditions were univariately tested to assess impact while reducing technical risk and complexity
- Plasmid DNA template quality had a significant impact on process performance (3rd iteration)
- Optimization of the initial concentrations of NTP and MgOAc further improved performance (4th & 5th iterations)

Summary

A hybrid model of the IVT reaction process was proposed and successfully applied to support process optimization for a specific mRNA sequence

From a modeling perspective:

- ✅ Model output capability: The hybrid modeling structure captured the evolution of NTP concentrations, total, full-length and fragmented mRNA concentrations over time for variations in (raw) materials as well as process parameters.
- ✅ Model-based DoE: The adopted design approach, in combination with the hybrid model, proved effective in leveraging the process region where integrity exceeds the set acceptance threshold
- ✅ Knowledge integration: The hybrid model takes into consideration the specific number of each NTPs present in the DNA Template molecule as well as the DNA molecular weight which corresponds to a specific mRNA product → allowing for potential applications to new mRNA products

Summary

A hybrid model of the IVT reaction process was proposed and successfully applied to support process optimization for a specific mRNA sequence

From an IVT process development perspective:

1. Lot-to-lot variations of starting materials (e.g. DNA template) seem to have a significant impact on the quality attributes of the produced mRNA, particularly integrity → 24% integrity increase
2. Lower incubation temperatures causes reduced transcription rates which lead to lower product yields; however thermal degradation rate is also decreased which leads to lower fragmentation% and higher product quality. Within the studied temperature ranges, significant variations in the process outcome could be observed, and specific temperature profiles could be designed should fragmentation increase or decrease for other mRNA sequences
3. The Mg/NTP ratio has a significant impact on the process outcome, which agrees with literature findings
4. Scale dependent variables, such as stirring rate, that are inputs to the machine-learning part could be replaced by engineering variables (such as volumetric power input) to render the model scale independent

Acknowledgements

- Datahow team

Name	Function
Moritz von Stosch	Project planning & conceptualization
Alice Rosa	Modeling / Data analysis / Reporting
Guilherme Ramos	Modeling / Data analysis / Reporting
Sofia Moreira	Modeling / Data analysis

- JnJ team

Name	Function
Sarah Touw-Mercier	Project planning & conceptualization
Robert Binek	Experimental execution
Jasmin Streur	Experimental execution
Hubert Boulic	Review of results & methods
Lodewijk De Jonge	Review of results & methods
Carmen Berends	Review of results & methods
Konstantinos Alexias	Experimental design / Review of results

The work presented will be submitted for publication in a scientific journal

Thank you

If you have more questions, please contact:
Konstantinos Alexias
kalexias@its.jnj.com

Johnson&Johnson
Innovative Medicine