Digital Discovery

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: M. Ahmed, S. Hamed, R. Cardoso, C. Kenyon, M. Pohare, M. Maamra, M. J. Dickman, J. Cordiner and Z. Kis, *Digital Discovery*, 2026, DOI: 10.1039/D5DD00417A.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/digitaldiscovery

View Article Online
DOI: 10.1039/D5DD00417A

Soft sensor based on pH for real-time monitoring of mRNA medicines production

Mahdi Ahmed[1], Shady Hamed[1], Ricardo Cardoso[1], Charley Kenyon[1], Manoj Pohare[1], Mabrouka Maamra[1], Mark Dickman[1], Joan Cordiner[1], Zoltan Kis[1,2*]

[1]School of Chemical, Materials and Biological Engineering, University of Sheffield, Sheffield S1 3JD, UK.

[2]Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK.

*Corresponding author. Email: z.kis@sheffield.ac.uk

Real-time monitoring of *in vitro* transcription (IVT) reactions is critical for enabling continuous manufacturing of high-quality mRNA vaccines and therapeutics for a wide spectrum of diseases. Compared to traditional batch manufacturing, continuous IVT production offers higher throughput, improved consistency, and reduced costs, but requires timely process monitoring to detect deviations and maintain product quality. Since pH is routinely measured in bioreactors, it can serve as a convenient, non-invasive input for real-time monitoring. We present the first IVT soft sensor based on H^+ release during NTP incorporation, using in-line pH data to infer up to 40 otherwise predominantly unobservable species in real time, without requiring additional sensors. Validated against a separate set of offline measurements (not used for model fitting), it delivers updates every 25 milliseconds via two complementary models. The first couples a mechanistic IVT model with an Unscented Kalman Filter (UKF) to dynamically infer ≈ 40 key indicators, including mRNA yield (R^2 = 0.95) and NTP depletion (R^2 = 0.84). The second applies the semi-empirical Henderson–Hasselbalch correlation to reconstruct mRNA yield (R^2 = 0.93) and NTP depletion (R^2 = 0.76) from buffer capac-

View Article Online
DOI: 10.1039/D5DD00417A

ity and pH change alone. This soft sensor enables continuous, real-time process monitoring by generating ≈1,600 concentration estimates per second, supporting Quality-by-Digital-Design and advanced control for continuous, disease-agnostic mRNA medicines manufacturing.

1 Introduction

The versatility of the messenger RNA (mRNA) platform technology resides in the unique ability of mRNA to function as a transient, programmable genetic template that instructs the host cell's translational machinery to produce virtually any protein encoded by the mRNA sequence (*1, 2*). The mRNA platform technology rose to prominence during the COVID-19 pandemic and is now driving the rapid development of hundreds of vaccine and therapeutic candidates against a broad range of diseases, including infectious diseases, cancers, autoimmune diseases, metabolic diseases, rare genetic disorders, and cardiovascular conditions (*3, 1, 4*). This powerful platform is also being harnessed for chimeric antigen receptor (CAR) T cell therapies, protein replacement and supplementation, and genome engineering applications (*1, 4, 5*).

This rapid clinical expansion creates a pressing need for highly productive and scalable multi-product manufacturing platforms capable of delivering high-quality mRNA drug substance in a cost-effective manner (*6, 7, 8*). Achieving such manufacturing advancements requires: (1) multi-product manufacturing capabilities; and (2) rapid, or ideally real-time, monitoring of both process performance and product quality attributes, which is vital for continuous production.

Quality-by-Design (QbD) principles have the potential to enable multi-product manufacturing capabilities by mapping the impact of critical material attributes (CMAs) and critical process parameters (CPPs) onto product critical quality attributes (CQAs) and manufacturing key performance indicators (KPIs). This can establish a multi-product design space defined by CMA and CPP ranges, within which products can be manufactured with optimal KPIs (cost effectively and rapidly) and with the desired CQAs (translating to patient safety and product efficacy) (*6, 9, 10, 11*). *Quality-by-Digital-Design* (QbDD) extends this capability by using mechanistic, data-driven, or hybrid models to guide manufacturing process development and automate the operation of the developed manufacturing process. This involves defining and optimizing the QbDD design space *in silico*

View Article Online
DOI: 10.1039/D5DD00417A

before a manufacturing run is executed and performing real-time optimization, feed-back control and feed-forward control (e.g. model-predictive control) during the operation of the manufacturing process (*6, 9, 11*). The implementation of this QbDD-aided multi-product mRNA manufacturing is currently constrained by the absence of real-time monitoring of manufacturing KPIs and product CQAs. Several spectroscopy-based Process Analytical Technology (PAT) tools have been explored for monitoring of the *in vitro* transcription (IVT) reaction. *In-situ* Raman spectroscopy, Fourier-transform infrared (FTIR) spectroscopy, flow-nuclear magnetic resonance (flow-NMR, aka. online NMR) spectroscopy, for example, resolve spectral signatures of NTPs, PPi and growing RNA chains, enabling chemometric reconstruction of reaction progress (*12, 13, 14*). These spectroscopy methods offer molecular specificity, but lack the required sensitivity and require expensive optics and often labor-intensive calibration and data analysis.

A simpler, more accessible monitoring strategy capable of quantifying both KPIs and CQAs in real-time would offer significant value by ensuring that CQAs and KPIs are kept within specification. Real-time (or near real-time) monitoring is especially important for continuous IVT processes, where raw materials are continuously fed and product is continuously generated. Without timely monitoring, process faults may only be detected hours or days later, potentially resulting in large quantities of off-specification products. Real-time monitoring improves process understanding and enables early detection of deviations, thereby enhancing efficiency, reducing costs, and improving consistency in product quality (*15,16,17,18*). Since pH is already routinely measured in bioreactors, it can serve as a convenient, non-invasive input for such a monitoring approach without requiring additional sensors or measurements.

The IVT reaction is at the core of the mRNA manufacturing process. The optimal outcome of the IVT reaction depends on carefully controlled reaction composition (e.g. NTP:Mg ratio, template DNA concentration, T7 RNA polymerase, etc.), pH, buffer strength, heat and mass transfer (*6, 19, 20, 21*). During IVT, the mRNA is assembled from nucleotide building blocks by bacteriophage enzymes (e.g. T7 RNA polymerase) based on a template DNA (*22, 19*). The T7 RNA polymerase is a formidable molecular machine capable of incorporating 200-250 NTPs per second into the nascent mRNA strand, each NTP incorporation consisting of 3 reversible and 1 irreversible sub-step (*22, 23, 24, 19*). During each NTP addition cycle, the 3′-hydroxyl group of the primer terminus is deprotonated to generate a 3′-O^- nucleophile. This nucleophile attacks the α-phosphorus atom

View Article Online
DOI: 10.1039/D5DD00417A

of the incoming NTP, generating a transient pentacoordinated phosphorane intermediate. In the canonical two-metal-ion mechanism, Mg^{A2+} (which acts as the general base to deprotonate the primer's 3′-OH) and Mg^{B2+} (which stabilizes the triphosphate moiety of the incoming NTP) work together to facilitate catalysis (*25, 26, 27*). The breakdown of the resulting pentacoordinate phosphorane intermediate yields a new phosphodiester bond (Figure S1). This reaction releases both pyrophosphate (PP_i) and a proton into the solution (*28, 29, 30*). Since one proton is released for each NTP incorporated, the cumulative proton release is stoichiometrically linked to NTP consumption. In IVT reactions, which are moderately buffered, this proton release results in a measurable pH change that can be detected with pH meters (*29, 28, 30, 31*). Because IVT models can account for proton balance, the pH trajectory can be linked to reaction progress (*32, 20*). This creates an opportunity for a low-cost IVT monitoring approach that relies on routine pH measurements and requires minimal calibration of both the sensor hardware and the model used to estimate IVT species concentrations.

The in-line, on-line, at-line or off-line measurement of pH in small volumes (e.g. 1 mL or less) can be achieved using micro pH sensors, such as fiber optic pH sensors. These microsensors were originally developed in the late 1970s (*33*) and have been used for rapid cell-culture monitoring (*34, 35, 36*). However, they remain under-used in cell-free RNA production. Several IVT kinetic models that account for pH and proton balance have been reported in the literature. Young *et al.* (*37*) developed an early model of batch IVT that included proton release and buffer equilibria to predict pH changes during transcription. Van de Berg *et al.* (*32*) extended this by building a QbD IVT model that accounts for NTP-Mg complexation and buffer speciation, enabling design space exploration. More recently, Ahmed *et al.* (*20*) developed a comprehensive mechanistic IVT model that captures ≈40 reaction and buffer species, including detailed Mg^{2+} complexation, pyrophosphate hydrolysis, and multiple buffer equilibria, while also accounting for enzyme kinetics and template-specific nucleotide composition. The key advance of the present work is the integration of such a mechanistic model into a real-time state estimation framework (Unscented Kalman Filter), enabling reconstruction of a high-dimensional reaction state from in-line pH measurements.

Here, we embed an in-line pH microprobe directly in the IVT reaction mixture, streaming frequent pH data to a soft sensor that computes real-time estimates of (i) individual NTP depletion (consumption), (ii) mRNA titer (yield) (both validated against offline assays), (iii) reaction rate,

View Article Online
DOI: 10.1039/D5DD00417A

and (iv) additional state variables predicted by the mechanistic model (e.g., magnesium–nucleotide complexes, pyrophosphate, orthophosphate and buffer species such as protonated and deprotonated forms of HEPES or TRIS). Two complementary computational models are evaluated using this soft sensor: (1) a mechanistic IVT model coupled to an Unscented Kalman Filter (UKF); (2) a semi-empirical Henderson–Hasselbalch (H-H) correlation that exploits buffer capacity and pH alone with no need for kinetic parameterization. These two approaches serve different but complementary purposes. The mechanistic IVT-UKF model can predict the future concentrations of a large number of reaction species and is regularly updated during the IVT reaction using pH measurements through the UKF. This enables estimation of otherwise unmeasurable reaction states and provides a framework suitable for advanced monitoring and future model-based control. However, it requires parameterization of kinetic models and higher computational effort. In contrast, the H-H approach relies only on pH measurements and buffer chemistry, making it simple, requiring minimal calibration, and easy to deploy with minimal computational requirements. However, it captures a smaller number of IVT species, provides more limited mechanistic insight, and does not enable forecasting of future concentration values. Using both models in parallel therefore provides a balance between practicality and predictive capability: the H-H model offers a minimal-data, rapid-deployment monitoring option, while the mechanistic IVT-UKF model provides deeper process insight and improved state estimation.

We demonstrate the application of this soft sensor in IVT reactions producing enhanced green fluorescent protein (eGFP) and SARS-CoV-2 spike protein (CSP) mRNA under two widely used buffer conditions (HEPES and TRIS). Predictions from both models are benchmarked against offline measurements of RNA yield and NTP depletion (hereafter used interchangeably with "NTP consumption" and "RNA titer," respectively), validating the utility of pH-driven soft sensing for real-time IVT monitoring.

Crucially, this work goes beyond conventional soft sensing and PAT approaches by demonstrating that a single, routinely measured variable (pH) can encode sufficient information to reconstruct a high-dimensional biochemical reaction network in real time, establishing a new paradigm in process analytics in which information is extracted from the underlying reaction stoichiometry rather than from multiple or complex sensor inputs. While the framework infers ≈40 IVT and buffer species through mechanistic relationships, only a subset (RNA yield and individual NTP concentrations)

View Article Online
DOI: 10.1039/D5DD00417A

is directly validated against independent experimental measurements.

2 Materials and Methods

2.1 *In Vitro* transcription reactions

IVT reactions were carried out to produce mRNA encoding either eGFP or CSP. Linearized plasmid DNA templates for both eGFP and CSP were supplied (by GenScript Biotech Corporation) at a final concentration of $0.05\,\mu g\,\mu L^{-1}$. Four reaction variants were obtained by pairing each of the two templates with one of two buffer systems using either HEPES (pH adjusted with NaOH, catalog number: Gibco^{TM} 15630080) or TRIS (pH adjusted using acetic acid) in a total volume of 1 mL. Each 1 mL reaction contained 10 mM ATP, CTP, GTP and UTP (Roche Diagnostics GmbH); T7 RNA polymerase at $330\,U\,\mu L^{-1}$ (Roche Diagnostics GmbH); inorganic pyrophosphatase at $0.05\,U\,\mu L^{-1}$ (Roche Diagnostics GmbH) to suppress pyrophosphate precipitation; and RNase inhibitor at $1\,U\,\mu L^{-1}$ (Roche Diagnostics GmbH) to protect nascent mRNA. Reactions were prepared in either a HEPES buffer (40 mM HEPES, pH 7.3; 42 mM magnesium acetate; 10 mM dithiothreitol; 2 mM spermidine; 0.01 % Triton X-100; 50 mM NaCl) or a TRIS buffer (40 mM TRIS, pH 7.9; 42 mM magnesium acetate; 10 mM dithiothreitol; 2 mM spermidine; 0.01 % Triton X-100; 50 mM NaCl). Reactions were incubated at 37 °C for up to 120 minutes.

2.2 In-line pH measurements

A fiber-optic pH system equipped with an IMP-HP5 fiber optic microsensor (World Precision Instruments) featuring a 250 µm diameter tip was used for pH measurements. Calibration was performed with standard buffers at pH 5.5, 6.5, and 7.1 according to the manufacturer's protocol. For each IVT reaction, 950 µL of the transcription mixture (minus DNA) was equilibrated in a 1.5 mL tube with the pH microsensor and temperature probe for 15 minutes. The remaining 50 µL of DNA template (pre-incubated at 37 °C) was then added to initiate the IVT reaction. The pH and temperature were recorded every 30 seconds at 37 °C via the vendor's software. This interval was chosen to align with the sensor's response time (of up to 30 seconds), ensure sufficiently frequent updates for the kinetic model, and prolong sensor lifespan relative to higher-

View Article Online
DOI: 10.1039/D5DD00417A

frequency measurements. The 30-second measurement interval is sufficiently frequent to capture all meaningful changes in the IVT reaction dynamics, which occur on the order of minutes. Recorded pH profiles, reflecting proton release from the 3′-OH of the ribose during NTP incorporation, were fed into both the mechanistic IVT model and a H-H-based mass balance soft sensor (Section 2.9) to estimate NTP consumption and RNA yield in real-time.

2.3 Sampling for offline quantification of NTP and RNA concentrations

Aliquots (20 µL) were withdrawn at defined intervals (see Table S1) to monitor reaction progress via offline anion-exchange high-performance liquid chromatography (AEX-HPLC), fluorometry, and UV-vis spectroscopy, enabling simultaneous quantification of NTPs and RNA (*31*). Sampling intervals were selected to correspond to approximately equal increments in pH decrease during the reaction, ensuring representative coverage of transcription kinetics. Each aliquot was immediately mixed with 20 µL of 100 mM EDTA to chelate Mg^{2+} and halt transcription. Samples were then kept on ice or stored at –20 °C until analysis.

2.4 RNA quantification by UV spectroscopy

Samples were purified using silica-based spin columns with the Monarch® RNA Cleanup Kit (NEB), diluting each 1:2.5 in RNase-free water before following manufacturer's instructions. Eluted RNA was quantified on a NanoDrop™ One^{c} (Thermo Fisher Scientific) at 260 nm, with A_{260}/A_{280} and A_{260}/A_{230} ratios checked for purity.

2.5 RNA quantification by fluorometry

RNA concentration was quantified by fluorometry using the Qubit^{TM} RNA XR (Extended Range) Assay Kit on a Qubit^{TM} 4 Fluorometer (Invitrogen, Thermo Fisher Scientific), according to the manufacturer's instructions. A working solution was prepared by diluting the Qubit^{TM} RNA XR Reagent 1:200 in the corresponding buffer. The instrument was calibrated using two standards, prepared by mixing 10 μL of the kit-provided Standard #1 (0 ng/μL in TE Buffer) and Standard #2 (1000 ng/μL in TE Buffer) with 190 μL of the working solution. For sample quantification, 2 μL

View Article Online
DOI: 10.1039/D5DD00417A

of each RNA sample was added to 198 μL of the working solution. All tubes were briefly vortexed and incubated at room temperature for 2 minutes prior to fluorescence measurement.

2.6 RNA and NTP quantification by AEX-HPLC

NTPs and RNA quantity were assessed on a Thermo Fisher U3000 system equipped with a DNAPac PA200 column (50 mm x 2.1 mm i.d.), following Welbourne *et al.* (*31*). Crude IVT samples were diluted 200–500× in RNase-free water and separated using 25 mM TRIS (pH 8.0) as Buffer A and 25 mM TRIS with 1 M NaCl (pH 8.0) as Buffer B, at a flow rate of 0.25 mL min^{-1}. UV detection at 260 nm captured both NTP and RNA peaks.

2.7 RNA size and integrity analysis

The size and integrity (intactness) of RNA were measured by capillary gel electrophoresis (CGE) using the 5200 Fragment Analyzer instrument (Agilent), together with the DNF-471 RNA Kit (15 nt) (Agilent). The kit includes an RNA separation gel, dsDNA inlet buffer (DNF-930, Agilent), TE rinse buffer, intercalating dye, RNA diluent marker (15 nt), RNA ladder (200–6,000 nt), and capillary conditioning solution. A 33 cm FA 12-Capillary Array Short cassette (Agilent) was employed. RNA was detected via fluorescence, with signal intensity reported in relative fluorescence units (RFU), as the intercalating dye in the separation gel binds specifically to RNA (*10*).

2.8 Materials and instrumentation

All reagents were analytical grade or higher: NTPs, enzymes, and buffers (Roche Diagnostics); linearized eGFP and CSP templates (GenScript); RNase-free water and cleanup kits (New England Biolabs). pH was monitored in real-time using a fiber-optic microsensor and RNA concentration by NanoDrop™ One^{c} UV-Vis Spectrophotometer. All reactions were held at 37 °C in a temperature-controlled block (1.5 mL Eppendorf tubes), with the pH probe calibrated as per the manufacturer's instructions.

View Article Online
DOI: 10.1039/D5DD00417A

2.9 Soft Sensor based on the kinetic model updated by the Unscented Kalman Filter

We implement the full mechanistic IVT model of Ahmed *et al.* (*20*), including all reaction species and kinetic steps, and integrate it with pH data from a fiber-optic probe to realize a soft sensor for real-time reaction monitoring. For brevity, this manuscript details only the subset of rate equations governing proton release and buffering that directly inform the pH measurement; the complete model formulation and parameter values follow Ahmed *et al.* and is an improved version of our previously published IVT kinetic QbD model (*32*). We then embed this model in a UKF, using each incoming pH reading (at 30 second intervals) as an indirect observation of overall reaction progress to iteratively correct the full state vector (NTPs, RNA, enzyme–substrate complexes, etc.) and improve the accuracy of real-time predictions.

2.9.1 IVT Model

Each NTP (ATP, CTP, GTP, UTP) is incorporated into the growing RNA chain, transitioning from an n-mer (RNA_n) to an $(n+1)$-mer (RNA_{n+1}^{-}), with pyrophosphate (PP_i^{4-}) and a free proton (H^+) also released:

$$\mathrm{NTP}^{4-} \;+\; \mathrm{RNA}_n \;\longrightarrow\; \mathrm{RNA}_{n+1}^{-} \;+\; \mathrm{PP}_i^{4-} \;+\; \mathrm{H}^+. \tag{1}$$

Accordingly, the net consumption of NTP and production of protons can be captured by

$$\frac{d[\mathrm{NTP}]}{dt} \;=\; -V_{\mathrm{tr}} \cdot N_{\mathrm{all}}, \tag{2}$$

$$\frac{d[\mathrm{H}^+]}{dt} \;=\; V_{\mathrm{tr}} \cdot N_{\mathrm{all}}, \tag{3}$$

where V_{tr} is the transcription rate and N_{all} is the total moles of NTP consumed per unit time, reflecting the specific template composition (Table S2).

The released protons are buffered by HEPES and other ionic complexes. For instance, the HEPES equilibrium is

$$\mathrm{HEPESH} \;\rightleftharpoons\; \mathrm{HEPES}^{-} \;+\; \mathrm{H}^+, \tag{4}$$

View Article Online
DOI: 10.1039/D5DD00417A

while the total proton balance involves multiple equilibria:

$$H_{\text{tot}} = [H^+] + [HNTP^{3-}] + [HPPi^{3-}] + 2\,[H_2PPi^{2-}] + [MgHNTP^-] + [MgHPPi^-] + [HHEPES] + [HPi^{2-}] + 2\,[H_2Pi^-] + 3\,[H_3Pi] + [MgHPi] + [HACET]\,. \tag{5}$$

All bound-species terms are expressed in free variables via dissociation constants; the complete set of equilibrium constants, kinetic parameters, and initial conditions is provided in (*20*). For systems with alternative buffering (e.g. TRIS), the corresponding equilibria can be trivially derived and adapted to the model as needed.

We collect the system's *differential* states in $\mathbf{x}(t)$ and the *algebraic* states in $\mathbf{z}(t)$. The model thus forms a set of differential-algebraic equations (DAEs),

$$\frac{dx}{dt} = f(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}), \tag{6}$$

$$\mathbf{0} = g(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}), \tag{7}$$

where $\boldsymbol{\theta}$ denotes the kinetic and equilibrium parameters. The UKF updates all model states in real-time (at 25 millisecond intervals) based on pH measurements recorded every 30 seconds, thus enabling a soft sensor for IVT progress. The list of state variables included in the IVT kinetic model (encompassing both IVT species) is shown in Table S3.

2.9.2 Unscented Kalman Filter

The UKF is particularly suited for the IVT model as it exhibits highly non-linear behavior that can be difficult to capture via simple linearization (as in an Extended Kalman Filter). The UKF estimates the joint distribution of $(\mathbf{x}, \mathbf{z})$ by generating *sigma points*, which are chosen vectors in state space that capture the mean and covariance of the current estimate. m refers to the *dimension* of the state, meaning that the filter produces $2m + 1$ sigma points for m state variables.

2.9.3 Sigma-point generation and prediction

At time t_k, the UKF forms $2m+1$ sigma points around the mean state $(\hat{\mathbf{x}}_{k|k}, \hat{\mathbf{z}}_{k|k})$ and the covariance $\mathbf{P}_{k|k}$. Each sigma point is advanced to t_{k+1} by integrating the system. The resulting predicted sigma points are then averaged to obtain $\hat{\mathbf{s}}_{k+1|k}$ and $\mathbf{P}_{k+1|k}$.

View Article Online
DOI: 10.1039/D5DD00417A

2.9.4 Updating the model with pH measurements

A fiber optic pH probe reports the measured pH at discrete time intervals t_k. Denoting these measurements as

$$\mathrm{pH}_{\mathrm{meas},k} \;=\; -\log_{10}\big([\mathrm{H}^+]\big) \;+\; v_k, \tag{8}$$

we assume that v_k is Gaussian measurement noise with mean zero and variance R. Since $[\mathrm{H}^+]$ depends on both $\mathbf{x}$ and $\mathbf{z}$, we view pH as a one-dimensional observation $h(\mathbf{x}, \mathbf{z})$. When a new pH value, $\mathrm{pH}_{\mathrm{meas},k+1}$, is received, each predicted sigma point is mapped through the measurement function

$$\mathrm{pH}^{(i)}_{k+1|k} \;=\; -\log_{10}\Big([\mathrm{H}^+]\big(\mathbf{x}^{(i)}_{k+1|k},\, \mathbf{z}^{(i)}_{k+1|k}\big)\Big). \tag{9}$$

From the ensemble $\{\mathrm{pH}^{(i)}_{k+1|k}\}$, the UKF computes the updated mean and covariance in measurement space, calculates the cross-covariance with the state, and applies the Kalman gain to refine $(\hat{\mathbf{x}}_{k+1|k+1},\, \hat{\mathbf{z}}_{k+1|k+1})$ and $\mathbf{P}_{k+1|k+1}$. Figure S2 depicts the overall algorithmic flow.

In a Kalman filtering framework, "process noise" ($\mathbf{Q}$) represents uncertainty in the underlying model. Although the IVT reactions are deterministic in principle, real-world operation always deviates from an idealized model, so the instantaneous differential states (e.g., NTP, RNA) will not follow the nominal dynamics exactly. In an ODE-based filter one can freely perturb all states; in a DAE system, however, arbitrary noise in the algebraic variables $\mathbf{z}(t)$ risks violating

$$g\big(\mathbf{x},\, \mathbf{z},\, \boldsymbol{\theta}\big) = 0.$$

We inject process noise only into the differential states $x(t)$ and selected kinetic parameters in θ. To ensure the noise scales with each state's order of magnitude using a single tuning parameter, we define

$$\mathbf{Q} \;=\; \varepsilon^2\, \mathrm{diag}\big(x_{\mathrm{nom}}\big)^2,$$

where x_{nom} is the vector of nominal state magnitudes and ε is a relative noise coefficient. We then fine-tune ε empirically by comparing short-interval model predictions to experimental pH measurements, balancing trust between the model dynamics and the sensor data.

The UKF is implemented in Python using FilterPy, with IDA/Assimulo solving the underlying DAE and a 30 second measurement interval for real-time tracking. Calibration is performed by iteratively adjusting four species to align the UKF's predicted distributions with offline pH analytics:

View Article Online
DOI: 10.1039/D5DD00417A

process noise (**Q**) assigned to key differential states or kinetic constants; measurement noise (**R**) inferred from the fiber-optic probe's empirical precision; unscented transform parameters (α, β, κ) tuned to control sigma-point spread; and the initial covariance ($\mathbf{P}_0$) set from prior state uncertainty. The kinetic model uses fixed kinetic parameters, calibrated against experimental data prior to runtime (these parameters need re-fitting when a new RNA molecule is produced based on a new template DNA); only state variables are updated by the UKF during operation. A summary of the final UKF parameter values appears in Table S4.

2.10 Soft sensor based on the semi-empirical Henderson–Hasselbalch correlation

The semi-empirical H–H equation was used to calculate NTP consumption in IVT reactions by converting pH changes to proton release, knowing that one H^+ is released per incorporated NTP. The list of IVT species computed by the H-H model is shown in Table S3.

At discrete times t_k, let

$$pH_k = -\log_{10}(c_{H,k}), \quad c_{H,k} = 10^{-pH_k}, \quad c_{H,0} = 10^{-pH_0}.$$

The change in proton concentration relative to t_0 is

$$\Delta c_{H,k} = c_{H,k} - c_{H,0}. \tag{10}$$

We relate pH_k to the HEPES buffer equilibrium via the H-H equation:

$$pH_k = pK_a(T) + \log_{10}\left([A^-]_k / [HA]_k\right), \tag{11}$$

Knowing that each NTP incorporation liberates one proton, the cumulative NTP count at step k is

$$n_{\mathrm{NTP},k} = V_{\mathrm{rxn}}\, \Delta c_{H,k}, \tag{12}$$

where V_{rxn} is the constant reaction volume.

The H–H correlation was also used to calculate the conjugate-base concentration [A^-] as a function of pH for commonly used IVT buffers (HEPES, TRIS–acetate, TRIS–HCl, HEPES titrated with NaOH, and TRIS–EDTA) as well as for various HEPES and TRIS–acetate buffer

View Article Online
DOI: 10.1039/D5DD00417A

concentrations (20-160 mM). For this, the pH range was discretized into fine intervals (0.001 pH units), and [A^-] was computed within each interval. Then the sum of [A^-] within each range was plotted as function of pH.

3 Results

3.1 Concept and architecture of the pH-driven soft sensor for real-time IVT monitoring

To enable real-time monitoring of IVT reactions, we developed a soft sensor that converts a simple, non-invasive pH signal into dynamic estimates of critical IVT and buffer species concentrations. The concept is illustrated in Figure 1 and comprises three integrated modules: (i) frequent pH measurement in the IVT reactor, (ii) digital processing via two complementary models, and (iii) output generation of key state variables. In the bioprocess setup, a fiber-optic pH microsensor is embedded directly into the IVT reactor, to record pH every 30 seconds. This pH signal reflects the stoichiometric release of protons from the 3′-OH of the ribose during each NTP incorporation into the growing RNA strand (cf. Figure S1), at a rate of up to 250 protons released per second (*22, 23, 24, 19*). Therefore, the temporal trajectory of pH serves as a real-time proxy for RNA synthesis and NTP depletion. In the soft sensor engine, the recorded pH values are processed using two complementary approaches. The first is a mechanistic kinetic IVT model, embedded within a UKF, which assimilates each incoming pH value to iteratively update the predicted concentrations of ≈40 species involved in the IVT reaction, including RNA, individual NTPs, Mg^{2+}:NTP complexes, pyrophosphate, orthophosphate, and buffer species. Model predictions are updated every 25 milliseconds, allowing high-frequency state estimation despite slower pH measurement rates (30 second frequency). The second approach is based on the semi-empirical H–H correlation, which calculates RNA yield and NTP consumption from the buffer equilibrium and pH trajectory, without requiring kinetic parameterization. This can enable rapid estimation of ≈15 IVT species. The output layer of the soft sensor delivers continuous, real-time predictions of KPIs, including RNA yield, NTP consumption, reaction rate, as well as IVT and buffer component dynamics, accounting for a total of ≈15 (H–H model) and ≈40 (kinetic model) IVT and buffer species, respectively, which are inferred

View Article Online
DOI: 10.1039/D5DD00417A

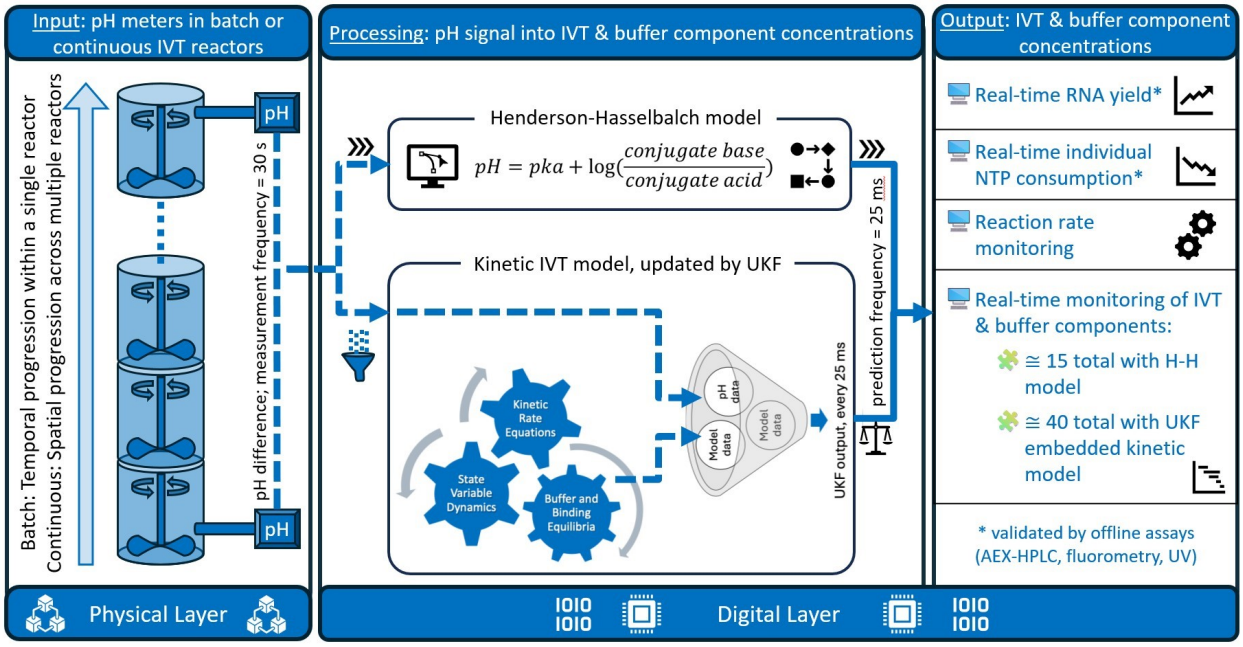


Figure 1: **Overview of the pH-driven soft sensor for continuous, real-time monitoring of IVT reactions.** This soft sensor integrates frequent in-line pH measurements with two computational models to enable continuous, real-time estimation of IVT and buffer species concentrations. The left panel (Physical Layer, bioprocess setup) consists of batch or continuous IVT reactors equipped with a fiber-optic pH microsensor. pH is measured every 30 seconds and serves as a proxy for proton release, which is stoichiometrically linked to NTP consumption and RNA synthesis. In the middle panel (Digital Layer, soft sensor engine), pH data are processed by either a Henderson–Hasselbalch model, which estimates buffer species concentrations and RNA yield based on acid–base equilibria, or a mechanistic IVT model embedded within an Unscented Kalman Filter, which dynamically updates the predicted states of ≈40 molecular species (NTPs, RNA, Mg-complexes, and buffer species) every 25 milliseconds. Please note that the estimated species are inferred through the mechanistic model constrained by pH measurements; only a subset (RNA yield and NTP consumption) is directly validated experimentally. The right panel (Digital Layer, output) includes continuous, real-time predictions of RNA yield, individual NTP depletion, reaction rate, and IVT and buffer component concentrations. Key predictions from both models are validated against offline assays (AEX-HPLC, UV spectrophotometry, fluorometry). This soft sensing framework supports advanced continuous process monitoring and control in mRNA, self-amplifying RNA (saRNA), and circular RNA (circRNA) manufacturing.

View Article Online
DOI: 10.1039/D5DD00417A

through the mechanistic framework constrained by pH measurements. Among these, RNA yield and NTP consumption are directly validated experimentally, using offline assays (AEX-HPLC, UV absorbance, and fluorometry). Together, these three modules form a model-based digital sensing framework that enables real-time tracking and prediction of IVT reaction progression, supporting QbDD, process development, and advanced control strategies.

3.2 Measured and model-predicted real-time pH as a proxy for IVT progress

The inherently low buffering capacity of IVT reactions (see Figure S3), arising from the use of weak buffers such as HEPES or TRIS at modest concentrations (e.g., 40 mM), results in a measurable pH drop as protons are released during the IVT. Figure 2 shows the in-line pH traces recorded every 30 seconds during IVT of eGFP (930 nt) and CSP (4283 nt) mRNA in HEPES and TRIS buffers. In the same figure, the overlaid dashed lines represent 25 millisecond interval predictions from the kinetic model tuned by the UKF, which closely track the measured pH trajectories while effectively filtering out sensor noise. The individual and averaged pH measurements, alongside the recorded temperature and RNA yield (measured by both UV spectrophotometry and fluorometry) are shown in Figure S4.

In every case, the pH drops sharply upon initiation, mirroring the NTP uptake and RNA produced. This indicates that pH provides a direct, non-invasive proxy for transcriptional progress. The period of maximum catalytic activity is clear in the first 30 minutes, where all four curves fall steeply. As NTPs become limiting, the slope flattens. Notably, larger net pH drop (especially in the eGFP–TRIS run) is associated with increased raw sensor noise. This is likely due to the measured pH values approaching the lower measurement limit (pH ≈ 5.5) specified by the manufacturer.

Buffer choice dictates the magnitude of the drop. TRIS ($pK_a \approx 7.8$ at 37 °C) lies farther from the initial reaction pH (≈ 6.67) than HEPES ($pK_a \approx 7.3$), so its buffering capacity is lower. RNA length modulates curve shape: the longer CSP mRNA transcript maintains an almost linear decline, indicative of either initiation, elongation or termination limitation when transcribing CSP mRNA, whereas eGFP mRNA shows a more pronounced initial fall. Based on H-H modeling results, a pH drop is expected in all commonly used IVT buffers (see Figure S3) and even at increased buffer concentrations (see Figure S5).

View Article Online
DOI: 10.1039/D5DD00417A

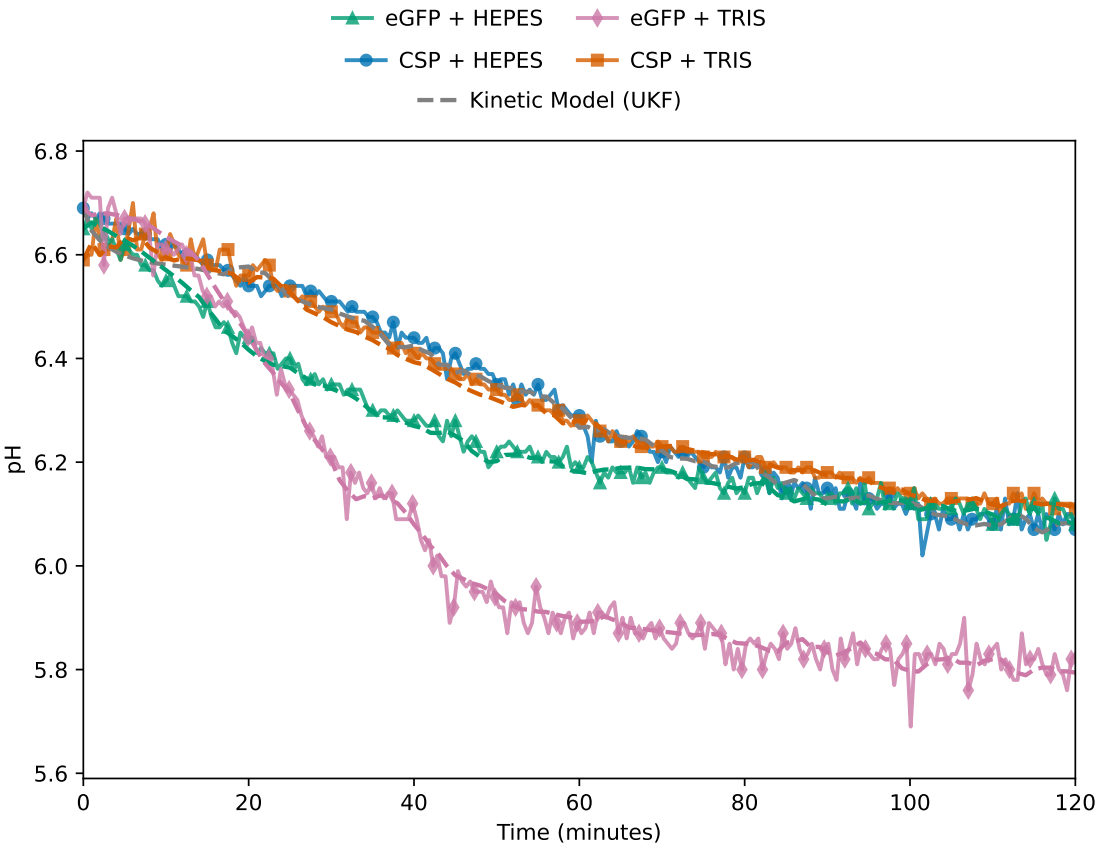


Figure 2: **Time-course of pH profiles during IVT reactions.** Solid lines with markers show measured pH every 30 seconds: blue circles for eGFP mRNA in HEPES buffer, orange squares for eGFP mRNA in TRIS buffer, green triangles for CSP mRNA in HEPES buffer, and pink diamonds for CSP mRNA in TRIS buffer. Overlaid dashed lines are pH trajectories predicted by the kinetic model and dynamically updated using a UKF, which assimilates experimental measurements to account for process noise and parameter uncertainty, improving prediction accuracy.

View Article Online
DOI: 10.1039/D5DD00417A

3.3 Real-time RNA quantification by the soft sensors based on the kinetic-UKF and H-H models

The principal goal of this work was to move beyond qualitative pH trends toward a quantitative, real-time soft sensor for monitoring RNA yield, NTP consumption, reaction rate, as well as IVT and buffer component dynamics, accounting for a total of ≈15 and ≈40 species with the H-H and kinetic-UKF models, respectively. To accomplish this, we integrated the IVT model with a UKF. This framework uses pH measurements (recorded every 30 seconds) to update and refine real-time model predictions (generated every 25 milliseconds) of key molecular species, such as RNA and individual NTP concentrations, among all other species modeled in the reaction.

As illustrated in Figure 3, the UKF-predicted RNA concentrations closely paralleled experimental measurements, with $R^2 > 0.90$ in every case. Comparing the two models across the four conditions, the H-H model generally tracks the experimental data well and in some cases (e.g., Figure 3A) appears to follow the experimental trend more closely than the UKF during mid-reaction; however, the UKF achieves a higher overall R^2 because it can dynamically correct its trajectory as new pH measurements arrive. The kinetic-UKF model is particularly advantageous in conditions where the reaction deviates from simple proton-balance assumptions (e.g., Figure 3D, CSP-TRIS), where the UKF's adaptive correction yields $R^2 = 0.983$ compared to 0.966 for the H-H model.

This confirms that a single, frequent pH measurement can reliably track overall reaction progression, substantially reducing the need for frequent, labor-intensive offline sampling that gives time-lagged readings and is subject to cumulative manual errors from aliquot withdrawal, EDTA quenching, spin-column purification, and instrument calibration. Indeed, the RMSE values for UKF-predicted RNA yield (0.40–1.15 $\mathrm{g\,L^{-1}}$, Table 1) are comparable to or smaller than the standard deviations of the offline UV absorbance measurements themselves, suggesting that a significant fraction of the apparent prediction "error" may originate from variability in the offline reference assay rather than from the soft sensor. Because the soft sensor derives its estimates from a single, continuous pH signal that bypasses all sample-handling steps, it can provide not only higher-frequency monitoring but also potentially more consistent estimates of reaction progress.

Occasional deviations of the UKF mean from individual experimental points align with regions of higher sensor noise or offline assay variability rather than a systematic model error. In addition,

View Article Online
DOI: 10.1039/D5DD00417A

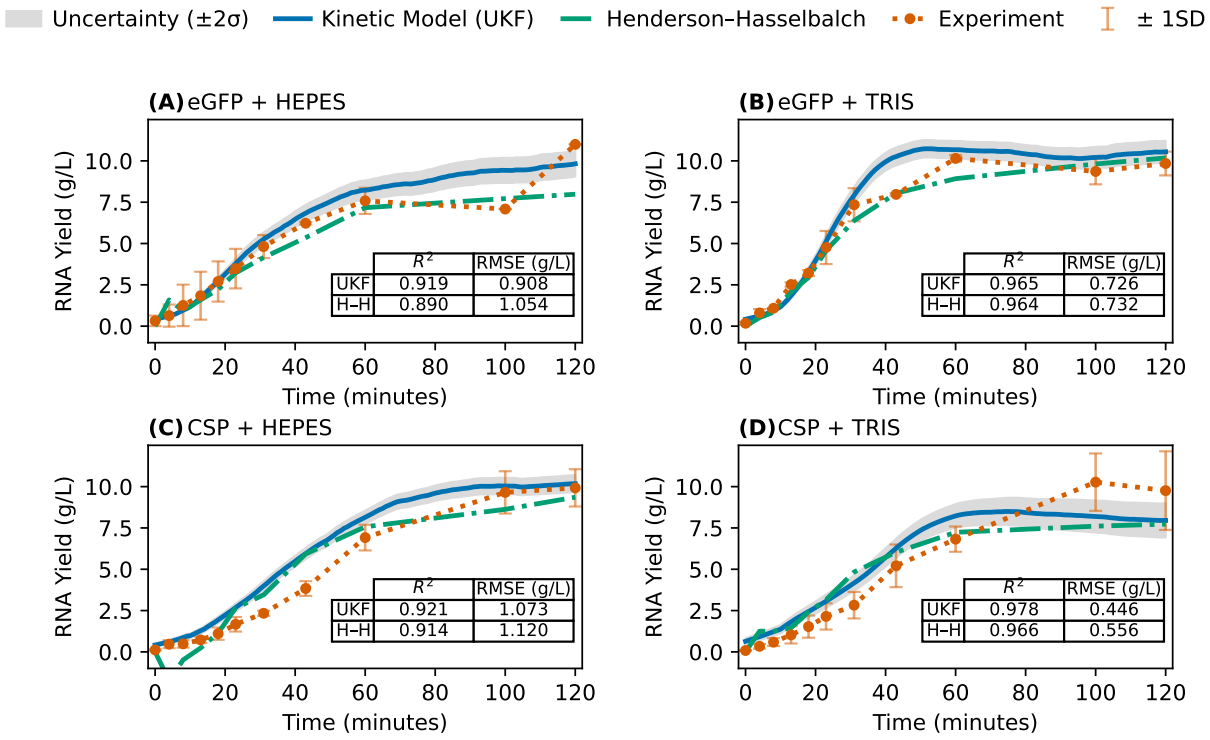


Figure 3: **Continuous, real-time RNA yield under four conditions** (A) eGFP mRNA in HEPES buffer, (B) eGFP mRNA in TRIS buffer, (C) CSP mRNA in HEPES buffer, (D) CSP mRNA in TRIS buffer. RNA yields are shown from 0 to 120 minutes following IVT initiation by the addition of template DNA. Red circles indicate experimental measurements obtained from RNA purified via silica spin columns and quantified by UV absorbance at 260 nm; error bars represent ± 1 SD from biological replicates ($n = 2$ for eGFP, $n = 3$ for CSP). The solid blue line shows the continuous pH prediction by the kinetic model, dynamically updated by the UKF, and the gray shaded band denotes $\pm 2\sigma$ confidence intervals. The green dash–dotted line shows the H-H estimates. Inset tables report the coefficient of determination (R^2) and root-mean-square error (RMSE) for each method in each condition.

View Article Online
DOI: 10.1039/D5DD00417A

some data points (e.g. at 100 and 120 minutes in Figure 3A) may be affected by experimental or analytical bias. Despite low standard deviation, this does not exclude systematic error (e.g. dilution or sample handling error), which would not necessarily be reflected in the error bars. The limited number of replicates further restricts definitive interpretation. The $\pm 2\sigma$ uncertainty bands around the UKF traces (largest in the eGFP–TRIS) are drawn directly from the filter's state-covariance update and mirror increases in pH measurement noise when the reaction drifts outside the sensor's optimal range. Notably, in the CSP–TRIS run the final RNA yield fell below the mechanistic model's initial forecast; the UKF responded by gradually adjusting its prediction downward from 60 to 120 minutes as the real-time pH began to diverge from the model's expected trajectory, illustrating the filter's ability to detect and correct systematic deviations. For additional details on the filter's adaptive weighting (process covariance trace and Kalman gain), see SI Figure S6A&B. Both the eGFP and CSP mRNA produced in the two different IVT buffers were of high integrity (intactness), as evidenced by CGE measurements, see Figure S7.

We also employed a H–H approach to estimate RNA yield solely based on proton balance and buffer equilibrium. As shown in Figure 3, overlaying the H–H predictions on the experimental data yields an average coefficient of determination of $R^2 > 0.90$ across all conditions, despite the absence of any fitted parameters or adjustments to the data by a Kalman filter. Under the assumption of perfectly accurate pH measurements, the H–H curve can be regarded as an internal "ground truth," such that any systematic deviation of the experimental RNA yields from this curve may reflect assay noise or bias. This interpretation is reinforced in Figure 3C: during the 20–60 minute window, which coincides with a large pH drop, both the H–H and UKF models overestimate the measured RNA, whereas the raw data remain comparatively flat, suggesting sampling artifacts rather than model failure. Of course, pH itself is challenging to measure with absolute accuracy and precision, and neither model captures every possible side reaction or ionic interaction. Nonetheless, the high R^2 values between the RNA yield predicted by the model and that measured by offline analytical assays indicate that the pH change reflects the proton balance associated with RNA formation in both models (kinetic-UKF and H-H). This shows that pH captures the dominant proton-release chemistry and supports its use as a robust, non-invasive input for the soft sensor.

View Article Online
DOI: 10.1039/D5DD00417A

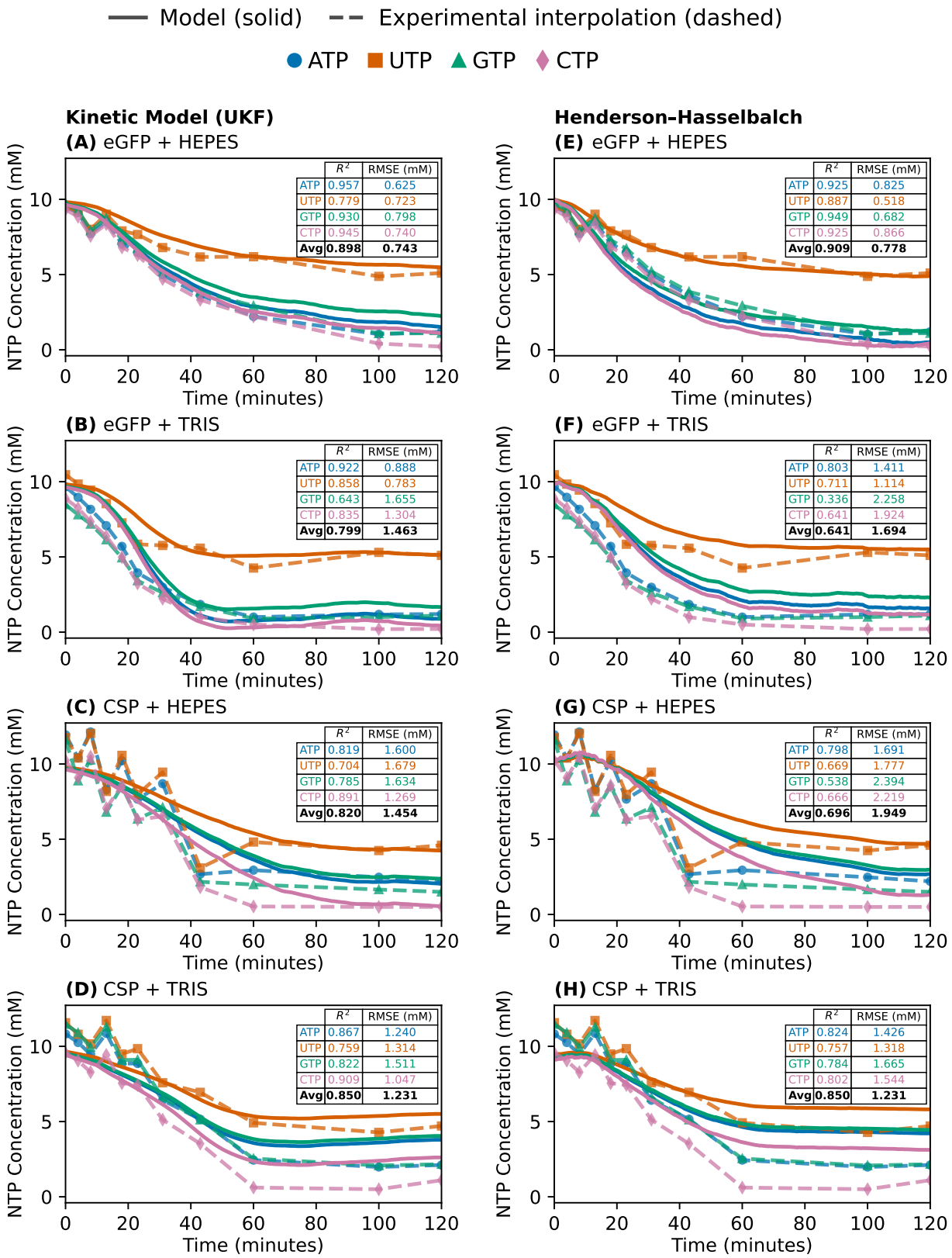
3.4 Real-time NTP quantification by the kinetic-UKF soft sensor

Given the UKF's strong performance in predicting RNA yield, it is reasonable to expect similarly accurate estimates for the depletion of the four NTPs (ATP, CTP, GTP, and UTP), which serve as the substrates for RNA synthesis. Figure 4 confirms this: the trajectories for ATP, CTP, GTP, and UTP predicted by the kinetic model and dynamically updated by the UKF closely track the offline AEX-HPLC measurements, yielding an average $R^2 \approx 0.84$. Although this R^2 is lower than that for RNA (see Figure 3), the discrepancy arises almost entirely from higher experimental error in the AEX-HPLC assay, as robust quantification of NTPs is notoriously challenging. In principle, given the known template sequence and proton stoichiometry, the exact mass-balance relationships should allow NTP predictions to match RNA predictions; hence the reduced R^2 can be ascribed to assay noise. For example, in Figure 4C, the overall NTP depletion pattern closely matches expectations despite noise in the experimental data. In contrast, Figure 4D shows apparent near-complete CTP depletion at late reaction times, which contradicts the observed RNA yield (lower than the expected 11 gL^{-1} based on the full consumption of the limiting NTP). This discrepancy likely arises from sample dilution or assay calibration errors rather than true variability in the IVT reaction or model inaccuracy. Similarly, in Figure 4B, C and D, the initial NTP concentrations vary between 8–11 mM (versus the nominal 10 mM), causing an apparent shift in the depletion slope that would vanish if the experimental assay data matched the model assumptions. Although stable enzyme:NTP complexes could theoretically hold up NTPs, their concentrations are expected to be negligible.

We also applied the H–H approach to infer the consumption of individual NTPs across the four template–buffer conditions, using only proton balance and buffer equilibria. As shown in Figure 4E-H, the H–H model predicted NTP consumption with an overall $R^2 > 0.76$ relative to offline AEX–HPLC measurements, without fitting parameters or Kalman filter updates. The lowest average R^2 (0.64) was obtained for eGFP mRNA production in TRIS buffer, while the highest average R^2 (0.91) was observed for eGFP mRNA production in HEPES buffer.

Overall, the kinetic model embedded in the UKF, and to a lesser extent the H-H model, capture the expected NTP consumption patterns and rates, and the modest drop in R^2 highlights the limits of offline assays. A full summary of the UKF's and H-H model's fit to the NTP data, including RMSE values, is provided in Table 1.

View Article Online
DOI: 10.1039/D5DD00417A



View Article Online
DOI: 10.1039/D5DD00417A

Figure 4: **Simulated versus measured NTP depletion in IVT reactions.** Time-courses from 0–120 minutes after initiating IVT (template DNA addition) are shown for four buffer–template combinations. Panels A-D show the *UKF model* (solid) versus offline AEX-HPLC measurements (dashed). Panels E-H show the *H-H model* (solid) versus the same experimental data. Colors and markers distinguish nucleotides: ATP (blue circles), UTP (orange squares), GTP (green triangles), and CTP (pink diamonds). Each data point represents a single AEX-HPLC measurement ($n = 1$). Inset tables report per-nucleotide goodness-of-fit (coefficient of determination, R^2, and root-mean-square error, RMSE) and the panel average.

Table 1: **RMSE and R^2 values for NTP and RNA predictions.** Comparison of UKF and H–H predictions against offline assays under four template and buffer conditions. *RMSE units: RNA in* gL^{-1}*; NTP in mM.*

	HEPES + eGFP		HEPES + CSP		TRIS + eGFP		TRIS + CSP	
Analyte	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
RNA (H–H)	1.054	0.890	1.120	0.914	0.732	0.964	0.556	0.966
RNA (UKF)	0.890	0.911	1.152	0.909	0.944	0.915	0.396	0.983
NTP (H-H)	0.778	0.909	1.949	0.696	1.694	0.641	1.231	0.850
NTP (UKF)	0.743	0.898	1.454	0.820	1.463	0.799	1.231	0.850

View Article Online
DOI: 10.1039/D5DD00417A

3.5 Real-time quantification of IVT and buffer species by the kinetic-UKF and H-H soft sensor

The UKF updates the kinetic model with frequent pH measurements (e.g. every 30 seconds). Between updates the kinetic model-based soft sensor is propagated to generate dynamic estimates of all IVT species at 25 millisecond intervals. While RNA yield and NTP concentrations are validated against independent experimental measurements, the remaining species should be interpreted as model-inferred quantities, whose accuracy depends on the validity of the underlying model structure and parameters. The complete kinetic model can compute ≈40 IVT species, whereas the H-H model can account for ≈15 of these (Table S3). By solving the time-dependent mass balances of the ≈40 IVT species at 25 millisecond intervals, the model generates approximately 1,600 concentration estimates per second. For a standard 2 hour IVT reaction this gives 11.5 million IVT and buffer species concentration estimates.

Such rapid and abundant measurements are otherwise not possible for these IVT species. This enables unique real-time, model-aided insights into the chemical speciation and progression of the IVT reaction. As an example, Figure 5 shows the time-course concentrations of pyrophosphate (PPi), orthophosphate (Pi), and magnesium–nucleotide complexes (Mg^{2+}:ATP, Mg^{2+}:CTP, Mg^{2+}:GTP, Mg^{2+}:UTP), as predicted by the kinetic model and dynamically updated using the UKF. The model predicts low and progressively decreasing concentrations of PPi, consistent with its enzymatic hydrolysis to Pi by pyrophosphatase, in accordance with the IVT experimental setup. As a result, Pi concentrations (in the 0-70 mM range) increase over time, as expected. The concentrations of Mg:NTP complexes remain low (in the μM range) and decline gradually as the four NTPs are consumed through incorporation into the growing RNA chain. The time-course concentration changes of the free and total H^+ alongside buffering species (namely, the protonated (acid) and deprotonated (base) forms of TRIS, HEPES, and acetate), predicted by both the H-H and UKF-embedded kinetic models are shown in Figures S8 and S9. As expected, the concentrations of the acid forms increase over time, while those of the base forms decrease, due to proton release during the IVT reaction. Among the conditions tested, the HEPES-buffered reactions (blue and green curves) exhibit the smallest variation in free H^+, indicating that HEPES provides stronger pH buffering compared to TRIS under these IVT conditions. This aligns with the pKa of HEPES

View Article Online
DOI: 10.1039/D5DD00417A

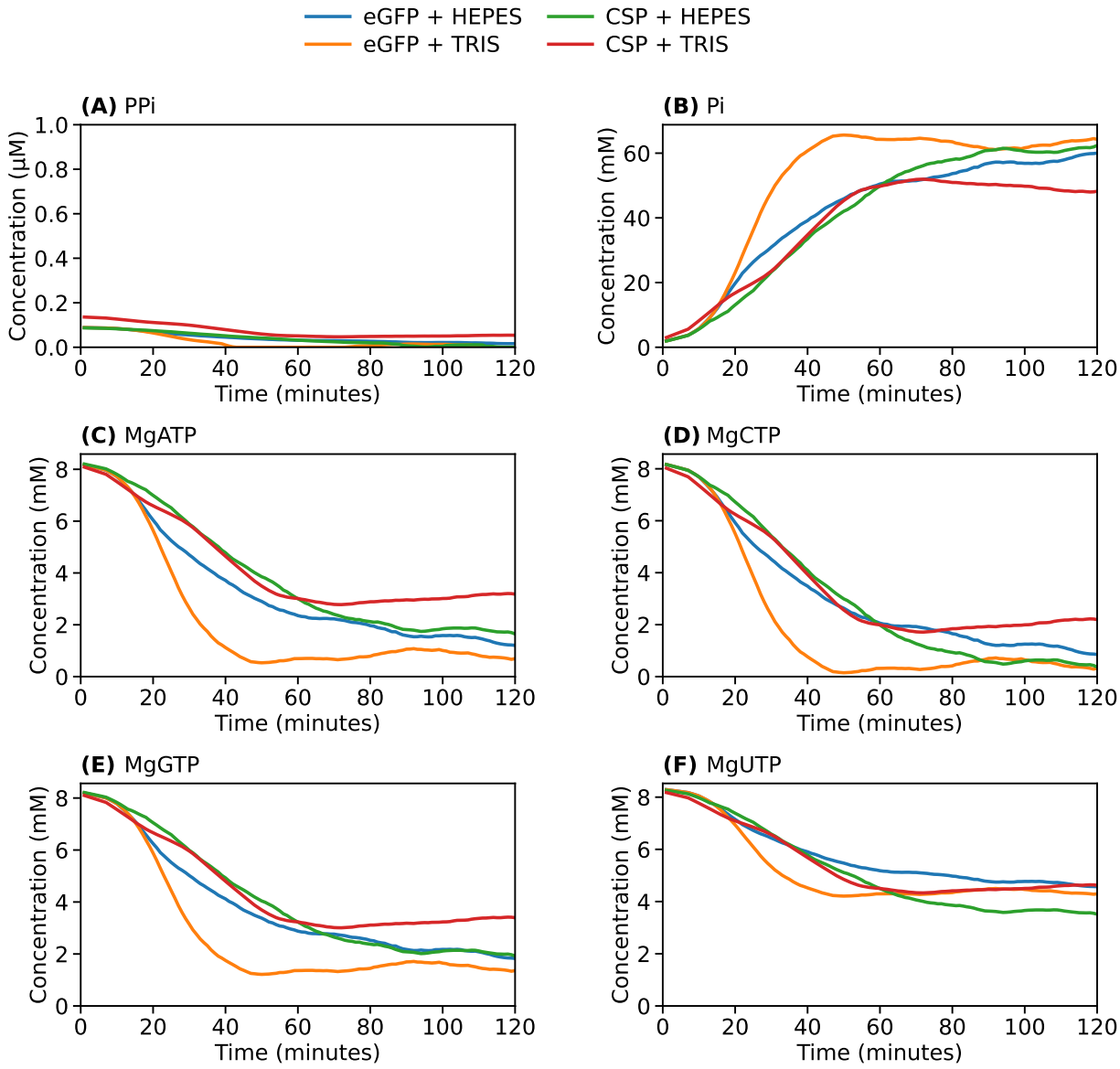


Figure 5: **Time-course profiles of Mg–nucleotide complexes, inorganic phosphate, and pyrophosphate predicted by a UKF-updated kinetic model across four IVT reactions varying template and buffer.** Blue = eGFP in HEPES, Orange = eGFP in TRIS, Green = CSP in HEPES, Red = CSP in TRIS. (A) PP_i from NTP polymerization, subsequently hydrolyzed by pyrophosphatase. (B) P_i released from PP_i. (C–F) MgATP, MgCTP, MgGTP, and MgUTP—the active Mg–NTPs consumed during transcription. Each plot spans 120 minutes and shows effects of template (eGFP vs. CSP) and buffer (HEPES vs. TRIS) on substrate depletion and by-product formation.

View Article Online
DOI: 10.1039/D5DD00417A

(7.3 at 37 °C) being closer to the IVT operating pH range (5.5–7.0) compared to that of TRIS (7.8 at 37 °C). The soft sensor, which operates by integrating the kinetic model within the UKF framework, supports real-time monitoring of numerous chemical species during the IVT reaction, providing mechanistic insight to aid the development, real-time optimization, and automation of the IVT process.

4 Discussion

This study demonstrates that a simple pH signal, when interpreted through a soft sensing framework, provides abundant and quantitative real-time insight into the dynamics of the IVT reaction. By interpreting the pH trajectory through either a mechanistic kinetic model coupled with a UKF or a H-H framework, it is possible to infer key IVT reaction states such as RNA yield, NTP depletion, and intermediate species concentrations. Although pH lacks the molecular specificity of direct assays like HPLC or spectrophotometry, it is low-cost, requiring minimal calibration, non-invasive, and already widely monitored in bioreactors. The mechanism underlying the pH decrease during IVT, caused by proton release, is well understood (cf. Figure S1) (*28, 26, 29, 37*). When combined with appropriate modeling, pH becomes a practical tool for real-time monitoring of IVT processes. This represents a shift from sensor-rich PAT strategies toward information-rich modeling of simple measurements, where reaction stoichiometry and physicochemical coupling are leveraged to extract latent process information.

The observed pH trajectories closely mirrored the reaction's progression across all four tested conditions. In each case, a rapid initial pH drop (Figure 2) corresponded to high transcriptional activity, RNA yield increase (Figure 3), and NTP consumption (Figure 4). The extent and shape of the pH curve were governed by both buffer capacity and RNA yield. For example, the eGFP-TRIS reaction showed the steepest pH drop, driven by a higher RNA yield ($\sim$ 11 gL^{-1}) and the use of TRIS buffer ($pK_a \approx 7.8$ at 37 °C), which is farther from the reaction pH ($\sim$ 5.5–7) than HEPES ($pK_a \approx 7.3$ at 37 °C). In contrast, CSP–TRIS yielded only $\sim$ 8 gL^{-1} RNA, resulting in a more modest pH decline despite the same buffer system. Meanwhile, both HEPES-buffered reactions (CSP–HEPES and eGFP–HEPES) achieved $\sim$ 11 gL^{-1} yields and similar final pH values. The initial pH of the IVT reactions was lower than that of the buffer, and this difference can be

View Article Online
DOI: 10.1039/D5DD00417A

attributed to the addition of NTPs to the reaction (*37*). Importantly, CGE confirmed that the mRNA products were of high integrity across all conditions (see Figure S7). These results support the use of pH as a robust indicator of IVT productivity, provided that buffer conditions are well characterized.

We developed a soft sensor that integrates pH measurements into a kinetic IVT model embedded within a UKF. The kinetic model continuously generates model predictions at 25 millisecond intervals and compares them to experimental pH measurements acquired at every 30 seconds, therefore state updates occur at a frequency of 30 seconds. This comparison allows the UKF to update the model state periodically, correcting for noise, uncertainty, and process variability. As a result, the soft sensor provides more accurate and robust live estimates of reaction progress. This soft sensor generates approximately 1,600 model predictions of IVT and buffer species concentrations; equivalent to 40 sets of time-resolved predictions per second. Importantly, the kinetic IVT model uses fixed kinetic parameters that were calibrated against experimental data prior to runtime; only state variables are updated by the UKF during operation. A separate set of experimental data (not used for model fitting) was used to validate both the IVT model (embedded into the UKF) and the semi-empirical H–H model, and the R^2 and RMSE values reported in this work were calculated against this separate validation dataset. A sensitivity analysis for the underlying kinetic IVT model parameters is provided in (*20*).

The UKF-based soft sensor achieved strong agreement with offline RNA measurements ($R^2 > 0.90$), demonstrating that pH input can be effectively used to infer RNA yield in real-time. The soft sensor provides a more continuous and potentially more accurate readout than intermittent offline assays, which are subject to manual sampling errors, dilution errors, and instrument calibration variability. By eliminating the need for repeated manual sampling, the soft sensor reduces the risk of such errors and provides high-frequency monitoring that captures transient process dynamics that would be missed by periodic offline measurements. Similarly, NTP depletion profiles predicted by the UKF showed a robust correlation with offline AEX-HPLC data (overall $R^2 \approx 0.84$), although deviations were more pronounced due to greater experimental noise in NTP assays (Figure 4). The H–H model also reproduced the overall trends in NTP concentration profiles, albeit with lower accuracy (overall $R^2 \approx 0.76$) relative to offline AEX-HPLC measurements. Nevertheless, this performance can be considered satisfactory given the model's simplicity, the absence of parameter fitting to experimental data, and the lack of state updates via the UKF. Notably, model

View Article Online
DOI: 10.1039/D5DD00417A

predictions reflected the known nucleotide composition of each template (Table S2): NTPs with higher fractional abundance in the RNA were consumed more rapidly, and the onset of substrate limitation was clearly mirrored in the flattening of the pH trajectory.

While the overall trends matched well, some discrepancies were observed. Initial NTP concentrations measured by AEX-HPLC varied between 8–11 mM despite nominally starting at 10 mM, suggesting pipetting, sampling or AEX-HPLC calibration errors. In CSP–TRIS, for example, offline data suggested near-complete CTP depletion, which was inconsistent with the final RNA yield. Variability in the experimental data (e.g. fluctuations in the AEX-HPLC measurements during the first 30-50 minutes in Figures 4C&D) may arise from several sources, including pipetting errors during IVT assembly, sampling inaccuracies (e.g., non-representative aliquots due to local heterogeneities in the reaction mixture), dilution errors during sample preparation for AEX-HPLC analysis, and calibration or operational issues related to the AEX-HPLC instrument. Overall, these deviations were attributed to experimental noise rather than model inaccuracy, further supporting the reliability of the soft sensor in noisy environments. The soft sensor thus provides a more continuous and potentially more reliable readout than intermittent offline assays, especially in the context of noisy or resource-intensive analytical workflows.

Beyond RNA and NTP quantification, the underlying kinetic IVT model can track the time-dependent concentrations of ≈ 40 IVT species (see Table S3). This results in around 1,600 concentration estimates per second and over 11.5 million model predictions across a 2-hour IVT reaction. This provides a rich dataset for IVT optimization and mechanistic insight into buffering dynamics, proton flux, and enzyme performance. For example, as concentrations and charge states evolve during the reaction, the IVT model–based soft sensor can identify which species contribute to buffering the released H^+. For instance, Mg^{2+}:NTP complexes ($pK_a \approx 6.5$) increasingly bind protons as the pH drops below 6.5 (*38*). However, by the time such low pH values are reached, Mg^{2+}:NTP concentrations have also declined (as NTPs are consumed to produce RNA), therefore there will be fewer Mg^{2+}:NTP complexes present to bind protons. The concentration of a subset of IVT species was plotted over time (cf. Figures 5, S8 and S9). These predictions offer real-time insight into the evolving chemical environment of the IVT reaction, useful for optimizing buffer formulations, magnesium usage, and enzyme loading. For example, the model correctly predicted near-complete degradation of PPi into Pi by pyrophosphatase, as well as micromolar-level depletion

View Article Online
DOI: 10.1039/D5DD00417A

of Mg:NTP complexes in line with nucleotide consumption (see Figure 5).

Importantly, buffer dynamics were also captured: HEPES-buffered reactions exhibited the smallest variation in free H^+ concentration, consistent with its p$K_a \approx 7.3$ at 37 °C being closer to the IVT operating pH range than that of TRIS (p$K_a \approx 7.8$ at 37 °C). These mechanistic insights highlight the soft sensor's utility not just for monitoring, but also for IVT process development and design space exploration.

It is important, however, to distinguish between directly validated outputs and model-inferred quantities: RNA yield was validated using orthogonal analytical methods (UV-Vis spectroscopy and fluorometry) and individual NTP concentrations using AEX-HPLC, whereas the remaining IVT and buffer species are inferred through the mechanistic framework, which relies on mass balances, equilibrium constants, and fixed kinetic parameters (Table S3). These inferred states are therefore contingent on the validity of the model structure and parameterization, and should not be interpreted as directly measured quantities. The associated uncertainty arises from model assumptions, parameter uncertainty, and potential unmodeled effects (e.g., side reactions or deviations from equilibrium); nonetheless, inference through literature-reported equilibrium constants represents the most accurate quantification currently achievable for these species, which cannot be directly measured in real-time during IVT reactions. Future work could validate additional species through orthogonal assays, complementary sensors or targeted experimental measurements, and further reduce uncertainty through parameter estimation.

The lower R^2 values observed for NTP predictions (compared to RNA yield) are largely attributable to higher experimental variability in the offline AEX-HPLC assay, as evidenced by the standard deviations of these measurements, rather than systematic model inaccuracy. On a related note, pH is not coupled directly to each species in the IVT model, therefore there are observability limitations. To minimize pH drift and maintain optimal enzyme activity, the IVT reaction could be initiated at a higher pH (e.g., pH 8–8.5), allowing the pH to decrease *into* the pK_a range of HEPES or TRIS during the reaction and thereby making more effective use of the buffer's capacity (*39, 37*). This would also provide a larger measurable pH window for the soft sensor. Additionally, the NTP counterion formulation may influence the sensor's performance: the sodium-salt NTPs used in this work do not contribute to buffering, whereas TRIS-salt NTPs would add buffering capacity, potentially reducing the initial pH dip and the magnitude of the subsequent pH signal.

View Article Online
DOI: 10.1039/D5DD00417A

Both the UKF and H–H models demonstrated robustness against measurement noise and offline assay variability. The UKF's ability to repeatedly reconcile model predictions with frequent pH measurements allows it to detect and correct for systematic deviations over time.

The H–H approach, while simpler, also achieved RNA yield predictions with $R^2 > 0.90$, and an overall NTP consumption prediction with $R^2 > 0.76$, despite using no kinetic parameters or experimental fitting. Accounting for ≈15 IVT species (Table S3), it provides a rapid, low-complexity benchmark or internal reference, particularly valuable when high-precision pH measurements are available and kinetic parameterization is not feasible. Consequently, the H-H quantification could serve to validate the kinetic-UKF model outputs within a model-based predictive control strategy, as it is independent of kinetic parameterization. Future improvements could focus on refining the process noise covariance to further improve UKF performance, particularly during early reaction phases. Additionally, while the current kinetic model reflects standard T7 polymerase behavior, adapting it to other polymerases or buffer systems might require reparameterization and revalidation.

This pH-based soft sensor is expected to enable real-time monitoring of RNA yield, NTP consumption, and buffering dynamics across all commonly used IVT buffers. This is supported by H-H modeling results (Figure S3), which shows that IVT buffers such as HEPES, TRIS–acetate, TRIS–HCl, and HEPES titrated with NaOH exhibit limited buffering capacity. Notably, even at elevated buffer concentrations, most likely suboptimal for IVT performance, a measurable pH drop should persist (Figure S5), indicating that the soft sensing approach remains applicable under a broad range of conditions. However, the sensor's sensitivity depends on the formulation producing a sufficient pH signal; formulations with substantially higher buffer concentrations or TRIS-salt NTPs could reduce this signal and should be validated experimentally. The H-H predictions in Figure S5 are theoretical and may not capture all interactions present in an assembled IVT reaction.

This soft sensing framework has direct implications for real-time bioprocess control. Because RNA yield and NTP depletion can be inferred continuously with low (25 millisecond) latency, it is feasible to implement closed-loop control strategies in which pH signals trigger automated feeding of NTPs, enzymes, or buffer, preventing premature substrate depletion or suboptimal reaction conditions. If deviations are detected from current measurements, feedback control can be applied to adjust and optimize the process in real time. Moreover, the kinetic model can forecast future trajectories of KPIs (e.g., mRNA yield and NTP concentrations over the next 5–10 minutes). If the

View Article Online
DOI: 10.1039/D5DD00417A

model predicts a potential decrease in RNA yield or other KPI deviation, corrective actions can be implemented through feed-forward control (e.g., model-predictive control) to prevent deviations before they occur. Such digital twin-based control approaches could be vital for maximizing yields and productivity while minimizing reagent costs and reaction times, especially in a continuous manufacturing setting.

The H-H model effectively represents a reduced-complexity alternative to the full mechanistic IVT-UKF model. Despite tracking only ≈15 species (compared to ≈40 for the kinetic model), the H-H model achieved comparable RNA yield predictions ($R^2 > 0.90$) without kinetic parameterization. This demonstrates that a simplified proton-balance approach can serve as a practical monitoring tool when kinetic parameterization is not feasible. However, the mechanistic IVT-UKF model provides additional value by explicitly representing reaction species and equilibria, enabling estimation of internal states (e.g., Mg:NTP complexes, PPi, Pi) and forecasting future reaction behavior—capabilities that are essential for advanced monitoring and potential model-based control. Additionally, the minimal, non-intrusive sampling requirements of the pH-driven soft sensor make it attractive for scale-up, as repeated offline assays can be resource-intensive. Furthermore, the approach aligns with the principles of QbDD, enabling *in silico* scenario analysis, real-time decision-making, and advanced process automation.

Finally, because proton release from the 3′–OH of the ribose is a universal feature of IVT for mRNA, self-amplifying RNA (saRNA), and circular RNA (circRNA) (*4, 40*), and IVT buffers have limited buffering capacity (cf. Figure S3), this pH-based soft sensing framework offers a broadly applicable tool for monitoring and optimizing the manufacturing of any mRNA, saRNA, and circRNA vaccines and therapeutics. It can also be extended to monitor DNA polymerization or other bioprocesses in which measurable pH changes correlate with substrate depletion, product formation, or reaction progress—such as enzymatic conversions or metabolic pathways in cell-free systems.

5 Conclusion

This work demonstrates that pH monitoring, implemented here with an *in-line* fiber-optic probe and coupled to either a detailed kinetic/UKF model or a H-H correlation, functions as an effective

View Article Online
DOI: 10.1039/D5DD00417A

soft sensing strategy for IVT reactions. The kinetic-UKF model achieved R^2 = 0.95 for mRNA yield and R^2 = 0.84 for per-NTP depletion, while the H-H model achieved R^2 = 0.93 for RNA yield and R^2 = 0.76 for NTPs, both validated against a separate set of offline measurements not used for model fitting.

Since pH is routinely logged in bioreactors, this soft sensor has strong potential as a widely applicable PAT, offering a convenient, non-invasive, single-variable route to continuously monitor KPIs in real-time, while reducing the need for offline assays. The pH-driven soft sensor was implemented here in a batch IVT reaction; however, in principle, it can also be deployed in continuous flow IVT reactors, where two or more pH sensors would feed into the models to monitor reaction progression along the reactor.

Future work will evaluate the suitability of this soft sensor across additional IVT buffers (e.g., TRIS, TRIS base, HEPES-NaOH), varied buffer concentrations, NTP counterion formulations (e.g., TRIS-salt NTPs), varying starting pH values, and RNA formats including saRNA and circRNA. H-H simulations could be used to systematically map the sensor's operating envelope across this formulation space, with targeted experimental validation of representative conditions. The pH measurement hardware will also be expanded to include microelectrode probes and pH flow cells, which may offer greater robustness for longer-duration and larger-scale continuous production runs. Overall, pH-driven soft sensing converts an already-monitored variable into actionable insight, enabling real-time control to improve throughput, quality and reproducibility in mRNA medicines manufacturing.

Acknowledgments

We thank Kesler Isoko (University College London (UCL) and University of Sheffield (UoS)), Joseph Middleton (UoS), Anna Leathard (UoS), Dr Pramuditha Mendis (UoS), and Dr Adithya Nair (UoS) for helpful discussions. Funding: We acknowledge funding from the Coalition for Epidemic Preparedness Innovations (CEPI). This study was co-funded by Innovate UK, Project Category: Small Business Research Initiative, Project Reference: 10085632. This funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript. This study was co-funded by Wellcome Leap R3 Program. This funder played no role in study

View Article Online
DOI: 10.1039/D5DD00417A

design, data collection, analysis and interpretation of data, or the writing of this manuscript. The work was supported by the School of Chemical, Materials and Biological Engineering (formerly Department of Chemical and Biological Engineering), University of Sheffield, UK.

6 Author contributions

Author contributions are listed using the CRediT (Contributor Roles Taxonomy) framework. **Conceptualization:** ZK, MA, SH, RC, CK. **Methodology:** MA, SH, RC, CK, MP, MD, MM, JC, ZK. **Software:** MA, SH, RC. **Validation:** MA, SH, RC, CK, ZK. **Formal analysis:** MA, SH, RC, CK, ZK. **Investigation:** MA, SH, RC, CK, ZK. **Data curation:** MA, SH, RC, CK. **Visualization:** MA, SH, RC, CK, MP, ZK. **Supervision:** ZK, MP, MM, MD, JC. **Project administration:** ZK. **Funding acquisition:** ZK, MD, MM, JC. **Writing—original draft:** MA, SH, RC, CK. **Writing—review & editing:** MA, SH, RC, CK, MP, MD, MM, JC, ZK.

7 Conflicts of Interest

ZK and MD are co-founders of RNA Forge Ltd. (UK company number: 16612680). All other authors declare no conflict of interest.

8 Data availability

8.1 Data

All data underlying the findings of this study are publicly available via the GitHub repository at: `https://github.com/mahdi1190/ph-ivt-soft-sensor` and via a preserved snapshot on Zenodo. The concept DOI (always resolving to the latest version) is `https://doi.org/10.5281/zenodo.19629299`, and the specific version used for this paper is archived at `https://doi.org/10.5281/zenodo.19629300` (v1.0.0). The archive contains: (i) raw input datasets, (ii) processed data required to recreate the figures and tables in the manuscript. See DATA.md for a complete inventory and licenses.

View Article Online
DOI: 10.1039/D5DD00417A

8.2 Code

The Python models, analysis code and scripts to regenerate all results are available at `https://github.com/mahdi1190/ph-ivt-soft-sensor` under the Academic and Research Use License (ARUL); see the LICENSE file in the repository for full terms. The version used for the paper is archived at Zenodo with concept DOI `https://doi.org/10.5281/zenodo.19629299` and version DOI `https://doi.org/10.5281/zenodo.19629300`, and is tagged as v1.0.0 in the repository.

8.3 Software and dependencies

A reproducible environment is specified in environment.yml and requirements.txt. Instructions to recreate the environment and reproduce the main figures are provided in the repository README.md.

8.4 Restrictions

If any dataset cannot be publicly shared (e.g., due to third-party licensing), DATA.md specifies the legal basis and provides a concrete route for qualified access (e.g., direct request to rightsholder, or controlled-access repository).

8.5 Corresponding author

Zoltan Kis

email: z.kis@sheffield.ac.uk;

webpage: `https://sheffield.ac.uk/cmbe/people/cbe-academic-staff/zoltan-kis`

View Article Online
DOI: 10.1039/D5DD00417A

References

1. U. Sahin, K. Karikó, Ö. Türeci, mRNA-based therapeutics—developing a new class of drugs. *Nature reviews Drug discovery* **13** (10), 759–780 (2014).

2. M. Janowski, A. Andrzejewska, The legacy of mRNA engineering: A lineup of pioneers for the Nobel Prize. *Molecular Therapy Nucleic Acids* **29**, 272–284 (2022).

3. E. Dolgin, How COVID unlocked the power of RNA vaccines. *Nature* **589**, 189–191 (2021).

4. S. Qin, *et al.*, mRNA-based therapeutics: powerful and versatile tools to combat diseases. *Signal transduction and targeted therapy* **7** (1), 166 (2022).

5. H.-H. Wei, L. Zheng, Z. Wang, mRNA therapeutics: New vaccination and beyond. *Fundamental Research* **3**, 749–759 (2023).

6. S. Daniel, Z. Kis, C. Kontoravdi, N. Shah, Quality by Design for enabling RNA platform production processes. *Trends in Biotechnology* **40**, 1213–1228 (2022).

7. A. Nair, K. A. Loveday, C. Kenyon, J. Qu, Z. Kis, Quality by digital design for developing platform RNA vaccine and therapeutic manufacturing processes, in *RNA Vaccines: Methods and Protocols* (Springer), pp. 339–364 (2024).

8. A. J. Geall, Z. Kis, J. B. Ulmer, Vaccines on demand, part II: future reality. *Expert Opinion on Drug Discovery* **18** (2), 119–127 (2023).

9. A. Nair, K. A. Loveday, C. Kenyon, J. Qu, Z. Kis, Quality by Digital Design for Developing Platform RNA Vaccine and Therapeutic Manufacturing Processes, in *RNA Vaccines* (Humana Press), vol. 2786 of *Methods in Molecular Biology*, pp. 339–364 (2024), doi: 10.1007/978-1-0716-3770-8_16.

10. J. Qu, *et al.*, Quality by design for mRNA platform purification based on continuous oligo-dT chromatography. *Molecular Therapy Nucleic Acids* **35** (4) (2024).

11. K. Isoko, J. L. Cordiner, Z. Kis, P. Z. Moghadam, Bioprocessing 4.0: A Pragmatic Review and Future Perspectives. *Digital Discovery* **3**, 1662–1681 (2024), doi:10.1039/D4DD00127C.

View Article Online
DOI: 10.1039/D5DD00417A

12. Merck KGaA, Darmstadt, Germany, Use of Raman Spectroscopy to Monitor *In Vitro* Transcription in mRNA Manufacturing, Technical article, Sigma-Aldrich (2024), accessed 27 Apr 2025.

13. Mettler Toledo, IVT (*In Vitro* Transcription) of mRNA Synthesis: Optimization and Scale-Up of mRNA Synthesis, Application note, Mettler Toledo (2024), https://www.mt.com/us/en/home/library/applications/automated-reactors/ivt-in-vitro-transcription-mrna-synthesis.html, accessed 27 Apr 2025.

14. A. Sarkar, G. Dong, J. Quaglia-Motta, K. Sackett, Flow-NMR as a process-monitoring tool for mRNA IVT reaction. *Journal of Pharmaceutical Sciences* **113** (4), 900–905 (2024).

15. A. Ouranidis, C. Davidopoulou, R.-K. Tashi, K. Kachrimanis, Pharma 4.0 continuous mRNA drug products manufacturing. *Pharmaceutics* **13** (9), 1371 (2021).

16. B. Y. Panah, *et al.*, Bioreactor for RNA *in vitro* transcription.

17. C. H. Bowen, *et al.*, Methods and compositions for continuous production of nucleic acids.

18. E. Nourafkan, Z. Yang, M. Maamra, Z. Kis, Advancing continuous encapsulation and purification of mRNA vaccines and therapeutics. *European Journal of Pharmaceutical Sciences* p. 107183 (2025).

19. A. Nair, Z. Kis, Bacteriophage RNA polymerases: catalysts for mRNA vaccines and therapeutics. *Frontiers in Molecular Biosciences* **11**, 1504876 (2024).

20. M. Ahmed, *et al.*, Enhancing mRNA Therapeutics Production: A Platform Technology Approach Through IVT Modeling Insights. *Results in Engineering* p. 110088 (2026), doi:10.1016/j.rineng.2026.110088, https://www.sciencedirect.com/science/article/pii/S2590123026011254.

21. J. Boman, *et al.*, Quality by design approach to improve quality and decrease cost of *in vitro* transcription of mRNA using design of experiments. *Biotechnology and Bioengineering* **121** (11), 3415–3427 (2024), doi:10.1002/bit.28806, https://doi.org/10.1002/bit.28806.

22. C. T. Martin, J. E. Coleman, Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry* **26** (10), 2690–2696 (1987).

View Article Online
DOI: 10.1039/D5DD00417A

764 23. D. Temiakov, *et al.*, Structural basis for substrate selection by T7 RNA polymerase. *Cell* **116** (3),
765 381–391 (2004).

766 24. Y. W. Yin, T. A. Steitz, The structural mechanism of translocation and helicase activity in T7
767 RNA polymerase. *Cell* **116** (3), 393–404 (2004).

768 25. A. Steitz, Thomas DNA- and RNA-dependent DNA polymerases. *Current Opinion in Struc-*
769 *tural Biology* **3** (1), 31–38 (1993), doi:10.1016/0959-440X(93)90198-T, https://www.
770 sciencedirect.com/science/article/abs/pii/0959440X9390198T.

771 26. A. Steitz, Thomas A mechanism for all polymerases. *Nature* **391** (6664), 231–232 (1998),
772 doi:10.1038/34542, https://www.nature.com/articles/34542.

773 27. T. Nakamura, Y. Zhao, Y. Yamagata, Y. Hua, W. Yang, Mechanism of the nucleotidyl-transfer
774 reaction in DNA polymerase revealed by time-resolved protein crystallography. *Biophysics*
775 *(Nagoya-shi)* **9**, 31–36 (2013), doi:10.2142/biophysics.9.31, https://www.ncbi.nlm.nih.
776 gov/pmc/articles/PMC4629682/.

777 28. C. Castro, *et al.*, Two proton transfers in the transition state for nucleotidyl transfer catalyzed by
778 RNA- and DNA-dependent RNA and DNA polymerases. *Proceedings of the National Academy*
779 *of Sciences of the United States of America* **104** (11), 4267–4272 (2007), doi:10.1073/pnas.
780 0608952104, https://www.pnas.org/doi/10.1073/pnas.0608952104.

781 29. C. Castro, *et al.*, Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nature*
782 *Structural & Molecular Biology* **16** (2), 212–218 (2009), doi:10.1038/nsmb.1540, https:
783 //www.nature.com/articles/nsmb.1540.

784 30. D. Roston, D. Demapan, Q. Cui, Extensive free-energy simulations identify water as the base
785 in nucleotide addition by DNA polymerase. *Proceedings of the National Academy of Sciences*
786 *of the United States of America* **116** (50), 25048–25056 (2019), doi:10.1073/pnas.1914613116,
787 https://www.pnas.org/doi/10.1073/pnas.1914613116.

788 31. E. N. Welbourne, *et al.*, Anion Exchange HPLC Monitoring of mRNA *In Vitro* Tran-
789 scription Reactions to Support mRNA Manufacturing Process Development. *Frontiers in*

View Article Online
DOI: 10.1039/D5DD00417A

Molecular Biosciences **11**, 1250833 (2024), doi:10.3389/fmolb.2024.1250833, https://www.frontiersin.org/articles/10.3389/fmolb.2024.1250833.

32. D. van De Berg, *et al.*, Quality by design modelling to support rapid RNA vaccine production against emerging infectious diseases. *npj Vaccines* **6** (1), 65 (2021).

33. K. O. Hill, Y. Fujii, D. C. Johnson, B. S. Kawasaki, Photosensitivity in optical fiber waveguides: Application to reflection filter fabrication. *Applied Physics Letters* **32**, 647–649 (1978).

34. J. Lin, Recent development and applications of optical and fiber-optic pH sensors. *Trends in Analytical Chemistry* **19**, 541–552 (2000).

35. A. S. Jeevarajan, S. Vani, T. D. Taylor, M. M. Anderson, Continuous pH monitoring in a perfused bioreactor system using an optical pH sensor. *Biotechnology and Bioengineering* **78**, 467–472 (2002).

36. C. Pendão, I. Silva, Optical Fiber Sensors and Sensing Networks: Overview of the Main Principles and Applications. *Sensors* **22**, 7554 (2022).

37. S. Young, J. F. Ramirez, W. H. Davis, R. Modeling and optimization of a batch process for *in vitro* RNA production. *Biotechnology and Bioengineering* **56** (2), 210–220 (1997), doi:10.1002/(SICI)1097-0290(19971020)56:2⟨210::AID-BIT10⟩3.0.CO;2-K, https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-0290(19971020)56:2<210::AID-BIT10>3.0.CO;2-K.

38. R. B. Stockbridge, R. Wolfenden, The intrinsic reactivity of ATP and the catalytic proficiencies of kinases acting on glucose, N-acetylgalactosamine, and homoserine: a thermodynamic analysis. *Journal of Biological Chemistry* **284** (34), 22747–22757 (2009).

39. J. L. Oakley, R. E. Strothkamp, A. H. Sarris, J. E. Coleman, T7 RNA polymerase: promoter structure and polymerase binding. *Biochemistry* **18** (3), 528–537 (1979).

40. E. Dolgin, Startups set off new wave of mRNA therapeutics. *Nature biotechnology* **39** (9), 1029–1031 (2021).

View Article Online
DOI: 10.1039/D5DD00417A

8 Data availability

8.1 Data

All data underlying the findings of this study are publicly available via the GitHub repository at: https://github.com/mahdi1190/ph-ivt-soft-sensor and via a preserved snapshot on Zenodo. The concept DOI (always resolving to the latest version) is https://doi.org/10.5281/zenodo.19629299, and the specific version used for this paper is archived at https://doi.org/10.5281/zenodo.19629300 (v1.0.0). The archive contains: (i) raw input datasets, (ii) processed data required to recreate the figures and tables in the manuscript. See DATA.md for a complete inventory and licenses.

8.2 Code

The Python models, analysis code and scripts to regenerate all results are available at https://github.com/mahdi1190/ph-ivt-soft-sensor under the Academic and Research Use License (ARUL); see the LICENSE file in the repository for full terms. The version used for the paper
is archived at Zenodo with concept DOI https://doi.org/10.5281/zenodo.19629299 and version DOI https://doi.org/10.5281/zenodo.19629300, and is tagged as v1.0.0 in therepository.

8.3 Software and dependencies

A reproducible environment is specified in environment.yml and requirements.txt. Instructions to recreate the environment and reproduce themain figures are provided in the repositoryREADME.md.

8.4 Restrictions

If any dataset cannot be publicly shared (e.g., due to third-party licensing), DATA.md specifies the legal basis and provides a concrete route for qualified access (e.g., direct request to rightsholder, or controlled-access repository).