

## STOCHASTIC MOLECULAR REACTION QUEUEING NETWORK MODELING FOR IN VITRO TRANSCRIPTION PROCESS

Keqi Wang  
Wei Xie  
Hua Zheng

Department of Mechanical and Industrial Engineering  
Northeastern University  
360 Huntington Ave, Boston, MA 02115, USA

### ABSTRACT

To facilitate a rapid response to pandemic threats, this paper focuses on developing a mechanistic simulation model for in vitro transcription (IVT) process, a crucial step in mRNA vaccine manufacturing. To enhance production and support industry 4.0, this model is proposed to improve the prediction and analysis of IVT enzymatic reaction network. It incorporates a novel stochastic molecular reaction queueing network with a regulatory kinetic model characterizing the effect of bioprocess state variables on reaction rates. The empirical study demonstrates that the proposed model has a promising performance under different production conditions and it could offer potential improvements in mRNA product quality and yield.

### 1 INTRODUCTION

In recent years, we have experienced several viral outbreaks, including the COVID-19 pandemic caused by the SARS-CoV-2 virus. Vaccines are highly effective in stopping epidemics and pandemics, but traditional vaccine production methods are too slow to respond to new viral outbreaks (Gebre et al. 2021). To overcome this issue, a rapid-response vaccine production platform is urgently needed. Compared to traditional vaccines, mRNA vaccines offer several key benefits. First, they can be produced rapidly at a large scale, making mRNA vaccines ideal for responding to new viral outbreaks. Second, they are highly effective in preventing infection, as shown with the success of new mRNA vaccines developed for COVID-19. Third, mRNA vaccines are safe and well-tolerated, with few side effects reported.

The mRNA manufacturing involves key steps, i.e., RNA synthesis, purification, and formulation. Basically, the synthesis of RNA strands encodes the desired antigenic protein. For mRNA and self-amplifying RNA (saRNA) vaccine platforms, cell-free DNA-templated RNA synthesis is used for this purpose. It utilizes in vitro transcription (IVT) reactions that are often catalyzed by the T7 RNA polymerase enzyme (T7RNAP). The synthesized RNA substance is then purified and formulated with a delivery system (such as lipid nanoparticles), which are filled into vials or other containers to create the final vaccine product.

This paper focuses on developing a mechanistic simulation model for the IVT process. Similar to the classical assembly lines, the IVT process synthesizes RNA chains based on the DNA templates. It is a complex reaction network involving many factors (such as enzymes, DNA templates, nucleoside triphosphates (NTPs), temperature, and pH) that interactively impact on the production outputs, including yield and product quality. Developing a mechanistic simulation model for the IVT process is crucial in advancing the understanding of underlying mechanisms, identifying key reaction parameters, and optimizing the manufacturing process. The existing partial differential equation/ordinary differential equation (PDE/ODE) kinetic models for in vitro mRNA synthesis (Arnold et al. 2001; van de Berg et al. 2021) incorporate mechanisms from multiple phases, i.e., enzyme binding, initiation, elongation, and termination. They take into account the dynamic changes in enzyme activity, DNA template availability, and substrate concentration and also consider the

effect of factors such as temperature, pH, and ionic strength. However, the existing kinetics modeling approaches are typically deterministic and ignore the intrinsic stochasticity of the IVT process. While there exist stochastic simulation models for therapeutic protein production (Wang et al. 2019; Xie et al. 2023) and mammalian cell culture (Zheng et al. 2023; Wang et al. 2023), a novel stochastic simulation model for the IVT process is currently lacking in the literature.

To support mechanism learning and accelerate automation of mRNA vaccine manufacturing, we propose a *stochastic molecular reaction queueing network model* for the IVT process with the regulatory mechanism accounting for the impact of state variables (e.g., the concentrations of DNA template and enzyme, environmental conditions) on the reaction rates. We establish a relationship between outputs (e.g., RNA yield, and product critical quality attributes (CQAs)) with key input factors, such as initial NTP, Magnesium (Mg) concentration, and buffer profile. The proposed model can improve the IVT process predictions through advancing the understanding of underlying reaction regulation mechanisms.

Therefore, the objective of this study is to develop a stochastic mechanistic model to support IVT process prediction and decision making, which can improve product quality consistency and increase yield. The key contributions are threefold. First, we develop a molecular reaction queueing network model to simulate the RNA synthesis process, accounting for inherent stochasticity. Second, we create a regulatory mechanistic model characterizing the dynamic impact of bioprocess state on the reaction rates. Third, the empirical study demonstrates that the proposed model can enhance mRNA product quality attributes and improve yield. Along with the advanced sensors, the proposed model has the potential to accelerate the development of flexible intensified mRNA vaccine manufacturing processes and support Industry 4.0.

The paper is organized as follows. We provide a brief introduction of the IVT process in Section 2, and describe the state variables and their relationships in terms of mass balance and equilibrium Section 3. Then, we propose a stochastic molecular reaction queueing network and model the regulatory mechanism of reaction rates in Section 4. This IVT process model is validated by using literature data and its performance is studied under various conditions in Section 5. Finally, we conclude the paper in Section 6.

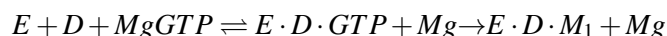
## 2 IVT PROCESS AND REACTION NETWORK

This section reviews the mechanisms of the IVT process and the associated simplified reaction network.

### 2.1 In Vitro Transcription Process

RNA molecules can be produced through in vitro transcription, utilizing the RNA polymerase enzymes (e.g., T7 RNAP), DNA template, nucleoside triphosphates (NTPs) as substrates, and a transcription buffer solution containing Mg and other factors (Young et al. 1997). The IVT process is typically divided into three stages: initiation, elongation, and termination (Cheetham et al. 1999; Dousis et al. 2023); see an illustration in Figure 1. The DNA template includes the RNA polymerase (RNAP) promoter for transcription initiation, the code for the RNA molecule elongation, and the terminator sequence. RNAP is a group of enzymes that catalyzes the synthesis reactions of RNA molecules from a DNA template through the process of transcription (Lodish et al. 2013). T7 RNAP, derived from the T7 bacteriophage, is a popular enzyme used for IVT process in mRNA vaccine manufacturing (Tabor and Richardson 1985).

**(1) Binding and initiation:** Initiation is the first stage of the IVT process, during which T7 RNAP enzyme (denoted as  $E$ ) binds to the promoter region of DNA template (denoted as  $D$ ) to form the transcription initiation complex (IC) (Cheetham et al. 1999). The promoter is a specific DNA sequence that signals RNAP where to bind upstream of a gene sequence. In the case of T7 RNAP, the initiator nucleotide is GTP, which is incorporated into the first position of the synthesized RNA molecule. Magnesium ions ( $Mg^{2+}$ ) play a critical role in the IVT process by supporting the activity of RNAP, stabilizing the RNAP-DNA complex, and promoting the binding of NTPs, including ATP, CTP, GTP, and UTP, to the RNAP. The initiation of enzymatic RNA synthesis reaction network can be written as:



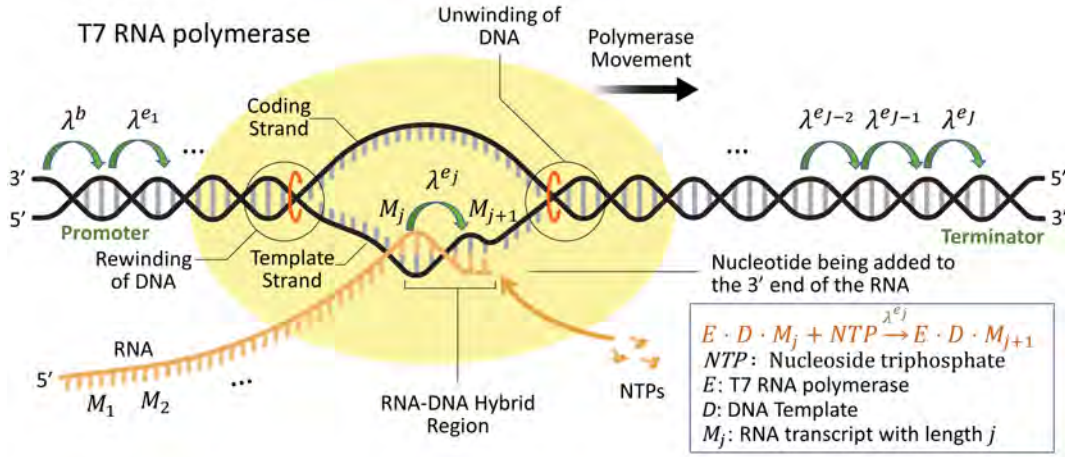


Figure 1: Topology of the mRNA synthesis mechanism with T7 RNAP in three stages of transcription (the plot is adapted from Dornell (2022)).

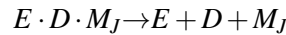
where  $M_1$  stands for an RNA transcript with length 1 for simplification,  $E \cdot D \cdot M_1$  represents an intermediate enzymatic complex containing the enzyme (E), DNA (D), and the nascent RNA ( $M_1$ ).

**(2) Elongation:** During the elongation stage, a stable and processive enzyme elongation complex (EC) is formed. RNA polymerase adds magnesium-complexed NTPs to the growing RNA chain based on the DNA coding sequence, following the rule of Watson-Crick base-pairing interactions. Each time a new NTP is added, one phosphodiester bond between the growing RNA molecule and the newly bound NTP is formed, leading to the release of one pyrophosphate ion (PPi) and one hydrogen ion (H) impacting on the pH level (Young et al. 1997). Thus, the overall enzymatic RNA synthesis reaction during the elongation stage of the IVT process can be described as:



where  $M_j$  stands for an RNA transcript with length  $j$  and  $PPi$  represents inorganic pyrophosphate.

**(3) Termination:** At the termination stage of transcription, the interaction between RNAP and DNA template is disrupted upon encountering a terminator sequence or signal. This halts the addition of complementary NTPs to the RNA strand, and the RNA transcript is released, marking the end of In-Vitro transcription process. The enzymatic RNA synthesis reaction during the termination stage can be expressed as:



where  $M_j$  stands for the full-length RNA transcript with length  $J$ .

The IC is unstable, which can produce short RNA transcripts known as abortive transcripts with 2 to 10 nucleotides in length, in a process called *abortive cycling* (Dousis et al. 2023). In addition, RNA degradation is a common issue that can have a negative impact on the yield and the RNA product CQAs. This process involves the breakdown of RNA molecules into smaller components, which can be caused by various factors, including inappropriate pH level, high temperature, buffer conditions, and contamination.

## 2.2 Critical Process Parameters

The RNA product quality is associated with the percentage of impurities coming from the synthesis of short abortive RNAs and the degradation of the full length RNAs. Based on the study in van de Berg et al. (2021), Table 1 summarizes the critical process parameters (CPPs) that impact the quality and yield of IVT production process. These CPPs include the pH level in the bioreactor, as well as the concentrations of

magnesium, DNA template, T7 RNAP, and NTPs. They are actively involved in the regulatory reaction network of the IVT process (see Figure 2). The regulatory mechanisms discussed below will be integrated into the proposed stochastic reaction network model in Section 4.

Table 1: The criticality of CPPs on the product CQAs. The magnitude of the impact is rated from 0 (low) to 3 (high). The direction and type of CPP-CQA relationship is characterized either by a positive impact labeled with plus “+”, a negative impact labeled with a minus “-”, or a peak behavior whereby the CQA increases with increasing the CPP reaches a peak and then decreases, labeled with plus-minus “±”.

Process parameter	RNA integrity and impurities	RNA yield
pH in transcription reactor	±2	±3
Total Mg concentration	±2	±3
DNA template concentration	0	+2
T7 RNAP concentration	±1	+3
Total NTP concentration	+2	±3

- (1) **pH:** The pH, reflecting the concentration of hydrogen (H) ions, plays a critical role in enzyme binding to DNA template, enzyme activity, and RNA degradation rate during the IVT process. T7 RNAP exhibits optimal activity at a pH range of 7.9 ~ 8.1 (Kartje et al. 2021), although it can still function within a pH range of 7.3 ~ 8.3. Deviations from this range may potentially decrease enzyme activity, negatively affecting the yield and RNA product quality. Therefore, it is crucial to continuously monitor and control the pH level during the IVT process and evaluate its impact.
- (2) **Magnesium:** Mg is an essential cofactor in the IVT reactions. It plays a critical role in stabilizing T7RNAP-DNA complex and promoting RNA synthesis (Vernon 1988). However, a too high concentration of  $Mg^{2+}$  also favors RNA degradation.
- (3) **PPi:** As PPi is continuously produced during the IVT process, it can result in the formation and precipitation of magnesium pyrophosphate ( $Mg_2PPi$ ). This can reduce the available free  $Mg^{2+}$  ions in the solution and ultimately lead to a decrease in the IVT process yield. To avoid this, an enzyme called inorganic pyrophosphatase (iPPase) (Tersteeg et al. 2022) can be utilized to hydrolyze the PPi, preventing the formation of magnesium pyrophosphate and ensuring that magnesium ions stay in the solution, thereby optimizing T7 RNAP enzyme activity.
- (4) **T7 RNAP, DNA template, and NTPs:** The concentrations of these elements are critical as they serve as key raw materials to support the RNA synthesis process. Insufficient concentrations and inappropriate proportions of these key components can result in low yield and poor quality of RNA product. Therefore, it is essential to carefully control and maintain the concentrations of DNA template, T7 RNAP, and NTPs during the IVT process.

To ensure consistent RNA product quality attributes and improved yield, monitoring and regulating the CPPs in the IVT process is essential. For instance, the pH level, influenced by hydrogen generation, affects enzyme binding to DNA template, enzyme activity, and RNA degradation rate; see the reaction network in Figure 2. Similarly, while magnesium plays a critical role in promoting RNA synthesis, the generation of PPi can reduce the free  $Mg^{2+}$  concentration and inhibit RNA synthesis. The control strategies, such as mechanistic model-based reinforcement learning (Zheng et al. 2022), can support *end-to-end IVT process control* so that these CPPs are optimized accounting for complex interactions and long-term effects.

### 3 STATE VARIABLES

In this section, we identify the system state variables of interest and present the mass balance and equilibrium equations as the constraints to the concentrations of different media components. Based on the information from Table 1 and van de Berg et al. (2021), at any time  $t$ , the system *state variables* include the concentrations

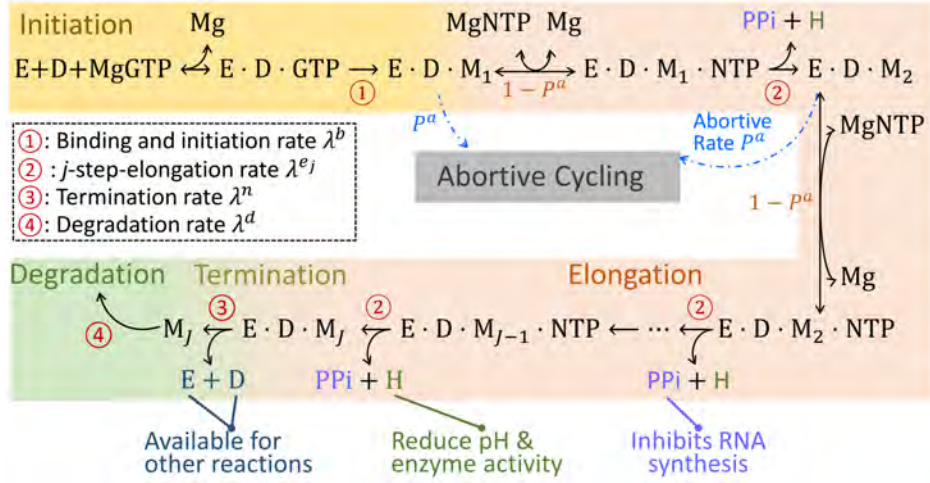


Figure 2: A simple illustration of the regulatory reaction network for the IVT process.

of molecules: T7 RNAP enzyme  $[T7RNAP]_t$ , DNA template  $[DNA]_t$ , free species  $\mathbf{s}_t^{free}$ , complexes  $\mathbf{s}_t^{comp}$ , total  $\mathbf{s}_t^{tot}$ , synthesized full-length mRNA  $[RNA]_t$  and undesirable impurities  $[Impurity]_t$ :

$$\mathbf{s}_t = \left\{ [T7RNAP]_t, [DNA]_t, \mathbf{s}_t^{free\top}, \mathbf{s}_t^{comp\top}, \mathbf{s}_t^{tot\top}, [RNA]_t, [Impurity]_t \right\}^\top,$$

where the bracket  $[\cdot]$  represents the concentration.

The main free species presented in the solution affecting transcription and degradation kinetics include:  $Mg^{2+}$ ,  $NTP^{4-}$  (i.e.,  $ATP^{4-}$ ,  $UTP^{4-}$ ,  $CTP^{4-}$ , and  $GTP^{4-}$ ),  $H^+$ ,  $HEPES^-$  (buffer) and  $PPi^{4-}$ , i.e.,

$$\mathbf{s}_t^{free} = \{ [Mg]_t, [ATP]_t, [UTP]_t, [CTP]_t, [GTP]_t, [H], [HEPES]_t, [PPi]_t \}^\top.$$

These 8 free solution components can form the 22 complexes:  $HATP^{3-}$ ,  $HUTP^{3-}$ ,  $HCTP^{3-}$ ,  $HGTP^{3-}$ ,  $MgATP^{2-}$ ,  $MgUTP^{2-}$ ,  $MgCTP^{2-}$ ,  $MgGTP^{2-}$ ,  $Mg_2ATP$ ,  $Mg_2UTP$ ,  $Mg_2CTP$ ,  $Mg_2GTP$ ,  $MgHATP^-$ ,  $MgHUTP^-$ ,  $MgHCTP^-$ ,  $MgHGTP^-$ ,  $MgPPi^{2+}$ ,  $Mg_2PPi$ ,  $HPPi^{3-}$ ,  $H_2PPi^{2-}$ ,  $MgHPPi^-$  and  $HHEPES$ . At any time  $t$ , we represent the set of complex's concentrations as a vector  $\mathbf{s}_t^{comp}$ . Then the total concentration of each species, defined as the sum of its free ion concentrations and the concentrations of its complexes, becomes:

$$\mathbf{s}_t^{tot} = \{ [Mg]_t^{tot}, [ATP]_t^{tot}, [UTP]_t^{tot}, [CTP]_t^{tot}, [GTP]_t^{tot}, [H]_t^{tot}, [PPi]_t^{tot}, [HEPES]_t^{tot} \}^\top.$$

In any short time period, we assume that the system is in equilibrium. That means the mass balance (outlined in Table 2) and equilibrium equations (outlined in Table 3) are used to describe the concentration constrains on the chemical species. Basically, the mass balance equations given in Table 2 describe the conservation of mass for each species presented in the system. Each equation defines the relationship between the total concentration of a species in the system, its free ion concentration, and the concentrations of its complexes. For example, in Equation **M1**, the total concentration of magnesium ions is equal to the sum of the concentration of free magnesium ions, magnesium ions bound to ATP, UTP, CTP, GTP, and their corresponding hydrolyzed products, and the concentrations of various magnesium-containing complexes.

The equilibrium equations given in Table 3 represent the chemical equilibrium conditions for the reactions involving various species in the system. Each equation describes an equilibrium constant for the reaction involving the species on the left- and right-hand sides of the equation. For example, in Equation **E1**, the equilibrium constant  $K_{eq,1}$  describes the equilibrium between ATP and its protonated form, HATP, in the presence of hydrogen ions.

Table 2: The mass balance equations (adapted from van de Berg et al. (2021))

<b>M1</b>	$[\text{Mg}]^{tot} = [\text{Mg}] + [\text{MgATP}] + [\text{MgUTP}] + [\text{MgCTP}] + [\text{MgGTP}] + 2 \times [\text{Mg}_2\text{ATP}] + 2 \times [\text{Mg}_2\text{UTP}] + 2 \times [\text{Mg}_2\text{CTP}] + 2 \times [\text{Mg}_2\text{GTP}] + [\text{MgHATP}] + [\text{MgHUTP}] + [\text{MgHCTP}] + [\text{MgHGTP}] + [\text{MgPPi}] + 2 \times [\text{Mg}_2\text{PPi}] + [\text{MgHPPi}]$
<b>M2</b>	$[\text{ATP}]^{tot} = [\text{ATP}] + [\text{MgATP}] + [\text{Mg}_2\text{ATP}] + [\text{MgHATP}] + [\text{HATP}]$
<b>M3</b>	$[\text{UTP}]^{tot} = [\text{UTP}] + [\text{MgUTP}] + [\text{Mg}_2\text{UTP}] + [\text{MgHUTP}] + [\text{HUTP}]$
<b>M4</b>	$[\text{CTP}]^{tot} = [\text{CTP}] + [\text{MgCTP}] + [\text{Mg}_2\text{CTP}] + [\text{MgHCTP}] + [\text{HCTP}]$
<b>M5</b>	$[\text{GTP}]^{tot} = [\text{GTP}] + [\text{MgGTP}] + [\text{Mg}_2\text{GTP}] + [\text{MgHGTP}] + [\text{HGTP}]$
<b>M6</b>	$[\text{H}]^{tot} = [\text{H}] + [\text{MgHATP}] + [\text{MgHUTP}] + [\text{MgHCTP}] + [\text{MgHGTP}] + [\text{HATP}] + [\text{HUTP}] + [\text{HCTP}] + [\text{HGTP}] + [\text{HPPi}] + 2 \times [\text{H}_2\text{PPi}] + [\text{MgHPPi}] + [\text{HHEPES}]$
<b>M7</b>	$[\text{PPi}]^{tot} = [\text{PPi}] + [\text{MgPPi}] + [\text{Mg}_2\text{PPi}] + [\text{HPPi}] + [\text{H}_2\text{PPi}] + [\text{MgHPPi}]$
<b>M8</b>	$[\text{HEPES}]^{tot} = [\text{HEPES}] + [\text{HHEPES}]$

Table 3: The equilibrium equations (adapted from van de Berg et al. (2021))

<b>E1</b>	$[\text{H}][\text{ATP}] = K_{eq,1}[\text{HATP}]$	<b>E2</b>	$[\text{H}][\text{UTP}] = K_{eq,2}[\text{HUTP}]$
<b>E3</b>	$[\text{H}][\text{CTP}] = K_{eq,3}[\text{HCTP}]$	<b>E4</b>	$[\text{H}][\text{GTP}] = K_{eq,4}[\text{HGTP}]$
<b>E5</b>	$[\text{Mg}][\text{ATP}] = K_{eq,5}[\text{MgATP}]$	<b>E6</b>	$[\text{Mg}][\text{UTP}] = K_{eq,6}[\text{MgUTP}]$
<b>E7</b>	$[\text{Mg}][\text{CTP}] = K_{eq,7}[\text{MgCTP}]$	<b>E8</b>	$[\text{Mg}][\text{GTP}] = K_{eq,8}[\text{MgGTP}]$
<b>E9</b>	$[\text{Mg}][\text{MgATP}] = K_{eq,9}[\text{Mg}_2\text{ATP}]$	<b>E10</b>	$[\text{Mg}][\text{MgUTP}] = K_{eq,10}[\text{Mg}_2\text{UTP}]$
<b>E11</b>	$[\text{Mg}][\text{MgCTP}] = K_{eq,11}[\text{Mg}_2\text{CTP}]$	<b>E12</b>	$[\text{Mg}][\text{MgGTP}] = K_{eq,12}[\text{Mg}_2\text{GTP}]$
<b>E13</b>	$[\text{Mg}][\text{HATP}] = K_{eq,13}[\text{MgHATP}]$	<b>E14</b>	$[\text{Mg}][\text{HUTP}] = K_{eq,14}[\text{MgHUTP}]$
<b>E15</b>	$[\text{Mg}][\text{HCTP}] = K_{eq,15}[\text{MgHCTP}]$	<b>E16</b>	$[\text{Mg}][\text{HGTP}] = K_{eq,16}[\text{MgHGTP}]$
<b>E17</b>	$[\text{Mg}][\text{PPi}] = K_{eq,17}[\text{MgPPi}]$	<b>E18</b>	$[\text{Mg}][\text{MgPPi}] = K_{eq,18}[\text{Mg}_2\text{PPi}]$
<b>E19</b>	$[\text{H}][\text{PPi}] = K_{eq,19}[\text{HPPi}]$	<b>E20</b>	$[\text{H}][\text{HPPi}] = K_{eq,20}[\text{H}_2\text{PPi}]$
<b>E21</b>	$[\text{Mg}][\text{HPPi}] = K_{eq,21}[\text{MgHPPi}]$	<b>E22</b>	$[\text{H}][\text{HEPES}] = K_{eq,22}[\text{HHEPES}]$

By solving these mass balance equations and equilibrium equations, the concentrations of any group (i.e.,  $s_t^{free}$ ,  $s_t^{comp}$ ,  $s_t^{tot}$ ) and equilibrium constants can be used to calculate the concentrations of the other two groups. Typically, the initial concentrations of the total group are known except for  $[\text{H}]_t^{tot}$ . However, the concentration of free hydrogen ions  $[\text{H}]_t$  can be determined by using the measured pH value, i.e.,  $\text{pH} = -\log([\text{H}])$ . It should be noted that the choice of free components and complexes is dependent on the specific IVT reaction and can be adapted based on the particular experimental setup. The modeling approach presented in this paper can be extended to incorporate additional or alternative components.

## 4 STOCHASTIC MOLECULAR REACTION QUEUEING NETWORK

Building upon recent research studies (Clement et al. 2020; Kloska et al. 2022), this section aims to develop a stochastic molecular reaction network model for the enzymatic IVT process using queueing theory. In Section 4.1, we model the IVT reaction process illustrated in Figure 2 as a queueing network producing the RNA molecule product. The queueing network includes multiple stages with the reaction rates (i.e., the number of molecular reactions occurring in unit time) depending on the state  $s_t$ . Then, we model the state transition for  $s_t$  characterizing the bioprocess macro-kinetics in Section 4.2.

### 4.1 Reaction Rate Modeling for Stochastic Molecular Reaction Queueing Network

The IVT process queueing network is composed of multiple steps: binding and initiation, abortive cycling, elongation, termination, and degradation. The proposed reaction regulation mechanistic model leverages the information from previously published studies (Akama et al. 2012; Arnold et al. 2001; van de Berg et al. 2021). After the binding and initiation stage, the initiated enzymatic complexes have two possible routes to follow: abortive cycling or elongation; see Figure 2. In abortive cycling, the initiated enzymatic

complexes prematurely release abortive RNA transcripts, before proceeding to the elongation stage. The probability of having an initiated complex entering abortive cycling, denoted by  $P^a(\mathbf{s}_t)$ , is influenced by the state variables, such as the pH in the transcription reactor and the concentrations of NTPs. The abortive RNA transcripts are considered as impurity that reduces IVT product quality as well as yield. If the initiated enzymatic complex successfully proceeds beyond the initiation stage, it enters the elongation stage. During this stage, the T7 RNAP enzyme moves along the DNA template, synthesizing the RNA transcript.

- (1) **Binding and initiation:** The synthesized RNA with length  $j$ , denoted as  $M_j$ , is treated as entities in the queueing system, where  $j = 1, 2, \dots, J$ . The generation rate of the nascent RNA  $M_1$  depends on the DNA template and T7 RNAP enzyme binding rate, denoted by  $\lambda^b(\mathbf{s}_t)$ . That means the DNA template and RNA polymerase are occupied, and GTP is consumed as the entities arriving at a rate of  $\lambda^b(\mathbf{s}_t)$ . The reaction rate of the binding and initiation phase  $\lambda^b(\mathbf{s}_t)$  is modeled as

$$\lambda^b(\mathbf{s}_t) = V_{max,b} \times \frac{[DNA]_t}{K_{M,DNA} + [DNA]_t} \times \frac{[MgGTP]_t}{K_{M,MgGTP}^b + [MgGTP]_t} \times \prod_{i \in \{MgATP, MgUTP, MgCTP\}} \frac{K_{I,i}}{K_{I,i} + [i]_t}, \quad (1)$$

where the parameters  $K_{I,\cdot}$ ,  $K_{M,\cdot}$ , and  $V_{max,\cdot}$  represent the Michaelis-Menten (MM) inhibition constant, the affinity constant, and the maximum specific reaction rate respectively.

The binding and initiation rate in Equation (1) is formulated as the product of three terms: a) the rate of promoter DNA binding, which is influenced by the concentration of DNA; b) the rate of initial GTP-Mg complex binding, which is influenced by the concentration of GTP-Mg complex; and c) a term that takes into account the competition among substrate NTP-Mg complexes. The latter term is given by the product of inverse MM constants ( $K_{I,i}$ ) and the concentration of the competing NTP-Mg complexes, denoted by  $[i]_t$ , where  $i$  belongs to the set of substrate NTP-Mg complexes including MgATP, MgUTP, and MgCTP. By appropriately quantifying the impact of each factor on the binding and initiation rate, Equation (1) can be used to predict the effect of changes in the concentrations of the various species on the efficiency of RNA synthesis.

- (2) **Abortive Cycling:** The production of short RNA or abortive transcripts can reduce RNA yield and purity (Dousis et al. 2023). To account for this, we incorporate the routing probability of abortive cycling into the model and denote it as  $P^a(\mathbf{s}_t)$ . The synthesized abortive RNA transcripts are considered impurities and reduce RNA product integrity. As shown in Table 1, the pH level in the transcription reactor and the concentrations of NTPs are the CPPs affecting the abortive RNA and impurity generation. Since the closed-form expression for the routing probability of abortive cycling, commonly referred to as the abortive rate, is not available in the existing literature, inspired from the work in Xie et al. (2022), we constructed a regulation mechanistic model to characterize the impact of pH (i.e., [H] and [OH]) and the concentrations of NTP-Mg complexes (i.e., [MgATP], [MgUTP], [MgCTP], [MgGTP]) on the abortive cycling routing probability  $P^a(\mathbf{s}_t)$ , i.e.,

$$P^a(\mathbf{s}_t) = f([H]_t, [OH]_t, [MgATP]_t, [MgUTP]_t, [MgCTP]_t, [MgGTP]_t) = \frac{1}{1 + e^{-(\kappa_t - 1)}},$$

where  $\kappa_t = \phi_1[H]_t + \phi_2[OH]_t + \phi_3[MgATP]_t + \phi_4[MgATP]_t + \phi_5[MgUTP]_t + \phi_6[MgCTP]_t + \phi_6[MgGTP]_t$  and the parameters  $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6$  characterize the influence from each CPP input. We assume no delay in releasing the occupied DNA template and enzyme during abortive cycling, i.e., making resources immediately available for other reactions.

- (3) **Elongation:** In the elongation stage, the T7 RNAP enzyme moves along the DNA template, synthesizing the RNA transcript. At each elongation step, one NTP unit is added into the growing transcript, i.e., the nascent RNA chain with length  $j$  having an increment rate denoted by  $\lambda^e(\mathbf{s}_t)$ . This rate depends on the state variables  $\mathbf{s}_t$ , such as reactant concentrations and environmental conditions. As the T7 RNAP enzyme progresses, the added NTP forms complementary base pairs

with the DNA template, effectively elongating the RNA chain. Throughout this stage, both the DNA template and enzyme remain occupied, ensuring that the elongation process continues uninterrupted until the RNA transcript is complete. The per-step-elongation rate depends on the concentration of the correct NTP-Mg complex, denoted by  $[i]_t$  with  $i \in \{\text{MgATP}, \text{MgUTP}, \text{MgCTP}, \text{MgGTP}\}$ , as well as the concentrations of incorrect NTP-Mg complexes with  $q \neq i$ . We model the elongation rate as,

$$\lambda^{e_j}(\mathbf{s}_t) = V_{max,i} \times \frac{[i]_t}{K_{M,i} + [i]_t} \times \prod_{q \in \{\text{MgATP}, \text{MgUTP}, \text{MgCTP}, \text{MgGTP}\} \& q \neq i} \frac{K_{I,q}}{K_{I,q} + [q]_t}. \quad (2)$$

Thus, the elongation rate in Equation (2) is modeled as the product of two terms: a) the rate of type  $i$ -th NTP-Mg complex associated with the ternary complex of T7 RNAP, DNA, and RNA, which is influenced by the concentration of  $[i]_t$ ; and b) a term that takes into account the competition among other substrate NTP-Mg complexes. The latter term is given by the product of inverse MM constants ( $K_{I,q}$ ) and the concentration of the competing NTP-Mg complexes,  $[q]_t$ . The parameter  $V_{max,i}$  represents the maximum possible elongation rate, while  $K_{M,i}$  is the concentration of the NTP-Mg complex that allows for half of the maximum elongation rate to be achieved.

- (4) **Termination:** In the termination stage, the T7 RNAP enzyme reaches to the end of the DNA template, signifying the successful completion of the transcription process. Then, the T7 RNAP enzyme dissociates from the DNA template, and both are released at a rate denoted by  $\lambda^n(\mathbf{s}_t)$ ,

$$\lambda^n(\mathbf{s}_t) = V_{max,n} \times \frac{[E \cdot D \cdot M_J]_t}{K_{M,J} + [E \cdot D \cdot M_J]_t}. \quad (3)$$

A full-length RNA transcript with the desired length  $J$  is synthesized and released, making it available for downstream applications or analyses. This termination rate in Equation (3) is influenced by factors including the concentration of enzyme ( $E$ ), DNA template ( $D$ ), and mRNA ( $M_J$ ), which are all present in the ternary transcription complex. This regulation model contains two parameters:  $V_{max,n}$  and  $K_{M,J}$ , where  $V_{max,n}$  represents the maximum possible rate of termination and  $K_{M,J}$  represents the concentration of the ternary complex that allows for half of the maximum termination rate to be achieved. Equation (3) is structured to exhibit an increasing termination rate as the concentration of the ternary transcription complex rises. However, there is a saturation point beyond which the termination rate remains at its maximum level.

- (5) **Degradation:** During the IVT process, the synthesized full-length RNA product inevitably undergoes degradation at a rate  $\lambda^d(\mathbf{s}_t)$ . We model the degradation rate  $\lambda^d(\mathbf{s}_t)$  of RNA at time  $t$  as follows:

$$\lambda^d(\mathbf{s}_t) = (k_{ac}[\text{H}]_t^{n_{ac}} + k_{ba}[\text{OH}]_t^{n_{ba}} + k_{Mg}[\text{Mg}]_t^{n_{Mg}})[\text{RNA}]_t^{n_{RNA}}, \quad (4)$$

where  $k_{ac}$ ,  $k_{ba}$ ,  $k_{Mg}$ ,  $n_{ac}$ ,  $n_{ba}$ ,  $n_{Mg}$ , and  $n_{RNA}$  are model-specific parameters. The RNA degradation rate in (4) takes into account the effects from two important environmental factors: the pH and the concentration of magnesium ions ( $\text{Mg}^{2+}$ ). The degradation rate is determined by three separate factors: a) the concentration of acidic hydrogen ions ( $\text{H}^+$ ) with a corresponding rate constant  $k_{ac}$  and exponent  $n_{ac}$ ; b) the concentration of basic hydroxide ions ( $\text{OH}^-$ ) with a rate constant  $k_{ba}$  and exponent  $n_{ba}$ ; and c) the concentration of magnesium ions with a rate constant  $k_{Mg}$  and exponent  $n_{Mg}$ . The RNA concentration at time  $t$  is also included in the equation with an exponent  $n_{RNA}$ .

In sum, four types of IVT *reaction rates* are identified and modeled as functions of system state  $\mathbf{s}_t$  at any time  $t$ : 1) the binding and initiation rate  $\lambda^b(\mathbf{s}_t)$ ; 2) the per-step-elongation rate  $\lambda^{e_j}(\mathbf{s}_t)$  with  $j = 2, 3, \dots, J$ ; 3) the termination rate  $\lambda^n(\mathbf{s}_t)$ ; and 4) the degradation rate  $\lambda^d(\mathbf{s}_t)$ . The proposed stochastic model for enzymatic molecular reaction networks can capture the inherent stochasticity of the IVT process with respect to reactant concentrations and reaction rates, as well as modeling the complex mechanisms and dynamic interactions of multiple CPPs to improve the prediction of yield and product quality attributes (PQAs).



## 4.2 Macro-kinetic State Transition Modeling

To model the dynamic evolution of the IVT process, at any time  $t$ , the reaction rates depend on state  $\mathbf{s}_t$ ,

$$\boldsymbol{\lambda}(\mathbf{s}_t) = \{\lambda^b(\mathbf{s}_t), \lambda^{e_2}(\mathbf{s}_t), \lambda^{e_3}(\mathbf{s}_t), \dots, \lambda^{e_J}(\mathbf{s}_t), \lambda^n(\mathbf{s}_t), \lambda^d(\mathbf{s}_t)\}^\top.$$

Let  $\mathbf{N}$  be a  $m \times (J+2)$  stoichiometry matrix that characterizes the structure of the IVT reaction network with  $m$  denoting the dimension of species or state  $\mathbf{s}_t$ . The  $(p, q)$ -th element of  $\mathbf{N}$ , denoted as  $\mathbf{N}(p, q)$ , represents the number of molecules of the  $p$ -th species that are either consumed (indicated by a negative value) or produced (indicated by a positive value) in each random occurrence of the  $q$ -th reaction.

Consequently, during the time interval  $[0, t]$ , the change in the solution mixture profile is derived as:

$$\mathbf{s}_t = \mathbf{s}_0 + \mathbf{N} \cdot \mathbf{R}_t,$$

where  $\mathbf{R}_t$  is a  $(J+2)$ -dimensional vector representing the accumulated number of occurrences of each reaction  $k$  until time  $t$  for  $k \in \{b, e_2, e_3, \dots, e_J, n, d\}$ , i.e.,  $\mathbf{R}_t = (R_t^b, R_t^{e_2}, R_t^{e_3}, \dots, R_t^{e_J}, R_t^n, R_t^d)^\top$ , and  $\mathbf{N} \cdot \mathbf{R}_t$  represents the net amount of IVT reaction outputs up to time  $t$ . Inspired by Anderson and Kurtz (2011), we assume that the occurrence times  $R_t^k$  for each reaction  $k$  follows a nonhomogeneous Poisson process. The intensity of this process is determined by the reaction rates. Therefore, the probability that the  $k$ -th reaction occurs  $r$  times during time interval  $(t, t + \Delta t]$  becomes,

$$P\left(R_{t+\Delta t}^k - R_t^k = r\right) = \frac{e^{-\int_t^{t+\Delta t} \lambda^k(\mathbf{s}_x) dx} \left(\int_t^{t+\Delta t} \lambda^k(\mathbf{s}_x) dx\right)^r}{r!} \triangleq \text{Poisson}\left(\int_t^{t+\Delta t} \lambda^k(\mathbf{s}_x) dx\right).$$

## 5 EMPIRICAL STUDY

In this section, simulation experiments were conducted to evaluate the performance of the proposed stochastic molecular reaction queueing network model. The proposed model was first validated in Section 5.1 with the data and observations from the study (Dousis et al. 2023) to ensure that it provides a reliable representation of the IVT process. In Section 5.2, a series of simulation experiments were performed to assess the performance of the IVT system under different conditions.

### 5.1 Model Validation

This section presents the validation of the proposed stochastic mechanistic model in terms of its ability to predict the purity and yield of RNA products by using a batch-based IVT process with a duration of 150 minutes (i.e.,  $T = 150$ ). The study conducted by Dousis et al. (2023) investigated the IVT process purity profiles by using eight different mRNAs with varying lengths and sequence compositions. The results revealed that the impurity levels for all eight mRNAs fell within the range of 10% to 20%. Our model considers two primary sources of impurity: the entry of initiated enzymatic complexes into abortive cycling and the degradation of RNA products.

The kinetic parameters of the proposed model were determined based on the existing research (Arnold et al. 2001; van de Berg et al. 2021). The values of dissociation equilibrium constants were taken to be  $10^{-6.95}$  for  $K_{eq,1}$  to  $K_{eq,4}$ ,  $10^{-4.42}$  for  $K_{eq,5}$  to  $K_{eq,8}$ ,  $10^{-1.69}$  for  $K_{eq,9}$  to  $K_{eq,12}$ ,  $10^{-1.49}$  for  $K_{eq,13}$  to  $K_{eq,16}$ , and  $10^{-5.42}$ ,  $10^{-2.33}$ ,  $10^{-8.94}$ ,  $10^{-6.13}$ ,  $10^{-3.05}$ ,  $10^{-7.5}$  mol/L for  $K_{eq,17}$  to  $K_{eq,22}$  respectively. The maximum specific reaction rates  $V_{max, \cdot}$  were set as  $1.8 \times 10^{-6}$ . The Michaelis-Menten affinity constant  $K_{M, DNA}$ ,  $K_{M, MgGTP}^b$ ,  $K_{M, MgATP}$ ,  $K_{M, MgUTP}$ ,  $K_{M, MgCTP}$ ,  $K_{M, MgGTP}$  were determined as  $6.3 \times 10^{-9}$ ,  $98 \times 10^{-6}$ ,  $88 \times 10^{-6}$ ,  $44 \times 10^{-6}$ ,  $44 \times 10^{-6}$ , and  $88 \times 10^{-6}$  respectively. The Michaelis-Menten inhibition constants  $K_{I, \cdot}$  were assigned a value of  $2 \times 10^3$ . The parameters related to abortive rate were set as  $25 \times 10^6$ ,  $0.3 \times 10^6$ , 0.5, 0.5, 0.5, 0.5 for  $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6$ . For degradation rate,  $n_{ac}$ ,  $n_{ba}$ ,  $n_{Mg}$  and  $n_{RNA}$  were set to 1 and  $k_{ac}$ ,  $k_{bc}$ ,  $k_{Mg}$  were determined as  $1.2 \times 10^6$ , 0, 0 respectively.

The full length for the target RNA product is set at 50 NTPs. To simplify the analysis, we assume that abortive cycling only occurs when the length of the growing RNA reaches 9 NTPs, considering the fact that the typical length of abortive transcripts ranges from 2 to 10 (Dousis et al. 2023). The main reason for this simplification is the lack of a comprehensive mechanistic model that accurately characterizes the decrease in abortive rate as the synthesized mRNA length increases. We conducted  $R = 30$  replications of the IVT process and each simulation run has run-length equal to  $T = 150$  minutes. Considering the relatively higher cost of NTPs compared to other raw materials (van de Berg et al. 2021), we conducted the simulation experiments with limited NTP resources, specifically 0.075 M Mg, 0.0015 M ATP, 0.0015 M UTP, 0.0015 M CTP, 0.0015 M GTP,  $10^{-5}$  M DNA template and  $10^{-5}$  M T7RNAP. The reactor solution was initially set at a pH level of 8. Figure 3 depicts the characteristics of 3 representative samples out of the 30 simulated processes.

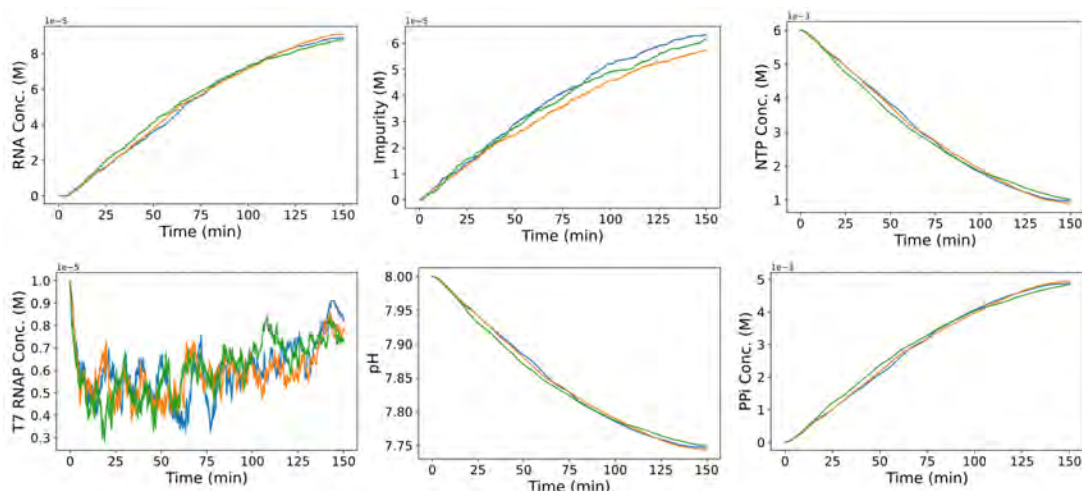


Figure 3: Batch-based IVT process characteristic prediction. (A) Synthesized RNA Concentration, (B) Impurity Concentration, (C) NTP Concentration, (D) T7 RNAP concentration, (E) pH, and (F) PPI concentration. Three colored lines denote three representative simulation output scenarios out of the total 30 replications.

Throughout the process, the production rate of full-length RNA gradually declines due to several reasons. Firstly, the accumulation of PPI leads to a decrease in the availability of the cofactor Mg in the solution. Secondly, the release of hydrogen ions (H) causes T7 RNAP enzyme to deviate from its optimal activity range. Third, the consumption of NTPs, which are raw materials, contributes to the reduction in the rate of RNA synthesis. The expected yield  $\eta = E[[RNA]_T]$  with 95% confidence interval (CI) across 30 replications was reported as  $97.56 \pm 0.82 \mu\text{M}$ . Since the utilized parameters are derived from existing literature estimation results, the reported results maintain validity and reasonableness. The expected impurity level  $\rho = E\left[\frac{9 \times [\text{Impurity}]_T}{9 \times [\text{Impurity}]_T + 50 \times [\text{RNA}]_T}\right]$  with 95% CI was reported as  $11.23 \pm 0.19\%$  weight/weight (w/w), which is consistent with the findings (i.e., 10% to 20%) in Dousis et al. (2023).

## 5.2 IVT Process Performance Analysis under Different Decisions

In this section, we examine the performance of the IVT process under different decision scenarios for the CPPs, considering the CPP-CQA relationship summarized in Table 1. It is worth noting that T7 RNAP, Mg, and NTPs play vital roles as raw materials. Insufficient concentrations of these components can result in inefficient IVT process performance. Conversely, excessively high concentrations of NTPs or imbalanced proportions among different types of NTPs can result in substrate nucleotide competition. This leads to a decrease in the transcription reaction rate. Therefore, we focus on the pH in the transcription reactor and the concentration of NTPs. We conducted 30 replications ( $R = 30$ ) for each experimental setting.

- (1) **pH:** The model was implemented at different pH levels, specifically 7, 7.5, 8, 8.5, and 9. All other experimental parameters remained the same as described in Section 5.1. The performance of the IVT process, with respect to the RNA yield and impurity level, is presented in Table 4.

Table 4: Batch-based IVT production system performance at different pH conditions.

pH	7	7.5	8	8.5	9
Yield ( $\mu\text{M}$ )	$24.38 \pm 0.74$	$84.75 \pm 0.38$	$97.56 \pm 0.82$	$95.44 \pm 0.62$	$71.63 \pm 0.50$
Impurity (%)	$56.01 \pm 0.90$	$16.48 \pm 0.22$	$11.23 \pm 0.19$	$12.72 \pm 0.03$	$23.27 \pm 0.18$

The efficiency of T7 RNAP is significantly influenced by the reactor environment, and the optimal condition is  $\text{pH} \approx 8$ . Under this condition, the process achieves the highest yield of  $97.56 \pm 0.82$  with the lowest impurity level of  $11.23 \pm 0.19$ . When the pH is maintained within the range of 7.5 to 8.5, the system demonstrates comparable performance. However, when the pH deviates from this range, such as pH equal to 7 or 9, the activity of the enzyme declines rapidly. As a result, the production rate decreases significantly, and impurities accumulate due to the elevated abortive rate.

- (2) **NTPs:** The simulation experiments were performed using different concentrations of NTPs, namely 0.006 M, 0.008 M, 0.01 M, and 0.012 M, with each NTP set at an equal concentration. The results provide insights into the system's performance, specifically in terms of RNA yield and impurity level, as summarized in Table 5.

Table 5: Batch-based IVT production system performance at different NTP concentrations.

NTP (M)	0.006	0.008	0.01	0.012
Yield ( $\mu\text{M}$ )	$97.56 \pm 0.82$	$118.88 \pm 0.77$	$131.38 \pm 1.96$	$134.38 \pm 1.46$
Impurity (%)	$11.23 \pm 0.19$	$10.95 \pm 0.13$	$10.56 \pm 0.38$	$10.75 \pm 0.20$

The yield exhibits an increase (from  $97.56 \pm 0.82$  to  $134.38 \pm 1.46$ ) as the NTPs concentrations increase, emphasizing their crucial role in the IVT process. However, the rate of increase diminishes due to two potential reasons: a) the higher NTPs concentrations lead to increased competition inhibition (Arnold et al. 2001), which is incorporated into the regulation mechanistic model in Equations (1) and (2); and b) the concentrations of Mg, DNA template, or T7 RNAP become limiting factors, hindering further improvements. Conversely, the impurity level remains relatively stable and does not show significant changes across the range of NTPs concentrations.

## 6 CONCLUSION

The urgency of rapid vaccine production has been highlighted by recent viral outbreaks. RNA vaccines, known for their advantages in terms of speed, effectiveness, and safety, offer a promising solution. In order to tackle the manufacturing challenges associated with RNA vaccine production, this study develops a mechanistic model that can support the prediction and analysis for the IVT process. This model combines a novel stochastic molecular reaction queueing network with a regulatory kinetic model characterizing the influence of system state (e.g., pH, the concentrations of DNA template, NTPs, magnesium, and T7 RNAP) on reaction rates and production outputs such as RNA yield and impurity levels. The empirical study shows that this model has a promising performance across different process conditions. Ultimately, it demonstrates the potential to enhance RNA product quality consistency, increase yield, and lower production costs.

## REFERENCES

- Akama, S., M. Yamamura, and T. Kigawa. 2012. "A Multiphysics Model of In Vitro Transcription Coupling Enzymatic Reaction and Precipitation Formation". *Biophysical Journal* 102(2):221–230.
- Anderson, D. F., and T. G. Kurtz. 2011. "Continuous Time Markov Chain Models for Chemical Reaction Networks". In *Design and Analysis of Biomolecular Circuits*, 3–42. Springer.

- Arnold, S., M. Siemann, K. Scharnweber, M. Werner, S. Baumann, and M. Reuss. 2001. “Kinetic Modeling and Simulation of In Vitro Transcription by Phage T7 RNA Polymerase”. *Biotechnology and Bioengineering* 72(5):548–561.
- Cheetham, G. M., D. Jeruzalmi, and T. A. Steitz. 1999. “Structural Basis for Initiation of Transcription from An RNA Polymerase–Promoter Complex”. *Nature* 399(6731):80–83.
- Clement, E. J., T. T. Schulze, G. A. Soliman, B. J. Wysocki, P. H. Davis, and T. A. Wysocki. 2020. “Stochastic Simulation of Cellular Metabolism”. *IEEE Access* 8:79734–79744.
- Dornell, J. 2022. “RNA Polymerase: Function and Definition”. *Technology Networks*. <https://www.technologynetworks.com/genomics/articles/rna-polymerase-function-and-definition-346823>, accessed 19<sup>th</sup> June 2022.
- Dousis, A., K. Ravichandran, E. M. Hobert, M. J. Moore, and A. E. Rabideau. 2023. “An Engineered T7 RNA Polymerase that Produces mRNA Free of Immunostimulatory Byproducts”. *Nature Biotechnology* 41(4):560–568.
- Gebre, M. S., L. A. Brito, L. H. Tostanoski, D. K. Edwards, A. Carfi, and D. H. Barouch. 2021. “Novel Approaches for Vaccine Development”. *Cell* 184(6):1589–1603.
- Kartje, Z. J., H. I. Janis, S. Mukhopadhyay, and K. T. Gagnon. 2021. “Revisiting T7 RNA Polymerase Transcription In Vitro with the Broccoli RNA Aptamer as A Simplified Real-time Fluorescent Reporter”. *Journal of Biological Chemistry* 296.
- Kloska, S. M., K. Pałczyński, T. Marciniak, T. Talaška, M. Miller, B. J. Wysocki, P. Davis, and T. A. Wysocki. 2022. “Queueing Theory Model of Pentose Phosphate Pathway”. *Scientific Reports* 12(1):1–9.
- Lodish, H., A. Berk, C. A. Kaiser, C. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira et al. 2013. *Molecular Cell Biology*. 7th ed. New York: Macmillan.
- Tabor, S., and C. C. Richardson. 1985. “A Bacteriophage T7 RNA Polymerase/Promoter System for Controlled Exclusive Expression of Specific Genes”. *Proceedings of the National Academy of Sciences* 82(4):1074–1078.
- Tersteeg, S., T. Mrozowich, A. Henrickson, B. Demeler, and T. R. Patel. 2022. “Purification and Characterization of Inorganic Pyrophosphatase for In Vitro RNA Transcription”. *Biochemistry and Cell Biology* 100(5):425–436.
- van de Berg, D., Z. Kis, C. F. Behmer, K. Samnuan, A. K. Blakney, C. Kontoravdi, R. Shattock, and N. Shah. 2021. “Quality by Design Modelling to Support Rapid RNA Vaccine Production against Emerging Infectious Diseases”. *NPJ Vaccines* 6(1):1–10.
- Vernon, W. 1988. “The Role of Magnesium in Nucleic-acid and Protein Metabolism”. *Magnesium* 7(5-6):234–248.
- Wang, B., W. Xie, T. Martagan, A. Akcay, and C. G. Corlu. 2019. “Stochastic Simulation Model Development for Biopharmaceutical Production Process Risk Analysis and Stability Control”. In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 1989–2000. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Wang, K., W. Xie, and S. W. Harcum. 2023. “Metabolic Regulatory Network Kinetic Modeling with Multiple Isotopic Tracers for iPSCs”. *arXiv preprint arXiv:2305.00165*. <https://arxiv.org/abs/2305.00165>, accessed 19<sup>th</sup> June 2022.
- Xie, W., R. R. Barton, B. L. Nelson, and K. Wang. 2023. “Stochastic Simulation Uncertainty Analysis to Accelerate Flexible Biomufacturing Process Development”. *European Journal of Operational Research* 310(1):238–248.
- Xie, W., B. Wang, C. Li, D. Xie, and J. Auclair. 2022. “Interpretable Biomufacturing Process Risk and Sensitivity Analyses for Quality-by-Design and Stability Control”. *Naval Research Logistics (NRL)* 69(3):461–483.
- Young, J. S., W. F. Ramirez, and R. H. Davis. 1997. “Modeling and Optimization of A Batch Process for In Vitro RNA Production”. *Biotechnology and Bioengineering* 56(2):210–220.
- Zheng, H., S. W. Harcum, J. Pei, and W. Xie. 2023. “Stochastic Biological System-of-Systems Modelling for iPSC Culture”. *arXiv preprint arXiv:2305.17787*. <https://arxiv.org/abs/2305.17787>, accessed 19<sup>th</sup> June 2022.
- Zheng, H., W. Xie, I. O. Ryzhov, and D. Xie. 2022. “Policy Optimization in Dynamic Bayesian Network Hybrid Models of Biomufacturing Processes”. *INFORMS Journal on Computing* 35(1):66–82.

## AUTHOR BIOGRAPHIES

**KEQI WANG** is Ph.D. candidate in Mechanical and Industrial Engineering (MIE) at Northeastern University. His research interests include machine learning, data analytics, and computer simulation. His email address is [wang.keq@northeastern.edu](mailto:wang.keq@northeastern.edu).

**WEI XIE** is an assistant professor in MIE at Northeastern University. Her research interests include interpretable AI/ML, computer simulation, model uncertainty, and stochastic optimization. Her email address is [w.xie@northeastern.edu](mailto:w.xie@northeastern.edu).

**HUA ZHENG** is Ph.D. candidate in MIE at Northeastern University with expertise in machine learning, computer simulation and stochastic optimization. His email address is [zheng.hua1@northeastern.edu](mailto:zheng.hua1@northeastern.edu). His website is <https://zhenghuazx.github.io/hua.zheng/>