Supplementary Information for:

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

Supplementary Tables

Table S1: **Sampling time points during IVT reactions.** Times at which 20 µL aliquots were withdrawn for offline NTP and RNA concentration measurement.

Sample #	Elapsed Time (minutes)
0	0
1	4
2	8
3	13
4	18
5	23
6	31
7	43
8	60
9	100
10	120

Table S2: **Nucleotide composition of CSP and eGFP mRNA.** Counts of each nucleotide in the two RNA transcripts produced using IVT.

Nucleotide	CSP Count	eGFP Count
Adenine (A)	1106	266
Guanine (G)	1061	243
Cytosine (C)	1315	279
Uracil (U)	801	142
Total	**4283**	**930**

Table S3: List of IVT species that can be theoretically accounted for by the kinetic IVT and H–H models.

Species included in the IVT kinetic model	**Species included in the H–H model**	**Species category***	**Validated****
ATP	Yes	IVT	Yes
CTP	Yes	IVT	Yes
GTP	Yes	IVT	Yes
UTP	Yes	IVT	Yes
Mg^{2+}	Yes (total not free)	IVT	No
ATP:Mg	No	IVT	No
CTP:Mg	No	IVT	No
GTP:Mg	No	IVT	No
UTP:Mg	No	IVT	No
ATP:Mg_2	No	IVT	No
CTP:Mg_2	No	IVT	No
GTP:Mg_2	No	IVT	No
UTP:Mg_2	No	IVT	No
H:ATP	Yes	IVT	No

Continued on next page. . .

Table S3: (continued)

Species included in the IVT kinetic model	Species included in the H–H model	Species category*	Validated**
H:CTP	Yes	IVT	No
H:GTP	Yes	IVT	No
H:UTP	Yes	IVT	No
H:ATP:Mg	No	IVT	No
H:CTP:Mg	No	IVT	No
H:GTP:Mg	No	IVT	No
H:UTP:Mg	No	IVT	No
Pi^{3-}	Yes	IVT	No
H:Pi	Yes	IVT	No
H_2:Pi	Yes	IVT	No
PPi	No (negligible)	IVT	No
H:PPi	No (negligible)	IVT	No
H_2:PPi	No (negligible)	IVT	No
Mg:PPi	No	IVT	No
Mg_2:PPi	No	IVT	No
Mg:H:PPi	No	IVT	No
Mg:H_2:PPi	No	IVT	No
T7RNAP	No	Initial condition	No
DNA	No	Initial condition	No
T7RNAP:DNA	No	Initial condition	No
PPase	No	IVT	No
Spd	No	IVT	No
T7RNAP:DNA:Spd	No	IVT	No
RNA	Yes	IVT	Yes
H^+	Yes	IVT & Buffer	No
HEPES	Yes	Buffer	No
H:HEPES	Yes	Buffer	No

Continued on next page. . .

Table S3: (continued)

Species included in the IVT kinetic model	Species included in the H–H model	Species category*	Validated**
TRIS	Yes	Buffer	No
H:TRIS	Yes	Buffer	No
Acetate	Yes	Buffer	No
H:Acetate	Yes	Buffer	No

* Primary role as an IVT species or a pH buffering species. Several IVT species can act as buffers since they can bind with protons.

** Validated in both the IVT and H-H models using offline assays such as: a) AEX-HPLC, b) fluorometry, c) UV spectroscopy preceded by silica spin column purification. For additional information see the Materials and Methods section.

Table S4: **UKF parameter values used in the soft sensor.**

Tuning Parameter	Value
State dimension (n_x)	36
Measurement dimension (n_z)	1
Sigma-point α	0.3
Sigma-point β	2.0
Sigma-point κ	0.0
Q_{diff} (ODE rows)	1×10^{-6}
Q_{alg} (alg. rows)	0.0
$P_{0,\text{ODE}}$	1×10^{-3}
$P_{0,\text{alg}}$	0.0
Measurement noise (R)	5×10^{-4}*

* $R_k = \sigma_R^2$ is recomputed online as a function of the current predicted pH:

$$\sigma_R = 0.01 + 0.04 \, \frac{|\text{pH}_k - 7.0|}{1.5},$$

so that noise grows from ≈ 0.01 pH (mid-range) up to ≈ 0.05 pH near the 5.5–8.5 calibration bounds. Since the reaction pH drifts downward over time, this state-dependent R_k helps the UKF account for measurement nonlinearity.

Supplementary Figures

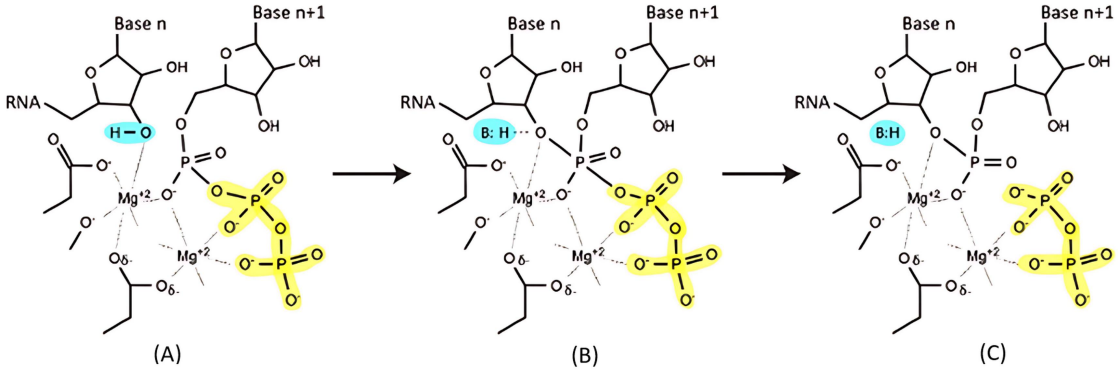

Figure S1: **Mechanism for a single NTP addition step during *in vitro* transcription of mRNA.** (**A**) Pre-catalytic complex with the 3′-hydroxyl protonated and coordinated to Mg^{2+}. (**B**) Deprotonation generates the 3′-O^- nucleophile, which attacks the α-phosphate of the incoming NTP. (**C**) Collapse of the pentacoordinate intermediate yields a new phosphodiester bond and releases pyrophosphate (PP_i) plus one H^+ to solution.

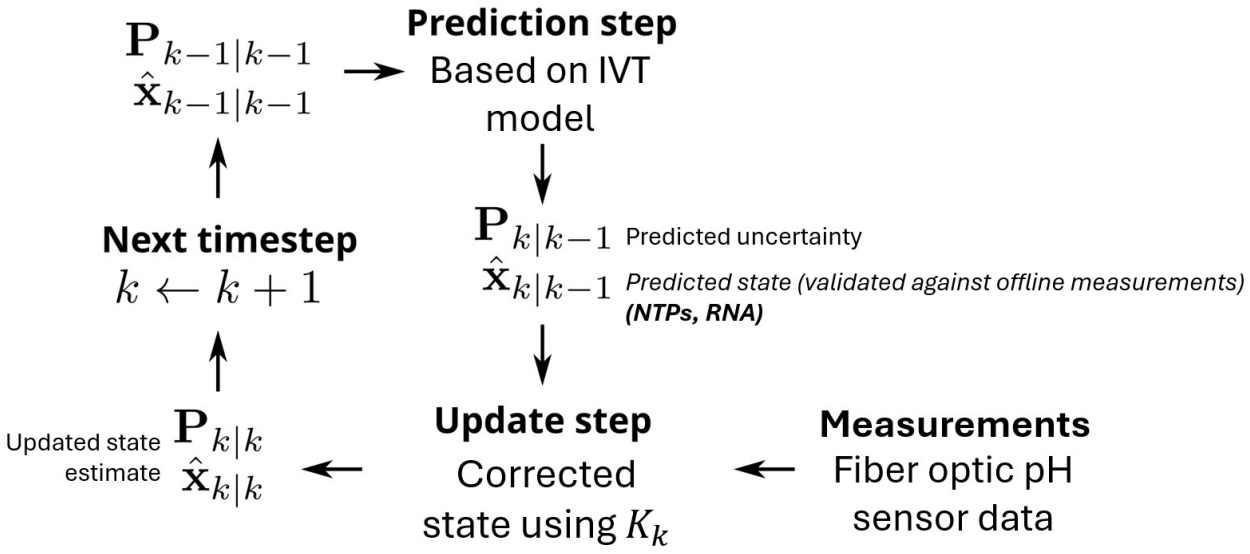

Figure S2: **UKF framework for real-time IVT state estimation.** The UKF iteratively estimates the state variables of the IVT reaction, including NTP concentrations, RNA yield, and pH. In the **prediction step**, the IVT model forecasts the next state ($\hat{x}_{k|k-1}$) and associated uncertainty ($P_{k|k-1}$). The **update step** corrects these predictions by incorporating fiber optic pH sensor measurements. The Kalman gain (K_k) adjusts the estimates based on measurement residuals, producing a refined state estimate ($\hat{x}_{k|k}$) and updated covariance ($P_{k|k}$). The process repeats iteratively for each new timestep ($k \leftarrow k + 1$), enabling continuous, real-time monitoring of IVT reaction progress.

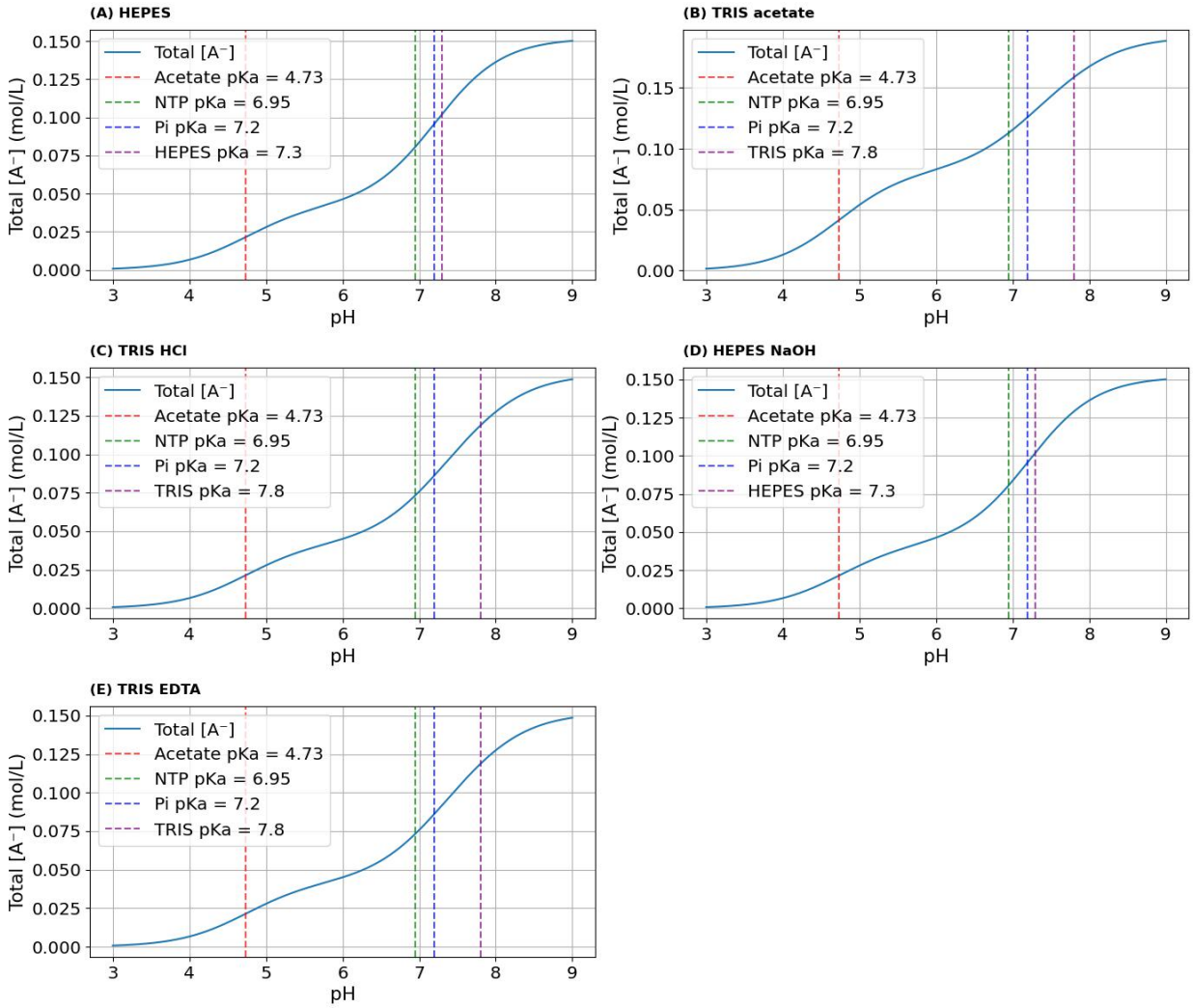

Figure S3: **Conjugate base concentrations [A$^-$] as a function of pH for five buffer systems commonly used in IVT reactions, calculated using the H-H equation.** Solid blue curves show the total [A$^-$] versus pH for (A) HEPES, (B) TRIS–acetate, (C) TRIS–HCl, (D) HEPES titrated with NaOH, and (E) TRIS–EDTA. Vertical dashed lines denote the acid dissociation constants (pKa) used in the equilibrium calculations: red=4.73 (Acetate), green=6.95 (NTP), blue=7.20 (Pi), and purple=7.30 (HEPES) or 7.80 (TRIS).

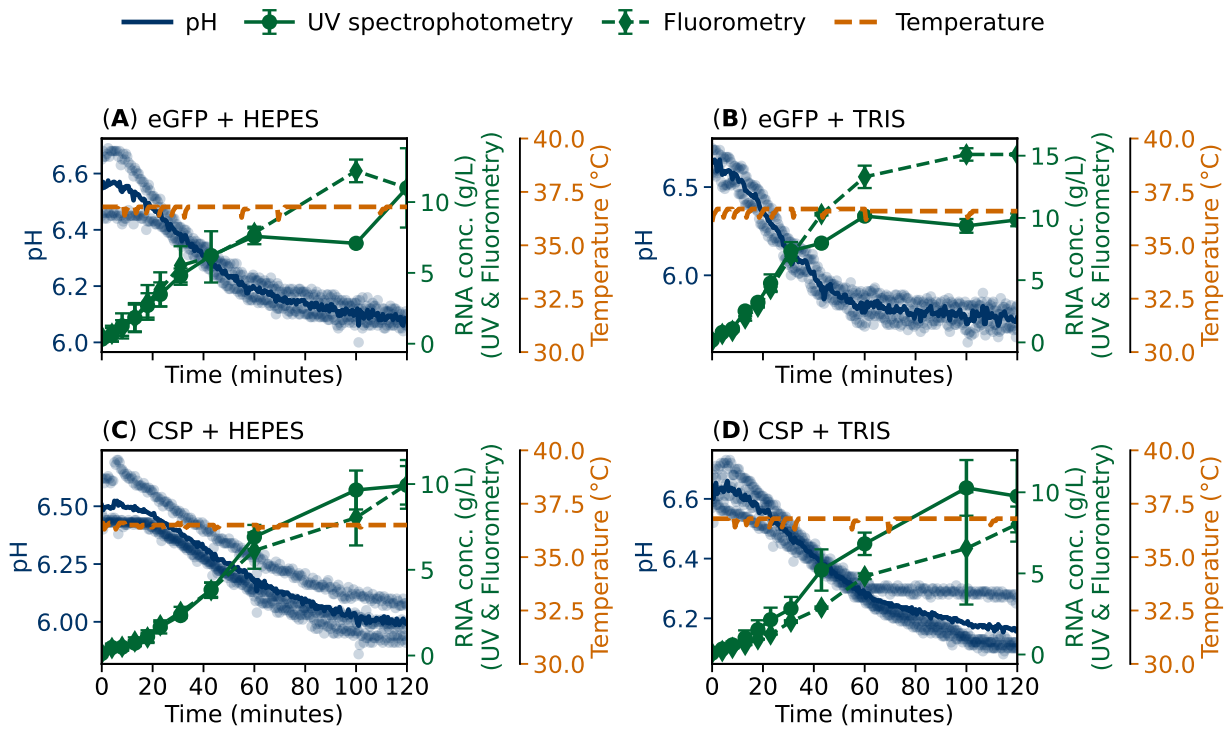

Figure S4: **Time-course profiles of average pH, RNA concentration (measured by UV spectrophotometry and fluorometry), and temperature in four *in vitro* transcription reaction conditions varying both template and buffer.** Panels: (A) eGFP mRNA in HEPES buffer, (B) eGFP mRNA in TRIS buffer, (C) CSP mRNA in HEPES buffer, (D) CSP mRNA in TRIS buffer. Solid blue lines show the average of pH values recorded over time (with faint blue circles indicating individual measurements). Solid green lines and green circles with error bars show RNA concentration in gL^{-1} measured by UV spectrophotometry using a NanoDrop™ One^{c} spectrophotometer (based on absorbance at 260 nm wavelength). Dashed green lines with green diamonds with error bars show RNA concentration in gL^{-1} measured by fluorometry using a Qubit™ 4 Fluorometer with the Qubit™ RNA XR (Extended Range) Assay Kit. Error bars represent standard deviations of duplicate (eGFP) or triplicate (CSP) measurements. Dashed orange lines track reaction temperature in °C.

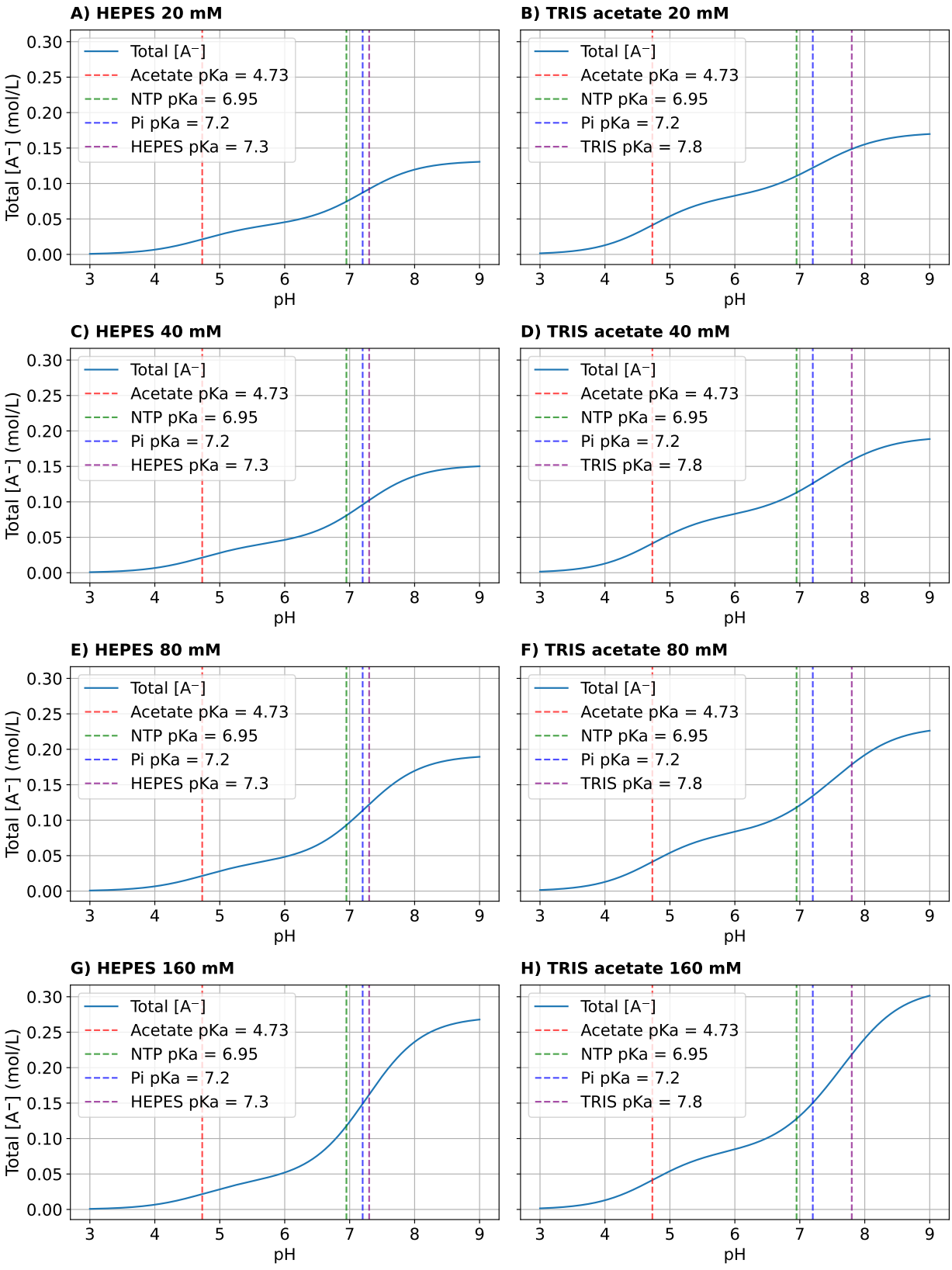
A) HEPES 20 mM
B) TRIS acetate 20 mM
C) HEPES 40 mM
D) TRIS acetate 40 mM
E) HEPES 80 mM
F) TRIS acetate 80 mM
G) HEPES 160 mM
H) TRIS acetate 160 mM
Total [A⁻]
Acetate pKa = 4.73
NTP pKa = 6.95
Pi pKa = 7.2
HEPES pKa = 7.3
TRIS pKa = 7.8
Total [A⁻] (mol/L)
pH

Figure S5: **Theoretical conjugate-base concentration profiles of HEPES and TRIS-acetate buffers, at various concentrations, as a function of pH, calculated using the semi-empirical H-H equation.** These are computational predictions based on nominal buffer compositions, not experimental measurements from assembled IVT reactions. The conjugate-base concentrations [A^-] were calculated based on the H–H correlation. Panels: (A) 20 mM HEPES, (B) 20 mM TRIS, (C) 40 mM HEPES, (D) 40 mM TRIS, (E) 80 mM HEPES, (F) 80 mM TRIS, (G) 160 mM HEPES, (H) 160 mM TRIS. Solid blue lines show the total buffer base concentration [A^-] versus pH for (left) HEPES and (right) TRIS-acetate at four concentrations: 20, 40, 80 and 160 mM (top to bottom). Vertical dashed lines mark the relevant pKa values: red=4.73 (Acetate), green=6.95 (NTP), blue=7.20 (Pi), and purple=7.30 (HEPES) or 7.80 (TRIS).

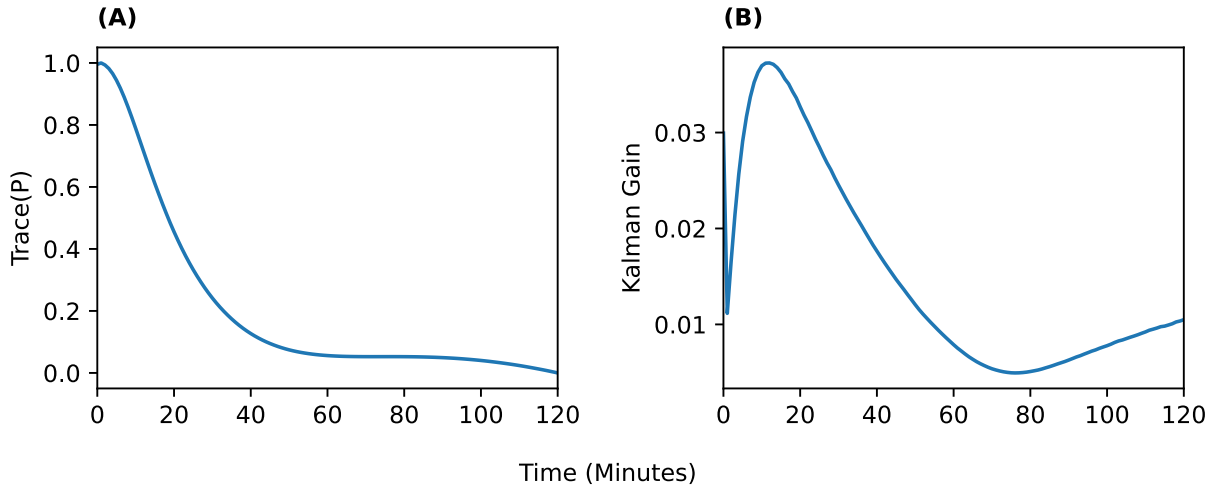

Figure S6: **Evolution of the Unscented Kalman Filter's key metrics over time.** (A) The normalized trace of the state-error covariance matrix, trace(P), falls rapidly from its large initial value toward a small steady-state level, indicating a fast reduction in estimation uncertainty as measurements are incorporated. (B) The Kalman gain K starts high, peaks during the first updates, and then decays toward its asymptotic value, reflecting the filter's progressive shift of trust from noisy measurements to its internal model dynamics.

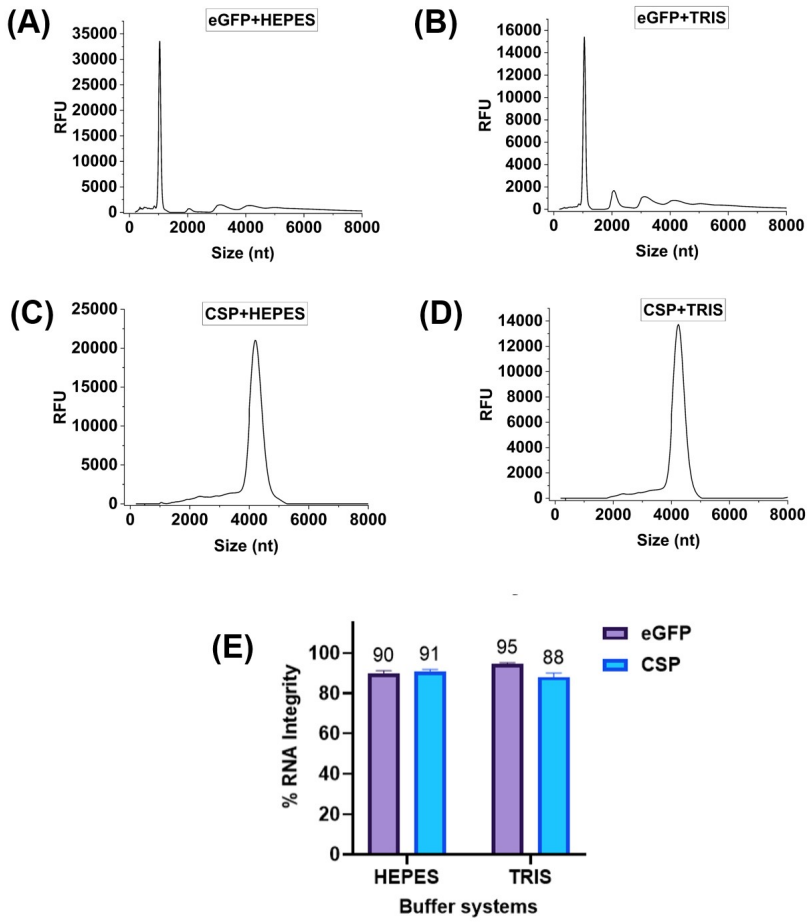

Figure S7: **Integrity (intactness) of eGFP and CSP mRNA produced in the two buffers, measured by CGE using the Agilent 5200 Fragment Analyzer instrument**. (A) Representative electropherogram of eGFP mRNA produced by IVT in HEPES buffer. (B) Representative electropherogram of eGFP mRNA produced by IVT in TRIS buffer. (C) Representative electropherogram of CSP mRNA produced by IVT in HEPES buffer. (D) Representative electropherogram of CSP mRNA produced by IVT in TRIS buffer. (E) Bar chart summarizing the mRNA integrity results (from A-D above) measured by CGE. Sky blue bars show percent mRNA integrity for eGFP, purple bars show percent mRNA integrity for CSP, measured in HEPES and TRIS buffers. Error bars show standard deviation of 2-5 technical (measurement) or experimental replicates, depending on the template-buffer combination. Average integrity values: eGFP—90% (HEPES), 95% (TRIS); CSP—91% (HEPES), 88% (TRIS).

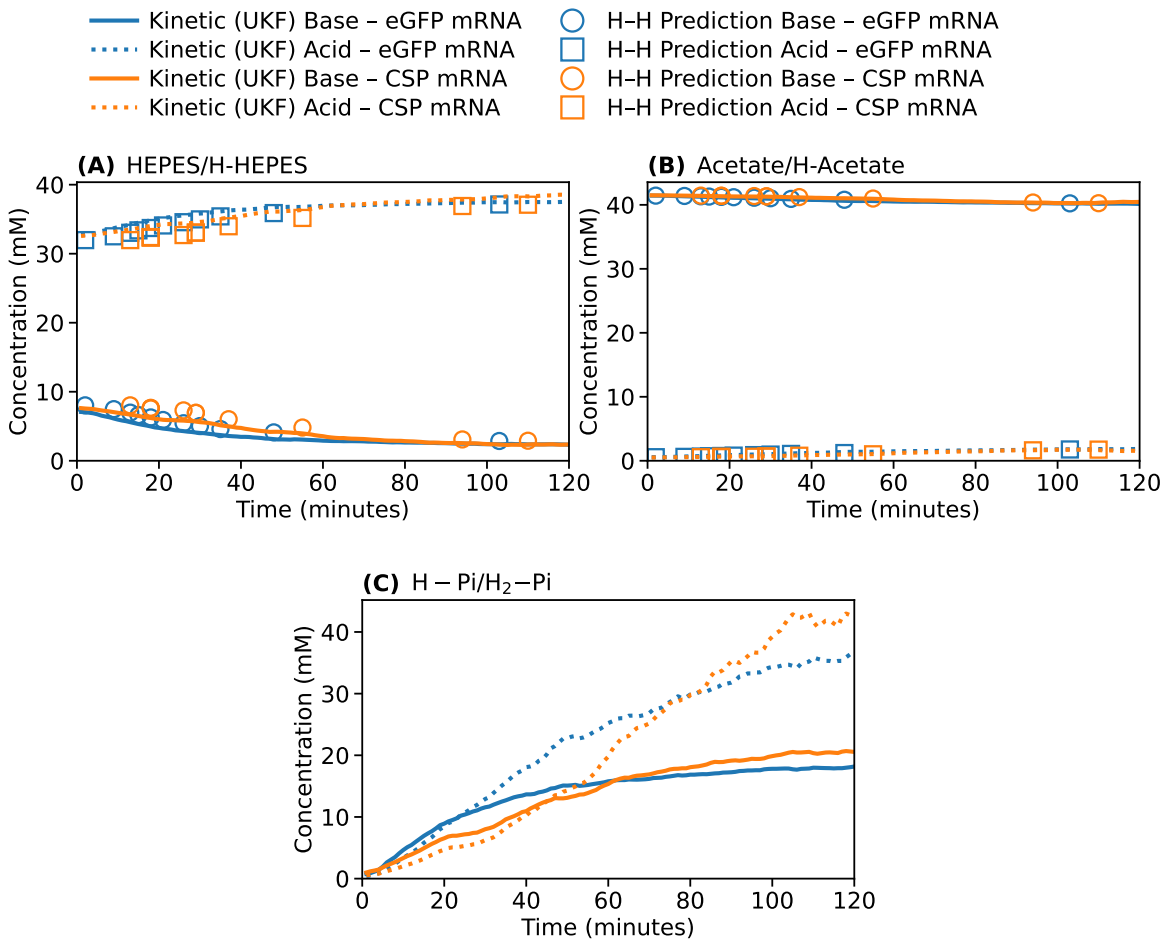

Figure S8: **Time-course profiles of buffer species in HEPES-buffered IVT reactions producing eGFP and CSP mRNA, as predicted by the H–H model and the UKF-updated kinetic model.** The buffers shown are the top three identified for minimizing pH drop. (A) HEPES buffer species (acid and conjugate base), (B) Acetate buffer species, and (C) Inorganic phosphate buffer (HPi/H_2Pi), where phosphate accumulates over the course of the reaction. Solid blue and orange lines denote the UKF-updated kinetic model predictions for HEPES conjugate base concentrations in eGFP and CSP mRNA production, respectively. Dotted blue and orange lines denote the UKF-updated kinetic model predictions for HEPES conjugate acid concentrations in eGFP and CSP mRNA production, respectively. Blue and orange circles denote the H-H model predictions for HEPES conjugate base concentrations in eGFP and CSP mRNA production, respectively. Blue and orange rectangles denote the H-H model predictions for HEPES conjugate acid concentrations in eGFP and CSP mRNA production, respectively.

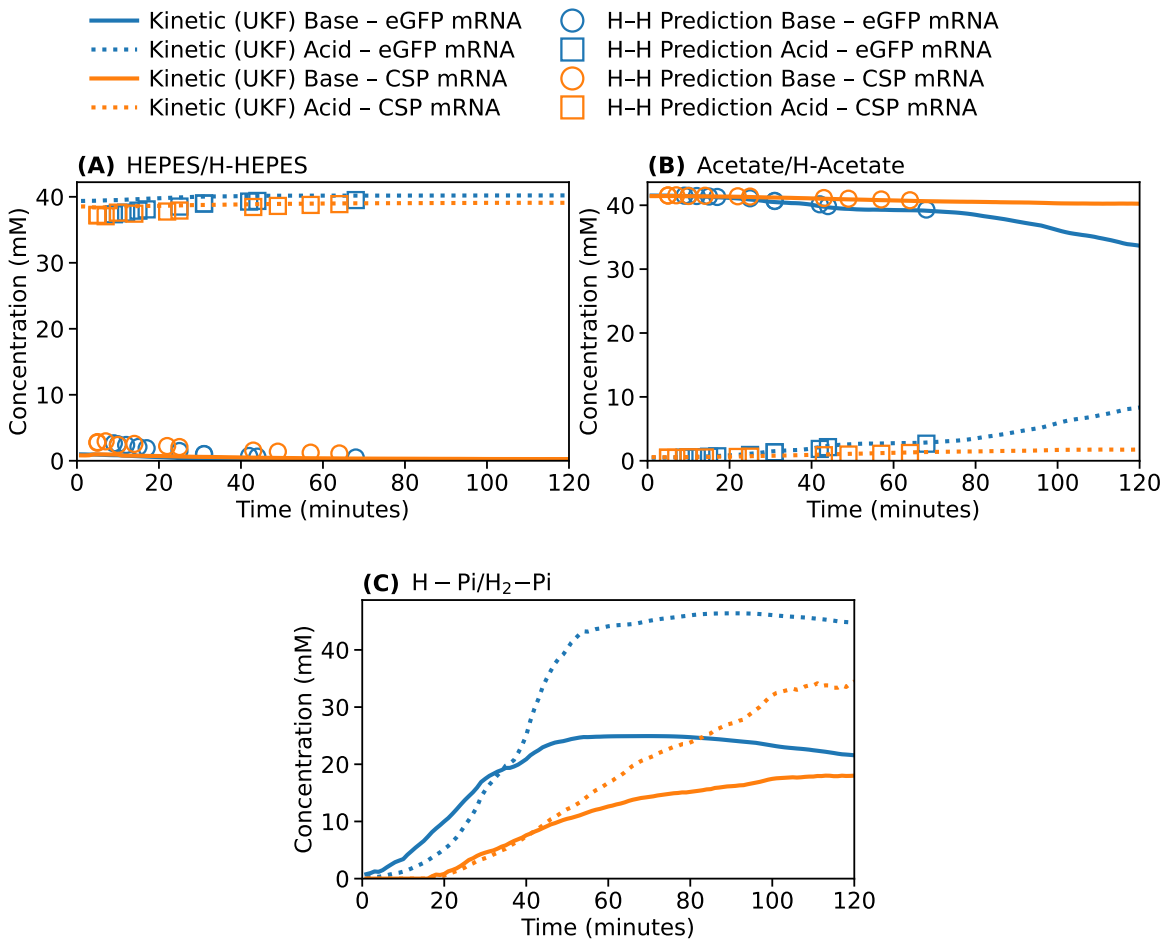

Figure S9: **Time-course profiles of buffer species in TRIS-buffered IVT reactions producing eGFP and CSP mRNA, as predicted by the H–H model and the UKF-updated kinetic model.** The buffers shown are the top three identified for minimizing pH drop. (A) TRIS buffer species (acid and conjugate base), (B) Acetate buffer species, and (C) Inorganic phosphate buffer (HPi/H_2Pi), where phosphate accumulates over the course of the reaction. Solid blue and orange lines denote the UKF-updated kinetic model predictions for TRIS conjugate base concentrations in eGFP and CSP mRNA production, respectively. Dotted blue and orange lines denote the UKF-updated kinetic model predictions for TRIS conjugate acid concentrations in eGFP and CSP mRNA production, respectively. Blue and orange circles denote the H-H model predictions for TRIS conjugate base concentrations in eGFP and CSP mRNA production, respectively. Blue and orange rectangles denote the H-H model predictions for TRIS conjugate acid concentrations in eGFP and CSP mRNA production, respectively.