Received: 24 August 2023 | Revised: 13 February 2024 | Accepted: 14 March 2024

DOI: 10.1002/bit.28699

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Mechanistic modeling of in vitro transcription incorporating effects of magnesium pyrophosphate crystallization

Nathan Merica Stover | Krystian Ganko | Richard D. Braatz

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Correspondence
Richard D. Braatz, Department of Chemical Engineering, Massachusetts Institute of Technology, E19-551, 77 Massachusetts Ave, Cambridge, MA 02139, USA.
Email: braatz@mit.edu

Funding information
U.S. Food and Drug Administration, Grant/Award Number: 75F40122C00200; U.S. Department of Energy; Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship, Grant/Award Number: DE-SC0022158

Abstract

The in vitro transcription (IVT) reaction used in the production of messenger RNA vaccines and therapies remains poorly quantitatively understood. Mechanistic modeling of IVT could inform reaction design, scale-up, and control. In this work, we develop a mechanistic model of IVT to include nucleation and growth of magnesium pyrophosphate crystals and subsequent agglomeration of crystals and DNA. To help generalize this model to different constructs, a novel quantitative description is included for the rate of transcription as a function of target sequence length, DNA concentration, and T7 RNA polymerase concentration. The model explains previously unexplained trends in IVT data and quantitatively predicts the effect of adding the pyrophosphatase enzyme to the reaction system. The model is validated on additional literature data showing an ability to predict transcription rates as a function of RNA sequence length.

KEYWORDS
cell-free synthesis, in vitro transcription, mechanistic modeling, mRNA production, pyrophosphatase

1 | INTRODUCTION

In recent years, messenger RNA (mRNA)-based vaccines and immunotherapies have shown clinical efficacy for COVID-19, seasonal influenza, Epstein–Barr virus, HIV, and some forms of cancer (Barbier et al., 2022). In addition, mRNA is a promising method of delivering therapeutic proteins, with several mRNA therapies in the process of early-stage clinical trials (Rohner et al., 2022). Producing mRNA vaccines at the scale needed for quickly immunizing populations, however, remains a challenge (Kis et al., 2021). Also, while mRNA therapies are targeted toward small population groups, 50–1000 times greater dosages are required than for mRNA vaccines, which adds to manufacturing costs (Barbier et al., 2022). Due to the broad reach of the mRNA platform, even modest advances in efficiency and quality control of mRNA production would have a significant impact on the availability of a wide variety of therapies. Consumption of reagents for the in vitro transcription (IVT) reaction used for RNA synthesis is a key source of cost of goods (Rosa et al., 2021). Mechanistic modeling of this biomanufacturing process can be useful to organize existing data, understand the dynamics of key processes, and design novel reaction schemes and reactors (Destro & Barolo, 2022; Hong et al., 2018).

The quantitative effect of pyrophosphatase (PPiase), an enzyme that degrades the pyrophosphate (PPi) byproduct of IVT and is heuristically included in most IVT schemes, remains unclear. While the mechanism of action and rate that PPiase catalyzes the degradation of PPi is well studied, its effect on the IVT system is poorly understood, and it remains unclear why PPi needs to be removed from the IVT system. Owing to this lack of mechanistic understanding, there remains disagreement as to whether PPiase is even useful for increasing IVT yields at all. Previous researchers have reported PPiase to both be one of the most important components in

their reaction schemes (Akama et al., 2012; Rosa et al., 2022), or to have no effect on yields (Kanwal et al., 2018; Samnuan et al., 2022). One phenomenon of interest to mechanistic modeling of IVT is the crystallization of the PPi byproduct in the form of magnesium pyrophosphate (Mg_2PPi). This process has been associated with decreased yields, but no mechanistic model has been published that can describe major trends in these data (Akama et al., 2012; Young et al., 1997).

In parallel to research on the manufacturing of RNA, recent efforts in the design of biological nanostructures for potential therapeutic and diagnostic uses have incidentally increased understanding of IVT. It has become well understood that Mg_2PPi crystallization in the presence of DNA forms Mg_2PPi–DNA composites that remove DNA from solution (Kim et al., 2019, 2017). In addition, Mg_2PPi–RNA composites have been observed and studied in conditions where crystallization of Mg_2PPi occurs in the presence of RNA (Baker et al., 2018; Shopsowitz et al., 2014). These DNA and RNA nanostructures have primarily been investigated with the goal of identifying synthesis–structure relationships for the engineering of delivery or diagnostic platforms. However, the knowledge that Mg_2PPi solid formation has this effect on both an essential reagent and the product of the reaction has profound implications for the engineering of IVT. This phenomenon has never been quantitatively modeled as part of a larger system for understanding the kinetics of IVT.

An additional unmet challenge in the mechanistic modeling of IVT is to develop generalizable models for arbitrary target RNA sequences. While predicting complex phenomena such as RNA secondary structure remains a grand challenge, incorporating simple characteristics like sequence length is a straightforward first step in developing generally applicable and easily translatable IVT models. Most previous work in applying mechanistic models to IVT data has ignored the effect of sequence length in both model development and data collection (Akama et al., 2012; van de Berg et al., 2021; Young et al., 1997), and the most complete past work in incorporating sequence length into mechanistic models of IVT was restricted by the limited data and fundamental understanding of the transcription process available at the time (Arnold et al., 2001).

In this work, a mechanistic model is developed for IVT that incorporates new quantitative descriptions of the crystallization of magnesium pyrophosphate, the sequestration of DNA due to crystal formation, and the degradation of PPi by PPiase. To generalize this model across multiple target RNA sequences of different lengths, our transcription rate law incorporates descriptions of both initiation and elongation steps. This mechanistic model is fit to a literature data set that is unique in the published literature in measuring the dynamics of PPi concentration (Akama et al., 2012). The inclusion of these new phenomena into the model enables it to capture important trends in this data set. In addition, the model quantitatively predicts the effect of adding PPiase to the IVT system. Our model for the IVT rate accurately predicts the effect of sequence length on the IVT rate measured for an independent set of experiments (Rosa et al., 2022).

2 | REVIEW OF PAST MODELS OF IVT REACTIONS

The primary process of IVT is the polymerization of RNA from four nucleoside triphosphate (NTP) monomers, which has the overall stoichiometry:

$$N_A(\mathrm{ATP}) + N_U(\mathrm{UTP}) + N_C(\mathrm{CTP}) + N_G(\mathrm{GTP}) \rightarrow +(N_{all} - 1)\mathrm{PPi} + (N_{all} - 1)\mathrm{H}, \quad (1)$$

where

$$N_{all} = N_A + N_U + N_C + N_G \quad (2)$$

and N_A, N_U, N_C, and N_G are the numbers of adenosine triphosphate, uridine triphosphate, cytidine triphosphate, guanosine triphosphate (ATP, UTP, CTP, and GTP) monomers incorporated into each RNA sequence. The reaction forms PPi and proton (H) byproducts. A typical IVT scheme requires a linearized template DNA of the target sequence, NTPs, T7 RNA polymerase (T7 RNAP), and a magnesium salt in an aqueous buffered reaction at 37°C and a pH around 7.5–8 (Beckert & Masquida, 2011). In addition, many IVT reaction schemes include PPiase, surfactants, spermadine, and dithiothreitol. However, past mechanistic models of IVT have only focused on modeling the concentrations and effects of NTPs, T7 RNAP, and Mg, and there is little to no public data describing the effect on IVT by the latter set of components.

In addition to the transcription reaction, past mechanistic models for IVT have included a number of secondary processes based on experimental observations and first principles. First, a network of equilibrium reactions between free species concentrations and complexes such as MgNTPs are described using a series of algebraic relations (Kern & Davis, 1997). In addition, past mechanistic models have included additional kinetic phenomena, including Mg_2PPi crystallization, RNA degradation, and T7 RNAP degradation (Akama et al., 2012; van de Berg et al., 2021; Young et al., 1997). While the latter two of these phenomena were introduced to help conform IVT models to individual data sets and are not directly observed in the context of IVT, Mg_2PPi crystallization is a confirmed phenomenon that is easily reproduced owing to the visibility of solid formation.

Past mechanistic models have focused on isolated operating regimes and design spaces of the IVT reaction due to the diversity of goals involved. The first mechanistic model for describing trajectories of solution concentrations in the IVT reaction was primarily focused on empirically modeling experimental data (Young et al., 1997). This work uniquely focused on modeling the presence of aborts, which are short transcription sequences that do not match the desired full sequence. A later work (Arnold et al., 2001) developed a mechanistic model of IVT with the goal of deriving rate expressions from the first principles of the known biochemistry of IVT. This work is unique in quantitatively including initiation, elongation, and termination of the RNA polymerization process into an IVT mechanistic model, and in including quantitative descriptions of the effect of DNA concentration on IVT rate. Another study (Akama et al., 2012) developed a mechanistic model to describe IVT in tandem with Mg_2PPi

crystallization. A recent study (van de Berg et al., 2021) built on two past models (Akama et al., 2012; Young et al., 1997) to fit a data set collected for a range of operating conditions.

The most comprehensive published data set on Mg_2PPi crystallization in IVT is by Akama et al. (2012), which is referred to as the *Akama data set* in this work. This data set includes the temporal evolution of both RNA and PPi concentrations, which is unique among published data sets. Despite the high quality and relevance of the Akama data set, no publications (not even Akama et al.) have fit these temporal reaction trajectories to a mechanistic model using the statistical techniques of parameter estimation. Our mechanistic model, which is described in Section 3, is fit to the Akama data set.

3 | MATERIALS AND METHODS

3.1 | Mechanistic model formulation

Our mechanistic model uses a set of differential equations:

$$\frac{d[\mathrm{DNA}]}{dt} = -V_{\mathrm{sequestration}}, \tag{3}$$

$$\frac{d[\mathrm{RNA}]}{dt} = V_{\mathrm{tr}}, \tag{4}$$

$$\frac{d[\mathrm{PPi}]_{\mathrm{tot}}}{dt} = (N_{\mathrm{all}} - 1)V_{\mathrm{tr}} - V_{\mathrm{solid}} - V_{\mathrm{PPiase}}, \tag{5}$$

$$\frac{d[\mathrm{ATP}]_{\mathrm{tot}}}{dt} = -N_{\mathrm{A}}V_{\mathrm{tr}}, \tag{6}$$

$$\frac{d[\mathrm{UTP}]_{\mathrm{tot}}}{dt} = -N_{\mathrm{U}}V_{\mathrm{tr}}, \tag{7}$$

$$\frac{d[\mathrm{CTP}]_{\mathrm{tot}}}{dt} = -N_{\mathrm{C}}V_{\mathrm{tr}}, \tag{8}$$

$$\frac{d[\mathrm{GTP}]_{\mathrm{tot}}}{dt} = -N_{\mathrm{G}}V_{\mathrm{tr}}, \tag{9}$$

$$\frac{d[\mathrm{Mg}]_{\mathrm{tot}}}{dt} = -2V_{\mathrm{solid}}, \tag{10}$$

$$\frac{d[\mathrm{Nuc}]}{dt} = V_{\mathrm{nuc}}, \tag{11}$$

$$\frac{d[\mathrm{Pi}]_{\mathrm{tot}}}{dt} = 2V_{\mathrm{PPiase}}, \tag{12}$$

which track the temporal evolution of 10 state variables representing the total concentration of species in the reaction: DNA, RNA, PPi, ATP, UTP, CTP, GTP, Mg, phosphate (Pi), and Mg_2PPi nuclei

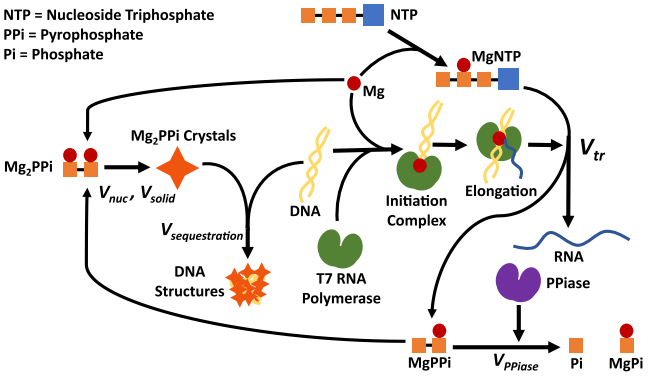

FIGURE 1 Major species in the IVT reaction model. The elongation of mRNA chains produces pyrophosphate (PPi) byproducts. This byproduct can complex with magnesium to form solid crystals, which sequester DNA, inhibiting transcription. The pyrophosphatase enzyme inhibits the formation of crystals by decreasing the free concentration of PPi. IVT, in vitro transcription; mRNA, messenger RNA.

(Figure 1). The model contains five kinetic processes: the transcription reaction (V_{tr}), nucleation and growth of Mg_2PPi crystals (V_{nuc}, V_{solid}), agglomeration of Mg_2PPi nuclei and DNA ($V_{\mathrm{sequestration}}$), and the degradation of PPi by PPiase (V_{PPiase}).

The process of transcription was modeled using a quasisteady-state assumption as the time constants associated with the transcription of a single transcript (3–30 s) are substantially lower than the time constant of substrate consumption in the data used in this work (0.25–0.5 h) (Koh et al., 2018; Tang et al., 2011). The transcription rate was modeled as a process of reversible binding of T7 RNAP (P) and DNA promoter (DNA), coupled with an irreversible initiation step and an elongation step dependent on the number of each base in the sequence (SI Section 1). The overall rate of transcription is equivalent to the rate of chain initiation, which is modeled as first order in the concentration of polymerase–DNA initiation complex (P · DNA)

$$V_{\mathrm{tr}} = k_{\mathrm{i}}[\mathrm{P} \cdot \mathrm{DNA}], \tag{13}$$

where k_{i} is an initiation rate constant. In addition to initiation, an elongation step is required for the formation of RNA. Without loss of generality, the effective rate constant for an elongating RNAP to incorporate an ATP is given as

$$k_{\mathrm{A}} = \frac{k_{\mathrm{e}}[\mathrm{MgATP}]}{K_1\left(1 + \frac{[\mathrm{MgPPi}]}{K_{\mathrm{i,PPi}}}\right) + [\mathrm{MgATP}]} \frac{[\mathrm{Mg}]}{K_2 + [\mathrm{Mg}]}. \tag{14}$$

This formulation is based on a dual Michaelis–Menten structure that has proved useful in previous work (Akama et al., 2012) and includes a term previously used in the literature describing the competitive inhibition of nucleoside addition by PPi (Arnold et al., 2001). Using a quasisteady assumption, the concentration of initiation complex is given as (SI Section 1)

$$[P \cdot DNA] = \frac{[P] + \alpha[DNA] + K_{MD} - \sqrt{([P] + \alpha[DNA] + K_{MD})^2 - 4\alpha[P][DNA]}}{2\alpha}, \quad (15)$$

where

$$\alpha = 1 + k_i\left(\frac{N_A}{k_A} + \frac{N_U}{k_U} + \frac{N_C}{k_C} + \frac{N_G}{k_G}\right), \quad K_{MD} = \frac{k_i + k_{off}}{k_{on}}, \quad (16)$$

and k_{on} and k_{off} are rate constants for T7 RNAP and the DNA promoter binding and unbinding, respectively (SI Section 2). Whereas Akama et al. (2012) modeled the rate of Mg_2PPi solid formation using empirical induction time models, we use classical nucleation theory (Myerson et al., 2019):

$$\hat{V}_{nuc} = \begin{Bmatrix} \hat{k}_{nuc}\exp\left(\frac{-\hat{B}}{\ln^2 S}\right) & \text{for } S > 1, \\ 0 & \text{for } S \le 1, \end{Bmatrix} \quad (17)$$

where S is the supersaturation,

$$S = \frac{[Mg_2PPi]}{[Mg_2PPi]_{eq}}, \quad (18)$$

and $\hat{B}$ and $\hat{k}_{nuc}$ are the dimensionless free energy barrier to nucleation and the nucleation rate constant, respectively. The total rate of a solid formation of PPi (in mol L^{-1} h^{-1}) is modeled in keeping with previous work as (Peng et al., 2015)

$$\hat{V}_{solid} = \begin{Bmatrix} \hat{k}_g[\hat{Nuc}]\ln S & \text{for } S > 1, \\ 0 & \text{for } S \le 1, \end{Bmatrix} \quad (19)$$

where $\hat{k}_g$ is a rate constant governing the growth of nuclei as a function of nuclei concentration, $[\hat{Nuc}]$. The number of model parameters can be reduced, by some algebraic manipulation (see SI Section 3), to give

$$V_{nuc} = \begin{Bmatrix} \exp\left(\frac{-B}{\ln^2 S}\right) & \text{for } S > 1, \\ 0 & \text{for } S \le 1, \end{Bmatrix} \quad (20)$$

$$V_{solid} = \begin{Bmatrix} k_{growth}[Nuc]\ln S & \text{for } S > 1, \\ 0 & \text{for } S \le 1, \end{Bmatrix} \quad (21)$$

where k_{growth} and B are the fitted parameters and [Nuc] is the rate-normalized concentration of nuclei. In addition, based on qualitative work demonstrating the agglomeration of Mg_2PPi crystals and DNA (Kim et al., 2019), a term was included to describe the rate of DNA sequestration in the solid phase:

$$V_{sequestration} = k_d[DNA][Nuc], \quad (22)$$

which hypothesizes that the rate is first order in both DNA concentration and rate-normalized nuclei concentration with a rate constant k_d. Past work has qualitatively shown that a similar phenomenon takes place in sequestering RNA (Shopsowitz et al., 2014). However, the experimental procedure used by Akama et al. (2012) redissolved any solid before measuring RNA concentrations, meaning that any RNA sequestration cannot be observed from the Akama data set.

Enzymatic degradation of PPi is modeled by

$$V_{PPiase} = k_{PPiase}[PPiase]\frac{[MgPPi]}{[MgPPi] + K_{M,PPiase}}, \quad (23)$$

with the rate law and parameters from a kinetic study of PPiase (Chao et al., 2006), where [PPiase] is in units of volume-based enzyme activity (U μL^{-1}).

The above rates are dependent on the concentration of complexes, such as MgATP and Mg_2PPi. The concentrations of these complexes over time are modeled by a set of algebraic equations that describe known equilibrium relations and material balances of the system. In this work, it is assumed that the equilibrium constants associated with all NTPs are the same, and NTPs are treated as a lumped state for the purpose of thermodynamic calculations, defining a total NTP concentration,

$$[NTP]_{tot} = [ATP]_{tot} + [UTP]_{tot} + [CTP]_{tot} + [GTP]_{tot}. \quad (24)$$

The material balances for the ionic species are

$$[Mg]_{tot} = [Mg] + [MgPPi] + [HMgPPi] + [MgNTP] + 2[Mg_2NTP] + 2[Mg_2PPi] + [HMgNTP] + [H_2MgPPi] + [MgPi], \quad (25)$$

$$[NTP]_{tot} = [NTP] + [HNTP] + [HMgNTP] + [MgNTP] + [Mg_2NTP], \quad (26)$$

$$[PPi]_{tot} = [PPi] + [MgPPi] + [Mg_2PPi] + [HPPi] + [HMgPPi] + [H_2PPi] + [H_2MgPPi], \quad (27)$$

$$[Buffer]_{tot} = [Buffer] + [HBuffer], \quad (28)$$

$$[Pi]_{tot} = [Pi] + [MgPi], \quad (29)$$

where the complex concentrations are defined by the equilibrium relations

$$[HNTP] = [H][NTP]K_{HNTP}, \quad (30)$$

$$[HMgNTP] = [HNTP][Mg]K_{HMgNTP}, \quad (31)$$

$$[HPPi] = [H][PPi]K_{HPPi}, \quad (32)$$

$$[HMgPPi] = [HPPi][Mg]K_{HMgPPi}, \quad (33)$$

$$[H_2PPi] = [HPPi][H]K_{H_2PPi}, \quad (34)$$

$$[H_2MgPPi] = [H_2PPi][Mg]K_{H_2MgPPi}, \quad (35)$$

$$[HBuffer] = [H][Buffer]K_{Buffer}, \quad (36)$$

$$[\mathrm{MgNTP}] = [\mathrm{Mg}][\mathrm{NTP}]K_{\mathrm{MgNTP}}, \tag{37}$$

$$[\mathrm{Mg_2NTP}] = [\mathrm{MgNTP}][\mathrm{Mg}]K_{\mathrm{Mg_2NTP}}, \tag{38}$$

$$[\mathrm{MgPPi}] = [\mathrm{Mg}][\mathrm{PPi}]K_{\mathrm{MgPPi}}, \tag{39}$$

$$[\mathrm{Mg_2PPi}] = [\mathrm{MgPPi}][\mathrm{Mg}]K_{\mathrm{Mg_2PPi}}, \tag{40}$$

$$[\mathrm{MgPi}] = [\mathrm{Mg}][\mathrm{Pi}]K_{\mathrm{MgPi}}. \tag{41}$$

Finally, the proton concentration is determined by a charge balance. While all species in the reaction contribute to the charge balance in theory, limited reporting of the exact counterions used in reaction feedstocks makes exact accounting of charges infeasible. The buffering salts are approximated to be the primary components.

$$[\mathrm{H}] + [\mathrm{HBuffer}] = [\mathrm{OH}] + [\mathrm{Cl}], \tag{42}$$

where

$$[\mathrm{OH}] = \frac{10^{-14}}{[\mathrm{H}]} \tag{43}$$

and [Cl] is the initial concentration of Cl added via the HCl–Tris buffer used for the reaction (SI Section 11). For calculation of transcription kinetics, without loss of generality, the concentration of [MgATP] is calculated as

$$[\mathrm{MgATP}] = [\mathrm{MgNTP}]\frac{[\mathrm{ATP}]_{\mathrm{tot}}}{[\mathrm{NTP}]_{\mathrm{tot}}}. \tag{44}$$

As for all first-principles models of complex reactions, some assumptions and simplifications are made: (1) The elongation rate law (14) ignores the effect of noncoding and coding sequence identity. (2) The transcription rate model ignores interactions between RNAP molecules and pausing of transcription. (3) Product quality variables, such as the presence of aborts and double-stranded RNA, are not considered in this work, as the literature data describing these byproducts are sparse. (4) Degradation of RNA and T7 RNAP are not considered as those effects were not essential for capturing the dynamics of the Akama data set. (5) The Michaelis–Menten description of PPiase action is a simplification of a more sophisticated network of reversible and irreversible reactions (Halonen et al., 2002; Tammenkoski et al., 2007). In addition, the rate law used in this work has only been shown for PPiase from *Helicobacter pylori*, which is not commonly used in IVT (Chao et al., 2006). However, as the PPi concentrations in the Akama data are relatively high, the most important part of this rate law is the maximum rate, which is quantitatively well understood and captured by such a simple model. (6) The nucleation-growth model does not take into account the effects of size heterogeneity of Mg_2PPi crystals and the contribution to total solid formation of the nuclei formation step. (7) Crystallization of magnesium hydrogen phosphate (MgHPi), which has been postulated as an additional process in the IVT system (Kim et al., 2019), is not considered.

3.2 | Computational methods

Model evaluation and parameter optimization are performed in the Julia language. The set of equations in the preceding section is combined into a system of differential algebraic equations that are solved forward in time using the high-order integrators available in the DifferentialEquations.jl package. Experimental measurements are assumed to have additive, uncorrelated measurement errors with a normal distribution of zero mean and diagonal measurement error covariance matrix V_y. Parameter estimation is performed in $\log_{10}$ space, to search the large numerical space and to best represent the prior distribution of parameters, where the vector k represents the $\log_{10}$ of the parameters. The prior distribution for k is assumed to follow a normal distribution with mean μ and covariance V_μ, which is equivalent to assuming a log-normal distribution of parameters.

MAP estimation of the vector k was carried out:

$$\min_k (y - u(k))^{\top} V_y^{-1}(k)(y - u(k)) + (k - \mu)^{\top} V_\mu^{-1}(k - \mu), \tag{45}$$

where y is the vector containing all of the experimental data used for estimating parameters and $u(k)$ is the vector of corresponding model outputs as a function of the $\log_{10}$ parameter vector k. The error covariance matrix V_{k*} of the best-fit estimate k^* is approximated by (Beck & Arnold, 1977)

$$\mathrm{cov}(k^* - k_{\mathrm{true}}) = V_{k^*} \approx \left(S^{\top} V_y^{-1} S + V_\mu\right)^{-1}, \tag{46}$$

where k_{true} denotes the true $\log_{10}$ of the parameters and S is the sensitivity of the model outputs with respect to the vector k. Additional details on the parameter estimation strategy can be found in SI Section 5.

Local gradient-based optimization is carried out with L-BFGS optimization using the ForwardDiff.jl and NLopt.jl packages in Julia to compute model output sensitivities to parameters and use those sensitivities in gradient-based optimizers, respectively (Liu & Nocedal, 1989). Multistart optimization using 4000 random starting points is performed to search for a global optimum (Martí, 2003). Best-fit parameter estimates in k^* are given in Table 1, and the parameter error covariance matrix V_{k*} is given in SI Section 10.

4 | RESULTS

4.1 | Fitting model to calibration data

The batch IVT reaction model is fit to the Akama data set, which consists of three parts (Akama et al., 2012). The first, and primary, source of data is a set of temporal trajectories of RNA and PPi concentrations for 13 different sets of Mg, NTP, and T7 RNAP input concentrations, each recorded at 9 timepoints (Figure 2). In addition, a set of data depicting the initial rate of RNA synthesis for 20 different sets of initial Mg and NTP concentrations was collected

TABLE 1 Model parameters.

Parameter	Units	Process	Prior value	Value after fitting
k_i	h^{-1}	Transcription initiation prior	$10^{2.97\pm0.6}$ (Koh et al., 2018)	$10^{3.61\pm0.06}$
k_e	h^{-1}	Transcription elongation prior	$10^{5.72\pm0.3}$ (Tang et al., 2011)	$10^{5.20\pm0.11}$
k_{off}	h^{-1}	T7 RNAP–DNA binding prior	$10^{3.64\pm0.5}$ (Koh et al., 2018)	$10^{3.74\pm0.43}$
k_{on}	$h^{-1}nM^{-1}$	T7 RNAP–DNA binding prior	$10^{2.31\pm0.1}$ (Koh et al., 2018)	$10^{2.30\pm0.10}$
K_1	M	Transcription MgNTP dependence	–	$10^{-3.59\pm0.18}$
K_2	M	Transcription Mg dependence	–	$10^{-3.59\pm0.20}$
$K_{i,PPi}$	M	Transcription PPi inhibition prior	$10^{-3.70\pm0.6}$ (Arnold et al., 2001)	$10^{-4.38\pm0.17}$
k_{growth}	moles h^{-1}	Mg_2PPi solid growth	–	$10^{0.59\pm0.30}$
B	arb. unit	Mg_2PPi solid growth prior	$10^{1.13\pm0.2}$ (Akama et al., 2012)	$10^{1.65\pm0.09}$
k_d	$h^{-1}M^{-1}$	DNA-Mg_2PPi agglomeration	–	$10^{4.99\pm0.34}$
K_{HNTP}	M^{-1}	Ion equilibrium prior	$10^{6.91\pm1.16}$	$10^{6.91\pm0.04}$
K_{HMgNTP}	M^{-1}	Ion equilibrium prior	$10^{2.08\pm1.16}$	$10^{2.08\pm0.10}$
K_{HPPi}	M^{-1}	Ion equilibrium prior	$10^{9.02\pm1.16}$	$10^{9.02\pm0.10}$
K_{HMgPPi}	M^{-1}	Ion equilibrium prior	$10^{3.32\pm1.16}$	$10^{3.32\pm0.10}$
K_{H_2PPi}	M^{-1}	Ion equilibrium prior	$10^{6.26\pm1.16}$	$10^{6.26\pm0.20}$
K_{H_2MgPPi}	M^{-1}	Ion equilibrium prior	$10^{2.11\pm1.16}$	$10^{2.11\pm0.20}$
K_{MgNTP}	M^{-1}	Ion equilibrium prior	$10^{4.54\pm1.16}$	$10^{4.10\pm0.14}$
K_{Mg_2NTP}	M^{-1}	Ion equilibrium prior	$10^{1.77\pm1.16}$	$10^{1.87\pm0.29}$
K_{MgPPi}	M^{-1}	Ion equilibrium prior	$10^{4.80\pm1.16}$	$10^{5.16\pm0.15}$
K_{Mg_2PPi}	M^{-1}	Ion equilibrium prior	$10^{2.57\pm1.16}$	$10^{3.99\pm0.11}$
K_{MgPi}	M^{-1}	Ion equilibrium	$10^{1.88}$	$10^{1.88}$
Mg_2PPi_{eq}	M	Ion equilibrium prior	$10^{-4.85\pm2}$ (Akama et al., 2012)	$10^{-3.89\pm0.07}$
k_{PPiase}	$Mh^{-1}(U/\mu L)^{-1}$	Degradation of PPi	60 (SI Section 7)	60
$K_{M,PPiase}$	M	Degradation of PPi	2.14×10^{-4} (Chao et al., 2006)	2.14×10^{-4}

Note: Strategy and sources for generating prior values of equilibrium constants are discussed in SI Section 6. Error on parameter priors represents a 95% confidence interval using standard deviation estimated from the literature. Error on parameter posteriors represents the 95% pointwise confidence intervals as approximated by drawing samples from the probability distribution defined by the parameter error covariance matrix. The unit U represents enzyme activity units, discussed in SI section 7.

Abbreviations: NTP, nucleoside triphosphate; PPi, pyrophosphate; RNAP, RNA polymerase.

(Figure 3a). Finally, Akama et al. (2012) conducted a set of experiments solely to assess the solubility of magnesium in the presence of NTP and PPi in the absence of transcription (Figure 3b). This data set contains 50 different sets of initial NTP and PPi inputs. The PPiase enzyme is not present in any of the calibration data used in this work.

The model features 24 parameters, which are fixed based on literature data, estimated from the calibration data using a Bayesian prior from the literature, or estimated from the calibration data without a prior (SI Section 5). The model demonstrates an ability to describe the trends of the calibration data set. PPi, a byproduct of transcription, initially grows rapidly as the transcription reaction progresses, but reaches a peak in concentration after which the PPi concentration decays rapidly due to competing Mg_2PPi solid formation (Figure 2a). Mg_2PPi solid formation is prevented in cases of high NTP and low Mg concentration (Figure 2b,c). In cases of solid formation, the reaction halts before reaching full conversion of NTPs (Figure 2d–f). At low NTP concentrations, increasing NTP concentration increases initial transcription rates. However, at higher NTP concentrations, this effect reverses as the addition of NTP decreases free Mg concentrations (Figure 3a). When small amounts of PPi (less than one equivalent of Mg) are added to aqueous Mg in the absence of the transcription reaction, Mg_2PPi solid formation decreases the concentration of Mg in solution after 24 h (Figure 3b). However, greater PPi input up to two equivalents of Mg decreases the amount of solid precipitate formed as the system is pushed toward the

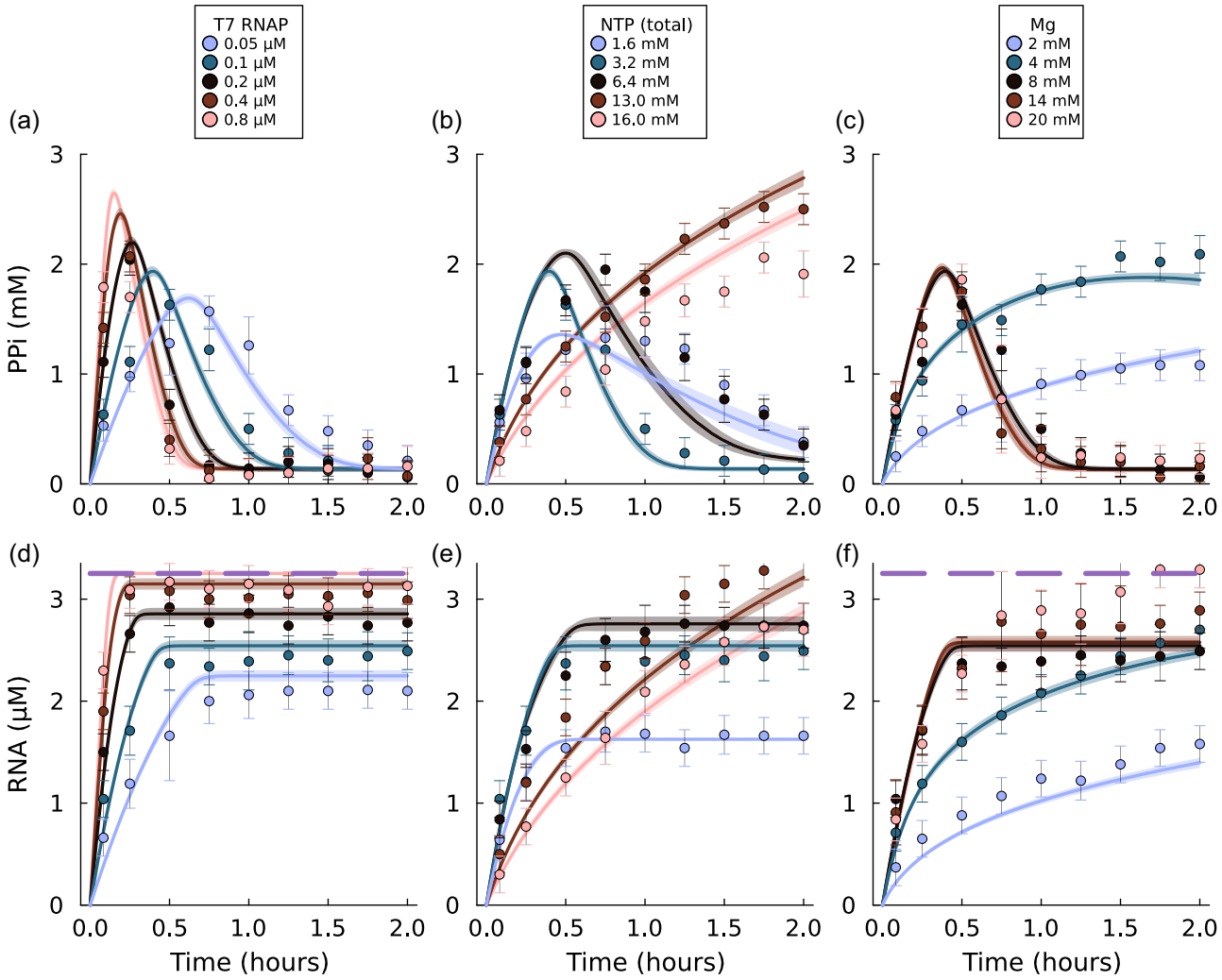

FIGURE 2 Model fitting results compared with fitting data set of dynamic concentration trajectories. Temporal trajectories of PPi (a–c) and RNA (d–f) concentration as a function of changing T7 RNA polymerase (a, d), NTP (b, e), and Mg (c, f) input concentrations. Purple dashed lines represent maximum possible mRNA yield based on stoicheometry (not shown in plot E as changing NTP input results in a range of maximum yields). Shaded areas about each model prediction are the 95% prediction interval (SI Section 4). NTP, nucleoside triphosphate; PPi, pyrophosphate; RNAP, RNA polymerase.

soluble MgPPi complex. The magnitude of precipitation is decreased upon the addition of NTP.

4.2 | Predicting effect of PPiase on the IVT system

The above calibration data are for experiments that did not include PPiase. In addition to these data, Akama et al. (2012) generated a small data set describing the effect of PPiase on reaction yields, which is used for model validation in this work (Figure 4). Akama and coworkers did not report the quantity of PPiase added but showed that when PPiase was used, the concentration of PPi was indistinguishable from zero over the course of the reaction.

Even though a new species is added to the reaction that is not present in the calibration data set, it is possible to use these data to validate the model as parameters describing the kinetics of PPiase activity are considered to be fixed from the literature (Table 1), and because the excess use of PPiase in this context renders the exact kinetics parameters of PPiase unimportant. PPiase addition was set to an excess value of 1 U μL^{-1} in the model to predict these results. Our model predictions for the effect of PPiase on the IVT reaction are within the experimental error bars (Figure 4).

4.3 | Predicting effect of sequence length

Rosa et al. (2022) collected trajectories of IVT yields for a set of three DNA constructs varying in length between 1195 and 5299 nucleotides (Figure 5), with excess (4 U mL^{-1}) PPi in the reaction. Our model is able to predict the dependency of the transcription rate on sequence length. This data set was not used in fitting the model parameters.

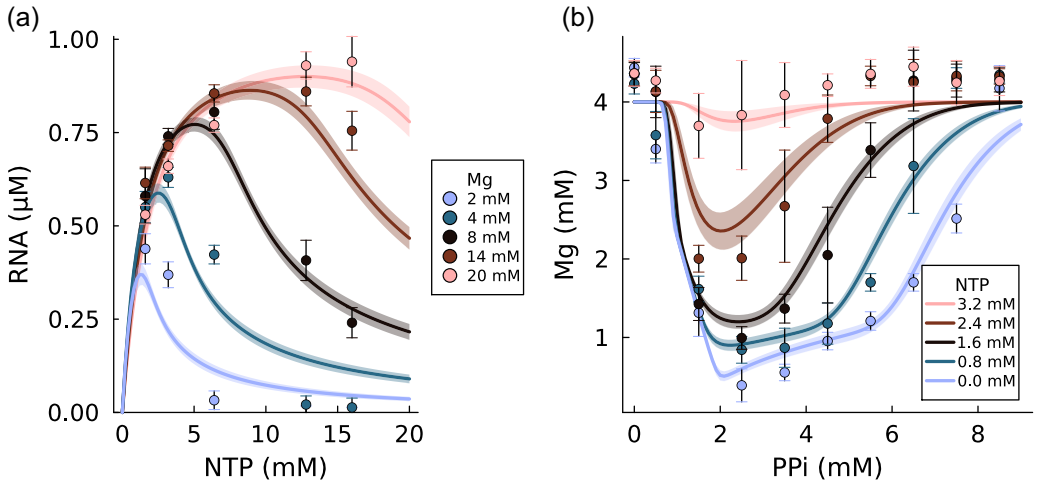

FIGURE 3 Model fitting results compared with a fitting data set of initial reaction rates and Mg_2PPi solubility. Initial transcription rates as a function of NTP and Mg input concentrations represented by RNA yields after 5 min of reaction (a). Magnesium (initially 4 mM) remaining in solution after 24 h as a function of input PPi and NTP input concentrations in the absence of the transcription reaction (b). Shaded areas about each model prediction are the 95% prediction interval (SI Section 4). NTP, nucleoside triphosphate; PPi, pyrophosphate.

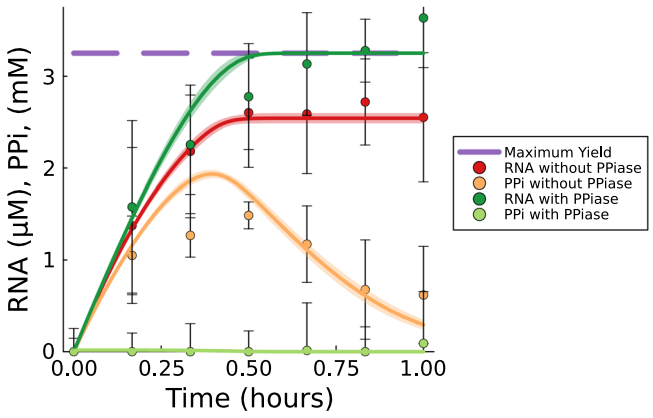

FIGURE 4 Model validation: effect of pyrophosphatase (PPiase). Model predictions are compared with the experimental results of Akama et al. (2012), showing the yields of otherwise identical reaction conditions with and without PPiase. Shaded areas about the model predictions are 95% prediction interval (SI Section 4). The error bar on each experimental data point is the 95% confidence interval based on the t-distribution of points taken in triplicate. PPi, pyrophosphate.

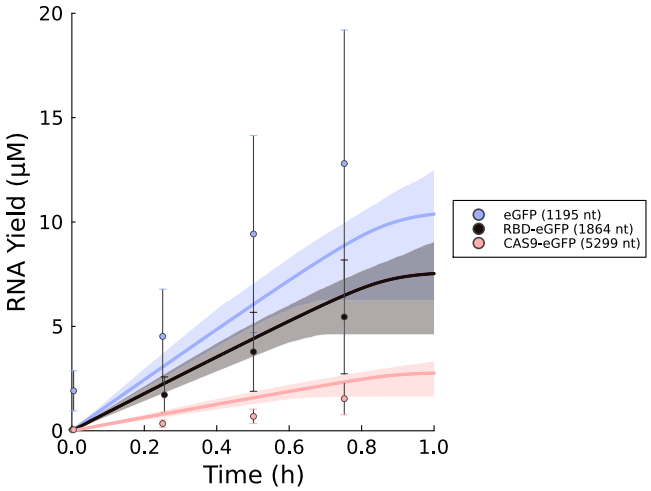

FIGURE 5 Representing various combinations of the extended green fluorescent protien (eGFP), SARS-CoV-2 spike protein receptor binding domain (RBD), and CAS9 Model validation: IVT yields of multiple DNA sequences of varying lengths (Rosa et al., 2022). Shaded areas about the model predictions are 95% prediction intervals (SI Section 4). The error bars on the data points are 95% confidence intervals based on the standard deviation estimated from the entire data set of Rosa et al. (2022). IVT, in vitro transcription. DNA sequences used in Rosa et al. are combinations of the extended green fluorescent protien (eGFP), SARS-CoV-2 spike protein receptor binding domain (RBD), and CAS9 genes.

5 | DISCUSSION

The primary result of this work is that adding a nucleation-growth model for Mg_2PPi crystallization, as well as a quantitative term describing the first-order agglomeration of DNA and Mg_2PPi nuclei, is a sufficient addition to past models to describe trends in experimental data. Our modeling of the additional phenomena is consistent with the qualitatively understood physics of the IVT system. This mechanism can additionally predict the effect of adding the PPiase enzyme on IVT yields, and is the first mechanistic model to do so. Validating this model on recent data demonstrates an ability to predict IVT rates across a range of input conditions and sequence lengths.

This work identifies Mg_2PPi solid formation as an important failure mode that is highly nonlinear, and the value of our model is its ability to predict these nonlinear effects (Figure 2). As such, our model is a suitable foundation for the development of model-based optimal design and control strategies. In addition, the incorporation

of sequence length into model predictions is a first step for the development of models that are easily adaptable to arbitrary RNA sequences.

5.1 | Mg_2PPi solid formation is a crucial element for describing IVT

Past experimental studies have shown that DNA agglomerates with Mg_2PPi nuclei during IVT, removing DNA from solution (Wang et al., 2019). The inclusion in the model of terms describing this sequestration of DNA (22) is able to describe trends in the calibration data, especially the early halting of reactions that do not go to full conversion (Figure 2). When paired with ion equilibrium laws describing the known thermodynamics of the system, this model is able to describe the conditions at which solid formation, and therefore reaction halting, occurs. At high NTP concentrations, NTP competes with PPi for the Mg ion, which leads to decreased Mg_2PPi solution concentrations and decreased solid formation (Figure 3b). This in turn prevents early halting of the reaction (Figure 2b,e). By the same mechanism, solid formation and reaction halting are prevented at low Mg concentrations (Figure 2c,f). The model also has the ability to describe the competition between the irreversible kinetic processes of transcription and DNA sequestration. Increased T7 RNAP concentration increases the initial rate of reaction (Figure 2d); while this causes solid formation to initiate earlier (Figure 2a), it ultimately leads to higher yields.

This model clears up some misconceptions and misprescriptions in the academic literature. Past studies in the mechanistic modeling of IVT have noted that the onset of crystallization is associated with a decrease or complete stoppage in transcription rates (Akama et al., 2012; Kern & Davis, 1997), but have attributed this stoppage to the onset of Mg_2PPi crystallization causing a decrease in magnesium concentration in the solution. Our analysis shows that this pathway cannot describe trends in data on its own (SI Section 9). Our model uses an entirely different explanatory pathway to describe this phenomenon.

While this development may seem like an academic distinction, the true cause of early halting is highly relevant for IVT process development and control. The general message from past work that has been transmitted to practitioners is that because Mg_2PPi solid formation decreases the solution concentration of Mg, reactions should be designed with a high concentration of Mg to preempt this effect. The literature contains many reports in the last 3 years of academic researchers justifying IVT reaction schemes and explaining results based on this idea (Pregeljc et al., 2023; Rosa et al., 2022; Samnuan et al., 2022). While the higher-order effects of magnesium on the IVT system remain poorly understood, one conclusion from this work is that the decrease in free Mg concentration due to Mg_2PPi solid formation cannot solely describe the early stopping of IVT as measured by Akama et al. (2012).

5.1.1 | Model describes mechanism of action of PPiase

The PPiase enzyme, which degrades PPi, is commonly added to IVT reaction schemes on a heuristic basis. The data set used to fit our model (Figure 2) did not include the use of the PPiase enzyme. Data from Akama et al. (2012) describing the effect of adding PPiase on IVT yields was used for model validation (Figure 4). When PPiase was added to the reaction, PPi was degraded to phosphate, preventing solid formation and sequestration of DNA as well as competitive inhibition of the transcription process. The quantitative predictions of the model that PPiase extends the length of the reaction without changing initial rates are consistent with the observed data (Figure 4).

As described in Section 4.2, the addition of PPiase to an IVT reaction system can lead to increases in reaction yield, depending on system inputs. However, despite the widespread adoption of PPiase based on heuristic observations, researchers provide conflicting explanations for the importance of both Mg_2PPi crystallization and PPiase to the IVT system, and many "rational" attempts at IVT optimization start by removing PPiase (Kanwal et al., 2018; Samnuan et al., 2022). We hypothesize that a key reason for disagreement in the literature is due to the input dependence of the crystallization process. Mg_2PPi is a key yield-limiting process, but only in a select set of regimes. Experiments performed in regimes in which crystallization does not occur will report yield to be insensitive to PPiase input. The model developed in this work represents a first step toward a unified understanding of the behavior of regimes sensitive and insensitive to PPiase.

5.2 | Mechanistic model predicts effect of sequence length on transcription rates

In addition to the small Akama data set describing the effect of PPiase (Figure 4), our model was validated on data from Rosa et al. (2022), which modulated sequence length as an independent variable. This data set is outside of the input and output range of the calibration data set used in this work (Table 2). Considering that it is heuristically understood amongst practitioners that the parameters of the IVT reaction are sequence dependent as well as the uncertainty associated with these data, we do not argue that our model is correct by virtue of correctly predicting the reaction rates of these experiments. These results should primarily be viewed as an evaluation of the model's ability to predict the general trend of the effect of sequence length on transcription rates.

These predictions may seem trivial, in the sense that they predict that sequence length has roughly no effect on the mass-based transcription rate of the IVT system. However, most previous models of IVT rely on the assumption of initiation–limitation and would predict that the initial rates of the three curves in Figure 5 should be identical (Akama et al., 2012; van de Berg et al., 2021; Young et al., 1997). The only past work on developing expressions for the transcription rate that attempts to model the effect of sequence

TABLE 2 Ranges of inputs and outputs in data explored in this work.

Author	T7 RNAP (nM)	DNA (nM)	NTP (mM)	Mg (mM)	PPiase (U mL^{-1})	Sequence length (nt)	RNA output (μM)	RNA output (g L^{-1})
Akama et al. (2012)	50–100	7.4	3.2–12.8	8–20	0	868	2–3.5	0.65–1.3
Rosa et al. (2022)	124	90	31	50	4	1195–5299	5–20	8–14

Note: The Akama data set was used for model fitting (Figure 2), with the exception of PPiase addition data that was used for model validation (Figure 4). The Rosa data (Figure 5) are used for model validation.

length was carried out using a relative paucity of data and concluded that the IVT reaction is primarily initiation-limited (Arnold et al., 2001). As such, the formulation presented in this work and demonstrated in Figure 5 is a break with past modeling conventions and a framework for the future development of IVT models.

5.3 | Limitations and directions for model improvement

The mechanism in our model that is most poorly understood in the literature is the physics by which Mg_2PPi solid formation inhibits the forward progress of the reaction. The hypothesis presented in this work—that Mg_2PPi nuclei agglomerate with DNA and decrease solution DNA concentration—is the best available explanatory mechanism based on findings of past qualitative work (Kim et al., 2019) and the consistency of quantitative predictions with data. However, the mechanistic understanding does not currently exist to rule out interactions of Mg_2PPi nuclei and other biomolecules, such as T7 RNAP, as an alternative cause. In addition, the two-step hypothesis presented in this work—that crystals nucleate and subsequently agglomerate with DNA—is currently indistinguishable from a mechanism by which crystals directly nucleate on DNA. In addition, a more developed understanding of how solution conditions affect the inhibition process is needed to accurately extend these results to different regimes. In the Akama data set, increasing magnesium concentrations at already high Mg concentrations (from 8 to 20 mM) has a nonnegligible effect on final yields without affecting the initial rate (Figure 2c). The model does not have the ability to describe this phenomenon, which is possibly due to Mg modulating the rate of sequestration. While this behavior is a limited component of the calibration data used in this work, capturing that effect would be needed for describing regimes of high Mg concentration that experience Mg_2PPi solid formation.

RNAP and DNA templates are two of the most costly components required for the IVT reaction, and understanding the highly nonlinear interactions of these components is essential for the development of a generalizable model. For example, the elongation of RNAP particles, approximated as a linear process in this work, is known to feature pauses (Janissen et al., 2022). In addition, exclusion between RNAP particles can decrease transcription rates due to crowding and is commonly described using totally asymmetric simple exclusion (TASEP) models (Wang et al., 2014). The effect of transcriptional pausing can amplify the effects of particle exclusion. Limitations due to polymerase exclusion should have an effect on the regime of high RNAP and low DNA concentrations.

6 | CONCLUSION

Process development of the IVT reaction continues to rely on the limited capabilities of heuristics-based design-of-experiments and data-driven modeling methods. Mechanistic models for IVT stand to provide rational and interpretable predictions of RNA yields outside of previously tested design spaces. In this work, we synthesized the first mechanistic model to feature an interpretable description of IVT alongside magnesium pyrophosphate crystallization, DNA-Mg_2PPi agglomeration, and PPiase enzyme activity. This model successfully describes trends observed in IVT experimental data, many of which lead to critically low RNA yields for previously unexplained reasons. Given that the IVT reaction is a foundational component for the manufacturing of a diverse and growing set of modern therapeutics, this model has the potential to provide insights for a variety of biomanufacturing systems.

AUTHOR CONTRIBUTIONS

Nathan Merica Stover: Conceptualization, methodology, software, validation, verification, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, and visualization. **Krystian Ganko**: Methodology, writing—review and editing. **Richard D. Braatz**: Conceptualization, methodology, resources, writing—review and editing, supervision, project administration, and funding acquisition.

ACKNOWLEDGMENTS

This research was supported by the U.S. Food and Drug Administration under the FDA BAA-22-00123 program, Award Number 75F40122C00200. Krystian Ganko was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0022158. The authors would like to thank Dr. Vivekananda Bal, Dr. Francesco Destro, and Christopher Canova for advice on methodology.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All software and data will be made available upon publication of the manuscript. All software and data used in this work, as well as additional tutorials for the use of this model, can be found at github.com/nathanmstover/IVTmechanisticmodel.

ORCID

Richard D. Braatz http://orcid.org/0000-0003-4304-3484

REFERENCES

Akama, S., Yamamura, M., & Kigawa, T. (2012). A multiphysics model of in vitro transcription coupling enzymatic reaction and precipitation formation. *Biophysical Journal*, *102*(2), 221–230. https://doi.org/10.1016/j.bpj.2011.12.014

Arnold, S., Siemann, M., Scharnweber, K., Werner, M., Baumann, S., & Reuss, M. (2001). Kinetic modeling and simulation of in vitro transcription by phage t7 RNA polymerase. *Biotechnology and Bioengineering*, *72*(5), 548-561.

Baker, Y. R., Chen, J., Brown, J., El-Sagheer, A. H., Wiseman, P., Johnson, E., Goddard, P., & Brown, T. (2018). Preparation and characterization of manganese, cobalt and zinc DNA nanoflowers with tuneable morphology, DNA content and size. *Nucleic Acids Research*, *46*(15), 7495–7505. https://doi.org/10.1093/nar/gky630

Barbier, A. J., Jiang, A. Y., Zhang, P., Wooster, R., & Anderson, D. G. (2022). The clinical progress of mRNA vaccines and immunotherapies. *Nature Biotechnology*, *40*(6), 840–854. https://doi.org/10.1038/s41587-022-01294-2

Beck, J. V., & Arnold, K. J. (1977). Parameter estimation in engineering and science. James Beck.

Beckert, B., & Masquida, B. (2011). Synthesis of RNA by in vitro transcription. In H. Nielsen (Ed.), *RNA: Methods and protocols*, *methods in molecular biology* (pp. 29–41). Humana Press. https://doi.org/10.1007/978-1-59745-248-9_3

Chao, T.-C., Huang, H., Tsai, J.-Y., Huang, C.-Y., & Sun, Y.-J. (2006). Kinetic and structural properties of inorganic pyrophosphatase from the pathogenic bacterium *Helicobacter pylori*. *Proteins: Structure, Function, and Bioinformatics*, *65*(3), 670–680. https://doi.org/10.1002/prot.21093

Destro, F., & Barolo, M. (2022). A review on the modernization of pharmaceutical development and manufacturing—Trends, perspectives, and the role of mathematical modeling. *International Journal of Pharmaceutics*, *620*, 121715. https://doi.org/10.1016/j.ijpharm.2022.121715

Halonen, P., Baykov, A. A., Goldman, A., Lahti, R., & Cooperman, B. S. (2002). Single-turnover kinetics of *Saccharomyces cerevisiae* inorganic pyrophosphatase. *Biochemistry*, *41*(40), 12025–12031. https://doi.org/10.1021/bi026018z

Hong, M. S., Severson, K. A., Jiang, M., Lu, A. E., Love, J. C., & Braatz, R. D. (2018). Challenges and opportunities in biopharmaceutical manufacturing control. *Computers & Chemical Engineering*, *110*, 106–114. https://doi.org/10.1016/j.compchemeng.2017.12.007

Janissen, R., Eslami-Mossallam, B., Artsimovitch, I., Depken, M., & Dekker, N. H. (2022). High-throughput single-molecule experiments reveal heterogeneity, state switching, and three interconnected pause states in transcription. *Cell Reports*, *39*(4), 110749. https://doi.org/10.1016/j.celrep.2022.110749

Kanwal, F., Chen, T., Zhang, Y., Simair, A., Rujie, C., Sadaf Zaidi, N., Guo, X., Wei, X., Siegel, G., & Lu, C. (2018). Large-scale in vitro transcription, RNA purification and chemical probing analysis. *Cellular Physiology and Biochemistry*, *48*(5), 1915–1927. https://doi.org/10.1159/000492512

Kern, J. A., & Davis, R. H. (1997). Application of solution equilibrium analysis to in vitro RNA transcription. *Biotechnology Progress*, *13*(6), 747–756. https://doi.org/10.1021/bp970094p

Kim, E., Agarwal, S., Kim, N., Hage, F. S., Leonardo, V., Gelmi, A., & Stevens, M. M. (2019). Bioinspired fabrication of DNA-inorganic hybrid composites using synthetic DNA. *ACS Nano*, *13*(3), 2888–2900. https://doi.org/10.1021/acsnano.8b06492

Kim, E., Zwi-Dantsis, L., Reznikov, N., Hansel, C. S., Agarwal, S., & Stevens, M. M. (2017). One-pot synthesis of multiple protein-encapsulated DNA flowers and their application in intracellular protein delivery. *Advanced Materials*, *29*(26), 1701086. https://doi.org/10.1002/adma.201701086

Kis, Z., Kontoravdi, C., Shattock, R., & Shah, N. (2021). Resources, production scales and time required for producing RNA vaccines for the global pandemic demand. *Vaccines*, *9*(1), 3. https://doi.org/10.3390/vaccines9010003

Koh, H. R., Roy, R., Sorokina, M., Tang, G.-Q., Nandakumar, D., Patel, S. S., & Ha, T. (2018). Correlating transcription initiation and conformational changes by a single-subunit RNA polymerase with near base-pair resolution. *Molecular Cell*, *70*(4), 695–706.e5. https://doi.org/10.1016/j.molcel.2018.04.018

Liu, D. C., & Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, *45*(1), 503–528. https://doi.org/10.1007/BF01589116

Martí, R. (2003). Multi-start methods. In F. Glover & G. A. Kochenberger (Eds.), *Handbook of metaheuristics* (pp. 355–368). Springer.

Myerson, A. S., Erdemir, D., & Lee, A. Y. (Eds.). (2019). *Handbook of industrial crystallization* (3rd ed., p. 100). Cambridge University Press. https://doi.org/10.1017/9781139026949

Peng, Y., Zhu, Z., Braatz, R. D., & Myerson, A. S. (2015). Gypsum crystallization during phosphoric acid production: Modeling and experiments using the mixed-solvent-electrolyte thermodynamic model. *Industrial & Engineering Chemistry Research*, *54*(32), 7914–7924. https://doi.org/10.1021/acs.iecr.5b01763

Pregeljc, D., Skok, J., Vodopivec, T., Mencin, N., Krušič, A., Ličen, J., Nemec, K. Š., Štrancar, A., & Sekirnik, R. (2023). Increasing yield of in vitro transcription reaction with at-line high pressure liquid chromatography monitoring. *Biotechnology and Bioengineering*, *120*(3), 737–747. https://doi.org/10.1002/bit.28299

Rohner, E., Yang, R., Foo, K. S., Goedel, A., & Chien, K. R. (2022). Unlocking the promise of mRNA therapeutics. *Nature Biotechnology*, *40*(11), 1586–1600. https://doi.org/10.1038/s41587-022-01491-z

Rosa, S. S., Nunes, D., Antunes, L., Prazeres, D. M. F., Marques, M. P. C., & Azevedo, A. M. (2022). Maximizing mRNA vaccine production with Bayesian optimization. *Biotechnology and Bioengineering*, *119*(11), 3127–3139. https://doi.org/10.1002/bit.28216

Rosa, S. S., Prazeres, D. M. F., Azevedo, A. M., & Marques, M. P. C. (2021). mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine*, *39*(16), 2190–2200. https://doi.org/10.1016/j.vaccine.2021.03.038

Samnuan, K., Blakney, A. K., McKay, P. F., & Shattock, R. J. (2022). Design-of-experiments in vitro transcription yield optimization of self-amplifying RNA. *F1000Research*, *11*, 333. https://doi.org/10.12688/f1000research.75677.1

Shopsowitz, K. E., Roh, Y. H., Deng, Z. J., Morton, S. W., & Hammond, P. T. (2014). Composite RNAi-microsponges form through self-assembly of the organic and inorganic products of transcription. Small (Weinheim an der Bergstrasse, Germany), *10*(8), 1623–1633. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4031615/

Tammenkoski, M., Moiseev, V., Lahti, M., Ugochukwu, E., Brondijk, T. H. C., White, S. A., Lahti, R., & Baykov, A. A. (2007). Kinetic and mutational analyses of the major cytosolic exopolyphosphatase from *Saccharomyces cerevisiae*. Journal of Biological Chemistry, *282*(13), 9302–9311. https://doi.org/10.1074/jbc.M609423200

Tang, G.-Q., Anand, V. S., & Patel, S. S. (2011). Fluorescence-based assay to measure the real-time kinetics of nucleotide incorporation during transcription elongation. *Journal of Molecular Biology*, *405*(3), 666–678. https://doi.org/10.1016/j.jmb.2010.10.020

van de Berg, D., Kis, Z., Behmer, C. F., Samnuan, K., Blakney, A. K., Kontoravdi, C., Shattock, R., & Shah, N. (2021). Quality by design modelling to support rapid RNA vaccine production against emerging infectious diseases. *npj Vaccines*, *6*(1), 1–10. https://doi.org/10.1038/s41541-021-00322-7

Wang, Y., Kim, E., Lin, Y., Kim, N., Kit-Anan, W., Gopal, S., Agarwal, S., Howes, P. D., & Stevens, M. M. (2019). Rolling circle transcription-amplified hierarchically structured organic–inorganic hybrid RNA flowers for enzyme immobilization. *ACS Applied Materials & Interfaces*, *11*(26), 22932–22940. https://doi.org/10.1021/acsami.9b04663

Wang, J., Pfeuty, B., Thommen, Q., Romano, M. C., & Lefranc, M. (2014). Minimal model of transcriptional elongation processes with pauses. *Physical Review E*, *90*(5), 050701. https://doi.org/10.1103/PhysRevE.90.050701

Young, J. S., Ramirez, W. F., & Davis, R. H. (1997). Modeling and optimization of a batch process for in vitro RNA production. *Biotechnology and Bioengineering*, *56*(2), 210–220. https://doi.org/10.1002/(SICI)1097-0290(19971020)56:2%3C210::AID-BIT10%3E3.0.CO;2-K

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Stover, N. M., Ganko, K., & Braatz, R. D. (2024). Mechanistic modeling of in vitro transcription incorporating effects of magnesium pyrophosphate crystallization. *Biotechnology and Bioengineering*, 1–12. https://doi.org/10.1002/bit.28699