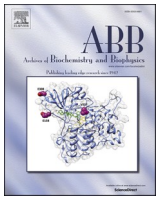
Archives of Biochemistry and Biophysics 778 (2026) 110737

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi



Systems analysis of the kinetics of in vitro transcription from interactions of T7 RNA polymerase and DNA

Nathan M. Stover [1], Marieke De Bock [1], Julie Chen, Jacob Rosenfeld, Maria del Carme Pons Royo, Allan S. Myerson, Richard D. Braatz [*]

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

ARTICLE INFO

Keywords:
Enzyme kinetics
Mathematical modeling
RNA polymerase
Transcription
Double-stranded RNA (dsRNA)

ABSTRACT

The in vitro transcription reaction (IVT) is of growing importance for the manufacture of RNA vaccines and therapeutics. While the kinetics of the microscopic steps of this reaction (promoter binding, initiation, and elongation) are well studied, the rate law of overall RNA synthesis that emerges from this system is unclear. In this work, we show that a model that incorporates both initiation and elongation steps is essential for describing trends in IVT kinetics in conditions relevant to RNA manufacturing. In contrast to previous reports, we find that the IVT reaction can be either initiation- or elongation-limited depending on solution conditions. This initiation-elongation model is also essential for describing the effect of salts, which disrupt polymerase-promoter binding, on transcription rates. Polymerase-polymerase interactions during elongation are incorporated into our modeling framework and found to have nonzero but unidentifiable effects on macroscopic transcription rates. Finally, we develop an extension of our modeling approach to quantitatively describe and experimentally evaluate RNA- and DNA-templated mechanisms for the formation of double-stranded RNA (dsRNA) impurities.

1. Introduction

The in vitro transcription (IVT) reaction for synthesis of RNA is a necessary step for the production of a growing number of vaccines and therapeutics. IVT is a cell-free biochemical reaction that requires a DNA template and a DNA-dependent RNA polymerase enzyme, most commonly T7 RNA polymerase. A kinetic law for the rate of RNA synthesis as a function of the concentrations of these catalysts can aid in design, optimization, and mechanistic understanding of the IVT process. However, a rate law for IVT in conditions relevant to RNA manufacturing has not been fully developed.

The elementary kinetic mechanisms that constitute IVT are well studied. Each of the promoter binding, initiation, promoter release, and elongation steps has been studied using a diverse and orthogonal set of tools, including thermodynamic measurements [1,2], kinetic modeling [3–6], structural analysis [7,8], and single-molecule experiments [9–11]. Despite this microscopic understanding, there is little research into the emergent macroscopic kinetics of systems in which these steps coexist. This knowledge gap is relevant in the context of manufacturing long RNA sequences (>1000 bp), such as mRNA and self-amplifying RNA (saRNA) vaccines.

The work of Arnold et al. [12] is the most complete past approach to modeling macroscopic transcription rates. Arnold et al. report that the estimated elongation rate constant is much greater than the estimated initiation rate constant and conclude that the IVT reaction is initiation limited for all industrially relevant sequence lengths. However, this work is incomplete and not representative of the IVT reaction in a bio-manufacturing context. As will be described in the results section, the kinetic model used by Arnold et al.is not appropriate for the synthesis of long RNA sequences. Secondly, this previous work estimated transcription and elongation rate constants using a poorly defined parameter estimation approach in which the number of fitted parameters nearly equaled the number of data points. It is well understood that this approach can be extremely sensitive to experimental noise or out-of-model effects. As a consequence, their estimated elongation rate constant ($5.8 \times 10^7 1\ /s$) is not only far greater than orthogonal estimates from single-molecule studies ($\sim 1.7 \times 10^2$ 1/s) [10], but is 10^3 times greater than the diffusion-limited rate constant, which implies that it is an aphysical artifact of errors in the parameter estimation process. These considerations are relevant for practitioners in the field of RNA manufacturing. For example, Boman et al. [13] rely on Arnold et al.'s conclusion of initiation limitation to estimate the effect of sequence

* Corresponding author.
E-mail addresses: stover@mit.edu (N.M. Stover), braatz@mit.edu (R.D. Braatz).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.abb.2026.110737
Received 19 June 2025; Received in revised form 14 January 2026; Accepted 16 January 2026
Available online 16 January 2026

length on IVT rates for process development. A kinetic modeling approach that appropriately incorporates all steps of the transcription process can serve as a useful tool for several aspects of RNA manufacturing, including accelerating process development of new sequences and designing dynamic models of IVT.

Beyond predicting the rates of product RNA formation, kinetic modeling can be a useful tool for understanding the formation of double-stranded RNA (dsRNA) impurities. Double-stranded RNA is a highly immunogenic byproduct of IVT which is costly to remove in downstream purification. Multiple mechanisms have been proposed to describe the formation of dsRNA, including RNA-templated 3' addition [14] and DNA-templated synthesis of antisense RNA followed by hybridization [15]. While both of these mechanisms have been experimentally observed within the context of model systems, their usefulness for the quantitative process development of the IVT reaction is unclear and there are no publications on attempting to quantify the kinetic predictions of these mechanisms.

In this work, we investigate a kinetic rate law that incorporates polymerase-promoter binding, initiation, and elongation steps from a first-principles standpoint. We show that this initiation-elongation model is necessary for describing the rate of transcription in regimes relevant to the manufacturing of mRNA. We demonstrate how the kinetic parameters of this model can be estimated from a minimal set of experiments, which allows for a comparative analysis between different DNA sequences and serves as a guide for practitioners on troubleshooting and understanding the application of IVT to novel sequences. We consider both the effect of polymerase-polymerase interactions and polymerase-DNA binding disruptions on the kinetic predictions of this model. Finally, we develop an extension of our modeling approach to quantitatively describe and experimentally evaluate RNA-templated and DNA-templated mechanisms for dsRNA formation.

2. Results

2.1. Analysis of initiation-elongation kinetic model

The structure of the initiation-elongation model used in this work is a variant of a model postulated (but not experimentally explored) in a past publication [16]. The IVT reaction involves a series of binding, initiation, and elongation kinetic steps. This reaction network was approximated to operate in a quasi-steady state as the time associated with the synthesis of a single transcript as measured by single-molecule experiments (3–30 s) is substantially lower than the time constant of substrate consumption for the data in this work (0.25–0.5 h) [10,17]. In this model, polymerase (P) and the DNA promoter (DNA_p) reversibly bind to form a complex ($P{\cdot}DNA_p$) that can undergo transcription initiation at a rate k_i. Here, initiation is defined as the transition in which the polymerase both begins translocation along the DNA promoter and dissociates from the promoter region. The initiation process involves a number of sequential kinetic sub-steps [17]. For the purpose of developing a macroscopically identifiable model in the quasi-steady limit, we lump these steps as a single first-order kinetic process. After initiation, the polymerase translocates across the DNA sequence in an elongation state (P_E). Similarly to initiation, the elongation of the RNA strand by a single base pair involves a number of sequential kinetic processes, which are repeated for each base pair in the sequence [4]. We lump these sequential kinetic processes as a single first-order kinetic process with an effective rate constant $k_{e,\mathrm{tot}}$. Initiation frees the DNA promoter to be further bound by incoming polymerase. Transcription termination was assumed to be instantaneous for the linearized DNA templates used in RNA manufacturing. In addition, this model neglects the formation of short sequences resulting from abortive transcription. Considering that these aborts comprise a negligible mass fraction of the IVT product of long transcripts, this abortion process can be considered part of the effective dissociation rate of the initiation complex, where k_{off} below incorporates both polymerase-promoter disassociation and abortion. Schematically, the transcription process is represented as

$$\mathrm{P} + \mathrm{DNA_p} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \mathrm{P{\cdot}DNA_pP_E} \xrightarrow[\text{-P}]{k_{e,\mathrm{tot}}} \mathrm{RNA}. \tag{1}$$

While transcription in the low-volume environment of a cell is commonly modeled as a stochastic process [18], this reaction is modeled as a deterministic process in this work owing to the large number of RNA polymerase molecules in a macroscopic IVT reaction ($>10^{12}$ for all experiments performed in this work). In addition, we assume that this reaction system is well-mixed. A consequence of describing the complex initiation and elongation processes, which involve and equilibrium process of NTP binding, is that these lumped initiation and elongation rate constants are dependent on solution conditions, notably temperature, pH, nucleoside triphosphate (NTP), and Mg concentrations. Here we model the IVT reaction rate as a function of DNA and RNA polymerase concentrations to serve as a framework for understanding the effects of other process variables. The rate of RNA synthesis is equivalent to the initiation rate,

$$\mathrm{R_{tr}} = k_i\left[\mathrm{P{\cdot}DNA_p}\right], \tag{2}$$

which is dependent on the concentration of the initiation complex. This complex concentration is derived using a quasi-steady state approximation (SI Section 1),

$$\left[\mathrm{P{\cdot}DNA_p}\right] = \frac{[\mathrm{P}]_{\mathrm{tot}} + \alpha[\mathrm{DNA}]_{\mathrm{tot}} + K_{\mathrm{MD}} - \sqrt{\left([\mathrm{P}]_{\mathrm{tot}} + \alpha[\mathrm{DNA}]_{\mathrm{tot}} + K_{\mathrm{MD}}\right)^2 - 4\alpha[\mathrm{P}]_{\mathrm{tot}}[\mathrm{DNA}]_{\mathrm{tot}}}}{2\alpha}, \tag{3}$$

where $[\mathrm{P}]_{\mathrm{tot}}$ and $[\mathrm{DNA}]_{\mathrm{tot}}$ are total polymerase and DNA concentrations and

$$\alpha = 1 + \frac{k_i}{k_{e,\mathrm{tot}}}, \qquad K_{\mathrm{MD}} = \frac{k_i + k_{\mathrm{off}}}{k_{\mathrm{on}}}. \tag{4}$$

This hypothesized model differs from the approach of Arnold et al. [12] in two key ways. Firstly, the removal of DNA promoter and RNA polymerase at different points in the process allow for a single DNA chain to feature multiple bound elongating polymerase molecules. Secondly, no assumptions are made regarding the relative concentration of DNA and polymerase during derivation of the rate law, which allows the model to operate across a broader space of species concentrations. For $\alpha = 1$, the proposed model converges to the structure of the rate law used by Martin and Coleman [19] in describing oligonucleotide transcription rates, which is based on an assumption that the effect of elongation is negligible. This rate law is referred to as an *initiation-limited* model in this work and is a special case of the presented initiation-elongation model.

The initiation-elongation model predicts that the limiting step depends on relative DNA and polymerase concentrations. In the regime where $\alpha[\mathrm{DNA}]_{\mathrm{tot}} \ll [\mathrm{P}]_{\mathrm{tot}}$, the limiting factor is the number of DNA promoter binding sites, and the overall rate converges to

$$R_{tr} = \frac{k_i[\mathrm{DNA}]_{tot}[\mathrm{P}]_{tot}}{[\mathrm{P}]_{tot} + K_{MD}}. \tag{5}$$

In this limit, the predictions of the initiation-limited model and the initiation-elongation model converge. We refer to this limit as the *initiation-limited regime*. Conversely, in the regime where $[\mathrm{P}]_{tot} \ll \alpha[\mathrm{DNA}]_{tot}$, the rate law converges to

$$R_{tr} = \left(k_i^{-1} + k_{e,tot}^{-1}\right)^{-1} \frac{[\mathrm{DNA}]_{tot}[\mathrm{P}]_{tot}}{[\mathrm{DNA}]_{tot} + \frac{K_{MD}}{\alpha}}, \tag{6}$$

in which the rate of transcription is limited by the combined timescale of initiation and elongation. Considering that for long (>1000 base pair) sequences, $k_{e,tot} < k_i$, we refer to this case as an *elongation-limited* regime.

2.2. Initiation-elongation model is necessary to describe IVT kinetic data

To validate the structure of the initiation-elongation model, the rate of transcription of a 2078 base pair DNA template encoding the firefly luciferase gene (Fluc) was measured as a function of DNA and T7 RNA polymerase concentrations (Fig. 1). Both the initiation-limited model (with parameters K_{MD} and k_i) and the initiation-elongation model (with parameters K_{MD}, k_i, and α) are fit to these data. As the solution concentrations of DNA and T7 RNA polymerase used in these experiments

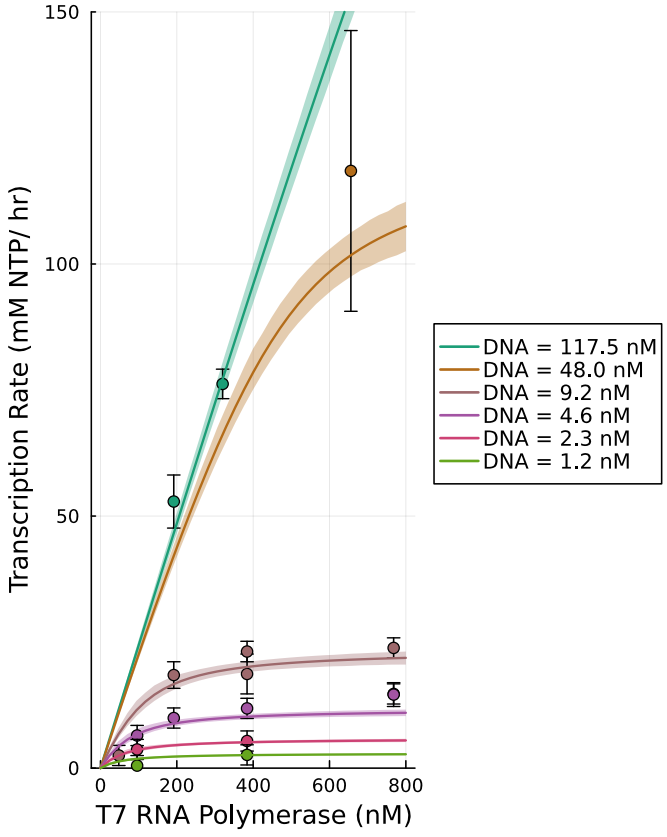


Fig. 1. Transcription Kinetics of Fluc Sequence.
Measured transcription rate as a function of T7 RNA polymerase and DNA concentrations with initiation-elongation model predictions after fitting. Other solution conditions are held constant as described in Methods section. Reaction rate is linear with respect to RNA polymerase in the regime of $\alpha[\mathrm{DNA}]_{tot} \gg [\mathrm{P}]_{tot}$ (points in upper left of graph) and linear with respect to DNA concentrations in regime of $\alpha[\mathrm{DNA}]_{tot} \ll [\mathrm{P}]_{tot}$ (lower right of graph). Error bars on data points represent estimated 1σ experimental error based on triplicate experiments. Shaded areas are 95 % prediction intervals of model based on estimated covariance matrix.

are too high to give identifiable estimates for K_{MD}, a Bayesian prior for K_{MD} of 50 ± 25 nM was used based on previous measurements [3]. This was acceptable for the fitting process as the main goal of these experiments was to estimate k_i and α. Bayesian information criterion analysis showed that the additional parameter of the initiation-elongation model provided a significant improvement in fitting over the initiation-limited model (SI Section 2.5). The initiation-elongation model (unlike the initiation-limited model) describes key trends in the data, including the linear relationship between RNA polymerase concentrations and reaction rate in the high-DNA regime and the linear relationship between DNA template concentration and reaction rate in the low-DNA regime. Uncertainty analysis indicates that the parameter estimates of the initiation-elongation model are practically identifiable and that the uncertainty region of k_i and α are not highly correlated with uncertainty in K_{MD}. This indicates that the exact choice for the prior value of K_{MD} has a minor effect on the estimates of k_i and α.

With the structure of the initiation-elongation model validated, a model-based design of experiments (MBDOE) approach using the D-optimal criterion was employed to choose experimental conditions that best identify the two key kinetic parameters (k_i and α) of three more DNA sequences with differing DNA sequence length and initiation sequence: a dodecamer sequence matching the first 12 base pairs of the Fluc sequence and sequences coding for the COVID spike protein and EGFP protein (Table 1). MBDOE analysis indicated that two experiments corresponding to the initiation and elongation limited regimes were sufficient to practically identify the two parameters. When necessary to achieve greater parameter precision after one round of data collection, the MBDOE process was iterated. Table 1 shows the sequences used, their length, first three initiating base pairs, and their estimated kinetic parameters. An average per-base pair elongation rate constant is calculated as

$$k_{e,bp} = N_{all} k_{e,tot} \tag{7}$$

to aid in comparison between sequences.

2.3. Kinetic modeling indicates that polymerase-polymerase interactions and pausing during elongation have nonzero but unidentifiable effects on macroscopic reaction rates

The initiation-elongation model is a minimal approach to understanding the kinetic trends of the IVT system and uses a number of approximations. One key approximation is that all polymerase molecules in the elongation state advance with the same rate constant regardless of their position on the chain or the local density of polymerase molecules. This approximation is not valid in the case where polymerase molecules can hinder each other's progress along the DNA sequence, which has been observed in the context of T7RNA polymerase [20]. In addition, this polymerase-polymerase exclusion can be exacerbated by the pausing of polymerase during elongation [21,22]. The possibility of polymerase-polymerase interactions and pausing raises several questions relevant for the engineering of the IVT system. In what regimes, if any, can these polymerase-polymerase interactions distort the predictions of the initiation-elongation model presented above? In addition, can the extent of these interactions be assessed using macroscopic rate measurements?

To answer these questions, we developed a kinetic model that extends the initiation-elongation model to represent elongation as a totally asymmetric simple exclusion process (TASEP). The model assumes that polymerase molecules elongate by unidirectional transitions between M L-nucleotide sized segments, where L is the estimated exclusion width of the polymerase molecule. Schematically, this model has the structure

$$\mathrm{P} + \mathrm{DNA_p} \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} \mathrm{P{\cdot}DNA_p} \xrightarrow[-\mathrm{DNA_p}]{k_i} \mathrm{P_1} \longrightarrow \mathrm{P_2} \cdots \mathrm{P_M} \xrightarrow[-\mathrm{P}]{} \mathrm{RNA} \tag{8}$$

Table 1
Transcription parameters of DNA sequences. Each sequence is characterized by its length and first three initiating nucleotides (init). Additional sequence information is shown in SI Section 7.

	Length	Init.	Parameter			
			k_i (s^{-1})	α	$k_{e,\mathrm{tot}} \times 10^{-2}$ (s^{-1})	$k_{e,\mathrm{bp}} \times 10^{-9}$ (s^{-1})
Fluc	2078	AGA	0.34 ± 0.03	9.9 ± 0.7	3.8 ± 0.2	7.8 ± 0.6
Fluc dodecamer	12	AGA	0.42 ± 0.08	0.7 ± 0.3	–	–
COVID	4243	AGA	0.36 ± 0.04	20 ± 4	1.9 ± 0.3	8.1 ± 0.1
EGFP	942	GGG	0.78 ± 0.14	15 ± 3	5.6 ± 0.8	5.3 ± 0.8

where P_i represents a polymerase molecule on the ith segment. Exclusion between particles is represented by the form of the rate law describing translocation between segments. Past work has focused on developing and validating mean-field approximations for these rates that account for both polymerase-polymerase interactions and pausing of polymerase molecules during elongation [22]. Using these rate laws, we developed an approximate analytical expression to predict macroscopic reaction rates (SI Section 2). In the most general case including pausing and polymerase-polymerase interactions, the derived rate law is

$$R_{\mathrm{tr}} = k_i\left[\mathrm{P{\cdot}DNA_p}\right]\frac{[\mathrm{DNA}]_{\mathrm{tot}} - \dfrac{\beta\left[\mathrm{P{\cdot}DNA_p}\right]}{K_{MD}-\theta}}{[\mathrm{DNA}]_{\mathrm{tot}} + \gamma\left[\mathrm{P{\cdot}DNA_p}\right]} \tag{9}$$

where the initiation complex $\left[\mathrm{P{\cdot}DNA_p}\right]$ is defined by

$$\alpha\gamma\left(\frac{\left[\mathrm{P{\cdot}DNA_p}\right]}{[\mathrm{DNA}]_{\mathrm{tot}}}\right)^3 + \left(\alpha + \frac{\beta}{[\mathrm{DNA}]_{\mathrm{tot}}} - \gamma\left(\frac{[\mathrm{P}]_{\mathrm{tot}}}{[\mathrm{DNA}]_{\mathrm{tot}}} + \frac{\theta}{[\mathrm{DNA}]_{\mathrm{tot}}} + \alpha\right)\right)\left(\frac{\left[\mathrm{P{\cdot}DNA_p}\right]}{[\mathrm{DNA}]_{\mathrm{tot}}}\right)^2 - \left(\frac{[\mathrm{P}]_{\mathrm{tot}}}{[\mathrm{DNA}]_{\mathrm{tot}}} + \alpha + \frac{K_{\mathrm{MD}}}{[\mathrm{DNA}]_{\mathrm{tot}}} - \gamma\frac{[\mathrm{P}]_{\mathrm{tot}}}{[\mathrm{DNA}]_{\mathrm{tot}}}\right)\left(\frac{\left[\mathrm{P{\cdot}DNA_p}\right]}{[\mathrm{DNA}]_{\mathrm{tot}}}\right) + \frac{[\mathrm{P}]_{\mathrm{tot}}}{[\mathrm{DNA}]_{\mathrm{tot}}} = 0 \tag{10}$$

with the composite parameters

$$\gamma = k_i\frac{f\tau^2}{1+f\tau}, \qquad \beta = \frac{Lk_i^2}{k_e k_{\mathrm{on}}}, \qquad \theta = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}}, \tag{11}$$

where f and τ represent the frequency of pausing events and the timescale of pauses, respectively, following the approach of Wang et al. [22] We refer to the model in this general case as the *long pause* (LP) model. Intuitively, the dimensionless parameter γ represents the relative importance of pausing to the transcription system, and setting γ to zero results in a simpler model that neglects the effect of pauses (which we call the *short pause* (SP) model). The parameter β represents the relative importance of polymerase-polymerase interactions on the system. Setting β to zero recovers the initiation-elongation model.

The complexity and number of parameters of the LP model raises a question of practical identifiability. If the transcription system truly behaved in accordance with equations (9) and (10), could it be distinguished from the initiation-elongation model with macroscopic measurements? Can macroscopic measurements be used to identify the parameters γ and β? Bayesian information criterion analysis indicates that neither of the SP or LP models fit the kinetic data collected in this work better than the initiation-elongation model (SI Section 2.5). Moreover, the same is true for synthetic data generated by the SP and LP models in the case of reasonable estimates for microscopic parameters and experimental noise. In fitting this synthetic data, neither the SP nor LP model can identifiably recover estimates for all of their kinetic parameters, resulting in highly correlated parameter confidence regions.

While the initiation-elongation model can describe the output of the more complicated LP and SP models, the resulting fitted kinetic parameters do not match the microscopic ground truth values used to generate the data. This implies that the measured kinetic parameters in Table 1 may serve as effective parameters that do not perfectly reflect microscopic rates of initiation and elongation. The effective initiation rate constant estimated from data generated by the SP model is approximately 80–90 % of the ground truth value. In the case of data generated by the LP model, the estimated initiation rate constant is dependent on values of γ but can be significantly lower (SI Section 2.5). The estimated elongation rate constant is not distorted by more than 10 % in either of these cases.

2.4. Initiation-elongation model describes sensitivity of transcription rates to salt addition

Past research has noted that the addition of salts, including the necessary magnesium cofactor, to the IVT system can decrease transcription rates by disrupting the binding between RNA polymerase and the DNA promoter [3,23]. The initiation-elongation model is a useful tool for quantitative understanding the effect of salt concentrations on transcription rates. Equations (5) and (6) show a non-obvious emergent result of the initiation-elongation model. While both initiation- and elongation-limited regimes can be described using a Michaelis-Menten structure, the effective Michaelis-Menten constant differs between the two by a factor of α. Intuitively, the elongation-limited regime is less sensitive than the initiation-limited regime to disruptions in polymerase-promoter binding.

In order to describe the effect of salt addition on overall transcription rates, a model for the effect of salt concentrations on K_{MD} is required. In this work, we assume that the inhibitory effects of salts and other reaction components on polymerase-promoter binding are due to rapid processes that reach equilibrium much faster than the rates of elongation, as opposed to longer-timescale processes that can inactivate DNA promoters during transcription in vivo [24]. The IVT system contains multiple salts, including NTPs, magnesium, buffers, and associated counterions. It has been shown that different salts affect transcription rates to different degrees [23]. While the literature on salt effects on protein-DNA binding is substantial [25], there is little published work on the practical problem of modeling this relationship in the context of mixed-salt systems relevant to IVT. We adapted a previously proposed approach that augments the predictions of counterion condensation (CC) theory with an effective salt concentration [26]:

$$K = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}} = K_0\left(\frac{[\mathrm{salt}]}{1\ \mathrm{mole}}\right)^n \tag{12}$$

where K_0 represents the intrinsic binding strength of the DNA promoter, and [salt] is an effective salt concentration calculated as

$$[\text{salt}] = \sum_{i=1}^{N_{\text{ion}}} \omega_{\text{ion},i}[\text{ion}_i] \tag{13}$$

where $\omega_{\text{ion},i}$ is a weighting factor specific to each cation and anion i in the IVT system (SI Section 3.2). While this relation was developed as an empirical extension of classical CC theory [27], we show that it emerges naturally from an extended treatment of CC theory that includes the presence of multiple salts (SI Section 3). While CC theory is a dramatically simplified rendering of the physics of ligand-DNA interactions that can be modeled with greater fidelity using computational techniques that incorporate biomolecule structure and diffuse ion binding [28,29], this relation describes key trends in experimental binding data of mixed salt systems and is useful in an engineering context. While this relation has previously been shown to represent trends in thermodynamic binding data [26], additional assumptions are required to extend it to describing K_{MD}, which combines binding thermodynamics with the kinetic processes of initiation and abortion. To adapt this thermodynamic relation to kinetic modeling, we assume that k_{on} is constant at a value of $5.67 \times 10^{-2}\ \text{s}^{-1}\ \text{nM}^{-1}$ [17]. While k_{on} has been shown to be sensitive to salt [26], this approximation is valid in the high salt limit in which K_{MD} is very large. Similarly, we assume here that salt concentrations do not affect the kinetics of initiation. One effect of these approximations is that the parameter K_0 incorporates the kinetic effects of abortion (through the rate constant k_{off}) and is thus best considered to be an empirical parameter that incorporates both thermodynamic and kinetic processes. According to CC theory, the constant n represents the number of cations displaced upon polymerase-DNA binding, a prediction which has been shown to correlate with experimental results [30]. Based on an approximate structural analysis of T7 RNA polymerase, we hypothesized that a reasonable estimate for n was 5.0^8. We validated this choice of n, fitting a single parameter K_0, on published data of K_{MD} as a function of sodium chloride addition (Fig. 2A) [3]. We find that this model suitably describes changes in K_{MD} in the high-salt limit.

To test the prediction of the initiation-elongation model that the salt sensitivity of the IVT reaction differs between the initiation and elongation limited regimes, we measured the transcription rate of the Fluc construct as a function of sodium chloride addition. Two reaction schemes with different DNA concentrations (117.5 and 9.2 nM) are tested with solution conditions (including 192 nM T7 RNA polymerase) otherwise held constant. The ratios $\alpha[\text{DNA}]_{\text{tot}}/[\text{P}]_{\text{tot}}$ of the two schemes are 5.6 and 0.5, reflecting that these experiments probed an elongation-limited regime and a region primarily governed by initiation limitation, respectively. The two reaction schemes exhibited significantly different responses to salt addition (Fig. 2B). After fitting a single parameter K_0, the initiation-elongation model described these trends. The difference in behavior between reaction conditions cannot be predicted by a model which only considers initiation limitations.

Table 2
Estimated binding parameters of measured sequences. Each sequence is characterized by its length and first three initiating nucleotides (init). Additional sequence information is shown in SI Section 7.

	Parameter			
	Length	Init.	K_0 (mM)	K_{MD} at [NaCl] added = 0 (nM)
Fluc	2078	AGA	1.5–3.0	40–70
EGFP	942	GGG	0.5–1.0	25–35

In order to understand how salt sensitivity varies between sequences, the same kinetic measurements are performed on the EGFP sequence using an IVT scheme analogous to the low-DNA Fluc kinetics discussed above (containing equal DNA concentrations by mass of transcribed region). If the parameter K_0 of the two sequences is the same, model prediction indicates that the salt sensitivity should be the same as the Fluc construct within the range of experimental precision. It was instead found that the EGFP sequence was significantly less sensitive to salt addition than the COVID sequence. In our modeling framework, this lower sensitivity is parametrized as a lower value for K_0. Table 2 shows estimated values of K_0 for the two tested sequences, as well as the implied K_{MD} in the absence of sodium chloride addition. These calculated values of K_{MD} validate the Bayesian prior used for the estimation of parameters in Table 1.

2.5. Extending modeling approach to the formation of dsRNA impurities

Double-stranded RNA (dsRNA) is an immunogenic byproduct of the IVT reaction. These dsRNA byproducts are heterogeneous in size and sequence, and a given RNA product molecule may contain both single and double stranded regions. As such, dsRNA is challenging to remove in downstream processing steps of RNA manufacturing. Two mechanisms for dsRNA formation in IVT have been proposed, which share undesired polymerase binding as a common feature. A mechanism of RNA self-templated extension has been shown to produce short double-stranded segments in oligomeric model systems [14], and has been used as the conceptual basis for strategies to decrease dsRNA formation based on immobilization and high salt concentrations [31]. Conversely, a mechanism of DNA-templated antisense RNA synthesis has been shown to form hybridized dsRNA structures for specific sequences [15]. While each of these mechanisms has been used as the conceptual basis for engineering strategies to reduce dsRNA formation, there is no work in understanding the quantitative implications of these models for input-output relations of dsRNA formation.

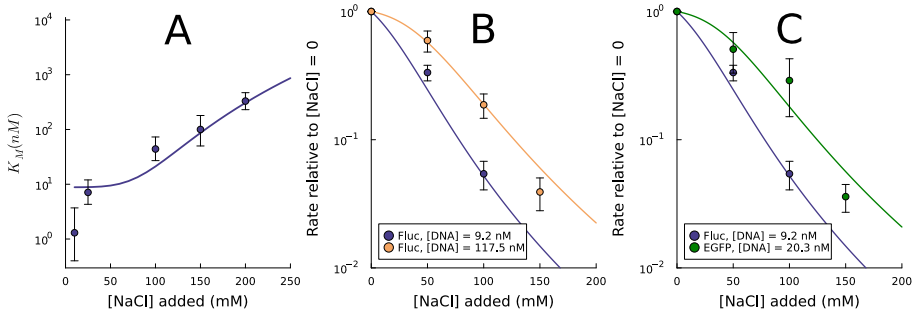


Fig. 2. Effect of salt addition on IVT kinetics.
(A) Semi-empirical model describing effect of salt addition of polymerase-promoter binding can describe trends in K_{MD}, particularly in the range of high salt concentrations (>150 mM NaCl added). Data from Maslak and Martin [3]. (B) Applying this model for K_{MD} to transcription kinetics of long sequences explains the difference in salt sensitivity between two reaction schemes that only differ in DNA concentration. The reaction scheme in the elongation-limited regime ($[\text{DNA}]_{\text{tot}} = 117.5$ nM) is less sensitive that the scheme in the initiation-limited regime ($[\text{DNA}]_{\text{tot}} = 9.2$ nM) to disruptions in polymerase-promoter binding due to salt addition. Model predictions are shown after fitting a single parameter K_0 (2.5 mM). (C) The EGFP sequence was measured to be less sensitive to salt addition than the Fluc sequence in analogous (equal DNA mass) reaction conditions. Model predictions are shown for best fit K_0 estimates for each sequence (0.75 mM for EGFP).

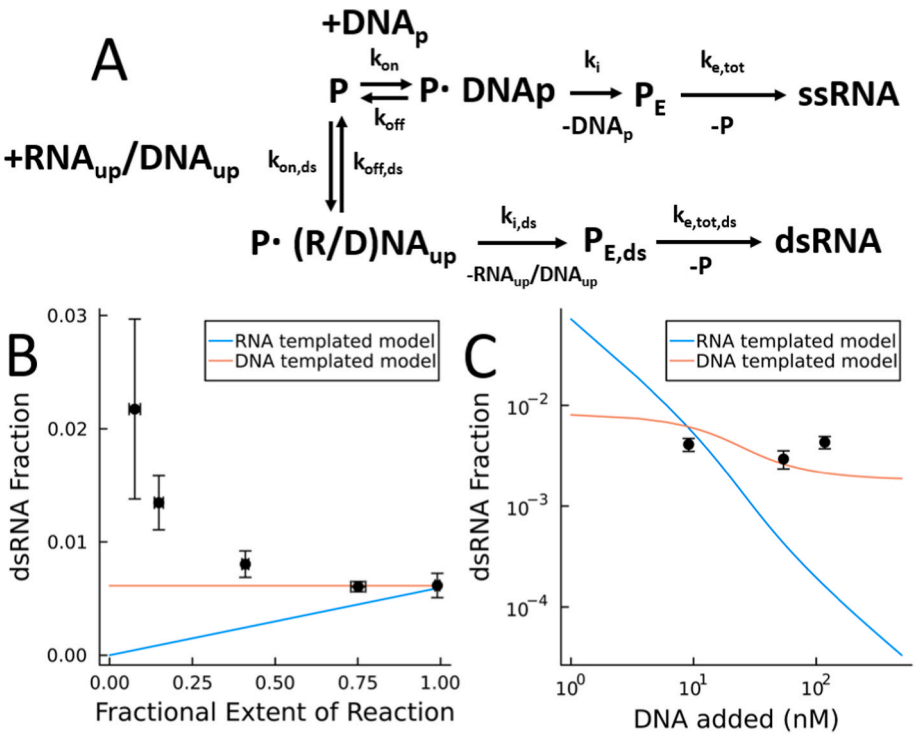


Fig. 3. Assessing Kinetic Models for dsRNA Formation.
(A) Kinetic models for dsRNA formation are based on polymerase binding and initiation at undesired sites, which competes with the formation of the desired single-stranded product. In an RNA-templated mechanism, RNA polymerase binds to transient loop-back RNA structures and elongates across the RNA sequence, synthesizing dsRNA. In a DNA-templated mechanism, RNA polymerase binds to an undesired antisense promoter and synthesizes antisense RNA. These antisense RNA products hybridize with the main product RNA, which our model assumes is an instantaneous process. (B) Fraction of dsRNA in the IVT product as a function of the extent of reaction. Model predictions are calibrated based on the final timepoint to show conceptual predictions. Reactions are performed with 192 nM of T7 RNA polymerase and 9.2 nM of Fluc DNA. (C) The dsRNA fraction after complete conversion is not significantly affected by DNA input concentrations. Model predictions are shown using the previous parameter calibration. Reactions are performed with 192 nM of T7 RNA polymerase.

Considering that undesired RNA polymerase binding is the foundation of both mechanisms, the modeling approach developed in this work, which explicitly considers both free polymerase solution concentrations and polymerase-DNA binding, serves as a necessary platform for modeling dsRNA formation kinetics. We developed an extension of the initiation-elongation model to incorporate the binding of free RNA polymerase to either an undesired promoter on RNA (RNA_{up}) or an undesired promoter on the antisense DNA strand (DNA_{up}) (Fig. 3A). This model considers dsRNA to be a homogeneous chemical species. While this description is not perfectly representative of the known heterogeneity of dsRNA, it is appropriate for understanding trends in macroscopic dsRNA quantities. Using this schematic model, we derived quantitative input-output relations to model the fraction of dsRNA in the IVT product (SI Section 4). In addition to the approximations used to derive the initiation-elongation model, we assume that the amount of dsRNA product is much less than the ssRNA product and that undesired binding is a relatively rare event compared to the desired binding.

For a mechanism of RNA-templated dsRNA formation, our modeling approach predicts that the product dsRNA fraction is proportional to

$$\frac{[\mathrm{dsRNA}]_t}{[\mathrm{R}]_t} \propto \frac{[\mathrm{R}]_t}{[\mathrm{DNA}]_{tot} - [\mathrm{P{\cdot}DNA_p}]} \tag{14}$$

where $[\mathrm{dsRNA}]_t$ and $[\mathrm{R}]_t$ are the concentrations of dsRNA and total RNA at a given extent of reaction, and $[\mathrm{P{\cdot}DNA_p}]$ is the same quantity given by equation (3). In the case of DNA-templated antisense synthesis, generating mechanistic predictions is more difficult given the dynamics of the sense-antisense hybridization step. Kinetic studies of analogous systems indicate that the rate constant of this hybridization is 10^{-5}–10^{-4} min^{-1} nm^{-1}, which implies a time constant of approximately 1–10 min for the reaction concentrations used in this work. Considering that the time constant of the IVT reactions studied in this work take place on time scales of about 20–600 min, the hybridization step was approximated as instantaneous. While this approximation may neglect these hybridization trends, it lends a dramatic simplification to model predictions. In the case of DNA-templated dsRNA formation, our modeling approach predicts that

$$\frac{[\mathrm{dsRNA}]_f}{[\mathrm{R}]_f} \propto \frac{[\mathrm{DNA}]_{tot}}{[\mathrm{DNA}]_{tot} - [\mathrm{P{\cdot}DNA_p}]} \tag{15}$$

Equations (14) and (15) can either be viewed as competing models, or as two components of a larger modeling strategy that includes both RNA- and DNA-templated pathways. In the context of this work, we focus on their evaluation as competing models.

The macroscopic predictions of these models differ in two key ways. First, the RNA-templated model predicts that the dsRNA fraction (the ratio of dsRNA concentration to total RNA concentration) is low at early timepoints and rises linearly with respect to reaction conversion, while the DNA-templated model predicts that the dsRNA fraction is constant with respect to reaction conversion. In addition, the RNA-templated model, which assumes a competition between RNA and DNA as binding sites, predicts that dsRNA formation should trend to zero as the concentration of DNA is increased. The DNA-templated model predicts some dependence between input DNA and dsRNA formation, but predicts a finite asymptotic value of dsRNA formation.

To evaluate these models, we measured the final mass fraction of dsRNA in the Fluc IVT product as a function of the extent of reaction (Fig. 3B). Above a fractional conversion of 40 %, the dsRNA fraction of the system was relatively constant. However, timepoints collected at earlier conversions showed a decreasing trend, which is not consistent

with the predictions of either model. The same measurements performed on the COVID and EGFP constructs showed a similar nonincreasing result (SI Section 5). In addition, we measured the dsRNA fraction at complete conversion as a function of input DNA (Fig. 3C). We found that these data did not exhibit a clear statistical trend. We found that varying the concentrations of polymerase enzyme and salts did not significantly affect final dsRNA fractions (SI Section 5).

3. Discussion

Given the ubiquity of in vitro transcription in industrial RNA manufacturing, a kinetic framework that incorporates both DNA template and RNA polymerase concentrations to predict rates of RNA synthesis is a crucial tool to effectively use both expensive catalysts. In this work, we find that a model which incorporates both initiation and elongation steps is required to describe kinetic data across a range of DNA and RNA polymerase concentrations. Contrary to previous reports, we show that the limiting step is dependent on solution conditions and that the reaction system can be limited by the rate of elongation.

A primary goal of this work is to inform rapid and economical IVT process development, which currently involves data-driven designs of experiments to assess input-output relations. Since different DNA sequences and lengths transcribe with different kinetics, these designs are often repeated for each manufactured sequence. In our modeling approach, differences between sequences are encoded by kinetic parameters. We demonstrate that the key kinetic parameters of transcription (k_i and α) can be identified with only two experiments (taken in the elongation and initiation limited regimes, respectively). The measured values of these parameters are correlated with physical intuition and prior literature (Table 1). Sequences with the same three initiating base pairs (AGA) had the same initiation rate constant within the uncertainty of our measurements ($0.3 - 0.5\,s^{-1}$). The EGFP construct, which contained the canonical initiation sequence (GGG), exhibited a significantly higher initiation rate constant ($0.6 - 0.9\,s^{-1}$). These values are in the general range of previous reports from both single molecule and oligomeric studies ($0.3 - 0.6\,s^{-1}$) [3,17] and are consistent with reports that mutations to the canonical initiation sequence decreased overall transcription rates [32]. We find that the parameter α, which describes the relative importance of elongation in the transcription process, is loosely correlated with sequence length. This in turn implies that the effective per-base-pair elongation rate is within 44–83 s^{-1} for all sequences tested. This is in the same order of magnitude as previously reported values from single molecule studies ($1.1 - 2.2 \times 10^2\,s^{-1}$) [11,33]. In addition, the estimated α of a tested oligomeric sequence is 1.0 within the interval of uncertainty, which confirms our intuition that transcription of oligomers is purely initiation limited.

Differences between the rate constants calculated in this work and those reported by previous researchers can be explained by two causes. These kinetic parameters are first-order approximations of multiple steps and as such are dependent on NTP concentrations, pH, temperature, and other solution conditions. Solution concentrations of NTPs, magnesium, and other salts are typically much higher in the context of industrial RNA manufacturing (and this work) than in most fundamental studies of transcription kinetics. In the case of elongation rate constants, additional phenomena not included in our kinetic model such as nonspecific polymerase binding and polymerase pausing may contribute to the mismatch between these results and single molecule values. In addition, kinetic modeling indicates that polymerase-polymerase interactions during elongation can manifest as a decrease in the effective initiation rate constant when analyzed with the initiation-elongation model. If the estimated initiation rate constant of a DNA sequence is significantly less than the initiation rate constant of its initiating oligomer sequence, these polymerase-polymerase interactions could be a cause. We do not observe a significant difference between the Fluc and Fluc dodecamer sequence in this work, however. Previous work has observed that T7RNA polymerase-polymerase interactions can lead to the displacement of leading polymerase molecules from the DNA sequence [34]. While our model neglects this effect, it can serve as a useful starting point for quantitative analysis of the effects of this phenomenon on IVT process outputs.

Recent trends in IVT reaction engineering have added new relevance to the effect of salts, which disrupt polymerase-promoter binding, on transcription rates. Salt addition has been proposed as a method to shift transcription away from dsRNA impurities, which introduces tradeoffs in the context of RNA manufacturing [31]. In addition, industrial IVT schemes, including fed-batch reactions, increasingly use high NTP and Mg concentrations, which increases salt concentrations. Intuitively, transcription in the elongation-limited regime should be less affected by binding disruptions than transcription in the initiation-limited regime. We find that a simple semi-empirical model can describe trends in K_{MD} as a function of salt concentration (Fig. 2A). Using this relation in combination with the initiation-elongation model, we predict the experimental result that transcription in the high-DNA elongation-limited regime is much less sensitive to salt addition than transcription in the low-DNA initiation-limited regime (Fig. 2B). Understanding this difference in sensitivity can inform reaction design in the context of industrial RNA manufacturing. In addition, we find that salt sensitivity varies between the Fluc and EGFP constructs studied in this work. In the context of our model, this difference is parametrized as a difference in the parameter K_0, which combines binding thermodynamics with the kinetic process of polymerase-DNA disassociation due to abortion (Fig. 2C–Table 2). These differences may be related to the different initiation sequences of the two constructs (SI Section 7). Additional work is required to quantitatively understand how each of the kinetic processes of promoter escape, initiation, and abortion are affected by salt concentrations and how sequence contributes to these salt effects. In addition, while this work neglected the effect of abortion on overall NTP consumption, future kinetic modeling of this abortion step can help to predict the consumption of expensive co-transcriptional capping agents.

A key application of kinetic modeling in the context of IVT is in understanding input-output relationships for impurity formation. In this work, we extend the initiation-elongation model to consider two proposed mechanisms for formation of double-stranded RNA (dsRNA): 3' RNA self-templated transcription and DNA-templated antisense transcription followed by hybridization (Fig. 3A). We evaluate the predictions of these models relative to a macroscopic binding assay that estimated the total concentration of dsRNA in the system. While these assays typically cannot detect small (~40 base pair) regions of dsRNA, they have been shown to correlate with in vivo immune response [35]. As such, we regard them as an effective measurement of the class of dsRNA (i.e., long dsRNA) that is of interest in the manufacturing process.

Understanding the dynamic trends of dsRNA in the IVT reaction is relevant for RNA process development (SI Section 6). The dsRNA fraction in our reaction system was constant or decreasing as a function of reaction conversion, in contrast to the prediction of the RNA-templated model that the dsRNA fraction should increase as more RNA is synthesized (Fig. 3B). The RNA-templated model predicts that adding more DNA to the IVT system should shift the kinetic competition for polymerase molecules away from RNA and proportionally decrease dsRNA formation. In contrast to these predictions, we find that increasing DNA concentrations did not significantly affect final dsRNA fractions (Fig. 3C). While the DNA-templated model does not diverge as dramatically from experimental results, it cannot describe the decrease in dsRNA fraction in early stages of the reaction. In addition, dsRNA fraction data collected by varying the concentration of polymerase enzyme and salt addition show ambiguous results which do not indicate that the predictions of the DNA-templated model are more effective than a constant null hypothesis (SI Section 5). A key approximation of our modeling approach is neglecting the kinetics of hybridization, which may be important for describing trends in these data. These ambiguous

results – that neither the RNA- nor DNA-templated models can quantitatively describe these process data – may owe to approximations made in model formulation, the presence of additional pathways for the formation of dsRNA structures detected by our assay, or the importance of confounding variables such as post-IVT processing in dsRNA detection.

While the presented results do not definitively identify a mechanism for explaining trends in macroscopic dsRNA formation, the modeling approach in this work serves as a platform for both future work and understanding trends in previously reported data. In both the RNA- and DNA-templated models, dsRNA formation is proportional to the ratio of Michaelis-Menten constants of desired and undesired promoter binding, respectively (SI Section 4). This implies that reaction engineering strategies which differentially impact these two bindings can be used to limit dsRNA formation. This modeling observation gives quantitative structure to an array of strategies that previous researchers (with either mechanism in mind) have used to decrease dsRNA formation, including the use of engineered polymerase enzymes [36], high temperatures [37], salts [31], and chaotropic agents [38]. Future work in understanding the kinetic pathways of dsRNA formation and the effects of sequence on binding, initiation, and abortion can add context to results indicating that changes in promoter sequence can effect dsRNA formation [35].

4. Experimental procedures

4.1. In vitro transcription kinetic measurements

All in vitro reactions studied in this work took place and pH 8.0 and contained 5 mM of each NTP (ATP, CTP, GTP, and N1-Methyl-pseudouridine-5′-Triphosphate), 21.075 mM $MgCl_2$, 45 mM of pH 7.9 Tris-HCl buffer, 2 mM spermidine, 10 mM DTT, 6 U/mL of inorganic pyrophosphatase, and 400 U/mL of RNAase inhibitor. All reaction materials were acquired from Hongene Biotech, other than $MgCl_2$, which was acquired from Thermo Fischer. Transcription reactions were assembled at volumes between 50 and 100 μL and incubated at 37 °C. Aliquots of 6 μL were periodically removed and quenched in 60 μL of 50 mM EDTA. These quenched samples were further diluted 36-fold (for a total dilution of 400-fold) and analyzed with the HPLC method of Welbourne et al. [39] to quantify the concentrations of the four NTPs. Linear regression analysis was used to quantify the rate of NTP decay. While an orthogonal analysis of the RNA product was possible, it was found that quantification of NTPs was less sensitive to both systematic and random experimental errors. In order to ensure that data points represented the initial rate of reaction, points collected at high conversion (below 2 mM of the limiting NTP remaining) were excluded from this analysis.

4.2. Measurement of dsRNA concentrations

The quantification of dsRNA was performed using the Lumit® dsRNA Detection Assay kit (Promega) according to the manufacturer's instructions. White 96-well plates were obtained from Thermo Fisher Scientific/Corning®. Diluted reaction samples were prepared as described in the previous method section. The diluted samples were subsequently mixed with the dsRNA assay buffer to achieve an expected final dsRNA concentration of 2 ng/mL per well. Three technical replicates of each reaction sample were measured using a Thermo Fisher Varioskan® Flash plate reader with an integration time of 500 ms, and the results were averaged. Experimental variance between these technical replicates was negligible relative to the variance between replicate reactions. Background luminescence was determined by averaging the readout from the 0 ng/mL dsRNA standard and was subtracted from all sample measurements.

To determine the dsRNA/mRNA fraction, mRNA concentration was quantified via HPLC using 400-fold diluted reaction samples.

4.3. Parameter estimation and model-based design of experiments for kinetic model

Model evaluation and parameter estimation are performed in the Julia language. In order to estimate the relevant kinetic parameters, the maximum likelihood estimate of the vector of parameters p was

$$\min_{p} (y - u(p))^{\top} V_y^{-1}(k)(y - u(p)) \tag{16}$$

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

$$\mathrm{cov}(p^* - p_{\mathrm{true}}) = V_{p^*} \approx \left(S^{\top} V_y^{-1} S\right)^{-1} \tag{17}$$

where p_{true} denotes the true parameters and S is the sensitivity of the model outputs with respect to the vector p.

To ensure precision in the estimated kinetic parameters, model-based design of experiments was performed to minimize the determinant of the estimated parameter covariance matrix, known as D-optimality. Given a hypothesized parameter set $\widehat{p}$ and a prior covariance matrix $\mathrm{cov}(\widehat{p})$, experimental points x were chosen by solving the optimization

$$\min_{x} \left| \left[S(x,\widehat{p})^{\top} V_p^{-1} S(x,\widehat{p}) \right) + \mathrm{cov}(\widehat{p})^{-1} \right]^{-1} \right| \tag{18}$$

where $S(x,\widehat{p})$ is the sensitivity matrix of the experimental points x and the estimated parameters $\widehat{p}$, and $|\cdot|$ is the determinant. For both optimizations (parameter estimation and design of experiments), the gradient-based L-BFGS optimization was performed 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.

CRediT authorship contribution statement

Nathan M. Stover: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marieke De Bock:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Julie Chen:** Software, Methodology, Investigation. **Jacob Rosenfeld:** Methodology, Investigation. **Maria del Carme Pons Royo:** Methodology, Investigation. **Allan S. Myerson:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Richard D. Braatz:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Funding information

This research was supported by the U.S. Food and Drug Administration under the FDA BAA-22-00123 program, Award Number 75F40122C00200.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Drs. Francesco Destro and Christopher Canova for

insightful conversations during the preparation of this manuscript and Andrew Hatas for aid in laboratory experiments. The authors thank Betsy Skrip of MIT CBI for the elements used in the graphical abstract. The authors thank Dr. Rok Sekirnik for correspondence in quantifying and adapting previous data on dynamic dsRNA trends. The authors also thank Prof. Mark Dickman of The University of Sheffield for useful correspondence in adapting the HPLC method used in this work. The authors dedicate this work to Prof. Anthony Sinskey (1940–2025) of MIT for his role as an advisor and mentor.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.abb.2026.110737.

Data availability

Data will be made available on request.

References

[1] A. Ujvári, C.T. Martin, Thermodynamic and kinetic measurements of promoter binding by T7 RNA polymerase, Biochemistry 35 (46) (1996) 14574–14582, https://doi.org/10.1021/bi961165g.
[2] S.I. Gunderson, K.A. Chapman, R.R. Burgess, Interactions of T7 RNA polymerase with T7 late promoters measured by footprinting with Methidiumpropyl-EDTA-Iron(II), Biochemistry 26 (6) (1987) 1539–1546, https://doi.org/10.1021/bi00380a007.
[3] M. Maslak, C.T. Martin, Effects of solution conditions on the steady-state kinetics of initiation of transcription by T7 RNA polymerase, Biochemistry 33 (22) (1994) 6918–6924, https://doi.org/10.1021/bi00188a022.
[4] Q. Guo, R. Sousa, Translocation by T7 RNA polymerase: a sensitively poised brownian ratchet, J. Mol. Biol. 358 (1) (2006) 241–254, https://doi.org/10.1016/j.jmb.2006.02.001.
[5] Y. Jia, S.S. Patel, Kinetic mechanism of transcription initiation by bacteriophage T7 RNA polymerase, Biochemistry 36 (14) (1997) 4223–4232, https://doi.org/10.1021/bi9630467.
[6] I. Kuzmine, C.T. Martin, Pre-steady-state kinetics of initiation of transcription by T7 RNA polymerase: a new kinetic model, J. Mol. Biol. 305 (3) (2001) 559–566, https://doi.org/10.1006/jmbi.2000.4316.
[7] T.A. Steitz, The structural changes of T7 RNA polymerase from transcription initiation to elongation, Curr. Opin. Struct. Biol. 19 (6) (2009) 683–690, https://doi.org/10.1016/j.sbi.2009.09.001.
[8] G.M.T. Cheetham, D. Jeruzalmi, T.A. Steitz, Structural basis for initiation of transcription from an RNA polymerase–promoter complex, Nature 399 (6731) (1999) 80–83, https://doi.org/10.1038/19999.
[9] G.M. Skinner, C.G. Baumann, D.M. Quinn, J.E. Molloy, J.G. Hoggett, Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. A single-molecule view of the transcription cycle, J. Biol. Chem. 279 (5) (2004) 3239–3244, https://doi.org/10.1074/jbc.M310471200.
[10] G.-Q. Tang, V.S. Anand, S.S. Patel, Fluorescence-based assay to measure the real-time kinetics of nucleotide incorporation during transcription elongation, J. Mol. Biol. 405 (3) (2011) 666–678, https://doi.org/10.1016/j.jmb.2010.10.020.
[11] P. Thomen, P.J. Lopez, U. Bockelmann, J. Guillerez, M. Dreyfus, F. Heslot, T7 RNA polymerase studied by force measurements varying cofactor concentration, Biophys. J. 95 (5) (2008) 2423–2433, https://doi.org/10.1529/biophysj.107.125096.
[12] S. Arnold, M. Siemann, K. Scharnweber, M. Werner, S. Baumann, M. Reuss, Kinetic modeling and simulation of in vitro transcription by phage T7 RNA polymerase, Biotechnol. Bioeng. 72 (5) (2001).
[13] J. Boman, T. Marušič, T.V. Seravalli, J. Skok, F. Pettersson, K.Š. Nemec, H. Widmark, R. Sekirnik, Quality by design approach to improve quality and decrease cost of in vitro transcription of mRNA using design of experiments, Biotechnol. Bioeng. 121 (11) (2024) 3415–3427, https://doi.org/10.1002/bit.28806.
[14] Y. Gholamalipour, A. Karunanayake Mudiyanselage, C.T. Martin, 3′ end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in Character—RNA-seq analyses, Nucleic Acids Res. 46 (18) (2018) 9253–9263, https://doi.org/10.1093/nar/gky796.
[15] X. Mu, E. Greenwald, S. Ahmad, S. Hur, An origin of the immunogenicity of in vitro transcribed RNA, Nucleic Acids Res. 46 (10) (2018) 5239–5249, https://doi.org/10.1093/nar/gky177.
[16] N.M. Stover, K. Ganko, R.D. Braatz, Mechanistic modeling of in vitro transcription incorporating effects of magnesium pyrophosphate crystallization, Biotechnol. Bioeng. 121 (9) (2024) 2636–2647, https://doi.org/10.1002/bit.28699.
[17] H.R. Koh, R. Roy, M. Sorokina, G.-Q. Tang, D. Nandakumar, S.S. Patel, T. Ha, Correlating transcription initiation and conformational changes by a single-subunit RNA polymerase with near base-pair resolution, Mol. Cell 70 (4) (2018) 695–706.e5, https://doi.org/10.1016/j.molcel.2018.04.018.
[18] M.B. Elowitz, A.J. Levine, E.D. Siggia, P.S. Swain, Stochastic gene expression in a single cell, Science 297 (5584) (2002) 1183–1186, https://doi.org/10.1126/science.1070919.
[19] C.T. Martin, J.E. Coleman, Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters, Biochemistry 26 (10) (1987) 2690–2696, https://doi.org/10.1021/bi00384a006.
[20] T. Nakano, R. Ouchi, J. Kawazoe, S.P. Pack, K. Makino, H. Ide, T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription, J. Biol. Chem. 287 (9) (2012) 6562–6572, https://doi.org/10.1074/jbc.M111.318410.
[21] K.C. Neuman, E.A. Abbondanzieri, R. Landick, J. Gelles, S.M. Block, Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking, Cell 115 (4) (2003) 437–447, https://doi.org/10.1016/S0092-8674(03)00845-6.
[22] J. Wang, B. Pfeuty, Q. Thommen, M.C. Romano, M. Lefranc, Minimal model of transcriptional elongation processes with pauses, Phys. Rev. E 90 (5) (2014) 050701, https://doi.org/10.1103/PhysRevE.90.050701.
[23] J.A. Kern, R.H. Davis, Application of solution equilibrium analysis to in vitro RNA transcription, Biotechnol. Prog. 13 (6) (1997) 747–756, https://doi.org/10.1021/bp970094p.
[24] J. Szavits-Nossan, R. Grima, Uncovering the effect of RNA polymerase steric interactions on gene expression noise: analytical distributions of nascent and mature RNA numbers, Phys. Rev. E 108 (3) (2023), https://doi.org/10.1103/PhysRevE.108.034405.
[25] L. Jen-Jacobson, L.A. Jacobson, Chapter 2. Role of water and effects of small ions, in: P.A. Rice, C.C. Correll (Eds.), Site-Specific Protein-DNA Interactions, Royal Society of Chemistry, Cambridge, 2008, pp. 13–46, https://doi.org/10.1039/9781847558268-00013.
[26] T. Łoziński, K.L. Wierzchowski, Evaluation of mixed-salt effects on thermodynamic and kinetic parameters of RNA polymerase-promoter DNA complexes in terms of equivalent salt concentrations. General applicability to DNA complexes, Acta Biochim. Pol. 56 (4) (2009) 695–702.
[27] M.T. Record, T.M. Lohman, P. de Haseth, Ion effects on ligand-nucleic acid interactions, J. Mol. Biol. 107 (2) (1976) 145–158, https://doi.org/10.1016/S0022-2836(76)80023-X.
[28] K.A. Sharp, R.A. Friedman, V. Misra, J. Hecht, B. Honig, Salt effects on polyelectrolyte–ligand binding: Comparison of Poisson–Boltzmann, and limiting law/counterion binding models, Biopolymers 36 (2) (1995) 245–262, https://doi.org/10.1002/bip.360360211.
[29] S.W. Chen, B. Honig, Monovalent and divalent salt effects on electrostatic free energies defined by the nonlinear Poisson−Boltzmann equation: application to DNA binding reactions, J. Phys. Chem. B 101 (44) (1997) 9113–9118, https://doi.org/10.1021/jp971521k.
[30] P.L. Privalov, A.I. Dragan, C. Crane-Robinson, Interpreting Protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components, Nucleic Acids Res. 39 (7) (2011) 2483–2491, https://doi.org/10.1093/nar/gkq984.
[31] E. Cavac, L.E. Ramírez-Tapia, C.T. Martin, High-salt transcription of DNA cotethered with T7 RNA polymerase to beads generates increased yields of highly pure RNA, J. Biol. Chem. 297 (3) (2021) 100999, https://doi.org/10.1016/j.jbc.2021.100999.
[32] D. Imburgio, M. Rong, K. Ma, W.T. McAllister, Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants, Biochemistry 39 (34) (2000) 10419–10430, https://doi.org/10.1021/bi000365w.
[33] V.S. Anand, S.S. Patel, Transient state kinetics of transcription elongation by T7 RNA polymerase, J. Biol. Chem. 281 (47) (2006) 35677–35685, https://doi.org/10.1074/jbc.M608180200.
[34] Y. Zhou, C.T. Martin, Observed instability of T7 RNA polymerase elongation complexes can be dominated by collision-induced “Bumping.”, J. Biol. Chem. 281 (34) (2006) 24441–24448.
[35] M. Wolczyk, J. Szymanski, I. Trus, Z. Naz, T. Tame, A. Bolembach, N. R. Choudhury, K. Kasztelan, J. Rappsilber, A. Dziembowski, G. Michlewski, 5′ terminal nucleotide determines the immunogenicity of IVT RNAs, Nucleic Acids Res. 53 (3) (2025), https://doi.org/10.1093/nar/gkae1252.
[36] A. Dousis, K. Ravichandran, E.M. Hobert, M.J. Moore, A.E. Rabideau, An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts, Nat. Biotechnol. 41 (4) (2023) 560–568, https://doi.org/10.1038/s41587-022-01525-6.
[37] M.Z. Wu, H. Asahara, G. Tzertzinis, B. Roy, Synthesis of low immunogenicity RNA with high-temperature in vitro transcription, RNA 26 (3) (2020) 345–360, https://doi.org/10.1261/rna.073858.119.
[38] X. Piao, V. Yadav, E. Wang, W. Chang, L. Tau, B.E. Lindenmuth, S.X. Wang, Double-stranded RNA reduction by chaotropic agents during in vitro transcription of messenger RNA, Mol. Ther. Nucleic Acids 29 (2022) 618–624, https://doi.org/10.1016/j.omtn.2022.08.001.
[39] E.N. Welbourne, K.A. Loveday, A. Nair, E. Nourafkan, J. Qu, K. Cook, Z. Kis, M. J. Dickman, Anion exchange HPLC monitoring of mRNA in vitro transcription reactions to support mRNA manufacturing process development, Front. Mol. Biosci. 11 (2024), https://doi.org/10.3389/fmolb.2024.1250833.